ARO-D Report 71-3

PROCEEDINGS OF THE SIXTEENTH CONFERENCE ON THE DESIGN OF EXPERIMENTS IN ARMY RESEARCH DEVELOPMENT AND TESTING

Sponsored by
The Army Mathematics Steering Committee
on Behalf of

THE CHIEF OF RESEARCH AND DEVELOPMENT

Unclassified

Security Classification

DOCUMENT CONTROL DATA - R & D

(Security classification of title, body of abstract and indexing annotation must be entered when the overall report is classified)

1. ORIGINATING ACTIVITY (Corporate author)	2a. REPORT SECURITY CLASSIFICATION
U. S. Army Research Office-Durham Box CM, Duke Station Durham, North Carolina 27706	Unclassified
	2b. GROUP N/A

3. REPORT TITLE

PROCEEDINGS OF THE SIXTEENTH CONFERENCE ON THE DESIGN OF EXPERIMENTS

4. DESCRIPTIVE NOTES (Type of report and inclusive dates)

Interim Technical Report

5. AUTHOR(S) (First name, middle initial, last name)

6. REPORT DATE	7a. TOTAL NO. OF PAGES	7b. NO. OF REFS
August 1971	611	

8a. CONTRACT OR GRANT NO.	9a. ORIGINATOR'S REPORT NUMBER(S)
b. PROJECT NO.	ARO-D Report 71-3
c.	9b. OTHER REPORT NO(S) (Any other numbers that may be assigned this report)
d.	

10. DISTRIBUTION STATEMENT

Approved for public release; distribution unlimited. The findings in this report are not to be construed as an official Department of the Army position, unless so designated by other authorized documents.

11. SUPPLEMENTARY NOTES	12. SPONSORING MILITARY ACTIVITY
None	Army Mathematics Steering Committee on behalf of the Office of the Chief of Research and Development.

13. ABSTRACT

This is the technical report resulting from the Sixteenth Conference on the Design of Experiments in Army Research, Development and Testing. It contains most papers presented at that meeting. These treat various Army statistical and design problems.

14. KEY WORDS:

field testing	multivariate analysis
discrimination information	control systems
split plot designs	target acquisition analysis
prediction theory	artillery fire
improper screening	calibration drift
economic payback ratio	initiators
empirical Bayes estimator	optimal designs
prediction theory	response surfaces
vulnerability analysis	biochemical trauma
reliability	distribution free analysis
wind profiles	multiple comparisons
serum titers	complex contingency tables
regression models	truncation theory
dynamic respiratory data	nonparametric analysis

DD FORM 1 NOV 1473 REPLACES DD FORM 1473, 1 JAN 64, WHICH IS OBSOLETE FOR ARMY USE.

Unclassified

Security Classification

REPRODUCTION QUALITY NOTICE

This document is the best quality available. The copy furnished to DTIC contained pages that may have the following quality problems:

- Pages smaller or larger than normal.
- Pages with background color or light colored printing.
- Pages with small type or poor printing; and or
- Pages with continuous tone material or color photographs.

Due to various output media available these conditions may or may not cause poor legibility in the microfiche or hardcopy output you receive.

☑ If this block is checked, the copy furnished to DTIC contained pages with color printing, that when reproduced in Black and White, may change detail of the original copy.

U. S. ARMY RESEARCH OFFICE-DURHAM

Report No. 71-3
August 1971

PROCEEDINGS OF THE SIXTEENTH CONFERENCE

ON THE DESIGN OF EXPERIMENTS

Sponsored by the Army Mathematics Steering Committee

Host

The U. S. Army Logistics Management Center
Fort Lee, Virginia

21-23 October 1970

U. S. Army Research Office-Durham
Box CM, Duke Station
Durham, North Carolina

FOREWORD

One of the outstanding events at the Sixteenth Conference on the Design of Experiments in Army Research, Development and Testing was the announcement that Professor George Snedecor was selected to receive the 1970 Wilks Award. The remarks made by Professor O. Kempthorne, who accepted this award on behalf of Professor Snedecor, are recorded in these Proceedings. Other important events were the addresses made by the invited speakers. The general areas covered by these gentlemen can be gleaned from the titles of their addresses:

Minimum Discrimination Information Estimation and Application
Professor Solomon Kullback, George Washington University

Field Testing
Dr. Richard J. Kaplan, The RAND Corporation

The Analysis of Complex Contingency Table Data from General Experimental Designs and Sample Surveys
Professor Gary G. Koch, University of North Carolina

Nonparametric Analysis of Covariance
Professor Dana Quade, University of North Carolina

There is little doubt that the most valuable phases of these conferences are the technical and clinical sessions. Army scientists take advantage of the technical sessions to announce their successes in conducting various types of experiments, and sharing their findings with persons in other installations. In the clinical sessions stimulating discussions take place, and these arguments often lead to suggestions on how better to analyze the troublesome design problems that are being considered. This year there were twenty-four (24) papers for the technical sessions and five (5) clinical papers. All these papers were well received and most of them appear as articles in these Proceedings.

Colonel W. L. Tate, Commandant of the U. S. Army Logistics Management Center, offered to hold the Sixteenth Conference on the Design of Experiments in the excellent facilities at his installation. He named Mrs. Virginia W. Perry to serve as the Chairman on Local Arrangements. She and members of her committee--Messrs. C. A. Correia, J. W. Griswold, E. A. Hartley, and R. L. Launer--are to be congratulated on the way they handled the many problems that arose before and during the course of the symposium.

On behalf of the Army Mathematics Steering Committee, sponsor of these conferences, let me thank the many speakers, chairmen and panelists for all the time and effort they contributed to this affair. Without their help and the many arguments and comments supplied by those in attendance, this meeting would not have served its main purposes. At this time let me also state that much credit for the success of this conference is due to the members of my Program Committee (Robert Bechhofer, Francis Dressel, Walter D. Foster, Fred Frishman, Murray Geisler, Boyd Harshbarger, Clifford Maloney, Henry B. Mann, William H. Marlow, George Nicholson, Virginia W. Perry, and Herbert Solomon). Finally, I would like to thank Francis Dressel in particular for the smooth accomplishment of another monumental job again for this conference year.

Frank E. Grubbs
Conference Chairman

TABLE OF CONTENTS

SIXTEENTH CONFERENCE ON THE DESIGN OF EXPERIMENTS
IN ARMY RESEARCH, DEVELOPMENT AND TESTING

21-23 October 1970

Wednesday, 21 October

0830-0900 REGISTRATION - Main Lobby of Bunker Hall (Building 12500)

0900-0930 OPENING OF THE CONFERENCE - Auditorium of Bunker Hall

Virginia Perry, Chairman on Local Arrangements

WELCOME

0930-1200 GENERAL SESSION I - Auditorium

Chairman: Dr. Clifford J. Maloney, Division of Biologics Standards, National Institutes of Health, Bethesda, Md.

MINIMUM DISCRIMINATION INFORMATION ESTIMATION AND APPLICATION

Professor Solomon Kullback, Department of Statistics, George Washington University, Washington, D. C.

FIELD TESTING

Dr. Richard J. Kaplan, Management Sciences Department, The RAND Corporation, Santa Monica, California

1200-1300 LUNCH

1300-1500 TECHNICAL SESSION 1

Chairman: James Kniss, Surveillance and Reliability Laboratory, Aberdeen Research and Development Center, Aberdeen Proving Ground, Maryland

A STATISTICAL ANALYSIS OF DYNAMIC RESPIRATORY DATA

Edward N. Fiske, System Analysis Office, Edgewood Arsenal, Edgewood Arsenal, Maryland

A COMPLEX SPLIT PLOT DESIGN FOR AN EXPERIMENT INVOLVING STANO SMALL UNIT LIVE FIRE

James DeGracie, David Faulkenberry and Timm Rodgers, Litton Scientific Support Laboratory, Fort Ord, California and Spec 5 Harvey Bunce, U. S. Army Combat Developments Command Experimentation Command

THE PREDICTION OF INDIVIDUAL MILITARY PERFORMANCE FROM LABORATORY MEASURES OF PERFORMANCE IN VOLUNTEERS EXPOSED TO INCAPACITATING AGENTS

James S. Ketchum, Philip Shiner, Lorence Gutterman, and Philip K. Kysor, Clinical Medical Sciences Department, Medical Research Laboratory, Research Laboratories, Edgewood Arsenal, Maryland

1300-1500 TECHNICAL SESSION 2

Chairman: Walter Mowchan, Surveillance and Reliability Laboratory, Aberdeen Research and Development Center, Aberdeen Proving Ground, Maryland

SOME EFFECTS OF AN IMPROPER SCREENING TECHNIQUE ON THE AOQ WHEN USING CSP-1

Fred L. Abraham, US Army Ammunition & Procurement Agency, Joliet, Illinois

SOME METHODOLOGIES FOR APPLICATION OF THE ECONOMIC PAYBACK RATIO

Eugene F. Dutoit, US Army Munitions Command, Cost Analysis Division, Dover, New Jersey

EMPIRICAL BAYES ESTIMATORS FOR SOME TIME SERIES PARAMETERS

Robert L. Launer, US Army Logistics Management Center, Fort Lee, Virginia

1300-1500 CLINICAL SESSION A

Chairman: Jadrig Kurkjian, US Army Materiel Command, Washington, D. C.

Panelists:
Robert Bechhofer, Cornell University
O. P. Bruno, US Army Ballistics Research and Development Center
Boyd Harshbarger, Virginia Polytechnic Institute
H. L. Lucas, North Carolina State University
Herbert Solomon, Stanford University

A STATISTICAL APPROACH TO OPTIMIZING THE MECHANICAL BEHAVIOR OF COMPOSITE MATERIALS

Donald L. Martin, Jr., Research & Engineering, Propulsion Mech Br, APL&C, US Army Missile Command Redstone Arsenal, Ala.

PREDICTION OF SHAPE CHARGE JET CHARACTERISTICS FROM OPTICAL MEASUREMENT OF LINEAR COLLAPSE VELOCITY

Glen Rander-Pherson, Engineering Science Laboratory, Feltman Research Labs, Picatinny Arsenal, Dover, New Jersey

1500-1530 BREAK

1530-1700 TECHNICAL SESSION 3

Chairman: Ronald L. Racicot, US Army Weapons Command, Benet R&E Labs, Watervliet Arsenal, Watervliet, New York

SYSTEMS VULNERABILITY DUE TO MULTIPLE COMPONENT DRIFT AND COMPONENT FAILURE

W. W. Happ, US Army Corps of Engineers, Champaign, Illinois

TIME CONSTRAINED RELIABILITY DATA DEVELOPMENT FOR HELICOPTER RADIO EQUIPMENT IN A GROUND-BASED LABORATORY

C. E. Deckard and T. K. DeClue, Wyle Laboratories, Huntsville, Alabama

1530-1700 TECHNICAL SESSION 4

Chairman: Eugene F. Dutoit, US Army Munitions Command, Cost Analysis Division, Dover, New Jersey

CHARACTERISTIC COEFFICIENTS, PROBABILITY AND CLASSSIFICATION OF WIND PROFILES (SURFACE TO 25 Km)

Oskar M. Essenwanger, Aerophysics Branch, Physical Sciences Laboratory, Research & Engineering Directorate, US Army Missile Command, Redstone Arsenal, Alabama

IDENTIFICATION OF WORKERS IN BIOLOGICALS THROUGH SERUM TITERS BY DISCRIMINANT FUNCTION

Walter D. Foster and Marian W. Jones, Analytical Sciences Directorate, Department of the Army, Fort Detrick, Frederick, Maryland

1530-1700 CLINICAL SESSION 5

Chairman: J. P. Bruno, Surveillance and Reliability Laboratory, Aberdeen Research and Development Center, Aberdeen Proving Ground, Maryland

Panelists:
Murray Geisler, The RAND Corporation
Bernard Greenberg, University of North Carolina
Boyd Harshbarger, Virginia Polytechnic Institute
George Nicholson, University of North Carolina

A BACKWARD ELIMINATION GENERAL SIGNIFICANCE REGRESSION MODEL

Charles E. Colvin, Plans and Analysis Directorate, US Army Missile Command, Redstone Arsenal, Alabama

TEST DESIGN AND DATA REQUIREMENTS FOR OPERATIONAL FIELD TESTING OF AIRCRAFT

Chauncey F. Bell, The RAND Corporation, Washington, D. C.

1800- BANQUET

Presentation of the Samuel S. Wilks Memorial Award by Dr. Frank E. Grubbs, US Army Aberdeen Research and Development Center, Aberdeen Proving Ground, Madyland

* *

Thursday, 22 October

0830-1030 TECHNICAL SESSION 5

Chairman: Jerome Johnson, Surveillance and Reliability Laboratory, Aberdeen Research and Development Center, Aberdeen Proving Ground, Maryland

A STATISTICAL HEIRARCHICAL MODEL FOR FLIGHT TEST DATA OF A VHF/FM DISTANCE MEASURING SYSTEM (DME)

Erwin Biser and Eddie Cornelious, Avionics Laboratory, USAECOM, Fort Monmouth, New Jersey

A MULTIVARIATE STATISTICAL MODEL FOR A SEMIAUTOMATIC FLIGHT OPERATIONS CENTER (SAFOC)

Sol Berg and William Patterson, American Electronics Laboratory, Colmar, Pennsylvania, and Edwin Biser, Arthur Coppola, and Edward Hansen, Avionics Lab, USAECOM, Fort Monmouth, New Jersey

ENVIRONMENTAL CONTROL SYSTEM ANALYSIS

Francis Brandi, Tank Systems Laboratory, US Army Weapons Command, Rock Island Arsenal, Rock Island, Illinois

0830-1030 TECHNICAL SESSION 6

Chairman: Siegfried H. Lehnigk, Physical Sciences Lab, Research and Development Directorate, US Army Missile Command, Redstone Arsenal, Alabama

NEW ANALYSIS AND METHODS LEADING TO IMPROVED TARGET ACQUISITION REQUIREMENTS INVOLVING SYSTEMS, GEODETIC AND RE-ENTRY ERRORS, AND INCREASED WEAPONS EFFECTIVENESS FOR CONVENTIONAL WEAPONS (PART II)

Hans Baussus-von Luetzow, US Army Topographic Laboratories, Fort Belvoir, Virginia

A MATHEMATICAL MODEL FOR ARTILLERY FIRE ADJUSTMENT ANALYSIS

Sidney Gerard, Reliability and Maintainability Division, US Army Material Systems Analysis Agency, Aberdeen Research and Development Center, Aberdeen Proving Ground, Maryland

QUICK-REACTION STUDY OF CALIBRATION DRIFT IN RADIACMETER IM-174()

Joseph Weinstein, Systems Costs Analysis Office, US Army Electronics Command, Fort Monmouth, New Jersey

0830-1030 CLINICAL SESSION C

Chairman: Clifford Cohen, University of Georgia, Athens, Georgia

Panelists:
Robert Bechhofer, Cornell University
Murray Geisler, RAND Corporation
Bernard Greenberg, University of North Carolina
George Nicholson, University of North Carolina
Herbert Solomon, Stanford University

RELIABILITY TESTING OF WEAPON SYSTEMS

Ronald L. Racicot, Army Weapons Command, Benet R&E Laboratories, Watervliet Arsenal, Watervliet, New York

TEST PROCEDURES FOR EVALUATION OF INITIATORS TO THE EFFECT OF NUCLEAR DEVICES

Robert E. Betts and W. B. Thomas, Solid Propellant Chemistry Branch, Army Propulsion Lab, Redstone Arsenal, Alabama

1030-1100 BREAK

1100-1200 TECHNICAL SESSION 7

Chairman: Mortimer Zinn, Electronics Components Laboratory, US Army Electronics Command, Fort Monmouth, New Jersey

OPTIMAL DESIGNS WITH A TCHEBYCHEFFIAN SPLINE REGRESSION

V. N. Murty, Graduate Center, The Pennsylvania State University, Middletown, Pennsylvania

1100-1200 TECHNICAL SESSION 8

Chairman: Boyd Harshbarger, Virginia Polytechnic Institute, Blacksburg, Virginia

ORDINARY AND EMPIRICAL BAYES APPROACH TO ESTIMATION OF RELIABILITY IN THE WEIBULL LIFE TESTING MODEL

George C. Canavos, NASA, Langley Research Center, Langley, Virginia, and Chris P. Tsokos, Virginia Polytechnic Institute, Blacksburg, Virginia

1200-1300 LUNCH

1300-1500 TECHNICAL SESSION 9

Chairman: Oskar M. Essenwanger, Aerophysics Branch, Physical Sciences Laboratory, Research and Engineering Directorate, US Army Missile Command, Redstone Arsenal, Ala.

SECOND-ORDER EQUI-RADIAL DESIGNS FOR WEIGHTED REGRESSION

John Cornell, Department of Statistics, University of Florida, Gainesville, Florida

SYSTEM PARAMETER OPTIMIZATION USING RESPONSE SURFACE METHODOLOGY

Gary Barnard, Plans & Analysis Directorate, US Army Missile Command, Redstone Arsenal, Alabama

THE APPLICATION OF BIOCELLULAR NUMBERS TO THE ASSESSMENT OF BIOCHEMICAL TRAUMA IN ANIMAL SYSTEMS

George L. Lavin, Vulnerability Laboratory, BRL, ARDC, Aberdeen Proving Ground, Maryland

1300-1500 TECHNICAL SESSION 10

Chairman: Robert Eisaner, Surveillance and Reliability Laboratory, Aberdeen Research and Development Center, Aberdeen Proving Ground, Maryland

A DISTRIBUTION FREE ALTERNATIVE TO PROBIT-TYPE ANALYSIS

Bert Levy and Gerald M. Schultz, Harry Diamond Laboratories, Washington, D. C.

MULTIPLE COMPARISONS REVISITED

Clifford J. Maloney, Bethesda, Maryland

DESIGN FOR ESTIMATING THE SLOPE OF A SECOND ORDER LINEAR MODEL

Lyman Ott and W. Mendenhall, Department of Statistics, University of Florida, Gainesville, Florida

1500-1530 BREAK

1530-1630 GENERAL SESSION II

Chairman: Professor Bernard Greenberg, School of Public Health, Department of Biostatistics, University of North Carolina, Chapel Hill, North Carolina

THE ANALYSIS OF COMPLEX CONTINGENCY TABLE DATA FROM GENERAL EXPERIMENTAL DESIGNS AND SAMPLE SURVEYS

Professor Gary G. Koch, Department of Biostatistics, School of Public Health, University of North Carolina, Chapel Hill, North Carolina

* *

Friday, 23 October

0830-0915 OPEN MEETING OF THE AMSC SUBCOMMITTEE ON PROBABILITY AND STATISTICS

Chairman: Dr. Walter D. Foster, Analytical Sciences Directorate, Department of the Army, Fort Detrick, Frederick, Maryland

0915-0945 BREAK

0945-1145 GENERAL SESSION III

Chairman: Dr. Frank E. Grubbs, US Army Aberdeen Research and Development Center, Aberdeen Proving Ground, Maryland (Chairman of this Conference)

ESTIMATION IN TRUNCATED POISSON DISTRIBUTIONS WITH CONCOMITANT INTERVALS AND TRUNCATION POINTS

Professor A. Clifford Cohen, The University of Georgia, Department of Statistics, Athens, Georgia

NONPARAMETRIC ANALYSIS OF COVARIANCE

Professor Dana Quade, Department of Biostatistics,
School of Public Health, University of North Carolina,
Chapel Hill, North Carolina

* *

PROGRAM COMMITTEE

Robert Bechhofer
Francis Dressel (Secretary)
Walter D. Foster
Fred Frishman
Murray Geisler
Boyd Harshbarger
Clifford J. Maloney
Henry B. Mann
William H. Marlow
George Nicholson
Virginia Perry
Herbert Solomon

Frank E. Grubbs, Chairman

Minimum Discrimination Information Estimation and Application

S. Kullback
The George Washington University
Washington, D. C. 20006

Abstract

This paper presents in some detail the application of the minimum discrimination information theorem to the analysis of multidimensional contingency tables. It is shown that the form of the minimum discrimination information estimate as a member of an exponential family provides a regression expression for the logarithm of the estimate. Computational procedures for the evaluation of the regression parameters and the minimum discrimination information estimates are described along with the tests for the hypotheses as provided by the minimum discrimination information statistics.

0. Introduction. This paper is related to [9] and [10] in which certain basic techniques and procedures were presented for the

---

[1] Supported in part by the Air Force Office of Scientific Research, Office of Aerospace Research, United States Air Force, under Grant AFOSR-68-1513.

This paper was reproduced photographically from the author's manuscript.

analysis of multidimensional contingency tables. In this paper we shall examine the underlying theory in greater detail and present one important area of application. In particular we shall detail the close analogy of this application with multivariate regression analysis. Although the ingredients of the underlying theory were discussed in [11] it seems necessary and desirable to present these ideas here in greater detail. We also remark that a more extensive computer program than that described in [6] and [9] has been prepared by Professor Ireland of The George Washington University. This new program can handle tables of higher dimension than four-way contingency tables and also provides the values of additional useful parameters.

It should be pointed out that there are other areas of application of minimum discrimination information estimation than that considered in detail in this paper, for example, [3], [4], [5], [7], [11], [12], [13], [14]. The particular application we shall consider here can be described as fitting the observed values in the cells of a contingency table in terms of a regression based on sets of observed marginals as explanatory variables.

1. Discrimination information. To make the discussion more specific we shall present it in terms of the analysis of four-way contingency tables. All the essential features of a more general presentation appear. Let us consider the space Ω

of four-way contingency tables R × S × T × U of dimension r × s × t × u so that the generic variable is $\omega = (i,j,k,\ell)$, $i = 1,\ldots,r$, $j = 1,\ldots,s$, $k = 1,\ldots,t$, $\ell = 1,\ldots,u$. Suppose there are two probability distributions or contingency tables (we shall use these terms interchangeably) defined over the space Ω, say $p(\omega)$, $\pi(\omega)$, $\sum_{\Omega} p(\omega) = 1$, $\sum_{\Omega} \pi(\omega) = 1$. The discrimination information is defined by

$$(1.1) \quad I(p:\pi) = \sum_{\Omega} p(\omega) \ln \frac{p(\omega)}{\pi(\omega)} .$$

The basis for this definition, its properties, and relation to other definitions of information measures may be found in [11], in the Proceedings of [13] and references therein. For the particular types of application of interest here the π-distribution, $\pi(\omega)$, in the definition (1.1) according to the problem of interest may either be specified, or it may be an estimated distribution, or it may be an observed distribution. The p-distribution, $p(\omega)$, in the definition (1.1) ranges over or is a member of a family of distributions of interest.

Of the various properties of $I(p:\pi)$ we mention in particular the fact that $I(p:\pi) > 0$ and $= 0$ if and only if $p(\omega) = \pi(\omega)$.

2. Minimum discrimination information estimation. Many problems in the analysis of contingency tables may be characterized as estimating a distribution or contingency table subject to certain restraints and then comparing the estimated table with an

observed table to determine whether the observed table satisfies a null hypothesis implied by the restraints. In accordance with the principle of minimum discrimination information estimation we select that member of the family of p-distributions satisfying the restraints which minimizes the discrimination information $I(p:\pi)$ over the family of pertinent p-distributions. We denote the minimum discrimination information estimate by $p^*(\omega)$ so that

$$(2.1)\quad I(p^*:\pi) = \Sigma\, p^*(\omega)\, \ell n\, \frac{p^*(\omega)}{\pi(\omega)} = \min I(p:\pi).$$

Unless otherwise stated, the summation is over Ω which will be omitted.

In one class of problems the restraints specify some requirement external to the observed values, for example, that a set of marginals have specified values as determined by genetic or other theory [4], [5], [12], or that marginals be homogeneous [3], [14], or that the distribution satisfy certain symmetry conditions [3]. In such problems $\pi(\omega)$ is taken to be an observed contingency table, that is, $x(\omega) = x(ijk\ell) = n\pi(ijk\ell)$, where $n = \Sigma\, x(\omega)$.

In another class of problems the restraints specify that the estimated distribution or contingency table have some set of marginals which are the same as those of an observed contingency table. In such cases $\pi(\omega)$ is taken to be either the uniform distribution $\pi(ijk\ell) = 1/rstu$ or a distribution already estimated subject to restraints contained in and implied by the restraints under examination. The latter case includes the classical

hypotheses of independence, conditional independence, homogeneity, conditional homogeneity and interaction, all of which can be considered as instances of generalized independence [3], [6], [7], [8], [9], [10], [13], and will be considered in some detail in this paper.

3. Minimum discrimination information statistic. To test whether an observed contingency table satisfies the null hypothesis as represented by the minimum discrimination information estimate we compute a measure of the deviation between the observed distribution and the appropriate estimate by the minimum discrimination information statistic. For notational convenience and later computational convenience let us denote the estimated contingency table in terms of occurrences by $x^*(\omega) = np^*(\omega)$, then for the first category of problems, that is, with restraints determined by external considerations, the minimum discrimination information statistic turns out to be

$$(3.1) \quad 2I(x^*:x) = 2\Sigma\, x^*(\omega)\, \ell n\, \frac{x^*(\omega)}{x(\omega)}$$

which is asymptotically distributed as a χ^2 with appropriate degrees of freedom under the null hypothesis. For the second category of problems, that is, with the restraints implied by a set of observed marginals, or those of a generalized independence hypothesis, the m.d.i. statistic is

$$(3.2)\quad 2I(x:x^*) = 2\Sigma\, x(\omega)\, \ell n\, \frac{x(\omega)}{x^*(\omega)}$$

which is asymptotically distributed as a χ^2 with appropriate degrees of freedom under the null hypothesis.

The statistic in (3.2) is also minus twice the logarithm of the likelihood ratio statistic but this is not true for the statistic in (3.1) or in other applications [11].

4. Minimum discrimination information theorem. We now present a theorem which is the basis for the principle of minimum discrimination information estimation and its applications. We shall present it in a form related to the context of this discussion on the analysis of contingency tables.

Let us consider the space Ω mentioned in section 1 and the discrimination information introduced in (1.1). Suppose now, for example, that we have three linearly independent statistics of interest defined over the space Ω

$$(4.1)\quad T_1(\omega),\ T_2(\omega),\ T_3(\omega).$$

Let us determine the value of $p(\omega)$ which minimizes the discrimination information

$$(4.2)\quad I(p:\pi) = \Sigma\, p(\omega)\, \ell n\, \frac{p(\omega)}{\pi(\omega)}$$

over the family of p-distributions which satisfy the restraints

$$\Sigma\, T_1(\omega)\, p(\omega) = \theta_1^*$$

(4.3) $$\Sigma\, T_2(\omega)\, p(\omega) = \theta_2^*$$

$$\Sigma\, T_3(\omega)\, p(\omega) = \theta_3^*$$

where θ_1^*, θ_2^*, θ_3^* are specified values.

If $\pi(\omega)$ satisfies the restraints (4.3) then of course the minimum value of $I(p:\pi)$ is zero and the minimizing distribution is $p^*(\omega) = \pi(\omega)$. More generally, the minimum discrimination information theorem [11] states that the minimizing distribution is given by

(4.4) $$p^*(\omega) = \frac{\exp(\tau_1 T_1(\omega) + \tau_2 T_2(\omega) + \tau_3 T_3(\omega)) \cdot \pi(\omega)}{M(\tau_1, \tau_2, \tau_3)}$$

where

(4.5) $$M(\tau_1, \tau_2, \tau_3) = \Sigma \exp(\tau_1 T_1(\omega) + \tau_2 T_2(\omega) + \tau_3 T_3(\omega))\, \pi(\omega)$$

and the τ's are parameters which are in essence undetermined Lagrange multipliers whose values are defined in terms of θ_1^*, θ_2^*, θ_3^* by

$$\theta_1^* = \frac{\partial}{\partial \tau_1} \ln M(\tau_1, \tau_2, \tau_3) =$$

$$= (\Sigma \exp(\tau_1 T_1(\omega) + \tau_2 T_2(\omega) + \tau_3 T_3(\omega)) T_1(\omega) \pi(\omega)) / M(\tau_1, \tau_2, \tau_3)$$

$$(4.6)\quad \theta_2^* = \frac{\partial}{\partial \tau_2} \ln M(\tau_1, \tau_2, \tau_3) =$$

$$= (\Sigma \exp(\tau_1 T_1(\omega) + \tau_2 T_2(\omega) + \tau_3 T_3(\omega)) T_2(\omega) \pi(\omega)) / M(\tau_1, \tau_2, \tau_3)$$

$$\theta_3^* = \frac{\partial}{\partial \tau_3} \ln M(\tau_1, \tau_2, \tau_3) =$$

$$= (\Sigma \exp(\tau_1 T_1(\omega) + \tau_2 T_2(\omega) + \tau_3 T_3(\omega)) T_3(\omega) \pi(\omega)) / M(\tau_1, \tau_2, \tau_3).$$

We can now state a number of consequences of the preceding.

We note first that $p^*(\omega)$ is a member of an exponential family of distributions generated by $\pi(\omega)$ and as such has the properties of members of an exponential family. In particular $p^*(\omega) = \pi(\omega)$ for $\tau_1 = \tau_2 = \tau_3 = 0$. We may also write (4.4)

$$(4.7)\quad \ln \frac{p^*(\omega)}{\pi(\omega)} = -\ln M(\tau_1, \tau_2, \tau_3) + \tau_1 T_1(\omega) + \tau_2 T_2(\omega) + \tau_3 T_3(\omega)$$

$$= L + \tau_1 T_1(\omega) + \tau_2 T_2(\omega) + \tau_3 T_3(\omega)$$

with $L = -\ln M(\tau_1, \tau_2, \tau_3)$. The regression expression in (4.7) for $\ln(p^*(\omega)/\pi(\omega))$ with $T_1(\omega)$, $T_2(\omega)$, $T_3(\omega)$ as the explanatory variables and τ_1, τ_2, τ_3 as the regression coefficients plays an important role in the analysis we shall consider.

We note next that the minimum value of the discrimination information (4.2) is

$$(4.8)\quad I(p^* : \pi) = \tau_1 \theta_1^* + \tau_2 \theta_2^* + \tau_3 \theta_3^* - \ln M(\tau_1, \tau_2, \tau_3)$$

where the θ^*'s are defined in (4.3) and the τ's are determined to satisfy (4.6). Using the value in (4.7) it may be shown that if $p(\omega)$ is any member of the family of distributions satisfying (4.3), then

$$(4.9)\quad I(p;\pi) = I(p;p^*) + I(p^*;\pi).$$

The pythagorean property (4.9) plays an important role in the analysis of information tables.

We note thirdly relations connecting the θ^*'s, the τ's, and the covariance matrix of the $T(\omega)$'s. If we define the matrices (vectors)

$$(\underline{d\theta^*})' = (d\theta_1^*,\ d\theta_2^*,\ d\theta_3^*),\quad (\underline{d\tau})' = (d\tau_1,\ d\tau_2,\ d\tau_3)$$

then [11, p.49]

$$(4.10)\quad (\underline{d\theta^*}) = \underline{\Sigma}^*(\underline{d\tau}),\quad (\underline{d\tau}) = \underline{\Sigma}^{*-1}(\underline{d\theta^*})$$

where $\underline{\Sigma}^*$ is the covariance matrix of $T_1(\omega)$, $T_2(\omega)$, $T_3(\omega)$ for the distribution $p^*(\omega)$, that is, with

$$\sigma_{ij}^* = \Sigma(T_i(\omega)-\theta_i^*)(T_j(\omega)-\theta_j^*)p^*(\omega),\ \underline{\Sigma}^* = (\sigma_{ij}^*),\ \underline{\Sigma}^{*-1} = (\sigma^{*ij})$$

$$(4.11)\quad \frac{\partial\theta_i^*}{\partial\tau_j} = \sigma_{ij}^*\ ,\ \frac{\partial\tau_i}{\partial\theta_j^*} = \sigma^{*ij}$$

From (4.5) it is seen that $M(\tau_1,\tau_2,\tau_3)$ is the moment-generating function of $T_1(\omega)$, $T_2(\omega)$, $T_3(\omega)$ under the distribution $\pi(\omega)$, hence

the cumulant-generating function is given up to quadratic terms by

$$(4.12)\quad \ln M(\tau_1,\tau_2,\tau_3) \approx \theta_1\tau_1 + \theta_2\tau_2 + \theta_3\tau_3 + \tfrac{1}{2}\sum\sum \sigma_{ij}\,\tau_i\tau_j$$

where

$$(4.13)\quad \theta_i = \Sigma\, T_i(\omega)\,\pi(\omega),\ \sigma_{ij} = \Sigma(T_i(\omega) - \theta_i)(T_j(\omega) - \theta_j)\,\pi(\omega).$$

Thus, using (4.12) in (4.6), we get

$$(4.14)\quad \begin{aligned} \theta_1^* &\approx \theta_1 + \sum_j \sigma_{1j}\,\tau_j \\ \theta_2^* &\approx \theta_2 + \sum_j \sigma_{2j}\,\tau_j \\ \theta_3^* &\approx \theta_3 + \sum_j \sigma_{3j}\,\tau_j \end{aligned}$$

and then using (4.14) in (4.8) yields

$$(4.15)\quad 2I(p^*:\pi) \approx (\underline{\theta}^* - \underline{\theta})'\,\underline{\Sigma}^{-1}\,(\underline{\theta}^* - \underline{\theta}) \approx \underline{\tau}'\,\underline{\Sigma}\,\underline{\tau}.$$

We have used three functions $T_1(\omega)$, $T_2(\omega)$, $T_3(\omega)$ thus far in the discussion merely as a matter of convenience. We note that (4.15) holds for a set of m functions $T_i(\omega)$, $i = 1,\ldots,m$ with appropriate meanings for the matrices. Let us partition the set of m functions $T_i(\omega)$ into a set H_a say of m_a and a set H_e of the remaining $m_e = m - m_a$ functions, where the functions in the set H_a have the property that

(4.16) $\theta_i^* = \theta_i$, $i = 1,\ldots,m_0$.

We have the related partitioning of the covariance matrix of the $T_i(w)$, $i = 1,\ldots,m$

$$(4.17)\quad \underline{\Sigma} = \begin{pmatrix} \underline{\Sigma}_{00} & \underline{\Sigma}_{01} \\ \underline{\Sigma}_{10} & \underline{\Sigma}_{11} \end{pmatrix}, \quad \underline{\Sigma}_{00} = \underline{\Sigma}_{00}'$$

and the $\underline{\theta}$, $\underline{\theta}^*$, and $\underline{\tau}$ matrices

$$(4.18)\quad \underline{\theta}^{*\prime} = (\underline{\theta}_0^{*\prime}, \underline{\theta}_1^{*\prime}),\ \underline{\theta}' = (\underline{\theta}_0', \underline{\theta}_1'),\ \underline{\tau}' = (\underline{\tau}_0', \underline{\tau}_1').$$

In terms of the partitionings in (4.17) and (4.18) the relations in (4.14) may be written as

$$(4.19)\quad \begin{aligned} \underline{\theta}_0^* &= \underline{\theta}_0 + \underline{\Sigma}_{00}\,\underline{\tau}_0 + \underline{\Sigma}_{01}\,\underline{\tau}_1 \\ \underline{\theta}_1^* &= \underline{\theta}_1 + \underline{\Sigma}_{10}\,\underline{\tau}_0 + \underline{\Sigma}_{11}\,\underline{\tau}_1 \end{aligned}$$

and using the fact that $\underline{\theta}_0^* = \underline{\theta}_0$, it is found that using these results in (4.8) now yields

$$(4.20)\quad 2I(p^*:\pi) = (\underline{\theta}_0^* - \underline{\theta}_0)'\,\underline{\Sigma}_{00\cdot 1}^{-1}\,(\underline{\theta}_0^* - \underline{\theta}_0) = \underline{\tau}_1'\,\underline{\Sigma}_{00\cdot 1}\,\underline{\tau}_1$$

where $\underline{\Sigma}_{00\cdot 1} = \underline{\Sigma}_{00} - \underline{\Sigma}_{01}\,\underline{\Sigma}_{11}^{-1}\,\underline{\Sigma}_{10}$ is an $m_0 \times m_0$ matrix. The results under the partitioning will help in interpreting the analysis of information values and are similar to those occurring in the testing of subhypotheses in the linear and multivariate linear hypothesis theory [11, p. 216, 259].

We note from (4.6) and (4.7) that

$$(4.21)\quad \frac{\partial}{\partial \tau_i} \ell n\, p^*(\omega) = T_i(\omega) - \frac{\partial}{\partial \tau_i} \ell n\, M(\tau_1, \tau_2, \ldots) = T_i(\omega) - \theta_i^*,$$

hence $T_i(\omega)$ is the maximum likelihood estimator of θ_i^*. Thus if we write $T_i(\omega) = \hat{\theta}_i^*$ and denote the values satisfying (4.6) or (4.14) with $\hat{\theta}_i^*$ in place of θ_i^* and $\hat{\tau}_i$ in place of τ_i, we have corresponding to (4.15)

$$(4.22)\quad 2I(\hat{p}^* : \pi) = 2 \sum_i \hat{\tau}_i \hat{\theta}_i^* - 2 \ell n\, M(\hat{\tau}_1, \hat{\tau}_2, \ldots) \approx (\underline{\hat{\theta}}^* - \underline{\theta})' \underline{\Sigma}^{-1} (\underline{\hat{\theta}}^* - \underline{\theta}) \approx \underline{\hat{\tau}}' \underline{\Sigma}\, \underline{\hat{\tau}}$$

and corresponding to (4.20)

$$(4.23)\quad 2I(\hat{p}^* : \pi) = (\underline{\hat{\theta}}_\bullet^* - \underline{\theta}_\bullet)' \underline{\Sigma}_{\bullet\bullet}^{-1} (\underline{\hat{\theta}}_\bullet^* - \underline{\theta}_\bullet) \approx \underline{\hat{\tau}}_\bullet' \underline{\Sigma}_{\bullet\bullet} \underline{\hat{\tau}}_\bullet .$$

We remark that the covariance matrix of the $\hat{\tau}$'s is the inverse of the covariance matrix of the $T_i(\omega)$'s.

If the $\hat{\theta}_i^*$ are the averages of n independent observations then we have for the minimum discrimination information statistics

$$(4.24)\quad 2n\, I(\hat{p}^* : \pi) \approx n(\underline{\hat{\theta}}^* - \underline{\theta})' \underline{\Sigma}^{-1} (\underline{\hat{\theta}}^* - \underline{\theta}) \approx n\, \underline{\hat{\tau}}' \underline{\Sigma}\, \underline{\hat{\tau}}$$

and in the partitioned case

$$(4.25)\quad 2n\, I(\hat{p}^* : \pi) \approx n(\underline{\hat{\theta}}_\bullet^* - \underline{\theta}_\bullet)' \underline{\Sigma}_{\bullet\bullet}^{-1} (\underline{\hat{\theta}}_\bullet^* - \underline{\theta}_\bullet) \approx n\, \underline{\hat{\tau}}_\bullet' \underline{\Sigma}_{\bullet\bullet} \underline{\hat{\tau}}_\bullet .$$

Under the null hypothesis $2n\, I(\hat{p}^* : \pi)$ in (4.24) or (4.25) is asymptotically distributed as χ^2 respectively with m or m_* degrees of freedom.

5. Computational procedures. An experiment has been designed and observations made resulting in a multidimensional contingency table with the desired classifications and categories. All the information the experimenter hopes to obtain from the experiment is contained in the contingency table. In the process of analysis, the aim is to express the observed table by a number of parameters depending on some or all of the marginals, that is, to find out how much of this total information is contained in a summary consisting of sets of marginals. Indeed, the relationship between the concept of independence or association and interaction in contingency tables and the role the marginals play is evidenced in the writings of Bartlett [1], Simpson [17], Roy and Kastenbaum [16], Lewis [15], Darroch [2] and others on the analysis of contingency tables. Thus, the θ's in the preceding discussion will be the marginals of interest.

5.1. The T(w) functions. The T(w) functions for the R x S x T x U table turn out to be a basic set of simple functions and their various products. Thus, for example, the T(w) function associated with the one-way marginal p(2...) is

$$(5.1)\quad T_2^R(ijk\ell) = 1 \quad \text{for } i = 2, \text{ any } j,k,\ell$$
$$= 0 \quad \text{otherwise}$$

since

$$(5.2)\quad \Sigma\, p(ijk\ell)\; T_{\beta}^{R}(ijk\ell) = p(2\ldots).$$

Similarly the T(ω) function associated with the one-way marginal p(..3.), for example, is

$$(5.3)\quad T_{\gamma}^{T}(ijk\ell) = 1 \qquad \text{for } k = 3, \text{ any } i,j,\ell$$
$$= 0 \qquad \text{otherwise}$$

since

$$(5.4)\quad \Sigma\, p(ijk\ell)\; T_{\gamma}^{T}(ijk\ell) = p(..3.).$$

Thus for the r x s x t x u table we have

$$(5.5)\quad \begin{array}{l} (r\text{-}1) \text{ linearly independent functions } T_{\alpha}^{R}(ijk\ell), \alpha = 1,\ldots,r\text{-}1 \\ (s\text{-}1) \text{ linearly independent functions } T_{\beta}^{S}(ijk\ell), \beta = 1,\ldots,s\text{-}1 \\ (t\text{-}1) \text{ linearly independent functions } T_{\gamma}^{T}(ijk\ell), \gamma = 1,\ldots,t\text{-}1 \\ (u\text{-}1) \text{ linearly independent functions } T_{\delta}^{U}(ijk\ell), \delta = 1,\ldots,u\text{-}1, \end{array}$$

since, for example,

$$\sum_{\alpha=1}^{r} \Sigma\; T_{\alpha}^{R}(ijk\ell) = rstu\ .$$

We have arbitrarly excluded the functions corresponding to $\alpha = r$, $\beta = s$, $\gamma = t$, $\delta = u$ as a matter of convenience, we could have selected $\alpha = 1$, $\beta = 1$, $\gamma = 1$, $\delta = 1$ or any other set of values.

The T(ω) function associated with the two-way marginal p(12..) say, is $T_1^R(ijk\ell)\ T_2^S(ijk\ell)$ since from the definition of $T_1^R(ijk\ell)$ and $T_2^S(ijk\ell)$ it may be seen that

$$(5.6)\quad T_1^R(ijk\ell)\ T_2^S(ijk\ell) = 1 \quad \text{for } i = 1,\ j = 2, \text{ any } k,\ell$$
$$= 0 \qquad \text{otherwise}$$

and

$$(5.7)\quad \Sigma\ p(ijk\ell)\ T_1^R(ijk\ell)\ T_2^S(ijk\ell) = p(12..).$$

Thus the T(ω) function associated with any two-way marginal is a product of two appropriate functions of the set (5.5).

Similarly the T(ω) function associated with any three-way marginal will be a product of three of the appropriate functions of the set (5.5), for example,

$$(5.8)\quad \Sigma\ p(ijk\ell)\ T_2^R(ijk\ell)\ T_1^T(ijk\ell)\ T_3^U(ijk\ell) = p(2.13).$$

Similarly the T(ω) function associated with any four-way marginal will be a product of four of the appropriate functions of the set (5.5), for example,

$$(5.9)\quad \Sigma\ p(ijk\ell)\ T_2^R(ijk\ell)\ T_1^S(ijk\ell)\ T_1^T(ijk\ell)\ T_2^U(ijk\ell) = p(2112).$$

We note that there are a total of

$$N_1 = (r-1) + (s-1) + (t-1) + (u-1)$$
$$N_2 = (r-1)(s-1)+(r-1)(t-1)+(r-1)(u-1)+(s-1)(t-1)+(s-1)(u-1)+(t-1)(u-1)$$

$$N_3=(r-1)(s-1)(t-1)+(r-1)(s-1)(u-1)+(r-1)(t-1)(u-1)+(s-1)(t-1)(u-1)$$

$$N_4=(r-1)(s-1)(t-1)(u-1)$$

respectively of the simple linearly independent functions and their products two, three, four at a time. It may be verified that

$$(5.10) \quad rstu - 1 = N = N_1 + N_2 + N_3 + N_4 .$$

These values are degrees of freedom in the analysis of information tables in [6], [10].

5.2. The $p^*(\omega)$ values. In the usual regression analysis procedure, one first computes the regression coefficients and then gets the values of the estimates. In this case however we reverse the procedure. Instead of trying to obtain the values of the τ's from (4.6) we shall first obtain the values of $p^*(\omega)$ by a straightforward convergent iterative procedure and then derive the values of the τ's from (4.7). We shall not discuss the details of the iteration here since they have been described in [4], [6], [9], [10]. The iteration may be described as successively cycling through adjustments of the marginals of interest starting with the $\pi(\omega)$ distribution until a desired accuracy of agreement between the set of observed marginals of interest and the computed marginals has been attained.

5.3. The τ values. From the definitions of the $T(\omega)$ functions in section 5.1 it is clear that they take on only the values 0 or 1 for each value of ω. From the nature of the $T(\omega)$ functions the set of regression equations (4.7) will have some

with a single τ value which can be determined. Then there will be a set with one additional unknown value and some of the τ's already determined. These new unknown τ values can be then determined. This process of successive evaluation is carried on until all the values of τ are determined.

6. Analysis of information. Although the preceding theoretical discussion has been in terms of probabilities, estimated probabilities or relative frequencies, in practice it has been found more convenient not to divide everything by n, the total number of occurrences, and deal with observed or estimated occurrences, that is, with $m(ijkl) = n/rstu$, $x(ijkl)$, $x(i...)$, $x(.jk.)$, $x^*(ijkl) = n\,p^*(ijkl)$ etc. The analysis of information is based on the fundamental relation (4.9) for the minimum discrimination information statistics. Specifically if $n\,p_a^*(\omega) = x_a^*(\omega)$ is the minimum discrimination information estimate corresponding to a set H_a of given marginals and $x_b^*(\omega)$ is the minimum discrimination information estimate corresponding to a set H_b of given marginals, where $H_b \subset H_a$, then the basic relations are

$$
\begin{aligned}
2I(x:m) &= 2I(x_a^*:m) + 2I(x:x_a^*) \\
2I(x:m) &= 2I(x_b^*:m) + 2I(x:x_b^*) \\
2I(x_b^*:m) &= 2I(x_a^*:m) + 2I(x_b^*:x_a^*) \\
2I(x:x_a^*) &= 2I(x_b^*:x_a^*) + 2I(x:x_b^*)
\end{aligned}
\tag{6.1}
$$

In terms of the representation in (4.4) as an exponential family, for our discussion, the two extreme cases are the uniform distribution for which all τ's are zero, and the observed contingency table or distribution for which all $N = rstu - 1$ τ's are needed.

Measures of the form $2I(x:x_a^*)$, that is, the comparison of an observed contingency table with an estimated contingency table, are called measures of interaction and measures of the form $2I(x_b^*:x_a^*)$, that is, the comparison of two estimated contingency tables, are called measures of effect, that is the effect of the marginals in the set H_b but not in the set H_a. From the results in (4.24) or (4.25) we see that $2I(x:x_a^*)$ tests a null hypothesis that the set of τ parameters in the representation of the observed contingency table $x(\omega)$ but not in the representation of the estimated table $x_a^*(\omega)$ are zero, and $2I(x_b^*:x_a^*)$ tests a null hypothesis that the additional set of τ parameters in the representation of the estimated table $x_b^*(\omega)$ but not in the representation of the estimated table $x_a^*(\omega)$ are zero.

Since the marginals of the estimated table $x_a^*(\omega)$ which form the set of restraints H_a used to generate $x_a^*(\omega)$ are the same as the corresponding marginals of the observed $x(\omega)$ table and all lower order implied marginals, $2I(x:x_a^*)$ is also approximately a quadratic in the differences between the remaining marginals of the $x(\omega)$ table and the corresponding ones as calculated from the $x_a^*(\omega)$ table.

Similarly $2I(x_b^*:x_a^*)$ is also approximately a quadratic in the differences between those additional marginals in H_b but not in H_a and the corresponding marginal values as computed from the

$x^*_\alpha(\omega)$ table.

As we shall see, because of the nature of the $T(\omega)$ functions described in section 5.1, the τ's are determined from the regression equations (4.7) as sums and differences of values of $\ln x^*(ijk\ell)$. A variety of statistics have been presented in the literature for the analysis of contingency tables which are quadratics in the marginal values or quadratics in the logarithms of the observed or estimated values. The principle of minimum discrimination information estimation and its procedures thus provides a unifying relationship since such statistics may be seen as opposite faces of the minimum discrimination information statistic.

We have presented the approximations in terms of quadratic forms in the marginals or the τ's to assist in understanding and interpreting the analysis of information tables as a bridge connecting the familiar procedures of classical regression analysis and the procedures proposed here. The covariance matrix of the $T(\omega)$ functions can be estimated for either the observed table or any of the estimated tables and the inverse of that matrix found should their values be desired.

7. <u>The 2 x 2 table</u>. Before we present an application of the preceding ideas to experimental data in a four-way contingency table, we shall reexamine the 2 x 2 table from the point of view of this paper. The algebraic details are simple in this case and exhibit the unification of the information theoretic development.

Suppose we have the observed 2 x 2 table in figure 7.1.

If we fit the one-way

x(11)	x(12)	x(1.)
x(21)	x(22)	x(2.)
x(.1)	x(.2)	n

Figure 7.1

marginals, the generalized independence hypothesis is the classical independence hypothesis and the minimum discrimination information estimate is $x^*(ij) = x(i.)x(.j)/n$. A convenient representation of the regression (4.7) is given in figure 7.2. The entries in the columns τ_1, τ_2, τ_3

i	j	L	τ_1	τ_2	τ_3
1	1	1	1	1	1
1	2	1	1		
2	1	1		1	
2	2	1			

Figure 7.2

are respectively the values of the functions $T_1(ij)$, $T_2(ij)$, $T_3(ij)$ associated with the marginals $\theta_1 = x(1.)$, $\theta_2 = x(.1)$, $\theta_3 = x(11)$, and the column headed L corresponds to the negative of the logarithm of the moment-generating function. For the observed distribution, recalling the regression (4.7), it is found that

(7.1) $L = \ell n\,(x(22)/n/4)$, $\tau_1 = \ell n\,(x(12)/x(22))$, $\tau_2 = \ell n\,(x(21)/x(22))$

$\tau_3 = \ell n\,(x(11)x(22)/x(12)x(21))$.

If we call $\underline{T}$ the matrix with columns the columns of Figure 7.2, that is,

$$(7.2)\quad \underline{T} = \begin{pmatrix} 1 & 1 & 1 & 1 \\ 1 & 1 & 0 & 0 \\ 1 & 0 & 1 & 0 \\ 1 & 0 & 0 & 0 \end{pmatrix}$$

and define a diagonal matrix $\underline{D}$ with main diagonal the elements $x(ij)$, that is,

$$(7.3)\quad \underline{D} = \begin{pmatrix} x(11) & 0 & 0 & 0 \\ 0 & x(12) & 0 & 0 \\ 0 & 0 & x(21) & 0 \\ 0 & 0 & 0 & x(22) \end{pmatrix}$$

then it may be verified that the estimate of the covariance matrix of the $T_i(w)$ for the observed contingency table is $\underline{\Sigma} = \underline{A}_{22.1}$ where

$$(7.4)\quad \underline{A} = \begin{pmatrix} \underline{A}_{11} & \underline{A}_{12} \\ \underline{A}_{21} & \underline{A}_{22} \end{pmatrix} = \underline{T}'\underline{D}\underline{T}$$

$$(7.5)\quad \underline{A}_{22.1} = \underline{A}_{22} - \underline{A}_{21}\,\underline{A}_{11}^{-1}\,\underline{A}_{12}$$

and $\underline{A}_{11}$ is 1×1, $\underline{A}_{22}$ is 3×3, $\underline{A}'_{21} = \underline{A}_{12}$ is 1×3. It is found that

$$(7.6)\quad \underline{\Sigma} = \begin{pmatrix} \dfrac{x(1.)x(2.)}{n} & x(11) - \dfrac{x(1.)x(.1)}{n} & \dfrac{x(11)x(2.)}{n} \\ x(11) - \dfrac{x(1.)x(.1)}{n} & \dfrac{x(.1)x(.2)}{n} & \dfrac{x(11)x(.2)}{n} \\ \dfrac{x(11)x(2.)}{n} & \dfrac{x(11)x(.2)}{n} & x(11) - \dfrac{x^2(11)}{n} \end{pmatrix}$$

Even for this simple case inverting the matrix in (7.6) is messy algebraically, however, it is easier to use the relations in (4.10) and (4.11). We have from (7.1)

$$(7.7)\quad \tau_1 = \ln x(12) - \ln x(22),\ \tau_2 = \ln x(21) - \ln x(22),$$
$$\tau_3 = \ln x(11) + \ln x(22) - \ln x(12) - \ln x(21)$$

and from $\theta_1 = x(1.)$, $\theta_2 = x(.1)$, $\theta_3 = x(11)$ and the relations implied in Figure 7.1 it is found that

$$(7.8)\quad x(11) = \theta_3,\ x(12) = \theta_1 - \theta_3,\ x(21) = \theta_2 - \theta_3,\ x(22) = n - \theta_1 - \theta_2 + \theta_3.$$

It then follows that

$$\frac{\partial \tau_1}{\partial \theta_1} = \frac{1}{x(12)} + \frac{1}{x(22)},\quad \frac{\partial \tau_1}{\partial \theta_2} = \frac{1}{x(22)},\quad \frac{\partial \tau_1}{\partial \theta_3} = -\frac{1}{x(12)} - \frac{1}{x(22)}$$

$$\frac{\partial \tau_2}{\partial \theta_1} = \frac{1}{x(22)},\quad \frac{\partial \tau_2}{\partial \theta_2} = \frac{1}{x(21)} + \frac{1}{x(22)},\quad \frac{\partial \tau_2}{\partial \theta_3} = -\frac{1}{x(21)} - \frac{1}{x(22)}$$

$$(7.9)$$

$$\frac{\partial \tau_3}{\partial \theta_1} = -\frac{1}{x(12)} - \frac{1}{x(22)},\quad \frac{\partial \tau_3}{\partial \theta_2} = -\frac{1}{x(21)} - \frac{1}{x(22)},$$

$$\frac{\partial \tau_3}{\partial \theta_3} = \frac{1}{x(11)} + \frac{1}{x(12)} + \frac{1}{x(21)} + \frac{1}{x(22)}$$

that is, the entries of Σ^{-1} since $\frac{\partial \tau_i}{\partial \theta_j} = \sigma^{ij}$.

Note that the value of the logarithm of the cross-product ratio as a measure of association appears in the course of the analysis as the value of τ_3, and that $\tau_3 = 0$ for $x^*(ij)$ whose representation as in Figure 7.2 does not involve the last column. The minimum discrimination information statistic to test the null hypothesis of independence is $2I(x:x^*)$. In this case $\varphi_1^* = \varphi_1$, $\varphi_2^* = \varphi_2$ and in accordance with (4.25)

$$(7.10)\quad 2I(x:x^*) = \left(x(11) - \frac{x(1.)x(.1)}{n}\right)^2 \left(\frac{1}{x^*(11)} + \frac{1}{x^*(12)} + \frac{1}{x^*(21)} + \frac{1}{x^*(22)}\right).$$

Remembering that $x^*(ij) = x(i.)x(.j)/n$, the right-hand side of (7.10) may also be shown to be

$$(7.11)\quad \Sigma\, (x(ij) - x(i.)x(.j)/n)^2 \Big/ \frac{x(i.)x(.j)}{n}$$

the classical X^2-test for independence with one degree of freedom. A test which has been proposed for the null hypothesis of no association or no interaction in the 2 x 2 table is

$$(7.12)\quad (\ln x(11) + \ln x(22) - \ln x(12) - \ln x(21))^2 \left(\frac{1}{x(11)} + \frac{1}{x(12)} + \frac{1}{x(21)} + \frac{1}{x(22)}\right)^{-1}$$

which is seen to be the approximation for $2I(x:x^*)$ in terms of the τ's with the covariance matrix estimated using the observed values and not the estimated values. We remark that if the observed values are used to estimate the covariance matrix then instead of the classical X^2- test in (7.11) there is derived the modified

Neyman chi-square

$$(7.13)\quad X'^2 = \Sigma(x(ij) - x(i.)x(.j)/n)^2/x(ij).$$

8. **Example with experimental data.** Consider the R × S × T × U table 8.1a representing the results of test shooting under three different conditions:

R:Gun barrel wear:i=1, new, i=2, moderate, i=3, excessive

S:Gun barrel temperature:j=1, cold, j=2, hot

T:Unit temperature:k=1, hot, k=2, ambient, k=3, cold

U:Number operative:ℓ=1, success, ℓ=2, failure.

We are indebted to Mr. B.M. Kurkjian of the Harry Diamond Laboratories for the data and his interest in the analytic procedure we have discussed. We note that 15 rounds each were fired under each of 18 experimental conditions. This is not necessary for the application of the analysis of information procedures but was required for the earlier application of Brandt's analysis to the data.

Figure 8.1 presents a graphic representation of the regression (4.7) and is similar to that in Figure 7.2 for the 2 × 2 table. The L column corresponds to the negative of the logarithm of the moment-generating function (a normalizing value) and each of the other columns is a $T(\omega)$ function with the associated τ value at the head of the column. Superscripts and subscripts are used to identify the factors and categories involved. The complete representation in Figure 8.1 with the 35 τ values will provide an exact representation for the observed values $x(\omega)$. Tables 8.2, 8.3, and 8.4 are analysis of information tables

presenting appropriate analyses as various sets of marginals of interest are introduced as explanatory variables.

In Figure 8.2 the columns corresponding to the τ parameters which enter into the various distributions appearing in tables 8.2, 8.3 and 8.4 have been checked. Note that for m, the uniform distribution, there are no checks, and for $x(\omega)$, the observed distribution all columns are checked. The degrees of freedom for any effect component is the difference in the number of columns checked for the corresponding estimates. The degrees of freedom for any interaction component is the difference in the number of columns checked for the observed x-distribution and the estimated distribution.

The null hypothesis for any effect component is that the additional τ parameters are zero, for example, the null hypothesis for the effect component $2I(x_\zeta^*:x_\zeta^*)$ in table 8.2 is that τ_{11}^{RU}, τ_{21}^{RU} are zero. The null hypothesis for any interaction component is that the set of parameters which are checked for the observed x-distribution but not for the estimated distribution are zero, for example, the null hypothesis for the third-order interaction component $2I(x:x_\beta^*)$ in table 8.2 is that τ_{1111}^{RSTU}, τ_{2111}^{RSTU}, τ_{1121}^{RSTU}, τ_{2121}^{RSTU} are zero.

Note that all the marginals implied for x_α^* in table 8.4 are $x(1\ldots)$, $x(.j..)$, $x(..k.)$, $x(...l)$, $x(1j..)$, $x(1.k.)$, $x(1..l)$, $x(.jk.)$, $x(.j.l)$, $x(1jk.)$, $x(1j.l)$ and the marginals implied for x_β^* in table 8.4 are $x(1\ldots)$, $x(.j..)$, $x(..k.)$, $x(...l)$, $x(1j..)$, $x(1.k.)$, $x(1..l)$, $x(.jk.)$, $x(.j.l)$, $x(..kl)$, $x(1jk.)$, $x(1j.l)$, $x(1.kl)$, hence the six parameters $\tau_{11}^{TU}, \tau_{21}^{TU}, \tau_{111}^{RTU}, \tau_{211}^{RTU}, \tau_{121}^{RTU}, \tau_{221}^{RTU}$ appear in x_α^* but not in x_β^*.

We draw the following conclusions from tables 8.2, 8.3, 8.4:

1. Success/Failure is not homogeneous over the 18 experimental situations, $x_s^*(ijk\ell) = x(ijk.)x(...\ell)/n$, $2I(x:x_s^*) = 34.371$, 17 D.F.

2. The effect of $x(i..\ell)$ in table 8.2 is almost significant, but those of $x(.j.\ell)$, $x(..k\ell)$ are not significant, hence we proceed as in table 8.4.

3. The marginals $x(ijk.)$, $x(ij.\ell)$, $x(i.k\ell)$ and the lower order marginals they imply provide an acceptable estimate for the original data since $2I(x:x_s^*) = 7.413$, 6 D.F., that is, we accept a null hypothesis that the set of six parameters τ_{111}^{STU}, τ_{1s1}^{STU}, τ_{1111}^{RSTU}, τ_{s111}^{RSTU}, τ_{11s1}^{RSTU}, τ_{s1s1}^{RSTU} are zero.

4. Using Figure 8.1 and Figure 8.2 we can express the logarithm of the ratio of the estimates for success to failure under all 18 experimental conditions, that is, the logit, as the linear combination of a constant term τ_1^U, a term depending on barrel wear τ_{11}^{RU}, τ_{s1}^{RU}, a term depending on the interaction of barrel wear and barrel temperature τ_{111}^{RSU}, τ_{s11}^{RSU}, and a term depending on the interaction of barrel wear and unit temperature τ_{111}^{RTU}, τ_{s11}^{RTU}, τ_{1s1}^{RTU}, τ_{ss1}^{RTU}.

$$\ell n \frac{x_s^*(1111)}{x_s^*(1112)} = \tau_1^U + \tau_{11}^{RU} + \tau_{11}^{SU} + \tau_{11}^{TU} + \tau_{111}^{RSU} + \tau_{111}^{RTU}$$

$$\ell n \frac{x_s^*(1211)}{x_s^*(1212)} = \tau_1^U + \tau_{11}^{RU} + \tau_{11}^{TU} + \tau_{111}^{RTU}$$

$$\ell n \frac{x_s^*(2111)}{x_s^*(2112)} = \tau_1^U + \tau_{s1}^{RU} + \tau_{11}^{SU} + \tau_{11}^{TU} + \tau_{s11}^{RSU} + \tau_{s11}^{RTU}$$

$$\ell n \frac{x_s^*(2211)}{x_s^*(2212)} = \tau_1^U + \tau_{s1}^{RU} + \tau_{11}^{TU} + \tau_{s11}^{RTU}$$

$$\ln \frac{x_3^*(3111)}{x_3^*(3112)} = \tau_1^U + \tau_{11}^{SU} + \tau_{11}^{TU}$$

$$\ln \frac{x_3^*(3211)}{x_3^*(3212)} = \tau_1^U + \tau_{11}^{TU}$$

$$\ln \frac{x_3^*(1121)}{x_3^*(1122)} = \tau_1^U + \tau_{11}^{RU} + \tau_{11}^{SU} + \tau_{21}^{TU} + \tau_{111}^{RSU} + \tau_{121}^{RTU}$$

$$\ln \frac{x_3^*(1221)}{x_3^*(1222)} = \tau_1^U + \tau_{11}^{RU} + \tau_{21}^{TU} + \tau_{121}^{RTU}$$

$$\ln \frac{x_3^*(2121)}{x_3^*(2122)} = \tau_1^U + \tau_{21}^{RU} + \tau_{11}^{SU} + \tau_{21}^{TU} + \tau_{211}^{RSU} + \tau_{221}^{RTU}$$

$$\ln \frac{x_3^*(2221)}{x_3^*(2222)} = \tau_1^U + \tau_{21}^{RU} + \tau_{21}^{TU} + \tau_{221}^{RTU}$$

$$\ln \frac{x_3^*(3121)}{x_3^*(3122)} = \tau_1^U + \tau_{11}^{SU} + \tau_{21}^{TU}$$

$$\ln \frac{x_3^*(3221)}{x_3^*(3222)} = \tau_1^U + \tau_{21}^{TU}$$

$$\ln \frac{x_3^*(1131)}{x_3^*(1132)} = \tau_1^U + \tau_{11}^{RU} + \tau_{11}^{SU} + \tau_{111}^{RSU}$$

$$\ln \frac{x_3^*(1231)}{x_3^*(1232)} = \tau_1^U + \tau_{11}^{RU}$$

$$\ln \frac{x_3^*(2131)}{x_3^*(2132)} = \tau_1^U + \tau_{21}^{RU} + \tau_{11}^{SU} + \tau_{211}^{RSU}$$

$$\ln \frac{x_3^*(2231)}{x_3^*(2232)} = \tau_1^U + \tau_{21}^{RU}$$

$$\ell n \frac{x_3^*(3131)}{x_3^*(3132)} = \tau_1^U + \tau_{11}^{SU}$$

$$\ell n \frac{x_3^*(3231)}{x_3^*(3232)} = \tau_1^U$$

5. Since the computer program provides not only the values of $x_3^*(ijk\ell)$ but also the values of $\ell n\,(x_3^*(ijk\ell)/x_3^*(3232))$, the values of the τ's in conclusion 4 above can be easily found.

$$\tau_1^U = \ell n \frac{x_3^*(3231)}{x_3^*(3232)} = 3.0281$$

$$\tau_{11}^{RU} = \ell n \frac{x_3^*(1231)}{x_3^*(1232)} - \tau_1^U = -1.6470$$

$$\tau_{21}^{RU} = \ell n \frac{x_3^*(2231)}{x_3^*(2232)} - \tau_1^U = -2.2870$$

$$\tau_{11}^{SU} = \ell n \frac{x_3^*(3131)}{x_3^*(3132)} - \tau_1^U = -0.6794$$

$$\tau_{11}^{TU} = \ell n \frac{x_3^*(3211)}{x_3^*(3212)} - \tau_1^U = -1.9759$$

$$\tau_{21}^{TU} = \ell n \frac{x_3^*(3221)}{x_3^*(3222)} - \tau_1^U = -0.7746$$

$$\tau_{111}^{RSU} = \ell n \frac{x_3^*(1131)}{x_3^*(1132)} - \tau_1^U - \tau_{11}^{RU} - \tau_{11}^{SU} = -0.2928$$

$$\tau_{211}^{RSU} = \ell n \frac{x_3^*(2131)}{x_3^*(2132)} - \tau_1^U - \tau_{21}^{RU} - \tau_{11}^{SU} = 1.7215$$

$$\tau_{111}^{RTU} = \ell n \frac{x_3^*(1211)}{x_3^*(1212)} - \tau_1^U - \tau_{11}^{RU} - \tau_{11}^{TU} = 2.3336$$

$$\tau_{211}^{RTU} = \ell n \frac{x_3^*(2211)}{x_3^*(2212)} - \tau_1^U - \tau_{21}^{RU} - \tau_{11}^{TU} = 1.4528$$

$$\tau_{1\,2\,1}^{RTU} = \ln \frac{x_s^*(1221)}{x_s^*(1222)} - \tau_1^U - \tau_{11}^{RU} - \tau_{21}^{TU} = 0.1639$$

$$\tau_{2\,2\,1}^{RTU} = \ln \frac{x_s^*(2221)}{x_s^*(2222)} - \tau_1^U - \tau_{21}^{RU} - \tau_{21}^{TU} = 0.5878$$

As a check we have, for example, $\ln (x_s^*(1111)/x_s^*(1112)) = 0.7666$ and $\tau_1^U + \tau_{11}^{RU} + \tau_{11}^{SU} + \tau_{11}^{TU} + \tau_{111}^{RSU} + \tau_{111}^{RTU} = 0.7666$.

6. The values of L and other τ parameters for the x_s^*-distribution can be obtained from Figure 8.1 and the computer listing of the values of $\ln(x_s^*(ijk\ell)/x_s^*(3232))$ and $\ln(x_s^*(3232)/m')$, in this case $m' = 270/(3\times2\times3\times2)$. Thus $L = -2.3822$, $\tau_1^R = \ln(x_s^*(1232)/x_s^*(3232)) = 1.4701$, etc.

7. The computer output for $x_s^*(ijk\ell)$ is listed as table 8.5. Five values are given for each i,j,k,ℓ, these are:

Observed: $x(ijk\ell)$

Predicted: $x_s^*(ijk\ell)$

Residual: $x(ijk\ell) - x_s^*(ijk\ell)$

Standardize: $2\,x(ijk\ell)\,\ln\,(x(ijk\ell)/x_s^*(ijk\ell))$

Log ratio: $\ln(x_s^*(ijk\ell)/x_s^*(3232))$.

There is also given the value of $2I(x:x_s^*)$ along with the degrees of freedom and a probability based on the χ^2- distribution and the value of L as log(x STAR/N/CELLS).

9. Acknowledgment. The interest and cooperation of Professor C.T. Ireland and Dr. H.H. Ku are gratefully acknowledged.

Table 8.1a Original Data $x(ijk\ell)$

	i	1						2						3						
	J	1			2			1			2			1			2			
	K	1	2	3	1	2	3	1	2	3	1	2	3	1	2	3	1	2	3	
ℓ	1	9	8	9	14	9	12	11	14	12	9	8	11	9	13	13	11	13	15	200
	2	6	7	6	1	6	3	4	1	3	6	7	4	6	2	2	4	2	0	70
		15	15	15	15	15	15	15	15	15	15	15	15	15	15	15	15	15	15	270

Table 8.1b Estimated Data $x_a^*(ijk\ell)$

Marginals Fitted: $x(ijk.)$, $x(ij.\ell)$, $x(i.k\ell)$

	i	1						2					
	J	1			2			1			2		
	K	1	2	3	1	2	3	1	2	3	1	2	3
ℓ	1	10.242	6.746	9.012	12.758	10.254	11.988	11.686	12.473	12.841	8.314	9.527	10.159
	2	4.758	8.254	5.988	2.242	4.746	3.012	3.314	2.527	2.159	6.686	5.473	4.841
		15.000	15.000	15.000	15.000	15.000	15.000	15.000	15.000	15.000	15.000	15.000	15.000

	i	3						
	J	1			2			
	K	1	2	3	1	2	3	
ℓ	1	8.883	12.425	13.692	11.117	13.575	14.308	200.000
	2	6.118	2.574	1.307	3.882	1.426	0.693	70.000
		15.001	14.999	14.999	14.999	15.001	15.001	270.000

Analysis of Information

Table 8.2

Component due to	Information	D.F.
a) $x(ijk.)$	$2I(x:x^*_6)$ -99.639	18
b) $x(ijk.), x(...l)$		
U-effect	$2I(x^*_5:x^*_4)$ -65.268	1
Interaction	$2I(x:x^*_5)$ -34.371	17
c) $x(ijk.), x(i..l)$		
RU-effect | RST	$2I(x^*_3:x^*_5)$ = 5.303	2
Interaction	$2I(x:x^*_3)$ -29.068	15
d) $x(ijk.), x(i..l), x(.j.l)$		
SU-effect | RST, RU	$2I(x^*_1:x^*_3)$ = 0.314	1
Interaction	$2I(x:x^*_1)$ -28.754	14
e) $x(ijk.), x(i..l), x(.j.l), x(..kl)$		
TU-effect | RST, RU, SU	$2I(x^*_2:x^*_1)$ = 2.705	2
Interaction	$2I(x:x^*_2)$ -26.049	12
f) $x(ijk.), x(..kl), x(ij.l)$		
RSU-effect | RST, RU, SU, TU	$2I(x^*_7:x^*_2)$ = 9.752	2
Interaction	$2I(x:x^*_7)$ -16.297	10
g) $x(ijk.), x(ij.l), x(i.kl)$		
RTU-effect | RST, RU, SU, TU, RSU	$2I(x^*_8:x^*_7)$ = 8.891	4
Interaction	$2I(x:x^*_8)$ = 7.406	6
n) $x(ijk.), x(ij.l), x(i.kl), x(.jkl)$		
STU-effect | RST, RU, SU, TU, RSU, RTU	$2I(x^*_6:x^*_8)$ = 4.543	2
Third-order interaction	$2I(x:x^*_6)$ = 2.863	4

$x^*_4 = x^*_0$

Table 8.3

Component due to	Information	D.F.
d) $x(ijk.), x(i..l), x(.j.l)$	$2I(x:x_1^*) = 28.754$	14
m) $x(ijk.), x(ij.l)$		
RSU-effect | RST, RU, SU	$2I(x_2^*:x_1^*) = 9.649$	2
Interaction	$2I(x:x_2^*) = 19.105$	12
f) $x(ijk.), x(ij.l), x(..kl)$		
TU-effect | RST, RU, SU, RSU	$2I(x_7^*:x_3^*) = 2.808$	2
Interaction	$2I(x:x_7^*) = 16.297$	10

Table 8.4

Component due to	Information	D.F.
b) $x(ijk.), x(...l)$	$2I(x:x_3^*) = 34.371$	17
c) $x(ijk.), x(i..l)$		
RU-effect | RST	$2I(x_6^*:x_5^*) = 5.303$	2
Interaction	$2I(x:x_6^*) = 29.068$	15
m) $x(ijk.), x(ij.l)$		
RSU-effect | RST, RU	$2I(x_8^*:x_6^*) = 9.963$	3
Interaction	$2I(x:x_8^*) = 19.105$	12
n) $x(ijk.), x(ij.l), x(i.kl)$		
RTU-effect | RST, RU, RSU	$2I(x_8^*:x_4^*) = 11.699$	6
Interaction	$2I(x:x_8^*) = 7.406$	6

$x_1^* \equiv x_8^*$

Table 8.5

Computer Output

RESIDUALS: R * S * T * U. FIRST 2 SUBSCRIPTS: 1 1

			1	2
1	OBSERVED	1	9.000000	6.000000
1	PREDICTED	2	10.241535	4.758002
1	RESIDUAL	3	-1.241535	1.241998
1	STANDARDIZE	4	-1.163043	1.391587
1	LOG RATIO	5	2.493764	1.927140
2	OBSERVED	6	8.000000	7.000000
2	PREDICTED	7	6.746210	8.254419
2	RESIDUAL	8	1.253790	-1.254419
2	STANDARDIZE	9	1.363679	-1.153870
2	LOG RATIO	10	2.276293	2.478661
3	OBSERVED	11	9.000000	6.000000
3	PREDICTED	12	9.012303	5.987520
3	RESIDUAL	13	-0.012303	0.012480
3	STANDARDIZE	14	-0.012295	0.012490
3	LOG RATIO	15	2.369903	2.156589

RESIDUALS: R * S * T * U. FIRST 2 SUBSCRIPTS: 1 2

			1	2
1	OBSERVED	1	14.000000	1.000000
1	PREDICTED	2	12.758455	2.241597
1	RESIDUAL	3	1.241545	-1.241597
1	STANDARDIZE	4	1.300076	-0.807367
1	LOG RATIO	5	2.913506	1.174679
2	OBSERVED	6	9.000000	6.000000
2	PREDICTED	7	10.253795	4.745572
2	RESIDUAL	8	-1.253795	1.254428
2	STANDARDIZE	9	-1.173809	1.407280
2	LOG RATIO	10	2.694960	1.924523
3	OBSERVED	11	12.000000	3.000000
3	PREDICTED	12	11.987688	3.012478
3	RESIDUAL	13	0.012312	-0.012478
3	STANDARDIZE	14	0.012308	-0.012452
3	LOG RATIO	15	2.851192	1.470075

RESIDUALS: R * S * T * U. FIRST 2 SUBSCRIPTS: 2 1

	1	2

1	OBSERVED	1	11.000000	4.000000
1	PREDICTED	2	11.685621	3.314257
1	RESIDUAL	3	-0.685621	0.685743
1	STANDARDIZE	4	-0.665103	0.752241
1	LOG RATIO	5	2.823670	1.565585
2	OBSERVED	6	14.000000	1.000000
2	PREDICTED	7	12.472881	2.527158
2	RESIDUAL	8	1.527119	-1.527158
2	STANDARDIZE	9	1.617001	-0.927095
2	LOG RATIO	10	2.890868	1.294807
3	OBSERVED	11	12.000000	3.000000
3	PREDICTED	12	12.841480	2.158594
3	RESIDUAL	13	-0.841480	0.841406
3	STANDARDIZE	14	-0.813287	0.987465
3	LOG RATIO	15	2.915992	1.136769

RESIDUALS: R * S * T * U, FIRST 2 SUBSCRIPTS: 2 2

	1	2

1	OBSERVED	1	9.000000	6.000000
1	PREDICTED	2	8.314383	6.685740
1	RESIDUAL	3	0.685617	-0.685740
1	STANDARDIZE	4	0.713138	-0.649305
1	LOG RATIO	5	2.485299	2.267289
2	OBSERVED	6	8.000000	7.000000
2	PREDICTED	7	9.527126	5.472840
2	RESIDUAL	8	-1.527126	1.527160
2	STANDARDIZE	9	-1.357613	1.722786
2	LOG RATIO	10	2.621455	2.067110
3	OBSERVED	11	11.000000	4.000000
3	PREDICTED	12	10.158509	4.841404
3	RESIDUAL	13	0.841491	-0.841404
3	STANDARDIZE	14	0.875411	-0.763642
3	LOG RATIO	15	2.685623	1.944516

RESIDUALS: R * S * T * U, FIRST 2 SUBSCRIPTS: 3 1

	1	2

1	OBSERVED	1	9.000000	6.000000
1	PREDICTED	2	8.882741	6.118252
1	RESIDUAL	3	0.117259	-0.118252
1	STANDARDIZE	4	0.118029	-0.117102
1	LOG RATIO	5	2.551422	2.178588
2	OBSERVED	6	13.000000	2.000000
2	PREDICTED	7	12.425370	2.574304
2	RESIDUAL	8	0.574630	-0.574304
2	STANDARDIZE	9	0.587705	-0.504864
2	LOG RATIO	10	2.887053	1.312891
3	OBSERVED	11	13.000000	2.000000

3	PREDICTED	12	13.691921	1.307411
3	RESIDUAL	13	-0.691921	0.692589
3	STANDARDIZE	14	-0.674136	0.850196
3	LOG RATIO	15	2.984118	-0.635361

RESIDUALS: R * S * T * U. FIRST 2 SUBSCRIPTS: 3 2

			1	2
1	OBSERVED	1	11.000000	4.000000
1	PREDICTED	2	11.117260	3.881743
1	RESIDUAL	3	-0.117260	0.118257
1	STANDARDIZE	4	-0.116640	0.120037
1	LOG RATIO	5	2.772810	1.723597
2	OBSERVED	6	13.000000	2.000000
2	PREDICTED	7	13.574622	1.425695
2	RESIDUAL	8	-0.574622	0.574305
2	STANDARDIZE	9	-0.562285	0.676974
2	LOG RATIO	10	2.975513	0.721572
3	OBSERVED	11	15.000000	0.000005
3	PREDICTED	12	14.308073	0.692593
3	RESIDUAL	13	0.691927	-0.692588
3	STANDARDIZE	14	0.708390	-0.000059
3	LOG RATIO	15	3.028135	-0.000001

HYPOTHESIS 4 2I(X:X*) = 7.406 DEGREES OF FREEDOM = 6 PROBABILITY OF A LARGER VALUE = 0.284956

LOG(XSTAR/N/CELLS) = -2.382215

FIGURE 8.1

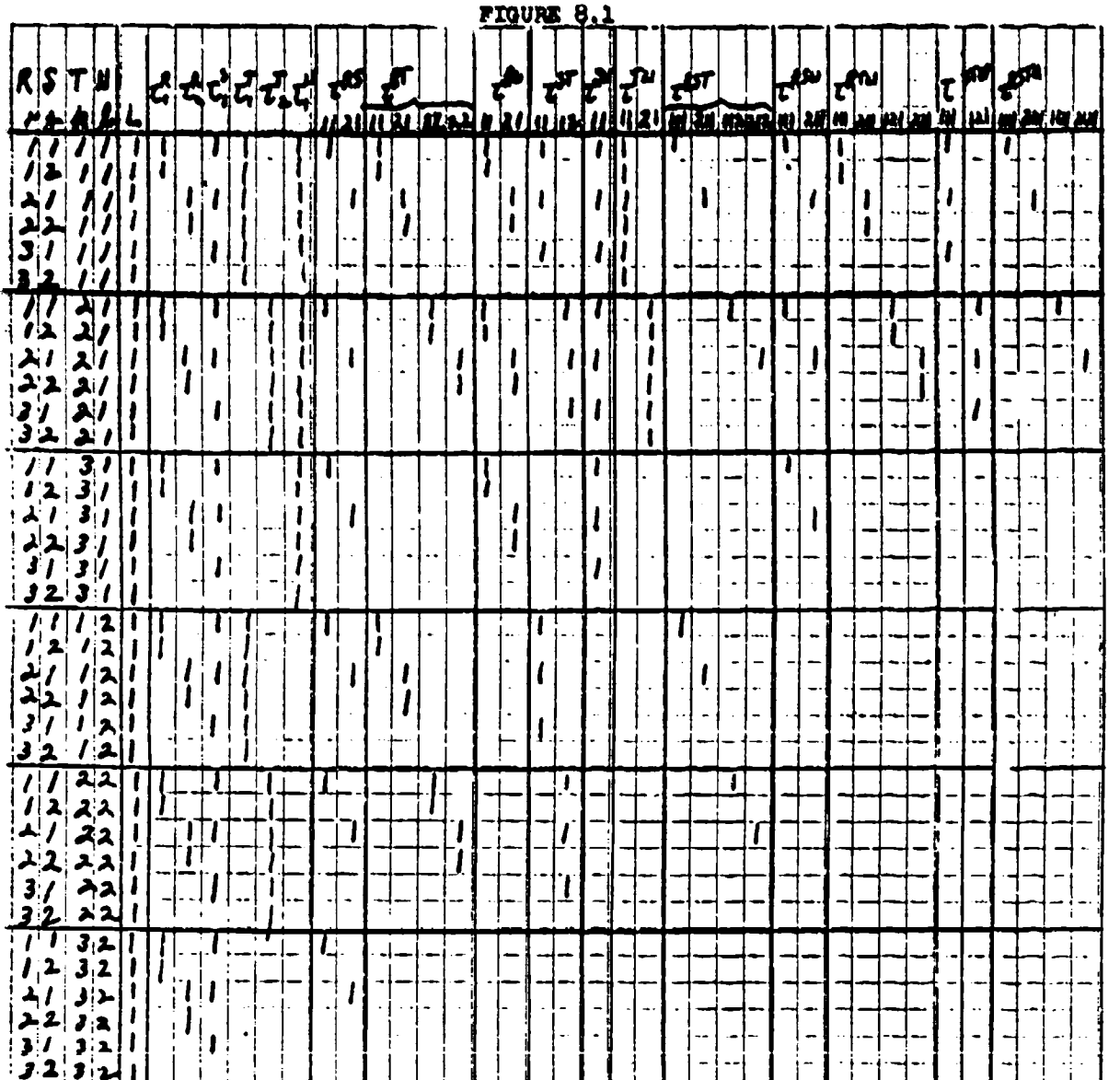
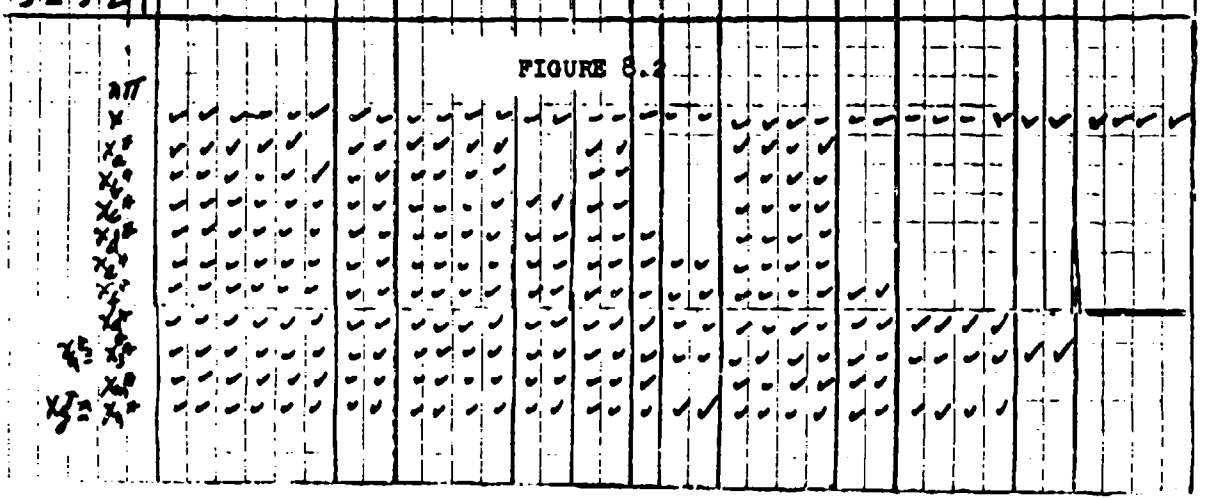


FIGURE 8.2



REFERENCES

[1] Bartlett, M.S. (1935), Contingency table interactions, J. Roy. Statist. Soc. Supplement 2, 248-252.

[2] Darroch, J.N. (1962), Interactions in multi-factor contingency tables, J.R. Statist. Soc. B. 24, 251-263.

[3] Ireland, C.T., Ku, H.H., and Kullback, S., (1969), Symmetry and marginal homogeneity of an r x r contingency table, J. Am. Statist. Assn. 64, 1323-1341.

[4] Ireland, C.T. and Kullback, S., (1968), Contingency tables with given marginals, Biometrika 55, 179-188.

[5] Ireland, C.T. and Kullback, S., (1968), Minimum discrimination information estimation, Biometrics 24, 707-713.

[6] Ku, H.H. and Kullback, S., (1968), Interaction in multi-dimensional contingency tables:an information theoretic approach, J. Res. Nat. Bur. Stand. B 72, 159-199.

[7] Ku, H.H., and Kullback, S., (1969), Approximating discrete probability distribution, I.E.E.E. Trans. on Information Theory, IT-15, 444-447.

[8] Ku, H.H., and Kullback, S. (1969), Analysis of multidimensional contingency tables:an information theoretic approach, I.S.I. Meeting, London, England, Sept. 1969, 156-158.

[9] Ku, H.H., Varner, R., and Kullback, S., (1968), Analysis of multi-dimensional contingency tables, Proceedings of the Fourteenth Conference on the Design of Experiments in Army Research Development and Testing, ARD-D Report 69-2, 141-180.

[10] Ku, H.H., Varner, R., and Kullback, S. (1971), On the analysis of multi-dimensional contingency tables, to appear in J. Am. Statist. Assn.

[11] Kullback, S., (1959), Information Theory and Statistics, Wiley, N.Y.; 1968, Dover Publications, Inc., N.Y.

[12] Kullback, S., (1968), Probability densities with given marginals, Annals Math. Statist. 39, 1236-1243.

[13] Kullback, S., (1970), Various applications of minimum discrimination information estimation, particularly to problems of contingency table analysis, Proceedings of The Meeting on Information Measures, University of Waterloo, Ontario, Canada.

[14] Kullback, S., (1971), Marginal homogeneity of multi-dimensional contingency tables, to appear in Annals of Math. Statist.

[15] Lewis, B.N., (1962), On the analysis of interaction in multi-dimensional contingency tables, J.R. Statist. Soc. A 125, 88-117.

[16] Roy, S.N. and Kastenbaum, M.A., (1956), On the hypothesis of no "interaction" in a multiway contingency table, Annals Math. Statist. 27, 749-57.

[17] Simpson, E.H., (1951), The interpretation of interaction in contingency tables, J.R. Statist. Soc. B 13, 238-241

FIELD TESTING: METHODOLOGICAL CONSIDERATIONS AND A SPECIFIC EXAMPLE

T. S. Donaldson and R. J. Kaplan*
The RAND Corporation
Santa Monica, California

I. INTRODUCTION

The graduate-school-trained experimentalist who goes out into the world of industry or government service generally runs into a severe period of intellectual and emotional stress. At first he will try to apply his education directly by considering the world to be just another laboratory where, with the application of proper controls, he can run experiments just like the ones he was used to doing in school or reading about in the published literature. It soon becomes apparent to him that he cannot, and that the world is just too dirty a place to accommodate his pure way of thinking. The need to do something constructive for his employer, however, quickly becomes a pressing matter, and at this stage his attitude is one of resignation to doing less than first-class work. He thinks you can go out into the world and observe things, even measure a few of the significant parameters with a reasonable degree of accuracy, but that there is no chance of exerting any substantial control over the situation. Any conclusions you draw from such activities are the result of good intuition or luck and not by virtue of any rigorous application of scientific methodology.

Most of us remain in that last stage, a fact which is both unfortunate and unnecessary. Our major point in this paper is that, while it is indeed true that the world out there is dirty and confused compared to our laboratory, we can do experiments in it. There are methods available to the sophisticated researcher which allow him both to exert a great deal of control over investigations taking place in a natural environment and to draw valid conclusions from his observations. Before discussing a specific example, let us look at some of the characteristics of what it is that we can do in the field to further the goals of our various establishments.

The first thing to be made clear is that there is not a dichotomy, but rather a continuum between laboratory experimentation and field testing. The world is dirty to different extents and therefore the degree of control that we can exert in our tests will be different. What must be thought out very carefully before any test is started is the

* Any views expressed in this paper are those of the author. They should not be interpreted as reflecting the views of the RAND Corporation or the official opinion or policy of any of its governmental or private research sponsors. Papers are reproduced by the RAND Corporation as a courtesy to members of its staff.

The remainder of this article was reproduced photographically from the author's manuscript.

tradeoff between what it is we have to have found out when we have completed the work and the rigor with which we conduct the research. The most carefully done and scientifically elegant study which answers totally uninteresting questions is as useless to us as the collection of poor and insufficient evidence about crucial topics. It is the task of the researcher in the field to find the happy medium between these equally undesirable extremes.

The different purposes for which we might want to conduct a field test must also be kept in mind. At one end of the spectrum is the notion of evaluation. The question to be answered here has to do with whether or not the equipment or system lives up to the expectation of its design specifications, and in this case the degree of unusual circumstances in which the system can be made to operate will be a determining factor in the utility of the results. The other end of the scale can be loosely defined by something called policy formulation. Here we already have an operating system and we are trying to learn something about its operational impact both on its own operating efficiency and on the interactions it has with the rest of the organization. The test here is not necessarily limited to the system itself which is under scrutiny as, for example, in a test to decide which of a number of different deployments of a new armored vehicle would be best suited to the accompanying infantry. The test we will use as a sample case later in this discussion is of the policy-formulation type, and we hope it will expose some of the intricacies of field-testing for this purpose.

The prime requirement laid on the researcher who enters the field-testing business is that of being imaginative. In this environment it is not possible to follow the "cookbook" techniques which can assure success in most graduate schools. You cannot change the problem to meet the specifications of your laboratory and you must, therefore, be able to adapt and bend your methods of investigating and the tools available to you to meet the task before you with varying degrees of exactitude. The matching of procedures and methods to a specific case must be done with full knowledge of what has been given up and what has been gained along the way. Nothing can replace the

creativity of the researcher s[illegible] the method must be worked out for each case and cannot be taugh[illegible] r all. He must be able to keep the whole problem in mind[illegible] the specific details together with the compromises which must be [illegible], and be constantly alert to the course of collecting data.

This brings us to the last point we want to make about field tests in general before getting on to a specific case in order to focus our discussion, and that is serendipity. Field testing is done necessarily in a very rich environment; too rich, in fact, and that is what causes all the problems relating to inadequate controls. We can be on the lookout, however, for instances where we can turn this richness to our advantage. Murphey's Law never fails to operate in a field test--the unexpected will always happen. Sometimes it will be disasterous and we will be lucky to salvage anything from the data we have so laborously collected. Other times, however, we will luck out if we are not too narrowly sighted and find things that we had not even been looking for. The good and most creative researcher will even go one step further and force the process of serendipity. Outside of the formal aspects of whatever design you adapt for a particular exercise, there are side effects which can be looked at in a less formal way. Sloppy? Yes, but also creative. As we started out by saying, it's a sloppy world out there and some degree of sloppiness in dealing with it may be of help, just as in some control processes, the introduction of white noise into the system has a stabilizing effect. Plan for the unexpected, then, and the surprise might well be a pleasant one.

II. THE PROBLEM

In the summer of 1964, Rand made an informal study of the Air Defense Command Dispersal Plan. One policy contained in that plan stated that aircraft at the dispersed site would be maintained on ground alert status for a maximum of 48 hours; they would then be flown back to their home bases. The requirement for a rotation period of no more than 48 hours between dispersed site and home base appears to have been based on a widespread belief in ADC that aircraft reliability deteriorates rapidly when aircraft are left on the ground for more than a few days. An inflexible requirement of this kind appears to have serious consequences when aircraft are at a dispersed site. The consequences at the home base are perhaps less serious, but are still present. If a longer ground alert were possible without degrading aircraft performance, flight-scheduling and maintenance-shop manning would be considerably easier because of greater flexibility. Further, a longer alert would lead to a number of cost reductions. For example, the time spent in uploading and downloading weapons for ground alert could be cut to between 25 and 50 percent of the present level. Since loading requires about 9 to 12 men, this saving alone could be considerable. As a result of these considerations, one of the suggestions made in the Rand study was that the effect of ground-alert duration on aircraft reliability be determined experimentally.

As a result of the Rand recommendation, Air Defense Command directed that a study be undertaken. A meeting was held at ADC headquarters between various Air Force officers and a Rand representative to develop the experimental design and procedures. The central issue in this first meeting concerned the tradeoffs between a highly controlled and a loosely controlled design for the test. This issue is focused sharply in consideration of sample size.

Estimating Sample Size. Because ADC was in a hurry to make a decision concerning length of ground alert it was necessary to keep the sample as small as possible. At the same time the Type I and II errors both had to be small. The Type I error is the probability of deciding that differences in aircraft performance exist (as a result of

differences/length of ground alert) when in fact there are no true differences. The Type II error is the probability that true differences exist but are not detected, i.e., the null hypothesis is not rejected. The errors define the two risks taken by the decision-maker, and both involve certain costs to him. If a Type I error occurs, he will decide not to go to longer alert periods (because differences in aircraft performance appear to be related to length of alert) and lose the benefits he would gain by using longer periods. If a Type II error occurs, he will decide in favor of longer alert periods (because he thinks there are no differences due to length of alert) and a degrading of aircraft performance will result.

It is desired that both of these errors be small and of known size. There were two methods available for controlling the size of these errors. The first is through the use of experimental controls which reduce random variations (i.e. the error variance, or error of measurement is smaller). The second is through the solution of an adequate sample size. The calculation of sample size, while controlling for Type I and II errors requires the following: (1) an estimate of error variance, (2) a decision on the magnitude of difference (in aircraft performance) that is meaningful and important, and (3) a statement of the size of risk the decisionmaker is willing to accept, i.e., the size of Type I and II errors.

It was decided that a change in aircraft break-rate (proportion of sorties on which aircraft malfunctioned) of 0.1 was important. It was also decided that the Type I error should be no larger than .05, and the Type II error should be approximately the same size for the most powerful tests, and could be less in some cases.

Using the normal approximation for tests on proportions with $\alpha = .05$ we have:

$$(1) \qquad \text{Prob}\left[\left|\frac{P_1 - P_2}{\sqrt{2\hat{p}\hat{q}/n}}\right| < 1.96\right] = .95,$$

where P_1 and P_2 are the observed proportions, $\hat{p}$ is the estimated value of the proportion in the population and is

$$\hat{p} = \frac{P_1 + P_2}{2} ;$$

and $\hat{q} = 1 - \hat{p}$. The number 1.96 is the value of a normal variate (z), such that Prob ($z < 1.96$) = .95, or it is the critical region for a test when $\alpha = .05$. For the sake of estimating sample size we let $p = 0.5$. This is in the range of values expected in the study. Furthermore, the product $\hat{p}\hat{q}$ (which is used for calculations) changes very little over a wide range of p as indicated in Table 1.

Table 1

$\hat{p}$	$\hat{q}$	$\hat{p}\hat{q}$
.3	.7	.21
.4	.6	.24
.5	.5	.25
.6	.4	.24
.7	.3	.21

Returning to Equation (1), and writing the quantity within the brackets as an identity:

$$(2) \qquad \frac{P_1 - P_2}{\sqrt{2\hat{p}\hat{q}/n}} = 1.96.$$

Solving for n,

$$n = \left[\frac{1.96\sqrt{2\hat{p}\hat{q}}}{P_1 - P_2}\right]^2$$

However, recall that $P_1 - P_2 = .1$, and $p = q = .5$, we have

$$(3) \qquad n = 136$$

Thus, if n = 136, the probability is 95% that $\hat{p}$ falls within the interval $\pm$ 0.1 of p.

The power of the test (1 - prop (Type II Error)) for a difference as big as 0.1 is of course .50, and for a difference as big as 0.2 it is .975. This is shown in Figure 1.

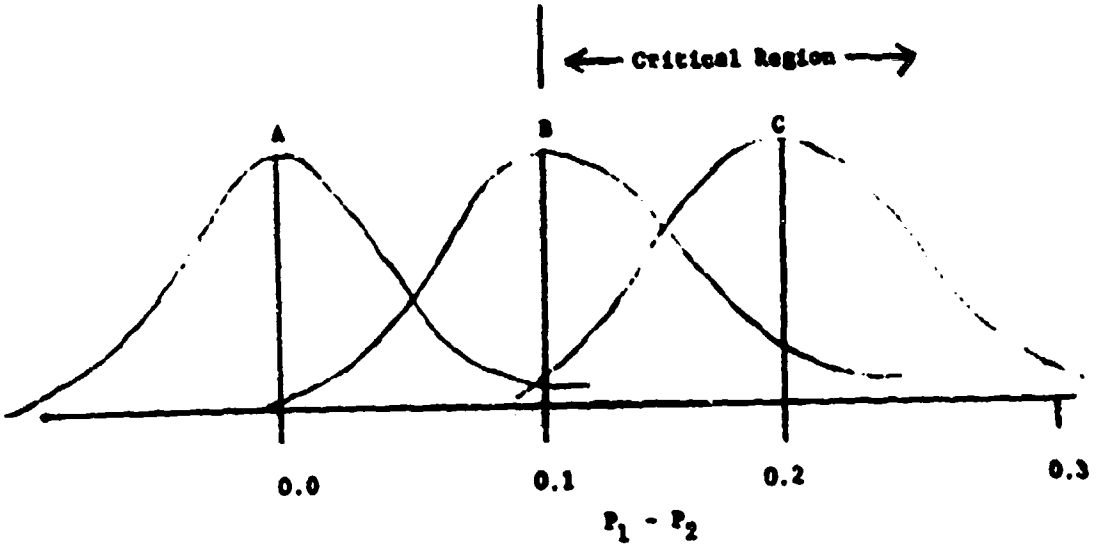


Fig. 1 - Distribution of $P_1 - P_2$ for three values of $\hat{p}$

For a true difference (between the mean of distribution A and B) of 0.1 the power is shown by curve B, and is represented by the area under B to right of 0.1 (the critical region). Obviously this is 50%. For a true difference of 0.2, the power is shown by curve C, and again is represented by the area to the right of 0.1. It is obvious that this is 97.5%, since 95% of the distribution of curve C falls between 0.1 and 0.3 . This follows very directly from the fact that the means of the distributions of $P_1 - P_2$ (0.0, 0.1 and 0.3) are spaced at 1.96 standard deviation intervals. The critical region specifies the values of $P_1 - P_2$ under H_0 for which

H_0 is rejected (0.1 in the example); conversely, the critical region specifies the values of $P_1 - P_2$ for which H_1 is accepted. Power is the probability of accepting H_1 given H_1 is true (i.e., H_1: $P_1 - P_2 = 0.1$ and H_1: $P_1 - P_2 = 0.2$).

Thus we conclude that samples of size approximately 136 will yield acceptable Type I and II errors. As will be pointed out in the following pages, the power of some tests in the design are better than others, and in some cases power is to be improved by pooling procedures.

General Design. Ground-alert periods of four different lengths were used in the study. Four lengths were used in order to determine the shape of the "reliability" curve over time. Further, an even number of periods allows one to combine the short versus long periods for more powerful statistical tests. This was particularly desirable for some indices of aircraft reliability (e.g., ground aborts) that have low frequencies of occurrence. The periods of ground alert used were 2, 4, 6, and 8 days. A two-day period was used because that is the length of alert proposed in the ADC dispersal plan. An eight-day period was used because it was a consensus that aircraft would probably show evidence of decreased reliability after a ground-alert period of this length.* As a result, one could then determine the maximum length of ground alert possible without degrading aircraft reliability. In order to determine the generality of results, three types of ADC aircraft (F-101, F-102, and F-106) and two bases (replications) for each type were used in the study.

Independent Variables. Several indicators of aircraft reliability are used in this study, the primary one being mission capability. The mission-accomplished rate, reported by the pilot and also scored by automatic devices, is the measure of mission capability. These data were obtained from the ADC 76-3 Pilot Debriefing Forms. Two other indicators of aircraft reliability are important, although not

*This was the opinion of some ADC and some Rand personnel, and was thought to reflect the opinion of many maintenance personnel in the field.

directly related to mission capability: pilot-reported malfunction (also on 76-3), and Air Force Manual 66-1 maintenance data. Experience has indicated that the number of malfunctions is a very sensitive measure and the resultant power of the tests would be considerably higher than for tests on break-rate.

In addition to reflecting aircraft reliability, the AFM 66-1 data allow an estimation of maintenance costs as a function of the duration of ground alert. For example, increased effort by maintenance crews and intensive ground testing could possibly maintain aircraft reliability for extremely long periods of ground alert, although the cost for doing so might be high. While aircraft mission capability (as a function of the length of alert) is the primary concern of the study, the cost for that capability cannot be disregarded.

III. PROCEDURES AND EXPERIMENTAL DESIGN

In this section the experimental controls which were introduced to increase test sensitivity are stated.

Aircraft of three types from six ADC bases were placed on ground alert for periods of 2, 4, 6, and 8 days. The effect of ground-alert duration on aircraft reliability was measured by aircraft performance on the first sortie following the ground alert. To reduce some of the random variations in the data, several rules and restrictions were imposed on test sorties:

(1) The sortie followed ground alert was required to take place within 12 hours of the end of the ground-alert period, and with no preventive maintenance prior to flight.

(2) During the alert period, minor maintenance could be performed as long as the aircraft was not removed from alert status.

(3) The sortie flown immediately following a test-alert period had to consist of at least two attempted intercepts.

(4) At least three sorties had to be flown between consecutive test-alert periods for a given aircraft.

(5) Whenever an aircraft flew a sortie following a test-alert period, it carried electronic evaluators in the weapons rails.

To control nonrandom variation between aircraft, the experimental design called for every aircraft to be placed on alert for each of the four alert periods. (Previous studies at both Rand* and ADC** indicate that significant differences exist between aircraft for short periods of time (several months). The order in which a given aircraft was placed in each of the alert periods might possibly have an effect, i.e., the alert period sequence of 2-4-6-8-days might have an effect different from the sequence of 8-4-6-2-days. This kind of order effect could be controlled by counterbalancing so that all

---

*W. H. McClothlin and T. S. Donaldson, Trends in Aircraft Maintenance Requirements (For Official Use Only), The Rand Corporation, RM-4049-PR, (DDC No. AD 447-880), June 1964; T. S. Donaldson and Anders Sweetland, Trends in F-101 Maintenance Requirements, The Rand Corporation, RM-4930-PR, April 1966 (For Official Use Only).

**Unpublished study by Captain John B. Abell, Headquarters Air Defense Command (ADMLP).

sequences occur with equal frequency. To have conducted the experiment with counterbalancing, however, would have imposed severe restrictions on aircraft scheduling. Since these order effects were expected to be small, they were controlled by randomization. This order randomization was carried out by ADC Headquarters.

Since there were two bases for each aircraft, the design might be thought of as a three-way layout involving three factors: alert period, aircraft type and replications (bases). The experimental design is shown in Table 2. The design suggests a 4 x 3 x 2 analysis of variance, with the within-cell error variance partitioned into a between-aircraft and residual error.* Differences between aircraft type and between replications would be tested against the between-aircraft error, and differences between alert periods would be tested against the residual error. This type of analysis with the kind of data observed in the study (primarily dichotomous) and with conservative sample sizes would be extremely risky since the true Type I and II errors would not be known. This is particularly true with respect to the interpretation of interaction effects, which, if present, would then require an analysis of one effect at each level of one of the other effects ("simple effects" analysis). Rather than follow this "classical" analysis, the above comments suggest that differences between alert periods should be analysed for each base separately. Pooling across bases would follow, whenever possible, to give more powerful tests. This is the analytic procedure used in this study.

---

*For example, see B.J. Winder, Statistical Principles in Experimental Design, McGraw-Hill Book Company, Inc., New York, 1962, Chap. 7

Table 2

EXPERIMENTAL DESIGN

Aircraft	Base No.	Alert Period (days)[a] 2	4	6	8
F-101	1				
	2				
F-102	3				
	4				
F-106	5				
	6				

[a]Measures are repeated on this factor.

Aircraft reliability was assessed by the following indices:

(1) Break rate by category: An aircraft is placed in one of three categories after a sortie.* These categories and codes are:

Code 1. The aircraft is operationally ready, i.e., all installed systems are fully operational.

Code 2. The capability of the aircraft to perform on air defense mission is degraded.

Code 3. The aircraft is not capable of performing an air defense mission.

The proportion of Code 2 or Code 3 sorties relative to the total number of sorties flown (or sum of Codes 1, 2, and 3) is the break rate for that respective code.

(2) MA success rate: MA denotes mission accomplished; it requires that a fire signal must register on the pilot's display scope and the target dot be in the center of the radar screen.

*ADC Regulation 66-28, 7 October 1964, p.7.

(3) WSEM/MSR success rate: This refers to the success of the aircraft systems in performing a mission as measured by the WSEM and MSR electronic evaluators.

(4) Number of aborts on ground alert; A ground abort refers to a cancelled sortie due to a critical failure occurring before takeoff. (Weather and operations ground aborts were not considered).

(5) Number of in-flight emergencies: An in-flight emergency occurs when the mission must be cancelled as the result of a critical failure while in flight.

(6) Number of discrepancies per sortie: This refers to the total number of malfunctions reported by the aircrew.

(7) Maintenance on the aircraft: This is defined as the unscheduled flight-line maintenance as recorded on the AFM 66-1, AFTO 210 and 211 forms. Units produced and maintenance manhours were used to measure the amount of maintenance required by the aircraft.

The first six of these measures are taken from information on ADC Form 76-3 (pilot debriefing). This form is made out after every sortie; it was modified for the present study to include a code indicating which test alert period the wortie followed. The seventh measure was obtained from AFM 66-1 forms. Unfortunately, these forms were not identified by sortie, making the resultant analysis weaker than it might otherwise have been.

IV. RESULTS

Sorties were attempted following 440 ground alerts. The number of alerts for each base and for each alert period are shown in Table 3. The experimental design required that every aircraft receive all treatments (length of ground alert), and that the order in which an aircraft was exposed to the treatments be random. Inspection of the actual dates that aircraft in the study ended their alert periods indicated that no biases in the order of treatment presentation were present, and experimental differences due to order effects should be absent (or at least extremely small). From this, we conclude that the bases were able to follow the scheduling requirements, and that in general, the experimental design was followed.

Table 3

DISTRIBUTION OF SAMPLE FOR BASES AND TREATMENTS

	Alert Period (days)				
Base	2	4	6	8	Total
1	15	17	14	14	60
2	16	16	16	15	63
3	22	22	22	22	88
4	20	18	20	17	75
5	21	21	22	22	86
6	17	17	17	17	68
Total	111	111	111	107	440

It was obvious by inspection of the means for each base that differences of a meaningful sort did not exist between alert periods. The actual data are presented in the study by Donaldson and Burke (1966), but are omitted in this paper because of their security classification. The data however did not reveal any increase in any of the indicators as a function of the length of ground alert, and in fact, more often than not the longer alerts indicated better aircraft performance. None of the differences however were significant when tested with χ^2.

The data across bases were pooled in order to increase the power of the tests, but again no significant differences were found.

A compilation arose in the analysis when it was discovered that some aircraft were sheltered and some were not. The two F-106 bases sheltered all aircraft, and these presented no problem: However, the other four bases sheltered some of their aircraft, but not all. An analysis of sheltered vs. nonsheltered aircraft showed a highly significant difference in all but one of the 76-3 measures in favor of the sheltered aircraft. The one measure which did not show this effect was the WSEM/MSR, although it did show that sheltered aircraft on long alerts were significantly better than those on shorter alerts. This was thought possibly to be due to drying out in the shelters.

Analysis of the 66-1 unscheduled maintenance man-hour data indicated that during long alerts, aircraft require only slightly more maintenance than during short ones. The cost of this increase is negligible. Further, there was no indication of a differential cost as a function of alert duration for a period of up to three days following the alert.

The results of the study clearly indicated that aircraft left on alert for eight days were as mission capable as those left on alert for two days.

V. CONCLUSIONS

This study is an example of how experimental controls can be used in an operating environment. A field test of this sort requires that the initial problem be well defined, and that the design of the test is clearly and simply related to the problem. Too often test designs become unwieldly and overly complex because the design incorporates too many operations. A field test of this kind is designed to obtain relatively precise information about a specific problem and a few variables rather than uncertain and confounded information about a wide ranging problem.

In all respects this study was considered to be a success, and ADC policy was changed within a few months of completion of the study.

What we have been talking about so far may be thought to fall under the general heading of the classical approach to experimental design. This approach derives mainly out of the thinking of R. A. Fisher during the early part of this century, and even until the present day it dominates the teaching in the areas not only of experimental design but also of statistics. Other approaches are beginning to emerge, however, and we should be aware of possibilities for different experimental strategies which are available to us because of these new ways of thinking. We will just mention some of these briefly in conclusion, without giving any specific examples.

The first of the two approaches which we want to mention we will call the Decision-Theoretic or Bayesian approach. We do experiments or field tests because we need to gather evidence for some decision we have to make, but we generally do not, as the classical theory would have us believe, go into the field or start an experiment in a state of total ignorance. We do have some prior opinions about what is going to happen and we want to collect data so we can upgrade those opinions. That is what the Bayesians are all about - modifying prior opinions (generally stated in terms of probabilities) by means of new evidence. They get the name Bayesians because the formally correct way of adjusting prior probabilities is by applying Bayes' Theorem, but a more distinctive

attribute of this group is the use they make of what has come to be called subjective or personal probability. The personal probabilities are concerned with the impact of each item of data that is gathered on the prior opinions about the hypotheses, and there is concern among this group with developing optimal stopping procedures for the collecting of information. In the decision theoretic sense, there is no point in collecting any more data when the cost of information exceeds the expected gain it reflects in the payoff. There is an analogy here to the optimal stopping techniques used in Wald's sequential testing procedures, except in the Bayesian analysis there is no concern with the null-hypothesis testing. The Bayesians, in fact, consider the null-hypothesis concept to be an artificial constraint which acts much to the detriment of most experimental design efforts.

Another, even further out, technology seems to be emerging in areas which are concerned primarily with the evaluation of programs in such fields as education, welfare, and social reforms. There is as yet no collective name for the procedures which are being developed, but we can refer to them under the designation of Environmental or Situational Theory. The distinctive feature of this activity is that the evaluation is generally done in the absence of anything we might call a true control group. Educational changes, for example, are so all pervading that the variables involved in studying new classroom techniques must be concerned not only with the techniques themselves, but also with the group selected for study, the teachers who administer the program, and the evaluation team which must gather performance measures for the program. The best we can do in cases like this is make an estimate of how the group would have performed under the old methods and compare this estimate against the results obtained from the innovations.

Marcia Guttentag of the City University of New York, in an as yet unpublished paper, advocates a legal model for the evaluation of research done in the field in areas where the institution of tight controls is all but impossible. This model would take as its method of ascertaining the "truth," the way in which our legal system operates. There are rules which govern the presentation and evaluation of evidence, individual cases are argued in terms of whether they do or do not fit a

particular precedent, with a citation of the characteristics of the case which leads to the inference that it falls under the precedent. Counter arguments are offered by the opposing side to show why such inferences are not tenable and other plausible alternative interpretations of the data are offered. Evidence is classified according to whether it is direct or circumstantial and assertions of fact are separated from direct evidence. The jury, or in this case the experimentor or the evaluation team, comes to a finding based on the "preponderance of evidence." We can look at this procedure as a new form of an experiment in which the experimental variables might be considered a form of advocacy <u>for</u> a certain position while the controls can be looked at as advocacy for <u>other</u> positions. This is all to say that there may be some quite reasonable alternatives to the classical experimental method for drawing inferences from controlled observations and these methods do not necessarily leave us in a mire of subjective chaos.

Let us consider one final example of an experimental or scientific method which, while far from the classical paradigm, still has a logic and rigor about it lending it significant respectability. The field is anthropology or archaeology. The scientist starts out in the field with only a vague notion of what he is looking for. When he finds something, he formulates hypotheses about the totality of what is there which gives him not only a picture of the history he is uncovering but also some direction of where and what to look for as he continues. His next finding may reinforce or negate his previous hypotheses, and he continues in this cycle of looking, finding, and formulating until a consistent picture emerges. An experiment in the classical sense is never formulated, but clear and unambiguous results often emerge from the process.

There is some indication from other disciplines, therefore, that something is to be gained by abandoning a rigid adherence to classical experimental methodology. The example we presented earlier represented an extension, a stretching, of the classical methods to meet the contingencies of experimentation under field conditions. With the world becoming more complex, it is incumbent upon us to expand our imaginations even further to meet the demands we will have to face.

A STATISTICAL ANALYSES OF DYNAMIC RESPIRATORY DATA GENERATED THROUGH PROTECTIVE MASK WEAR

Edwards N. Fiske
Systems Analysis Office
Edgewood Arsenal, Maryland

I. INTRODUCTION. For years Defense Development and Engineering Laboratories at Edgewood Arsenal desired a means of measuring the dynamic performance of a protective mask, specifically, breathing resistance under field conditions. A system, designed and built by Sanborn-Statham, has been used with reasonable success in the laboratory; however, this system is too heavy and bulky for field use. After several designs had been submitted for acceptance, Edgewood Arsenal decided to develop a system called Dynaper, manufactured by Space Laboratories, Inc., Van Nuys, California. The system consists of the following units: (1) a backpack having a transmitter and a rechargeable battery; and (2) a ground receiving console.

When Edgewood Arsenal Physical Protection Laboratory received the Dynaper system, a series of tests were conducted to determine its reproductibility over repeated runs and to determine its accuracy for measuring mask resistance as compared to that of the Sanborn laboratory system.

The results of these tests showed that there is a 95 percent chance that the Dynaper system will measure a given pressure drop or flow rate within 1 percent for repeated runs. The difference between the Dynaper and the Sanborn system over the flow range of -250 to +250 liters per minute averages approximately 3 percent. The results revealed that the Dynaper system is an acceptable system for field use.

Since the preliminary tests on the Dynaper system were considered a success, the Physical Protection Laboratory initiated a test program to measure the mask resistance of three protective masks under both field and laboratory conditions. The three masks tested are designated as the M17, XM28, and the M17A1. The test program was statistically designed by the author.

II. METHODOLOGY. A 5x5 factorial design was chosen for the test program. The two factors were flow rate and breathing systems. Flow rates of 50, 100, 150, 200, and 250 liters per minute were chosen. The following systems were utilized: (1) a reciprocating piston-type engine called a breathing machine attached to the Sanborn-Statham recorder; (2) the breathing machine attached to the Dynaper backpack; (3) a human subject on an ergometer carrying

a Dynaper backpact wired to the ground receiving console in the laboratory; (4) a human subject exposed to field conditions carrying a Dynaper backpack containing a radio transmitter that transmits impulses to the ground receiving console in the laboratory; and (5) a human subject on a ergometer attached to the Sanborn-Statham recorder. The first four devices are considered the four basic systems that were compared to the standard, the fifth system, for the statistical analyses. Three separate experiments were conducted consisting of 30 dynamic runs covering a continuous flow range of -250 to +250 liters per minute were performed and the flow rates and pressure drops were recorded on strip charts. The data on the strip chart were reduced to peak pressure drops at the five flow rates chosen. Since the early tests indicated that variation between runs was not significantly different, data extracted from the strip charts were averaged over runs (Table 1 and 2).

III. ANALYSIS. Analyses of variance (Table 3 and 4) carried out on the data revealed that there is a significant variation between the five measuring systems for each of the three masks, and inhalation and exhalation respectively. Therefore, regression analyses were performed on the data to determine the functional relationship between the standard and the four basic systems. The regression analyses revealed that there is a linear function, $Y = MX$, between the pressure drop value for each of the basic systems (X) and the pressure drop value for the standard system (Y). The additional variation, that would be explained by a higher order fit over the linear fit is not significant. The M values for the model $Y = MX$ are listed in Table 5. These M values are all significantly different from one.

Additional regression analyses were carried out to determine the functional relationship, $Y = MX$ between the two breathing machine (X) systems (Dynaper and Sanborn) and their respective human subject systems (Y) for each of the three masks. These regression analyses revealed that the curves for both comparisons were not significantly different. Hence their respective M values were pooled for each of the three masks. These values were also found to be significantly different from one. These values are listed in Table 6.

IV. DISCUSSION. The analyses revealed that the Dynaper field system, on the average, measures values 1 percent higher than the Dynaper laboratory system would measure but the overall difference between values measured by the Dynaper laboratory system and the Sanborn system is 2 percent.

When the data were plotted the mask resistance curves indicated that the relationship between the pressure drop and flow rate might have a second order fit for all masks and measuring systems; but the analyses indicate the curves

were linear, since the coefficient of determination (r^2) for pressure drop and flow rate, which represents the total amount of variation explained by a linear curve, is greater than 98 percent.

Table 5 can be used to evaluate the performance of other protective masks with the same characteristics as either the XM28, M17A1, or M17 masks. If one designs a mask with the same characteristics as the XM28 mask, one would use the figures listed under Dynaper Human subject (laboratory) system to determine what the standard would be. For example, if the Dynaper system measures a pressure drop of 90mm of water across the mask (inhalation) the standard pressure drop would be 87.3 [(0.97) x (90)] mm of water.

Table 6 can be used to convert a breathing system reading to its corresponding human subject system reading. For example, if a mask similar to the M17 is fitted to the mechanical breathing head set to produce a flow rate of 160 liters per minute, and the Dynaper system measures the pressure drop across the mask to be 80mm of water, the same Dynaper backpack if carried by a human subject breathing at 160 liters per minutes through the particular mask, would measure a pressure drop of 96 [(120) x (80)] mm of water.

VI. CONCLUSION. It can be concluded from the test program that: (1) The Dynaper system is a reliable means of measuring the dynamic performance of any protective mask exposed to field conditions; (2) the system can measure the mask resistance of a mask under field conditions within 3 percent of the mask resistance determined in the laboratory with the standard equipment; (3) the system has a 95 percent chance of measuring a given mask resistance within 1 percent for repeated runs; (4) the mask resistance curves for each mask are linear; (5) the mask resistance for the XM28 is significantly higher than for either the M17A1 or M17 mask; and (6) the factors established to correlate field data with data generated in the laboratory can be used to evaluate the performance of other protective masks with the same characteristics as the three masks tested

Table 1

MEAN PRESSURE DROP AT SEVERAL FLOW RATES, M17, XM28, M17A1 PROTECTIVE MASKS, FIVE MEASURING SYSTEMS-INHALATION

Mask Type	Flow Rates (Liters/Min) / Measuring Systems	50	100	150	200	250
		mm of Water				
M17	Sanborn-Breathing Machine	22	47	75	105	143
	Dynaper-Breathing Machine	24	50	80	112	147
	Dynaper-Human Subject (Laboratory)	30	63	98	138	180
	Dynaper-Human Subject (Field)	27	57	89	124	163
	Sanborn-Human Subject	30	60	93	130	173
XM28	Sanborn-Breathing Machine	25	55	88	125	165
	Dynaper-Breathing Machine	29	60	96	135	175
	Dynaper-Human Subject (Laboratory)	30	65	105	145	190
	Dynaper-Human Subject (Field)	33	63	103	143	187
	Sanborn-Human Subject	25	56	93	134	180
M17A1	Sanborn-Breathing Machine	24	50	81	118	153
	Dynaper-Breathing Machine	26	58	86	123	162
	Dynaper-Human Subject (Laboratory)	29	61	96	135	175
	Dynaper-Human Subject (Field)	33	65	106	149	163
	Sanborn-Human Subject	31	65	101	145	190

Table 2

MEAN PRESSURE DROP AT SEVERAL FLOW RATES, M17, XM28, M17A1 PROTECTIVE MASKS, FIVE MEASURING SYSTEMS-EXHALATION

Mask Type	Flow Rates (Liters/Min) / Measuring Systems	-50	-100	-150	-200	-250
		mm of Water				
M17	Sanborn-Breathing Machine	-10	-22	-30	-40	-48
	Dynaper-Breathing Machine	-15	-27	-37	-44	-50
	Dynaper-Human Subject (Laboratory)	-10	-21	-30	-37	-43
	Dynaper-Human Subject (Field)	-12	-23	-31	-40	-45
	Sanborn-Human Subject	- 9	-20	-28	-35	-40
XM28	Sanborn-Breathing Machine	-10	-20	-30	-40	-50
	Dynaper-Breathing Machine	-11	-24	-33	-43	-53
	Dynaper-Human Subject (Laboratory)	-10	-19	-28	-36	-46
	Dynaper-Human Subject (Field)	-11	-21	-31	-41	-51
	Sanborn-Human Subject	-12	-22	-31	-40	-49
M17A1	Sanborn-Breathing Machine	-10	-23	-40	-60	-85
	Dynaper-Breathing Machine	-13	-30	-48	-69	-93
	Dynaper-Human Subject (Laboratory)	-12	-25	-41	-58	-76
	Dynaper-Human Subject (Field)	-11	-24	-38	-55	-73
	Sanborn-Human Subject	- 9	-23	-37	-54	-68

Table 3

ANALYSES OF VARIANCE OF DATA IN TABLE 1

M17 Mask

Source of Variation	dF	SS	MS	F Ratio*
Between Flows	4	56528.0	14132.0	552.9
Between Systems	4	1918.0	479.5	18.7
Error	16	410.0	25.6	-
Total	24	58856.0		

XM28

Source of Variation	dF	SS	MS	F Ratio*
Between Flows	4	71903.6	17975.9	166.4
Between Systems	4	798.8	199.7	18.5
Error	16	173.6	10.8	-
Total	24	73876.0		

M17A1

Source of Variation	dF	SS	MS	F Ratio*
Between Flows	4	67274.8	16818.7	535.6
Between Systems	4	2050.4	512.6	16.3
Error	16	512.8	31.4	-
Total	24	69838.0		

Table 4

ANALYSES OF VARIANCE OF DATA IN TABLE 2

M17 Mask

Source of Variation	dF	SS	MS	F Ratio*
Between Flows	4	3626.2	906.6	647.6
Between Systems	4	186.6	46.7	33.3
Error	16	21.8	1.4	-
Total	24	3834.6		

XM28 Mask

Source of Variation	dF	SS	MS	F Ratio*
Between Flows	4	4686.6	1171.7	1065.2
Between Systems	4	65.8	16.4	14.9
Error	16	17.8	1.1	-
Total	24	4770.2		

M17A1

Source of Variation	dF	SS	MS	F Ratio*
Between Flows	16	14556.4	3639.1	273.6
Between Systems	16	446.8	111.7	8.4
Error	4	212.8	13.3	-
Total	24	15216.0		

*All F Ratios significant at 95% level.

Table 5

M VALUES FOR RELATING PRESSURE DROPS MEASURED BY FOUR BASIC MEASURING SYSTEMS (X) TO THOSE MEASURED BY THE STANDARD SYSTEM (Y) THROUGH THE FUNCTIONAL RELATIONSHIP, Y = MX

Breathing Method	Mask Type	Sanborn Breathing Machine	Dynaper Breathing Machine	Dynaper-Human Subject (Laboratory)	Dynaper-Human Subject (Field)
Inhalation	M17	1.19	1.16	0.95	1.05
	XM28	1.11	1.06	0.97	1.00
	M17A1	1.22	1.18	1.09	0.99
Exhalation	M17	0.82	0.87	0.94	0.93
	XM28	0.92	0.89	1.03	0.92
	M17A1	0.80	0.75	0.92	0.95

Table 6

M VALUES FOR RELATING PRESSURE DROPS MEASURED BY THE BREATHING MACHINE (X) TO THOSE MEASURED BY THE HUMAN SUBJECT SYSTEM (Y) THROUGH THE FUNCTIONAL RELATIONSHIP (Y = MX)

Breathing Method	Protective Masks		
	M17	XM28	M17A1
Inhalation	1.20	1.10	1.14
Exhalation	0.87	0.89	0.81

A COMPLEX SPLIT PLOT DESIGN FOR AN EXPERIMENT INVOLVING STANO SMALL UNIT LIVE FIRE

James S. De Gracia, G. David Faulkenberry, Tima R. Rodgers
SP5 Harvey Bunce, III, U.S. Army, CDCEC
Litton Scientific Support Laboratory
Mellonics Systems Development Division
Litton Industries
Fort Ord, California

1. GENERAL DISCUSSION OF EXPERIMENT.

a. Objective of the Experiment. The objective, with respect to performance, of the STANO Small Unit Live Fire Experiment was to obtain fire effectiveness data of a small unit firing night live fire with selected STANO devices under varying levels of illumination, varying terrain, employing various types of ammunition (including tracer) against representative personnel targets. This objective was then paraphrased into a series of questions, termed Essential Elements of Analysis (EEA), which could be answered analytically. The main EEA for this experiment was: Does a ten man fifle squad reinforced with a 3 man machinegun section have a significant increase in fire effectiveness when using one test mix of STANO devices as compared to its fire effectiveness when using another test mix, or when using no night vision aids?

(1) In natural low light?
(2) In natural mid light?
(3) In natural high light?

To answer this EEA four variables were selected to be analysed. These variables, termed Measures of Effectiveness (MOE), were:

(1) MOE 1: THTP - Target Hit per Targets Presented. The proportion of targets presentod that were hit measures the target effects achieved and therefore is the basic measure of fire effectiveness.

(2) MOE 2: TFE - Time to First Effect. The time from the appearance of the first target in any array to the registration of the first hit of near miss on any target in that array. This MOE measures the acquisition time plus the time required to achieve a target effect and thus will provide an indication of the relative target acquisition time.

(3) MOE 3: FRPET - Fractional Reduction of Programmed Exposure Time. Reduction of Programmed Exposure Time/Adjusted (for malfunctions) Programmed Exposure Time (RPET/APET). The proportional reduction of programmed exposure time due to target hits measures the timeliness of the effective fire of the squad and reflects the number of targets that were hit as well as the time to hit.

(4) MOE 4: THSAE - Target Hits per round of Ammunition Expended. The average single round hit probability is a basic measure of the squad's efficiency in achieving target effects.

b. Field Experiment. The field experiment was conducted at Hunter Ligget Military Reservation. Two target arrays were utilized during each trial. A total of three events were conducted to simulate the movements of the enemy. The appearance of each target array was accomplished in a two phase operation. The use of this phasing allowed the target arrays to be exposed initially to test the firer's ability to detect targets at varying ranges. Firers were instructed to fire at targets once detected. If the targets were not detected during the initial phase, a period of "forced detection" was included to enable the evaluators to determine the fire units' ability to place and adjust fire on a detected target. The phasing of the targets was as follows:

(1) Phase I. This phase was a silent period in which targets appeared at approximately three second intervals until they were all exposed. It was during this phase that target detection capability was determined.

(2) Phase II. This phase was initiated by a forced detection sequence in which selected targets were programmed to fire their simulator. This insured that the targets had been detected so that ability to place and adjust fire could be measured. The phase consisted of both target exposure and simulator fire. Targets during this phase were programmed to move up and down at different intervals while displaying varying degrees of simulator fire. The amount of target exposure and simulator fire were based on the distance the targets were located from the friendly element in an attempt to approach combat realism within the constraints of the experimental design.

(3) During Phase I, if a firer was able to detect a target, to fire, and to obtain a hit or near miss, the Phase was terminated and Phase II was automatically initiated. If a hit or near miss did not occur, Phase II was initiated automatically upon completion of the time programmed for Phase I.

c. Design Considerations. In planning statistical aspects of the experimental design, factors and constraints considered were the following:

(1) Basis of Issue (BOI) The main object of the experimental design was to allow comparisons of selected STANO device mixes in a rifle squad, reinforced with a 3 man machinegun section, in terms of their fire effectiveness. The different mixes of STANO devices, BOI's, were determined on the basis of military rationale so that reasonable differences in the BOI's could be detected.

(2) Light Levels. Light intensity was of primary concern in the BOI comparisons. Rather than carry light as a continuous covariate, it was considered easier analytically and operationally to categorize it into three levels as follows:

High Light:	Ambient light $\geq 3.0 \times 10^{-3}$ footcandles
Mid Light:	2.5×10^{-4} footcandles $\leq$ Ambient light $< 3.0 \times 10^{-3}$
Low Light:	Ambient light $< 2.5 \times 10^{-4}$ footcandles.

In the time frame set for the experiment the first mid light occurred on the evening of 10 July and ended 28 July, high light began 15 July and ended 24 July, and consistent low light first occurred 26 July and continued into the first part of August. It was evident from this that simple randomization of light levels over trials could not have been done.

(3) Time. Since light level occurrence was dictated as noted above, the completion of the experiment was restricted. If the record trials were not completed in the time allotted the experiment would have to have been extended until the needed light level reoccurred. Results were also needed as soon as possible to assist in a decision relating to the production of the selected STANO devices under test.

(4) Squads. Drawing upon available personnel, it was decided that 12 squads of 13 men would be used in the experiment. An attempt was made to balance the squads with respect to military background (e.g. combat versus non combat) and human factor considerations (e.g. players wearing glasses versus not wearing of glasses). Although the attempt was made to balance the squads, it was realized from previous experimentation that large variation in the performance of the squads could occur.

(5) Sample Size. Considering the above factors and constraints, it was decided that four BOI's would be used in the experiment, appearing with each of the three light levels. The basic experimental matrix then consisted of 12 cells. The need for 12 cells in the experimental matrix, as well as results from previously conducted night live fire experiments, dictated a sample size of 12 observations per cell of the design matrix in order to obtain reasonable probabilities of Type I and Type II errors. The number of light levels and BOI's and the sample size determined the experimental matrix below:

Light Level	BOI 1	BOI 2	BOI 3	BOI 4
Low	12	12	12	12
Mid	12	12	12	12
High	12	12	12	12

(6) Learning/Motivation. Because of the above time, player, and sample size considerations, it was decided that the main experiment would begin on 10 July and trials would be conducted every night until completion on 30 July. The number of players and the time available to complete the experiment dictated that the players would be used repeatedly for a period of time which included weekends. Twelve trials were scheduled for each squad. It was thought that, with the players continually repeating the trials, they would learn when and where to fire to obtain a target hit. It was also anticipated that the players' motivation to perform well would decline as the number of repeated trials increased. These carry-over effects (learning and motivation), which would have been counfounded with light levels and/or BOIs if a simple randomization of BOIs over or within light levels had been used, were a major consideration in the selection of the statistical design.

2. DESIGN STRUCTURE - MODEL.

a. Design Layout. The statistical design selected for this experiment was basically a split-plot design where light levels corresponded to whole plots. Twelve squads were formed and randomly divided into three groups of four squads each. One group fired only during low light, another during mid light only, and the third group only at high light. Within a light level, three factors were examined. These were BOI/AMMO Mix combinations, Squads, and Blocks. For this experiment, a block was defined as a group of four trials in which each squad fired once and each BOI/AMMO Mix combination was used. The three factors were run in three Latin Square arrangements where each squad fired each BOI/AMMO Mix combination three times for a total of twelve trials per squad. The first Latin Square under mid light had squads 1, 2, 3, and 4 as rows, blocks 1, 2, 3, and 4 as columns, and BOIs 1, 2, 3, and 4 as the randomly assigned treatments within the square. The second Latin Square under the same light level had the same squads as rows, blocks 5, 6, 7, and 8 as columns and the same BOIs again randomly placed in the rows and columns. The third Latin Square under the mid light level continued in the same manner. This method of randomization was used for each light level, squads 5, 6, 7 and 8, being used for high light and squads 9, 10, 11 and 12 being used for low light. It should be noted that since the block number (1 through 12) corresponded to the number of times each squad fired in a given light level, this block effect was closely related to both motivation and learning. The experiment was expressly designed to balance block and squad differences in order to minimize their counfounding effect on the BOI and light level comparisons. Additionally, the design enabled the compilation of useful information as to the nature and extent of the block and squad differences.

b. Model. The statistical model for the basic design was

$$y_{ijkl} = \mu + L_i + S_{j(i)} + B_k + BA_l + (BxL)_{ik} + (BAxL)_{il} + \varepsilon_{ijkl}$$

where

y_{ijkl} = observation corresponding to the ith light level, jth squad in the ith light level, kth block, and lth basis-of-issue/ammo mix combination

μ = true overall mean

L_i = ith light level effect, i = 1, 2, 3

$S_{j(i)}$ = jth squad effect in ith light level, j = 1, 2, 3, 4

B_k = kth block effect, k = 1, 2, ... , 12

BA_l = lth basis-of-issue/ammo mix combination effect, l = 1, 2, 3, 4

$(BxL)_{ik}$ = block by light level interaction

$(BAxL)_{il}$ = basis-of-issue/ammo mix combination by light level interaction

ε_{ijkl} = random error corresponding to the y_{ijkl} observation,

and

$$\sum_{i=1}^{3} L_i = 0; \; S_{j(i)} \sim N(0, \sigma_s^2); \; \sum_{k=1}^{12} B_k = 0$$

$$\sum_{l=1}^{4} BA_l = 0; \; \sum_{i} (BxL)_{ik} = \sum_{k} (BxL)_{ik} = 0;$$

$$\sum_{i} (BAxL)_{il} = \sum_{l} (BAxL)_{il} = 0;$$

$$e_{ijkl} \sim N(0, \sigma_e^2); \; E(S_{j(i)}) = E(e_{ijkl}) = 0.$$

It was felt that this model would fit the observed data reasonably well and that the underlying assumptions of the model were justifiable.

3. ANALYSIS.

a. Residual Analysis. The model was fit to the observed data for each of the MOE's. ($Y = \arcsin \sqrt{x}$ was used for THSAE). Analysis of the residuals showed the model and corresponding assumptions to be reasonable.

b. Analysis of Variance. The analysis of variance table used in the analysis of the observed data corresponding to the model described above was:

Source of Variation	Degrees of Freedom	Mean Square	Expected Mean Square
Total	143	$\frac{1}{143} \sum_{ijkl} (y_{ijkl} - \bar{y}_{....})^2$	
Light Levels	2	$\frac{1}{2} [48 \sum_{i} (\bar{y}_{i...} - \bar{y}_{....})^2]$	$\sigma_e^2 + 12\sigma_s^2 + 48 (\sum_i L_i^2/2)$
Squads Within Light Levels	9	$\frac{1}{9} [12 \sum_{ij} (\bar{y}_{ij..} - \bar{y}_{i...})^2]$	$\sigma_e^2 + 12\sigma_s^2$
BOI/Ammo Mixes	3	$\frac{1}{3} [36 \sum_{l} (\bar{y}_{...l} - \bar{y}_{....})^2]$	$\sigma_e^2 + 36 (\sum_l BA_l^2/3)$
Blocks	11	$\frac{1}{11} [12 \sum_{k} (\bar{y}_{..k.} - \bar{y}_{....})^2]$	$\sigma_e^2 + 12 (\sum_k B_k^2/11)$
BAxLL (Interaction)	6	$\frac{1}{6} [12 \sum_{il} (\bar{y}_{i..l} - \bar{y}_{i...} - \bar{y}_{...l} + \bar{y}_{....})]$	$\sigma_e^2 + 12 (\sum_{il} (BAxLL)_{il}/6)$
BxLL (Interaction)	22	$\frac{1}{22} [4 \sum_{ik} (\bar{y}_{i.k.} - \bar{y}_{i...} - \bar{y}_{..k.} + \bar{y}_{....})^2]$	$\sigma_e^2 + 4 (\sum_{ik} BxLL)_{ik}/22)$
Error	90	$\frac{1}{90}$ (residual SS)	σ_e^2

The means and variances of respective cells and cell differences were as follows:

(1) Mean of observations at ith light level, lth BOI/Ammo mix

$$\bar{y}_{i..l} = \mu + L_i + \bar{S}_{(i)} + BA_l + (BAxL)_{il} + \bar{e}_{i..l},$$

$$\text{Var}\ (\bar{y}_{i..l}) = \frac{1}{12}\ (\sigma_e^2 + 3\sigma_s^2)$$

(2) Difference in two cell means (different BOI/Ammo mix) at same light level ($l \neq l'$)

$$\bar{y}_{i..l} - \bar{y}_{i..l'} = (BA_l - BA_{l'}) + [(BAxL)_{il} - (BAxL)_{il'}]$$

$$+ (\bar{e}_{i..l} - \bar{e}_{i..l'})$$

$$\text{Var}\ (\bar{y}_{i..l} - \bar{y}_{i..l'}) = \frac{2\sigma_e^2}{12}$$

(3) Difference in two cell means (different BOI/Ammo mix) at different light levels ($i \neq i'$, $l \neq l'$)

$$\bar{y}_{i..l} - \bar{y}_{i'..l'} = (L_i - L_{i'}) + (BA_l - BA_{l'}) + [(BAxLL)_{il}$$

$$- (BAxLL)_{i'l'}] + (\bar{S}_{(i)} - \bar{S}_{(i')}) + (\bar{e}_{i..l} - \bar{e}_{i'..l'})$$

$$\text{Var}\ (\bar{y}_{i..l} - \bar{y}_{i'..l'}) = \frac{1}{6}\ (\sigma_e^2 + 3\sigma_s^2)$$

(4) Mean of BOI/Ammo mix averaged over light levels

$$\bar{y}_{...l} = \mu + \bar{S}_{.(.)} + BA_l + \bar{e}_{...l}$$

$$\text{Var}\ (\bar{y}_{...l}) = \frac{1}{36}\ (\sigma_e^2 + 3\sigma_s^2)$$

(5) Mean difference of same BOI/Ammo mix at different light levels ($i \neq i'$)

$$\bar{y}_{i..l} - \bar{y}_{i'..l} = (L_i - L_{i'}) + (\bar{S}_{(i)} - \bar{S}_{(i')}) + [BAxLL)_{il}$$

$$- (BAxLL)_{i'l}] + (\bar{e}_{i..l} - \bar{e}_{i'..l})$$

$$\text{Var}\ (\bar{y}_{i..l} - \bar{y}_{i'..l}) = \frac{1}{6}\ \ (\sigma_e^2 + 3\sigma_s^2)$$

4. RESULTING ADVANTAGES OF THE DESIGN USED. The ability of the design to make use of the Latin Square arrangement, where squads and blocks (measure of learning/motivation) formed the rows and columns of the square, was an important advantage of its selection. Extending blocks over the three squares in a given light level gave the advantage of yielding information on learning/motivation over a greater period of time. The design then, not only reduced the error mean square by removing squads within light levels and blocks as sources of error but also gave insight into the variability among squads and, most important, it gave information as to the effect of learning/motivation on players who were used over extended periods of time. Another advantage of this design was that it lent itself to continuous analysis of the data. Upon the completion of a square, analysis could be done and in this way the data could be monitored and analyzed in successive steps.

REFERNCES

1. W. G. Cochran and G. M. Cox, Experimental Designs. John Wiley and Sons, Inc., New York (1957).

2. D. R. Cox, Planning of Experiments. John Wiley and Sons, Inc., New York (1958).

3. Allen, L. Edwards, Experimental Design in Psychological Research. Holt, Rinehart, and Winston, Inc., New York (1968).

4. E. F. Lindquist, Design and Analysis of Experiments in Psychology and Education. Houghton Mifflin Company, Boston (1953).

5. B. J. Winer, Statistical Principles in Experimental Design. McGraw-Hill Book Company, New York (1962).

EMPIRICAL BAYES ESTIMATORS FOR SOME TIME SERIES PARAMETERS

Robert L. Launer
US Army Logistics Management Center
Fort Lee, Virginia

Richard G. Krutchkoff
Virginia Polytechnic Institute
and State University
Blacksburg, Virginia

Summary

When estimating the parameters of a time series, one ordinarily has available an entire set of time series, or a multiple series. Some obvious examples are economic and aerologic series. This paper treats the situation in which the time series of interest is one of a set of observed series whose parameters are "generated" independently by a random mechanism, i.e., independent realizations of a random variable θ with unknown (prior) distribution. This suggests the Empirical Bayes approach to the estimation problem. The specific models studied are the first order auto-regressive process with zero mean, and the time series regression model with auto-correlated error term. Although exact Empirical Bayes estimators do not exist for the various situations, approximate Empirical Bayes estimators are developed and used to obtain smaller mean squared error than with the usual estimators.

1. Decision Theory: Bayes and Empirical Bayes

Let us suppose that we are interested in estimating a parameter θ, and that this estimation situation occurs repeatedly for different parameter values. Further, suppose that the specific (but unknown) values of θ are independent realizations of a random variable θ, with distribution $G(\theta)$. The observations $x_1, x_2, \ldots, x_n$ which are used to estimate θ follow the conditioned distribution $f(x|\theta)$. The decision (in this case our estimate, $\hat{\theta}$) is characterized by a loss function $l(\theta,\hat{\theta})$. This function represents our "loss" when we say that $\hat{\theta}$ is the parameter value, when its value is θ. In this situation, we are naturally interested in

finding an estimation procedure which minimizes our loss. We thus seek that procedure $\tilde{\theta}$ which will minimize our overall expected loss, or regret,

$$E_G(\hat{\theta}) = \int_x \int_\theta l(\theta,\hat{\theta})\, h(x,\theta)\, d\theta dx$$

$$= \int_x \int_\theta l(\theta,\hat{\theta})\, f(x|\theta)\, g(\theta)\, d\theta dx,$$

where $g(\theta) = G'(\theta)$ and $h(x,\theta)$ is the bivariate density of x and θ. If $\tilde{\theta}$ represents the "minimizing procedure," i.e.,

$E_G(\tilde{\theta}) = \min_{\hat{\theta}} E_G(\hat{\theta})$, then $\tilde{\theta}$ is called the Bayes estimate of θ and $E_G(\tilde{\theta})$ is the Bayes regret.

The most widely used loss function in estimation is the quadratic loss, $l(\theta,\hat{\theta}) = k(\theta-\hat{\theta})^2$. This yields the very appealing Bayes estimator,

$$\tilde{\theta} = E(\theta|x) \text{ with regret } E[\mathrm{Var}(\theta|x)].$$

Let us consider the special case, when $\theta \sim N(\mu,\sigma^2)$ and $x \sim N(\theta,\sigma^2)$. For the quadratic loss, we obtain,

$$E(\theta|x) = \frac{x\tau^2 + \mu\sigma^2}{\sigma^2 + \tau^2}$$

Notice that $f(x) = \int_\theta f(x|\theta)g(\theta)d\theta$, and

$$E(\theta|x) = \int_\theta \frac{\theta f(x|\theta)g(\theta)d\theta}{f(x)}$$

It is apparent that if the density function $g(\theta)$ is not known, then it is impossible to obtain $E(\theta|x)$. For the special case, $x \sim N(\theta, \sigma^2)$, $E(\theta|x)$ can be rewritten as:

$$E(\theta|x) = x + \sigma^2 \cdot \frac{f'(x)}{f(x)} \quad , \text{where}$$

$$f'(x) = \frac{d\, f(x)}{dx} \, . \qquad \text{The marginal density } f(x) \text{ is}$$

not known exactly, but it can be estimated with several past observations from several different (random) values of θ. In that case we have,

$$\widehat{E(\theta|x)} = x + \sigma^2 \cdot \frac{\widehat{f'(x)}}{f(x)} \qquad \text{which is} \qquad (1)$$

an approximation to the Bayes estimate called the Empirical Bayes estimate. The density ratio in the right hand side of equation (1) can be estimated well by estimating numerator and denominator separately. (See Parzen.) Under certain mild conditions, the risk incurred by using the Empirical Bayes estimator approaches that of the true Bayes estimator as n, the number of experiences increases.

2. <u>The First Order Auto-Regressive Parameter</u>

The simplest auto-regressive model is:

$$x_t = \alpha x_{t-1} + \varepsilon_t, \text{ where the } \varepsilon_t \text{ are} \qquad (2)$$

independent errors. If the ε_t are assumed normal, then the M.L.E. of α is:

$$\hat{\alpha} = \frac{\sum_{t=1}^{n} x_t x_{t-1}}{\sum_{t=1}^{n} x_{t-1}^2} \, , \text{ and} \qquad (3)$$

$$\hat{\sigma}^2 = s^2 = \frac{1}{n} \sum_{t=1}^{n} (x_t - \hat{\alpha} x_{t-1})^2.$$

It has been shown (Mann & Wald, 1943; and Durbin, 1960) that if $|\alpha| < 1$ then $\xi_n = \sqrt{n}(\hat{\alpha} - \alpha)$ is asymptotically $N(0, 1-\alpha^2)$ regardless of the distribution of the ϵ_t. The likelihood function $L = \prod_{i=1}^{n} f(x_i|\alpha,\sigma)$ can be written

$$L = (2\pi\sigma^2)^{-n/2} \left[\exp \frac{-1}{2\sigma^2} \quad (ns^2 + (\hat{\alpha}-\alpha)^2 \ \Sigma x_{t-1}^2) \right],$$

which is of the general form:

$$f(\underset{\sim}{x}|\alpha,\sigma^2) = f(\hat{\sigma}, s^2, u|\alpha,\sigma^2),$$

where $u = \Sigma x_{t-1}^2$ and $\underset{\sim}{x}' = (x_1, x_2, \ldots, x_n)$. It can be shown that

$$E(\Sigma x_{t-1}^2) = E(\Sigma \alpha^{2t} \epsilon_{n-t}) \xrightarrow[n]{} \sigma^2/(1-\alpha^2),$$

and thus

$$L = (2\pi\sigma^2)^{-n/2} \exp \left[\frac{-1}{2} \left(\frac{ns^2}{\sigma^2} + (\hat{\alpha}-\alpha)^2/(1-\alpha^2) \right) \right].$$

so that $\hat{\alpha}$ is "asymptotically" sufficient for α. For small samples we have,

$$f(\underset{\sim}{x}|\alpha,\sigma^2) = f(\hat{\alpha}, s^2, u|\alpha,\sigma^2). \quad \text{But}$$

$$f(\underset{\sim}{x},\alpha,\sigma^2) = f(\underset{\sim}{x}|\alpha,\sigma^2) \cdot g(\alpha,\sigma^2)$$

$= f(\hat{\alpha},s^2,u|\alpha,\sigma^2).g(\alpha,\sigma^2)$. Thus,

$$f(\underset{\sim}{x}) = \int\int_{\sigma^2}\int_{\alpha} f(\hat{\alpha},s^2,u|\alpha,\sigma^2) \cdot g(\alpha,\sigma^2)\, d\alpha d\sigma^2$$

$= f(\hat{\alpha},s^2,u)$. Therefore

$$f(\alpha,\sigma^2|\underset{\sim}{x}) = \frac{f(\alpha,\sigma^2\underset{\sim}{x})}{f(\underset{\sim}{x})}$$

$= f(\alpha,\sigma^2|\hat{\alpha},s^2,u)$, and after integrating w.r.t. σ^2, we obtain,

$f(\alpha|\underset{\sim}{x}) = f(\alpha|\hat{\alpha},s^2,u)$. Note that $E(\alpha|\underset{\sim}{x}) = E(\alpha|\hat{\alpha},s^2,u) \neq E(\alpha|\hat{\alpha})$.

We shall ignore the dependence of Var $(\hat{\alpha})$ on α and propose to use a "marginal" empirical Bayes estimator (Clemmer and Krutchkoff, 1968) for α.

Assume that α is a random variable with unknown density function $g(\alpha)$, and that we have observed $k>1$ time series involving the independently generated parameters $\alpha_1,\alpha_2,\ldots,\alpha_k$. If we have computed the M.L.E.'s $\hat{\alpha}_1, \hat{\alpha}_2,\ldots,\hat{\alpha}_k$ then the marginal empirical Bayes estimator (Clemmer and Krutchkoff, 1968, and Rutherford and Krutchkoff, 1969) which we propose for α_k is:

$$E_k(\alpha_k|\hat{\alpha}_k) = \hat{\alpha}_k + \{(1-\hat{\alpha}_k^2)/n\}\frac{f_k'(\hat{\alpha}_k)}{f_k(\hat{\alpha}_k)}, \text{ where} \qquad (4)$$

$$\frac{f_k'(\hat{\alpha}_k)}{f_k(\hat{\alpha}_k)} = \frac{\sum_{i=1}^{k}\left\{\left(\frac{\sin G_i}{G_i}\right)^2 - \left(\frac{\sin F_i}{F_i}\right)^2\right\}}{h\sum_{i=1}^{k}\left(\frac{\sin F_i}{F_i}\right)^2} ,$$

and where

$$G_i = (\hat{\alpha}_k - \hat{\alpha}_i + h)/2h$$

$$F_i = (\hat{\alpha}_k - \hat{\alpha}_i)/2h ,$$

$$h = k^{-1/5}\,\mathrm{Max}[\,(\tfrac{1}{k}\sum(\hat{\alpha}_i - \bar{\hat{\alpha}})^2)^{1/2},\ \{(1 - \hat{\alpha}_k^2)/n\}^{1/2}],$$

and n = length of present (kth) time series.

3. Prediction of future values of Auto-Regressive Model.

The A.R. model is useful in predicting future observations if a good estimate of α is available since,

$$x_{t+1} = \alpha x_t + \epsilon_t$$

and

$$E(x_{t+1}|x) = \alpha x_t .$$

Assume that x_{t+1} is an unknown parameter, θ, and that we desire to minimize the loss function $L(\theta,\hat{\theta}) = (\theta - \hat{\theta})^2$. Furthermore, assume a multivariate distribution $f(x,\alpha,\theta)$. Let f represent the density function for the associated random variables. Formally, then, the Bayes estimate of θ is:

$$E(\theta|x) = \frac{\int\int_{\alpha\theta}\theta f(x,\alpha,\theta)\,d\theta\,d\alpha}{\int\int_{\alpha\theta} f(x,\alpha,\theta)\,d\theta\,d\alpha}$$

$$= \int\int_{\alpha\theta}\theta f(\theta|x,\alpha)\,\frac{g(x,\alpha)}{g(x)}\,d\theta\,d\alpha$$

$$= \int\int_{\alpha\theta}\theta f(\theta|x,\alpha)g(\alpha|x)\,d\theta\,d\alpha$$

$$= \int_{\alpha} g(\alpha|x)\left\{\int_{\theta} f(\theta|x,\alpha)\,d\theta\right\}d\alpha$$

$$= \int_{\alpha} g(\alpha|x) \cdot E(\theta|x,\alpha)\,d\alpha$$

$$= \int_{\alpha} g(\alpha|x) \cdot \alpha x_t\,d\alpha$$

$$= E(\alpha|x) \cdot x_t$$

Thus, to minimise the squared error of the estimate of x_{t+1}, it is only necessary to use the Bayes estimate of α, $E(\alpha|x)$ in the usual estimator. In other words, to estimate the future value x_{t+1} well, just estimate the parameter α well.

4. The Regression Model.

The A.R. process may appear as the error term in regression models. Suppose that $x_1, x_2, x_3, \ldots$ are fixed and known numbers (investment expenditures for example), and $u_t = \alpha u_{t-1} + \epsilon_t$ is an A.R. process. Then

$y_t = \beta x_t + u_t$ is called a time series regression model with auto-regressive error term. The model can be written (5)

$$y_t - \alpha y_{t-1} = \beta(x_t - \alpha x_{t-1}) + \epsilon_t$$

yielding the M.L.E.'s for α, β:

$$\hat{\alpha} = \frac{\sum_{t=1}^{n} (y_t - \hat{\beta}x_t)(y_{t-1} - \hat{\beta}x_{t-1})}{\sum_{t=1}^{n} (y_{t-1} - \hat{\beta}x_{t-1})^2}, \quad (6)$$

$$\beta = \frac{\sum_{t=1}^{n} (x_t - \hat{\alpha}\, x_{t-1})(y_t - \hat{\alpha}\, y_{t-1})}{\sum_{t=1}^{n} (x_t - \hat{\alpha}\, x_{t-1})^2} \quad (7)$$

while the estimate of σ^2 is:

$$s^2 = \frac{1}{n} \sum_{t=1}^{n} \{y_t - \hat{\alpha}\, y_{t-1} - \hat{\beta}(x_t - \hat{\alpha}\, x_{t-1})\}^2$$

The M.L.E.'s for α and β (6), (7) are non-linear and require an iterative scheme for solution (Cochran & Orcutt, 1949). It may be shown that,

$$E(\sigma^2 | \alpha,\beta) = \sigma^2,$$

$E(\hat{\beta} | \alpha,\beta) = \beta$, and using the result of Mann & Wald,

$E(\hat{\alpha} | \alpha,\beta) = \alpha$ asymptotically.

Furthermore, if $\epsilon_t \sim NID(0,\sigma^2)$ then $\hat{\beta} \sim N\{\beta, \sigma^2/\Sigma(x_t - \alpha x_{t-1})^2\}$.

Let $Q(\alpha,\beta) = \sum_{t=1}^{n} [y_t - \alpha y_{t-1} - \beta(x_t - \alpha x_{t-1})]^2$.

Then $\frac{\partial^2 Q(\alpha,\beta)}{\partial \alpha \partial \beta} = \sum[(x_t - \alpha x_{t-1}) u_{t-1} + \epsilon_t x_t]$.

Now if $E(u_o) = E(\epsilon_o) = 0$, we note that

$$u_n = \sum_{i=0}^{n} \alpha^{n-1} \epsilon_i, \quad \text{so that } E(\partial^2 Q/\partial \alpha \partial \beta) = 0.$$

Thus, we shall assume that $\hat{\alpha}$ and $\hat{\beta}$ are independent.

Note that $\hat{\beta}$ is sufficient for β when α is known but that $\hat{\alpha}$ is not sufficient for α when β is known. When α and β are both unknown, then the likelihood function involves several statistics. Thus if $y = (y_1, y_2, \ldots, y_n)$, the likelihood function will take the form,

$$f(\underline{y} | \alpha,\beta,\sigma^2) = f(\hat{\alpha},\hat{\beta},\underline{u} | \alpha,\beta,\sigma^2).$$

Now suppose that α and β are random variables with joint density $g(\alpha,\beta)$, and that the estimation problem has occurred k-1 times previously.

It may be shown that the posterior densities for α and β may be written:

$$f(\alpha|y) = f(\alpha | \hat{\alpha},\hat{\beta},s^2,u)$$

and $$f(\beta|y) = f(\beta | \hat{\alpha},\hat{\beta},s^2,u) \quad .$$

We therefore propose to use "marginal" estimators for α and β.

Then,

$$E_k(\alpha_k|\hat{\alpha}_k) = \hat{\alpha} + \{(1 - \hat{\alpha}^2)/n\} \cdot \left\{ \frac{f_k'(\hat{\alpha}_k)}{f_k(\hat{\alpha}_k)} \right\},$$

where

$$f_k(\hat{\alpha}_k) = \frac{1}{2\pi k h} \cdot \sum_{i=1}^{k} \left\{ \frac{\sin\left(\frac{\hat{\alpha}_k - \hat{\alpha}_i}{2h}\right)}{(\hat{\alpha}_k - \hat{\alpha}_i)/2h} \right\}^2 ,$$

$$f_k'(\hat{\alpha}_k) = \{f_k(\hat{\alpha}_k + h) - f_k(\hat{\alpha}_k)\}/h ,$$

$$h = k^{-1/5} \text{ Max } [\{\frac{1}{k} \sum(\hat{\alpha}_i - \bar{\hat{\alpha}})^2\}^{1/2}, \{(1 - \hat{\alpha}^2)/n\}^{1/2}],$$

and n = length of kth times series, and

$$E_k(\beta_k|\hat{\beta}_k) = \hat{\beta}_k + \frac{s^2}{\sum_{t=1}^{n} (x_t - \hat{\alpha}_k x_{t-1})^2} \cdot \frac{f_k'(\hat{\beta}_k)}{f_k(\hat{\beta}_k)} ,$$

$$h = k^{-1/5} \text{ Max}\{[\frac{1}{k} \sum (\hat{\beta}_i - \bar{\hat{\beta}})^2]^{1/2}, s\} .$$

5. Prediction with the Regression Model.

As in the case with the A.R. model, we shall assume that the future value, y_{t+1} is the parameter θ, and that there exists a distribution

$f(y,\theta,a,b)$, where $y_t = bx_t + u_t$,

$u_t = au_{t-1} + \varepsilon_t$, and ε_t are independent errors. Then,

$y_{t+1} = ay_t + b(x_{t+1} - ax_t) + \varepsilon_t$ and

$E(y_{t+1}|a,b,y_t) = ay_t + b(x_{t+1} - ax_t)$. Then, formally,

$E(\theta|y_t) = \int \theta f(\theta|y) d\theta$

$= \int \theta f(\theta,y) d\theta / \int f(\theta,y) d\theta$

$$= \frac{\iiint \theta f(\theta,y,a,b)\,da\,db\,d\theta}{\iiint f(\theta,y,a,b)\,da\,db\,d\theta}$$

$$= \iiint \theta\, f(\theta|y,a,b)g(y,a,b)\,d\theta\,da\,db/g(y)$$

$$= \iint E(\theta|y_t,a,b)\frac{g(y,a,b)}{g(y)}\,da\,db$$

$$= \iint E(\theta|y_t,a,b)g(a,b|y)\,da\,db$$

$$= \iint g(a,b|y)[ay_t + b(x_{t+1} - ax_t)]\,da\,db$$

$$= E(a|y_t)y_t + E(b|y_t)[x_{t+1} - E(a|y_t) \cdot x_t]$$

$$= \bar{a}y_t + \bar{b}(x_{t+1} - \bar{a}x_t).$$

The last two lines follow if a and b are independent.

6. General Simulation Procedures and Results.

The Monte-Carlo simulations in this research were conducted on the IBM 360 computer at Virginia Polytechnic Institute and State University, Blacksburg, Virginia. For each of the estimation situations, 50 values of the parameter θ were generated from either a uniform distribution or one of the following Pearson distributions: Bell-shaped, U-shaped, J-left shaped, or J-right shaped. These 50 values of θ were used to simulate up to 50 independent, consecutive experiences. The classical estimate for each value of θ was obtained, and the E.B. estimate was computed for the 2nd, 4th, 6th, 10th, 12th, 25th, and 50th experiences. The entire procedure was repeated 500 times, with 50 different values of θ on each repetition. The M.S.E.'s were estimated for each of the past experience numbers listed above, as follows:

$$\text{MS}(\tilde{\theta}) = \frac{1}{500}\sum_{i=1}^{500} (\tilde{\theta}_i - \theta_i)^2$$

$$\text{MS}(\hat{\theta}) = \frac{1}{500}\sum_{i=1}^{500} (\hat{\theta}_i - \theta_i)^2 .$$

For economy of computer time, the errors were almost exclusively uniform. Most of the estimates involved were "asymptotically" independent of error distribution. A number of simulation runs were duplicated with both normal and uniform errors, but almost no difference was noted in the quality of the empirical Bayes relative to the classical estimator. (See Table II).

Some typical examples of the relative improvement of empirical Bayes procedures over classical are given in the accompaning Tables. As is usual in empirical Bayes, the improvement is given in terms of

$R_k = EMS(\tilde{\theta})/EMS(\hat{\theta})$, for k experiences and for several different z vlaues with

$Z = E\,Var(\hat{\theta}|\theta)/Var(\theta)$ and for each type of prior distribution. It has been found in previous research that for a fixed number of past experiences, R_k invariably decreases with z. All of the simulations were conducted using equi-length series. The empirical Bayes estimator does not require this, but the computer time and storage were reduced considerably in this way. Incidentally, it was noted in the simulations that the R values were not noticably different for series of length 12 and 36 observations.

In the regression model, a change in the Z value of one of the parameters did not change the R value of the estimate of the other parameter. This is attributable to the independence of the M.L.E.'s. Furthermore, the improvement in R of the e.B. estimate of b with respect to experiences, is far greater than that for a up to about 5 or 6 experiences and after that, the situation is reversed. This produces a "crossover" effect in the R-values of a and b.

In all of the situations, for $Z \geq .75$, a tremendous improvement is noted in R values up to about 6 experiences. Thereafter there is a "tapering off" in efficiency noted. The optimum value in most situations seems to be about 6 experiences.

TABLE I

VALUES OF E. B. ESTIMATES OF A. R. PARAMETER

EXPERIENCES	2	5	10	25	50
UNIFORM PRIOR					
Z= .36	-	-	.96	.83	.79
.73	.96	.86	.81	.73	.69
1.1	.87	.74	.69	.63	.58
5.2	.8	.63	.57	.49	.43
BELL SHAPE PRIOR					
Z= .43	.97	.9	.86	.8	.77
.77	.93	.81	.78	.73	.68
1.6	.86	.72	.68	.61	.55
4.6	.83	.65	.59	.49	.41
U-SHAPE PRIOR					
Z= .23	-	-	.94	.89	.86
.42	1.0	.93	.87	.83	.8
.9	.93	.81	.76	.74	.7
2.6	.78	.67	.6	.54	.5
J-SKEWED LEFT PRIOR					
Z= .36	-	-	.94	.84	.78
.73	.97	.88	.81	.73	.7
1.65	.89	.73	.67	.61	.55
4.9	.83	.61	.55	.46	.4

TABLE II

R VALUES OF REGRESSION PARAMETERS FOR N=12

Experiences	Uniform Errors		Normal Errors	
	$\tilde{a}$	$\tilde{b}$	$\tilde{a}$	$\tilde{b}$
2	.957	.890	.954	.895
4	.872	.760	.873	.770
6	.811	.710	.812	.711
10	.790	.653	.786	.665
25	.714	.618	.715	.622
50	.644	.563	.646	.592
Z	2.9	3.8	2.9	3.8

TABLE III

R-VALUES FOR THE REGRESSION MODEL PARAMETER ESTIMATES

PARAMETER a - UNIFORM PRIOR

EXPERIENCES		2	5	10	25	50
Z_a =	.48	-	.94	.92	.85	.8
	.91	.95	.87	.82	.76	.72
	3.74	.85	.73	.65	.57	.5

PARAMETER b - UNIFORM PRIOR

Z_b =	.47	-	.90	.88	.85	.81
	1.05	-	.84	.78	.73	.68
	1.67	-	.82	.71	.62	.57
	5.8	1.0	.67	.56	.51	.45

PARAMETER a - BELL SHAPE PRIOR

Z_a =	1.28	.88	.78	.73	.69	.65
	5.25	.84	.71	.63	.53	.47

PARAMETER b - BELL SHAPE PRIOR

Z_b =	2.39	-	.74	.65	.58	.51
	8.35	.88	.61	.54	.5	.46

PARAMETER a - U-SHAPE PRIOR

Z_a =	.675	.92	.86	.82	.8	.78

PARAMETER b - U-SHAPE PRIOR

Z_b =	1.22	1.0	.81	.73	.65	.6

REFERENCES

Clemmer, B. A. and Krutchkoff, R. G., 1966; The Use of Empirical Bayes Estimates in a Linear Regression Model; Biometrika 55, p. 525.

Cochran, D. and Orcutt, G. H., 1949; Application of Least-Squares Regression to Relationships Containing Auto-Correlated Error Terms; Journ. Amer. Stat. Assn., 44, p. 32.

Durbin, J., 1960; Estimation of Parameters in Time Series Regression Models; JRSS, B, 22, p. 139.

Mann, H. B. and Wald, A., 1943; On the Statistical Treatment of Linear Stochastic Difference Equations; Econometrica, 11, 173.

Rutherford, J. R. and Krutchkoff, R. G., 1964; ε Asymptotic Optimality of Empirical Bayes Estimators; Biometrika, 56, p. 220.

SOME EFFECTS OF AN IMPROPER SCREENING TECHNIQUE ON THE AOQ WHEN USING CSP-1

Fred L. Abraham
U.S. Army Ammunition Procurement and Supply Agency
Joliet, Illinois

1.0 INTRODUCTION

1.1 Continuous sampling procedures are acceptance/rectification inspection plans developed for use where production is continuous and where the formulation of inspection lots for lot-by-lot acceptance may be impractical or undesirable (such as in the inspection of ammunition), or artifical as is often the case with conveyor line production. While there exist many continuous sampling procedures, they all have similar characteristics; namely, inspection is carried out by alternate sequences of consecutive item inspection, called 100% inspection or screening and sequences of production from which only samples are taken. The objective of continuous sampling procedures is to provide assurance that the long run percentage of defective units in accepted product will be held down to some prescribed limiting value which will not be exceeded no matter what quality of product is submitted to inspection. This limiting value is called the Average Outgoing Quality Limit of simply, the AOQL.

1.2 The simplest and perhaps the most widely used continuous sampling procedure is Dodge's CSP-1. It is applicable only to quality characteristics subject to non-destructive inspection and on a go-no-go basis - that is, classification of units is either good or bad. The procedure is defined by two parameters, f, the sampling frequency, and i, the clearance number which is the number of consecutive good units that must be found defect free before sampling is to be resumed. Figure 1 gives the procedural flow.

1.3 For given values of f, i and some fraction defective of incoming product, there will result for product of a statistically controlled process a definite average outgoing fraction defective, called the AOQ. A statistically controlled process is defined here as Bernoullian, that is, each unit has an equal probability of being defective. For these values of f and i, the AOQ will have a maximum, or limit, at some particular fraction defective of incoming product. As noted, earlier, this maximum is the AOQL.

1.4 There are 3 other properties besides the AOQ and AOQL which can be useful in describing a particular sampling plan. Those are:

u, the expected number of units inspected on screening following the finding of a defect.

v, the expected number of units passed on sampling before a defect is found, and

---

This paper was reproduced photographically from the author's manuscript.

F, the average fraction of total product inspected.

Appendix A gives the mathematical formulae for CSP-1.

1.5 Inspection under CSP-1 can be performed either by judging a class of defects or by judging individual defects independently. A class of defects is treated in the inspection operation as if the class of defects were a single entity - that is, the finding of any one of the listed defects of the class causes screening to be initiated for all listed defects of the class. When I refer to a class of defects, I am not making specific reference to the classification of defects according to there seriousness such as critical, major, or minor classes as found in Military Standards and Handbooks on inspection. An example of inspection by a class of defects would be the inspection of a cylinder for three quality characteristics - inside diameter - outside diameter - and length of cylinder. Screening would be initiated for all 3 characteristics if any one of the three were found to be nonconforming or defective. As might be expected, when inspection is by a class of defects, inspection during the screening operation is sometimes incorrectly limited to those characteristics triggering the screening requirement.

1.6 Since the choice of f, i and AOQL to be used to inspect product is often based on whether inspection is to be by a class of defects or on an individual basis, we are concerned with the effects of this improper screening technique on the average outgoing quality. The purpose of this paper is to demonstrate the effects on the AOQ, and therefore the AOQL, when the improper technique is used.

2.0 SUMMARY

2.1 The effects of the improper screening technique on the AOQ will be demonstrated by using the following CSP-1 plans from MIL-STD-1235:

(a)	Code J	i = 200	f = 1/100
		AQL = 1.0%	AOQL = 1.35%
(b)	Code K	i = 65	f = 1/200
		AQL = 4.0%	AOQL = 4.96%
(c)	Code H	i = 15	f = 1/25
		AQL = 10%	AOQL = 10.70%

Using these plans, three sets of conditions are considered:

Case I — All defects in the class are considered to have equal probability of occurrence.

Case II — One specific defect has twice the probability of occurrence of any of the others.

Case III — One specific defect has ten time the probability of occurrence of any of the others.

For demonstration purposes it was arbitrarily decided to consider having five defects in the class.

2.2 AOQ curves for the three plans with the three conditions imposed were constructed to demonstrate visually the differences in AOQ when using the correct screening procedure as opposed to using the improper techniques. These curves are exhibited as Figures 2, 3, and 4.

3.0 DISCUSSION

3.1 For relatively small values of the incoming fraction defective of submitted product, p, within a particular inspection plan, there is little difference in effect upon the AOQ between correct and improper techniques. As p increases, the difference in effect upon the AOQ between the correct and improper technique becomes quite significant. It is at these larger values of p where the most concern is focused.

3.2 Visual inspection of Figure 2, 3, and 4 shows that the improper screening technique allows a much larger percentage of defective material to be accepted than does the proper technique. Consider plan (a), Code J, AOQL = 1.35% (See Figure 2). If the submitted product were 4% defective, the Average Outgoing Quality would only be .2% defective when the correct screening technique is used. For the same percent defective, improper screening technique, Case I, would allow the outgoing product to contain 3.2% defective or 16 times more than would the correct technique; Case II, 15.5 times more; and Case III, 3.7 times more than the correct technique. The same analysis will yield similar results from the two other plans considered.

3.3 For the three plans considered in this paper, Table I gives the values of the AOQL, the percent defective of the submitted product where the AOQL occurs, and compares these to the values of the correct method. It clearly can be seen that each case of improper screening has a much larger AOQL value than the correct technique and that these AOQL values occur at larger values of p. The effect, then, of the improper screening technique is the changing of the specified plan to one with a higher AOQL which allows more defective product to pass inspection.

3.4 Appendices A, B, C and D contain the mathematical derivation of formulae for each of the conditions considered. Definitions of symbols and terms, unless otherwise stated, are those given in MIL-STD-1235.

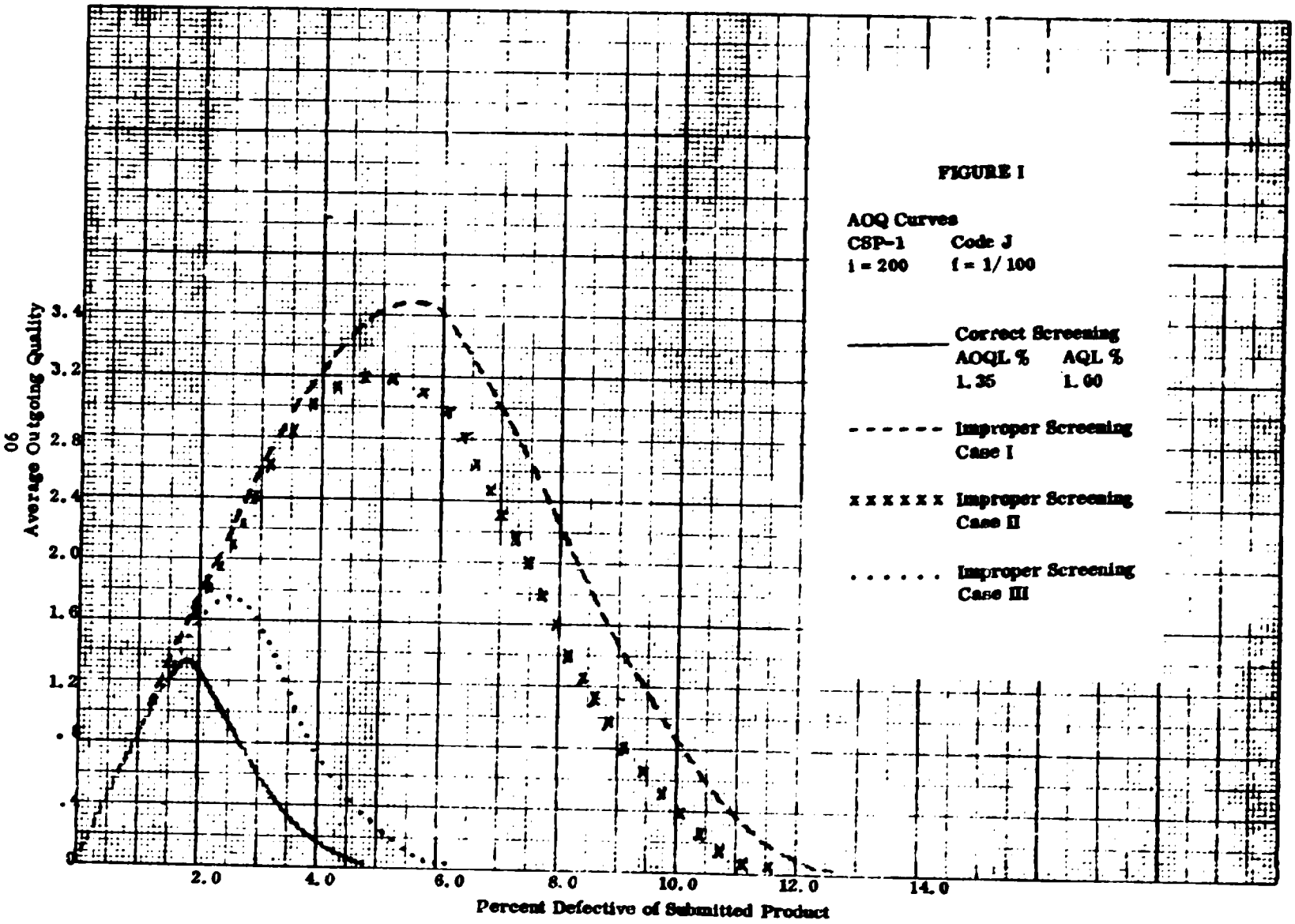


FIGURE I
AOQ Curves
CSP-1 Code J
i = 200 f = 1/100
Correct Screening
AOQL % AQL %
1.35 1.00
Improper Screening
Case I
Improper Screening
Case II
Improper Screening
Case III
Average Outgoing Quality
Percent Defective of Submitted Product
0
.4
1.2
1.6
2.0
2.4
2.8
3.2
3.4
2.0
4.0
6.0
8.0
10.0
12.0
14.0

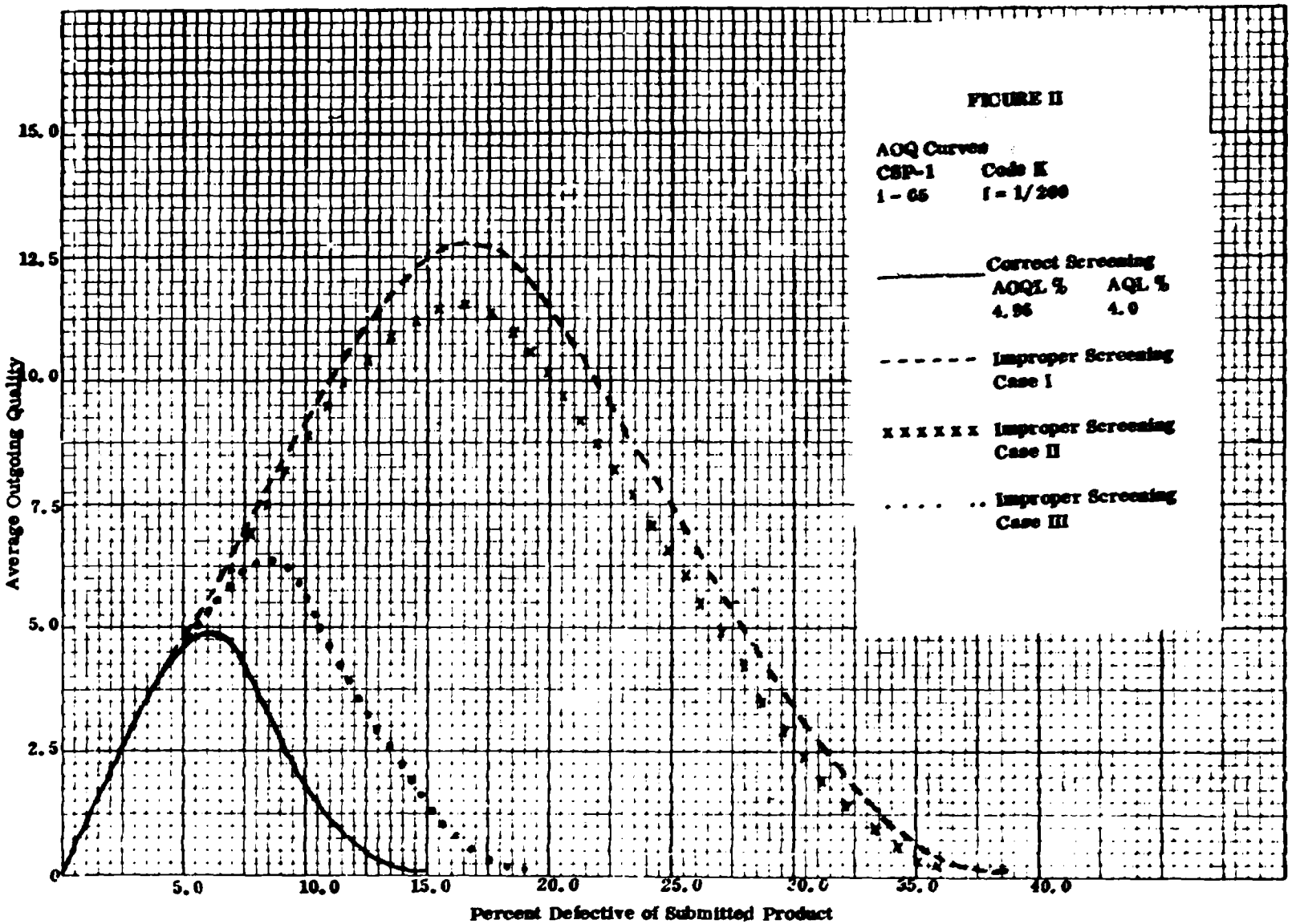


FIGURE II
AOQ Curves
CSP-1 Code K
1 - 65 f = 1/200
Correct Screening
AOQL % AQL %
4.96 4.0
Improper Screening
Case I
Improper Screening
Case II
Improper Screening
Case III
Average Outgoing Quality
15.0
12.5
10.0
7.5
5.0
2.5
0
5.0
10.0
15.0
20.0
25.0
30.0
35.0
40.0
Percent Defective of Submitted Product

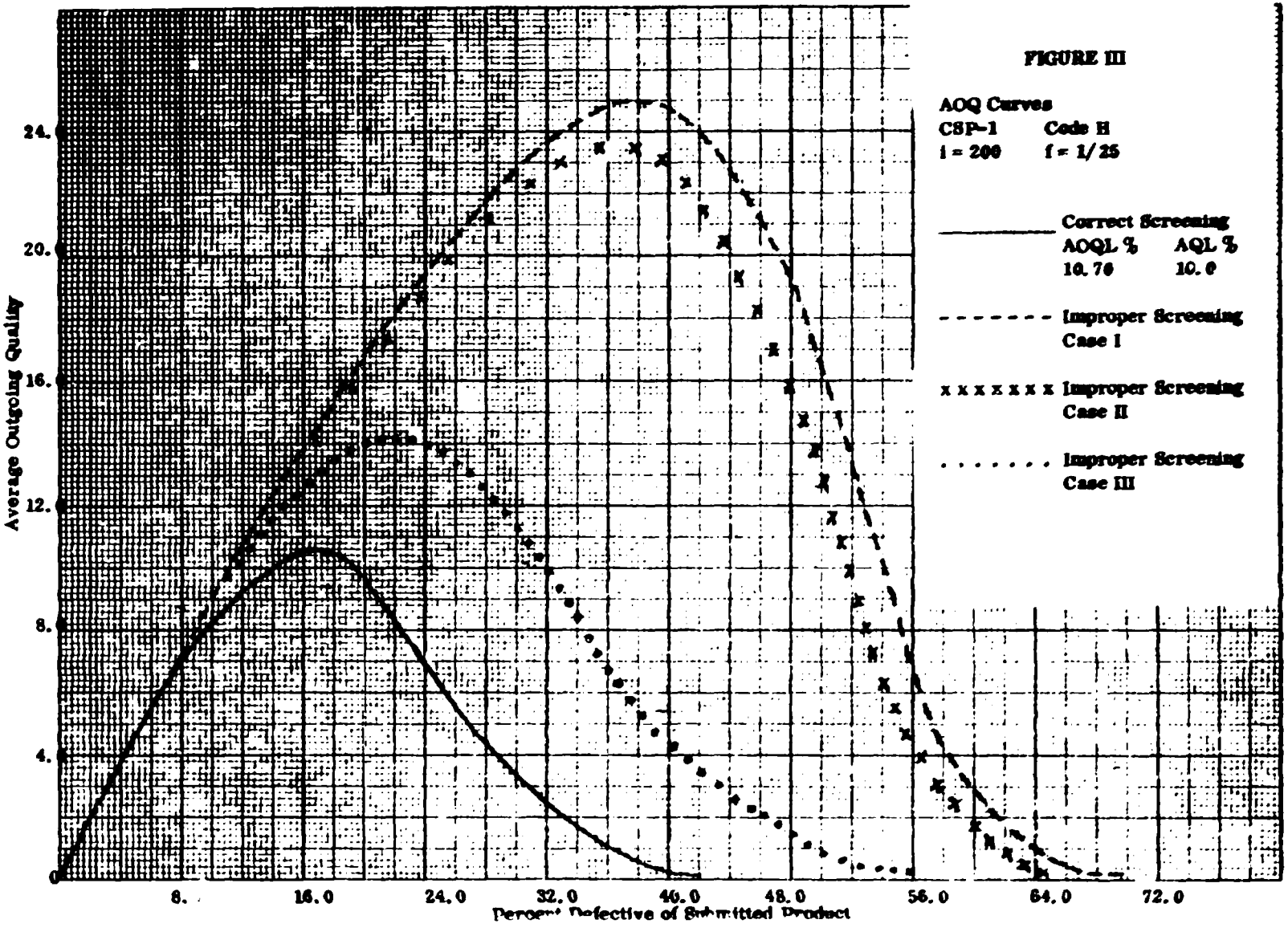


FIGURE III
AOQ Curves
CSP-1 Code H
i = 200 f = 1/25
Correct Screening
AOQL % AQL %
10.76 10.0
Improper Screening
Case I
Improper Screening
Case II
Improper Screening
Case III
Average Outgoing Quality
24.
20.
16.
12.
8.
4.
0
8.
16.0
24.0
32.0
40.0
48.0
56.0
64.0
72.0

TABLE I

CSP-1 PLANS

	i = 200 f = 1/100		i = 66 f = 1/100		i = 15 f = 1/35	
	AOQL	p	AOQL	p	AOQL	p
Case I	3.71	5.45	12.75	17.05	24.95	38.0
Case II	3.25	4.95	11.33	16.75	23.68	36.5
Case III	1.79	2.25	6.30	8.50	14.22	21.9
Correct Method	1.35	1.75	4.96	6.25	10.70	16.5

4.0 BIBLIOGRAPHY

4.1 Dodge, H. F., "A Sampling Inspection Plan for Continuous Production", Annals of Mathematical Statistics, Vol. 14 (1943), pp. 264-279.

4.2 MIL-STD-1235, Single and Multilevel Continuous Sampling Procedures and Tables for Inspection by Attributes, (17 July 1962), Department of Defense.

APPENDIX A

DERIVATION OF FORMULAE FOR THE AOQ

A.1 For given values of i and p, there will be an expected average number of units, u, inspected following the finding of a defect. For given values of f and p, there will be an expected average number of units, v, that will be passed under the sampling procedures before a defect is found.

A.2 The average fraction of total product units inspected in the long run is defined by Dodge as being

$$\text{(A I)} \qquad F = \frac{u + fv}{u + v}$$

where: $$u = \frac{1 - q^i}{p\, q^i} \quad , \quad v = \frac{1}{fp}$$

A.3 It is further assumed that

(a) the inspection operation itself never overlooks a defect

(b) each defective unit found is removed from the operation and is not replaced by a good unit (a departure from Dodge).

The average outgoing quality is related as follows to the incoming quality p:

$$\text{(A II)} \qquad AOQ = p\left(\frac{1 - F}{1 - pF}\right)$$

A.4 Now consider u, the average number of units screened following the finding of a defect.

From Dodge,

$$u = \frac{1 - q^i}{p\, q^i} \quad , \quad q = 1 - p$$

or

$$u = \frac{1 - (1 - p)^i}{p\,(1 - p)^i} \qquad \text{(A III)}$$

However, when the improper screening technique defined is used, Equation (A III) becomes

$$u = \sum_{J_1' = 0}^{J_1} \sum_{J_2' = 0}^{J_2} w_{J_1'\ J_2'}\ u_{J_1'\ J_2'} \qquad \text{(A IV)}$$

where

$$u_{J_1'\ J_2'} = \frac{1 - (1 - A_{J_1'\ J_2'})^i}{A_{J_1'\ J_2'}\,(1 - A_{J_1'\ J_2'})^i}$$

See Appendix B for explanation and derivation of $A_{J_1'\ J_2'}$ and Appendix C for $w_{J_1'\ J_2'}$

A.5 The expected average number of units, v, that will be passed under the sampling procedure before a defect is found, is not affected by improper screening, and remains as $v = 1/fp$.

A.6 Consider Equation (A II),

$$AOQ = p\left(\frac{1 - F}{1 - pF}\right) .$$

Substituting (A I) for F and simplifying

$$\text{(A V)} \qquad AOQ = \frac{p\,(1 - f)}{u\,f\,p\,(1 - p) + 1 - fp}$$

where the value of u in (A V) is obtained by using Equation (A IV). The AOQ under the three conditions of improper screening technique given can now be calculated using Equation (A V) after first solving for u.

A.7 The Average Outgoing Quality Limit, AOQL, being the maximum value of the AOQ that will result for any given values of f and i considering all possible values of p, can be determined by differentiating Equation (A V) with respect to p, equating the derivative to 0, and solving for p.

A.7.1 This was done, but the resulting equation in p was found to be too complex for practical use; therefore, it is not contained in this paper. Instead, Equation (A v) was used to calculate the AOQ for several values of p, and curves were constructed from these points.

APPENDIX B

DERIVATION OF FORMULAE FOR $A_{J_1' \ldots J_k'}$ and $u_{J_1' \ldots J_k'}$

B.1 Let $j = 1, 2, \ldots, J$ be the defects listed in the class, p_{o_j} be the probability that a unit displays the j^{th} defect listed in the class, and E_j the occurrance of a unit displaying the j^{th} defect, then the probability, p, that a unit displays at least one defect in the class is

$$\text{(B I)} \qquad p = \Pr\left\{\bigcup_{j=1}^{J} E_j\right\}$$

The relationship in (B I) can be expressed in another manner if we group those events with equal probabilities of occurrances.

B.2.1 Let $i = 1, 2, \ldots, K$ be the number of groups formed such that there are k groupings with J_i defects in each group, and let E_{j_i} denote the occurrance or a unit displaying the j^{th} member of the i^{th} group, then

$$\text{(B II)} \qquad p = \Pr\left\{\bigcup_{j_1=1}^{J_1} E_{j_1} \cdots \bigcup_{j_k=1}^{J_K} E_{j_K}\right\}$$

where $j_1 + \ldots J_K = J$

B.2.2 When the improper screening method discussed in this paper is used, it becomes apparent that screening may be done for only one defect of the class, for two defects of the class, or for any specific number of defects listed in the class. If screening is done for J' defects of the class, then the probability of finding at least one of the J' defects on a particular unit is

$$\text{(B III)} \qquad p' = A_{J_1' \ldots J_K'} \quad , \quad \text{where}$$

(B IV) $$A_{J_1'} \ldots J_K' = \Pr\left\{\bigcup_{j_1=0}^{J_1'} E_{j_1} \cdots \bigcup_{j_K=0}^{J_K'} E_{j_K}\right\}$$

and $$\sum_{i=1}^{K} J_i' = J'$$

B.3 Case I

B.3.1 The assumption is made that each defect within the class has an equal probability of occurrence, that is

$$p_{o_j} = p_{o_{j+1}} \quad , \quad j = 1, 2, \ldots, J-1$$

B.3.2 For this case only, the general Equation (B II) becomes

(B V) $$p = \Pr\left\{\bigcup_{j_1=1}^{J} E_{j_1}\right\} = \sum_{j=1}^{J} \binom{J}{j_1} (-1)^{j_1-1} p_{o_{j_1}}^{j_1}$$

$$= 1 - q_{o_{j_1}}^{J} \quad , \quad \text{where } q_{o_{j_1}} = 1 - p_{o_{j_1}}$$

and Equation (B IV) becomes

(B VI) $$A_{J'} = \Pr\left\{\bigcup_{j_1=1}^{J_1} E_{j_1}\right\}$$

B.4 Case II and III

B.4.1 Case II assumes that one specific defect has twice the probability of occurrence of the other four listed in the class. This one defect how has a fraction defective $2\,p_o$.

B.4.2 Case III assumes that one specific defect has ten times the probability of occurrence of the other four defects listed in the class. The fraction defective of this one specific defect is now 10 p_o.

B.4.3 Since the specific defect in both cases differ by the constants, 2 and 10, only one set of equations need to be derived from the general equations, (B II) and (B IV):

$$\text{(B VII)} \qquad p = Pr\left\{ \bigcup_{j_1=0}^{J_1} E_{j_1} \bigcup_{j_2=0}^{J_2} E_{j_2} \right\}$$

and

$$\text{(B VIII)} \qquad A_{J_1' \; J_2'} = Pr\left\{ \bigcup_{j_1=0}^{J_1'} E_{j_1} \bigcup_{j_2=0}^{J_2'} E_{j_2} \right\}$$

where $p_{o_{j_2}} = C\, p_{o_{j_1}}$, C = either 2 or 10 and the state $J_1' = 0$, $J_2' = 0$ does not exist, since screening would not be reinitiated unless at least one defect were found on the sampled unit.

B.5 For each state $A_{J_1' \ldots J_K'}$, screened following the finding of a defect.

$$u_{J_1' \ldots J_K'} = \frac{1 - (1 - A_{J_1' \ldots J_K'})^i}{A_{J_1' \ldots J_K'} (1 - A_{J_1' \ldots J_K'})^i}$$

or for the cases presented in this paper

$$u_{J_1' \; J_2'} = \frac{1 - \left(1 - A_{J_1' \; J_2'}\right)^i}{A_{J_1' \; J_2'} \left(1 - A_{J_1' \; J_2'}\right)^i} \, i \quad .$$

APPENDIX C

DERIVATION OF FORMULAE FOR $w_{J_1' \ldots J_k'}$

C.1 It has been shown that the improper screening technique described in this paper allows screening to be done for one defect in the class, for two defects in the class, or for any specific number of defects listed. We must now consider the probability of being on screening at each of these particular states.

C.1.1 Appendix B showed for each of the above states, how the expected average number of units, $u_{J_1' \ldots J_k'}$, screened in one sequence was found. These values of $u_{J_1' \ldots J_k'}$ must be weighted by the probability of that particular state occurring.

C.1.2 One would not re-initiate screening unless a defect in the class were found. Since a sampled unit may contain one, two or all the listed defects, it can be said that the probability of going on screening is the probability of finding at least one of the listed defects.

C.1.3 Now consider the different ways or states to be on screening for the J defects listed in the class. It follows that the probability of being on screening for a particular state, denoted by $w_{J_1' \ldots J_k'}$, given that screening is in effect, is

$$(C\,1)\quad w_{J_i' \ldots J_k'} = \frac{\text{Probability of screening for } J_1' \ldots J_k' \text{ defects}}{\text{Probability of screening for at least one defect in the class}}$$

where the symbols and their definitions are the same as those developed in Appendix B.

C.2 In Case I, only where the listed defects have equal probability of occurrence, the probability that a sample unit contains J_1' defects in the class, resulting in screening for these J_1' defects, is

$$\binom{J}{J_1'} p_o^{J_1'} q_o^{J - J_1'}$$

and the probability of screening for at least one defect in the class is

$$\sum_{J_1' = 1}^{J} \binom{J}{J_1'} p_o^{J_1'} q_o^{J - J_1'}$$

Then (C I) becomes

$$(C\ II) \qquad w_{J_1'} = \frac{\binom{J}{J_1'} p_o^{J_1'} q_o^{J - J_1'}}{\sum_{J_1' = 1}^{J} \binom{J}{J_1'} p_o^{J_1'} q_o^{J - J_1'}} = \frac{\binom{J}{J_1'} p_o^{J_1'} q_o^{J - J_1'}}{1 - q_o^{J}}$$

The five weights for Case I are then

$$w_1 = \frac{5 p_o q_o^4}{1 - q_o^5} \qquad w_4 = \frac{5 p_o^4 q_o}{1 - q_o^5}$$

$$w_2 = \frac{10 p_o^2 q_o^3}{1 - q_o^5} \qquad w_5 = \frac{p_o^5}{1 - q_o^5}$$

$$w_3 = \frac{10 p_o^3 q_o^2}{1 - q_o^5}$$

C.3 Using the constant "C" defined in B.4.3, a general expression for the weights in Cases II and III can be found by using the same method of C.2.

(C I) becomes

$$(C\ III)\qquad w_{J_1'\ J_2'} = \frac{\binom{J_1}{J_1'}\ p_o^{J_1'}\ q_o^{J_1 - J_1'}\ \binom{J_2}{J_2'}\ (C\,p_o)^{J_2'}\ (1 - C\,p_o)^{J_2 - J_2'}}{D}$$

where $D = \sum_{J_1'=0}^{J_1} \sum_{J_2'=0}^{J_2} \binom{J_1}{J_1'} \binom{J_2}{J_2'}\ p_o^{J_1'}\ q_o^{J_1 - J_1'}\ (C\,p_o)^{J_2'}\ (1 - C\,p_o)^{J_2 - J_2'}$

$$- \binom{J_1}{0} \binom{J_2}{0}\ p_o^0\ (C\,p_o)^0\ q_o^{J_1}\ (1 - C\,p_o)^{J_2}$$

and where $J_1 = 4$, $J_2 = 1$, $C = 2$ for Case II, $C = 10$ for Case III,

and the term $\binom{J_1}{0}\binom{J_2}{0}\ p_o^0\ (C\,p_o)^0\ q_o^{J_1}\ (1 - C\,p_o)^{J_2}$ in the denominator represents the state $J_1' = 0$, $J_2' = 0$, which will not exist in screening.

The denominator in (C III) above will simplify to

$$1 - q_o^4\ \ (1 - C\,p_o\ ).$$

Then from (C III),

$$w_{10} = \frac{(1 - C p_o)\, 4 p_o q_o^3}{D} \qquad w_{01} = \frac{C p_o q_o^4}{D}$$

$$w_{20} = \frac{(1 - C p_o)\, 6 p_o^2 q_o^2}{D} \qquad w_{11} = \frac{C p_o\, 4 p_o q_o^3}{D}$$

$$w_{30} = \frac{(1 - C p_o)\, 4 p_o^3 q_o}{D} \qquad w_{21} = \frac{C p_o\, 6 p_o^2 q_o^2}{D}$$

$$w_{40} = \frac{(1 - C p_o)(p_o^4)}{D} \qquad w_{31} = \frac{C p_o\, 4 p_o^3 q_o}{D}$$

$$w_{41} = \frac{C p_o p_o^4}{D}$$

where $D = 1 - q_o^4 (1 - C p_o)$.

C. 4 Note that (C III) reduces to (C II) when $J_2 = 0$. Then (C III) may be considered the general expression for Case I, II and III.

APPENDIX D

SPECIFIC TO GENERAL

D.1 Throughout this memorandum the word "specific" has been used as an adjective to modify a defect whose probability of occurrence is different from the other, equal probability of occurrences. This restriction was made to prevent the mathematical derivation from becoming needlessly lengthy.

D.2 It will be shown here that this restriction is not needed, that the results will be the same if the particular defect having the odd probability of occurrence is not specified.

D.3 Assume we have J defects in the class, where the jth defect has probability p (j) of being the odd one, where the sum p (1) + p (2) + ... + p (j) = 1.

D.4 Any result, R, in the memorandum, would have to be made subject to the condition that the ith defect was the odd one. However, it is immediately apparent that $(R \mid 1) = (R \mid 2) = \ldots = (R \mid j) = R$, since the given condition would be equivalent to specification.

D.5 It can therefore be stated that p (1) $(R \mid 1)$ + p (2) $(R \mid 2)$ + ... + p (j) $(R \mid j)$ = (p (1) + p (2) + ... + p (j)) R = R.

EMPIRICAL BAYES ESTIMATION FOR THE APPLIED STATISTICIAN

Richard Krutchkoff
Virginia Polytechnic Institute and State University

In the past few years my students and I have published several articles (see the references) illustrating the practicality of the Empirical Bayes approach to statistical estimation. Unfortunately, these articles were written in a way which does not easily allow for extension to other applications. I am, therefore, writing this non-technical paper so that Empirical Bayes estimation can be both understood and used without one having to be an expert in the field.

When to Use Empirical Bayes

Let us say that you want to estimate the parameter θ (or a vector of parameters θ) from some specified distribution. You take your sample and obtain the best classical estimate, $\hat{\theta}$, for θ. This is usually sufficient for your purposes. If, however, you feel that this estimate is not good enough, perhaps due to large experimental error or small sample size, then you should consider other alternatives.

Let us say at this point you realize that you have done all this several times before. That is, you have considered similar situations and have estimated the parameter θ from the same distributional form. The first question you should ask yourself is, "are the previous values of the estimated parameter the same as the parameter presently being estimated". If the answer to this is "yes", then all the data should be pooled. This is usually done by weighting the estimates proportional to the number of observations contained therein. If the variation of θ between experiments is less than the variance of your unpooled estimator, then you do improve your estimate in this way. If, on the other hand, the variation of θ between experiments is larger than the variance of your unpooled estimator, then this pooled approach affords an estimator with larger mean squared error than the unpooled estimator. What is needed in this case is a weighted average in which data from distributions with parameters close to the present parameter value receive more weight than data from distributions with parameters not so close. A natural measure of closeness is provided by the probability density or mass function itself. The estimators obtained in this manner are called Empirical Bayes estimators. They have been derived in the literature from another point of view (see the references), to exhibit optimal asymptotic properties and have been shown by extensive stochastic simulation experiments to exhibit excellent small sample properties. In particular these estimators are never worse than the classical estimators and often provide up to a 50 percent decrease in mean squared error.

There is a good deal of evidence which indicates that the Empirical Bayes estimators are robust to most everything. If the previous experiments contain little evidence about this experiment, then the Empirical Bayes estimate differs little from the classical estimate with correspondingly little decrease in mean squared error. On the other hand, the more closely the experiments resemble the present one the greater the decrease in mean squared error.

What Estimator to Use

Although I have, in the literature, proposed different Empirical Bayes estimators for different situations, I will here propose only one easily used procedure for any parametric situation. Let us say that $f(\underset{\sim}{x}|\theta)$ represents the probability density function (or the probability mass function) of the data for a given value of the parameter. If there is a sufficient statistic for θ, then let $\underset{\sim}{x}$ represent the sufficient statistic itself rather than the vector of observations; and let $f(\underset{\sim}{x}|\theta)$ represent the probability density of that sufficient statistic. Let $\hat{\theta}$ be the classical estimator for the present value of θ. Now consider the last several past experiences (several being anywhere from five to fifteen; fifteen when available). Use $\hat{\theta}_i$, $i = 1, 2, \ldots, 16$, to represent the classical estimates in the past fifteen (or as few as five) experiences plus the present one ($i = 16$). The Empirical Bayes estimate for the present value of θ is given by

$$\tilde{\theta} = \frac{\sum_{i=1}^{16} \hat{\theta}_i f(\underset{\sim}{x}|\hat{\theta}_i)}{\sum_{i=1}^{16} f(\underset{\sim}{x}|\hat{\theta}_i)} .$$

This is a weighted average of the estimates of the last sixteen θ values. The weights, however, are not functions of how well $\hat{\theta}$ is estimating θ_i, i.e., a function of the number of observations in the ith experiment, but rather a function of how likely the present set of data, $\underset{\sim}{x}$, would be if θ_i were the true value of the parameter. It should be noted that the present unpooled estimate $\hat{\theta}_{16}$ recieves the greatest weight. This is more easily seen when $\underset{\sim}{x} = \hat{\theta}_{16}$ and the weight is proportional to $f(\hat{\theta}_{16}|\hat{\theta}_{16})$.

The MSA of the Empirical Bayes Estimator

The decrease in mean squared error using this Empirical Bayes procedure depends on the ratio of the variance of the unpooled estimator to the variation of the parameter between experiments. Unlike ordinary pooling, however, the Empirical Bayes estimators never have a larger mean squared error than the unpooled estimator.

Since

$$\text{Var } \hat{\theta} = E \text{ Var}(\hat{\theta}|\theta) + \text{Var } E(\hat{\theta}|\theta)$$

we have

$$\text{Var } \theta = \text{Var } \hat{\theta} - E \text{ Var}(\hat{\theta}|\theta)$$

when $\hat{\theta}$ is unbiased for θ. We can, therefore, often present

$$Z = \frac{\text{Var}(\hat{\theta}|\theta)}{\text{Var } \hat{\theta} - E \text{ Var}(\hat{\theta}|\theta)}$$

as the ratio required for use with Table 1 in obtaining the approximate mean squared error for your Empirical Bayes estimator. Here $\text{Var}(\hat{\theta}|\theta)$ is estimated as the variance of the unpooled estimator for your present experiment. $E \text{ Var}(\hat{\theta}|\theta)$ is estimated by the average of the variances of the unpooled estimators for the present and fifteen past experiments, while Var θ is estimated by

$$\frac{1}{16} \sum_{i=1}^{16} (\hat{\theta}_{16} - \frac{1}{16} \sum_{j=1}^{16} \hat{\theta}_j)^2 .$$

Table I is an approximate table which has been compiled by using extensive stochastic simulations. One finds the Z value and obtains the proper multiplier, interpolating when necessary. The variance of the unpooled estimator is then multiplied by this multiplier to obtain the mean squared error for the Empirical Bayes estimator. If Z is very small, under 0.1, then there is little advantage in obtaining the Empirical Bayes estimator. If, on the other hand, Z is very large, over 4.0, then one should seriously consider pooling in the usual manner. For $0.1 \leq Z \leq 4.0$ one should use the estimator given here.

When Z is based on only five past experiences, the multipliers are closer to unity. A reasonable estimate would be to split the difference between them and unity, e.g., a multiplier of 0.8 would become 0.9 when only 5 past experiences are available.

Table I: Variance Multipliers

Z	Multiplier
0.0	1.0
0.2	0.9
0.5	0.8
1.0	0.7
2.5	0.6
5.0	0.5

A Nonparametric Estimator

Sometimes we feel that our data is almost, but not quite normal or not quite poisson or not quite gamma, etc. In spite of this, we often make the assumption that our data fits close enough in order to avoid having to use a nonparametric procedure, which is itself very powerful. The nonparametric Empirical Bayes procedure which I am presenting here is, however, remarkably powerful and competes well with its parametric counterparts.

Let $\underset{\sim}{x}_i$, $i = 1, 2, \ldots, 16$, represent the last sixteen vector observations, including the present one with $i = 16$. That is, $\underset{\sim}{x}_i$, $i = 1, 2, \ldots, 16$, represents the sets of random samples obtained in the last sixteen experimental situations with $\underset{\sim}{x}_{16}$ the present set of random samples. Let x_{ij}, $j = 1, 2, \ldots, m_i$, represent the components of the ith vector. Note that the vector lengths need not be the same but m_i should be at least two.

Let us now say that we have a classical estimator for whatever it is we want to estimate. In the nonparametric situation this would be something like the mean, the variance, the standard deviation, or the percentiles.

Call the estimates in the last sixteen situations $\hat{\theta}_i$, $i = 1, 2, \ldots, 16$, where $\hat{\theta}_{16}$ is the present classical estimate of what we are interested in estimating. Then the Empirical Bayes estimator is given by

$$\tilde{\theta} = \frac{\sum_{i=1}^{16} \hat{\theta}_i f(\underset{\sim}{x}_{16}; \underset{\sim}{x}_i)}{\sum_{i=1}^{16} f(\underset{\sim}{x}_{16}; \underset{\sim}{x}_i)}$$

where

$$f(\underset{\sim}{x}_{16}; \underset{\sim}{x}_i) = \frac{1}{(2\pi m_i h_i)^{m_{16}}} \prod_{j=1}^{m_{16}} \sum_{k=1}^{m_i} \left[\frac{\sin\left(\frac{x_{16j} - x_{ik}}{2h_i}\right)}{\frac{x_{16j} - x_{ik}}{2h_i}} \right]^2$$

with

$$h_i = (1/m_i)^{7/10} \sqrt{\sum_{k=1}^{m_i} (x_{ik} - \bar{x}_i)^2}$$

and $\bar{x}_i = (1/m_i) \sum_{k=1}^{m_i} x_{ik}$ and where $\frac{\sin \theta}{\theta} = 1$. The $f(\underset{\sim}{x}_{16}; x_i)$

given here is simply a nonparametric estimate of the appropriate likelihood function. Preliminary results indicate that this procedure is also remarkably robust. In particular, the lengths of the vectors from experiment to experiment need not be the same.

REFERENCES

1. Bennett, G. K., and Martz, Jr., H. F. (1969). A Continuously Smooth Empirical Bayes Estimator. Proceedings of the Texas Tech Symposium on Statistics, August 1969, pp. 93-109.

2. Clemmer, B. A. and Krutchkoff, R. G. (1968). Use of Empirical Bayes Estimators in a Linear Regression Model, Biometrika, Vol. 55, pp. 515-524.

3. Gabbert, J. T. and Krutchkoff, R. G. (1970). Supplementary Sample Non-Parametric Empirical Bayes in a Quality Control Situation, Biometrika, Vol. 57, pp. 67-69.

4. Griffin, B. S. and Krutchkoff, R. G. (1971). Optimal Linear Estimators: An Empirical Bayes Version with Application to the Binomial Distribution, Biometrika in press.

5. Griffin, B. S. and Krutchkoff, R. G. (1971). An Empirical Bayes Estimator for P (success) in the Binomial Distribution, Sankya in press.

6. Krutchkoff, R. G. (1967). Supplementary Sample Nonparametric Empirical Bayes Biometrika, Vol. 54, pp. 451-458.

7. Krutchkoff, R. G. (1969). An Introduction to Empirical Bayes, published in "Empirical Bayes Estimation and Computing in Statistics." Proceedings of the Texas Tech Symposium on Statistics, August 1969, pp. 1-15.

8. Krutchkoff, R. G. (1969). The Future of Empirical Bayes, published in "Empirical Bayes Estimation and Computing in Statistics." Proceedings of the Texas Tech Symposium on Statistics, August 1969, pp. 150-

9. Lemon, G. H. and Krutchkoff, R. G. (1969). An Empirical Bayes Smoothing Technique, Biometrika, Vol. 56, pp. 361-365.

10. Lemon, G. H. (1969). Smooth Empirical Bayes Estimators: With Results for the Binomial and Normal Situations, Proceedings of the Texas Tech Symposium on Statistics, August 1969, pp. 110-140.

11. Lemon, G. H. (1969). A Note on Some Empirical Bayes Techniques in Point Estimation, Proceedings of the Texas Tech Symposium on Statistics, August 1969, pp. 41-48.

12. Martz, H. F. and Krutchkoff, R. G. (1969). Empirical Bayes Estimators in a Multiple Linear Regression Model, Biometrika, Vol. 56, pp. 367-374.

13. Martz, Jr., H. F. (1969). Multivariate Extended Empirical Bayes Estimation, Proceedings of the Texas Tech Symposium on Statistics, August 1969, pp. 49-60.

14. Martz, Jr., H. F. (1969). A Note on Estimating the Derivative of a Density Function with Empirical Bayes Applications, Proceedings of the Texas Tech Symposium on Statistics, August 1969, pp. 141-149.

15. Miller, D. M. (1969). Empirical Bayes Estimates of Queuing Parameters, Proceedings of the Texas Tech Symposium on Statistics, August 1969, pp. 71-92.

16. Rencher, A. C. (1969). Some Empirical Bayes Estimators in Regression and Analysis of Variance, Proceedings of the Texas Tech Symposium on Statistics, August 1969, pp. 16-40.

17. Rutherford, J. R. (1969). An Empirical Bayes Approach to the Non-Central T Distribution, Proceedings of the Texas Tech Symposium on Statistics, August 1969, pp. 61-70.

18. Rutherford, J. R. and Krutchkoff, R. G. (1967). The Empirical Bayes Approach: Estimating the Prior Distribution. Biometrika, Vol. 54, pp. 326-328.

19. Rutherford, J. R. and Krutchkoff, R. G. (1969a). ε Asymptotic Optimality of Empirical Bayes Estimators, Biometrika, Vol. 56, pp. 220-223.

20. Rutherford, J. R. and Krutchkoff, R. G. (1969b). Some Empirical Bayes Techniques in Point Estimation. Biometrika, Vol. 56, pp. 133-137.

A STATISTICAL APPROACH TO OPTIMIZING THE MECHANICAL BEHAVIOR OF COMPOSITE MATERIALS

Donald L. Martin, Jr.*

ABSTRACT

The filler particle size distribution is shown to have a significant effect on the mechanical properties of CTPB composite materials. The experimental compositions are selected in accordance with a simplex-centroid lattice design, and the regression equation for this statistical model is shown to adequately fit the experimental data. The various mechanical properties estimated from the regression equation for a fixed filler content are represented by contour plots on triangular coordinate graphs. These plots predict the various possible mixtures that will result in the same material behavior, and should prove valuable in optimizing mechanical properties of composite materials for specific applications.

INTRODUCTION

Composite materials are currently being used in many applications where the strength-to-weight ratio becomes a significant factor. The extreme loading conditions imposed on these materials has prompted many investigations of methods for optimizing the mechanical properties for specific designs.

The work described here concerned a carboxyl-terminated polybutadiene (CTPB) composite material that consists of a rubbery matrix highly filled with rigid filler particles. The size distribution of these filler particles and the total filler content have a pronounced effect on the rheological characteristics of an uncured slurry [1-4],** and may drastically alter the mechanical response of the cured composite material. Such characteristics as the equilibrium modulus, stress-strain response, and ultimate tensile properties [5-8] of a cured composite material may be significantly changed by altering the particle size distribution for a fixed filler content.

---

*Aerospace Engineer, U. S. Army Missile Command, Redstone Arsenal, Alabama.

**Numbers in brackets pertain to references listed at the end of this paper.

Since high strength composites usually require a high solids content, the solids must be efficiently packed to obtain optimum properties. The packing fraction as defined here is the volume fraction of solids in the minimum sedimentation volume, which is thus independent of the actual total filler content in the formulation. When the particle shape, average size, or size distribution are changed such that the packing fraction is increased, less binder is required to fill the voids between the filler particles. Thus there is a direct relationship between the packing fraction of the mixture of filler particles and the potential solids-loading capability of the resulting composite. For a fixed filler content, the largest packing fraction usually results in the minimum mix viscosity, the lowest initial modulus, and the greatest strain capability.

Optimizing a mixture of filler particle sizes from the standpoint of rheological, physical, and thermal properties requires knowledge of the mechanical response of all possible mixtures that might be considered. Since the experimental determination of all the various responses is prohibitively expensive and time-consuming, a method for accurate prediction is highly desirable.

BACKGROUND AND THEORY

The concept of simplex lattice designs introduced by Scheffé [9] for experimental situations involving multicomponent systems is valid when the experimental response depends only on the proportion of components in the system. To illustrate, let q be the number of components considered and X_i the proportion of the i^{th} component that is subject to the constraints

$$X_i \geq 0 \quad (i = 1, 2, \ldots q) \tag{1}$$

and

$$X_1 + X_2 + X_3 + \ldots + X_q = 1 \, . \tag{2}$$

The factor space is a (q-1)-dimensional simplex satisfying these constraints (triangle for q = 3, tetrahedron for q = 4). Scheffé uses the nomenclature {q, m} to describe the designs, where m is the degree of the polynomial regression function in X_i used to describe components. The {q, m} simplex lattice (m = 1, 2, 3, ...) consists of the $\left(\frac{m+q-1}{m}\right)$ experimental points in the simplex representing all possible mixtures in which the proportion of each component has the m+1 equally spaced values:

$$X_i = 0, \frac{1}{m}, \frac{2}{m}, \ldots 1 \, . \tag{3}$$

The simplex lattice method has two key features: (1) Properties or responses are measured at lattice composition points and (2) polynomial equations having a special correspondence to the lattice points are then used to represent the measured responses at these points.

Simplex-Centroid Design

The {3,2} simplex-centroid lattice design proposed by Scheffé [10] was satisfactorily applied by Alley and Dykes [11] to estimate the packing fractions of multicomponent mixtures of ammonium perchlorate and aluminum particles.

In the simplex-centroid statistical model, (2^q-1) observations are taken, one at each of the following: the q pure components, the $(q/2)$ binary mixtures with equal proportions, the $(q/3)$ ternary mixtures with equal proportions, ..., and the q-nary mixture with equal proportions. These observations correspond to points $X_1, X_2, \ldots, X_q$ of the simplex, obtained by making q permutations of $(1,0,0,\ldots,0)$, $(q/2)$ permutations of $(1/2, 1/2, 0, \ldots, 0)$, and $(q/3)$ permutations of $(1/3, 1/3, 1/3, 0, \ldots, 0)$, and the point $(1/q, 1/q, \ldots, 1/q)$. Thus the design consists of the centroid of the {q,2} simplex and the centroids of all the lower dimensional simplexes it contains.

The polynomial regression equation has as many coefficients as there are points in the design, and is given by

$$\eta = \sum_{i=1}^{q} \beta_i X_i + \sum_{i=1}^{q-1} \sum_{j=i+1}^{q} \beta_{ij} X_i X_j + \sum_{i=1}^{q-2} \sum_{j=i+1}^{q-1} \sum_{k=j+1}^{q} \beta_{ijk} X_i X_j X_k + \cdots\cdots + \beta_{12\ldots q} X_1 X_2 \cdots X_q \,. \tag{4}$$

When the response satisfies (4), the (2^q-1) coefficients β are uniquely determined by the values of the response at the (2^q-1) points of the simplex-centroid design [10]: The coefficient β_i in (4) is the response of the pure component i, called the linear blending value; β_{ij} is the coefficient of binary synergism for components i and j; β_{ijk} is the coefficient of ternary synergism for components i, j, and k; etc.

Reference Mixtures

In many experiments dealing with mixtures, a physical, economic, or chemical limitation may restrict the concentration of one or more components of a mixture. For these cases Scheffé suggested the use of pseudocomponents, or what were later termed "reference mixtures," which are mixtures of the pure components and can be substituted for the pure components in the simplex lattice designs.

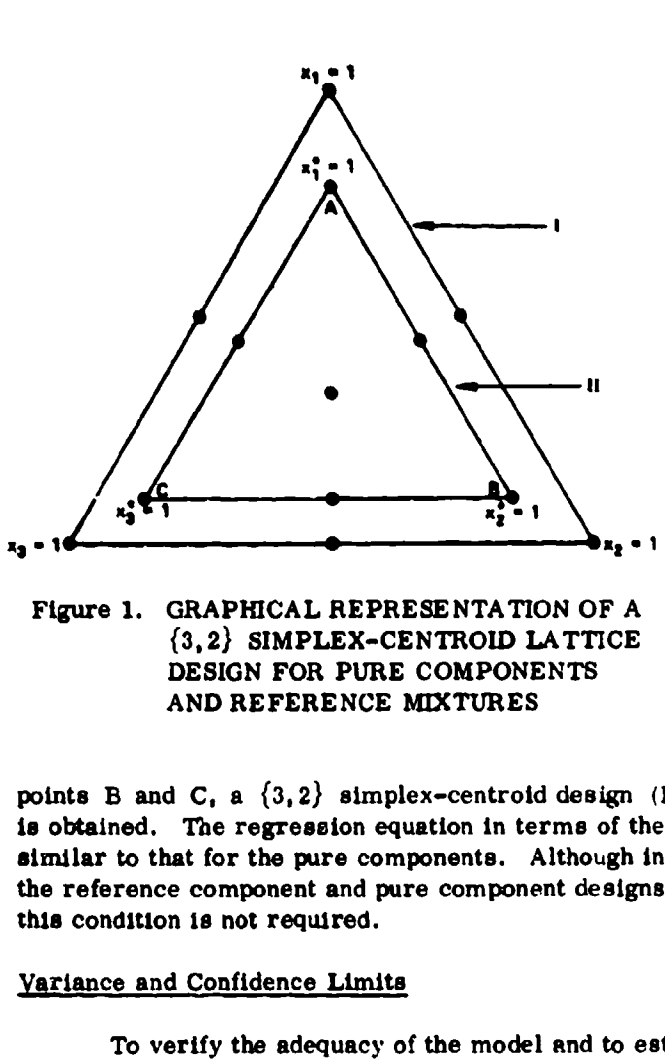


Figure 1. GRAPHICAL REPRESENTATION OF A {3,2} SIMPLEX-CENTROID LATTICE DESIGN FOR PURE COMPONENTS AND REFERENCE MIXTURES

Womeldorph [12] and Alley and Meyers [13] incorporated reference mixtures into the simplex-centroid design for multicomponent x-ray analysis. Figure 1 illustrates two {3,2} simplex-centroid designs, one for mixtures of pure components (designated by I) and the other for reference mixtures (II). Point A is a mixture of pure components, but this mixture is also 100 percent reference mixture 1. By a similar definition for points B and C, a {3,2} simplex-centroid design (II) using reference mixtures is obtained. The regression equation in terms of the reference components is similar to that for the pure components. Although in this particular illustration the reference component and pure component designs have the same midpoint, this condition is not required.

Variance and Confidence Limits

To verify the adequacy of the model and to establish confidence intervals, the variance of the values predicted by the model must be estimated. Variances dealing with any number of components are discussed in detail by Scheffé [9] and Womeldorph [12] and by Gorhman and Hinman [14], who extended the simplex lattice design and applied it to properties of gasolines and other multicomponent systems. The discussion here will concern the expression for the variance of values predicted from the {3,2} simplex-centroid model.

The coefficients β_i, β_{ij}, and β_{123} in the $\{3,2\}$ simplex-centroid model,

$$\eta = \sum_{i=1}^{3} \beta_i X_i + \sum_{i=1}^{2} \sum_{j=i}^{3} \beta_{ij} X_i X_j + \beta_{123} X_1 X_2 X_3 \ , \tag{5}$$

are estimated from linear combinations of the measured response at the lattice points [9]. Replacing the β's by their estimates in terms of the measured responses $\hat{\eta}$ yields

$$\bar{\eta} = \sum_{i=1}^{3} b_i \hat{\eta}_i + \sum_{i=1}^{2} \sum_{j=i}^{3} b_{ij} \hat{\eta}_{ij} + b_{123} \hat{\eta}_{123} \ . \tag{6}$$

If an independent estimate of the error variance, σ^2, is not available, the design must be replicated to obtain an estimate of σ^2:

$$\hat{\sigma}^2 = \frac{\sum_{i=1}^{p} \sigma_i^2}{p} \ , \tag{7}$$

where δ^2 is the average variance for the model, σ_i^2 is the variance of the i^{th} mixture, and p is the number of distinct design points in the model considered.

The X_i's are assumed to be known without error and the $\hat{\eta}$'s result from replicated observations at each lattice point. The expected variance in the predicted response is then given by

$$\mathrm{Var}(\bar{\eta}) = \sigma^2 \left(\sum_{i=1}^{3} b_i^2 / r_i + \sum_{i=1}^{2} \sum_{j=i+1}^{3} b_{ij}^2 / r_{ij} + b_{123}^2 / r_{123} \right) \ . \tag{8}$$

When there are an equal number of observations, r, at each lattice point composition,

$$r = r_i = r_{ij} = r_{123}$$

and

$$\mathrm{Var}(\bar{\eta}) = \frac{\sigma^2}{r} \zeta \ , \tag{9}$$

where

$$\zeta = \sum_{i=1}^{3} b_i^2 + \sum_{i=1}^{2} \sum_{j=i+1}^{3} b_{ij}^2 + b_{123}^2 \tag{10}$$

$$b_i = \left(X_i/2\right)\left(6X_i^2 - 2X_i + 1\right) - 3 \sum_{i=1}^{3} X_i^2 \tag{11}$$

$$b_{ij} = 4X_iX_j \left(3X_i + 3X_j - 2\right) \tag{12}$$

$$b_{123} = 27\, X_1 X_2 X_3 \, . \tag{13}$$

The quantity σ^2/r is dependent on the precision of the experimental observations and ζ is dependent only on the composition being considered. Thus, if an estimate is given of σ^2 and the number of replications r for the simplex-centroid system, the estimated variance for any composition in the factor space can be obtained by using (9) through (13). For 95 percent confidence intervals [11]:

$$Pr\left(\overline{\eta} - \Delta < \eta < \overline{\eta} + \Delta\right) = 0.95 \, , \tag{14}$$

where

$$\Delta = t_{\alpha/(2k),\, f} \sqrt{\mathrm{Var}\left(\overline{\eta}\right)}$$

$\alpha = 0.05$

k = number of constants in model

f = number of degrees of freedom.

When desirable, a contour plot of confidence intervals for the predicted values may be prepared by using (14).

EXPERIMENTAL PROCEDURE

Material

The composite system selected for this investigation consisted of a CTPB binder filled with rigid filler particles. The filler material was obtained from Rohm & Haas Company, Redstone Arsenal, Alabama. Table I presents the nominal size, weight-mean diameter, and specific surface area for the filler particles. In the ensuing discussion, the nominal size will be used to identify the filler component. The particle size distribution of the 200 μ particles was determined by Ro-tap screen analysis; the particle size distributions of the 40 μ and 10 μ particles were determined by a liquid sedimentation technique using a Mine Safety Appliances analyzer.

Table I. PARTICLE SIZE DATA FOR FILLER PARTICLES

MATERIAL NO.	NOMINAL SIZE (μ)	WEIGHT-MEAN DIAMETER (μ)	SPECIFIC SURFACE AREA (CM^2/G)
1	40	44.5	1449
2	200	191.7	190
3	10	16.6	1989

The cumulative particle size distributions are plotted in Figure 2. The 200 μ particle sizes ranged from 74 μ to 380 μ and had a weight-mean diameter of 191.7 μ. The 40 μ particle sizes ranged from 3 μ to 80 μ, with a weight-mean diameter of 44.5 μ. The 10 μ particle sizes ranged from 3 μ to 50 μ with a weight-mean diameter of 16.6 μ. Calculation of the specific surface areas and the weight-mean diameters was based on particle size distribution data and assumed spherical particles.

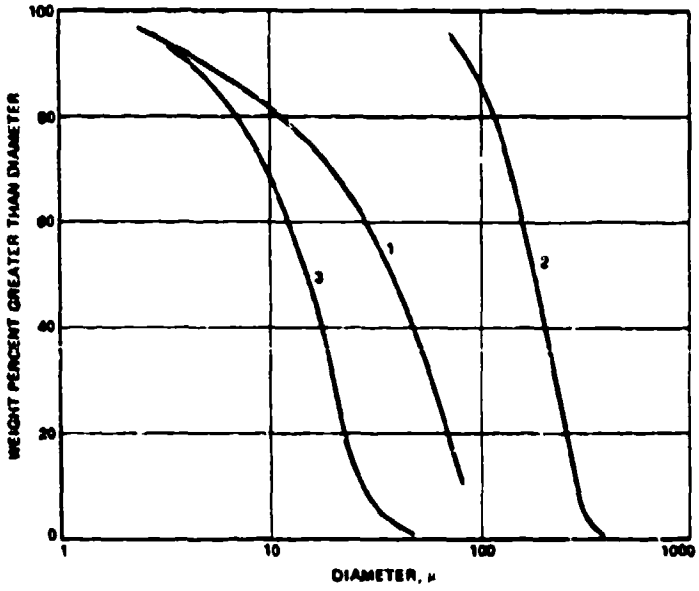


Figure 2. PARTICLE SIZE DISTRIBUTION CURVES FOR FILLER PARTICLES

The raw materials used in the binder formulation of all compositions investigated were obtained from Thiokol Chemical Corporation, Huntsville, Alabama. The chemical analyses of the CTPB prepolymer and curing agents are listed in Table II. The CTPB system uses the carboxyl groups at the terminal positions and the butadiene polymer as the backbone. The ERLA is a trifunctional epoxide compound and the MAPO is a trifunctional imine compound.

Table II. RAW MATERIALS

POLYMER	LOT NO.	EQUIVALENT WEIGHT
CTPB-ZL434	3511	2080.0
ERLA-0510	2327	98.85
MAPO	2565	73.53

The equivalence ratio of the binder component in all compositions investigated was maintained constant at 1.5 equivalents of curing agents to one equivalent of CTPB prepolymer. The ratio of curing agents was maintained constant at 3.8 equivalents of MAPO per one equivalent of ERLA. The composite formulations investigated are summarized in Table III.

Table III. COMPOSITE FORMULATIONS

MIX NO.	VOLUME PERCENTAGE			
	BINDER	40μ	200μ	10μ
314	0.40	0.60	0.0	0.0
311	0.40	0.0	0.60	0.0
317	0.40	0.0	0.0	0.60
318	0.40	0.30	0.30	0.0
319	0.40	0.35	0.0	0.25
320	0.40	0.0	0.35	0.25
321	0.40	0.22	0.22	0.16

Mechanical Response

Uniaxial stress-strain characteristics were determined utilizing an Instron tensile machine at a constant crosshead displacement rate of 1.0 inch/minute and 75°F. Bonded tab-end specimens approximately 3/8 inch wide, 1/4 inch thick, and 2.8 inches long were used to obtain the mechanical response data. Three specimens from each composition were tested and the average value for each property was used in the subsequent correlation.

Crosslink Density

The mechanical properties of polymeric binders may change with the addition of rigid filler particles due to the reinforcing effect of filler particles, changes produced in the crosslink density of the binder, or a combination of the two. For some binder systems the addition of particular filler materials results in an increase in the apparent crosslink density of the binder, while other filler materials hinder the cure reaction and cause a decrease in the apparent crosslink density. Therefore, to adequately account for influence of filler materials on a polymeric binder, it is necessary to determine the relative effects of changes in the filler particles and crosslink density.

The crosslink density of each lattice point composition was determined from compression and tensile properties at equilibrium swelling conditions in benzene. The number of moles of effective crosslinks per unit volume, ν_e, for elastomeric based materials was then calculated from the theory of rubber elasticity which predicts [15]

$$\frac{F}{A_i} = \nu_e RT \left(\lambda - \lambda^{-2}\right) ,$$

where

F = force (dynes)

A_i = the initial cross-sectional area of the unstrained, unswollen specimen (cm^2)

R = universal gas constant, 8.314×10 dynes/mole (°K)

T = test temperature (°K)

λ = extension ratio.

Packing Fraction

The packing fraction or minimum sedimentation volume of filler particles was determined according to the following procedure: Samples of the filler material of each of the lattice point compositions were placed in 15-milliliter centrifuge tubes and the weight of each sample was accurately determined. Each mixture sample was blended with a Pica blender and enough Twitchell base n-heptane was added to completely cover the material. After the slurry was stirred with a microspatula to completely wet the particles and dispel air bubbles, it was centrifuged at 2900 rpm and the volume of the solids measured at 20-minute intervals until a constant volume was obtained. In most cases this determination required 60 minutes or less, with the samples containing the

largest proportion of smaller particle sizes requiring the longest time. The packing fraction, ϕ_m, of the sample was then determined by the relationship

$$\phi_m = \frac{\beta}{\rho},$$

where β is the bulk density of the sample at the minimum sedimentation volume and ρ is the theoretical density of filler particle mixture. Thus the packing fraction is the fraction of the minimum sedimentation volume that is actually occupied by the filler particles, and $1 - \phi_m$ would represent the minimum amount of binder that would theoretically be required to completely fill the voids between the filler particles.

RESULTS

The results of the investigation indicated that the mechanical properties of the composites considered are strongly influenced by the particle size distribution of the filler material. Listed in Table IV are the experimentally determined packing fraction, ϕ_m; maximum corrected stress, σ_m*; strain at maximum corrected stress, ϵ_m; maximum initial tangent modulus, E; crosslink density, ν_e; and normalized tangent modulus, E/ν_e.

Table IV. MEASURED RESPONSES OF LATTICE POINT COMPOSITIONS

MIX NO.	ϕ_m	σ_m (PSI)	$\epsilon_m \times 10^2$ (IN/IN)	E (PSI)	$\nu_e \times 10^5$ (MOLES/CM^3)	$E/\nu_e \times 10^{-6}$ (PSI/MOLE-CM^{-3})
314	0.730	147.4	8.86	1780	19.9	8.95
311	0.680	100.7	8.86	1510	11.2	13.50
317	0.735	260.5	15.1	1990	44.6	4.46
318	0.778	114.7	8.0	2080	15.8	13.20
319	0.754	174.3	12.0	1580	25.0	6.30
320	0.834	118.2	11.7	1545	13.9	11.10
321	0.804	132.7	11.7	1410	7.2	7.22

* The maximum corrected stress σ_m as used in this investigation is defined as the nominal stress multiplied by the extension ratio.

From these values the coefficients $(\beta_i, \beta_{ij}, \beta_{ijk})$ were determined and are presented in Table V.

Table V. COEFFICIENTS OF REGRESSION POLYNOMIAL

COEFFICIENT	ϕ_m	σ_m	ϵ_m	E	ν_e	E/ν_e
β_1	0.730	147.4	8.86	1780	19.9	8.95
β_2	0.680	100.7	8.86	1510	11.2	13.50
β_3	0.735	260.5	15.1	1090	44.6	4.46
β_{12}	0.292	-37.4	-3.44	1740	1.0	7.90
β_{13}	0.089	-87.8	1.97	-1184	-21.94	-2.98
β_{23}	0.533	-207.9	0.691	-696	-47.36	6.00
β_{123}	0.253	174.8	38.76	-10539	-263.1	-99.50

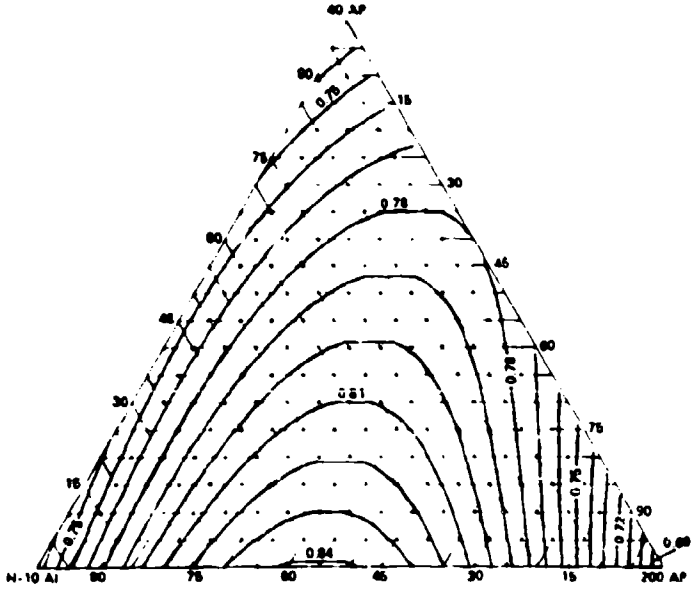


Figure 3. PACKING FRACTION VERSUS FILLER PARTICLE SIZE DISTRIBUTION

The numerical coefficients were substituted into the regression equation (4) to predict the various compositions that will exhibit the same mechanical behavior characteristics. This information is displayed on triangular coordinate graphs as contour plots of constant response with respect to the individual properties investigated.

The first of these plots (Figure 3) indicates the contours of constant packing fraction, which is dependent only on the proportion of the various particle sizes.

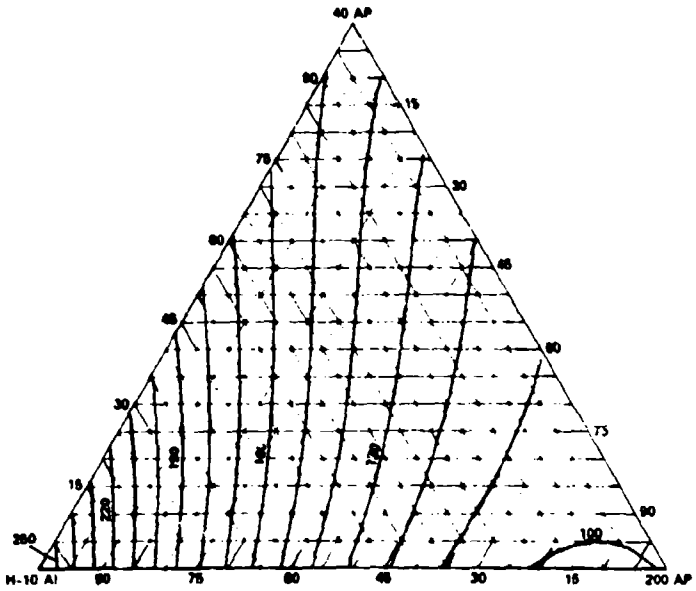


Figure 4. MAXIMUM CORRECTED STRESS AS INFLUENCED BY FILLER PARTICLE SIZE DISTRIBUTION (σ_m, psi)

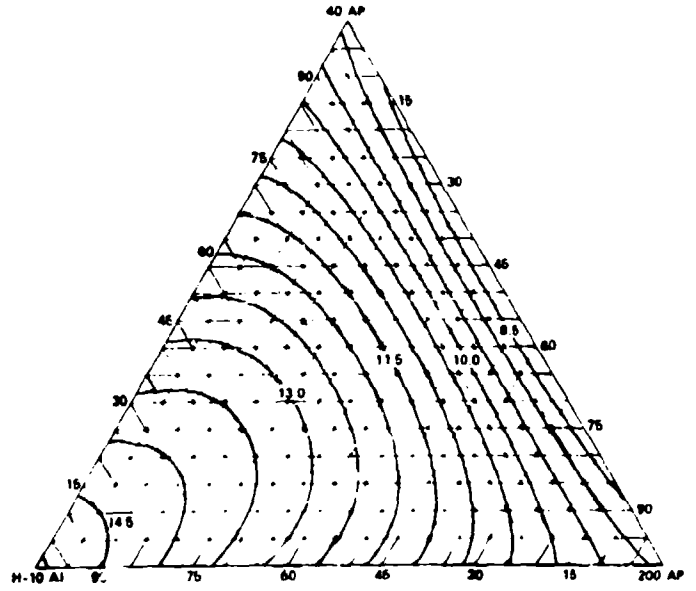


Figure 5. STRAIN CAPABILITY OF CTPB COMPOSITE MATERIALS VERSUS FILLER PARTICLE SIZE DISTRIBUTION (ϵ_m, in./in. × 10^2)

The next three graphs (Figures 4, 5, and 6) show the effect of filler particle size distribution on the maximum corrected stress, strain at maximum corrected stress, and maximum initial tangent modulus, respectively. In all compositions the filler content was maintained constant at 60 percent by volume. The coordinates of the graphs represent the proportions of 40 μ, 200 μ, and 10 μ particle sizes in a given mixture; the vertices of the graphs represent the composite formulations in which the total filler content consisted of 100 percent of the nominal particle size designated.

The maximum corrected stress (Figure 4) varies from 100 to 260 psi, depending on the filler particle size distribution, and in general increases with increasing content of 40 μ and 10 μ filler particles, with the largest stress value occurring near the 100 percent 10 μ point.

The strain at maximum corrected stress (Figure 5) varies from 0.08 to 0.151 inch per inch depending on the filler particle distribution, and increases with the increase in 10 μ content.

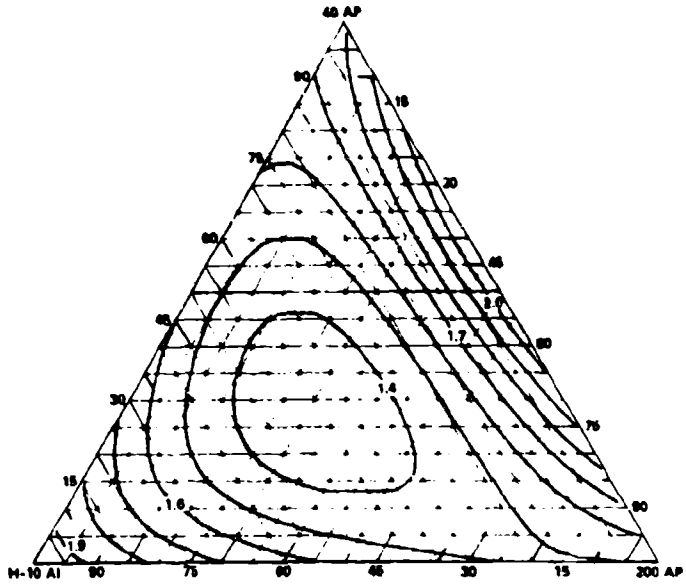


Figure 6. MAXIMUM INITIAL TANGENT MODULUS VERSUS FILLER PARTICLE SIZE DISTRIBUTION ($E \times 10^{-3}$, psi)

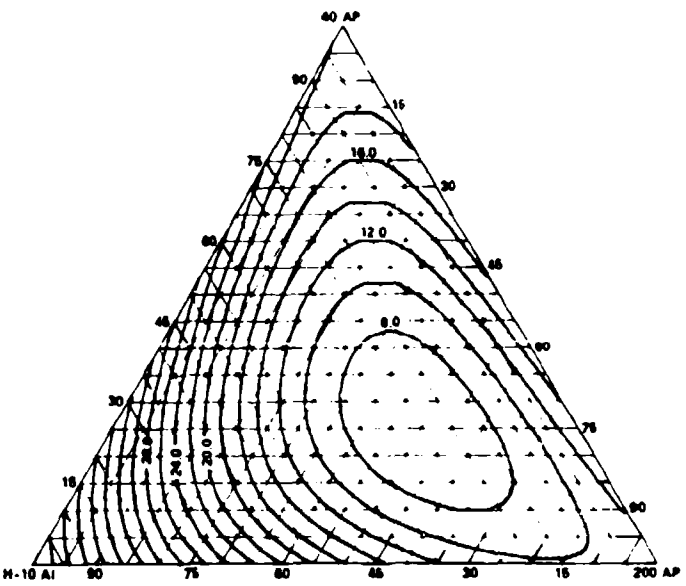


Figure 7. APPARENT CROSSLINK DENSITY VERSUS FILLER PARTICLE SIZE DISTRIBUTION ($\nu_e \times 10^5$, moles/cm^3)

The maximum initial tangent modulus (Figure 6) varies from 1400 to 2080 psi, depending on particle size distribution, and reaches a minimum value when the mixture of filler particles consists of 32 percent 200 μ, 28 percent 40 μ, and 40 percent 10 μ particles.

The stress, strain, and modulus contours bear no resemblance to the constant packing fraction contours, indicating that the filler material has an effect on the binder cure reaction. To determine the extent of this effect, the apparent crosslink density was determined for each lattice point composition, using the procedure described previously. The resulting values appear in the sixth column of Table IV. The corresponding column of Table V gives the coefficients determined for the regression equation. Plotted in Figure 7 are the estimated contours of constant crosslink density values within the factor space. These contours are very similar to those of the initial tangent modulus, with the exception that the point where the minimum crosslink value occurs indicates a greater proportion of 200 μ particles.

Figure 7 provides further evidence that the filler material alters the cure of the binder: the crosslink density varies from 7.2×10^{-5} to 44.6×10^{-5} moles/cm^3, depending on the particle size distribution used.

As was indicated earlier, the maximum initial tangent modulus for crosslinked polymeric binders varies in direct proportion to the crosslink density [15], so that the additional step of normalizing the modulus by the crosslink density was necessary to clearly reflect the reinforcing effect of particle size distribution. The normalized values of E/ν_e for each lattice point composition are included in Table IV (last column), and the corresponding column of Table V gives the coefficients, β, determined for the regression polynomial. The graphic plot of the information (Figure 8) represents contours of constant values of initial tangent modulus, normalized to unit crosslink density, as a function of the particle size distribution.

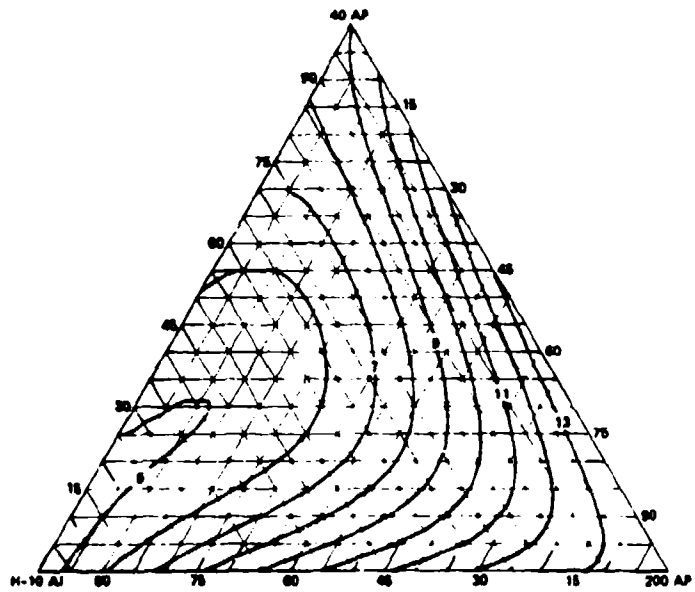


Figure 8. TANGENT MODULUS NORMALIZED UNIT CROSSLINK DENSITY VERSUS FILLER PARTICLE SIZE DISTRIBUTION $\left[\left(E/\nu_e\right) \times 10^{-6}, \text{ psi/mole-cm}^{-3}\right]$

The normalized initial tangent modulus was found to vary from 4.46×10^6 to 13.5×10^6 psi/mole-cm^{-3}, depending on the particle size distribution used. The value was lowest at the 100 percent 10 μ size and increased with increasing content of 40 μ and 200 μ sizes.

CONCLUSIONS

With only a small amount of experimental data, the effect of filler particle distribution on the mechanical properties of CTPB composites can be rapidly and simply approximated by simplex-centroid experimental designs. The most significant feature of this lattice design is the ease with which the regression equation can be derived. The values of the coefficients are uniquely determined and are equivalent

to the estimates that would be obtained by a standard least squares analysis. It should be noted that the estimated coefficient values are not changed by the addition or deletion of components in a design.

The technique is subject to some risks since it is used to estimate wide variations in unknown compositions with a limited data base; however, the results obtained in this investigation indicated that the mechanical property-composition contours are well-behaved and can be adequately represented by the simplex-centroid model.

Although the subject system contained a fixed volume fraction of binder, the method can theoretically be extended to estimate the response for variations in total filler content by using the binder as a fourth component. In addition, a similar design can be devised to incorporate other component sizes of filler particles by merely adding the appropriate coefficients to the regression equation.

The extension of this method to more than three components, however, presented certain difficulties. The comments and suggestions of the clinical session panelists are particularly solicited with respect to the following difficulties encountered:

(1) When investigating composite materials of the type used in this study, we are restricted to the use of reference components. However, the formulator of such composite materials requires the information in terms of the pure components. Therefore, for optimum application of this technique, it becomes necessary to transform the response function in terms of reference mixtures to a response function with pure component concentrations as the so-called independent variables. Womeldorph [12] presented the resulting transformations for a three-component system. The same reasoning used by Womeldorph can theoretically be applied to any number of reference mixture components but the mathematical complexity of the transformation increases by orders of magnitude with the addition of extra components to the model. The simplex-centroid design appears less attractive in these cases and it is doubtful that the designs would be widely used for more than three components. To overcome this difficulty it is planned to use equation (4) and express the composite formulation in terms of the pure components (X) instead of reference mixture components (X*). The $\{q, 2\}$ simplex-centroid concept would thus result in a set of q simultaneous equations which could easily be solved for the unknown coefficients using a high speed computer. The system of experimental points designated in Figure 1 was utilized as a check on this technique. Identical plots were obtained for systems I and II. The actual composition of the points on II in terms of the pure components X_i were used instead of X_i^*. Will this give any confidence that the same results will be obtained for more than three components? What limitations may be anticipated when using this technique with more than three components?

(2) Provided the actual response and composite formulation are used in equation (4) even though it may be slightly different from the lattice point composition, what would be the anticipated effect on the coefficients β? What limitations should be placed on this difference to insure that the response equation would still represent the true behavior of the material? What effect, if any, would this have on the expected confidence limits of the response predicted by the regression equation?

(3) Particular formulations of composite materials are difficult and sometimes almost impossible to obtain. For example, when a small amount of 200 μ filler particles is used in a composite formulation containing no 40 μ or 10 μ sizes, the 200 μ particles will tend to settle to the bottom of the mold before the material is completely cured, resulting in a non-uniform material. Provided a given lattice point composition is unobtainable, would it be more desirable to use an estimated response for the lattice point composition or to neglect this point entirely in the regression equation?

(4) Often it becomes desirable to obtain an estimate for the response of composite formulations that are outside the factor space represented by the reference mixtures tested. How may the confidence limits be estimated for those composition points that are not contained within the lattice of experimental points? Will the procedure discussed on variances and confidence limits with equations (5)-(13) still apply in this case; if not, how would one estimate the confidence limits for these situations?

The range of information available from the contour plots, which in this study indicated the various possible mixtures of filler particle sizes that will result in the same material behavior characteristics, offers the potential of broad application in Army research programs, provided the difficulties encountered can be resolved. Simultaneous consideration of a similar series of plots during the development of any composite material should significantly improve the chances of optimizing the mechanical properties for a specific design and should greatly reduce the time and effort required.

REFERENCES

1. LANDEL, R. F., MOSER, B. G., and BAUMAN, A. J., "Rheology of Concentrated Suspension: Effect of a Surfactant," Proceedings of the 4th International Congress on Rheology, Brown University, Providence, Rhode Island, August 1964, Interscience Publishers, New York, 1965.

2. [redacted]

3. LEEMING, H., and GILLIS, T. W., "The Mechanics of Highly-Filled Propellants," Bulletin of the 4th Meeting of the ICRPG Working Group on Mechanical Behavior, U.S. Naval Training Center, San Diego, California, CPIA No. 944, 1965, p. 1.

4. [redacted]

5. LANDEL, R. F., and SMITH, T. L., "Viscoelastic Properties of Rubber-like Composite Propellants and Filled Elastomers," American Royal Society Journal, v. 31, 1961, p. 599.

6. BREE, H. W., SCHWARZL, F. R., and STRUIK, C. E., "The Mechanical Behavior of Rubbers Containing Fillers of Bimodal Size Distribution," Bulletin of the 5th Meeting of the ICRPG Working Group on Mechanical Behavior, Applied Physics Laboratory, The Johns Hopkins University, Silver Spring, Maryland, CPIA No. 119, v. 1, 1966.

7. MARTIN, D. L., Jr., "The Influence of Microstructural Characteristics on the Maximum Strain Energy Density of a Composite Propellant," ICRPG Mechanical Behavior Working Group, v. 1, 1966, p. 277.

8. MARTIN, D. L., Jr., "Microstructural Characteristics and Failure Mechanism for a Carboxyl Terminated Polybutadiene Composite System," Report No. RK-TR-68-16, U. S. Army Missile Command, Redstone Arsenal, Alabama, September 1968.

9. SCHEFFÉ, H., "Experiments with Mixtures," Royal Statistical Society Journal, Series B, v. 20, 1958, p. 344.

10. SCHEFFÉ, H., "The Simplex-Centroid Design for Experiments with Mixtures," Royal Statistical Society Journal, Series B, v. 25, 1963, p. 235.

11. ALLEY, B. J., and DYKES, H. W. H., "Packing of Multicomponent Mixtures of Ammonium Perchlorate and Aluminum Powders," Report No. RK-TR-66-6, U. S. Army Missile Command, Redstone Arsenal, Alabama, March 1966.

12. WOMELDORPH, D. E., Jr., "Estimation in the Use of the X-Ray Fluorescence Method and Use of Reference Components in Mixture Experimental Designs," Doctoral Dissertation, Virginia Polytechnic Institute, Blacksburg, Virginia, May 1966 (R. H. Myers, Professor).

13. ALLEY, B. J., and MYERS, R. H., "Corrections for Matrix Effects in X-Ray Fluorescence Analysis Using Multiple Regression Methods," Analytical Chemistry, v. 37, No. 1685, 1965.

14. GORMAN, J. W., and HINMAN, J. E., "Simplex Lattice Designs for Multicomponent Systems," Technometrics, v. 4, No. 4, 1962, p. 463.

15. FLORY, P. J., and REHNER, J., Jr., "Statistical Mechanics of Crosslinked Polymer Networks," Journal of Chemical Physics, v. 11, 1943, p. 512.

SYSTEMS VULNERABILITY DUE TO MULTIPLE COMPONENT DRIFT AND COMPONENT FAILURE

W. W. Happ
U. S. Army Construction Engineering Research Laboratory
Champaign, Illinois

ABSTRACT. Two distinct aspects of vulnerability of a lumped system are examined: (i) a small performance drift per unit component drift, commonly referred to as elasticity or sensitivity, and (ii) catastrophic failure of a component, commonly referred to as performance failure or error. Sensitivity analysis and analysis of failure are increasingly employed in computer-oriented analysis, optimization and design of systems and circuits. A systematic derivation of second-order effects is presented, as the current literature is restricted to first-order analysis. Sensitivity of sensitivity, also referred to as second-order sensitivity, and error of error, or second-order error, are evaluated as are sensitivity of error and error of sensitivity. Results are presented in tabular form for quick reference. The implementation of these concepts by data manipulation techniques such as binary base tagging is illustrated. A tutorial introduction to vulnerability concepts is provided to make this survey self-contained.

INTRODUCTION. This investigation was performed under the auspices of the Coordenação dos Programas de Pós-Graduação de Engenharia (COPPE) da Universidade Federal do Rio de Janeiro. The majority of the material is presented here for the first time although some of the material was published as informal course notes for a graduate seminar on "Modelling of System" at the University of Wisconsin in 1966 and for a similar course at the Universidad de Chile, 1969.

The challenge to apply systems concepts effectively to local problems originated at COPPE and provided the impetus for the development of unified, formal exposition of sensitivity and failure analysis as a single topic: vulnerability.

The systems analyst faces radically different kinds of problems in countries with a long tradition of successful technological development - such as Britain or Germany - compared to problems faced by countries with accelerated technological development - such as Puerto Rico or Venezuela - because the latter expose themselves to increased technological vulnera-

Acknowledgement: Part of this work was performed under sponsorship of USAID, United States Agency for International Development, as part of a ten-week teaching and research assignment in Brazil at COPPE, Coordenacao dos Programs de Pos-Graduacao de Engenheria, Universidade Federal do Rio de Janeiro. Part of this material was presented in Publicacao No. 10.70 May 1970 of COPPE and was issued as lecture notes in a graduate course Sistemas multiterminais. This paper has also been submitted for publication to the Journal of System Science, England.

bility. Increased vulnerability results when a relatively small failure in one of the minor components of the system has a major effect on overall system performance.

Vulnerability to component failure or component drift is usually adequately, and nearly always quickly, recognised and compensated for if an appropriate feed back component is available to perform this function. Nations with a long technological tradition have developed this feedback mechanism. Countries without this tradition have not developed it, and are therefore vulnerable, since the pace of technological advance is artificially accelerated. Often this speed-up exceeds by orders of magnitude the "natural mode" of technological development. It is then that the vulnerability to technological failure or dislocation increases rapidly.

The resultant cause-effect relationship can be formulated concisely in the language of systems analysts: the system can respond only in its natural mode, the delay for remedial action is too long, and the forced accelerated pace of technological development is at variance with the feedback mechanism.

Because of the universality of the laws governing system behavior, the systems analyst is destined to make a significant contribution in resolving this problem. The methodology of systems science requires only minor adaptations to switch from the complex, large systems in developed countries to entirely different systems as they exist in developing countries. Beyond the intrinsic academic incentives of analyzing new systems, the systems analyst will tend to study systems of immediate practical value. This type of analysis will thus be of interest to industry and to governmental planning agencies, provided existing methodology of systems science can be effectively extended. This is exactly the aim of examining system vulnerability.

Systems science applies analytical techniques to physical, economic, ecological and many other types of systems, such as communications systems or project planning. The systems approach usually involves three phases:

(i) Modelling: a description of the system in a form suitable for analysis

(ii) Evaluation: includes analysis, optimization, vulnerability and information display

(iii) Strategy: search for implementation, definition of alternatives, decision risks, validation of models and verification of process improvements.

The effective utilization of systems science by developing countries, hand-in-hand with their acceptance of computer technology, or more specifically the effective utilization of even small computers requires increasing familiarity with systems science and thereby provides a challenge for countries with accelerated technological development to adopt systems science to their own needs. Clearly, if computer technology is capable of increasing the pace of remedial action, the aim must be to develop a strategy to identify and to off-set technological vulnerability.

This investigation of vulnerability has the primary aim of making a small and modest step in this direction: to extend systems concepts to unexplored areas of vulnerability, a topic of secondary interest to research agencies in highly developed countries with practically invulnerable technology.

In summary, computer-oriented algorithms for highly vulnerable systems are needed for a simple reason: rapidly developing economies are vulnerable systems, and no scholarly studies on this topic are available.

This report is also intended to serve as a guideline for a graduate-level seminar which aims to apply vulnerability analysis to electrical and industrial engineering. Thus a secondary aim of this investigation is to stimulate seminar projects in which problems for graduate theses in systems science and research projects of practical interest are formulated. Vulnerability investigations were selected because they provide a useful link between (i) the urgent need by local engineers for systems-oriented investigations, and (ii) the capability of the academic community to provide computer-oriented analytical techniques to meet these needs.

The reader who is not primarily interested in mathematical techniques may proceed directly to the last section of the report. The conclusions derived from this investigation are stated specifically for management review, requiring no specialized technical background. The utilization of the results by economists, scientists, engineers, and others, who are primarily interested in possible applications rather than mathematical techniques, will thereby be facilitated.

The remainder of this article was reproduced photographically from the author's manuscript.

1. Basic Concepts of Vulnerability

1.1 Review of Current Literature

During the past five years, techniques for sensitivity and failure analysis were developed for computer-oriented analysis of lumped parameter networks. These techniques can be applied to optimization in systems analysis, automatic control, and stochastic networks. It is therefore desirable to summarize existing results and to extend them in order to formulate a systematic approach. The first part of this investigation consists of a survey of existing results based on several recent investigations:

(a) Development of basic concepts, analytical approaches with emphasis on scope and limitation of procedures may be grouped according to major techniques:

- binary base tagging for lumped systems (1,2,3)
- sensitivity analysis - first-order and higher order formulae (4,5,6)
- failure analysis - error and preassigned accuracy techniques (7,8,9)

(b) Applications of tagging procedures to lumped parameter systems:

- electrical circuits with active elements (10,11, 12)
- stochastic networks applied to models of proces-

sea (13,14,15)

- tolerance analysis and Monte Carlo techniques (16,17,18)

(c) Surveys and summaries primarily with tutorial objectives:

- textbooks and reference books (19,20,21)
- topical reviews and expositions (22-29)
- tutorial surveys (30-35)

1.2 Terminology for Lumped Systems

Consider the circuit in Fig. 1, in which Z_1, Z_2 and Z_3 are known impedance, from which the unknown impedance

$$-Z_4 = \frac{Z_1Z_3 + Z_2Z_3}{Z_1 + Z_2 + Z_3}$$

is to be found. The arbitrary minus sign for the unknown will permit us to write the result as $Z = 0$ where

$$Z = Z_1Z_3 + Z_1Z_4 + Z_2Z_3 + Z_2Z_4 \text{ or } Z = \Sigma Z_1^{d(1)}Z_2^{d(2)}Z_3^{(3)}Z_4^{(4)}$$

where d(i) is zero or one and obeys the code pattern $Y = \overline{Z}$ or $\overline{Y} = Z$. Symbolically,

$$Z^* = \begin{vmatrix} 1 & 0 & 1 & 0 \\ 1 & 0 & 0 & 1 \\ 0 & 1 & 1 & 0 \\ 0 & 1 & 0 & 1 \\ 0 & 0 & 1 & 1 \end{vmatrix} \qquad \text{and} \qquad Y^* = \begin{vmatrix} 0 & 1 & 0 & 1 \\ 0 & 1 & 1 & 0 \\ 1 & 0 & 0 & 1 \\ 1 & 0 & 1 & 0 \\ 1 & 1 & 0 & 0 \end{vmatrix}$$

Note that:

(a) Inspection of Fig. 1 shows that the code Z can be obtained directly as the cut set from the geometry. Similarly code Y above represents the tie set.

(b) The unknown Z_4 plays the same role as the known quantities. Unknowns are isolated by expansion of $Z = 0$; namely,

$$Z = Z(Z_4 = 0) + Z_4 Z(Z_4 = 1)$$

where $Z(Z_4 = 0) = Z_1 Z_3 + Z_2 Z_3$ by scanning Z^* for Z_4

and $Z(Z_4 = 1) = Z_1 + Z_2 + Z_3$ also scanning Z^* for Z_4

hence $Z_4 = -Z(Z_4 = 0)/Z(Z_4 = 1)$

(c) Similarly $Y = 0$ yields $Y_4 = -Y(Y_4 = 0)/Y(Y_4 = 1)$

This is conveniently written as $Y_4 = -Y(\overline{4})/Y(4')$

1.3 Illustrative Example of Vulnerability

To find Z_4 when $Z_2 = 0$ obtain from $Z = 0$ and Z^*

$$Z_4(Z_2 = 0) = - \frac{Z(Z_4 = 0,\ Z_2 = 0)}{Z(Z_4 = 1,\ Z_2 = 0)} = \frac{Z_1 Z_3}{Z_1 + Z_3}$$

The system performance when component 2 fails is:

$$Z(4, \overline{2}) = \frac{Z(\overline{4}, \overline{2})}{Z(4', \overline{2})} \qquad \text{where} \quad Z(\overline{4}, 2) = Z_1 Z_3$$

$$\text{and} \quad Z(4', \overline{2}) = Z_1 + Z_3$$

Generally if P and Q belong to a system H = 0 then H(P,Q) = 0 yield for P with Q set to zero

$$P(\bar{Q}) = - \frac{H(\bar{P},\bar{Q})}{H(P',\bar{Q})}$$

Similarly P with 1/Q set to zero yields

$$P(Q') = - \frac{H(\bar{P},Q')}{H(P',\bar{Q})}$$

Thus Z_4 with $1/Z_2$ set equal to zero yields

$$Z_4(Z_2 = 1) = - \frac{Z(\bar{4},2')}{Z(4',2')} = Z_3$$

where $Z(\bar{4},2') = Z_3$ and $Z(4',2') = 1$, by scanning Z*.

Since H(P,Q) = 0 yields

$$P = - H(\bar{P})/H(P') \quad \text{and} \quad Q = - H(\bar{Q})/H(Q')$$

and dH = 0 yields $H(Q')dQ + H(P')dP = 0$

$$\frac{d\ lnP}{d\ lnQ} = \frac{dP}{dQ}\frac{Q}{P} = - \frac{H(Q')}{H(P')}\frac{H(\bar{Q})}{H(Q')}\frac{H(P')}{H(\bar{P})}$$

hence $S(P,Q) = -H(\bar{Q})/H(\bar{P})$

For example $S(Z_4,Z_2) = - \dfrac{Z_1Z_3 + Z_3Z_4 + Z_1Z_4}{Z_2Z_3 + Z_1Z_3}$

since $Z(\bar{4}) = Z_2Z_3 + Z_1Z_3$

and $Z(\bar{2}) = Z_1Z_3 + Z_3Z_4 + Z_1Z_4$

These are first-order vulnerability effects:

(a) <u>sensitivity</u> or drift of Z_4 with Z_2 is defined by the fractional or percentage change in Z_4, if Z_2 is changed by one percent, namely

$$(dZ_4/Z_4)/(dZ_2/Z_2) = Z(\bar{2})/Z(\bar{4})$$

Often also referred to as elasticity, the first-order effect is

$$\frac{d\ \ln P}{d\ \ln Q} = -\frac{H(\bar{Q})}{H(\bar{P})}$$

(b) <u>failure</u> of component Z_2 affects the performance parameter Z_4 of the system either by "open" or "short" of the element. Thus errors due to two kinds of catastrophic failure can be readily computed from Z^*, namely

$$Z(4,\bar{2}) = \frac{Z(\bar{4},\bar{2})}{Z(4',\bar{2})} \quad \text{and} \quad Z(4,2') = \frac{Z(\bar{4},2')}{Z(4',2')}.$$

A sharper definition of these concepts will now be attempted, followed by an extension to second-order concepts.

2. Systematics of First-Order Relationships

2.1 Sensitivity

Consider a linear system in which two elements P and Q satisfy the equation $H(P,Q) = 0$ where

$$H = H(\bar{P},\bar{Q}) + PH(P',\bar{Q}) + QH(\bar{P},Q') + PQH(P',Q') = 0 \quad (1)$$

The sensitivity of P with respect to Q is defined as

$$S(P,Q) = \frac{d \ln P}{d \ln Q} \quad (2)$$

Since $H(\bar{Q}) = H(\bar{P},\bar{Q}) + PH(P',\bar{Q})$ it is often desirable to eliminate P from the expression for sensitivity:

$$S(P,Q) = \frac{-H(\bar{Q})}{H(\bar{P})} = \frac{H(P',\bar{Q})\cdot H(\bar{P}) - H(\bar{P},\bar{Q})\cdot H(P')}{H(\bar{P})\cdot H(P')}$$

$$= \frac{H(P',\bar{Q})}{H(P')} - \frac{H(\bar{P},\bar{Q})}{H(\bar{P})}$$

In circuit analysis, it is often necessary to find the sensitivity with respect to 1/P with respect to Q. It is convenient to denote

$S(P',\bar{Q}) = H(P',\bar{Q})/H(P')$ and

$S(\bar{P},\bar{Q}) = H(\bar{P},\bar{Q})/H(\bar{P})$ hence

$S(P,Q) = S(P',\bar{Q}) - S(\bar{P},\bar{Q})$ and (3)

$S(\frac{1}{P},Q) = -S(P,Q) = S(\bar{P},\bar{Q}) - S(P',\bar{Q})$ (4)

In equations (3) and (4) the terms on the right hand side do not depend on the value of P and therefore make available by tagging only known symbols. To derive expressions of this type for sensitivity and error functions is the purpose of this investigation. Note the similarity in the function $S(\bar{P},\bar{Q})$ and $E(\bar{P},\bar{Q})$.

2.2 Error Due to Single Failure

Define the error in P, with Q = 0 (a short circuit if Q is an impedance) by

$$E(P',\bar{Q}) = \frac{P - P(\bar{Q})}{P} \tag{5}$$

and the error in P, with Q = (an open circuit if Q is an impedance) by

$$E(P',Q') = \frac{P - P(Q')}{P} \tag{6}$$

It follows that

$$E(P',\bar{Q}) = \frac{H(\bar{Q})}{H(\bar{Q}) - H(\bar{P},\bar{Q})} \tag{7a}$$

and

$$E(P',Q') = \frac{H(Q')}{H(Q') - H(\bar{P},Q')} \tag{7b}$$

Similarly, define the errors in $\frac{1}{P}$ by

$$E(\bar{P},\bar{Q}) = \frac{\frac{1}{\bar{P}} - \frac{1}{P(\bar{Q})}}{\frac{1}{\bar{P}}} \tag{8a}$$

$$E(\bar{P},Q') = \frac{\frac{1}{P} - \frac{1}{P(Q')}}{\frac{1}{P}} \tag{8b}$$

leading to

$$E(\bar{P},\bar{Q}) = \frac{H(\bar{Q})}{H(\bar{P},\bar{Q})} \tag{9a}$$

$$E(\bar{P},Q') = \frac{H(Q')}{H(\bar{P},Q')} \tag{9b}$$

These expressions are not yet useful for tagging since $H(\bar{Q})$ and $H(Q')$ depend on the value of P, but substituting

$$H(\bar{Q}) = H(\bar{P},\bar{Q}) - \frac{H(\bar{P})}{H(P')} \cdot H(P',\bar{Q}) \tag{10a}$$

$$H(Q') = H(\bar{P},Q') - \frac{H(\bar{P})}{H(P')} \cdot H(P',Q') \tag{10b}$$

yields the following useful result:

$$E(P',X) = 1 - \frac{H(\bar{P},X) \cdot H(P')}{H(P',X)\ H(\bar{P})} \tag{11a}$$

$$E(\bar{P},X) = 1 - \frac{H(P',X) \cdot H(\bar{P})}{H(\bar{P},X) \cdot H(P')} \tag{11b}$$

where X can be either $\bar{Q}$ or Q'.

2.3 Error Due to Multiple Failure

It is useful to determine the error in P or $\frac{1}{P}$ if two elements, Q and R, were to fail.

In general,

$$E(P';X,Y) \triangleq \frac{P - P(X,Y)}{P} \tag{12a}$$

$$E(\bar{P};X,Y) \triangleq \frac{\frac{1}{\bar{P}} - \frac{1}{P(X,Y)}}{\frac{1}{\bar{P}}} \tag{12b}$$

where $X = \bar{Q}$ or Q' and $Y = \bar{R}$ or R'.

The final result is analogous to the single failure case:

$$E(P';X,Y) = 1 - \frac{H(\bar{P},X,Y)}{H(P',X,Y)} \cdot \frac{H(P')}{H(\bar{P})} \tag{13a}$$

$$E(\bar{P};X,Y) = 1 - \frac{H(P',X,Y)}{H(\bar{P},X,Y)} \cdot \frac{H(\bar{P})}{H(P')} \tag{13b}$$

2.4 Sensitivity-Error Relations

Consider $E(\bar{P},\bar{Q}) = \frac{H(\bar{Q})}{H(\bar{P},\bar{Q})}$

$$E(\bar{P},Q') = \frac{H(Q')}{H(\bar{P},Q')}$$

Then

$$\frac{1}{E(\bar{P},Q')} - \frac{1}{E(\bar{P},\bar{Q})} = \frac{H(\bar{P},Q')}{H(Q')} - \frac{H(\bar{P},\bar{Q})}{H(\bar{Q})} = S(Q,P)$$

so that

$$\frac{1}{E(\bar{P},Q')} - \frac{1}{E(\bar{P},\bar{Q})} = \frac{1}{S(P,Q)} \tag{14}$$

3. Second-Order Relationships

3.1 Higher Order Sensitivity

Define the "sensitivity of a sensitivity" by

$$S(S(P,Q),R) = \frac{d\,\frac{d\ln P}{d\ln Q}}{d\,\ln R} \tag{15}$$

It has been shown (9) that

$$S(S(P,Q),R) = \frac{H(\bar{P},\bar{R})}{H(\bar{P})} - \frac{H(\bar{Q},\bar{R})}{H(\bar{Q})} \tag{16}$$

yielding, after simplifying, the result given in table 2.

3.2 Higher Order Error

Define the error of an error by

$$E(E(P',Q),R) = \frac{E(P',\bar{Q}) - E(P',\bar{Q},R)}{E(P',\bar{Q})} \tag{17}$$

Recall that $E(P',\bar{Q})$ is the error in P when Q = 0, and $E(P';\bar{Q},\bar{R})$ is the error in P when Q = R = 0.

In general,

$$E(E(P',X),Y) \triangleq \frac{E(P',X) - E(P';X,Y)}{E(P',X)} \tag{18a}$$

and

$$E(E(\bar{P},X),Y) \triangleq \frac{E(\bar{P},X) - E(\bar{P};X,Y)}{E(\bar{P},X)} \tag{18b}$$

Substituting (7) and (9) in these expressions and then expanding all terms which depend on P by the formula

$$H(X) = H(\bar{P},X) - \frac{H(\bar{P})}{H(P')} H(P',X) \tag{19}$$

yields expressions useful for computer analysis, given in table 2.

3.3 Sensitivity of Error

Define the sensitivity of an error by

$$S(E(P',Q),R) \triangleq \frac{d \ln E(P',\bar{Q})}{d \ln R} \tag{20}$$

Thus,

$$S(E(P',\bar{Q}),R) = \frac{d E(P',\bar{Q})}{d R} \cdot \frac{R}{E(P',\bar{Q})}$$

Using

$$E(P',\bar{Q}) = \frac{H(\bar{Q})}{H(\bar{Q}) - H(\bar{P},\bar{Q})} \quad \text{and} \quad \frac{d H(X)}{d R} = H(X,\bar{R})$$

$$\frac{dE(P',\bar{Q})}{d R} = \frac{H(\bar{Q},\bar{R}) \cdot \{H(\bar{Q}) - H(\bar{P},\bar{Q})\} - H(\bar{Q}) \{H(\bar{Q},\bar{R}) - H(\bar{P},\bar{Q},\bar{R})\}}{\{H(Q) - H(P,Q)\}^2}$$

$$= \frac{H(\bar{Q})\ H(\bar{P},\bar{Q},\bar{R}) - H(\bar{P},\bar{Q})\ H(\bar{Q},\bar{R})}{\{H(\bar{Q}) - H(\bar{P},\bar{Q})\}^2}$$

$$\frac{dE(P',\bar{Q})}{d R} \cdot \frac{R}{E(P',\bar{Q})} = \frac{\{H(\bar{Q}) H(\bar{P},\bar{Q},\bar{R}) - H(\bar{P},\bar{Q}) H(\bar{Q},\bar{R})\} \cdot R}{\{H(\bar{Q}) - H(\bar{P},\bar{Q})\}\{H(\bar{Q})\}}$$

To get a result useful for tagging, simply expand $H(\bar{Q})$ and $H(\bar{Q},\bar{R})$ by (20) and use

$$R = \frac{-H(\bar{R})}{H(R')}$$

and then expand $H(\bar{R})$ and $H(R')$ by (19).

The final result is given in table 3, with similar results for $S(E(\bar{P},\bar{Q}),R)$ as well as the analogous case for Q' in place of $\bar{Q}$.

3.4 Error of Sensitivity

Define the error of a sensitivity by

$$E(S(P,Q),\bar{R}) \triangleq \frac{S(P,Q) - S(P,Q,\bar{R})}{S(P,Q)} \qquad (21)$$

where

$$S(P,Q,\bar{R}) \triangleq S(P,Q) \Big|_{R=0}$$

Through computations analogous to those used in deriving the expressions for error of error, the final results in table 1 are obtained.

4. Discussion of Binary Base Tagging

Table 5 lists terms obtained from an expansion of $H = 0$ in terms of three parameters P, R and Q. If the expression $H(P,Q,R)$ is tagged in terms of these parameters, the second-order effects evaluated in Tables 1-4 are obtained.

5. Perspective: System Vulnerability due to Unreliable Components

The vulnerability of a system consisting of numerous interacting components is measured by the change of a specified performance parameter of the system (P) caused by a unit change in a component value (Q). The classical example, examined by Bode (1) in 1941 is the change in current amplification (P) of an electronic circuit, if the transconductance of the vacuum tube (Q) is changed. Although Bode does not explicitly define two distinct types of vulnerability, two interrelated aspects of vulnerability are recurring frequently in his work:

> Sensitivity or drift S(P;Q) is defined as the fractional or logarithmic change d(ln P)/d(ln Q) of the performance parameter (P) of the system caused by corresponding change of the component parameter (Q). Mathematics texts refer to S(P;Q) as "elasticity", while the electronics terminology refers to small changes of component values affecting system performance as "drift".
>
> Failure of a component (Q) affects the system performance differently and is measured by the error in systems performance E(P;Q). The error E(P;Q) is caused by catastrophic failure of Q and affects the system performance parameter (P). Failure of a component in

an energy transforming system usually results in one of two predictable extremum conditions of the component. In circuits these conditions consist of open-circuit or short circuit failure of a component. In probabilistic networks the extrema are certainty of occurrence or of non-occurrence of an event.

Although a large number of investigators have drawn heavily on Bode's work, the overwhelming majority of investigations are devoted exclusively to one type of vulnerability, primarily sensitivity. Despite an extensive literature on sensitivity, most investigation fails to correlate advances in sensitivity with corresponding advances in failure or "error" analysis.

About twenty years elapsed between Bode's fundamental work and the next major step (2,3): the exploitation of the isomorphism between graph theory - referred to by engineers as "topological techniques" - and vulnerability - frequently examined by engineers as part of "reliability analysis".

During the past decade (1960-1970), automated evaluation techniques have extended vulnerability calculations significantly. The size of a system, which can be analyzed by an economically purposeful effort has increased by two orders of magnitude. For systems with lumped components, for example an electronic circuit, conventional techniques are concerned with systems of 5 to 10 components, while computer-oriented tech

niques are attempting corresponding calculations for systems with 20 to 50 components. This change in design philosophy has triggered a vigorous search to extend vulnerability concepts. Three examples will illustrate this:

Preassigned accuracy techniques: A computer-aided or computer-oriented selection obtains from the original system a simplified or an approximate system by eliminating components. The approximate system must be capable of supplying perfcrmance data with a pre-specified accuracy. To illustrate, assume that a system with 200 components is described by a performance parameter, which is needed with only 10% accuracy. Except for pathological cases, less than 10 components can each contribute more than 10% to the systems performance. Thus it is likely that 20 components will provide a model to describe the system within the preassigned accuracy. An aucquate, more economical model is thus established.

Binary base tagging: A scanning procedure for efficiently identifying groups of binary symbols, is used to replace far lengthier numerical calculations. An application was presented here using this technique for vulnerability calculation. Since vulnerability calculations by their very nature involve second order effects in which small differences are of paramount im-

portance, numerical calculations are subject to significant round-off errors. If these errors are cumulative, iterations may lead to numerical instability. Symbol manipulations lead to calculations which in practice exclude the possibility of numerical instability and its inherent difficulty of detection. Tagging and similar symbol manipulation techniques are preferable

(i) when the numerical stability of calculations is in doubt

(ii) when parametric solutions are required, thus maintaining some symbols and replacing others by numbers

(iii) when special techniques, such as vulnerability methods, can be formulated more efficiently by symbol manipulation than by numerical approaches, for example differentiation of a function.

Worst case design: The effect of all possible simultaneous parameter variations are combined to ascertain which combination yields the widest fluctuations for a specified performance parameter. It is difficult to anticipate the effect of the variations of one parameter, on a specified system performance parameter, since several components are statistically likely to cause

changes which compensate each other. Conventional optimization techniques employ successive approximations to arrive at a unique selection of parameters. Alternatively it is possible to develop appropriate properties of vulnerability, which will assure in each instant that the "worst case" for each component is selected for obtaining the specified effect on the system.

These three examples merely illustrate the range of applications of vulnerability analysis, which emerges as an expanding but nearly unexplored frontier in systems science. The definitions and concepts of second order effects were examined and basic formalisms were developed. Higher order effects are still one of the unexplored areas deserving attention.

This investigation aims at being the first in a sequence of reports on vulnerability studies. Subsequent investigations will apply these formalisms developed here to multiterminal networks, thermodynamic fluctuations and similar physical systems as well as to modelling of production flow and to process control.

REFERENCES

1. H. W. Bode, Network Analysis and Feedback Amplifier Design, Van Nostrand, 1941.

2. R. Tomovic, Sensitivity Analysis of Dynamic Systems, McGraw-Hill, 1963.

3. W. W. Happ and D. E. Moody, "Topological Techniques for Sensitivity Analysis", IEEE Trans. on Aerospace and Navigational Electronics, Vol. ANE-11, pp. 249-254, Dec. 1964.

4. W. W. Happ, "Flowgraph Techniques for Closed Systems" IEEE Trans. AES 2, No. 3, pp. 252-264, 2 May, 1966.

5. W. W. Happ, "Correspondence Principle for Enumeration: Trees and Loops", Proceedings, Eleventh Midwest Symposium on Circuit Theory, Notre Dame, Ind., p. 204-213, May 1968.

6. M. F. Goldberg and J. Vaccaro (Editors), Physics of Failure in Electronics, U.S. Dept. of Commerce, Office of Technical Services, Washington, D.C., AD-434/349, 1964.

7. R. D. Lamb, E. R. Robbins and W. W. Happ, "Circuit Transformations based upon Topology", Proceedings of the Allerton Conference, pp. 508-529, University of Illinois, Sept. 1964.

8. W. W. Happ, "Present Capabilities of a Maintained Program", IEEE International Convention Record, Part 5. pp. 64-88, March 1967.

9. W. W. Happ and M. C. Weinberg, "Variance Analysis Applied to Performance Error from Component Failure in Microcircuits", Proceedings, Southeastern Symposium on System Theory, V.P.I., Blacksburg, Va., pp. EI-1 to EI-11, May 1969.

10. H. Potash and L. P. McNamee, "Applications of Unilateral and Graph Techniques to the Analysis of Linear Circuits", ACM Journal, Oct. 1968.

11. J. B. Compton and W. W. Happ, "Flowgraph Models of Integrated Circuits and Film-Type Systems", IEEE Trans. AS-2, pp. 259-271, 1964.

12. J. B. Compton and W. W. Happ, "Modelling of Radiation Effects in Semiconductor Devices and Circuits", IEEE Trans. Vol. AS-3, pp. 372-378, June 1965.

13. A. A. B. Pritsker and W. W. Happ, "GERT: Graphical Evaluation and Review Technique Part I, Development", Journal of Industrial Engineering, Vol. 17, pp. 267-274, May 1966.

14. G. E. Whitehouse and A. A. B. Pritsker, "GERT: Graphical Evaluation and Review Technique Part II, Applications", Journal of Industrial Engineering, Vol. 17, pp. 293-301, June 1966.

15. G. E. Whitehouse and A. A. B. Pritsker, "GERT: Graphical Evaluation and Review Technique Part III, Generating Functions, Counters, Renewal Times, and Correlations", AIIE Transactions, Vol. I, No. 1, pp. 45-50, 1969.

16. M. J. Devaney, G. W. Zobrist, and W. W. Happ, "Sensitivity Analysis", Proceedings, Tenth Midwest Circuit Symposium, Purdue University, May 1967.

17. J. Staudhammer and W. W. Happ, "Topological Techniques for the Derivation of Models with Pre-Assigned Accuracy", Archiv der Elektrischen Ubertragung, Vol. 20, HEFT 6, pp. 329-335, 20, June 1966.

18. G. J. Herskowitz, Computer-Aided Integrated Circuit Design, McGraw-Hill, Dec. 1968.

19. G. W. Zobrist, Computer-Aided Network Design, Boston Technical Publishers, May 1969.

20. A. A. B. Pritsker, "The Status of GERT" in H.J.M. Lombaers (Editor) Project Planning by Network Analysis, North Holland Publishing Co., Amsterdam, London, 1969.

21. R. M. Carpenter and W. W. Happ, "Algorithms for the Dichotomous Representation of Systems", IEEE Trans., Vol. SSC-3 1, pp. 39-46, June 1967.

22. R. M. Carpenter and W. W. Happ, "Computer-Aided Design: Part 3 - Analyzing Circuits with Symbols", Electronics, pp. 92-98, Dec. 1966.

23. E. R. Robbins, D. E. Moody and W. W. Happ, "Modelling Piecewise-Linear Systems by Topological Techniques", IEEE Trans. AES, Vol. AES-2, No. 4, pp. 370-375, July 1966.

24. P. V. Kokotovic and R. S. Rutman, "Sensitivity of Automatic Control Systems", Automatica i Telemekhamika, Vol. 26, No. 4, pp. 730-750, April 1965.

25. W. W. Happ, "Flowgraphs as a Teaching Aid", IEEE Trans. on Education, Vol. E-9, No. 2, pp. 69-79, June 1966.

26. W. W. Happ and S. C. Gupta, "Time Domain Analysis and Measurement Techniques for RC Structures, Part I, Analysis in the Reciprocal Time Domain", J. Appl. Physics, Vol. 40, No. 1, pp. 109-117, Jan. 1969.

27. R. C. Peirson and E. C. Bertnolli, "Time Domain Analysis and Measurement Techniques for RC Structures, Part II, Impulse Measurement Techniques", J. Appl. Physics, Vol. 40, No. 1, pp. 118-122, Jan. 1969.

28. J. P. Roth, "An Application of Algebraic Topology to Numerical Analysis: on the existence of a solution to the network problem", Proc. National Acad. Sciences, Vol. 41, pp. 518-521, 1955.

29. G. Kron, "A Set of Principles to Interconnect the Solut' .n of Physical Systems", J. Appl. Physics, Vol. 24, pp. 6 ,- 980, 1953.

30. A. Krajewski and W. W. Happ, "Dewiacje - Nowa Metodc Preksztatcew Parametrow Elektrycznych", Rosprawy Electrotechneczne, Tom XII 4, pp. 675-706, Warsaw, Poland, October 1966.

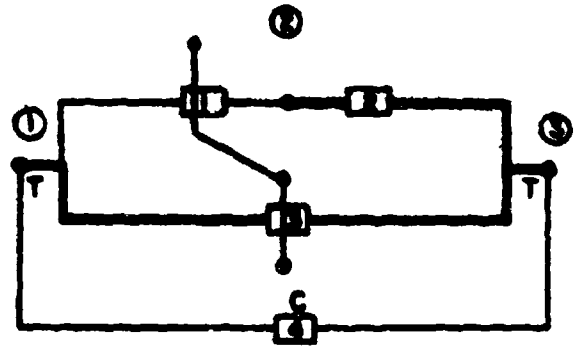
31. W. W. Happ, (Editor) "Special Issue on Educational Aspects of Circuit Design by Computer", IEEE Trans. on Education, Sept/Dec. 1969.

32. L. DePian, Linear Active Network Theory, Prentice-Hall, 1962.

33. P. M. Chirlian, Integrated and Active Network Analysis and Synthesis, Prentice-Hall, 1967.

34. G. Kron, Tensor Analysis of Networks, Wiley, 1939.

35. M. C. Weinstein and W. W. Happ, Symbol-oriented Failure Analysis in Multicomponent System, NASA TN D-5112, 14 pp., CFSTI, Springfield, Va., April 1969.



T: TERMINALS OF OPEN SYSTEM

C: CLOSING ELEMENT

FIGURE 1

Closed System Showing Cut (1010) and Tie (0110)

TABLE 1

Basic Formulas for Error and Sensitivity

N/D	Numerator (N)	Denominator (D)
$1-E(P';\overline{Q})$	$H(\overline{P},\overline{Q}) \cdot H(P')$	$H(P',\overline{Q}) \cdot H(\overline{P})$
$1-E(P';Q')$	$H(\overline{P},Q') \cdot H(P')$	$H(P',Q') \cdot H(\overline{P})$
$1-E(\overline{P};\overline{Q})$	$H(P',\overline{Q}) \cdot H(\overline{P})$	$H(\overline{P},\overline{Q}) \cdot H(P')$
$1-E(\overline{P};Q')$	$H(P',Q') \cdot H(\overline{P})$	$H(\overline{P},Q') \cdot H(P')$
$S(P;Q)$	$\begin{vmatrix} H(P',\overline{Q}) & H(\overline{P},\overline{Q}) \\ H(P') & H(\overline{P}) \end{vmatrix}$	$H(\overline{P}) \cdot H(P')$

TABLE 2

Second-Order Error and Sensitivity

N/D	Numerator (N)	Denominator (D)
$1-E(E(P';\overline{Q});\overline{R})$	$H(P',\overline{Q}) \cdot \begin{vmatrix} H(\overline{P},\overline{Q},\overline{R}) & H(P',\overline{Q},\overline{R}) \\ H(\overline{P}) & H(P') \end{vmatrix}$	$H(P',\overline{Q},\overline{R}) \cdot \begin{vmatrix} H(\overline{P},\overline{Q}) & H(P',\overline{Q}) \\ H(\overline{P}) & H(P') \end{vmatrix}$
$1-E(E(\overline{P};\overline{Q});\overline{R})$	$H(\overline{P},\overline{Q}) \cdot \begin{vmatrix} H(\overline{P},\overline{Q},\overline{R}) & H(P',\overline{Q},\overline{R}) \\ H(\overline{P}) & H(P') \end{vmatrix}$	$H(\overline{P},\overline{Q},\overline{R}) \cdot \begin{vmatrix} H(\overline{P},\overline{Q}) & H(P',\overline{Q}) \\ H(\overline{P}) & H(P') \end{vmatrix}$
$1-E(E(P';X);Y)$	$H(P',X) \cdot \begin{vmatrix} H(\overline{P},X,Y) & H(P',X,Y) \\ H(\overline{P}) & H(P') \end{vmatrix}$	$H(P',X,Y) \cdot \begin{vmatrix} H(\overline{P},X) & H(P',X) \\ H(\overline{P}) & H(P') \end{vmatrix}$
$1-E(E(\overline{P};X);Y)$	$H(\overline{P},X) \cdot \begin{vmatrix} H(\overline{P},X,Y) & H(P',X,Y) \\ H(\overline{P}) & H(P') \end{vmatrix}$	$H(\overline{P},X,Y) \cdot \begin{vmatrix} H(\overline{P},X) & H(P',X) \\ H(\overline{P}) & H(P') \end{vmatrix}$
$S(S(P;Q);R)$	$H(\overline{P},\overline{R}) \cdot \begin{vmatrix} H(\overline{P},\overline{Q}) & H(P',\overline{Q}) \\ H(\overline{P}) & H(P') \end{vmatrix} - H(\overline{P}) \cdot \begin{vmatrix} H(\overline{P},\overline{Q},\overline{R}) & H(P',\overline{Q},\overline{R}) \\ H(\overline{P}) & H(P') \end{vmatrix}$	$H(\overline{P}) \cdot \begin{vmatrix} H(\overline{P},\overline{Q}) & H(P',\overline{Q}) \\ H(\overline{P}) & H(P') \end{vmatrix}$

TABLE 3

Sensitivity of Error

N/D	Numerator (N)	Denominator (D)
$S(E(P';\overline{Q});R)$	$H(P')\begin{vmatrix} H(\overline{P},\overline{Q},R') & H(P',\overline{Q},R') \\ H(\overline{P},\overline{Q}) & H(P',\overline{Q}) \end{vmatrix} \cdot \begin{vmatrix} H(\overline{P}) & H(P') \\ H(\overline{P},\overline{R}) & H(P',\overline{R}) \end{vmatrix}$	$H(P',\overline{Q})\begin{vmatrix} H(\overline{P},\overline{Q}) & H(P',\overline{Q}) \\ H(\overline{P}) & H(P') \end{vmatrix} \cdot \begin{vmatrix} H(\overline{P},R') & H(P',R') \\ H(\overline{P}) & H(P') \end{vmatrix}$
$S(E(P';Q');R)$	$H(P')\begin{vmatrix} H(\overline{P},Q'R') & H(\overline{P}) \\ H(\overline{P},Q') & H(P',Q') \end{vmatrix} \cdot \begin{vmatrix} H(\overline{P}) & H(P') \\ H(\overline{P},\overline{R}) & H(P',\overline{R}) \end{vmatrix}$	$H(P',Q')\begin{vmatrix} H(\overline{P},Q') & H(P',Q') \\ H(\overline{P}) & H(P') \end{vmatrix} \cdot \begin{vmatrix} H(\overline{P},R') & H(P',R') \\ H(\overline{P}) & H(P') \end{vmatrix}$
$S(E(\overline{P};\overline{Q});R)$	$H(\overline{P})\begin{vmatrix} H(\overline{P},\overline{Q},R') & H(P',\overline{Q},R') \\ H(\overline{P},\overline{Q}) & H(P',\overline{Q}) \end{vmatrix} \cdot \begin{vmatrix} H(\overline{P}) & H(P') \\ H(\overline{P},\overline{R}) & H(P',\overline{R}) \end{vmatrix}$	$H(\overline{P},\overline{Q})\begin{vmatrix} H(\overline{P},\overline{Q}) & H(P',\overline{Q}) \\ H(\overline{P}) & H(P') \end{vmatrix} \cdot \begin{vmatrix} H(\overline{P},R') & H(P',R') \\ H(\overline{P}) & H(P') \end{vmatrix}$
$S(E(\overline{P};Q');R)$	$H(\overline{P})\begin{vmatrix} H(\overline{P},Q',R') & H(P',Q',R') \\ H(\overline{P},Q') & H(P',Q') \end{vmatrix} \cdot \begin{vmatrix} H(\overline{P}) & H(P') \\ H(\overline{P},\overline{R}) & H(P',\overline{R}) \end{vmatrix}$	$H(\overline{P},Q')\begin{vmatrix} H(\overline{P},Q') & H(P',Q') \\ H(\overline{P}) & H(P') \end{vmatrix} \cdot \begin{vmatrix} H(\overline{P},R') & H(P',R') \\ H(\overline{P}) & H(P') \end{vmatrix}$

TABLE 4

Error in Sensitivity

N/D	Numerator (N)	Denominator (D)
1-E(S(P;Q);$\bar{R}$)	$H(\bar{P})\cdot H(P')\cdot \begin{vmatrix} H(\bar{P},\bar{R}) & H(P',\bar{R}) \\ H(\bar{P},\bar{Q},\bar{R}) & H(P',\bar{Q},\bar{R}) \end{vmatrix}$	$H(\bar{P},\bar{R})\cdot H(P',\bar{R})\cdot \begin{vmatrix} H(\bar{P}) & H(P') \\ H(\bar{P},\bar{Q}) & H(P',\bar{Q}) \end{vmatrix}$
1-E(S(P;Q);R')	$H(\bar{P})\cdot H(P')\cdot \begin{vmatrix} H(\bar{P},R') & H(P',R') \\ H(\bar{P},\bar{Q},R') & H(P',\bar{Q},R') \end{vmatrix}$	$H(\bar{P},R')\cdot H(P',R')\cdot \begin{vmatrix} H(\bar{P}) & H(P') \\ H(\bar{P},\bar{Q}) & H(P',\bar{Q}) \end{vmatrix}$

TABLE 5

Symbols Needed in Computation of Sensitivity and Error

	$H(\bar{P})$	$H(P')$	$H(\bar{P},\bar{Q})$	$H(\bar{P},Q')$	$H(\bar{P},\bar{R})$	$H(\bar{P},R')$	$H(P',\bar{Q})$	$H(P',Q')$	$H(P',\bar{R})$	$H(P',R')$	$H(\bar{P},\bar{Q},\bar{R})$	$H(\bar{P},\bar{Q},R')$	$H(\bar{P},Q',\bar{R})$	$H(\bar{P},Q',R')$	$H(P',\bar{Q},\bar{R})$	$H(P',\bar{Q},R')$	$H(P',Q',\bar{R})$	$H(P',Q',R')$
E(P';$\bar{Q}$)	X	X	X				X											
E(P';Q')	X	X		X				X										
E(P;Q)	X	X	X				X											
E(P';Q')	X	X		X				X										
E(P';Q,R)	X	X									X				X			
E(P';Q,R')	X	X										X				X		
E(P';Q,R')	X	X											X				X	
E(P';Q',R')	X	X												X				X
E(P;Q,R)	X	X									X				X			
E(P;Q,R')	X	X										X				X		
E(P;Q',R)	X	X											X				X	
E(P;Q'R')	X	X												X				X
S(P,Q)	X	X	X				X											
E(E(P';$\bar{Q}$);$\bar{R}$)	X	X	X				X				X				X			
E(E(P';Q);R')	X	X	X				X					X				X		
E(E(P';Q');R)	X	X		X				X					X				X	
E(E(P';Q');R')	X	X		X				X						X				X
E(E(P;Q);R)	X	X	X				X				X				X			
E(E(P;Q);R')	X	X	X				X					X				X		
E(E(P,Q');R)	X	X		X				X					X				X	
E(E(P;Q')R')	X	X		X				X						X				X
S(S(P;Q);R)	X	X	X		X		X				X				X			
S(E(P';Q);R)	X	X	X		X	X	X		X	X		X				X		
S(E(P';Q');R)	X	X		X	X	X		X	X	X				X				X
S(E(P;Q);R)	X	X	X		X	X	X		X	X		X				X		
S(E(P;Q');R)	X	X		X	X	X		X	X	X				X				X
E(S(P;Q);$\bar{R}$)	X	X	X		X		X		X		X				X			
E(S(P;Q);R')	X	X	X			X	X			X		X				X		

TIME CONSTRAINED RELIABILITY DATA DEVELOPMENT FOR RADIO EQUIPMENT IN A GROUND-BASED LABORATORY

C.E. Deckard and T.K. DeClue
Wyle Laboratories
Huntsville, Alabama

ABSTRACT

A ground-based laboratory test program was conceived for the purpose of economically acquiring a qualitative measure of the MTBF of radio equipment subjected to a helicopter combat flight environment. The initial portion of the program was time constrained with a very small number (two) of test specimens available. This dictated departure from the normal single-axis testing technique. A system designed to provide testing in the three orthogonal axes, simultaneously, was implemented and utilized to acquire the desired data.

INTRODUCTION

A requirement for determining the system effectiveness of the AN/ARC 115 Radio Set (used in the OH-58 Helicopter) was recently set forth by the United States Aviation Test Board, Fort Rucker, Alabama. At present, flight tests are being conducted to obtain the data required for establishing the equipment MTBF. The execution of these flight tests demands:

- Extensive use of personnel and aircraft.
- Schedule difficulties due to aircraft and flight crew priorities.
- High costs.

It is felt that system effectiveness may be economically determined by subjecting the radio set to a controlled series of dynamics tests in a ground-based laboratory. The advantages of laboratory testing are

- Control of conditions
- Control of test scheduling
- Simplification of analysis
- Repeatability
- Economy
- Safety

A high degree of simulation realism would be imperative in such an approach; the primary system requirement would be the development of a test system capable of imposing the actual flight dynamics inputs in the specimen's three orthogonal axes simultaneously. Such a system should provide the data necessary for developing an accurate measure of the equipment MTBF since it would effectively reproduce the service environment. The program under discussion can be called a "pilot program," the primary purpose of which is to determine both the feasibility and cost savings to be realized by ground-based laboratory testing.

The purpose of this paper is to describe the test system designed to provide the dynamic inputs that would effectively simulate actual combat flight conditions. This system is presently operational, having provided approximately 775 hours of flight simulation to the radio equipment. Though more data are required to make a final, detailed, comparison between the flight and ground-based tests, excellent results have been obtained in the laboratory. At present, the number of radio sets under test has been increased from two to six.

Though the primary intent of the paper is to describe the test system, the general design philosophy used in developing the system is also discussed. Certain portions of the design presented herein are proprietary with patent rights pending; although the techniques inherent in the design philosophy are available to those who wish to use them.

TEST SYSTEM DESIGN PHILOSOPHY

The procedures employed in the design of a dynamics test system begins with an evaluation of the important aspects of the service environment, viz. the origins, transmission paths and coupling modes of the forces acting on the specimen during service. The primary objective is to assess the magnitudes and frequency spectra of these forces and couple them to the specimen with impedances similar to those of the mountings and supports used in actual service. When possible, the actual support hardware should be used in the testing configuration in order to provide a more exact duplication of the service boundary conditions. The more complex the test item and its support hardware are, the greater is the need for accurate boundary reproduction and the more difficult it is to provide simulation of the boundary through manipulation of the inputs and test fixtures.

An initial step in the design of the test system is to determine how large a section of the entire system, of which the specimen is a part, can be accommodated within the laboratory space and exciter force limits. Secondly, determination of the number of degrees of freedom of motion that the specimen experiences must be made. Where, through analyses or data examination of the actual environment, it is found that motion is insignificant or does not occur in or about one or more of the axes, those degrees of freedom may be restrained without adversely effecting the test system or resulting test data. The design and construction of a six degrees of freedom system, although conceivable with present day technology tends to be extremely expensive. Since the rotational axes in typical flight environments generally experience minimal frequency and amplitude response, consideration should be given to the three translational axes. Thus, the design of a test system capable of excitation in the three translational axes (while restraining the three rotational axes) would allow the use of available single degree of freedom electrodynamic exciters at modest expense.

DETERMINING THE ENVIRONMENT

An OH-58 Helicopter was instrumented on and about the AN/ARC 115 radio set with a total of seven piezoelectric accelerometers for the purpose of monitoring the in-flight vibratory environment. Figure 1 illustrates the helicopter flight control panel and the accelerometer locations.

Figure 1. Helicopter Flight Control Panel; AN/ARC 115 Radio and Monitoring Accelerometers

A 14-channel tape recorder was installed onboard the aircraft to record all accelerometer output signals. During the flight tests, one channel was reserved for voice commentary to cover both the general test conditions and the event time. A 1k Hz acoustic calibrator was used to produce a constant amplitude audible tone for periodic transmitter modulation checks during the flight tests.

A total of two complete flight profiles were performed during flight testing. These profiles were:

- Flight Profile No. 1
  - One minute hover (3 feet)
  - Seven minute flight at maximum velocity attainable, not to exceed VNE (velocity never exceed) at 1500 ft MSL (mean sea level)
  - Simulated diving firing with 60 degree bank, 180 degree side entry and right pull up
  - Twenty minute loitering at 60-70 knots
  - Diving firing with 60 degree bank, 180 degree side entry and right pull up
  - Twenty minute loitering at 60-70 knots
  - Simulated diving firing with 60 degree bank, 180 degree side entry and right pull up
  - Twenty minute loitering at 60-70 knots
  - Diving firing with 60 degree bank, 180 degree side entry and right pull up
  - Twenty minute loitering at 60-70 knots
  - Simulated diving firing with 60 degree bank, 180 degree side entry and right pull up
  - Twenty minute loitering at 60-70 knots
  - Diving firing with 60 degree bank, 180 degree side entry and right pull out
  - Seven minute flight at maximum velocity attainable not to exceed VNE at 1500 ft MSL
  - One minute hover (3 feet)
- Flight Profile No. 2

  This mission was conducted in the same manner as Flight Profile No. 1 with the following exceptions:

  a. The simulated diving firing and diving firing order were reversed.

  b. All diving pull ups were made to the left.

Vibration data were continuously recorded during take-off dash to VNE, each phase of the flight profiles and landings, transmitter carrier output and modulation were periodically measured and recorded throughout each flight test. At the conclusion of each flight test, the taped data were played back into a recording oscillograph for visual display. This allowed examination of the recording process and provided a means of determining the need for possible retest.

The vibration data obtained during the actual flight test were reduced to X-Y plots, oscillograph records, and computer analyzed power spectral density (PSD) plots. These data were analyzed and reviewed with cognizant Fort Rucker personnel. Figure 2 shows typical PSD plots of accelerometer data.

Those channels that displayed the highest levels of vibration in the three orthogonal axes of the radio set were isolated and selected for providing the dynamic inputs into the test system. A master tape recording was prepared from these selected channels. The master tape contained the most dynamically severe portions of the two flight profiles flown. Thus, a basic test system input profile, approximately two hours in length, was developed. As anticipated, the most severe dynamic environment occurred when the helicopter was diving and firing its weapons system simultaneously. This flight phase occupied approximately 0.5 per cent of the two hour test profile.

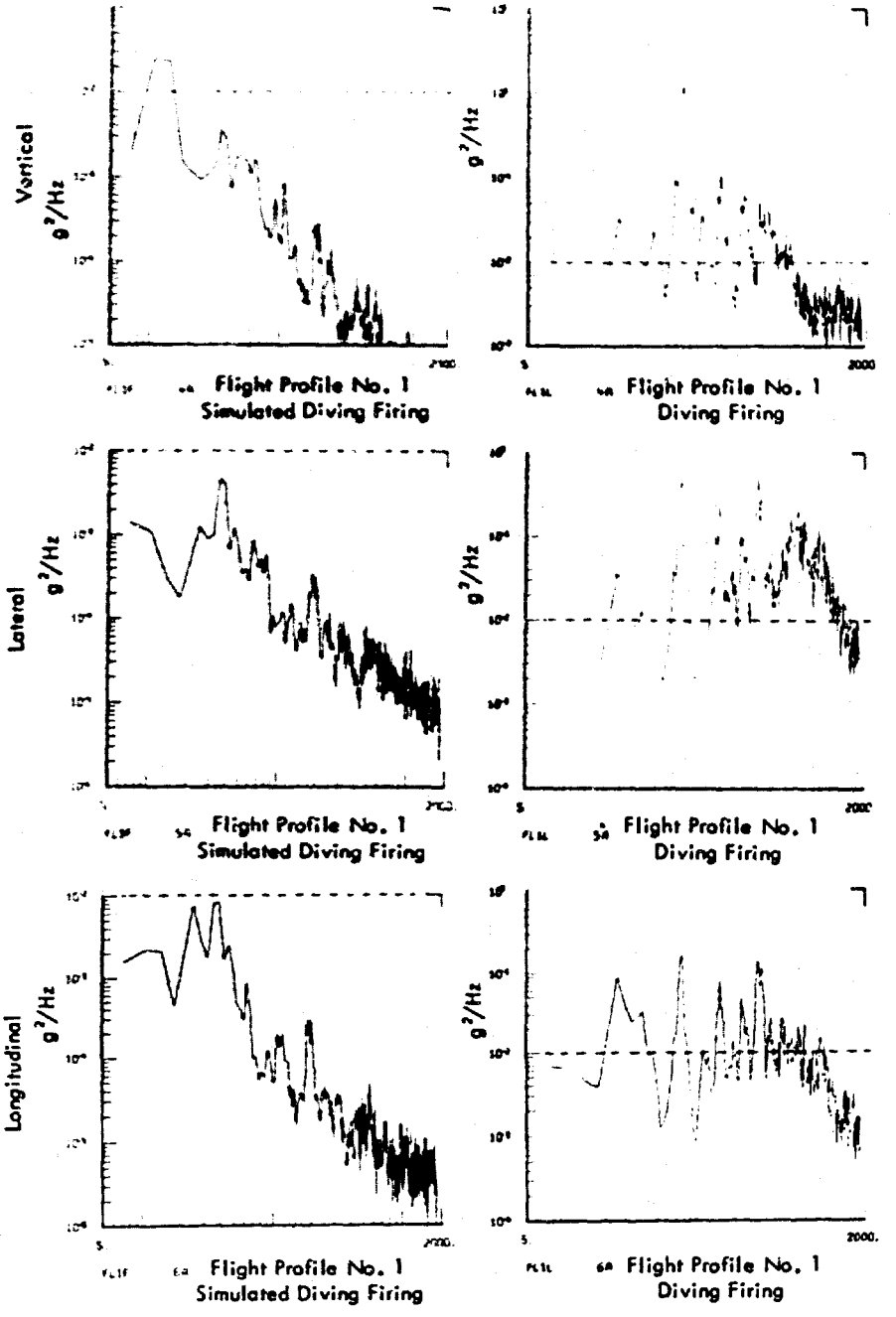


Figure 2. Typical PSD Plots of Accelerometer Data

EVALUATION OF THE ENVIRONMENT

Having defined the dynamic environment, it now became necessary to evaluate the environment and those techniques that would most realistically reproduce the environment in the laboratory. It was realized that an ordinary set of three, uni-axial vibration tests would not provide the data that is required for estimating the service life or MTBF of an item, in as much as cross-axes coupling can cause failures not reproducible in single degree of freedom testing. In reality a set of tests of N hours duration in the three orthogonal axes may result in either a greater or less severe environment for sub-components located within the test item than would N hours of the actual service environment. Also, where it is possible that "rattle space" problems might occur, uni-axial testing does not exercise all the internal degrees of freedom of the device simultaneously, as does multi-axial testing.

Consider a typical electronic component mounted to a surface (Figure 3) by bending its axial leads at right angles to the body and securing the free ends.

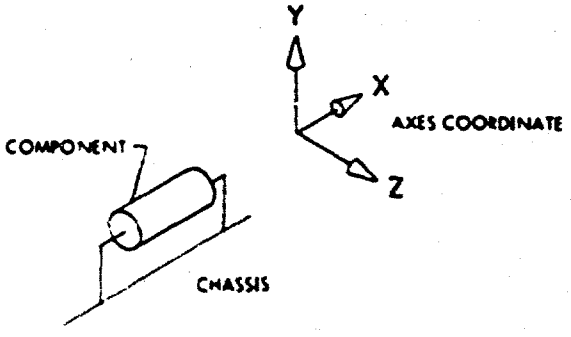


Figure 3. Electronic Component and Axes Coordinate

Due to the configuration of the support, its stiffness is different in the three orthogonal axes; it is most stiff in the vertical (Y) axis, moderately stiff in the longitudinal (X) axis, and softest in the transverse (Z) axis. The ratio of the three stiffnesses might be of the order 100:10:1 and, since resonance frequencies are related to the square root of stiffness, the resonance frequencies would have the ratio of 10:3.16:1.

The deflection of the body, and, therefore the stress, is proportional to the inverse of frequency squared and would be of the ratio 1:10:100 (inversely proportional to stiffness). Fatigue, being a function of the product of stress and number of cycles, would be in the ratio of 10:31.6:100, assuming that the inputs in all three axes were of the same duration same amplitude and of a flat acceleration versus frequency spectrum.

If the component is mounted such that its body axes are parallel to the three orthogonal vibration axes ordinarily selected for sets of single-axis tests, and the single axis vector inputs are applied to the three axes for an equal length of time, N, the fatigue is proporational to Σ_σ = (10 KNZ) + (3.16 KNX) + (KNY).

If however, the component body axes were all rotated, say 45°, from the input axes, the fatigue would be proportional to:

$$\Sigma_\sigma = 3(10\ KNZ \sin\alpha + 3.16\ KNX \sin\beta + 1\ KNY \sin\gamma)$$

The Σ_σ for the first case is 14.16 KN whereas, for the second case, where $\sin\alpha = \sin\beta = \sin\gamma = .707$, the sum is 29.89 KN. The difference of 2:1 is, of course, explained by the resolution of the input vectors into body reaction vectors and responses.

If the three inputs in the three axes were imposed simultaneously, and were random both in phase and frequency, or were reproductions of the actual service environment, the vectors would resolve properly to indicate a true service life. However, the application of a synthetic environment by testing in the three axes separately does not allow proper resolution of the input vector forces in such a way as to provide a quantitative measure of actual service life.

SIMULATION OF THE ENVIRONMENT

The primary principle on which the design of the three degrees of freedom test system rests is illustrated in Figure 4. This device, * (called the Deckard device) while providing a high degree of mechanical stiffness in the axial direction is compliant in all radial directions. It should be noted that when a force is applied radially at the top of the drive rod, the drive rod will translate only, viz. it will not move in an arcing motion.

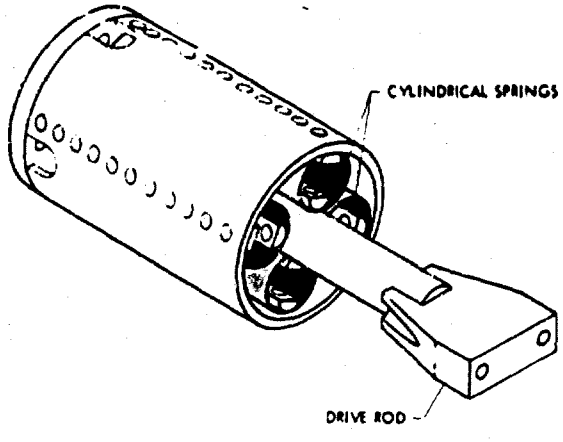


Figure 4. Cylindrical Spring Drive Assembly

* Invented by C.E. Deckard, patent applied for.

Figure 5 further illustrates this concept combined to provide a three degrees of freedom motion system. This particular concept provides freedom of motion in the three orthogonal axes only, while providing restraint of motion in the three rotational axes.

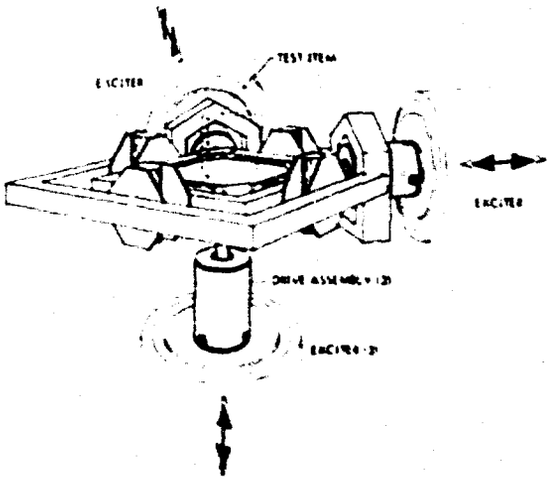


Figure 5. Three-Degrees-of-Freedom Vibration System

Figure 6. Three-Degrees-of-Freedom Vibration System

The two radio sets were mounted by their normal bracketry to the vibration test system shown in Figures 6 and 7. Three electrodynamic exciters were used to provide the three dynamic vibratory inputs to the system. The master tape, containing the test profile, was played into the control system shown in Figure 8. These input signals were amplified and passed through filtering networks that were tuned to compensate for fixture and exciter resonances. A master control unit was incorporated to provide total control of the entire system.

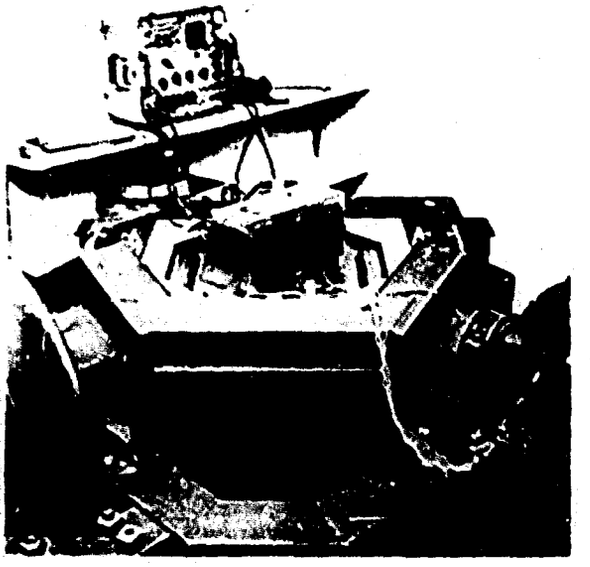


Figure 7. Three-Degrees-of-Freedom Vibration System

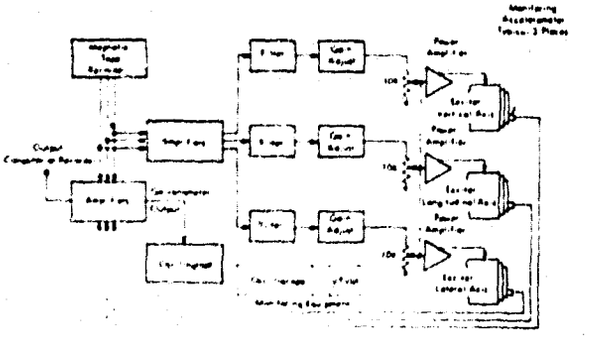


Figure 8. Control System for Three-Degrees-of-Freedom Vibration System

Three control accelerometers were located on the vibration fixture (one for each exciter). The output signals from the accelerometers were monitored during testing to insure proper vibration levels and observe any test anomalies. Periodically during the test, the output signals from the control accelerometers and the tape input signals to the three axis system were recorded and analyzed on a digital computer. The data obtained from the laboratory tests compared favorably with the data obtained from the flight tests. This comparison was extremely good to 100 Hz. The energy recorded at the higher frequencies was believed to be a result of acoustic coupling between the Helicopter drive system, the weapons system (when firing) and the radio set. Since this energy was not mechanically coupled through the aircraft structure and mounting bracketry, emphasis was placed on the lower, mechanically coupled, frequencies in the design of the test system. The total Grms values of the taped input and fixture response spectra were quite close. Table I lists the comparisons of the total Grms levels for the most severe portion of the simulated profile.

TABLE I

DIVING FIRING PROFILE

AXIS	TAPED INPUT	FIXTURE RESPONSE
VERTICAL	1.9 GRMS	1.8 GRMS
LATERAL	2.8 GRMS	2.7 GRMS
LONGITUDINAL	1.5 GRMS	1.2 GRMS

The electrical operation of the two radio test sets was checked at the beginning and ending of each test profile. This functional test consisted of measuring the transmitter carrier output, the transmitter carrier modulation and the input voltage.

TEST RESULTS

To date, a total of 775 hours of laboratory induced vibration has been imposed on two radio test sets. The receiver section of the first radio set malfunctioned after 42 hours of testing. An internal examination of the test specimen showed that the first IF stage on the main receiver card had detached. It was necessary to realign the RF section to conform to standard TM11-5021-260-35. The radio set was repaired by USAAVNTBD and functioned for the next 733 hours with no malfunctions. The second radio set was also subjected to 775 hours of laboratory testing with no resulting discrepancies.

The majority of reliability criteria assumes the exponential failure rate specified in MIL-STD-781B. This is expressed by the equation:

$$f(t) = \frac{1}{\theta} e^{\frac{-t}{\theta}}$$

where

$f(t)$ = Failure density

θ = Mean life

t = Time period of interest

MTBF is defined as the total test time divided by the total number of failures. Thus, the MTBF developed thus far in the laboratory is computed to be:

$$\theta = \frac{750 + 750}{1} = 1500$$

Since the MTBF determined experimentally is essentially only an estimate of the true MTBF, statistical confidence levels must be placed on it. The MTBF of an exponential failure function is distributed as a Chi-Square with 2N degrees of freedom where N is the number of failures. There exists standard reliability tables of confidence levels wherein these data may be inserted for the purpose of determining the confidence bounds. Naturally the greater the number of test hours imposed on the specimen, the more these confidence bounds may be narrowed, particularly if more failures are realized.

At present, the test system has been redesigned for the purpose of accommodating four additional radio sets. These six radio sets shall be subjected to 750 hours of testing. Thus, at the conclusion of the program, a representative, 6000 hours of time, will exist for MTBF determination.

SUMMARY

To more closely reproduce in a ground-based laboratory, the actual, dynamic flight environment, a multi-degree of freedom test system should be used. The data resulting from such a test system may be reliably utilized for the determination of the test item MTBF.

The program described has proven the feasibility of such a testing approach by using commercially, available, testing equipment. The only item requiring special design was the drive and restraint device mentioned previously.

Though present indications of the test data obtained in the laboratory seem realistic, final confirmation of the test system validity will rest upon a favorable comparison between the laboratory generated MTBF and the MTBF generated in the field. Should this comparison be favorable, it will be possible to perform future reliability tests of this type at a greatly reduced cost and without the hazards and schedule difficulties associated with actual flight tests.

CHARACTERISTIC COEFFICIENTS, PROBABILITY AND CLASSIFICATION OF WIND PROFILES (Surface to 25 Km)

Oskar M. Essenwanger
Aerophysics Function
Physical Sciences Directorate
Directorate for Research, Development, Engineering
and Missile Systems Laboratory
U.S. Army Missile Command
Redstone Arsenal, Alabama

ABSTRACT. An attempt is made to derive a limited number of realistic wind profiles on a global basis under consideration of probability thresholds for analysis of the environmental influence upon missiles.

In the past, synthetic wind profiles were available for certain selected probability levels. They do not reflect the true altitude relationship of the wind profile and association with probability is selective and ambiguous. Individual wind profiles or correlation matrices overcome problems of unrealistic description but input into computer programs is generally bulky and computer runs may be very costly and time consuming. Random selection of individual wind profiles limited in number introduces bias by persistence and/or gaps in present meteorological upper air data collection.

A new technique developed by the author overcomes the above cited shortcomings. The individual wind profile is approximated by a number of characteristic coefficients, which can be reduced to one variable. The global condition can now be described by three equations and a set of constants. This serves as the basis to classify wind profiles into typical families.

Five major groups whose subtypes depend on the desired accuracy and the particular missile problem to be solved have been found sufficient. In our case 14 subgroups have been established with a total of 49 wind profile models. For many first survey goals it is only necessary to analyze the effect of 5 - 6 wind profiles.

The seasonal and geographic variation of the groups is discussed.

INTRODUCTION. The establishment of appropriate wind profiles for design and trajectory analysis has always been an intricate problem in missile climatology. It is difficult to derive proper wind profiles with inclusion of the true vertical relationship and assessment of the proper probability thresholds, and not to end up with bulky inputs of data for computations in the analysis of the environmental influence upon missiles.

In the past several solutions have been recommended. As a trivial approach, individual wind profiles have been employed. Let us assume key stations reflecting global representations can be selected; this method has solved the problem of true vertical relationship. Remaining is the voluminous input into

computer programs which can cause storage problems, and the analysis may be very costly and suffer from inhomogeneity and data bias, as upper air observations decrease considerably in number with altitude, and data gaps exist. If only a limited number of profiles were selected (even by Monte Carlo methods) the results would lack completeness and would not be conclusive.

Data input is curtailed to some degree by the establishment of intra and interlevel correlation matrices (i.e. Court, 1957). Their calculation, however, is costly and complex, and probability thresholds cannot readily be derived. Virtually little is gained over the previous method to simplify design analysis.

The third technique is based upon limited number of mostly synthetic wind profiles (i.e. Sissenwine, 1954; Handbook of Geophysics and Space Environment, etc.). This cuts down on analysis cost, but in many of these synthetic profiles the true vertical relationship is neglected. Further, profiles must be established for any number of thresholds, association with wind shear, gust, and turbulence is difficult, and a true probability is hard to assess. Modern tools of statistical analysis such as a Monte Carlo method could not be applied.

A new approach has been suggested by the author. The individual wind profile is mathematically described by a number of coefficients. Their interrelationship was studied and it proved that interdependence permits reduction to one characteristic coefficient. The global wind conditions can now be described by three equations and a set of constants varying by month. One of the equations, the frequency distribution of the characteristic coefficient, constitutes the link between wind profile and probability of occurrence. With it any selected threshold can be studied without modification of the input which is minimized to a fraction of less than one thousandth of the original data.

The new technique serves as the basis for classification of the wind profile into essentially five typical families, whose submodels depend on the accuracy and the individual missile problem to be solved. A quick survey of the wind influence upon missile systems would necessitate a maximum of 14 profile models for detailed evaluation, but six extreme profile models would suffice as a first approach. Since probability of the models on a global basis is known, the proper assessment of the wind influence upon the missile system under study can be made.

The technique may have potential for other design and analysis problems, and details may be presented in the following that other applications can be made.

THE DEVELOPMENT OF WIND MODELS

a. Mathematical Representation of the Wind Profile

In previous reports (Essenwanger, 1964 and 1970, Essenwanger and Boyd, 1970) the mathematical representation of the wind profile has been treated. It was pointed out that the wind speed profile from surface to about 25 km can be best described by terms of a Fourier Series, while the wind direction profile is more readily approximated by (orthogonal) polynomials. Although other forms, i.e. expressing the wind profile by mathematical terms for the zonal and meridional components, could be employed (see Essenwanger, 1964), the wind speed and direction system was adopted in the following as best suited for input into missile design.

The interrelationship between coefficients was studied by Essenwanger and Boyd (1970), and the conclusion was drawn that only one characteristic coefficient is necessary to represent the wind speed profile V_h with sufficient accuracy. This provides the following equation for a wind profile system from surface through 25 km

$$V_h - \bar{V}_h = (A_o - \bar{A}_o)[1 + k_1 \sin(\alpha_h + \bar{\alpha}_1) + k_2 \sin(2\alpha_h + \bar{\alpha}_2) + k_3 \sin(3\alpha_h + \bar{\alpha}_3)] \quad (1)$$

where h denotes the altitude dependency, H the top altitude and

$$\alpha_h = \frac{2\pi h}{H} = \frac{360h}{H}, \quad h = 0, 1, \ldots (H-1) \quad (1a)$$

with $$\bar{V}_h = \bar{A}_o [1 + k_1 \sin(\alpha_h + \bar{\alpha}_1) + \ldots]. \quad (1b)$$

The $\bar{A}_o$, k_1, k_2, k_3, $\bar{\alpha}_1$, $\bar{\alpha}_2$, $\bar{\alpha}_3$ stand for constants varying with season and geographic location for sets of individual wind profiles.

The α_h, although different at individual altitude levels, assumes the same value for every profile and for an individual time interval or particular location, and the A_o is the only variable parameter left. Therefore the frequency distribution of A_o takes on the role of a probability scale for wind profiles, and any desired threshold of exceedance of design criteria can thus be established from the frequency distribution of A_o. As a suitable form the Weibull distribution has been chosen, although other curves may fit, too. This provides the equation for the cumulative distribution.

$$F(x) = 1 - e^{-\left(\frac{x-\gamma}{\theta}\right)^{\beta}} \quad (2)$$

where the x is identical with the A_o and the γ, θ, δ are parameters of the distribution. Estimation of the parameters can be made by various methods (see Essenwanger, 1968 and 1971).

The profile for any desired threshold can be obtained by calculating A_o for this threshold. This is equivalent to determining

$$x_{th} = \gamma + \theta[\ell n\, 1/(1 - F_x)]^{1/\delta} \tag{2a}$$

and then computing the profile by equation (1) with this threshold value. The F(x) denotes the probability of the specified threshold.

The third equation relates the wind direction θ to altitude and A_o. Since the correlation between A_o and the wind direction profile is almost non existent, a mean wind direction profile can be associated with the above system. This leads to

$$\bar{\theta}_h = \bar{\theta}_w + c_1\Phi_{1h} + c_2\Phi_{2h} + \cdots c_5\Phi_{5h} \tag{3}$$

where $\Phi_1 \ldots \Phi_5$ are orthogonal (Tchebycheff) polynomials (see Essenwanger, 1964 and 1970), the $\bar{\theta}_h$ is the mean wind direction profile by altitude and $\bar{\theta}_w$, c_1, $c_2 \ldots c_5$ vary with season and location commensurate with the constants of equation (1). It should be noticed that the parameter θ of equation (2) and θ_w of equation (3) are not identical.

The three equations satisfy the postulated conditions: a realistic description of the true wind profile under inclusion of the vertical relationship, an association with probability of occurrence ans a limitation in the numbers of input quantities.

The method has been successfully applied to evaluation of the wind influence in the development of new missile systems.

b. Concepts for Deriving Sets of Wind Models

The mathematical description serves as the basis for the development of a global set of wind models which can be utilized in analysis and tactical operation of missile systems. The extreme models of the set would be available for quick analysis and simple studies of design and evaluation of wind environment upon missile concepts.

The derivation of a set of wind models is generally a subjective task depending on accuracy requirement, practical application and limitations set by the configuration of a missile system. Thus the number of models will vary with the posted goals. The technique of derivation is general, however, and will subsequently be discussed. These techniques may have potential application beyond the narrow field of wind analysis.

It has been demonstrated that equation (1) contains constants varying with season and location and thus a global combination with one frequency distribution and a global set of constants is not advisable, although formalistic computation of constants for this global set would be possible. This can also be interpreted that the complexity of environmental wind conditions cannot be reduced to one simple system. Thus the trivial solution of wind models by class intervals of A_o on a global basis is not very promising.

It would be possible to derive a set of models for every month at specified representative locations. These models could be combined into a global set. Duplicate models could be eliminated, thus reducing the number of models. The variety of conditions and the existence of models with very low probability of occurrence make this method little attractive.

Analysis of equation (1) leads to the conclusion that the wind speed profile is a mixture of waves, in which the constants k_1, k_2 and k_3 are the key to the mixture or the weight of the individual waves. This can be quickly illustrated as for the system

$$2\sigma^2 = \sum_1^3 k_i^2 \tag{4}$$

with the percentage reduction

$$\varepsilon_j^2 = k_j^2 / \sum_1^3 k_i^2 \tag{4a}$$

This scheme is free of A_o and depends on the constants k_i only. It may therefore be possible to build a set of models from class divisions of wave dominance or mixture. A similar technique has been utilized for sound speed profiles where the vertical structure of the profile was the deciding factor for appearance of acoustic focusing (see Essenwanger, 1966).

In an earlier investigation (Essenwanger, 1970b) the mixture of waves was based on monthly mean conditions, and model sets were derived from there. Since some individual profiles were poorly classified, however, a new scheme was tried with individual profiles.

First, a schematic objective classification was set up with criteria like 100% of a single wave ($k_1 = 1.0$, $k_2 = k_3 = 0$), a double wave, triple wave, mixture of waves, etc., and a first frequency count of the individual profiles falling into these classes followed. As expected, many of the classes remained empty, although a total of four representative stations with about 18,000 profiles was employed. Then the number of conditions was reduced until finally only eight typical groups (families) remained. It was recognized from analysis of the individual profile fitting that the wind models had to be based on a slightly modified form of (1) by subtracting (1b) and introducing a factor B

$$V_h = A_o \{1 + B[k_1 \sin (\alpha_h + \alpha_1) + k_2 \sin (2\alpha_h + \alpha_2) + k_3 \sin (3\alpha_h + \alpha_3)]\} \qquad (5)$$

The B serves as an adjustment between A_o and k_1. Although the percentage reduction is not changed, and with it the mixture of waves (harmonic) remains the same, this permits a shift of the mean value A_o (profile reference). In the prior scheme (Essenwanger, 1970b) based on the monthly mean, the B assumes the value 1.0. Thus some individual profiles displayed systematic deviations from the best fitting prototype. This deficiency needed correction.

The factor $B = A_1/A_o$ was computed for all individual profiles and a frequency distribution established. (The A_1 is the coefficient of the first harmonic, see Essenwanger and Boyd, 1970). Although B varied from 0.4 to 1.4, the mode was 1.0. Significant deviations from the regular models appeared for $B \geq 1.3$ and $B \leq 0.6$. There were only a few cases out of the 18,000 profiles with $B \geq 1.3$. They could readily be accommodated into other prototypes of models. Thus the condition was cancelled and only $B \leq 0.6$ was kept. It proved that only two families of models needed this modification.

The phase angle was added from the frequency distribution within the model groups. Although numerous combinations are possible, only a few significant prototypes emerged. Elimination of non-extreme rare profile types close to other models resulted in the scheme of Table 1 with 48 individual wind speed profile models. A model $A_o = 0$ as number 49 was added. Since the closest models next to it employed an amplitude A_o of 6 m/sec, only 1% of the individual profiles occupied class 49. This would be different, if other spacing between models were chosen than the one illustrated in Table 1, i.e., the first amplitude would be 8 m/sec. Thus 49 rounds out the set of models.

Table 1 displays the final arrangement of models. Only five main groups (families) were left with varying subtypes. Within a subtype several models have been derived by classes of A_o. Although in the beginning more subgroups with varying α_1 had been established, occupancy of the model classes by observed models, proximity to other models and consistency within families reduced the

Table 1

Configuration of Models based on Equation (5)

Model Group	B	α_1	α_2	α_3	k_1	k_2	k_3	A_o = 6	12.5	19.5	27	35	40 m/sec
								Model Number					
11	1.0	240	-	-	1.0	-	-	1	2	3	4		
12	1.0	300	-	-	1.0	-	-	-	-	5	6	7	
10	0.6	300	-	-	1.0	-	-	8	9	10	11	12	13
21	1.0	-	200	-	-	1.0	-	14	15	16			
22	1.0	-	300	-	-	1.0	-	17	18				
31	1.0	300	160	-	.7	.7	-	-	19	20	21		
32	1.0	340	210	-	.7	.7	-	-	22	23	24		
41	1.0	330	210	240	.9	.3	.3	-	25	26	27		
42	1.0	300	90	240	.9	.3	.3	-	28	29	30	31	
43	1.0	270	90	240	.9	.3	.3	-	32	33	34	35	
45	1.0	0	240	240	.75	.35	.35	36	37	38			
51	1.0	240	210	230	.5	.5	.5	-	39	40	41		
50	0.6	160	210	230	.5	.5	.5	-	42	43	44	45	
52	1.0	160	150	230	.5	.5	.5	-	46	47	48		
	calm, A_o = 0 model 49												

number to the 48 types shown in Table 1. The models with B ≠ 1.0 were given the code index zero in the second digit. Others were then numbered as displayed in Table 1. More details on the goup of wind models will be discussed in the next section.

c. The Families of Wind Models

Table 1 summarizes the respective constants utilized in the computation of the wind speed models. It should be noticed that only five main conditions remained from a variety of initially postulated profile mixtures. This does not mean that other groups would not exist. It was decided, however, that all groups which placed less than five observed profiles out of the 18,000 studied were considered negligible.

Within the group several subgroups were derived by alteration of the α (phase angle). Within the subgroups models were established with the spacing of the A_o as given in the headings of Table 1. Originally equal spacing of A_o was considered. Studies of the proximity of models, frequency of occurrence and accuracy of the distributed observed profiles led to the present setting. Equal spacing of A_o would have resulted in at least 10 to 12 more models with virtually no gain in accuracy.

The first group contains profiles with a single maximum, while the second group exhibits a double maximum profile. The third group is a mixture between the first and second harmonic. Following is a group with a main minimum, as the single wave dominates in this mixture. The last group is an equally balanced combination of the first three harmonics.

Although the Σk_i^2 for this last group would result in a sum of 0.75, while all other groups add up to 1.0, this deviation is intentional. It can be readily seen by equation (4a) that the change of the sum does not influence the percentage reduction. The modification determines spacing of the models within the group. A study of the closeness of fit within the individual groups revealed that the largest deviations between observed and analytic profile were found in this last group. The selection of k_i as given by Table 1 reduced the spacing of the models and made it commensurable with the distance in the other groups.

Within a group initially all models as obtained from the spacing of A_o were computed and run through a comparison program. The mean squared difference ϵ^2 between models*) was computed, and all models with $\epsilon^2 < 16$ were eliminated. Although these wind speed profile models would certainly exist, proximity to other prototypes make them surplus. When there was a choice between models, the one with the lower subgroup number was kept. Thus various models with amplitude $A_o = 6.0$ were omitted.

*) See Equation (6).

Figure 1 illustrates the model types in groups 11 and 12. They differ by the altitude of the maximum. In group 11 the maximum appears at 14 - 16 km, a property found in the tropics (see Table 3), while group 12 shows a maximum around 10 - 12 km, expected in subtropical and midlatitude regions.

Figure 2 serves for comparison of the groups 10, 11, and 12. The series of models 8 through 13, although similar to group 12, includes a displacement of the profile reference A_o, which makes some of the observed profiles better than the series 5 through 7.

Figure 3 illustrates the profile types with double and triple waves. The extreme profile models were chosen from each subgroup to demonstrate the individual features better than models with smaller amplitude A_o. Again, the difference between subgroups 21 and 22 is the altitude of the maximum and the minimum wind speed. Profile subgroup 21 displays a shift towards higher altitude.

The difference in phase angle for subgroups 50, 51, and 52 causes the side-maxima in the troposphere and lower stratosphere to vary in strength compared with one another, while the main maximum speed stays above 20 km.

Figure 4 finally exhibits a dominant single wave (groups 3 and 4). The major difference from the single wave pattern subgroups 10 through 12 is the narrowing of the maximum. Some shift in altitude for the maximum of the individual subgroups is visible, too, as should be expected from the phase angles given in Table 1.

The geographic and seasonal distribution follows in section 2d. Before detailed consideration of time and space variability a short look at the closeness of fit may be appropriate.

Since the spacing of the models is an arbitrary decision of selecting a few from the numerous possible sequence within one family, some objective criteria should be derived for indication of the quality of the model sets. This can be accomplished by computing

$$\varepsilon^2 = [\frac{1}{n} \Sigma (y_{hj} - y_{hk})^2]^{1/2} \qquad (6)$$

where the h denotes the altitude and j and k two different models, one of these can be the observed profile. The mean squared difference ε^2 can also be employed to place the observed profile into one model type. We select the closest to the observed speed profile, i.e., the minimum ε^2. This is the same procedure by which the observations were associated with the former set of models (Essenwanger, 1970b). Comparison with these 43 models derived from monthly average conditions shows that the present set of models fits the individual

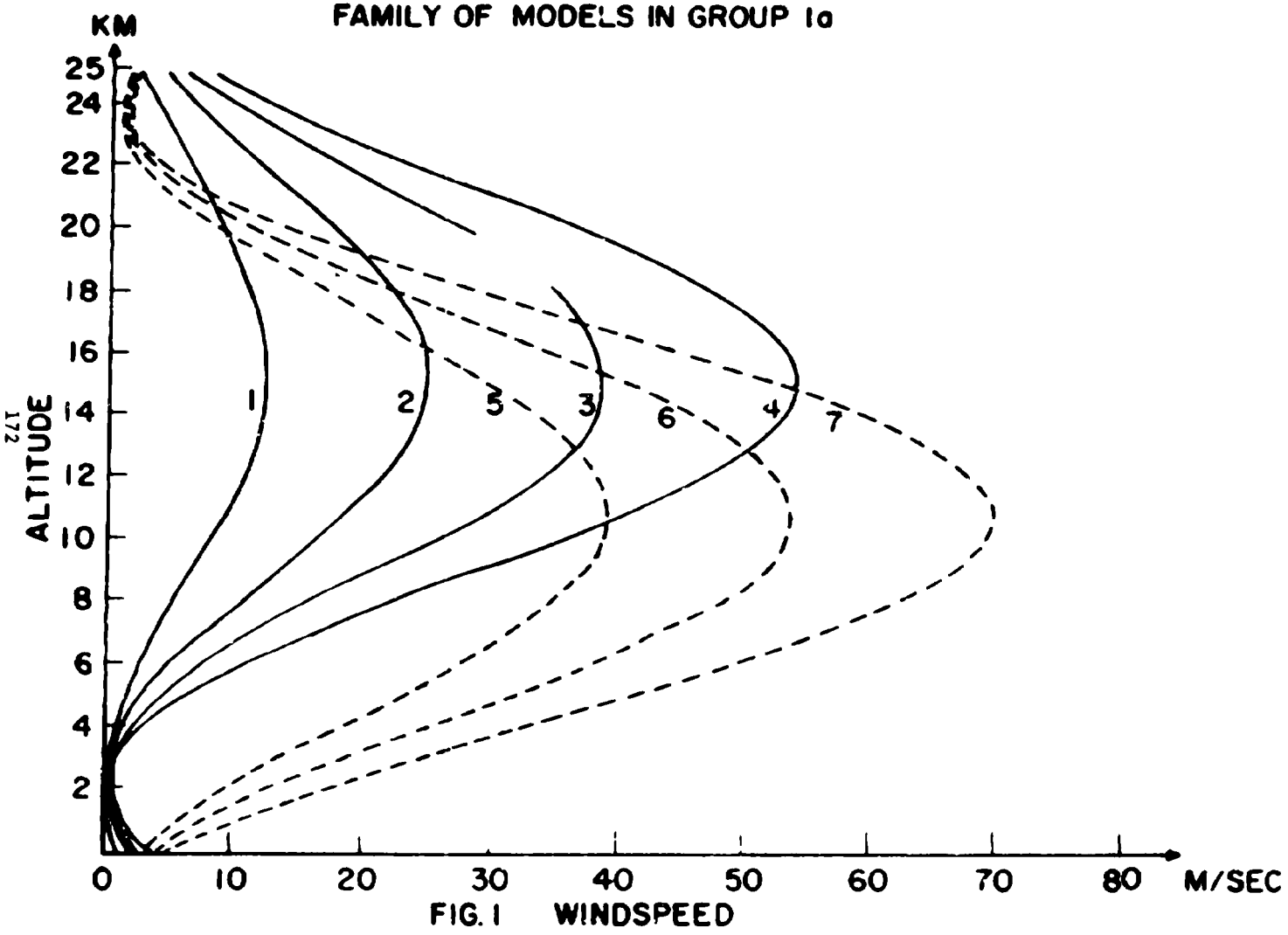


FIG. 1 WINDSPEED

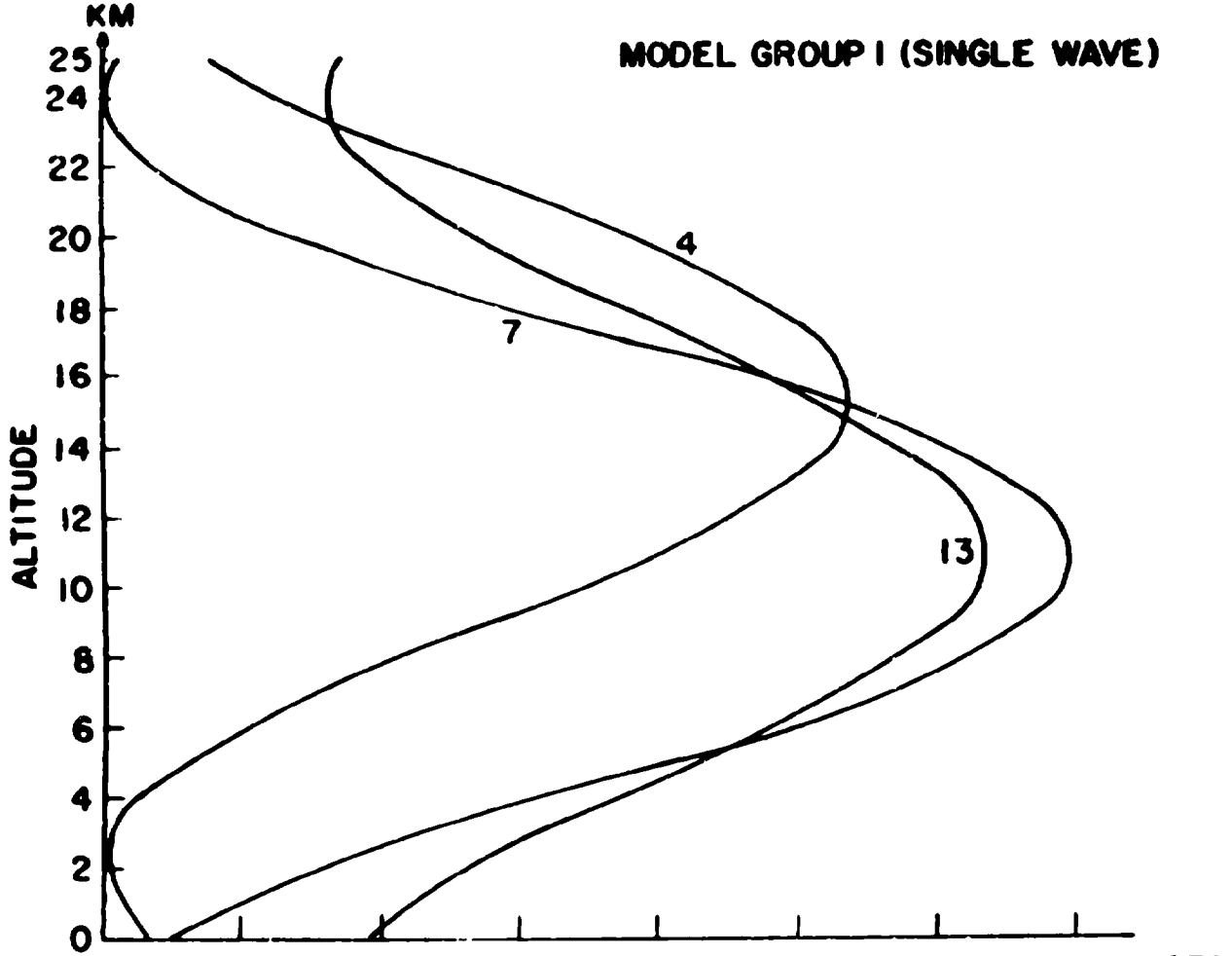

MODEL GROUP I (SINGLE WAVE)
KM
ALTITUDE
25
24
22
20
18
16
14
12
10
8
6
4
2
0
4
7
13
10
20
30
40
50
60
70 M/SEC
WINDSPEED

FIG. 2

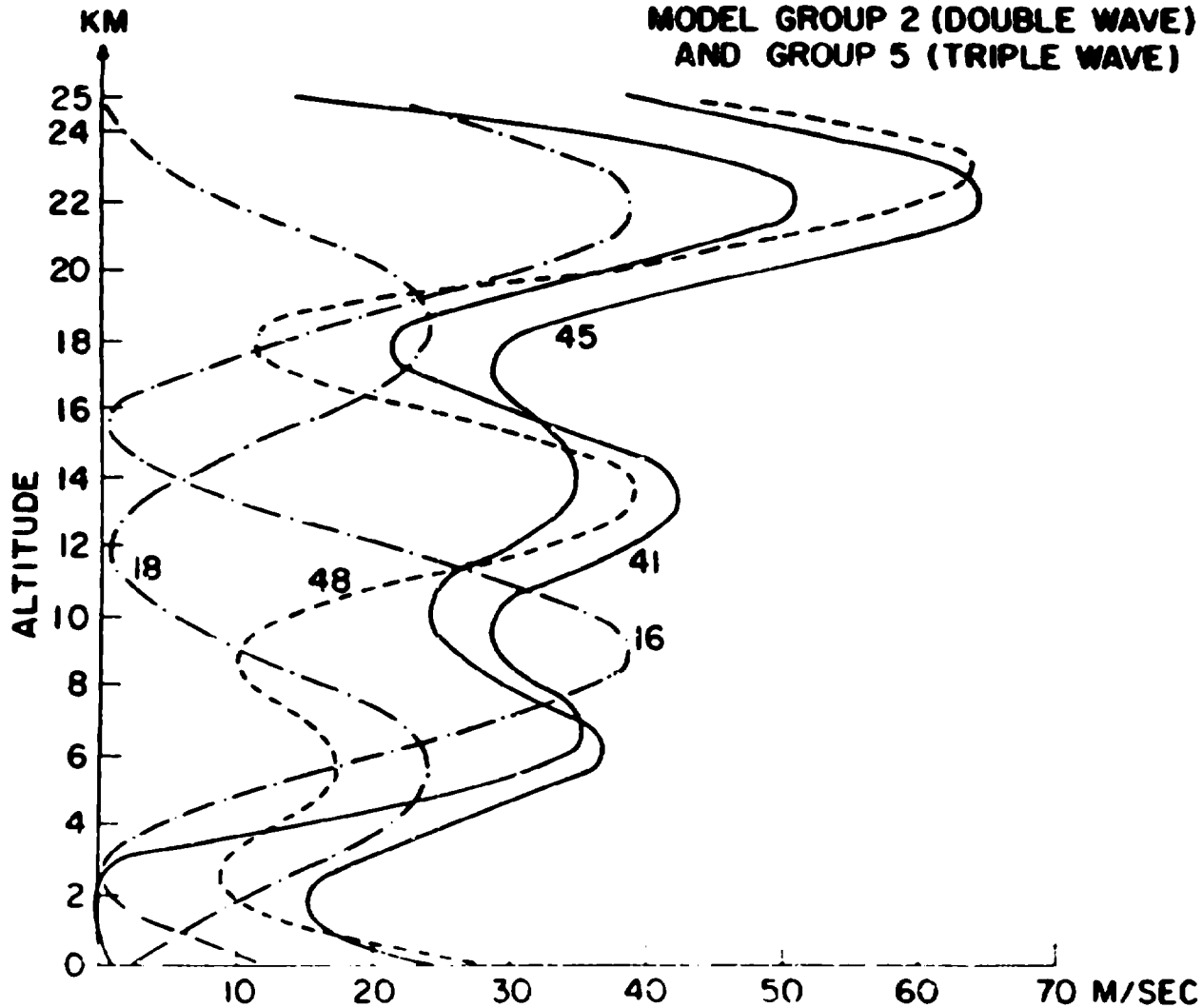


FIG 3 WINDSPEED

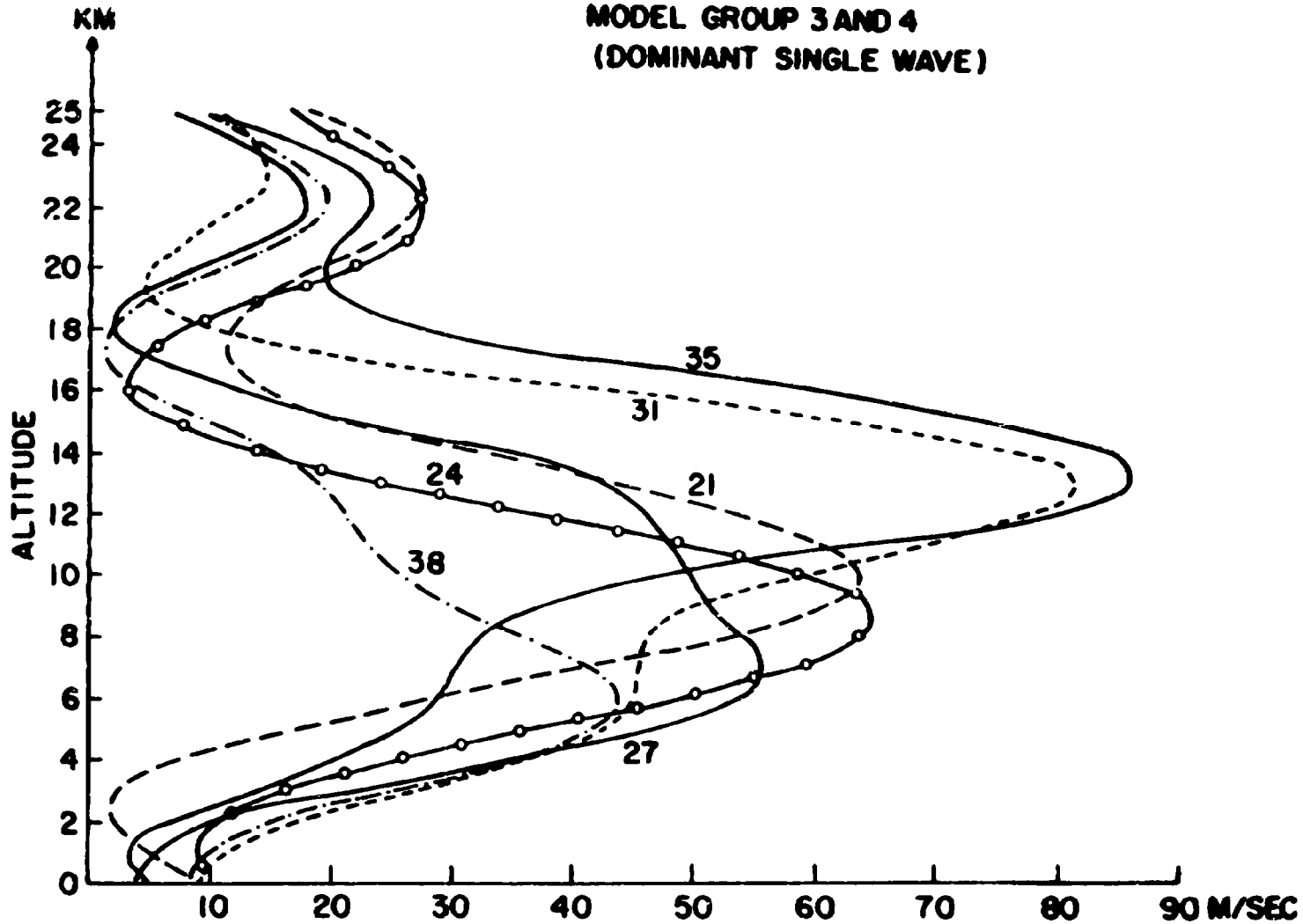


FIG. 4 WINDSPEED

profiles closer, although only six more models were necessary. Table 2 discerns the distribution of the ε^2 by station and season. The highest dispersion takes place in the subtropical zone (Montgomery) and in winter. This result parallels the conclusion from the 43 model set. The minimum ε^2 per station and season emerges for Albrook and the summer.

d. Geographic and Seasonal Variation of Models

The discussion on sets of wind models would be incomplete without the analysis of the time and space variability of the models. A summary for the four stations is given in Table 3, the seasonal variation by station in Table 4. Since the number of observations varied by station and season, (see Table 4) a conversion to 0.1% was made and the station summary and global combination was based on the homogenized number. This would not eliminate the bias introduced by data gaps within the seasons or the period of record utilized, but balances the disproportionate weight of summer data within the year and the four daily observations at Thule compared with two daily recordings at the other stations in the global summary.

The last column in Table 3 summarizes the global conditions. No subgroup contains less than 1%. The only frequency less than 1% is the profile model 49, (A_o = 0) which can be considered calm air for practical purposes. Explanation was given in the preceeding section that the number of profiles placed into this model type depends on the spacing of the next A_o. The type is necessary as a boundary to complete the series of models.

The most frequent subgroup is 10, a profile with a single maximum at 10 - 12 km. As expected, this is a midlatitude type and comprises about 50% of the wind profiles at Chateauroux. It is also the type with the maximum frequency at Thule and Montgomery and the most frequent one in all seasons.

The next subgroup in number of placed profiles is 50, a mixture of all three harmonics with the major maximum above 20 km (Figure 3). It should be noticed that the decrease in wind speed towards 25 km indicated in Figure 3 is a result of the mathematical representation of the profile by three terms only and the selection of the altitude range from surface through 25 km. In most empirical profiles this decrease of the wind speed around 25 km cannot be observed and the maximum actually lies at a higher level than 25 km. The limitation in the mathematical description produces the maximum around 23 km. The importance is, however, that this subgroup displays an increase of the wind speed in the altitude range between 20 to 25 km exceeding the maximum speed at the tropospheric level (jet stream). This type of wind profile is typical for the polar and tropical region as can be noticed from the almost 25% frequency at Albrook and Thule. Table 4, giving the seasonal breakdown by station, exhibits a dominance of 53% at Thule in winter, while at Albrook this subgroup persists in summer with 36%. It explains the high occurrence for the year at both stations. The peak season at Thule and Albrook is opposite, however.

Table 2

Average Squared Deviation Between Wind Model and Observed Profile

(in 0.1%)

ϵ^2	Al	Mo	Cha	Thu	Wi	Sp	Su	Fa	Total
≤ 1			1	1	1	1	1	1	1
2 - 4	18	10	35	46	10	3[illegible]	35	30	[illegible]7
5 - 9	140	90	164	[illegible]05	66	157	[illegible]00	177	150
10 - 16	309	244	[illegible]75	[illegible]80	[illegible]0	[illegible]97	312	[illegible]97	[illegible]77
17 - 25	363	310	267	[illegible]21	[illegible]82	[illegible]85	[illegible]95	[illegible]79	[illegible]85
26 - 36	162	184	130	118	19[illegible]	136	121	144	148
37 - 49	[illegible]6	86	74	57	110	56	27	50	61
50 - 64	2	36	3[illegible]	29	59	19	7	14	[illegible]5
65 - 81		24	12	[illegible]0	40	10	1	5	14
82 - 100		9	5	8	16	3	1	2	5
> 100		7	5	15	22	4		1	7
Average	4.30	5.01	4.62	4.63	5.47	4.49	4.15	4.37	4.65

Al = Albrook (Canal Zone)
Mo = Montgomery (Alabama)
Cha = Chateauroux (France)
Thu = Thule (Greenland)
Wi = December - February
Sp = March - May
Su = June - August
Fa = September - November

Table 3

FREQUENCY OF MODEL GROUPS (in 0.1%)

Model Group	Alb	Montg	Chat	Thule	Wi	Sp	Su	Fa	Total
10	221	296	488	263	246	331	320	371	317
11	148	61	13	25	58	67	59	62	62
12	-	164	43	-	85	67	21	34	52
21	135	94	36	46	48	71	120	73	78
22	57	3	2	11	5	19	37	12	18
31	3	26	64	12	24	21	32	29	26
32	-	0	36	59	22	26	24	23	24
41	-	13	94	37	26	52	33	34	36
42	3	121	51	-	18	56	40	61	44
43	35	146	8	-	88	41	16	44	47
45	17	20	61	254	47	117	114	73	88
50	225	38	71	224	238	83	124	112	139
51	23	16	26	27	49	11	12	20	23
52	132	1	4	18	46	30	33	48	39
Calm	1	1	3	24	0	8	15	4	7
Weak	558	269	302	498	193	440	571	424	407

Table 4

SEASONAL VARIATION OF WIND PROFILE GROUPS FREQUENCY (in 0.1%)

Model	Albrook				Montgomery				Chateauroux				Thule			
Group	Wi	Sp	Su	Fa	Wi	Sp	Su	Fa	Wi	Sp	Su	Fa	Wi	Sp	Su	Fa
10	223	243	151	266	347	299	187	370	346	530	504	571	90	250	439	476
11	170	162	122	137	21	51	105	66	17	16	6	13	35	37	3	34
12	-	-	-	-	330	229	3	94	10	38	82	43	-	-	-	-
21	94	216	83	148	0	2	336	37	50	21	36	35	47	44	23	74
22	7	50	138	35	0	0	6	4	0	1	1	3	13	21	2	6
31	9	2	0	0	13	14	54	26	67	60	66	64	7	7	9	25
32	-	-	-	-	0	1	0	0	47	58	15	25	47	46	80	67
41	-	-	-	-	9	18	6	20	50	141	100	87	47	48	27	47
42	1	0	0	11	62	210	29	184	10	16	129	48	-	-	-	-
43	139	1	0	1	209	162	52	160	5	2	1	14	-	-	-	-
45	13	34	8	12	2	10	43	24	67	78	43	55	105	346	361	203
50	178	166	364	194	8	2	133	8	39	20	1	25	528	146	0	221
51	59	17	9	9	20	0	41	4	80	12	0	11	38	16	0	54
52	107	109	125	185	0	0	5	1	12	4	0	0	58	8	0	6
Calm	0	0	0	2	0	0	0	2	0	3	4	6	0	30	56	9
Weak	163	686	495	590	11	158	613	293	175	347	330	356	122	568	845	457
N	973	951	870	854	655	840	1085	1018	871	1031	983	916	599	1951	2535	1774

Other subgroups show less than 10% in the total summary. Noticeable is the 25% of subgroup 45 at Thule, which is a profile type with predominance of a single wave but a narrow altitude region for the maximum. The altitude of the maximum is rather low (Fig. 4, profile 38). The subgroup has its peak frequency in spring and summer.

Since the first amplitude is 6 m/sec only, profile types in all subgroups comprise situations when the wind speed is not very strong. They have been combined under "weak" wind profiles irrespective of the subgroups and the model 49 has been included. It is noteworthy that Albrook and Thule emerge as stations where about 50% of the wind profiles are placed into these weak wind speed profile groups. As expected, the subtropical and midlatitude regions have by far stronger winds. The seasonal change should also be pointed out, which varies between about 60% in summer and 20% in winter. Although this result is a well known fact, the confirmation in objective numbers is satisfying, and it demonstrates that the system of wind profile models is reasonably rational. The season of the peak for the number of weak wind profiles at Albrook, the spring, may be a surprise. Somewhat unexpected may also be the low 1% in winter at Montgomery. The high 85% in summer in Thule is striking. Otherwise the seasonal and station distribution is as anticipated.

Table 5 summarizes the distribution by combining the subgroups 10 - 12 and 40 - 45 into one group of single maximum profiles and lists each of the other three main groups. Since the "calm" model has been omitted, the percentage figures do not add up to 100% for all seasons or situations.

Table 5 reveals that the single maximum profile type is dominant in the midlatitudes and subtropics. Montgomery shows more than 90% of this group in all but the summer season, while only 50% of this group is left for the winter season in Chateauroux. This agrees with the expected seasonal shift of the jetstream.

The tropical region (Albrook) appears with a split between three major groups. The peak frequency shifts from a single wave structure in winter with a maximum around the tropopause to the group with a maximum in the stratosphere in summer. The reverse change takes place in the polar region (Thule).

The double wave pattern (21 - 22) seems important only in the tropical zone during all seasons and in the subtropical regime during summer. The group 31 - 32, a mixture between single and double wave, appears significant in the midlatitude and polar region.

Table 5

Summary of Profile Groups (in °/o)

a. Seasonal Variation

Model Group	Albrook				Montgomery				Chateauroux				Thule			
	Wi	Sp	Su	Fa	Wi	Sp	Su	Fa	Wi	Sp	Su	Fa	Wi	Sp	Su	Fa
Single	55	44	28	43	96	98	43	92	51	82	88	83	27	68	83	54
21 - 22	10	27	22	18	0	1	34	4	5	2	4	4	6	7	2	8
31 - 32	1	0	0	0	1	1	5	3	11	12	8	9	5	5	9	9
50 - 52	34	29	50	39	3	0	18	1	33	4	0	3	62	17	0	28

b. Station and Season Survey

Model Group	Stations				Seasons				Total
	Al	Mo	Ch	Th	Wi	Sp	Su	Fa	
Single	42	82	76	58	57	73	60	68	64
21 - 22	19	10	4	6	5	9	16	9	10
31 - 32	1	3	10	7	5	5	6	5	5
50 - 52	36	5	10	27	33	13	17	18	20

The predominance of one group at most of the stations during the seasons makes it plausible that the mathematical representation of the wind speed profile and the subsequent reduction to one coefficient (Essenwanger and Boyd, 1970) succeeds for monthly summaries. It can be recognized, however, that a global combination of the frequencies of one parameter will not render a good representation of the structure of the wind profile, as the average constants for a global combination would not express the variety of the profile groups.

SUMMARY AND CONCLUSIONS. The mathematical description of the wind profile by one variable characteristic coefficient and monthly sets of constants by location (Essenwanger and Boyd, 1970) have served as the foundation for the derivation of a global set of wind speed profile models. Groups of models are established by utilizing the percentage reduction (equation 4 and 4a) of three coefficients. Although numerous class groups could have been established, about 18,000 profiles from four representative stations of four climatic zones were divided into essentially five major typical families of profiles with a total of 14 subgroups. The subgroups and the subsequent individual prototypes of profiles depend on accuracy and individual goals for missile systems.

The presented groups divide into 49 wind profile models, whose 14 extreme profiles of the individual subgroups permit quick calculation of configuration for missile systems and evaluation of problem areas. The 14 extreme profiles could be further reduced to essentially 5 - 6 wind speed profiles, representing the major groups. Since the global probability of occurrence for the individual groups is known (Tables 3, 4, 5), evaluation of the wind influence upon missile systems or in trajectory analysis, etc., should be possible with relatively little cost.

The developed set of 49 profiles can also be employed in tactical operations or other systems of operational purpose. The set of 49 models fits the individual wind speed profiles better than an earlier set of 43 models derived from mean monthly conditions. The set of 49 models serves further to investigate the association with other wind parameters such as wind shear, turbulence, gust, etc., and association with density and temperature profiles. A further application lies in the study of predictability of the models for tactical purposes.

Although the 49 models presented here are synthetic profiles, their origin from the description of the vertical structure of the wind speed profile makes them a realistic approximation of the true conditions. Therefore system analysis with them should lead to maximum efficiency and accuracy with minimum cost to missile systems, avoiding over - or under design. The problem of voluminous data input in studies of the environmental effect is solved and data bias and inhomogeneity is largely reduced.

The seasonal and geographic variation of the profile groups shows excellent agreement with known interaction in the general circulation. This supports the fact that the wind models are capable of expressing the true structure of the atmosphere. A sophisticated modification of the models would be the correction of some deficiencies of the three term harmonic recomputation by comparison with the empirical profiles placed into the prototype division.

The technique of deriving models from a set of global descriptors or a countless number of observations has general application in statistical analysis and is not restricted to the narrow field of wind analysis.

ACKNOWLEDGMENT. I am indebted to Mrs. Helen Boyd, who aided the investigation by establishing the various IBM 7094 computer programs for the development of the final set of profiles and the seasonal and geographic distribution of the profile types. I wish to express my gratitude to Dr. Dorathy A. Stewart for reviewing the manuscript. Mrs. B. Wright deserves the credit for typing the manuscript in record time.

LIST OF REFERENCES

1. Court, Al, 1957, Vertical Correlations of Wind Components, Scientific Report No. 1, Contract AF 19(604)-2060.

2. Essenwanger, O. M., 1964, Mathematical Characteristics of Individual Wind Profiles, Army Missile Command Report No. RR-TR-64-12.

3. Essenwanger, O. M., 1966, Profile Types of Sound Speed in the Lower Atmosphere and Their Relationship to Acoustic Focusing, Army Missile Command Report No. RR-TR-66-6.

4. Essenwanger, O. M., 1968, On Deriving 90 to 99% Wind and Wind Shear Thresholds from Statistical Parameters, Proceedings of Third National Conference on Aerospace Meterology, New Orleans, P. 145 - 153.

5. Essenwanger, O. M., 1970a, On Wind Profile Representation for Missile Design and Environment Analysis, Proceedings of the Fourth National Conference on Aerospace Meteorology, Las Vegas, Nev., p. 246 - 253.

6. Essenwanger, O. M., 1970b, The Characteristic Coefficients Technique for Probability Models of Wind Profiles in Missile Design and Environment Analysis, Proceedings of Army Science Conference, 16 - 19 June 1970, Vol. I, p. 291 - 305.

7. Essenwanger, O. M., 1971, Elements of Statistical Analysis in Climatology, World Survey of Climatology, Vol. 3, General Climatology, Elsevier Publishing Company, Amsterd., in print (1971).

8. Essenwanger, O. M. and H. Boyd, 1970, On Methods to Establish a Set of Global Characteristic Wind Profiles, Army Missile Command Report No. RR-TR-70-25.

9. Handbook of Geophysics and Space Environment, McGraw Hill Company, New York, 1965.

10. Sissenwine, N., 1954, Wind Speed Profiles, Wind Shear and Gusts for Design of Guidance Systems for Vertical Rising Vehicles, Air Force Surveys in Geophysics, No. 57, AFCRL, 1954.

IDENTIFICATION OF WORKERS IN BIOLOGICALS THROUGH SERUM TITERS BY DISCRIMINANT FUNCTION

Walter D. Foster
Marian D. Jones
Biomathematics Division
Fort Detrick
Frederick, Maryland

ACKNOWLEDGMENT. It would not have been possible to perform the analyses described here without the excellent support of James F. Jacobs, who wrote or adapted the necessary computer programs. His perception of the problems, insight in analysis, and superlative computerizations were indispensable. We should like also to acknowledge the help of Major Thomas J. Keefe, Mr. Joseph A. Mangiafico, Mr. Arthur E. Hougland, and Captain W. K. Taylor in their patient and detailed statement of the problem.

SUMMARY. Researchers who work with infectious biological material such as Pasteurella tularensis generally are immunized lest they become infected as a result of a laboratory mishap. The standard measure of immunity is the worker's serological antibody titer -- the higher titers generally thought to confer a greater degree of immunity. A recent experiment compared 201 workers in biologicals who received immunological procedures designed to protect them in their areas of specialization with 100 Army draftees who received the standard Army immunization series. Serological titers to 19 different biological agents were assayed for each of the persons in the two populations. It was the purpose of this paper to discover if the persons could be correctly classified into the appropriate group according to their serological titers only.

A preliminary screening for agents that showed group differences was performed by using the t-test on group means for each titer separately and by using the chi-square statistic as a test for a shift in frequency. Candidate discriminators were those agents for which the t-test or both the t-test and the chi-square statistic were significant at the .10 level. Agents 1, 2, 4, 6, 7, 8, 15, 17, 19 survived the screening procedure.

A modified step-wise discriminant function procedure was employed to create a battery of discriminant functions that competed for first place in ability to correctly classify persons in the two groups. Because the sum of squares attributable to the discriminant function depends upon which agents were included, the addition of another agent frequently resulted in an improvement in the percent classification in one group but a decrease for the other.

Three other agents 5, 8, and 13 (two of which had not survived the preliminary screening) were observed to have distributions such that all "control" group members (Army draftees) and all but a few in the "treatment" (biological workers) group had negative titers. Three "decision rules" were added to the classification technique: a positive titer for any of the three agents 5, 8, or 13 automatically classified the person as "treatment"; a negative titer for all three called up the discriminant function. A small but distinct improvement in the percent classification was noted with the addition of the decision rules.

A cost function was developed as a criterion for selecting the best of the competing discriminant functions. Costs of misclassifying a true treatment person, misclassifying a true control person, and the cost of assay were included in the function. Because actual costs were difficult to estimate, it was valid to recast the cost function in terms of relative costs -- the ratio of costs. The battery of competing discriminants was evaluated for each of 14 combinations of values allotted to the three cost ratios by a panel of experts.

All discriminants in a battery were ranked for each of the 14 parameter combinations. The discriminant, 1-6-4-15, plus the decision agents 5, 8, 13 was found to be the winner for 99% of the range included in the cost ratio parameters. This discriminant was able to classify correctly 81% for the control group and 98.5% for the treatment group. Discriminant 1-6-4-2-15-17 was also consistently near the top for 99% of the range of the cost ratios; its percentages were 88% and 97.5% respectively. That it was not a winner was due to the emphasis on correct classification for the treatment group.

When the misclassification costs were equal, the ranking of the discriminants was sensitive to the relative cost of the assay. For expensive assays, agent 1 by itself was best; for inexpensive assays, discriminant 1-6-4-2-15-17 plus 5, 8, and 13 was best. The consistently high-ranked discriminant mentioned above, 1-6-4-2-15-17 was also high-ranked for both assay costs.

INTRODUCTION. The extent to which a serological profile might "fingerprint" workers with biological materials was the criterion for using two groups of persons in the Department of Army for extensive serum testing.

Persons known to be working actively with biologicals were selected, 100 at the Army Biological Laboratories at Fort Detrick; and 101 from workers at Dugway Proving Ground, a testing site. A control group of 100 enlisted men receiving the standard army immunization series but known not to be involved with biologicals was selected at Edgewood Arsenal, Maryland. It is the purpose of this part of the investigation to discover if a combination of titers is more effective in identifying workers in biolcgicals than individual titers and to what extent such identification is predictable, using initially the data from these two groups.

Individuals in the Detrick-Dugway "treatment" group were selected on the basis of their active participation with biologicals and the immunization programs provided for the protection of the workers in both locations. Serum titers to 19 microbiological agents were assessed for HI, CF, or SN antibodies for each individual; an extensive questionnaire was requested for each participant designed to provide information that might account for titers not explainable by either immunization or laboratory contact.

PRELIMINARY SCREENING. If the distributions of titers for a particular agent were exactly the same for the two population groups, there would be no basis for using that agent to classify a person into either group. A perfect basis would be an agent for which there was always a titer in the treatment group and never a titer in the control group. Another perfect basis would exist if the titer distributions for the two groups did not overlap. The problem becomes one of finding how much the treatment group distribution of titers for an agent must differ from that for the control group to be an acceptable basis for classification.

A two-stage procedure was devised to answer the question of which agents should be considered as effective classifiers. The first stage has been called a preliminary screening stage and consisted of computing t-tests on the mean positive titers for the two groups for every agent and a chi-square test on the frequency of specific titers for every agent. The results of this preliminary screening can be seen in Table 2. These two tests were used in conjunction with each other because of the definitions of "positive" and "specific" titer. A positive titer was any titer that was not considered negative. When the t-test was performed on the means of positive titers, persons with negative titers were not included in the t-test computation for that agent. However, they were included in the frequency test afforded by the chi-square statistic.

Values of "specific" titers were defined arbitrarily for each agent; it was intended that they should represent a threshold titer not likely to have occurred because of assay variability. Values of both are shown in Table 2.

The probability levels for these two statistical tests were not stringent; as a screening device, the .10 level was used as opposed to the .05 or .01 levels. Those agents showing sufficient distributional differences between the two groups on the basis of both tests or of the t-test alone were included as candidate discriminators. Agents 1, 2, 4, 6, 7, 8, 15, 17, 19 survived the screening procedure. A parenthetical note is added here to note that differences between the Dugway and Detrick workers were also examined before they were pooled to form the "treatment" group. Several differences did occur, but they were generally explainable on the basis of known subject history. There was no valid reason for not putting these two groups together because they did represent workers in microbiology.

The second stage consisted of using a discriminant function -- a combination of individual agents -- as a basis for classification. Before this approach is described, it is worthwhile examining the discriminating ability of the candidate agents individually. Table 3 lists some of the individual agents and indicates the percent correct classification for each group.

The procedure for classification was simple. The mean of the Agent #1 titer (in log scale) for the control group was .9289 and 2.920 for the treatment group. Halfway between the two means is the midmean value 1.925, which was used as the classifying rule: a person with a titer less than the midmean was classified as control; a person with a higher titer was classified as treatment. Since each person was known to be in one group or the other, the percent correct classification for each group was easy to compute for one agent individually.

Some of the agents with extremely high percentages for one group but low for the other deserve a word of comment. Agent 6, for example, gave 98 and 10 percent correct for control and treatment groups respectively. Actually, only 22% of the subjects in the control group and 10% of the subjects in the treatment group had positive titers. Thus 98% of the control group was correctly classified; those in the treatment group with titers less than the midmean were also classified as control and therefore incorrectly. This agent could not be considered very effective according to the midmean rule; nevertheless, it was retained for possible use either in the discriminant function or in the decision rule approach.

DISCRIMINANT FUNCTION APPROACH. A linear discriminant function is a combination of predictors (agents), each with a weight corresponding to its relative effectiveness in discrimination:

$$F = B_1X_1 + B_2X_2 + \ldots + B_kX_k$$

The X_i are the titer values in logarithms for the various agents, 1--k, used in the discriminant; the weights, B_i, are estimated from the data. As applied, a person's titers (in log scale) are substituted as the X values; if the resulting F value is closer to that for the control group than it is for the treatment group, the subject is classified as belonging to control. Theory guarantees that the methodology for estimating the weights results in the best possible linear discriminant using these predictors.

The individual, one-agent-only approach described earlier with the results shown in Table 3 cannot apply to a combination of agents. The presence of a second and more discriminators in conjunction with the first will have definite effects on the behavior of the discriminant function. The method of estimating the weights for the linear discriminant above recognizes these covariable effects and optimizes their joint function.

APPLICATION OF THE DISCRIMINANT FUNCTION. Perhaps the easiest way to look at the results of the discriminant functions as constituted by a variety of predictor combinations is to look at the percent correct classification for each group. Table 4 identifies the various discriminants in terms of the agents used as predictors and shows the percent correct classification for each; they may be regarded as competing for optimum performance. The most obvious result was the clear superiority of all discriminants with agent #1 over those without #1. This finding was true regardless of the number of other agents or which ones. Thus, the next step was to select the best discriminant, having discarded all discriminants without agent #1.

The question of which discriminant to choose was complicated by the fact that some discriminants did better for one group, and others did better for the other group; none was clearly best for both groups. A second complication arose by virtue of the inherent characteristics of discriminant function estimation. For example, as can be seen in Table 4, the addition of agent #4 to discriminant function 1-6 resulted in going from percent correct classification 94.0 and 88.6 for the control and treatment groups respectively to 90.0 and 91.0. Thus the addition of another variable to a discriminant function was not necessarily an improvement in both categories. Further study of Table 4 does not allow an easy identification of the winning discriminant. A discussion of criterion for choosing is deferred until the approach involving decision rules has been added to the discriminant function approach.

DECISION RULES AS AN ADJUNCT DISCRIMINATION DEVICE. An additional device, denoted here as "decision rules," utilized two agents not already incorporated in any of the discriminant functions. This approach is analogous to the frequency basis for the preliminary screening; the discriminant function corresponds to the comparison of means. Two of the agents, #5 and 13, were not deemed to be potentially useful for the discriminant function approach because of the generally low incidence of positive titers in either group. However, these two agents with agent #8 shared one characteristic in common -- all titers for all control subjects were negative (with the exception of 2 persons for agent #8). To take advantage of the information afforded by the distributional form, the following decision rules were set up as an adjunct to the discriminant function approach as an enlarged basis for classifying subjects into the control and treatment groups:

1) If a person has a positive titer to agent #5, classify him as treatment and remove the subject from further classifying rules. If the subject has a negative titer, do not classify but go on to agent #8.

2) If the subject has a positive titer to agent #8, classify as treatment and remove subject from further classifying rules. If he has a negative titer, do not classify but go on to agent #13.

3) If subject has a positive titer to agent #13, classify as treatment and remove subject from further classifying rules. If he has a negative titer, do not classify but apply the discriminant function.

This decision-rule approach is essentially probabilistic -- one that might be developed much further by means of joint and conditional probability statements. This opportunity has not been exploited beyond the rules described above.

The effectiveness of the decision rule approach compared with the discriminant function only can be seen in Table 5. The results were consistent. The control group percentages uniformly decreased by 2%, obviously because of the two control subjects with positive titers to agent #8. However, the percent correct classification for the treatment group increased without exception. If there is greater importance attached to correct classification for the treatment group, then the addition of the decision rules generally was a benefit. The problem of selecting a best combination among the discriminants with or without the decision rules remains to be answered. Again, there has been no obviously established winner after the addition of the decision rules. A criterion for selection is clearly needed.

RELATIVE COST AS A BASIS FOR A CRITERION OF EVALUATION.

A. Development of the Cost Function

A simple rule or criterion would be to consider the percent correct classification for each group to be of equal importance. If no other rule were added, then the winner along the first six functions listed in Table 4 would be function 1-8, whose average percent correct classification for the two groups was 92.0, higher than any other average in the first six. However, the average percent correct classification for function 1-6-4-2 was 92.5, but it required the addition of two more agents. The question arises whether the increase in the average percentage was worth the cost of the additional assays. A general model of cost as a criterion is developed in this section, with application to these questions: (1) was the addition of the decision rules worthwhile and (2) what is the most efficient discriminant when performance is assessed against cost?

In this development of a cost function, it will be convenient to define certain costs without stating or defining their dollar value. We begin by defining error rates as another way of stating successful classification:

Let P_c = proportion correct classification in the control group and

P_t = proportion correct classification in the treatment group.

Then the error rates or rates of misclassification are defined as

$q_c = 1 - p_c$ = proportion of incorrect classification for control and

$q_t = 1 - p_t$ = proportion of incorrect classification for treatment.

Let Q_c = cost of misclassifying a control subject and

Q_t = cost of misclassifying a treatment subject.

In addition to converting error rates into cost of misclassification as shown above, we proceed to include the cost of assay in the cost function.

Let B = cost of one bacterial assay and

V = cost of one viral (or rickettsial) assay by hemagglutination inhibition.

Let b = number of bacterial agents used in discriminator and

v = number of viral (or rickettsial) agents used in discriminator. Therefore,

bB = cost of b bacterial assays and

vV = cost of v viral assays.

Thus Cost = $q_c Q_c + q_t Q_t + vV + bB$.

This statement adds both costs of misclassification to the cost of assay. The reduction of any of the quantities in the cost equation results in reducing the overall cost. Specifically, if the quantities q_c and q_t are reduced, the resulting improvement in correct classification can be translated into reduced costs.

But it may be difficult to specify the actual dollar values for the quantities Q_c, Q_t, V, and B. Therefore, it may be more convenient and still relevant to define relative costs as follows:

Let $Q_t/Q_c = k$, $0 < k < \infty$, be the cost of misclassifying a treatment subject relative to the cost of misclassifying a control subject. Likewise,

Let $V/B = c$, $0 < c < \infty$, be the cost of a viral (or rickettsial) assay relative to the cost of a bacterial assay. And finally,

Let $Q_c/B = r$, $0 < r < \infty$, be the cost of misclassifying a control subject relative to the cost of a bacterial assay. Upon substitution of these terms into the cost equation above, we have

$$\text{Relative Cost} = q_c Q_c + k q_t Q_c + cvB + bB$$

$$= r q_c B + r k q_t B + cvB + bB \text{ and}$$

$$(\text{Relative Cost})/B = r(q_c + k q_t) + cv + b.$$

This expression for (Relative Cost)/B, a dimensionless quantity, now depends only on the two error rates, q_c and q_t, and the number of assays. It continues to be a criterion for excellence; the lower the cost, the better the discriminator. The result of dividing Relative Cost by the factor B has no effect on the criterion because B has been defined as a constant. Evaluation of the various discriminants does depend upon the values selected for the constants, r, k, and c. Of these, it was easy to give a value to only one ratio, c. It was deemed that the cost of an HI test was essentially equivalent to the costs of other methods of assay (excluding serum neutralization); therefore the value of c was set as unity. Rather than use a single value for r, two values were suggested by a panel to represent a reasonable range; these were 50 and 250. Because the ratio of the cost of misclassification was both difficult to assign a value to and because the quantity, (Relative Cost)/B, was very sensitive to it for low values, three values were assigned initially to k — k = 1, 30, 1000-- to cover a wide range.

B. Evaluation Results for the Discriminant Functions Only

To evaluate every one of the discriminants listed in Table 4 would include many discriminants with obviously poor performance, namely those without agent #1. Therefore, only those discriminants that included agent #1 were considered for the cost function analysis. Every discriminant that included Agent #1 but omitted the decision rules was evaluated for all combinations of the ratio constants c, k, and r. Results of these evaluations are given in Table 6. There it can be seen that for both r values and the two higher values of k, the discriminant consisting of agents 1-6-15 was consistently top ranked. Running second and third were discriminants 1-6-4-15 and 1-6-4-15-17. The performance of these first three discriminants as ranked in cost and measured by percentage correct classifications were 80-96, 83-95.5, and 80-95.5 respectively for the control and treatment groups.

There was no effect in changing the value of r from 50 to 250 for the two higher values of k on the ranking of the first five discriminant functions:

1-6-15
1-6-4-15
1-6-4-15-17
1-6-4-2
1-15

which had the same order in the four rankings involved.

However, there was a drastic reversal of the rankings when the value of unity was assigned to k. The first three functions referred to above did not appear at all in the first 20 ranks for either r = 50 or r = 250. Because of this reversal, further investigation of the effect of k on the ranking of the discriminants was deferred until the addition of the decision rules has been evaluated by the cost function.

C. Evaluation Results for Discriminant Functions Plus Decision Rules

The same cost function developed above was applied to the results of the discriminant function plus decision rules for all combinations of c, k, and r. These are ranked by increasing cost in Table 7. Two results are immediately apparent. There was an improvement in classification according to the cost function criterion with the addition of the decision rules for the values of k = 30 and k = 1,000 by virtue of the decreased costs. This was not so for values of k = 1 and both values of r, where only the discriminant function had generally lower costs than discriminant plus decision rules and therefore was a more effective discriminator. This result stemmed from comparing the entries in Table 6 vs Table 7 rank by rank for each of the six categories.

The second result was the fact that the top rankings with or without the decision rules tended strongly to be the same for values of k = 30 and k = 1,000. For example in the four cases k = 30, k = 1,000 for both values of r, the top three rankings were composed of the same discriminant functions. For the top-ranked six discriminant function plus decision rules in Table 7, five of these six discriminant functions were also top-ranked in the first six places for discriminant function without the decision rules; this result held for the same four cases of k = 30, k = 1000 for both values of r. Note that costs are not to be compared across categories of k and r; they are relevant only within a given category.

The results including the decision rules also encountered the previous experience of changing drastically when k = 1. Therefore it is worthwhile investigating the behavior of the cost function as it is affected by changes in the parameter k, especially in the region between k = 30 and k = 1,000 where the large reversals occurred as noted above. Arbitrarily, the value of r was kept fixed at r = 250 while intermediate values of k = 20, 10, 5, and 2 were added to the original set of k = 1,000, 30, and 1. The ranking of the discriminant functions plus decision rules was obtained for each value of k. These are shown in Table 8 where it is obvious that the ranking at k = 20 remained identical to that for k = 30 and k = 1000. A small shift in the ranks occurred at the value k = 10; much larger disruptions occurred for values of k lower than 10. As a method of looking quantitatively at the degree of change in the ranks with the change in k, it was convenient to compute Spearman's Coefficient of Rank Correlation with k = 1,000 as the base line successively with rankings for all other values of k. A plot of Spearmen's Rho versus k is shown in Figure 1. It may be concluded for r = 250 that the rankings of the discriminants are essentially unchanged for values of k larger than 10. Thus the choice of a winning discriminant plus decision rule for those considered

here can be based on a simple judgment as to whether k should be considered smaller than 10. For the case of k = 10 or more, the consistent number one performer is based on the discriminant composed of agents 1-6-4-15 and the decision rules based on agents 5, 8, and 13.

The temptation to reach down to discriminant function 1-15, which is ranked in fourth place for the higher values of k and use it on the basis of the implied economy of only two assays (plus the three for the decision rule) should be resisted. Since the cost function has already included the cost of the assays, a selection of other than the winning combination would be based on false premises.

TABLE 1

LIST OF AGENTS AND TYPES OF TITER ASSAY USED IN THE SERO-SURVEY

AGENT #	TYPE OF TITER ASSAY
1	Rapid Tube Agglutination
2	Rapid Tube Agglutination
3	Agar-gel Precipitin Inhibition
4	Hemagglutination Inhibition
5	Complement Fixation
6	Complement Fixation
7	Complement Fixation
8	Agar-gel Precipitin Inhibition
9	Hemagglutination Inhibition
10	Hemagglutination Inhibition
11	Hemagglutination Inhibition
12	Hemagglutination Inhibition
13	Hemagglutination Inhibition
15	Hemagglutination Inhibition
16	Hemagglutination Inhibition
17	Hemagglutination Inhibition
18	Hemagglutination Inhibition
19	Hemagglutination Inhibition

TABLE 2. PRELIMINARY SCREENING BY GEOMETRIC MEANS AND FREQUENCY

Agent #	Definition of Positive Titer[-1]	Definition of Specific Titer[-1]	Control G-mean Positive Titer[-1]	Treatment G-Mean Positive Titer[-1]	t-test on G-means	χ^2-test on Frequency	Tentative Use
1	10	80	24.3	201.	19.0**	188.**	X
2	25	200	39.9	52.6	2.9**	2.6(1)	X
3	2	2	7.6	9.3	.8	19.4**	
4	5	80	128.	251.	3.1**	7.4**	X
5	5	10	0.	13.5		29.7**	
6	5	10	5.0	8.9	4.3**	4.7*	X
7	5	10	8.6	5.7	1.7(1)	2.3(2)	X
8	2	2	2.8	6.0	2.1*	46.5**	X
9	20	40	0.	30.3		.2	
10	20	40	44.1	41.6	.6	2.7(1)	
11	20	40	0.	26.4		.3	
12	20	40	20.0	24.8	1.3(2)	.3	
13	20	40	0.	32.6		13.8**	
14							
15	20	40	22.4	27.8	2.6**	2.3(2)	X
16	20	40	0.	23.3		.7	
17	20	40	20.0	71.3	2.4*	1.1	X
18	20	40	0.	0.			
19	20	40	20.0	29.0	1.8(1)	1.1	X

** Statistically significant at .01 level
* Statistically significant at .05 level
(1) Statistically significant at .10 level
(2) Statistically significant at .20 level

G-mean: Geometric mean of positive titers.

Frequency: χ^2 on frequency of specific titers.

TABLE #3. PERCENT CORRECT CLASSIFICATION BY SOME AGENTS INDIVIDUALLY

Agent #	Percent correct classification control	Percent correct classification treatment
1	94.0	88.1
6	98.0	10.0
4	52.0	63.7
2	49.0	59.7
15	43.0	61.2
17	99.0	3.0
8	98.0	39.3

TABLE 4. PERCENT CORRECT CLASSIFICATION FOR COMPETING DISCRIMINANT FUNCTIONS

Agent #'s	Percent correct classification control	Percent correct classification treatment
1-6	94.0	88.6
1-4	83.0	94.5
1-2	89.0	94.0
1-15	81.0	95.0
1-17	94.0	88.6
1-8	94.0	90.0
6-4	51.0	67.2
6-2	48.0	64.2
6-15	98.0	10.0
6-17	97.0	12.4
6-8	96.0	44.8
4-2	51.0	66.2
4-15	47.C	70.6
4-17	51.0	65.7
4-8	90.0	49.3
1-6-4	90.0	91.0
1-6-2	90.0	94.0
1-6-15	80.0	96.0
1-6-17	94.0	89.1
6-4-2	58.0	65.7
6-4-15	56.0	68.2
6-4-17	50.0	68.2
4-2-15	52.0	69.7
4-2-17	57.0	64.7

Agent #'s	Percent Correct classification control	Percent correct classification treatment
1-6-4-2	90.0	95.0
1-6-4-15	83.0	95.5
1-6-4-17	90.0	91.5
1-6-4-8	94.0	90.0
1-6-4-19	90.0	90.5
1-6-2-15	89.0	94.0
1-6-2-17	90.0	94.5
1-4-2-15	89.0	94.0
1-4-2-17	90.0	94.5
6-4-2-15	64.0	62.7
6-4-2-17	60.0	64.7
6-4-15-17	62.0	66.7
6-2-15-17	65.0	44.8
4-2-15-17	58.0	67.7
1-6-4-2-17	91.0	93.5
1-6-4-15-17	80.0	95.5
1-6-2-15-17	90.0	94.0
1-4-2-15-17	90.0	92.5
6-4-2-15-17	61.0	62.2
1-6-4-2-15-17	92.0	93.5
1-6-4-2-15-17-8	94.0	94.5
1-6-4-2-15-17-8-19	92.0	94.5
1-6-4-2-15-17-8-19-7	92.0	94.5

TABLE 5. COMPARISON OF CORRECT CLASSIFICATION: DECISION RULES V8 DISCRIMINANT FUNCTION ONLY

Agent #'s	Discriminant Function Only		Discriminant Function + Decision Rules	
	Control	Treatment	Control	Treatment
1-4	83.0	94.5	81.0	97.0
1-2	89.0	94.0	87.0	97.0
1-15	81.0	95.0	79.0	98.0
1-6-2	90.0	94.0	88.0	97.0
1-6-15	80.0	96.0	78.0	98.5
1-6-4-2	90.0	95.0	88.0	97.5
1-6-4-15	83.0	95.5	81.0	98.5
1-6-2-15	89.0	94.0	87.0	97.5
1-6-2-17	90.0	94.5	88.0	97.0
1-4-2-17	90.0	94.5	88.0	97.0
1-6-4-2-15	90.0	94.0	88.0	97.5
1-6-4-2-17	91.0	93.5	89.0	96.5
1-6-4-15-17	80.0	95.5	78.0	98.5
1-6-2-15-17	90.0	94.0	88.0	97.5
1-4-2-15-17	90.0	92.5	88.0	96.5
1-6-4-2-15-17	92.0	93.5	90.0	97.5
1-6-4-2-15-17-8	94.0	94.5	92.0	97.0
1-6-4-2-15-17-8-19	92.0	94.5	90.0	97.0
1-6-4-2-15-17-8-19-7	92.0	94.5	90.0	97.0

TABLE 6. RANKING OF DISCRIMINANT FUNCTIONS BY COST

Discriminant Function	c=1,k=1,000 r=250	Discriminant Function	c=1,k=1,000 r=50	Discriminant Function	c=1,k=1 r=250
	X10				
1-6-15	1000	1-6-15	2.003	1-6-4-2-15-17-8	35.7
1-6-4-15	1124	1-6-4-15	2,251	1-6-4-2	41.4
1-6-4-15-17	1125	1-6-4-15-17	2,254	1-6-4-2-15-17-8-19	41.7
1-6-4-2	1247	1-6-4-2	2,497	1-8	41.9
1-15	1249	1-15	2.499	1-6-4-2-15-17	42.2
1-6-4-2-15-17-8	1370	1-6-2-17	2.745	1-6-2-17	42.7
1-6-4-2-15-17-8-19	1371	1-4-2-17	2.745	1-4-2-17	42.7
1-6-4-2-15-17-8-19-7	1371	1-6-4-2-15-17-8	2.746	1-6-4-2-15-17-8-19-7	42.7
1-6-2-17	1371	1-4	2.747	1-6-2	42.9
1-4-2-17	1371	1-6-4-2-15-17-8-19	2.748	1-6-4-2-17	43.7
1-4	1373	1-6-4-2-15-17-8-19-7	2.749	1-6-4-8	43.9
1-2	1495	1-2	2.993	1-2	44.4
1-6-2	1495	1-6-2	2.993	1-6-4-2-15	44.9
1-6-4-2-15	1496	1-6-2-15	2.995	1-6-2-15-17	44.9
1-6-2-15-17	1496	1-4-2-15	2.995	1-6-17	45.4
1-6-2-15	1496	1-6-4-2-15	2.995	1-6	45.6
1-4-2-15	1496	1-6-2-15-17	2.995	1-17	45.6
1-6-4-2-15-17	1620	1-6-4-2-17	3.243	1	45.9
1-6-4-2-17	1620	1-6-4-2-15-17	3.244	1-6-2-15	46.4
1-4-2-15-17	1869	1-4-2-15-17	3.741	1-4-2-15	46.4
1-6-4-17	2117	1-6-4-17	4.238	1-4-2-15-17	48.7
Discriminant Function	**c=1, k=30 r=250**	**Discriminant Function**	**c=1,k=30 r=50**	**Discriminant Function**	**c=1, k=1 r=50**
1-6-15	351	1-6-15	73	1	9.970
1-6-4-15	382	1-6-4-15	80	1-8	9.975
1-6-4-15-17	390	1-6-4-15-17	82	1-2	10.485
1-6-4-2	402	1-6-4-2	84	1-6	10.721
1-15	422	1-15	86	1-17	10.721
1-6-4-2-15-17-8	432	1-6-2-17	91	1-6-2	10.985
1-6-4-2-15-17-8-19	438	1-4-2-17	91	1-6-17	11.473
1-6-2-17	439	1-6-4-2-15-17-8	92	1-6-4-2	11.488
1-4-2-17	439	1-4	93	1-6-2-17	11.736
1-6-4-2-15-17-8-19-7	439	1-6-4-2-15-17-8-19	94	1-4-2-17	11.736
1-4	455	1-6-4-2-15-17-8-19-7	95	1-6-4-8-19	11.975
1-6-2	476	1-2	97	1-6-4	12.478
1-2	477	1-6-2	98	1-6-2-15	12.485
1-6-4-2-15	478	1-6-2-15	99	1-4-2-15	12.485
1-6-2-15-17	478	1-4-2-15	99	1-6-4-2-17	12.734
1-6-2-15	479	1-6-4-2-15	100	1-6-4-2-15-17-8	12.736
1-4-2-15	479	1-6-2-15-17	100	1-6-4-2-15	12.985
1-6-4-2-15-17	511	1-6-4-2-17	107	1-6-2-15-17	12.985
1-6-4-2-17	513	1-6-4-2-15-17	107	1-6-4-17	13.229
1-4-2-15-17	590	1-4-2-15-17	122	1-6-4-2-15-17	13.234
				1-4	13.236

Note: Values tabled are (Relative Cost)/B

TABLE 7. RANKING OF DISCRIMINANT FUNCTIONS PLUS DECISION RULES BY COST

Discriminant Function	c=1,k=1,000 r=250	Discriminant Function	c=1,k=1,000 r=50	Discriminant Function	c=1,k=1 r=250
	X10				
1-6-4-15	378.6	1-6-4-15	762.7	1-6-4-2-15-17-8	37.5
1-6-15	379.2	1-6-15	763.3	1-6-4-2-15-17	40.2
1-6-4-15-17	379.4	1-6-4-15-17	765.3	1-6-4-2	43.2
1-15	500.3	1-15	1010.5	1-6-2	43.5
1-6-4-2-15-17	625.3	1-6-4-2	1256.8	1-6-4-2-15-17-8-19	43.5
1-6-4-2	625.6	1-6-2-15	1257.3	1-6-4-2-17	44.2
1-6-4-2-15	625.7	1-6-4-2-15	1257.8	1-6-4-2-15	44.2
1-6-2-15-17	625.7	1-6-2-15-17	1257.8	1-6-2-15-17	44.2
1-6-2-15	625.8	1-6-4-2-15-17	1257.8	1-6-2-17	44.5
1-6-4-2-15-17-8	749.3	1-2	1504.0	1-4-2-17	44.5
1-6-2	749.9	1-6-2	1504.5	1-6-4-2-15-17-8-19-7	44.5
1-6-4-2-15-17-8-19	749.9	1-6-2-17	1505.5	1-2	45.0
1-6-2-17	750.0	1-4-2-17	1505.5	1-6-2-15	45.7
1-4-2-17	750.0	1-6-4-2-15-17-8	1506.5	1-4-2-15-17	46.7
1-6-4-2-15-17-8-19-7	750.0	1-4	1507.0	1-6-4-15	58.2
1-2	750.0	1-6-4-2-15-17-8-19	1508.5	1-4	60.0
1-4	751.5	1-6-4-2-15-17-8-19-7	1509.5	1-15	62.5
1-6-4-2-17	874.2	1-6-4-2-17	1754.8	1-6-15	64.7
1-4-2-15-17	874.4	1-4-2-15-17	1755.3	1-6-4-15-17	66.7
Discriminant Function	**c=1,k=30 r=250**	**Discriminant Function**	**c=1,k=30 r=50**	**Discriminant Function**	**c=1,k=1 r=50**
1-6-4-15	166	1-6-4-15	38.9	1-2	13.0
1-6-15	173	1-6-15	39.4	1-6-2	13.5
1-6-4-15-17	175	1-6-4-15-17	41.4	1-6-4-2	14.2
1-15	207	1-15	45.4	1-6-2-17	14.5
1-6-4-2-15-17	221	1-6-4-2	50.3	1-4-2-17	14.5
1-6-4-2	224	1-6-2-15	50.8	1-6-2-15	14.7
1-6-4-2-15	225	1-6-4-2-15	51.3	1-6-4-2-15	15.2
1-6-2-15-17	225	1-6-2-15-17	51.3	1-6-4-2-17	15.2
1-6-2-15	226	1-6-4-2-15-17	51.3	1-6-2-15-17	15.2
1-6-4-2-15-17-8	254	1-2	56.3	1-6-4-2-15-[illegible]7	15.2
1-6-2	260	1-6-2	56.8	1-6-4-2-15-17-8	15.5
1-6-4-2-15-17-8-19	260	1-6-2-17	57.8	1-4-2-15-17	15.7
1-6-2-17	261	1-4-2-17	57.8	1-4	16.0
1-4-2-17	261	1-6-4-2-15-17-8	58.8	1-15	16.5
1-6-4-2-15-17-8-19-7	261	1-4	59.3	1-6-4-15	17.2
1-2	261	1-6-4-2-15-17-8-19	60.8	1-6-4-2-15-17-8-19	17.5
1-4	276	1-6-4-2-15-17-8-19-7	61.8	1-6-15	17.8
1-6-4-2-17	297	1-6-4-2-17	65.7	1-6-4-2-15-17-8-19-7	18.5
1-4-2-15-17	299	1-4-2-15-17	66.2	1-6-4-15-17	19.8

Note: Values tabled are (Relative cost)/B

TABLE 8. CHANGE IN RANKINGS OF DISCRIMINANT FUNCTIONS PLUS DECISION RULES BY COST FOR CHANGES IN k

c = 1; r = 250

Agent #'s	Values of k						
	1000	30	20	10	5	2	1
1-6-4-15	1	1	1	1	7	15	15
1-6-15	2	2	2	3	15	18	18
1-6-4-15-17	3	3	3	7	17	19	19
1-15	4	4	4	10	18	17	17
1-6-4-2-15-17	5	5	5	2	1	2	2
1-6-4-2	6	6	6	4	3	3	3
1-6-4-2-15	7.5	7.5	7.5	5.5	4.5	4.5	7.5
1-6-2-15-17	7.5	7.5	7.5	5.5	4.5	4.5	7.5
1-6-2-15	9	9	9	8	6	11	13
1-6-4-2-15-17-8	10	10	10	9	2	1	1
1-6-2	11.5	11.5	11.5	11.5	8.5	6.5	4.5
1-6-4-2-15-17-8-19	11.5	11.5	11.5	11.5	8.5	6.5	4.5
1-6-2-17	14	14	14	14	11	9	10
1-4-2-17	14	14	14	14	11	9	10
1-6-4-2-15-17-8-19-7	14	14	14	14	11	9	10
1-2	16	16	16	16	13	12	12
1-4	17	17	17	19	19	16	16
1-6-4-2-17	18	18	18	17	14	13	6
1-4-2-15-17	19	19	19	18	16	14	14
Values of Spearman's Rho		1.00	1.00	.928	.298	-.049	-.168

Figure 1 - Spearman's Rho for Rankings of Discriminant Function plus Decision Rules for Intermediate Values of k with N = 1,000 as base, r = 250

SIXTH SAMUEL S. WILKS AWARD

Presentation made by
Dr. Frank E. Grubbs

Acceptance remarks made by
Professor O. Kempthorne

SNEDECOR AWARDED THE 1970 SAMUEL S. WILKS MEMORIAL MEDAL

Professor George Snedecor, now retired from Iowa State University, has been awarded the Samuel S. Wilks Memorial Medal for 1970 The announcement of Professor Snedecor's selection for the 1970 Wilks Award was one of the highlights of the Sixteenth Annual Conference on the Design of Experiments in Army Research, Development and Testing, which was held at the U. S. Army Logistics Management Center, Fort Lee, Virginia, 21-23 October 1970. Professor Snedecor has long been recognized as a pioneer, an international authority and one of the outstanding applied statisticians in both the U. S. A. and countries abroad as well, having made many fundamental contributions to statistical methodology and analysis of statistical experiments. The citation for Professor Snedecor reads as follows:

> "To George W. Snedecor for his pioneer contributions in the development and use of statistical methods, including applications of experimental design to research investigations, and for introducing several generations of statisticians and research workers to the subject of statistics through teaching and the six editions of his world-renowned book, Statistical Methods."

Previous recipients of the Samuel S. Wilks Memorial Medal include John W. Tukey of Princeton University (1965), Major General Leslie E. Simon (1966), William G. Cochran of Harvard University (1967), Jerzy Neyman of the University of California (1968), and Jack Youden (1969) retired from the National Bureau of Standards.

The Samuel S. Wilks Memorial Medal Award, initiated in 1964 by the U. S. Army and the American Statistical Association jointly, is administered by the American Statistical Association, a non-profit, educational and scientific society founded in 1839. The Wilks Award is given each year to a statistician and is based primarily on his contributions to the advancement of scientific or technical knowledge in Army statistics, ingenious application of such

knowledge, or successful activity in the fostering of cooperative scientific matters which coincidentally benefit the Army, the Department of Defense, the U. S. Government, and our country generally.

The Award consists of a medal, with a profile of Professor Wilks and the name of the Award on one side, the seal of the American Statistical Association and name of the recipient on the reverse, and a citation and honorarium related to the magnitude of the Award funds. The annual Army Design of Experiments Conferences, at which the Award is given each year, are sponsored by the Army Mathematics Steering Committee on behalf of the Office of the Chief of Research and Development, Department of the Army.

The funds for the S. S. Wilks Memorial Award were donated by Philip G. Rust, retired industrialist, Thomasville, Georgia.

With the approval of President T. A. Bancroft of the American Statistical Association, the Wilks Memorial Medal Committee for 1970 consisted of the following:

Professor Robert E. Bechhofer	- Cornell University
Professor William G. Cochran	- Harvard University
Dr. Francis G. Dressel	- Duke University and the Army Research Office-Durham
Dr. Churchill Eisenhart	- National Bureau of Standards
Professor Oscar Kempthorne	- Iowa State University
Dr. Alexander M. Mood	- University of California
Major General Leslie E. Simon	- Retired
Dr. John W. Tukey	- Princeton University
Dr. Frank E. Grubbs, Chairman	- U.S. Army Aberdeen Research & Development Center Aberdeen Proving Ground, Maryland

George W. Snedecor is an international authority who was a pioneer in his field of statistics and related sciences.

Born October 2, 1881 in Memphis, Tennessee, he received his education at the Alabama Polytechnic Institute, the University of Alabama (B.S., 1905, mathematics and physics) and the University of Michigan (A.M., 1913, physics). He joined the Iowa State University faculty as an assistant professor of mathematics in 1913, and remained on campus for 45 years.

Shortly after his arrival at Iowa State, Professor Snedecor took the lead in helping people with problems in statistical applications. In 1924 he participated in weekly sessions conducted by Henry A. Wallace, then editor of Wallace's Farmer. At these sessions rapid machine calculation was demonstrated as a method of solving basic statistical needs.

As a result of these get-togethers a bulletin, "Correlation and Machine Calculation," was written in 1925 by Snedecor and Wallace. It promptly attained worldwide distribution as a pioneer publication in statistics, and remained in use for many years.

Because of his interest and enthusiasm in the field, Professor Snedecor continued statistical experimentation. In 1927 he and A. E. Brandt were put in charge of a newly created Mathematics Statistical Service at Iowa State University. This Service was established to supply the demand for professional help in statistics.

Professor Snedecor had an appreciation for the importance of the work of the eminent British statistician Sir Ronald Fisher, who was to become known as the greatest figure in the history of statistics. As a result, Sir Ronald was brought to Iowa State University as a visiting lecturer during the summers of 1931 and 1936. Iowa State was the first U. S. institute to award Sir Ronald an honorary degree.

In 1933 Iowa State's Statistical Laboratory was established as a research institute under the president, with George W. Snedecor as its first director. It was the first statistical center of its kind in the United States, and its organization provided the impetus for other universities to establish similar research and service institutes in statistics.

As an administrator, Professor Snedecor collected a staff which earned a reputation for excellence. His diligence in developing cooperative agreements between the Statistical Laboratory and the U. S. government provided funds for the expansion of Laboratory personnel staff and its projects.

As a teacher, Professor Snedecor inspired his students to achieve their highest goals, and encouraged them to establish even higher goals. The work and contributions of his students alone have given him a reputation as one of the foremost teachers of statistics during the development of the field.

Professor Snedecor served as director of the Statistical Laboratory until college regulations forced him to relinquish administrative responsibilities in 1947 at the age of 65. During that same year a separate Department of Statistics was established. Snedecor continued as a professor of statistics on a part-time basis until his retirement in 1958, and still is affiliated with the university as professor emeritus.

Professor Snedecor is recognized throughout the world for his pioneer work in the systematic development and wide-spread use of statistical methods. His book, Statistical Methods, first published in 1937, is now in its sixth edition and has sold more than 100,000 copies. It has been translated into

Spanish, Hindi, Japanese and Rumanian. A French translation is now in progress and an Indian reprint has been published in English. Snedecor also is the author of three other books and some 50 papers on statistics.

In 1948 Professor Snedecor was named president of the American Statistical Association. His national and international reputation is indicated by his election as a Fellow of the American Statistical Association, the Institute of Mathematical Statistics, and the American Association for the Advancement of Science. He was elected an Honorary Fellow of the British Royal Statistical Society, an honor given to only a very few Americans.

He holds membership in the International Statistical Institute, the Biometrics Society, Iowa Academy of Science, Phi Beta Kappa, Sigma Xi, Gamma Alpha, Pi Mu Epsilon and Kappa Sigma.

George W. Snedecor received the Iowa State University Faculty Citation in 1955, and honorary Doctorates of Science from North Carolina State University in 1956 and Iowa State in 1958. On May 18, 1970 the building which houses Iowa State's statistical center was formally renamed Snedecor Hall in his honor.

GEORGE W. SNEDECOR: A CHRONOLOGY

October 20, 1881 - born in Memphis, Tennessee

1899-1901 - attended Alabama Polytechnic Institute
1901-1905 - attended University of Alabama
1905 - received B.S., University of Alabama (mathematics and physics)
1905-1907 - instructor, Selma Military Academy
1907-1910 - professor, mathematics, Austin College, Sherman, Texas
1908 - married Gertrude Crosier
1910-1913 - graduate assistant, physics, University of Michigan
1913 - received A.M., University of Michigan (physics)
1913-1914 - assistant professor, mathematics, Iowa State University
1914-1930 - associate professor, mathematics, Iowa State University
1915 - first courses in statistics offered at Iowa State University
1924 - attended a ten-week course conducted by Henry A. Wallace on rapid machine calculation of correlation coefficients, partial correlation and the calculation of regression lines
1925 - co-author, with Wallace, of "Correlation and Machine Calculation," a bulletin which attained worldwide distribution
1927 - in charge, with A. E. Brandt, of a newly created Mathematics Statistical Service at Iowa State University
1931-1947 - professor, mathematics, Iowa State University
1931 - first degree in statistics at Iowa State University, an M.S., was awarded through the Department of Mathematics to his student, Gertrude M. Cox.

1931 - brought R. A. Risher to Iowa State University for the first time, as a visiting professor of statistics during the summer
1933 - organized the Statistical Laboratory as an institute under the president's office
1933-1947 - director, Statistical Laboratory, Iowa State University
1933 - appointed first Station Statistician of the Iowa Agricultural Experiment Station
1934 - Analysis of Variance published by the Iowa State University Press
1935 - named head of the newly created statistical section of the Agricultural Experiment Station
1935 - chairman of the Osborne Club (local staff research club)
1936 - established the custom of tea-drinking at a weekly statistics seminar which provided an opportunity for dialogue between staff and students on research problems presented by persons in substantive areas
1936 - president of Sigma Xi (Iowa State University Chapter)
1937 - first edition of Statistical Methods published by Iowa State University Press
1938 - second edition of Statistical Methods published by Iowa State University Press
1938 - established a cooperative agreement between the Statistical Laboratory and the United States Department of Agriculture which provided funds for the expansion of Laboratory personnel and projects
1938 - brought W. G. Cochran to Iowa State University as a visiting professor- he joined the staff in 1939; Charles P. Winsor joined the staff
1939 - the Statistical Laboratory was moved into the newly built Service Building (now known as Snedecor Hall)
1939 - elected Fellow, American Statistical Association
1940 - third edition of Statistical Methods published by Iowa State University Press
1941 - guest lecturer, North Carolina Institute of Statistics, summer
1942 - elected Fellow, Institute of Mathematical Statistics
1943 - established a cooperative agreement between the Statistical Laboratory and the Bureau of the Census; Master Sample Project began
1944 - instituted a contractural project with the Unites States Weather Bureau
1945-1958 - editor of the Queries Section of Biometrics
1946 - guest lecturer, North Carolina Institute of Statistics, summer
1946 - fourth edition of Statistical Methods published by Iowa State University Press
1947 - retired as director of the Statistical Laboratory, Iowa State University
1947 - Department of Statistics was organized at Iowa State University
1947-1958 - professor, statistics, Iowa State University
1947 - vice president, American Statistical Association

1947 - guest lecturer, Statistical Summer Session, Virginia Polytechnic Institute
1948 - visiting research professor of statistics, Alabama Polytechnic Institute, spring
1948 - president, American Statistical Association
1948 - introduced a pioneering basic introductory course sequence in statistics for undergraduates at Iowa State University, especially designed for students majoring in natural and social sciences
1949 - guest lecturer, North Carolina Institute of Statistics, summer
1950 - limited first edition of Everyday Statistics published by William C. Brown, Dubuque
1951 - limited second edition of Everyday Statistics published by William C. Brown, Dubuque
1951 - guest lecturer, North Carolina Institute of Statistics, summer
1953 - consultant in experimental statistics, Alabama Polytechnic Institute and Florida University, January-June
1954 - elected honorary Fellow, British Royal Statistical Society
1955 - consultant in experimental statistics, Woman's College of North Carolina, January-April
1955 - awarded faculty citation, Iowa State University
1956 - consultant in agricultural statistics, Brazil, January-June
1956 - fifth edition of Statistical Methods published by Iowa State University Press
1956 - received honorary D.Sc. degree, University of North Carolina
1957-1958 -visiting professor, Department of Experimental Statistics, North Carolina State University
1958 - received honorary D.Sc. degree, Iowa State University
1958-1963 - consultant in experimental statistics (part-time), U.S. Navy Electronics Laboratory, San Diego
1959-present - professor emeritus, Iowa State University
1967 - sixth edition of Statistical Methods published by Iowa State University Press
1969 - Service Building renamed Snedecor Hall; formal ceremonies held May 18, 1970

ACCEPTANCE REMARKS OF O. KEMPTHORNE

ON BEHALF OF G. W. SNEDECOR

My assignment is a little strange in that in addition to accepting the award for Professor Snedecor, I have to make some appropriate remarks. As I was not sure what to include in these, I wrote to Professor Snedecor. He replied, "I leave it to you to decide what is suitable to say. I think you know me better than I know myself. You won't make any false or silly claims."

I do not recall hearing an acceptance speech for an occasion similar to the present so I have no precedent to follow.

I am very glad to accept the award for George Snedecor. I see George Snedecor as one of a very select group of people whose lives had a certain definite aim—the interpretation of data. My list of these is as follows: Quetelet, who was the first, perhaps, to apply ideas of accurate statistical description to a wide variety of populations; Galton, Weldon, and then Karl Pearson, whose energy in quantifying populations was fabulous; Raymond Pearl of the United States, similarly; R. A. Fisher, who, I believe, viewed statistics primarily as a tool for the improvement of human knowledge; and Snedecor.

It is interesting to me that Snedecor became interested in statistics before Fisher did his great work. In fact, Snedecor introduced statistics at Iowa State College in 1915. This is surely remarkable because before that time statistics was regarded almost entirely as a tool for economics. Unfortunately it is still so regarded at many institutions of higher education.

The big public event of Snedecor's work was the publication in 1937 of Statistical Methods. I rate this as second in importance to one book only, Fisher's Statistical Methods for Research Workers. It is most interesting to note that at this time Snedecor was already 56. This should surely encourage many of us that there is still time in our lives to do some outstanding action. I note parenthetically that this book has brought recognition and fame to the statistical effort at Iowa State University, and was partly instrumental in my leaving the "Sceptred isle," and joining its Statistical Laboratory.

I am indebted to Dr. Malin of the Institute for Scientific Information for the following information. This Institute prepares the very valuable Science Citation Index. The Book "Statistical Methods" of G. W. Snedecor was the most cited book in 1961 and from 1964 to 1967 (the last year for which the Statistics are available) was the second most cited book. The first most cited book in these years was Methods of Enzymology edited by S. P. Colowick and N. O. Kaplan.

The number of citations of the book, given to me by Dr. Malin are also interesting:

1961	361
1964	621
1965	684
1966	806
1967	880

So the usage of Snedecor also grew though not necessarily at the same rate as the scientific literature. This book, now with W. G. Cochran as coauthor, has gone through six editions since 1937 and has become to a remarkable extent, the "pharmacopoeia" of collectors of data in the sciences.

For the past fifteen years or so, the field of statistics has undergone controversies of the deepest kind. There are workers in the "foundations of statistics" (a subject which Irwin Bross says should be banned from professional statistical meetings) who appear to take the view that "Statistical Methods" is utter nonsense, and that the people who follow these methods are stupid or have been misled by proponents of the methods. These people are, it is implied, caught up in mores of scientific procedure that have no real logical basis, just as in past centuries peoples were totally involved in the worship of Homeric gods which determined modes of thought and accommodation to the real world.

To understand Snedecor's life and work, it is critical, I believe, to distinguish at least five aspects of the processes of knowledge:

(i) Interaction with already available data to develop ideas;
(ii) Collection of new data according to a thought-out plan;
(iii) Rational processes for interacting with the new data, and with old and new data together;
(iv) Processes for drawing conclusions;
(v) Processes for choice of action in the face of the existent uncertainty.

The dilemmas of statistical philosophy are not new. The recent resurgence of Bayesian processes is due, I believe, to a failure on the part of some groups to recognize that even a very crude categorization of processes of human knowledge must include <u>all</u> the aspects I mention above. This resurgence was undoubtedly stimulated by an earlier effort to force the whole matter into the single last aspect I mention.

Speaking now as George Snedecor, I would say the following:

The aim of my life and writings has been to take part in and to contribute to the processes by which the human mind approaches situations of ignorance, and to develop, teach and practice reasonable ways to collect data and useful ways to interact with data. The material of my book "Statistical Methods" should not be regarded as giving final conclusive answers. It is suggestive but not exhaustive. It is rather an outline of how a rational mind can approach a situation of partial knowledge and partial ignorance in a search for the truth, which is the essential but not totally achievable goal of human endeavor.

A STATISTICAL HIERARCHICAL MODEL FOR FLIGHT TEST DATA OF A VHF-FM DISTANCE MEASURING SYSTEM (DME)

Erwin Biser and Eddie Cornelious
Avionics Laboratory
U.S. Army Electronics Command
Fort Monmouth, New Jersey

ABSTRACT

This paper describes a statistical hierarchical model of flight test data of a VHF-FM distance measuring modification kit for conventional communications FM transceivers. The distance measuring equipment modification kit consists basically of three separate units: an attachment which connects an existing airborne VHF-FM transceiver to a ranging system; an attachment which converts an existing ground VHF-FM transceiver to a ranging transponder; and a display unit.

The analysis of the data is to determine the accuracy of the system under dynamic conditions.

TABLE OF CONTENTS

ACKNOWLEDGMENT

The authors gratefully acknowledge the unstinted effort and aid of Miss Gloria Fornataro in the preparation and typing of this paper.

The remainder of this article has been reproduced photographically from the author's manuscript.

Preceding page blank

I. OBJECTIVE:

This paper describes a statistical hierarchical model for flight test data of a VHF-FM Distance Measuring Equipment (DME) Modification Kit for conventional communication fm transceivers.

II. DESCRIPTION:

The Distance Measuring Equipment Modification Kit consists of three separate units: an attachment which converts an existing aircraft VHF/FM transceiver to a ranging system; an attachment which converts an existing ground VHF/FM transceiver to a ranging transponder; and a display unit which displays the range between the two transceivers in kilometers. There are no internal modifications made to either of the transceivers. The modification can be applied through the external connections of the existing transceiver without the need of any disassembly of either unit.

The aircraft units have a built-in homing capability with which they can determine when the aircraft is pointed directly toward or away from a ground beacon which it is receiving. In many applications, particularly in environments where visibility to the ground is restricted, such as nighttime operation or operation in heavy foliage, it is also desirable for the aircraft to have an indication of the range from its position to the beacon (transmitter) on the ground (fig. 1). The DME modification kit, in conjunction with an airborne and ground transceiver, will provide the ranging function. By modifying the transmitted signal which is normally used to carry voice, the propagation delay and, therefore, the range can be measured using such equipment as the AN/ARC-131 and AN/PRC-77. In this way no additional ground or aircraft transmitters, receivers, antennas, etc., are required and only a small additional attachment and some cabling need be carried. An advanced development model (feasibility type) was developed by RCA for USAECOM. The flight testing of the system was conducted jointly by USAECOM and RCA.

III. FUNCTIONAL DESCRIPTION:

The airborne DME initiates a ranging cycle by using the transmit divider to phase-modulate the transmitter output signal through a phase shift keyer inserted in the antenna line (see figure 2). After the end of the airborne transmit period, the airborne transmit frequency divider acts as a phase memory by continuing to count at its normal rate, thereby generating a phase-coherent reference signal. The ground DME receives the transmitted phase-modulated signal delayed in time by an amount equal to the one-way range propagation time and locks the ground receive divider in-phase with the received signal. The ground receiver is then turned off. This output is used to phase-modulate the output of the ground transmitter. Then, a phase-coherent signal is retransmitted to the airborne DME system. After propagation delay, the airborne unit receives this signal and by means of its phase-locked loop, locks its receive divider in-phase with the received signal. The modulation frequency (1.5 kHz) was chosen such that its period is equal to the maximum DME round trip range time (maximum range of the system is 100 km). Since the transmit divider is generating a phase-coherent reference signal, range may be determined by measuring the phase delay between the airborne receive and transmit divider outputs.

The phase keyer provides a periodic phase shift of 90 degrees in the transmitter carrier which is unmodulated, and a -90 degrees phase shift at the receiving end.

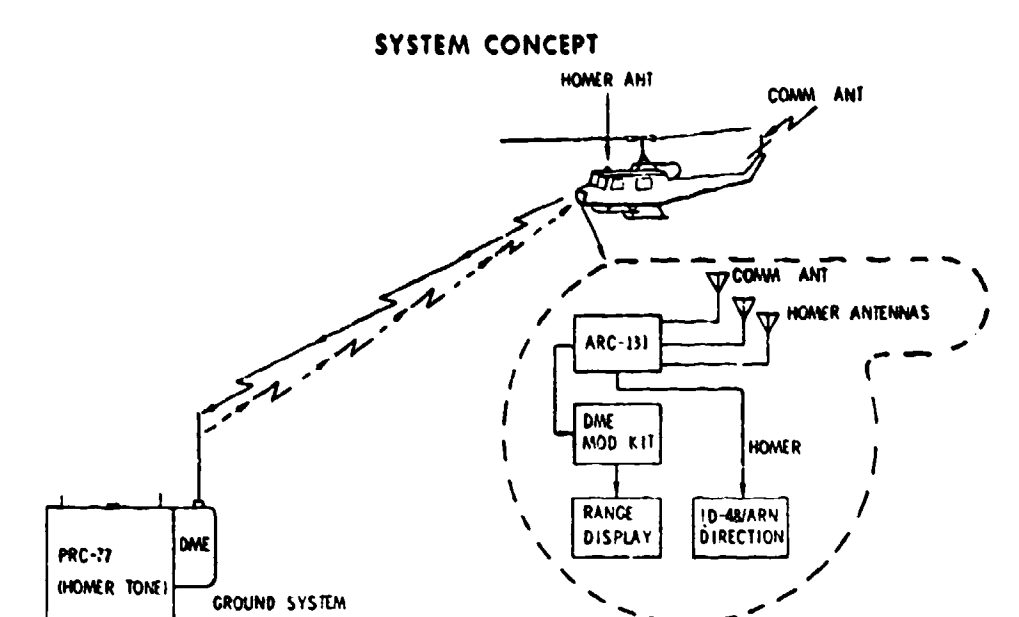


Figure 1.

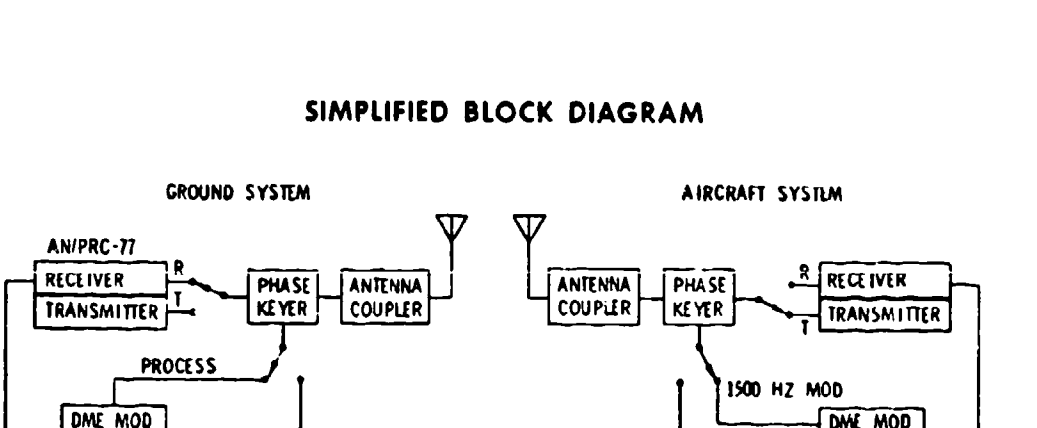


Figure 2

The transceivers associated with the DME operate only in simplex modes; that is, they cannot transmit and receive simultaneously. Therefore, it was necessary to set up a timing cycle that would enable the ground unit to respond at the time the airborne transceiver is in the receive mode. A timing cycle for the airborne unit consists of three periods: transmit, switchover, and receive; and, conversely, the ground unit must cycle through receive, switchover, and transmit. During the switchover phase of this cycle the DME ground unit is not range tracking; it is working on memory. The airborne unit must remember the phase of the signal it has transmitted while the ground unit must remember the phase of the signal it has received. In addition, the aircraft is moving and the range is changing.

IV. OBJECTIVE OF DME TEST PROGRAM:

The tests were designed to show to what extent the equipment meets the goals and requirements set forth in the specifications. Three types of test were conducted: static (tower tests), hover, and dynamic (flight tests).

A. Tower Tests (Static). The tower tests were conducted for the purpose of determining slant range accuracy under static conditions and, to obtain some data on phenomena such as interfering signals, signal fading, and multipath. Furthermore, the tower tests were used in debugging the equipment prior to flight testing. These tests made use of Radio Sets AN/PRC-77 and AN/ARC-131. The ground DME system was installed on top of a 400-foot tower and the airborne DME system was installed in an S-153/U equipment shelter mounted on a 3/4-ton truck. The airborne system was moved to various sites of a known distance from the tower. The DME airborne system measured slant range to the tower. This data is compared with the actual slant range computed from US C&G markers given in x-y coordinates, and the known height of tower and the elevation of each site. Test results are shown in table 1.

Table 1. Static Test

(To calibrate system)

MEASURED SLANT RANGE (km)	DME MEASURED SLANT RANGE (mean) (km)	ERROR (meters)	SAMPLE SIZE	STD DEV	FREQUENCY (MHz)
1.69	1.632	- 58	40	.03741	30.5
	1.818	- 72		.0389	36.3
	1.699	+ 9		.0407	41.6
	2.108	+218		.0356	46.7
6.85	6.97	+120	40	.0428	30.5
	6.82	- 20		.0332	36.3
	6.94	+ 90		.0529	41.6
	7.02	+170		.1570	46.7

B. Hover Test. In order to check the system in an aircraft environment, hover tests were performed. The ground unit was placed at known x-y coordinates and the aircraft hovered at several altitudes over a known location. The computed slant range is compared with the DME measured slant range. This was performed both on a calibrated channel and on an uncalibrated channel. The results of this test are indicated in table 2.

Table 2. Hover Test

ALTITUDE	DISTANCE TO GROUND SITE	MEAN: $\bar{X}$	STD DEV	SAMPLE SIZE	FREQUENCY (MHz)
160	160	54.687	16.211	32	49.7
320	2679	29.166	41.173	24	40.0

It was concluded from the results of both the tower and hover tests that the system was operating properly. Therefore, this brings us to the most important test, the dynamic test to which the nested model is applied.

C. Dynamic Flight Tests. The flight tests are used to evaluate the performance of the equipment under dynamic conditions where appreciable velocities and accelerations exist. A UH-1D helicopter was used for this test. The airborne unit was installed aboard the aircraft and the ground system was located at a known (surveyed) location where an M-33 Fire Control Radar System was also located. The radar system is used as a reference standard for measuring the accuracies of the DME. The ground system was located 30 meters north of the radar system in order to minimize the effect of rf interference.

In this test, the flight paths were chosen so that the aircraft flew toward and away from the ground DME unit in a southward-northward direction. This flight path was chosen because of the location of the airborne antenna. As indicated in figure 3, the airborne communication antenna is located in the tail section of the aircraft. Therefore, the data was grouped according to direction (toward and away from the ground DME unit). This test was performed for nine frequencies where data was recorded at five distances from the ground DME unit. See tables 3 and 4.

In this test we are dealing with a situation involving sampling within each of several major groups. To put it correctly, the directions (south and north) are two levels of one (qualitative) major factor, direction. The other factor, frequency, is nested (or subsumed under) within each level of the major group. Thus, direction is a major classification and frequency is a minor classification.

There are two major groups, south and north directions. The frequencies are nested within the directions. A hierarchical model is characterized by the fact that the levels of one factor (minor classification) are nested within each of the levels of another factor (major classification).

It is well known that the number of levels of the nested factor (frequency in this experiment) need not be the same for all levels of the major factor. We assume that the frequencies are selected randomly, i. e., frequency is a random factor. The directions are also considered random. Our model is given by the following structure.

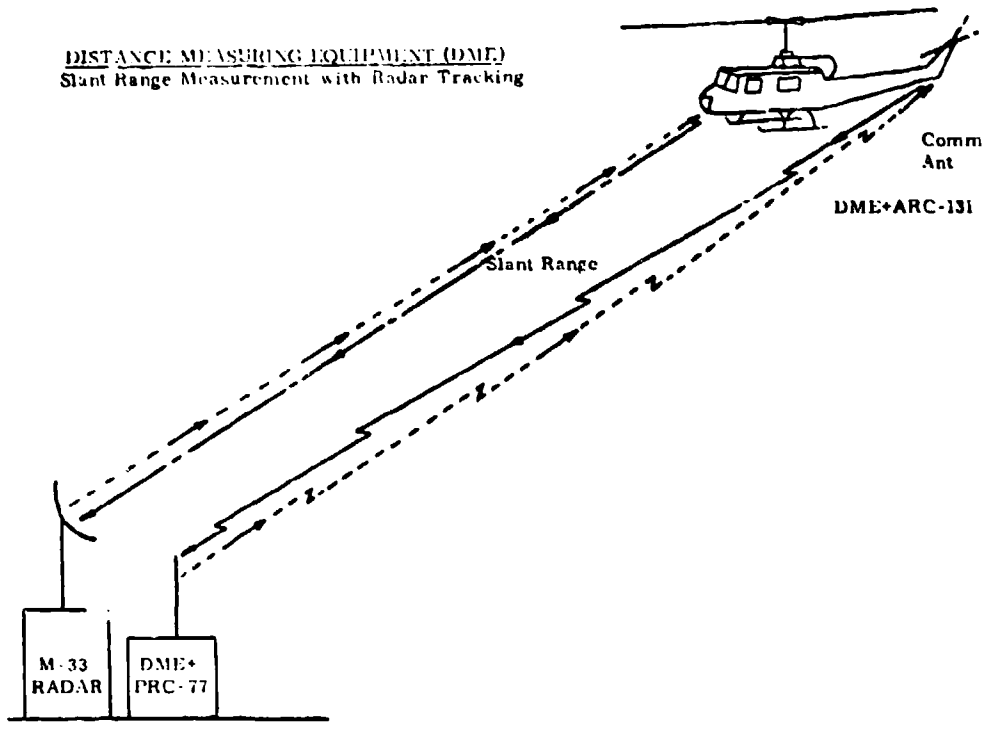


Figure 3.

Table 3. Test Frequencies for Southward and Northward Flights

SOUTHWARD FLIGHT	NORTHWARD FLIGHT
F_1 = 49.7 MHz	F_6 = 46.7 MHz
F_2 = 42.9 MHz	F_7 = 41.6 MHz
F_3 = 38.1 MHz	F_8 = 36.3 MHz
F_4 = 34.1 MHz	F_9 = 32.1 MHz
F_5 = 30.1 MHz	

Table 4. Distance (D) from Radar (M-33) to Aircraft in Kilometers (km)

TEST POINTS	
1	$D_1 = 7.74$
2	$D_2 = 4.32$
3	$D_3 = 1.09$
4	$D_4 = 3.52$
5	$D_5 = 6.81$

V. THE STRUCTURE OF THE MODEL

$$X_{ijk} = u + A_i + B_{j(i)} + e_{k(ij)}$$

The symbolx X_{ijk} and $e_{k(ij)}$ are defined in the sequel i = 1, 2; j = 1, 2, . . ., 5 for the B's in A_1; and i = 6, 7, 8, 9 for the B's in A_2.

A_1 = South (S); A_2 = North (N)

The B's are the levels of frequency nested under the respective levels of directions, north and south.

This model has no interactions since the frequencies are not crossed with the directions.

u = the population mean

A_i = the effects associated with the major groups. (i = 1, 2)

u is defined so that --

$$\sum_{i=1}^{I} A_i = 0$$

$B_{j(i)}$ = the fluctuations between the J random samples drawn from each of the I major groups (I = 2).

The A_i are so defined that --

$$\sum_{j=1}^{J} B_{j(i)} = 0$$

e_{ijk} = the random errors. They are assumed to be normally distributed with the variance equal to σ_e^2.

The $B_{j(i)}$ are so defined that $\sum_{k(ij)} = 0$ for any ij. Denote the estimates of u, A_i, $B_{j(i)}$, and $e_{k(ij)}$ by $\hat{u}, \hat{\alpha}_i, \hat{B}_{j(i)}$ and $\hat{e}_{k(ij)}$ respectively. Denote the sample unbiased estimates of u, A_i, $B_{j(i)}$ by $\bar{x}$; $(\bar{x}_i - \bar{x})$ and $(\bar{x}_{ij} - \bar{x}_i)$ respectively. For each i, we obtain:

$$\sum_{i=1}^{I} \hat{\alpha}_i = \sum_{i=1}^{I} (\bar{x}_i - \bar{x}) = 0$$

$$\sum_{j=1}^{J} \hat{B}_{j(i)} = (\bar{x}_{ij} - \bar{x}_i) = 0$$

$$\sum_{k} \hat{e}_{k(ij)} = \sum_{k} (x_{ijk} - \bar{x}_{ij}) = 0$$

$$\sum_{i}^{I} \sum_{j}^{J} \sum_{k}^{K} (x_{ijk} - \bar{x})^2 = \sum_{i=1}^{I} KJ\,(x_i - \bar{x})^2 + \sum_{i=1}^{I} \sum_{j=1}^{K} k(\bar{x}_{ij} - \bar{x}_i)^2 + \sum_{i=1}^{I} \sum_{j=1}^{J} \sum_{k=1}^{K} (x_{ijk} - \bar{x}_{ij})^2$$

This shows that the total sum of squares is equal to the sum of squares between the levels of A, plus the sum of squares between the levels of B nested within each level of A, plus the sum of squares of the errors. See table 6.

SS_{BM} = sum of squares between major groups.

SS_{Bm} = sum of squares between minor groups (within major groups).

SS_{wm} = sum of squares within minor groups.

$$SS_{BM} = \sum_{i} N_i (\bar{x}_i - \bar{x})^2 = \sum N_i m_i^2$$

where $m_i \equiv (\bar{x}_i - \bar{x})$ is the sample estimate of the major group effect

$$SS_{BM} = \sum_{i} \sum_{j} n_{ij} (\bar{x}_{ij} - \bar{x}_i) = \sum_{i} \sum_{j} n_{ij} t_{ij}^2$$

where $t_{ij} \equiv (\bar{x}_{ij} - \bar{x}_i)$ is the sample estimate of the minor group effect

Table 5. A Two-Factor Hierarchical (Nested) Table for DME Experiment

(Entries are errors (in meters): DME Slant Range-Radar Range.)

Direction	S					N			
Frequencies	F_1	F_2	F_3	F_4	F_5	F_6	F_7	F_8	F_9
Distance from Ground Station									
D_1	190	30	40	- 50	490	-270	300	-400	-390
D_2	180	- 20	230	- 180	350	80	- 90	-	90
D_3	140	+ 60	60	- 180	310	-160	-110	-240	70
D_4	100	-130	140	- 420	210	- 60	-390	-210	-130
D_5	130	- 60	-100	- 470	560	180	-450	-100	-170
Frequency Totals	740	-120	-370	-1300	1920	-230	-740	-950	-530
Sample Size	5	5	5	5	5	5	5	5	5
Direction Total	+1610					-2450			

Table 6. General Analysis of Variance Table for a Nested Experimental Model

SOURCE OF VARIATION	DEGREE OF FREEDOM	SUM OF SQUARES: SS	MEAN SUM OF SQUARES: MS
A_i	I-1	$\sum_{i}^{I} \frac{T_i^2}{KJ} - \frac{T^2}{KIJ}$	$SS_{A/I-1}$
$B_{j(i)}$	I(J-1)	$\sum_{i}^{I}\sum_{j}^{J} \frac{T_{ij}^2}{K} - \sum_{i}^{J} \frac{T_i^2}{KJ}$	$SS_{B/I(J-1)}$
$e_{k(ij)}$	IJ(k-1)	$\sum_{i}^{I}\sum_{j}^{J}\sum_{k}^{K} x^2_{ijk} - \sum_{i}^{I}\sum_{j}^{J} \frac{T_{ij}}{K}$	$SS_{e/IJ(k-1)}$
Totals = (IJK - 1)		$\sum_{i}^{I}\sum_{j}^{J}\sum_{k}^{K} x^2_{ijK} - \frac{T_{..}^2}{KIJ}$	

$$SS_{wm} = \sum_i \sum_j \sum_k (x_{ijk} - \bar{x}_{ij})^2 = \sum_i \sum_j \sum_k e_{ijk}$$

where $e_{ijk} = (x_{ijk} - \bar{x}_{ij})$

Let CF (correction factor) = C = $\sum T^2/N$

where $T = \sum_i \sum_j \sum_k x_{ijk}$ = the total magnitude of all the observations, and

$N = \sum_i N_i = \sum_i \sum_j n_{ij}$; N = the grand total number of observations of the experiment.

The sum of squares for the i^{th} direction is:

$$D_{SS} = \sum_i \left[(T_i^2)/JK\right] - C$$

In this experiment one major group A_1 (South) has five minor groups (frequencies), with five observations for each frequency. The other major group (North) has four minor groups with five observations for each of three frequencies and four observations for one frequency (F_6).

The sum of squares for the directions (D_{SS}) is:

$$D_{SS} = \left[(T_1^2)/25 + (T_2)^2/19\right] - C$$

where T_1 and T_2 are the magnitudes of the total number of observations of the major groups 1 and 2 respectively.

VI. THE ANALYSIS OF VARIANCE FOR HIERARCHICAL CLASSIFICATION

a. For the total sum of squares, the degrees of freedom are 44 - 1 = 43.

b. For the direction sum of squares, the degrees of freedom are 2-1 = 1.

c. For the frequency sum of squares, the degrees of freedom are the number of frequencies minus the number of directions: 9 - 2 = 7.

d. For the error sum of squares, the degrees of freedom are the number of observations minus the number of frequencies: 44 - 9 = 35.

Let us examine the analysis of variance table and analyze the F-test for significance of various sources of variation. We compare the frequency mean square with the error mean square:

$$F = \frac{169539.972}{25457.57} = 6.659$$

The tabular value for 7 and 35 degrees of freedom at the 5 percent significance level is:

$F_{7,\ 35;\ 05} = 2.29$

$F_{7,\ 35;\ 01} = 3.20$ (at the 1 percent level)

In either case there is a difference between the frequencies (not highly significant).

For the significance of directions, we compare the direction mean square with the error mean square:

$$F = \frac{403568.689}{25457.57} = 15.852$$

The tabular values for 1 and 35 degrees of freedom are:

$F_{1,\ 35;\ 05} = 4.12$

$F_{1,\ 35;\ 01} = 7.42$

There is a significant difference between direction (south and north). The observations (the errors) vary both from frequency to frequency and from direction to direction.

VII. FORMULAS FOR HIERARCHICAL MODEL

n_{ij} = number of observations in j^{th} minor group (frequencies) of the i^{th} major group (direction)

= size of j^{th} minor group

$N_i = \sum_{j=1}^{J} n_{ij}$ = number of observations in the i^{th} major group

= size of major group

$N = \sum_{i=1}^{I} N_i = \sum_{i=1}^{I} \sum_{j=1}^{J} n_{ij} = N_1 + N_2$ = grand total number of observations

X_{ijk} = the k^{th} individual observation in the j^{th} minor group of the i^{th} major group

$\bar{X}_{ij} = \sum_{i} \sum_{j} X_{ij}/n_{ij} = T_{ij}/n_{ij}$ = mean value of observations of j^{th} subgroup (minor group) of i^{th} major group

T_{ij} = total value of k observations of j^{th} subgroup of i^{th} major group

where $T_{ij} = \sum_{k=1}^{K_j} X_{ijk}$ = the total value of the observations in j^{th} subgroup in the i^{th} major group

$T_i = \sum_{j=1}^{J} X_{ij}$ = the total value of the observations of the i^{th} major group

$T \equiv X_{...} = \sum_{i=1}^{I} T_i$ = the total value of the observations in the I major group

$\bar{X}_{...} = \bar{X}$ = grand mean of the entire set of observations

$\bar{X} = \bar{X}_{...} = T/N$

VIII. CONCLUSION

In this experiment one major group (South) has five minor groups (frequencies) with five observations for each frequency. The other major group (North) has four minor groups with five observations for each three frequencies.

Let us examine the analysis of the variance table and analyze the F-test for significance of various sources of variation. It is found that there is a difference between frequencies (not highly significant) and there is a significant difference between direction North and South. Therefore, the system varies both from frequency to frequency and from direction to direction.

The errors associated with the frequencies are due to the instability of the transceivers. It was determined during laboratory testing that when the system was calibrated on one frequency, this calibration did not hold for all frequencies in the 30 to 70 MHz band.

When the data was collected for the North direction, the aircraft was going away from the ground transponder and the data for the South direction was collected when the aircraft was going toward the ground transponder. Therefore from figure 1 it can be seen that the communications antenna (DME used) is located in the tail of the aircraft. It may be concluded that the orientation of the aircraft influences the antenna pattern and received and transmitted signal levels. Therefore the decrease in signal level when the aircraft was going toward the transponder caused more jitter in the system. The results are indicated in table 8 where the mean is 168.4 meters for the S-Freq and -134.375 meters for the N-Freq.

Table 7. Mean and Standard Deviation for the Test Frequencies F_1 through F_9

	S					N			
	F_1	F_2	F_3	F_4	F_5	F_6	F_7	F_8	F_9
MEAN	148.00	-24.00	74.00	260.00	384.00	- 45.00	-148.00	-237.50	-106.00
STANDARD DEVIATION	37.013	75.033	118.659	177.904	414.463	180.499	298.026	123.928	196.672

Table 8. Mean and Standard Deviation

MEAN FOR S-FREQUENCIES	MEAN FOR N-FREQUENCIES	OVERALL MEAN FOR $F_1 - F_9$	STANDARD DEVIATION FOR S-FREQUENCIES	STANDARD DEVIATION FOR N-FREQUENCIES	OVERALL STD DEVIATION FOR $F_1 - F_9$
168.400	-134.375	17.012	473.824	418.853	632.414

$S^2_{F_1} + S^2_{F_2} + S^2_{F_3} + S^2_{F_4} = 224510$ $\quad S^2_{F_6} + S^2_{F_7} + S^2_{F_8} + S^2_{F_9} = 175438.323$

Note: The calculations assume statistical independence of the observations of these frequencies.

Table 9. The Analysis of Variance for Hierarchical Classification of Data

SOURCE OF VARIATION	SOURCE OF VARIATION	df	SS	MSS	F-ratios
Between directions	Between major groups	1	403568.69	403568.69	15.85
Between frequencies	Between minor (within major) groups	7	1186779.95	169539.99	6.66
Error within frequencies	Error within minor groups	35	891015.00	25457.57	
	TOTAL	43	2481363.64	57706.13	

Table 10. Standard Deviations from Analysis of Variance

S_D	S_F	S_E	S_T
635.270	411.752	159.554	240.221

Where:

S_D = DMS, DMS is the Direction Mean Square

S_F = FMS, FMS is the Frequency Mean Square

S_E = EMS, EMS is the Error Mean Square

S_T = TMS, TMS is the Total Mean Square

REFERENCES:

Anderson, R. L. and T. A. Bancroft, Statistical Theory in Research, McGraw-Hill Company, Inc., New York, N. Y., 1952

Snedecor, George W., Statistical Methods, The Iowa State College Press, Ames, Iowa, Fifth Edition, 1956

Shelton, C. T., A. T. Crowley, and D. W. Wern, ECOM-0677-F, Development of Distance Measuring Equipment Modification Kit, Phase II

A MULTIVARIATE STATISTICAL MODEL FOR A SEMIAUTOMATIC FLIGHT OPERATIONS CENTER (SAFOC)

Sol Berg and William Peterson
American Electronic Laboratories, Inc.
Colmar, Pennsylvania

Erwin Bieer, Arthur Coppola, and Edward Hansen
Avionics Lab, USAECOM
Fort Monmouth, New Jersey

ABSTRACT

A multivariate model for a design of experiments to obtain information for testing the Semiautomatic Flight Operations Center (SAFOC) for air traffic control is presented.

A scheme for ranking four different methods of operation to determine the best method of operating the system is discussed using confidence intervals constructed from Bonferroni's inequalities.

The interactions among controller teams, traffic density levels, and methods of operation are discussed. Tests for significance of interactions are presented.

An alternative ranking scheme designed to handle traffic densities levels interactions is also presented.

The remainder of this article has been reproduced photographically from the author's manuscript.

A MULTIVARIANT STATISTICAL MODEL FOR A SEMIAUTOMATIC FLIGHT OPERATIONS CENTER (SAFOC)

1 INTRODUCTION:

1.1 OBJECTIVES

The purpose of this paper is to discuss the design of experiments for testing the performance and capabilities of a Semiautomatic Flight Operations Center (SAFOC).

The general purpose of this series of tests is to determine the effectiveness of SAFOC in performing its mission in a realistic environment. The specific purpose of the tests described in this paper, denoted as Phase I tests, is to determine the best method for operating SAFOC and to find the system performance measures and system effectiveness measures that can be attained under these optimum operational conditions.

The Phase II test effort, not described in this paper, will complete the task of evaluating SAFOC in a realistic environment and will determine those changes to SAFOC which will improve its mission effectiveness.

1.2 BACKGROUND:

SAFOC will be tested and evaluated at the Federal Aviation Agency (FAA), National Aviation Facilities Experimental Center (NAFEC), Atlantic City, N.J., by the U.S. Army Electronics Command (ECOM), Avionics Laboratory (AMSEL-VL-G), Fort Monmouth, N.J. The evaluation testing of the SAFOC will be performed in two phases designated Phase I and Phase II. Phase I tests are scheduled for the period from 1 March through 31 August 1970. Phase II tests are to be conducted between 1 October 1970 and 31 January 1971.

These tests are designed to determine system operating characteristics in a realistic environment for evaluation of system merits. This method of testing is usually termed a "Military Potential" test. The SAFOC system was originally tested during the acceptance test by ECOM to determine whether the equipment satisfied the contractual requirement described in the systems description. These tests were intended to validate the design as called out by the specifications.

The test flow diagram, figure 1, shows that the SAFOC system is tested with a number of Army air traffic controller teams at various air traffic densities for the purpose of giving the system a fair test over a broad spectrum of conditions. The SAFOC equipment was designed for flexibility in operation and can be operated in a number of ways. For this reason, various operational methods or doctrines were incorporated into the test scheme.

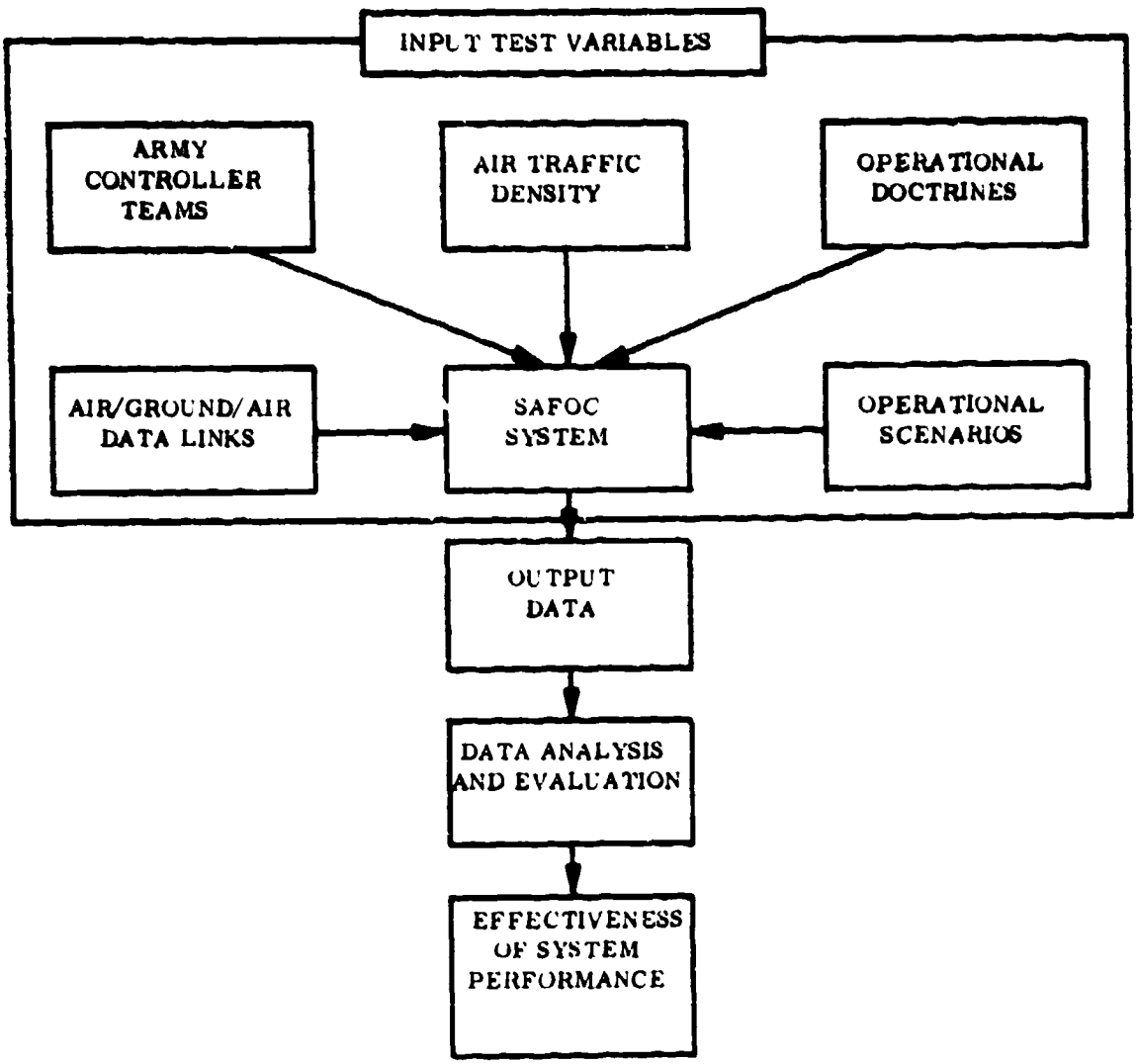


Figure 1. SAFOC Design of Experiments, Test Flow Diagram

These tests of the SAFOC system will provide data for analysis and evaluation. The evaluation, as described later in this paper, will measure system effectiveness and variations.

1.3 SAFOC TECHNICAL OBJECTIVES:

The technical objectives of the SAFOC are:

1. To regulate Army air traffic in an efficient and orderly manner under instrument flight rules (IFR).

2. To provide flight following capability under visual flight rules (VFR).

3. To improve information flow and transfer among system elements and units being supported.

4. To provide computer facilities which can automatically assimilate and analyze air traffic data for display and decision-making by an operator.

5. To provide a rapid and reliable means for making commanders aware of the current air traffic situation for integration into overall tactical planning.

6. To perform specific functions required in an air traffic regulation system, including flight plan processing, flight following, position determination, rescue, traffic corridor planning, and exchange of information with other control centers.

2. GENERAL SYSTEM DESCRIPTION:

2.1 DESCRIPTION OF EQUIPMENT

Figure 2 represents the actual SAFOC configuration within the van. Included within the SAFOC system are the following subsystems:

1. Data processing subsystem
2. Radar processing subsystem
3. Display subsystem
4. Manual backup subsystem

Figure 3 is a block diagram showing the interconnections of the above subsystems within the SAFOC system.

2.2 SYSTEM CAPABILITIES

2.2.1 System Functions

The SAFOC provides an air traffic regulation service by collecting, analyzing, and disseminating the information necessary to regulate the movement of Army aircraft under instrument flight rules (IFR) and monitor the movement of cooperating aircraft under visual flight rules (VFR). The air controller(s) perform the air traffic regulation function using data processing and display equipment to provide the following capabilities:

a. Flight Data Processing**

1. Flight data entry-air filed, ground filed
2. Flight plan clearance
3. Flight plan activation
4. Aircraft position determination
5. Conflict prediction
6. Collision avoidance

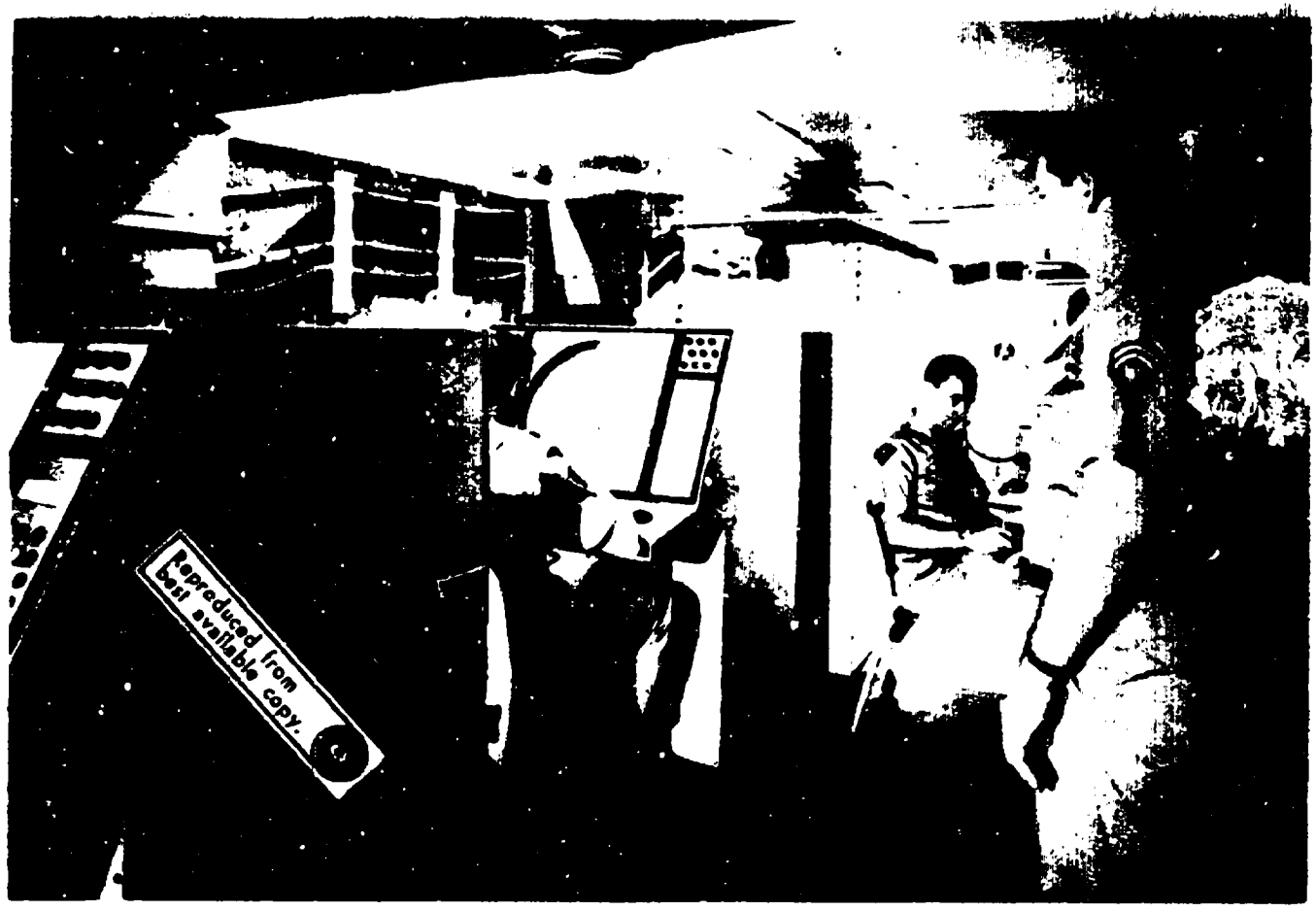


Figure 2. Semiautomatic Flight Operations Van (SAFOC)

7. Accommodating flight plan changes

8. Flight plan deactivation

b. Flight Following

c. Flight Handoff

d. Identification Assistance

e. Emergency Assistance

f. Air/Ground Coordination

g. Ground/Ground Coordination

2.2.2 Flight Tracking Methods

SAFOC provides five methods of flight tracking which, in order of their assigned priority, are:

a. Data Link (auto-tracking only)

b. Radar Beacon (auto-tracking)

c. Radar Beacon (rate-aided manual tracking)

d. Radar Skin Return (rate-aided manual tracking only)

e. Plan Follow

2.3 SAFOC TEST CONFIGURATION:

Figure 4 shows the SAFOC test operations and information flow diagram. As illustrated, the scenario generator program, prior to actual testing, generates scenarios and scripts based on random processes. At the time of testing, the scripts are given to the flight simulator pilots, who keep comprehensive logs of all actions they undertake.

The pilots simulate actual flights using target generators which are part of NAFEC's data link simulation.

Using a pre-determined operational-procedural mode, SAFOC controls the simulated flights and produces exhaustive time histories on magnetic tape. These histories include all actions performed by the equipment or by the controller.

The raw output data tape and the target generator history tape are processed using a series of formatting and editing programs.

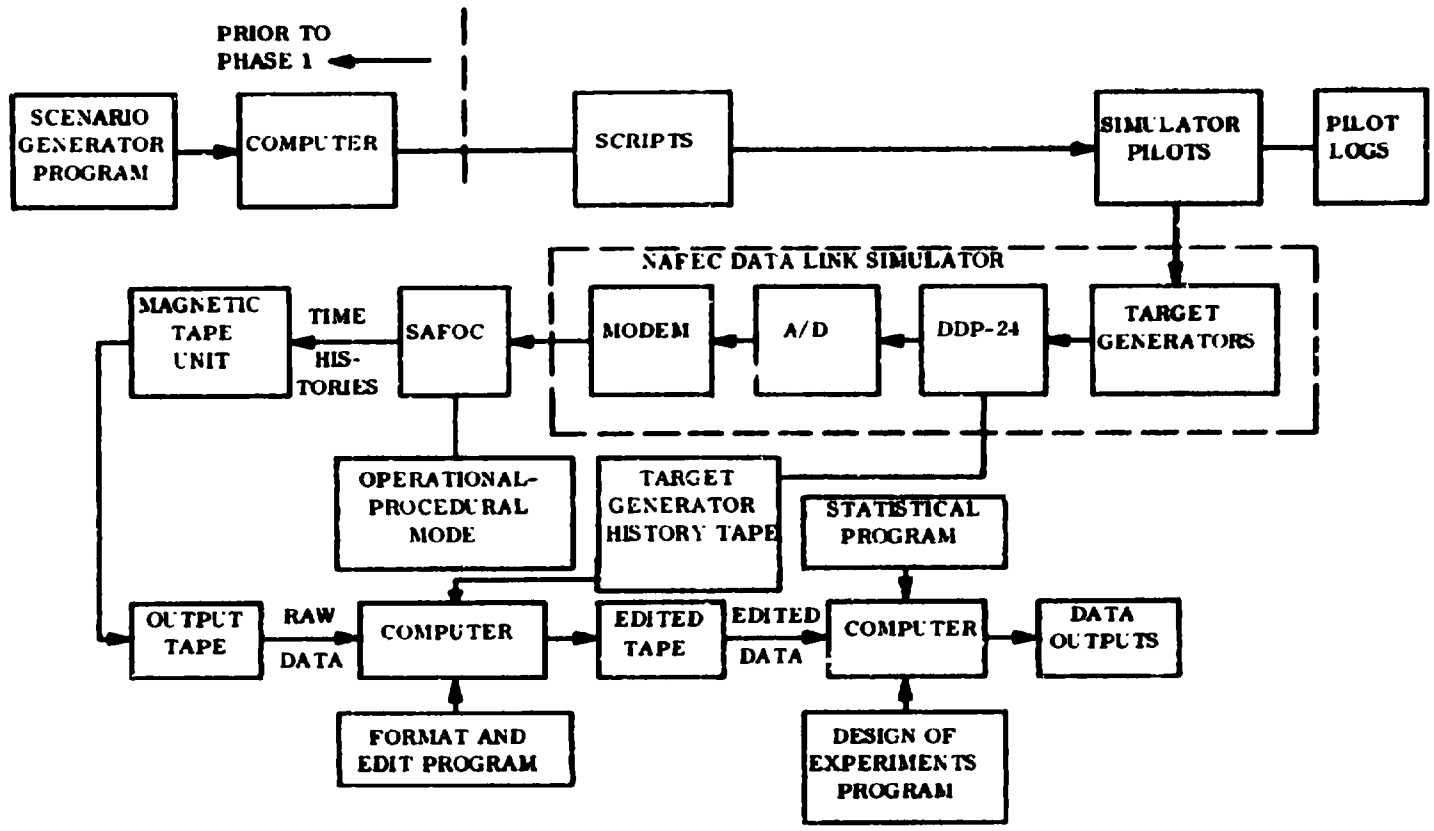


Figure 4. SAFOC Test Operations and Information Flow Diagram

The processed tapes are then operated upon using statistical programs in accordance with the design of experiments, providing the desired data outputs to be described later.

3. SAFOC SYSTEM EVALUATION:

3.1 OBJECTIVE

The purpose of the SAFOC system evaluation to be described is to determine the best operational-procedural combination for SAFOC. The tests will consist of a series of scenarios of three different traffic levels. Each of four controller teams will operate the SAFOC according to four different operational-procedural combinations. The outputs, consisting of system effectiveness measures, will be ranked to determine the best operational-procedural mode.

3.2 TEST FLOW

Figure 5 illustrates the overall statistical test flow for Phase I testing of SAFOC. As shown, the ultimate objective is to determine the best operational-procedural mode. Since the ranking techniques assume that all controller teams are of approximately equal proficiency, it is first necessary to test the validity of this assumption. If it is found that controller performances are not uniform, further training and retesting is necessary. The techniques for testing controller performance are described in paragraph 3.7.

When it is found that all controllers have been trained to a level where any existing differences have no significant effects on the system effectiveness measures (see paragraph 3.5), the controller team tests are begun on the SAFOC system. The performance and effectiveness measures generated from these tests will be used to evaluate SAFOC capability and to determine the best operational-procedural combination.

3.3 STATISTICAL FACTORIAL DESIGN MATHEMATICAL MODEL:

The basic model for this test is used in all 48 runs of the series. The model assumes that the controller teams have essentially the same performance levels. (The validity of this assumption is to be checked prior to the initiation of record runs according to the tests described in 3.7.) The effects of random, or possibly even systematic, changes in performance among teams and from day-to-day is minimized by randomizing the order of performance of the 48 test runs.

The purpose of the initial experimental design is to determine the variation of the dependent variables (in the mean) with the independent variables and to rank these outputs. Further, the purpose is to determine what sort of functional dependencies may exist between inputs and outputs. When such dependencies are discovered, it may be possible to perform regression analysis, using an appropriate model, to obtain analytical expressions for any empirical variations.

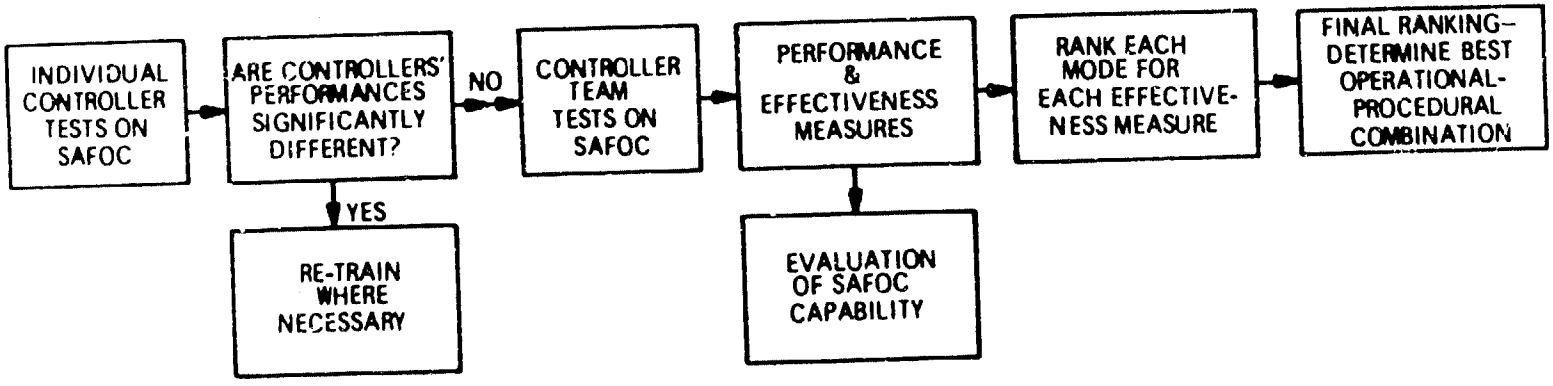


Figure 5. Flow Diagram of Phase I Statistical Testing

Reference runs will be generated, in which aircraft will enter the system and fly through without any air traffic control or regard for conflicts. Entrance rates (generated by a random number process on a computer) will be controlled and changed from run to run as an independent variable. The entire set of reference runs will be available as history. The same "input" conditions are then used in runs in which air traffic control is exercised. Comparisons between the "flythrough" and the controlled runs can then indicate the effectiveness of the SAFOC method of operation.

Reference runs will consist of the flight path charts and scripts from the AEL scenario generator. These contain the time required for a flight to begin and end the mission and represent the shortest time possible for a flight to be in the system. Control functions performed on the flight may require additional flight time. Thus, the reference runs represent an orderly account of operation under "ideal" conditions as a start towards understanding performance under "field" conditions.

3.4 INPUT TEST PARAMETERS:

Four methods of operating the SAFOC System are defined as follows:

a. O_1. Operate the system with two adjoining ATC sectors, one at each display. Each controller is then independently responsible for his own sector.

b. O_2. Operate the system with one sector, of the same area as in O_1, and divide the various ATC functions between the two displays. One controller is the senior; the other is the junior.

c. P_1. An active procedural mode where all flights are displayed to the operators.

d. P_2. A passive procedural mode with displays only as a result of alerts to which the operator reacts.

From the above we generate four operational concepts, namely:

$$O_1P_1,\ O_1P_2,\ O_2P_1,\ O_2P_2$$

These are the four methods which will be tested by the various performance measures.

Three levels of traffic will be generated to test the four methods. Four teams of two controllers each will operate the SAFOC for each method at each traffic level. Figure 6 shows the interactions of the test inputs. Table 1 gives the experimental design in the notation to be used hereafter.

The order of performance of the test runs is randomized as shown in Table 2 to eliminate the effects of the learning process of the operators in the course of performing the test runs.

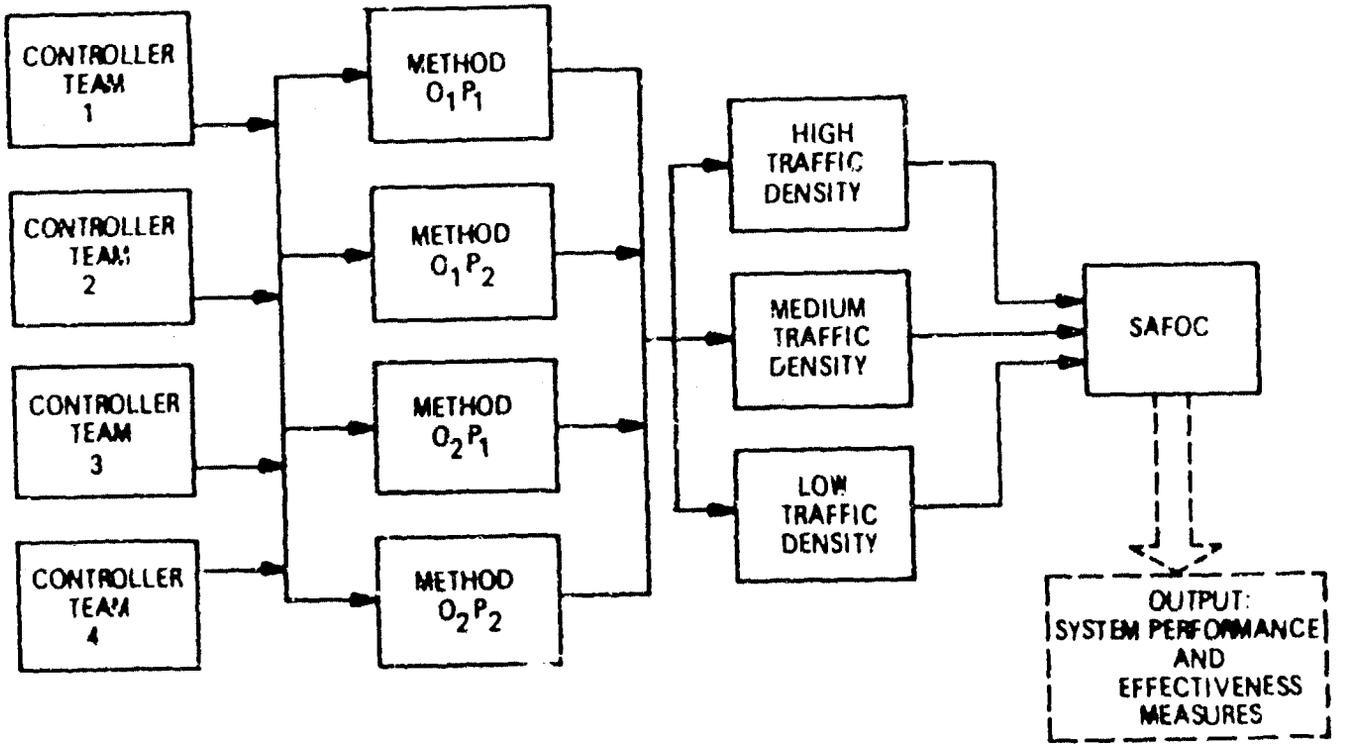


Figure 6. Experimental Design

Table 1. EXPERIMENTAL DESIGN FOR THREE TRAFFIC LEVELS AND FOUR CONTROLLER TEAMS

	Traffic Level 1				Traffic Level 2				Traffic Level 3			
	Operational X Procedural Modes				Operational X Procedural Modes				Operational X Procedural Modes			
Team	O_1P_1	O_1P_2	O_2P_1	O_2P_2	O_1P_1	O_1P_2	O_2P_1	O_2P_2	O_1P_1	O_1P_2	O_2P_1	O_2P_2
1	X_{1111}	X_{1211}	X_{2111}	X_{2211}	X_{1121}	X_{1221}	X_{2121}	X_{2221}	X_{1131}	X_{1231}	X_{2131}	X_{2231}
2	X_{1112}	X_{1212}	X_{2112}	X_{2212}	X_{1122}	X_{1222}	X_{2122}	X_{2222}	X_{1132}	X_{1232}	X_{2132}	X_{2232}
3	X_{1113}	X_{1213}	X_{2113}	X_{2213}	X_{1123}	X_{1223}	X_{2123}	X_{2223}	X_{1133}	X_{1233}	X_{2133}	X_{2233}
4	X_{1114}	X_{1214}	X_{2113}	X_{2214}	X_{1124}	X_{1224}	X_{2124}	X_{2224}	X_{1134}	X_{1234}	X_{2134}	X_{2234}
Sum	S_{111}	S_{121}	S_{211}	S_{221}	S_{112}	S_{122}	S_{212}	S_{222}	S_{113}	S_{123}	S_{213}	S_{223}

$X_{i\,jk\ell}$ is some effectiveness measure (e.g., workload)

$i = 1, 2$ Operational modes

$j = 1, 2$ Procedural modes

$k = 1, 2, 3$ Traffic levels

$\ell = 1, 2, 3, 4$ Teams (four distinct teams throughout)

$$S_{ijk} = \sum_{\ell=1}^{4} X_{ijk\ell}$$

Table 8. ORDER OF TEST PERFORMANCE

(Randomized)

1. X_{2211}	17. X_{1112}	33. X_{1131}
2. X_{2122}	18. X_{2134}	34. X_{1113}
3. X_{1111}	19. X_{1114}	35. X_{2223}
4. X_{1232}	20. X_{1123}	36. X_{1234}
5. X_{1223}	21. X_{2222}	37. X_{2214}
6. X_{2131}	22. X_{1212}	38. X_{1133}
7. X_{1121}	23. X_{1231}	39. X_{2213}
8. X_{2124}	24. X_{2114}	40. X_{2212}
9. X_{1211}	25. X_{1233}	41. X_{1122}
10. X_{2133}	26. X_{1132}	42. X_{1134}
11. X_{2123}	27. X_{2231}	43. X_{2232}
12. X_{2221}	28. X_{2233}	44. X_{1224}
13. X_{1221}	29. X_{2112}	45. X_{1213}
14. X_{2111}	30. X_{2234}	46. X_{1222}
15. X_{1214}	31. X_{2121}	47. X_{2113}
16. X_{2132}	32. X_{2224}	48. X_{1124}

If practical considerations, such as controller availability and/or equipment availability, require deviations from this order of testing, attempts will be made to estimate the resulting effects upon test results. In such cases the test plan will be modified accordingly.

Specifically, if four independent but equally proficient controller teams are not available for the test duration, not enough data will be obtained to make significant evaluations of the system. If necessary, the test plan will be modified to test performance using only one controller per team and the two sector mode of operation will not be utilized.

3.5 OUTPUT TEST PARAMETERS:

This section describes the system performance measures and system effectiveness measures which are obtained by reducing the data generated from test runs on SAFOC.

3.5.1 System Performance Measures

These are measures of the individual items which, in total, influence the system effectiveness. The performance measures represent, for example, the actual contributors to total workload, and it is through improvements in the performance measures that the effectiveness measures can be improved.

The following are the system performance measures:

1. Time to perform a service for each service
2. Service rate
3. Waiting time for service
4. Event time history
5. False dismissal probability
6. Actual density history
7. Queue lengths
8. Typewriter errors
9. Near miss history
10. Communication time history
11. Number of impossible requests

12. Altitude change history

13. Closest approach history

Service time is time between initiation of service and completion of service.

The services to be considered are:

1. Typewriter
2. Handoff
3. Activate
4. Clear flight plan
5. Coordinate
6. Conflict resolution
7. Alert servicing
   (a) Clearance alert
   (b) Conflict 1 alert
   (c) Conflict 2 alert
   (d) Coordinate alert
   (e) Emergency alert
   (f) Flight overdue alert
   (g) Flight plan complete alert
   (h) Flight not active alert
   (i) Poor tracking alert
   (j) Handoff alert

For each service mentioned above, the service rate is determined by dividing the number of times service is performed in a given time span by the time span duration.

Waiting time for service is the time between initiation of a request for service and the initiation of that service.

Queue length is the number of flights waiting for service at any time.

False dismissal probability is the number of actual near misses which did not register a conflict 1 or 2 alert divided by the total number of near misses.

Impossible request alerts are generated whenever a controller asks the computer to perform an impossible task. These will be counted by controller.

Actual density history is a time record of actual flight entrance and exits from the system.

Near miss history is a record of all flight pairs with actual separations less than the minimum required.

Event time history is a complete history of all events in order of occurrence for each flight.

Typewriter errors are counts of the number of typing errors committed by the typist.

Altitude change history records all altitude changes and the time of change.

Closest approach history records all flight pairs within a specified distance, the closest approach of each flight pair and the time of closest approach.

Communication time history records each time a communication is performed and the duration of the communication.

3.5.2 System Effectiveness Measures

The system effectiveness measures are used to provide relative rankings of the operational procedural mixes and to evaluate relative controller team performance. The following measures are chosen because it is believed they represent the characteristics most important to the user:

a. Safety

b. Controller Workload

c. Communications Workload

d. Delays

e. Throughput

f. Capacity

g. Uncontrolled Time

Safety is defined as the number of near misses per aircraft mile flown. This is measured by computing the number of flights separated by less than the minimum distance during a run divided by the total number of flight miles flown during that run. For the purpose of this measure, the run begins when a steady state density is reached.

Controller workload is defined as total time for all flight hooks in a given time span divided by the time span. Controller workload can be used to estimate controller replacement rates, controller scheduling, etc. Since some tasks the controller must perform may be more taxing than others, workload will be measured by task (handoff, conflict resolution, etc.) as well as by total time that work of any kind is done.

Communications workload is defined as the time spent in communications during a run divided by the duration of the run.

Delays are defined as the actual departure time delay from the planned departure time. Delays are believed to be important in a tactical situation. It does little good to get aircraft safely to a particular location if they arrive too late to be of use in some situations. This measure is easily computed by subtracting actual departure times from the planned departure time.

Throughput is defined as the actual number of flights entered during study state divided by the number of planned entries in that time period. The number of flights entered in a given time divided by the number of planned entrances measures the ability of the system to obtain a desired throughput rate.

Capacity is defined as the peak flight density safely handled by the system. This is obtained by performing a regression analysis on near misses vs. peak density. System capacity is that peak density where the resultant regression curve first exhibits significant non-zero slope.

Uncontrolled time is defined as the total time of flights within the SAFOC control area without being controlled by SAFOC.

3.6 MATHEMATICAL MODEL: Operational-Procedural Mode Ranking

Figure 7 shows the flow of statistical tests necessary to arrive at a ranking of the four procedural modes (operational-procedural combinations).

A series of 48 tests, as discussed in 3.4, will be run using the NAFEC flight simulators as input to SAFOC. The performance and effectiveness measures listed in 3.5 will be extracted. The X_{ijkl} discussed in the following sections will refer to any one of the system effectiveness measures. The four procedural modes will be ranked according to each effectiveness measure, and the final ranking will be performed as in 3.6.6.

Returning to figure 7, it is assumed (as stated previously) that the controllers have attained approximately equal levels of proficiency. This assumption will be tested as described in 3.7.

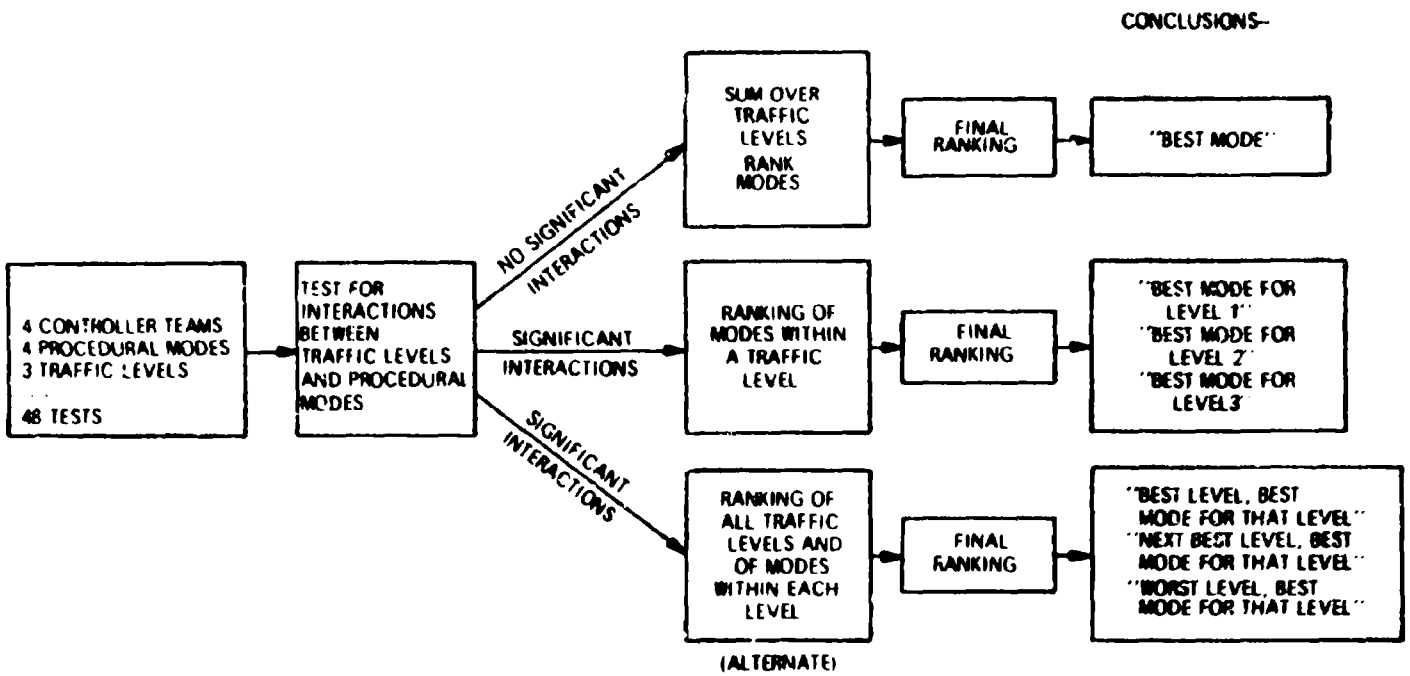


Figure 7. Statistical Testing: Comparison of Operational-Procedural Combinations

After this assumption is validated, the first test on the 48 runs is a test for interactions between traffic levels and the procedural modes. This test is described in 3.6.1.

If significant interactions are absent, the ranking described in 3.6.2 will be attempted.

If significant interactions are detected, one of the two ranking schemes described in 3.6.3 and 3.6.4 will be attempted.

A final ranking will then be done as described in 3.6.6.

The total experiment is a mixed model of fixed treatments consisting of three traffic levels and four modes together with random blocks consisting of the four controller teams.* Another view of the model is that we make four random observations (controller team) on a 12-component vector consisting of the four methods and the three traffic levels. Ranking will be performed using the system performance criteria discussed earlier.

3.6.1 Operational-Procedural Combination Interactions with Traffic Level

The multiple comparison of means to be described in 3.6.2 assumes no interactions between traffic level and the operational-procedural combinations. This section discusses a statistical technique of testing that assumption. If significant interactions are detected, one of the two ranking methods described in 3.6.3 and 3.6.4 can be used.

It is desired to test the multiple hypothesis of no traffic level interactions with the four procedural modes:

$$H_o: \mu_{111} - \mu_{112} - \mu_{121} - \mu_{122} \quad \mu_{211} - \mu_{212} \quad \mu_{221} - \mu_{222}$$

$$\mu_{112} - \mu_{113} = \mu_{122} - \mu_{123} = \mu_{212} - \mu_{213} = \mu_{222} - \mu_{223}$$

where μ_{ijk} is mean performance for the ij operational-procedural combination at traffic level k.

The procedure followed is outlined in Morrison (1967), pp. 186-196 for profile analysis of q responses grouped according to g treatments. In our case there are 3 responses corresponding to the three traffic levels and 4 treatments corresponding to the 4 operational-procedural modes. For each traffic level and mode there are 4 observations, one from each team.

Let $x_{ijk\ell}$ be the system effectiveness using the ij operational-procedural combination at traffic level k with team ℓ.

---

*This is discussed more fully in reference 6 pages 159-196. Tables are found in reference 6 pages 312-319.

Form the differences between adjacent traffic levels

$$C_{ij2\ell} = x_{ij1\ell} - x_{ij2\ell} \text{ and } C_{ij2\ell} = x_{ij2\ell} - x_{ij3\ell}$$

Compute the following sums

$$T_{ijh} = \sum_{\ell=1}^{4} C_{ijh\ell} = \text{sum over all teams on adjacent traffic level differences with the ij operational-procedural combination}$$

and

$$G_h = \sum_{j=1}^{2} \sum_{i=1}^{2} T_{ijh} = \text{grand total of all differences between adjacent traffic levels summed over ij operational-procedural combinations}$$

$$H_{rs} = \sum_{j\ 1}^{2} \sum_{i\ 1}^{2} \frac{1}{4} T_{ijr}\ T_{ijs} - \frac{1}{16} G_r G_s \ ; \ r = 1, 2; s = 1, 2$$

$$E_{rs} = \sum_{i=1}^{2} \sum_{j=1}^{2} \sum_{\ell\ 1}^{4} C_{ijr\ell}\ C_{ijs\ell} - \sum_{i=1}^{2} \sum_{j\ 1}^{2} \frac{1}{4} T_{ijr}\ T_{ijs}; \ r = 1, 2; s = 1, 2$$

It can be shown that if λ_m is the maximum root of

$$\lambda^2 (E_{11}E_{22} - E_{12}^2) + \lambda (2E_{12}H_{12} - E_{11}H_{22} - H_{11}E_{22}) + H_{11}H_{22} - H_{12}^2 = 0,$$

then $\frac{\lambda_m}{1+\lambda_m}$ follows the Heck maximum characteristic root distribution* with parameters 2, 0, 4.5. For a level of significance α, the hypothesis of no traffic mode interaction is accepted if

$$\frac{\lambda_m}{1+\lambda_m} \leq \lambda\ (\alpha)\ 2,\ 0,\ 4.5$$

where $\lambda\ (\alpha)$ is obtained from tables of the Heck distribution*.
For $\alpha = .01$ $\quad \lambda\ (\alpha) = .775$

---

*This is discussed more fully in reference 6 pages 159-196. Tables are found in reference 6 pages 312-319.

3.6.2 Multiple Comparisons of Means among the Operational and Procedural Methods

If it is found that no significant interactions exist between the traffic levels and the operational-procedural methods, the statistical tests given in this section can be used.

Table 3 shows a portion of the total experiment represented by Table 1. The entries for the other two traffic levels differ only in the subscript k representing traffic level.

A 2 x 2 table of averages of the four combinations of the operational and procedural methods for the aggregate data from the three traffic levels can be formed:

		Procedural Mode	
		P_1	P_2
Operational Mode	O_1	$\bar{x}_{11\cdot\cdot}$	$\bar{x}_{12\cdot\cdot}$
	O_2	$\bar{x}_{21\cdot\cdot}$	$\bar{x}_{22\cdot\cdot}$

In the notation of Table 3

$$\bar{x}_{11\cdot\cdot} = \frac{1}{12} \sum_{k=1}^{3} \sum_{\ell=1}^{4} x_{11k\ell}$$

$$= \frac{1}{12} (S_{111} + S_{112} + S_{113}),$$

.

.

.

$$\bar{x}_{22\cdot\cdot} = \frac{1}{12} \sum_{k\ 1}^{3} \sum_{\ell-1}^{4} x_{22k\ell}$$

$$= \frac{1}{12} (S_{221} + S_{222} + S_{223}).$$

Table 3. DETAILS OF EXPERIMENTAL DESIGN FOR TRAFFIC LEVEL 1

Methods of Operating SAFOC				
O_1P_1	O_1P_2	O_2P_1	O_2P_2	
X_{1111}	X_{1211}	X_{2111}	X_{2211}	(Repeat this for each traffic level)
X_{1112}	X_{1212}	X_{2112}	X_{2212}	
X_{1113}	X_{1213}	X_{2113}	X_{2213}	
X_{1114}	X_{1214}	X_{2114}	X_{2214}	
S_{111}	S_{121}	S_{211}	S_{221}	

where $X_{11k\ell}$ denotes the O_1P_1 effectiveness measure
$X_{12k\ell}$ denotes the O_1P_2 effectiveness measure
$X_{21k\ell}$ denotes the O_2P_1 effectiveness measure
$X_{22k\ell}$ denotes the O_2P_2 effectiveness measure

Each column is summed over all four teams for each effectiveness measure

where $S_{111} = \sum_{\ell=1}^{4} X_{111\ell}$, $S_{121} = \sum_{\ell=1}^{4} X_{121\ell}$ etc.

The practice of summing (aggregating) over the three traffic levels assumes that the interaction of Traffic by Operational, Traffic by Procedural, and other similar interaction effects are not statistically significant.

Assume that the vector of means

$$\bar{\underset{\sim}{x}} = (\bar{x}_{11\cdots}, \bar{x}_{12\cdots}, \bar{x}_{21\cdots}, \bar{x}_{22\cdots})$$

is an observation on a four-variate multinormal random variable with mean vector

$$\underset{\sim}{\mu} = (\mu_{11\cdots}, \mu_{12\cdots}, \mu_{21\cdots}, \mu_{22\cdots})$$

and symmetric positive definite covariance matrix

$$\frac{1}{N}\Sigma = \frac{1}{4}\begin{bmatrix} \sigma_{11} & \sigma_{12} & \sigma_{13} & \sigma_{14} \\ \sigma_{12} & \sigma_{22} & \sigma_{23} & \sigma_{24} \\ \sigma_{13} & \sigma_{23} & \sigma_{33} & \sigma_{34} \\ \sigma_{14} & \sigma_{24} & \sigma_{34} & \sigma_{44} \end{bmatrix}$$

$N = 4$ is the number of controller teams, while the elements σ_{ij} are the variances and covariances of the sums

$$x_{ij1\ell} + x_{ij2\ell} - x_{ij3\ell}$$

of Team ℓ summed across the three traffic levels. It is assumed that these variances and covariances are the same for each team

From the elements of μ we can form $\binom{4}{2} = 6$ distinct mean differences. These differences and their estimates are shown in Table 4.

Table 4. DISTINCT MEAN DIFFERENCES AND THEIR SAMPLE ESTIMATES

Population Mean	Sample Estimate
$\delta_1 = \mu_{11\cdot\cdot} - \mu_{12\cdot\cdot}$	$\bar{d}_1 = \bar{x}_{11\cdot\cdot} - \bar{x}_{12\cdot\cdot}$
$\delta_2 = \mu_{11\cdot\cdot} - \mu_{21\cdot\cdot}$	$\bar{d}_2 = \bar{x}_{11\cdot\cdot} - \bar{x}_{21\cdot\cdot}$
$\delta_3 = \mu_{11\cdot\cdot} - \mu_{22\cdot\cdot}$	$\bar{d}_3 = \bar{x}_{11\cdot\cdot} - \bar{x}_{22\cdot\cdot}$
$\delta_4 = \mu_{12\cdot\cdot} - \mu_{21\cdot\cdot}$	$\bar{d}_4 = \bar{x}_{12\cdot\cdot} - \bar{x}_{21\cdot\cdot}$
$\delta_5 = \mu_{12\cdot\cdot} - \mu_{22\cdot\cdot}$	$\bar{d}_5 = \bar{x}_{12\cdot\cdot} - \bar{x}_{22\cdot\cdot}$
$\delta_6 = \mu_{21\cdot\cdot} - \mu_{22\cdot\cdot}$	$\bar{d}_6 = \bar{x}_{21\cdot\cdot} - \bar{x}_{22\cdot\cdot}$

In order to test for homogeneity of means, test simultaneously the six null hypotheses:

$$H_o: \delta_i = 0, \quad i = 1, \ldots, 6,$$

of no treatment combination mean difference at some overall level α (e.g., 0.05 or 0.01). Individual tests will not satisfy the maintenance of a controlled significance level α because in making such multiple tests, the effective overall significance level is raised. Furthermore, the possible

dependencies of the test statistics also obscure the true error rate. These difficulties for a set of hypotheses such as the six suggested by the δ_l can be overcome by constructing confidence intervals from a theorem in probability known as Bonferroni's inequalities (Feller 1957, p. 100). Details of its application to simultaneous statistical inference have been described by Miller (1966). Essentially, for a set of confidence intervals on m means with confidence coefficient at least $1 - \alpha$, one uses the usual intervals with the distribution critical value $t_{\alpha/2;n}$ replaced by $t_{\alpha/2m;n}$. In our case we would replace $\alpha = 0.05$ by 0.00833 if we wished a probability of covering all six δ_l in excess of 0.95. The general Bonferroni interval is

$$\bar{d}_h - t_{\alpha/12;n} s_{\bar{d}_h} \leq \delta_h \leq \bar{d}_h + t_{\alpha/12;n} s_{\bar{d}_h}.$$

where

$$\bar{d}_h = \frac{1}{4} \sum_{\ell=1}^{4} \left(\frac{d_{h\ell}}{3} \right) \; ; h = 1, 2.., 6$$

$d_{h\ell}$ is the hth mean difference between operational and procedural combinations for team ℓ. h = 1, ..., 6

$$d_{1\ell} = x_{111\ell} - x_{121\ell} + x_{112\ell} - x_{122\ell} + x_{113\ell} - x_{123\ell}$$

likewise for $d_{2\ell}$, $d_{3\ell}$ ...., $d_{6\ell}$ in the general Bonferroni interval.

Here

n = degrees of freedom = number of teams (4) minus 1 = 3

and

$s_{\bar{d}_h}$ = standard error of $\bar{d}_h$ h = 1, 2..6

i.e. the standard error of the hth operational procedural combination

= $\frac{1}{\sqrt{N}}$ (times the standard deviation of the differences of the team scores for the hth combination).

$$s_{\bar{d}_h} = \frac{1}{\sqrt{N}} \sqrt{\frac{1}{N-1} \left[\sum_{\ell=1}^{4} \left(\frac{d_{h\ell}}{3} \right)^2 - \frac{1}{N} \left(\sum_{\ell=1}^{4} \frac{d_{h\ell}}{3} \right)^2 \right]}$$

For N = 4

$$s_{\bar{d}_h} = \frac{1}{2}\sqrt{\frac{1}{3}\left[\sum_{\ell=1}^{4}\frac{\left(d_{h\ell}\right)^2}{9} - \frac{1}{4}\frac{\left(\sum_{\ell=1}^{4} d_{h\ell}\right)^2}{9}\right]}, \quad h = 1, 2 \ldots, 6$$

where $d_{h\ell}$ = hth mean difference between operational and procedural combinations for team ℓ; h = 1, 2, ... 6.

For n = 3 degrees of freedom we have the following critical values (where n = number of teams) (by linear interpolation in tables of the t distribution):

α	$t_{\alpha/12;3}$	$t_{\alpha/2;3}$
0.05	6.23	3.18
0.01	12. (approximate)	5.84

For the ranking of the four combination means, proceed in this manner:

(1) Arrange the four combination treatment means in ascending order:

$$\bar{x}_{ij(\min)} \leq \ldots \leq \bar{x}_{ij(\max)}$$

or, introducing a new subscript notation [i] for the ith smallest mean,

$$\bar{x}_{[1]} \leq \bar{x}_{[2]} \leq \bar{x}_{[3]} \leq \bar{x}_{[4]}.$$

(2) Compute the standard deviations of the six pairs of differences.

(3) Compute the six confidence intervals for the distinct population differences $\delta_1, \ldots \delta_6$. If the jth confidence interval does not contain zero, reject the null hypothesis

$$H_o: \delta_j = 0 \text{ for all } j, \; j = 1, \ldots, 6$$

and conclude that the means of the two combinations in δ_j are different.

(4) Draw horizontal lines under each pair of means $x_{[i]}$, $x_{[j]}$ whose differences are not significantly different. If a mean is statistically different from each of the other three, denote this by a single horizontal line under it. A perfectly unique ranking would have no joining lines. If the smallest two means were not significantly different, yet the remaining means were significantly different from each other and the first two means, the configuration would be

$$\underline{\bar{x}_{[1]}\ \bar{x}_{[2]}}\ \underline{\bar{x}_{[3]}\ \bar{x}_{[4]}}.$$

It should be emphasized that the multiple tests may not lead to a distinct ranking. For example the following ranking is not distinct.

$$\underline{\bar{x}_{[1]}\ \bar{x}_{[2]}}\ \underline{\bar{x}_{[3]}\ \bar{x}_{[4]}}$$

3.6.3 Multiple Comparisons within a Traffic Level

The method of paragraph 3.6.1 can be used to determine if significant Operational-Procedural-Traffic level interactions exist. If these interactions are significant, rankings of operational procedural combinations must be made within a traffic level. The method is the same as that used in 3.6.2 to rank the four $O_i\ P_j$

combination means. The degrees of freedom and critical values remain unchanged, but the various differences and means must be appropriately redefined. For example, the combination means are:

$$\bar{x}_{11k\cdot} = \frac{1}{4} \sum_{\ell=1}^{4} x_{11k\ell}; \; k = 1, 2, 3$$

.

.

.

$$\bar{x}_{22k\cdot} = \frac{1}{4} \sum_{\ell=1}^{4} x_{22k\ell}; \; k = 1, 2, 3$$

To test the multiple hypothesis of no difference between the population means of the combinations O_1P_1 and O_1P_2 at Traffic level 1, we calculate the four differences

$$e_{111} = x_{1111} - x_{1211}$$

.

.

.

$$e_{114} = x_{1114} - x_{1214}$$

their mean

$$\bar{e}_{11\cdot} = \bar{x}_{111\cdot} - \bar{x}_{121\cdot}$$

and the standard error of that mean as in the previous section with N-1 and $x_i = e_{hk\ell}$

$$s_{\bar{e}_{hk}} = \frac{1}{2} \sqrt{\frac{1}{3} \left[\sum_{\ell=1}^{4} e_{hk\ell}^2 - \frac{1}{4} \left(\sum_{\ell=1}^{4} e_{hk\ell} \right)^2 \right]} \quad h = 1, 2, \ldots, 6; \; k = 1, 2, 3$$

The test and the subsequent ranking of the four combinations is as described before.

3.6.4 Multiple Comparisons and Ranking of the Twelve Traffic-Operational-Procedural Combinations

In order to eliminate interaction effects, it is theoretically possible to rank the twelve means of the complete experimental design, but since this leads to $\binom{12}{2} = 66$ pairs, the effect on the t critical value is such that the resulting tests would be inordinately conservative. A more reasonable approach might consist of a ranking of the Traffic levels, followed by rankings within each Traffic level.

In the same notation, the three Traffic level means are

$$\bar{x}_{..1.} = \frac{1}{16} \sum_{\ell=1}^{4} R_{1\ell}$$

$$\bar{x}_{..2.} = \frac{1}{16} \sum_{\ell=1}^{4} R_{2\ell}$$

$$\bar{x}_{..3.} = \frac{1}{16} \sum_{\ell=1}^{4} R_{3\ell}$$

where $R_{k\ell}$ = sums of observations on operational and procedural combinations for the kth traffic level and ℓth team.

$$R_{k\ell} = \sum_{j=1}^{2} \sum_{i=1}^{2} x_{ijk\ell} = x_{..k\ell}$$

These means can be ranked by computing the Bonferroni simultaneous confidence intervals [ ] on their mean differences

$$\bar{u}_1 = \bar{x}_{..1.} - \bar{x}_{..2.}$$

$$\bar{u}_2 = \bar{x}_{..2.} - \bar{x}_{..3.}$$

$$\bar{u}_3 = \bar{x}_{..1.} - \bar{x}_{..3.}$$

Here,

$$\bar{u}_3 = \frac{1}{16} \left[\sum_{1}^{4} R_{1\ell} - \sum_{1}^{4} R_{3\ell} \right] = \frac{1}{16} \sum_{1}^{4} (R_{1\ell} - R_{3\ell})$$

If we denote the successive level differences for the teams by the u_{ij} symbol used in the matrix of level differences and introduce the new differences

$$u_{3\ell} = R_{1\ell} - R_{3\ell}; \ell = 1, \ldots, 4,$$

then letting

$$\bar{u}_3 = \frac{1}{16} \sum_{\ell=1}^{4} u_{3\ell} = \sum_{\ell=1}^{4} \frac{(u_{3\ell})}{16}$$

apply:

$$s_{\bar{x}} = \frac{1}{\sqrt{N}} \sqrt{\frac{1}{N-1} \left[\Sigma x_i^2 - \frac{(\Sigma x_i)^2}{N} \right]} .$$

For

$$N = 4, \text{ let } x = \frac{u_{3\ell}}{4}$$

The three standard errors of the level mean differences can be written as

$$s_{\bar{u}_1} = \frac{1}{8} \sqrt{\frac{1}{3} \left[\sum_1^4 u_{1\ell}^2 - \frac{1}{4} \left(\sum_1^4 u_{1\ell} \right)^2 \right]},$$

$$s_{\bar{u}_2} = \frac{1}{8} \sqrt{\frac{1}{3} \left[\sum_1^4 u_{2\ell}^2 - \frac{1}{4} \left(\sum_1^4 u_{2\ell} \right)^2 \right]},$$

$$s_{\bar{u}_3} = \frac{1}{8} \sqrt{\frac{1}{3} \left[\sum_1^4 u_{3\ell}^2 - \frac{1}{4} \left(\sum_1^4 u_{3\ell} \right)^2 \right]},$$

Since there are three mean differences the critical value is $t_{\alpha/6;3}$:

α	$t_{\alpha/6;3}$	
0.05	4.87	(linear interpolation)
0.01	8.6	(linear extrapolation)

For example, the 95% confidence interval for the population mean μ_1 of $\bar{u}_1$ is

$$\bar{u}_1 - 4.87\, s_{\bar{u}_1} \leq \mu_1 \leq \bar{u}_1 + 4.87\, s_{\bar{u}_1}$$

If such an interval contains zero we would accept the hypothesis

$$H_o: \mu_1 = 0 \text{ for each } i;\ i = 1, 2, 3$$

and conclude that the Traffic level grand means are not significantly different. We would rank the Traffic level means in the same manner as in paragraph 3.6.2. A distinct rank is assigned to those means that are statistically significantly different from the other two means.

3.6.5 Justification of the Bonferroni Inequalities Approach

The usual multiple comparison method for repeated measurements data is one commonly due to Scheffe´ (1959). When the number of multiple tests is large, the Scheffe´ confidence intervals will be shorter on the average than those given by Bonferroni's inequalities, provided the sample size is moderate to large. Since our sample of N = 4 teams is exceptionally small, and we are only interested in a few comparisons of the means, the Bonferroni approach is decidedly preferable.

3.6.6 Final Ranking

The system effectiveness measures, listed in paragraph 3.6.2, are the X_{ijki} parameters used in the preceding discussion. Operational procedural combination rankings will be performed for each effectiveness measure separately. The final ranking will be made by summing the effectiveness measure ranks for each operational-procedural combination. The combinations will then be ranked according to lowest rank sum to determine the best operational-procedural combination.

3.7 MATHEMATICAL MODEL AND CONTROLLER TEAM PERFORMANCE

3.7.1 Objective

The objective of this test is to determine when the SAFOC controllers have reached an asymptotic limit in learning the SAFOC controls. This will be performed by measuring how well each of the four teams performs. This test will be repeated to determine the learning curve of operators performing similar tasks. The effectiveness measures listed in paragraph 3.5.2 will be used for controller performance evaluation.

It is possible to rank the controllers on some measurable characteristics. For this ranking the C teams in R replications using different but equivalent scenarios will be used. In essence, random trials of the controllers' ability to handle repeated scenarios will be performed to determine whether the controllers are significantly different in their abilities. If they are different, they will be ranked in order of their abilities to determine those needing additional training.

If new personnel must be introduced during the tests, attempts must be made to train them to the approximate proficiency levels of the experienced controllers for the results to have significance.

3.7.2 Statistical Factorial Design Mathematical Model

Table 5 demonstrates the symbology of the effects (effectiveness measures) resulting from replications of different scenarios by each controller and indicates the sums to be performed. Table 6 indicates the usual analysis of variance for a two-way classification based on a mixed-model of fixed teams and random samples from the hypothetical population of replication observations.

When the scenarios are run, the data (such as the workload times of each controller) consisting of the X_{ij}'s filled in on Table 5 is operated on by performing the column and row sums, followed by the operations called out in Table 6. Then the sums of squares and mean squares called for in Table 6 are computed. The F ratio is computed:

$$(1) \quad F = \frac{S_2 (R - 1)}{S_3}$$

where S_2, S_3 and R are defined in Table 6. Reject the null hypothesis of no team difference if (1) exceeds the α point of the F distribution with (C-1) and (R-1) (C-1) degrees of freedom. For example, if α equals 0.05, C equals four and R equals two, then the critical value is:

Table 5. Effects of Replications

TEAMS

Replications	1	2	.	.	C	Row Sum
1	X_{11}	X_{12}	.	.	X_{1C}	$\sum_{j}^{C} X_{1j} = C\bar{X}_{1.}$
2	X_{21}	X_{22}	.	.	X_{2C}	$\sum_{j}^{C} X_{2j} = C\bar{X}_{2.}$
.	.	.	.	.	.	
.	.	.	.	.	.	
R	X_{R1}	X_{R2}	.	.	X_{RC}	$\sum_{j}^{C} X_{Rj} = C\bar{X}_{R.}$
Column Sum	$R\bar{X}_{.1} = \sum_{i}^{R} X_{i1}$	$\sum_{i}^{R} X_{i2} = R\bar{X}_{.2}$			$R\bar{X}_{.C} = \sum_{i}^{R} X_{iC}$	$\sum_{i}^{R}\sum_{j}^{C} X_{ij} = RC\bar{X}_{..}$

C = Number of teams
R = Number of replications

Table 6. Analysis of Variance for a Two-Way Classification

SOURCE	SUMS OF SQUARES	DEGREES OF FREEDOM	MEAN SQUARE
Replications	$S_1 = \sum_{i}^{R} (\bar{X}_{i.} - \bar{X}_{..})^2$	R-1	$S_1/(R-1)$
Teams	$S_2 = \sum_{j}^{C} (\bar{X}_{.j} - \bar{X}_{..})^2$	C-1	$S_2/(C-1)$
Error	$S_3 = \sum_{i}^{R}\sum_{j}^{C} (X_{ij} - \bar{X}_{i.} - \bar{X}_{.j} + \bar{X}_{..})^2$	(R-1) (C-1)	$S_3/(R-1)(C-1)$
TOTALS	$S_4 = \sum_{i}^{R}\sum_{j}^{C} (X_{ij} - \bar{X}_{..})^2$	RC-1	---

(2) $F_{0,\ 05,\ 3,\ 3} = 9.28$

If the computed value of F was larger than 9.28 one could say that the controllers performed their tasks with significantly different capabilities and that similar tests would repeat this conclusion with a 5% risk of being in error.

If we wish to rank each of the teams with respect to every other team, Tukey's multiple comparison procedure can be used. This consists of taking the difference between every pair of team performance means computed during the test:

(3) $B_1 - B_2 = Z_1$

$B_1 - B_3 = Z_2$

$B_1 - B_4 = Z_3$

$B_2 - B_3 = Z_4$

$B_3 - B_4 = Z_6$

where B_ℓ is the mean workload performance parameter of ℓth team

$$B_\ell = \frac{1}{R} \sum_{i}^{R} X_{i\ell}$$

We then compute:

(4) $$Z_1 - \frac{S}{\sqrt{2}} q_\alpha \leq \delta \leq Z_1 + \frac{S}{\sqrt{2}} q_\alpha$$

where $S = \sqrt{S_3 / (C-1)(R-1)}$.

q_α = critical value for the Studentized Range.

The value of q_α is extracted from the indicated table at some value of α (say 0.05) and for the error degrees of freedom (3 in the example). If (4) contains zero (that is, if the left hand term is negative and the right hand term is positive) then the difference Z_1 is not significant. This is performed for the remaining Z's and rankings are achieved on the basis of the relative values of (4). For example,

(5) $\underline{B_2\ B_1}\ \underline{\underline{B_4\ B_3}}$

where the underlines might indicate no significance between differences in B_2 and B_1 but significant differences in B_2 and B_4. We would then conclude that Team 2 is not significantly better than Team 1 (at the 5% level of significance) but that the difference in performance between Team 2 and Team 4 is significant. This result could be used to indicate that Teams 3 and 4 needed additional training to bring them up to Team 2 and Team 1.

For computational purposes, the row and column sums will be summed and then squared, rather than summing the squares as shown by the operations in Table 6:

$$S_1 = R \frac{\sum_i^R (\sum_j^C X_{ij})^2 - (\sum_i^R \sum_j^C X_{ij})^2}{RC}$$

$$S_2 = C \frac{\sum_j^C (\sum_i^R X_{ij})^2 - (\sum_i^R \sum_j^C X_{ij})^2}{RC}$$

$$S_3 = S_4 - S_1 - S_2$$

$$S_4 = RC \frac{\sum_i^R \sum_j^C X_{ij}^2 - (\sum_i^R \sum_j^C X_{ij})^2}{RC}$$

3.7.3 Criteria

a. INDEPENDENT VARIABLES

The independent variables for the controller team comparison test are:

1. Traffic Density Level

The total number of flights simultaneously in the SAFOC control area.

2. Replications

The repetition of similar (but not necessarily identical) scenarios.

3. Controllers

Tests will be performed with four teams of two controllers each.

b. DEPENDENT VARIABLES

1. Safety

2. Controller Workload

3. Communications Workload

4. Delays

5. Throughput

6. Capacity

7. Uncontrolled Time

c. CONSTANTS

The same map will be used by all teams and for all replications. Altitudes and other flight plan data will be held constant. Mix of aircraft speeds, aircraft types, aircraft equipment will be constant.

The same number of scenario replications will be played to each team. Similar scenarios will be used.

A fixed traffic level will be used in this experiment.

APPENDIX A
NOTATION CONVENTIONS

For subscripted variables such as $x_{ijk\ell}$ the following conventions apply:

If a subscripted variable is summed over the entire range of a subscript, the result is denoted by the subscripted variable with a dot replacing the index over which the sum has been taken.

For example:

$$X_{.jk\ell} = \sum_{i=1}^{R} X_{ijk\ell}$$

$$X_{i.k\ell} = \sum_{j=1}^{N} X_{ijk\ell}$$

$$X.... = \sum_{\ell=1}^{P} \sum_{k=1}^{M} \sum_{j=1}^{N} \sum_{i=1}^{R} X_{ijk\ell}$$

If a variable is to be averaged, a bar appears over the variable

thus,

$$\bar{X}_{i.k\ell} = \frac{1}{N} X_{i.k\ell} = \frac{1}{N} \sum_{j=1}^{N} X_{ijk\ell}$$

or

$$\bar{X}.... = \frac{1}{PMNR} X.... = \frac{1}{PMNR} \sum_{\ell=1}^{P} \sum_{k=1}^{M} \sum_{j=1}^{N} \sum_{i=1}^{R} X_{ijk\ell}$$

APPENDIX B
DEFINITION OF SYMBOLS

Symbol	Definition
B_ℓ: $B_1, \ldots, B_4$	Performance Parameters for four teams
	$B_\ell = \frac{1}{R} \sum_{i=1}^{R} X_{i\ell}$ $\ell = 1, 2, 3, 4$
	B_ℓ - mean workload performance parameter of ℓ th team
C	Number of columns (teams)
$\bar{D}_{12}$	Mean difference $= \frac{1}{4} \sum_{i=i}^{4} u_{il}$
$\bar{d}_h$: $\bar{d}_1, \ldots, \bar{d}_6$	Mean differences $\bar{d}_1 = \bar{x}_{11..} - \bar{x}_{12..}$
	Note that $\bar{d}_h = \frac{1}{12} \sum_{\ell=1}^{4} d_{h\ell}$
$d_{h\ell}$	$h^{\underline{th}}$ mean difference between operational and procedural combinations for team ℓ . h = 1, ... , 6
$e_{hk\ell}$: $e_{111}, \ldots, e_{634}$	h^{th} difference among the four Operational and Procedural combination observations for team ℓ at traffic level k:
$\bar{e}$	Mean of operational and procedural differences averaged over the teams.
F	Ratio used in statistical tests
h	Index indicating one of six mean differences
H_o	Null hypotheses

Symbol	Definition
i = 1, 2 Operational modes	
j = 1, 2 Procedural modes	
k = 1, 2, 3 Traffic levels	
ℓ = 1, 2, 3, 4 Teams	
m	Number of means for which simultaneous confidence intervals are to be computed
N	Number of events or measures
n	Degrees of freedom (no. of teams - 1)
O_1	Operational method with two sectors
O_2	Operational method with one sector
P_1	Procedural method-normal
P_2	Procedural method-passive
O_1P_2, O_1P_j	Operational procedural combinations
q_α	Critical value for the Studentized Range
$R_{k\ell}$	Row sums of observations on the Operational and Procedural mode combinations for the kth Traffic level and ℓth team. $R_{k\ell} = \sum_{j=1}^{2} \sum_{i-1}^{2} X_{ijk\ell}$
S_{ijk}	Column sums of the basic data matrix (observations under Procedural mode i, Operational mode j, and Traffic level k summed over the four teams). $S_{ijk} = \sum_{\ell-1}^{4} X_{ijk\ell}$

Symbol	Definition
$s_{\bar{d}_h}$	Standard Error of the h̲th Operational Procedural Combination Mean Difference: $s_{\bar{d}_h} = \frac{1}{6}\sqrt{\frac{1}{3}\left(\sum_{\ell=1}^{4} d_{h\ell}^2 - \frac{1}{4}\left(\sum_{\ell=1}^{4} d_h\right)^2\right)}$ $h = 1, \ldots, 6$
$s_{\bar{e}_{hk}}$	Standard error of the h̲th Operational Procedural Combination Mean Difference for Traffic level k: $s_{\bar{e}_{hk}} = \frac{1}{2}\sqrt{\frac{1}{3}\left(\sum_{\ell=1}^{4} e_{hk\ell}^2 - \frac{1}{4}\left(\sum_{\ell=1}^{4} e_{hk\ell}\right)^2\right)}$ $h = 1, \ldots, 6$ $k = 1, 2, 3$
$s_{\bar{u}_g}$	Standard error of the gth traffic level combination mean difference: $s_{\bar{u}_g} = \frac{1}{8}\sqrt{\frac{1}{3}\left(\sum_{i-1}^{4} u_{g\ell}^2 - \frac{1}{4}\left(\sum_{\ell=1}^{4} u_{g\ell}\right)^2\right)}$ $g = 1, 2, 3$
$s^2_{\bar{x}}$	Standard error of mean of x
$S_1 \ldots S_4$	Sums of Squares
S	Root-mean-square error
T^2	Statistic used for testing hypotheses
t̲	Critical value for statistical test
$t_{\alpha/6;3}$	$100\,\alpha/6\%$ critical value of Student-Fisher distribution with 3 degrees of freedom
u_{ij}	Symbol used to denote the successive level differences

$\bar{u}_i$	Average of level differences (u_{ij} averaged over j)
$X_{ijk\ell}$	is some effectiveness measure (e.g., workload)
$x_{ijk\ell}$	Observed value of variable X for Operation mode i, Procedural mode j, Traffic levek k, and Team ℓ.
	i = 1, 2 j = 1, 2 k = 1, 2, 3 ℓ = 1, 3, 4
$\bar{x}_{ij}$	Combination treatment mean for ith operational and jth procedural combination.
$\bar{x}_{[1]} \cdots \bar{x}_{[4]}$	$\bar{x}_{[1]}\ \bar{x}_{[2]} \cdots \bar{x}_{[m]}$
	Ranks - ordered set of means
$\underset{\sim}{\bar{x}}$	is an observation on a four-variate multinormal random variable with mean vector
	$\underset{\sim}{\bar{x}} = (\bar{x}_{11}\cdot\cdot, \bar{x}_{12}\cdot\cdot, \bar{x}_{21}\cdot\cdot, \bar{x}_{22}\cdot\cdot)$
$\bar{x}_{\cdot\cdot k\ell}$	the 4 x 3 matrix of team sums for each traffic level
	$\bar{x}_{\cdot\cdot k\ell} = \sum_{i=1}^{2} \sum_{j=1}^{2} x_{ijk\ell}$
$\bar{x}_{\cdot\cdot 1\cdot}$	Traffic level mean
	$\bar{x}_{\cdot\cdot 1\cdot} = \frac{1}{16} \sum_{\ell=1}^{4} R_{1\ell}$
$X_{R1}, X_{R2} \cdots X_{RC}$	The data (such as the workload times of each controller) consisting of the X_{ij}'s
$\bar{X}_R$	Mean performance measure of Rth replication (averaged over C teams)
X_{11}	Replication one for team 1 of performance measure for controller team evaluation

Symbol	Definition
$Z_1, Z_2, \ldots Z_6$	Differences in team performance means
δ_i	Operational-Procedural Combination Differences (population mean)
δ_1	$\delta_1 = E(\bar{d}_1) = \mu_{11..} - \mu_{12..}$
δ_2	$\delta_2 = E(\bar{d}_2) = \mu_{11..} - \mu_{21..}$
δ_3	$\delta_3 = E(\bar{d}_3) = \mu_{11..} - \mu_{22..}$
δ_4	$\delta_4 = E(\bar{d}_4) = \mu_{12..} - \mu_{21..}$
δ_5	$\delta_5 = E(\bar{d}_5) = \mu_{12..} - \mu_{22..}$
δ_6	$\delta_6 = E(\bar{d}_6) = \mu_{21..} - \mu_{22..}$
$\underset{\sim}{\mu}$	Mean four-variate multinormal random vector
	$= (\mu_{11\cdots}, \mu_{12\cdots}, \mu_{21\cdots}, \mu_{22\cdots})$
$\mu_{11\cdots}, \mu_{22\cdots}$	Components of $\underset{\sim}{\mu}$
μ_1	Population mean
α	Controlled significance level
σ_{ij}	ij element of covariance matrix

APPENDIX C

REFERENCES

1. "Phase I Test Plan for a Semiautomatic Flight Operations Center (SAFOC) (Design of Experiments Program)," 1 July 1970, Contract No. DAAB07-69-C-0040, American Electronic Laboratories.

2. "System Description and Operators Manual for Semiautomatic Flight Operations Center (SAFOC) (U)" 14 Mar 69, Contract DAAB07-68-C-0354, Hughes Aircraft Co.

3. R. Steele and J. Torrie (1960), "Principles and Procedures of Statistics", McGraw Hill Co., N.Y., P. 109-110.

4. Feller, W. (1957). "An Introduction to Probability Theory and its Applications," John Wiley & Sons, Inc., New York.

5. Miller, R. G., Jr. (1966). "Simultaneous Statistical Inference," McGraw-Hill Book Company, New York.

6. Morrison, D. F. (1967). "Multivariate Statistical Methods," McGraw-Hill Book Company, New York.

7. Scheffé, H. (1959). "The Analysis of Variance," John Wiley and Sons, Inc., New York.

PART II

NEW ANALYSES AND METHODS LEADING TO IMPROVED TARGET ACQUISITION REQUIREMENTS INVOLVING SYSTEMS, GEODETIC AND RE-ENTRY ERRORS, AND INCREASED WEAPONS EFFECTIVENESS FOR CONVENTIONAL WEAPONS

Hans Baussus-von Luetzow
U. S. Army Engineer Topographic Laboratories
Fort Belvoir, Virginia

SUMMARY. The paper represents a supplemental analysis for height bursts as well as vertical target location errors, considering flat and contoured terrain, and thus completes the development of optimal methods for weapons research and development and a broad spectrum of requirement analyses.

FORWORD. This study, appearing as Part II under the same title as a prior one presented at the Fifteenth Conference on the Design of Experiments in Army Research, Development and Testing, in October 1969, completes the rigorous treatment through extension to three geometrical dimensions of the fragment damage problem pertaining to tube and missile artillery. It may be followed by a similar analysis for conventional cratering and for nuclear weapons. As emphasized before, implementation of the methods and concepts developed would undoubtedly lead to a significant increase of Army weapons effectiveness. In addition, the new methods are expected to have some ramifications pertaining to a variety of research and development and combat development activities. The technical responsibility for this study is exclusively that of the author, who appreciates the U. S. Army Engineer Topographic Laboratories' continued interest in this kind of effort.

CONTENTS

The remainder of this article has been reproduced photographically from the author's manuscript.

NEW ANALYSES AND METHODS LEADING TO IMPROVED TARGET ACQUISITION REQUIREMENTS INVOLVING SYSTEMS, GEODETIC AND RE-ENTRY ERRORS, AND INCREASED WEAPONS EFFECTIVENESS FOR CONVENTIONAL WEAPONS

1. Introduction. The author has shown in a prior paper, referenced as R70-2,[1] that the artillery problem in two geometrical dimensions can be formulated in terms of four-dimensional integrals and that optimization for multiple volleys is possible. In this way, two-dimensional weapons dispersion, meteorological errors (winds, air densities), target location errors (including identification, intelligence, map), size, shape, and target distribution (density) can be considered simultaneously, i.e., an integrated effect in terms of expected casualty results as a function of multiple volleys.

For three geometrical dimensions, it is necessary to utilize the burst height distribution parameters for proximity and time fuzes (for a constant range and impact angle) and the vertical target-location error (mean height error of target distribution) and their associated statistical distributions, assumed to be normal.

Fortunately, the former analytical approach can be systematically extended to six dimensions, and respective optimization conditions can be written.

In order to simplify the problem and permit evaluation of improvements as well as sensitivity analyses, the target may be assumed to be circular and homogeneous and referenced to a center of gravity under consideration of target-density indices.

If the target mean height variance is small, the optimization pattern for n volleys to be invariant with respect to a rotation is practically two-dimensional. In case of greater mean height variances, competitive three-dimensional optimization configurations arise, although they are generally restricted to the center of gravity, e.g., three competitive aiming heights. Comparison of numerical optimization results easily permits establishment of the correct ranking order.

Inclusion of the shape of the terrain does not provide any difficulties and is provided for in the general formulas, although, in practical situations, it can hardly be considered and its effect should rather be added to the target mean height variance.

1. New Analyses and Methods Leading to Improved Target Acquisition Requirements Involving Systems, Geodetic and Re-entry Errors, and Increased Weapons Effectiveness for Conventional Weapons (Part I), ARO-D Report 70-2, Proc. of the Fifteenth Conf. on the Design of Experiments in Army Research, Development and Testing. The paper is also available as USAETL Research Note No. 35, AD 703923.

It is clear that the inclusion of two additional dispersions results in lower mean casualties or requires more volleys, respectively. For this reason, the reduction of target identification and intelligence errors by direct target acquisition, including relatively accurate location of controlled weapon delivery systems and forward observers and the utilization of the contemplated Position and Azimuth Determining System (PADS) and the Long Range Positioning Determining System (LRPDS), is of importance. For longer ranges and under a variety of conditions, these systems might not be available on a timely basis or might not function properly, so that optimal methods for multiple volleys to be incorporated in TACFIRE are indispensable. It follows also from marginal utility considerations, as outlined in R70-2, that four or more volleys to be fired in optimal patterns are advantageous for shorter ranges, where effectiveness is relatively high. In this way, the problem of map accuracy per se, at least with respect to class 1 and 2 maps, can be completely resolved. Fortunately, target mean height inaccuracies, often envisioned at longer ranges despite relatively precise determination of horizontal coordinates, can also be overcome by optimal volley patterns.

Based upon strategic consideration, it appears that a combination of improved positioning/target-acquisition devices or systems and optimal multiple volley fire based on a set of input descriptors is particularly cost effective in case of a favorable force ratio and cost effective as well as decisive in case of a force ratio of about one. Such a combination can, of course, not compensate for a highly unfavorable force ratio. Furthermore, such a combination is expected to be most effective in small- to medium-intensity conflict.

Since optimal multiple volleys are highly effective—when there is ample ammunition supply and strong compensation for errors of all kinds—and cost effective when ammunition is limited, preparations for their incorporation in TACFIRE should start as soon as possible. The great advantage of optimal multiple volleys is their flexibility, pertaining to the combat situation, including availability of position and target-acquisition equipment and, last but not least, they eliminate the need for stringent, rigid, and often inachievable design specifications for new equipment.

2. Formal Extension to Include Burst Height and Vertical Target Location Variability. The definitions and nomenclature of R70-2 are maintained and extended.

As to weapon distribution parameters, we add the vertical standard deviation S_z.

In case of a proximity fuze, we have then the complete range variance

$$S_{r,p}^2 = S_r^2 + \cot n^2 \beta \, S_z^2 \qquad (1)$$

with β as trajectory impact angle (cf. Fig. 1 in R70-2).

The lateral variance S_ζ^2 remains unchanged.

If a time fuze is applied, we obtain

$$S_{z,T}^2 = v_E^2 \tan^2 \beta \operatorname{var} T \tag{2a}$$

$$S_{r,T}^2 = S_r^2 + v_E^2 \cot n^2 \beta \operatorname{var} T \tag{2b}$$

where v_E is the terminal projectile velocity, var T the time fuze variance, and $S_{r,T}^2$ the composite range variance. It should be emphasized again that all equations refer to a constant range.

The vertical target-location distribution (with reference to the mean height) is

$$f_{\bar{\zeta}}(\bar{\zeta}) = \frac{1}{\sqrt{2\pi}\,\sigma_{\bar{\zeta}}} \exp. \left(-\frac{1}{2}\frac{\bar{\zeta}^2}{\sigma_{\bar{\zeta}}^2}\right) d\bar{\zeta} \tag{3}$$

in supplementation of eq. (2) in R70-2.

The aiming point for the first volley now has the coordinates a_1, b_1, d_1 (the letter c has already been used as a configuration parameter).

After these preparations, and denoting the weapon distribution parameters for simplicity by S_r, S_ζ, S_z; eqs. (3) -- (8) of R70-2 can be generalized as follows:

For the first volley, we have four burst distributions designated by

$$\lambda_1 \left(x,y; a_1 + \frac{c}{2}, b_1 + \frac{c}{2}; S_r, S_\zeta\right) dx dy\, \delta\,(z; d_1; S_z)\, dz,$$

$$\lambda_2 \left(x,y; a_1 - \frac{c}{2}, b_1 + \frac{c}{2}; S_r, S_\zeta\right) dx dy\, \delta\,(z; d_1; S_z)\, dz, \text{ etc.} \tag{4}$$

where δ represents the supplemental normal height of burst distribution.

For sufficiently small volume elements $\Delta x \Delta y \Delta z$, we arrive then in integral form at an intermediate average probability of hitting T_1 at least once.

$$P_{il}^{1} = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\int_{0}^{\infty}\sum_{1}^{1} \lambda_{\nu} \cdot \delta \cdot P_{1}(x,y,z;\bar{\xi},\bar{\eta},\bar{\zeta};\xi_{1},\eta_{1},\zeta_{1})\,dx\,dy\,dz \tag{5}$$

and, since 0_3 obeys a distribution law, at

$$P_{1}^{1} = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\int_{0}^{\infty}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\sum_{1}^{1} \lambda_{\nu} \cdot \delta \cdot P_{1}(x,y,z;\bar{\xi},\bar{\eta},\bar{\zeta};\xi_{1},\eta_{1},\zeta_{1}) \cdot$$

$$\cdot f(\bar{\xi},\bar{\eta})\, f_{2}(\bar{\zeta})\,dx\,dy\,dz\,d\bar{\xi}\,d\bar{\eta}\,d\bar{\zeta} \tag{6}$$

For k volleys and p individual targets, we obtain the total expected casualty result

$$n_{1}\sum_{1}^{k} = \sum_{\tau=1}^{\tau=k}\sum_{\mu=1}^{\mu=N}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\int_{0}^{\infty}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\sum_{1}^{1} \lambda_{\nu}\,(x,y;a_{\tau},b_{\tau})\,\delta\,(z;d_{\tau})\,P_{1}(x,y,z;\bar{\xi},\bar{\eta},\bar{\zeta};\xi_{\mu},\eta_{\mu},\zeta_{\mu}) \cdot$$

$$\cdot f(\bar{\xi},\bar{\eta})\, f_{2}(\bar{\zeta})\,dx\,dy\,dz\,d\bar{\xi}\,d\bar{\eta}\,d\bar{\zeta} \tag{7}$$

The optimization conditions can be formulated as

$$\frac{\partial}{\partial a_{\tau}}\, n_{1}\sum_{1}^{k} = 0\,; \quad \frac{\partial}{\partial b_{\tau}}\, n_{1}\sum_{1}^{k} = 0\,; \quad \frac{\partial}{\partial d_{\tau}}\, n_{1}\sum_{1}^{k} = 0 \tag{8}$$

Optimal aiming coordinate triple sets have, of course, to be obtained by numerical methods and appropriate ranking.

For a circular and homogeneous (uniform) target distribution, the conditions (7) reduce to fewer equations, i.e., the respective aiming pattern, consisting of a set of k origins 0_2, would be invariant under a rotation about 0_1.

After the first volley, which has in many cases a surprise effect, a degradation with respect to P_1 can be expected which can be expressed as a transition from ground to air burst[2] and by an empirical reduction factor $R < 1.00$. Taking this into consideration, a more general result corresponding to eq. (7) reads as

2. Cf. Footnote 3 of R70-2.

$$n_1 \sum_{1}^{k} = \sum_{\mu^{(1)}=1}^{\mu^{(1)}=N} \int_{-\infty}^{+\infty} \int_{0}^{\infty} \int_{-\infty}^{+\infty} \sum_{\nu=1}^{\nu=l} \lambda_\nu \cdot \delta \cdot P_1(x,y,z;\bar{\xi},\bar{\eta},\bar{\zeta};\xi_{\mu^{(1)}},\eta_{\mu^{(1)}},\zeta_{\mu^{(1)}}) \cdot$$

$$\cdot f(\bar{\xi},\bar{\eta})\, f_2(\bar{\zeta})\, dx dy dz d\bar{\xi} d\bar{\eta} d\bar{\zeta}$$

$$+ \sum_{\tau=2}^{\tau=k} \sum_{\mu^{(2)}=1}^{\mu^{(2)}=N} R \int_{-\infty}^{+\infty} \int_{0}^{\infty} \int_{-\infty}^{+\infty} \sum_{\nu=1}^{\nu=l} \lambda_\nu \cdot \delta \cdot \pi_{\mu^{(2)}} \cdot P_1^{c}(x,y,z;\bar{\xi},\bar{\eta},\bar{\zeta};\xi_{\mu^{(2)}},\eta_{\mu^{(2)}},\zeta_{\mu^{(2)}}) \cdot$$

$$\cdot f(\bar{\xi},\bar{\eta})\, f_2(\bar{\zeta})\, dx dy dz d\bar{\xi} d\bar{\eta} d\bar{\zeta} \qquad (9)$$

It should be noted that the coupling term P_1 or P_1^{c}, respectively, in eqs. (7) - (9) is a nonfactorable function of six independent variables, so that the optimization problem is a composite one. For numerical evaluations, it is recommended to hold a point P(x,y,z) fixed, perform a summation with respect to μ with fixed $\bar{\xi}$, $\bar{\eta}$, $\bar{\zeta}$ coordinates, integrate over the $\bar{\xi}$, $\bar{\eta}$, $\bar{\zeta}$ domain, and thereafter integrate with respect to x, y, z. A numerical analysis and a subsequent computer program does not present formidable difficulties. The integration pertaining to z has to start at zero with an associated cumulative probability of the distribution δ (from minus infinity to zero).

3. Simplifications and Ramifications of Six-Dimensional Analytical Approach. A considerable simplification results if the target coordinates relate to a center of gravity easily computed through the use of standardized target areas and actual target-density descriptors (indices). Thereafter, more or less satisfactory homogeneity is assumed and, in general, a circular target area which may, however, vary in diameter.

As a further simplification, the terrain is assumed to be negligible, i.e., expressed only by an average height $\bar{d}$. An increased $S^2_{z,P}$ or $S^2_{z,T}$ would satisfactorily compensate herefore.

If $S_{z,P}$ or $S_{z,T}$ is relatively small, optimization is practically two-dimensional. As shown in R70-2, the respective patterns are invarient under a rotation.

In the two-dimensional case, we would thus have optimal coordinates a_1, b_1, $\bar{d}$; a_2, b_2, $\bar{d}$; a_3, b_3, $\bar{d}$, etc., where a_1, b_1 may also be 0, 0. Although the existence of two distributions involving the random variables z and $\bar{\zeta}$ is disregarded, as far as optimization is concerned, integration with respect to six variables is still necessary in order to obtain correct parameters for the casualty probability distributions, which depend

on the number of volleys k (for constant range and a given set of target descriptors such as size).

The three-dimensional optimization is required only in case of a greater vertical mean target height uncertainty, which is mostly associated with longer ranges, including missile applications.

Optimal configurations characterized by coordinate triples involving different d's might then appear to be competitive, such as 0, 0, d_1; 0, 0, d_2 or 0, 0, d_1; 0, 0, d_3; 0, 0, d_3; and, in exceptional circumstances, triples referring to the first optimization ring, such as $a_1^{(1)}$, $b_1^{(1)}$, $d_1^{(1)}$; $a_2^{(1)}$, $b_2^{(1)}$, $d_1^{(1)}$, ... and $a_1^{(1)}$, $b_1^{(1)}$, $d_2^{(1)}$; $a_2^{(1)}$, $b_2^{(1)}$, $d_2^{(1)}$; . . . . . Because of the symmetries involved, the optimization structure is well discernible and the numerical-computational effort reduced. It should be kept in mind that optimizations and evaluations involving the vertical geometrical dimensions depends on fragment damage tables for various heights of burst.

In conclusion, it should be emphasized that a considerable reduction of computational work is expected after a thorough numerical study has been conducted for various ranges, target sizes, number of volleys, etc., for a given weapon system. The results of such a study would yield much more insight into weapons effects and their possible optimization than hitherto available.

A MATHEMATICAL MODEL FOR ADJUSTMENT PHASE PROCEDURES AND MONTE CARLO INVESTIGATION OF THE EFFECT OF THESE PROCEDURES ON FIRE FOR EFFECT ROUNDS

Sidney Gerard
Reliability and Maintainability Division
U.S. Army Materiel Systems Analysis Agency
Aberdeen Research & Development Center
Aberdeen Proving Ground, Maryland

ABSTRACT

The development of a mathematical model for the adjustment phase procedures utilized by artillery and mortar groups and a Monte Carlo simulation of these adjustment phase procedures is presented herein. Investigations on the number of rounds required to accomplish the adjustment phase of the mission and the errors of the centers of impact of rounds fired for effect upon accomplishment of the adjustment phase are also presented in this paper. The above investigations are dependent on many factors the most important being the following:

1. First round aiming errors.
2. Observer to target distance.
3. The angle between the gun to target (GT) line and observer to target (OT) line.
4. Target size (registration area).
5. Round-to-round dispersions.
6. Observer's error in locating impact points of adjustment phase rounds.

Finally a sensitivity analysis has been conducted on these factors and results of the significant findings presented here.

Most of this paper was reproduced photographically from the manuscript supplied by the author.

INTRODUCTION

During a military engagement (battle) where artillery, mortar or nuclear weapon systems are being employed there are usually four basic methods of accuracy fire employed. These are (1) predicted, (2) meteorological plus velocity errors, (3) adjusted and (4) K transfer (includes adjusted fire) fire. However, prior to use of these weapon systems, a target must be located and a suitable elevation and azimuth determined which will enable rounds to be fired effectively on the target. For approximately 75% of the accuracy fire missions the determination of this elevation and azimuth is accomplished by adjustment procedures. Adjusted fire is the method of fire in which a number of rounds are expended and through successive corrections the suitable elevation and azimuth are obtained. These corrections are directed by an observer usually forward of the gun and set off at an angle less than $\pm$ 500 mils from the gun to target (GT) line of fire.

This paper presents the development of a mathematical model and a Monte Carlo simulation of these adjustment phase procedures. Errors in the center of impact of rounds fired for effect and the number of rounds expended in the adjustment phase of the mission are obtained from the results of the simulation.

ACCOMPLISHMENT OF THE ADJUSTMENT PHASE

The basic objective of all weapon systems is to fire rounds effectively on a designated target. The adjustment phase is the part of an adjusted fire mission in which a number of rounds are expended and through successive corrections a suitable elevation and azimuth are obtained to achieve this objective. Questions posed concerning the use of this technique of fire are: (1) what is the expected number of rounds expended during the adjustment phase and (2) what determines the final suitable elevation and azimuth. The adjustment phase is complete and the above questions answered upon fulfillment of any of the three conditions below.

1. The first or any subsequent round during the adjustment phase is observed to impact on the target registration area which is defined by a predetermined radius, R_t, from the center of the target.

2. A y (usually 50, 100, or 200) meter bracket has been split with reference to the observer to target (OT) line. Splitting a y meter bracket means that two consecutive rounds impact within y meters of each other and bracket (that is one round impacting short and other beyond the target center) the center of the target. However, one-half the indicated correction is applied before firing rounds for effect upon the target.

3. The indicated OT range correction is less than or equal to R_t and the OT deflection correction is less than or equal to 50 meters. However, the deflection correction must be applied prior to firing rounds for effect.

ADJUSTMENT PHASE RANGE AND DEFLECTION CORRECTIONS

Corrections made by an observer during the adjustment phase are based on his ability to estimate the impact points of the rounds fired. Bracketing the target (that is one round impacting short and another beyond the target) on the first two rounds of the mission is a critical factor in determining the number of rounds required to accomplish the mission. During an observer's training he is severely penalized if he fails to do so and for that reason, he over adjusts prior to firing the second round. Table I below shows minimum observer range corrections depending on the observer's estimated distance to the target.

The observer senses the round impact points and makes his corrections on the OT line. These corrections are transformed to the GT line and thus enable weapon elevation and azimuth corrections to be applied. Azimuth corrections are made to change the deflection right or left to the nearest ten meters. Since range is much more difficult, corrections are made so that the elevation is changed to increase or decrease range according to Table II.

Table I

Observer's Estimated Distance to the Target Center (Meters)	Minimum Observer Range Correction Prior to Firing Second Round (Meters)
0 - 999	100
1000 - 1999	200
2000 - over	400

Table II

Observed Impact Point Deviation (GT line), Δ, from Target Center (Meters)	Amount of Correction (r^*) (Meters)
$R_t \leq \Delta < 100$	100
$100 \leq \Delta < 200$	200
$200 \leq \Delta < 300$	400
$300 \leq \Delta < 500$	600
$500 \leq \Delta < 700$	800
$700 \leq \Delta < 900$	1000
$900 \leq \Delta < \infty$	1200

* It should be noted that if Δ is positive then r is negative and vice versa.

DEVELOPMENT OF A MATHEMATICAL MODEL TO REPRESENT THE ADJUSTMENT PHASE PROCEDURES

An adjusted fire mission can be separated into two basic phases which are (1) the adjustment and (2) fire for effect phases. This paper is concerned with the adjustment phase effects on the fire for effect phase.

The first step in any adjustment phase is to aim the weapon at the desired target (i.e., an elevation and azimuth to fire the first round of the mission). Since it is nearly impossible to aim any weapon system on the center of the target, aiming errors are inherent in all firing missions. The aiming errors involve (1) meteorological (2) target location and (3) other related predicted fire errors (such as estimated velocity, elevation, and azimuth measurements, etc.). However, this paper is concerned with adjusted fire missions only and thus aiming errors only affects the first round of every individual adjustment phase. Since the range aim point ($\bar{y}$) has true range ($\mu_{\bar{x}}$) and standard deviation, $\sigma_{\bar{y}}$, and the deflection aim point ($\bar{x}$) has expected deflection ($\mu_{\bar{x}}$) and standard deviation, $\sigma_{\bar{x}}$, some method of characterizing the inaccuracies of estimating the above expectations must be devised. It is assumed that the expected range and deflection follow some arbitrary distribution and therefore the density of impacts at ranges $\bar{y}_j$ and

deflection $\tilde{x}_j$ can be expressed by the functions, $\rho(y:\mu_{\tilde{y}}, \sigma_{\tilde{y}})$ and $\rho(x:\mu_{\tilde{x}}, \sigma_{\tilde{x}})$, respectively.

Upon having aimed the weapon system the first round is fired and if the system had perfect accuracy the round impact point coordinates would be $(\tilde{y},\tilde{x})$. Since no weapon systems have been devised with perfect accuracy, some means of characterizing the inaccuracies must be developed. It is assumed that the range delivery errors follow some arbitrary distribution, $\rho(y:\tilde{y}, \sigma_y)$ where $\tilde{y}$ is the expected range and σ_y is the round-to-round range standard deviation. Similarly, the delivery errors existing in deflection can be expressed by the distribution $\rho(x:\tilde{x}, \sigma_x)$.

The coordinate of the actual impact point for the adjustment phase round is represented by (y_1, x_1) with respect to the GT line or (y_1', x_1') with respect to the OT line. Since the observer utilizes his estimate of (y_1', x_1') to make OT line deviations Δ_1 (range) and d_1 (deflection) an error in the adjustment phase is attributable to his mislocation of the impact point. A method must be devised to characterize these errors. It is assumed that his range errors, ye, can be expressed by some arbitrary distribution, $\rho(ye:\mu_{ye}, \sigma_{ye})$ and the observer's deflection errors, xe, are expressed by $\rho(xe:\mu_{xe}, \sigma_{xe})$ and are similarly characterized in each subsequent adjustment phase round.

If none of the previously defined conditions necessary for accomplishment of the adjustment phase have been satisfied the

corrections r_1 and d_1 for range and deflection respectively are applied and the second round is fired. The distribution of the delivery errors associated with an individual weapon system does not vary, however, due to the application of the above corrections the expected range and deflection for the second round are $y+r_1$ and $x+d_1$ respectively. The distributions of range and deflection for the second round are expressed by the functions, $\rho(y:\bar{y}+r_1, \sigma_y)$ and $\rho(x:\bar{x}+d_1, \sigma_x)$, respectively. In a similar manner the expected range and deflection impact points changes for each subsequent adjustment round depends on the corrections applied. For the i^{th} round of any adjustment mission the expected range and deflection can be expressed by

Range:

$$\bar{y}_i = \bar{y} \qquad \text{For } i = 1 \qquad (1)$$

$$\bar{y}_i = \bar{y} + \sum_{k=1}^{i-1} r_k \qquad \text{For } i > 1 \qquad (2)$$

Deflection:

$$\bar{x}_i = \bar{x} \qquad \text{For } i = 1 \qquad (3)$$

$$\bar{x}_i = \bar{x} + \sum_{k=1}^{i-1} d_k \qquad \text{For } i > 1 \qquad (4)$$

where $\bar{y}$ ans $\bar{x}$ are the expected range and deflection for the first round of an adjustment mission and the second terms of each expression are total range and deflection corrections applied before firing the i^{th} round.

Since all of these previously described errors affect any adjustment mission the question posed upon completion of the adjustment phase is that of the center of impact of rounds fired for effect. The ultimate goal of the adjustment phase or any method fire is to obtain an elevation and azimuth which will have an aim point with the coordinates (μ_y, μ_x) where μ_y is the true range from weapon system to target center and there is no deflection error. However, this is a practical impossibility and upon completion of the adjustment phase the rounds fired for effect will have aim point coordinates (y^*, x^*) where y^* and x^* are the range and deflection centers of impact of these rounds.

Because it was assumed that range and deflection errors are independently distributed it is possible to work with y^* and x^* individually. The adjustment phase is terminated if condition (1) any round is observed to impact on the target registration area or (2) a y meter bracket has been split or (3) the OT line range correction is $\leq R_t$ and the OT line deflection correction is ≤ 50 meters is met. Assuming the adjustment phase is terminated on the n^{th} round then using equations (1), (2), (3), and (4) the expected range and deflection for this round are

Range:

$$\bar{y}_n = \bar{y} \qquad \text{For } n = 1 \qquad (5)$$

$$\bar{y}_n = \bar{y} + \sum_{k=1}^{n-1} r_k \qquad \text{For } n > 1 \qquad (6)$$

Deflection:

$$\bar{x}_n = \bar{x} \qquad \text{For } n = 1 \qquad (7)$$

$$\bar{x}_n = \bar{x} + \sum_{k=1}^{n-1} d_k \qquad \text{For } n > 1 \qquad (8)$$

If termination of the adjustment phase was due to fulfillment of conditions (1) or (3) the deflection corrections are applied and the fire for effect phase is begun thus the center of impact of these rounds will be the following

Range:

$$y^* = \bar{y}_n = \bar{y} \qquad \text{For } n = 1 \qquad (9)$$

$$y^* = \bar{y}_n = \bar{y} + \sum_{k=1}^{n-1} r_k \qquad \text{For } n > 1 \qquad (10)$$

Deflection:

$$x^* = \bar{x}_n = \bar{x} + \sum_{k=1}^{n} d_k \qquad (11)$$

However, if condition (2) is met the deflection center of impact remains as shown by equation (11) but since half the indicated range correction

is applied the range center of impact of the fire for effect phase rounds is now defined by

$$y^* = \bar{y} + 1/2\ r_1 \qquad \text{For } n = 1 \qquad (12)$$

$$y^* = \bar{y} + \sum_{k=1}^{n-1} r_k + 1/2\ r_n \qquad \text{For } n > 1 \qquad (13)$$

Thus equations (9), (10), (12), or (13) is the mathematical representation of the range center of impact of the fire for effect phase rounds and equation (11) is the deflection representation of these rounds.

It is apparent that estimates of the parameters y^*, x^*, n, σ_y^*, σ_x^*, and σ_n are a necessity in determining the effects the adjustment phase have on the fire for effect phase.

ESTIMATES OF THE ADJUSTMENT PHASE PARAMETERS BASED ON THE NORMAL DISTRIBUTION

A method had to be devised to obtain estimates of these parameters due to the fact that there was a lack of data and extreme difficulty in conducting experiments to obtain this data. It was decided that Monte Carlo techniques would be used to obtain estimates of these parameters.

Prior to using these Monte Carlo techniques it was necessary to assume that the (1) aiming, (2) delivery and (3) observer errors previously discussed follow some particular distribution or distributions. For the purpose of this paper these errors were assumed to be normally distributed.

Since the aiming errors were assumed to be normally distributed, the density of impacts at ranges $\bar{y}_j$ and deflection $\bar{x}_j$ can be expressed by the functions

Range:

$$\hat{\rho}(\bar{y}) = \frac{1}{\sqrt{2\pi}\,\hat{\sigma}_{\bar{y}}} \exp\left[-(\bar{y} - \mu_{\bar{y}})^2 / 2\,\sigma_{\bar{y}}^2\right] \quad (14)$$

Deflection:

$$\hat{\rho}(\bar{x}) = \frac{1}{\sqrt{2\pi}\,\hat{\sigma}_{\bar{x}}} \exp\left[-(\bar{x} - \mu_{\bar{x}})^2 / 2\,\hat{\sigma}_{\bar{x}}^2\right] \quad (15)$$

where $\mu_{\bar{y}}$ and $\mu_{\bar{x}}$ are expected range and deflection, and $\hat{\sigma}_{\bar{y}}$ and $\hat{\sigma}_{\bar{x}}$ are the standard deviations of the aiming errors associated with range and deflection, respectively.

The second type of error associated with the adjustment phase rounds is the delivery error. The distributions of range and deflection can be expressed by the functions

Range:

$$\hat{\rho}(y) = \frac{1}{\sqrt{2\pi}\,\hat{\sigma}_y} \quad \exp\left[-(y-\bar{y})^2 / 2\,\hat{\sigma}_y^{\,2}\right] \qquad (16)$$

Deflection:

$$\hat{\rho}(x) = \frac{1}{\sqrt{2\pi}\,\hat{\sigma}_x} \quad \exp\left[-(x-\bar{x})^2 / 2\,\hat{\sigma}_x^{\,2}\right] \qquad (17)$$

where $\bar{y}$ and $\bar{x}$ are the expected range and deflection, and $\hat{\sigma}_{\bar{y}}$ and $\hat{\sigma}_{\bar{x}}$ are the round-to-round range and deflection standard deviations, respectively.

The third error associated with the adjustment phase rounds is the observer error in locating impact points. Since it was assumed his errors are normally distributed, the density of the observer errors in range, ye, and deflection, xe, can be expressed by the functions.

Range:

$$\hat{\rho}(ye) = \frac{1}{\sqrt{2\pi}\,\hat{\sigma}_{ye}} \exp\left[-(ye - \mu_{ye})^2 / 2\,\hat{\sigma}_{ye}^2\right] \quad (18)$$

Deflection:

$$\hat{\rho}(xe) = \frac{1}{\sqrt{2\pi}\,\hat{\sigma}_{xe}} \exp\left[-(xe - \mu_{xe})^2 / 2\,\hat{\sigma}_{xe}^2\right] \quad (19)$$

where μ_{ye} and μ_{xe} are the expected observer errors and $\hat{\sigma}_{ye}$ and $\hat{\sigma}_{xe}$ are the round-to-round standard deviations in observer errors for range and deflection, respectively.

MONTE CARLO TECHNIQUES

In order to use Monte Carlo techniques, it was necessary to draw random samples from normal distributions. The Ballistic Research Laboratories Electronic Scientific Computers (BRLESC I and II) at Aberdeen Proving Ground, Maryland, were used to generate these random numbers.

A random number, Z, was generated from a normal distribution with a mean, 0, and standard deviation, 1, i.e., N(0,1). This normal random, number, Z, was then converted to a normal random number, Z^*, from N (μ_{Z^*}, σ_{Z^*}). As an example let us consider the sampling of the range aim point for the first adjustment round, which is known to have mean, μ_y, (assumed to be the actual target center) and standard deviation, $\sigma_{\bar{y}}$. A normal random number, Z, from a N(0,1) is generated and converted to a normal random number, Z^*, from a N ($\mu_{\bar{y}}$, $\sigma_{\bar{y}}$). This procedure for generating aim points, impact points and observer errors was used throughout these adjustment procedures.

Estimates of the means and standard deviations of y^* and x^* for samples of size 5000 were generated using the above procedures.

A PRACTICAL APPLICATION FOR THE RESULTS GENERATED

The broad problem of gun tube life termination involves the study of several criteria to establish the controlling factors which correspond to the lowest limit in terms of number of rounds corresponding to acceptable performance of a gun system. A major factor in determining gun tube life termination is accuracy-cost. Accuracy-cost can be best described as such, the accuracy of a gun system is degraded as the number of rounds fired from a tube increases. The wear characteristics of the tube cause greater round-to-round dispersion. As the number of rounds fired from a tube increases there is a balance between the tube cost per round and ammunition costs as the number of rounds required to hit a target increases. A point is reached where the increased ammunition costs more than compensate for the tube cost savings effected by extending the tube life.

A method for determining gun tube life termination based on accuracy-cost was developed by Bruno, Kniss and Gerard (1). This method involves the concept of an effective round which is defined as a round which hits a target during either the adjustment or fire for effect phases.

The probability of an effective round was defined as

$$P(\text{effective round}) = C(1-A) + A(B) \qquad (20)$$

where A is the ratio of the rounds used in an adjustment phase to the total number of rounds fired in a mission (except for emergency missions the number of rounds fired are generally predetermined), B is the probability of a round hitting the target during the adjustment phase and C is the probability of a round hitting the target during the fire for effect phase. Estimates of the mean number of rounds expended during the adjustment phase were used to obtain estimates of A and σ_y^* and σ_x^* were used to determine the overall dispersion of the given weapon system in obtaining estimates of C.

The ammunition cost per effective round was then defined as

$$Y = \frac{S}{C(1-A)+A(B)} \qquad (21)$$

where S is the ammunition cost per round and the denominator of equation (21) is the probability of an effective round as previously defined by equation (20).

By utilizing equations (20 and (21) and results of these Monte Carlo simulations optimum gun tube life termination round numbers based on accuracy-cost could be determined for several gun systems.

SENSITIVITY ANALYSIS

Upon having developed a mathematical model to represent these adjustment phase procedures, Monte Carlo techniques were used to generate data to obtain estimates of the mean and dispersions for three basic parameters associated with these procedures. These parameters were (1) the number of rounds required to accomplish an adjustment mission, (2) the range center of impact and (3) the deflection center of impact.

Prior to conducting a sensitivity analysis, investigations to determine which factors would possibly effect these adjustment phase procedures were conducted. Since these investigations revealed there were numerous factors their minimum and maximum estimates were used to investigate their effect on the adjustment phase procedures. For example data was generated for deflection round-to-round standard deviations of 1 and 20.

The sensitivity analysis was conducted to determine the effects of seven factors on the desired parameter estimates. The first and second factors, target size (registration area) and range round-to-round dispersion, significantly effected the results obtained for all the above parameters. The third factor investigated was the deflection round-to-round dispersion. Results indicated

that increasing the deflection round-to-round standard deviation resulted in a significant increase of the deflection center of impact disperaions, but had no effect on the estimates of the remaining parameters.

Factor four, the angle between th[illegible] line and OT line, significantly effected the dispersion estimates of the range and deflection center of impacts. The range estimates were significantly decreased while the deflection estimates were significantly increased as the angle was increased. The fifth factor, OT distance did not significantly effect the parameter estimates, but significantly effects the observer's errors in locating impact points of adjustment phase rounds, the sixth factor. The estimates of observer errors used in this paper were based on the miss distance of the impact point from the target center which was weighted according to the OT distance. These estimates were based on judgment only due to the lack of available data. However, efforts are currently being made to conduct experiments to generate data to estimate the observer's errors. It should be noted that these adjustment phase procedures are sensitive, depending on the magnitude, to these observer errors. However, it is felt that data from the above experiments will probably not yield observer errors of sufficient magnitude to effect the results. The final factor, the first round aiming errors, had no practical significance on the results.

Although the sensitivity analysis data could be utilized to obtain some optimum policy for these adjustment phase procedures, this optimum policy could not be achieved since most of the factors are uncontrollable, i.e., such as the observer's ability to control his angle or distance to the target because of environmental factors.

Results of the data generated by the Monte Carlo techniques are summarized in Tables I through VI.

CONCLUSIONS

It has been shown, by performing a sensitivity analysis on data obtained, that six of the seven factors investigated effected the estimates of the parameters of these adjustment phase procedures. However, these parameter estimates have been shown to be more sensitive to the target size (registration area) and the round-to-round range dispersion.

These results indicate that employing adjusted fire techniques the dispersion of the fire for effect rounds is smaller than for any other method of fire. Using these results it has been shown the probabilities of hitting a given target is greater for adjusted fire (even including the rounds for the adjustment phase) than for any other method of fire. Therefore, these results strengthens the artillerymen's position of employing adjusted fire whenever possible.

Furthermore, estimates of these parameters are being used to improve current estimates of the overall accuracies and optimum tube life of weapon systems.

BIBLIOGRAPHY

1. Bruno, O. P., Kniss, J. R., and Gerald, S., Operation Research Criteria for Accuracy Life of Gun Tubes, 175mm, M113E1, Aberdeen Research and Development Center, Aberdeen Proving Ground, Maryland, July, 1969.
2. Davies, O. L., The Design and Analysis of Industrial Experiments, Hafner Publishing Company, New York, 1960.
3. Department of the Army Field Manual, FM 23-90, 81-MM Mortar, M29, Washington, D. C.: Headquarters, Department of the Army, January, 1967.
4. Department of the Army Field Manual, FM 6-40, Field Artillery Cannon Gunnery, Washington, D. C.: Headquarters, Department of the Army October, 1967.
5. Mood, A. M., Introduction to the Theory of Statistics, McGraw Hill Book Co., Inc., 1950.

APPENDIX A

TABLES OF MEANS AND DISPERSIONS OF THE NUMBER OF ROUNDS EXPENDED DURING AN ADJUSTMENT MISSION

TABLE III

Mean and Dispersions of the Number of Rounds Expended during an Adjustment Phase Mission For a Target of Radius 1 Meter

$\hat{\sigma}_y$ = 25 meters

Deflection Rd.-to-Rd. Std. Dev. ($\hat{\sigma}_x$)	Observer* Standard Deviations ($\hat{\sigma}_{ye}$ & $\hat{\sigma}_{xe}$)	OT Angle (Radians)			
		0		.4	
		Average No. of Rds. (M)	Std. Dev. of No. of Rds. ($\hat{\sigma}_n$)	Average No. of Rds. (M)	Std. Dev. of No. of Rds. ($\hat{\sigma}_n$)
1	OE_1	3.74	1.44	4.09	1.67
	OE_2	5.64	2.54	6.74	2.91
	OE_3	4.10	1.75	4.75	2.15
	OE_4	7.84	3.65	8.58	3.65
20	OE_1	3.99	1.57	4.20	1.76
	OE_2	5.86	2.69	6.78	2.94
	OE_3	4.23	1.92	4.78	2.15
	OE_4	7.99	3.77	8.74	3.68

$\hat{\sigma}_y$ = 50 meters

1	OE_1	4.36	2.07	5.07	2.60
	OE_2	5.90	2.91	7.31	3.20
	OE_3	4.61	2.30	5.73	2.98
	OE_4	7.83	3.71	8.68	3.67
20	OE_1	4.54	2.23	5.20	2.71
	OE_2	6.30	3.05	7.67	3.55
	OE_3	4.94	2.43	5.70	2.78
	OE_4	8.07	3.79	8.78	3.81

* The observer Standard Deviations are defined in footnotes at the end of this article.

TABLE III(Cont'd)

Mean and Dispersions of the Number of Rounds Expended during an Adjustment Phase Mission For a Target of Radius 1 Meter

$\hat{\sigma}_y$ = 100 meters

Deflection Rd.-to-Rd. Std. Dev. ($\hat{\sigma}_x$)	Observer* Standard Deviations ($\hat{\sigma}_{ye}$ & $\hat{\sigma}_{xe}$)	OT Angle (Radians) 0		.4	
		Average No. of Rds. (ṁ)	Std. Dev. of No. of Rds. ($\hat{\sigma}_n$)	Average No. of Rds. (ṁ)	Std. Dev. of No. of Rds. ($\hat{\sigma}_n$)
1	OE_1	3.77	1.78	5.02	2.71
	OE_2	5.05	2.56	6.87	3.35
	OE_3	4.05	2.03	5.62	3.02
	OE_4	7.25	3.73	8.07	3.79
20	OE_1	3.93	1.90	5.04	2.72
	OE_2	5.55	2.85	6.97	3.41
	OE_3	4.30	2.19	5.65	3.06
	OE_4	7.32	3.79	8.06	3.86

$\hat{\sigma}_y$ = 150 meters

1	OE_1	4.65	2.61	6.40	3.54
	OE_2	5.52	3.02	7.36	3.61
	OE_3	4.69	2.62	6.73	3.62
	OE_4	7.34	3.78	8.06	3.91
20	OE_1	4.92	2.81	6.38	3.51
	OE_2	5.83	3.13	7.42	3.63
	OE_3	5.10	2.88	6.72	3.60
	OE_4	7.42	3.88	8.10	3.91

* The observer Standard Deviations are defined in footnotes on Page 326

TABLE III (Cont'd)

Mean and Dispersions of the Number of Rounds Expended

during an Adjustment Phase Mission For a Target of Radius 1 Meter

$\hat{\sigma}_y$ = 200 meters

Deflection Rd.-to-Rd. Std. Dev. ($\hat{\sigma}_x$)	Observer* Standard Deviations ($\hat{\sigma}_{ye}$ & $\hat{\sigma}_{xe}$)	OT Angle (Radians) 0		.4	
		Average No. of Rds. ($\bar{n}$)	Std. Dev. of No. of Rds. ($\hat{\sigma}_n$)	Average No. of Rds. ($\bar{n}$)	Std. Dev. of No. of Rds. ($\hat{\sigma}_n$)
1	OE_1	5.42	3.17	7.30	3.82
	OE_2	5.77	3.20	7.63	3.79
	OE_3	5.29	3.07	7.36	3.85
	OE_4	7.34	3.83	7.89	3.87
20	OE_1	5.84	3.34	7.27	3.80
	OE_2	6.32	3.52	7.91	3.81
	OE_3	5.59	3.24	7.25	3.79
	OE_4	7.56	3.92	8.14	3.91

* The observer Standard Deviations are defined in footnotes on Page 326

TABLE II

Mean and Dispersions of the Number of Rounds Expended
during an Adjustment Phase Mission For a Target Radius of 25 Meters

$\hat{\sigma}_y$ = 25 meters

Deflection Rd.-to-Rd. Std. Dev. (σ_x)	Observer* Standard Deviations ($\hat{\sigma}_{ye}$ & $\hat{\sigma}_{xe}$)	OT Angle (Radians) 0		.4	
		Average No. of Rds. ($\bar{n}$)	Std. Dev. of No. of Rds. ($\hat{\sigma}_n$)	Average No. of Rds. ($\bar{n}$)	Std. Dev. of No. of Rds. ($\hat{\sigma}_n$)
1	OE_1	3.32	1.19	3.37	1.26
	OE_2	4.55	2.10	5.03	2.15
	OE_3	3.50	1.39	3.76	1.58
	OE_4	6.81	3.76	7.27	3.94
20	OE_1	3.35	1.17	3.40	1.30
	OE_2	4.60	2.10	5.20	2.40
	OE_3	3.53	1.43	3.78	1.67
	OE_4	6.95	3.76	7.45	3.81

$\hat{\sigma}_y$ = 50 meters

1	OE_1	3.71	1.57	3.94	1.88
	OE_2	4.72	2.27	5.51	2.66
	OE_3	3.76	1.73	4.23	2.09
	OE_4	6.93	3.73	7.43	3.83
20	OE_1	3.78	1.65	4.01	1.94
	OE_2	4.85	2.37	5.49	2.65
	OE_3	3.83	1.75	4.28	2.10
	OE_4	6.98	3.75	7.43	3.82

* The observer Standard Deviations are defined in footnotes on Page 326

TABLE II (Cont'd)

Mean and Dispersions of the Number of Rounds Expended

during an Adjustment Phase Mission For a Target Radius of 25 Meters

$\hat{\sigma}_y$ = 100 meters

Deflection Rd.-to-Rd. Std. Dev. ($\hat{\sigma}_x$)	Observer* Standard Deviations ($\hat{\sigma}_{ye}$ & $\hat{\sigma}_{xe}$)	OT Angle (Radians) 0		.4	
		Average No. of Rds. ($\bar{n}$)	Std. Dev. of No. of Rds. ($\hat{\sigma}_n$)	Average No. of Rds. ($\bar{n}$)	Std. Dev. of No. of Rds. ($\hat{\sigma}_n$)
1	OE_1	3.48	1.57	4.01	2.12
	OE_2	4.33	2.23	5.30	2.82
	OE_3	3.60	1.75	4.26	2.24
	OE_4	6.45	3.69	7.24	3.87
20	OE_1	3.60	1.68	4.02	2.12
	OE_2	4.44	2.28	5.35	2.82
	OE_3	3.75	1.89	4.29	2.31
	OE_4	6.60	3.75	7.22	3.89

$\hat{\sigma}_y$ = 150 meters

1	OE_1	4.17	2.27	5.16	3.04
	OE_2	4.56	2.50	5.89	3.24
	OE_3	4.12	2.32	5.15	3.02
	OE_4	6.52	3.78	7.25	3.89
20	OE_1	4.39	2.48	5.19	3.08
	OE_2	4.87	2.79	5.86	3.22
	OE_3	4.28	2.39	5.25	3.05
	OE_4	6.58	3.82	7.16	3.93

* The observer Standard Deviations are defined in footnotes on Page 326

TABLE II (Cont'd)

Mean and Dispersions of the Number of Rounds Expended

during an Adjustment Phase Mission For a Target Radius of 25 Meters

$\hat{\sigma}_y$ = 200 Meters

Deflection Rd.-to-Rd. Std. Dev. (σ_x)	Observer* Standard Deviations ($\hat{\sigma}_{ye}$ & $\hat{\sigma}_{xe}$)	OT Angle (Radians) 0		.4	
		Average No. of Rds. ($\bar{n}$)	Std. Dev. of No. of Rds. ($\hat{\sigma}_n$)	Average No. of Rds. ($\bar{n}$)	Std. Dev. of No. of Rds. ($\hat{\sigma}_n$)
1	OE_1	4.58	2.89	6.07	3.52
	OE_2	4.90	2.86	6.25	3.49
	OE_3	4.57	2.74	5.87	3.42
	OE_4	6.61	3.82	7.28	4.00
20	OE_1	5.10	3.03	6.04	3.52
	OE_2	5.20	3.05	6.25	3.52
	OE_3	4.90	2.91	5.74	3.38
	OE_4	6.75	3.89	7.22	3.98

* The observer Standard Deviations are defined in footnotes on Page 326

TABLE III

Mean and Dispersion of the Number of Rounds Expended during an Adjustment Phase Mission for a Target of Radius 100 Meters

$\hat{\sigma}_y$ = 25 meters

Deflection Rd.-to-Rd. Std. Dev. ($\hat{\sigma}_x$)	Observer* Standard Deviations (σ_{ye} & σ_{xe})	OT Angle (Radians) 0 Average No. of Rds. ($\bar{n}$)	0 Std. Dev. of No. of Rds. ($\hat{\sigma}_n$)	.4 Average No. of Rds. ($\bar{n}$)	.4 Std. Dev. of No. of Rds. ($\hat{\sigma}_n$)
1	OE_1	2.74	1.07	2.55	0.99
	OE_2	3.24	1.75	3.36	1.83
	OE_3	2.79	1.26	2.80	1.29
	OE_4	5.08	3.58	5.34	3.61
20	OE_1	2.76	1.06	2.61	1.01
	OE_2	3.22	1.72	3.38	1.84
	OE_3	2.72	1.21	2.80	1.30
	OE_4	5.14	3.56	5.25	3.59

$\hat{\sigma}_y$ = 50 meters

1	OE_1	2.84	1.21	2.68	1.14
	OE_2	3.24	1.84	3.48	1.97
	OE_3	2.83	1.36	2.85	1.40
	OE_4	5.13	3.60	5.39	3.64
20	OE_1	2.84	1.19	2.71	1.14
	OE_2	3.30	1.85	3.49	1.91
	OE_3	2.84	1.37	2.86	1.39
	OF_4	5.10	3.50	5.39	3.67

* The observer Standard Deviations are defined in footnotes on Page 326

TABLE III (Cont'd)

Mean and Dispersion of the Number of Rounds Expended during an Adjustment Phase Mission for a Target of Radius 100 Meters

$\hat{\sigma}_y$ = 100 meters

Deflection Rd.-to-Rd. Std. Dev. ($\hat{\sigma}_x$)	Observer* Standard Deviations ($\hat{\sigma}_{ye}$ & $\hat{\sigma}_{xe}$)	OT Angle (Radians) 0		.4	
		Average No. of Rds. ($\bar{n}$)	Std. Dev. of No. of Rds. ($\hat{\sigma}_n$)	Average No. of Rds. ($\bar{n}$)	Std. Dev. of No. of Rds. ($\hat{\sigma}_n$)
1	OE_1	2.81	1.31	2.76	1.33
	OE_2	3.19	1.83	3.41	1.99
	OE_3	2.73	1.39	2.88	1.47
	OE_4	5.07	3.55	5.25	3.63
20	OE_1	2.80	1.29	2.78	1.34
	OE_2	3.13	1.76	3.40	1.94
	OE_3	2.78	1.40	2.90	1.51
	OE_4	4.93	3.48	5.24	3.66

$\hat{\sigma}_y$ = 150 meters

1	OE_1	3.19	1.84	3.22	1.88
	OE_2	3.30	2.02	3.57	2.21
	OE_3	3.07	1.80	3.17	1.87
	OE_4	5.00	3.51	5.39	3.75
20	OE_1	3.26	1.84	3.28	1.93
	OE_2	3.36	2.07	3.59	2.22
	OE_3	3.07	1.80	3.22	1.84
	OE_4	4.95	3.51	5.27	3.65

* The observer Standard Deviations are defined in footnotes on Page 326

TABLE III (Cont'd)

Mean and Dispersion of the Number of Rounds Expended during an Adjustment Phase Mission for a Target of Radius 100 Meters

$\hat{\sigma}_y$ = 200 meters

Deflection Rd.-to-Rd. Std. Dev. ($\hat{\sigma}_x$)	Observer* Standard Deviations ($\hat{\sigma}_{ye}$ & $\hat{\sigma}_{xe}$)	OT Angle (Radians) 0		.4	
		Average No. of Rds. ($\bar{n}$)	Std. Dev. of No. of Rds. ($\hat{\sigma}_n$)	Average No. of Rds. ($\bar{n}$)	Std. Dev. of No. of Rds. ($\hat{\sigma}_n$)
1	OE_1	3.67	2.31	3.71	2.37
	OE_2	3.45	2.26	3.83	2.52
	OE_3	3.30	2.05	3.48	2.15
	OE_4	5.09	3.60	5.39	3.77
20	OE_1	3.75	2.36	3.78	2.44
	OE_2	3.58	2.33	3.76	2.44
	OE_3	3.35	2.09	3.57	2.28
	OE_4	5.14	3.70	5.42	3.85

* The observer Standard Deviations are defined in footnotes on Page 326

APPENDIX B

TABLES OF MEANS AND DISPERSION OF RANGE AND DEFLECTION CENTER OF IMPACTS

TABLE IV

Means and Dispersions of Range and Deflection Center of Impacts For a Target of Radius 1 Meter

Range

$\hat{\sigma}_y$ = 25 meters

Deflection Rd.-to-Rd. Std. Dev. ($\hat{\sigma}_x$)	Observer* Standard Deviation ($\hat{\sigma}_{ye}$ & $\hat{\sigma}_{xe}$)	OT Angle (Radians) 0		.4	
		Average Center of Impact	Standard Deviation of CI's	Average Center of Impact	Standard Deviation of CI's
1	OE_1	-1.89	34.91	2.73	32.22
	OE_2	-0.28	110.92	1.15	53.78
	OE_3	-7.45	59.66	0.75	40.16
	OE_4	-17.79	365.50	8.72	95.83
20	OE_1	-1.94	36.47	2.24	34.39
	OE_2	-0.93	107.50	0.62	58.26
	OE_3	-5.55	50.39	1.04	39.30
	OE_4	-6.07	282.13	8.27	99.15

Deflection

$\hat{\sigma}_y$ = 25 meters

1	OE_1	-0.09	3.00	5.45	19.69
	OE_2	-0.05	4.33	5.34	30.44
	OE_3	0.04	3.31	6.46	22.55
	OE_4	-0.23	10.55	6.61	60.26
20	OE_1	0.26	17.89	2.63	24.38
	OE_2	0.28	18.79	6.47	36.20
	OE_3	-0.12	18.06	39.30	3.30
	OE_4	0.47	27.91	4.96	63.84

* The observer Standard Deviations are defined in footnotes on Page 326

TABLE IV (Cont'd)

Means and Dispersions of Range and Deflection

Center of Impacts For a Target of Radius 1 Meter

Range

$\hat{\sigma}_y$ = 50 meters

Deflection Rd.-to-Rd. Std. Dev. (σ_x)	Observer* Standard Deviation ($\hat{\sigma}_{ye}$ & $\hat{\sigma}_{xe}$)	OT Angle (Radians) 0		.4	
		Average Center of Impact	Standard Deviation of CI's	Average Center of Impact	Standard Deviation of CI's
1	OE_1	-4.68	48.19	0.69	43.96
	OR_2	-3.48	138.96	-0.18	69.65
	OE_3	-11.25	67.12	-1.46	50.04
	OE_4	-8.29	359.06	1.47	131.12
20	OE_1	-2.13	46.85	1.69	46.46
	OE_2	-4.19	121.85	-0.62	60.92
	OE_3	-7.26	61.91	-1.09	49.47
	OE_4	-15.57	306.11	6.87	102.23

Deflection

$\hat{\sigma}_y$ = 50 meters

1	OE_1	-0.01	2.96	4.87	18.83
	OE_2	-0.10	4.48	6.63	35.71
	OE_3	0.01	3.16	5.07	23.46
	OE_4	-0.06	9.55	6.38	65.27
20	OE_1	0.24	17.86	2.55	25.23
	OE_2	-0.05	18.60	6.02	41.13
	OE_3	-0.48	18.00	4.37	28.46
	OE_4	0.86	26.88	4.08	79.59

* The observer Standard Deviations are defined in footnotes on Page 326

TABLE IV (Cont'd)

Means and Dispersions of Range and Deflection

Center of Impacts For a Target of Radius 1 Meter

Range

$\hat{\sigma}_y$ = 100 meters

Deflection Rd.-to-Rd. Std. Dev. ($\hat{\sigma}_x$)	Observer* Standard Deviation ($\hat{\sigma}_{ye}$ & $\hat{\sigma}_{xe}$)	QT Angle (Radians) 0 Average Center of Impact	0 Standard Deviation of CI's	.4 Average Center of Impact	.4 Standard Deviation of CI's
1	OE_1	-17.01	112.71	-3.25	99.38
	OE_2	-16.46	207.18	-0.19	131.95
	OE_3	-28.24	139.57	-9.05	104.40
	OE_4	-0.99	415.32	13.01	185.87
20	OE_1	-16.61	110.16	-2.94	99.65
	OE_2	-19.68	210.51	-10.85	129.69
	OE_3	-28.26	136.29	-6.99	108.57
	OE_4	-27.08	453.10	2.22	186.91

Deflection

$\hat{\sigma}_y$ = 100 meters

1	OE_1	0.05	2.97	6.56	30.35
	OE_2	0.10	5.05	11.51	53.64
	OE_3	0.02	3.22	10.12	38.88
	OE_4	-0.20	12.76	5.79	85.32
20	OE_1	0.53	18.05	5.95	34.21
	OE_2	0.65	19.16	13.22	58.10
	OE_3	0.29	17.98	8.49	43.35
	OE_4	0.53	26.30	9.42	84.65

* The observer Standard Deviations are defined in footnotes on Page 326

TABLE IV (Cont'd)

Means and Dispersions of Range and Deflection

Center of Impacts For a Target of Radius 1 Meter

Range

$\hat{\sigma}_y$ = 150 meters

Deflection Rd.-to-Rd. Std. Dev. ($\hat{\sigma}_x$)	Observer* Standard Deviation (σ_{ye} & σ_{xe})	OT Angle (Radians) 0		4	
		Average Center of Impact	Standard Deviation of CI's	Average Center of Impact	Standard Deviation of CI's
1	OE_1	-13.90	79.64	0.54	70.57
	OE_2	-13.11	176.90	-1.28	100.50
	OE_3	-22.51	108.40	-3.63	79.55
	OE_4	-15.35	379.91	7.93	155.59
20	OE_1	-11.36	80.01	-2.04	72.00
	OE_2	-11.57	167.31	-1.83	106.66
	OE_3	-18.00	104.82	-3.86	81.02
	OE_4	-4.29	360.70	10.78	158.12

Deflection

$\hat{\sigma}_y$ = 150 meters

1	OE_1	-0.01	2.97	5.99	28.60
	OE_2	-0.01	4.97	8.58	47.43
	OE_3	0.07	3.56	7.76	35.10
	OE_4	-0.01	12.38	6.61	80.76
20	OE_1	0.56	17.81	5.11	32.90
	OE_2	0.49	18.76	9.36	51.54
	OE_3	-0.31	18.01	7.32	39.38
	OE_4	-0.25	26.38	7.86	81.48

* The observer Standard Deviations are defined in footnotes on Page 326

TABLE IV (Cont'd)

Means and Dispersions of Range and Deflection

Center of Impacts For a Target of Radius 1 Meter

Range

$\hat{\sigma}_y$ = 200 meters

Deflection Rd.-to-Rd. Std. Dev. (σ_x)	Observer* Standard Deviation ($\hat{\sigma}_{ye}$ & $\hat{\sigma}_{xe}$)	OT Angle (Radians) 0		.4	
		Average Center of Impact	Standard Deviation of CI's	Average Center of Impact	Standard Deviation of CI's
1	OE_1	-22.48	143.15	-9.20	129.36
	OE_2	-17.02	241.27	-5.70	164.29
	OE_3	-35.99	173.23	-8.65	136.83
	OE_4	-17.40	445.22	2.88	232.92
20	OE_1	-19.02	141.06	-3.26	126.60
	OE_2	-18.02	231.10	-10.02	166.50
	OE_3	-33.37	170.03	-11.72	137.54
	OE_4	-11.87	411.32	9.83	251.30

Deflection

$\hat{\sigma}_y$ = 200 meters

1	OE_1	0.03	3.00	8.60	30.76
	OE_2	0.09	4.71	11.85	62.45
	OE_3	0.02	3.28	11.25	42.26
	OE_4	-0.07	12.22	9.49	99.04
20	OE_1	0.17	18.07	7.76	34.36
	OE_2	-0.20	18.81	13.09	64.72
	OE_3	0.01	18.08	12.35	44.98
	OE_4	0.06	28.36	9.32	97.04

* The observer Standard Deviations are defined in footnotes on Page 326

TABLE V

Means and Dispersions of Range and Deflection

Center of Impacts For a Target of Radius 25 Meters

Range

$\hat{\sigma}_y$ = 25 meters

Deflection Rd.-to-Rd. Std. Dev. ($\hat{\sigma}_x$)	Observer* Standard Deviation ($\hat{\sigma}_{ye}$ & $\hat{\sigma}_{xe}$)	OT Angle (Radians) 0		.4	
		Average Center of Impact	Standard Deviation of CI's	Average Center of Impact	Standard Deviation of CI's
1	OE_1	-4.79	30.34	-1.10	34.60
	OE_2	4.71	149.51	16.15	106.96
	OE_3	-8.99	64.86	-4.69	64.96
	OE_4	1.83	319.45	-9.39	328.52
20	OE_1	-3.38	29.02	-2.04	36.47
	OE_2	-0.08	138.41	0.78	122.83
	OE_3	-6.45	65.38	-2.65	67.98
	OE_4	-3.31	305.29	9.17	317.04

Deflection

$\hat{\sigma}_y$ = 25 meters

1	OE_1	-0.04	2.97	3.07	15.68
	OE_2	-0.25	5.77	0.00	43.08
	OE_3	-0.05	3.67	3.18	26.69
	OE_4	-0.02	16.09	0.83	106.77
20	OE_1	0.19	19.50	1.78	24.85
	OE_2	0.09	22.41	1.68	54.55
	OE_3	-0.12	20.65	1.85	33.73
	OE_4	0.62	35.86	0.83	112.95

* The observer Standard Deviations are defined in footnotes on Page 326

TABLE V (Cont'd)

Means and Dispersions of Range and Deflection

Center of Impacts For a Target of Radius 25 Meters

Range

$\hat{\sigma}_y$ = 50 meters

Deflection Rd.-to-Rd. Std. Dev. ($\hat{\sigma}_x$)	Observer* Standard Deviation ($\hat{\sigma}_{ye}$ & $\hat{\sigma}_{xe}$)	OT Angle (Radians) 0		.4	
		Average Center of Impact	Standard Deviation of CI's	Average Center of Impact	Standard Deviation of CI's
1	OE_1	-4.41	46.01	-2.60	49.27
	OE_2	-3.71	145.65	3.71	134.65
	OE_3	-9.54	80.28	-4.91	80.62
	OE_4	2.05	369.61	5.59	305.90
20	OE_1	-4.86	45.81	-2.61	49.81
	OE_2	2.03	141.42	2.39	131.97
	OE_3	-5.33	81.49	-4.61	72.97
	OE_4	3.93	314.51	-0.13	294.26

Deflection

$\hat{\sigma}_y$ = 50 meters

1	OE_1	-0.00	3.13	2.39	15.80
	OE_2	-0.06	6.07	0.28	55.27
	OE_3	-0.01	3.71	2.17	29.83
	OE_4	0.34	18.51	1.39	107.87
20	OE_1	-0.85	19.61	1.80	25.31
	OE_2	-0.14	23.34	0.32	58.18
	OE_3	0.09	20.19	2.52	34.40
	OE_4	-0.13	36.52	3.60	109.38

* The observer Standard Deviations are defined in footnotes on Page 326

TABLE V (Cont'd)

Means and Dispersions of Range and Deflection
Center of Impacts For a Target of Radius 25 Meters

Range

$\hat{\sigma}_y$ = 100 meters

Deflection Rd.-to-Rd. Std. Dev. ($\hat{\sigma}_x$)	Observer* Standard Deviation ($\hat{\sigma}_{ye}$ & $\hat{\sigma}_{xe}$)	OT Angle (Radians) 0		.4	
		Average Center of Impact	Standard Deviation of CI's	Average Center of Impact	Standard Deviation of CI's
1	OE_1	-14.92	84.25	-4.62	80.55
	OE_2	-3.98	185.85	-0.13	154.45
	OE_3	-21.11	117.37	-6.51	101.89
	OE_4	-0.29	424.59	6.52	305.82
20	OE_1	-10.97	84.46	-3.82	82.69
	OE_2	-1.43	188.44	-3.89	148.52
	OE_3	-18.46	111.65	-7.40	102.81
	OE_4	0.66	394.30	13.71	335.32

Deflection

$\hat{\sigma}_y$ = 100 meters

1	OE_1	-0.06	3.03	4.96	24.21
	OE_2	-0.11	7.00	5.23	62.81
	OE_3	-0.06	3.65	4.44	38.12
	OE_4	-0.65	22.94	1.67	115.06
20	OE_1	0.51	18.89	3.51	30.40
	OE_2	-0.30	22.28	4.07	63.07
	OE_3	-0.17	19.45	5.35	42.43
	OE_4	-0.20	34.41	1.73	117.94

* The observer Standard Deviations are defined in footnotes on Page 326

TABLE V (Cont'd)

Means and Dispersions of Range and Deflection

Center of Impacts For a Target of Radius 25 Meters

Range

$\hat{\sigma}_y$ = 150 meters

Deflection Rd.-to-Rd. Std. Dev. ($\hat{\sigma}_x$)	Observer* Standard Deviation ($\hat{\sigma}_{ye}$ & $\hat{\sigma}_{xe}$)	OT Angle (Radians) 0		.4	
		Average Center of Impact	Standard Deviation of CI's	Average Center of Impact	Standard Deviation of CI's
1	OE_1	-14.16	117.10	-5.88	110.30
	OE_2	-2.60	218.09	-3.12	177.98
	OE_3	-25.06	142.83	-12.32	135.53
	OE_4	-8.78	424.66	11.48	368.45
20	OE_1	-15.35	113.26	-7.44	111.62
	OE_2	-6.04	215.40	-3.63	191.63
	OE_3	-18.78	144.91	-12.97	134.99
	OE_4	-10.02	404.01	15.94	405.14

Deflection

$\hat{\sigma}_y$ = 150 meters

1	OE_1	0.05	3.00	4.53	25.19
	OE_2	0.00	7.84	4.73	66.53
	OE_3	-0.06	3.60	7.25	43.80
	OE_4	0.21	18.21	0.83	120.54
20	OE_1	0.28	19.33	3.44	31.24
	OE_2	-0.37	22.59	4.96	73.02
	OE_3	0.17	19.98	5.12	47.59
	OE_4	0.44	34.81	0.34	125.48

* The observer Standard Deviations are defined in footnotes on Page 326

TABLE V (Cont'd)

Means and Dispersions of Range and Deflection

Center of Impacts For a Target of Radius 25 Meters

Range

$\hat{\sigma}_y$ = 200 meters

Deflection Rd.-to-Rd. Std. Dev. ($\hat{\sigma}_x$)	Observer* Standard Deviation ($\hat{\sigma}_{ye}$ & $\hat{\sigma}_{xe}$)	Of Angle (Radians) 0		.4	
		Average Center of Impact	Standard Deviation of CI's	Average Center of Impact	Standard Deviation of CI's
1	OE_1	-18.14	148.54	-4.76	139.40
	OE_2	-14.21	247.62	-2.85	214.57
	OE_3	-28.24	178.58	-12.45	164.59
	OE_4	-8.60	471.61	2.74	397.00
20	OE_1	-13.14	146.54	-7.57	141.25
	OE_2	-10.14	246.22	2.98	217.09
	OE_3	-26.63	177.32	-7.57	168.67
	OE_4	4.33	453.49	3.07	395.03

Deflection

$\hat{\sigma}_y$ = 200 meters

1	OE_1	-0.04	3.00	4.74	25.43
	OE_2	-0.01	7.47	5.63	73.69
	OE_3	-0.03	3.59	6.40	48.04
	OE_4	0.00	19.59	-0.20	131.96
20	OE_1	-0.32	19.03	4.20	31.20
	OE_2	0.39	22.66	3.14	75.80
	OE_3	-0.19	19.58	4.48	51.31
	OE_4	0.65	35.86	1.27	134.61

* The observer Standard Deviations are defined in footnotes on Page 326

TABLE VI

Means and Dispersions of Range and Deflection

Center of Impacts For a Target of Radius 100 Meters

Range

$\hat{\sigma}_y$ = 25 meters

Deflection Rd.-to-Rd. Std. Dev. ($\hat{\sigma}_x$)	Observer* Standard Deviation ($\hat{\sigma}_{ye}$ & $\hat{\sigma}_{xe}$)	OT Angle (Radians) 0		.4	
		Average Center of Impact	Standard Deviation of CI's	Average Center of Impact	Standard Deviation of CI's
1	OE_1	0.69	59.81	-0.64	75.95
	OE_2	12.06	205.04	17.76	193.44
	OE_3	-0.33	129.27	1.17	115.27
	OE_4	20.59	432.14	20.79	462.34
20	OE_1	-0.51	59.28	0.25	72.41
	OE_2	10.79	214.46	10.34	192.52
	OE_3	2.84	128.65	-0.70	118.71
	OE_4	3.33	427.99	19.17	453.15

Deflection

$\hat{\sigma}_y$ = 25 meters

1	OE_1	0.04	3.94	0.10	26.26
	OE_2	0.28	16.34	-5.02	75.11
	OE_3	-0.08	6.91	-0.80	43.50
	OE_4	0.24	31.07	-4.08	144.18
20	OE_1	-0.35	21.05	0.20	32.29
	OE_2	0.27	26.65	-3.24	78.14
	OE_3	-0.26	22.06	0.71	47.30
	OE_4	1.57	43.85	-2.52	139.04

* The observer Standard Deviations are defined in footnotes on Page 326

TABLE VI (Cont'd)

Means and Dispersions of Range and Deflection

Center of Impacts For a Target of Radius 100 Meters

Range

$\hat{\sigma}_y$ = 50 meters

Deflection Rd.-to-Rd. Std. Dev. (σ_x)	Observer* Standard Deviation (σ_{ye} & σ_{xe})	Of Angle (Radians) 0 Average Center of Impact	0 Standard Deviation of CI's	.4 Average Center of Impact	.4 Standard Deviation of CI's
1	OE_1	-0.46	69.83	-0.39	82.86
	OE_2	11.19	206.22	17.92	204.53
	OE_3	-1.34	131.12	0.26	125.24
	OE_4	20.10	435.05	8.83	413.07
20	OE_1	0.04	70.42	-0.06	84.46
	OE_2	14.62	211.28	15.87	199.47
	OE_3	-0.12	133.63	0.05	122.20
	OE_4	12.03	400.08	13.99	440.17

Deflection

$\hat{\sigma}_y$ = 50 meters

1	OE_1	-0.00	3.88	-0.08	26.54
	OE_2	0.32	16.13	-5.12	77.96
	OE_3	-0.22	7.14	0.53	45.46
	OE_4	-0.05	30.21	-1.29	135.40
	OE_1	0.11	20.99	0.01	32.95
	OE_2	-0.29	26.70	-4.83	79.53
	OE_3	0.14	22.10	0.27	48.32
	OE_4	-0.31	44.13	-2.68	143.17

* The observer Standard Deviations are defined in footnotes on Page 326

TABLE VI (Cont'd)

Means and Dispersions of Range and Deflection

Center of Impacts For a Target of Radius 100 Meters

Range

$\hat{\sigma}_y$ = 100 meters

Deflection Rd.-to-Rd. Std. Dev. (σ_x)	Observer* Standard Deviation ($\hat{\sigma}_{ye}$ & $\hat{\sigma}_{xe}$)	Qʻl Angle (Radians) 0		.4	
		Average Center of Impact	Standard Deviation of CI's	Average Center of Impact	Standard Deviation of CI's
1	OE_1	-9.58	101.55	-7.15	106.64
	OE_2	4.75	231.91	9.61	219.01
	OE_3	-10.68	160.91	-6.00	148.55
	OE_4	17.33	470.25	15.20	449.81
20	OE_1	-8.34	100.41	-7.16	105.14
	OE_2	9.27	239.88	14.30	228.55
	OE_3	-11.43	157.05	-5.24	147.40
	OE_4	-1.75	445.52	5.20	443.25

Deflection

$\hat{\sigma}_y$ = 100 meters

1	OE_1	0.05	3.88	2.98	29.44
	OE_2	-0.07	15.43	-1.68	80.72
	OE_3	-0.00	7.02	0.93	49.97
	OE_4	0.01	28.27	-4.13	145.69
	OE_1	-0.17	20.48	3.58	35.01
	OE_2	0.77	26.66	-2.19	83.33
	OE_3	0.29	21.75	2.51	53.71
	OE_4	-0.25	41.96	-0.91	144.85

*The observer standard Deviations are defined in footnotes on Page 326

TABLE VI (Cont'd)

Means and Dispersions of Range and Deflection

Center of Impacts For a Target of Radius 100 Meters

Range

$\hat{\sigma}_y$ = 150 meters

Deflection Rd.-to-Rd. Std. Dev. ($\hat{\sigma}_x$)	Observer* Standard Deviation ($\hat{\sigma}_{ye}$ & $\hat{\sigma}_{xe}$)	OT Angle (Radians) 0 Average Center of Impact	0 Standard Deviation of CI's	.4 Average Center of Impact	.4 Standard Deviation of CI's
1	OE_1	-7.77	132.62	-6.61	136.23
	OE_2	8.70	256.10	12.93	241.38
	OE_3	-9.37	180.40	-6.63	171.43
	OE_4	13.78	456.82	13.01	458.73
20	OE_1	-6.93	130.43	-3.93	135.69
	OE_2	1.25	250.75	15.09	243.93
	OE_3	11.86	176.42	-7.71	173.14
	OE_4	10.98	480.80	18.54	479.77

Deflection

$\hat{\sigma}_y$ = 150 meters

1	OE_1	0.03	3.84	2.08	30.68
	OE_2	0.17	15.17	-2.22	81.85
	OE_3	0.02	5.85	1.87	53.71
	OE_4	0.73	33.31	-1.39	147.95
20	OE_1	0.05	20.24	1.27	35.28
	OE_2	-0.00	26.80	-2.26	86.86
	OE_3	-0.08	21.80	1.07	58.02
	OE_4	0.56	44.52	-2.03	152.15

* The observer Standard Deviations are defined in footnotes on Page 326

TABLE VI (Cont'd)

Means and Dispersions of Range and Deflection

Center of Impacts For a Target of Radius 100 Meters

Range

$\hat{\sigma}_y$ = 200 meters

Deflection Rd.-to-Rd. Std. Dev. (σ_x)	Observer* Standard Deviation ($\hat{\sigma}_{ye}$ & $\hat{\sigma}_{xe}$)	OT Angle (Radians) 0 Average Center of Impact	0 Standard Deviation of CI's	.4 Average Center of Impact	.4 Standard Deviation of CI's
1	OE_1	-11.13	166.80	3.40	167.65
	OE_2	7.52	285.29	10.84	260.13
	OE_3	-11.25	208.85	-8.22	201.08
	OE_4	14.69	528.87	12.77	509.48
20	OE_1	-9.22	161.78	-1.75	166.42
	OE_2	2.04	279.94	15.98	267.05
	OE_3	-8.76	207.45	-8.67	201.25
	OE_4	17.54	498.78	4.12	540.15

Deflection

$\hat{\sigma}_y$ = 200 meters

1	OE_1	-0.02	3.43	1.91	30.02
	OE_2	0.13	14.51	0.39	87.71
	OE_3	-0.02	6.13	2.67	59.49
	OE_4	0.84	32.19	-3.65	154.20
20	OE_1	-0.13	20.38	0.87	36.00
	OE_2	0.13	26.71	-3.96	89.88
	OE_3	-0.67	21.27	1.91	60.26
	OE_4	-0.69	44.17	-0.67	156.92

* The observer Standard Deviations are defined in footnotes on Page 326

FOOTNOTES

* The observer Errors are defined below:

OE_1: $\hat{\sigma}_{ye} = .2\ (y)$; $\hat{\sigma}_{xe} = .1\ (x)$

OE_2: $\hat{\sigma}_{ye} = .2(4y)$; $\sigma_{xe} = .1\ (4x)$

OE_3: $\hat{\sigma}_{ye} = .5\ (y)$; $\hat{\sigma}_{xe} = .2\ (x)$

OE_4: $\hat{\sigma}_{ye} = .5(4y)$; $\hat{\sigma}_{xe} = .2\ (4x)$

where: y and x are the range and deflection miss distances of the impact point from the target center, respectively. OE_1 and OE_3 are used when the OT distance is less than 1000 meters while OE_2 and OE_4 are used when the OT distance is between 3000 and 4000 meters.

QUICK REACTION STUDY OF CALIBRATION DRIFT IN RADIACMETER IM-174()

Joseph Weinstein
Systems/Cost Analysis Office
US Army Electronics Command
Fort Monmouth, New Jersey

1. BACKGROUND

My purpose is to show how the use of an experimental design:

a. Verified an allegation that a problem of "drift" in calibration existed in the Radiacmeter IM-174().

b. Estimated the size of the drift (which was useful in recommending a "quick fix" on the problem).

c. Satisfied a time constraint given to the commodity manager for responding to the allegation.

d. Identified the causitive factor for corrective action attention.

e. Saved engineering test samples, facility test time and time costs of skilled personnel of the order of 90% of the conventional approach that would have been taken without the "short fuse" time constraint.

The urgency for resolution of the problem rested on the fact that the instrument was authorized for issue and was in short supply. Two production lines had been stopped by the contracting officer when the allegation of fault was made and he was assuming penalty payments while waiting for engineering guidance for termination, engineering changes or continuance of the production.

It would be incorrect to think that our engineer-managers immediately recognized that they had a problem where the services of a design statistician were needed. Like those engineers in other laboratories and engineering organizations they had prepared a conventional plan of study. It involved some 1500 instruments which would be tested under "crash" programs and a minimum time of six months would be used to obtain the data necessary for a recommendation. Since the penalty costs were in the neighborhood of $1000 or $1500 a day on each contract the commodity manager turned to the last source of help - the statistician. We held several short orientation meetings and studied

background documents, test reports and some recent test data from the Lexington Army Depot on 12 instruments which had concluded that the "drift" problem was real and primarily found in instruments from Depot Stock (IM 108/U) which had been modified and "improved" at the depot and had been recalibrated shortly before these tests were made. This report also thought that the causes were to be found with the source of the instrument, the length of time it had been stored and the "aging" cycle that had been experienced by the #5886 electrometer tube during its manufacturing cycle. ("Expert" tube engineers were so sure of this cause that Raytheon Corp. offered to speedily redesign the tube itself if we at EOM would assume the costs.)

A careful review of the data on these 12 instruments in conjunction with a physical examination showed that the readings were made on a logarithmic scale (covering 2.5 decades of radiation intensity in 2.25 inches of arc.) - strongly suggested that the "drift" could easily be associated with biased reading error.

Since the readings were made at a radiation level where large reading errors could occur in drift-free instruments and since a "run" of 6 readings, each smaller than its predecessor, had a small, but not infinitesimal, probability of occurrance with drift-free instruments, there was sufficient basis for deferring judgment until additional information from a well-controlled test was available, if it could be acquired quickly. It was agreed that a verification experiment ought to be done before funding the redesign of the tube. If possible, it was also desirable to estimate the influence of the source of manufacture, the aging cycle used in the 5886 tube, the time spent in storage prior to test, and (as an after thought) a simulation of the experience of solar heating in storage depots by a temperature cycling of 4 hours at 160°F in every 24 hours of storage.

2. THE TEST PLAN:

The test plan developed is given below. It uses 128* instruments instead of the 1500 considered and the entire program of testing is completed in 8 weeks with a report submitted by the 10th week instead of the 26 weeks of testing proposed. It considers more factors, estimating their joint effects, and gives reasonably precise estimates of each factor's influence on the drift observation. All of these outputs being obtained with roughly 10% of the effort proposed by the engineering group were completed and submitted to the program officer by the 10th week after initiation.

*These were to be from 2 producing sources however a third source was developed and an additional 64 instruments were made available.

A detailed test plan had to be prepared for the remote testing facility (at the Lexington (Ky) Signal Depot). The plan covered 5 pages of single spaced information with step-by-step procedures for selection, pre-conditioning initial calibration, and the test environment to be assigned to each sample unit. Since the facilities for test performance were limited (and the time for completion as well) it was necessary to consider tight time scheduling for each test sample unit and do partial data analysis while tests were continuing. All details relating to these considerations are shown in plates 1-9 inclusive.

The factors and levels considered were:

Storage time	1	(1)	8	weeks
Producing source	1,	2,	3	(8 units from each source per week)
Aging schedule of	24	(24)	96	hours of additional aging

Temperature cycling room temperatures vs 4 hours at 160°F

Plan of the Data Analysis.

The measure of "instability" for the test sample unit was the absolute value of the change in reading of the instrument between the calibration value at the start of the experiment and the measurement at the end of the scheduled test time. After this procedure had established the existence of a "drift", the arithmetic value of the change could be examined for the direction and magnitude of "drift." These measures of "drift" were made for three levels of the radiation 400 r/hr, 200 r/hr, and 50 r/hr. Results from the experiment became available weekly providing the analyst with 16 (24) randomly chosen instruments which had been exercised by the test plan "w" weeks (w = 1 to 8). Of these 8 instruments were from each producing source; 2 of each 8 had electrometer tubes at each of the tube aging schedules and 4 of each 8 (1 at each of the 4 tube aging schedules) were given the temperature cycle exposure. These were subjected to analysis of variance techniques for the estimation of each factor's influence on the drift phenomenon.

3. TEST RESULTS AND ANALYSIS.

(a) Pre-test Calibration data.

Prior to exposure of the test units to the schedule of test conditions developed, it was necessary to verify their compliance with specified performance requirements. This is indicated graphically by plates 10, 11, and 12 each for a different radiation exposure level and each depicting the distribution of the samples from the three producing

sources. With minor exceptions the test units comply with specific values. It is also observed that whereas the 200 r/hr data are closely grouped and symmetric about the design value the 400 r/hr data and the 50 r/hr data exhibit broader spreads and asymmetry. The depot test units suggest the possibility of design differences from the other two producing sources since the direction of skewness for these units is opposite to that of the other producers.

(b) Catastrophic Failures.

An unexpected but quite important result was the observation that a large proportion of the instruments failed to perform after exposure to the test environment. The weekly data seemed to be independent of time but heavily dependent on the temperature cycling exposure. Summarized data is given in Table 1.

TABLE I: Catastrophic Failures (32 Tested at each condition):

SOURCE OF TEST UNIT	TEMPERATURE	
	ROOM	4 HRS at 160°F Daily
Depot	0	15
LF&C	2	13
Victory	1	10
Percent Failure	$\frac{3}{96}$ = .031	$\frac{38}{96}$ = .395

(c) Temperature Cycling and Time.

Plate 13 illustrates the strong effect of the temperature exposure over the 8 weeks of testing. This effect is independent of the radiation level and seems to be disappearing at the 8th week of exposure. Unfortunately the testing was so tightly scheduled and the need for decision about the drift problem limited the opportunity to pursue this conjecture to a decision. However, the importance of temperature cycling to change in calibration and as a cause of catastrophic failures is unequivocal. The design also protected against bias due to scale shifts with time which appear sinusoidal.

(d) Production Sources and Time.

Plate 14 illustrates the relatively random influence of the producing source as a cause of change from calibration values.

(e) Aging Schedule and Time.

Plate 15 shows how drift with time was influenced by the prior aging of the 5886 electrometer tube before incorporation in the test sample. Here again we find relative shifting of position of the symbols associated with prior tube aging so that no indication of a tube effect appears to exist.

(f) Calibration "Drift" with Time by Producing Source and Radiation Level.

The change between pre-test calibration and post test instrument reading was regressed on the time axis for data derived from the 8 weeks of testing. This was done for each level of radiation exposure and producing source with separate regression lines for the instruments held at room temperature and for those having been temperature cycled. The results are given in plates 16, 17, & 18 for a 400 r/hr radiation level and in plates 19, 20, & 21 for 200 r/hr radiation level. A listing of the regression coefficients in the Table II below shows positive valued coefficients for time and negative constants (initial values) throughout, with the exception of the Victory room temperature data at 400 r/hr.

4. SUMMARY:

The graphic studies of the averaged absolute changes in instrument readings with time at 3 points on the radiation scale (400, 200 and 50 rads/hour) are presented as:

Plate 13, DM-174 Study of Temperature Cycling)
Plate 14, DM-174 Study of Source Differences) These figures are based on Depot units and Landers, Frary & Clark units only. However, they are believed to apply as well to Victory units.
Plate 15, DM-174 Study of Electrometer Tube Aging)

From these plates - 13, 14, 15, it is quite apparent, from the weekly rearrangement of the order of the averaged absolute changes for different amounts of aging (Plate 15) and for sources (Plate 14), that there is no consistent difference in performance of the IM-174() instruments for these factors. From Plate 13, however, it is quite apparent that temperature cycling plays a very important role in changes from original calibration values. These appraisals are independently confirmed by Analyses of Variance performed on each group of 16 reports which are available in the files of the undersigned but are not submitted as part of this report. Having established that a well defined effect was exhibited between instruments temperature cycled and those held at room temperature over the 8 week study the nature of this effect was studied more exactly by fitting trend lines to the observed changes including the sign of the change. This is shown in Plates 16, 17, 18, and 19, 20, 21. It is apparent that a positive* time drift exists. The validity of these trend coefficients - in terms of adequately describing the effect of the assumption that a linear drift with time existed in the data - was tested by performing a variance analysis on the fitted curves. The results imply a high degree of verification of the assumed linear trend in the case of units stored at room temperature with residual (unexplained) variances which appear to be reasonable as estimates of the instrument reading error. However, in the case of units temperature cycled during storage, such verification is somewhat clouded by the loss of a large proportion of the instruments and by the potential of the temperature cycling to have stimulated several different modes of change (or failure) not present in the room temperature environments. In any event there is a considerable increase in residual variation about the trend lines fitted to such data and suggesting a reading error of the order 2 to 3 times the values obtained in the room temperature fittings. These results are tabulated in Table II.

5. RECOMMENDATIONS AND CONCLUSIONS:

The alleged "drift" of calibration with storage time is confirmed by this experiment. However its effect is so small relative to influence of temperature cycling and possibly poor manufacturing control as evidenced by Table I, that no further modification effort in this area is recommended until the gross effects to be reported in the ongoing engineering study by Atomics Branch, USAELRDL are corrected and controlled. The utility of the present instrument in field use can be measurably improved by: (1) scheduling a 3-4 month recalibration cycle instead of the present 6 month cycle. This will keep the instruments within their specified tolerances, (2) maintain them in storage conditions which will avoid the thermal cycling exposure of the test specification until modifications proposed by the engineering report can be made.

*The Lexington Army Depot data on 12 instruments tested in the spring of 1963 and reported in May 1963 correspondence to USAEMSA indicated a negative drift.

TABLE II: CHANGE AS A FUNCTION OF TIME, TYPE OF STORAGE, AND LEVEL OF RADIATION

Type of Storage	Source of Units	Fitted Trend Line Equation Change in (r/hr)	Are Slope Coefficients Statistically Significant?	Estimate of Reading Error (r/hr)
Room Temp.	400 4/hr			
	Landers, Frary & Clark	5.9W - 20.70	Yes	17
	Depot	3.98W - 4.32	Yes	15
	Victory	7.50W + 5.21	Yes	22
	200 4/hr			
	Landers, Frary & Clark	1.24W - 5.21	Yes	7
	Depot	1.32W - 2.28	Yes	6
	Victory	2.48W - 2.28	Yes	8

Type of Storage	Source of Units	Fitted Trend Line Equation Change in (r/hr)	Are Slope Coefficients Statistically Significant?	Estimate of Reading Error (r/hr)
	400 r/hr			
Temp Cycling at 160°F	Landers, Frary & Clark	4.28W - 58.26	No	22
	Depot	6.14W - 55.28	Marginal	36
	Victory	5.18W - 77.38	Marginal	30
	200 r/hr			
	Landers, Frary & Clark	4.34W - 44.13	Yes	12
	Depot	3.00W - 33.03	No	22
	Victory	1.71W - 34.00	No	13

TEST PLAN FOR IM-174 RADIACMETER

The 128 units are to be conditioned before measurement in accordance with the attached schedule which has been designed so that: (1) 64 units will come from Depot and 64 from Landers, Frary & Clark; (2) 64 units will be preconditioned in a normal warehouse temperature environment and 64 will be preconditioned in accordance with the elevated temperature cycle at 160° F for 4 hours of every 24 hours during its storage period; (3) 32 units will be preconditioned by ageing or "burn-in" of the instrument for 24 hours, 32 for 48 hours, 32 for 72 hours, and 32 for 96 hours; (4) 16 units will be preconditioned by having been stored without use since last calibration for 1 week similarly 16 each for 2, 3 ---7, and 8 weeks.

The specific preconditioning of each test unit is obtained from the schedule attached by reading across a row. It is extremely important that there be no deviations made from these preconditioning requirements. A procedure for selecting the units for the required preconditioning willl be given late in these instructions. The data sheets* to be kept for each instrument during the testing should show the: (1) date of calibration of the instrument; (2) the calibration points and the instrument reading for each point; (3) name of operator making the calibration; (4) any comments by operator about adjustments to test jigs, source, or other conditions of the test environment.

*Edge-punched cards would be very useful in the data analysis phase.

Plate 1

As soon as one sub-group is filled by the randomization scheme used, no further assignments are to be made to that sub-group so that we will finally have 32 instruments in each ageing sub-group.

The procedure to be followed in making this assignment to sub-groups should be applied first to the group coded 00XX so that the third code digit can be assigned (0 to 24 hours, 1 for 48 hours, 2 for 72 hours, and 3 for 96 hours) giving

000X for 8 test instruments
001X for 8 test instruments
002X for 8 test instruments
003X for 8 test instruments

In turn, applying this procedure to the three other codes 01XX, 10XX, and 11XX will give 16 sub-groups of 8:

010X	100X	110X
011X	101X	111X
012X	102X	112X
013X	103X	113X

Plate 2

The measurement of all 128 units is to be completed on two work days with a particular block of 32 to be done on the morning of the first day, a second assigned block in the afternoon of that day, a third block on the morning of the second day and the last block in the afternoon of the second day. It is understood that up to 12 units can be set up simultaneously for measurement, and that readings can be taken at three points of the scale (radiation rates) on all twelve instruments in one hour. The instruments provided for this test could best be subdivided into four groups of eight with the remaining four positions filled with other calibration activities of the Lexington Depot if they desire to utilize all twelve positions available. If the Lexington Depot plans to run other instruments at the same time as those involved in this test procedure, great care must be exercised to avoid mixing in data from the other instruments.

To obtain suitable instruments for this test procedure:

A. Selection. Acquire at least 128 IM-174 units which have had no prior "burn-in" or ageing. Sixty-four of these should be obtained from Landers, Frary & Clark and 64 should be from the Depot modification of IM-108U's.

Plate 3

B. Identification Marking. Assign a four-digit code number to each unit. The coding procedure will be developed as we go through the preconditioning procedure. At this point we may use a oxxx to identify Landers, Frary & Clark instruments and ixxx to identify Depot instruments. (The x's will not be marked but merely indicate places that will be marked.)

C. Storage Temperature. For the 64 units coded oxxx--On a purely random basis (say by tossing a coin) separate them into two groups of 32 units each and mark those in one group as ooxx and those in the other as oixx. For the 64 units coded ixxx, again using a random assignment basis, separate them into two groups which will be marked ioxx and iixx.

Those instruments with a "i" in the second digit position from the left will be exposed to 160^0 F. temperatures continuously for four hours out of every 24 while being stored.

D. Ageing Time. Prior to storage, all 128 test units shall have been aged for one of the following four periods: 24 hours, 48 hours, 72 hours, 96 hours. The units to be assigned to a particular ageing schedule will be assigned randomly to four sub-groups (say by tossing two coins and assigning head-head 24 hours, head-tail 48 hours, tail-head 72 hours and tail-tail 96 hours.)

Plate 4

E. Storage Time. The assignment of test units for various periods of storage will again be on a random basis. (This time by tossing three coins--say a penny, nickel, and quarter.) The randomization should be applied in turn to each of the 16 subgroups (of 8) so that each subgroup of the type say 0101X will now be separated into 8 units, one to be stored for each of the following time periods:

1 for 1 week before it is rechecked
1 for 2 weeks before it is rechecked
1 for 3 weeks before it is rechecked
1 for 4 weeks before it is rechecked
1 for 5 weeks before it is rechecked
1 for 6 weeks before it is rechecked
1 for 7 weeks before it is rechecked
1 for 8 weeks before it is rechecked

If we assign codes 0 for 1 week storage
1 for 2 weeks storage
2 for 3 weeks storage
3 for 4 weeks storage
4 for 5 weeks storage
5 for 6 weeks storage
6 for 7 weeks storage
7 for 8 weeks storage

as the 4th code digit we can now identify the treatment to be given to each of the test instruments and also by use of this coding group together, those that are to be tested at any one time in accordance with the accompanying schedule of 4 blocks (I-IV).

RECHECK SCHEDULE BY WEEKS OF STORAGE

Storage Codes: Cell items are run #'s within blocks on calibration test design.

BLOCK	0	1	2	3	4	5	6	7
	1	3	4	11	6	2	15	8
I	14	10	5	12	7	13	16	9
	19	23	28	21	26	20	17	24
	32	30	29	22	27	31	18	25
	3	1	11	4	2	6	8	15
II	10	14	12	5	13	7	9	16
	23	19	21	28	20	26	24	17
	30	32	22	29	31	27	25	18
	4	11	1	3	15	8	6	2
III	5	12	14	10	16	9	7	13
	28	21	19	23	17	24	26	20
	29	22	32	30	18	25	27	31
	11	4	3	1	8	15	2	6
IV	12	5	10	14	9	16	13	7
	21	28	23	19	24	17	20	26
	22	29	30	32	25	18	31	27

At any recheck time the order or recheck will be assigned at random.

Plate 6

RECHECK SCHEDULE BY WEEKS OF STORAGE

Storage Codes (by weeks of storage): Cell items are other preconditionings of the test item.

BLOCK	0	1	2	3	4	5	6	7
	000	001	002	003	110	000	112	002
I	012	110	010	011	001	111	100	010
	111	013	113	112	102	012	003	113
	103	102	101	100	013	103	011	101
	001	000	003	002	000	110	002	112
II	110	012	011	010	111	001	010	100
	013	111	112	113	012	102	113	003
	102	103	100	101	103	013	101	011
III	002	003	000	001	112	002	110	000
	010	011	012	110	100	010	001	111
	113	112	111	013	003	113	102	012
	101	100	103	102	011	101	013	103
	003	002	001	000	002	112	000	110
IV	011	010	110	012	010	100	111	001
	112	113	013	111	113	003	012	102
	100	101	102	103	101	011	103	013

At any recheck time the order of recheck will be assigned at random.

IM-174 TEST DESIGN: CALIBRATION

Layout of Coded Runs by Blocks

RUN	BLOCK #	I	II	III	IV
1	----------	0000	0001	0002	0003
2	----------	0005	0004	0007	0006
3	----------	0011	0010	0013	0012
4	----------	0022	0023	0020	0021
5	----------	0102	0103	0100	0101
6	----------	1104	1105	1106	1107
7	----------	0014	0015	0016	0017
8	----------	0027	0026	0025	0024
9	----------	0107	0106	0105	0104
10	----------	1101	1100	1103	1102
11	----------	0033	0032	0031	0030
12	----------	0113	0112	0111	0110
13	----------	1115	1114	1117	1116
14	----------	0120	0121	0122	0123
15	----------	1126	1127	1124	1125
16	----------	1006	1007	1004	1005
17	----------	0036	0037	0034	0035
18	----------	0116	0117	0114	0115
19	----------	1110	1111	1112	1113
20	----------	0125	0124	0127	0126
21	----------	1123	1122	1121	1120
22	----------	1003	1002	1001	1000
23	----------	0131	0130	0133	0132
24	----------	1137	1136	1135	1134
25	----------	1017	1016	1015	1014

Plate 8

IM-174 TEST DESIGN: CALIBRATION (Continued)

RUN	BLOCK #	I	II	III	IV
26	----------	1024	1025	1026	1027
27	----------	0134	0135	0136	0137
28	----------	1132	1133	1130	1131
29	----------	1012	1013	1010	1011
30	----------	1021	1020	1023	1022
31	----------	1035	1034	1037	1036
32	----------	1030	1031	1032	1033

Runs are to be assigned at random within any block.

Plate 9

400 r/hr

IM - 174

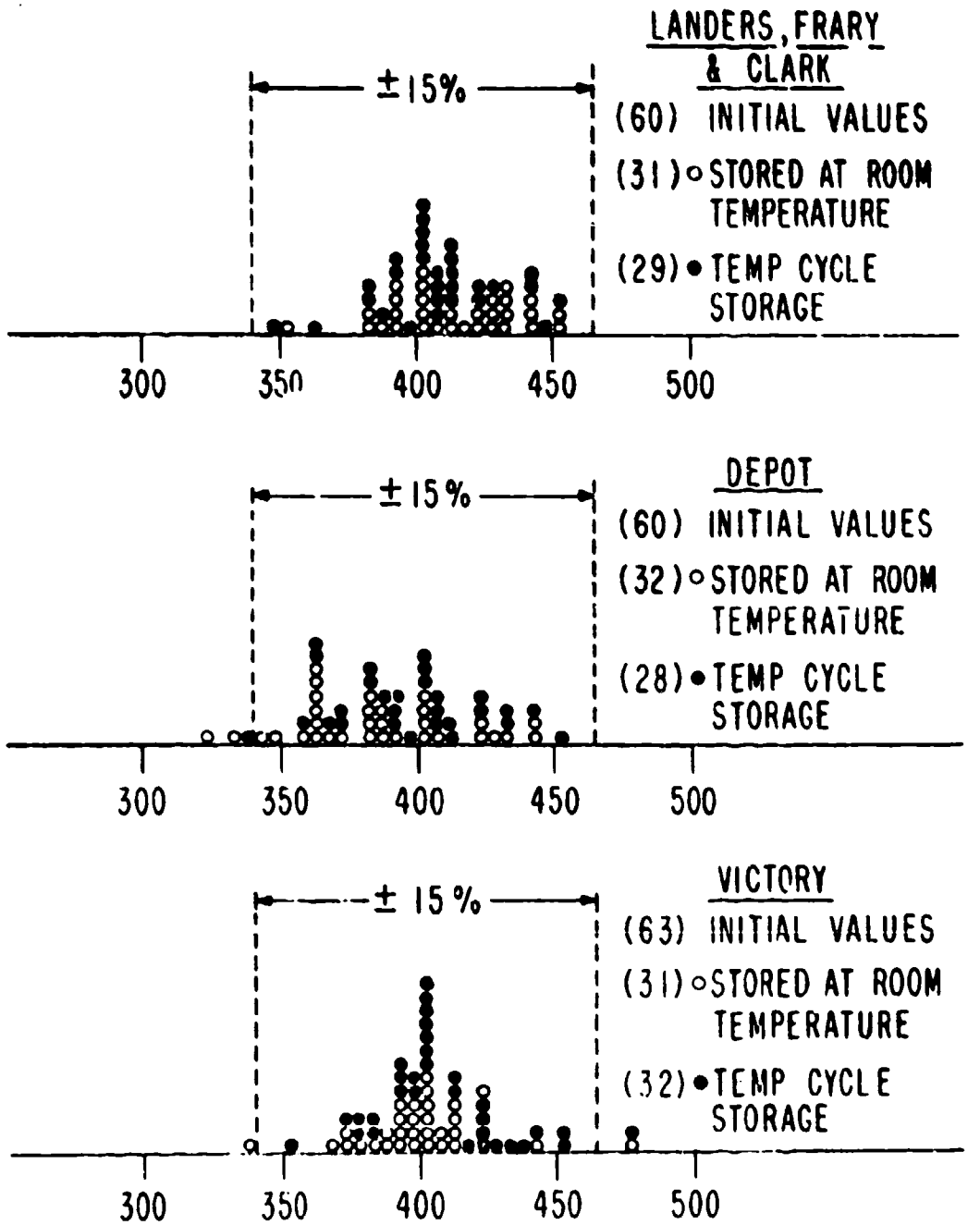


Plate 10

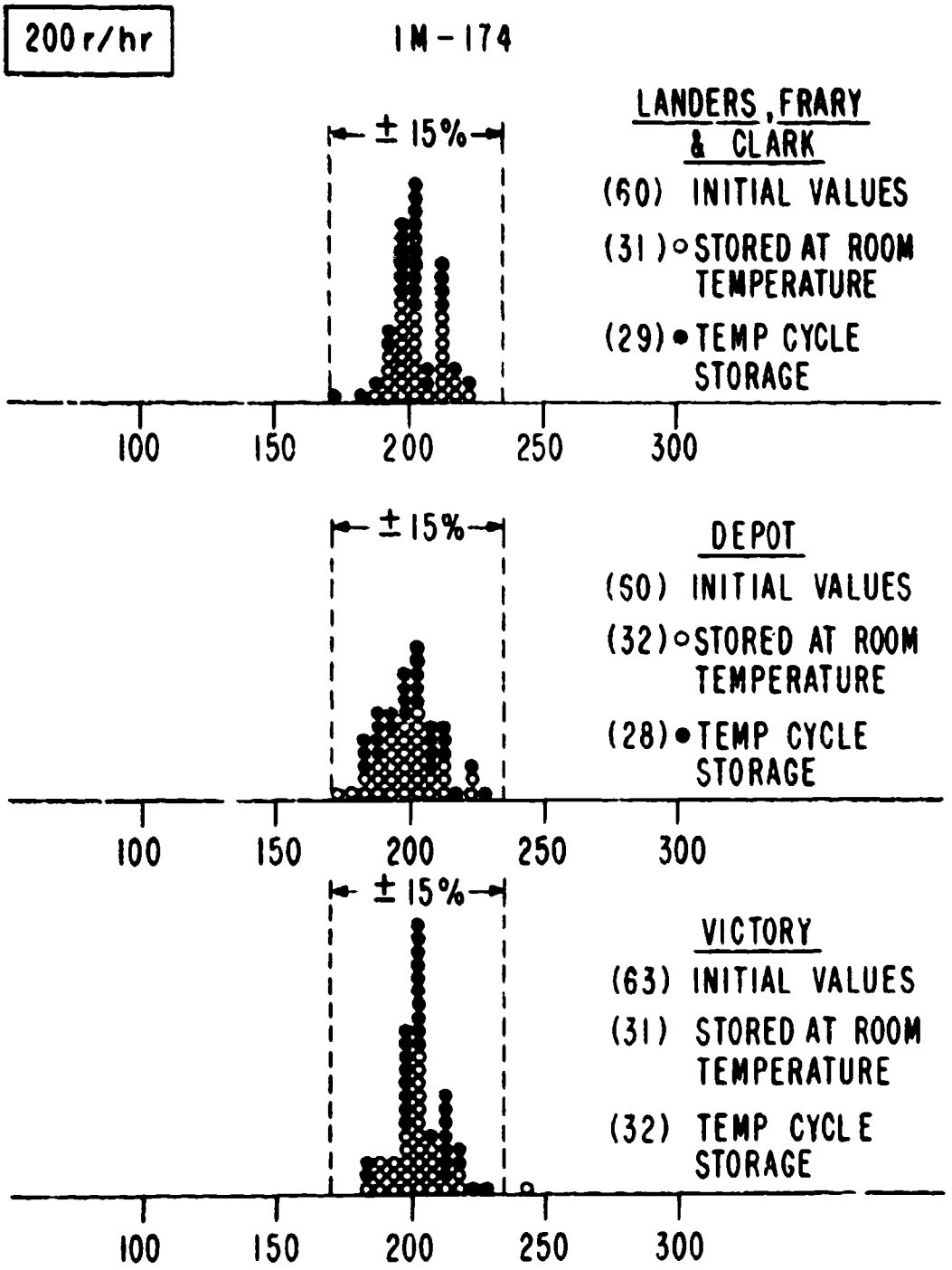

200 r/hr
IM-174
LANDERS, FRARY & CLARK
(60) INITIAL VALUES
(31) ○ STORED AT ROOM TEMPERATURE
(29) ● TEMP CYCLE STORAGE
± 15%
100
150
200
250
300
DEPOT
(50) INITIAL VALUES
(32) ○ STORED AT ROOM TEMPERATURE
(28) ● TEMP CYCLE STORAGE
± 15%
100
150
200
250
300
VICTORY
(63) INITIAL VALUES
(31) STORED AT ROOM TEMPERATURE
(32) TEMP CYCLE STORAGE
± 15%
100
150
200
250
300

Plate 11

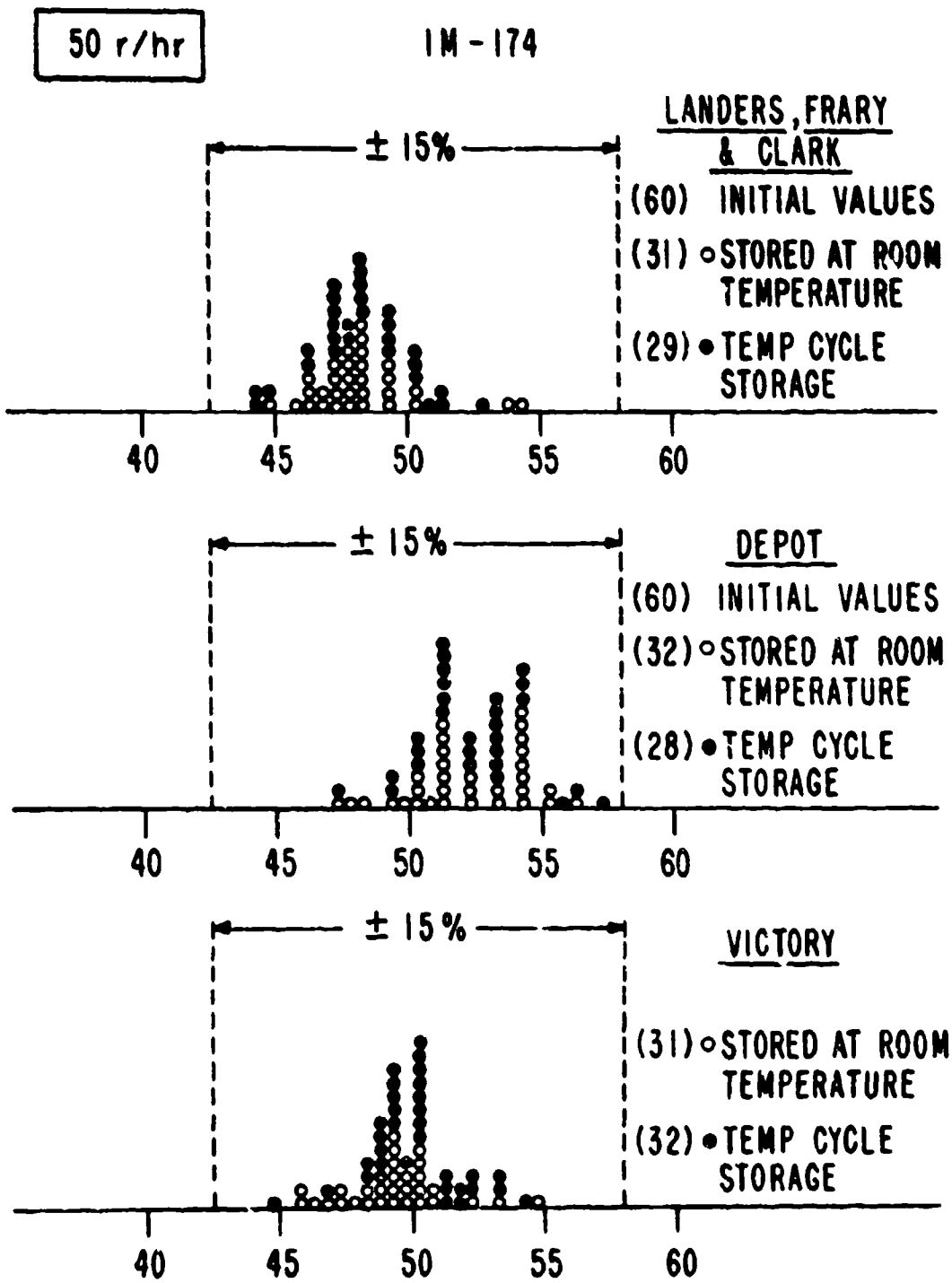

50 r/hr
IM-174
LANDERS, FRARY & CLARK
± 15%
(60) INITIAL VALUES
(31) ○ STORED AT ROOM TEMPERATURE
(29) ● TEMP CYCLE STORAGE
40
45
50
55
60
DEPOT
± 15%
(60) INITIAL VALUES
(32) ○ STORED AT ROOM TEMPERATURE
(28) ● TEMP CYCLE STORAGE
40
45
50
55
60
VICTORY
± 15%
(31) ○ STORED AT ROOM TEMPERATURE
(32) ● TEMP CYCLE STORAGE
40
45
50
55
60

Plate 12

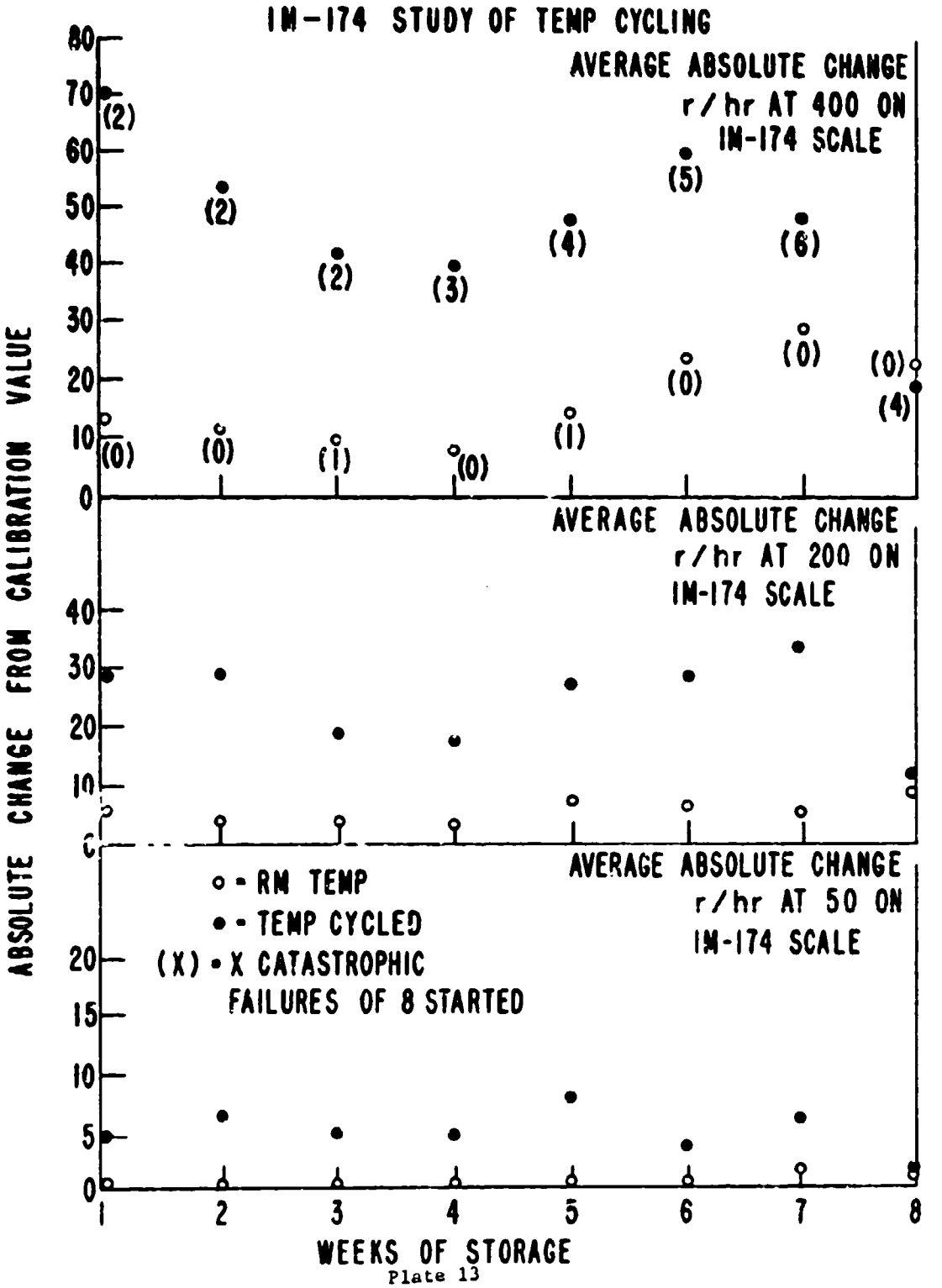

IM-174 STUDY OF TEMP CYCLING
AVERAGE ABSOLUTE CHANGE r/hr AT 400 ON IM-174 SCALE
AVERAGE ABSOLUTE CHANGE r/hr AT 200 ON IM-174 SCALE
AVERAGE ABSOLUTE CHANGE r/hr AT 50 ON IM-174 SCALE
ABSOLUTE CHANGE FROM CALIBRATION VALUE
o - RM TEMP
• - TEMP CYCLED
(X) • X CATASTROPHIC FAILURES OF 8 STARTED
WEEKS OF STORAGE

Plate 13

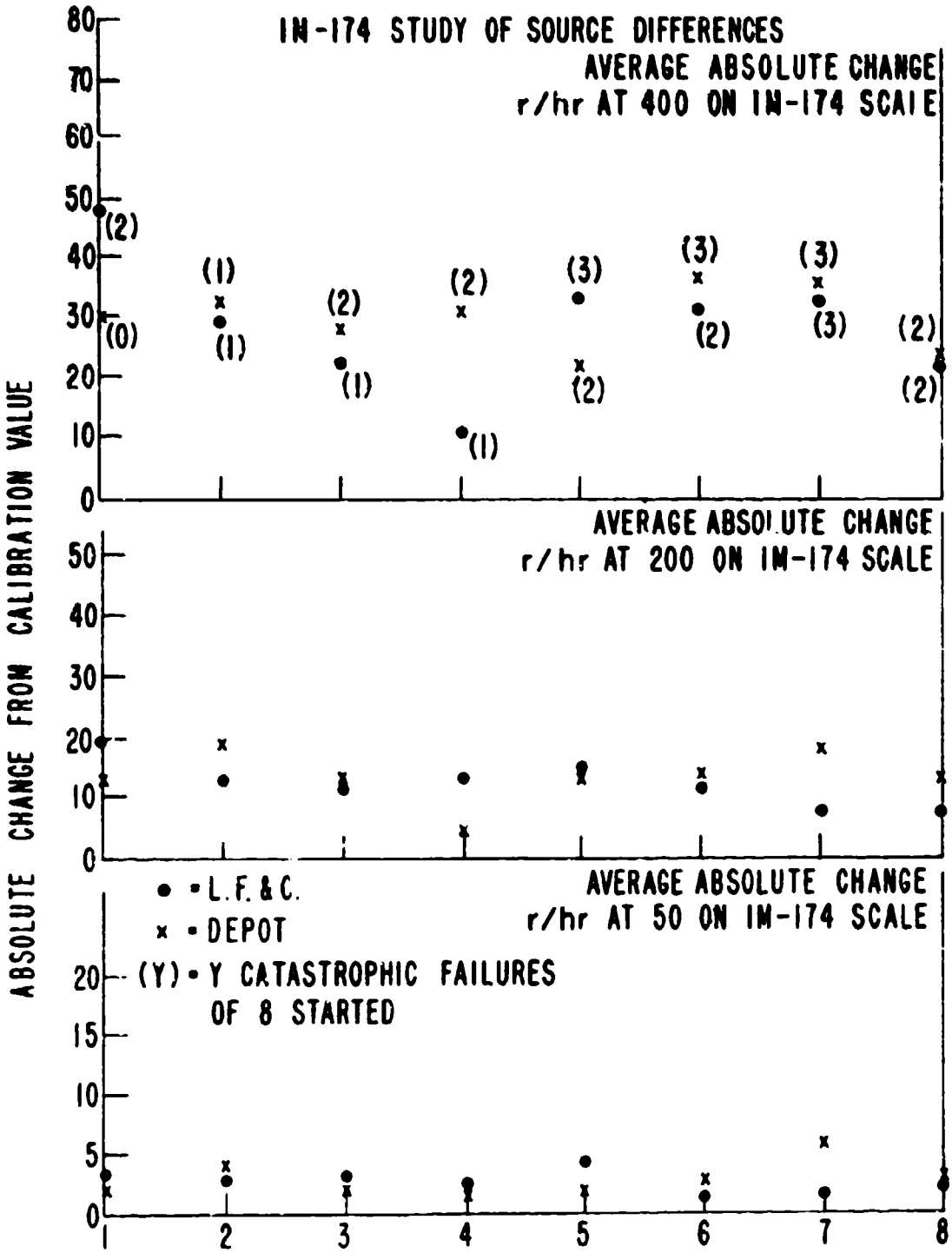

IM-174 STUDY OF SOURCE DIFFERENCES
AVERAGE ABSOLUTE CHANGE
r/hr AT 400 ON IM-174 SCALE
AVERAGE ABSOLUTE CHANGE
r/hr AT 200 ON IM-174 SCALE
AVERAGE ABSOLUTE CHANGE
r/hr AT 50 ON IM-174 SCALE
ABSOLUTE CHANGE FROM CALIBRATION VALUE
● = L.F.&C.
x = DEPOT
(Y) = Y CATASTROPHIC FAILURES OF 8 STARTED
WEEKS OF STORAGE

Plate 14

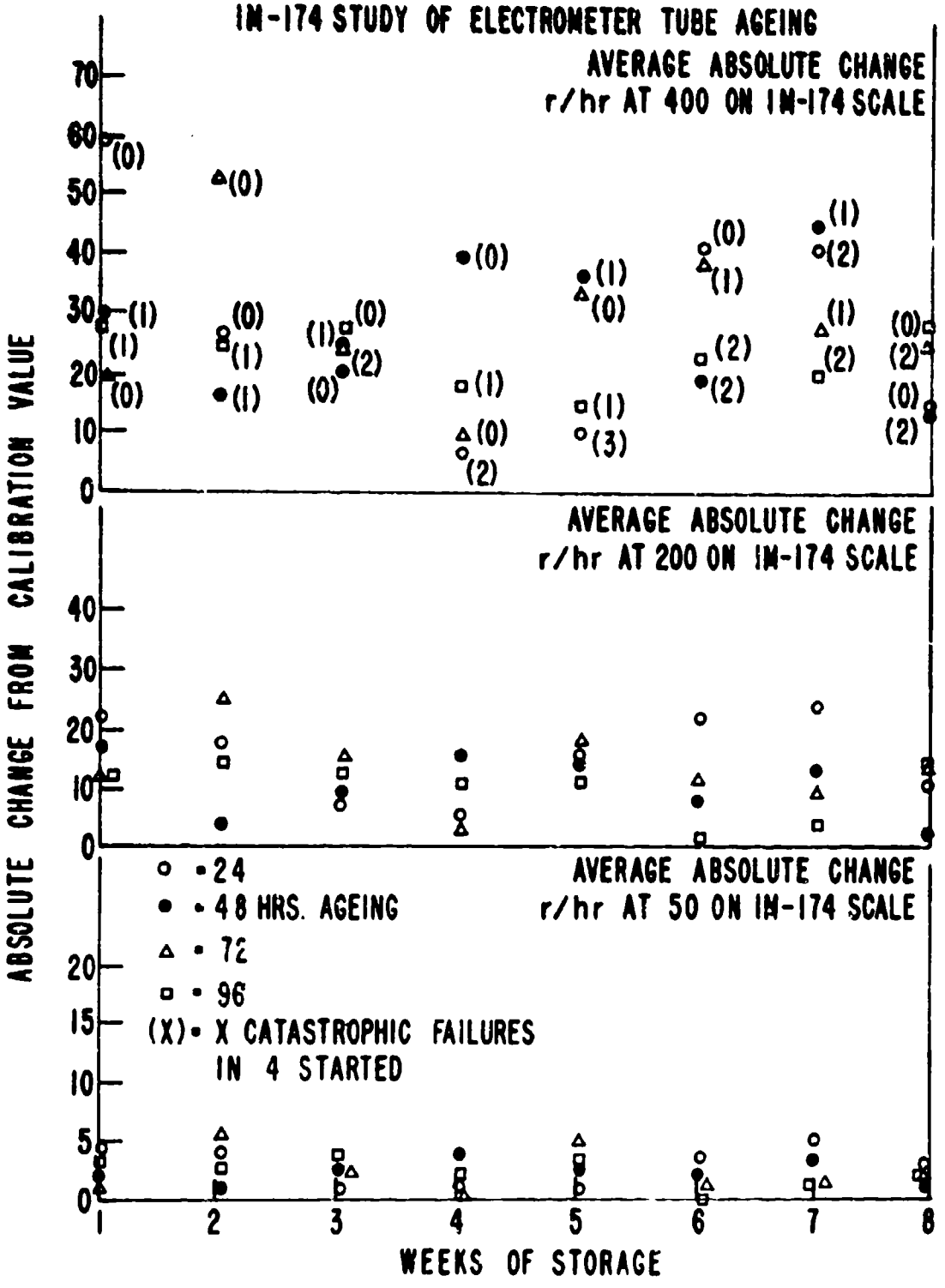

IM-174 STUDY OF ELECTROMETER TUBE AGEING
AVERAGE ABSOLUTE CHANGE
r/hr AT 400 ON IM-174 SCALE
AVERAGE ABSOLUTE CHANGE
r/hr AT 200 ON IM-174 SCALE
AVERAGE ABSOLUTE CHANGE
r/hr AT 50 ON IM-174 SCALE
ABSOLUTE CHANGE FROM CALIBRATION VALUE
○ • 24
● • 48 HRS. AGEING
△ • 72
□ • 96
(X) • X CATASTROPHIC FAILURES
IN 4 STARTED
WEEKS OF STORAGE

Plate 15

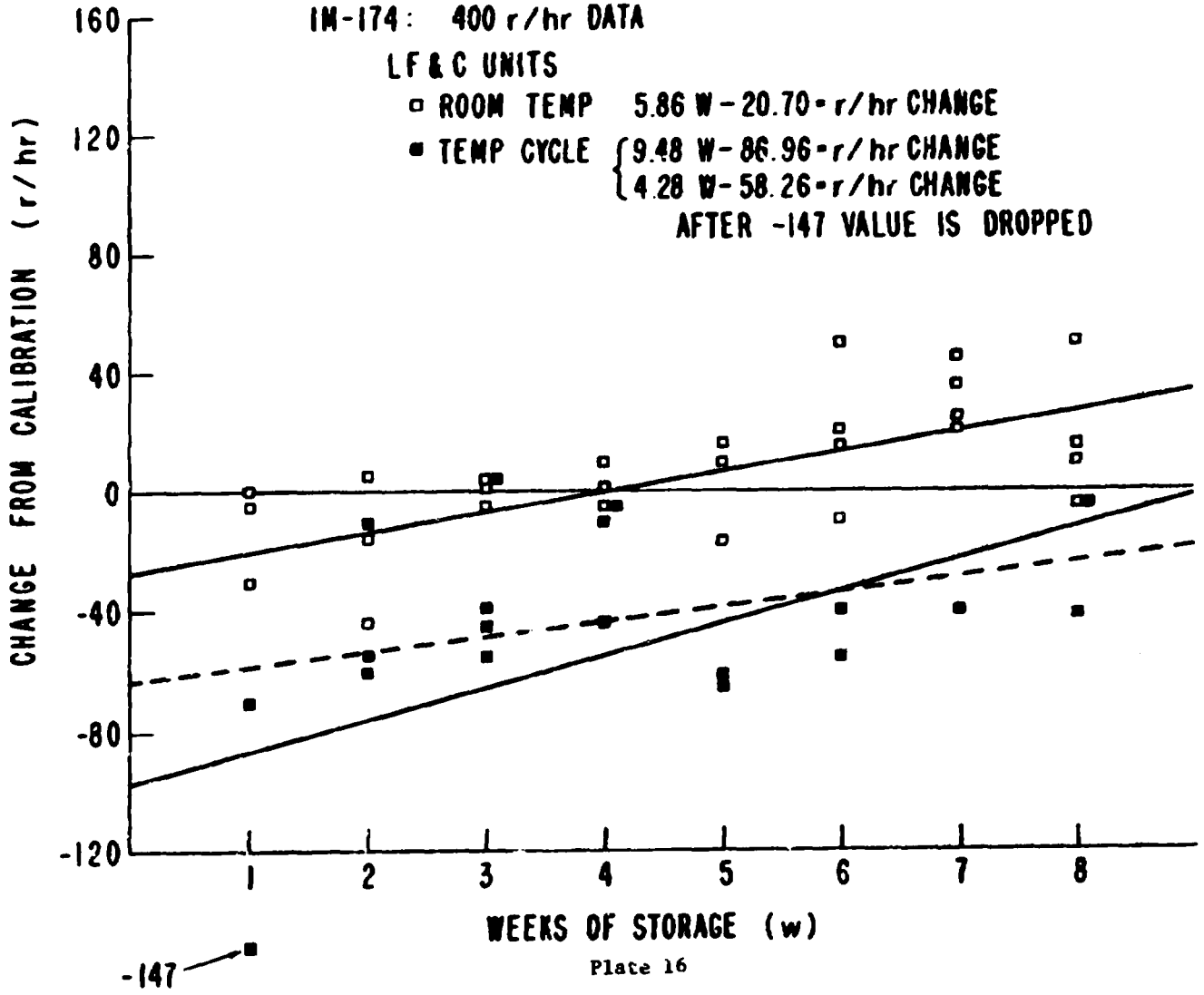

IM-174: 400 r/hr DATA
LF & C UNITS
□ ROOM TEMP 5.86 W - 20.70 • r/hr CHANGE
■ TEMP CYCLE {9.48 W - 86.96 • r/hr CHANGE
{4.28 W - 58.26 • r/hr CHANGE
AFTER -147 VALUE IS DROPPED
CHANGE FROM CALIBRATION (r/hr)
160
120
80
40
0
-40
-80
-120
1
2
3
4
5
6
7
8
WEEKS OF STORAGE (w)
-147

Plate 16

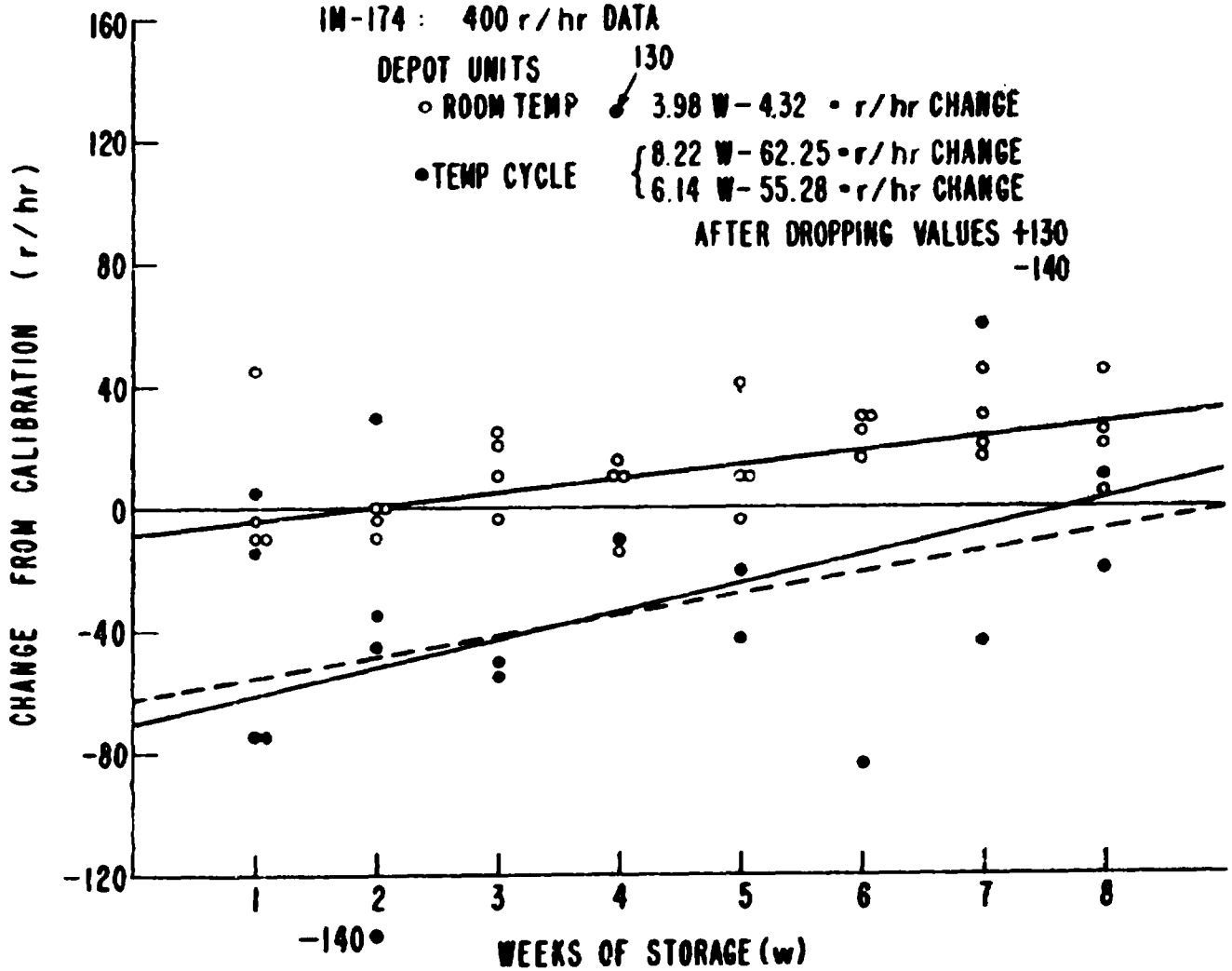

IM-174 : 400 r/hr DATA
DEPOT UNITS
130
o ROOM TEMP
3.98 W - 4.32 • r/hr CHANGE
•TEMP CYCLE
8.22 W - 62.25 • r/hr CHANGE
6.14 W - 55.28 • r/hr CHANGE
AFTER DROPPING VALUES +130
-140
CHANGE FROM CALIBRATION (r/hr)
160
120
80
40
0
-40
-80
-120
1
2
3
4
5
6
7
8
-140
WEEKS OF STORAGE (w)

Plate 17

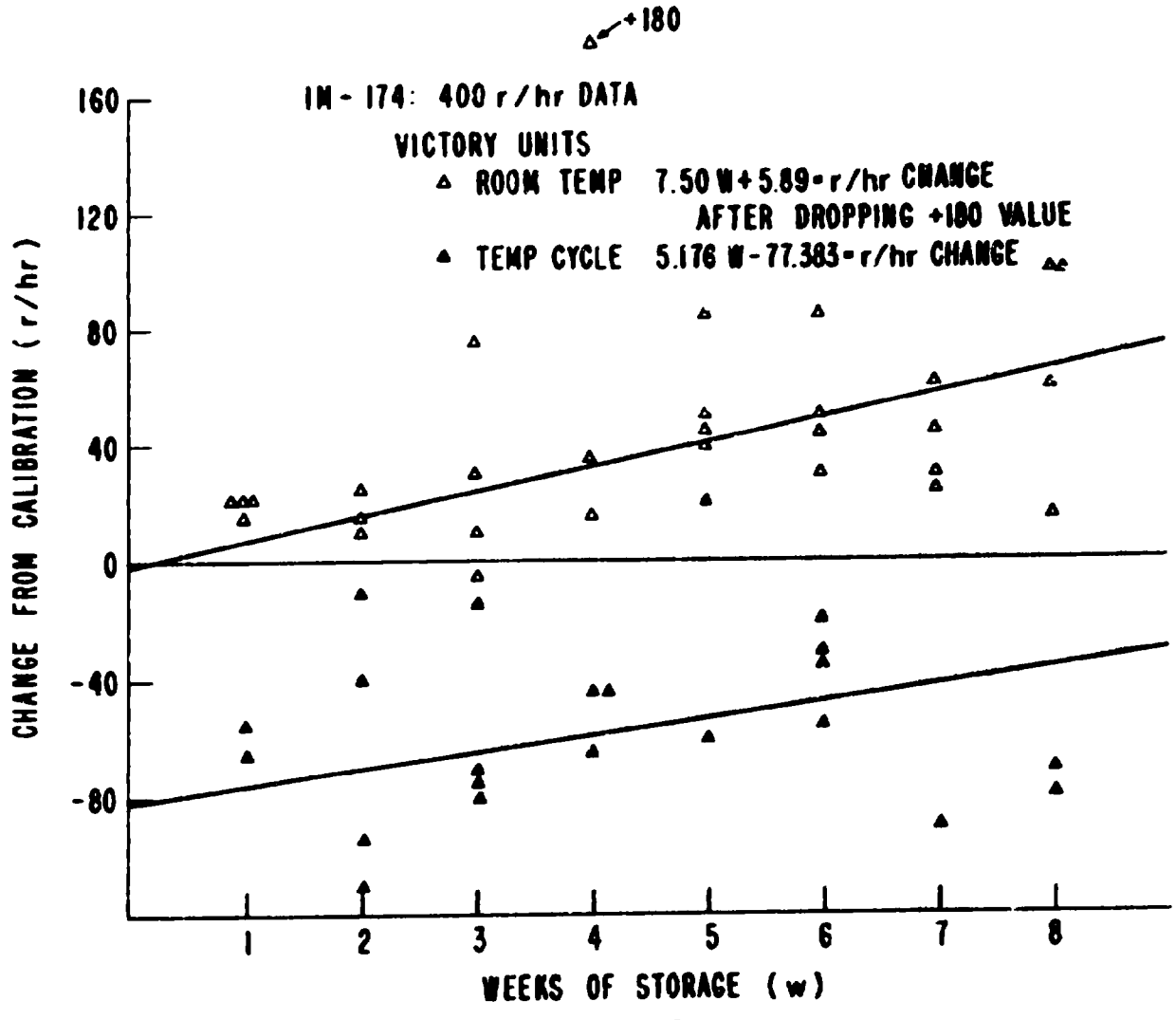

+180
IM - 174: 400 r/hr DATA
VICTORY UNITS
ROOM TEMP 7.50 W + 5.89 - r/hr CHANGE
AFTER DROPPING +180 VALUE
TEMP CYCLE 5.176 W - 77.383 - r/hr CHANGE
160
120
80
40
0
-40
-80
CHANGE FROM CALIBRATION (r/hr)
1
2
3
4
5
6
7
8
WEEKS OF STORAGE (w)

Plate 18

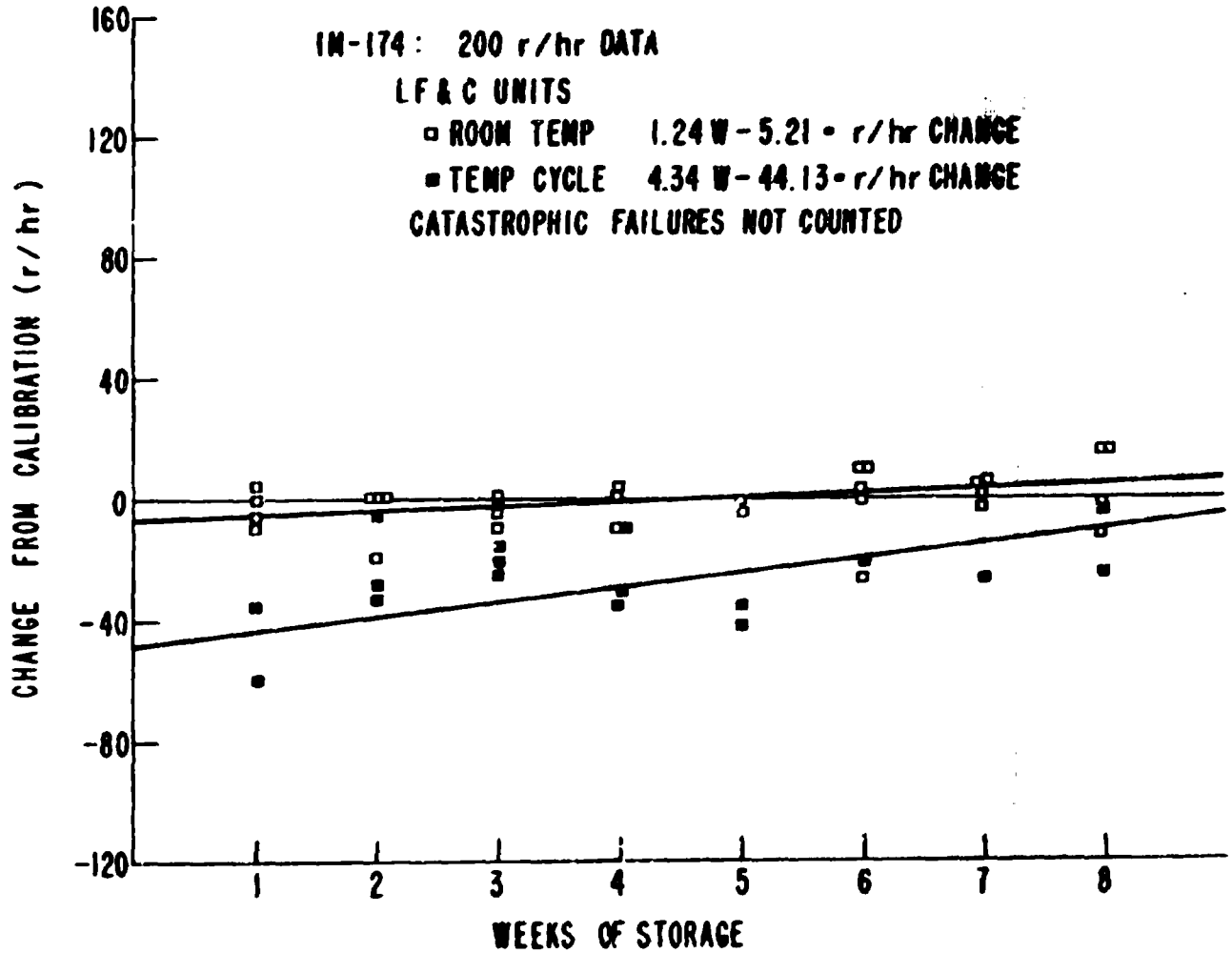

IM-174 : 200 r/hr DATA
LF & C UNITS
□ ROOM TEMP 1.24 W - 5.21 • r/hr CHANGE
■ TEMP CYCLE 4.34 W - 44.13 • r/hr CHANGE
CATASTROPHIC FAILURES NOT COUNTED
CHANGE FROM CALIBRATION (r/hr)
160
120
80
40
0
-40
-80
-120
1
2
3
4
5
6
7
8
WEEKS OF STORAGE

Plate 19

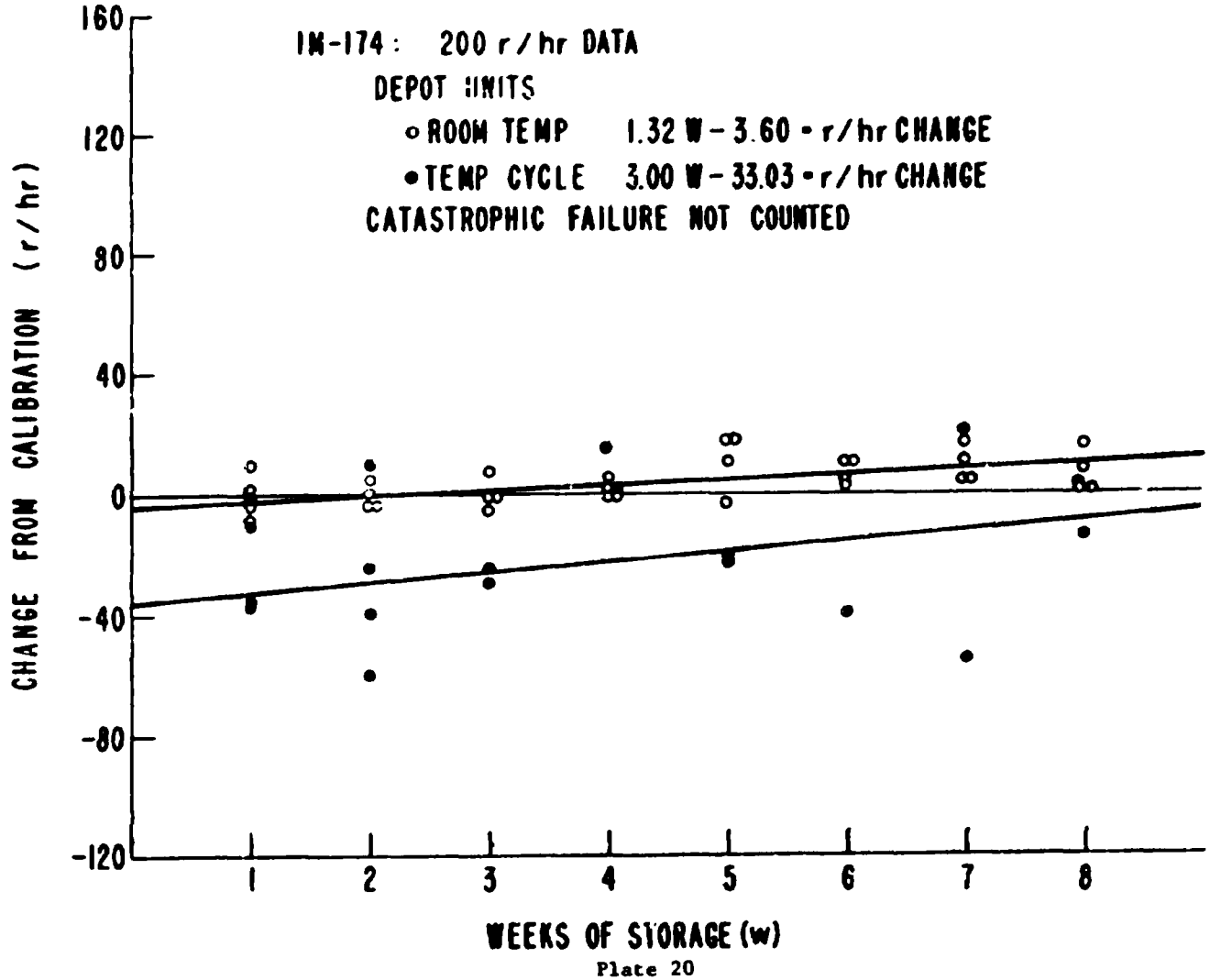

IM-174 : 200 r / hr DATA
DEPOT UNITS
o ROOM TEMP 1.32 W - 3.60 • r/hr CHANGE
• TEMP CYCLE 3.00 W - 33.03 • r/hr CHANGE
CATASTROPHIC FAILURE NOT COUNTED
CHANGE FROM CALIBRATION (r / hr)
160
120
80
40
0
-40
-80
-120
1
2
3
4
5
6
7
8
WEEKS OF STORAGE (w)

Plate 20

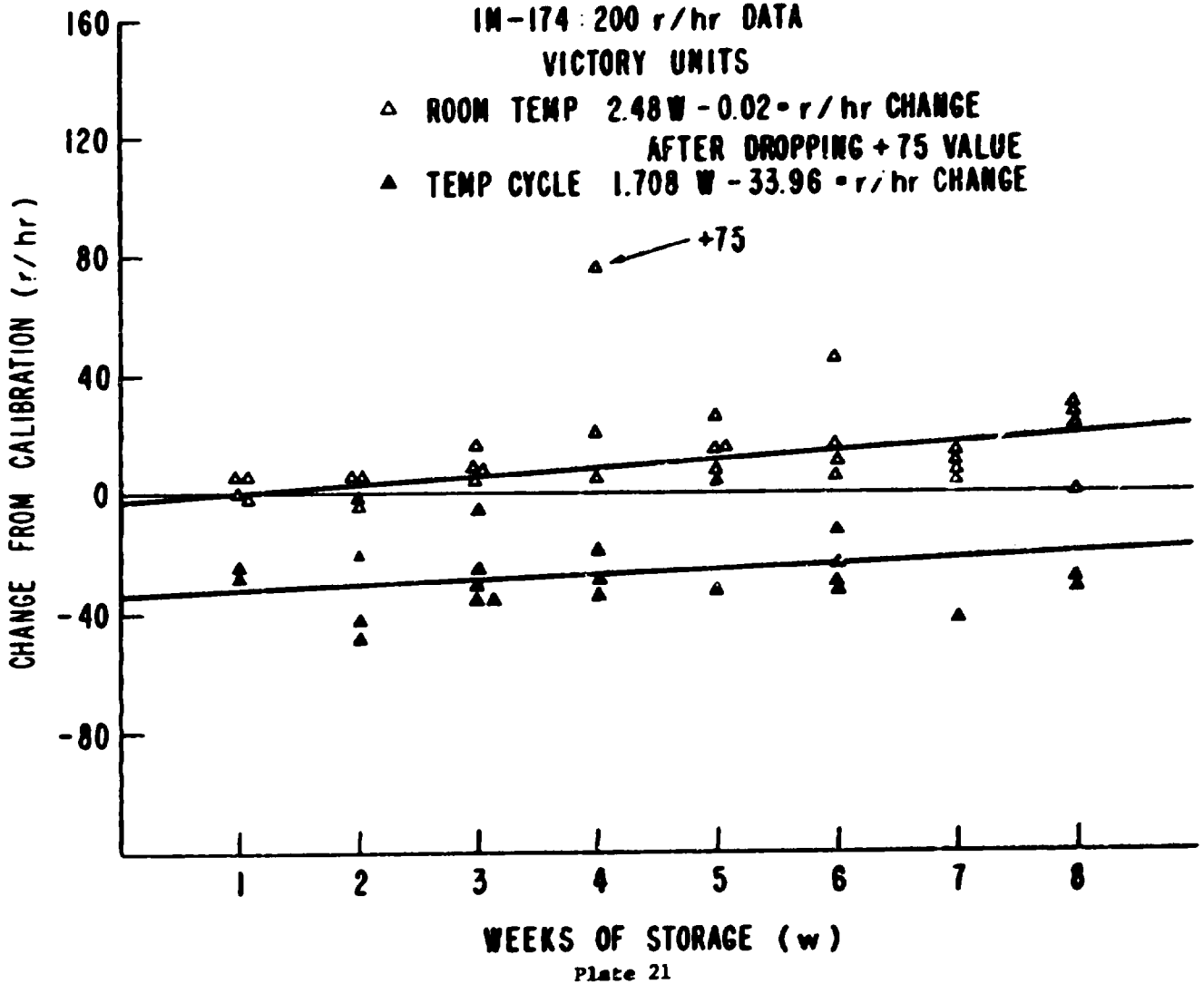

IM-174 : 200 r/hr DATA
VICTORY UNITS
△ ROOM TEMP 2.48 W - 0.02 • r/hr CHANGE
AFTER DROPPING + 75 VALUE
▲ TEMP CYCLE 1.708 W - 33.96 • r/hr CHANGE
+75
160
120
80
40
0
-40
-80
CHANGE FROM CALIBRATION (r/hr)
1
2
3
4
5
6
7
8
WEEKS OF STORAGE (w)

Plate 21

RELIABILITY TESTING OF WEAPON SYSTEMS

Ronald L. Racicot
Benet R&E Laboratories
Watervliet Arsenal
Watervliet, New York

1. Introduction:

The basic problem which I would like to offer for discussion and comments stems from my personal involvement in the preparation of Reliability/Maintainability Program and Test Plans for a particular close support weapon system. One of the high priority requirements for this weapon system, as well as other proposed Army Materiel, is increased reliability. Quantified reliability requirements have consequently been included in the QMR, System Specifications and System Development Plans for this weapon system. This will, of course, necessitate the adequate testing and assessment of the achieved reliability of the weapon system and its components at various stages in its development cycle.

In the preparation of the reliability test plans, a number of difficulties were encountered, particularly in the methods to be used for establishing confidence intervals on the reliability of components for a special class of idealized physical situations. What I seek here, then, are comments and possible practical approaches for resolving these apparent difficulties.

2. Definition of Reliability:

I would first like to define, as clearly as I can, one of the basic reliability indices which is specified in the weapon system requirements, that is, the "mission" or interval reliability under the assumption of "ideal repair" (Ref. 1). A mission here is defined in terms of a given number of rounds and miles which is a small fraction of the total expected life of the weapon system. Reliability is defined then as the probability that the weapon system will perform its intended function for the specified interval (mission) under stated conditions which are given in the mission profile for the weapon system. The assumption is made that at the beginning of a mission, the weapon system is in an operable state.

The term "ideal repair" just means that given a failure of the weapon system, it is subsequently repaired or renewed with any part or component replaced or repaired being put in "as new" condition. The weapon system is then available for further combat use. Between missions, preventive maintenance actions can be performed which include the replacement of parts

or components which are worn and are approaching a high failure rate condition (preventive maintenance parts replacement).

I emphasize the above points to make it clear that it is not the reliability for first mission or the conditional reliability for first failure which is specified here but rather the reliability for an interval of time at any point during the weapon life at which time the weapon system could have previously experienced one or more failures. In general, this reliability is transient with the specifications consequently being given as an average reliability over the entire life of the weapon system.

3. Reliability Theory:

It might do well at this point to introduce summarily without proof, some of the mathematics involved in computing the mission reliability assuming that all required parameters are known exactly. To start with, consider only a single part or component. Theoretically, if we know the reliabilities of the individual components, the system reliability can be computed using an appropriate reliability model for the system. Let $f(t)$ be the probability density distribution of times to first failure. We could write $f(t)$ as $f(t;\theta)$ where θ in this case represents one or more population parameters which completely define the density distribution f.

The initial problem usually considered in elementary reliability theory is the so-called "first failure" reliability problem. In this case

$R(t)$ = Probability that a given component will not fail in time $(0,t)$;

$= 1 - F(t)$;

$$= \int_t^\infty f(t)dt. \qquad (1)$$

The "conditional" interval reliability for an interval $(t, t + \Delta t)$ is given as

$R(t, \Delta t)$ = Probability that a given component will not fail in an interval $(t, t + \Delta t)$ given that the component has not failed up to time t.

$$= \frac{R(t+\Delta t)}{R(t)} \qquad (2)$$

It is useful at this point to define the failure rate or force of mortality, $\lambda(t)$ (Ref. 1):

$\lambda(t)dt$ = Probability that a component will fail in the incremental time interval $(t,t+dt)$. (3)

For the first failure reliability problem, $\lambda(t)$ is often called the "conditional" failure rate where $\lambda(t)dt$ is the conditional probability of failure given that the component has not failed up to time t. As will become evident shortly, however, the term conditional faiure rate is not appropriate for the ideal repair case since failures and replacements prior to time t are allowed. In terms of the conditional failure reliabilities previously given as equations (1) and (2) can be written as

$$R(t) = \exp\left(-\int_0^t \lambda(t)dt\right)$$

$$R(t,\Delta t) = \exp\left(-\int_t^{t+\Delta t} \lambda(t)dt\right) \quad (4)$$

where $\Delta(t) = f(t)/(1 - F(t))$
$= f(t)/R(t)$

Consider next the case of ideal repair. Clearly this case represents a renewal process where a part or component is renewed after each failure. In this instance, the probability distribution of times to the 1st, 2nd,...,nth failure for n arbitrarily large are required to establish the failure rate $\Delta(t)$ at any given time t. Let $f_j(t)$ be the probability density distribution of the time to the jth failure where $f_1(t) = f(t)$, the fundamental density of time to first failure. From renewal theory, the following expression relates $f_j(t)$ to the fundamental first failure frequency f(t) (Ref. 1, page 143):

$$L\{f_j(t)\} = [L\{f(t)\}]^j \quad (5)$$

where $L\{\cdot\}$ = Laplace or Fourier transform.

Also, the failure rate $\lambda(t)$ in this case can be written down as follows (Ref. 1, page 143):

$$\lambda(t) = \sum_{j=1}^{\infty} f_j(t) \quad (6)$$

This is, of course, the unconditional failure rate.

Taking the transform of equation (6) and using (5) finally gives the following equation relating $\lambda(t)$ to the fundamental first failure frequency $f(t)$:

$$L\{\lambda(t)\} = \frac{L\{f(t)\}}{1 - L\{f(t)\}} \tag{7}$$

Theoretically then, given $f(t)$, we can compute $\lambda(t)$. Interval or mission reliability in this instance is then given as

$$R(t,\Delta t) = \exp\left\{-\int_{t}^{t+\Delta t} \lambda(t)dt\right\} \tag{8}$$

A problem of considerable interest in reliability theory involves parts or components which exhibit wear-out phenomena caused by such mechanisms as fatigue, corrosion, erosion, abrasion and other similar mechanisms. For this case, the conditional failure rate for the first failure problem is an increasing function of time. In this instance, it is possible to increase interval or mission reliability by replacing components between missions before they enter the high-failure rate, wear-out regions of their lives. Let us assume that a given part is to be replaced after it has been in operation for a time period T, assuming it does not fail prior to this time. A replacement in this case is not considered a failure. The fundamental frequency distribution of times to first failure is now changed but can be determined from the underlying distribution of first failure times without replacement, $f(t)$, using the expression (Ref. 1, page 153):

$$f_p(t) = \sum_{j=0}^{\infty} [1 - F(T)]^j \, f_m(t - jT)$$

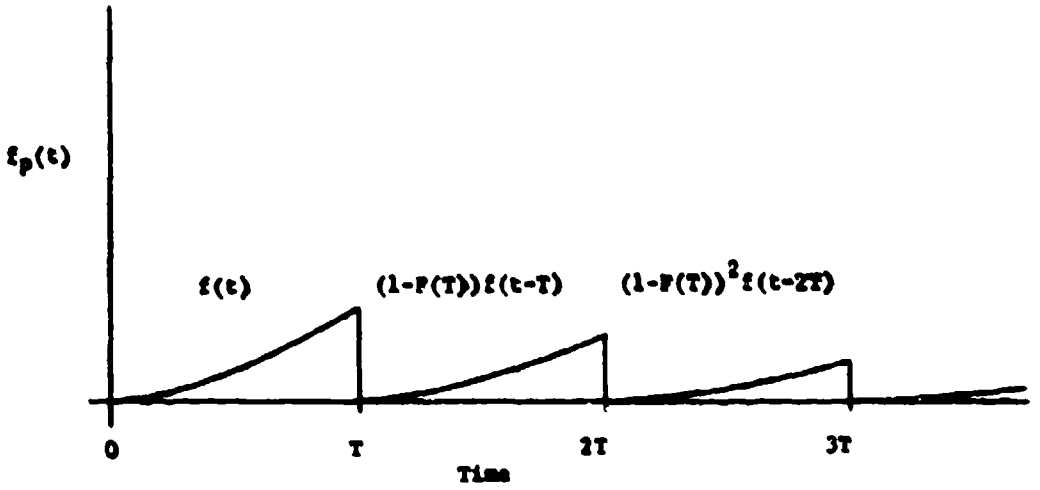
$$\text{where } f_m(z) = \begin{cases} f(z), & \text{for } 0 \leq z \leq T; \\ 0, & \text{otherwise} \end{cases} \tag{9}$$

$$F(T) = \int_{0}^{T} f(t)dt;$$

T = scheduled replacement time.

Figure 1 is a typical plot of $f_p(t)$.

Figure 1



For the combined case of ideal repair and preventive maintenance parts replacement, the same expressions as were previously presented for just the ideal repair case apply here with the density f(t) now being replaced with the preventive parts replacement density $f_p(t)$. We now have at least the theoretical means of computing interval (mission) reliability as a function of time for the ideal repair and parts replacement cases assuming that we know exactly the underlying first failure distribution f(t).

4. Reliability Testing Problems:

   a. Determination of the Confidenced Reliability of Components from Failure Data

In the initial stages of weapon system development, tests are to be conducted on prototype systems. The general philosophy during these early tests will be to use as little preventive maintenance as possible in order to purposely generate failures. In addition to the prototype tests, laboratory simulation tests will be conducted on a number of critical weapon system components, again with the intent of generating failures. The general purpose of this type of approach is to attempt to pinpoint as early as possible significant design problems as well as to generate failure data to establish preventive maintenance schedules including parts replacement. One type of data collected during these tests will be the failure times for each part which exhibited failures. It is desirable at the end of these initial tests to assess the achieved reliability of the weapon system including a projected parts replacement policy. We currently plan to do

this on a component-per-component basis and then to compute system reliability using an appropriate reliability model for the system: for example, for the series reliability model, system reliability $R_S(t)$ is determined from

$$R_S(t) = \prod_{i=1}^{n} R_i(t) \qquad (10)$$

where n = total number of components

$R_i(t)$ = reliability of the ith component

In general

$$R_S(t) = h(R_i(t)); \qquad (11)$$

= given function of the individual component reliabilities.

The usual assumption of independence of failures between different components is made in arriving at equations (10) and (11).

The specific problem here is how to compute the confidence interval for the ideal repair mission reliability for individual components given the failure times (i.e., a sample from the population f(t)) and, if applicable, the replacement time T. Assume that the form of the underlying first failure distribution is known; e.g., Weibull, Gamma, Normal, exponential, etc. For the general case, the mission reliability is given as

$$R(t,\Delta t;\theta) = g(f(t,\theta)) \qquad (12)$$

where $g(\cdot)$ is a known function of $f(t;\theta)$, the first failure density. For the preventive parts replacement case, only the lower tail of $f(t;\theta)$ enters into the reliability computation. θ represents one or more unknown population parameters.

Special Cases:

(1) Exponential or Constant Failure Rate Case

In this instance, failures are purely random and do not depend in any way on the previous history of failures. There is no advantage to be gained in using a parts replacement schedule and determination of confidenced mission reliability in this case is straightforward even for the ideal repair case.

(2) Low MTBF (mean time between failures) Case.

For system times of approximately 3 to 4 times the MTBF or greater, the failure rate $\lambda(t)$ approaches a constant value for the ideal repair case and is equal to 1/MTBF. For the preventive parts replacement case, the MTBF must be based on the density $f_p(t)$ rather than f(t). Mission reliability then for large system times can be computed from the relation

$$R(t,\Delta t) = \exp(-\Delta t/MTBF). \qquad (13)$$

What is required in this instance then is a confidence incerval on the MTBF which in turn depends on the underlying first failure distribution (either f(t) of $f_p(t)$). This seems to be a particularly bothersome problem for the preventive parts replacement case where only the lower tails of the original first failure distribution f(t) enter into the determination of $f_p(t)$. It should be emphasized here that even though the failure rate is constant, the flow of failures in the time direction is not a purely random process; that is, the interarrival times are not exponentially distributed.

b. Determination of the Confidence Interval for Total System Mission Reliability from Individual Component Tests

Assume that the system reliability can be computed from the component reliabilities using equation (11), i.e.

$$R_S(t) = h(R_i(t)),\ i = 1,\ldots n.$$

where $R_i(t) = R_i(t;\theta_i)$

θ_i = population parameters for the ith component

n = number of components

h(·) = known function of the component reliabilities.

Given the results of individual component failure times, the problem here is how to compute the confidence interval for the system mission reliability.

c. Determination of the Confidence Interval for System Mission Reliability from Total System Tests

In the latter stages of weapon system development, advanced production engineering prototype systems will be subjected to full scale life tests. Although failure times for each component comprising the weapon system will be recorded in the course of the tests, it may be advantageous in computing confidenced mission reliability to treat the system as if it were a single unit. The system failures in this case would then be a composite or overlay of all the mission stopping component failures that are experienced during the test. Under certain idealized conditions, it may be possible to treat such system failures as a simple Poisson flow, i.e. purely random occurrence of failure (ref. 2, pages 129-135). In this instance, confidenced mission reliability can be readily determined.

Based on prior experience, however, the frequency of system failures generally increases toward the latter stages of system life. This situation is common, for example, in the case of the auto. After a large number of miles have been accumulated, failures begin to occur at an increasing rate usually through the failure of components not previously failed (e.g. radiator, transmission, motor parts, etc.). In this case, the flow of failure is no longer simple Poisson. The basic problem then is how to compute confidenced mission reliability for this particular situation.

REFERENCES:

(1) Erich Pieruschka, "Principles of Reliability," Prentice-Hall, Inc., Englewood Cliffs, New Jersey, 1963.

(2) B. V. Gnedenko, Yu. K. Belyayev, and A. D. Solovyev, "Mathematical Methods of Reliability Theory," Academic Press, New York, 1969.

COMMENTS BY MURRAY A. GEISLER
on
"Reliability Testing of Weapon Systems"
by
Ronald L. Racicot

1. From previous day's discussion on field testing, we understand that the reliability of a weapon system is much more complex than getting the reliability of individual components. Weapons are used in a much more complex environment than is represented by the serial product of probabilities.

2. A more meaningful measure of the capability of a weapon system would be operational readiness more than just reliability alone, since operational readiness takes account of the downtime after the system becomes nonoperative and it must be repaired. Such an approach takes into account cost of repair and the cost of downtime, as well as the reliability before being launched. This approach thus incorporates the notion of life cycle cost, which is very important in the selection of weapon systems, since the total resource requirements must be considered in making the selection.

3. You have ignored the burn-in or infant mortality in your replacement decision. The need is not only to be in the wearout portion of the curve, but also to be far enough up it to outweigh the risk of infant mortality.

4. Also the choice in maintenance is sometimes between inspect the weapon system or replace particular components on the basis of age or activity. You may want to inspect a component before you replace it, especially if there is uncertainty as to the rate of wearout. There is a well developed theory in this area.

5. However, there are still complications in applying a parts replacement policy during a test period. This is because, typically, for the usual length of test experience, we may get no failures, yet we know that the parts will fail. This problem is equivalent to that of observing very low demands in inventory theory. To deal with the problem of 0 demand, we usually have to inject some prior probability distribution, like a negative exponential to get a practical parts replacement policy.

6. The Barlow-Proschan theory picks up the notion of monotonic failure distributions. They have a book on Mathematical Theory of Reliability, plus many papers, including empirical analyses, to illustrate their technique.

7. It would also be worthwhile to explore the possibility of using computer simulation to get at the desired confidence intervals by sampling.

8. Subsequent to the conference, two references were located that it is felt should be helpful to the author in his work. These are: *Mathematical Methods of Reliability Theory*, by B. V. Gnedenko, Yu K. Belyayev, and A. D. Solovyev, published by the Academic Press, 1969; and *Confidence Interval Estimation of the Reliability of Multicomponent Systems Using Component Test Data*, ARDC Technical Report No. 3, by Jerome R. Johnson, published by the Aberdeen Research and Development Center in November 1969.

TEST PROCEDURES FOR EVALUATION OF NUCLEAR EFFECTS ON INITIATORS

W. B. Thomas and R. E. Betts
Solid Propellant Chemistry Branch
Army Propulsion Lab
Redstone Arsenal, Alabama

The exoatmospheric explosion of a nuclear device produces radiation which is absorbed by materials. The amount of radiation absorbed by any given material is roughly equivalent to the fourth power of its atomic number. The absorption phenomena produces a high temperature gradient in such a short time that it generates a shock wave in the material causing spallation and fracturing of metal parts, and activation of explosive components.

The following VU-GRAPHS show graphically this phenomena. Figures 1 and 2 show by relative areas the amount of this energy absorbed by given materials with respect to iron. The division being made to those materials used by conventional initiators and those materials proposed for use with the Low-Z SQUIB.

Figure 3 shows the deposition phenomena and accompanying reaction with respect to time; the deposition time occurring in fractions of microseconds. Figures 4 and 5 show the reaction as it would occur in an initiator.

Electric initiators have been exposed to simulated nuclear environments in underground testing. Both Picatinny Arsenal and Redstone Arsenal have participated in these tests. Since conventional electric initiators are composed of noble metal electric bridgewires, and in some instances explosives with high atomic number, such as lead azide or lead styphante, it is obvious that these initiators would be quite susceptible to the radiation of the bomb. Tests have proven this to be the case.

With a relative amount of energy in the order of (1), the lead compound initiators explosives detonate. Between the relative numbers of 3 to 10, explosives of low atomic numbers will initiate if they are adjacent to the bridgewire which are of high atomic number such as gold or platinum. In the relative number 30 range, these wires will actually be vaporized to the extent that they will be detrimental even if no explosive were in contact with the wires.

The design approach of the Propulsion Lab is to design an initiator which is composed of low atomic numbered elements. Primarily the goal would be to use elements of atomic number (13 Aluminum) and below.

A prototype design of this initiator was assembled and tested in an underground test, and shown to be able to survive the relative 30 level. This design was considered primeval in nature and proved only the feasibility of the concept. Development effort on this initiator has remained dormant for approximately 2 years due to the inability to schedule underground tests. However, within recent weeks scheduling for such tests have been acquired and developmental effort has been reactivated.

This effort will now commence with a more sophisticated design using a metal to ceramic header with low-Z bridgewires or spark gap. The materials of primary interest are aluminum-alumna header and beryllium-berylla header with beryllium or aluminum bridgewire. In addition to the electrical initiator approach, effort is also being expended to develop an initiator which is activated by laser light. The concept is to use laser energy piped through fiber-optics. The laser energy is transmitted into the initiator through a quartz window. Hopefully, the element silicon will be the highest Z number.

In the past, in order to obtain data as to the survivability of these initiators to the nuclear stimulii, a given number of initiators were placed at a given distance from the source. Then, based upon the number of initiators which survived at a particular station, conclusions were drawn as to the survivability of the item. For example, if all the items survived at 1 and 3 relative numbers but one or more of the initiators were damaged or were activated at the 10 relative number, then it was concluded that at relative level 3 the initiators were safe; somewhere above 3 but under 10 they were not. At present, we have no methods to determine the qualititative effect of the nuclear environment on the initiators. Therefore, we are basing our survivability numbers on GO-NO-GO data only.

The shortcoming of the testing program appear to be four-fold. 1. The underground test does not give a true reproduction of an exoatmospheric environment. 2. Qualititative effects cannot be measured. 3. The expense of performing the test is enormous. 4. The long term effects are unknown.

When the two initiators now under development reach the stage of completion for testing, our plans are to do testing similar to that which has already been done. That is, to use a sample size of 10 to 20 initiators and station them at 3 to 10 different stimuli. Figure 6 is a schematic of a test set-up showing the radiation as it impinges on the various stations. The lowest station at which damage or activation does not occur will be assumed as the safety level, if after exposure the initiator performance does not change. The space in an underground test will probably be limited to approximately 25 sp. in. per station, for 3 to 10 stations. This space limitation fixes the 10 to 20 sample size. The question posed here is, "Is our rationale justified in testing with such low numbers in determing threshold levels or survivability to a given nuclear environment?" We also

ask which are the most advantageous stations for testing if we estimate prior to testing, the level of survival, i.e., if we estimate that the initiator will survive the 30 level would it be better to test two stations below 30 and five stations above or vice versa?

If development time permits, prior to underground testing, then simulation techniques with sophisticated machines will be used to determine the threshold level and survivability point. This testing will be accomplished by the Bruceton method in which approximately 25 to 50 initiators will be tested. The energy level will be varied and the 50% point calculated. The 0.5% and the 99.5% reliability will be calculated at a 95% confidence level using the Bruceton type data.

We conclude by asking what is the most advantageous test method and the best method of handling test data for determining the adverse nuclear effects on initiators.

RELATIVE AMOUNT OF ELECTROMAGNETIC RAY ENERGY ABSORBED

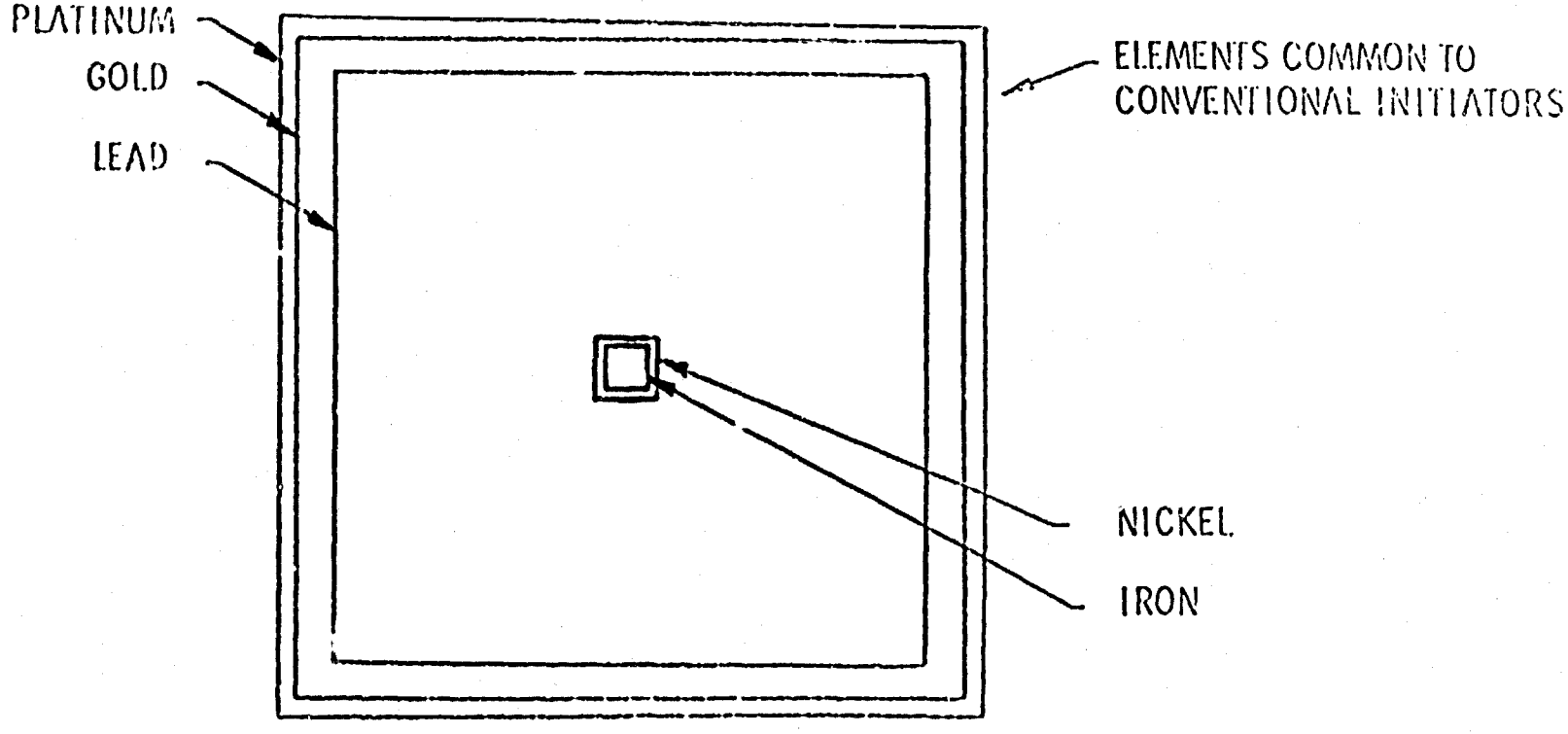


FIGURE 1

FIGURE 2

FIGURE 3

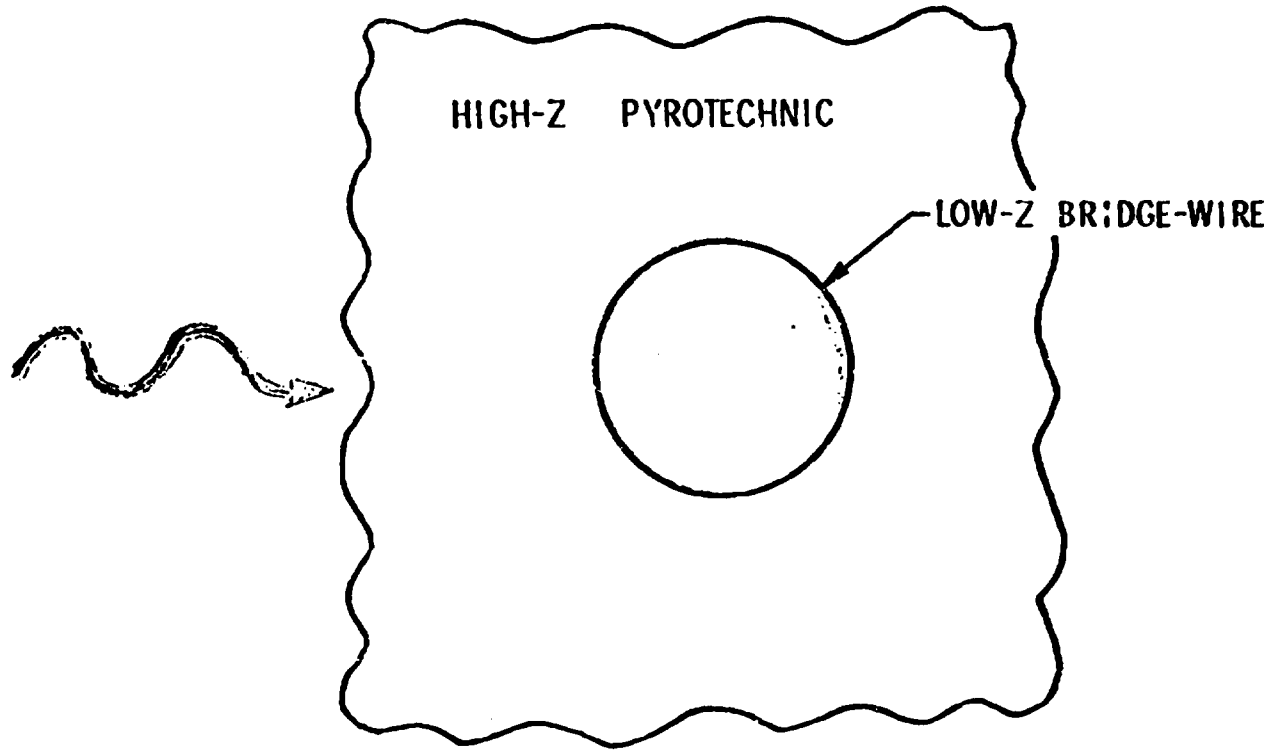


FIGURE 4

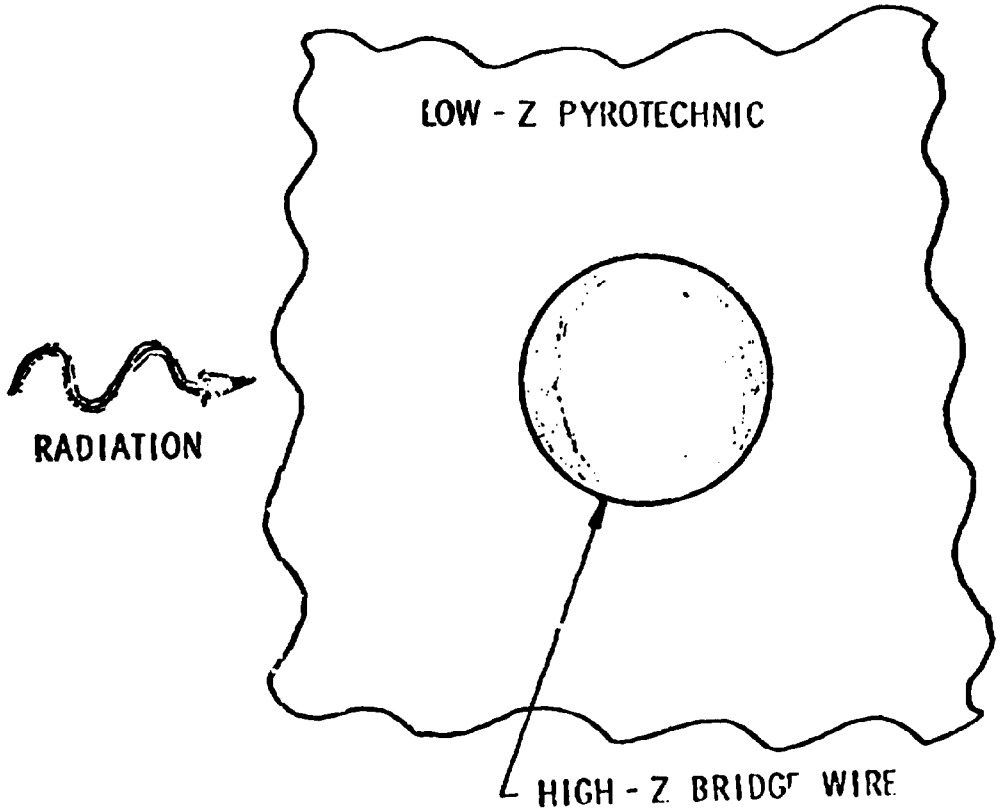


FIGURE 5

TEST SET-UP FOR EXPOSURE OF SAMPLES TO RADIANT ENERGY FROM QUARTZ ENVIRONMENTAL

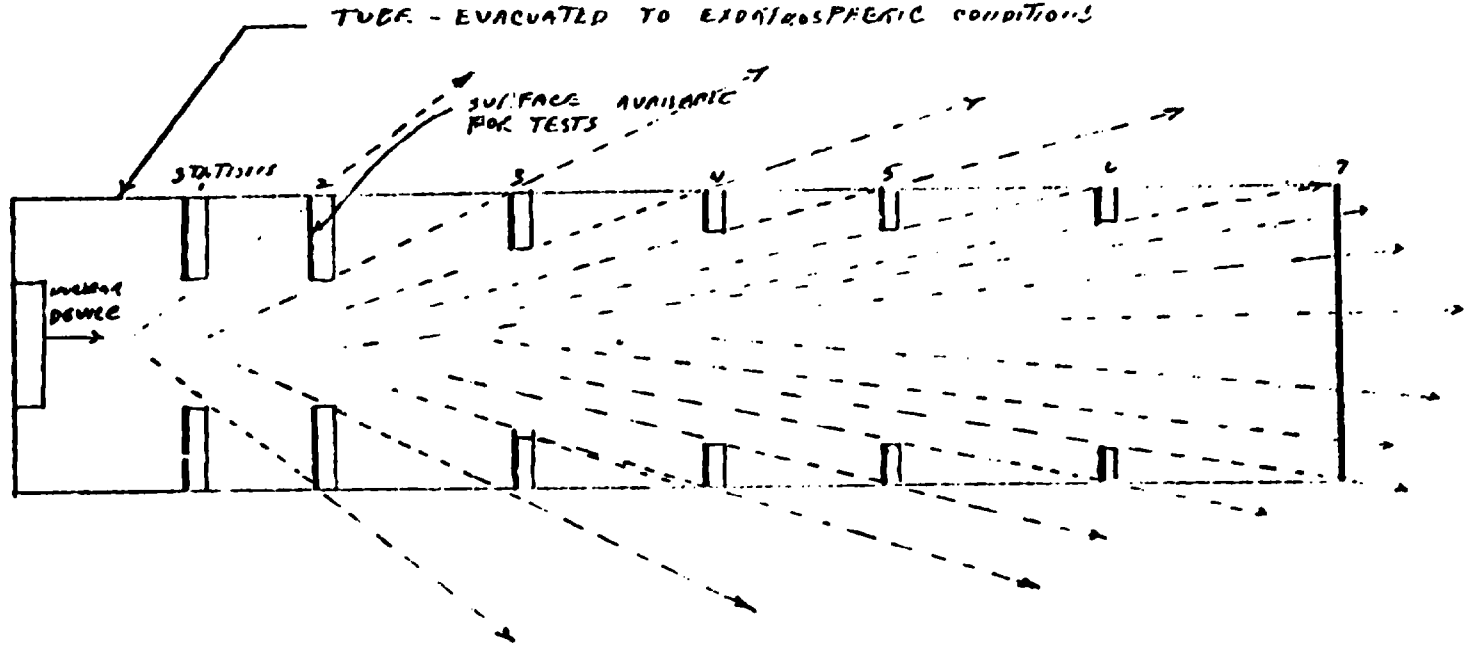


FIGURE 6

COMMENTS BY MURRAY A. GEISLER
on
"Test Procedures for Evaluation of Nuclear Effects on Initiators"
by
W. B. Thomas & R. E. Betts

1. The sample sizes seem very small for the complexity of phenomena to which you are extrapolating. I would want supplementary data and evidence to develop the final test plan.

2 The possibility of using simulation, as described in next to last paragraph of the paper, seems intriguing. If this simulation has validity, I would want to use it to plan test. If simulation data do help to determine the threshold value and survivability point, I would use simulation predictions to help set up the test plan. Simulation data could also help to estimate the test confidence.

3 Using simulated data, I would concentrate tests around where the 30 level seems indicated, with somewhat more tests distributed above the 30 level, to be more confident that you have reached it. Concentration around where you think the 30 level is, will depend on how confident you are in the simulation data.

4. I would also try to calibrate simulation from the test data, and might try to design the experiment for obtaining test data to accomplish this, since apparently you can replicate the simulations more readily. I would use this approach as a primary source of data and analysis.

5. You still have the problem of extrapolating from the test to the real world. You have go-no-go data only, and only a small sample. How are initiators distributed as to material? Can you estimate effect from one material to another using this 4th power rule? Do distance effects get realistically translated through using this rule across materials? I guess this part of the theory is still very tenuous.

OPTIMAL DESIGNS WITH A TCHEBYCHEFFIAN SPLINE REGRESSION

V. N. Murty
The Pennsylvania State University
Middletown, Pennsylvania

This talk will be a summarized version of three papers by the author. [See references below as to where the full papers can be found.]

Studden (1968) showed that the optimal design for estimating any specified regression coefficient or parameter is supported by one of two sets of points for Tchebycheffian systems with certain symmetry properties. In the first paper we consider a Tchebycheffian Spline Regression Function, defined on an interval, and show that the optimal design for estimating any specified regression coefficient is supported on the same set of points.

When an experimenter is interested in more than one parameter in the regression model, and tries to obtain a design that minimizes the maximum variance of the individual regression coefficients he is looking for a minimax design with respect to the single parameter, a concept introduced by Elfving (1959). In our second paper we explicitly present the minimax s.p. designs for the ordinary polynomial regression, when the degree is ≤ 12. A general solution of this problem is still open, but the results obtained do indicate the direction in which one could look for a possible general solution.

In the third paper it is shown that the optimal design for estimating any specified parameter in a polynomial spline regression with a single multiple knot at the center is supported by one of two sets of points.

REFERENCES

1. Murty, V. N., OPTIMAL DESIGNS WITH A TCHEBYCHEFFIAN SPLINE REGRESSION, Vol. 42, No. 2, April 1971 issue of Annals. Math. Stat.

2. Murty, V. N., OPTIMAL DESIGNS WITH A POLYNOMIAL SPLINE REGRESSION WITH A SINGLE MULTIPLE KNOT AT THE CENTRE, Vol 42, No. 3, June 1970 issue of Annals. Math. Stat.

3. Murty, V. N., MINIMAX DESIGNS, June 1971 issue of Jour. Amer. Stat. Assn.

ORDINARY AND EMPIRICAL BAYES APPROACH TO ESTIMATION OF RELIABILITY IN THE WEIBULL LIFE TESTING MODEL

George C. Canavos
NASA Langley Research Center
and
Chris P. Tsokos
Virginia Polytechnic Institute

SUMMARY

An ordinary as well as an empirical Bayesian analysis for a random scale parameter in the Weibull distribution is developed with respect to the usual life testing procedures. The results of Bhattacharya (1967) are generalized to include the flexibility of the shape parameter in the Weibull case. Empirical Bayes estimators are developed for the scale parameter and the reliability function based upon prior experiences. A simulation procedure is carried out and a comparison is made between Bayes, empirical Bayes, and minimum variance unbiased estimators.

1. INTRODUCTION

For the Weibull failure distribution with probability density function

$$f(x|\theta,\xi) = \frac{\xi}{\theta} x^{\xi-1} e^{-\frac{1}{\theta}x^{\xi}} \quad (0 < x < \infty,\ \theta,\ \xi > 0), \tag{1.1}$$

we develop ordinary as well as empirical Bayes estimators of the scale parameter and the reliability function

$$R(t|\theta,\xi) = e^{-\frac{1}{\theta}t^{\xi}} \quad t > 0 \tag{1.2}$$

with respect to the usual life-testing procedures. The shape parameter ξ is assumed to be known.

This paper has been reproduced photographically from the author's manuscript.

Since in many life-testing situations it is not unlikely to note the unpredictable fluctuation of the scale parameter in a failure model, it is justifiable to consider such a parameter as a random variable and, thus, appeal to a Bayesian analysis. In specific, let Θ denote the random variable associated with the scale parameter and θ its realization. Obviously, a Bayesian analysis depends on the utilization of prior information which, in this case, we assume to exist either in the form of a prior distribution of Θ or a sequence of sufficient statistics from past experiments. For the ordinary Bayes approach, we appeal to a well-known transformation to generalize the results of Bhattacharya (1967) for the one-parameter exponential model so as to include the flexibility provided by the shape parameter ξ in the Weibull distribution. In fact, the Weibull failure model has an increasing ($\xi > 1$) or decreasing ($\xi < 1$) failure rate and, thus, is likely to describe the life span of items with variable failure rates.

The empirical Bayes estimation technique was largely motivated by Robbins (1955) who assumed the existence of a prior distribution for an unknown parameter but not the knowledge of its form. Instead, he substitutes past information which he assumes to exist as a result of the repetitive nature in the problem of estimation. Thus, in the absence of knowledge concerning the form of the prior distribution, we appeal to an empirical Bayes approach to estimate the scale parameter. By using this estimate, an estimate of the reliability function is made possible.

2. ORDINARY BAYES ESTIMATION

The usual procedure in life-testing is to place n items on a life-test and to terminate the experiment after a predetermined $r \leq n$ number

of failures have been observed. Inferences are usually based on a sufficient statistic which for the Weibull model is given by $T_r = \sum_{i=1}^{r} x_i^{\xi}$ where $(x_1, x_2, \ldots, x_r) = \underline{x}$ denote the r lifetimes of the censored sample. Clearly, the complete sample is realized for $r = n$. With respect to this procedure, Bhattacharya (1967) considered a random scale parameter Θ in the exponential failure distribution with density

$$f(y|\theta) = \frac{1}{\theta} e^{-\frac{1}{\theta}y} \quad (0 < y < \infty, \ \theta > 0) \tag{2.1}$$

and for the usual assumptions derived Bayesian estimates of θ, $R(t) = e^{-\frac{1}{\theta}t}$, and their variances for a uniform, an exponential, and an inverted gamma prior distributions. Here, we extend his results to include the Weibull failure model and the flexibility provided by its shape parameter.

It is well known--see Tate (1959)--that if the random variable Y has the exponential density given by (2.1) then the random variable $X = Y^{1/\xi}$ follows a Weibull distribution with density given by (1.1). Hence, for the Weibull case, the Bayesian estimates and their variances which are given in table 2.1 follow immediately from this transformation and the results of the exponential model. For detail of proof, see Bhattacharya (1967). We remark that in table 2.1,

$$\gamma(n,z) = \int_0^z e^{-t} t^{n-1} \, dt$$

and

$$\gamma^*(m,v) = \gamma\left(m, \frac{v}{\alpha}\right) - \gamma\left(m, \frac{v}{\beta}\right)$$

represent relations involving the incomplete gamma function while $K_\nu(au)$ is the modified Bessel function of the third kind of order ν.

3. EMPIRICAL BAYES ESTIMATION

We implement here an empirical Bayes approach for the estimation of the scale parameter and the reliability function in the Weibull distribution. The empirical Bayes method that we consider was introduced by Robbins (1955), and its theory and techniques, were developed by Robbins (1964), and Rutherford and Krutchkoff (1969), among others. The difference between empirical Bayes and ordinary Bayes is that the former does not make explicit the form of the prior distribution in order to make possible a Bayes solution. Instead, the empirical Bayes method depends on the existence of prior information in the form of past estimates of either the parameter in question or some close variation of it.

It is well known that for a squared error loss, the Bayes risk is achieved if the decision function is the posterior expectation, $E\{\theta|\underline{x}\}$, where

$$E\{\theta|\underline{x}\} = \int_0^\infty \theta h(\theta|\underline{x}) dG(\theta) \tag{3.1}$$

and

$$h(\theta|\underline{x}) = \frac{\ell(\underline{x}|\theta) dG(\theta)}{\int_0^\infty \ell(\underline{x}|\theta) dG(\theta)} . \tag{3.2}$$

The motivation of the empirical Bayes technique, therefore, is to utilize past experiences to provide a consistent estimator of the posterior expectation (3.1) whose risk in general converges to that of ordinary Bayes as the number of past experiences used tends toward ∞.

In the current situation, the existence of prior information is based on the assumption that life-tests have been conducted periodically on identical items and, thus, past failure information has accrued. In specific, we assume the existence of $k > 1$ past sufficient statistics. Let $(\underline{x}_1,T_1)$, $(\underline{x}_2,T_2)$, ..., $(\underline{x}_k,T_k)$ represent the sequence of previous information where the vector $\underline{x}_j = (x_{1j}, x_{2j}, \ldots, x_{nj})$ denotes a sample of $n > 1$ random lifetimes from (1.1), and $T_j = \sum_{i=1}^{n} x_{ij}^{\xi}$ represents the corresponding sufficient statistic. We use the fact that for known shape parameter ξ, T is a sufficient statistic of fixed dimensionality--Raiffa and Schlaifer (1961)--and thus find an equivalent form of the posterior density (3.2) conditional on the sufficient statistic T rather than upon $\underline{x}$. We determine first the probability density function of T conditional on θ which is given by

$$q(T|\theta) = \frac{1}{\Gamma(n)\theta^n} T^{n-1}e^{-T/\theta} \quad (0 < T < \infty,\ \theta > 0). \tag{3.3}$$

This follows from the fact that if X is Weibull then X^{ξ} follows an exponential distribution; therefore, the density of T is the convolution of n exponentials. See Tate (1959).

Notice that the likelihood of a sample of lifetimes from (1.1) can be written as

$$l(\underline{x}|\theta) = \frac{T^{n-1}e^{-T/\theta}}{\Gamma(n)\theta^n} \cdot \frac{\xi^n\Gamma(n)\prod_{i=1}^{n} x_i^{\xi-1}}{\left(\sum_{i=1}^{n} x_i^{\xi}\right)^{n-1}}$$

which reduces to

$$l(\underline{x}|\theta) = q(T|\theta)r(\underline{x}).$$

If the prior density of θ is $g(\theta)$, the joint probability density of $\underline{x}$ and θ is

$$\ell'(\underline{x},\theta) = q(T|\theta)g(\theta)r(\underline{x}) = q'(T,\theta)r(\underline{x}). \tag{3.4}$$

Integrating (3.4) over θ, we obtain

$$\ell^*(\underline{x}) = q^*(T)r(\underline{x})$$

where

$$q^*(T) = \int_0^\infty q(T|\theta)\, dG(\theta).$$

As a result, the posterior density $h(\theta|\underline{x})$ can be written as

$$h(\theta|\underline{x}) = \frac{\ell'(\underline{x},\theta)}{\ell^*(\underline{x})} = \frac{q(T|\theta)g(\theta)}{q^*(T)} = h(\theta|T).$$

Therefore,

$$E\{\theta|\underline{x}\} = E\{\theta|T\} = \int_0^\infty \theta h(\theta|T)\, d\theta. \tag{3.5}$$

The conditional density of T_k where T_k denotes the current sufficient statistic can be written as

$$q_n(T_k|\theta) = \frac{1}{\Gamma(n)}\left(\frac{T_k}{\theta}\right)^n \frac{e^{-T_k/\theta}}{T_k} \tag{3.6}$$

where the subscript n implies that T_k is a function of n random life times. Multiplying equation (3.6) by θ and after some manipulation, the equation is reduced to

$$\theta q_n(T_k|\theta) = \frac{T_k}{n-1} \cdot q_{n-1}(T_k|\theta) \tag{3.7}$$

where

$$q_{n-1}(T_k|\theta) = \frac{1}{\Gamma(n-1)}\left(\frac{T_k}{\theta}\right)^{n-1} \frac{e^{-T_k/\theta}}{T_k}$$

is the conditional density of T_k as a function of the first $(n-1)$ lifetimes of the current sample. Substituting equation (3.7) in equation (3.5) we obtain

$$E\{\theta|T_k\} = \frac{T_k}{n-1} \cdot \frac{\int_0^\infty q_{n-1}(T_k|\theta)\, dG(\theta)}{q_n^*(T_k)}$$

or

$$E\{\theta|T_k\} = \frac{T_k}{(n-1)}\left\{\frac{q_{n-1}^*(T_k)}{q_n^*(T_k)}\right\}. \tag{3.8}$$

Obviously, the marginal densities $q_{n-1}^*(T_k)$ and $q_n^*(T_k)$ depend on the form of the prior distribution $G(\theta)$. However, by substituting consistent estimates $q_{n-1,k}^*(T_k)$ and $q_{n,k}^*(T_k)$, respectively, we obtain an empirical Bayes estimator of θ. Based on empirical evidence, the risk of such an estimator is reduced if the quantity $T_k\left\{\frac{q_{n-1,k}^*(T_k)}{q_{n,k}^*(T_k)}\right\}$ is divided by n rather than by $(n-1)$. Hence, the recommended empirical Bayes estimator of the current realization θ_k becomes

$$\tilde{\theta}_k = \frac{T_k}{n}\left\{\frac{q_{n-1,k}^*(T_k)}{q_{n,k}^*(T_k)}\right\}. \tag{3.9}$$

Substituting (3.9) in (1.2), we obtain an estimator of the reliability function based on $\tilde{\theta}_k$ and expressed by

$$R(t|\tilde{\theta}_k, \xi) = e^{-\frac{1}{\xi} t^\xi} \qquad t > 0. \tag{3.10}$$

For the computation of $q^*_{n-1,k}(T_k)$ and $q^*_{n,k}(T_k)$, we appeal to a result of Parzen (1962) who has provided a class of consistent estimators of probability density functions. In specific, we recommend an estimator for $q^*_{n-1}(T_k)/q^*_n(T_k)$ of the form

$$\frac{q^*_{n-1,k}(T_k)}{q^*_{n,k}(T_k)} = \frac{\sum_{j=1}^{k} e^{-|\varphi'_j|}}{\sum_{j=1}^{k} e^{-|\varphi_j|}}, \tag{3.11}$$

where

$$\varphi'_j = \frac{T_{n,k} - T_{(n-1),j}}{\delta},$$

$$j = 1,2, \ldots, k$$

$$\varphi_j = \frac{T_{n,k} - T_{n,j}}{\delta},$$

and

$$\delta = k^{-1/5}\sqrt{\sum_{j=1}^{k} (T_{n,j} - \bar{T}_n)^2}.$$

4. A SIMULATION PROCEDURE

From the results of the previous section, the validity of an empirical Bayes estimator rests on the notion that its risk approaches the risk of ordinary Bayes as the amount of past information becomes substantial. However, if the risk of empirical Bayes does not converge to that of ordinary Bayes within a relatively small number of past experiments, the practical application of the former in life testing seems doubtful. This is primarily because of the expected limitation on the size of past

information. Moreover, if the size of past information is indeed substantial, say 15 or more, the reasonable approach in such a situation should be based on an attempt to identify the prior distribution form and use ordinary Bayes procedures rather than empirical Bayes methods. Hence, we find it necessary to appeal to a Monte Carlo simulation of random lifetimes from the Weibull distribution so that a comparison of small-sample behavior of ordinary Bayes and empirical Bayes be made possible. At the same time, we investigate the degree of improvement of the empirical Bayes estimators (3.9) and (3.10) as functions of the number k of past estimates over the corresponding minimum variance unbiased (MVU) estimators developed by Tate (1959).

The simulation procedure is as follows: a complete sample of n lifetimes is generated by first simulating a realization θ according to a prior distribution. This task is repeated 15 times. At each time, the squared deviations from θ and the reliability function of ordinary Bayes and MVU estimates are computed and stored with the reliability estimates calculated for relatively small increments of t. For $k = 2, 5, 10,$ and 15, the corresponding empirical Bayes estimates and their squared deviations are determined as functions of the k previous estimates. Finally, the entire process is repeated 500 times and average squared-errors are computed. The criterion for comparison is determined by considering mean-squared error (MSE) ratios of empirical Bayes to ordinary Bayes and to MVU estimators.

The results of the Monte Carlo simulation are systematically given in the tables below. The situation under which the contents

of each table were realized is made explicit and need not be repeated here. We remark that an index to the improvement of the empirical Bayes estimator over the corresponding MVU estimator is provided by the quantity

$$A = \frac{E\{Var(\hat{\theta}|\theta)\}}{Var(\theta)},$$

where $Var(\hat{\theta}|\theta)$ is the conditional variance of the sufficient estimator, $\hat{\theta} = \frac{T}{n}$, and $Var(\theta)$ is the prior variance.

The following general conclusions drawn from the results given below are apparent for a random scale parameter in the Weibull distribution:

(1) As expected, the ordinary Bayes estimators have the smallest MSE irrespective of prior distribution form, prior parameter values, or shape parameter value.

(2) The MSE of the empirical Bayes estimators gets closer to that of ordinary Bayes as the number of past estimates increases. However, it is doubtful whether further squared-error improvement can be made with more than 15 past estimates.

(3) Generally, for $k \geq 5$ past estimates, the empirical Bayes estimators have smaller MSE's than corresponding MVU estimators. Moreover, the rate of improvement is not a function of prior distribution form but is directly proportional to the index A.

REFERENCES

Bhattacharya, S. K. (1967). Bayesian approach to life testing and reliability estimation. J. Amer. Statist. Assoc. 62, 48-62.

Parzen, E. (1962). On estimation of a probability density function and mode. Ann. Math. Statist. 33, 1065-76.

Raiffa, H. and Schlaifer, R. (1961). Applied Statistical Decision Theory. Harvard Graduate School of Business Administration.

Robbins, H. (1955). An empirical Bayes approach to statistics. Proc. Third Berkeley Symp. 1, 157-63.

Robbins, H. (1964). The empirical Bayes approach to statistical decision problems. Ann. Math. Statist. 35, 1-20.

Roy, L. K. and Wasan, M. T. (1968). The first passage time distribution of brownian motion of positive drift. Math. Biosci. Vol. 3, 191-204.

Rutherford, J. R. and Krutchkoff, R. G. (1969). Some empirical Bayes techniques in point estimation. Biometrika 56, 221-3.

Tate, R. F. (1959). Unbiased estimation: functions of location and scale parameters. Ann. Math. Statist. 30, 341-66.

TABLE 2.1.- ORDINARY BAYES ESTIMATORS FOR A RANDOM SCALE PARAMETER IN THE WEIBULL FAILURE DISTRIBUTION

Prior density	Symbol	Bayes estimator
$g(\theta) = \frac{1}{\beta - \alpha}$ $(0 < \alpha < \theta < \beta)$	$E(\theta \mid \underline{x})$	$\frac{\gamma^*(r-2, T_r)}{\gamma^*(r-1, T_r)} \cdot T_r$
	$\mathrm{Var}(\theta \mid \underline{x})$	$\frac{\gamma^*(r-3, T_r)\gamma^*(r-1, T_r) - [\gamma^*(r-2, T_r)]^2}{[\gamma^*(r-1, T_r)]^2} \cdot T_r^2$
	$E[R(t) \mid \underline{x}]$	$\frac{\gamma^*(r-1, T_r + t^b)}{\gamma^*(r-1, T_r)} \cdot \left(1 + \frac{t^b}{T_r}\right)^{1-r}$
	$\mathrm{Var}[R(t) \mid \underline{x}]$	$\frac{1}{[\gamma^*(r-1, T_r)]^2} \left\{ \frac{\gamma^*(r-1, T_r)\gamma^*(r-1, T_r + 2t^b)}{\left(1 + \frac{2t^b}{T_r}\right)^{r-1}} - \frac{[\gamma^*(r-1, T_r + t^b)]^2}{\left(1 + \frac{t^b}{T_r}\right)^{2(r-1)}} \right\}$
$g(\theta) = \frac{1}{\lambda^{1/\theta}} \theta^{-\frac{1}{\theta}}$ $(0 < \theta < \infty, \lambda > 0)$	$E(\theta \mid \underline{x})$	$(\lambda T_r)^{1/2} \cdot \frac{K_{r-2}(2\sqrt{T_r/\lambda})}{K_{r-1}(2\sqrt{T_r/\lambda})}$
	$\mathrm{Var}(\theta \mid \underline{x})$	$\frac{\lambda T_r}{K_{r-1}(2\sqrt{T_r/\lambda})} \left\{ K_{r-3}(2\sqrt{T_r/\lambda}) - \frac{K_{r-2}^2(2\sqrt{T_r/\lambda})}{K_{r-1}(2\sqrt{T_r/\lambda})} \right\}$
	$E[R(t) \mid \underline{x}]$	$\left\{ K_{r-1}\left(2\sqrt{\frac{T_r + t^b}{\lambda}}\right) \Big/ K_{r-1}(2\sqrt{T_r/\lambda}) \right\} \cdot \frac{1}{\left(1 + \frac{t^b}{T_r}\right)^{(r-1)/2}}$
	$\mathrm{Var}[R(t) \mid \underline{x}]$	$\frac{1}{K_{r-1}(2\sqrt{T_r/\lambda})} \left\{ \frac{K_{r-1}\left(2\sqrt{\frac{T_r + 2t^b}{\lambda}}\right)}{\left(1 + \frac{2t^b}{T_r}\right)^{(r-1)/2}} - \frac{K_{r-1}^2\left(2\sqrt{\frac{T_r + t^b}{\lambda}}\right)}{K_{r-1}(2\sqrt{T_r/\lambda})\left(1 + \frac{t^b}{T_r}\right)^{r-1}} \right\}$
$g(\theta) = \frac{(u/\theta)^{v+1} e^{-u/\theta}}{\theta \Gamma(v)}$ $(0 < \theta < \infty, u, v > 0)$	$E(\theta \mid \underline{x})$	$\frac{T_r + u}{r + v - 1}$
	$\mathrm{Var}(\theta \mid \underline{x})$	$\frac{(T_r + u)^2}{(r + v - 1)^2 (r + v - 2)}$
	$E[R(t) \mid \underline{x}]$	$\frac{1}{\left(1 + \frac{t^b}{T_r}\right)^{r+v}}$
	$\mathrm{Var}[R(t) \mid \underline{x}]$	$\frac{1}{\left(1 + \frac{2t^b}{T_r + u}\right)^{r+v}} - \frac{1}{\left(1 + \frac{t^b}{T_r}\right)^{2(r+v)}}$

TABLE 4.1.- COMPARISON OF EMPIRICAL BAYES TO THAT OF BAYES AND MVU ESTIMATORS OF

THE SCALE PARAMETER θ IN THE WEIBULL MODEL

Uniform Prior Distribution for θ

Ratio	Empirical Bayes MSE/ Bayes MSE				Empirical Bayes MSE/ MVU MSE			
$E\{Var(\hat{\theta}\|\theta)\}/Var(\theta)$; n; (a,β)	0.45; 10; (1,25)		1.37; 3; (1,100)		0.45; 10; (1,25)		1.37; 3; (1,100)	
No. of Past Experiences	$\xi = 1$	$\xi = 2.5$	$\xi = 1$	$\xi = 2.5$	$\xi = 1$	$\xi = 2.5$	$\xi = 1$	$\xi = 2.5$
2	3.70	3.99	4.22	3.92	2.00	1.86	1.35	1.28
5	1.70	1.87	2.24	2.79	0.83	0.76	0.76	0.79
10	1.65	1.60	2.06	2.10	0.82	0.79	0.67	0.62
15	1.77	1.85	1.90	2.04	0.82	0.84	0.66	0.68

Exponential Prior Distribution for θ

Ratio	Empirical Bayes MSE/Bayes MSE				Empirical Bayes MSE/MVU MSE			
$E\{Var(\hat{\theta}\|\theta)\}/Var(\theta)$; n; λ	0.2; 10; 30		0.67; 3; 15		0.2; 10; 30		0.67; 3; 15	
No. of Past Experiences	$\xi = 1$	$\xi = 2.5$	$\xi = 1$	$\xi = 2.5$	$\xi = 1$	$\xi = 2.5$	$\xi = 1$	$\xi = 2.5$
2	2.00	2.13	1.92	1.86	1.13	1.30	1.17	0.98
5	1.21	1.42	1.76	1.31	0.98	0.92	0.67	0.80
10	1.35	1.37	1.25	1.27	0.98	0.91	0.77	0.84
15	1.41	1.35	1.82	1.27	0.96	0.89	0.74	0.75

Inverted Gamma Prior Distribution for θ

Ratio	Empirical Bayes MSE/Bayes MSE				Empirical Bayes MSE/ MVU MSE			
$E\{Var(\hat{\theta}\|\theta)\}/Var(\theta)$; n; (μ,ν)	0.2; 10; (10,3)		1.0; 3; (80,4)		0.2; 10; (10,3)		1.0; 3; (80,4)	
No. of Past Experiences	$\xi = 1$	$\xi = 2.5$	$\xi = 1$	$\xi = 2.5$	$\xi = 1$	$\xi = 2.5$	$\xi = 1$	$\xi = 2.5$
2	1.84	2.02	2.40	2.40	1.46	1.68	1.31	1.01
5	1.17	1.05	1.28	1.65	0.92	0.95	0.77	0.71
10	1.08	1.04	1.47	1.26	0.91	0.94	0.73	0.72
15	1.10	1.00	1.89	1.48	0.90	0.93	0.69	0.64

TABLE 4.2.- COMPARISON OF MSE OF EMPIRICAL BAYES TO THAT OF BAYES AND MVU ESTIMATORS OF RELIABILITY IN THE WEIBULL MODEL

Uniform Prior Distribution for Θ

Ratio	Empirical Bayes MSE/Bayes MSE				Empirical Bayes MSE/MVU MSE			
$E\{Var(\hat{\theta}\|\theta)\}/Var(\theta)$; n; (α,β)	0.45; 10; (1,25)		1.37; 3; (1,100)		0.45; 10; (1,25)		1.37; 3; (1,100)	
No. of Past Experiences	$\xi = 1$	$\xi = 2.5$	$\xi = 1$	$\xi = 2.5$	$\xi = 1$	$\xi = 2.5$	$\xi = 1$	$\xi = 2.5$
2	4.74	5.98	3.11	5.99	3.40	4.07	1.81	2.94
5	1.28	1.28	1.30	1.60	0.92	0.85	0.75	0.83
10	1.18	1.20	1.28	1.31	0.87	0.84	0.67	0.68
15	1.19	1.28	1.12	1.37	0.86	0.86	0.62	0.72

Exponential Prior Distribution for Θ

Ratio	Empirical Bayes MSE/Bayes MSE				Empirical Bayes MSE/MVU MSE			
$E\{Var(\hat{\theta}\|\theta)\}/Var(\theta)$; n; λ	0.2; 10; 30		0.67; 3; 15		0.2; 10; 30		0.67; 3; 15	
No. of Past Experiences	$\xi = 1$	$\xi = 2.5$	$\xi = 1$	$\xi = 2.5$	$\xi = 1$	$\xi = 2.5$	$\xi = 1$	$\xi = 2.5$
2	3.49	2.95	1.62	3.13	3.20	2.48	1.25	2.11
5	1.12	1.08	0.96	1.22	1.06	0.92	0.70	0.76
10	1.03	1.05	0.96	1.21	0.91	0.93	0.69	0.76
15	1.05	1.03	0.97	1.20	0.95	0.90	0.70	0.75

Inverted Gamma Prior Distribution for Θ

Ratio	Empirical Bayes MSE/Bayes MSE				Empirical Bayes MSE/MVU MSE			
$E\{Var(\hat{\theta}\|\theta)\}/Var(\theta)$; n; (μ,ν)	0.2; 10; (10,3)		1.0; 3; (80,4)		0.2; 10; (10,3)		1.0; 3; (80,4)	
No. of Past Experiences	$\xi = 1$	$\xi = 2.5$	$\xi = 1$	$\xi = 2.5$	$\xi = 1$	$\xi = 2.5$	$\xi = 1$	$\xi = 2.5$
2	3.15	5.30	3.80	8.30	2.25	3.74	1.51	2.68
5	1.22	1.26	1.71	2.41	0.91	0.94	0.71	0.74
10	1.17	1.23	1.49	2.01	0.89	0.89	0.60	0.66
15	1.17	1.18	1.50	1.71	0.86	0.87	0.58	0.68

EQUI-RADIAL DESIGNS FOR WEIGHTED REGRESSION ANALYSIS: FITTING A SECOND-DEGREE MODEL.

John A. Cornell

University of Florida

When constructing design configurations for the purpose of exploring a response surface, one often assumes that the observations possess homogeneous or constant variation within the experimental region. Designs introduced by Box and Draper [1959, 1962] and others have received considerable attention.

We relax the assumption of homogeneous variation among the observations by partitioning the region of interest into two variance zones where if an observation is collected from zone i (i=I,II) it has variance σ_i^2 and $\sigma_{II}^2 = k\sigma_I^2$, k = .25, .5, 1.0, 2.0, 4.0. The designs discussed are second and third-order equi-radial rotatable designs. A weighted least-squares analysis is used enabling us to compare configurations which are optimum when $k \neq 1.0$ to optimum configurations when it is assumed the variance of the observations is constant or k = 1.0.

1. Introduction

When exploring a response surface using the method of Box and Wilson [5], the following assumptions are usually made:

(i) the true surface can be expressed as a function of k controllable factors $X_1, X_2, \ldots, X_k$ and at the u^{th} sample point (u=1,2,...,N), the true surface is written as

$$\eta_u = \phi(X_{1u}, X_{2u}, \ldots, X_{ku}) \quad , \tag{1.1}$$

where X_{iu} is the value (setting) of the i^{th} factor at the u^{th} sample point;

(ii) within a given region of interest, the response surface may be represented by its Taylor's expansion to terms of order d,

$$\eta = \beta_0 X_0 + \beta_1 X_1 + \ldots + \beta_{11} X_1^2 + \ldots + \beta_{12} X_1 X_2 + \ldots + \beta_{111} X_1^3 + \ldots \tag{1.2}$$

where in the subscript of β, the number of times each factor-number appears is the appropriate power of that factor (and $X_0=1$);

(iii) the observed response y_u varies about a mean of η_u with a common variance σ^2 for all values of u, these N errors being uncorrelated, i.e.,

$$y_u = \eta_u + \varepsilon_u \tag{1.3}$$

where

$$E(\varepsilon_u) = 0, \quad E(\varepsilon_u^2) = \sigma^2 , \quad E(\varepsilon_u \varepsilon_{u'}) = 0 \quad u \neq u'; \; u,u'=1,2,\ldots,N \tag{1.4}$$

(iv) the <u>estimate</u> of the value of the response at some point $\underset{\sim}{x}' = (X_1, X_2, \ldots, X_k)$ in the region of interest can be expressed in model (vector) form as,

$$\hat{y}(\underset{\sim}{x}) = \underset{\sim}{x}_1' \underset{\sim}{b}_1 \tag{1.5}$$

where $\underset{\sim}{x}_1' = (1, X_i, \ldots, X_1^2, \ldots)$, $\underset{\sim}{b}_1' = (b_0, b_1, \ldots, b_{11}, \ldots)$ and the b_i's are estimates of the β_i's in (1.2).

In an attempt to approximate the relationship (1.2) with a polynomial model, one performs experiments at N predetermined combinations of the levels of the k controllable factors. These combinations are referred to as the experimental design. Once the N observations are collected, the b_i's in (1.5) may be obtained by the method of least squares using the formula,

$$\underset{\sim}{b}_1 = (\underset{\sim}{X}_1'\underset{\sim}{X}_1)^{-1}\underset{\sim}{X}_1'\underset{\sim}{y} \quad , \tag{1.6}$$

where $\underset{\sim}{X}_1$ is the matrix of values taken by the terms in $\underset{\sim}{x}_1'$ over the N experimental combinations of the k factors and $\underset{\sim}{y}$ is the column vector of the N observed response values. (We have assumed $\underset{\sim}{X}_1'\underset{\sim}{X}_1$ to be non-singular.)

For response surface exploration, a great deal of information is available about a certain class of designs called rotatable designs. Such information may be found in the following papers, [1,2,6,7,8,9,10]. For the present work, however, the papers [8,10,11] provide most of the necessary background in that we shall be concerned primarily with second-order and third-order rotatable designs in two controllable factors.

2. An application

In an industrial combustion stack, the variability of the particulate matter is not uniform across the stack as it approaches the opening at the top of the stack. This lack of uniformity of the variance is owing to the swirling motion of the particles. The swirling action of the particles causes not only a greater concentration of particles but also greater particle variation near the sides of the stack than at the center. An illustration of the variation profile for a particular case is shown in Figure 1. (On a scale of $\sigma^2 = 1$)

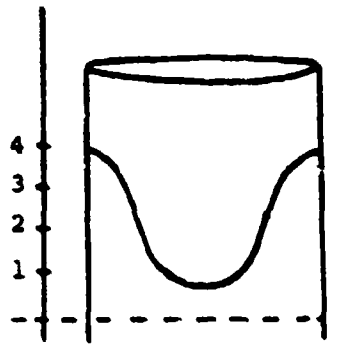


Figure 1. Variance profile across a stack.

The particle concentration distribution across the stack is of interest to us for sampling reasons. Hence, if we wish to use response surface methods in an attempt to describe the distribution, we must be careful about making the assumption of homogeneous variation (1.4). We now consider an alternative to assumption (iii).

3. Development of present work

We shall relax the assumption of homogeneous variation among the N observed responses $\underset{\sim}{y}$ and instead consider the following,

$$E(\underset{\sim}{\varepsilon}) = \underset{\sim}{0} \quad , \tag{3.1}$$

$$\text{Var-cov}(\underset{\sim}{\varepsilon}) = E(\underset{\sim}{\varepsilon}\underset{\sim}{\varepsilon}') = \underset{\sim}{W}\sigma^2 \quad ,$$

where

$$\underset{\sim}{W} = \text{diagonal}\left(\frac{1}{w_1}, \frac{1}{w_2}, \ldots, \frac{1}{w_N}\right) \quad , \tag{3.2}$$

w_u is not necessarily equal to $w_{u'}$, $u \neq u'$.

(When $w_u = 1$ for all $u = 1, 2, \ldots, N$, (3.1)=(1.4) and (3.3)=(1.6)). The form of the variance-covariance matrix (3.2) will amend the estimation equation (1.6) to read,

$$\underset{\sim}{b}_1 = (\underset{\sim}{X}_1'\underset{\sim}{W}^{-1}\underset{\sim}{X}_1)^{-1}\underset{\sim}{X}_1'\underset{\sim}{W}^{-1}\underset{\sim}{y} \quad . \tag{3.3}$$

Furthermore, the variance-covariance structure of the estimates $\underset{\sim}{b}_1$ will take the form,

$$\text{Var-cov } (\underset{\sim}{b}_1) = (\underset{\sim}{X}_1'\underset{\sim}{W}^{-1}\underset{\sim}{X}_1)^{-1}\sigma^2 \tag{3.4}$$

The formula (3.4) arises in the method of <u>weighted least squares</u> (see [12], p. 77-81).

We shall not be concerned directly with the magnitudes of the individual w_u's in (3.2), but rather with how the difference in the magnitudes of different sets of the w_u influences the construction of the design. For example, in Figure 2 a cross-section of the combustion stack of Figure 1 is shown.

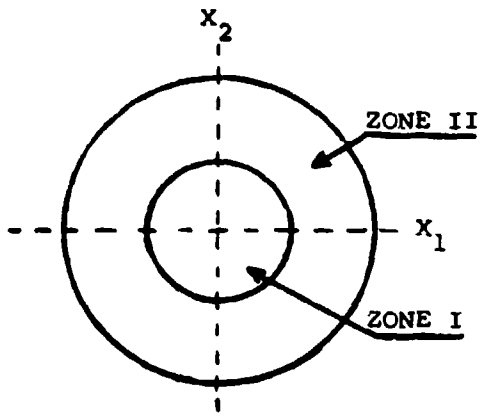


Figure 2. <u>Cross-section of the stack</u>.

Since we know that the observations across the stack do not have constant variability, we attempt to simplify the problem of design construction by first considering the cross-sectional area as being comprised of two variance zones, denoted by ZONE I and ZONE II. If an observation is taken from ZONE I, it possesses a variance σ_I^2; whereas an observation taken from ZONE II possesses a variance of σ_{II}^2. (Consideration of more than two zones is discussed later in Section 5.) The question we wish to answer is this:

If $\sigma_{II}^2 = k\sigma_I^2$ where k is some known value in the interval

(.25, 4.0) how should the experimental design configuration be oriented so as to minimize the quantity,

$$J = \frac{\Omega}{\sigma^2} \int_R E[\hat{y}(\underset{\sim}{x}) - \eta(\underset{\sim}{x})]^2 d\underset{\sim}{x} \tag{3.5}$$

where

$$\Omega^{-1} = \int_R d\underset{\sim}{x} \quad , \tag{3.6}$$

$\hat{y}(\underset{\sim}{x})$ and $\eta(\underset{\sim}{x})$ are defined in (1.5) and (1.2) respectively and R is the region of interest (stack cross-section in our example). The vector $\underset{\sim}{x} = (X_1, X_2)'$ contains the values of the position coordinates across the stack. The quantity J in (3.5) is the mean square error of $\hat{y}(\underset{\sim}{x})$ averaged over the region R.

By configuration orientation is meant; for a given value of k where $\sigma^2_{II} = k\sigma^2_I$, and using two or more sets of equi-radial points, should the design points be placed at the extremes of the two zones? Should the sets of points contain large n_i where n_i is the number of points comprising set i? We now discuss the minimization of J using the method of minimum bias estimation (see [11]).

Development and Minimization of J

It is easy to show that the quantity J in (3.5) is the sum of the average squared bias (B) of the model and the average variance (V) of the model, that is,

$$J = B + V \tag{3.7}$$

where

$$B = \frac{\Omega}{\sigma^2} \int_R [E[\hat{y}(\underset{\sim}{x})] - \eta(\underset{\sim}{x})]^2 d\underset{\sim}{x} \tag{3.8}$$

and

$$V = \frac{\Omega}{\sigma^2} \int_R \text{Var } \hat{y}(\underset{\sim}{x}) d\underset{\sim}{x} \quad . \tag{3.9}$$

As expected, we can concentrate on the minimization of J in the region of interest through the choice of the experimental design. The orientation and size (number of points) of the design will of course depend on the relative magnitudes of the B and V contributions. In fact, Box and Draper [7,8] show that to minimize V alone, one spreads the design points to the boundary of the region of interest. To minimize B alone, however, one decreases the distance from the center of the design to the non-central points. Consequently, when there is the possibility of a contribution from both V and B, one compromises in the size of spread of the points.

A different approach to minimizing J is discussed by Karson, Manson and Hader [11]. This approach involves estimating the parameters in the fitted model to achieve minimum B in (3.8). Subject to achieving minimum B, one then concentrates on minimizing V in (3.9) through the orientation of the design. We now discuss this approach called minimum bias estimation.

Let the fitted equation (second-order model) be written as

$$\hat{y}(\underset{\sim}{x}) = \underset{\sim}{x}_1'\underset{\sim}{b}_1 \tag{3.10}$$

where

$$\underset{\sim}{x}_1' = (1, x_1, x_2, x_1^2, x_2^2, x_1x_2)' \; ; \quad \underset{\sim}{b}_1' = (b_0, b_1, b_2, b_{11}, b_{22}, b_{12})' \; . \tag{3.11}$$

Assume the true response surface is represented by

$$\eta = \underset{\sim}{x}_1'\underset{\sim}{\beta}_1 + \underset{\sim}{x}_2'\underset{\sim}{\beta}_2 \quad , \tag{3.12}$$

where

$$\underset{\sim}{x}_2' = (x_1^3, x_1x_2^2, x_2^3, x_1^2x_2)' \; ; \quad \underset{\sim}{\beta}_2' = (\beta_{111}, \beta_{122}, \beta_{222}, \beta_{112}) \quad . \tag{3.13}$$

Letting

$$E[\hat{y}(\underset{\sim}{x})] = \underset{\sim}{x}_1'\underset{\sim}{\gamma}_1 \quad , \text{ say,} \tag{3.14}$$

it is easy to show (see [11], p. 464, [8], p. 349-350), that, to

minimise the value of B in (3.8), that is, to obtain

$$\text{Min } B = \frac{1}{\sigma^2}\underset{\sim}{\beta}_2'\left[\underset{\sim}{\mu}_{22} - \underset{\sim}{\mu}_{12}'\underset{\sim}{\mu}_{11}^{-1}\underset{\sim}{\mu}_{12}\right]\underset{\sim}{\beta}_2$$

where

$$\underset{\sim}{\mu}_{ij} = \Omega\int_R \underset{\sim}{x}_i\underset{\sim}{x}_j' \, d\underset{\sim}{x} \qquad i\leq j=1,2 \qquad (3.15)$$

we have for $\underset{\sim}{\gamma}_1$,

$$\underset{\sim}{\gamma}_1 = \underset{\sim}{\beta}_1 + \underset{\sim}{\mu}_1^{-1}\underset{\sim}{\mu}_{12}\underset{\sim}{\beta}_2 \ . \qquad (3.16)$$

The matrices $\underset{\sim}{\mu}_{ij}$ in (3.15) are called region moment matrices. (See [7,8]). Rewriting $\underset{\sim}{\gamma}_1$ as

$$\underset{\sim}{\gamma}_1 = \underset{\sim}{A}\underset{\sim}{\beta} = \left[\underset{\sim}{I}\vdots\underset{\sim}{\mu}_{11}^{-1}\underset{\sim}{\mu}_{12}\right]\begin{bmatrix}\underset{\sim}{\beta}_1\\ \underset{\sim}{\beta}_2\end{bmatrix} \qquad (3.17)$$

we see that a necessary and sufficient condition for the minimization of B is simply,

$$E(\underset{\sim}{b}_1) = \underset{\sim}{A}\underset{\sim}{\beta} \ . \qquad (3.18)$$

Since $E(\underset{\sim}{y}) = \underset{\sim}{X}\underset{\sim}{\beta}$ where $\underset{\sim}{X} = (\underset{\sim}{X}_1\vdots\underset{\sim}{X}_2)$, and if we express $\underset{\sim}{b}_1$ as a linear combination of the observations,

$$\underset{\sim}{b}_1 = \underset{\sim}{T}'\underset{\sim}{y} \qquad (3.19)$$

then from (3.18) the matrix $\underset{\sim}{T}'$ must satisfy,

$$\underset{\sim}{T}'\underset{\sim}{X} = \underset{\sim}{A} \ . \qquad (3.20)$$

Thus Min B can be obtained as long as $\underset{\sim}{A}\underset{\sim}{\beta}$ is estimable.

Since Var-cov$(\underset{\sim}{y}) = \underset{\sim}{W}\sigma^2$, it is worth mentioning that designs satisfying the condition,

$$(\underset{\sim}{X}_1'\underset{\sim}{W}^{-1}\underset{\sim}{X}_1)^{-1}\underset{\sim}{X}_1'\underset{\sim}{W}^{-1}\underset{\sim}{X}_2 = \underset{\sim}{\mu}_{11}^{-1}\underset{\sim}{\mu}_{12} \qquad (3.21)$$

and further if we use the least squares estimator (see (3.4)),

$$\underset{\sim}{T}' = (\underset{\sim}{X}_1'\underset{\sim}{W}^{-1}\underset{\sim}{X}_1)^{-1}\underset{\sim}{X}_1'\underset{\sim}{W}^{-1} \qquad (3.22)$$

so that $\underset{\sim}{T}'\underset{\sim}{X} = \underset{\sim}{A}$, the weighted least squares estimator (3.4) does achieve Minimum B.

Subject to achieving Min B using the estimator (3.19) where

$$\underset{\sim}{T}' = \underset{\sim}{A}(\underset{\sim}{X}'\underset{\sim}{W}^{-1}\underset{\sim}{X})^{-1}\underset{\sim}{X}'\underset{\sim}{W}^{-1} \tag{3.23}$$

we now attempt to minimize V in (3.9) through the orientation of the design. The quantity is minimized when $\hat{y}(\underset{\sim}{x})$ is the least squares estimate of its expectation at all points in R. Since $E[\hat{y}(\underset{\sim}{x})] = \underset{\sim}{x}_1'\underset{\sim}{\gamma}_1 = \underset{\sim}{x}_1'\underset{\sim}{A}\underset{\sim}{\beta}$, then V is minimized when

$$\begin{aligned} \underset{\sim}{x}_1'\underset{\sim}{b}_1 &= \underset{\sim}{x}_1'\underset{\sim}{T}'\underset{\sim}{Y} \\ &= \underset{\sim}{x}_1'\underset{\sim}{A}(\underset{\sim}{X}'\underset{\sim}{W}^{-1}\underset{\sim}{X})^{-1}\underset{\sim}{X}'\underset{\sim}{W}^{-1}\underset{\sim}{y} \end{aligned} \tag{3.24}$$

that is, when $\underset{\sim}{T}'$ is as shown in (3.23). Furthermore, the variance of the fitted equation is,

$$\text{Var } \hat{y}(\underset{\sim}{x}) = \underset{\sim}{x}_1'\underset{\sim}{A}(\underset{\sim}{X}'\underset{\sim}{W}^{-1}\underset{\sim}{X})^{-1}\underset{\sim}{A}'\underset{\sim}{x}_1\sigma^2 \tag{3.25}$$

and V in (3.9) can be written as

$$V = \text{trace}[\underset{\sim}{A}(\underset{\sim}{X}'\underset{\sim}{W}^{-1}\underset{\sim}{X})^{-1}\underset{\sim}{A}'\underset{\sim}{\mu}_{11}] \ . \tag{3.26}$$

We now discuss the design configurations to be used for the purpose of minimizing J in (3.5). The designs are a special class of third-order rotatble designs in two variables (see[10]).

4. Designs

As we mentioned previously, we shall consider fitting a model of the second degree and assume the true polynominal relationship η is of the third degree. The design configurations used will be rotatable designs consisting of two sets of equi-radial points lying on concentric circles. The inner set will consist of $n_1 \geq 7$ equi-radial points located at a distance $\rho_1 > 0$ from the center ($X_1 = 0$, $X_2 = 0$) of the design space whereas the outer set of points, $n_2 \geq 7$ in number, will be located at a distance $\rho_2 > \rho_1$ from the center of the design space. We shall also

require $\rho_1 \leq$ boundary of ZONE I and $\rho_2 \leq$ boundary of ZONE II. For simplicity of development, we set $0 < \rho_1 \leq .5$ and $.5 < \rho_2 \leq 1.0$.

The use of this particular class of designs enables us to specify the form of the symmetrical matrix $\underset{\sim}{X}'\underset{\sim}{W}^{-1}\underset{\sim}{X}$ of Equations (3.23) and (3.26) to be as follows:

$$\underset{\sim}{X}'\underset{\sim}{W}^{-1}\underset{\sim}{X} = \begin{bmatrix} a & 0 & 0 & b & b & 0 & 0 & 0 & 0 & 0 \\ & b & 0 & 0 & 0 & 0 & 3d & d & 0 & 0 \\ & & b & 0 & 0 & 0 & 0 & 0 & 3d & d \\ & & & 3d & d & 0 & 0 & 0 & 0 & 0 \\ & & & & 3d & 0 & 0 & 0 & 0 & 0 \\ & & & & & d & 0 & 0 & 0 & 0 \\ & & & & & & 5f & f & 0 & 0 \\ & & (\text{Same}) & & & & & f & 0 & 0 \\ & & & & & & & & 5f & 0 \\ & & & & & & & & & f \end{bmatrix} \tag{4.1}$$

where the quantities a, b, d and f are,

$$a = \Sigma w_u \ , \quad b = \Sigma x_{iu}^2 w_u \ , \quad d = \Sigma x_{1u}^2 x_{2u}^2 w_u \tag{4.2}$$

$$f = \Sigma x_{1u}^4 x_{2u}^2 w_u = \Sigma x_{1u}^2 x_{2u}^4 w_u \qquad (i=1,2)$$

and the summations are over u=1,2,...,N. Furthermore, using n_i and ρ_i in place of X_i (i=1,2) in (4.2), we have for a, b, d and f:

$$a = n_1 w_1 + n_2 w_2 \ , \quad b = \frac{1}{2} \sum_{i=1}^{2} n_i \rho_i^2 w_i$$

$$d = \frac{1}{8} \sum_{i=1}^{2} n_i \rho_i^4 w_i \quad \text{and} \quad f = \frac{1}{16} \sum_{i=1}^{2} n_i \rho_i^6 w_i \tag{4.3}$$

since the w_i (i=1,2) are constant for all n_i points located at a distance ρ_i from the center of the design.

Pausing briefly, the form of the matrix $\underset{\sim}{X}'\underset{\sim}{W}^{-1}\underset{\sim}{X}$ is as specified in (4.1) as long as we use at least two equi-radial sets of points where the radii of two or more sets are greater than zero and $n_i \geq 7$. Of course, to obtain the estimates using the formula (3.3) we would need only one equi-radial set of points, the radius of which is greater than zero, consisting of $n \geq 5$ points and $n_0 \geq 1$ additional

center point replicates. With this latter design configuration however, the matrix $\underset{\sim}{X}'\underset{\sim}{W}^{-1}\underset{\sim}{X}$ in (4.1) is singular. To obtain a value of V in (3.26), the use of a generalized universe matrix, the form of which is not unique, would be necessary. This would complicate the design problem considerably.

A further note to consider is that in addition to the $n_1+n_2 \geq 14$ points mentioned above, we could run n_0 center point replicates. These additional n_0 points would help us to reduce V in (3.26). However, since we are talking already about designs consisting of at least 14 points, we dismiss the notion of additional n_0 center points.

The symmetric matrix $\underset{\sim}{\mu}_{11}$ (see (3.15)) takes the form,

$$\underset{\sim}{\mu}_{11} = \begin{bmatrix} 1 & 0 & 0 & \rho^{*2}/4 & \rho^{*2}/4 & 0 \\ & \rho^{*2}/4 & 0 & 0 & 0 & 0 \\ & & \rho^{*2}/4 & 0 & 0 & 0 \\ & & & 3\rho^{*4}/24 & \rho^{*4}/24 & 0 \\ & \text{(Same)} & & & 3\rho^{*4}/24 & 0 \\ & & & & & \rho^{*4}/24 \end{bmatrix} \qquad (4.4)$$

where ρ^* is the radius of the region of interest (from the center to the boundary of ZONE II). Since we previously set $\rho^* = 1.0$ and from the description of the matrix $\underset{\sim}{A}$ in (3.17), the average variance V in (3.26) can now be written as, (using (4.1), (4.3) and (4.4),

$$V = \frac{1}{12d} + \frac{a+24d-6b}{12(2ad-b^2)} + \frac{18f+b-12d}{72(bf-2d^2)} + \frac{90f+7b-84d}{72(5bf-14d^2)} \quad . \qquad (4.5)$$

Minimizing V in (4.5) then is analogous to minimizing J in (3.5).

Values of V in (4.5) are presented in Tables 1-5 corresponding to the following values of n_i, ρ_i and k:

$$n_1 = 7,8,10 \qquad n_2 = 7,8,10$$

$$\rho_1 = .1,\ .25,\ .50 \qquad \rho_2 = .80,\ .95,\ 1.0 \quad .$$

$$k = 0.25,\ 0.5,\ 1.0,\ 2.0,\ 4.0 \quad .$$

Only included in Tables 1-5 however, is the best ρ_2 value for each ρ_1 value defined above. For example, when $k = 0.5$, the values of V when $\rho_1 = 0.1$, $\rho_2 = 0.8$ are less than the values of V when $\rho_1 = 0.1$ and $\rho_2 = 0.95$ or $\rho_1 = 0.1$ and $\rho_2 = 1.0$ for all corresponding values of n_1 and n_2. Thus x represents V at $\rho_1 = 0.1$, $\rho_2 = 0.8$ in Table 2.

In Table 3, the value of k is $k = 1.0$. Table 3 provides us with an insight to the behavior of V for the above n_1, n_2, ρ_1 and ρ_2 values when it is assumed the variance of the observations is constant in the region of interest.

Table 1. Values of V when $k = 0.25$.

n_1 =	7			8			10		
n_2 =	7	8	10	7	8	10	7	8	10
x:	.78	.71	.61	.75	.68	.58	.72	.65	.55
Δ:	.77	.69	.59	.74	.67	.57	.71	.64	.54
o:	.73	.70	.66	.67	.64	.60	.59	.56	.51

x: ρ_1=0.1, ρ_2=0.8 Δ: ρ_1=0.25, ρ_2=0.8 o: ρ_1=0.5, ρ_2=0.95

Table 2. Values of V when $k = 0.50$.

n_1 =	7			8			10		
n_2 =	7	8	10	7	8	10	7	8	10
x:	.68	.61	.51	.67	.59	.49	.65	.58	.48
Δ:	.55	.52	.47	.52	.48	.44	.47	.43	.39
o:	.48	.45	.42	.44	.42	.38	.39	.37	.33

x: ρ_1=0.1, ρ_2=0.8 Δ: ρ_1=0.25, ρ_2=0.95 o: ρ_1=0.5, ρ_2=1.0

Table 3. Values of V when k = 1.0.

n_1 =	7			8			10		
n_2 =	7	8	10	7	8	10	7	8	10
x:	.63	.56	.46	.62	.55	.45	.61	.54	.44
Δ:	.41	.38	.33	.40	.36	.31	.37	.34	.29
o:	.33	.31	.28	.32	.29	.26	.29	.27	.23

x: ρ_1=0.1, ρ_2=0.8 Δ: ρ_1=0.25, ρ_2=0.95 o: ρ_1=0.5, ρ_2=1.0 ρ_2=1.0

Table 4. Values of V when k = 2.0.

n_1 =	7			8			10		
n_2 =	7	8	10	7	8	10	7	8	10
x:	.55	.51	.47	.52	.48	.43	.47	.43	.38
Δ:	.31	.29	.25	.30	.27	.24	.28	.26	.22
o:	.26	.24	.20	.25	.23	.19	.23	.21	.18

x: ρ_1=.1, ρ_2=.95 Δ: ρ_1=0.25, ρ_2=1.0 o: ρ_1=.5, ρ_2=1.0

Table 5. Values of V when k = 4.0.

n_1 =	7			8			10		
n_2 =	7	8	10	7	8	10	7	8	10
x:	.41	.38	.33	.40	.36	.31	.37	.34	.29
Δ:	.26	.23	.20	.25	.23	.19	.24	.22	.18
o:	.20	.19	.16	.15	.18	.16	.23	.26	.14

x: ρ_1=0.1, ρ_2=0.95 Δ: ρ_1=0.25, ρ_2=1.0 o: ρ_1=0.5, ρ_2=1.0

Discussion of Tables 1. - 5.

From the entries of V in Table 1., we notice little difference in the value of V for ρ_1=0.1 and ρ_1=0.25 when ρ_2=0.8 for all n_1 and n_2. Increasing n_1 results in a reduction of V when both ρ_1 and ρ_2 are large; thus, if we can afford to increase n_1, we should spread both radii. Increasing n_2 results in a reduction of V for all n_1, ρ_1 and ρ_2. Finally, for small ρ_1(ρ_1=0.1, 0.25), ρ_2=0.8 is better than ρ_2=0.95 or ρ_2=1.0.

From Table 2., we see that the smaller the radii ρ_1 and ρ_2, the larger the reduction in V resulting from increasing n_2. Increasing n_2 reduces V for all n_1, ρ_1 and ρ_2. When both ρ_1 and ρ_2 are large, increasing n_1 results in a reduction of V.

When the variance of the observations is constant throughout the experimental region (Table 3.), the value of V decreases as we increase ρ_1 and ρ_2. Very little affect on the value of V as we increase n_1. However, as we increase n_2, the following reductions in the value of V occur: on a scale of 100 percent, approximately 6% reduction per one unit increase in n_2 using x; approximately 3% reduction per one unit increase in n_2 using Δ, and, approximately 2% per one unit increase in n_2 using o.

From Table 4., we see that the smallest value of V results from using large ρ_1 and ρ_2. Increasing n_1 has little affect on V except for small ρ_1. Approximately a 2% decrease in the value of V for each increasing value of n_2 when ρ_1=0.25 or ρ_1=0.50 when ρ_2=1.0.

As expected, in Table 5., we notice that ρ_2 should be large. Little change in the value of V results from increasing n_1 when ρ_1=0.1 or ρ_1=0.25. However, when ρ_1=0.5, the value of V became erratic when we increase n_1 and n_2. Increasing n_2 results in a reduction of the value of V except when $n_1 \geq 8$ and ρ_1 is large (ρ_1=.25, 0.50).

5. Summary

In this paper, we have discussed the problem of design construction (using rotatable designs) for fitting a 2nd order polynomial in two variables when it is known that the observations do not possess homogeneous or constant variability throughout the experimental region. We divided the region of interest into two circular variance zones where if an observation is taken from zone i (i = I, II), it has variance σ_i^2 and $\sigma_{II}^2 = k\sigma_I^2$ where k takes one of the values 0.25, 0.5, 1.0, 2.0, 4.0. The case of only two variance zones was discussed owing to the limited presentation time. More than two zones is presently under investigation.

Although more extreme values of k (i.e., k = 0.05, 0.1, 10.0) were used in addition to the values above in the investigation leading to the writing of this paper, it is hoped that the values of k used in this paper lend some light on the problem of design construction when $k \neq 1.0$. From the entries of the value of V presented in Tables 1.-5., we conclude the following: When $k > 1.0$, and we spread both design radii to the extremes, we would be using an acceptable design. However, in assuming k = 1.0 and calculating the value of V for the above design, we <u>overestimate</u> the value of V by the following amounts:

when k = 20, using $n_1 = 7$, we overestimate V by approx. 7%
$n_1 = 8$, " " " " " 7%
$n_1 = 10$, " " " " " 6%

when k = 4.0, using $n_1 = 7$, we overestimate V by approx. 13%
$n_1 = 8$, " " " " " 13%
$n_1 = 10$, " " " " " 15%

when $k < 1.0$, the above design (both radii spread to the extremes) might be adequate when k is close to 1.0 (i.e., $.4 \leq k < 1.0$) and $n_1 > 8$.

However, when $k < .4$, we should reduce ρ_2 to approximately 0.8 and also use $.1 \leq \rho_1 \leq .3$. In using the design with $\rho_1 = 0.5$ and $\rho_2 = 1.0$, we <u>underestimate</u> the value of V by the following quantities:

when $k = 0.50$, using $n_1 = 7$, we underestimate V by approximately 14%
$n_1 = 8$, " " " " " 12%
$n_1 = 10$, " " " " " 10%

when $k = 0.25$ and $n_1 = 10$, when we use
$n_2 = 7$, we underestimate V by approximately 30%
$n_2 = 8$, " " " " " 29%
$n_2 = 10$, " " " " " 28%

References.

[1] R. C. Bose and N. R. Draper, (1959). Second-order rotatable designs in three dimensions., Ann. Math. Statist., Vol. 30, p. 1097-1112.

[2] R. C. Bose and R. L. Carter, (1959). Complex representation in the construction of rotatable designs., Ann. Math Statist., Vol. 30, p. 771-780.

[3] G. E. P. Box, (1952). Multi-factor designs of first order. Biometrika, Vol. 39, p. 49-57.

[4] G. E. P. Box, (1954). The exploration and exploitation of response surfaces: Some general considerations and examples. Biometrics, Vol. 10, p. 16-60.

[5] G. E. P. Box and K. B. Wilson, (1951). On the experimental attainment of optimum conditions. J. Royal Statist. Soc., Series B, Vol. 13, p. 1-45.

[6] G. E. P. Box and J.S. Hunter, (1957). Multi-factor experimental designs for exploring response surfaces. Ann. Math. Statist., Vol. 28, p. 195-241.

[7] G. E. P. Box and N. R. Draper, (1959). A basis for the selection of a response surface design. J. Amer. Statist. Assoc., Vol. 54, p. 622-654.

[8] G. E. P. Box and N. R. Draper, (1962). The choice of a second-order rotatable design. Biometrika, Vol. 50, p. 335-352.

[9] R. L. Carter, (1957). New designs for the exploration of response surfaces. Technical Report No. 172, Institute of Statistics, Univ. of North Carolina.

[10] D. A. Gardiner, A. H. E. Grandage and R. J. Hader, (1959). Third order rotatable designs for exploring response surfaces. Ann. Math. Statist., Vol. 30, p. 1082-1096.

[11] M. J. Karson, A. R. Manson and R. J. Hader, (1969). Minimum bias estimation and experimental designs for response surfaces. Technometrics, Vol. 11, No. 3, p. 461-475.

[12] N. R. Draper and H. Smith (1966). Applied Regression Analysis. John Wiley & Sons, Inc.,

SYSTEM PARAMETER OPTIMIZATION USING RESPONSE SURFACE METHODOLOGY

Mr. Garry W. Barnard
U.S. Army Missile Command

I. OPTIMIZATION

This paper considers the problem of determining the optimum set of missile system performance parameters using results from computer simulations. The tool used for determining this optimum is Response Surface Methodology. The concept of response surface methodology involves a dependent variable (Y), such as percent aircraft targets killed, and several independent or controlled variables ($X_1, X_2, X_3, \ldots X_k$), such as radar detection range, single shot kill probability, reaction time, etc. If all of these variables are assumed measurable, the response surface can be expressed as

$$Y = f(X_1, X_2, X_3, \ldots X_k). \tag{1}$$

The empirical response function, expressed in the standard linear regression notation, is given by

$$\underline{Y} = X\underline{\beta} + \underline{\varepsilon} \tag{2}$$

where

i $\underline{Y}$ is an N × 1 vector of observations or responses

ii X is an N × k matrix of constants which are the preselected levels of the independent variables

iii $\underline{\beta}$ is a k × 1 vector of unknown coefficients

iv $\underline{\varepsilon}$ is an N × 1 vector of the error or deviations from the true response to the estimated value.

The estimator for the unknown coefficient vector $\underline{\beta}$ is determined by least squares, which is given by

$$\underline{b} = (X'X)^{-1}X'\underline{Y} \tag{3}$$

where $\underline{b}$ is the estimated vector for $\underline{\beta}$.

A fitted second order response function, given by equation (2) can be written as

$$\hat{\underline{Y}} = b_o + \underline{X}'\underline{b}^* + \underline{X}'B\underline{X} \tag{4}$$

This article has been reproduced photographically from the author's manuscript.

where b_o is the intercept and

$$\underline{X} = \begin{bmatrix} X_1 \\ X_2 \\ \cdot \\ \cdot \\ \cdot \\ X_k \end{bmatrix} \qquad \underline{b}^* = \begin{bmatrix} b_1 \\ b_2 \\ \cdot \\ \cdot \\ \cdot \\ b_k \end{bmatrix} \qquad B = \begin{bmatrix} b_{11} & b_{12/2} \cdots b_{1_{k/2}} \\ & b_{22} \cdots b_{1_{k/2}} \\ & \cdot \\ & \cdot\cdot \\ & \cdot \\ & b_{kk} \end{bmatrix}.$$

Notice that the $\underline{X}'\underline{b}^*$ portion gives the first order terms in the response function and the quadratic form $\underline{X}'B\underline{X}$ gives the quadratic contribution. The stationary or critical point is where the derivatives $\partial\hat{Y}/\partial X_1$, $\partial\hat{Y}/\partial X_2 \ldots, \partial\hat{Y}/\partial X_k$ are simultaneously zero, which is given by the vector

$$\underline{X}_o = -1/_2 B^{-1}\underline{b}^*. \tag{5}$$

The stationary point $\underline{X}_o$ can be a maximum, a minimum, or a saddle point. Therefore, it will be necessary to determine the exact nature of the stationary point. It is convenient to express the second order response function (4) in a different form which can be more clearly interpreted by the analyst. This involves translating the response function from the origin ($X_1 = 0$, $X_2 = 0, \ldots X_k = 0$) to the stationary point $\underline{X}_o$. Then the response function can be expressed in terms of new variables W_1, $W_2, \ldots W_k$. Figure 1 illustrates this for two variables where the contours represent constant responses.

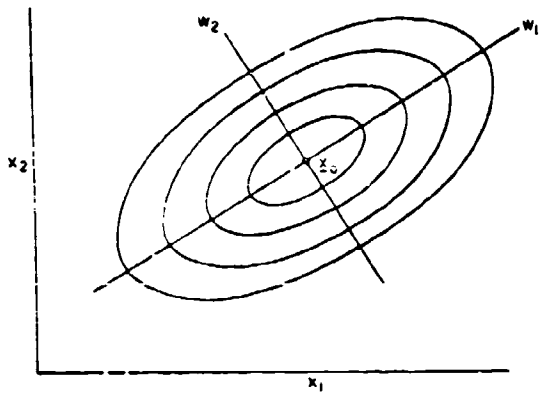


Figure 1. Canonical Form for a Response Surface in Two Variables.

The function in terms of the new variables is called the CANONICAL FORM and is given by

$$\hat{Y} = \hat{Y}_o + \lambda_1 W_1^2 + \lambda_1 W_2^2 + \ldots + \lambda_k W_k^2 \qquad (6)$$

where $\hat{Y}$ is the estimated response at the stationary point and λ_i's are the characteristic roots or eigenvalues of the B matrix. The new variables $W_1, W_2, \ldots W_k$ are related to the original variables $X_1, X_2, \ldots X_k$ as follows

$$\underline{W} = M'[\underline{X} - \underline{X}_o] \qquad (7)$$

where M is a k × k orthogonal matrix the columns of which are the normalized eigenvectors associated with the eigenvalues λ_i. Reducing the response function (4) to the canonical form (6) is called a CANONICAL ANALYSIS.

Upon completing the canonical analysis, the analyst can more readily interpret the response surface. By observing the sign of the eigenvalues λ_i which, of course, are the roots of the following equation

$$\left| B - \lambda I_k \right| \qquad (8)$$

the nature of the stationary point can be determined.

It can be seen from the canonical form (6) that if $(\lambda_1, \lambda_2, \ldots \lambda_k)$ are all negative, a move in any direction from the stationary point results in a decrease in $\hat{Y}$. Therefore, the stationary point is a point of maximum response for the fitted surface. If $(\lambda_1, \lambda_2, \ldots \lambda_k)$ are all positive, $\underline{X}_o$ represents a minimum for the fitted surface. For the case where the λ's differ in sign, the stationary point is a saddle point. Figure 2 illustrates for two variables, a situation where the stationary point is a saddle point.

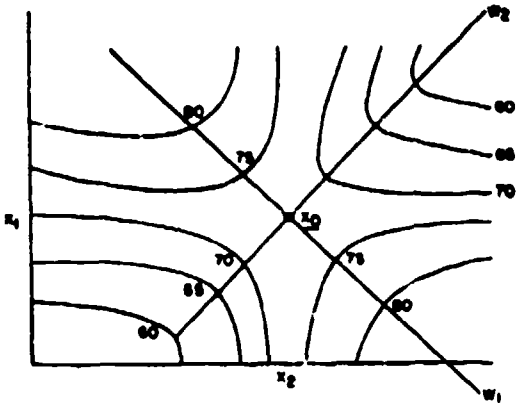


Figure 2. Saddle Point.

In this illustration $\lambda_1 > 0$ and $\lambda_2 < 0$; a move along the W_2 axis away from the stationary point in either direction results in a decrease in estimated response. Moving along the W_1 axis, a corresponding increase in response is obtained. The same indications could be gained in this illustration by looking only at the canonical form. Equation (6) for the example would be of the form

$$\hat{Y} = \hat{Y}_o + \lambda_1 W_1^2 - \lambda_2 W_2^2.$$

Clearly, increases in response would be expected the smaller W_2 becomes, say zero, and the larger W_1 becomes. This is equivalent to moving along the W_1 axis away from X_o in either direction. The point being made here is that when there are more than two independent variables involved, the analyst has to rely on the canonical analysis rather than on drawings of the response surface to do his analysis.

Two examples will be given here to illustrate this analytical tool. The data used for these examples are results from an air defense simulation of hypothetical air defense systems against a low altitude aircraft attack in mountainous terrain. In the first example, the two parameters to be optimized are the air defense system's radar detection range and radar tracking range.

The input values for the two independent variables and the corresponding response (percent aircraft kills) from the simulation are given in Table I. Input data for these examples are coded such that

i the midpoint value of any parameter is 0,

ii the smallest value of any parameter is -1,

iii the largest value of any parameter is +1.

Table I. X_1, X_2, and Y for Example 1.

Detection Range (X_1)	Tracking Range (X_2)	Percent Aircraft Killed (Y)
-1	-1	30
0	-1	50
1	-1	51
-1	0	32
1	0	48
-1	1	32
0	1	44
1	1	45
0	0	54
0	0	56
0	0	53
0	0	54
0	0	52

The midpoint for the two variables was replicated five times to get an estimate of the experimental error. Using the data given in Table I, the following response function

$$\hat{Y} = 52.93 + 8.3333X_1 - 1.667X_2 - 10.759X_1^2 - 3.759X_2^2 - 2.0X_1X_2$$

was obtained by least squares. The analysis of variance in Table II below tests the significance of the linear and quadratic contributions to the response and the significance of the lack of fit.

Table II. Analysis of Variance, Example 1

Source	Degrees of Freedom	Sum of Squares	Mean Squares	F-Test
Regression	5	968.468		
Linear	2	433.333	216.667	98.485*
Quadratic	3	535.135	178.378	81.081*
Lack of Fit	3	33.039	11.013	5.006
Error	4	8.8	2.20	
Total	12	1010.308		

The tabulated F values at the 0.05 significance level for the degrees of freedom indicated in the table vary from 6.04 to 6.94, which indicates that there is a very strong linear contribution, a significant quadratic contribution, and an insignificant lack of fit. This implies that the fitted model is adequate, namely, that it is not necessary to consider higher order terms.

Applying equation (5) results in the following stationary point

DETECTION RANGE = 0.418

TRACKING RANGE = -0.333

The eigenvalues given by equation (8) are

$$\lambda_1 = -3.618 \text{ and } \lambda_2 = -10.899$$

which means that the stationary point is the point of maximum response. The estimated response at the stationary point is 54.39. This system is depicted in Figure 3.

The drawing shows that tracking range is the more sensitive parameter.

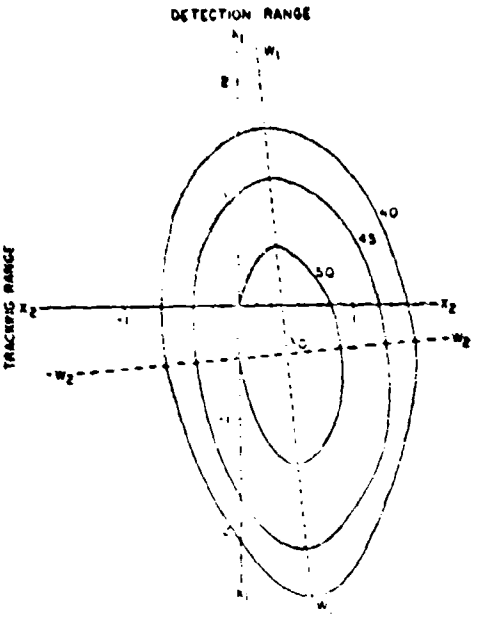


Figure 3. Example 1.

The next example attempts to determine the optimum values for the average missile velocity and radar tracking range for a hypothetical missile system against the same type threat and in the same type environment as the first example. The input data and the results from the simulation are coded as before and given in Table III.

Table III. X_1, X_2, and Y for Example 2.

Average Velocity (X_1)	Tracking Range (X_2)	Percent Aircraft Killed (Y)
-1	-1	34
0	-1	38
1	-1	42
-1	0	50
1	0	60
-1	1	50
0	1	56
1	1	62
0	0	56
0	0	54
0	0	56
0	0	58
0	0	57

Again the midpoint ($X_1 = 0$, $X_2 = 0$) was replicated five times to estimate the experimental error. Applying the methods of least squares resulted in the following response function

$$\hat{Y} = 56.034 + 5.0X_1 + 9.0X_2 - 0.6207X_1^2 - 8.6207X_2^2 + X_1X_2$$

The analysis of variance performed on this example is given in Table IV.

Table IV. Analysis of Variance, Example 2.

Source	Degrees of Freedom	Sum of Squares	Mean Squares	F-Test
Regression	5	894.514		
Linear	2	635.999	317.999	158.99*
Quadratic	3	285.516	86.171	39.169
Lack of Fit	3	0.993	0.331	0.150
Error	4	8.8	2.2	
Total	12	904.308		

The same types of conclusions are reached with this analysis of variance as in the previous example. The linear contribution is very strong, the quadratic contribution is significant, and the lack of fit is very small.

The stationary point for this system is

AVERAGE VELOCITY = 4.666

TRACKING RANGE = 0.7926

which is far removed from the experimental region. The eigenvalues are

$\lambda_1 = -8.6518$ and $\lambda_2 = -0.5896$

which again indicates a maximum; but since the stationary point is remote from the region of experimentation, no valid conclusions can be drawn about the surface around it. However, by looking at equation (6) expressed for this example

$$\hat{Y} = \hat{Y}_0 - 8.6518W_1^2 - 0.5896W_2^2$$

it can be seen that increases in response can be obtained by W_1 equal to zero and moving along the W_2 axis toward the stationary point. This case is illustrated in Figure 4.

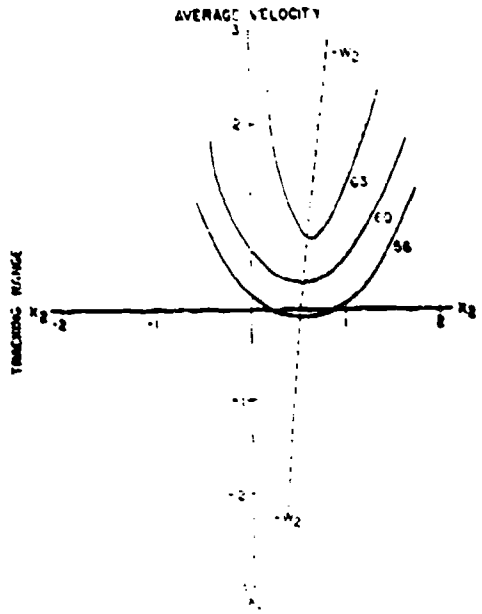


Figure 4. Example 2.

In the next section this example will be used for illustrating a constrained optimum, since the stationary point is far removed from the experimental region.

II. CONSTRAINED OPTIMIZATION

If an absolute optimum cannot be obtained within the ranges of the system parameters, it becomes necessary to determine a constrained optimum. The usual mathematical or quadratic programming techniques are not used here since they require a concave function for maximization, and this is not always the case in this type of optimization. In fact, saddle points occur very frequently. A method will be given here for determining a constrained optimum based on work done by Hoerl [1959] and Draper [1963]. The essential requirements for this method are (i) the experimental error should be small, and (ii) the lack of fit should not be too great.

Using coded variables as described in the previous section, local optima will be obtained for each sphere formed by varying radii R from the center of the experimental region (0,0,...0) to points $(X_1, X_2, \ldots X_k)$. The procedure can easily be illustrated for a two-variable case given in Figure 5.

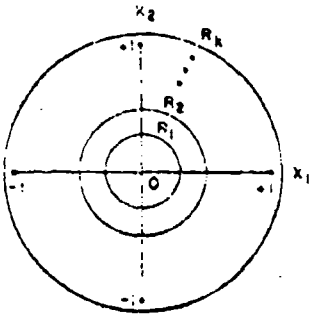


Figure 5. Constrained Optimization for Two Variables.

The spheres for this two-variable case reduce to circles. This method, by restricting itself to spheres of varying radii, finds a local maximum on each of the spheres (circles in Figure 5). When these restricted optima are determined, several two-dimensional plots can be drawn of $\hat{Y}$ against $X_1, X_2, \ldots X_k$, the coordinates that give rise to these optima, and $\hat{Y}$ against R. The analyst can then observe these plots and find the radius that gives the maximum response and the corresponding value of each of the independent variables that yields this maximum.

For each radius R, the condition on the independent variables $(X_1, X_2, \ldots X_k)$ which will maximize $\hat{Y}$ subject to the constraint

$$\sum_{i=1}^{n} X_i^2 = R^2 \tag{9}$$

will be determined. In matrix notation, this constraint is equivalent to

$$(\underline{X}'\underline{X} - R^2). \tag{10}$$

To maximize $\hat{Y}$ subject to the constraint (10), consider the function

$$F = b_0 + \underline{X}'\underline{b}^* + \underline{X}'B\underline{X} - \lambda(\underline{X}'\underline{X}-R^2) \tag{11}$$

where λ is a Lagrange multiplier and $\underline{X}' = [X_1, X_2, \ldots X_k]$. The derivative of F with respect to the vector $\underline{X}$ is given by

$$\partial F/\partial X = \underline{b}^* + 2B\underline{X} - 2\lambda\underline{X}. \tag{12}$$

Equating (12) to zero and solving for $\underline{X}$ gives

$$(B-\mu I_R)\underline{X} = -1/2\underline{b}^*. \qquad (13)$$

By selecting the proper value of the Lagrange multiplier μ and inserting in equation (13), the condition on the independent variables $(X_1, X_2, \ldots X_k)$ can be determined, which will maximize $\hat{Y}$ for this sphere. The radius of the sphere can be determined from equation (10) and $\hat{Y}$ from equation (4) of the previous section. It should be pointed out here that this is only the maximum on this one particular sphere of radius R. With several preselected values of μ, this procedure would give local maxima on spheres of varying radii. Then plots as suggested earlier could be given as in Figure 6.

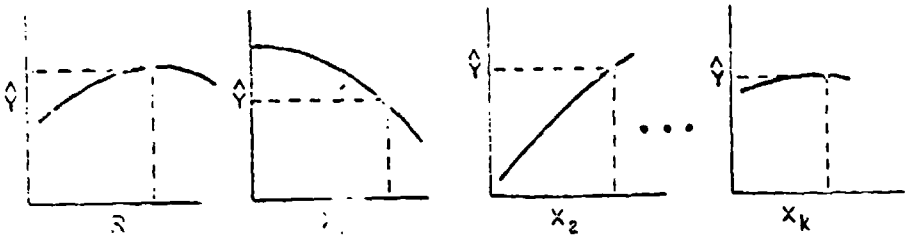


Figure 6. Two Dimensional Plots.

The first plot gives the radius R from the center of the experimental region where the maximum response is expected while the other plots show the values of the independent variables that yield this maximum. Not only do these plots give the X values that produce the maximum response, but the most important result from this analysis is that a path can be determined which, if followed, should give the quickest gains in response. Again, confirmatory runs of the simulation should be made along this path to insure that the prediction is correct.

The second example considered in the previous section will be used here to illustrate this constrained optimization procedure. Recall that in the previous example the stationary point was a considerable distance from the experimental region.

Table V gives the results of this analysis.

Table V. Constrained Optimization Results.

μ	Average Velocity (X_1)	Tracking Range (X_2)	R	Y
0.5	2.5121	0.6311	2.5902	68.5098
1.0	1.7136	0.5568	1.8018	66.0725
1.5	1.2988	0.5088	1.3949	64.4898
2.0	1.0441	0.4728	1.1462	63.4004
2.5	0.8722	0.4439	0.9786	62.6069
3.0	0.7484	0.4194	0.8579	62.0010
3.5	0.6550	0.3983	0.7666	61.5212
4.0	0.5821	0.3796	0.6950	61.1300
5.0	0.4757	0.3478	0.5893	60.5256
7.0	0.3477	0.2992	0.4587	59.7229
10.0	0.2471	0.2483	0.3503	58.9965
20.0	0.1251	0.1594	0.2026	57.8859

The μ's in Table V are the preselected Lagrange multipliers and the X_1 and X_2 values were computed by equation (13) for each μ. The R's are the square roots of equation (9) for each X_1 and X_2 combination, and the Y's are the estimated responses obtained from equation (4).

This table gives a path for which increases in response should be obtained more quickly by moving along the path from well within the experimental region to the outer fringes of the experimental region. Confirmatory runs were made of the simulation with the first, second, and fourth points of Table V with the following results.

X_1	X_2	Y
2.512	0.6310	67.57
1.714	0.5570	64.85
1.044	0.4728	61.10

Each value in the Y column is the average of five replications. Notice that each response from the simulation is slightly lower than the predicted response in Table V. However, the differences are very small and goodly increases in response are obtained by moving along this path as predicted. It should be pointed out here that the response function becomes less reliable farther and farther away from the experimental region from which the function was based. If further analysis is desired outside the experimental region, a new response surface analysis should be performed.

The results of this analysis are also plotted in Figure 7.

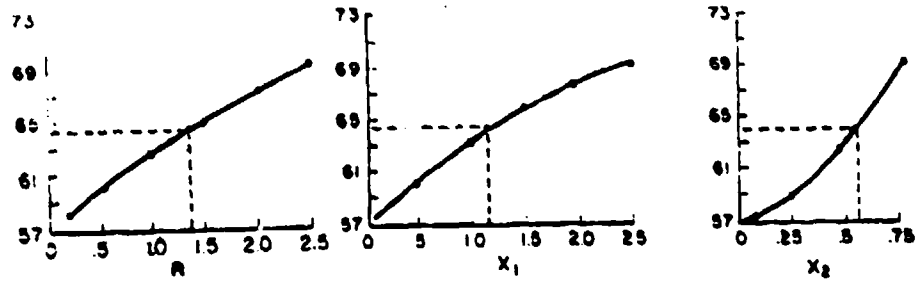


Figure 7. Two Dimensional Plots, Example 2.

From these plots the maximum response can be obtained constrained to the experimental region. Since $\hat{Y}$ continues to increase as R increases, the maximum response within the experimental region should correspond to the maximum R within the experimental region which is equal to 1.414. The estimated maximum response at this point is approximately 64.0 while $X_1 = 1.11$ and $X_2 = 0.55$ are the values of the independent variables (average missile velocity and radar tracking range) that yield this constrained maximum.

It was mentioned earlier that if the proper Lagrange multipliers μ were selected, local maximums could be found on each of the spheres. Meyers [1970] shows that if μ's are selected so that they are greater than the largest eigenvalue λ_1, local maximums will be obtained. On the other hand, if μ's are chosen so that they are less than the smallest λ_1, local minimums will be obtained.

In this paper, only two variable examples have been given for illustrative purposes due to ease of presentation. However, the methodology is very general and applies for any number of independent variables.

REFERENCES

Draper, N.R., "Ridge Analysis" of Response Surfaces, Technometrics 5, No. 4, p.469, 1963.

Hoerl, A.E., Optimum Solution of Many Variables Equations, Chemical Engineering Progress 55, No. 11, p.69, 1959.

Myers, R.H., Response Surface Analysis and Design, Allyn and Bacon, Inc., Boston, 1970.

CRITERIA FOR A BALLISTIC MODEL

George I. Lavin

Vulnerability Lab., Ballistic Research Lab., ARDC
Aberdeen Proving Ground, Md.

ABSTRACT - A survey is being made of work on the biochemistry and biophysics of animal systems in relation to the specific task performance of such systems.

Bio-chemically important systems which contain materials such as proteins, nucleic acids, polysaccharides, lipoids, complex mineral salts, etc., are thought to be responsible for particular animal functionality and behavior. This applies to muscle, nerve, circulatory systems, hard tissue, etc.

The incapacitation of animal systems (shock) can be accounted for by a modification of cellular systems (enzyme inactivation, cell wall destruction, etc.) or removal of tissue (wound tract).

Consideration is being given to the use of Biocellular Numbers, which characterise the absorption, equipartition and consequences of energy uptake by animal systems and which reflect the traumatic nature of the reactions in terms of cellular biochemistry....Also to protect against these effects.. It is possible that these numerical values can be used in the machine evaluation of the probability of occurence of certain stated events.

The design of the model should take into consideration the application of the material mentioned above - in dimensionality.

Multiple Comparisons Revisited

Clifford J. Maloney
Bethesda, Maryland

I. Introduction: In the years since World War II, a proliferation of approaches has arisen to the statistical analysis of data from simultaneously conducted trials of several materials or treatments (3). The existence of this cacophony of analyses strongly suggests that the true key to the situation has not yet appeared. In the hope of discovering that true key, a very detailed examination will be made of the applicable ideas in the following section.

II. Geometry of Multiple Comparisons: If but a single number is at hand the statistician has no contribution to make, though the applied and pure mathematician can point out respectively that (1) the number may represent a measurement of a particular property of a particular object or event on the one hand and (2) may be represented by a certain segment on the "number line" on the other.

If, next, two numbers are available, the possibilities are enormously enlarged. First, the two numbers may be equal or different. Next, whether equal or not they may correspond to (a) the same property of the same object or event, (b) the same property of different objects or events or (c) different properties of the same or different objects or events.

In case (a) the pure mathematician may elect to represent them as intervals on the number line which overlap or abut. In the latter case, the combined interval corresponds to the sum of the two numbers (Figure 1). This total interval corresponds to the sum of the two numbers irrespective of how the total is divided between them.

In either of cases (b) and (c) it would be normal to represent each number on a separate (not necessarily orthogonal) axis (Figure 2). Two quite different characteristics of that figure can be fixed for attention but either characteristic applies to each case. The sum of the squares of the two numbers gives the square of the diagonal of the rectangle of which the two numbers are the sides and their product gives its area. Here again an infinity of different pairs of numbers gives the same length of diagonal or alternatively the same area though at most two rectangles yield both a given length of diagonal and a given area.

By dividing the sum or the sum of squares by two or by taking the square root of the product we obtain the linear mean, the quadratic mean, or the geometric mean of the given two numbers respectively. In the special case where the two numbers are equal all these means are likewise equal. So far, this discussion has been restricted to purely mathematical aspects.

Now that we have two numbers to consider, the statistician has free scope for the exercise of his talents. Of course, the statistician's

approach will vary depending on whether these two numbers are taken to fall under case (a), (b) or (c). We have seen that, if the two numbers were obtained under the conditions of case (a), i.e., each is a measure of the same characteristic of a single (or at least similar) object or event, then the statistician has the responsibility of explaining why the two numbers differ (if they do). This he does by further separating each of the two numbers concerned into two parts. In the case of the linear or quadratic mean those two components are construed as entering additively to yield the observed measurement

$$x = m + e. \qquad (1)$$

It is assumed that, so far as the model of Figure 1 is concerned, e contributes no information and that the "true" lengths of the two segments are equal and are equal to m. The pure mathematician graphs the two m's contributed by both readings as intervals on a one-dimensional axis since each is an example of the same characteristic. But the statistician employs a different condition to justify plotting the two readings as abutting intervals on a single axis. It is that the two m's, being numerically equal, are perfectly correlated with each other. In contrast, it is assumed that the model in Figure 1 is <u>not</u> applicable to the two quantities e_i but that Figure 2 supplies the correct model. The e_i contributed by each x_i is plotted on a separate axis because these values are independent of (orthogonal to) each other. Nevertheless each e_i "estimates" a true

constant quantity, σ, which is best estimated from the quadratic mean, i.e., the diagonal of the rectangle in Figure 2. In the case of the geometric mean all rectangles are counted as equivalent which have the same area (not the same diagonal). If, in any particular case, e_1 and e_2 differ in magnitude, this variation provides no information about the magnitude of σ, whether estimated from the diagonal or from the area but σ is the same for both measurements. Were it possible to ascertain and plot σ for x_1 and x_2, Figure 2 would take the form of a square of side σ and hence the side estimated by the diagonal would exactly equal that estimated by the area and vice versa. Since e_1 and e_2 are not in general equal, these two estimates will not in general be either. But since, by hypothesis, the variation between e_1 and e_2 is due to chance, the difference in the estimate of σ from the diagonal and from the area is likewise a result of chance.

So far we have been considering just two measurements under case (a), but no change in concept is introduced when the number of measurements is increased to n. Any difference in magnitude of the several numbers is attributed to the influence of chance. But that chance itself is characterized by a number, the scale factor σ. This quantity is estimated by a mean just as the linear mean, m, is. But this mean is a quadratic mean, since its components are mutually independent, the same characteristic which calls for a geometric representation as a rectangle in Figure 2 or in general an n-dimensional rectangular parallelepiped.

Bartlett's test for homogeneity of variances exploits the contrast in the two approaches of the quadratic and geometric mean to estimates of the side of n-dimensional cubes. The length of side estimated from the observed diagonal length is compared with the side length estimated from the hyper-volume. Unless these two magnitudes are "acceptably" close the hypothesis of homogeneity, i.e., that the model of a hyper-cube applies, is rejected. It should be noted, however, that the presumption that a hyper-cubical model applies to each within class variance is not challenged.

But the situation changes radically under cases (b) and (c). Now, if our two or more numbers differ we attribute those differences to real differences in the characteristic or characteristics being measured and no chance effect is appealed to. Case (c) is beyond the scope of this paper. Case (b) involves a situation in which, if two measurements are at hand, both are appropriately graphed on the same axis, but any difference in their magnitude is attributed not to chance as in equation (1) but to a real difference in the two quantities of the characteristic being measured in the two instances. It is true that the square of their difference is numerically equal to and is even obtained by the same arithmetic as the estimate of σ^2 is obtained in case (a) which is treated so fully above. But the physical interpretation is entirely different. With two readings and with each reading interpreted as applying to a different object or event there is no scope for the action of probability and no chance for the statistician to exercise his talents.

The contrast between the statistician's approach to case (a) and case (b) is dramatically evident as soon as three measurements are at hand. For now, in addition to an estimate of σ being available under case (a) the numerical algorithm by which it is estimated is now different from the numerical process for estimating meaningful differences applicable to case (b). Further, in the model of the normal distribution for case (a) no further parameters are introduced by this third or any subsequent reading; whereas in case (b) every new reading introduces K-1 new parameters, the meaningful difference between its value and the K-1 preceding values. A problem in pure mathematics thus arises in this case that is meaningless in case (a). This is the problem of selecting a "basis" for the assemblage of $^1/_2$ K (K-1) differences between k measurements. A choice which will prove serviceable later is set out in Figure 3.

In the figure, three values are diagrammed which may be regarded either as individual measurements or as means of measurements which are not regarded as differing only by chance influences but as differing due to real influences of the conditions applying to each measurement. While there are three measurements or means, and hence three differences between elements of pairs of readings we may choose, as a basis for this set of differences, the two abutting segments labelled x_1 and x_2 in the figure. Of course, this model extends directly to k readings or means, yielding k-1 abutting intervals as a suitable basis. Clearly, in the case of k means, if any one of the x_i differs from zero, then all values of y to its left

differ from all values to its right. These differences, x_i, are of course not independent, so this choice is not an orthogonal decomposition of the k-1 degrees of freedom for treatments. Orthogonality is an attractive but by no means essential property of a basis for a set of vectors.

In short, the geometric representation of case (a) is an n-dimensional rectangular parallelepiped, whereas the appropriate geometric model for case (b) is a set of n points on the number line, each point falling at that value which corresponds to the Dedekind correlate of the number obtained by the measurement. Of course, if the graphing problem had originally arisen in an analytic context devoid of probability overtones this approach would be immediately adopted as it is every day in the mathematical classroom. Again, for k treatments, a basis for the $^1/_2$ K (K-1) differences which they yield can be found in the K-1 differences between adjacent measurements.

III. Least Squares or Analysis of Variance?: We use a linear mean to estimate the common component (m in equation 1) in a set of measurements, but the quadratic mean to estimate the scale factor of the random component (e in equation 1). This is not an arbitrary choice, but is forced upon us. The whole basis for statistics, for the application of probability in practical affairs, is that the contribution of chance factors tends to decrease in a mean. But a contribution is a "chance" contribution or a

real difference depending on one's point of view. The relation is not symmetrical. The quadratic mean has no way of discriminating between systematic and chance factors, or between chance components ascribable to varying conditions under which the several measurements are secured, but retains them all indiscriminantly. The influence of the constant factor, m, is retained in the quadratic mean, which latter yields an estimate of the scale factor only when that of the constant factor is subtracted out.

No problem of computation or difficulty in interpretation arises when but one mean and/or one variance have to be estimated. Of course, no matter how many means or parameters require estimation, if no variance estimate is involved, we have a straight-forward, if elaborate, problem of simultaneous equations (at most). If any number of parameters are to be estimated, just one variance, and, possibly, the standard error or confidence interval of one or more parameters or of linear combinations of them, the appropriate least squares procedures which were so highly developed in the nineteenth century apply.

The usual multiple comparisons situation fits this model directly. There is the task therefore of accounting for the entanglement of this problem with analysis of variance, if an explanation exists over and above the usual temptation to see every problem as an instance of the technique most prestigious at the moment. The discussion of the last section shows how this temptation is at least plausible. A few further remarks appear appropriate.

Given any set of numbers from whatever source, it is arithmetically possible to calculate their quadratic mean and to infer from it the numerical value of the scatter of the numbers. An appropriate mean value is a useful quantity and that utility exists irrespective of the significance or lack of significance of the individual values or of any other mean calculated from the numbers. It does not however command a preeminent position. Thus, only statisticians would insist on answering the question: "Is any current governor of a state of the Union a wealthy man?" by first forming the mean wealth of all governors and examining that value. The latter procedure would be appropriate however as an answer to the question whether possession of wealth is, on the average, associated with election as governor. That is, to examine a question about an average, a long run tendency, we need to form an average and examine it. To answer a question about an individual, even an individual mean, the average is not merely of no help, but is more detrimental the more values enter into that comprehensive mean. Now, the main effect for treatments in an analysis of variance is a quadratic mean. It answers the question: "On the average, do these treatments differ in effectiveness?" It is irrelevant to the question: "Do a specified pair of these treatments differ?" It is quite true that if a large number of tests are made some will be erroneous. That is what the selection of a significance level is all about. What is at issue is whether or not, and why, if so, the

combining of several tests into one experiment should influence that choice of level. If an experimenter carried out trials in treatment pairs, one trial at a time, presumably he would not be complimented for his efficiency but, he would be expected to use normal t-table significance levels. If he shared a laboratory with a colleague also performing comparative trials, one pair at a time, his choice of tables would not be influenced by the presence or any action of his colleague. The freedom from influence would exist for his colleague as well. Only, if one of the experimenters conducted all of the trials, and simultaneously at that, or if both continued to work but their results were pooled would the issue of altering the significance level arise; and then, in such a way, as to discourage experimenters from choosing such a combined experimental design, since now any observed differences are sharply penalized.

The confusion seems to stem from viewing the F-test as a required preliminary test which alone can justify introduction of a multiple comparisons analysis. But the F-test is a test of a mean. It is justified and valuable where a mean is called for; where a small but varying influence inheres separately in each member of the group, and where meaningful differences between them are non-existent or are to be ignored for the purposes of the particular analysis.

Since, in this case, the numerator of F is a quadratic mean--that is a variance-- this mode of viewing the individual treatment means as random

selections from a common pool for a null hypothesis was uncritically adopted. The consequence has been confusion, a plethora of rival techniques and an affront to common sense. It may be conjectured that the confusion would never have arisen if the converse of the principle of the mean were kept in view. The principle is that all of the information about the common element in a set of readings is given by the mean--and none by differences between the individual readings. The converse is that no information about intrinsic (non-chance) differences between the individual readings is given by the mean, which latter are the province of comparisons between partial means. Any of these may be real when the overall mean is not and vice versa.

IV. Least Squares Solution: The argument of Sections II and III purports to show that for the solution of what Eisenhart (2) has called Model I, the Analysis of Variance is a sterile cul-de-sac. This is clearly revealed by the confused status of the multiple comparisons problem. When, however, the problem is recognized as a straight-forward least squares task in terms of a natural choice of a basis set for all possible pairwise differences the difficulty disappears.

An example involving six treatment means is diagrammed in Figure 4. The observed means are denoted by y_i and the adjacent differences by x_i. The figure illustrates what may have been one source of the failure to recognize that the multiple comparisons problem is an instance of least squares in that, what is under study is not the observations directly but comparisons between them. To form observations equations in terms of

an orthogonal system of contrasts seems artificial since these are not the contrasts of inte[illegible] contrasts of interest are dependent and it is not usual to [illegible]dent observational equations. There would still be the problem of choosing which set to use. Figure 5 shows the equations which result if all possible (hence dependent) differences are used. There is one further difference between this set and the usual set of observational equations. Any k-1 set (in our case 5) yields x_i which satisfy <u>exactly</u> all other equations in Figure 5.

Our chief concern however lies not in the estimates of the differences but in tests of their significance and/or in confidence limits for them. For this purpose we may proceed with a conventional least squares solution having regard for the lack of independence in the equation of Figure 5. We obtain the normal equations of Figure 6. The solutions for the x_i are given in Figure 7. While these results are "obvious" our purpose is, by connecting the work with the general least squares approach, to establish that the procedure of that method <u>also</u> yields (1) the variance of every one of the $^1/_2$ K (K-1) difference, (2) tests of significance for each, and (3) confidence regions for each.

Formally, or directly by sight it is clear that the variance of each difference in Figure 7 is $2\sigma^2$. Any chosen difference can be formed as a linear combination of x_i's. Remembering that these are not independent the usual formula for the variance of a sum yields the same quantity, $2\sigma^2$, for <u>each</u> difference.

In using Figure 4 as a guide in our reasoning we have ignored the fact that the observed order of the means, y_i, may well differ from the order of the population values which they estimate. But whatever order is employed, the value $2\sigma^2$ is obtained as the variance of any difference. The true order must be included in the set of all possible orders. Hence, the result is established.

V. A Related Problem: The above treatment of the multiple comparisons problem historically grew out of a related but distinct problem (1). When an animal is exposed to a particular disease agent it will in general develop antibodies against a subsequent attack. These antibodies are highly but not completely specific so that while protection against other attacks of the identical organism are most completely prevented (at a fixed level of challenge strength) a greater or less level of protection is also afforded against closely related organisms. By sensitizing test animals with several related organisms and then challenging each animal with equivalent doses of each antigen complex a matrix of responses is obtained, some of which are homologous (same challenge as sensitizing organism) and some heterologous. In the paper referenced, the least square technique is applied to ascertain the relative strength of common antigens contained in the several challenge suspensions.

REFERENCES

1. Chaparas, Sotiros D., Maloney, Clifford J., and Hedrick, Sally R. (1970). Specificity of Tuberculins and Antigens from Various Species of Mycrobacteria, American Review of Respiratory Disease, Vol. 101, pp. 74-83.

2. Eisenhart, Churchill (1947). The Assumptions Underlying the Analysis of Variance, Biometrics, Vol. 3, No. 1, pp. 1-21.

3. Miller, Rupert G., Jr. (1966). Simultaneous Statistical Inference, McGraw-Hill Book Company.

Figure 1. Univariate representation of numbers as segments on a single axis. The sum of any number of segments is a representation of the sum of the numbers.

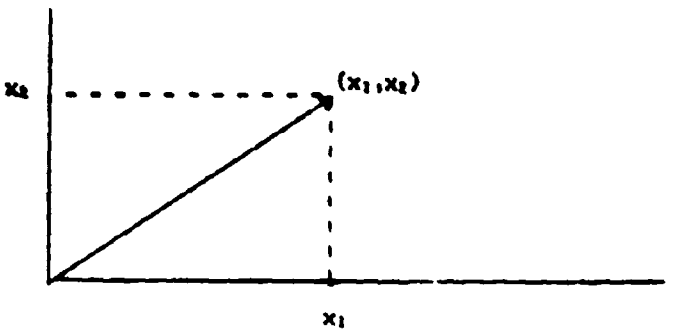


Figure 2. Orthogonal representation of two numbers as sides of a rectangle. The diagonal of the rectangle is their quadratic mean. The side of a square of equal area is their geometric mean.

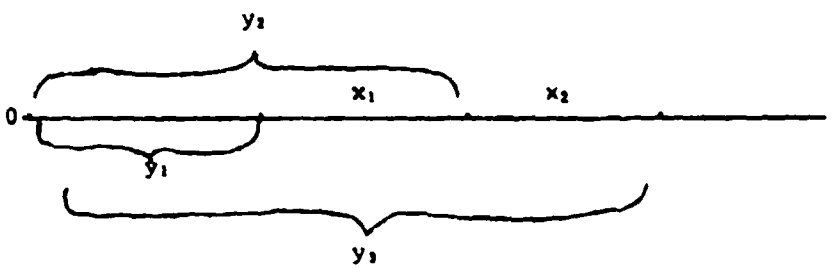


Figure 3. A linear plot of the magnitudes of three measurements (or means) of three different treatments, objects, or events. The two differences between adjacent end points are x_1, and x_2, the third is their sum.

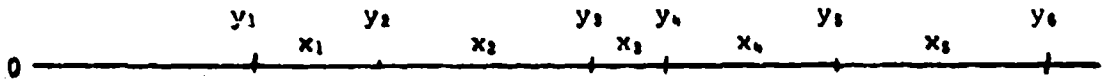


Figure 4. A linear plot of six treatment means, y_i. The resulting five adjacent differences x_i are shown.

$$
\begin{aligned}
x_1 + x_2 + x_3 + x_4 + x_5 &= y_6 - y_1 \\
x_2 + x_3 + x_4 + x_5 &= y_6 - y_2 \\
x_3 + x_4 + x_5 &= y_6 - y_3 \\
x_4 + x_5 &= y_6 - y_4 \\
x_5 &= y_6 - y_5 \\
x_1 + x_2 + x_3 + x_4 &= y_5 - y_1 \\
x_2 + x_3 + x_4 &= y_5 - y_2 \\
x_3 + x_4 &= y_5 - y_3 \\
x_4 &= y_5 - y_4 \\
x_1 + x_2 + x_3 &= y_4 - y_1 \\
x_2 + x_3 &= y_4 - y_2 \\
x_3 &= y_4 - y_3 \\
x_1 + x_2 &= y_3 - y_1 \\
x_2 &= y_3 - y_2 \\
x_1 &= y_2 - y_1
\end{aligned}
$$

Figure 5. Observational equations for the five parameter values x_i of Figure 4.

$$5x_1 + 4x_2 + 3x_3 + 2x_4 + x_5 = y_6 + y_5 + y_4 + y_3 + y_2 - 5y_1$$

$$4x_1 + 8x_2 + 6x_3 + 4x_4 + 2x_5 = 2(y_6 + y_5 + y_4 + y_3) - 4(y_1 + y_2)$$

$$3x_1 + 6x_2 + 9x_3 + 6x_4 + 3x_5 = 3(y_6 + y_5 + y_4) - 3(y_1 + y_2 + y_3)$$

$$2x_1 + 4x_2 + 6x_3 + 8x_4 + 4x_5 = 4(y_6 + y_5) - 2(y_1 + y_2 + y_3 + y_4)$$

$$x_1 + 2x_2 + 3x_3 + 4x_4 + 5x_5 = 5y_6 - (y_1 + y_2 + y_3 + y_4 + y_5)$$

Figure 6. Normal equations for the parameters x_i formed from the observational equations of Figure 5.

$$x_5 = y_6 - y_5$$

$$x_4 = y_5 - y_4$$

$$x_3 = y_4 - y_3$$

$$x_2 = y_3 - y_2$$

$$x_1 = y_2 - y_1$$

Figure 7. Solutions of equations (6) for parameters x_i.

AN OPTIMAL U_2 (MIN) DESIGN FOR ESTIMATING THE SLOPE OF A SECOND ORDER LINEAR MODEL

L. Ott and W. Mendenhall
Department of Statistics
University of Florida
Gainesville, Florida

The problem presented here concerns estimation of the slope of a second order linear model. Previous results are used to obtain the experimental design which minimizes a linear cost function subject to the restriction that the variance of the estimated slope is constant for a given setting of the independent variable.

1.1 Introduction

Experimenters in chemistry, biology, and engineering are frequently concerned with the examination of second order response models. The objective of such studies may involve estimation of the rates of change of the response for given values of the independent variable. For example, a chemist in the pharmaceutical industry might want to examine the rates of change in the response of rats to different doses of a drug product. Or, an engineer might be interested in studying velocities of a vehicle using a particular solid fuel propellant.

In this paper we will assume that the slope of the second order response curve is estimated using the method of least squares. We then consider the experimental design which minimizes a linear cost function subject to the restriction that the variance of the least square estimator of the slope at a given setting of the independent variable is equal to a fixed value, B.

1.2 Background

We will consider the following second order model relating a response, y, to an independent variable, x.

$$y = \alpha_0 + \alpha_1 x + \alpha_2 x^2 + \varepsilon \qquad (1)$$

A three-point experimental design is to be used with design points $x_1 < x_2 < x_3$. We will also allocate n_i observations to design point x_i $(i = 1,2,3)$ where

$$\sum_{i=1}^{3} n_i = n$$

It is convenient to code the independent variable, x, about the midpoint of the design region. Let

$$u = \frac{x - \theta}{r} \tag{2}$$

Hence the design points for the new independent variable, u, can be labeled -1, u_2, and 1 and equation (1) can be rewritten as

$$y = \beta_0 + \beta_1 u + \beta_2 u^2 + \epsilon \tag{3}$$

The estimator of the slope for a given value of the independent variable, u, is

$$\hat{s}|u = \hat{\beta}_1 + 2\hat{\beta}_2 u \tag{4}$$

and the variance of $\hat{s}|u$ is

$$V(\hat{s}|u) = V(\hat{\beta}_1) + 4u^2 V(\hat{\beta}_2) + 2u\ \mathrm{Cov}(\hat{\beta}_1, \hat{\beta}_2) \tag{5}$$

Equation (5) traces a parabola as a function of u and can be rewritten as

$$V(\hat{s}|u) = \sigma^2 C(u - h)^2 + \sigma^2 k \tag{6}$$

see (Ott and Mendenhall, 1970). The coordinates of the vertex of the variance parabola, (h, k), can be shown to be the expressions

$$h = \frac{-n_2(1 - u_2^2)(bu_2 - a)}{2W}, \tag{7}$$

$$k = \frac{b}{4n_1n_3} - \frac{n_2(bu_2 - a)^2}{4Wn_1n_3}, \tag{8}$$

with $$C = \frac{W}{n_1n_2n_3(1 - u_2^2)^2} \tag{9}$$

where $a = n_3 - n_1$, $b = n_3 + n_1$ and

$$W = n_2bu_2^2 - 2n_2au_2 + n_2b + b^2 - a^2$$

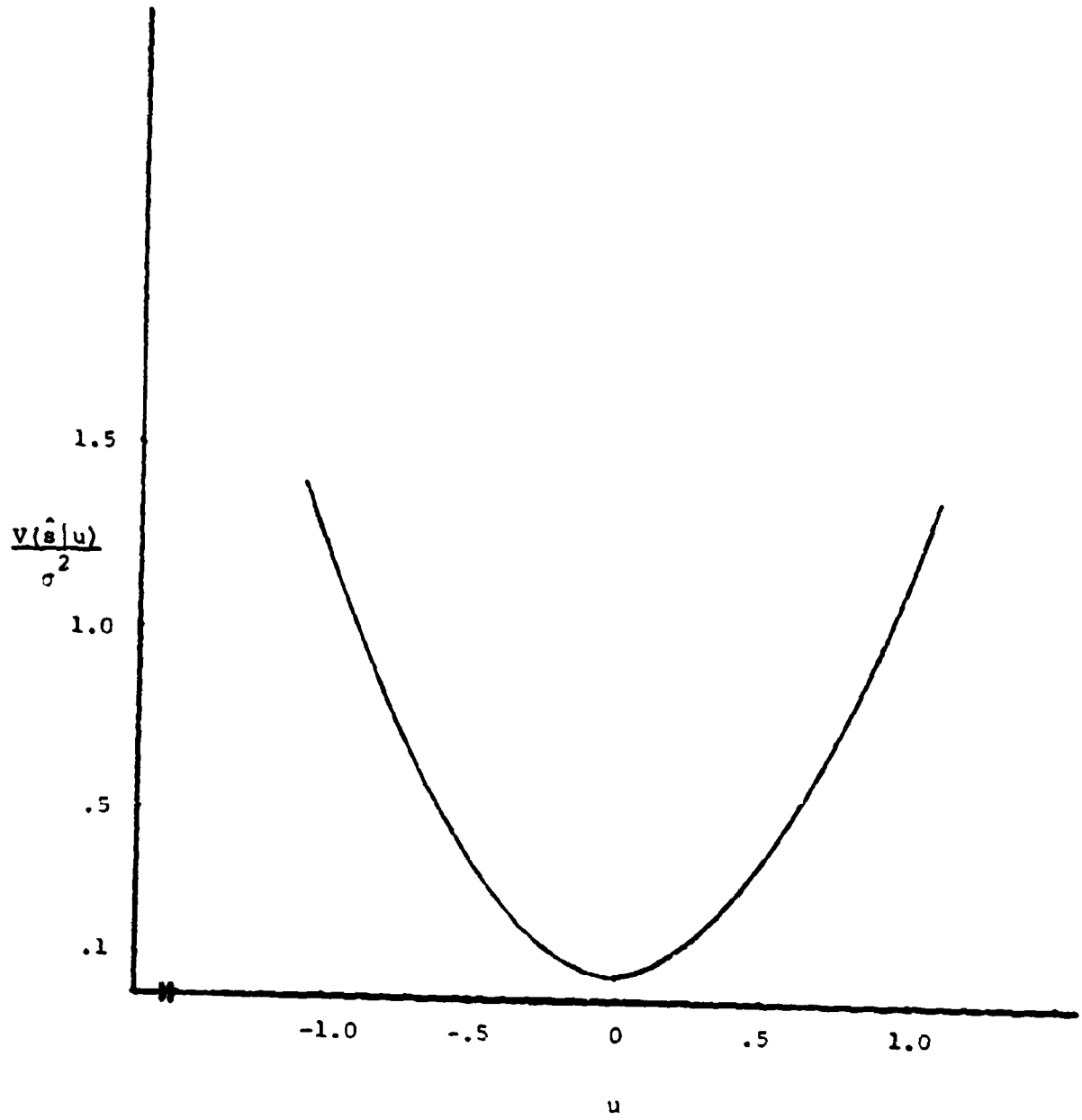
We have illustrated $\frac{V(\hat{s}|u)}{\sigma^2}$ for the design $u_2 = 0$, $n_1 = n_3 = 8$ and $n_2 = 4$ in Figure 1.

Examining equation (6) we see that the variance parabola, $V(\hat{s}|u)$ is determined by the location of its vertex, (h, k), and the rate of curvature, C. Changes in the quantity, σ^2, merely shift the entire parabola upward or downward.

From a practical standpoint, we would like the vertex of the variance parabola to be near the center of the experimental region. That is, we wish h to be near (or equal to) zero. We would also like k, the variance at the vertex of the parabola, to be small. Ott and

Figure 1

$\frac{V(\hat{s}|u)}{\sigma^2}$ For $u_2 = 0$, $n_1 = n_3 = 8$ and $n_2 = 4$



Mendenhall (1970) showed that the interior design point, $u_2(\min)$, which minimizes the variance at the vertex of the parabola is given by

$$u_2(\min) = \frac{a}{b} .$$

It can be shown that $h = 0$ for all $u_2(\min)$ designs (designs of the form $-1, \frac{a}{b}, 1$). Hence for a given sample size and allocation (n_1, n_2, and n_3) a $u_2(\min)$ design selects a value for the interior design point, u_2, which not only minimizes the variance of the vertex of the variance parabola but also locates the vertex, (h, k), in the center of the experimental region. The classification of $u_2(\min)$ designs can be refined even further. Ott and Mendenhall (1970) showed that a symmetrical $u_2(\min)$ design ($a = n_3 - n_1 = 0$) is to be preferred to a general $u_2(\min)$ design since both k and C increase as the design loses its symmetry. We shall restrict our attention to symmetrical $u_2(\min)$ designs in the remainder of this paper.

1.3 Minimizing the Cost of Experimentation

An optimal experimental design is one that purchases a specified quantity of information at minimum cost. We shall determine the symmetrical $u_2(\min)$ design which minimizes the cost of sampling subject to the restriction that $V(\hat{s}|x)$ equals a specified value, B. We will assume a cost function of the form

$$\text{Cost} = n_1c_1 + n_2c_2 + n_3c_3 \qquad (10)$$

where c_i is the cost of obtaining an observation at the i-th design point ($i = 1,2,3$). For symmetrical $u_2(\min)$ designs we rewrite equation (10) as

$$\text{Cost} = n_1d_1 + n_2d_2 \qquad (11)$$

with $d_1 = 2c_1$ and $d_2 = c_2$.

We define a function, Φ, which will be used to obtain the symmetrical $u_2(\min)$ design that minimizes equation (11) subject to the restriction that $V(\hat{s}|x) = B$.

$$\Phi = n_1 d_1 + n_2 d_2 + \lambda[V(\hat{s}|x) - B] \tag{12}$$

Substituting into equation (6) with $n_3 = n_1$ we find

$$V(\hat{s}|x) = \frac{\sigma^2 C}{r^2}(x - \theta)^2 + \sigma^2 k \tag{13}$$

where $C = \dfrac{2n}{n_1 n_2}$ and $k = \dfrac{1}{2n_1}$

Hence $$\Phi = n_1 d_1 + n_2 d_2 + \lambda\left[\frac{4\sigma^2(x-\theta)^2}{r^2 n_2} + \frac{2\sigma^2(x-\theta)^2}{r^2 n_1} + \frac{\sigma^2}{2n_1} - B\right] \tag{14}$$

We obtain a system of three equations in the three unknowns (n_1, n_2 and λ) by taking partial derivatives of Φ with respect to these unknowns.

$$\frac{\partial \Phi}{\partial n_1} = d_1 - \lambda\left[\frac{2\sigma^2(x-\theta)^2}{r^2 n_1^2} + \frac{\sigma^2}{2n_1^2}\right] = 0 \tag{15}$$

$$\frac{\partial \Phi}{\partial n_2} = d_2 - \lambda\left[\frac{4\sigma^2(x-\theta)^2}{r^2 n_2^2}\right] = 0 \tag{16}$$

$$\frac{\partial \Phi}{\partial \lambda} = \frac{4\sigma^2(x-\theta)^2}{r^2 n_2} + \frac{2\sigma^2(x-\theta)^2}{r^2 n_1} + \frac{\sigma^2}{2n_1} - B = 0 \tag{17}$$

Solving equations (15), (16) and (17), we obtain

$$n_1 = \frac{\sigma^2}{B}(2D + 1/2)\left[\frac{d_2}{d_1} + 1\right] \qquad (18)$$

and $$n_2 = 4n_1\left[\left(\frac{d_1}{d_2}\right)\frac{D}{(2D + 1/2)}\right] \quad \text{where } D = \frac{(x - \theta)^2}{r^2} \qquad (19)$$

It can be easily shown that the results presented in equations (18) and (19) provide the sample size $(2n_1 + n_2)$ and allocation of the sample size (n_1, n_2, n_1) for a symmetrical u_2(min) design which minimize the cost function

$$\text{Cost} = n_1 d_1 + n_2 d_2 \qquad (20)$$

subject to the restriction that the variance of the estimated slope at a point, x, equals B. Note the solution is location invariant. If $d_1 = d_2$, i.e., the costr are equal, the optimal solution is

$$n_1 = \frac{2\sigma^2}{B}(2D + 1/2) \quad \text{and} \quad n_2 = \frac{4n_1 D}{(2D + 1/2)} .$$

1.4 Example

Consider the following example. We wish to estimate the slope of a second order model where the range of tho experimental region is r = 1.0 and the variance of the response, y, is $\sigma^2 = 1.0$. We will assume that the experimenter is interested in estimating the slope at a distance of .5 from the midpoint of the experimental region

so that the variance of the estimated slope at this distance is $B = .1$. Further assume that it costs $d_1 = \$1.00$ to obtain an observation at either of the exterior design points and $d_2 = \$2.00$ at the interior point.

We use equation (21) with $(x - \theta)^2 = (.5)^2$ and $D = .25$ to obtain n_1.

$$n_1 = \frac{[2(.25) + 1/2]}{.1}\left[\left(\frac{2}{1}\right) + 1\right] = 30 \qquad (21)$$

Substituting into equation (19) we have

$$n_2 = 30\left\{4\left(\frac{1}{2}\right)\left[\frac{.25}{2(.25) + 1/2}\right]\right\} = 15 \qquad (22)$$

The symmetrical u_2(min) design (in units of u) is then $n_1 = 30$, $n_2 = 15$, and $n_3 = 30$ observations at the design points -1, 0, and 1 respectively. We can easily convert this to units of x using the relationship

$$u = \frac{x - \theta}{r}$$

Thus if the midpoint of the experimental region for this example is $\theta = 5$, we obtain the design points 4, 5, 6 with allocations 30, 15 and 30, respectively.

1.5 Summary

In this paper we have discussed the use of $u_2(\min)$ designs for estimating the slope of a second order linear model. We then determined the symmetrical $u_2(\min)$ design which minimizes a cost function of the form

$$\text{Cost} = n_1 d_1 + n_2 d_2$$

subject to the restriction that $V(\hat{s}|x) = B$.

REFERENCES

Chernoff, H. (1953). "Locally optimum designs for estimating parameters." Ann. Math. Statist., 24, 586-602.

David, H.A. and B.E. Arens (1959). "Optimum spacing in regression analysis." Ann. Math. Statist., 30, 1072-1081.

Folks, J.L. (1958). "Comparison of designs for exploration of response surface relationships." Ph.D. dissertation, Iowa State College.

Hoel, P.G. (1958). "Efficiency Problems in Polynomial Estimation." Ann. Math. Statist., 29, 1134-1145.

Kiefer, J. (1959). "Optimum experimental designs." J.R. Statist. Soc. B, 22, 272-304.

Kiefer, J. and J. Wolfowitz (1959). "Optimum designs in regression problems." Ann. Math. Statist., 30, No. 2, 271-294.

Ott, L. and W. Mendenhall (1970). "Designs for Estimating the Slope of a Second Order Linear Model" Technical Report No. 13, Department of Statistics, University of Florida. (to appear in Technometrics).

THE ANALYSIS OF COMPLEX CONTINGENCY TABLE DATA FROM GENERAL EXPERIMENTAL DESIGNS AND SAMPLE SURVEYS

Gary G. Koch and Donald W. Reinfurt
Department of Biostatistics
University of North Carolina at Chapel Hill
Chapel Hill, North Carolina

SUMMARY

Initially, a number of asymptotically equivalent approaches to the analysis of complex contingency table data are outlined and contrasted. These include maximum likelihood, minimum discrimination information, and the general linear model procedures. Various test criteria are presented along with a number of methods for estimation. Due to the generality of the linear model approach and its familiarity from the continuous case, it is the procedure chosen for the sequel.

The notation, assumptions, and methods of inference for the general linear model approach are detailed as in Grizzle, Starmer, and Koch [1969] and are illustrated for the familiar test of independence for two-way contingency tables.

A more general categorical data model is developed for the situation in which the data are incomplete in the sense that not all of the experimental units are classified according to each of the dimensions of the table. Since this information relative to a subset of the dimensions of the table is assumed to arise by design rather than by chance, it is referred to as "supplemental information" and the relevant margins of the table as "supplemented margins." This supplemental information is used to improve the precision of the estimates of the marginal probabilities over the supplemented dimension(s), and it is incorporated into the estimation of the individual cell probabilities in order to improve certain tests of hypotheses, especially those involving marginal symmetry. Note is made of the obvious resemblance to double sampling.

Special notation for the case of supplemental data is developed and a weighted least squares estimation procedure outlined. A general two-stage test procedure resembling that for the unsupplemented case is indicated.

Contingency table data often arise from sample surveys of finite populations. It is noted herein that, with certain modifications consistent with the principles of survey sampling (viz, modifications to the estimate of the covariance matrix), the more general linear model approach applies directly.

Finally, a number of detailed examples are presented. These examples illustrate the general linear model approach to categorical data arising in the following contexts:

(1) a simple quantal biological assay

(2) a factorial design-type contingency table including interaction (Dyke and Patterson, [1952])

(3) a split-plot contingency table on social mobility trends for different countries (Mosteller, [1968])

(4) a comparison of two drugs where there is supplemental information for both drugs

(5) a sample survey situation involving questionnaires sent to a sample of newly-licensed drivers in North Carolina

1. SOME HISTORICAL REMARKS

1.1 Introduction

Categorical data and its analysis have been of interest to statisticians ever since the earliest origins of the subject. For example, in the area of vital statistics, Graunt presented data during the early 1600's on causes of death (e.g., consumption, diseases of infancy, tooth diseases) which appeared in the form of frequency tabulations according to the sex and marital status of the deceased. Since 1790, census studies have collected and analyzed numerous cross-tabulations on a wide variety of demographic variables. Another example is in genetics where Mendel collected data in the mid-1800's on garden pea varieties (one-way tables) leading to the now-famous Mendelian heritability ratios. Finally, in the field of epidemiology, Greenwood and Yule in the early 1900's had considerable data on typhoid attacks on inoculated and non-inoculated individuals.

The remainder of this article has been reproduced photographically from the author's manuscript.

This research was partially supported by National Institutes of Health, Institute of General Medical Sciences, Grant No. GM 12868-06.

Aside from this historical background, categorical data appearing in the form of frequency tabulations in one-way, two-way, and multi-way tables are a familiar data array to virtually all statisticians in today's society. The specific examples range from experiments in medicine dealing with all or nothing responses to various stimuli or methods of therapy, to investigations of highway safety data indicative of degree of personal injury severity in motor vehicle accidents, and to sample surveys dealing with various aspects of life including even such personal matters as family planning. In most of these situations which cut across the social sciences, the life sciences, and even the physical sciences, the data are expressed in terms of nominal or ordinal scales. Moreover, even quantitative data are occasionally represented in a categorical form when grouped into class intervals.

Although many statisticians are aware of the various interrelationships that may exist among the classification type variables in a multi-dimensional contingency table, such data are often analyzed in terms of the traditional Pearson chi-square statistic as applied to some subset of the totality of two-dimensional tables. Since these statistics are directed in many cases at fragmented hypotheses, a definite need becomes apparent for comprehensive methods of statistical analysis which are directly applicable to qualitative data and which are analogous in scope and power to multiple regression and multivariate analysis of variance as applied to quantitative data. Unfortunately, because categorical data has been of interest to a wide variety of researchers with divergent areas of application, only a limited amount of attention has been directed at general models and methods of inference in the published literature. Instead, more consideration has been given to the development of miscellaneous techniques which are discussed from somewhat different philosophical points of

view and which are applicable to a series of not always well-defined special cases. Thus, the practitioner has often had to search the literature for a procedure appropriate to his problem; when he has failed to find such a technique, he has either been required to use an inappropriate analysis for his data or leave his data unanalyzed.

Another difficulty in the analysis of categorical data has been the existence of numerous methods of estimation and corresponding test criteria upon which statistical inferences may be based. As a result, even when the researcher has decided upon the hypotheses of interest, he still is confronted with a wide choice of particular methods to apply. In most cases, this choice is not crucial since many of the methods have been shown to be asymptotically equivalent. However, since various authors use different methods, the variety of choices is often confusing to the uninitiated.

1.2. Some Recent Developments in General Methods of Analysis for Categorical Data

Pearson [1947] was among the first to note that, even in the simple 2×2 contingency table, the same configuration could have arisen from different sampling schemes. Thus, it appeared necessary to carefully specify the underlying probability model since different models could lead to different statistical procedures with obviously different interpretations.

The importance of model construction and hypothesis formulation in complex multi-dimensional contingency tables has been discussed in some detail by Bhapkar and Koch [1968a, 1968b]. Their work emphasized the relationship between certain problems in categorical data situations

with corresponding problems in multivariate analysis pertaining to patterns of correlations and factorial analysis of effects. As such, it was a direct outgrowth of earlier work pioneered by Roy and Mitra [1956], Roy and Kastenbaum [1956], Roy and Bhapkar [1960], and Bhapkar [1961, 1966, 1968]. Underscoring the results of this research has been the careful consideration of the sampling scheme from which the data arose. The particular factor-response structure (i.e., sampling scheme) of the model has dictated the hypotheses of relevance just as in univariate and multivariate analysis of variance for continuous variables.

The application of this approach to data from a variety of experimental situations through the use of linear regression models has been described by Grizzle, Starmer, and Koch [1969], Grizzle and Williams [1970], Johnson and Koch [1970a], Koch and Reinfurt [1970], and Reinfurt [1970]. These authors base their analyses on the use of a fairly powerful and widely applicable computer program which has been documented by Forthofer, Grizzle, and Starmer [1969]. The resulting test statistics are derived through weighted least squares procedures and correspond identically to the modified minimum chi-square statistics (χ^2_1) due to Neyman [1949] or equivalently the generalized quadratic form criteria due to Wald [1943].

Another line of development has been pursued by researchers who are particularly interested in multiplicative models for multinomial populations. Here a maximum likelihood approach based on iterative proportional fitting of appropriate marginal sub-tables has been emphasized. A comprehensive discussion of this approach has been given by Goodman [1969, 1970]. These papers bring together a number of results obtained

previously by Birch [1963], Goodman [1963a, 1963b, 1964a, 1964b], Mosteller [1968], Bishop [1969], Fienberg [1969] and Plackett [1969]. A substantial part of this effort has been motivated by the resemblance of multi-dimensional contingency tables to certain complex factorial designs. Maximum likelihood methods are then applied to the assumed multiplicative underlying models. Finally, Bishop and Fienberg [1969] and Goodman [1968] have demonstrated that this approach is a particularly useful one for handling incomplete contingency tables (i.e., tables containing cells with zero frequencies).

Finally, a third general approach which bears a definite resemblance to the maximum likelihood methodology previously discussed is based on the principle of minimum discrimination information. Certain aspects of this type of analysis are summarized in Ku and Kullback [1968]. Some other relevant results are given in Kullback [1959], Kullback, Kupperman, and Ku [1962], Good [1963], Ireland and Kullback [1968a, 1968b], and Ireland, Ku, and Kullback [1969].

For the most part, all of these approaches are asymptotically equivalent in the sense of being based on BAN estimates (see Section 2.3) as described by Neyman [1949]. Certain results of Hoeffding [1965] and Rao [1963] suggest that the maximum likelihood approaches are more efficient. However, the authors here believe the minimum χ^2_1-approach to be computationally simpler and to be more robust against departures from underlying model assumptions. Hence, it will be the one which is emphasized in the remainder of this paper. Finally, it should be noted that the relative merits of the various approaches have been discussed a number of times by Berkson [1955, 1957, 1968] in the context of data from the health and medical sciences.

This discussion has focused attention primarily on the three general methods of analysis. Other important contributions which have cut across these lines of research and supported their development include the work of Bartlett [1935], Simpson [1951], Lancaster [1951], Plackett [1962], Lewis [1962], and Darroch [1962] on contingency table interaction; Cochran [1952] and Maxwell [1961] on methods which pertain to the refining and strengthening of the applicability of the traditional Pearson type χ^2-test; and Berkson [1944, 1953, 1954, 1968] and Grizzle [1961] on applications to data arising from bioassay models.

2. MODELS FOR CATEGORICAL DATA

2.1. The Characterization of Multi-dimensional Contingency Tables

In the analysis of any multi-dimensional contingency table, there are three fundamental aspects to be considered. These include

(i) the specification of the underlying model;

(ii) the formulation (in terms of this model) of the hypothesis to be tested and the calculation of the corresponding test statistic;

(iii) the interpretation (with respect to the model) of the results.

In order to specify the underlying model, it must be realized that each subject (or experimental unit) may give rise to two types of data. These are

(i) a description of the experimental conditions which the subject undergoes (or of the sub-population of units to which the experimental unit belongs)--henceforth referred to as "factors" or "populations";

(ii) a description of what subsequently happens to each subject--henceforth referred to as "responses".

In other words, the data in any multi-dimensional contingency table are

the frequencies with which subjects belonging to the same sub-population or combination of factor categories yielded the same combination of responses. The dimension, d, of a contingency table refers to the total number of factors and responses while the levels represent the sub-categories within the factors and responses.

More specifically, consider r independent random samples taken from r multinomial populations where n_{io} represents the size of the sample from the i-th population and n_{ij} the observed number falling in the j-th category of the i-th sample where $i = 1, 2, \ldots, r$ and $j = 1, 2, \ldots, s$. It should be noted that i and/or j may be multiple subscripts as in the case of multi-factor and/or multi-response models. For example,

$\underset{\sim}{i} = \underset{\sim}{i}' = (i_1, i_2, \ldots, i_{d_f})$ with $i_\alpha = 1, 2, \ldots, r_\alpha$, $\alpha = 1, 2, \ldots, d_f$

$\underset{\sim}{j} = \underset{\sim}{j}' = (j_1, j_2, \ldots, j_{d_r})$ with $j_\beta = 1, 2, \ldots, s_\beta$, $\beta = 1, 2, \ldots, d_r$.

In this context, the model is that of a d_f-factor, d_r-response contingency table with the factors at levels i_α, $\alpha = 1, 2, \ldots, d_f$, and the responses at levels j_β, $\beta = 1, 2, \ldots, d_r$. Here, it will be assumed that all response-level combinations occur with positive probability; however, situations where this assumption does not hold may be handled by methods analogous to those given by Goodman [1968]. On the other hand, not all factor-level combinations are required to appear. In particular, if the sample design is an incomplete block design, not all factor-level combinations appear.

From these considerations, it follows that four types of multi-dimensional contingency table models can be identified. These are as follows:

Model I. No factor, multi-response tables

Model II. Uni-factor, multi-response tables

Model III. Multi-factor, uni-response tables

Model IV. Multi-factor, multi-response tables

Obviously, for one-way tables, only Model I can occur; for two-way tables, only Models I and II can occur;for three-way tables, only Models I, II, and III can occur. Otherwise ($d > 3$), all four models can arise in various situations.

To identify the hypotheses appropriate to the different models, consider the case of three-dimensional tables. In the case of "no factor, three response" tables (Model I), the primary interest lies in the relationships among the three responses which are analogous to problems of independence and correlation in normal multivariate analysis of variance. Questions of interest include total independence of the three responses, independence of any one response from the other two, pairwise independence, and partial association between two responses within given levels of the third. In addition, hypotheses of total symmetry and marginal symmetry are often of considerable interest.

If the experiment is of the "one factor, two response" type (Model II), interest lies in the association between the responses as well as the effect of the factor on the responses. Here the questions posed are similar to those encountered in one-way normal multivariate analysis of variance. Appropriate hypotheses for this case include independence of the responses within each factor level, homogeneity for each marginal response (i.e., does the factor level affect the marginal distribution of the response?), and joint homogeneity (i.e., does the

factor level affect the joint distribution of the responses?).

Finally, if the experiment is of the "two factor, one response" type (Model III), the questions of interest are similar to those arising in univariate normal analysis of variance, i.e., how do the factors (cf. treatments or independent variables) combine to determine the response (cf. "yield" or dependent variable)? Here the hypotheses of interest include total homogeneity (i.e., do both factors affect the distribution of the response?), partial homogeneity (i.e., does one factor affect the distribution of the response?), and "no interaction" between factors in the way they affect the response (i.e., do the factors determine the response in a purely additive or multiplicative way or are the relationships between the factors and the response more complex?).

For completeness, it should be noted that for Model IV situations ("multi-factor, multi-response" experiments and hence $d > 3$), interest centers in both the relationships among the responses and in the way the factors combine to affect the responses. Thus, the structure of the table (which is not always obvious or unambiguous) dictates the types of questions that are of interest in a given experimental situation. In the next section, these concepts will be formulated in terms of a general mathematical model from which test statistics may be derived. It is at this point that various differences arise among the three general methodologies based on minimum χ^2, maximum likelihood, and minimum discrimination information.

2.2. Hypothesis Formulation

Let π_{ij} represent the probability of an observation from the i-th population (or factor level) falling in the j-th response category (i.e.,

the π_{ij} represent individual cell probabilities). If the sampling is either from a very large population or with replacement, then the probability distribution of the observed frequencies, n_{ij}, is given by the product-multinomial distribution. This model has the form

$$\phi = \prod_{i=1}^{r} n_{io}! \prod_{j=1}^{s} \frac{(\pi_{ij})^{n_{ij}}}{n_{ij}!} \qquad (2.2.1.)$$

where $\sum_{j=1}^{s} \pi_{ij} = 1$ and $\sum_{j=1}^{s} n_{ij} = n_{io}$ (fixed) for all i and where it is assumed tnat $\pi_{ij} > 0$ for all i, j. This basic model allows various hypotheses of interest to be expressed as functions of the unknown parameters in (2.2.1). Hence, it is the primary basis of inference in the remainder of this paper. However, it should be noted that, if sampling is without replacement from a relatively small population and hence the sampling fraction is large (say, exceeding 10 per cent), methods of analysis based on the theory of sampling from finite populations are more appropriate. The details of the analysis for this situation which requires adjustments for finite population correction factors have been considered by Johnson and Koch [1970b] and are illustrated in Section 5.5.

A general formulation for many hypotheses of interest in the analysis of categorical data (e.g., in two-way tables, independence in the Model I situation, homogeneity in the case of Model II) is given by the following set of constraints on the cell probabilities:

$$H_{01}: \underset{t\times 1}{f}(\underset{\sim}{\pi}) = \underset{\sim}{0} \qquad (2.2.2)$$

where

$$(\underset{\sim}{f}(\underset{\sim}{\pi}))' = (f_1(\underset{\sim}{\pi}), f_2(\underset{\sim}{\pi}), \ldots, f_t(\underset{\sim}{\pi})) \quad t \leq r(s-1) \qquad (2.2.3)$$

with

$$\underset{\sim}{\pi}' = (\pi_{11},\pi_{12},\ldots,\pi_{1s}; \pi_{21},\pi_{22},\ldots,\pi_{2s}; \ldots; \pi_{r1},\pi_{r2},\ldots,\pi_{rs})$$
$$= (\underset{\sim}{\pi}_1',\underset{\sim}{\pi}_2',\ldots,\underset{\sim}{\pi}_r')$$

Here, the functions, $f_k(\underset{\sim}{\pi})$, are assumed to be functionally independent of each other and independent of the constraints, $\sum_{j=1}^{s} \pi_{ij} = 1$, $i = 1, 2,\ldots,r$. In addition, the $f_k(\underset{\sim}{\pi})$ must have continuous first and second partial derivatives with respect to the π_{ij}'s and $\underset{t\times rs}{H}(\underset{\sim}{u}) = (\partial f_k(\underset{\sim}{\pi})/\partial \pi_{ij})_{\underset{\sim}{\pi}=\underset{\sim}{u}}$ must be of rank t (i.e., full rank) for any $\underset{\sim}{u}$ in the neighborhood of $\underset{\sim}{\pi}$.

For example, consider the hypothesis of independence in a two-way contingency table of the "no factor, two response" type (i.e., Model I). This hypothesis is usually formulated as

$$\pi_{j_1 j_2} = \pi_{j_1 o}\,\pi_{o j_2}$$

where

$$\pi_{j_1 o} = \sum_{j_2=1}^{s_2} \pi_{j_1 j_2}$$

and

$$\pi_{o j_2} = \sum_{j_1=1}^{s_1} \pi_{j_1 j_2}$$

represent the appropriate marginal probabilities. Alternatively, this hypothesis may be written in the form specified by (2.2.2) as

$$\log_e \left\{ \frac{\pi_{j_1 j_2}\,\pi_{j_1' j_2'}}{\pi_{j_1 j_2'}\,\pi_{j_1' j_2}} \right\} = 0 \qquad j_1 \neq j_1',\ j_2 \neq j_2'.$$

It should be noted that since this situation involves a single multinomial population, the i-subscript has been suppressed. Alternatively, if the two-way contingency table is of the "one factor, one response" type (i.e., Model II), then the hypothesis of homogeneity across subpopulations may be written in the form specified by (2.2.2) as

$$\pi_{ij} - \pi_{i'j} = 0 \qquad \text{for all } i \neq i' \text{ and each } j.$$

In other cases, the $f_k(\underset{\sim}{\pi})$ pertain to any type of function about which the researcher desires to draw inferences.

More generally, hypotheses involving the functions $\underset{\sim}{f}(\underset{\sim}{\pi})$ can be formulated in terms of linear regression models

$$H_{02}: \underset{t\times 1}{\underset{\sim}{f}}(\underset{\sim}{\pi}) = \underset{t\times u}{\underset{\sim}{X}}\ \underset{u\times 1}{\underset{\sim}{\beta}} \qquad (2.2.5)$$

where $\underset{\sim}{f}(\underset{\sim}{\pi})$ is as before, $\underset{\sim}{X}$ is a known design matrix of rank $u \leq t$ and $\underset{\sim}{\beta}$ is a vector of unknown parameters. Given that a linear model for the functions $\underset{\sim}{f}(\underset{\sim}{\pi})$ applies, hypotheses concerning various constraints on the model parameters can be formulated as follows:

$$H_{03}: \underset{c\times u}{\underset{\sim}{C}}\ \underset{u\times 1}{\underset{\sim}{\beta}} = \underset{c\times 1}{0} \qquad (2.2.6)$$

where $\underset{\sim}{C}$ is a matrix of known constants of rank $c \leq u$. Proper choices of $\underset{\sim}{C}$ allow hypotheses pertaining to various main effects and lower order interactions to be examined. This aspect of analysis is illustrated by several examples in Section 5.

2.3 Estimation Procedures

If the hypothesis of interest is formulated in terms of constraints on the cell probabilities as in (2.2.2), then it is necessary to obtain estimates of the cell probabilities subject to those constraints. These constrained estimates are then incorporated into one of several asymptotically equivalent test criteria and the test is performed.

In the framework of (2.2.2), there are essentially two different classes of estimation procedures. These are

(1) Maximum likelihood and minimum discrimination information;

(2) Minimum chi-square (minimum χ^2) and modified minimum chi-square (minimum χ_1^2) in the sense of Neyman.

Both classes of estimation procedures satisfy a certain optimality property; namely, they all yield BAN (best asymptotically normal) estimates. Thus, these estimates, $\hat{\pi}_{ij}$, are

(i) Consistent and hence asymptotically unbiased (i.e., they converge in probability to the π_{ij});

(ii) Asymptotically normal as $N = \sum_{i=1}^{r} n_{io} \to \infty$ with $\frac{n_{io}}{N} \to \xi_i$ where ξ_i is a constant;

(iii) Efficient (i.e., the variance of any other consistent, asymptotically normal estimate is at least as large as the variance of $\hat{\pi}_{ij}$);

(iv) Sufficiently regular (i.e., $\frac{\partial \hat{\pi}_{ij}}{\partial \pi_{ij}}$ exist and are continuous in π_{ij} for all i,j).

Historically, maximum likelihood estimates (MLE's) were probably the first estimates derived for (2.2.2) and have been fairly popular in certain special contexts like quantal response bioassays. To obtain MLE's of the π_{ij}, (2.2.1) must be maximized with respect to the π_{ij}'s and subject to (2.2.2). This leads to a system of simultaneous equations which are often non-linear in the π_{ij} and consequently difficult to solve. Although various iterative schemes have been proposed (e.g., Roy and Kastenbaum, [1956]; Kastenbaum and Lamphiear, [1959]), the variety of possible hypotheses has precluded the formulation of a general solution of the resulting systems of equations. However, due to the results of Birch [1963], fairly general maximum likelihood procedures have recently appeared. These include the various iterative proportional fitting schemes presented in Mosteller [1968], Goodman [1969, 1970], Bishop and Fienberg [1969], and Bishop [1969]. These are based on the realization that for certain

hypotheses, the MLE's of the expected cell frequencies are uniquely determined by the marginal totals being equal to the MLE's of their expectations.

Alternatively, Kullback and his associates have recommended the use of minimum discrimination information estimates. These BAN estimates are obtained by minimizing, with respect to the π_{ij}'s,

$$I(\underset{\sim}{\pi};\underset{\sim}{p}^*) = \sum_{i=1}^{r}\sum_{j=1}^{s} \pi_{ij} \ln\left(\frac{\pi_{ij}}{p^*_{ij}}\right) \qquad (2.3.1)$$

subject to (2.2.2) where the p^*_{ij} are fixed by hypothesis, observed, or estimated. Similar to maximum likelihood estimation, this procedure also usually requires iterative procedures.

Minimum χ^2 estimates are obtained by minimizing

$$\sum_{i=1}^{r}\sum_{j=1}^{s} \frac{(n_{ij} - n_{io}\pi_{ij})^2}{n_{io}\pi_{ij}} \qquad (2.3.2)$$

with respect to the π_{ij}'s and subject to (2.2.2). Again, this technique usually involves the solution of a system of complicated, non-linear simultaneous equations, and, since minimum χ^2 estimates are not known to possess any desirable properties that MLE's lack, these estimates have seldomly been used.

If the denominator of (2.3.2) is replaced by the observed cell frequency, n_{ij}, and the resulting sum

$$\sum_{i=1}^{r}\sum_{j=1}^{s} \frac{(n_{ij} - n_{io}\pi_{ij})^2}{n_{ij}} \qquad (2.3.3)$$

minimized with respect to the π_{ij}'s and subject to (2.2.2), the resulting minimum χ^2_1 estimates are obtained by solving only linear

equations provided the $f_k(\underset{\sim}{\pi})$ are linear in the π_{ij}'s. Neyman [1949] proved that, if the $f_k(\underset{\sim}{\pi})$ are not linear in the π_{ij}'s, the $f_k(\underset{\sim}{\pi})$ may be replaced by their Taylor series expansion about the point, $\underset{\sim}{\pi} = \underset{\sim}{p}$, where $p_{ij} = n_{ij}/n_{io}$, $i = 1, \ldots, r$; $j = 1, \ldots, s$, namely

$$f_k^*(\underset{\sim}{\pi}) = f_k(\underset{\sim}{p}) + \sum_{i=1}^{r}\sum_{j=1}^{s} \left(\frac{\partial f_k(\underset{\sim}{\pi})}{\partial \pi_{ij}}\right)_{\underset{\sim}{\pi}=\underset{\sim}{p}} \cdot (\pi_{ij} - p_{ij}) \qquad (2.3.4)$$

$$k = 1, 2, \ldots, t$$

thereby reducing the problem to the linear case. In either case, the resulting minimum χ_1^2 estimates are BAN estimates and do not require iterative procedures for their solution.

In the framework of (2.2.5) where a linear model is fitted to the data, Grizzle et al. [1969] estimate $\underset{\sim}{f}(\underset{\sim}{\pi})$ by replacing $\underset{\sim}{\pi}$ by its unrestricted MLE, $\underset{\sim}{p}$, and then estimate the parameters $(\underset{\sim}{\beta})$ of the model by ordinary weighted least squares procedures applied to

$$\underset{t\times 1}{\underset{\sim}{f}}(\underset{\sim}{p}) = \underset{t\times u}{X}\ \underset{u\times 1}{\underset{\sim}{\beta}}. \qquad (2.3.5)$$

This procedure yields the familiar weighted least squares estimate (WLSE), $\underset{\sim}{b}$, of $\underset{\sim}{\beta}$, namely

$$\underset{u\times 1}{\underset{\sim}{b}} = (X'\underset{\sim}{S}^{-1}X)^{-1}\, X'\underset{\sim}{S}^{-1}\underset{\sim}{f}(\underset{\sim}{p}) \qquad (2.3.6)$$

where

$\underset{txt}{\underset{\sim}{S}}$ = a consistent estimate of the covariance matrix for $\underset{\sim}{f}(\underset{\sim}{p})$

This estimation procedure is discussed in more detail in Section 3.

2.4 Test Statistics

A number of asymptotically equivalent test statistics have been proposed for testing the hypotheses specified by (2.2.2) and (2.2.5). Asymptotic equivalence requires that, regardless of whether the null hypothesis holds, the probability of any two tests being contradictory tends to zero as $N = \sum_{i=1}^{r} n_{io} \to \infty$ with $\frac{n_{io}}{N} = \xi_i$ where ξ_i is a constant. The test statistics for (2.2.2) fall into three essentially different classes. They are

(1) Pearson's chi-square statistic

$$X_P^2 = \sum_{i=1}^{r} \sum_{j=1}^{s} \frac{(n_{ij} - n_{io} \hat{\pi}_{ij})^2}{n_{io} \hat{\pi}_{ij}} \tag{2.4.1}$$

(2) (a) Neyman-Pearson's likelihood ratio chi-square statistic

$$X_L^2 = \sum_{i=1}^{r} \sum_{j=1}^{s} [-2 \ln (\frac{n_{io} \hat{\pi}_{ij}}{n_{ij}})^{n_{ij}}]. \tag{2.4.2}$$

(b) Minimum discrimination information statistic

$$X_I^2 = \sum_{i=1}^{r} \sum_{j=1}^{s} [2N \hat{\pi}_{ij} \ln(\frac{\hat{\pi}_{ij}}{p_{ij}^*})]. \tag{2.4.3}$$

(3) (a) Neyman's chi-square statistic

$$X_N^2 = \sum_{i=1}^{r} \sum_{j=1}^{s} \frac{(n_{ij} - n_{io} \hat{\pi}_{ij})^2}{n_{ij}} \tag{2.4.4}$$

(b) Wald's statistic

$$X_W^2 = (\underset{\sim}{t}(\underset{\sim}{p}))' \underset{\sim}{S}^{-1} (\underset{\sim}{f}(\underset{\sim}{p})) \tag{2.4.5}$$

where $\underset{\sim}{p}$, $\underset{\sim}{f}(\underset{\sim}{p})$, $\underset{\sim}{S}$, p_t and N are defined in Section 2.3, the $\hat{\pi}_{ij}$ are any BAN estimates of the π_{ij}, and the $\hat{\hat{\pi}}_{ij}$ are the minimum discrimination information estimates obtained by minimizing (2.3.1).

For testing the fit of the linear model (2.2.5), the usual analysis of variance error sum of squares is used, i.e.,

$$X_f^2 = SS(\underset{\sim}{f}(\underset{\sim}{\pi}) = \underset{\sim\sim}{X}\underset{\sim}{\beta}) = (\underset{\sim}{f}(\underset{\sim}{p}))' \underset{\sim}{S}^{-1}(\underset{\sim}{f}(\underset{\sim}{p})) - (\underset{\sim\sim}{X}\underset{\sim}{b})' \underset{\sim}{S}^{-1} (\underset{\sim\sim}{X}\underset{\sim}{b}) \tag{2.4.6}$$

where $\underset{\sim}{b}$ is defined in Section 2.3. Then, given that the model adequately fits the data, tests involving contrasts of the model parameters ($\underset{c\times u}{C}\ \underset{u\times 1}{\beta} = \underset{c\times 1}{0}$) are produced by usual analysis of variance sums of squares, i.e.,

$$X_C^2 = SS(\underset{\sim\sim}{C}\underset{\sim}{\beta} = \underset{\sim}{0}) = (\underset{\sim\sim}{C}\underset{\sim}{b})'[\underset{\sim}{C}(\underset{\sim}{X}'\underset{\sim}{S}^{-1}\underset{\sim}{X})^{-1}\underset{\sim}{C}']^{-1}(\underset{\sim\sim}{C}\underset{\sim}{b}) \tag{2.4.7}$$

where $\underset{\sim}{C}$ is a matrix of constants of rank $c \leq u$.

Under H_{01} in (2.2.2), Neyman [1949] has shown that X_P^2, X_{LR}^2 and X_N^2, using any BAN estimates of the π_{ij}, are asymptotically χ^2 variates with t degrees of freedom (D.F.) and hence are asymptotically equivalent provided that

(i) $\frac{n_{io}}{N}$, i = 1, 2, ..., r, remain constant as $N \to \infty$,

(ii) $f_k(\underset{\sim}{\pi}) = 0$, k = 1, 2, ..., t, has at least one solution such that $\pi_{ij} > 0$ for all i, j.

Bhapkar [1966] has shown that X_W^2 (Wald's statistic) for testing linear hypotheses in categorical data is algebraically identical to X_N^2 whenever X_N^2 is defined, i.e., whenever all the n_{ij} are positive. Similarly, for

testing non-linear hypotheses, X^2_W is identical to X^2_N using Neyman's linearization technique on the hypothesis constraints as in (2.3.4). Kullback [1959] has shown that X^2_I (the minimum discrimination information statistic), under H_{01}, is also asymptotically χ^2 with D.F. = t so that all of these tests used for H_{01} are asymptotically equivalent when using the appropriate estimates of the individual cell probabilities.

In the context of the linear model as in (2.2.5), the tests are derived by conventional methods of weighted least squares and hence are the same as Neyman's chi-square tests when translated into constraints. Under H_{02}, X^2_F for testing the fit of the model is asymptotically χ^2 with D.F. = (t - u); given the model, X^2_C for testing contrasts of the model parameters is asymptotically χ^2 with D.F. = c under the null hypothesis, $\underset{\sim}{C}\underset{\sim}{\beta} = \underset{\sim}{0}$.

3. THE LINEAR MODEL APPROACH

3.1 Notation and Assumption

In this section, the general methodology for categorical data discussed by Grizzle et. al.[1969] will be outlined in more detail. The underlying probability model is that defined by (2.2.1) with the notation referring to the expected cell probabilities and hypothetical data shown in Table 3.1.1.

Table 3.1.1 Expected cell probabilities (cell frequencies) for the standard contingency table

Populations (factors)	Response Categories 1	2	...	s	Totals
1	π_{11}	π_{12}	...	π_{1s}	1
	(n_{11})	(n_{12})		(n_{1s})	(n_{1o})
2	π_{21}	π_{22}	...	π_{2s}	1
	$(n_{21}$	(n_{22})		(n_{2s})	(n_{2o})
.	.	.	...	.	.
.	.	.	...	.	.
.	.	.	...	.	.
r	π_{r1}	π_{r2}	...	π_{rs}	1
	(n_{r1})	(n_{r2})		(n_{rs})	(n_{ro})

Let $\underset{\sim}{\pi}'$ be defined as in (2.2.4) and let $\underset{\sim}{p}'$ be the corresponding vector of unrestricted MLE's of the respective components of $\underset{\sim}{\pi}'$; i.e.,

$$\underset{1\times rs}{\underset{\sim}{p}'} = (p_{11}, p_{12}, \ldots, p_{1s}; p_{21}, p_{22}, \ldots, p_{2s}; \ldots; p_{r1}, p_{r2}, \ldots, p_{rs}) = (\underset{\sim}{p}_1', \underset{\sim}{p}_2', \ldots, \underset{\sim}{p}_r') \tag{3.1.1}$$

where $p_{ij} = (n_{ij}/n_{io})$ for $i = 1, 2, \ldots, r$ and $j = 1, 2, \ldots, s$.

Let $\underset{\sim}{V}(\underset{\sim}{\pi}_i)$ be the variance-covariance matrix of the unrestricted MLE's from the i-th population; hence,

$$\underset{s\times s}{V}(\underset{\sim}{\pi}_i) = \text{var}(\underset{\sim}{p}_i) = \frac{1}{n_{io}}\begin{bmatrix} \pi_{i1}(1-\pi_{i1}) & -\pi_{i1}\pi_{i2} & \cdots & -\pi_{i1}\pi_{is} \\ & \pi_{i2}(1-\pi_{i2}) & \cdots & -\pi_{i2}\pi_{is} \\ & & \ddots & \\ \text{(symmetric)} & & & \pi_{is}(1-\pi_{is}) \end{bmatrix} \qquad (3.1.2)$$

A consistent estimator for $\underset{\sim}{V}(\underset{\sim}{\pi}_i)$ independent of any hypothesis on the $\underset{\sim}{\pi}_i$ is

$$\underset{s\times s}{\hat{V}_i} = \underset{\sim}{V}(\underset{\sim}{p}_i) = \underset{\sim}{V}(\underset{\sim}{\pi}_i)_{\underset{\sim}{\pi}_i = \underset{\sim}{p}_i} = \text{sample estimate of } \underset{\sim}{V}(\underset{\sim}{\pi}_i) \qquad (3.1.3)$$

Hence, a consistent estimator for the variance-covariance matrix of $\underset{\sim}{p}$ is

$$\underset{rs\times rs}{\hat{V}} = \underset{\sim}{V}(\underset{\sim}{p}) = \text{block diagonal matrix with diagonal blocks of the form } \frac{1}{n_{io}}(\underset{\sim}{D}_{\underset{\sim}{p}_i} - \underset{\sim}{p}_i\underset{\sim}{p}_i') \qquad (3.1.4)$$

where $\underset{\sim}{D}_{\underset{\sim}{p}_i} = \text{diagonal}\ (p_{i1}, p_{i2}, \ldots, p_{is})$.

Let $\underset{\sim}{f}(\underset{\sim}{\pi})$ be defined as in (2.2.3) and define $\underset{\sim}{H}$ and $\underset{\sim}{S}$ by

$$\underset{t\times rs}{H} = \underset{\sim}{H}(\underset{\sim}{p}) = \left(\frac{\partial f_k(\underset{\sim}{\pi})}{\partial \pi_{ij}}\right)_{\underset{\sim}{\pi}=\underset{\sim}{p}} \qquad (3.1.5)$$

$$\underset{t\times t}{S} = \underset{\sim}{H}\hat{\underset{\sim}{V}}\underset{\sim}{H}' = \text{sample estimate of the variance-covariance matrix of } \underset{\sim}{f}(\underset{\sim}{p}) \text{ after linearization as in (2.3.4)} \qquad (3.1.6)$$

It is assumed that the $f_k(\underset{\sim}{\pi})$ are functionally independent of each other and of the constraints, $\sum_{j=1}^{s} \pi_{ij} = 1$, for $i=1,2,\ldots,r$; hence $\underset{\sim}{H}$ and $\underset{\sim}{H}\hat{\underset{\sim}{V}}\underset{\sim}{H}'$ are asymptotically of full rank. Finally, if certain $n_{ij} = 0$, the authors suggest that they be replaced, in certain cases, by $(1/s)$ to eliminate possible singularities in $\underset{\sim}{S}$. (See Berkson [1955]).

3.2. Methods of Inference

Assume that $\underset{\sim}{f}(\underset{\sim}{\pi}) = \underset{\sim\sim}{X}\underset{\sim}{\beta}$ where the f_k's are possibly non-linear and $\underset{\sim}{X}$ and $\underset{\sim}{\beta}$ are defined following (2.2.5). For the no factor, multi-response case (i.e., one population problems), the relevant hypotheses are those of various types of independence and symmetry. These are tested using

$$X_F^2 = SS(\underset{\sim}{f}(\underset{\sim}{\pi}) = \underset{\sim}{0}) = \underset{\sim}{f}'\underset{\sim\sim}{S}^{-1}\underset{\sim}{f} = X_W^2 \qquad (3.2.1)$$

which, under H_0: $\underset{\sim}{f}(\underset{\sim}{\pi}) = \underset{\sim}{0}$, is asymptotically χ^2 with D.F. = t. If $\underset{\sim}{f}(\underset{\sim}{\pi})$ is linear, i.e., $\underset{\sim}{f}(\underset{\sim}{\pi}) = \underset{\sim\sim}{A}\underset{\sim}{\pi}$ where

$$\underset{t\times rs}{\underset{\sim\sim}{A}} = \begin{bmatrix} a_{(1)11} \cdots a_{(1)1s}; a_{(1)21} \cdots a_{(1)2s}; \cdots; a_{(1)r1} \cdots a_{(1)rs} \\ a_{(2)11} \cdots a_{(2)1s}; a_{(2)21} \cdots a_{(2)2s}; \cdots; a_{(2)r1} \cdots a_{(2)rs} \\ \cdot \quad \cdot \quad \cdot \quad \cdot \quad \cdot \quad \cdot \\ \cdot \quad \cdot \quad \cdot \quad \cdot \quad \cdot \quad \cdot \\ \cdot \quad \cdot \quad \cdot \quad \cdot \quad \cdot \quad \cdot \\ a_{(t)11} \cdots a_{(t)1s}; a_{(t)21} \cdots a_{(t)2s}; \cdots; a_{(t)r1} \cdots a_{(t)rs} \end{bmatrix}$$

$$= \begin{bmatrix} \underset{\sim}{a}'(1) \\ \underset{\sim}{a}'(2) \\ \cdot \\ \cdot \\ \cdot \\ \underset{\sim}{a}'(t) \end{bmatrix}$$

is an appropriate matrix of constants of rank t, then (3.2.1) is more explicitly given by

$$X_F^2 = SS(\underset{\sim}{f}(\underset{\sim}{\pi}) = \underset{\sim\sim}{A}\underset{\sim}{\pi} = \underset{\sim}{0}) = (\underset{\sim\sim}{A}\underset{\sim}{p})'(\underset{\sim\sim\sim}{\hat{A}\hat{V}A}')^{-1}(\underset{\sim\sim}{A}\underset{\sim}{p}). \qquad (3.2.2)$$

A second class of functions often arising in categorical data problems is the class consisting of the logarithmic functions which can be expressed by $\underset{\sim}{f}(\underset{\sim}{\pi}) = \underset{\sim}{K}\ \underset{\sim}{\ell n}(\underset{\sim}{A}\underset{\sim}{\pi})$ where $\underset{\sim}{\ell n}(\underset{\sim}{A}\underset{\sim}{\pi})$ denotes the vector of natural (or Naperian) logarithms of the elements of $\underset{\sim}{A}\underset{\sim}{\pi}$, and $\underset{v\times t}{K} = (k_{cv})$ is an appropriate matrix of constants of rank $v \leq t$. The corresponding test statistic for this case is given by

$$X_F^2 = SS(\underset{\sim}{f}(\underset{\sim}{\pi}) = \underset{\sim}{K}\ \underset{\sim}{\ell n}(\underset{\sim}{A}\underset{\sim}{\pi}) = \underset{\sim}{0}) = \tag{3.2.3}$$

$$(\underset{\sim}{K}\ \underset{\sim}{\ell n}(\underset{\sim}{A}\underset{\sim}{p}))'(\underset{\sim}{K}\underset{\sim}{D}^{-1}\underset{\sim}{A}\hat{\underset{\sim}{V}}\underset{\sim}{A}'\underset{\sim}{D}^{-1}\underset{\sim}{K}')^{-1}(\underset{\sim}{K}\ \underset{\sim}{\ell n}(\underset{\sim}{A}\underset{\sim}{p}))$$

where $\underset{t\times t}{D} = \text{diagonal}\ (\underset{\sim}{a}'_{(1)}\cdot\underset{\sim}{p}, \underset{\sim}{a}'_{(2)}\cdot\underset{\sim}{p}, \ldots, \underset{\sim}{a}'_{(t)}\cdot\underset{\sim}{p})$. Under H_0, X_F^2 is asymptotically χ^2 with D.F. = v. It should be noted that caution must be taken in constructing $\underset{\sim}{A}$ so that no element of the vector, $\underset{\sim}{A}\underset{\sim}{p}$, is zero. Often when logarithmic functions are used, there is interest in multiplicative models for $\underset{\sim}{\pi}$ and hence $\underset{\sim}{A}=\underset{\sim}{I}$=(identity matrix). Since empty cells are to be replaced by 1/s, each element of $\underset{\sim}{I}\underset{\sim}{p}$ (= p) will be positive so that no difficulty is encountered with this common application of the logarithmic functions.

For example, consider the test of independence for the two-way case cited in section 2.2. Let $j_1 = 1, 2$; $j_2 = 1, 2$. Then

$$H_0: \underset{\sim}{f}(\underset{\sim}{\pi}) = \underset{\sim}{K}\ \underset{\sim}{\ell n}(\underset{\sim}{A}\underset{\sim}{\pi}) = \underset{\sim}{0}$$

is given by

$$f(\underset{\sim}{\pi}) = \ln\left(\frac{\pi_{11}\pi_{22}}{\pi_{12}\pi_{21}}\right)$$

$$= \ln\pi_{11} - \ln\pi_{12} - \ln\pi_{21} + \ln\pi_{22} = 0$$

so that

$$\underset{\sim}{A} = \underset{\sim}{I}_4 = \text{identity matrix}$$

$$\underset{\sim}{K} = (1, -1, -1, 1)$$

$$\hat{\underset{\sim}{V}} = \underset{\sim}{V}(\underset{\sim}{p}) = \frac{1}{N}\begin{bmatrix} p_{11}(1-p_{11}) & -p_{11}p_{12} & -p_{11}p_{21} & -p_{11}p_{22} \\ & p_{12}(1-p_{12}) & -p_{12}p_{21} & -p_{12}p_{22} \\ & & p_{21}(1-p_{21}) & -p_{21}p_{22} \\ \text{(symmetric)} & & & p_{22}(1-p_{22}) \end{bmatrix}$$

where

$$N = \sum_{j_1=1}^{2}\sum_{j_2=1}^{2} n_{j_1 j_2}$$

$$\underset{\sim}{D} = \text{diagonal } (p_{11}, p_{12}, p_{21}, p_{22})$$

and hence

$$\underset{\sim\sim}{K}\underset{\sim}{D}^{-1}\underset{\sim}{A} = \left(\frac{1}{p_{11}}, -\frac{1}{p_{12}}, -\frac{1}{p_{21}}, \frac{1}{p_{22}}\right)$$

$$\underset{\sim\sim}{K}\underset{\sim}{D}^{-1}\underset{\sim\sim}{A}\hat{\underset{\sim\sim}{V}} = \frac{1}{N}(1, -1, -1, 1)$$

$$\underset{\sim\sim}{K}\underset{\sim}{D}^{-1}\underset{\sim\sim}{A}\hat{\underset{\sim\sim}{V}}\underset{\sim\sim}{A}'\underset{\sim}{D}^{-1}\underset{\sim}{K}' = \frac{1}{N}\left(\frac{1}{p_{11}} + \frac{1}{p_{12}} + \frac{1}{p_{21}} + \frac{1}{p_{22}}\right)$$

Thus, the test statistic with D.F. = 1 is given by

$$X_F^2 = SS(\underset{\sim}{f}(\pi) = \underset{\sim}{K}\,\ell n(\underset{\sim\sim}{A}\pi) = \underset{\sim}{0}) \qquad (3.2.4)$$

$$= \frac{N(\ell n\, p_{11} - \ell n\, p_{12} - \ell n\, p_{21} + \ell n\, p_{22})^2}{\frac{1}{p_{11}} + \frac{1}{p_{12}} + \frac{1}{p_{21}} + \frac{1}{p_{22}}}$$

For the uni- or multi-factor, uni- or multi-response case (i.e., several population problems), $\underset{\sim}{X}$ is known and of rank $u \leq t$. A test of fit of the model $\underset{\sim}{f}(\underset{\sim}{\pi}) = \underset{\sim}{X}\underset{\sim}{\beta}$, is given by the residual sum of squares

$$X_F^2 = SS(\underset{\sim}{f}(\underset{\sim}{\pi}) = \underset{\sim}{X}\underset{\sim}{\beta}) = \underset{\sim}{f}'\underset{\sim}{S}^{-1}\underset{\sim}{f} - (\underset{\sim}{X}\underset{\sim}{b})'\underset{\sim}{S}^{-1}(\underset{\sim}{X}\underset{\sim}{b}) \qquad (3.2.5)$$

where $\underset{\sim}{b}$ is the vector that minimizes $(\underset{\sim}{f} - \underset{\sim}{X}\underset{\sim}{b})'\underset{\sim}{S}^{-1}(\underset{\sim}{f} - \underset{\sim}{X}\underset{\sim}{b})$ and is given by

$$\underset{\sim}{b} = (\underset{\sim}{X}'\underset{\sim}{S}^{-1}\underset{\sim}{X})^{-1}\underset{\sim}{X}'\underset{\sim}{S}^{-1}\underset{\sim}{f}. \qquad (3.2.6)$$

If the model fits the data, X_F^2 is asymptotically χ^2 with D.F. = t - u. Recalling that for the linear case, $\underset{\sim}{f} = \underset{\sim}{A}\underset{\sim}{p}$, $\underset{\sim}{S} = \underset{\sim}{A}\hat{\underset{\sim}{V}}\underset{\sim}{A}'$ and for the logarithmic case, $\underset{\sim}{f} = \underset{\sim}{K}\,\ell n(\underset{\sim}{A}\underset{\sim}{p})$, $\underset{\sim}{S} = \underset{\sim}{K}\underset{\sim}{D}^{-1}\underset{\sim}{A}\hat{\underset{\sim}{V}}\underset{\sim}{A}'\underset{\sim}{D}^{-1}\underset{\sim}{K}'$, explicit expressions for the test statistic given in (3.2.5) can easily be obtained for these two classes of functions.

Given that the model fits, various hypotheses concerning constraints on the model parameters, $\underset{\sim}{\beta}$, may be of interest. The choice of hypotheses is illustrated in considerable detail in section 5. A test of the hypothesis, H_{03}: $\underset{\sim}{C}\underset{\sim}{\beta} = \underset{\sim}{0}$, is produced by

$$X_C^2 = SS(\underset{\sim}{C}\underset{\sim}{\beta} = \underset{\sim}{0}) = (\underset{\sim}{C}\underset{\sim}{b})'[\underset{\sim}{C}(\underset{\sim}{X}'\underset{\sim}{S}^{-1}\underset{\sim}{X})^{-1}\underset{\sim}{C}']^{-1}(\underset{\sim}{C}\underset{\sim}{b}) \qquad (3.2.7)$$

where $\underset{\sim}{C}$ is a (cxu) matrix of arbitrary constants of rank $c \leq u$. Under H_{03}, X_C^2 is asymptotically χ^2 with D.F. = c. Again, explicit expressions for (3.2.7) can easily be obtained for the linear and logarithmic classes of functions.

4. MORE GENERAL CATEGORICAL DATA MODELS

4.1. Contingency Tables with Supplemented Margins

As has been indicated previously, the approach of Grizzle et.al. [1969] is based upon applying weighted least squares analysis to certain appropriately formulated linear models similar to those arising in univariate and/or multivariate analysis. As such, the following assumptions are made:

(i) All cells are assumed to have positive probability of occurrence.

(ii) The data is complete in the sense that every experimental unit is classified according to each of the d dimensions of the table.

In many practical problems, some of these assumptions must be relaxed. Goodman [1968] and Bishop and Fienberg [1969], among others, have considered (i) by dealing with the implications of having a priori certain empty cells. Their approach utilizes multiplicative models with maximum likelihood estimation. The relaxation of (i) for the linear model approach is handled by suitable definition of $\underset{\sim}{\Gamma}(\pi)$ and is discussed in Grizzle and Williams [1970]. The relaxation of (ii) is the primary concern of this section.

In the usual categorical data situation, there is complete information available for each entry in the corresponding contingency table. For instance, in a four-dimensional table of the two-factor, two-response type, each experimental unit has been classified according to all four dimensions. However, in many situations, the experimenter is not equally interested in the response variates or perhaps additional information on one particular dimension could be obtained relatively easily and economically.

Alternatively, perhaps, as described in Kleinbaum [1970] for the continuous case, there is missing data for some individuals in the sense that classifications on only $d' < d$ dimensions have been made. Retaining and utilizing this partial information is distinctly different from the "apparently" similar problem of empty cells (see Assumption(i)).

Here, the additional information is assumed to arise by design rather than by chance and as such is referred to as "supplemental information" and the corresponding margins of the contingency table as "supplemented margins". This supplemental information is used to improve the precision of the estimates of the marginal probabilities over the supplemented dimension(s), and it is also incorporated into the estimation of the individual cell probabilities in order to improve certain tests of hypotheses.

Examples where such supplemental information might arise include questionnaire surveys similar to the recently-completed national census. In such surveys, every respondent would be expected to answer a basic set of questions. In addition, a certain proportion would be asked to respond to a second set of questions judged to be less important or more difficult to answer than the questions in the basic set. Supplemental information would derive from the answers to questions in the basic set from those respondents not requested to consider the second set of questions. Similarly, in automobile crash investigations, the investigating officer might routinely obtain certain biographical information (e.g., race, sex, age) for all drivers but, due to the difficult conditions of the interview, obtain, only for a specified subset, additional information such as sobriety of the driver, purpose

of the trip, etc. In this case, the supplemental information would consist of the answers to those biographical questions for those drivers who were not examined in detail.

It should be noted here that the use of supplemented margins in the analysis of contingency tables resembles the technique of "double sampling" as described in Cochran [1963]. In "double sampling", a preliminary sample (cf. supplemental information) is used to estimate the distribution of an auxiliary variable (cf. marginal distribution of particular response or combination of responses) and a second stratified sample is used to estimate the characteristic of interest (cf. cell probability).

Some work in this area has recently been done by Blumenthal [1968] who treats essentially the same problem as described here. However, he restricts attention to one-way tables where each of the I main categories has J_i, i = 1, 2, ..., I, sub-categories of classification. In addition, Blumenthal assumes that a sample unit belonging in cell (i,j) has probability α_{ij} of being only "partially classified" as opposed to having a secondary sample for which supplementary information on a subset of the study variables is obtained. Imposing certain simplifying assumptions for this one-way situation, Blumenthal obtains MLE's of the individual cell probabilities as well as the bias and variance of these estimates.

It should be noted that supplementation is considered only for margins that are not fixed (i.e., response margins rather than factor margins). Supplementing a factor margin is of no interest since that margin being considered fixed a priori yields complete information about its composition.

In order to see more clearly the implications of using supplemental data, let us consider the special case corresponding to the simple 2x2 contingency table of the "no factor, two response" type. Let A denote one of the response variables and let B denote the other. If n_{oo} subjects are classified according to both A and B while n_{o*} are classified only according to A and n_{*o} are classified only according to B, the resulting data can be displayed as in Table 4.1.1. It is assumed that there is

Table 4.1.1.

	Response B Levels 1	2	Sub-total	Supplement A	Total
Response 1	n_{11}	n_{12}	n_{1o}	n_{1*}	N_{1o}
A Levels 2	n_{21}	n_{22}	n_{2o}	n_{2*}	N_{2o}
Sub-total	n_{o1}	n_{o2}	n_{oo}	n_{o*}	N_A
Supplement B	n_{*1}	n_{*2}	n_{*o}		
Total	N_{o1}	N_{o2}	N_B		N

no interaction between subjects and the presence of supplemental data; i.e., for the responses classified, the joint marginal distributions for subjects with complete data and for those with supplemental data are the same. Conditional on the number of individuals with various types of

complete or supplemental data being fixed by design a priori (or by assumption a posteriori), the basic probability model is the product of several multinomials like (2.2.1) with marginal and/or cell probabilities as fundamental parameters. For the example in Table 4.1.1, this model implies that

$$E\begin{bmatrix} n_{11} \\ n_{12} \\ n_{21} \\ n_{1*} \\ n_{*1} \end{bmatrix} = \begin{bmatrix} n_{oo} & 0 & 0 \\ 0 & n_{oo} & 0 \\ 0 & 0 & n_{oo} \\ n_{o*} & n_{o*} & 0 \\ n_{*o} & 0 & n_{*o} \end{bmatrix} \begin{bmatrix} \pi_{11} \\ \pi_{12} \\ \pi_{21} \end{bmatrix} \tag{4.1.1}$$

and that a consistent estimate for the variance of the various frequencies is $\hat{V}$ as given by

$$\begin{bmatrix} n_{oo}p_{11}(1-p_{11}) & -n_{oo}p_{11}p_{12} & -n_{oo}p_{11}p_{21} & 0 & 0 \\ & n_{oo}p_{12}(1-p_{12}) & -n_{oo}p_{12}p_{21} & 0 & 0 \\ & & n_{oo}p_{21}(1-p_{21}) & 0 & 0 \\ & \text{symmetric} & & n_{o*}p_{1*}(1-p_{1*}) & 0 \\ & & & & n_{*o}p_{*1}(1-p_{*1}) \end{bmatrix} \tag{4.1.2}$$

where $p_{j_1 j_2} = (n_{j_1 j_2}/n_{oo})$, $p_{j_1 *} = (n_{j_1 *}/n_{o*})$, $p_{*j_2} = (n_{*j_2}/n_{*o})$ with $j_1, j_2 = 1, 2$. Reinfurt [1970] indicates how the supplemental information can be used to obtain improved estimators for the parameters $\pi_{11}, \pi_{12}, \pi_{21}, \pi_{22}$. He discusses both maximum likelihood estimation and weighted least squares estimation (which if iterated leads to estimators which are equivalent to the MLE's). Either of these estimators for the $\pi_{j_1 j_2}$ can be used to play the same role as the unrestricted MLE's in the unsupplemented case. This means that one can test hypotheses of the form (2.2.2) or (2.2.5) by using the linear model approach of Section 3. All that is required is to formulate appropriate definitions of the A, K, X, and C matrices and to replace $\hat{V}$ as in (3.1.4) by the consistent estimator of the asymptotic covariance matrix for the estimators of the $\pi_{j_1 j_2}$ which have been adjusted for supplementation. All of this can be put together in terms of a two-stage test procedure utilizing the computer program described by Forthofer et.al.[1969] as follows. The first stage consists of the following:

I.1 Input the observed cell and marginal frequencies.

I.2 Obtain estimates of the cell probabilities corresponding to the unrestricted MLE's for the case with complete data:

(a) If the table is fairly simple, obtain the MLE's. (This is the closest analogue to the unrestricted MLE's for the complete data version and is most consistent with tests using Wald statistics.)

(b) For most cases, compute the weighted least squares estimators by iterative techniques.

(c) If interested in estimating functions of the cell probabilities (say, $A\underset{\sim}{\pi}$), rather than of $\underset{\sim}{\pi}$ obtain weighted least squares estimators of these expressions directly. This is particularly appropriate when sample sizes are small.

I.3 Obtain an estimate, $\hat{\Omega}$ (or $A\hat{V}A'$ for (c)), of the asymptotic covariance matrix using the estimates derived in (I.2).

The second stage consists of the following:

II.1 Input the A, K, X, and C matrices required for the analysis as in the existing program.

II.2 Use the adjusted (for supplemental data) estimates of the cell probabilities as obtained in (I.2) of the first stage in place of the usual estimates along with the adjusted estimate of the asymptotic covariance matrix in (I.3) of the first stage in place of the usual estimate. Proceed with the testing of various hypotheses exactly as in the existing linear models framework.

An example of this two-stage procedure is given in Section 5.4.

4.2 Contingency Tables from Sample Surveys of Finite Populations

The general methods of analysis discussed by Grizzle et al. [1969], Goodman [1970], and Ku and Kullback [1968] presume that the underlying model for complex contingency tables is the product multinomial distribution. This model does not directly apply to categorical data arising from sample survey situations since the selected subjects give responses which can be often either determined with complete certainty (e.g., sex, amount of education, year of birth, years of experience) or which have widely varying probability distributions

(e.g., opinions regarding suitability of certain types of health services or programs, the occurrences of accidents or diseases during specified time periods, the occurrence of a positive or other type of response to various methods of therapy). In addition, if the actual survey design is a complex highly stratified cluster scheme with varying probabilities of selection, the underlying probability distribution becomes considerably more difficult to handle.

In spite of these problems, it does follow from the theory of sample survey methodology that, in large samples, the estimated proportions in corresponding complex contingency tables have approximately multivariate normal distributions with a covariance matrix which can be consistently estimated by straightforward and well-known methods (as developed in Cochran [1963]). Since this fact is really the crucial assumption underlying the use of the approach of Grizzle et al. [1969] in an ultimate sense, all that is required to apply their results to categorical data from sample surveys is to make certain appropriate modifications consistent with the principles of survey sampling.

Some progress has been made in this direction for the case of stratified simple random samples. A number of results based on the linear models approach have been described in Johnson and Koch [1970b] and are illustrated in terms of health services evaluation data obtained from a sample survey of hospital administrators. Alternatively, Nathan [1969, 1970] has proceeded from a somewhat different point of view and concentrated substantial effort on tests for the hypothesis of independence in two way contingency tables derived from stratified samples.

Finally, it should be noted that for this situation the analysis may be viewed in a two-stage fashion with essentially the same structure as that described in Section 4.1. However, here the estimates of the parameters at the first-stage are developed according to the principles of sample survey methodology. For simple survey designs, such estimates are obtained in a straightforward fashion; in more complex situations with unequal selection probabilities, their derivation as well as the corresponding estimated variance-covariance matrix is more difficult; however, once the first stage is completed, the second stage is essentially the same as outlined previously and involves only the proper specification of $\underset{\sim}{A}$, $\underset{\sim}{K}$, $\underset{\sim}{X}$ and $\underset{\sim}{C}$ matrices. Finally, an example of a sample survey situation is described in Section 5.5.

5. EXAMPLES

5.1. A simple quantal biological assay.

One type of experimental situation in which categorical data often arise is the biological assay. The purpose of such investigations is to measure the potency of some given treatment relative to some standard treatment in terms of the magnitudes of their effects on responses from living material (see Finney [1964]). If the measured response is expressed in all or nothing fashion like lives or dies, then the experiment is called a quantal bioassay.

An hypothetical example of the results of a quantal bioassay is given in columns 1-4 of Table 5.1.1. The experimental design here involves assigning 20 animals to receive each of the specified doses

TABLE 5.1.1

Preparation	Dose	Alive	Dead	Proportion	Logit	Estimated Var. Logit	Predicted Logit X_2	Residual
Standard	1	19	1	0.05	-2.94	1.05	-2.60	-0.34
Standard	2	15	5	0.25	-1.10	0.27	-1.30	0.20
Standard	4	11	9	0.45	-0.20	0.20	0.00	-0.20
Standard	8	4	16	0.80	1.39	0.31	1.30	0.09
Standard	16	1	19	0.95	2.94	1.05	2.61	0.33
Treatment	1	16	4	0.20	-1.39	0.31	-1.58	0.19
Treatment	2	12	8	0.40	-0.41	0.21	-0.28	-0.13
Treatment	4	5	15	0.75	1.10	0.27	1.03	0.07
Treatment		2	18	0.90	2.20	0.56	2.33	-0.13

of the two preparations which have been labelled here as "standard" and "treatment". It can be noted that as the dose of either preparation is increased, the proportion of animals which die also increases. It is this relationship between the proportion of dead animals and dose for which a categorical data model is needed. One model which has received extensive attention in the literature is based on the logistic function. In this case, it is assumed that as dose is increased on the log dose scale, the proportion of dead animals increases according to the cumulative probability function of the logistic distribution; i.e., if P(x) denotes the proportion of dead animals as a function of log dose x, then the assumed model is

$$P(x) = \{1 + \exp[-(\alpha + \beta x)]\}^{-1}. \qquad (5.1.1)$$

A second assumption which is made is that the treatment preparation acts like a dilution of the standard in the sense that there exists a parameter ρ called the relative potency of the treatment such that for every dose u_T of treatment, the effect on the proportion of dead animals is the same as the dose $u_S = \rho u_T$ of standard. In terms of the model (5.1.1) with subscripts T and S referring to treatment and standard, this means

$$\begin{aligned} P_T(x_T = \log_{10} u_T) &= P_S(x_S = \log_{10}(\rho u_T)) \qquad (5.1.2) \\ &= \{1 + \exp[-(\alpha_S + \beta x_S)]\}^{-1} \\ &= \{1 + \exp[-(\alpha_S + \beta \log_{10} \rho + \beta x_T)\}^{-1} \\ &= \{1 + \exp[-(\alpha_T + \beta x_T)]\}^{-1}. \end{aligned}$$

In other words, the functions P_S and P_T have different parameters α_S and α_T but the same β. Moreover the relative potency ρ is related to the parameters α_S, α_T, β in the sense that

$$\log_{10}\rho = (\alpha_T - \alpha_S)/\beta \tag{5.1.3}$$

It should be noted here that log dose $x = \log u$ and $\log \rho$ can be expressed in terms of any base which is convenient. Here, $\log_{10}$ has been used for simplicity since $\log_{10}(2) \simeq 0.3$.

The model implied by (5.1.1) or (5.1.2) can be fitted to the data in a very simple manner as soon as it is recognized that

$$\begin{aligned} -\ln\{\frac{1}{P(x)} - 1\} &= \ln\{\frac{P(x)}{1 - P(x)}\} \\ &= -\ln\{\exp[-(\alpha + \beta x)]\} \\ &= \alpha + \beta x \end{aligned} \tag{5.1.4}$$

where, as before, $\ln \equiv \log_e$. The quantity $\ln\{\frac{P(x)}{1-P(x)}\}$ is often referred to as the "logit." For the given example, its values appear in column 6 of Table 5.1.1.

The validity of the various assumptions can be tested by fitting the following linear model to the vector of logits

$$\underset{\sim}{X}_1 = \begin{bmatrix} 1 & 0 & 0.0 & 0.0 \\ 1 & 0 & 0.3 & 0.0 \\ 1 & 0 & 0.6 & 0.0 \\ 1 & 0 & 0.9 & 0.0 \\ 1 & 0 & 1.2 & 0.0 \\ 0 & 1 & 0.0 & 0.0 \\ 0 & 1 & 0.0 & 0.3 \\ 0 & 1 & 0.0 & 0.6 \\ 0 & 1 & 0.0 & 0.9 \end{bmatrix} \tag{5.1.5}$$

A statistical test for the goodness of fit of the logistic model (5.1.1) is provided by the weighted residual sum of squares which in this example is $X^2 = 0.75$ with D.F. = 5; hence it is reasonable to conclude that the logistic model is appropriate for these data. The assumption that the treatment acts like a dilution of the standard is equivalent to the hypothesis that P_S and P_T have the same β or equivalently that the lines fitted to the two sets of logits in (5.1.5) are parallel. This can be formulated here in terms of the $\underset{\sim}{C}$-matrix [0 0 1 -1] as described in Section 3. The resulting $X^2 = 0.76$ with D.F. = 1 and hence justifies the dilution assumption.

Since the previously described tests support the validity of the model specified by (5.1.1) and (5.1.2) together, the final step is to fit the reduced model with

$$\underset{\sim}{X}_2 = \begin{bmatrix} 1 & 0 & 0.0 \\ 1 & 0 & 0.3 \\ 1 & 0 & 0.6 \\ 1 & 0 & 0.9 \\ 1 & 0 & 1.2 \\ 0 & 1 & 0.0 \\ 0 & 1 & 0.3 \\ 0 & 1 & 0.6 \\ 0 & 1 & 0.9 \end{bmatrix} \qquad (5.1.6)$$

to the vector of logits. The goodness of fit test for this model has $X^2 = 0.85$ with D.F. = 6 which further verifies our previous conclusions with respect to $\underset{\sim}{X}_1$. Estimates for the parameters in (5.1.6) are

$$\begin{aligned} \hat{\alpha}_S &= -2.60 \\ \hat{\alpha}_T &= -1.58 \\ \hat{\beta} &= 4.34 \end{aligned} \qquad (5.1.7)$$

and predicted values for the logits based on these estimates are given in column 8 of Table 5.1.1; the corresponding residuals appear in column 9. From the estimates in (5.1.7), it follows that an estimate for $\log_{10}\rho$ is

$$\widehat{\log_{10}\rho} = (\hat{\alpha}_T - \hat{\alpha}_S)/\hat{\beta} = (1.02)/(4.34) = 0.235.$$

Finally, by using the $\underset{\sim}{C}$-matrices [1 -1 0] and [0 0 1], test statistics for the hypothesis of equality of preparation effects and equality of dose effects can be tested; the results here are $X^2 = 6.49$ with D.F. = 1 and $X^2 = 46.62$ with D.F. = 1 respectively. For a further discussion of quantal bioassays, the reader is referred to Berkson [1955] and Finney [1964].

5.2 A factorial design type contingency table with interaction.

Many higher-dimensional contingency tables arise from experimental or survey situations which have a structure analogous to the typical multi-way factorial (cross-classification) design for continuous data. One such example which has been analyzed a number of times in the literature (e.g. Lombard and Doering [1947], Dyke and Patterson [1952], Cox and Snell [1968], Bishop [1969], and Goodman [1970]), appears in Table 5.2.1 and corresponds to data from a sample of 1729 individuals who were cross classified according to the following five categorical variables:

1. whether they read newspapers or not
2. whether they listened to the radio or not
3. whether they read books and magazines (solid reading) or not
4. whether they attended lectures or not
5. whether their knowledge of cancer was good or poor

TABLE 5.2.1.

Read Newspapers	Listen to Radio	Solid Reading	Attend Lectures	Cancer Knowledge Good	Cancer Knowledge Poor	Proportion Good Knowledge	Estimate of Variance	Predicted Value X_3	Predicted Value X_5	Predicted Value X_6
yes	yes	yes	yes	23	8	0.74	0.0062	0.74	0.74	0.69
yes	yes	yes	no	102	67	0.60	0.0014	0.60	0.60	0.61
yes	yes	no	yes	8	4	0.67	0.0185	0.55	0.54	0.47
yes	yes	no	no	35	59	0.37	0.0025	0.41	0.40	0.39
yes	no	yes	yes	27	18	0.60	0.0053	0.66	0.66	0.63
yes	no	yes	no	201	177	0.53	0.00066	0.52	0.52	0.54
yes	no	no	yes	7	6	0.54	0.0191	0.47	0.46	0.41
yes	no	no	no	75	156	0.32	0.00095	0.33	0.32	0.32
no	yes	yes	yes	1	3	0.25	0.0469	0.38	0.20	0.55
no	yes	yes	no	16	16	0.50	0.0078	0.53	0.50	0.46
no	yes	no	yes	4	3	0.57	0.0350	0.33	0.50	0.33
no	yes	no	no	13	50	0.21	0.0026	0.19	0.20	0.24
no	no	yes	yes	3	8	0.27	0.0180	0.30	0.31	0.48
no	no	yes	no	67	83	0.45	0.0016	0.44	0.45	0.40
no	no	no	yes	2	10	0.17	0.0116	0.18	0.18	0.26
no	no	no	no	84	393	0.18	0.00030	0.18	0.18	0.18

Strictly speaking, this is a "no factor, five response" situation since only the sample size n = 1729 is fixed and for each individual all classifications are random variables. On the other hand, the parameters of greatest practical interest pertain to the conditional probabilities of good or poor knowledge of cancer, given fixed categories for exposure to newspapers, radio, solid reading and lectures. This conditional point of view allows the situation to be considered for analytical purposes in terms of a "four factor, one response" model with knowledge of cancer being the response. Hence, the data are considered to represent 16 independent binomial samples as displayed in the 16 x 2 array in columns 5 and 6 of Table 5.2.1. The marginal total for each row is assumed to be fixed and the data from different rows are assumed to be statistically independent. These assumptions are essentially justified by the conditional point of view. The vector of conditional relative frequencies p associated with good knowledge of cancer in the various newspaper x radio x solid reading x lectures categories appears in column 7 of Table 5.2.1 and the corresponding estimated variances appear in column 8. It is appropriate to note that even without the assumptions associated with the "four factor, one response" model, these conditional relative frequencies are uncorrelated with one another. Hence, their analysis in terms of linear models and weighted least squares still applies independent of the conditional point of view. In other words, the only aspect of this discussion which is actually "conditional" is the set of parameters which are conditional probabilities of good knowledge of cancer as related to exposure to various media. However, this is not unreasonable since these are the parameters which can be dealt

with in practical terms because an individual's exposure to media can be controlled like a factor a priori even though here it was measured a posteriori.

Several linear models were fitted to the set of conditional relative frequencies in column 7 of Table 5.2.1. The first of these was

$$\underset{\sim}{X}_1 = \begin{bmatrix}
1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \\
1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & -1 & -1 & -1 & -1 & -1 & -1 & -1 & -1 \\
1 & 1 & 1 & 1 & -1 & -1 & -1 & -1 & 1 & 1 & 1 & 1 & -1 & -1 & -1 & -1 \\
1 & 1 & 1 & 1 & -1 & -1 & -1 & -1 & -1 & -1 & -1 & -1 & 1 & 1 & 1 & 1 \\
1 & 1 & -1 & -1 & 1 & 1 & -1 & -1 & 1 & 1 & -1 & -1 & 1 & 1 & -1 & -1 \\
1 & 1 & -1 & -1 & 1 & 1 & -1 & -1 & -1 & -1 & 1 & 1 & -1 & -1 & 1 & 1 \\
1 & 1 & -1 & -1 & -1 & -1 & 1 & 1 & 1 & 1 & -1 & -1 & -1 & -1 & 1 & 1 \\
1 & 1 & -1 & -1 & -1 & -1 & 1 & 1 & -1 & -1 & 1 & 1 & 1 & 1 & -1 & -1 \\
1 & -1 & 1 & -1 & 1 & -1 & 1 & -1 & 1 & -1 & 1 & -1 & 1 & -1 & 1 & -1 \\
1 & -1 & 1 & -1 & 1 & -1 & 1 & -1 & -1 & 1 & -1 & 1 & -1 & 1 & -1 & 1 \\
1 & -1 & 1 & -1 & -1 & 1 & -1 & 1 & 1 & -1 & 1 & -1 & -1 & 1 & -1 & 1 \\
1 & -1 & 1 & -1 & -1 & 1 & -1 & 1 & -1 & 1 & -1 & 1 & 1 & -1 & 1 & -1 \\
1 & -1 & -1 & 1 & 1 & -1 & -1 & 1 & 1 & -1 & -1 & 1 & 1 & -1 & -1 & 1 \\
1 & -1 & -1 & 1 & 1 & -1 & -1 & 1 & -1 & 1 & 1 & -1 & -1 & 1 & 1 & -1 \\
1 & -1 & -1 & 1 & -1 & 1 & 1 & -1 & 1 & -1 & -1 & 1 & -1 & 1 & 1 & -1 \\
1 & -1 & -1 & 1 & -1 & 1 & 1 & -1 & -1 & 1 & 1 & -1 & 1 & -1 & -1 & 1
\end{bmatrix} \qquad (5.2.1)$$

This is a complete model since the number of effects (i.e., columns of $\underset{\sim}{X}_1$) is equal to the number of distinct factor combinations (i.e., the number of rows of $\underset{\sim}{X}_1$). A description of the effects corresponding to the columns of $\underset{\sim}{X}_1$ together with significance tests of their importance appear in the respective rows of Table 5.2.2.

with in practical terms because an individual's exposure to media can be controlled like a factor *a priori* even though here it was measured *a posteriori*.

Several linear models were fitted to the set of conditional relative frequencies in column 7 of Table 5.2.1. The first of these was

$$\underset{\sim}{X}_1 = \begin{bmatrix}
1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \\
1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & -1 & -1 & -1 & -1 & -1 & -1 & -1 & -1 \\
1 & 1 & 1 & 1 & -1 & -1 & -1 & -1 & 1 & 1 & 1 & 1 & -1 & -1 & -1 & -1 \\
1 & 1 & 1 & 1 & -1 & -1 & -1 & -1 & -1 & -1 & -1 & -1 & 1 & 1 & 1 & 1 \\
1 & 1 & -1 & -1 & 1 & 1 & -1 & -1 & 1 & 1 & -1 & -1 & 1 & 1 & -1 & -1 \\
1 & 1 & -1 & -1 & 1 & 1 & -1 & -1 & -1 & -1 & 1 & 1 & -1 & -1 & 1 & 1 \\
1 & 1 & -1 & -1 & -1 & -1 & 1 & 1 & 1 & 1 & -1 & -1 & -1 & -1 & 1 & 1 \\
1 & 1 & -1 & -1 & -1 & -1 & 1 & 1 & -1 & -1 & 1 & 1 & 1 & 1 & -1 & -1 \\
1 & -1 & 1 & -1 & 1 & -1 & 1 & -1 & 1 & -1 & 1 & -1 & 1 & -1 & 1 & -1 \\
1 & -1 & 1 & -1 & 1 & -1 & 1 & -1 & -1 & 1 & -1 & 1 & -1 & 1 & -1 & 1 \\
1 & -1 & 1 & -1 & -1 & 1 & -1 & 1 & 1 & -1 & 1 & -1 & -1 & 1 & -1 & 1 \\
1 & -1 & 1 & -1 & -1 & 1 & -1 & 1 & -1 & 1 & -1 & 1 & 1 & -1 & 1 & -1 \\
1 & -1 & -1 & 1 & 1 & -1 & -1 & 1 & 1 & -1 & -1 & 1 & 1 & -1 & -1 & 1 \\
1 & -1 & -1 & 1 & 1 & -1 & -1 & 1 & -1 & 1 & 1 & -1 & -1 & 1 & 1 & -1 \\
1 & -1 & -1 & 1 & -1 & 1 & 1 & -1 & 1 & -1 & -1 & 1 & -1 & 1 & 1 & -1 \\
1 & -1 & -1 & 1 & -1 & 1 & 1 & -1 & -1 & 1 & 1 & -1 & 1 & -1 & -1 & 1
\end{bmatrix} \qquad (5.2.1)$$

This is a complete model since the number of effects (*i.e.*, columns of $\underset{\sim}{X}_1$) is equal to the number of distinct factor combinations (*i.e.*, the number of rows of $\underset{\sim}{X}_1$). A description of the effects corresponding to the columns of $\underset{\sim}{X}_1$ together with significance tests of their importance appear in the respective rows of Table 5.2.2.

Table 5.2.2.

Source of Variation	D.F.	X^2
Newspaper	1	17.94
Radio	1	4.10
Newspaper x Radio Interaction	1	0.03
Solid Reading	1	4.78
Newspaper x Solid Reading Interaction	1	0.29
Radio x Solid Reading Interaction	1	0.75
Newspaper x Radio x Solid Reading Interaction	1	1.10
Lectures	1	2.34
Newspaper x Lectures Interaction	1	3.44
Radio x Lectures Interaction	1	1.13
Newspapers x Radio x Lectures Interaction	1	0.12
Solid Reading x Lectures Interaction	1	6.55
Newspaper x Solid Reading x Lectures Interaction	1	1.28
Radio x Solid Reading x Lectures Interaction	1	1.19
Newspapers x Radio x Solid Reading x Lectures Interaction	1	1.09
Simultaneous effects of all sources	15	261.04

The analysis in Table 5.2.2 is analogous to that for a 2x2x2x2 factorial design. It indicates that there are significant effects due to newspapers ($\alpha = .01$), radio ($\alpha = .05$), solid reading ($\alpha = .05$), solid reading x lectures interaction ($\alpha = .05$), and newspaper x lecture interaction ($\alpha = .10$). However, once these conclusions have been noted, the next objective is to find a model with as few parameters as possible, but which still accounts for most of the variation in the set of conditional relative frequencies reflecting good knowledge of cancer. The results in Table 5.2.2. warn the analyst that he must pay careful attention to the interaction of lectures with newspapers and solid reading in formulating this model. Some insight for this purpose can be attained by looking at Table 5.2.3 which shows the effect of each of the factors within fixed levels of the other three factors.

Table 5.2.3

Read Newspapers	Listen to Radio	Solid Reading	Attend Lectures	Newspaper Effect	Radio Effect	Solid Reading Effect	Lectures Effects
Yes	Yes	Yes	Yes	0.49	0.14	0.07	0.14
Yes	Yes	Yes	No	0.10	0.07	0.23	*
Yes	Yes	No	Yes	0.10	0.13	*	0.30
Yes	Yes	No	No	0.16	0.05	*	*
Yes	No	Yes	Yes	0.33	*	0.06	0.07
Yes	No	Yes	No	0.08	*	0.21	*
Yes	No	No	Yes	0.37	*	*	0.22
Yes	No	No	No	0.14	*	*	*
No	Yes	Yes	Yes	*	-0.02	-0.32	-0.25
No	Yes	Yes	No	*	0.05	0.29	*
No	Yes	No	Yes	*	0.40	*	0.36
No	Yes	No	No	*	0.03	*	*
No	No	Yes	Yes	*	*	0.10	-0.18
No	No	Yes	No	*	*	0.27	*
No	No	No	Yes	*	*	*	-0.01
No	No	No	No	*	*	*	*

In this table, the difference between the proportion with good knowledge and the proportion with poor knowledge for a given factor is displayed for every combination of the other three factors in those rows of the table for which the given factor is at the "yes" value.

From Table 5.2.3, it is apparent that

1. In all cases, newspapers have a positive effect varying from moderate to large in magnitude.
2. Radio, for the most part, has a positive effect varying from small to large in magnitude.
3. Except for one instance where it showed a large negative effect, solid reading has a positive effect varying from small to large in magnitude.

4. The effects of lectures are fairly complex.

   a. For newspaper readers, it has a positive effect varying from small to large in magnitude.

   b. For newspaper non-readers but solid readers, it has a moderate negative effect.

   c. For newspaper non-readers, radio listeners, and non-solid readers, it has a large positive effect.

   d. For newspaper non-readers, radio non-listeners and non-solid readers, it has essentially no effect.

As a consequence of the preceding set of conclusions, the second linear model considered was

$$\underset{\sim}{X}_2 = \begin{bmatrix} 1 & 1 & 1 & 1 & 1 & 0 & 0 & 0 \\ 1 & 1 & 1 & 1 & -1 & 0 & 0 & 0 \\ 1 & 1 & 1 & -1 & 1 & 0 & 0 & 0 \\ 1 & 1 & 1 & -1 & -1 & 0 & 0 & 0 \\ 1 & 1 & -1 & 1 & 1 & 0 & 0 & 0 \\ 1 & 1 & -1 & 1 & -1 & 0 & 0 & 0 \\ 1 & 1 & -1 & -1 & 1 & 0 & 0 & 0 \\ 1 & 1 & -1 & -1 & -1 & 0 & 0 & 0 \\ 1 & -1 & 1 & 1 & 0 & 1 & 0 & 0 \\ 1 & -1 & 1 & 1 & 0 & -1 & 0 & 0 \\ 1 & -1 & 1 & -1 & 0 & 0 & 1 & 0 \\ 1 & -1 & 1 & -1 & 0 & 0 & -1 & 0 \\ 1 & -1 & -1 & 1 & 0 & 0 & 0 & 1 \\ 1 & -1 & -1 & 1 & 0 & 0 & 0 & -1 \\ 1 & -1 & -1 & -1 & 0 & 0 & 0 & 0 \\ 1 & -1 & -1 & -1 & 0 & 0 & 0 & 0 \end{bmatrix} \qquad (5.2.2)$$

This is an incomplete model since $\underset{\sim}{X}_2$ has $8 < 16$ columns. Its use must be justified by a goodness of fit test which measures the model's ability to account for the variation in the data. The test statistic for this is the weighted residual sum of squares shown in Table 5.2.4 along with significance tests of the respective effects corresponding to the columns of $\underset{\sim}{X}_2$.

Table 5.2.4

Source of Variation	D.F.	X^2
Newspaper	1	37.47
Radio	1	7.21
Solid Reading	1	44.88
Lectures within Newspaper Readers	1	7.46
Lectures within Newspaper Non-Readers, Radio Listeners, and Solid Readers	1	0.63
Lectures within Newspaper Non-Readers, Radio Listeners, and Non-Solid Readers	1	1.58
Lectures within Newspaper Non-Readers, Radio Non-Listeners, and Solid Readers	1	2.79
Simultaneous effects of above sources	7	256.34
Residual	8	4.70

Since the residual sum of squares is not significant ($\alpha = .25$), one may say that the model (5.2.2) provides an adequate fit to the data. In addition, significant effects are detected for newspapers ($\alpha = .01$), solid reading ($\alpha = .01$), and lectures within newspaper readers ($\alpha = .01$). Although the other effects of lectures are not detected as significant ($\alpha = .05$), it is not really appropriate to eliminate them from the model because the following estimates of the effects in $\underset{\sim}{X}_2$ indicate that they are essentially of the same magnitude as the significant effects of lectures within newspaper readers.

Overall Mean	0.424
Newspaper Effect	0.109
Radio Effect	0.041
Solid Reading Effect	0.096
Lectures within newspaper readers	0.068
Lectures within newspaper non-readers, radio listeners, and solid readers	-0.069
Lectures within newspaper non-readers, radio listeners, and non-solid readers	0.072
Lectures within newspaper non-readers, radio non-listeners, and solid readers	-0.078

As a result, the third model considered was

$$X_3 = \begin{bmatrix} 1 & 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 & -1 \\ 1 & 1 & 1 & -1 & 1 \\ 1 & 1 & 1 & -1 & -1 \\ 1 & 1 & -1 & 1 & 1 \\ 1 & 1 & -1 & 1 & -1 \\ 1 & 1 & -1 & -1 & 1 \\ 1 & 1 & -1 & -1 & -1 \\ 1 & -1 & 1 & 1 & -1 \\ 1 & -1 & 1 & 1 & 1 \\ 1 & -1 & 1 & -1 & 1 \\ 1 & -1 & 1 & -1 & -1 \\ 1 & -1 & -1 & 1 & -1 \\ 1 & -1 & -1 & 1 & 1 \\ 1 & -1 & -1 & -1 & 0 \\ 1 & -1 & -1 & -1 & 0 \end{bmatrix} \qquad (5.2.3)$$

in which all of the varying lecture effects have been placed into a single complex interaction term which reflects the different directions the effects display at different combinations of the other factors. The appropriate X^2-statistics for testing the effects in the model (5.2.3) are given in Table 5.2.5.

Table 5.2.5.

Source of Variation	D.F.	X^2
Newspaper	1	46.62
Radio	1	9.18
Solid Reading	1	54.38
Lectures as interacting with other factors	1	12.32
Simultaneous effects of above sources	4	256.30
Residual	11	4.74

Since the residual sum of squares is not significant ($\alpha = .25$), one may say that the model (5.2.3) provides a satisfactory fit to the

data. Moreover, all of the effects included in this model are significant at the $\alpha = .01$ level. Finally, estimates for the various effects are

Overall Mean	0.426
Newspaper Effect	0.109
Radio Effect	0.041
Solid Reading Effect	0.097
Lectures as interacting with other factors	0.071

Predicted values for the conditional probabilities of good knowledge of cancer within each factor combination are given for the model (5.2.3) in column 9 of Table 5.2.1.

Although the preceding discussion has led to a fairly satisfactory analysis, some additional considerations are worth noting. First of all, although the residual sum of squares for the model specified by X_3 is not significant ($\alpha = .25$) when compared to the chi-square distribution with D.F. = 11, its value 4.74 does exceed 3.84 which is the $\alpha = .05$ critical point of the chi-square distribution with D.F. = 1. Hence, it is possible that this residual conceals a significant single degree of freedom which is important in understanding the variation in the data. Whether this is actually the case or not is difficult to ascertain, although recognition of patterns in the differences between observed relative frequencies and corresponding predicted values provides some guidance. Alternatively, one can proceed to fit other models until a reasonable one is found for which the residual sum of squares is not only non-significant but does not exceed 3.84 and hence cannot conceal a significant component. For the data in this example, such a model was

$$\underset{\sim}{X}_4 = \begin{bmatrix} 1 & 1 & 1 & 1 & 1 & 0 & 0 & 0 \\ 1 & 1 & 1 & 1 & -1 & 0 & 0 & 0 \\ 1 & 1 & 1 & -1 & 1 & 0 & 0 & 0 \\ 1 & 1 & 1 & -1 & -1 & 0 & 0 & 0 \\ 1 & 1 & -1 & 1 & 1 & 0 & 0 & 0 \\ 1 & 1 & -1 & 1 & -1 & 0 & 0 & 0 \\ 1 & 1 & -1 & -1 & 1 & 0 & 0 & 0 \\ 1 & 1 & -1 & -1 & -1 & 0 & 0 & 0 \\ 1 & -1 & 1 & 0 & 0 & -1 & 0 & 0 \\ 1 & -1 & 1 & 0 & 0 & 1 & 0 & 0 \\ 1 & -1 & 1 & 0 & 0 & 1 & 0 & 0 \\ 1 & -1 & 1 & 0 & 0 & -1 & 0 & 0 \\ 1 & -1 & -1 & 0 & 0 & 0 & 1 & 1 \\ 1 & -1 & -1 & 0 & 0 & 0 & 1 & -1 \\ 1 & -1 & -1 & 0 & 0 & 0 & -1 & 0 \\ 1 & -1 & -1 & 0 & 0 & 0 & -1 & 0 \end{bmatrix} \qquad (5.2.4)$$

with the residual sum of squares $X^2 = 2.57$ with D.F. = 8. Because of certain similar patterns in magnitudes of effects the model based on $\underset{\sim}{X}_4$ was modified to

$$\underset{\sim}{X}_5 = \begin{bmatrix} 1 & 1 & 1 & 1 & 1 & 0 \\ 1 & 1 & 1 & 1 & -1 & 0 \\ 1 & 1 & 1 & -1 & 1 & 0 \\ 1 & 1 & 1 & -1 & -1 & 0 \\ 1 & 1 & -1 & 1 & 1 & 0 \\ 1 & 1 & -1 & 1 & -1 & 0 \\ 1 & 1 & -1 & -1 & 1 & 0 \\ 1 & 1 & -1 & -1 & -1 & 0 \\ 1 & -1 & 1 & 0 & 0 & -1 \\ 1 & -1 & 1 & 0 & 0 & 1 \\ 1 & -1 & 1 & 0 & 0 & 1 \\ 1 & -1 & 1 & 0 & 0 & -1 \\ 1 & -1 & -1 & 1 & -1 & 0 \\ 1 & -1 & -1 & 1 & 1 & 0 \\ 1 & -1 & -1 & -1 & 0 & 0 \\ 1 & -1 & -1 & -1 & 0 & 0 \end{bmatrix} \qquad (5.2.5)$$

by the same reasoning that led from $\underset{\sim}{X}_2$ to $\underset{\sim}{X}_3$. The appropriate X^2-statistics for testing the effects in the model (5.2.5) are given in Table 5.2.6.

Table 5.2.6

Source of Variation	D.F.	χ^2
Newspapers	1	40.66
Radio	1	7.63
Solid Reading as modified	1	56.51
Lectures as modified	1	9.71
Solid Reading x Lecture Interaction in Newspaper non-readers and radio listeners	1	11.37
Simultaneous effects of above sources	5	258.45
Residual	10	2.59

Hence, this model provides a satisfactory fit to the data and all of the effects included in it are significant at the $\alpha = .01$ level. Estimates for the various effects in (5.2.5) are

Overall Mean	0.424
Newspaper effect	0.109
Radio effect	0.039
Solid reading as modified effect	0.100
Lecture as modified effect	0.070
Solid reading x lecture interaction in newspaper non-readers and radio listeners	0.149

Predicted values for the conditional probabilities of good knowledge of cancer based on the model (5.2.5) are given in Column 10 of Table 5.2.1. It is apparent that this model has the advantage of not only accounting for the obvious complex effects of lectures but the more subtle ones due to solid reading.

Finally, some further justification for the objective of finding models for which the residual sum of squares does not exceed 3.84 may be seen in noting what happens when a simple no interaction model is fitted to the data; namely,

$$\underset{\sim}{X}_6 = \begin{bmatrix} 1 & 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 & -1 \\ 1 & 1 & 1 & -1 & 1 \\ 1 & 1 & 1 & -1 & -1 \\ 1 & 1 & -1 & 1 & 1 \\ 1 & 1 & -1 & 1 & -1 \\ 1 & 1 & -1 & -1 & 1 \\ 1 & 1 & -1 & -1 & -1 \\ 1 & -1 & 1 & 1 & 1 \\ 1 & -1 & 1 & 1 & -1 \\ 1 & -1 & 1 & -1 & 1 \\ 1 & -1 & 1 & -1 & -1 \\ 1 & -1 & -1 & 1 & 1 \\ 1 & -1 & -1 & 1 & -1 \\ 1 & -1 & -1 & -1 & 1 \\ 1 & -1 & -1 & -1 & -1 \end{bmatrix} \qquad (5.2.6)$$

The X^2-statistics for testing the effects in the model (5.2.6) are given in Table 5.2.7.

Table 5.2.7

Source of Variation	D.F.	X^2
Newspapers	1	34.37
Radio	1	6.08
Solid Reading	1	79.24
Lectures	1	4.34
Simultaneous effects of above sources	4	248.32
Residual	11	12.72

Estimates for the various effects in (5.2.6) are

Overall mean	0.436
Newspaper effect	0.072
Radio effect	0.033
Solid reading effect	0.111
Lecture effect	0.043

and predicted values based on these appear in column 11 of Table 5.2.1. The important point to note here is that the residual sum of squares is definitely not significant (α = .25) when compared with the chi-square

distribution with D.F. = 11. Moreover, from the discussion of the model (5.2.1) based on $\underset{\sim}{X}_1$, it is apparent that this residual definitely conceals the significant interactions which lecture effects have with newspaper and solid reading effects. One consequence of this is the presence of fairly large residuals for several of the factor combinations.

In summary, when fitting models to complex categorical data situations, some caution must be exercised in interpreting the residual sum of squares as a goodness of fit test. The safest criterion is to require that this residual sum of squares does not exceed 3.84 and hence, does not include any individually significant degrees of freedom. When this goal cannot be realistically achieved, then effort should be directed at finding a model for which the residual sum of squares is non-significant (when tested using the appropriate D.F.) and as small as possible. Further justification can be given to such models if there are no striking patterns in the differences between observed and predicted values. Finally, once a model is fitted, it is desirable that the test of significance on each separate effect is individually significant at the $\alpha = .05$ level. Otherwise, an attempt should be made to reduce the number of effects in the model by refining their definitions. This last guideline requires careful judgement because all tests of significance on the parameters of a given linear model are adjusted in a covariance sense for all effects which are not specified in the corresponding hypotheses. This important point follows from the fact that in a weighted analysis, orthogonal effects are no longer uncorrelated and hence overlap each

other in a considerably more subtle fashion than in the usual unweighted analysis of variance situation. In particular, one consequence of this which is readily visible is that in each of the tables 5.2.2-5.2.7, the simultaneous test of effects has an X^2-value which is considerably larger than the sum of the X^2-statistics for the individual component degrees of freedom. Other than this warning, essentially the only other advice that the researcher needs is that, since this type of analysis is based on linear models and weighted least squares, whatever types of analysis that can be obtained by using multiple regression procedures either in a stepwise fashion in the sense considered here or otherwise apply equally well for categorical data as for continuous data. A further discussion of this topic in the context of another example is given in Johnson and Koch [1970a].

5.3 A split-plot contingency table.

In this section, we shall consider an example of a categorical data model analogous to split-plot experiments in analysis of variance. These models have been discussed by Koch and Reinfurt [1970] and correspond to a "uni-factor, multi-response" situation with the following structure: Within each sub-population (whole plot factor), each member of the corresponding group of subjects is exposed to each of the d levels of some treatment (split-plot factor) with the resulting responses being classified into one of s levels. The resulting data may then be represented in an r x s x s x ... x s contingency table of (d+1) dimensions where r denotes the number of sub-populations (levels of the whole-plot factor) and s^d denotes the number of possible multivariate responses. The distinguishing feature of these models is that the d responses are measured in essentially

the same terms and factor-like comparisons across them are of interest. Specifically, the hypotheses of interest are

1. There are no differences among the sub-populations with respect to the distribution of the d-variate response vector (i.e., there are no whole-plot factor effects).

2. There are no differences among the effects of the d levels of the treatment in each sub-population (i.e., there are no split-plot factor effects).

3. There are no differences across sub-populations with respect to differences among the effects of the d levels of the treatment (i.e., there is no whole-plot x split-plot interaction).

Certain of these concepts become more clear if the s response categories may be quantitively scaled. In this case, hypotheses can be formulated in terms of mean scores for each sub-population x treatment combination (where the d mean scores within a given sub-population are correlated with one another since each is based on the same set of subjects). As a result, test statistics for hypotheses pertaining to treatment effects (split-plot factor) are analogous to Hotelling T^2 procedures while the others have more complex analogues in multivariate analysis of variance (see Cole and Grizzle[1966] and Koch [1969]).

One type of data which can be interpreted from the point of view of split-plot contingency table models are social mobility tables for different countries. In Table 5.3.1 , data which were recently described in Mosteller [1968] are presented. These data originally arose from Glass's British study and Svalastoga's Danish study of occupational mobility. The frequencies in the table correspond to the joint distributions of father's and son's occupation on a five point ordinal scale. These data may be regarded as arising from a "one factor, two response" split-plot contingency table model in

which country is the whole-plot factor and father vs. son is the split-plot factor.

Table 5.3.1

British Data		Son's Occupation Status 1	2	3	4	5	Total
Father's Occupation Status	1	50	45	8	18	9	129
	2	28	174	84	154	55	495
	3	11	78	110	223	96	518
	4	14	150	185	714	447	1510
	5	3	42	72	320	411	848
	Total	106	489	459	1429	1017	
Danish Data							
Father's Occupation Status	1	18	17	16	4	2	57
	2	24	105	109	59	21	318
	3	23	84	289	217	95	708
	4	8	49	175	348	198	778
	5	6	8	69	201	246	530
	Total	79	263	658	829	562	

On applying the $\underset{\sim}{A}$-matrix

$$\underset{\sim}{A} = \begin{bmatrix} 1\,1\,1\,1\,1 & 2\,2\,2\,2\,2 & 3\,3\,3\,3\,3 & 4\,4\,4\,4\,4 & 5\,5\,5\,5\,5 \\ 1\,2\,3\,4\,5 & 1\,2\,3\,4\,5 & 1\,2\,3\,4\,5 & 1\,2\,3\,4\,5 & 1\,2\,3\,4\,5 \end{bmatrix}$$

to the sub-table corresponding to each country arranged in a single row vector by placing the existing rows one after the other, mean status scores for father's and son's occupation within each country may be determined. The resulting estimates and the corresponding estimated variance-covariance matrix are given in Table 5.3.2.

Table 5.3.2

Country	Person	Mean Score	Covariance with Score for: Brit. Father	Brit. Son	Den. Father	Den. Son
Britain	Father	3.70	0.00034	0.00015		
Britain	Son	3.79	0.00015	0.00035		
Denmark	Father	3.59			0.00046	0.00022
Denmark	Son	3.64			0.00022	0.00047

where the "blank" covariances are zero because the data from the two separate countries involve different subjects and hence are independent of each other. If the linear model corresponding to the design matrix

$$\underset{\sim}{X} = \begin{bmatrix} 1 & 1 & 1 \\ 1 & 1 & -1 \\ 1 & -1 & 1 \\ 1 & -1 & -1 \end{bmatrix} \qquad (5.3.1)$$

is fitted to the mean scores in Table 5.3.2 by weighted least squares, test statistics for hypotheses pertaining to the effects of country and father vs. son and their interaction may be tested according to the methodology outlined in Section 3. The appropriate X^2-statistics are given in Table 5.3.3.

Table 5.3.3

Source of Variation	D.F.	X^2
Country	1	28.78
Father vs. Son	1	24.68
Residual: Interaction of Country and Father vs. Son	1	1.46

Hence, it is apparent that there is no interaction between the effects of country and father vs. son and that the mean occupation status score is significantly greater for Britain than for Denmark and significantly greater for sons than for fathers. Finally, the estimates for the parameters in the linear model defined by (5.3.1) are

Overall Mean = 3.680
Country Effect = 0.065
Father vs. Son Effect = 0.036

These estimates can be used to determine predicted values for mean occupation status scores for fathers and sons in the two countries.

In the preceding discussion, the data were viewed as a "one factor, two response" situation with comparisons of country effects and father vs. son effects being the topics of primary interest. Alternatively, if there is more interest in studying the way in which the mean status score for son's occupation depends on country and the corresponding level of father's occupation, it is more appropriate to

regard the data in terms of a "two factor, one response" model in which son's occupation is the response, and country and father's occupation are the factors. In this case, the $\underset{\sim}{A}$-matrix

$$\underset{\sim}{A} = [1 \quad 2 \quad 3 \quad 4 \quad 5]$$

is applied to each row of Table 5.3.1 (except the "total" rows) to determine mean scores for each country x father's occupation status combination. These scores together with their corresponding estimated variances are given in columns 3 and 4 respectively of Table 5.3.4. It can be noted that these mean scores are uncorrelated with one another since they are based on the conditional distributions of son's occupation status given father's occupation status. Three linear models were fitted to the set of mean scores for son's occupation status. The first of these was

$$\underset{\sim}{X}_1 = \begin{bmatrix} 1 & 1 & -2 & 4 & -8 & 16 \\ 1 & 1 & -1 & 1 & -1 & 1 \\ 1 & 1 & 0 & 0 & 0 & 0 \\ 1 & 1 & 1 & 1 & 1 & 1 \\ 1 & 1 & 2 & 4 & 8 & 16 \\ 1 & -1 & -2 & 4 & -8 & 16 \\ 1 & -1 & -1 & 1 & -1 & 1 \\ 1 & -1 & 0 & 0 & 0 & 0 \\ 1 & -1 & 1 & 1 & 1 & 1 \\ 1 & -1 & 2 & 4 & 8 & 16 \end{bmatrix} \qquad (5.3.2)$$

where the status scores for father's occupation are being treated, for simplicity, as lying on an equally spaced scale centered at the middle level of "3". The appropriate X^2-statistics for testing the effects in the model (5.3.2) are given in Table 5.3.5.

TABLE 5.3.4

Country	Father's Occupation Status	Mean Score for Son's Occupation Status	Estimated Variance	Predicted Value X_1	Predicted Value X_2	Predicted Value X_3
Britain	1	2.14	0.0120	2.20	2.33	2.28
Britain	2	3.07	0.0027	3.01	3.01	2.99
Britain	3	3.61	0.0020	3.54	3.56	3.56
Britain	4	3.95	0.00059	3.96	3.98	3.99
Britain	5	4.29	0.00084	4.32	4.27	4.27
Denmark	1	2.21	0.0202	2.10	2.20	2.27
Denmark	2	2.84	0.0033	2.91	2.84	2.86
Denmark	3	3.39	0.0013	3.44	3.39	3.39
Denmark	4	3.87	0.0010	3.86	3.87	3.86
Denmark	5	4.27	0.0013	4.22	4.27	4.27

Table 5.3.5

Source of Variation	D.F.	X^2
Country	1	15.53
Father's Occupation Status	4	1[illegible]6.94
Residual: Interaction of Country with Father's Occupation Status	[illegible]	[illegible]1.31

Since the interaction of the two sources of variation is significant, ($\alpha = .05$), the model corresponding to $\underset{\sim}{X}_1$ does not quite appropriately fit the data although it does explain a substantial proportion of the variation. The second model considered was

$$\underset{\sim}{X}_2 = \begin{bmatrix} 1 & 1 & -2 & 4 & -2 & 4 \\ 1 & 1 & -1 & 1 & -1 & 1 \\ 1 & 1 & 0 & 0 & 0 & 0 \\ 1 & 1 & 1 & 1 & 1 & 1 \\ 1 & 1 & 2 & 4 & 2 & 4 \\ 1 & -1 & -2 & 4 & 2 & -4 \\ 1 & -1 & -1 & 1 & 1 & -1 \\ 1 & -1 & 0 & 0 & 0 & 0 \\ 1 & -1 & 1 & 1 & -1 & -1 \\ 1 & -1 & 2 & 4 & -2 & -4 \end{bmatrix} \qquad (5.3.3)$$

Only linear and quadratic components of father's occupation status together with their interactions with country were included in (5.3.3) because for the model $\underset{\sim}{X}_1$ in (5.3.2) a test of significance for the cubic and quartic components had $X^2 = 3.51$ with D.F. = 2. From this result and the fact that such effects do not have any real practical interpretation for such data showing a monotone relationship, neither the second nor the third models considered here will include cubic or quartic terms. The appropriate X^2-statistics for testing the effects in the model (5.3.3) are given in Table 5.3.6. Hence, the model does fit the data since the residual is not significant ($\alpha = .05$).

Table 5.3.6

Source of Variation	D.F.	X^2
Country	1	20.10
Linear Component of Father's Status	1	1005.48
Quadratic Component of Father's Status	1	25.62
Country x Linear Component Interaction	1	0.93
Country x Quadratic Component Interaction	1	1.65
Residual	4	7.78

In addition, it appears as if neither interaction term is significant at the $\alpha = .05$ level. However, it must be borne in mind that with weighted least squares, apparently orthogonal effects are no longer uncorrelated. Hence, all tests reflect the contribution of a given effect conditional on all other effects being in the model. Thus since the model based on $\underset{\sim}{X}_1$ told us that interaction was present, the primary conclusion from Table 5.3.6 is that this interaction can be explained in terms of essentially one component, i.e., either the country x linear component or the country x quadratic component. The estimates of the parameters in the model $\underset{\sim}{X}_2$ defined by (5.3.3) are

Overall Mean	3.476
Country Effect	0.083
Linear Component of Father's Status	0.500
Quadratic Component of Father's Status	-0.052
Country x Linear Component Interaction	-0.015
Country x Quadratic Component Interaction	-0.013

Since the country x quadratic component interaction has a larger relative magnitude with respect to the quadratic component main effect than the country x linear component interaction has relative to the linear component main effect, the third model considered is equivalent to that defined by $\underset{\sim}{X}_2$ with the country x linear component interaction eliminated. This model was

$$\underset{\sim}{X}_3 = \begin{bmatrix} 1 & 1 & -2 & 4 & 0 \\ 1 & 1 & -1 & 1 & 0 \\ 1 & 1 & 0 & 0 & 0 \\ 1 & 1 & 1 & 1 & 0 \\ 1 & 1 & 2 & 4 & 0 \\ 1 & -1 & -2 & 0 & 4 \\ 1 & -1 & -1 & 0 & 1 \\ 1 & -1 & 0 & 0 & 0 \\ 1 & -1 & 1 & 0 & 1 \\ 1 & -1 & 2 & 0 & 4 \end{bmatrix} \qquad (5.3.4)$$

The appropriate X^2-statistics for testing the effects in the model (5.3.4) are given in Table 5.3.7. First it can be noted that the model does fit the data since the residual is not significant (for $\alpha = .05$). In addition, every effect is individually significant ($\alpha = .05$). Finally, the hypothesis of equality of the two separate

Table 5.3.7

Source of Variation	D.F.	X^2
Country	1	19.49
Linear Component of Father's Status	1	1033.90
Quadratic Component of Father's Status in Britain	1	32.31
Quadratic Component of Father's Status in Denmark	1	5.21
Residual	5	8.71

quadratic effects can be tested with the resulting $X^2 = 6.11$ where D.F. = 1 implying that there is an interaction here which is significant ($\alpha = .05$). The estimates of the effects in the model specified by (5.3.4) are

Overall Mean	3.475
Country Effect	0.081
Linear Component of Father's Status	0.500
Quadratic Component of Father's Status in Britain	-0.070
Quadratic Component of Father's Status in Denmark	-0.031

These estimates to some extent show the somewhat different relationships between son's occupation status and father's occupation status in the two countries. Finally, predicted values for the mean scores for son's occupation status for each country x father's occupation status combination are given for the linear models based on $\underset{\sim}{X}_1$, $\underset{\sim}{X}_2$, $\underset{\sim}{X}_3$ in columns 5, 6, 7, respectively, of Table 5.3.4. In comparing observed and predicted values, it should be noted that the criteria for weighted analyses forces the residuals to be smallest where the sample sizes are the largest.

5.4 Drug Example with Supplemented Margins

This is an hypothetical example of a categorical data model with supplemental information on both margins. Suppose a certain pharmaceutical company wished to compare the effects of drugs A and B. Of the 80 patients available for study, 50 patients (Group I) received both drugs under essentially the same conditions including a suitable time delay between applications (with the order of presentation of A and B random); 16 patients (Group II) received drug A only; and 14 patients (Group III) received drug B only. In each case, the patient's response to the particular drug was observed and recorded as "none", "slight", or "strong". The results are given in Table 5.4.1. The main question of interest is whether the distribution of strength of response is the same for both drugs.

Table 5.4.1

[Group I] Response to A	Response to B None (1)	Slight (2)	Strong (3)	Sub-total	Response to A only [Group II]	Total
None (1)	4	6	0	10	4	14
Slight (2)	4	8	9	21	7	28
Strong (3)	0	5	14	19	5	24
Sub-total	8	19	23	50	16	66
Response to B only [Group III]	3	4	7	14		
Total	11	23	30	64		80

Within this framework, let

$$\underset{\sim}{\pi}' = (\pi_{11}, \pi_{12}, \pi_{13}, \pi_{21}, \pi_{22}, \pi_{23}, \pi_{31}, \pi_{32}, \pi_{33})$$

$$\underset{\sim}{n}_G' = (\underset{\sim}{n}', \underset{\sim}{n}'_{(B)}, \underset{\sim}{n}'_{(A)}) = (4,6,0,4,8,9,0,5,14;4,7,5;3,4,7)$$

For this example, it appears reasonable to assign scores to the three levels of the two responses and to test for equality of mean scores over the two marginals. Namely

$$H_0: \alpha_A = \alpha_B \qquad (5.4.1)$$

where α_A and α_B are the mean scores for the responses to drugs A and B, respectively. Therefore, assigning scores $a_1 = 1$, $a_2 = 2$, and $a_3 = 3$ to the response categories for both drugs A and B and using initial

relative frequencies ($\hat{\underset{\sim}{\pi}}_G$) in place of the observation vector ($\underset{\sim}{n}_G$), the first stage estimation involves the process referred to as part I.2(c) of the first stage. Specifically, weighted least squares estimators are obtained using the model

$$E(\underset{4\times15}{\underset{\sim}{A}_1}\ \underset{15\times1}{\hat{\underset{\sim}{\pi}}_G}) = \underset{4\times2}{\underset{\sim}{X}_1}\ \underset{2\times1}{\underset{\sim}{\alpha}} \tag{5.4.2}$$

where

$$\hat{\underset{\sim}{\pi}}_G' = (\hat{\underset{\sim}{\pi}}', \hat{\underset{\sim}{\pi}}_{.o}', \hat{\underset{\sim}{\pi}}_{o.}') \tag{5.4.3}$$

with

$$\hat{\underset{\sim}{\pi}}' = \frac{1}{50}\,(4,6,0,4,8,9,0,5,14)$$

$$\hat{\underset{\sim}{\pi}}_{.o}' = \frac{1}{16}\,(4,7,5)$$

$$\hat{\underset{\sim}{\pi}}_{o.}' = \frac{1}{14}\,(3,4,7)$$

$$\underset{\sim}{\alpha}' = (\alpha_A, \alpha_B)$$

$$\underset{4\times15}{\underset{\sim}{A}_1} = \begin{bmatrix} 1 & 1 & 1 & 2 & 2 & 2 & 3 & 3 & 3 & 0 & 0 & 0 & 0 & 0 & 0 \\ 1 & 2 & 3 & 1 & 2 & 3 & 1 & 2 & 3 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 2 & 3 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 2 & 3 \end{bmatrix} \tag{5.4.4}$$

$$\underset{\sim}{X}_1 = \begin{bmatrix} 1 & 0 \\ 0 & 1 \\ 1 & 0 \\ 0 & 1 \end{bmatrix} \tag{5.4.5}$$

The inverse ($\hat{\underset{\sim}{V}}$) of the weight matrix is given by $\underset{\sim}{A}_1 \hat{\underset{\sim}{V}}_{\pi_G} \underset{\sim}{A}_1'$ where

$$\underset{15\times15}{\hat{V}_{\pi_G}} = \begin{bmatrix} \hat{V}_{(\ )_\pi} & & 0 \\ & \hat{V}_{(B)_\pi} & \\ 0 & & \hat{V}_{(A)_\pi} \end{bmatrix} \qquad (5.4.6)$$

with

$$\underset{9\times9}{\hat{V}_{(\ )_\pi}} = \frac{1}{n_{oo}} (D_\pi - \hat{\pi}\hat{\pi}')$$

$$\underset{3\times3}{\hat{V}_{(B)_\pi}} = \frac{1}{n_{o*}} (D_{\hat{\pi}_{.o}} - \hat{\pi}_{.o}\hat{\pi}_{.o}')$$

$$\underset{3\times3}{\hat{V}_{(A)_\pi}} = \frac{1}{n_{*o}} (D_{\hat{\pi}_{o.}} - \hat{\pi}_{o.}\hat{\pi}_{o.}')$$

Finally, a test with D.F. = 1 for (5.4.1) is obtained (as in Grizzle et al. [1969]) using $\hat{\alpha}_A$ and $\hat{\alpha}_B$ from the first stage with the contrast matrix, $C = (1, -1)$. Alternatively, if the linear model

$$E(\hat{\alpha}) = X_2\mu \quad \text{where } X_2 = \begin{bmatrix} 1 \\ 1 \end{bmatrix} \qquad (5.4.7)$$

is fitted using weighted least squares, the residual sum of squares provides the required test statistic with D.F. = 1.

It should be noted that, for the test of equality of mean scores over the marginals, empty cells are no problem since, as long as the marginal totals are non-zero, the required estimate of the covariance matrix is non-singular.

The usual format of the linear model approach can be modified to produce this analysis as follows. First the data are considered

as having arisen from three populations (Groups I, II, and III). Thus, the following matrices comprise the basic input.

$$\underset{3\times 9}{\underset{\sim}{n}} = \begin{bmatrix} 4 & 6 & 0 & 4 & 8 & 9 & 0 & 5 & 14 \\ 4 & 7 & 5 & 0 & 0 & 0 & 0 & 0 & 0 \\ 3 & 0 & 0 & 4 & 0 & 0 & 7 & 0 & 0 \end{bmatrix}$$

$$\underset{4\times 27}{\underset{\sim}{A}_2} = \begin{bmatrix} 1 & 1 & 1 & 2 & 2 & 2 & 3 & 3 & 3 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 1 & 2 & 3 & 1 & 2 & 3 & 1 & 2 & 3 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 2 & 3 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 2 & 0 & 0 & 3 & 0 & 0 \end{bmatrix}$$

The model as given by (5.4.2) fits the data (i.e., the test of fit gives $X_F^2 = 0.30$ with D.F. = 2). To obtain the test in (5.4.1), namely, H_0: $\alpha_A = \alpha_B$, use $\underset{\sim}{X}_1$ as given by (5.4.5) and $\underset{\sim}{C} = (1, -1)$. The corresponding test statistic with D.F. = 1 is found to be $X_M^2 = 2.1$ indicating that the mean scores for drugs A and B are the same.

5.5 A sample survey situation

In this section, we shall briefly consider an example of categorical data obtained from a sample survey situation. These data arose from a survey of drivers born in 1950 and licensed in North Carolina in 1969. The sample design was stratified with the three strata and the corresponding sampling rates being

Stratum	Sample Size	Sampling Rate
White Male	1034	(1/30)
White Female	994	(1/25)
Non-White Male	1300	(1/8)

For all drivers in the sample, information was obtained from his driver record pertaining to numbers of violations and accidents. In

addition, the drivers were sent a questionnaire concerned with their exposure to risk (e.g., mileage, night driving, etc.) which they either returned immediately, after follow-up, or not at all. Hence, there is interest in whether those who returned the questionnaire tended to have better driver records than those who did not. As a result, the data in Table 5.5.1 indicate the joint distribution of the responses pertaining to questionnaire return or not and at least one accident or not for each of the three strata.

Table 5.5.1

	Quest. Returned		Quest. Not Returned	
	No Acc.	1+ Acc.	No Acc.	1+ Acc.
White Males	429	194	269	142
White Females	529	111	303	51
Non-White Males	535	165	460	140

In this analysis, attention is focused on the proportion of no accident drivers within each questionnaire return status x stratum combination. The estimates of these proportions are obtained in the usual way and are given in Table 5.5.2. However, the variances are adjusted in the sense of being multiplied by the appropriate finite population correction factor (i.e., by $(1-f_i)$ where f_i is the sampling rate). These are also given in Table 5.5.2.

Table 5.5.2

Stratum	Quest. Returned	Prop. No. Acc.	Unadj. Est. Variance	Finite Pop. Corr.	Adj. Est. Variance
White Males	yes	0.69	0.000344	(29/30)	0.000333
White Males	no	0.65	0.000550	(29/30)	0.000532
White Females	yes	0.83	0.000224	(24/25)	0.000215
White Females	no	0.86	0.000348	(24/25)	0.000334
Non-White Males	yes	0.76	0.000257	(7/8)	0.000225
Non-White Males	no	0.77	0.000298	(7/8)	0.000261

Once the estimated variances have been adjusted for finite population sampling, the usual analysis applies. Hence, the linear model

$$\underset{\sim}{X} = \begin{bmatrix} 1 & 1 & 0 & 1 \\ 1 & 1 & 0 & -1 \\ 1 & 0 & 1 & 1 \\ 1 & 0 & 1 & -1 \\ 1 & -1 & -1 & 1 \\ 1 & -1 & -1 & -1 \end{bmatrix}$$

was fitted to the data. Test statistics for the effects in $\underset{\sim}{X}$ appear in Table 5.5.3.

Table 5.5.3

Source of Variation	D.F.	X^2
Strata	2	78.61
Questionnaire Returned or Not	1	0.08
Residual: Interaction	2	2.98

Hence, it appear that there are little if any differences among those drivers who returned the questionnaire and those who did not with respect to the proportion of no accidents. There are definite differences among the strata.

References

Bartlett, M. S. [1935]. Contingency table interactions. Journal of the Royal Statistical Society Supplement 2, pp. 248-52.

Berkson, J. [1944]. Application of the logistic function to bio-assay. Journal of the American Statistical Association, 39, pp. 357-65.

Berkson, J. [1953]. A statistically precise and relatively simple method of estimating the bio-assay with quantal response based on the logistic function. Journal of the American Statistical Association, 48, pp. 565-99.

Berkson, J. [1954]. Estimation by least squares and by maximum likelihood. Pp. 1-11 in: Proceedings of the Third Berkeley Symposium on Mathematical Statistics and Probability, 1. University of California Press, Berkeley and Los Angeles.

Berkson, J. [1955]. Maximum likelihood and minimum χ^2 estimates of the logistic function. Journal of the American Statistical Association, 50, pp. 130-62.

Berkson, J. [1957]. Tables for the use in estimating the normal distribution function by normit analysis. Part I. Description and use of tables. Part III. Comparison between minimum normit analysis estimate and the maximum likelihood estimate. Biometrika, 44, pp. 411-35.

Berkson, J. [1968]. Application of minimum logit χ^2 estimate to a problem of Grizzle with a notation on the problem of no interaction. Biometrics, 24, pp. 75-95.

Bhapkar, V. P. [1961]. Some tests for categorical data. Annals of Mathematical Statistics, 32, pp. 72-83.

Bhapkar, V. P. [1966]. A note on the equivalence of two test criteria for hypotheses in categorical data. Journal of the American Statistical Association, 61, pp. 228-35.

Bhapkar, V. P. [1968]. On the analysis of contingency tables with a quantitative response. Biometrics, 24, pp. 329-38.

Bhapkar, V. P. and Koch, G. G. [1968a]. Hypotheses of 'no interaction' in multi-dimensional contingency tables. Technometrics, 10, pp. 107-23.

Bhapkar, V. P. and Koch, G. G. [1968b]. On the hypotheses of 'no interaction' in contingency tables. Biometrics, 24, pp. 567-94.

Birch, M. W. [1963]. Maximum likelihood in three-way contingency tables. Journal of the Royal Statistical Society, Series B, 25, pp. 220-33.

Bishop, Y. M. M. [1969]. Full contingency tables, logits, and split contingency tables. Biometrics, 25, pp. 383-99.

Bishop, Y. M. M. and Fienberg, S. E. [1969]. Incomplete two-dimensional contingency tables. Biometrics, 25, pp. 119-28.

Blumenthal, S. [1968]. Multinomial sampling with partially categorized data. Journal of the American Statistical Association, 63, pp. 542-51.

Cochran, W. G. [1952]. The χ^2 test of goodness of fit. Annals of Mathematical Statistics, 23, pp. 315-45.

Cochran, W. G. [1963]. Sampling Techniques. Wiley, New York. Second Edition.

Cole, J. W. L. and Grizzle, J. E. [1966]. Applications of multivariate analysis of variance to repeated measurements experiments. Biometrics, 22, pp. 810-27.

Cox, D. R. and Snell E. J. [1968]. A general definition of residuals. Journal of the Royal Statistical Society, Series B, 30, pp. 248-65.

Darroch, J. N. [1962]. Interactions in multi-factor contingency tables. Journal of the Royal Statistical Society, Series B, 24, pp. 251-63.

Dyke, G. V. and Patterson, H. D. [1952]. Analysis of factorial arrangements when the data are proportions. Biometrics, 8, pp. 1-12.

Fienberg, S. E. [1969]. Preliminary graphical analysis and quasi-independence for two-way contingency tables. Applied Statistics, 18, pp. 153-68.

Finney, D. J. [1964]. Statistical Method in Biological Assay. Charles Griffin & Company Limited, London, Second Edition.

Forthofer, R. N., Starmer, C. F. and Grizzle, J. E. [1969]. A program for the analysis of categorical data by linear models. University of North Carolina Institute of Statistics Mimeo Series No. 604. January, 1969.

Good, I. J. [1963]. Maximum entropy for hypothesis formulation, especially for multidimensional contingency tables. Annals of Mathematical Statistics, 34, pp. 911-34.

Goodman, L. A. [1963a]. On Plackett's test for contingency table interactions. Journal of the Royal Statistical Society, Series B, 25, pp. 179-88.

Goodman, L. A. [1963b]. On methods for comparing contingency tables. Journal of the Royal Statistical Society, Series A, 126, pp. 94-108.

Goodman, L. A. [1964a]. Interactions in multidimensional contingency tables. Annals of Mathematical Statistics, 35, pp. 632-46.

Goodman, L. A. [1964b]. Simple methods for analysing three-factor interaction in contingency tables. Journal of the American Statistical Association, 59, pp. 319-52.

Goodman, L. A. [1968]. The analysis of cross-classified data: Independence, quasi-independence, and interactions in contingency tables with or without missing entries. Journal of the American Statistical Association, 63, pp. 1091-131.

Goodman, L. A. [1969]. The analysis of multidimensional contingency tables: Stepwise procedures and direct estimation methods for building models for multiple classifications. To appear in Technometrics.

Goodman, L. A. [1970]. The multivariate analysis of qualitative data: Interactions among multiple classifications. Journal of the American Statistical Association, 65, pp, 226-56.

Grizzle, J. E. [1961]. A new method of testing hypotheses and estimating parameters for the logistic model. Biometrics, 17, pp. 372-85.

Grizzle, J. E. Starmer, C. F. and Koch, G. G. [1969]. Analysis of categorical data by linear models. Biometrics, 25, pp. 489-504.

Grizzle, J. E. and Williams, O. D. Multivariate analysis of categorized data by linear models. Presented at the international meeting of the Biometric Society, Hanover, Germany, August, 1970.

Hoeffding, W. [1965]. Asymptotically optimal tests for multinomial distributions. Annals of Mathematical Statistics, 36, pp. 369-401.

Ireland, C. T. and Kullback, S. [1968a]. Contingency tables with given marginals. Biometrika, 55, pp. 179-88.

Ireland, C. T. and Kullback, S. [1968b]. Minimum discrimination information estimation. Biometrics, 24, pp. 707-13.

Ireland, C. T., Ku, H. H. and Kullback, S. [1969]. Symmetry and marginal homogeneity of an r×r contingency table. Journal of the American Statistical Association, 64, pp. 1323-41.

Johnson, W. D. and Koch, G. G. [1970a]. A note on the weighted least squares analysis of the Ries-Smith contingency table data. Submitted to Technometrics.

Johnson, W. D. and Koch, G. G. [1970b]. Some aspects of the statistical analysis of qualitative data from sample surveys. Part I: Proportions, mean scores, and other linear functions. Submitted to Health Services Research.

Kastenbaum, M. A. and Lamphiear, D. E. [1959]. Calculation of chi-square to test the no three-factor interaction hypothesis. Biometrics, 15, pp. 107-15.

Kleinbaum, D. G. [1970]. A general method for obtaining test criteria for multivariate linear models with more than one design matrix and/or incomplete in response variates. Ph.D. thesis, Department of Statistics, University of North Carolina at Chapel Hill. University Microfilms, Inc., Ann Arbor, Michigan.

Koch, G. G. [1969]. Some aspects of the statistical analysis of 'split plot' experiments in completely randomized layouts. Journal of the American Statistical Association, 64, pp. 485-505.

Koch, G. G. and Reinfurt, D. W. [1970]. The analysis of categorical data from mixed models. To appear in Biometrics.

Ku, H. H. and Kullback, S. [1968]. Interaction in multidimensional contingency tables: An information theoretic approach. Journal of Research of the National Bureau of Standards - Mathematical Sciences, 72B, pp. 159-99.

Kullback, S., Kupperman, M., and Ku, H. H. [1962]. Tests for contingency tables and Markov chains. Technometrics, 4, pp. 573-608.

Kullback, S. [1959]. Information Theory and Statistics. Wiley, New York.

Lancaster, H. O. [1951]. Complex contingency tables treated by the partition of χ^2. Journal of the Royal Statistical Society B, Series B, 13, pp. 242-49.

Lewis, B. N. [1962]. On the analysis of interaction in multidimensional contigency tables. Journal of the Royal Statistical Society, Series A, 125, pp. 88-117.

Lombard, H. L. and Doering, C. R. [1947]. Treatment of the four-fold table by partial correlation as it relates to public health problems. Biometrics, 3, pp. 123-28.

Maxwell, A. E. [1961]. Analyzing Qualitative Data. Methuen, London.

Mosteller, F. [1968]. Association and estimation in contingency tables. Journal of the American Statistical Association, 63, pp. 1-28.

Nathan, G. [1969]. Tests of independence in contingency tables from stratified samples. Pp. 578-600 in: New Developments in Survey Sampling (N. L. Johnson and H. Smith, Jr. - editors). Wiley, New York.

Nathan, G. [1970]. Approximate tests of independence in contingency tables from complex stratified cluster samples. Unpublished manuscript.

Neyman, J. [1949]. Contribution to the theory of the χ^2 test. Pp. 239-73 in: Proceedings of the Berkeley Symposium on Mathematical Statistics and Probability, University of California Press, Berkeley and Los Angeles.

Pearson, E. S. [1947]. The choice of statistical tests illustrated on the interpretation of data classed in a 2×2 table. Biometrika, 34, pp. 139-67.

Plackett, R. L. [1962]. A note on interactions in contingency tables. Journal of the Royal Statistical Society, Series B, 24, pp. 162-66.

Plackett, R. L. Multidimensional contingency tables: A survey of models and methods. Presented at the meeting of ISI in London, England, September, 1969.

Rao, C. R. [1963]. Criteria of estimation in large samples. Pp. 345-62 in: Contributions to Statistics. Pergamon Press, New York.

Reinfurt, D. W. [1970]. The analysis of categorical data with supplemented margins including applications to mixed models. Ph.D. Thesis, North Carolina State University at Raleigh. University Microfilms, Inc., Ann Arbor, Michigan.

Roy, S. N. and Bhapkar, V. P. [1960]. Some nonparametric analogs of normal ANOVA, MANOVA, and of studies in normal association. pp. 371-87 in: Contributions to Probability and Statistics. Stanford University Press, Stanford, California.

Roy, S. N. and Kastenbaum, M. A. [1956]. On the hypothesis of no interaction in a multiway contingency table. Annals of Mathematical Statistics, 27, pp. 749-57.

Roy, S. N. and Mitra, S. K. [1956]. An introduction to some non-parametric generalizations of analysis of variance and multivariate analysis. Biometrika, 43, pp. 361-76.

Simpson, E. H. [1951]. The interpretation of interaction in contingency tables. Journal of the Royal Statistical Society, Series B,13,pp.238-41.

Wald, A. [1943]. Tests of statistical hypotheses concerning several parameters when the number of observations is large. Transactions of the American Mathematical Society, 54, pp. 426-82.

ESTIMATION IN TRUNCATED POISSON DISTRIBUTIONS WITH CONCOMITANT EXPOSURE INTERVALS AND TRUNCATION POINTS

A. Clifford Cohen

The University of Georgia

1. INTRODUCTION

Estimation in the truncated Poisson distribution has previously been considered by numerous writers including Bliss [1], David and Johnson [4], Moore [6], Plackett [7], Subrahmanian [8], the writer [2,3], and others. Our concern here is with estimation of the Poisson parameter when the $i^{\underline{th}}$ observation of the discrete random variable X extends over a specified exposure interval of magnitude t_i, subject to the restriction that $x_i \geq c_i$, where c_i is a specified truncation point. More specifically, our attention is directed toward maximum likelihood estimation based on a random sample consisting of n independent observations $\{x_i, t_i, c_i\}$, where x_i is the number of occurrences of the event of interest observed during the $i^{\underline{th}}$ interval of exposure, while t_i and c_i are corresponding concomitant values. Of course, x_i and c_i must be non-negative integers, but t_i is permitted to assume any positive value. Accident data in which reports are rendered for time intervals of varying size with truncation as stipulated fall into the category of samples under consideration here. Inspection data in which defects per unit are reported for units of differing sizes might also result in samples of the type considered here.

With concomitant values t_i and c_i given, the applicable Poisson probability function may be written as

$$\Pr(X=x_i \mid t_i, c_i) = \begin{cases} \dfrac{e^{-\lambda t_i}(\lambda t_i)^{x_i}}{x_i!\,\bar{F}(c_i)}, & x_i = c_i, c_i+1, \ldots, \\ 0, & \text{elsewhere,} \end{cases} \tag{1}$$

where

$$\bar{F}(c_i) = \sum_{x_i=c_i}^{\infty} \frac{e^{-\lambda t_i}(\lambda t_i)^{x_i}}{x_i!} = 1 - F(c_i - 1) . \tag{2}$$

2. MAXIMUM LIKELIHOOD ESTIMATION OF λ

The likelihood function for a sample consisting of the n observed values $\{x_i, t_i, c_i\}$ for which the probability function (1) is applicable, follows as

$$L = \prod_{i=1}^{n} \frac{e^{-\lambda t_i}(\lambda t_i)^{x_i}}{x_i!\,\bar{F}(c_i)} . \tag{3}$$

On taking logarithms of both sides of (3) and differentiating with respect to λ, we obtain

$$\frac{\partial \ln L}{\partial \lambda} = -\sum_{1}^{n} t_i + \frac{1}{\lambda} \sum_{1}^{n} x_i - \sum_{1}^{n} \frac{1}{\bar{F}(c_i)} \frac{\partial \bar{F}(c_i)}{\partial \lambda} . \tag{4}$$

It follows from (2) that

$$\frac{\partial \bar{F}(c_i)}{\partial \lambda} = \frac{t_i e^{-\lambda t_i} (\lambda t_i)^{c_i - 1}}{(c_i - 1)!} = t_i f(c_i - 1) . \tag{5}$$

When (5) is substituted into (4) and the resulting expression is equated to zero and simplified, we obtain the estimating equation

$$\sum_{1}^{n} x_i = \lambda \left[\sum_{1}^{n} t_i + \sum_{1}^{n} t_i f(c_i - 1)/\bar{F}(c_i) \right] , \tag{6}$$

in which $f(c_i - 1)$ and $\bar{F}(c_i)$ are defined by (5) and (2) respectively.

With the aid of an ordinary table of individual and cumulative values of the Poisson function such as those of Molina [5], equation (6) can be solved with relative ease for the required estimate λ, using trial and error techniques. Once two values λ_1 and λ_2 have been found in a sufficiently narrow interval such that $G(\lambda_1) < \sum_{1}^{n} x_i < G(\lambda_2)$, where $G(\lambda)$ has been written for the right side of (6), the required estimate λ can be obtained by interpolating linearly between λ_1 and λ_2. Of course, standard iterative procedures

might also be employed in the calculation of $\hat{\lambda}$.

3. SOME SPECIAL CASES

By appropriately restricting the values of t_i and c_i in (6), we obtain results that apply in various cases of special interest. Certain of these specialized results are, of course, quite well known.

No Truncation — Unequal Intervals

In this case, $c_i = 0$ for all i, and equation (6) leads to the estimator

$$\hat{\lambda} = \sum_{1}^{n} x_i / \sum_{1}^{n} t_i \ . \tag{7}$$

No Truncation — Equal Intervals

If in addition to the restriction, $c_i = 0$ for all i, we impose the further restriction that $t_i = t$ for all i, the applicable estimator reduces to

$$\hat{\lambda} = \bar{x}/t \ , \tag{8}$$

where $\bar{x} = \sum_{1}^{n} x_i/n$. When $t = 1$, as in the usual case of estimation in the Poisson distribution, then (7) leads to the familiar estimator $\hat{\lambda} = \bar{x}$.

Only Zero Class Truncated In Making Each Observation — Unequal Intervals

In this case $c_i = 1$ for each i, and the estimating equation (6) reduces to

$$\sum_{1}^{n} x_i = \sum_{1}^{n} [t_i/(1-e^{-\hat{\lambda} t_i})] \ , \tag{9}$$

a result which was obtained independently by one of the writer's graduate assistants, Mr. S. S. Srivastava.

Only Zero Class Truncated In Making Each Observation – Equal Intervals

When we further specialize the preceeding result by requiring that $t_i = t$ for all i, then (9) becomes

$$\bar{x} = \lambda t/[1-e^{-\lambda t}] \ . \tag{10}$$

If we specialize still further by requiring that $t = 1$, then (10) reduces to

$$\bar{x} = \lambda/[1-e^{-\lambda}] \ , \tag{11}$$

as previously given in [3], where a table of the function on the right side of (11) is provided in order to facilitate solution for the required estimate λ.

4. VARIANCE OF ESTIMATES

The asymptotic variance of $\hat{\lambda}$ can be expressed as

$$V(\hat{\lambda}) = -\left[E\left(\frac{\partial^2 \ln L}{\partial \lambda^2}\right)\right]^{-1} \doteq -\left[\frac{\partial^2 \ln L}{\partial \lambda^2}\right]^{-1}_{\lambda=\hat{\lambda}} \tag{12}$$

for each of the cases considered here. The second derivative in the most general case under consideration follows from (4) as

$$\frac{\partial^2 \ln L}{\partial \lambda^2} = -\frac{\sum_1^n x_i}{\lambda^2} - \sum_1^n t_i \left[\frac{f(c_i-2) - f(c_i-1)}{\bar{F}(c_i)}\right]' + \sum_1^n \left[\frac{t_i f(c_i-1)}{\bar{F}(c_i)}\right]^2 \ , \tag{13}$$

a result which was previously given in [2]. With the aid of tables such as those of Molina [5], this derivative can easily be evaluated once (6) has been solved for $\hat{\lambda}$. The variance then follows from (12).

5. AN ILLUSTRATIVE EXAMPLE

In order to illustrate the practical application of estimators derived here, we consider inspection results on the number of surface defects per unit found in products of differing surface areas but which were manufactured on the same production line. In some of the inspections, no truncation takes place. In others, the zeros are eliminated, and in still others, both zeros and ones are eliminated. Based on the total sample which results from combining the several sub-samples involved, we proceed to estimate the Poisson parameter λ which characterizes the underlying common production process, and to determine the asymptotic variance of this estimate. The total sample available for this purpose consists of inspection results on 422 separate manufactured units. These data are presented in Table 1, with the total sample broken down into eight sub-samples having values of (t,c) as follows: (1,0),(1.5,1),(2,1),(2.5,1),(3,1),(1,2),(2,2), and (3,2). The value t = 1 indicates a surface area of one unit, while t = 2 indicates a surface area of two units, etc. For the total number of defects found in the 422 units inspected, we have $\sum_{1}^{422} x_i = 1597$. For the sum of the t's, we have $\sum_{1}^{422} t_i = 100(1)$ + 94(1.5) + 50(2) + 32(2.5) +28(3) + 30(1) + 34(2) + 54(3) = 767. These values are substituted into (6) which must then be solved for

TABLE 1

NUMBER OF SURFACE DEFECTS FOUND IN

FINAL INSPECTION OF 422 MANUFACTURED UNITS

No. Defects	Frequencies n_x								Row Totals
	c=0	c=1				c=2			
x	t=1	t=1.5	t=2	t=2.5	t=3	t=1	t=2	t=3	
0	13								13
1	28	16	5	1	0				50
2	27	21	6	3	1	14	6	3	81
3	17	22	9	5	3	9	5	3	73
4	10	16	11	6	5	5	7	7	67
5	3	11	9	5	4	1	6	7	46
6	2	6	4	5	5	1	5	9	37
7		1	3	4	4		3	8	23
8		1	-	2	3		1	6	15
9			1	1	1		1	5	9
10					2			3	5
11								2	2
12								1	1
Column Totals	100	94	50	32	28	30	34	54	422
Exn_x	200	294	203	154	162	86	153	345	1597

the required estimate $\hat{\lambda}$. As a first approximation, we try $\lambda_1 = 2.0000$, which comes from the sub-sample of size 100 with t = 1 and c = 6. In general, satisfactory first approximations might be obtained from any one or a combination of the available sub-samples. In some circumstances, a judicious guess might even suffice. We subsequently arrive at the final estimate $\hat{\lambda} = 2.005$, by linear interpolation as follows

λ		$G(\lambda)$
2.100		1663.72
2.005	$\sum_{1}^{422} x_i$	= 1597.00
2.000		1593.22

where $G(\lambda)$ has been written for the right side of the estimating equation (6). For the variance of the above estimate, we employ (12) and (13) to calculate $V(\hat{\lambda}) = 0.0028$.

It is of interest to note that the upper bound on λ as given by (6a) for this example becomes

$$\lambda_{UB} = 1597/765 = 2.0876 ,$$

a result which further justifies the selection of $\lambda_1 = 2.000$ as our first approximation.

REFERENCES

[1] Bliss, C. I., (1948). Estimation of the mean and its error from incomplete Poisson distributions, Connecticut Agricultural Experiment Station Bulletin 513, 12 p.

[2] Cohen, A. C., Jr. (1954). Estimation of the Poisson parameter from truncated samples and from censored samples, J. Am. Stat. Assn. 49, 158-68.

[3] Cohen, A. C., Jr. (1960). Estimating the parameter in a conditional Poisson distribution, Biometrics 16, 203-11.

[4] David, F. N., and Johnson, N. L., (1952). The truncated Poisson, Biometrics 8, 275-85.

[5] Molina, E. C., (1942). Poisson's Exponential Binomial Limit, D. Van Nostrand Co., Inc.

[6] Moore, P. G., (1952). The estimation of the Poisson parameter from a truncated distribution, Biometrika, 39, 247-51.

[7] Plackett, R. L., (1953). The truncated Poisson distribution, Biometrics 9, 485-88.

[8] Rider, Paul R., (1953). Truncated Poisson distributions, J. Am. Stat. Assn. 48, 826-30.

[9] Subrahmanian, K., (1965). A note on estimation in the truncated Poisson, Biometrika 52, 279-82.

MYTHS IN THE DESIGN OF MEASUREMENTS

Keats A. Pullen, Jr.
U.S. Army Aberdeen Research and Development Center
Ballistic Research Laboratories
Aberdeen Proving Ground, Maryland

SUMMARY. There are at least two different theories which must be taken into account in the design of experiments for making of measurments, namely, the theories of statistics, and the theories of fuzzy sets. In addition, there are some important basic principles which must be taken into account in the use of either of the basic theories.

The purpose of this paper is to explore some of the gray areas created by the mingling of the two theories and by the lack of ideal conditions which must be accepted with the basic theories by themselves. Such questions as small differences, and the unavoidable biases in supposedly random data will be considered.

Further, the information engineering of experiments to assure that meaningful deterministic parameters are given proper weighting in the design of experiments will be discussed.

INTRODUCTION. The basis for measurement of scientific data required in any kind of engineering design is intimately dependent on the theory of design of experiments. In fact, the proper structure for an experiment for making measurements on which design must be based is dependent on a diverse set of disciplines, some of the more important being:

1. The theory of statistics
2. The theory of fuzzy sets
3. The theory of Greco-Roman squares
4. Principles of information engineering
5. General mathematical limitations

Interestingly, it is important to examine NOT ONLY conventional theory of these disciplines as normally developed, but it is possible even more important that the study be enlarged in each case to determine the consequences of built-in biases.

MYTHS DUE TO STATISTICAL BIAS. The theory of statistical analysis is usually invoked when errors of a random nature can be expected to be encountered which render the making of deterministic evaluations of precision difficult if not impossible. This assumption of the use of statistical methods (particularly

Gaussian statistics) for handling errors postulates that there is a very large collection of different error sources, each of which has approximately equal probability of occurrence. In fact, this collection of errors behaves something like a multi-dimensional random-walk problem. (The magnitudes of the random steps and their probabilities of occurrence are by implication equal.) Clearly, the validity of the statistical "sharpening" of the final results is intimately dependent on the validity of the implication of equality of the probabilities.

Statistical processes, of course, are not limited to Gaussian processes based on $\exp(-h^2 x^2)$, but can include various kinds of skew distributions and other shaped distributions. In this respect, it is important to note that if one attempts to develop a statistical weighting system where the weighting of the data points converges, one obtains a divergent, and hence useless, statistical system. The obvious conclusion that this leads to is that the discarding of data point SOLELY on the basis that they differ by more than some number of sigma units from the balance of the set is not permissible. The discard of data points can only be justified IF there are known physical or other reasons which throw the data themselves into legitimate question!

It is generally assumed that an increase in the number of data points leads to an unending increase in available precision for a given measurement. While this assumption may be valid, it need not be valid. The causes-of-error data, for the assumption to be truly valid, must be parts of near infinite sets; all must be essentially equally likely; and the contributions of each must be roughly equal for a valid increase in precision to be possibly obtainable. (Other conditions may have to be met as well.) Otherwise, there is an optimum number N of data which will yield as much precision as can be achieved, and use of more data will only yield an apparent improvement. (The behavior is somewhat similar to that of an asymptotic function.) What this means, of course, is that the results achieved mathematically no longer represent reality since the bias in the original data are obscuring the true values.

Unfortunately, it is seldom possible to determine what the optimum value of N to be selected is. It is a function of all the parameters which could not be processed in a deterministic fashion. The variations in probability of occurrence of the different causes of error, and the variations in magnitude of errors all contribute to this indeterminacy, and lead to the limitation on the available amount of improvement of precision.

Statistical weighting and evaluation is applied whenever the scatter on data points indicates that random processes are disturbing an experiment. Where the range of data variation is small, and there do not appear to be data points which fall significantly away from the normal grouping, this procedure probably is acceptable. When many of the data points do group well, but there are some falling substantially away from the remaining group for no readily apparent reason, use of caution and a careful review of the experiment both are essential to make certain that unsuspected factors are not disturbing the experimental measurements.

The same theory which is used to justify the discarding of an apparently discordant datum actually should warn the user that there is something needing more investigation, or the point would have been compatible with the rest of the set. Typically, the existance of such data is a warning that there may be deterministic sources of error, ones of much larger influence than the rest and occurring more frequently than the rest, still to be separated and studied. These data cannot be ignored with impunity!

THEORY OF FUZZY SETS. Many of the measurements which one must make in the development or utilisation of an experiment are of dubious value at least as much or more because of measurement problems than because of random errors. When this is true, it is often possible to develop an improved understanding of the measurement problem and its nature through consideration in terms of fuzzy sets. Whereas statistical approaches are based on the idea that the number is precise, but cannot be measured precisely, the theory of fuzzy sets implies that the number is only determinable to a limited precision because it is ill-defined by nature. In this case there is a priori a limit to the amount of improvement which can be obtained by use of statistical methods. (In short, a measurement problem is by nature probably partly statistical and partly "fuzzy-set".)

It has been known for at least fifty years that the making of a measurement can disturb the value of the parameter being measured, and that there is in fact a minimum measurable element as a consequence. This idea is basic to the development of quantum physics of the structure of matter, but it has a much wider range of applicability. It becomes clearly evident that this is true when the measurement of energy in an electromagnetic wave is attempted. At high energy levels, the measuring system removes a negligible part of the incident energy, but as the incident level of energy is decreased, what first appeared to be a wave motion begins to look more like a stream of particles (photons). This transition takes place as the sensor requires a major part of the incident energy to activate the measurement. In short, the process of making a measurement can make a "fuzzy set" appear even fuzzier!

INFORMATION ENGINEERING LIMITATIONS. It is almost axiomatic that experiments for measurement of parameters are only as valid as their mathematical and physical foundations are sound. Since in real-life experiments, a variety of kinds of parameters and data must be measured in a meaningful way, it is desirable to review some of the more important mathematical and physical criteria which must be taken into account, and it is also important that the criteria of information engineering be taken into account. The discipline of information engineering, one largely ignored in at least some technological fields, is fundamental to proper application of information obtained by experiment, and for that reason will be considered first.

The information engineering of parameter measurements made in an experiment consists of two principal steps, the first being the separation of the information into basic categories, and the second the establishment of the "order of importance" of each of the kinds of information involved. The separation into categories is required in order that the user may have a better understanding of the environment and detail behavior, and typically the two most important categories are frequently the ones defining the basic operating climate to which the experiment is to be subjected and the actual active measurements to be made in this climate.

In a problem involving active electronic devices, the environmental data would include such parameters as applied voltages and currents and absolute temperature, and the active measurements would include carefully selected small-signal characteristics of the device. In a chemical reaction, the environmental specifications might include such parameters as pressure, temperature, state of materials, catalytic agents, moderators, iso-parameters, applied energy fields, etc., and the detail behavior might involve specifications in terms of derivatives, etc. of the iso-contours specifying the reaction. Stated in electrical terms, it is important to know equipotentials and field gradients in terms of flow rate, or the dual functions, in addition to special environmental restrictions.

The second step in information engineering is one of ordering on the basis of importance of the parameters to the application at hand. This is sometimes considered in terms of a series of experiments starting with a zero-order experiment followed successively by first-order, second-order, etc., in which each successive experiment represents a refinement on the previous one. In another sense, of course, the measurements made should be organized to measure and weight data on the basis of their influence on the application at hand. The problem of ordering these data is a problem in determining the parameters leading to the "steepest descent" and organizing in terms of relative "rates of descent".

As an example of such a characterization, one can examine the typical plate characteristic curves of a pentode electron tube as compared to the corresponding screen characteristic curves. In fact, if the slopes of plate current vs. the available control voltages, plate, grid one, grid two, and grid three are determined, the respective values of the transconductance order as follows:

$$g_{m1} > g_{m2} > g_{m3} > g_p$$

with many tubes. (The order of g_p and g_{m3} may be reversed, but when they are, the reason is a greatly reduced effectiveness of control due to the suppressor grid. Typically, the value of g_p may be taken to approach zero under normal use conditions, and the value of g_{m3} may also be negligible.

The values of g_{m1} and g_{m2} are ALWAYS important.

Order of importance has one very unusual aspect, in that it can be used to show that a small-signal parameter may be more important than an environmental parameter. For example, with a typical pentode tube type, measurements of the characteristics will frequently show that a wide discrepancy may exist between the sets of static curves for different samples of the same type tube, but a small-signal parameter may be specifiable in terms which will apply with ALL of the samples. Because of just such a condition, it is possible to specify pentode conductances in terms of an equation of the form:

$$g_{mj} = \chi_j \Lambda I_b \qquad (2)$$

where the particular chi specifies an efficiency factor for a given element of the tube, the Fermi parameter, Λ, having the value (q/kT) is familiar from solid-state physics, and the current I_b is the plate current. Equation two may also be used for determination of plate conductance by the appropriate substitutions.

It should be noted that Eq. 2 explains why so much design of tube circuits has been based on stabilization of output current.

Other factors in addition to order of importance must be considered in properly organizing the information to be determined by experiment. It is important to determine the stability of the measurements in terms of possible measuring parameters, and to provide data in terms of the most dependable configuration of parameters.

For example, historically, values for transconductance for tubes have been measured as a function of plate current and voltage, and of grid and plate voltages. Examination of the stability considerations involved (particularly in terms of Eq. 2) has shown that the proper selection of characterizing variables must lead to use of plate current and either the plate voltage or the screen voltage (for the multigrid tube), and the experiment should be designed for measurement of transconductance in terms of them.

MATHEMATICAL LIMITATIONS. Tolerances and precision are normally expressed in mathematical terms, and they are of particular concern in the design of experiments. They often enter in unexpected ways into experiments in conjunction with differencing errors in particular. For reasons of convenience, the examples to follow are primarily drawn from the physics of solid-state active devices like transistors.

One of the most important mathematical limitations which must be recognized and taken into account is the principle of small differences. This principle states that use of relations dependent on small differences of large numbers is to be avoided to the greatest extent possible. Since in a sense a derivative is a ratio of two "small differences", it is immediately clear that great caution is mandated in the development of small-signal data to be used with nonlinear devices.

Examination of the nature of the beta for a transistor clearly shows the characteri of the problem. By definition, beta can be given in terms of the equation:

$$\beta = \partial I_c / \partial I_b \tag{3}$$

where the base current is further defined by the equation:

$$I_b + I_c + I_e = 0 \text{ identically} \tag{4}$$

Since for any good transistor, the value of I_b is small compared with either I_c or I_e, it is immediately evident that beta is defined in terms of a derivative taken with respect to a small difference of large numbers! It can be shown that the mu for a triode tube is defined by a similarly uncontrollable derivative with respect to an unstable variable, the grid voltage.

An examination of the problem of coordinating small-signal and static, of environmental, operating conditions shows that the sole function played by the environment is to assure that the required point-by-point small-signal operating conditions can be developed. This means that the environment should be stabilized and controlled by the same parameters or variables that control the small-signal behavior, and the small-signal behavior should also be specified in terms of the same variables. Otherwise, the measurements obtained from the experiment will be of less than optimum usefulness to the worker.

Order-of-importance of stability considerations can in fact be rephrased in terms of the "small-difference" limitation. In particular, the small-signal parameters which may be of use with any nonlinear device should be examined one-by-one to determine which ones can be expressed in the most stable form in terms of physical parameters and important device variables. A tabulation of input and forward immittance properties for the bipolar transistor then might take the form:

Parameter	"Y"	"H"	"Z"
Input	ΔI	$(\Delta I)^{-1}$	$(\Delta I)^{-1}$
Forward	ΔI_c	$(\Delta I)^{-1}$	$(\Delta I)^{-1}$

It immediately becomes clear that ALL input immittances depend on ΔI, AND both the h_f and z_f parameters are also dependent on the same ΔI. Only the y_f is not. In other words, not only the input but also the forward immittance parameters as well are dependent on small differences of large currents with solid-state devices whose characteristics are expressed in terms of the H and Z configurations. It is possible to minimize the consequences of dependence on ΔI in the input immittance, but not in BOTH the input and the forward immittances.

CONCLUSIONS. It is evident from the above discussion that the design of an experiment whose purpose is to provide meaningful measurements is a relatively complex problem. The experiment for making the measurements must be designed to yield meaningful measurement which can effectively describe the characteristics of the object being measured, and it must separate deterministic factors from both statistical and fuzzy-set factors. In addition, the parameters measured must be ones having both mathematical and physical validity, and they must be stated in the most deterministic form possible. Above all, the basic criteria for use of statistical and fuzzy-set methods must be kept clearly in mind.

A BACKWARD ELIMINATION GENERAL SIGNIFICANCE REGRESSION MODEL

Charles E. Colvin
and
Maureen E. Machtolff
Systems Analysis Division
Plans and Analysis Directorate (Prov.)
U.S. Army Missile Command
Redstone Arsenal, Alabama

ABSTRACT

This report develops a statistical procedure for selecting a proper set of independent variables in a linear regression model. The procedure is a backward elimination procedure in which initially a large model is hypothesized and systematically non-significant variables are eliminated one by one.

Two different situations were investigated concerning sample estimates of the error:(1) pure error, and (2) using lack-of-fit as an error estimate, with the effects on the testing procedure for each case.

This is a reproduction of Report No. DA-TR-70-1, produced under DA Project No. 11B15011, AMC Management Structure Code No. 4210.18.0841.1.25.

The remainder of this article was reproduced photographically from the author's manuscript.

Preceding page blank

A BACKWARD ELIMINATION GENERAL SIGNIFICANCE REGRESSION MODEL

Section I. INTRODUCTION

During weapon system studies and investigations, there is no shortage of "output information." At the completion of most field tests, measurements and many forms of recorded data abound. Likewise a large volume of output data is generated during a study where many "computer runs" are made. To properly analyze this output information, analysis techniques should be developed for extracting, from the masses of test results and output data, the main features of the relationships implied or hidden in the data.

In investigating any process or system in which variable quantities change, it will be beneficial to examine the effects that some variables exert (or appear to exert) on others. There may be a functional relationship between variables: in electricity Ohm's law states that voltage (V) in a circuit is equal to the product of the resistance (R) of the wire and the current (I), $V = RI$; or Boyle's law which states that at a constant temperature the pressure (P) and the volume (V) of a given quantity of gas are related by the equation $PV =$ constant. A functional relationship between variables is frequently too complicated to grasp or to describe in simple terms. In this case, we may wish to approximate this functional relationship by some simple mathematical function. By examining such an approximating function we may be able to learn more about the underlying true relationship between important variables. Even where no sensible physical (functional) relationship exists between variables, we may desire to relate them by some sort of mathematical equation. While this equation may be physically meaningless, it may nevertheless be worthwhile for predicting the values of some variables from knowledge of other variables.

Frequently, during many weapons system studies and investigations, the use of large-scale long-running (machine time) computer simulations may be necessary; and, in many cases, field experimentation tests are mandatory. Examples of output information from these simulations or field tests might include factors such as the following:

a. Attrition rate of a postulated enemy force from a computerized wargame simulation

b. Miss distances of missiles fired at targets during field tests

c. Detection ranges of an acquisition radar for a particular type jet aircraft

d. Visual detection ranges for an observer versus a low flying, high speed jet aircraft

In many cases, for each value of these factors, there are a number of concomitant variables which are known and, in many cases, fixed. In the case of b above, fixed and known variables associated with each test shot could be (1) target altitude, (2) range of the target at intercept, (3) target speed, and (4) target type. For d above, the fixed variables could be (1) target altitude, (2) target speed, (3) unmask range of the target, (4) search sector size for the observer, and (5) flight profile type.

Frequently, once the computer simulations are run or the field tests are over, there are a multitude of "what if" questions. Consider example d above and suppose that during this field test a total of three different target altitudes (60, 200, and 1000 meters) are used, as well as two target speeds (400 and 600 knots). Shortly thereafter questions may arise concerning visual detection ranges against a jet aircraft flying at an altitude of 2500 meters and a speed of 750 knots.

The desire to address these "what if" questions, as well as to learn more about the underlying relationships between important variables, necessitates the development of a procedure or model for such types of analysis. This report is the results of such an investigation into the development of a "prediction model" which will be a useful analysis tool for use in weapon system studies and investigations.

Section II. BACKGROUND AND GENERAL METHODOLOGY DEVELOPMENT

Consider a system or some process whereby for a given set of independent variables a response (or output) is obtained; this response will generally be referred to as the dependent variable. Suppose that for a set S of t independent variables, denoted by $S \equiv \{X_1, X_2, \ldots, X_t\}$, it is highly desirable to determine or estimate a relationship for eventually predicting the dependent variable. Consider a total of n responses, each denoted by Y_i, and the appropriate independent variables associated with each response, denoted by $\{X_{1i}, X_{2i}, \ldots, X_{ti}\}$, for $i = 1,2,\ldots,n$. Suppose it is desirable to approximate the relationship (unknown)

$$Y_i = g(X_{1i}, X_{2i}, \ldots, X_{ti}) \tag{1}$$

with a relationship given by the model

$$Y_i = \beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + \ldots + \beta_k X_{ki} + \varepsilon_i \tag{2}$$

for $i = 1,2,\ldots,n$, where the β's are unknown parameters and $k \leq n-1$. The ε_i term of (2) has conventionally become known as the "error" term in the model; it can represent experimental errors (measurement errors, incorrectly calibrated equipment, leaks in the system, etc.), model misclassification errors, etc., or a sum of all errors from several sources. Also note that the parameters are all linear, i.e., no parameters such as β_i^2, β_j^3, etc., which implies a "linear model." The basic assumptions are made for the model given by (2) that

a. ε_i is a random variable with mean zero and variance σ^2 (unknown)

b. ε_i and ε_j are uncorrelated, $i \neq j$, so that

$$\text{Covariance}(\varepsilon_i, \varepsilon_j) = 0,$$

and for hypothesis testing purposes later on

c. ε_i is a normally distributed random variable with mean zero and variance σ^2 (unknown).

The system of observations (responses) given by (2) can be written in vector form as

$$\underline{Y} = X\underline{\beta} + \underline{\varepsilon} \qquad (3)$$

where

$\underline{Y}$ is an (n×1) vector of responses,

X is an (n×p) matrix of known fixed quantities,

$\underline{\beta}$ is a (p×1) vector of unknown parameters,

$\underline{\varepsilon}$ is an (n×1) vector of errors,

where $p = k+1$.

At this stage we assume that the model given by (2) is true. However, at a later stage we must test to determine if indeed it is true. Draper and Smith [3] points out that we are "considering or tentatively entertaining our model." Our goal is not only to obtain estimates for the parameters of $\underline{\beta}$, but also to determine which independent variables are contributing significantly to the response of our mathematical model. Variables that are non-significant can be left out of the model for prediction purposes, and usually they should be deleted.

1. Development of a Prediction Equation

References [3] and [4] show a generally known fact that based on the previous assumptions concerning the ε_i in our model, the least squares and maximum likelihood estimates for $\underline{\beta}$ are given by

$$\hat{\underline{\beta}} = (X'X)^{-1}X'\underline{Y} \qquad (4)$$

where $\hat{\underline{\beta}}$ is an unbiased estimator, i.e., $E(\hat{\underline{\beta}}) = \underline{\beta}$, for $\underline{\beta}$, and where X and $\underline{Y}$ are defined in equation (3). The $(')$ and $(^{-1})$ superscripts on the matrices are the standard notations denoting "transpose" and inverse, respectively. Since $\hat{\underline{\beta}}$ is a random vector, each element being a random variable, it has an associated variance-covariance matrix, denoted by $\mathrm{Var}(\hat{\underline{\beta}})$, where

$$\mathrm{Var}\,(\hat{\underline{\beta}}) = \sigma^2(X'X)^{-1}.$$

Based on our previous assumptions concerning our model it follows that $\underline{\hat{\beta}}$ is a normally distributed random vector with mean $\underline{\beta}$ and variance-covariance matrix $\sigma^2(X'X)^{-1}$; notationally this property is written as

$$\underline{\hat{\beta}} \sim N(\underline{\beta}, \sigma^2(X'X)^{-1}). \tag{5}$$

The variances of the elements in $\underline{\hat{\beta}}$ are equal to the product of σ^2 and the diagonal elements of $(X'X)^{-1}$, and the covariances of the elements in $\underline{\hat{\beta}}$ are equal to the product of σ^2 and the off diagonal elements of $(X'X)^{-1}$. Notationally we have

$$\text{Var}\ (\hat{\beta}_i) = \sigma^2 C_{ii}$$

$$\text{Cov}\ (\hat{\beta}_i, \hat{\beta}_j) = \sigma^2 C_{ij}, i \neq j$$

where C_{ij} represents the i,j^{th} element of $(X'X)^{-1}$.

One statistical technique that could be used to determine which variables in (2) are significant would be to test the hypothesis

$$H_0: \beta_i = 0$$

$$H_1: \beta_i \neq 0 \tag{6}$$

for $i = 1,2,\ldots,k$. For prediction purposes, a variable that is highly non-significant (fail to reject H_0) could be left out of our model; variables that are significant (reject H_0) would be retained. Using this type elimination procedure, one could develop a prediction model as depicted by (2).

Since the β_i, for $i=1,2,\ldots,k$, might not be independent, the results of (6) could very well be dependent on each other. This could lead to the conclusion that a particular variable is significant when, in fact, it only appears that way because it is highly correlated with some other important variable. We need some way of assessing how much "good" a variable is doing over and above the other variables. Namely, we need significance tests of the type given by (6) where, in fact, the test in some fashion adjusts the contribution for a particular variable taking into consideration the fact that it is correlated with others. This can be accomplished by the "General Regression Significance Test" for which the methodology is given below.

2. General Regression Significance Test

The first step is to fit a "full model" as given by (2) and to determine the estimate $\hat{\underline{\beta}}$ for $\underline{\beta}$ as given by (4). The model with the sample estimates for the unknown parameters $\beta_0, \beta_1, \ldots, \beta_k$ has the form

$$y_i = \hat{\beta}_0 + \hat{\beta}_1 X_1 + \hat{\beta}_2 X_2 + \ldots + \hat{\beta}_k X_k + \epsilon_i \qquad (7)$$

where $\hat{\beta}_i$ is the estimate of β_i, for $i = 0,1,2,\ldots,k$. Associated with each fitted regression equation is the so-called sums of squares due to regression which conventionally is denoted by SSres or $R(\beta_0, \beta_1, \ldots, \beta_k)$. For our purposes the latter will be chosen and at times will be referred to as R(all), i.e., the regression due to all of the variables. In essence the sums of squares due to regression is the variation in the responses due to the variation in the independent variables. The $R(\beta_0, \beta_1, \ldots, \beta_k)$ can be shown as

$$R(\beta_0, \beta_1, \ldots, \beta_k) = \hat{\underline{\beta}}' X' \underline{Y} = Y'[X(X'X)^{-1}X']Y \qquad (8)$$

where the X matrix and $\underline{Y}$ vector are defined in (3). Associated with any sums of squares is a parameter known as the degrees of freedom; its importance will be discussed later. For $R(\beta_0, \beta_1, \ldots, \beta_k)$ the degrees of freedom will be denoted by ν where

$$\nu = k + 1.$$

Suppose we are interested in testing the significance of some β_i, the hypothesis which is given in (6). As stated in paragraph 1 above, we need the regression due to β_i adjusted for all of the other variables, which is denoted by $R(\beta_i/\beta_0, \beta_1, \ldots, \beta_{i-1}, \beta_{i+1}, \ldots, \beta_k)$, where $0 < i \leq k$, or $R(\beta_i/\text{all the rest})$. This can be determined from the equation

$$R(\beta_i/\text{all the rest}) = R(\beta_0, \beta_1, \ldots, \beta_k) - R(\beta_0, \beta_1, \ldots, \beta_{i-1}, \beta_{i+1}, \ldots, \beta_k) \qquad (9)$$

where $R(\beta_0,\beta_1,\ldots,\beta_k)$ is given by (8), $R(\beta_0,\beta_1,\ldots,\beta_{i-1},\beta_{i+1},\ldots,\beta_k)$ is the sums of squares due to regression in a "reduced model" with the variable X_i left out, and $0 < i \leq k$. The degrees of freedom for $R(\beta_i/\text{all the rest})$ is unity. Recall from Section I that the error term in our model is a random variable with a variance of σ^2 which is unknown. To test the hypothesis given by (6) we must have an estimate of the unknown variance σ^2. Given an estimate of σ^2 (or error estimate) we can form a test statistic, whose distribution is known under the null hypothesis, and can test the appropriate hypothesis as depicted by (6). Two different situations concerning estimates of the error will be discussed: (1) the case where there is an estimate of "pure error" (Case A), and (2) the case where there is no "pure error" estimate in which specific assumptions must be made (Case B).

a. Estimating the Pure Error (Case A)

Recall that we have a total of n responses. Let there be a total of s distinct sets of independent variables, $s_i = \{X_{1i},X_{2i},\ldots,X_{ti}\}$, and suppose for each set there are n_i responses, for $i = 1,2,\ldots,s$. Thus we have n_i replications per set of independent variables. This can be written in vector form as

$$\underline{Y}_{n\times 1} = \begin{bmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_{n_1} \\ \hline Y_{n_2+1} \\ Y_{n_1+2} \\ \vdots \\ Y_{n_1+n_2} \\ \hline \vdots \\ \hline \vdots \\ Y_{n_1+n_2+\ldots+n_s} \end{bmatrix} \begin{matrix} \\ \\ \\ n_{1\times 1} \\ \\ \\ \\ n_{2\times 1} \\ \\ \vdots \\ \\ n_{s\times 1} \end{matrix} \qquad (10)$$

where $n_1+n_2+\ldots+n_s=n$. With replications, as we have in this case, an unbiased estimate for the error exists. This estimate is said to represent "pure error"; because, with the independent variables identical for two responses, only the random variation can influence the results and provide differences between them. Let the sums of squares due to pure error be denoted by SSPE. The computational formula for SSPE is not given here, but it can be found in many textbooks. Denote the degrees of freedom associated with SSPE by dFPE which is given by

$$dFPE = (n_1-1)+(n_2-1)+\ldots+(n_s-1). \tag{11}$$

We can now find the mean square of the pure error (MSE) by the ratio

$$MSE = SSPE/dFPE \tag{12}$$

where $E(MSE) = \sigma^2$. With this estimate of the error we can now test the hypothesis given by (6) for some β_i of interest. We form the test ratio

$$F_i = R(\beta_i/\text{all the rest})/MSE \tag{13}$$

where $R(\beta_i/\text{all the rest})$ is given by (9) and MSE by (12) The distribution of the test ratio under the null hypothesis is an "F" random variable with degrees of freedom 1 and dFPE. Therefore, based on the value of F_i in (13) and some nominal significance level, a decision is made concerning retaining β_i in our prediction model. A test statistic can be formulated for all β_i using the equations given by (9), (12), and (13) with separate decisions being made as to whether or not to retain each β_i in our model for $i = 1,2,\ldots,k$.

b. No Estimate of Pure Error (Case B)

As in the previous case, we have a total of n responses, as well as s distinct sets of independent variables. In this case, however, we have no replications, which means that $n_1=n_2=\ldots=n_s=1$. In this case, we have no estimate of pure error which means that the error estimate must come from some other source.

Consider the model proposed by (2) with its (k+1) parameters $\beta_0, \beta_1, \ldots, \beta_k$. Theoretically, in any regression model the total number of parameters can equal, but not exceed, the number of responses (observations). If the postulated model (2) contains less than the maximum allowable number of parameters it is said to suffer from "lack-of-fit." The total sums of squares associated with our n responses can be partitioned in such a way as to provide a lack-of-fit sums of squares, denoted by SSLF. The SSLF, in the case where there are no replications, can be determined from the equation

$$SSLF = \underline{Y}'\underline{Y} - \underline{\hat{\beta}}'X'\underline{Y} \qquad (14)$$

where $\underline{Y}'\underline{Y}$ is the total sums of squares and

$\underline{\hat{\beta}}'X'\underline{Y}$ is the total regression sums of squares given by (8).

The SSLF divided by its corresponding degrees of freedom, SSLF/dFLF, will be denoted by MSLF which provides us an estimate for the unknown error (σ^2). The dFLF can be found from the relation

$$dFLF = n-(k+1)$$

where n is the total number of responses and (k+1) is the total number of parameters in our postulated model.

Given an error estimate, i.e., MSLF, we can proceed exactly as we did in Case A. That is, we form the test ratios given by (13) for each β_i and draw conclusions concerning each postulated β_i in our model for $i = 1,2,\ldots,k$. Here we use MSLF instead of MSE in the test ratios.

It will be mentioned briefly that using the lack of fit for an error estimate has one major shortcoming. The mean square (MSLF) for the lack of fit estimates the error (σ^2) only in the case where our postulated model (2) is true. If it is untrue, then the mean square for the lack of fit (MSLF) contains not only random error but, also, systematic error. This can be expressed in the expected value of MSLF by

$$E(MSLF) = \sigma^2+k$$

where k is some positive constant whose magnitude is dependent upon how "untrue" the model is. If the model is untrue we have no estimate of error and, thus, no valid test statistic. This fact demonstrates the need for replications, and in most situations (field tests and computer simulations), replications are a planned part of the study.

In summary, the following steps were taken during the development of the prediction model:

(1) Fit a "full model" as given by (2)

(2) Obtain the regression due to each β_i adjusted for all of the other variables, for $i = 1,2,\ldots,k$, given by (9).

(3) Obtain an estimate for the unknown variance (σ^2), Case A or Case B

(4) Conduct k separate test of hypothesis as to the significance of each β_i in the prediction model, for $i = 1,2,\ldots,k$.

Section III. SELECTING THE BEST REGRESSION EQUATION

Recall from Section II that a linear regression equation, or prediction equation, was established for a particular response Y in terms of independent variables $X_1, X_2, \ldots, X_k$. A full model was fit with a resulting regression equation given by (7). Methods were developed to determine if any variable in the full model could be left out as far as prediction purposes are concerned. Initially one might be inclined to select a regression equation with as many X's (independent variables) as possible so that reliable fitted or predicted values could be determined. This could be disadvantageous from a practical point of view in that excessive costs and energy might be required in obtaining information on a large number of X's. Also, it is extremely important to remember that each predicted response is a random variable itself with an expected value and variance. Suppose, for discussion purposes, that the full model given by (7) is the final model we select after going through the procedures discussed in Section II. The variance for a predicted value of Y_i given a set of fixed X's is given by

$$\mathrm{Var}(\hat{Y}_i) = \mathrm{Var}[\hat{\beta}_0 + \hat{\beta}_1 X_1 + \ldots + \hat{\beta}_k X_k + \varepsilon_i]$$

or

$$\mathrm{Var}(\hat{Y}_i)=\mathrm{Var}(\hat{\beta}_0)+\mathrm{Var}(\hat{\beta}_1 X_1)+\ldots+\mathrm{Var}(\hat{\beta}_k X_k)+2X\ \mathrm{Cov}(\hat{\beta}_0,\hat{\beta}_1)$$
$$+\ldots+2X_{k-1}X_k\mathrm{Cov}(\hat{\beta}_{k-1},\hat{\beta}_k)+\mathrm{Var}(\varepsilon_i) \qquad (15)$$

Note that as the X's increase in the model, the precision about each predicted response decreases, i.e., the variance increases. Since this is true, it would not be desirable to include a large number of X's in the model when, in fact, a smaller number might be sufficient. The inclusion of too many variables in the model has become known as "over-fitting" the model. The procedure for selecting an ample number of variables, guarding against over-fitting, is called "selecting the best equation."

1. Selecting the Best Equation

There is no unique statistical procedure for selecting the "best" regression equation; personal judgment and an understanding of the system being modeled are necessary in choosing the best regression equation. A literature search provides a rather large list of procedures that can be used in developing a regression equation. Unfortunately, they do not necessarily lead to the same solution. The procedure used in this study for choosing the best equation is called "the backward elimination procedure" and is discussed below.

The Backward Elimination Procedure

Many steps and procedures have been discussed in Section I and II that are applicable in this backward elimination procedure. Before this procedure is described, one principle pertaining to the partitioning of the total sums of squares, $\underline{Y}'\underline{Y}$, will be presented. The total sums of squares can be partitioned into two major separate components called the regression and residual sums of squares. This can be expressed in the equation

$$\underline{Y}'\underline{Y} = \hat{\underline{\beta}}'\underline{X}'\underline{Y}+[\underline{Y}'\underline{Y}-\hat{\underline{\beta}}'\underline{X}'\underline{Y}] \qquad (16)$$

where

$\underline{Y}'\underline{Y}$ is the total sums of squares

$\hat{\underline{\beta}}'\underline{X}'\underline{Y}$ is the total regression sums of squares and

$\underline{Y}'\underline{Y}-\hat{\underline{\beta}}'\underline{X}'\underline{Y}$ is known as the residual sums of squares.

Equation 16 can be written for brevity as

$$SS_{total} = SS_{reg} + SS_{res} \tag{17}$$

where SS_{reg} is the same as $R(\beta_0,\beta_1,\ldots,\beta_k)$ in (8). Theoretically, the SS_{res} can be broken down into two separate quantities -- the sums of squares due to "pure error" and the sums of squares due to "lack of fit," denoted by SSPE and SSLF, respectively.

Thus we have in equation form

$$SS_{res} = SSPE + SSLF \tag{18}$$

Suppose that a regression model is being "entertained" in a situation in which there are true replications in the data and further that the model has some lack of fit. That is, we did not include the maximum allowable number of parameters in our model. The possibility exists that this lack of fit may be statistically significant, which would indicate that our model should be "enlarged." Thus we need a test to determine if the lack of fit is significant. If it is insignificant we would conclude that our model suffers from no lack of fit, which would imply that the model is large enough. If the lack of fit is significant we would conclude that our model should be enlarged, in which case we would entertain a new, enlarged model. From (18) we see that

$$SSLF = SS_{res} - SSPE \tag{19}$$

where SS_{res} is given in (18) and SSPE was discussed in Section II. It follows that, with an estimate of pure error (equation (12)), a test ratio can be found to test for the significance of the lack of fit. We determine the mean square for the lack of fit, MSLF, given by

$$MSLF = SSLF/dFLF \tag{20}$$

where SSLF is given by (19) and the degrees of freedom due to lack of fit, dFLF, is given by

$$dFLF = (n-(k+1)) - dFPE. \tag{21}$$

Recall that dFPE is given by (11). The null hypothesis we wish to test is

$$H_0: \text{Lack of fit} = 0 \qquad (22)$$

$$H_1: \text{Lack of fit} \neq 0.$$

We form the test statistic

$$F_{LF} = MSLF/MSE \qquad (23)$$

which under the null hypothesis is an F random variable with degrees of freedom dFLF and dFPE. Since this is a one-sided upper-tailed test, a critical value, F_0, at some specified significance level is found from the appropriate table of critical values for the F distribution. If $F_{LF} > F_0$, we would conclude that the lack of fit is significant; otherwise we would fail to reject the null hypothesis and conclude that there is no lack of fit.

It must be noted that no test for the lack of fit is possible if we have no replications. With no replications we have SSPE $\equiv 0$ and from (18) we see that

$$SS_{res} = SSLF. \qquad (24)$$

Recall that in the case where we have no replications, the mean square for the lack of fit had to be used as our estimate for the unknown variance (σ^2). Therefore, if we have no replications it follows that no test exists to determine the significance of the lack of fit.

We are in a position now to describe the backward elimination procedure developed in this study. The basic steps are as follows:

(1) Fit a "full model" as given by (7) that you feel is purposely too large.

(2) Determine the regression for each β_i variable adjusted for all other variables given by (9).

(3) Determine the test ratio for the lack of fit if we have replications.

(4) Form test ratios for β_i as given by (13), for $i = 1,2,\ldots,k$.

(5) Determine whether or not the lack of fit is significant. If it is significant enlarge the model and repeat Steps (1) thru (5) again.

(6) The lowest F-test value from Step (4), F_L, is compared with a preselected F_o critical value at some nominal significance level.

(7) If $F_L < F_o$, remove the variable X_L, which gave rise to F_L, from the model, and recompute the regression equation in the remaining variables and repeat steps (1) thru (7).

(8) If $F_L > F_o$, accept the regression equation as calculated.

The analysis of variance (ANOVA) table for the first time through Steps (1) thru (7) is shown in Table I. Once Step (8) is reached the backward elimination selection procedure is terminated and those remaining variables not previously eliminated are the ones in the final regression model.

Table I. ANOVA Table for Elimination Procedure

Source	Degrees of Freedom	Sums of Squares	F Ratios
Total (uncorrected)	n	$\underline{Y}'\underline{Y}$	
Regression	k+1	$\hat{\underline{\beta}}'X'\underline{Y}$	
$R(\beta_0/\text{all the rest})$	1	$R(\beta_0/\text{all the rest})$	
$R(\beta_1/\text{all the rest})$	1	$R(\beta_1/\text{all the rest})$	F_1
.	.	.	.
.	.	.	.
.	.	.	.
$R(\beta_k/\text{all the rest})$	1	$R(\beta_k/\text{all the rest})$	F_k
Residual	n-(k+1)	$\underline{Y}'\underline{Y}-\hat{\underline{\beta}}'X'\underline{Y}$	
Lack of Fit	Subtraction by	$(Y'Y-\hat{\underline{\beta}}'X'Y)$-SSPE	F_{LF}
Pure Error	$(n_1-1)+\ldots+(n_s-1)$	SSPE	

The backward elimination procedure has been computerized in order for the selection procedure to be "automated." The Appendix contains descriptions of the computer program along with the input format for the data. The program prints out the ANOVA table each time through steps (1) thru (7) along with a calculated quantity known as the square of the multiple correlation coefficient, denoted by R^2. It is calculated from the equation

$$R^2 = \frac{\hat{\underline{\beta}}'\underline{X}'\underline{Y} - n\bar{Y}^2}{\underline{Y}'\underline{Y} - n\bar{Y}^2} \qquad (25)$$

where $\hat{\underline{\beta}}'\underline{X}'\underline{Y}$ is the regression sums of squares after the completion of step (7) each time through the procedure,

$\bar{Y}$ is merely the mean of all the responses, and

n is the total number of responses.

R^2 is a measure of the usefulness of the terms, other than β_0, in the model, where $0 < R^2 \leq 1$. Generally, large values of R^2, $R^2 > 0.90$, are associated with models that fit the data very well. R^2 might be used as an easily obtainable measure of the success of the regression equation in explaining the variation in the data. One must remember, however, that it can be made large merely by including in our model near the maximum allowable terms, the maximum being the number of observations. The multiple correlation coefficient is calculated and displayed, not from the standpoint as a measure of success of our model, but rather from the statisticians viewpoint of seeing how little R^2 decreases from the initial "full model" to the final selected model. This will be shown in the following example.

2. Example

A rocket engine's performance is affected by a number of environmental factors and conditions. One measure of normal performance that can be selected for analysis purposes is chamber pressure. A typical set of such data is shown in Table II, which we will use to develop an empirical prediction model for chamber pressure. With the data given in Table II let us entertain or hypothesize a model which includes the variables X_1, X_2, X_3, and X_4, as well as all of

Table II. Rocket Engine Data

Chamber Pressure (Y)	Temperature of Cycle (X_1)	Vibration (X_2)	Drop Shock (X_3)	Static Fire (X_4)
1.4	-75	0	0	-65
26.3	175	0	0	150
26.5	0	-75	0	150
5.8	0	175	0	-65
23.8	0	-75	0	150
7.4	0	175	0	-65
29.4	0	0	-65	150
9.7	0	0	165	-65
32.9	0	0	0	150
26.4	-75	-75	0	150
8.4	175	175	0	-65
28.8	0	-75	-65	150
11.8	0	175	165	-65
28.4	-75	-75	-65	150
11.5	175	175	165	-65
26.5	0	-75	0	150
5.8	0	175	0	-65
1.3	0	0	-65	-65
21.4	0	0	165	150
0.4	0	-75	-65	-65
22.9	0	175	165	150
26.4	0	-75	-65	150
11.4	0	175	165	-65
3.7	0	0	0	-65

the first order cross-products between variables, i.e., $X_1, X_2, \ldots, X_3X_4$. If this is the case, our model will take on the form

$$Y_i = \beta_0+\beta_1X_1+\beta_2X_2+\beta_3X_3+\beta_4X_4+\beta_5X_1X_2+\beta_6X_1X_3 \qquad (26)$$

$$+\beta_7X_1X_4+\beta_8X_2X_3+\beta_9X_2X_4+\beta_{10}X_3X_4+\varepsilon_i \ ,$$

Where the β's are unknown coefficients, and will be estimated from our sample of 24 responses. The X matrix for this situation (equation 3) is given in Table III. When we look at our model along with the vector of responses and the X matrix we see there are some variables, especially cross-products, that probably will be insignificant in our model. When you do not know what variables should be in the model, you purposely hypothesize a large model (perhaps over-fit) and then let the backward elimination procedure choose those variables that should be eliminated and those that should be retained.

Tables IV thru X (ANOVA Tables 1 thru 7) are copies from computerized model printouts of the backward elimination procedure. ANOVA Table 1 shows the regression for each variable in the full model adjusted for all of the other variables. The lack of fit is tested the first time through and is found to be insignificant; thus, we conclude that our model does not suffer from any lack of fit. We see that variable $X_5 \equiv X_1X_2$ is eliminated as being insignificant. Note that the square of the multiple correlation coefficient is 0.984769. This table also shows the sample vector of the unknown parameters with β_5 eliminated.

We see in ANOVA Table 6 that variable $X_8 \equiv X_2X_3$ is eliminated and, in fact, X_8 is the last variable to be eliminated as indicated by ANOVA Table 7. Since no variable is eliminated in Table 7, the backward elimination procedure has determined that, of the ten original independent variables included in the full model, only variables X_2, X_3, X_4 and X_3X_4 should be retained in our prediction model. The sample estimates of the unknown parameters $\beta_0, \beta_2, \beta_3, \beta_4$, and β_{10} are given in this order in ANOVA Table 6. The final prediction equation (regression model) is written as

$$\hat{Y}_i = 10.6376+0.0183X_2+0.0106X_3+0.115X_4-0.0003X_3X_4 \qquad (27)$$

Table III. Matrix of Known Quantities.

X_0	X_1	X_2	X_3	X_4	X_1X_2	X_1X_3	X_1X_4	X_2X_3	X_2X_4	X_3X_4
1	-75	0	0	-65	0	0	4875	0	0	0
1	175	0	0	150	0	0	26250	0	0	0
1	0	-75	0	150	0	0	0	0	-11250	0
1	0	175	0	-65	0	0	0	0	-11375	0
1	0	-75	0	150	0	0	0	0	-11250	0
1	0	175	0	-65	0	0	0	0	-11375	0
1	0	0	-65	150	0	0	0	0	0	-9750
1	0	0	165	-65	0	0	0	0	0	-10725
1	0	0	0	150	0	0	0	0	0	0
1	-75	-75	0	150	5625	0	-11250	0	-11250	0
1	175	175	0	-65	30625	0	-11375	0	-11375	0
1	0	-75	-65	150	0	0	0	4875	-11250	-9750
1	0	175	165	65	0	0	0	28875	-11375	-10725
1	-75	-75	-65	150	5625	4875	-11250	4875	-11250	-9750
1	175	175	165	-65	30625	28875	-11375	28875	-11375	-10725
1	0	-75	0	150	0	0	0	0	-11250	0
1	0	175	0	-65	0	0	0	0	-11375	0
1	0	0	-65	-65	0	0	0	0	0	4225
1	0	0	165	150	0	0	0	0	0	24750
1	0	-75	-65	-65	0	0	0	4875	4875	4225
1	0	175	165	150	0	0	0	28875	26250	24750
1	0	-75	-65	150	0	0	0	4875	-11250	-9750
1	0	175	165	-65	0	0	0	28875	-11375	-10725
1	0	0	0	-65	0	0	0	0	0	0

Table IV. ANOVA Table 1

Source of Variation	Degrees of Freedom	Sum of Squares	Mean Square	F Ratio
Total (Uncorrected)	24	9308.4500		
Total Regression	11	9267.1495		
Due to β 0/rest	1	995.1493	995.1493	539.2139
Due to β 1/rest	1	3.8214	3.8214	2.0706
Due to β 2/rest	1	31.3958	31.3958	17.0116
Due to β 3/rest	1	19.1695	19.1695	10.3868
Due to β 4/rest	1	1763.5096	1763.5096	955.5440
Due to β 5/rest	1	1.7787	1.7787	0.9638
Due to β 6/rest	1	3.0182	3.0182	1.6354
Due to β 7/rest	1	7.0241	7.0241	3.8059
Due to β 8/rest	1	5.2367	5.2367	2.8374
Due to β 9/rest	1	2.7710	2.7710	1.5014
Due to β 10/rest	1	121.2713	121.2713	65.7099
Residual	13	41.3004		
Lack of Fit	7	30.2271	4.3182	2.3398
Pure Error	6	11.0733	1.8456	

Percentage Variation Explained by Regression - 98.4769

Variable Leaving X_5

Vector of Unknown Parameters with Variable X_5 Deleted

10.9908
0.0070
0.0196
0.0160
0.1134
-0.0001
-0.0001
-0.0001
0.0001
-0.0004

Table III. Matrix of Known Quantities.

X_0	X_1	X_2	X_3	X_4	X_1X_2	X_1X_3	X_1X_4	X_2X_3	X_2X_4	X_3X_4
1	-75	0	0	-65	0	0	4875	0	0	0
1	175	0	0	150	0	0	26250	0	0	0
1	0	-75	0	150	0	0	0	0	-11250	0
1	0	175	0	-65	0	0	0	0	-11375	0
1	0	-75	0	150	0	0	0	0	-11250	0
1	0	175	0	-65	0	0	0	0	-11375	0
1	0	0	-65	150	0	0	0	0	0	-9750
1	0	0	165	-65	0	0	0	0	0	-10725
1	0	0	0	150	0	0	0	0	0	0
1	-75	-75	0	150	5625	0	-11250	0	-11250	0
1	175	175	0	-65	30625	0	-11375	0	-11375	0
1	0	-75	-65	150	0	0	0	4875	-11250	-9750
1	0	175	165	65	0	0	0	28875	-11375	-10725
1	-75	-75	-65	150	5625	4875	-11250	4875	-11250	-9750
1	175	175	165	-65	30625	28875	-11375	28875	-11375	-10725
1	0	-75	0	150	0	0	0	0	-11250	0
1	0	175	0	-65	0	0	0	0	-11375	0
1	0	0	-65	-65	0	0	0	0	0	4225
1	0	0	165	150	0	0	0	0	0	24750
1	0	-75	-65	-65	0	0	0	4875	4875	4225
1	0	175	165	150	0	0	0	28875	26250	24750
1	0	-75	-65	150	0	0	0	4875	-11250	-9750
1	0	175	165	-65	0	0	0	28875	-11375	-10725
1	0	0	0	-65	0	0	0	0	0	0

Table V. ANOVA Table 2

Source of Variation	Degrees of Freedom	Sum of Squares	Mean Square	F Ratio
Total (Uncorrected	24	9308.4500		
Total Regression	10	9265.3708		
Due to β 0/rest	1	1150.6872	1150.6872	623.4910
Due to β 1/rest	1	2.6233	2.6233	1.4214
Due to β 2/rest	1	31.6996	31.6996	17.1762
Due to β 3/rest	1	20.0527	20.0527	10.8654
Due to β 4/rest	1	1803.7715	1803.7715	977.3596
Due to β 6/rest	1	4.9871	4.9871	2.7022
Due to β 7/rest	1	8.7484	8.7484	4.7402
Due to β 8/rest	1	4.6604	4.6604	2.5252
Due to β 9/rest	1	2.5714	2.5714	1.3933
Due to β 10/rest	1	121.2032	121.2032	65.6730
Residual	14	43.0791		
Lack of Fit	8	32.0058	4.0007	2.1678
Pure Error	6	11.0733	1.8456	

Percentage Variation Explained by Regression - 98.4113

Variable Leaving X_9

Vector of unknown Parameters with Variable X_9 Deleted

10.7164
0.0068
0.0183
..0155
0.1125
-0.0001
-0.0001
-0.0001
-0.0003

Table VI. ANOVA Table 3

Source of Variation	Degrees of Freedom	Sum of Squares	Mean Square	F Ratio
Total (Uncorrected)	24	9308.4500		
Total Regression	9	9262.7994		
Due to β 0/rest	1	1517.2513	1517.2513	822.1109
Due to β 1/rest	1	2.4660	2.4660	1.3362
Due to β 2/rest	1	29.2196	29.2196	15.8324
Due to β 3/rest	1	18.9586	18.9586	10.2726
Due to β 4/rest	1	1847.6881	1847.6881	1001.1555
Due to β 6/rest	1	4.1765	4.1765	2.2630
Due to β 7/rest	1	6.6966	6.6966	3.6285
Due to β 8/rest	1	2.7748	2.7748	1.5035
Due to β 10/rest	1	179.5070	179.5070	97.2645
Residual	15	45.6506		
Lack of Fit	9	34.5772	3.8419	2.0817
Pure Error	6	11.0733	1.8456	

Percentage Variation Explained by Regression - 98.3165

Variable leaving X_1

Vector of Unknown Parameters with Variable X_1 Deleted

10.7134
0.0206
0.0154
0.1132
-0.0001
-0.0001
-0.0001
-0.0003

Table VII. ANOVA Table 4

Source of Variation	Degrees of Freedom	Sum of Squares	Mean Square	F Ratio
Total (Uncorrected)	24	9308.4500		
Total Regression	8	9260.3334		
Due to β 0/rest	1	1516.4824	1516.4824	821.6943
Due to β 2/rest	1	45.3953	45.3953	24.5971
Due to β 3/rest	1	18.6350	18.6350	10.0972
Due to β 4/rest	1	1928.5725	1928.5725	1044.99821
Due to β 6/rest	1	2.0264	2.0264	1.0980
Due to β 7/rest	1	4.4023	4.4023	2.3854
Due to β 8/rest	1	4.5745	4.5745	2.4787
Due to β 10/rest	1	182.0314	182.0314	98.6323
Residual	16	48.1166		
Lack of Fit	10	37.0433	3.7043	2.0072
Pure Error	6	11.0733	1.8456	

Percentage Variation Explained by Regression - 98.2255

Variable Leaving X_6

Vector of Unknown Parameters with Variable X_6 Deleted

10.6867
0.0200
0.0137
0.1129
-0.0000
-0.0001
-0.0003

Table VIII. ANOVA Table 5

Source of Variation	Degrees of Freedom	Sum of Squares	Mean Square	F Ratio
Total (Uncorrected)	24	9308.4500		
Total Regression	7	9258.3070		
Due to β 0/rest	1	1515.9869	1515.9869	821.4258
Due to β 2/rest	1	43.6365	43.6365	23.6441
Due to β 3/rest	1	16.6087	16.6087	8.9993
Due to β 4/rest	1	1932.7930	1932.7930	1047.2690
Due to β 7/rest	1	2.9298	2.9298	1.5875
Due to β 8/rest	1	3.6566	3.6566	1.9813
Due to β 10/rest	1	190.3714	190.3714	103.1513
Residual	17	50.1430		
Lack of Fit	11	39.0696	3.5518	1.9245
Pure Error	6	11.0733	1.8456	

Percentage Variation Explained by Regression - 98.1508

Variable Leaving X_7

Vector of Unknown Parameters with Variable X_7 Deleted

10.7079
0.0202
0.0138
0.1127
-0.0001
-0.0003

Table IX. ANOVA Table 6

Source of Variation	Degrees of Freedom	Sum of Squares	Mean Square	F Ratio
Total (Uncorrected)	24	9308.4500		
Total Regression	6	9255.3771		
Due to β 0/rest	1	1525.1281	1525.1281	826.3789
Due to β 2/rest	1	44.3778	44.3778	24.0458
Due to β 3/rest	1	16.9036	16.9036	9.1591
Due to β 4/rest	1	1929.9972	1929.9972	1045.7540
Due to β 8/rest	1	3.5612	3.5612	1.9296
Due to β 10/rest	1	199.1669	199.1669	107.9170
Residual	18	53.0728		
Lack of Fit	12	41.9995	3.5000	1.8964
Pure Error	6	11.0733	1.8456	

Percentage Variation Explained by Regression - 98.0427

Variable Leaving X_8

Vector of Unknown Parameters with Variable X_8 Deleted

10.6376
0.0183
0.0106
0.1115
-0.0003

Table X. ANOVA Table 7

Source of Variation	Degrees of Freedom	Sum of Squares	Mean Square	F Ratio
Total (Uncorrected)	24	9308.4500		
Total Regression	5	9251.8159		
Due to β 0/rest	1	1533.4719	1533.4719	830.8999
Due to β 2/rest	1	40.9456	40.9456	22.1861
Due to β 3/rest	1	13.3751	13.3751	7.2472
Due to β 4/rest	1	2011.4026	2011.4026	1089.8629
Due to β 10/rest	1	197.6084	197.6084	107.0726
Residual	19	56.6340		
Lack of Fit	13	45.5606	3.5047	1.8990
Pure Error	6	11.0733	1.8456	

Percentage Variation Explained by Regression - 97.9114

where the estimates for the β's are given in ANOVA Table 7. Observe in ANOVA Table 7 that the square of the multiple correlation coefficient is 0.979114, and recall from ANOVA Table 1 that this coefficient is 0.984769. Thus, we see that even after eliminating six variables, there is only a slight reduction in percentage of variation explained by regression. This implies that our final equation is doing a good job in "fitting" the data.

3. Concluding Remarks

This report has presented the backward elimination procedure developed for selecting the proper variables that should be included in a linear regression model for prediction purposes. Several assumptions (explicit and implicit) were made that were necessary for the development of the elimination procedure. If these assumptions in practice are incorrect, the backward elimination procedure or any general regression procedure will be invalid.

Two major assumptions were: (1) the normality assumption on our error term (ε_i) in our model (explicit assumption), and (2) the necessity for the ($X'X$) matrix to be a non-singular matrix (implicit assumption). If assumption (1) is incorrect, we have no valid tests for our variables in the ANOVA table. This assumption does not affect the sample estimate for the unknown parameters $\beta_0, \beta_1, \ldots, \beta_k$.

The topic of what to do if the observed responses are not normally distributed has not been addressed in this study. Needless to say, it is an important topic; however, the literature has ample examples pertaining to various transformation techniques that can be used to attain the basic requirements of normality (References [1], [2], [3], and [5]).

Assumption (2) means in practice that our normal equations

$$X'X\hat{\underline{\beta}} = X'\underline{Y} \tag{28}$$

must involve as many independent equations as there are parameters to be estimated. If the ($X'X$) matrix is singular, then steps must be taken to correct this. In most designed experiments the ($X'X$) matrix is always non-singular and thus no problem exists.

Another topic not discussed in this study is the treating of qualitative or deterministic variables. In the model proposed in this study all of the independent variables have been quantative over some continuous range. Suppose we wish to include variables such as machine number, factor number, vat number, missile modification number, etc. in our model. These variables need to be quantified if they are included in our prediction model. Reference [3] has a good discussion on the use of dummy (qualitative) variables in multiple regression. Generally, there is no problem in including these types of variables in a regression model.

LITERATURE CITED

[1] Bartlett, M.S. "The Use of Transformations," Biometrics 3. ;957, pp. 39-52.

[2] Box, G.E.P., and Cox, D.R. "An Analysis of Transformations," Journal of the American Statistical Association, B-26, 1964, pp. 211-243, Discussion 244-252.

[3] Draper, N.R., and Smith, H. Applied Regression Analysis, John Wiley & Sons, Inc., New York, 1961.

[4] Graybill, F.A. An Introduction to Linear Statistical Models, McGraw-Hill, Inc., New York, 1961.

[5] Snedecor, George W. Statistical Methods, 5th Ed., Iowa State University Press, Ames, Iowa, 1960.

APPENDIX

GENERAL PROGRAM DESCRIPTION

Section I. INTRODUCTION

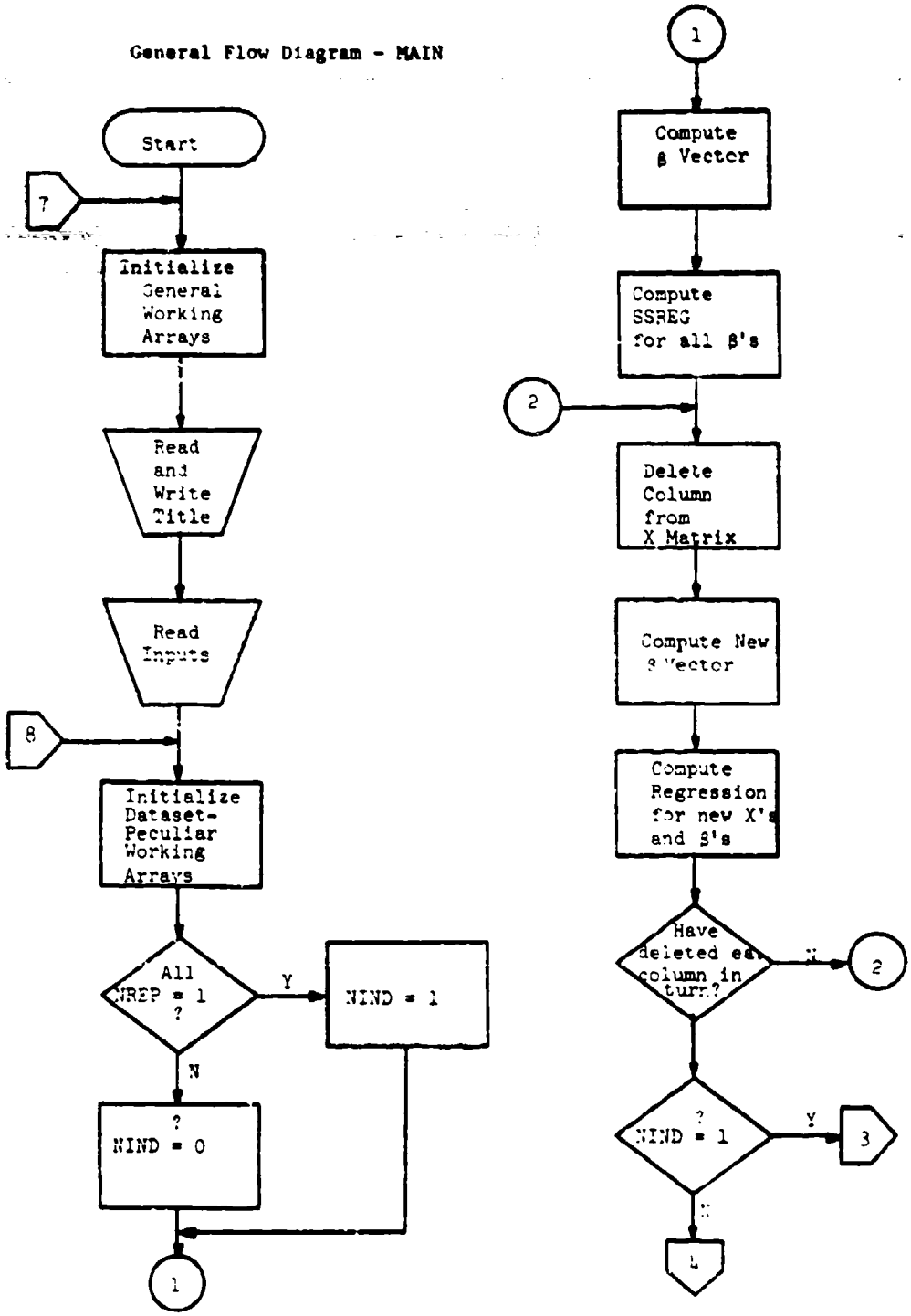
The Prediction Model, written in Fortran IV, uses double precision arithmetic for the greatest degree of accuracy possible. The program was written for the IBM 7094 and runs approximately two minutes on this machine.

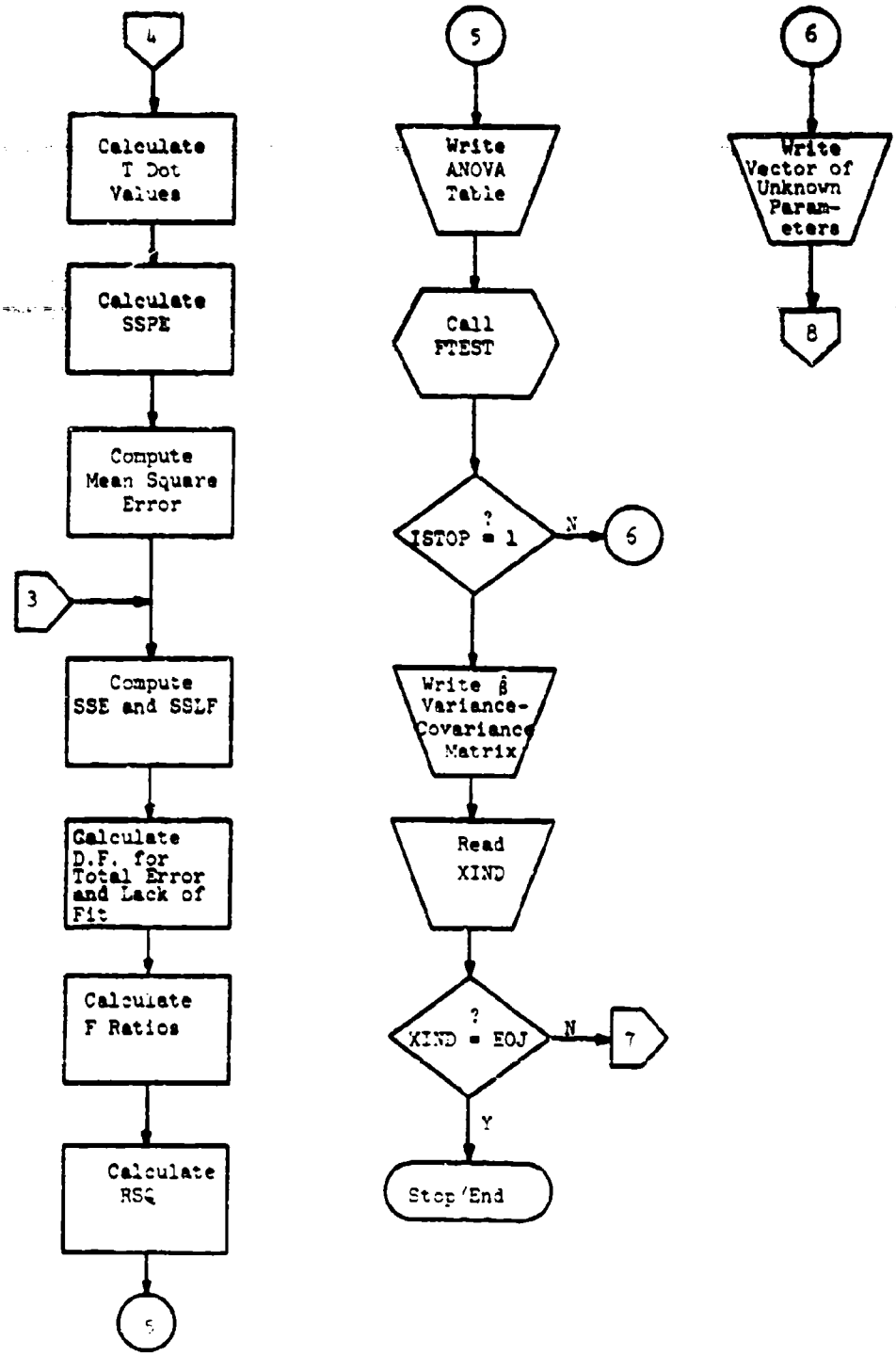
The General Flow Diagrams for MAIN and FTEST subroutines are included herein.

The model has the capability of computing answers for any number of cases in a single run, provided the deck setup is as described in Section II, Data Inputs. A complete set of deck cards must be inserted for each case desired.

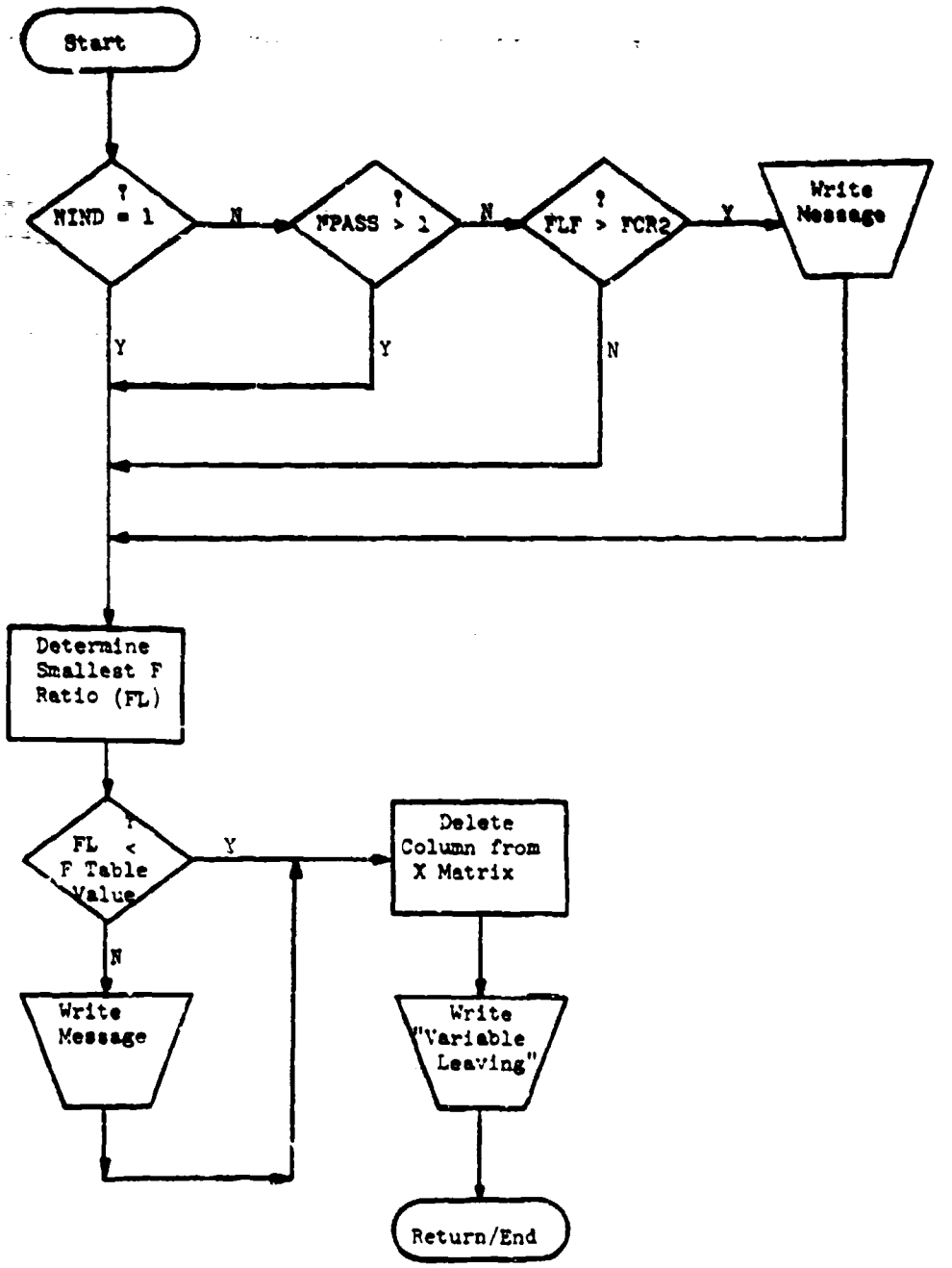
The Fortran IV Source Statements are given in Section III.

General Flow Diagram - MAIN




4
Calculate T Dot Values
Calculate SSPE
Compute Mean Square Error
3
Compute SSE and SSLF
Calculate D.F. for Total Error and Lack of Fit
Calculate F Ratios
Calculate RSQ
5
5
Write ANOVA Table
Call FTEST
ISTOP = 1 ?
N
6
Write β̂ Variance-Covariance Matrix
Read XIND
XIND = EOJ ?
N
7
Y
Stop 'End
6
Write Vector of Unknown Param-eters
8

General Flow Diagram - FTEST



Section II. DATA INPUTS

The data inputs for the Prediction Model are illustrated in the following figures. A complete data set consists of

(1) 1 Title Card

(2) 1 Control Parameters Card

(3) 1 Replications Identification Card

(4) 1 set Matrix Entries Card(s)

(5) 1 set Vector Entries Card(s)

(6) 1 F Critical Value Card for Testing Lack-of-Fit

(7) 1 KF Table Identification Card

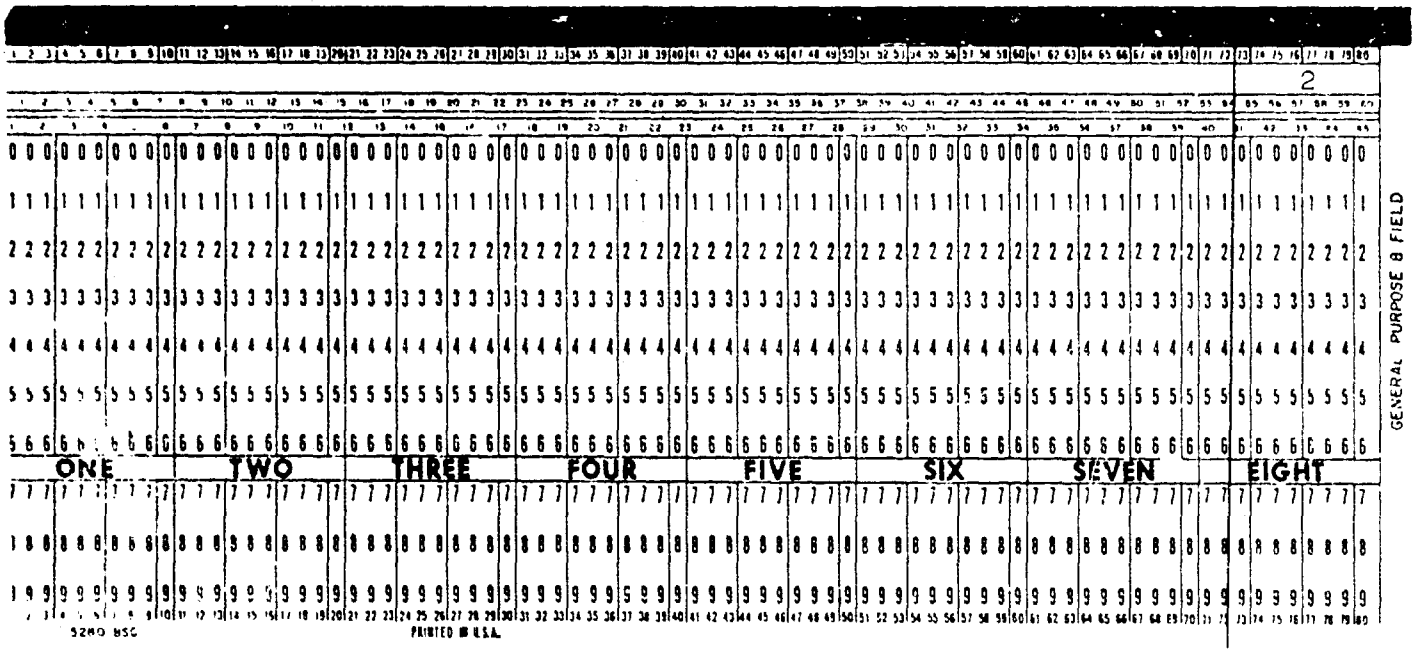
(8) 1 set F Table Entries Card(s)

(9) 1 End of Deck Indicator Card

The cards must be input in the order listed above, and a complete set must be inserted for each run desired.

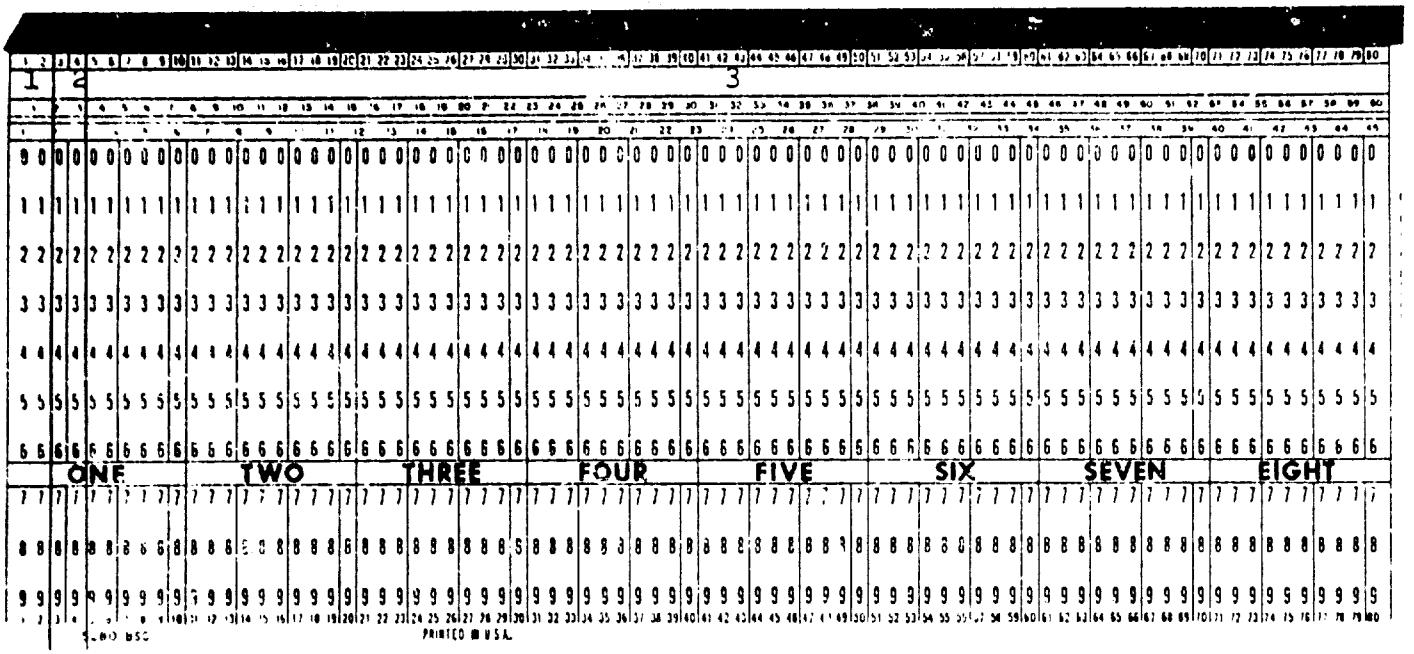
The input card formats and descriptions are as follows:

TITLE CARD



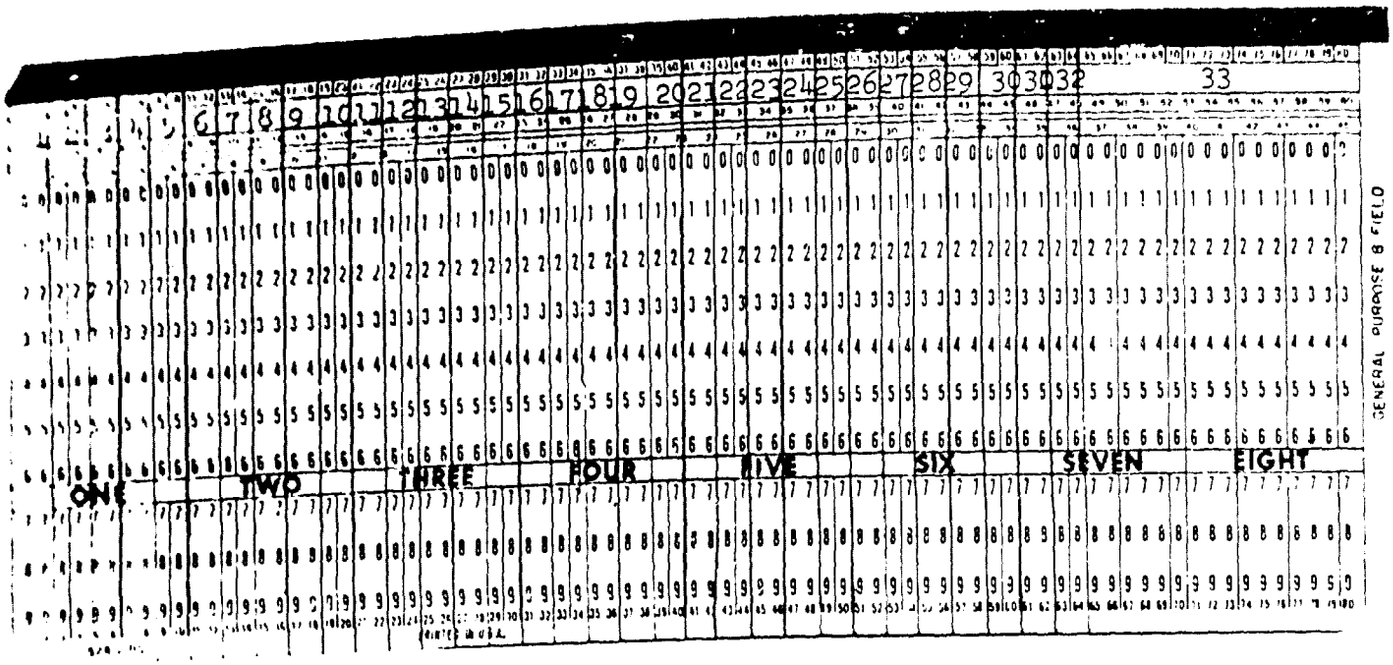
Field	Column	Description
1	1-72	Any alphanumeric character set; printed out as run description.
2	73-80	Not used.

CONTROL PARAMETERS CARD



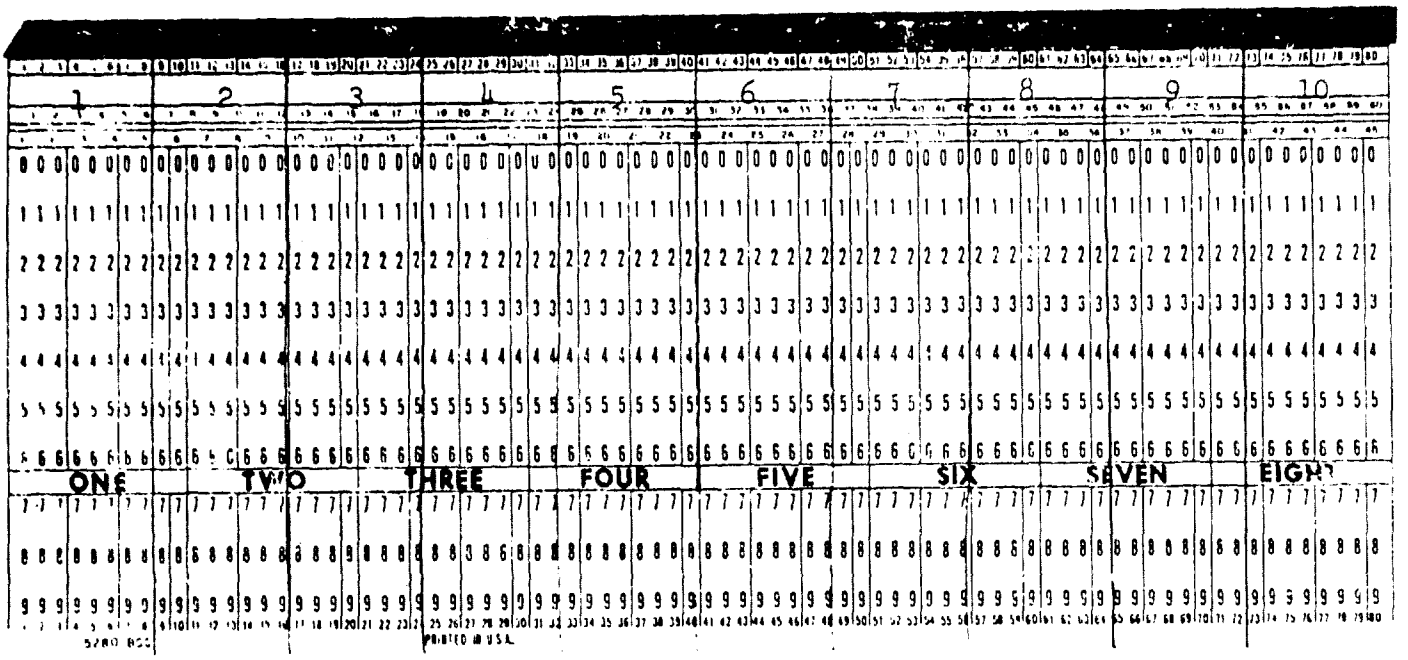
Field	Column	Description
1	1-2	NROW - number of rows in input matrix; Maximum value = 30.
2	3-4	NCOL - number of columns in input matrix; Maximum value = 30.
3	5-80	Not used.

REPLICATIONS IDENTIFICATION CARD



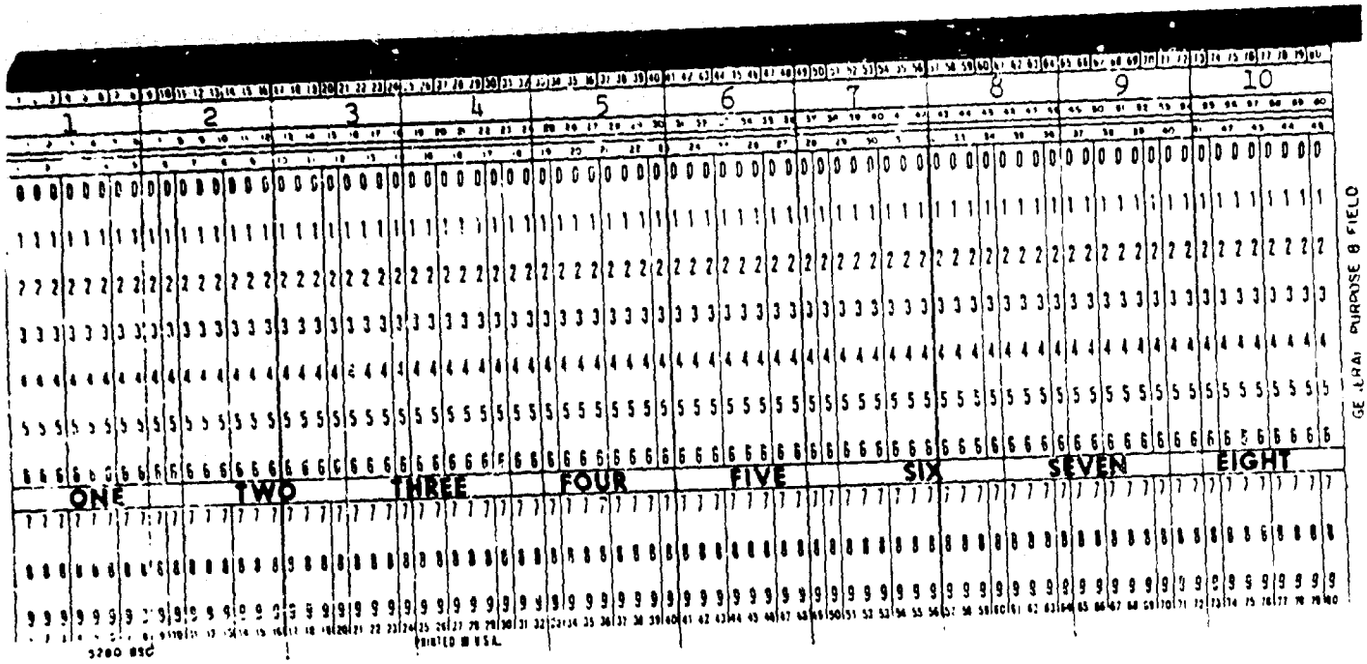
Field	Column	Description
1	1-2	K - number of partitions of Y vector.
2	3-4	NREP array - K values, each of which represents the number of replications within each partition of the Y vector.
3	5-6	
4	7-8	
5	9-10	
.	.	
.	.	
32	63-64	
33	65-80	Not used.

MATRIX ENTRIES CARD(S)



Field	Column	Description
1	1-8	Entries for input matrix in row order. All ten fields of each card must be filled, with the exception of the last card, which may or may not be filled, depending on the dimensions of the matrix. For example, a 5x5 matrix would have 3 matrix entries cards. Card 1 would contain the values in Rows 1 and 2; Card 2 would contain the values in Rows 3 and 4; and Card 3 would contain the values in Row 5 (with the last 5 fields blank).
2	9-16	
3	17-24	
4	25-32	
5	33-40	
6	41-48	
7	49-56	
8	57-64	
9	65-72	
10	73-80	

VECTOR ENTRIES CARD(S)

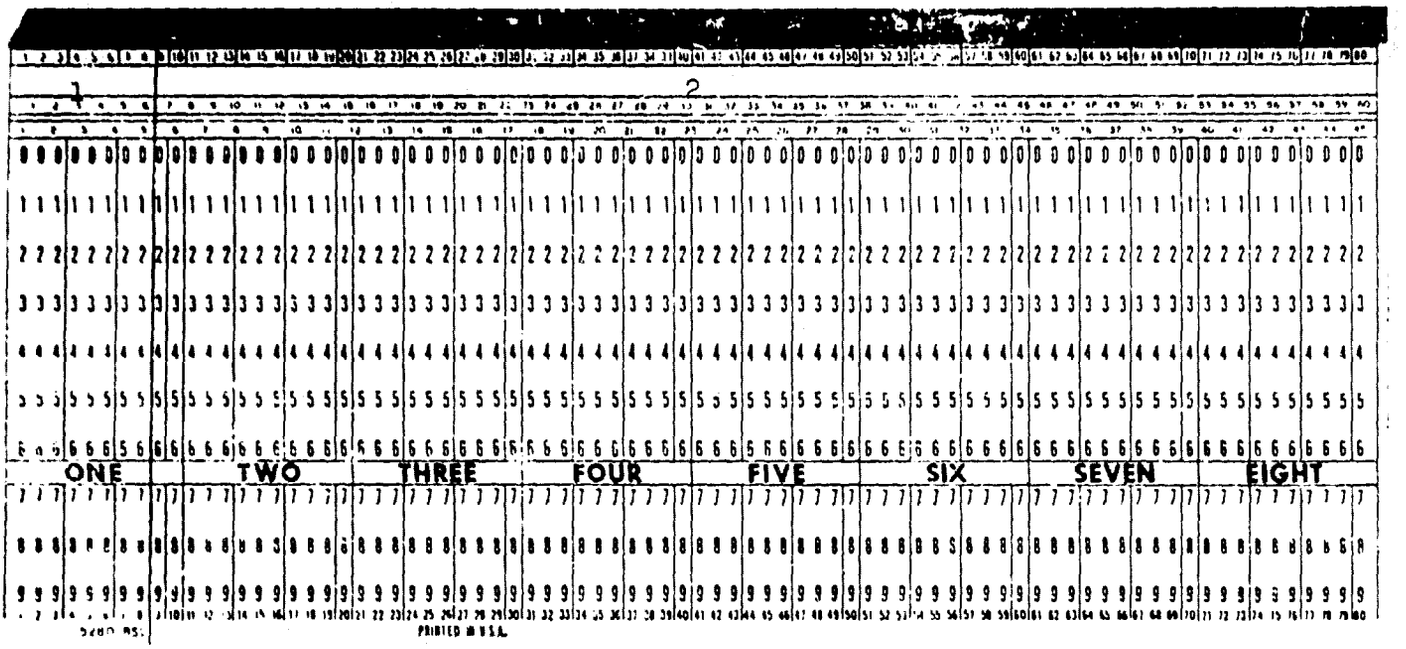


Field	Column
1	1-8
2	9-16
3	17-24
4	25-32
5	33-40
6	41-48
7	49-56
8	57-64
9	65-72
10	73-80

Description

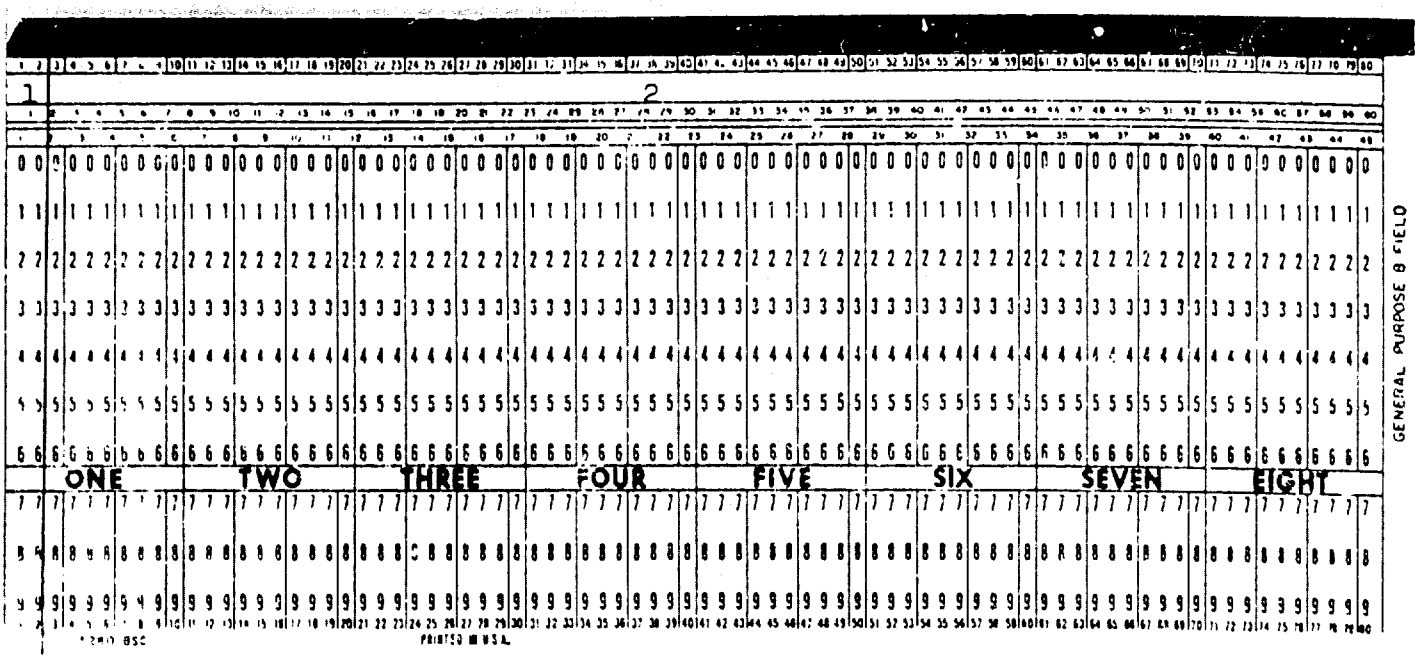
Components of input vector; all ten fields must be filled with the exception of the last card, which may or may not be filled, depending on the dimension of the vector.

F CRITICAL VALUE CARD FOR TESTING LACK-OF-FIT



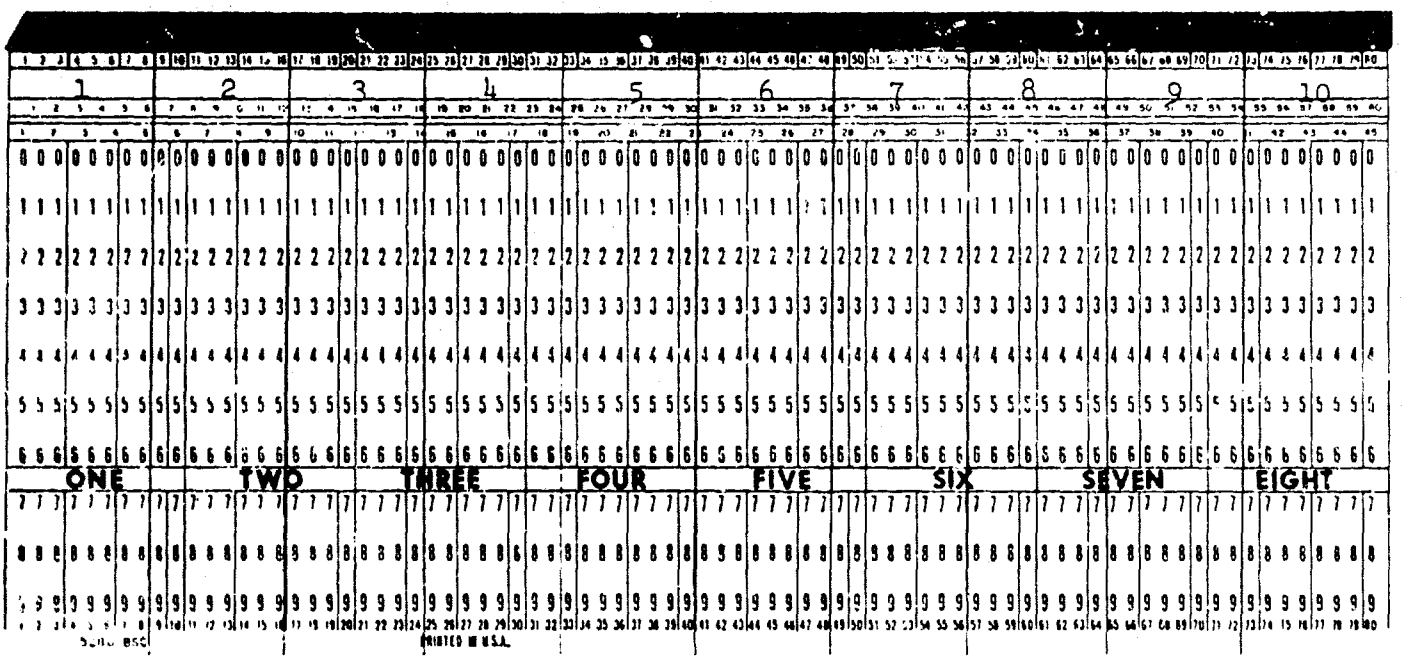
Field	Column	Description
1	1-8	Preselected significance level critical point (from F Table).
2	9-80	Not used.

KF TABLE IDENTIFICATION CARD



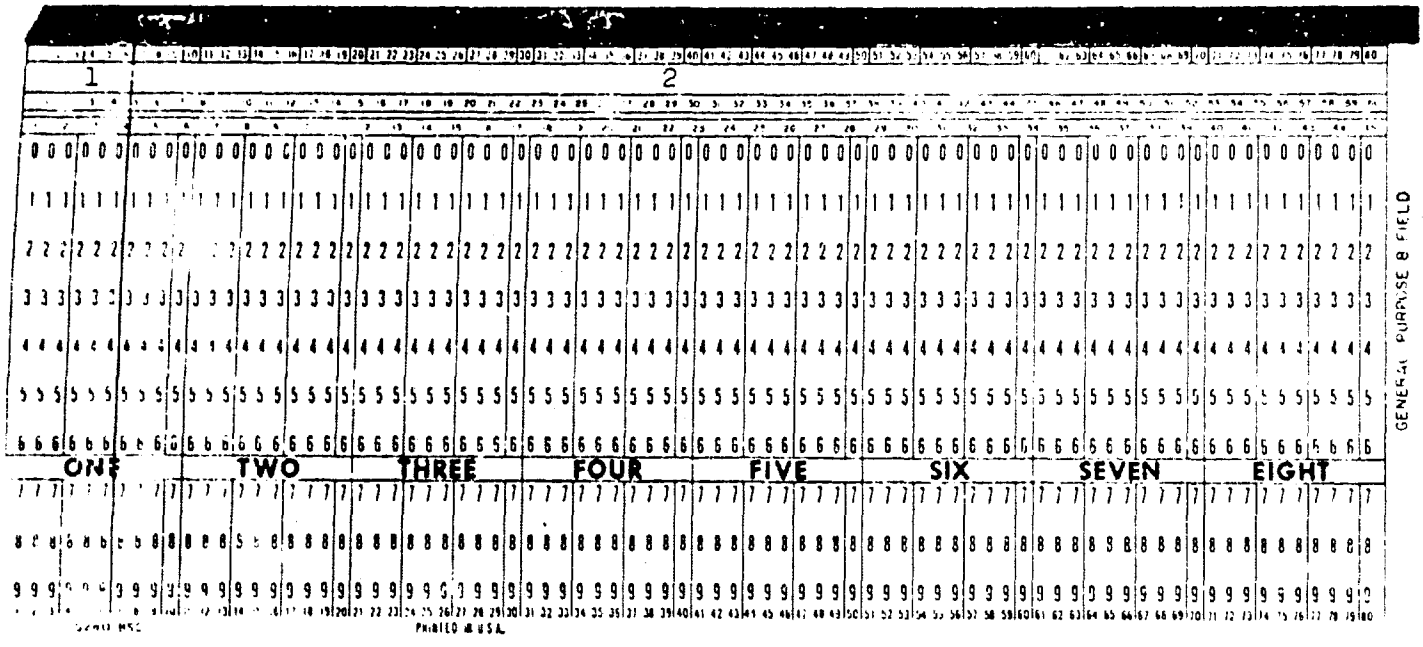
Field	Column	Description
1	1-2	KF - Number of F Table entries to be made.
2	3-80	Not used.

F TABLE ENTRIES CARD(S)



Field	Column	Description
1	1-8	Values from F Table using 1 and n degrees of freedom. Entries must begin with n=1 and may go as far as n=20. (The upper limit is defined as KF in the previous card.)
2	9-16	
3	17-24	
4	25-32	
5	33-40	
6	41-48	All ten fields of each card must be filled, with the exception of the last card, which may or may not be filled, depending on the value of K.
7	49-56	
8	57-64	
9	65-72	
10	73-80	

END OF DECK INDICATOR CARD



Field	Column	Description
1	1-6	Indicates procedure program should follow: Value = EOJbbb if this is the last or only data set. Value = EORbbb if another data set follows.
2	7-80	Not used.

Section III. FORTRAN IV SOURCE STATEMENTS

```
REGR
    MAIN        - EFN    SOURCE STATEMENT  -  IFN(S)  -

      COMMON A(30,30),F(30),FTABL(20),WORK(60), FLF,FCR2,NIND,KIND,NPASS
     1       ,NCOL,NN),M,IDET7,IDEPL,IF,ISTOP,INDEX(10),INDIC(20)
     2       ,MS,MN1,IHLD(60,
      DIMENSION TITLE(12),Y(30),TA(20,20),XM(30,60),
     1          XMY(30),B(30),XNEW(30,30),BNEW(30),
     2           TAR(30,30),XMR(30),SSR1(30),T(30),XREP(30),BTX(30)
     3          ,TANEW(30,30),XMN(30,30),XMYN(30),NREP(30),MALPHA(20)
      DOUBLE PRECISION X,Y,TA,XM,WORK,C,XMY,B,SSREG,XNEW,BNEW,TXR,XMR,
     *          SSR,SSR1,T,SUMY,SUMT,SSPE,XDFPE,RMSE,F,YSUM,SSLF,XMSLF,
     1          FLF,XREP,BTX,SSE,XMYN,TXNEW,XMN,RMSR,FTABL,FCR2
     C          ,RMS,YS,XCOL
      DATA EOJ/6HEOJ   /
      DATA (MALPHA(I),I=1,20)/0,1,2,3,4,5,6,7,8,9,10,11,12,13,14,15,16,
     * 17,18,19/
    1 FORMAT(12A6)
    2 FORMAT(2I2)
    3 FORMAT(3I2)
    4 FORMAT(1CD8.0)
    5 FORMAT(1H1,9X,17HGRAND BETA VECTOR  )
    6 FORMAT(14X,F12.4)
    7 FORMAT(////,11X,4HBETA,I2,7H VECTOR  )
   12 FORMAT(A6)
   13 FORMAT( 1H1,24X,46HPERCENTAGE VARIATION EXPLAINED BY REGRESSION -
     2  ,F12.4)
   14 FORMAT(//////////,24X,44HVECTOR OF UNKNOWN PARAMETERS WITH VARIABL
     *E X ,I2,8H DELETED )
   16 FORMAT(40X,F12.4)
   17 FORMAT(1H1,10X,12A6)
   18 FORMAT(///)
  123 FORMAT(1H1,10X,24HNCOL GT 5, CHECK FORMATS )
  180 FORMAT(1H1,59X,5HANOVA)
  181 FORMAT(////,29X,74HSOURCE OF           DEGREES OF           SUM OF
     1          MEAN              F )
  182 FORMAT(29X,76HVARIATION              FREEDOM              SQUARES
     2   SQUARE           RATIO )
  183 FORMAT(/,24X,84H-------------------------------------------------
     3--------------------------------)
  184 FORMAT(///,24X,19HTOTAL (UNCORRECTED),5X,I2,11X,F12.4)
  185 FORMAT(///,24X,18HTOTAL REGRESSION  ,6X,I2,11X,F12.4)
  186 FORMAT(//,27X,6HDUE TO B,I2,14H,REST       1 ,13X,F10.4,5X,F10.4,5
     5X,F10.4)
  187 FORMAT(///,24X,8HRESIDUAL,16X,I2,13X,F10.4)
  188 FORMAT(//,27X,11HLACK OF FIT,10X,I2,13X,F10.4,5X,F10.4,5X,F10.4)
  189 FORMAT(//,27X,11HPURE ERROR ,10X,I2,13X,F10.4,5X,F10.4)
  191 FORMAT(1H1,29X,37HBETA HAT VARIANCE - COVARIANCE MATRIX )
  201 FORMAT(//,42X,F10.4)
  402 FORMAT(//,35X,2(F10.4,5X))
  403 FORMAT(//,27X,3(F10.4,5X))
  404 FORMAT(//,20X,4(F10.4,5X))
  405 FORMAT(//,13X,5(F10.4,5X))
      CALL ELIM                                                        2
12000 DO 23 I = 1,20
   23 INDIC(I) = MALPHA(I)
C     INITIALIZE
      MS = 1
```

Fortran IV Source Statements (contd)

```
REGR
     MAIN        - EFN   SOURCE STATEMENT  - IFN(S)  -

C      MN1 IS THE ROW DIMENSION OF X AS DEFINED IN THIS ROUTINE
       MN1 = 30
       NPASS = 0
       ISTOP = 0
       NIND = 0
       NFLAG = 0
C      READ INPUTS
       READ(5,1) (TITLE(I),I=1,12)                                          17
       WRITE(6,17) (TITLE(I),I=1,12)                                        24
       READ(5,2) NROW,NCOL
C      K MUST BE LE 30                                                      31
       READ(5,3)K,(NREP(I),I=1,K)                                           34
       READ(5,4) ((X(I,J),J=1,NCOL),I=1,NROW)                               42
       READ (5,4)(Y(I),I=1,NROW)
C      FCR2 = 0 IF ALL NREP = 1                                             52
       READ(5,4) FCR2                                                       59
       READ(5,2) KF                                                         60
       READ(5,4) (FTABL(I),I=1,KF)                                          62
       NC = NCOL
       DO 22 I = 1,10
    22 INDEX(I) = 0
C      CHECK NREP(I).  IF ALL = 1, SET NIND TO 1.
       DO 25 I = 1,K
       IF(NREP(I).NE.1) GO TO 35
    25 CONTINUE
       NIND = 1
    35 CONTINUE
C      ZERO OUT ARRAYS
11111  DO 10 I = 1,30
       DO 10 J = 1,30
       AMY(I)= 0.00
       B(I)=0.00
       XME(I,J)=0.00
       XMN(I,J) = 0.00
       XMYM(I) = 0.00
       SSR1(I)=0.00
       BTA(I) = 0.00
       T(I)=0.00
    10 CONTINUE
       DO 15 I = 1,30
       DO 15 J = 1,60
       XM(I,J) = 0.00
    15 WORK(J) = 0.00
       YSUM=0.00
       IDFRE=0
       SSREG=0.00
       INDI=0
       IL=0
       SUMY=0.00
       SUMT=0.00
       IFLAG = 1
       ICNT = 0
       YS = 0.00
C
C      COMPUTE BETA VECTOR
C
```

Fortran IV Source Statements (contd)

```
REGR
MAIN        - EFN   SOURCE STATEMENT  - IFN(S)  -

C       TRANSPOSE INPUT MATRIX
        DO 100 I =1,NCOL
        DO 100 J =1,NROW
  100 TX(I,J) = X(J,I)
C       MULTIPLY INPUT MATRIX BY ITS TRANSPOSE
        DO 200 I =1,NCOL
        DO 200 J =1,NCOL
        DO 200 L =1,NROW
  200 XM(I,J) = XM(I,J) + TX(I,L) * X(L,J)
C       DETERMINE INVERSE OF XM
        CALL SINV(XM,NCOL,NCOL,MS,MN1,D,R,E,WORK,IHLD)                    149
C       MULTIPLY Y VECTOR BY X TRANSPOSE
        DO 300 I =1,NCOL
        DO 300 J =1,NROW
  300 XMY(I) = XMY(I) + TX(I,J) * Y(J)
C       MULTIPLY XM INVERSE TIMES XMY
        DO 400 I = 1,NCOL
        DO 400 J = 1,NCOL
  400 B(I) = B(I) + XM(I,J) * XMY(J)
        IF(IFLAG.EQ.1) GO TO 2222                                         180
 31111 WRITE(6,5)
        WRITE(6,5)(B(I),I =1,NCOL)
C
C
C       GENERAL REGRESSION SIGNIFICANCE TEST
C                                                                         181
C       COMPUTE SSREG FOR ALL VALUES OF BETA
        DO 450 J = 1,NROW
        DO 450 I = 1,NCOL
  450 BTX(J) = BTX(J) + B(I) * TX(I,J)
        DO 500 I = 1,NROW
  500 SSREG = SSREG + BTX(I) * Y(I)
C
C       COMPUTE REGRESSION DUE TO EACH BETA(I)
C
C       DELETE ONE COLUMN FROM X MATRIX
 20000 SSR = 0.00
        DO 725 I = 1,NCOL1
        BTX(I) = 0.00
        XMY(I) = 0.00
        DO 725 J = 1,NCOL1
        XMN(I,J) = 0.00
  725 CONTINUE
        DO 730 I = 1,NROW
  730 XMN(I) = 0.00
        NCOL1 = NCOL - 1
        DO 550 I = 1,NROW
  550 XMN(I) = 0.00
        IF(IFLAG .NE. 1) GO TO 575
C       DETERMINE XNEW FOR GO
        DO 525 I = 1,NROW
        DO 525 J = 1,NCOL1
  525 XNEW(I,J) = X(I,J+1)
        GO TO 720
  575 IND1 =IND1 + 1
        DO 600 I = 1,NROW
```

Fortran IV Source Statements (contd)

```
REGR
    MAIN        -  SEN   SOURCE STATEMENT  -  IFN(S)  -

      DO 600 J = 1,IND1
  600 XNEW(I,J) = X(I,J)
      IND2 = IND1 + 1
      IF(IND2.EQ.NCOL) GO TO 720
      DO 700 I = 1,NROW
      DO 700 J = IND2,NCOL1
  700 XNEW(I,J) = X(I,J+1)
C
C     COMPUTE NEW BETA VECTOR
C     TRANSPOSE XNEW
  720 DO 750 I = 1,NROW
      DO 750 J = 1,NCOL1
  750 TXNEW(J,I) = XNEW(I,J)
C     MULTIPLY XNEW BY ITS TRANSPOSE
      DO 775 I = 1,NCOL1
      DO 775 J = 1,NCOL1
      DO 775 L = 1,NROW
  775 XMN(I,J) = XMN(I,J) + TXNEW(I,L) * XNEW(L,J)
C     DETERMINE INVERSE OF XMN
      CALL SINV(XMN,NCOL1,NCOL1,MS,PN1,D,R,E,WORK,IHLD)
C     MULTIPLY Y VECTOR BY TXNEW
      DO 780 I = 1,NCOL1
      DO 780 J = 1,NROW
  780 XMYN(I) = XMYN(I) + TXNEW(I,J) * Y(J)
C     MULTIPLY XMN INVERSE BY XMYN
      DO 785 I = 1,NCOL1
      DO 785 J = 1,NCOL1
  785 BNEW(I) = BNEW(I) + XMN(I,J) * XMYN(J)
      IF(NPASS.EQ.0) GO TO 791
      IL4 = IL + 1
      WRITE(6,7)INDIC(IL4)
      GO TO 792
  791 WRITE(6,7) IL
  792 WRITE(6,6) (BNEW(I),I=1,NCOL1)
C     COMPUTE REGRESSION FOR NEW X-S AND BETA-S
      DO 800 I = 1,NCOL1
      DO 800 J = 1,NROW
  800 TXR(I,J) = XNEW(J,I)
      DO 900 J = 1,NROW
      DO 900 I = 1,NCOL1
  900 XMR(J) = XMR(J) + BNEW(I) * TXR(I,J)
      DO 1000 I = 1,NROW
 1000 SSR = SSR + XMR(I) * Y(I)
      IL = IL + 1
      IFLAG = IFLAG + 1
      SSR(IL) = SSNEG -SSR
      IF(IL.EQ.NCOL ) GO TO 1100
      GO TO 10000
 1100 IF(NUMD .EQ. 1) GO TO 30
C
C     DETERMINE PURE ERROR
C      CALCULATE T DOT VALUES
      LL = 1
      LU = 0
      DO 1300 KK = 1,K
      NAP = NREP(KK) + LU
```

```
REGR
     MAIN        -  EPS   SOURCE STATEMENT  -  IFN(S)  -

      LU = NXP
      DO 140 I = LL,LU
  140 T(KK) = T(KK) + Y(I)
      LL = LL + NREP(KK)
  150 CONTINUE
C
C     CALCULATE SUM SQUARES DUE TO PURE ERROR
      DO 155 I = 1,NROW
  155 SUMY = SUMY + Y(I)**2
      DO 156 I = 1,K
  156 XNREP(I) = NREP(I)
      DO 157 I = 1,K
  157 SUMT = SUMT +(T(I)**2)/XNREP(I)
      SSPE = SUMY - SUMT
C     COMPUTE DEGREES OF FREEDOM FOR SSPE
      DO 170 KK = 1,K
  170 IDFPE = IDFPE + NREP(KK) - 1
C     COMPUTE MEAN SQUARE ERROR
      XDFPE = IDFPE
      RMSE = SSPE/XDFPE
C
C     APPLY F TEST
C
   33 CONTINUE
C     COMPUTE SSE AND SSLF
      DO 190 I = 1,NROW
  190 YSUM = YSUM + Y(I)**2
      SSE = YSUM -SSREG
      IDFTE = NROW - NCOL
      IF(IDFPE.EQ.1) GO TO 33
      SSLF = SSE - SSPE
C     CALCULATE D.F. FOR TOTAL ERROR AND LACK OF FIT - PERFORM F TEST
      IDFLF = IDFTE - IDFPE
      XDFLF = IDFLF
      XMSLF = SSLF/XDFLF
      FLF = XMSLF/RMSE
      DO 200 I = 1,NCOL
  200 F(I) = SSR1(I)/RMSE
      GO TO 32
   33 XDFTE = IDFTE
      XMSE = SSE / XDFTE
      DO 34 I = 1,NCOL
   34 F(I) = SSR1(I) / RMSE
C
C      CALCULATE RSQ
C
   32 DO 35 I = 1,NROW
   35 YS = YS + Y(I)
      YS = YS**2
      XNUM = NROW
      RSQ = (SSREG - YS/XNROW) / (YSUM - YS/XNROW)
      RSQ = 1.0 * RSQ
      WRITE(6,180)
      WRITE(6,181)
      WRITE(6,182)
      WRITE(6,183)
```

Fortran IV Source Statements (contd)

```
      MAIN      - EFN   SOURCE STATEMENT  - IFN(S) -

      WRITE(6,184) ARGW,YSUM
      WRITE(6,185) NCOL,SSREG
      DO 193 I = 1,NCOL
  193 WRITE(6,186) INDIC(I),SSR1(I),SSR1(I),F(I)
  197 WRITE(6,187) IDFTE,SSE
      IF(XIND.EQ.1) GO TO 1101
      WRITE(6,188) IDFLF,SSLF,XMSLF,FLF
      WRITE(6,189) IDFPE,SSPE,PMSE
 1101 WRITE(6,13) RSQ
      CALL FTEST
      IF(ISTOP.EQ.1) GO TO 121
      WRITE(6,14) INDEX(NPASS)
      NFLAG = 1
      GO TO 11111
 2222 WRITE(6,16) (B(I),I=1,NCOL)
      NFLAG = 0
      GO TO 31111
C
C       WRITE BETA HAT VARIANCE - COVARIANCE MATRIX
  121 WRITE(6,191)
      IF(NCOL.GT.5) GO TO 122
      GO TO (201,202,203,204,205),NCOL
  201 DO 301 I = 1,NCOL
  301 WRITE(6,401) (XM(I,J),J=1,NCOL)
      GO TO 120
  202 DO 302 I = 1,NCOL
  302 WRITE(6,402) (XM(I,J),J=1,NCOL)
      GO TO 120
  203 DO 303 I = 1,NCOL
  303 WRITE(6,403) (XM(I,J),J=1,NCOL)
      GO TO 120
  204 DO 304 I = 1,NCOL
  304 WRITE(6,404) (XM(I,J),J=1,NCOL)
      GO TO 120
  205 DO 305 I = 1,NCOL
  305 WRITE(6,405) (XM(I,J),J=1,NCOL)
      GO TO 120
  122 WRITE(6,123)
  120 READ(5,12) XIND
      IF(XIND.EQ.EOJ) GO TO 3333
      GO TO 12000
 3333 STOP
      END
```

```
      REAL
     1    FTST        - REAL   SOURCE STATEMENT  -  IFN(S)  -


      SUBROUTINE FTEST
      COMMON X(30,3),F(3),FTABLE(20),WORK(30), FLF,FCR2,NIND,KIND,NPASS
     1       ,NCOL,NPAR,IOPTS,IOPTC,IF,ISTOP,INDEX(10),INDIC(20)
     2       ,S,R,I,IHLS(30)
      DIMENSION XN(30,30)
      DOUBLE PRECISION A,F,FTABLE,FCR2,XN,FL,FLF,D,WORK
    1 FORMAT(1H1,10X,'ENLARGE  DF FIT SIGNIFICANT, RE-EXAMINE THE REGRESSI
     1ON EQUATION  ')
    2 FORMAT(1H1,10X,'CONGRATULATIONS, YOUR REGRESSION HAS BEEN FOUND
     1ADEQUATE  ')
    3 FORMAT(/////////////,24X,'INVARIABLE LEAVING X,I2)
    4 FORMAT(1H1,'FTABLE NOT SUFFICIENT, INDX =  ,I3)
      ICNT = 0
      NPASS = NPASS + 1
      IF(NPASS.GT.1) GO TO 5.
C
C     CHECK FOR SIGNIFICANCE OF LACK OF FIT
C
      IF(NPASS.GT.1) GO TO 10
      IF(FLF.GT.FCR2) GO TO 20
      GO TO 15
   20 WRITE(6,1)
   10 CONTINUE
C
C     DETERMINE SMALLEST F RATIO
C
    5 FL = F(2)
      IF = 2
      DO 30 I = 3,NCOL
      IF(F(I).GT.FL) GO TO 30
      FL = F(I)
      IF = I
   30 CONTINUE
      INDEX(NPASS) = IF + NPASS - 2
      IF(NPASS.EQ.1) GO TO 35
      NP = NPASS - 1
      DO 33 I = 1,NP
      IF(INDEX(I).GE.INDEX(NPASS)) ICNT = ICNT + 1
   33 CONTINUE
      INDEX(NPASS) = INDEX(NPASS) - ICNT
C
C     COMPARE FL WITH CRITICAL VALUE FROM FTABL
C
   35 CONTINUE
      DO 37 I = 1,NCOL
      IF(I.NE.INDEX(NPASS).OR.INDIC(I)) GO TO 37
      IFN = I
      GO TO 39
   37 CONTINUE
C     DELETE INDEX FROM INDIC ARRAY
   38 NCL = NCOL - 1
      DO 34 I = IFN,NCL
   34 INDIC(I) = INDIC(I+1)
      IF(IFN.EQ.1) GO TO 60
      INDX = IOPTS
```

Fortran IV Source Statements (contd)

```
REGR
    FTST        - EPN   SOURCE STATEMENT  - IFN(S)  -

      GO TO 70
   60 INDX = IDFTE
   70 CONTINUE
      IF(FTABL(INDX).EQ.0.D0) GO TO 160
      IF(FL.LT.FTABL(INDX)) GO TO 80
      IF(NPASS.NE.1) GO TO 65
      WRITE(6,2)
    2 FORMAT(' ALL VARIABLES HAVE BEEN DELETED')
   65 ISTOP = 1
   80 CONTINUE
C
C     DELETE COLUMN FROM X MATRIX
      IND = IF - 1
      DO 90 I = 1,NROW
      DO 90 J = 1,IND
   90 XN(I,J) = X(I,J)
      NCOL = NCOL - 1
      DO 100 I = 1,NROW
      DO 100 J = IF,NCOL
  100 XN(I,J) = X(I,J+1)
      DO 110 I = 1,NROW
      DO 110 J = 1,NCOL
  110 X(I,J) = XN(I,J)
      IF(ISTOP.EQ.1) GO TO 120
      WRITE(6,3) INDEX(NPASS)
      GO TO 120
C
  160 WRITE(6,4) INDX
      ISTOP = 1
  120 RETURN
      END
```

```
REGN
     SESJSS     - LEV   SOURCE STATEMENT  - IFN(S)  -


C      SUBROUTINE SESCMI(X,N,NB,MS,MN1,D,K,E,WORK,IHLD)
C      CATEGORY
C          MATHEMATICAL
C      PURPOSE
C          SOLVE SIMULTANEOUS EQUATIONS OR INVERT MATRICIES
C      DESCRIPTION
C          THIS SUBROUTINE WILL SOLVE AN N BY N SYSTEM OF SIMULTANEOUS
C          EQUATIONS WITH AN ARBITRARY NUMBER OF RIGHT HAND SIDES OR
C          INVERT A MATRIX OF ORDER N.  IN THE PROCESS, THE RANK OF
C          THE MATRIX AND ITS DETERMINANT ARE EVALUATED.  THE METHOD
C          USED IS THAT OF GAUSS-JORDAN WITH TOTAL PIVOTING.  THE
C          ROUTINE MAY BE RUN IN SINGLE OR DOUBLE PRECISION ARITHMETIC.
C      INPUT
C          1  X        FIRST LOCATION OF INPUT COEFFICIENT MATRIX,X(1,1)
C                      AUGMENTED BY NB RIGHT HAND SIDES.  FOR MATRIX
C                      INVERSE, X IS FIRST LOCATION OF THE MATRIX TO BE
C                      INVERTED.  I.E. X(1,1).  X MUST BE DIMENSIONED
C                      TO (MN1,MN1+NB) IN THE CALLING PROGRAM IN EITHER
C                      CASE.
C          2  N        NUMBER OF SIMULTANEOUS EQUATIONS TO BE SOLVED,
C                      OR ORDER OF MATRIX TO BE INVERTED.
C          3  NB       NB = NUMBER OF RIGHT HAND SIDES FOR SIMULTANEOUS
C                      EQUATION SOLUTION.  NB = N FOR MATRIX INVERSE.
C          4  MS       MS = 0 FOR SIMULTANEOUS EQUATION SOLUTION.
C                      MS = 1 FOR MATRIX INVERSE.
C          5  MN1      ROW DIMENSION OF X AS DEFINED IN CALLING PROGRAM.
C          6  WORK     WORKING ARRAY DIMENSIONED AS FOLLOWS IN CALLING
C                      PROGRAM... WORK(MN1*NB).
C          7  IHLD     WORKING ARRAY DIMENSIONED AS FOLLOWS IN CALLING
C                      PROGRAM... IHLD(MN1).
C      OUTPUT
C          1  X        X(1,1) THROUGH X(N,1) CONTAIN FIRST SOLUTION
C                      VECTOR.  X(1,2) THROUGH X(N,2) CONTAIN SECOND
C                      SOLUTION VECTOR, ETC.  FOR MATRIX INVERSE, THE
C                      ARRAY X CONTAINS THE INVERSE MATRIX.
C          2  D        DETERMINANT OF INPUT X.
C          3  K        RANK OF INPUT X.
C          4  E        ERROR CHECK.  IF E = 0. O.K., IF E=1., THE
C                      MATRIX OF COEFFICIENTS ARE SINGULAR.  IF E = 2.
C                      SOLUTION IS ATTEMPTED BUT EQUATIONS MAY BE
C                      SINGULAR OR ILL CONDITIONED.
C      REMARKS
C      REFERENCE
C          INTERNAL TECHNICAL NOTE 40
C          DOUBLE PRECISION SIMULTANEOUS EQUATION SOLVER OR MATRIX
C          INVERSION ROUTINE NO. 1.15 AUG 1963.
C          B. G. GIBBS
C          SCIENTIFIC DIGITAL PROGRAMMING BRANCH
C          ARMY COMPUTATION CENTER
C          REDSTONE ARSENAL, ALABAMA
C      CONTACT
C          CHIEF,THEORETICAL PROBLEMS SECTION
C          SCIENTIFIC DIGITAL PROGRAMMING BRANCH
C          ARMY COMPUTATION CENTER
C
```

Fortran IV Source Statements (contd)

```
HEGR
     SESDSB     - FPN   SOURCE STATEMENT  -  IPN(S)  -

      SUBROUTINE SESOM1(X,N,NB,MS,MN1,C,R,E,WORK,IHLD)
      DIMENSION X(MN1,1),WORK(1),IHLD(1)
C        IF DOUBLE PRECISION IS DESIRED INSERT THE FOLLOWING CARD
C     DOUBLE PRECISION X,WORK,Y,D,SUM,X1
      DOUBLE PRECISION X,WORK,Y,D,SUM,X1
      X1=1.                                                          SES00220
      E=0.                                                           SES00230
      R=0.                                                           SES00240
      DO 21 I=1,N                                                    SES00260
   21 IHLD(I)=I                                                      SES00270
      IF(MS)6,4,6                                                    SES00280
    6 MM=M+N                                                         SES00290
      NB=M                                                           SES00300
      M1=M+1                                                         SES00310
      DO 14 I=1,N                                                    SES00320
      DO 14 J=M1,MM                                                  SES00330
   14 X(I,J)=0.00                                                    SES00340
      DO 15 I=1,N                                                    SES00350
      J=I+M                                                          SES00360
   15 X(I,J)=1.00                                                    SES00370
      GO TO 16                                                       SES00380
    4 MM=M+NB                                                        SES00390
   16 JJ=M+N                                                         SES00400
      MMN=M-1                                                        SES00420
      D=1.00                                                         SES00430
      DO 36 I=1,N
      Y=X(I,1)
      DO 35 J=2,N
      IF(ABS(Y).LT.ABS(X(I,J)))Y=X(I,J)
   35 CONTINUE
      D=D*Y
      DO 36 J=1,MM
   36 X(I,J)=X(I,J)/Y
      DO 5 I=1,M                                                     SES00440
      KK=M-I                                                         SES00450
      IF(KK)10,10,26                                                 SES00460
   26 LL=KK+1                                                        SES00470
      IJJ=1                                                          SES00480
      L=I                                                            SES00490
      WORK=X                                                         SES00500
      DO 17 II=1,LL                                                  SES00510
      DO 17 J=1,LL                                                   SES00520
      IF(ABS(WORK)-ABS(X(II,J)))18,17,17                             SES00530
   18 WORK=X(II,J)                                                   SES00540
      L=J+I-1                                                        SES00550
      IJJ=J                                                          SES00560
   17 CONTINUE                                                       SES00570
      IF(IJJ-1)2,2,19                                                SES00580
   19 DO 20 II=1,N                                                   SES00590
      Y=X(II,1)                                                      SES00600
      X(II,1)=X(II,IJJ)                                              SES00610
   20 X(II,IJJ)=Y                                                    SES00620
      IY=IHLD(I)                                                     SES00630
      IHLD(I)=IHLD(L)                                                SES00640
      IHLD(L)=IY                                                     SES00650
      D=-D                                                           SES00660
```

Fortran IV Source Statements (concluded)

```
 [illegible]
 [illegible]      -  [illegible]   SOURCE STATEMENT  -  IFN(S)  -

  2 DO 9 L=1,KK                                                       SES00670
    IF(ABS(X)-ABS(X(L+1,1)))7,1,1                                     SES00680
  7 X=[illegible]                                                     SES00690
    DO 9 J=1,JJ                                                       SES00700
    Y=X(L,J)                                                          SES00710
    X(L,J)=X(L+1,J)                                                   SES00720
  9 X(L+1,J)=Y                                                        SES00730
  1 CONTINUE                                                          SES00740
 10 JJ=JJ-1                                                           SES00750
    [illegible]
    IF(X.[illegible])GO TO [illegible]
 31 IF(ABS(ABS((X1-X)/X1)-1.).LT.1.E-7)F=2.
    [illegible]
 11 [illegible]                                                       SES00780
    DO 12 J=1,JJ                                                      SES00790
 12 WORK(J)=X(1,J+1)/X                                                SES00800
    KK=JJ+1                                                           SES00810
    IF([illegible])GO TO 33                                           A7  C815
    DO 3 K=1,[illegible]                                              SES00820
    DO 3 J=2,KK                                                       SES00830
  3 X(K,J-1)=X(K+1,J)-X(K+1,1)*WORK(J-1)                              SES00840
 33 DO 5 J=1,JJ                                                       A7  C845
  5 X([illegible],J)=WORK(J)                                          SES00860
    [illegible]=N-1                                                   SES00870
    IF([illegible])GO TO 13
    DO 22 I=1,[illegible]                                             SES00880
    L=I+1                                                             SES00890
    DO 22 J=L,N                                                       SES00900
    IF(IHLD(I)-IHLD(J))22,22,23                                       SES00910
 23 IY=IHLD(I)                                                        SES00920
    IHLD(I)=IHLD(J)                                                   SES00930
    IHLD(J)=IY                                                        SES00940
    DO 25 K=1,[illegible]                                             SES00950
    Y=X(I,K)                                                          SES00960
    X(I,K)=X(J,K)                                                     SES00970
 25 X(J,K)=Y                                                          SES00980
 22 CONTINUE                                                          SES00990
 13 RETURN                                                            SES01050
  6 E=1.                                                              SES01060
    GO TO 13                                                          SES01070
    END                                                               SES01080
```

Fortran IV Source Statements (concluded)

```
      ISN   SOURCE STATEMENT   -   IFN(S)   -

    2 DO 3 L=1,N                                                     SFSU0670
      IF(ABS(Y)-ABS(A(L+1,I)))7,1,1                                  SFSU0680
    7 J=L                                                            SFSU0690
      DO 6 J=1,JJ                                                    SFS00700
      Y=A(I,J)                                                       SES00710
      A(I,J)=A(L+1,J)                                                SES00720
    6 A(L+1,J)=Y                                                     SFS00730
    3 CONTINUE                                                       SFS00740
   30 JJ=JJ-1                                                        SFS00750
      J=0                                                               
      IF(A.[illegible])GO TO [illegible]
   31 IF(ABS(ABS(A-X)/X)-1.).LT.1.E-7)F=2.
      [illegible]
   11 [illegible]                                                    SFS00[illegible]
      DO 12 J=1,JJ                                                   SES00[illegible]
   12 [illegible](J)=A(I,J+1)/X                                      SES[illegible]
      [illegible]J+1                                                 SES[illegible]
      IF([illegible])GO TO 31                                        [illegible]
      DO 3 K=1,[illegible]                                           SFS[illegible]
      DO [illegible] J=2,[illegible]                                 SFS[illegible]
      A(K,J-1)=A(K+1,J)-X(K+1,1)*[illegible]K(J-1)                   SFS[illegible]
   33 DO 5 J=1,JJ                                                    [illegible]
    5 A([illegible],J)=[illegible]K(J)                               SES00860
      [illegible]-1                                                  SES00[illegible]
      IF([illegible])GO TO 13
      DO [illegible] I=1,[illegible]                                 SES00880
      L=I+1                                                          SFS00890
      DO 22 J=L,N                                                    SFS00900
      IF(IHLD(I)-IHLD(J))26,22,23                                    SES00910
   23 IY=IHLD(I)                                                     SES00920
      IHLD(I)=IHLD(J)                                                SES00930
      IHLD(J)=IY                                                     SES00940
      DO 25 K=1,NN                                                   SFS00950
      Y=X(I,K)                                                       SES00960
      X(I,K)=X(J,K)                                                  SFS00970
   25 X(J,K)=Y                                                       SES00980
   22 CONTINUE                                                       SES00990
   13 RETURN                                                         SFS01050
    0 E=1.                                                           SES01060
      GO TO 13                                                       SFS01070
      END                                                            SFS01080
```