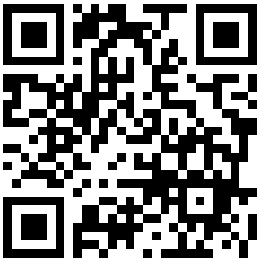
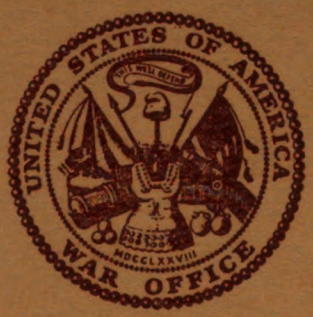


ARO-D Report 69-2

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



Sponsored by

The Army Mathematics Steering Committee

on Behalf of

THE OFFICE OF THE CHIEF OF RESEARCH AND DEVELOPMENT

FOREWORD

In a letter under date of 12 December 1967, Dr. Charles A. Reynolds, Technical Director of Edgewood Arsenal, issued an invitation to hold the Fourteenth Conference on the Design of Experiments in Army Research, Development and Testing at Edgewood Arsenal, Maryland. In his letter, Dr. Reynolds set the dates for this meeting as 23-25 October 1968, and he appointed Messrs. Joseph Mandelson and Raymond Schnell to serve as Co-Chairmen on Local Arrangements. These conferences are sponsored by the Army Mathematics Steering Committee and they come under the supervision of the AMSC Subcommittee on Probability and Statistics. Dr. Walter Foster, the Chairman of this Subcommittee, was happy to accept this invitation and started laying the groundwork for this conference. He and other members of the AMSC would like to thank Messrs. Mandelson and Schnell, as well as many other employees of Edgewood Arsenal, who helped to make the Fourteenth Conference such an enjoyable and successful meeting.

These conferences are open to scientific personnel of all Government agencies, and the participation on the program by staff members of various agencies has been gratifying. In this, and in past meetings, scientists from the National Bureau of Standards have contributed a great deal to the tone of these symposia. It seems appropriate that we point out some of the Bureau participants in this Edgewood Arsenal Conference. Dr. Joseph Cameron served as a member of the Program Committee; and he, along with Dr. Joan R. Rosenblatt, served as panelists in several of the clinical sessions. Messrs. H. H. Ku and Roy H. Wampler each presented technical papers. Further, there was presented a paper which was authored jointly by David Hogben and John Mandel. We are pleased to be able to publish most of these papers in this technical manual.

Those attending the conference had the pleasure of hearing the following invited speakers talk on the topics noted below:

Broadening the Horizons of Experimental Design
Lieutenant General William B. Bunker
U. S. Army Material Command

Structure and Classification of Patterns
Professor Rolf E. Bargmann
University of Georgia

Bulk Sampling
Professor Acheson J. Duncan
Johns Hopkins University

Time Series
Professor Emanuel Parzen
Stanford University

The keynote speaker, General Bunker, died before these Proceedings could be issued. His passing is a heavy loss to the scientific community, and to me, a special loss, as he was a warm personal friend.

An outstanding feature of the program of the Fourteenth Conference was a panel on Bulk Sampling. This is an area of statistics of special interest to the scientific personnel of the host installation. Dr. Walter Foster served as chairman and organizer of this phase of the agenda. He selected Professor A. J. Duncan to serve as a Discussant and Advisor to the following Panel Members: Henry Ellner; Boyd Harshbarger; G. R. Lowrimore; Joseph Mandelson; and, V. H. Rechmeyer. Another outstanding feature of these conferences is the awarding of the Wilks Memorial Medal. This year, it was my pleasure to announce that Professor Jerzy Neyman, of the University of California at Berkeley, was selected to receive the Fourth Samuel S. Wilks Memorial Medal.

Members of the Army Mathematics Steering Committee think that the papers presented at the conference have made valuable contributions to the fields of the design of experiments, statistics, and reliability, and have requested that these articles be published in these Proceedings. They wish to thank the many speakers, chairmen, and panelists for their help in conducting this symposium.

The conference had an attendance of 163 scientists, and 50 organizations were represented. Speakers and panelists came from: Cornell Aeronautical Lab; Duke University; Federal Electric Corporation/ITT; Hercules, Inc.; Johns Hopkins University; National Bureau of Standards; Stanford University; Thiokol Chemical Corporation; University of Chicago; University of Georgia; and Virginia Polytechnic Institute; and nineteen army facilities.

Colonel Paul R. Cerar, Commanding Officer of Edgewood Arsenal, gave the Welcoming Remarks for the host installation. In his talk, he gave many interesting and historical facts about Edgewood Arsenal. His address is published here for the edification of those who were not able to hear him speak.

Formulation of the outstanding features of this conference and the selection of the invited speakers were made by the members of the Program Committee (Joseph Cameron, Francis Dressel, Walter D. Foster, Fred Frishman, Boyd Harshbarger, William Kruskal, H. L. Lucas, Jr., Clifforn Maloney, Joseph Mandelson, Henry Mann, Raymond B. Schnell, and Herbert Solomon). The Chairman wishes these individuals to know that he appreciated their assistance and valued their comments on the various phases of the program.

Frank E. Grubbs
Conference Chairman

U. S. ARMY RESEARCH OFFICE - DURHAM
REPORT NO. 69-2
SEPTEMBER 1969

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

Proceedings of the ... Conference on the Design of Experiments.

SPONSORED BY THE ARMY MATHEMATICS STEERING COMMITTEE

Host

U. S. Army Edgewood Arsenal
Edgewood Arsenal, Maryland

23 - 25 October 1968

FOREWORD

In a letter under date of 12 December 1967, Dr. Charles A. Reynolds, Technical Director of Edgewood Arsenal, issued an invitation to hold the Fourteenth Conference on the Design of Experiments in Army Research, Development and Testing at Edgewood Arsenal, Maryland. In his letter, Dr. Reynolds set the dates for this meeting as 23-25 October 1968, and he appointed Messrs. Joseph Mandelson and Raymond Schnell to serve as Co-Chairmen on Local Arrangements. These conferences are sponsored by the Army Mathematics Steering Committee and they come under the supervision of the AMSC Subcommittee on Probability and Statistics. Dr. Walter Foster, the Chairman of this Subcommittee, was happy to accept this invitation and started laying the groundwork for this conference. He and other members of the AMSC would like to thank Messrs. Mandelson and Schnell, as well as many other employees of Edgewood Arsenal, who helped to make the Fourteenth Conference such an enjoyable and successful meeting.

These conferences are open to scientific personnel of all Government agencies, and the participation on the program by staff members of various agencies has been gratifying. In this, and in past meetings, scientists from the National Bureau of Standards have contributed a great deal to the tone of these symposia. It seems appropriate that we point out some of the Bureau participants in this Edgewood Arsenal Conference. Dr. Joseph Cameron served as a member of the Program Committee; and he, along with Dr. Joan R. Rosenblatt, served as panelists in several of the clinical sessions. Messrs. H. H. Ku and Roy H. Wampler each presented technical papers. Further, there was presented a paper which was authored jointly by David Hogben and John Mandel. We are pleased to be able to publish most of these papers in this technical manual.

Those attending the conference had the pleasure of hearing the following invited speakers talk on the topics noted below:

Broadening the Horizons of Experimental Design
Lieutenant General William B. Bunker
U. S. Army Material Command

Structure and Classification of Patterns
Professor Rolf E. Bargmann
University of Georgia

Bulk Sampling
Professor Acheson J. Duncan
Johns Hopkins University

Time Series
Professor Emanuel Parzen
Stanford University

The keynote speaker, General Bunker, died before these Proceedings could be issued. His passing is a heavy loss to the scientific community, and to me, a special loss, as he was a warm personal friend.

An outstanding feature of the program of the Fourteenth Conference was a panel on Bulk Sampling. This is an area of statistics of special interest to the scientific personnel of the host installation. Dr. Walter Foster served as chairman and organizer of this phase of the agenda. He selected Professor A. J. Duncan to serve as a Discussant and Advisor to the following Panel Members: Henry Ellner; Boyd Harshbarger; G. R. Lowrimore; Joseph Mandelson; and, V. H. Rechmeyer. Another outstanding feature of these conferences is the awarding of the Wilks Memorial Medal. This year, it was my pleasure to announce that Professor Jerzy Neyman, of the University of California at Berkeley, was selected to receive the Fourth Samuel S. Wilks Memorial Medal.

Members of the Army Mathematics Steering Committee think that the papers presented at the conference have made valuable contributions to the fields of the design of experiments, statistics, and reliability, and have requested that these articles be published in these Proceedings. They wish to thank the many speakers, chairmen, and panelists for their help in conducting this symposium.

The conference had an attendance of 163 scientists, and 50 organizations were represented. Speakers and panelists came from: Cornell Aeronautical Lab; Duke University; Federal Electric Corporation/ITT; Hercules, Inc.; Johns Hopkins University; National Bureau of Standards; Stanford University; Thiokol Chemical Corporation; University of Chicago; University of Georgia; and Virginia Polytechnic Institute; and nineteen army facilities.

Colonel Paul R. Cerar, Commanding Officer of Edgewood Arsenal, gave the Welcoming Remarks for the host installation. In his talk, he gave many interesting and historical facts about Edgewood Arsenal. His address is published here for the edification of those who were not able to hear him speak.

Formulation of the outstanding features of this conference and the selection of the invited speakers were made by the members of the Program Committee (Joseph Cameron, Francis Dressel, Walter D. Foster, Fred Frishman, Boyd Harshbarger, William Kruskal, H. L. Lucas, Jr., Clifforn Maloney, Joseph Mandelson, Henry Mann, Raymond B. Schnell, and Herbert Solomon). The Chairman wishes these individuals to know that he appreciated their assistance and valued their comments on the various phases of the program.

Frank E. Grubbs
Conference Chairman

WELCOME*

Colonel Paul R. Cerar
Commanding Officer, Edgewood Arsenal

General Bunker, distinguished guests and speakers, ladies and gentlemen .....

Edgewood Arsenal is proud and gratified to have been chosen to act as your host for this, the Fourteenth Annual Conference on the Design of Experiments in Army Research, Development, Testing and Evaluation. I consider it a privilege to welcome you on behalf of the arsenal and its personnel. It is particularly fitting that our arsenal should be given this opportunity as part of its scientific program for this year of our existence, a half century of work and achievement as a significant element in the defense structure of our country.

In October of 1917 the War Department acquired this reservation, later to become the Infant Gas Warfare Service's first home, and in May 1918 named the installation Edgewood Arsenal. During the lean years between world wars Edgewood Arsenal struggled to prepare the military arm, offensively and defensively, in the area of chemical warfare. Despite the meager resources allotted, especially during the depression years, somehow the installation survived to provide the basic cadre for the enormous expansion to over 7000 military and 8000 civilian personnel in the peak years of World War II. Through their devoted efforts, our military forces were provided with a capability in research, development, procurement and supply of chemical offensive and defensive materiel.

Existing industrial and manufacturing facilities were rehabilitated and new ones built. Necessary support facilities such as utilities, an airstrip, and an expanded rail network were added. The chemical warfare school was expanded and a modern laboratory complex was built to house consolidated research and development activities. In May 1942 the installation was re-designated the Chemical Warfare Center. In August 1946 the name was changed to Army Chemical Center but in 1963 we reverted to the original title: Edgewood Arsenal.

In a re-organization approved 7 July 1966, Edgewood Arsenal was designated the U.S. Army's Chemical Commodity Center with responsibility for all chemical weapons and defense materiel research and development, subordinate to U.S. Army Munitions Command. Its previous administrative control over Fort Detrick was relinquished and Fort Detrick became a separate commodity center with responsibility for biological weapons and defense research and development. However, because certain of our responsibilities overlap those of Fort Detrick the old cooperation between the two installations is still in existence both by necessity and choice.

---

*Colonel Cerar gave the Welcoming Remarks at the start of the Conference and also served as Chairman of General Session I.

Two sub-posts fall under the command jurisdiction of the Edgewood Arsenal Commander: Pine Bluff Arsenal, Arkansas and Rocky Mountain Arsenal, Colorado. These two arsenals are engaged in various aspects of procurement manufacture and testing of chemical materiel.

Over the years, then, Edgewood Arsenal has grown to represent about $115 million in fixed investments, to include $9.6 million in land and improvements; $78.1 million in buildings and facilities; and $27 million in machinery and equipment. These figures do not include our sub-posts. The installation employs over 3,800 civilians and over 1,600 military personnel with a combined gross payroll of some $40 million.

Among our civilian employees more than 900 hold bachelor degrees; over 190 have master degrees; and 75 have attained their doctorates. In connection with the subject which is basic to the purpose of this conference - statistics as it is employed in research, development, testing and evaluation - Edgewood Arsenal can point to a long, and a still growing interest and participation in this highly specialized field. Starting about 1942, statistics of this type began to be used in preparing specification requirements and later in the development of certain theoretical concepts upon which our surveillance and other quality assurance activities are based. Much of this work found its way into the literature and our personnel were actively engaged in the development of important sampling standards. Interest in, and utilization of statistics, soon spread from our quality assurance elements to our research, development and testing activities. At a later date, an Operations Research Group was formed in whose work, as you know, statistical principles play a major role. This group was recently incorporated into the U. S. Army Munitions Command but it remains physically located on this post.

The Chemical Corps Engineering Command sponsored several conferences on Statistical Engineering in the 1950's which some of you may have attended. It has been our policy to encourage our personnel to take an active part in all professional activities - delivering and publishing technical papers and acting as chairmen and moderators of technical sessions.

Our background dates back some 26 years, when, as you may recall, the work of Professors Fisher and Pearson in England on the Design of Experiments and even the work of Shewhart, Dodge, and Romig in this country in Statistical Quality Control were practically unknown. You can see why Edgewood Arsenal feels so proud to act as your hosts for the next three days.

At this point, I am pleased to acknowledge our indebtedness to the Army Research Office and to its arrangements committee for inviting us to host this conference and to extend my thanks through Dr. Francis Dressel, the Secretary, to this committee for the excellent work they have done in securing such outstanding speakers and in arranging so interesting

a technical program. We are especially honored and pleased to have as our keynote speaker, a distinguished soldier who has taken a very keen and active interest in the subject to be discussed.

Lieutenant General William B. Bunker is a graduate of the United States Military Academy, Class of '34. He attended the Massachusetts Institute of Technology receiving his degree of Master of Science in Engineering. During World War II, General Bunker served as Deputy in Charge of the Transportation Corps' Supply Program and, in 1945, as 7th Army Transportation Officer, during the occupation of Germany.

When the Berlin Airlift began in 1948 the General was put in charge of Terminal Operations governing gathering of shipments, loading in the United States zone, unloading and distributing cargo in Berlin. He organized a similar system between Korea and Japan when hostilities erupted in 1950.

In 1950 the Chief of Transportation named General Bunker to be Chief, Air Transport Division, investigating the application of the helicopter to Army transportation. The result of this investigation was an immediate large scale expansion of this activity. General Bunker was appointed Commandant of the U. S. Army Transportation School in 1954 and the following year was assigned as Commander, U. S. Army Transportation Materiel Command, responsible for logistic support of Army aviation. He was promoted to Major General 1 June 1961.

In February 1962 he became a member of the planning group which developed the organization for the Army Materiel Command and in June was assigned as its Comptroller and Director of Programs. On 1 April 1962 he became Deputy Commanding General, U. S. Army Materiel Command and was thereupon promoted to Lieutenant General on 9 May 1966.

General Bunker has been the recipient of many decorations for his outstanding work in a long and honorable career, not only from his own grateful country but also from the United Kingdom and Nicaragua.

He is a member of Professional Societies and published various articles in technical journals, and has developed a keen interest in the use of statistics in Army Research, Development, Testing and Evaluation.

It is with great pleasure that I introduce our keynote speaker, Lieutenant General William B. Bunker.

The title of his address is: "Broadening the Horizons of Experimental Design."

.....Thank you, General Bunker for your very interesting and informative address.

One of the most important objectives of these conferences has been to afford the conferees an opportunity to explore with authorities in the field those aspects of the subject matter which had most recently received major attention and development. When such areas have been determined, it has become the practice to invite experts in these various areas to speak on the topics selected.

Our next speaker is Professor Rolf Erwin Bargmann of the University of Georgia and the Thomas J. Watson Research Center of IBM. He has had a varied career, having been a Rockefeller Foundation Fellow prior to taking his Doctorate in Mathematical Statistics at the University of North Carolina. He was associated with our State Department in Germany and served as an interpreter during the Nuremberg Trials. He was Assistant Professor of Statistics and Head of the Department at Frankfurt, later Associate Professor of Statistics at Virginia Polytechnic Institute. He achieved full professorship in 1959. He was a consultant to White Sands Proving Ground in the summers of 1957 and 1959. He is a Fellow of the American Association for the Advancement of Science and a member of several statistical societies.

It gives me great pleasure to present Professor Bargmann, who will speak on, "The Structure and Classification of Patterns."

TABLE OF CONTENTS

*These papers were presented at the Conference, but do not appear in the Proceedings.

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

23 - 25 October 1968

Wednesday, 23 October

0830-0930 REGISTRATION - Building 4470

0930-0940 OPENING OF THE CONFERENCE - Post Theatre

Joseph Mandelson and LT Philip Watters
Chairmen on Local Arrangements

WELCOME

Colonel Paul R. Cerar, Commanding Officer,
Edgewood Arsenal

0940-1215 GENERAL SESSION I - Post Theatre

Chairman: Colonel Paul R. Cerar

BROADENING THE HORIZONS OF EXPERIMENTAL DESIGNS
Lieutenant General William B. Bunker, Deputy Commanding General, U. S. Army Materiel Command, Washington, D. C.

STRUCTURE AND CLASSIFICATION OF PATTERNS
Professor Rolf E. Bargmann, Department of Statistics, University of Georgia, Athens, Georgia

1230 LUNCH - Edgewood Arsenal Officers' Open Mess

1400-1520 TECHNICAL SESSION 1

Chairman: Ira A. DeArmon, Jr., Operations Research Group, U. S. Army Chemical Corps, Edgewood Arsenal, Maryland

STATISTICAL ANALYSIS OF CUTTING FLUID PERFORMANCE DATA
Lanny D. Wells, U. S. Army Weapons Command, Research and Engineering Directorate, Rock Island Arsenal, Rock Island, Illinois

MEASUREMENT OF ONE ASPECT OF VEHICULAR MOBILITY
Carol D. Rose and Raymond Owens, U. S. Army Tank-Automotive Command, Warren, Michigan

Wednesday Afternoon (Continued)

1400-1520 TECHNICAL SESSION 2

Chairman: Gideon A. Culpepper, Quality Control Division, White Sands Missile Range, New Mexico

THE DETERMINATION OF CONFIDENCE LIMITS FOR THE LOCATION OF THE MAXIMA, MINIMA, OR SADDLE POINT OF TWO VARIABLE, SECOND ORDER, MULTIPLE REGRESSION EQUATIONS
Lt. L. L. Lehn, U. S. Army Weapons Command, Research and Engineering Directorate, Rock Island Arsenal, Rock Island, Illinois,

BIONOMIAL WEIGHTED FOURIER TRANSFORM OF DESCRETE DATA
Hal Germond, Joint Task Force Two, Sandia Base, New Mexico

1400-1520 TECHNICAL SESSION 3

Chairman: Morton Shavitt, Quality Assurance Directorate, U. S. Army Edgewood Arsenal, Maryland

PROBABILITY OF A NON-REPEATABLE OBSERVATION - AN EXAMINATION OF THE UTILITY CONCEPT AND THE NATURE OF QUEUEING SEQUENCES
Mikiso Mizuki, Systems Performance Analysis and Evaluations, Federal Electric Corporation/ITT, Vandenberg Air Force Base, California

AN APPLICATION OF SIGNAL FLOW GRAPH THEORY TO A STOCHASTIC PROCESS
R. G. Stimson, Strategic Air Defense Systems Team, Weapons Systems Analysis Division, Office, Chief of Staff, Department of the Army, Washington, D. C.

1520-1550 BREAK

1550-1730 TECHNICAL SESSION 4

Chairman: Morris Rhian, Quality Assurance Directorate, U. S. Army Edgewood Arsenal, Maryland

DERIVATION OF CHARACTERISTIC CURVE FORMULAE FOR CSP-R
Harold W. Kelley and Fred L. Abraham, U. S. Army Ammunition Procurement and Supply Agency, Joliet, Illinois

APPROXIMATIONS TO CONTROL CHART CONSTANTS FOR ELECTRONIC COMPUTERS
David Hogben and John Mandel, National Bureau of Standards, Washington, D. C.

1550-1730 TECHNICAL SESSION 5

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

THE ROLE OF THE ANALOG COMPUTER IN THE RDT&E EFFORT AT EDGEWOOD ARSENAL
William Shulman, Weapons Development and Engineering Laboratories, U. S. Army Edgewood Arsenal, Edgewood Arsenal, Maryland

AN EVALUATION OF LINEAR LEAST SQUARES COMPUTER CODES
Roy H. Wampler, National Bureau of Standards, Washington, D. C.

1550-1730 TECHNICAL SESSION 6

Chairman: Robert M. Eissner, Surveillance and Reliability Laboratory, Aberdeen Research and Development Center, Aberdeen Proving Ground, Maryland

A NOTE ON THE ERROR ANALYSIS OF A FIRE CONTROL SYSTEM
T. H. Slook, U. S. Army Frankford Arsenal, Philadelphia, Pennsylvania

ANALYSIS OF MULTI-DIMENSIONAL CONTINGENCY TABLES
H. H. Ku, National Bureau of Standards, Washington, D. C.

1830- SOCIAL HOUR AND BANQUET

Announcement of the recipient of the Fourth Samuel S. Wilks Memorial Award by Dr. Frank E. Grubbs

Thursday, 24 October

9099-1040 CLINICAL SESSION A

Chairman: Bernard Harris, Mathematics Research Center, U. S. Army, University of Wisconsin, Madison, Wisconsin

Panelists:

Joseph M. Cameron, Statistical Engineering Laboratory, National Bureau of Standards, Gaithersburg, Maryland

Walter D. Foster, Biomathematics Division, U. S. Army Biological Laboratories, Fort Detrick, Frederick, Maryland

H. L. Lucas, Jr., Institute of Statistics, North Carolina State University, Raleigh, North Carolina

PROBLEMS IN EVALUATING TREATMENT RESPONSE OVER UNEQUALLY SPACED TIME INTERVALS

Gerhard J. Isaac, Statistics Branch, U. S. Army Medical Research & Nutrition Laboratory, Fitzsimons General Hospital, Denver, Colorado

ANALYSIS OF DATA FROM THE WOUND DATA AND MUNITIONS EFFECTIVENESS TEAM IN VIETNAM

W. Bruchey, L. Sturdivan, and R. Whitmire, Research Laboratories, Edgewood Arsenal, Maryland, and Aberdeen Research and Development Center, Aberdeen Proving Ground, Maryland

0900-1040 CLINICAL SESSION B

Chairman: A. C. Cohen, Jr., Institute of Statistics, University of Georgia, Athens, Georgia

Panelists:

O. P. Bruno, Surveillance and Reliability Laboratory, Aberdeen Research and Development Center, Aberdeen Proving Ground, Maryland

Frank E. Grubbs, Operations Research Analyst, Aberdeen Research and Development Center, Aberdeen Proving Ground, Maryland

W. G. Kruskal, Department of Statistics, University of Chicago, Chicago, Illinois

Joan R. Rosenblatt, Statistical Engineering Laboratory, National Bureau of Standards, Gaithersburg, Maryland

DECENTRALIZED GRADIENTS IN EXPERIMENTAL OPTIMIZATION
David R. Howes, U. S. Army Strategy & Tactics Analysis Group, Bethesda, Maryland

EFFECT OF NUMBER OF OBSERVING STATIONS ON FLIGHT MEASUREMENT PRECISION
Fred S. Hanson, White Sands Missile Range, New Mexico

0900-1040 TECHNICAL SESSION 7

Chairman: Henry Ellner, Quality Assurance Directorate, U. S. Army Materiel Command, Washington, D. C.

EVALUATION OF NICKEL-IRON AND NICKEL-ZINC BATTERIES
Martin J. Sulkes, U. S. Army Electronics Command, Fort Monmouth, New Jersey

STATISTICALLY DESIGNED EXPERIMENTS IN THE STUDY OF FACTORS AFFECTING BREAKDOWN OF HIGH VOLTAGE IN A VACUUM
M. M. Chrepta, G. W. Taylor, and M. H. Zinn, Electronic Tubes Division, Electronic Components Laboratory, U. S. Army Electronics Command, Fort Monmouth, New Jersey

DESIGN OF EXPERIMENT AND A STATISTICAL PERFORMANCE MODEL FOR A RADAR ALTIMETER
Erwin Biser, Operations Research Analyst, Avionics Laboratory, U. S. Army Electronics Command, Fort Monmouth, New Jersey

1040-1120 BREAK

1120-1215 CLINICAL SESSION C

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

Panelists:

Robert Bechhofer, Operations Research, Cornell University, Ithaca, New York

Frank E. Grubbs, Operations Research Analyst, Aberdeen Research and Development Center, Aberdeen Proving Ground, Maryland

Boyd Harshbarger, Department of Statistics, Virginia Polytechnic Institute, Blacksburg, Virginia

Herbert Solomon, Department of Statistics, Stanford University, Stanford, California

AN EXPERIMENT USING NUMERICAL ANALYSIS TO MODEL A FUNCTIONAL RELATIONSHIP BETWEEN ABM SYSTEM SENSOR RESPONSES AND REENTRY VEHICLE CHARACTERISTICS

Andrew H. Jenkins, U. S. Army Missile Command, Redstone Arsenal, Alabama

1120-1215 TECHNICAL SESSION 8

Chairman: James R. Gebert, C-E-I-R, Biomathematics Division, U. S. Army Biological Laboratories, Fort Detrick, Frederick, Maryland

SOME EXPERIENCES IN LABORATORY CONTROL INVESTIGATION

S. P. Zobel, Cornell Aeronautical Laboratory, Buffalo, New York. Representing U. S. Army Edgewood Arsenal, Maryland.

1120-1215 TECHNICAL SESSION 9

Chairman: Hal Germond, Analyses and Reports Directorate, The Joint Chiefs of Staff Joint Task Force Two, Sandia Base, New Mexico

STATISTICAL ANALYSIS MADE EASY: A SUCCESSFUL MARRIAGE OF STATISTICAL THEORY AND COMPUTERS

Melvin O. Braaten and C. Frank Starmer, North Carolina Operations Analysis Standby Unit, HQ Tactical Air Command (OA) and Duke University, Durham, North Carolina

1215-1315 LUNCH - Edgewood Arsenal Officers' Open Mess

1315-1400 OPEN MEETING OF THE AMSC SUBCOMMITTEE ON PROBABILITY AND STATISTICS - [To be held at the Officers' Open Mess Building]

Presided over by: Dr. Walter D. Foster, Biomathematics Division, U. S. Army Biological Laboratories, Fort Detrick, Frederick, Maryland

Thursday afternoon (Continued)

1400-1500 GENERAL SESSION II - Post Theatre

Chairman: Fred Frishman, Mathematics Branch, Office of the Chief of Research and Development, Department of the Army, Washington, D. C.

BULK SAMPLING
Professor Acheson J. Duncan, Department of Operations Research and Industrial Engineering, The Johns Hopkins University, Baltimore, Maryland

1500-1530 BREAK

1530-1715 PANEL DISCUSSION ON BULK SAMPLING

Chairman: Dr. Walter D. Foster, U. S. Army Biological Laboratories

Discussant: Professor Acheson J. Duncan, The Johns Hopkins University

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

Henry Ellner, U. S. Army Materiel Command, Washington, D. C.

Gene Ray Lowrimore, Hercules, Inc., Radford Army Ammunition Plant, Radford, Virginia

Joseph Mandelson, U. S. Army Edgewood Arsenal, Maryland

Vernon H. Rechmeyer, Thiokol Chemical Corporation, Huntsville Division, Redstone Arsenal, Alabama

****

Friday, 25 October

0900-1040 CLINICAL SESSION D

Chairman: Boyd Harshbarger, Department of Statistics, Virginia Polytechnic Institute, Blacksburg, Virginia

Panelists:
Robert Bechhofer, Operations Research, Cornell University, Ithaca, New York

A. C. Cohen, Jr., Institute of Statistics, University of Georgia, Athens, Georgia

W. G. Kruskal, Department of Statistics, University of Chicago, Chicago, Illinois

H. L. Lucas, Jr., Institute of Statistics, North Carolina State University, Raleigh, North Carolina

Herbert Solomon, Department of Statistics, Stanford University, Stanford, California

MEASUREMENT ON-THE-MOVE OF TANK WEAPON POINTING ACCURACY
Stanley M. Birley, U. S. Army Weapons Command, Rock Island Arsenal, Rock Island, Illinois

A COMPARISON OF THE SAMPLING ERRORS OF A SYSTEMATIC SAMPLING METHOD WITH COMPLETE RANDOM SAMPLING IN THE ESTIMATION OF TOTAL NUMBER OF NERVE FIBERS
Douglas B. Tang, Department of Epidemiology and Biostatistics, Walter Reed Army Institute of Research, Washington, D. C.

0900-1040 TECHNICAL SESSION 10

Chairman: H. Gill Hilton, Biomathematics Division, U. S. Army Biological Laboratories, Fort Detrick, Frederick, Maryland

SOME STATISTICAL ASPECTS OF ASSURANCE OF STERILIZATION
Francis M. Wadley, Consultant, Fort Detrick, Frederick, Maryland

RESEARCH AND DEVELOPMENT MATHEMATICAL EQUATIONS AS RELATES TO AN ARMY AIRCRAFT SYSTEM
Tony N. O'Truk, Directorate of Systems and Cost Analysis, U. S. Army Aviation Materiel Command, St. Louis, Missouri

HYPOTHESIS TESTING AND CONFIDENCE INTERVALS FOR PRODUCTS AND QUOTIENTS OF POISSON PARAMETERS WITH APPLICATIONS TO RELIABILITY
Bernard Harris, Mathematics Research Center, U. S. Army, University of Wisconsin, Madison, Wisconsin

0900-1040 TECHNICAL SESSION 11

Chairman: Edwin Fiske, Quality Assurance Directorate, U. S. Army Edgewood Arsenal, Maryland

Friday Morning (Continued)

METHODOLOGY OF ASSESSMENT OF BIOCELLULAR PERFORMANCE EFFICIENCY
George I. Lavin, Terminal Ballistic Laboratory, Aberdeen Research and Development Center, Aberdeen Proving Ground, Maryland

MONTE CARLO INVESTIGATION OF THE ROBUSTNESS OF DIXON'S CRITERIA FOR TESTING OUTLYING OBSERVATIONS
Jerry Thomas, Surveillance and Reliability Laboratory, Aberdeen Research and Development Center, Aberdeen Proving Ground, Maryland

A METHOD FOR APPROXIMATING PROBABILITY FUNCTIONS DEFINED ON FINITE DOMAINS
Joseph S. Tyler, Jr., Systems Analysis Office, U. S. Army Edgewood Arsenal, Maryland

1040-1120 BREAK

1120-1220 GENERAL SESSION III

Chairman: Colonel D. F. Burton, Commanding Officer, U. S. Army Research Office-Durham, Durham, North Carolina

TIME SERIES
Professor Emanuel Parzen, Department of Statistics, Stanford University, Stanford, California

CLOSING OF THE CONFERENCE
Dr. Frank E. Grubbs, Conference Chairman, Aberdeen Research and Development Center, Aberdeen Proving Ground, Maryland

1230-1330 LUNCH

TOUR of the U. S. Army Edgewood Arsenal

PROGRAM COMMITTEE

Joseph Cameron	H. L. Lucas, Jr.
Francis Dressel (Secretary)	Clifford Maloney
Walter D. Foster	Joseph Mandelson
Fred Frishman	Henry Mann
Boyd Harshbarger	Raymond B. Schnell
William Kruskal	Herbert Solomon

Frank E. Grubbs (Chairman)

BROADENING THE HORIZONS OF EXPERIMENTAL DESIGN

LT General William B. Bunker, Deceased
U.S. Army Materiel Command
Washington, D.C.

From its early beginnings, statistics has been an important vehicle with which reasonable men have attempted to seek an understanding of the problems which confront them. Some of the earliest developments and applications of statistical concepts occurred in response to problems at the gaming tables. In fact, I have been told that more than one early statistician earned his keep by calculating odds for a wealthy gambler. The basic orientation of statistics toward the solution of practical problems can be found as the motivation for many major developments in statistics. For example, Thomas Bayes in his often quoted and controversial essay stressed his desire to provide a more efficient procedure for the estimation of probabilities. More recently, the contributions of Professor R.A. Fisher in the area of small sample statistics were motivated by a desire to improve the analytic tools available in biomedical research.

The essential point is that many of the important developments in statistics were motivated by a desire to solve real world problems. I am concerned that in some quarters this orientation to problem-solving has been replaced with a tendency toward self contemplation and a primary interest in statistical purity. There is a need to re-examine the direction of current efforts and to confront our major problems head-on. Only through broadening the horizons of experimental design can we hope to deal effectively with our most pressing problems.

Today, as a first step toward broadening the horizon, I would like to spend the remainder of my time discussing several areas that are amenable to the application of the concepts of experimental statistics.

SYSTEM TESTING AND DEVELOPMENT. One important area in which much work is needed involves the statistical issues in equipment testing. At the offset, I want to stress that our test programs are not and in fact cannot be scientific experiments. One reason for this is that the traditional requirements for the design of experiments are infeasible within the context of a test and development program. For example, a basic principle of design of experiments involves the control or minimization of the variation in the experimental situation. This is an almost impossible requirement to satisfy for two reasons. First, due to modification in the system during development, the basic heterogeneity of experimental units is high. This inherent variability represents a violation of a basic statistical assumption. Second, the dimensions of the problem frequently preclude control or even measurement of extraneous sources of variation. The problem was illustrated in the test program for our new AAFSS.

The status of a scientific experiment also is denied to our development and test programs because of the fact that we just can't afford the large

number of data points that are required in a classical experimental design. In practice, testing is done on a small number of prototype systems. If an attempt was made to gather the number of observations required to achieve the desired level of statistical significances, no development would ever take place.

The statistical aspects of testing programs are further compounded by our difficulties in specification of the model. In many of our test programs it is difficult to begin to select the relevant variables and logically impossible to identify the important interactions and nonlinearities.

Our recent experience with the development of 152 ammunition for the Sheridan provides a case in point. The variable of interest in this case is binominal, either the round fires or it does not fire. We know that reliability of this ammunition is a function of a number of variables including quality control, the efficiency scavenger system, the ammunition case, and the storage environment, but we also realize that there are n other important dimensions of the problem which remain to be identified. For example, through observation we have established an interaction between the degree of moisture in the powder and the quantity of residue. Experience has demonstrated that higher moisture content resulted in more residue. In response to this finding we have lowered the moisture content, but this change raises a question concerning other yet unknown interactions that are at work in determining the reliability of the ammunition.

Changing the moisture content also illustrates another problem that pervades the testing programs. When the nature of an item is altered as a matter of course in testing and development, how does one aggregate the test data that were generated prior to the change with that data which have been gathered after the change? In a strict sense, the modification has changed the basic structure of the situation that is being modeled, and has made the two sets of data incommensurable. In reality, we are measuring a series of separate probability curves and are reporting the envelope of these curves. This is analogous to developing a baseball batting average by combining performance in the preliminary grapefruit league with that in standard league play. In both cases, the cumulative measure of performance combines early and tentative results with those that have been obtained after the system has been brought up to working order. The net effect of this procedure is to substantially understate the reliability of the system.

Given this situation, how can we give our customer a valid statement of quality assurance? Upon examining the results of the testing program, the statistician would say that we have a ratio of approximately 1 to 52,000, but what we really need to satisfy the customer is a ratio of 1 - 1,000,000. At this point I can say, qualitatively, that the real reliability of the system is understated; however, it is impossible to specify the absolute magnitude of the error. Naturally, the customer is not satisfied with the statement about reliability of the ammunition, and something must be done to improve the situation. The statisticians' answer to this dilemma is more testing to develop the required observations. This is an extremely costly procedure and it would have been better to have done more work on estimating the initial function. Ad hoc testing at this juncture is not a feasible solution to the problem.

An alternative approach can be found in the area of statistical decision theory. Resolution of this dilemma may be achieved through the combination of the subjective judgment of the experts and objective experimental results.

A second area in our testing program that requires attention involves the development of large, expensive systems. The Main Battle Tank provides a good illustration of the problem. We really have only a vision of the MBT. In this situation, the problem is that there is no real testing of the whole system. Instead, tests are conducted on different vehicles with various configurations. This means that most of the parameters of interest vary from test to test and that very little remains constant among the tests. What we are attempting to model then is really a function of functions. Casual factors can no longer be expressed simply as numeric values but themselves must be represented as functions, the values of which are in turn dependent upon the value of the total function.

One analytic technique that has been utilized to attempt to model a function of functions is dynamic programming. In the development of the basic algorithm Bellman used a recursive scheme to reflect the method of sequential calculation that is the essence of the approach. For example, consider an aerial weapon system consisting of a navigational subsystem, a target acquisition subsystem, and a weapon subsystem. It is desired to determine the optimal characteristics of all three subsystems, but all these decisions are interdependent. The thing we do know is that whatever navigational and target acquisition subsystems are chosen, the characteristics of the weapon system, e.g., the rate of fire must be optimal with respect to the effectiveness of the whole system. Using the principle of optimality proposed by Bellman, we can say that the optimum rate of fire is a function of the effectiveness of the aerial weapons system. Since we do not know the optimal characteristics for the other two subsystems, the optimal rate of fire and total system effectiveness must be found for all feasible outputs of the subsystem. This technique may provide a clue regarding the way to handle complex equations without knowing their specific form.

The essential point is that we must move away from concepts that require the testing of a static system. Pressures imposed by necessary modifications of systems in the development process do not allow all other things to remain equal and this dynamic aspect of the environment cannot be ignored.

On balance, it appears that increased emphasis on rigor in the design of experiments has diverted our attention from the ultimate objectives. Efforts must be undertaken to develop techniques which provide feasible solutions to problems of quality assurance and the manipulation of more complex dynamic models. We need to soften the science of experimental design to make it a more useful tool in test and development programs. The alternative to this change is to continue to strive for more technically precise answers which are even less meaningful in the decision making process. Unless a conscious effort is made to avoid this plight, experimental statistics may create a paradox similar to that caused by

managerial accounting. As a tool of management, the discipline of accounting has experienced an increase in the precision with which financial information is analyzed and reported, but it still does not provide much assistance in the decision making process. Decision makers can safely rely on accounting to identify the loss after the investment has failed, but it is of no help in forecasting the likelihood of this occurrence. It is an after the fact discipline, and our requirements are for knowledge before the fact.

While reflecting on these challenges that lie ahead, it may be useful to reconsider the role of statistical analysis in the decision making process. The decision maker is concerned with choosing between two or more alternatives; the value of which remains to be established by events in the future. Statistical analysis is valuable only to the extent to which it raises the level of understanding of the problem and in so doing provides an improved basis for fixing beliefs about the future. In contrast, analyses that provide interesting expositions, but no additional understanding, are of little value. It, therefore, is essential for the analyst to be attuned to informational requirements of the decision maker if real progress is to be made.

MANAGEMENT INFORMATION SYSTEMS. A second area which could benefit from the attention of statisticians is the design of management information systems. Even a cursory examination of the recent attempts to design and implement management information systems reveals the opportunity for substantial improvement through the infusion of the concepts of experimental statistics. Many of these efforts reflect a lack of understanding of the available techniques for summarizing and annalyzing data. The result of this naivete has been inefficiency in system design and confusion regarding the purpose and value of the output of the system. For example, the operation readiness of our hawk units throughout the world must be monitored daily by phone. Since this information is vital to decision makers at the highest levels, one would have hoped that a less cumbersome communication system could have been planned.

To provide you with more background on the problem area, it may be useful to examine briefly the origin of our current dilemma. The root of the problem can be found in our recently acquired capacity to process and transmit rapidly information. In the last thirty years technological progress has resulted in the development of three generations of computers; each of which represented a dramatic improvement over the current state-of-the-art. Equipped with the exciting abilities to process in a real time mode and to directly access data banks, the designers of these systems have moved in the direction of including everything about everything in the system.

One example of the problem is provided by the periodic Army readiness report that is prepared for the Chief-of-Staff. Included in this report, in great detail, is information on not only major items such as tanks and jeeps but also on many minor items as well. Once attention was drawn to equipment readiness at this level of specificity it became apparent that the number and status of most of the items were subject to continual change.

This meant that the job of preparing a large scale report was further compounded by the fact that the information had to be updated and published frequently, if it was to be of value in its current form. A question can be asked as whether or not this is a worthwhile or even feasible effort. This same point should be raised in every management information system.

In nearly all phases of our business today one can observe information being translated into electronic impulses for transmission up to higher levels of authority. It is important to note that once data is separated from its traditional hard copy vehicle, e.g., the DA Form; it can be sorted, summarized, or transmitted at almost unbelievable speeds. It is this speed and the low per unit cost of processing information which have caused many of the current problems with management information systems.

These rapid changes in communications technology have caused some rather traumatic experiences in most large organizations. To begin with, many management theorists and most managers of today are still thinking in terms of the traditional forms of organization structure. These concepts generally involve pyramidal configurations of the different layers of authority. The problem is that these organizations reflect a certain state of information processing technology and this level of technology is rapidly becoming obsolete. There is no doubt that a certain disparity has always existed between the institutionalized organization structure and information technology; however, recent innovations have aggravated and accentuated the problem. It is useful to examine the factors that are important to this problem in order to better evaluate alternative solutions.

One important factor is the heterogeneity in the speed with which different types of information are processed through the organization. While it is not possible to rapidly analyse and summarize information on personnel strength through the organization, it is still necessary to individually monitor the progress of many R&D programs. So within the same large organization, new information processing techniques have dramatically affected the form and function of some activities while others remain essentially unchanged. This phenomenon has made the traditional concepts of a centralized and decentralized organization obsolete in that both tendencies are apparent within many phases of our business.

The increasing magnitude of the upward flow of information also serves to exacerbate the disparity between information processing technology and organizational structure. Too frequently, our concept of the informational requirements that must be transmitted up to top management reflects a lack of appreciation for the objectives of the system. Most communication that an individual has with the higher levels of the organization is through his immediate superior. Communication at this level is intimate and detailed and this is as it should be between superior and subordinate. This is not, however, the appropriate level of communication between a first line supervisor and top management. The top level manager has neither the need to know nor the capability to assimilate the large volumes of specific information; and, therefore, it makes little sense to send information at this level of detail up through

the information system.

In addition to being illogical, this tendency has serious implications for the organization and the decision maker. If the trend continues, middle management will of necessity be relegated, in large measure, to the job of expediting the flow of information up the line of authority. More important, however, is the effect of this tendency on the performance of the decision maker. From his point of view, this tremendous flow of information provides an all encompassing yet fragmentary view of reality. While the decision maker has easy access to information regarding every significant dimension of the problem and some trivial ones as well; he may still find himself in a quandary over the nature of the situation. The reality of any situation is extremely complex when viewed in its entirety. Most of us have learned through experience in situations to suppress those aspects of reality which are superfluous to the problem at hand; however, the ability to do this effectively depends on an intimate understanding of the particular problem and environment. This point illustrates a major impetus for specialization of interest and talent but raises a serious question concerning the relationship between the top level decision maker and the information system. It is obvious that no top manager, regardless of his ability, can begin to accumulate experience comparable to the new sum of that possessed by the specialists in his organization. It should be equally obvious that the detail and format of information required by the manager is markedly different from that which is required in the lower echelons. This is, however, only half the problem.

The sorting and evaluating of information by the decision maker is further complicated by the fact that the information has been abstracted from the environment to which it is indigenous. No longer is it possible to view the situation in its totality or to make inferences from the juxtaposition of the various elements. The information is now presented in a homogenous package and there is little effort made to illustrate the relative importance of the various bits of information. This format encourages the tendency to limit the analysis to what are apparently obvious relationships in the data, and all too often, these obvious relationships depict only a superficial view of the problem. When confronted with such a situation, the decision maker is tempted to feel that his evaluation is profound when it in fact may be obvious and trivial or even worse incorrect.

The question then arises as to what alternatives are available to aid us in resolving this dilemma. One answer to the problem may be found in the imaginative and effective application of the techniques of statistical analysis. Concepts and procedures that have been used successfully for years by statisticians offer the means by which meaningful order can be restored in our information systems.

Returning to the example of the Army readiness reports, in this information system the emphasis has been placed on reporting the status of practically every item in the inventory. A moment's reflection reveals that this approach is a violation of the principle of parsimony. Why is it necessary to report data on the status of every item, when we are really

only interested in those items in a particular status? It is encouraging to note that all information systems have not proceeded down the same path. The New York City Department of Public Health, for example, does not attempt to measure the health status of the city by directly estimating the proportion of the total population who are well. Instead, their attention is focused only on those who are sick. Their approach is to monitor the population of the hospitals throughout the city. Through observation of this one accessible indicator, they are able to maintain an adequate estimate of the general level of health of the community.

The principle is to replace the real variable of interest with surrogate which is more easily measured and analyzed. This has been a relatively common practice among statisticians and it should have application in the design of our information systems. In the case of the readiness report, a substantial increase in the value of the effort would be realized by reporting exceptions rather than the status of the whole system. This scheme would substantially reduce the upward flow of information and focus attention on the real variable of interest. In another phase of the operation, perhaps the status of a particular maintenance operation could be gauged more efficiently and accurately through the examination of the re-enlistment rates rather than the number of items serviced per month. The kind of changes suggested would not only reduce the upward flow of information but also place the information in a form and format that is more useful in the decision making process.

The concepts of sampling offer yet another statistical tool that appears to have application in the design of information systems. Even if modern technology can provide us with the machine capability to process information at very high speeds, this capability has a significant, positive cost. It is therefore necessary to examine alternative ways to economize in the operation. Sampling theory provides the basic notions for efficiently and economically gathering data about a particular population of interest. For example, the mean cost of procuring an item could be estimated accurately and at a mere fraction of the cost of total enumeration through the use of a self-weighting, stratified sample. It should also be remembered that in many cases, sample estimates might be even better than would usually be expected because our concern is primarily with finite populations.

A more general perspective for design of an information system may be gained from the philosophy of analysis that pervades among statisticians. While many of the designers of information systems have been content to concentrate on the preparation and reporting of data, the interest of most statisticians continues through analysis and interpretation. Efforts must be made to bring the analysis phase into the design of a system. Up to this point system designers have emphasized performance measures such as speed or cost per calculation as measures of effectiveness, but we have seen that this approach ignores the important question about system effectiveness, i.e., what is the value of information? Timeliness of information is important; however, in our effort to obtain more current data we have ignored certain other important aspects of the problem. Is it really worth anything to the organization to spend additional money to send information more quickly if much of the information in the system is

already redundant or nonuseable? Does it make sense to publish figures in a daily report if it will require several weeks worth of observation to verify whether a change in the data is real or simply an aberration? The answer to both questions is obviously no! Both queries suggest that, in the future, major payoffs will accrue to advances in the analysis of data that can be incorporated within the system. Further analysis will take additional time; however, it should also substantially increase the informational value of reports. When examining this tradeoff it is essential to remember that most changes that take place within a large organization are gradual and occasionally painfully slow. Given this situation, it is reasonable to expect that the opportunity cost of the time lost during further analysis may be substantially less valuable than the increased understanding which would be generated.

In summary, there is a genuine need to apply the philosophy of experimental design to the design of management information and control systems. Statistical techniques can help to determine which variables should be measured and which should be ignored, as well as facilitating the analysis and forecasting of trends. Up to now, there has been little feedback between those interested in experimental design and those involved in information system design. Much of what we know in the latter area has been the result of a trial and error process, and as I am sure you are well aware, this can be a very expensive way to learn. If some of the statistical notions of sampling and analysis can be communicated to system designers, then substantial payoffs will be realized. A response in this direction now will encourage efficiency and progress. If no response is forthcoming, however, and decision making continues to escalate, a requirement for total information reporting will demand a huge organization just for purposes of processing. In many ways, the dilemma of the decision maker is analogous to that of an individual who attempts to examine the behavior of a particle suspended in liquid. The more the individual studies the particle the more confused he becomes of the random effect of brownian motion. The perception of both the hypothetical individual and the decision maker could be improved through the use of certain basic statistical notions.

CONCLUSION. As we have seen there are a number of opportunities to broaden the horizons of experimental design through reduced emphasis on rigor and increased attention to current problems; be they in testing or systems design. The next move is up to you.

THE STRUCTURE AND CLASSIFICATION OF PATTERNS*

Rolf E. Bargmann
University of Georgia
Athens, Georgia

TERMINOLOGY.**

Logical Pattern: A set of p diagnostic events is observed. Occurrence is marked by 1, non-occurrence by 0. Such single observation results in a row of 0's and 1's. Observations are repeated, and several such rows constitute a pattern. If rows are dependent (e.g., observation at consecutive times), a cyclical autocorrelation dependence is assumed.

Major Event: One or very few underlying artificial events, each of which may assume two or more states, which influence the probability of occurrence of each diagnostic event.

Calibration Pattern: A logical pattern, consisting of several rows, containing observation of occurrences and non-occurrences of all diagnostic events if the major event (or, rather, some physical event closely related to the artifical major event) is in a known state (e.g., repeated observation of symptoms of a patient who suffers from a known disease).

Model Assumption (leading to a variant of the Latent Class Model): The state of the major event determines the probability of occurrence or non-occurrence of each diagnostic event. Except for this influence, the diagnostic events are assumed to be independent (principle of conditional independence).

Sample Pattern: A logical pattern consisting of one or more rows, describing a situation where the state of the major event is unknown. Its distance (Euclidean distance or, better, -2 log likelihood) from each of the calibration patterns determines the proximity of the current state of the major event to each of the known states represented by the calibration patterns.

Note that extensive calculations are required on calibration patterns only. Determination of the distances of a sample pattern from each calibration pattern is a very simple matter, and can even be done by hand calculation.

---

*A handout at the conference served as a basis for this paper.

**Reference, R. E. Bargmann, "A Method of Classification Based upon Dependent 0-1 Patterns," IBM Research Report No. RC-677, April, 1962).

SOME OF THE SIMPLER FORMULAS. The following quantities must be obtained from each calibration pattern:

x_{ti} = 0 or 1 -- the entry in row t and column i

of the calibration pattern. N = number of rows, p = number of columns

$$S_i = \sum_{t=1}^{N} x_{ti} \qquad \hat{p}_i = S_i/N \text{ (column averages)}$$

$$S_{ij} = \sum_{t=1}^{N} x_{ti}x_{tj} \qquad \hat{p}_{ij} = S_{ij}/N \text{ (average number of 1-matches in columns i,j)}$$

If rows are assumed to be independent, then

$$\hat{\sigma}_{ii} = [\hat{p}_i(1-\hat{p}_i)]/N\text{-}1$$

$$\hat{\sigma}_{ij} = [\hat{p}_{ij} - \hat{p}_i\hat{p}_j]/N\text{-}1$$

If rows are assumed to be time-dependent (cyclical, autocorrelation of lag 1) the following additional quantities are needed

$$C_i = \sum_{t=1}^{N} x_{ti}x_{t+1,i} \qquad (x_{1i} = x_{N+1,i})$$

$$D_{ij} = \sum_{t=1}^{N} x_{ti}x_{t+1,j} \qquad \text{(1-matches, down)}$$

$$U_{ij} = \sum_{t=1}^{N} x_{t+1,i}x_{tj} \qquad \text{(1-matches, up)}$$

$$\hat{r}_i = (C_i - S_i^2/N)/(S_i - S_i^2/N) \quad \text{(autocorrelation)}$$

(If N is even, and if a perfectly alternating sequence occurs in a column--i.e., 010101... or 101010..., $\hat{r}_i$ should be replaced by $2/(N+1)-1$).

Then

$$\hat{\sigma}_{ii} = \frac{\hat{p}_i(1-\hat{p}_i)}{N-1} \cdot \left(1 + \frac{2\hat{r}_i}{1-\hat{r}_i}\right)$$

$$(= (N-2)/N^3 \quad , \text{ if a column consists of perfect zeros or ones})$$

$$\hat{\sigma}_{ij} = \frac{\hat{p}_{ij}-\hat{p}_i\hat{p}_j}{N-1} + \frac{\hat{r}_i}{(1-\hat{r}_i)(1-\hat{r}_j)} \cdot \frac{S_{ij}-D_{ij}}{N(N-1)}$$

$$+ \frac{\hat{r}_j}{(1-\hat{r}_i)(1-\hat{r}_j)} \cdot \frac{S_{ij}-U_{ij}}{N(N-1)}$$

Subject the matrix $\hat{\sum}$ (or the corresponding correlation matrix, --most computer programs do the conversion automatically) to a Factor Analysis. If the major event is assumed to have 2 states, extract one factor, if k + 1 states, extract k factors. A crude technique (e.g., Centroid) or even cruder ones (e.g., principal components which, alas, some computer programs call "Factor Analysis") can be expected to yield satisfactory results. For the special case of two states of the major event, a single vector $\underline{f}$ (elements f_i) will be reported. From each calibration pattern, the weights $w_i = 1/(1 - f_i^2)$ should be calculated.

Now, to calculate the distance (or rather, the -2 log likelihood quantity) of a sample pattern from calibration pattern q obtain the average of each column in the sample pattern, call it a_i.

Then

$$d_q = \sum_{i=1}^{p} \log \hat{\sigma}_{iiq} + \log \left[1 + \sum_{j=1}^{p} (w_{jq}-1)\right] - \sum_{i=1}^{p} \log w_{iq}$$

$$+ k \sum_{i=1}^{p} \frac{(a_i-\hat{p}_{iq})^2 w_{iq}}{\hat{\sigma}_{iiq}} -$$

$$- k \left[\sum_{i=1}^{p} \frac{(a_i - \hat{p}_{iq}) f_{iq} w_{iq}}{\sqrt{\hat{\sigma}_{iiq}}} \right]^2 / \left[1 + \sum_{j=1}^{p} (w_{jq} - 1) \right]$$

where k = number of rows in the sample pattern, and all logs are to base e. The last subscript q indicates that the corresponding value is to be taken from the q'th calibration pattern.

IMPLICATIONS OF MODEL ASSUMPTIONS ON THE STRUCTURE OF THE COVARIANCE MATRIX: If the major event has only two states, and α is the probability that the major event is in state 1

$$N \cdot \sum = (\alpha - \alpha^2) \; \underline{p} \, \underline{p}' + \text{diagonal}$$

where the vector $\underline{p}$ has elements $(p_{i/1} - p_{i/0})$; i.e., the difference between the conditional probabilities of occurrence of diagnostic event i, given that the major event is in state 1 or 0.

If there are k + 1 states (or, with restrictions, several major events), the covariance matrix has the structure

$$N \cdot \sum = P \begin{bmatrix} \alpha_1(1-\alpha_1) & -\alpha_1\alpha_2 & \cdots & -\alpha_1\alpha_k \\ -\alpha_1\alpha_2 & \alpha_2(1-\alpha_2) & \cdots & -\alpha_2\alpha_k \\ \cdot & \cdot & \cdots & \\ -\alpha_1\alpha_k & -\alpha_2\alpha_k & \cdots & \alpha_k(1-\alpha_k) \end{bmatrix} P' + \text{diagonal}$$

where α_m denotes the probability that the major event is in state m, and the matrix P has k columns (number of states minus 1). The element in row i and column m is $(p_{i/m} - p_{i/0})$.

These are standard factor analysis models. The matrices are easily inverted, and the determinant is easily found -- thus, the calculation of distances from a sample pattern to each of the calibration patterns can be most easily effected.

A direct evaluation of the conditional probabilities can be made only if assumptions can be made relative to the probabilities that the artificial major event is in a given state. Such assumptions are somewhat tenuous, inasmuch as the physical major event is not identical (though hopefully highly correlated) with the artificial major event.

Example: Frequency of Repair records, 6 consecutive years, 5 characteristics 0 = below average or average, 1 = above average. Calibration patterns for Make A and Make B

Make A	Make B
00111	10111
10111	01111
01010	01111
11010	10101
01101	10101
11111	00101

Estimates of covariance matrices (assuming row dependence)

$$\frac{1}{1080}\begin{bmatrix} 9 & 0 & 0 & 0 & 0 \\ 0 & 80 & -40 & -12 & -40 \\ 0 & -40 & 80 & -22 & 200/3 \\ 0 & -12 & -22 & 20 & -22 \\ 0 & -40 & 200/3 & -22 & 80 \end{bmatrix} = \hat{\sum}_A$$

$$\frac{1}{1080}\begin{bmatrix} 27 & -33 & 27/2 & -18 & -27/2 \\ -33 & 80 & -16 & 54 & 12 \\ 27/2 & 54 & -9 & 108 & 18 \\ -27/2 & 12 & -6 & 18 & 20 \end{bmatrix} = \hat{\sum}_B$$

Factor Analysis results:

$$\underline{f}_A' = [0,\ -.500,\ .912,\ -.563,\ .924]$$

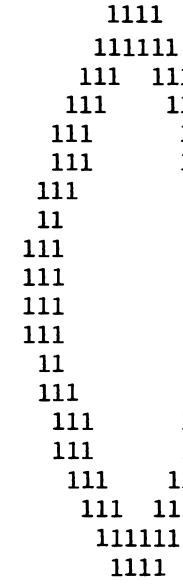
$$\underline{f}_B' = [-.972,\ .730,\ -.591,\ .377,\ .588]$$

Use each row as a "sample pattern"

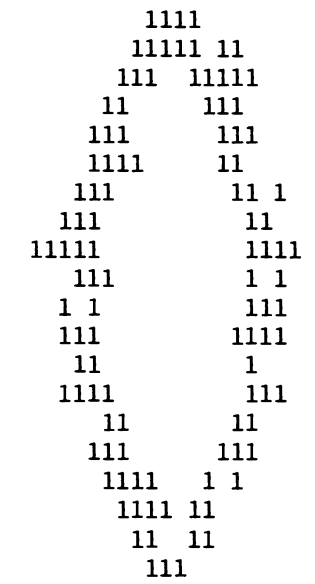
Row	d_A	d_B	Decision	Comments
00111	23.0534	24.2761	to A	
10111	23.0534	11.0242	to B	Occurs in A and B, assigned to B
01010	18.1141	166.8874	to A	
11010	18.1141	216.8186	to A	
01101	60.1230	9.8467	to B	Misclassification
11111	22.4190	50.9441	to A	
01111	22.4190	1.7716	to B	See comment below
10101	58.3983	3.2090	to B	
00101	58.3983	31.5718	to B	

The 01101 sequence shows the importance of a dependence or row assumption. Since, on the first diagnostic variables, the averages are equal for A and B, there would have been no difference in assignment between this and 11111, if independence had been assumed. The present classification is correct.

Calibration Pattern 0



Sample from 0 with 20% error

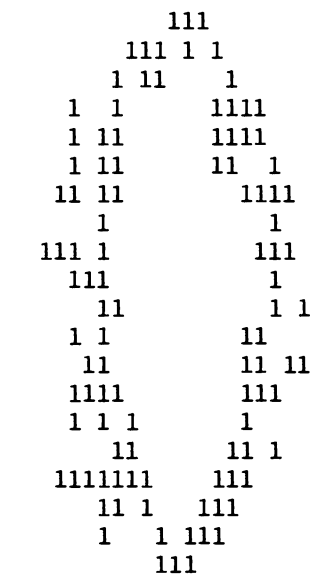


- 2 log likelihood

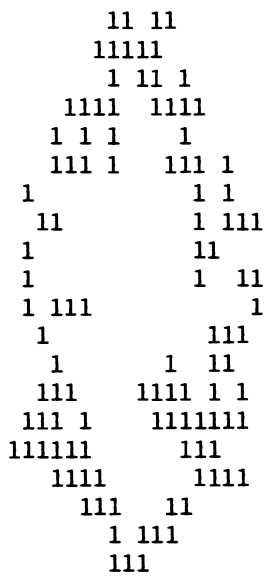
from 0	- 97.7
from Q	- 43.7
from A	920.2
from E	8931.5
from I	27429
from 1	28026

from 0	72.5
from Q	105.6
from A	1526.4
from E	8630.7
from I	27875
from 1	28417

Sample from 0 with 40% error



Sample from 0 with 50% error

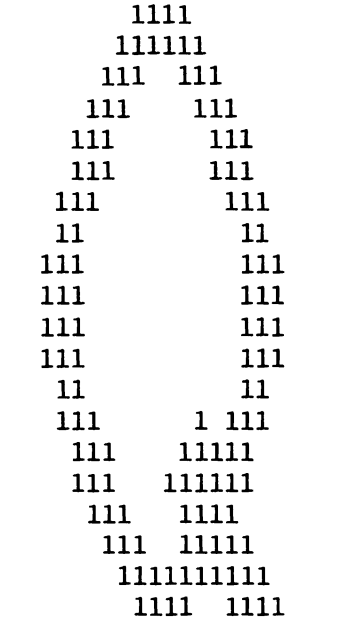


- 2 log likelihood

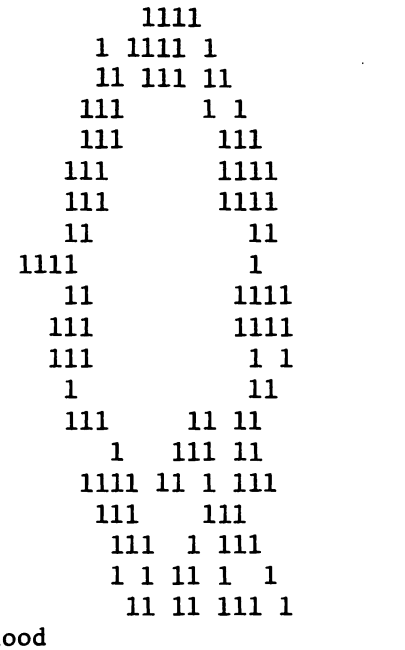
from 0	437.9
from Q	617.2
from A	2908.5
from E	8423.5
from I	29973
from 1	30022

from 0	227.9
from Q	346.8
from A	1463.8
from E	11686.1
from I	21388
from 1	21578

Calibration Pattern Q



Sample from Q with 20% error

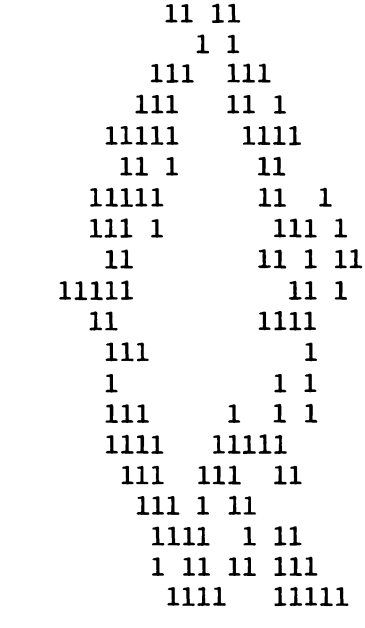


- 2 log likelihood

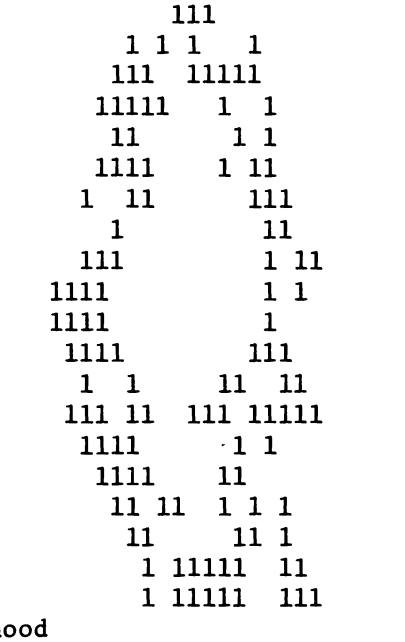
from Q	- 98.4
from O	- 39.9
from A	1379.1
from E	10466
from I	31516
from 1	31969

from Q	- 13.4
from O	- 4.2
from A	937.9
from E	9600.7
from I	28003
from 1	28551

Sample from Q with 40% error



Sample from Q with 50% error

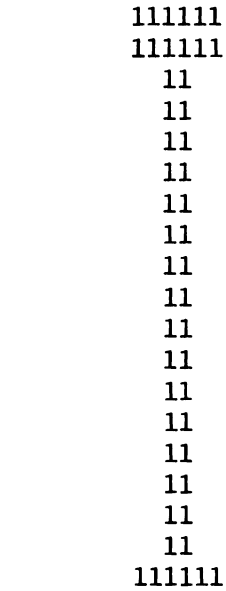


- 2 log likelihood

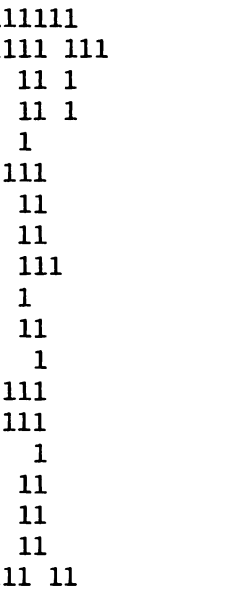
from O	511.5
from Q	529.8
from A	1585.9
from E	10848
from I	28720
from 1	29341

from Q	385.2
from O	421.7
from A	1925.5
from E	11749
from I	28467
from 1	29087

Calibration Pattern I



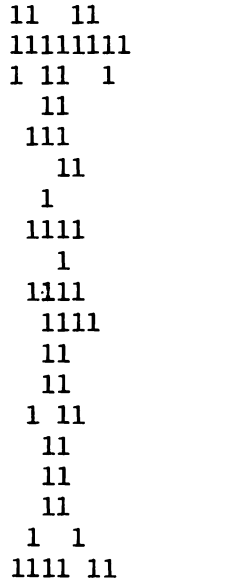
Sample from I with 20% error



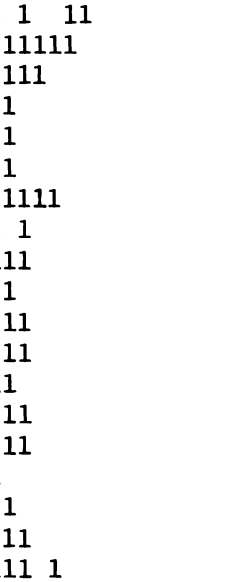
- 2 log likelihood

from I	- 141.3	from 1	453.6
from 1	- 60.5	from I	488.1
from 0	1215.1	from 0	930.3
from Q	2197.3	from Q	1827.3
from A	3037.1	from A	1964.3

Sample from I with 40% error



Sample from I with 50% error

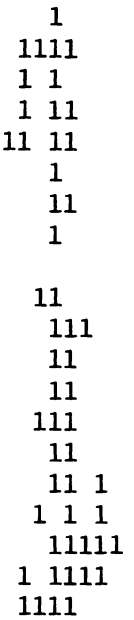


- 2 log likelihood

from 0	911.6	from 0	840.3
from A	1171.8	from A	1285.5
from I	1398.1	from Q	1675.8
from Q	1815.6	from 1	2011.4
from 1	1835.8	from I	2142.0

Calibration Pattern 1

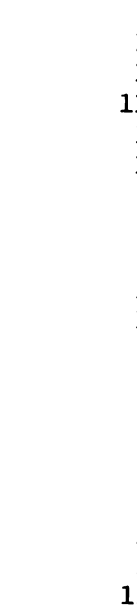
Sample from 1 with 30% error



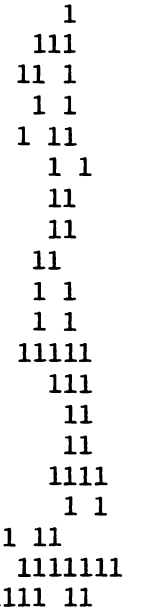
- 2 log likelihood

from 1	- 140.4	from I	907.7
from I	57.3	from 0	916.6
from 0	1231.6	from 1	937.1
from Q	2217.6	from Q	1789.6
from A	3218.2	from A	1976.3

Sample from 1 with 40% error



Sample from 1 with 50% error



- 2 log likelihood

from 0	911.3	from 0	802.8
from 1	1264.7	from A	956.0
from A	1408.6	from Q	1643.6
from I	1630.9	from I	2867.3
from Q	1773.1	from 1	3316.5

STATISTICAL ANALYSIS OF CUTTING FLUID PERFORMANCE DATA

Lanny D. Wells
U.S. Army Weapons Command
Research and Engineering Directorate

ABSTRACT. A 2^4 factorial experiment was conducted to determine the effects of 4 factors in a single-point tool, turning operation. Factors considered were A (tool material), B (cutting fluid type), C (fluid application method), and D (fluid concentration). Factor B (cutting fluid type) was of primary interest in this experiment.

An analysis of variance was performed using Yates' technique to test significance of the different factors and interactions and to determine the relative importance of these different effects.

The results of this analysis indicate that the type of cutting fluid is a relatively unimportant factor compared with the method of application and the concentration of the cutting fluid.

INTRODUCTION. Cutting fluids are applied to various metal cutting tools to help prevent excessive heat buildup and to reduce friction at the tool-chip interface. A number of beneficial effects can be obtained if a cutting fluid can perform these functions. Tool life can be extended; or, higher cutting speeds can be used while maintaining the same tool life; or, some combination of higher speed and longer tool life can be obtained. Tolerances and surface finish may improve or be easier to maintain with an effective cutting fluid.

Various users and manufacturers of cutting fluids have developed formal performance tests to evaluate and compare different cutting fluids, mainly for their own special interests. Unfortunately, these tests have not been standardized; no specific procedure has been widely accepted; and, rarely, is any formal significance test made. Also, the importance of optimizing the cutting fluid is not usually determined relative to the importance of optimizing other parameters such as tool geometry or material. In many cases elaborate programs are set up for cutting fluid selection; but, in the same shop no organized effort is made to optimize cutting speeds and feeds or any of the other parameters affecting the machining operation. In fact, experimental design and statistical analysis have been notoriously lacking in the whole field of metal cutting research. A typical comment overheard in a conversation between some colleagues went something like this: "Statistics is fine, but we can't run that many tests in metal cutting." The idea that a great number of test runs is necessary to facilitate statistical analysis is complete nonsense! Experiments can often be reduced in size by proper design and consideration of the analysis to be performed. It is certainly uneconomical to make experiments larger than necessary.

A 2^4 FACTORIAL EXPERIMENT. As an illustration of a type of experimental design which can be used in the metal cutting field, the remainder of this paper describes a 2^4 factorial experiment (4 factors each at 2 levels). This experiment was conducted to determine the effects of four factors in a single-point, lathe, turning operation. These factors were:

Factor A - Tool material
Factor B - Cutting fluid type
Factor C - Method of fluid application
Factor D - Fluid concentration

Each factor was tested at two levels, thus, making an experiment of 16 observations. The two cutting fluids tested were Fluid A (a heavy-duty, chlorinated, soluble fluid) and Fluid B (a fluid specially formulated for mist application). Each fluid was used at two different dilutions (20:1 and 35:1), and the two different methods of application were conventional flood and mist.

It should be understood that an experiment of 16 observations is certainly a small experiment; but, it could be readily expanded by adding more factors and/or using more than 2 levels. The mathematical model of this experimental design was:

$$Y_{tijk} = M + A_t + B_i + C_j + D_k + AB_{ti} + AC_{tj} + BC_{ij} + AD_{tk} + BD_{ik} +$$

$$CD_{jk} + ABC_{tij} + ABD_{tjk} + BCD_{ijk} + ABCD_{tijk}$$

The tool life was obtained for each of the 16 different treatment combinations at 4 different cutting speeds. A computerized regression analysis gave a tool life vs. cutting speed relationship of the form $V = V_1 T^n$. Where T = tool life (minutes), V = cutting speed (surface speed of workpiece in feet per minute), V_1 = cutting speed for 1 minute tool life, and n = a determined exponent. Estimates of V_{20} (the cutting speed corresponding to a 20 minute tool life) was obtained from these equations.

These estimates of V_{20} are presented in Table I. This data was then used in a formal analysis of variance using Yates' technique (Table II).

The Yates' Technique gives the sums of squares for all the effects without the need of memorizing or looking up any equations and, thus, is a powerful tool for analysis of variance. The ANOVA table is shown in Table III. The 4-three factor and the four factor interactions have been pooled to form a residual term with 5 degrees of freedom. This is justified in this case since all of these terms are of the same order of magnitude.

TABLE I

Test Results (V_{20})

		A_o (Carbide)		A_1 (Cast alloy)	
		b_o (fluid A)	b_1 (fluid B)	b_o (fluid A)	b_1 (fluid B)
C_o (flood)	d_o (20:1)	231	232	68	80
	d_i (35:1)	225	202	69	49
C_1 (Mist)	d_o (20:1)	217	219	88	94
	d_1 (35:1)	219	196	89	95

TABLE II

Yates' Technique

Treatment	Yield	(1)	(2)	(3)	(4)	SS $(4)^2 \div 16$
(1)	231	299	611	1229	2373	------
a	68	312	618	1144	-1109	76,867.5625
b	232	305	545	-569	-39	95.0625
ab	80	313	599	-540	47	138.0625
c	217	294	-315	21	61	232.5625
ac	88	251	-254	-60	139	1207.5625
bc	219	308	-309	15	21	27.5625
abc	94	291	-231	32	19	22.5625
d	225	-163	13	7	-85	451.5625
ad	69	-152	8	54	29	52.5625
bd	202	-129	-43	61	-81	410.0625
abd	49	-125	-17	78	17	18.0625
cd	219	-156	11	-5	47	138.0625
acd	89	-153	4	26	17	18.0625
bcd	196	-130	3	-7	31	60.0625
abcd	95	-101	29	26	33	68.0625
totals	2373					79,807.4375

TABLE III

ANOVA Table

Source	SS	DF	MSR
A (tool material)	76,867.5625	1	2,057.345***
B (fluid)	95,0625	1	2.544
C (Application Method)	232.5625	1	6.224*
D (Concentration)	451.5625	1	12.086*
AB	138.0625	1	3.695
AC	1207.5625	1	32.320**
AD	52.5625	1	1.407
BC	27.5625	1	.738
BD	410.0625	1	10.975*
CD	138.0625	1	3.695
ABC 22.5625			
ABD 18.0625	186.8125	5	
ACD 18.0625			
BCD 60.0625	$\frac{186.8125}{5} = 37.3625$		
ABCD 68.0625			

$F_{1,5,.95} = 6.61$ $F_{1,5,.99} = 16.26$ $F_{1,5,.999} = 47.18$

*significant at 95% confidence level
**significant at 99% confidence level
***significant at 99.9% confidence level

INTERPRETATION OF ANOVA TABLE. The significant AC (tool material X application method) interaction indicates that the application method best for one tool material may not work well on the other tool material. Also, the high BD (fluid X concentration) interaction indicates that the best concentration depends upon the fluid used.

The cutting fluid type (Factor B) appears to be a relatively unimportant factor compared with the application method and the concentration.

The tool material (Factor A) was a very highly significant factor, as expected, since carbide and cast alloy are quite different in character. This factor was so dominant that it appeared to be desirable to analyze the data for carbide and cast alloy as two separate experiments. This was done, and the results of this analysis are presented in Table IV and V, respectively.

INTERPRETATION OF ANOVA TABLE FOR CARBIDE. Analysis of data using carbide tools shows that all of the main effects were formally significant in the following order:

1. Factor D - (Concentration)
2. Factor B - (Fluid type)
3. Factor C - (Application method)

The best combination for carbide was flood application of fluid A at the 20:1 concentration.

INTERPRETATION OF ANOVA TABLE FOR CAST ALLOY. Considering the cast alloy tool material alone, only Factor C (method of fluid application) was formally significant. Mist application was much better with cast alloy tools.

CONCLUSION. As this paper clearly illustrates, statistical design and analysis can be effectively used in metal cutting experiments. The factorial design is particularly well suited to these experiments. Yates' Technique, applied to a factorial experiment, is not difficult and can be carried out without any computational equipment.

TABLE IV

Yates' Table (Carbide)

Treatment	Yield	(1)	(2)	(3)	$(3)^2 \div 8$ SS
(1)	231	463	899	1741	--------
b	232	436	842	-43	231.125
c	217	427	3	-39	190.125
bc	219	415	-46	1	.125
d	225	1	-27	-57	406.125
bd	202	2	-12	-49	300.125
cd	219	-23	1	15	28.125
bcd	196	-23	0	-1	.125
total	1741				1155.875

ANOVA TABLE (Carbide)

Effect	SS	DF	MSR
B	231.125	1	1849*
C	190.125	1	1521*
D	406.125	1	3249*
BC	.125	1	1
BD	300.125	1	2401*
CD	28.125	1	225*
BCD	.125	1	

$F_{1,1,.95} = 161.4$

Factors
B - Fluid
C - Application method
D - Concentration

TABLE V

Yates' Table (Cast Alloy)

Treatment	Yield	(1)	(2)	(3)	$(3)^2 \div 8$
(1)	68	148	330	632	-------
b	80	182	302	4	2
c	88	118	18	100	1250
bc	94	184	-14	20	50
d	69	12	34	-28	98
bd	49	6	66	-32	128
cd	89	-20	-6	32	128
bcd	95	6	-26	32	128
total	632				1784

ANOVA Table (Cast Alloy)

Effect	SS	DF	MS	MSR
B	2	1	2	.018
C	1250	1	1250	11.521*
D	98	1	98	.903
BC 50 BD 128 CD 128 BCD 128	434	4	108.5	---

$F_{1,4,.95} = 7.71$

Factors
B - Fluid
C - Application method
D - Concentration

MEASUREMENT OF ONE ASPECT OF VEHICULAR MOBILITY

Carol D. Rose and Raymond Owens
U. S. Army Tank Automotive Command
Vehicular Components and Materials Laboratory
Laboratories Support Division
Design of Experiments Branch
Warren, Michigan

ABSTRACT. Measurement of One Aspect of Vehicular Mobility.

Measurements of vehicular mobility have usually been conducted as "go-no-go" tests, in which vehicles are matched against obstacles until they can no longer proceed, or "jury system" tests which rely upon qualitative judgments based on opinions of observers and/or drivers of the vehicles under test. As a new approach this project investigates the feasibility of using a statistically designed test which is reasonably unbiased and provides some measurement of precision for evaluating mobility of the vehicles.

The paper describes the design problems presented for developing a test program, the experimental design selected, the field conduct of the test program, and results of the test. Test data were limited to time required for a vehicle-driver combination to traverse a prescribed course. The report covers a total of 450 runs, using 18 drivers, ten vehicles, and 27 test courses over three different terrains.

ACKNOWLEDGMENT. The assistance of Project 07312, Willow Run Laboratories, Institute of Science and Technology, the University of Michigan, under Contract No. DA-20-113-AMC-05927(T), with Emil H. Jebe as Principal Investigator, is acknowledged in the design of the experiment, preparation of computer programs for analysis of the test data, interpretation of the experimental results and the preparation of this report.

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

INTRODUCTION

Mobility has long been a major aspect of consideration in warfare. In the year 218 B. C., Hannibal crossed the Alps and subsequently won the first of many battles from the Romans. In addition to horses Hannibal utilized a few elephants which apparently increased his overall mobility of materiel.

With the advent of motorized vehicles considerable progress was made in the transportation of men and materiel. This progress was due in great part to the roads and highways which were built as part of the transportation complex.

Roads are often not available to supply front line troops during wartime or for other use during national emergencies. In recent years then, a prime consideration in the design of a military vehicle has been off-the-road mobility.

Measurements of vehicular mobility generally have been grouped into two types, the "go-no-go" and the "jury system". In the "go-no-go" type, the vehicles are pitted against various obstacles- ditches, steep inclines, swamps, etc., until they can no longer proceed. The "jury system" uses the combined opinions of the drivers and observers for evaluation. These tests give useful results but are subject to certain weaknesses. For example, the courses are usually well defined, not properly replicated, and performance of a vehicle can be greatly influenced by the driver.

As a new approach, this project investigated the feasibility of using a statistically designed test which is reasonably unbiased and provides some measurement of precision for evaluating mobility of the vehicles in a tactical cross-country situation. In a tactical situation, the driver often may not be familiar with the area, and paths to follow are not defined. Roads may be mined. The driver may avoid obstacles if possible, and the time required to reach a destination may be an important factor for the successful completion of a mission.

CONSIDERATIONS OF THE TEST

At an early stage in the development of the statistical design, some basic issues were resolved.

1. These tests were intended to measure only one aspect

of mobility. This was the time required for a vehicle-driver combination to traverse from point A to point B where the course is defined only by the points A and B except where auxiliary markers may be needed to keep the driver on course.

2. The experimental unit was the course. It was not practical to provide the number of courses required to perform all the desired tests and still have the drivers limited to only one traverse of a course. This aspect was desired; otherwise a learning factor would be introduced when a driver traversed a course more than once. As an alternative perhaps many drivers could be used and thus reduce the number of courses required.

3. The courses selected would be about the same length, approximately measured, and not accurately surveyed. A course length of somewhere between 5 and 20 miles seemed reasonable. [Examination of the data showed that the actual lengths varied from 0.6 mile in Terrain III to 2.7 miles in Terrain I, approximately].

4. The tests were to be conducted in Nevada with the cooperation of the Nevada Automotive Test Center. Three types of terrain were selected to give greater meaning to any results or conclusions obtained. The terrains were defined as follows:

a. Terrain I: Flat and open with small irregularities in the form of dry washes, and scattered areas of sagebrush one to two feet in height. Obstacles were minor in nature.

b. Terrain II: Hilly and open with rolling hills, and areas of deep washes and sheer drops. This area contained outcrops of rock and scattered areas of sagebrush similar to Terrain I.

c. Terrain III: Hilly and timber covered. Areas of trees were scattered between open spaces of sagebrush and grass. The trees were closely spaced pine ranging between five and twenty-five feet tall. This was the most difficult of the three terrains.

5. The supply of drivers was not a problem. However, the supply of experienced drivers was limited. By definition, a driver was classified as experienced or novice according to his own statements as to his ability and/or experience to drive on the highway and cross country.

6. It was planned that the drivers would be instructed to traverse the course at the fastest speed they felt they could go without damaging the vehicle or injuring personnel.

7. A referee was to ride in each vehicle. The referee was the official timekeeper. He would also record any other information that might effect interpretation of the data. For example, a driver may become bogged down, or lost, or the vehicle may not be performing properly. The referees were also responsible for the safety of the vehicles and occupants by having the driver avoid any maneuver which could result in damage to the vehicle or occupants. The referees were to be familiar with the particular courses to which assigned.

8. Nine or ten vehicles were expected to be available for this test. The ones used would be those available at the time of the test.

CONSIDERATIONS IN THE DESIGN

The primary interest in these tests was to determine if a designed experiment could be useful for evaluating factors that affect the mobility of vehicles. This objective could be met if it were possible to design a test which could differentiate between vehicles, at a specified confidence level. Any other information obtained would be useful for designing future tests.

Considerations were as follows:

1. Vehicle effect
2. Course effect
3. Driver effect
   a. experienced
   b. novice
4. Terrain effect
5. Marking of courses
6. Order of testing. The tests were expected to require several weeks. The weather could be a factor.
7. Tracks left from a previous run on the same course.
8. Referee effect
9. Interaction effects
   a. Vehicle - course
   b. Driver - course
   c. Vehicle - driver
   d. Vehicle - terrain
   e. Driver - terrain

SELECTION OF THE DESIGN

In the considerations of the design there were at least five major factors that had to be accounted for in the design. These were order of testing (runs), courses, vehicles, drivers, and terrain. The other factors would have to be controlled by conducting the test with care or, considered not significant. Comments are as follows:

1. Marking of the courses should present no problem in Terrain I but in the hilly and/or timber covered Terrains II and III care should be exercised so that a driver could easily determine the course by following the check point markers.

2. After a course was used once there would then be a path to follow. It was decided that before a test run was made, each course would be traversed once. In addition, each course would have two or three false trails at the start. The purpose of this was to give the first driver an environment similar to that of the following drivers. Drivers were instructed not to follow previous tracks unless absolutely necessary. Generally there were no roads to follow but in case a driver did come across an established road he was instructed to assume it was mined, in which case his maximum speed could not exceed the two or three miles per hour of mine sweeping operations.

3. The referee effect was to be controlled by careful selection and uniform instruction to those selected as referees. Also, the referees were to establish the courses so they could become familiar with them before the tests were started.

4. Each course could have been laid out across all three terrain types. This would still satisfy the primary objective of the experiment, but it would give no information on terrain effect nor on the interaction effects of vehicle-terrain and driver-terrain.

5. One way to cope with a problem of this size is to adopt the Graeco-Latin square as the basic structure for the experimental plan. With this choice only four factors can be used. The basic structure would include runs, courses, vehicles, and drivers. To obtain any evaluation of terrain effect, each square would have to be repeated for each terrain. The Latin-square and Graeco-Latin square have the limitation that no interaction effects can be measured. It seems reasonable that

there probably are some inter-action effect. If present, these effects would inflate the error sum of squares and decrease the sensitivity of the test. In retrospect, one driver was unable to complete some of the runs in Terrain III because of his inability to handle the vehicle on these courses. Nevertheless, it was assumed that interaction effects would not seriously affect the analysis and the Graeco-Latin square was adopted as an acceptable design for this experiment.

6. Information was desired on experienced driver versus novice driver. The test was designed such that nine of each were assigned. Drivers were randomly divided into two groups with the requirement that one group contain four experienced and five novice drivers and the second group contain five experienced and four novice drivers. Each group was then assigned to either the first or second square for each terrain.

THE GRAECO-LATIN SQUARE

A Graeco-Latin square of side N is defined as a square layout of N rows and N columns with N Latin and N Greek letters filling the N^2 cells with the following restriction: each letter (Latin or Greek) may appear only once in each row and once in each column, and each Latin-Greek combination may appear only once. Graeco-Latin squares do not exist for all sizes. A square of size six is not known. One of side ten was only recently determined.

In this experiment the Latin treatment represents vehicles and are designated by capital letters. The Greek treatment represents drivers and are designated by numbers. Rows and columns represent order of run and course, respectively.

Correct randomization procedures must be used when conducting the experiment using a Graeco-Latin square design. The general procedure is as follows: Randomly select a square of the size required from a listing of the squares that are different from one another; that is, they are not convertible into one another by permuting rows and/or columns. After selection of a basic square, the rows are permuted randomly, then the columns are permuted randomly. Finally, the Latin and Greek treatments are randomly assigned.

PROBLEMS ENCOUNTERED DURING THE TEST

1. It was anticipated that some drivers would become disoriented while traversing a course. One task of the referee was to prevent this when it appeared the driver was in the process of becoming lost. There were a few incidents of this nature, including one where the referee also became disoriented. These incidents were recorded by the referee. Upon completion of the test the project engineer and the referee discussed the individual incidents and made a decision whether or not to accept the elapsed time as a data point or discard the data as an outlier. In general, when a driver became lost for more than four or five minutes, that time datum was rejected, since this was not the fault of the vehicle, and the vehicle was the factor of primary interest. A discarded test run was not rerun.

2. In a few instances, a vehicle bogged down. Again, when excessive time was required for the vehicle to again get under way, the time datum for that run was rejected.

3. A driver-terrain interaction effect or more precisely, a driver-course interaction effect became evident during the test. In particular, one driver lost confidence in controlling some of the vehicles during the tests in Terrain III. In these instances, the referee had to drive the vehicle back to camp. Since the time datum for these runs were not used any analysis for driver-terrain interaction would be biased.

4. Some of the courses within the same terrain were more difficult to negotiate than others in the same terrain. Differences in vehicles contributed to an apparent interaction effect. For example, Vehicle I was an armored car, and this vehicle has a high center of gravity which could be dangerous in the hilly courses of Terrains II and III. Two vehicles were driven with the hatch closed and vision was limited to that obtainable through the vision blocks. Conditions of this nature did result in a few uncompleted runs (as previously mentioned), or data which were subsequently not used.

5. The referees did not react equally to hazardous situations. During off-duty hours, the drivers would discuss actions of the referees. Thus the drivers obtained an insight into how a referee would react under certain conditions. As a result, the drivers had a tendency to modify their driving according to who the referee was.

6. The referees also were not uniform in controlling the test when a driver wandered off course. The time allowed before a referee gave the driver instructions in these cases apparently varied considerably. Since the data obtained were the times required to traverse the courses, the referees did influence the outcome of the tests.

7. Drivers were instructed not to follow trails left by previous vehicles. By the time the second square was begun courses were covered with trails and it became increasingly more difficult to keep the drivers off these trails. The subsequent analysis of variance data did not show a significant run effect at the five percent significance level.

8. Vehicle E1, a 5000-gallon fuel tanker-truck was withdrawn from the test after completion of the runs of Terrain I. This vehicle was difficult to control over the basically flat terrain of Terrain I. It was judged best for the safety of the drivers and vehicle not to use that vehicle for Terrains II and III. A 1-1/4-ton cargo truck designated E2 replaced Vehicle E1 for Terrains II and III.

9. There were instances of mechanical breakdown of a vehicle during a test run, which required varying amount of time to repair. There were also instances when a vehicle performed below par. This again would result in a judgment by the project engineer and referee whether to accept or reject the time datum for that run.

10. Because of mechanical difficulties Vehicle C proved inadequate in Terrain III. Tests with this vehicle were stopped after the first square in Terrain III. A replacement vehicle was not available so an 8 X 8 Graeco-Latin Square had to be designed for the second square of Terrain III, in lieu of the 9 X 9 size used for the previous five squares.

11. Vehicle I broke an axle and did not finish tests in Square 2 of Terrain III.

THE DATA

Data to be analyzed were data for Squares 1 and 2 for each of Terrains I, II, and III.

Data for one of these six Graeco-Latin squares is shown in Table I. The small squares indicate where data are missing. The

minimum and maximum times required for traversing a course are shown within the two circles. The extremes for this square give a range of 5.1 to 52.8 minutes. Other recorded data were used to compute average vehicle speed and the average vehicle miles driven per course. A summary of these data are shown in Table II. The courses as laid out were much shorter than originally suggested. The courses are listed as miles driven rather than length. In Terrain III especially some of the larger vehicles with a large turning radius had to detour around some obstacles that smaller vehicles could negotiate.

The overall data obtained for analysis showed the following:

Terrain I, Square 1 had one empty cell.
Terrain I, Square 2, Terrain II, Squares 1 and 2 were complete.
Terrain III, Square 1 had eleven empty cells.
Terrain III, Square 2 had seven empty cells in the 8 X 8 square.

ANALYSIS OF THE DATA

Analysis of Variance Tables for the squares having no empty cells were computed in a straightforward manner. Analyses were performed in two ways for the three squares having empty cells. The first analyses were obtained by estimating the missing data, then performing the standard analysis of variance computations. This method results in an upward bias for the treatment sum of squares, so the data were also analyzed by regression analysis to obtain an unbiased value for the treatment sum of squares.

To determine the missing values, these missing data were designated as a, b, c, ... etc. Then steps were set up for an analysis of variance. The error sum of squares is defined in the usual manner; that is, it is the remainder after the treatments sums of squares are subtracted from the corrected total sum of squares. The error sum of squares is thus determined in terms of the unknowns. Partial differentiation is performed on the error term with respect to each of the unknown missing values and derivatives are set equal to zero. The resulting set of equations is solved for the missing values. Since the error term was minimized, the remainder sum of squares is unbiased.

This analysis was also obtained using the experimental design model:

$$y = \sum_{i=1}^{p} \beta_i X_i + e \qquad i = 1, 2, \ldots, 37 \text{ (for 9 X 9 square)}$$

This is a general linear model of less than full rank. The X_i's take only values of 0 or 1. In the 9 X 9 square, the 9 levels for each of the four factors, plus b_1 for the mean, gives an X matrix of size n X 37 with n equal to the number of observations. Square 1 of Terrain III, with eleven missing values gave an X matrix of size 70 X 37, and $\hat{\beta}$ was solved from the normal equations for this model of

$$X^T X\hat{\beta} = X^T Y$$

A solution was derived by arbitrarily equating to zero the β_i's corresponding to the ninth level for each of the four factors, partitioning the matrices, and solving the reduced X matrix of size 70 X 33, which was of full rank. (reparametrization). Now, one of the conditions that may be applied in solving the regression equation is that the sum of the β's for each factor is equal to zero. A linear transformation was imposed on the $\hat{\beta}$'s to meet this condition as follows: The $\hat{\beta}$'s for each factor were summed, the result divided by nine, and this amount subtracted from each of the nine $\hat{\beta}$'s. The general mean was also adjusted the same amount. The $\hat{\beta}$'s or b's were thus departures from mean time and could be interpreted directly. A large negative b meant that this level of the factor had the effect of traversing the course in a much shorter time than the average time.

Additional computations were performed on the squares with missing data to obtain the sum of squares for vehicles and drivers for an ANOVA table.

The items of main interest were vehicle effects and driver effects. For these effects, differences of the means were tested using Duncans Multiple Range Test at the five percent significance level.

RESULTS

Primary analyses of the data were summarized in ANOVA tables for the six squares. One of these tables is shown as Table III. It is noted that only vehicles and drivers were randomized. The courses are assumed to be a random sample from a population of courses. Then a significance test for courses is valid. Runs cannot be randomly assigned. Thus the sum of

squares for runs can be computed but a significance test for runs is not valid. The magnitude of the ratio of mean square for runs to remainder mean square was small for all six squares. When tested at a significance level of five percent, the course, vehicle, and driver effects were always significant at this level except in two cases and in these two cases the significance level was less than 10 percent. Results of the analyses are summarized in Table IV, which shows the F-ratios and their respective significance levels.

Although two squares were run in each terrain, true replication was not obtained because a different set of drivers were used in each square. Under the assumption that each set of nine drivers per square were approximately equal, an analysis of variance was made on the combined Squares 1 and 2 for Terrains I and II. The squares of Terrain III were not combined. Combining the squares made the tests more sensitive for differences between vehicles.

A comparison of experienced versus novice drivers was made by partitioning the sum of squares for drivers. The F-ratios did not show a significant difference between experienced and novice drivers for any square, nor for the combined squares, at the five percent level.

Application of Duncan's Multiple Range Test applied to the means gave separation of vehicles into groups which were significantly different from one another, at the five percent significance level. See Table V. Some vehicles fall between two adjacent groups and cannot be considered different from either group. These vehicles are indicated by connecting lines to the main groups in the table. For example in the upper left square, the group D, F, and A was the fastest, followed by the group B, I, G, C, and then vehicle E1. Vehicle H can be associated with either of the two groups indicated.

COMBINED ANALYSIS FOR TERRAINS

The vehicle mobility test was designed around the individual Graeco-Latin Square. It was not designed so the six squares over the three terrains could be pooled in a straightforward manner. Any analysis over the three terrains is further complicated by the missing values in Terrain III, and grouping of the drivers into experienced or novice drivers. Main items

of interest in combining the terrains are measures of the relative performance of (1) vehicles over terrains and (2) drivers over terrains.

(1) Vehicles Over Terrains

The problem of missing values was minimized by using only data on vehicles for which information was available for all six squares. A table of means of size 6 X 7 representing six squares and seven vehicles was obtained by omitting data for vehicles C, E1, and E2. See Table VI. The table is still slightly biased because of the inclusion of estimated values for missing data on individual runs. The bias was not considered serious because an estimated value in general was incorporated into an average of eight or nine numbers.

An analysis of variance was performed on the table of means for vehicles. This data is shown in Table VII. The F-ratios confirm that terrain and vehicle effects were highly significant. The main item of interest in this Table, the vehicle X terrain interaction effect, had on F-ratio of 0.93 and thus was not significant.

With the understanding that squares are to be thought of as replicates, for vehicles at least, the entries in the ANOVA for [Sq 1] versus [Sq 2] and [Sq 1] minus [Sq 2] may be used as some measure of "learning". This is a "pseudo-learning" since a different set of drivers was always used in the second square for each terrain. It does indicate, however, that drivers were able to increase speeds by utilizing evidence of trails from earlier runs. The [Sq 1] verus [Sq 2] mean square provides an estimate of "learning" over the whole experiment (all three terrains). The [Sq 1] minus [Sq 2] comparison provides an estimate of the variation in this learning from Square 1 to Square 2 within each terrain. In both cases the probability of these F-ratios occurring by chance under H_o is less than 0.005. The significant "learning" effect appears to be contradictory to the conclusion of no run effect within each square. That is, if this "learning" effect is the result of tracks or trails left from the previous vehicle, then the "learning" effect should commence immediately after the first run.

The run effect for the six squares was investigated further as follows: First, the run totals were plotted with the order of runs as the abscissa. A least squares linear regression line was added. Although the points appeared

scattered the slopes were negative for five of the six squares, indicating less time to traverse a course as the run number increased. But one square had a positive slope.

Analysis of variance tables were constructed to show the reduction in sum of squares due to linear regression, with one degree of freedom, and the deviation from linear regression, with seven degrees of freedom for the 9 X 9 squares and six degrees of freedom for the 8 X 8 square. If the order of runs is a real effect, then the mean square ratios for reduction in sum of squares due to linear regression should be large. A tabulation of these ratios for the six squares, Terrain I, Square 1 through Terrain III, Square 2, follows:

d.f.	RED. M.S./REM. M.S.
1, 47	6.16
1, 48	0.46
1, 48	1.10
1, 48	4.02
1, 37	0.34
1, 28	2.68

These ratios do not give strong support for the conclusion that drivers used evidence of previous trails to reduce their traverse time within a square. A possible explanation of the pseudo-learning between squares is that during the runs of the first square the drivers were able to comply with the requirement of not using trails left from previous runs; however, compliance to this requirement broke down during tests for the second square.

Performance of vehicles over the three terrains was shown graphically by first subtracting the means for the individual square from the vehicle means. The performance for the vehicles were then algebraically added over the six squares. The departure from mean time for each vehicle could then be plotted as shown in Figure 1. Vehicle F shows the best overall performance. This vehicle is the 1/4-ton M151 A1 Utility Truck. This is a jeep type vehicle. Vehicle D was the commercial Kaiser jeep. Vehicle E1 was the slowest vehicle. This was a 16-ton payload vehicle with an unusual control system and no suspension other than the tires. In Figure 1 it must be remembered that vehicle E1 was used only in Terrain I, Vehicle E2 was used only in Terrains II and III, and Vehicle C was not used in square 2 of Terrain III.

(2) Drivers Over Terrains

There was no replication of drivers within terrains; hence, no direct test for driver X terrain interaction can be obtained if an ANOVA is performed on a table of means for drivers. An ANOVA operation was computed and the mean square term for driver X terrain interaction was 77,915. The magnitude of this term is not large in relation to other relevant mean squares obtained from the data. The remainder sum of squares term in the previous ANOVA table for vehicles over terrains was 22,325. A denominator of this magnitude for determination of the F-ratio would indicate that the driver X terrain interaction effect is significant at about the 0.025 level. The remainder sum of squares for the six basic squares, however, ranged from 35, 938 to 283,164 and it is concluded that the driver-terrain interaction effect cannot be properly assessed.

Data for drivers were summarized in the same manner as for vehicles. Differences between the fastest and slowest drivers within a square ranged from 5.5 to 11.2 minutes except in Terrain III Square 1 the maximum difference was 25.8 minutes. The mean time for all drivers in this square was 30.2 minutes. The major portion of the difference can be attributed to driver Number 3. Overall performance of this driver was poor, as shown graphically in Figure 2.

The total difference in elapsed time, over all terrains, between experienced and novice drivers was 49.9 seconds, or less than one minute. This difference was not significant at the five percent significance level, nor even at the 20 percent significance level. It can be safely concluded that although differences exist among drivers, the differences are between individual drivers and not the subclassification of experienced and novice as defined for this experiment.

BIAS IN TREATMENT SUM OF SQUARES

The percent upward bias of treatment sum of squares were calculated for the three squares having missing values. Results are shown in Table VIII. The percent bias was determined from the ratio of sum of squares determined by supplying estimated values, to the unbiased sum of squares as determined by the regression analysis. It can be seen that the bias for vehicles for Terrain III square 1, with eleven missing values was over 52 percent. Actually, all conclusions for vehicles were the same as both F-ratios were significant at less than the 0.01 level.

However, the 21.4 percent bias did modify the conclusions for drivers in Terrain III, Square 2. (The F-ratio obtained using the unbiased sum of squares was not significant at the five percent level whereas the F-ratio had been significant when computed from the biased sum of squares).

CONCLUSIONS

1. The Graeco-Latin square design for this experiment did allow the separation of various factors so that relative effects of each could be estimated.

2. The linear run effect within squares was not consistent throughout the experiment. However, there was some evidence that the order of testing may sometimes be significant.

3. The course effect was large throughout the experiment and the largest overall contributor to the sum of squares. This means that the courses within a terrain were not homogeneous with respect to time required for a vehicle to traverse the courses.

4. Vehicle effect was significant. It was possible to assess relative vehicle performance by separating vehicles of similar performance into different groups.

5. Driver effect was significant. Individual drivers could be separated into groups of similar performance. However, there was no signifcant difference between the subclassifications of experienced and novice drivers as defined for this experiment.

6. Interaction between vehicles and terrain was not significant.

7. Interaction between drivers and terrain could not be properly assessed.

8. A pseudo "learning" effect between the two squares within a terrain was highly significant. The cause of this effect was not accurately described.

SUMMARY AND RECOMMENDATIONS

The number of possible combinations of the four factors with nine levels is nine to the fourth power or 6561. Since

only 81 observations were taken by using the 9 X 9 Graeco-Latin square, the actual test is equivalent to a 1/81 replicate. Results of these tests and the information obtained were considered very satisfactory. This type of test appears useful for other tests involving mobility of vehicles. Specific points for consideration are as follows:

1. There was some evidence, although not conclusive, that trails left by previous runs influenced subsequent runs. It is also reasonable to expect that variations in the weather and other environmental conditions would affect the outcome of a test run. It is therefore recommended that the order of testing (runs) be built into the design for future tests of this nature.

2. The design must allow for analysis of the effects of differences in courses and differences in drivers.

3. The referee effect was not measured during these tests. Ancillary information picked up during these tests indicate the referee effect may be significant. In a future experiment of this type it may be appropriate to superimpose an additional orthogonal square onto the two orthogonal squares of the Graeco-Latin design to assess the referee effect, i.e., add another language to the design.

4. Since there was no significant vehicle-terrain interaction effect, the size of most future experiments could be reduced by limiting tests to one terrain. As an alternative, courses may be laid out over a varying type of terrain.

5. This general type of statistically designed vehicular mobility test may be extended to determine differences among features of vehicles. Examples:

(a) Different power plants, transmissions, or other components in the same vehicle.

(b) Effects of payload

(c) Tracked versus wheeled vehicles over a particular type of terrain.

REFERENCES

1. Davies, Owen L., "The Design and Analysis of Industrial Experiments", Oliver and Boyd, New York, 1956

2. Duncan, D. B., "Multiple Range and Multiple F Tests", Biometrics, No. 11 (1956).

3. Graybill, F. A., "An Introduction To Linear Statistical Models, Volume I", McGraw-Hill Book Company, Inc., New York, 1961.

4. Kempthorne, O., "The Design and Analysis of Experiments", John Wiley & Sons, Inc., New York, 1952.

5. Fisher, R. A., and Yates, F., "Statistical Tables For Biological, Agricultural and Medical Research", Oliver and Boyd, London, 1948.

6. Owens, Ryamond, "Statistical Design and Analysis For Military Vehicle Mobility Study", TACOM Technical Report Number 10132, 26 July 1968.

COURSES	1	2	3	4	5	6	7	8
RUNS								
	926	855	1323	1228	1931	1491	1878	1425
12	I14	D10	B18	(E2)17	G16	H15	A12	F13
	429	616	977	935	2429	1618	1631	763
13	D12	I16	(E2)13	B15	A10	F17	G14	H18
	630	842	1662	[I10]	662	1554	742	1629
14	(E2)18	B17	D14		F15	A16	H13	G12
	700	544	1164	1597	[I18]	(3169)	968	1075
15	G15	A13	H16	F12		B14	D17	(E2)10
	462	594	1409	741	1972	[I13]	878	896
16	H10	F14	G17	A18	B12		(E2)16	D15
	450	1185	[I12]	728	1144	1968	814	1402
17	B13	(E2)15		D16	H17	G10	F18	A14
	768	712	622	[H14]	1604	1901	[I15]	1098
18	A17	G18	F10		D13	(E2)12		B16
	(304)	541	1749	1173	1594	1202	1285	[I17]
19	F16	H12	A15	G13	(E2)14	D18	B10	

Letters = Vehicles
Numerals = Drivers
Numerical Data = Time to traverse course, in seconds.

VEHICULAR MOBILITY TEST DATA

TERRAIN III, SQUARE 2

TABLE I

	Minutes Per Run	Average Vehicle Speed, MPH	Average Miles Driven Per Course
Terrain I			
Square 1	10.0 - 46.9	5.7 - 12.7	2.5 - 2.7
Square 2	8.0 - 35.9	5.8 - 14.9	
Terrain II			
Square 1	9.2 - 46.6	4.8 - 7.1	1.5 - 2.5
Square 2	7.7 - 34.4	5.5 - 7.8	
Terrain III			
Square 1	7.1 - 74.1	2.3 - 4.1	0.6 - 1.7
Square 2	5.1 - 52.8	2.7 - 5.2	

VEHICLE SPEED AND COURSE DATA

TABLE II

SOURCE	d.f.	S.S.	M.S.	F-Ratio	p
Mean	1	87,405,878			
Runs	8	191,972	23,996	-	-
Courses	8	750,559	93,820	2.56	0.025
Vehicles	8	5,337,965	667,246	18.25	0.0005
Drivers	8	815,749	101,969	2.78	0.025
Error	48	1,754,467	36,551		
Total	81	96,256,590			

ANOVA, Terrain I, Square 2

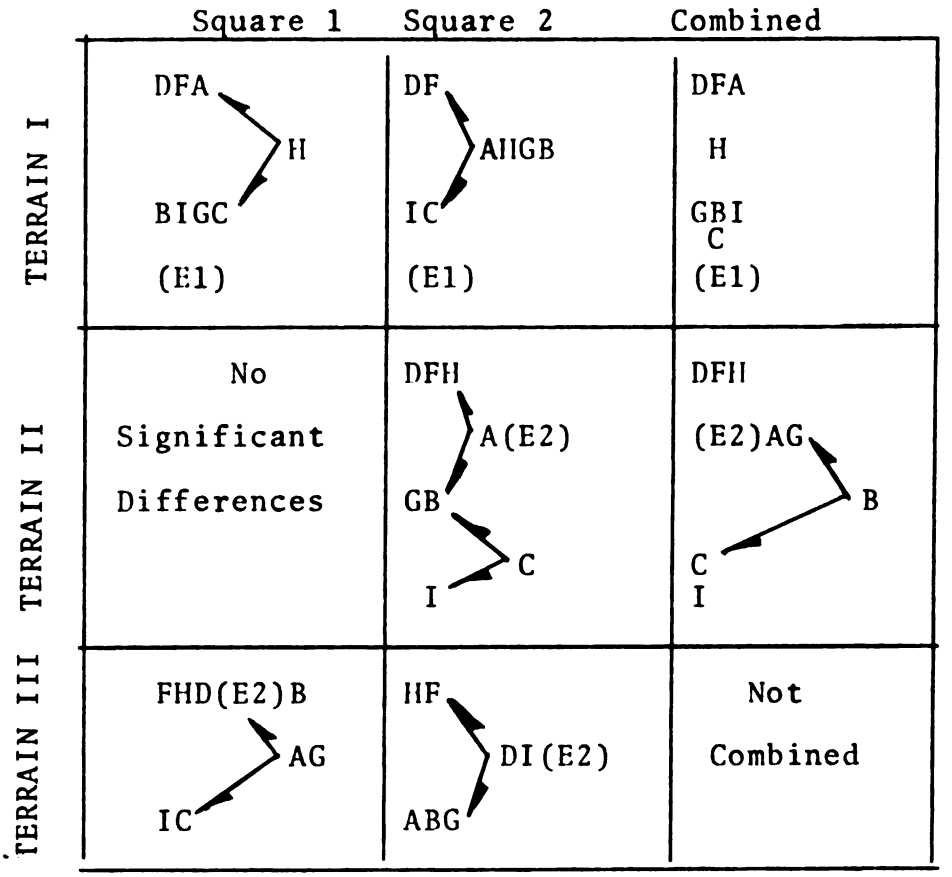
TABLE III

	I-1*	I-2	II-1	II-2	III-1	III-2
	F-RATIOS					
Runs **	1.70	0.65	1.81	0.94	0.84	1.04
Courses	4.90	2.56	4.38	14.26	11.39	10.66
Vehicles	16.68	18.25	1.99	5.13	3.38	2.40
Drivers	3.20	2.78	3.22	10.35	3.91	2.07
	SIGNIFICANCE LEVELS					
Runs **	-	-	-	-	-	-
Courses	.001	.025	.001	.001	.001	.001
Vehicles	.001	.001	.100	.001	.005	.050
Drivers	.010	.025	.010	.001	.005	.100

* Example: I-1 = Terrain I, Square 1
** Significance Test Not Valid

F-Ratios and Their Respective Significance Levels

TABLE IV



DISTINGUISHABLE VEHICULAR GROUPS
AT 5 PERCENT SIGNIFICANCE LEVEL

TABLE V

		A	B	D	F	G	H	I
TER #1	SQ 1	932.9	1140.9	901.5	928.6	1160.2	1067.2	1149.4
	SQ 2	843.7	1070.9	760.1	766.8	1025.6	926.3	1149.9
TER #2	SQ 1	1270.0	1327.2	975.1	1025.6	1260.4	1090.7	1535.0
	SQ 2	1003.2	1077.1	855.8	857.0	1061.3	908.6	1288.7
TER #3	SQ 1	1816.7	1635.9	1578.9	1372.4	2013.5	1416.3	2416.2
	SQ 2	1383.1	1384.2	1043.0	954.5	1394.1	917.2	1060.0

TABLE VI: TABLE OF MEANS FOR VEHICLES OVER TERRAINS

SOURCE	d.f.	M.S.	F-Ratio
Squares	(6)		
Mean	1	58,919,874	
Terrain	2	827,562	37.07
[Sq 1] vs [Sq 2]*	1	940,056	42.11
[Sq 1] minus [Sq 2]**	2	227,064	10.17
Vehicles	6	172,334	7.72
Vehicles X Terrain	12	20,788	0.93
Remainder ***	(18)	22,325	
Vehicle X Square	6	18,641	
Veh X Ter X Sq	12	24,167	
Total	42		

*Over three terrains
**Within Terrains
** For estimate of experimental error

ANOVA For Vehicles Over Terrains

TABLE VII

	PERCENT BIAS		
	Terrain I Square 1 (1 Missing Value)	Terrain III Square 1 (11 Missing Values)	Terrain III Square 2 (7 Missing Values)
Runs + Courses	2.4	12.0	2.0
Vehicle	0.8	52.2	9.1
Driver	2.2	47.0	21.4

UPWARD BIAS IN TREATMENT SUM OF SQUARES
WHEN MISSING VALUES WERE ESTIMATED

TABLE VIII

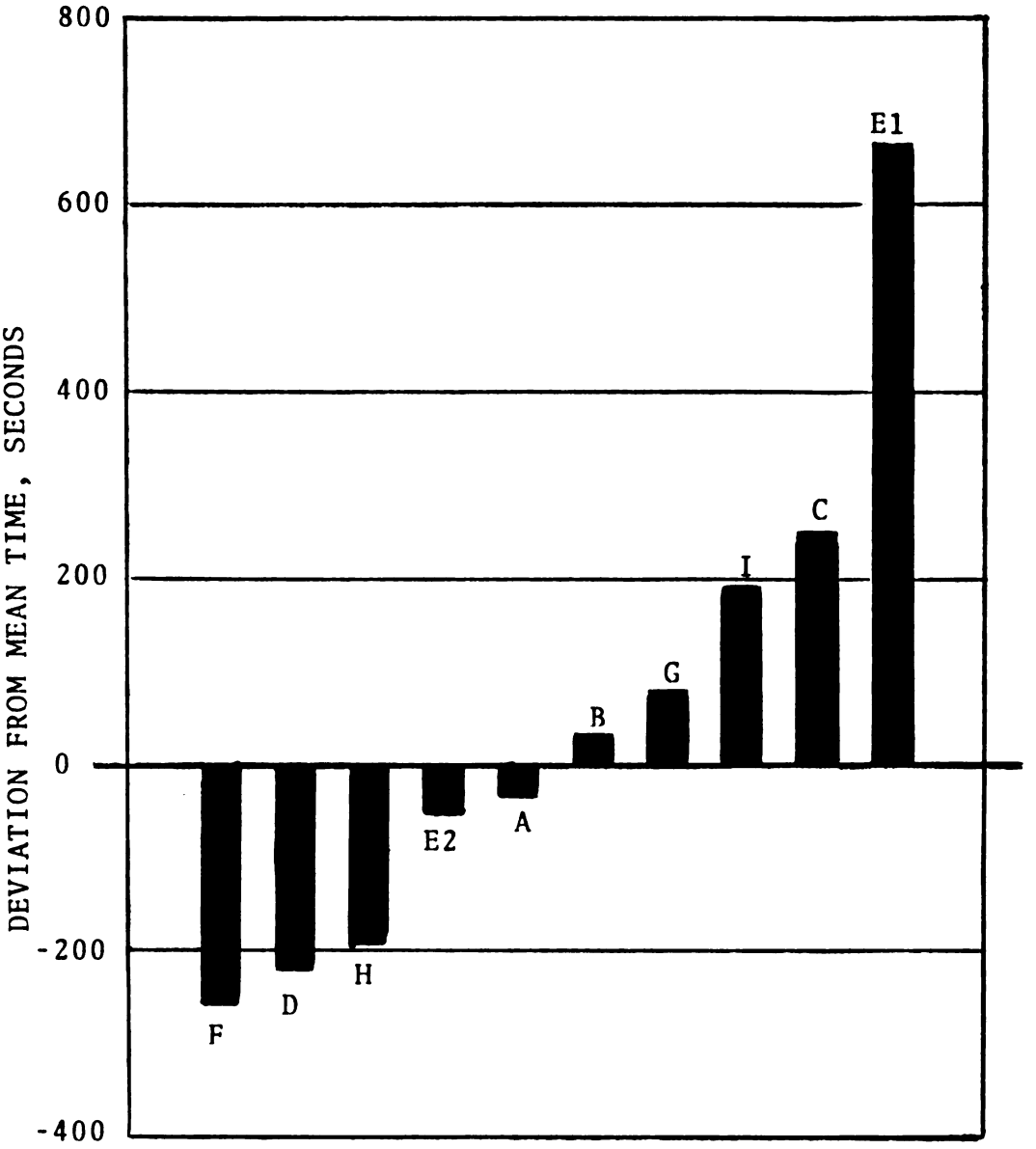


Figure 1

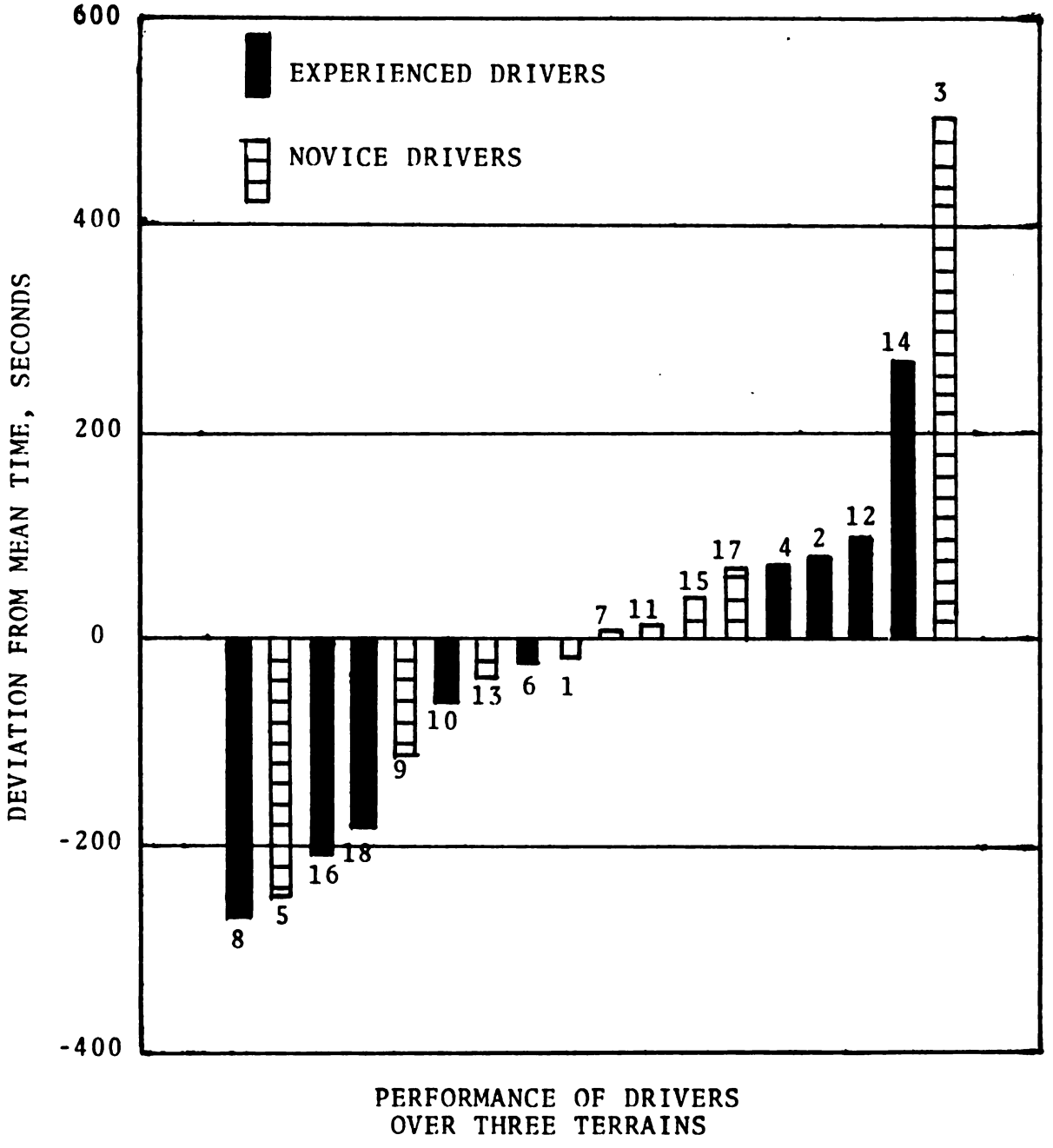


Figure 2

PROBABILITY OF A NON-REPEATABLE OBSERVATION - AN EXAMINATION OF THE UTILITY CONCEPT AND THE NATURE OF QUEUEING SEQUENCES

Mikiso Mizuki
Federal Electric Corporation/ITT
Vandenberg AF Base, California

0. INTRODUCTION. The subjective probability is often defined using the utility concept of gambles and lotteries, cf., de Finetti [2] and Savage [8]. Such an approach gives the only tangible means of measuring the personal assessment of subjective probabilities. However, the basis for this approach seems to be the unstated premise that the gambles are to be played or can be played repeatedly. The expected utility or the weighted mean of gains with the weighting of the probabilities of particular outcomes has a clearly defined meaning under such conditions. On the contrary, the same weighted mean does not possess any practical meaning for a non-repeatable observation. Fishburn [3] concedes that in order to define subjective probability coherently using the utility concept, it is essential to have consequences that can occur under more than one state. This indicates the possibility of modifying the utility theory for non-repetitive random events.

As the second topic of this paper, the nature of queueing sequences is investigated from the same point of view. The queueing sequences constitute non-repeatable observations for each particular service system. An observable queue size sequence is dependent on its companion sequence of arrival/service events. By the above argument, the probability discussed in queueing models of a particular system cannot be interpreted as subjective probability. An investigation on the characteristics of ensembles of queueing sequences is made.

1. UTILITY THEORY AND SUBJECTIVE PROBABILITY. The utility theory is constructed using a mixture space, for instance as defined in [3]. A mixture set consists of a set $\Omega = \{A,B,C,\ldots\}$ and operation $\alpha A + (1-\alpha)B$ which define an associating element of Ω with each $\alpha \in [0,1]$ and each ordered pair $(A,B) \in \Omega^2$ such that, if $A,B \in \Omega$, $\alpha, \beta \in [0,1]$, then

$$1A + 0B = A \tag{1.1}$$

$$\alpha A + (1-\alpha)B = (1-\alpha)B + \alpha A \tag{1.2}$$

$$\alpha[\beta A + (1-\beta)B] + (1-\alpha)B = \alpha\beta A + (1-\alpha\beta)B. \tag{1.3}$$

In repeatedly played gambles, the expression $\alpha A + (1-\alpha)B$ corresponds to the gain (or loss) of mixed outcomes of A's in $100\alpha\%$ of plays and of B's in $100(1-\alpha)\%$ of plays. In particular, if the gain A is set equal to unity and the gain B is set to zero, the utility of the mixture of A's and B's; namely, $\alpha \cdot 1 + (1-\alpha) \cdot 0 = \alpha$, represents the subjective probability that A occurs. The generalization of this gambling situation to non-repetitive random events requires the substitution of the uncertainty of a single random outcome by an aggregate of random observations.

Some of the difficulties are typified by the examples of non-constant valued consequences. For instance, the utility in the sense of social justice of a judge's sentence varies depending on his choice of act of taking the side that the accused did or did not commit the crime; Fishburn, loc. cit. In the risk taking acts of Russian roulette and dangerous mountain climbing, the mental elation, if survives, after the acts gives a different value of being alive from that of not taking the chances. Under such conditions, the linear combinations of utilities of consequences do not have any meaning. And, this is the basis that Fishburn made the statement that subjective probability cannot be discussed for such cases.

The probability assigned to a non-repeatable observation is best formulated as a set of real numbers distributed over an exhaustive set of mutually exclusive possible outcomes. Denote the possible outcomes by A_i, $i = 1,\ldots,n$, and the real numbers assigned to A_i by $P(A_i)$, satisfying $P(A_i) \geq 0$ and $\sum P(A_i) = 1$. Suppose a gain of A_i is made when A_i is observed, where all the gains may be bounded. Then, if A_i is observed, no other A_j's ($j \neq i$) can add to the gain after observing A_i. Because of this, there exist no logical bases for associating a gain of A_i with those of A_j's in the form of the expected utility, $\sum_i$ (gain of A_i) $P(A_i)$.

2. EXPECTATION AND EXPECTED UTILITY. Define a variable which takes on x_i when A_i is observed, and define the indicator function

$$I_{A_i} = \begin{cases} 1 & \text{if the observed outcome is } A_i, \\ 0 & \text{otherwise.} \end{cases} \tag{2}$$

Then, the simple random variable X is given by, cf. Loève [4],

$$X = \sum_i x_i \, I_{A_i} \, . \tag{3}$$

The expectation of X is defined by

$$E(X) = \sum_i x_i \; P(A_i). \tag{4}$$

Suppose the utility of the constant consequences are represented by x_i when the state A_i obtains. Then, (4) is the expected utility of the outcomes. However, as mentioned earlier, the same expression is not adequate for the representation of the utility of a non-repeatable and non-constant valued consequence. In order to circumvent this difficulty, Mizuki [6] suggested an alternative definition of expectation for a non-repeatable observation of the form

$$ENR(X) = \sum_i x_i \; I_{A_i} \; (PA_i) = x_i \; P(A_i) \; I_{A_i} \tag{5}$$

$i=1,\ldots,n$, yielding n different expected values of each possible outcome x_i. The ENR expectation introduced here is consistent with Bayes' definition of probability [1] of any event to be the ratio between the value at which an expectation depending on the happening of the event ought to be computed and the value of the thing expected upon its happening. The use of (5) leads to an interesting modification of the utility theory for a non-repeatable event.

The above formulation is slightly generalized. Suppose there exist a chosen act, denoted by H, and n mutually exclusive states, A_j, $j=1,\ldots,n$, and consequences, u_j^H, measured in some utility, when H is chosen and A_j obtains. The probability that A_j obtains when H is chosen is defined by real numbers $P_H(A_j)$, satisfying $P_H(A_j) \geq 0$ and $\sum P_H(A_j) = 1$. In order to account for the non-constant values of consequences, u_j^H is not necessarily equal to $u_j^{H'}$ for $H' \neq H$. The familiar use of mixed acts is not justified for a non-repeatable situation and will be excluded from the subsequent development. The connotation is that in spite of the mixing operations prior to the final choice of act, the chosen act is unique, thus losing all of its random attributes unlike the case of repeatable events. This eliminates the necessity of defining the probability P(H) assigned over different choice of H's. Under this set of conditions, a simple random variable of utility U_H of a chosen act H is defined by

$$U_H = \sum_{j=1}^{n} u_j^H \; I_{A_j} \; . \tag{6}$$

The ENR expectation of U_H is then given by

$$(7) \qquad ENR(U_H) = u_j^H \; P_H(A_j) \; I_{A_j} \quad ,j=1,\ldots,n.$$

For a choice problem of a non-repeatable event, the expectation given in (7) can be used as the optimizing criterion.

3. A CRITERION FOR PREFERENCE. The individual expectation $u_j^H \; P_H(A_j)$ of (7) may be interpreted as the psychological incentive force acting on a lever at the point of distance u_j^H from the fulcrum with the mass $P_H(A_j)$, whereas the incentive force should be measured at a fixed point on the lever always. Since there exists only one outcome event A_j, the incentive forces can act only individually, but not collectively, for any given decision problem. Application of such an interpretation is considered below.

Savage discusses an example on the choice between two pairs of gambles, pp. 101-103 of [8]. Savage prefers Gamble 1 to 2, and Gamble 3 to 4 after reversing his initial intuitive choice of Gamble 4 over 3 by applying the sure-thing principle. However, the utility theory being developed simply as a normative theory, it is natural to seek an augmented normative theory which explains his initial intuitive choice. The specifications of Savage's gambles are as follows. For the choice between Gambles 1 and 2,

Gamble 1: \$500,000 with probability 1; and

Gamble 2: \$2,500,000 with probability 0.1,
\$500,000 with probability 0.89,
\$0 with probability 0.01.

Similarly, for Gambles 3 and 4,

Gamble 3: \$500,000 with probability 0.11,
\$0 with probability 0.89; and,

Gamble 4: \$2,500,000 with probability 0.1,
\$0 with probability 0.9.

For the sake of simplicity, suppose one acts based on a linear utility function over the range of zero to \$2,500,000 of the form $u(x) = kx$, $k > 0$ for x dollars gain. Using (7), it is immediately seen that the expected utility term of \$500,000 of Gamble 1 is greater than any of

the three expected utility terms of $250,000, $445,000 and $0 of Gamble 2. Likewise, the combination of the expected utility terms of $250,000 and $0 of Gamble 4 is more attractive than the $55,000 and $0 combination of Gamble 3. A similar preference pattern is obtained even in the general case of usual concave utility functions. When the utility function becomes sharply convex, an individual inclines to prefer Gamble 2 to 1, and at the same time, he remains to prefer Gamble 4 to 3. This is a clear indication that the leverage system model can explain the general intuitive choice patterns.

The preference rule examined above can be summarized by:
Dominance of Expectations: Act H is preferred to act G, if

$$u_j^H \; P_H(A_j) > u_j^G \; P_G(A_j) \qquad \text{for corresponding } A_j\text{'s.}$$

This is a partitioned version of the familiar Bayes' principle which maximizes the expected utility (or utilities in this case). The other familiar rules of dominance principle, minimax regret, and maxmin principles remain unchanged for such non-repeatable events. For the details of this development, the readers are referred to [7].

4. ABOUT QUEUEING SEQUENCES. A queueing model is specified by the input process, service time distribution, and the number of servers. The most elementary example is that of Poisson arrivals (M) and negative exponential distribution (M) of service times with a single server (1), or M/M/1 system, which will be examined in the following.

For a particular system, a pair of sequences of customers' arrival and departure times, or equivalently a pair of sequences of queue sizes and arrival/service events can be observed. In the latter pair, the queue size sequence is functionally dependent on the observed sequence of arrival/service events. These sequences are random in nature prior to the observation, but are unique and fixed when it is observed. In other words, these sequences constitute a pair of non-repeatable observations from an ensemble of such pairs. A subjective probability may be used to describe the uncertainties of such samplings. However, there exists a complete analogy with the utility of non-constant valued consequences of non-repeatable event of Section 2. If we use the Fishburn's example of a judge's sentence, the arrival/service events sequence corresponds to the judge's taking the side that the accused did or did not commit the crime, and the queue size sequence corresponds to the social justice. This puts the problem right back to the start.

The M/M/1 models are often analyzed using the birth-and-death process models. Consider a simple birth-and-death process of Poisson input with a constant parameter λ and a negative exponential service time with a constant parameter μ. By denoting the probability that the queue size is n at time t by $P_n(t)$, the standard differential difference equations are introduced, i.e.,

$$(8) \qquad P_n'(t) = -(\lambda + \mu)P_n(t) + \lambda P_{n-1}(t) + \mu P_{n+1}(t) \qquad \text{for } n > 0,$$

$$P_0'(t) = -\lambda P_0(t) + \mu P_1(t) .$$

The original balancing equation is given by

$$(9) \qquad P_n(t+\Delta t) = (1-\lambda\Delta t - \mu\Delta t)P_n(t) + \lambda\Delta t P_{n-1}(t) + \mu\Delta t P_{n+1}(t)$$

for $n > 0$. Notice that there exist two classes of probabilities in (9), namely, one class of $\lambda\Delta t$, $\mu\Delta t$, and $(1 - \lambda\Delta t - \mu\Delta t)$, and the other of $P_n(t)$. The former designates the probability of arrival/service events, and the latter designates the probability of queue sizes. The queue size of a particular M/M/1 system is, however, by definition a step function in time. If the queue size at time t, denoted by q(t), is known, for suitably small Δt,

$$(10) \qquad q(t+\Delta t) = \begin{cases} q(t) & \text{with probability } 1-\lambda\Delta t-\mu\Delta t, \\ q(t)-1 & \text{with probability } \mu\Delta t, \\ q(t)+1 & \text{with probability } \lambda\Delta t. \end{cases}$$

In fact, q(t) may not be known unless it is observed, but q(t) is not a random variable. Rather, q(t) is an observation which is a constant; and, furthermore, q(t) cannot be observed repeatedly for any given t. Thus, q(t) is a single non-repeatable observation. Equation (10) defines that q(t) is a function dependent on another non-repeatable observation over Δt of a new arrival, a departure, or no events.

In the original formulation of the birth-and-death process models, $P_n(t)$ is defined as the proportion of n items in existence at time t with respect to a set of simultaneously observable ensembles, such as bacterial cultures, and particles in chambers. Our primary interest in the behavior of a particular queueing system differs from these cases, and $P_n(t)$ is a representation of the uncertainty for the value q(t) prior to its non-repeatable observation. Since q(t) is known to be unique at t, it is sensible to construct a parametric model shown below:

Define Q(N), Q(A), and Q(L) to be three matrices satisfying

$$Q(N) = (\delta_{ij}) \qquad i,j=0,1,2,\ldots$$

$$Q(A) = (\delta_{i,j-1}) \qquad i=0,1,2,\ldots;\ j=1,2,\ldots$$

(11)

$$Q(L) = \begin{cases} (\delta_{0,j}) & j=0,1,2,\ldots \\ (\delta_{i-1,j}) & i=1,2,3,\ldots;\ j=0,1,2,\ldots \end{cases}$$

At time t the queue size of a particular M/M/1 system is given by a vector $q(t) = (q_0(t), q_1(t), q_2(t), \ldots)$, $q_i(t) = 1$ for some i, $q_j(t) = 0$ for $i \neq j$. The queue size at $t + \Delta t$ is then given by $q(t)Q(x)$, $x = N,A,L$, which will occur with the probability $\Pi(x)$ respectively such that

$$\Pi(N) = 1 - (\lambda + \mu)\Delta t$$

(12)

$$\Pi(A) = \lambda\Delta t$$

$$\Pi(L) = \mu\Delta t.$$

In this formulation Q(x) is a random matrix which takes on Q(N), Q(A), or Q(L) with the probability of $\Pi(N)$, $\Pi(A)$, or $\Pi(L)$.

The two different notions of expectations of (4) and (5) can be applied to the above argument. Let x_i and $P(A_i)$ correspond to Q(x) and $\Pi(x)$, respectively. Then, we can define a simple random matrix

(13)
$$Q = \sum_x Q(x)\ I_x$$

where x = N,A,L. Then, from (4) we obtain

$$E(Q) = \sum_x Q(x)\ \Pi(x)$$

(14)

$$= (1-\lambda\Delta t-\mu\Delta t)Q(N) + \lambda\Delta t Q(A) + \mu\Delta t Q(L).$$

Consider some arbitrary ensemble of q(t)'s, and define the expectation $E(q(t)) = p(t)$ over this ensemble to be a probability vector such that $p(t) = (p_i(t))$, $i = 0,1,2,\ldots$, $0 \leq p_i(t) \leq 1$, $\sum p_i(t) = 1$. Define the entry of p(t)Q(N) for queue size n to be $P_n(t)$, of p(t)Q(A) to be $P_{n-1}(t)$, and of p(t)Q(L) to be $P_{n+1}(t)$. Then, the entry of E(p(t)Q(x)) for queue size n is given by the Equation (9) of the birth-and-death process model.

On the other hand, the use of (5) obtains

$$ENR(Q) = \sum_{x} Q(x) \quad II(x)\ I_x$$

$$(15) \qquad = \begin{cases} (1-\lambda\Delta t-\mu\Delta t)Q(N) , & \text{if } x = N, \\ \lambda\Delta t Q(A) , & \text{if } x = A, \\ \mu\Delta t Q(L) , & \text{if } x = L. \end{cases}$$

This definition of ENR(Q) satisfied a one-to-one correspondence with (10) except for the fact II(x)Q(x) is given instead of Q(x) with its associated II(x).

Another queueing model of Poisson arrivals (M) and general service time distribution (G) of a single server (1), or M/G/1 system can also be analyzed using the approach of queueing sequences. It can be shown that the convergence properties defined for the overall ensemble of queueing sequences do not hold for the conditional subensemble of M/G/1 sequences, cf. [5].

REFERENCES

[1] Bayes, T., An essay solving a problem in the doctrine of chances, with Richard Prices' foreward and discussion, Phil. Trans. Royal Soc., 1763, 370-418.

[2] deFinetti, B., La previsione; ses lois logiques, ses sources subjectives, Ann. of Inst. of H. Poincare, 7, 1-68, 1937.

[3] Fishburn, P. C., Preference-based definitions of subjective probability, Ann. Math. Stat., 1605-1716, 1967.

[4] Loeve, M., Probability Theory, Foundations, Random Sequences, Van Nostrand, Princeton, 1955.

[5] Mizuki, M., Analyses of subensembles of M/G/1 queueing sequences, Israel J. of Math., 3 236-247, 1966.

[6] Mizuki, M., Probability model of a non-repetitive conditional event, submitted to a journal.

[7] Mizuki, M., Preference based on distributions of expected utilities of a non-repetitive event, submitted to a journal.

[8] Savage, L. J., The Foundations of Statistics, Wiley, 1954.

APPLICATION OF SIGNAL FLOW GRAPH THEORY TO A STOCHASTIC PROCESS

R. G. Stimson
Office, Chief of Staff, Army
Office, Director, Weapon Systems Analysis
Washington, D. C.

ABSTRACT. A method is presented for calculating the probability of killing a multiple target aircraft formation attacking a missile battery as a function of engagement parameters and missile firing strategy. The stochastic processes engendered by various firing strategies are represented by signal flow graphs, facilitating the calculations. Results are utilized to optimize missile firing strategy. Although developed for analysis of firing strategies, the method can be applied to many analogous problems involving stochastic duels and programming under conditions of uncertainty, where the situation can be resolved into discrete states with transition probabilities dependent on both the state and the path by which it was reached.

INTRODUCTION. When an air defense system using missiles, which home on energy furnished by an illuminating radar and reflected by the targets, attempts to engage a formation of aircraft (or missiles) which are grouped closely enough in position and velocity that they appear as a single target to the homing missiles until the latter are close to the formation, a question arises as to the optimal firing strategy. The choice of a strategy for any particular situation depends on several factors which affect the conditional probability of success at any particular point in the process and which must be accounted for in formulating a generalized framework for assessing various strategies. When a missile engages the formation, it initially homes on the centroid of reflected energy. At some point, the return from a single target will override the centroid, and the missile may have to perform a relatively high-g maneuver in the end game, degrading its kill probability. The effects are worst for the case of two targets, where the energy centroid may move back and forth rapidly, and become less detrimental as the number of targets increases, since the energy centroid tends to remain closer to the center of the formation in this case. Therefore, in analyzing the effectiveness of various missile firing strategies, it is necessary to assign a weighting factor to the single shot kill probability (or SSKP) in accordance with the number of targets in the formation. Since the magnitude of the weighting factor increases as the number of targets increases, it might seem advantageous to fire as many missiles as possible in the first volley. However, as the number of simultaneously fired missiles is increased, the probability of two or more missiles locking on the same target increases, and at some point a further increase becomes unattractive.

In the following method of analysis, the number of attacking aircraft is taken as k, and the stochastic process of shooting them down is represented by a system having k states, the number of each state denoting the number of aircraft which have been killed at that point in the process. The system is depicted by a single flow graph for each firing strategy. The paths leaving each node represent all possible ways to go from each state to succeeding states, each path value being the conditional probability of reaching state n+p via that path given that state n has been reached. In order to illustrate how this technique is used to determine an optimal firing strategy, the number of targets is taken as four, and the following four strategies, in which $Sn_1, n_2, \ldots n_j$, refers to n_1 missiles fired in the first volley, n_2 the second time, etc., and n_j the jth and all succeeding times, are analyzed, using the signal flow diagrams depicted in Figure 1 through 4 in conjunction with Mason's signal flow graph rule to effect the calculations.

MISSILE FIRING STRATEGIES

S1, 1, 1, ...
S2, 2, 2, ...
S3, 3, 3, ...
S3, 2, 1, ...

SIGNAL FLOW GRAPHS FOR UNIFORM STRATEGIES. For strategy S1, 1, 1, ..., the engagement process is represented by Figure 1 in a manner suggested by Hall [1]. The four states are represented by nodes 1 to 4, each state representing the number of planes which have been shot down at that point in the process. Each firing of a missile is a Bernoulli trial with the probability of success, equal to the product of the single shot kill probability and the multiple target weighting factor for that state, determining the value of the path to the next state, and the probability of failure determining the path value of the self-loop to the same state. Path values are multiplied by a dimensionless parameter x. Since the system function or ratio of output to input, from the input to a specific node is a multiplicative function of the node-to-node path values, the exponent of x in the calculated system function, or gain, to that node is equal to the number of missiles fired to reach the state represented by that node via that path. The self-loop in state four is necessary to account for any missiles fired or still in transit after all targets are killed. It is seen that the engagement sequence in this case is a Markov chain with as many states as there are aircraft, each state representing the number of aircraft which have been killed. Although, in this simple case, it is feasible to solve the problem using transition matrices, it will be seen later that this technique will become increasingly tedious for more complicated strategies. For instance, a "non-uniform" strategy, where successive volleys may contain different numbers of

missiles, constitutes a system with memory, in which the conditional probability of transition to the next state depends not only on the present state, but also on how one arrived in it; i.e., how many missiles were fired in reaching the present state. The stochastic process then ceases to be represented by a Markov chain, and the flow graph becomes very useful as an aid both in calculation and in understanding the physical implications of the situation. The actual calculations are carried out using Mason's gain formula [2, 3].

In the signal flow diagram, the value at each node is equal to the sum of the values of all paths leading to that node. Each path value is the product of the value of the node at the beginning of the path and the transfer function associated with that path. Signal flow diagrams find their greatest application in electrical engineering in connection with differential equations, representing control systems, which are first Laplace-transformed, then depicted as flow diagrams, solved using Mason's rule as described below, and then transformed back to the time domain. In the present case, the "input signal" is simply unity probability of reaching state 0; i.e., of shooting down at least zero aircraft. The nodes represent states which are defined by the number of aircraft which have been shot down, state k representing a point in the process at which k aircraft have been shot down. The "transfer functions" are simply the conditional probability of reaching a certain state, or number of aircraft downed, given that a certain other state had been reached previously. In order to represent these probabilities as functions of the number of missiles being shot, the conditional probabilities are multiplied by x^n, where n is the number of missiles which are shot in each volley when attempting to go from a node to a succeeding node. As will be seen below, when Mason's rule is used to find the output signal, given the input signal and the signal flow diagram, the path values between successive nodes are multiplied. Therefore, the highest exponent of x in the system function, or ratio of output, represents the total number of missiles fired to reach the final node, or number of aircraft downed, since it was arrived at by traversing a series path from node to node, with the path values multiplying and therefore with the exponents of x in each path adding. Thus, considering Figure 1, it is obvious that the probability of shooting down four planes; i.e., of reaching Node IV, by firing only four missiles is

$$P(IV, 4) = \prod_{i=0}^{3} P_i$$

This is true since, in order to down four aircraft with four missiles, one must traverse the paths representing the conditional probability of reaching the next state (getting a hit) directly from Node 0 to Node IV without traversing any self-loops, which represent the conditional probability of remaining in the same state (getting a miss). It is seen that by

multiplying each P_i by x, the system function, or gain, for reaching node IV by firing only four missiles is

$$G(IV, 4) = \prod_{i=0}^{3} P_i \; x = x^4 \prod_{i=0}^{3} P_i$$

The exponent of x is seen to represent the number of missiles fired. If one was to miss with the jth shot, however, it would take five missiles to shoot down the four aircraft, and the probability would be

$$P(IV, 5, \text{miss jth shot}) = q_j \prod_{\substack{i=0 \\ i \neq j}}^{3} P_i$$

In this case the system function arrived at by multiplying each P_i by x, would be

$$G(IV, 5, \text{miss jth shot}) = q_j \; x \prod_{\substack{i=0 \\ i \neq j}}^{3} P_i \; x = x^5 \; q_j \prod_{\substack{i=0 \\ i \neq j}}^{3} P_i$$

Of course, the total probability of shooting down four aircraft by firing five missiles is the sum of five such probabilities, arrived at by considering a miss on the jth shot and letting j range from 0 to 5. However, the system gain will still contain an x^5 term. As will be seen below, use of Mason's rile in conjunction with a particular diagram will produce a polynomial in x in which the coefficient of x in each term will indicate the probability of shooting down all the aircraft, using the strategy associated with that diagram, by firing the number of missiles indicated by the exponent of x in that term. The calculations may be carried out to any desired power of x (number of missiles fired) and the probability of shooting down the aircraft approaches unity as the number of missiles is increased without limit. If it were desired to find the probability of reaching a lesser state, say, state k (k aircraft downed), then the signal flow graph could be used by omitting all paths which lead to higher nodes than Node k.

Mason's signal flow graph gain formula is a technique for utilizing a signal flow graph to obtain the gain of the system instead of directly solving the equations describing the system. It makes use of the gains, or transfer functions, associated with forward paths and loops, the gain of a forward path being the product of the gains of each segment of the path, where each segment leads from one node to another. A loop is simply a forward path which closes on itself. The formula is:

$$G = \sum_k \frac{G_k \Delta_k}{\Delta}$$

where

G = system gain, or ratio of output to input

G_k = gain of the kth forward path

Δ = system determinant

= 1 - (sum of all individual loop gains)

+ (sum of products of gains of all possible combinations of two non-touching loops) - (sum of products of gains of all possible combinations of three non-touching loops)

+ - ...

Δ_k = value of Δ for that part of graph not touching the kth forward path

While the mechanics of the formula are simple, they frequently are tedious for a signal flow diagram which has many loops and forward paths. Fortunately, this type of repetitive calculation is easily carried out with a digital computer; one only needs to identify the individual forward paths and their respective non-touching loops on a particular graph in order to be able to use a standard program.

For the strategy of firing successive volleys of two missiles, S2, 2, 2, ..., depicted by Figure 2, it is seen that several results may ensue from the firing of a volley. First, one may score two hits, not on the same target, and will, therefore, go from state n to state n+2. Secondly, one may score only one hit, and will, therefore, reach state n+1. Thirdly, one may score two hits, both on the same target, and will, therefore, reach state n+1 by a different path. Fourthly, one may score no hits and remain in state n. In order to assign the correct value to each path, it is necessary to know the probability of ℓ missiles homing on the same target when each of m missiles homes on one of n targets. This will be

$$b_{\ell, m, n} = n \binom{m}{\ell} \left(\frac{1}{n}\right)^{\ell} \left(1 - \frac{1}{n}\right)^{m-\ell}$$

The effects of these probabilities are seen in the signal flow graph. The technique of forming the graph is straightforward; all possible transitions from one state to the next are given a path, which is assigned the appropriate probability and multiplied by x^2, since two

missiles are being fired. The binomial coefficients are also necessary, since the paths result from Bernoulli trials and follow the binomial probability law. The values of all paths leaving each state node will, of course, sum to unity; if they do not, a mistake has been made. It is seen that the process is still a Markov chain, since two missiles will be fired with the same probability of success when the system is in a particular state, regardless of how that state was reached. The extension to strategy S3, 3, 3, ..., is straightforward and results in the flow graph in Figure 3.

FLOW GRAPHS FOR NON-UNIFORM STRATEGIES. When one uses a non-uniform strategy such as S3, 2, 1, 1, ..., the system becomes somewhat more complicated, as seen in Figure 4. The number of missiles to be fired when the system is in a particular state now depends on how many were fired in reaching it. Therefore, the transition probabilities are dependent not only on the present state, but also on how that state was reached. The system has now developed a memory and can no longer be represented by a Markov chain. Fortunately, the flow graph remains quite simple even for this type of stochastic process. Node N, representing the state in which n targets have been killed, is merely split into m nodes, where each node is reached from a prior node either by firing a volley composed of a different total number of missiles, m being the number of different total numbers of missiles, or by proceeding from a different node such that the same point in the firing strategy is reached. Consider a "keyed" firing strategy S3, 2, 1, 1, 1, ..., where the transition between volley sizes; e.g., between the volley of three missiles and the volley of two, is not made until there has been a change of state; i.e., until at least one plane has been shot down as a result of firing missiles in volleys of three. This information is normally available from a continuous wave illuminating radar, since a falling tone indicates that one or more (but not how many) planes has been killed. The keyed strategy does, of course, require the operator to wait until the present volley of missiles has reached the target area before firing the next volley. For the case of four targets and firing strategy S3, 2, 1, 1, 1, ..., this necessitates two nodes for state two and two nodes for state three, as can be seen from the diagram. Although self-loops and forward paths are thereby added to the flow diagram, the calculations do not become conceptually more complicated, but merely more voluminous. Since an electronic computer would ordinarily be used to evaluate system gain, using Mason's rule, for situations involving a large number of aircraft or a complicated firing strategy, this is not a serious drawback. Indeed, the chief advantage of the method is that the complexity of the calculations does not increase in proportion to the number of states in the system and the complexity of the strategy. If instead of a keyed non-uniform strategy, one uses a "pure" strategy, in which the transition between volley sizes is independent of changes in state, it is necessary to provide additional split nodes to accommodate the paths representing misses by all missiles in a volley. This type of

path will no longer be a self-loop to the same node, but will lead to a separate node representing the same state but requiring that a different number of missiles be fired. For example, for the strategy S3, 2, 1, 1, 1, ..., the self-loop to node 0 would now lead to subsequent nodes by the strategy S2, 1, 1, 1, ..., finally reaching node IV. Similarly, the particular path from node OB which represents two misses would not be a self-loop, but would lead to node OC, which would then lead to subsequent nodes by the S1, 1, 1, ... strategy depicted in Figure 1. The paths representing misses by all missiles at other nodes would be treated in the same manner. This case is not worked out here since it adds nothing to the explanation of the technique, merely representing a straightforward extension of the diagram with no difference in the manner of solution except that it requires more steps in the computer program.

The above four strategies were analyzed for a formation of four attacking aircraft. The SSKP was taken as .75, and the multiple target weighting factors were taken as 0.9, 0.8, 0.5, and 1.0 for states 0, 1, 2, and 3, respectively, in accordance with the fact, explained above, that the multiple target effects become less pronounced as the number of targets increases. Therefore, the resulting kill probabilities, p_n, were .675, 600, .375, .750, and 0 for states 0, 1, 2, 3, and 4, respectively, being 0 in state 4 since there are no remaining aircraft at this point. The diagrams were used to calculate, for each firing strategy, the probability of killing all four targets as a function of the number of missiles fired. As an example, the flow diagram for S3, 2, 1, 1, 1, ..., after assigning path values and combining parallel paths, is shown in Figure 5. The system gain can now be found by node absorption, Mason's rule, or a combination thereof. For example, the quotient of polynomials obtained for S3, 3, 3, 111, was:

$$G = \sum_{i=1}^{n} \frac{G_i \Delta_i}{\Delta} = \frac{0.224x^6 + 0.418x^9 + 0.052x^{12}}{1 - 1.360x^3 + 0.391x^6 - 0.034x^9 + 0.001x^{12}}$$

$$= .224x^6 + .722x^9 + .946x^{12} + .963x^5 + .964x^{18}.$$

Thus, the probability of killing all four targets with, for example, twelve missiles fired three at a time was .946. The results of the calculations are shown in Figure 6. It is seen that S1, 1, 1, ..., provides a higher probability of killing all four targets than do the other strategies when the number of missiles to be fired is seven or less. However, the probability then levels off rather sharply, and a great many missiles would be necessary in order to exceed a probability of .8. The curves for S2, 2, 2, ..., and S3, 3, 3, ..., have the same

general shape as that for S1, 1, 1, ..., except that they tend to level off at a higher range of values. However, it would still be necessary to fire a large number of missiles in order to attain a probability in excess of .9. The curve for S3, 2, 1, 1, 1, ... tends to level off at a high range of values, and it has the advantage of rising more quickly to this range. The reason for this is fairly obvious, since this strategy calls for firing a large volley at first and then smaller volleys, taking advantage of the fact that the energy centroid of the targets tends to remain more in the center of the formation if the number of targets is large and that the probability of more than one missile locking on the same target is lower for a large number of targets. Although these effects are intuitively clear, the exact manner in which they interact is not, and it is apparent that further analysis along the lines suggested by Figure 6 would lead, by a kind of dynamic programming process, to the optimal firing strategy for any given situation if one is trying to maximize the probability of killing all four targets by firing a certain number of missiles. If one is interested in the probability of killing some specific number of the attackers instead of all of them, as a function of firing strategy and number of missiles fired, it is necessary only to delete all flow graph nodes representing a number of kills greater than this.

If one is attempting to optimize some other aspect of the situation, the information is generally available from Figure 6. For instance, the expected kills per missile are plotted for each strategy in Figure 7. It is seen that strategy S3, 2, 1, 1, 1, ... provides a higher number of expected kills per missile than the others if five or more missiles are fired. In order to obtain the true mathematical expectation, of course, one would also need the probability of killing three, two, and one of the attacking aircraft, which would necessitate calculations of the system functions to nodes I, II, and III. The main contribution, however, is provided by the probability of killing all four of the aircraft, and the true expectations, although somewhat higher than the ones in Figure 7, would not differ from them qualitatively, and one would not ordinarily require a refinement of this nature until it was apparent that the optimal strategy had been approached.

Although the above technique, utilizing representation of transition probabilities by signal flow graphs and subsequent application of Mason's rule to calculate system functions which indicate the effectiveness of the relevant strategies, was used in conjunction with missile firing strategies in this case, it is readily seen that it is applicable to a variety of problems arising in military operations research and in other situations involving stochastic duels and programming under conditions of uncertainty. It also provides a facile method for analyzing, by means of an electronic computer, the effects of a change in strategy (or programming) or of engagement parameters or program elements and therefore is amenable to gaming.

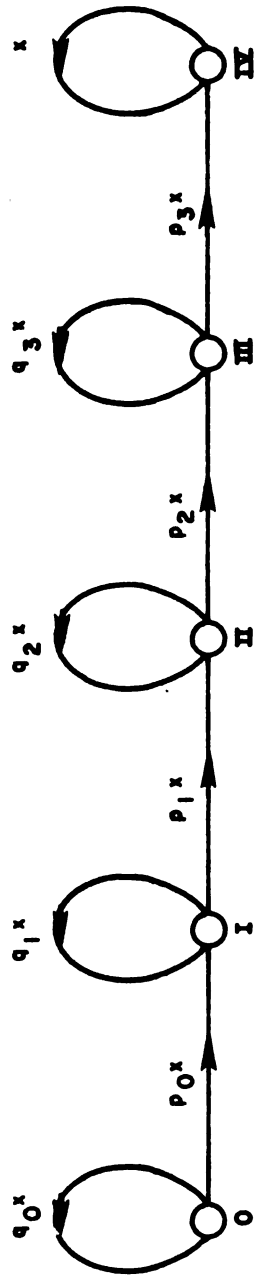


FIG. 1. FLOW DIAGRAM REPRESENTING STRATEGY $S_j = I, I, I, \cdots$ EMPLOYED AGAINST FOUR TARGETS.

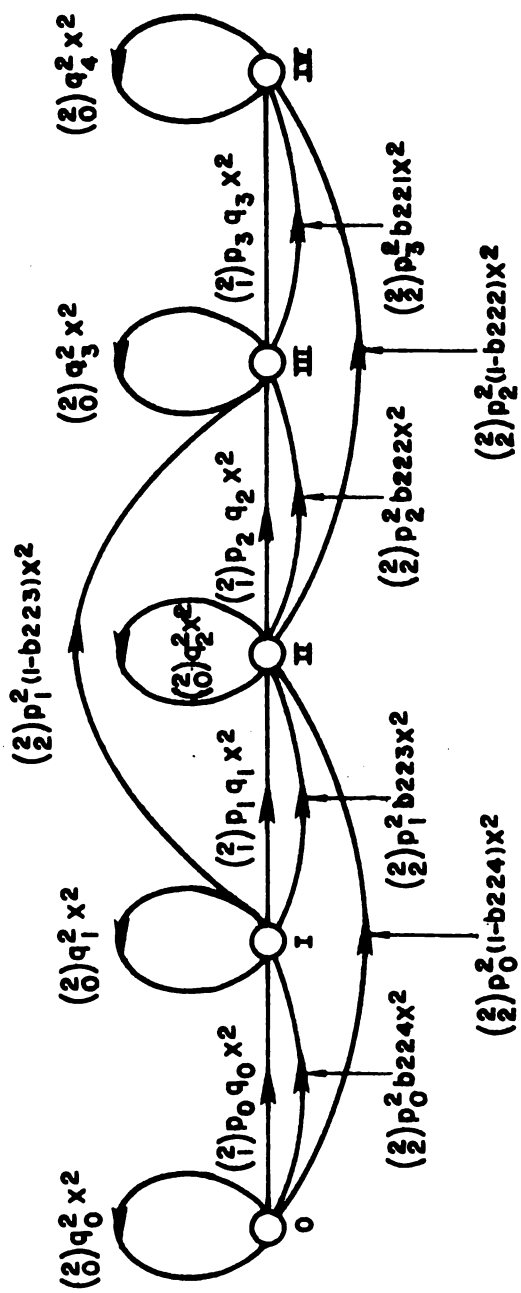


FIG. 2. FLOW DIAGRAM REPRESENTING STRATEGY $S_2 = 2,2,2,\cdots$ EMPLOYED AGAINST FOUR TARGETS.

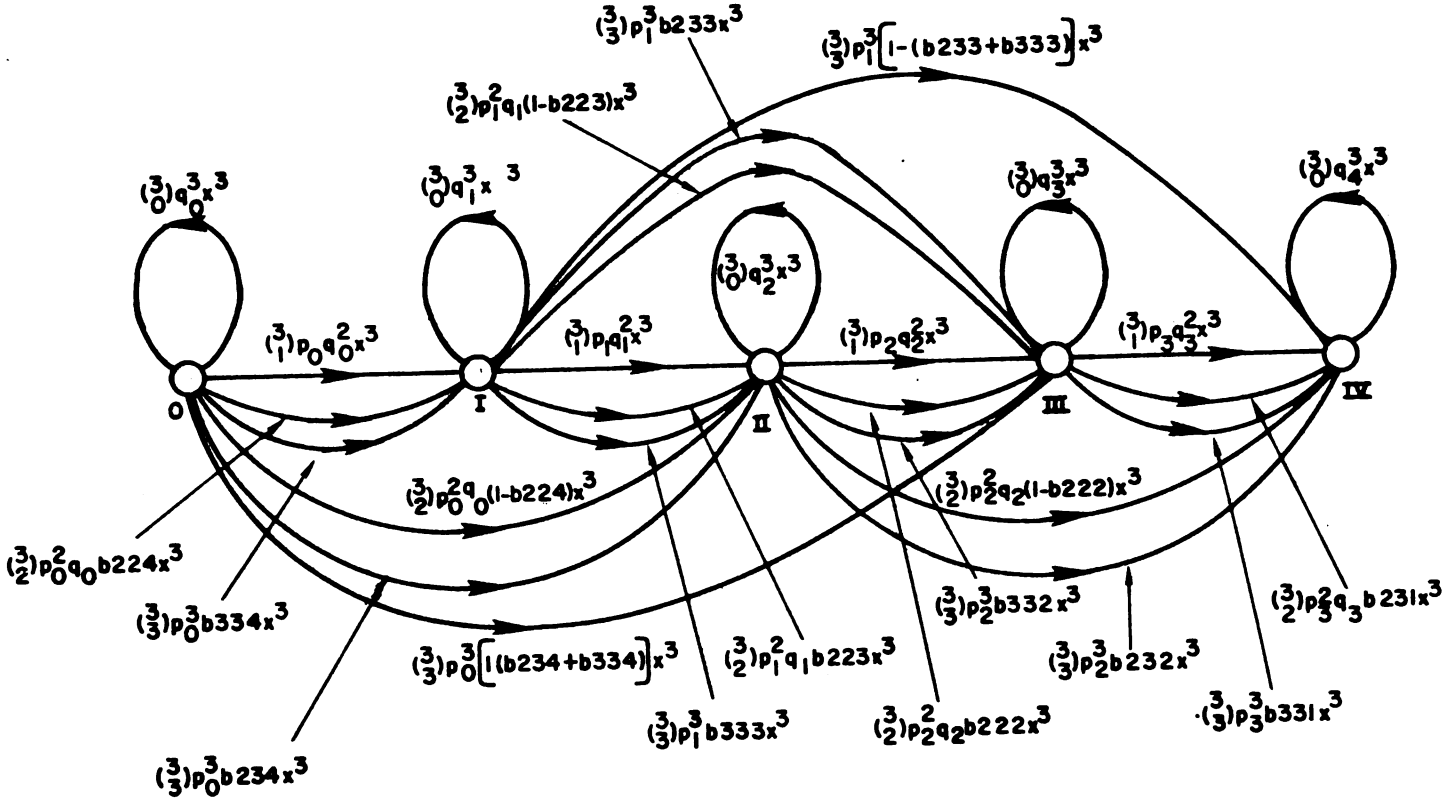


FIG. 3. FLOW DIAGRAM REPRESENTING STRATEGY $S_3 = 3,3,3,\cdots$ EMPLOYED AGAINST FOUR TARGETS.

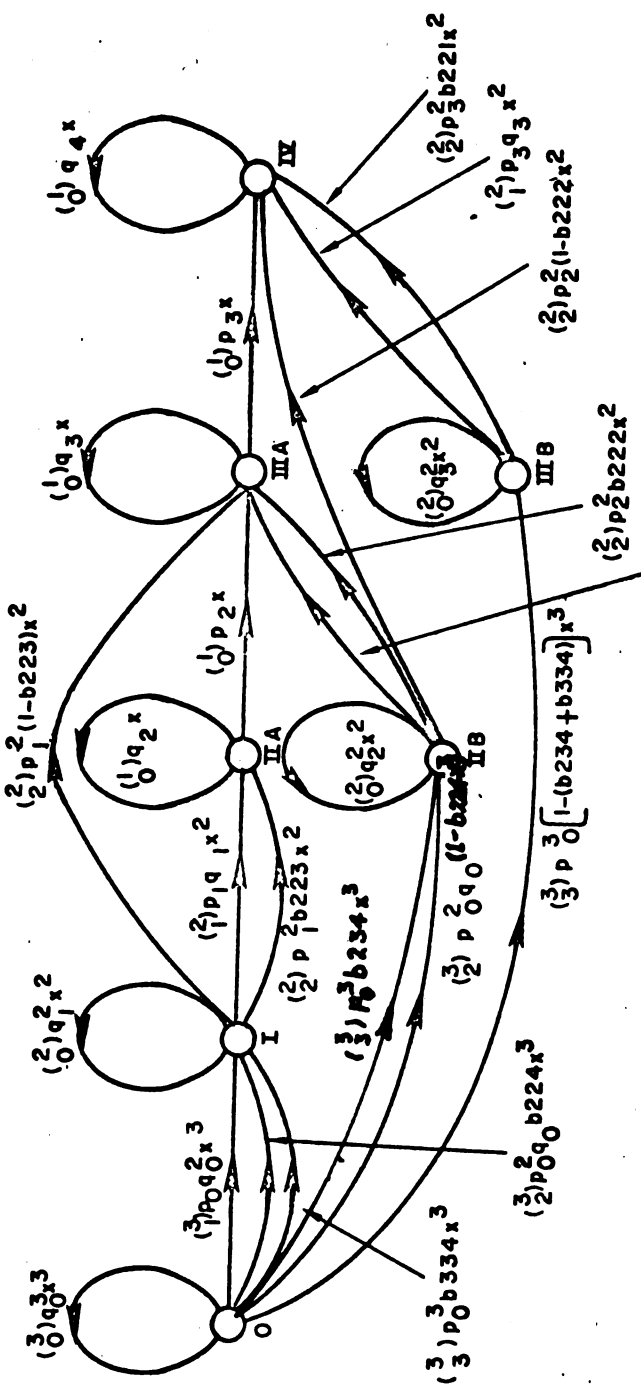


FIG. 4. FLOW DIAGRAM REPRESENTING STRATEGY $S_4 = 3,2,1,\cdots$ EMPLOYED AGAINST FOUR TARGETS.

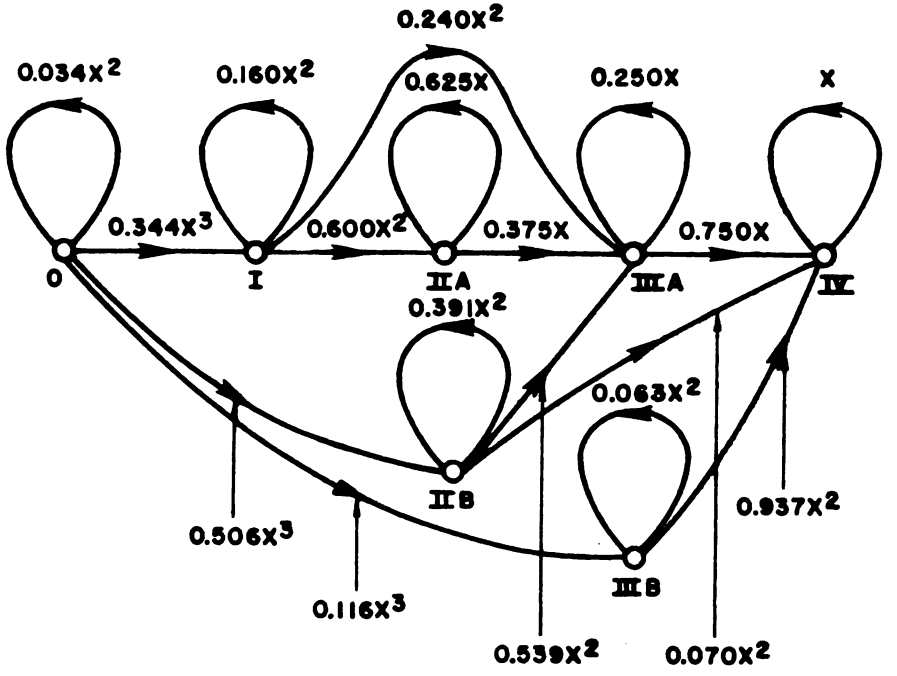


FIG. 5. DIAGRAM OF FIG. 4 AFTER USING VALUES FOR SSKP AND WEIGHTING FACTOR, AND FORMULA FOR PROBABILITY OF MORE THAN ONE MISSILE HOMING ON SAME TARGET, GIVEN IN TEXT. PARALLEL PATHS HAVE BEEN COMBINED.

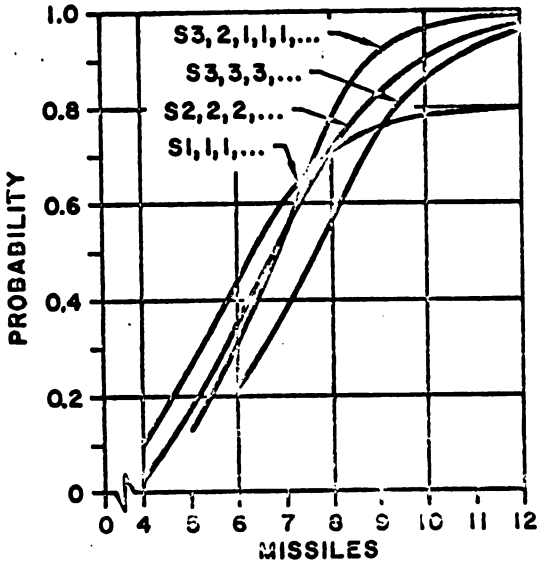


FIG 6. PROBABILITY OF KILLING ALL FOUR TARGETS AS A FUNCTION OF THE NUMBER OF MISSILES FIRED USING VARIOUS ENGAGEMENT STRATEGIES

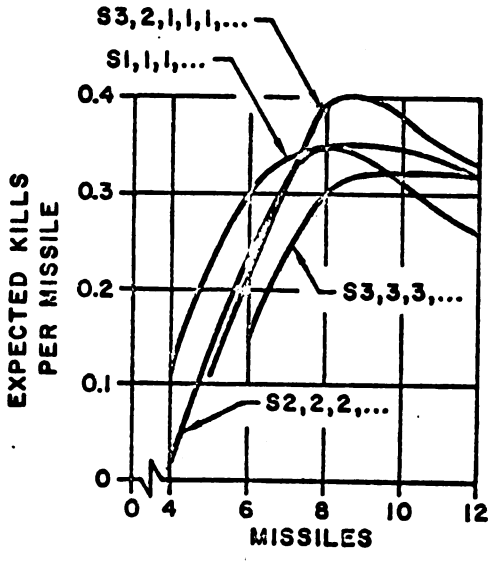


FIG. 7. COMPARISON OF EXPECTED KILLS PER MISSILE FOR VARIOUS STRATEGIES. PROVIDES HEURISTIC INDICATION OF MORE OPTIMAL STRATEGIES, ENABLING ITERATIVE OPTIMIZING PROCEDURE.

THEORY AND ASSUMPTIONS UNDERLYING THE DEVELOPMENT OF CSP-R*

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

1.0 INTRODUCTION

This memorandum discusses the development of CSP-R, a continuous sampling procedure involving normal, tightened, and reduced sampling inspection. The memorandum discusses some of the considerations that led to its development and the objectives set for the procedure during development. It also provides the necessary mathematical derivations used in the development. CSP-R plans will appear in MIL-STD-1235A, "Continous Sampling Procedures and Tables for Inspection by Attributes."

2.0 BACKGROUND

2.1 Reduction in Sampling Inspection

When confidence has been established that a manufacturing process is stable and is producing a small percentage of defective material, the user of continuous sampling plans often has the desire to reduce the amount of sampling inspection being done.

2.2 CSP-M

MIL-STD-1235 contains a multi-level sampling plan, CSP-M, which allows such reduction in sampling inspection. In spite of this feature, a survey of Army Ammunition Plant inspection elements indicated that CSP-M was considered too complicated in terms of its administration to be useful. For this reason, the CSP-M plans were generally ignored.

From a technical point of view, CSP-M contains another weakness; it is not very responsive to a deterioration in quality if one of the reduced sampling states has been reached. As an example, suppose that we are inspecting at sampling rate level number five, AQL = .25%, i = 287. Suppose that a previously low process average shifted to 1%, or four times the AQL. The probability of continuing[1] on one hundred percent inspection after finding a defect is only .00000000946. In fact, there is only an 80% probability that the 100% inspection

[1]-that is, going progressively through the checking states to the 100% inspection level.

---

*This article has previously appeared as Technical Memorandum QEM 21-230-6. The remainder of this paper has been reproduced photographically from the author's copy.

level will be reached without first reaching a certain level R or star state (say, level 3) and then reverting to a lower level (level 4).

2.3 CSP-1

The simplest continuous sampling plan is, of course, CSP-1 wherein the finding of i consecutive defect free units on 100% inspection allows sampling inspection to begin, during which the finding of a defect causes a reversion to 100% or screening inspection. CSP-1, however, does not allow a decrease in sampling inspection. Using a CSP-1 plan with the same AOQL but with a smaller sampling frequency may be a solution, but indiscriminate shifting between plans without specified rules based upon the mathematical impact of such shifting is, of course, not desirable.

2.4 CSP-2

CSP-2, while not allowing a reduction in the sampling frequency, does delay the resumption of screening inspection under certain circumstances. This feature is desirable in those situations where an alert of the screening crew seems necessary, but it offers no special advantages insofar as allowing a reduction in sampling inspection.

2.5 MIL-STD-105D

MIL-STD-105D allows a reduction in lot-by-lot sampling via the reduced sampling technique. A history of good product quality allows a reduction in sample sizes for subsequent inspections. At the same time, a history of marginal product quality causes a tightened inspection to be initiated. This tightened inspection sometimes requires a larger sample size, but in all cases the probability of accepting a lot with a given percent defective[2] is lower under tightened sampling inspection.

3.0 OBJECTIVES

Consideration of the points mentioned above led to some general ideas about what kinds of characteristics a continuous sampling procedure should have, if this continuous sampling procedure were to allow a reduction in sampling inspection after demonstration of a low process average.

3.1 Responsiveness

The procedure should be responsive to an undesirable shift in the process average. This feature could be obtained by requiring a screening sequence after finding a defect on a sampling sequence.

---

[2]-other than 0% or 100% defective.

3.2 Simplicity

The procedure should be both simple in design and relatively easy to administer. Although simplicity is a somewhat subjective concept, it would seem that, generally speaking, the fewer inspection states a procedure has, the simpler the procedure would be. Likewise, a procedure with simple rules for switching between sampling and screening states[3] is simpler than one which requires check states or similar devices. It was felt, therefore, that a procedure with a relatively few number of states, with the switching rules similar to those of CSP-1, would satisfy the objective of simplicity.

3.3 Average Outgoing Quality Limit

The development of the procedure should be based on the concept of an average outgoing quality limit (AOQL), not only to provide a limit to average outgoing quality which will not be exceeded no matter what quality of product is submitted for inspection, but also to establish correspondence with CSP-1 plans and other continuous sampling plans from which a user can make a choice.

3.4 Relationship with CSP-1

Common sense dictated that the procedure require less inspection than some norm for product of high quality and more inspection for product of marginal quality. Accordingly, it appeared reasonable that the first step of the development would be establishment of a norm. CSP-1 was selected as this norm because it is the most widely used of existing CSP's by Army Ammunition Plants inspection elements.

The attainment of this objective could be demonstrated by a comparison of Average Fraction Inspected (AFI) curves for the developed plans with AFI curves for corresponding CSP-1 plans[4]. An AFI curve shows the percentage of units inspected over the long run when the process average is of a certain value.

3.5 Relationship with the Normal-Tightened-Reduced Concept of MIL-STD-105

Purely as a matter of standardization, it was decided to develop the procedure along the lines of the normal-tightened-reduced concept of MIL-STD-105D. Users of MIL-STD-105D could adapt easily, therefore, should they have occasion to use this procedure in MIL-STD-1235A.

[3]CSP-1, for example, is the epitome of simplicity in this regard.
[4]A graphical illustration of this comparison is given in [7.1].

4.0 THE DEVELOPMENT

With the objectives above in mind, development of the procedure began. Several models were formulated and weighed against the objectives stated. Actually, most of the objectives could be satisfied simply by designing them into the procedure. The steps used to evaluate each model in terms of its statistical properties are discussed below.

4.1 Determining the Parameters

After a general procedure was defined, which would satisfy, by its construction, most of the objectives, it became necessary to investigate the procedure's relationship with CSP-1. In order to satisfy the objective concerned with this relationship, representative examples of CSP-1 plans were selected. The AOQL's for these plans were used in determining the parameters (sampling frequencies and clearance numbers for the plans based upon the procedure under investigation. Accordingly, the AOQ formula for each procedure had to be developed[5] and the parameters subjected to variation until the maximum resulting AOQ for any value of the process average, p, was close to the target AOQL. In general, the sampling frequencies were held fixed and the clearance numbers were allowed to vary. As can be seen from a study of Appendices A and B, this was no small task.

4.2 Computing the AFI Curves

Upon the determination of the parameters of the plan, the AFI formula developed prior to developing the AOQ formula[6] was used to find several points of the AFI curve for the plan. The AFI curves were then drawn on graph paper.

4.3 Comparing AFI Curves

After determining the AFI curve for the plan under test, the AFI curve for the corresponding[7] CSP-1 plan was drawn on the same sheet of graph paper, and the results were compared.

As discussed in 3.4 above, it was desired that a plan based on the developed procedure require less inspection than a corresponding CSP-1 plan for product of good quality and more inspection than CSP-1 for product of marginal quality. Expressing this mathematically, we want

AFI (of CSP-1) > AFI (of developed plan) for $p \leq p_o$, and

AFI (of CSP-1) < AFI (of developed plan) for $p > p_o$, where p_o

would be the "dividing line" of good and marginal quality. It was desired to

[5] See Appendices A and B for the work involved in deriving the AOQ formula for the selected procedure.

[6] See Appendix B for the AFI formula of the selected procedure.

[7] The method of establishing the correspondence was defined for each procedure but in each case depended on the AOQL.

keep p_o within the interval $(0, p_L)$, where p_L is the value of the process average for which the AOQ is equal to the AOQL. This choice, though arbitrary, seemed reasonable.

4.4 Selection of Procedure

A procedure was finally selected which most satisfactorily fulfilled the objectives. This procedure was designated CSP-R, and is described in block diagram form in Figure I. This procedure, while generally satisfying all of the objectives, does not strictly satisfy the objective relative to the AFI curves when the clearance number is very small and at the same time the sampling frequency is very large[8]. Since plans with these parameters are not used extensively, this limitation did not seem restrictive.

5.0 THE PROCEDURE

Although Figure I seems very self-explanatory, discussion of some of the features of CSP-R seems in order.

There are three sampling states: normal, tightened, and reduced, and three screening states: qualification, retrial, and tightened. It can be seen that the three sampling states are parallel to the normal-tightened-reduced concept of MIL-STD-105D, and this is, in fact, why they are labelled as such. The rationale for the three screening states can be found in the discussion below.

5.1 How the Procedure Operates

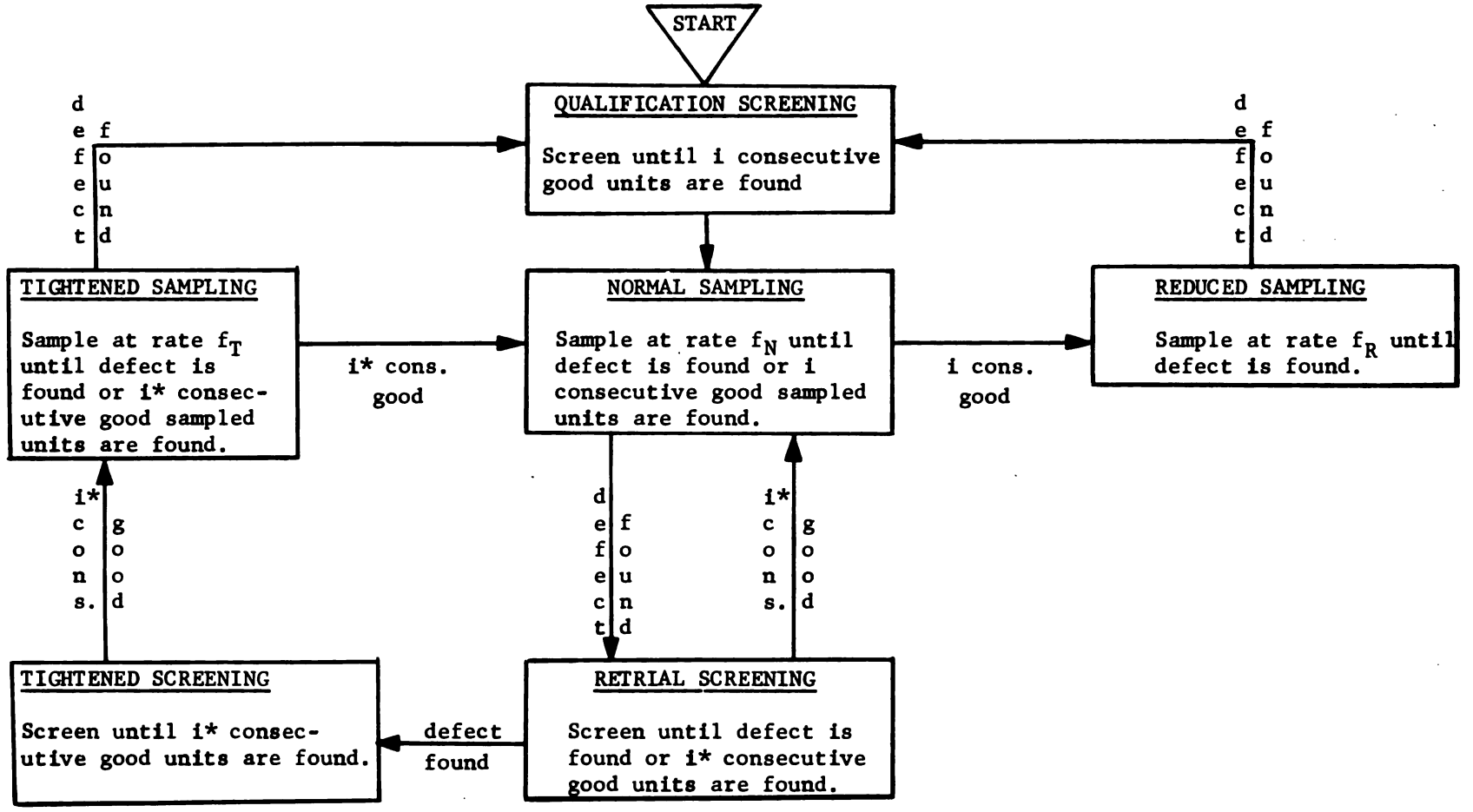
Entrance into the inspection states was designed to be dependent upon the demonstrated capability of the production process, as evidenced by favorable or unfavorable inspection results. Under the system, the qualification state is initially entered. When evidence indicates the quality of an item has stabilized at a satisfactory level, normal sampling is initiated. Continued evidence of the process's capability to produce satisfactory or better quality permits the reduced sampling state to be entered. Once reduced sampling is initiated, sampling remains in effect until a defect is found, at which time the system immediately invokes its qualification screening provisions.

The tightened inspection phase of the system was also designed to be entered from the normal inspection phase. However, tightened inspection provisions are invoked only when defect(ive)s fall too closely together; that is, when the separation of defect(ive)s is less than a prescribed minimum spacing. Tightened screening remains in effect until sufficient evidence indicates the process is capable of generating an item of at least marginal quality. Once this evidence is established, tightened sampling is initiated. The normal sampling state may then be re-entered if evidence of favorable inspection

[8]-those plans in MIL-STD-1235 associated with large AQL's and the lower code letters.

FIGURE I

CSP-R FLOW DIAGRAM



continues. If not, the system invokes its qualification screening provisions and continues as before.

Similarity of provisions governing transitions between states in CSP-R and those associated with the MIL-STD-105D scheme is apparent. However, under the MIL-STD-105D scheme there is a transition from reduced to normal sampling not only upon an unfavorable inspection result (rejection of a lot), but also upon acceptance under the procedures of 10.1.4 of that Standard[9]. We therefore see that the reduced state is entered with difficulty, but left immediately should doubt arise as to the continued high quality of material. The analogous CSP-R provision is the requirement of re-entrance into the qualification screening state. This provision, though admittedly drastic, was established to assure performance of sufficient screening to guarantee that the previously good quality level had not deteriorated.

The retrial screening provision of CSP-R was designed to represent a reasonable balance between: (1) the need for assurance of the previously established quality level for normal sampling and (2) a desire to avoid a premature decision to invoke the tightened provisions.

5.2 Properties of the Parameters

In common with most CSP plans, those of CSP-R were developed to be based on AOQL and defined by the parameters f_j and i_k, where f_j is the sampling frequency in the jth sampling state and i_k is the clearance number in the kth screening state. Also, in common with most CSP plans, the parameters f_j and i_k of CSP-R plans determine the AOQ function, as discussed previously.

To maintain the normal, tightened and reduced inspection concept, the following relationship among sampling frequencies was used: $f_T > f_N > f_R$; where the subscripts T, N, and R refer to tightened, normal, and reduced sampling, respectively. Since CSP-1 had been established as the norm, it was decided to equate f_N of CSP-R to f of CSP-1 for equal AOQL and production interval size. Consequently, sampling rates f_T, f_N and f_R in CSP-R could be the frequencies for any three consecutive code letters under CSP-1 for a given AOQL. This was conducive to simplicity.

Two values of i_k were established for the procedure: i and i*. The relationship between i and i* is i* = i/2 (with a few exceptions). The choice of this relationship between i and i* was predicated upon the need for more stringent requirements for entering reduced sampling than for entering tightened inspection. It had been noted that the MIL-STD-105D scheme generally requires ten consecutively accepted lots (plus the defects in these ten lots

[9]-that is, when there is not strong evidence that quality is superior.

being less than a prescribed minimum number) under normal sampling to qualify for reduced sampling, but only five consecutively accepted lots on tightened to re-enter normal. Hence, the relationship between i and i* followed by analogy. In addition, it was noted that the MIL-STD-105D scheme invokes tightened inspection provisions if any two of five (analogous to i*) consecutive lots are rejected on normal. CSP-R was designed to require tightened inspection when defect(ive)s are separated by fewer than i* units; one defect(ive) being permitted in normal sampling but not another in re-trial screening.

6.0 DERIVATION OF FORMULAE

As mentioned previously, the development of CSP-R required, upon setting up a hypothetical procedure, the determination of the mathematical properties of the procedure, so that appropriate comparisons could be made.

6.1 The Flow Diagram

The first step in constructing the appropriate mathematical model would be to outline the procedure in flow diagram form. Figure I is the flow diagram of CSP-R.

6.2 Events Causing a New State/Phase to be Entered

The next step is to look over each of the blocks in the flow diagram and determine the events causing a state and/or phase to be entered. As used herein, "state" refers to either qualification, retrial, or tightened screening, or normal, tightened or reduced sampling. "Phase" refers to either the units inspected during a sampling state, or the units skipped during a sampling state.

Figure II shows the events laid out in matrix form. The following notation has been used in Figure II:

O = Qualification state

N = Normal sampling state

T = Tightened sampling state

R = Reduced sampling state

N* = Retrial screening state

T* = Tightened screening state

The subscripts I and S in Figure II pertain to phases of sampling states. I denotes the phase when a unit is being inspected, and S denotes the phase when the units are being skipped.

FIGURE II

EVENTS CAUSING A STATE/PHASE TO BE ENTERED/REENTERED

FROM \ TO	0	N_S	N_I	R_S	R_I	T_S	T_I	N*	T*
0	on a defect	on i cons. good units	-	-	-	-	-	-	-
N_S	-	-	after $(1/f_N)-1$ units	-	-	-	-	-	-
N_I	-	on a good unit	-	on i cons. good units	-	-	-	on a defect	-
R_S	-	-	-	-	after $(1/f_R)-1$ units	-	-	-	-
R_I	on a defect	-	-	on a good unit	-	-	-	-	-
T_S	-	-	-	-	-	-	after $(1/f_T)-1$ units	-	-
T_I	on a defect	on i* cons. good units	-	-	-	on a good unit	-	-	-
N*	-	on i* cons. good units	-	-	-	-	-	-	on a defect
T*	-	-	-	-	-	on i* cons. good units	-	-	on a defect

6.3 State Probabilities

The next step is to develop formulae for determining the percentage of units, over the long run, which will reach the point of inspection during each of the states. The development of these formulae for CSP-R is shown in Appendix A.

6.4 The AFI

Next, the AFI formula must be developed. This development is shown in Appendix B for CSP-R. The resultant formula is

$$AFI = P_0 + P_{N*} + P_{N_I} + P_{T*} + P_{T_I} + P_{R_I}$$

where P_j is the state probability of state j, where the subscripts are defined as in 6.2 above.

6.5 The AOQ

Upon determining the expression for the AFI, the AOQ formula can be constructed rather simply. This is shown in Appendix B. The resultant formula is

$$AOQ = \frac{p[1 - AFI]}{1 - p(AFI)} \quad ;$$

where p is the probability of a defective unit.

6.6 Determining the Parameters

Using a certain value of AOQL and establishing values for the sampling frequencies, the AOQ formula, through an iterative process, was used to develop the values of i and i* for the CSP-R plans which will appear in MIL-STD-1235A. It should be pointed out here that the AOQL's used in MIL-STD-1235A are generally less than the corresponding values in MIL-STD-1235. This is because the AOQL's in MIL-STD-1235A have been matched (with certain limitations) to the AOQL's of the MIL-STD-105D single sampling schemes (with the same AQL), treating the scheme as encompassing normal, tightened, and reduced inspection. The effect of tightened inspection caused the resultant AOQL's to be lower.

6.7 Computing the Curve Points

During development, the curve points (AFI and AOQ) were computed for certain representative plans. Upon selection of the CSP-R procedure, curves for each of the plans were computed on the Agency's RCA 501 digital computer. Additionally, Operating Characteristic (OC) Curves were computed. The derivation of the formula for the OC Curves appears in Appendix C. These curves, should they appear in MIL-STD-1235A, will show the percentage of units accepted on a sampling basis, for each value of the process average, p.

6.8 Assumptions Used in the Derivations

Throughout this discussion, we will assume Figure I defines an ergodic Markov process. Thus, after many steps have occurred in the system the probability of being in any given state of CSP-R tends to become a steady state probability which is independent of the number of steps but dependent upon the state in which the system was at the last step and upon the transitional probabilities.

We will further assume:

(1) All items are classified correctly, i.e. defect(ive) or non-defect(ive);

(2) The production process is in statistical control;

(3) When sampling is in effect, every $1/f_j$th unit is inspected with screening required to begin with the next unit after a defective is observed (see below); and

(4) Defective units found are removed but not replaced by non-defectives.

We will digress here to briefly discuss the effect of these assumptions.

The assumptions above have been adopted largely because they lead to the simplest mathematics. However, the use of these assumptions does not imply that CSP-R plans are invalid if conditions other than those assumed apply. What their use does imply is simply that the plans have been designed with these conditions in mind. Deviations from the stated conditions will, in general, affect the AFI function and result in values of AOQL higher than the theoretical values computed from formulae derived herein. Although the modifications of the theoretical AOQL values resulting from such deviations have not been thoroughly explored, some treatment of alternatives has been made [7.11], [7.12], [7.13], [7.14], [7.15], [7.16].

Assumption (3) above has been adopted solely for mathematical convenience. It is recognized that the theoretically best method of sampling would be probabilistic, i.e., each unit would be inspected with probability f_j, independent of other units. However, strict adherence to this method in an actual production situation would be impractical, if not impossible. In some instances, block (or group) sampling may be required; in others, probabilistic or the assumed systematic sampling method may be in order. Thus, MIL-STD-1235 provides for the selection of sample units "so as to give each unit of product an equal chance of being inspected" with the inspector allowing the interval between sample units to vary somewhat rather than drawing "sample units according to a rigid pattern." The effect of assumption (3) is to provide AOQL values of the same magnitude as those computed under the assumption of probability sampling.

7.0 BIBLIOGRAPHY

7.1 Kelley, H. W., and Abraham, F. L., "CSP-R: A Continuous Sampling Plan with Provisions for Normal, Tightened, and Reduced Inspection," QEM 21-230-5, Ammunition Procurement and Supply Agency, Sep 1967.

7.2 Koopman, Linda L., "Three CSP Models with Three Levels of Inspection Severity,: QEM 21-230-1, Ammunition Procurement and Supply Agency, 17 March 1965.

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

7.4 MIL-STD-105D, Military Standard, Sampling Procedures and Tables for Inspection by Attributes, Dept. of Defense, 29 April 1963.

7.5 Proposed MIL-STD-1235A (MUCOM), Military Standard, Continuous Sampling Procedures and Tables for Inspection by Attributes, 1966.

7.6 Brugger, Richard M., and Kelley, Harold W., "The Computation of AOQ for the Single Sampling Schemes of MIL-STD-105D," QEM 21-230-7, Ammunition Procurement and Supply Agency, to be published 1969.

7.7 Dodge, H. F., "A Sampling Inspection Plan for Continuous Production," Annals of Mathematical Statistics, Vol. XIV, No. 3, Sep 1943.

7.8 Lieberman, G., and Solomon, H., "Multi-level Continous Sampling Plans," Technical Report No. 17, Sep 1954, Applied Mathematics and Statistics Laboratory, Stanford University, Stanford, California.

7.9 Girshick, M. A., "A Sequential Inspection Plan for Quality Control," Technical Report No. 16, July 1954, Applied Mathematics and Statistics Laboratory, Stanford University, Stanford, California.

7.10 Savage, I. R., "A Three Decision Continuous Sampling Plan for Attributes," Technical Report No. 20, Applied Mathematics and Statistics Laboratory, Stanford University, Stanford, California.

7.11 White, L. S., "The Evaluation of H106 Continuous Sampling Plans under the Assumption of Worst Conditions," Journal of American Statistical Association, No. 61, Sep 1966.

7.12 White, L. S., "On Finding the AOQL's of a Large Class of Continuous Sampling Plans," Technical Report No. 22, Statistical Engineering Group, Columbia University, New York, N. Y., May 1964.

7.13 Endres, Allen, "The Computation of the Unrestricted AOQL When Defective Material is Removed but Not Replaced," QEM 21-240-8, Ammunition Procurement and Supply Agency, November 1967.

7.14 Brugger, Richard M., "The UAOQL of CSP-1, CSP-2, and CSP-R Under the Nonreplacement Assumption," QEM 21-240-9, Ammunition Procurement and Supply Agency, December 1967.

7.15 Murphy, R. B., "A Criterion to Limit Inspection Effort in Continuous Sampling Plans," The Bell System, Technical Journal, Vol. XXXVII, No. 1, January 1958.

7.16 LeMaster, V., and McKeague, R., "Stopping Rules Used in the Proposed DOD Handbook on Single-Level CSP," ESM 401-0-29, Ammunition Procurement and Supply Agency, 28 November 1958.

7.17 Dodge, H. F., and Romig, H. G., Sampling Inspection Tables, Single and Double Sampling, John Wiley and Sons, New York, New York, 2nd Ed., 1959.

APPENDIX A

DERIVATION OF STATE PROBABILITY FORMULAE

A.1 GENERAL

A.1.1 In deriving the steady state and state entrance probabilities, we will define

p = Probability of a defective unit;

q = 1-p = probability of a non-defective unit;

i = clearance number for states 0, N;

i^* = clearance number for states N*, T*, T;

and for j = 0, N, N*, T, T*, R, let

P_j = Prob. (being in state j on the present step);

P'_j = Prob. (entering state j);

f_j = the sampling rate for state j.

A step will be defined as the inspection of a unit of product.

A.1.2 When the process is in states j = N, R, T, some units are being skipped (passed) while others are being sampled and inspected. In the derivations the skipped unit possibilities in these states will be considered. It is convenient, therefore, to partition states j = N, R, T, into skipping and sampling phases.

Let

P_{j_S} = Prob. (being in the skipping phase of state j);

and

P_{j_I} = Prob. (being in the sampling phase of state j).

Then,

$$P_j = P_{j_S} + P_{j_I}.$$

Moreover, it is convenient to partition the skipping phase, j_S, into skip unit phase one and skip unit phase two. Therefore, let

$P_{j_{S_o}}$ = Prob. (being in skip unit phase one of state j);

and

$P_{j_{S_i}}$ = Prob. (being in skip unit phase two of state j).

Then,

$$P_{j_S} = P_{j_{S_o}} + P_{j_{S_i}}$$

for

j = N, R, T.

Skip unit phase one will be defined as that phase of j initially entered, and skip unit phase two will be defined as that phase of j_S in which all subsequent skips occur. Skip unit phase one may therefore be viewed as being a "transitional" phase between the last step in some previous state and the first step in the present state.

The preceding state/phase symbols with primes will be used to denote the probability of entering a given state/phase on the present step.

A.2 EXPRESSIONS FOR THE STEADY STATE PROBABILITIES

(1) P_0 = Prob. (just entering state 0 on the last step) + Prob. (entering 0, two steps ago, and inspecting a non-defective on the last step) + . . . + Prob. (entering 0, i steps ago, and inspecting i-1 consecutive non-defective units)

$$= P'_0 + P'_0 q + P'_0 q^2 + \ldots + P'_0 q^{i-1}$$

$$= P'_0 (1-q^i)/p$$

(2) P_N = Prob. (being in the sampling phase of state N) + Prob. (being in the skip unit phase of state N)

$$= P_{N_I} + P_{N_S} .$$

(3) P_{N_I} = Prob. (just entering phase N_I on the last step) + Prob. (entering phase N_I two steps ago and inspecting a good unit on the last step) + . . . + Prob. (entering N_I, i steps ago, and inspecting i-1 consecutive non-defective units)

$$= P'_{N_I} + P'_{N_I}q + P'_{N_I}q^i + \ldots + P'_{N_I}q^{i-1}$$

$$= P'_{N_I}(1-q^i)/p.$$

(4) P_{N_S} = Prob. (entering skip unit phase of N and passing (skipping) the next $(1/f_N)-1$ units)

$$= P'_{N_S}[(1/f_N)-1].$$

Similarly then,

(5) $P_T = P_{T_I} + P_{T_S}$

(6) $P_{T_I} = P'_{T_I}(1-q^{i*})/p$

(7) $P_{T_S} = P'_{T_S}[(1/f_T)-1]$

(8) $P_R = P_{R_I} + P_{R_S}$

(9) $P_{R_I} = P'_{R_I}(1/p)$

(10) $P_{R_S} = P'_{R_S}[(1/f_R)-1]$

(11) $P_{T*} = P'_{T*}(1-q^{i*})/p$

(12) $P_{N*} = P'_{N*}(1-q^{i*})/p$

A.3 EXPRESSIONS FOR THE STATE ENTRANCE PROBABILITIES

With the aid of FIGURE II we obtain

(13) P'_0 = Prob. (being in state 0 and finding a defective unit) + Prob. (being in state R and finding a defective unit) + Prob. (being in state T and finding a defective unit)

$= P_0 \cdot p + P_{R_I} \cdot p + P_{T_I} \cdot p.$

Combining (1), (6), and (9) with the above, yields

(14) $P'_0 = P'_0 (1-q^i) + P'_{R_I} + P'_{T_I} (1-q^{i*}).$

In a similar manner the other P'_j 's are obtained:

(15) $P'_{N*} = P'_{N_I} (1-q^i)$

(16) $P'_{T*} = P'_{T*} (1-q^i) + P'_{N*} (1-q^{i*})$

(17) $P'_{N_{S_o}} = P'_0 q^i + P'_{T_I} q^{i*} + P'_{N*} q^{i*}$

(18) $P'_{N_{S_i}} = P'_{N_I} q + P'_{N_I} q^2 + P'_{N_I} q^3 + \ldots + P'_{N_I} q^{i-1}$

(19) $P'_{N_I} = P'_{N_{S_o}}$

(20) $P'_{T_{S_o}} = P'_{T*} q^{i*}$

(21) $P'_{T_{S_i}} = P'_{T_I} q + P'_{T_I} q^2 + \ldots + P'_{T_I} q^{i*-1}$

(22) $P'_{T_I} = P'_{T_{S_o}}$

(23) $P'_{R_{S_o}} = P'_{N_I} q^i$

$$(24) \qquad P'_{R_{S_i}} = P'_{R_I} q + P'_{R_I} q^2 + \ldots + P'_{R_I} q^i + P'_{R_I} q^{i+1} + \ldots$$

$$(25) \qquad P'_{R_I} = P'_{R_{S_o}} .$$

By definition

$$(26) \qquad P'_{j_S} = P'_{j_{S_o}} + P'_{j_{S_i}} .$$

Then, from (17) and (18) we obtain

$$(27) \qquad P'_{N_S} = P'_{N_{S_o}} + P'_{N_{S_i}}$$

$$= P'_{N_I} + P'_{N_I} q + P'_{N_I} q^2 + \ldots + P'_{N_I} q^{i-1}$$

$$= P'_{N_I} (1-q^i)/p .$$

Similarly,

$$(28) \qquad P'_{T_S} = P'_{T_I} (1-q^{i*})/p$$

$$(29) \qquad P'_{R_S} = P'_{R_I} (1/p) .$$

A.4 EXPRESSIONS FOR THE STEADY STATE AND STATE ENTRANCE PROBABILITIES IN TERMS OF KNOWN PARAMETERS.

A.4.1 Equations (14), (15), (16), (19), (22), (25), (27), (28), (29) define nine equations in nine unknown entrance probabilities. These equations and their associated steady state probabilities may be expressed in terms of parameters p, q, i, i*, and f_j, which are assumed known. This section discusses the derivation of such expressions.

A.4.2 In lieu of solving explicitly for each P'_j and P_j, it was convenient to first express each state entrance probability, P_j, in terms of P'_{N_I}.

Equations (14) through (29) were then used to obtain:

$$(14') \quad P'_0 = P'_{N_I} [q^i + (1-q^i)(1-q^{i*})^2]/q^i$$

$$(15') \quad P'_{N*} = P'_{N_I} (1-q^i)$$

$$(16') \quad P'_{T*} = P'_{N_I} (1-q^i)(1-q^{i*})/q^{i*}$$

$$(22') \quad P'_{T_I} = P'_{N_I} (1-q^i)(1-q^{i*})$$

$$(25') \quad P'_{R_I} = P'_{N_I} q^i$$

$$(27') \quad P'_{N_S} = P'_{N_I} (1-q^i)/p$$

$$(28') \quad P'_{T_S} = P'_{N_I} (1-q^i)(1-q^{i*})^2/p$$

$$(29') \quad P'_{R_S} = P'_{N_I} q^i/p$$

When the preceding primed equations are substituted into equations (1), (12), (11), (6), (9), (4), (7), and (10) respectively, the following steady state probability equations are obtained in terms of P'_{N_I}:

$$(1') \quad P_0 = P'_{N_I} [q^i + (1-q^i)(1-q^{i*})^2] (1-q^i)/p\, q^i$$

$$(3') \quad P_{N_I} = P'_{N_I} (1-q^i)/p$$

$$(4') \quad P_{N_S} = P'_{N_I} [(1-q^i)/p] [(1/f_N)-1]$$

$$(6') \quad P_{T_I} = P'_{N_I} (1-q^i)(1-q^{i*})^2/p$$

$$(7') \quad P_{T_S} = P'_{N_I} [(1-q^i)(1-q^{i*})^2] [(1/f_T)-1]/p$$

$$(9') \qquad P_{R_I} = P'_{N_I} q^i/p$$

$$(10') \qquad P_{R_S} = P'_{N_I} q^i [(1/f_R)-1]/p$$

$$(11') \qquad P_{T*} = P'_{N_I} (1-q^i) (1-q^{i*})^2/q^{i*} p$$

$$(12') \qquad P_{N*} = P'_{N_I} (1-q^i) (1-q^{i*})/p$$

Since $\Sigma P_j = 1$, equations (1'), (3'), (4'), (6'), (7'), (9'), (10'), (11'), and (12') can be combined to obtain

$$(30) \qquad P'_{N_I} = p\, q^{i*} q^i/D \ ;$$

where

$$D = q^{i*} (1-q^i) [q^i + (1-q^i) (1-q^{i*})^2] + q^i (1-q^i) (1-q^{i*}) + q^i (1-q^i) (1-q^{i*})^2 + q^{i*} q^i (1-q^i)/f_N + q^{2i} q^{i*}/f_R + q^i q^{i*} (1-q^i) (1-q^{i*})^2/f_T.$$

Expressions for the steady state probabilities in terms of known parameters can now be obtained by substituting equation (30) into the primed number equations.

$$(1'') \qquad P_0 = q^{i*} (1-q^i) [q^i + (1-q^i) (1-q^{i*})^2]/D$$

$$(3'') \qquad P_{N_I} = q^{i*} q^i (1-q^i)/D$$

$$(4'') \qquad P_{N_S} = q^{i*} q^i (1-q^i) [(1/f_N)-1]/D$$

$$(6'') \qquad P_{T_I} = q^{i*} q^i (1-q^i) (1-q^{i*})^2/D$$

$$(7'') \qquad P_{T_S} = q^{i*} q^i (1-q^i) (1-q^{i*})^2 [(1/f_T)-1]/D$$

$$(9'') \qquad P_{R_I} = q^{2i} q^{i*}/D$$

(10") $P_{R_S} = q^{2i}\, q^{i*}\, [(1/f_R)-1]/D$

(11") $P_{T*} = q^{i}\, (1-q^{i})\, (1-q^{i*})^2/D$

(12") $P_{N*} = q^{i*}\, q^{i}\, (1-q^{i})\, (1-q^{i*})/D$

APPENDIX B

DERIVATION OF AFI AND AOQ

B.1 THE AFI FUNCTION

B.1.1 By definition AFI is the expected ratio of the total number of units inspected to the total number of units inspected or passed. Thus, by letting

K_j = the number of units passing through the inspection system in state j;

K_{j_I} = the number of units inspected in state j;

K_{j_S} = the number of units skipped (passed) in state j;

$K_j = K_{j_I} + K_{j_S}$;

and $K = \sum K_j$; where j = O, N, R, T, N*, T*,

we may write

$$(1) \qquad AFI = \lim_{\sum K_j \to \infty} \frac{\sum K_{j_I}}{\sum K_{j_I} + \sum K_{j_S}}$$

$$= \lim_{K \to \infty} \frac{K_O + K_{N*} + K_{T*} + K_{N_I} + K_{T_I} + K_{R_I}}{K_O + K_{N*} + K_{T*} + K_{N_I} + K_{T_I} + K_{R_I} + K_{N_S} + K_{T_S} + K_{R_S}}$$

$$= P_O + P_{N*} + P_{N_I} + P_{T*} + P_{T_I} + P_{R_I} \quad .$$

Using (1"), (3"), (4"), (6"), (7"), (9"), (10"), (11"), and (12") of A.4, APPENDIX A, expressions for the AFI in terms of the parameters p, q, i, i*, and f_j are obtained.

B.2 THE AOQ FUNCTION

Dodge and Romig [7.17], have given expressions for the AOQ functions under two assumptions:

Case I: Defective units are removed and replaced by non-defective units.

Case II: Defective units are removed but not replaced.

Case two (II) is consistent with standard operating procedure in most ammunition inspection situations. Accordingly, appealing to the Dodge and Romig expression, we used the following:

$$(1) \qquad AOQ = \frac{p[1 - AFI]}{1 - p(AFI)} \quad ; \text{ where}$$

AFI is as defined by (1) of B.1 and p is the probability of a defective unit.

APPENDIX C

DERIVATION OF O.C.

By definition the fraction of product accepted on a sampling basis is the expected ratio of the number of units accepted on a sampling basis to the total number of units inspected or passed. Now, recalling the assumption that defective units are removed but not replaced, the number of units accepted on a sampling basis must obviously consist of only those units inspected and found non-defective plus those units passed (skipped and therefore accepted) in the sampling inspection states. Thus, appealing to the notation of A.1 of APPENDIX A, and B.1 of APPENDIX B, we write for j = N, R, and T

$$(1)\quad \text{O.C. } (\%) = \lim_{K \to \infty} \frac{\sum_j \left(K_{j_S} + qK_{j_I} \right)}{K} \times 100$$

$$= \left[P_{N_S} + P_{T_S} + P_{R_S} + qP_{N_I} + qP_{T_I} + qP_{R_I} \right] \times 100 .$$

Using the equations of A.2 and A.3 of APPENDIX A, it can be shown that

$$P_{N_S} = P_{N_I} [(1/f_N)-1],$$

$$P_{R_S} = P_{R_I} [(1/f_R)-1], \text{ and}$$

$$P_{T_S} = P_{T_I} [(1/f_T)-1] .$$

Therefore, equation (1) above can be written as

$$(2)\quad \text{O.C. } (\%) = 100 \{[(1/f_N)-1 + q] P_{N_I} + [(1/f_R)-1 + q] P_{R_I} + [(1/f_T)-1 + q] P_{T_I}\}$$

$$= 100 \{[(1/f_N)-p] P_{N_I} + [(1/f_R)-p] P_{R_I} + [(1/f_T)-p] P_{T_I}\} .$$

AN EVALUATION OF LINEAR LEAST SQUARES COMPUTER PROGRAMS: A SUMMARY REPORT

Roy H. Wampler
National Bureau of Standards
Washington, D. C.

ABSTRACT. Two linear least squares test problems based on fifth degree polynomials have been run on more than twenty different computer programs in order to assess their numerical accuracy. Among the programs tested were representatives from various statistical packages as well as some from the SHARE library. Essentially four different algorithms were used in the various programs to obtain the coefficients of the least squares fits. The tests were run on several different computers, in double precision as well as single precision. By comparing the coefficients reported, it was found that those programs using orthogonal Householder transformations or Gram-Schmidt orthonormalization were much more accurate than those using elimination algorithms. Programs using orthogonal polynomials (suitable only for polynomial fits) also proved to be superior to those using elimination algorithms. The most successful programs accumulated inner products in double precision and made use of iterative refinement procedures. In a number of programs, the coefficients reported in one test problem were sometimes completely erroneous, containing not even one correct significant digit.

1. INTRODUCTION. Since the time when the electronic computer began to supplant the desk calculator as the chief tool for solving linear least squares problems, numerous least squares computer programs have been written. These programs have utilized a variety of computational algorithms. Because least squares problems are by their very nature frequently ill-conditioned, the numerical accuracy achieved by a least squares program strongly depends upon the choice of the algorithm. Many programs have been written which use methods appropriate for desk calculators but inappropriate for computers. Anscombe [1] has aptly remarked: "Textbooks of statistical method display a wonderful unanimity in recommending computational procedures that are suited to desk calculators but are perilous for computers. Only with some determination can the statistician break himself of bad habits and become adequately informed about round-off error."

The present study was undertaken to assess the numerical accuracy of representative least squares programs from a variety of sources. Two test problems, both fifth degree polynomials, have been run on more than twenty different programs. Included in the study were programs from the BMD Biomedical Computer Programs collection [14], the C-E-I-R Multi-Access Computing Services library [10], the IBM SHARE library [23], the IBM System/360 Scientific Subroutine Package [22], the Univac MATH-PACK [33] and STAT-PACK [34] collections, and the Project MAC 7094 disk files [28]. A listing of the sources of the programs is given in Appendix A, together with a brief description of each program.

For a number of programs, the test problems were run in double precision as well as in single precision. This, of course, necessitated certain changes in the original programs.

The programs included in this study used essentially four different algorithms: orthogonal Householder transformations; Gram-Schmidt orthonormalization; orthogonal polynomials; and, Gaussian or Jordan elimination.

The linear least squares problem may be briefly stated as follows: One has n observations or measurements of a "dependent" variable y, which are statistically independent with common variance σ^2, whose expected values are given by a linear function of the corresponding values of k "independent" variables, $x_1, x_2, \ldots, x_k$, $k \leq n$. In matrix notation we say that the n observations have expected values $E(Y) = X\beta$, where Y is an n x 1 vector, X is an n x k matrix, and β is a k x 1 vector of unknown coefficients. Assuming that X is of rank k, the least squares estimates of the coefficients are given by $\hat{\beta} = (X'X)^{-1}X'Y$. Other quantities of interest are $\hat{Y} = X\hat{\beta}$, the vector of predicted values; $\delta = Y - \hat{Y}$, the vector of residuals; and $s^2 = \delta'\delta/(n-k)$, an estimate of the variance σ^2.

In running certain programs, modifications were occasionally made to input and output formats. Other changes were made in five of the programs using elimination algorithms because the original versions of these programs failed to give solutions to the fifth degree polynomial problems. The nature of these changes will be described in the discussion of the individual programs in Section 7.

Three computers were used: the GE 235, the IBM 7094, and the Univac 1108. The 1108 which was used is located at the National Bureau of Standards, and the 7094 which was chiefly used is located at Harry Diamond Laboratories, Washington, D. C. The programs run on the 235 and the Project MAC 7094 utilized consoles at the National Bureau of Standards connected to computers at other locations.

Previous studies appraising linear least squares programs and comparing the results of different algorithms have been made by Cameron [9], Freund [18], Bright and Dawkins [7], Zellner and Thornber [38], Longley [25], and Jordan [24]. The present study differs from the earlier ones mainly by including a larger selection of widely used and easily accessible programs.

A more detailed report of the present study is given in Wampler [36]. The more detailed version contains an appendix giving the individual coefficients obtained in running each program, an investigation into the effect of rounded input on the solution of a least squares problem, additional details pertaining to certain programs, and results from some additional test problems. The longer report also includes several programs designed not specifically for solving least squares problems but for solving n equations in n unknowns, thus forcing one to use X'X and X'Y as input. Since it is well known that this is not, in general, a good method for solving least

squares problems, these programs are omitted from the present summary report. (There was one outstanding exception among the programs requiring X'X and X'Y as input. This was Newman's program, described in [30], which requires integer input and uses integer arithmetic and congruential methods to obtain exact solutions.) The present report gives results of one program (BJORCK-GOLUB) not included in the more detailed report.

It was outside the scope of the present study to make a detailed comparison of algorithms with respect to efficiency of computation time and storage requirements. The programs which were included in this study exhibited considerable variation in what quantities were calculated as well as in the methods of calculation, and output ranged from meager to copious. Moreover, no comparative examination of the outputs provided by the programs was made. Rather, this investigation focused attention on the performance of existing programs.

2. THE TEST PROBLEMS. The two test problems which were used throughout this investigation are identified as Y1 and Y2. Both were fifth degree polynomials, with the values of x being the integers 0, 1, 2, ..., 20. The "observations," Y1 and Y2, were calculated from the following equations:

Y1: $y = 1 + x + x^2 + x^3 + x^4 + x^5$, $x = 0(1)20$

Y2: $y = 1 + .1\,x + .01\,x^2 + .001\,x^3 + .0001\,x^4 + .00001\,x^5$, $x = 0(1)20$

Thus, the values of Y1 were integers having from one to seven digits, and those of Y2 were five-decimal numbers ranging from 1.00000 to 63.00000.

If the least squares solutions were computed with no rounding error, one would obtain

$$\hat{\beta}(Y1) = \begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \\ 1 \\ 1 \end{bmatrix}, \qquad \hat{\beta}(Y2) = \begin{bmatrix} 1. \\ .1 \\ .01 \\ .001 \\ .0001 \\ .00001 \end{bmatrix},$$

and for both problems the residual standard deviation would be zero.

For some programs the input required was the 21 values of x and y. Other programs required, in addition, the powers x^2, x^3, x^4, and x^5 to be entered as input. The input is listed in Table 5, along with the matrices X'X and X'Y associated with the test problems.

The two test problems, Y1 and Y2, were chosen because they are so highly ill-conditioned that some programs fail to obtain correct solutions while other programs succeed in obtaining reasonably accurate solutions. Polynomial problems were chosen because polynomial fitting is an important type of linear least squares problem which occurs frequently in practice.

The ill-conditioning of the two test problems can be described more explicitly. One measure of the condition of a matrix A is the P-condition, defined as

$$P(A) = \left| \frac{\lambda}{\mu} \right|$$

where λ is the numerically largest eigenvalue of A and μ is the numerically smallest eigenvalue of A. (See Newman [29, p. 240]).

For A = X'X, the 6 x 6 matrix associated with Y1 and Y2, the P-condition is 4.095×10^{13}. In this respect, it is similar to the Hilbert matrix of order 10, whose P-condition is 1.603×10^{13} (see Fettis and Caslin [16]). The P-condition of the Hilbert matrix of order 11 is 5.231×10^{14}. The relation between the Hilbert matrix and the matrix X'X which arises in a polynomial fit is discussed in Forsythe [17].

Most of the programs which were tested obtained more accurate solutions for Y2 than for Y1. If we let A denote the 7 x 7 matrix

$$A = \begin{bmatrix} X'X & X'Y \\ Y'X & 0 \end{bmatrix}$$

we find that for Y2, $P(A) = 4.095 \times 10^{13}$, whereas for Y1, $P(A) = 6.829 \times 10^{13}$, indicating that the system involving Y1 is more ill-conditioned than that involving Y2.

The test problem used by Longley [25] was also highly ill-conditioned. For the 7 x 7 matrix X'X of his problem, the P-condition is 2.361×10^{19}.

3. SUMMARY OF THE RESULTS. Tables 1 to 4 present a brief summary of the main results. A count, C_j, of the number of correct significant digits in each computed coefficient was obtained as follows:

Let β_j (j = 1, 2, ..., 6) denote the "true" value of the coefficient -- that is, the value computed with no rounding error. Let $\hat{\beta}_j$ denote the value calculated by the computer. Then

$$C_j = \begin{cases} -\log_{10}\left|\dfrac{\beta_j - \hat{\beta}_j}{\beta_j}\right|, \text{ if } \left|\beta_j - \hat{\beta}_j\right| \neq 0 \text{ and } \beta_j \neq 0 \\ -\log_{10}\left|\beta_j - \hat{\beta}_j\right|, \text{ if } \left|\beta_j - \hat{\beta}_j\right| \neq 0 \text{ and } \beta_j = 0 \\ D, \text{ the approximate number of decimal digits with which the machine computes, if } \beta_j - \hat{\beta}_j = 0. \end{cases}$$

The above approach to counting the number of correct digits in a computed value has been used by Jordan [24] and others.

Tables 1 to 4, in the columns headed "Average Number of Correct Digits" report $C = \frac{1}{6}\sum_{j=1}^{6} C_j$.

From the above definition, a negative count can occur. For example, if $\beta_j = 1.0$, and $\hat{\beta}_j = 136.0$, we get $C_j = -2.130$. This indicates that $\hat{\beta}_j$ is wrong by roughly two orders of magnitude.

For two programs reported in Table 1, BMD03R run on the 7094 and DAM run on the 7094, the count for several coefficients was made in a different manner. The BMD03R program printed the coefficients in a fixed-decimal format, with five decimals. The DAM program used a floating-point format with only three decimals printed. A coefficient printed as .00010, when the true coefficient was .0001, was given a count of 2, and 0.100E01, when the true coefficient was 1., was given a count of 3. In such cases the assigned count may have been too small, since the coefficients may have been calculated accurately to more digits than were printed. In running these two programs on the 1108, the output format was changed so that eight significant digits were printed.

Each of the tables (1 through 4) summarizes a set of results for a particular machine precision. Within each table the various programs are given a numerical rank for each of the two test problems, with rank 1 denoting the best performance according to the count C.

4. PROGRAMS USING ORTHOGONAL HOUSEHOLDER TRANSFORMATIONS. LSTSQ is a program written by Peter A. Businger using orthogonal Householder transformations. This algorithm is described by Golub [19], and Businger and Golub [8]. The program applies a sequence of orthogonal transformations

to the n x k least squares matrix X to obtain a decomposition X = QR, where R is upper triangular and $Q'Q = I_k$. A pivoting strategy is used so that at each step the column with the largest sum of squares is reduced next. Once an initial solution is obtained, the program iterates to obtain a (possibly) improved solution.

The BJORCK-GOLUB program uses the Householder transformation algorithm described by Björck and Golub [6]. This algorithm takes advantage of the fact that $X'\delta = 0$, where δ is the vector of residuals, to obtain the solution β in $X\beta = Y$ from the augmented system of n + k equations:

$$\begin{bmatrix} I & X \\ X' & 0 \end{bmatrix} \cdot \begin{bmatrix} \delta \\ \beta \end{bmatrix} = \begin{bmatrix} Y \\ 0 \end{bmatrix}$$

Here δ as well as β is included in the iterative refinement procedure.

Of all the programs included in this study, LSTSQ and BJORCK-GOLUB appear to have given the best performance. In Table 3, which reports the performance of eleven double precision programs, we see that LSTSQ ranked first for Y1 and second for Y2, and that BJORCK-GOLUB ranked first for Y2 and second for Y1. In Table 1, which reports the performance of 20 single precision programs, we see that LSTSQ ranked first for Y1 and fourth for Y2, and that BJORCK-GOLUB ranked second for Y1 and third for Y2. Ranks 1 and 2 for the Y2 problem were obtained by ORTHOL and OMNITAB (using ORTHO), two programs using Gram-Schmidt orthonormalization which will be discussed in the next section. Table 4 reports the performance of four programs which used single precision arithmetic except for the accumulation of inner products, where double precision arithmetic was used. Here we see that LSTSQ and BJORCK-GOLUB tied to obtain the top rank for Y1 (having perfect scores of 8.000), but ranked third and fourth, respectively, for Y2. In Table 4, we note that all four programs obtained similar scores for the Y2 problem, with rank 1 corresponding to 6.530 and rank 4 to 6.227. In the Businger-Golub and Björck-Golub algorithms, it is recommended that all inner products be accumulated in double precision. By comparing Tables 4 and 1 we see that when LSTSQ included this feature, the average counts increased from 4.528 to 8.000 for Y1 and from 5.840 to 6.279 for Y2. With <u>all</u> operations performed in double precision (see Table 3), the counts increased to 14.643 and 16.293, respectively. The BJORCK-GOLUB program displayed similar improvements in accuracy when inner products were accumulated in double precision and when all operations were carried out in double precision.

Another program using Householder transformations was ALSQ, a program containing no pivoting and no iteration. In Tables 1, 3, and 4 we see that ALSQ performed not quite as well as the LSTSQ and BJORCK-GOLUB programs which included these features, except in one instance. In this one instance, Y2 in Table 4, we note that its performance was slightly better than that of the other programs in this category.

5. PROGRAMS USING GRAM-SCHMIDT ORTHONORMALIZATION. ORTHO is a program written by Philip J. Walsh using a Gram-Schmidt orthonormalization process. This algorithm is described by Davis and Rabinowitz [13], Davis [12], and Walsh [35]. ORTHO exists as a FORTRAN program, an ALGOL procedure, a BASIC program, and as a routine of the OMNITAB program [21].

Starting with the n x k matrix X, the Gram-Schmidt process of ORTHO obtains $\phi = XT'^{-1}$ and $\hat{\beta} = T'^{-1}\,\phi'Y$, where T'^{-1} is upper triangular and $\phi'\phi = I_k$. This algorithm includes a feature of reorthonormalizing the vectors of ϕ, proceeding from a first approximation $\overline{\phi}_j$ to a (usually) better approximation ϕ_j. From Table 1 it is clear that this reorthonormalizing is vital to the algorithm, for ORTHO's good performance in handling Y1 and Y2 deteriorated when this iteration was omitted. For Y1, the count of correct digits dropped from 4.137 to -1.976, and for Y2 the drop was from 5.464 to 0.419. In Table 3, also, we see that in double precision the omission of the iteration resulted in a loss of about five correct digits for both problems.

Of the six programs in Table 2, LSFITW***, written in BASIC, ranked first on both problems. We note that Table 2 includes no Householder transformation programs.

The ORTHO program was also run in a version using single precision except for the accumulation of inner products, where double precision was used. In Table 4 we see that there were four programs in this category, and ORTHO ranked third for Y1 and second for Y2.

ORTHOL is a program using a modification of the Davis-Rabinowitz algorithm. It differs from ORTHO in two respects: (1) the iteration procedure includes the dependent variable as well as the independent variables; and, (2) before any other operations are applied to the matrix X, from each element of each vector of X, the truncated mean of that vector is subtracted. (The "truncated mean" denotes the largest integer less than or equal to the mean if the mean is nonnegative, and the smallest integer greater than or equal to the mean if the mean is negative.) ORTHOL obtained the top rank for Y2 in single precision, but ranked sixth for Y1 (Table 1). In double precision (Table 3), it ranked third on both problems.

6. PROGRAMS USING ORTHOGONAL POLYNOMIALS. Since the two test problems are both polynomial fits, we were able to test programs in which the algorithm used orthogonal polynomials. This method, described by Forsythe [17], is attractive because it generally requires many fewer operations than other methods.

Two such programs were included in this study. One was the UNIVAC 1108 MATH-PACK ORTHLS routine [33]. The other was POLFIT, an anonymous program written in BASIC.

In Tables 1, 2, and 3, we see that the performance of the orthogonal polynomial programs is not as good as that of the Householder transformation and the Gram-Schmidt programs (with iteration), but the performance is better than that of any of the programs using elimination algorithms.

7. PROGRAMS USING ELIMINATION ALGORITHMS. The majority of the programs tested in this investigation used some form of an elimination algorithm. Although this was the most popular method, it was the least successful. None of these programs performed as well as those using Householder's transformations, Gram-Schmidt orthonormalization (with iteration), or orthogonal polynomials.

Within this class of programs, there were several variations in the method of obtaining the least-squares coefficients. In some cases, the matrix X'X was inverted, after which the inverse was postmultiplied by X'Y. One program inverted the matrix Z'Z where the vectors of Z were obtained from the vectors of X by subtracting the mean of each vector from every element of that vector. A number of programs obtained the solution by inverting a matrix of correlation coefficients. The five stepwise regression programs made use of matrix partitioning in connection with inverting a matrix of correlation coefficients.

The five stepwise regression programs were BMD02R, MPR3, the STAT-PACK program RESTEM, WRAP, and STAT20***. They all, to a greater or lesser extent, follow Efroymson's algorithm [15]. Tables 1, 2, and 3 give the results of these five programs.

In running the two test problems on three of the stepwise programs, namely, BMD02R, RESTEM and STAT20***, calculations stopped before the solutions were obtained. These programs at various steps calculate an F-level in connection with entering or removing variables, and a point was reached where this F-level was calculated to be negative because of rounding error. Since this condition caused the calculations to stop, certain steps of the algorithm had to be bypassed to obtain the final solution. These steps were not, however, connected with the calculation of the least squares coefficients.

WRAP, the program with the lowest rankings in Table 1, computed coefficients which were exceptionally far from the true values. These coefficients are listed below.

Y1		Y2	
True $\hat{\beta}$	Computed $\hat{\beta}$	True $\hat{\beta}$	Computed $\hat{\beta}$
1.	2991622.	1.	-33.84546
1.	-6065892.	.1	71.54880
1.	2218821.	.01	-26.16913
1.	-296194.5	.001	3.493256
1.	16462.20	.0001	-.1936966
1.	-322.5731	.00001	.003812985

Two other BMD programs, in addition to BMD02R mentioned earlier, were tested. These were BMD03R, Multiple Regression with Case Combinations, which inverts a matrix of correlation coefficients, and BMD05R, Polynomial Regression, which inverts the matrix Z'Z where the vectors of Z are formed from the vectors of X by subtracting the mean of each vector from every element of that vector. All the crucial operations of BMD05R, such as the forming of inner products and matrix inversion, are carried out in double precision. The performance of BMD03R and BMD05R is shown in Tables 1 and 3, respectively.

DAM is a general-purpose computer program for data processing and multiple regression [31]. In running the two test problems on DAM on the 1108, computations stopped after a fourth degree polynomial was fitted. It was found that a computed variance was zero and that this condition causes the computations to stop. By bypassing the checks on this computed variance, results for fifth degree fits were obtained. On the 7094, however, the fifth degree results were reached without any such difficulties. DAM's performance on the two computers is given in Table 1.

The program POLRG is the polynomial regression program of the IBM System/360 Scientific Subroutine Package [22]. We see from Table 1 that the single precision version of POLRG obtained rather low scores on both test problems. A double precision version of POLRG was also run, and the performance here as reported in Table 3 was comparable to other programs using similar elimination algorithms.

The user of POLRG specifies m, the highest degree polynomial to be fitted, and the program automatically reports the results of fitting polynomials of successively increasing degrees, starting with the first degree. If there is no reduction in the residual sum of squares between two successive degrees of polynomials, the program stops the problem before completing the analysis for the highest degree specified. In running both test problems in single precision the analysis stopped after degree four, and in lieu of a fifth degree polynomial fit, the message "NO IMPROVEMENT" was printed. In order to complete the calculations for the fifth degree, the checks on "improvement" were bypassed. In the double precision version, fifth degree results were obtained without any such alterations.

Each of the two STAT-PACK programs, GLH, General Linear Hypotheses, and REBSOM, Back Solution Multiple Regression, has its individual features, but for the two test problems the solutions were carried out in the same manner, so that the coefficients obtained from the two programs were identical, as is indicated in Table 1. Both programs invert X'X by calling the same matrix inversion subroutine which uses a Gauss-Jordan elimination scheme with maximal column pivoting and row scaling.

The BASIC program LINFIT*** in order to obtain $\hat{\beta}$ inverts the matrix

$$A = \begin{bmatrix} X'X & X'Y \\ Y'X & \sum y_i^2 \end{bmatrix}$$

whose inverse, if it exists, is

$$\begin{bmatrix} (X'X)^{-1} + \dfrac{\hat{\beta}\hat{\beta}'}{\sum y_i^2 - Y'\hat{Y}} & \dfrac{-\hat{\beta}}{\sum y_i^2 - Y'\hat{Y}} \\ \dfrac{-\hat{\beta}'}{\sum y_i^2 - Y'\hat{Y}} & \dfrac{1}{\sum y_i^2 - Y'\hat{Y}} \end{bmatrix}.$$

When $\hat{Y} = Y$, the matrix A is singular. In the two test problems $\hat{Y} = Y$, so that the matrix A, if it were formed in the computer without any rounding error, would be singular. But A, for Y1 and Y2, contains 14-digit numbers, whereas the 235 computer works with approximately nine-digit numbers, so that rounding of the elements of A is inevitable, and the version of A contained in the computer is not singular. An "inverse" was obtained, and from this $\hat{\beta}$ was immediately computed. Table 2 gives the results.

LSCF--*** and STAT21*** are two BASIC programs available in the C-E-I-R Multi-Access Computer Service; results are given in Table 2. LSCF--***, which obtains the coefficients by inverting X'X and then post-multiplying the inverse by X'Y, had the lowest rankings of Table 2. STAT21*** obtains $(X'X)^{-1}$ and $\hat{\beta}$ by applying Jordan elimination to X'X and X'Y.

The LINFIT program included in Table 1 is one of eighteen statistical routines described by Miller [28] which exist in the Project MAC* 7094 disk files. The two test problems were run on the LINFIT program on a time-

*A description of Project MAC is given in Crisman [11].

shared computer via a remote console communicating with Project MAC. The method used by the LINFIT program is not given. By conjecture, it has been included in this section among programs using elimination algorithms.

8. OTHER RECENT ALGORITHMS. Some other algorithms apparently of high quality which have been published in the last few years were not included in this study. Two such algorithms are given by Bauer [2] and Björck [5].

Bauer [2] gives an ALGOL procedure using iterative refinement for finding the least squares solution of $X\beta = Y$, where X is n x k ($k \leq n$) of rank k and Y is n x p. The procedure is based on the decomposition of X into UDR where U is n x k with orthogonal columns, $D = (U'U)^{-1}$, and R is upper triangular. This decomposition yields a triangular system $R\beta = U'Y$ which is solved by back substitution. The reduction to $R\beta = U'Y$ is carried out by a Gaussian elimination scheme, but with a suitably weighted combination of rows used for elimination instead of a single row.

Björck's algorithm [5] (see also Björck [3], [4]) using a modified Gram-Schmidt orthogonalization process, has certain features in common with the Björck-Golub algorithm discussed in Section 4 above. Two such features are solving the system of n + k equations

$$\begin{bmatrix} I & X \\ X' & 0 \end{bmatrix} \cdot \begin{bmatrix} \delta \\ \beta \end{bmatrix} = \begin{bmatrix} Y \\ 0 \end{bmatrix}$$

to obtain β and δ, and inclusion of δ as well as β in the iterative refinement procedure.

Both the classical Gram-Schmidt orthogonalization process and the modified Gram-Schmidt orthogonalization process, as described by Björck [3], decompose the matrix X into QR where Q'Q is diagonal and R is upper triangular. In the classical procedure, at the i-th stage, the i-th column vector is made orthogonal to each of the i - 1 previously orthogonalized column vectors; this is done for column indices i = 2, 3, ..., k. In the modified procedure which Björck uses, at the i-th stage, the (k - i + 1) column vectors indexed i, i + 1, ..., k are made orthogonal to the (i - 1)-th column vector; this is done for column indices i = 2, 3, ..., k. Jordan [24] shows why the modified procedure is superior to the classical procedure. Björck [3] states that his modified Gram-Schmidt procedure is equivalent to Bauer's method using weighted row combinations mentioned above. Björck's algorithm is generalized to handle the case where X is of less than full rank; here, linear constraints are entered.

Björck [3], [5] discusses the number of operations and the storage requirements of his algorithm, and he compares the number of operations needed with the corresponding number needed in the Björck-Golub algorithm [6].

9. CONCLUSIONS.

(1) Computational procedures appropriate for desk calculators may be perilous for computers.

(2) Of the four procedures which were included in this study, orthogonal Householder transformations and Gram-Schmidt orthonormalization proved to be the best. Orthogonal polynomials ranked next. Elimination methods were the least successful but the most popular.

(3) Programmers who have been writing least squares programs, especially for statistical packages, have often not been taking advantage of the advances in this area made by numerical analysts in recent years.

(4) The importance of accumulating inner products in double precision cannot be overstressed. A number of recent papers on least squares computations have emphasized this point. These include Businger and Golub [8], Bauer [2], Golub and Wilkinson [20], Björck and Golub [6], and Björck [5]. On many third-generation computers which have double precision built into the hardware, double precision arithmetic is quite efficient.

(5) Iterative refinement is another valuable feature of recent algorithms. Five programs included in the present study (BJORCK-GOLUB, LSFITW***, LSTSQ, ORTHO and ORTHOL) made effective use of iterative refinement, and the two algorithms described in Section 8 both include this feature. Golub and Wilkinson [20] give a discussion of this topic.

(6) The users of least squares programs can take certain precautionary steps to gain an awareness of whether or not a rounding error problem exists. Among the suggestions which have been made here are the following:

(a) Run test problems where the coefficients are known (Cameron [9]).

(b) Transform the data; e.g., by subtracting means (Freund [18], Longley [25]).

(c) Do the calculations several times, scaled differently each time (Zellner and Thornber [38], Longley [25]).

(d) Shuffle the columns of X and run the problem more than once (Longley [25]).

(e) Check whether $X'\delta = 0$ (Longley [25]).

(f) Use double precision arithmetic (Freund [18]).

(g) Follow the initial fit by a fit to $\hat{Y}$, the predicted values (suggested by J. M. Cameron; see Wampler [36]).

(7) In any mathematical calculation carried out on a computer, it is desirable to know whether an accurate solution has been obtained or whether the result of a calculation is contaminated by rounding error to such an extent that it is worthless. This goal has been achieved in some areas. Martin, Peters, and Wilkinson [27], in their paper giving an algorithm for solving $Ax = b$, where A is an n x n positive definite matrix, state that their procedure "either produces the correctly rounded solutions of the equation $Ax = b$ or indicates that A is too ill-conditioned for this to be achieved without working to higher precision (or is possibly singular)." Similarly, Wilkinson's program [37] for the solution of an ill-conditioned n x n system of equations $Ax = b$, "gives either a solution of the system which is correct to working accuracy or alternatively indicates that the system is too ill-conditioned to be solved without working to higher precision and may even be singular."

It appears that the goal set out above has now been achieved in the linear least squares programs of Björck and Golub [6] and Björck [5]. The authors state that their procedures may be used to compute accurate solutions and residuals to linear least squares problems, but that the procedures will fail when X modified by rounding errors has less than full rank, and that they will also fail if X is so ill-conditioned that there is no perceptible improvement in the iterative refinement. The user is easily informed of these situations.

ACKNOWLEDGMENTS. I would like to express my appreciation to Joseph M. Cameron who suggested this investigation and made many valuable contributions, to Joan R. Rosenblatt for helpful discussions, and to Joseph Hilsenrath, Russell A. Kirsch and Thomas Hoover for use of their time-shared computer facilities to run several problems. Thanks are due to Gene H. Golub, Stanford University, for his constructive remarks.

REFERENCES

[1] Anscombe, F. J., Topics in the investigation of linear relations fitted by the method of least squares, Journal of the Royal Statistical Society, Series B, 29, p. 1-52, 1967.

[2] Bauer, F. L., Elimination with weighted row combinations for solving linear equations and least squares problems, Numerische Mathematik, 7, p. 338-352, 1965.

[3] Björck, Ake, Solving linear least squares problems by Gram-Schmidt orthogonalization, BIT (Nordisk tidskrift for informations-behandling), 7, p. 1-21, 1967.

[4] Björck, Ake, Iterative refinement of linear least squares solutions, I, BIT, 7, p. 257-278, 1967.

[5] Björck, Ake, Iterative refinement of linear least squares solutions, II, BIT, 8, p. 8-30, 1968.

[6] Björck, Ake and Gene Golub, ALGOL Programming, Contribution no. 22: Iterative refinement of linear least square solutions by Householder transformation, BIT, 7, p. 322-337, 1967.

[7] Bright, J. W. and G. S. Dawkins, Some aspects of curve fitting using orthogonal polynomials, Industrial and Engineering Chemistry Fundamentals, 4, p. 93-97, 1965.

[8] Businger, Peter and Gene H. Golub, Linear least squares solutions by Householder transformations, Numerische Mathematik, 7, p. 269-276, 1965.

[9] Cameron, J. M., Some examples of the use of high speed computers in statistics, Proceedings of the First Conference on the Design of Experiments in Army Research, Development and Testing, Office of Ordnance Research, Report No. 57-1, p. 129-135, 1957.

[10] C-E-I-R Multi-Access Computer Services, Library programs documentation, MAC 71-7-1, Nov. 17, 1967. Addendum, MAC 71-7-1, A 12-368, Mar. 1, 1968.

[11] Crisman, P. A. (ed.), The compatible time-sharing system: A programmer's guide, second edition, M. I. T. Computation Center, M. I. T. Press, Massachusetts Institute of Technology, Cambridge, Massachusetts, 1965.

[12] Davis, Philip J., Orthonormalizing codes in numerical analysis, Chapter 10 in [32], 1962.

[13] Davis, Philip J. and Philip Rabinowitz, Advances in orthonormalizing computation, in Franz L. Alt (ed.), Advances in Computers, vol. 2, Academic Press, New York, p. 55-133, 1961.

[14] Dixon, W. J. (ed.), BMD Biomedical Computer Programs, Health Sciences Computing Facility, University of California, Los Angeles, 1964. Revised 1965 and 1967.

[15] Efroymson, M. A., Multiple regression analysis, in Anthony Ralston and Herbert S. Wilf (editors), Mathematical Methods for Digital Computers, vol. 1, John Wiley and Sons, New York, Ch. 17, p. 191-203, 1960.

[16] Fettis, Henry E. and James C. Caslin, Eigenvalues and eigenvectors of Hilbert matrices of order 3 through 10, Mathematics of Computation, 21, p. 431-441, 1967.

[17] Forsythe, George E., Generation and use of orthogonal polynomials for data-fitting with a digital computer, Journal of the Society for Industrial and Applied Mathematics, 5, p. 74-88, 1957.

[18] Freund, R. J., A warning of roundoff errors in regression, The American Statistician, 17, p. 13-15, 1963.

[19] Golub, G., Numerical methods for solving linear least squares problems, Numerische Mathematik, 7, p. 206-216, 1965.

[20] Golub, G. H. and J. H. Wilkinson, Note on the iterative refinement of least squares solutions, Numerische Mathematik, 9, p. 139-148, 1966.

[21] Hilsenrath, Joseph, Guy G. Ziegler, Carla G. Messina, Philip J. Walsh, and Robert Herbold, OMNITAB, A Computer Program for Statistical and Numerical Analysis, National Bureau of Standards Handbook 101, U. S. Government Printing Office, Washington, D. C., 1966. Reissued January 1968, with corrections.

[22] IBM Application Program, System/360 Scientific Subroutine Package (360A-CM-03X) Version III, Programmer's Manual, H20-0205-3, International Business Machines Corporation, 1968.

[23] IBM Systems Reference Library, Catalog of programs for IBM 704-709-7040-7044-7090 and 7094 Data Processing Systems, Form C20-1604-3, International Business Machines Corporation, December, 1965.

[24] Jordan, T. L., Experiments on error growth associated with some linear least-squares procedures, Mathematics of Computation, 22, p. 579-588, 1968.

[25] Longley, James W., An appraisal of least squares programs for the electronic computer from the point of view of the user, Journal of the American Statistical Association, 62, p. 819-841, 1967.

[26] Longley, James W., personal communication, 1968.

[27] Martin, R. S., G. Peters, and J. H. Wilkinson, Iterative refinement of the solution of a positive definite system of equations, Numerische Mathematik, 8, p. 203-216, 1966.

[28] Miller, James R., On-line analysis for social scientists, MAC-TR-40, Project MAC, Massachusetts Institute of Technology, Cambridge, Massachusetts, 1967.

[29] Newman, Morris, Matrix computations, Chapter 6 in [32], 1962.

[30] Newman, Morris, Solving equations exactly, Journal of Research of the National Bureau of Standards, 71B (Mathematics and Mathematical Physics), p. 171-179, 1967.

[31] Rhomberg, Rudolf R., Lorette Boissonneault, and Leonard Harris, A general purpose computer program for data processing and multiple regression (DAM), Revision 3, Research and Statistics Department, International Monetary Fund, Washington, D. C., September, 1965.

[32] Todd, John (ed.), Survey of Numerical Analysis, McGraw-Hill Book Company, New York, 1962.

[33] Univac 1108 Multi-Processor System, MATH-PACK Programmers Reference, UP-7542, Univac Division of Sperry Rand Corporation, 1967.

[34] Univac 1108 Multi-Processor System, STAT-PACK Programmers Reference, UP-7502, Univac Division of Sperry Rand Corporation, 1967.

[35] Walsh, Philip J., Algorithm 127, ORTHO, Communications of the ACM, 5, p. 511-513, 1962.

[36] Wampler, Roy H., An evaluation of linear least squares computer programs, Journal of Research of the National Bureau of Standards, 73B (Mathematical Sciences), p. 59-90, 1969.

[37] Wilkinson, J. H., The solution of ill-conditioned linear equations, in Anthony Ralston and Herbert S. Wilf (editors), Mathematical Methods for Digital Computers, vol. 2, John Wiley and Sons, New York, Ch. 3, p. 65-93, 1967.

[38] Zellner, A. and H. Thornber, Computational accuracy and estimation of simultaneous equation econometric models, Econometrica, 34, p. 727-729, 1966.

TABLE 1 SUMMARY OF PROGRAMS RUN IN SINGLE PRECISION - 8 Digits

Program	Computer	Algorithm*	Average Number of Correct Digits Y1	Y2	Rank Y1	Y2
ALSQ	1108	HT	4.098	5.368	4	6
BJORCK-GOLUB	1108	HT	4.393	5.950	2	3
BMD02R	1108	E	-0.106	1.981	13	15
BMD03R	7094	E	0.742	1.721	9	17
BMD03R	1108	E	-0.123	2.287	14	13
DAM	7094	E	1.389	2.312	8	12
DAM	1108	E	-0.264	2.622	17	10
LINFIT (Miller)	7094	?	-2.756	-0.301	19	19
LSTSQ	1108	HT	4.528	5.840	1	4
MATH-PACK, ORTHLS	1108	OP	2.118	4.363	7	7
MPR3	7094	E	-0.140	1.856	15	16
OMNITAB (Ortho)	7094	GS	3.954	5.968	5	2
OMNITAB (Ortho)	1108	GS	4.137	5.464	3	5
ORTHO (no iteration)	1108	GS	-1.976	0.419	18	18
ORTHOL	1108	GS	3.593	6.197	6	1
POLRG	1108	E	-0.191	2.280	16	14
STAT-PACK, GLH	1108	E	0.066	2.767	11½	8½
STAT-PACK, REBSOM	1108	E	0.066	2.767	11½	8½
STAT-PACK, RESTEM	1108	E	0.651	2.407	10	11
WRAP	7094	E	-5.300	-2.871	20	20

*E = Elimination method; GS = Gram-Schmidt orthonormalization; HT = Orthogonal Householder transformations; OP = Orthogonal polynomials.

TABLE 2 SUMMARY OF PROGRAMS RUN IN SINGLE PRECISION - 9 Digits

Program	Computer	Algorithm*	Average Number of Correct Digits Y1	Y2	Rank Y1	Y2
LINFIT***	235	E	0.905	2.894	4	5
LSCF---***	235	E	0.308	2.483	6	6
LSFITW***	235	GS	4.102	6.354	1	1
POLFIT	235	OP	3.349	5.922	2	2
STAT20***	235	E	0.612	2.920	5	4
STAT21***	235	E	1.169	3.183	3	3

*E = Elimination method; GS = Gram-Schmidt orthonormalization;
OP = Orthogonal polynomials.

TABLE 3 SUMMARY OF PROGRAMS RUN IN DOUBLE PRECISION - 18 Digits

Program	Computer	Algorithm*	Average Number of Correct Digits Y1	Y2	Rank Y1	Y2
ALSQ	1108	HT	12.667	15.322	5	5
BJORCK-GOLUB	1108	HT	13.580	17.057	2	1
BMD02R	1108	E	9.645	12.865	7	7
BMD05R	1108	E	9.368	11.791	9	10
LSTSQ	1108	HT	14.643	16.293	1	2
MATH-PACK, ORTHLS	1108	OP	12.098	14.461	6	6
ORTHO	1108	GS	13.188	15.514	4	4
ORTHO (no iteration)	1108	GS	7.963	10.354	11	11
ORTHOL	1108	GS	13.212	15.604	3	3
POLRG	1108	E	9.290	11.806	10	9
STAT-PACK, RESTEM	1108	E	9.494	12.019	8	8

TABLE 4 SUMMARY OF PROGRAMS RUN IN SINGLE PRECISION (8 Digits) WITH INNER PRODUCTS ACCUMULATED IN DOUBLE PRECISION (18 Digits)

Program	Computer	Algorithm*	Average Number of Correct Digits Y1	Y2	Rank Y1	Y2
ALSQ	1108	HT	3.506	6.530	4	1
BJORCK-GOLUB	1108	HT	8.000	6.227	1½	4
LSTSQ	1108	HT	8.000	6.279	1½	3
ORTHO	1108	GS	3.904	6.459	3	2

*E = Elimination method; GS = Gram-Schmidt orthonormalization; HT = Orthogonal Householder transformations; OP = Orthogonal polynomials.

TABLE 5 INPUT FOR FIFTH DEGREE POLYNOMIALS

X	Y1	Y2
0.	1.	1.00000
1.	6.	1.11111
2.	63.	1.24992
3.	364.	1.42753
4.	1365.	1.65984
5.	3906.	1.96875
6.	9331.	2.38336
7.	19608.	2.94117
8.	37449.	3.68928
9.	66430.	4.68559
10.	111111.	6.00000
11.	177156.	7.71561
12.	271453.	9.92992
13.	402234.	12.75603
14.	579195.	16.32384
15.	813616.	20.78125
16.	1118481.	26.29536
17.	1508598.	33.05367
18.	2000719.	41.26528
19.	2613660.	51.16209
20.	3368421.	63.00000

MATRIX X'X ASSOCIATED WITH THE TEST PROBLEMS

21.	210.	2870.	44100.	722666.	12333300.
210.	2870.	44100.	722666.	12333300.	216455810.
2870.	44100.	722666.	12333300.	216455810.	3877286700.
44100.	722666.	12333300.	216455810.	3877286700.	70540730666.
722666.	12333300.	216455810.	3877286700.	70540730666.	129915527990.
12333300.	216455810.	3877286700.	70540730666.	129915527990.	2416357168085O.

MATRIX X'Y FOR Y1	MATRIX X'Y FOR Y2
13103167.	310.39960
229558956.	5058.55410
4106845446.	87258.40800
74647573242.	1549291.38666
1373802809082.	28043466.66600
25537373767266.	514843723.46850

APPENDIX A

SOURCES OF THE PROGRAMS, WITH BRIEF DESCRIPTIONS

ALSQ. A FORTRAN IV subroutine to solve the linear least squares problem, written by G. W. Stewart, III, Union Carbide Corp., Oak Ridge, Tennessee (present address: University of Texas, Austin, Texas). This program uses a modification of the Businger-Golub algorithm [8].

BJORCK-GOLUB. A FORTRAN V program to solve the linear least squares problem, written by Roy H. Wampler, National Bureau of Standards, using the Björck-Golub algorithm [6].

BMD02R, Stepwise Regression. One of the Biomedical Computer Programs, written in FORTRAN [14].

BMD03R, Multiple Regression with Case Combinations. One of the Biomedical Computer Programs, written in FORTRAN [14].

BMD05R, Polynomial Regression. One of the Biomedical Computer Programs, written in FORTRAN [14].

DAM. A general purpose computer program for data processing and multiple regression, written in FORTRAN by Rudolf R. Rhomberg, Lorette Boissonneault, and Leonard Harris, International Monetary Fund [31].

LINFIT. A program which fits a linear function to collected data via least squares. Optional constraints may be applied to the fitting coefficients to make them non-negative, add to a constant, etc. One of eighteen statistical routines written by James R. Miller [28]. This library of routines exists in the Project MAC 7094 in the disk files of user number T169 2750.

LINFIT***. A program written in BASIC for linear least squares curve fitting and computing correlations. Origin: Dartmouth College, Hanover, N. H. Available in the C-E-I-R Multi-Access Computer Services library [10].

LSCF--***. A least squares polynomial curve fitting subroutine written in BASIC. Origin: Dartmouth College, Hanover, N. H. Available in the C-E-I-R Multi-Access Computer Services library [10].

LSFITW***. A least squares curve fitting program written in BASIC. Adapted by John B. Shumaker, National Bureau of Standards, from Philip J. Walsh's ORTHO algorithm [35]. Available in the C-E-I-R Multi-Access Computer Services library [10].

LSTSQ. A FORTRAN IV subroutine which solves for X the overdetermined system AX = B of m linear equations in n unknowns for p right-hand sides. Written by Peter Businger, Computation Center, University of Texas (present address: Bell Telephone Laboratories, Murray Hill, N. J.), using the Businger-Golub algorithm [8].

MATH-PACK, ORTHLS, Orthogonal Polynomial Least-Squares Curve Fitting. One of the Univac 1108 MATH-PACK programs, written in FORTRAN V [33].

MPR3, Stepwise Multiple Regression with Variable Transformations. A FORTRAN II program written by M. A. Efroymson, Esso Research and Engineering Co., Madison, N. J., using the Efroymson algorithm [15]. Available in the SHARE library: 7090-G2 3145MPR3 [23].

OMNITAB, a general-purpose computer program for statistical and numerical analysis. Developed at the National Bureau of Standards by Joseph Hilsenrath et al [21]. Now available in an A. S. A. FORTRAN version, OMNITAB allows the user to communicate with a computer in an efficient manner by means of simple English sentences.

ORTHO. A program written by Philip J. Walsh, National Bureau of Standards (present address: University Computing Co., East Brunswick, N. J.), which uses a Gram-Schmidt orthonormalization process for least squares curve fitting. ORTHO exists as an ALGOL procedure [35], a FORTRAN program, a BASIC program (see LSFITW*** above), and as a routine of OMNITAB [21], where it is called by the commands FIT and POLYFIT.

ORTHOL. A modification of the Davis-Rabinowitz orthonormalization algorithm [12], [13], written in FORTRAN II by James W. Longley, Bureau of Labor Statistics, Washington, D. C., and Roger A. Blau, Bureau of Labor Statistics and Carnegie-Mellon University, Pittsburgh, Pa. [26].

POLFIT. An anonymous program written in BASIC for least squares polynomial curve fitting using orthogonal polynomials.

POLRG, Polynomial Regression. One of the programs of the IBM System/360 Scientific Subroutine Package written in FORTRAN IV [22].

STAT-PACK, GLH, General Linear Hypotheses. One of the Univac 1108 STAT-PACK programs, written in FORTRAN V [34].

STAT-PACK, REBSOM, Back Solution Multiple Regression. One of the Univac 1108 STAT-PACK programs, written in FORTRAN V [34].

STAT-PACK, RESTEM, Stepwise Multiple Regression. One of the Univac 1108 STAT-PACK programs, written in FORTRAN V [34].

STAT20***. A program written in BASIC for stepwise multiple linear regression. Written by Thomas E. Kurtz, Dartmouth College, Hanover, N. H. Available in the C-E-I-R Multi-Access Computer Services library [10].

STAT21***. A program written in BASIC for multiple linear regression with detailed output. Written by Gerald Childs, Dartmouth College, Hanover, N. H. Available in the C-E-I-R Multi-Access Computer Services library [10].

WRAP, Weighted Regression Analysis Program. A FORTRAN II program written by M. D. Fimple, Sandia Corp., Albuquerque, New Mexico. Available in the SHARE library: 7090-G2 3231WRAP [23].

ERROR ANALYSIS FOR CONTROL SYSTEMS

T. H. Slook
Temple University and Frankford Arsenal
Philadelphia, Pennsylvania

I. INTRODUCTION. From the days following World War II to the present time, many research papers and books have been written on feedback control systems. In almost every case, these publications emphasize the analysis and design of such systems. Relatively few pages have been devoted to error analysis techniques for control systems. The important contributions which this paper makes are:

A. To exhibit an error analysis technique for an arbitrary control system; and,

B. To prove, in a general setting, three theorems relating the variances and power spectral densities of the inputs and outputs of such systems.

II. MEASURES OF EFFECTIVENESS. Every measure of effectiveness for a control system involves, either directly or indirectly, some knowledge of system errors. To demonstrate this point and to make this paper more meaningful and less abstract, let us consider a fire control system (FCS). Such a control system includes tracking servos, data transmission devices, conversion elements, analog and/or digital computing components and weapon pointing servos, each, of which, possesses errors and contributes to the overall system output errors. Clearly, the magnitude and frequency of the output errors determine the control system's effectiveness.

Two of the many measures of effectiveness for a FCS are hit probability and kill probability. To be specific, the single shot engagement hit probability is obtained by evaluating

$$P = 1 - \int_o^1 \left[1 - (A/(2\pi \cdot \sigma_d^2)) \times {\sigma_b^2}/{\sigma_d^2}\right]^n dx \qquad [II\text{-}1]$$

where

n = the number of rounds for an engagement,

A = target area,

σ_b^2 = variance of the bias, and

σ_d^2 = variance of the dispersion.

In this paper, the bias b is the deviation of the center of impact of n rounds from the target center, and the dispersion d is the square root of the average value of the square of the deviations of the rounds from the center of impact.

Observe that P, defined above, is a function of σ_b^2 and σ_d^2. These variances, whether used to calculate P or any other measure of effectiveness, depend upon the variance in the error in the elevation $\sigma_{\varepsilon_E}^2$ and the variance in the error in train $\sigma_{\varepsilon_T}^2$ of the gun tube or launcher throughout the firing interval and each of these statistical measures depend upon:

a. errors in the inputs to the control system,
b. non-ideal system element errors,
c. system function approximations, and
d. vehicle-target paths.

Let us agree to call the above error sources the system input errors for a FCS. Observe that (a), (b), and (c) are system input errors for every control system, and that (d) is an additional error source that must be considered in a FCS error analysis.

The fact that every control system consists of an assemblage of a finite number of components, each of which has measurable characteristics, generates, in a natural way, a finite number of equations relating the inputs and outputs of the control system. These equations are called the system equations. Some of the system equations may be empirical. For example, the ballistic functions are empirical equations in a FCS.

A relatively easy and straightforward error analysis is possible when the system equations are not differential equations. However, many control systems and most FCS generate an independent set of differential equations. The inclusion of differential equations complicates the solution of the system of error equations. This we now explain.

III. SYSTEM ERROR EQUATIONS. Consider a FCS of q system elements having s independent inputs. This means that at each instant of time, every system element will have at most $\underline{X} = \{x_1, x_2, \ldots, x_s\}$ inputs from outside the FCS and at most $\underline{Y} = \{y_1, y_2, \ldots, y_q\}$ inputs from within the FCS; see Figure 1. Observe that the external inputs x_j (j = 1, 2, ..., s) and the system element outputs y_k (k = 1, 2, ..., q) are functions of time and it is customary to assume that these inputs and outputs have continuous first derivatives.

In Figure 1, the output at time t of the i^{th} system element having $\{x_{i_1}, x_{i_2}, \ldots x_{i_{r_i}}\}$, a subset of $\underline{X}$, as external inputs and

$\{y_{i_1}, y_{i_2}, \ldots y_i, \ldots y_{i_{p_i}}\}$, a subset of $\overline{Y}$, as internal inputs is

described by

$$y_i = g_i(x_{i_1}, x_{i_2}, \ldots, x_{i_{r_i}}, y_{i_1}, y_{i_2}, \ldots, y_i, \ldots, y_{i_{p_i}}). \qquad [III-1]$$

Those x's and y's which are not inputs to the i^{th} system element are not in the domain of g_i. The function g_i is called the performance operator of the i^{th} system element, and it determines the output y_i of this system element. Figure 1 shows that the output of the i^{th} system element is also an input and for a feedback loop, we prefer to write the performance equation in the implicit form,

$$f(x_{i_1}, x_{i_2}, \ldots, x_{i_{r_i}}, y_{i_1}, y_{i_2}, \ldots, y_i, \ldots, y_{i_{p_i}}) = 0. \qquad [III-2]$$

The only change in the performance equation of the i^{th} system element for a non-feedback loop would be the deletion of y_i as an input variable in equation [III-1].

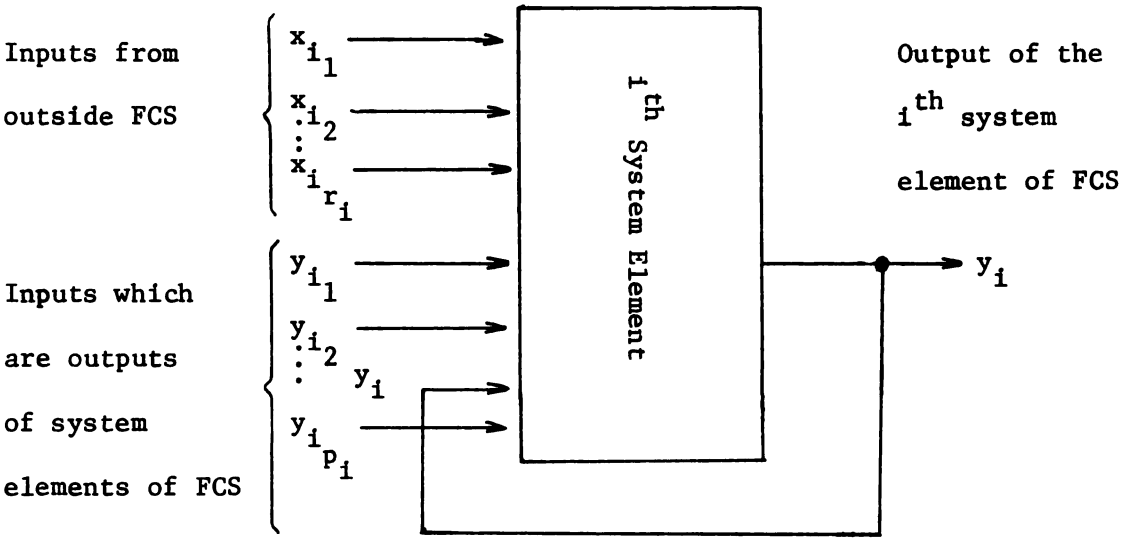


FIGURE 1

In practice, each input to the i^{th} system element may possess an error and we denote the error by $\varepsilon_{x_{i_j}}$ $(j = 1, 2, \ldots, r_i)$ and $\varepsilon_{y_{i_k}}$ $(k = 1, 2, \ldots, p_i)$. Since the system element is not an ideal element, the output of this imperfect element is the correct output y_i plus the system element error m_i. Each of these errors may also be considered as functions of time; see Figure 2.

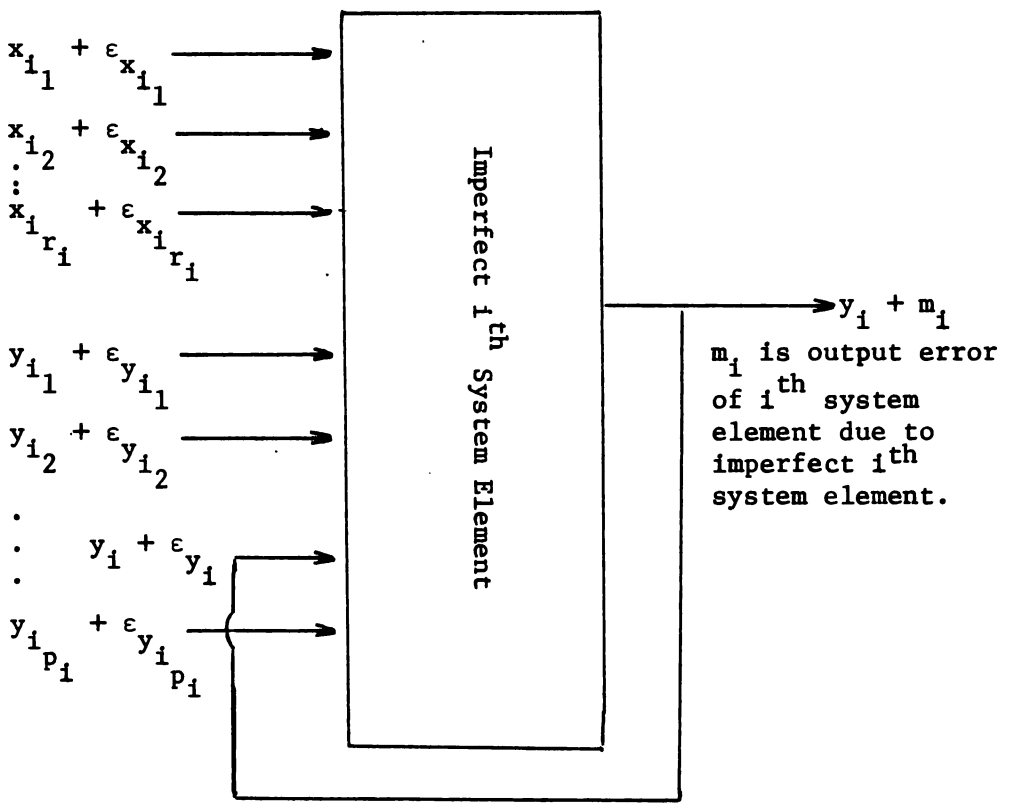


FIGURE 2

Since performance operators are smooth functions, then each g_i (= 1, 2,..., q) possesses continuous first partial derivatives with respect to the external and internal inputs. This implies that every f_i possesses this property. Hence, the change in f_i, produced by increment changes $\varepsilon_{x_{i_j}}$, $\varepsilon_{y_{k_i}}$, and m_i is

$$\Delta f_i = f_i(Q_i) - f_i(P_i)$$

$$= \sum_{j=1}^{r_i} \left.\frac{\partial f_i}{\partial x_{i_j}}\right|_{P_i} \cdot \varepsilon_{x_{i_j}} + \sum_{k=1}^{p_i} \left.\frac{\partial f_i}{\partial y_{i_k}}\right|_{P_i} \cdot \varepsilon_{y_{i_k}} + \left.\frac{\partial f_i}{\partial y_i}\right|_{P_i} \cdot m_i \qquad \text{[III-3]}$$

where $P_i = (x_{i_1}, \ldots, x_{i_{r_i}}, y_{i_1}, \ldots, y_i, \ldots, y_{i_{p_i}})$ and

$$Q_i = (x_{i_1} + \varepsilon_{x_{i_1}}, \ldots, x_{i_{r_i}} + \varepsilon_{x_{i_{r_i}}}, y_{i_1} + \varepsilon_{y_{i_1}},$$

$$\ldots, y_i + \varepsilon_{y_i} + m_i, \ldots, y_{i_{p_i}} + \varepsilon_{y_{i_{p_i}}}).$$

The points P_i and Q_i are in the domain of f_i, thus $\Delta f_i = 0$ and equation [III-3] becomes

$$\sum_{k=1}^{p_i} \left.\frac{\partial f_i}{\partial y_{i_k}}\right|_{P_i} \cdot \varepsilon_{y_{i_k}} = - \sum_{j=1}^{r_i} \left.\frac{\partial f_i}{\partial x_{i_j}}\right|_{P_i} \cdot \varepsilon_{x_{i_j}} - \left.\frac{\partial f_i}{\partial y_i}\right|_{P_i} \cdot m_i \qquad \text{[III-4]}$$

This is the error equation for the i^{th} system element, describable by other than a differential equation. Therefore, the set of error equations for a FCS with s external inputs and q system elements describably by other than differential euqations is the linear system of equations $A\,\varepsilon_y = B$. This matrix equation we prefer to write:

$$\left\{\begin{array}{l}
\displaystyle\sum_{k=1}^{p_1} \frac{\partial f_1}{\partial y_{1_k}}\bigg|_{P_1} \cdot \varepsilon_{y_{1_k}} = -\sum_{j=1}^{r_1} \frac{\partial f_1}{\partial x_{1_j}}\bigg|_{P_1} \cdot \varepsilon_{x_{1_j}} - \frac{\partial f_1}{\partial y_1}\bigg|_{P_1} \cdot m_1 \\
\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad [III\text{-}5] \\
\displaystyle\sum_{k=1}^{p_q} \frac{\partial f_q}{\partial y_{q_k}}\bigg|_{P_i} \cdot \varepsilon_{y_{q_k}} = -\sum_{j=1}^{r_q} \frac{\partial f_q}{\partial x_{q_i}}\bigg|_{P_q} \cdot \varepsilon_{x_{q_j}} - \frac{\partial f_q}{\partial y_q}\bigg|_{P_q} \cdot m_q
\end{array}\right.$$

The above technique may be employed to generate:

$$\left\{\begin{array}{l}
\displaystyle\sum_{k=1}^{p_1} \left\{\frac{\partial f_1}{\partial y_{1_k}}\bigg|_{P_1} \cdot \varepsilon_{y_{1_k}} + \frac{\partial f_1}{\partial \dot{y}_{1_k}}\bigg|_{P_1} \cdot \dot{\varepsilon}_{y_{1_k}} + \ldots\right\} = -\frac{\partial f_1}{\partial y_1}\bigg|_{P_1} \cdot m_1 - \\
\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad [III\text{-}6] \\
\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad \displaystyle\sum_{j=1}^{r_1} \left\{\frac{\partial f_1}{\partial x_{1_j}}\bigg|_{P_1} \cdot \varepsilon_{x_{1_j}} + \frac{\partial f_1}{\partial \dot{x}_{1_j}}\bigg|_{P_1} \cdot \dot{\varepsilon}_{x_{1_j}} + \ldots\right\} \\
\displaystyle\sum_{k=1}^{p_q} \left\{\frac{\partial f_q}{\partial y_{q_k}}\bigg|_{P_q} \cdot \varepsilon_{y_{q_k}} + \frac{\partial f_q}{\partial \dot{y}_{q_k}}\bigg|_{P_q} \cdot \dot{\varepsilon}_{y_{q_k}} + \ldots\right\} = -\frac{\partial f_q}{\partial y_q}\bigg|_{P_q} \cdot m_q - \\
\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad \displaystyle\sum_{j=1}^{r_q} \left\{\frac{\partial f_q}{\partial x_{q_j}}\bigg|_{P_q} \cdot \varepsilon_{x_{q_j}} + \frac{\partial f_q}{\partial \dot{x}_{q_j}}\bigg|_{P_q} \cdot \varepsilon_{x_{q_j}} + \ldots\right\}
\end{array}\right.$$

which is the set of error equations for a FCS with s external inputs and q system elements describable by differential equations.

Using [III-5] or [III-6] and a given set of system input errors, one can determine the system output error vector

$$\varepsilon_y = (\varepsilon_{y_1}, \varepsilon_{y_2}, \ldots, \varepsilon_{y_q})^t,$$

providing the given set of system equations can be solved. Observe that the coefficients in both systems of equations are functions of the arbitrary but fixed points P_i. Thus, [III-5] is easily solved for ε_y, but [III-6] is not easily solved for ε_y when one or more infinite series expansions occur. It is not the purpose of this paper to discuss conditions for a solution to [III-6] because the external input errors for a control system are given as variances $\sigma_{x_i}^2$ and not as ε_{x_i} $(i = 1, 2, \ldots, q)$. Hence, the main problem is to express the variance in the output errors as functions of the variances in the system input errors. For a FCS this means: express the variance in elevation error $\sigma_{\varepsilon_E}^2$ and the variance in train error $\sigma_{\varepsilon_T}^2$ as functions of the variances in system input errors. This we now discuss.

IV. STATISTICAL MEASURE OF OUTPUTS. To express $\sigma_{\varepsilon_E}^2$ and $\sigma_{\varepsilon_T}^2$ as functions of the variances in the system input errors require that we prove several remarkable theorems. One may omit the proofs if he so desires, because the theorems are proved only for the sake of completeness.

Let $L'(\mu)$ be the Banach space of summable functions defined on $\overline{X} = \{t: -\infty < t < +\infty\}$ with μ as Lebesque measure and $||x||_1 = \int_{\overline{X}} |x|\,dx$. The following theorem exhibits a relationship between the derivative of the variance in the variable x with respect to frequency and its power spectral density.

Theorem 1. Let $x \in L'(\mu)$. Then

$$\frac{d}{d\omega}\{\sigma_x^2(\omega)\} = \frac{1}{2\pi}\Phi_{xx}(\omega) - \frac{d}{d\omega}\{(\widetilde{x(\omega)})^2\} \qquad [IV\text{-}1]$$

where

a) $\sigma_x^2(\omega)$ is the variance of x,

b) $\Phi_{xx}(\omega)$ is the power spectral density of x,

c) $\widetilde{x(\omega)}$ is the mean of x.

Proof: For arbitrary $x \in L'$, the autocorrelation function of x

$$\phi_{xx}(\tau) = \lim_{T\to\infty} \frac{1}{2T} \int_{-T}^{T} x(t)\, x(t+\tau)dt, \quad -\infty < \tau < \infty$$

exists, and for $\tau = 0$ the autocorrelation function reduces to

$$\phi_{xx}(0) = \lim_{T\to\infty} \frac{1}{2T} \int_{-T}^{T} x^2(t)dt = \widetilde{x^2},$$

the expected value of x^2. Since

$$\phi_{xx}(\tau) = \frac{1}{2\pi} \lim_{T\to\infty} \int_{-T}^{T} \Phi_{xx}(\omega)\ e^{j\omega\tau}\ .\ d\omega\ , -\infty < \tau < \infty$$

then for $\tau = 0$

$$\phi_{xx}(0) = \frac{1}{2\pi} \int_{\infty}^{\infty} \Phi_{xx}(\omega)\ d\omega\ .$$

Thus, we may conclude that

$$\widetilde{x^2} = \frac{1}{2\pi} \int_{\infty}^{\infty} \Phi_{xx}(\omega)\ d\omega\ . \qquad \text{[IV-2]}$$

Observe that $\widetilde{x^2}$ is constant. However, equation [IV-2] permits us to define $\widetilde{x^2(\omega)}$ as follows:

$$\widetilde{x^2(\omega)} = \frac{1}{2\pi} \int_{\infty}^{\omega} \Phi_{xx}(\mu)\ d\mu\ , \qquad -\infty < \omega < \infty$$

This implies that

$$\frac{d}{d\omega}\{\widetilde{x^2(\omega)}\} = \frac{1}{2\pi} \Phi_{xx}(\omega)$$

for all real ω. Using the well-known statistical equation

$$\sigma^2 = \widetilde{x^2} - (\widetilde{x})^2$$

we obtain

$$\frac{d}{d\omega}\{\sigma_x^2(\omega)\} = \frac{d}{d\omega}\{\widetilde{x^2(\omega)}\} - \frac{d}{d\omega}\{(\widetilde{x(\omega)})^2\}$$

$$= \frac{1}{2\pi}\Phi_{xx}(\omega) - \frac{d}{d\omega}\{(\widetilde{x(\omega)})^2\}$$

where

$$\widetilde{x(\omega)} = \int_{-\infty}^{\omega} x(t)dt, \qquad -\infty < \omega < +\infty .$$

In the above theorem, x may be thought of as an input to a system element having response r. The output y for this system element, in its most general form, is given by the convolution integral

$$y(t) = \int_{-\infty}^{\infty} x(z)r(t - z)\, dz, \qquad -\infty < t < +\infty.$$

Theorem 2 gives the relationship between the power spectral density of x and the power spectral density of y.

Theorem 2. If r and x belong to $L'(\mu)$, then the power spectral density of y is given by

$$\Phi_{yy}(\omega) = |R(\omega)|^2 \cdot \Phi_{xx}(\omega) \qquad [IV-3]$$

where $R(\omega)$ is the transfer function of the system element.*

Proof. Since the output y is defined by a convolution function whose determining functions r and x belong to $L'(\mu)$, then $y \in L'(\mu)$ and the autocorrelation function of y

---

*It was brought to my attention by a member of the audience that this property was known to Norbert Weiner. However, it should be mentioned that in August of 1958, the team of Tappert, Pfeilsticker, and Slook, having no knowledge of Weiner's result, proved this property in two entirely different ways.

$$\phi_{yy}(\tau) = \lim_{T\to\infty} \frac{1}{2T} \int_{-T}^{T} y(t)\ y(t+\tau)\ dt$$

exists. Replacing $y(t)$ and $y(t+\tau)$ by their respective convolution integrals we obtain

$$\phi_{yy}(\tau) = \lim_{T\to\infty} \frac{1}{2T} \int_{-T}^{T} \{ \int_{\infty}^{\infty} x(u)\ r(t-u)du \} \{ \int_{\infty}^{\infty} x(v) r(t+\tau - v)dv \} dt$$

Let $-v = t - u$ and $-\rho = t + \tau - v$, then the above equation becomes

$$\phi_{yy}(\tau) = \lim_{T\to\infty} \frac{1}{2T} \int_{-T}^{T} \{ \int_{\infty}^{\infty} x(t+v)\ r(-v)dv \} \{ \int_{\infty}^{\infty} x(t+\tau+\rho) r(-\rho)\ d\rho \} dt$$

$$= \int_{\infty}^{\infty} \int_{\infty}^{\infty} r(-v)\ r(-\rho) \{ \lim_{T\to\infty} \frac{1}{2T} \int_{-T}^{T} x(t+v)\ x(t+\tau+\rho)dt \} dvd\rho \qquad [IV-4]$$

$$= \int_{\infty}^{\infty} \int_{\infty}^{\infty} r(-v)\ r(-\rho) \{ \lim_{T\to\infty} \frac{1}{2T} \int_{-T+v}^{T+v} x(u)\ x(u+\tau+\rho-v)\ du \} dvd\rho$$

$$= \int_{\infty}^{\infty} \int_{\infty}^{\infty} r(-v)\ r(-\rho)\ \phi_{xx}(\tau+\rho-v)\ dvd\rho\ .$$

The change in the order of integration is possible because the conditions of the Fubini theorem are satisfied.

The Fourier integral of $\phi_{yy}(\omega)$ defines the power spectral density of y, that is

$$\Phi_{yy}(\omega) = \int_{\infty}^{\infty} \phi_{yy}(\tau)\ e^{-j\omega\tau}\ d\tau$$

In this integral, replace $\phi_{yy}(\tau)$ by the last relation described in equation [IV-4]. Thus $\Phi_{yy}(\omega)$ becomes

$$\Phi_{yy}(\omega) = \int_{\infty}^{\infty} \{ \int_{\infty}^{\infty} \int_{\infty}^{\infty} r(-v)\ r(-\rho)\ \phi_{xx}(\tau+\rho-v)\ dv\ d\rho \}\ e^{-j\omega\tau}\ d\tau$$

$$= \int_{-\infty}^{\infty} \{\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} r(-v)\ e^{j\omega v}\ r(-v)\ e^{-j\omega v}\ \phi_{xx}(\tau + \rho - v)\ dv\ d\rho\}\ e^{-j\omega(\tau+\rho-v)}\ d\tau$$

$$= \Phi_{xx}(\omega)\{\int_{-\infty}^{\infty} r(-v)\ e^{j\omega v}\ dv\}\ \{\int_{-\infty}^{\infty} r(-\rho)\ e^{-j\omega\rho}\ \rho\}\ .$$

The change in order of integration given above is possible as the conditions of the Fubini theorem hold.

The Fourier transform of the response r,

$$R(\omega) = \int_{-\infty}^{\infty} r(v)\ e^{-j\omega v}\ dv$$

is called the transfer function of the system element. Substituting for the integral forms in equation [IV-4] their equivalent functions $R(\omega)$ and $\overline{R(\omega)}$, one obtains the desired functional relationship.

$$\Phi_{yy}(\omega) = R(\omega)\ \overline{R(\omega)}\ \Phi_{xx}(\omega)$$

$$= |R(\omega)|^2\ \Phi_{xx}(\omega)\ .$$

As a consequence of theorems 1 and 2, we obtain a useful relationship between the variance of the input and the variance of the output of a system element. This result we embody in the following theorem.

Theorem 3. Let x be the input to a system element having response r and output y. If r and x belong to $L'(\mu)$ and $\widetilde{x(\omega)} = \widetilde{y(\omega)}$, then

$$\sigma_y^2 = \int_{-\infty}^{\infty} |R(\omega)|^2\ \frac{d\{\sigma_x^2(\omega)\}}{d\omega}\ d\omega$$

Proof: Combining the relationships of theorems (1) and (2), we obtain

$$2\pi\left\{\frac{d\{\sigma_y^2(\omega)\}}{d\omega} + \frac{d\{(\widetilde{y(\omega)})^2\}}{d\omega}\right\} = |R(\omega)|^2\left(2\pi\left\{\frac{d\{\sigma_x^2(\omega)\}}{d\omega} + \frac{d\{(\widetilde{x(\omega)})^2\}}{d\omega}\right\}\right)$$

which reduces to

$$\frac{d\{\sigma_y^2(\omega)\}}{d\omega} = |R(\omega)|^2\ \frac{d\{\sigma_x^2(\omega)\}}{d\omega}\ .$$

Hence

$$\sigma_y^2 = \int_{-\infty}^{\infty} \frac{d\{\sigma_y^2(\omega)\}}{d\omega}\, d\omega = \int_{-\infty}^{\infty} |R(\omega)|^2 \frac{d\{\sigma_x^2(\omega)\}}{d\omega}\, d\omega .$$

Theorem 3 established for a system element consisting of a single input and a single output the functional relationship between the variances of these variables. Applying this technique to the system element illustrated in Figure 2 which has $r_i + p_i + 1$ sources of error $\varepsilon_{x_{i_j}}$ $(j = 1, 2, \ldots, r_i)$, $\varepsilon_{y_{i_k}}$ $(k = 1, 2, \ldots, p_i)$ and m_i, we obtain

$$\frac{d\{\sigma_{\varepsilon_{y_i}}^2(\omega) + \sigma_{m_i}^2(\omega)\}}{d\omega} = |R_i(\omega)|^2 \left(\sum_{j=1}^{r_i} \frac{d\{\sigma_{\varepsilon_{x_{i_j}}}^2(\omega)\}}{d\omega} + \sum_{k=1}^{p_i} \frac{d\{\sigma_{\varepsilon_{y_{i_k}}}^2(\omega)\}}{d\omega} \right)$$

which reduces to

$$\frac{d\{\sigma_{\varepsilon_{y_i}}^2(\omega)\}}{d\omega} - |R_i(\omega)|^2 \sum_{k=1}^{p_i} \frac{d\{\sigma_{\varepsilon_{y_{i_k}}}^2(\omega)\}}{d\omega} = |R_i(\omega)|^2 \sum_{j=1}^{r_i} \frac{d\{\sigma_{\varepsilon_{x_{i_j}}}^2(\omega)\}}{d\omega} - \frac{d\{\sigma_{m_i}^2(\omega)\}}{d\omega}$$

Therefore, the set of variance error equations for a FCS with s external inputs and q system elements may be written:

$$\begin{cases} s_1(\omega) \cdot \sum_{k=1}^{p_1} \frac{d\{\sigma_{\varepsilon_{y_{1_k}}}^2(\omega)\}}{d\omega} = |R_1(\omega)|^2 \sum_{j=1}^{r_1} \frac{d\{\sigma_{\varepsilon_{x_{1_j}}}^2(\omega)\}}{d\omega} - \frac{d\{\sigma_{m_1}^2(\omega)\}}{d\omega} \\ \vdots \\ s_q(\omega) \cdot \sum_{k=1}^{p_q} \frac{d\{\sigma_{\varepsilon_{y_{q_k}}}^2(\omega)\}}{d\omega} = |R_q(\omega)|^2 \sum_{j=1}^{r_q} \frac{d\{\sigma_{\varepsilon_{x_{q_j}}}^2(\omega)\}}{d\omega} - \frac{d\{\sigma_{m_q}^2(\omega)\}}{d\omega} \end{cases} \qquad \text{[IV-6]}$$

Solve this system of linear equations in $\frac{d}{d\omega}\{\sigma^2_{\varepsilon_{y_{i_k}}}(\omega)\}$ $(i = 1, 2, \ldots, q)$

for the outputs desired. In the case of a FCS one would solve for $\frac{d}{d\omega}\{\sigma^2_{\varepsilon_E}(\omega)\}$ and $\frac{d}{d\omega}\{\sigma^2_{\varepsilon_T}(\omega)\}$. Thus, the variance in elevation and the variance in train may be calculated by using equations

$$\sigma^2_{\varepsilon_E} = \int_{-\infty}^{\infty} \frac{d}{d\omega}\{\sigma^2_{\varepsilon_E}(\omega)\}\, d\omega$$

$$\sigma^2_{\varepsilon_T} = \int_{-\infty}^{\infty} \frac{d}{d\omega}\{\sigma^2_{\varepsilon_T}(\omega)\}\, d\omega$$

Observe that the above technique provides a means for determining variances of the output errors of a control system describable by differential equations. These variances, as demonstrated in the fire control example, may be used to determine a measure of effectiveness for the control system.

REFERENCES

1. Tappert, J. G., "General Theory of Propagation of Errors in Analog Computers," Tech. Note No. TN-1106, Fkd. Arsenal, Phila., Pa., September 15, 1958.

2. Glynn, J. C., Pfeilsticker, R. C., "Hit Probability of the Vigilante System," vols. I, II, and III, Fkd. Arsenal Report No. 1510-1177-8, Frankford Arsenal, Phila., Pa., August 1959.

3. Lee, Y. W., "Statistical Theory of Communication," John Wiley & Sons, Inc., New York, 1960.

ANALYSIS OF MULTI-DIMENSIONAL CONTINGENCY TABLES

H. H. Ku and R. Varner
National Bureau of Standards
Washington, D. C.

and

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

ABSTRACT. This is an expository paper on the analysis of contingency tables given at the Fourteenth Conference on the Design of Experiments. The principle of minimum discrimination information estimation is described and used to generate estimates for tests of hypotheses concerning second-order and higher-order interactions. All classical hypotheses for contingency tables can be generated by the use of this principle when certain marginals are considered as fixed.

Examples are given and two available computation programs are described in detail.

I. INTRODUCTION. In the January issue of the Journal of Royal Statistical Society, there is a paper by M. G. Kendall (1968) entitled, "On the Future of Statistics - A Second Look." A particular paragraph in his paper concerns the topic under discussion today. We quote:

> 19. It is rather a hazardous task to try to forecast the future lines of development of theoretical statistics, but there seem to me to be two major growing points and I should like to consider them in some detail. The first concerns the bridging of the gap between theory and practical requirements in multivariate analysis. The problems which are encountered in nearly all statistical enquiries concerned with this subject are very far from being solved. I will cite a few examples from what might be a very long list:
>
> (1) Multiple contingency tables. The problems of manifold classification in p dimensions are of three kinds: the pure problem of display so that one can look at the results as a whole; the problem of empty cells, or small frequencies, which are apt to arise on the edges of a table even for large samples; and, perhaps the most difficult of all, a method of analysis which will bring out the various inter-relationships among the classificatory variables.

*Supported in part by the Air Force Office of Scientific Research, Office of Aerospace Research, United States Air Force, under grant AFOSR-68-1513.

We agree with Kendall on both counts: that the problem needs further investigation and the problem is a difficult one. The procedure we present today proposes a unified treatment of multi-dimensional contingency tables, and we believe it to be a step in the right direction.

A. Formulation of the Hypothesis. The formulation of a meaningful hypothesis of no interaction in a multi-way table is not as simple as one might expect at first. For the simplest case beyond a two-way table, a 2 x 2 x 2 table, with modified conventional notations as shown in Figure 1.1, Bartlett (1935) defined "no second-order interaction" as implying

$$(1.1) \qquad \frac{p(111)p(221)}{p(121)p(211)} = \frac{p(112)p(222)}{p(122)p(212)} .$$

Bartlett's Definition

No Second-Order Interaction for a 2 x 2 x 2 Table

	D_1			D_2		
	C_1	C_2		C_1	C_2	
R_1	p(111)	p(121)	p(1.1)	p(112)	p(122)	p(1.2)
R_2	p(211)	p(221)	p(2.1)	p(212)	p(222)	p(2.2)
	p(.11)	p(.21)	p(..1)	p(.12)	p(.22)	p(..2)

$$H_0 : \frac{p(111)p(221)}{p(121)p(211)} = \frac{p(112)p(222)}{p(122)p(212)}$$

Figure 1.1

This definition is essentially an extension of the cross-product ratio definition of independence in a 2 x 2 table. The hypothesis proposed is the equality of association between classifications R and C within D_1 and D_2. How would one go about to extend this formulation to higher dimension tables with more than two categories within each classification? How many relations of the form (1.1) does one need to express the hypothesis of no second-order interaction in such cases? These questions were studied by Roy and Kastenbaum (1955).

B. Computation of Expected Frequencies. Once a null hypothesis is decided upon, the next step is to estimate the expected cell frequencies under the null hypothesis using the marginal frequencies, in the same way as we estimate cell frequencies under the independence hypothesis in an r x c table, using

$$\hat{p}(ij) = \hat{p}(i.)\hat{p}(.j) = \frac{x(i.)}{n} \cdot \frac{x(.j)}{n},$$

where $x(i.) = \sum_j x(ij)$, $x(.j) = \sum_i x(ij)$, $\sum_{ij} x(ij) = n$, and $x(ij)$ is the observed frequency in the ij-th cell. For expression (1.1), Bartlett proposed to solve for Δ in the expression

$$(1.2) \qquad \frac{[x(111) + \Delta][x(221) + \Delta]}{[x(121) - \Delta][x(211) - \Delta]} = \frac{[x(112) - \Delta][x(222) - \Delta]}{[x(122) + \Delta][x(212) + \Delta]}$$

which is a third degree equation in Δ. Note that this implies that the two-way marginals are unchanged. Then a statistic $X^2 = \Delta^2 \sum_{ijk}[1/x(ijk)]$, asymptotically distributed as χ^2 under the null hypothesis, can be computed for a test with one degree of freedom. For a three-way r x c x d table, one has to solve (r-1) (c-1) (d-1) third-degree simultaneous equations in as many unknowns. The computation involved is not a trivial one!

C. Interpretation of Results. Once we have formulated the hypothesis and performed the computations, we need to interpret the results in terms of the actual physical variables. What does no second-order interaction in a four-way table mean? How about no third-order interaction? In some cases the interpretation may be quite natural, in other cases the interpretation would be rather stretched. A general interpretation that may apply to a majority, if not all, of the cases would be extremely desirable.

II. SUMMARY OF A PROPOSED PROCEDURE FOR THE ANALYSIS OF MULTI-DIMENSIONAL CONTINGENCY TABLES. We now discuss a procedure for the analysis of multi-dimensional contingency tables which we believe has general applicability. We shall sketch the principle and structure of the proposed analysis and then illustrate the procedure with a four-way table. For details see Ku and Kullback (1968) and Ireland and Kullback (1968). The one by Ireland and Kullback contains the proofs of the main results and applies the procedure to a problem of data adjustment. The one by Ku and Kullback applies the procedure to the testing of hypotheses, in particular the formulation, estimation and testing of second-order and higher-order interactions. We shall discuss the procedure for a three-way table, using a modified form of conventional notation.

For any observed contingency table of interest, we may visualize three associated tables as follows:

(1) The π-table $\{\pi(ijk)\}$. The π-table may be specified by the null-hypothesis, given by observations, or estimated. For example, the π-table may specify equal probability in all the cells, three-way independence, etc.

(2) The class of p-tables denoted by $\{p(ijk)\}$. A p-table is a contingency table that satisfies certain conditions of interest, usually a specification of the marginals, for instance, the one-way marginals $p(i..)$, $p(.j.)$ and $p(..k)$.

(3) The p*-table $\{p^*(ijk)\}$. The p*-table is that member of the class of p-tables which most closely "resembles" the π-table in the sense of minimum discrimination information; i.e., the p*-table minimizes the discrimination information:

$$I(p:\pi) = \sum p \cdot \ell n \frac{p}{\pi} \tag{2.1}$$

over the class of p-tables.

With these three tables fixed in mind, we shall summarize the main results given in the two references.

A. If we set $\pi(ijk) = \frac{1}{rcd}$, the uniform r x c x d table, then the classical hypotheses of independence, homogeneity, conditional independence, no second-order interaction, etc. are represented by p*-tables when certain marginals are considered as fixed, and can be considered as "generalized" independence hypotheses. Thus, when $p(i..)$, $p(.j.)$, $p(..k)$ are fixed, the p*-table has the form (for any π-table)

$$p^*(ijk) = a(i)b(j)c(k)\pi(ijk) \tag{2.2}$$

where $a(i)$, $b(j)$, $c(k)$ are determined to satisfy the marginal restraints. It turns out that for $\pi(ijk) = \frac{1}{rcd}$,

$$p_1^*\text{-table:} \qquad p_1^*(ijk) = p(i..)p(.j.)p(..k) \; . \tag{2.3}$$

When two of the two-way marginals, say $p(ij.)$ and $p(i.k)$ are specified then the p*-table takes the form

$$p^*(ijk) = a(ij)b(ik)\pi(ijk)$$

$$= \frac{p(ij.)p(i.k)}{p(i..)} \quad \text{for } \pi = \frac{1}{rcd} \; . \tag{2.4}$$

When all three two-way marginals are considered fixed, the p*-table has the form

$$(2.5) \qquad p_2^* \text{ - table: } p_2^*(ijk) = a(ij)b(jk)c(ik)\pi(ijk)$$

and the p_2^*-table satisfies Bartlett's definition on no second-order interaction for $\pi = 1/rcd$, since

$$\frac{p_2^*(111)p^*(221)}{p_2^*(121)p^*(211)} = \frac{a(11)b(11)c(11)\pi\, a(22)b(21)c(21)\pi}{a(12)b(21)c(11)\pi\, a(21)b(11)c(21)\pi}$$

$$= \frac{a(11)a(22)}{a(12)a(21)} = \frac{p_2^*(112)p_2^*(222)}{p_2^*(122)p_2^*(212)} .$$

A straight forward convergent iterative procedure is given later to determine $p_2^*(ijk)$.

A pictorial representation may be visualized as shown in Figure 2.1. Let the ordinate represent some measure of association or dependence. Then the uniform table π would be at the zero datum. Now let the p-tables be represented by the series of regions above π. If there is no restriction on p, p will include π and p^* is π. With one-way marginal restraints, the class p becomes smaller and shrinks away from π. Then the p_1^* table is the one closest to π yet satisfies these one-way marginal restraints. With all two-way marginals fixed (and hence also the one-way marginals), the region shrinks further and p_2^* is the table closest to π, and is also the table closest to p_1^*. The observed sample table is represented by a point $\hat{p}$ in the picture. The closeness of the resemblance from one table to another table is measured by the discrimination information, and the following relationships hold.

		effects of marginals		measure of interaction
$I(\hat{p}:\pi)$	=	$I(p_1^*:\pi)$	+	$I(\hat{p}:p_1^*)$
$I(\hat{p}:\pi)$	=	$I(p_2^*:\pi)$	+	$I(\hat{p}:p_2^*)$
$I(p_2^*:\pi)$	=	$I(p_1^*:\pi)$	+	$I(p_2^*:p_1^*)$
$I(\hat{p}:p_1^*)$	=	$I(p_2^*:p_1^*)$	+	$I(\hat{p}:p_2^*)$

(2.6)

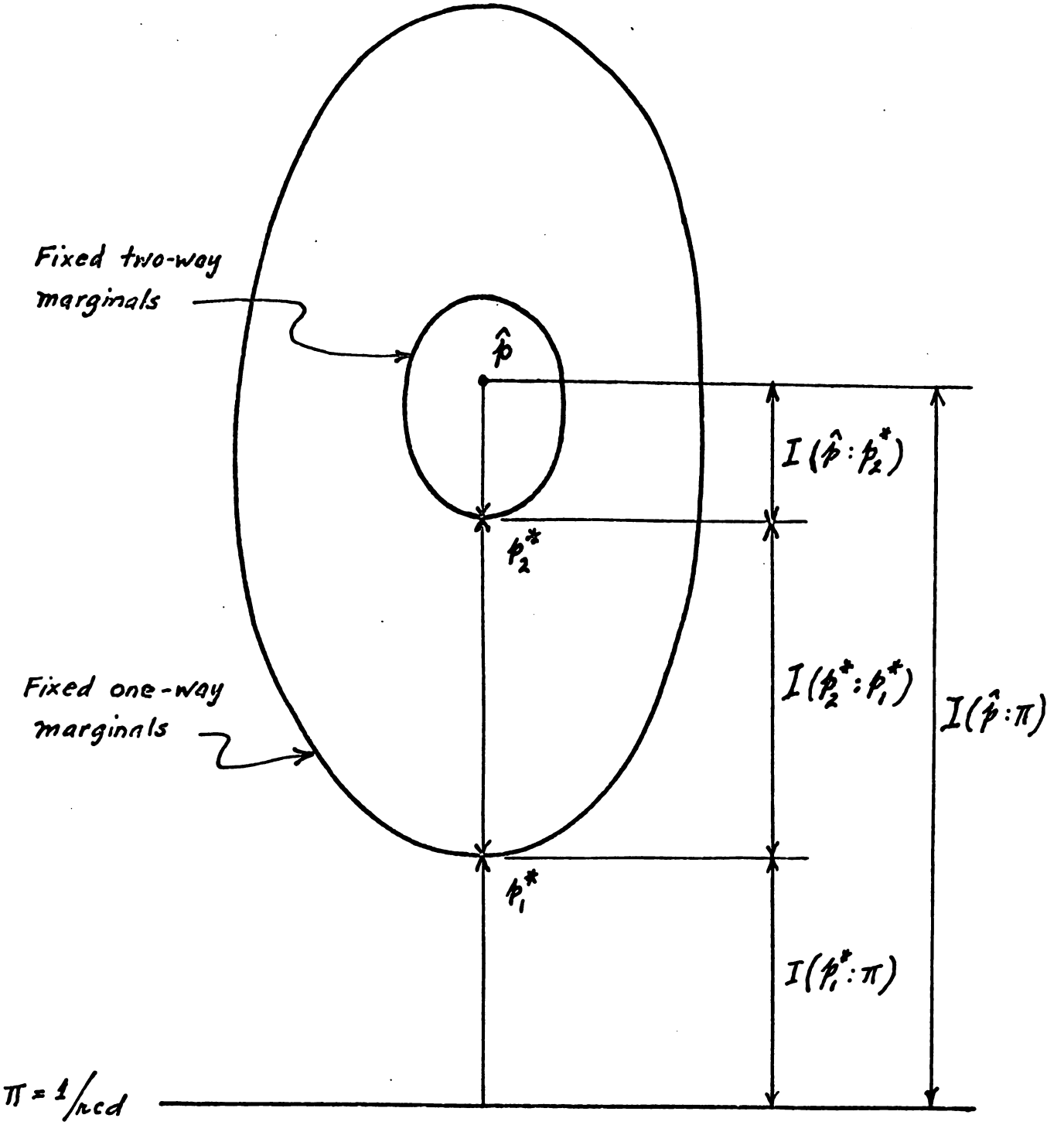


Figure 2.1

Schematic Diagram of Components of Information

In general, if p^*_a corresponds to a set H_a of given marginals and p^*_b corresponds to a set H_b of given marginals where $H_b \Rightarrow H_a$, then

$$I(\hat{p}:p^*_a) = I(p^*_b:p^*_a) + I(\hat{p}:p^*_b).$$

B. The values of the p*-table can be computed by an iterative scheme which adjusts the $\pi(ijk)$ to satisfy successively the given marginal restraints. For a three-way table when all two-way marginals are given, we cycle through

$$p^{(0)}(ijk) = \pi(ijk) = \frac{1}{rcd}$$

$$p^{(3n+1)}(ijk) = \frac{p(ij.)}{p^{(3n)}(ij.)}\, p^{(3n)}(ijk)$$

(2.7)

$$p^{(3n+2)}(ijk) = \frac{p(i.k)}{p^{(3n+1)}(i.k)}\, p^{(3n+1)}(ijk)$$

$$p^{(3n+3)}(ijk) = \frac{p(.jk)}{p^{(3n+2)}(.jk)}\, p^{(3n+2)}(ijk), \quad n = 0, 1, \ldots$$

It can be shown that the iteration converges to p* and p* is unique. For (2.4) the iteration is completed at the end of the first cycle.

C. The p*-table provides RBAN (Regular Best Asymptotically Normal) estimates under the given constraints, and

$$2nI(\hat{p}:p^*) = 2I(x:x^*) = 2\sum x(ijk) \ell n \frac{x(ijk)}{x^*(ijk)}$$

is asymptotically distributed as χ^2 under the corresponding null hypothesis, including the no second-order interaction hypothesis.

In practice, the iteration is done on the cell frequencies

$n\pi = \frac{n}{rcd}$, $x = n\hat{p}$, and $x^* = np^*$. Usually 5 to 7 cycles are sufficient to obtain agreement between marginals to within .01 or .001, when more than one cycle is required.

Now let us consider these results with respect to the three problems raised at the beginning of this paper; i.e., the problems in the formulation of the hypothesis, the computation of expected cell frequencies, and the interpretation of results.

First, we have defined a measure of "closeness" between two discrete distributions by the discrimination information given in (2.1). A hypothesis of interest is usually concerned with independence or association between various classifications. By necessity, the expected cell frequencies under such hypotheses will have to be estimated from observed marginal frequencies. Hence, all these hypotheses are members of the "generalized" independence, or no interaction hypothesis, represented by the table which is closest to the uniform π-table, subject to various marginal restraints. These tables are the p^*-tables in our procedure.

Second, we have an iterative scheme for the computation of p^* or np^*. There are two computer programs available which we shall discuss later.

We shall dwell on the third problem, the interpretation, at some length, since this is the aspect in which we are most interested. We shall give first a general interpretation and then details.

We may consider the complete sample table to contain all the "information" available from the particular experiment. In the process of analysis, we aim to express the sample table in a reduced number of parameters represented by some or all of the marginal totals. In other words, we are interested in knowing how much of this total information is contained in a summary consisting of sets of marginal tables.

If there is no first-order interaction, i.e., there is independence of all classifications, then all the information is contained in the first-order marginals in the sense that given these marginals, the complete table can be constructed to within sampling error. If the first-order interaction is significant, but there is no second-order interaction, then the set of two-way marginals will be required to summarize the data adequately. The use of two-way tables to summarize multi-way classification data is a rather common practice, and the implied assumption is therefore "no second- and higher-order interactions."

A direct consequence of this interpretation is that the analysis can be reduced to that of the set of marginal tables if there is no interaction of the same order.

We remark that the set of marginal tables must be considered jointly for proper interpretation, and if one or more of these tables show significant interactions, the results of tests of the remaining tables could lead to erroneous conclusions. An example of such a case was given in Simpson (1951).

The above interpretation is not restricted to complete sets of marginals. If the p*-table computed from three out of the six two-way marginals in a four-way table is found to be "close enough" to the p-table by our test, the three two-way marginal tables could be considered as containing essentially all the information in the four-way table. The analysis can therefore be performed on these marginal tables and the complexity of the problem reduced. For example, the analysis for a four-way table may be reduced to that of one two-way and two three-way tables, or to that of three two-way tables and one three-way table, provided that the corresponding interactions are found to be not significant.

On the other hand, we may also wish to estimate the effects, or contributions, of the specified marginal tables. An analysis of information table can be constructed using the relationships given in (2.6) wherein all the components of information are additive as well as the associated degrees of freedom. The interpretation of such a table is very similar to that of an analysis of variance table.

We remember that

$$p_1^*(ijk) = a(i)b(j)c(k)\pi(ijk), \text{ or}$$

$$\ell n\, p_1^* = \ell n\, a(i) + \ell n\, b(j) + \ell n\, c(k) + \ell n\, \pi(ijk) \tag{2.9}$$

which compares with

$$E(y) = \mu + \rho_i + c_j + \delta_k \tag{2.10}$$

the usual model for analysis of variance. Thus each model can be expressed as the sum of a grand mean, a row effect, a column effect, and a depth effect. Instead of a linear additive model, we have a logarithmic linear additive model. This fact is interesting in the sense that we did not assume such a model to start with as others have, but ended up with this model by minimizing the discrimination information.

We can also compare this procedure to multivariate regression analysis. We are in fact fitting the observed frequencies using the marginals as variables. The "a's," "b's," and "c's" are the fitted coefficients. If the effects of the two-way marginals are small, then $p_2^* \doteq p_1^*$, and the values of a(ij), b(jk), c(ik) are close to unity.

The additional effect of the two-way marginals given the one-way marginals is represented by

$$2nI(p_2^*:p_1^*) = 2\sum_{ijk} x_2^*(ijk) \ln \frac{x_2^*(ijk)}{x_1^*(ijk)},$$

(2.11)

$$= 2nI(\hat{p}:p_1^*) - 2nI(\hat{p}:p_2^*),$$

or the difference between the information statistics measuring the first-order interaction and the second-order interaction. Since

$$\ln \frac{p_2^*}{p_1^*} = \ln a(ij) + \ln b(ik) + \ln c(jk),$$

we could write also

$$2nI(p_2^*:p_1^*) = 2\sum_{ij} x(ij.) \ln a(ij) + 2\sum_{ik} x(i.k) \ln b(ik) + 2\sum_{jk} x(.jk) \ln c(jk), \quad (2.12)$$

where a(ij), b(ik) and c(jk) can be computed as products of ratios of marginals in the iteration process using p_1^* as the starting table $p^{(0)}(ijk)$.

While (2.12) is algebraically correct, and the value of $2nI(p_2^*:p_1^*)$ is unique, the three components appearing on the right side of the equation are not necessarily independent, and the computed values of these terms depend on the order of the two-way marginals within the cycle of iteration. The properties of these components need further investigation.

Hence, if a breakdown of the two-way marginal effect is desired, a conditional approach is necessary; i.e., the two-way marginal restraints are considered in successive sets, where each set implies the

preceding one, and the effect of a particular two-way marginal is computed conditioned on the preceding set of two-way marginals that had been fixed. An example is given for a 2 x 2 x 2 x 2 table in Appendix A, together with computer print-out and details of computation. This procedure is discussed in further detail in Ku and Kullback (1968), and the results of computing these effects by different approaches are compared for two four-way tables.

The main advantage in using an analysis of information table such as that given in Appendix A is that the table presents an additive analysis of the complete contingency table, rather than just a special segment of the analysis, say the hypothesis of no second-order interaction. Therefore it aids in seeing the picture as a whole and in understanding its underlying structure.

We list also, in Table 2.1, a number of results from examples appearing in current literature.

III. THE COMPUTER PROGRAMS. There are two computer programs available for the analysis of contingency tables by the procedure described above. These programs, designated KKV68A and KKV68B, respectively, are written in double precision FORTRAN V language and are now on FASTRAND at the National Bureau of Standards for use with its 1108 computer.*

The two programs are basically similar and can be used for the analysis of up to four-way tables. Whereas program A allows more categories for each classification than B, program B allows some options that are not available in A. We shall describe these programs in detail and note the computation and options that are available.

---

*Because of difference in compiler and peculiarities of behavior of different models of computers, certain minor changes may have to be made before these programs will work on other computers. We would be happy to furnish, to persons interested in using these programs, card images on a blank tape to be sent to the first two authors at:

Statistical Engineering Laboratory
National Bureau of Standards
Washington, D. C. 20234

Data from Berkson [1968]

Three-Way Tables

Comparison of $2I(x:x^*)$, χ^2 and Minimum logit χ^2 Values

Example from	$2I$ Min. Disc. Inf.	χ^2 MLE and Min. Disc. Inf.	χ^2 Min. Logit	D.F.
Cochran, 2x2x3	.854	.851	.849	2
Woolf, 2x2x3	2.9655	2.9839	2.9811	2
Norton, 2x12x2	7.71	7.59	7.37	11
Bartlett, 2x2x2	2.2945	2.2704	2.2641	1
Kastenbaum and Lamphiear, 2x3x5	3.160	3.158	3.128	8

Data from Ku and Kullback [1968], Bhapkar and Koch [1968]

Four-Way Tables

Example from	Second-Order Interaction	D.F.	Third-Order Interaction	D.F.
Hoyt, Krishniah, and Torrance, 7x4x3x2	172.257	108	44.793	36
Ries and Smith, 2x2x2x3	9.847	9	.739	2
Kihlberg, Narragon and Campbell, 2x2x2x2	7.33	5	.67	1

Table 2.1

(1) Dimension limitations.

FOR	R	C	D	T
KKV68A	$r \leq 9$	$c \leq 19$	$d \leq 9$	$t \leq 4$
KKV68B	$r \leq 7$	$c \leq 9$	$d \leq 4$	$t \leq 3$

(1A) Requirements for computer memory locations.

FOR	Code	Data	Total
KKV68A	7323	41495	48818
KKV68B	9587	11670	21257

(2) Input data and options.

A. Title cards are provided for each of the classifications.

B. Tables of sampled data: $X(IJKL) = N\hat{p}(ijk\ell)$. These data are punched on cards in 7D 10.0 format, and read in by columns within each row, row x column within each depth, and row x column x depth within each level.

C. In program A, $N\pi(IJKL) = F(IJKL)$ is always taken to be equal to n/rcdt and iterations begin with these numbers for each of the three iterations to compute $Np^*_1(ijk\ell)$, $Np^*_2(ijk\ell)$ and $Np^*_3(ijk\ell)$.

In program B, there are two options: (i) The first option uses $F(IJKL) = n/rcdt$ same as in program A. The computation sequence in program B, however, differs from that of A in that the iteration begins with F(IJKL) to compute the cell frequencies for the no first-order interaction NP^*_1, then uses NP^*_1 as input to compute the cell frequencies for no second-order interaction NP^*_2, and uses NP^*_2 to compute cell frequencies under the no third-order interaction NP^*_3. This computation sequence allows the calculation of the effects of the sets of marginals such as $2NI(p^*_1:\pi)$, $2NI(p^*_2:p^*_1)$ and $2NI(p^*_3:p^*_2)$, and their components. (ii) The second option allows the input of a table of F(IJKL), after and in the same manner as X(IJKL). This choice is useful in adjusting data to fit specified marginals - a topic not discussed in this paper. The fixed marginals will be those of X(IJKL). The table F(IJKL) is the observed table to be adjusted to fit the marginals of X(IJKL). An example is given in Ireland and Kullback (1968).

D. For program A only, option is provided for the choice of sets of marginals if these marginals are not a complete set of one-, two-, or three-way marginals. Iterative computations for the complete sets of marginals are always automatically performed.

E. Options are provided to specify the maximum number of cycles of iteration for the computation of each iteration, and also for the specification of the tolerance desired between the original marginals and the computed marginals. Experience has shown that 20 cycles and agreement to 0.01 are usually sufficient for most problems.

(3) Outputs and options.

The following notations are used in the output:

X(IJKL)	Observed cell frequencies
Y(IJKL)	Cell frequencies NP^*_1
Z(IJKL)	Cell frequencies NP^*_2
W(IJKL)	Cell frequencies NP^*_3
U(IJKL)	Cell frequencies corresponding to specified marginals
R(I), C(J), etc.	These are equivalent to a(i), b(j), etc. used in the text
RC(IJ), RD(IK), etc.	These are equivalent to a(ij), b(ik), etc. used in the text
RCD(IJK), RCT(IJL), etc.	These are equivalent to a(ijk), b(ijℓ), etc used in the text

A. In program A, there is no option, the print-out is arranged in the following order:

(i) Titles of classification.

(ii) Original tables of X(IJKL) in the form of two-way tables.

(iii) All marginal three-way, two-way, and one-way tables and the grand total. These tables are useful for inspection if there are no higher-order interactions.

(iv) All 16 sums of quantities of the form $2\sum X(IJKL)LNX(IJKL)$ computed in double precision. These sums are useful in testing certain hypotheses as illustrated in Kullback, Kupperman and Ku (1962).

(v) Number of complete cycles of iterations performed for each interaction computed, and the tolerance specified for the marginal agreement.

(vi) Tables Y(IJKL)

$2\sum Y(IJKL)LNY(IJKL)$

First-order interaction $2NI(\hat{p}:p_1^*)$

Chi-squared = $\sum \frac{(X-Y)^2}{Y}$

Tables of residuals = X-Y

Tables of normalized residuals $\frac{X-Y}{\sqrt{Y}}$

(vii) Print-outs under (vi) are repeated corresponding for Z and W, and U when specified.

B. Options available in print-out of program B. Print-outs described in Paragraph A(i)-(vii) above for program A remain the same, except for the following options:

(1) If tables of coefficients R(I), ..., RC(IJ), ..., RCD(IJK), ..., and quantities such as 2SUM X(I...)LNR(I...), ..., are computed, then these numbers will be printed out. Both tables of residuals and normalized residuals will be suppressed in this case.

(2) Options are provided to print either the residuals or the normalized residuals, or both if the coefficients are not computed and printed.

A sample computer print-out is given in Appendix A, and the setup for data cards is given in Appendix B.

REFERENCES

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

Berkson, Joseph (1968), Application of minimum logit χ^2 estimate to a problem of Grizzle with a notation on the problem of no interaction. Biometrics, 24, 75-95.

Bhapkar, V. P. and Koch, Gary G. (1968), Hypothesis of "no interaction" in multidimensional contingency tables. Technometrics, 10, 107-123.

Darroch, J. N. (1962), Interaction in multi-factor contingency tables. J. Roy. Statist. Soc. B, 24, 251-263.

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

Kastenbaum, Marvin A. (1965), Contingency tables: a review, MRC Technical Summary Report No. 596, Mathematical Research Center, The University of Wisconsin.

Kendall, M. G. (1968), On the future of statistics - a second look. J. Roy. Statist. Soc. A, 131, Part 2, 182-192.

Ku, H. H. and Kullback S. (1968), Interaction in multidimensional contingency tables: an information theoretic approach. J. of Research of the National Bureau of Standards, 72B, 159-199.

Mosteller, Frederick (1968), Association and estimation in contingency tables. J. Am. Statist. Assoc., 63, 1-28.

Pearson, Karl (1900), On the Criterion that a given system of deviations from the probable in the case of a correlated system of variables is such that it can be reasonably supposed to have arisen from random sampling. Phil. Mag. (Ser. 5) 50, 157-175.

Roy, S. N. and Kastenbaum, M. A. (1956), On the hypothesis of no "interaction" in a multiway contingency table. Ann. Math. Statist., 27, 749-757.

Simpson, C. H. (1951), The interpretation of interaction in contingency tables. J. Roy. Statist. Soc. B, 13, 238-241.

APPENDIX A

Data used in this example are taken from the Kihlberg, Narragon, and Campbell study of the relationship between car size and accident injuries as quoted by Bhapkar and Koch (1968).

There are four classifications as follows:

R: Driver ejection - not ejected or ejected

C: Accident severity - not severe or severe

D: Accident type - collision or rollover

T: Car weight - small or standard

The data are shown in the 2 x 2 x 2 x 2 table at the beginning of the print-out.

Before we discuss our procedure of analysis on this set of data, we wish to make two remarks:

First, Bhapkar and Koch condensed the original data into a 2 x 2 x 2 x 2 table presumably for convenience. The original data has 4 categories in accident type, 3 in severity and 3 in car weight, and is a 2 x 3 x 4 x 3 table for drivers who were alone at the time of accidents. Hence, all conclusions and interpretations given below are strictly for illustrative purposes, and are based on data as condensed by Bhapkar and Koch.

Second, in many problems of data analysis, there is usually additional knowledge available which should be taken into account. In this example for instance, there is a time element linking the four classifications in the sequence: car weight → accident type → accident severity → ejection. The dependence of one classification on another can only go from right to left and not in the reverse order. In addition, there is the distinction of a "cause and effect" relationship between two classifications (by logic or by law-like long past experience) or a mere "association" relationship. Severe accidents are likely to cause ejections of driver, is an example of the former; ejection of driver and car weight is an example of the latter. We hope to use these additional bits of knowledge to make our analysis more meaningful.

Analysis of information Table A.1 represents a preliminary scan of these data using our procedure. Neither the third-order nor the second-order interaction reached significance. The value for the no third-order interaction hypothesis of .67 checks with Bhapkar and Koch's results.

By our interpretation given earlier in the text, the six 2 x 2 marginal tables jointly contain essentially all the statistical information available in the four-way table, or, given the six two-way marginal tables, we could approximate the four-way table to within sampling error. Thus, the analysis is reduced to a breakdown of the six two-way marginal effects into individual degrees of freedom. We shall do this in two ways for illustration.

If we compute the <u>independence component</u> for each 2 x 2 table, we get the values shown in the <u>second column</u> in Table A.2. These computations can be performed easily using the SUM 2X(xxxx)LN X(xxxx) values given in the print-out. For example, the R x C independence component is:

$$2\hat{I} = \sum_{ij} X(IJ..) \ln \frac{NX(IJ..)}{X(I...)X(.J..)}$$

and is the difference of two sums.

$$\begin{aligned} 2X(IJ..)LN\ X(IJ..) &= 71559.893 \\ 2N\quad LN\quad N &= \underline{81960.898} \\ &\quad 153520.791 \end{aligned}$$

$$\begin{aligned} 2X(I...)LN\ X(I...) &= 77938.434 \\ 2X(.J..)LN\ X(.J..) &= \underline{75296.372} \\ &\quad 153234.806 \end{aligned}$$

$$2\hat{I} = 153520.791 - 153234.806 = 285.99 .$$

We note that the sum of these six components, 1351.62, is much larger than the two-way marginal effects value of 1185.78.

This result shows up the danger of looking at the marginal tables one at a time, even if there is no second-order interaction. Since the six two-way marginal tables are interrelated, interaction in one two-way table could conceivably affect the magnitude of interaction in a neighboring two-way table, and thus masks the actual relationship between these classifications.

Next we use the step-wise approach; i.e., the cumulative addition of one two-way marginal at a time, and compute the <u>discrimination information value</u> of the effect of the m-th two-way marginal over the marginals that had been fixed up to that time. The values so computed for the selected sequence of marginals given in column 1 are shown in

column 3 of Table A.2. The total effect checks with the value of $2nI(p_2^*:p_1^*)$ as it should.

With six two-way marginals, we have a large number (6!) of ways to order these marginals in the sequence. Corresponding to each sequence we could compute a set of information value effects. Obviously many of these sequences are without meaningful interpretation.

Here we shall appeal to the additional knowledge inherent in this set of data; i.e., we shall order the marginals to be fixed in the same order as the sequence in time, and order the "cause and effect" relationship ahead of "association" relationship as follows:

2 x 2

Marginal tables	Association of	Marginals fixed
DT	accident type - car weight	(..KL),(I...),(.J..)
CD	accident severity - accident type	(..KL),(.JK.),(I...)
RC	ejection - accident severity	(..KL),(.JK.),(IJ..)
RD	ejection - accident type	(..KL),(.JK.),(IJ..),(I.K.)
CT	accident severity - car weight	(..KL),(.JK.),(IJ..), (I.K.),(.J.L.)
RT	ejection - car weight	all two-way marginals

In choosing this particular sequence of ordering, we realize that the logic for the selection may not be entirely free from criticism. Nonetheless, this ordering appears to be reasonable for the particular problem. In comparing columns 2 and 3 in Table A.2, we note that the same conclusions will be reached for the first four effects, but exactly opposite conclusions are evidenced by the values of the last two effects.

We give in Analysis of Information Table A.3 a detailed analysis of the no first-order interaction component $2nI(\hat{p}:p_1^*) = 2I(x:x_1^*)$, using $x_a^*, \ldots, x_e^*$ to denote expected cell frequencies with 1, 2, ..., 5 two-way marginals fixed, and x_2^* to denote that of no second-order interaction. We note that

$$x_1^*(ijk\ell) = x(i...)x(.j..)x(..k.)x(...\ell)/n^3,$$

and, by (2.7),

$$x_a^* = \frac{x(..k\ell)}{x_1^*(..k\ell)}\, x_1^*(ijk\ell) = x(..k\ell)x(i...)x(.j..)/n^2,$$

hence,

$$2I(x_a^*:x_1^*) = 2\sum_{k\ell} x(..k\ell) \ln \frac{nx(..k\ell)}{x(..k.)x(...\ell)} .$$

Similarly, $x_b^*(ijk\ell)$ and $x_c^*(ijk\ell)$ can also be expressed explicitly as functions of the marginals, and the iteration process (2.7) ends at the first cycle. x_d^*, x_e^*, however, cannot be so expressed and a number of cycles of iteration are necessary to obtain the desired agreement among the marginals.

One of the useful features in these programs is that the residuals and the normalized residuals are printed out for examination. Column 1 of Table A.4 shows the normalized residuals after all two-way marginals have been fixed. All these residuals are small in magnitude. The largest two are R(2121) = 1.084 and R(2122) = -1.819. The addition of the three-way marginals x(.jkℓ), x(ijk.) and x(i.kℓ) in that sequence did not change the residuals by much. The addition of the three-way marginal x(ij.ℓ), however, improves the overall picture of these residuals. The information value of 2.928 with 1 d.f. suggests that association between ejection-accident severity-car weight may merit further investigation.

There are many ways to construct analysis of information tables for a four-way table - the choice of which depends mainly on the purposes of the experiment and the hypotheses of interest. In analysis of information Table A.5 we give an example for the analysis involving the following hypotheses:

H_1: Given accident type and car weight, ejection is independent of accident severity, or RxC|DT. The appropriate marginals to be considered fixed here are x(i.kℓ) and x(.jkℓ). Let the expected frequencies under this hypothesis be denoted as x_m^*. It can be verified that $x_m^*(ijk\ell) = \frac{x(i.k\ell)x(.jk\ell)}{x(..k\ell)}$, and can be computed directly.

Since $\sum_{k\ell} x_m^*(ijk\ell) \neq x(ij..)$, the effect of the marginal restraint x(ij..) has not been taken in account in H_1. Addition of x(ij..) to x(i.kℓ) and x(.jkℓ) as restraints yields x_n^* as expected cell frequencies of independence of R and D classifications given the three marginals. The statistic $2I(x_n^*:x_m^*)$ measures the conditional independence between R and C classifications as in H_1, but with the added restraint x(ij..), i.e., the sum of the expected cell frequencies over the last two classifications must satisfy the observed frequencies. Hence, we have

H_2: Given accident type, car weight and the observed ejection-severity frequencies, ejection is independent of accident severity. The difference between the two components represents

H_3: The association between ejection and accident severity is independent of accident type and car weight.

We note that x^*_n cannot be computed directly since $x(i.k\ell)$, $x(.jk\ell)$ and $x(ij..)$ imply all six two-way marginals. In analysis of information Table A.6 we show that $2I(x:x^*_n)$ is in fact a component of three-way marginal effect.

Analysis of Information Table A.1

Traffic accidents data from Bhapkar and Koch (1968)

Components due to	Information	Value	D.F.
No first-order interaction	$2nI(\hat{p}:p^*_1)$	1193.11	11
Effect: all two-way marginals	$2nI(p^*_2:p^*_1)$	1185.78	6
No second-order interaction	$2nI(\hat{p}:p^*_2)$	7.33	5
Effect: all three-way marginals	$2nI(p^*_3:p^*_2)$	6.66	4
No third-order interaction	$2nI(\hat{p}:p^*_3)$	.67	1

Table A.2

Analysis of Effect of All Two-Way Marginals $2nI(p^*_2:p^*_1)$

Two-way marginal tables	Independence in each 2 x 2 table	Information value, marginals fixed cumulatively in the sequence given at left	D.F.
DT	52.96	52.96	1
CD	601.42	601.42	1
RC	285.99	286.00	1
RD	401.69	229.33	1
CT	.76	14.38	1
RT	8.80	1.69	1
	1351.62	1185.78	6

Analysis of Information Table A.3

First-order Interaction $2I(x:x_1^*) = 1193.11$

	Two-way marginals given	Information	D.F.
a)	DT effect	$2I(x_a:x_1^*) = 52.96$	1
	interaction	$2I(x:x_a^*) = 1140.15$	10
b)	DT,CD effect	$2I(x_b^*:x_a^*) = 601.42$	1
	interaction	$2I(x:x_b^*) = 538.73$	9
c)	DT,CD,RC effect	$2I(x_c^*:x_b^*) = 286.00$	1
	interaction	$2I(x:x_c^*) = 252.73$	8
d)	DT,CD,RC,RD effect	$2I(x_d^*:x_c^*) = 229.33$	1
	interaction	$2I(x:x_d^*) = 23.40$	7
e)	DT,CD,RC,RD,CT effect	$2I(x_e^*:x_d^*) = 14.38$	1
	interaction	$2I(x:x_e^*) = 9.02$	6
f)	DT,CD,RC,RD,CT,RT effect	$2I(x_f^*:x_e^*) = 1.69$	1
	interaction	$2I(x:x_2^*) = 7.33$	5

Table A.4

Normalized Residuals

	All two-way marginals		Add x(.jkℓ) x(ijk.) x(i.kℓ)		All three-way marginals	
R(IJ11)						
	-.482	.825	-.126	.426	.086	-.131
	.473	-.576	.487	-.470	-.306	.348
R(IJ21)						
	.351	-.306	-.551	.426	-.205	.153
	1.084	-.387	1.158	-.481	.386	-.178
R(IJ12)						
	.068	-.112	.055	-.074	-.037	.050
	.374	-.252	-.222	.188	.154	-.126
R(IJ22)						
	.293	-.143	.369	-.219	.133	-.080
	-1.819	.608	-.860	.274	-.331	.099

Components due to	Information	D.F.
Second-order interaction	$2I(x:x_2^*) = 7.328$	5
CDT, RCD, RDT effect	3.733	3
RCT effect	2.928	1
Third-order interaction	.667	1

Analysis of Information Table A.5

Components due to	Information	D.F.
RxC|DT	$2I(x:x_m^*) = 120.665$	4
RxC|DT,RC	$2I(x_n^*:x_m^*) = 115.137$	1
(RxC)(DT)	$2I(x:x_n^*) = 5.528$	3

Analysis of Information Table A.6

Second-order interaction	$2I(x:x_2^*) = 7.328$	5
	$2I(x_n^*:x_2^*) = 1.800$	2
	$2I(x:x_n^*) = 5.528$	3
	$2I(x_3^*:x_n^*) = 4.859$	2
Third-order interaction	$2I(x:x_3^*) = .669$	1

```
R  DRIVER EJECTED   I=1 NO           I=2    YES
C  SEVERITY         J=1 NOT SEVERE   J=2    SEVERE
D  ACCIDENT TYPE    K=1 COLLISION    K=2    ROLLOVER
T  CAR WEIGHT       L=1 SMALL        L=2    STANDARD


ORIGINAL TABLES


X(IJ11)


   350   150
    26    23


X(IJ21)


    60   112
    19    80


X(IJ12)


  1878  1022
   111   161


X(IJ22)


   148   404
    22   265


MARGINAL TABLES


THREE-WAY TABLES


X(IJ1*)


  2228  1172
   137   184
```

X(IJ2*)

208 516
41 345

X(.J1L)

376 1989
173 1183

X(.J2L)

79 170
192 669

X(1J.L)

410 2026
262 1425

X(2J.L)

45 133
103 426

X(I.1L)

500 2900
49 272

X(I.2L)

172 552
99 287

TWO-WAY TABLES

X(IJ..)

2436	1688
178	529

X(I.K.)

3400	724
321	386

X(I..L)

672	3452
148	559

X(.JK.)

2365	249
1356	861

X(.J.L)

455	2159
365	1852

X(..KL)

549	3172
271	839

ONE-WAY TABLES

X(I...)

```
  4124    707

X(.J..)

  2614   2217

X(..K.)

  3721   1110

X(...L)

   820   4011

TOTAL

X(....)

  4831

PRINT OF SUMS

SUM 2X(IJKL)LNX(IJKL)=     .6285850425756393+005

SUM 2X(IJK.)LNX(IJK.)=     .6718460467309836+005

SUM 2X(.JKL)LNX(.JKL)=     .6635806581957651+005

SUM 2X(IJ.L)LNX(IJ.L)=     .6717184045852968+005

SUM 2X(I.KL)LNX(I.KL)=     .6878513736964094+005
```

SUM 2X(IJ..)LNX(IJ..)= .71559893135795l1+005

SUM 2X(I.K.)LNX(I.K.)= .7313235841855975+005

SUM 2X(I..L)LNX(I..L)= .73546476450096l0+005

SUM 2X(.JK.)LNX(.JK.)= .706900353910l703+005

SUM 2X(.J.L)LNX(.J.L)= .7089638610024700+005

SUM 2X(..KL)LNX(..KL)= .7240536400142l9+005

SUM 2X(I...)LNX(I...)= .7793842367626649+005

SUM 2X(.J..)LNX(.J..)= .752963715522335S+005

SUM 2X(..K.)LNX(..K.)= .767531425477157l+005

SUM 2X(...L)LNX(...L)= .7756015555323097+005

2N LN N= .8196089828312467+005

NO OF ITERATIONS= 1 CYCLES

AGREEMENT BETWEEN MARGINALS TO .100-02

Y(I,J,K,L)

Y(IJ11)

291.734 247.427
 50.014   42.418

Y(IJ21)

 87.026   73.809
 14.919   12.654

Y(IJ12)

1427.005 1210.279
244.639 207.485

Y(IJ22)

425.685 361.035
 72.978   61.894

2Y LN Y= .616653984800727I+005

FIRST-ORDER INTERACTION= .1193105777+004

CHI-SQUARED= .1601729312+004

NO OF ITERATIONS= 8 CYCLES

AGREEMENT BETWEEN MARGINALS TO .100-02

Y(I,J,K,L)

Y(IJ11)

359.131 140.235
23.698 25.935

Y(IJ21)

57.345 115.289
14.826 83.540

Y(IJ12)

1875.040 1025.594
107.132 164.235

Y(IJ22)

144.484 406.882
32.345 255.289

TABLE OF RESIDUALS

R(IJ11)

-9.131 9.765
2.302 -2.935

R(IJ21)

2.655	-3.289
4.174	-3.540

R(IJ12)

2.960	-3.594
3.868	-3.235

R(IJ22)

3.516	-2.882
-10.345	9.711

TABLE OF NORMALIZED RESIDUALS

R(IJ11)

-.482	.825
.473	-.576

R(IJ21)

.351	-.306
1.084	-.387

R(IJ12)

.068	-.112
.374	-.252

R(IJ22)

.293	-.143

```
-1.819     .608

2Z LN Z=     .6285117708584006+005

SECOND-ORDER INTERACTION=          .7327171723+001

CHI-SQUARED=          .7014267537+001

NO OF ITERATIONS=     7 CYCLES

AGREEMENT BETWEEN MARGINALS TO          .100-02

Y(I,J,K,L)

Y(IJ11)

348.393  151.607
 27.607   21.393

Y(IJ21)

 61.608  110.392
 17.392   81.608

Y(IJ12)

1879.608 1020.392
 109.392  162.608

Y(IJ22)

146.392  405.608
 23.608  263.392
```

TABLE OF NORMALIZED RESIDUALS

R(IJ11)

.086	-.131
-.306	.348

R(IJ21)

-.205	.153
.386	-.178

R(IJ12)

-.037	.050
.154	-.126

R(IJ22)

.133	-.080
-.331	.099

2W LN W= .628578345802709+005

THIRD-ORDER INTERACTION= .6696095368+000

CHI-SQUARED= .6714044680+000

SPECIFIED MARGINALS ..KL .JK. IJ.. I.K.

NO OF ITERATIONS= 6 CYCLES

AGREEMENT BETWEEN MARGINALS TO .100-02

Y(I,J,K,L)

Y(IJ11)

329.621 172.019
19.314 28.047

Y(IJ21)

49.293 127.467
11.499 82.741

Y(IJ12)

1904.476 993.885
111.589 162.050

Y(IJ22)

152.609 394.630
35.599 256.161

2(U4) LN (U4)= .6283510899375868+005

INTERACTIONS (U4)= -.2339526380+002

CHI-SQUARED= .2352315589+002

```
SPECIFIED MARGINALS      ..KL      .JK.      IJ..      I.K.      .J.L

NO OF ITERATIONS=      8 CYCLES

AGREEMENT BETWEEN MARGINALS TO          .100-02

Y(I,J,K,L)

Y(IJ11)

 361.187   143.288
  21.163    23.363

Y(IJ21)

  58.909   120.276
  13.742    78.073

Y(IJ12)

1872.911  1022.615
 109.740   166.735

Y(IJ22)

 142.994   401.821
  33.356   260.829

2(US) LN (US)=      .6284948373955333+005

INTERACTIONS (US)=          .902051801D+001

CHI-SQUARED=          .876727215D+001
```

SPECIFIED MARGINALS .JKL IJK. I.KL

NO OF ITERATIONS= 4 CYCLES

AGREEMENT BETWEEN MARGINALS TO .100-02

Y(I,J,K,L)

Y(IJ11)

552.368 147.632
23.632 25.368

Y(IJ21)

64.420 107.580
14.580 84.420

Y(IJ12)

1075.632 1024.368
113.368 158.632

Y(IJ22)

143.580 408.420
26.421 260.579

TABLE OF NORMALIZED RESIDUALS

R(IJ11)

-.126	.195
.437	-.470

R(IJ21)

-.551	.426
1.156	-.431

R(IJ12)

.055	-.074
-.222	.188

R(IJ22)

.369	-.219
-.360	.274

Σ(U4) LN (U4)= .628549090432104E+005

INTERACTIONS (U4)= .3595214353+001

CHI-SQUARED= .3660335627+001

SPECIFIED MARGINALS I*KL *JKL IJ**

NO OF ITERATIONS= 6 CYCLES

AGREEMENT BETWEEN MARGINALS TO .100-02

Y(I,J,K,L)

Y(IJ11)

353.366 146.634
22.634 26.366

Y(IJ21)

62.331 109.669
16.669 82.331

Y(IJ12)

1881.086 1018.914
107.914 164.087

Y(IJ22)

139.217 412.783
30.783 256.217

2(U1) LN (U1)= .6285297612235454+005

INTERACTIONS (U1)= .5528135209+001

CHI-SQUARED= .5276909156+001

APPENDIX B

Setup for Data Cards for KKV68A and KKV68B

All data values must be right adjusted in the fields specified.

(1) First card has the value of N punched in columns 1-5.

N = 2 for the two-way table
N = 3 for the three-way table
N = 4 for the four-way table

(2) The next set of data consists of N descriptive title cards. The information may be punched in columns 1-72.

(3) The next card consists of data in the following columns:

Col. 1-5 M1 = number of categories in the row classification
Col. 6-10 M2 = number of categories in the column classification
Col. 11-15 M3 = number of categories in the depth classification
Col. 16-20 M4 = number of categories in the level classification
Col. 21-25 NSETS = number of specified sets of marginals ≤ 5
For program KKV68B, NSETS $\equiv 0$.
Col. 26-30 ITMAX = maximum cycles of iterations
Col. 31-50 CONST = tolerance required between marginals
Col. 51-55 IFCR = 1, compute F(IJKL) = n/rcdt
IFCR = 2, input F(IJKL)
Col. 56-60 IRCD = 1, coefficients not calculated
IRCD = 2, coefficients calculated
Col. 61-65 IPRINT = 1, print residuals
IPRINT = 2, print normalized residuals
IPRINT = 3, print both

If IRCD = 2, IPRINT has no effect.

For program KKV68A, cols- 51-65 are not used.

(4) The next group of data cards consists of the specified marginals if NSET $\neq 0$ in program KKV68A.

The first card contains NMARGS in col. 1-5. NMARGS ≤ 6.

The second card contains MARGN in 6A4 format.

MARGN is specified as, e.g., IJ..

(5) The next set of cards is the input X(IJKL) specified using format 7D10.0.

(6) If IFCR = 2 in program KKV68B, the next set of cards are the F(IJKL) values in 7D10.0 format. The cards for (5) and (6) are punched by column with each row, row x column within each depth, and row x column by depth within each level.

Repeat from (1) through (6) for each set of data to be analyzed.

NEYMAN AWARDED THE 1968 SAMUEL S. WILKS MEMORIAL MEDAL

The Recipient of the Fourth Samuel S. Wilks Award Announced by Frank E. Grubbs

Professor Jerzy Neyman of the University of California, Berkeley, has been awarded the Samuel S. Wilks Memorial Medal for 1968. The announcement of Professor Neyman's selection for the 1968 Wilks Award was one of the highlights of the Fourteenth Annual Conference on the Design of Experiments in Army Research, Development, and Testing, which was held at the Army Chemical Center, Maryland, 23-25 October 1968. Professor Neyman has long been recognized as one of the foremost statisticians in the entire world, having made many fundamental contributions to the theory and application of statistical methodology. The citation for Professor Neyman reads as follows:

"To Professor Jerzy Neyman, whose extensive contributions both to the theory and practice of statistics have led to fundamental changes in the thinking and methodology of scientists all over the world. He has inspired and led more than a generation of students and his continued leadership is effective today. Both by precept and by example, he is one of the foremost statisticians in the entire world."

The Samuel S. Wilks Memorial Medal Award 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 contribution 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, and the Government.

The Award consists of a medal, with a profile of Professor Wilks and the name of the Award on one side, and the seal of the American Statistical Association and name of the recipient on the reverse; a citation, and an honorarium related to the magnitude of the Award funds. The Annual 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 Wilks Memorial Award were donated by Philip G. Rust, Thomasville, Georgia.

With the approval of President Geoffrey Moore of the American Statistical Association, the Wilks Award Committee for 1968 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. Frank E. Grubbs, Chairman	Aberdeen Research and Development Center

Professor Jerzy Neyman was born in Bendery, Bessarabia of Polish parents. He was educated in Russia at the University of Kharkov and when Poland again became an independent state, he went to Warsaw where he received the Ph.D. degree from the University of Warsaw. He held several positions in Poland: as a Lecturer at the University of Warsaw and the University of Cracow and was Head of the Biometric Laboratory of the Nencki Institute in Warsaw. Dr. Neyman received a Rockefeller Fellowship which allowed him to study at the University of Paris and at University College, London. In 1934, he became a member of the staff at University College, remaining there until 1938 when he went to the University of California, Berkeley as Professor of Mathematics. He has remained at Berkeley for 30 years as the Director of the Statistical Laboratory and Professor first of Mathematics and then, in 1955, when the Statistics Department was established, as Professor of Statistics. Professor Neyman has been a Visiting Lecturer at many universities in the United States and abroad. He is now Professor Emeritus recalled to active duty and Director of the Statistical Laboratory.

Professor Neyman has received many awards and honors including an honorary degree from the University of Chicago, the University of California, and the University of Stockholm. He has also received the Guy Medal in Gold of the Royal Statistical Society, (London, England) the Newcomb Cleveland Award of the American Association for the Advancement of Science and the Centennial Award of the Academic Senate of the University of California at Berkeley. In 1963, he was elected to the National Academy of Sciences, U.S.A., and as a foreign member of the Royal Swedish Academy. In 1966, he was elected to foreign membership in the National Academy of Science of Poland. He was elected an honorary member of the International Statistical Institute, a Fellow of the American Statistical Association, of which he was Vice President 1947-48, and a Fellow of the Institute of Mathematical Statistics of which he was President in 1949. Dr. Neyman is a Fellow of several other societies including the Econometric Society, the Biometric Society, the Mathematical Society of France, and the Polish Mathematical Society. He is a member of several other mathematical societies and of several astronomical societies including the International Astronomical Union. He is President-elect of the International Association for Statistics in Physical Sciences.

Professor Neyman's research can be divided into three parts. First, he worked in pure mathematics, but, then beginning in the late 1920's, he turned to the theory of statistics. He developed, jointly with E. S. Pearson, the theory of testing hypotheses and also developed the theory of confidence intervals. Even while Dr. Neyman was in Poland and in England, he was concerned with the application of statistics. After he came to the United States, his interests turned more towards applications of statistics, especially sampling theory and applications in the various sciences including astronomy, biology, and health and weather modification. His principal recent theoretical work has been the development of the $C(\alpha)$ test for testing composite hypotheses.

During World War II, Professor Neyman and the Berkeley Statistical Laboratory worked on tactical problems of the Air Force under the National Defense Research Committee. He has served on many committees concerned with statistics in branches of the Government, in scholarly societies and in education.

Professor Neyman is the author of Lectures and Conferences and of First Course in Probability and Statistics. He is the editor of Bernoulli-Bayes-Laplace Jubilee Volume and of the Proceedings of the Berkeley symposia, which now amounts to 17 volumes running to more than 7,600 pages. In addition, he is the author or co-author of more than 200 scientific papers in scholarly journals. As noted above, his publications form the very basis of modern testing hypotheses and interval estimation. Indeed, they are now regarded as classical and the earlier papers have been republished jointly by the Cambridge University Press and the University of California Press in two volumes: one contains the paper joint with E. S. Pearson; the other contains the remaining important papers published before 1945. Several of his books and papers have been translated into Spanish, Polish, and Russian.

It is probably correct to state that Professor J. Neyman is one of the most outstanding statisticians in the world today, due not only to his extremely important basic contributions, but also to his great activity in using the fundamental concepts in many fields of application and in constructing stochastic models with such diverse and important phenomena as a two-stage theory of carcino-genesis and the distribution of galaxies in space.

Professor Neyman has many students and by now grand students and great-grand students all over the world. Almost all of his students in Poland were killed by the Nazi invasion. However, since he attracts students to Berkeley from every country, there is by now a new generation of Polish students of Professor Neyman's. Today, his doctoral students are working in theoretical statistics, in problems arising in the design of weather modification experiments, in carcino-genesis, in the transfer of memory, and in several intricate problems in cell biology.

Professor Neyman is admired by his colleagues and his students in creating a stronger science, a more meaningful education, and a better world in which to live. In addition to the high esteem of his colleagues and students, Neyman enjoys their affection. Distinguishing characteristics are his intellectual inspiration and dedication.

PROBLEMS IN EVALUATING TREATMENT RESPONSE OVER UNEQUALLY SPACED TIME INTERVALS

Gerhard J. Isaac
U.S. Army Medical Research and Nutrition Laboratory
Denver, Colorado

ABSTRACT. Test subjects are first conditioned and then various physiological parameters are measured at a sea level location to provide base-line values. After moving to high altitude the measurements are repeated several times at unequally spaced time intervals. Final measurements are made upon return to sea level. Interest exists in such findings as the initial impact of high altitude exposure, possible adjustment to altitude and effect of return to sea level. What statistical analysis will provide the most appropriate basis for inferences about the questions of interest? Analyses considered include analysis of variance and paired t tests against control.

The problem I am going to outline for you became of special interest to me in connection with certain high altitude studies carried out by our Laboratory. Basically these studies involved the moving of test subjects from sea level to high altitude (14,100 ft.) and back to sea level. A primary interest was in the effect of altitude on performance. Also of interest were possible explanations of the physiological basis for changes in performance and in ways of ameliorating the effects of abrupt movement to high altitude.

Parameters selected for measurement include those which prior studies, or knowledge of physiological processes, suggest may be responsive to changes in altitude. Initial measurements made at sea level, after a period of training, provide the control of base-line values for each subject. The effects of altitude are reflected in subsequent measurements on selected days at altitude. This might follow a pattern like days 1, 3, 7 and 14 after arrival at altitude. Final measurements are made upon return to sea level.

An appropriate statistical analysis is desired to provide a basis for answers to a series of questions about the parameters measured. Inference drawn will reflect not only comparisons among the findings, but also will deal with the physiological aspects of the parameters. In a given case the fact of signigicant change may be more important than the direction of change, though direction also may be of concern.

It may be useful to list some of the questions that arise in a study of this kind.

1. What is the initial impact of a move to high altitude?
   a) Is the parameter significantly modified in any way by the change in environment?

2. What is the effect of remaining at high altitude?
   a) If the initial impact is a modification of the control values, do they tend to subsequently return to normal, do they tend to modify further or do they remain about as initially modified?
   b) If there is no significant initial impact at altitude, is there a tendency for values to change gradually as exposure to altitude continues? Is there reason to believe there is a training effect, an adjustment to altitude, or that some other factor is operating?

3. What is the effect of returning to sea level after a period at altitude?
   a) Is there an immediate return to control levels?
   b) Is there a delayed return to control levels?
   c) Do values find a level different from both control and altitude levels, as in the case of a continuing training effect?

The obvious overall problem is that of selecting statistical procedures that are both valid and appropriate for testing the various hypotheses implied in the series of questions posed. Also to be considered is, which statistical procedures will contribute the most toward extracting the maximum amount of useful information from the data.

Let us consider first a statistical evaluation that starts with an analysis of variance. This permits inference regarding the presence of significant differences among the means for the measurement days. At this point, however, there is no direct information as to which differences between days are significant. For example, we have to look further for information about the significance of the initial impact of moving to altitude as reflected in the control measurement and the first one at altitude. Furthermore, the size of the difference between these measurements, or any other pair, may reflect both treatment effect and the number of days elapsed between measurements.

At this point I became disturbed at the implications of using an overall anova as the basis for some critical difference which would be used to test for significant differences between various means. The difficulties seemed to be much the same whether I thought of unusual variability among the subjects at this time because of accidents of selection, variability in the state of conditioning or adjustment to the test procedures. Of course, especially when numbers are small, the usual observations about the paired t procedure are in order. A minus factor is the loss of degrees of freedom, and a plus factor is the incorporation into the calculation of the correlation between the two sets of response data. Thus this analysis places a premium upon consistency of direction and extent of change among the subjects. As we well know, if all subjects tend to move in the same direction and in about the same proportion, even a small relative change may show up as highly significant. But such a comparison utilizes only a portion of the data in determining the error component, whereas anova utilizes the entire set of data.

The main argument for the paired t test seemed to lie in the directness of the inferences that could be drawn. The test between control and initial altitude values would provide an answer to the question about possible significant changes due to the initial impact of altitude. Comparisons with subsequent days would reveal if significant changes persisted and for how long. Or, in the case of a delayed reaction, when a significant difference developed. Comparisons between control and final sea level values would reveal the extent to which there was, or was not, a return to original levels. This would reflect possible training effects or carryover effects related to the stay at altitude. Similar paired t tests made between altitude days would throw light on the effect of sustained living at altitude. Or comparisons could be made between final altitude and final sea level measurements.

It would seem that in the paired t approach the emphasis is on changes in the levels of the measurements under the various test and there is a minimum concern over the length of time intervals between measurements.

An extension of the problem occurs when the test subjects are subdivided into treatment groups. A common procedure is to put all subjects through a conditioning program at sea level before making control measurements. When the subjects are moved to altitude one group may be fed a diet, or a drug, that it is hoped will mitigate some of the undesirable responses to altitude exposure. If the randomization process used to make assignments to the treatment groups is successful, the control values of the groups will be in close agreement. By the same token, if the treatment is successful, there will be divergence at altitude, and possibly a return to agreement when again measured at sea level.

A two-way anova can be performed but questions of logic arise because of the patterns in the responses. Often, the response curve is essentially parabolic in form and is anchored at control and final sea level values. Only the "middle" is really subject to treatment response. It would seem that this would lead to understating the average difference between groups because the "end" values, by design, have minimal variation, whereas, treatment response, if present, is concentrated in the "middle" values.

It would appear that there are several options for approaching this problem. If there are only two treatments, it would seem appropriate to first run a t test of differences between treatments at control and at return to sea level. Non-significant differences at final sea level might be said to confirm this. On the other hand there could be significant differences because of treatment carry-over effects on a particular parameter. Or for that matter, training effects not related to treatment could also be present.

By means of t tests, the differences between treatment groups could be evaluated at any time-point. It is possible to determine if significant differences between treatments appear as a reflection of the initial impact of altitude, staying at altitude, or returning from altitude to sea level.

If more than two treatments are involved, the problem is somewhat more complex. It would appear appropriate to evaluate differences at each time period separately. Anova is a possibility. This could be followed by some of the tests of all differences between means, to find which treatment differences contribute the most to overall variability.

Under either approach, when treatment groups differ significantly at control, no simple answer follows from comparisons at later times. If significance disappears at subsequent dates, it may be that the passage of time, training or adaptation to environment, tend to make the response in the treatment groups the same. If the groups are significantly different on all measurement days, the interpretation is at best ambiguous. It could be that all of the test subjects happen to be responsive to altitude in the same way regardless of treatment. Quantitatively the values may be at different levels for the various treatments. Again this could be due to chance, or poor judgment (or lack of randomness) in assigning the test subjects to treatment groups.

The problem we have been considering is not unique to the experiments I have been using as examples. There are parallels in other areas. A common experimental procedure in nutrition research is to feed test and normal diets to groups of rats during their most active growth period. This may cover a period of 8 to 12 weeks immediately after weaning. Comparison of the growth curves during this period is one way of evaluating the response to the test diet. Usually initial group-average weights are very close together. This is partly by intent, and is accomplished in any of several ways. The experimental animals may be purchased under specifications limiting the weights to a fairly narrow range. The animals may be assigned to treatment groups entirely at random, or arrayed by weight and weight pairs distributed to treatments randomly. Either method usually results in treatment averages that agree closely.

The experimenter may be interested in either the final weights or in the route by which they got there. With initial weights not significantly different, the final weights for the two treatment groups can be examined. A t test of the difference between the group means seems appropriate. If the difference is significant, there may be interest in when this became apparent. In a 12-week experiment, differences might be tested at 6, 8 and 10 weeks to locate when the divergence became significant. Actual times could be selected from examination of the raw data in chart form. In most of these experiments the precise shape of the growth curves is less interest than evidence of significant divergence. In any event, it would be possible to establish whether differences became significant early or late in the experiment. Also in the case of non significance of final differences, it might be useful to know if significant differences appeared at midpoint and then disappeared as the laggards "caught up." In general, however, the primary emphasis has been on differences and not on levels of weight achieved at any particular time. In some experiments, such as involving mature animals, there could be interest in level changes within treatments as well as in differences between treatments.

To get back to the altitude problem, are there other alternatives? One that occurs to me is to express a parameter in some other form to facilitate comparison. Perhaps values for each subject expressed relative to his control as 100 percent might permit meaningful evaluation. Or would it suffice in a given case simply to note that under one dietary regimen, values at altitude are not significantly different from control while under another regimen, they are.

I have not found a satisfactory and definitive answer of universal application in experiments of the kinds I have used as illustrations. It would seem that a large dose of judgment is essential to guide a statistical evaluation of this kind. The reaction of the panel to the various possibilities is solicited. Suggestions for entirely different approaches also are in order.

ANALYSIS OF DATA FROM THE WOUND DATA AND MUNITIONS EFFECTIVENESS TEAM IN VIETNAM

W. Bruchey, L. Sturdivan, and R. Whitmire
Terminal Ballistics Laboratory
U. S. Army Aberdeen Research and Development Center
Aberdeen Proving Ground, Maryland

I. INTRODUCTION. Efforts of the U. S. Army to gather data on wound ballistics dates back into the 19th century. In "modern" times, the laboratory experiments have been supplemented by data gathered in the battlefield. We refer, of course, to the U. S. Army Surgeon General's report on a number of engagements in the second World War and Korea, entitled, "Wound Ballistics." Until the Vietnam conflict, however, efforts to collect field data were rather limited in both breadth and depth. In August, 1966, the Vice Chief of Staff, U. S. Army, made known the requirements that a study be undertaken to gather data pertinent to evaluating the effectiveness of antipersonnel munitions deployed in Vietnam, including a comprehensive study of wounds and post-wounding behavior of resultant casualties. The Wound Data and Munitions Effectiveness Team, called WDMET, was organized to fulfill this mission. A data collection format with eleven (11) sections dealing with specific areas of interest was compiled from the requirements of relevant government agencies. A team of forty-three (43) men with various military specialties was given training in ballistics, wound ballistics and collection procedures in late April and early May, 1967 at the Army Chemical Center, Edgewood Arsenal, Maryland. By late July, 1967, the Team was in operation in Vietnam. Another group of about ten (10) men was assigned to Edgewood Arsenal as a center for receiving, processing and analyzing the data from the Vietnam Teams. In conjunction with the Wound Ballistics Group of the Ballistic Research Laboratories (BRL), Aberdeen Proving Ground, a complete system for storage, retrieval, and analysis of the WDMET data was designed for the BRL electronic computer, BRLESC.

The WDMET Team in Vietnam was organized into four (4) sections: (1) Headquarters and Support Section in Saigon; (2) Section I in An Kae, following units of the 1st Air Cavalry; (3) Section II at Cu Chi, covering elements of the 25th Infantry Division; and, (4) the Pathology Section at the Saigon Mortuary. Each section studies American casualties from a battalion-sized unit. Section I reported on 100% of the casualties in its selected units, while Section II covered all casualties in selected engagements. The Pathology Section autopsied the "killed-in-action" and "died-of-wounds" casualties which had been covered by the field teams. In addition, they performed autopsies on selected cases not covered by the field sections, but which could contribute to fulfilling the WDMET mission.

Due to the irregular character of the conflict in Vietnam, most cases contribute useful information to only a small part of the WDMET area of interest. The cases are not randomly selected, and they are "typical" only insofar as the war in the two areas covered is typical of Vietnam as a whole. The enemy's weapons are often improvised or not seen, making identification or characterization of the weapon difficult or impossible in those cases. This type of data was collected under most unfavorable conditions, to say the least. The WDMET personnel, of course, were never allowed to interfere with the mission of the units being covered or with the proper medical treatment of the wounded. Data could not be obtained until the engagement was terminated.

In addition to the problems inherent in the method of data acquisition, biases are present in the data selection procedures. As stated previously, Field Section I attempted to get information on all casualties from a selected unit. This was done whenever feasible. However, due to the limited number of personnel available for data acquisition and the nature of the Vietnam conflict, there were periods of intense activity during which it was not possible to cover every casualty. There was necessarily some case selection on an individual basis with the team member forced to select the "most valuable" cases. Field Section II, on the other hand, selected incidents from which all casualties were covered. Incidents were generally selected on the basis of weapon or weapons involved and the availability of information.

As the completed casualty reports are received at WDMET(C), Edgewood Arsenal, they are coded onto punch cards and are submitted to the Wound Ballistics Group at BRL for processing on the BRLESC computer. The actual processing of the coded information is handled by three separate computer programs: (1) an error checking program; (2) a print-out program; and, (3) an analysis program. The error checking program checks for selected punching errors in the input data; the print-out program reads the input data and produces a narrative print-out for each casualty report. It is the third program which is of interest here.

At present, the analysis program produces a simple enumeration of the frequency of occurrence of the various factors contained in the study and correlations among the factors (frequency of occurrence of two or more factors in the same case). Presented here is a selection of correlations generated by this computer program. These specific correlations were chosen because of their interesting content and the likelihood that they would contain the largest number of data points for our limited size sample.

Our problem centers about the interpretation of selected correlations and methods of determining if significant differences exists between the distribution of groups of data as it is received from Vietnam.

II. FREQUENCY OF OCCURRENCE OF FACTORS. The eleven (11) sections of the data collection format are composed primarily of coded information. For example, the type weapon carried by the casualty was coded as follows:

Code No.	Weapon
0	Unknown
1	M16
2	M14
3	M79
4	M60
5	Mortar
6	Rocket
7	Other

In this manner, it was possible to determine if a factor occurred and how often. The following three tables represent this type of enumeration of factors.

The data presented in Table 1 represents the number of casualties associated with each injury type for the first 930 cases received and coded onto punch cards by WDMET(C). The totals on injury type come to 904 cases; adding to this the 26 cases which had no information on injury type bring the total to 930. It should be understood that this distribution is not truly representative of the Vietnam conflict as a whole in that there is a higher percentage of fatalities than is found in the casualty distributions as compiled by the Office of the Surgeon General. This is due primarily to the fact that the personnel located at Saigon Mortuary performed a number of autopsies in wound pathology studies apart from cases studied in the field. When this is taken into consideration, much of the difference between the true distribution and the WDMET distribution is removed.

The upper half of Table 2 lists the types and frequency of occurrence of body armor encountered in the study thus far. The total number of casualties who were wearing body armor was 139; those known not to be wearing armor, 565. This gives a ratio of 4 to 1 of armor not worn to armor worn.

The number of hits on the body armor and helmet is shown in the upper half of Table 3. In general, one would expect the quantity of hits on the helmet and body armor to be large. Out of 139 casualties wearing body armor, more hits should have been on the armor. Likewise, there were 585 casualties known to be wearing a helmet; this is 80% of the 721 cases which contained the body armor set. Using information compiled from the analysis program, it was found that the average number of hits per casualty was 3.4. Using this information, the following table may be constructed:

	B.A.	Helmet
No. Known to be Wearing Equipment	139	585
Avg. No. Hits Per Casualty	3.4	3.4
Expected No. of Hits	473	1989
% Body Area Covered by Protective Gear	23%	7%
Expected No. of Hits on Equipment	109	139
Actual No. of Hits on Equipment	49	66
Actual No. as Percent to Expected No.	45%	47%

From these calculations, it is concluded that more than half the armor and helmets are not available for examination or that hits are not noticed on the equipment examined (e.g., in a helmet which is badly battered from driving tent stakes or armor with worn or frayed spots concealing small hits).

The lower half of the table shows the quantity and boot type worn. The table for the "no boot" category suggests that during the shelling of base camps many troops are in bed or relaxing with their boots off.

III. TWO-WAY CORRELATIONS OF FACTORS. In addition to simple enumeration of the various factors studied, correlations between pairs of factors were also found, an example of which is presented in Table 4. The correlation is between wound location by six body areas and activities accomplished or not accomplished. The intent in gathering this information was to explore the relationship between wound location and incapacitation. The speed with which the casualties are evacuated seldom leaves the soldier time to attempt any task. The data seems to show that the soldier seldom tries something he cannot do; as activities accomplished outnumber those not accomplished by over nine to one.

Another two factor correlation, weapon versus location of hit, is contained in Table 5. The right hand column is the total of the row for each weapon. To circumvent the overwhelming quantity of numbers, two major groups of weapons were extracted to make the last two rows. These two groups will be referred to as rifles and fragments hereafter. To further simplify matters, the wound distribution for rifles and fragments is transformed into percentages in Table 6.

Table 6 also shows the distribution of wounds of body area correlated with a number of other factors. The first column shows the wound distribution compiled by accumulating the total number of hits in a body area then transforming those totals into percentages. The second column was compiled by accumulating presence of a hit in each body area over all casualties then transforming these totals into percentages. For instance, if a casualty received twelve (12) hits in the head or thorax, the sample number under "total number of hits" would be increased by 12; however, the sample number under "presence of a hit" would be increased

by 2, one for presence of a hit in the head and one for presence of a hit in the thorax. As is evident from a comparison of the percentages in these two columns, there is no great difference between the two methods of accumulating the wound distribution. Evidently, the body area which receives the most hits is also the area most likely to be hit (one or more times). The one possible exception is the lower extremity. When it is hit, the lower extremity seems to get more hits than other areas. This could be because the lower extremity tends to have more shielding from fragmenting munitions than the upper parts of the body, so when a fragmenting munition does detonate near enough to the man that the legs are exposed, the probability of multiple hits, especially to the lower extremity, is quite high.

The next pair of columns was derived from Table 5 where the numbers of hits by rifle bullets and fragments have been converted into percentages. The percentages for fragmenting munitions are almost identical to those of presence of a hit (by any weapon), but the increased percentage of hits in the combined head, neck, and thorax areas for rifles might well be an indication of aimed fire.

The next group of correlation in Table 6 shows wound location versus three categories of body position, upright, "doubled-up," and lying (which is 90% prone). Percentages do not differ enough, column to column, to be highly significant. However, in each case the small difference is in the direction which one would expect. For instance, in moving from the upright into a doubled-up position the head, neck, thorax, and upper extremities do not change in presented area; however, the lower abdomen, pelvis, and lower extremity are those parts which are doubled-up, proving shielding to each other, and thus losing presented area. The relative percentages of hits under these two areas reflect these observations.

The mean presented area of the head and neck to horizontal hits is about 6.5% of the total presented area of the upright man. Why, then, is there such a large percentage of hits on the head and neck of the man in an upright position? Part of the reason has already been mentioned: in a ground burst the fragment sprays will be limited by earth, irregularities in the surface flora, stones, or other low-flying cover, so that there will be some angle with the horizontal below which few or no fragments will be found. Thus, the upper parts will receive more hits, on the average, than their mean presented area warrants. If, on the other hand, the fragments are from direct fire artillery or proximity fuse munitions the burst is considerably above the ground. In this case, the presented area of a man is like his appearance if one is standing on a building looking down at him. The majority of his presented area in this case is head, thorax and shoulders. For the soldier in the lying position, cover is a major factor in wound distribution. More cover is offered near the ground, and that is usually the reason the man is lying there in the first place, to take advantage of whatever cover is available. Therefore, we would expect that with this highly variable factor, the

wound distribution would be very erratic, which it is. We also would expect that the tendency toward greater numbers of hits in the higher parts of the body would disappear, which it does.

The last three columns of Table 6 correlates injury type, fatal and non-fatal and cause of death to body area wounded. As expected, for the WIA's the larger proportion of hits occur on the least vulnerable parts, the extremities; while for the KIA and DOW's the larger percentages occur in the head, neck, and thorax. The last line in the table shows that the average casualty received 3.50 hits in 1.82 body areas. Since each casualty received (on the average) several wounds in two body parts, the distribution of wounds in the KIA and DOW's does not show the true distribution of cause of death. For instance, a fatality may have a bullet wound of the leg, but it was the bullet through the heart that killed him. When only the wound causing death is considered, the last column of Table 6 results.

In Table 7, the rifle bullet and fragment wound distributions of Table 6 are further broken down into hits on the front or back of the body. The purpose therein was to determine if hits about the body are truly random for bullets or fragments. The differences which stand out in this comparison are the front and back of the head and neck for fragments and the front and back of the upper extremity for bullets. Considering the latter first, we note that the body diagrams used in this study consider the man to be standing in the standard anatomical position; i.e., with the palms of the hands facing forward. In the battlefield, the soldier can be envisioned to be holding his arms in almost any other position rather than the standard anatomical position, resulting in considerable ambiguity in what is the front and what is the back of the arm. As a matter of fact, the palms are usually turned toward the body with the result that a large part of what is called the front of the lower arm is usually shielded from being hit by the trunk.

The much lower incidence of fragment hits on the back of the head and neck, mentioned earlier, is due to the ability of the helmet to defeat most incoming fragments and the much greater shielding that the helmet provides the back of the head and neck. When the effect of helmet protection from fragments and the ambiguity of which is the front or the back of the arm are removed, the lower part of Table 7 follows. Here it is seen that there is an insignificantly small difference between the percentage of fragments striking the front or back of the soldier. For bullets, on the other hand, a difference was expected because of the highly directional nature of rifle fire; viz., in a fire fight the troops are usually facing each other across some kind of battle front.

Table 8 shows the number of injuries and fatalities due to enemy fire as a function of weapon. The first two columns contain the numbers of fatal and non-fatal cases while the third column lists the ratio for

the two. The quantity NF/F will be referred to as the "survival index." Thus, weapons such as rifles and the Claymore mine have a low survival index (or, conversely, a high fatality rate). The recoilless rifle and large land mine display a moderate survival index while most other fragmenting munitions show a moderate to high survival with hand grenade and artillery showing a very high index. In the lower part of the table, the rifle and fragment combination groups are listed.

IV. THREE-WAY CORRELATION OF FACTORS. Figures 1 through 6 all show the results of one correlation of three factors, injury type (KIA, DOW, WIA, etc.), weapon, and distance between the casualty and weapon (or detonation). Figure 1 displays graphically the cumulative distribution of fragment and rifle bullet injuries regardless of injury type. The curves show that fragmenting munitions are much shorter range weapons than rifles, as would be expected. To quantitate this, note that 90% of the fragment wounds occur at 40 meters or less whereas 90% of the bullet wounds are accumulated only at 160 meters. Figures 2 and 3 split the data in Figure 1 into fatal and non-fatal wounds. Ninety percent (90%) of the fatal fragment wounds occur at ranges less than 30 meters. Ninety percent (90%) of fatal bullet wounds occur at ranges of 125 meters or less. Figures 4 and 5 show the same data presented in a slightly different manner.

Figure 6 shows a slightly different method of presenting a three-way correlation. The factors correlated are weapon, range, and number of hits a casualty received. Specifically, the curve in Figure 6 is the cumulative distribution of number of hits for fragments. The family of curves represents categories of range. Similar data was obtained for bullets, but the sample sizes were too small to give a coherent form to the curves. These last five curves indicate the difficulty one has in displaying multiple correlations even when the weapons were combined into only two categories. Presenting a full four-way correlation in a reasonable space borders on the impossible.

In the future, as more cases are entered on punch cards, a similar "analysis" will be conducted using more factors and more extensive correlations. It is hoped that sample sizes will be large enough, and definite enough, to clear up the inconclusiveness in some of the data presented here.

Frequency of Injury Type Sustained

Injured as a Result of Hostile Action (IRHA):

Killed in Action (KIA)	195
Died of Wounds (DOW)	27
Wounded in Action (WIA)	580

NonBattle Injury (NBI):

Killed in Action (KIA)	18
Died of Wounds (DOW)	4
Wounded in Action (WIA)	80

TABLE 1

Body Armor Worn:	Frequency
Army 12-ply Ballistic Nylon; M52-A, -2A, -3A	80
Marine Ballistic Nylon and Doron; M1955	5
Army Aircrew Small Arms (ceramic fiberglass), front plate	19
Same as above plus back plate	8
Army Infantry Frag (ceramic fiberglass); NTK T66-1	1
Army Lightweight Nylon Felt; T66-1	1
Army Titanium/Nylon Composite; T65-1	0
Other Types	11
Unknown Type	14
Total	139
No Body Armor Worn	565

TABLE 2

No. of Hits on Body Armor and Helmets

		No. of times hit			
		1	2	3	more than 3
Body Armor	- not perforated	5	0	1	3
	- perforated	13	5	1	0
	- perforated w/spalling	0	0	0	0
Helmet	- not perforated	5	2	1	3
	- perforated	16	5	0	1
	- perforated w/spalling	4	2	0	0

Frequency of Boot Types

Stitched or molded cloth or leather boot	507
Boot with Steel Insole	119
Boot with Protective Shank	34
Mine Clearance Sabot	0
Overboot with Protective Features	0
No Boot	44

TABLE 3

Wound Loc.	Crawl		Stand		Walk		Run		L. Weapon		F. Weapon		T. Grenade		Perf. Exp. Norm. Duty	
	A	NA	A	NA	A	NA	A	NA	A	NA	A	NA	A	NA	A	NA
Head, Face, Neck	21	2	32	6	29	3	12	1	8	-	14	1	2	1	17	-
Thorax, Shoulders	26	3	48	7	45	4	15	2	5	1	7	1	3	-	13	1
U. Abdomen	8	3	12	3	12	2	4	2	4	1	2	1	1	-	3	2
L. Abdomen, Pelvis Hips, Buttocks	14	2	19	3	19	1	4	-	6	-	7	-	3	-	4	2
L. Extremity	42	4	51	8	54	8	13	6	12	-	21	1	4	2	21	2
U. Extremity	39	4	53	6	56	4	19	1	18	1	21	1	7	-	27	2

A = Accomplished
NA = Not Accomplished

TABLE 4

TABLE 5

Presence of a Hit:

Weapon:	Head, Face, Neck	Thorax, Shoulder	Upper Abdomen	Lower Abdomen Pelvis, Hip, Buttocks	Thigh, Knee Leg, Foot	Arm, Elbow Forearm, Hand	Total
Rifle - M16	0	3	0	0	5	1	9
Rifle - M14	0	1	1	0	0	0	2
Rifle - AK47	11	17	2	6	12	11	59
Rifle - Other	41	45	11	9	39	31	176
Pistol	3	3	1	1	10	6	24
Machine Gun	2	2	2	2	7	7	22
Improvised	31	37	12	30	48	41	199
*Wpn Launch Grenade	20	22	7	9	31	21	110
*Hand Grenade	32	46	20	24	55	58	235
*Mortar	31	31	9	15	26	23	135
*Claymore	1	8	4	7	9	7	36
*A-P Mine	0	1	0	1	7	2	11
*Large Land Mine	5	5	2	1	9	6	28
*Artillery	7	15	3	6	16	8	55
Bomb	0	2	0	0	3	3	8
Blade	0	0	0	0	0	0	0
Puncture	0	0	0	1	3	1	5
Toxic	0	0	0	0	0	0	0
Flame	1	0	0	0	0	3	4
Other	47	57	20	30	57	50	261
*Recoilless Rifle	5	6	2	6	12	7	38
Total	237	301	96	148	349	286	1417
Rifle, AK47, Other	52	62	13	15	51	42	235
*Fragment	101	134	47	69	165	132	648

Wound Distribution: Percentage of Hits in the Various Body Areas

	Total No. of Hits	Presence of a Hit	Correlations with Presence of Hit: Fragments	Rifle: AK47 and Other	Body Position (Upright)	B.P. (Sitting) (Crouching) (Kneeling)	B.P. (Lying)	IRHA-Fatal (KIA+DOW)	IRHA-Nonfatal (WIA)	Cause of Death
Head, Face, Neck	15	18	16	22	14	18	11	23	16	44
Thorax, Shoulder	20	21	21	26	18	21	17	27	18	39
Upper Abdomen	5	7	7	5	7	4	5	8	7	8
L. Abd, Pelvis, Hips, Buttocks	9	11	11	6	11	7	14	12	11	4
Thigh, Knee, Leg, Foot	33	24	25	22	28	23	36	16	27	5
Arm, Elbow, Forearm, Hand	18	19	20	18	22	27	18	14	22	0
Total %	100	100	100	99	100	100	101	100	101	100
Total No. of Sample	2,978	1,548	648	235	502	228	148	485	886	201
No. of Casualties	850	850								
Ratio	3.50	1.82								

TABLE 6

Wound Distribution: Presence in Various Body Area Versus Weapon

Body Area Hit	Weapon: Fragment, Front #	Fragment, Front %	Fragment, Back #	Fragment, Back %	Bullet, Front #	Bullet, Front %	Bullet, Back #	Bullet, Back %
Head, Face, Neck	88	21.52	28	9.00	37	22.84	18	18.37
Thorax, Shoulder	78	19.07	72	23.15	44	27.16	24	24.49
Upper Abdomen	25	6.11	24	7.72	11	6.79	2	2.04
L. Abd., Pelvis Hips, Buttocks	38	9.29	34	10.93	9	5.56	6	6.12
Thigh, Knee, Leg Foot	104	25.43	85	27.33	38	23.46	25	25.51
Arm, Elbow, Forearm Hand	76	18.58	68	21.86	23	14.20	23	23.47
Total	409	100	311	100	162	100	98	100

Note: Eliminating the Bias Due to the Helmet Protection From Fragments and the Ambiguity of Front or Back Arm Hit, the Following Distribution Results:

	Fragment, Front #	Fragment, Front %	Fragment, Back #	Fragment, Back %	Bullet, Front #	Bullet, Front %	Bullet, Back #	Bullet, Back %
Total	409	51.9	379	48.1	162	62.31	98	37.69

TABLE 7

No. Wounded as a Result of Hostile Action

Weapon:	Fatal	Non-fatal	NF/F
Rifle - M16	2	1	0.50
Rifle - M14	1	0	-
*Rifle - AK47	21	18	0.86
*Rifle - Other	47	73	1.6
Pistol	2	3	1.5
Machine Gun	0	10	-
Improvised	16	67	4.2
**Wpn. Lnch. Grenade	9	48	5.3
**Hand Grenade	14	90	6.4
**Mortar	11	44	4.0
**Claymore	6	6	1.0
**A-P Mine	0	7	-
**Lg. Lnd. Mine	4	10	2.5
**Artillery	3	22	7.3
Bomb	0	4	-
Blade	0	0	-
Puncture	0	4	-
Toxic	0	0	-
Flame	0	1	-
Other	37	91	2.5
**Recoilless Rifle	7	15	2.1
Total	180	515	2.9
*Rifles	68	91	1.3
**Fragments	54	242	4.5

TABLE 8

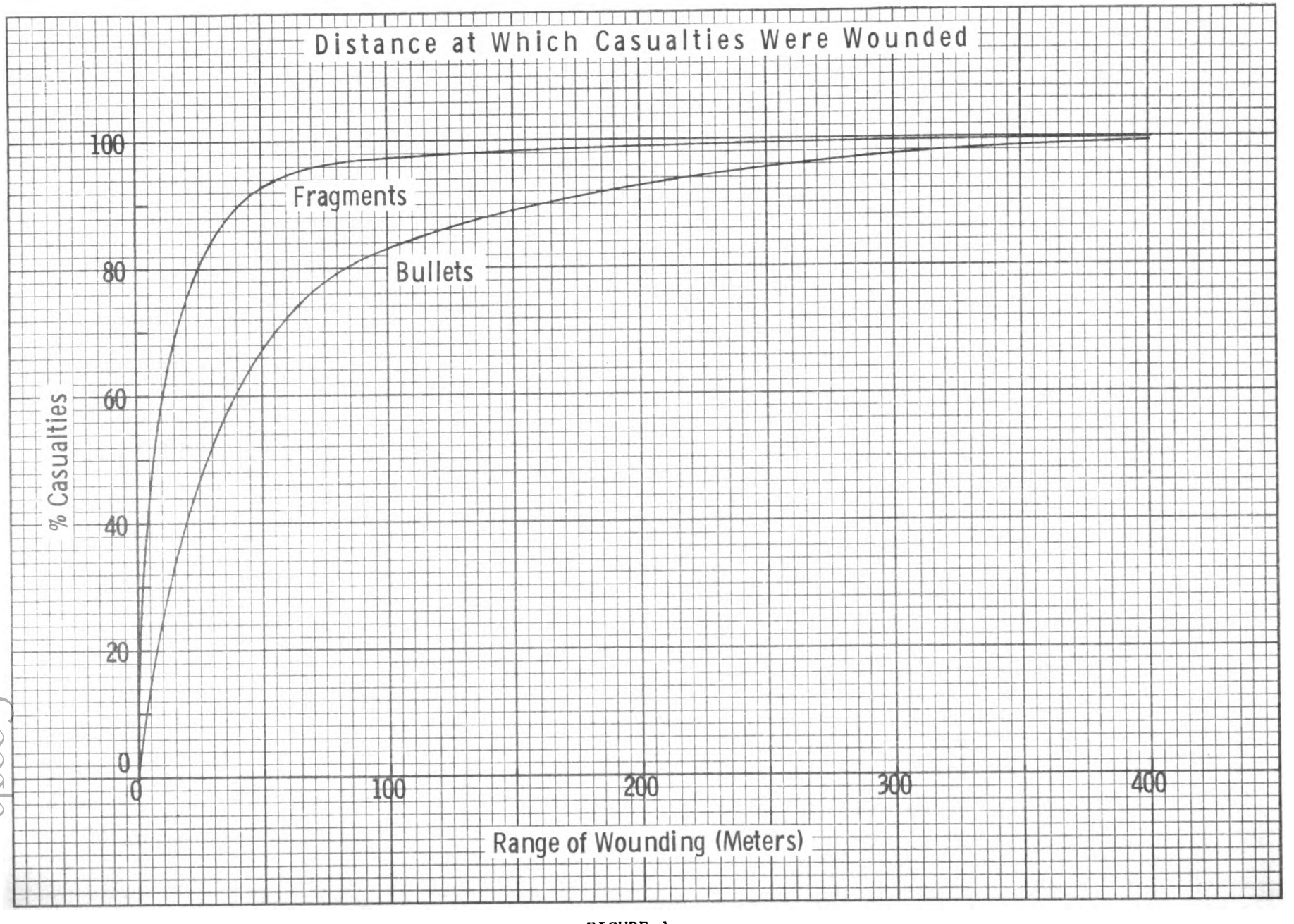


FIGURE 1

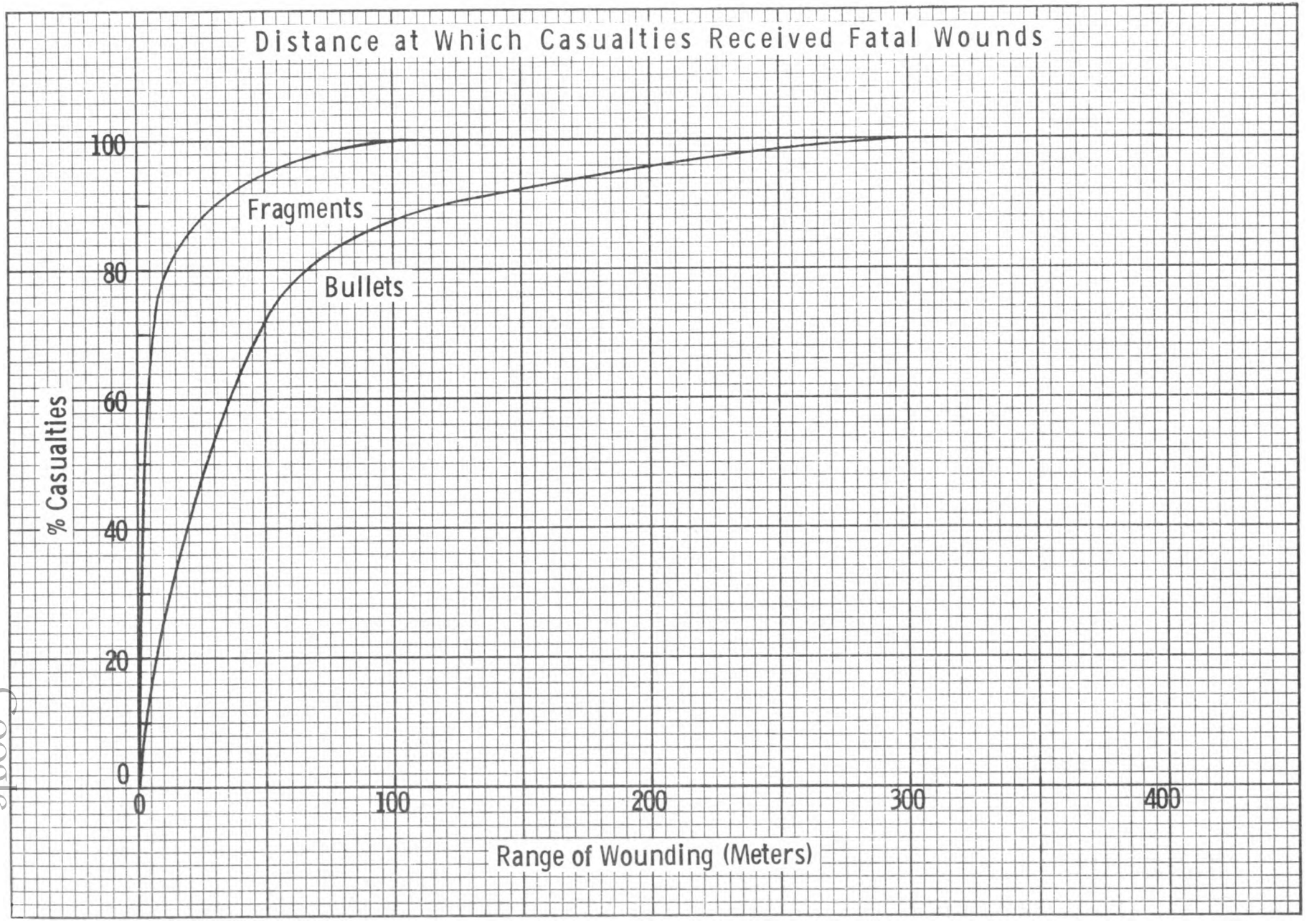


FIGURE 2

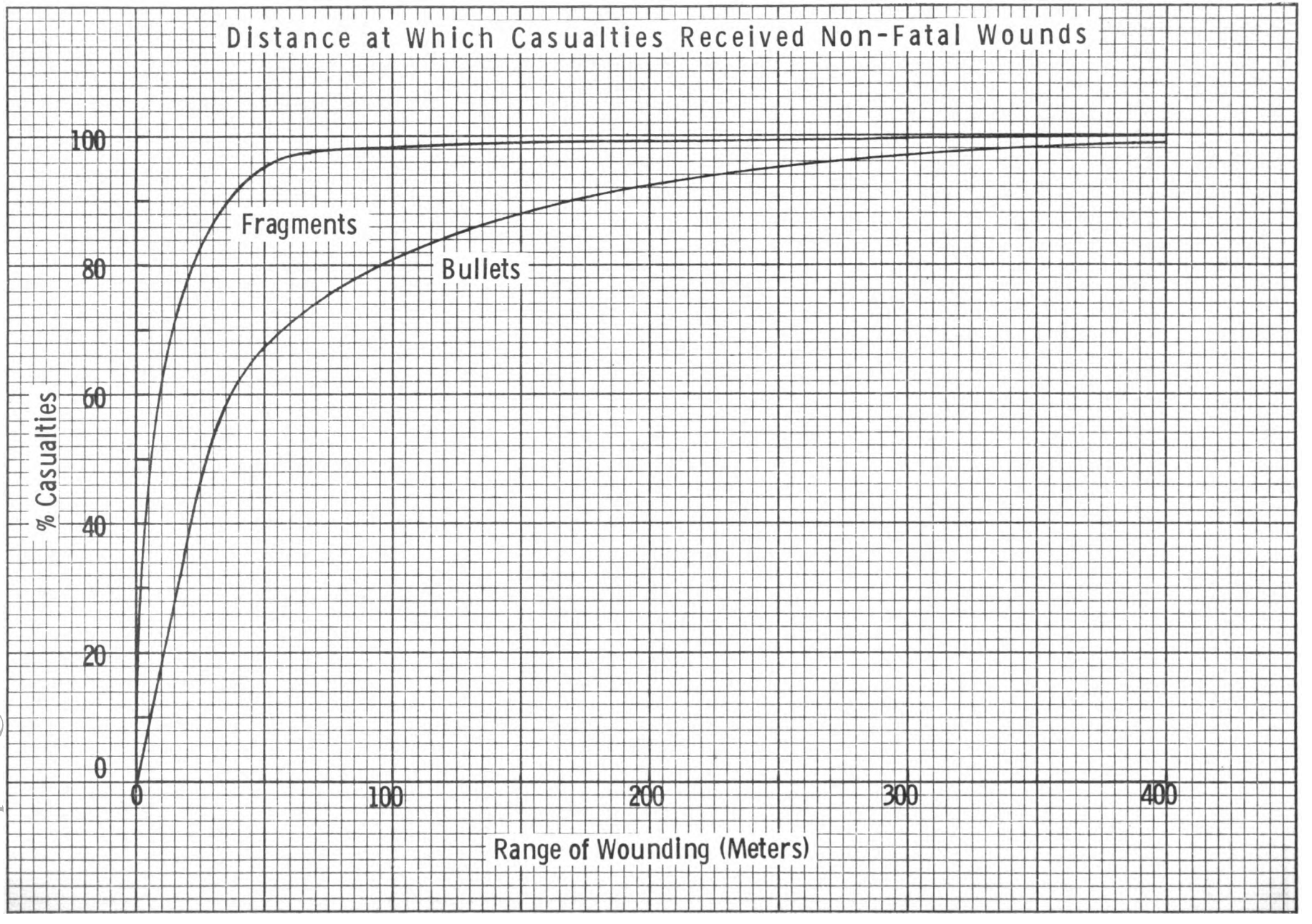


FIGURE 3

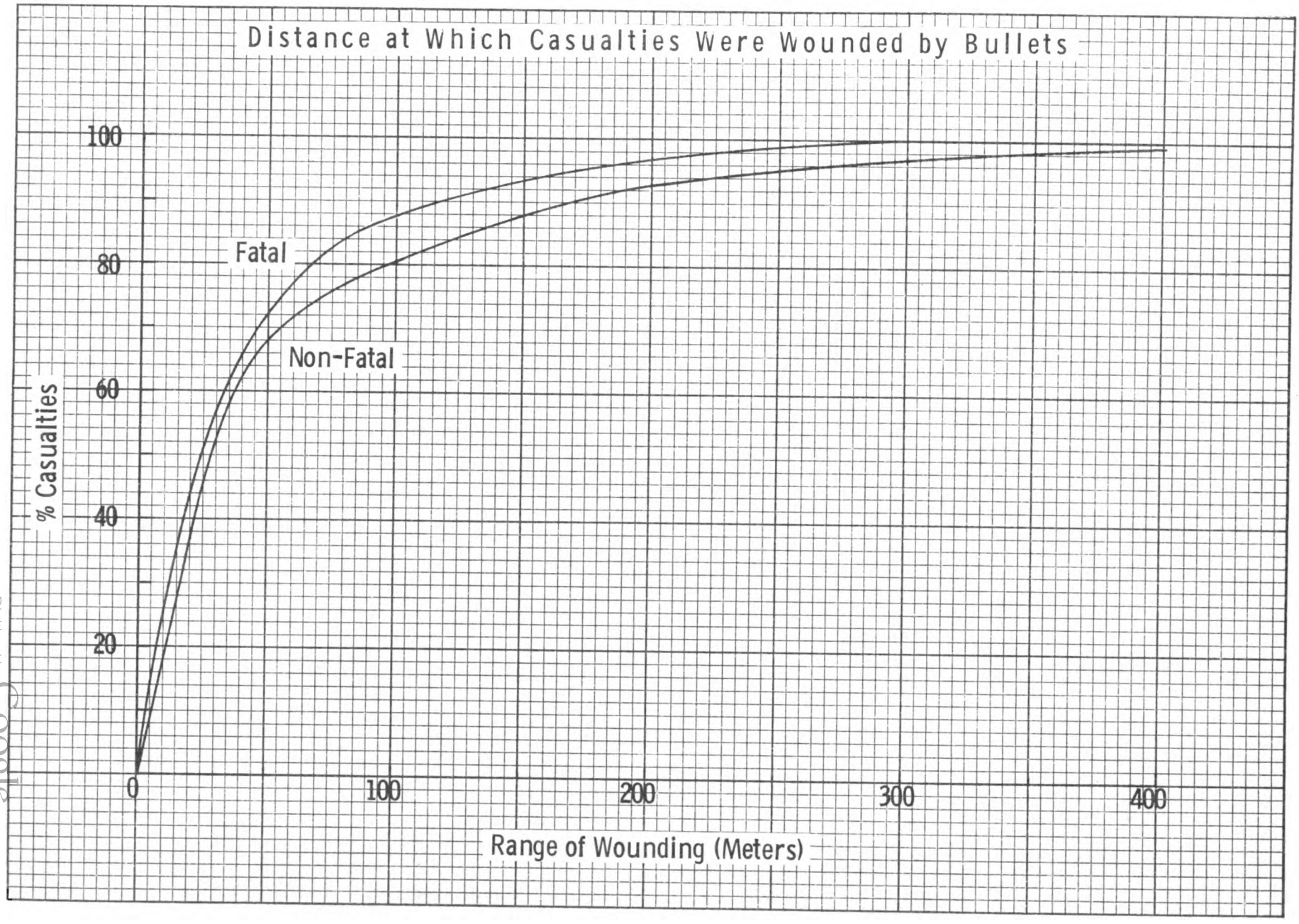


FIGURE 4

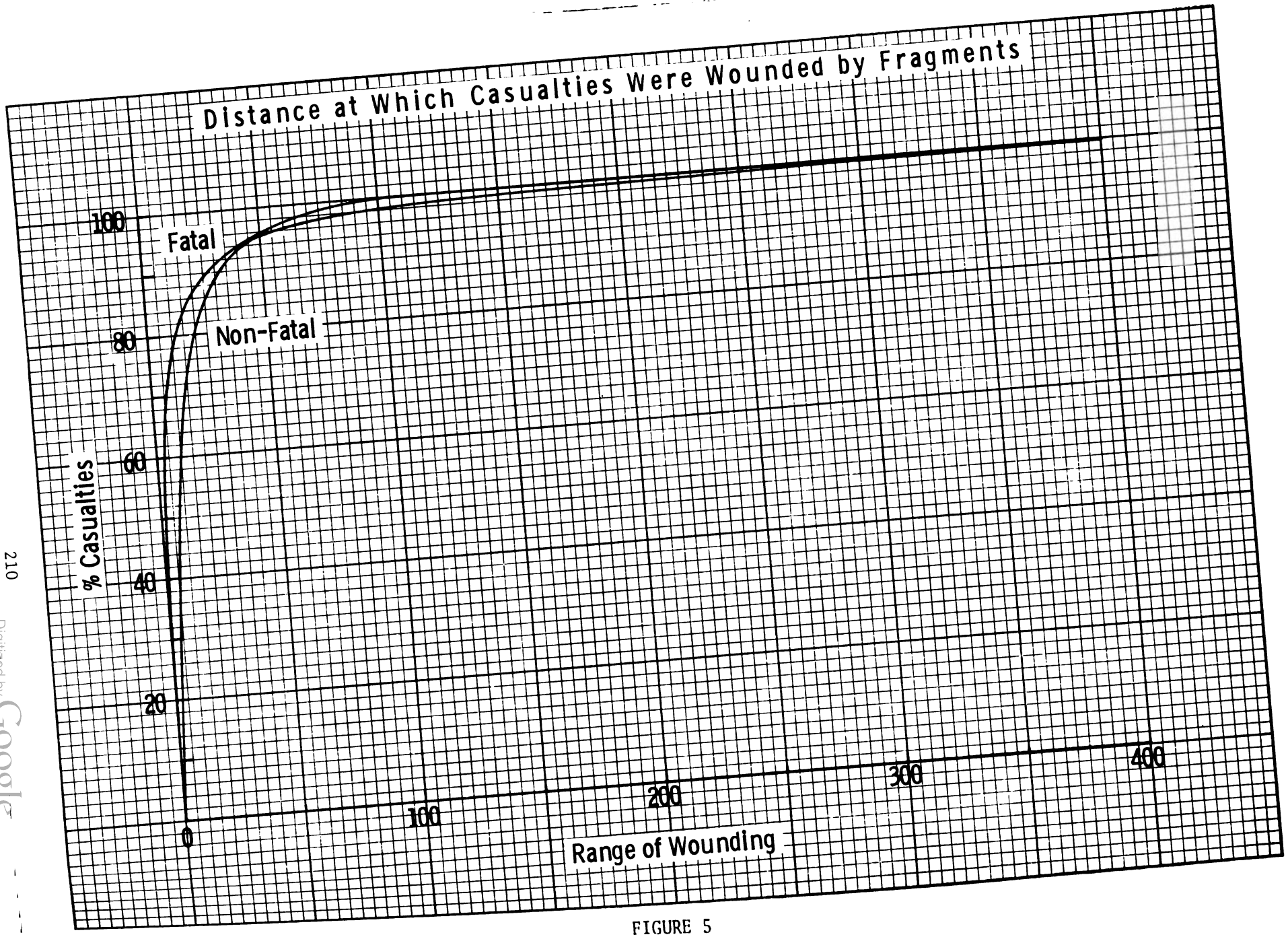


FIGURE 5

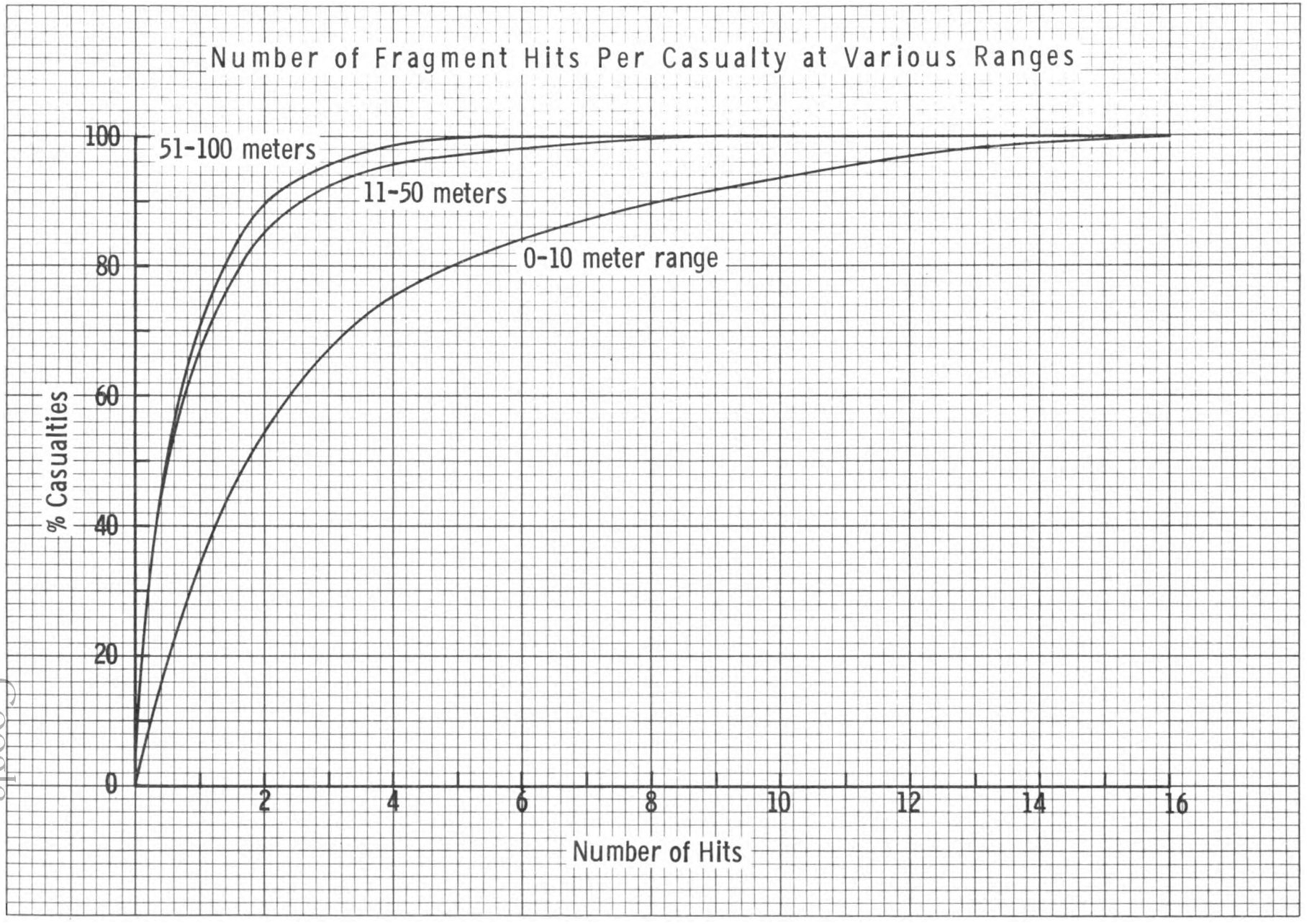


FIGURE 6

EFFECT OF NUMBER OF OBSERVING STATIONS ON FLIGHT MEASUREMENT PRECISION

Fred S. Hanson
National Range Operations
White Sands Missile Range, New Mexico

ABSTRACT. Estimation of the standard error of a measured space-position is reviewed. Pooling such standard deviations for the portion of a trajectory covered by a given measuring system - and for a series of tests on the same missile - is discussed. Results are presented showing the dependence of average position-precision on number of stations used in the solution. The correlation of these variables in operating data is dominant and the magnitude of the effect is profound. The *exponential* improvement of position-precision by increasing stations can be as much as four times the effect of increased sample-size on the standard-error-of-the-mean of a normal distribution. Mechanisms considered embrace: geometric convergence, observational constraints, methodological deficiencies, and statistical considerations. The exponential dependence of position-precision on number of cinetheodolites may be an index of the measurability of the object ('readability' of its point-of-reference). Statistical measures-of-goodness of geometric convergence are derived. A procedure is suggested for rating test-configurations. It is shown that calculating observationally-redundant precision of *nonredundant* solutions is a generalization of the classical calculation of the precision of single observations from the precision-of-the-mean of a sample of a given size. A need is suggested for a statistics of observations which define geometric surfaces in space. (This may be a generalization of numerical statistics.) Results are also presented showing the dependence of precisions of derived velocity and acceleration on number of stations. A probabilistic improvement of physical accuracy (bias) by increasing stations in flight-measurement is hypothesized. A summary is appended.

INTRODUCTION. Our reviews of over- and under-meeting of quality requirements made it necessary to investigate the relationship between quality and resources in our data-support operation.

We had been aware of the statistical improvement of precision in which ordinary averages bunch closer together in proportion to the square root of the number averaged.

Much of the following work was published in internal memoranda during the fall of 1967.

BACKGROUND. This paper is clinical in the sense that it is exploratory.

Since January 1963, White Sands Missile Range has built a sufficient, standard basis for a data-precision spec into its user-document format - in the interest of comparability. (We say what we mean by our numbers, and that we assume the user's numbers mean the same thing - unless he makes it very clear otherwise.) WSMR-standard precision is the average standard-error of component values of data, obtained by propagation from the previous stage of the collection-reduction process. This precision-index can be tied back directly to station quality, and film-reader quality. It applies to the data in the form in which it is reported.

Our data-precision is many things. It's the radius of confusion of a data value, due to the disagreement among the stations. It's how well we can know from the observations what the value is. When related to a valid requirement, it's a measure of Range effectiveness. Precision is available in current operation. It affords a means for operational control, and suffices for some of the user's needs. Apparently, in flight-measurement, optimizing system precision *tends* to optimize system accuracy. We furnish our users the precision of each data value. And, we use root-mean-square *average* precisions - by segment, by test, by month, and by program - for operating and management control.

I gave a historical and exploratory paper on data-support quality control three years ago at the Design of Experiments Conference (Ref. 1). Our monthly Data Quality reports (Ref. 2) give actual average precisions - by station, by measuring system, and by missile - along with the requirements and commitments. Averages are monthly, and cumulative for the fiscal year. Being definite and quantitative, keeping usable scores on data quality, and controlling closely on the basis of results are a bit of a departure from missile-range tradition. Personally, I feel it is to the *advantage* of mission personnel to provide management quantitative bases for decisions.

ESTIMATION OF PRECISION. Figure 1 is a summary of the math we use to calculate precision of observed position for *cinetheodolites.*

This is from our Data Reduction Handbook - 'R. C. Davis' method (Ref. 3). We solve for position by minimizing the sum-of-squares of the deviations of the stations in azimuth and elevation, from their least-squares point. In the first equation, $\cos \epsilon_i$ allows for the fact azimuth circles get smaller as one goes up - until 'the universe comes to a point directly over each of our stations'. As the azimuth circles get small we lose resolution; the *azimuth* error becomes something between ungodly and unknown. So, we temper it by the cosine of the elevation angle[1]. We average the angular deviations as their squares. The square root of

[1] Reference 14 is a formal explanation of this correction.

that (σ_A) is our measure of the *average* disagreement among the angular observations. Use of the 3-dimensional degrees-of-freedom (2N-3) yields an estimate of the population standard deviation of angular observations (of what it would be if we could repeat them indefinitely).

Now, we least-squares in 3 dimensions. But, forever after we treat the components separately; as though they had been independently determined (we'll come back to that). Multiplying the variance of the *angular* deviations by the cofactor of the proper element of the (principal) diagonal of the determinant of the least-squares matrix, then dividing by the (value of the) determinant is said (by a liberal interpretation) to transform the angular variance to a linear-component variance of the least-squares position *mean*. It is planned to *verify* this last point by: converting angular residuals to their linear-component equivalents; calculating standard-deviation *and* standard-error-of-mean of each set of these (about the least-squares mean); noting which is closer to the matrix result.

We finally multiply each standard deviation by the proper value of the 't' statistic, to correct for the relatively small departure from normality at the 68.3% probability level due to the small sample-size alone (to the small degrees-of-freedom). Additionally, we usually obtain a precision of *smoothed* position by dividing the standard-deviation of a series of points, about its 2nd-degree fit, by the proper reduction-factor in terms of the number-of-points. (We are not yet incorporating lack-of-fit into smoothed-position precision.) Our precisions of velocity and acceleration are obtained by propagating smoothed-position precision thru the lst and 2nd derivatives of smoothed position.

AVERAGING PRECISION. It is physically necessary to describe measurement quality in terms of frequency distributions. It is operationally and managerially necessary to describe data quality in job lots (segments or tests) - and *wholesale* (series-of-tests). Since January 1963, WSMR has officially defined data quality as the average precision for the firings covered by the documentation.

When a requirement or commitment is met as an rms-average, approximately 68.3% of the data values fall within the stated tolerance of their statistically-true values. (i.e., When compliance of the individual standard deviations is 50%, average compliance of the data values which they *characterize* approximates 68.3%.)

In root-mean-squaring a component precision for a segment or a test, our denominator is the number of component values. Then, our test-average quality is the root-mean-square of the 3 test-average component precisions. (Yielding the radius of that sphere which conventionally approximates the average error-ellipsoid.) In our monthly and cumulative project averages, tests are given equal weight.

What constitutes a statistical population is an *operational* decision. Our average precisions are calculated by the same procedures each time; so they have an operational validity (we're not in the rigor business). We *are* interested in knowing the magnitude and direction of the errors resulting from our nonivorytower applications of ivorytower methods.

One purpose of statistics is to numerically characterize errors. This paper suggests that includes numerically characterizing the errors incurred in *applying* statistics to the real world.

Let's look at some operational findings.

EFFECT OF NUMBER OF STATIONS. In this investigation, test-average precisions were sorted by average-number-of-stations (to the nearest integer) in the solution. Each group of precisions was then root-mean-squared; the denominator being the number of rounds.

Figure 2 shows average precision of position of Navy bombs vs average number of Askania (cinetheodolite) stations computed (Jan-Jul 1967). The horizontal ticks are the plotted points. Numbers *on* the graph are the number of rounds (tests) represented by each plotted point. The number of position-points in one of our tests varies widely. Typically, it is a few hundred, times 3 components. The Navy bomb-drop is a highly diversified program. In a manner of speaking, the range and user 'did their worst'; but the average quality depended on only one variable (except for the limited 5-station data). *That* data is shown both as is, and after deleting the worst round. It may indicate they were running out of reasonably well-located stations for covering the *bomb*, which impacts the ground.

GENERALIZATION OF MODEL. Please turn to Figure 3. This merely looks hard. It's only one equation (equation (5)) - transformed (equation (1)) - and generalized (in both forms). If our only effect were *numerical* redundancy, position quality would improve in proportion to the square-root of the number-of-observations (Ref. 4). Or, to the square-root of the number-of-stations, since the number-of-observations per station is constant for a system. Equation (5) is the basic form, for optics, of this classical theorem (for x normally distributed). For radar the 2 becomes 3; for DOVAP it becomes 1.

When the only effect is *number* of observations, the population standard deviation of individual observations (σ_x) is of course unchanged. So, if equation (5) is used twice in constructing a curve, it reduces to its working form, equation (1). You might guess that (6), (7), (8) and (2), (3), (4) were *empirically* derived by generalizing the exponent in (5) and in (1). Equations (5), (6), (7), and (8) are hyperbolas of the respective types:

$$x^{1/2}\ y = k,\ xy = k,\ x^{3/2}\ y = k,\ x^2\ y = k.$$

In Figure 4, the operational curve is the solid one. The working form of the classical-statistical equation was used both ways from a midpoint to construct the curve shown as a string-of-beads. This didn't do the job so we had to look further. The upper half of the Navy-bomb curve is closest to improvement of precision as $N^{3/2}$. The lower half is closest to N^2. Overall, it's a tossup between those two. Physical interpretations of these 3 hyperbolas are *possible* at each ½-station, for optics. At 1-½ stations, the curves represent standard deviations of *single* observations of a point-in-3-dimensional-space. The interpretations of infinitely-poor precision for zero stations and of infinite stations for perfect precision are obvious.

VARIABILITY AT A GIVEN NUMBER OF STATIONS. Curvilinear correlation coefficients might be calculated for equation (7) applied to the upper half of the Navy-bomb data and for equation (8) applied to the lower half. Differences between these and unity would estimate the *relative* influence of all factors other than number-of-stations. Further, the standard deviations of individual-round precisions about the pooled values, at each number of stations, could be calculated from the round-average data. Or, they could be calculated about the corresponding points on the fitted curves. These sigmas could be used to set current individual-round tolerances for the models, at each number of stations (it is planned to use this approach).

The corresponding tolerances of pooled (cumulative) precisions could be estimated as the appropriate standard deviation of individual rounds divided by the square root of the number in the average. In calculating correlation of round-average precision, or in controlling a cumulative average, the average number of stations should be carried to the first decimal place (it is *available* to 4 places).

Dr. H. H. Germond suggests plotting average precision vs number of stations on log-log paper. That, in this way, the data may be fitted with straight lines whose slopes are negatives of the corresponding powers of N of the hyperbolas (Ref. 5). Straight lines fit people better. This investigation has emphasized direct study of the relationships. Linear correlation coefficients would of course apply to log precision vs log number-of-stations, rather than to precision vs number-of-stations.

FURTHER DATA. Figure 5 shows average precision of position of Navy aircraft vs average number of Askania stations computed (Jan-Jul '67). The curve is less steep than for the bombs. There is no clear indication they were running out of stations well-located for covering the *aircraft*. Figure 6 shows that, for the aircraft, precision came closest to improving in direct proportion to number-of-stations.

Mr. Frank Hemingway suggested we look at the vertical component separately. In Figure 7, the dashed curve is the average quality of Navy aircraft x, y, and z from Figure 5. The solid curve is the precision of the vertical component only. From inspection of other WSMR cinetheodolite data, this better precision of the vertical component appears to be a general result. Without z in the composite, the difference would be half as much again. Please note that these are parallel, except that the z-curve is a bit flatter near the right-hand end.

Figure 8 shows the effect of number-of-stations on Redeye Contraves (cinetheodolite) precision-of-position was dominant and profound, even with less data (Jan-May '67) than on the Navy tests. In Figure 9, the upper part of the Redeye curve was about a tossup between $N^{3/2}$ and N^2. The lower part was N^2. I gave N^2 a little edge, overall.

Figure 10 shows the precision curve was slightly less steep for the Redeye *target* (Jan-May '67). In Figure 11, this relationship was a tossup between N and $N^{3/2}$ for the upper half, $N^{3/2}$ lower and overall.

Figure 12 shows precision of Askania position-measurement on the PEARL aircraft radome (Jan-Jul '67) improved approximately as $N^{1/2}$, if we lightly regard the poorer 4-station data. If we delete the *worst* rounds, as indicated by Figure 13, the 3-station data was better than the trend of the rest; the 4-station data was on the curve; the upper half improved only as $N^{1/3}$.

In Figure 14, if we didn't take our result too seriously, we might approximate this Redeye fixed-camera data (Jan-May '67) by the solid curve - which turned out to be nearest to improvement of precision in direct proportion to number-of-stations.

Figure 15 shows a similar situation held for this limited DOVAP data on Lance (Oct '66 - June '67). Because of the very small amount of data, it appeared desirable to *also* look at 2 of the averages on the assumption that they might not be representative samples. In Figure 16, our solid approximation turned out to fall closest to improvement of precision in direct proportion to number-of-stations.

Preliminary indications were that the average standard deviation of least-squares-mean space-positions determined by FPS-16 radar typically improved as the 3/2 power of the number-of-stations.

Apparently, our cumulative-average component-precision of observed position converges rapidly to correlation with number of stations.

It was noted from plotting the foregoing data-curves on the same graph that the exponential rate of proportional (percentage) improvement of precision, with increase in number-of-stations, depends on the magnitude of the precision values as well as on the steepness of the curve. Smaller numerical values (better precision) require less numerical improvement (less steepness) for a given exponential rate of proportional improvement.

VALUE OF APPROACH. Findings of this paper are being used in management reviews to express over- and undermeeting of requirements in terms of resources (resource-equivalent ratios of precisions). The foregoing data-curves provide specific resource-capability relationships by project. These are relevant to data-support committing, control, and planning. The data will lend itself to further structuring of our capabilities by graphing in various ways. Also, to advancing the state-of-the-art and the state-of-the-understanding - as the following pages indicate.

SUMMARY FOR CINETHEODOLITES. Figure 17 is a summary for Askania and Contraves position-precision, based on all the cine-position data I have plotted to date.

The upper half of this table shows the (approximate) spectrum of dependence-of-precision on various powers of N which was produced by the interaction of cinetheodolite systems with various flight-measurement tasks. Relative point-of-reference difficulty *may* explain the broad precision-response spectrum of the cines. The PEARL radar pod is a large, black hemisphere without distinguishable markings. It is more difficult to establish a consistent reference on large aircraft than on small, and more difficult on small aircraft than on small or medium missiles or bombs. The exponential dependence is *apparently* an index of the measurability of an object.

Our management reviews show the factor of over- or undermeeting of requirements *in terms of* resources (stations) - where this is not identical with effective ratio of average-precision to its requirement. The proportionalities of numbers of stations to precision are, of course, the *inverses* of the upper table in Fig. 17 - as shown in the lower table. We use relationships specific to projects where they are available. If dependence must be obtained from the table, we show the resource-equivalent ratio in parentheses.

Now, it's easy to say precision can improve in proportion to as much as N^2 in flight-measurement *because* it's a 3-dimensional process. But what are the *mechanisms* by which this takes place - and what is the relative importance of each? We have turned over a few stones, and here is a list:

MECHANISMS.

1. The decrease through increase in sample-size alone of our uncertainty as to what the observations say the data value would be if we could repeat the same measuring process indefinitely. This is

the well-known statistical convergence (Ref. 4). It is generally accepted, locally, that our 'Davis' method for cine data effectively takes this into account.

So, $\bar{\bar{\sigma}}_C = f(N^m)$, where $m = f$ (sample size;

Let's call this one improvement by overcoming small sample-size.

2. Improvement, through increase in sample-size, of the *average* goodness of the intersection-angles of the lines-of-sight from the stations (less chance of only bad intersections). The relationship of intersection-angle to both precision and accuracy was *treated* in my earlier clinical paper (Ref. 6). The angle-of-intersection mechanism has *some* diminishing returns as the useful cones (with vertex at the missile) become divided into smaller, less desirable intersection-angles. In multi-station measurement, the projected intersections of each station-line with *each* of the others are relevant to linear precision.

Simulations, informally communicated by Mr. W. V. Hereford of Sandia Corp. (Ref. 7), showed a *first-power* improvement in rms position-error while overcoming his 'worst-case' geometries; only a half-power improvement while interacting with his 'best-case' geometries. (Gradual expanding of a narrow baseline *vs* gradual spacing-in of a wide one.) This establishes *an* effect of geometry on the power of the precision-response.

Multidimensional measurement depends on geometric convergence as well as on numerical convergence.

Figure 18 shows an acute convergence-angle (*smaller* than the 90° convergence which Reference 6 deduced to be the general optimum). Figure 18 can be any plane through 2 stations and the missile. Let the ± angular dispersion about the direction lines be an average angular standard-deviation (σ_A).

The diagonals of the smaller almost-diamond are its linear-standard-deviation subtenders of σ_A in the directions perpendicular to and parallel to the baseline.

Solving either error-triangle, SMO, which contains the *vertical* diagonal (given the angles and the slant range):

$$\sigma_\perp = \frac{r \sin \sigma_A}{\sin (180^\circ - \theta/2 - \sigma_A)} = \frac{r \sin \sigma_A}{\sin (\theta/2 + \sigma_A)} \approx \frac{r \sin \sigma_A}{\sin \theta/2}$$

The horizontal diagonal may be obtained from:

$$\cot \theta/2 \approx \frac{\sigma_{\perp/2}}{\sigma_{\parallel/2}} \approx \frac{\sigma_\perp}{\sigma_\parallel}$$

whence:

$$\sigma_{\parallel} \approx \frac{\sigma_{\perp}}{\cot \theta/2} \approx \frac{r \sin \sigma_A}{\sin \theta/2 \cot \theta/2} \approx \frac{r \sin \sigma_A}{\cos \theta/2} \approx \frac{r \sin \sigma_A}{\sin \phi/2}$$

where ϕ is the parallel convergence-angle.

$\sigma_{\perp}$ may be taken as a measure of the badness of the *convergence* for measuring *perpendicular* to the baseline. $\sigma_{\parallel}$ may be taken as a measure of the badness of the convergence for measuring *parallel* to the baseline. (They are inverse measures of the goodness of the convergence for these purposes.)

If Figure 18 is a vertical plane parallel to the baseline, $\sigma_{\perp}$ measures the goodness of the vertical projection of the convergence (in terms of the other projected quantities) for measuring the z coordinate of the missile. This measurement depends only on the elevation readings. Here, $\sigma_{\perp}$ *measures* σ_z (of linear observations [1a], not our standard precision of the least-squares mean). Its units are those of the slant range.

Per the first of the above 3 equations:

Vertical convergence *Precision* and *Variance* Factors

Vertical Convergence (θ)	$\frac{1}{\sin \theta/2}$	$\left(\frac{1}{\sin \theta/2}\right)^2$
180°	1.00	1.00
135°	1.08	1.17
90°	1.41	2.00
45°	2.61	6.82
15°	7.66	58.6
5°	22.9	524.
1°	114.4	13,120.

This table compares the goodness of vertical convergence at any constant value of projected slant range, which avoids confusing the effect of range on linear precision with the effect of convergence-angle. It indicates that the optimum *vertical* convergence, per se, is *not* 90°. That the most precise measurement of the *vertical*, as far as geometry goes, is when the object is *in* the line-of-sight between the stations.

Vertical convergence, in Fig. 18, can be easily calculated for a given case as:

$$\theta = 2 \text{ arc tan } \frac{\text{baseline}}{2 \text{ (altitude)}}$$

The keys to good vertical convergence are, of course, long baselines.

[1a] The mode of $\sigma_{\perp}$ is defined by the mode of σ_A (e.g., observations, mean, series, curve, etc.).

If Figure 18 is a horizontal plane, $\sigma_\perp$ measures the goodness of the horizontal projection of the convergence for measuring perpendicular to the baseline, and $\sigma_\parallel$ measures the goodness of the horizontal projection of the convergence for measuring parallel to the baseline. (When the baseline is east-west, $\sigma_\perp$ measures σ_x and $\sigma_\parallel$ measures σ_y.) Since the two horizontal measurements share their plane of projection, a gain in convergence for one is a loss in the other. In general, it's not sound practice to improve data in one coordinate by making it worse in another. So, the practical optimum *horizontal* convergence is 90°.

The separate formulas for $\sigma_\perp$ and $\sigma_\parallel$ are of relative value for horizontal measurement.

Otherwise, by the second of the above 4 equations:

Horizontal Convergence Precision And Variance Disparity

Perpendicular Convergence	$\frac{\sigma_\perp}{\sigma_\parallel}$	$\left(\frac{\sigma_\perp}{\sigma_\parallel}\right)^2$
179°	1/114.6	1/13,140
175°	1/22.9	1/525
165°	1/7.60	1/57.6
135°	1/2.41	1/5.83
120°	1/1.73	1/3.00
90°	1.00	1.00
60°	1.73	3.00
45°	2.41	5.83
15°	7.60	57.6
5°	22.9	525
1°	114.6	13,140

This table compares the disparity of the two horizontal convergences at any given ground range. The ratios of precisions approximate the ratios of width/length (or length/width) of the actual horizontal linear-observation-error *ellipses*. The ratios of variances approximate the ratios of areas of circles whose diameters are the two diameters of the ellipses. (These ratios would be the same for the least-squares means as for their linear observations.)

Perpendicular convergence in Fig. 18 is, of course, calculated for a given case as:

$$\theta = 2 \arctan \frac{\text{baseline}}{2\,(\text{perpendicular distance})}$$

One key to optimum horizontal convergence is the set of optimum configurations in Figure 19, which were demonstrated in Reference 6. The *least* that should be done is to compensate a narrow convergence in any horizontal direction with another that is roughly perpendicular to it.

The numbers in the above 2 tables are similar, but only the first table is a direct measure of loss. The second is the ratio of loss in one direction to loss in the direction at right angles.

A direct measure of the *combined* goodness of horizontal convergences is the root-mean-square average of their perpendicular and parallel precisions.

Per the first and third of the above 5 equations:

$$\sigma_H \approx r \sin \sigma_A \sqrt{\frac{\frac{1}{\sin^2 \theta/2} + \frac{1}{\cos^2 \theta/2}}{2}}$$

Horizontal Average Precision And Variance Factors

Perpendicular Convergence (θ)	$\sigma_{H/r \sin \sigma_A}$	$\left(\sigma_{H/r \sin \sigma_A}\right)^2$
179°	81.0	6561
175°	16.2	263
165°	5.46	29.8
135°	2.00	4.0
90°	1.41	2.0
45°	2.00	4.0
15°	5.46	29.8
5°	16.2	263
1°	81.0	6561

This table compares the goodness of horizontal convergences, at any given ground range, in terms of the *average* quality of horizontal-component data. It confirms the 90° optimum. Average horizontal quality is somewhat less affected than vertical quality by a given narrowness of convergence, but it enters twice into data quality. Comparing twice the variance in this table with the variance in the first of the above tables, it is seen to be equally important to avoid narrowing horizontal and vertical convergences in the region from 45° to 0°. Over convergences from 45° to 90°, the ~~vertical~~ horizontal goes from being equally important to being *twice* as important.

The case of the missile in the plane normal to a 2-station baseline at its midpoint (Figure 18) was picked to simplify the *math* - in the interest of physical understanding.

I think we can say our *net* station configuration is *as much* a chance proposition as our net number of stations. Of course, our results reflect our *average* station-configurations.

The above analytical approach has demonstrated ample *potential* for improving position-precision by improving the average goodness of intersection-angles, through increase in sample-size. (Toward a happy-medium convergence.) For the z coordinate, the first of the above tables indicates the direct effect of average convergence-angle on precision. For x and y, the last of the above tables indicates the direct effect of average convergence-angle on *their* average precision. (2 optical stations have 1 convergence, 3 stations 3, 4 stations 6, etc.).

The 'readability' mechanism speculated under SUMMARY FOR CINETHEODOLITES would differ from the others in this paper in being a *degradation* of gains that would otherwise be made. I am inclined, now, to explain our cine precision spectrum as due to readability *and/or* to increasingly *vertical* trajectories in going from left to right in the upper half of Figure 17.

So, $\bar{\bar{\sigma}}_C = f(N^m)$, where $m = f$ (; test configuration;

Let's call this mechanism improvement by overcoming nonoptimum *test* configuration.

3. Another possibility is that limitations of our least-squares *methods* are being averaged-out by (sheer) number of stations. I have attempted to state what might be called practical theorems about *optimum* least-squares (Ref. 8). These are given in Figure 20.

(1) Statistical optimum (minimizing variation among *individual* observations). WSMR's Davis methods for cines and for DOVAP (Ref. 9) are optimum by this criterion. Its Odle and Bodwell (Ref. 10) cine methods, used to some extent in the past, crippled their own ability to estimate statistically-true population-means and population-variances, by deterministically cutting their sample-sizes in half (least-squaring deviations of direction-lines). Until recently, WSMR's multistation-radar method deterministically cut its sample-sizes to one-third (by least-squaring deviations of 1-station solutions). Our new multistation-radar method is optimum by this criterion. Our Davis cine method is not statistically-optimum in estimation-of-quality, because it propagates an *average* angular-precision.

(2) General optimum (transforming residuals from station-variable to missile-variable *before* optimizing). Our former Odle and Bodwell cine methods and our recent radar method were optimum by this criterion, in the sense that they optimized *linear* deviations of 'observations'. This paper suggests that our Davis cine method is not optimum by this criterion, because it optimizes only similarity *at* the stations - not *overall* congruence; that is, it treats stations equally regardless of their *slant-ranges* from the missile and of the *convergences* of their lines-of-sight with those of the other stations. It optimizes angular quality of the *stations;* it does not optimize linear-position quality of the *missile.* (Some work has been done at WSMR toward a linear least-squares method for cinetheodolites.)

(3) Quality optimum (avoiding the probable loss of accuracy inherent in geometric-averaging of angular observations). Per right triangles: If 2 azimuth planes both miss a least-squares position solution (in general they will), their *intersection* will miss it farther than either plane (hypotenuse vs perpendiculars). Ditto for 2 elevation cones. If the azimuth plane and elevation cone of a station both miss a least-squares solution (in general they will), their intersection (the missile direction) will miss it farther than either surface. It follows that the linear *errors* of our former Odle and Bodwell, and recent radar, methods were probably larger than those of methods which least-square the original observations. So, those methods probably degraded physical accuracy. Our Davis cine and DOVAP and our new multistation-radar methods are optimum by this criterion.

(4) Summary. It *seems* clear that the criterion of a totally-optimum reduction can be met only by minimizing the sums-of-squares of the *linear* perpendiculars to: azimuth-planes, elevation-cones, range-spheres and loop-range-ellipsoids. That results by such optimum methods will be somewhat different, *more* precise, and probably more accurate than by our current methods.

Mr. Darold Comstock suggested accomplishing the proposed least-squaring of *linear* residuals by multiplying elevation residual by slant-range and azimuth residual by ground-range. This is radius times (angle in) radians. Algebraically it is a 'weighting'. Geometrically (and physically) it is a conversion. The subsequent algebra must be changed accordingly. (It is still necessary to propagate these linear standard deviations into those of the components of the least-squares point.)

It should be apparent that mechanism 3 interacts with mechanism 2 and with differences in slant range.

Let's *sample* the potential of linear least-squares for improving precision (and probably accuracy).

The configuration at the top of Figure 21 can show; a vertical projection of 2 stations at any point on their *respective* 'azimuth' circles about OM; or, a horizontal projection in any direction about point 0. (S′ can also 'flop' 180°). In the graphical *representation*, next: starting with the angular-LS solution, the 'direction-line' is allowed to swing until the *linear* residuals are equal. Total linear error, summation to point 'T', is also shown for each method. In doing this arithmetic, I actually used the slant range (or ground range) in the radius-times-radians *approximation* of arc for perpendicular to projection of El-cone or Az-plane.

Taking it slowly:

Angular least-squares yields the (arithmetic) mean of the angular residuals. (In this simple case, it makes the angular residuals equal.)

Linear least-squares yields the mean of the linear residuals. (In this simple case, it makes the linear residuals equal.) The angular subtends of the linear residuals are *not* equal, because their unequal scales of observation have been taken into account. By incurring a little bigger error with respect to S′, we minimize the *total* error. (Optimizing our *end-result* seems a *little* unnatural in our range environment.) The improvement is 38% in precision; 61% in variance. In a redundant case, use of degrees of freedom in calculating these would *increase* the differences. Finally, the *change* in the position component would generally lie between the minus 10.1 and plus 3.0 ft ballparks (of the changes in the linear precisions). On the *average*, I feel it *should* represent that much improvement in physical accuracy.

If one faces the linear errors behind our angular errors, to the point of a *linear* LS, the foregoing dependence of linear error on slant-range nullifies an assumption of 'rigorous' derivations of the least-squares principle - that the variances are not significantly different. Least-squares still yields the *minimum vector-resultant of the observed errors.*

It has been suggested that the *large* linear deviations should be weighted. Perhaps inversely as their slant-ranges! In *effect*, angular least-squares does that. This paper suggests we may be *deceiving* ourselves, in linear measurement, if we least-square angular residuals *in order to* perform our least-squares with 'nearly equal' variances. Slant-range is a physical variable - not a statistical weight. (The differences which it causes in linear error are not due to random sampling.) Our example indicated which procedure yields the smaller variance of our end-result. *Implicit inequality of variances* in our angular-LS apparently does more harm to our result than if this inequality were *minimized* by linear-LS.

Mr. Comstock's r σ_A is a very good approximation of the actual 'linear-equivalent' of an angular residual (r sin σ_A - that subtender of σ_A which is perpendicular to the El-cone or Az-plane).

In Figure 7, you saw our linear precision of a cine vertical component *better* than the linear precision of the horizontal components (FURTHER DATA, above). This does *not* hold true for the corresponding angular residuals (the angular precision of the vertical runs a bit worse). It appears this anomaly is all due to the interaction of our angular 'Davis' method with convergence, slant-range, and the azimuth-elevation system.

As you can see in Figure 22, the vertical subtender of a given angular error *increases* with elevation angle, for a given horizontal range. In the top *drawing* this actually overcompensates the effect of the difference in slant-range - which is the cause of the difference between angular and linear 'Davis' methods. This compensative case is the *whole* story for the *vertical* plane - that half the story for the horizontal plane wherein we subtend with the normal-to-the-line-between-the-stations. The bottom, *additive* case, is the other half of the story for the horizontal plane. What we can win in one direction we more than lose in the other. The net is an average *un*compensation of the equality of x and y. So, our Davis precisions of x and y do not approach the optimum of a linear least-squares.

This appears to explain the better precision of WSMR's vertical component. But, how much of the effect of number of stations can it account for? It turns out, the left end of the curve of σ_Z is slightly steeper relative to *its* N-to-the-first-power curve than is the left end of the curve of the composite σ. The right end is slightly shallower; but the curves are equally close to N-to-the-first-power, *overall.* So, the deficiency of our angular Davis does not *appear* to be a large part of the *particular* answer we set out to find in this paper (a little more on this under mechanism 5).

Let's *combine* the effects of slant-range and convergence to evaluate the *net* implications for linear-vs-angular LS and for z vs x, y.

Our first equation under mechanism 2 approximates *any* station's *separate* contribution to projected baseline-perpendicular measures-of-goodness of its measuring-convergence, at *any* point within-or-between the baseline-normal planes which pass through it and through any other station - regardless of the separate angles of their projected lines-of-sight with the baseline-normal through the projected object. Repeating this equation:

$$\sigma_\perp \approx \frac{r \sin \sigma_A}{\sin \theta/2}$$

For the top drawing of Figure 22, it turns out that the relationship of the normal subtenders of σ_A is (sliderule calculation):

$$\sigma_{\perp 60^\circ} \approx 1.002\sigma_{\perp 30^\circ}$$

(The relationship of the corresponding chords which reflect only slant-range is:

$$C_{60°} = 0.583C_{30°} \cdot)$$

So, in the 60° - 30° case, the *normal* compensation of slant-range by observing-angle is virtually perfect.

When the stations are not equidistant, the opposite diagonals of the convergence error-figure are no longer perpendicular. The third equation under mechanism 2 still approximates the *parallel subtenders* in the bottom drawing of Figure 22. It turns out that:

$$S_{\parallel 30°} \approx 2.945S_{\parallel 60°}$$

The chords are unchanged and their relationship may be written:

$$C_{30°} = 1.714C_{60°} \cdot)$$

Considered with the normal, this gives some feeling for the *net* uncompensation of slant-range by observing-angle in the horizontal plane.

For our LS example of Figure 21, it turns out that the normal subtenders of σ_A is:

$$\sigma_{\perp 60°} \approx 0.578\sigma_{\perp 15°}$$

(The relationship of the corresponding chords which reflect only slant-range, is:

$$C_{60°} = 0.299C_{15°} \cdot)$$

So, in the 60° - 15° case, the normal compensation of slant-range by observing-angle is quite inadequate. (In the 60° - 45° case, there is *over-compensation* - by 1.16.)

It appears that if we took slant-range into account in our *estimation* of position, we could produce more precise and accurate data (from the same records).

The above analytical approach has demonstrated *potential* for improving position-precision by overcoming the deficiency of an angular LS, through increase in sample-size. (Toward a happy-medium slant-range.)

So, $\bar{\sigma}_C = f(N^m)$, where m = f (; ; optimization criterion;

Let's call this mechanism improvement by overcoming nonoptimum choice of variable-to-be-optimized.

4. The decrease, through increase in sample-size, of the uncertainty of the directional aspect of position and position-quality. This is associated with the *increasing* probability (as we increase stations) that the 3-dimensional least-squares optimum will also be the least-squares optimum for *each* component.

In simpler language, we are talking about the probability (at each data-point) that the 3-way-average position will also be the average position in x, the average position in y, and the average position in z. A small sample that satisfies these *3* conditions seems about as likely as 3-cherries-in-a-row (on a slot machine).

An analytical approach to 3-dimensional vs 1-dimensional sampling is not within the scope of this paper. The following table gives some feeling for the extent to which these differ.

NUMBER OF SIGMAS	% LINEAR PROBABILITY	% SPHERICAL PROBABILITY[2)]
1	68.3	19.9
2	95.5	73.9
3	99.7	97.1

In connection with verifying that our matrix algebra yields the standard-error of a mean (ESTIMATION OF PRECISION, above), the standard-deviation and standard-error-of-mean of each set of linear-component equivalents of angular residuals will be calculated about its own mean (as well as about the least-squares mean). The differences in the *variances* of each component for the two means will sample the statistical-error in our assumption that the 3-dimensional least-squares optimum is also the least-squares optimum for each component. The component differences in the two *means* will sample the bias of measurement of each component relative to measurement of the 3-dimensional quantity (or vice versa).

The above discussion has indicated a *potential* for better optimizing the precision of *each* component, through increase in sample-size. (Toward a happy 3-dimensional medium.)

So, $\bar{\bar{\sigma}}_C = f(N^m)$, where m = f (; ; ; 3-dimensional sampling;

Let's call this one improvement by overcoming our inability to optimize *each* dimension.

5. WSMR's Final Data Reports show how the normal-distribution value of 1.65σ at 90% probability increases for individual standard-deviations at the *small* numbers of stations used for a position-estimate. As tempered by dividing out the corresponding smaller increase of the factor 1.00 at 68.3% probability, routinely introduced by Data Analysis Directorate ('t' statistic, ESTIMATION OF PRECISION, above). Figure 23 lists values of the 't' correction at 68.27% probability (Ref. 11).

This paper suggests that our $t_{.6827}$ correction to *individual* standard-deviations should also be divided out when the degrees-of-freedom are increased by averaging. That the average position-precision should

[2)] These values were obtained independently by Mr. Gideon Culpepper and by Dr. H. H. Germond.

then be multiplied by the value of $t_{.6827}$ for the degrees-of-freedom *in* the average (where this value of t is not negligibly close to unity).

We are concerned with the quality of our data in the form that it is reported to the user. Our basic unit of position-quality is the population-estimate of the quality of each component-value. This is properly normalized for sample-size and reported to the user in our Final Data Reports.

Our *test* (or segment) qualities are clearly the average quality of their individual component-values. If these *data-values* were independent, normally-distributed, and equivariant, the degrees-of-freedom in the average-quality of a round (or segment) would be: $2\bar{N}$-3 times the-number-of-component-qualities-averaged. Regardless of limitations of the validity of these statistical assumptions, this *apparent* degrees-of-freedom in the average is still our best-available estimate - incomparably better than using $2\bar{N}$-3 (the average degrees-of-freedom of an individual component-value). Averaging is a normalizing process.

Our various cumulative-average qualities, in which *rounds* are given equal weight, are practical approximations of the cumulative-average quality of their individual component-values. The best-available estimate of *their* degrees-of-freedom is $2\bar{N}$-3 times the-average-number-of-component-values-in-a-round times the-number-of-rounds-averaged.

Our (equivalent) linear-component deviations are *not* independent, because we least-square observational-deviations *simultaneously* in 3 dimensions. Component *variances* are further lacking in independence, because we propagate them from an *average* angular-variance. Now, *effective* degrees-of-freedom of these *less* than their 2N-3 means that our individual precision estimates are *too good.* Also, trajectory-measurements closely-spaced-in-time cannot reasonably be assumed independent. *Effective* sample-size of averages *less* than their number-of-component-values means that our estimated average-precisions are *too good.*

It is planned to follow-up Mr. Charles Bicking's suggestion to rms *random* samples of 25-or-less position-component sigmas from a given test; also, from a given series-of-tests - to get around lack of normality, independence, and homogeneity among our *component-values.* Then, to compare these and our regular-average sigmas with ASTM control-chart limits for their respective 'sample-sizes' - to sample the error in our assumption that our data-values have the above properties. *Suggestions: That* sample-size which places a regular-average sigma in the same proportional relationship to *its* control-chart limits as the corresponding random-sample sigma will be its effective sample-size. (If upper and lower limits yield different values, these may be averaged.) An appropriate set of sample-size conversion-factors for our average-sigmas can be generated in this manner. Use of *effective* sample-size will yield valid average-sigmas. (An existing average may be corrected by dividing by the square-root of its sample-size conversion-factor.) *One* minus the *ratio* of an average-sigma calculated from its apparent sample-size to its value calculated from its effective sample-size will be an *estimate* of the coefficient-of-linear-correlation of its component-values. (More simply obtained as one minus the square-root of its conversion-factor.)

Prof. William Kruskal's Comments on my paper show how negative[3)] correlations of observations increase the precision-response exponent.

The 't' correction normalizes the distributions of samples. In Figure 23, the larger values of t at low degrees-of-freedom reflect our inability to *know* the quality represented by one small sample. But, we know the round- and cumulative-*average* qualities of our low-degrees-of-freedom data quite well.

The foregoing implies that the WSMR average-precision-vs-number-of-stations data in Figures 2 and 4-16 should be divided by the values of t in Figure 23. That the *average*-precision of WSMR's 2-station (1 degree-of-freedom) optical position-data is substantially better than has been realized. 3-station (3 degrees-of-freedom), somewhat better. And, so on. (Similarly for DOVAP.)

On this basis, Figure 24 is the corrected version of Figure 9 (Redeye Contraves-precision vs number-of-stations and its exponential models). Comparing: Between 2 and 3 stations, our curve went from somewhere among its N^2 and $N^{3/2}$ models to somewhere among its $N^{3/2}$ and N models. Between 3 and 4 stations, our curve went from N^2 to $N^{3/2}$. Between 4 and 5 and 5 and 6 stations, there was little change. So, mechanism 5 accounted for $N^{3/4}$ between 2 and 3 stations and for $N^{1/2}$ between 3 and 4 stations. (For nothing between 5 and 6 and 6 and 7 stations.)

Figure 25 is the correspondingly corrected version of Figure 11 (Redeye target Contraves-precision vs number-of-stations and its models). Our 'fitted' curve went from being closest to $N^{3/2}$ to being closest to N.

It turns out that mechanism 5 changes the upper-half of our Lance DOVAP-precision curve (Figures 15 and 16) from proportionality to N to proportionality to $N^{1/2}$. The lower-half of the same curve remains at proportionality to N.

Mechanism 5 leads me to put more weight on the right-half of my comparisons of precision of z and of x, y, z (Figure 7) with their models. The right-half of the curve of σ_z is slightly shallower relative to its N-to-the-first power curve than is the right-half of the curve of the composite. This appears to show some effect of mechanism *3* on the power of the precision response (more room for overcoming the *horizontal* plane's net-uncompensation of differences in slant-range).

So, $\bar{\bar{\sigma}}_C = f(N^m)$, where m = f (; ; ; ; estimating degrees-of-freedom;

Let's call this one improvement by overcoming nonoptimum estimate of degrees-of-freedom.

6. The open-ended residue of other possible mechanisms of improvement of flight-measurement precision by increasing stations.

In this paper, we're talking about how fine a tolerance we can meet by increasing (and optimizing) *comparisons* in flight-measurement. Number of possible minimum *solutions* expresses the number of possible comparisons of observations on the common basis of our end-result. For optics:

3) When one increases the other decreases. This is commonly the case in 3-dimensional LS solutions. These are usually influenced by measurement in one dimension only at the expense of measurement in the others.

NUMBER OF STATIONS	NUMBER OF OBSERVATIONS	D. F. OF OBSERVATIONS	NUMBER OF SOLUTIONS	D. F. OF SOLUTIONS
½	1	− 2	1/3	−2/3
1	2	− 1	2/3	−1/3
1½	3	0	1	0
2	4	1	4	3
2½	5	2	9 or 10	8 or 9
3	6	3	19	18
etc.				

It turned out that replacing number-of-observations by degrees-of-freedom (degrees-of-redundancy) of observations in the models of Figure 3 gave curves whose shapes were *less* like our operational curves. Curves based on degrees-of-freedom (degrees-of-redundancy) of solutions, and on number-of-solutions, would be still less so.

However, number of possible comparisons and number of independent[4] comparisons (D.F. of solutions) may be relevant to learning the *degree* to which stations are capable of mutual calibration (to improve system-precision) and of calibrating one another (EFFECT ON PHYSICAL ACCURACY, below).

Figure 26 shows the frequency-distributions of the *squares* of the precisions which were sorted by number-of-stations in Figure 2. It appears that correlation of other determinants of position-quality with number-of-observations clearly *corresponds* to contraction of extended frequency-distributions of variance through increasing[5] degrees-of-freedom chi-squared (or less-skewed normal) distributions, as number-of-stations increases. This change of shape is an *effect*, not a cause.

Figure 27 summarizes the five clear-cut mechanisms of this paper, plus its open-ended 'catchall'. My present *guess* is that the order of descending (but real) magnitude of the effects of the first five mechanisms is: 2; 1; 5; tie between 3 and 4.

RATING CONFIGURATIONS OR STATIONS. Suggested procedure:

1. Generate a more-complete version of the 1st and 3rd columns of the *first* table under mechanism 2, above. Also, a more complete version of the 1st column and of twice the 3rd column of the *third* table under mechanism 2. This is properly done with a trig. table and a slide-rule.

2. On a reproduction of a map or scale-drawing of a given (or proposed) configuration, draw *all possible* sides and diagonals. (A diagonal *can* be external.) *Estimate*, and 'x', the configuration's center-of-gravity on the basis of visual judgment - aided somewhat by little-circles around all intersections of its diagonals. The configuration should be chosen so the nominal trajectory will pass near its center-of-gravity.

[4] In the sense of affording a choice.

[5] But effectively *very* low.

3. For some given (or proposed) altitude above this C.G. - *or* above the nearest point to it on the nominal trajectory - approximate the projected vertical convergence for *each possible* (2-station) baseline as:

$$\theta = 2 \text{ arc tan } \frac{\text{baseline}}{2\,(\text{altitude})}$$

where each baseline is measured graphically. The altitude will generally be that of the nominal trajectory at the midpoint of this planned segment.

4. Look up in the first table, list, and add the variances of these vertical convergences.

5. Approximate the projected horizontal convergence for *each possible* baseline as:

$$\theta = 2 \text{ arc tan } \frac{\text{baseline}}{2\,(\text{perpendicular distance})}$$

where its ⊥ distance is measured graphically *from* the C.G. - or, *better,* from the nearest-point-to-it under the nominal trajectory. Or, *preferably,* read these horizontal-convergences with a protractor. (Using the 'nearest-point' includes the *nearness* of the configuration's C.G. to the trajectory in the rating.)

6. Look up in the second table, list, and add the double-variances of these horizontal-convergences.

7. Combine totals from steps 4 and 6.

8. For a *given number* of stations, pick the configuration with the smallest total variance.

9. In adding a station to a given configuration, pick the station for which the sum of *its* convergence-variances (with all the others) is the smallest.

10. In deleting a station, drop the one/sum of whose convergence-variances is the biggest.

The above proximate method should work fairly well, because of the big variances of bad intersections. It is valid for optics and radar. Possible refinements include:

(1) Finding the true C.G. by graphical or analytical methods.

(2) Multiplying each vertical-convergence variance by the square of its projected *slant-range:*

$$r^2 = \left(\frac{\text{baseline}}{2}\right)^2 + (\text{altitude})^2$$

And, multiplying each horizontal-convergence variance by the square of its projected ground-range:

$$r^2 = \left(\frac{\text{baseline}}{2}\right)^2 + (\perp \text{distance})^2$$

(The latter r may also be obtained by averaging the graphical distances from the ends of the baseline to the C.G. - or to the nearest-point.)

The two tables described under step 1 are optimum *figures-of-merit* of the various convergence-angles. As such they are suitable for general use. The overall measuring-effectiveness of most actual convergences will fall short of their projected-equal-baseline optima. But, losses in going to unequal slant-ranges and to convergences-external-to-the baseline should be reflected well enough for ordinary purposes through the badness of such convergences.

COMPONENTS OF COVARIANCE (GDOP MADE EASY). 'Geometric dilution of precision' most usefully refers to the *geometric components* of position-measurement variance. Somewhat less definitively: '. . . the magnitude of the position errors caused by random measurement errors . . . depends on the particular parameters measured, the measurement system, location of the measuring equipment and the location of the missile with respect to the equipment. Variation of the effect of random errors is measured by a quantity defined as the Geometrical Dilution of Precision (GDOP).' (Ref. 12).

From MECHANISM 2, above, our equations for the vertical and horizontal projections of linear-measurement variance are:

$$\sigma_V^2 \approx \frac{r^2 \sin^2 \sigma_A}{\sin^2 \theta/2}$$

$$2\sigma_H^2 \approx r^2 \sin^2 \sigma_A \left(\frac{1}{\sin^2 \theta/2} + \frac{1}{\cos^2 \theta/2}\right)$$

These apply when the missile is in the plane normal to a 2-optical-station baseline at its midpoint. The first equation also applies to any optical station's *separate* contribution at any point within-or-between the baseline-normal planes which pass through it and through any other station - regardless of the separate angles of their projected lines-of-sight with the baseline-normal through the projected object. These equations also serve fairly well for radar under these conditions.

σ_V^2 and $2\sigma_H^2$ are the vertical and horizontal components of position-measuring variance under the conditions described.

$\sin^2 \sigma_A$ is the instrument-component of both vertical and horizontal position-measuring covariance, under these conditions. The remainders of the right-hand sides of the above equations are the geometric-components of vertical and horizontal position-measuring covariance (under the conditions described). r^2 is the 'scale' subcomponent of the geometric-components of both vertical and horizontal

position-measuring covariance. $1/\sin^2 \theta/2$ and the parenthetic trig. expression are the 'configuration' subcomponents of the geometric-measuring covariance. The latter are, then, simple trig.-functions of convergence-angle. (The square roots of all the above are the 'components' of position-measuring precision or *coprecision.)* Our Davis cine method does not optimize the *geometric* components of position-measurement covariance.

For a *given* instrument-precision, the above 2 equations are the vertical and horizontal 'GDOP' equations - for these conditions. They are a bridge between partial-differential equations, their matrix representation, and nonspecialists - for the 2-station optimum (baseline-midpoint) set.

REDUNDANT PRECISION OF NONREDUNDANT SOLUTIONS. It is operationally necessary to compare the quality of nonredundant solutions for missile-position with the quality of redundant solutions. (Examples of nonredundant solutions are: 1-station radar, 1½-station optics, and 3-station DOVAP.)

The uncertainty of a nonredundant (zero-degrees-of-freedom) solution is not a problem if one has sufficient info on the statistical population of which the solution is a sample (of size one) - or on *comparable* populations.

The commonest example of a nonredundant solution is a single observation of a one-dimensional quantity. The concept of the standard deviation of (single) observations is the best-known of all precision concepts. It is, of course, the characteristic statistical uncertainty of a single observation.

Usually, the parameters of the population-of-observations are estimated from a sample of (several) observations. The precision of the sample-mean is then given by the relationship:

$$\hat{\sigma}_{\bar{x}_n} = \frac{\hat{\sigma}_x}{n^{1/2}}$$

where σ_x is the estimate of the standard-deviation-of-observations, and n is the sample size (Ref. 4).

If only the precision of the mean of a given number of someones observations (of a one-dimensional quantity) *is* available, it is routine to use this equation to calculate the uncertainty of a single (nonredundant) observation.

In flight-measurement terms, the above equation takes the form of equation (5) of our Figure 3. Or its working form - equation (1) of the same figure. Empirical generalization of the latter may be written:

$$\hat{\sigma}_{C_2} = \left(\frac{N_1}{N_2}\right)^m \cdot \hat{\sigma}_{C_1}$$

And, we have found that we can determine, by trial-and-error, a power of the ratio of numbers of stations which makes this equation closely fit a segment of a given plot of average-precision vs number-of-stations.

From the foregoing, this relationship for 1 and 2 degrees-of-freedom position-data can clearly be used to *calculate* the *average* component-precision of the *corresponding* zero-degrees-of-freedom (nonredundant) solution. As our reductions are *presently* carried out, this means using 2- and 3-station precisions of radar or optics or 4- and 5-station precisions of DOVAP. (3-station optics is actually 3 degrees-of-freedom and 3-station radar is 6.) The result will be as representative as the 1 and 2 (or 3) degrees-of-freedom data. The solution of the above equation for the exponent reduces to:

$$m = \log \frac{\bar{\sigma}_{c_2}}{\bar{\sigma}_{c_1}} \div \log \frac{N_1}{N_2}$$

which can be carried out to the degree-of-fineness desired (three digits are ample). The value of m for 1 thru 2 degrees-of-freedom is, of course, used in the previous equation with the 1 degree-of-freedom data to calculate the zero-degrees-of-freedom precision. If the 1 degree-of-freedom data are limited to a particular combination of stations, the result will be average for *those* stations - slightly influenced by the makeup of the second (or third) degree-of-freedom.

If one desires to calculate average x, y, and z precisions of nonredundant solutions, the above may be carried out separately for average x, average y, and average z precisions.

If one desires to calculate average precisions of the observations of azimuth, etc. in a nonredundant solution, the above may be carried out separately for average azimuth, etc. residuals from 1 and 2 degrees-of-freedom solutions. (Preliminary indications are that, in our process, time-rmsd deviations of an observational parameter are *not* independent of number-of-stations.)

If one desires to calculate average component-precisions, or x, etc. precisions, or azimuth, etc. precisions of a *particular* nonredundant solution, the above may be carried out separately for average component residuals, average x, etc. residuals, or average azimuth, etc. residuals of *that particular* nonredundant solution, from 1 and 2 degrees-of-freedom solutions. These precisions will of course be somewhat influenced by the makeup of the first (and even second) added degrees-of-freedom.

Combining x, y, and z precisions of nonredundant solutions calculated by any of the above, as an rms, then comparing with the similarly-calculated corresponding average *component* precision can serve to check the coherent *execution* of the methods. (When desired coherence is not attained, the *best* result is the rms of the two.)

The foregoing applies equally well to nonredundant solutions for attitude - or for any other measured missile-flight variable.

The above method is empirical only to the *degree* that improvement of the particular precision by increasing stations exceeds the purely repetitive (statistical) amount.

So, there are no conceptual difficulties - physical or statistical - in calculating the average-precision of nonredundant solutions on the same (WSMR standard) basis as the average-precision of redundant

solutions. The former may be calculated from the latter, for the same or comparable measuring situations. Where there is overlap, the result may be used in quality-control of the non-redundant solutions. Where redundant precisions from comparable current measuring situations must be used, the result is still a basis for stating nonredundant capability.

GEOMETRICAL STATISTICS. Geometry is used here in its plane, solid, and mensuration senses.

I am inclined to consider geometrical statistics a boundary-discipline of geometry and numerical statistics. Our concern here is with the statistics of geometry (not with the geometry of statistics). Regardless of terminology, our subject *includes* the following elements:

1. Individual measurements of object-position define geometric surfaces in space. (Or geometric curves in the plane of object and observer.) The particulars of these for a station are its own coordinate system. The particulars of these for a system are its reduction geometry. In general, 3 individual measurements of magnitude or direction determine a space-position[6]. The σ_A of our Davis cine method is a parameter of an error-ellipsoid which is definable only in terms of the particular measurement-configuration. This paper suggests our use of an average angular precision evades the issue of how the 3 degrees of freedom used *by* the LS determination of a *particular* 3-dimensional-Cartesian position *are* effectively distributed between azimuth and elevation; and, hence, how they should be distributed for estimating the separate *effective* precisions of azimuth and elevation (which would be more useful). This is a question of measuring the *relative* degree to which x is determined from azimuth and from elevation; ditto for y. (z competes with x and y for elevation. We can consider it subtracts 1 degree-of-freedom from what they leave.) Because of the smaller *separate* sample-sizes, and because of the physical imbalance of 3-dimensional degrees-of-freedom between azimuth and elevation[6], it may be concluded our average-angular precision makes our separate angular measurements look a little better and a lot more alike than they are.

It is suggested that, in *evaluating* effective azimuth and elevation precisions, one would apply their separate 't' corrections. And, that one would bypass or divide out said corrections for quality-control averages of these separate precisions. Perhaps the method of generating 't' tables will work for a fraction of one degree-of-freedom! (And for fractional interpolation.)

The above suggestion differs from present practice in 4 ways: The problem of determining the separate effective degrees-of-freedom of azimuth and elevation; QC-averaging of estimates of station-quality which *would* be based on degrees-of-freedom; estimating separate effective-precisions of azimuth and elevation for *each* data-point; the problems of propagating these precisions of azimuth and elevation into effective-precision of the x-component and effective-precision of the y-component, and of this elevation precision into effective-precision of the z-component. It may also be concluded our average-angular precision - as far as it goes - makes our 3 precisions of linear-measurement look a little better and a lot more alike than they are.

---

[6] Azimuths alone cannot determine the vertical coordinate. Although it is not *widely* realized, elevation residuals have x and y components as well as z. (At high elevation-angles, they are *mostly* x and/or y!) Of course (in the physical sense that is relevant here), azimuth residuals have only x and y components.

This paper recommends the above suggestion be carried out for one high-altitude round, reduced as 2-, 3-, 4-, etc-sta. solutions - to see *where* we stand and *whether* a change-of-procedure is needed.

Suggested *approach* to relative degree x depends on azimuth and on elevation; ditto for y. (And to how they divide up azimuth.): Calculate separate precisions of azimuth and elevation based on N - 3/2 degrees-of-freedom. Propagate *each* of these sigmas into the x-component of the LS-solution by multiplying it by the rms of the partial-derivatives of the x-component with respect to *its* set-of-angular-measurements. The ratio of the variance of the x-component propagated from azimuth to the sum of the variances of the x-component from azimuth and from elevation is the fractional dependence of x on azimuth. One minus that is its fractional dependence on elevation. Ditto for y. (The ratio of the variance of the x-component from azimuth to the sum of the variances of the x- and y-components from azimuth is the x-fraction of azimuth. One minus that is its y-fraction.) N minus the fractional dependences of x and y on azimuth is the degrees-of-freedom to use in recalculating the effective-precision of azimuth. (N - 1) minus the fractional dependences of x and y on elevation is the degrees-of-freedom to use in recalculating the precision of elevation. Iterate propagation of these revised sigmas into x- and y- components - as many times as necessary. (One iteration may suffice.) Effective-sigma of the x-component is the rms of its final effective-sigmas from azimuth and from elevation. Ditto for effective-sigma of the y-component. Propagate final effective-sigma of elevation into the z-component, to get the latter's effective-sigma.

One of the operations of geometrical statistics, then, is transformation of precision indices from instrument-coordinates to data-coordinates. Our Davis cine method accomplishes this, in its way, simultaneously with (multivariate) statistical transformation of the precision indices from observations to data. This paper suggests that *its* procedure, above, would do the same thing *much* more validly.

The foregoing suggestion would also apply to the totally-optimum linear least-squares proposed under MECHANISM 3, above. The computing would simplify, since propagations would be only (multivariate) statistical transformations of precision indices from observations to data.

A second element:

2. Geometric convergence. This element was treated under: MECHANISMS 2, 3; RATING CONFIGURATIONS OR STATIONS; and COMPONENTS OF COVARIANCE (GDOP MADE EASY). These effectively *disimbed* GDOP from its matrices, in forms that are more easily understood and used. This was done by graphical representation and elementary trigonometry. Jeffreys (Ref. 13) has shown that (the) trig. functions are *laws* of (physical) mensuration, not merely mathematical definitions.

It was feasible, in this paper, to deal with the projected *planar* convergence of the linear *intersections* of the two angular observations of *each* of two optical (or radar) stations. This *approach* is certainly *one* of the practical optimums. From MECHANISM 3, dealing simultaneously with the multiple *spatial* convergence of the *individual* observational-surfaces of the *total* number of simultaneously-tracking stations would be the statistical and quality optimum (criteria 1 and 3 of Figure 20). In Figure 18, σ_A can be σ_{Az} *or* σ_{El}. The values of these may differ, but the value of each must be the same for both stations.

A third element:

3. (Theorem) Totally-optimum *measurement* analysis optimizes the *geometric* components of position-measurement covariance, as well as the instrument components. Taking slant-range into account - transforming residuals from station-variable to missile-variable - before least-squaring was treated under MECHANISM 3. Taking convergence-angle into account before least-squaring is treated in this paper only to the point of station-selection (RATING CONFIGURATIONS OR STATIONS).

4. (Theorem) The multidimensional-optimum position is generally not optimum for any single dimension. This element was treated under MECHANISM 4.

As long as we deal with vectors in terms of their *components*, it is valid to characterize precision by three 1-dimensional frequency distributions. But, if one deals with a 3-dimensional vector as an entity (e.g., in vector analysis), the standard deviation of *its* value (quality of the vector) is:[7)]

$$\sigma_V = \sqrt{\sigma_x^2 \lambda^2 + \sigma_y^2 \mu^2 + \sigma_z^2 \nu^2}$$

where λ, μ, and ν are the vector's respective direction cosines. Our $t_{.6827}$ corrections to individual component-sigmas should be divided out beforehand.[8)] This formula applies to any performance variable. For the case where $\sigma_x^2 = \sigma_y^2 = \sigma_z^2$, the direction cosines drop out. For the position vector, σ_V is how well we know the distance of the missile from its launcher reference. It's the precision of the radius vector (r) of spherical coordinates. (So, the above is one of the 3 geometrical-statistics formulas required to transform precision indices from Cartesian coordinates to *spherical*.) For the velocity or acceleration vector, σ_V is (correspondingly) the precision of the missile's radial-velocity or radial-acceleration. The formula without the direction-cosines is also the vector representation of 3-dimensional precision (vector of the quality).

Though not commonly considered a geometric dimension, overcoming effective *time-measurement* differences among station-observations - toward a happy-medium difference - is part of our open-ended MECHANISM 6. This is not limited to differences in the time-signal, but includes *all* differences in synchronization throughout the data-process (Ref. 6). It shows up as an *unbalanced* contribution to position-measurement variance (maximum in the direction of the velocity-vector - zero, normal to that).

5. (Finding) In flight-measurement, curves of average precision vs number-of-observing-stations may be fitted by *generalizing*, to a variable degree, the exponential dependence of the standard-error-of-the-mean-of-a-normal-distribution on sample-size. The *exponential* improvement of position-precision by increasing stations can be as much as 4 times that of the classical relationship of numerical statistics. (2 projects in which the improvement was 5 times are disregarded, because data were somewhat limited.)

In this sense, at least, geometrical statistics is a generalization of numerical statistics. In addition to numerical convergence, the above exponential dependence of precision-response is clearly a function of

---

[7)] This equation was derived for the writer by Mr. W. E. Mimmack. It checks a classical source (Ref. 15).

[8)] To normalize, the 3-dimensional sigma should be multiplied by a 3-dimensional 't'. These *can* be deduced, for either the general or equivariant case, from Reference 16.

geometric convergence (the configuration subcomponent of the geometric component of position-measuring covariance); is apparently a function of slant-range (the scale subcomponent of the geometric component); and is probably a function of multidimensional *sampling.* It *may* also depend to some extent on precision-indices being variant under coordinate-transformation - and on time-measurement.

A negative view might be that *classical* numerical statistics is not applicable to flight-measurement, because observations and component-values of flight-data lack normality, independence, and homogeneity.

This paper suggests that the 'random-error' concept of numerical statistics is not adequate *by itself* for flight-measurement. That, besides more concern with the physical and geometrical meanings of our math, more formal development of the boundary-discipline of geometrical statistics might be helpful.

The findings of this paper *logically* lead to the following statistical 'heresies':

(1) We should average our precisions in whatever exponential form is proportional to sample-size for that vehicle. This calls for an averaging-spectrum ranging from averaging variances to averaging square-roots of standard-deviations.

(2) Our results call for an *optimizing* spectrum ranging from least-squares to least-square-roots.

We aren't doing these heretical things because: it's inconvenient; we want to maintain our bridge to standard methodology; it's desirable for a data-quality index to be sensitive to bad data (as variance is).

Each of the foregoing elements of geometrical statistics is also *relevant* to station and system calibration and residual bias, and *their* statistical uncertainties. (See EFFECT ON PHYSICAL ACCURACY, below.)

EFFECT ON PRECISIONS OF VELOCITY AND ACCELERATION. Figure 28 shows average precision of velocity of Chaparral missile (Feb - Nov '67) vs average number of Contraves stations computed. Its hyperbolic models show it closest to improvement of precision in direct proportion of number-of-stations.

Over 6 sets of Contraves and one set of DOVAP data, there was a *slight* tendency for velocity precision to depart from proportionality to N in the direction of $N^{3/2}$. Also, a *slight* tendency for the curves to be straighter than the hyperbolic model.

Figure 29 shows average precision of *acceleration* of Redeye target (Mar - May '67) vs average number of Contraves stations computed. The upper half of the curve was closest to improvement in proportion to N; the lower half closest to $N^{3/2}$. Overall, it was a tossup between those two.

Over 6 sets of Contraves and one set of DOVAP data, *acceleration* precision was evenly divided between proportionality to N and proportionality to $N^{3/2}$. More than half the curves tended to be straighter than their closest hyperbolic model. The tendency for average quality to be influenced by a bad round increased in going from position to velocity to acceleration.

It is planned to look at the above relationship for *smoothed* position, which occurs in our process between measured-position and velocity.

Propagating the effect of number-of-stations on precision-of-measured-position through: time series, polynomial fit, lack-of-fit, 1st derivative, and 2nd derivative is not attempted in this paper. It is reasonable that the precision of a position time-series reflects to a considerable degree the precision of its values. (And, so on.)

Some of the variability of our observations is 'short-period', so their numerical convergence (MECHANISM 1, above) should be reflected to some extent in time-varying precisions. The difference between 3-dimensional and 1-dimensional sampling (MECHANISM 4) should also have a 'short-period' component. Mechanisms 2 and 3 are time-varying, but not clearly short-period.

The 't' correction to position-precision (MECHANISM 5) does not enter time-varying precisions. These 'start over' with the time-series 'observations' of position. (t corrections have not been applied to our time-varying precisions.) Calculating effective position-component precisions from effective-precisions of azimuth and elevation (under element 1 of GEOMETRICAL STATISTICS) would not enter time-varying precisions.

Effective time-and-sync. differences among station-observations (under element 4 of GEOMETRICAL STATISTICS) should have a short-period component. Overcoming this through increase in sample-size - toward a happy-medium difference - would be doubly reflected in actual precisions of velocity and acceleration. (Indirectly through space-measurement and directly through time-measurement.) We are not yet propagating time-precision through the derivatives.

Our smoothing interval is usually constant for a project. Number-of-stations should influence the *relationships* between smoothing-interval and time-varying precisions.

EFFECT ON PHYSICAL ACCURACY. The effect of number of observing stations on flight measurement accuracy (bias) is a subject for further investigation.

I am willing to postulate a probabilistic improvement of physical accuracy by increasing stations in flight-measurement - because we thereby increase the probability of mutual compensation of station biases in both magnitude and direction.

Churchman (Ref. 17) notes that 'the true value is not a random variate, that it is a unique element among the real numbers, and that the probability of its lying in *any* interval is therefore either exactly one or exactly zero.' However, such absolute knowledge is not granted us. Our *estimates* of physically-true values - or of bias therefrom - are random variates. (In the sense that physical *info* is unavoidably probabilistic.)

System precision (data precision) is a collective measure of the *mutual* calibration of its stations in space and time. System *bias* (data accuracy) is the net (uncompensated) sum of its station biases (in space

and time). On the average, improving *both* individual and mutual station calibrations (station accuracy and system *precision*) should improve the *net* calibration (system accuracy) *more* than would individual calibrations alone.

MECHANISM 2 treated the *error* in Figure 18 as a dispersion (or precision) index. Let's consider it a discrete bias. Then, in Figure 30b, if the discrete angular errors happen to have *opposite* signs, only the baseline-normal diagonal of the smaller almost-diamond in Fig. 18 exists. From MECHANISM 2:

$$\Delta_{\perp} \approx \frac{r \sin \Delta_A}{\sin \theta/2}$$

And, the first two columns of the first table under MECHANISM 2 *also* compare the *accuracy* of perpendicular convergences for either horizontal or vertical planes. In Fig. 30a, if the discrete angular errors happen to have the *same* sign, only the baseline-parallel diagonal of the smaller almost-diamond in Fig. 18 exists. From MECHANISM 2:

$$\Delta_{\parallel} \approx \frac{r \sin \Delta_A}{\cos \theta/2}$$

This equation is meaningful *only* for the horizontal plane.

The following table averages the accuracies of the vertical convergences for the (even) chances that station biases will have the same or opposite signs[9]:

Vertical Average *Accuracy* Factors

VERTICAL CONVERGENCE (θ)	$\frac{1}{2}\left(\frac{1}{\sin \theta/2}\right)$
180°	0.50
135°	0.54
90°	0.71
45°	1.31
15°	3.84
5°	11.5
1°	57.2

The above values are half those of the first table under MECHANISM 2. (When signs are opposite, there is *no* net bias.) So, vertical convergences *rank* the same for average-accuracy as for average-precision, but the *effect* of a given departure from the 180° optimum is only half as great.

[9] Even chance, because - to the extent bias of a given-type instrument *consistently* has the *same* sign it is more likely to be adjusted or corrected for.

The following table averages the accuracies of the horizontal convergences for the (even) chances that station biases will have the same or opposite signs. (In each case, a zero enters for the *other* diagonal.)

Horizontal Average Accuracy Factors

PERPENDICULAR CONVERGENCE (θ)	$\frac{1}{4}\left(\frac{1}{\sin \theta/2} + \frac{1}{\cos \theta/2}\right)$
179°	28.9
175°	6.0
165°	2.17
135°	0.92
90°	0.71
45°	0.92
15°	2.17
5°	6.0
1°	28.9

Horizontal convergences rank the same for average-accuracy as for average-precision (third table of MECHANISM 2), but the effect of a given departure from the 90° optimum is less than half as great.

Still, there is plenty of potential for improving position *accuracy* by improving the average goodness of intersection-angles through increase in sample-size. Mechanisms 3 and 4 (Fig. 27), and our time-measurement possibility, also apply to accuracy as well as precision.

SUMMARY. Supplementing the ABSTRACT: WSMR-standard precision of a measured component-value reflects the agreement among its station-observations at that point in time. We use rms-average precisions for operating and management control. It was found curves of average precision vs number-of-observing-stations may be fitted by *generalizing* the exponential dependence of the standard-error-of-the-mean of a normal distribution on sample-size. The precision-response of cinetheodolite position-data ranged from proportionality to $N^{1/2}$ to proportionality to N^2. Five mechanisms seemingly involved in this profound effect, plus an open-ended catchall, are summarized in Fig. 27. (My present guess as to their descending magnitude: 2; 1; 5; tie between 3 and 4.) This investigation has led to methods-improvement suggestions for collection, reduction, and quality-estimation. A marriage of geometry and statistics has been partly consummated, on simple terms. Previous work on optimum convergence (Ref. 6) was extended to *quantitative* evaluation of the precision and accuracy of *all* linear convergence-angles - for measuring the vertical and horizontal components of space-position. It appears that if we incorporated slant-range ahead of our least-squares estimate of position, we would produce more precise and accurate data. *(Implicit* inequality of variances in angular least-squares apparently does more harm than if this inequality were minimized by linear least-squares.) A sufficient reason for using least-squares: Even when *all* the rigorous assumptions of the Least-Squares Principle are violated, least-squares still yields the minimum vector-resultant of the observed errors. A method was given for evaluating our assumptions that propagating variance into a least-squares position-component yields the standard-error of a mean, and that a 3-dimensional optimum is optimum for each component. It was

suggested the 't' correction to *individual* position-precisions should be removed before averaging these. An approach to evaluating the effective sample-size of our average-precisions has been described. The procedure suggested for rating measuring-ability of test-configurations can be set up with trig.-table and slide-rule, and operated with an adding-machine. Components-of-position-measuring-covariance and GDOP (geometrical-dilution-of-precision) were presented in forms easily understood and used. Some elements of the boundary-discipline of geometrical statistics have been discussed. A way was suggested of taking the physical imbalance of degrees-of-freedom between azimuth and elevation into account - to calculate more valid angular and linear-component precisions. Relationships between component-quality and vector-quality were touched on. The entire paper is *relevant* to geodesy as well as flight-measurement.

REFERENCES

1. Hanson, F. S. 'Control of Data-Support Quality', Proceedings of the Eleventh Conference on the Design of Experiments in Army Research Development and Testing, 1965, pp. 313-336.

2. 'An Evaluation of the Quality of Flight Test Data for August 1968', Analysis and Computation Directorate, National Range Operations, White Sands Missile Range, New Mexico.

3. Davis, R. C. 'Techniques for the Statistical Analysis of Cinetheodolite Data', NAVORD Report 1299 (NOTS 369), China Lake, California, 1951.

4. Hoel, *Introduction to Mathematical Statistics,* Wiley, 1954, p.103.

5. Letter from H. H. Germond, Albuquerque, New Mexico, January 1969.

6. Hanson, F. S. 'System Configuration Problems and Error Separation Problems', Proceedings of the Tenth Conference on the Design of Experiments in Army Research, Development, and Testing, 1964, pp. 119-143.

7. Hereford, W. V. 'Discussion on Improvement in Position Data Accuracy as a Function of the Number of Tracking Contraves Stations', Sandia Corp., Albuquerque, New Mexico, October 1968.

8. Hanson, F. S. Minutes of 106th Meeting of NRO Quality Assurance Committee, 10 May '68 (published 28 August '68), pp. 4-5.

9. Davis, R. C. 'Techniques for the Statistical Analysis of Continuous-Wave Doppler Data', NAVORD Report 1312 (NOTS 383), China Lake, California, 1951.

10. Bodwell, C. A. 'A Least-Squares Solution of the Cinetheodolite Problem', FDL-WSPG Report No. MTHT-138, Holloman Air Force Base, New Mexico, 1951.

11. Pearson and Hartley, *Biometrika Tables for Statisticians,* Vol. 1, Cambridge University, 1954.

12. Pyne, R. J. 'Geometrical Dilution of Precision - Spherical-Polar System (SCR-584 Radar)', Air Force Technical Report No. 26, AFMTC, Patrick Air Force Base, Cocoa, Florida, 2 November 1951, p. *v.*

13. Jeffreys, *Scientific Inference,* Cambridge University, 1957, p. 111.

14. Germond, H. H. 'Accuracies Obtainable from Redundant Data When Position and Orientation of Radars is Subject to Error', PAA Technical Staff Technical Memo No. 14, AMR, Patrick Air Force Base, Florida, 3 April 1964, Appendix A.

15. Schols, C. M. 'Theory of Errors in a Plane and in Space', Annals of the Delft Polytechnical School, *3,* 1886, pp. 123-179. (Translated from a French version by Schmidt, Wozencraft, and Hayes, WSPG, 1958.)

16. Owen, *Handbook of Statistical Tables,* Addison-Wesley, 1962, pp. 202-224.

17. Churchman, *Theory of Experimental Inference,* MacMillan, 1954, p. 36.

$$\sigma_A = \sqrt{\frac{\Sigma[(\delta\alpha_i)\cos\epsilon_i]^2 + \Sigma(\delta\epsilon_i)^2}{2N-3}}$$

$$\sigma_x = \sqrt{\frac{\Delta_{11}\,\sigma_A^2}{\Delta}}$$

$$\sigma_y = \sqrt{\frac{\Delta_{22}\,\sigma_A^2}{\Delta}}$$

$$\sigma_z = \sqrt{\frac{\Delta_{33}\,\sigma_A^2}{\Delta}}$$

FIGURE 1

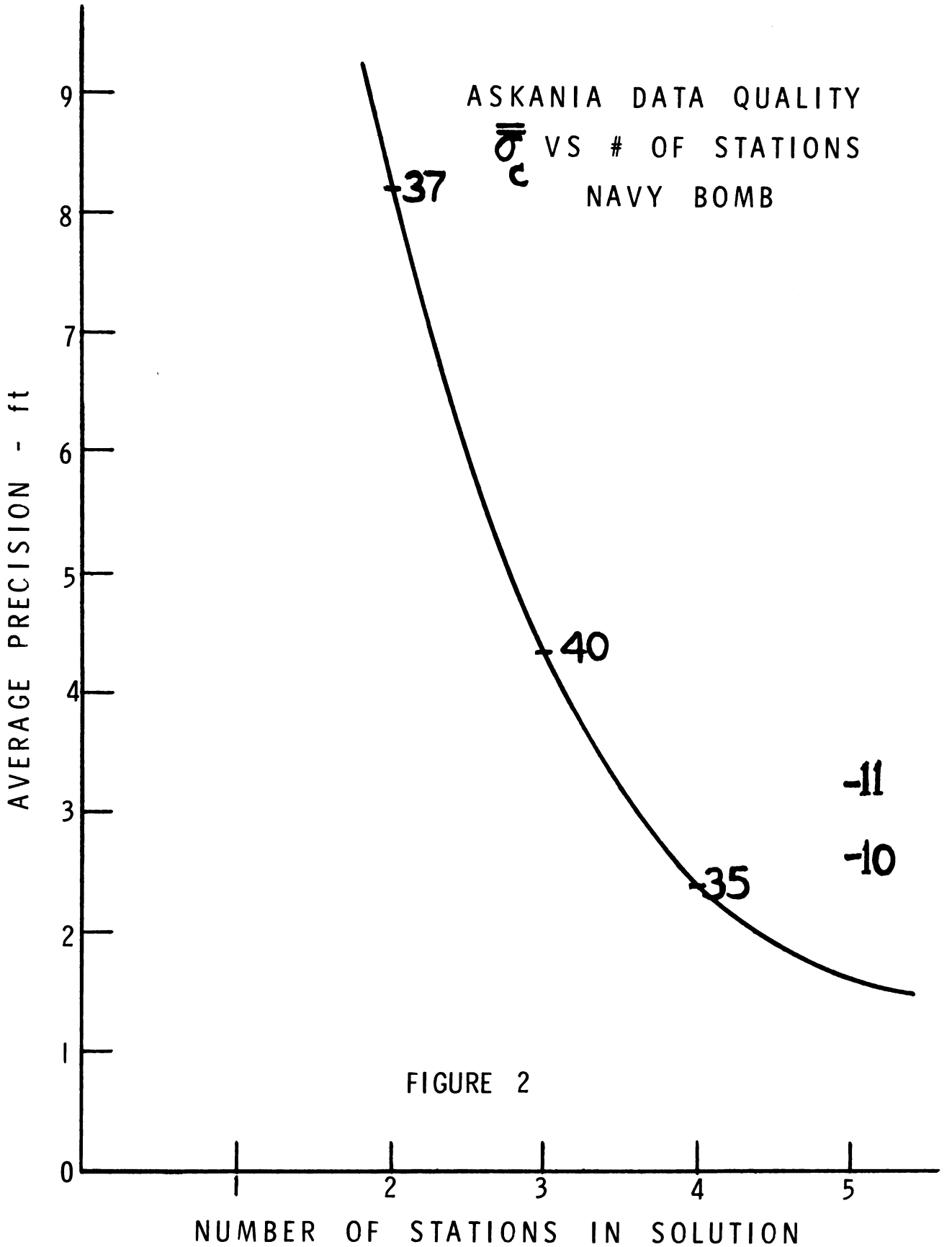


FIGURE 2

$$\bar{\bar{\sigma}}_{c_2} = \left(\frac{N_1}{N_2}\right)^{\frac{1}{2}} \cdot \bar{\bar{\sigma}}_{c_1} \qquad (1)$$

$$\bar{\bar{\sigma}}_{c_2} = \left(\frac{N_1}{N_2}\right) \cdot \bar{\bar{\sigma}}_{c_1} \qquad (2)$$

$$\bar{\bar{\sigma}}_{c_2} = \left(\frac{N_1}{N_2}\right)^{\frac{3}{2}} \cdot \bar{\bar{\sigma}}_{c_1} \qquad (3)$$

$$\bar{\bar{\sigma}}_{c_2} = \left(\frac{N_1}{N_2}\right)^{2} \cdot \bar{\bar{\sigma}}_{c_1} \qquad (4)$$

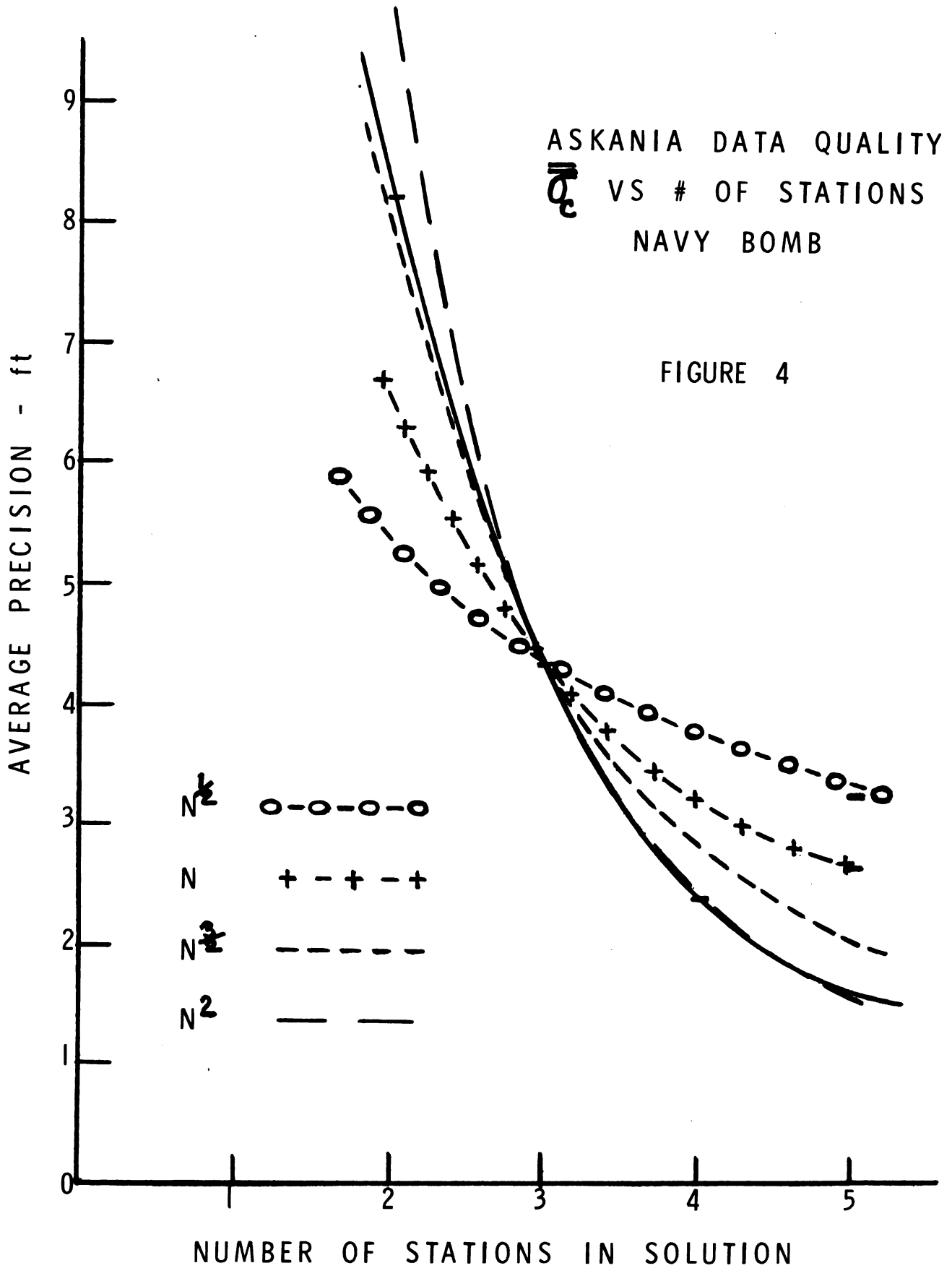
$$\bar{\bar{\sigma}}_{\bar{x}} = \frac{\sigma_x}{(2N)^{\frac{1}{2}}} \qquad (5)$$

$$\sigma_{\bar{x}} = \frac{\sigma_x}{2N} \qquad (6)$$

$$\sigma_{\bar{x}} = \frac{\sigma_x}{(2N)^{\frac{3}{2}}} \qquad (7)$$

$$\sigma_{\bar{x}} = \frac{\sigma_x}{(2N)^{2}} \qquad (8)$$

FIGURE 3



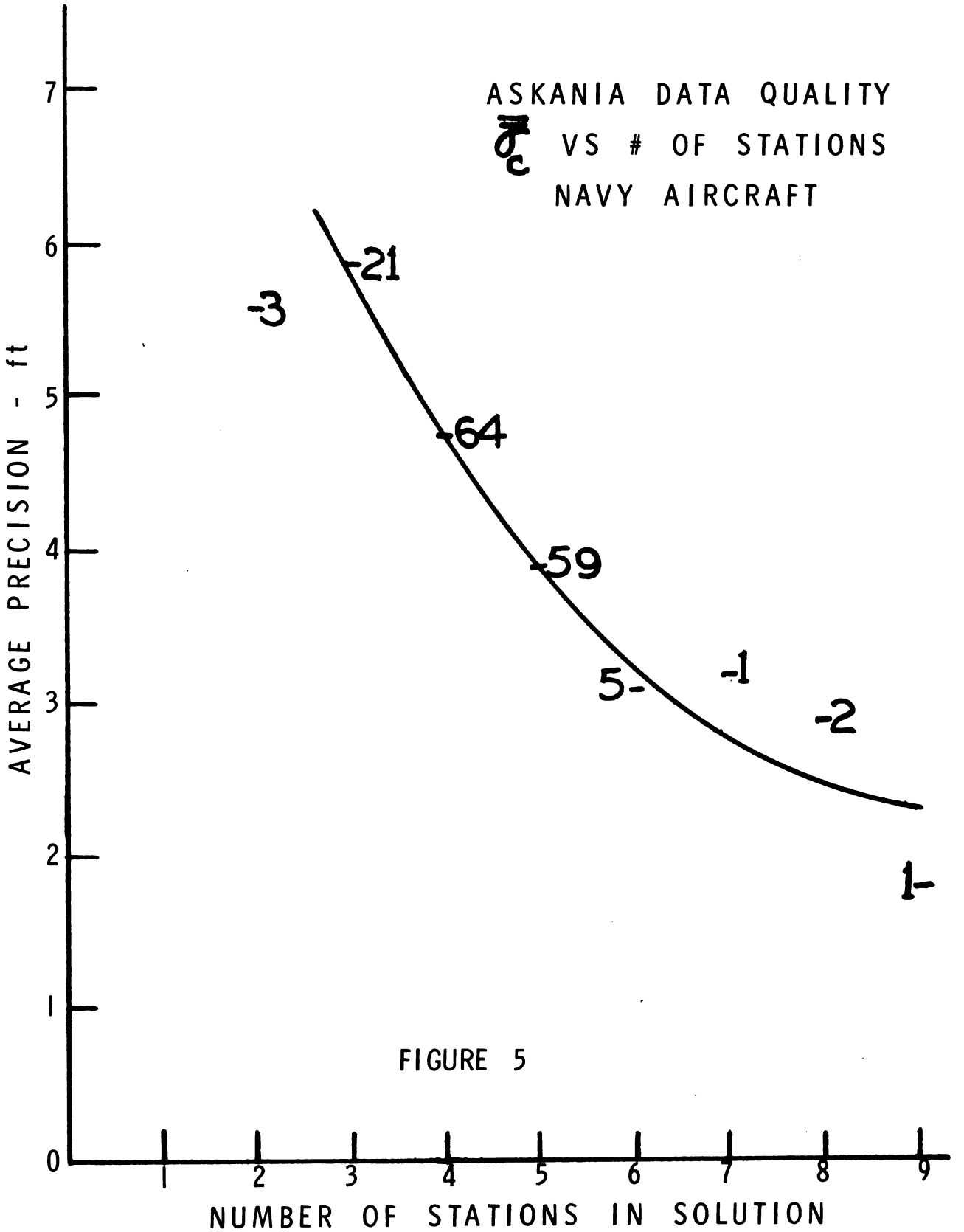


FIGURE 5

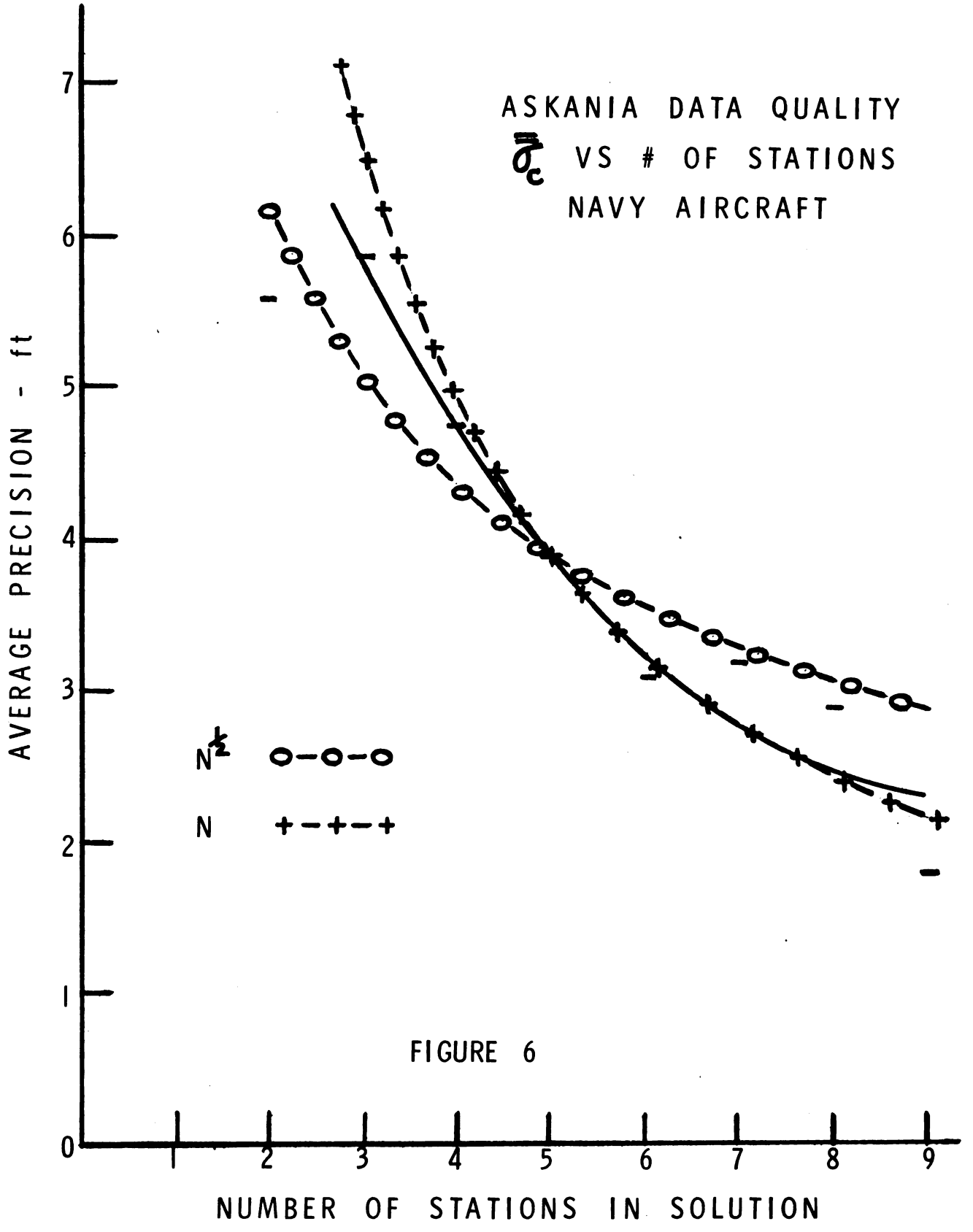


FIGURE 6

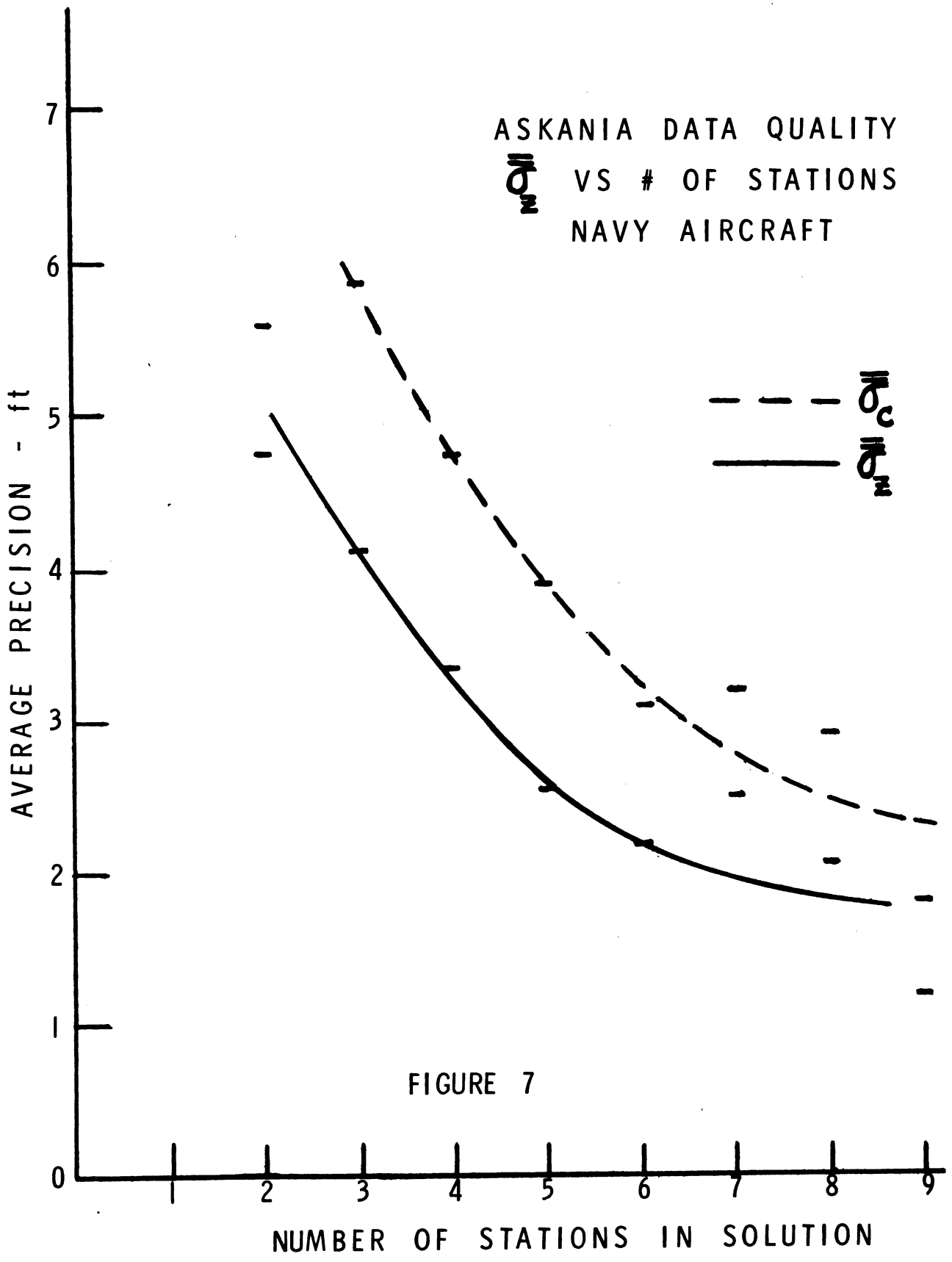


FIGURE 7

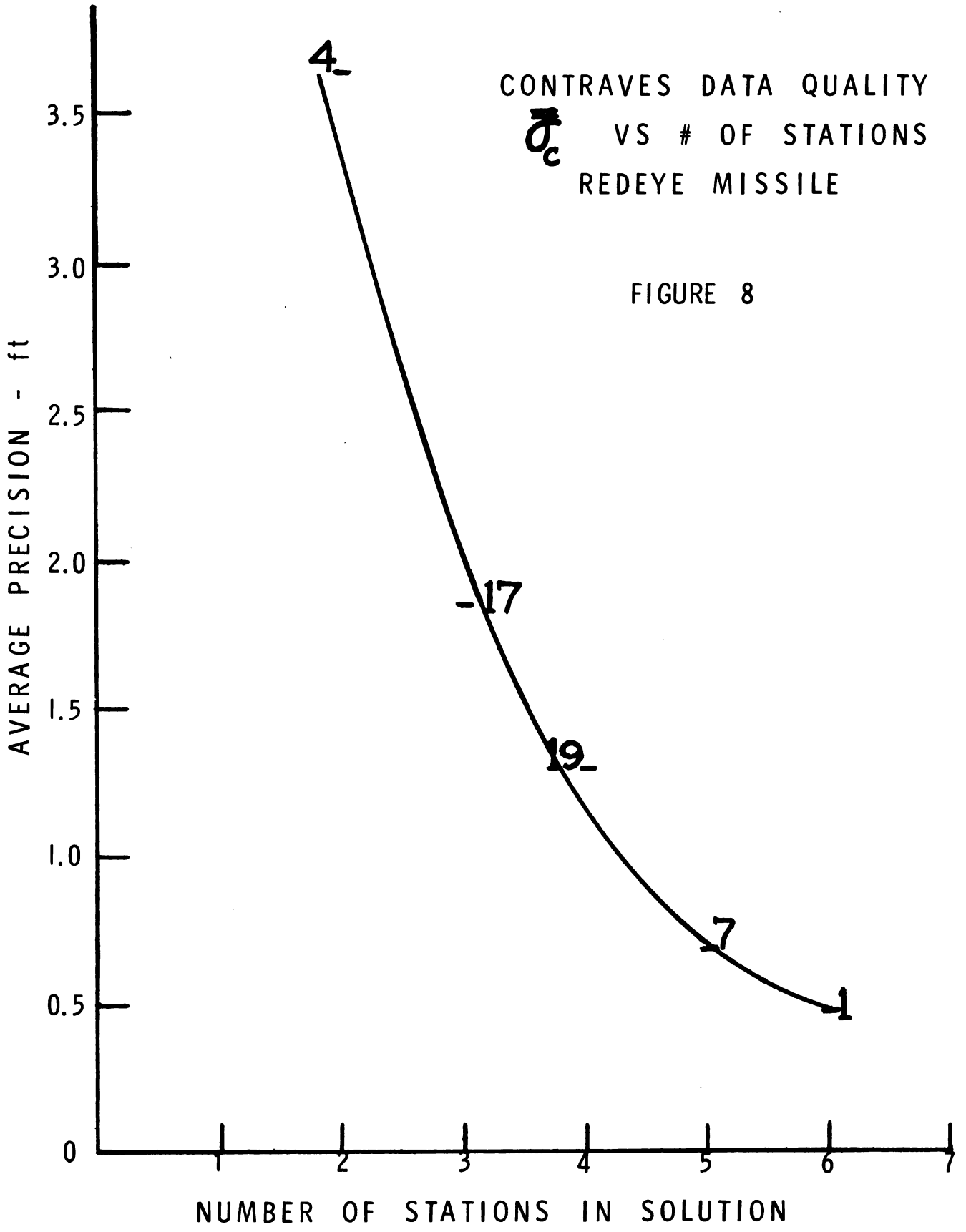


FIGURE 8

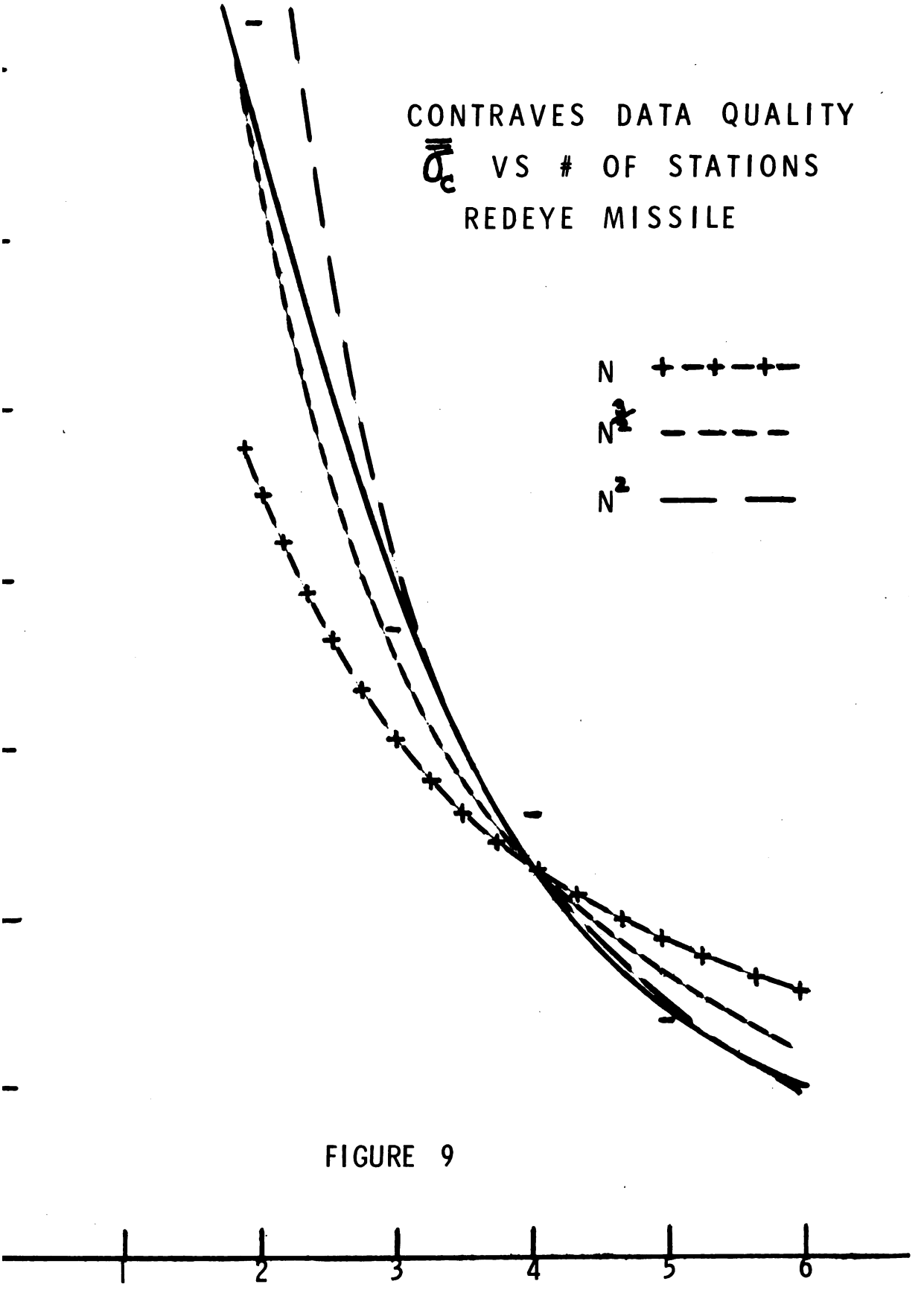


FIGURE 9

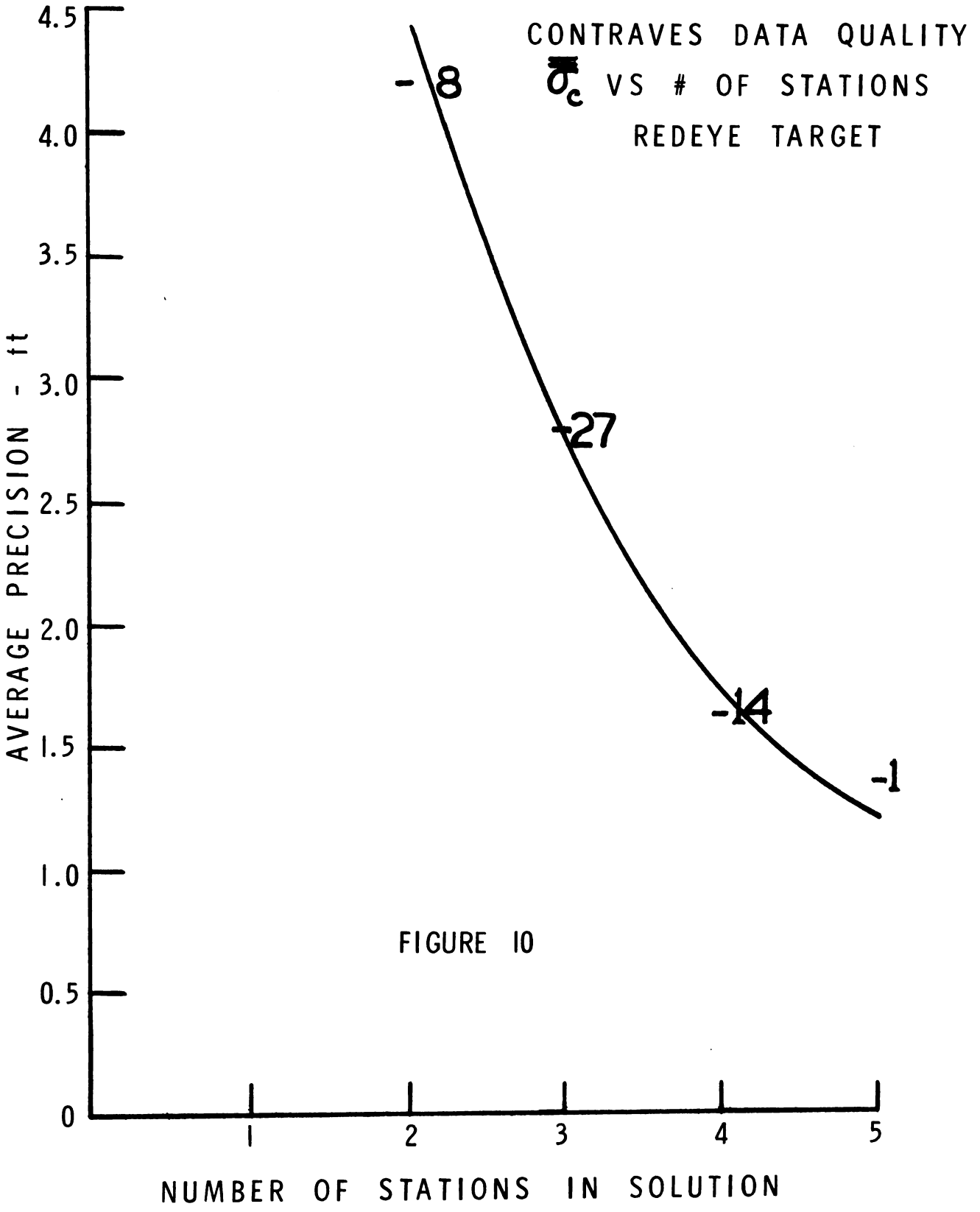


FIGURE 10

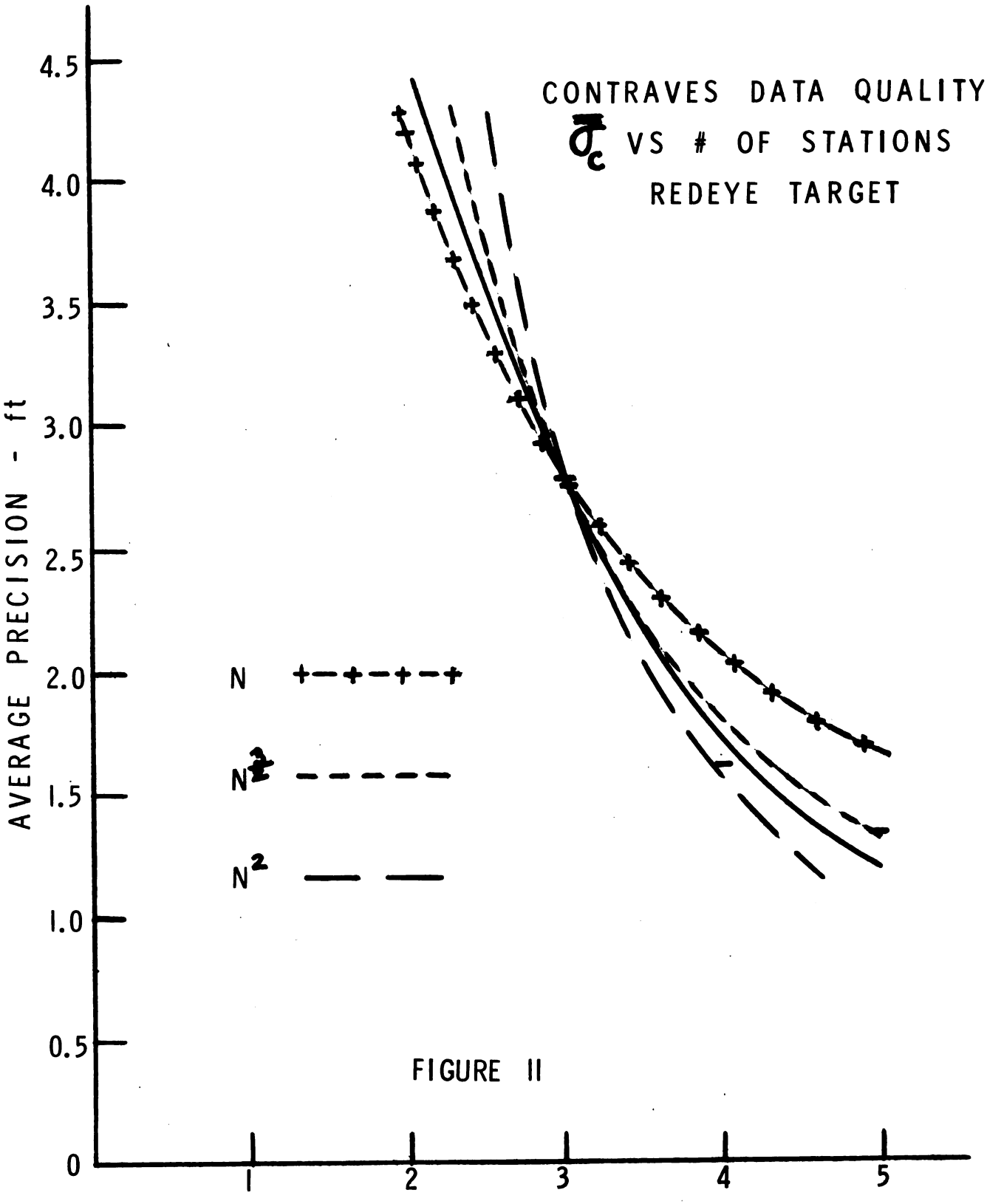


FIGURE II

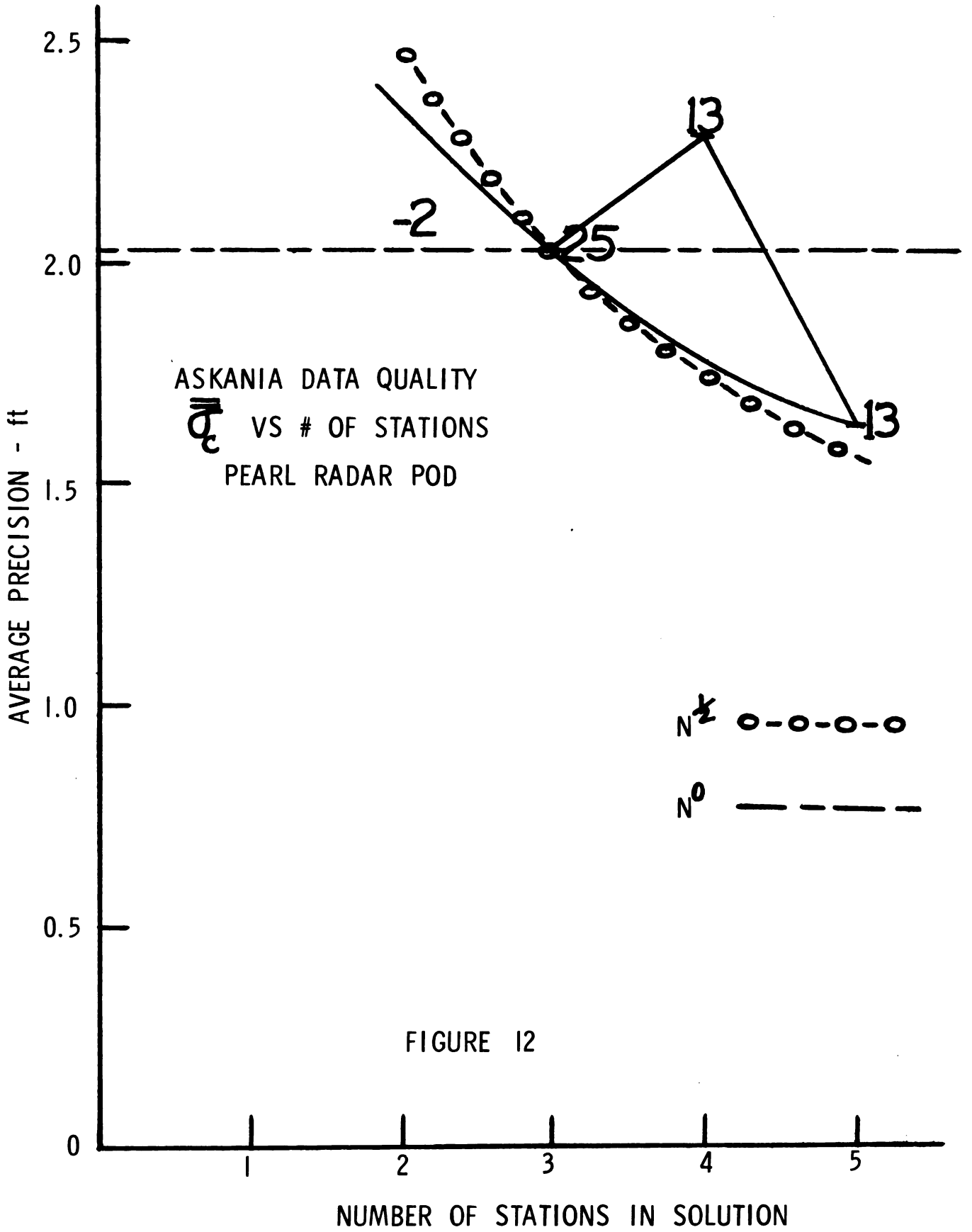


FIGURE 12

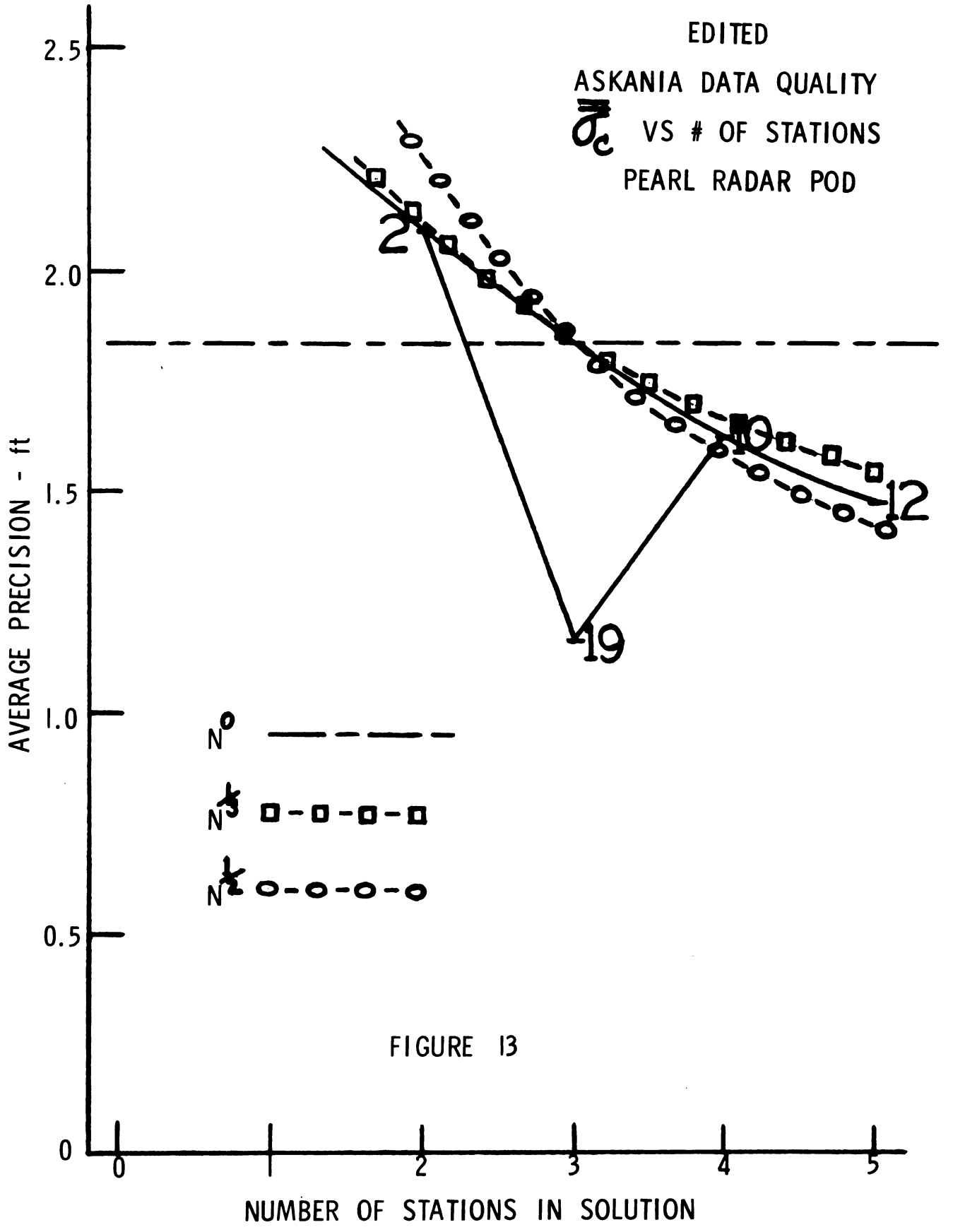


FIGURE 13

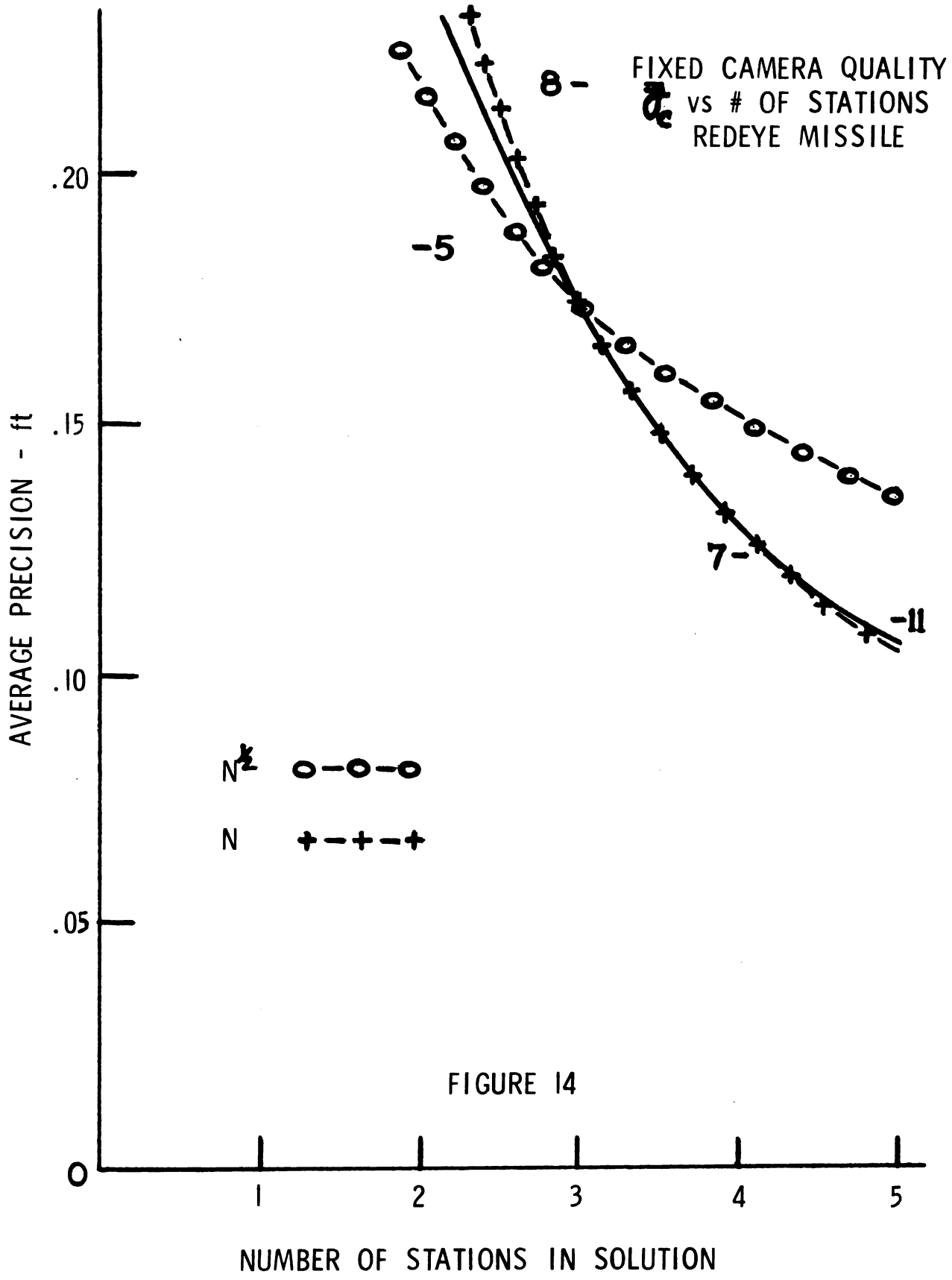
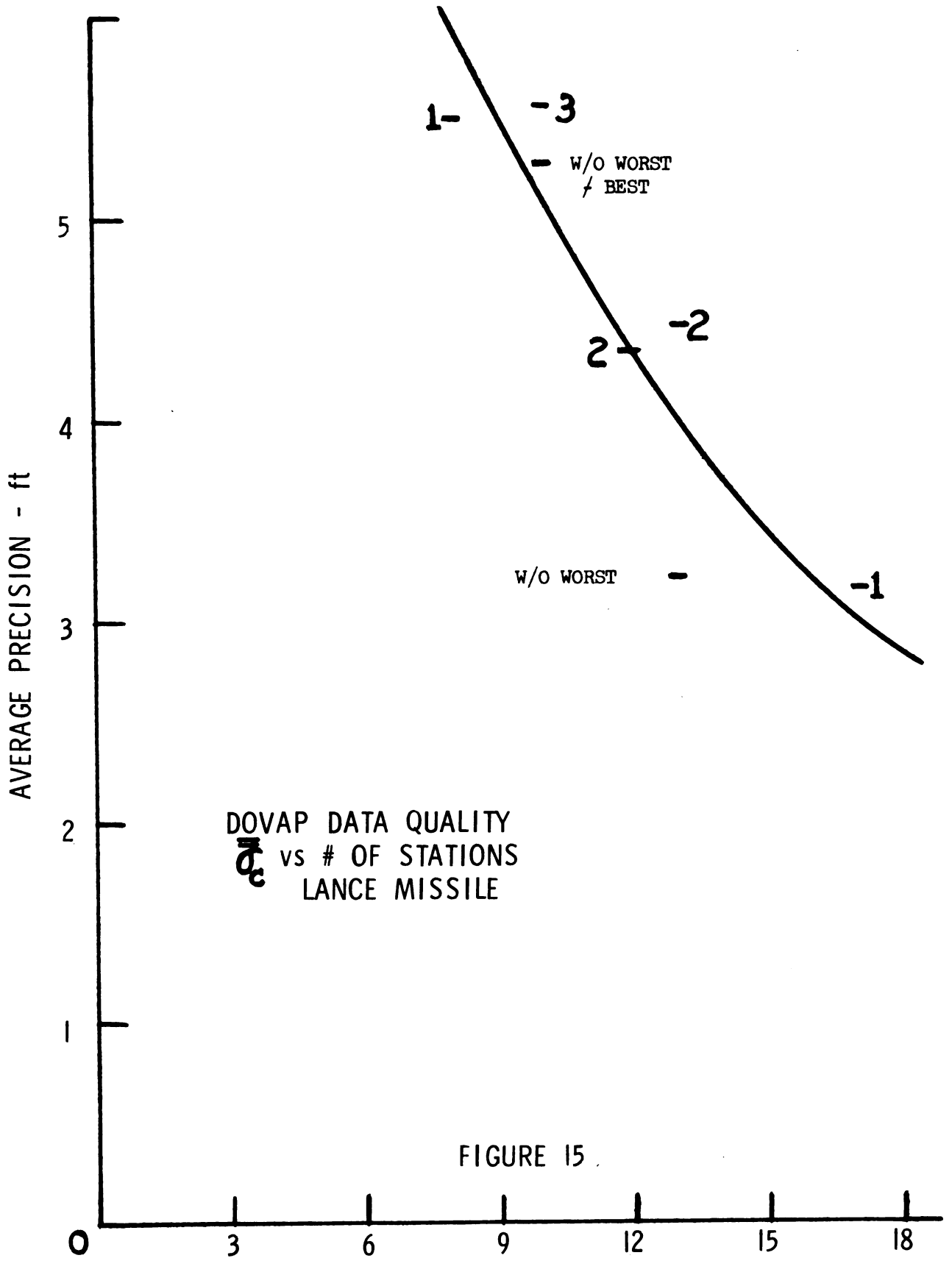


FIGURE 14



DOVAP DATA QUALITY
$\bar{\sigma}_c$ vs # OF STATIONS
LANCE MISSILE

FIGURE 15

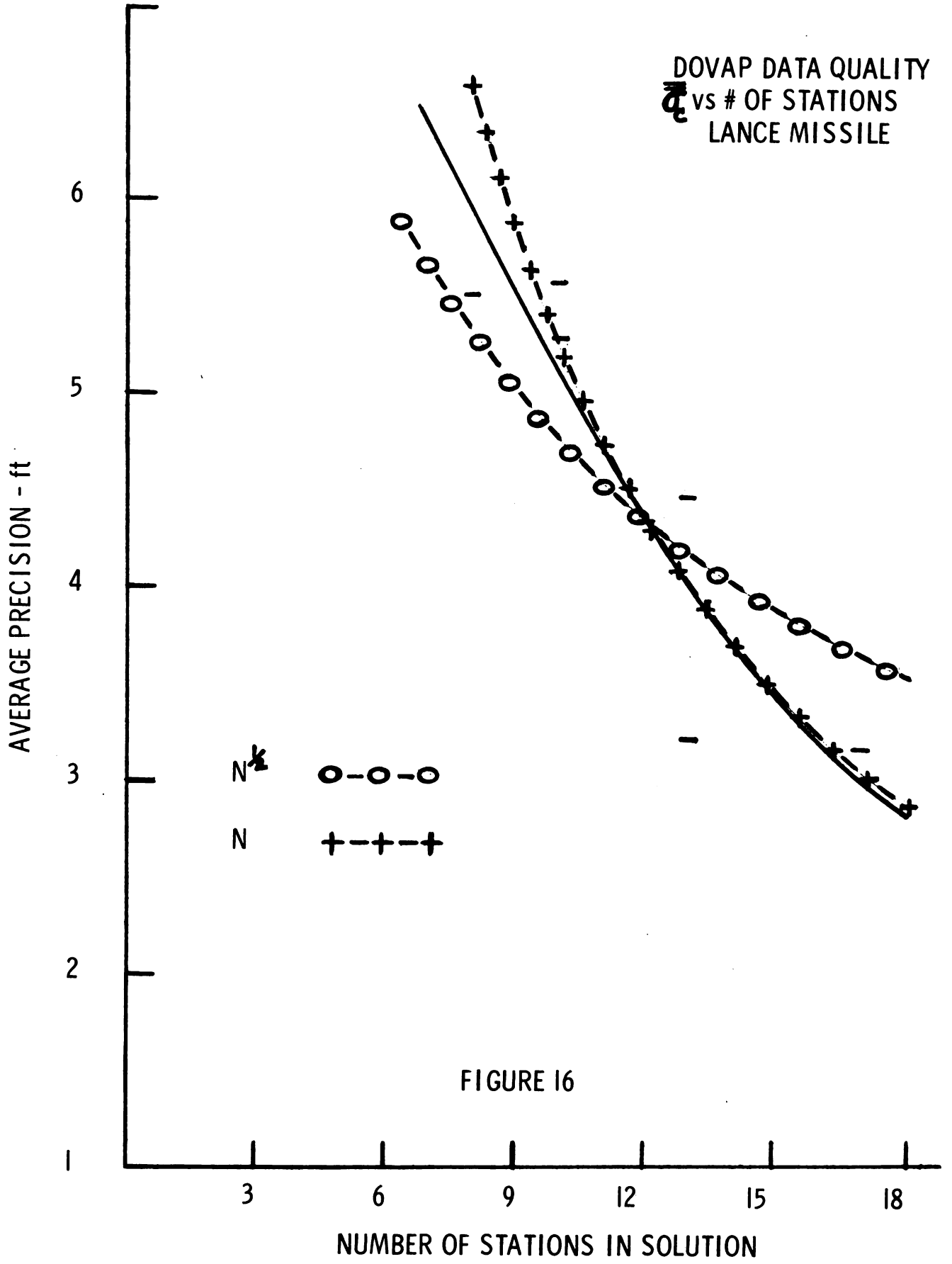
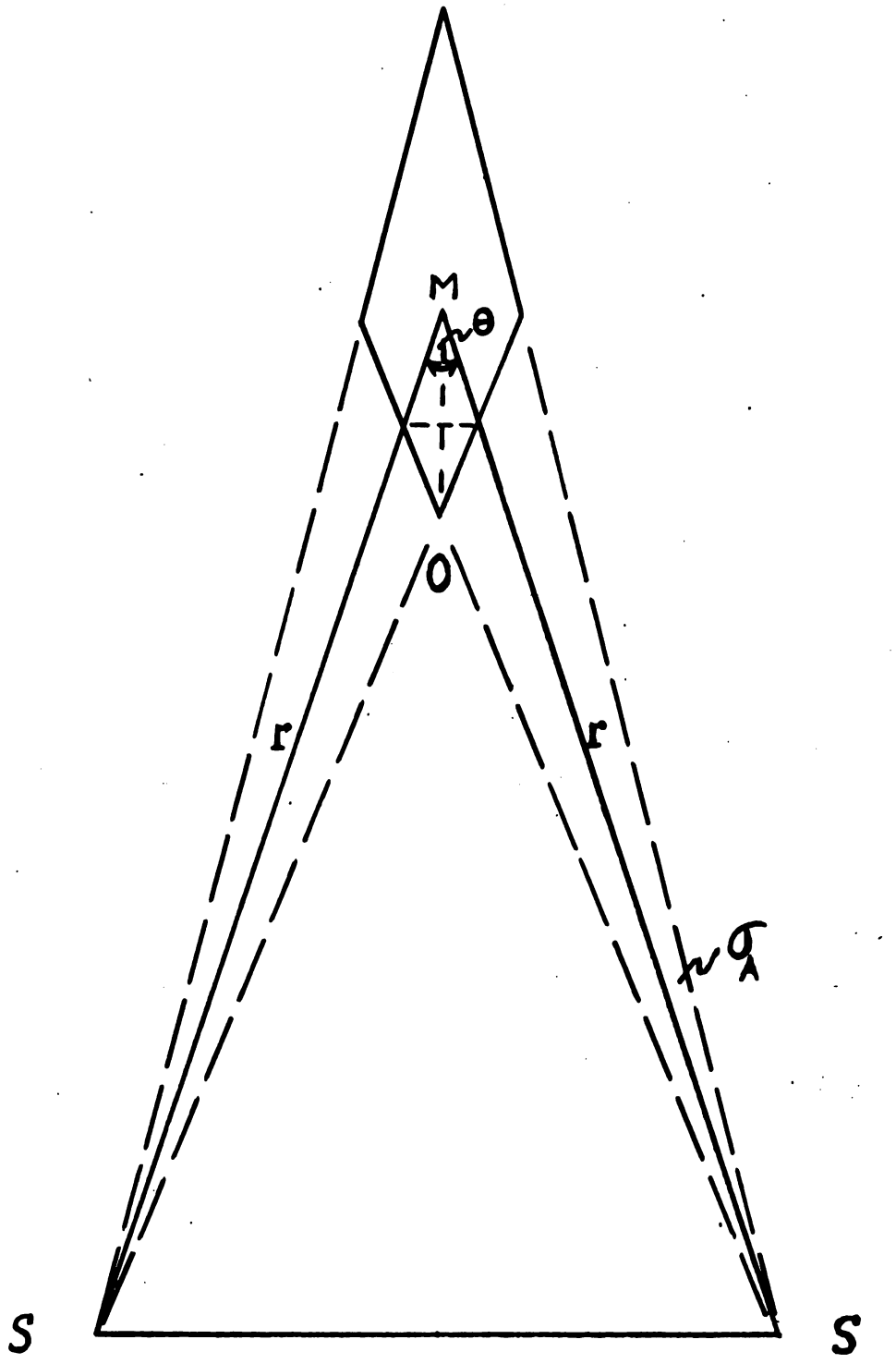


FIGURE 16

CINE POSITION PRECISION PROPORTIONAL TO:	$N^{\frac{1}{2}}$	N	$N^{\frac{3}{2}}$	N^{2}
	radomes (on A/C)	large A/C	drones (small A/C)	missiles bombs

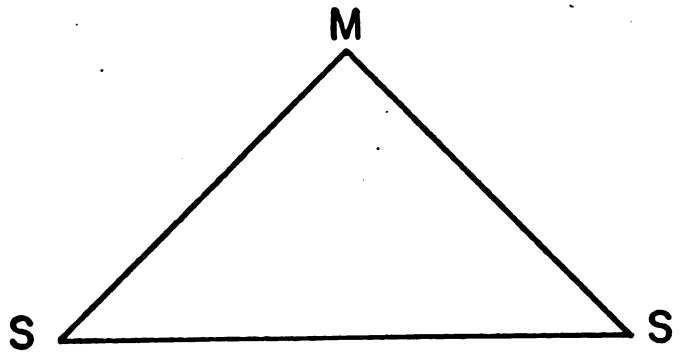
RATIO OF NUMBER OF STATIONS PROPORTIONAL TO:	RATIO BY WHICH REQ'D PRECISION IS BEING MET	$Ratio^{\frac{1}{2}}$	$Ratio^{\frac{3}{2}}$	$Ratio^{2}$
CINE POSITION	large A/C	missiles bombs	drones (small A/C)	radomes (on A/C)

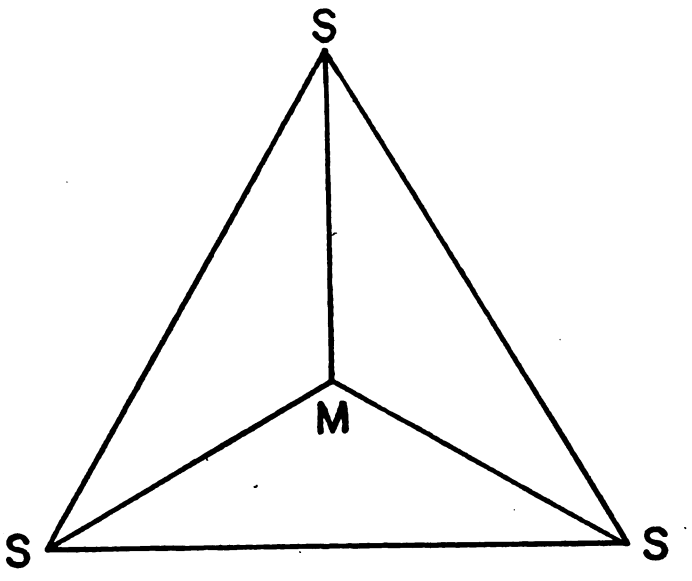
FIGURE 17

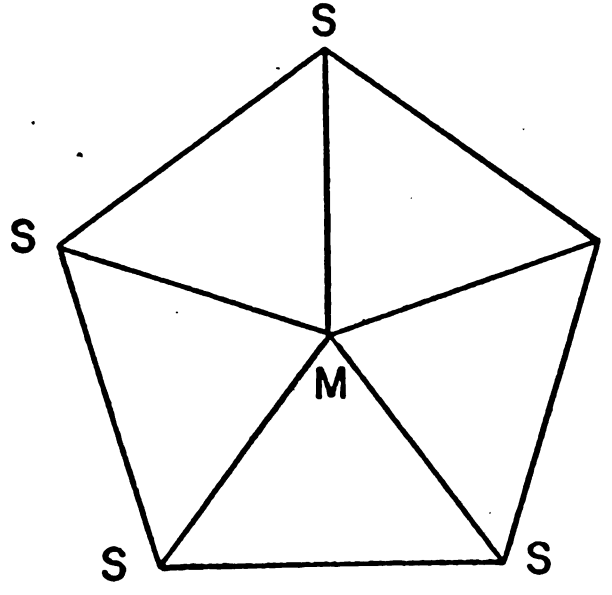


PRECISION OF POSITION FROM ANGULAR OBSERVATIONS

FIGURE 18







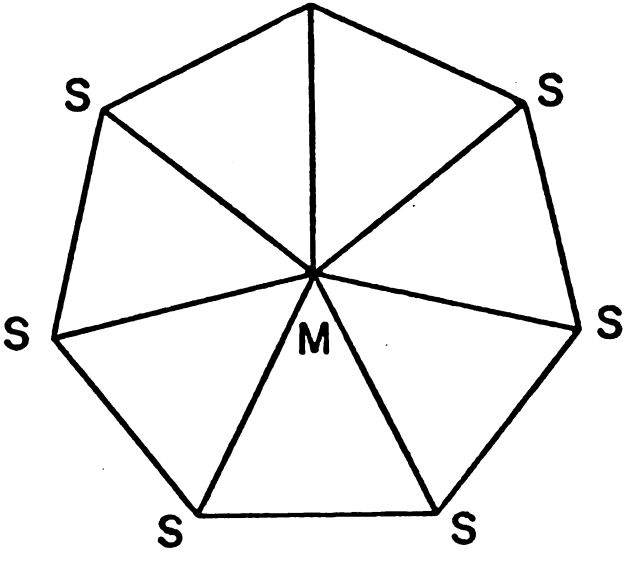


Figure 19

1. STATISTICALLY-OPTIMUM REDUCTIONS OF OBSERVATIONAL DATA TAKE ACCOUNT OF THE TOTAL VARIATION AMONG THE OBSERVATIONS; THEY OPTIMIZE THE MAXIMUM POSSIBLE DEGREES-OF-FREEDOM.

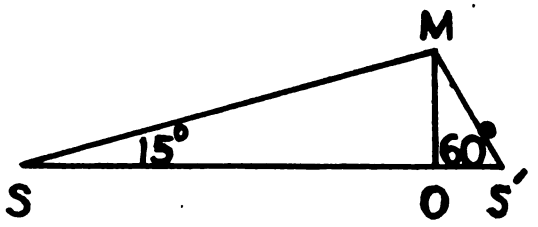
2. PHYSICALLY-, GEOMETRICALLY-, AND USEFULLY-OPTIMUM REDUCTIONS TREAT 'OBSERVATIONS" IN THE MOST RELEVANT FRAME-OF-REFERENCE - WHICH IS THAT OF THE DATA AS REPORTED TO THE USER. i.e., PHYSICALLY-, GEOMETRICALLY-, AND USEFULLY-OPTIMUM METHODS OPTIMIZE THE QUALITY OF THE USER'S DATA.

3. THE MOST VALID PRECISION OF AN OBSERVED DATA-VALUE IS THE ONE THAT MOST CLOSELY REFLECTS OUR ABILITY TO OBSERVE.

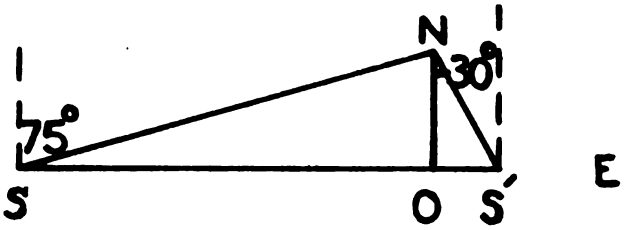
SUMMARY. TOTALLY-OPTIMUM REDUCTIONS ARE THOSE WHOSE DATA-QUALITY REFLECTS AGREEMENT AMONG THE ORIGINAL OBSERVATIONS IN THE FRAME-OF-REFERENCE OF THE USER'S DATA.

FIGURE 20

FIGURE 21

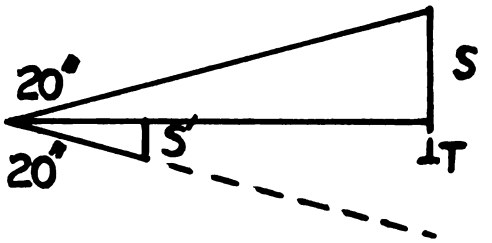


vertical

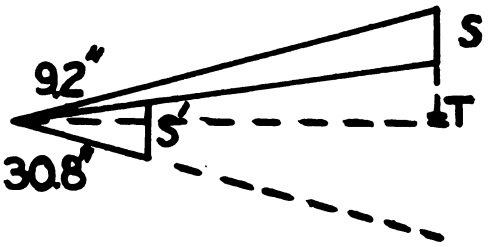


horizontal

Angular LS



Linear LS



For: OM = ON = 50,000 ft

$(5.6\ ft)^2 + (18.7\ ft)^2 = 2(13.8\ ft)^2$

$\Sigma = 24.3$ ft rms = 13.8 ft

$\Delta S = 18.7 - 8.6 = 10.1$ ft

OM = ON = 50,000 ft

$(8.6\ ft)^2 + (8.6\ ft)^2 = 2(8.6\ ft)^2$

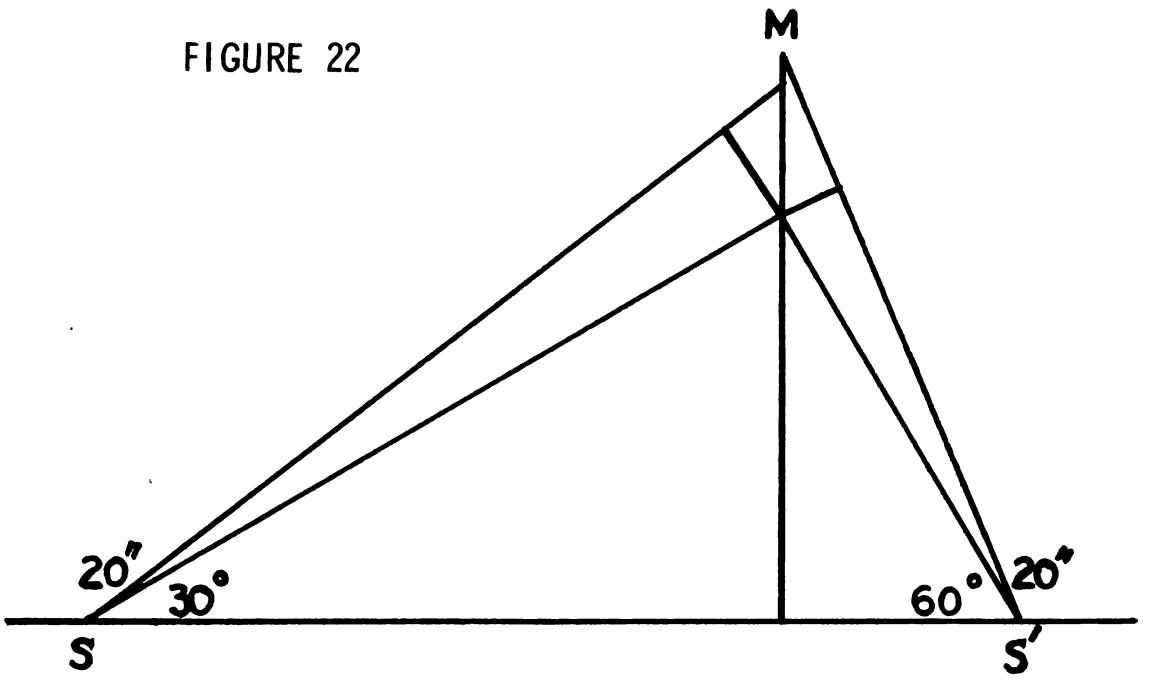
$\Sigma = 17.2$ ft rms = 8.6 ft

$\Delta S' = 8.6 - 5.6 = 3.0$ ft

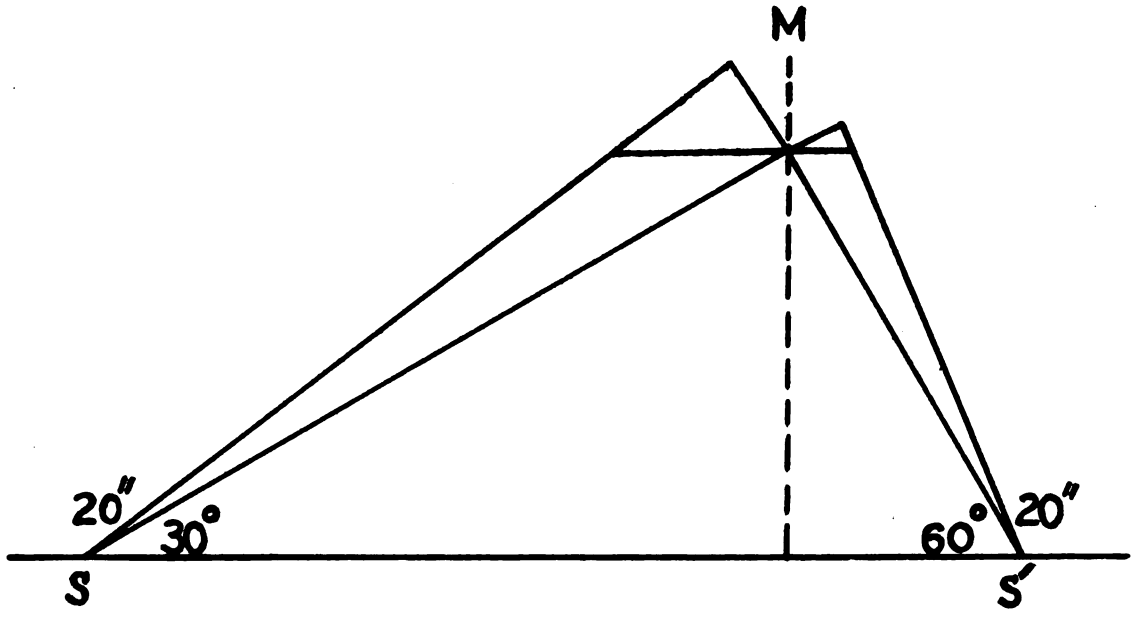
$$\frac{13.8 - 8.6}{13.8} \times 100\% = 38\% \text{ in rms}$$

$$\frac{13.8^2 - 8.6^2}{13.8^2} \times 100\% = 61\% \text{ in ms}$$

FIGURE 22



SUBTENDERS ON NORMAL



SUBTENDERS ON PARALLEL

From Pearson and Hartley

VALUES OF t FOR 68.27% CONFIDENCE LEVEL

DEGREE OF FREEDOM	t.6827
1	1.838
2	1.322
3	1.197
4	1.143
5	1.111
6	1.091
7	1.077
8	1.068
9	1.060
10	1.053
11	1.049
12	1.045
13	1.041
14	1.038
15	1.035
16	1.033
17	1.031
18	1.029
19	1.028
20	1.027
21	1.025
22	1.024
23	1.023
24	1.022
25	1.021
26	1.020
27	1.019
60	1.009
∞	1.00

FIGURE 23

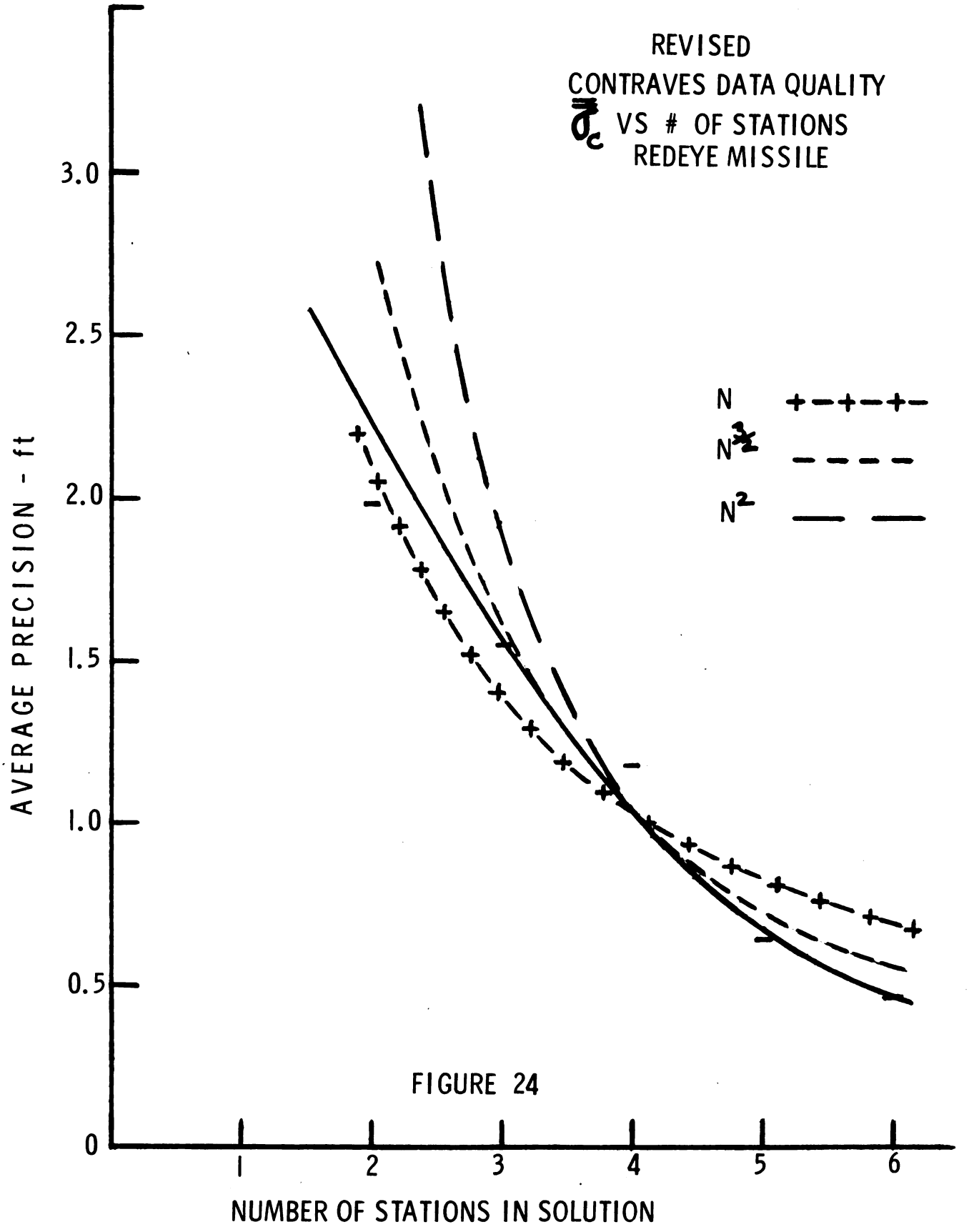


FIGURE 24

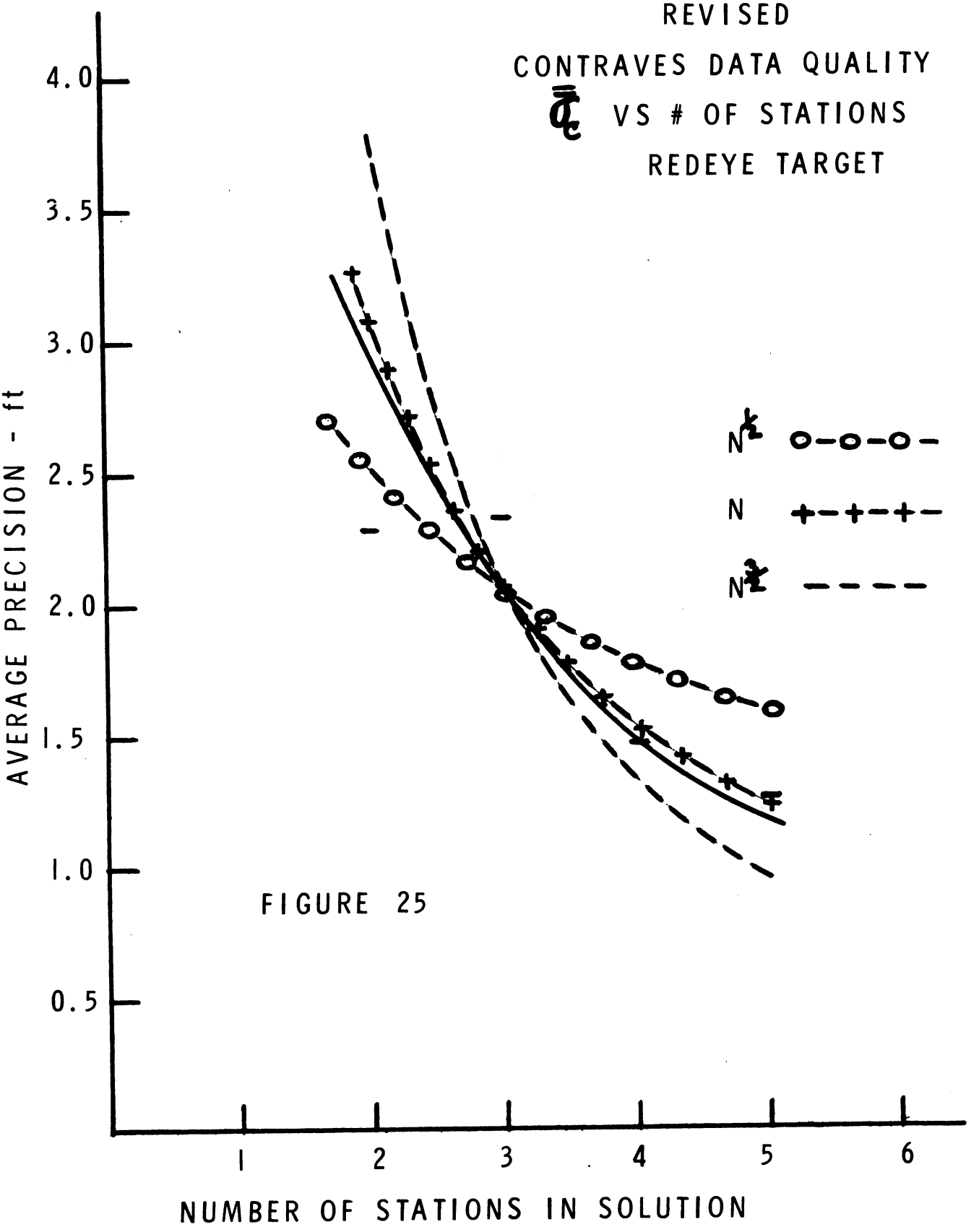


FIGURE 25

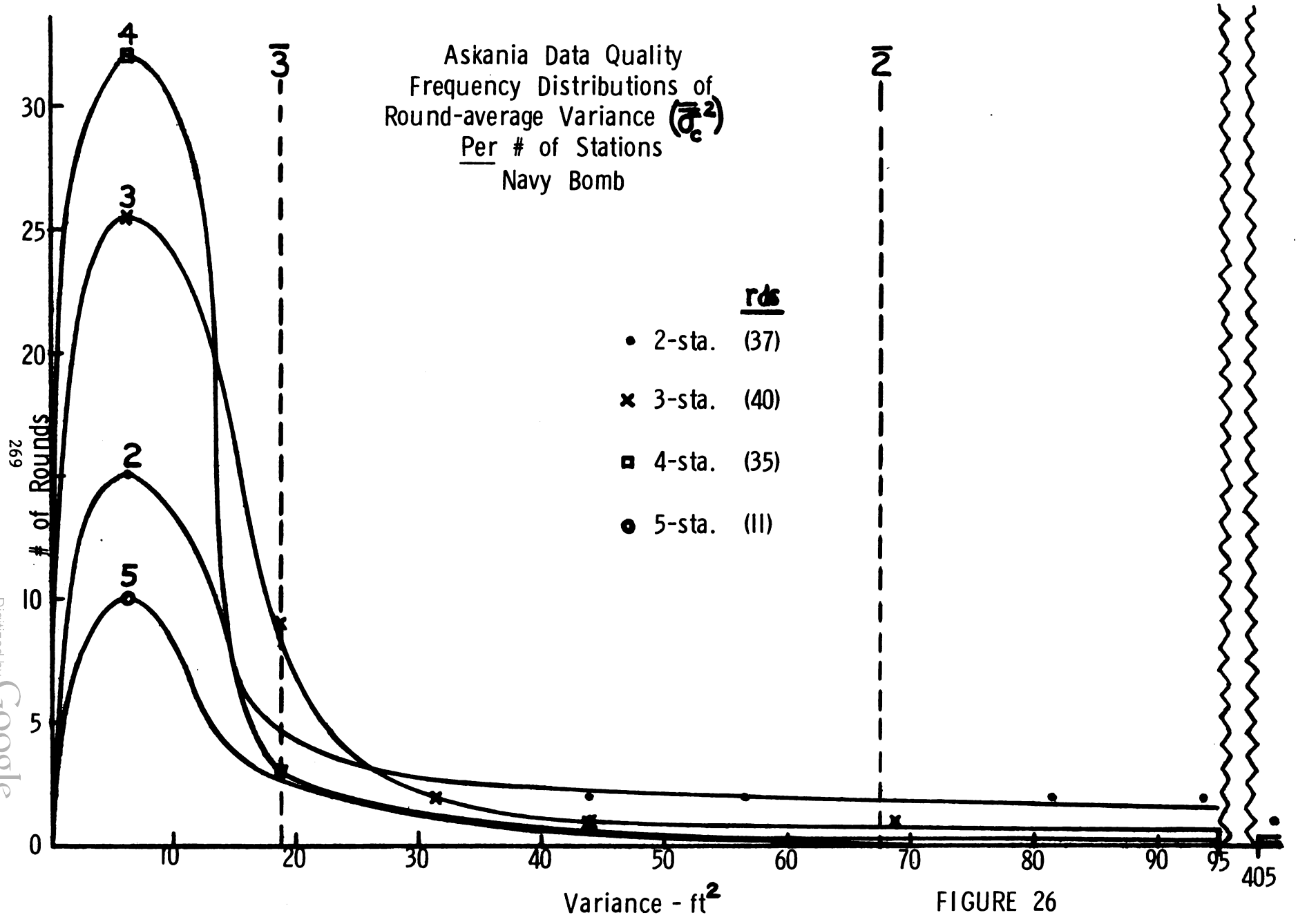


FIGURE 26

FIGURE 27

MECHANISMS OF IMPROVEMENT OF FLIGHT MEASUREMENT PRECISION BY INCREASING STATIONS

1. BY OVERCOMING SMALL SAMPLE-SIZE.
2. BY OVERCOMING NONOPTIMUM TEST CONFIGURATION.
3. BY OVERCOMING NONOPTIMUM CHOICE OF VARIABLE TO BE OPTIMIZED.
4. BY OVERCOMING OUR INABILITY TO OPTIMIZE EACH COORDINATE.
5. BY OVERCOMING NONOPTIMUM ESTIMATE OF DEGREES-OF-FREEDOM.
6. BY OVERCOMING ERRORS INCURRED IN APPLYING STATISTICS TO FLIGHT-MEASUREMENT.

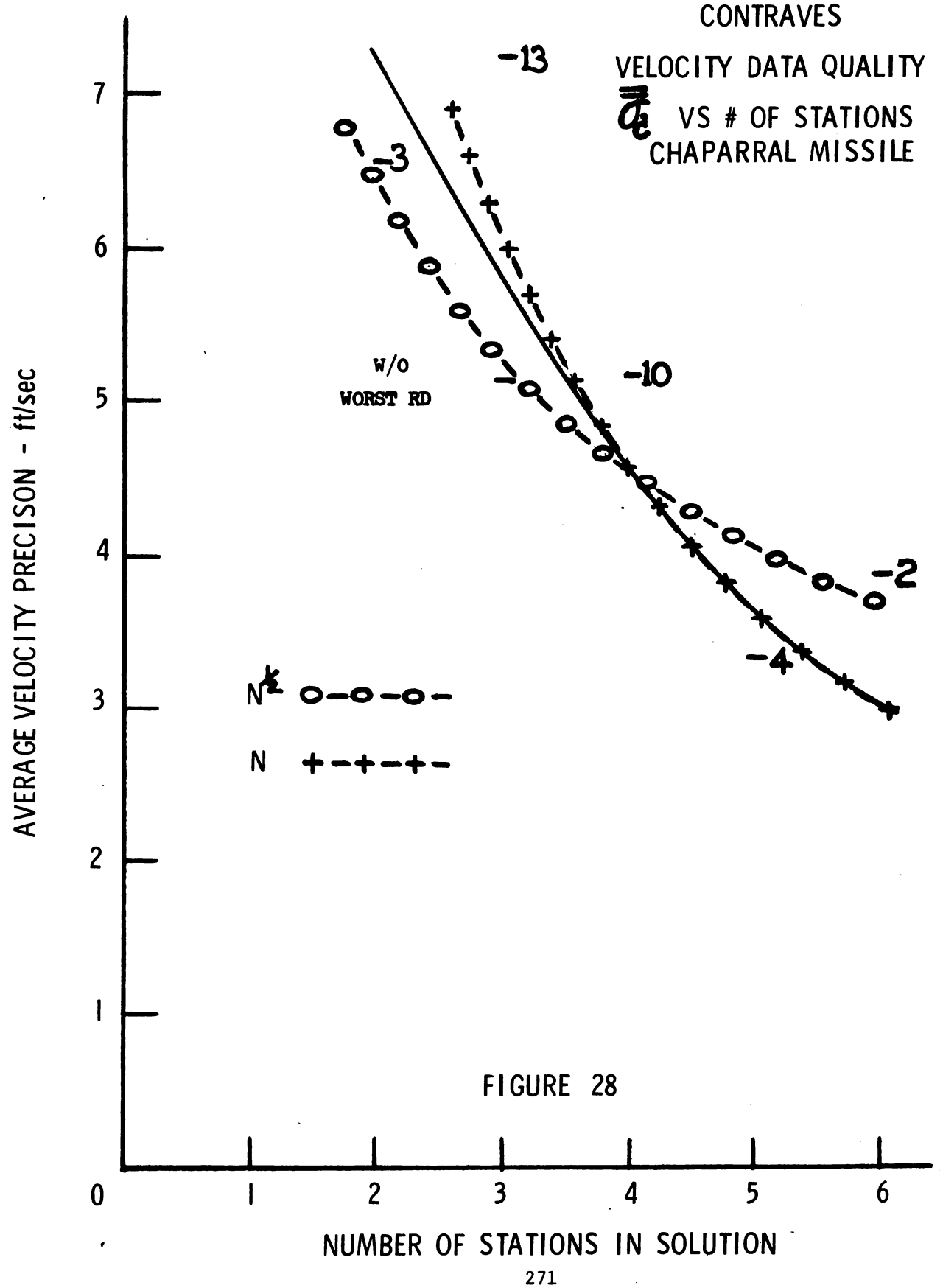


FIGURE 28

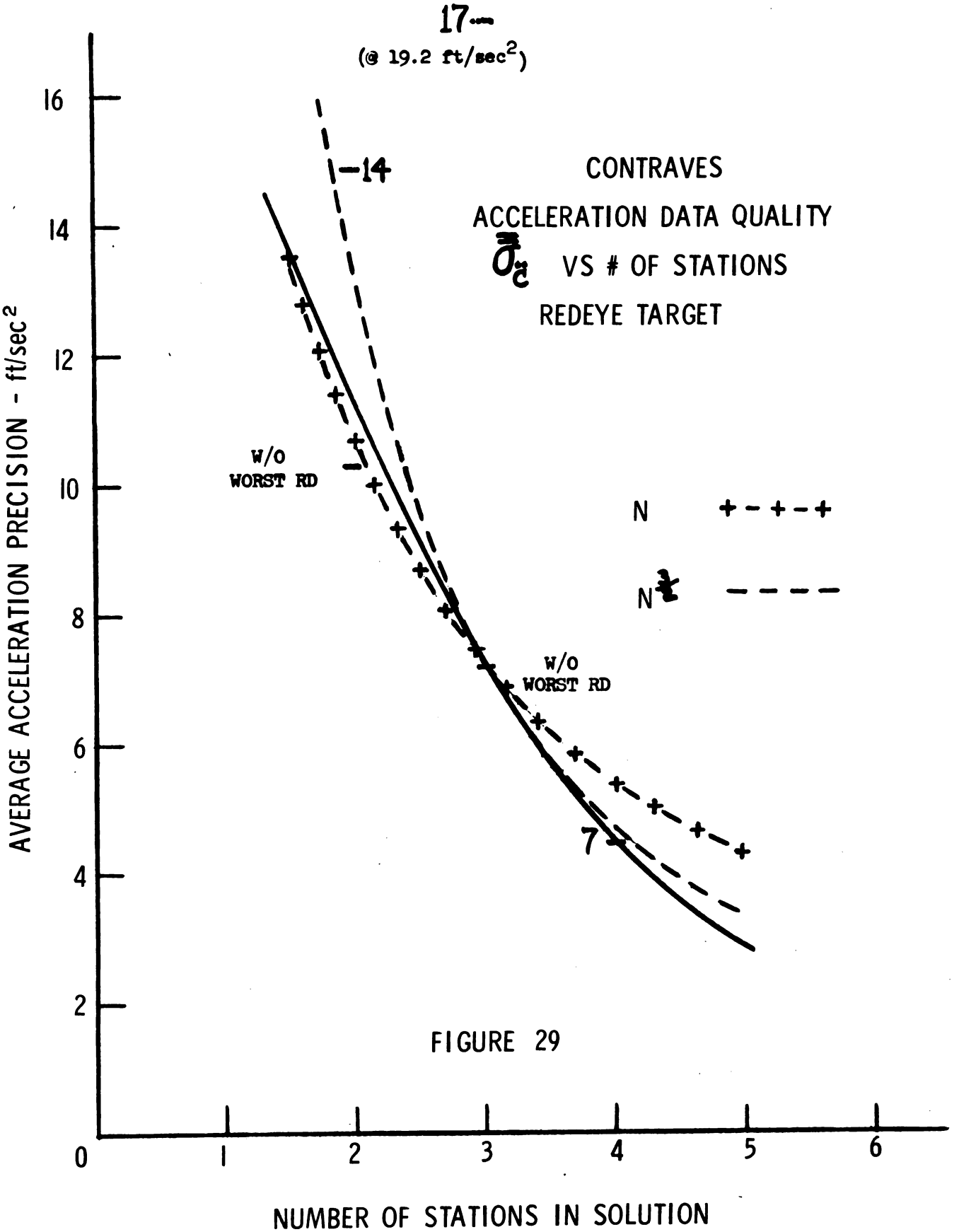


FIGURE 29

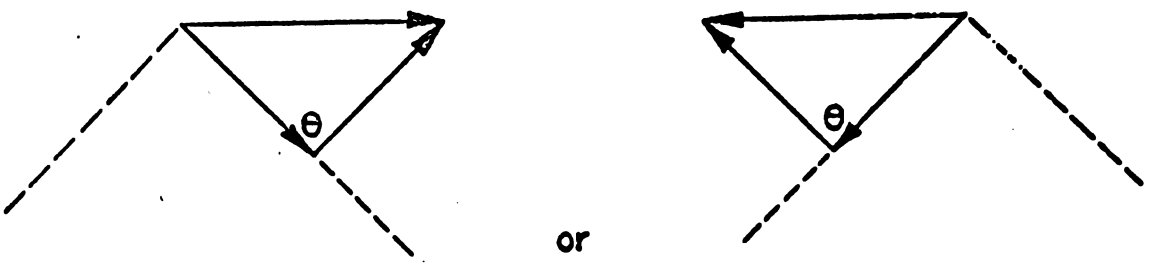


Figure 30a Same sign

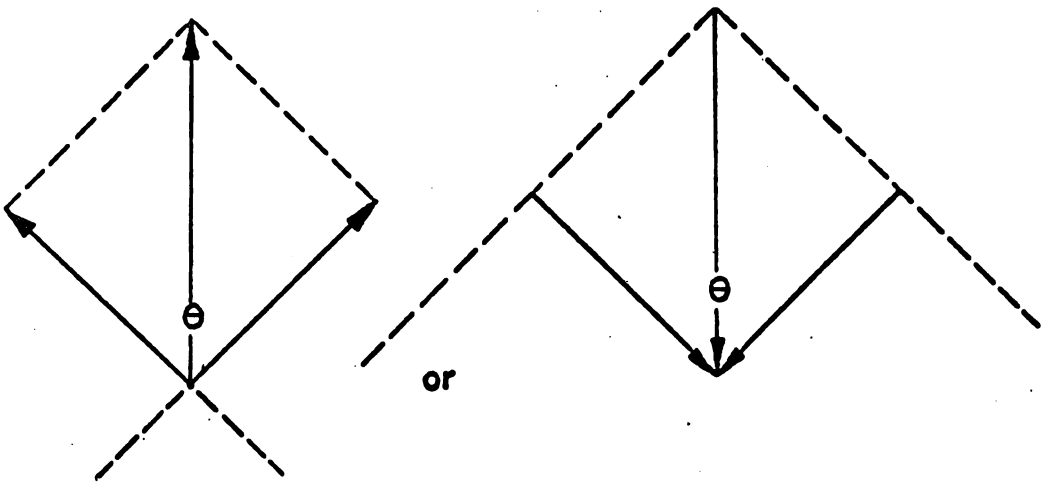


Figure 30b Opposite signs

COMMENTS ON PRESENTATION BY FRED HANSON

William Kruskal
Department of Statistics
The University of Chicago
Chicago, Illinois

You will recall that, when we met at the Edgewood Arsenal conference, I expressed regret that I could not attend the session at which you presented your problem, and I added that I would look at the materials you had sent to me.

Not everything in those materials is clear to me, but I take it that your major worry is that the empirically determined standard deviations of position determinations, as a function of number of stations, N, decreases faster when N grows than the $N^{-1/2}$ rate that would be expected under standard circumstances. You apparently have evidence that the rate is more like $N^{-3/2}$.

The first thought that comes to mind is that the standard $N^{-1/2}$ rate depends squarely on the assumption that the observations are uncorrelated and have equal variances. In particular, if the observations have equal variances but *negative* correlations, then the standard deviation of the sample mean is *less* than that expected under the standard assumptions.

Let me make this specific. Suppose, for simplicity, that we are dealing with N random variables, all with variance σ^2 and such that any pair of variables has correlation ρ. It is a standard fact that ρ cannot be less than $-1/(N-1)$.

Under these circumstances, the standard deviation of $\bar{X}$ (the average of X_i) is

$$\sigma \Big/ \sqrt{\frac{N}{1 + (N-1)\rho}} .$$

Suppose that

$$\rho = \frac{c - N^2}{N^2 (N-1)}$$

for some positive constant c. Then, substituting back, we would have for the standard deviation of $\bar{X}$,

$$\frac{\sigma}{\sqrt{N^3/c}} .$$

It seems to me conceivable that something like the above may be taking place for your radar measurements. Suppose that a measurement error comes from small changes in the angular orientations of the object measured. Then the effect of such a small change on one radar station might be nearly linearly related to the effect on another station, and with a negative slope.

Of course this is all speculation because I do not understand the measurement set-up and the data reduction method. In particular, it would be strange for ρ to depend strongly on N.

With cinetheodolites, it is hard for me to see offhand how large negative correlations could be effective.

More basically, it is not clear to me how your empirical standard deviations were obtained. Is it possible that your results are a result of something about that method?

EVALUATION OF NICKEL-IRON AND NICKEL-ZINC BATTERIES

Martin J. Sulkes
Power Sources Division, Electronics Components Laboratory
USAECOM, Fort Monmouth, N. J.

Most Army communication missions requiring secondary batteries are presently being met by the nickel-cadmium (Ni-Cd) system, with silver-zinc (Ag-Zn) filling the remainder of the missions that require low weight. Both of these systems contain expensive, limited-supply materials. Namely, silver at $32/lb. or cadmium at $3.25/lb.

The nickel-iron (Ni-Fe) and nickel-zinc (Ni-Zn) systems are potential low cost replacements for the Ni-Cd and Ag-Zn systems, since Zn and Fe are less than $0.15/lb. Ni-Fe and Ni-Zn batteries have been known for many years, however, until the present they have not developed the energy densities and life of which they are theoretically capable. For example, iron has a theoretical capacity of 0.98 Ah/gm compared to 0.47 Ah/gm for Cd and approximately 5% higher voltage. However, the Ni-Fe (Edison) cell has low energy density (8 Wh/lb) compared to 12-15 Wh/lb for Ni-Cd. In addition, the Edison cells low temperature and high rate performance are poor. Its cycle life, however, is excellent. Since much of the Edison cells' poor performance is due to the iron electrode, an improved iron such as was developed by GT & E labs could make this an attractive system.

The nickel-zinc system has had limited cycle life because of shorting by zinc dendrites and loss of capacity due to zinc electrode shape change. Energy density has been limited by the need to include a large excess of zinc. Recent work on Ag-Zn batteries and fundamental investigations of the zinc electrode have indicated how dendrite formation could be controlled and zinc shape change reduced. It was, therefore, estimated that through the use of an improved zinc electrode and the contractor's high energy nickel electrode a battery capable of delivering up to 30 Wh/lb for 200 or more cycles could be developed. However, a great deal of investigation of the various interrelated cell construction factors was required to successfully achieve the desired goals.

The objective of this work, therefore, was to optimize a design for nickel-zinc and nickel-iron and evaluate such cells in standard line configuration as possible low cost replacements for existing systems. A comparison of the discharge curves for an equal weight of all 4 electrochemical systems is shown in Figure 1. Specifically, this evaluation explored the construction of Ni-Fe and Ni-Zn cells for various design parameters, and tested them over a variety of rates and temperatures.

The nickel-iron system was investigated in two, 2^4 - design experiments, while the nickel-zinc system was studied in a replicated 2^3

experiment. Each experiment had 16 cells. In all cases, the assembly of cells, electrolyte fill and position on various tests were carried out in random order as determined from a table of random numbers. All experiments were analyzed by the technique of multiple linear regression. The calculations were made on a Scientific Data Systems 930 computer using the Multiple Linear Regression program from the IBM 360 SSP (Scientific Subroutine Package). Surprisingly, when the calculations from the first experiment were checked by the Yates method, an error was discovered in the IBM program. This was then corrected and the results obtained through manual and computer calculation were then equal.

The multiple linear regression technique used for data analysis assumes that the total response; i.e., the faradaic capacity, is a linear function of the independent variables (factors) being studied. The general equation is

$$Y = b_o + b_i X_1 + b_2 X_2 + \ldots + b_n X_n = b_o + \sum_{i=1}^{n} b_i K_i$$

where Y is the dependent variable (response) and X_1, X_2, . . . X_n, are the factors in the experimental study. The coefficients b_1, b_2 . . . b_n (partial regression coefficients) were determined by fitting the response data to the general equation. Each coefficient then became an effect value and an indicator of the effect of its factor on the total response, independent of the other factors. The sign of the coefficient ($\pm$) determined the direction of the effect in going from one level to another of the factor. The constant b_o is the intercept on the Y axis.

The first nickel-iron experiment consisted of 16 cells made with four variable construction factors each at two levels as shown in Figure 2. These cells were given a total of 16 charge-discharge cycles. Based on the pre-tested capacity of the positive electrodes, it was expected that these cells would have a nominal capacity of 6Ah in the normal, positive limiting design. However, when these cells were cycled, lower than normal capacities were obtained after several cycles. This low capacity was traced to difficulties in control of the chemical activation process for the iron electrodes used in these cells. To eliminate this problem, the next experiments were assembled with iron electrodes made by a controlled electrochemical activation process.

Despite this setback with the Experiment 1 cells, valuable experience was gained by the contractor on cell assembly techniques. Furthermore, the data from cycle one (Figure 3) upon statistical analysis did demonstrate the dependence of cell capacity upon the variable factors studied in the experiment. This analysis, shown in

Figure 4 was run after non-significant interactions were eliminated from the analysis. The F value of 38.188 indicates that the data fit the assumed linear relationship. With 9 and 6 degrees of freedom an F value exceeding 18.69 is significant at the 0.999 probability level. Since the computed t values are not mutually independent, they could only be used for ranking the order of importance of the variable factors and for showing the direction of effect of the variable levels. Thus, it can be said that variable C (electrolyte concentration) with a computed t value of 11.544 had the major effect on the cell capacity. The + sign shows that the high level (31% KOH) of this variable gives more capacity than the low level (21% KOH); in fact, the 31% KOH yielded 37% more capacity than the 21% KOH. Figure 5 ranks the variables in order of importance and shows the preferred variable level. Similar analyses of data from later cycles of the cells in Experiment 1 gave similar results, thus strengthening these conclusions. However, it should be pointed out that each succeeding cycle is not independent of the first cycle data, since the cell construction is fixed.

Preliminary studies of the dependence of cell capacity on charge rate and Ah input indicated that higher charge rates (C/2 to C/4 were more beneficial and more efficient than low charge rates (< C/8). Since the cells were positive limiting, this effect is a function of the positive plate, verifying previous experience with positive-limiting nickel-iron and nickel-cadmium batteries. Further studies are necessary to determine optimum charging conditions.

A second, 16 cell nickel-iron experiment was setup in accordance with the design shown in Figure 6. These cells all contained electrochemically activated iron electrodes as opposed to the chemically activated ones in the first Ni-Fe experiment. Because of the change in the iron electrodes, it was thought necessary to repeat the two most significant factors found in experiment 1.

A total of 8 charge-discharge cycles were run in accordance with the regime given in Figure 7 and analyzed. In the analysis the variable factors and their first-order interactions were the independent variables, while the Ah capacity was the dependent variable. In addition, percent capacity retention in Ah was analyzed by comparing the Ah capacities on cycle 4 with that obtained on cycle 5 after a 7 day charged stand. Figure 8 gives the actual effect values of the various factors on the dependent variable during the eight cycles run.

It is apparent that LiOH content (D), KOH concentration (C) and the interaction of these two variables have the greatest effect on Ah capacity, at C/4 rates, with the saturated LiOH better than no LiOH and 31% KOH better than 21% KOH. It is also interesting to note that the charge retention cycle (~ 5) and the high rate cycle (~ 7) disrupted the relative ranking of effects on subsequent cycles. Also, with respect

to percentage Ah charge retention (~ 5), the KOH concentration was the variable with the major contributing effect.

Therefore, on the basis of the two nickel-iron experiments the following design features were chosen for the optimum Ni-Fe cell design.

1.	Electrolyte Concentration	31% KOH
2.	Additive	LiOH saturated electrolyte
3.	Electrode Geometry	End plates are positives
4.	Separator	Nylon-Cellophane-Nylon
5.	Electrode Thickness	0.037"

These factors by no means completely define the system and, therefore, additional experiments will have to be run to determine the influence of such factors as positive to negative-capacity ratios; and quantity of overcharge per cycle which were at fixed levels in the experiments.

Nickel-zinc Cell Experiments

One experiment was run to date on Ni-Zn cells. The design of this experiment, shown in Figure 9, contains only 3 variable factors, each at two levels. Replication was provided since it was known that zinc systems are more erratic, particularly with regard to cycle life than the long-lived Ni-Fe or Ni-Cd system. Therefore, additional cells are required to achieve more reliable data analysis and also to provide for substitute cells in the case of premature failure.

Seven cycles were given to the cells in Ni-Zn experiment 1. The first cycle analysis is given in Figure 10. An F value exceeding 2.75 is significant at the 0.90 level, and a value exceeding 3.73 is significant at the 0.95 level. Thus, the fit of the Ni-Zn first cycle data to the regression curve is only fair. However, if non-significant interactions are eliminated the fit is greatly improved.

The computed t values indicate that interaction between the zinc electrode substrate thickness and electrode geometry (AB) is the major contributor to the ampere-hour variation observed. Electrode geometry (B) is the second most important contributor with the 15 negative, 14 positive cells (B+) producing more ampere-hour capacity than the 14 positive cells (B-). This result is to be expected, since the outer two positive electrodes in the 15 negative cells are probably more completely utilized and would show as increased capacity in these positive limiting cells. The zinc electrode substrate thickness (A) is the third most important variable, and the negative t value shows that the low level (0.0025 inch thick) is better than the high level (0.005 inch thick) of this variable. The excess ZnO variable apparently had little effect on initial Ah capacity. This was expected as its effect, if any, was more likely to show on cycle life.

The large interaction effect shown in Figure 11, required some study to explain, since the factor levels should not have been sufficient to produce changes of the magnitude found. However, since the main effect of this interaction was to reduce the performance below acceptable level, the cause had to be determined to avoid repeating it in future designs. Cell teardown analysis determined that all cells had been constructed so as to be tight. However, with the thick substrate (A+) and the lower number of negative electrodes (B-) there was an insufficient amount of compressible zinc to prevent excessive tightness in the cell, which was responsible for the significant reduction in cell capacity.

On the second and succeeding test cycles on the Ni-Zn experiment 1 cells, an intermittent internal shorting problem became apparent. No further data is presented, since this shorting problem made statistical data analysis unreliable. The shorts were particularly evident after a seven-day charge retention test. Only 6 of the 16 cells showed appreciable charge retention (from 33 to 77 percent). Examination of the internal structure of the shorted cells indicated that it was caused by zinc growth at the top edge of the electrode shorting over to the adjacent positive electrodes. In future experiments, this will be corrected in three ways: (1) coating the edges of zinc electrodes with an inert film forming agent; (2) additional separator height above electrodes; and, (3) less initial electrolyte fill.

Though this initial experiment did very little toward achieving optimization, it did succeed in pointing out several critical design parameters that must be considered before satisfactory performance can be obtained in a high energy density Ni-Zn cell. Two additional design experiments are planned to evaluate such construction factors as the negative to positive capacity ratios, the total number of plates (plate thickness), separator type and number of layers, amount of amalgamation of the negative, etc. These factors must all be explored before a Ni-Zn battery meeting the required goals can be fielded.

In Summary: The use of factorial design experiments has greatly reduced the number of cells required for the evaluation of these two electrochemical systems. This reduction in the number of cells is particularly important for secondary batteries, since by their nature, each cell can tie up testing space for many months as it repeatedly cycles. Several important design factors have been optimized for both systems though much more work remains.

This work was carried out by General Telephone and Electronics Laboratories, Inc., under Contract DAAB07-68-C-0102. Complete data for the experiments reported on, may be found in R & D Technical Report ECOM-0102-1 by Mr. T. Blickwedel of GT&E Labs issued in September 1968. The suggestions and assistance of Mr. Joseph Weinstein of the Electronic Components Laboratory, USAECOM is gratefully acknowledged.

Figure 1

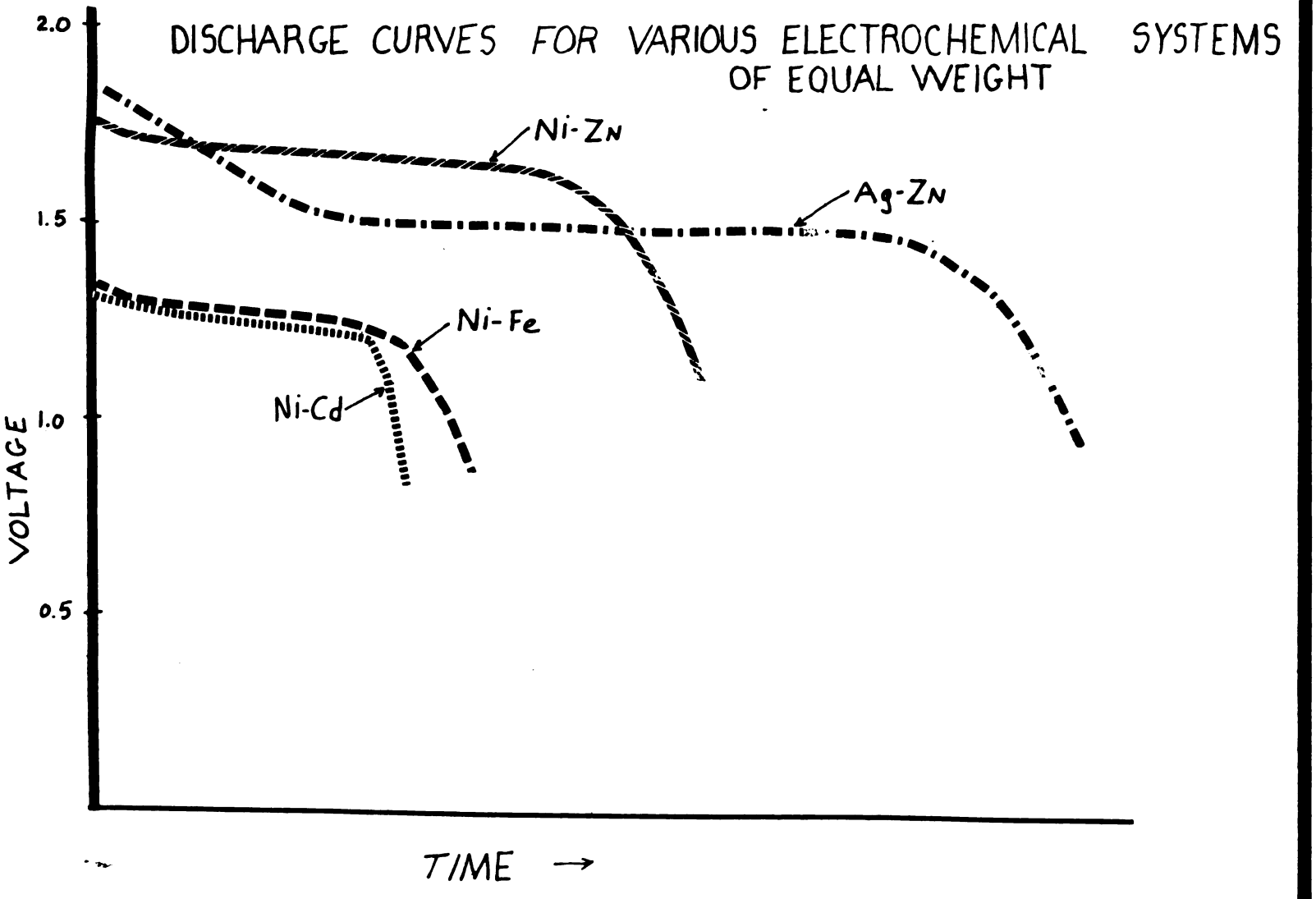


Figure 2

Ni-Fe Experiment 1 -- Factorial Design

Variable Factor	Low Level (-1)	High Level (+1)
(A) Separator	Nylon-Cellophane-Nylon	Polypropylene
(B) Electrode Geometry	9 pos, 10 neg	10 pos, 9 neg
(C) KOH Concentration	21%	31%
(D) LiOH Additive	No LiOH	LiOH Saturated

Cell Number	FACTOR LEVEL A	B	C	D	Statistical Analysis Designation
1	-1	-1	-1	-1	(1)
2	1	-1	-1	-1	a
3	-1	1	-1	-1	b
4	1	1	-1	-1	ab
5	-1	-1	1	-1	c
6	1	-1	1	-1	ac
7	-1	1	1	-1	bc
8	1	1	1	-1	abc
9	-1	-1	-1	1	d
10	1	-1	-1	1	ad
11	-1	1	-1	1	bd
12	1	1	-1	1	abd
13	-1	-1	1	1	cd
14	1	-1	1	1	acd
15	-1	1	1	1	bcd
16	1	1	1	1	abcd

Figure 3

Data From Ni-Fe Experiment One - Cycle One

Cell No.	Statistical Analysis Designation	Ah to 1.0 V
1	(1)	3.33
2	a	3.12
3	b	4.04
4	ab	3.88
5	c	5.44
6	ac	5.33
7	bc	5.95
8	abc	5.41
9	d	5.17
10	ad	4.33
11	bd	5.73
12	abd	5.15
13	cd	7.01
14	acd	6.00
15	bcd	6.28
16	abcd	6.64

Figure 4

Multiple Linear Regression Analysis

Ni-Fe Experiment 1, Cycle 1

Variable (Independent)	Standard Deviation	Correlation	Regression Coefficient	Std. Error of Reg. Coef.	Computed t value
A	1.0328	-0.174	-0.193	0.072	-2.684
B	1.0328	0.187	0.208	0.072	2.892
C	1.0328	0.748	0.832	0.072	11.544
BC interaction	1.0328	-0.132	-0.147	0.072	-2.036
D	1.0328	0.552	0.613	0.072	8.513
CD interaction	1.0328	-0.124	-0.138	0.072	-1.920
(dependent)					
Ah 5.177	1.148				
INTERCEPT	5.177				
MULTIPLE CORRELATION		0.981			
STD. ERROR OR ESTIMATE		0.288			

ANALYSIS OF VARIANCE FOR THE REGRESSION

Source of Variation	Degrees of Freedom	Sum of Squares	Mean Squares	F Value
Attributable to Regression	6	19.028	3.171	38.188
Deviation from Regression	9	0.747	0.083	
TOTAL	15	19.776		

Figure 5

Rank Of Variable Importance and Preferred Level
Ni-Fe Experiment I, Cycle I

Rank in Order of Importance	Variable	Preferred Level	Percent Increase of Preferred over Non-Preferred Level
1st	(C) Electrolyte concentration	31%	37.
2nd	(D) LiOH additive	Saturated LiOH	26
3rd	(B) Electrode geometry	10 positive, 9 negative	8.3
4th	(A) Separator	Nylon-Cellophane-Nylon	7.7

Figure 6

Ni-Fe Experiment 2 -- Factorial Design

Variable Factor	Low Level	High Level
(A) Electrode Thickness	0.030 "	0.037 "
(B) Electrode Geometry	pos, (+1) neg	(+1) pos, neg
(C) KOH Concentration	21%	31%
(D) LiOH Additive	No LiOH	LiOH Saturated

Cell Number	Factor - Level A	B	C	D	Statistical Analysis Designation
17	-1	-1	-1	-1	(1)
18	1	-1	-1	-1	a
19	-1	1	-1	-1	b
20	1	1	-1	-1	ab
21	-1	-1	1	-1	c
22	1	-1	1	-1	ac
23	-1	1	1	-1	bc
24	1	1	1	-1	abc
25	-1	-1	-1	1	d
26	1	-1	-1	1	ad
27	-1	1	-1	1	bd
28	1	1	-1	1	abd
29	-1	-1	1	1	cd
30	1	-1	1	1	acd
31	-1	1	1	1	bcd
32	1	1	1	1	abcd

Figure 7

Test Regime For Ni-Fe Cells

Cycle	Charge Rate	Discharge Rate	Stand Conditions
1	1.5A	1.5A	None
2	6.0A	"	"
3	3.0A	"	"
4	1.5A	"	"
5	1.5A	"	7 day at RT
6	0.75A	"	None
7	1.5A	12.0A	"
8	1.5A	1.5A	"

Figure 8

Ni-Fe Experiment 2 - - Magnitude of effect of the Independent Variables on the Dependent Variable

Cycle No	Dependent Variable	Mean	Independent Variable Name A	B	C	AC	D	AD	BD	CD	F Value for the Analysis of Variance for the Regression
1	Ah	4.773	0.454		0.278		0.904				11.825
2	Ah	4.589		0.234	0.278		0.890			0.265	17.248
3	Ah	4.525	0.225	0.237	0.262		0.968			0.331	18.900
4	Ah	4.409	0.231	0.225			0.975			0.350	18.970
5(Charge retention)	Ah	3.189			0.453		0.741			0.228	17.712
6	Ah	4.162	0.293	0.293			0.887			0.362	29.508
7	Ah	3.888	0.275		0.325		0.837			0.542	11.808
8	Ah	4.484			0.443		0.731	0.369	0.393		1.490
-	-	-	-	-	-	-	-	-	-	-	-
5% retention(Ah)		72.21	-3.06		7.46	-2.11					7.686

Figure 9

Ni-Zn Experiment I - - Factorial Design

VARIABLE FACTOR	LOW LEVEL (-I)	HIGH LEVEL (+I)
(A) Zinc Electrode Substrate Thickness	0.0025 "	0.005
(B) Electrode Geometry	14 pos, 13 neg	14 pos, 15 neg
(C) ZnO Additive	Chemically saturated	2 g ZnO excess/cell

Cell Number	Factor - Level A	B	C	Statistical Analysis Designation
1	-I	-I	-I	(I)
2	-I	-I	-I	(I)
3	I	-I	-I	a
4	I	-I	-I	a
5	-I	I	-I	b
6	-I	I	-I	b
7	I	I	-I	ab
8	I	I	-I	ab
9	-I	-I	I	c
10	-I	-I	I	c
11	I	-I	I	ac
12	I	-I	I	ac
13	-I	I	I	bc
14	-I	I	I	bc
15	I	I	I	abc
16	I	I	I	abc

Figure 10

MULTIPLE REGRESSION - - Ni-Zn Experiment I, Cycle I

Independent Variable	Standard Deviation	Correlation Indep. vs Dep.	Regression Coefficient	Computed T Value
A - Zn substrate thickness	1.0328	-0.34523	-0.39583	-1.89132
B - Electrode geometry	1.0328	0.37430	0.42917	2.05059
AB	1.0328	0.63231	0.72500	3.46410
C - ZnO additive	1.0328	0.13082	0.15000	0.71671
AC	1.0328	-0.08358	-0.09583	-0.45970
BC	1.0328	0.07631	0.08750	0.41808
ABC	1.0328	-0.21077	-0.24167	-1.15470
Dependent Variable				
Ah (Mean) 5.94167	1.18419			

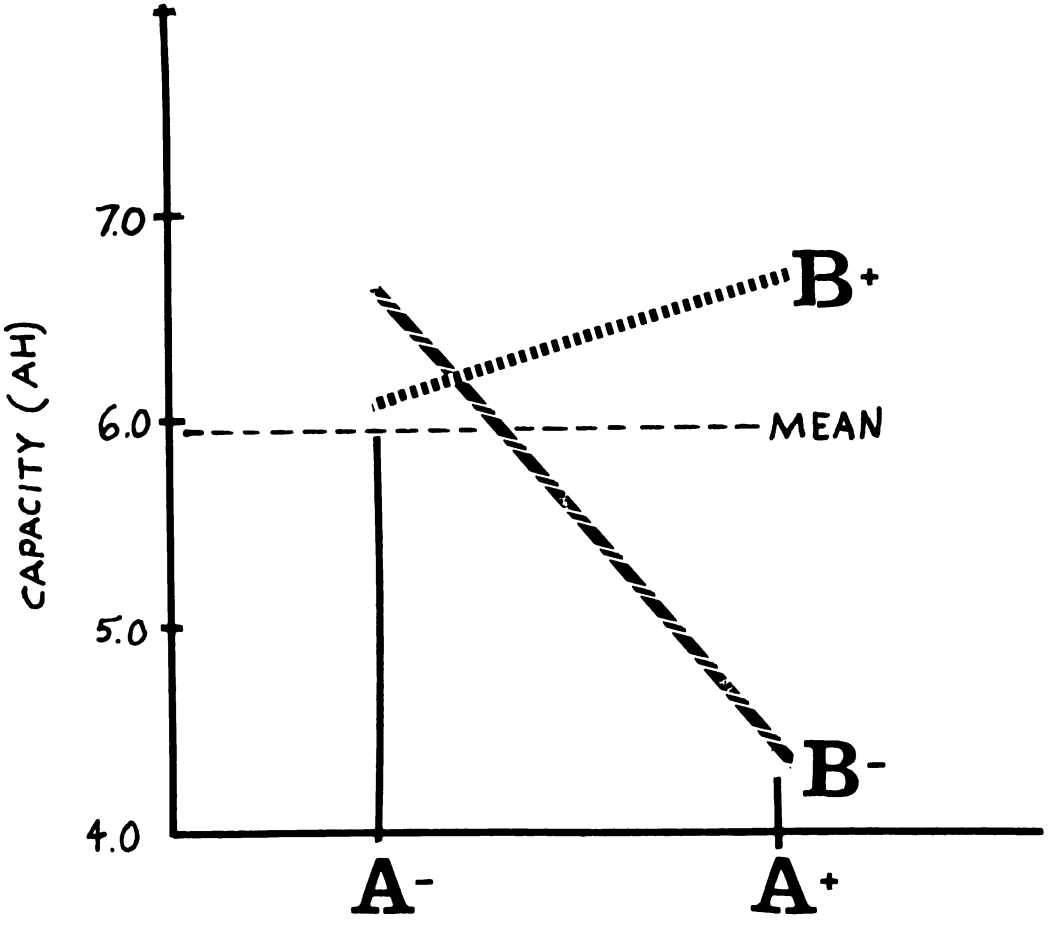
Intercept = 5.94167, Multiple correlation = 0.85642, Std. Error of Estimate = 0.83716, Std Error of Reg Coef = 0.2093

ANALYSIS OF VARIANCE FOR THE REGRESSION

Source of Variation	Degrees of Freedom	Sum of Squares	Mean Squares	F Value
Attributable to Regression	7	15.42778	2.20397	3.14478
Deviation from Regression	8	5.60667	0.70083	
Total	15	21.03444		

Figure II

Significant Calculated Mean Effects from Ni-Zn Experiment



DESIGN OF EXPERIMENTS AND A STATISTICAL PERFORMANCE MODEL FOR A RADAR ALTIMETER

Erwin Biser
Avionics Laboratory, U. S. Army Electronics Command
Fort Monmouth, New Jersey

GLOSSARY OF TERMS AND SYMBOLS.

n	Number of observations
$N(+d)$	Number of positive deviations
$N(-d)$	Number of negative deviations
H_m	Height observation measured by Honeywell Altimeter, Model 7091-A (Modified); Test Item 1
H_o	Height observation measured by AN/APN-22 Altimeter; Test Item 2
H_R	Reference height observation measured by RCA Laser Range-Finder AN/GVS-1 (XE-6)
d_{mR}	$= H_m - H_R$
d_{oR}	$= H_o - H_R$
$\bar{d}_{mR}$	Average deviation of the RCA Laser Range-Finder AN/GVS-1 (XE-6) height observations from the Honeywell Altimeter, Model 7091-A (Modified) height observations
$\bar{d}_{oR}$	Average deviation of the RCA Laser Range-Finder AN/GVS-1 (XE-6) height observations from the AN/APN-22 Altimeter height observations
θ_p	Angle of pitch measured by the attitude indicator from the vertical (90°) as established by the Honeywell Vertical Gyro; positive angles of pitch indicate the nose of the aircraft is up; negative angles of pitch indicate the nose of the aircraft is down.
θ_r	Angle of roll measured by the attitude indicator from the vertical (90°) as established by the Honeywell Vertical Gyro; positive angles of roll indicate the aircraft rolling to the right; negative angles of roll indicate the aircraft rolling to the left.

$\overline{\theta}_p$ Average deviation of the pitch angle observations from the vertical (90°) as established by the Honeywell Vertical Gyro

$\overline{\theta}_r$ Average deviation of the roll angle observations from the vertical (90°) as established by the Honeywell Vertical Gyro

$s_{(\theta_p)}$ Sample standard error of the deviations of the angle of pitch observations from the vertical (90°) as established by the Honeywell Vertical Gyro

$s_{(\theta_r)}$ Sample standard error of the deviations of the angle of roll observations from the vertical (90°) as established by the Honeywell Vertical Gyro

$s_{(d_{mR})}$ Sample standard error of the deviations of the RCA Laser Range-Finder measured observations from the 7091-A Altimeter (Modified) measured observations

$s_{(d_{oR})}$ Sample standard error of the deviations of the RCA Laser Range-Finder measured observations from the AN/APN-22 Altimeter measured observations

$N_T(d)$ Total number of positive and negative deviations

REFERENCES

Hart, B. I., "Significance Levels for the Ratio of the Mean Square Successive Difference to the Variance;" Annals of Mathematical Statistics, Vol XIII, 1942, pp. 445-447.

von Neumann, J., "Distribution of the Ratio of the Mean Square Successive Difference to the Variance;" Annals of Mathematical Statistics, Vol XII, 1942, pp. 367-395.

ACKNOWLEDGMENTS. The author gratefully acknowledges his indebtedness to Mr. Raymond Cruickshank, Multifunction Sensor Team Leader of the Enviornment Sensing Technical Area, and the Project Engineer for this task. His collaboration and advice given in the process of designing the experiment is fully appreciated. The tests were carried out under his supervision, and with his participation.

Thanks are due to Mr. Edward Sender of the Operations Research Office for his aid in this effort and for the computations performed to obtain significant results.

Many thanks to my secretary, Mrs. Christine Snyder, for her aid in the preparation of this report.

BACKGROUND. Experiments conducted in 1963-4 in the Arctic region by the U. S. Army Electronics Command Avionics Laboratory have confirmed that the electrical properties of polar ice and snow do, in fact, cause microwave frequencies to suffer high surface reflection losses, and low transmission losses within the medium. These results correlate well with the theoretical predictions. Specifically, these experiments revealed that, for standard 4.3GHz radar altimeter frequencies, normally incident electromagnetic waves impinging upon essentially uncontaminated snow surfaces effectively penetrate the snow/ice media to depths of several hundred feet. In many instances, sub-surface interfaces provide signal reflections which are of substantial amplitude and are readily detectable at the radar altimeter receiver. These sub-surface reflections can be as much as 20 db stronger than the surface reflections.

Findings further showed that radar altimeters employing nanosecond pulse leading-edge-tracking techniques are significantly more accurate than those utilizing frequency modulation. The accuracies of these techniques differed greatly because the pulse system measured altitude from the closest terrain surface (the leading edge of the reflected RF pulse) whereas the FM-CW system integrated all the surface and sub-surface signal returns, with no discrimination against the more distant radar echoes.

Specifically, CW altimeter errors as great as 150 feet were recorded for an actual altitude of 300 feet. Pulse altimeter errors were considered negligible; in fact, they were not measurable since they did not exceed the instrumentation error inherent in the experiment.

In April 1967, Research and Development personnel of the U. S. Army Electronics Command Avionics Laboratory conducted radar altimeter tests in a Choctaw CH-34C helicopter over the three-story high rain forests of the Panama Canal Zone. These tests, the first of their kind, were made possible through the use of specially designed instrumentation including an air-portable range finder with a height-measuring accuracy of one meter.

During the tests, altimeter data was continuously gathered and recorded while the project aircraft was flown 1000 miles over dense jungle. The project personnel previously conducted experiments in Greenland, which first showed the unique potential of this nanosecond pulse radar, with its leading-edge-tracking technique, for providing accurate height measurements over deep ice and snow of the Arctic region.

While most radar altimeters provide relatively accurate height information over large, flat airstrips they typically become highly unreliable and grossly in error when employed over varying terrain such as deep polar ice and snow or high jungle foliage.

In addition to providing radar altimeter performance data not heretofore obtained, the technical information resulting from these tests proved a significant factor in selection of radar altimeter design techniques most suitable for Army aircraft applications. More recently, procurement of the AN/APN-171 Radar Altimeter, employing this recommended design concept, has been initiated for Mohawk aircraft applications.

1. INTRODUCTION. Through the support of USATTC (USA Tropic Test Center), Ft. Clayton, Canal Zone, including the Army Aviation Detachment at Albrook AFB, radar altimeter tests were conducted over the high jungle canopy in the vicinity of Rio Chagres and Rio Pina, with three flights on April 4, 5, and 7. The tests were conducted in accordance with the procedure and objective as stated in the test plan prepared by the Office of Operations Research entitled: Design of Experiments for Radar Altimeter Techniques at the Tropical Test Center, Panama Canal Zone. Data was obtained from 16 hours of flight time at altitudes of 600 feet and 1000 feet, using a CH-34C helicopter bearing tail #34508.

Altogether, approximately 500 bits of data were obtained, each representing a comparison of the indicated attitude of one of the test item radar altimeters with actual aircraft height measured through use of a precise laser distance measuring equipment with a 1-meter accuracy.

2. SUMMARY AND CONCLUSIONS. It is to be understood that these conclusions are primarily statistical in character, and hence, are (statistical) inferences drawn from the evidence based solely on the observations. Furthermore, the observations in this analysis are deviations of measurements from the reference measurements of altitude, pitch, and roll.

The following conclusions emerge from this analysis:

2.1 The 7091-A Altimeter observations were predominantly negative and, hence, the readings were consistently less than the respective RCA Laser reference readings on all flights (see Table 2).

2.2 The AN/APN-22 Altimeter readings were predominantly positive and, hence, the readings were consistently greater than those of the respective RCA Laser reference readings with the exception of readings taken at a height of 600 feet and a velocity of 70 knots (see Table 2).

The following plausible explanation for conclusions 2.1 and 2.2 is offered by the Project Engineer: It appears that the narrow (one milli-radian) laser beam penetrated some appreciable distance through openings in the rain forest canopy before striking the uppermost foliage layer.

2.3 The analysis of variance technique shows that the population means of the test items, namely 7091-A and AN/APN-22 Altimeters, are significantly different at a level of significance of .01 on all flights for which data were obtained.

2.4 The standard deviation of pitch angle observations taken during the use of the AN/APN-22 Altimeter were consistently less than the respective standard deviations of pitch angle observations taken during use of the 7091-A Altimeter (see Table 3).

2.5 For the combined positive and negative observations at a at a height of 1000 feet at both velocities, the absolute magnitudes of the means and the standard deviations of the 7091-A observations are consistently less than the respective means (absolute magnitudes) and standard deviations of the AN/APN-22 observations. The opposite results are obtained at a height of 600 feet and at a velocity of 70 knots (see Table 1).

ALTIMETER / FACTORS			7091-A			AN/APN-22		
Height (feet)	Flt #	Velocity (kts±5)	Deviations (feet)			Deviations (feet)		
			$N_T(d)$	$\bar{d}_{mR}$	$s_{(d_{mR})}$	$N_T(d)$	$\bar{d}_{oR}$	$s_{(d_{oR})}$
600	1	55	59	-27.20	23.65	(SEE NOTE BELOW)		
	3	70	29	-33.75	29.56	70	.29	28.55
1000	2	55	91	-34.45	28.42	91	37.09	37.17
	3	70	52	-29.18	27.44	58	61.45	37.10

MEANS AND STANDARD DEVIATIONS OF ALL (both positive and negative) MEASUREMENTS FOR EACH TEST ITEM

Velocity was not treated as a factor.

TABLE 1

NOTE: No experimentation was performed with the AN/APN-22 Altimeter at a height of 600 feet and a velocity of (55±5) knots

ALTIMETER / FACTORS			7091-A						AN/APN-22					
			Deviations (Feet)						Deviations (Feet)					
			Positive			Negative			Positive			Negative		
Height (feet)	Flt #	Velocity (kts±5)	N(+d)	$\bar{d}_{mR}$	$s_{(d_{mR})}$	N(-d)	$\bar{d}_{mR}$	$s_{(d_{mR})}$	N(+d)	$\bar{d}_{oR}$	$s_{(d_{oR})}$	N(-d)	$\bar{d}_{oR}$	$s_{(d_{oR})}$
600	1	55	8	4.03	3.18	51	-32.10	21.39		(See Note Below)				
	3	70	4	11.30	10.96	25	-40.96	24.15	35	22.61	16.82	35	-22.02	18.10
1000	2	55	12	8.42	6.80	79	-40.96	24.35	78	46.73	28.01	13	-20.78	13.48
	3	70	7	10.44	6.83	45	-35.34	24.01	53	67.96	29.52	5	- 7.70	8.29

MEANS AND STANDARD DEVIATIONS
OF ALTIMETER MEASUREMENTS

TABLE 2

Velocity was not treated as a factor.

NOTE: No experimentation was performed with the AN/APN-22 Altimeter at a height of 600 feet and a velocity of (55±5) knots.

Height (feet)	Flight #	Velocity (kts±5)	Deviations (Degrees) Pitch Angle			Roll Angle			Deviations (Degrees) Pitch Angle			Roll Angle		
			$N_T(d)$	$\bar{\theta}_p$	$s_{(\theta_p)}$	$N_T(d)$	$\bar{\theta}_r$	$s_{(\theta_r)}$	$N_T(d)$	$\bar{\theta}_p$	$s_{(\theta_p)}$	$N_T(d)$	$\bar{\theta}_r$	$s_{(\theta_r)}$
600	1	55	63	-.22	.86	63	.22	.95		(See Note Below)				
	3	70	33	.15	1.62	33	.58	1.28	81	.31	1.43	81	.37	1.73
1000	2	55	100	.01	1.12	100	.16	1.30	97	.22	1.03	97	.14	1.22
	3	70	58	.61	1.12	58	.55	1.07	60	1.24	.98	60	.57	1.29

MEANS AND STANDARD DEVIATIONS OF
PITCH AND ROLL MEASUREMENTS

TABLE 3

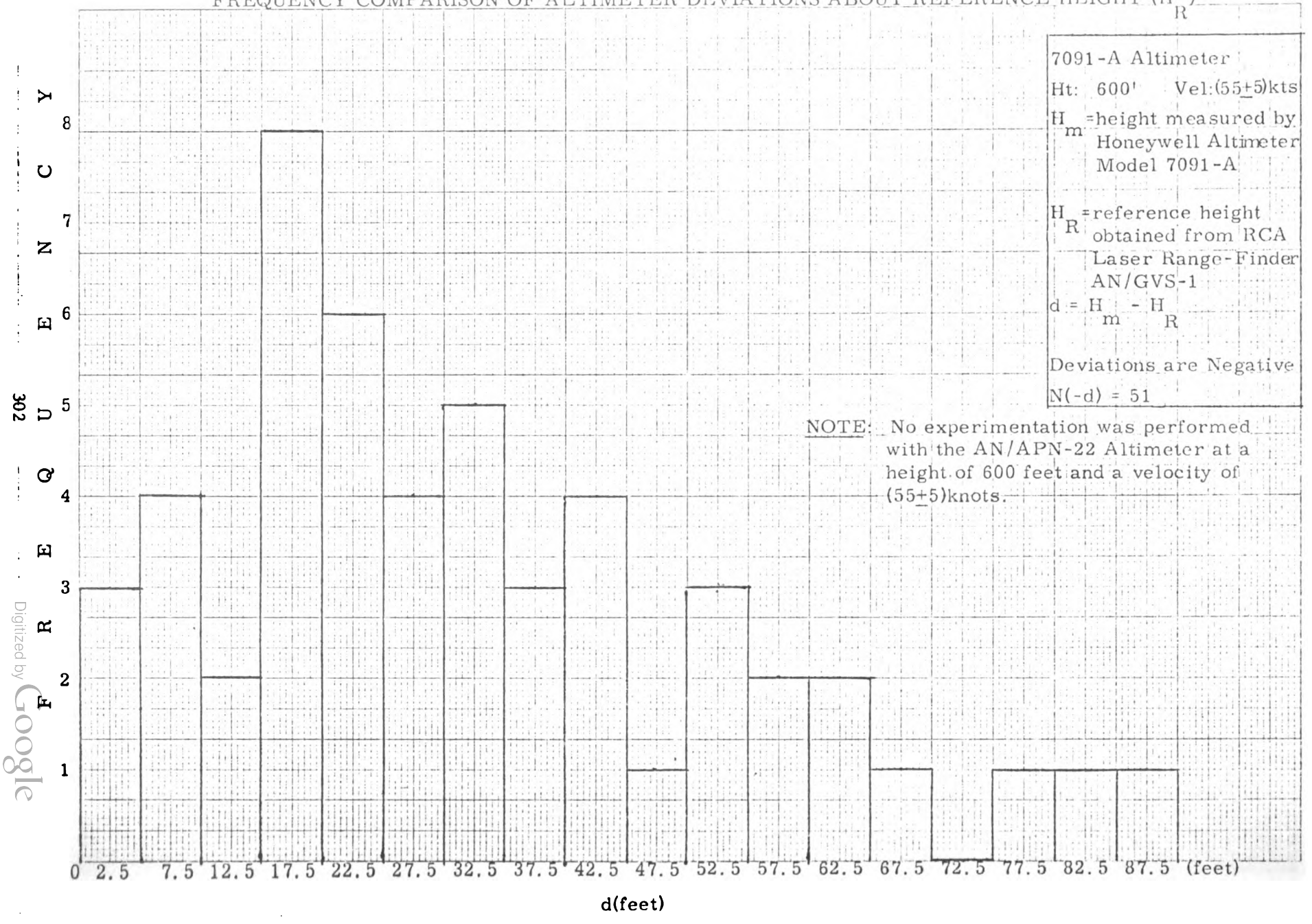
Velocity was not treated as a factor.

NOTE: No experimentation was performed with the AN/APN-22 Altimeter at a height of 600 feet and a velocity of (55±5) knots.

GRAPHS of OBSERVATIONS

of the

7091-A and AN/APN-22 ALTIMETERS



FREQUENCY COMPARISON OF ALTIMETER DEVIATIONS ABOUT REFERENCE HEIGHT (H_R)

7091-A Altimeter	APN-22 Altimeter
Ht: 600' Vel: (70±5)kts	Ht: 600' Vel: (70±5)kts
H_m = height measured by Honeywell Altimeter Model 7091-A	H_o = height measured by AN/APN-22 Altimeter
H_R = reference height obtained from RCA Laser Range-Finder AN/GVS-1	H_R = reference height obtained from RCA Laser Range-Finder AN/GVS-1
Deviations are Negative	Deviations are Positive
N(-d) = 25	N(+d) = 35
$d = H_m - H_R$	$d = H_o - H_R$

7091-A

APN-22

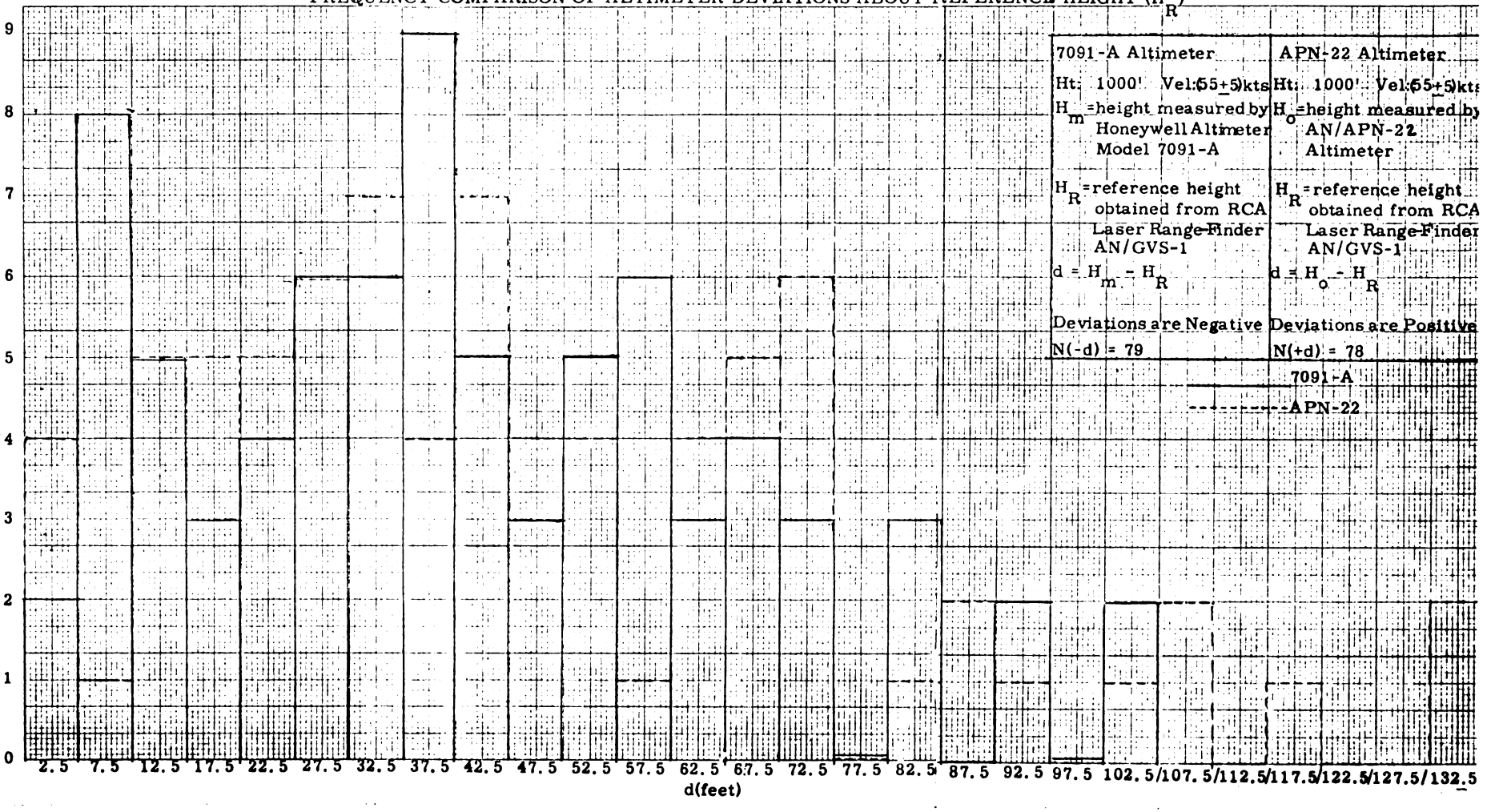
FREQUENCY

0 1 2 3 4 5 6 7 8

2.5 7.5 12.5 17.5 22.5 27.5 32.5 37.5 42.5 47.5 52.5 57.5 62.5 67.5 72.5 77.5 82.5 87.5 (feet)

d(feet)

FREQUENCY COMPARISON OF ALTIMETER DEVIATIONS ABOUT REFERENCE HEIGHT (H_R)



FREQUENCY COMPARISON OF ALTIMETER DEVIATIONS ABOUT REFERENCE HEIGHT (H_R)

7091-A Altimeter	APN-22 Altimeter
Ht: 1000' Vel(70±5)kts	Ht: 1000' Vel(70±5)kts
H_m = height measured by Honeywell Altimeter Model 7091-A	H_o = height measured by AN/APN-22 Altimeter
H_R = reference height obtained from RCA Laser Range-Finder AN/GVS-1	H_R = reference height obtained from RCA Laser Range-Finder AN/GVS-1
$d = H_m - H_R$	$d = H_o - H_R$
Deviations are Negative	Deviations are Positive
N(-d) = 45	N(+d) = 53

FREQUENCY OF OCCURRENCE IN 305 SAMPLES

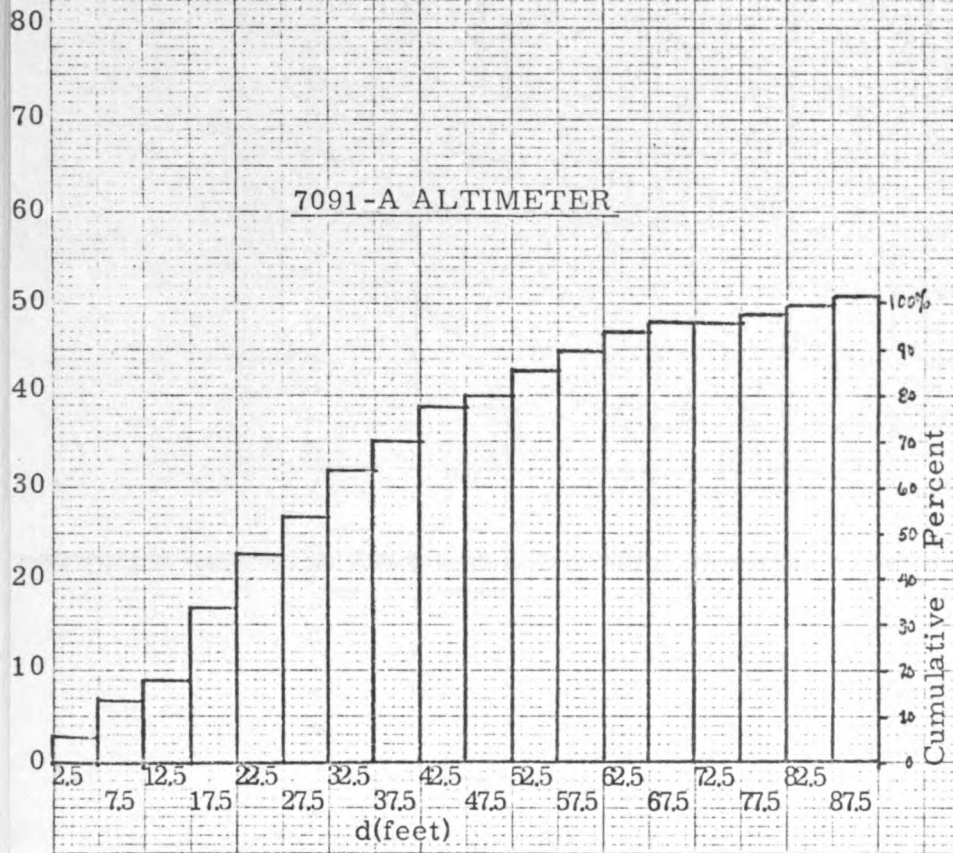
0 1 2 3 4 5 6 7 8

.25 .75 2.5 7.5 12.5 17.5 22.5 27.5 32.5 37.5 42.5 47.5 52.5 57.5 62.5 67.5 72.5 77.5 82.5 87.5 92.5 97.5 102.5 107.5 112.5 117.5 122.5 127.5 132.5

d(feet)

7091-A
APN-22

CUMULATIVE FREQUENCY COMPARISON OF ALTIMETER DEVIATIONS:



Ht: 600' Vel:(55±5)kts

H_m = height measured by Honeywell Altimeter 7091-A

H_R = reference height obtained from RCA LaserRange-Finder AN/GVS-1

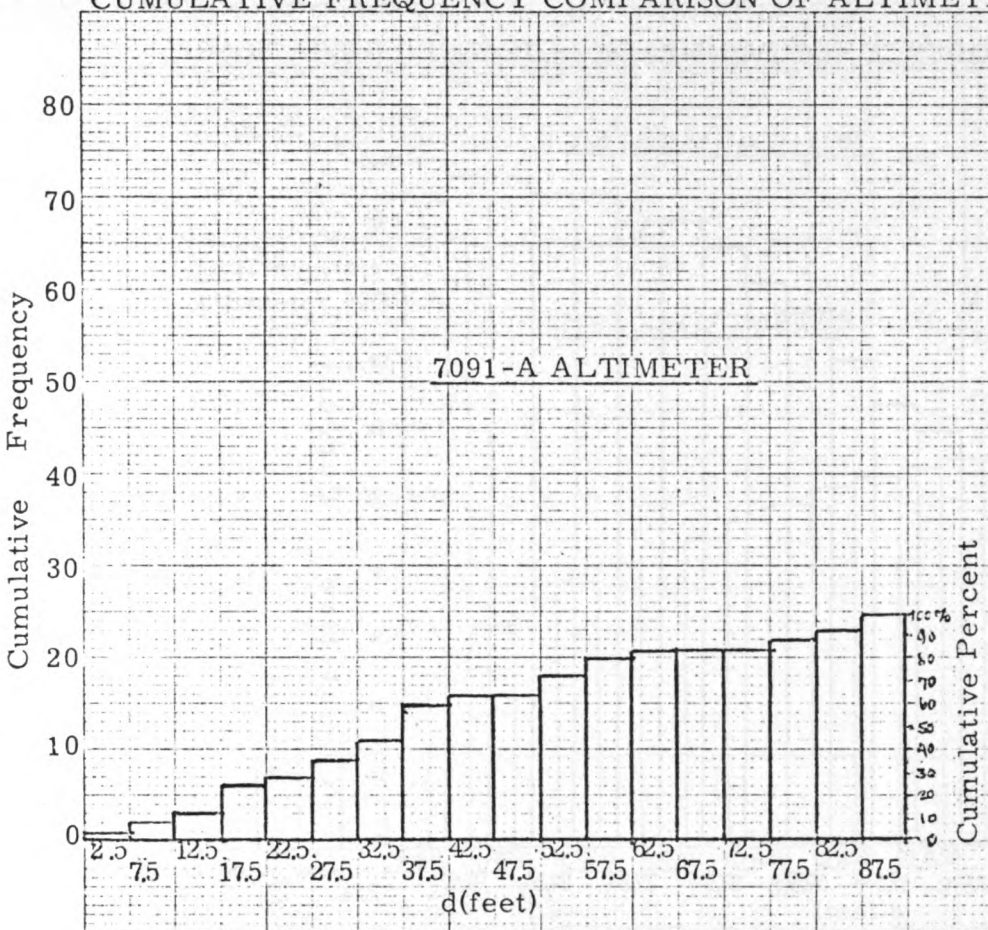
$d = H_m - H_R$

Deviations are Negative

N(-d) = 51

NOTE: No experimentation was performed with the AN/APN-22 Altimeter at a height of 600 feet and a velocity of (55±5) knots.

CUMULATIVE FREQUENCY COMPARISON OF ALTIMETER DEVIATIONS:



Ht: 600' Vel:(70±5)kts

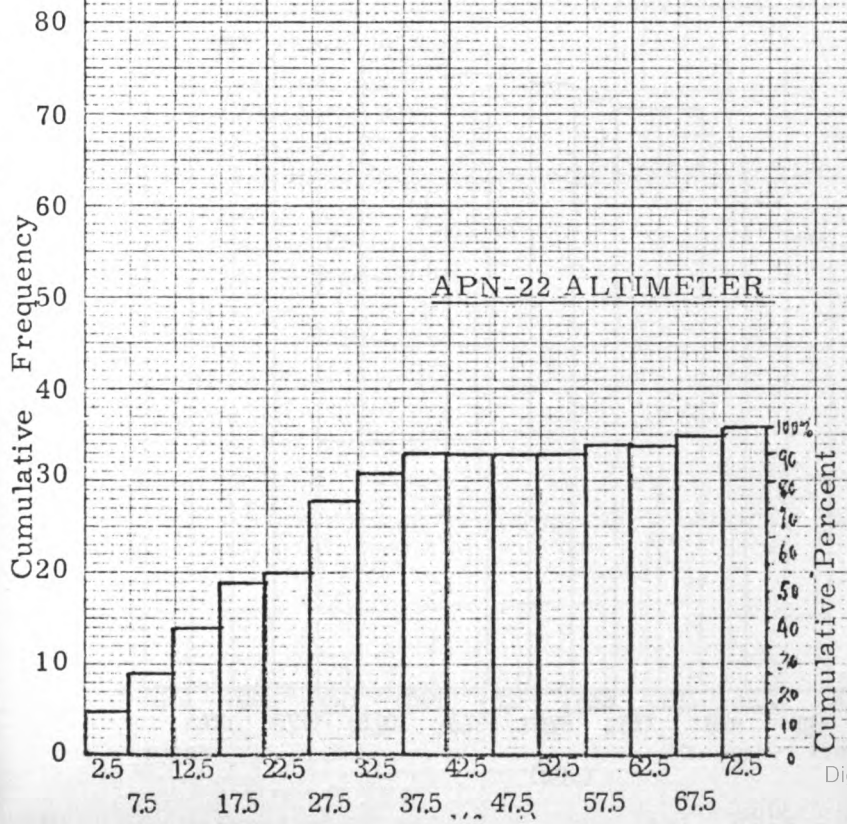
H_m = height measured by Honeywell Altimeter 7091-A

H_R = reference height obtained from RCA LaserRange-Finder AN/GVS-1

$d = H_m - H_R$

Deviations are Negative

N(-d) = 25



Ht: 600' Vel:(70±5)kts

H_R = reference height obtained from RCA LaserRange-Finder AN/GVS-1

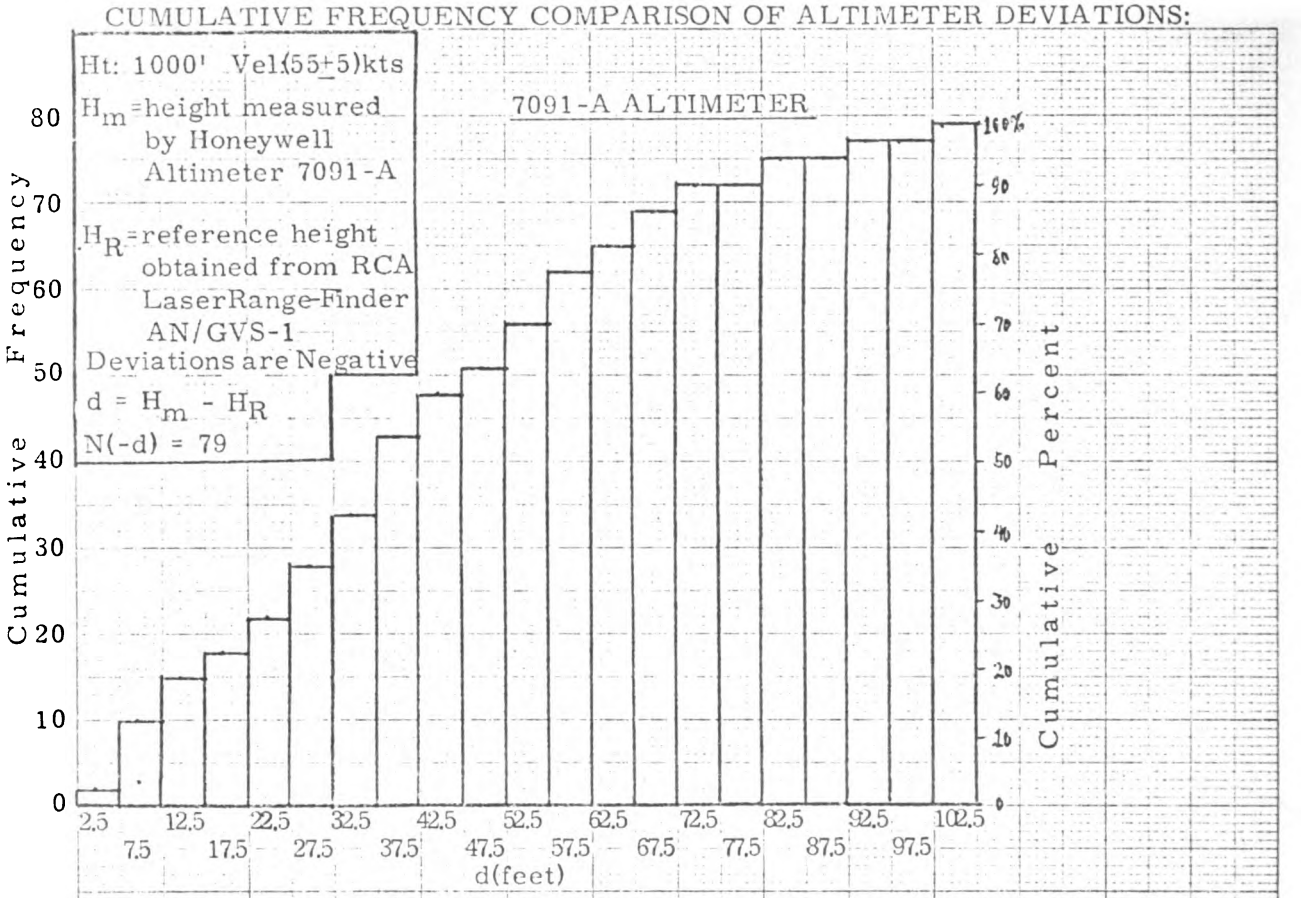
H_o = height measured by AN/APN-22 Altimeter

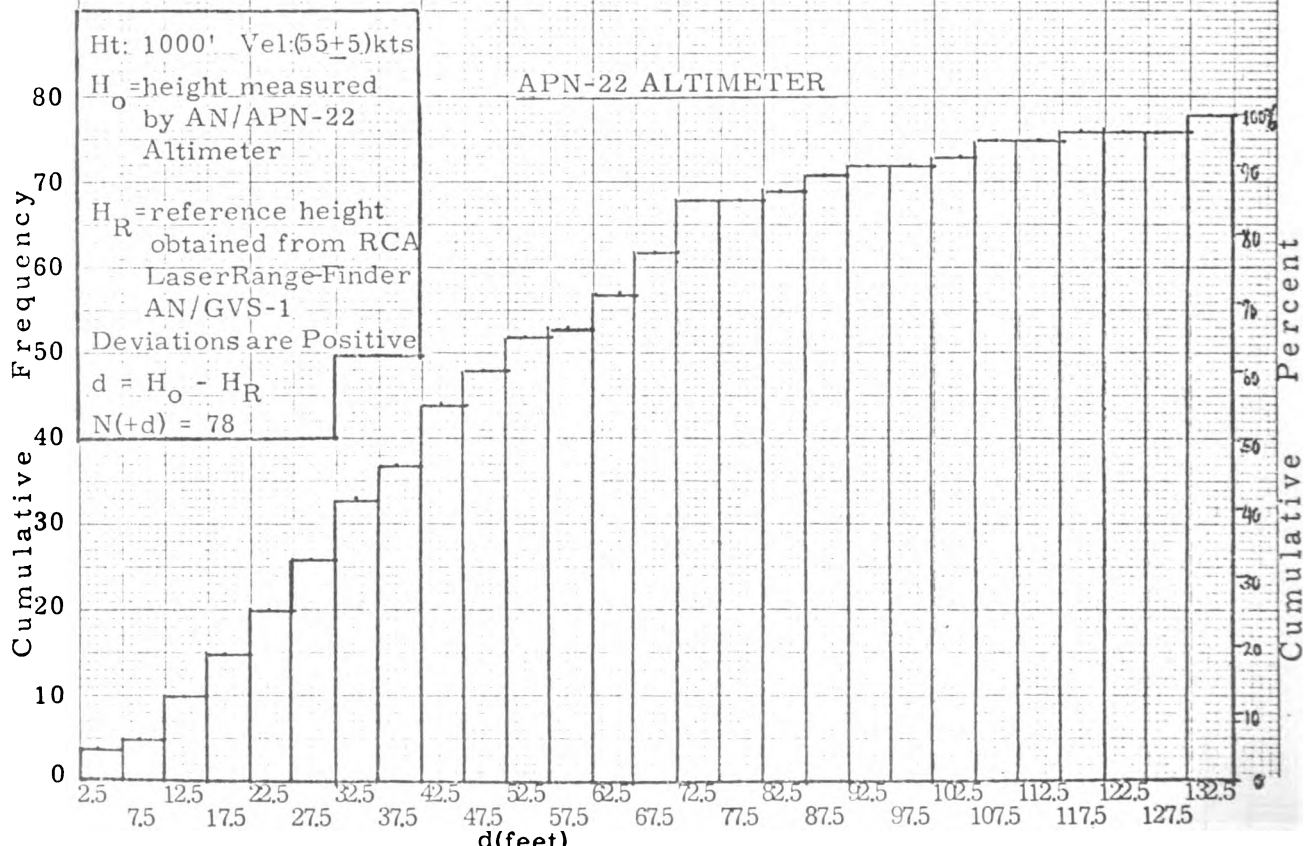
$d = H_o - H_R$

Deviations are Positive

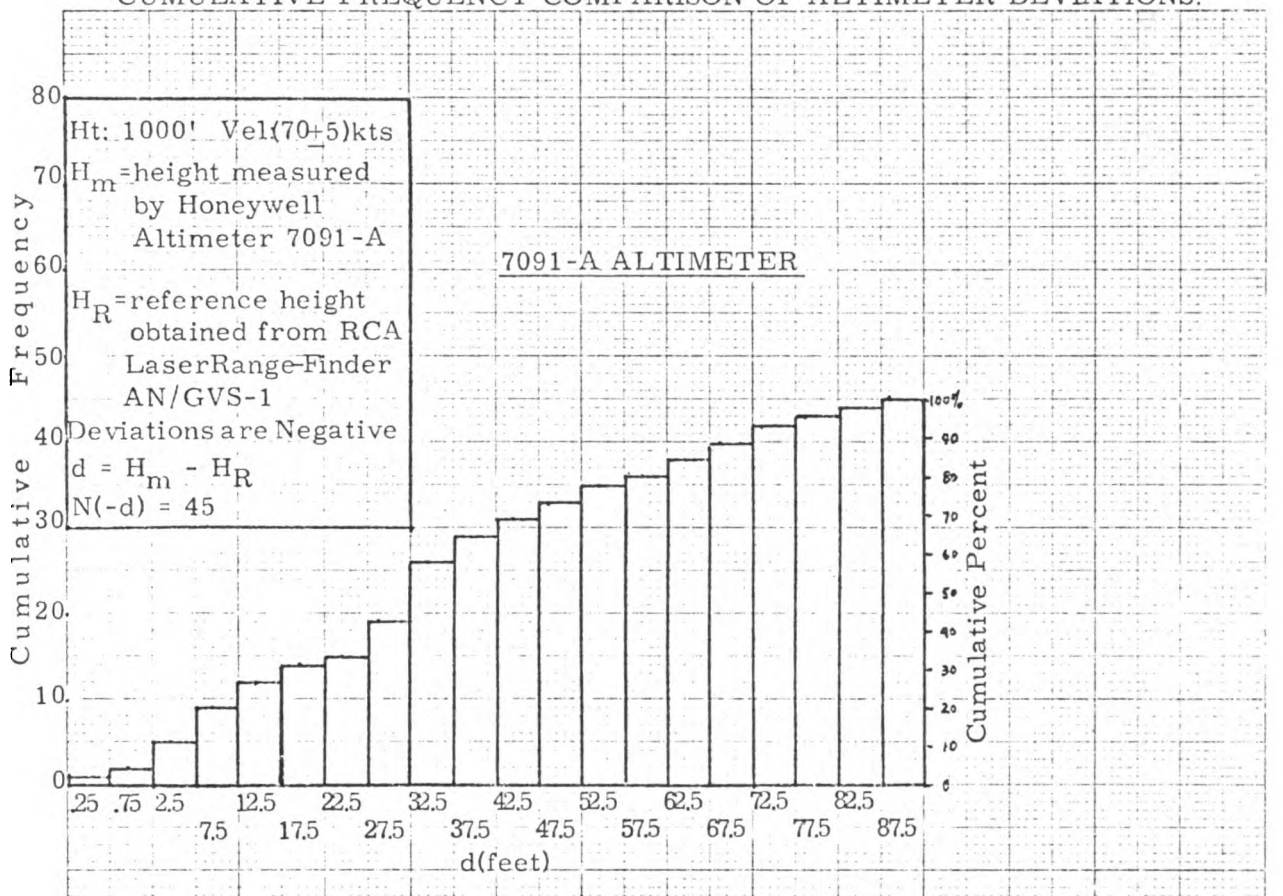
N(+d) = 35

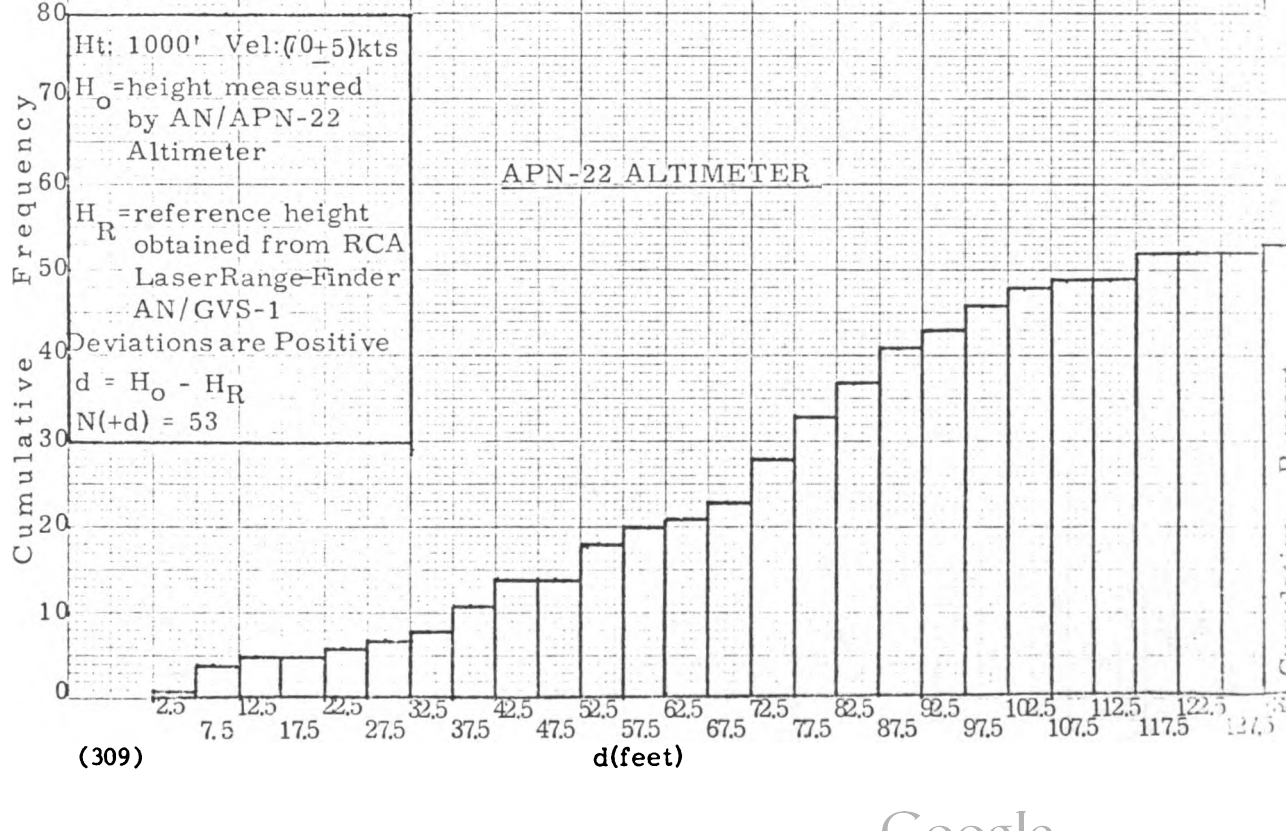
CUMULATIVE FREQUENCY COMPARISON OF ALTIMETER DEVIATIONS:

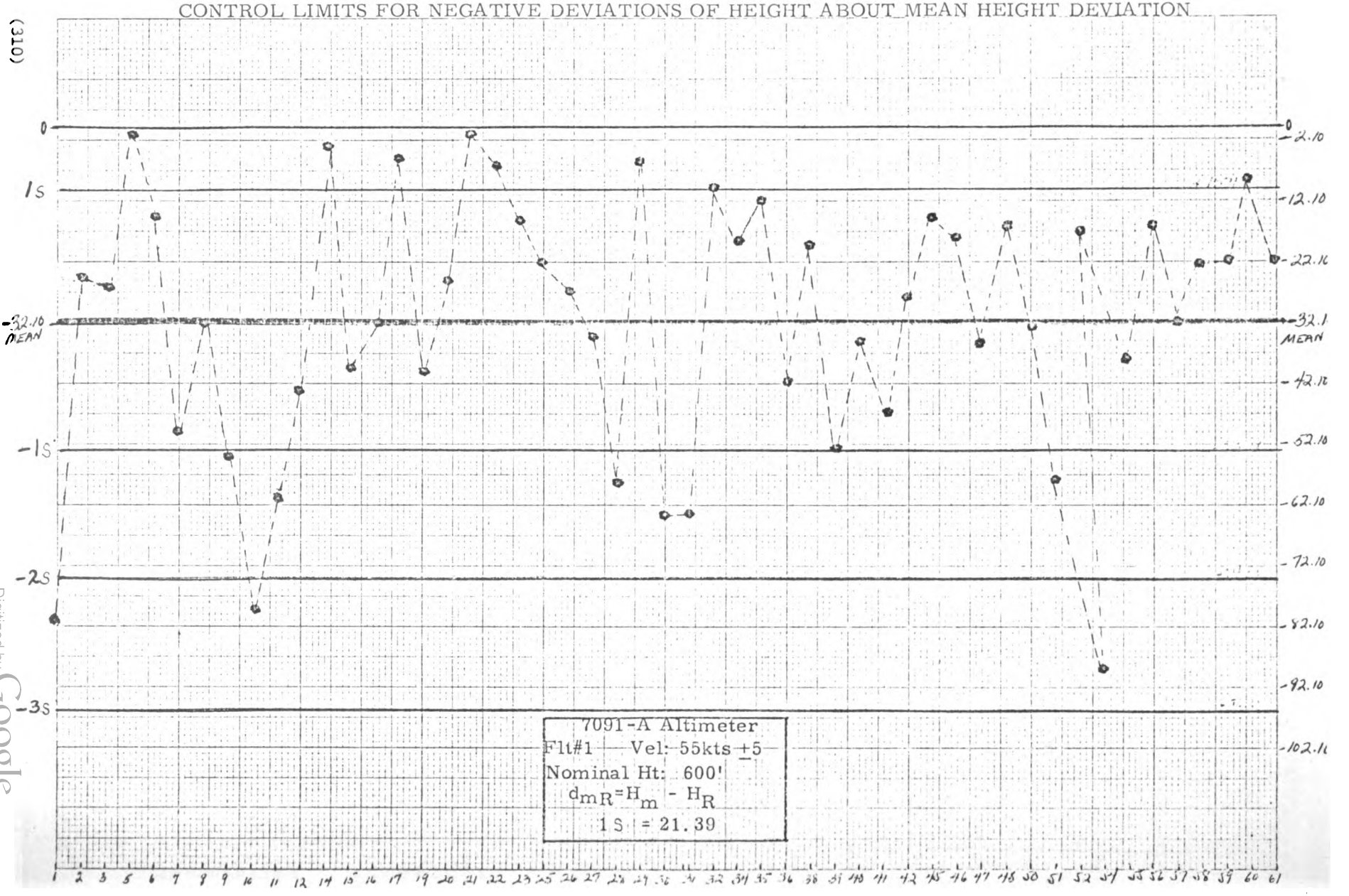


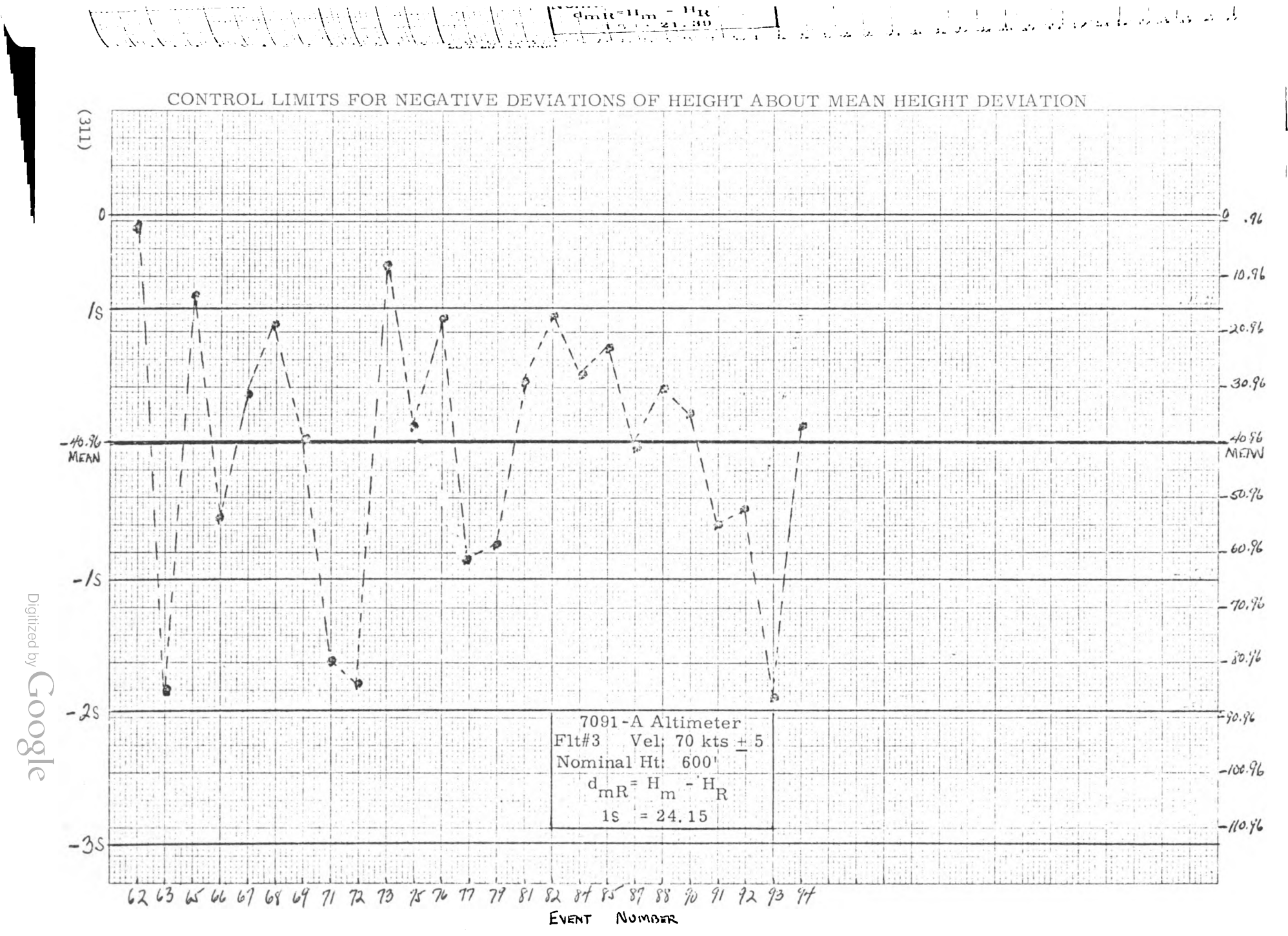


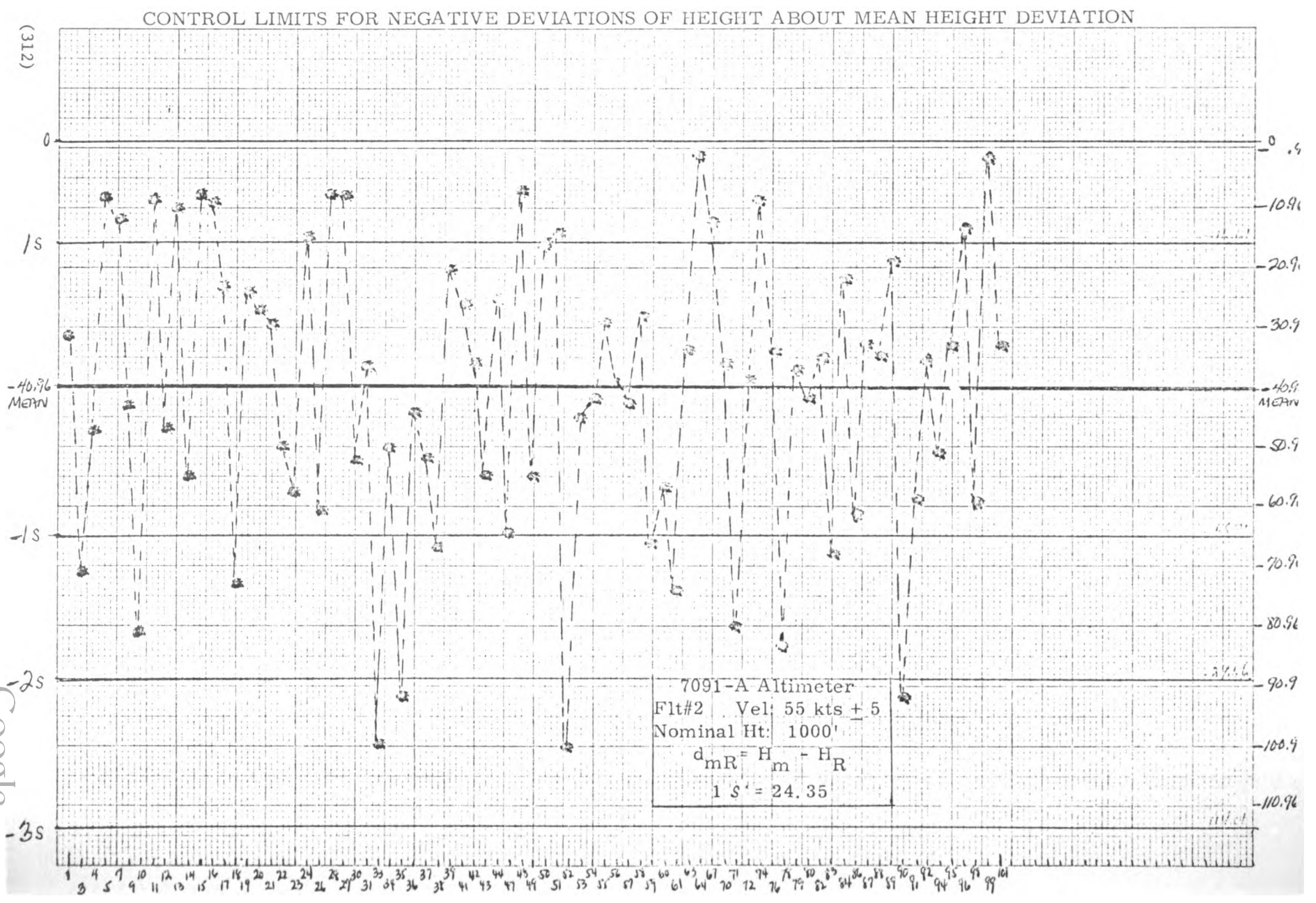
CUMULATIVE FREQUENCY COMPARISON OF ALTIMETER DEVIATIONS:

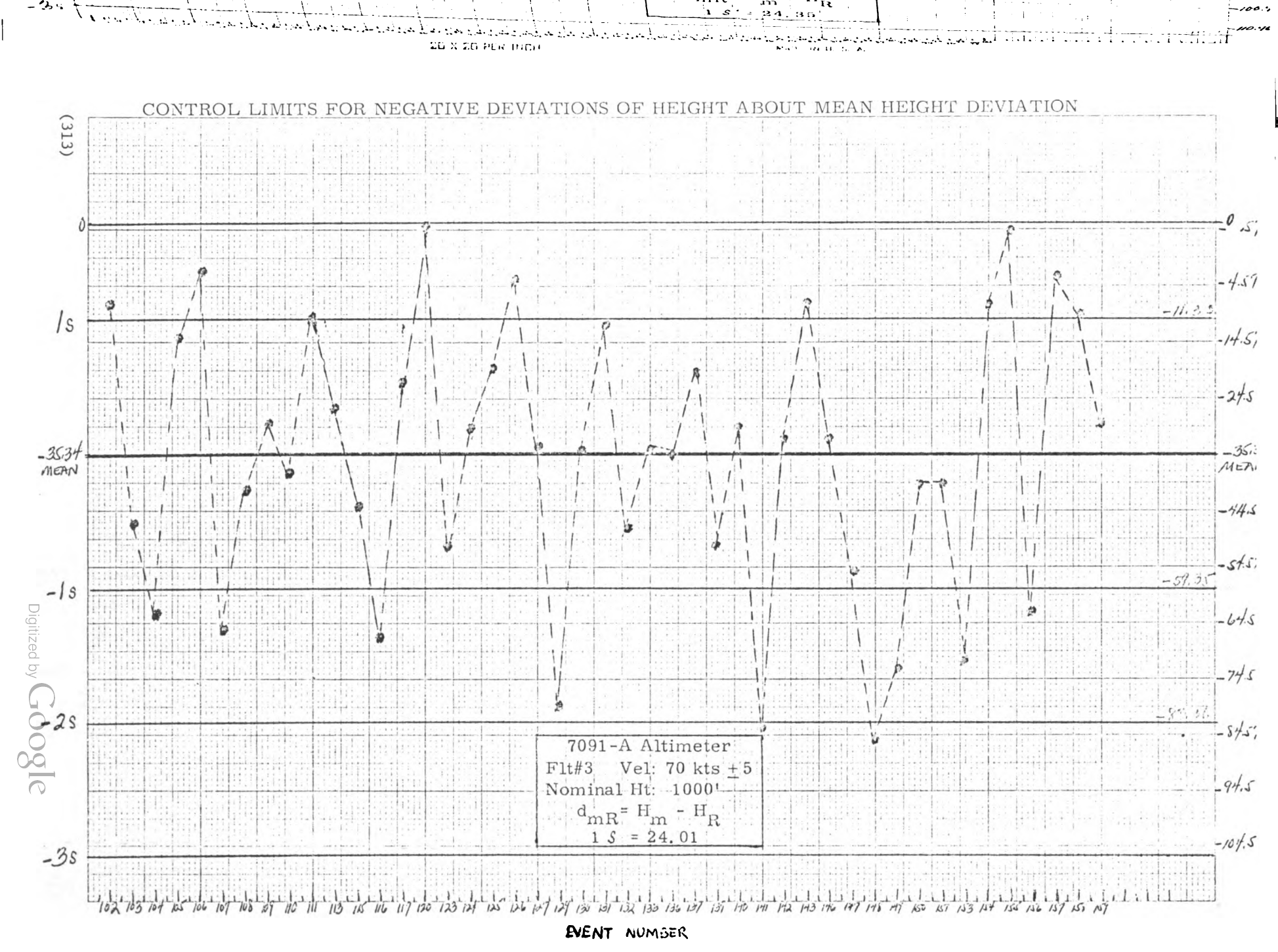


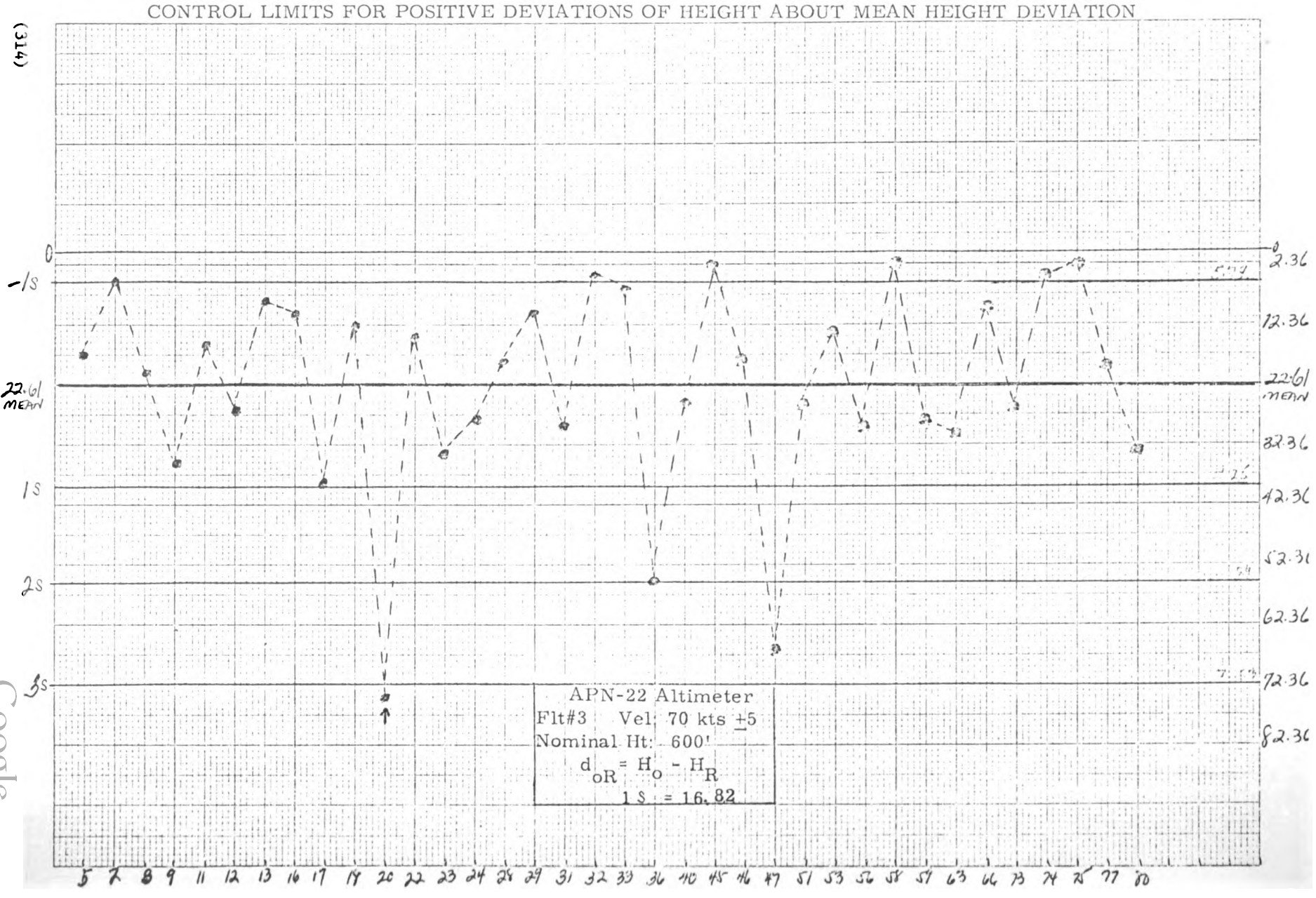


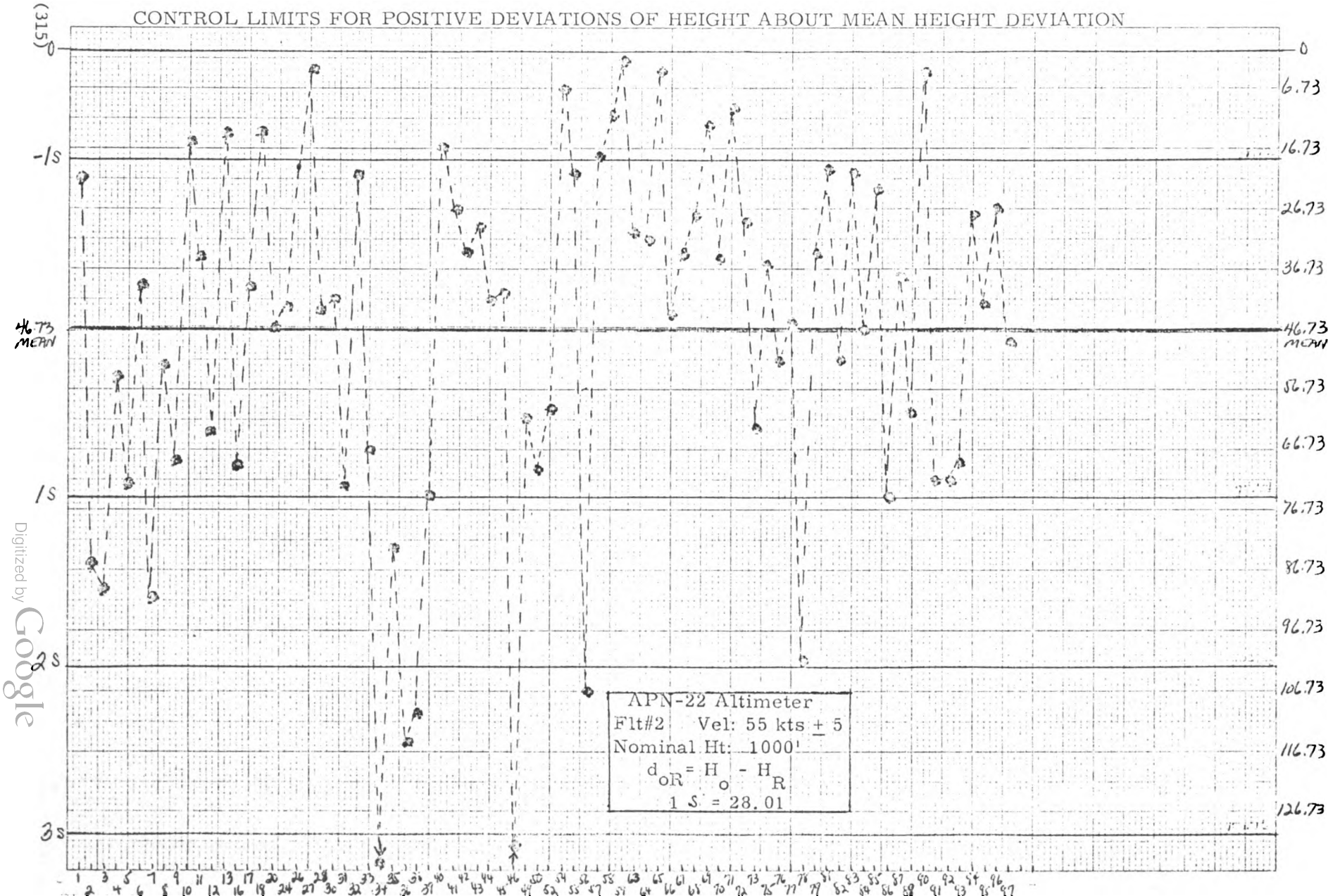


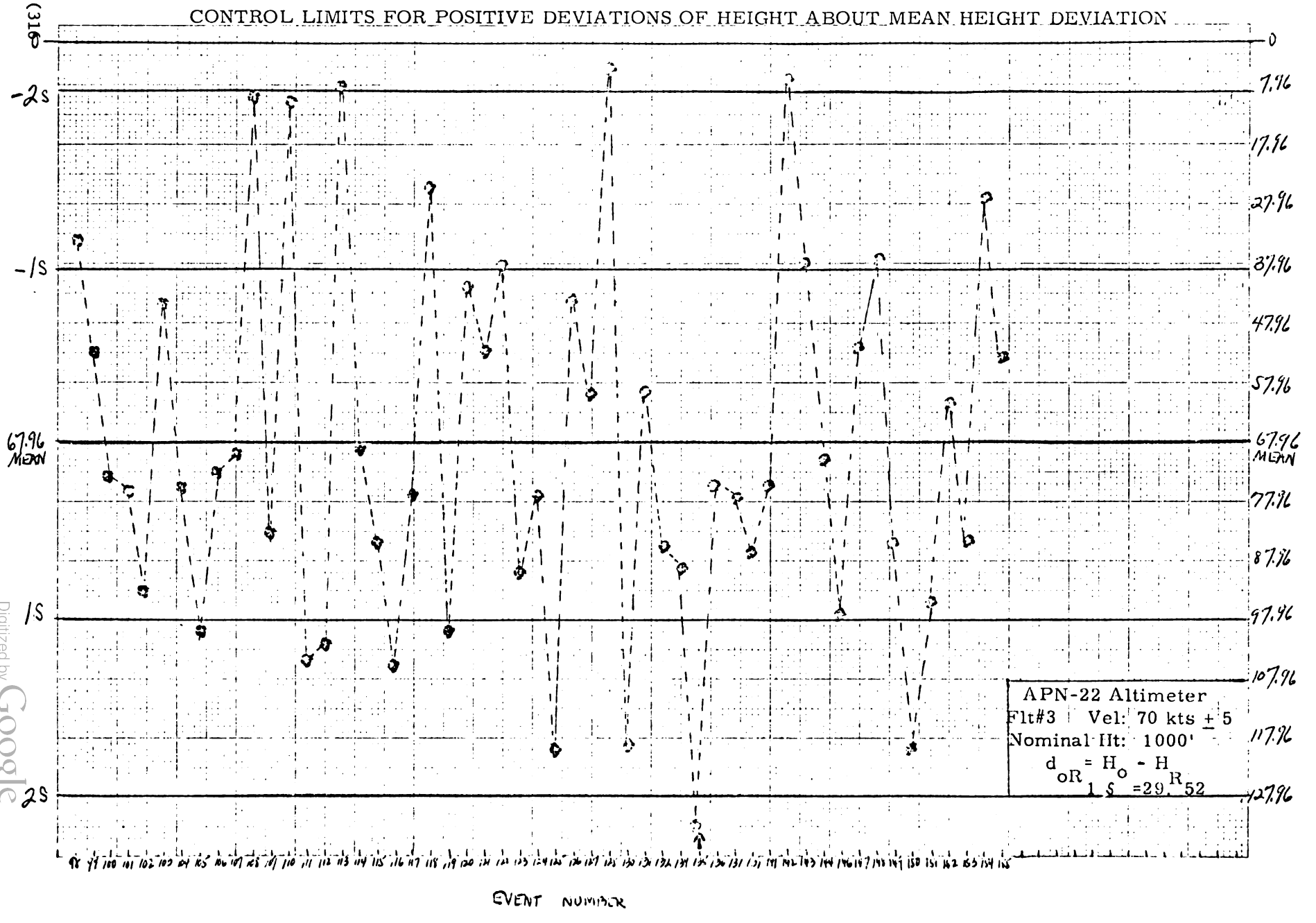


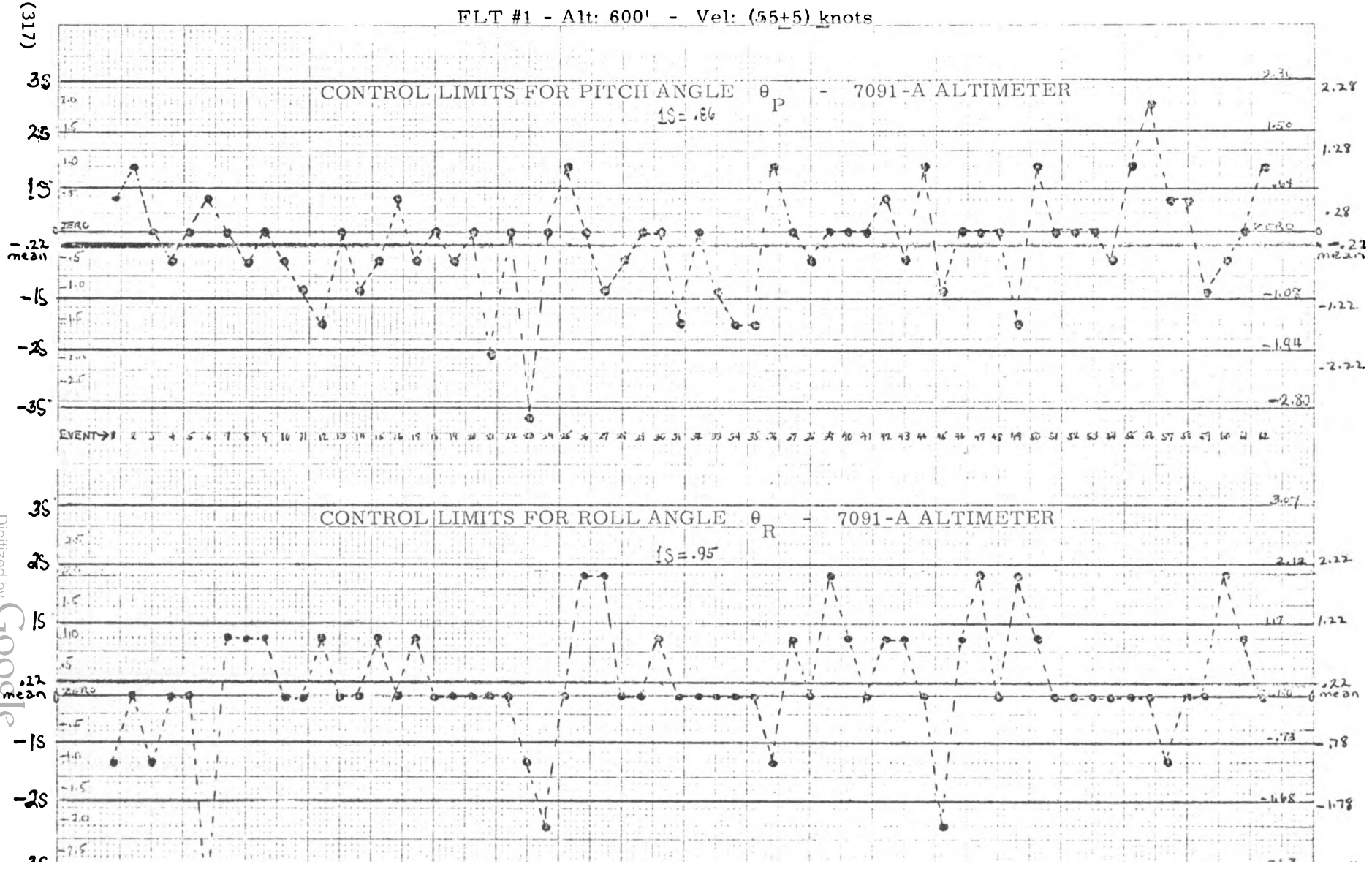


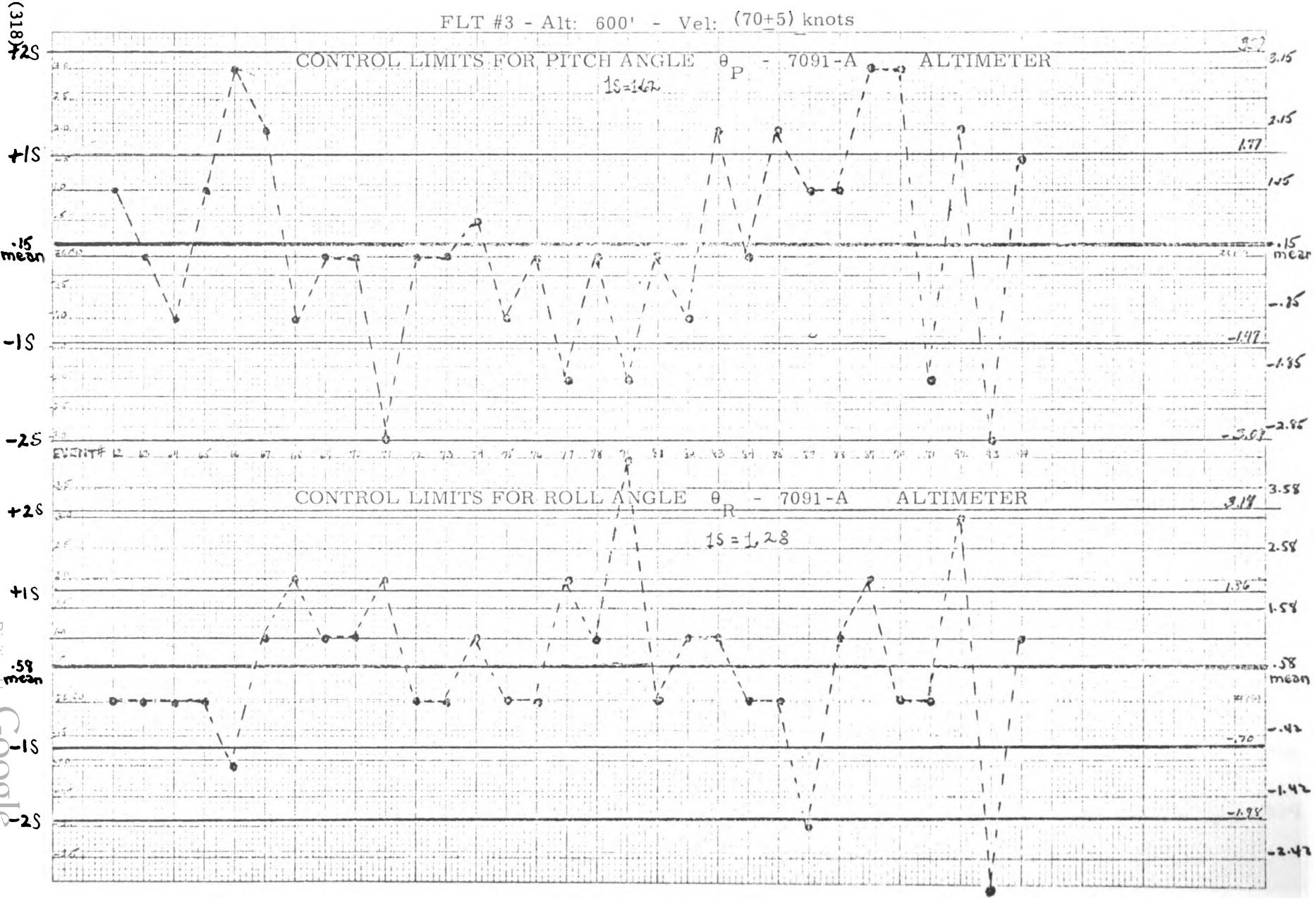

CONTROL LIMITS FOR NEGATIVE DEVIATIONS OF HEIGHT ABOUT MEAN HEIGHT DEVIATION
(313)
0
1S
-3534
MEAN
-1S
-2S
-3S
-0 S;
-4.5;
-11.2 S
-14.5;
-24.5
-35.3
MEAN
-44.5
-54.5;
-59.35
-64.5
-74.5
-84.5;
-94.5
-104.5
7091-A Altimeter
Flt#3 Vel: 70 kts ±5
Nominal Ht: 1000'
$d_{mR} = H_m - H_R$
1 S = 24.01
EVENT NUMBER

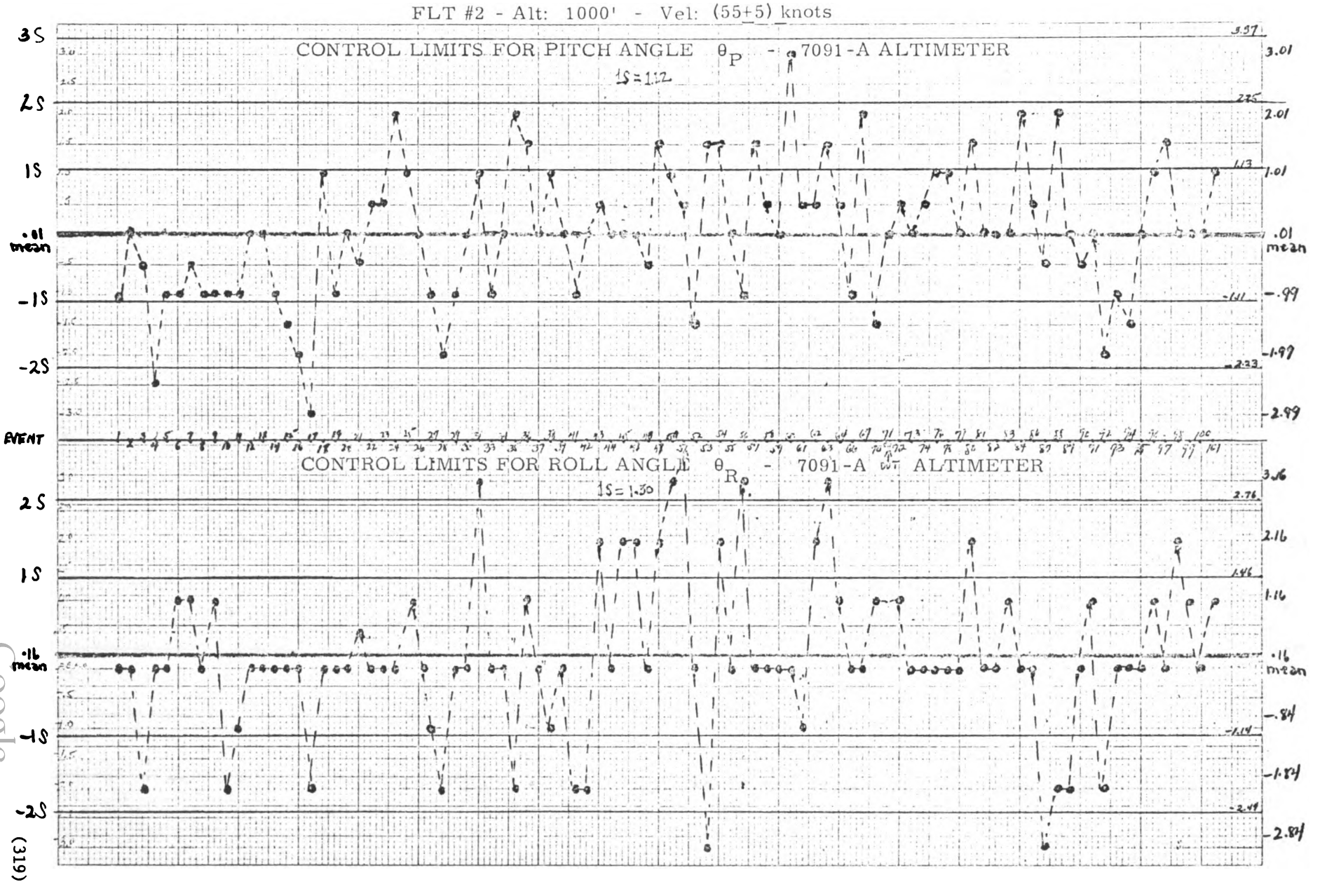




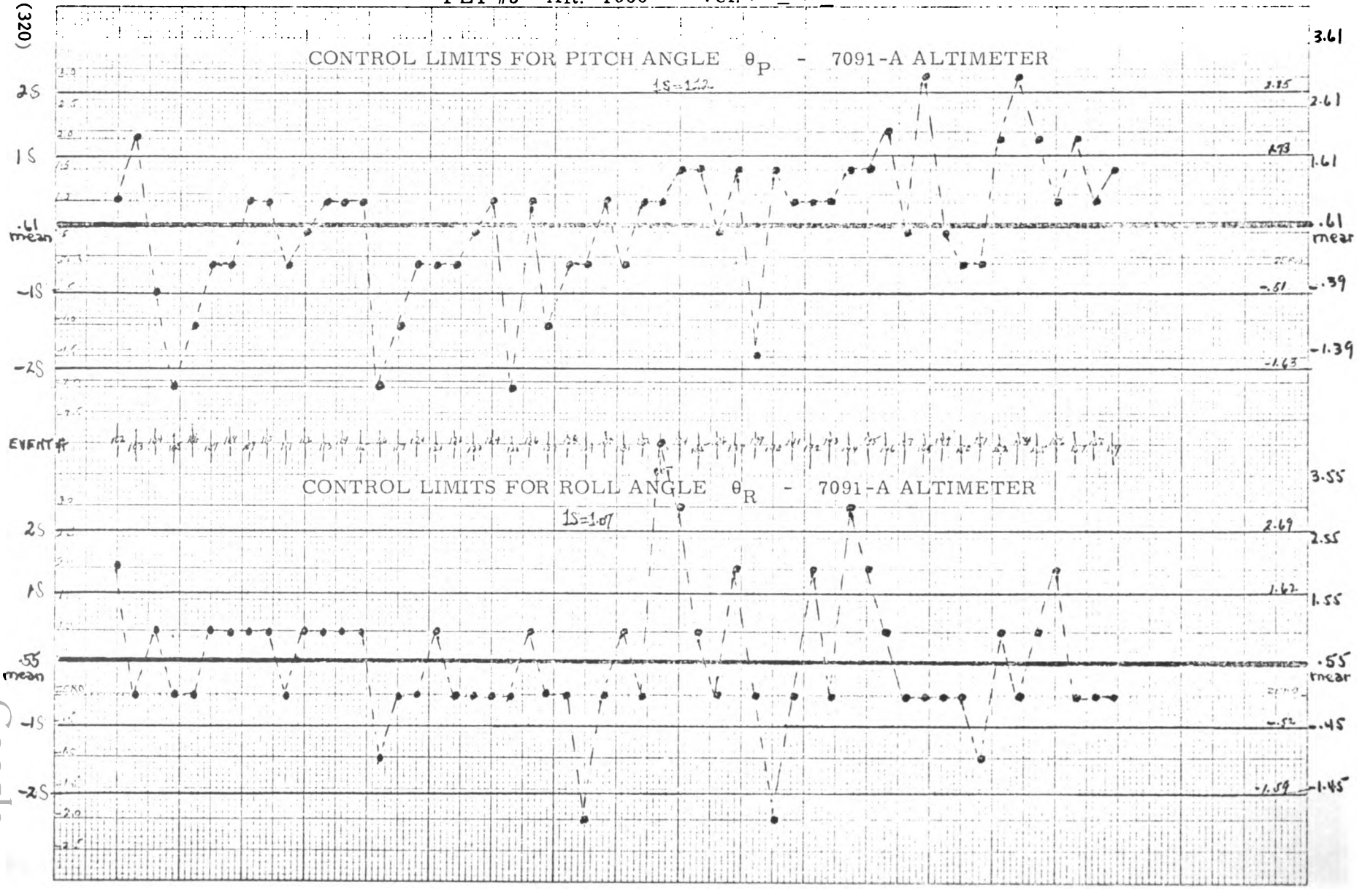




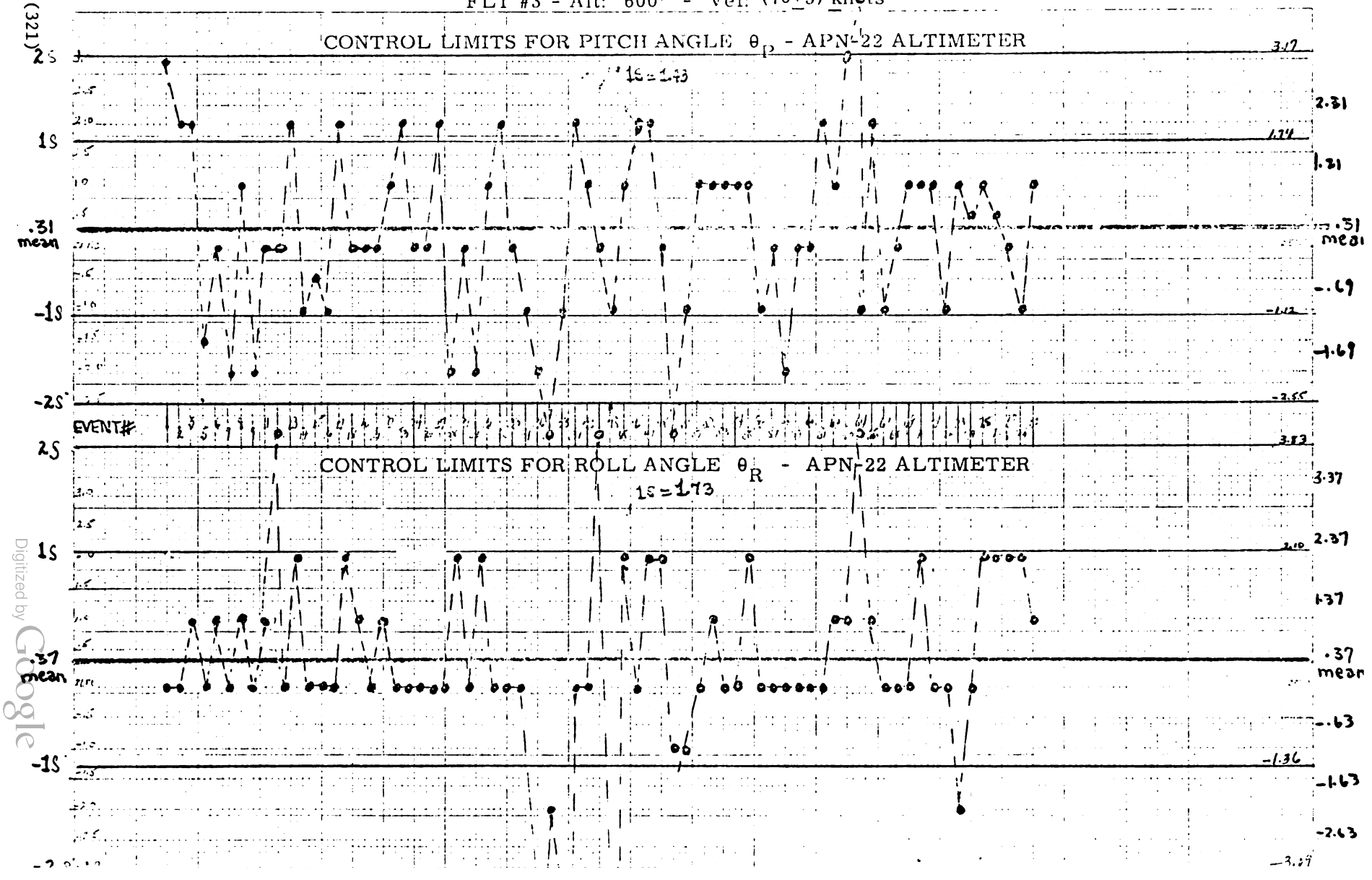




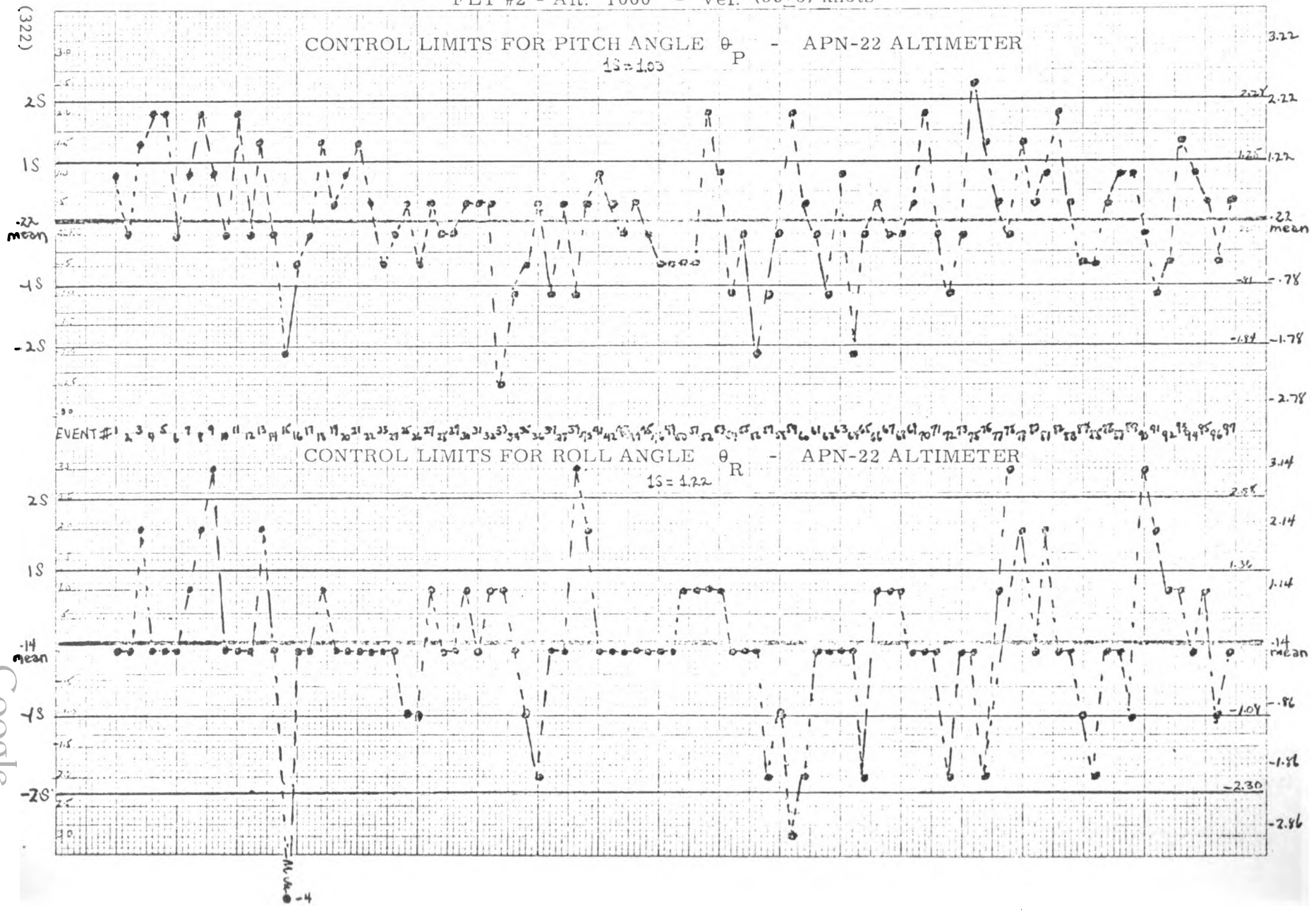
FLT #3 - Alt: 1000' - Vel: (70±5)knots



FLT #3 - Alt: 600' - Vel: (70±5) knots

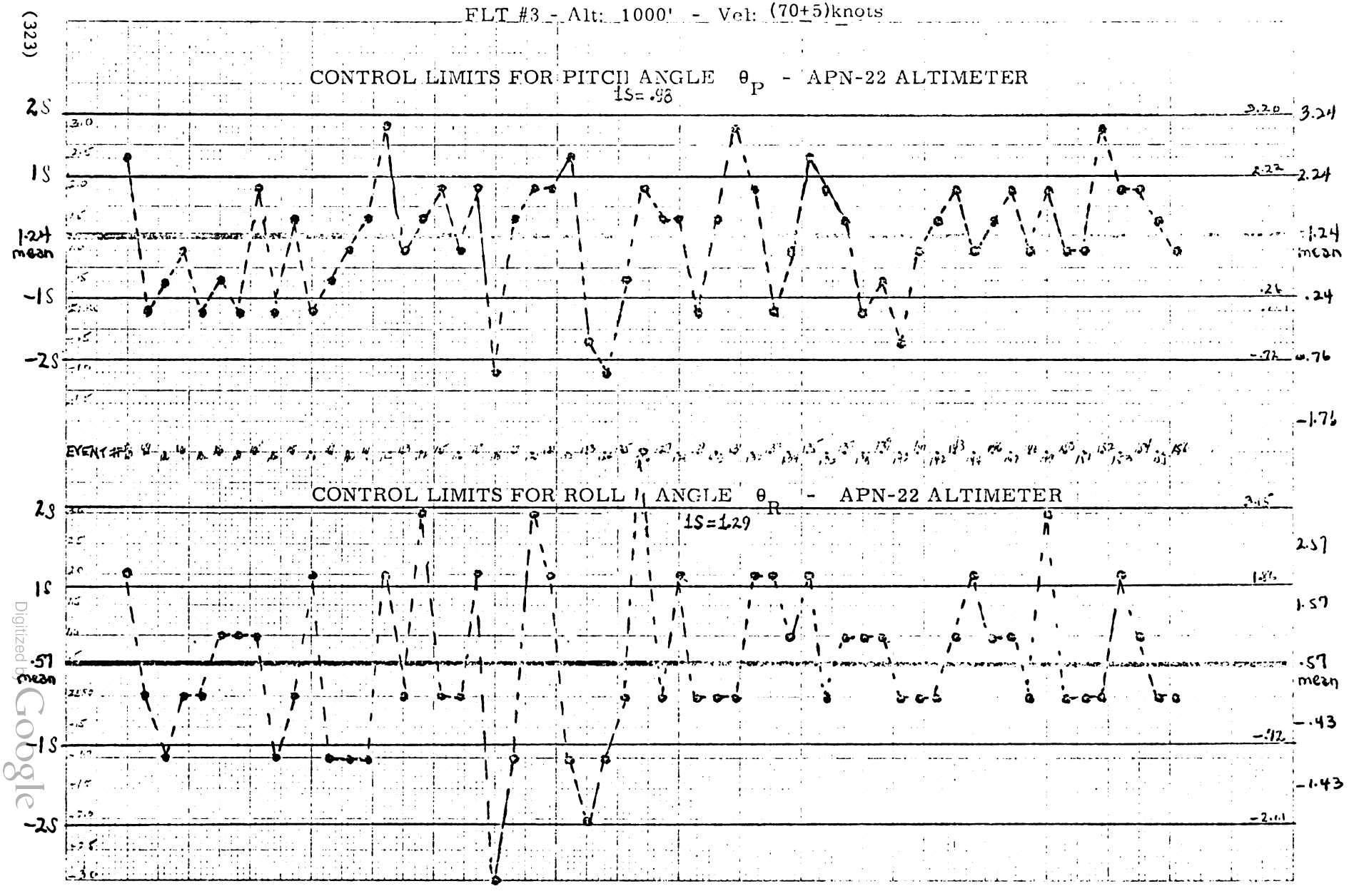


FLT #2 - Alt: 1000' - Vel: (55±5) knots



(323)

FLT #3 - Alt: 1000' - Vel: (70±5)knots



3. DESIGN OF EXPERIMENT.

3.1 Objective of Experiment: The objective of the experimental test is to evaluate the accuracy of the Honeywell Model 7091-A Altimeter (Modified) in a tropical zone, utilizing the nanosecond pulse leading-edge-tracking technique.

3.2 Test Item 1: The Honeywell Model 7091-A Altimeter (Modified) utilizing the nanosecond pulse, leading-edge-tracking technique appeared capable of reasonably good accuracy in previous tests in the temperate and arctic zones. (See "Radar Altimeter Techniques in the Arctic Environment"...R. J. Lucas & R. C. Cruickshank, Presentation at 1966 Meeting of the AGARD Avionics Panel (NATO), Avionics Laboratory, USAECOM, Ft. Monmouth, N. J.) Thus, the same altimeter was selected for testing to determine the accuracy of its design technique in a tropical zone. The Honeywell Model 7091-A Altimeter (Modified) will be referred to as Test Item 1.

3.3 Test Item 2: In the statistical analysis of this experiment, the accuracy of the AN/APN-22 Altimeter, utilizing a frequency modulation, continuous-wave design technique is compared with the accuracy of the Honeywell Model 7091-A Altimeter (Modified). The AN/APN-22 will be referred to as Test Item 2.

3.4 Standard of Reference: The standard of reference for evaluating and comparing the accuracies of the two test items is the RCA Laser Range-Finder AN/GVS-1(XE-6) which has a one (1) meter error (one sigma). The test items and the instrumentation were installed in the CH-34C (CHOCTAW), a helicopter capable of seating twelve people.

3.5 Measured Observations: The measured observations consisted of height readings above the closest foliage at height levels of 600 feet and 1000 feet for the following pieces of equipment:

a. Honeywell Model 7091-A Altimeter (Modified)

   (1) The observations were measured in feet
   (2) The measured observations of height using the Honeywell Model 7091-A Altimeter are symbolized by H_m.

b. AN/APN-22 Altimeter

   (1) The observations were measured in feet
   (2) The measured observations of height using the AN/APN-22 Altimeter are symbolized by H_o.

c. RCA Laser Range-Finder AN/GVS-1(XE-6)

   (1) The observations of height from this piece of equipment serve as a standard reference to determine the deviations of height for both the Honeywell Model 7091-A Altimeter and the AN/APN-22 Altimeter.

(2) The measured observations of height using this piece of equipment are symbolized by H_R.

(3) The observations were measured in meters and converted to feet in the calculations.

(4) The Laser Range-Finder has an accuracy of $\pm$ 1 meter (= 1 sigma).

(5) Environmental specifications:

$$1\ \sigma \text{ error} \simeq \pm \left[5 \text{ ft} + 3\%H + 5^\circ \left(\frac{\partial H}{\partial t}\right)\right]$$

where $\frac{\partial H}{\partial t}$ is the altitude rate.

(6) It was intended to obtain measured observations of height at the following levels of height:

H_1	H_2	H_3	H_4	H_5	H_6
400	600	800	1000	1200	1400

However, the data was obtained only at the 600 ft. and 1000 ft. height levels in the actual experimentation. Data was recorded at the rate of one observation every minute at the 600 ft. and 1000 ft. height levels.

The means and standard deviations of d_{mR} were computed and also the number (N) of positive and negative deviations from H_R:

H_1 = 600 ft.

H_m	H_R	$d_{mR} = H_m - H_R$

Likewise for 1000 ft., etc.

d_{mR} ; $\sigma_{(d_{mR})}$; $N(+d_{mR})$; $N(-d_{mR})$

3.6 Controlled Parameters (constraints):

a. Velocity of Aircraft (CH-34C)

(1) (55±5) knots

(2) (70±5) knots

(3) The quantal error is 5; i.e., the velocity is between 50 and 60 knots or between 65 and 75 knots.

b. Aircraft Attitude

(1) An attempt was made to maintain the aircraft attitude to within ±2° from the vertical (90°); i.e., the deviations of pitch angle (θ_p) and the deviations of roll angle (θ_r) should each be within ±2° of the vertical.

(2) For each RCA Laser Range-Finder measurement taken, corresponding measurements of θ_p and θ_r were taken.

(3) A positive measurement of θ_p indicates the nose of the aircraft is in an upward position; a negative measurement of θ_p indicates the nose of the aircraft is in a downward position.

(4) A positive measurement of θ_r indicates the aircraft is rolling to the right; a negative measurement of θ_r indicates the aircraft is rolling to the left.

(5) It was intended to select from the data a set of observations H_m corresponding to the low attitude angles θ_p and θ_r arranged in order of magnitude for each level of height (H):

H_1 = 600 ft.

θ_p	H_m	θ_r
2° 2.5° . . .		2° . . . .

(Likewise for H_2 = 1000 ft.)

c. Aircraft Height (H) - Levels of height are 600 ft. and 1000 ft. However, it was intended to use height levels of 400 ft. to 1400 ft. in steps of 200 ft.

3.7 Comparison Between Test Items 1 and 2: For each of the two levels of H, height readings were to be taken as recorded from the AN/APN-22. These will be compared with the output of 7091-A (Test Item 1) and, of course, with the Standard of Reference, the RCA Laser:

H_1 = 600 ft.

LASER	7091-A	AN/APN-22 (old equipment)		
H_R	H_m	H_o	d_{oR}	d_{mR}
.	.	.	.	.
.	.	.	.	.
.	.	.	.	.

Where H_o ≡ Height recorded by old equipment (AN/APN-22 Altimeter)

The standard deviations of the observations from the reference data was computed to obtain the distribution of the errors. However, in the actual experimentation, data was obtained only at the 600 ft. and 1000 ft. levels of height.

3.8 List of Equipments:

a. Equipment Items to be Tested:

(1) Test Item 1 - Honeywell Model 7091-A Altimeter (Modified)

(2) Test Item 2 - AN/APN-22 Altimeter

b. Test Instrumentation and Accuracies:

(1) Meter (1 σ) RCA Laser Range-Finder AN/GVS-1 (XE-6) - the output of the Laser is recorded on a decimal drum readout in digital form.

(2) Five-foot (5 ft.) Recorder, Mark 280, Brush (Precision Servo Penmotor Recorder), 2-Channel with two events channels

(a) Two DC analog channels recorded test item altimeters' height data

(b) One event mark channel were synchronized with the laser-firing (once every minute).

(3) 0.5° 1 σ Vertical Gyro, Cageable, Honeywell Mfr. Part No. JG 7044 A-35, SN.04

(a) Used to establish vertical standard for measurements of the deviations of pitch angle and roll angle from the vertical (90°)

(b) Pitch attitude deviations recorded as positive or negative

(c) Roll attitude deviations recorded as positive or negative

(d) The above gyro outputs are displayed on a zero-center meter and are recorded with each laser firing on the decimal drum readout.

(4) TS-352/U Multimeter, Tektronix Model 422 scope, HP Model G382A Variable Attenuator (Precision). The HP Model is used to check the sensitivity (loop gain) of the Test Items.

NOTE: It is to be emphasized that velocity was not treated as a factor, since the radar response (with 1000 pulses per seconds) would not be affected by velocities below 300-400 knots. This is the reason that no interactions were computed.

---

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

4 Analysis of Variance Computations:

(For Testing the Hypothesis of Equal Means Between Test Item Height Observations)

Flight #3 Height ≈ 600 feet Velocity (70±5) knots

7091-A (Test Item 1)	AN/APN-22 (Test Item 2)
$n_1 = 4 + 25 = 29$	$n_2 = 35 + 35 = 70$
$X_{1.} = -979.0$	$X_{2.} = 20.6$
$\bar{X}_{1.} = \frac{-979.0}{29} = -33.75$ feet	$\bar{X}_{2.} = \frac{20.6}{70} = .29$ feet
$\sum_{j}^{n_1} X^2_{ij} = 57517.12$	$\sum_{j}^{n_2} X^2_{ij} = 56241.78$

$$T \equiv N. = 29 + 70 = 99$$

$$\bar{X}.. = \frac{-979.0 + 20.6}{99} = -9.68$$

$$\sum_{i}^{k} \sum_{j}^{n_i} X^2_{ij} = 57517.12 + 56241.78 = 113758.90$$

$$T\bar{X}^2.. = 99(-9.68)^2 = 9276.54$$

Sum of Squares Between Groups (SSB):

$$SSB = \sum_{i}^{k} n_i \bar{X}^2_{i.} - T\bar{X}^2.. = 29(-33.75)^2 + 70(.29)^2 - 9276.54 = 23762.16$$

Sum of Squares Within Groups (SSW):

$$SSW = \sum_{i}^{k} \sum_{j}^{n_i} X^2_{ij} - \sum_{i}^{k} n_i \bar{X}^2_{i.} = 113758.90 - [29(-33.75)^2 + 70(.29)^2] = 80720.$$

Total Sum of Squares (SST):

$$SST = \sum_{i}^{k} \sum_{j}^{n_i} X^2_{ij} - T\bar{X}^2.. = 113758.90 - 99(-9.68)^2 = 104482.36$$

Source of Variation	df	SS	MS
Between Groups	1	23762.16	23762.16
Within Groups	97	80720.20	832.17
Total	98	104482.36	

$$F_{computed}(1, 97) = \frac{23762.16}{832.17} = 28.55$$

$F_{.99}(1, 100) = 6.90$ (tabular value of F-distribution)

Since $28.55 >> 6.90$, the population means are highly significantly different at a significance level of .01 .

Flight # 3 Height ≈ 1000 feet Velocity (70±5) knots

7091-A (Test Item 1)

$n_1 = 7 + 45 = 52$

$X_{1\cdot} = -1517.20$

$\bar{X}_{1\cdot} = \frac{-1517.20}{52} = -29.18$ feet

$\sum_{j}^{n_1} X^2_{ij} = 82670.32$

AN/APN-22 (Test Item 2)

$n_2 = 53 + 5 = 58$

$X_{2\cdot} = 3563.90$

$\bar{X}_{2\cdot} = \frac{3563.90}{58} = 61.45$ feet

$\sum_{j}^{n_2} X^2_{ij} = 291629.06$

$$T \equiv N_{\cdot} = 52 + 58 = 110$$

$$\bar{X}_{\cdot\cdot} = \frac{-1517.20 + 3563.90}{110} = 18.61$$

$$\sum_{i}^{k} \sum_{j}^{n_i} X^2_{ij} = 82670.32 + 291629.06 = 374299.38$$

$$T\bar{X}^2_{\cdot\cdot} = 110(18.61)^2 = 38096.30$$

Sum of Squares Between Groups (SSB):

$$SSB = \sum_{i}^{k} n_i \bar{X}^2_{i\cdot} - T\bar{X}^2_{\cdot\cdot} = 52(-29.18)^2 + 58(61.45)^2 - 38096.30 = 225193.94$$

Sum of Squares Within Groups (SSW):

$$SSW = \sum_{i}^{k} \sum_{j}^{n_i} X^2_{ij} - \sum_{i}^{k} n_i \bar{X}^2_{i\cdot} = 374299.38 - 263290.24 = 111009.14$$

Total Sum of Squares (SST):

$$SST = \sum_{i}^{k} \sum_{j}^{n_i} X^2_{ij} - T\bar{X}^2_{\cdot\cdot} = 374299.38 - 38096.30 = 336203.08$$

Source of Variation	df	SS	MS
Between Groups	1	225193.94	225193.94
Within Groups	108	111009.14	1027.86
Total	109	336203.08	

$$F_{computed}(1,108) = \frac{225193.94}{1027.86} = 219.09$$

$F_{.99}(1,100) = 6.90$ (tabular value of F-distribution)

Since $219.09 >> 6.90$, the population means are highly significantly different at a significance level of .01.

Flight #2 Height ≈ 1000 feet Velocity (55±5) knots

7091-A (Test Item 1)

$n_1 = 12 + 79 = 91$

$X_{1.} = -3135.00$

$\bar{X}_{1.} = \frac{-3135.00}{91} = -34.45$ feet

$\sum_{j}^{n_1} X_{ij}^2 = 180825.10$

AN/APN-22 (Test Item 2)

$n_2 = 78 + 13 = 91$

$X_{2.} = 3374.80$

$\bar{X}_{2.} = \frac{3374.80}{91} = 37.09$ feet

$\sum_{j}^{n_2} X_{ij}^2 = 239507.26$

$$T \equiv N_{.} = 91 + 91 = 182$$

$$\bar{X}_{..} = \frac{-3135.0 + 3374.8}{182} = 1.32$$

$$\sum_{i}^{k} \sum_{j}^{n_i} X_{ij}^2 = 180825.10 + 239507.26 = 420332.36$$

$$T\bar{X}^2_{..} = 182\,(1.32)^2 = 317.12$$

Sum of Squares Between Groups (SSB):

$$SSB = \sum_{i}^{k} n_i \bar{X}_{i.}^2 - T\bar{X}^2_{..} = 91\,(-31.45)^2 + 91\,(37.09)^2 - 317.12 = 232867.70$$

Sum of Squares Within Groups (SSW):

$$SSW = \sum_{i}^{k} \sum_{j}^{n_i} X_{ij}^2 - \sum_{i}^{k} n_i \bar{X}_{i.}^2 = 420332.36 - 233184.82 = 187147.54$$

Total Sum of Squares (SST):

$$SST = \sum_{i}^{k} \sum_{j}^{n_i} X_{ij}^2 - T\bar{X}^2_{..} = 420332.36 - 317.12 = 420015.24$$

Source of Variation	df	SS	MS
Between Groups	1	232867.70	232867.70
Within Groups	180	187147.54	1039.71
Total	181	420015.24	

$$F_{computed}(1, 180) = \frac{232867.70}{1039.71} = 223.97$$

$F_{.99}(1, 150) = 6.81$ (tabular value of F-distribution)

Since $223.97 >> 6.81$, the population means are highly significantly different at a significance level of .01 .

5. t-Test Computations:

Testing the Hypothesis of Equal Means Between Test Item Height Observations

Only negative deviations for the 7091-A Altimeter and positive deviations for the AN/APN-22 Altimeter were used in the following calculations because of their predominant occurrence in the data.

Flight #3	Height ≈ 600 feet	Velocity: (70±5) knots
7091-A (Test Item 1)		AN/APN-22 (Test Item 2)
$n_1 = 25$		$n_2 = 35$
$\bar{X}_1 = -40.96$		$\bar{X}_2 = 22.61$
$s_1 = 24.15$		$s_2 = 16.82$

One-Sided Test:

$$t = (\bar{X}_1 - \bar{X}_2) \sqrt{\frac{n_1 n_2 (n_1+n_2-2)}{(n_1+n_2)(n_1 s_1^2 + n_2 s_2^2)}}$$

$$t = [-40.96 - 22.61] \cdot \sqrt{\frac{(25x35)(25+35-2)}{(25+35)[25(24.15)^2+35(16.82)^2]}}$$

$$t = -20.91$$

For a significance level of $\alpha = .05$, the tabular value of the t-distribution table is: $t_{1-.05}(50) = t_{.95}(50) = 1.67$.

Since $-20.91 << -1.67$, the population means of the test items are highly significantly different at $\alpha = .05$ level of significance.

Furthermore, for a significance level of $\alpha = .005$, the tabular value of the t-distribution table is: $t_{1-.005}(50) = t_{.995}(50) = 2.68$.

Therefore, since $-20.91 << -2.68$, the population means of the test items are also highly significantly different at $\alpha = .005$ level of significance.

Two-Sided Test:

For a significance level of $\alpha = .05$, the tabular value of the t-distribution table is: $t_{1-\frac{.05}{2}}(50) = t_{.975}(50) = 2.01$.

Since $|t| = 20.91 >> 2.01$, the population means of the test items are highly significantly different at $\alpha = .05$ level of significance.

Furthermore, for a significance level of $\alpha = .01$, the tabular value of the t-distribution table is: $t_{1-\frac{.01}{2}}(50) = t_{.995}(50) = 2.68$.

Since $|t| = 20.91 >> 2.68$, the population means of the test items are also highly significantly different at $\alpha = .01$ level of significance.

Flight #2 Height ≈ 1000 feet Velocity: (55±5) knots

7091-A (Test Item 1)

$n_1 = 79$

$\bar{X}_1 = -40.96$

$s_1 = 24.35$

AN/APN-22 (Test Item 2)

$n_2 = 78$

$\bar{X}_2 = 46.73$

$s_2 = 28.01$

One-Sided Test:

$$t = [-40.96-46.73]\cdot\sqrt{\frac{(79x78)(79+78-2)}{(79+78)[79(24.35)^2+78(28.01)^2]}}$$

$t = -20.80$

For a significance level of $\alpha = .05$, the tabular value of the t-distribution table is: $t_{1-.05}(100) = t_{.95}(100) = 1.66$.

Since $-20.80 << -1.66$, the population means of the test items are highly significantly different at $\alpha = .05$ level of significance.

Furthermore, for a significance level of $\alpha = .005$, the tabular value of the t-distribution table is: $t_{1-.005}(100) = t_{.995}(100) = 2.63$.

Therefore, since $-20.80 << -2.63$, the population means of the test items are also highly significantly different at $\alpha = .005$ level of significance.

Two-Sided Test:

For a significance level of $\alpha = .05$, the tabular value of the t-distribution table is: $t_{1-\frac{.05}{2}}(100) = t_{.975}(100) = 1.98$.

Since $|t| = 20.80 >> 1.98$, the population means of the test items are highly significantly different at $\alpha = .05$ level of significance.

Furthermore, for a significance level of $\alpha = .01$, the tabular value of the t-distribution table is: $t_{1-\frac{.01}{2}}(100) = t_{.995}(100) = 2.63$

Since $|t| = 20.80 >> 2.63$, the population means of the test items are also highly significantly different at $\alpha = .01$ level of significance.

Flight #3 Height ≈ 1000 feet Velocity: (70±5) knots

7091-A (Test Item 1)	AN/APN-22 (Test Item 2)
$n_1 = 45$	$n_2 = 53$
$\bar{X}_1 = -34.57$	$\bar{X}_2 = 67.96$
$s_1 = 24.01$	$s_2 = 29.52$

One-Sided Test:

$$t = [-34.57 - 67.96] \cdot \sqrt{\frac{(45 \times 53)(45+53-2)}{(45+53)[45(24.01)^2+53(29.52)^2]}}$$

$$t = -18.42$$

For a significance level of $\alpha = .05$, the tabular value of the t-distribution table is: $t_{1-.05}(80) = t_{.95}(80) = 1.66$.

Since $-18.42 << -1.66$, the population means of the test items are highly significantly different at $\alpha = .05$ level of significance.

Furthermore, for a significance level of $\alpha = .005$, the tabular value of the t-distribution table is: $t_{1-.005}(80) = t_{.995}(80) = 2.64$;

Since $-18.42 << -2.64$, the population means of the test items are also highly significantly different at $\alpha = .005$ level of significance.

Two-Sided Test:

For a significance level of $\alpha = .05$, the tabular value of the t-distribution table is: $t_{1-\frac{.05}{2}}(80) = t_{.975}(80) = 1.99$.

Since $|t| = 18.42 >> 1.99$, the population means of the test items are highly significantly different at $\alpha = .05$ level of significance.

Furthermore, for a significance level of $\alpha = .01$, the tabular value of the t-distribution table is: $t_{1-\frac{.01}{2}}(80) = t_{.995}(80) = 2.64$.

Since $|t| = 18.42 >> 2.64$, the population means of the test items are also highly significantly different at $\alpha = .01$ level of significance.

6. HOMOGENEITY OF VARIANCE COMPUTATIONS:

F-Test Computations for Testing Homogeneity of Variance Between Test Item Height Observations

Only negative deviations for the 7091-A Altimeter and positive deviations for the AN/APN-22 Altimeter were used in the following calculations because of their predominant occurrence in the data.

Flight #3 Height ≈ 600 feet Velocity (70±5) knots

7091-A (Test Item 1)	AN/APN-22 (Test Item 2)
$n_1 = 25$	$n_2 = 35$
$s_{(d_{mR})} = 24.15$	$s_{(d_{oR})} = 16.82$
$s^2_{(d_{mR})} = 583.22$	$s^2_{(d_{oR})} = 282.91$

$$F_{computed}(24, 34) = \frac{583.22}{282.91} \quad 2.06$$

$F_{.95}(24, 34) = 1.84$ (tabular value of F-distribution)

$F_{.99}(24, 34) = 2.38$ (tabular value of F-distribution)

Since $2.06 > 1.84$, the hypothesis that $\sigma^2_{(d_{mR})} = \sigma^2_{(d_{oR})}$ is contradicted by the observed data at a significance level $\alpha = .05$.

Since $2.06 < 2.38$, the hypothesis that $\sigma^2_{(d_{mR})} = \sigma^2_{(d_{oR})}$ is not contradicted by the observed data at a significance level $\alpha = .01$.

Flight #3 Height ≈ 1000 feet Velocity (70±5) knots

7091-A (Test Item 1)	AN/APN-22 (Test Item 2)
$n_1 = 45$	$n_2 = 53$
$s_{(d_{mR})} = 24.01$	$s_{(d_{oR})} = 29.52$
$s^2_{(d_{mR})} = 576.48$	$s^2_{(d_{oR})} = 871.43$

$$F_{computed}(52, 44) = \frac{871.43}{576.48} = 1.51$$

$F_{.95}(50, 44) = 1.63$ (tabular value of F-distribution)

$F_{.99}(50, 44) = 2.00$ (tabular value of F-distribution)

Since $1.51 < 1.63$ and $1.51 < 2.00$, the hypothesis that $\sigma^2_{(d_{mR})} = \sigma^2_{(d_{oR})}$ is not contradicted by the observed data at either significance level of $\alpha = .05$ or $\alpha = .01$.

Flight #2 Height ≈ 1000 feet Velocity (55±5) knots

7091-A (Test Item 1)	AN/APN-22 (Test Item 2)
$n_1 = 79$	$n_2 = 78$
$s_{(d_{mR})} = 24.35$	$s_{(d_{oR})} = 28.01$
$s^2_{(d_{mR})} = 592.92$	$s^2_{(d_{oR})} = 784.56$

$$F_{computed}(77, 78) = \frac{784.56}{592.92} = 1.32$$

$F_{.95}(75, 80) = 1.45$ (tabular value of F-distribution)

$F_{.99}(75, 80) = 1.70$ (tabular value of F-distribution)

Since $1.32 < 1.45$ and $1.32 < 1.70$, the hypothesis that $\sigma^2_{(d_{mR})} = \sigma^2_{(d_{oR})}$ is not contradicted by the observed data at either significance level of $\alpha = .05$ or $\alpha = .01$.

FACTORS			7091-A ALTIMETER						
			Negative Deviations (feet)			Mean Confidence Interval (feet)		Standard Deviation Confidence Interval(ft)	
Height (feet)	Flt. No.	Velocity (kts±5)	N(-d)	$\bar{d}_{mR}$	$s_{(d_{mR})}$	95%	99%	95%	99%
600	1	55	51	-32.10	21.39	-38.18 — -26.02	-40.21 — -23.99	17.90 — 26.59	16.96 — 28.59
	3	70	25	-40.96	24.15	-51.11 — -30.81	-54.76 — -27.16	18.86 — 33.60	17.53 — 37.63
1000	2	55	79	-40.96	24.35	-46.45 — -35.47	-48.24 — -33.68	20.83 — 28.45	19.94 — 30.06
	3	70	45	-35.34	24.01	-42.63 — -28.05	-45.08 — -25.60	20.67 — 32.22	19.49 — 35.00

CONFIDENCE INTERVALS
of μ and σ for 7091-A Altimeter
Measurements

TABLE 5

Velocity not treated as a factor.

FACTORS			AN/APN-22 ALTIMETER						
			Positive Deviations (feet)			Mean Confidence Interval (feet)		Standard Deviation Confidence Interval(ft)	
Height (feet)	Flt No.	Velocity (kts±5)	N(+d)	$\bar{d}_{oR}$	$s_{(d_{oR})}$	95%	99%	95%	99%
600	1	55		(S E E N O T E B E L O W)					
	3	70	35	22.61	16.82	16.75 — 28.47	14.75 — 30.47	14.31 — 23.93	13.39 — 26.41
1000	2	55	78	48.73	28.01	40.38 — 53.08	38.30 — 55.16	23.80 — 32.51	22.79 — 34.36
	3	70	53	67.96	29.52	59.73 — 76.19	56.99 — 78.93	25.19 — 37.42	23.88 — 40.24

CONFIDENCE INTERVALS
of μ and σ for AN/APN-22 Altimeter
Measurements

TABLE 6

Velocity not treated as a factor.

NOTE: No experimentation was performed with the AN/APN-22 Altimeter at a height of 600 feet and a velocity of (55±5) knots.

Computations: Confidence Intervals for Means of 7091-A Altimeter Measurements

The following confidence interval computations were performed only with the negative deviations of the 7091-A Altimeter because of their predominant occurrence in the data.

$$\text{Confidence Limits} = \bar{X} \pm t_{\alpha} \frac{s}{\sqrt{n-1}}$$

Flight #1 Height ≈ 600 feet Velocity: (55±5) knots

$n = 51$ $s = 21.39$ $\bar{X} = -32.10$

95% Confidence Interval $= -32.10 \pm 2.01 \left(\frac{21.39}{\sqrt{50}}\right) = -32.10 \pm 6.0802$

$= -38.18$ to -26.02

99% Confidence Interval $= -32.10 \pm 2.68 \left(\frac{21.39}{\sqrt{50}}\right) = -32.10 \pm 8.1070$

$= -40.21$ to -23.99

Flight # 3 Height ≈ 600 feet Velocity: (70±5) knots

$n = 25$ $s = 24.15$ $\bar{X} = -40.96$

95% Confidence Interval $= -40.96 \pm 2.06 \left(\frac{24.15}{\sqrt{24}}\right) = -40.96 \pm 10.1549$

$= -51.11$ to -30.81

99% Confidence Interval $= -40.96 \pm 2.80 \left(\frac{24.15}{\sqrt{24}}\right) = -40.96 \pm 13.8028$

$= -54.76$ to -27.16

Flight # 2 Height ≈ 1000 feet Velocity: (55±5) knots

$n = 79$ $s = 24.35$ $\bar{X} = -40.96$

95% Confidence Interval $= -40.96 \pm 1.99 \left(\frac{24.35}{\sqrt{78}}\right) = -40.96 \pm 5.4866$

$= -46.45$ to -35.47

99% Confidence Interval $= -40.96 \pm 2.64 \left(\frac{24.35}{\sqrt{78}}\right) = -40.96 \pm 7.2787$

$= -48.24$ to -33.68

Flight #3 Height ≈ 1000 feet Velocity: (70±5) knots

$n = 45$ $s = 24.01$ $\bar{X} = -35.34$

95% Confidence Interval $= -35.34 \pm 2.015 \left(\frac{24.01}{\sqrt{44}}\right) = -35.34 \pm 7.2934$

$= -42.63$ to -28.05

99% Confidence Interval $= -35.34 \pm 2.69 \left(\frac{24.01}{\sqrt{44}}\right) = -35.34 \pm 9.7367$

$= -45.08$ to -25.60

Computations: Confidence Intervals for Means of AN/APN-22 Altimeter Measurements

The following confidence interval computations were performed only with the positive deviations of the AN/APN-22 Altimeter because of their predominant occurrence in the data.

$$\text{Confidence Limits} = \bar{X} \pm t_{\alpha} \frac{s}{\sqrt{n-1}}$$

Flight # 3 Height ≈ 600 feet Velocity: (70±5) knots

$n = 35$ $s = 16.82$ $\bar{X} = 22.61$

95% Confidence Interval $= 22.61 \pm 2.03 \left(\frac{16.82}{\sqrt{34}}\right)$ $= 22.61 \pm 5.8557$

$= 16.75$ to 28.47

99% Confidence Interval $= 22.61 \pm 2.725 \left(\frac{16.82}{\sqrt{34}}\right)$ $= 22.61 \pm 7.8605$

$= 14.75$ to 30.47

Flight #2 Height ≈1000 feet Velocity: (55±5) knots

$n = 78$ $s = 28.01$ $\bar{X} = 46.73$

95% Confidence Interval $= 46.73 \pm 1.99 \left(\frac{28.01}{\sqrt{77}}\right)$ $= 46.73 \pm 6.3520$

$= 40.38$ to 53.08

99% Confidence Interval $= 46.73 \pm 2.64 \left(\frac{28.01}{\sqrt{77}}\right)$ $= 46.73 \pm 8.4268$

$= 38.30$ to 55.16

Flight #3 Height ≈1000 feet Velocity: (70±5) knots

$n = 53$ $s = 29.52$ $\bar{X} = 67.96$

95% Confidence Interval $= 67.96 \pm 2.01 \left(\frac{29.52}{\sqrt{52}}\right)$ $= 67.96 \pm 8.2281$

$= 59.73$ to 76.19

99% Confidence Interval $= 67.96 \pm 2.68 \left(\frac{29.52}{\sqrt{52}}\right)$ $= 67.96 \pm 10.9708$

$= 56.99$ to 78.93

Computations: Confidence Intervals for Standard Deviations of 7091-A Altimeter Measurements

The following confidence interval computations were performed only with the negative deviations of the 7091-A Altimeter because of their predominant occurrence in the data.

$\frac{\nu s^2}{\chi^2_{\nu, 1-\alpha}} > \sigma^2$ is the $100(1-\alpha)\%$ upper confidence interval for σ^2

$\frac{\nu s^2}{\chi^2_{\nu, \alpha}} < \sigma^2$ is the $100(1-\alpha)\%$ lower confidence limit for σ^2

Two-sided confidence interval for the unknown σ^2 :

$$\frac{\nu s^2}{\chi^2_{\nu, \frac{\alpha}{2}}} < \sigma^2 < \frac{\nu s^2}{\chi^2_{\nu, 1-\frac{\alpha}{2}}}$$

where ν represents the number of degrees of freedom, α is the level of significance, and χ is the Chi-Square distribution.

Flight #1 Height ≈ 600 feet Velocity: (55±5) knots

$n = 51$ $s = 21.39$

95% Confidence Interval: $\frac{50(21.39)^2}{71.4202} < \sigma^2 < \frac{50(21.39)^2}{32.3574}$

$320.31 < \sigma^2 < 706.9976$

$17.90 < \sigma < 26.59$

99% Confidence Interval: $\frac{50(21.39)^2}{79.4900} < \sigma^2 < \frac{50(21.39)^2}{27.9907}$

$287.7922 < \sigma^2 < 817.2930$

$16.96 < \sigma < 28.59$

Flight #3 Height ≈ 600 feet Velocity: (70±5) knots

$n = 25$ $s = 24.15$

95% Confidence Interval: $\frac{24(24.15)^2}{39.3641} < \sigma^2 < \frac{24(24.15)^2}{12.4001}$

$355.5864 < \sigma^2 < 1128.8086$

$18.86 < \sigma < 33.60$

99% Confidence Interval: $\frac{24(24.15)^2}{45.5585} < \sigma^2 < \frac{24(24.15)^2}{9.88623}$

$307.2388 < \sigma^2 < 1415.8463$

$17.53 < \sigma < 37.63$

Computations: Confidence Intervals for Standard Deviations of AN/APN-22 Altimeter Measurements

The following confidence interval computations were performed only with the positive deviations of the AN/APN-22 Altimeter because of their predominant occurrence in the data.

The confidence interval for the unknown σ^2 is:

$$\frac{\nu s^2}{\chi^2_{\nu,\frac{\alpha}{2}}} < \sigma^2 < \frac{\nu s^2}{\chi^2_{\nu,1-\frac{\alpha}{2}}}$$

Flight # 3 Height ≈ 600 feet Velocity: (70±5) knots

$n = 35$ $s = 16.82$

95% Confidence Interval: $\frac{34(16.82)^2}{46.9792} < \sigma^2 < \frac{34(16.82)^2}{16.7908}$

$204.7506 < \sigma^2 < 572.8745$

$14.31 < \sigma < 23.93$

99% Confidence Interval: $\frac{34(16.82)^2}{53.6720} < \sigma^2 < \frac{34(16.82)^2}{13.7867}$

$179.2186 < \sigma^2 < 697.7029$

$13.39 < \sigma < 26.41$

Flight #3 Height ≈ 1000 feet Velocity: (70 ±5) knots

$n = 45$ $s = 24.01$

95% Confidence Interval: $\frac{44(24.01)^2}{59.3417} < \sigma^2 < \frac{44(24.01)^2}{24.4331}$

$427.4418 < \sigma^2 < 1038.1459$

$20.67 < \sigma < 32.22$

99% Confidence Interval: $\frac{44(24.01)^2}{66.7659} < \sigma^2 < \frac{44(24.01)^2}{20.7065}$

$379.9113 < \sigma^2 < 1224.9836$

$19.49 < \sigma < 35.00$

Flight #2 Height ≈ 1000 feet Velocity: (55±5) knots

n = 78 s = 28.01

95% Confidence Interval: $\frac{77(28.01)^2}{106.629} < \sigma^2 < \frac{77(28.01)^2}{57.1532}$

$566.5543 < \sigma^2 < 1057.0034$

$23.80 < \sigma < 32.51$

99% Confidence Interval: $\frac{77(28.01)^2}{116.321} < \sigma^2 < \frac{77(28.01)^2}{51.1720}$

$519.3484 < \sigma^2 < 1180.5504$

$22.79 < \sigma < 34.36$

Flight # 2 Height ≈ 1000 feet Velocity: (55±5) knots

n = 79 s = 24.35

95% Confidence Interval: $\frac{78(24.35)^2}{106.629} < \sigma^2 < \frac{78(24.35)^2}{57.1532}$

$433.7277 < \sigma^2 < 809.1927$

$20.83 < \sigma < 28.45$

99% Confidence Interval: $\frac{78(24.35)^2}{116.321} < \sigma^2 < \frac{78(24.35)^2}{51.1720}$

$397.5890 < \sigma^2 < 903.7746$

$19.94 < \sigma < 30.06$

Flight # 3 Height ≈ 1000 feet Velocity: (70±5) knots

n = 53 s = 29.52

95% Confidence Interval: $\frac{52(29.52)^2}{71.4202} < \sigma^2 < \frac{52(29.52)^2}{32.3574}$

$634.4756 < \sigma^2 < 1400.4333$

$25.19 < \sigma < 37.42$

99% Confidence Interval: $\frac{52(29.52)^2}{79.4900} < \sigma^2 < \frac{52(29.52)^2}{27.9907}$

$570.0639 < \sigma^2 < 1618.9084$

$23.88 < \sigma < 40.24$

8. MEAN SQUARE SUCCESSIVE DIFFERENCE: A Test for Randomness

One of the tests used to detect randomness is the mean square difference method. A brief discussion of this method is deemed advisable because of the sensitivity of this method to non-random fluctuations. This method is particularly sensitive in detecting long-term trends, periodic or excessively rapid oscillations in observed data.

Let us assume that $X_1, X_2, \ldots, X_n$ represent n successive observations from a population which obeys the normal distribution law:

$$f(X) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left[-(X-\mu)^2/2\sigma^2\right]$$

with the mean μ and standard deviation σ. The sample mean and standard deviation are defined respectively:

$$\bar{X} = \frac{1}{n}\sum_1^n X_i \quad ; \qquad s = \sqrt{\frac{1}{n-1}\sum_1^n (X_i - \bar{X})^2}$$

The mean square difference is:

$$\delta^2 = \frac{1}{n-1}\sum_i^n (X_{i+1} - X_i)^2$$

i.e., we compute the mean of the squares of the n-1 successive differences between the observations.

It can be shown that:

$$E\left[\frac{\delta^2/2}{s^2}\right] = 1 \quad ,$$ and thus $\delta^2/2$ is an unbiased estimate of σ^2.

The variance is: $$V\left[\frac{\delta^2/2}{s^2}\right] = \frac{n-2}{(n-1)(n+1)}$$

In this test we are comparing the values of δ^2 and s^2 ; we are particularly interested in the ratio: δ^2/s^2 , since the disparity between the values of δ^2 and s^2 will indicate the trend or short period oscillations in the observations. It is assumed that the value of δ^2 will not be increased by the trend as appreciably as s^2 ; hence a small value of the ratio δ^2/s^2 will indicate trends in the observations. In the case of short periods of oscillations both δ^2 and s^2 will increase; and the increases in δ^2 will be proportionately greater.

The distribution of $1 - \frac{\delta^2}{2s^2} \equiv \theta$

is symmetrical with average value zero for random samples drawn from a normal population. For values of $n > 25$, θ is very nearly normally distributed with average zero and variance equal to

$$\frac{n-2}{(n-1)(n+1)} \quad .$$

We can use the statistic: $t = \theta/\sigma_\theta$ and the percentage point, for a standard normal deviate in testing for significance of θ for large values of n . Long term trends in the observations would be indicated by high negative values of t ; and high positive values of t would be symptomatic of short rapid oscillations in the observations. Significance levels for the δ^2/s^2 ratios have been tabulated by B. I. Hart (Significance Levels for the Ratio of the Mean Square Successive Difference to the Variance, Annals of Mathematical Statistics, Vol. XIII, 1942, pp. 445-447).

It is clear from Tables 7 and 8 on the following pages that the values of $t\left(=\frac{\theta}{\sigma_\theta}\right)$ at the 95% confidence level are not statistically significant. This means that the hypothesis of randomness of the data is not rejected at the 5% significance level.

FACTORS			7091-A ALTIMETER						
Height (feet)	Flt #	Velocity (kts ±5)	n	s^2	δ^2	$\eta = \frac{\delta^2}{s^2}$	$\theta = 1 - \frac{\eta}{2}$	$\sigma^2_\theta = \frac{n-2}{(n-1)(n+1)}$	$t = \frac{\theta}{\sigma_\theta}$
600	1	55	59	559.16	847.49	1.52	0.240	0.0163	1.8798
	3	70	29	873.84	1884.28	2.16	-0.080	0.0321	-0.4465
1000	2	55	91	807.62	1588.78	1.97	0.015	0.0107	0.1450
	3	70	52	753.00	1341.40	1.78	0.110	0.0184	0.8109

SUMMARY OF THE MEAN SQUARE SUCCESSIVE DIFFERENCES

Velocity not taken as a factor.

TABLE 7

FACTORS			AN/APN-22 ALTIMETER						
Height (feet)	Flt #	Velocity (kts±5)	n	s^2	δ^2	$\eta = \frac{\delta^2}{s^2}$	$\rho = 1 - \frac{\eta}{2}$	$\sigma_\rho^2 = \frac{n-2}{(n-1)(n+1)}$	$t = \frac{\rho}{\sigma_\rho}$
600	1	55		(SEE NOTE BELOW)					
	3	70	70	815.01	1764.91	2.17	-0.085	0.0138	-0.7236
1000	2	55	91	1381.67	1917.77	1.39	0.305	0.0107	2.9486
	3	70	58	1376.16	2977.51	2.16	-0.080	0.0166	-0.6209

SUMMARY OF THE MEAN SQUARE SUCCESSIVE DIFFERENCES

Velocity not taken as a factor.

TABLE 8

NOTE: No experimentation was performed with the AN/APN-22 Altimeter at a height of 600 ft. and a velocity of (55±5 Knots).

AN EXPERIMENT USING NUMERICAL ANALYSIS TO MODEL A FUNCTIONAL RELATION BETWEEN ABM SYSTEM SENSOR RESPONSES AND REENTRY VEHICLE CHARACTERISTICS

Andrew H. Jenkins
U. S. Army Missile Command
Redstone Arsenal, Alabama

INTRODUCTION. It is believed that prediction models can be developed by the analysis of experimental data in light of the known physical laws pertinent to high speed reentry. The development of the model is accomplished by numerical analysis of the full-scale reentry experimental data obtained on the eastern and western test ranges.

In the past, considerable effort has been expended to rigorously and theoretically describe the interdependent and interacting phenomena of hypervelocity reentry. This is a very complex and difficult job. In general, the basic theoretical relations are not adequately described for ideal conditions. Of more importance, the real case of reentry is usually described with even less precision than the ideal case. This is not to say that progress has not been made in the purely theoretical approach nor is it to imply that it should not continue. The selection of the proper variables and stratification of empirical models depends upon such efforts.

The phenomenological processes which occur during reentry couple with the radar sensor to produce gross effects in the measurable responses. These gross effects are considered to be typical from test to test, and differ only in the degree or level of effect on the response. The empirical determination of the degree or level of effect is to relate the sensor responses to the body parameters and the trajectory parameters by experimental observation without the full benefit of a complete theoretical knowledge to describe the underlying physics and chemistry of the phenomena. This is graphically depicted in Figure 1.

The variables used are those which are recorded by the radar system on data tape or published in data reports and are, of course, representative of the real-case responses in a real time frame. Accurate estimates of body characteristics made continuously in real time are the ultimate goal of this approach. Also, it is desired that the prediction models contain sufficient physical variables representative of the sensor, body, and trajectory parameters that the simultaneous masking of all measurables becomes economically and practically infeasible for the offense.

In the development of empirical prediction models, the operational conditions should not be ignored. The final utility of any techniques of target identification depends upon the capability of the model to make

accurate real-time estimates. Also, the models should be fairly easy and economical to incorporate into the defense system. It should be one which can be improved in accuracy and updated as more knowledge of the problem is accumulated.

In the real operational situation, the defense system essentially stands alone. The identification of the objects in the reentry complex must be done in real time. This can be less than 60 seconds. The models should be able to provide continuous estimates of the characteristics of the objects in the reentry complex. It is also highly desirable that the model estimation process converge as soon as possible in the real-time track to the best estimate of the true value of the particular body parameter (for example, weight). This provides a longer time for decision making or for intercept at the highest possible point. The estimation of as many body parameters as possible is obviously highly desirable. The body parameters can be used for cross-checks on the estimated values of each other. Every object in a reentry complex will not be a simple decision case of "warhead-decoy" even with a very precise model. There will be grey areas. Therefore, it is believed that several models for different characteristics of the body will be essential in the final decision to commit an interceptor.

It has not been determined just how more than one body parameter estimate will be made in real time. It may be required to tabulate the data in the form of discrete time (for example, altitude) intervals and develop prediction models for each time increment and body characteristic rather than use one model throughout the reentry track.

It is mentioned that the material presented in this report represents a minimal effort which is neither complete nor concrete. Some of the variables used in this "first cut" numerical analysis were selected because of expediency and availability in order to make a beginning in this approach.

FORMULATING THE MODEL. The most common physical characteristics of the body are weight (W), diameter (D), and length (L). The drag area product used in conjunction with weight can provide an estimate of ballistic coefficient (β). Shape is one characteristic that affects the drag area product (C_DA) for a given set of reentry trajectory conditions and is reflected in the value of C_DA. This value in turn is reflected in the ballistic coefficient.

Some measurements that can be made by the radar are radar cross section (σ), velocity (V), time derivative of velocity ($\dot{V}$), and altitude (h). There are characteristics of radar such as wavelength (λ) and aspect angle (ϕ) on which the above measurements depend.

The first body characteristic selected to empirically determine the functional dependence is vehicle weight (W). It is hypothesized

that the estimated weight W' is not a function of sensor characteristics and trajectory parameters as measured by the sensor system. That is

$$H_0: W' \neq f(S,T),$$

and similarly for the other physical characteristics of the body

$$H_0: D' \neq f(S,T) \tag{1}$$

$$H_0: L' \neq f(S,T).$$

The alternate hypotheses are

$$H_1: W' = f(S,T),$$

and similarly

$$H_1: D' = f(S,T) \tag{2}$$

$$H_1: L' = f(S,T)$$

where W', D', L' = estimates of the true values,

S = sensor parameters,

T = trajectory parameters.

The null hypothesis H_0 is tested against the alternate hypothesis H_1 by deriving a model of W, D, and L as a function of S and T by regression analysis.

The general multivariate linear regression analysis is written

$$y' = a_0 + a_1x_1 + a_2x_2 + \ldots a_px_p\ , \tag{3}$$

where x_p = the p^{th} independent variable

a_o = the true intercept

a_p = the p^{th} true coefficient

y' = the regression estimate.

An analysis of actual range data is made to test the null hypotheses that an estimate of a physical characteristic is independent of sensor, trajectory, and body variables. The range data are analyzed with a computer program that calculates the regression of the dependent variables on the independent variables by a stepwise technique. The regression program analysis is a linear relationship, but it can be made to accommodate nonlinear functions by any one of 20 different transformations, such as logarithms. The analysis first calculates the simple correlation coefficients between each independent variable and the dependent variable. The variable with the highest correlation is selected for the first regression calculation. The linear regression of the form

$$y = a_0 + a_1 x_1 \tag{4}$$

is therefore calculated for one of the physical characteristics; say W, as y, and the independent variable with the highest simple correlation as x_1. Each of the remaining independent variables was then correlated with y and x_1. The variable (x_2) was then selected as the variable that produced the highest of these correlations. A second step regression was then calculated for the form

$$y = a_0 + a_1 x_1 + a_2 x_2 \ . \tag{5}$$

If the correlation of regression relationship should be reduced by the addition of another variable, this variable was removed. If, however, the correlation increased, the variable was retained and the step procedure is repeated for another variable up to the p^{th} variable and coefficient as shown in Eq. (3).

Currently, it is believed that the best body parameter for target identification is the weight of the reentry vehicle. Therefore, weight was selected for the initial effort. Quantitative measurements of parameters obtainable from the field sensor are V, $\dot{V}$, h, and σ. The operational problem requires that the prediction model be expressed in terms of the parameters measured by the sensor. The radar cross section is dependent on the ratio of the plasma frequency to the incident radar frequency. The plasma frequency is in turn dependent on the strength of the shock front and viscous forces. The viscous forces determine velocity, acceleration, and altitude changes as a function of time. The interaction and interdependency of these parameters (as well as others) determine the effects of the entire reentry environment perturbations on the magnitudes of these parameters as measured by the sensor, as well as their histories.

Hence, the change in the inertial force (F_A) of the body is caused by the drag force (F_D) acting on the body as it penetrates the earth's atmosphere. As an initial effort in the development of a prediction model, these forces were assumed proportional, neglecting gravity. That is

$$F_A \approx F_D. \tag{6}$$

This simple assumption is used as a basis to postulate an equation in which the constants and coefficients are assumed unknown or at least different from the Newtonian values. F_A and F_D can be expressed as

$$F_A = ma = \frac{W}{g}\dot{V} \tag{7}$$

$$F_D = \frac{1}{2}\rho V^2 C_D A. \tag{8}$$

Equating (7) and (8) and solving for W, it is found that

$$W = \frac{\rho V^2 C_D A\, g}{2\dot{V}}. \tag{9}$$

The independent variables of Eq. (9) are ρ, V, C_DA, and $\dot{V}$. Operationally the radar cannot provide estimates of ρ and C_DA directly. Therefore, these variables must be expressed in terms of measurements available from the radar. The density ρ can be expressed as

$$\rho = \rho_0 e^{-Bh}, \tag{10}$$

where

ρ_0 = standard density,

B = a constant,

h = the altitude.

Hence, density is expressed as a function of altitude, a variable which can be obtained from the radar. The remaining variable C_DA can be expressed as a function of the radar cross section (σ).

Bethe, Edwards and McDonald, and Martin have studied the functional relationship of σ and C_DA. The relationship developed is of the form

$$\sigma \approx K\ (C_D A)^N \tag{11}$$

where K = a constant,

N = the exponent.

Substituting Eq. (10) and (11) into Eq. (9) and rearranging terms, the following is obtained

$$W = \frac{\rho_0 g}{2(K)^{1/N}} \cdot \frac{V^2\ \sigma^{1/N}}{\dot{V}\ e^{Bh}} . \tag{12}$$

In the above expression, W is expressed as the function of a constant times the ratio of V^2 and $\sigma^{1/N}$ to V and h. Since the relationship is nonlinear, it must be linearized for the regression program. The equation is expressed as

$$W = a_0\ \frac{V^{a_1}\ \sigma^{a_3}}{\dot{V}^{a_2}\ e^{a_4 h}} \tag{13}$$

where a_0 = the regression constant

a_1, a_2, a_3, a_4 = regression coefficients.

Since Eq. (13) is nonlinear and the regression program is linear, then the equation must be linearized. Natural logarithms (which may not be ideal) were used to linearize the equation. It can be expressed as

$$\ln W = \ln a_0 + a_1 \ln V + a_2 \ln \dot{V} + a_3 \ln \sigma + a_4 h. \tag{14}$$

SELECTION AND REDUCTION OF REENTRY TEST DATA. Analysis of actual full-scale reentry test data requires that a historical sample of tests be selected. The selection of the tests requires the establishment of certain criteria.

The criterion for numerical analysis is that the test data be essentially complete throughout a prescribed trajectory range. That is, the radar must have maintained nearly continuous track. Also, it was imperative that each channel of track be accurately identified as to the type of information. (This is to avoid a mixing of the data sets.)

The trajectory criteria were established on the basis of the deposition of momentum energy into the disturbed medium through which the vehicle passes. The deposition of energy begins to be appreciable in the continum flow regime when the shock forms and viscous effects come into play on the body. The effects produced in this regime provide measurable variables which relate to the underlying physics of the interchange of momentum and energy between the body and its environment and the coupling of these phenomena with the sensor responses. If these phenomena are assumed to be typical, then it remains only to relate the amplitudes of the sensor returns to the levels of body parameters. For a typically sized reentry body at typical reentry velocity these energy interchange effects become pronounced at about 300 kft. However, the portion of the trajectory selected is from 150 kft. to 60 kft. in order to bracket a region of maximum kinetic and dynamic interchange for the sample of vehicles selected.

The body criteria were selected simply to obtain a sample with wide ranges of body characteristics such as weight, length, diameter, drag-area, and shape. One restraint placed on the initial selection of sample bodies is that all bodies be of the ablative type. The rationale behind this on the initial study was to have all bodies of the type which would at least unintentionally and somewhat randomly contaminate the flow field with typical reentry vehicle materials for data consistency.

One other constraint placed on the selection of the data sample was a constant radar frequency. Future analyses could relax this constraint and the data of several different frequencies could be used to develop a more universal model. In the operational mode bistatic measurements may be made. It would be desirable to have the frequency variable included. The prediction model could be adjusted for each particular discrete radar frequency used by the system.

After the criteria for selection of the sample were established, the data had to be actually selected and reduced to a usable format. The radar data tapes were located and presumably the proper information channels identified. A coupling program was developed so that the data could be directly machine fed from the tapes into the regression program. A printout of the smoothed values of the sensor measurements recorded on the tapes was programmed for a check on the data tapes input to the regression program. The data were taken from the data tapes at the appropriate time after lift-off, corresponding to the established trajectory altitude limits. The data were smoothed to obtain discrete values in 0.5-sec. intervals. (The intervals could be shortened to, say 0.25-sec. intervals.) The 0.5-sec. intervals provide an average of about 15 matched set data points for each reentry test selected. A total of 10 reentry bodies were included in the first analysis. The range of body characteristics selected is shown in Table I.

TABLE I

Body shapes: Simple spheres to complex sphere-cone-cyl-flare

Body diameters: 7.5 inches to 90 inches

Body lengths: 7.5 inches to 169 inches

Body weights: 17.5 pounds to 7181 pounds

ANALYSIS RESULTS. The reentry data sets as compiled were subjected to analyses. The values were programmed into the regression analysis on the computer, a total of 150 matched data sets. As mentioned above, the data sets are over the altitude regime of approximately 150 kft. to 60 kft. They represent about 7.3 sec. of reentry time.

These data were run on two different types of regression programs which computed the same values for the constants and coefficients. The general regression equation is

$$\ln W' = \ln a_0 + a_1 \ln V + a_2 \ln \dot{V} + a_3 \ln \sigma + a_4 h. \tag{15}$$

Each regression coefficient was statistically tested for significance.

Let α_1, α_2, α_3, and α_4 be the true values of the regression coefficients whose estimates are a_1, a_2, a_3, and a_4, respectively. The following hypotheses are tested:

$$\begin{aligned}
&H_0 : \alpha_1 = 0 \\
&H_1 : \alpha_1 > 0 \\
&H_0 : \alpha_2 = 0 \\
&H_1 : \alpha_2 > 0 \\
&H_0 : \alpha_3 = 0 \\
&H_1 : \alpha_3 > 0 \\
&H_0 : \alpha_4 = 0 \\
&H_1 : \alpha_4 > 0.
\end{aligned} \tag{16}$$

The t test used is as follows:

$$t_i = \frac{a_i - \alpha_i}{S_i} \tag{17}$$

where S_i = standard deviation of regression coefficient, a_i, i=1,2,3, and 4. The computed values for S are as follows:

$$a_1: S_1 = 1.218$$
$$a_2: S_2 = 0.0761$$
$$a_3: S_3 = 0.4234$$
$$a_4: S_4 = 0.0206.$$

The calculations of t are:

$$t_1 = \frac{a_1 - \alpha_1}{S_1} = \frac{8.741 - 0}{1.218} = 7.175 \tag{18}$$

$$t_2 = \frac{a_2 - \alpha_2}{S_2} = \frac{-0.08688 - 0}{0.0761} = -1.14 \tag{19}$$

$$t_3 = \frac{a_3 - \alpha_3}{S_3} = \frac{0.2640 - 0}{0.4234} = 0.6236 \tag{20}$$

$$t_4 = \frac{a_4 - \alpha_4}{S_4} = \frac{-0.1414 - 0}{0.0206} = -6.842 \tag{21}$$

For a 95 percent confidence level ($\alpha = 0.05$), the critical value for t is $|1.960|$. Therefore

$t_1 = 7.175 > |1.960|$, Reject H_0,
$t_2 = -1.141 > |1.960|$, Accept H_0,
$t_3 = 0.6236 > |1.960|$, Accept H_0,
$t_4 = -6.842 > |1.960|$, Reject H_0.

Therefore, the regression coefficients a_2 and a_3 are not significant. The regression equation is now recalculated as

$$\ln W' = a_0 + a_1 \ln V + a_4 h, \tag{22}$$

which expresses the weight of a reentry body in terms of the velocity and altitude as determined from the radar sensor. The regression equation was calculated to be as shown in Eq. (23). The actual values are not shown for security reasons.

$$\ln W' = \ln a_0 + a_1 \ln V - a_4 h, \tag{23}$$

The correlation coefficient is

$$\nu = 0.602$$

The final equation (23) was used to calculate point-by-point estimates of 13 independent reentry object weights. The range of characteristics of these objects is shown in Table II.

TABLE II

Shape: Simple sphere to sphere-cone-cylinder-flare

Diameters: 4 inches to 40 inches

Lengths: 12 inches to 167 inches

Weights: 7.5 pounds to 3,390 pounds

The calculated values are shown in point plots of estimated weight versus altitude simulating real time estimates of object weight. These plots are Figures 2 through 14.

A composite plot of all thirteen bodies is shown in Figure 15. This is a semi-log plot of the best estimate of weight versus altitude which comes out of a reasonably straight line which is expected in view of the transformation of the data to fit the hypothesized equation.

You have noticed that the plots show positive or negative slopes indicating increasing or decreasing weight estimates as a function of altitutde (for example, time). Only two plots have indicated both positive and negative slopes where the true weight was estimated twice during the time-frame of calculation. The desirable shape of the real time plots of individual objects is shown in Figure 16. It would be desirable to have the estimate converge to an asymptote to the true weight within some established confidence limits. All bodies displaying these curve forms could be classified as decoy or RV. Those outside the confidence limits would be engaged.

DISCUSSION. The results of the regression analysis computations indicate that the two variables, acceleration and radar cross section, do not significantly relate to the body dimension weight. This is not to say that they are not significant, but rather that with the data sample used they could not be established as significant. There are reasons which could account for the failure to establish significance of these two variables. One reason could be the poor distribution of these variables in the sample of data. The poor distribution could be due to the error of estimating these quantities by the radar sensors. Considerable error could be contained in the estimates of the negative acceleration of the body because of the azimuth and elevation rate changes of the antenna caused by shifts in tracking the electromagnetic centroid of the reentry complex. The error contained in the radar cross section is possibly caused by the inherent error in the C_DA radar cross section relationship used in the development of the hypothesized equation.

Another reason is that the 95 percent confidence level may be too high for the degree of precision in making the measurements. A further stratification of the data could be made that would provide a range of more consistent variation in the acceleration and radar cross section readings. However, this would be useful only for study purposes and would not improve the inherent inaccuracy of the radar system estimates in the real operational case. The improvement in the accuracy of the values would establish their significance and raise the present correlation coefficient of 0.602. Weight estimates of the reentry vehicle would be more accurate with an improved correlation coefficient.

ILLUSTRATIONS

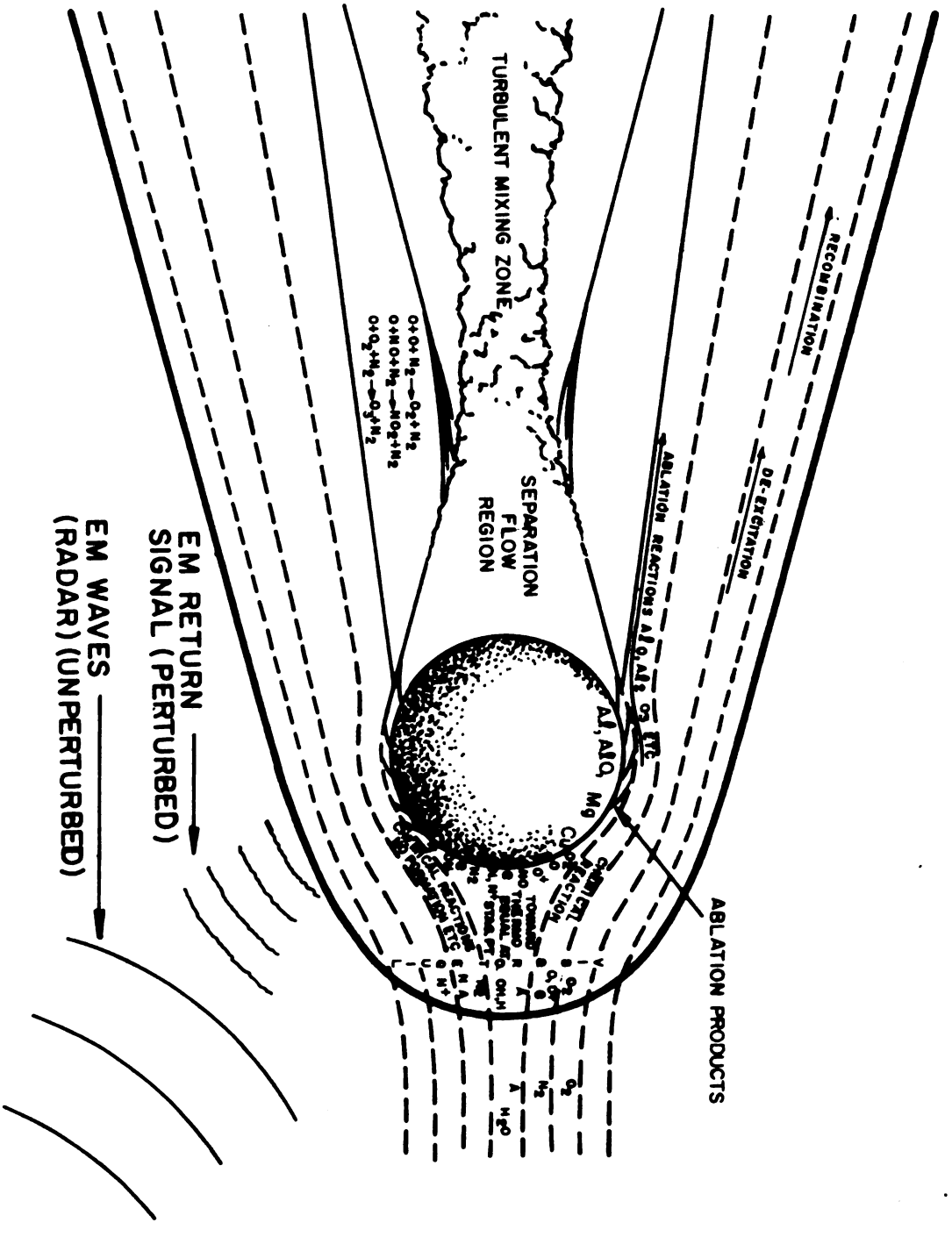


FIGURE 1

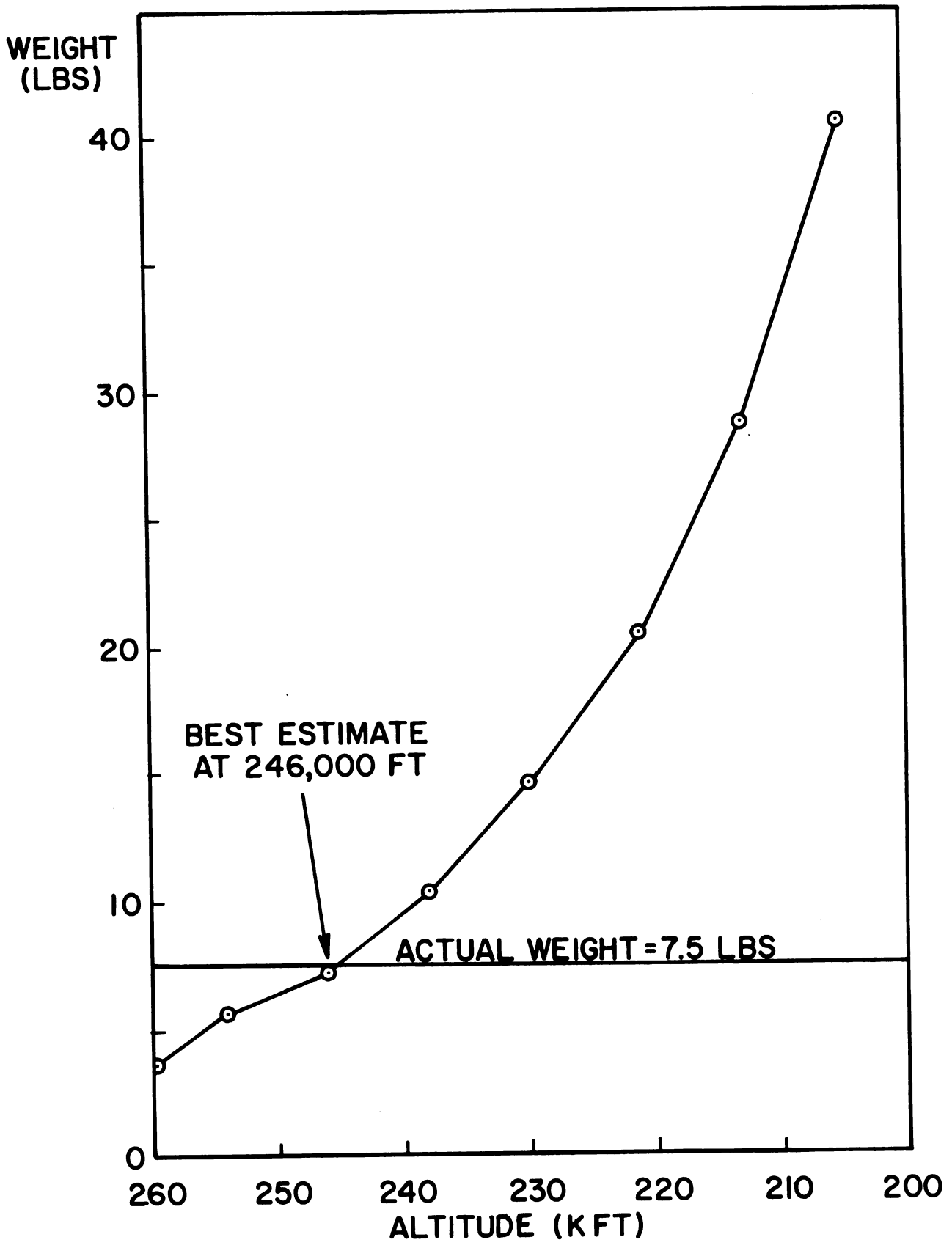


FIGURE 2

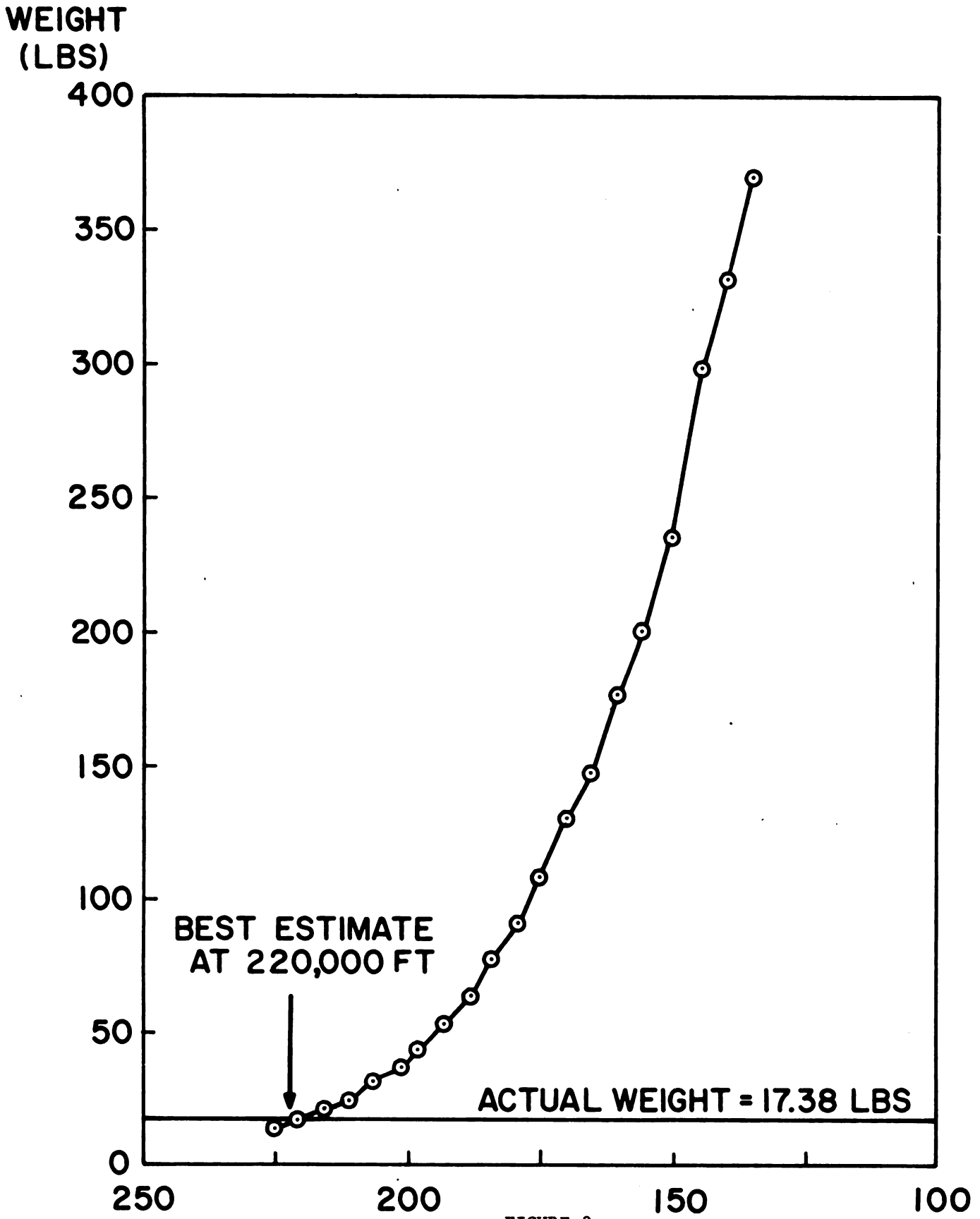


FIGURE 3

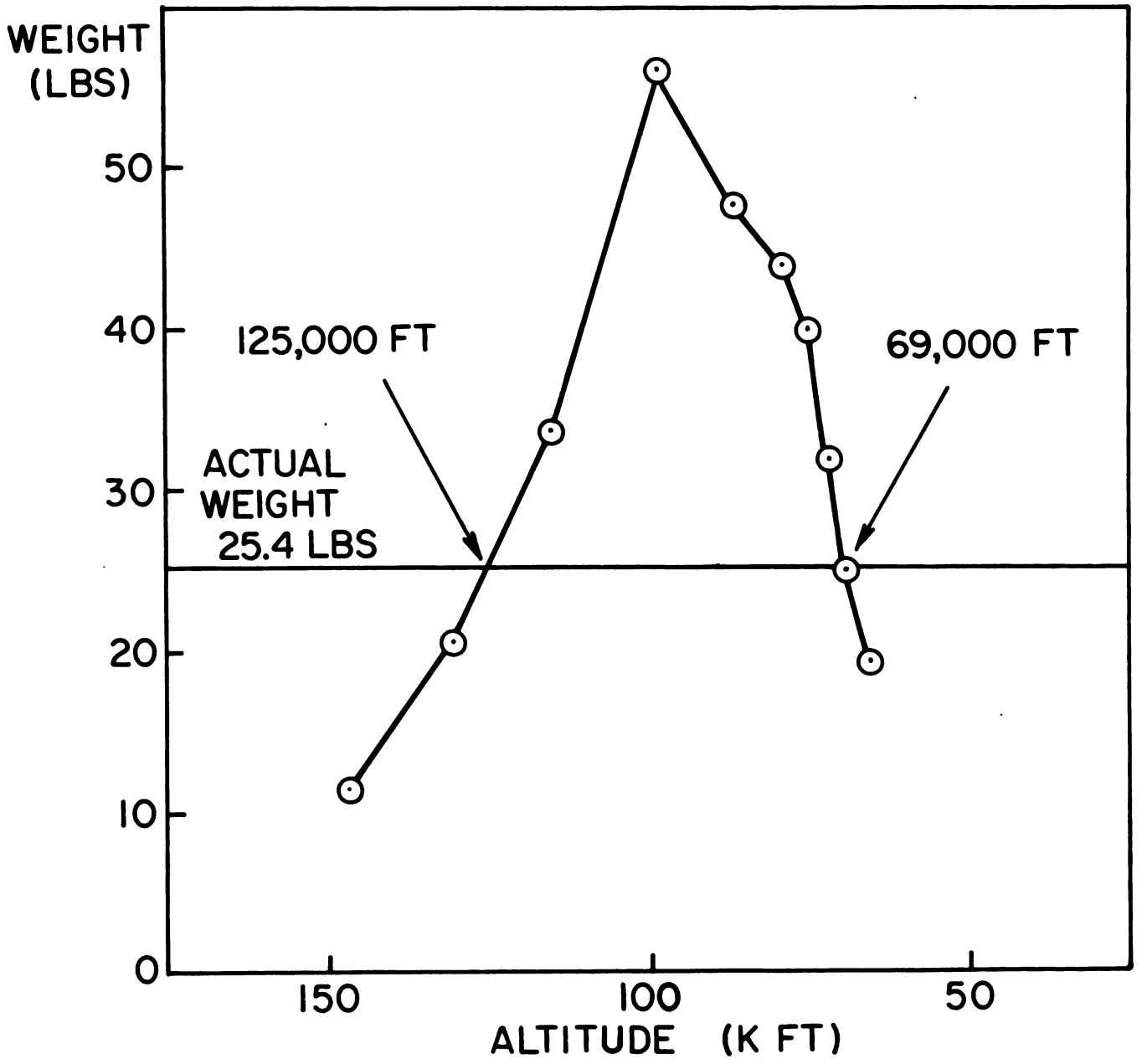


FIGURE 4

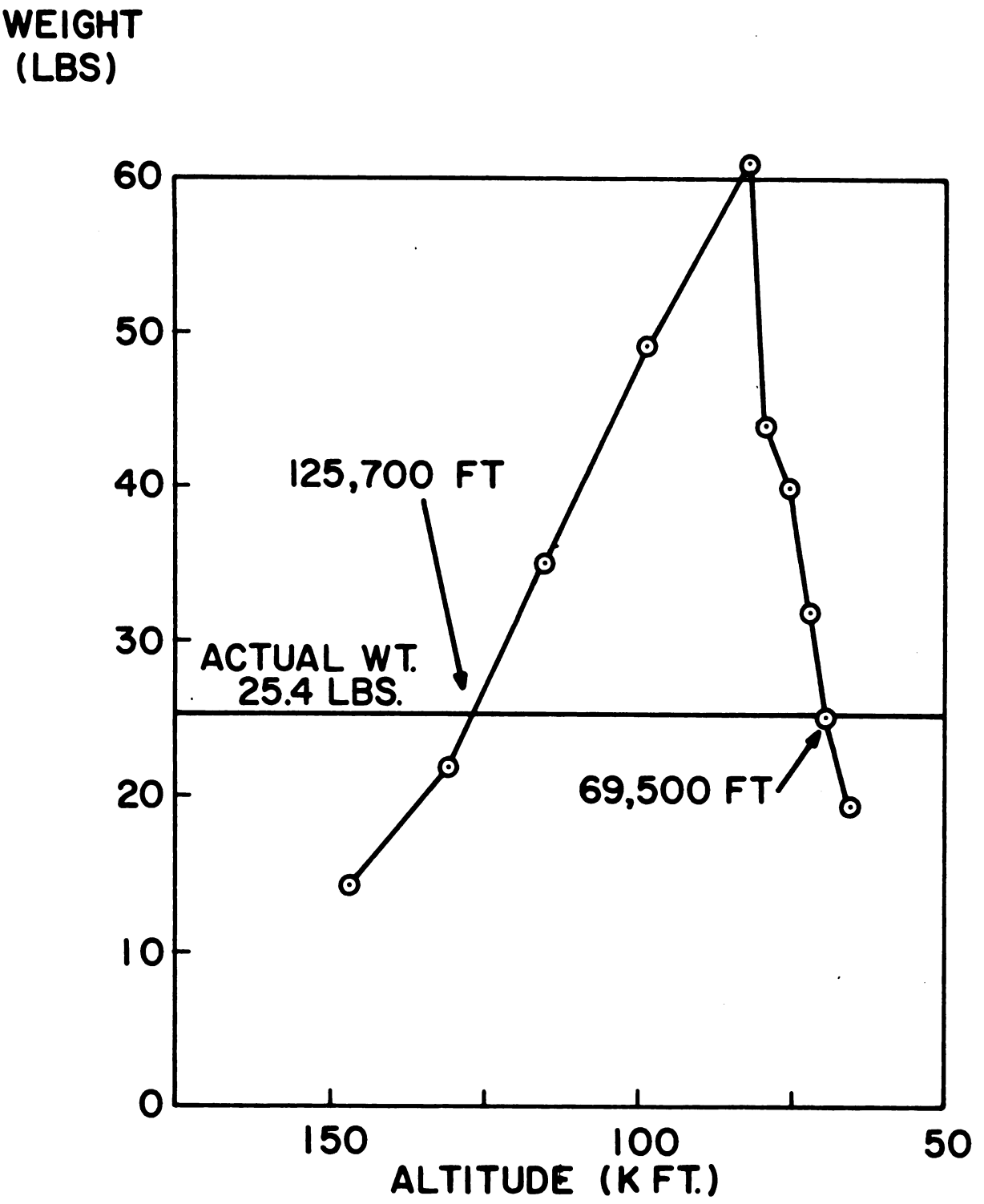


FIGURE 5

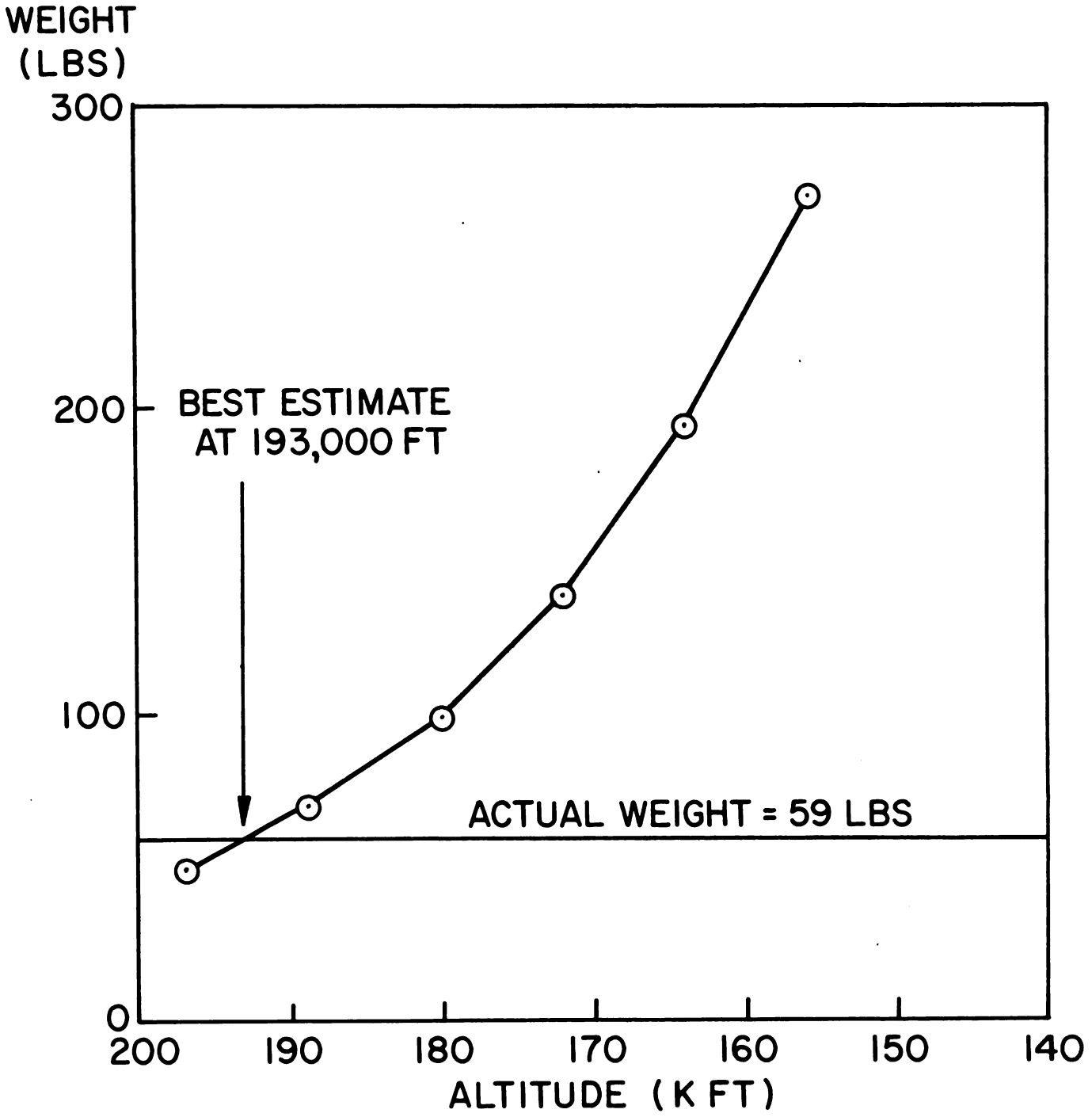


FIGURE 6

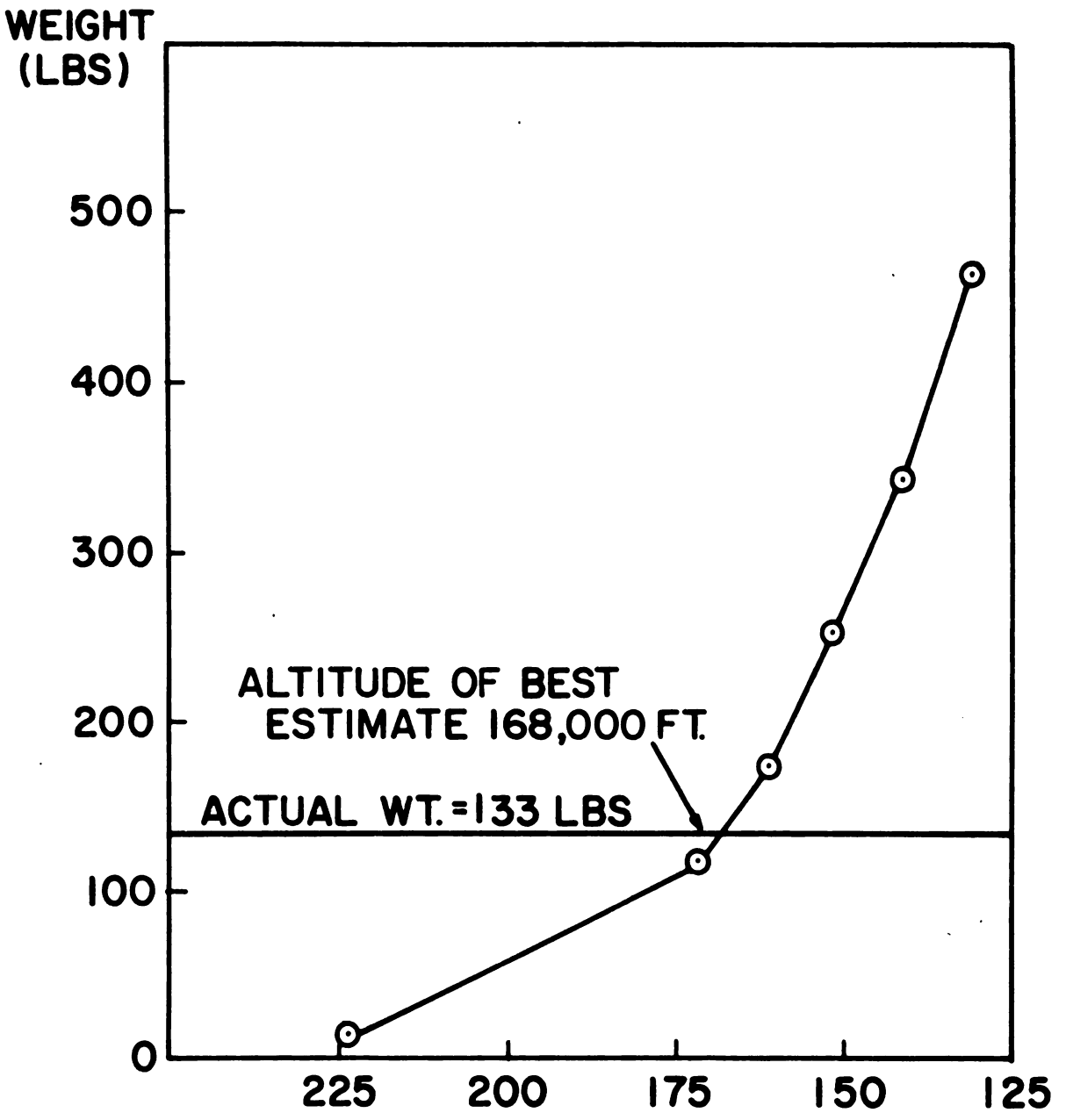


FIGURE 7

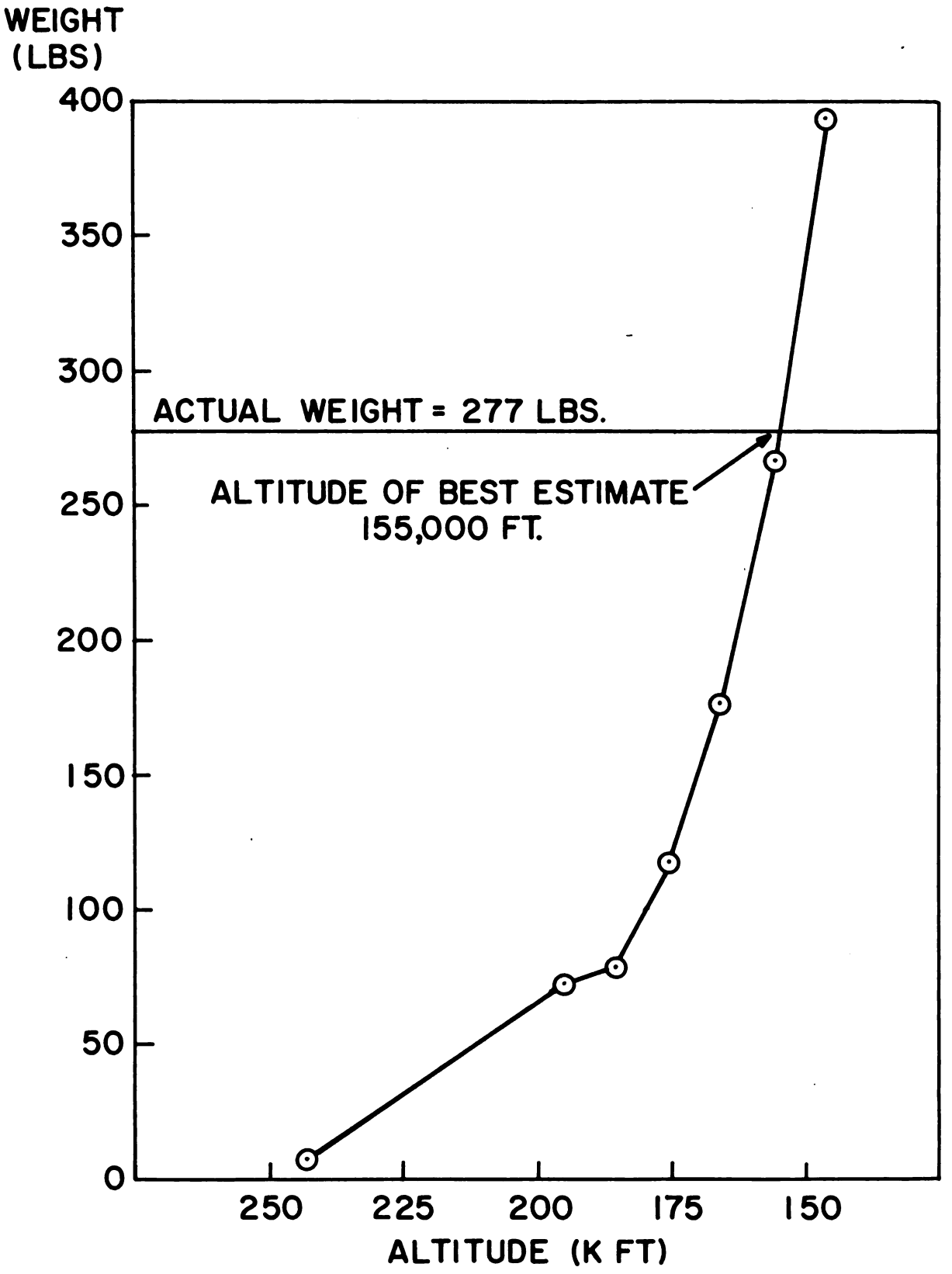


FIGURE 8

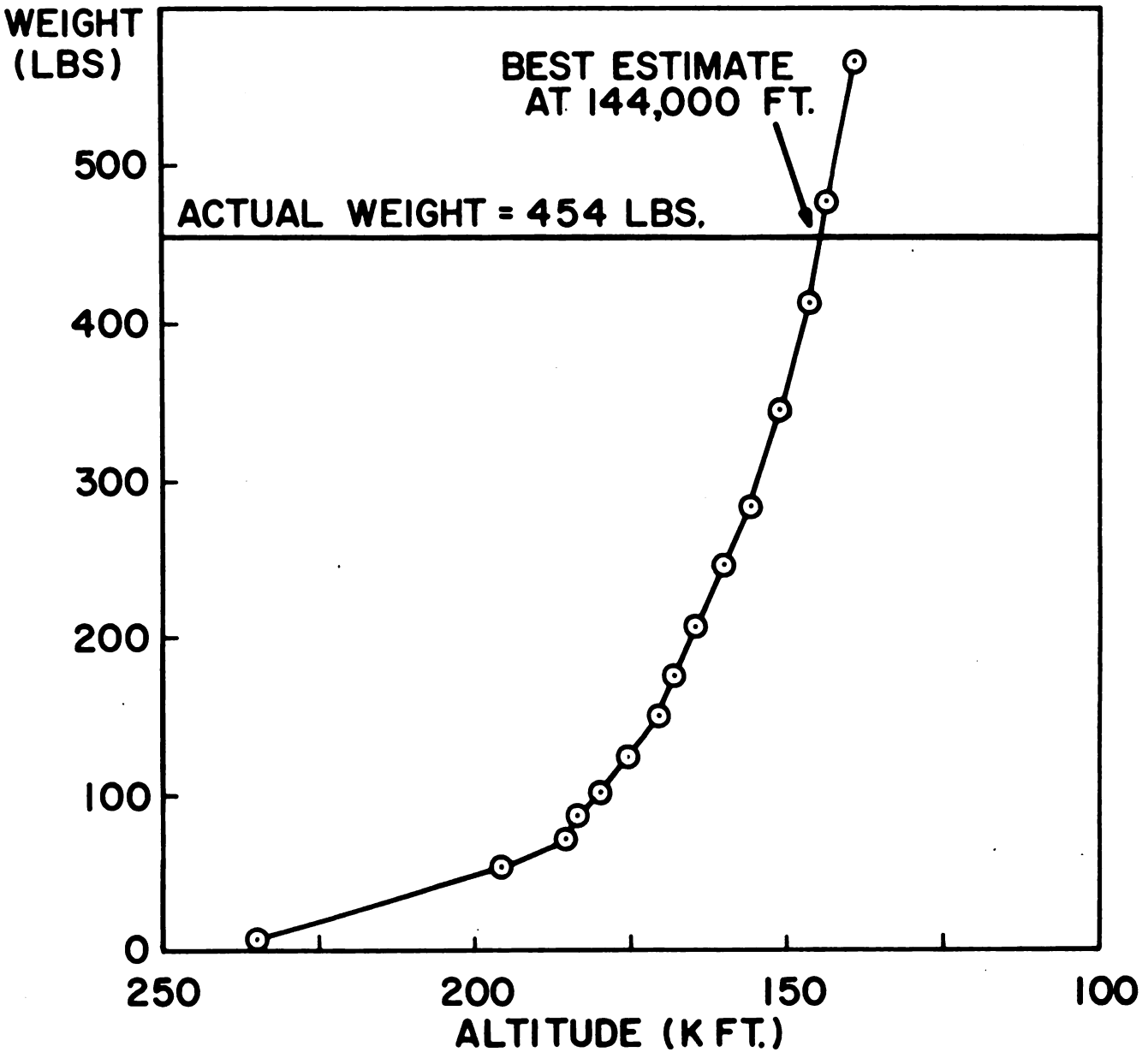


FIGURE 9

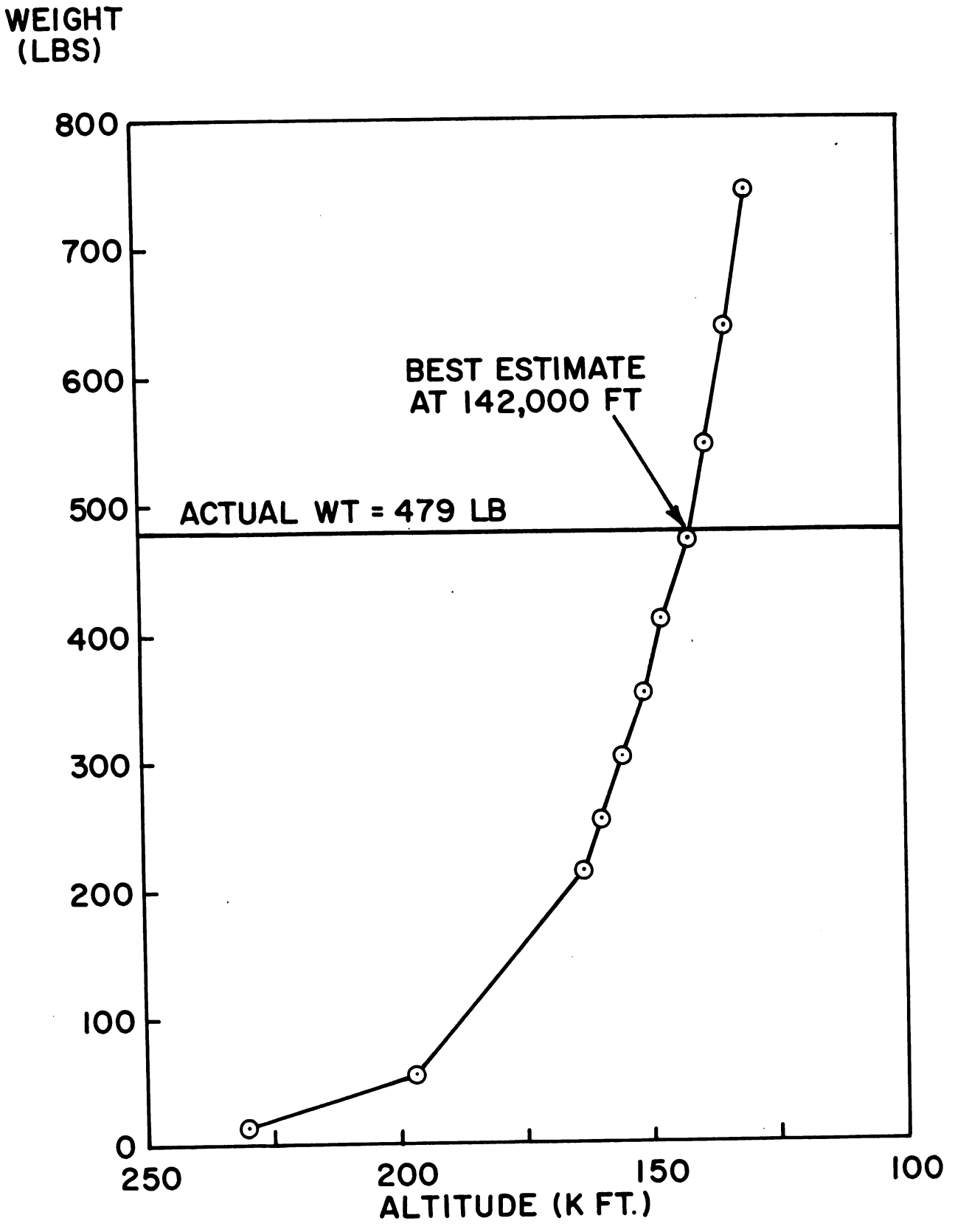


FIGURE 10

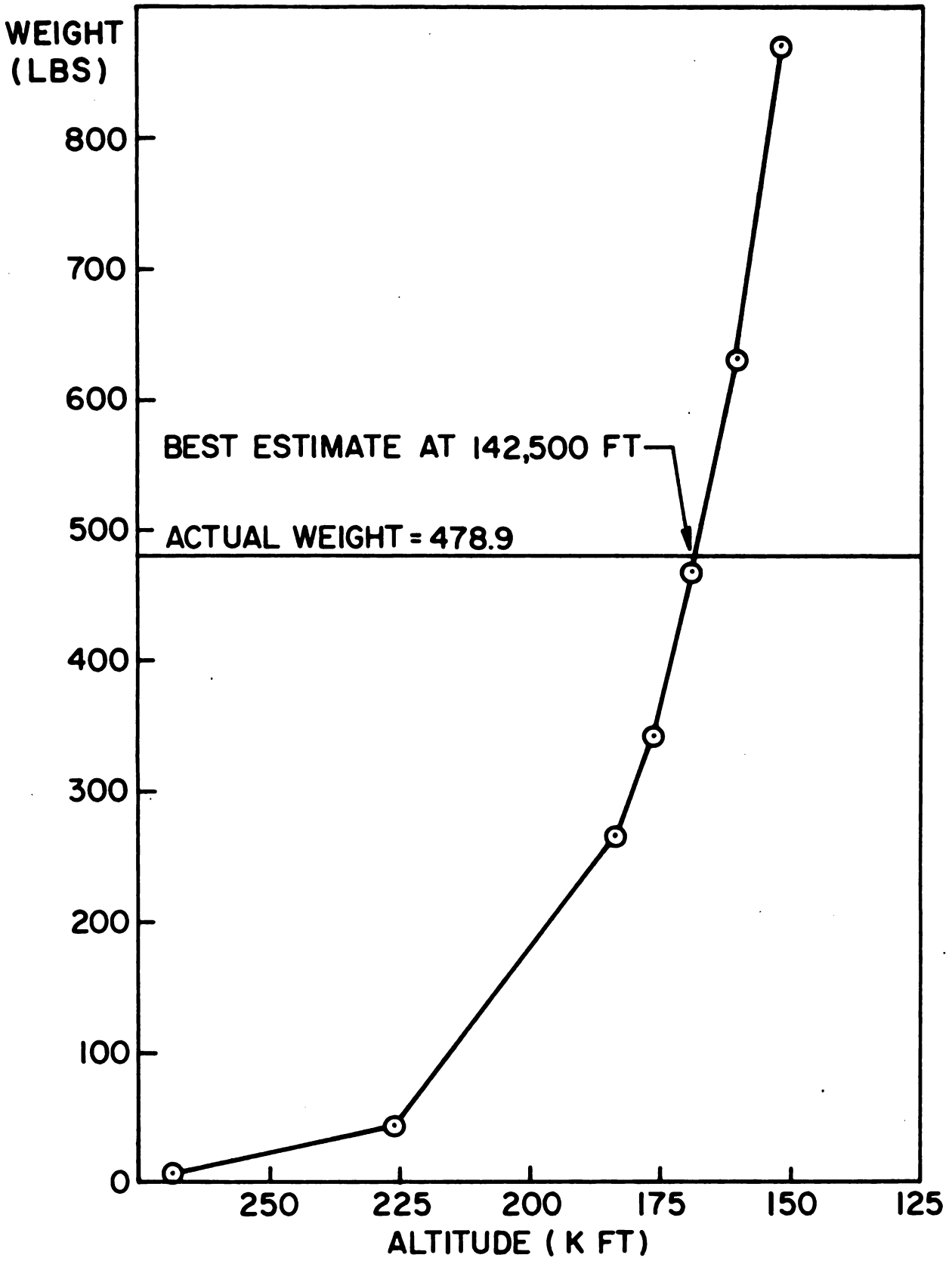


FIGURE 11

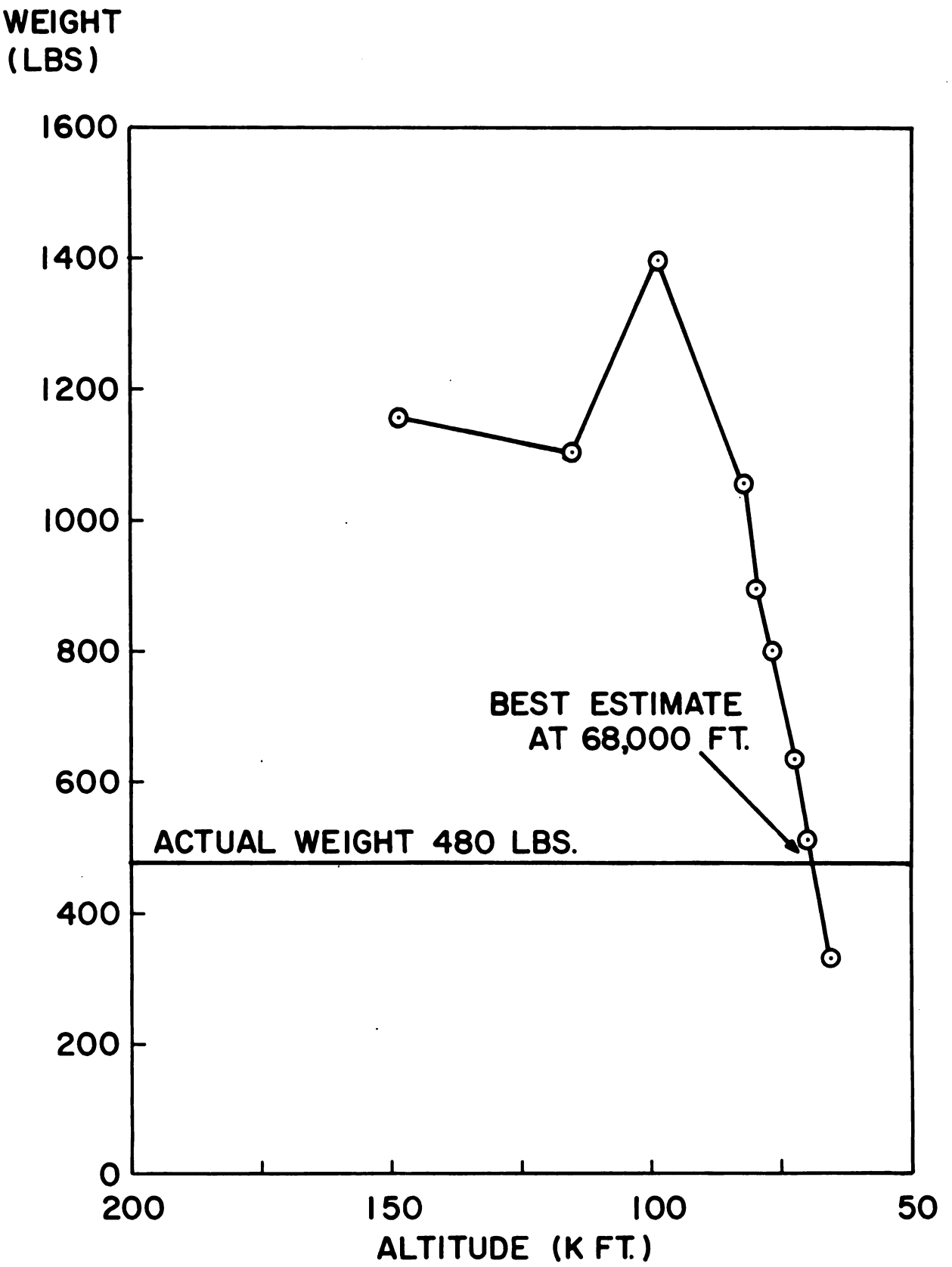


FIGURE 12

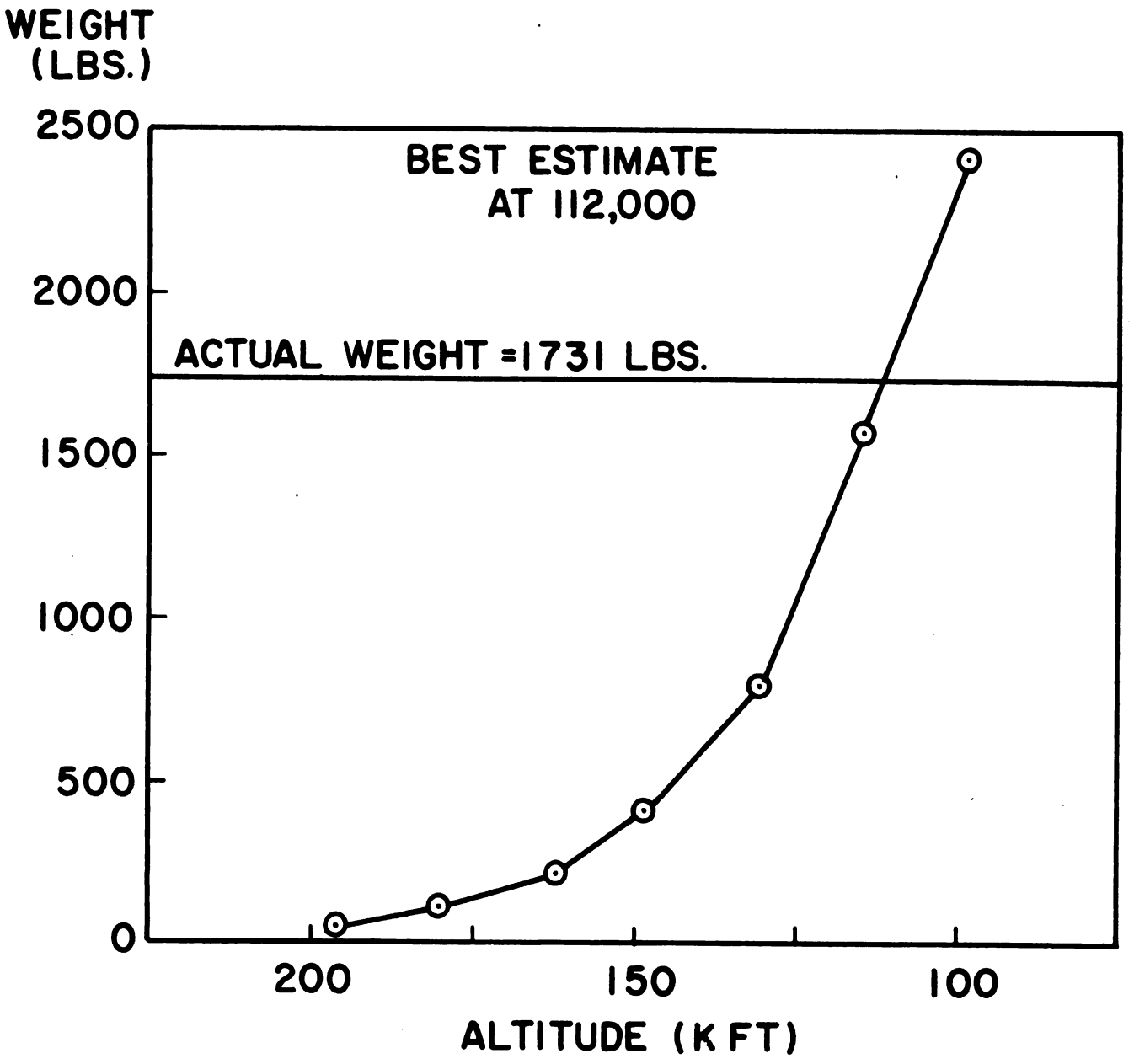


FIGURE 13

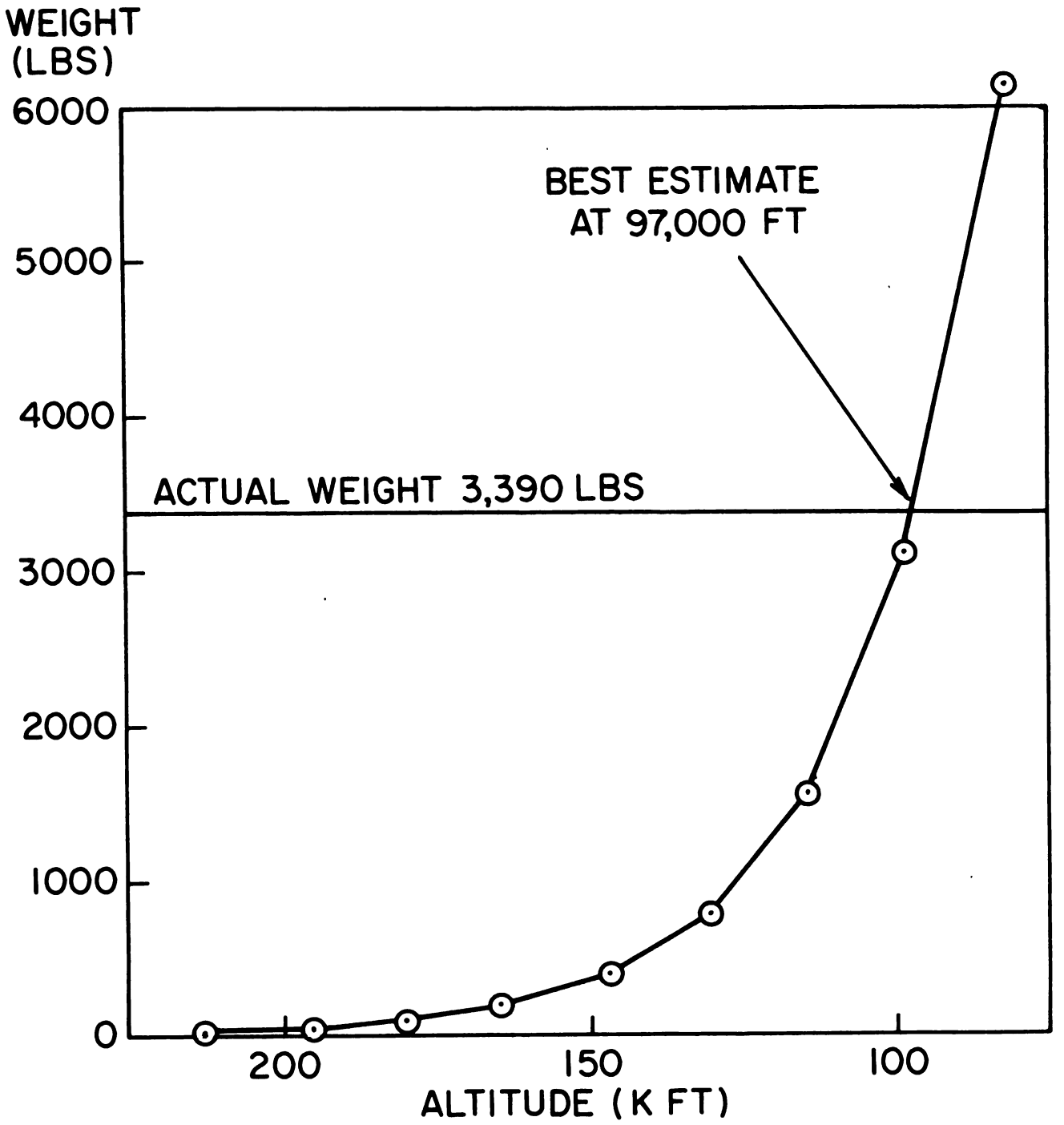


FIGURE 14

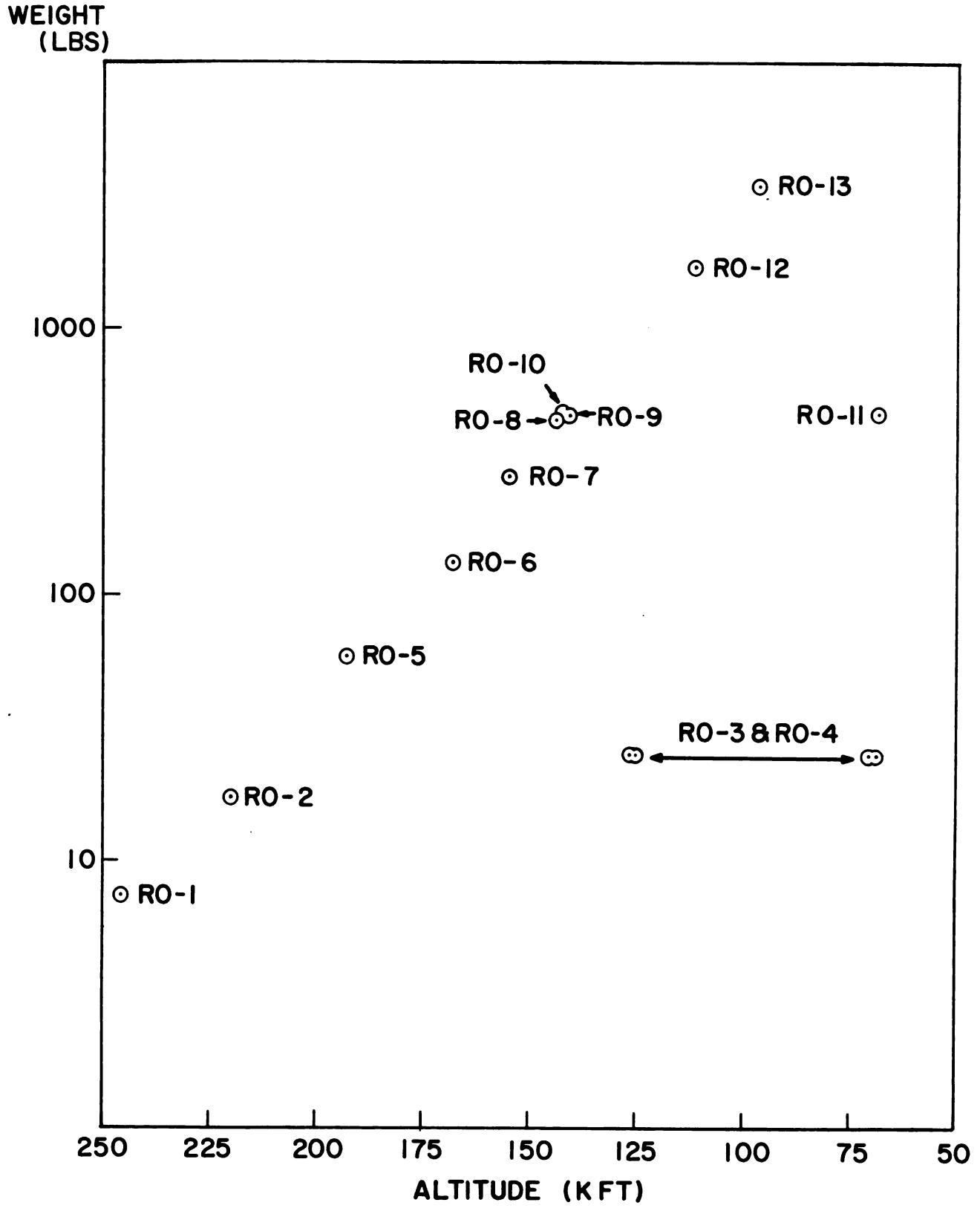


FIGURE 15

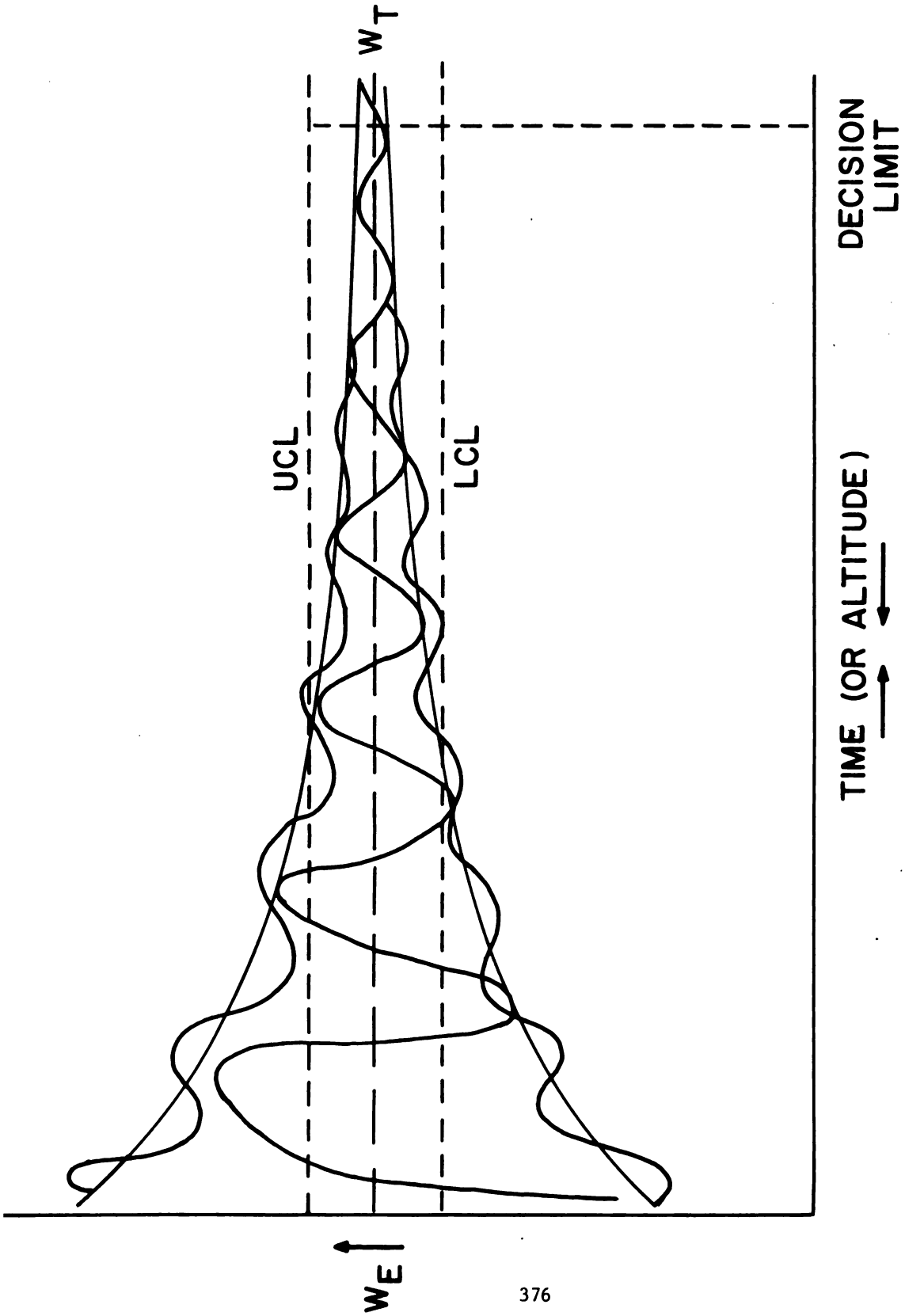


FIGURE 16

SOME EXPERIENCES IN LABORATORY CONTROL INVESTIGATION*

Sigmund P. Zobel
Cornell Aeronautical Laboratory, Inc.
Buffalo, New York

1. INTRODUCTION. Statistical quality control as a way of life in American, Canadian, and British industry is over twenty years of age. Further, as a logical extension of the scope of applied statistics in industry, statistically designed experiments and analysis of variance have been increasingly used over, perhaps, the last fifteen years. One phase of industrial operations, however, still appears to be rather slow in making use of the applications of statistical methods to the analysis and control of its routine activities. I refer to the typical chemical laboratory, wherein there are frequently many ways in which the operation can be made more efficient in terms of precision and accuracy and overall reliability of analysis. Much of the potential improvement in effectiveness can be delineated, achieved, and preserved through the use of statistical experimental and control techniques.

In the past few years, my colleagues and I have had the opportunity to participate in the study of one such laboratory, whose director realized he had many problems connected with the achievement of improved precision and accuracy in the analytical results, and was quite cooperative in permitting experiments designed to shed light on the specific problem areas, so that control measures may be instituted and/or procedures changed. Since more or less standard techniques of analysis were used, describing some of our experiences in that laboratory may be of possible use to others confronted with similar situations. Objectives, means of accomplishment, and conclusions will be discussed to the exclusion of technical details of computations, or theory.

Let me describe the setting for you. A wet chemistry laboratory routinely analyzes samples resulting from a particular event. Several hundred samples, over a wide concentration range, are delivered to the laboratory for analysis. The analysis is performed colorimetrically, and the goal is to optimize the precision and accuracy of the results. There are two different colorimetric instruments, with two units of each, in service. One instrument is mainly automatic in its sample preparation and analysis; the other is largely manual. There are four analysts available, and any combination or all four may be assigned to the job when it comes into the laboratory. The number of samples which may result from the event is usually larger than can be handled by all four analysts during a regular 8 hour shift. When the budget permits,

*This work was performed under Contract No. DA 18-035-AMC-280(A), Field Evaluation Division, Technical Support Directorate, U. S. Army Edgewood Arsenal, Maryland.

overtime is used to complete the analyses within one calendar day. Otherwise, the work may require two, or even three calendar days. Since the reagents and standard solutions have varying degrees of perishability, delays may be deleterious to the yield estimated from the analysis.

With this background in mind, the need will be rather obvious for the various experiments to be discussed in this paper. The first two or three will be presented in greater detail than the latter illustrations.

2. ANALYSIS OF ALIQUOT VOLUMES. One factor contributing to bias and variation in results was believed to be lack of uniformity in the aliquot volumes in the test tubes containing the samples. To obtain the aliquot volumes, the samples are originally collected in larger vessels, an amount in excess of that prescribed is poured into a test tube, and a suction apparatus is used to draw off excess liquid to a purportedly reproducible level. The test tubes are in racks which hold 40 tubes in a 4 x 10 rectangular array. Since many racks are used for the analysis of a given event, it was suspected that rack to rack variation* and tube to tube variation may be to blame for some bias and variation in the analytical results.

A components of variance model approach was selected, since the interest lay in estimating variances.

2.1 Estimation of Rack Variance.

For the estimation of rack variance, ten test tube racks were randomly selected. Similarly, forty test tubes were randomly selected and tared. The test then proceded as follows:

The forty test tubes were placed in one rack, filled and drawn down to volume. The tubes and their contents were weighed, and the total weight of liquid determined by subtracting the tare weight of the empty test tubes. The test tubes were then returned to the same rack, refilled, and again drawn down to volume. Reweighing of the test tubes and correction for tare weight then provided a duplicate weight determination for the given rack. Upon repeating the above procedure for each of the ten racks, one obtained ten pairs of determinations. The measure of variance provided by variations among the means of the ten pairs includes rack variance as well as other random effects. On the otherhand, differences between duplicate determinations made in the same rack provided a variance estimate from which the rack combination was eliminated. In this way, it was possible to isolate the rack component and compare its magnitude with

*Since the suction apparatus is applied to tubes positioned in a rack, variations in racks due to nonuniform depths of tube bottom recesses may contribute to nonuniformity in residual volumes.

the magnitude of the residual variance. Weight, of course, is being used as a proxy for volume. Table I contains the results of the analysis of variance.

ANALYSIS OF VARIANCE OF CODED DATA

Source of Variance	Sum of Squares	Degrees of Freedom	Mean Squares	Expected Mean Squares
Between Racks	18,122.8	9	2013.6	$\sigma^2 + 2\,\sigma_R^2$
Within Racks	524.0	10	52.4	σ^2

TABLE I

From the above, one can solve for σ_R^2, since σ^2 is given as 52.4. σ_R^2 proves to be 980.6. Additional calculations provided the interesting result that the coefficient of variation representing the total variance for a random determination on a random test tube in a random rack was 11.4%, while if determinations are constrained to the same rack, or if rack variance is eliminated, the coefficient of variation could be reduced to 2.6%.

The data also permitted an analysis to be made of error contributions by the individual racks. Noting that the error contribution from any rack will appear as a bias for all test tubes within that rack, the rack bias may be estimated by examination of the mean weights of the 40 aliquots in each rack. Table II shows some interesting data.

One way to eliminate the error contributions from a rack is to establish a correction factor, as function of the bias, for each rack. A second and immediately applicable method would consist of isolating the most heavily biased racks, such as those starred in Table II, and either retiring them from service or making some physical adjustment to eliminate the bias. The actual outcome was an even better corrective measure. When the laboratory management was made cognizant of the facts, it obtained a specially made rack, and required that all tubes were to be drawn down to volume only in that rack, although this necessitated one extra handling step.

RACK BIASES

Rack Number	Deviation from Mean Weight for all Racks
1	+ 4.9
2	+ 1.9
3	+ 20.4*
4	- 2.6
5	+ 2.9
6	+ 14.4*
7	- 85.6*
8	+ 31.4*
9	+ 4.4
10	+ 7.9

*A Significant Deviation

TABLE II

2.2 Estimation of Test Tube Variance.

For the estimation of test tube contribution to the variance of aliquot weights, forty test tubes were randomly selected, tared, and placed in a particular rack. These test tubes were then filled, drawn down to volume, and weighed individually. Correction for the tare thus yielded an estimate of the volumes of 40 randomly chosen test tubes. The same test tubes were then replaced in the same rack, refilled, and the process repeated. Thus, a pair of determinations was obtained for each of the forty test tubes in the rack. The measure of variance provided by variation among the means of the forty pairs includes test tube variability and lack of reproducibility of the suction device, but does not include variance introduced by racks since only one rack was employed. On the other hand, differences between duplicate determinations made on the same test tube provides an estimate of the variance attributable only to lack of reproducibility of the suction process.

An analysis of variance similar to that discussed above was performed to investigate test tube effects. The between test tubes mean square was significant compared to the within test tubes mean square. Now, the within racks variance found earlier is another independent estimate of the between tubes variance. That the two such estimates are in excellent agreement may be seen from Table III, which shows the pertinent coefficients of variation. In addition, the measure reflecting the degree of reproducibility of the suction process is included, as is the rack to rack measure.

COEFFICIENTS OF VARIATION

Based on Rack to Rack	11.4%
Based on Within Racks	2.6%
Based on Between Test Tubes	2.4%
Based on Within Test Tubes	0.4%

TABLE III

Table III affords a summary of the two experimental studies. Obviously, the rack to rack variation is the largest. As has been noted above, however, appropriate corrective action was taken to eliminate this effect. The fact that the two independent estimates of the tube to tube variation are so close to each other clearly establishes this as a real source of variation in volumes. Further investigation revealed differences in tube diameters and bottom curvatures. These are standard laboratory supplies, however, and could not be ordinarily obtained at better quality levels. But, a policy was established to request that new orders be filled from one production lot whenever possible, and some inspection procedure was to be set up to examine receipts of new tubes. The data also reveal that, since the within test tubes coefficient of variation is so small, and it reflects the filling reproducibility of the suction device, there is probably no problem on that account.

3. STUDY OF COLORIMETERS AND ANALYSTS. In the study of any operation, for the purpose of enhancing its effectiveness, all sensitive phases must be considered. In the preceding section, the drawing down to volume step was examined, and placed under a better state of control. This section will be directed at consideration of the equipments and operators.

As noted earlier, there are two instruments of each of two types, and four analysts. Thus, an analysis or group of analyses may be performed on any one of four instruments. Unless all four possess the same intrinsic properties of variation, color perception, translation of color perception to signal output, etc., each machine represents a different analytical system, and hence, as its output, produces results which may not be completely comparable to the outputs of the others. That is, a set of results may be high or low, more dispersed or less, as a consequence of the particular instrument which generated it. Such a situation dilutes the effort of a laboratory, since, on the one hand real differences between batches may be masked by the analytical system, or, at the other extreme, minor differences might be exaggerated by the system.

Much the same might be said with regard to the several analysts who share the bench work responsibilities.

Because of these possibilities, the four instruments and the four analysts were studied through a 3-factor experiment designed to yield information on accuracy and precision of each instrument and analyst. The third factor was concentration level, since this also was suspected of contributing to bias and loss of precision. The data output was combined with data obtained earlier in another connection.

3.1 Drift and Bias.

Earlier in the study program, the presence of instrument drift was suspected, and reaffirmed in subsequent data analyses. Accordingly, one output of the experiment was a set of data deliberately designed to provide evidence pertaining to instrument drift. This was accomplished by comparing results from each of three concentrations which were used. Comparisons were made between first and second members of pairs of consecutive samples of the same concentration, and early in the run and late in the run analysis.

The first comparison type revealed clearly that there is a carry-over effect in the automatic analyzer. Rather conclusive evidence showed that if two samples of the same concentration followed a higher concentration, the first of the two showed a higher concentration than the second. If two samples of the same concentration followed a lower concentration, the first of the two showed a lower concentration than the second. And, when two samples of the same concentration were first in a series of unknowns, the first was lower than the second. Thus, it may be concluded that there is indeed a carry-over effect.

The same study showed that the manual instruments did not exhibit a significant drift, except in one instance of an apparent interaction between one of the analysts and one instrument. Since the samples presented to the manual device are each in its own test tube, this is further confirmation of the possibility that the common test cell in the automatic is not sufficiently purged before entry of the next sample. As still further evidence to support this thesis, it should be noted that all four analysts' work showed the upward drift in the calibration groups on automatic No. 2., and all but one did so on the automatic No. 1. On manual No. 1, all four analysts had little or no drift indications, while on manual No. 2 three of the four did so. It may also be noteworthy that one analyst was the exception in each of the two cases cited. The difference between his performance and that of the others will also be evident below.

Analysis for instrument bias revealed that the two manuals and one of the automatics had positive bias, while the other automatic had a negative bias. If only one analyst had made all determinations, on only one concentration, the above conclusion would be relatively firm. However, since several concentrations were used, interactions may have influenced the results. That is, the amount and direction

of bias on each instrument may vary with the concentrations, or with the analyst. The instrument-concentration interaction is shown in Table IV.

AVERAGE BIAS BY INSTRUMENT-CONCENTRATION COMBINATIONS

	Concentration Level		
Instrument	850	2540	5070
Automatic 1	+ 48	+ 100	- 51
Automatic 2	+ 48	+ 15	- 114
Manual 1	+ 94	+ 48	- 18
Manual 2	+ 85	+ 62	+ 20
All	+ 69	+ 56	- 27
Bias as % of Concentration	8%	2%	- 5%

TABLE IV

It can readily be seen that, with but one exception, bias is an inverse function of level, so that the low concentration has the highest bias and the high concentration has either a negative bias or the least positive bias for a given instrument. Stating it differently, lower concentrations tend to be measured higher than actual levels, while higher concentrations tend to measure lower than actual values. However, the order of magnitude of the bias does not appear to be any clear function of the particular instruments.

The ultimate breakdown of bias is according to instrument, concentration, and analyst. Since this three factor interaction was also statistically significant, further insight can be gained by examination of the three factor bias components, as in Table V.

AVERAGE BIAS BY ANALYST-INSTRUMENT-CONCENTRATION COMBINATIONS

Concentration	Analyst	Instrument M_1	M_2	A_1	A_2
850	1	123	167	97	103
	2	70	70	37	17
	3	130	75	25	22
	4	30	60	40	30
2540	1	140	113	153	117
	2	73	60	107	93
	3	- 07	- 07	170	120
	4	- 13	80	- 30	- 83
5070	1	57	- 27	-133	70
	2	- 20	93	93	-293
	3	-155	- 11	82	- 45
	4	83	117	- 90	-157

TABLE V

A quick confirmation of the concentration effect noted above may be obtained from the first appearance of the negative biases in the middle concentration, and the greater number of negatives in the high concentration.

More importantly, however, Table V provides an entree' for drawing inferences on analyst bias. For example, there are 12 instrument concentration combinations. In 9 of these, analyst 1 has an extreme amount of bias, and in 2 others is close to an extreme. Thus, he is outlying, or out of line, in 11 of 12 possible cases. Similarly, analyst 3 is out of line in 7 cases, followed by analyst 4 who is out of line 5 times, and analyst 2, with 3 times. Further, from the original data analyst 1 has an over-all bias of +82, analyst 3 has +25, analyst 2 has +18, and analyst 4, +06. Analyst 1 is the same one who was the exception to the general pattern of drift shown by the analysts, as discussed earlier.

While analyst 1 is almost consistently high, having the largest positive bias on both manuals and automatic 2, the others were quite inconsistent. Analyst 2 has the largest negative bias, under analyzing on automatic 2, although his biases on the other three instruments are positive; analyst 3 has negative bias on both manuals, and positive on both automatics. Analyst 4 shows a bias pattern just opposite so that of analyst 3.

3.2 Precision.

A similarly detailed analysis was made of precision. In the interests of brevity only the findings are presented herein. Analyst 1 stood out again, this time for having the largest variances. The other three analysts showed some inconsistencies, but not approaching the degree indicated by No. 1. In terms of the instruments, the two manual instruments had higher variances than the two automatics, but not significantly so. An important finding, and one that corroborated early suspicions, was that variance increased with concentration levels. Indeed, a good linear fit was obtained by taking the logarithm of the variance as a function of the logarithm of the concentration level.

4. STABILITY OF STANDARDS. By now, the overall program has indicated the need for control over supplies, instrument operating characteristics, and efficiency of the individual analysts. One important phase of the operation which had not yet been examined was the stability of standard solutions used for obtaining calibration curves and checks on machine drift.

Happily, a large test run was in the offing, and we were able to design an experiment using several standard solution concentrations. Samples were inserted as blind samples in each rack of 40 tubes of ostensible production samples. The quantity of samples, coupled with no overtime allowed, dictated that three working days be required to perform the chemical analyses. Since by now the laboratory management was convinced of the upward drift trait of the automatic colorimeters, they decided to use only the manual instruments for this particular job. Hence, one variable was eliminated from the areas of concern.

The results of the analysis were again in part confirmatory of other findings, and in part substantively directed at support of an important hypothesis. Once more, for the third or fourth independent case, variance was found to be related to concentration level. But more importantly, the data showed that over a three day period, concentrations were not stable, definite and significant losses in levels were determined on the second day as compared to the first day, and the third day compared to the second. This effect was present independently of instrument or analyst. In the course of any one day, there was some hint of deterioration from start to end, but it was not sufficiently clear cut and persistent to permit a positive assertion.

5. CONCLUSION. I would like to conclude this paper by summarizing the various applications of statistical methods which we have found useful in laboratory control investigations, and the kinds of answers that were obtained. And, since this audience has its primary interests in the applications of statistical quality control to laboratory problem areas, we may note the implied laboratory controls which were recommended to management in this particular case.

5.1 Summary.

It is obvious from my remarks in Sections 2-4 that designed experiments and analysis of variance are well suited to the study of laboratory operations for the purpose of pinpointing problem areas. Less obvious are several other techniques which were used to good advantage in this particular program.

Significance tests on means and variances helped to evaluate the merits of an analyst versus another, or one instrument versus another. Nonparametric tests for trend assisted in the investigation of instrument drift and stability of solutions. Regression and correlation analysis were also used in the study of the volumes and concentration level. In an ancillary part of the overall investigation response surface analysis was also used to good advantage.

On the strength of the findings resulting from the applications of these techniques, it was possible to confirm many conjectures which had previously existed, as well as absolve of responsibility for bias and variation one or two aspects of the operation. On net balance, many recommendations were tendered the management, including those discussed in the next section.

5.2 Statistical Control Recommendations.

A laboratory control program administered by a suitably trained individual would be highly desirable. It would serve the purposes of keeping management informed, pointing out where corrective action is required, and helping the analysts to do their best.

As a minimum, the following elements should comprise the control effort:

1. Control charts for each colorimetric instrument, to maintain surveillance over bias and precision. A multi-vary chart may be useful here, or a combination of differences and range charts.*

2. Control charts on each automatic colorimeter, for drift control. Individuals and moving ranges charts may be useful here.

3. Control charts on selected reagents and other critical solutions, to avoid using one which has been degraded. Averages and ranges control charts should be useful here, application being made to reagents obtained from vendors as well as those prepared in house. In the former case, the procedures can be related to acceptance sampling.

---

*Those unfamiliar with multi-vary charts may find explanations in either reference given above.

4. Control charts on each laboratory analyst, using standards inserted into the production stream as unknowns. Differences and range charts are again useful, and should be maintained for each regular procedure.

Finally, I would like to point out the findings and recommendations discussed above are not unique to the particular laboratory concerned. I have had very similar experiences in working with several establishments of a completely different nature. The same kinds of problems were found in all of them; all of them could be helped by an appropriate statistical control program. I'm sure your laboratory can, too.

Acknowledgement: I am grateful to H. T. McAdams, a colleague at CAL, for his permission to draw upon one of his studies for a portion of the above.

REFERENCES

1. Duncan, Acheson, J., QUALITY CONTROL AND INDUSTRIAL STATISTICS: Richard D. Irwin, Inc., Homewood, Illinois, Third Edition, 1965.

2. J. M. Juran, Editor, QUALITY CONTROL HANDBOOK; McGraw-Hill Book Company- Inc., New York, Second Edition, 1962.

SOME STATISTICAL ANALYSIS WITH RESPECT TO COMPOSITING IN THE SAMPLING OF BULK MATERIAL

A. J. Duncan, The Johns Hopkins University

1. INTRODUCTION

The sampling of bulk material differs in a number of respects from the sampling of individual items. These differences are discussed at some length in reference (1). One difference is that bulk material can be physically composited whereas individual items cannot. It is thus possible in the case of bulk material to take a physical average in lieu of an arithmetic average. Although mixing and reduction may be expensive, the great decrease in the number of tests that have to be run with physical compositing is likely to yield considerable economy.

To illustrate bulk sampling with compositing consider the following example. An inspector wishes to determine the percent nitrogen in a given lot of fertilizer. The lot contains 200 bags. He selects 20 bags at random and with a sampling tube draws a small portion of fertilizer from each of the 20 bags. These portions are poured on to a rubber mat, are thoroughly mixed and hand-quartered until there is just enough to fill a laboratory bottle. Two tests are run on the reduced composite sample.

It is to be noted that the reduction of the composite sample is a form of subsampling and is thus accompanied by sampling variability. The variance of this is called the "reduction variance" (σ_r^2). There is thus a greater variability with compositing than with arithmetic averaging. In the analysis that follows σ_r^2 will be one of the terms in the sampling variance whenever there is compositing. Generally it is assumed that the reduction variance is the same whether we are reducing a large or somewhat smaller quantity. Of course, if the composite sample that is being reduced is not large relative to the part retained, a finite population correction factor may have to be applied.

Before we discuss the statistical aspects of compositing, let us look at the statistical procedures pertinent to a case in which there is no compositing.

2. SAMPLING WITH NO COMPOSITING

Let us consider a modification of the ASTM Tentative Recommended Practice for Sampling Industrial Chemicals (E 300-66 T). Let us consider only a single stage instead of the two-stage plan that is actually discussed in the Recommended Practice. To keep a concrete example in mind suppose the bulk material comes in cans and is homogeneous within cans, but varies in quality from can to can. Assume that we have an isolated lot of this material and are interested in the mean of the lot. With our modification of E 300-66 T the procedure would go as follows:

1. Take a preliminary sample of n_1 (e.g. 10) cans and measure the quality characteristic of the contents of each can.
2. Compute $\bar{X}_1 = \Sigma X_i/n_1$ and

$$s_1^2 = \sum_i (X_i - \bar{X}_1)^2/(n_1 - 1)$$

3. Use these data to determine on overall sample size (n) that would yield certain desired criteria.
4. Take an additional $n - n_1$ cans and test the contents of each.
5. Compute $\bar{X} = \sum_i X_i/n$ and $s^2 = \Sigma_i (X_i - \bar{X})^2/(n - 1)$.
6. Determine 0.95 confidence limits for the mean of the lot. Thus 0.95 confidence limits for μ would be $\bar{X} \pm t_{0.025} s/\sqrt{n}$ where $t_{.025}$ is the 0.025 point of a t distribution with $n-1$ degrees of freedom.

3. SAMPLING WITH COMPOSITING

Now consider the above example if after the n cans are selected they are physically composited and a single test made on this composite. Assume that the n_1 preliminary cans are measured as before. By this compositing we have reduced the cost of inspection by the cost of $n - n_1 - 1$ tests. We have added, however, the cost of compositing n cans and reducing this for running a single test.

3.1 WHEN BASIC VARIANCES ARE KNOWN

The variance of the single composite measurement (X_c) will be

$$\sigma_{X_c}^2 = \frac{\sigma_X^2}{n} + \sigma_r^2 + \sigma_a^2 \qquad (1)$$

where σ_X^2 is the product variance, σ_r^2 is the variance of reduction and σ_a^2 is the variance of analysis. If we knew all three of the variances,

0.95 confidence limits for μ would be given by

$$X_c \pm 1.96 \left(\frac{\sigma_X^2}{n} + \sigma_r^2 + \sigma_a^2\right)^{1/2}$$

3.2 WHEN BASIC VARIANCES ARE UNKNOWN

Let us see what we can do if we do not know the basic variances. Note that

$$\frac{X_c - \mu}{\left(\frac{\sigma_X^2}{n} + \sigma_r^2 + \sigma_a^2\right)^{1/2}} = \frac{\sqrt{n}(X_c - \mu)}{\left(\sigma_X^2 + n\sigma_r^2 + n\sigma_a^2\right)^{1/2}}$$

is normally distributed with zero mean and unit variance. Also note that s_1^2 contains both the product variance (σ_X^2) and the analytical variance (σ_a^2) which we will assume are independent so that

$$E(s_1^2) = \sigma_X^2 + \sigma_a^2$$

Then $\frac{(n_1 - 1)s_1^2}{\sigma_X^2 + \sigma_a^2}$ has a χ^2 distribution with $n_1 - 1$ degrees of freedom.
It follows that

$$\frac{\sqrt{n}(X_c - \mu)}{(\sigma_X^2 + n\sigma_r^2 + n\sigma_a^2)^{1/2}} \Bigg/ \frac{s_1}{(\sigma_X^2 + \sigma_a^2)^{1/2}}$$

has a t-distribution with $n_1 - 1$ degrees of freedom.

The above statistic would appear to be of little use to us since we do not know the basic variances. Note, however, that it can be rewritten in the form:

$$\frac{\sqrt{n}(X_c - \mu)}{s_1} \left(\frac{1 + \sigma_a^2/\sigma_X^2}{1 + n\sigma_r^2/\sigma_X^2 + n\sigma_a^2/\sigma_X^2}\right)^{1/2} \qquad (2)$$

Thus if we know the ratios σ_r^2/σ_X^2 and σ_a^2/σ_X^2, the statistic could be used and good guesses as to the ratios might work out fairly well.

For example, with n = 20 variation in the ratios from .8 to 1.2 causes the factor in brackets to vary only from .208 to .234, so that a 20% margin of error in estimating the ratios would cause only a deviation of 0.02 to 0.03 in the value of t, hardly enough to have a significant effect on the probabilities involved.

If we wish to set up approximate 0.95 confidence limits for μ in this case, we would have

0.95 confidence limits for μ

$$= X_c \pm \frac{s_1}{\sqrt{n}} \cdot \left[\frac{1 + n\sigma_r^2/\sigma_X^2 + n\sigma_a^2/\sigma_X^2}{1 + \sigma_a^2/\sigma_X^2}\right]^{1/2} \cdot t_{0.025\ (n_1 - 1)}$$

3.3 WHEN OUTSIDE ESTIMATES OF VARIANCE ARE AVAILABLE

If we do not know σ_X^2, σ_r^2, σ_a^2 but have independent estimates, we can proceed as follows. Let s_r^2 be an estimate of the reduction variance based on f_r degrees of freedom and let s_a^2 be an independent estimate of the analytical variance based on f_a degrees of freedom. Then we will have

$$E(s_1^2 - s_a^2) = \sigma_X^2 + \sigma_a^2 - \sigma_a^2 = \sigma_X^2$$

and an estimate of the variance of X_c will be

$$\frac{s_1^2 - s_a^2}{n} + s_a^2 + s_r^2 = \frac{s_1^2}{n} + \frac{(n-1)}{n} s_a^2 + s_r^2$$

This is a weighted sum of variances so that following Satterthwaite [4]

$$\frac{\upsilon\left[s_1^2 + \frac{(n-1)}{n} s_a^2 + s_r^2\right]}{\frac{\sigma_X^2}{n} + \sigma_a^2 + \sigma_r^2}$$

will have approximately a χ^2 distribution with degrees of freedom given by

$$\upsilon = \frac{\left[\frac{s_1^2}{n} + \frac{(n-1)}{n} s_a^2 + s_r^2\right]^2}{\frac{(s_1^2/n)^2}{n_1 - 1} + \frac{((n-1)s_a^2/n)^2}{f_a} + \frac{(s_r^2)^2}{f_r}}$$

Hence the statistic

$$\frac{(X_c - \mu)}{\left(\frac{s_1^2}{n} + \frac{(n-1)}{n} s_a^2 + s_r^2\right)^{1/2}}$$

will have approximately a t distribution with υ degrees of freedom and 0.95 confidence limits for μ will be given approximately by

$$X_c \pm \left(\frac{s_1^2}{n} + \frac{(n-1)}{n} s_a^2 + s_r^2\right)^{1/2} \cdot t_{.025\ (\upsilon)} \qquad (3)$$

4. SAMPLING A STREAM OF LOTS

The preceding discussion was concerned with an isolated lot and what kind of inferences we could make about the mean of the lot under varying circumstances. The approach to the analysis was strictly classical. Here we shall consider ASTM Tentative Methods for Mechanical Sampling of Coal (D2234-65T) and the point of view will be Bayesian.

4.1 ASTM D2234-65T

Sampling of coal on a conveyor belt consists of taking n increments of coal systematically from the belt, compositing these increments, reducing the composite sample to a laboratory sample and making a determination of quality such as ash content or the like. D2234-65T offers a solution to the problem of how many increments should make up the sample. The solution calls for a preliminary determination of the basic variances and is based on the assumption that the values so determined continue to be valid for routine sampling of subsequent lots.

4.1.1 THE STOCHASTIC ASSUMPTIONS ABOUT THE STREAM OF COAL

The program offered by this standard is based on a hypothesis regarding the nature of the coal being sampled. This is that the variations of quality in the stream consist of two kinds; one is a local variance, the other a "trend" or "segregation" variance. It is as if* the coal came in large segments which varied in average quality from segment to segment, a measure of which is the trend variance σ_t^2, while within the segments there is random variability the variance of which is designated as σ_1^2 (since it applies to one lb. increments). The within-segment variance is assumed to be the same for all segments. Measurements of the quality of individual

*Note that this is considered to be an approximate working model, not a true model. cf.[5].

increments of w lbs. of coal taken at random from various segments with no more than one increment per segment would thus have a variance equal to $\frac{\sigma_1^2}{w} + \sigma_t^2 + \sigma_r^2 + \sigma_a^2$ (where in this case σ_r^2 refers to the variance resulting from the reduction of the w lbs. of coal.)

4.1.2 THE PILOT STUDY

D2234-65T calls for a pilot study for determining the basic variance components for a given coal. The study provides for the collection of 30 sets of two samples from a stopped conveyor belt. "Each of the 30 sets of samples includes a very small sample, to furnish data for the random variance, and a large sample, to furnish data for the system (trend) variance. Since one of the important components of variance is that due to segregation it is essential that the 30 sets of samples be so distributed with respect to time that coverage of all subtypes of coal are represented" [6]. The samples are to be taken by a two section Belt Divider. "One of the sections should be approximately the width corresponding to three times the top size of the coal and should trap a sample of between 4 and 20 lb. The other section should be approximately the width corresponding to 20 times the top size of the coal and should trap a sample between 80 and 150 lb." [6]. Designate the small samples by the letter A and the large samples by the letter B. The subsamples A are reduced say by a riffle to laboratory samples of between 100 and 200 grams. These are ground to -60 mesh for analysis. The subsamples B are also worked down to laboratory samples and ground to -60 mesh for analysis.

The variances of the A and B results are measured by the usual formula

$$s^2 = \frac{\sum_i (X_i - \bar{X})^2}{n - 1}$$

Now if the weight of the A samples is w_1 and that of the B samples is w_2, then assuming w_2/w_1 to be an integer and letting $\sigma_{w_1}^2$ be the random variance for increments of coal, weighing w_1 lbs., we have

$$\text{Expected value of } s_A^2 = \sigma_t^2 + \sigma_{w_1}^2 + \sigma_r^2 + \sigma_a^2$$

and

$$\text{Expected value of } s_B^2 = \sigma_t^2 + \frac{w_1}{w_2}\sigma_{w_1}^2 + \sigma_r^2 + \sigma_a^2$$

An unbiased estimate of $\sigma_{w_1}^2$ will be given by

$$\delta_{w_1}^2 = \frac{w_2(s_A^2 - s_B^2)}{w_2 - w_1}$$

and for 1 lb. increments, the random variance can be estimated at

$$\delta_1^2 = w_1 \delta_{w_1}^2$$

To estimate the trend variance σ_t^2, multiply s_A^2 by w_1 and s_B^2 by w_2 and subtract. This yields

$$\delta_t^2 = \frac{w_2 s_B^2 - w_1 s_A^2}{w_2 - w_1} - s_r^2 - s_a^2$$

where s_r^2 and s_a^2 are estimates of the reduction and analytical variances obtained from another pilot study that need not be reviewed here. The above estimates can be used to obtain an estimate of the variance of an increment of any weight w. Thus

$$\delta_w^2 = \frac{\delta_1^2}{w} + \delta_t^2 + s_r^2 + s_a^2$$

A composite of n increments would have an estimated variance equal to

$$\frac{\frac{\delta_1^2}{w} + \delta_t^2}{n} + s_r^2 + s_a^2 \qquad (4)$$

4.1.3 DETERMINATION OF n FOR SUBSEQUENT SAMPLING

The last formula is employed by the ASTM standard to determine how many increments should be used in future routine sampling of a lot. Thus, proceeding in a crude manner, the standard notes that 0.95 confidence limits for the mean of any given portion of coal from which n increments of weight w have been taken, composited, reduced and tested would be given roughly by

$$X_c \pm 1.96 \left(\frac{\frac{\delta_1^2}{w} + \delta_t^2}{n} + s_r^2 + s_a^2 \right)^{1/2}$$

so if we wish the confidence interval to be of width 2Δ, then we would take

$$\Delta = 1.96 \left(\frac{\frac{\delta_1^2}{w} + \delta_t^2}{n} + s_r^2 + s_a^2 \right)^{1/2}$$

and solve for n. The standard takes $\Delta = .10\,\mu$ where μ is a good guess as to the mean quality of the consignment of coal.

4.1.4 ESTIMATION OF THE MEAN OF THE CURRENT LOT

With the number of increments determined as described in 4.1.3, the mean of a current lot is estimated by taking the prescribed number of increments from the current lot, compositing them, reducing the composite to a laboratory sample and analyzing a specimen from this sample. The result X_c is taken as an estimate of the quality of the lot. From the way in which the number of increments was decided this estimate is expected to be within 10% of the true mean of the lot. If confidence limits are desired, they could be crudely determined by

$$\mu = X_c \pm 1.96 \left(\frac{\hat{\sigma}_1^2}{w} + \frac{\hat{\sigma}_t^2}{n} + s_r^2 + s_a^2 \right)^{1/2}$$

4.2 USE OF PRIOR INFORMATION IN ESTIMATING THE MEAN OF A CURRENT LOT

It will be noted that while the prior information in variances is used in Section 4.1.4 to set up crude confidence limits for the lot mean, no use is made of the mean of the pilot study. The question may be raised, however, as to whether the mean of the current lot would not be better estimated by a weighted average of the pilot study mean and the measurement X_c made from the current lot. The argument would run like this. If the various lots of coal were really large samples from the stream of coal, their individual means would probably differ very little from the mean of the whole stream and the best estimate we could make of the mean of a current lot would be an estimate of the mean of the stream based on all the information available for making such an estimate. Suppose, for example, that in the pilot study the mean of the 30 large (w_2 lb.) samples was $\overline{X}_{w_2}$ and the mean of the 30 small (w_1 lb.) samples was $\overline{X}_{w_1}$, then an estimate of the mean of a subsequent lot on which we have a composite sample measurement X_c could be taken as

$$\overline{X}_1 = \frac{30w_2 \overline{X}_{w_2} + 30w_1 \overline{X}_{w_1} + nwX_c}{30w_2 + 30w_1 + nw}$$

where w is the weight of the n increments composited in the sampling of the current lot. An alternative estimate that omits the small samples of the pilot study would be

$$\overline{X}_2 = \frac{30w_2 \overline{X}_{w_2} + nwX_c}{30w_2 + nw}$$

This would be based on less data. The reduction in the amount of data would not be great, however, and as discussed below, it might be feasible to use $\overline{X}_2$ in a supplementary test of significance but not $\overline{X}_1$.

Pooling the pilot study with the current lot measurement should be preceded by a statistical test to determine whether the assumption that the two sets of data came from the same population is a reasonably valid one. Actually it will be sufficient to test whether the means differ signigicantly. If we plan to use $\bar{X}_1$ above, the test would be based on a comparison of $\frac{30w_2\bar{X}_{w_2} + 30w_1\bar{X}_{w_1}}{30w_2 + 30w_1}$ with X_c or if we plan to use $\bar{X}_2$, on a comparison of $\bar{X}_{w_2}$ with X_c.

It is necessary at this point to interrupt the argument and to note that if the sampling and analyticial procedures used is as described in Section 4.1.4 above, there will be only one measurement made on the current lot, viz., X_c, and this will provide us with no information on the variance of the current lot. In order to be able to run the suggested significance test, it will be necessary for the sampling and analyticial procedures to be modified to give some information on variability of the current lot. To accomplish this it is recommended that in lieu of a single composite sample, 4 separate composite samples be formed* and measured separately. The mean of the four separate composite measures ($\bar{X}_c$) would take the place of the single composite measure X_c and the variance of the 4 separate composite means would yield an estimate of the variance of the current lot. Thus, we would have

$$\bar{X}_c = \frac{X_{c1} + X_{c2} + X_{c3} + X_{c4}}{4}$$

and

$$s_c^2 = \frac{\Sigma(X_{ci} - \bar{X}_c)^2}{4 - 1}$$

and $\frac{n}{4} s_c^2$ would be an estimate of the variance of increments of weight w from the current lot.

Returning to the discussion of the significance tests, it is to be noted that if we set $a = 30w_2/(30w_2 + 30w_1)$ and $b = 30w_1/(30w_2 + 30w_1)$, the variance of the weighted mean of the pilot study data would be

$$\sigma^2_{\text{weighted mean}} = a^2\sigma_{\bar{X}_{w_2}}^2 + b^2\sigma_{\bar{X}_{w_1}}^2 + 2abr\sigma_{\bar{X}_{w_1}}\sigma_{\bar{X}_{w_2}}$$

where r is the correlation between $\bar{X}_1$ and $\bar{X}_2$. In practice this would be estimated by $a^2s_B^2/30 + b^2s_A^2/30 + 2abrs_Bs_A/30$ where

---

*If the increments are taken systematically from the lot then increments 1,5,9,...could be mixed to form composite 1, increments 2,6,10,... could be mixed to form composite 2, increments 3,7,11,...could be mixed to form composite 3, and increments 4,8,12,...could be mixed to form composite 4.

$$r = \sum_i \frac{\left(X_{iw_2} - \bar{X}_{w_2}\right)\left(X_{iw_1} - \bar{X}_{w_1}\right)}{29 s_B s_A}$$

and s_B^2 and s_A^2 are as defined in Section 4.1.2 above. The variance of $\bar{X}_c$ would be estimated by $s_c^2/4$. A crude test of the difference between the two means would therefore be given by treating

$$\frac{\dfrac{30w_2\bar{X}_{w_2} + 30w_1\bar{X}_{w_1}}{30w_2 + 30w_1} - \bar{X}_c}{\left(\dfrac{a^2 s_B^2 + b^2 s_A^2 + 2abrs_B s_A}{30} + \dfrac{s_c^2}{4}\right)^{1/2}} \tag{5}$$

as if it were normally distributed. If this does not fall beyond +1.96 or -1.96, we could conclude that it is safe to make the pooled estimate. If the given statistic falls beyond ±1.96, pooling is not recommended and $\bar{X}_c$ alone should be taken to estimate the mean of the current lot.

The quantity r would have to be computed from the original pilot study data. If this information is not available, then

$$\bar{X}_2 = \frac{30w_2\bar{X}_{w_2} + nw\bar{X}_c}{30w_2 + nw}$$

could be used in place of $\bar{X}_1$. In this instance a crude significance test would be given by treating

$$\frac{\bar{X}_{w_2} - \bar{X}_c}{\left(\dfrac{s_B^2}{30} + \dfrac{s_c^2}{4}\right)^{1/2}} \tag{5a}$$

as if it were normally distributed, again comparing it with ± 1.96.

If there is a series of lots to be inspected from the stream of coal, the mean of each lot that passes the significance test could be pooled with the pilot study mean and other past lot means that have met the significance tests. This pooled mean would then become the point of reference with which the mean of the current lot ($\bar{X}_c$) would be compared. The significance test would in this case be carried out by treating

$$\frac{\bar{X}_{\text{Pooled}} - \bar{X}}{\left(s^2_{\text{Pooled Mean}} + \dfrac{s_c^2}{4}\right)^{1/2}} \tag{5b}$$

as if it were normally distributed, where for m past means

$$\bar{X}_{Pooled} = \frac{30w_2\bar{X}_{w_2} + nw\sum_{i=1}^{m}\bar{X}_{ci}}{30w_2 + mnw}$$

and for $g = \frac{30w_2}{30w_2 + mnw}$ and $h = \frac{nw}{30w_2 + mnw}$

$$s^2_{Pooled} = g^2 s^2_B/30 + h^2 \sum_{i=1}^{m} s^2_{ci}/4$$

Comparison would again be with ± 1.96, although owing to truncation the α-risk would now be less than 0.05. The basic variances are assumed unchanged.

4.3 A MORE GENERAL BAYESIAN PROCEDURE

An all around and somewhat more sophisticated approach as to how to use the pilot study data in making inferences about the lot mean would be to use Bayes Theorem in which the prior distributions are based on the pilot study data. Our probabilities would now become rational degrees of relief, but they would not be entirely subjective in that if the mathematical procedure is accepted, the degrees of belief, i.e. the probabilities, follow directly from the analysis. The procedure will be to note the various likelihoods and prior distributions and then apply Bayes theorem to get the posterior distribution for μ, the lot mean.

We shall begin by taking the composite mean $\bar{X}_c$ to be distributed normally about the lot mean μ with variance equal to

$$\left[\frac{\sigma_1^2/w + \sigma_t^2}{n/4} + \sigma_r^2 + \sigma_a^2\right] \Big/ 4$$

If we set $\sigma_X^2 = \sigma_1^2/w + \sigma_t^2$, this becomes

$$\frac{\sigma_X^2}{n} + \frac{\sigma_r^2 + \sigma_a^2}{4}$$

Then the density function for $\bar{X}_c$ given μ is

$$f(\bar{X}_c \mid \mu) = \frac{e^{-(\bar{X}_c - \mu)^2/2\left(\frac{\sigma_X^2}{n} + \frac{\sigma_r^2 + \sigma_a^2}{4}\right)}}{\left(\frac{\sigma_X^2}{n} + \frac{\sigma_r^2 + \sigma_a^2}{4}\right)^{1/2}\sqrt{2\pi}} \qquad (6)$$

Next, let the variance of the four composites X_{ic} be $s_c^2 = \Sigma(X_{ic} - \bar{X}_c)^2/(4 - 1)$. The expected value of s_c^2 will be

$E(s_c^2) = \frac{4\sigma_X^2}{n} + \sigma_r^2 + \sigma_a^2$ and we shall assume that $3s_c^2$ divided by this expected value has a χ^2 distribution with 3 degrees of freedom. Thus the distribution of s_c^2, given the variances σ_X^2, σ_r^2 and σ_a^2, will be

$$h(s_c^2|\sigma_X^2, \sigma_r^2, \sigma_a^2) = K \frac{(s_c^2)^{(4-2)/2} \exp(-3s_c^2/2([4\sigma_x^2/n] + \sigma_r^2 + \sigma_a^2))}{\left[\frac{4\sigma_X^2}{n} + \sigma_r^2 + \sigma_a^2\right]^{4/2}} \quad (7)$$

where K is a factor of proportionality. Since the two densities are independent of each other their joint density will be the product of the two individual densities.

We shall express our degrees of belief about prior distributions as follows:

Prior distribution for μ is proportional to $\frac{1}{\sigma_X} \cdot e^{-M(\mu - \mu_0)^2/2\sigma_X^2}$ (8)

where M is the size of the lot and μ_0 is the grand mean of the pilot study. The assumption here is that the lots are merely large samples from the stream of coal.

Prior distribution for σ_X^2 is proportional to

$$\frac{e^{-f_{oX}(\sigma_{oX}^2)/2\sigma_X^2} \cdot (\sigma_{oX}^2)^{f_{oX}/2}}{(\sigma_X^2)^{1 + f_{oX}/2}} \quad (9)$$

where σ_{oX}^2 is the pilot study estimate of $\frac{\sigma_1^2}{w} + \sigma_t^2$ and f_{oX} is the degrees of freedom on which it is based. This is a conjugate prior for the distribution of s_c^2.

Prior distribution for σ_r^2 is proportional to

$$\frac{e^{-f_{or}(\sigma_{or}^2)/2\sigma_r^2} \cdot (\sigma_{or}^2)^{f_{or}/2}}{(\sigma_r^2)^{1 + f_{or}/2}} \quad (10)$$

Prior distribution for σ_a^2 is proportional to

$$\frac{e^{-f_{oa}(\sigma_{oa}^2)/2\sigma_a^2} \cdot (\sigma_{oa}^2)^{f_{oa}/2}}{(\sigma_a^2)^{1 + f_{oa}/2}} \tag{11}$$

These are similar in form to the density assumed for σ_X^2.

The product of all the above densities would be the joint distribution of the likelihoods and priors. Unfortunately the expression is too complicated to handle analytically. It is possible to do something, however if we neglect σ_r^2 and σ_a^2 and merely include μ and σ_X^2. Since σ_r^2 and σ_a^2 are likely to be small relative to σ_X^2, the approximation may not be bad. Confidence limits based on it will indicate limits that are less than the true ones. They will offer a lower bound, however.

Proceeding as indicated, we would have, omitting non-relevant factors, the following joint distribution of likelihoods and priors. (Note that f_o and σ_o are now used instead of f_{oX} and σ_{oX} since there is no longer a need to distinguish pilot study variances.) Thus $f(\bar{X}_c, s_c^2, \mu, \sigma_X^2)$ would be proportional to

$$\frac{e^{-n(\bar{X}_c - \mu)^2/2\sigma_X^2}}{\sigma_X/\sqrt{n}} \cdot \frac{(s_c^2)e^{-3s_c^2/2(4\sigma_X^2/n)}}{(4\sigma_X^2/n)^{4/2}} \cdot \frac{1}{\sigma_X} e^{-M(\mu - \mu_o)^2/2\sigma_X^2} \cdot \frac{e^{-f_o(\sigma_o^2)/2\sigma_X^2}}{(\sigma_X^2)^{1 + f_o/2}} \cdot (\sigma_o^2)^{f_o/2} \tag{12}$$

where μ_o^2 is the pilot study mean and σ_o^2 is the pilot study variance.

If we integrate (12) over μ, this leaves, except for factors not containing σ_X^2,

$$\frac{\exp\left(- \left[\frac{(n\bar{X}_c + M\mu_o)^2}{n + M} + M\mu_o^2 + n\bar{X}_c^2 + \frac{3ns_c^2}{4} + f_o\sigma_o^2 \right] \Big/ 2\sigma_X^2 \right)}{(\sigma_X^2)^{1/2} \cdot (\sigma_X^2)^{1 + f_o/2} \cdot (4\sigma_X^2/n)^{4/2}}$$

Set the expression in brackets equal to H, so that the quantity becomes, except for factors not involving σ_X^2,

$$e^{-H/2\sigma_X^2} \cdot (\sigma_X^2)^{-(f_o + 7)/2}$$

The next step is to integrate over σ_X^2. If we set $\chi^2 = H/\sigma_X^2$ so that $d\sigma_X^2 = -\frac{H}{(\chi^2)^2}\, d\chi^2$, the integral becomes

$$\int_0^\infty e^{-\chi^2/2} \cdot H^{-(f_o + 5)/2} \cdot (\chi^2)^{-(f_o + 3)/2} \cdot d\chi^2$$

which is proportional to

$$H^{-(f_o + 5)/2}$$

The joint posterior distribution is therefore given (except for a proportionality factor) by Expression (12) $\cdot$ $H^{(f_o + 5)/2}$.

To get the marginal distribution for μ, which is the posterior distribution of μ, we integrate the joint posterior distribution over σ_X^2. The part of Expression (12) $\cdot$ $H^{(f_o + 5)/2}$ containing σ_X^2 is

$$\frac{e^{-\left[n(\bar{X}_c - \mu)^2 + \frac{3ns_c^2}{4} + M(\mu - \mu_o)^2 + f_o\sigma_o^2\right]/2\sigma_X^2}}{(\sigma_X^2)^{(f_o + 8)/2}}$$

If we set G equal to the expression in the brackets and put $\chi^2 = G/\sigma_X^2$, the integral becomes

$$\int_0^\infty \frac{e^{-\chi^2/2}}{(G/\chi^2)^{(f_o + 8)/2}} \cdot \frac{G}{(\chi^2)^2}\, d\chi^2$$ which is proportional to

$G^{-(f_o + 6)/2}$. Accordingly, except for a proportionality factor, $G^{-(f_o + 6)/2}$ is the posterior distribution of μ.

Now G can be put in the form

$$G = (n + M)\left[(\mu - \frac{n\bar{X}_c + M\mu_o}{n + M})^2 + V\right] \qquad (13)$$

where $V = -\left[\frac{n\bar{X}_c + M\mu_o}{n+M}\right]^2 + \frac{\frac{3ns_c^2}{4} + n\bar{X}_c^2 + M\mu_o^2 + f_o\sigma_o^2}{n+M}$ (14)

Dividing by V and absorbing $V^{-\frac{f_o+6}{2}}$ in the proportionality factor, we have

$G^{-\frac{f_o+6}{2}}$ is proportional to $\left\{\frac{\left[\mu - \frac{n\bar{X}_c + M\mu_o}{n+M}\right]^2}{\sqrt{V/(f_o+5)}} \Big/ (f_o+5) + 1\right\}^{-\frac{f_o+5+1}{2}}$

Since the above expression is of the form $(\frac{t^2}{\nu} + 1)^{-\frac{\nu+1}{2}}$, it follows that the posterior distribution of a function of μ (not μ itself) has the form of a t-distribution and that Bayesian confidence intervals for μ can be obtained from this.

Thus the probability that $\frac{\mu - \frac{n\bar{X}_c + M\mu_o}{n+M}}{\sqrt{V/(f_o+5)}}$

lies between $-t_{.025}$ and $+t_{.025}$ for $f_o + 5$ degrees of freedom equals 0.95.

Hence the probability is 0,95 that μ lies between

$$\frac{n\bar{X}_c + M\mu_o}{n+M} \pm \sqrt{V(f_o+5)} \cdot t_{.025} \text{ for } f_o + 5 \text{ degrees of freedom} \quad (15)$$

These are consequently the Bayesian 0.95 confidence limits for μ. The limits given by (15) are tighter than the true ones since it will be recalled that σ_r^2 and σ_a^2 have been neglected.

It is interesting to note that the sample size n appears in V in a peculiar way. If we divide both the numerators and denominators of the two terms of V by n, we get in the numerators the quantities $\frac{M\mu_o}{n}$, $\frac{M\mu_o^2}{n}$, and $\frac{f_o\mu_o^2}{n}$ and in the denominators M/n. If we increase n, M/n goes down less rapidly than $\frac{M\mu_o}{n}$ or $\frac{M\mu_o^2}{n}$ and the reduction due to $\frac{f_o\mu_o^2}{n}$ is gratis.

This verifies what should be the case, viz., that as n increases the confidence interval becomes smaller. The same applies to f_o the degrees of freedom for the pilot study of σ_X^2. Likewise for the lot size M.

Finally it should be noted that the analysis indicates that the lot size M should be taken as the weighting factor for μ_o in getting the average estimate of μ. This stems from the special assumption that was made about the prior distribution, viz., that the lots were merely large samples from the process of size M and thus had a variance of σ_X^2/M. If the special assumption about the prior distribution of μ is that it is normal with mean μ_o but with variance σ_x^2/M where M is an arbitrary constant, the model would still hold. In either case the weighting would be inversely proportional to the variances, as would be expected.

4.4 A COMPARISON OF APPROACHES

It is of interest to conclude with a comparison of the results yielded by the formula of Section 4.1.4 with Bayesian confidence limits yielded by formula (15). The comparison will be made by numerical examples.

Suppose that 20 increments of 50 lbs. each are taken from a current lot to form a composite sample the measurement on which (X_c) is 10. Suppose that the random variance for 1 lb. increments (σ_1^2) and the trend variance (σ_t^2) have been estimated from a pilot study to be 7.6 and 1.2 respectively.* And suppose another pilot study yields an estimate of the sum of the reduction and analytical variances ($s_r^2 + s_a^2$) as equal to 0.0465.* Then using the formula of Section 4.1.4 above, approximate 0.95 confidence limits for the mean of the lot would be

$$\mu = 10 \pm 1.96 \left(\frac{\frac{7.6}{50} + 1.2}{20} + 0.0456 \right)^{1/2}$$

$$= 10 \pm 0.66$$

Suppose now that instead of a single composite, 4 composite samples are taken from the current lot each based on 5 increments of 50 lbs. each so that n still equals 20. Let the mean of the 4 composite measurements ($\bar{X}_c$) be 10 and let their variance (s_c^2) be 0.32. Let the mean of the pilot study data (μ_o) be 12 and let the size of the lot (M), measured in 50 lbs. increments,

---

*These figures are taken from the illustrative material given in the ASTM standard Methods for Sampling of Coal (D2234).

be 800 such increments. Finally, let the number of degrees of freedom (f_o) on which the pilot study estimate of the product variance σ_o^2 is based be 30.* Then with the same values for δ_1^2, δ_t^2 and s_r^2 and s_a^2 as before we will have

$$\sigma_o^2 = \frac{\delta_1^2}{w} + \delta_t^2 = \frac{7.6}{50} + 1.2 = 1.35$$

and the 0.95 Bayesian confidence limits for the mean of the current lot would be (according to formulas (14) and (15))

---

*The number of degrees of freedom f_o is derived by Satterthwaite's approximation as follows: The pilot study referred to in the coal sampling standard D 2234 yielded $s_A^2 = 29.2$ and $s_B^2 = 1.3$ (See Section 4.1.2 of this paper). In deriving these variances, $w_1 = 0.27$ lb. was taken as the size of the small samples and $w_2 = 10.6$ lbs. was taken as the size of the large samples. With w = 50, these figures yielded (in accordance with Section 4.1.2 above)

$$\sigma_o^2 = \frac{\delta_1^2}{50} + \delta_t^2 = \frac{0.27(1.00255)(s_A^2 - s_B^2)}{50}$$

$$+ 1.00255\, s_B^2 - 0.002554\, s_A^2 - s_{ra}^2$$

where s_{ra}^2 is the sum of the reduction and analytical variances. This can be put in the form

$$\sigma_o^2 = 0.004725\, s_A^2 + 0.9825\, s_B^2 - s_{ra}^2$$

Each of the estimates of variance was based on 29 degrees of freedom so that the degrees of freedom for σ_o^2 [following Satterthwaite (See A.J. Duncan, Quality Control and Industrial Statistics, 3rd ed., p. 605)] was

$$f_o = \frac{(0.004725\, s_A^2 + 0.9825\, s_B^2 - s_{ra}^2)^2}{\frac{(0.004725\, s_A^2)^2}{29} + \frac{(.9825\, s_B^2)^2}{29} + \frac{(s_{ra}^2)^2}{29}}$$

which with $s_A^2 = 29.2$, $s_B^2 = 1.3$ and $s_{ra}^2 = 0.0465$, gives $f_o = 30$.

$$\mu = \frac{20(10) + 800\ (12)}{20 + 800} \pm t_{.025(30)} \sqrt{V/35}$$

where

$$V = -\left[\frac{20(10) + 800(12)}{20 + 800}\right]^2 + \frac{\frac{3(20)(0.32)}{4} + 20(10)^2 + 800(12)^2 + 30(1.35)}{20 + 800}$$

$$= -142.80 + \frac{4.8 + 117200 + 4.05}{820}$$

$$= -142.80 + 142.93 + \frac{4.8 + 40.5}{820} = .13 + \frac{45.3}{820} = 0.19$$

This yields

$$\mu = 11.95 \pm 2.04\sqrt{.0054} = 11.95 \pm 0.15.$$

As could have been anticipated, the relatively heavy weighting given μ_o as compared with that given $\bar{X}_c$ causes the Bayesian limits to be centered close to μ_o. Further, the assumption that the lots are merely random samples from the process yields a prior distribution for the lot mean that has a variance σ_x^2/M which for large M is small. This is what accounts for the much tighter confidence limits. With large M therefore there will be a marked difference in the results yielded by the two procedures.

It is not necessary, however, for the validity of the Bayesian model that the various lots be assumed to be random samples from the process with a variance equal to σ_x^2 divided by the lot size. As noted above it is possible to view M simply as an arbitrary constant which expresses the assurance we have about the location of the lot mean. Thus, if we take M = 80 instead of 800 as in the previous example, the prior distribution for μ will have much greater dispersion which means that our prior knowledge as to the value of μ is much less certain. In this case, the Bayesian confidence limits for the mean of the current lot will be

$$\mu = \frac{20(10) + 80(12)}{100} \pm t_{.025(30)} \sqrt{V/35}$$

where

$$V = -\left[\frac{20(10) + 80(12)}{100}\right]^2 + \frac{20(10)^2 + 80(12)^2}{100} + \frac{45.3}{100}$$

$$= -134.56 + 135.20 + .453 = .64 + .453 = 1.093.$$

This yields

$$\mu = 11.6 \pm 2.04\sqrt{0.0512} = 11.6 \pm .46$$

a result that is much closer to that yielded by the formula of Section 4.1.4 above.

If now we do not wish to assume any prior knowledge regarding the mean of the current lot (even the mean of the pilot study data is considered irrelevant), but we are willing to assume a prior distribution for the variance, then we can modify the Bayesian analysis by putting M = 0. If we do this, our confidence limits for μ become

$$\mu = X_c \pm 2.04\sqrt{V/35}$$

where

$$V = \frac{3ns_c^2/4 + f_o\sigma_o^2}{n}$$

which for the example in hand becomes

$$V = \frac{45.3}{20} = 2.27$$

This yields

$$\mu = 10 \pm 2.04\sqrt{0.0649} = 10 \pm .52.$$

If we allow for the omission of the variances of reduction and analysis, this is almost the same result as that given by the formula of Section 4.1.4. The conclusion seems warranted therefore, that the formula of Section 4.1.4 is the practical equivalent of a Bayesian confidence interval when we are willing to use the pilot study data to give us prior distributions for the basic variances but are unwilling to make any prior assumptions about the lot mean.

Selected References

(1) A.J. Duncan, "Bulk Sampling: Problems and Lines of Attack," Technometrics, Vol. 4, 1962, pp. 319-44.
(2) A.J. Duncan, "Contributions of ASTM to the Statistical Aspects of the Sampling of Bulk Materials", Materials Research and Science, Vol. 4, 1967, pp.477-85.
(3) D.V. Lindley, Introduction to Probability and Statistics, Pt. 2 (Cambridge University Press).
(4) F.E. Satterthwaite, "An Approximate Distribution of Estimates of Variance", Biometrics Bulletin, Vol. II (1946), pp. 110-14.
(5) J. Visman, Towards A Common Basis for the Sampling of Materials, (Dept. of Mines and Technical Surveys, Ottawa, Canada, Mines Branch Research Report R-93.)
(6) 1965 Book of ASTM Standards, Pt. 19, pp. 414-15.

PANEL DISCUSSION ON BULK SAMPLING

Chairman: Walter D. Foster, Biological Laboratories, Fort Detrick, Frederick, Maryland

Discussant: Acheson J. Duncan, The Johns Hopkins University, Baltimore, Maryland

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

Henry Ellner, U. S. Army Materiel Command, Washington, D. C.

Gene Ray Lowrimore, Hercules, Inc., Radford Army Ammunition Plant, Radford, Virginia

Joseph Mandelson, Edgewood Arsenal, Maryland

Vernon H. Rechmeyer, Thiokol Chemical Corporation, Huntsville Division, Redstone Arsenal, Alabama

Since the host installation for the Fourteenth Conference on the Design of Experiments has a special interest in chemical and other forms of bulk sampling, the Program Committee decided to have a group discussion in this area of statistics. Dr. Walter Foster agreed to serve as chairman of the panel and to select several experts to help him explore this field.

Three papers on bulk sampling appear in this technical manual. The preceding article by Professor Acheson Duncan, and the next two papers, one by Joseph Mandelson and the other by Gene Lowrimore.

CHEMICAL SAMPLING

Joseph Mandelson, Edgewood Arsenal, Maryland

The problem of sampling of chemical materials has never been solved on an overall basis and is not likely ever to be solved in this manner. By an "overall basis," I mean the establishment of a standard such as Military Standard 105 applicable to all materials which contain classifiable quality characteristics and to which an AQL can meaningfully be assigned. In the past, a number of standards have been prepared governing the sampling, inspection, and test of chemicals (e.g., ASTM, AOAC, etc.), but each standard is specific for one material and usually applicable to only one type or grade of that material. Thus, an ASTM standard for testing quicklime will tell you nothing about sampling of reagent grade CaO. And that is as it should be.

I believe we can handle the problem only by recognizing, in detail, what our objectives are and by indicating what can be done to handle each type of objective.

First off, there are three general ideas which are of the greatest significance in this area.

a. The concept of percent defective, which is basic to Military Standards 105 and 414, and in terms of which AQL's are expressed, has no meaning in connection with testing chemicals per se. Of course, inspection of factors such as packaging, packing, and marking of chemicals can be accomplished using AQL's and Military Standard 105, but not the actual testing of the chemical.

b. Probably the most important characteristic to be defined in planning chemical sampling is the degree of lot homogeneity required. This must be determined within the framework of the actual way the material is used and by the importance of the material in that usage. Examples of this will be indicated later on.

c. While Military Standards 105 and 414 assume that no inspection error occurs and their OC curves are plotted accordingly, the actual existence of error merely results in the translation of the OC to the right or left depending on the kind of error made. In chemical sampling, we have no OC curve (because the abcissa is a percent defective), but we do have an experimental error which may or may not be large enough to be significant. In any case, the size of experimental error can be determined (assuming competent testing personnel) and the causes thereof ascertained. In every case, the acceptance criteria set must reflect the irreducible experimental error while the actual sampling and test procedures must be hedged about with specified technical precautions to hold these errors to as near these minima as possible.

Now, let us see how these general ideas affect the problem of chemical sampling:

a. Military Standard 105 has appeared to many, if not most Quality Assurance engineers like Lydia Pinkham's pills, a cure-all for whatever ailments you have. They prescribe its use for anything and everything - including chemical sampling and sampling for destructive test. The small sample sizes contained in the S levels of Military Standard 105 are particularly cited though unsuited for these two purposes. For chemical sampling, the sample size indicated in the S levels depends upon the AQL prescribed and, as already indicated, AQL is rarely of significance in chemical testing. For example, for lot size of 1000 (packages, I suppose) level S-2, Military Standard 105 prescribes sample code letter C, which calls for a sample of 3 for 4.0% AQL, a sample of 8 for 6.5% AQL and a sample of 5 for all other AQL's. Of course, the allowable number of defects,

which has little if any meaning in chemical sampling, differs with sample size and AQL. So what do we do? Obviously, we had best avoid quoting Military Standard 105 and put a sampling table of our own into the specification. As an afterthought: suppose we write a specification for technical grade acetone. This can come in any number of commercial packages, from 5-gallon cans, 55-gallon drums, to tank cans. Imagine using level S-2 or any other quote from Military Standard 105!

b. Before we discuss the problem of homogeneity, I'd like to point out that chemical tests are or should be specified for accomplishment in replicate (that is, in 2 or more parallel determinations). Results are expected to vary due to experimental error so it is possible (and it frequently occurs) that one replicate will appear to fail with respect to one or more quality characteristics while others may meet the requirement. We usually allow the average to govern. But this is not always spelled out in the specification. Furthermore, the use of such undeclared decision criteria ignores the fact that certain requirements are far more important than others so that the average, by itself, may be insufficient to insure a desirable product. In fact, in many cases, an exact parallel exists with the concept of classification of defects as used in sampling and inspection in accordance with Military Standard 105.

For instance, for a vaccine, acceptance will require that no living virus be observed in any of the many replicate samples taken from the batch. This corresponds to the Military Standard 105 critical defect. Further, the number of units per gram or ml. of material is very important since dosage depends on precise control of this figure. It may be possible to admit of some variation such that one or more of the replicates may be permitted to fall somewhat below the specified minimum provided the average is not less than this minimum while the variation, measured as a standard deviation, is not greater than some prescribed maximum. This corresponds to the major defect concept. There may be additional requirements (e.g. specific gravity, etc.) of lesser importance where the average alone may be permitted to govern. These are equivalent to the minor defect.

We can see, therefore, that the more important the requirement, the greater the need for the lot to be homogeneous and the more stringent the evidence required to prove it. Also, you must now be aware that the requirement for homogeneity stems from the way the material is used and what it is supposed to accomplish.

By contrast with vaccine, let us consider FS: chlorosulfonic acid - SO_3 solution. The most important requirement is total acidity. However, in its use as a smoke agent, if the total acidity were 5% below the specified minimum, it is doubtful that you could see any difference in the smoke it made. So this would be a minor characteristic, even though it is the most important one, and the average of replicate determinations on a composite sample would be sufficient to govern.

In determining homogeneity, it is usually possible to test a single characteristic, not necessarily the one which is of critical interest, to prove it. Major characteristics require a number of individual samples and replicate tests of each but all minor characteristics can be determined on a composite sample. With this general guidance and your knowledge of the material and how and why it is used, meaningful, economic chemical sampling can be devised. One way of insuring a degree of homogeneity is to prescribe that product shall contain material from not more than one batch of chemicals. The batch is defined as that quantity of material manufactured by some unit chemical process or subjected to some physical mixing operation intended to make the final product substantially uniform. This is a minimum requirement in production of a homogeneous product.

c. In all chemical measurements explicit consideration must be given to the experimental error of the specified procedure. It is frequently taken that, in a well-run laboratory, the most common source of error lies in reading the instrument; e.g., 0.02 ml for ordinary burettes, 0.1 mg for the analytical balance, etc. Any experienced analyst knows this premise is highly optimistic and that reading errors comprise only a fraction, perhaps, but a small fraction, of the total error. Most important, however, is recognition of the fact that the assumption of a constant laboratory-wide error is pure fantasy, that every procedure has its own inherent error, and that this is modified by the personal error of the analyst, sampling, and the like. For this reason, specification criteria can be intelligently and fairly established only when and if a valid estimate of experimental error is provided. This is easier said than done.

If we want to determine the experimental error of a procedure, we must ask whether this will be done under "ideal" conditions or under those obtaining in an ordinary laboratory using "normal" precautions; whether to use the most proficient analyst or the journeyman. Merely to state the question is to indicate how difficult it is to implement the decision.

So you see, there is no quick and easy answer to chemical sampling. Each case must be considered by itself. Frequently, a recognized sampling standard for a material of similar characteristics may be used as a guide but considerable technical soul-searching is required before you snatch at this straw. The excellent specifications put out by ASTM, AOAC, etc., are based on long experience with the specified commercial chemicals but each refers only to the specific material covered. They provide excellent guidance - but they are only guides not answers to all problems.

The importance of proper sampling is stressed in many texts on chemical analysis but the advice given is frequently ignored in practice. It is well known that a sample, improperly taken, can vitiate the results obtained by the most competent analyst using the most sophisticated methods

and apparatus available. Yet, in practice, because the actual preparation of a sample usually requires considerable physical exertion, the task is allowed to devolve upon laborers, operating under vague, imprecise instructions which they understand imperfectly, if at all. For example, what do laborers, indeed many professional analysts, know of the special connotations hidden in the deceptively simple requirement "take a random sample?" What do they know of the techniques and tools which must be employed to insure true randomicity?

Ideally, the analyst will be thoroughly trained in the art of taking samples, in seeing and knowing how to overcome the many unforeseen difficulties which arise in every sampling environment. Such a man should take and prepare the sample himself, but this is rarely practical. As an alternative, there is no objection to the expedient of having the sample taken by non-professional personnel provided always they are under the direct, personal supervision of a competent individual. Indeed, if they have been suitably trained in every aspect of the task under the conditions they will face, the continual presence of the supervisor may not be required. However, assurance must be given in all cases; that the individual taking the sample is himself knowledgeable or that he is acting in accordance with the explicit instructions of a competent person, preferably an experienced analyst. All too often environmental changes, not necessarily always meteorological in nature, produce conditions not envisioned by the specification writer, which must be overcome to produce a proper sample. Only a competent, knowledgeable supervisor of sampling personnel can be entrusted with the responsibility for devising necessary additions to and modifications of the prescribed procedure (and documenting these) to insure that a proper sample is taken in the circumstances.

A great weakness in many analytical chemists is their lack of familiarity with the statistical considerations involved in the phenomenon of experimental error. This is not to say that chemists are unaware of or underestimate the importance of experimental error. It is simply the case that so many of them do not know how to handle it or even know it can be handled. Fortunately, modern curricula have replaced old-fashioned, inefficient statistics (e.g., average deviation, etc.) with more modern, efficient concepts such as standard deviation but it remains a matter of concern whether sufficient emphasis has been placed on teaching the student the dangers of bias and how to avoid them, the true meaning of randomicity and how to effect it, the components of variance and how to calculate them and, more generally, how to employ statistics in analytical chemistry.

Sampling error (σ_s) is a significant factor in overall experimental error. When determined as part of a factorial experiment, σ_s will frequently turn out to be surprisingly high as compared with other components of experimental error. For this reason, the reduction of σ_s to a minimum is an important factor in improving chemical testing. To effect this

objective, it is essential to use valid statistical methods to determine σ_s so that alternative methods of sampling may be evaluated by quantitative determination of σ_s and that procedure adopted which has demonstrated the lowest sampling error. It is interesting to note that normal statistical test methods (e.g., analysis of variance) will not only measure σ_s, but will usually identify the causes of error, thus furnishing leads as to what can be done to reduce or eliminate them.

BIBLIOGRAPHY

Japanese Industrial Standard JIS-M8100-1965, "Common Rules for Methods of Sampling Bulk Materials of Mining Products."

ASTM Designation E300-66T, "Tentative Recommended Practice for Sampling Industrial Chemicals."

Duncan, Acheson J., Quality Control and Industrial Statistics, 3rd Edition, 1965.

Juran, J. M., Quality Control Handbook, 2nd Edition 1962.

Zobel, Sigmund P., Some Experience in Laboratory Control Investigation.

COMMENTS ON BULK SAMPLING

Gene R. Lowrimore
Hercules, Inc., Radford Army Ammunition Plant
Radford, Virginia

Professor Duncan, in his presentation, discussed compositing as an integral part of the methodology of bulk sampling. My comments will not be directed to bulk sampling, per se, but, I think they are pertinent to the question of what happens when we composite, or blend. Normally, when we draw a test unit in a bulk sampling situation, we assume that it consists of a very large number of, say, particles. In contrast, if the test unit consisted of only one particle, we would be in a discrete sampling situation.

At Radford Army Ammunition Plant, we manufacture a number of cannon propellants. The smallest identifiable unit of one of these propellants is a grain or fairly large particle, for example, .1" by .8". These propellants are manufactured in a stream of batches and a large number of batches are combined through a blending process to form a lot. Test units are drawn from the lot and, consequently, contain grains from a number of batches.

Because a charge weight correction is made for every lot at firing, the lot mean is of secondary interest to us. The within-lot variance is our primary concern, since it is directly related to landing round after round on target.

We have undertaken a mathematical investigation of the test-to-test or within-lot variability in terms of the batch-to-batch and within-batch variability. In our investigation, we assumed that the true value for the test unit is the sum of the values for the particles making up that test unit. This assumption allowed us to exploit the analogy between this situation and the situation in sample survey theory where we are estimating a total from a stratified sample. We have some results for the case where the number of grains in the test show, N_t, is much greater than the number of batches blended, N_b.

We are currently studying the situation where N_t may actually be less than N_b. All batches cannot now be represented in the test unit. We hope to gain some insight into what happens to the within-lot variability in this case.

These investigations have provided us with valuable insight into the relationship between discrete and bulk sampling and what compositing does in some bulk sampling situations.

SOME STATISTICAL ASPECTS OF ASSURANCE OF STERILIZATION

F. M. Wadley
Consultant to Fort Detrick, Frederick, Maryland

In biological research, we often deal with assurance of sterilization or disinfection, especially in microbiological work and in pest quarantines. We desire assurance that our procedures will give protection against subsequent infection.

Often, we cannot be entirely sure of 100% kills; circumstances of treatment may not be perfect, or the population treated may be very slow in approach to 100% mortality. The probit transformation, widely used in dosage-mortality studies, does not allow mathematically for 100% kill, though it can be approached as closely as desired. Some well-qualified workers in the mortality field prefer to define experimentally a very small risk, which can be accepted. The assurance is then that the probability of any survival is very small indeed, and that with ordinary numbers treated, survival of even one individual will be rare. This viewpoint is discussed by A. C. Baker (1939). It seems more realistic than speaking of 100% kill, and helps to keep preliminary tests to a manageable volume.

For this reason, studies of assurance may deal in very low probabilities; perhaps one survival in thousands or millions. The probabilities are defined by preliminary work, which must obviously be quite extensive and involve great numbers of individuals. Sometimes a limited extrapolation to greater numbers or lower survivals is used. It is desirable to be thorough in preliminary tests without going to a prohibitive amount of work.

Very low percentage counts are involved, and these can be treated as binomially distributed if care is used. The close relation of the Poisson distribution to the binomial can be utilized with some gain in convenience, where percentages are near zero or 100, and numbers are large. For example, suppose an estimate of 3 per 10,000 average survival, or a proportion of 0.0003. Using the binomial estimates of distribution of survivals can be made from several terms of the binomial $(0.0003 + 0.9997)^{10,000}$. Using the Poisson, distribution of survivals can be estimated simply by expanding the Poisson with mean 3, or by looking in published tables. This is true for survival estimates of 3 per 10,000; 3 per 1,000; or 3 per million. Student (1907) showed that the binomial approaches the Poisson at its extreme proportions with $\underline{n}$ large.

A recent inquiry to the Physical Defense Department at Fort Detrick, referred to establishing an assurance that chance of contamination be not over 1 in 1 million. The material in question was a biological fluid to be transported under stressful conditions. The frequently used method

of heat sterilization could not be used because heat would alter the fluid. Filtration was to be used. The treatment is described by Portner, Phillips, and Hoffman (1967).

Extensive tests were made with reusable and disposable filters, dealing with large populations of Serratia marcescens. The best filters gave no survival out of an estimated total of 240,000,000 organisms in replicated trials. Referring to the Poisson, it is found that populations averaging 3 will give an occasional zero; with means of 4 or more, zero is rare. Thus, a tentative maximum of 3 passing per 240,000,000; or 1 for 80 million, is reached. If the sirvivors average 1 in 80 million, and there are only 80 organisms in the material, the chance of only 1 in 1 million is tentatively reached. Other good filters gave occasional survivals of 1 or 2, and seemed to be in the same class.

The material seems likely to have much more than 80 in a typical sample. The solution reached was to use a second filtration, which would seem to give ample assurance. This second filtration also aids in the question of possibly defective filters. An occasional defective filter in a disposable lot, or a proven but deteriorating filter from a reusable lot, is to be avoided. The second filtration with new filters from good lots seems to reduce this hazard to insignificance without an inordinate amount of work.

Another case of use of very small probabilities is given by Baker, in the case of fruit sterilization by moderate heat, to kill fruit fly stages. This was in quarantine work. Populations were estimated by rearing the adults out, both in a check sample and in treated samples. A graded series of time exposures was used, and time was treated as dosage in a probit analysis. Several thousand individuals per dose were used, and probit values up to more than 8 were secured. The lines were extrapolated to estimate dosage required for 9 probits (3 survivors per 100,000), which the author believed to be an acceptable risk.

REFERENCES

Baker, A. C., 1939. The basis for treatment of products where fruit-flies are involved, as a condition for entry into the United States. U. S. Dept. Agr. Circ. 551.

Portner, D. C., Phillips, C. R., and Hoffman, R. K., 1967. Certification of probability of sterilization of liquid by filtration. Appl. Microbiol. 15: 800-807 "Student."

Student., 1907. On the errors of counting with a haemocytometer. Biometrika 5: 351-360.

RESEARCH AND DEVELOPMENT MATHEMATICAL EQUATIONS AS RELATES TO AN ARMY AIRCRAFT SYSTEM

Tony N. O'Truk*
U. S. Army Aviation Materiel Command
St. Louis, Missouri

ABSTRACT. This paper covers the life cycle of the Research and Developmental Phase of an Army aircraft system. It also covers the preparation of mathematical equations as pertains to the hardware under the prototype aircraft, as well as the training, maintenance support, and administration of the prototype aircraft system.

This article has been issued as Technical Report C/A 20-68 as of 11 March 1968. A copy of this paper can be obtained by requesting it from the following address:

Cost Analysis Division
Directorate, Systems and Cost Analysis (Prov)
U. S. Army Aviation Materiel Command
Twelfth and Spruce Streets
St. Louis, Missouri 63166

---

*Mr. O'Truk is no longer at the U. S. Army Aviation Materiel Command. His new address is Belleville Area College, 2555 West Boulevard, Belleville, Illinois 62221.

HYPOTHESES TESTING AND CONFIDENCE INTERVALS FOR PRODUCTS AND QUOTIENTS OF POISSON PARAMETERS WITH APPLICATIONS TO RELIABILITY

Bernard Harris
Mathematics Research Center, United States Army,
The University of Wisconsin, Madison, Wisconsin

ABSTRACT. $X_1, X_2 \ldots, X_{k_1}, Y_1, Y_2, \ldots, Y_{k_2}$ are $k_1 + k_2$ mutually independent Poisson random variables with parameters $\lambda_1, \lambda_2, \ldots, \lambda_{k_1}, \mu_1, \mu_2, \ldots, \mu_{k_2}$ respectively. Confidence intervals and tests of hypotheses for the parameter $\theta = \lambda_1 \lambda_2 \cdots \lambda_{k_1} / \mu_1 \mu_2 \cdots \mu_{k_2}$ are obtained. Under suitable conditions these procedures may be used to obtain approximate confidence intervals and tests of hypotheses of the parameter $\rho = p_1 p_2 \cdots p_{k_1} / p_{k_1+1} p_{k_1+2} \cdots p_{k_1+k_2}$, where the p_i's, $i = 1,2,\ldots, k_1+k_2$ are binomial parameters. This problem is of importance in reliability analysis and some applications to reliability analysis are exhibited.

The remainder of this article has been reproduced photographically from the author's copy. It was issued by the Mathematics Research Center as MRC Technical Summary Report No. 923.

HYPOTHESES TESTING AND CONFIDENCE INTERVALS FOR PRODUCTS AND QUOTIENTS OF POISSON PARAMETERS WITH APPLICATIONS TO RELIABILITY

Bernard Harris

1. Introduction and Summary. Let $X_1, X_2, \ldots, X_{k_1}; Y_1, Y_2, \ldots, Y_{k_2}$ be $k_1 + k_2$ mutually independent Poisson random variables with parameters $\lambda_1, \lambda_2, \ldots, \lambda_{k_1}$; $\mu_1, \mu_2, \ldots, \mu_{k_2}$ respectively. In this paper, we obtain confidence intervals for the parameter $\theta = \lambda_1\lambda_2 \ldots \lambda_{k_1}/\mu_1\mu_2 \ldots \mu_{k_2}$ and the corresponding tests of hypotheses. The required theoretical development is given in section 2. In section 3, we examine the particular case $k_1 = 2$, $k_2 = 0$ because of the specific nature of the answer obtained in this case. In section 4, some of the concrete situations which lead to this problem are pointed out and some numerical illustrations are given. In particular, the reader should note that for $k_2 = 0$, the parameter θ is a product of Poisson parameters and the solution to the present problem can be interpreted as an approximate solution to the corresponding problem of finding confidence intervals for the product of binomial parameters. Estimation of the product of binomial parameters has been investigated by A. Madansky [2] and R. J. Buehler [1]. Their results and methods will be compared with those of the present paper in section 4.

2. Determining Confidence Intervals for θ. The joint distribution of $X_1, X_2, \ldots, X_{k_1}; Y_1, Y_2, \ldots, Y_{k_2}$ is given by

Sponsored by the Mathematics Research Center, United States Army, Madison, Wisconsin, under Contract No.: DA-31-124-ARO-D-462.

(1) $$p(x_1, x_2, \ldots, x_{k_1}; y_1, y_2, \ldots, y_{k_2}; \lambda_1, \lambda_2, \ldots, \lambda_{k_1}; \mu_1, \mu_2, \ldots, \mu_{k_2}) =$$

$$e^{-\sum_{i=1}^{k_1} \lambda_i - \sum_{j=1}^{k_2} \mu_j} \prod_{i=1}^{k_1} \lambda_i^{x_i}/x_i! \prod_{j=1}^{k_2} \mu_j^{y_j}/y_j!$$

where $\lambda_i > 0$, $x_i = 0, 1, 2, \ldots$, $i = 1, 2, \ldots, k_1$; $\mu_j > 0$, $y_j = 0, 1, 2, \ldots$, $j = 1, 2, \ldots, k_2$

Assume $k_1 > 0$. Then, let $U_1 = X_1$, and for $i = 2, 3, \ldots, k_1$, define $U_i = X_i - X_1$; for $j = 1, 2, \ldots, k_2$, define $V_j = Y_j + X_1$. The joint distribution of $U_1, U_2, \ldots, U_{k_1}, V_1, V_2, \ldots, V_{k_2}$ is then given by

(2) $$p_1(u_1, u_2, \ldots, u_{k_1}; v_1, v_2, \ldots, v_{k_2}; \lambda_1, \lambda_2, \ldots, \lambda_{k_1}, \mu_1, \mu_2, \ldots, \mu_{k_2}) =$$

$$e^{-\sum_{i=1}^{k_1} \lambda_i - \sum_{j=1}^{k_2} \mu_j} \left(\lambda_1^{u_1}/u_1! \right) \left(\prod_{i=2}^{k_1} \lambda_i^{u_i+u_1}/(u_i+u_1)! \right) \left(\prod_{j=1}^{k_2} \mu_j^{v_j-u_1}/(v_j-u_1)! \right) ,$$

$u_1 = 0, 1, 2, \ldots$, $u_i = -u_1, -u_1+1, -u_1+2, \ldots$, $i = 2, 3, \ldots, k_1$; $v_j = u_1, u_1+1, u_1+2, \ldots$, $j = 1, 2, \ldots, k_2$.

Consequently, the conditional distribution of U_1 given $U_2 = u_2$, $U_3 = u_3$, $\ldots$, $U_{k_1} = u_{k_1}$; $V_1 = v_1$, $V_2 = v_2$, $\ldots$, $V_{k_2} = v_{k_2}$ is

(3) $$p_\theta(u_1 | u_2, u_3, \ldots u_{k_1}, v_1, v_2, \ldots, v_{k_2}) = \frac{\theta^{u_1 h - 1}(u_2, u_3, \ldots, u_{k_1}, v_1, v_2, \ldots v_{k_2}; \theta)}{u_1! \prod_{i=2}^{k_1} (u_i+u_1)! \prod_{j=1}^{k_2} (v_j - u_1)!} ,$$

where $\max(0, \max_{2 \le i \le k_1} (-u_i)) \le u_1 \le \min_{1 \le j \le k_2} v_j$ and

$$(4)\quad h(u_2,u_3,\ldots,u_{k_1},v_1,v_2,\ldots,v_{k_2};\theta) = \sum_r \theta^r/(r!\prod_{i=2}^{k_1}(u_i+r)!\prod_{j=1}^{k_2}(v_j-r)!)\ ,$$

the sum running from $\max(0,\ \max_{2\le i\le k_1}(-u_i))$ to $\min_{1\le j\le k_2} v_j$. In particular, note that the probability distribution (3) depends only on θ and not on the individual λ's and μ's . Since the probability distribution (3) is a member of a one-parameter exponential family, one and two-sided tests of size α of hypotheses concerning θ can be written down as follows.

To test $H:\theta=\theta_0$, against alternatives $\theta>\theta_0$, reject H if $U_1=k$ and

$$(5)\qquad \sum_{u_1\ge k} p_{\theta_0}(u_1|u_2,u_3,\ldots,u_{k_1},v_1,v_2,\ldots,v_{k_2}) \le \alpha\ .$$

To test $H:\theta=\theta_0$ against alternatives $\theta<\theta_0$, reject H if $U_1=k$ and

$$(6)\qquad \sum_{u_1\le k} p_{\theta_0}(u_1|u_2,u_3,\ldots,u_{k_1},v_1,v_2,\ldots,v_{k_2}) \le \alpha\ .$$

To test $H:\theta=\theta_0$ against the alternative $\theta\ne\theta_0$, reject H if $U_1=k$ and either

$$(7a)\qquad \sum_{u_1\le k} p_{\theta_0}(u_1|u_2,u_3,\ldots,u_{k_1},v_1,v_2,\ldots,v_{k_2}) \le \alpha/2$$

or

$$(7b)\qquad \sum_{u_1\ge k} p_{\theta_0}(u_1|u_2,u_3,\ldots,u_{k_1},v_1,v_2,\ldots,v_{k_2}) \le \alpha/2\ .$$

The tests given by (5) and (6) are uniformly most powerful similar tests. The test given by (7) is similar, but in specifying the right hand sides of both

(7a) and (7b) as $\alpha/2$, this will generally not be a uniformly most powerful similar test. This choice is suggested for ease of computation, since the "optimal" choices for the right hand sides of (7a) and (7b) will depend on $(u_2, u_3, \ldots, u_{k_1}, v_1, v_2, \ldots, v_{k_2})$. It should be noted that since $p_{\theta_0}(u_1 | u_2, u_3, \ldots, u_{k_1}, v_1, v_2, \ldots, v_{k_2})$ is discrete, the tests given above actually are tests of size not exceeding α. In order to produce tests of exact size α, randomized tests will usually have to be employed. The required modifications can easily be carried out.

Confidence intervals of confidence coefficient $1-\alpha$ can be easily obtained for each of the above tests.

Upon observing $U_1 = k$, the $1-\alpha$ upper confidence limit θ_2 for θ conditional on $U_2 = u_2, U_3 = u_3, \ldots, U_{k_1} = u_{k_1}, V_1 = v_1, V_2 = v_2, \ldots, V_{k_2} = v_{k_2}$ is

$$\theta_2 = \sup \{\theta : \sum_{u_1 \leq k} p_\theta(u_1 | u_2, u_3, \ldots, u_{k_1}, v_1, v_2, \ldots, v_{k_2}) \geq \alpha\} . \tag{8}$$

Similarly, the corresponding lower confidence limit θ_1 for θ after observing $U_1 = k$ is

$$\theta_1 = \inf \{\theta : \sum_{u_1 \geq k} p_\theta(u_1 | u_2, u_3, \ldots, u_{k_1}, v_1, v_2, \ldots, v_{k_2}) \geq \alpha\} . \tag{9}$$

From (7a) and (7b), we can obtain a two-sided $1-\alpha$ confidence interval upon observing $U_1 = k$ by

$$P\{\theta_1(k) < \theta < \theta_2(k)\} \geq 1 - \alpha , \tag{10}$$

where

(10a) $$\theta_1(k) = \inf\{\theta: \sum_{u_1 \geq k} p_\theta(u_1 | u_2, u_3, \ldots, u_{k_1}, v_1, v_2, \ldots, v_{k_2}) \geq \alpha/2\}$$

and

(10b) $$\theta_2(k) = \sup\{\theta: \sum_{u_1 \leq k} p_\theta(u_1 | u_2, u_3, \ldots, u_{k_1}, v_1, v_2, \ldots, v_{k_2}) \geq \alpha/2\} \ .$$

If $k_1 = 0$, then $\theta = (\mu_1\mu_2\cdots\mu_{k_2})^{-1}$ and upon defining $\theta^* = \theta^{-1}$, the preceding tests and confidence intervals ((6) through (10)) are readily transformed to provide the corresponding results for this case. That is, let $U_1 = Y_1, U_i = Y_i - Y_1$, $i = 2, 3, \ldots, k_2$. Then, in precisely the same manner as before, we obtain tests for θ^* and confidence intervals for θ^* which are completely equivalent to tests and confidence intervals for θ. These facts are briefly summarized below.

The conditional distribution of U_1 given $U_2 = u_2$, $U_3 = u_3$, ..., $U_{k_2} = u_{k_2}$ is

(11) $$p_{\theta^*}(u_1 | u_2, u_3, \ldots, u_{k_2}) = \frac{\theta^{*u_1} h_1^{-1}(u_2, u_3, \ldots, u_{k_2}; \theta^*)}{u_1! \prod_{i=2}^{k_2} (u_i + u_1)!}, \quad u_1 \geq \max(0, \max_{2 \leq i \leq k_2}(-u_i)) \ ,$$

where

(12) $$h_1(u_2, u_3, \ldots, u_{k_2}; \theta^*) = \sum_r \theta^{*r} / [r! \prod_{i=2}^{k_2} (u_i + r)!] \ ,$$

the sum running from $\max(0, \max_{2 \leq i \leq k_2}(-u_i))$ to ∞. Then a size α test of the hypothesis $H: \theta = \theta_0$ against the alternative $\theta < \theta_0$ is given by the rule: reject H if $U_1 = k$ and

(13) $$\sum_{u_1 \geq k} p_{\theta_0^*}(u_1 | u_2, u_3, \ldots, u_{k_2}) < \alpha \ ,$$

where $\theta_0^* = \theta_0^{-1}$.

Similarly, a test of the hypothesis $H: \theta = \theta_0$ against the alternative $\theta > \theta_0$ is given by the rule: reject H if $U_1 = k$ and

(14) $$\sum_{u_1 \leq k} p_{\theta_0^*}(u_1 | u_2, u_3, \ldots, u_{k_2}) \leq \alpha \ .$$

Finally, to test $H: \theta = \theta_0$ against the alternative $\theta \neq \theta_0$, reject H if $U_1 = k$ and either

(15a) $$\sum_{u_1 \leq k} p_{\theta_0^*}(u_1 | u_2, u_3, \ldots, u_{k_2}) \leq \alpha/2$$

or

(15b) $$\sum_{u_1 \geq k} p_{\theta_0^*}(u_1 | u_2, u_3, \ldots, u_{k_2}) \leq \alpha/2 \ .$$

Upon observing $U_1 = k$ the $1 - \alpha$ upper confidence limit θ_2 for θ conditional on $U_2 = u_2, U_3 = u_3, \ldots, U_{k_2} = u_{k_2}$ is given by

(16) $$\theta_1^* = \inf \{\theta^* : \sum_{u_1 \geq k} p_{\theta^*}(u_1 | u_2, u_3, \ldots, u_{k_2}) \geq \alpha\} \ ,$$

and $\theta_2 = 1/\theta_1^*$.

Similarly, the corresponding lower confidence limit θ_1 for θ after observing $U_1 = k$ is

(17) $$\theta_2^* = \sup \{\theta^* : \sum_{u_1 \leq k} p_{\theta^*}(u_1 | u_2, u_3, \ldots, u_{k_2}) \geq \alpha\} \ ,$$

and $\theta_1 = 1/\theta_2^*$.

Finally, the two-sided $1-\alpha$ confidence interval upon observing $U_1 = k$ is

$$P\{\theta_1(k) < \theta < \theta_2(k)\} \geq 1-\alpha \quad , \tag{18}$$

where

$$\theta_1^{-1}(k) = \sup\{\theta^*: \sum_{u_1 \leq k} p_{\theta^*}(u_1|u_2, u_3, \ldots, u_{k_2}) \geq \alpha/2\} \tag{18a}$$

and

$$\theta_2^{-1}(k) = \inf\{\theta^*: \sum_{u_1 \geq k} p_{\theta^*}(u_1|u_2, u_3, \ldots, u_{k_2}) \geq \alpha/2\} \quad . \tag{18b}$$

Remark. When $k_1 = 0$ we could also have proceeded by letting $V_1 = -Y_1$, $V_j = Y_j - Y_1$, $j = 2, 3, \ldots, k_2$; then the conditional distribution of V_1 given $V_2 = v_2$, $V_3 = v_3, \ldots, V_{k_2} = v_{k_2}$ would depend on $\mu_1, \mu_2, \ldots, \mu_{k_2}$ only through θ. The tests and confidence intervals obtained by repeating the analysis leading to (3) through (10) would give precisely the same results as (11) through (18).

3. Tests and Confidence Intervals for the Product of Two Poisson Parameters. In this section we exhibit some specific properties of the particular case $k_1 = 2$, $k_2 = 0$; that is $\theta = \lambda_1\lambda_2$. In this case,

$$p_\theta(u_1|u_2) = \theta^{u_1}/(u_1!\,(u_1+u_2)!\,h(u_2;\theta)), \quad u_1 \geq \max(0, -u_2) \quad , \tag{19}$$

where

$$h(u_2;\theta) = \sum_{r=\max(0,-u_2)}^{\infty} \theta^r/(r!\,(u_2+r)!) \quad . \tag{20}$$

Define

$$I_\nu(t,\kappa) = (\kappa/2)^\nu \sum_{k=0}^{t} \frac{(\kappa^2/4)^k}{k!\,(\nu+k)!} \quad ,$$

where $I_\nu(\infty, \kappa) = I_\nu(\kappa)$ is the modified Bessel function of order ν .

Then, if $u_2 \geq 0$,

$$h(u_2;\theta) = \sum_{r=0}^{\infty} \theta^r/(r!\,(u_2+r)!) = \theta^{-u_2/2} I_{u_2}(2\sqrt{\theta})$$

and

$$\sum_{u_1=0}^{t} \theta^{u_1}/(u!\,(u_1+u_2)!) = \theta^{-u_2/2} I_{u_2}(t, 2\sqrt{\theta}) \quad .$$

Thus, for t an integer ≥ 0 ,

$$P_\theta\{U_1 \leq t \mid U_2 = u_2\} = I_{u_2}(t, 2\sqrt{\theta})/I_{u_2}(2\sqrt{\theta}) \quad . \tag{21}$$

Similarly, if $u_2 < 0$, let $v = -u_2$; then

$$h(u_2;\theta) = \sum_{r=v}^{\infty} \theta^r/(r!\,(r-v)!) = \sum_{r=0}^{\infty} \theta^{r+v}/(r!\,(r+v)!) = \theta^{v/2} I_v(2\sqrt{\theta}) \quad .$$

Further

$$\sum_{u_1=v}^{t} \theta^{u_1}/(u_1!\,(u_1-v)!) = \sum_{r=0}^{t-v} \theta^{r+v}/(r!\,(r+v)!) = \theta^{v/2} I_v(t-v, 2\sqrt{\theta}) \quad .$$

Thus, for t an integer $\geq -u_2$,

$$P_\theta\{U_1 \leq t \mid U_2 = u_2\} = I_{-u_2}(t+u_2, 2\sqrt{\theta})/I_{-u_2}(2\sqrt{\theta}) \quad . \tag{22}$$

Combining (21) and (22), the cumulative distribution function of U_1 given $U_2 = u_2$ is

$$(23) \qquad P_\theta\{U_1 \leq t \mid U_2 = u_2\} = \begin{cases} I_{u_2}(t, 2\sqrt{\theta})/I_{u_2}(2\sqrt{\theta}), & u_2 \geq 0 \\ I_{-u_2}(t+u_2, 2\sqrt{\theta})/I_{-u_2}(2\sqrt{\theta}), & u_2 < 0 \end{cases},$$

where t is an integer ≥ 0, if $u_2 \geq 0$, and an integer $\geq -u_2$, if $u_2 < 0$.

It seems natural to name this distribution the "incomplete modified Bessel function:.

Returning to the tests and confidence intervals given earlier, the $1-\alpha$ upper confidence limit θ_2 for θ conditional on $U_2 = u_2$ may be written

$$(24a) \qquad \theta_2 = \sup\{\theta : I_{u_2}(k, 2\sqrt{\theta})/I_{u_2}(2\sqrt{\theta}) \geq \alpha\}, \qquad u_2 \geq 0$$

and

$$(24b) \qquad \theta_2 = \sup\{\theta : I_{-u_2}(k+u_2, 2\sqrt{\theta})/I_{-u_2}(2\sqrt{\theta}) \geq \alpha\}, \quad u_2 < 0,$$

where k is the observed value of U_1.

The other confidence intervals and tests given in (5) to (10) admit of similar representations, which will not be explicitly given here. In this case, it is also quite reasonable to tabulate this distribution and we hope to produce such a tabulation in the near future.

4. Applications. Despite the fact that the problem of hypothesis testing or confidence intervals for the parameter $\theta = \lambda_1\lambda_2 \cdots \lambda_{k_1}/\mu_1\mu_2 \cdots \mu_{k_2}$, where

λ_i, μ_j, $i = 1, 2, \ldots, k_1$, $j = 1, 2, \ldots, k_2$ are each Poisson parameters, may arise as a problem of interest in its own right, the procedures described in this paper may be of more interest and will presumably be applied more often as approximate techniques for statistical inference questions concerning products and quotients of binomial parameters. We proceed to give some illustrations of this usage. Throughout the subsequent discussion we shall assume that the relevant parameters of all binomial distributions being considered are such that the Poisson approximation to the binomial distribution is satisfactory to the user.

Consequently, assume that we have $k_1 + k_2$ binomial populations with parameters $(n_1, p_1), (n_2, p_2), \ldots, (n_{k_1}, p_{k_1}), (n_{k_1+1}, p_{k_1+1}), (n_{k_1+2}, p_{k_1+2}), \cdots, (n_{k_1+k_2}, p_{k_1+k_2})$ respectively and that the mutually independent binomial random variables $X_1, X_2, \ldots, X_{k_1}, X_{k_1+1}, X_{k_1+2}, \ldots, X_{k_1+k_2}$ have been observed. Then, let

$$\rho = \frac{p_1 p_2 \cdots p_{k_1}}{p_{k_1+1} p_{k_1+2} \cdots p_{k_1+k_2}}, \quad p_i > 0, \quad i = 1, 2, \ldots, k_1 + k_2 . \tag{25}$$

Replace $n_i p_i$ by λ_i, $i = 1, 2, \ldots, k_1$, and for $i = k_1+1, k_1+2, \ldots, k_1+k_2$ replace $n_i p_i$ by μ_j, where $j = i - k_1$. Then, assuming that $X_1, X_2, \ldots, X_{k_1+k_2}$ are each approximately Poisson distributed, we have from (10),

$$P\{\theta_1 < \theta = \frac{\lambda_1 \lambda_2 \cdots \lambda_{k_1}}{\mu_1 \mu_2 \cdots \mu_{k_2}} < \theta_2\} \sim 1 - \alpha . \tag{26}$$

This is equivalent to

$$P\{\theta_1 < \prod_{i=1}^{k_1} n_i p_i \Big/ \prod_{i=k_1+1}^{k_1+k_2} n_i p_i < \theta_2\} \sim 1 - \alpha . \tag{27}$$

and from (27), we obtain an approximate confidence interval for ρ by

$$(28) \qquad P\{\theta_1 \prod_{i=k_1+1}^{k_1+k_2} n_i / \prod_{i=1}^{k_1} n_i < \rho < \theta_2 \prod_{i=k_1+1}^{k_1+k_2} n_i / \prod_{i=1}^{k_1} n_i\} \sim 1-\alpha \ .$$

The process for getting approximate upper (lower) confidence limits for ρ is quite similar to the derivation of (28) and will not be explicitly stated here. In addition, in testing hypotheses, we clearly have that a test of any hypotheses concerning θ is an approximate test for the corresponding hypotheses for ρ .

We now turn to some concrete illustrations.

In reliability analysis, a mechanism may fail if and only if each of k components fail. Let E_i be the event that the ith component fails, $i = 1,2,\ldots,k$ and assume that the events E_i are mutually independent. Then, the probability of failure $= P(\bigcap_{i=1}^{k} E_i) = \prod_{i=1}^{k} P(E_i) = \prod_{i=1}^{k} p_i$. If each component is tested separately in n_i Bernoulli trials, and if the p_i's are "small" and the n_i's are "large", then (28) or the equivalent formula for upper (lower) confidence limits for ρ applies. For this problem R. J. Buehler [1] gave a procedure employing a Poisson approximation. However, Buehler's procedure does not readily extend to products of more than two binomial parameters without introducing extensive computational difficulties. On the other hand, for $k \geq 2$, the series (4) introduced in this paper, whose individual terms give the conditional distribution (when normalized by (4)), converges more rapidly than the exponential series and can be easily evaluated in any specific case by hand computation. The individual terms can each be computed recursively. A. Madansky [2] employed the likelihood ratio statistic $L(\rho)$ and used the approximate distribution theory, namely that

$-2 \log L(\rho)$ has asymptotically the χ^2 distribution with one degree of freedom. He compared this with the approximate confidence regions that would be obtained by "linearization" methods. Madansky also noted that the application of the asymptotic distribution theory for either the likelihood ratio statistic or the "linearized" statistic is not too satisfactory for the case of very high reliabilities. However, this last concern is precisely what motivated the present investigation.

To see how one may obtain ratios $(k_2 > 0)$, we state the specific problem which was posed to the author. Let E_1, E_2, E_3, E_4 be arbitrary events. A confidence interval for $P(E_2 \cap E_3 \cap E_4 | E_1)$ is required, which we write as

$$(29)\quad P(E_2 \cap E_3 \cap E_4 | E_1) = P(E_1 \cap E_2)\, P(E_3 | E_1 \cap E_2)\, P(E_4 | E_1 \cap E_2 \cap E_3)/P(E_1) \ .$$

Separate sequences of Bernoulli trials are conducted for each of the four factors in (29). Thus, we seek to obtain a confidence interval for a parameter of the form $\rho = p_2 p_3 p_4 / p_1$, and (28) applies. In this illustration, we have $k_2 = 1$; clearly, the above illustration can be extended to exhibit experiments with other values for k_2 .

Experiments such as the type leading to (29) are useful in situations requiring very high reliability, inasmuch as the conditioning appearing in terms like $P(E_4 | E_1 \cap E_2 \cap E_3)$ may be needed in order that the probability of occurrence of a failure will be sufficiently high so that a failure may be observable in a moderate number of trials. In addition, this type of experiment may also be used to eliminate the need for assuming independence in reliability problems. However, it does introduce the difficulty of requiring conditional experiments.

Acknowledgements. This problem has been suggested by Mr. Allan Albright of Watervliet Arsenal, New York and Mr. G. T. Dobrindt of Aberdeen Proving Grounds.

REFERENCES

[1] Buehler, R. J., "Confidence Intervals for the Product of Two Binomial Parameters," Journal of the American Statistical Association, 52 (1957), 482-493.

[2] Madansky, A., "Approximate Confidence Limits for the Reliability of Series and Parallel Systems", Technometrics, 7 (1965), 495-503.

METHODOLOGY OF ASSESSMENT OF BIOCELLULAR PERFORMANCE

George I. Lavin

Terminal Ballistic Laboratories, Aberdeen Research and Development Center, Aberdeen Proving Ground, Maryland

ABSTRACT. Our laboratory is interested in problems which are concerned with the assessment of the effect of absorbed energy on the efficiency of performance of bio-cellular systems as modified by the absorption of external energy. The type of specific, non-destrictive analytical procedures which are designed for this purpose and which have been the subject of previous presentations to The Design of Experiments in Army Research Development and Testing, are listed below.

Microscopy - A spectrum line (Mercury 2537 A°) is used as the light source - for better optical resolution.

Spectroscopy - A continuous light source (hydrogen discharge tube) together with a spectrograph of low dispersion. The combination allows the detection and identification of large molecules in a mixture.

Microspectroscopy - Both sources are used. The line source for miscoscopical structure. The continuous source brings our absorption band details which is needed for compound (amino acids etc.) differentiation and identification.

Model Simulation - A three dimensional model is described which simulates the action of an animal which senses the presence of an object and then reaches for it. The many unrealities of task performance of this model are pointed out. These include the lack of biochemical reality which means no biochemical feedback with no replacement of material as action performance continues.

Biochemistry of Tissue Systems - The relationship of specific task performance to the chemical composition of the particular tissue system. Subjects considered: Proteins, Nucleic Acids, Lipoids, Carbohydrates, Polysaccharides, Enzymes, etc.

Bionics and Cybernetics - A consideration of the application of systems analysis in relation to animal performance. Feedback effects.

Mechanism of Energy Absorption by Cellular Systems - An analog is drawn between the origin of optical spectra and the amount of energy absorbed by the system on exposure to ultraviolet, visible or infrared radiation.

Consequences of Energy Absorbed by Biocellular Systems - Initiation of atom and free radical chain reactions which result in the formation of wound tracts and stress. Levels of damage.

The last presentation was a summary of the above.

MONTE CARLO INVESTIGATION OF THE ROBUSTNESS OF DIXON'S CRITERIA FOR TESTING OUTLYING OBSERVATIONS

Jerry Thomas
Surveillance and Reliability Laboratory
Aberdeen Research and Development Center
Aberdeen Proving Ground, Maryland

ABSTRACT. An investigation of the effect of non-normality on the distribution of Dixon's criteria for detecting outlying observations is presented here. Monte Carlo techniques were used to determine the distributions of the Dixon statistics when observations are selected from specific non-normal distributions with varying degrees of abnormality. Two such distributions whose degree of abnormality, as determined by the coefficient of skewness, may be varied by changes in the parameters of the distributions are the beta and gamma distributions.

A measure of the lack of robustness, that is the sensitivity to departures from normality, in the Dixon criteria may be determined by comparison of the frequency distributions of the Dixon type statistics computed from sampling the non-normal distributions with those values obtained by Dixon when sampling from the normal distribution.

Based on the distributions of the Dixon statistics computed from the non-normal distributions, it has been shown that Dixon's criteria is not robust and its wide use may result in incorrect decisions when the underlying distribution is asymmetric or skewed.

I. INTRODUCTION. After experimental data has been collected, and before it can be analyzed, the observations must be carefully screened to determine if they come from the same population. If any of these observations appear to be radically different from the majority of the other values obtained in the experimentation, it is necessary to determine if the suspect value is an extreme value or an outlying observation (commonly called an outlier). By an outlier, we mean an observation that did not come from the same population as the remaining values. In order to do this, a knowledge of the testing procedures, the manner in which the data was collected and recorded, and some prior knowledge as to what the range of the observations should be, are very helpful in deciding whether a value should be retained in the analyses or be thrown out as an outlier.

To be consistent in this process, statistical procedures have been developed to determine whether a value is an outlier or not. One of these procedures was developed by W. J. Dixon (1). Dixon's statistics have the advantage of being easily computed and are thus widely used in applied statistics. However, Dixon's statistics were developed for

normally distributed variates. The question was posed as to whether or not Dixon's tests for outliers were robust tests. By this, we mean, are the tests insensitive to deviations from normality. In order to check the robustness of Dixon's test statistics, the coefficient of skewness was chosen to measure the degree of departure from normality. Two distributions whose coefficients of skewness may be varied by changes in the parameters of the distributions are the beta and gamma distributions. Thus, these two distributions were chosen to be used in this paper.

II. TEST OF ROBUSTNESS OF DIXON'S CRITERIA.

2.1 Definitions of Statistics to be Investigated. The four statistics proposed by Dixon for testing extreme values are defined below, where the X's are the observed values from a normal distribution arranged in ascending order such that, $X_1 \leq X_2 \leq X_3 \ldots\ldots\ldots \leq X_{n-1} \leq X_n$.

For a single outlier, X_1

$$r_{10} = \frac{X_2 - X_1}{X_n - X_1} \qquad [1a]$$

or for a single outlier, X_n

$$r_{10} = \frac{X_n - X_{n-1}}{X_n - X_1} \qquad [1b]$$

For a single outlier X_1, avoiding X_n

$$r_{11} = \frac{X_2 - X_1}{X_{n-1} - X_1} \qquad [2a]$$

or for a single outlier X_n, avoiding X_1

$$r_{11} = \frac{X_n - X_{n-1}}{X_n - X_2} \qquad [2b]$$

For outlier X_1, avoiding X_2 and X_n

$$r_{21} = \frac{X_3 - X_1}{X_{n-1} - X_1} \qquad [3a]$$

or for outlier X_n, avoiding X_1 and X_{n-1}

$$r_{21} = \frac{X_n - X_{n-2}}{X_n - X_2} \qquad [3b]$$

For outlier X_1, avoiding X_2, X_{n-1} and X_n

$$r_{22} = \frac{X_3 - X_1}{X_{n-2} - X_1} \qquad [4a]$$

or for outlier X_n, avoiding X_1, X_2, and X_{n-1}

$$r_{22} = \frac{X_n - X_{n-2}}{X_n - X_3} \qquad [4b]$$

These computations are widely used in applied statistics. One of the main advantages in using these statistics is the ease with which these tests for outliers may be performed. It is a simple matter, especially for small samples, to visually order the data such that the values needed for the test statistic, i.e., X_1, X_2, X_{n-1}, X_n, can be determined. Then using these values, $r_{j,\ i-1}$ is computed and compared to the critical value listed in tables that are readily available. If $r_{j,\ i-1}$ (the computed value) is greater than R (the critical value), at the desired risk level, α, then X_k (k = 1 or n) is determined to be an outlier with $1 - \alpha$ confidence.

Since Dixon's critical values were derived using the normal distribution, the question was posed as to how departure from normality would affect these critical values. In order to investigate this, the Pearson Type I curve (2) or beta distribution was chosen as the underlying distribution. This distribution was used with various α's and β's to give distributions with various degrees of skewness.

Dixon computed the distribution of the ratio, $r_{j,\,i-1} = (X_n - X_j) / (X_n - X_i)$ using the following function:

$$\int_{-\infty}^{\infty}\int_{0}^{\infty} \frac{n!}{(i-1)!\,(n-j-i-1)!} \left(\int_{-\infty}^{x-v} f(t)\,dt \right)^{i-1} f(x-v) \cdot \left(\int_{x-v}^{x-rv} f(t)\,dt \right)^{n-j-i-1} \cdot f(x-rv)\, f(x) \left(\int_{x-rv}^{x} f(t)\,dt \right)^{j-1} dvdx \qquad [5]$$

Where $j = 1, 2$; $i = 1, 2, 3$; $v = X_n - X_i$; $rv = X_n - X_j$; $X = X_n$

$f(t) = \frac{1}{\sqrt{2\pi}} e^{-\frac{t^2}{2}}$. If instead of the normal distribution, the beta distribution is used in [5], the following function is obtained:

$$\int_{0}^{1}\int_{0}^{x} \frac{n!}{(i-1)!\,(n-j-i-1)!\,(j-1)!} \left(\int_{0}^{x-v} f(t)\,dt \right)^{i-1} \cdot f(x-v) \cdot \left(\int_{x-v}^{x-rv} f(t)\,dt \right)^{n-j-i-1} \cdot f(x-rv)\, f(x) \left(\int_{x-rv}^{x} f(t)\,dt \right)^{j-1} dvdx \qquad [6]$$

Where $j = 1, 2$; $i = 1, 2, 3$; $v = X_n - X_i$; $rv = X_n - X_j$; $X = X_n$ and

$f(t) = \frac{(\alpha + \beta + 1)!}{\alpha!\,\beta!}\, t^{\alpha}\, (1-t)^{\beta}$.

It is apparent that this integration is very difficult for sample sizes of n = 3 and becomes more difficult as n increases. In fact Dixon used numerical integration for only a few sample sizes and interpolated to obtain the remaining values. Thus, due to the problems of integration and the fact that Mowchan (3) has demonstrated that using Monte Carlo techniques for obtaining the distributions of the $r_{j,\,i-1}$'s were very accurate, it was decided that Monte Carlo techniques would be used in this paper.

2.2 Monte Carlo Techniques. In order to use Monte Carlo techniques, it was necessary to draw random samples from the beta distribution. Since beta random numbers are not usually readily available in the form of subroutines, the following method was used.

The Ballistic Research Laboratories Electronic Scientific Computer (BRLESC) at Aberdeen Proving Ground, Maryland was used to generate a random number, y, from the uniform distribution over the unit interval. This uniform random number, y, was considered to be the area of interest from a cumulative distribution, F (X). The cumulative form of the distribution was integrated from 0 to X, where X is the point on the distribution that would define an area equal to y. For the beta distribution this is as follows:

$$F(X) = \begin{cases} 0 & X < 0 \\ \displaystyle\int_0^X \frac{(\alpha + \beta + 1)!}{\alpha!\ \beta!}\, t^{\alpha}\ (1-t)^{\beta}\ dt & 0 < X < 1 \\ 1 & X > 1 \end{cases}$$

thus

$$= \int_0^X \frac{(\alpha + \beta + 1)!}{\alpha!\ \beta!}\, t^{\alpha}\ (1-t)^{\beta}\ dt\,.$$

This procedure for generating X's was used to obtain samples of size n = 6, 10, and 15 for this paper.

2.3 Determination of Critical Values. An extreme value may occur as either a high value or a low value. Thus, since the beta distribution is generally not symmetric, it was necessary to construct test criteria for testing either high or low values. To do this, both forms of equations [1] through [4] were used.

Six hundred samples of size n were drawn. Each sample of size n was ordered such that $X_1 \leq X_2 \ldots \leq X_n$. Then using the appropriate X's the test statistics were computed using each of the formulas to obtain the $r_{j,\ i-1}$'s. After 600 $r_{j,\ i-1}$'s were obtained for each test statistic, the cumulative distribution of these $r_{j,\ i-1}$'s was constructed. Various percentiles were computed ranging from 10 to 99.5. These percentiles, along with Dixon's percentiles from the normal distribution (4) are

given in tables I through XII. These percentiles are given in terms of α, where α is equal to one minus the various percentiles. Thus α is equal to the significance level of the test at which the suspect outlier is being tested, with the values in the tables being the critical values at the given significance level. These critical values are tabulated for r_{10}, r_{11}, r_{21}, and r_{22} for both upper and lower tails using the following parameters of the beta distribution with their skewness coefficient, γ_1, as identification.

α	β	γ_1
5	5	0
7	4	-0.24
8	3	-0.42
9	2	-0.64
10	1	-0.96
19	1	-1.14

These skewness coefficients were computed using $\gamma_1 = \dfrac{E\,(X - \mu)^3}{\sigma^3}$.

Where $E\,(X - \mu)^3 = \mu'_3 - 3\,\mu'\,\mu'_2 + 2\,(\mu'_1)^3$. For the beta distribution

$$E\,(X - \mu)^3 = \left[\frac{(\alpha + 1)\,(\alpha + 2)\,(\alpha + 3)}{(\alpha + \beta + 2)\,(\alpha + \beta + 3)\,(\alpha + \beta + 4)}\right] - 3\left[\frac{(\alpha + 1)}{(\alpha + \beta + 2)}\right]\left[\frac{(\alpha + 1)\,(\alpha + 2)}{(\alpha + \beta + 2)\,(\alpha + \beta + 3)}\right] + 2\left[\frac{(\alpha + 1)^3}{(\alpha + \beta + 2)^3}\right]$$

and

$$\sigma^3 = \left[\frac{(\alpha + 1)\,(\alpha + 2)}{(\alpha + \beta + 2)\,(\alpha + \beta + 3)} - \frac{(\alpha + 1)^2}{(\alpha + \beta + 2)^2}\right]^{3/2}$$

These values of α and β were chosen so as to give various degrees of skewness. When $\alpha = \beta$ then the beta distribution is symmetric and the skewness coefficient is equal to zero. In order to minimize computer time, it was desired to keep the sum of $\alpha + \beta$ as small as possible, since as this sum increases, so does the computing time. However, it was desired to get various degrees of skewness, thus α was increased and β decreased. By choosing to do this, negative skewness coefficients were obtained. Positive skewness coefficients could have been obtained by increasing β and decreasing α. However, the only difference a positive skewness coefficient would make is that the skewed tail would be on the right instead of on the left. Thus, if the skewness coefficients were positive the upper and lower tail values would be reversed.

Since the beta variates range in value only from 0 to 1, the question arises as to how sampling from a distribution which has an infinite limit on one tail would affect the critical values. Thus, the Pearson Type III Curve or gamma distribution, which has as its limits 0 to ∞, was chosen.

The cumulative distribution for the gamma distribution is

$$F(X) = \begin{cases} 0 & X < 0 \\ \displaystyle\int_0^X \frac{1}{\alpha!\ \beta^{\alpha+1}}\ t^{\alpha}\ e^{-t/\beta} dt & X > 0 \end{cases}$$

with $\alpha > -1$ and $\beta > 0$. Since a change in β only changes the scale and not the general shape of the curve, without loss of generality $\beta = 1$ was used with $\alpha = 0, 1, 2, 3, 4, 5$. γ_1 was computed for the gamma distribution using $\gamma_1 = \dfrac{E(X-\mu)^3}{\sigma^3}$

where

$$E(X-\mu)^3 = (\alpha + 1)(\alpha + 2)(\alpha + 3) - 3(\alpha + 1)^2(\alpha + 2) + 2(\alpha + 1)^3$$

and

$$\sigma^3 = (\alpha + 1)^{3/2}$$

β	α	γ_1
1	0	2.00
1	1	1.41
1	2	1.15
1	3	1.00
1	4	0.84
1	5	0.82

The same general techniques described previously were used. Again 600 samples of sizes n = 6, 10, and 15 were drawn from the gamma distributions. Both forms of equations [1] through [4] were used in computing the test statistics since the gamma distribution is also usually not symmetric but skewed.

The cumulative distributions of these test statistics were formed and the critical values were computed for the various levels of confidence. These critical values from the normal distribution for the same levels of confidence, are given in tables XII through XXIV.

From the Dixon Statistics that were computed using the beta and gamma distributions, it is apparent that for a given confidence level the critical values in the skewed tail (the lower tail for the beta distributions and the upper tail for the gamma distributions) increase as the absolute value of γ_1 increases. Vice versa, in the tail opposite the skewness, the critical values tend to decrease as the absolute value of γ_1 decreases.

The reason for this might be described in the following manner:

Let us look at the Dixon test which uses the statistic r_{10} where

$$r_{10} = \frac{X_n - X_{n-1}}{X_n - X_1}$$ for the test of an observation that appears to be

larger than the other observations in the sample. It is obvious that for r_{10} to become smaller, the numerator (the difference between the largest and the next largest observation) must become smaller faster than the denominator (the difference between the largest and the smallest observation). It should also be noted that for the beta and gamma variates used in this paper the absolute value of γ_1 increases as the variance decreases.

Beta Distribution		Gamma Distribution	
γ_1	σ^2	γ_1	σ^2
0	.019	0.82	6.0
-0.24	.017	0.89	5.0
-0.42	.015	1.00	4.0
-0.64	.013	1.15	3.0
-0.96	.009	1.41	2.0
-1.14	.004	2.00	1.0

This can be intuitively demonstrated when considering the fact that as the skewness increases, the distribution becomes clustered at one end of the range of the distribution, with only a small portion of the distribution lying in the skewed tail. For example, using the beta distribution with a large skewness coefficient, let the suspect outlier to be a value larger than the other sample observations. Thus, the majority of the values are generally clustered in the upper tail, close to the upper limit of one. It would, therefore, be very unlikely for the difference between the largest and the second largest observation to be very large. On the other hand, since the skewed tail of the distribution goes to 0, it is likely that in a sample, at least one of the observations will be small in comparison with the other samples. Therefore, when a distribution is markedly skewed, it is expected that the values of r_{10} will be small for the skewed tail.

The critical values obtained using the beta and gamma distributions were compared to Dixon's critical values by using the Kolmogorov-Smirnov statistic (5). The empirical distributions were tested against those derived by Dixon and the level at which these tests of equality were rejected is given in tables XXV and XXVI. The distributions of the $r_{j,\ i-1}$ were listed as not significantly different from those obtained by Dixon for the normal distribution at the .10 level.

It can be seen for the beta distribution variates, that the significance level generally decreases as the absolute value of the skewness coefficients increase. For gamma distribution variates, the significance level is generally .01 for all degrees of skewness. Since the distributions of the $r_{j,\ i-1}$'s obtained using the beta and gamma distributions are significantly different from those obtained by Dixon using the normal distribution, some examples are given to show how it is possible to make the wrong decision in deciding whether or not an observation is an outlier.

2.4 Examples. Suppose for example we had the following observation from a beta distribution with a skewness coefficient of $\gamma_1 = -0.42$

$X_1 = .2319$ $X_2 = .6516$ $X_3 = .7453$

$X_4 = .7555$ $X_5 = .8547$ $X_6 = .9690$

Let X_1 be our suspect outlier and using $r_{10} = \frac{.6516 - .2319}{.9690 - .2319} = 0.569$. Comparing this with Dixon's critical value of 0.560, we would designate X_1 as an outlier at the .05 level of significance. However, using the critical values in table 1, Lower Tail under $\gamma_1 = -0.42$, we see that the critical value is 0.6085 at the .05 significance level. Thus, X_1 would not be an outlier.

As the second example, take the 10 observations drawn from a beta distribution with $\gamma_1 = -0.64$

$X_1 = 0.2306$ $X_2 = 0.3312$ $X_3 = 0.4317$

$X_4 = 0.4814$ $X_5 = 0.5489$ $X_6 = 0.5806$

$X_7 = 0.6548$ $X_8 = 0.6637$ $X_9 = 0.73626$

$X_{10} = 0.9701$

Using $r_{11} = \frac{X_{10} - X_9}{X_{10} - X_2}$ as our test statistic, we test X_{10} to see if it is an outlier.

$$r_{11} = \frac{.9701 - .7363}{.9701 - .3312} = .3660$$

Using Dixon's criteria, X_{10} would not be an outlier at the .05 significance level. However, using the critical value listed in table VI, Upper Tail, under -0.64, we see that its critical value is 0.3466. Thus, X_{10} would be an outlier.

For example three, let us look at a sample drawn from a gamma distribution with $\gamma_1 = 1.15$.

$X_1 = 0.4790$ $X_2 = 0.9628$ $X_3 = 1.4398$

$X_4 = 1.8540$ $X_5 = 2.5660$ $X_6 = 2.8963$

$X_7 = 3.4193$ $X_8 = 3.6188$ $X_9 = 6.6278$

$X_{10} = 9.0$[illegible]

Using $r_{21} = \frac{X_{10} - X_8}{X_{10} - X_2}$ to test X_{10}, we get $r_{21} = \frac{9.0973 - 3.6188}{9.0973 - 0.9628} = 0.6735$ which is significant at the .05 significance level, using Dixon's critical value of 0.612. However, using table XIX, Upper Tail, under $\gamma_1 = 1.15$, we see that the critical value at this .05 significance level is 0.7514. Thus, X_{10} would not be an outlier.

As example four, take a sample of size n = 15 from a gamma distribution with $\gamma_1 = 1.15$.

$X_1 = 0.2129$ $X_2 = 1.1867$ $X_3 = 2.3271$

$X_4 = 2.7486$ $X_5 = 2.8934$ $X_6 = 3.0924$

$X_7 = 3.4674$ $X_8 = 3.6631$ $X_9 = 3.8998$

$X_{10} = 4.1009$ $X_{11} = 4.3123$ $X_{12} = 4.5184$

$X_{13} = 5.6396$ $X_{14} = 5.7802$ $X_{15} = 6.0301$

Using X_1 as our suspect outlier, and $r_{22} = \frac{X_3 - X_1}{X_{13} - X_1}$ we get

$r_{22} = \frac{2.3271 - 0.2129}{5.6396 - 0.2129} = 0.3896$. Dixon's critical value at .05 significance level is 0.525. Thus, X_1 would not be an outlier at the .05 significance level. However, using table XXIV, Lower Tail, under $\gamma_1 = 1.15$, we see that the critical value is 0.353. Thus, X_1 would be an outlier at the .05 significance level.

From these examples, it is easy to see that there are two types of errors that can be made if the sample observations are not from a normal distribution and if Dixon's critical values are used for testing extreme values. These values can be called outliers when, in fact, they are not outliers at the chosen significance level or they can be outliers at a chosen significance level and not be so designated. Thus, from these examples, it can be seen that the Type I or α errors, i.e., the rejection of the hypothesis when it is in fact true and the Type II or β errors, i.e., the acceptance of the hypothesis when it is false, are not what they are specified to be when operating under the assumption of normality when in fact, the observations come from a non-normal distribution.

III. CONCLUSIONS. It has been shown, on an empirical basis, and using the Kolmogorov-Smirnov goodness-of-fit test that there is a difference in the cumulative distributions of the $r_{j, i-1}$ statistic obtained using the normal distribution as opposed to distributions that are non-normal. These differences are usually significant at a low risk level.

Also, it has been shown that the effect of departure from normality is dependent on whether the suspect outlier is a large or small value. Thus, it is necessary to have critical values for testing either large or small values.

It is also evident that the degree of skewness of the distributions affects the critical values. That is, these critical values tend to depart more from those values derived by Dixon for the normal distribution as the skewness increases.

For a symmetric distribution, that is, one for which the skewness coefficient is zero, Dixon's criteria is robust. However, as the distribution becomes asymmetric and the absolute value of the skewness coefficient increases, Dixon's criteria becomes less robust.

BIBLIOGRAPHY

1. Dixon, W. J., Analysis of Extreme Values, Annals of Math. Stat., Vol. 21 (1950, pp 488 - 506).

2. Mood, A. M., Introduction to the Theory of Statistics, McGraw-Hill Book Co., Inc., New York, 1950.

3. Mowchan, W. L., "Monte Carlo Investigation of the Probability Distributions of Dixon's Criteria for Testing Outlying Observations," Ballistic Research Laboratories, Memo. Report No. 1716, 1966.

4. Dixon, W. J., Ratios Involving Extreme Values, Annal of Math. Stat., Vol. 22 (1951, pp 68 - 78).

5. Lindgren, B. W., Statistical Theory, The Macmillan Co., New York, 1960.

APPENDIX A

TABLES OF PERCENTAGE POINTS OF DIXON'S CRITERIA FROM BETA AND GAMMA DISTRIBUTIONS

TABLE I

BETA DISTRIBUTION

N = 6

$PR(r_{10} > R) = \alpha$

UPPER TAIL

α/γ_1	D.C.*	0.00	-0.24	-0.42	-0.64	-0.96	-1.14
.900	.038	.0435	.0378	.0244	.0243	.0215	.0199
.800	.079	.0809	.0668	.0530	.0560	.0453	.0430
.700	.121	.1214	.1077	.0888	.0895	.0713	.0694
.600	.164	.1639	.1427	.1280	.1226	.0972	.0908
.500	.210	.2089	.1827	.1650	.1596	.1250	.1214
.400	.261	.2479	.2147	.2083	.2011	.1600	.1507
.300	.318	.3109	.2733	.2572	.2486	.1923	.1972
.200	.386	.3703	.3399	.3271	.3112	.2429	.2686
.100	.482	.4797	.4280	.4260	.4180	.3201	.3663
.050	.560	.5756	.5043	.4984	.5045	.4014	.4386
.020	.644	.6453	.6014	.5810	.5903	.5334	.5806
.010	.698	.6865	.7085	.6515	.6039	.5896	.6342
.005	.740	.7289	.7444	.6716	.6124	.6304	.6601

LOWER TAIL

α/γ_1	D.C.	0.00	-0.24	-0.42	-0.64	-0.96	-1.14
.900	.038	.0353	.0376	.0537	.0474	.0530	.0508
.800	.079	.0753	.0837	.0997	.1074	.1137	.1130
.700	.121	.1160	.1339	.1412	.1614	.1773	.1700
.600	.164	.1559	.1817	.1810	.2046	.2394	.2363
.500	.210	.1934	.2311	.2274	.2715	.3072	.2948
.400	.261	.2517	.2787	.2899	.3317	.3750	.3575
.300	.318	.3124	.3276	.3418	.4004	.4369	.4364
.200	.386	.3667	.4079	.4253	.4729	.5135	.5203
.100	.482	.4637	.4975	.5363	.5555	.6045	.6258
.050	.560	.5416	.5800	.6085	.6296	.6760	.6868
.020	.644	.6264	.6696	.7150	.7011	.7472	.7565
.010	.698	.6889	.7125	.7491	.7252	.7794	.8308
.005	.740	.7161	.7571	.8002	.7600	.8249	.8571

* Dixon's critical values from normal distribution

TABLE II

BETA DISTRIBUTION

N = 6

$PR(r_{11} > R) = \alpha$

UPPER TAIL

α/γ_1	D.C.	0.00	-0.24	-0.42	-0.64	-0.96	-1.14
.900	.056	.0584	.0510	.0366	.0403	.0312	.0327
.800	.113	.1074	.0987	.0861	.0951	.0715	.0695
.700	.169	.1647	.1559	.1274	.1391	.1105	.1109
.600	.227	.2192	.2007	.1786	.1815	.1579	.1500
.500	.288	.2746	.2538	.2356	.2334	.1961	.1994
.400	.350	.3419	.3087	.3044	.3022	.2456	.2460
.300	.420	.4035	.3704	.3566	.3644	.3084	.3099
.200	.502	.4905	.4585	.4474	.4526	.3789	.3906
.100	.609	.6250	.5621	.5490	.5610	.4838	.5049
.050	.689	.6865	.6345	.6513	.6272	.5957	.5934
.020	.763	.7818	.7146	.7322	.7489	.6832	.7051
.010	.805	.8203	.8122	.7815	.7750	.7773	.7651
.005	.839	.8369	.8446	.8166	.8434	.8016	.8072

LOWER TAIL

α/γ_1	D.C.	0.00	-0.24	-0.42	-0.64	-0.96	-1.14
.900	.056	.0460	.0557	.0724	.0660	.0681	.0660
.800	.113	.1022	.1144	.1332	.1387	.1468	.1411
.700	.169	.1602	.1765	.1820	.2106	.2161	.2162
.600	.227	.2135	.2397	.2340	.2657	.2897	.2851
.500	.288	.2661	.2954	.2910	.3339	.3767	.3684
.400	.350	.3265	.3516	.3610	.4186	.4453	.4443
.300	.420	.3929	.4192	.4332	.4873	.5155	.5095
.200	.502	.4803	.5150	.5319	.5655	.5900	.5911
.100	.609	.5946	.6201	.6398	.6722	.6952	.7030
.050	.689	.6859	.7093	.7371	.7336	.7719	.7638
.020	.763	.7686	.7809	.8101	.8011	.8256	.8649
.010	.805	.8506	.8452	.8311	.8738	.8515	.8858
.005	.839	.8775	.8538	.8674	.8837	.8762	.8960

TABLE III

BETA DISTRIBUTION

$N = 6$

$PR(r_{21} > R) = \alpha$

UPPER TAIL

α/γ_1	D.C.	0.00	-0.24	-0.42	-0.64	-0.96	-1.14
.900	.268	.2530	.2521	.1961	.1903	.1866	.1851
.800	.364	.3614	.3321	.2969	.2917	.2584	.2522
.700	.439	.4396	.3876	.3703	.3668	.3238	.3145
.600	.504	.4966	.4634	.4430	.4356	.3848	.3840
.500	.563	.5486	.5199	.5064	.5012	.4401	.4571
.400	.621	.6162	.5842	.5698	.5623	.5203	.5176
.300	.680	.6797	.6412	.6450	.6138	.5819	.5828
.200	.745	.7527	.7083	.7049	.7051	.6475	.6654
.100	.821	.8364	.7992	.7836	.7985	.7547	.7534
.050	.872	.8823	.8605	.8416	.8643	.8138	.8079
.020	.924	.9169	.9147	.9121	.9031	.8809	.8852
.010	.951	.9393	.9303	.9428	.9296	.9045	.9248
.005	.970	.9553	.9476	.9509	.9607	.9291	.9623

LOWER TAIL

α/γ_1	D.C.	0.00	-0.24	-0.42	-0.64	-0.96	-1.14
.900	.268	.2345	.2667	.2757	.2735	.2971	.3122
.800	.364	.3402	.3803	.3781	.3768	.4300	.4151
.700	.439	.4195	.4617	.4685	.4729	.5388	.5003
.600	.504	.4802	.5074	.5290	.5405	.6001	.5748
.500	.563	.5359	.5682	.5941	.6000	.6595	.6428
.400	.621	.5951	.6332	.6491	.6739	.7254	.6994
.300	.680	.6662	.6888	.7174	.7263	.7639	.7544
.200	.745	.7398	.7476	.7819	.7883	.8168	.8161
.100	.821	.8160	.8127	.8407	.8606	.8812	.8780
.050	.872	.8589	.8730	.8814	.9139	.9202	.9039
.020	.924	.9185	.9091	.9228	.9507	.9484	.9354
.010	.951	.9442	.9281	.9599	.9639	.9566	.9560
.005	.970	.9533	.9477	.9669	.9725	.9741	.9671

TABLE IV

BETA DISTRIBUTION

N = 6

$PR(r_{22} > R) = \alpha$

UPPER TAIL

α/γ_1	D.C.	0.CO	-0.24	-0.42	-0.64	-0.96	-1.14
.900	.410	.4132	.3692	.3419	.3234	.3042	.3144
.800	.540	.5269	.4974	.4803	.4703	.4300	.4177
.700	.640	.6116	.5838	.5869	.5654	.5550	.5145
.600	.720	.6912	.6613	.6578	.6406	.6316	.5943
.500	.780	.7548	.7273	.7213	.7002	.7059	.6815
.400	.830	.7981	.7816	.7806	.7682	.7799	.7467
.300	.880	.8577	.8445	.8458	.8269	.8335	.8321
.200	.930	.9128	.8998	.9011	.8939	.8955	.8867
.100	.965	.9521	.9614	.9573	.9539	.9410	.9472
.050	.983	.9793	.9788	.9776	.9763	.9646	.9757
.020	.992	.9904	.9932	.9920	.9905	.9849	.9921
.010	.995	.9972	.9959	.9949	.9966	.9912	.9966
.005	.998	.9985	.9968	.9961	.9982	.9960	.9984

LOWER TAIL

α/γ_1	D.C.	0.CO	-0.24	-0.42	-0.64	-0.96	-1.14
.900	.410	.3726	.4277	.4209	.3983	.4205	.4233
.800	.540	.4999	.5369	.5539	.5300	.5921	.5606
.700	.640	.5979	.6289	.6390	.6321	.7011	.6735
.600	.720	.6793	.7110	.7135	.7088	.7608	.7423
.500	.780	.7617	.7588	.7731	.7876	.8263	.7979
.400	.830	.8093	.8147	.8324	.8385	.8759	.8523
.300	.880	.8543	.8721	.8861	.8806	.9082	.8990
.200	.930	.9026	.9160	.9273	.9243	.9345	.9366
.100	.965	.9563	.9567	.9630	.9664	.9678	.9690
.050	.983	.9786	.9796	.9820	.9878	.9838	.9872
.020	.992	.9903	.9901	.9909	.9941	.9926	.9962
.010	.995	.9980	.9956	.9938	.9969	.9958	.9987
.005	.998	.9990	.9968	.9972	.9990	.9967	.9992

TABLE V

BETA DISTRIBUTION

$N = 10$

$PR(r_{10} > R) = \alpha$

UPPER TAIL

α/γ_1	D.C.	0.00	-0.24	-0.42	-0.64	-0.96	-1.14
.900	.025	.0267	.0210	.0173	.0157	.0118	.0123
.800	.051	.0509	.0384	.0379	.0325	.0223	.0249
.700	.080	.0784	.0596	.0552	.0523	.0346	.0348
.600	.110	.1054	.0873	.0838	.0696	.0510	.0462
.500	.142	.1354	.1158	.1107	.0894	.0680	.0579
.400	.178	.1690	.1506	.1324	.1115	.0868	.0774
.300	.219	.2132	.1832	.1568	.1384	.1092	.1015
.200	.273	.2611	.2252	.2079	.1734	.1372	.1269
.100	.349	.3340	.2841	.2550	.2326	.1881	.1762
.050	.412	.3791	.3471	.2969	.2801	.2283	.2115
.020	.483	.4545	.4109	.3813	.3385	.2882	.3081
.010	.527	.4923	.4359	.4154	.3773	.3249	.3437
.005	.568	.5219	.4626	.4232	.3968	.3566	.3624

LOWER TAIL

α/γ_1	D.C.	0.00	-0.24	-0.42	-0.64	-0.96	-1.14
.900	.025	.0252	.0255	.0326	.0421	.0382	.0464
.800	.051	.0534	.0556	.0710	.0650	.0755	.0905
.700	.080	.0766	.0827	.1037	.1004	.1140	.1280
.600	.110	.0941	.1142	.1413	.1391	.1527	.1829
.500	.142	.1209	.1470	.1823	.1823	.1957	.2286
.400	.178	.1535	.1763	.2211	.2288	.2426	.2733
.300	.219	.1865	.2263	.2636	.2704	.3032	.3404
.200	.273	.2446	.2718	.3163	.3410	.3742	.4006
.100	.349	.3257	.3431	.3939	.4069	.4686	.4986
.050	.412	.3894	.4193	.4352	.4666	.5311	.5751
.020	.483	.4486	.4922	.5065	.5528	.5955	.6392
.010	.527	.4718	.5186	.5296	.5791	.6286	.6877
.005	.568	.4970	.5596	.5513	.6208	.6832	.6931

TABLE VI

BETA DISTRIBUTION

N = 10

$PR(r_{11} > R) = \alpha$

UPPER TAIL

α/γ_1	D.C.	0.00	-0.24	-0.42	-0.64	-0.96	-1.14
.900	.030	.0322	.0273	.0222	.0234	.0168	.0151
.800	.063	.0597	.0457	.0479	.0432	.0296	.0333
.700	.098	.0940	.0753	.0747	.0672	.0472	.0496
.600	.134	.1261	.1091	.1063	.0896	.0693	.0659
.500	.173	.1621	.1400	.1343	.1158	.0884	.0842
.400	.216	.2008	.1818	.1638	.1412	.1157	.1059
.300	.265	.2528	.2269	.2050	.1774	.1430	.1353
.200	.325	.3085	.2758	.2592	.2217	.1808	.1736
.100	.409	.3880	.3573	.3328	.2739	.2389	.2256
.050	.477	.4544	.4241	.3903	.3466	.2822	.2873
.020	.551	.5257	.4845	.4279	.4217	.3671	.3753
.010	.597	.5641	.5097	.4491	.4585	.3852	.3999
.005	.639	.5729	.5625	.5128	.4911	.4312	.4262

LOWER TAIL

α/γ_1	D.C.	0.00	-0.24	-0.42	-0.64	-0.96	-1.14
.900	.030	.0301	.0304	.0384	.0489	.0429	.0531
.800	.063	.0649	.0664	.0820	.0762	.0848	.0979
.700	.098	.0932	.0999	.1166	.1157	.1237	.1421
.600	.134	.1149	.1364	.1672	.1615	.1696	.2022
.500	.173	.1461	.1703	.2064	.2056	.2147	.2503
.400	.216	.1831	.2124	.2520	.2488	.2687	.3021
.300	.265	.2307	.2662	.3031	.3093	.3333	.3672
.200	.325	.2939	.3251	.3687	.3721	.3947	.4297
.100	.409	.3625	.4069	.4349	.4577	.4910	.5333
.050	.477	.4463	.4873	.5091	.5113	.5635	.6081
.020	.551	.5322	.5709	.5786	.6211	.6361	.6765
.010	.597	.5748	.6072	.6099	.6770	.6800	.7071
.005	.639	.5951	.6521	.6374	.6882	.7080	.7297

TABLE VII

BETA DISTRIBUTION

$N = 10$

$PR(r_{21} > R) = \alpha$

UPPER TAIL

α/γ_1	D.C.	0.00	-0.24	-0.42	-0.64	-0.96	-1.14
.900	.130	.1226	.1066	.1048	.0960	.0748	.0676
.800	.189	.1804	.1653	.1515	.1297	.1079	.1082
.700	.240	.2343	.2080	.1937	.1683	.1367	.1348
.600	.286	.2746	.2481	.2283	.2055	.1636	.1618
.500	.329	.3156	.2839	.2651	.2431	.1910	.1909
.400	.374	.3555	.3186	.3049	.2746	.2156	.2212
.300	.420	.4058	.3603	.3468	.3179	.2532	.2529
.200	.474	.4608	.4158	.3938	.3556	.2986	.2962
.100	.551	.5420	.4953	.4719	.4269	.3809	.3605
.050	.612	.5987	.5534	.5296	.4894	.4334	.4193
.020	.678	.6454	.6264	.6133	.5712	.5145	.5279
.010	.726	.7241	.6717	.6282	.6346	.5731	.5455
.005	.760	.7353	.7402	.7318	.6637	.6049	.5853

LOWER TAIL

α/γ_1	D.C.	0.00	-0.24	-0.42	-0.64	-0.96	-1.14
.900	.130	.1243	.1340	.1545	.1540	.1651	.2016
.800	.189	.1776	.1946	.2194	.2270	.2502	.2846
.700	.240	.2225	.2447	.2700	.2947	.3111	.3384
.600	.286	.2643	.2940	.3213	.3425	.3595	.3917
.500	.329	.3065	.3373	.3706	.3847	.4036	.4381
.400	.374	.3479	.3795	.4169	.4349	.4613	.4898
.300	.420	.3904	.4314	.4689	.4840	.5114	.5375
.200	.474	.4458	.4879	.5157	.5403	.5761	.5955
.100	.551	.5281	.5766	.5969	.6104	.6559	.6747
.050	.612	.5742	.6287	.6523	.6644	.7005	.7221
.020	.678	.6381	.6877	.7063	.7140	.7462	.7662
.010	.726	.6722	.7427	.7641	.7532	.7709	.8032
.005	.760	.7129	.7652	.7784	.8059	.8101	.8526

TABLE VIII

BETA DISTRIBUTION

$N = 10$

$PR(r_{22} > R) = \alpha$

UPPER TAIL

α/γ_1	D.C.	0.00	-0.24	-0.42	-0.64	-0.96	-1.14
.900	.150	.1492	.1327	.1320	.1206	.0979	.0905
.800	.231	.2192	.1988	.1929	.1648	.1364	.1379
.700	.285	.2718	.2540	.2325	.2058	.1706	.1768
.600	.335	.3194	.2960	.2783	.2545	.2096	.2090
.500	.384	.3690	.3398	.3158	.2935	.2432	.2470
.400	.433	.4201	.3822	.3660	.3364	.2790	.2804
.300	.483	.4613	.4251	.4130	.3814	.3289	.3215
.200	.543	.5202	.4925	.4752	.4291	.3857	.3765
.100	.620	.6175	.5767	.5632	.5146	.4560	.4499
.050	.682	.6593	.6439	.6220	.5922	.5046	.5078
.020	.749	.7297	.7070	.6896	.6801	.5705	.6303
.010	.791	.7710	.7555	.7193	.7235	.6604	.6660
.005	.826	.8133	.7619	.7586	.7506	.6813	.6853

LOWER TAIL

α/γ_1	D.C.	0.00	-0.24	-0.42	-0.64	-0.96	-1.14
.900	.150	.1438	.1538	.1871	.1802	.1876	.2247
.800	.231	.2047	.2257	.2570	.2628	.2789	.3154
.700	.285	.2575	.2845	.3086	.3224	.3378	.3778
.600	.335	.3135	.3330	.3599	.3809	.3967	.4246
.500	.384	.3593	.3817	.4191	.4311	.4448	.4786
.400	.433	.4085	.4332	.4671	.4797	.4941	.5315
.300	.483	.4613	.4960	.5289	.5312	.5533	.5834
.200	.543	.5220	.5608	.5766	.5944	.6246	.6352
.100	.620	.6032	.6407	.6538	.6724	.6995	.7106
.050	.682	.6596	.6928	.7115	.7212	.7431	.7680
.020	.749	.7206	.7636	.7780	.7803	.7955	.8182
.010	.791	.7325	.7892	.8072	.8196	.8321	.8528
.005	.826	.7565	.8023	.8214	.8609	.8578	.8757

TABLE IX

BETA DISTRIBUTION

N = 15

$PR(r_{10} > R) = \alpha$

UPPER TAIL

α/γ_1	D.C.	0.CO	-0.24	-0.42	-0.64	-0.96	-1.14
.900	.019	.0172	.0167	.0106	.0125	.0111	.0069
.800	.040	.0365	.0353	.0202	.0269	.0206	.0136
.700	.062	.0565	.0496	.0364	.0379	.0281	.0204
.600	.085	.0772	.0642	.0525	.0514	.0394	.0282
.500	.111	.1002	.0817	.0694	.0660	.0495	.0410
.400	.141	.1268	.1059	.0891	.0836	.0601	.0543
.300	.175	.1557	.1326	.1142	.1017	.0719	.0717
.200	.220	.1983	.1749	.1451	.1301	.0923	.0909
.100	.285	.2527	.2277	.1912	.1636	.1267	.1180
.050	.338	.2976	.2652	.2371	.2018	.1599	.1442
.020	.399	.3428	.3140	.2830	.2497	.2103	.1790
.010	.438	.3702	.3319	.3160	.2789	.2605	.1981
.005	.475	.4068	.3591	.3398	.2919	.2819	.2040

LOWER TAIL

α/γ_1	D.C.	0.CO	-0.24	-0.42	-0.64	-0.96	-1.14
.900	.019	.0164	.0211	.0266	.0280	.0318	.0290
.800	.040	.0366	.0418	.0461	.0503	.0619	.0676
.700	.062	.0554	.0620	.0738	.0802	.0870	.1086
.600	.085	.0763	.0840	.0971	.1071	.1264	.1511
.500	.111	.0960	.1123	.1255	.1432	.1744	.2011
.400	.141	.1266	.1412	.1603	.1765	.2212	.2343
.300	.175	.1565	.1828	.2012	.2216	.2607	.2862
.200	.220	.2062	.2265	.2579	.2669	.3230	.3476
.100	.285	.2610	.3071	.3353	.3459	.3891	.4248
.050	.338	.3155	.3570	.3746	.3984	.4651	.4919
.020	.399	.3735	.4239	.4221	.4625	.5318	.5524
.010	.438	.3943	.4390	.4543	.5034	.5651	.6045
.005	.475	.4060	.4800	.5262	.5150	.6813	.6121

TABLE X

BETA DISTRIBUTION

N = 15

$PR(r_{11} > R) = \alpha$

UPPER TAIL

α/γ_1	D.C.	0.00	-0.24	-0.42	-0.64	-0.96	-1.14
.900	.023	.0205	.0194	.0127	.0149	.0144	.0083
.800	.047	.0422	.0404	.0264	.0313	.0256	.0184
.700	.072	.0646	.0593	.0433	.0465	.0364	.0275
.600	.099	.0870	.0769	.0640	.0630	.0495	.0375
.500	.129	.1168	.0986	.0816	.0800	.0611	.0530
.400	.164	.1439	.1267	.1043	.1006	.0748	.0696
.300	.203	.1773	.1583	.1359	.1225	.0960	.0886
.200	.253	.2278	.1995	.1741	.1535	.1182	.1097
.100	.323	.2809	.2606	.2258	.2025	.1603	.1472
.050	.381	.3313	.3062	.2687	.2566	.1938	.1897
.020	.445	.4104	.3579	.3252	.2753	.2595	.2179
.010	.486	.4276	.3792	.3595	.3147	.3204	.2709
.005	.522	.4608	.3844	.3921	.3303	.3688	.2818

LOWER TAIL

α/γ_1	D.C.	0.00	-0.24	-0.42	-0.64	-0.96	-1.14
.900	.023	.0188	.0231	.0291	.0314	.0329	.0326
.800	.047	.0408	.0458	.0525	.0560	.0665	.0764
.700	.072	.0627	.0704	.0821	.0863	.0944	.1163
.600	.099	.0888	.0931	.1075	.1170	.1384	.1609
.500	.129	.1132	.1259	.1408	.1569	.1867	.2126
.400	.164	.1485	.1622	.1771	.1940	.2314	.2476
.300	.203	.1773	.2090	.2237	.2340	.2797	.3048
.200	.253	.2368	.2565	.2845	.2865	.3439	.3696
.100	.323	.2933	.3301	.3627	.3707	.4201	.4453
.050	.381	.3451	.3868	.4089	.4316	.4803	.5164
.020	.445	.4093	.4518	.4624	.4890	.5664	.5992
.010	.486	.4398	.4840	.4951	.5226	.5877	.6197
.005	.522	.4650	.5251	.5339	.5471	.6842	.6389

TABLE XI

BETA DISTRIBUTION

N = 15

$PR(r_{21} > R) = \alpha$

UPPER TAIL

α/γ_1	D.C.	0.CO	-0.24	-0.42	-0.64	-0.96	-1.14
.900	.094	.0850	.0788	.0600	.0675	.0460	.0398
.800	.138	.1317	.1139	.0922	.0946	.0653	.0623
.700	.175	.1679	.1421	.1189	.1171	.0848	.0802
.600	.208	.1976	.1742	.1449	.1390	.1073	.1006
.500	.245	.2274	.2011	.1694	.1603	.1278	.1168
.400	.280	.2637	.2255	.2052	.1848	.1455	.1328
.300	.319	.2909	.2621	.2322	.2168	.1716	.1511
.200	.366	.3326	.2988	.2673	.2534	.2018	.1779
.100	.431	.4038	.3604	.3324	.3004	.2467	.2263
.050	.483	.4481	.4069	.3666	.3391	.2911	.2722
.020	.537	.4925	.4773	.4163	.3869	.3444	.3145
.010	.574	.5187	.5111	.4462	.4036	.3908	.3367
.005	.607	.5741	.5415	.4856	.4320	.4633	.3957

LOWER TAIL

α/γ_1	D.C.	0.CO	-0.24	-0.42	-0.64	-0.96	-1.14
.900	.094	.0861	.0867	.1046	.1161	.1230	.1270
.800	.138	.1336	.1455	.1453	.1635	.1782	.2043
.700	.175	.1649	.1848	.1874	.2050	.2328	.2500
.600	.208	.1940	.2202	.2281	.2444	.2830	.3080
.500	.245	.2232	.2545	.2717	.2758	.3304	.3568
.400	.280	.2553	.2861	.3090	.3154	.3731	.4031
.300	.319	.2974	.3353	.3576	.3593	.4255	.4435
.200	.366	.3417	.3926	.4084	.4171	.4739	.5041
.100	.431	.4063	.4504	.4715	.4780	.5378	.5812
.050	.483	.4523	.4822	.5188	.5337	.5837	.6307
.020	.537	.4950	.5491	.5768	.5846	.6532	.6842
.010	.574	.5309	.5883	.6181	.6237	.7050	.7200
.005	.607	.5556	.6138	.6430	.6332	.7393	.7437

TABLE XII

BETA DISTRIBUTION

N = 15

$PR(r_{22} > R) = \alpha$

UPPER TAIL

α/γ_1	D.C.	0.00	-0.24	-0.42	-0.64	-0.96	-1.14
.900	.109	.0935	.0967	.0746	.0764	.0525	.0480
.800	.156	.1440	.1322	.1078	.1068	.0780	.0734
.700	.196	.1903	.1623	.1356	.1373	.1007	.1001
.600	.234	.2172	.1988	.1660	.1617	.1298	.1194
.500	.273	.2522	.2291	.2004	.1820	.1504	.1371
.400	.312	.2864	.2633	.2319	.2113	.1744	.1586
.300	.353	.3259	.2916	.2565	.2484	.2012	.1801
.200	.402	.3661	.3305	.3084	.2888	.2325	.2120
.100	.472	.4455	.3913	.3600	.3457	.2859	.2669
.050	.525	.4975	.4657	.4032	.3860	.3402	.3047
.020	.579	.5352	.4990	.4545	.4375	.3985	.3508
.010	.616	.5765	.5373	.4935	.4584	.4508	.3846
.005	.647	.6157	.5717	.5311	.4852	.5004	.4271

LOWER TAIL

α/γ_1	D.C.	0.00	-0.24	-0.42	-0.64	-0.96	-1.14
.900	.109	.0955	.0984	.1144	.1285	.1307	.1365
.800	.156	.1480	.1584	.1627	.1793	.1915	.2163
.700	.196	.1863	.2039	.2032	.2208	.2448	.2600
.600	.234	.2192	.2455	.2512	.2644	.3033	.3192
.500	.273	.2472	.2773	.2965	.2965	.3497	.3757
.400	.312	.2874	.3153	.3343	.3435	.3968	.4224
.300	.353	.3243	.3577	.3773	.3864	.4484	.4671
.200	.402	.3755	.4292	.4353	.4437	.4944	.5239
.100	.472	.4412	.4838	.5058	.5056	.5599	.5964
.050	.525	.4944	.5248	.5442	.5552	.6056	.6498
.020	.579	.5470	.5657	.6177	.6204	.6858	.7056
.010	.616	.5795	.6172	.6424	.6446	.7139	.7390
.005	.647	.5923	.6477	.6462	.6607	.7508	.7640

TABLE XIII

GAMMA DISTRIBUTION

N = 6

$PR(r_{10} > R) = \alpha$

UPPER TAIL

α/γ_1	D.C.	0.82	0.89	1.00	1.15	1.41	2.00
.900	.038	.0463	.0535	.0483	.0499	.0717	.0710
.800	.079	.1023	.1137	.1025	.1047	.1339	.1455
.700	.121	.1556	.1735	.1493	.1539	.1926	.2214
.600	.164	.2107	.2348	.1990	.2099	.2538	.2928
.500	.210	.2632	.2922	.2576	.2709	.3179	.3687
.400	.261	.3262	.3437	.3185	.3360	.3915	.4389
.300	.318	.4074	.4275	.3871	.4221	.4645	.5308
.200	.386	.4701	.4977	.4930	.4994	.5414	.6132
.100	.482	.5695	.5965	.5875	.5885	.6302	.7130
.050	.560	.6702	.6773	.6616	.6507	.7092	.8088
.020	.644	.7417	.7380	.7333	.7249	.7895	.8497
.010	.698	.7618	.7761	.8036	.7929	.8144	.8756
.005	.740	.7854	.8056	.8274	.8434	.8408	.9024

LOWER TAIL

α/γ_1	D.C.	0.82	0.89	1.00	1.15	1.41	2.00
.900	.038	.0232	.0255	.0305	.0237	.0172	.0117
.800	.079	.0508	.0529	.0503	.0507	.0354	.0222
.700	.121	.0732	.0773	.0792	.0714	.0546	.0352
.600	.164	.1096	.1160	.1037	.0994	.0823	.0518
.500	.210	.1452	.1528	.1367	.1297	.1048	.0646
.400	.261	.1830	.1900	.1699	.1679	.1310	.0884
.300	.318	.2255	.2347	.2119	.2044	.1683	.1262
.200	.386	.2874	.3101	.2813	.2514	.2117	.1826
.100	.482	.3639	.4029	.3748	.3469	.2842	.2574
.050	.560	.4517	.4853	.4640	.4037	.3444	.3270
.020	.644	.5573	.5556	.5464	.4859	.4338	.4399
.010	.698	.6253	.5813	.5946	.5366	.4762	.5322
.005	.740	.6444	.6919	.6387	.5852	.5325	.5828

TABLE XIV

GAMMA DISTRIBUTION

N = 6

$PR(r_{11} > R) = \alpha$

UPPER TAIL

α/γ_1	D.C.	0.82	0.89	1.00	1.15	1.41	2.00
.900	.056	.0616	.0681	.0662	.0604	.0881	.0841
.800	.113	.1286	.1515	.1309	.1230	.1632	.1746
.700	.169	.1972	.2296	.1888	.1836	.2294	.2624
.600	.227	.2647	.3034	.2552	.2638	.3019	.3369
.500	.288	.3319	.3685	.3162	.3240	.3668	.4211
.400	.350	.4058	.4345	.3839	.4211	.4559	.5027
.300	.420	.4766	.5090	.4853	.5003	.5254	.5798
.200	.502	.5607	.5941	.5706	.5787	.6051	.6580
.100	.609	.6829	.7026	.6788	.6534	.7193	.7684
.050	.689	.7579	.7837	.7589	.7375	.7961	.8425
.020	.763	.7993	.8418	.8325	.8170	.8323	.8971
.010	.805	.8220	.8768	.8870	.8793	.8723	.9283
.005	.839	.8513	.9231	.8995	.9071	.8982	.9429

LOWER TAIL

α/γ_1	D.C.	0.82	0.89	1.00	1.15	1.41	2.00
.900	.056	.0383	.0448	.0452	.0383	.0304	.0219
.800	.113	.0747	.0857	.0758	.0726	.0615	.0426
.700	.169	.1159	.1276	.1136	.1162	.0959	.0661
.600	.227	.1665	.1775	.1584	.1552	.1260	.0875
.500	.288	.2147	.2187	.2069	.1947	.1646	.1288
.400	.350	.2773	.2885	.2608	.2510	.2135	.1726
.300	.420	.3281	.3717	.3164	.3064	.2661	.2259
.200	.502	.3995	.4374	.4100	.3809	.3244	.2894
.100	.609	.4970	.5627	.5121	.4744	.4176	.4128
.050	.689	.5893	.6559	.6126	.5625	.5347	.5008
.020	.763	.6855	.7403	.7372	.6683	.6351	.5640
.010	.805	.7770	.8347	.7731	.7477	.7134	.7671
.005	.839	.8150	.8578	.8318	.7666	.7767	.7840

TABLE XV

GAMMA DISTRIBUTION

N = 6

$PR(r_{21} > R) = \alpha$

UPPER TAIL

α/γ_1	D.C.	0.82	0.89	1.00	1.15	1.41	2.00
.900	.268	.3243	.3152	.3060	.3070	.3420	.3612
.800	.364	.4033	.4553	.4153	.3940	.4599	.3085
.700	.439	.4917	.5220	.4885	.4939	.5460	.5929
.600	.504	.5552	.5791	.5604	.5788	.6066	.6575
.500	.563	.6211	.6367	.6314	.6475	.6671	.7118
.400	.621	.6830	.6988	.6838	.6949	.7103	.7665
.300	.680	.7453	.7500	.7321	.7506	.7505	.8136
.200	.745	.8105	.8009	.7991	.8102	.8164	.8676
.100	.821	.8616	.8658	.8671	.8739	.8852	.9252
.050	.872	.8928	.8960	.9181	.9170	.9262	.9512
.020	.924	.9231	.9338	.9587	.9559	.9543	.9731
.010	.951	.9462	.9664	.9662	.9645	.9671	.9761
.005	.970	.9662	.9736	.9742	.9762	.9749	.9792

LOWER TAIL

α/γ_1	D.C.	0.82	0.89	1.00	1.15	1.41	2.00
.900	.268	.2056	.2025	.1843	.1913	.1522	.1044
.800	.364	.2805	.2902	.2700	.2697	.2271	.1846
.700	.439	.3545	.3601	.3220	.3419	.2882	.2486
.600	.504	.4161	.4369	.3836	.4062	.3366	.2900
.500	.563	.4761	.4915	.4450	.4568	.3928	.3477
.400	.621	.5448	.5526	.5163	.5193	.4682	.4156
.300	.680	.5998	.6130	.5772	.5865	.5454	.5002
.200	.745	.6813	.6846	.6611	.6599	.6267	.5766
.100	.821	.7765	.7736	.7809	.7838	.7386	.6932
.050	.872	.8552	.8418	.8382	.8403	.8275	.7644
.020	.924	.9049	.8903	.8916	.9030	.8942	.8429
.010	.951	.9487	.9164	.9160	.9385	.9153	.8824
.005	.970	.9745	.9434	.9340	.9568	.9253	.9108

TABLE XVI

GAMMA DISTRIBUTION

N = 6

$PR(r_{22} > R) = \alpha$

UPPER TAIL

α/γ_1	D.C.	0.82	0.89	1.00	1.15	1.41	2.00
.900	.410	.4385	.4711	.4499	.4411	.4684	.4646
.800	.540	.5787	.5968	.5518	.5719	.6240	.6420
.700	.640	.6702	.6703	.6579	.6811	.6939	.7326
.600	.720	.7426	.7385	.7329	.7628	.7594	.7916
.500	.780	.7944	.7927	.7941	.8161	.8061	.8432
.400	.830	.8390	.8461	.8441	.8592	.8495	.8825
.300	.880	.8884	.8829	.8774	.9023	.8867	.9180
.200	.930	.9310	.9182	.9192	.9381	.9297	.9488
.100	.965	.9679	.9584	.9610	.9676	.9614	.9758
.050	.983	.9853	.9818	.9819	.9866	.9763	.9904
.020	.992	.9933	.9941	.9933	.9955	.9884	.9970
.010	.995	.9950	.9974	.9965	.9976	.9925	.9989
.005	.998	.9980	.9986	.9982	.9984	.9963	.9993

LOWER TAIL

α/γ_1	D.C.	0.82	0.89	1.00	1.15	1.41	2.00
.900	.410	.3562	.3423	.3097	.3433	.2939	.2369
.800	.540	.4840	.4730	.4327	.4763	.4011	.3656
.700	.640	.5595	.5607	.5443	.5657	.4918	.4612
.600	.720	.6335	.6352	.6148	.6340	.5688	.5418
.500	.780	.6933	.7068	.6868	.7082	.6477	.6278
.400	.830	.7679	.7657	.7516	.7758	.7198	.7067
.300	.880	.8276	.8247	.8182	.8382	.7764	.7703
.200	.930	.8880	.8779	.8876	.8949	.8458	.8490
.100	.965	.9489	.9293	.9371	.9495	.9188	.9188
.050	.983	.9769	.9693	.9619	.9776	.9516	.9628
.020	.992	.9898	.9902	.9847	.9899	.9780	.9909
.010	.995	.9941	.9944	.9921	.9947	.9902	.9962
.005	.998	.9978	.9979	.9990	.9966	.9922	.9980

TABLE XVII

GAMMA DISTRIBUTION

$N = 10$

$PR(r_{10} > R) = \alpha$

UPPER TAIL

α/γ_1	D.C.	0.82	0.89	1.00	1.15	1.41	2.00
.900	.025	.0400	.0447	.0382	.0359	.0400	.0552
.800	.051	.0739	.0817	.0747	.0821	.0842	.1190
.700	.080	.1069	.1221	.1210	.1292	.1383	.1723
.600	.110	.1504	.1631	.1635	.1807	.1827	.2259
.500	.142	.1924	.2021	.2054	.2273	.2388	.3019
.400	.178	.2377	.2461	.2672	.2770	.3036	.3607
.300	.219	.2901	.3009	.3226	.3393	.3612	.4280
.200	.273	.3608	.3814	.4021	.4184	.4486	.4977
.100	.349	.4319	.4715	.5030	.5276	.5359	.6162
.050	.412	.5027	.5266	.5511	.5898	.6073	.6979
.020	.483	.5787	.6053	.6274	.6585	.6652	.7458
.010	.527	.6351	.6357	.6589	.6819	.7121	.7871
.005	.568	.6445	.6862	.6882	.7243	.7281	.8013

LOWER TAIL

α/γ_1	D.C.	0.82	0.89	1.00	1.15	1.41	2.00
.900	.025	.0151	.0179	.0163	.0150	.0096	.0033
.800	.051	.0362	.0347	.0299	.0289	.0204	.0077
.700	.080	.0532	.0517	.0447	.0393	.0303	.0130
.600	.110	.0741	.0726	.0621	.0560	.0426	.0202
.500	.142	.0983	.0918	.0840	.0721	.0550	.0277
.400	.178	.1229	.1168	.1027	.0932	.0678	.0376
.300	.219	.1449	.1423	.1291	.1177	.0885	.0491
.200	.273	.1852	.1719	.1658	.1532	.1147	.0643
.100	.349	.2420	.2291	.2301	.2036	.1619	.0922
.050	.412	.3073	.2860	.2743	.2441	.2025	.1270
.020	.483	.3354	.3631	.3371	.2802	.2671	.1575
.010	.527	.3639	.3993	.3774	.3394	.3120	.1735
.005	.568	.3690	.4483	.3974	.3512	.3445	.1835

TABLE XVIII

GAMMA DISTRIBUTION

N = 10

$PR(r_{11} > R) = \alpha$

UPPER TAIL

α/γ_1	D.C.	0.82	0.89	1.00	1.15	1.41	2.00
.900	.030	.0435	.0507	.0438	.0405	.0422	.0576
.800	.063	.0829	.0924	.0881	.0911	.0914	.1265
.700	.098	.1220	.1384	.1376	.1478	.1503	.1802
.600	.134	.1678	.1839	.1886	.2006	.2009	.2346
.500	.173	.2220	.2335	.2315	.2521	.2624	.3059
.400	.216	.2675	.2791	.2944	.3105	.3234	.3745
.300	.265	.3275	.3404	.3551	.3797	.3870	.4441
.200	.325	.3927	.4221	.4365	.4508	.4742	.5172
.100	.409	.4991	.5175	.5454	.5653	.5767	.6305
.050	.477	.5494	.5794	.6118	.6340	.6404	.7276
.020	.551	.6458	.6443	.6619	.6952	.7094	.7714
.010	.597	.6959	.6893	.7169	.7348	.7278	.7899
.005	.639	.7200	.7412	.7473	.7725	.7815	.8183

LOWER TAIL

α/γ_1	D.C.	0.82	0.89	1.00	1.15	1.41	2.00
.900	.030	.0211	.0236	.0244	.0219	.0138	.0052
.800	.063	.0433	.0478	.0418	.0405	.0303	.0130
.700	.098	.0721	.0709	.0618	.0579	.0449	.0203
.600	.134	.0949	.0961	.0852	.0795	.0634	.0293
.500	.173	.1277	.1230	.1134	.1019	.0774	.0406
.400	.216	.1537	.1568	.1385	.1314	.0974	.0588
.300	.265	.1923	.1844	.1702	.1619	.1231	.0760
.200	.325	.2469	.2304	.2108	.2081	.1624	.0970
.100	.409	.3180	.3013	.2985	.2598	.2182	.1421
.050	.477	.3803	.3769	.3358	.3127	.2619	.1858
.020	.551	.4277	.4634	.4714	.3633	.3816	.2323
.010	.597	.4716	.5138	.5006	.3858	.4028	.2845
.005	.639	.4994	.5336	.5449	.4244	.4391	.3072

TABLE XIX

GAMMA DISTRIBUTION

$N = 10$

$PR(r_{21} > R) = \alpha$

UPPER TAIL

α/γ_1	D.C.	0.82	0.89	1.00	1.15	1.41	2.00
.900	.130	.1646	.1656	.1894	.1819	.2151	.2265
.800	.189	.2349	.2521	.2599	.2771	.3004	.3280
.700	.240	.2934	.3166	.3206	.3431	.3582	.3983
.600	.286	.3456	.3716	.3699	.3992	.4153	.4738
.500	.329	.3998	.4207	.4209	.4507	.4629	.5256
.400	.374	.4614	.4674	.4669	.5023	.5220	.5831
.300	.420	.5116	.5135	.5255	.5575	.5766	.6354
.200	.474	.5619	.5728	.5955	.6197	.6296	.6948
.100	.551	.6423	.6456	.6606	.6988	.6931	.7689
.050	.612	.7103	.6878	.7209	.7514	.7591	.8166
.020	.678	.7755	.7550	.7660	.7975	.8004	.8510
.010	.726	.7890	.7686	.8088	.8196	.8281	.8827
.005	.760	.8346	.7959	.8174	.8414	.8501	.8926

LOWER TAIL

α/γ_1	D.C.	0.82	0.89	1.00	1.15	1.41	2.00
.900	.130	.0923	.0986	.0933	.0886	.0763	.0351
.800	.189	.1290	.1306	.1301	.1222	.1028	.0567
.700	.240	.1743	.1726	.1617	.1580	.1286	.0749
.600	.286	.2048	.2055	.1931	.1818	.1484	.0935
.500	.329	.2401	.2408	.2230	.2185	.1749	.1104
.400	.374	.2934	.2698	.2586	.2512	.2067	.1335
.300	.420	.3387	.3071	.2960	.2847	.2384	.1632
.200	.474	.3894	.3656	.3385	.3293	.2818	.2006
.100	.551	.4510	.4429	.4116	.3876	.3538	.2569
.050	.612	.5099	.5048	.4791	.4488	.4288	.3124
.020	.678	.6164	.5888	.5640	.5054	.5024	.3930
.010	.726	.6751	.6391	.6322	.5609	.5521	.4209
.005	.760	.7087	.6638	.6746	.6026	.5785	.4629

TABLE XX

GAMMA DISTRIBUTION

N = 10

$PR(r_{22} > R) = \alpha$

UPPER TAIL

α/γ_1	D.C.	0.82	0.89	1.00	1.15	1.41	2.00
.900	.150	.1946	.1882	.2198	.2127	.2404	.2424
.800	.231	.2716	.2816	.2954	.3122	.3273	.3561
.700	.285	.3300	.3548	.3544	.3766	.3911	.4327
.600	.335	.3856	.4138	.4109	.4367	.4511	.5006
.500	.384	.4480	.4672	.4622	.4943	.5039	.5520
.400	.433	.5000	.5147	.5179	.5503	.5590	.6084
.300	.483	.5610	.5648	.5705	.6078	.6183	.6648
.200	.543	.6192	.6214	.6285	.6685	.6670	.7226
.100	.620	.6881	.6870	.7081	.7353	.7484	.8029
.050	.682	.7500	.7348	.7662	.7907	.7917	.8502
.020	.749	.8078	.8062	.8169	.8329	.8400	.8788
.010	.791	.8349	.8319	.8398	.8579	.8691	.8999
.005	.826	.8810	.8597	.8527	.8708	.8870	.9082

LOWER TAIL

α/γ_1	D.C.	0.82	0.89	1.00	1.15	1.41	2.00
.900	.150	.1140	.1258	.1198	.1159	.1050	.0523
.800	.231	.1697	.1653	.1598	.1573	.1385	.0807
.700	.285	.2164	.2136	.2050	.2017	.1700	.1043
.600	.335	.2551	.2530	.2413	.2381	.1974	.1317
.500	.384	.3082	.2940	.2795	.2698	.2327	.1576
.400	.433	.3510	.3303	.3239	.3092	.2646	.1896
.300	.483	.4116	.3752	.3645	.3592	.3086	.2214
.200	.543	.4624	.4388	.4222	.4082	.3638	.2674
.100	.620	.5397	.5175	.4883	.4868	.4433	.3542
.050	.682	.5994	.6095	.5649	.5460	.5229	.4250
.020	.749	.6866	.6629	.6499	.6136	.5971	.5172
.010	.791	.7281	.7042	.7076	.6519	.6555	.5842
.005	.826	.8212	.7063	.7444	.6746	.6745	.6113

TABLE XXI

GAMMA DISTRIBUTION

N = 15

$PR(r_{10} > R) = \alpha$

UPPER TAIL

α/γ_1	D.C.	0.82	0.89	1.00	1.15	1.41	2.00
.900	.019	.0340	.0263	.0339	.0330	.0351	.0431
.800	.040	.0653	.0568	.0645	.0691	.0741	.0990
.700	.062	.0981	.0865	.0962	.1056	.1184	.1401
.600	.085	.1365	.1201	.1361	.1418	.1600	.1935
.500	.111	.1749	.1557	.1760	.1857	.2033	.2409
.400	.141	.2189	.1901	.2211	.2293	.2570	.2970
.300	.175	.2586	.2425	.2756	.2863	.3162	.3707
.200	.220	.3140	.3159	.3358	.3611	.3861	.4452
.100	.285	.3890	.3947	.4249	.4423	.4764	.5327
.050	.338	.4499	.4646	.4854	.5055	.5561	.5910
.020	.399	.5068	.5371	.5095	.5535	.6232	.6915
.010	.438	.5414	.5649	.5781	.5941	.6678	.7512
.005	.475	.5769	.5859	.6267	.6126	.6997	.7628

LOWER TAIL

α/γ_1	D.C.	0.82	0.89	1.00	1.15	1.41	2.00
.900	.019	.0123	.0103	.0116	.0090	.0069	.0027
.800	.040	.0269	.0202	.0237	.0190	.0145	.0046
.700	.062	.0416	.0315	.0370	.0284	.0214	.0076
.600	.085	.0578	.0433	.0511	.0394	.0287	.0117
.500	.111	.0705	.0572	.0639	.0513	.0366	.0161
.400	.141	.0893	.0751	.0796	.0638	.0464	.0210
.300	.175	.1099	.0915	.0982	.0798	.0623	.0276
.200	.220	.1352	.1205	.1226	.1017	.0834	.0350
.100	.285	.1819	.1564	.1598	.1349	.1108	.0555
.050	.338	.2208	.1964	.1900	.1681	.1376	.0698
.020	.399	.2813	.2480	.2407	.1930	.1723	.1007
.010	.438	.3124	.2704	.2757	.2272	.1985	.1307
.005	.475	.3543	.2839	.2854	.2377	.2332	.1463

TABLE XXII

GAMMA DISTRIBUTION

N = 15

$PR(r_{11} > R) = \alpha$

UPPER TAIL

α/γ_1	D.C.	0.82	0.89	1.00	1.15	1.41	2.00
.900	.023	.0382	.0297	.0370	.0364	.0372	.0449
.800	.047	.0731	.0608	.0693	.0759	.0795	.1003
.700	.072	.1114	.0964	.1011	.1129	.1248	.1471
.600	.099	.1528	.1313	.1494	.1552	.1684	.1982
.500	.129	.1920	.1674	.1907	.1984	.2185	.2480
.400	.164	.2350	.2079	.2386	.2413	.2686	.3031
.300	.203	.2822	.2582	.2923	.3066	.3254	.3801
.200	.253	.3454	.3356	.3616	.3811	.4030	.4518
.100	.323	.4178	.4242	.4483	.4593	.4954	.5432
.050	.381	.4776	.4841	.5126	.5383	.5820	.5945
.020	.445	.5309	.5595	.5587	.5778	.6515	.7012
.010	.486	.5796	.6098	.5988	.6237	.6976	.7572
.005	.522	.6138	.6292	.6508	.6413	.7217	.7703

LOWER TAIL

α/γ_1	D.C.	0.82	0.89	1.00	1.15	1.41	2.00
.900	.023	.0171	.0123	.0170	.0127	.0100	.0040
.800	.047	.0332	.0263	.0317	.0244	.0201	.0072
.700	.072	.0525	.0406	.0465	.0365	.0295	.0117
.600	.099	.0742	.0536	.0632	.0520	.0373	.0166
.500	.129	.0923	.0715	.0788	.0685	.0508	.0227
.400	.164	.1113	.0944	.1032	.0814	.0653	.0300
.300	.203	.1360	.1156	.1261	.1019	.0801	.0376
.200	.253	.1605	.1404	.1590	.1335	.1082	.0493
.100	.323	.2210	.1935	.2005	.1745	.1436	.0719
.050	.381	.2778	.2433	.2453	.2067	.1805	.1024
.020	.445	.3375	.2923	.2879	.2541	.2188	.1290
.010	.486	.3841	.3295	.3081	.2885	.2411	.1564
.005	.522	.4126	.3547	.3214	.3081	.2938	.1602

TABLE XXIII

GAMMA DISTRIBUTION

N = 15

$PR(r_{21} > R) = \alpha$

UPPER TAIL

α/γ_1	D.C.	0.82	0.89	1.00	1.15	1.41	2.00
.900	.094	.1296	.1246	.1329	.1342	.1670	.1959
.800	.138	.1926	.1787	.1862	.1950	.2347	.2663
.700	.175	.2485	.2379	.2401	.2446	.2854	.3232
.600	.208	.2833	.2831	.2902	.2852	.3337	.3816
.500	.245	.3251	.3251	.3409	.3353	.3896	.4286
.400	.280	.3726	.3746	.3919	.3928	.4392	.4771
.300	.319	.4247	.4228	.4410	.4493	.4884	.5296
.200	.366	.4754	.4710	.4959	.5038	.5443	.5940
.100	.431	.5352	.5410	.5609	.5850	.6212	.6622
.050	.483	.5886	.5843	.6202	.6312	.6769	.7143
.020	.537	.6353	.6808	.6645	.6677	.7445	.7718
.010	.574	.6750	.7164	.6973	.7007	.7836	.7999
.005	.607	.7013	.7227	.7663	.7264	.7977	.8134

LOWER TAIL

α/γ_1	D.C.	0.82	0.89	1.00	1.15	1.41	2.00
.900	.094	.0658	.0593	.0555	.0521	.0432	.0172
.800	.138	.0979	.0828	.0829	.0785	.0600	.0262
.700	.175	.1293	.1077	.1058	.0967	.0770	.0368
.600	.208	.1529	.1303	.1283	.1123	.0944	.0451
.500	.245	.1712	.1494	.1516	.1339	.1109	.0544
.400	.280	.1956	.1769	.1793	.1555	.1283	.0657
.300	.319	.2232	.2051	.2088	.1798	.1547	.0807
.200	.366	.2552	.2444	.2453	.2142	.1791	.0995
.100	.431	.3203	.2918	.2936	.2512	.2271	.1360
.050	.483	.3573	.3419	.3273	.2859	.2554	.1594
.020	.537	.4169	.3777	.3612	.3589	.2927	.2033
.010	.574	.4545	.4146	.4038	.3752	.3084	.2335
.005	.607	.5040	.4885	.4327	.4313	.3541	.2409

TABLE XXIV

GAMMA DISTRIBUTION

N = 15

$PR(r_{22} > R) = \alpha$

UPPER TAIL

α/γ_1	D.C.	0.82	0.89	1.00	1.15	1.41	2.00
.900	.109	.1417	.1392	.1421	.1416	.1683	.2030
.800	.156	.2146	.1966	.2001	.2081	.2515	.2746
.700	.196	.2607	.2585	.2553	.2603	.2994	.3365
.600	.234	.3077	.3045	.3086	.3051	.3530	.3902
.500	.273	.3489	.3553	.3630	.3563	.4116	.4359
.400	.312	.3989	.3939	.4093	.4136	.4573	.4979
.300	.353	.4520	.4514	.4647	.4764	.5096	.5404
.200	.402	.5051	.4986	.5249	.5339	.5694	.6056
.100	.472	.5717	.5665	.5886	.6150	.6459	.6776
.050	.525	.6190	.6294	.6390	.6528	.6980	.7298
.020	.579	.6732	.6866	.6936	.6967	.7551	.7765
.010	.616	.6985	.7294	.7287	.7269	.7926	.8036
.005	.647	.7514	.7752	.7781	.7348	.8019	.8178

LOWER TAIL

α/γ_1	D.C.	0.82	0.89	1.00	1.15	1.41	2.00
.900	.109	.0786	.0735	.0649	.0636	.0537	.0223
.800	.156	.1179	.1022	.1018	.0919	.0755	.0356
.700	.196	.1521	.1283	.1278	.1137	.0992	.0474
.600	.234	.1765	.1524	.1541	.1357	.1178	.0589
.500	.273	.1993	.1822	.1764	.1592	.1384	.0720
.400	.312	.2261	.2160	.2087	.1827	.1616	.0846
.300	.353	.2577	.2448	.2402	.2171	.1870	.1023
.200	.402	.3027	.2916	.2820	.2493	.2184	.1256
.100	.472	.3617	.3478	.3447	.2877	.2635	.1628
.050	.525	.4246	.3841	.3843	.3530	.3014	.1908
.020	.579	.4701	.4433	.4280	.4177	.3656	.2501
.010	.616	.5227	.4800	.4676	.4403	.3855	.2833
.005	.647	.5442	.5258	.4969	.4579	.4055	.3124

APPENDIX B

TABLES LISTING SIGNIFICANCE LEVEL AT WHICH BETA AND GAMMA STATISTICS DIFFER FROM NORMAL STATISTICS

TABLE XXV

KOLMOGOROV-SMIRNOV GOODNESS-OF-FIT TEST RESULTS

γ_1	0	-.24	-.42	-.64	-.96	-1.14

N = 6

UPPER TAIL

r_{10}	N.S.	.01	.01	.01	.01	.01
r_{11}	N.S.	.01	.01	.01	.01	.01
r_{21}	N.S.	.01	.01	.01	.01	.01
r_{22}	.05	.01	.01	.01	.01	.01

LOWER TAIL

r_{10}	N.S.	N.S.	.10	.01	.01	.01
r_{11}	N.S.	N.S.	N.S.	.01	.01	.01
r_{21}	N.S.	N.S.	.05	.01	.01	.01
r_{22}	.10	N.S.	N.S.	N.S.	.01	N.S.

N = 10

UPPER TAIL

r_{10}	N.S.	.01	.01	.01	.01	.01
r_{11}	N.S.	.01	.01	.01	.01	.01
r_{21}	N.S.	.01	.01	.01	.01	.01
r_{22}	N.S.	.01	.01	.01	.01	.01

LOWER TAIL

r_{10}	.01	N.S.	.01	.01	.01	.01
r_{11}	.01	N.S.	.01	.01	.01	.01
r_{21}	.05	N.S.	.01	.01	.01	.01
r_{22}	.10	N.S.	.01	.01	.01	.01

N = 15

UPPER TAIL

r_{10}	.05	.01	.01	.01	.01	.01
r_{11}	.05	.01	.01	.01	.01	.01
r_{21}	.01	.01	.01	.01	.01	.01
r_{22}	.01	.01	.01	.01	.01	.01

LOWER TAIL

r_{10}	.05	N.S.	.05	.01	.01	.01
r_{11}	.05	N.S.	N.S.	.01	.01	.01
r_{21}	.01	N.S.	.01	.01	.01	.01
r_{22}	.01	N.S.	.05	.01	.01	.01

* N.S. MEANS NOT SIGNIFICANTLY DIFFERENT FROM DIXON'S VALUES AT .10 RISK LEVEL OR LOWER

TABLE XXVI

KOLMOGOROV-SMIRNOV GOODNESS-OF-FIT TEST RESULTS

γ_1	0.82	0.89	1.00	1.15	1.41	2.00
N = 6						
UPPER TAIL						
r_{10}	.01	.01	.01	.01	.01	.01
r_{11}	.01	.01	.01	.01	.01	.01
r_{21}	.01	.01	.01	.01	.01	.01
r_{22}	.01	.01	.01	N.S.	.05	N.S.
LOWER TAIL						
r_{10}	.01	.01	.01	.01	.01	.01
r_{11}	.01	.01	.01	.01	.01	.01
r_{21}	.01	.01	.01	.01	.01	.01
r_{22}	.01	.01	.01	.01	.01	.01
N = 10						
UPPER TAIL						
r_{10}	.01	.01	.01	.01	.01	.01
r_{11}	.01	.01	.01	.01	.01	.01
r_{21}	.01	.01	.01	.01	.01	.01
r_{22}	.01	.01	.01	.01	.01	.01
LOWER TAIL						
r_{10}	.01	.01	.01	.01	.01	.01
r_{11}	.01	.01	.01	.01	.01	.01
r_{21}	.01	.01	.01	.01	.01	.01
r_{22}	.01	.01	.01	.01	.01	.01
N = 15						
UPPER TAIL						
r_{10}	.01	.01	.01	.01	.01	.01
r_{11}	.01	.01	.01	.01	.01	.01
r_{21}	.01	.01	.01	.01	.01	.01
r_{22}	.01	.01	.01	.01	.01	.01
LOWER TAIL						
r_{10}	.01	.01	.01	.01	.01	.01
r_{11}	.01	.01	.01	.01	.01	.01
r_{21}	.01	.01	.01	.01	.01	.01
r_{22}	.01	.01	.01	.01	.01	.01

* N.S. MEANS NOT SIGNIFICANTLY DIFFERENT FROM DIXON'S VALUES AT .10 RISK LEVEL OR LOWER

APPENDIX C

MACHINE PROGRAMMING OF DISTRIBUTIONS

```
                    JERRY THOMAS - BETA DISTRIBUTION
        BLOC(RR-RR600)
        BLOC(QQ-QQ599)HH-HH599)BB-BB599)$
        BLOC(CC-CC599)DD-DD599)EE-EE599)
        BLOC(ZZ-ZZ1300)ZZZ-ZZZ13)
        BLOC(Q-Q600)
        BLOC(N1-N80)Y1-Y40)GG1-GG400)A1-A40)B1-B40)$
        SYN (X1=N2)GX=GG1)GY=GG101)GZ=GG201)GW=GG301)
START    ENTER(SETDP0)
         READ(RCC)$ ENTER(CVFTOI)(RCC)(RC)
         ZZZ=.90$ ZZZ1=.80$ ZZZ2=.70$ ZZZ3=.60$ ZZZ4=.50
         ZZZ5=.40$ ZZZ6=.30$ ZZZ7=.20$ ZZZ8=.10$
         ZZZ9=.05$ ZZZ10=.02$ ZZZ11=.01$ ZZZ12=.005$
          SET(HP=1)W=15)G=0)$
AGAIN    READ-FORMAT(H)-(80)NOS.AT(N1)$ STOR=08
         INC(HP=HP+1)$
          NEX=09$   ENTER(PRINT B)
         MM=0$ MMM=0$SET(WW=0)$ ENTER(PRINT B)$
         ENTER(ZEROCC)$ SET(ALP=ALP1)(BTA=BTA1)$
ALPHA    SET(K=0)L=0)$ EPS=.00001$
1.0      N=N1,K$ X=X1,K$ D=C$ IF(X<100)GOTO(1.1)$
         IF(N-X<2)GOTO(1.1)$  Y=X/N$  GOTO(1.7)
1      A=1/(N+1)$  B=1$C=X+1$J=N-X$I=1$INT(RB=RB*RA)$
         BRLESC$B8(RB)(/OOMMM)(RB)$
         ENTER(CVXTOF)(RB)(R)
1.2      IF(B>J)GOTO(1.3)$
         A=A*B/C$  B=B+1$ C=C+1$ GOTO(1.2)
1.3       A'=A$GOTO(,BTA)$
1.4      YY=Y/(1-Y)$ F=(1-Y)$
         IF(Y<0)OR(Y>1)GOTO(2.3)
1.41     SS=EXP(X*LOG(Y)+J*LOG(F))
         D=D+1$ IF(D>50)GOTO(2.0)
         IF(D>47)GOTO(2.1)
1.43     IF(Y>.95)GOTO(1.8)
         S=S0=SS*Y/(X+1)$ I=1
1.5      IF(I=J+1)WITHIN(.001)GOTO(1.6)
         S=S*YY(J-I+1)/(X+I+1)
         S0=S0+S$ I=I+1$ GOTO(1.5)
1.6       GOTO(,ALP)$
ALP3      ET=(S0-A*R)/SS$IF(D>2)GOTO(1.61)$  ET=ET/2
1.61     Y=Y-ET$IF(Y=1)WITHIN(.0001)GOTO(1.7)$
         IF(Y>1)GOTO(2.4)
```

```
1.65        IF-ABS(ET>EPS)GOTO(1.4)
1.7         IF(Y>1)GOTO(2.4)    $  Y1,L=Y
            INC(K=K+2)$ INC(L=L+1)
            IF-INT(L<STOR)GOTO(1.0)$ INC(RESC=RESC+1)
            SET(M=0)
17.8        MM=MM +Y1,M$ MMM=MMM+(Y1,M*Y1,M)$
            COUNT(W)IN(M)GOTO(17.8)$ SET(M=0)$
TSTFWD      IF(Y1,M<=Y2,M)GOTO(10.22)
            Y50=Y1,M$ Y1,M=Y2,M$ Y2,M=Y50$ P=M
10.44       IF-INT(P=0)GOTO(TSTFWD)$ INC(P=P-1)
            IF(Y1,P<=Y2,P)GOTO(TSTFWD)$
            Y50=Y1,P$ Y1,P=Y2,P$ Y2,P=Y50$ GOTO(10.44)$
10.22       COUNT(W-1)IN(M)GOTO(TSTFWD)
             Q,WW=(Y,W-Y,(W-1))/(Y,W-Y2) $
            BB,WW=(Y,W-Y,(W-2))/(Y,W-Y2)$
            QQ,WW=(Y,W-Y,(W-1))/(Y,W-Y1)$
            HH,WW=(Y2-Y1)/(Y,W-Y1)$
            EE,WW=(Y3-Y1)/(Y,(W-2)-Y1)
            CC,WW=(Y3-Y1)/(Y,(W-1)-Y1)$
            DD,WW=(Y,W-Y,(W-2))/(Y,W-Y3)$
            RR,WW=(Y2-Y1)/(Y,(W-1)-Y1)$ INC(WW=WW+1)$
            IF-INT(RESC<RC)GOTO(ALPHA)
11.55       ENTER(CVITOF)(W)(FQ)$DID=MM/(FQ*600)$
            SID=(FQ*600*MMM-MM*MM)/(FQ*600(FQ*600-1))$
            PRINT<MEAN = >DID< VARIANCE = >SID
3.11        SET(P=0)
3.01        SET(M=0)
3.0         IF(QQ,M<=QQ1,M)GOTO(3.22)
            Y50=QQ,M$ QQ,M=QQ1,M$ QQ1,M=Y50$ L=M
3.1         IF-INT(L=0)GOTO(3.0)$ INC(L=L-1)
            IF(QQ,L<=QQ1,L)GOTO(3.0)$
            Y50=QQ,L$QQ,L=QQ1,L$ QQ1,L=Y50$ GOTO(3.1)$
3.22        COUNT(599)IN(M)GOTO(3.0)
            ZZ,G=QQ59$ ZZ1,G=QQ119$ ZZ2,G=QQ179$ ZZ3,G=QQ239
            ZZ4,G=QQ299$ ZZ5,G=QQ359$ ZZ6,G=QQ419$ZZ7,G=QQ479
            ZZ8,G=QQ539$ ZZ9,G=QQ569$ ZZ10,G=QQ587$
            ZZ11,G=QQ593$ ZZ12,G=QQ596$ INC(G=G+13)$
            INC(P=P+1)$ IF-INT(P> 7)GOTO(18.69)$
            IF-INT(P=1)GOTO(3.3)$ IF-INT(P=2)GOTO(3.4)$
            IF-INT(P=3)GOTO(3.5)$ IF-INT(P=4)GOTO(3.6)
            IF-INT(P=5)GOTO(3.7)$
            IF-INT(P=6)GOTO(3.03)$IF-INT(P=7)GOTO(3.8)$
3.3         MOVE(600)NOS.FROM(HH)TO(QQ)$ GOTO(3.01)
3.4         MOVE(600)NOS.FROM(Q )TO(QQ)$GOTO(3.01)
3.5         MOVE(600)NOS.FROM(RR)TO(QQ)$GOTO(3.01)
3.6         MOVE(600)NOS.FROM(BB)TO(QQ)$ GOTO(3.01)
3.7         MOVE(600)NOS.FROM(CC)TO(QQ)$ GOTO(3.01)
3.03        MOVE(600)NOS.FROM(DD)TO(QQ)$ GOTO(3.01)
```

```
3.8         MOVE(600)NOS.FROM(EE)TO(QQ)$ GOTO(3.01)
18.69        PRINT<  RB  >
            ENTER(SEXAPR)(RB)(RB)
            CLEAR(400)NOS.AT(GG1)$ SET(RESC=0)$ V=0$ VA=0$
            M3=0$ UA=0$ UB=0$ UC=0$ UX=0$ UY=0$ UZ=0$ M4=0$
            M1=0$ M2=0$UD=0$ UW=0$ SET(K=0)L=0$
            IF-INT(HP>6)GOTO(16.11)$ GOTO(AGAIN)$
 .8         S=S0=SS*F/(J+1)$ I=1
1.9         S=S(X-I+1)/YY*(J+I+1)
            S0=S0+S$ I=I+1$ IF(I=<X+.001)GOTO(1.9)
            S0=A-S0$ GOTO(1.6)
2.0         PRINT<       ERROR>
            PRINT(R)Y)N)X)ET)$ GOTO(1.0)
2.1         PRINT(R)Y)N)X)ET)D)$ GOTO(1.43)
2.2         Y=EXP(LOG(R)/(N+1))$GOTO(1.7)$
2.3         PRINT(Y)N)X)R)ET)D)$  GOTO(1.41)
2.4         SET(PP=0)
2.5         ET=ABS(ET/2)$ Y=Y-ET$ INC(PP=PP+1)$
            IF-INT(PP>20)GOTO(2.6)$ IF(Y>1)GOTO(2.5)$
            IF(Y=1)WITHIN(.0001)GOTO(1.7)$ GOTO(1.65)
2.6         CLEAR(400)NOS.AT(GG1)$ SET(RESC=0)$ V=0$ VA=0$
            M1=0$ M2=0$UD=0$ UW=0$
            M3=0$ UA=0$ UB=0$ UC=0$ UX=0$ UY=0$ UZ=0$M4=0$
            IF-INT(HP>6)GOTO(16.11)$ GOTO(AGAIN)$
F5      FORM(3-14)12-4)3-2)1-1)12-6)3-1)1-6)
H       FORM(10-10)10-10)
RB      SEXABRLESC(00L7NJ68L3K5LS003)$
RA      SEXABRLESC(00422NK8S0KC0K425)
BTA1        Y=(X-1)/N$A=A'(N+1)/X$X=X-1$ GOTO(1.4)$
BTA2        X=X+1$ IF(N<X+1+.001)GOTO(BT2.1)$ Y=(X+1)/N
            A=A'(N+1)/(N-X)$J=J-1       $ GOTO(1.4)$
BT2.1         S0=A=1$GOTO(,ALP)$
BTA3        Y=X/N$IF(N=X)WITHIN(.001)GOTO(2.2)$ GOTO(1.4)$
            Z1=(X-1)/N$Z2=(X+1)/N$IF(Z2<1)GOTO(2.7)$ Z2=1
2.7               Y=Z1+(Z2-Z1)(R-A1,L)/(B1,L-A1,L)$ GOTO(1.7)
ALP1         A1,L=S0/A$SET(ALP=ALP2)GOTO(BTA2)$
ALP2         B1,L=S0/A$ INC(K=K+2)(L=L+1)$
            SET(ALP=ALP1)(BTA=BTA1)$
            IF-INT(L<STOR)GOTO(1.0)$
            SET(ALP=ALP3)(BTA=BTA3)GOTO(ALPHA)$
16.11       SET(M=0)K=0)$
14.4        PRINT-(F5)-(ZZZ,K)(6)NOS.AT(ZZ,M/208)$
            INC(K=K+1)$ INC(M=M+1)$
            IF-INT(M>207)GOTO(N.PROB)$
            IF-INT(K<13)GOTO(14.4)$ SET(K=0)$ GOTO(14.4)
        LIST
        END GOTO(START)
```

```
                    JERRY THOMAS - GAMMA DISTRIBUTION
        BLOC(U-U600)V-V600)W-W600)XX-XX600)
        BLOC(Q-Q600)RR-RR600)
        BLOC(WW-WW600)VV-VV600)X-X700)EE-EE12)
        BLOC(S-S50)Y-Y50)
        SYN (Z=096)(CV=087)(ZZ=G1=088)(G2=089)(J=02)$
        SYN (Z1=08S)(T1=08K)$
START   READ-(F1)-(NR)M)$
        ENTER(SETDP0)$
        EE7=.2$ EE8=.1$ EE9=.05$ EE10=.02$
        EE=.9$ EE1=.8$ EE2=.7$ EE3=.6$ EE4=.5$
        EE5=.4$ EE6=.3$
        EE11=.01$ EE12=.005$
        SET(HP=0)$SET(H=0)$
        PRINT(23><R>14><ALPHA>19><Z>17>$$
2.1     READ-(F2)-(A)$
        INC(HP=HP+1)
        SET(E=0)P=0)D=0)$
        IF-INT(09=0)GOTO(10.0)$ SET(K=0)$
        IF(A=0)GOTO(2.2)$
        SET(FL=1)$ A'=A+1$ ENTER(LGAMMA)A')LGA)$
        ENTER(CVFTOI)A)AI)$
        GOTO(2.3)$
2.2     SET(FL=2)$
2.25    SET(K=0)$
2.3     B4(IR2)(/LLZ)(IR2)
2.3     MXR(IR1)IR2)0)$ SHR(0)0)IR2)$ R=IR2$
        TP(/Z7L)/ZLL)R)$ A(R)050)R)$
        IF(R>1)OR(R<0)GOTO(11.0)$
        GOTO,FL(2.4)2.8)$
2.4     Y'=A$ CFY=1-R$
2.5     S=EXP(A*LOG(Y')-Y'-LGA)$ SET(I=1)$ CON=0$SUM=S$
2.6     S,I=(A-CON)*S,(I-1)/Y'$ SUM=SUM+S,I$$
        CON=CON+1$ COUNT(AI+1)IN(I)GOTO(2.6)$
        DY=(CFY-SUM)/S$ Y'=Y'-DY$
        IF-ABS-NOT(DY<.0001)GOTO(2.5)$ GOTO(3.0)$
2.8     Y'=-LOG(1-R)$
3.0     Y,K=Y'$
        COUNT(M)IN(K)GOTO(2.3)$ SET(K=0)$
TSTFWD  IF(Y,K<=Y1,K)GOTO(10.22)
        Y60=Y,K$ Y,K=Y1,K$ Y1,K=Y60$ P=K$
10.44   IF-INT(P=0)GOTO(TSTFWD)$ INC(P=P-1)$
        IF(Y,P<=Y1,P)GOTO(TSTFWD)$
        Y60=Y,P$ Y,P=Y1,P$ Y1,P=Y60$ GOTO(10.44)$
```

```
10.22       COUNT(M-1)IN(K)GOTO(TSTFWD)
10.33       Q,E=(Y,(M-1)-Y,(M-2))/(Y,(M-1)-Y)
            U,E=(Y,(M-1)-Y,(M-2))/(Y,(M-1)-Y1)$
            V,E=(Y1-Y)/(Y,(M-2)-Y)$
            W,E=(Y,(M-3)-Y,(M-1))/(Y1-Y,(M-1))$
            XX,E=(Y2-Y)/(Y,(M-2)-Y)
            WW,E=(Y,(M-3)-Y,(M-1))/(Y2-Y,(M-1))$
            VV,E=(Y2-Y)/(Y,(M-3)-Y)
            RR,E=(Y1-Y)/(Y,(M-1)-Y)$ INC(E=E+1)$
            COUNT(NR)IN(D)GOTO(2.25)$
            PRINT<  R10  LOWER>
            SET(P=0)
9.5         SET(E=0)
11.1        IF(RR,E<=RR1,E)GOTO(11.22)
            Y60=RR,E$ RR,E=RR1,E$ RR1,E=Y60$ L=E
11.44       IF-INT(L=0)GOTO(11.1)$ INC(L=L-1)
            IF(RR,L<=RR1,L)GOTO(11.1)$
            Y60=RR,L$ RR,L=RR1,L$ RR1,L=Y60$ GOTO(11.44)$
11.22       COUNT(NR-1)IN(E)GOTO(11.1)
            X,H=RR59$ X1,H=RR119$ X2,H=RR179$ X3,H=RR239$
            X4,H=RR299$ X12,H=RR596$
            X5,H=RR359$ X6,H=RR419$ X7,H=RR479$ X8,H=RR539$
            X9,H=RR569$ X10,H=RR587$ X11,H=RR593$
            INC(H=H+13)$
            INC(P=P+1)$ IF-INT(P>7)GOTO(6.3)
            IF-INT(P=1)GOTO(3.3)$ IF-INT(P=2)GOTO(3.4)$
            IF-INT(P=3)GOTO(3.5)$ IF-INT(P=4)GOTO(3.6)$
            IF-INT(P=5)GOTO(3.7)$
            IF-INT(P=6)GOTO(3.8)$ IF-INT(P=7)GOTO(3.9)$
3.3         MOVE(600)NOS.FROM(Q)TO(RR)$ GOTO(9.5)
3.4         MOVE(600)NOS.FROM(U)TO(RR)$ GOTO(9.5)
3.5         MOVE(600)NOS.FROM(V)TO(RR)$ GOTO(9.5)
3.6         MOVE(600)NOS.FROM(W)TO(RR)$ GOTO(9.5)
3.7         MOVE(600)NOS.FROM(XX)TO(RR)$ GOTO(9.5)
3.8         MOVE(600)NOS.FROM(WW)TO(RR)$ GOTO(9.5)
3.9         MOVE(600)NOS.FROM(VV)TO(RR)$ GOTO(9.5)
6.3         IF-INT(HP<6)GOTO(2.1)
            SET(H=0)(P=0)
13.0        PRINT-(F6)-(EE,P)(6)NOS.AT(X,H/104)
            INC(H=H+1)$ INC(P=P+1)$ IF-INT(H>103)GOTO(10.0)$
            IF-INT(P<13)GOTO(13.0)$ SET(P=0)$ GOTO(13.0)$
10.0        ENTER(SEXAPR)IR2)IR2)$
            GOTO(N.PROB)$$
11.0        PRINT<R=>(R)$ GOTO(N.PROB)$
ERROR       PRINT-(F3)-(R)A)Y')DY)$ GOTO(N.PROB)$
GAMMA       TP6(SELF+1)(045)(ALPH)$ J(AL1)(AL2)(SELF+2)
LGAMMA      TP11(SELF-1)(046)(ALPH)$ AX(1)(EX1)(EXIT)$
```

```
          TP(3)(047)(STOR)$
          Z=,2$ Z1=,2$ CV=1$
1.1       IF(Z>D1)GOTO(1.3)$
1.2       CV=CV*Z$ Z=Z+051,$ GOTO(1.1)
1.3       CB12(CV) )E.1)$ IF-ABS(CV<1)GOTO(1.2)$
          ZZ=1/Z*Z$ SET(J=0)$ B12(C0))$
1.4       PMA(ZZ)(0)(C1,J)$ LPI(,J)(6)(1.4)$ D(0)(Z)(G1)$
          G2=LOG(Z)$ S(Z)(D2)(0)$ M(G2)(0)(G2)$
          AA(G1)(D3)(G2)$ G2=G2-Z$
ALPH      GOTO(AL  1)$
AL2       G1=ABS(CV)$ G1=LOG(G1)$ S(G2)(G1)(0)$ GOTO(STOR)$
AL1       IF(G2>D4)GOTO(E.1)$ G1=EXP(G2)$ D(G1)(CV)(0)$
STOR      B12(0)(0)(0)$
EXIT      GOTO(EXIT)$
EX1       GOTO(1)$
E.1       SX(EXIT)(EX1)(1)$ SET(2=EWD)(3=Z1)GOTO(060)
EWD   ALFNGAMMA
D1    DEC (10.)
D2    DEC (.5)
D3    DEC (.91893853320467267301)
C0    DEC (.00641025641)
C1    DEC (-.001917526918)(.C008417508418)(-.0005952381)
      DEC (.0007936507936508)(-.0027777777777777777)
      DEC (.083333333333333333)
D4    DEC (350.)
IR1   DEC (I74505805969238281 25)
IR2   SEXA(0579K2F59S9820KS6)
F1    FORM(4-10)1-2)
F2    FORM(10-10)$
F5    FORM(12-6)3-2)1-10)
F3    FORM(12-6-13)3-2)12-2-6)3-4)12-6-13)3-2)12-6-13)$
F4    FORM(3-14)12-6-13)3-10)12-2-6)3-10)12-6-13)3-14)$
F6    FORM(3-14)12-4)3-2)1-1)12-6)3-1)1-6)$
      LIST
      END GOTO(START)$$
```

A METHOD FOR APPROXIMATING PROBABILITY FUNCTIONS DEFINED ON FINITE DOMAINS

Joseph S. Tyler, Jr.
Systems Analysis Office
Edgewood Arsenal, Maryland

I. INTRODUCTION. The incentive for this paper arose from the requirement to determine approximately the probability density function h(d) of the random variable D from a knowledge of the first r-moments (about the origin) of that variate.

Specifically, the moments are computed from equation (1.0).

$$(1.0) \quad M_r(D) = \iiiint_{\Omega} D^r(\zeta, \eta, u, v)\, f(\zeta, \eta, u, v)\, d\zeta d\eta du dv \qquad \text{for } (r=0,1,2,\ldots,N).$$

Where D is a known continuous function and f is a known continuous probability density function of the variates ζ, η, u, v. Moreover, the range of D is known, $0 \leq D \leq 1$, and the integration is performed over the Euclidean four-dimensional space Ω.

It has been demonstrated, by H. Hamburger 1920 (Ref 1), that when the domain of definition of a probability function is finite then that function is uniquely determined by the set of all of its moments. A method of constructing a probability density function defined on [-1,1], from the infinite set of its moments, has been published by Philip Davis in his book, INTERPOLATION AND APPROXIMATION, 1961 (Ref 2). The method is essentially an infinite series expansion in Legendre polynomials. However, from a statistical viewpoint, it is not practical to construct the required function h from the entire set of its moments. The purpose of this paper, therefore, is to present a method, employing only the first r-moments, by which nonnegative approximations of probability density functions on [0,1] can be constructed.

Essentially, the approximation method is based on an iterative procedure. The first step utilizes the first r-moments of the random variable D to specify the initial approximation to the function h. Secondly, successive improvements over the initial approximation are achieved by applying a modified version of the classical method for representing continuous functions by orthonormal polynomials. The error of the approximation is measured in terms of the given original first r-moments of the variate D.

II. ESSENTIAL ASPECTS FROM THE CLASSICAL THEORY. In general, any continuous function g(x) defined on the finite interval [0,1], can be expanded in a series of weighted orthonormal polynomials $w(x) \sum c_i \theta_i(x)$.

Specifically, it will be required that the following set of conditions be satisfied:

A. $g(x) \in C'[0,1]$, (i.e., the function g and its first derivative be continuous on the closed interval [0,1]).

B. A sequence of polynomials $\{\theta_n(x)\}$ continuous, bounded and orthonormal with respect to some weight function $w(x) \geq 0$ [0,1] are known.

C. $w(x)\ g(x)$ and $w(x)\ g^2(x)$ be integrable on [0,1].

The sequence of polynomials $\{\theta_n(x)\}$ satisfy the following properties:

$$[P_1] \qquad \int_0^1 w(x)\ \theta_i(x)\ \theta_j(x)\ dx = 0, \qquad i \neq j$$

$$[P_2] \qquad \int_0^1 w(x)\ \theta_i^2(x)\ dx = 1$$

$$[P_3] \qquad \int_0^1 x^r\ w(x)\ \theta_i(x)\ dx = 0, \qquad r<i$$

$$[P_4] \qquad \int_0^1 P_m(x)\ w(x)\ \theta_i(x)\ dx = 0, \qquad m<i$$

$a_n>0$, denotes the coefficient of x^n in $\theta_n(x)$

$[P_5]$ $a_{n+1}>0$, denotes the coefficient of x^{n+1} in $\theta_{n+1}(x)$

($P_m(x)$ denotes any polynomial of degree m).

Under the orthogonality conditions on $\{\theta_i\}$, the expansion of $g(x)$ can be expressed as:

$$(1) \qquad g(x) \simeq \sum_{i=0}^{\infty} c_i \cdot \theta_i(x),\ c_i = \int_0^1 w(x)\ g(x)\ \theta_i(x)\ dx.$$

Let $S_n(x)$ denote the partial sum,

$$(1.1) \qquad S_n(x) = \sum_{i=0}^{n} c_i \ \theta_i(x).$$

Then by the definition of c_i's and the orthonormal properties of the θ_i's, we have

$$(1.2) \qquad \int_0^1 w(x)[g(x) - S_n(x)]^2 \, dx = \int_0^1 w(x)[g(x)]^2 dx - \sum_{i=0}^{n} c_i^2 .$$

Now that the first member of equation (1.2) is nonnegative, the same is true of the second member and,

$$(1.3) \qquad \sum_{i=0}^{n} c_i^2 \leq \int_0^1 w(x)[g(x)]^2 \, dx, \text{ for all values of } n.$$

Consequently,

$$(1.4) \qquad \sum_{i=0}^{n} c_i^2 \text{ is convergent, for } n \to \infty, \text{ and}$$

$$\lim_{i \to \infty} c_i = 0.$$

Hence, we conclude that $S_n(x)$ converges to $g(x)$ in the least square sense over the finite interval $[0,1]$.

Under the assumption that $g'(x)$ is continuous on $[0,1]$, it can be demonstrated that $S_n(x)$ converges to $g(x)$ for every $x \varepsilon [0,1]$ as n increases without bound.

The Christoffel-Darboux identity (Ref 3) provides the following symmetric kernel function $K_n(x,t)$, of order n, for the system of polynomials $\theta_i(x)$. That is,

$$(2) \qquad K_n(x,t) = K_n(t,x) = \sum_{i=0}^{n} \theta_i(t) \ \theta_i(x)$$

or

$$K_n(x,t) = \frac{a_n}{a_{n+1}} \frac{\theta_{n+1}(t)\,\theta_n(x) - \theta_n(t)\,\theta_{n+1}(x)}{t - x},$$

where

$a_n > 0$ is the coefficient of x^n in $\theta_n(x)$

$a_{n+1} > 0$ is the coefficient of x^{n+1} in $\theta_{n+1}(x)$

$a_{n+1} > a_n$.

By utilizing this identity, we may express $S_n(x)$, equation (1.1), as

$$S_n(x) = \int_0^1 w(t)\, g(t)\, K_n(x,t)\, dt, \tag{2.1}$$

and from the orthonormal properties of the θ_i's we have,

$$1 = \int_0^1 w(t)\, K_n(x,y)\, dt. \tag{2.2}$$

Multiplication of equation (2.2) by $g(x)$, which is constant with respect to the variable of integration gives

$$g(x) = \int_0^1 w(t)\, g(x)\, K_n(x,t)\, dt. \tag{2.3}$$

Hence, by subtraction of equation (2.3) from (2.1),

$$S_n(x) - g(x) = \int_0^1 w(t)[g(t) - g(x)]\, K_n(x,t)\, dt. \tag{2.4}$$

Then, by substitution of $K_n(x,t)$, equation (2), in equation (2.4) one obtains, for an arbitrary $x \varepsilon [0,1]$, the relation.

$$S_n(x) - g(x) = \left[\frac{a_n}{a_{n+1}} \theta_n(x) \int_0^1 w(t) \frac{g(t)-g(x)}{t-x} \theta_{n+1}(t)\, dt \right.$$

(2.5)

$$\left. -\theta_{n+1}(x) \int_0^1 w(t) \frac{g(t)-g(x)}{t-x} \theta_n(t)\, dt \right] .$$

The proof that $S_n(x)$ converges to $g(x)$ on $[0,1]$ consists of showing that equation (2.5) approaches zero as n becomes infinite.

Since, by hypothesis, the θ_i's are bounded and $a_n > 0$, $a_{n+1} > 0$, then (a_n/a_{n+1}) is also bounded. Moreover, the derivative g' is continuous on $[0,1]$.

That is,

$$g'(x) = \lim_{t \to x} \frac{g(t)-g(x)}{t-x} , \tag{2.6}$$

and from equations (1) and (1.4), it follows that

$$\lim_{j \to \infty} c_j' = \lim_{\substack{t \to x \\ j \to \infty}} \int_0^1 w(x) \frac{g(t)-g(x)}{t-x} \theta_j(x)\, dx = 0. \tag{2.7}$$

where the index j denotes either n or n+1 in equation (2.5). Therefore, $S_n(x)$ approaches $g(x)$ for every $x \varepsilon [0,1]$ and the expansion of $g(x)$ can be written as

$$g(x) = \sum_{i=0}^{\infty} c_i \cdot \theta_i(x), \text{ for } x \varepsilon [0,1]. \tag{2.8}$$

In the derivation of the method of approximating probability functions defined on finite domains, the following theorem for weighted orthonormal polynomials will be required:

Theorem 1. Let $H(x)$ denote a polynomial of degree m that is nonnegative on $[0,1]$. Let $\theta_i(x)$, $i=0,1,\ldots$, be the orthonormal polynomials corresponding to the weight function $w(x)$ on $[0,1]$. Let $q_i(x)$, $i = 0,1,\ldots$, be the orthonormal polynomials associated with the weight function $H(x)w(x)$. Then boundedness of the θ_i's assures the boundedness of the q_i's.

Proof. The product $H(x)\ q_n(x)$ is a polynomial of degree $n+m$, and can be expressed in the form

$$H(x)\ q_n(x) = \sum_{i=0}^{n+m} c_{ni}\ \theta_i(x)\ , \tag{3}$$

where

$$c_{ni} = \int_0^1 w(x)\ H(x)\ q_n(x)\ \theta_i(x)\ dx.$$

If $i < n$, then $c_{ni} = 0$ as a consequence of the orthogonality properties of $q_n(x)$ with respect to the weight function $H(x)\ w(x)$. So that,

$$H(x)\ q_n(x) = \sum_{i=n}^{n+m} c_{ni}\ \theta_i(x). \tag{3.1}$$

As for the coefficients c_{ni} which do not vanish,

$$|c_{ni}| \leq \int_0^1 [w(x)]^{1/2}\ H(x)|q_n(x)|\cdot[w(x)]^{1/2}|\theta_i(x)|\ dx \tag{3.2}$$

$$c_{ni}^2 \leq \int_0^1 w(x)[H(x)]^2\ [q_n(x)]^2\ dx \cdot \int_0^1 w(x)[\ \theta_i(x)]^2\ dx.$$

The last expression follows from Schwarz's inequality, and the last integral is equal to 1, since the θ_i's are normalized.

Let $G = \underset{x\in[0,1]}{\text{Max}}\ [H(x)]$, then

$$c_{ni}^2 \leq G\int_0^1 w(x)\ H(x)\ [q_n(x)]^2\ dx = G. \tag{3.3}$$

So that, $|c_{ni}| \leq G^{1/2}$, and since θ_i is bounded, that is $|\theta_i| \leq A$, for all $x\in[0.1]$, we have from equation (3.1).

(3.4) $$|H(x)\, q_n(x)| = |H(x)| \cdot |q_n(x)| \leq G^{1/2} A (m+1).$$

The polynomial $H(x)$ by hypothesis has a lower positive bound on $[0,1]$; therefore, the polynomials $q_n(x)$ are also bounded on $[0,1]$.

III. APPROXIMATING PROBABILITY DENSITY FUNCTIONS ON [0,1]. The information and results discussed in Section II, is next utilized in the formulation of a method for constructing nonnegative approximations of continuous probability density functions defined on the closed domain $[0,1]$.

A. Assumptions and Notations

1. Let $f(x)$ denote a probability density function, and $f'(x)$ its first derivative, and assume that both $f(x)$ and $f'(x)$ are continuous on $[0,1]$.

2. It is assumed that the first r-moments, m_j, $j = 0,1, \ldots,r$, of $f(x)$ are known.

3. Let $P_r(x)$ denote a polynomial of degree r.

4. It is assumed that the orthonormal polynomials $\{\theta_i(x)\}$, associated with weight function $w(x)$, are known.

B. The Initial Approximation

The probability density function $f(x)$, by equation (2.8), can be represented by the following expansion:

(4) $$\frac{f(x)}{w(x)} = g(x) = \sum_{i=0}^{\infty} c_i \cdot \theta_i(x), \text{ for } x \epsilon [0,1]$$

or

$$f(x) = w(x) \sum_{i=0}^{\infty} c_i \cdot \theta_i(x).$$

The coefficients c_i's are computed from the relation

(4.1) $$c_i = \int_0^1 \theta_i(x)\, f(x)\, dx.$$

Now that $\theta_i(x)$ is a polynomial degree i, it can be written as

$$\theta_i(x) = \sum_{j=0}^{i} a_{ij}x^j, \quad a_{ii} > 0, \tag{4.2}$$

and equation (4.1) can, therefore, be expressed as

$$c_i = \sum_{j=0}^{i} a_{ij} \int_0^1 x^j f(x)\, dx \tag{4.3}$$

or by

$$c_i = \sum_{j=0}^{i} a_{ij}\, m_j, \tag{4.4}$$

where

$$m_j = \int_0^1 x^j f(x)\, dx, \quad (j=0,1, \ldots,r).$$

The finite set of moments m_j, (j=0,1, ...,r), are reproducible from the expansion given by equation (4). That is,

$$m_j = \int_0^1 x^j f(x)\, dx = \sum_{i=0}^{j} c_i \int_0^1 x^j w(x)\ \theta_i(x)\, dx$$

$$+ \sum_{i=j+1}^{\infty} c_i \int_0^1 x^j w(x)\ \theta_i(x)\, dx. \tag{4.5}$$

By property $[P_3]$, the last integral is zero; therefore,

$$m_j = \sum_{i=0}^{j} c_i \int_0^1 x^j w(x)\ \theta_i(x)\, dx. \quad (j=0,1,\ldots,r) \tag{4.5.1}$$

As a consequence of equation (4.5), the initial approximation for the density function f(x), based on its r-moments m_r, has the following form

$$f(x) \simeq w(x) \sum_{i=0}^{r} c_i \cdot \theta_i(x). \tag{4.6}$$

Moreover, it is observed that the reproducibility of the moments possessed by f(x), is independent of the choice of the weight function.

C. Successive Improvements Over Initial Approximation

It may happen that the initial approximation may become negative on [0,1], and in this event it is not a satisfactory representation of the given probability density finction f(x). The following approximation scheme is introduced so as to remove the possibility of obtaining a negative approximation for f(x). For this purpose, the initial approximation can be rewritten as

$$f(x) \simeq f_o(x) = w(x) \sum_{i=0}^{r} c_i \theta_i(x). \tag{4.7}$$

If $f_o(x)$ is nonnegative on [0,1], then the approximation of f(x) by $f_o(x)$ possesses the same first r-moments as possessed by f(x), by equation (4.5), and the process is therefore terminated at this step. However, if this is not the case, then a positive constant h_1 can be determined such that

$$w_1(x) = \frac{w(x)}{h_1+1} \left[h_1 + \sum_{i=0}^{r} c_i \theta_i(x) \right] > 0, \text{ for } x \varepsilon [0,1]. \tag{4.7.1}$$

(The method by which $h_1 > 0$ is determined is presented in subsection F.)

The first improvement over the initial approximation $f_o(x)$ is obtained by constructing a new sequence of orthonormal polynomials $\{q_i^{(1)}(x), i=0,1, \ldots, r\}$ with respect to the new weight function $w_1(x)$ (the sequence $\{q_i^{(1)}\}$ can be obtained by applying the Schmidt orthonormalization process (Ref 3) and then computing a new approximation by applying equations (4.4) and (4). The new approximation has the following form

$$(4.8) \qquad f(x) \simeq f_1(x) = w_1(x) \sum_{i=0}^{r} c_i^{(1)} q_i^{(1)}(x)$$

for

$$c_i^{(1)} = \int_0^1 q_i^{(1)}(x) f(x)\, dx = \sum_{j=0}^{i} a_{ij}^{(1)} m_j .$$

Moreover,

$$(4.8.1) \quad w_1(x) \sum_{i=0}^{r} c_i^{(1)}, \quad q_i(x) = w(x) \left[\frac{h_{1+} \sum_{i=0}^{r} c_i \theta_i(x)}{h_1+1} \right] .$$

$$\sum_{i=0}^{r} c_i^{(1)} q_i^{(1)}(x)$$

or

$$f_1(x) = w(x) P_{2r}(x),$$

and since $P_{2r}(x)$ is a polynomial of degree 2r we can write equation (4.8) as

$$(4.8.2) \qquad f(x) \simeq f_1(x) = w(x) \sum_{i=0}^{2r} d_i \theta_i(x)$$

where,

$$d_i = \int_0^1 \theta_i(x) f(x)\, dx .$$

If $f_1(x)$, equation (4.8), is nonnegative on [0,1], then the process is terminated with the first improvement over the initial approximation. If, however, $f_1(x)$ becomes negative on [0,1], then a second positive constant h_2 is determined and the computations indicated by equations (4.7.1) and (4.8) are repeated.

The results obtained after repeating the above process n times is expressed by the following relations:

$$f(x) \simeq f_n(x) = w_n(x) \sum_{i=0}^{r} c_i^{(n)} q_i^{(n)}(x) \tag{4.9}$$

where,

$$c_i^{(n)} = \int_0^1 q_i^{(n)}(x) f(x)\, dx = \sum_{j=0}^{i} a_{ij}^{(n)} m_j$$

$$w_n(x) = \frac{w_{n-1}(x)}{h_n+1} \left[h_n + \sum_{i=0}^{r} c_i^{(n-1)} q_i^{(n-1)}(x) \right] \tag{4.9.1}$$

$$w_n(x) = w_{n-1}(x) \cdot P_r(x)$$

$$w_n(x) = w(x) \cdot P_{nr}(x)$$

$$f(x) \simeq f_n(x) = w(x) \cdot P_{(n+1)r}(x) \tag{4.9.2}$$

$$f_n(x) = w(x) \sum_{i=0}^{(n+1)r} d_i \theta_i(x)$$

where

$$d_i = \int_0^1 \theta_i(x) f(x)\, dx.$$

D. Convergence of Process.

The convergence of the above process can be demonstrated by applying equation (1.2) along with the following replacements or substitutions:

1. $g(x) = \dfrac{f(x)}{w(x)}$, $c_i = d_i$

2. $S_n(x) = f_n(x)$, $n = (n+1)r$

Equation (1.2) then becomes,

$$(4.10) \qquad \int_0^1 w(x)\left[\frac{f(x)}{w(x)} - f_n(x)\right]^2 dx = \int_0^1 \frac{f^2(x)}{w(x)}\,dx - \sum_{i=0}^{(n+1)r} d_i^2 \geq 0$$

and $\sum_{i=0}^{(n+1)r} d_i^2$ is convergent provided $\dfrac{f^2(x)}{w(x)}$ is integrable over $[0,1]$ as $n \to \infty$, also $\lim_{n \to \infty} d_i = 0$.

E. The Error E_r^n.

Having assumed that the finite set of moments $\{m_j,\ j=0,1,\ \ldots,r\}$ are known, we then essentially carry out the approximation $f_n(x)$ by applying equation (4.9). The coefficients $c_i^{(n)}$ are computed from the given set of moments and it appears natural to measure the error of the approximation in terms of the moments.

The error $E_r^{(n)}$ is defined by

$$(4.11) \qquad E_j^{(n)} = \left|\lambda_j^{(n)} - m_j\right|, \text{ for } j=0,1,\ldots,r$$

with

$$\lambda_j^{(n)} = \int_0^1 x^j f_n(x)\,dx,$$

$$\delta_j^{(n)} = \int_0^1 x^j w_{(n)}(x)\,dx.$$

Next let,

$$(4.11.1) \qquad f_1(x) = w_1(x) = \frac{1}{h_1+1}\left[h_1 w(x) + w(x) \sum_{i=0}^{r} c_i \cdot \theta_i(x)\right]$$

then

$$\lambda_j^{(1)} = \frac{1}{h_1+1}\, h_1 \left[\delta_j^{(o)} + m_j\right]$$

and

$$E_j^{(1)} = \frac{h_1}{h_1+1} \left| \delta_j^{(o)} - m_j \right| .$$

The errors $E_j^{(2)}$ are determined as follows:

$$(4.11.2) \qquad f_2(x) = w_2(x) = \frac{1}{h_2+1}\left[h_2\, w_1(x) + w_1(x) \sum_{i=0}^{r} c_i^{(1)}\, q_i^{(1)}\, (x)\right] .$$

So that

$$\lambda_j^{(2)} = \frac{1}{h_2+1}\left[h_2 \;\; \delta_2^{(1)} + m_j\right]$$

and

$$E_j^{(2)} = \frac{h_2}{h_2+1} \left| \delta_2^{(1)} - m_j \right| .$$

The error $E_j^{(n)}$, associated with the n-th approximation has the form

$$(4.12) \qquad E_j^{(n)} = \frac{h_n}{h_n+1} \left| \delta_j^{(n-1)} - m_j \right| , \quad (j=0,1,\ldots,r)$$

and this error approaches zero, as $n \to \infty$, provided that $h_n \to 0$.

F. Construction of Positive Constants h_n.

It has previously been shown that the approximation $f_n(x)$, equation (4.9), has the same first r-moments that are possessed by $f(x)$. In order that $f_n(x)$ represent a probability density function it is necessary that $f_n(x)$ be nonnegative on the domain [0,1].

The nonnegativity of the approximation is next considered. The approximation, at the n-th step, can be expressed as follows:

$$f(x) \simeq w(x)\, h(x) \quad , \; x \epsilon [0,1] \tag{5}$$

where,

$$h(x) = \sum_{i=0}^{r} c_i \, q_i(x) \tag{5.1}$$

and

$$q_i(x) = \sum_{j=0}^{i} a_{ij} \, x^j \quad . \tag{5.2}$$

By substituting $q_i(x)$ equation (5.2) in equation (5.1), $h(x)$ can be written in the form of a polynomial. This is,

$$h(x) = \sum_{i=0}^{r} A_{ri} \, x^i \tag{5.3}$$

for

$$A_{ri} = \sum_{j=i}^{r} c_i \, a_{ij} \, . \tag{5.4}$$

By definition $w(x)$ is nowhere negative on [0,1], and if the approximation becomes negative on (0,1) it is due to $h(x)$ being negative on that interval. Therefore, if the polynomial $h(x)$ has a real root, of order one, on (0,1) it implies there exist at least one

point $x_o \in (0,1)$ such that $h(x_o) < 0$. The following two theorems can be applied to determine the possibility of a real root of h(x) on (0,1).

Theorem 2: On the upper bound for the real roots of a polynomial.

Let R(x) be a polynomial of degree r, and let $R^{(k)}(x=1) \geq 0$ for $(k=0,1,\ldots,r)$, where $R^{(k)}$ denotes the k-th derivative of polynomial R(x). Then the point x=1 is an upper bound for the real roots of R(x).

Proof: By Taylor's formula, the polynomial R(x) can be expanded about the point x=1. That is,

$$R(x) = R(x=1) + \sum_{k=1}^{r} \frac{(x-1)^k}{k!} R^{(k)}(x=1). \tag{5.5}$$

By hypothesis

$$R^{(k)}(x=1) \geq 0, \text{ for } (k=0,1,\ldots,r). \tag{5.6}$$

Hence, R(x) for x > 1, by equation (5.5) is also positive. Therefore, x=1 is an upper bound for all the real roots of R(x).

Theorem 3: On the lower bound for the positive real roots of a polynomial.

Let h(x) be a polynomial of degree r, and denote its real roots by the set $\{x_1, \ldots, x_r\}$.

That is,

$$h(x) = b_o x^r + b_1 x^{r-1} + \ldots + b_r, \text{ or} \tag{5.7}$$

$$h(x) = b_o(x-x_1)(x-x_2)\ldots(x-x_r). \tag{5.8}$$

Next, let polynomial R(x) be defined in the following manner

$$R(x) = x^r h(1/x), \text{ or} \tag{5.9}$$

(5.10) $$R(x) = x^r [b_o x^{-r} + \ldots + b_r], \text{ or}$$

(5.11) $$R(x) = b_o + b_1 x + \ldots + b_r x^r, \text{ or}$$

(5.12) $$R(x) = b_o (1-x_1 x)\ (1-x_2 x) \ldots (1-x_r x).$$

The polynomial R(x) equation (5.11) simply reverses the order of the coefficients of h(x) and the roots of R(x) equation (5.12) are simply the reciprocals of the roots of h(x).

If no real root of h(x) lies in (0,1), then no real root of R(x) lies in the interval $(1,\infty)$ and by theorem 2, x=1 is an upper bound for the real roots of R(x). Moreover, x=1 is a lower bound for the positive real roots of h(x).

Proof: The roots of R(x) are the elements of the set $[1/x_i : (i=1,2,\ldots,r)]$. Moreover,

(5.13) (1). $x_i < 0$ implies $\dfrac{1}{x_i} < 0$, and

(5.14) (2). $x_i > 1$ implies $\dfrac{1}{x_i} < 1$.

Hence, (1) and (2) together imply that R(x) has no real roots in the open interval $(1, \infty)$. Therefore, x=1 is an upper bound for the real roots of R(x). Moreover, x = 1 is a lower bound for the positive real roots of h(x) which implies that h(x) has no real roots on the interval (0,1).

By the application of theorems (2) and (3), a test can be constructed to determine the positivity of the polynomial h(x) on (0,1). From equation (5.3) h(x) is defined as

(5.15) $$h(x) = \sum_{i=0}^{r} A_{ri}\, x^i$$

Let R(x) be the polynomial

(5.16) $$R(x) = x^r\, h\left(\frac{1}{x}\right) = \sum_{i=0}^{r} A_{ri}\, x^{r-i} \quad .$$

Then the k-th derivative of R(x) is denoted by

$$R^{(k)}(x) = \sum_{i=0}^{(r-k)} \frac{(r-i)!}{(r-i-k)!} A_{ri} x^{r-i-k} , \tag{5.17}$$

or

$$R^{(j)}(x) = \frac{r!}{j!} A_{ro} X^{j} + \sum_{i=1}^{j} \frac{(r-i)!}{(j-i)!} A_{ri} x^{j-i} , \tag{5.18}$$

and if h(x) has no real roots on (0,1), then by theorem (3),

$$R^{(j)}(x=1) = \frac{r!}{j!} A_{ro} + \sum_{i=1}^{j} \frac{(r-i)!}{(j-i)!} A_{ri} \geq 0. \tag{5.19}$$

That is,

$$A_{ro} \geq - \frac{j!}{r!} \sum_{i=1}^{j} \frac{(r-i)!}{(j-i)!} A_{ri} \tag{5.20}$$

for (j=0,1, ..., r).

When the relation on A_{ro} (equation 5.20) is satisfied the approximation, equation (5), is positive on (0,1) and represents the probability density function.

However, if $R^{(j)}(x=1) < 0$, for any value of j, then h(x) must be modified such that its modified form becomes a positive function on (0,1). Essentially, the constant term A_{ro} is increased by some positive constant h until equation (5.20) is satisfied for all values of j=0,1,...,r.

The product of w(x) and the modified positive function h(x) produces a new weight function and this new weight function is then used to generate a new set of orthonormal polynomials $\{q_i(x): i=0,1,...,r\}$ needed to obtain the next improved approximation. The process is

terminated when the succeeding approximation becomes positive on (0,1) and the succeeding weight function has moments that are arbitrarily close to the moments of f(x).

REFERENCES

1. Kendall, Maurice G., The Advanced Theory of Statistics, Vol. I, Hafner Publishing Company, New York, Fifth Edition, 1952.

2. Davis, Philip J., Interpolation and Approximation, Blaisdell Publishing Company, New York, 1963.

3. Jackson, Dunham, Fourier Series and Orthogonal Polynomials, The Mathematical Association of America, University of Minnesota, Fourth Edition, 1957.

LIST OF ATTENDEES

Adams, G.	Edgewood Arsenal, Maryland
Alexander, G. D.	Edgewood Arsenal, Maryland
Angers, Claude	Canadian Armanent Research and Development Establishment, Quebec City, Canada
Arnold, W. N.	Aberdeen Proving Ground, Maryland
Atkinson, John	Harvard Computing Center, Massachusetts
Baker, C.	Edgewood Arsenal, Maryland
Bargmann, R. E.	University of Georgia, Athens, Georgia
Bates, Carl B.	Fort Belvoir, Virginia
Bechhofer, Robert	Cornell University, Ithaca, New York
Beck, J.	Edgewood Arsenal, Maryland
Bell, R.	Edgewood Arsenal, Maryland
Benjamin, W.	Edgewood Arsenal, Maryland
Benton, A. W.	Edgewood Arsenal, Maryland
Birley, Stanley M.	Rock Island Arsenal, Illinois
Biser, Erwin	Fort Monmouth, New Jersey
Braaten, Melvin O.	Duke University, Durham, N. C.
Brennan, J.	Edgewood Arsenal, Maryland
Brown, James R.	UAAMC, Washington, D. C.
Bruchey, Wm. J., Jr.	Aberdeen Proving Ground, Maryland
Bunker, W. B.	Dep. Command Gen., USAMC, Washington, D.C.
Burton, D. F.	Army Research Office - Durham, N. C.
Cameron, Joseph M.	National Bureau of Standards, Washington
Chrepta, M. M.	Fort Monmouth, New Jersey
Chichowicz, J.	Edgewood Arsenal, Maryland
Clark, O.	Edgewood Arsenal, Maryland
Cohen, A. C., Jr.	Univers ty of Georgia, Athens, Georgia
Condon, G.	Edgewood Arsenal, Maryland
Connolly, Walter	Watervliet Arsenal, New York
Coon, Helen J.	Aberdeen Proving Ground, Maryland
Cousin, T.	Edgewood Arsenal, Maryland
Craw, A. R.	Natl. Bureau of Standards, Washington
Culpepper, G. A.	White Sands Missile Range, New Mexico
Davis, H. T., III	Sandia Base, Albuquerque, New Mexico
DeArmon, Ira A., Jr.	Edgewood Arsenal, Maryland
DeFranks, S.	Aberdeen Proving Ground, Maryland
DeGracie, J. S.	Dugway Proving Ground, Utah
DeSmidt, G.	Edgewood Arsenal, Maryland
Delp, J.	Edgewood Arsenal, Maryland
Dobrindt, G. T.	Aberdeen Proving Ground, Maryland
Dressel, F. G.	Army Research Office - Durham, N. C.
Duff, James B.	Ft. Belvoir, Virginia
Duncan, Acheson J.	Johns Hopkins University, Baltimore, Md.
Dutoit, Eugene	Picatinny Arsenal, Dover, New Jersey

Eberhart, L. M.	HQS, DA, Washington, D. C.
Eissner, V. M.	Aberdeen Proving Ground, Maryland
Ellner, H.	USAMC, Washington, D. C.
Fichter, L. S.	Picatinny Arsenal, Dover, New Jersey
Finanger, K.	Aberdeen Proving Ground, Maryland
Fiske, E.	Edgewood Arsenal, Maryland
Frishman, Fred	Army Research Office, Arlington, Va.
Galbraith, A. S.	Army Research Office, Durham, N. C.
Gebert, J. R.	Ft. Detrick, Frederick, Maryland
Gerard, S.	Aberdeen Proving Ground, Maryland
Germond, Hal H.	Albuquerque, New Mexico
Gray, B. C.	Ft. Detrick, Frederick, Maryland
Greenwood, Joseph	Justice Dept., Arlington, Virginia
Gross, L.	Edgewood Arsenal, Maryland
Hanson, Fred S.	White Sands Missile Range, New Mexico
Harris, Bernard	University of Wisconsin, Madison, Wis.
Harshbarger, Boyd	Virginia Polytechnic Institute, Blacksburg, Virginia
Heyderman, A. J.	Picatinny Arsenal, Dover, New Jersey
Hilton, H. G.	Fort Detrick, Frederick, Maryland
Hoeben, D.	Natl. Bureau of Standards, Washington
Hopkins, F. C.	Edgewood Arsenal, Maryland
Howes, David R.	Strategy & Tactics Analysis, Bethesda, Md.
Huk, K. G.	Picatinny Arsenal, Dover, New Jersey
Inselmann, Edmund	Frankford Arsenal, Pennsylvania
Isaac, G. J.	Fitzsimons General Hospital, Denver, Col.
Jamison, J.	Edgewood Arsenal, Maryland
Jebe, Emil H.	University of Michigan, Ann Arbor, Mich.
Jenkins, Andrew	Redstone Arsenal, Alabama
Jessup, Gordon L., Jr.	National Center for Radiological Health, Rockville, Maryland
Johnson, J. R.	Aberdeen Proving Ground, Maryland
Jones, Marian W.	Fort Detrick, Frederick, Maryland
Jones, Van R.	Aberdeen Proving Ground, Maryland
Kelley, Harold W.	USAPSA, Joliet, Illinois
Kingsley, Edward	George Washington University, Alexandria, Virginia
Kleinkauf, H.	Hercules, Inc., Cumberland, Maryland
Klotz, Jerome H.	University of Wisconsin, Madison, Wisc.
Kniss, J. R.	Aberdeen Proving Ground, Maryland
Kone, D. M.	Edgewood Arsenal, Maryland
Kruskal, William	University of Chicago, Illinois
Ku, H. H.	Natl. Bureau of Standards, Washington
Kullback, Solomon	George Washington University, Washington, D. C.
Kurkjian, B.	USAMC, Washington, D. C.
Kysor, K.	Edgewood Arsenal, Maryland
Lavin, George I.	Aberdeen Proving Ground, Maryland
Lehn, L. L.	Rock Island Arsenal, Illinois

Lieblein, J.	Natl. Bureau of Standards, Washington, D.C.
Liuzza, C. J.	Picatinny Arsenal, Dover, New Jersey
Lockhart, Lynn	Edgewood Arsenal, Maryland
Louis, Roberta	Edgewood Arsenal, Maryland
Lowrimore, G. R.	Hercules, Inc., Radford, Virginia
McGauley, R. M.	Aberdeen Proving Ground, Maryland
McIntosh, W. B.	Aberdeen Proving Ground, Maryland
McLaughlin, Georges J.	Canadian Armanent Research and Development Establishment, Quebec City, Canada
Mandelson, J.	Edgewood Arsenal, Maryland
Mann, Nancy R.	Rocketdyne, Canago Park, California
Martin, Stanley M.	Walter Reed Army Institute of Research, Washington, D. C.
Mayberry, John P.	MATHEMATICA, Bethesda, Maryland
Mizuki, M.	Federal Electric Co./ITT, Vandenberg AFB. California
Mowchan, W. L.	Aberdeen Proving Ground, Maryland
Mydosh, J.	Picatinny Arsenal, Dover, New Jersey
Nedelka, R.	Frankford Arsenal, Pennsylvania
Nickens, P. G.	Ft. Belvoir, Virginia
Ocko, A.	Edgewood Arsenal, Maryland
Ortynsky, Mr.	Edgewood Arsenal, Maryland
Parrish, Gene B.	Army Research Office, Durham, N. C.
Parzen, Emanuel	Stanford University, Stanford, Calif.
Pasini, H. C., Jr.	Aberdeen Proving Ground, Maryland
Pennsyle, R.	Edgewood Arsenal, Maryland
Pillersdorf, Arthur	Aberdeen Proving Ground, Maryland
Reckmeyer, V. H.	Thiokol Chemical Corporation, Redstone Arsenal, Alabama
Reddy, W.	Edgewood Arsenal, Maryland
Rednor, A.	Edgewood Arsenal, Maryland
Richards, J. C.	Edgewood Arsenal, Maryland
Rogers, F.	Edgewood Arsenal, Maryland
Rose, G.	Edgewood Arsenal, Maryland
Rose, Carol D.	USATCOM, Warren, Michigan
Rosenblatt, Harry M.	Bureau of the Census, Washington, D. C.
Rosenblatt, Joan R.	Natl. Bureau of Standards, Washington, D.C.
Ross, L.	Edgewood Arsenal, Maryland
Roylance, K. O.	Hercules, Inc., Magna, Utah
Sacco, William	Aberdeen Proving Ground, Maryland
Schiff, M.	Edgewood Arsenal, Maryland
Schnell, R. B.	USAAMCA, Washington, D. C.
Selig, S. M.	Office of Naval Research, Washington, D.C.
Shavitt, M. D.	Edgewood Arsenal, Maryland
Shulman, W.	Edgewood Arsenal, Maryland
Silverthorn, C.	Edgewood Arsenal, Maryland
Slook, T. H.	Frankford Arsenal, Pennsylvania
Smith, Kitty Lou	Watervliet Arsenal, New York

Solomon, Herbert	Stanford University, Stanford, Calif.
Speckman, Janace A.	Natl. Bureau of Standards, Washington
Stimson, R. G.	Weapons Systems Analysis Dir., Washington, D. C.
Sturdivan, Larry M.	Aberdeen Proving Ground, Maryland
Sulkes, M. J.	Fort Monmouth, New Jersey
Sutton, James	Hercules, Inc., Cumberland, Maryland
Tang, D. B.	St. Paul, Minnesota
Thomas, D.	Edgewood Arsenal, Maryland
Thomas, J.	Aberdeen Proving Ground, Maryland
Thomas, P. D.	Naval Res. Lab., Washington, D. C.
Travsky, P.	Edgewood Arsenal, Maryland
Trieger, S.	Edgewood Arsenal, Maryland
Tyler, J.	Edgewood Arsenal, Maryland
Varner, Ruth N.	Natl. Bureau of Standards, Washington
vonKaenel, P. A.	Aberdeen Proving Ground, Maryland
Wampler, Roy H.	Natl. Bureau of Standards, Washington
Waring, J.	Edgewood Arsenal, Maryland
Weinstein, Joe	Fort Monmouth, New Jersey
Wells, Lanny D.	Rock Island Arsenal, Illinois
Wenger, W. S.	Aberdeen Proving Ground, Maryland
West, L. L.	Fort Lee, Virginia
Whitmire, R. E.	Edgewood Arsenal, Maryland
Willis, N.	Edgewood Arsenal, Maryland
Willoughby, W. F.	Aberdeen Proving Ground, Maryland
Winslow, M.	Aberdeen Proving Ground, Maryland
Wood, W. O.	Edgewood Arsenal, Maryland
Zobel, S. P.	Cornell Aeronautical Lab., New York

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)*
U. S. Army Research Office-Durham
Box CM, Duke Station
Durham, North Carolina 27706

2a. REPORT SECURITY CLASSIFICATION
Unclassified

2b. GROUP
N/A

3. REPORT TITLE
PROCEEDINGS OF THE FOURTEENTH 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
September, 1969	529	

8a. CONTRACT OR GRANT NO.	9a. ORIGINATOR'S REPORT NUMBER(S)
b. PROJECT NO.	AROD Reprot 69-2
c.	9b. OTHER REPORT NO(S) *(Any other numbers that may be assigned this report)*
d.	

10. DISTRIBUTION STATEMENT
This document is subject to special export controls and each transmittal to foreign governments or foreign nationals may be made only with prior approval of the U. S. Army Research Office - Durham.

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 Fourteenth Conference on the Design of Experiments in Army Research, Development, and Testing. It contains 26 papers presented at that meeting. These treat various Army statistical and design problems.

14. KEY WORDS:

reliability
pattern classification
cutting fluid performance
vehicular mobility
confidence limits
multiple regression
queueing sequences
stochastic processes
control charts
analogue computers
wound data analysis
multi-dimensional contingency tables
optimization problems
laboratory control investigations
design of experiments
sampling error analysis
hypothesis testing
Monte Carlo investigation of robustness
approximating probability functions
binomial weighted transforms of discrete data
treatment response
bulk sampling

DD FORM 1 NOV 68 1473 REPLACES DD FORM 1473, 1 JAN 64, WHICH IS OBSOLETE FOR ARMY USE.

*3558-8-1-A
5-47
CC