Interface

Interface Conference on Applied Statistics

December 9 – 12, 2013

American Statistical Association, Alexandria, VA

Interface Foundation of North America, Inc.
P.O. Box 7460
Fairfax Station, VA 22039-7460

Schedule at a Glance

Start	**End**	**Monday - December 9**	**Start**	**End**	**Tuesday - December 10**
9:00	12:00	Tutorial	9:00	12:00	Tutorial
12:00	1:00	LUNCH	12:00	1:00	LUNCH
1:00	4:00	Tutorial	1:00	4:00	Tutorial
Start	**End**	**Wednesday - December 11**	**Start**	**End**	**Thursday - December 12**
9:00	10:00	General Session - Keynote Chair: Edward Wegman Speaker: Kirk Borne	9:00	10:00	General Session III Chair: Edward Wegman Speaker: Colin Wu
10:00	10:30	BREAK	10:00	10:30	BREAK
10:30	1200	Cont 1, Allan Mense (Chair) Speakers: Savitsky, Loh, Xing	10:30	12:00	Cont 3, Wendy Martinez (Chair) Speakers: Spalding, Samaniego, Mense Cont 4, Amanda King (Chair) Speakers: Zhang, Wei, Thomas
12:00	1:00	LUNCH	12:00	1:00	LUNCH
1:00	2:30	Special 1, Terrance Savitsky (Chair): Text Mining Cont 2, Fei Xing (Chair) Speakers: Zimmer, Park, Yang	1:00	2:30	Special 2, Alyson Wilson (Chair): ABMs Pinelis Cont 5, Frank Samaniego (Chair) Speakers: King, Cho, Eltinge
2:30	3:00	BREAK	2:30	3:00	BREAK
3:00	4:00	General Session II Chair: Allan Mense Speaker: Edward Wegman	3:00	4:00	General Session IV Chair: David Spalding Speaker: Bill Heavlin
5:00	7:00	Wilks Award Banquet A la Lucia Restaurant			

ABSTRACTS

Wednesday, December 11

General Session I: Edward J. Wegman (Chair)

Keynote: Big Data: Concepts and Missteps
Kirk Borne, George Mason University

Abstract: Data are plural, but Big Data is a concept. It is driving revolutionary changes in nearly all social, public, and institutional settings. It has democratized the data sciences, including statistics. Now everyone and every organization is "doing it". I will describe some of the most prevalent characteristics and concepts surrounding the current research and developments associated with Big Data. I will then describe the importance of data science as a research discipline and as an application methodology for making advances in Big Data analytics. Following the discussion of these concepts (steps toward discovery from Big Data), I will describe some of the statistical misconceptions (missteps on the road to discovery) that can afflict current Big Data efforts.

Contributed Session I: Allan Mense (Chair)

Collaborative Learning Under a Large-Scale Elicitation
Terrance Savitsky, Bureau of Labor Statistics

Abstract: Our application data are produced from a scalable, on-line expert elicitation process that incorporates hundreds of participating raters to score the importance of research goals for the prevention of suicide with the purpose to inform policy-making. We develop a Bayesian formulation for analysis of ordinal multi-rater data motivated by our application. Our model employs a non-parametric mixture distribution over rater-indexed parameters for a latent continuous response under a Poisson-Dirichlet process mixing measure that allows inference about distinct rater behavioral and learning typologies from realized clusters.

A Regression Tree Approach to Identifying Subgroups with Differential Treatment Effects
Wei-Yin Loh, University of Wisconsin, Madison

Abstract: In the search for new drugs to treat diseases such as cancer, it is very hard to find one that has a large beneficial effect for all subjects. The focus in recent years is to identify subgroups of the subject population where a drug has a substantial above-average effect. Because a regression tree model naturally partitions the data into subgroups, it is an ideal solution to this problem. We review several previous attempts and compare them with some new ones. Our methods are applicable to treatment variables with two or more levels, ordinal and categorical predictor variables with and without missing values, and ordinal response variables with or without censoring. They extend the GUIDE algorithm by employing three key ideas: (i) use of the treatment variable as a linear predictor in the node models, (ii) analysis of node residuals to ensure unbiased variable selection, and (iii) for censored responses, fitting proportional hazards tree models via Poisson regression. Each solution also yields importance scores for identifying influential variables that may not appear in the tree structures. Real and simulated data are used to evaluate the methods.

Seeking Better Supercomputer Experience for Kraken Users -- Evidence Based Assessment Using Statistical Tools

Fei Xing, Statistician, Mathematica Policy Research Inc.; Haihang You, Computational Scientist, National Institute for Computational Sciences, ORNL; Yang Shen, Corporate Credit Risk Analytics Project Manager, BB&T Bank.

Abstract: Supercomputer Kraken, located in the Oak Ridge National Laboratory, is a PetaFLOP scale supercomputer which used to be the most powerful computer in the world managed by academia and currently ranks No. 30th fastest supercomputer all over the world. Each year, Kraken operates large amount of computational intensive scientific projects from all over the US. From a user's perspective, choosing suitable parameter values to shorten the queuing time is of great practical importance. Unfortunately, most users make parameter choices based on word of mouth or just pure guess without guide, which lead to unnecessary long time waiting.

In this paper, we study the workload data of Kraken from January 2011 to June 2013 with over 1,800,000 user records and discover the more efficient parameter combination rule for a user to achieve shorter queuing time. Moreover, by applying Bayesian frame work to predict the temporal trend of long time waiting percentage, we are able to provide Kraken users a monthly updated reference chart for choosing optimal parameters in order to avoid long time waiting in queue. After testing the history data, the model could make over 85% of the correct predictions for all the cases we have tested.

Special Session I: Text Mining
Chair: Terrance Savitsky

Inferential Variability in Latent Dirichlet Allocation

Ming Sun, Amazon

Abstract: Topic modeling works to discover latent topics in a collection of documents. Latent Dirichlet Allocation (LDA) is a widely used generative model, which considers documents as a mixture of topics, and each topic as a distribution on words. We investigate the inferential variability of LDA for two inference methods: variational Expectation-Maximization (variational EM) and Gibbs sampling. Our method quantifies the inferential variability of LDA based on adjusted rand index and multidimensional scaling, which results in a d- dimensional representation with points representing multiple LDA repetitions. We also demonstrate significant effects of this inferential variability inherent in LDA on the performance of a downstream inference task, namely vertex nomination.

IMPLICATION (Implied Causation): Locating and Extracting Causation Statements in Unstructured Text

Joseph A. Marr, Muffarah G. Jahangeer, Trevor S. Crawford, Angeline P. Chau, Hyun Soo Choi, Jennifer J. Jones, Aneela B. Wadan, and Edward J. Wegman, George Mason University

Abstract: Causation statements are knowledge carrying structures that connect sets of facts and thus enable reasoning about these facts. In unstructured text, causation statements occur in a variety of formats. One of these formats involves explicit signal phrases that identify the nearby presence of causation statement components (antecedents and consequents), and enable automated extraction of these components. Downstream processing of antecedents and consequents supports the automated assembly of customized knowledge bases and, ultimately, expert systems. Here we report initial results on the identification and characterization of text-based causation signals, and the use of these signals in

extraction algorithms that process news media reporting and published medical research. We also discuss performance evaluations for our algorithms.

Contributed II: Fei Xing (Chair)

Tolerance Limits Under Mixture Models

Zachary Zimmer, Army Test and Evaluation Center; Thomas Mathew, University of Maryland Baltimore County; DoHwan Park, University of Maryland Baltimore County

Abstract: The tolerance intervals are widely used in industrial applications since they provide information concerning on the entire population. So far, the main focus has been done for a single distribution in the tolerance interval context, but we investigate the upper tolerance limits under mixture models; two components mixtures such as normal/normal and log-normal/normal. We used bootstrap approach and delta method to compute upper confidence limit and convert it to tolerance limit for the mixture models. We use Pearson-type goodness-of-fit test with bootstrap maximum likelihood estimation, suggested by Lin and Ma (2013), to select the proper mixture models for the data. We compare our method to nonparametric tolerance limit approach to illustrate. The performance of the proposed approach is illustrated by simulation studies and the application to the peak cladding temperature (PCT) in computer simulated nuclear accidents.

Point-Wise and Simultaneous Tolerance Limits Under Logistic Regression

Zachary Zimmer, Army Test and Evaluation Center; Dr.Thomas Mathew, University of Maryland Baltimore County; Dr. DoHwan Park, University of Maryland Baltimore County

Abstract: Computation of point-wise and simultaneous tolerance limits is investigated under the logistic regression model for binary data. The data consist of n binary responses where the probability of a positive response depends on covariates via the logistic regression function. Upper tolerance limits are constructed for the number of positive responses in m future trials for fixed as well as varying levels of the covariates. The former provides point-wise upper tolerance limits, and the latter provides simultaneous upper tolerance limits. The upper tolerance limits are obtained from upper confidence limits for the probability of a positive response, modeled using the logistic function. To compute point-wise upper confidence limits for the logistic function, likelihood based asymptotic methods, small sample asymptotics, as well as bootstrap methods are investigated and numerically compared. To compute simultaneous upper tolerance limits, a bootstrap approach is investigated. The problems have been motivated by an application of interest to the U.S. Army, dealing with the testing of ballistic armor plates for protecting soldiers from projectiles and shrapnel, where the success probability depends on covariates such as the projectile velocity, size of the armor plate, etc. Such an application is used to illustrate the tolerance interval computations in the article.

Impact of Topcoding on the Utility of Consumer Expenditure Microdata

Daniel K. Yang and Daniell Toth, Bureau of Labor Statistics, Office of Survey Methods Research (BLS, OSMR), 2 Massachusetts Ave. N.E. Suite 1950, Washington, DC 20212

Abstract: The Consumer Expenditure (CE) Survey implements a statistical disclosure limitation method known as “topcoding” to conceal sensitive and identifiable information in the publicly released data as a way to protect the household’s confidentiality. This method requires that the high (low) end household annual income, for example, be replaced by the average of all high (low) end households’ annual income in the microdata for public users. Topcoding will have a numerical impact on microdata’s utility and data

quality, especially for analysis that is dependent on the tails of the distribution. In this study, we evaluate the numerical impact of topcoding on CE microdata utility focusing on the analysis of expenditure items that are: 1) highly correlated with income and highly topcoded; 2) highly correlated with income but not highly topcoded; 3) not highly correlated with income but highly topcoded. A data utility measure, related to confidence interval overlap of coefficient estimates, is applied to assess the impact on the utility of the topcoded microdata.

General Session II: Allan Mense (Chair)

Big Data: Technology and Analysis

Edward Wegman, School of Physics, Astronomy, and Computational Science and Department of Statistics, George Mason University

Abstract: On March 29, 2012, the Obama administration announced the Big Data Research and Development Initiative. A number of U.S. federal agencies including the National Science Foundation, the National Institutes of Health, the Department of Defense, the Department of Energy and the U.S. Geological Survey have committed substantial additional funds to Big Data projects. The White House press release described the goals of the Big Data Initiative: "to advance the state-of-the-art core technologies needed to collect, store, preserve, manage, analyze, and share huge quantities of data; harness these technologies to accelerate the pace of discovery in science and engineering; to strengthen our national security, and transform teaching and learning; and to expand the work force to needed to develop and use Big Data technologies."

It should be noted that the scale of what is considered Big Data has been increasing steadily. Kilobytes (10^3), megabytes (10^6), gigabytes (10^9), and terabytes (10^{12}) by now are familiar to any researcher using modern computer resources. The Earth Observing System of the Jet Propulsion Laboratory introduced serious consideration of petabytes (10^{15}). Data collection systems looming on the horizon such as the Large Synoptic Survey Telescope promise data on the scale of exabytes (10^{18}). It is conceivable that data collection methods in the future may generate data sets of the scale of zettabytes (10^{21}) and yottabytes (10^{24}). The issue with big data is that while computing power doubles every 18 months (Moore's Law) and I/O bandwidth increases about 10% every year, the amount of data doubles every year. It is clear that conventional distributed systems such as those employed by Google, Facebook, and JPL (distributed active archive centers) must be expanded to include such new technologies as hadoop and new analysis methods. In this lecture, I will focus on aspects of these Big Data issues.

Wilks Award Banquet

Location: A la Lucia Restaurant
Time: 5:00 – 7:00 PM
Walking Directions: Left out the front door of the ASA. Turn right on Madison and walk 3 blocks (under 2/10 mile). A la Lucia is diagonally across the intersection with N. Royal. It is a white, 1-story building.

Thursday, December 12

General Session III: Ed Wegman (Chair)

Estimation of Longitudinal Concomitant Intervention Effects with Shared Parameter Models

Colin O. Wu, Office of Biostatistics Research, National Heart, Lung and Blood Institute, National Institutes of Health, wuc@nhlbi.nih.gov

Abstract: We investigate a change-point approach for modeling and estimating the regression effects caused by a concomitant intervention in a longitudinal study. Since a concomitant intervention is often introduced when a patient's health status exhibits undesirable trends, statistical models without properly incorporating the intervention and its starting time may lead to biased estimates of the intervention effects. We propose a class of shared parameter change-point models to evaluate the pre- and post-intervention time trends of the response and develop a likelihood-based method for estimating the intervention effects and other parameters. Application and statistical properties of our method are demonstrated through a longitudinal clinical trial in depression and heart disease and a simulation study.

Contributed Session III: Wendy Martinez (Chair)

A Method to Estimate Fix Effectiveness for Known Failure Modes

Dr. David Spalding, Institute for Defense Analyses, Studies and Analyses Center

Abstract: For a developmental test program, Fix Effectiveness Factors (FEFs) quantify the effectiveness of system modifications ("fixes") that increase reliability. FEFs are ratios of post-fix to pre-fix failure rates for individual failure mechanisms ("failure modes"). Both individual FEFs and average FEFs are used to estimate reliability. Many current methods that estimate FEFs apply prospectively at the start of testing, typically using expert opinion in various ways, or retrospectively after extensive testing that provides pre- and post-fix data for statistical analysis. The method presented here applies in the middle of a test program, when the prospective success of a program is the basis for proceeding with further development. The method estimates the effectiveness of future fixes for failure modes that have occurred in past testing. Building on concepts already in use,e.g. at AMSAA, it defines criteria that relate to specific information in Test Incident Reports (TIRs), Failure Analysis and Corrective Action Reports (FACARS) and other sources, with the intent to improve the transparency, repeatability and soundness of FEF estimates. The criteria are based on general principles of scientific explanation and are arrayed in a matrix, with a dimension for ranking root cause robustness and another dimension for ranking probability of fix success. The matrix provides a qualitative scatter plot of the fixes that highlights trends in the processes that determine root causes and fixes. The matrix is also the basis for assigning quantitative fix effectiveness values to sample data whose distribution can support further analysis. For several sets of FEF data, bootstrap analysis is used to determine confidence intervals for average FEF. Despite small sample sizes, confidence intervals based on sample mean and standard deviation agree well with the intervals from bootstrap analysis.

Estimating Component Reliability based on Failure Time Data from a System of Unknown Design*
Francisco J. Samaniego, University of California, Davis

Abstract: Suppose that N identical systems are tested until failure and that each system is based on n components whose lifetimes are i.i.d. with common continuous distribution function F(t) and survival function $\bar{F}(t)$ = 1 − F(t). Under the assumption that the system design is known, Bhattacharya and Samaniego (2010) obtain the nonparametric maximum likelihood estimate of F based on the observed system failure times, and they characterized its asymptotic behavior. Their estimator has the form $\hat{\bar{F}}_1(t) = h^{-1}(\hat{\bar{F}}_T(t)),$ where h(·) is the system's reliability polynomial and $\hat{\bar{F}}_T(t)$ is the empirical survival function of the system lifetimes {T_1,…, T_N}. Here, we treat this estimation problem when the system design is unknown. An unknown design may arise, for example, in military operations, where it is not uncommon to capture or gain control of a collection of like systems whose precise design is unknown. To estimate the component reliability function $\bar{F}(t),$ the system's design must be estimated from data. We assume that auxiliary data in the form of a variable K, defined as the number of failed components at the time of system failure, is available along with the system's lifetime. The data (T_1, K_1),…, (T_N, K_N) permits the estimation of the reliability polynomial h. Denoting the estimated polynomial as $\hat{h},$ we study the properties of the estimator $\hat{\bar{F}}_2(t) = \hat{h}^{-1}(\hat{\bar{F}}_T(t)).$ Our main results include (1) $\hat{\bar{F}}_2(t)$ is a consistent and asymptotically normal estimator of the component reliability function $\bar{F}(t)$ and (2) the asymptotic variance of $\hat{\bar{F}}_2(t)$, based on the augmented data {(T_i, K_i)}, is shown to be uniformly less than or equal to the asymptotic variance of $\hat{\bar{F}}_1(t)$, based on the data {T_i} and the assumption that h is known. The latter result is somewhat surprising! An elementary example of this type of phenomenon provides some helpful intuition.

*This work is joint with Peter Hall and Yin Jin.

Comparison of Classical and Bayesian Storage Reliability Models
Allan T. Mense, Ph.D., PE, CRE, Principal Engineering Fellow, Raytheon Missile Systems, Tucson, AZ;
Grant Olsen, M.S., Engineer, Raytheon Missile Systems, Tucson, AZ

Abstract: One of the most important reliability issues in product support and warranty claims is the need for a predictive model to estimate how many units of any given product will fail when in storage. –Calculating the number of failures after a period of storage is the challenge and very dependent upon available data. Analyses also depend upon availability of evidence of clear failure (degradation) mechanisms and understanding of the storage environment. Use of Binary Linear Regression models using classical and Bayesian analyses will be discussed and examples given.

Contributed Session IV: Amanda King (Chair)

High Dimension Data Analysis with Application on Schizophrenia
Hongda Zhang, George Washington University; Yuanzhang Li, Walter Reed Army Institute of Research

Abstract: When using biomarkers in the analysis, regression of high dimensional data is particularly difficult, especially if the number of observations is limited. The dependency among the variety of

biomarkers can't be avoided. The collinearity generates difficulties to make unbiased conclusion. The goal of this study is to find a particular partition of the space X, such that each subspace consists of all independent factors. Results: We propose an approach, which consists of three steps: 1. decomposing the sample space; 2. finding the most significant hyper plane consisting of biomarkers, which has the most significant effect on outcome; 3. using general linear regression based on the vectors generated in step 2, to evaluate the multiple biomarker effect and eliminate the biomarkers with weakest effect. A simulation shows our approach is stable and efficient. We also use this approach to identify the biomarkers, which have stronger effect on schizophrenia.

An Investigation of Quantile Function Estimators Relative to Quantile Confidence Interval Coverage

Lai Wei, Digital System, Inc., Chevy Chase, MD and State University of New York at Buffalo,Department of Biostatistics, Buffalo, NY; Dongliang Wang, State University of New York Upstate Medical University, Department of Public Health and Preventive Medicine, Syracuse, NY; Alan D. Hutson, State University of New York at Buffalo,Department of Biostatistics, Buffalo, NY.

Abstract: In this article, we investigate the limitations of traditional quantile function estimators and introduce a new class of quantile function estimators, namely, the semi-parametric tail-extrapolated quantile estimators, which has excellent performance for estimating the extreme tails with finite sample sizes. The smoothed bootstrap and direct density estimation via the characteristic function methods are developed for the estimation of confidence intervals. Through a comprehensive simulation study to compare the confidence interval estimations of various quantile estimators, we discuss the preferred quantile estimator in conjunction with the confidence interval estimation method to use under different circumstances. Data examples are given to illustrate the superiority of the semi-parametric tail-extrapolated quantile estimators. The new class of quantile estimators is obtained by slightly modification of traditional quantile estimators, and therefore, should be specifically appealing to researchers in estimating the extreme tails.

Comparison of Taguchi L9 and JMP 9 run Definitive Design

William C. Thomas MS., Raytheon Co., Tucson, AZ

Abstract: This paper will discuss the attributes of the JMP 9 run 4 factor 3 level design with a traditional Taguchi L9 4 factor 3 level design. The emphasis will be on the orthogonality of the 2 designs when you look at the quadratic effects and the 2 way interactions. I will discuss the methodology and the interesting results that will provide Engineers with the most effective way to estimate the effects.

Special Session II (ABMs): Alyson Wilson (Chair)

Projecting Inventories for Navy Community Managers using IMPACT-AC

Yevgeniya (Jane) Pinelis (CNA), Warren Sutton (CNA), and CDR Arjay J. Nelson (BUPERS-34B)

Abstract: The Office of Naval Research (ONR) is sponsoring the research and development of a new Navy Manpower, Personnel, Training, and Education (MPT&E) modeling tool designed for Bureau of Naval Personnel (BUPERS-3 - Navy Community Management). The agent-based modeling system, called Integrated Manpower Agent-Based Computer Tool – Active Component (IMPACT-AC), can be used to model and project sailor and officer inventories within their respective communities. Using algorithms built based on the Navy's rules, IMPACT-AC uses user-supplied inputs of gains, losses, advancements, and promotions to forecast the overall behavior of the Navy population by simulating individual agent

responses to various decision scenarios. The software's predictive capabilities are enhanced by accurate representation of the Navy's processes as well as usage of various statistical and econometric analyses to form the algorithms. The ultimate purpose of the tool is decision support to Community Manager end users who will be able to create, modify, and examine the effects of MPT&E policies on Navy communities. Ultimately, IMPACT-AC will provide the ability for Navy Community Managers to better integrate broad MPT&E policy goals and community-level planning, both in execution year monitoring and strategic long-term forecasting.

Contributed Session 5: Frank Samaniego (Chair)

An Evaluation of Joint Models Using Different Feature Extraction Metrics for Structural Health Monitoring (SHM) of Aircraft.

Amanda Sue King*, Christine M. Schubert Kabban*, Cheryl P. Edelmann *, Mark Derriso, Air Force Institute of Technology*, AFRL/RQ

Abstract: Feature extraction in SHM can be performed using a variety of different methods. This paper presents a comparison of several different features (Damage Indices (DI)) and their ability to accurately represent data which will jointly model crack length in both the horizontal and vertical directions using a Generalized Linear Mixed Model (GLMM). A comparison of different GLMM models is performed; models which combine different DIs into the same model will be compared with other models which use only one DI. A discussion of the importance of experimental design and feature selection prior to the execution of the experiment will be included. An emphasis will be placed on the importance of proper statistical technique when collecting data for analysis.

Identifying Common Groups for Variance Function Models in Complex Sample Design

MoonJung Cho, John L. Eltinge, Julie Gershunskaya and Larry Huff

Abstract: Due to relatively high levels of sampling variability in standard direct design-based variance estimators, analyst often seeks to develop a model that is relatively parsimonious and that produces variance estimators that are approximately unbiased and relatively stable. This development and validation work often begin with regression of initial variance estimators (computed through standard design-based methods) on one or more candidate explanatory variables based on the assumption that the coefficient vector is constant. In this paper, we explore the possible heterogeneity of the coefficient and present some simple methods of identifying groups within which we can fit a shared model. Some of the proposed diagnostics are applied to data from the U.S. Current Employment Statistics survey.

Assessment of Quality, Cost and Risk Factors in Statistical Work With Administrative Record Data

John Eltinge, U.S. Bureau of Labor Statistics Eltinge.John@bls.gov

Abstract: In recent years, large-scale statistical organizations have expressed increased interest in the use of administrative records to supplement standard sample survey data. Practical decisions on use of administrative data involve a complex set of factors that affect the balance of quality, cost and risk in thestatistical production process. Many of these factors are qualitative in nature, e.g., ones that involve legal, regulatory, contractual or operational constraints on use of administrative data. Other factors are quantitative, e.g., measures of missing-data rates, misclassification rates, or cost components. This paper reviews and synthesizes previous literature on assessment of the abovementioned quality, cost and risk factors. Seven issues receive special attention: (1) Conditions under which primary interest may center on, respectively, qualitative or quantitative assessment of these factors. (2) Inferential goals for the proposed use of administrative data; and linkage of these goals with the information needs of primary stakeholders. (3) Administrative-record extensions of concepts and methods developed primarily within the context of sample survey methodology. These include “total survey error” models; broader assessments of survey data quality; integration of multiple data sources; and adjustments for incomplete data, reporting errors, aggregation effects and definitional effects. (4) Measurement and modeling of fixed and variable components of cost; and amortization of some cost components across time and across multiple production systems. (5) Distinctions between aggregate risks (associated with the cumulative effects of a large number of independent events) and systemic risks (associated with a small number of high-impact events). (6) The degree of control that the statistical organization exercises over specified dimensions of quality, cost and risk. (7) The extent to which issues (1) through (6) may be affected by the architecture of a statistical production system.

General Session IV: David Spalding (Chair)

Ranking Performance Videos

William D Heavlin (bheavlin@google.com), Google, Inc.

Abstract: As American Idol attests, where one performance simply entertains, two performances implicitly compete. In 2011-2012 YouTube presented a few series of pairs of video performances (“slams”), and invited its viewers to vote for their favorite. The first-order goal was to determine better videos: the resulting rankings guided YouTube viewing suggestions. A secondary goal was to entertain the viewers voting: greater participation both increases data volume and enhances the YouTube brand.

Here we consider three statistical issues: (1) how to pair and play the videos, the *tournament design problem;* (2) how to rank the videos, the (*primal*) *estimation problem;* and (3) how to rank the acuity of the viewers, a (*dual*) *estimation problem.* Our approach likewise has three components: (a) a linearized objective function to schedule performance pairs, (b) a Bradley-Terry model to estimate performance quality; and (c) Fisher scores to assess voter acuity.