

# The Fairest of Them All: Using Variations of Beta-Binomial Distributions to Investigate Robust Scoring Methods

Mary Good - Bluffton University  
Christopher Kinson - Albany State University  
Karen Nielsen - University of Oklahoma  
Malin Rapp-Olsson - University of Arizona  
Mame Fatou Thiam - University of Wisconsin River Falls

## Abstract

Contestants in subjective competitions often encounter the issue of unfair or inconsistent judging. Scores are awarded to contestants and winners are determined, but how do we know that the best contestants win? By assuming there exists a well-defined order of the contestants' abilities, we can explore which scoring method best captures the true order of the contestants. We use variations of beta-binomial distributions to model the contestants' abilities and the judges' scoring tendencies. We use Monte Carlo simulations to investigate seven scoring methods (mean, z-scores, median, raw rank, z rank, Olympic, and outliers) and determine which method, if any, works best for various judge panels. We apply our model to the scenario of a scholarship competition with 20 contestants, where the top three contestants receive equal monetary awards.

# The Traveling Salesman at a Home Near You: Investigating the Traveling Salesman Problem with Varying Home Points

Natalie Collar  
Ashlee Edwards  
Jaime Jeke  
Shana Johnson  
Kayse Lee  
Cathalee Soergel

Manchester College  
Virginia Wesleyan College  
University of Pittsburgh  
Lynchburg College  
Bethel University  
Miami University

Advisor: Dr. Lakeshia Legette  
Graduate Assistant: Brett Jefferson

## Abstract

Have you ever wondered how MapQuest<sup>®</sup> works? It has the capability of finding the shortest route from one location to another. Suppose you want to visit  $n$  cities. In which order would you visit these cities to travel the shortest distance? We explore the Traveling Salesman Problem with both three and six cities on a two-dimensional plane. In both cases, we consider a varying Home point,  $\mathbf{H}$ . It turns out that the plane is divided into distinct Home point regions such that for each region there is a unique minimal circuit to travel. We find that for three cities, some regions become disconnected, or *split*. This split suggests that it is not always optimal to travel the most intuitive circuit. We find all the city arrangements where this phenomenon occurs. Furthermore, we find that a particular arrangement of six cities guarantees only six distinct minimal circuits of the 720 possible circuits.

### **Abstract #3**

Eloise Hilarides

"Using Weighting Schemes to Account for Coverage Bias in Internet Surveys"

Advisor: Michael Sheard (msheard@stlawu.edu)

Over the past decade, Internet surveys have become a popular method for collecting data about the general population. In 2005, the Harris Poll published findings which claimed that 74% of the United States Population had access to the Internet somewhere. While this number has steadily risen over recent years, bias still may be introduced if the population without Internet access is different from the Internet population in regards to the variables of interest. In this research we studied whether Internet users that only have access to the Internet outside their home can be useful in reducing bias by assuming that they are more similar to those without Internet access than the Internet population as a whole. We outline several weighting adjustment schemes aimed at reducing coverage bias. Data for this study was taken from the Computer and Internet Use Supplement of October 2003 administered by the Current Population Survey. We evaluate the schemes based on overall accuracy by considering the reduction in bias for ten variables of interest and the variability of estimates from the schemes. We find that several of the proposed schemes are successful in improving accuracy.

## **Abstract #4**

Jamie Wolff

"Performance vs. Pick: A Study of the NBA Draft"

Advisor: Michael Schuckers (schuckers@stlawu.edu)

Millions of dollars are invested in the top draft picks of the National Basketball Association (NBA). A significant amount of deliberation and analysis is put into determining which athlete to select. Often teams make trades in order to better their position in the draft, or they "trade down" meaning trade away an early draft pick for more draft picks later on. By giving another team money, current players, draft picks, a team is hoping this lower number pick will be more productive than their higher number. This investigation will explore any significant differences among draft picks and what would be the advantage, if any, in drafting at one pick over another. We will use NBA career statistics (points per game, minutes per game, games played, all-star appearances) to assess draft decisions based on player productivity. Data from the 1994 through the 2007 seasons was compiled from Basketball-Reference.com. This study will divide the draft picks into eight zones (four in the first round and four in the second) and compare zones to find any significant differences. Results suggest that lottery picks are valuable and trading up or down may not be productive.

## Abstract #5

Title:	<i>A <math>p</math>-adic Euclidean Algorithm</i>
Author:	Cortney Lager , Winona State University
Abstract:	The rational numbers can be completed with respect to the standard absolute value and this produces the real numbers. However, there are other absolute values on the rationals besides the standard one. Completing the rationals with respect to one of these produces the $p$ -adic numbers. In this paper, we take some basic number theory concepts and apply them to rational $p$ -adic numbers. Using these concepts, a $p$ -adic division algorithm is developed along with a $p$ -adic Euclidean Algorithm. These algorithms produce a generalized greatest common divisor in the $p$ -adics along with a $p$ -adic simple continued fraction.

## **Investigating the Validity of the Survey of Attitudes Toward Statistics**

Winona State University  
Department of Mathematics and Statistics

*The Survey of Attitudes Toward Statistics (SATS) is a tool widely used by statistics educators to help gain insight into students' attitudes and how they impact teaching and learning in introductory statistics courses. Three instructors at Winona State University have been administering this survey to students both at the beginning and end of several semesters since 2011. This study involved an analysis of the data collected in these courses to investigate students' attitudes towards statistics and how they change throughout the semester. The results were also compared to national norms. Finally, an exploratory factor analysis was conducted using the data collected from Winona State University courses to investigate the construct validity of the SATS tool. It was found that this national survey might not actually be measuring what the creators of the survey intended.*

## Abstract #7

# Text Mining and Data Information Analysis for Network Public Opinion

**Author:** Yan Hu 

## Abstract

Network public opinion information is massive and complex, and it is difficult to make effective use of manual means. In this paper, a method based on pattern matching and machine learning (PMML) was proposed to analyze the emotional tendencies of network public opinion. Firstly, the key words in public opinion were extracted, then the patterns were extracted and matched, and the emotional tendencies of words were calculated to obtain the pattern sequence vectors. Support vector machine (SVM) classifier was used to classify emotional tendencies. The Internet reviews of Meituan hotel were taken as the experimental subject. PMML method was found to have a high classification performance, with a maximum accuracy of 86.75%. It suggested the effectiveness of the proposed method. Then PMML method was used to classify the emotional tendencies of the collected reviews, and the results showed that the negative emotional tendency was greater than the positive tendency, which showed the inadequacy of Meituan hotel. The experiments in this paper provide some basis for the application of PMML in sentiment analysis of Internet public opinion.

## Abstract #8

# The Challenges of Data Quality and Data Quality Assessment in the Big Data Era

**Authors:** [Li Cai](#) , [Yangyong Zhu](#)

## Abstract

High-quality data are the precondition for analyzing and using big data and for guaranteeing the value of the data. Currently, comprehensive analysis and research of quality standards and quality assessment methods for big data are lacking. First, this paper summarizes reviews of data quality research. Second, this paper analyzes the data characteristics of the big data environment, presents quality challenges faced by big data, and formulates a hierarchical data quality framework from the perspective of data users. This framework consists of big data quality dimensions, quality characteristics, and quality indexes. Finally, on the basis of this framework, this paper constructs a dynamic assessment process for data quality. This process has good expansibility and adaptability and can meet the needs of big data quality assessment. The research results enrich the theoretical scope of big data and lay a solid foundation for the future by establishing an assessment model and studying evaluation algorithms.