**Proceedings of the Conference on
Promoting Undergraduate Research in Mathematics**

The Rice University Summer Institute of Statistics (RUSIS)

Javier Rojo

The Rice University Summer Institute of Statistics (RUSIS) has been generously supported by The National Science Foundation (NSF) and The National Security Agency (NSA) for five years. A total of 17 undergraduate students per summer spend 10 weeks at Rice University. Twelve students are supported by NSF and five students are supported by NSA.

The main goal of RUSIS is to encourage undergraduate students to pursue Ph. D. work in the statistical sciences. Selected junior and senior underrepresented minority students and students with no easy access to a career experience at their institutions are recruited. Requirements for admission to the program include: three semesters of calculus, one semester of linear algebra or matrix analysis, a minimum 3.0 GPA, and at least two letters of recommendation. However, students with a GPA below the 3.0 threshold, or students who have not met the mathematical requirements, are not automatically rejected. Rather, when the letters of reference and statement of purpose give an indication that the student is capable of pursuing graduate work, telephone interviews are conducted to better ascertain the students potential. Whenever feasible (e.g., local students) applicants are invited to Rice for an interview. RUSIS believes that grades constitute but one indicator of the potential for a successful career in graduate school. Other important dimensions that impact success in graduate school derive from personal qualities such as perseverance and creativity. The program also enrolls the help of the advisory board in the search for qualified candidates.

Through intensive short courses in areas of current research interest, intensive seminars in computation, close mentoring and supervision by faculty, students develop skills that help them during their graduate careers. In addition, attendance to a series of lectures by top scientists keep their interest and focus on science and graduate school. By the end of the summer students have participated in at least one research project analyzing data, running computer simulations, developing algorithms, and, when appropriate, engaging in theoretical work. Every student is encouraged and expected to prepare their research findings in close collaboration with their faculty mentors and peers for presentation at student sessions at national

Received by the editor December 18, 2006.

meetings and subsequent submission for publication.

At the end of the summer, all students meet with the advisory committee, in the absence of any personnel associated with the RUSIS, to provide feedback on ways to improve the program.

1. Institute Activities. The first day of the Institute is reserved for introductions of students, mentors, and support staff; tours of the campus and Library, and applying for Library cards. Students register, at no charge, as Rice students and earn one hour of credit for Independent Study. Students fill out questionnaires related to their backgrounds, and expectations for the program. A pre-test and a post-test are given, and the information is utilized as part of a set of indicators to measure the impact of the program.

As most students come without the needed background in probability and statistics, an intensive course in probability, stochastic processes, statistical inference, and survival analysis, is taught for the first three weeks of the summer. In addition, an afternoon short course in computation is taught during the first three to four weeks of the institute. Beginning in the fourth week students start working on their research projects. Mentors, with the support of postdoctoral and graduate students, work closely with students to provide them the background material specific to their group projects and the needed research direction as the projects evolve.

Throughout the summer, invited speakers from MD Anderson Cancer Center, the Michael E. DeBakey Department of Surgery at the Baylor College of Medicine, the University of Texas Health Sciences Center at Houston, lecture on survival analysis applications in cancer research, liver transplant, and other health-related and environmental applications. In addition, Sallie Keller-McNulty, Dean of the School of Engineering at Rice, and President of the American Statistical Association, visits with the students. For example, during the summer of 2005, Dean Keller-McNulty discussed opportunities for graduate work in Statistics and the job market. Students commented in the exit questionnaires that they felt that RUSIS really cared about them as a result of the visits by the Dean and the amiable conversations with the advisory committee members. Visits by NSA have been motivational and elicit very positive reactions by the students.

Friday afternoons are dedicated to meet as a group to discuss improvements of the program and, more importantly, discuss progress of the research projects as well as requirements and strategies for graduate school. Statistical/mathematical videos (Fermat's last theorem, If Copernicus had a computer, MSRI on-line lectures) are presented to spark the interest of students in pursuing graduate work.

At least two field trips are scheduled during the summer. Students spend a day at NASA facilities where the group is allowed to operate the flight simulators and tour the facilities including Mission Control. David Mains and Daniel Adamo have provided excellent educational tours. Students also visit The MD Anderson Cancer Center. Gary Rosner organizes a session where various researchers present their work on cutting-edge cancer biostatistical research.

In addition, a consultant from the Cain project www.owlnet.rice.edu/∼cainproj helps the students improve their writing skills. As part of their final projects, students write letters to their Congresspersons describing their summer experience and encouraging them to continue supporting the federal funding for these activities.

2. Teaching and computational facilities. All teaching, computational, and mentoring activities take place at Rice University. RUSIS focuses explicitly on Statistics and its applications. Rice University is committed to facilitate the development of nontraditional groups in undergraduate science and engineering programs. Several well-equipped classrooms exist in Duncan Hall, the home of the School of Engineering and the Statistics Department.

Classrooms are equipped with the latest technology: A 32-Button Wireless Remote Control, One Video Projector (1024 x 768 native, 1280 x 1024 max.), one VCR (VHS / S-VHS) with Cable TV; PowerMac G4 w/ MacOS 10.4 (with floppy, CD / DVD); Pentium w/ Windows 2000 (Floppy, CD/DVD); Laptop connection with video, audio, network (DHCP); and a Microphone with Distributed speaker system.

The Computer Laboratoy has room for 22 students. The laboratory is equipped with one VCR (VHS), a DVD Player, a Document Camera, a video projector, and a Distributed speaker system. Instructors work from two Compaq Pentium PC with Windows 2000, and the students have access to twenty-two active network ports (DHCP) and twenty Compaq Pentium PC with windows 2000. A wireless network is available, and students are provided access to all the library electronic holdings in addition to their Rice students library privileges. Laser printing facilities are available in the room.

The Statistics Department has access to the Rice Virtual Laboratory in Statistics (http://www.ruf.rice.edu/~lane/rvls.html) to support teaching. The site contains: HyperStat Online – An online statistics book with links to other statistics resources; Simulations/Demonstrations: Java applets that demonstrate various statistical concepts; Case Studies: Examples of real data with analyses and interpretation; Analysis Lab: Some basic statistical analysis tools. RUSIS students use Mathematica, R, SAS, Gauss, and Matlab to support their research projects. The School of Engineering provides network, hardware, and software support through the Information Technology Department.

3. Examples of Projects. It is understood that not all students come with the same background and training. Our goal is to provide them with a valuable experience that is helpful during their graduate school career. Therefore, several projects are presented in class through the first 3 or 4 weeks and students are allowed to choose their project(s). The current need to understand, develop, and assess the merit of new methodologies in the areas of multivariate survival analysis, multivariate extreme value theory, analysis of microarray data, and analysis of massive data sets presents a great opportunity for the training of selected undergraduate students. Problems of current interest are used to motivate the students and serve as a point of departure for the research projects. Some examples of projects follow.

Extreme Value Theory: Extreme events have large impact on various areas of engineering, science and economics. Events such as extreme waves, rainfall, and floods are of fundamental importance, as are high wind speeds and extreme temperatures, and extreme value theory provides a foundation to study corrosion and metal fatigue. Health hazards develop as a result of high concentrations of pollutants, and damages to the economy develop from extreme changes in the market. The development of new statistical and probabilistic methods for extremes is an active area of research and one that has potential to impact risk management, pollution and weather forecasting. Bivariate extreme value distributions can always be transformed so that their marginal distributions are exponentially distributed. Once this is done, the bivariate distribution is characterized by the dependence function which must satisfy certain constraints. Rojo et al. (2001), proposed a nonparametric estimator for the dependence function. The students have the opportunity to work on various computational aspects of these problems. Ozone level data from various National Parks covering over a period of fifteen years is available and provides an interesting set of theoretical and applied challenges. Questions of interest include: Are extreme ozone levels, as measured by exceedances over the 12 ppm threshold level, decreasing in size and/or in frequency? Are there any significant differences among the parks? Can the impact of the Clean Air Act be observed in the ozone level trends? These issues are explored through a mixture of statistical computer modeling and computer graphics exploration.

Multiple comparisons: The renewed impetus in this area is due in part to the arrival of microarray data. The statistical interest in microarray experiments derives (See, e.g., http://www.sbm.temple.edu/ ~sanat/cbmsconf/lectures.html, and http://elib.zib.de/mailing-lists/public/st-net/2002/msg00045.html) from its much larger data sets and coincides with the arrival of computer technology which allows for computer-intensive re-sampling based methods. (e.g., Westfall, P. H. and Young, S. S. (1993)). Genomics and bioinformatics have spawned challenging problems. Modern biotechnology allows researchers to collect high-throughput genetic data which leads to thousands of significance tests. Advanced undergraduate students, in the context of differential gene expression, learn multiple comparison issues and methods relatively quickly, and they compare the various testing procedures that are in current practice using computer simulations. Substantive problems and (real) data is available through various publicly accessible websites.

Dimension Reduction: One major challenge in the analysis of survival microarray data derives from the number of replicates being very small, while the number of genes is usually very large (10,000-20,000). There are at least two procedures used for dimension reduction before fitting a survival analysis model. Principal Component Analysis (PCA), and Partial Least Squares (PLS). The latter has been heralded by their proponents as a clear winner over PCA. Students in the last two RUSIS have investigated the properties of both methods in terms of Prediction Mean Squared Error, and the conclusion has been that in the modest simulation studies performed by the students, there is no clear distinction between the two. Both tend to select the same genes, and both tend to have similar Mean Squared Error properties. This area continues to be explored.

Other areas of interest have included: multivariate survival analysis with censored data; ROC curves as a way to identify differentially expressed genes; Linkage analysis with sib-pair data; and inverse problems in geophysics.

4. Program assessment. An advisory committee provides an annual assessment of RUSIS. The purpose of the committee is two-fold: **(i)** Provide advice for evaluation and improvement of the program and, **(ii)** To serve as ambassadors of the program at their own institutions and help in the recruitment of students.

The following currently serve on the advisory board: Arturo Bronson, University of Texas at El Paso; Willie Pearson, Georgia Institute of Technology, William Velez, University of Arizona; Anna Baron, University of Colorado at Denver; Cristina Villalobos, University of Texas Pan American; Maria Acosta, Texas State University San Marcos. The advisory board convenes during the tenth week of the program to assess the merits of the program and the progress of the students, and to provide insights and feedback on ways to improve the Institute. In addition, board members provide an invited lecture on a statistical aspect of their discipline. These lectures have received a lot of praise from the students who think that lectures are inspirational. These lectures also provide a different perspective on science and engineering and on opportunities for career development and employment.

Student Recruitment and selection Program information and application packets are mailed to mathematics and statistics departments throughout the country. A website in the Statistics Department provides detailed information on application deadlines and requirements for admission. World Wide Web sites that specialize in mathematics and statistics are utilized to post announcements about RUSIS. Similar electronic announcements are sent to most of the universities located in predominantly underrepresented minority areas in the United States. In addition, SACNAS conferences offer excellent opportunities for recruitment, and various committees of the American Statistical Association are contacted seeking their help in making the information available to undergraduate advisors.

Project evaluation and reporting The most important goal of the RUSIS is to motivate students to pursue graduate work in the Statistical sciences, through a series of activities designed to entice them and excite them to engage in research. As such, the real impact of the program is difficult to gauge, and the outcomes of interest will be observed only until after a few years have passed. The program keeps track of all students after their participation in RUSIS. Students are asked to provide permanent contact information, and faculty and staff reciprocate to encourage the students to stay in touch. During their visit, each member of the advisory committee presents a fifteen-minute talk that will highlight how statistics play a role in their discipline. In addition, the students present their research projects and interact with members of the board. These interactions are valuable for the board in formulating their recommendations for improvement of the program. These recommendations are communicated to the RUSIS Director during an exit interview, and a formal site visit report is written. Through student questionnaires and exit interviews, students provide feedback on various aspects of the program.

5. Summary of RUSIS accomplishments. The number of applicants has been stable at 51-55 per year. Most of the students so far have come from the southwestern and southern states, although several students from the Midwest and New England have participated. Roughly 56% of the students have graduated. (The program accepts sophomores-seniors). Of these, 85% have expressed a desire to pursue graduate work, and 75% are currently in Ph. D. programs or are applying for admission to Ph. D. programs. For example, four students are applying for admission to Rice University, while three students are already members of the statistics department at Rice University. Che Smith (Spelman) has begun her Ph. D. in biostatistics at the University of North Carolina, Chapel Hill. Sarah Williams is a Ph. D. student in environmetrics at Colorado State University. Tahira Saleem was accepted to the Ph. D. program at Rice but opted for a job with NSA. Daisy Wang is a Ph. D. statistics student at UC Berkeley. These are examples of the many success stories of former RUSIS participants. There is, however, an example of a student who has blossomed as a result of his RUSIS participation. Juan Gallegos, the first student in his family to attend college received a bachelors degree from the downtown campus of the University of Houston. Juan took advantage of the opportunities offered by RUSIS and presented his work at several student conferences. His self-confidence grew and he was accepted and offered a fellowship by the University of Texas at Houston to pursue Ph. D. work in Epidemiology.

Students present their work at national conferences, and many have participated in SACNAS, American Mathematical Society meetings, and the Young Mathematicians Conference. Students have won recognition for their work. Jean Kongpinda and Venessa Tavares received an award at the 2004 AMS meeting; Stacey Ackerman, Israel Cabello, Cyrus Aghili, and John Ratana won second place at the 2005 SACNAS meeting; and David Kahle and Darren Homrighausen won a poster award at the 2006 AMS/MAA meeting. Tollie Thigpen writes in a recent email: "I have had the opportunity of presenting it (RUSIS project) at conferences in Mississippi and North Carolina. I was awarded 2nd and 3rd places at two of the conferences in which I was unaware of being judged."

The RUSIS program has been featured in the Rice News University paper, (http://senews.rice.edu/hotnews.cfm?mode=details&status=Archived#1058), and in a short article that appeared in Math Horizons in September 2005. Student photos and abstracts, as well as the students' home institutions, may be found at the RUSIS webpage – http://www.stat.rice.edu/RUSIS03.

References

[1] Rojo, J., Villa, E., and Flores, M. (2001). Nonparametric Estimation of the Dependence Function in Bivariate Extreme Value Distributions, *Journal of Multivariate Analysis*, pp 159-191.

[2] Rojo, J. (2005). REU Spotlight: Rice University Summer Institute of Statistics, *Math Horizons*, Vol 13, pp 30.

[3] Westfall, P. H. and Young, S. S. (1993). *Resampling-Based Multiple Testing: Examples and Methods for P-value Adjustment*, Wiley, New York.

Statistics Department, Rice University, 6100 Main Street, Houston, TX 77005
E-mail address: jrojo@rice.edu