Planning the Statistics Education of Future Teachers
by Jerry Moreno

One of the projects that the Advisory Committee for Teacher Enhancement (ACTE) of the American Statistical Association has created is called TEAMS—Teachers Education: Awareness, Methods, and Strategies. TEAMS is co-chaired by statistics professors at the University of Georgia (UGA), Christine Franklin (chris@stat.uga.edu) and Robert Taylor, (bob@stat.uga.edu). This article is based on notes that Chris sent me.

ASA with further sponsorship and organizational support from the University of Georgia’s Department of Statistics, College of Arts and Sciences, and the Department of Mathematical Education, College of Education is developing the inaugural conference on statistics in teacher preparation. It will be held October 30 to November 2, 2003, at the University of Georgia. TEAMS is being planned by representatives from ASA, The Conference Board of the Mathematical Sciences (CBMS), the Mathematical Association of America (MAA), the National Council of Teachers of Mathematics (NCTM), and other leaders in statistics and mathematics education.

With the emphasis that is placed on probability and statistics in both the 1989 and 2000 versions of the NCTM Principles and Standards for School Mathematics, the National Assessment of Educational Progress (NAEP) framework which recommends more data analysis at the k-12 level, in addition to the explosive growth in the Advanced Placement Statistics course, incredible demands have been placed on many k-12 teachers. Many of these teachers have had little or no coursework in statistics during their own preparation to be teachers. While many k-12 textbooks now include statistical topics, and a number of organizations have provided in-service programs for present classroom teachers, there has been no organized effort to address pre-service teachers and those pursuing graduate degrees. With this training deficiency in mind, ASA and UGA are coordinating the effort to plan TEAMS.

The target audience for the conference is teacher educators (whether they are in schools of education, mathematics departments, statistics departments, or other departments). A primary goal is to create awareness among teacher educators of what is needed to better prepare k-12 teachers to give instruction in probability and statistics at the k-12 level. Another goal is to build a leadership team that can periodically meet to examine issues and provide guidance in methods to improve teacher preparation. Realizing that many pre-service teachers begin their undergraduate education at two-year colleges, there is also a need to bring an awareness to the instructors of probability and statistics courses at these types of institutions. The support of the American Mathematical Association of Two Year Colleges (AMATYC) is being enlisted to bring about this awareness.

Specific topics to be addressed at the 2003 TEAMS conference relative to what future teachers need to learn about statistics include the following:

Also In This Issue...

Enrichment for Advanced Placement Statistics Students.............................3
Teaching Matrices From A Data Point Of View .....................................4
• The subject matter of statistics, including the deeper meaning of statistical thinking,

• Connections between statistics and other subjects,

• The state of research on how students learn statistics, including similarities and differences with mathematics,

• Appropriate scope and sequence of statistics in k-12 using as a guideline the NCTM Standards and the NAEP framework,

• The advantage of stand alone courses versus integration with other mathematics (and perhaps the sciences and social sciences) courses,

• The pedagogy of statistics with an emphasis on active learning in which students explore statistical concepts through hands on activities and through the use of technology,

• The role of statistics for all teachers as reflective professionals (for example, interpretation of standardized tests and results of educational research).

The Mathematics Education of Teachers (MET) Report, released by the Conference Board of the Mathematical Sciences (CBMS) in fall 2001, outlines recommendations for how mathematical science departments can help educate future teachers of mathematics at the k-12 level. (See www.cbmsweb.org/MET_Document/index.htm for details.) The 2003 TEAMS conference aims to build connections between academic programs in statistics and teacher education for the purpose of fulfilling these recommendations in the areas of probability and statistics. That statistics is a major area of concern among the writers of the MET Report can be seen in quotes from the document itself:

“Statistics is the science of data, and the daily display of data by the media notwithstanding, most elementary teachers have little or no experience in this vitally important field.”

“Of all the mathematical topics now appearing in the middle grades curricula, teachers are least prepared to teach statistics and probability.”

The conference program will consist of:

• Plenary sessions presented by well known individuals who use statistics in their professions

• Presentations about current model programs

• Presentations about current research on how students learn

• Workshops that provide the participants a sampling of activities used in current model programs and that would allow interaction between the presenter and the team members.

In this inaugural conference, it is planned that participant selection will be by application and consequent invitation. The organizers of the 2003 TEAMS conference encourage teams of educators to apply. These teams ideally consist of a statistician, mathematician, math educator, science and/or social science educator, curriculum director or instructional leader from a local school district, and a statistics educator from a two or four year college that is a feeder school for the school of education or department providing the teacher preparation program. Individuals from industry and/or business who use statistics are encouraged to participate as part of the invited participants. It is hopeful that funds will be found to cover most of the expenses for the invited participants. The process for applying will be available for viewing in February at the ASA web site, www.amstat.org. Individuals can also email Chris at chris@stat.uga.edu with details of a possible team to bring to the conference.

Teams must express a strong interest in following through with the goals of the conference. The conference committee hopes to find funds that teams may apply for to establish programs at their home institutions which will promote the integration of statistical education in the curriculum for teacher preparation. The conference will draft guidelines for the statistics education of teachers and will formulate a plan to have these adopted by the lead organizations involved with teacher education and statistics education (ASA, MAA, NCTM, AMATYC). The guidelines will be disseminated to departments teaching statistics and to Colleges of Education by means of a network of educators. This network will encourage feedback and will build support mechanisms for implementing the guidelines. Support mechanisms will include designated groups working on materials review and development, pedagogical improvements, ties to business and industry, and the implementation of research findings related to how students learn statistics and probability. Educational researchers will be encouraged to have their findings realize practical implementation and to help classroom teachers become aware of the research findings. The teams receiving grant monies will be required to document the progress of their program based on the guidelines provided. A team may reapply for continuing grant monies based on a successful evaluation.
The Statistics Teacher Network

Enrichment for Advanced AP Students

by Jerry Moreno

The seed of an idea:

I must share with you one of the most exhilarating educational experiences I have ever had. Last January or so, Jim Matis, a statistician in the Department of Statistics at Texas A&M and a dear friend of mine from working together on the ASA/NCTM Joint Committee a few years ago, called me. For a couple summers, he had been giving a workshop for students who had already taken advanced placement statistics. The purpose of the course was to enrich their experience beyond what they had learned in the AP class and to introduce them to research. There was a possibility of continuing and expanding the program through a National Science Foundation grant, but he first needed a replica or two to build support.

So because he knew that the Cleveland Chapter of the ASA had been doing workshops in northeast Ohio for a decade or so, he called me to see if I might be interested in doing something a little different.

RAPS:

First I’ll very briefly describe Jim’s program in 2000 that he called RAPS, Research for AP Students. (The one in 2001 investigated a different topic but was equally exciting and intense.) The program in the summer of 2000, Stochastic Modeling, ran for five days July 31 through August 4 for around fifteen students primarily from the Science Academy in Mercedes, Texas. The objectives were to expose the students to some current research, involve them in disseminating research results to subject matter researchers, and to challenge them to participate in possible extensions. Shortly before RAPS 2000, Tom Kiffe, a mathematician at Texas A&M, had published a module Stochastic Population Models with Jim. A workshop version of this module was written to be used as the material for RAPS. From the module, the research focused on changes in population sizes for various biological populations. The models, mostly sets of differential equations, seek to understand the basic kinetic structure of the process that generates the observed data. Ideally such kinetic models could be used to analyze population data, to make statistical inferences relating to population size, and ultimately to predict, or even help manage, population size. They worked with stochastic versus determinis-
it is very difficult if not impossible to say "no" to him. And so, I was caught a bit between a rock and a hard place—very unsure that I could pull this off in the short time I had to organize such a workshop (four months at best), while talking to someone who I swear could sell ice to Eskimos! I said "yes!"

**RAPSS:**

Building on Jim’s RAPS, I chose to call our version RAPSS, Research for Advanced Placement Statistics Students. My first two tasks were to find instructors and to find funding. Jim had mentioned that he felt in order to get students to sign up for the experience, that they would have to be paid. After all, most if not all would have summer jobs that they would have to leave for a week in mid-summer. Also, some will no doubt have graduated and were looking forward to entering college so going to school in the summer may not be considered a “fun” thing to do.

Regarding the instructors, I personally did not have any current research to offer, certainly not on the order of that which Jim did. So I took myself out of the teaching role. I decided to build the workshop on one instructor for six students. How many such groups I could offer depended on two things, finding the instructors, and finding the students! I wasn’t honestly sure I could get any students to bite, stipend or no stipend. Besides, many of my statistics friends work in industry or business and would no doubt find it very difficult not only to get away from work for a week, but also to find a real top of current interest that they could present that would not jeopardize confidentiality of their client(s). I set my goal to find two instructors and twelve students.

I figured that I would need $7300: $250 stipends for 12 students; $1500 remuneration with $150 in expenses per instructor; around $500 in lunches for everyone; and $500 in workshop materials, supplies, copying, mailing, etc. Although we have a couple foundations in the Cleveland area that might be receptive to such an idea, even if I could write the grant in a couple days the turnaround response time would be at least two months. So I turned to our Development Office at John Carroll, and asked them if they would be able to help me find funding. Sure enough, they identified an organization that was interested and hurdle one was done. The second hurdle was finding instructors. This proved to be as difficult as I thought it might be for the reasons previously mentioned, including some who reacted with the skepticism as I had that it really wouldn’t be that worthwhile. But I found two wonderful colleagues who were open to the challenge.

Dr. Janet Larsen is a colleague in the Psychology Department of John Carroll University. Her main research interest is human memory. She teaches courses in sensation and perception, human memory and cognition, and research methods in psychology. Her thoughts were to involve her six RAPSS’ students in design, carrying out and analyzing data from an experiment related to human perception, memory or thinking. The statistics learned would be in the area of analysis of variance, factorial and repeated measures designs. Dr. Harold “Smitty” Haller is president of his own statistical consulting firm, has over 30 publications, an excellent speaker and teacher. His firm specializes in helping organizations implement business, manufacturing, and research oriented continual improvement efforts through training and direct involvement in projects. He thought that he would get his six RAPSS’ students involved in experimental design, particularly fractional factorial.

I sent flyers to the teacher participants who had taken my advanced placement statistics I, and advanced placement statistics II workshops. Students had to apply to RAPSS including why they wanted to be considered, and a letter of recommendation from their teacher. The information made it clear that we were looking for very able students, independent thinkers who could also work in a group, and who would have an interest in an intense week of enrichment in statistical research. Not knowing if I would get even one application, I was floored that within a month, I had 21 outstanding students to consider! Pleasantly surprised, I was then confronted with the problem of how to choose 12 participants from the 21. To give you an idea of my problem, one applicant had 1600 as his total SAT scores, two others had 800 SAT mathematics scores, two others were eighth graders! But before I made any decisions I returned to the Development Office in hope that they could possibly find another $3000 so that I could afford eighteen students and one more instructor. If additional funds were needed beyond that, I knew that the ASA Cleveland Chapter would be willing to help financially. A couple days later, they called and said that they had found the funds for me. Thank you Development Office!

Of course, that meant that I had to find another instructor. Fortunately, Dr. Ralph O’Brien, Director of the Collaborative Biostatistics Center at the Cleveland Clinic Foundation, a member of the ASA Cleveland Chapter and an active member in the Teaching of Statistics in the Health Sciences Section of ASA thought that RAPSS was a very exciting idea. However, because of job commitments, he would have to work with his six students at the Cleveland Clinic rather than at Carroll; I didn’t see that to be a problem and it turned out fine. He was sure that he could get his six RAPSS students deeply involved in clinical trials research.

Two students were not able to be released from their summer employment so we had nineteen student participants, two groups of six and one of seven.
The week's agenda (8:30 – 4:30 daily):

Monday—Introductions: hour presentations by each of the three instructors concerning what they do and what they were suggesting their RAPSS research problems to be; Selection of teams. I had the students list their preference for project. Some of them had no preference. I was able to assign first choices to all who specified a preference. A short time was spent with instructor and student team.

Tuesday—Friday morning: Each instructor split their six students into two groups of three. Each group worked on a different problem under the guidance of the instructor mentor. There was a considerable amount of interaction between the students and mentors. Material for the projects was learned from lectures, the web, handouts and articles. It was a delight for me to visit the various groups through the week and observe the learning process. Incredibly exciting.

Friday afternoon: Each of the six groups gave half-hour power point presentations of their work. They did a superb job as parents, their teachers, their RAPSS instructors and I would attest.

Topics:

One of Dr. Larsen’s groups presented How Does Facial Expression Affect the Emotions of an Individual? This group based their project on a theory by Israel Waynbaum on vascular theory of emotional efference in which it is conjectured that facial expression can produce changes in brain temperature, therefore changing emotions. They created an instrument to test whether or not facial expressions actually have an effect on a person’s emotions, collected data, and analyzed it by a two-factor factorial analysis of variance design. The other group’s presentation Clothing Style Effects on First Impressions of Females investigated the direction and strength of the effect that clothing type at three levels of business, casual, and skimpy has upon impressions of a female as well as assessing the differences in the direction and strength of these impressions between male and female subjects. They studied two research papers plus Dr. Larsen’s own research before they developed a semantic differential scale, interviewed 85 subjects, and used a two-factor analysis of variance to analyze their data.

Dr. Haller gave both of his groups the following problem to investigate. “Your company provides a water-base solution of closely controlled properties. The property of interest is the velocity of a plastic bead as it falls through the fluid. The velocity is measured by dropping a 3/16 inch nylon bead into a 100 ml graduated cylinder and measuring the time the bead takes to drop the distance between the 100 ml mark and the bottom. The specification is 8 plus/minus 1 seconds at 20 degrees C. The solution is to contain all of the following ingredients in addition to water: NaCl, NaHCO3, and Sucrose. The solution must contain at least 1 percent of each ingredient, but otherwise the amounts are unspecified, and it must be the most cost effective.” The students learned about sequential design of experiments, the theory of fractional factorial designs, multiple regression analysis, and the analysis of means. Because of the complexity and interest of the problem, both groups of three worked on it. Their papers were Three Level Fractional Factorial Designs of Experiments, and Studies in Design of Experiments Implementing the Central Composite and Fractional Factorial Designs.

Dr. O’Brien has done substantial work in power analysis having written a terrific piece of software Unifypow. Both groups studied power and sample size determination in depth. One group researched a paper by Tascilar et al. to formulate the following research question. “People with the active SMAD4 Gene, which makes the body create the SMAD4 Protein, live longer after a surgery to remove a pancreatic tumor using the Whipple method of resection than patients without the activated gene. The SMAD4 protein seems to suppress the tumor from growing again after resection. If a person with pancreatic adenocarcinoma, but without the activated gene, receives a supplement of the protein along with chemotherapy, will they live longer than a patient receiving chemotherapy without the supplement?” This group learned about double blind clinical trials, blocking, Cox regression, Kaplan-Meir survival curves, and Cox proportional hazards models. They used the NCSS statistical software. The other group studied a double-blind, randomized clinical trial comparing the effects of two beta-blockers on the blood pressures of hypertension patients of ages 55-80 as well as the effects of these drugs on gender. Beta-blockers are drugs that are used to treat high blood pressure, usually in patients with heart problems such as hypertension (systolic blood pressure in excess of 130 or diastolic in excess of 90). They learned two factor analysis of variance and used NCSS software to conduct the analysis. They included a budget analysis as their study had to fit a fixed amount. The title of their presentation was The Effects of Beta-Blocker Aolol and Beta-Blocker Bolol on the Blood Pressures of Hypertension Patients.

Concluding remark:

Unfortunately, the NSF category which Jim and I were interested in pursuing never materialized. Nonetheless, I thank Jim for suggesting the workshop. It was one of the most exciting educational exercises I have ever been a part of. If any of you have an interest in conducting a similar workshop for advanced placement statistics students, or any statistics students for that matter, please don’t hesitate to contact me. I am more than willing to help.
As statistics continues to increase its presence in the school curriculum, particularly the mathematics one, teachers complain that if statistics must be included, then something must go. One suggestion to solve the problem is to combine the topics of statistics and mathematics so that both are presented together. The NSF-funded project *Data-Driven Mathematics* has done precisely that. The series of eleven modules motivates mathematics topics found in pre-algebra, algebra, geometry, advanced algebra, and advanced mathematics from a data point of view that involves students. Each module includes the mathematics and statistics content that is covered, why the content is important, a pacing and planning guide for the units, and quizzes and exams. There is a student and a teacher edition. Previous issues of STN covered data-driven algebra (48) and data-driven geometry (47). This article will introduce one idea in the module *Modeling with Matrices* by G. Burrill, J. Burrill, J. Landwehr, J. Witmer, Dale Seymour Publ, ISBN 1-57232-259-4.

The Use of Matrices to Represent Ratings

Studies are often conducted to determine the most desirable city to live in based on a number of variables such as cost of living, quality of education, availability and security of jobs, extent of cultural and recreation opportunities, transportation, health facilities, air quality, safety and crime, and climate. Each city in the study could be given a ranking per variable and the resulting rankings for all the variables weighted somehow to determine the most desirable city in which to live. Note that if rankings are used per criterion, the advantage is that all variables are on the same scale, but the variability of the original data has been lost.

To solve the difficulty of what to do in the case of the variables of interest not having the same scale of measurement, ratings from 1 (low) to 100, say (high) might be used. Yet other situations may involve variables that are not all on the same scale and it is desired to use neither ratings nor rankings to solve the problem.

For example, consider the problem of having to determine a textbook for a course in which the variables of interest are the quality of graphics and text, the accuracy of content material, the topics covered, the quality of problems in the exercises, the quantity of problems, and whether or not there are supplementary materials to aid teacher and student. Teachers at a high school known to the author had to choose a text series for their mathematics curriculum from six available that will be called T1 through T6. There were eight teachers in the department that rated each series from a low score 1 to a high score 10 per each of the seven variables previously listed. The rounded mean of the teachers ratings were taken and are shown in the following table.

<table>
<thead>
<tr>
<th>Topics Covered</th>
<th>Graphics Quality</th>
<th>Text Quality</th>
<th>Problems Quality</th>
<th>Problems Quantity</th>
<th>Content Accuracy</th>
<th>Supplements</th>
</tr>
</thead>
<tbody>
<tr>
<td>A B C D E F G</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>T1 7 9 9 7 8 9 7</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>T2 9 8 9 8 9 8 8</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>T3 8 8 9 10 6 8 8</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>T4 8 7 9 8 8 9 8</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>T5 9 9 8 9 7 8 8</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>T6 9 10 9 7 8 10 8</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

For purposes of brevity, suppose that variables B, D, and E are considered the most important and that D is three times as important as B, and E is twice as important as B. A model for the problem is then written as textbook rating \( R = B + 3D + 2E \). To determine the highest rated textbook, students would probably compute \( R \) for each \( Ti \) by using the weighted formula. In so doing they should easily recognize that each computation involves the \( (B,D,E) \) values being multiplied respectively by \( (1,3,2) \). They can then be shown that their observation is organized mathematically by the use of matrices and that the definition of multiplication of matrices follows directly from their observation.

Let \( S_{6 \times 3} = \begin{bmatrix} 9 & 7 & 8 \\ 8 & 8 & 8 \\ 8 & 10 & 6 \\ 7 & 8 & 8 \\ 9 & 9 & 7 \\ 10 & 7 & 8 \end{bmatrix} \quad W_{3 \times 1} = \begin{bmatrix} 1 \\ 3 \\ 2 \end{bmatrix} \quad R_{6 \times 1} = \begin{bmatrix} r_{T1} \\ r_{T2} \\ r_{T3} \\ r_{T4} \\ r_{T5} \\ r_{T6} \end{bmatrix} \)

Which textbook is ranked first according to the chosen weighting scheme?

The answer is the ratings vector

\[
\begin{bmatrix} 46 \\ 48 \\ 50 \\ 47 \\ 47 \end{bmatrix}
\]

---

Statistics in the Classroom...

**Teaching Matrices from a Data Point of View**

by Jerry L. Moreno

moreno@jcu.edu

Teachers at a high school known to the author had to choose a text series for their mathematics curriculum from six available that will be called T1 through T6. There were eight teachers in the department that rated each series from a low score 1 to a high score 10 per each of the seven variables previously listed. The rounded mean of the teachers ratings were taken and are shown in the following table.
So, we see that publishers T3 and T5 are tied for being best according to the specified weighting scheme. If there were other weighting schemes, then \( W \) would be expanded with each new scheme occupying a new column and the size of \( R \) would change accordingly, one ratings column for each weighting scheme in \( W \).

The student should also recognize that the weighting scheme \[
\begin{bmatrix}
\frac{1}{3} \\
\frac{1}{3} \\
\frac{1}{3}
\end{bmatrix}
\]
yields the same ratings as \[
\begin{bmatrix}
1 \\
1 \\
1
\end{bmatrix}
\]
does except the ratings are one-third as large.

**The Use of Graphs to Represent Ratings**

A graphical geometric solution is introduced to answer the question as to which text is best if one were interested only in Graphics Quality (B) and Problems Quality (D).

If a straightedge were taken vertically, sweeping from the right of the graph toward the Graphics axis, then clearly T3 is the best on Problems Quality. If the straightedge were swept horizontally from the top toward the Problems axis, then T6 is the sole choice for Graphics Quality. Note that T5 is better than T2 with respect to both variables. But overall no text dominates all texts.

The equally weighted model would be \( R = B + D \), or \( B = -D + R \), where B represents Graphics Quality and D represents Problems Quality. If the line \( B = -D + R \) were swept from the upper right hand corner across the scatterplot toward the origin, then one would see that T3 and T5 are met first at the same time. This result clearly coincides with the matrix result of ratings, written in its transpose for convenience of space, \[
\begin{bmatrix}
16 & 16 & 15 & 18 & 17 \\
17 & 17 & 15 & 18 & 17 \\
18 & 17 & 15 & 18 & 17 \\
17 & 18 & 15 & 18 & 17 \\
17 & 18 & 15 & 18 & 17
\end{bmatrix}^t.
\] T3 and T5 are tied.

Students could be asked that if they were employees of the publisher of T6, then what model(s) would they be happy with? Note that the slope of the line through T6 and T5 is -1/2. The students should suggest that they would be pleased with any model whose slope is larger than -1/2. Then ask them that for such models how would the customer be viewing the relationship between Graphics Quality and Problems Quality? Suppose the model with slope -5/16, say, were chosen. Then, \( B + 5/16 D = R \). The customer would interpret this model as one in which Problems Quality is considered 5/16 as important as Graphics Quality, or Graphics Quality is 16/5 as important as Problems Quality.

**PREDICTION**

The statistics topic of prediction or regression analysis is rich in the use of matrices. The module does an excellent job in presenting the regression module in matrix notation and determining the least squares estimates of the parameters. A companion module Modeling with Logarithms investigates nonlinear models and transformations to linearity using logarithms.

I encourage teachers who are looking for material that combines standard mathematics topics and data analysis to look at the Data-Driven Mathematics set of modules.
**Keep Us Informed ...**

The Statistics Teacher Network is a newsletter published three times a year by the American Statistical Association—National Council of Teachers of Mathematics Joint Committee on Curriculum in Statistics and Probability Grades K-12.

Send articles by email to the Editor:

**Editor** - Jerry L. Moreno  
Dept. of Mathematics and Computer Science  
John Carroll University  
University Hts. OH 44118  
moreno@jcu.edu • fax: 216-397-3033

Send mailing address corrections and additions to: Madge Haven: madge@amstat.org