STATS

STATS

The Magazine for Students of Statistics

Winter 2003 • Number 36

Editors

Beth L. Chance
email: bchance@calpoly.edu
Department of Statistics
California Polytechnic State University
San Luis Obispo, CA 93407

Allan J. Rossman
email: arossman@calpoly.edu
Department of Statistics
California Polytechnic State University
San Luis Obispo, CA 93407

Editorial Board

Patti B. Collings
email: collingp@byu.edu
Department of Statistics
Brigham Young University
Provo, UT 84602

Gretchen Davis
email: davis@stat.ucla.edu
Department of Statistics
UCLA
Los Angeles, CA 90095-1554

E. Jacquelin Dietz
email: dietz@stat.ncsu.edu
Department of Statistics
North Carolina State University
Raleigh, NC 27695-8203

David Fluharty
email: fluharty_david@hotmail.com
Continental Teves
One Continental Drive
Auburn Hills, MI 48326

Robin Lock
email: rlock@stlawu.edu
Department of Math, CS, and Stat
Saint Lawrence University
Canton, NY 13617

Chris Olsen
email: colsen@esc.cr.k12.ia.us
Department of Mathematics
George Washington High School
Cedar Rapids, IA 53403

Production

Megan Murphy
email: megan@amstat.org
American Statistical Association
1429 Duke Street
Alexandria, VA 22314-3415

STATS: The Magazine for Students of Statistics (ISSN 1053-8607) is published three times a year, in the winter, spring, and fall, by the American Statistical Association, 1429 Duke St., Alexandria, Virginia 22314-3415 USA; (703) 684-1221; fax: (703) 684-2036; Web site: *www.amstat.org*

STATS is published for beginning statisticians, including high school, undergraduate, and graduate students who have a special interest in statistics, and is distributed to student members of ASA as part of the annual dues. Subscription rates for others: $13.00 a year to members; $20.00 a year to nonmembers.

Ideas for feature articles and material for departments should be sent to the Editors; addresses of the Editors and Editorial Board are listed above.

Requests for membership information, advertising rates and deadlines, subscriptions, and general correspondence should be addressed to *STATS* at the ASA office.

Copyright © 2003 American Statistical Association.

Features

Departments

Editors' Column

Beth Chance Allan Rossman

Our feature article in this issue comes from Chris Palmer, Director of the Centre for Applied Medical Statistics at the University of Cambridge in the United Kingdom. Chris makes a compelling argument that statistics and ethics are closely aligned, particularly in the planning and analysis of clinical trials. In addition to summarizing the key aspects of clinical trials and how they have developed historically, Chris describes strategies for adapting the trial design as data are collected. This strategy allows for treating current patients more ethically, without sacrificing the accumulation of information for the benefit of future patients. He argues that today's students and tomorrow's statisticians will have opportunities to apply their skills to develop more ethical clinical trials. We especially appreciate Chris' admonition that patients should be treated less like fertilizer and more like ammunition; please read the article to decipher what he means by this statement.

Mary Ann Ritter, a retired statistician and manager from General Motors, shares her experiences by offering practical advice to help bachelor's-level statisticians to land jobs in industry. Her suggestions are especially valuable since they are based on responses to a survey of eight non-academic statisticians about what they look for in their hiring practices.

One of the highlights of the year for us is always the annual Joint Statistical Meetings, the largest conference for statisticians, every August. Curious about the experiences of students who attend this conference, we asked several students (both graduate and undergraduate) to write about their reactions to last year's meeting in New York City. In this issue we are pleased to present the perspectives of Middle Tennessee State's Leann Hooge, Keeley Hutchinson, and Ileah McKee, as well as North Carolina State's Venita DePuy. We hope that their tales will encourage more students to attend this summer's conference to be held in San Francisco. Further enticement is provided by the return of the Stat Bowl competition to the JSM this summer, with ASA helping to provide travel costs for student participants, as described in Mark Payton's article.

Students of statistics quickly learn that our field involves so many symbols that we have to make heavy use of the Greek alphabet to make room for them all! Elena Papanastasiou, a native of Greece and a faculty member at the University of Kansas, offers suggestions for dealing with this potential confusion by learning how to pronounce Greek letters correctly.

This issue's Student Project article comes from Katherine Van Schaik, a junior in the Discovery Math and Science Magnet Program at Spring Valley High School in Columbia, South Carolina. Katherine won several prizes at the International Science and Engineering Fair last spring, including the ASA Award for Best Use of Statistics. In this article she describes her study of the effects of using arsenic to treat lumber used in playgrounds. Joe Ward, a tireless devotee of statistics education from San Antonio, introduced Katherine to *STATS* magazine, and Joe has written a sidebar to accompany Katherine's article that offers advice about science fair competitions for both students and statisticians.

Our statistical sport fan Robin Lock is joined by his daughter Kari Frazer Lock in writing this issue's column. Kari is a junior Mathematics major at Williams College and is also an accomplished figure skater who has performed throughout the United States and abroad. The Locks address a topic very near to Kari's heart by using a bootstrap approach to analyze the consistency of ratings awarded by figure skating judges, using the controversial results of the 2002 Salt Lake City Olympic competition as their example.

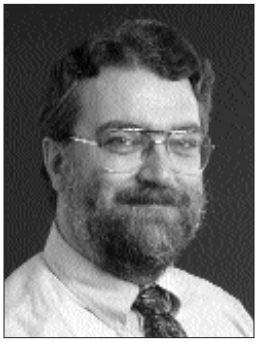
We are very pleased to introduce a new member of our editorial board with this issue. Josh Tabor of Wilson High School in Hacienda Heights, California joins our team and contributes an AP Statistics column about interpreting computer output for regression analysis. While we welcome Josh aboard, we also want to thank his predecessor Gretchen Davis for her contributions to STATS in the past year.

Chris Olsen offers new µ-sings in this issue, revealing that STATS readers who are also fans of detective fiction may enjoy a novel by Stanislaw Lem in which the protagonist who solves the case is actually a statistician!

This issue's Data Sleuth Mystery was contributed by Michael Kahn of Wheaton College, who provides data appearing to suggest that smokers have greater lung capacity than non-smokers. Please see if you can solve this mystery, and consider sending us your own exam-

An Interplay Between Statistics and Ethics

Data-Dependent Designs in Clinical Trials



Chris Palmer

Statistics, Ethics and Trials

Some might think of statistics as a mathematical science all about data analysis and not at all about people. Well, in medical statistics in particular, such is far from the truth. In fact, matters of health (including ethics, which promotes the well being of people by seeking to 'do the right thing') and statistics come together in the field of biostatistics, a term I use interchangeably with 'medical statistics.' Statistics and ethics may seem poles apart as a pair of disciplines, but I believe they are more closely linked than one would first think: one is all about pursuing numeric-based truth amid uncertainty as to what may, might or could happen; the other is all about pursuing what is right, also amid uncertainty, as to what must, ought or should happen. This link between ethics and statistics is seen most clearly in the way researchers, including biostatisticians, contribute to medical science by means of the controlled clinical trial (Figure 1), for it is here that mathematics and people meet head on.

In this article, I review what statisticians have done for trials so far, including a whistle-stop tour of the history of trials and the necessity of randomization for allowing sound inference; how trials strive to keep the quality of clinical research as pure as possible, if not pure as gold; some present-day pressures that are mounting against trials in general; and finally some discussion about what statisticians may be able to do for trials in the future as a way of alleviating these pressures. Here I describe the role of *data-dependent designs,* which have been around on paper as theoretical options for quite some time, but have yet to make their mark on research practice. In short, and in contrast to trials based on fixed sample sizes, these modern designs seek to make fullest use of emerging data within, not just following, a clinical trial, and for this reason are also known as "learn-as-you-go" designs.

One cannot and must not ignore the fact that in clinical research the 'units of experimentation' are actually priceless human beings. It is imperative to get the balance right between, on the one hand, individuals who are in the trial and, on the other, those who stand to benefit from the results of the trial. Incidentally, the term 'guinea pigs' would be unfair to describe those in the former group, as trial participants generally fare better than contemporaries having the same illness outside of trials. Those benefiting subsequently are the people yet to get the disease being studied, so the dichotomy can be thought of as best serving the needs of current and future patients. These are, respectively, the domain of *individual ethics* and *collective ethics*, terms reflecting primary concerns with today's volunteers and tomorrow's society at large. As noted by Pocock (1983) among others, clinical trials are a delicate balance between these often-conflicting priorities.

Chris Palmer (chris.palmer@medschl.cam.ac.uk) is an academic medical statistician and founding Director of Centre for Applied Medical Statistics, University of Cambridge. Originally a mathematician at Oxford, Chris lived in US 1982-88 while a post-doc at Harvard's Biostatistics Department following his PhD in Statistics at UNC-Chapel Hill where his thesis topic was an ethically-motivated (actually a decision-theoretic) clinical trials model.

Randomization and Controlled Trials

What is Randomization?

Clinical trials, also known as randomized controlled trials (RCTs), are regarded as the "gold standard of medical research". The process of randomization—allocating treatments to patients by an external chance mechanism—is alone responsible for allowing doctors to conclude that perhaps a new treatment A is truly better than an old treatment B, with results that are beyond mere chance variation. Why is this so? Essentially, it is because two groups of patients are created, each receiving one treatment, that are effectively 'identical' bar one thing: the treatment received. Hence, any difference in health outcome recorded at the end of the trial, beyond what would be expected by chance, whether a matter of days or years later, can be ascribed to the treatment received.

Figure 1. Medicine and mathematics, in particular ethics and statistics, meet head on in clinical trials. [Image from University of Massachusetts Medical School website, permission to use requested]

Why is Randomization Necessary?

It takes randomization, which could literally be as simple as the tossing of a coin to decide group membership, to ensure the two groups are homogeneous. If we allowed human intervention to decide who went into which group, we would not be able to draw such powerful conclusions. This may seem counter-intuitive, but it is not. Suppose we thought a person's health outcome was related to their age, sex, previous medical history, body mass index (a measure combining weight and height), smoking behavior, *etc*., then wouldn't we be better off forcing the groups to be balanced with respect to this list of items? Why leave it to chance?

Well, the key word in the list just cited was "*etc*." We will never know all the other factors that influence someone's health. At best, we can measure and record, albeit imperfectly in many cases, a number of factors that we think are related. We may even have evidence that some or all of them really do influence long-term health. But we can never be confident we have an exhaustive list of such factors, which is why we have to let the chance mechanism balance things out for those that are known and (the endless list of) those that are unknown. So, randomization is indispensable within clinical trials.

Trial Conduct

Avoiding Bias

In practice, trials are conducted with some important safeguards to supplement the cornerstone of randomization. If, for instance, it is desired to create groups that have an equal number of men and women, or of young and old people, or whatever, then there are ways to ensure this. Ideally subjects are also randomly selected from some larger population so the results can be generalized to a variety of people groups and not limited to some subgroup of the population. It is also necessary to ensure sample sizes are adequate, for if not then the law of large numbers will not come into play and the statistical inference may miss effects that are truly present.

A major difficulty in clinical trials is the need to overcome bias—any systematic tendency for results to be different from the true relationship. Sample selection bias is the most common problem, in that those who enter into a trial in a given disease area may not be truly representative of the entire population of people with that particular disorder. But bias can be more insidious than that, and anytime human interventions or decisions are made, there is potential for bias to creep into a study. This is why the design stage of a study is even more important than the analysis stage, for then it is too late to correct any adverse influences on the manner in which data were collected and entered into the computer database.

Blinding

Keeping a trial "double-blind" is another commonly employed safeguard to the quality of the research when it is both feasible and ethically acceptable to do so. It means not telling the patients which treatment group they are in, and also withholding this information from the doctors (or whoever is responsible for) assessing their health outcomes. The first aspect of blinding is to prevent self-delusion in thinking "I'm on the new treatment, so I must get better." This is related to the so-called *placebo effect* whereby people do get better just by being prescribed a dummy tablet, of no intrinsic effect, from someone wearing a white coat and stethoscope! The second aspect of blinding is subtler, but it ensures the research study has not been influenced by doctors' subjective judgments and preferences. If a doctor was reading an x-ray, say, and had to decide whether a tumor had shrunk, remained the same, or grown, then borderline decisions could be affected—either consciously or sub-consciously—by knowledge of whether the patient was on old or new treatment.

Double-blinding guards against this possibility and prevents skeptical readers of the trial results, once published, from levying such accusations. Good science is all about good quality research and can be seen as an example, if ignored, of how "bad statistics is bad ethics," a point serving to underscore the link between the disciplines.

Safety Monitoring

During a trial's recruitment phase, the conduct stage between design and analysis, patients are randomized and their chosen treatment initiated. In long-term trials, as often is the case for instance in cardiology or in many cancers, there can be a lengthy period before

each patient reaches his or her endpoint (which is just as well if "failure" is really a statistical euphemism for death), or until the study reaches its pre-determined termination date. Especially if mortality is an issue, such trials would typically have a Data and Safety Monitoring Board (DSMB) checking updated trial results on a regular basis. The DSMB is yet another safeguard built into trials, since they consider any deaths and serious adverse events at each meeting. Their charge is to inform the trial steering committee if important differences are emerging between the treatment groups, with a view to stopping a trial "prematurely" if warranted. This is a decision invoked quite rarely, and after serious debate, as there are various penalties incurred for stopping too early, particularly if the trial may never be repeated again perhaps for ethical reasons. Generally, DSMB members do not know which group is which, dealing instead with generic labels, e.g., "Group A" and "Group B," with the code only being broken if necessary. Sometimes this is referred to as "triple-blinding"—one can only hope no trial has ever accidentally been quadruple-blind, in which no one has a clue about who was in which group!

Brief History of Trials

Ancient History

A brief history of trials can extend as far back as the Biblical book of Daniel, circa 2600 B.C. It is reported (see Daniel, Chapter 1) that four young men requested and received a simple diet of water and vegetables so as to avoid 'unclean' food. Their appearance ten days later turned out to be better than that of other young men given a rich diet including wine and inappropriately idol-sacrificed meat. The study can perhaps be criticized through modern eyes for using a very small sample, having a short follow-up duration, being unblended, and especially, being non-randomized. Then again, it can hardly be faulted for choice of publication for reaching the maximum readership!

Recent History

Although there are some examples of comparative studies in the last three centuries, the history of *randomized* experiments is much shorter, dating only from the mid-1920's and not even within medicine, but in agriculture. The eminent 20th century statistician, and indeed founding father of much of modern statistics, Ronald A Fisher conducted a series of crop field trials in the UK (1926) to determine the best fertilizer to apply to a field of wheat (Figure 2).

Fisher's methods required the use of randomization so that plots of land were equally distributed among fertilizer treatments regarding aspect, slope, rainfall, soil composition, (and that all-important) *etc*. Notice

Figure 2. Fields of wheat in controlled studies to determine most effective fertilizers involve non-precious resources. [Image from Texas A & M University website, permission to use requested]

that in such a trial the outcome is the plant yield obtained at a suitably pre-defined harvest time after the growing season.

This trial was the pre-cursor to the first use of randomization in human beings some two decades later. Well-known British medical statistician Austin Bradford Hill convinced doctors, in part because the novel treatment was in short supply, to employ randomization of the drug streptomycin together with bed-rest or else bed-rest alone in a study involving patients with pulmonary tuberculosis. This has been acknowledged as the first ever RCT to be published, appearing in the *British Medical Journal* in 1948.

Present-Day Trials

Today's trials have evolved and matured somewhat since the middle of last century. You may be surprised to learn that there is a journal devoted to *Controlled Clinical Trials* and even an international organization of interested professionals called the *Society for Clinical Trials*, set up some 25 years ago. In drug development, there are four accepted stages or *phases* of trials, labelled from I-IV, ranging from first tentative use in man, up to large-scale, post-marketing surveillance studies to detect rare or long-term side effects. The RCT is firmly established at the heart of "evidence-based medicine," with randomized evidence rightly considered to be the best available type for deciding whether interventions work and hence for justifying doctors' treatment decisions for their patients.

However, at the same time, trials of today are fundamentally quite similar to those of fifty years ago, in that they typically involve equal allocation of treatments to patients, generally after performing a power calculation to determine a target number of patients to be recruited. So, in a two-treatment comparative trial, half the patients customarily receive the standard, half the experimental treatment. With the possible exception of DSMB committee members and a statistician conducting an interim analysis, no one looks at the results until

Figure 3. In World War II, ammunition had to be tested but obviously one only wanted to perform controlled explosions for a limited number of shells to conserve precious resources. [Image from Dr Robert E Sterling Collection, Joliet Junior College, Illinois, used with permission]

all the patients have been randomized and followed up. At the end of the trial it is possible that the experimental treatment is declared a statistically significant improvement and heralded as a clinical success.

It is an ethical problem, however, if as mentioned before 'failure' means the patient died and one can look back with some remorse wondering "if only we had come to this conclusion sooner perhaps we could have saved some lives." Even if the outcome is not as serious as death, the argument persists: could fewer patients in the study have suffered on the way to reaching a valid conclusion?

This last question has motivated much research by ethically minded statisticians. Ironically, this work dates back at least as far as the first modern clinical trial, for the whole area of "sequential analysis" traces its history to the 1940's, World War II and U.S. government-contracted statistician Abraham Wald (1947). His work was also not in the medical area of application, but in ammunition testing (Figure 3), an altogether different example of seeking to cope with precious and limited resources. Medical application of sequential methods does seem entirely appropriate. After all, patients arrive to be treated sequentially (they are not all waiting in line outside the doctor's office or hospital clinic at the start of a trial!) and similarly, results from some are available sooner than from others.

The rationale for sequential trials involves looking carefully at data *as they accrue* with a view to stopping just in time. Hence, the number of experimental units required is not fixed in advance but is a random variable. Theory shows that the expected number of patients involved in a sequentially analyzed randomized controlled trial is less than the corresponding fixed sample size trial, for any given power and level of significance (Jennison and Turnbull, 2000). It is possible, when treatment groups fare broadly equally well, for a sequential trial to need slightly more patients overall compared with a trial using traditional design, but this would be quite unusual.

For better or worse, the clinical trial as conducted and analyzed today is not in Wald's style of testing ammunition but rather in Fisher's application of fertilizer to fields of wheat. These two metaphors illustrate the fundamental difference between the statistics behind clinical trials that strive to learn-as-they-go and those that wait, literally, until harvest time before beginning to make scientific inferences. Personally, I believe that, contrary to normative practice, wherever possible, people who volunteer for clinical trials should be treated with the same respect as afforded the ammunition, and not the fertilizer.

Data-Dependent Designs

Here I outline four broad categories of data-dependent designs, all of which are in the spirit of learning-as-they-go, as opposed to ignoring intermediate data. I describe each in turn along with some specific features when applied to clinical trials.

Sequential Designs

Following Wald's pioneering research, sequential designs have evolved as sophisticated tools to assist those on DSMBs and hence can be considered mainstream, in contrast to the remaining design types discussed below. It should be said though that these methods are not routinely implemented as primary analytical tools for driving trials. Instead, at best, they are used as 'back seat drivers' to exert indirect influence on trial conduct. How do they do this? Essentially, as data accumulate, a test statistic can be plotted on a graph of treatment difference vs. time, and trial recruitment can be recommended to terminate just as soon as a pre-determined boundary is crossed.

This boundary may take on various shapes, the simplest being triangular with two possible options (see Figure 4). Either treatment A or B is declared better depending on which side of the triangle is crossed first. To allow for a third, non-conclusive option with a pre-determined maximum trial sample size the boundary outline is modified to include a vertical line at a given point on the time (strictly "information") axis. The idea is to stop the trial in favor of treatment A, say, if the upper line of the boundary is crossed first; B if the lower line; or else, conclude no clinically relevant difference between A and B if the vertical line is reached first.

There are variations on this theme with rules such

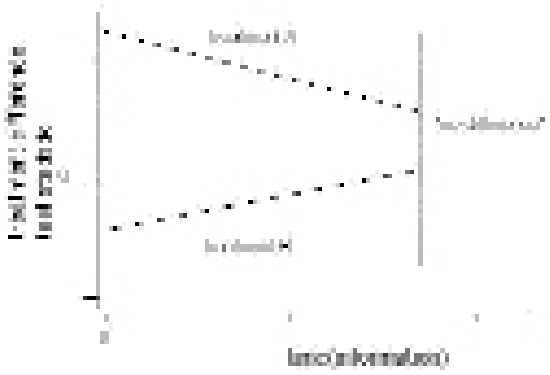


Figure 4. Sample sequential design to choose between Treatment A and Treatment B

as those derived by Pocock and by O'Brien and Fleming being popular examples. Thus it is not necessary to update the graph after every single observation. One can apply rules, called "group sequential methods," that update after small batches of results become available. For more details refer to Jennison and Turnbull's helpful textbook (2000). Statistical software for implementing these rules is readily available in several commercial packages (e.g., EaSt, PEST, S+SeqTrial).

One disadvantage with sequentially designed experiments is that their usefulness, namely their potential to learn while in progress, is self-limiting to trials having relatively rapid endpoints. Thus a sequential trial offers little benefit over a traditional, fixed sample size trial if the outcome remains unknown until years after randomization. For example, this may be so in breast cancer, but is not a limitation in emergency medicine or in rapidly fatal diseases.

Bayesian Designs

Here investigators start by eliciting a prior distribution, either from a panel of clinical experts or from a reasonable selection of available theoretical distributions thought to mimic reality in terms of treatment success distributions. For example, a beta distribution with suitably chosen parameters can represent initial beliefs about a treatment's efficacy ranging from negatively skewed to uniformly distributed to positively skewed. In practice, there is virtue in choosing a prior that makes the experimental treatment appear initially a weak contender, so that positive results in favor of the treatment are not too dependent on initial choice of prior. As the patients' results accumulate, the conditional distribution given the data thus far is evaluated—the so-called posterior distribution, amalgamating the prior and the likelihood. Inference is based on the posterior, including the evaluation of credible intervals, analogous to confidence intervals in the frequentist context.

An advantage is the ease of interpretation of these intervals for they have more intuitive meaning to clinicians and patients. A disadvantage is the general lack of awareness of Bayesian methods since these are, when it comes to schools of statistical thought, the kindergarten in contrast to the senior high of frequentist methods. This is reflected in the comparative lack of statistical textbooks, courses and software aligned to the Bayesian paradigm. Spiegelhalter *et al* (1994) provide an excellent overview of Bayesian methodology applied to clinical trials. Some see the subjective or arbitrary nature of the prior distribution involved as a weakness; others regard it as a positive opportunity to incorporate provisional information about the potential new treatment. Whether Bayesian or frequentist analyses should be preferred within the context of clinical trials is, I believe, another ethical matter discussed further elsewhere (e.g., Palmer, 1993).

Decision-Theoretic Designs

Some experimental studies can be conducted with the resulting inference, in terms of how the information will be used to reach a practical decision concerning which treatment to recommend, as the driving force. For example, one can specify a criterion such as minimizing expected successes lost, or maximizing successes gained, over the course of a pre-determined number of future patients, the "horizon," within and outside a comparative trial. Another criterion could be maximizing the probability of correct selection of superior treatment. Either way, the focus is on the pragmatic need to make a decision to use one of the treatments or not once the trial is over, in a direct attempt to balance the needs of current and future patients. Colton (1963) was among the first to advocate decision-theoretic trials.

It is possible to discount future patients by putting more weight on present results, although this whole area can become mathematically quite intricate, especially when modelling with unconstrained "multi-armed bandits" in the context of deciding among several treatments (Berry and Fristedt, 1985). Nevertheless, practical simplifications can be incorporated, such as limiting designs to equal allocation among remaining treatments. In the case of just two treatments this amounts to allocating pairs of treatments until it is optimal, by whatever criterion, to cease the comparative stage and switch all remaining patients within the horizon to the preferred treatment.

Objections to the subjective nature of prior distributions involved in this type of decision-theoretic framework can be alleviated, for example, by appealing to minimax criteria. This means implementing a design that has good theoretical properties across a broad range of priors. Computer software allowing such designs to be implemented is limited, contributing to the current lack of use of such methods in practice.

Response-Adaptive Designs

The (response-) adaptive design is the most extreme type of data-dependent design, for it incorporates the information accruing from the data to modify the treatment allocation probabilities away from 50:50 in the case of two treatments. For example, the trial would start with equal allocation, but as the data begin to favor one treatment even slightly, the adaptive design increases the odds of the next allocation being in the favorable treatment. In practice it works like this. Imagine a bag containing an equal number of red and blue balls. A red ball drawn indicates the next allocation is to treatment A; a blue ball, treatment B. If a success occurs a ball of the appropriate color is added to the bag before the next drawing, and hence treatment allocation, takes place.

One disadvantage is illustrated by the rather poor prototypical example of a mid-1980's trial (Bartlett, Roloff, and Cornell, 1985), involving extra-corporeal membrane oxygenation (ECMO) therapy, which has received much attention in the statistical and medical literature. Ethicists, clinicians and statisticians have all contributed to the debate about this particular trial (Royall, 1991). It involved critically ill new-born babies and the relevant outcome in question really was a matter of life and death. In retrospect, it was clearly a mistake to begin this trial with precisely one ball of each color in the bag instead of, say, ten of each. What ensued was a highly unbalanced distribution of treatment allocation (for ECMO babies generally lived, unlike many of those not on ECMO therapy) rendering sensible inference difficult, if not impossible.

However, this early example should not be a reason to abandon adaptive designs, just as clinical trials were not abandoned after a famous 1930's experiment in Lanarkshire, Scotland went awry (Student, 1931). In a study to assess growth, researchers had unwisely allowed teachers to choose which children would receive free, supplementary milk. Teachers systematically assigned the poorer children to the "extra milk" group. This confounded the treatment with socio-economic background, making it impossible to isolate any treatment effect.

Adaptive designs are the most controversial of the four types presented here. This is largely because they appear to react too quickly to early data, which may be subject to systematic bias or time trends. Also, if not careful, they can begin to adapt too swiftly even to chance results. There is also the criticism that if one treatment happens to be a placebo, why should anything change after a success or a failure on such an inert substance? Nevertheless, with suitable cautions and awareness of the issues involved, adaptive designs can be a highly effective and ethically appealing design, despite the relative dearth of positive examples of their actual use so far. For further reading on adaptive designs, refer for example to some conference proceedings edited by Flournoy and Rosenberger (1995).

Discussion

Here I raise three key questions of direct relevance to tomorrow's generation of statisticians. Respectively, these concern why, when and how to implement data-dependent designs in the practice of clinical trials. More details and further arguments can be found elsewhere (Palmer 1999; 2002).

Why Use Data-Dependent Designs?

The primary reason for using data-dependent designs is for the ethical advantage in terms of how patients in trials are regarded, without compromising the scientific rigor or usefulness of studies for the sake of future patients. There are also secondary benefits, notably derived from the side effect of expecting fewer patients to be involved in trials.

Pharmaceutical companies, and other trial sponsors, would be glad to spend less money on this aspect of research and development, assuming that regulatory bodies, such as the US Food and Drug Administration, can be provided with evidence from non-standard, randomized designs of sufficient reliability.

- Patients and their families would be pleased to know that their needs were being considered foremost within the trial.
- Volunteers in trials may benefit in some cases from increased chances of receiving the superior treatment.
- Research physicians may find it easier to persuade people to take part in clinical trials in the first place, helping doctor-patient relationships become less one-sided.
- Medical science itself may benefit from expediting the drug development process.

One can add to this list some negatively-motivated reasons. For instance it seems a shame that modern clinical trials have not moved on from the agricultural trials pre-dating them. Failing to put into practice the fruits of five decades of statistical research is almost criminal--why should trial designs and analyses today be limited to techniques available in the 1920's? Would it not be called scandalous if medicine, or any other discipline for that matter, were so constrained?

It is more than a pity that much of modern statistics and computing goes untapped in such a key area of application. Armitage (1985) put it this way (my italics): "[Learn-as-you-go-designs] *are just the sort of contribution that statistics should be making to the design, execution and analysis of clinical trials*…." He goes on to lament "*such lack of contact between theory and practice seems to me quite deplorable*." These sentiments speak even louder now that the best part of two more decades has elapsed, yet things are only slowly beginning to

progress.

A related disappointment is that data-dependent designs depend on not only modern theory, but they also exploit modern technology since they require rapid communication of data back and forth between the statistical centre and the point of contact between doctors and patients. It used to be argued that learn-as-you-go designs were too cumbersome since it is true they carry heavier logistical, practical and administrative burdens. However, with the advent of mobile communications technology (e-mail, cellphones, laptop computers, *etc.*) now so commonplace in many parts of the world, many of these previous barriers are diminishing. Admittedly, there remain minor difficulties associated with drug distribution and budgeting for unknown numbers of trial participants, but these objections are but poor excuses to neglect implementing modern designs in those circumstances when they are relevant and appropriate. So, when is this the case?

When to Use Them?

There are two separate aspects to this question. First, for what sort of clinical trials are data-dependent designs best suited? (Clearly, there are certain situations for which they would be unsuitable.) Secondly, when in real time is a suitable opportunity to start making more routine use of them?

To answer the latter question first, I suggest the time is now ripe, in this, the first decade of the 21st century. There are numerous reasons for this related to the feasibility of rapid data transfer as just mentioned, but more importantly the growing pressure mounting on today's clinical trials. For the RCT is not without critics. Increasingly, influential scientists are realising the clinical trial as conducted today is not necessarily the best thing since sliced bread (Horton, 2001).

When considering the changing nature of doctor-patient relationships; the role of the Internet raising expectations of 21st century patients; the growth of patient support or advocacy groups demanding that their needs be given higher attention than ever before; the perceived and actual threats of medical malpractice lawsuits extending into the realm of trials that have allegedly harmed individuals; problems experienced with recruiting sufficient numbers of patients in large trials; and so the list goes on, it really does seem prime time for a paradigm shift in the way trials are performed. I reckon that introducing data-dependent designs more routinely would help alleviate many of these problematic pressures (Palmer, 2002).

To answer the former question about the type of trials amenable to data-dependent designs, in short, arguments supporting their use are strongest whenever individual ethics outweigh collective ethics. For example, this is the case for early phase trials seeking to combat a serious or life-threatening illness (Palmer 1993; 1999). Data-dependent designs are most appropriate for rare diseases that have rapidly known outcomes, at least relative to the patient accrual rate. Hence the rarer the disease, the longer the anticipated interval before the next patient is recruited, allowing an outcome slightly more distant in time than *immediate*. Furthermore, the greater concern ought to be given to those within the trial if it is a relatively rare disease (recall the balance of today's vs. tomorrow's patients).

However, once familiar with the application of data-dependent designs there is no intrinsic reason why they cannot be used in a wide variety of settings, including later trial phases and more common, serious illnesses. This is particularly true if studies form part of a prospectively planned program of research and contribute, therefore, to a future meta-analysis of a series of high quality randomized clinical trials covering a diversity of patient types, geographical and temporal settings, and study designs.

How to Use Them?

Given the backdrop of very few practical examples thus far of trials that have been governed by data-dependent designs one cannot expect an overnight conversion among trialists. What will help the cause, without doubt, would be one or two good examples, wherein it is clear for all to see the benefits of applying learn-as-you-go techniques in specific clinical trial contexts. I am aware of several such ongoing examples and eagerly await feedback from the medical and, in due course, regulatory authorities, not to mention the oft-neglected patients and their family members. For regulatory purposes, there may need to be an intermediate step between present-day extremes of considering a new drug as 'investigational' one day, yet 'fully licensed' for marketing the next, which seems a remarkably abrupt change. This could be accomplished by introducing a graduated licensing scheme for new drugs, as I suggest in Palmer (2002).

Conclusion

My hope is that there will soon be a watershed and that once the methods of data-dependent designs are seen to be workable and advantageous, there will follow a flood of applications across many disease areas in those situations where they are most appropriate. Perhaps there will even be a day when statisticians look back with amazement at how long it took to start routinely conducting clinical trials in a data-dependent manner, and if that happens, who knows what else might happen? Just maybe, non-statisticians of tomorrow may even get to think of our subject as all about people, and not just data analysis…

References

Armitage, P. (1985), "The search for optimality in clinical trials," *International Statistical Review*, 53,

15–24.

Bartlett, R.H., Roloff, D.W., Cornell, R.G., Andrews, A.F., Dillon, P. W., & Zwischenberger, J.B. (1985), "Extracorporeal Circulation in Neonatal Respiratory Failure: A Prospective Randomized Study," *Pediatrics*, 76, 479–87.

Berry, D.A., and Fristedt, B. (1985), *Bandit Problems—Sequential Allocation of Experiments*. London: Chapman and Hall.

Colton, T. (1963), "A Model for Selecting One of Two Medical Treatments," *Journal of the American Statistical Association*, 58, 388–401.

Fisher, R.A. (1926), "The Arrangement of Field Experiments," *Journal of the Ministry of Agriculture, Great Britain*, 503, 13.

Flournoy, N. and Rosenberger. W.F. (eds). (1994), *Adaptive Designs*, Hayward, CA: Institute of Mathematical Statistics.

Horton, R. (2001), "The Clinical Trial: Deceitful, Disputable, Unbelievable, Unhelpful, and Shameful—What Next?" *Controlled Clinical Trials*, 22, 593–604.

Jennison, C. and Turnbull, B.W. (2000), *Group Sequential Methods with Applications to Clinical Trials*. London: Chapman and Hall/CRC.

MRC Streptomycin in Tuberculosis Trials Committee (1948), "Streptomycin treatment for pulmonary tuberculosis," *British Medical Journal*, vol ii, 769–82.

Palmer, C.R. (1993), "Ethics and Statistical Methodology in Clinical Trials," *Journal of Medical Ethics*, 19, 219–22.

Palmer, C.R. (2002), "Ethics, Data-Dependent Designs and the Strategy of Clinical Trials: Time to Start Learning-As-We-Go?," *Statistical Methods in Medical Research*, 11, 381–402.

Palmer, C.R., and Rosenberger, W.F. (1999), "Ethics and Practice: Alternative Designs for Phase III Randomized Clinical Trials," *Controlled Clinical Trials*, 20, 172–86.

Pocock, S.J. (1983), *Clinical Trials: A Practical Approach*. Chichester: John Wiley.

Royall, R.M. (1991), "Ethics and Statistics in Randomized Clinical Trials (with discussion)," *Statistical Science*, 6,52–62.

Smith, R. (1998), "Fifty Years of Randomised Controlled Trials" (Editorial), *British Medical Journal,* 317.

Spiegelhalter, D.J., Freedman, L.S., and Parmar, M.K.B. (1994), "Bayesian Approaches to Randomized Trials," *Journal of the Royal Statistical Society, Series A,* 157, 357–416.

Student (Gossett WS). (1931), "The Lanarkshire Milk Experiment," *Biometrika*, 23, 398–406.

Wald, A. (1943), *Sequential Analysis of Statistical Data*, Report submitted to Applied Mathematics Panel National Defense Research Committee, September, 1943, (de-classified in Wald, A. (1947), *Sequential Analysis*. New York: John Wiley).

Software Packages

East, Cytel Software Corporation, Cambridge MA. Information available from

http://www.cytel.com/new.pages/EAST.2.html.

Planning and Evaluation of Sequential Trials (PEST), Medical and Pharmaceutical Statistics Research Unit, University of Reading. Information available from http://www.rdg.ac.uk/mps/mps_home/software/pest4/pest4.htm.

S+SeqTrial, Insightful Corporation, Seattle WA. Information available from http:// www.insightful.com/products/product.asp?PID=16.

Employment Advice for Undergraduate Statistics Graduates

Mary Ann Ritter

Introduction

Traditionally, students pursued graduate degrees before embarking on careers in statistics. Despite this focus on advanced degrees, many people are now graduating from college with bachelor's degrees in statistics. In the United States there were 504 bachelor's degrees granted in statistics in 1996-97. People receiving bachelor's degrees in statistics have choices when they graduate—they may go to graduate school, seek immediate employment, start their own business, or they may attend a professional school such as medical, law or business school. This article offers advice for students who will seek employment immediately after graduation. While these are the skills most likely sought by employers, rest assured that no employer expects a student to have all of them.

Where Does This Advice Come From?

Eight non-academic statisticians were asked six questions about employment for new statistics graduates. They answered the questions as well as offered their own ideas. The contributors came from a variety of companies and fields of application:
—Baxter Healthcare Corporation (pharmaceutical)
—Delphi Automotive (quality)
—Ernst & Young (management consulting)
—Intel Corporation (manufacturing)
—Merck & Company (pharmaceutical basic research)
—National Agricultural Statistics Service, US Department of Agriculture (government)
—Southwest Technology Consultants (consulting)
—Westat (consulting)

The six questions about positions for new statistics graduates were

—What jobs are out there? (Position titles)

—What do these statisticians do? ("Key job elements")

—What do students need to know? (Candidate qualifications)

—What do they REALLY need to know? (Detailed qualifications)

—What's most important? (Rating of qualifications)

—Advice or recommendations

Mary Ann Ritter (ritterma@ix.netcom.com) is retired from General Motors where she held management positions in strategic planning, production, materials management and engineering. Before working for General Motors she was employed by MIT and SRI International. She holds a B.S. degree in mathematics and an M.S. degree in statistics from Stanford University, and an S.M. degree in management from MIT. Mary Ann is currently assistant principal violist with the Livingston Symphony Orchestra and violist with the Ypsilanti Symphony Orchestra.

This article was adapted from "Advice From Prospective Employers on Training BS Statisticians," The American Statistician, *February 2001, Vol. 55, No. 1, 14-18. by M. A. Ritter, R. R. Starbuck and R. V. Hogg.*

Jobs For New Graduates

The rest of this article summarizes the answers to these questions about jobs for new statistics graduates. It is based on the opinions of a few knowledgeable people rather than a large survey. The surprising agreement among the contributors lends weight to their advice.

What Jobs Are Out There?

Very few positions were identified as exclusively for new statistics graduates. Instead, new graduates qualify for a number of positions for which statistics is just one of several appropriate backgrounds. The position title most commonly mentioned was (no surprise here!) "statistician" modified by an adjective describing the field of work. Some of these titles were:
—Statistician
—Biostatistcian
—Agricultural statistician
—Survey statistician

—Mathematical statistician
—Automated data processing statistician

Other common positions were:
—Staff (in a consulting organization)
—Programmer
—Analyst

What Do These Statisticians Do?

Employers often describe what an employee is expected to do by listing "key job elements" that are the major responsibilities assigned to the person in the position. These are often used by employers to describe positions at job fairs, in on-campus recruiting notices or newspaper advertisements, or in formal personnel records.

Many companies use terms in a special way inside their company or industry. To find the best employment opportunity, students should understand the specific terms as different employers use them. For example, if a job posting reads "responsible for programming PC's," does this mean programming personal computers in an office environment or programmable controllers on the shop floor? Or does it mean responding to problem communications from suppliers or customers? The way to find out is to ask someone knowledgeable in the industry. Company interviewers, faculty advisors and recent graduates in the same industry are good sources for this information.

The contributors identified key job elements in three categories:
—Statistical (specific statistics theory or methods)
—Technical (mathematical, engineering or computer-related activities) and
—Non-statistical (activities outside statistics methods or theory).

The common elements were:

Statistical

—Apply statistical methods (the specific method varied with the industry)
—Apply statistical theory
—Collect, analyze, interpret data
—Perform general statistical consulting
—Review and diagram processes
—Prepare sampling frames
—Draw samples

Technical

—Write SAS computer programs
—Use databases
—Conduct web-based searches

Non-statistical

—Write reports
—Make presentations
—Participate in teams

What Do Students Need to Know?

Key job elements describe a position. Qualifications describe a candidate or what a candidate needs to know. Usually the list of qualifications describes the "ideal" candidate. Employers often do not find a person who is an exact match, but they attempt to select someone who comes as close as possible on the most important or required areas.

The formal qualifications mentioned by the contributors were
—Bachelor's degree with two to four years' experience
—Major in math, statistics or operations research
—Master's degree strongly preferred (MS with no experience was seen as equivalent to bachelor's with two to four years experience)
—Minor in the field of application (examples: a science, social science or engineering)
—Specific statistics course work (more about this in the next section)
—Communication skills (written and oral)
—Computer programming skills (SAS was mentioned most often)

On these qualifications there was considerable agreement among the contributors despite their widely different experience. Almost all of them mentioned that a statistics degree was only one of several qualifying degrees, that an advanced degree was highly desirable, and that knowledge of the subject area was very important. Students can demonstrate subject matter knowledge by taking additional courses (more than just the introductory course--perhaps even a concentration or a minor) in the area or by summer or co-op employment in the field while still in school. The best solution would be to have both course work and related employment.

This list begs the question about whether a statistics bachelor's degree really is enough to find a first job after graduation. The answer is a definite yes based on the contributors' and author's experience. It also immediately raises the big question for all people looking for their first job: how do I get the experience required for the first job until I've had my first job? This question is not unique to statistics degree holders and the advice section at the end of this article offers some answers.

What Do They REALLY Need to Know?

The formal qualifications listed above are very general but were mentioned by almost everyone. The contributors also offered more detailed descriptions of qualifications, but these tended to be specific to a firm or industry.

The detailed qualifications below (in no particular order) were mentioned by more than one contributor, but not by all. They suggest specific courses, class projects or work study experiences that students might have during college.

Statistical

—Analysis of variance/general linear models
—Simple analysis methods
—Reliability statistics
—Survival statistics
—Variance component analysis
—Variance propagation
—Acceptance sampling
—Exponentially weighted moving average
—Design of experiments
—Nonstandard experimental design
—Graphical analysis (box and whiskers, etc.)
—Statistical process control
—Sampling
—Principles of statistics and variation
—Survey methods and techniques
—Research methods and techniques
—Data collection/handling
—Limitations of methods
—Statistical experience/hands-on work

Technical

—Tolerancing
—Measurement capability analysis
—Calibration
—Statistical package (especially SAS, although S-plus and Minitab were also mentioned)
—Database programming/structure/large database experience
—Mathematics including advanced calculus, linear algebra
—Subject matter knowledge

Non-statistical

—Written communication
—Oral communication
—Work organization
—Consulting (practical experience preferred)
—Meeting participation (agendas, minutes, etc.)
—Team membership/collaboration
—Interpretation of statistics to non-statisticians

That is quite a list! No single student can or will graduate knowing about all of the topics. How can students know what are the most important items for them?

What's Most Important?

To help answer this question, the contributors were asked to rate the qualifications in importance on a scale from 1 to 5, with 5 being the most important. They consistently gave two qualifications the highest importance:

—The statistical methods most often used in their field of application.This was a different method in different fields.

—Communication skills. Written and oral communications were equally important.

Advice

The last question for the contributors was open ended. It asked them to give any comments or advice they felt would be useful to graduating statistics students. Their answers covered a lot of issues both statistical and non-statistical and were offered with great energy and belief. Their advice can be summarized in eight points.

Experience

Get as much experience with applied statistics as is possible while still in school.This may be in the form of class projects, co-op experiences, work-study, or internships. Work in a campus consulting center if at all possible.

Employers often use summer hiring to check out prospective full-time employees. Students can use summer employment in a similar manner by looking for summer work in a field or at a firm that holds long-term interest for them. Doing this also helps overcome the "experience required" hurdle that faces people seeking their first full-time job.

Field of application

Learn about statistics as used in a specific field. Take a minor or introductory course sequence in a field of application such as biology or marketing.

Find out what fields the statistics faculty members have worked in. Ask them which methods are most used in that field. Ask statistics instructors what fields use the methods they are teaching. When visitors come to campus to give talks, attend the talks and notice what methods they use. Ask them what methods are used in their fields.

Acquire experience in the specific field. It is most useful if the experience involves statistics, but it is also useful to acquire any experience in an industry or firm that is of long-term interest. For example, it is valuable experience to have worked in the tooling or production department of a manufacturing plant in a summer or co-op job if a student wishes to seek a position as a quality statistician after graduation.

Teams

Get experience working in project teams. Learn to fill the different roles in teams (member, leader, facilitator, etc.) Take a class in organizational behavior that explores team dynamics. Sign up for classes with term projects assigned to teams. Learn to become comfortable in roles that are not familiar. "Take charge" people should practice facilitating rather than leading; "followers" should volunteer for leadership roles.

Communication

Develop excellent written and oral communication

skills. Take a writing class, then take another! Learn to write long, detailed reports and one-page summaries. Prepare presentations with bullet points. Become comfortable speaking in front of groups. Offer to make the presentation for team projects. Learn to convey statistical information to non-statisticians.

Learning

Plan to continue to learning. Take personal responsibility for continuing to learn after graduation. Any technical field such as statistics is continually growing. To stay useful, students must plan to grow with their field. This can be done after graduation by joining professional societies and attending meetings, taking classes offered at work by their employers, taking evening classes either through distance learning or local colleges, reading professional journals. Find out whether an employer values specific degrees or certification and work towards one. Ask for new assignments on the job that require expanded skills.

Programming

Learn to use a statistics application package and a high level programming language. The most commonly mentioned package was SAS. Each employer usually has a statistics package that is used at the company. Learn to use this package as quickly as possible. Learn to use as much of the capability of application packages as possible. Develop a careful, error-free approach to programming.

Data

Develop good data collection and management skills, including the use of a database management program. Work with the largest data sets available. While still in school, work on projects that require the design of an analysis plan and the collection of the data to support it and the construction of the database from the collected data. Learn data quality assurance and documentation methods. Learn to move data between application packages. If possible, take a class in data base management to supplement hands-on experience constructing and maintaining databases.

Graphical methods

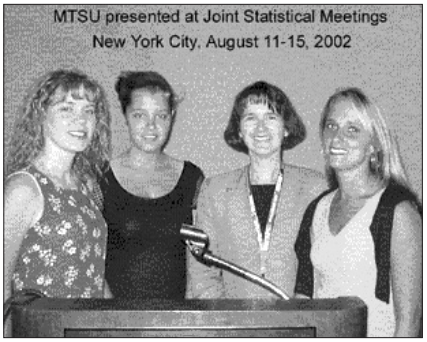
Learn to think about and explain statistical analyses using graphical methods. Learn to use the graphing routines of statistics application packages. Learn to present an analysis using only graphical methods. Learn as many graphical forms as possible, their appropriate use and their limitations. Incorporate graphical summaries in any analysis prepared for class, projects or work-study, even if this is not a specific requirement.

This advice is not significantly different from advice offered in previous articles. What may be different is the increased emphasis on computing and database skills and the very heavy emphasis on the need for real-world experience and the non-statistical skills of communication and team participation. Without these skills and experiences in their toolkits, new statistics graduates will not be competitive for employment, regardless of the field in which they seek employment or the statistical methods they have learned.

Summary

The undergraduate statistics degree has received relatively little emphasis at most institutions compared to statistics graduate degrees. However, it is a degree with great potential and one that offers students much flexibility in their career choices. The advice offered in this article should make the job-seeking process a little easier for students graduating with undergraduate degrees in statistics.

Student Perspectives from the Joint Statistical Meetings



Leann Hooge, Keeley Hutchison, Ileah McKee, and Venita DePuy

Foreword by Ginger Holmes Rowell

Last August I had the privilege of taking three students to attend the Joint Statistical Meetings (JSM) in New York City. If you are a student and would like to go to the JSM now is a good time to start planning. Next summer, the JSM will be held in San Francisco, California, August 3-7. The JSM offers several important opportunities for students to learn more about careers related to statistics. You can attend sessions related to your potential career, or you can visit the exhibit booths set up by statistical companies ranging from the Center for Disease Control to insurance companies. If you are actively searching for a job, then you might want to participate in the Career Placement Service. Prospective employers and employees can review job openings and resumes as well as set up interviews. My students were all very glad for the opportunity to attend the JSM. It was important for their education as students interested in math and statistics, and for their exposure to a world of statistics that previously they did not even know existed. Traditionally, few undergraduate and graduate students participate in the JSM; however, with so many sessions on different statistical topics, ways to learn more about statistical careers, and opportunities for networking there really is something for every student interested in attending the JSM.

Before you go to the JSM meeting, you can study an online version of the conference program. This can help you sort through the many different types of sessions you can attend. A teacher might be able to suggest appropriate level sessions to match your interests and your background knowledge. There are social opportunities at the conference as well. Each year there is a student mixer in the evening. This is a good opportunity to meet other students attending the conference, and you might even win a door prize. Another social activity is the annual informal dance party. And if you want to learn more about the area where the conference is located, you can usually take a tour with other JSM attendees.

Attending such a conference can be expensive, especially if the location is far away and if you want to participate in the placement program and tours. The JSM tries to make attending the conference more affordable for students. For JSM 2003, the reduced-rate registration fee is $50 for students and $60 for K-12 teachers regardless of whether you register early or at the conference. Many times, the college or university you attend might provide funds for the registration fee and/or travel expenses. Check with your appropriate academic departments and with your school Student Government Association for possible ways to help fund student travel to academic conferences. Guidelines will differ at different schools, but generally, you may have to complete an application for the travel funds and write a report of your experience. Be certain to ask in advance what receipts you need to retain for reimbursement. Also, there are usually some student accommodation rates at the hotels, and the JSM provides a forum to help you locate another JSM attendee to share a hotel room and related expenses with, as an additional way to cut costs.

Students interested in learning more about attending the conference in San Francisco can log onto *http://www.amstat.org/meetings/jsm/2003*.

My students were all very glad for the opportunity to attend the JSM. It was important for their education as students interested in math and statistics, and for their exposure to a world of statistics that previously they did not even know existed. Traditionally, few undergraduate and graduate students participate in the

Ginger Holmes Rowell is an Associate Professor of Mathematics at Middle Tennessee State University. She primarily teaches courses in statistics and looks for opportunities to involve students in statistical projects and research. She is interested in assessing technology tools used to teach statistics and is working on a project that has been supported by NASA's Marshall Space Flight Center, Engineering Cost Office and MTSU to develop web-based regression tutorials.

JSM; however, with so many sessions on different statistical topics, ways to learn more about statistical careers, and opportunities for networking there really is something for every student interested in attending the JSM.

The vignettes below were written by students attending JSM 2002, as they reflected on their experiences:

An Undergraduate Perspective, by Ileah McKee

Ileah is a senior undergraduate math major at Middle Tennessee State University. She plans to continue her education to earn a Ph.D. in mathematics. She would like to teach mathematics on the college level. Ileah is minoring in foreign languages and spent much of her summer in France. One of Ileah's many hobbies is scuba diving.

A First Year Graduate Student Perspective, by Keeley Hutchison

Keeley is a graduate teaching assistant at MTSU where she teaches college algebra and helps students with questions and homework in the mathematics laboratory. In May 2003, she will receive her Masters degree in applied mathematics with an emphasis in statistics. Following graduation, she plans to teach either onat? the high school or collegiate level. Keeley is active in her church and community during her spare time.

A High School Teacher Perspective, by Leann Hooge

Leann is a mathematics teacher at Hume-Fogg Magnet High School in Nashville, Tennessee where she has taught for five years. Last year she took a one-year leave of absence from public school teaching to finish her Master of Science in Teaching, with a concentration in secondary mathematics education at Middle Tennessee State University. She graduated in August and returned to teaching, tutoring and coaching an award winning cheerleading squad.

An Experienced Graduate Student Perspective, by Venita DePuy

need Venita's bio sketch

An Undergraduate Perspective

by Ileah McKee

I am a senior undergraduate math major at MTSU and plan to continue my education and earn a Ph.D. in mathematics so that I can teach mathematics at the college level. Attending the New York City JSM was a very enlightening experience. My views of the professional world of statistics were opened dramatically since my prior statistics experience had primarily been in the undergraduate classroom at the undergraduate level.

We began the conference by registering and receiving an assortment of materials such as the conference program, which lists the events with times and locations, and a book containing the abstracts of the presentations and panel discussions. On the first day, Dr. Rowell selected a session for us to attend to give us an understanding of how the meeting was to progress. Each of us then referred to our programs and abstract guides to plan the rest of our schedules to suit our individual interests. There were many choices among the sessions that were offered. An attendee could not only hear lectures on a wide range of interesting topics but could also see numerous poster presentations showcasing a variety of statistical applications as well as exhibit booths with employment information.

Many interesting topics caught my eye, but since the lectures overlapped, it was sometimes a difficult task to decide which session to attend. I could choose from approximately 29 different sessions conducted simultaneously from 8:30 a.m. until around 6 p.m. each day along with opportunities to attend luncheon lectures for a fee. There were even special events in the evenings. The subjects of the numerous lectures included many topics like biomedical warfare and air pollution, education and assessment, census taking, insurance statistics and economics, and even medicine and genetics. I was surprised to see so many different areas of expertise where statisticians provided the cornerstone of the work.

The atmosphere of the conference varied depending on the lecture attended. Some lectures were advanced such as Bayesian analysis and Voronoi tessellation. Others focused on the teaching and learning of statistics. Since I had been working with Dr. Rowell on a project using statistical teaching tools available on the web, I made it a point to attend several lectures regarding education, statistics, web-based teaching methods, computer-based education, and instructional technology, to name a few.

All of the lectures I attended were interesting. I will mention a few that were particularly applicable to my interest. The lecture by Andrew Gelman was based on his and Deborah Nolan's book "Bag of Tricks for Teaching Statistics." He was a very energetic and organized speaker. He not only had his presentation projected from a computer slide show onto a wide screen, but he also distributed printed copies of his slides and excerpts from his book. Another interesting presentation I attended was given by Alexander Kugeshev. He demonstrated the use of his website for teaching statistics and gave some thoughts on how to enhance the web-based statistical concepts with teaching techniques in the classroom.

I was glad the conference gave me an opportunity to experience such insightful lectures on similar topics

from very different perspectives. Overall, I feel that my experience at this year's JSM was a complete success. I gained some very helpful hints for my future as a teacher; however, should my career plans change, I now have seen other interesting statistics-related disciplines and can make more informed career and educational choices.

A First Year Graduate Student Perspective

by Keeley Hutchison

I am currently a graduate teaching assistant at MTSU where I teach college algebra and work in the mathematics laboratory. In May 2003, I will receive my Master's degree in applied mathematics with an emphasis in statistics. I am planning on teaching when I graduate, but I have not yet decided whether I want to teach at the secondary or collegiate level.

This conference opened my eyes to many different ways that statistics is used in the "real world." The sessions that I chose to attend broadened my understanding of what to do with statistics. One of my favorite sessions was "Data Visualization in the Media: Infographics at *The New York Times*." This session, which lasted for an hour and a half, emphasized incorporating statistics with the design and graphics of *The New York Times*. Most statisticians at *The New York Times* are involved in producing and editing the editorial page whereas only a few statisticians are involved in the graphic design department. Pressed for time, the graphic designers are presented with a wide variety of data, and they have to decide how to present it in an appealing and eye-catching way that still communicates the information clearly. It was very interesting to learn all that is involved in displaying statistical graphics in *The New York Times*.

Another session I attended was "The Use of Web-Based Methods to Broaden the Reach of Statistics." This session was very applicable to my interest in teaching math and statistics. One of the speakers spoke on the use of Java applets for Internet-based statistical computing. Java applets are small computer programs that can add a dynamic component to instruction on introductory statistical concepts that otherwise might be hard to teach. This session was enjoyable because I was exposed to ideas for teaching statistics in the classroom using the Internet. Another speaker presented ways to use the Internet to maintain constant communication with students in the class. Ideas presented encouraged faculty members to use the Internet to post grades, to use anonymous e-mail for student comments, and to use software tutorials. This session was very beneficial to me and my future plans for teaching.

Another aspect of the JSM was the Exhibits. This part of the conference was a little overwhelming at first. The many booths representing different companies were not particularly geared toward students. However, with close inspection one could find beneficial information about different careers that use statistics.

Attending the Joint Statistical Meeting is definitely a great learning experience. Even though this was the first meeting I attended, I plan to return to the JSM next year.

A High School Teacher Perspective

by Leann Hooge

I am currently a mathematics teacher at Hume-Fogg Magnet High School in Nashville, Tennessee. Taking a one-year leave of absence from teaching, I was able to attend MTSU full-time and graduated in August with my Master of Science in Teaching degree, with a concentration in secondary mathematics education. I have taught five years in secondary public education and taught college algebra and mathematics for elementary educators as a graduate teaching assistant at MTSU for one year.

As an educator, I viewed attending this conference as a professional development activity, and I hoped the topics discussed and information I learned would be applicable to my teaching. I was not disappointed. The JSM had a wide variety of options for educators at both the high school and collegiate levels. Even though the majority of the sessions were not focused on education, a selective number of workshops did focus upon this topic. Education related mini-seminars included using technological resources in the classroom, the changing teaching principles, new trends and current issues in education at all levels of instruction, and the future of statistics education.

I was quite impressed by the plethora of diverse workshops and the fast-paced world of presentations. Conference presentations varied in length and format. In many of the sessions I attended, as many as five presenters individually demonstrated their knowledge on a specific topic, each taking a slightly different angle on the same subject. Each person received a short amount of time (about 12 minutes) to develop their topic but was allowed to elaborate later during the open discussion time at the end of the session. This manner of presentation seemed to restrict some presenters but at the same time allowed a number of people to present their current research. This format was beneficial to me because I was able to hear several topics with which I was not previously familiar. From the sessions, one could learn about the predictions of Wall Street, how to use statistics to detect diseases such as the West Nile Virus, the statistical significance within sports, and various other topics appealing to students.

If these options were not enough for an educator, there were numerous other topics to be discovered. I was astonished by the vast number of topics with which

I was not familiar. However, with the open door policy, I was free to find the sessions that best suited my educational needs and found the New York City JSM conference to be a delight!

An Experienced Graduate Student Perspective

by Venita DePuy

The story of my trip to the 2002 Joint Statistical Meetings actually begins a few days before. Like the rest of my class at North Carolina State University, I had studied all summer for the Master's Exam, which was Thursday, August 8. So when I boarded my flight for New York City at noon on Friday, my mind had been everywhere but on the conference. While the majority of my classmates were sweating their way through the Ph.D. qualifier on Friday, I was packing and trying to figure out how to make the most out of my first visit to the JSM. Since I plan to graduate in May, I was especially looking forward to networking and meeting with various companies to lay the groundwork for next spring.

A few months earlier, I had answered an advertisement in the *Amstat News* for student volunteers to help with the Continuing Education classes before and during the JSM. In return, volunteers could attend a class of equal length for free. My sister and brother-in-law live near Times Square, so the expense of an extra night's stay was not an issue for me. After looking through the variety of classes offered, I decided to take the Applied Spatial Statistics class on Sunday and to volunteer to help with a class the day before. I chose those two days because I wanted to leave myself free to see as much as possible of the conference, which didn't really get into full swing until Monday.

I began my first morning by waking up at 6:45 on Saturday, questioning the wisdom of volunteering to help with the Continuing Education classes at that early hour. The morning started with a quick orientation; then we moved downstairs to begin checking people in. The hour until the class started was a frantic rush of signing in participants, distributing handouts and textbooks, and answering questions. Once the class began, things calmed down to the extent that I wished I had brought something to read, to take my mind off debating why the hotel air conditioning was set to 'Antarctica.' My co-monitor and I alternated sitting in on the regression methods class and, in the meantime, we or I? got to know the other class monitors.

Through these conversations I learned about the Survey Methodology curriculum in Maryland and undergraduate research opportunities in Boston. During breaks, I also had the opportunity to learn what statisticians do at the National Cancer Institute and that my department at NC State hasn't really changed too much in recent years, based on the reminiscences of a graduate from the late 90s.

I also learned that name recognition can play a big part in getting to know statisticians. I can't remember how many people said, "Oh, you go to NC State? Do you know such-and-such person?" In spite of the strenuous curriculum in my program, I learned that it's worth it to go to such an excellent school. I can tell a statistician who I've never met before that I've taken Tsiatis' clinical trials class (at NC State), and he will recognize the name and continue to tell me in greater detail about his own work, while relating it to a recent paper published by Dr. Tsiatis.

My first day in the Big Apple was made even more complete when, while walking the few blocks from the hotels to my sister's apartment, I discovered a street fair taking up ten blocks of Times Square. Looking down ten blocks of solid people was a little intimidating, but the shopping was a blast – everything from $1 Thai food (very good chicken satay) to jewelry to pottery. After spending an hour or two there, I was more than happy to wrap up the day with a manicure and pedicure.

Sunday started off early with my spatial statistics class at 8:15. It was a very informative class, and the number one thing I learned was how little I know! The teacher was quite good and combined his PowerPoint presentation with demonstrations of SAS, SPSS, and GIS. During lunch, I was lucky enough to have the chance to ask the instructor and another student about environmental and wildlife statistics. They encouraged me to become active in the Wildlife Society if I wanted to pursue a career in wildlife statistics, and they provided some insight into the availability of jobs.

During the afternoon break in the class, I struck up a conversation with an industrial statistician who works for a large food company, and I learned a little about what she does. She also told me about how she had gone through a recruiting firm when obtaining her job and how happy she was with them. She talked about how they had filled her in on the background of the company and the people she would meet, got her the schedule for the interview, and told her answers to many questions that she might not have thought to ask. I also (during my morning break) spoke with a gentleman about the National Institute of Standards and Technology, where he was employed, and also learned a little about the insurance industry, where he had worked previously.

After the class was over, I went to the Career Placement Service, which I had registered for online. I was given another name tag, and I sat down to look through books of employers' job openings. There was a computer system that allowed employers and potential employees to contact each other easily – just typing someone's ID number in a special field caused their name to pop up. Although there were not enough computers during busy times, it was a very useful system.

I found my way to the informal reception spon-

sored by the Caucus for Women in Statistics and learned a little more about that organization over hors d'oeuvres. The party was open to everyone, and I'm sure the few men that were there enjoyed themselves! The speakers shared good tips about getting involved and making the most of the meetings, and they were very welcoming. After the reception, I went to the opening mixer. It was rather large and intimidating until I found a faculty member from my department and then some students that I knew. It was a little daunting seeing so many statisticians in one room!

I was a little anxious anticipating the New York City crowds at 8am Monday morning, but they turned out to be minimal. I started out the day listening to a talk by a Ph.D. student from NC State. I followed that up with breakfast with a researcher who works for the federal government in DC, and I learned about using reverse propensity scores to account for biases in clinical trials. I filled the rest of my morning with interviews with various organizations and then had lunch with a street corner vendor. I squeezed in a few minutes during the lunch hour to look at the posters on display. I enjoyed seeing the exhibitors' booths after lunch, where I got free samples of software, looked at publishers' discounts to find the course textbooks I'd need for fall semester, nibbled on free candy, and picked up enough pens to last me through my second year of grad school. The best freebies (in my opinion) were the squeezable pill – like a stress relief ball – from Eli Lilly and the retractable phone cords from Pfizer.

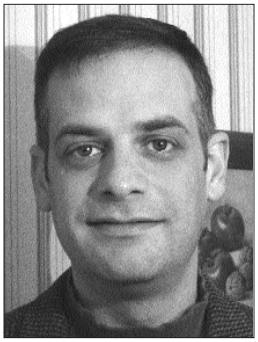
On Monday evening, I didn't have time for all the receptions I wanted to attend. I was unable to make it to the reception for the Caucus for Women in Statistics and the student mixer because I attended two of the pharmaceutical companies' receptions. The food was good, and I enjoyed making contacts. Students who had graduated from my department were kind enough to introduce me to people at both companies. I learned about some of the differences between clinical trials and pre-clinical studies; the latter seemed to be less regulated, as they deal with cells or animals, and to involve more statistical analyses and less red tape. All in all, it was an exhausting day.

Tuesday was a slight bit calmer. I started out the day by checking in at the Placement Service, happy to find a few more messages requesting or confirming interviews. I checked the binders of available jobs and employers, and I was pleased to see that they were still putting more in the binder every evening. After listening to a few speakers, I went by the exhibit booths for a late "breakfast" of free chocolate before my 11am interview, where I learned that insurance companies also hire programmers and biostatistical analysts in addition to actuaries. I had an incredible lunch at the Hallo Berlin pushcart, followed by another interview before I listened to more speakers in the afternoon. Another interview kept me busy until the NC State reception, after which several of us took the subway to Little Italy for an excellent dinner.

By the time Wednesday morning rolled around, I was getting worn out! Things were beginning to slow down a little. I'd settled into my routine of scheduling the speakers I most wanted to see, around which I checked with the Placement Service, met with people, and tried to absorb everything I could. I finally managed to fit in a little shopping before an evening meeting. By the time I flew home on Thursday, I was exhausted.

Since the JSM, I've made a point to follow up on job opportunities and personal contacts. One of those has led to a data set I'm analyzing for a class project. All in all, attending the JSM was a wonderful experience that I would highly recommend to any student.

It□s Back! ASA Stat Bowl to Make Return at JSM in 2003



Mark Payton

I'm pleased to announce the reincarnation of a popular event at the Joint Statistical Meetings (JSM). The ASA Stat Bowl (formerly known as "College Bowl") will make its return in San Francisco in 2003. Preparations are being made for the event with some exciting new aspects.

The biggest change? MONEY! ASA will be reimbursing students who compete in the Bowl to the tune of $500 for travel and registration. So if finances were keeping you at home at JSM time, here's your chance to help fund your way to the meetings.

Another big change is the elimination of teams. The College Bowls of the past were like most academic team competitions. (Recent results appear in Table 1.) Eight teams of four students represented university statistics programs in a single-elimination tournament. For the new version of the competition, individual players will compete and represent their university. So you no longer have to round up three teammates to compete in the Bowl. Multiple players from a single university can compete, and though individuals will be playing, team points will be awarded based upon individual performance. Think of the collegiate golf championship, where individual performances determine team standings. Individual awards as well as a team award will be presented at the conclusion of the event.

Mark Payton is

Students will be accepted into the tournament on a first come, first in basis. Notification of a willingness to participate will serve as entry. Inquires about the Bowl or requests to be registered as a contestant can be made to Mark Payton, Oklahoma State University, mpayton@okstate.edu. A maximum of sixteen players will be allowed in the contest. In the event that the field of contestants fills to capacity, each university will be restricted to two players to assure diversity. A waiting list will be established to fill unexpected vacancies should they occur at game time.

The contest will be held on the Tuesday of JSM in two sessions. Session 1 will consist of four games, each with four contestants. The winners of these four games plus two at large contestants will advance to Session 2. Players who score the most points in Session 1 without winning their game will be the at large winners. Session 2 will consist of the six players who have advanced, playing two games each with three players, and the two winners meeting head-to-head in a championship game. The question format for the ASA Stat Bowl has not been finalized as of yet, but the questions will focus on the ASA and on statistical history and methodology. (Sample questions appear in Table 2.)

Any student interested in playing should contact us ASAP before the player positions are filled. We encourage all players to register before July 1, 2003, but players will be allowed up to game time provided space is available. Hope to see you in San Francisco!

Table 1: A Brief History of Past Statistics Academic Bowls

Year	Event	Location	Champion	Runner Up
1992	ASA Winter Conf.	Louisville, KY	Bowling Green	Virginia Tech
1994	ASA Winter Conf.	Atlanta, GA	Iowa State	Florida
1995	ASA Winter Conf.	Raleigh, NC	Nebraska	Bowling Green
1996	JSM	Chicago, IL	Iowa State	Chicago
1997	JSM	Anaheim, CA	Iowa	UC-Santa Barbara
1998	JSM	Dallas, TX	Iowa	Texas A&M
1999	JSM	Baltimore, MD	Florida	Maryland

Table 2: Sample Questions (with Answers Below)

1. If 5 cards are drawn at random from a standard deck of 52, what is the probability that the last card drawn is a diamond?
2. Which Sesame Street character is undoubtedly a closet statistician?
3. What major statistician made Rothamstead famous?
4. When the sampling distribution of the estimator is insensitive to changes in the distribution of the population, we say the estimator is...what?
5. What is the value of the third moment of a standard normal distribution?
6. Give the name of the international society devoted to the mathematical and statistical aspects of biology.

(Answers: 1. 1/4 2. The Count 3. R.A. Fisher 4. Robust 5. 0 6. Biometric Society)

Solutions to Data Sleuth Mystery

Question #1: From the boxplots it appears that the smokers have, in general, higher FEV scores. The median FEV for the smokers is over 3 liters while the median FEV for the nonsmokers is less than 2.5 liters.

Question #2: In any observational study in which we are looking to assess the relationship between a self-selected category (smoking), and some other measure (FEV), we should always be aware of other confounding reasons that might help clarify and/or explain the apparent relationship. In this case, since these data include so many very young children who do not smoke and whose lung capacity are unlikely to be as large as the older children, we should consider accounting for the age of each subject. In the actual dataset from Rosner, the subjects' height and gender are also included and make for an even more complete and interesting story.

Question #3: For 16-year-old nonsmokers the estimated average FEV is just under four liters, say 3.8 liters, while smokers have an average FEV of just over three liters, say 3.2 liters. Similarly, for 19-year-old nonsmokers the estimated average FEV is over four liters, say 4.5 liters, while smokers have an estimated average FEV of about 3.5 liters. Finally, for 10-year-old nonsmokers the estimated average FEV is under three liters, say 2.8 liters, while smokers have an estimated average FEV of about 2.5 liters. Though the standard errors of these estimates have yet to be expressed, these data do indicate that the older subjects who do not smoke have, on average, larger FEV than the smokers. The inconsistency arises when considering the youngest smokers since the estimates suggest that for 10-year-olds smoking is associated with larger FEV. This has a variety of possible explanations, some of which include the very small number of young smokers, the short time and relatively small amount that these children are likely to have smoked and errors in self-reporting.

Unseen, Unfelt, and Understated

The Dangers Posed To Children By The Use Of Arsenic-Treated Lumber In Playgrounds

Katherine D. Van Schaik

Background

In Tallahassee, Florida, the Environmental Working Group and the Healthy Building Network recently demanded that the Florida government immediately ban the use of lumber that is pressure-treated with a preservative called chromated copper arsenate. On May 23, 2001, these two groups, in conjunction with many other environmental groups, released a study that enumerated the risks posed to children due to exposure to lumber that is treated with preservatives containing arsenic ("Coalition: Ban treated wood," 2001). Chromated copper arsenate, or CCA, is commonly used to treat lumber to prevent decay due to insects and fungi. The uses of CCA-treated lumber are many; it's used for everything from bridges to playgrounds to picnic tables.

The preservative CCA is a mixture of chromium trioxide (CrO_3), copper oxide (CuO), and arsenic pentoxide (As_2O_5). The arsenic is a pesticide, the copper is a fungicide, and the chromium fixes the arsenic and the copper to the wood.

Lumber is treated with CCA according to its use. The more a piece of wood is exposed to the ground and to the elements, the more preservative is impregnated into the wood. For example, lumber that has no contact with the ground contains 0.25 pounds of preservative per cubic foot, lumber that contains 0.40 pounds of preservative per cubic foot is used to build playgrounds and picnic tables, and lumber that is immersed in salt water contains 2.50 pounds of preservative per cubic foot (Florida Hazardous Waste and Waste Management Department, 2000).

D.E. Stilwell and K.D. Gorny of the Connecticut Agricultural Experiment Station determined that the arsenic in CCA-treated wood leaches out of the lumber into the surrounding environment (1997). They collected 85 soil samples from below seven decks, aged four months to 15 years. Concentrations of arsenic in the soil ranged from 3 mg/kg to 350 mg/kg. The mean arsenic concentration was 76 mg/kg.

The chemical profile of arsenic from the U.S. Environmental Protection Agency indicates that chronic exposure to small amounts of arsenic may cause decreased blood cell production and nerve damage. There is sufficient evidence that inorganic arsenic compounds are skin and lung carcinogens in humans (USEPA, 1987). Direct skin contact with inorganic arsenic compounds can cause swelling, redness, and irritation (USDHHS Toxicological Profile for Arsenic, 2000).

Minimal Risk Levels (MRLs) are developed by the Agency for Toxic Substances and Disease Registry (ATSDR) to establish "an estimate of the daily human exposure to a hazardous substance that is likely to be without appreciable risk of adverse noncancer health effects over a specified duration of exposure" (ATSDR, Minimal Risk Levels, 2001). MRLs exist for acute (1–14 days), intermediate (>14–365 days), and chronic (>365 days) exposure through inhalation and oral exposure routes. As of December 2001, a MRL for the dermal route of exposure had not been identified because the ATSDR was unable to find a suitable method for deriving a dermal MRL. The MRL for acute oral exposure to arsenic is 0.005 mg/kg/day, and the MRL for chronic oral exposure to arsenic is 0.0003 mg/kg/day.

MRLs are especially applicable to sensitive individuals, such as children (ATSDR, Minimal Risk Levels, 2001). Children are at a greater risk for arsenic poisoning than adults because they are more likely to ingest soils that contain arsenic. Information also suggests that

Katherine D. Van Schaik is a junior in the Discovery Math and Science Magnet Program at Spring Valley High School in Columbia, South Carolina. Last school year, she presented her science research project at the South Carolina Junior Academy of Science Spring Meeting, the National Junior Science and Humanities Symposium in San Diego, and the Intel ISEF in Louisville, KY. At the Intel ISEF, she received awards for first place, Best Use of Statistics, first place in the category Environmental Science from the U.S. Air Force, and second place overall in the category Environmental Science. She also enjoys tennis, fishing, volunteer work, and church activities.

children are less efficient than adults at internally converting inorganic arsenic into less harmful organic arsenic (Toxicological Profile for Arsenic, 2000).

The purpose of this research was to determine the effect of exposure time and concentration of arsenic on the amount of the preservative absorbed into skin. It is possible that children who are playing on playground structures and picnic tables constructed with CCA-treated lumber are absorbing arsenic into their skin. Children exposed to CCA-treated wood for long periods of time could absorb enough arsenic through their skin to be potentially hazardous to their health. Chicken skin was used to simulate human skin because the structural arrangements of its epidermis, dermis, and collagen fibers are similar to that of a young human child (Dr. Glenda George, personal communication, September 3, 2001). I hypothesized that the higher the concentration of arsenic and the longer the time of exposure, the greater the amount of arsenic that would be absorbed by the chicken skin.

Method

I obtained fresh chicken skins from Amick Farms in Batesburg, South Carolina, and I cut the skins into 66 squares approximately 2 cm by 2 cm. Next, I obtained sheets of glass and ten petri dishes and rinsed them with distilled water and 10% nitric acid. I purchased the pieces of lumber with preservative concentrations of .25, .40, and 2.50 pounds per cubic foot, and I sawed them into pieces approximately 2.5 cm by 3.5 cm. Finally, I collected bricks and broke them so that each piece weighed approximately 3.0 pounds, or 1.35 kg. This weight was chosen to simulate the weight of a small child sitting on a playground structure or picnic table: 3 pounds of brick pressing on a 2.5 cm x 3.5 cm wood surface with chicken skin underneath the wood is proportional to a 40 pound child with 18 in^2 of bare skin exposed to the wood.

Each trial consisted of an exposure duration (2, 8, or 12 hours) and a preservative concentration (.25, .40, or 2.50 pounds per cubic foot). Seven repetitions were conducted for each of the 9 (3 x 3) treatments. For each time duration, I used 22 squares of chicken skin. I placed one square (the control) into a petri dish, and I put the other twenty-one squares on top of the glass. I put seven of the pieces of the 0.25 treated wood, the 0.40 treated wood, and the 2.50 treated wood on top of the twenty-one squares of chicken skin. I placed the bricks on top of each piece of wood (see Figure 1).

I designed this setup to simulate a small child, aged 2–6, sitting on a picnic bench. The chicken skin simulated the child's skin, and the squares of wood simulated the picnic bench. The brick added weight that is equal to that of a small child, as distributed evenly over a surface area of 18 in^2, which is the approximate area of a small child's thighs. The glass upon which the chicken skins were placed was a smooth, clean, unreactive surface.

I removed the bricks, chicken skins, and wood after the indicated amount of time elapsed. I then placed the chicken skins into three separate petri dishes according to the concentration to which they were exposed and labeled the dishes.

I acid-digested the chicken skin samples with a medium that was 25% nitric acid, 25% sulfuric acid, and 50% distilled water. I further digested the samples with 30% hydrogen peroxide. The digested samples were then analyzed for arsenic with a Perkin-Elmer Atomic Absorption Spectrophotometer.

Results

The experimental data partially supported my hypothesis. While the amount of arsenic absorbed by the chicken skin did increase as time of exposure increased, increases in preservative concentration did not significantly increase arsenic absorption. The scatterplots in Figures 2, 3, and 4 display the arsenic absorption amounts vs. duration times for .25 wood (Figure 2), .40 wood (Figure 3), and 2.50 wood (Figure 4). The mean arsenic absorption values for the exposure times and the amounts of preservative are illustrated in Table 2; the standard deviations are reported in parentheses.

I used two-way analysis of variance (ANOVA), conducted with the Minitab software package, to analyze the data. The results appear in Table 3. The null hypothesis that the chicken skin would not absorb more arsenic over increased periods of time was rejected. The F-statistic was 23.74, and the p-value was less than .001. The effect of the level of preservative concentration was not significant (F = 2.17, p = .124), and the effect of an interaction term was also not significant (F = 0.84, p =

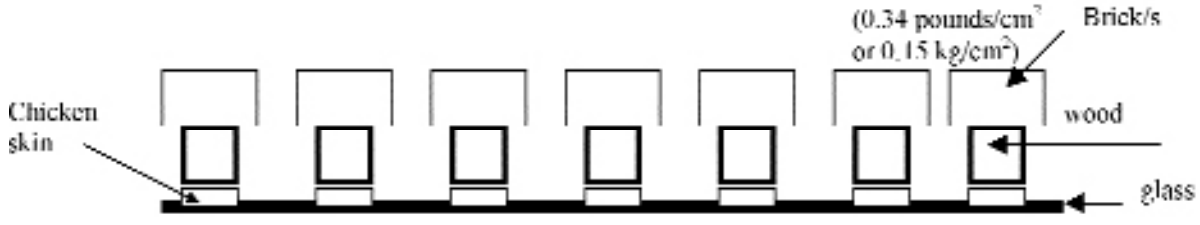


Figure 1: Diagram of Experimental Setup

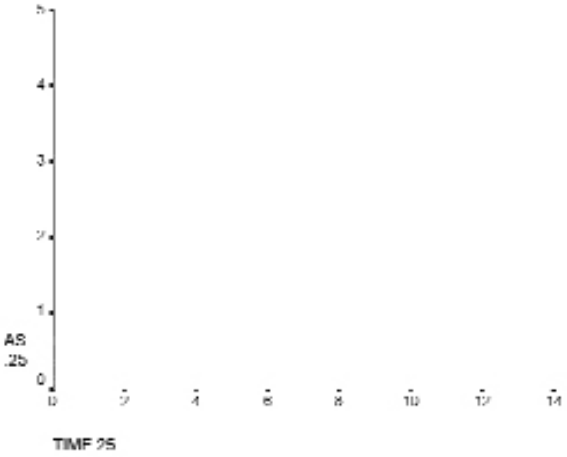


Figure 2: Scatterplot of Absorption vs. Duration for 0.25 Wood

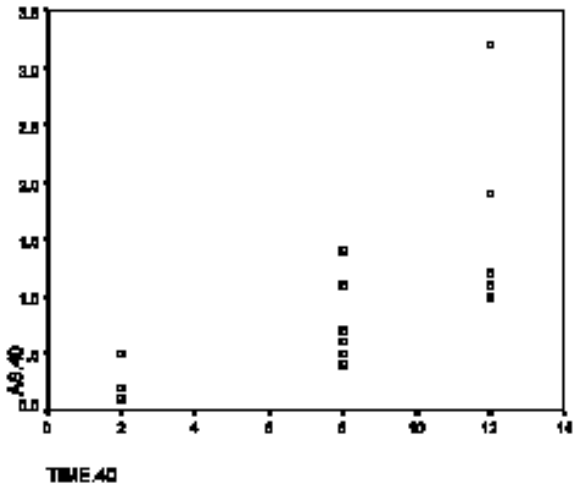


Figure 3: Scatterplot of Absorption vs. Duration for 0.40 Wood

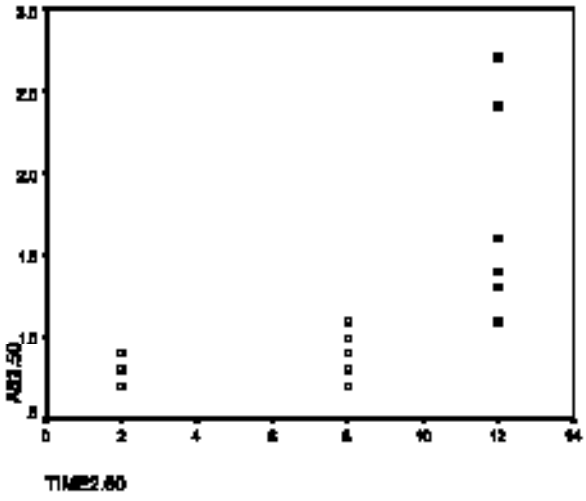


Figure 4: Scatterplot of Absorption vs. Duration for 2.50 Wood

Table 2: Mean (and standard deviation) of the amount of arsenic absorbed by skin (ug/cm^2)

Time of Exposure	Preservative Concentration (n_i = 7) 0.25	0.40	2.50
2 hours	0.27 (.20)	0.19 (.15)	0.81 (.09)
8 hours	1.2 (.59)	0.74 (.37)	1.53 (.93)
12 hours	2.04 (1.48)	1.53 (.80)	1.84 (.64)

.505).

Figure 5 displays an interaction plot. All of the lines are approximately parallel and the differences between the levels of the dependent variable, amount of arsenic absorbed, are roughly the same distance for each value of the significant independent variable, time of exposure. In the graph we see that time of exposure had a significant impact on the amount of arsenic absorbed.

Because time of exposure was statistically significant, I used a Pearson Product Moment Correlation Coefficient (r) to determine the strength of the relationship between the time of exposure and the amount of arsenic absorbed by the chicken. The correlation for the 0.25 treated wood was $r = 0.643$, with a two-sided p-value of 0.002. For the 0.40 treated wood, $r = 0.740$ with $p < 0.001$. For the 2.50 treated wood, $r = 0.659$ with $p = 0.001$.

Extension

D.E. Stilwell (1998) used a wipe test with a polyester cloth to determine the amount of arsenic picked up by the cloth after exposure to Type C 0.40 treated wood, the same type of wood used in this experiment. He placed a polyester cloth under a cement block, and pulled this structure across the surface of the CCA-treated wood five times. The amount of arsenic dislodged onto the polyester cloth is shown below. I extrapolated his values, as shown on the right, since my values are given in ug As/cm^2, and his values are given in ug/100 cm^2.

These results are fairly consistent with my mean values of 0.19, 0.74, and 1.53 micrograms As/cm^2 for each of the exposure levels for the 0.40 treated wood.

Table 5 shows the values after they were converted

Table 3: Two-Way ANOVA Summary

Source	DF	SS	MD	F	p
Exposure time	2	20.370	10.185	23.74	< 0.001
Preservative Conc	2	1.864	0.932	2.17	0.124
Interaction	4	1.444	0.361	0.84	0.505
Error	54	23.171	0.429		

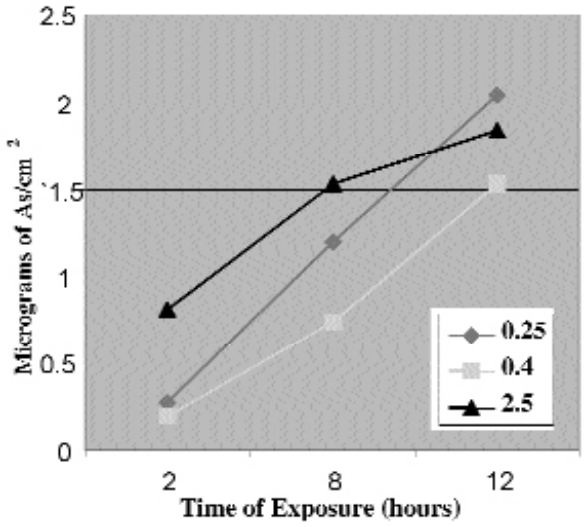


Figure 5: Interaction Plot.

from ug As/cm^2 into mg As/kg of the child's body weight. This allowed me to compare my results to the Minimal Risk Levels for both acute and chromic oral exposure to arsenic.

With the exception of two, all of the values shown are greater than or equal to the minimal risk level for acute oral exposure to arsenic (0.005 mg/kg/day). All of the values are greater the minimal risk level for chronic oral exposure to arsenic (0.0003 mg/kg/day). This indicates that the skin absorbed arsenic in quantities that exceeded the threshold at which the effects of arsenic exposure begin to occur in sensitive individuals.

Discussion

The purpose of this research was to determine the effect of time of exposure and concentration of the CCA preservative on the amount of arsenic absorbed into chicken skin. I found that as time of exposure increased, the amount of arsenic absorbed by the skin increased. For the 0.25 wood, for example, the mean amount of arsenic absorbed by the skin increased from 0.27 micrograms As/cm^2 (2 hours), to 1.2 micrograms As/cm^2 (8 hours), to 2.04 micrograms As/cm^2 (12 hours). The data supported the above hypothesis.

The data did not support the hypothesis that the greater the amount of preservative, the greater the arsenic absorption, but this could be due to the small sample sizes (7) failing to detect an effect that may actually be present. Also, several uncontrollable variables could account for this. Since I purchased the wood from two different states, the treatment process could have been different. (I could not buy the 2.50 and 0.40 wood in my hometown of Columbia, SC, so I bought them from a treatment facility in Florida.) Also, some of the wood could have been older than other pieces, or could have been exposed to more weathering.

Science Fair Excitement

When my pre-Research teacher, Dr. Glenda George, told me and the rest of my classmates in the Discovery Math and Science Magnet Program that we would need to finalize our research project ideas before the end of the summer, I was nervous but excited at the thought of such a challenge. So, in the summer between my freshman and sophomore year of high school, I spoke with Dr. George and found a topic that both worried me and intrigued me: arsenic-treated wood was being used to build playgrounds.

Over the next six months, I wrote my paper, developed my experimental design, and carried out the experimentation and analysis. At this point, I came to what my Research teacher, Mrs. Marilyn Senneway, called the "meat" of the entire research paper: the statistics. After several attempts at finding a way to input the data so as to yield results that would appropriately reflect the data, I finally found a way, and, in doing so, learned the "significance" of statistics in research. Mrs. Senneway had continually told us that, without statistics, we would have no way of knowing what our results really were. After looking at the Two-Way ANOVA table, the Interaction Graph, and the Pearson Product Moment Correlation Table, I knew she was absolutely right. My resulting graph and tables told me more about what had actually happened between the variables than I had ever thought possible.

In April, I competed at our local science fair and won the top award, and with it, the chance to compete at the Intel International Science and Engineering Fair in Louisville, Kentucky in mid-May 2002. To say that the Intel ISEF was the highlight of my sophomore year would be a gross understatement – it was the experience of a lifetime. Over 2000 students and teachers from 39 nations attended, as well as Nobel Prize recipients who spoke with students and answered questions from the audience.

The first of three award ceremonies was the day after the judging, and it was at this award ceremony that the American Statistical Association presented the award for the Best Use of Statistics. When the announcement was made, my initial reactions were shock and overwhelming excitement. The director of my magnet program, Mrs. Jennifer Richter, immediately used her cell phone to call Spring Valley High School and my principal, Dr. Greg Owings. As I shook Chapter President Bill Wunderlin's hand, I was trembling with excitement. All the hours of work and, at times, frustration, were worth it. I knew I was hooked on science, math, and exploring areas of the environment where I felt I could have an impact. I felt I had found a career path that was challenging, exciting, and very worthwhile.

Table 4: Results of Stilwell□s Wipe Test

Set Number	Amount of As Dislodged in Stilwell□s Wipe Test (ug/100 cm²)	Amount of As Dislodged after being multiplied by 0.01 (ug/cm²)
1	15–31	0.15–0.31
2	6–33	0.06–0.33
3	56–122	0.56–1.22
4	15–26	0.15–0.26

Table 5: Comparison to ATSDR Minimal Risk Levels (mg As/kg of the average child□s body weight)

Time of Exposure	Preservative Concentration 0.25	0.40	2.50
2 hours	0.002	0.001	0.005
8 hours	0.008	0.005	0.010
12 hours	0.013	0.010	0.012

Discrepancies in findings also could have resulted from the uneven surface of the chicken skin and the fat content in the chicken skin. It is possible that an increased fat content could affect absorption (Gensie Waldrop, personal communication, November 3, 2001).

In the future, I could dip the skins in water to remove surface arsenic before they are digested. Also, I could analyze sand samples from local playgrounds for arsenic content.

Acknowledgements

I would like to thank Ms. Gensie Waldrop, Mrs. Marilyn Senneway, Dr. Glenda George, Dr. David Stilwell, Ms. Debbie Easler, Mr. and Mrs. M.K. Weingarth, Amick Farms, and Mr. and Mrs. Douglas L. Van Schaik.

References

Agency for Toxic Substances and Disease Registry (ATSDR). (2001), "Miminal Risk Levels (MRLs) for Hazardous Substances," on the web at www.atsdr.cdc.gov/mrls.html.

American Wood Preservers' Association (AWPA) (2001), on the web at www.preservedwood.com.

Atomic Absorption Spectrophotometry (2001), on the web at campus.murraystate.edu/academic/faculty/judy.ratliff/graphite.htm.

"Coalition: Ban treated wood," (2001). *St. Petersburg Press*, May 23, A1.

Florida Hazardous Waste and Waste Management Department (2000), "What is Treated Wood," on the web at www.ccaresearch.org.

George, G. (2001), personal communication, September 3.

Gradient Corporation (2001), "Focused Evaluation of Human Health Risks Associated with Exposure to Arsenic from CCA-Treated Wood," on the web at www.preservedwood.com/safety/ccafocus.pdf.

Stilwell, D.E. (1998), "Arsenic from CCA-treated wood can be reduced by coating," *Frontiers of Plant Science*, 51, 6–8.

Stilwell, D.E. and Gorny, K.D. (1997), "Contamination of soil with copper, chromium, and arsenic under decks built from pressure-treated wood," *Bulletin of Environmental Contamination and Toxicology*, 58, 22–29.

U.S. Department of Health and Human Services (2000), "Toxicological Profile for Arsenic," Syracuse Research Corporation under contract no. 205-1999-00024.

U.S. Environmental Protection Agency (USEPA) (1987), on the web at www.epa.gov/swercepp/ehs/profile/1303282p.txt.

Waldrop, G. (2001), personal communication, November 3.

Getting Involved With Science Fairs

by Joe Ward

Science Fairs and other science competitions are held across the United States each year with the goals of encouraging students to share ideas, motivating innovation, and showcasing cutting-edge science projects. These are very important goals not only for individual students but also for the future scientific progress of our nation. Students compete for awards and scholarships in these competitions, and the statistical analysis incorporated in their projects can play an important role in the final quality of their work. Judges at these contests frequently include statisticians.

For several years I have attended the International Science & Engineering Fair (ISEF). I always visit the ASA Special Awards winners to congratulate them. Katherine Van Schaik won not only the ASA First Special Award of $500 and a plaque, but also a U.S. Air Force First Award of $3,000, and an Environmental Protection Agency Second Award of $1,500.

I also give a presentation titled "Combining the Power of Statistics and Computers to Enhance Science Fair Projects" to Fair Directors, Teachers, Parents and ISEF Finalists. I present some ideas about applying computer-based statistical analysis techniques to improve research projects.

Also, of most interest is a discussion of ways to obtain statistical support for student research. I encourage participants to go to www.amstat.org to identify ASA Chapters and Chapter officers to locate statisticians in their area who might assist with statistical design and data analysis techniques. Some students have not only contacted nearby assistance but have also received valuable long-distance guidance by phone and email.

I also believe that statisticians can do more to help high school science fair participants. I suggest that statisticians consider:

- assisting high school students with statistical design and data analysis techniques that will enhance the quality of their research projects;
- providing special statistics awards and judges to select the award winners at local science-related fairs;
- offering students information about the association between winning statistics special awards and winning specific science-category awards;
- encouraging special statistics award-winners to enter the ASA Poster & Project Competitions (see www.amstat.org/education/index.html#K12) and to submit their research papers for publication consideration in journals and magazines such as STATS.

These activities can contribute much toward "selling" science fair students on the value of using quality statistics in their research projects.

To learn about science-related competitions, consult sources such as Science Service at www.sciserv.org (to locate the Science Fairs in nearby locations), an internet search of Junior Academy of Science (to locate Junior Academy competitions) and the Junior Science and Humanities Symposium at www.ijas.org/. Since many school districts have web pages, it is easy to locate nearby schools and teachers who are already engaged in science and engineering research projects.

Greek Letters in Measurement and Statistics. Is it All Greek to You?

Elena C. Papanastasiou

Because I am Greek, and work in the field of measurement and statistics in the United States, many people approach me to ask questions about the Greek alphabet, which are usually followed by the comment, 'It's all Greek to me!" More specifically, students, professors and other colleagues have questions about how to properly pronounce the names of, and sounds of letters in the Greek alphabet. In the past, when they used to ask me these questions, I would not be sure how to answer. I was not sure if I should tell them how to pronounce the letters in the proper, Greek way, or in the Americanized version of the Greek alphabet. However, as time passed by, I found myself using the American pronunciation of the Greek alphabet more often than the proper Greek way in the USA. It was just a lot easier to say that the letter 'μ' is called 'miu', instead of 'mee', which is the way it is pronounced in Greek. It was also a lot easier to call χ^2 chi-square, and not 'hee-square.' However, since people are still asking me about how to pronounce the Greek letters, I decided to create a table that would answer all of those questions. This table includes information on how the names of the Greek letters are actually pronounced in Greek and what sounds they actually represent. This table also provides the corresponding adapted International Phonetic Alphabet (IPA) sounds for each letter.

Elena C. Papanastasiou is an assistant professor at the University of Kansas in the department of Educational Psychology and Research. She obtained her Ph.D. in Measurement and Quantitative Methods from Michigan State University, and her area of expertise is that of Item Response Theory. Her research interests include the issues of Computer Adaptive Testing, as well as comparative international research in the subject areas of science and mathematics education. She previously held an appointment in the Department of Education at the University of Cypress. **<Cyprus perhaps??>**

The letters are presented in Table 1, in the order they appear in the Greek alphabet. The first column presents the upper-case Greek letters, and the second column presents the lower-case Greek letters. The third column presents the Roman letters that most closely approximate each Greek letter. The Greek alphabet contains some letters for sounds that are not represented by single letters in the Roman alphabet, such as the letters δ, θ, ξ, and ψ. In addition, in some cases, the Greek alphabet has more than one letter that represents the same sound. For example, the letters ι, υ, and η all correspond to the sound 'ee'. In addition, the letters ο and ω both represent a sharp version of the sound 'o.'

A note should be made that the fourth and fifth columns that will be discussed next, represent the letters in a way that would allow people with an American accent to pronounce the letters and their names in the appropriate manner. More specifically, the fourth column includes English words that contain the sound of the letter that is represented in that row. The sound each letter represents is bolded in that word. For example, the Greek letter θ sounds like the 'th' sound in the word '**th**ink', while the Greek letter δ sounds like the 'th' sound in the word "**th**en". American readers should also note that when the letter 't' or 'τ' is used in Greek, or in Table 1, it represents a soft version of the sound 't', which is a sound between the English sounds of 't' and 'd'. Consequently, the letter 'τ' was used in this paper as part of the IPA, to represent this softer sound. This was the only adaptation that was made to the IPA for this paper. The fifth column presents how the name of each letter should be pronounced. For example, the letter φ is called 'fee' in Greek and not 'fy', while the letter ψ is called 'ppsee' in Greek and not 'sy'.

The sixth column represents the correct way of pronouncing the name of each letter, based on an adaptation of the IPA. The column that follows includes the corresponding phonetic (IPA) sounds of what each letter sounds like. The last column includes a few

Table 1. Greek Letters and Their Proper Pronunciation

Capital Greek Letter	Small Greek Letter	Corresponding English letters	Sounds like	Letter□s name	Phonetic (IPA) pronunciation of letter□s name	Phonetic (IPA) sounds	Examples of what it represents in statistics/ measurement
Α	α	A	**a**nother	halfa	[ál.fa]	[a]	Significance level
Β	β	V	**v**alidity	veeta	[ví:.τa]	[v]	Beta weight
Γ	γ	Y	**y**es	yhahmma	[γá.ma]	[γ]	Lower asymptote, or pseudo-guessing parameter
Δ	δ	Th	**th**en	thelta	[δεl. τa]	[δ]	Difference
Ε	ε	E	eff**e**ct	hepsilon	[ε.psi.lon]	[ε]	Residual
Ζ	ζ	Z	**z**-score	zeeta	[zí:.τa]	[z]	Proportion-correct score
Η	η	Ee	m**e**dian	eeta	[í:.ta]	[i]	Eta-squared
Θ	θ	Th	**th**ink	theeta	[θí:. τa]	[θ]	IRT Ability
Ι	ι	Ee	m**ea**l	yiota	[ió.τa]	[i]	
Κ	κ	K	fa**c**tor	gkappa	[ká:.pa]	[k]	Kappa coefficient
Λ	λ	L	**l**ength	lamtha	[lám.δa]	[l]	Wilk's lambda
Μ	μ	M	**m**ean	mee	[mi]	[m]	Mean
Ν	ν	N	**n**ormal	nee	[ni]	[n]	
Ξ	ξ	X	ta**x**i	ksee	[ksi]	[ks]	Number-right true score
Ο	ο	O	**o**rdinal	omikron	[ó.mi.kron]	[o]	
Π	ϖ	P	a**p**e	pea	[pi]	[p]	Pi = 3.14159…
Ρ	ρ	R	mo**r**e	rho	[Rç]	[r]	Correlation, Reliability
Σ	σ	S	**s**tatistics	seehyma	[síγ.ma]	[s]	Sum, SD
Τ	τ	T	a**t**e	tahf	[τaf]	[τ]	True score
Υ	υ	Ee	m**e**dian	eepseelon	[í.psi.lon]	[i]	
Φ	φ	F	**f**igure	fee	[fi]	[f]	Phi-coefficient, Normal cumulative distribution function
Χ	χ	H	**h**istogram	hee	[hi]	[h]	Chi-square
Ψ	ψ	Pps	i**ps**ilon	ppsee	[psi]	[ps]	Logistic distribution of a density function
Ω	ω	O	**o**rdinal	omehya	[o.mε.γa]	[o]	Omega-square

examples of what the Greek letters represent in measurement or statistics. This column is not intended to be a comprehensive list that includes all the meanings of each letter. In addition, there are some letters (e.g., the letters **ι, ν, υ**) that do not represent any specific symbol in measurement or statistics because they are quite similar to letters that are used in the Roman alphabet.

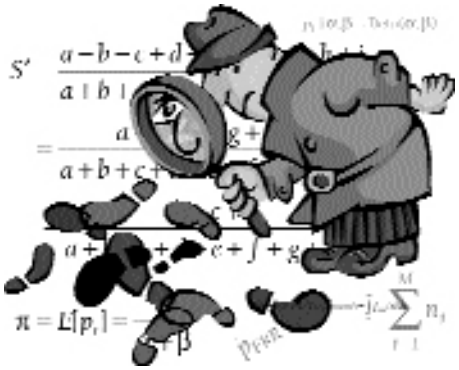
Finally, I would like to make clear that the purpose of creating this table was not to 'force' people to use the correct pronunciation of these letters. This is especially the case since I find myself pronouncing them in the Americanized way. I created this table to provide a valuable resource for people in measurement or statistics who would like to learn about the Greek letters, without having to open a dictionary to do so. However, determining the most convenient way to pronounce the letters (with the Greek or USA pronunciation) in the fields of measurement and statistics is beyond the scope of this paper.

Acknowledgements

I would like to sincerely thank Dr. Antônio R. M. Simões for his assistance with the use of the International Phonetic Alphabet (IPA).

Data Sleuth

An Exhalent Problem



Contributed by Michael Kahn, Wheaton College (MA)

A dataset from Rosner's *Fundamentals of Biostatistics* concerns the relationship between forced expiratory volume (FEV, a measure of respiratory function) and smoking, along with several other variables. The data include information from 654 children and young adults, ranging from 3 to 19 years of age. The variables considered here are FEV (in liters), self-reported smoking status, and age (in years).

The boxplots in Figure 1 compare the distributions of the smokers' FEV with the nonsmokers' FEV.

Question #1: Using the boxplots, do nonsmokers appear to have, on average, higher FEV scores than smokers?

Question #2: Is it sensible to use these data to discuss, in isolation, the effects of smoking on FEV? In particular, would you conclude that smoking causes young people to strengthen their respiratory function? If not, can you suggest an alternative explanation for the differences in the boxplots?

Figure 2 compares the nonsmokers' and smokers' relationships between FEV and age. The "curves" are computed using *lowess* (Cleveland, 1979); they provide estimates of the (conditional) average FEV for a given age.

Question #3: Using the scatterplot and lowess curves, do the nonsmoking 16-year-olds appear to have, on average, stronger respiratory function than those 16-year-olds who smoke? 19-year-olds? 10-year-olds? Suggest some possible explanations for the inconsistencies in your answers to these questions.

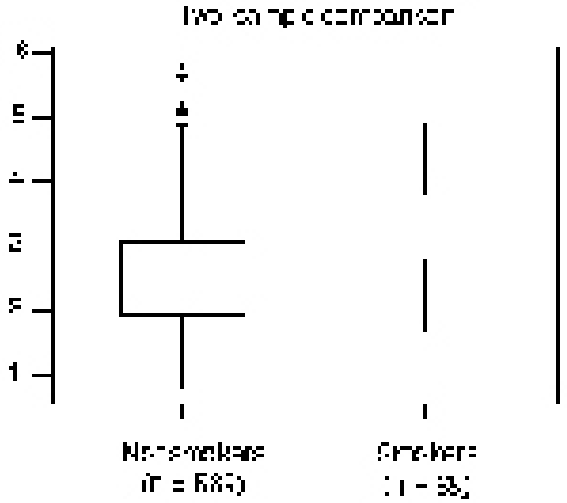


Figure 1: Forced Expiratory Volume by Smoking Status

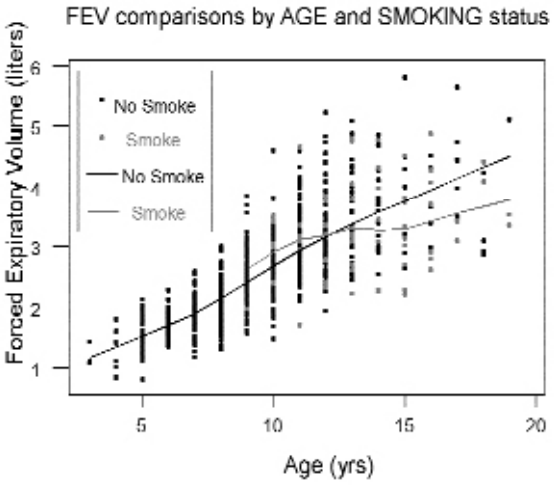


Figure 2: Forced Expiratory Volume by Age and Smoking Status

References

Cleveland, W.S. (1979), "Robust locally weighted regression and smoothing scatterplots," *Journal of the American Statistical Association*, 74, 829–836.

Rosner, B. (2000) *Fundamentals of Biostatistics* (5th ed.), Pacific Grove, CA: Duxbury Press.

The Statistical Sports Fan

Judging Figure Skating Judges

Robin Lock

Kari Frazer Lock

Figure skating is a wonderful sport, combining athleticism with artistry. However, unlike most other sports, figure skating judging is subjective. It is not like hockey, soccer, or lacrosse in which the team that scores the most goals wins, or like swimming or running in which the fastest person wins. Rather, a panel of judges decides the winner of a figure skating competition. The winner is chosen based on the subjective opinions of human beings who rank the performances of each of the skaters in the competition. This subjectivity sometimes causes skaters, coaches, commentators or fans to question the fairness of the judging for a particular event. How can we determine *statistically* when a specific figure skating judge produces a ranking of the skaters that is significantly different from the rankings of the other judges? We describe a technique, using a bootstrap distribution, for identifying an inconsistent judge and apply the method to competitions from the 2002 Winter Olympic Games.

Measuring a Judge„s Judgments

The system for determining the final placement of skates in a competition is somewhat complex (see sidebar) due to safeguards that are designed to prevent any one judge from exerting too much influence on the final result. See Basset & Persky (1994) or Russell (1997) for additional discussion of the merits of the "best of majority" method. The primary feature of this judging system that is important for our comparison of judges is that each judge produces a rank ordering of all of the skaters in a competition. For example, Table 1 gives the rankings for each of the nine judges (and the final placements) for the ladies free skate event at the 2002 Winter Olympics. Judges are human, they each have their own tastes and preferences, they may notice different elements of a particular performance and therefore we should not expect them to produce identical rankings for a particular set of skating performances. A certain degree of variability in the judge's rankings is inevitable, particularly in a close competition where the distinctions between the quality of the performances are small. But, occasionally one judge appears to stand out as being in noticeable disagreement with the other judges. Our task is determine when the deviations of one judge's rankings are significantly larger than one would expect to see, given the variability of the rankings for all the judges in that competition.

Since the methods for determining the final placement of skaters should not be unduly influenced by one inconsistent judge, we can determine that one judge is in significant disagreement with the other judges if that judge's rankings differ significantly from the final placement of the skaters. Although disagreeing with the other judges might not necessarily be bad, we will refer to the extent that a judge's rankings match the final placement of the skaters as the "success" of that judge. The rankings of a successful judge will closely match the final placement of the skaters, while the rankings of an unsuccessful judge will disagree with the final placements. To determine how much an individual judge agrees with the final placement, we will look at the Spearman rank correlation (just a correlation between the ranks of the data) between that judge's ranking and the final placement of the skaters. A high correlation will indicate a successful judge, while a lower correlation will indicate a less successful judge.

Table 2 gives the rank correlations of each judge with the final placements, from the ladies free skate event at the 2002 Winter Olympics.

The first thing to notice when looking at these correlations is that they are all extremely high. A perfect

Kari Frazer Lock is a junior majoring in mathematics and psychology at Williams College. She has earned United States Figure Skating Association gold medals in freestyle, moves in the field, and ice dancing. She skates professionally in shows across the U.S. and abroad.

Table 1: Ranks given by nine judges of the Ladies Free Skate at the 2002 Winter Olympics

Final Placement	Skater		J1	J2	J3	J4	J5	J6	J7	J8	J9
1	HUGHES Sarah	USA	1	4	3	4	1	2	1	1	1
2	SLUTSKAYA Irina	RUS	3	1	1	1	4	1	2	3	2
3	KWAN Michelle	USA	2	3	2	2	2	3	3	2	3
4	COHEN Sasha	USA	5	2	4	3	3	4	4	4	4
5	SUGURI Fumie	JPN	4	8	5	5	5	7	5	5	5
6	BUTYRSKAYA Maria	RUS	6	5	8	7	12	5	8	7	6
7	ROBINSON Jennifer	CAN	7	7	7	9	6	8	10	6	7
8	SEBESTYEN Julia	HUN	8	10	12	8	7	6	12	8	8
9	KETTUNEN Elina	FIN	9	9	13	6	12	10	7	11	14
10	VOLCHKOVA Viktoria	RUS	10	6	14	11	10	12	6	9	15
11	MANIACHENKO Galina	UKR	13	12	11	12	16	11	11	10	9
12	FONTANA Silvia	ITA	14	11	18	16	9	15	9	12	10
13	LIASHENKO Elena	UKR	15	13	6	10	8	14	13	14	16
14	ONDA Yoshie	JPN	11	14	10	15	15	13	15	13	11
15	HUBERT Laetitia	FRA	12	17	17	13	11	16	14	15	13
16	MEIER Sarah	SUI	16	16	9	14	14	9	16	16	12
17	GUSMEROLI Vanessa	FRA	17	15	15	17	17	18	17	17	17
18	SOLDATOVA Julia	BLR	19	18	22	20	21	17	18	18	19
19	HEGEL Idora	CRO	20	21	16	22	18	19	21	19	18
20	GIUNCHI Vanessa	ITA	18	19	20	21	19	20	20	20	20
21	BABIAKOVA Zuzana	SVK	22	20	19	19	20	21	19	22	22
22	KOPAC Mojca	SLO	21	22	23	18	22	22	22	21	21
23	LUCA Roxana	ROM	23	23	21	23	23	23	23	23	23

correlation (where the judge agrees exactly with the final placements) is 1.0, and only one of these correlations is even below 0.9. This tells us that each of the judges is in general agreement with the final placement of the skaters (and so in general agreement with each other). This is good news! It tells us that judges are basing the judging more on the skaters than on individual preferences (if they were judging based on individual preferences, the correlations would not be so high). However the skaters are being ranked (hopefully it is based on their skating performance, but it could also be based on their reputation, among other things), they are being ranked in a way in which the judges agree. The rankings are not arbitrary from judge to judge, and thus the final placements are more credible and meaningful.

Looking at the correlations for each judge, you will notice that the correlation between Judge #3's rankings and the final placement is lower than the correlations of the other judges. Obviously, one judge must have the lowest correlation. How do we know if the correlation for Judge #3 is simply the lowest correlation of these judges, or if it is *significantly* lower than the correlations of the other judges?

Ordinarily, to determine if a sample statistic (such as Judge #3's correlation) is statistically significant, we would compare the statistic to some underlying distribution, look at how far out on the distribution it lies, and calculate the probability of a sample statistic being that far out if all of the judges were consistent. If this probability is small, then the sample statistic is statistically significant and we would conclude that the judge is inconsistent with the others in the panel. But what is our underlying distribution in this case? We can't really compare the sample results to a population of all possible judges, or to all events, since each event is different and some events are harder to judge than others (sometimes there is a clear order in which the skaters should be ranked, which would give extremely high correlations, and sometimes all the skaters are about the same, which would give low correlations). Here we are only comparing the correlation of Judge #3 to the correlations of the other eight judges of the event, so its difficult to determine a probability for how "unusual"

Table 2: Rank correlations for judges of the Ladies Free Skate

J1	J2	J3	J4	J5	J6	J7	J8	J9
.979	.970	.876	.953	.928	.960	.966	.993	.951

the correlation of 0.876 is in this case.

Constructing a Bootstrap Distribution

A general technique for using the data in a sample to produce a reference distribution for a sample statistic is called the *bootstrap*. The basic idea is to randomly select elements of the sample itself to generate new samples and then to examine the distribution of the statistic in question for all of these new samples. Thus we don't need to make assumptions about the distribution of the underlying population itself; instead we let the bootstrap samples reveal relevant structure. In our case of figure skating judges, we use the rankings provided by the nine actual judges to produce a much larger "population" of judges with similar rankings. We can then create a whole distribution of rank correlations of the rankings of these simulated judges with the actual final placements of the skaters in the competition and see where the correlation of Judge #3 fits in this distribution. For more information on bootstrap techniques see Efron and Tibshirani (1993).

We start with the 9 ordinals (rankings) each skater received (one from each "real" judge). To create a simulated judge's score for a particular skater, randomly choose one of the 9 ordinals actually received by that skater. For example, since Michelle Kwan received five 2nd's and four 3rd's, each simulated judge has a 5/9 probability of giving Michelle 2nd, and a 4/9 probability of giving Michelle 3rd. (Thus, the simulated judge is judging the event as a real judge would) Thus we are simulating the behavior of a real judge, since the simulated ordinal given each skater was actually received by the skater from a real judge, and if more actual judges gave a skater one ordinal, more simulated judges would tend to give a skater that ordinal. Then repeat this random selection from the ordinals received by the other skaters. In actual competitions, a judge rarely gives two skaters the same ordinal (an exact tie); but this could easily occur for our simulated judges (for example, the random selection might choose ordinals of "3" for both Irina Slutskaya and Sasha Cohen). Ties are routinely handled in a rank correlation by averaging, so Slutskaya and Cohen would both be given a 3.5 rank and the next best ordinal would get rank 5.

If you'd like to generate your own new "judge" by hand, use 23 digits from a random number table (ignoring zeros) to determine which of the nine judge's ordinal you'll choose for each of the 23 skaters in Table 1, then determine the new judge's ranking by ranking those ordinals and averaging ties. For example, suppose that you enter a random number table and find the first five digits to be: 26885. Then your simulated judge would assign judge 2's ranking to the first skater (4 for Hughes), judge 6's ranking to the second skater (1 for Slutskaya), judge 8's ranking to skaters 3 and 4 (2 for Kwan, 4 for Cohen), and judge 5's ranking to the fifth

After watching an individual or pair skate, each judge awards two scores (on a 0-6.0 scale); one for the *technical merit* of the performance and the other score for its *artistic presentation*. These scores are then added together to give a combined score for each skater. After all the skaters have competed, they are ranked for each judge according to these combined scores, with the skater with the highest combined score receiving a first from that judge. In the case of a tie, the skater with the higher presentation mark is ranked higher. This results in each skater receiving an ordinal (rank) from each judge, with 1 being the top rank. The ordinals for the ladies free skate from the 2002 Winter Olympics are shown in Table 1. Final placements are determined by comparing the ordinals of the skaters.

The key factor in determining most final placements is a skater□s *median ordinal*, the position where a majority of the judges (at least 5 of 9 in our case) place the skater at or better. A skater with a better median ordinal will always finish ahead of a skater with a worse (larger) median ordinal. When skaters finish with the same median ordinal, preference is given to the skater who has more judges giving that rank or better (called the *size of the majority*). If a tie still exists, the actual ordinals on the majority side (those at or better than the median) are added, with preference given to the smaller sum. If the tie is still unbroken, the ordinals for all of the judges are added. If that fails to determine a winner, the skaters are officially listed as tied.

Let□s see how these rules apply to determine some of the final placements for the ladies free skate shown in Table 1. Five of the nine judges (a majority) placed Sarah Hughes in first, so she won the top spot (and the gold medal). Irina Slutskaya and Michelle Kwan each had a median ordinal of 2, but Kwan only got five judges at 2nd or better, while Slutskaya had six, so Slutskaya finishes second and Kwan drops to third. Sasha Cohen and Fumie Suguri had median ordinals of 4 and 5 respectively, so they are easily placed in those positions. Next come Maria Butyrskya and Jennifer Robinson, each with a median ordinal of 7 and exactly six of the nine judges placing them at 7th or better. Butyrskya□s majority included ranks of 6, 5, 7, 5, 7, 6 while Robinson□s were 7, 7, 7, 6, 6, 7, so Butyrskya□s sum is smaller and she gets the sixth spot and Robinson goes to seventh in the final placements.

An important goal of this system is to prevent a lone judge from single-handedly helping or hurting a skater by giving them a mark much lower or higher than they deserve. For example, if you look at the results for Sasha Cohen in Table #1, all that mattered in her ranking was that 8 of her 9 ordinals were 4th or better. She would have received the same final placement if Judge #2 had not been so generous giving her 2nd or if Judge #1 had ranked her 17th instead of 5th. While the judges□ opinions should certainly determine the final outcome of the event, no rogue judge should be able to unduly influence

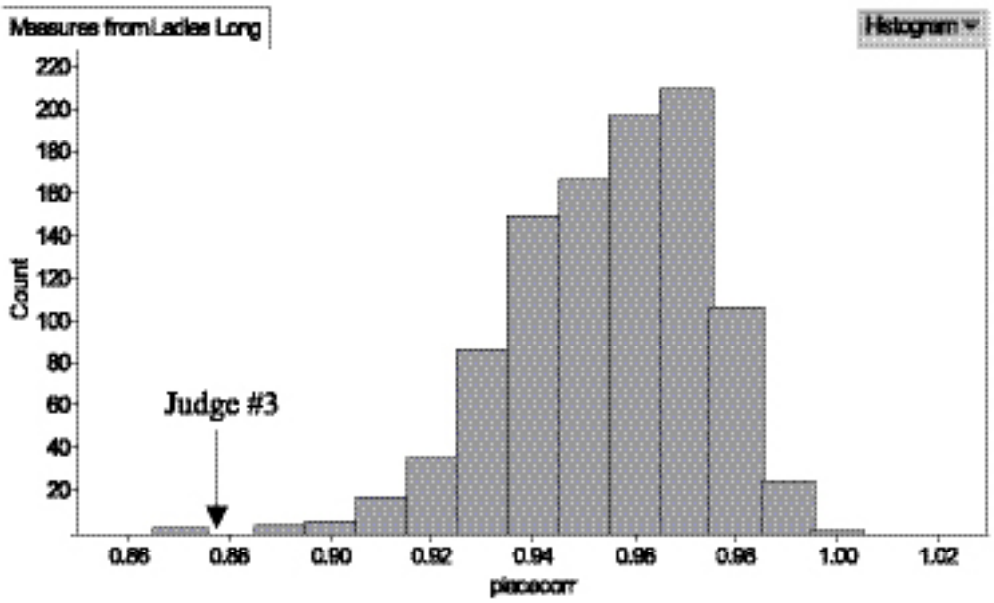


Figure 1. Rank correlations for 1,000 simulated judges of the Ladies Free Skate

skater (5 for Suguri) and so on. Accounting for the tie (between Hughes and Cohen) and assuming that no later skater is ranked in the top five, the "random" judge's rankings would start with (1) Slutskaya, (2) Kwan, (3.5) Hughes, (3.5) Cohen and (5) Suguri.

Using random selections (with the software package Fathom), we generated 1000 simulated judges, and found the correlation of the rankings of each simulated judge with the final placements. Since the correlation is a measure of the "success" of the judge in matching the actual final placements, this set of simulated correlations provides a yardstick for identifying where a typical judge's correlation should lie for this particular event, considering random variation. A histogram of these bootstrap correlations is shown in Figure 1.

We can see that Judge #3's correlation of 0.876 falls toward the extreme lower end of the distribution. The great majority (998 out of 1000 to be exact) of the generated judges have correlations higher than that of Judge #3. If Judge #3 really was judging in the same way the other judges were, it would be very rare for her to be that far out in the distribution. The approximate p-value from the bootstrap distribution is the proportion of simulated judges with a correlation as low or lower than the judge in question. Thus, the p-value for Judge #3 is about 0.002 and we can conclude that the correlation of Judge #3 is significantly low and, therefore, Judge #3's rankings are not consistent with the other judges. This could indicate a bias, poor quality judging, a mistake by the judge, or most likely just a difference in opinion. But, whatever the reason, Judge #3's assessment of the Ladies Free Skate event at the 2002 Winter Olympics was significantly different from the other 8 judges of that event.

The Notorious French Judge

What about the famous case of the French judge of the pairs long program at the 2002 Winter Olympics? She was accused of being biased to favor the Russian pair over the Canadians. But did she really judge the event significantly differently than the other judges? Below are the correlations for the 9 judges of the pair long program (Table 3).

The first thing you should notice about these correlations is that they are extremely high! Every one of them is above 0.98, which means that the judges agreed very much on the ranking of the skaters in that event. These strong correlations could be due to very accurate judging by the panel, clear differences between the performances of the skaters, or a tendency to pre-judge a competition and placed skaters according to their past reputations. Where was the notorious French judge among these correlations? Surprisingly, she was Judge #4, who had the highest correlation of the whole panel. In fact, when we did the bootstrap (see Figure 2), she was significantly *more* in agreement with the final place-

Table 3: Rank correlations for judges of the Pairs Long Program

J1	J2	J3	J4	J5	J6	J7	J8	J9
.994	.994	.988	.998	.997	.986	.992	.994	.983

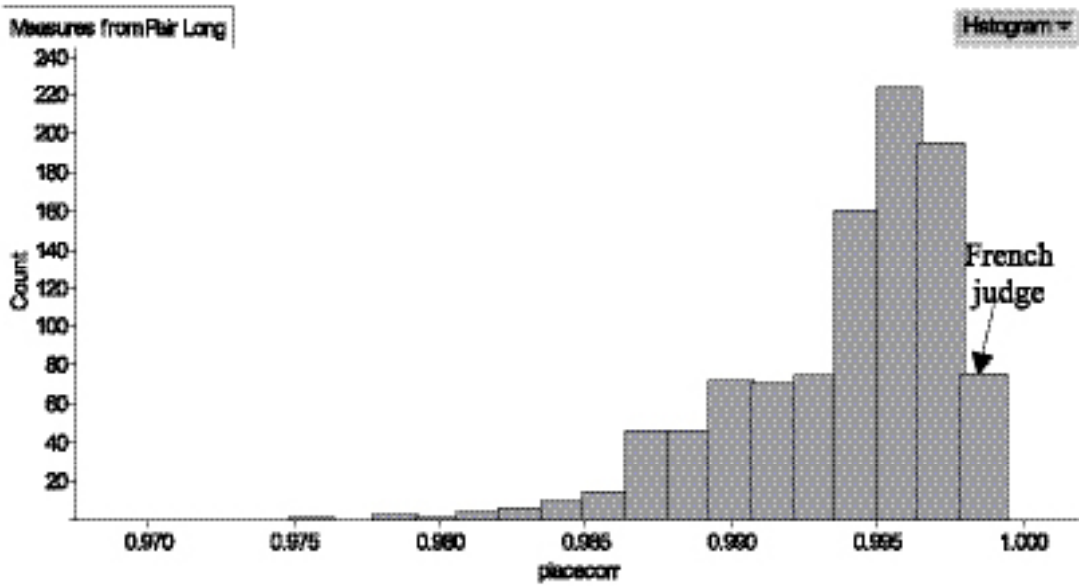


Figure 2. Rank correlations for 1000 simulated judges of the Pairs Long Program

ments than the other judges. Her rankings differed from the final placements by just a single permutation of the 8th and 9th place positions. By our definition of "successful" she judged the event extremely accurately, yet she was accused of bias. Why? Her rankings of the 1st and 2nd place skaters were the same as the original final placements, but were critical for determining those placements since the other 8 judges had split evenly, giving four firsts and four seconds each to the Russian and Canadian pairs. Many observers of the competition believed that the Canadian pair had skated a superior program, so the judging was scrutinized with extra attention. The controversy erupted with allegations that the French judge had been pressured or agreed to a deal to favor the Russian pair over the Canadians before the competition even started. In light of these allegations, the French judge's rankings were disregarded, producing an exact tie for the top spot and duplicate gold medals were awarded. While the distinction between first and second in a close competition is very important to the skaters, their fans and countries, a single permutation may have relatively little affect on the rank correlations between an individual judge and the final placements. So, our bootstrap procedure would not detect that sort of bias in judging.

Conclusion

The bootstrap technique provides a means to assess when the correlation between an individual judge's rankings of the skaters in a competition and the final placements of those skaters by the entire panel of judges is unusually low. While we have applied these ideas to two events from the 2002 Winter Olympics, they could also be applied to other figure skating competitions at various levels or to other "judged" events such as gymnastics, diving, or freestyle skiing. Interesting avenues for future work would be to try to characterize the distribution of rank correlations among judges for different events, levels of competition, numbers of judges, or types of sports. Is the skewed shape of the bootstrap distribution that we see in our two examples typical of most cases? Do pairs skating competitions tend to produce higher correlations than individual events? Can we follow the same judge over several competitions to determine a consistent pattern of disagreement? These methods can help determine whether a judge is really inconsistent with the rest of the judges or just exhibiting the sort of random variation in rankings that one would naturally expect for that particular competition.

References

Bassett, G.W. and Persky, J. (1994), "Rating Skating," *Journal of the American Statistical Association*, 89, 1075–1079.

Efron, B. and Tibshirani, R. (1993), *An Introduction to the Bootstrap*, New York: Chapman and Hall.

Russell, E. (1997), "Choosing a Rating System," *STATS: The Magazine for Students of Statistics*, 19, 13–17.

Web Resource

Judging results for figure skating at the 2002 Winter Olympic Games can be found at the United States Figure Skating Associations website *http://www.usfsa.org/olys02/results/*

Understanding Regression Output

Josh Tabor

In our society, it is usually considered impolite to ask how much money a person makes. However, suppose that you are single and are interested in dating a particular person. Of course, salary isn't the most important factor when considering whom to date, but it is certainly nice to know (especially if it is high!). Well, in this case, the person you are interested in happens to be a teacher, so you know a high salary isn't an issue. Still, you would like to know how much she makes, so you take an informal survey of 13 teachers that you know. Their salaries (in thousands of dollars) are listed below:

salary (in thousands) 39.9, 47.6, 49.3, 51.6, 47, 46.2, 48.5, 51.7, 58.1, 56.1, 63.7

Based on this data, what can you conclude? Well, absent any other information, your best estimate for her salary would be the average ($50,882). However, it is not likely that your estimate will be correct. To get an idea of how far off you might be, you can calculate the standard deviation:

$$s = \sqrt{\frac{\sum_{i=1}^{11}(x_i - \bar{x})^2}{n-1}} = \sqrt{\frac{421.36}{11-1}} = 6.49$$

Thus, your best estimate for her salary is $50,882 but a typical estimate will be off by about $6500.

So, how can you improve your guess? You remember that one of the teachers told you that teachers' salaries are somewhat dependent on how many years a teacher has been teaching. So, you go back to each of the original 11 teachers you surveyed and ask them for their years of experience.

salary (in thousands) 39.9, 47.6, 49.3, 51.6, 47, 46.2, 48.5, 51.7, 58.1, 56.1, 63.7
Years of experience 4, 8, 5, 9, 1, 4, 4, 6, 8, 7, 11 (source: Hacienda La Puente Unified School District Salary Schedule)

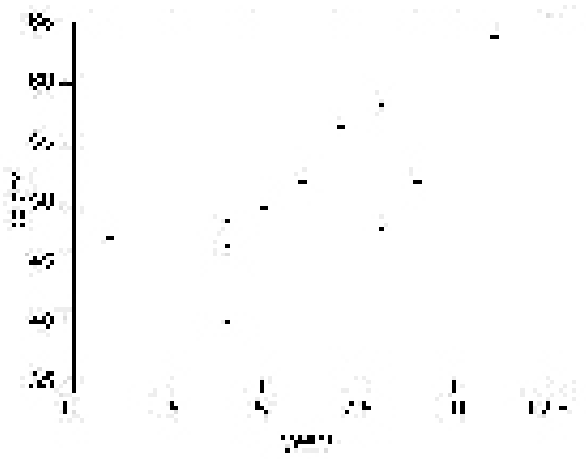


Figure 1: Scatterplot of salary ($1000s) and years of experience

Making a scatterplot of salary (y) vs. years of experience (x) you see that the teacher was right. There is a positive association between years of experience and salary.

You happen to know that the person you are interested in has been teaching for 8 years. Using this information, how can you predict this potential mate's salary? Using JMP-Intro (or some other software package), you can calculate the least squares regression line. In this case, $\hat{salary} = 40.61 + 1.686\ years$. Therefore, you predict that she makes: $\hat{salary} = 40.61 + 1.686(8) = 54.098$ or $54,098.

When you ask most statistics packages to calculate a regression line, however, you get a lot of other information besides the regression line. All of the output can look intimidating at first, but with a little time and patience most of it can (and should) be understood by a student in an AP Statistics class. In most cases, spending a little time with the output will greatly enrich your understanding of the relationship between the two variables.

The output below is from JMP-Intro, but it is very similar to the output from other software packages:

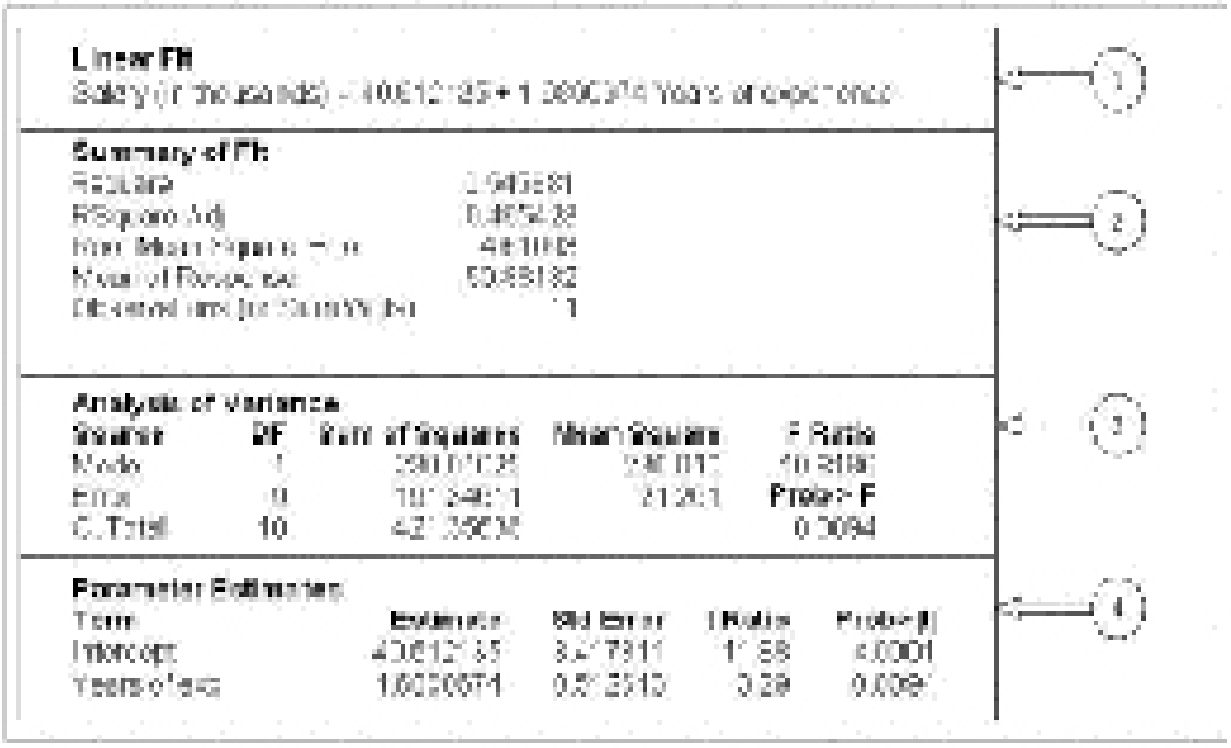


Figure 2: JMP-Intro Regression Analysis

The first section simply reports the regression line.

The second section gives a summary of how well the line fits the data. We will come back to this section after we have looked at the others.

The third section is called "Analysis of Variance." As the name suggests, the results in this section analyze the variance (in this case, the variance of the response variable, salary). First, let's consider the last row, labeled "C. Total." Under the column "Sum of Squares" we see a familiar number: 421.36. This was the same number that we used when we calculated the standard deviation of y. Thus, the Total Sum of Squares (or SSTotal) is

$$\sum_{i=1}^{n}(y_i - \bar{y})^2$$

This is a measure of the total variability in salary (from the mean salary). To find the standard deviation we took the square root after we divided by 11 – 1 = 10, which is the number we see under the column called "DF" (degrees of freedom). Note: For a discussion of degrees of freedom, see Gretchen Davis' article in *STATS* issue 33 (2002).

Moving up one line in the "Analysis of Variance" section we see a row called "Error." Under the Sum of Square column, we see the number 191.35. This is a measure of the variability in salary from the regression line:

$$SSError = \sum_{i=1}^{n}(y_i - \hat{y}_i)^2$$

Since the regression line is usually better for making predictions than the mean (see Figure 3, points tend to lie much closer to the $\hat{y}$ line than to the $\bar{y}$ line), SSError should be less than SSTotal (and will never be more).

The quantity SSError is also useful because we can use it to calculate the standard deviation about the regression line:

$$s = \sqrt{\frac{\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}{n-2}} = \sqrt{\frac{191.35}{11-2}} = \sqrt{21.26} = 4.61$$

Thus, when we use the regression line to make predictions, the typical prediction will be off by only \$4610 (as compared to \$6490 when we used the mean as an estimate). Notice that the number 4.610 shows up in the "Summary of Fit" table as the "Root Mean Square Error" (in many packages this is simply called "s"). Thus, in the regression setting, the symbol s represents the typical deviation from the least squares regression line. However, in the univariate case, s represents the typical deviation from the mean. In some books, s_e is used for the regression setting and s_x for the univariate case.

Also, notice that several of the numbers in our

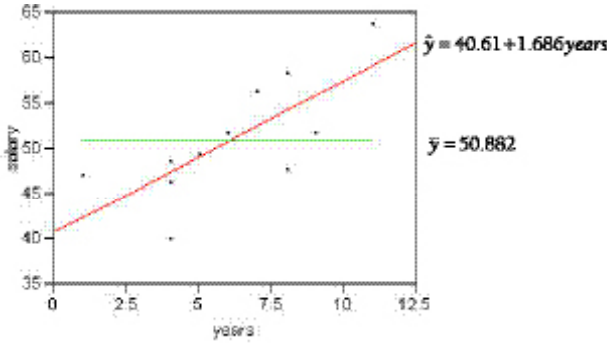


Figure 3: Comparison of Regression Line to $\hat{y}$ line.

equation show up in other places on the table. This is no coincidence! The denominator in the square root (11) is the same as the DF Error from the table. The quotient of the Sum of Squares and Degrees of Freedom (SS/DF = 21.261) is called the Mean Square (Error).

The first line in the "Analysis of Variance" table is called "Model." In this case, DF Model = 1 since we are only using 1 predictor variable (years). In multiple regression, DF Model = number of predictor variables. The Sum of Squares for the Model is a measure of the variability in salary that is accounted for by the relationship between years and salary:

$$SSModel = \sum_{i=1}^{n} (\hat{y}_i - \bar{y})^2 = 230.01 .$$

This brings us to an important relationship between the various Sums of Squares: SSTotal = SSModel + SSError. That is, the total variability in salary (SSTotal) is partitioned into 2 parts: the part that is explained by the model using years (SSModel), and the part that is still unexplained by the model using years (SSError).

In addition to the standard deviation about the regression line (s, or root mean square error), another common way to judge the quality of a model is to determine what percentage of the variability in salary is accounted for by the model. This is called the coefficient of determination, often denoted by r^2 or R^2. To determine the value of r^2, we examine the ratio of the variability explained by the model and the variability in the response:

$$\frac{SSModel}{SSTotal} = \frac{230.01}{421.36} = .546$$

which can also be found in the "Summary of Fit" table. Thus, 54.6% of the variability in salary can be explained its relationship with years.

A few final notes about the "Analysis of Variance" table: The F statistic is calculated to determine the overall "usefulness" of the model

$$\left(F = \frac{SSModel / DFModel}{SSError / DFError} = \frac{MSModel}{MSError} = 10.82 \right)$$

The "Prob > F" is the probability that we would get an F ratio this big by random chance assuming that the model is not useful (i.e., that knowing the years of experience is of no help in predicting salary). Thus, since the p-value is pretty small (.0094) we can say that the model is useful.

Now it is time to go back to the "Summary of Fit" table. As described earlier, "RSquare" is the coefficient of determination (r^2) and "Root Mean Square Error" is the standard deviation about the regression line (s). Also included in the table are the "Mean of Response" which is the average salary ($\bar{y}$) and the number of "Observations" or sample size (n). Finally, "RSquare Adj" is a statistic used when evaluating multiple regression models. So, in the case of simple linear regression (one predictor variable), we can usually ignore the adjusted r-squared value.

The last section (Section 4) in the computer output is the table of "Parameter Estimates." In simple linear regression we find models in the form $\hat{y} = a + bx$, where "a" is the y-intercept (or constant) and "b" is the slope (or coefficient of the dependent variable). In our model, the estimate of the intercept is 40.61 and the estimate of the slope is 1.686. Notice that these are the same numbers that appear at the very top of the output under "Linear Fit." Based on these numbers, we can predict that a beginning teacher (that is, a teacher with 0 years of experience) will make about $40,610 and that for each additional year of experience, the salary will increase by about $1,686 on average.

Of course, these values are only estimates based on the 11 observations in your sample. Other samples from the same population of teachers would certainly give different estimates (although hopefully they will be close!). How precise are the estimates? The "Std. Error" column gives us this information. For example, for the sample slope the std. error = 0.5126, thus we can say that in sample like ours, the typical estimate of the slope will be off from the population slope by about 0.5. Taking different samples will give different slope estimates, but they should typically vary by only about .5126

We can also use the standard error make inferences about the population slope. For example, the 95% Confidence Interval with 9 degrees of freedom would be 1.686 ± 2.262 (0.5126) = (.5265, 2.846). Also, the t-ratio is the t-statistic for a test of slope = 0

$$\left(t = \frac{b - 0}{s_b} = \frac{1.686 - 0}{.5126} = 3.29 \right)$$

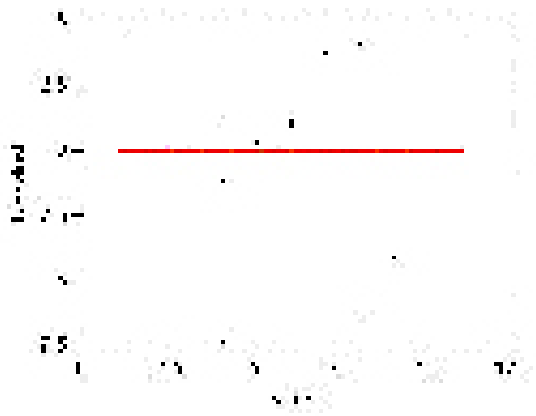


Figure 4: Residual Plot

With 9 degrees of freedom, the p-value is .0094, which indicates that we have strong evidence that the population slope is not 0. Notice that this is the same p-value we observed from the F-test in the "Analysis of Variance" table. Also, notice that $t^2 = F$ ($3.29^2 = 10.82$). Note however, that these relationships only hold in the simple linear regression case (one predictor variable).

It is important to note that the standard computer output does not allow you to check the conditions for making inferences about the slope (e.g., the linear model is appropriate). To check the conditions, it is important to look at the residuals ($y - \hat{y}$). The residual plot for our data is shown below. There is no obvious pattern, so the linear model seems appropriate. Also, the variability of the residuals seems relatively constant across all values of x.

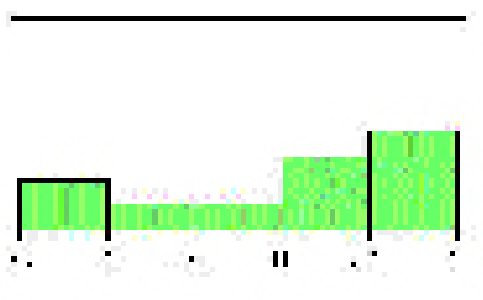


Figure 5: Histogram of Residuals

Finally, it is important to check if the distribution of residuals is approximately normal. Based on the histogram below, the normality of the residuals seems plausible.

The first line in the parameter estimates table gives similar information about the intercept. The estimate of the population intercept is \$40,612 with a standard error of \$3418. The t-ratio (11.88) is for a test that the population intercept is 0. Since the p-value is very small (<.0001) we can safely conclude that the population intercept is not 0.

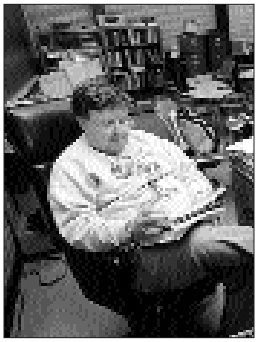
So, what have we learned from our investigation? We have learned that knowing how many years a teacher has been teaching allows us to make better estimates than simply using the average salary (the typical error went down from \$6490 to \$4610 and about 54.6% of the variability in salary was explained by years). We also learned that this relationship was not due to chance (p-value = .0094). We didn't learn the exact salary of your prospective spouse, but we came up with a pretty good estimate using statistics. And besides, money isn't everything, right??

References

Davis, G. (2002), "Some Thoughts about Degrees of Freedom," *STATS: The Magazine for Students of Statistics*, 33, 18–20.

μ-Sings

The Investigation



Chris Olsen

I guess I might just as well confess. I like a good mystery, and — just so you know — am somewhat partial to Agatha Christie. To those who know a little about statistics and enjoy a good murder mystery as much as a well-crafted random variable, it might seem that solving statistical problems is a lot like detective work, except that statistics problems are solved even when nobody gets killed, kidnapped, robbed, defrauded, or even mildly offended.

The practicing statistician, working on a case, must round up the usual suspects, query the reluctant data, discredit the offending outliers, and in the end unmask that pesky parameter with 95% confidence on the last page. So too, must the working detective query the reluctant witnesses, discredit the red herring suspects, and unmask the appropriate miscreant with 100% confidence. As the mystery aficionado will appreciate, I'm excluding from this analogy that outlier, *Trent's Last Case*, wherein a different appropriate miscreant is unmasked with 100% confidence at least 5 times; the detective equivalent of data mining, I suppose.

Despite the obvious parallels between doing detecting and doing statistics, one would be hard pressed to find any statistically literate crime fighter after Sherlock Holmes. That famous sleuth was apparently a data analyst, as recorded by the faithful Dr. Watson in *The Adventure of the Blue Carbuncle*:

> "I can see nothing," said I, handing it back to my friend. "On the contrary, Watson, you can see everything. You fail, however, to reason from what you see. You are too timid in drawing your inferences."

While more modern fictional detectives have run the gamut from Clouseau to Colombo (laughable to lovable) they seem uniformly devoid of statistical knowledge – until now. The single exception would seem to be one Csiss…Doctor Csiss. Hmm, not quite the same resonance as Bond…James Bond. (Now that I think about it, Csiss…..Doctor Csiss sounds an awful lot like a python looking at the latest strain of tasty laboratory mouse in a Charles River Laboratories catalog.) Being that as it may, Csiss IS a statistician, and he has been called in by Scotland Yard to help solve a baffling case in a mystery I just finished, Stanislaw Lem's *The Investigation*.

The case at hand is baffling not so much because people have been murdered on English moors – heaven knows, that has been going on since the Hound of the Baskervilles was haunch high to a Chihuahua. No, what makes the case baffling is that the bodies seem to be coming back to life. (Take THAT, Professor Moriarty!) This returning to the land of the living invites all sorts of scientific, philosophical, and theological speculation in the course of the investigation, and Sciss…Doctor Sciss is one of those doing the speculating. It is not always clear to this reader which of those strands of thought he represents, but his dominant *modus operandi* seems to be represented by this exchange. Csiss speaks:

> "As we have seen, the classical methods of investigation – the collection of evidence and the search for motives—have failed completely. Consequently, I have utilized the statistical method of investigation. It offers obvious advantages. […] Thus, we proceed by preparing a statistical breakdown of all the phenomena. Until now this method has almost never been used in a criminal investigation, and I am very pleased that I now have an opportunity to

introduce it to you gentlemen, together with my preliminary findings…."

With this, Sciss proceeds to calculate a geographic confidence interval, pausing only to toss out an outlier in the form of one body at the Medical School that "doesn't fit this pattern." The Scotland Yard inspector assigned to the case – apparently not a confidence interval sort – recasts Sciss' analytical results in the form of a hypothesis: "Are you trying to tell us," he exploded, "that an invisible spirit of some kind came up out of those damned moors, flew through the air, and snatched the bodies?"

As you may imagine, this is not your typical mystery. At times Sciss seems more a metaphysician than an empiricist, and his arguments seem to border on fantasy, but that's just what Karl Pearson thought about Fisher's methods. So if you pick up *The Investigation* for an afternoon of escape from your statistics assignment, be prepared for something of a different kind of detective story. But then, who knows? Perhaps Csiss' statistical brand of solving crimes will escape from the pages of fiction and we will see his like being interviewed on Larry King!

If so, I hope his real life counterpart has a more euphonious name. Perhaps something like Rumplestiltskin…Cornelius Rumplestiltskin.

References

Bentley, E. C. (1997), *Trent's Last Case*. Mineola, NY: Dover Mystery Classics.

Doyle, A. C. (1994), *The Hound of the Baskervilles*. Mineola, NY: Dover Press.

Lem, S. (1974), *The Investigation*. San Diego: Harcourt Brace & Company.