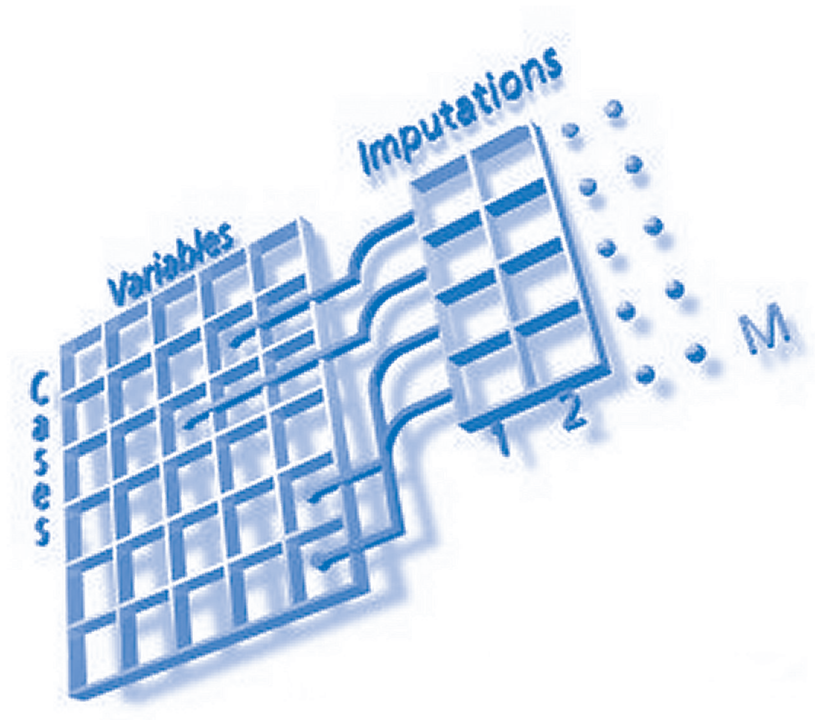
STATS

THE MAGAZINE FOR STUDENTS OF STATISTICS

WINTER 2001 • ISSUE # 30



INSIDE:

Some Advice for Advisors

Student Project: Effects of Color and Gender on Short-Term Memory

Multiple Imputation Using SOLAS for Missing Data Analysis

8th Annual Student Paper Competition

The Biopharmaceutical Section of the American Statistical Association announces its 8th Annual Student Paper Competition for papers presented by students at the 2001 Joint Statistical Meetings in Atlanta, GA, August 5–9, 2001. Double-blind judging of the papers by the Awards Committee is based upon four categories: relevance, contribution, clarity, and applications, as motivated by problems in the different areas of biopharmaceutical research. Papers that include practical examples are particularly valued. A list of previous student winners and paper titles can be found at www.best.com/~asabp/award.htm. Typically three to five awards are given each consisting of a plaque and $1,000. In order to enter the competition and be eligible for an award, the student must be:

- An ASA member (or must join at the time of abstract submission).
- A degree candidate (bachelors, masters, doctorate) during the 1999-2000 and/or 2000-2001 academic year.
- The first author of the abstract and paper that is submitted for presentation at a contributed paper session of the Biopharmaceutical Section.
- Willing to attend and present the paper at the 2001 JSM.

Submit the abstract to ASA by February 1, 2001 (use the official ASA form and check off the Biopharmaceutical Section box).

Submit the manuscript and endorsement of the advisor or department head indicating the student's contributions to the paper by May 1, 2001, to the Biopharmaceutical Section Program Chair:

Keith A. Soper, Ph.D.
Director, Biometrics Research
Merck & Co., Inc.
WP37C-305
West Point, PA 19486
e-mail: keith_soper@merck.com

STATS

STATS

The Magazine For Students Of Statistics

Winter 2001 • Number 30

Editor

Jerome P. Keating
email:
jkeating@utsa.edu

College of Business
University of Texas at San Antonio
6900 North Loop 1604 West
San Antonio, TX 78249

Editorial Board

Beth Chance
email:
bchance@calpoly.edu

Department of Statistics
California Polytechnic State University
San Luis Obispo, CA 93407

E. Jacquelin Dietz
email:
dietz@stat.ncsu.edu

Department of Statistics
North Carolina State University
Raleigh, NC 27695-8203

Rudy Guerra
email:
rguerra@mail.smu.edu

Department of Statistical Science
Southern Methodist University
Dallas, TX 75275

Robert L. Mason
email:
rmason@swri.edu

Statistical Analysis Section
Southwest Research Institute
San Antonio, TX 78228

Allan J. Rossman
email:
rossman@dickinson.edu

Department of Mathematics
Dickinson College
P.O. Box 1773
Carlisle, PA 17013

W. Robert Stephenson
email:
wrstephe@iastate.edu

Statistics Department
Iowa State University
327 Snedecor Hall
Ames, IA 50011-1210

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 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 Editor, Jerome P. Keating. Addresses of the Editor 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.

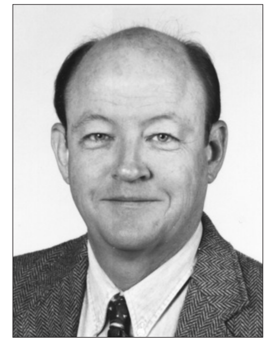
Copyright © 2001 American Statistical Association.

Produced by Oakland Street Publishing • Printed in U.S.A.

Features

Departments

Editor's Column



Jerome P. Keating

As I write this editorial column, we are completing our fall semester and none of us could have foreseen the daunting task of counting ballots in the State of Florida and its national repercussions. In forthcoming volumes, the statistical issues associated with counting, recounting, and manual recounting of ballots will become a common theme. In this edition of STATS, we explore the topic of proper treatment of observations that are missing. To place this topic in the context of an election, what would one do if the ballots from an entire precinct were lost, destroyed, etc.?

How should you analyze data when observations are missing? What assumptions or strategies are involved when values are inserted for missing observations? The insertions are imputed. Fiona O'Callaghan presents an interesting and timely discussion of multiple imputation. She discusses the concepts and strategies behind imputation as well as the use of a major statistical package, SOLAS. I recommend that you try the online guides available through SOLAS to become facile in this emerging area of imputation.

Let me recommend that everyone in our profession read with care and consideration the article entitled, *Some Advice for Advisors.* The article confronts the issues of advising from the eyes of our underrepresented students, i.e., Women, African, Hispanic, and Native Americans.

The authors note that the article evolved from a roundtable discussion at an Eastern North American Regional (ENAR) meeting of the International Biometric Society. They propose that advisors can serve a much greater role for the underrepresented student by acting as mentors. Since the workforce of the future will consist of new workers, who are less likely (15%) to be white males, this is a message that should be embraced by the entire statistics community. Note that seven of the eight authors from this roundtable discussion are members of underrepresented groups. This article reflects the theme of the Joint Statistical Meetings (JSM): *"Celebrate Diversity in Statistics."*

Rudy Guerra, an associate editor of Stats, adds some helpful comments for advisees and an alternative perspective for our readers.

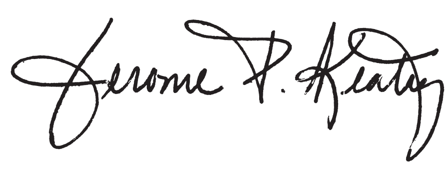
Column Articles

In this issue, we present a *Student Project* by Dominique Shelton of Health Careers High School in San Antonio. She discusses the effects of gender and color in short term memory. This application of statistical methods into the domain of cognitive psychology explores the transfer of sensory storage of a stimulus into short-term memory. To see what effects gender and color have on memory transfer, please read her contribution.

Bob Stephenson in the *AP STATS* column explores the concept of generating sampling distributions of statistics. Bob's discussion, which is directed at our audience of AP Statistics teachers and students, follows on the heels of Albert Madansky's article in the preceding volume. Try the class activities and see new ways to teach and convey fundamental inferential concepts.

The *Outlier…s* column discusses outliers in a variety of common situations. From outliers in elections, to outliers in birth and death rates, to outliers in athletic performances, Allan Rossman presents many new and challenging exercises. Try your hand at estimating the likelihood of some performances of Tiger Woods.

The Biopharmaceutical Section of the American Statistical Association has announced its 8th Annual Student Paper Competition for papers presented by students at the 2001 Joint Statistical Meetings in Atlanta, GA, August 5–9, 2001. Papers that include practical examples are particularly valued. To enter the competition and be eligible for an award, please see the announcement in this issue.



Multiple Imputation using SOLAS for Missing Data

Fiona O'Callaghan

Missing data is a problem that confronts every data analyst. Missing values lead to less efficient estimates because of the reduced size of the database, and standard complete-data methods of analysis no longer apply. For example, analyses such as multiple regression use only cases that have complete data, so including a variable with numerous missing values would severely reduce the sample size.

Different values in a dataset may be missing for different reasons. A laboratory value might be missing because it was below the level of detectability, above the level measurable by the assay, not done because the patient did not come in for a scheduled visit, not done because the test tube was dropped or lost, not done because the patient died or was lost to followup, or numerous other possible causes.

Until recently, the only missing-data methods available to most data analysts have been relatively ad-hoc practices such as list-wise or pairwise deletion, which omit entire records, or pairs of variables, with missing values.

These ad-hoc methods, though simple to implement, have serious draw-backs which have been well documented. For example, list-wise deletion, or complete-case analysis, is easy to implement, but it is inefficient, discarding data is bad, and the complete cases are often a biased sample as there may be systematic differences between responders and non-responders.

Single Imputation refers to any method whereby each missing value in a dataset is filled in with one value, yielding one complete dataset. The important disadvantage of single imputation is that the single value being imputed cannot itself reflect the uncertainty about the actual value. The imputed dataset will fail to provide accurate measures of variability because subsequent analyses would fail to account for missing-data uncertainty.

For example, consider mean imputation: if we have two variables X_1 and X_2, and X_2 contains some missing values, then one possible imputation method would be to calculate the mean of the observed values in X_2 and substitute this for the missing values (see *Figure 1* below).

Imputing the mean would systematically underestimate the variability of the resulting imputed distribution because there would be no residual variance. Also, the correlation between X_1 and X_2 would not be maintained.

Another example is regression imputation, where the missing values in X_2 are imputed by filling in the predicted values from a regression of X_2 on X_1 (*Figure 2*).

When the imputed values all lie on the best fit regression line, the residuals for all of these values would be zero. Also, substituting regression

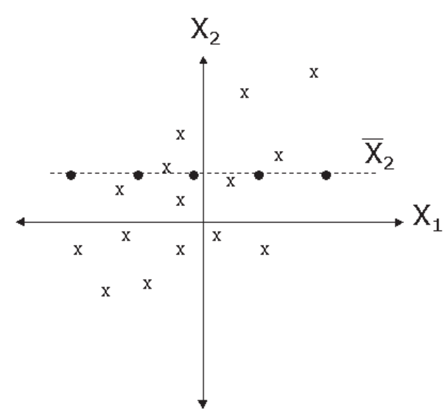


Figure 1. Illustration of mean imputation.

Fiona O'Callaghan has worked as Customer Support Manager with Statistical Solutions Ltd. since 1996. She was extensively involved in the development of SOLAS. She has a masters degree in statistics from University College Cork, Ireland.

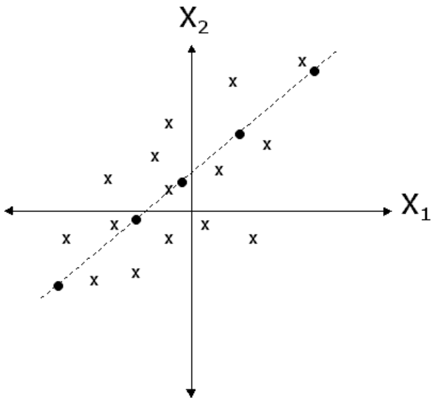


Figure 2. Illustration of regression imputation.

predictions would artificially inflate the correlation between X_1 and X_2.

No matter what the imputation method is—even if missing values could be imputed in such a way that the distributions of variables and relationships among them were perfectly preserved—imputed values are still only estimates of the unknown true values. Any analysis which ignores the uncertainty of missing data prediction will lead to standard errors that are too small, p-values that are artificially low, and confidence intervals that systematically cover less than their nominal coverages.

Multiple Imputation

Multiple Imputation (MI) is a technique that replaces each missing datum with a set of $m > 1$ plausible values instead of just one (see *Figure 3*).

The m versions of the complete data are analysed by standard complete-data methods, and the results are combined using simple rules to yield estimates, standard errors, and p-values that formally incorporate missing data uncertainty.

The variation among the m imputations reflects the uncertainty with which the missing values can be predicted from the observed ones with a set of $m > 1$ plausible values.

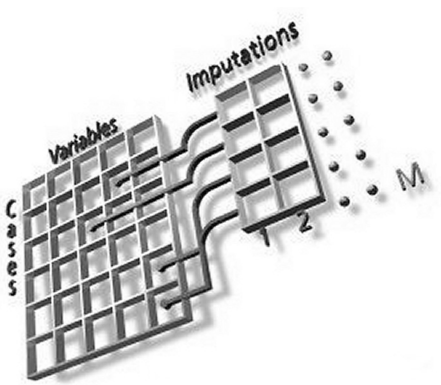


Figure 3. Multiple imputation replaces each missing value

Once the MI's have been created, the datasets may be analysed by any method that would be appropriate if the data were complete. Any analysis would have to be run m times, once for each imputed dataset, and the results across these datasets will vary as a reflection of missing data uncertainty. An overall set of results can be obtained by combining the m sets of results using the rules given by Rubin (1987).

SOLAS for Missing Data Analysis

SOLAS is the only commercially available software package that performs Multiple Imputation. SOLAS 2.0 provides two multiple imputation approaches; a Propensity Score-based approach, and a Model-based approach using multiple linear regression. In both, the multiple imputations are independent repetitions from a posterior predictive distribution for the missing data, i.e., their conditional distribution, given the observed data.

Before the imputations are actually generated in SOLAS, the missing data pattern is sorted as close as possible to a monotone missing data pattern, and each missing data entry is labeled as either monotone missing or non-monotone missing, according to where it fits in the sorted missing data pattern.

A monotone data pattern occurs when the variables can be ordered, from left to right, such that a variable to the left is at least as observed as all variables to the right. For example, if variable A is fully observed and variable B is sometimes missing, A and B form a monotone pattern. Or if A is only missing when B is also missing, A and B form a monotone pattern. If A is sometimes missing when B is observed, and B is sometimes missing when A is observed, then the pattern is not monotone.

A monotone pattern of missingness, or a close approximation to it, can be quite common. For example, in longitudinal studies, subjects often drop out as the study progresses so that all subjects have time 1 measurements, a subset of subjects have time 2 measurements, only a subset of those have time 3 measurements, and so on. Monotone patterns are useful because the resulting imputation is completely principled since only observed/real data are used in the models to generate the imputed values. See Rubin (1987), Chapter 5.

Propensity Score Approach

In the Propensity Score-based approach, cases are grouped according to their probability of being

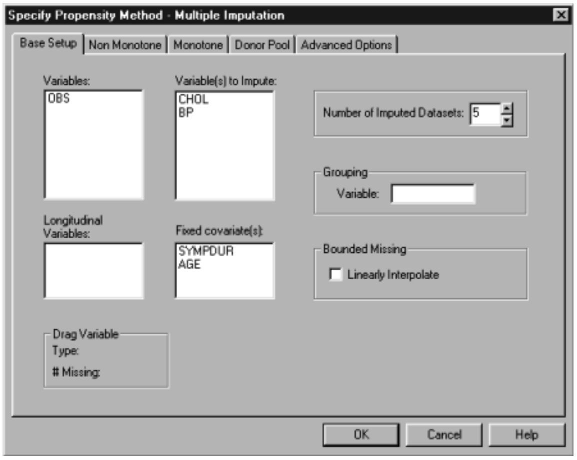


Figure 4. Base Setup tab.

missing (i.e., propensity score) and then observed values are drawn from within each group to impute the missing values. So in general terms, a missing value will be replaced with a value that was observed in another case that had a similar probability of being missing. Or another way of thinking about this is that missing values will be imputed with observed values sampled from cases with similar "histories." When you have several covariates that you want to use for imputation, stratifying on history can be difficult, but using the propensity score in this way is a method of reducing a multivariate stratification into a univariate one.

For each variable that contains missing values, a logistic regression is used to compute a propensity score for every case in the dataset. The dependent variable in the logistic regression will be an indicator variable that corresponds to the presence or absence of the variable that is being

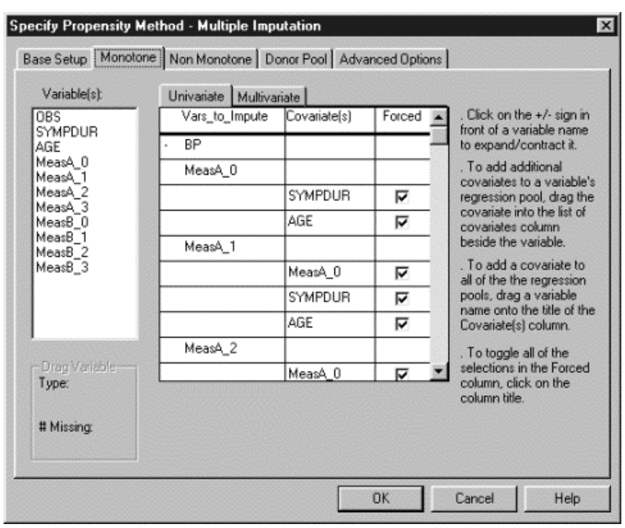


Figure 5. Monotone and Non-montone tab.

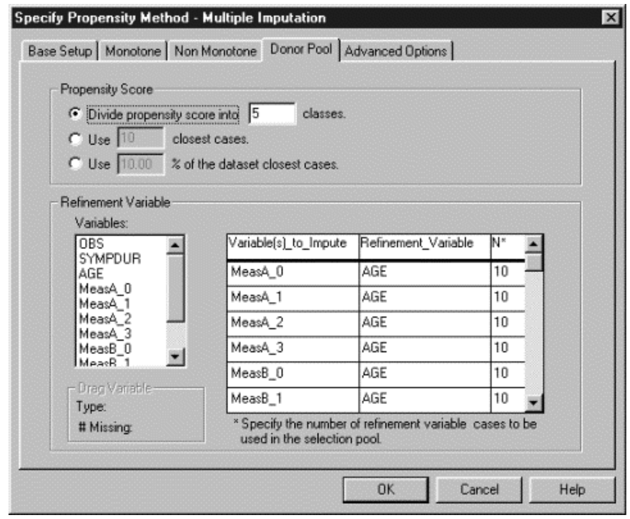


Figure 6. Donor pool tab.

imputed. The independent variables, or covariates, will be a set of variables in the dataset that are predictive of missingness. From the resulting regression equation, a propensity score can be calculated for every case. The dataset is then divided into a user-specified number of subgroups based on propensity score (the default being 5) such that all of the cases within a given subgroup will have a similar probability of being missing. An approximate Bayesian bootstrap is then applied within each subgroup so that the values that are sampled to impute the missing values will come from another case in the dataset that had a similar propensity score.

The Base Setup dialog box allows the user to choose which variables are to be imputed, which variables are to be used as covariates, and also to set options such as the number of imputed datapages (*m*) that are to be created.

The Monotone and Non-monotone tabs allow the user control over the variables to be used as predictors in the regression models.

By default, all of the covariates are forced into the regression models, but the user has the option to "unforce" (switch off the forced option) covariates if he/she wishes. A backward stepping logistic regression is performed, so covariates which are not forced may be removed from the model.

The donor pool page provides the user with more control over which cases will be included in the pool from which the imputed values will be drawn. These donor pools or subgroups are derived from the propensity scores that are calculated for each case, so that within a given subgroup, all of the cases will have a similar probability of being missing.

In addition, if there is a variable in the dataset that is highly correlated with the variable to be imputed, then the donor pool can be further refined by matching on this "refinement variable" within each subgroup.

Model-based Approach

In the model-based approach, the predictive information contained in a user-specified set of covariates is used to predict the missing values in the variables to be imputed. First, a linear regression is estimated from the observed data. Using this estimated model, a new linear regression model is randomly drawn from its Bayesian posterior distribution. This randomly drawn model is then used to generate the imputations, which include random deviations from the model's predictions. Drawing the model from its posterior distribution ensures that the extra uncertainty about the unknown true model is reflected.

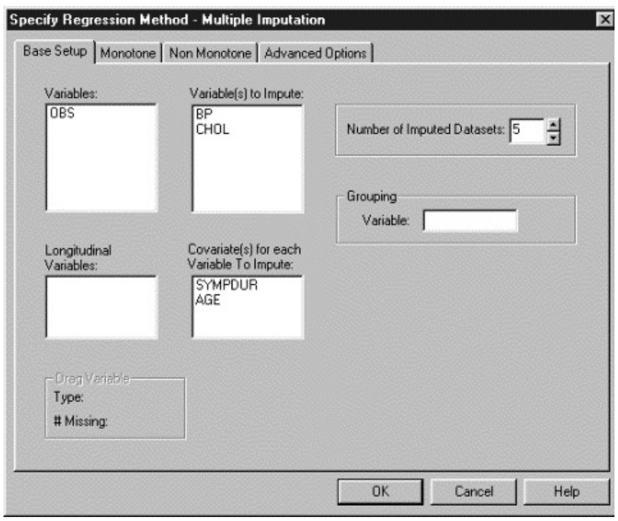


Figure 7. Base Setup tab.

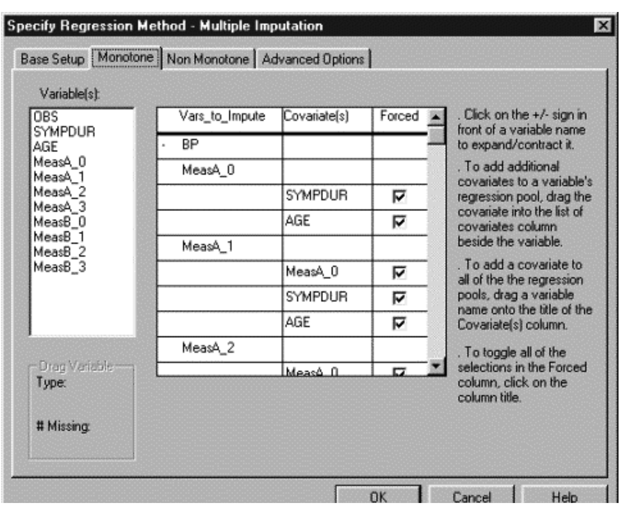
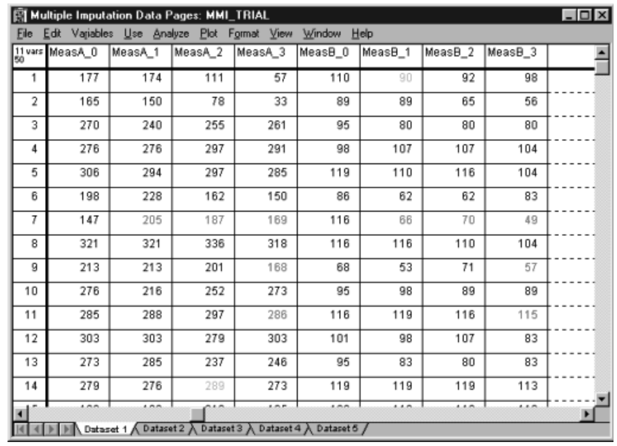


Figure 8. Monotone tab.



11 vars 50	MeasA_0	MeasA_1	MeasA_2	MeasA_3	MeasB_0	MeasB_1	MeasB_2	MeasB_3
1	177	174	111	57	110	90	92	98
2	165	150	78	33	89	89	65	56
3	270	240	255	261	95	80	80	80
4	276	276	297	291	98	107	107	104
5	306	294	297	285	119	110	116	104
6	198	228	162	150	86	62	62	83
7	147	205	187	169	116	66	70	49
8	321	321	336	318	116	116	110	104
9	213	213	201	168	68	53	71	57
10	276	216	252	273	95	98	89	89
11	285	288	297	286	116	119	116	115
12	303	303	279	303	101	98	107	83
13	273	285	237	246	95	83	80	83
14	279	276	289	273	119	119	119	113

Figure 9. Multiply Imputed datasets.

The Base Setup dialog box is much the same as that for the Propensity Score-based approach. It allows the user to decide which variables to impute, and which variables to use as predictor variables, or covariates.

The Monotone and Non-monotone tabs allow control over the variables to be used as predictors in each of the regression models.

Again, the default is that all covariates are forced into the regression model. Variables can be removed from the regression model by just dragging the variable from the list of covariates, back to the Variables list on the left-hand side of the dialog box.

Once the multiple imputation has run, the imputed datasets appear with the imputed values appearing in blue. The default for the number of imputations is 5, but this can be set to between 2 and 10 imputations. Each of these datasets can be saved for later analysis or exported to any of a variety of other statistical packages.

Multiple Imputation Combined Statistics

When the multiple imputation has run, the user will have multiple complete datasets instead of just one. The idea behind multiple imputation is that any analysis should be performed multiple times (once for each imputed dataset) and then the results of these analyses are combined to give one overall set of results (repeated-imputation inference) which formally incorporates the missing data uncertainty.

The combined estimate of any parameter of interest, θ, for a particular variable is simply the mean of the estimates from each of the *m* imputed datasets. For example, the combined estimate of the mean of a particular variable, or a regression coefficient in a model, is simply the mean of the estimates for that parameter across the *m* imputed datasets. To estimate the variance of the combined

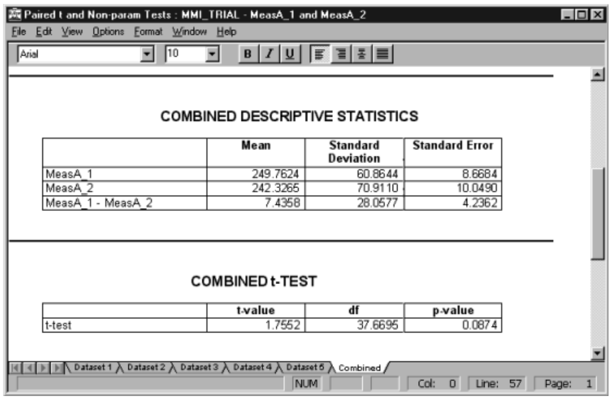


Figure 10. Combined output.

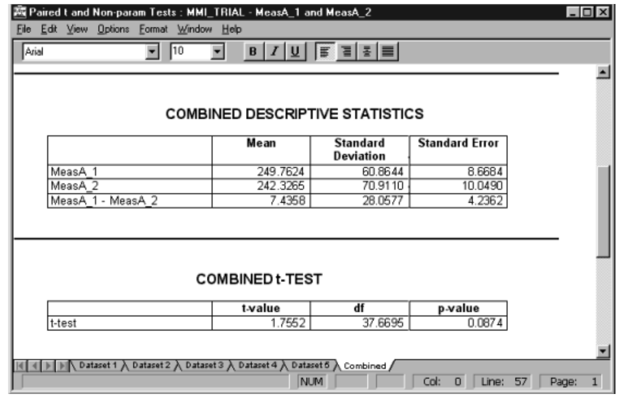


Figure 11. Script language window.

parameter estimate, you must combine the variance that is estimated from the combined parameter estimates from within each imputed dataset, with the variability of the estimate across the *m* imputed datasets, in other words, the within imputation variance and the between imputation variance.

In SOLAS 2.0, any analysis performed on a set of multiply imputed datasheets will be automatically combined and presented in a Combined Output Window. For example if you run a two group t-test, you will get *m* sets of t-test results, plus an overall combined set which is the result of pooling across the *m* pages.

Script Language Facility

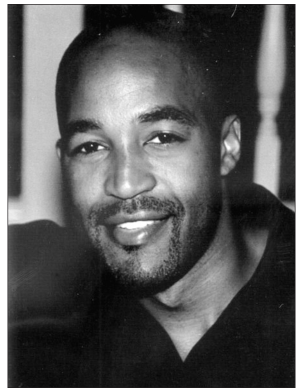
When you run a multiple imputation in SOLAS 2.0, the selections made are recorded and the corresponding commands are written into an imputation window that can be viewed and modified. This ability to save and access imputation set-ups can help to simplify the documentation when submitting to a regulatory agency. It also facilitates running simulations, as you can save your imputation set-up and run the same multiple imputation on different datasets.

A tutorial of SOLAS is available for download from *http://www.statsolusa.com,* and a demo version is also available on request.

References

1. Rubin, D. (1987), *Multiple Imputation for Nonresponse in Surveys*, Wiley, New York.
2. Rubin, D. and Schenker, N. (1991), “Multiple imputation in health-care data bases: an overview and some applications”, *Statistics in Medicine*, 10, 585–598.
3. Lavori, P., Dawson, R. and Shera, D. (1995), “A multiple imputation strategy for clinical trials with truncation of patient data”, *Statistics in Medicine*, 14, 1913-1925.
4. Rubin, D. (1996), “Multiple imputation after 18+ years”, *Journal of the American Statistical Association*, 91, 473-489.
5. SOLAS for Missing Data Analysis v. 2.1, Statistical Solutions, Stonehill Corporate Center, Suite 104, 999 Broadway, Saugus MA 01906 (*http://www.statsolusa.com*).

Some Advice for Advisors



DuBois Bowman

Angelita Bush

David Cowart

Dionne Price

Carol Redmond

Louise Ryan

June Stone-Chestnut

Belinda Young

It is no secret that graduate school can be a grueling and difficult experience. Most who have been through it already look back with a mixture of emotions: fond recollections of time spent with fellow grad students, wistful memories of what it was like to actually have time to read a journal article, perhaps a shudder at the all too common experience of feeling inadequate for the task of finding a good topic and completing a respectable thesis. For some students, making the transition to graduate school can border on traumatic. This can especially be the case for students who have been out in the workforce for several years and away from their study habits. The transition can also be difficult for students coming straight from a small undergraduate college to a large graduate institution. Coping with the relatively impersonal environment of a large graduate institution can be difficult when such students are used to a more supportive and nurturing setting. All these difficulties with the transition to graduate school can easily lead to a sense of failure. Adjusting to graduate school can also be especially difficult for

---

DuBois Bowman graduated with a Ph.D. in Biostatistics from University of North Carolina, Chapel Hill, in 2000, and is now an Assistant Professor in the Department of Biostatistics at Emory University. Angelita Bush graduated with a Masters degree in Biostatistics from University of North Carolina, Chapel Hill, in 2000 and now works at Georgetown University, DC. June Stone-Chestnut graduated with a Masters degree in Statistics from Temple University in 1993 and now works at Wyeth-Ayerst Research Laboratories. David Cowart graduated with a Ph.D. in Biostatistics from Emory University in 2000, and now works at PricewaterhouseCoopers in Boston. Dionne Price is a graduate student in the Department of Biostatistics at Emory University. Carol Redmond is a Professor of Biostatistics at University of Pittsburgh. Louise Ryan is a Professor of Biostatistics at Harvard School of Public Health. Belinda Young graduated with her M.S. in Biostatistics from Emory University in 1998 and now works at the University of California at San Francisco in the Department of Medicine, Divisions of Cardiology and Nephrology.

students who are underrepresented minorities, especially those coming from small undergraduate schools which have a predominantly minority enrollment, for example, HBCUs (Historically Black Colleges and Universities). Along with the difficulties in making the transition from a more personal and supportive environment, underrepresented minority students face additional issues such as isolation. This is especially true in fields like statistics and biostatistics where there are few minority role models, and where it is easy for students to find themselves in a situation of being the only minority student in the classroom.

These issues were discussed in depth at a round table luncheon sponsored by the Eastern North American Region (ENAR) of the International Biometric Society at the 1999 Spring meetings in Atlanta. There were eight of us attending the lunch, including two faculty members, one professional from the pharmaceutical industry and six students. Five of us were members of an underrepresented group. Our discussion quickly turned to the role of the graduate advisor. A good advisor can make all the difference in helping a student to adjust and make a successful transition to graduate school. Unfortunately, however, it seems that all too often, graduate school academic advisors seem to see their role as simply signing off on their students' course plan. Our assumption is that most advisors would be more than willing to help, if only they were aware of the issues facing their advisees and of the important role that they could play. At our luncheon discussion, we were easily able to generate the following list of things that schools and advisors can do to help new students:

1. There is no reason why graduate school academic advisors cannot be assigned in the spring at the time a student is admitted. It would be helpful if prospective advisors could contact their advisee by phone and/or email to begin a dialogue well before the student starts classes in the fall. Alternatively, the department or the advisor could send a letter informing the student of the name, e-mail address, and phone number of his/her academic advisor. This letter could encourage the student to call or e-mail their advisor to begin a dialogue prior to the start of classes in the fall. The letter could also include tips (i.e., contacting other students in the department for advice over the summer, obtaining course syllabi for perusal, reviewing texts used in first semester classes) to ensure a successful beginning to their graduate career.

2. The first thing a prospective advisor should do is go over the student's undergraduate record to identify any potential gaps. Has the student had enough linear algebra and calculus? For more theoretical programs, does the student need to learn some additional analysis? Does the student have good computing skills? Are there any poor grades, perhaps indicating the need for some brushing up of skills? The next step will be to discuss these issues with the student. It is also a good idea for the advisor to ask for more detail about specific courses. Calculus III, for example, can vary considerably from school to school! The advisor should ask the student in depth about important prerequisite courses. For example, asking what the topics were and what book was used for each course. If this evaluation reveals any gaps, the advisor might work with the student to identify a suitable program of reading or study over the summer. For example, perhaps the student might benefit by reviewing a linear algebra text. A more mathematically prepared student might benefit from reading an applied statistics book that introduces them to statistical applications. In some cases, a student might decide to enroll for summer school.

3. When it comes to picking courses for the fall semester, advisors should take the time to make sure that their advisee is prepared for the standard set of classes. In some cases, a student may be much better off with a slower start that allows them to fill in gaps in their background. It is much better to identify gaps and explore alternative course sequences at the beginning of the semester, rather than waiting until half-way through a course when a student may already be in serious trouble.

   We encourage the advisor and the student to be conservative when selecting courses for the fall semester. Conservative course selection can be especially important for minorities or other students whose transition into graduate school may involve more difficulties than the normal challenges faced by all students. The initial semester of graduate school often sets the stage for the entire experience of course work and possibly dissertation work. Although there may not be a great deal of flexibility in the selection of fall courses, the course schedule should be constructed on an individual basis and should maximize the chance of excellent performance. For the student with apparent gaps or deficiencies in undergraduate mathematics and statistics courses, perhaps a program of course work should be developed that allows the student to complete a normal two-year sequence in two and a half or three years. The student with average preparation may reduce a typical load of four courses to three courses for the first semester. Finally, the student with superior background may take the usual set of courses or even an accelerated set of courses, if appropriate.

The costs associated with poor performance from taking a course load that is too ambitious during the fall semester may include severely reduced confidence, feelings of inadequacy, feelings of intimidation toward professors and possibly classmates, and insufficient understanding of basic material that will affect future performance. These are only a few of the potential effects of a poor start in graduate school that often lead to an overall unpleasant experience. In addition, these effects immediately become obstacles to continuing on to advanced graduate work or perhaps other career plans. Our recommendation of conservatively selecting fall courses is thus based on the belief that the ramifications are far less severe than the consequences of poor performance during the first semester of graduate training.

4. All of us at the round table lunch felt strongly that the role of the academic advisor should not stop at signing the course schedule each semester. Ideally, the academic advisor should be a mentor— someone the student can grow to trust and rely on for advice and encouragement. Having a caring and reliable mentor can be particularly important for a minority student who might be feeling isolated in the department or whose confidence might be affected by the kinds of subtle racism that can often exist at large majority institutions. Advisors should be encouraged to see their advisee regularly throughout the semester, especially in a student's first year. Although it can be difficult sometimes to draw a student out, advisors should try to communicate to the student that they *really* want to know how the student is doing. Often, unless an advisor pushes a little, many students will simply answer "fine" when asked how things are going. A student needs to be reassured that the advisor is there to help and will not judge the student in a negative way if the student expresses that they are struggling. Sometimes, it might help to talk away from the office setting. For example, talking over lunch can be a good way to break the ice and encourage the student to talk more freely about how they are doing.

   Academic advisors can also play an important role in terms of providing resources for their advisees. For example, an advisor might make sure that the student knows where to go to find out about health insurance, housing, childcare, etc. Minority students may be interested to know about on-campus minority student associations. While the advisor may not necessarily have all this knowledge themselves, they should be prepared to refer the student to people who do.

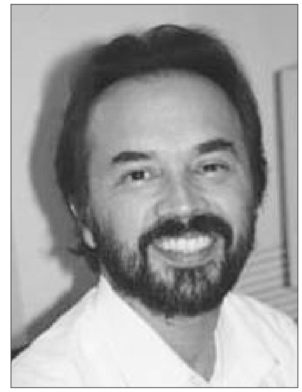
5. The academic advisor should encourage the student to actively seek out study groups and assistance from instructors in addition to doing independent work. Independent work is obviously critical to student development and is one of the primary goals of graduate training. However, regular communication with professors can not only contribute to a student's understanding of course material, but it can also begin to eradicate some of the perceived barriers between students and faculty. Fellow students can be one of the most valuable resources available to a graduate student. During the initial phases of graduate training, when relationships between students have not been well developed, students should be encouraged to take an active role in establishing or participating in study groups. A sensitive advisor can play an important role in helping a student to seek out and establish effective collaborations. For example, an advisor can help a student identify a group involving the right level of challenge, as well as peer support. Sometimes a student may need to try out several different study groups before finding the right one. As relationships between students progress over time, such student collaborations will arise more naturally.

The issues we have raised in this article represent just a few of the ways that a school can help ease a student's path through graduate school. Often, a little extra effort can go a long way in improving a student's life. The academic advisor can play a key role, but there is potential for developing a mentoring relationship that can be rewarding to both student and advisor.

Further useful tips for advisors can be found in "Advisor, Teacher, Role Model, Friend," published by the National Academy Press, Washington, DC.

Comments on Some Advice to Advisors

A Word to the Advisees



Rudy Guerra

Graduate school is a grueling experience with rigorous courses, qualifying exams, thesis writing, and stress. Additional factors such as coming from an underrepresented group make the entry and adjustment to graduate school even more difficult. A faculty advisor or mentor in a seemingly impersonal environment can make all the difference in the world to someone who's feeling isolated. Although it may appear that professors are insensitive, the fact is that most want to see students succeed. For this reason, I provide brief commentary on some issues raised by Bowman et al. Few of us have forgotten what graduate school was like (sleepless, as I recall). We spend a lot of time discussing graduate student issues, both academic and personal, and it is very important that students know this.

Bowman et al. make very good suggestions about how academic advisors can help students have a more positive experience in graduate school, especially in the first year:

- Communicate with new students prior to matriculation.
- Advise students on potential gaps or weaknesses in their backgrounds.
- Advise thoughtfully on course selection and course load.
- Provide regularly scheduled mentoring beyond courses.
- Encourage student to participate in formal and informal departmental activities, especially study groups.

Their points are very important and demand serious attention by department faculty. Although we may not be doing as well as we, or our students would like, we are in fact quite aware of these issues. Points 1-3 are components of the graduate admissions process at the departmental level. Of particular interest is the student's background and suitability for a given program. During the admissions process, much time is spent discussing the probable success of applicants. In most departments, students with special circumstances are strongly advised and often required to follow an academic plan tailored to increase their chances of success. An advisor per se is not yet assigned but faculty members involved with recruiting communicate regularly with applicants. This is the first step in establishing a working relationship with a student, whereby weaknesses, strengths, and potential programs are outlined for and with the student. Upon admission and arrival to a new environment most students will have at least one faculty member, who is familiar with their background.

The potential benefits of points 4 and 5 depend on the defined duties of academic advisors, how advisors are assigned to graduate students, and the personalities of the individuals involved. Sometime during the first year, students will become comfortable with at least one classroom professor, who can and often does serve as an academic advisor and mentor. It is mainly through these bonds that students receive most direction. Many of these self-selected mentors are aware of the newly formed relationships, realize they are a vital part of the student's academic life, and nurture them accordingly. In many instances this mentor will be the unofficial spokesperson for the student in faculty meetings, TA assignments, and so forth. I agree with

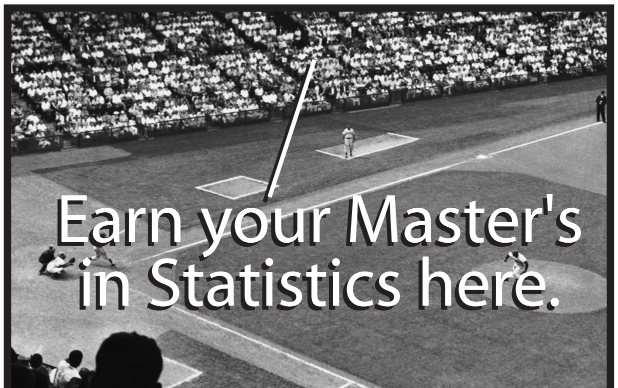
Rudy Guerra was a graduate student in the department of statistics at the University of California, Berkeley and is currently Associate Professor of statistics at Rice University. For several years he served as member and chair of the admissions committee at Southern Methodist University, where he taught for 9 years.

Points 4 and 5 and I suspect that the advisor-advisee relationship could be more beneficial by more thoughtful assignments. Having students meet with their advisors or mentors on a regular basis (2-3 times a year) can be immensely helpful to both parties. It is important for the student to know, however, that this is a two-way street and they, too, have a responsibility to communicate. Indeed, if not asked to come in, as Bowman et al. note, many students will not seek out advice they wish they were getting. Although not specifically noted, feedback should be a component of point 4. It is absolutely critical that students receive honest feedback, be it from their mentors, graduate advisors, or department chairs. Systematic yearly reviews (preferably written) monitor progress, while frank opinions and recommendations can prevent difficulties or hardships down the road.

Let me comment about minority affairs as they relate to the topic at hand. As a Mexican-American, I've had my share of "minority experiences" as a graduate student. As a faculty member, I've also had the benefit of observing minorities apply to, gain admission to, and progress through graduate school. For various reasons there are added difficulties for many of us from minority backgrounds. It is unlikely that you will get good advice on such minority matters from your advisor or mentor. In large part because they won't be minority, but also because most have not been trained to detect problems related to minority issues or it simply is not in their comfort zone. I wouldn't be too disappointed or angry if they don't broach the subject with you; at best expect an empathetic ear. It is a difficult issue that needs attention. Articles such as that by Bowman et al. alert departments about concerns, but I do not see much changing in the near future until the composition of the faculty changes.

There are those (students and faculty) that do not wish to claim their minority identity or attribute anything to it. Perhaps this is one reason why people don't bring it up. Why should an advisor or mentor presume that I want to discuss the difficulties of being Mexican-American in graduate school just because my last name is Guerra? There's a lot of gray area here, no doubt. You will have to navigate this one very carefully. Most universities have various minority organizations and there will be faculty across campus that have a reputation of being sensitive to the issues. That may be a good place to start if you're not sure how to proceed with a minority related concern.

Just like graduate students, faculty and departments vary in their personalities; many serve students well. Bowman et al. have some very good advice for advisors and I hope that after reading my article students know that for the most part we are aware of the issues. However, actions speak louder than words and I appreciate the reminder that many of my own concerns as a graduate student belong to every new generation of students. The next time a student walks into your office ask them how they're *really* doing. Or, the next time you walk into your advisor's office, ask him or her if they're up for a coffee break.



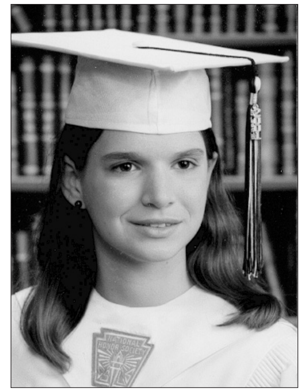
Or in your office. Or at home on your sofa. With Colorado State University's distance education program you can earn your Master's anywhere, using a combination of videotape and on-line resources.

Call (Toll free) 877-491-4336 for information.

Colorado State University

New courses now available.

The Effects of Color and Gender on Short-term Memory



Dominique Shelton

1. Background

An adequate memory is vital to the life-long process of learning. Cognitive psychologists have researched information processing greatly over the past 20 years. As a result, our understanding of this complex topic has dramatically increased. Regarding memory, people are most interested in how their memory may be improved.

Three distinct stages of memory have been identified: sensory storage, short-term memory, and long-term memory. Sperling (1960) conducted classic studies in the area of sensory storage that demonstrate that humans have an extremely accurate and complete memory for visual stimulation—but it is very brief. In order to increase the chances of transferring a stimulus to memory, the image must be salient. The more salient an image is, the greater chance it has of transferring into the short-term memory, and eventually the long-term memory. Peterson and Peterson (1959) studied short-term memory and demonstrated the impact of interference. When subjects were presented a stimulus, required to do something else (an interference task), then asked to recall the stimulus (retrieve it from short-term memory) their performance was far less accurate than for subjects not given the interference task. This leads to the serial position effect where in a list of stimuli (such as a set of numbers) the first and final items are more likely to be recalled than the middle.

Many factors may influence the saliency of a stimulus, and thus the chances of forming a memory. This study was concerned specifically with the color of the stimulus and the gender of the subject. Linear models were used to analyze the data (Ward & Jennings, 1979).

Dominique Shelton, Health Careers High School, San Antonio, Texas. (Dom_Shelton@hotmail.com) This project won 1st place in the 1999 American Statistical Association Project competition, 1st place at the San Antonio Alamo Regional Science Fair (in the Behavioral category), 2nd place at the Texas State Science Fair (in the Behavioral category), and 2nd place at the Texas Junior Academy of Science.

2. Project

The objective of this experiment was to determine the effects of color (blue, green, red) and gender (male, female) on short-term memory. Of the 30 high school students participating in the experiment, 15 were male and 15 were female. The first set of stimuli included a short string of seven random two-digit numbers (DATA7). The second set of stimuli included a long string of 14 random two-digit numbers (DATA14). Black was the control color, and blue, green, and red were the test colors. For both stimuli conditions the test numbers were presented initially in black, and then the subject immediately was given a test where the numbers were presented in blue, green, or red. The experiment was administered and the data were analyzed using Visual Basic, Excel and SYSTAT 6.0.

Of the 15 subjects for each gender, five subjects were presented blue-colored numbers, five were shown red-colored numbers, and five were given green-colored numbers. This assignment was random until each color cell was full. All of the subjects performed the control task with black color. Each subject first selected his/her gender and grade level on the computer. Then they clicked "Continue" to begin the experiment. This presented them with seven random two-digit numbers for ten seconds. After ten seconds, the screen automatically changed to the next one. The subjects then recalled as many of the numbers as they could remember. They were instructed to fill in all of the blanks, even if they were required to guess. When the subject filled in the last box, they were automatically sent to the next screen where they indicated how confident they were with their answers on a scroll bar (one end indicating "Not Confident" and the other "Confident"). When the subject was done, they clicked "Continue" to repeat the procedure with the next set of numbers. Every subject took a control and colored test for DATA 7 and DATA14.

The response variable is the measure of accuracy of memory (the score). The score of the colored numbers less the score of the black (control) numbers produces a difference score. The primary interest is determining if color makes a difference in the score, determining if gender makes a difference in score, and determining if there is interaction between color and gender. By interaction, I mean that the differences between the expected score for males and females are not the same for all three colors.

3. Methodology

I used a general linear model approach to investigate the research question of interest. This involves first developing an ASSUMED model that allows for the investigation of the hypothesis of interest. The natural language hypothesis implies certain restrictions on the parameters of the assumed model resulting in a RESTRICTED model.

The first step in creating the statistical models is to define the following vectors:

SCORE = the number of correct responses using the test stimuli (colored numbers) minus the number of correct responses using the control stimuli (black numbers). This yields a positive value if the color improved memory or a negative value if the color hindered memory.

U = 1 for all 30 elements

FEMALE = 1 if the corresponding element of SCORE is from a Female; 0 otherwise

MALE = 1 if the corresponding element of SCORE is from a Male; 0 otherwise

GREEN = 1 if the corresponding element of SCORE is from a Green stimulus; 0 otherwise

BLUE = 1 if the corresponding element of SCORE is from a Blue stimulus; 0 otherwise

RED = 1 if the corresponding element of SCORE is from a Red stimulus; 0 otherwise

MG = MALE*GREEN

MB = MALE*BLUE

MR = MALE*RED

FG = FEMALE*GREEN

FB = FEMALE*BLUE

FR = FEMALE*RED

With the collected data, the following five models were created:

1. $SCORE = b_1*U + b_f*FEMALE + b_r*RED + b_b*BLUE + E_1$
2. $SCORE = b_2*U + E_2$
3. $SCORE = b_{mg}*MG + b_{mb}*MB + b_{mr}*MR + b_{fg}*FG + b_{fb}*FB + b_{fr}*FR + E_3$
4. $SCORE = b_m*MALE + b_f*FEMALE + E_4$
5. $SCORE = b_b*BLUE + b_g*GREEN + b_r*RED + E_5$

where the b_j represent the unknown model coefficients, and the E_j represent the random error terms.

Model 1 is the No-interaction model. This model is used to test the hypothesis that the differences between the expected values (MEANS) for MALES and FEMALES are the SAME for ALL THREE COLORS.

Model 2 is the Unit vector model of all 1s. The least-squares coefficient of U is the GRAND MEAN of the elements of the dependent vector, SCORE.

Model 3 contains predictor vectors for all SIX mutually exclusive categories. The SIX least-squares coefficients of this model are the AVERAGES (MEANS) of the SCORES for the five subjects in each of the SIX categories.

Model 4 contains TWO predictor vectors for the GENDER attribute. The TWO least-squares coefficients of this model are the AVERAGES (MEANS) of the SCORES for the 15 Males and fifteen Females.

Model 5 contains THREE predictor vectors for the COLOR attribute. The THREE least-squares coefficients of this model are the AVERAGES (MEANS) of the SCORES for the ten subjects in each of the BLUE, GREEN and RED categories.

These models are compared through the use of the F statistic:

$$F = \frac{\dfrac{\text{(sum of squares of error of restricted - sum of squares of error of assumed)}}{\text{(\# of predictors in assumed - \#of predictors in restricted)}}}{\dfrac{\text{(sum of squares of error of assumed)}}{\text{(\# of cases - \# of predictors of assumed)}}}$$

While graphs give us a graphical representation of the means, the F statistic can be used to make statistical decisions about the significance of the hypotheses.

With the models, the following hypotheses were investigated for each set of test conditions (i.e., DATA 7 and DATA14):

Hypothesis 1: No interaction (There are constant differences between MALE-FEMALE MEANS across the three colors.)

- Assumed Model = six-vector Model (3)
- Restricted Model = four-vector, No-interaction Model (1)

Hypothesis 2: Assuming No interaction, there are no differences between the MALE-FEMALE means across the three colors.

- Assumed Model = No interaction Model (1)
- Restricted Model = Color Model (5)

Hypothesis 3: Assuming No interaction, there are no differences between the BLUE, GREEN and RED means for both the MALE and FEMALE categories.

- Assumed Model = No interaction Model (1)
- Restricted Model = Gender Model (4)

4. Results

Overall, a blue color for the digits improved scores while a green or red color generally lowered scores. The average score for the DATA7 tests, for males, was 2.2 for blue digits, 0.0 for green digits, –0.2 for red digits; for females the average score was 0.8 for blue digits, –1.6 for green digits, –1.6 for red digits. The average score for the DATA14 tests, for males was 2.8 for blue digits, -0.6 for green digits, –0.8 for red digits; for females the average score was 2.8 for blue digits, –2.2 for green digits, 1.2 for red digits (see Figure 1).

DATA7

The hypothesis of no interaction between gender and color failed to be rejected for DATA7 tests (i.e., F = 0.02, p=0.9814). You can see from Figure 2 that no interaction is present. Because the hypothesis of no interaction failed to be rejected, the no-interaction model (1) was used as the Assumed Model for the DATA7 tests. The test of no gender differences was rejected (F= 9.79, p= .0043) and the test of no color differences was rejected (F= 11.18, p=.0003).

DATA14

The hypothesis of no interaction was rejected for the DATA14 test (i.e, F= 3.97, p=.0325). We can observe this conclusion from the six means shown in Figure 2 and the results given in Table 1.

Blue improved scores for both genders. I noticed that it was improved most notably in the beginning of the string of numbers (between positions two and four of the sequence). In the control (the black string of numbers), 10% of the subjects gave a correct response for position four. In the blue stimulus, 70% gave a correct response. Red was inconsistent, decreasing the score for males but increasing or decreasing the score of females. Green, on the other hand, decreased scores across all categories.

Discussion

The DATA14 tests may be a more reliable measure of short-term memory because the longer string of numbers allows for more variability

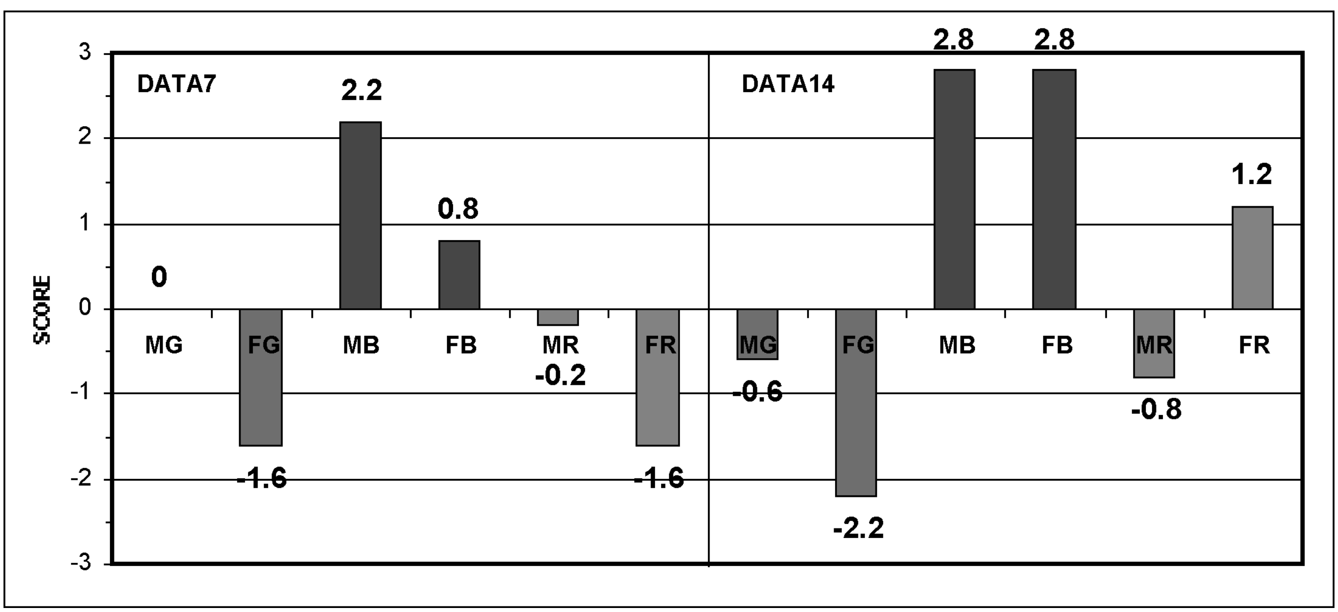


Figure 1. Mean Scores

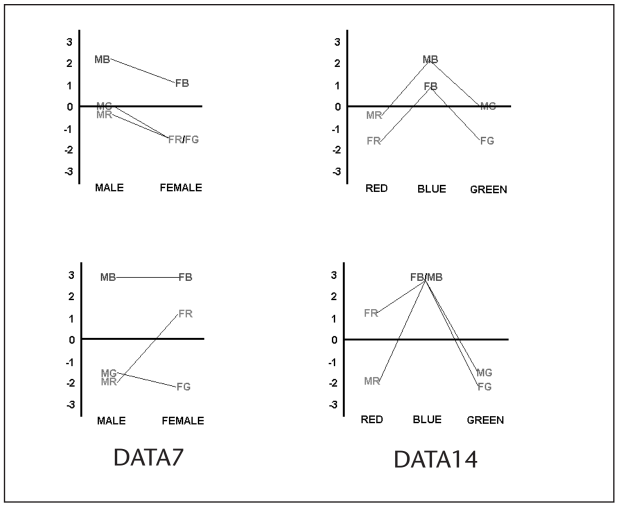


Figure 2. Means for DATA7 and DATA14

(variance for data7 = 3.30, versus 5.36 for data14). Many subjects were able to remember most of the numbers in the string of 7, but not in the string of 14. By presenting more numbers than the subject is capable of remembering, we allow for improvement.

It is interesting to note that of the three colors, blue has the lowest level of relative energy (450 to 500 nm) and green has the highest relative energy. It would be interesting to investigate whether there is a relationship between the wavelength of the color and short-term memory. It is known that in fast food restaurants and other places, bright colors such as red and bright green are used to get people in a “hurry up and spend money” mood. On the other hand, soft colors or pleasant shades of blue are not used because these colors tend to have a calming effect. This could be applied to the concept of short-term memory or the process of learning in general.

5. Conclusion

The data demonstrate a significant relationship between color, gender, and short-term memory and shows the potential for future research. In general, females are affected more by color than males. Overall, blue color appears to improve short-term memory while the colors red and green generally hinder it. Longer strings of numbers allow for more variability and so are more reliable measures of the effects of certain factors on short-term memory.

6. Future

In the future, more subjects will have to be tested over a wider range of colors to verify the validity of the results. With the results, there are many possible applications for studies involving short-term memory in the field of psychology. For example, there is great interest in learning how to study more effectively and remember more reliably. If there were a simple method to improve memory, such as writing with a certain color pen, then it would be easy for many people to benefit from it.

7. References

Best, J. B. (1992), *Cognitive Psychology.* St. Paul: West Publishing Company.

Berk, Kenneth N. (1998), *Introductory Statistics with Systat*. New Jersey: Prentice Hall.

Miller, G. A. (1956), The Magical Number Seven, Plus or Minus Two: Some limits on our capacity for processing information. *Psychological Review,* 63, 81-97.

Moore, David S. (1995), *The Basic Practice of Statistics.* New York: W.H. Freeman & Company.

Peterson, L.R., & Peterson, M.J. (1959), Short-term retention of individual verbal items. *Journal of Experimental Psychology,* 58, 193-198.

Sperling, G. (1960), The information available in brief visual presentations. *Psychological Monographs*, 74, 498.

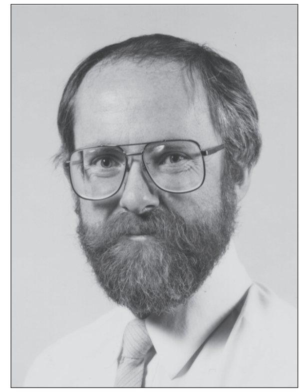
Ward, Joe H. & Jennings, Earl. (1979), *Introduction to Linear Models.* Englewood Cliffs, NJ: Prentice-Hall, Inc.

8. Acknowledgements

I would like to thank all those who supported me in this study, including Dr. Joe Ward Jr. for helping with my statistical analysis, Dr. Robert L. Mason for reviewing my paper, the subjects for participating in the study, and my family for their unwavering support.

Table 1. F Statistics of DATA7 and DATA14

Model	F value	Prob. F	Conclusion
DATA7 No-Interaction	0.02	0.9814	FAIL TO REJECT
DATA7 Gender Equal	9.79	0.0043	REJECT AT .05
DATA7 Color Equal	11.18	0.0003	REJECT AT .05
DATA14 No-Interaction	3.97	0.0325	REJECT AT .05



W. Robert Stephenson

Sampling Distributions

The sampling distribution of a statistic is one of the most difficult concepts encountered by students in an introductory statistics course. It is also one of the most fundamental. Without an understanding of sampling distributions, statistical inference becomes a mysterious mix of formulas and steps learned by rote. Sampling distributions are hard to understand because there is so much going on. Students must grapple with not just one idea, but several.

This article presents a series of activities, some in class and some out of class, that allow students to experience and explore the sampling distribution of the sample mean. The in class activity is presented using a hypothetical class of 25 students. However it can easily be adapted to both larger and smaller class sizes. The out of class activity involves the use of the internet and a Java applet developed by David Lane as part of the Rice Virtual Lab in Statistics at Rice University (*www.ruf.rice.edu/~lane/rvls.html*).

In Class Activity

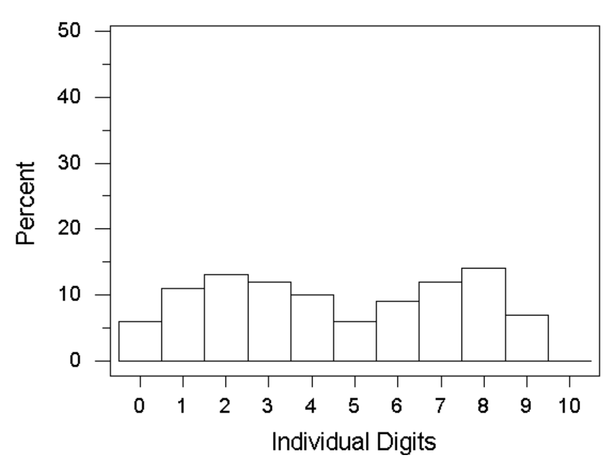
The understanding of sampling distributions begins with a clear distinction between a population and a sample. The population consists of all items of interest, where a sample is simply a few of those items selected from the population. One problem with in class activities is having materials conveniently available. For this activity the students provide the materials by way of their telephone numbers. Specifically, the last four digits of their telephone numbers. Below is a hypothetical class of 25 students and the last four digits of their telephone numbers.

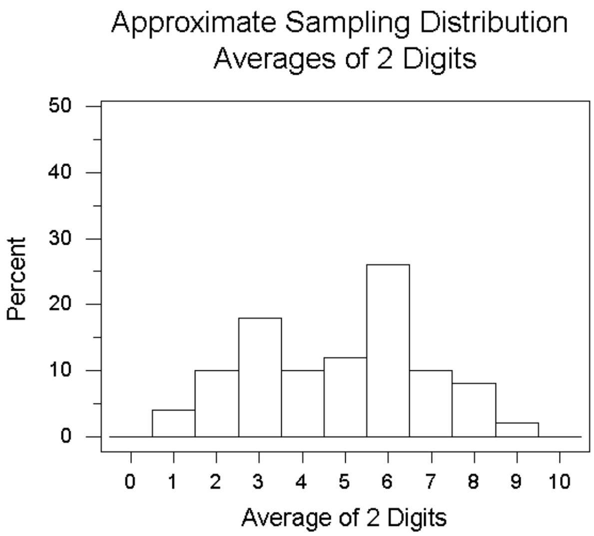
Amanda	2078	Doug	1849
Kathryn	5293	Nathan	4198
Sarah B.	1213	Amy	1176
Jamie	0885	Kelly	8652
Nathaniel	2647	Sarah S.	1325
Angela	7877	Jeffrey	7314
Kimberly	9281	Patrick	2456
Shelley	4839	Audra	0679
Jennifer	5748	Mark	2338
Peter	3678	Theresa	4227
Brianne	3083	Jessie	1492
Miki	3806	Robert	2741
Virginia	3066		

The collection of 100 phone digits will be our population of interest. If you have a large class (100 or more, as I do) each student contributes only the last digit of her/his phone number to the population. For a smaller class you can go to a phone book and get additional numbers to create a larger population. The first thing to consider is how these digits are distributed. Have students come to the board and tally their digits, create a histogram and comment on the histogram's center, spread and shape.

The distribution of the individual digits is centered around 4.5 with values ranging from 0 to 9. The shape is somewhat bimodal with relatively more 2's and 8's and relatively few 0's, 5's and 9's. This shape is not necessarily typical of the distribution of telephone digits. An important point to make is that this distribution does not change. Any time you look at the entire population you will get the same distribution of values.

Distribution of Individual Phone Digits

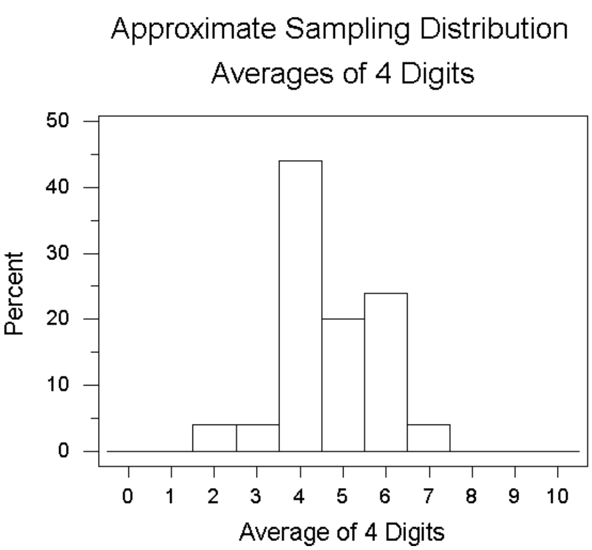




What happens when we sample from our population of 100 digits? Consider samples of size 2. It is convenient to have each student take the first 2 digits of her/his 4 digit number as a sample of size 2. She/he can also use the last two digits as a sample of size 2. By finding the average value for each of these samples, one can have 50 realizations of the average of samples of size 2. These are only 50 of the thousands of possible samples of size 2. Once each student has her/his averages of size 2, the results can be tallied and a histogram of these averages can be constructed and described. Above is the histogram for our class of 25.

The center of this simulated sampling distribution is around 4.5 (similar to that of the original population distribution) but its spread is somewhat less, sample means range from 1 to 9. The shape is tending to mound more towards the center. The percentage of values away from the center is reduced.

Now consider samples of size 4. It is convenient to have each student take his/her 4 numbers as a sample of size 4. By finding the average value for each of these samples, one can have 25 realizations of the average of samples of size 4. Again, the students can go to the board, tally their averages and construct a histogram of those averages. At the bottom of the first column is the histogram for our class of 25.



The center of this simulated sampling distribution is around 4.5 (similar to that of the original population distribution) but its spread is much less, sample means range from 2 to 7. The shape is much more clustered (a single mound) near the center.

Discussion

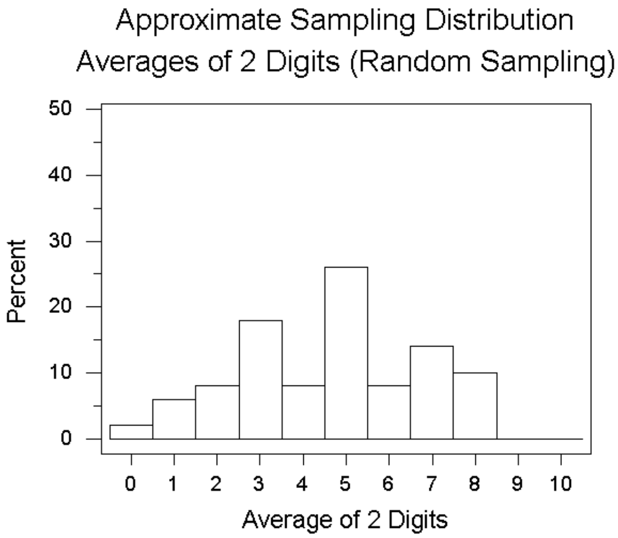
With a class of 100, this activity takes about 30 minutes to complete. Rather than have students come to the board, I have a show of hands for the number of values in each interval class.

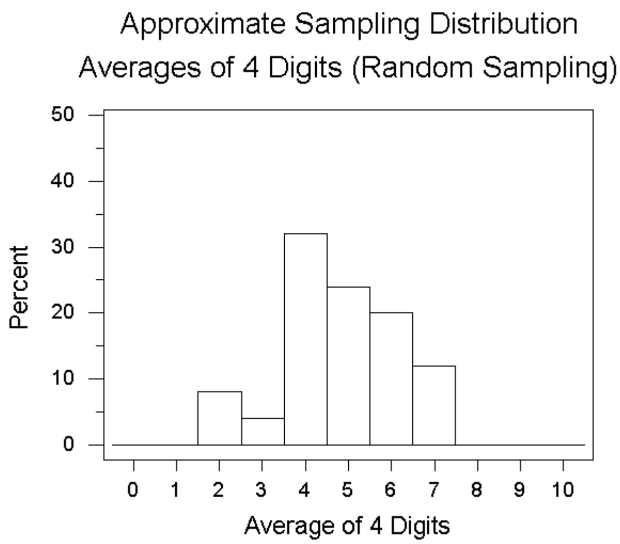
This activity helps to highlight the relationship between the center and spread of the population of values and the center and spread of the sampling distribution of the sample average. It also begins to hint at the Central Limit Theorem. The bimodal shape seen in the population distribution is not apparent in the sampling distribution of the sample average values as the sample size increases.

Many students are confused about the number of samples versus the sample size. We have taken 50 samples of size 2 and 25 samples of size 4 to construct the latter two histograms. It is the size of the sample that is important, not the number of samples. By using a percent scale for the vertical axis on the histograms, the number of samples is de-emphasized.

For larger classes, I have students form groups of 2 and 4. Each member of the group contributes the last digit of his/her telephone number. Most students will form groups by proximity (who is closest to them).

Either with the four digits from each student or groups formed by proximity, it appears that we have violated an important rule for the construction of sampling distributions. We have used convenience samples instead of random samples. Does this make a difference? Have the students come up with a method for randomly sampling from the population. This can be done by having each student write their digits on individual slips of paper. The slips can then be put into a bag, mixed and sampled from. Create multiple random samples of size 2 and multiple random samples of size 4 and construct histograms of the averages. The results you get will be different from the histograms below but should show the same narrowing of spread and clustering (mounding) in



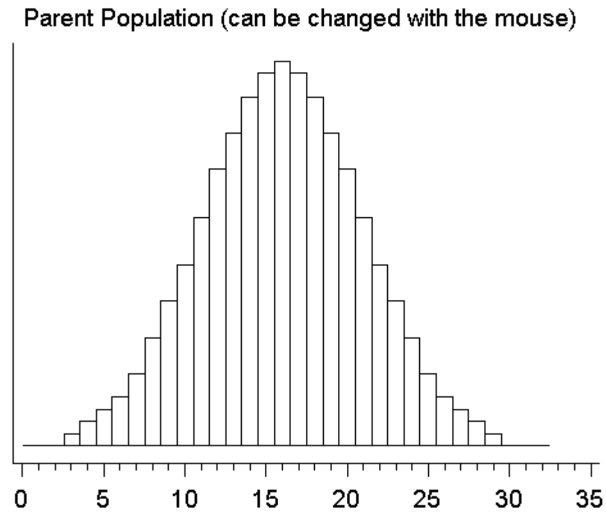


the middle.

What the students will find is that in this case the convenience samples work as well as random samples since the assignment of the last 4 digits of a phone number is essentially a random assignment. With most practical sampling situations, convenience samples are prone to bias and random samples should be used.

Exploring Sampling Distributions on the World Wide Web

Once students have done the in class activity it is time to turn them loose (with a little guidance) to explore sampling distributions using a program developed by David Lane at Rice University. This program allows students to quickly draw many, many samples and to easily change the characteristics of the population and the size of samples. Students will need access to the World Wide Web via a browser that supports frames and



Java.

Students should go to the URL: *www.ruf.rice.edu/~lane/stat_sim/sampling_dist/index.html*

Instructions for the use of the sampling distribution Java applet appear on the right side of the page. Click on the Begin button on the left side of the page. This will bring up a Java applet Window with four sets of axes. The top set displays a mounded, symmetric distribution of the parent population. To the left are the population parameters, a mean of 16 and a standard deviation of 5. To the right is a pull down menu where you may select a few other shapes. Make sure that a normal distribution is selected for the first activity. The parent population should look like the histogram above.

1. Click on the Animated Sample button located at the right of the Sample Data axes. The results of a simple random sample of size 5 taken from the parent population will appear on the Sample Data graph. The mean of this sample of size 5 will appear in blue on the Distribution of Means graph.

- In what ways does the sample look like the parent population?
- What could you do to increase the likelihood that the sample would look more like the population?
- Looking at the summary statistics to the left of the Sample Data graph, is the mean of the sample near the population mean?
- If the Animated Sample button is clicked again will the new sample be the same as the current sample? Briefly explain your answer.
- Click on the Animated Sample button to

Table 1: Distribution of Means, n=5

Mean	
St. Dev	
Sketch	

confirm your answer.

Let's focus on the Distribution of Means graph. This graph should contain two blocks representing the means of the two random samples of size 5 that have been selected. The mean of these two sample means is given to the left of the graph. Press the Animated Sample button several times. Now use the 5 samples and 1000 samples buttons to build up the sampling distribution. Each time one of these buttons is pressed, more samples of size 5 are selected from the population. It is important to remember that the sample size (5) does not change. Each sample contains 5 observations from the parent population. We have to simulate enough samples before the Distribution of Means becomes apparent. Reset the applet by clicking on **Clear lower 3**. Use the 10,000 samples button to simulate the sampling distribution of the sample mean for samples of size 5. In Table 1, sketch the Distribution of Means and give the mean and standard deviation (sd).

- Does the mean of the Distribution of Means differ from the center of the parent population by a lot?
- How does the spread of the Distribution of Means differ from the spread of the parent population?
- Does the shape of the Distribution of Means differ from the shape of the parent population distribution? Concentrate on the shape. Where does the distribution mound? Is the distribution symmetric?

2. Repeat activity 1 only this time use the N = 5 pull down menu next to the Distribution of Means graph to change the sample size to N = 25. Record the description of the Distribution of Means in Table 2. What do you notice that is different when using random samples of size 25 instead of 5?

Table 2: Distribution of Means, n=25

Mean	
St. Dev	
Sketch	

Change the population shape to skewed. Report the mean and standard deviation (sd) and shape (it may help to sketch the distribution) of the parent population in Table 3. Select a sample size of 2 for the Distribution of Means on the third set of axes. Select the Mean and a sample size of 5 for the Distribution of Means on the fourth set of axes. Take 10,000 samples. Report the means and standard deviations and describe and/or sketch the shapes for the Distributions of Means in Table 3. Now select a sample size of 10 for the Distribution of Means on the third set of axes. Select the Mean and a sample size of 25 for the Distribution of Means on the fourth set of axes. Take 10,000 samples. Report the means and standard deviations and describe and/or sketch the shapes for the Distributions of Means in Table 3.

Compare the Distribution of Means for each sample size to the parent population distribution. How does the mean compare? How does the standard deviation compare? How does the shape compare? Summarize your findings about the relationship between the distribution of the parent population, sample size and the sampling distribution of the sample mean in one or two sentences. How does your summary compare to the statement of the Central Limit Theorem?

Discussion

The purpose of the first activity is to have students understand the idea of random sampling from a parent population. The Animated Sample is quite good at visually displaying the selection of 5 values from the parent population. These 5 values produce one realization of the sample mean. Another Animated Sample provides a different set of 5 values and a different realization of the sample mean. By building up the Distribution of the Mean slowly at first, students can see how different samples produce different sample means (the basis

Table 3, Skewed distribution

Sample size	Parent	2	5	10	25
Mean					
Standard Deviation					
Shape					

of sampling variability). Being able to generate multiple samples (10,000 samples button) greatly speeds the process of building the Distribution of the Mean.

Students can sometimes confuse the number of samples with the sample size. It is important to differentiate these two ideas. The Java applet uses Reps to indicate the number of times samples are taken and N to indicate the size of each sample. By sketching the Distribution of Means once for each sample size, and comparing the distributions for several sample sizes, the confusion with the number of samples generated should be lessened.

The first two computer activities deal with a parent population that is symmetric and mounded in the middle (a.k.a. normal). The Central Limit Theorem does not enter in here because the Distribution of Means is normal, no matter what the size of the sample, when sampling from a normal parent population. The Central Limit Theorem only comes into play when we are sampling from a parent population that does not have a normal shape. The third activity does address the Central Limit Theorem and how large a sample one needs before the Distribution of Means is approximately normal shaped. The activity is nice in that the parent population is always displayed and its shape never changes throughout the activity. It is the shape of the Distribution of Means that changes with the sample size.

There are several other web sites that you can go to in order to explore sampling distributions. Some, but certainly not all, are:

- *statweb.calpoly.edu/chance/applets/applets.html*

There are several applets at this site that deal with sampling distributions.

—*Sampling Senators* This applet allows you to draw samples of U.S. Senators and look at characteristics such as: gender, party, and years in office.

—*Sampling Pennies* The population of interest consists of pennies and the year they were minted. Samples are taken from the population and the distribution of the sample mean year is displayed.

—*Reeses Pieces* This applet has a clever animation that looks at the sampling distribution of a binomial proportion.

- *www.stat.sc.edu/~west/javahtml/CLT.html*

This applet simulates the rolling of 1 to 5 dice to demonstrate the Central Limit Theorem.

- *www.gen.unm.edu/faculty_staff/delmas/ stat_tools/sampling_distributions/samp_dist_tools.htm*

This website contains tools for exploring sampling distributions. There is computer software, that you can down load, for construction of the sampling distribution of the sample mean. There are also pre-tests and post-tests and instructions for using the software.

Acknowledgement

The author would like to thank Beth Chance for her helpful comments that greatly improved the quality of this column.

Outlier...s

Allan Rossman

The theme for this issue's column is the name itself: outliers. One of the most interesting aspects of any statistical investigation is looking for outliers. Even more interesting is deciding what to do with them, which inevitably involves non-statistical issues of determining whether they are recording errors or important anomalies. In this column I will introduce you to some of my favorite outliers. [**Assignment 1**: Before you even read further, think about some outliers that you have encountered and found to be interesting. Please send me a message telling me about your favorite.]

Outliers In Elections

By the time this issue of *Stats* appears in print, the United States will have inaugurated its 43rd President. As I write this, though, during the Thanksgiving holiday a full seventeen days after the election, neither I nor anyone else in the country knows for sure who that will be.

An intriguing part of the vote tabulation controversy in Florida involved the question of whether large numbers of people in Palm Beach County who voted for Pat Buchanan had really intended to vote for Al Gore but were confused by the infamous "butterfly" ballot. Statistical analyses conducted on the vote results revealed that the number and percentage of Buchanan votes in that county were extreme outliers compared to results in the other 66 counties in Florida. Whether to attribute the cause for the outlier to voter confusion is a question of much debate. [See *http://madison.hss.cmu.edu* for an analysis by Greg Adams of Carnegie Mellon and Chris Fastnow of Chatham College, along with links to many other analyses.]

I will pursue a much less controversial analysis of outliers related to this election. My home state of Pennsylvania was also a key "battleground" state, attracting as much attention as Florida prior to the election, but far less afterward. The boxplot in Figure 1 displays for the 67 counties in Pennsylvania (the same number as in Florida!) the percentage of a county's votes that went to Al Gore, as reported by CNN's website *www.cnn.com/ELECTION/2000/results/*. [**Assignment 2**: Before reading further, make a guess for the proportion of Pennsylvania counties that Gore won. **Assignment 3**: Try to identify the outlier county.]

The boxplot reveals the upper quartile of this

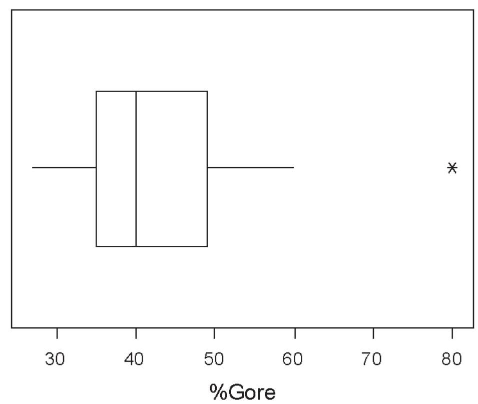


Figure 1.

distribution to be below 50%, but with so many minor-party candidates in the race, it was possible to win a plurality of votes in a county with less than 50% of the votes. Gore won 18 of Pennsylvania's 67 counties, about 27%. The outlier county in which 80% of the votes went to Gore is Philadelphia County, including the city of Philadelphia. This county is not only an outlier in its overwhelming support for Gore but also in terms of number of votes cast. Far more votes were cast in Philadelphia County than in any other. The combination of these two factors, of course, explains Gore's winning the state of Pennsylvania with 51% of the vote to Bush's 47% despite Bush winning more than 70% of the counties, with an average margin of more than 23 percentage points over Gore in those counties.

If Bush had won 80% of the vote in a county, would that have been classified as an outlier like

Gore's 80% in Philadelphia? The IQR of Bush's vote percentages is 63-47 = 16, so according to the convention that outliers are those values falling more than 1.5 IQR's from their nearer quartile, Bush would have had to win more than 87% of the votes in a county for it to qualify as an outlier. [**Assignment 4**: Perform a similar analysis of your state's results to see if any counties are outliers.]

Outliers In Life And Death

The boxplots in Figures 2 and 3 reveal the distributions of birth rates (Fig. 2) and death rates (Fig. 3) for the fifty states. These are rates per 1000 population, based on 1997 data, as reported by the National Center for Health Statistics and reprinted in *The World Almanac*. The states are grouped by whether the state is primarily east or west of the Mississippi River. [**Assignment 5**: Identify the two states through which the river runs. **Assignment 6**: Before you read further, try to identify which two states are the outliers.]

Very reasonable explanations can be found for the two outliers in these distributions. The state with the high birth rate is Utah, which has a high density of people in religions that do not condone birth control. The state with the low death rate is Alaska, a state to which very few people retire to spend their twilight years.

This example also illustrates that outliers are determined by the company they keep. If we pool all fifty states together, then Arizona joins Utah as an outlier on the high end of the birth rates and Utah joins Alaska as an outlier on the low end of the death rates, according to the conventional 1.5IQR rule.

Outliers In Sports

The biggest outlier in sports these days is unquestionably Tiger Woods. He won three of golf's four major championships in 2000. Most impressively, he won the U.S. Open at Pebble Beach and the British Open at St. Andrews in runaway, record-setting fashion. The boxplot in Figure 4 displays the total strokes in the U.S. Open for the 63 golfers who made the cut. Woods' total of 272 was 12 strokes under par and a full 15 strokes better than the runner-up's score. How much of an outlier is this? The quartiles are 293 and 299, so the IQR is 6, so Woods' 272 lies a full 3.5 IQR's below the lower quartile. If you prefer to measure extremeness in z-scores, the mean of these scores is 296.08 and the standard deviation is 5.85, so Woods' 272 corresponds to a z-score of –4.12. For comparison, the runner-up's z-score is –1.55.

A round-by-round analysis reveals that Woods had the best score in the first round (65), tied for the best in the second round (69), and the best in the fourth round (67). His third round score of 71 was actually bettered by three of the 63 golfers. None of these individual round scores classifies as an outlier, but of course the cumulative effect of the four rounds of excellence produces one of the most impressive outliers in the history of the game.

Of course, Woods is not the only outlier appearing in the boxplot. The last-place golfer among those who made the cut finished with a pair of 84 rounds for a total of 313, 29 strokes over par and 41 behind Woods. This golfer's total is 2.33 IQR's above the upper quartile and corresponds to a z-score of 2.89. Both of these would usually be impressive, but they pale in comparison to the degree to which Woods is an outlier. [**Assignment 7**: Identify the player with this high outlying score.]

Woods won the British Open with a score of 269, 19 strokes under par on the venerable links of St. Andrews. The runner-up's score was 281. The quartiles for the 73 golfers who made the cut were 281 and 289, so Woods' score of 269 is exactly 1.5 IQR's below the lower quartile. While this precludes Woods' score from appearing as an outlier on the boxplot, I doubt if this diminishes his achievement in his eyes of the golf world. Woods' total corresponds to a z-score of –3.34 and the runner-up's to a z-score of –1.69. [**Assignment 8**: Determine the mean and standard deviation of these scores.]

Another sports example reveals another one of my favorite outliers. Beth Chance submitted to the JSE Data Archive data on the weights of rowers on the 1996 U.S. men's Olympic rowing team (*www.amstat.org/publications/jse/archive.htm*). A boxplot of the weights of the members of the eight-man event appears in Figure 5. [**Assignment 9**: Before reading further, explain the outlier in this boxplot.]

This outlier can be explained not by a recording error or even by voter confusion. The outlier among these rowers is the coxswain, who does not row but rather calls out instructions to help the rowers to work in unison. In order not to add much weight to the boat, an ideal coxswain is therefore very light (and loud!). This example also serves as a reminder that outliers depend on their company. When one examines weights of the entire team, the six rowers who participate in "lightweight" events have small enough weights to pull down the value of the lower quartile to where the coxswain's weight is less than one IQR below the lower quartile. [**Assignment 10**: Analyze the data on weights of the 2000 Olympic rowing team to see if a similar outlier emerges (see *http://rowing.about.com/recreation/rowing/library/blolympics00usteam.htm*).

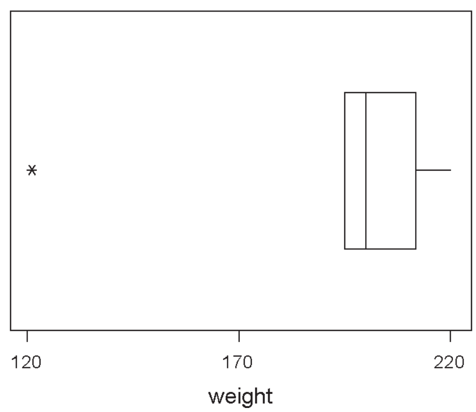


Figure 5.

Outliers From Mathematical Models

One of my favorite assignments to give to students who are studying the normal distribution for the first time is to find the probability that an observation from a normal distribution will be classified an outlier by the 1.5IQR rule. [**Assignment 11**: Before you read further, go ahead and do this calculation.]

One can find that the quartiles of a normal distribution fall at $\mu \pm 0.6745\sigma$. Thus, the IQR is 1.349σ. Outliers are therefore values that fall below $\mu - 2.698\sigma$ or above $\mu + 2.698\sigma$. The probability of each of these is .0035, so the probability that a value from a normal distribution is classified as an outlier is .0070. How does this probability compare to other probability distributions? [**Assignment 12**: Before reading further, perform this calculation for the continuous uniform and exponential distributions.]

The uniform distribution with endpoints α and β is quite easy to work with. The quartiles occur one-fourth and three-fourths of the way along the interval, so the IQR is one-half of the interval. Thus, subtracting 1.5IQR from the lower quartile falls below α, and adding 1.5IQR to the upper quartile exceeds β. The probability is therefore zero that an observation from a uniform distribution will be labeled an outlier.

For an exponential distribution with mean θ, the quartiles can be shown to occur at $-\theta\ln(.75)$ and $-\theta\ln(.25)$. The IQR is therefore $\theta\ln(3)$, or about 1.099θ. The right skewness of the exponential distribution ensures no outliers on the low end, but a high outlier can be found to be a value above roughly 3.0342θ, and the probability of a high outlier can be calculated to be about .048, a much higher probability than with the normal distribution.

As a check on this analysis, I simulated 10,000 values from a normal distribution. I found that 71 of the 10,000 fell below –2.698 or above 2.698. I then simulated 10,000 values from an exponential and found that 481 of the 10,000 fell above 3.0342θ.

These calculations are unsatisfying in some respects, though. They pertain to the theoretical model and not to samples of data generated from the model. The IQR calculation on which the outlier test depends would be based on the sample data, not on the quartiles of the model. For instance, is the probability really zero of encountering an outlier in a sample from a uniform distribution? There is some positive probability, admittedly quite small, that enough sample values would fall near the center of the distribution to have a small enough IQR that some values near the endpoints could be labeled as outliers. Similarly, with a sample from a normal distribution, the IQR based on the sample data will naturally not equal the IQR of the model itself, so the calculation above will not be valid for a sample of data. A more substantial analysis, either by studying order statistics or through simulation, is well beyond the scope of this modest column that set out to examine Al Gore's votes and Tiger Woods' strokes. [**Assignment 13**: Please conduct more thorough analyses of these outlier problems, and write up your findings as a submission to *Stats*. **Assignment 1 reminder**: Please send me your favorite examples of stories of outliers.]

[Answers to assignments not given in the column: 5. The Mississippi River runs through the states of Minnesota and Louisiana, but in both cases most of the state lies to the west of the river. 7. Robert Damron was the outlier who finished in last place in the 2000 U.S. Open. 8. The mean of the scores was 285.18, and the standard deviation was 4.85.]

Please send your assignments, and suggestions for future columns, to Allan Rossman at *rossman@dickinson.edu*.

Start Your Career Path Here

A GROWING MARKET
FOR YOUR
TALENTS

SMITH HANLEY
Associates, Inc.

Current Nationwide Opportunities

- Clinical Statistician
- Pre-Clinical Statistician
- Quality Control Statistician
- Research & Development Statistician
- SAS Programmer

Pharmaceutical/Industrial Statistics
Contact: Barbara Day
Phone: 800-989-5627

- Market Research Analyst
- Database Marketing Analyst
- Marketing Scientist
- Forecast Analyst
- Credit Risk Analyst

Business/Marketing Statistics
Contact: Janet Kasper
Phone: 800-768-3957

www.smithhanley.com

At Smith Hanley, we're not content with simply finding you work – our mission is to help you capitalize on the best career options available.
This means not only representing you to some of the top companies in your chosen industry, but also offering advice on career paths and industry trends. We have worked with hundreds of students from statistics departments across the country. Ask your alumni about our services.

As leaders in the placement of Statisticians, with specialized recruiters in each industry, Smith Hanley is able to locate the position that's ideal for you – one that matches your unique skills, goals and professional profile.

E-mail resume as a Word attachment to: stats@smithhanley.com

Get FIRED up for the 2001

JOINT STATISTICAL MEETINGS in **HOT 'LANTA**

JSM2001
ATLANTA
statistical science for the information age

August 5–9, 2001

Atlanta Hilton & Towers and
Atlanta Marriott Marquis Hotels

photo by Kevin Rose