

TEACHING of STATISTICS in the HEALTH SCIENCES

WALTER AMBROSIUS

From the Section Chair



As with most of the ASA, much of our activity as a Section is related to the JSM. We sponsor many sessions and roundtable lunches as well as our business meeting/mixer. Our current TSHS Program Chair, Patrick Arbogast, described the sessions he's organized in the last TSHS Newsletter. The specific

schedule is on page 7 and on the ASA website: <http://www.amstat.org/meetings/jsm/2005/onlineprogram>

Please note that the business meeting/mixer is on Monday evening this year rather than the traditional Wednesday. This was done in order to improve attendance as some don't stay until the end of the JSM. It will be from 5:30-7pm in the Hilton. Please add this to your calendar and plan to attend! We've ordered as much good food and drink as our budget will allow and want to make sure we have plenty of people to enjoy it!

Congratulations to Marleen Egger and Gloria Caldito. Both were winners in the recent TSHS elections. A complete description appears on page 5.

Elsewhere in this newsletter you'll find several interesting articles:

- Book reviews by Bob Oster and Dan Byrne
- Bill Seltzer discusses the ethics of data collection for classroom research. Most IRBs do not review classroom projects but the data

collected can sometimes be used for purposes other than teaching. Bill's article provides some interesting examples and raises the question of whether IRBs should review classroom data collection.

- Bob Oster discusses the recently approved revisions to the Section's Charter and the very deserving recipient of the Section's first Distinguished Achievement Award.
- Janet Tooze, the 2006 Program Chair, requests ideas for invited and topic contributed sessions for next summer. The invited sessions need to be finalized soon so please contact her with ideas!

2005 DISTINGUISHED ACHIEVEMENT AWARD COMMITTEE REPORT

Bob Oster

The 2005 Distinguished Achievement Award Committee is pleased to announce that the winner of the 2005 TSHS Distinguished Achievement Award is Stephen Looney. Congratulations, Stephen! (Picture below).



This award is intended to be the highest honor that can be bestowed on an individual by TSHS, and is given for outstanding service to TSHS over a period of several years. Guidelines for this award are posted on the TSHS website under the "Awards" link.

Stephen has been very active in TSHS for more than ten years. He has served the section as Program Chair, Chair, Publications Officer, and as Newsletter Editor (for six years). In addition, he has given several TSHS presentations during recent Joint Statistical Meetings. In 2000, he won

an award for Best Invited Paper. Stephen has also contributed articles to the TSHS Newsletter.

Stephen has truly served as an inspiration to current and recent officers of TSHS. Most of us who have served TSHS recently, or who are currently serving TSHS, can honestly say that TSHS is a much better section due to Stephen's efforts and hard work. Please congratulate Stephen in Minneapolis this year at the JSM.

The 2005 Distinguished Achievement Award Committee consisted of Walter Ambrosius, Cyndy Long, and myself (as committee chair). I want to thank Walter and Cyndy for their assistance with the work of this committee. The next Distinguished Achievement Award will not be given until 2007 (according to the guidelines for this award).

BOOK REVIEW 1

Regression Modeling Strategies: With Applications to Linear Models, Logistic Regression, and Survival Analysis

Frank E. Harrell, Jr.
Springer-Verlag, 2001

Reviewed by Robert A. Oster, Ph.D., of the University of Alabama at Birmingham (Pictured)



The emphasis of this text is on problem-solving strategies that address issues found in various types of multivariable modeling rather than on a comprehensive description of modern data analysis techniques.

The text attempts to bring linear regression, logistic regression, and survival analysis models under a predictive modeling umbrella that focuses on how best to use these models and when the use of these models is appropriate.

The target audience of this text is Masters-level or PhD-level graduate students, and data analysts and statistical methodologists, who have had a general introductory probability and statistics course and who are well versed in

ordinary multiple regression and intermediate algebra. This book should be of interest to TSHS members that have the required mathematical and statistical background, and can be useful for teaching or research purposes. However, this text will be less useful to those not having a solid background in applied statistics.

The book contains 20 chapters, one appendix, a list of references, and an index. Some of the topics covered in this text are as follows: 1) general aspects of fitting regression models, including splines and assessment of model fit, 2) missing data, including single and multiple imputation, 3) multivariable modeling strategies, including data reduction techniques, 4) resampling, validating, describing, and simplifying the model, 5) various modeling functions in S-Plus, 6) an overview of maximum likelihood estimation, 7) binary logistic regression, 8) ordinal logistic regression, including proportional odds models and continuation ratio models, 9) models using nonparametric transformations, and 10) survival analysis, including parametric survival models such as parametric proportional hazards models and accelerated failure time models, and Cox proportional hazards regression models. The text also contains detailed case studies in 1) least squares fitting and interpretation of a linear model, 2) imputation and data reduction, 3) binary logistic regression (two separate case studies), 4) ordinal logistic regression, 5) modeling using nonparametric transformations, 6) parametric survival modeling, and 7) Cox regression modeling.

There is much to like about this text. The case studies are very helpful in illustrating the various modeling techniques that are discussed. The case studies also cover a wide range of subject areas, such as federal elections (the 1992 presidential election in particular), estrogen therapy for the treatment of prostate cancer (a large randomized trial), the survival of Titanic passengers, predictors of diagnosis and severity of illness related to pneumonia, meningitis, and sepsis, cardiovascular risk factors, and an examination of critically ill hospitalized adults for in-hospital outcomes and for long-term survival. Numerous graphics and examples of computer output are included in descriptions of the results for these studies.

Several chapters include a well-written section on further reading. Each of these sections contains additional helpful information relevant to the specific chapter and to the specific references that are cited.

The problems provided at the end of the chapters are useful for teaching purposes (when teaching to those who have the required mathematical and statistical background). These problems concern applied statistics, and focus on real-world examples or datasets given in the chapters, or on the various case studies themselves. Working through these problems will definitely give the student a greater and clearer understanding of the modeling techniques described in the text.

The organization of this volume is helpful for learning about regression modeling strategies. The first part of the text discusses multivariable linear regression modeling; the middle part discusses binary and ordinal logistic regression modeling; and the latter part discusses survival analysis, including parametric and semi-parametric (Cox) regression modeling. This is the order in which many statisticians learn these techniques, and is one that I find to be logical.

All of the datasets cited in the text, as well as the S-PLUS libraries that are written or created by the author, are freely downloadable from the Web site mentioned in the text. The author also provides the reader 466 citations relevant to regression modeling strategies; most of these citations are to peer-reviewed journal articles.

I have only three criticisms of the text. The first one is that some of the explanations appear to be too terse for the reader. I sometimes wanted to see more details for certain topics. Additional details may help the reader learn more about modern regression modeling strategies.

The second is that coverage of the topics is somewhat variable. The coverage that is provided is reasonably clear, but much more material is provided for some topics than for others.

Regarding these two points, the text is written as a reference text and as a text for graduate students (primarily statistics and/or biostatistics graduate students). It is not intended to contain a lot of introductory material on the regression modeling strategies, so perhaps it was the author's intention for some of the topics to be covered tersely, while providing additional coverage to selected strategies.

The final criticism is that most of the results are reproduced using only S-Plus. This is fine for those who are proficient in and enjoy using S-Plus. However, it would have been very helpful for some of the results to have been reproduced using SAS and Stata for readers who use either of these packages but who may not use (or may not want to learn) S-Plus. The author states that SAS

macros and notes on SAS usage are available from the Web site cited in the text, so this is something that SAS users will want to look into.

I recommend "Regression Modeling Strategies: With Applications to Linear Models, Logistic Regression, and Survival Analysis" for use to anyone who considers themselves in the target audience for this text. I believe that it is useful as a reference for applied statisticians (for research and consulting purposes). The text will also be useful for teaching purposes (for Masters-level or PhD-level graduate students with the appropriate mathematical and statistical background).

BOOK REVIEW 2

Introductory Statistics with R
Peter Dalgaard
Springer-Verlag, 2002

Reviewed by Daniel W. Byrne,
Department of Biostatistics, Vanderbilt
University Medical Center (Pictured)



Many of us who teach statistics in the health sciences have recognized the growing importance of the open source language/statistical software R and the need for a book that clearly explains how to begin using this software. Peter Dalgaard has filled this niche with his book

Introductory Statistics with R, which is based on notes from a course he teaches at the University of Copenhagen in Denmark. Not surprisingly, the writing style is that of a teacher coaching and encouraging students to try R commands and explaining the output.

The R software, which can be downloaded free of charge (www.r-project.org), is appealing to teachers for several reasons. In addition to the free access, R provides flexibility in producing graphics and statistical model building. This flexibility does come at a price – a steep learning curve, which necessitates an introductory text such as this. To truly benefit from this book, however, the reader should reproduce all, or most, of the examples in R while reading. The author

facilitates this by providing the data sets used in the book in the ISwR package, which can be downloaded from the R web site.

Chapter 1 (Basics) is a 44-page overview of the R system, with examples on every page for the reader to try. Although this overview succeeds in introducing many important features of R to the reader, it includes advanced programming techniques that could exhaust and frustrate a novice. And yet several essential techniques are not given adequate attention. For example, readers will need to know how to import data sets that were saved in other formats: SAS, SPSS, Excel, etc. Unfortunately, there is insufficient detail to help readers accomplish this essential step. Future editions of this book could be improved by expanding this section into a complete chapter and placing this early in the book. Instructors who adopt this book for a course would need to provide supplemental material on importing data sets.

Chapter 2 provides a brief overview of probability functions. Chapter 3 introduces graphics. By demonstrating the power and flexibility of the graphical routines in R, an instructor can help students see the value of investing the time to learn this system. The professional-quality graphs produced in R can help students increase their odds of publishing a paper in a more prestigious journal. Instructors should, however, consider teaching students how to use some of the functions found in the add-in packages rather than in the base R system. For example, on page 63, Dalgaard explains how to create an empirical cumulative distribution function plot in the following 2 steps:

```
n <- length(x)
plot (sort(x), (1:n)/n, type= "s", ylim=c(0,1))
```

Instead, instructors could teach how to load the Hmisc package and use the following command:

```
ecdf(x)
```

This is not a criticism of the author, since there are many R packages, and some routines may have been created after the publication of this book. Instructors will, however, want to find the most efficient methods of accomplishing these tasks not only to help with the learning curve, but also to help with the 'forgetting curve'. Another valuable graphical technique from the Hmisc package that would be worth teaching as a supplement to this book is the plot smoothing function (plsmo). Graphical functions, such as these, will help to motivate students to learn R.

The book covers most of the topics typically included in an introductory biostatistics book: t-test, chi-square, nonparametric statistics, ANOVA, linear and logistic regression, survival analysis, and sample size calculations. For most introductory courses, however, I would recommend that instructors also require a book such as Kirkwood and Sterne's "Essential Medical Statistics" to provide a more comprehensive foundation.

In summary, for instructors who wish to learn R themselves, I would recommend Dalgaard's book but strongly encourage experimenting with R for every example given. For instructors who teach biostatisticians, this book would be an appropriate text to introduce R, however, one would also need to require an introductory biostatistics text. Instructors who teach nonbiostatisticians, such as those in an MPH program, and wish to expose students to R, might find Dalgaard's text appropriate but they should not underestimate the amount of preparation that will be required to create the lessons and supplemental material to make such a course successful. Having used this book in an introductory biostatistics graduate course for physicians, I would encourage instructors to budget sufficient time for this preparation. The key is to make it easy for the students to see the value of R and provide them with clear step-by-step instructions to avoid frustration. Introductory Statistics with R can be an important component of a statistics course but it must be accompanied by adequate supplemental material and an instructor who is sufficiently prepared help students understand the techniques that are not explained in sufficient detail in this book. The errors and notes on the following web page are useful place to begin: <http://www.biostat.ku.dk/~pd/ISwR.html>.

JSM 2006: Advance Notice



It's not too early to start thinking about JSM 2006 in Seattle! If you are interested in submitting a proposal for an invited session for JSM 2006, please contact me, **Janet Tooze** (336-716-3833), jtooze@wfubmc.edu, by July 11. Suggestions on all topics are welcome.

Classroom Research: A Note on Professional Ethics

William Seltzer

Chair, ASA Committee on Professional Ethics

I am writing as Chair of the ASA Committee on Professional Ethics to bring a matter of concern to your attention. I refer to the occasional misuse of what is sometimes referred to as “classroom research.” In undergraduate and graduate courses, students are sometimes given assignments involving interviewing or observing others in order to gain experience in certain aspects of research methods. Such practical classroom exercises are often an element of methods courses in the health, social and behavioral sciences. The degree to which statisticians and ASA members are directly involved in such exercises varies markedly by institution and sponsoring department. Regardless of who is doing the teaching, however, these exercises should be designed to embody sound and ethical statistical practices.

One problem that has arisen with respect to sample surveys or other forms of data gathering carried out as classroom exercises is that they seem to fall in a sort of regulatory no man’s land, often unreviewed by the college or university’s Institutional Review Board (IRB). According to my understanding of the federal Office of Research Integrity’s position, at least as stated at JSM 2004, such studies are outside the normal area of its responsibility, as they are seen as “training” rather than “research.” Of course the bulk of such training exercises are just that. However, in a few instances such classroom training exercises have been used by the instructors for their own personal, non-academic purposes, which can compromise both the ethical responsibilities of the instructors and their performance as responsible researchers. Within the context of academic training in research, the modeling of poor research methods by an instructor is particularly reprehensible.

Because of the media attention they received in the past two years, the Ethics Committee became aware of two examples where surveys carried out by students as specific classroom assignments were used to serve other, non-academic purposes

to the possible detriment of at least some of the students involved. (In neither case, did we attempt any independent verification of these instances. Our descriptions are based on media reports and other postings on various websites as described below. Indeed, our concern was not the detailed facts of any specific case, but the underlying problem which these two incidents appeared to reflect.)

In the first example, students at the College of Charleston were given credit by a tenured Associate Professor in the School of Business and Economics to carry out a political poll in connection with a forthcoming local election. According to the reports available, the faculty member giving the assignment was the campaign manager for one of the candidates and at least some of the questions used in the survey were of the nature of a “push poll,” a reprehensible survey technique designed not to gather data, but to influence the views of respondents. The newspaper story reporting this incident, which appeared in the Charleston (SC) Post and Courier, noted in passing that “Professors routinely use students for polling and marketing data at the college, officials said. [The same professor] previously has used students to study the ethnic greeting card industry, the feasibility of a fitness center for the Charleston peninsula, and dealership satisfaction for a local boat-building company.”

(http://www.bio.ri.ccf.org/ASA_TSHS/pdf/2003PushPollStory.pdf)

The second example involved a survey carried out as a classroom assignment by students in a sociology class at California State University, Stanislaus, and because it played a role in the Scott Peterson murder trial, it has and continues to receive considerable media attention. In this case, which involved a tenured full professor in the criminology program, the faculty member was called as an expert witness on behalf of the defense to support a successful change of venue motion. His testimony primarily involved presenting the results of a public opinion survey his students carried out as a formal class assignment, concluding that the results showed that “a change of venue was warranted.” The judge, in part citing this survey, approved the change of venue motion. After this was reported by the media, several of the students came forward to the press to say that because of the difficulty of the assignment and the limited time available they faked some of the survey results.

The consequences for the trial were quickly resolved, but the resulting academic fall-out for the students involved, the professor, and the University continues. (http://www.bio.ri.ccf.org/ASA_TSHS/pdf/ScottPetersonMurderCase.pdf contains a series of stories, announcements, and other materials related to this incident.)

As portrayed in the media reports, both of these incidents appeared to be a disservice to the students involved, and both were clearly an embarrassment to their educational institutions. The faculty members who made the assignments in question, their department chairs, and at least some of the deans involved did not appear to be aware, until after the fact, of the ethical issues posed by these assignments. Moreover, regardless of the circumstances of these two examples, it is clear that the potential for serious misuse exists when surveys carried out as an academic assignment are also used to further the non-academic interests of the instructor.

In addition, stories about these two examples were posted on AAPORNET, a list serve maintained by American Association for Public Opinion Research (AAPOR). The related comments posted in response stressed (a) an awareness of additional examples where students were assigned survey taking roles for the benefit of others, (b) the importance of obtaining IRB review for any such surveys even if not required by federal research regulations, and (c) a recognition that a number of academic and other survey research centers had long-standing procedures to separate classroom training exercises from operational surveys. (Permission was granted by AAPOR to access AAPORNET for the purpose of this inquiry.)

When the Ethics Committee discussed this matter at its 2004 meeting, it noted that such practices also reflected badly on the entire statistical profession, and suggested that I raise the matter with those ASA Sections principally involved in the teaching of statistics. I would like to recommend that all ASA members in general and TSHS members in particular work within their academic institutions to establish clear guidelines relating to such research, probably through their institution's existing IRB mechanism.

I and my colleagues on the Ethics Committee would be happy to discuss this further. In particular, we would welcome suggested ways to address this issue.

TSHS News and Notes

TSHS Election Results

On May 23 the ASA announced the results of the election of officers. Marlene Egger from the University of Utah School of Medicine will be the Section Chair for 2007 and Gloria Caldito from Louisiana State University Health Sciences Center will be the Secretary Treasurer for 2006-2007. Congratulations to the two of them! The proposed revisions to the Section Charter were approved and can be viewed on the Section's website at http://www.bio.ri.ccf.org/ASA_TSHS/html/charter.htm.

2004-2005 TSHS CHARTER REVIEW COMMITTEE FINAL REPORT

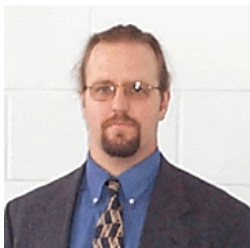
Bob Oster

On behalf of the 2004-2005 TSHS Charter Review Committee, I want to say "thank you" to all section members for approving the proposed revisions to the TSHS charter in the most recent ASA election. We believe that the revised charter is now up-to-date, and more accurately reflects the way that we currently govern ourselves as a section. The revised charter, which will be referred to as the 2005 version of the TSHS charter, is now posted on the TSHS website http://www.bio.ri.ccf.org/ASA_TSHS/html/charter.htm and on the ASA's own website at <http://www.amstat.org/sections/blocks/TSHSCharter.pdf>

This document can be downloaded as a pdf file from either of these links. The charter review committee consisted of Jim Leeper, Ralph O'Brien, and myself (as committee chair). I once again want to thank Jim and Ralph for their assistance with the work of this committee.

Previewing JSM 2005

**From the 2005 Program Chair:
Patrick G. Arbogast
Vanderbilt University**



Hello from Music City! At JSM 2005 in Minneapolis, our illustrious section is sponsoring one stimulating invited session, one engaging contributed papers session, three fascinating contributed posters, three thought-provoking roundtable luncheons, and an outstanding CE course. In addition, we are also having our annual TSHS members meeting and mixer (food and drink provided). Below is a listing of all of these activities. For additional details, go to the ASA website. I look forward to seeing you in August!

Activity	Title	Type	Date/ Time
74	Contributed Poster Session 1	Contributed Posters	08/07/2005 8:00 PM to 9:50 PM MCC-Level 2 Lobby B
CE 13C	Sample-Size Analysis for Study Planning	CE Course	08/08/2005 8:00 AM to 4:00 PM MCC-L100 G
147	Contributed Poster Session 2	Contributed Posters	08/08/2005 10:30 AM to 12:20 PM MCC-Level 2 Lobby B
160	Assessment of Medical Students and Postgraduate Physicians in Biostatistics Courses	Luncheons Roundtable Lnch	08/08/2005 12:30 PM to 1:50 PM MCC-Ballroom B
200531	TSHS Business Meeting and Mixer	Other Cmte/Business	08/08/2005 5:30 PM to 7:00 PM H-Director's Row 2
272	Issues and Methods of Program and Course Development in Teaching Statistics for the Health Sciences	Contributed Papers	08/09/2005 10:30 AM to 12:20 PM MCC-103 E
289	Integrating Statistical Software in Health Sciences Graduate Courses	Luncheons Roundtable Lnch	08/09/2005 12:30 PM to 1:50 PM MCC-Ballroom B
418	Distance Education in Biostatistics	Luncheons Roundtable Lnch	08/10/2005 12:30 PM to 1:50 PM MCC-Ballroom B
429	Distance Learning in the Health Sciences	Invited Papers	08/10/2005 2:00 PM to 3:50 PM MCC-205 C

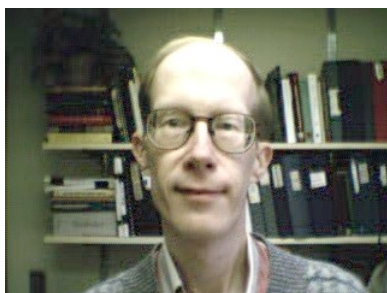
MCC = Minneapolis Convention Center, H = Hilton Minneapolis

2005 Section Officers for TSHS

Chair	Walter Ambrosius	(336) 716-6281	wambrosi@wfubmc.edu
Chair Elect	Cyndy Long	(319) 884-5157	LONG_C@palmer.edu
Chair-Elect-Elect	Marleen Egger	(801)587-3335	MEGGER@DFPM.UTAH.EDU
Past-Chair	Robert Oster	(205) 934-3679	oster@uab.edu
2005 Program Chair	Patrick Arbogast	(615) 343-0701	patrick.arbogast@vanderbilt.edu
2006 Program Chair	Janet Tooze	(336) 716-3833	jtooze@wfubmc.edu
Secretary-Treas. Elect	Gloria Caldito	(318) 675-6803	GCaldi@lsuhsc.edu
Secretary-Treas.	Reena Deutsch	(619) 543-6898	rdeutsch@ucsd.edu
Publications Officer	Renee Stolove	(718) 543-6657	CatLadyPT@netscape.net
Council of Sections Rep.	Lynn Ackerson	(510) 891-3556	lma@dor.kaiser.org
Newsletter Editor	Ed Gracely	(215) 991-8466	egracely@drexelmed.edu
Book Review Editor	Daniel Byrne	(615) 322-4228	daniel.byrne@vanderbilt.edu
ASA Staff Liaison	Monica Clark	(703) 684-1221	monica@amstat.org

American Statistical Association
 TSHS Newsletter
 Ed J. Gracely, Ph.D. Editor
 Family, Community, and Preventive Medicine
 Drexel College of Medicine
 2900 Queen Lane
 Philadelphia, PA 19129
 215-991-8466 (phone)
 215-843-6028 (fax)
egracely@drexelmed.edu

FROM THE EDITOR



A very full issue, with lots of good reading. Dan will be reviewing books on S, T, U, and V in future editions. His reviews of A through Q are available in the archives -- NOT!

Hope you find plenty of items of interest.

As always, if you have something to contribute, send it along!

Ed