INSTRUCTORS

Alan Agresti (Distinguished Professor), is author of four books, including *Categorical Data Analysis*. He has conducted numerous short courses on categorical data analysis for industry and professional organizations in about 20 countries around the world. This one-day tutorial won the 2002 award for Excellence in Continuing Education from the American Statistical Association when it was presented at the 2002 Joint Statistical Meetings in New York City. Professor Agresti was named Statistician of the Year by the Chicago Chapter of the ASA.

James Booth (Professor, Associate Chair, College of Liberal Arts and Sciences) previously served for several years as graduate program director and has had great success developing and teaching a graduate level course in generalized linear models. Professor Booth is an active researcher in statistical methodology and has published numerous articles on a variety of statistical topics including bootstrap methods, Monte Carlo model fitting, saddlepoint approximation and exact conditional tests.


Cynthia Wilson Garvan (Research Assistant Professor) is graduate program coordinator of the Department of Statistics in the College of Medicine. She has received two teaching awards for her innovative work in teaching statistics courses to disabled students. Professor Garvan is the chief statistician in a prospective study examining the effects of cocaine exposure on child development, and is also the statistician for a large prospective study investigating the incidence of post-operative cognitive dysfunction. Her numerous publications include many applications of longitudinal data analysis methods. She is currently the associate editor of *Pain Physician*.


Ramon Littell (Professor, Associate Chair, Institute of Food and Agricultural Sciences) is a consulting statistician with the Florida Agricultural Experiment Station. He is the co-author of the texts *SAS System for Linear Models* and *SAS System for Regression*. Professor Littell is a Fellow of the American Statistical Association (ASA), and is an associate editor for the *Journal of Agricultural, Biological, and Environmental Statistics*.

Jonathan Shuster (Adjunct Professor) has over 20 years experience as the founding Group Statistician of the Pediatric Oncology Group, a federally funded consortium of over 100 medical centers conducting Phase I, II, and III clinical trials and biological studies in Childhood Cancer. He has been Principal Investigator on over $25 million in federal grants and has published over 200 peer reviewed scientific articles, mostly on the design, conduct and analysis of clinical trials. He is on the editorial board of the journals *Blood* and *Sequential Analysis*, has extensive experience on National Institutes of Health Grant Review Boards, and has assisted drug and device companies with IND and IDE submissions to the FDA.

Alex Trindade (Assistant Professor) is an active researcher in time series and quantitative finance. He teaches time series at the masters level in the department, structuring courses into an applied and computer-intensive format. In his consulting and collaborative efforts, he regularly applies time series methodology to solve a variety of problems in biological and engineering sciences.

Mark Yang (Professor) has worked extensively in applied statistics in addition to his research and teaching duties at the University of Florida. He has worked with Bell Labs, the Department of Energy, the Naval Research Lab, NASA and the Redstone Arsenal. His main activities outside the university have been in image processing and time series model building. He has published numerous papers in statistical and IEEE journals and is a Fellow of the American Statistical Association. He received the University Excellence in Teaching (TIP) Award in 1998. A list of his publications can be found on his webpage http://www.stat.ufl.edu/~yang.
COURSE DESCRIPTIONS

Analysis of Repeated Categorical Measurement Data

March 8: Alan Agresti

This course surveys methods for correlated categorical data, which occur with repeated measurement and other forms of clustering. After reviewing standard methods for matched pairs, the main focus is on two types of models. One type models the marginal distributions, with parameter estimation often handled with generalized estimating equation (GEE) methodology. The other type uses random effects to describe subject-specific conditional distributions. For each type, emphasis is on logit models for binary responses but with some discussion of ordinal responses. Examples use SAS (PROC GENMOD and NLMIXED). The presentation, which follows Chapters 10-13 of the text Categorical Data Analysis (2nd ed., 2002), emphasizes concepts rather than technical details. However, attendees should have some background in logistic regression.

1. Review of logistic regression
   - Interpretation of model parameters
   - Maximum likelihood model fitting
   - Extensions for multinomial responses

2. Models for Matched Pairs
   - Comparing dependent proportions: McNemar test et al.
   - Conditional vs marginal models for binary matched pairs
   - Comparing margins of square contingency tables

3. Marginal Models
   - Marginal logit models for repeated binary response
   - Maximum likelihood (ML) and weighted least squares (WLS)
   - Generalized estimating equations (GEE) approach
   - Cumulative logit modeling of repeated ordinal responses

4. Conditional Models: GLMs with Random Effects
   - Conditional logistic regression of clustered binary data
   - Generalized linear mixed models (GLMMs): ML fitting and inference
   - Logistic GLMMs for clustered binary data

5. Additional Topics about Mixture Models
   - GLMMs for ordinal data
   - Multivariate GLMMs for multi-center clinical trials
   - Beta-binomial model for clustered binary data: Teratology studies

Generalized Linear Models
March 10-11: James Booth

The generalized linear model is one of the most important developments in practical statistical methodology in the last thirty years. Generalized linear models provide a versatile modeling framework in which a function of the mean response is "linked" to appropriate covariates through a linear predictor and in which variability is described by a distribution in the exponential dispersion family. These models include logistic regression and log-linear models for counts together with normal, gamma and inverse Gaussian models for continuous responses. Recently generalized linear models have been extended to allow for dependent responses either by the use of generalized estimating equations or by the inclusion of random effects in the linear predictor.

Attendees will be introduced to different topics and will learn how to perform the analyses using the SAS and R computer packages. No previous knowledge of generalized linear models will be assumed, but a background in normal theory linear models is required. No textbook is required but a useful reference is Generalized Linear Models, McCullagh and Nelder, Chapman & Hall 1989, ISBN 0-412-31760-5.

Main topics covered are:
1. Review of normal theory linear models
2. Definition of generalized linear models (GLMs)
3. Inference and diagnostics for GLMs
4. Binomial regression
5. Poisson regression
6. Methods for handling overdispersion
7. Generalized estimating equations (GEEs)
8. Generalized linear mixed models: (time permitting)

Modeling and Analysis Using Monte Carlo Methods

March 8: Mark Yang

Monte Carlo statistical methods, particularly those based on Markov chains, have matured to the point that they are now part of the standard set of techniques used by statisticians. This short course serves as an introduction to the application and underlying workings of Monte Carlo methods. A variety of interesting statistical applications are used to illustrate how these methods enhance statistical practice.

The course begins with basics of random number generation and illustrations of how simulation approaches often supply easy methods for solving difficult problems. We explore techniques for Monte Carlo integration and optimization, and the more recent Markov chain Monte Carlo techniques such as the Gibbs sampler and the Metropolis-Hastings Algorithm. These methods are particularly useful in analyzing data that are modeled using a hierarchical structure. We will use examples from economics, engineering, biostatistics and other areas to provide details of the construction and analysis of such models, and illustrate algorithms such as EM and Data Augmentation.

Each attendee is strongly urged to bring a laptop computer that has a copy of R installed on it. (R is available free of charge, and can be downloaded from http://www.r-project.org/) A number of examples will be worked out and, time permitting, each attendee will have the opportunity to analyze and interpret the data and models used.

Attendees are also encouraged to bring their own data sets for analysis. Much of the course material and many of the examples used are from Monte Carlo Statistical Methods, Robert & Casella, Springer-Verlag 1999. ISBN 0-387-98707-X, and it is strongly recommended that the attendees have a copy of the book.

We do not assume that attendees have any familiarity with
Monte Carlo techniques, or with any Markov chain theory. We do assume familiarity with basic theoretical statistical concepts such as densities, distributions, probability and expectations, the Law of Large Numbers, the Central Limit Theorem, and maximum likelihood estimation. Hierarchical models are often analyzed using Bayesian methods, and familiarity with these methods is desirable but not essential, as the basics will be covered. Some necessary background can be gained from the text *Statistical Inference*, Casella & Berger, Duxbury 2002, ISBN 0-534-24312-6.

Copies of all course slides and example output discussed will be provided.

**Longitudinal Data Analysis**

**March 8 & 9: Cynthia Wilson Garvan**

As researchers strive to understand both the cross sectional and time dependent relationships of explanatory variables and outcomes, longitudinal studies have become increasingly popular in biomedical applications. In longitudinal studies, repeated measures are made on the same subject (or other experimental unit) over time, inducing the complexity of dependence among measures. Advances in statistical methodology and available software continue to make analysis of such studies feasible. This course will outline the broad issues in longitudinal data analysis and give an overview of methods. Practical applications with illustrating examples will be stressed.

Participants are expected to have the background equivalent of a first year graduate course in statistical theory. Familiarity with rudimentary matrix notation and operations is also assumed. The course is designed to give an overview of the material covered in the following texts: *Analysis of Longitudinal Data*, Diggle, Liang, Zeger, Oxford University Press 1994, ISBN 0-198-52484-6 and *Linear Mixed Models for Longitudinal Data*, Verbeke and Molenberghs, Springer-Verlag 2000, ISBN 0-387-95027-3. There is no required text for the course. Copies of all course slides and example output discussed will be provided.

Main topics covered are:
1. Introduction and examples
2. Objectives of longitudinal analysis
3. Models for longitudinal data
4. Methods for estimating model parameters
5. Inference - confidence intervals and tests of hypotheses
6. Diagnostics to check model fit
7. Graphical presentation of longitudinal data
8. SAS programming
9. Report writing
10. Design consideration
11. Handling missing data

**Bayesian Stochastic Simulation Using WinBUGS**

**March 9: Jeff Gill**

This course is an applied introduction to implementing Markov chain Monte Carlo methods in WinBUGS and R. We will evaluate in detail the process of designing and coding Bayesian statistical models that do not have closed-form analytical solutions. The emphasis will be on applying MCMC theory in a practical manner and producing useful results. Specific topics include: varieties of Markov chains, BUGS syntax and coding, convergence diagnostics, and hybrid solutions.


**Mixed Model and Hierarchical Data Analysis**

**March 10-11: Ramon Littell**

Data sets from designed experiments, sample surveys, and observational studies often contain correlated observations due to random effects and repeated measures. These data sets must be analyzed carefully in order to make valid statistical inferences. Mixed models can be used to accommodate the correlation structure and thereby produce efficient estimates of means and differences between means, and to provide valid estimates of standard errors. Repeated measures and longitudinal data require special attention because they involve correlated data that arise when the primary sampling units are measured repeatedly over time or under different conditions. Normal theory models for random effects and repeated measures ANOVA will be used to introduce the concept of correlated data. PROC GLM and PROC MIXED in the SAS system will be illustrated using practical examples from pharmaceutical studies, clinical trials, environmental studies and laboratory experiments.


Main topics covered are:
1. Split-Plot Designs
2. Cross-Over
3. Hierarchical Models
4. Repeated Measures
5. Variance Component Estimation
6. Best Linear Unbiased Estimation (BLUP)
7. Computer Implementation Using the SAS System

**Clinical Trials, Design, Management, Analysis and Oversight**

**March 10-11: Jonathan Shuster**

The gold standard for demonstrating whether or not a new therapy is effective is the clinical trial. While these trials have contributed much to the advancement of medicine, those conducting the trials and those who oversee their conduct have an enormous responsibility to make sure that the trials are rigorously con-
ducted and that subjects are properly protected from unnecessary risks. Conflicts of interest must be managed to the highest possible ethical standards. Course participants will become familiar with all aspects of setting up a “Protocol” (manual of procedures), which deals with the objectives, background (with preliminary data motivating the trial), eligibility, registration (including randomization), data collection, statistical methods for interim and final analysis, regulatory issues (such as Institutional Review Board Approval, informed consent, reporting adverse events), and data and safety monitoring (including on-site audits, other quality assurance measures, and trial oversight by a Data and Safety Monitoring Committee).

The course is designed for multiple audiences, including physicians, clinical research associates, nurses, clinical trial administrative managers, and biostatisticians. Knowledge of statistical methods will be helpful in only about 20% of the course, but those statistical topics will be discussed at a level such that non-statisticians will gain an appreciation for the underlying principles involved in the design and analysis of the trials.

There is no required text for the course. Copies of all course slides and example output discussed will be provided.

Main topics covered are:

1. A Template for the Design, Conduct, Monitoring, and Analysis of Clinical Trials
2. Types of Trials (Phase I, Phase II, or Phase III)
3. Design issues: Randomization vs. Historical Controls (including Stratification, Blinding and Phase)
4. How Many Patients do we Need?
5. Interim Monitoring of Safety and Efficacy (Treatment Differences)
6. Data Safety and Monitoring Committees
7. Data Management Issues (Including Confidentiality)
8. Quality Assurance and Audits
9. Methods of Analysis of Data (Group Sequential Methods for Phase I, II, III)
10. Bioequivalence Studies (Attempting to Demonstrate No Difference)
11. Subset Conclusions (Do the overall results apply to subsets, such as females or subjects over 60 years of age? Understanding Interaction)
12. Interpreting the Statistical Results for the Lay Population

**Applied Analysis of Time-dependent Data Canceled**

March 8-9: Alex Trindade

Time-dependent data is encountered in a variety of fields of application, including: finance (stock prices, risk management), utilities (seasonal power and water demand), business & marketing (sales, inventory control, production planning), economics (budgeting, economic planning, strategy evaluation), environmental & biological sciences (monitoring pollutant levels, weather forecasting, earthquake analysis, hydrological control, disease and infection rates, longitudinal studies) and engineering (loads and resonant frequencies, electrical circuit modeling).

Data collected over time usually exhibit a high degree of correlation. Its analysis therefore requires an approach different than the traditional linear statistical model framework with i.i.d. errors. This short course will provide a “hands-on” introduction to modern applied time series analysis and forecasting. The main focus will be on the modeling and forecasting of time series software package to demonstrate applications.

No previous knowledge of time series is required but attendees should be familiar with elementary distributions (normal, t, chi-square) and inferential concepts (confidence intervals, hypothesis tests, p-values). Knowledge of trigonometric functions, matrix notation and exposure to multiple linear regression is also necessary. Ideally, attendees will have taken a two-semester undergraduate mathematical statistics sequence, such as might be covered in for example *Mathematical Statistics with Applications*, Wackerly, Mendenhall & Scheaffer, Duxbury 2002, ISBN: 0-534-37741-6.

The text, *Introduction to Time Series and Forecasting*, Brockwell & Davis, Springer-Verlag 2002, ISBN 0-387-95351-5 and accompanying software ITSM2000, are required. If you do not have a copy, please order one from your local bookstore, an internet provider, or on the registration form. Although not a requirement, maximum benefit from the interactive exercises will be derived if attendees bring along a laptop computer outfitted with ITSM2000. Copies of all course slides and example output discussed will be provided.

Topics covered include:

1. Trend and seasonality; testing and estimated noise sequence
2. ARMA(p,q) processes, the ACF and PACF, modeling and forecasting with ARMA processes
3. Non-stationary and seasonal time series models, ARIMA and SARIMA models
4. Regression with time series errors
5. Transfer function modeling
6. Forecasting techniques, ARAR and Holt-Winters algorithms
7. Spectral theory and linear filtering, spectral densities of ARMA processes

**Genetic Data Analysis**

March 9: Mark C.K. Yang

Modern molecular genetics plays a crucial role in biology and medicine and consequently in biostatistics. To really understand how statistics works in this area, one needs to understand certain basic facts in biology and biochemistry. This one class will start from the basic molecular genetics to the different forms of genetics data that require substantial statistical help, including gene hunting, heritability estimation, microarray and genomics. The emphasis is at the conceptual level but with mathematics details whenever necessary.


Topics covered include:

1. Biological & chemical background for Modern Statistical Genetics
2. Linkage Analysis with Qualitative Trait
3. Genetics of Quantitative Trait (QTL)
4. Small Area Gene Mapping by Linkage Disequilibrium
5. Forensic Evidence Using Genetic Markers
6. Statistical Methods in Microarray Analysis
Statistics Short Courses include expert instruction with printed
detailed course notes, continental breakfast, lunch in the Le Jardin Cafe,
morning and afternoon refreshment breaks, and evening receptions.
Required textbooks may be included at an additional cost. One 1-day
course is $700, two 1-day courses are $1,100, one 2-day course is $1,100
and the cost for the complete 4 days of courses is $2,000.

Lodging
A block of rooms has been reserved at the Adam’s Mark Hotel, at the
Florida Mall, Orlando, Florida. The special room rate is $109 single or
double, plus taxes.
The Florida Mall has over 200 specialty stores, department stores and
30 eateries. The Mall also includes two fine restaurants and a Lobby
Louge. There is a pool, a whirlpool and an exercise facility available for
Adam’s Mark guests who stay in one of the 496 oversized custom deco-
rated guest rooms in this 11-story high-rise hotel.
The Adam’s Mark is an ideal meeting headquarters within easy reach
of all the excitement and attractions that Central Florida has to offer.
Situated in south Orlando at Sand Lake Road and U.S. 441, the Adam’s
Mark is six miles west of Orlando International Airport and seven miles
from downtown. The hotel offers free parking, on-site car rental, attraction
tickets and transportation is available to: Walt Disney World,
Universal Studios Florida, Sea World, Church Street Station and the
Kennedy Space Center.
Make reservations by calling (407) 859-1500 [Fax: (407) 855-1585]
or mail to:

In order to receive the special room rate ($109, single or double) you
(or your agent) must clearly indicate that you are attending the
University of Florida Statistics Week of Short Courses. The Hotel
requires a one night deposit to confirm a reservation. Deadline for hotel
reservations is February 5, 2004. After this date, rooms will be on a
space and rate availability basis.

NOTE
Please make your hotel and travel reservations early. The Orlando
area is a major conference and tourist area, and March is a very popular
month to visit Florida. Airlines, auto rentals, and hotels fill up quickly at
this time of the year.

REFUNDS
Full refunds, less a $75 processing fee, will be made if a written
request is postmarked by January 30, 2004. The University of Florida
reserves the right to cancel individual courses or this event. In the event
the University of Florida cancels a course or this event, fees paid for the
courses and texts will be fully refunded. However, the University of
Florida will not be responsible for any other expenses paid by partici-
pants, including, but not limited to travel and hotel costs, regardless of the
date of cancellation. In compliance with the ADA act, participants with
special needs can be reasonably accommodated by contacting Carol
Rozear in the Department of Statistics at the University of Florida before
January 30, 2004. She can be reached by phone at (352) 392-1941 ext.
207, by fax at (352) 392-5175, or by calling 1-800-955-8771 (TDD).