- As part of the **8th Purdue Symposium on Statistics: Diversity in the Statistical Sciences for the 21st Century; June 20-24th, 2012** there will be educational Workshops (June 20-21, 2012).

- The majority of these Workshops are of interest to users and non-statisticians. Please distribute this announcement widely, and refer anyone(students) interested in registering to: http://www.stat.purdue.edu/symp2012/

- Please see details below about Workshops on:

  - **"Bioinformatics"**
  - **"Diversity and Leadership in the Statistical Sciences"**
  - **"Introduction to Clinical Trials in Drug Development"**
  - **"iPlant"**
  - **"Introduction to R for Non-statisticians"**
  - **"Smoothing Splines ANOVA Models"**
  - **"Quantitative Finance"**

- Refreshments (10am and 3pm breaks are included, as are notes for each Workshop)

- Workshop Registration is required: http://www.stat.purdue.edu/symp2012/

June 20, 2012

Wednesday, June 20, 8:30am - 5:00pm and Thursday, June 21, 8:30am - 5:00pm | Chair: Olga Vitek

Statistical design of experiments and linear mixed models, and their applications in the context of bioinformatics (2 days; $300)

Speaker(s)

- Bruce Craig (Purdue University)
- Olga Vitek (Purdue University)

Abstract
The 2-day workshop discusses the use of linear and generalized linear models in design and analysis of experiments, and their applications in modern quantitative high-throughput biological investigations. The morning sessions focus on general aspects of design and analysis of experiments. They cover topics such as fixed effects and mixed effects linear models, analysis of incomplete block designs, analysis of repeated measures data, generalized linear mixed effects models and power determination. The sessions present extensive case studies, and examples of analysis using SAS. The afternoon sessions discuss the use and the extensions of linear and generalized linear modeling for high-throughput quantitative biological applications. The topics include Empirical Bayes linear mixed models for continuous response (e.g. gene expression microarrays and quantitative proteomics and metabolomics), and models for experiments with count response (e.g. RNA-seq and spectral counts proteomics). Additional related topics are normalization, methods to control False Discovery Rate, power determination with multiple tests, and differential expression of pre-defined sets of genes. The discussion is illustrated with case studies, and examples of analysis using R and Bioconductor.

Wednesday, June 20, 1:30pm - 5:00pm | Chair: Chong Gu *(half day; $125)*

Introduction to R for non-statisticians

Speaker(s)

- Chong Gu (Purdue University)

Abstract
R is an open-source software environment for statistical computing and graphics. It runs on all major computer platforms including UNIX, Windows and MacOS. A powerful feature of R is its package system, which facilitates the modular expansion of its functionality via packages developed by both the R developers and the R users. In this tutorial, we will introduce R as a powerful "graphing calculator" with limited reference to its statistical functions. To be discussed are primary R objects and their manipulations, the "vector arithmetic," methods, etc. R resources are archived at CRAN (www.cran.r-project.org), the Comprehensive R Archive Network, from where the source code, the executables, the manuals, the FAQs, and 3500+ packages can be downloaded. Participants are strongly encouraged to visit CRAN prior to the workshop, and download and install R on their laptops.

Wednesday, June 20, 1:30pm - 5:00pm | Chair: Steve Ruberg

Introduction to Clinical Trials in Drug Development (half day; $125)

Speaker(s)

- Steve Ruberg (Purdue University)

Abstract
Drug development by pharmaceutical companies is the major source of new medications entering the marketplace, and clinical trials done by pharma are the primary evidence for marketing approval by regulatory agencies around the world. As such, these trials require the highest level of scientific rigor. This short course will cover the overall structure and process for clinical drug development as well as highlight important scientific/statistical considerations. The course will be presented at a level that does not require in-depth knowledge of statistical theory and methods, but will describe principles for good scientific and statistical practice. Real examples from the pharmaceutical industry will be used to illustrate issues and concepts.

June 21, 2012

Thursday, June 21, 8:30am - 5:00pm | Chair: Chong Gu

Smoothing Spline ANOVA Models (full day; $250)

Speaker(s)

- Chong Gu (Purdue University)

Abstract

In this one-day short course, we will provide an overview of smoothing spline ANOVA models. To be discussed are the model construction via functional ANOVA decomposition with versatile marginal configurations as well as modeling/inferential tools such as cross-validation, Kullback-Leibler projection, and Bayesian confidence intervals. Software implementation in the R package gss will be illustrated using real data examples in Gaussian and non-Gaussian regression, density estimation, and hazard estimation. The intended audience includes applied statisticians looking to expand their tool-box, as well as advanced graduate students looking to broaden their knowledge bases; basic training in statistical modeling and experience with R are assumed. Prospective participants are encouraged to send their own data, of moderate size and complexity, to the instructor. One or two such data sets may be selected to discuss possible modeling options if time permits.

Thursday, June 21, 8:30am - 12:00pm | Chair: José E. Figueroa-López

Quantitative Finance (half day; $125)

Speaker(s)

- Rene Carmona (Princeton University)

Abstract

The quantitative study of uncertainty in finance is decades old, with mathematical underpinnings that trace back to the work of Bachelier in 1900. The last 15 years have seen extraordinary development of probabilistic and statistical tools for the study of stock markets and investment finance, aided by the new availability of powerful computational platforms and data streams which increase in frequency and volume. Some of the most established current foci in quantitative finance include the stochastic theory of absence of arbitrage, and the time series analysis of financial econometrics. In the former, Brownian motion and other continuous-time processes are used to study derivative pricing and extensions to incomplete markets and insider trading, while in the latter, volatility modeling in empirical finance helps understand value-at-risk and many other portfolio risk management tools. These two dominant tool-sets do not typically coexist in modern treatments. This Workshop will foster an environment in which this gap can be addressed, by bringing in world-class specialists and new and rising researchers representing these diverse points of view. Other emerging directions, such as market microstructure, high-frequency finance, statistical inference for stochastic finance, and the study of market crashes, will further expand the scope of this wide-reaching workshop.

Thursday, June 21, 8:30am - 5:00pm | Chair: Mary Ellen Bock

Diversity and Leadership in the Statistical Sciences (free send email to martindl@purdue.edu to register for this Workshop only)

Speaker(s)

- Mary Ellen Bock (Purdue University)

Abstract
Whether in academia, government or industry, today's leaders seek to take advantage of the full range of talents and insights that a diverse workforce brings. This workshop considers strategies for diversifying workforce and leadership in the statistical sciences.

Thursday, June 21, 1:30pm - 5:00pm | Chair: Eric Lyons

iPlant Data Store and iPlant Discovery Environment (half day; $125)

Speaker(s)

- Eric Lyons (University of Arizona)

Abstract
1. iPlant Data Store: Cloud data storage that gives researchers terabytes of storage that can be connected to any online computational resource (cloud computing, supercomputing, local computing). Also provides the means for high-performance and large file transfers (currently a limitation with traditional http/ftp/scp methods) and methods to share and collaborate on large data. This resource is of use for anyone managing large (and small) sets of data.

2. iPlant Discovery Environment: iPlant's web-based system for managing data (provided by the Data Store), integrating analytical tools, and running analyses on cluster and high-performance computing environments. This resources might be of interest for people at the conference wanting to make their tools available for other users in an easy-to-use framework.