

# **8th ANNUAL PROBABILITY AND STATISTICS DAY AT UMBC**

FUNDED BY NATIONAL SECURITY AGENCY

Workshop on  
***Analysis of Overdispersed Data using SAS®***

Presented by

***Dr. Jorge Morel***

***Procter and Gamble***

***Professor Nagaraj K. Neerchal***

***University of Maryland Baltimore County***

**Friday, April 18, 2014 ♦ 2:00 P.M. – 5:45 P.M. ♦ Room 105, Public Policy Building**

Registration is FREE but required! ♦ Registration Deadline: Friday, April 11, 2014

Registration Website: [www.umbc.edu/circ/hosting/ProbStatDay2014](http://www.umbc.edu/circ/hosting/ProbStatDay2014)

## ***Speakers***

Dr. Morel, a PhD in Statistics from Iowa State University, is currently a Principal Statistician and technical leader in the Quantitative Sciences Department of Procter and Gamble (P&G). He is a former Assistant Professor at the Department of Epidemiology and Biostatistics at the University of South Florida. Dr. Morel has played a key role in the development of study designs and analysis plans encompassing many of the current projects at P&G. His research topics include logistic regression with clustered data and modeling of correlated multinomial responses.

Professor Neerchal is likewise a PhD in Statistics from Iowa State University, and currently the Chair of the Department of Mathematics and Statistics at UMBC and a Fellow of the American Statistical Association. Professor Neerchal's main areas of research interest are time series analysis and methods of analyzing correlated categorical data, and he has extensive experience in the applications of statistical methodology to environmental, biological and engineering projects.

## ***Abstract***

Overdispersion (extra variation) arises in binomial/multinomial/count data when variances are larger than those allowed by the Binomial, Multinomial, and Poisson models. This phenomenon is caused by clumping, presence of excess of zeros in the data, lack of independence, or clustering. Commonly used overdispersion models include the Beta-binomial, Random-clumped Binomial, Zero-inflated Binomial, Negative Binomial, Zero-inflated Poisson and Negative Binomial, Dirichlet-Multinomial, Random-clumped Multinomial and Poisson and Negative

Binomial Hurdle models. When covariates are available, the mean and overdispersion parameters can be modeled using appropriate link functions as in Generalized Linear Models (GLM). Such models will be called Generalized Linear Overdispersion Models (GLOM). GLOM do not always belong to the exponential family, and therefore not usually covered under GLM expositions. The aim of the course is to introduce GLOM using several real-life examples, and illustrate the main methods of estimation such as Quasi-likelihood, Maximum Likelihood and Generalized Estimating Equations. Examples will be analyzed using the SAS<sup>®</sup> procedures LOGISTIC, GENMOD, GLIMMIX, COUNTREG, NLMIXED, FMM and SURVEYLOGISTIC.

### **Outline**

#### **Lecture 1 (2:15pm-3:45pm):**

- 1.1 Overdispersion: To be or not to be
- 1.2 Quasi-likelihoods and Generalized Estimating Equations
- 1.3 Example: All Mice Are Created Equal, but Some Are More Equal
- 1.4 Overdispersion Models for Binomial-type of Data: BB, RCB
- 1.5 All Mice Are Created Equal Example Revisited

#### **Lecture 2 (4:15pm-5:45pm):**

- 2.1 Overdispersion Models for Count Data: ZIP, Hurdle Models
- 2.2 Example: Urinary Tract Infections (UTI) in Men Infected with HIV
- 2.3 Example: Milk Does Your Body Good
- 2.4 Generalized Linear Overdispersion Models (GLOM)
- 2.5 Final remarks

This workshop is based on the following recent SAS publication by Dr. Morel & Dr. Neerchal:

