Workshop Title: Monte Carlo Simulation using R

Dr. Gordon Brooks

Dr. Gordon Brooks received the Ph.D. in Educational Research and Evaluation from Ohio University. He holds the B.A. in Computer Science and Economics from the College of William and Mary in Virginia and the M.A. in Interpersonal and Organizational Communication from Ohio University. Dr. Brooks teaches Educational Research and Evaluation courses as an Associate Professor in the Patton College of Education at Ohio University. His primary scholarship examines statistical issues using Monte Carlo simulation techniques, including

topics such as sample size, statistical power, Type I error rates, principal components analysis, and differential item functioning. He has developed computer software with the primary purpose of helping students learn relatively abstract statistical concepts, but that also serves more practical purposes (e.g., *TAP: Test Analysis Program*, *MCMR: Monte Carlo Multiple Regression*, and *FISH: Friendly Introductory Statistics Help*). Dr. Brooks currently serves as the Immediate Past President of the Mid-Western Educational Research Association, serves on the Editorial Board of the journal *Multiple Linear Regression Viewpoints*, and has been an invited reviewer for seven other journals and several textbook publishers. He recently finished a 3-year term as Department Chair for Educational Studies at Ohio University, currently serves as Program Coordinator for the Educational Research and Evaluation program, and also serves on the Board of Education of the Alexander Local School District.

Workshop Description (2 sessions):

Morning session: The primary purpose of this hands-on workshop is to introduce a free and powerful software package called R and to prepare workshop participants to use R for research and applied statistics. After a brief introduction of R, which will include such things as installing new packages, using front-end programs and GUIs, and keeping the program updated, the workshop will be organized into the following sections where participants can follow along on their own computers: (a) data management in R, (b) descriptive statistics in R, and (c) basic and intermediate statistical analyses in R. Basic data management will include how to import data into R from other programs (e.g., Excel and SPSS) and manipulate data (e.g., computing and recoding variables), including a description of how R handles various data types, data structures, and missing values. We will demonstrate creating and obtaining R data using several methods. Next, basic statistical analyses will be performed in R, including calculating descriptive statistics (e.g., mean, median, standard deviation, frequency tables, and cross-tabulations), creating common graphs (e.g., histograms, scatterplots), and performing some basic measurement analyses (e.g., reliability). Third, several common statistical analyses will be presented, including correlations, t tests, ANOVA, and multiple regression. Several methods for performing measurement analyses will be covered, including classical test theory analyses, IRT analyses, and applications such as DIF analysis will also be covered. Relevant analyses for screening data (e.g., outliers) and testing assumptions will also be demonstrated.

Afternoon session: Because simulation studies provide useful information about the effectiveness of statistical and measurement procedures, Monte Carlo simulation studies are important tools in both statistical and psychometric research. For example, bias, precision, Type I errors, statistical power can all be studied through Monte Carlo simulation methods.   The R statistical program provides a programming interface that enables researchers with moderate programming skills to develop and perform Monte Carlo simulations for research. R scripts can also be written to

interface with other programs (e.g., BILOG). This workshop will demonstrate how R can be used to generate data and perform both statistical and measurement simulation studies. Participants will be able to run the examples on their own computers during the presentations. Topics to be covered include (a) data generation, (b) basic programming (e.g., logic, loops), and (c) running Monte Carlo simulations. First, we will demonstrate how to generate both independent and correlated data for statistical simulations. We will also demonstrate how IRT data can be generated for measurement Monte Carlo simulations. Second, basic programming required in Monte Carlo simulations will be demonstrated. For example, participants will learn how to obtain the values they need for their research, how to store the values necessary for their analyses, how to create loops to repeat the analyses over many replications, and how to use logic to test conditions within the simulations (e.g., if-then-else). Third, we will demonstrate through several examples how to create the logic and run the simulations needed to answers the Monte Carlo research questions. We will provide illustrations of both statistical simulations and measurement simulations. Finally, we will present ways to organize the data and information in useful ways for reporting the results of the simulation.

Intended Audience:

Participants in this workshop should have a foundation in statistical and measurement methods. No previous experience with R will be assumed. No previous computer programming experience is required. To receive full benefit from the workshop, participants should bring their own laptop computers with R already installed (R is available free at http://www.r-project.org/ by following the "Download, Packages" CRAN link). If participants cannot install R prior to the workshop, a CD-ROM will be provided at the workshop from which they can run R on their own computers during the workshop.