Bringing multivariate statistics to the masses: concepts and pedagogy for post-introductory courses in statistics

Nathan L. Tintle*
Dordt College, Sioux Center, Iowa, USA, nathan.tintle@dordt.edu

Beth Chance
California Polytechnic University, San Luis Obispo, California, USA, bchance@calpoly.edu

George Cobb
Mt. Holyoke College, South Hadley, Massachusetts, USA, gcobb@mtholyoke.edu

Allan Rossman
California Polytechnic University, San Luis Obispo, California, USA, arossman@calpoly.edu

Soma Roy
California Polytechnic University, San Luis Obispo, California, USA, soroy@calopoly.edu

Todd Swanson
Hope College, Holland, Michigan, USA, swansont@hope.edu

Jill VanderStoep
Hope College, Holland, Michigan, USA, vanderstoepj@hope.edu

The demands for a statistically literate society are increasing, and the introductory statistics course ("Stat 101") remains the primary venue for learning statistics for the majority of high school and undergraduate students. After three decades of very fruitful activity in the areas of pedagogy and assessment, but with comparatively little pressure for rethinking the content of this course, the statistics education community has recently turned its attention to use of randomization-based methods to illustrate core concepts of statistical inference. This new focus not only presents an opportunity to address documented shortcomings in the standard Stat 101 course (for example, improving students' reasoning about inference), but provides an impetus for re-thinking the timing of the introduction of multivariable statistical methods (for example, multiple regression and general linear models). Multivariable methods dominate modern statistical practice but are rarely seen in the introductory course. Instead these methods have been, traditionally, relegated to second courses in statistics for students with a background in calculus and linear algebra. Recently, curricula have been developed to bring multivariable content to students who have only taken a Stat 101 course. However, these courses tend to focus on models and model-building as an end in itself. We have developed a preliminary version of an integrated one to two semester curriculum which introduces students to the core-logic of statistical inference through randomization-methods, and then introduces students to approaches for protecting against confounding and variability through multivariable statistical design and analysis techniques. The course has been developed by putting primary emphasis on the development of students' conceptual understanding in an intuitive, cyclical, active-learning pedagogy, while continuing to emphasize the overall process of statistical investigations, from

asking questions and collecting data through making inferences and drawing conclusions. The curriculum successfully introduces introductory statistics students to multivariable techniques in their first or second course.

Key words: variability; Stat 101; randomization-based; second course