BOSTON UNIVERSITY STATISTICS AND PROBABILITY SEMINAR SERIES

Using Regression Models to Analyze Randomized Trials: Asymptotically Valid Hypothesis Tests Despite Incorrectly Specified Models

Michael Rosenblum

Center for AIDS Prevention Studies (CAPS)

University of California, San Francisco

January 29, 4:30-5:30pm
Mathematics and Computer Science (MCS) Building, Room 149
111 Cummington Street, Boston
Tea and Cookies at 3:30pm in MCS 153

Abstract: Regression models are often used to test for cause-effect relationships from data collected in randomized trials or experiments. This practice has deservedly come under heavy scrutiny, since commonly used models such as linear and logistic regression will often not capture the actual relationships between variables, and incorrectly specified models potentially lead to incorrect conclusions. In this paper, we focus on hypothesis tests of whether the treatment given in a randomized trial has any effect on the mean of the primary outcome, within strata of baseline variables such as age, sex, and health status. Our primary concern is ensuring that such hypothesis tests have correct Type I error for large samples. Our main result is that for a surprisingly large class of commonly used regression models, standard regression-based hypothesis tests (but using robust variance estimators) are guaranteed to have correct Type I error for large samples, even when the models are incorrec! tly specified. To the best of our knowledge, this robustness of such model-based hypothesis tests to incorrectly specified models was previously unknown for Poisson regression models and for other commonly used models we consider. Our results have practical implications for understanding the reliability of commonly used, model-based tests for analyzing randomized trials.

For directions and maps, please see *http://math.bu.edu/research/statistics/statseminar.html.*