Linear Regression Analysis in Non-linear Populations

Lawrence D. Brown*, R. Berk, A. Buja, E. George, D. Mc Carthy, E. Pitkin, K. Zhang and L. Zhao.
Statistics Department, Wharton School, Univ. of Pennsylvania
Philadelphia, PA, 19104, USA
* = presenting author and corresponding author.

email: lbrown@wharton.upenn.edu. All other authors can be reached via noellef@wharton.upenn.edu

Abstract:

Linear regression analysis is often applied to populations that do not satisfy the conventional assumptions of linearity, homoscedasticity and normality of residuals. Furthermore, the classical theory considers the covariates as fixed constants, whereas in most applications in the social sciences (and elsewhere) they are actually random variables. This talk surveys a comprehensive theory of linear regression that is free of any of these classical assumptions. Although much of the talk is "expository" some new results will be included related to the structure of statistical errors in such situations. In particular, alternate forms of the familiar Huber-Eicker-White "sandwich" estimator are described as well as a coordinate specific test for conventional assumptions. Results concern both inference about the linear regression coefficients and about best linear predictive inference. The basic perspectives thus involve assumption-lean statistical populations that include random covariates that may have an unspecified distribution, and then applying assumption-rich statistical models as approximations. If time permits, some further developments built on this approach will be discussed. These include an extension of the perspective to settings often described as semi-supervised learning in which additional observations are available on the covariates unaccompanied by the response of interest. The perspective also leads to a simple but effective estimator for the Average Treatment Effect in randomized trials involving covariates. Results in these additional directions will be surveyed as time allows.

Key words: Least-squares estimates, Robust regression analysis, Random design