Cook's Distance

Cook (1977) greatly expanded the study of residuals and influence measures. Cook's key observation was the effects of deleting each observation in turn could be computed without undue additional computational expense. Consequently deletion diagnostics have become an integral part of assessing linear models.

Cook's Distance

In classical linear regression, a commonly used meausre of influence is Cook's distance. It is used as a measure of influence on the regression coefficients.

Cook's Distance

Cooks Distance (D_i) is an overall measure of the combined impact of the ith case of all estimated regression coefficients. It uses the same structure for measuring the combined impact of the differences in the estimated regression coefficients when the $i-$th case is deleted.

Importantly, $D_{(i)}$ can be calculated without fitting a new regression coefficient each time an observation is deleted.

0.1 Cooks's Distance

Cook's D statistics (i.e. colloquially Cook's Distance) is a measure of the influence of observations in subset U on a vector of parameter estimates.

$$\delta_{(U)} = \hat{\beta} - \hat{\beta}_{(U)}$$

If V is known, Cook's D can be calibrated according to a chi-square distribution with degrees of freedom equal to the rank of $\boldsymbol{X}$.

For LME models, Cook's distance can be extended to model influence diagnostics by definining.

It is also desirable to measure the influence of the case deletions on the covariance matrix of $\hat{\beta}$.