

1 Cook's Distance

1.1 Cook's Distance

Cooks Distance (D_i) is an overall measure of the combined impact of the i th 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 k th case is deleted. $D_{(k)}$ can be calculated without fitting a new regression coefficient each time an observation is deleted.

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 is a well known diagnostic technique used in classical linear models, extended to LME models. For LME models, two formulations exist; a Cook's distance that examines the change in fixed fixed parameter estimates, and another that examines the change in random effects parameter estimates. The outcome of either Cook's distance is a scaled change in either β or θ .

1.2 Cook'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 \mathbf{X} .

1.3 Cook's Distance

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

For linear mixed effects models, Cook's distance can be extended to model influence diagnostics by defining.

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