

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 (1986) gave a completely general method for assessing influence of local departures from assumptions in statistical models.

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.

Cook's Distance

Cook's Distance (D_i) is an overall measure of the combined impact of the i th case on 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.

0.0.1 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} .

For LME 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 $\hat{\beta}$.

0.1 Cook's Distance for LMEs

Diagnostic methods for fixed effects are generally analogues of methods used in classical linear models. Diagnostic methods for variance components are based on ‘one-step’ methods. *cook86* gives a completely general method for assessing the influence of local departures from assumptions in statistical models.

For fixed effects parameter estimates in LME models, the Cook's distance can be extended to measure influence on these fixed effects.

$$CD_i(\beta) = \frac{(c_{ii} - r_{ii}) \times t_i^2}{r_{ii} \times p}$$

For random effect estimates, the Cook's distance is

$$CD_i(b) = g_{(i)}(I_r + \text{var}(\hat{b})D)^{-2}\text{var}(\hat{b})g_{(i)}.$$

Large values for Cook's distance indicate observations for special attention.

0.1.1 Change in the precision of estimates

The effect on the precision of estimates is separate from the effect on the point estimates. Data points that have a small Cook's distance, for example, can still greatly affect hypothesis tests and confidence intervals, if their influence on the precision of the estimates is large.