0.1 Measures of Influence

The impact of an observation on a regression fitting can be determined by the difference between the estimated regression coefficient of a model with all observations and the estimated coefficient when the particular observation is deleted. The measure DFBETA is the studentized value of this difference.

Influence arises at two stages of the LME model. Firstly when V is estimated by $\hat{V}$, and subsequent estimations of the fixed and random regression coefficients β and u, given $\hat{V}$.

0.1.1 DFBETA

$$
\begin{aligned}
DFBETA_a &= \hat{\beta} - \hat{\beta}_{(a)} & (1) \\
&= B(Y - Y_{\bar{a}} & (2)
\end{aligned}
$$

0.1.2 DFFITS

DFFITS is a statistical measured designed to a show how influential an observation is in a statistical model. It is closely related to the studentized residual.

$$
DFFITS = \frac{\widehat{y_i} - \widehat{y_{i(k)}}}{s_{(k)}\sqrt{h_{ii}}}
$$

0.1.3 PRESS

The prediction residual sum of squares (PRESS) is an value associated with this calculation. When fitting linear models, PRESS can be used as a criterion for model selection, with smaller values indicating better model fits.

$$
PRESS = \sum (y - y^{(k)})^2 \tag{3}
$$

- $e_{-Q} = y_Q - x_Q \hat{\beta}^{-Q}$

- $PRESS_{(U)} = y_i - x\hat{\beta}_{(U)}$

0.1.4 DFBETA

$$DFBETA_a = \hat{\beta} - \hat{\beta}_{(a)} \qquad (4)$$
$$= B(Y - Y_{\bar{a}} \qquad (5)$$

0.1.5 Influential Observations : DFBeta and DFBetas

Cook's distance refers to how far, on average, predicted y-values will move if the observation in question is dropped from the data set. dfbeta refers to how much a parameter estimate changes if the observation in question is dropped from the data set. Note that with k covariates, there will be k+1 dfbetas (the intercept,β_0, and 1 β for each covariate). Cook's distance is presumably more important to you if you are doing predictive modeling, whereas dfbeta is more important in explanatory modeling.