0.1 Outliers and Leverage

The question of whether or not a point should be considered an outlier must also be addressed. An outlier is an observation whose true value is unusual given its value on the predictor variables. The leverage of an observation is a further consideration. Leverage describes an observation with an extreme value on a predictor variable is a point with high leverage. High leverage points can have a great amount of effect on the estimate of regression coefficients.

Influence can be thought of as the product of leverage and outlierness. An observation is said to be influential if removing the observation substantially changes the estimate of the regression coefficients. The `R` programming language has a variety of methods used to study each of the aspects for a linear model. While linear models and GLMS can be studied with a wide range of well-established diagnostic technqiues, the choice of methodology is much more restricted for the case of LMEs.

0.1.1 Residual diagnostics

For classical linear models, residual diagnostics are typically conducted using a plot of the observed residuals and the predicted values. A visual inspection for the presence of trends inform the analyst on the validity of distributional assumptions, and to detect outliers and influential observations.

0.1.2 Leverage

In statistics, leverage is a term used in connection with regression analysis and, in particular, in analyses aimed at identifying those observations that are far away from corresponding average predictor values. Leverage points do not necessarily have a large effect on the outcome of fitting regression models.

Leverage points are those observations, if any, made at extreme or outlying values of the independent variables such that the lack of neighboring observations means that the fitted regression model will pass close to that particular observation.

Modern computer packages for statistical analysis include, as part of their facilities for regression analysis, various quantitative measures for identifying influential observations: among these measures is partial leverage, a measure of how a variable contributes to the leverage of a datum.

0.1.3 Leverage

Leverage can be defined through the projection matrix that results from a transformation of the model with the inverse of the Cholesky decomposition of $\boldsymbol{V}$, or an oblique projector: $\boldsymbol{Y} = \boldsymbol{H}\hat{\boldsymbol{Y}}$.

While H is idempotent, it is generally not symmetric and thus not a projection matrix in the narrow sense.

$$h_{ii} = x_i'(X'X)^{-1}x_i$$

The trace of $\boldsymbol{H}$ equals the rank of $\boldsymbol{X}$. If V_{ij} denotes the element in row i, column j of $\boldsymbol{V}^{-1}$, then for a model containing only an intercept the diagonal elements of $\boldsymbol{H}$.

$$h_{ii} = \frac{\sum v_{ij}}{\sum\sum v_{ij}}$$

0.1.4 Leverage

The leverage of an observation is based on how much the observation's value on the predictor variable differs from the mean of the predictor variable. The greater an observation's leverage, the more potential it has to be an influential observation.

For example, an observation with a value equal to the mean on the predictor variable

has no influence on the slope of the regression line regardless of its value on the criterion variable. On the other hand, an observation that is extreme on the predictor variable has the potential to affect the slope greatly.

Calculation of Leverage (h)

The first step is to standardize the predictor variable so that it has a mean of 0 and a standard deviation of 1. Then, the leverage (h) is computed by squaring the observation's value on the standardized predictor variable, adding 1, and dividing by the number of observations.

0.1.5 Leverage in LME models SAS

For the general mixed model, leverage can be defined through the projection matrix that results from a transformation of the model with the inverse of the Cholesky decomposition of , or through an oblique projector (Schabenberger, 2004). The MIXED procedure follows the latter path in the computation of influence diagnostics.

The leverage value reported for the th observation is the th diagonal entry of the matrix

which is the weight of the observation in contributing to its own predicted value, . While is idempotent, it is generally not symmetric and thus not a projection matrix in the narrow sense.

The properties of these leverages are generalizations of the properties in models with diagonal variance-covariance matrices. For example, , and in a model with intercept and , the leverage values

are and . The lower bound for is achieved in an intercept-only model, and the upper bound is achieved in a saturated model.

The trace of equals the rank of . If denotes the element in row , column of , then for a model containing only an intercept the diagonal elements of are

Because is a sum of elements in the th row of the inverse variance-covariance matrix, can be negative, even if the correlations among data points are nonnegative. In case of a saturated model with , .

Nobre Singer : Mixed Model Residuals

Usually one assumes

- $b_i \sim N_q(0, G) i = 1, ..., m$
- $e_i \sim N_{n_i}(0, \sigma_i)$
- b_i and e_i independent
- G and σ_i are $(q \times q)$ and $(n_i \times n_i)$ positive denite matrices with elements expressed as functions of a vector of covariance parameters θ not functionally related to β
- If $\sigma_i = I_{n_i}\sigma^2$: homoskedastic conditional independence model

$$\begin{bmatrix} \boldsymbol{b} \\ \boldsymbol{e} \end{bmatrix} \sim \mathcal{N}_{qm+n}$$

$$\boldsymbol{Q} = \boldsymbol{V}^{-1} - \boldsymbol{V}^{-1}\boldsymbol{X}(\boldsymbol{X}^T\boldsymbol{V}^{-1}\boldsymbol{X})^{-1}$$

Sensitivity and residual analysis of the underlying assumptions constitute important tools for evaluating the fit of any model to given data.

Bibliography

Schabenberger, O. (2004). Mixed model influence diagnostics. 18929.