STAT 574 Linear and Nonlinear Mixed Models

Lecture 10: Diagnoses and Influence Analysis

Chencheng Cai

Washington State University

Influence analysis

- Influence analysis is a set of techniques used to identify and assess the impact of individual data points on the overall model fit and parameter estimates.
- In this lecture, we consider the **influence analysis** as a **sensitivity analysis** of the model fit to the data.
- Data influence: the sensitivity of the model to a infinitesimal purturbation in the data.
- Model influence: the sensitivity of the model to the assumptions.

Linear Regression Model

Consider a linear regression model:

$$y_i = \boldsymbol{\beta}^T \boldsymbol{x}_i + \epsilon_i \quad \forall i.$$

The OLS estimator of $\boldsymbol{\beta}$ is given by

$$\hat{\boldsymbol{\beta}} = (\boldsymbol{X}^T \boldsymbol{X})^{-1} \boldsymbol{X}^T \boldsymbol{y} = \left(\sum_i \boldsymbol{x}_i \boldsymbol{x}_i^T \right)^{-1} \sum_i \boldsymbol{x}_i y_i,$$

where $\boldsymbol{X}$ is the design matrix and $\boldsymbol{y}$ is the response vector.

Leverage

The **leverage** of the i-th observation is defined as the i-th diagonal element of the hat matrix

$$\boldsymbol{H} = \boldsymbol{X}(\boldsymbol{X}^T\boldsymbol{X})^{-1}\boldsymbol{X}^T.$$

That is

$$h_i = (\boldsymbol{H})_{ii} = \boldsymbol{x}_i^T(\boldsymbol{X}^T\boldsymbol{X})^{-1}\boldsymbol{x}_i.$$

Leverage

The sum of the leverages is equal to the number of parameters in the model:

$$\begin{aligned}
\sum_{i=1}^{n} h_i &= \mathrm{tr}(\boldsymbol{H}) \\
&= \mathrm{tr}(\boldsymbol{X}(\boldsymbol{X}^T\boldsymbol{X})^{-1}\boldsymbol{X}^T) \\
&= \mathrm{tr}((\boldsymbol{X}^T\boldsymbol{X})^{-1}\boldsymbol{X}^T\boldsymbol{X}) \\
&= \mathrm{tr}(\boldsymbol{I}) \\
&= m,
\end{aligned}$$

where m is the number of parameters in the model.

- Observations with high leverage are called **influential observations**.
- It measures how the predicted value is influenced by the i-th observation.

Leave-one-out

Another measure of influence is to check the change in the estimates of the parameters when the i-th observation is removed from the data set.
The estimated parameter $\hat{\boldsymbol{\beta}}_{(i)}$ when the i-th observation is removed is given by

$$\hat{\boldsymbol{\beta}}_{(i)} = (\boldsymbol{X}_{(i)}^T \boldsymbol{X}_{(i)})^{-1} \boldsymbol{X}_{(i)}^T \boldsymbol{y}_{(i)},$$

where $\boldsymbol{X}_{(i)}$ and $\boldsymbol{y}_{(i)}$ are the design matrix and response vector with the i-th observation removed.

Leave-one-out

We notice that

$$(\boldsymbol{X}_{(i)}^T\boldsymbol{X}_{(i)})^{-1} = (\boldsymbol{X}^T\boldsymbol{X} - \boldsymbol{x}_i\boldsymbol{x}_i^T)^{-1} = (\boldsymbol{X}^T\boldsymbol{X})^{-1} + \frac{(\boldsymbol{X}^T\boldsymbol{X})^{-1}(\boldsymbol{x}_i\boldsymbol{x}_i^T)(\boldsymbol{X}^T\boldsymbol{X})^{-1}}{1 - \boldsymbol{x}_i^T(\boldsymbol{X}^T\boldsymbol{X})^{-1}\boldsymbol{x}_i}$$

Then

$$\begin{aligned}
\hat{\boldsymbol{\beta}}_{(i)} &= \left(\boldsymbol{X}^T\boldsymbol{X} - \boldsymbol{x}_i\boldsymbol{x}_i^T\right)^{-1}(\boldsymbol{X}^T\boldsymbol{y} - \boldsymbol{x}_i y_i) \\
&= \left[(\boldsymbol{X}^T\boldsymbol{X})^{-1} + (1-h_i)^{-1}(\boldsymbol{X}^T\boldsymbol{X})^{-1}\boldsymbol{x}_i\boldsymbol{x}_i^T(\boldsymbol{X}^T\boldsymbol{X})^{-1}\right](\boldsymbol{X}^T\boldsymbol{y} - \boldsymbol{x}_i y_i) \\
&= \hat{\boldsymbol{\beta}} + (1-h_i)^{-1}(\boldsymbol{X}^T\boldsymbol{X})^{-1}\boldsymbol{x}_i\boldsymbol{x}_i^T(\boldsymbol{X}^T\boldsymbol{X})^{-1}\boldsymbol{X}^T\boldsymbol{y} - (\boldsymbol{X}^T\boldsymbol{X})^{-1}\boldsymbol{x}_i y_i \\
&\quad - (1-h_i)^{-1}(\boldsymbol{X}^T\boldsymbol{X})^{-1}\boldsymbol{x}_i\boldsymbol{x}_i^T(\boldsymbol{X}^T\boldsymbol{X})^{-1}\boldsymbol{x}_i y_i \\
&= \hat{\boldsymbol{\beta}} + (1-h_i)^{-1}(\boldsymbol{X}^T\boldsymbol{X})^{-1}\boldsymbol{x}_i\hat{y}_i - (\boldsymbol{X}^T\boldsymbol{X})^{-1}\boldsymbol{x}_i y_i - \frac{h_i}{1-h_i}(\boldsymbol{X}^T\boldsymbol{X})^{-1}\boldsymbol{x}_i y_i \\
&= \hat{\boldsymbol{\beta}} - \frac{y_i - \hat{y}_i}{1-h_i}(\boldsymbol{X}^T\boldsymbol{X})^{-1}\boldsymbol{x}_i.
\end{aligned}$$

Cook's distance

The confidence region for the estimator $\hat{\boldsymbol{\beta}}$ is given by

$$\left\{\boldsymbol{\beta} : (\boldsymbol{\beta} - \hat{\boldsymbol{\beta}})^T (\boldsymbol{X}^T \boldsymbol{X})^{-1} (\boldsymbol{\beta} - \hat{\boldsymbol{\beta}}) \leq m s^2 F_{\alpha,m,n-m}\right\}$$

The **Cook's distance** is defined as

$$D_i = \frac{1}{ms^2} (\hat{\boldsymbol{\beta}}_{(i)} - \hat{\boldsymbol{\beta}})^T (\boldsymbol{X}^T \boldsymbol{X})^{-1} (\hat{\boldsymbol{\beta}}_{(i)} - \hat{\boldsymbol{\beta}})$$

By previous result, we have

$$D_i = \frac{(y_i - \hat{y}_i)^2}{ms^2} \frac{h_i}{(1 - h_i)^2}$$

Large values of D_i indicate that the i-th observation has a large influence on the fitted model.

Infinitesimal Influence (I-influence)

- $\boldsymbol{D}$: data vector including all the observed values.
- $\boldsymbol{t}(\boldsymbol{D})$: a statistic of interest.
- The **infinitesimal data influence** is

$$\lim_{\Delta D \to 0} \frac{\boldsymbol{t}(\boldsymbol{D} + \Delta D \boldsymbol{e}_i) - \boldsymbol{t}(\boldsymbol{D})}{\Delta D} = \frac{\partial \boldsymbol{t}(\boldsymbol{D})}{\partial D_i}$$

Infinitesimal Influence (I-influence)

- Let $\ell(\boldsymbol{\theta})$ be the log-likelihood function of the model.
- Consider a more general model $\ell(\boldsymbol{\theta} \mid \boldsymbol{\omega})$ such that $\boldsymbol{\omega} = 0$ corresponds to the model of interest.
- The **infinitesimal model influence** is defined as

$$\left. \frac{\partial \boldsymbol{t}}{\partial \boldsymbol{\omega}} \right|_{\boldsymbol{\omega}=0}$$

Influence of the Dependent Variable

We consider the influence of y_i on the estimated parameters $\hat{\boldsymbol{\beta}}$.

- The OLS solution is

$$\hat{\boldsymbol{\beta}} = (\boldsymbol{X}^T\boldsymbol{X})^{-1}\boldsymbol{X}^T\boldsymbol{y} = (\boldsymbol{X}^T\boldsymbol{X})^{-1}\sum_i \boldsymbol{x}_i y_i.$$

- The influence of y_i on $\hat{\boldsymbol{\beta}}$ is given by

$$\frac{\partial\hat{\boldsymbol{\beta}}}{\partial y_i} = (\boldsymbol{X}^T\boldsymbol{X})^{-1}\boldsymbol{x}_i.$$

- Two possibilities that the size of the influence is large:
  - $|\boldsymbol{x}_i|$ is large: the i-th observation is far from the center of the data.
  - The direction of $\boldsymbol{x}_i$ is close to the direction of the eigenvector of $\boldsymbol{X}^T\boldsymbol{X}$ corresponding to the smallest eigenvalue.

Influence of the Continuous Explanatory Variable

In particular, we are interested in

$$\frac{\partial \hat{\boldsymbol{\beta}}}{\partial x_{ik}},$$

where x_{ik} is the k-th element of $\boldsymbol{x}_i$.
Use matrix calculus, we have (by considering $\boldsymbol{x}_i$ as the only variable)

$$\begin{aligned}
d\hat{\boldsymbol{\beta}} &= (\boldsymbol{X}^T\boldsymbol{X})^{-1} d(\boldsymbol{X}^T\boldsymbol{y}) + \left(d(\boldsymbol{X}^T\boldsymbol{X})^{-1}\right)(\boldsymbol{X}^T\boldsymbol{y}) \\
&= (\boldsymbol{X}^T\boldsymbol{X})^{-1} y_i d\boldsymbol{x}_i - (\boldsymbol{X}^T\boldsymbol{X})^{-1}(\boldsymbol{x}_i d\boldsymbol{x}_i^T + (d\boldsymbol{x}_i)\boldsymbol{x}_i^T)(\boldsymbol{X}^T\boldsymbol{X})^{-1}(\boldsymbol{X}^T\boldsymbol{y}) \\
&= (\boldsymbol{X}^T\boldsymbol{X})^{-1} y_i d\boldsymbol{x}_i - (\boldsymbol{X}^T\boldsymbol{X})^{-1}\boldsymbol{x}_i (d\boldsymbol{x}_i)^T\hat{\boldsymbol{\beta}} - (\boldsymbol{X}^T\boldsymbol{X})^{-1}(d\boldsymbol{x}_i)\hat{y}_i \\
&= (\boldsymbol{X}^T\boldsymbol{X})^{-1}(y_i - \hat{y}_i) d\boldsymbol{x}_i - (\boldsymbol{X}^T\boldsymbol{X})^{-1}\boldsymbol{x}_i \hat{\boldsymbol{\beta}}^T d\boldsymbol{x}_i.
\end{aligned}$$

Therefore,

$$\frac{\partial \hat{\boldsymbol{\beta}}}{\partial \boldsymbol{x}_i} = (\boldsymbol{X}^T\boldsymbol{X})^{-1}\left((y_i - \hat{y}_i)\boldsymbol{I} - \boldsymbol{x}_i\hat{\boldsymbol{\beta}}^T\right)$$

Influence of the Continuous Explanatory Variable

Now we have

$$\frac{\partial \hat{\boldsymbol{\beta}}}{\partial x_{ik}} = (\boldsymbol{X}^T\boldsymbol{X})^{-1}\left((y_i - \hat{y}_i)\boldsymbol{e}_k - \hat{\beta}_k \boldsymbol{x}_i\right)$$

- First component: normalized residual.
- Second component: the influence through the dependent variable.

It is also connected to Cook's local influence, which is measured by

$$\text{local influence of } x_{ik} \text{ on } \hat{\beta}_k = y_i - \hat{y}_i - \hat{\beta}_k q_i$$

where q_i is the residual of the regression of $\boldsymbol{x}^{(k)}$ on the other variables.

Influence of the Binary Explanatory Variable

Now we assume x_{ik} is a binary variable.

- A binary variable can be misclassified.
- We assume x_{ik} is an observation for the true binary variable z_{ik} such that miscalssification occurs with probability q_i.
- The true model should be

$$E(y_i \mid z_{ik}) = \alpha + \beta_k z_{ik}$$

- But now

$$E(y_i \mid x_{ik}) = \alpha + \beta_k(x_{ik} + (1 - 2x_{ik})q_i)$$

- The influence of the misclassification is given by

$$\left.\frac{\partial \hat{\boldsymbol{\beta}}}{\partial q_i}\right|_{q_i=0} = (1 - 2x_{ik})(\boldsymbol{X}^T\boldsymbol{X})^{-1}\Big((y_i - \hat{y}_i)\boldsymbol{e}_i - \hat{\beta}_k \boldsymbol{x}_i\Big)$$

Influence on the Predicted Value

The predicted value is connected to the estimated parameters by

$$\hat{y}_i = \hat{\boldsymbol{\beta}}^T \boldsymbol{x}_i.$$

The influence can be transferred to the influence on the predicted value. That is

$$\begin{aligned}
\frac{\partial \hat{y}_i}{\partial x_{ik}} &= \frac{\partial \hat{y}_i}{\partial \hat{\boldsymbol{\beta}}} \frac{\partial \hat{\boldsymbol{\beta}}}{\partial x_{ik}} + \left. \frac{\partial \hat{y}_i}{\partial x_{ik}} \right|_{\hat{\boldsymbol{\beta}}} \\
&= \boldsymbol{x}_i^T (\boldsymbol{X}^T \boldsymbol{X})^{-1} \Big((y_i - \hat{y}_i) \boldsymbol{e}_k - \hat{\beta}_k \boldsymbol{x}_i \Big) + \hat{\beta}_k \\
&= \boldsymbol{x}_i^T (\boldsymbol{X}^T \boldsymbol{x})^{-1} \boldsymbol{e}_k (y_i - \hat{y}_i) + (1 - h_i) \hat{\beta}_k.
\end{aligned}$$

Influence on Regression Characteristics

Y-influence on

- **Coefficient of determination** R^2:

$$\frac{\partial R^2}{\partial y_i} = \frac{2}{\mathrm{SST}} \left[(1 - R^2)(y_i - \bar{y}) - (y_i - \hat{y}_i)\right]$$

- **t-statistics** with $\boldsymbol{t} = s^{-1}\boldsymbol{D}^{-1/2}\hat{\boldsymbol{\beta}}$ where $\boldsymbol{D} = \mathrm{diag}((\boldsymbol{X}^T\boldsymbol{X})^{-1})$:

$$\frac{\partial \boldsymbol{t}}{\partial y_i} = s^{-1}\boldsymbol{D}^{-1/2}(\boldsymbol{X}^T\boldsymbol{X})^{-1}\boldsymbol{x}_i - \frac{y_i - \hat{y}_i}{\mathrm{RSS}}\boldsymbol{t}_i$$

Nonlinear Regression Models

Consider the nonlinear regression model:

$$y_i = f_i(\boldsymbol{\beta}, \boldsymbol{x}_i) + \epsilon_i$$

The LS estimator (LSE) of $\boldsymbol{\beta}$ is given by

$$\hat{\boldsymbol{\beta}} = \arg\min_{\boldsymbol{\beta}} \sum_i (y_i - f_i(\boldsymbol{\beta}, \boldsymbol{x}_i))^2.$$

The estimating equation is given by

$$\sum_i (y_i - f_i(\boldsymbol{\beta}, \boldsymbol{x}_i)) \frac{\partial f_i(\boldsymbol{\beta}, \boldsymbol{x}_i)}{\partial \boldsymbol{\beta}} = 0.$$

Influence of the dependent variable on the LSE

The influence of y_i on the LSE $\hat{\boldsymbol{\beta}}$ is

$$\frac{\partial \hat{\boldsymbol{\beta}}}{\partial y_i}$$

Taking partial derivative of the estimating equation w.r.t. y_i, we have

$$\sum_j \left(\delta_{ij} - \left(\frac{\partial f_j(\hat{\boldsymbol{\beta}}, \boldsymbol{x}_j)}{\partial \hat{\boldsymbol{\beta}}} \right)^T \frac{\partial \hat{\boldsymbol{\beta}}}{\partial y_i} \right) \frac{\partial f_j(\hat{\boldsymbol{\beta}}, \boldsymbol{x}_j)}{\partial \hat{\boldsymbol{\beta}}} + \sum_j (y_j - f_j(\hat{\boldsymbol{\beta}}, \boldsymbol{x}_j)) \frac{\partial^2 f_j(\hat{\boldsymbol{\beta}}, \boldsymbol{x}_j)}{\partial \hat{\boldsymbol{\beta}}^2} \frac{\partial \hat{\boldsymbol{\beta}}}{\partial y_i} = 0$$

This gives

$$\frac{\partial \hat{\boldsymbol{\beta}}}{\partial y_i} = \boldsymbol{H}^{-1} \frac{\partial f_i(\hat{\boldsymbol{\beta}}, \boldsymbol{x}_i)}{\partial \hat{\boldsymbol{\beta}}},$$

with

$$\boldsymbol{H} = \sum_{j=1}^n \left[\frac{\partial f_j(\hat{\boldsymbol{\beta}}, \boldsymbol{x}_j)}{\partial \hat{\boldsymbol{\beta}}} \left(\frac{\partial f_j(\hat{\boldsymbol{\beta}}, \boldsymbol{x}_j)}{\partial \hat{\boldsymbol{\beta}}} \right)^T - (y_j - f_j(\hat{\boldsymbol{\beta}}, \boldsymbol{x}_j)) \frac{\partial^2 f_j(\hat{\boldsymbol{\beta}}, \boldsymbol{x}_j)}{\partial \hat{\boldsymbol{\beta}}^2} \right]$$

Influence of the explanatory variable on the LSE

We can follow the same procedure to get

$$\frac{\partial \hat{\boldsymbol{\beta}}}{\partial \boldsymbol{x}_i} = \boldsymbol{H}^{-1} \left[(y_i - \hat{y}_i) \frac{\partial^2 f_i(\hat{\boldsymbol{\beta}}, \boldsymbol{x}_i)}{\partial \hat{\boldsymbol{\beta}} \partial \boldsymbol{x}_i} - \frac{\partial f_i(\hat{\boldsymbol{\beta}}, \boldsymbol{x}_i)}{\partial \hat{\boldsymbol{\beta}}} \left(\frac{\partial f_i(\hat{\boldsymbol{\beta}}, \boldsymbol{x}_i)}{\partial \boldsymbol{x}_i} \right)^T \right]$$

Influence on the predicted value

Let $\boldsymbol{f}(\hat{\boldsymbol{\beta}}) = (f_1(\hat{\boldsymbol{\beta}}, \boldsymbol{x}_1), f_2(\hat{\boldsymbol{\beta}}, \boldsymbol{x}_2), \cdots, f_n(\hat{\boldsymbol{\beta}}, \boldsymbol{x}_n))^T$ be the vector of predicted values. The influence of y_i on the predicted value is given by

$$\frac{\partial \boldsymbol{f}(\hat{\boldsymbol{\beta}})}{\partial y_i} = \frac{\partial \boldsymbol{f}(\hat{\boldsymbol{\beta}})}{\partial \hat{\boldsymbol{\beta}}} \frac{\partial \hat{\boldsymbol{\beta}}}{\partial y_i} = \boldsymbol{G}\boldsymbol{H}^{-1} \frac{\partial f_i(\hat{\boldsymbol{\beta}}, \boldsymbol{x}_i)}{\partial \hat{\boldsymbol{\beta}}}$$

where

$$\boldsymbol{G} = \frac{\partial \boldsymbol{f}(\hat{\boldsymbol{\beta}})}{\partial \hat{\boldsymbol{\beta}}}$$

This can be generalized to

$$\frac{\partial \boldsymbol{f}(\hat{\boldsymbol{\beta}})}{\partial \boldsymbol{y}} = \boldsymbol{G}\boldsymbol{H}^{-1}\boldsymbol{G}^T,$$

which is called the **Jacobian leverage**.

Logistic Regression

Consider the logistic regression model:

$$P(y_i = 1) = \frac{e^{\boldsymbol{\beta}^T \boldsymbol{x}_i}}{1 + e^{\boldsymbol{\beta}^T \boldsymbol{x}_i}}$$

For logistic regressions, we use MLE.
The log-likelihood function is given by

$$\ell(\boldsymbol{\beta}) = \sum_{i=1}^{n} \left[y_i \boldsymbol{\beta}^T \boldsymbol{x}_i - \log(1 + e^{\boldsymbol{\beta}^T \boldsymbol{x}_i}) \right]$$

The score function is given by

$$\frac{\partial \ell(\boldsymbol{\beta})}{\partial \boldsymbol{\beta}} = \sum_{i=1}^{n} y_i \boldsymbol{x}_i - \sum_{i=1}^{n} \frac{e^{\boldsymbol{\beta}^T \boldsymbol{x}_i}}{1 + e^{\boldsymbol{\beta}^T \boldsymbol{x}_i}} \boldsymbol{x}_i = \sum_{i=1}^{n} \frac{1}{1 + e^{\boldsymbol{\beta}^T \boldsymbol{x}_i}} \boldsymbol{x}_i - \sum_{i=1}^{n} (1 - y_i) \boldsymbol{x}_i$$

Influence of the covariate on the MLE

Take the partial derivative of the score function w.r.t. $\boldsymbol{x}_i$, we have

$$-(1-y_i)\boldsymbol{I} + \frac{1}{1+e^{\hat{\boldsymbol{\beta}}^T\boldsymbol{x}_i}}\boldsymbol{I} - \frac{e^{\hat{\boldsymbol{\beta}}^T\boldsymbol{x}_i}}{(1+e^{\hat{\boldsymbol{\beta}}^T\boldsymbol{x}_i})^2}\hat{\boldsymbol{\beta}}\boldsymbol{x}_i^T - \sum_j \frac{e^{\hat{\boldsymbol{\beta}}^T\boldsymbol{x}_j}}{(1+e^{\hat{\boldsymbol{\beta}}^T\boldsymbol{x}_j})^2}\frac{\partial\hat{\boldsymbol{\beta}}}{\partial\boldsymbol{x}_i}\boldsymbol{x}_j\boldsymbol{x}_j^T = 0$$

This gives the influence of the covariate on the MLE:

$$\frac{\partial\hat{\boldsymbol{\beta}}}{\partial\boldsymbol{x}_i} = \left[(y_i-\hat{p}_i)\,\boldsymbol{I} - \frac{e^{\hat{\boldsymbol{\beta}}^T\boldsymbol{x}_i}}{(1+e^{\hat{\boldsymbol{\beta}}^T\boldsymbol{x}_i})^2}\hat{\boldsymbol{\beta}}\boldsymbol{x}_i^T\right]\boldsymbol{H}^{-1},$$

where

$$\boldsymbol{H} = \sum_{j=1}^n \frac{e^{\hat{\boldsymbol{\beta}}^T\boldsymbol{x}_j}}{(1+e^{\hat{\boldsymbol{\beta}}^T\boldsymbol{x}_j})^2}\boldsymbol{x}_j\boldsymbol{x}_j^T, \quad \hat{p}_i = \frac{e^{\hat{\boldsymbol{\beta}}^T\boldsymbol{x}_i}}{1+e^{\hat{\boldsymbol{\beta}}^T\boldsymbol{x}_i}}$$

Influence on the predicted probability

The predicted probability is given by

$$\hat{p}_i = \frac{e^{\hat{\boldsymbol{\beta}}^T \boldsymbol{x}_i}}{1 + e^{\hat{\boldsymbol{\beta}}^T \boldsymbol{x}_i}}$$

The influence of $\boldsymbol{x}_i$ on the predicted probability is given by the chain rule:

$$\frac{\partial \hat{p}_i}{\partial \boldsymbol{x}_i} = \frac{\partial \hat{p}_i}{\partial \hat{\boldsymbol{\beta}}} \frac{\partial \hat{\boldsymbol{\beta}}}{\partial \boldsymbol{x}_i} + \left. \frac{\partial \hat{p}_i}{\partial \boldsymbol{x}_i} \right|_{\hat{\boldsymbol{\beta}}} = \frac{e^{\hat{\boldsymbol{\beta}}^T \boldsymbol{x}_i}}{(1 + e^{\hat{\boldsymbol{\beta}}^T \boldsymbol{x}_i})^2} \left[\left[(y_i - \hat{p}_i)\, \boldsymbol{I} - \frac{e^{\hat{\boldsymbol{\beta}}^T \boldsymbol{x}_i}}{(1 + e^{\hat{\boldsymbol{\beta}}^T \boldsymbol{x}_i})^2} \hat{\boldsymbol{\beta}} \boldsymbol{x}_i^T \right] \boldsymbol{H}^{-1} \boldsymbol{x}_i + \hat{\boldsymbol{\beta}} \right]$$

Influence of measurement error

Now we consider the assumption that the covariate $\boldsymbol{x}_i$ are fixed and known.
We consider a broader model with measurement error in the covariate such that

$$x_i = z_i + \sigma s_i,$$

where z_i is the designed covariate, s_i is the standardized measurement error, and σ is the standard deviation of the measurement error.
The model is given by

$$P(y_i = 1) = H(\boldsymbol{\gamma}^T \boldsymbol{u}_i + \tau x_i)$$

for some probability function H.
The observed model is

$$P(y_i = 1) = \mathbb{E}_s[H(\boldsymbol{\gamma}^T \boldsymbol{u}_i + \tau z_i + \tau \sigma s)]$$

The parameters is $\boldsymbol{\beta} = (\boldsymbol{\gamma}, \tau)^T$.

Influence of measurement error

The influence of the measurement error on the MLE is given by

$$\left.\frac{\partial \hat{\boldsymbol{\beta}}}{\partial \sigma^2}\right|_{\sigma^2=0} \approx -\frac{1}{\tau^2}\boldsymbol{H}^{-1}\left(\sum_{i=1}^{n}\frac{H_i''H_i'}{H_i(1-H_i)}\begin{bmatrix}\boldsymbol{u}_i\\ z_i\end{bmatrix}\right)$$

with

$$\boldsymbol{H} = \sum_{i=1}^{n}\frac{H_2'^2}{H_i(1-H_i)}\begin{bmatrix}\boldsymbol{u}_i\\ z_i\end{bmatrix}\begin{bmatrix}\boldsymbol{u}_i\\ z_i\end{bmatrix}^T$$

Proof: see textbook Chapter 9.10.

Influence analysis for the LME model

Recall the LME model:

$$\boldsymbol{y}_i = \boldsymbol{X}_i\boldsymbol{\beta} + \boldsymbol{Z}_i\boldsymbol{b}_i + \boldsymbol{\epsilon}_i$$

with $\boldsymbol{b}_i \sim N(0, \sigma^2\boldsymbol{D})$ and $\boldsymbol{\epsilon}_i \sim N(0, \sigma^2\boldsymbol{I})$.
The MLE of $\boldsymbol{\beta}$ is given by

$$\hat{\boldsymbol{\beta}} = \boldsymbol{H}^{-1}\boldsymbol{s},$$

where

$$\boldsymbol{H} = \sum_{i=1}^{N}\boldsymbol{X}_i^T\boldsymbol{V}_i^{-1}\boldsymbol{X}_i, \quad \boldsymbol{s} = \sum_{i=1}^{N}\boldsymbol{X}_i^T\boldsymbol{V}_i^{-1}\boldsymbol{y}_i, \quad \boldsymbol{V}_i = \boldsymbol{I} + \boldsymbol{Z}_i\hat{\boldsymbol{D}}\boldsymbol{Z}_i^T$$

Influence analysis for the LME model

The **leverage matrix** for the LME model is defined as

$$\boldsymbol{P}_i = \frac{\partial \hat{\boldsymbol{y}}_i}{\partial \boldsymbol{y}_i}$$

with $\hat{\boldsymbol{y}}_i = \boldsymbol{X}_i \hat{\boldsymbol{\beta}}$.
Then we have

$$\boldsymbol{P}_i = \boldsymbol{X}_i \boldsymbol{H}^{-1} \boldsymbol{X}_i^T \boldsymbol{V}_i^{-1}$$

and we can verify that

$$\mathrm{tr}(\boldsymbol{P}_i) = m.$$

Influence analysis for the LME model

The influence of the response variable on the MLE is

$$\frac{\partial \hat{\boldsymbol{\beta}}}{\partial \boldsymbol{y}_i} = \boldsymbol{H}^{-1} \boldsymbol{X}_i^T \boldsymbol{V}_i^{-1}$$

The influence of the covariate on the MLE is given by

$$\frac{\partial \hat{\boldsymbol{\beta}}}{\partial x_{ijk}} = \boldsymbol{H}^{-1} \left(\boldsymbol{E}_{ijk}^T \boldsymbol{V}_i^{-1} (\boldsymbol{y}_i - \hat{\boldsymbol{y}}_i) - \boldsymbol{X}_i^T \boldsymbol{V}_i^{-1} \boldsymbol{E}_{ijk} \hat{\boldsymbol{\beta}} \right)$$

where $\boldsymbol{E}_{ijk}$ is an $n_i \times m$ matrix with only (j, k)-th element equal to 1.

Influence analysis for the LME model

Now we consider removing the i-th cluster from the dataset. The leave-one-out estimator is given by

$$\begin{aligned}\hat{\boldsymbol{\beta}}_{(i)} &= [\boldsymbol{H} - \boldsymbol{X}_i^T \boldsymbol{V}_i^{-1} \boldsymbol{X}_i]^{-1} [\boldsymbol{s} - \boldsymbol{X}_i^T \boldsymbol{V}_i^{-1} \boldsymbol{y}_i] \\ &= \boldsymbol{\beta} - \boldsymbol{H}^{-1} \boldsymbol{X}_i^T \boldsymbol{V}_i^{-1} (\boldsymbol{I} - \boldsymbol{P}_i)^{-1} (\boldsymbol{y}_i - \hat{\boldsymbol{y}}_i)\end{aligned}$$

Therefore, we can define the generalized Cook's distance for LME as

$$D_i = \frac{1}{m\hat{\sigma}^2} (\boldsymbol{y}_i - \hat{\boldsymbol{y}}_i)^T (\boldsymbol{I} - \boldsymbol{P}_i)^{-1} \boldsymbol{V}_i^{-1} \boldsymbol{X}_i \boldsymbol{H}^{-1} \boldsymbol{X}_i^T \boldsymbol{V}_i^{-1} (\boldsymbol{I} - \boldsymbol{P}_i)^{-1} (\boldsymbol{y}_i - \hat{\boldsymbol{y}}_i)$$