Ve406 Lecture 11

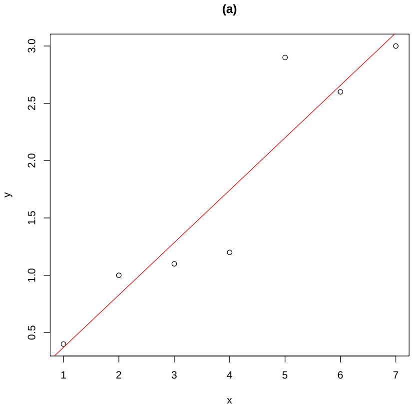
Jing Liu

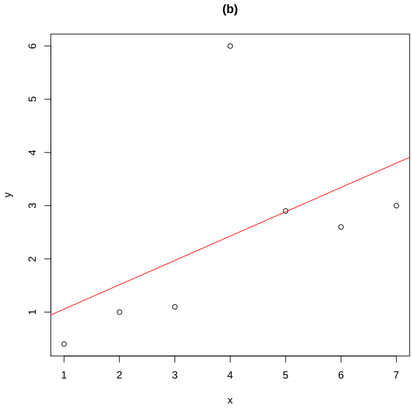
UM-SJTU Joint Institute

June 20, 2018

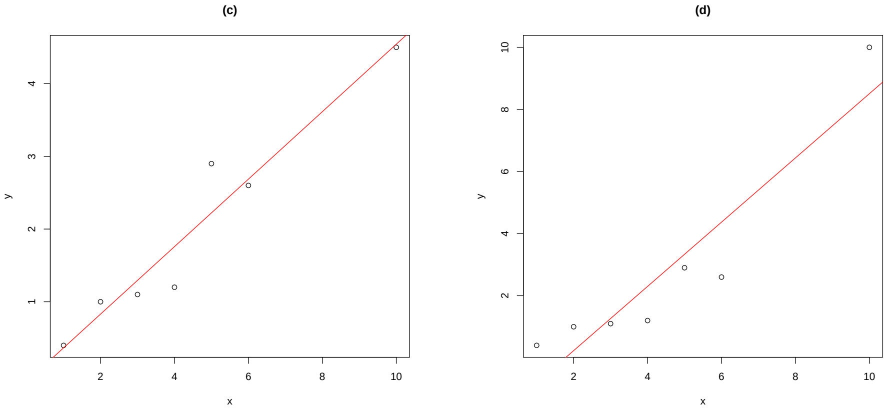
- Outliers are points with extreme response values $y_i \mid x_i$, so possibly large $\hat{e}_i$.
- High leverage points are points with extreme x_{ij}-values relative to others.

Q: Do we have any outlier or high leverage point in the following?





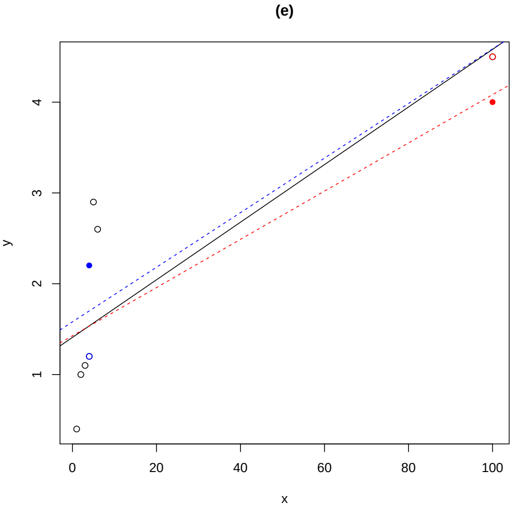
- Notice the difference between the following two cases.



- High leverage points do not necessarily have large residuals, so that it is occasionally difficult to recognise them from a residual plot.

- Outliers need to be discussed because outliers can distort the regression.

Q: Why do we care about high leverage points? Consider the following



- Having different response values for high leverage points

$$y_i$$

will alter the regression surface more in comparison to low leverage points.

- This observation leads to a common detection and quantification method.
- Notice the fitted values are always on the regression surface,

$$\hat{y}_i$$

so if they change values, regression surface will change.

- Recall we derived the following when we started multiple regression

$$\hat{\mathbf{y}} = \mathbf{X}\hat{\boldsymbol{\beta}} = \mathbf{X}\left(\mathbf{X}^{\mathrm{T}}\mathbf{X}\right)^{-1}\mathbf{X}^{\mathrm{T}}\mathbf{y} = \mathbf{P}\mathbf{y}$$

where $\mathbf{P}$ is known as the project or hat matrix.

- The leverage score is defined as the partial derivative

$$\frac{\partial \hat{y}_i}{\partial y_i} = [\mathbf{P}]_{ii} = p_{ii} \qquad \text{since} \qquad \hat{y}_i = \sum_{j=1}^{n} p_{ij} y_j$$

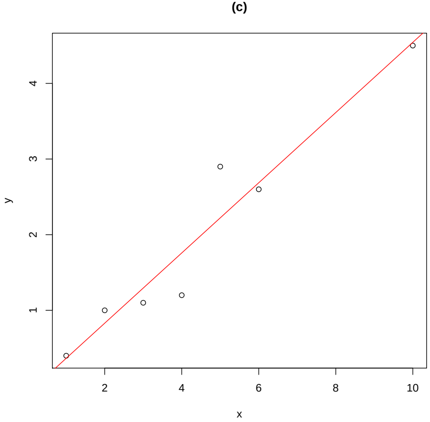
- The leverage score is entirely depended on the design matrix $\mathbf{X}$, not only $\mathbf{y}$.
- For the model having one predictor x_i and no intercept, it can be shown

$$\frac{\partial \hat{y}_i}{\partial y_i} = p_{ii} = \frac{x_i^2}{\sum_{j=1} x_j^2}$$

- When we add back the intercept, we have

$$\frac{\partial \hat{y}_i}{\partial y_i} = p_{ii} = \frac{1}{n} + \frac{(x_i - \bar{x})^2}{\sum_{j=1} (x_j - \bar{x})^2}$$

- Intuitively, the leverage score p_{ii} measures the "the amount of support" that the ith data point provided to the regression surface.
- A large leverage score for the ith point means the ith point is high leverage point, and regression surface alters significantly if y_i takes a different value.
- A high leverage point is not necessarily influential, or an influential point



- An influential point is a point whose deletion would significantly alter the regression surface. There are several detection or quantification methods:

1. Standardised difference in coefficients

$$\frac{\hat{\beta}_j - \hat{\beta}_j(-i)}{\mathrm{SE}\left(\hat{\beta}_j\right)}$$

   where $\hat{\beta}_j(-i)$ is the estimate of β_j after the ith data point has been deleted.

2. Standardised difference in fitted values

$$\frac{\hat{Y}_j - \hat{Y}_j(-i)}{\mathrm{SE}\left(\hat{Y}_j\right)}$$

- The standard errors are based on an estimate of σ without the ith data.

检验是否之前这个点，
对beta, 或者fitted value产生了影响，前两种办法

3. Cook's distance, D_i, is based on the idea of confidence ellipsoid

$$\left\{ \boldsymbol{\beta} \colon \frac{\left(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}\right)^{\mathrm{T}} \mathbf{X}^{\mathrm{T}} \mathbf{X} \left(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}\right)}{(k+1)\hat{\sigma}^2} \leq F_{k+1,n-k-1}(\alpha) \right\}$$

which gives $100 \cdot (1-\alpha)\%$ confidence ellipsoid for $\boldsymbol{\beta}$.

- The idea is to define

$$D_i = \frac{\left(\hat{\boldsymbol{\beta}} - \hat{\boldsymbol{\beta}}(-i)\right)^{\mathrm{T}} \mathbf{X}^{\mathrm{T}} \mathbf{X} \left(\hat{\boldsymbol{\beta}} - \hat{\boldsymbol{\beta}}(-i)\right)}{(k+1)\hat{\sigma}^2}$$

to quantify the whether $\hat{\boldsymbol{\beta}}(-i)$ is essentially the same as $\hat{\boldsymbol{\beta}}$ by comparing

$$D_i \qquad \text{with} \qquad F_{k+1,n-k-1}(\alpha)$$

合起来考虑，判断beta有没有意义

- Recall one reason to look at the plot of

  Residuals Vs Fitted Values

  is to check the equal variance assumption.
- However, when all assumptions are satisfied, it can shown that

$$\operatorname{Var}[\hat{e}_i \mid \mathbf{X}] = (1 - p_{ii})\sigma^2$$

- The implication is that high influence points tend to have smaller variances.
- Because of this, it is common practice to standardise the residuals
- Internally studentised residuals/standardised residual are defined by

$$\hat{e}_i' = \frac{\hat{e}_i}{\hat{\sigma}\sqrt{1 - p_{ii}}}$$

- Externally studentised residuals/studentised residual are defined by

$$\hat{e}_i^* = \frac{\hat{e}_i}{\hat{\sigma}(-i)\sqrt{1 - p_{ii}}}$$