Checking Assumptions: Residuals and Influential Observations

Merlise Clyde

STA721 Linear Models

Duke University

October 2, 2012

Outline

Topics

- Distribution of Residuals
- Leverage
- Standardized Residuals
- Cook's Distance
- Example: Stackloss data

Readings: Christensen Chapter 13

Linear Model

Linear Model:

$$\mathbf{Y} = \boldsymbol{\mu} + \boldsymbol{\epsilon}$$

Linear Model

Linear Model:

$$\mathbf{Y} = \boldsymbol{\mu} + \boldsymbol{\epsilon}$$

Assumptions:

Linear Model

Linear Model:

$$\mathbf{Y} = \boldsymbol{\mu} + \boldsymbol{\epsilon}$$

Assumptions:

$$\boldsymbol{\mu} \in C(\mathbf{X}) \quad \Leftrightarrow \quad \boldsymbol{\mu} = \mathbf{X}\beta$$

Linear Model:

$$\mathbf{Y} = \boldsymbol{\mu} + \boldsymbol{\epsilon}$$

Assumptions:

$$\begin{aligned} \boldsymbol{\mu} \in C(\mathbf{X}) \quad &\Leftrightarrow \quad \boldsymbol{\mu} = \mathbf{X}\beta \\ \boldsymbol{\epsilon} \quad &\sim \quad \mathrm{N}(\mathbf{0}_n, \sigma^2 \mathbf{I}_n) \end{aligned}$$

Linear Model:

$$\mathbf{Y} = \boldsymbol{\mu} + \boldsymbol{\epsilon}$$

Assumptions:

$$\begin{aligned} \boldsymbol{\mu} \in C(\mathbf{X}) \quad &\Leftrightarrow \quad \boldsymbol{\mu} = \mathbf{X}\beta \\ \boldsymbol{\epsilon} \quad &\sim \quad \mathrm{N}(\mathbf{0}_n, \sigma^2 \mathbf{I}_n) \end{aligned}$$

What could go wrong?

- Wrong mean

Linear Model:

$$\mathbf{Y} = \boldsymbol{\mu} + \boldsymbol{\epsilon}$$

Assumptions:

$$\begin{aligned} \boldsymbol{\mu} \in C(\mathbf{X}) & \Leftrightarrow & \boldsymbol{\mu} = \mathbf{X}\beta \\ \boldsymbol{\epsilon} & \sim & \mathsf{N}(\mathbf{0}_n, \sigma^2 \mathbf{I}_n) \end{aligned}$$

What could go wrong?

- Wrong mean
- Wrong covariance

Linear Model:

$$\mathbf{Y} = \boldsymbol{\mu} + \boldsymbol{\epsilon}$$

Assumptions:

$$\begin{aligned} \boldsymbol{\mu} \in C(\mathbf{X}) \quad &\Leftrightarrow \quad \boldsymbol{\mu} = \mathbf{X}\beta \\ \boldsymbol{\epsilon} \quad &\sim \quad \mathrm{N}(\mathbf{0}_n, \sigma^2 \mathbf{I}_n) \end{aligned}$$

What could go wrong?

- Wrong mean
- Wrong covariance
- Wrong distribution for $\boldsymbol{\epsilon}$

Residuals

Residuals (MLE of ϵ) are

$$\mathbf{e} = \mathbf{Y} - \mathbf{X}\hat{\beta}$$

Residuals

Residuals (MLE of ϵ) are

$$
\begin{aligned}
\mathbf{e} &= \mathbf{Y} - \mathbf{X}\hat{\beta} \\
&= (\mathbf{I}_n - \mathrm{P}_{\mathbf{X}})\mathbf{Y}
\end{aligned}
$$

Residuals

Residuals (MLE of ϵ) are

$$
\begin{aligned}
\mathbf{e} &= \mathbf{Y} - \mathbf{X}\hat{\beta} \\
&= (\mathbf{I}_n - \mathrm{P}_{\mathbf{X}})\mathbf{Y}
\end{aligned}
$$

$$
\mathrm{E}[\mathbf{e}] = (\mathbf{I}_n - \mathrm{P}_{\mathbf{X}})\boldsymbol{\mu}
$$

Residuals

Residuals (MLE of ϵ) are

$$\begin{aligned}\mathbf{e} &= \mathbf{Y} - \mathbf{X}\hat{\beta} \\ &= (\mathbf{I}_n - \mathsf{P}_{\mathbf{X}})\mathbf{Y}\end{aligned}$$

$$\begin{aligned}\mathsf{E}[\mathbf{e}] &= (\mathbf{I}_n - \mathsf{P}_{\mathbf{X}})\boldsymbol{\mu} \\ &= \mathbf{0}_n \text{ if } \boldsymbol{\mu} \in C(\mathbf{X})\end{aligned}$$

Residuals

Residuals (MLE of ϵ) are

$$
\begin{aligned}
\mathbf{e} &= \mathbf{Y} - \mathbf{X}\hat{\beta} \\
&= (\mathbf{I}_n - \mathrm{P}_{\mathbf{X}})\mathbf{Y}
\end{aligned}
$$

$$
\begin{aligned}
\mathrm{E}[\mathbf{e}] &= (\mathbf{I}_n - \mathrm{P}_{\mathbf{X}})\boldsymbol{\mu} \\
&= \mathbf{0}_n \text{ if } \boldsymbol{\mu} \in C(\mathbf{X})
\end{aligned}
$$

Mean will not be zero if we have left out terms

Variance Residuals

Under model with constant variances $\mathrm{Cov}[\mathbf{Y}] = \sigma^2 \mathbf{I}_n$

Under model with constant variances $\mathrm{Cov}[\mathbf{Y}] = \sigma^2 \mathbf{I}_n$

$$\mathrm{Cov} = [\mathbf{e}] \quad = \quad \mathrm{Cov}[(\mathbf{I} - \mathrm{P}_{\mathbf{X}})\mathbf{Y}$$

Variance Residuals

Under model with constant variances $\textsf{Cov}[\mathbf{Y}] = \sigma^2 \mathbf{I}_n$

$$\begin{aligned} \textsf{Cov} = [\mathbf{e}] &= \textsf{Cov}[(\mathbf{I} - \textsf{P}_{\mathbf{X}})\mathbf{Y} \\ &= (\mathbf{I} - \textsf{P}_{\mathbf{X}})^T \textsf{Cov}[\mathbf{Y}](\mathbf{I} - \textsf{P}_{\mathbf{X}}) \end{aligned}$$

Variance Residuals

Under model with constant variances $\text{Cov}[\mathbf{Y}] = \sigma^2 \mathbf{I}_n$

$$\begin{aligned}
\text{Cov} = [\mathbf{e}] &= \text{Cov}[(\mathbf{I} - \text{P}_{\mathbf{X}})\mathbf{Y} \\
&= (\mathbf{I} - \text{P}_{\mathbf{X}})^T \text{Cov}[\mathbf{Y}](\mathbf{I} - \text{P}_{\mathbf{X}}) \\
&= \sigma^2 (\mathbf{I} - \text{P}_{\mathbf{X}})^T (\mathbf{I} - \text{P}_{\mathbf{X}})
\end{aligned}$$

Variance Residuals

Under model with constant variances $\text{Cov}[\mathbf{Y}] = \sigma^2 \mathbf{I}_n$

$$\begin{aligned}
\text{Cov} = [\mathbf{e}] &= \text{Cov}[(\mathbf{I} - \mathrm{P}_{\mathbf{X}})\mathbf{Y} \\
&= (\mathbf{I} - \mathrm{P}_{\mathbf{X}})^T \text{Cov}[\mathbf{Y}](\mathbf{I} - \mathrm{P}_{\mathbf{X}}) \\
&= \sigma^2 (\mathbf{I} - \mathrm{P}_{\mathbf{X}})^T (\mathbf{I} - \mathrm{P}_{\mathbf{X}}) \\
&= \sigma^2 (\mathbf{I} - \mathrm{P}_{\mathbf{X}})
\end{aligned}$$

Variance Residuals

Under model with constant variances $\mathrm{Cov}[\mathbf{Y}] = \sigma^2 \mathbf{I}_n$

$$
\begin{aligned}
\mathrm{Cov} = [\mathbf{e}] &= \mathrm{Cov}[(\mathbf{I} - \mathrm{P}_{\mathbf{X}})\mathbf{Y} \\
&= (\mathbf{I} - \mathrm{P}_{\mathbf{X}})^T \mathrm{Cov}[\mathbf{Y}](\mathbf{I} - \mathrm{P}_{\mathbf{X}}) \\
&= \sigma^2 (\mathbf{I} - \mathrm{P}_{\mathbf{X}})^T (\mathbf{I} - \mathrm{P}_{\mathbf{X}}) \\
&= \sigma^2 (\mathbf{I} - \mathrm{P}_{\mathbf{X}})
\end{aligned}
$$

$$
\mathbf{e} \sim \mathrm{N}(\mathbf{0}, \sigma^2 (\mathbf{I} - \mathrm{P}_{\mathbf{X}}))
$$

Variance Residuals

Under model with constant variances $\mathrm{Cov}[\mathbf{Y}] = \sigma^2 \mathbf{I}_n$

$$\begin{aligned}
\mathrm{Cov} = [\mathbf{e}] &= \mathrm{Cov}[(\mathbf{I} - \mathrm{P}_{\mathbf{X}})\mathbf{Y} \\
&= (\mathbf{I} - \mathrm{P}_{\mathbf{X}})^T \mathrm{Cov}[\mathbf{Y}](\mathbf{I} - \mathrm{P}_{\mathbf{X}}) \\
&= \sigma^2 (\mathbf{I} - \mathrm{P}_{\mathbf{X}})^T (\mathbf{I} - \mathrm{P}_{\mathbf{X}}) \\
&= \sigma^2 (\mathbf{I} - \mathrm{P}_{\mathbf{X}})
\end{aligned}$$

$$\mathbf{e} \sim \mathrm{N}(\mathbf{0}, \sigma^2(\mathbf{I} - \mathrm{P}_{\mathbf{X}}))$$

- $\mathrm{P}_{\mathbf{X}}$ is the "hat" matrix (sometimes called $\mathbf{H} = [h_{ij}]$)

Variance Residuals

Under model with constant variances $\text{Cov}[\mathbf{Y}] = \sigma^2 \mathbf{I}_n$

$$\begin{aligned}
\text{Cov} = [\mathbf{e}] &= \text{Cov}[(\mathbf{I} - \text{P}_{\mathbf{X}})\mathbf{Y} \\
&= (\mathbf{I} - \text{P}_{\mathbf{X}})^T \text{Cov}[\mathbf{Y}](\mathbf{I} - \text{P}_{\mathbf{X}}) \\
&= \sigma^2 (\mathbf{I} - \text{P}_{\mathbf{X}})^T (\mathbf{I} - \text{P}_{\mathbf{X}}) \\
&= \sigma^2 (\mathbf{I} - \text{P}_{\mathbf{X}})
\end{aligned}$$

$$\mathbf{e} \sim \text{N}(\mathbf{0}, \sigma^2 (\mathbf{I} - \text{P}_{\mathbf{X}}))$$

- $\text{P}_{\mathbf{X}}$ is the "hat" matrix (sometimes called $\mathbf{H} = [h_{ij}]$)
- diagonal elements of $\text{P}_{\mathbf{X}}$ are denoted as h_{ii}

Variance Residuals

Under model with constant variances $\text{Cov}[\mathbf{Y}] = \sigma^2 \mathbf{I}_n$

$$\begin{aligned}
\text{Cov} = [\mathbf{e}] &= \text{Cov}[(\mathbf{I} - \text{P}_\mathbf{X})\mathbf{Y} \\
&= (\mathbf{I} - \text{P}_\mathbf{X})^T \text{Cov}[\mathbf{Y}](\mathbf{I} - \text{P}_\mathbf{X}) \\
&= \sigma^2 (\mathbf{I} - \text{P}_\mathbf{X})^T (\mathbf{I} - \text{P}_\mathbf{X}) \\
&= \sigma^2 (\mathbf{I} - \text{P}_\mathbf{X})
\end{aligned}$$

$$\mathbf{e} \sim \text{N}(\mathbf{0}, \sigma^2 (\mathbf{I} - \text{P}_\mathbf{X}))$$

- $\text{P}_\mathbf{X}$ is the "hat" matrix (sometimes called $\mathbf{H} = [h_{ij}]$)
- diagonal elements of $\text{P}_\mathbf{X}$ are denoted as h_{ii}
- leverage of the ith observation is h_{ii}

Variance Residuals

Under model with constant variances $\text{Cov}[\mathbf{Y}] = \sigma^2 \mathbf{I}_n$

$$\begin{aligned}
\text{Cov} = [\mathbf{e}] &= \text{Cov}[(\mathbf{I} - \text{P}_{\mathbf{X}})\mathbf{Y} \\
&= (\mathbf{I} - \text{P}_{\mathbf{X}})^T \text{Cov}[\mathbf{Y}](\mathbf{I} - \text{P}_{\mathbf{X}}) \\
&= \sigma^2 (\mathbf{I} - \text{P}_{\mathbf{X}})^T (\mathbf{I} - \text{P}_{\mathbf{X}}) \\
&= \sigma^2 (\mathbf{I} - \text{P}_{\mathbf{X}})
\end{aligned}$$

$$\mathbf{e} \sim \text{N}(\mathbf{0}, \sigma^2(\mathbf{I} - \text{P}_{\mathbf{X}}))$$

- $\text{P}_{\mathbf{X}}$ is the "hat" matrix (sometimes called $\mathbf{H} = [h_{ij}]$)
- diagonal elements of $\text{P}_{\mathbf{X}}$ are denoted as h_{ii}
- leverage of the ith observation is h_{ii}

$$\text{var}(e_i) = \sigma^2(1 - h_i)$$

Parameterization

Let $\mathbf{X} = [\mathbf{1}_n \mathbf{Z}]$ and $\mathsf{P}_{\mathbf{1}} = \mathbf{1}(\mathbf{1}^T\mathbf{1})^{-1}\mathbf{1}^T$ denote the projection on $C(\mathbf{1})$

Parameterization

Let $\mathbf{X} = [\mathbf{1}_n \mathbf{Z}]$ and $\mathsf{P}_{\mathbf{1}} = \mathbf{1}(\mathbf{1}^T\mathbf{1})^{-1}\mathbf{1}^T$ denote the projection on $C(\mathbf{1})$

$$\mathbf{Y} = \mathbf{X}\beta + \epsilon$$

Parameterization

Let $\mathbf{X} = [\mathbf{1}_n \mathbf{Z}]$ and $\mathrm{P}_{\mathbf{1}} = \mathbf{1}(\mathbf{1}^T\mathbf{1})^{-1}\mathbf{1}^T$ denote the projection on $C(\mathbf{1})$

$$
\begin{aligned}
\mathbf{Y} &= \mathbf{X}\beta + \boldsymbol{\epsilon} \\
&= \mathbf{1}_n \alpha_0 + \mathbf{Z}\boldsymbol{\alpha} + \boldsymbol{\epsilon}
\end{aligned}
$$

Parameterization

Let $\mathbf{X} = [\mathbf{1}_n \mathbf{Z}]$ and $\mathsf{P}_{\mathbf{1}} = \mathbf{1}(\mathbf{1}^T\mathbf{1})^{-1}\mathbf{1}^T$ denote the projection on $C(\mathbf{1})$

$$\begin{aligned}
\mathbf{Y} &= \mathbf{X}\beta + \boldsymbol{\epsilon} \\
&= \mathbf{1}_n\alpha_0 + \mathbf{Z}\boldsymbol{\alpha} + \boldsymbol{\epsilon} \\
&= \mathbf{1}_n\alpha_0 + \mathsf{P}_{\mathbf{1}}\mathbf{Z}\boldsymbol{\alpha} + (\mathbf{I}_n - \mathsf{P}_{\mathbf{1}})\mathbf{Z}\boldsymbol{\alpha} + \boldsymbol{\epsilon}
\end{aligned}$$

Parameterization

Let $\mathbf{X} = [\mathbf{1}_n \mathbf{Z}]$ and $\mathsf{P}_{\mathbf{1}} = \mathbf{1}(\mathbf{1}^T\mathbf{1})^{-1}\mathbf{1}^T$ denote the projection on $C(\mathbf{1})$

$$
\begin{aligned}
\mathbf{Y} &= \mathbf{X}\beta + \boldsymbol{\epsilon} \\
&= \mathbf{1}_n\alpha_0 + \mathbf{Z}\boldsymbol{\alpha} + \boldsymbol{\epsilon} \\
&= \mathbf{1}_n\alpha_0 + \mathsf{P}_{\mathbf{1}}\mathbf{Z}\boldsymbol{\alpha} + (\mathbf{I}_n - \mathsf{P}_{\mathbf{1}})\mathbf{Z}\boldsymbol{\alpha} + \boldsymbol{\epsilon}
\end{aligned}
$$

$$
= \mathbf{1}_n\alpha_0^* + (\mathbf{Z} - \mathbf{1}_n\bar{\mathbf{Z}}^T)\boldsymbol{\alpha} + \boldsymbol{\epsilon}
$$

Let z_i^T denote the vector of explanatory variables for the ith case

Mahalanobis Distance & Leverage

Let z_i^T denote the vector of explanatory variables for the ith case

$$\mathsf{P}_{\mathbf{X}} = \mathsf{P}_{\mathbf{1}} + \mathsf{P}_{\mathbf{Z}-\mathbf{1}\bar{\mathbf{z}}^T}$$

Mahalanobis Distance & Leverage

Let z_i^T denote the vector of explanatory variables for the ith case

$$
\begin{aligned}
\mathsf{P}_{\mathbf{X}} &= \mathsf{P}_{\mathbf{1}} + \mathsf{P}_{\mathbf{Z}-\mathbf{1}\bar{\mathbf{Z}}^T} \\
h_{ii} &= \frac{1}{n} + (z_i - \bar{\mathbf{Z}})^T \left((\mathbf{Z} - \mathbf{1}\bar{\mathbf{Z}}^T)^T(\mathbf{Z} - \mathbf{1}\bar{\mathbf{Z}}^T)\right)^{-1}(z_i - \bar{\mathbf{Z}})
\end{aligned}
$$

Mahalanobis Distance & Leverage

Let z_i^T denote the vector of explanatory variables for the ith case

$$
\begin{aligned}
\mathrm{P}_{\mathbf{X}} &= \mathrm{P}_{\mathbf{1}} + \mathrm{P}_{\mathbf{Z}-\mathbf{1}\bar{\mathbf{Z}}^T} \\
h_{ii} &= \frac{1}{n} + (z_i - \bar{\mathbf{Z}})^T \left((\mathbf{Z} - \mathbf{1}\bar{\mathbf{Z}}^T)^T (\mathbf{Z} - \mathbf{1}\bar{\mathbf{Z}}^T) \right)^{-1} (z_i - \bar{\mathbf{Z}}) \\
&= \frac{1}{n} + \frac{D_i^2}{n-1}
\end{aligned}
$$

Mahalanobis Distance & Leverage

Let z_i^T denote the vector of explanatory variables for the ith case

$$\begin{aligned} \mathsf{P}_{\mathbf{X}} &= \mathsf{P}_{\mathbf{1}} + \mathsf{P}_{\mathbf{Z}-\mathbf{1}\bar{\mathbf{Z}}^T} \\ h_{ii} &= \frac{1}{n} + (z_i - \bar{\mathbf{Z}})^T \left((\mathbf{Z} - \mathbf{1}\bar{\mathbf{Z}}^T)^T(\mathbf{Z} - \mathbf{1}\bar{\mathbf{Z}}^T)\right)^{-1} (z_i - \bar{\mathbf{Z}}) \\ &= \frac{1}{n} + \frac{D_i^2}{n-1} \end{aligned}$$

Leverage is a function of the estimated Mahalanobis distance D_i^2 of z_i from $\bar{\mathbf{Z}}$

Residual Plots

Under normality and constant variance assumptions, the residuals $\mathbf{e}$ and fitted values $\hat{\mathbf{Y}}$ are jointly normal but independent.

Residual Plots

Under normality and constant variance assumptions, the residuals $\mathbf{e}$ and fitted values $\hat{\mathbf{Y}}$ are jointly normal but independent. Under the correct model

Residual Plots

Under normality and constant variance assumptions, the residuals $\mathbf{e}$ and fitted values $\hat{\mathbf{Y}}$ are jointly normal but independent. Under the correct model

- $\mathsf{E}[\mathbf{e}] = \mathbf{0}_n$: scatter plots of residuals versus any combination of terms in the mean function will have a constant mean of $\mathbf{0}$

Residual Plots

Under normality and constant variance assumptions, the residuals $\mathbf{e}$ and fitted values $\hat{\mathbf{Y}}$ are jointly normal but independent. Under the correct model

- $\mathsf{E}[\mathbf{e}] = \mathbf{0}_n$: scatter plots of residuals versus any combination of terms in the mean function will have a constant mean of $\mathbf{0}$
- $\text{var}(e_i) = \sigma^2(1 - h_{ii})$, so even if fitted model is correct, the variance function is not have different variances constant. Lower variability for points with high leverage with $h_{ii} \approx 1$

Residual Plots

Under normality and constant variance assumptions, the residuals $\mathbf{e}$ and fitted values $\hat{\mathbf{Y}}$ are jointly normal but independent. Under the correct model

- $\mathsf{E}[\mathbf{e}] = \mathbf{0}_n$: scatter plots of residuals versus any combination of terms in the mean function will have a constant mean of $\mathbf{0}$
- $\mathsf{var}(e_i) = \sigma^2(1 - h_{ii})$, so even if fitted model is correct, the variance function is not have different variances constant. Lower variability for points with high leverage with $h_{ii} \approx 1$
- The residuals are correlated, but this correlation is usually not visible in residual plots

Residual Plots

Under normality and constant variance assumptions, the residuals $\mathbf{e}$ and fitted values $\hat{\mathbf{Y}}$ are jointly normal but independent. Under the correct model

- $\textsf{E}[\mathbf{e}] = \mathbf{0}_n$: scatter plots of residuals versus any combination of terms in the mean function will have a constant mean of $\mathbf{0}$
- $\textsf{var}(e_i) = \sigma^2(1 - h_{ii})$, so even if fitted model is correct, the variance function is not have different variances constant. Lower variability for points with high leverage with $h_{ii} \approx 1$
- The residuals are correlated, but this correlation is usually not visible in residual plots

When the model is not correct we hope to see structure in the residual plots to help us re-model the data

Residual Plots

Under normality and constant variance assumptions, the residuals $\mathbf{e}$ and fitted values $\hat{\mathbf{Y}}$ are jointly normal but independent. Under the correct model

- $\mathrm{E}[\mathbf{e}] = \mathbf{0}_n$: scatter plots of residuals versus any combination of terms in the mean function will have a constant mean of $\mathbf{0}$
- $\mathrm{var}(e_i) = \sigma^2(1 - h_{ii})$, so even if fitted model is correct, the variance function is not have different variances constant. Lower variability for points with high leverage with $h_{ii} \approx 1$
- The residuals are correlated, but this correlation is usually not visible in residual plots

When the model is not correct we hope to see structure in the residual plots to help us re-model the data
Use standardized residuals

Standardized Residuals

Since $\text{var}(e_i) = \sigma^2(1 - h_{ii})$ has non-constant variance

Standardized Residuals

Since $\text{var}(e_i) = \sigma^2(1 - h_{ii})$ has non-constant variance

- Divide by standard deviation:

$$\frac{e_i}{\sigma\sqrt{1 - h_{ii}}}$$

(constant variance)

Standardized Residuals

Since $\text{var}(e_i) = \sigma^2(1 - h_{ii})$ has non-constant variance

- Divide by standard deviation:

$$\frac{e_i}{\sigma\sqrt{1 - h_{ii}}}$$

  (constant variance)

- (Internally) Standardized residuals

$$r_i = \frac{e_i}{\hat{\sigma}\sqrt{1 - h_{ii}}}$$

  where $\hat{\sigma}$ is root MSE

Standardized Residuals

Since $\text{var}(e_i) = \sigma^2(1 - h_{ii})$ has non-constant variance

- Divide by standard deviation:

$$\frac{e_i}{\sigma\sqrt{1 - h_{ii}}}$$

  (constant variance)

- (Internally) Standardized residuals

$$r_i = \frac{e_i}{\hat{\sigma}\sqrt{1 - h_{ii}}}$$

  where $\hat{\sigma}$ is root MSE

Properties of Leverage

- $\sum_i h_{ij} = 1$

Properties of Leverage

- $\sum_i h_{ij} = 1$

$$\mathbf{1}^T \mathsf{P}_{\mathbf{X}} = \mathbf{1}^T \mathsf{P}_{\mathbf{1}} + \mathbf{1}^T \mathsf{P}_{\mathbf{Z} - \mathbf{1}\bar{\mathbf{Z}}^T}$$

Properties of Leverage

- $\sum_i h_{ij} = 1$

$$\begin{aligned}
\mathbf{1}^T \mathsf{P}_{\mathbf{X}} &= \mathbf{1}^T \mathsf{P}_{\mathbf{1}} + \mathbf{1}^T \mathsf{P}_{\mathbf{Z}-\mathbf{1}\bar{\mathbf{Z}}^T} \\
&= \mathbf{1}^T \mathbf{1}(\mathbf{1}^T \mathbf{1})^{-1} \mathbf{1}^T + \mathbf{0}^T
\end{aligned}$$

Properties of Leverage

- $\sum_i h_{ij} = 1$

$$
\begin{aligned}
\mathbf{1}^T \mathsf{P}_{\mathbf{X}} &= \mathbf{1}^T \mathsf{P}_{\mathbf{1}} + \mathbf{1}^T \mathsf{P}_{\mathbf{Z}-\mathbf{1}\bar{\mathbf{Z}}^T} \\
&= \mathbf{1}^T \mathbf{1}(\mathbf{1}^T \mathbf{1})^{-1} \mathbf{1}^T + \mathbf{0}^T \\
&= \mathbf{1}^T
\end{aligned}
$$

Properties of Leverage

- $\sum_i h_{ij} = 1$

$$
\begin{aligned}
\mathbf{1}^T \mathrm{P}_{\mathbf{X}} &= \mathbf{1}^T \mathrm{P}_{\mathbf{1}} + \mathbf{1}^T \mathrm{P}_{\mathbf{Z}-\mathbf{1}\bar{\mathbf{Z}}^T} \\
&= \mathbf{1}^T \mathbf{1}(\mathbf{1}^T \mathbf{1})^{-1} \mathbf{1}^T + \mathbf{0}^T \\
&= \mathbf{1}^T
\end{aligned}
$$

- $0 \leq h_{ii} \leq 1$

Properties of Leverage

- $\sum_i h_{ij} = 1$

$$\begin{aligned}
\mathbf{1}^T \mathsf{P}_{\mathbf{X}} &= \mathbf{1}^T \mathsf{P}_{\mathbf{1}} + \mathbf{1}^T \mathsf{P}_{\mathbf{Z}-\mathbf{1}\bar{\mathbf{Z}}^T} \\
&= \mathbf{1}^T \mathbf{1}(\mathbf{1}^T \mathbf{1})^{-1} \mathbf{1}^T + \mathbf{0}^T \\
&= \mathbf{1}^T
\end{aligned}$$

- $0 \leq h_{ii} \leq 1$

$$\mathsf{P}_{\mathbf{X}} = \mathsf{P}_{\mathbf{X}}^2$$

Properties of Leverage

- $\sum_i h_{ij} = 1$

$$\begin{aligned}
\mathbf{1}^T \mathsf{P}_{\mathbf{X}} &= \mathbf{1}^T \mathsf{P}_{\mathbf{1}} + \mathbf{1}^T \mathsf{P}_{\mathbf{Z}-\mathbf{1}\bar{\mathbf{Z}}^T} \\
&= \mathbf{1}^T \mathbf{1}(\mathbf{1}^T\mathbf{1})^{-1}\mathbf{1}^T + \mathbf{0}^T \\
&= \mathbf{1}^T
\end{aligned}$$

- $0 \leq h_{ii} \leq 1$

$$\begin{aligned}
\mathsf{P}_{\mathbf{X}} &= \mathsf{P}_{\mathbf{X}}^2 \\
h_{ii} &= \sum_j h_{ij} h_{ji}
\end{aligned}$$

Properties of Leverage

- $\sum_i h_{ij} = 1$

$$\begin{aligned} \mathbf{1}^T \mathsf{P}_{\mathbf{X}} &= \mathbf{1}^T \mathsf{P}_{\mathbf{1}} + \mathbf{1}^T \mathsf{P}_{\mathbf{Z}-\mathbf{1}\bar{\mathbf{Z}}^T} \\ &= \mathbf{1}^T \mathbf{1}(\mathbf{1}^T\mathbf{1})^{-1}\mathbf{1}^T + \mathbf{0}^T \\ &= \mathbf{1}^T \end{aligned}$$

- $0 \le h_{ii} \le 1$

$$\begin{aligned} \mathsf{P}_{\mathbf{X}} &= \mathsf{P}_{\mathbf{X}}^2 \\ h_{ii} &= \sum_j h_{ij} h_{ji} \\ &= \sum_j h_{jj}^2 \end{aligned}$$

Properties of Leverage

- $\sum_i h_{ij} = 1$

$$
\begin{aligned}
\mathbf{1}^T \mathrm{P}_{\mathbf{X}} &= \mathbf{1}^T \mathrm{P}_{\mathbf{1}} + \mathbf{1}^T \mathrm{P}_{\mathbf{Z}-\mathbf{1}\bar{\mathbf{Z}}^T} \\
&= \mathbf{1}^T \mathbf{1}(\mathbf{1}^T \mathbf{1})^{-1} \mathbf{1}^T + \mathbf{0}^T \\
&= \mathbf{1}^T
\end{aligned}
$$

- $0 \leq h_{ii} \leq 1$

$$
\begin{aligned}
\mathrm{P}_{\mathbf{X}} &= \mathrm{P}_{\mathbf{X}}^2 \\
h_{ii} &= \sum_j h_{ij} h_{ji} \\
&= \sum_j h_{ij}^2 \\
&= h_{ii}^2 + \sum_{j \neq i} h_{ij}^2
\end{aligned}
$$

Properties of Leverage

- $\sum_i h_{ij} = 1$

$$\begin{aligned}
\mathbf{1}^T \mathsf{P}_{\mathbf{X}} &= \mathbf{1}^T \mathsf{P}_{\mathbf{1}} + \mathbf{1}^T \mathsf{P}_{\mathbf{Z}-\mathbf{1}\bar{\mathbf{Z}}^T} \\
&= \mathbf{1}^T \mathbf{1}(\mathbf{1}^T \mathbf{1})^{-1} \mathbf{1}^T + \mathbf{0}^T \\
&= \mathbf{1}^T
\end{aligned}$$

- $0 \leq h_{ii} \leq 1$

$$\begin{aligned}
\mathsf{P}_{\mathbf{X}} &= \mathsf{P}_{\mathbf{X}}^2 \\
h_{ii} &= \sum_j h_{ij} h_{ji} \\
&= \sum_j h_{jj}^2 \\
&= h_{ii}^2 + \sum_{j \neq i} h_{jj}^2 \\
h_{ii}(1 - h_{ii}) &\geq 0
\end{aligned}$$

Predictions

$$\hat{Y}_i = h_{ii} Y_i + \sum_{j \neq i} h_{ij} Y_j$$

Predictions

$$\hat{Y}_i = h_{ii} Y_i + \sum_{j \neq i} h_{ij} Y_j$$

If $h_{ii} \approx 1$ then $\hat{Y}_i \approx Y_i$

Predictions

$$\hat{Y}_i = h_{ii} Y_i + \sum_{j \neq i} h_{ij} Y_j$$

If $h_{ii} \approx 1$ then $\hat{Y}_i \approx Y_i$

- $\sum_j h_{ij} = 1$

Predictions

$$\hat{Y}_i = h_{ii} Y_i + \sum_{j \neq i} h_{ij} Y_j$$

If $h_{ii} \approx 1$ then $\hat{Y}_i \approx Y_i$

- $\sum_j h_{ij} = 1$
- $\sum_i h_{ij}^2 = 1$

Predictions

$$\hat{Y}_i = h_{ii} Y_i + \sum_{j \neq i} h_{ij} Y_j$$

If $h_{ii} \approx 1$ then $\hat{Y}_i \approx Y_i$

- $\sum_j h_{ij} = 1$
- $\sum_i h_{ij}^2 = 1$

$$\mathbf{1}^T \mathsf{P}_{\mathbf{X}} \quad = \quad \mathbf{1}^T (\mathsf{P}_{\mathbf{1}} + \mathsf{P}_{\mathbf{X}})(\mathsf{P}_{\mathbf{1}} + \mathsf{P}_{\mathbf{X}})^T$$

Predictions

$$\hat{Y}_i = h_{ii} Y_i + \sum_{j \neq i} h_{ij} Y_j$$

If $h_{ii} \approx 1$ then $\hat{Y}_i \approx Y_i$

- $\sum_j h_{ij} = 1$
- $\sum_i h_{ij}^2 = 1$

$$\begin{aligned} \mathbf{1}^T \mathsf{P}_{\mathbf{X}} &= \mathbf{1}^T (\mathsf{P}_{\mathbf{1}} + \mathsf{P}_{\mathbf{X}})(\mathsf{P}_{\mathbf{1}} + \mathsf{P}_{\mathbf{X}})^T \\ &= \mathbf{1}^T \mathbf{1}(\mathbf{1}^T \mathbf{1})^{-1} \mathbf{1}^T + \mathbf{0}^T \end{aligned}$$

Predictions

$$\hat{Y}_i = h_{ii} Y_i + \sum_{j \neq i} h_{ij} Y_j$$

If $h_{ii} \approx 1$ then $\hat{Y}_i \approx Y_i$

- $\sum_j h_{ij} = 1$
- $\sum_i h_{ij}^2 = 1$

$$\begin{aligned} \mathbf{1}^T \mathsf{P}_{\mathbf{X}} &= \mathbf{1}^T (\mathsf{P}_{\mathbf{1}} + \mathsf{P}_{\mathbf{X}})(\mathsf{P}_{\mathbf{1}} + \mathsf{P}_{\mathbf{X}})^T \\ &= \mathbf{1}^T \mathbf{1}(\mathbf{1}^T \mathbf{1})^{-1} \mathbf{1}^T + \mathbf{0}^T \\ &= \mathbf{1}^T \end{aligned}$$

Predictions

$$\hat{Y}_i = h_{ii} Y_i + \sum_{j \neq i} h_{ij} Y_j$$

If $h_{ii} \approx 1$ then $\hat{Y}_i \approx Y_i$

- $\sum_j h_{ij} = 1$
- $\sum_i h_{ij}^2 = 1$

$$\begin{aligned} \mathbf{1}^T \mathsf{P}_{\mathbf{X}} &= \mathbf{1}^T (\mathsf{P}_{\mathbf{1}} + \mathsf{P}_{\mathbf{X}})(\mathsf{P}_{\mathbf{1}} + \mathsf{P}_{\mathbf{X}})^T \\ &= \mathbf{1}^T \mathbf{1}(\mathbf{1}^T \mathbf{1})^{-1} \mathbf{1}^T + \mathbf{0}^T \\ &= \mathbf{1}^T \end{aligned}$$

- $h_{ij} \approx 0$ for $i \neq j$ if $h_{ii} \approx 1$

Predictions

$$\hat{Y}_i = h_{ii} Y_i + \sum_{j \neq i} h_{ij} Y_j$$

If $h_{ii} \approx 1$ then $\hat{Y}_i \approx Y_i$

- $\sum_j h_{ij} = 1$
- $\sum_i h_{ij}^2 = 1$

$$\begin{aligned} \mathbf{1}^T \mathsf{P}_{\mathbf{X}} &= \mathbf{1}^T (\mathsf{P}_{\mathbf{1}} + \mathsf{P}_{\mathbf{X}})(\mathsf{P}_{\mathbf{1}} + \mathsf{P}_{\mathbf{X}})^T \\ &= \mathbf{1}^T \mathbf{1} (\mathbf{1}^T \mathbf{1})^{-1} \mathbf{1}^T + \mathbf{0}^T \\ &= \mathbf{1}^T \end{aligned}$$

- $h_{ij} \approx 0$ for $i \neq j$ if $h_{ii} \approx 1$
- $\text{var}(e_i) = \sigma^2 (1 - h_{ii}) \approx 0$ if $h_{ii} \approx 1$

Illustration

Predicted Residuals

Let $\mathbf{X}_{(i)}$ and $\mathbf{Y}_{(i)}$ denote the design matrix and response vector with the ith row $\mathbf{x}_i^T$ deleted

Predicted Residuals

Let $\mathbf{X}_{(i)}$ and $\mathbf{Y}_{(i)}$ denote the design matrix and response vector with the ith row $\mathbf{x}_i^T$ deleted

- $\hat{\beta}_{(i)} = (\mathbf{X}_{(i)}^T \mathbf{X}_{(i)})^{-1} \mathbf{X}_{(i)}^T \mathbf{Y}_{(i)}$

Predicted Residuals

Let $\mathbf{X}_{(i)}$ and $\mathbf{Y}_{(i)}$ denote the design matrix and response vector with the ith row $\mathbf{x}_i^T$ deleted

- $\hat{\beta}_{(i)} = (\mathbf{X}_{(i)}^T \mathbf{X}_{(i)})^{-1} \mathbf{X}_{(i)}^T \mathbf{Y}_{(i)}$
- Predicted residual $\mathbf{e}_{(i)} = y_i - \mathbf{x}_{(i)}^T \hat{\beta}_{(i)}$

Predicted Residuals

Let $\mathbf{X}_{(i)}$ and $\mathbf{Y}_{(i)}$ denote the design matrix and response vector with the ith row $\mathbf{x}_i^T$ deleted

- $\hat{\beta}_{(i)} = (\mathbf{X}_{(i)}^T \mathbf{X}_{(i)})^{-1} \mathbf{X}_{(i)}^T \mathbf{Y}_{(i)}$
- Predicted residual $\mathbf{e}_{(i)} = y_i - \mathbf{x}_{(i)}^T \hat{\beta}_{(i)}$
- Variance $\mathrm{var}(e_{(i)}) = \sigma^2 (1 + \mathbf{x}_{(i)}^T (\mathbf{X}_{(i)}^T \mathbf{X}_{(i)})^{-1} \mathbf{x}_{(i)})$

Predicted Residuals

Let $\mathbf{X}_{(i)}$ and $\mathbf{Y}_{(i)}$ denote the design matrix and response vector with the ith row $\mathbf{x}_i^T$ deleted

- $\hat{\beta}_{(i)} = (\mathbf{X}_{(i)}^T \mathbf{X}_{(i)})^{-1} \mathbf{X}_{(i)}^T \mathbf{Y}_{(i)}$
- Predicted residual $\mathbf{e}_{(i)} = y_i - \mathbf{x}_{(i)}^T \hat{\beta}_{(i)}$
- Variance $\mathrm{var}(e_{(i)}) = \sigma^2 (1 + \mathbf{x}_{(i)}^T (\mathbf{X}_{(i)}^T \mathbf{X}_{(i)})^{-1} \mathbf{x}_{(i)})$

Updating Formula

How to compute without re-fitting model for each case deletion?

Updating Formula

How to compute without re-fitting model for each case deletion?

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

Updating Formula

How to compute without re-fitting model for each case deletion?

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

Special case of the Binomial Inverse Theorem or Woodbury Theorem

Updating Formula

How to compute without re-fitting model for each case deletion?

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

Special case of the Binomial Inverse Theorem or Woodbury Theorem

$$\hat{\beta}_{(i)} = (\mathbf{X}_{(i)}^T\mathbf{X}_{(i)})^{-1}\mathbf{X}_{(i)}^T\mathbf{Y}_{(i)}$$

Updating Formula

How to compute without re-fitting model for each case deletion?

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

Special case of the Binomial Inverse Theorem or Woodbury Theorem

$$\begin{aligned}\hat{\beta}_{(i)} &= (\mathbf{X}_{(i)}^T\mathbf{X}_{(i)})^{-1}\mathbf{X}_{(i)}^T\mathbf{Y}_{(i)} \\ &= (\mathbf{X}_{(i)}^T\mathbf{X}_{(i)})^{-1}(\mathbf{X}^T\mathbf{Y} - \mathbf{x}_i y_i)\end{aligned}$$

Updating Formula

How to compute without re-fitting model for each case deletion?

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

Special case of the Binomial Inverse Theorem or Woodbury Theorem

$$\begin{aligned} \hat{\beta}_{(i)} &= (\mathbf{X}_{(i)}^T\mathbf{X}_{(i)})^{-1}\mathbf{X}_{(i)}^T\mathbf{Y}_{(i)} \\ &= (\mathbf{X}_{(i)}^T\mathbf{X}_{(i)})^{-1}(\mathbf{X}^T\mathbf{Y} - \mathbf{x}_i y_i) \\ &\quad \text{some algebra} \end{aligned}$$

Updating Formula

How to compute without re-fitting model for each case deletion?

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

Special case of the Binomial Inverse Theorem or Woodbury Theorem

$$\begin{aligned}
\hat{\beta}_{(i)} &= (\mathbf{X}_{(i)}^T\mathbf{X}_{(i)})^{-1}\mathbf{X}_{(i)}^T\mathbf{Y}_{(i)} \\
&= (\mathbf{X}_{(i)}^T\mathbf{X}_{(i)})^{-1}(\mathbf{X}^T\mathbf{Y} - \mathbf{x}_i y_i) \\
&\quad \text{some algebra} \\
&= \hat{\beta} + \frac{\mathbf{X}^T\mathbf{X}\mathbf{x}_i e_i}{1 - h_{ii}}
\end{aligned}$$

Updating Formula

How to compute without re-fitting model for each case deletion?

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

Special case of the Binomial Inverse Theorem or Woodbury Theorem

$$\begin{aligned}
\hat{\beta}_{(i)} &= (\mathbf{X}_{(i)}^T\mathbf{X}_{(i)})^{-1}\mathbf{X}_{(i)}^T\mathbf{Y}_{(i)} \\
&= (\mathbf{X}_{(i)}^T\mathbf{X}_{(i)})^{-1}(\mathbf{X}^T\mathbf{Y} - \mathbf{x}_i y_i) \\
&\quad \text{some algebra} \\
&= \hat{\beta} + \frac{\mathbf{X}^T\mathbf{X}\mathbf{x}_i e_i}{1 - h_{ii}}
\end{aligned}$$

Predicted Residuals

Predicted residual

$$e_{(i)} = y_i - \mathbf{x}_i^T \hat{\beta}(i) = \frac{e_i}{1 - h_{ii}}$$

Predicted Residuals

Predicted residual

$$e_{(i)} = y_i - \mathbf{x}_i^T \hat{\beta}(i) = \frac{e_i}{1 - h_{ii}}$$

with variance

$$\text{var}(e_{(i)}) = \frac{\sigma^2}{1 - h_{ii}}$$

Predicted Residuals

Predicted residual

$$e_{(i)} = y_i - \mathbf{x}_i^T \hat{\beta}(i) = \frac{e_i}{1 - h_{ii}}$$

with variance

$$\text{var}(e_{(i)}) = \frac{\sigma^2}{1 - h_{ii}}$$

Standardized predicted residual is

$$\frac{e_{(i)}}{\sqrt{\text{var}(e_{(i)})}} = \frac{e_i/(1 - h_{ii})}{\sigma/\sqrt{1 - h_{ii}}} = \frac{e_i}{\sigma\sqrt{1 - h_{ii}}}$$

Predicted Residuals

Predicted residual

$$e_{(i)} = y_i - \mathbf{x}_i^T \hat{\beta}(i) = \frac{e_i}{1 - h_{ii}}$$

with variance

$$\text{var}(e_{(i)}) = \frac{\sigma^2}{1 - h_{ii}}$$

Standardized predicted residual is

$$\frac{e_{(i)}}{\sqrt{\text{var}(e_{(i)})}} = \frac{e_i/(1 - h_{ii})}{\sigma/\sqrt{1 - h_{ii}}} = \frac{e_i}{\sigma\sqrt{1 - h_{ii}}}$$

Same as before!

External estimate of σ^2

Estimate $\hat{\sigma}^2_{(i)}$ using data with case i deleted

External estimate of σ^2

Estimate $\hat{\sigma}^2_{(i)}$ using data with case i deleted

$$\text{SSE}_{(i)} = \text{SSE} - \frac{e_i^2}{1 - hii}$$

External estimate of σ^2

Estimate $\hat{\sigma}^2_{(i)}$ using data with case i deleted

$$
\begin{aligned}
\mathrm{SSE}_{(i)} &= \mathrm{SSE} - \frac{e_i^2}{1 - hii} \\
\hat{\sigma}^2_{(i)} = \mathrm{MSE}_{(i)} &= \frac{\mathrm{SSE}_{(i)}}{n - p - 1}
\end{aligned}
$$

External estimate of σ^2

Estimate $\hat{\sigma}^2_{(i)}$ using data with case i deleted

$$\begin{aligned} \text{SSE}_{(i)} &= \text{SSE} - \frac{e_i^2}{1 - hii} \\ \hat{\sigma}^2_{(i)} = \text{MSE}_{(i)} &= \frac{\text{SSE}_{(i)}}{n - p - 1} \end{aligned}$$

Externally Standardized residuals

External estimate of σ^2

Estimate $\hat{\sigma}^2_{(i)}$ using data with case i deleted

$$\begin{aligned} \mathrm{SSE}_{(i)} &= \mathrm{SSE} - \frac{e_i^2}{1 - hii} \\ \hat{\sigma}^2_{(i)} = \mathrm{MSE}_{(i)} &= \frac{\mathrm{SSE}_{(i)}}{n - p - 1} \end{aligned}$$

Externally Standardized residuals

$$t_i = \frac{e_{(i)}}{\sqrt{\hat{\sigma}^2_{(i)}/(1 - h_{ii})}} = \frac{y_i - \mathbf{x}_i^T \hat{\beta}_{(i)}}{\sqrt{\hat{\sigma}^2_{(i)}/(1 - h_{ii})}}$$

External estimate of σ^2

Estimate $\hat{\sigma}^2_{(i)}$ using data with case i deleted

$$\begin{aligned} \text{SSE}_{(i)} &= \text{SSE} - \frac{e_i^2}{1 - hii} \\ \hat{\sigma}^2_{(i)} = \text{MSE}_{(i)} &= \frac{\text{SSE}_{(i)}}{n - p - 1} \end{aligned}$$

Externally Standardized residuals

$$t_i = \frac{e_{(i)}}{\sqrt{\hat{\sigma}^2_{(i)}/(1 - h_{ii})}} = \frac{y_i - \mathbf{x}_i^T \hat{\beta}_{(i)}}{\sqrt{\hat{\sigma}^2_{(i)}/(1 - h_{ii})}}$$

May still miss extreme points with high leverage, but will pick up unusual y_is

Outlier Test

H_0: $\mu_i = \mathbf{x}_i^T \boldsymbol{\beta}$ versus H_a: $\mu_i = \mathbf{x}_i^T \boldsymbol{\beta} + \alpha_i$

Outlier Test

H_0: $\mu_i = \mathbf{x}_i^T \boldsymbol{\beta}$ versus H_a: $\mu_i = \mathbf{x}_i^T \boldsymbol{\beta} + \alpha_i$

- Show that t-test for testing H_0: $\alpha_i = 0$ is equal to t_i (HW)

Outlier Test

H$_0$: $\mu_i = \mathbf{x}_i^T \boldsymbol{\beta}$ versus H$_a$: $\mu_i = \mathbf{x}_i^T \boldsymbol{\beta} + \alpha_i$

- Show that t-test for testing H$_0$: $\alpha_i = 0$ is equal to t_i (HW)
- if p-value is small declare the ith case to be an outlier: E$[Y_i]$ not given by $\mathbf{X}\beta$ but $\mathbf{X}\boldsymbol{\beta} = \delta_i \alpha_i$

Outlier Test

H$_0$: $\mu_i = \mathbf{x}_i^T \boldsymbol{\beta}$ versus H$_a$: $\mu_i = \mathbf{x}_i^T \boldsymbol{\beta} + \alpha_i$

- Show that t-test for testing H$_0$: $\alpha_i = 0$ is equal to t_i (HW)
- if p-value is small declare the ith case to be an outlier: E$[Y_i]$ not given by $\mathbf{X}\beta$ but $\mathbf{X}\beta = \delta_i \alpha_i$
- control for multiple testing

Outlier Test

H_0: $\mu_i = \mathbf{x}_i^T \boldsymbol{\beta}$ versus H_a: $\mu_i = \mathbf{x}_i^T \boldsymbol{\beta} + \alpha_i$

- Show that t-test for testing H_0: $\alpha_i = 0$ is equal to t_i (HW)
- if p-value is small declare the ith case to be an outlier: $\mathrm{E}[Y_i]$ not given by $\mathbf{X}\beta$ but $\mathbf{X}\beta = \delta_i \alpha_i$
- control for multiple testing
- Extreme case $\boldsymbol{\mu} = \mathbf{X}\beta + \mathbf{I}_n \boldsymbol{\alpha}$ all points have their own mean!

Outlier Test

H$_0$: $\mu_i = \mathbf{x}_i^T \boldsymbol{\beta}$ versus H$_a$: $\mu_i = \mathbf{x}_i^T \boldsymbol{\beta} + \alpha_i$

- Show that t-test for testing H$_0$: $\alpha_i = 0$ is equal to t_i (HW)
- if p-value is small declare the ith case to be an outlier: $\mathsf{E}[Y_i]$ not given by $\mathbf{X}\beta$ but $\mathbf{X}\beta = \delta_i \alpha_i$
- control for multiple testing
- Extreme case $\boldsymbol{\mu} = \mathbf{X}\beta + \mathbf{I}_n \boldsymbol{\alpha}$ all points have their own mean!

Cook's Distance

Cook's Distance measure of how much predictions change with ith case deleted

Cook's Distance measure of how much predictions change with ith case deleted

$$D_i = \frac{\|\mathbf{Y}_{(i)} - \mathbf{Y}\|^2}{p\hat{\sigma}^2} = \frac{(\hat{\beta}_{(i)} - \hat{\beta})^T \mathbf{X}^T \mathbf{X} (\hat{\beta}_{(i)} - \hat{\beta})}{p\hat{\sigma}^2}$$

Cook's Distance

Cook's Distance measure of how much predictions change with ith case deleted

$$
\begin{aligned}
D_i &= \frac{\|\mathbf{Y}_{(i)} - \mathbf{Y}\|^2}{p\hat{\sigma}^2} = \frac{(\hat{\beta}_{(i)} - \hat{\beta})^T \mathbf{X}^T \mathbf{X} (\hat{\beta}_{(i)} - \hat{\beta})}{p\hat{\sigma}^2} \\
&= \frac{r_i^2}{p} \frac{h_{ii}}{1 - h_{ii}}
\end{aligned}
$$

Cook's Distance

Cook's Distance measure of how much predictions change with ith case deleted

$$
\begin{aligned}
D_i &= \frac{\|\mathbf{Y}_{(i)} - \mathbf{Y}\|^2}{p\hat{\sigma}^2} = \frac{(\hat{\beta}_{(i)} - \hat{\beta})^T \mathbf{X}^T \mathbf{X} (\hat{\beta}_{(i)} - \hat{\beta})}{p\hat{\sigma}^2} \\
&= \frac{r_i^2}{p} \frac{h_{ii}}{1 - h_{ii}}
\end{aligned}
$$

Flag cases where $D_i > 1$ or large relative to other cases

Cook's Distance

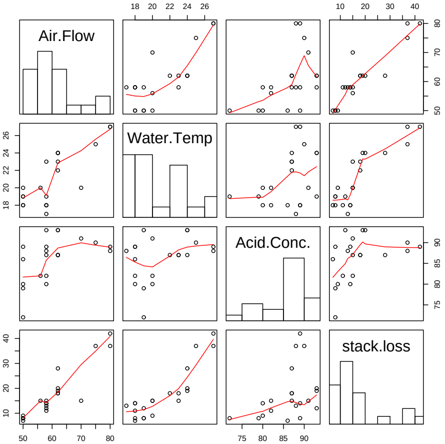
Cook's Distance measure of how much predictions change with ith case deleted

$$
\begin{aligned}
D_i &= \frac{\|\mathbf{Y}_{(i)} - \mathbf{Y}\|^2}{p\hat{\sigma}^2} = \frac{(\hat{\beta}_{(i)} - \hat{\beta})^T \mathbf{X}^T \mathbf{X} (\hat{\beta}_{(i)} - \hat{\beta})}{p\hat{\sigma}^2} \\
&= \frac{r_i^2}{p} \frac{h_{ii}}{1 - h_{ii}}
\end{aligned}
$$

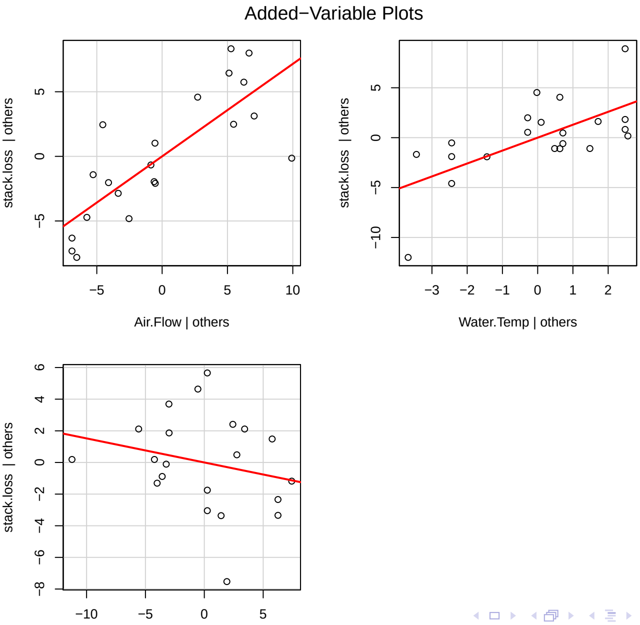
Flag cases where $D_i > 1$ or large relative to other cases Influential

Cases

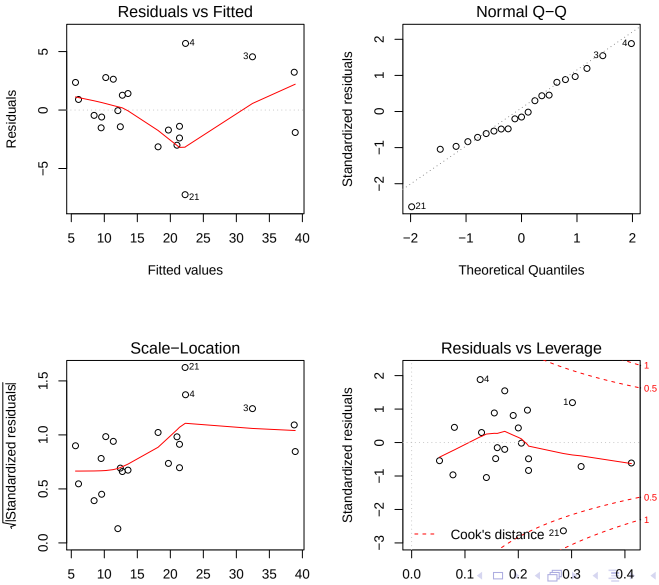
Stackloss Data



Stackloss Added Variable Plot



Stackloss Data



Case 21

- Leverage 0.285
- Cooks'd Distance .69
- p-value t_{21} is 0.0042
- Bonferonni adjusted p-value is 0.0024
- Other points? Masking?
- Refit without Case 21

Others have suggested that cases (1, 3, 4, 21) are outliers

To Remove or Not?

- For suspicious cases, check data sources

To Remove or Not?

- For suspicious cases, check data sources
- Check that points are not outliers because of wrong mean function or distributional assumptions

To Remove or Not?

- For suspicious cases, check data sources
- Check that points are not outliers because of wrong mean function or distributional assumptions
- Influential cases - report results with and without cases (resutls may change)

To Remove or Not?

- For suspicious cases, check data sources
- Check that points are not outliers because of wrong mean function or distributional assumptions
- Influential cases - report results with and without cases (resutls may change)
- Outlier test - suggests alternative population; if not influential may in keep analysis, but will inflate $\hat{\sigma}^2$ and interval estimates

To Remove or Not?

- For suspicious cases, check data sources
- Check that points are not outliers because of wrong mean function or distributional assumptions
- Influential cases - report results with and without cases (resutls may change)
- Outlier test - suggests alternative population; if not influential may in keep analysis, but will inflate $\hat{\sigma}^2$ and interval estimates
- Document steps - reproducibility!
- Robust Regression Methods

To Remove or Not?

- For suspicious cases, check data sources
- Check that points are not outliers because of wrong mean function or distributional assumptions
- Influential cases - report results with and without cases (resutls may change)
- Outlier test - suggests alternative population; if not influential may in keep analysis, but will inflate $\hat{\sigma}^2$ and interval estimates
- Document steps - reproducibility!
- Robust Regression Methods

Are there outliers in the Stackloss Data?