Lecture 13

Model Diagnostics

STAT 8020 Statistical Methods II September 18, 2019



Leverage Values

Studentized &
Studentized Deleted
Residuals

DFFITS

Variance Inflation Factor

Non-Constant Variance &

Whitney Huang Clemson University

Agenda

- Model Diagnostics

 CLEMS
- Leverage Values
- Studentized &
 Studentized Deleted
 Residuals
- FFITS
- Variance Inflation Factor
- Non-Constant /ariance &

- Leverage Values
- Studentized & Studentized Deleted Residuals
- **3 DFFITS**
- **4** Variance Inflation Factor
- 5 Non-Constant Variance & Transformation

Leverage

Recall in MLR that $\hat{Y} = X(X^TX)^{-1}X^TY = HY$ where H is the hat-matrix

Model Diagnostics



Leverage Values

Studentized & Studentized Deleted Residuals

DFFITS

Variance Inflation Factor

Leverage

Recall in MLR that $\hat{Y} = X(X^TX)^{-1}X^TY = HY$ where H is the hat-matrix

• The leverage value for the i_{th} observation is defined as:

$$h_i = \boldsymbol{H}_{ii}$$

Model Diagnostics



Leverage Values

Studentized & Studentized Deleted Residuals

DFFITS

ariance Inflation actor

Leverage

Recall in MLR that $\hat{Y} = X(X^TX)^{-1}X^TY = HY$ where H is the hat-matrix

• The leverage value for the i_{th} observation is defined as:

$$h_i = \boldsymbol{H}_{ii}$$

Model Diagnostics



Leverage Values

Studentized & Studentized Deleted Residuals

DFFITS

ariance Inflation actor

Model Diagnostics CLEMS N V E R S I T Y

Leverage Values

Studentized &
Studentized Deleted
Residuals

DFFITS

/ariance Inflation

Non-Constant Variance & Transformation

Recall in MLR that $\hat{Y} = X(X^TX)^{-1}X^TY = HY$ where H is the hat-matrix

• The leverage value for the i_{th} observation is defined as:

$$h_i = \boldsymbol{H}_{ii}$$

• Can show that $\text{Var}(e_i) = \sigma^2(1-h_i)$, where $e_i = Y_i - \hat{Y}_i$ is the residual for the i_{th} observation

Model Diagnostics CLEMS N V E R S I T Y

Leverage Values

Studentized &
Studentized Deleted
Residuals

DFFITS

/ariance Inflation

Non-Constant Variance & Transformation

Recall in MLR that $\hat{Y} = X(X^TX)^{-1}X^TY = HY$ where H is the hat-matrix

• The leverage value for the i_{th} observation is defined as:

$$h_i = \boldsymbol{H}_{ii}$$

• Can show that $\text{Var}(e_i) = \sigma^2(1-h_i)$, where $e_i = Y_i - \hat{Y}_i$ is the residual for the i_{th} observation

Model Diagnostics

CLEMS#N UNIVERSITY

Leverage Values

Studentized &
Studentized Deleted
Residuals

DFFITS

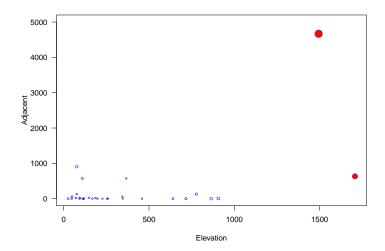
/ariance Inflation Factor

- Recall in MLR that $\hat{Y} = X(X^TX)^{-1}X^TY = HY$ where H is the hat-matrix
 - The leverage value for the i_{th} observation is defined as:

$$h_i = \boldsymbol{H}_{ii}$$

- Can show that $\text{Var}(e_i) = \sigma^2(1 h_i)$, where $e_i = Y_i \hat{Y}_i$ is the residual for the i_{th} observation
- $\frac{1}{n} \leq h_i \leq 1$, $1 \leq i \leq n$ and $\bar{h} = \sum_{i=1}^n \frac{h_i}{n} = \frac{p}{n} \Rightarrow$ a "rule of thumb" is that leverages of more than $\frac{2p}{n}$ should be looked at more closely

Leverage Values of Species \sim Elev + Adj



Model Diagnostics



Leverage Value

Studentized &
Studentized Deleted
Residuals

FFITS

/ariance Inflation

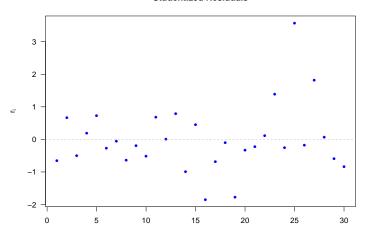
Variance Inflation

Non-Constant /ariance &

- As we have seen ${
 m Var}(e_i)=\sigma^2(1-h_i)$, this suggests the use of $r_i=rac{e_i}{\hat{\sigma}\sqrt{(1-h_i)}}$
 - r_i 's are called **studentized residuals**. r_i 's are sometimes preferred in residual plots as they have been standardized to have equal variance.
 - If the model assumptions are correct then ${\rm Var}(r_i)=1$ and ${\rm Corr}(e_i,e_j)$ tends to be small

Studentized Residuals of Species \sim Elev + Adj

Studentized Residuals



Model Diagnostics



Leverage Values

Studentized &
Studentized Deleted
Residuals

OFFITS

Variance Inflation Factor

on-Constant ariance &

Studentized Deleted Residuals

• For a given model, exclude the observation i and recompute $\hat{\beta}_{(i)}$, $\hat{\sigma}_{(i)}$ to obtain $\hat{Y}_{i(i)}$

Model Diagnostics



Leverage Values

Studentized &
Studentized Deleted
Residuals

DFFITS

Variance Inflation

Studentized Deleted Residuals

• For a given model, exclude the observation i and recompute $\hat{\beta}_{(i)}$, $\hat{\sigma}_{(i)}$ to obtain $\hat{Y}_{i(i)}$

Model Diagnostics



Leverage Values

Studentized &
Studentized Deleted
Residuals

DFFITS

Variance Inflation

Studentized &
Studentized Deleted
Residuals

DFFITS

ariance Inflation

Non-Constant
Variance &
Fransformation

ullet The observation i is an outlier if $\hat{Y}_{i(i)} - Y_i$ is "large"

Studentized &
Studentized Deleted
Residuals

DFFITS

ariance Inflation

Non-Constant
Variance &
Fransformation

ullet The observation i is an outlier if $\hat{Y}_{i(i)} - Y_i$ is "large"

FFITS

Variance Inflation

- For a given model, exclude the observation i and recompute $\hat{\beta}_{(i)}$, $\hat{\sigma}_{(i)}$ to obtain $\hat{Y}_{i(i)}$
- The observation i is an outlier if $\hat{Y}_{i(i)} Y_i$ is "large"
- $\begin{array}{l} \bullet \; \; \mathsf{Can} \; \mathsf{show} \\ \mathsf{Var}(\hat{Y}_{i(i)} Y_i) = \sigma_{(i)}^2 \left(1 + \boldsymbol{x}_i^T (\boldsymbol{X}_{(i)}^T \boldsymbol{X}_{(i)})^{-1} \boldsymbol{x}_i \right) = \frac{\sigma_{(i)}^2}{1 h_i} \end{array}$

FFITS

Variance Inflation

- For a given model, exclude the observation i and recompute $\hat{\beta}_{(i)}$, $\hat{\sigma}_{(i)}$ to obtain $\hat{Y}_{i(i)}$
- The observation i is an outlier if $\hat{Y}_{i(i)} Y_i$ is "large"
- $\begin{array}{l} \bullet \; \; \mathsf{Can} \; \mathsf{show} \\ \mathsf{Var}(\hat{Y}_{i(i)} Y_i) = \sigma_{(i)}^2 \left(1 + \boldsymbol{x}_i^T (\boldsymbol{X}_{(i)}^T \boldsymbol{X}_{(i)})^{-1} \boldsymbol{x}_i \right) = \frac{\sigma_{(i)}^2}{1 h_i} \end{array}$

Variance Inflation

Non-Constant Variance & Transformation

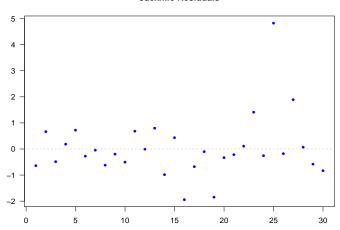
- For a given model, exclude the observation i and recompute $\hat{\beta}_{(i)}$, $\hat{\sigma}_{(i)}$ to obtain $\hat{Y}_{i(i)}$
- ullet The observation i is an outlier if $\hat{Y}_{i(i)} Y_i$ is "large"
- $\begin{array}{l} \bullet \; \; \mathsf{Can} \; \mathsf{show} \\ \mathsf{Var}(\hat{Y}_{i(i)} Y_i) = \sigma_{(i)}^2 \left(1 + \boldsymbol{x}_i^T (\boldsymbol{X}_{(i)}^T \boldsymbol{X}_{(i)})^{-1} \boldsymbol{x}_i \right) = \frac{\sigma_{(i)}^2}{1 h_i} \end{array}$
- Define the Studentized Deleted Residuals as

$$t_i = \frac{\hat{Y}_{i(i)} - Y_i}{\hat{\sigma}_{(i)}^2 / 1 - h_i} = \frac{\hat{Y}_{i(i)} - Y_i}{\mathsf{MSE}_{(i)} (1 - h_i)^{-1}}$$

which are distributed as a t_{n-p-1} if the model is correct and $\varepsilon \sim \mathrm{N}(\mathbf{0}, \sigma^2 \pmb{I})$

Jackknife Residuals of Species $\sim \mathtt{Elev} + \mathtt{Adj}$

Jacknife Residuals



Model Diagnostics



Leverage Values

Studentized & Studentized Deleted Residuals

OFFITS

Variance Inflation

on-Constant ariance &

DEFITS

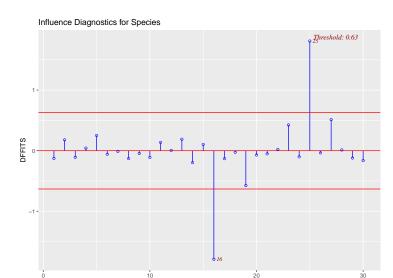
ariance Inflation

Non-Constant Variance & Transformation

DFFITS

- \bullet Difference between the fitted values \hat{Y}_i and the predicted values $\hat{Y}_{i(i)}$
- $\quad \bullet \; \mathsf{DFFITS}_i = \frac{\hat{Y}_i \hat{Y}_{i(i)}}{\sqrt{\mathsf{MSE}_{(i)} h_i}}$
- Concern if absolute value greater than 1 for small data sets, or greater than $2\sqrt{p/n}$ for large data sets

DFFITS of Species \sim **Elev** + **Adj**



Observation

Model Diagnostics



Leverage Values

Studentized & Studentized Deleted Residuals

FFITS

Variance Inflation

Variance Inflation Factor (VIF)

Leverage Values

Studentized &
Studentized Deleted
Residuals

DFFITS

/ariance Inflation actor

Non-Constant Variance & Transformation

 $\mathsf{VIF}_k = \frac{1}{1 - R_k^2},$

where R_k^2 is the coefficient of determination when X_k is regressed on the remaining p-2 other predictors.

> vif(step_gala)

Elevation Adjacent

1.404074 1.404074

> vif(full)

Area Elevation Nearest

2.928145 3.992545 1.766099

Scruz Adjacent

1.675031 1.826403

Residual Plot of Species $\sim \mathtt{Elev} + \mathtt{Adj}$

Model Diagnostics



Leverage Values

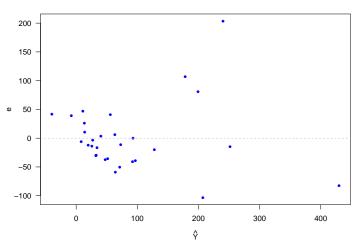
Studentized & Studentized Deleted Residuals

FITS

ariance Inflation

Non-Constant Variance & Transformation

Residuals



Residual Plot After Square Root Transformation

Model Diagnostics



Leverage Values

Studentized & Studentized Deleted Residuals

DEFITS

ariance Inflation

