Lecture 8

Multiple Linear Regression IV

Reading: Chapter 13

STAT 8020 Statistical Methods II
September 15, 2020

Whitney Huang
Clemson University

Agenda

1. **Variable Selection**
2. **Model Diagnostics: Residual Plots**
3. **Model Diagnostics: Influential Points**
4. **Non-Constant Variance & Transformation**

Other Approaches: Information criteria

- Akaike's information criterion (AIC)

$$n \log(\frac{\text{SSE}_k}{n}) + 2k$$

- Bayesian information criterion (BIC)

$$n \log(\frac{\text{SSE}_k}{n}) + k \log(n)$$

- Can be used to compare **non-nested** models

Automatic Search Procedures

- Forward Selection
- Backward Elimination
- Stepwise Search
- All Subset Selection

Model Assumptions

Model:

$$Y_i = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \cdots + \beta_{p-1} X_{p-1} + \varepsilon_i, \quad \varepsilon_i \overset{i.i.d.}{\sim} \mathrm{N}(0, \sigma^2)$$

We make the following assumptions:

- Linearity:

$$\mathrm{E}(Y|X_1, X_2, \cdots, X_{p-1}) = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \cdots + \beta_{p-1} X_{p-1}$$

Model Assumptions

Model:

$$Y_i = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \cdots + \beta_{p-1} X_{p-1} + \varepsilon_i, \quad \varepsilon_i \overset{i.i.d.}{\sim} \mathrm{N}(0, \sigma^2)$$

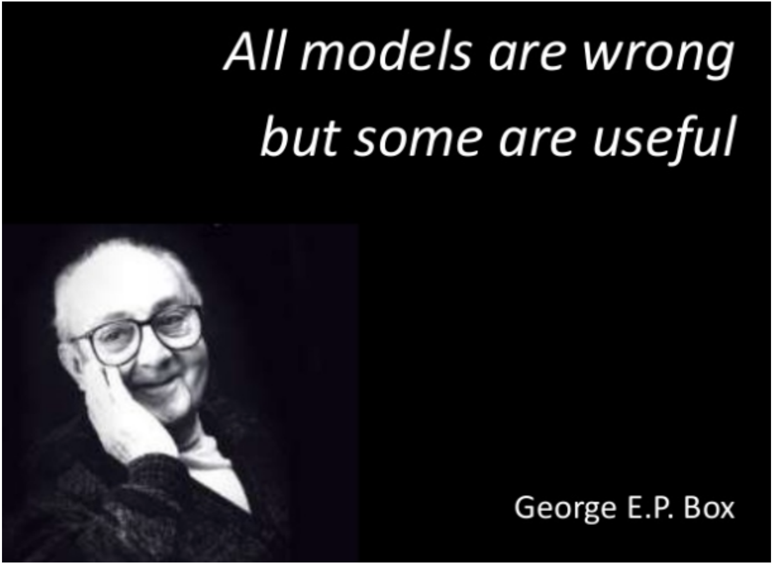
We make the following assumptions:

- Linearity:

$$\mathrm{E}(Y|X_1, X_2, \cdots, X_{p-1}) = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \cdots + \beta_{p-1} X_{p-1}$$

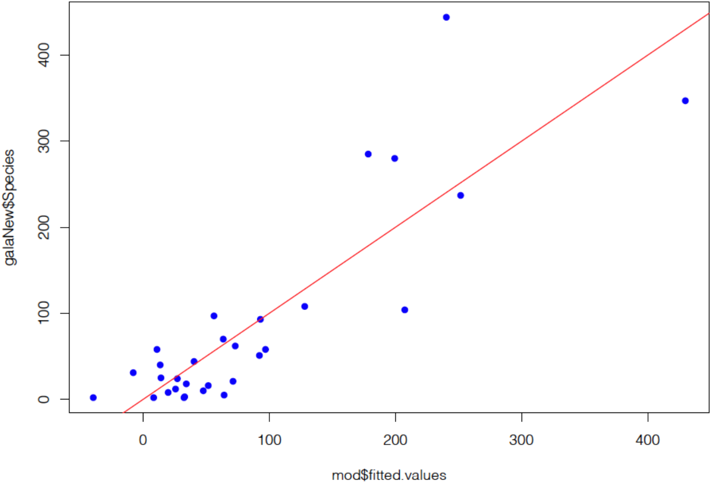
- Errors have constant variance, are independent, and normally distributed

$$\varepsilon_i \overset{i.i.d.}{\sim} \mathrm{N}(0, \sigma^2)$$


All models are wrong
but some are useful
George E.P. Box

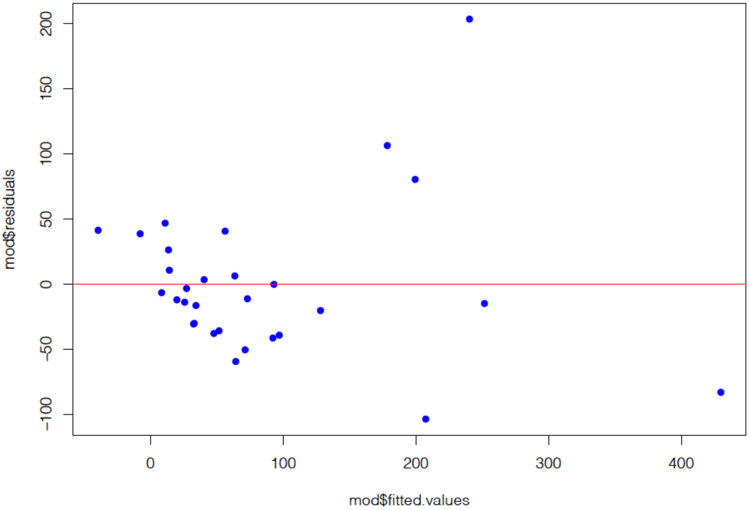
Observed Values versus Fitted Values Plot

```r
mod <- lm(Species ~ Elevation + Adjacent, data = galaNew)
plot(mod$fitted.values, galaNew$Species, pch = 16, col = "blue")
abline(0, 1, col = "red")
```



Residuals versus Fits Plot

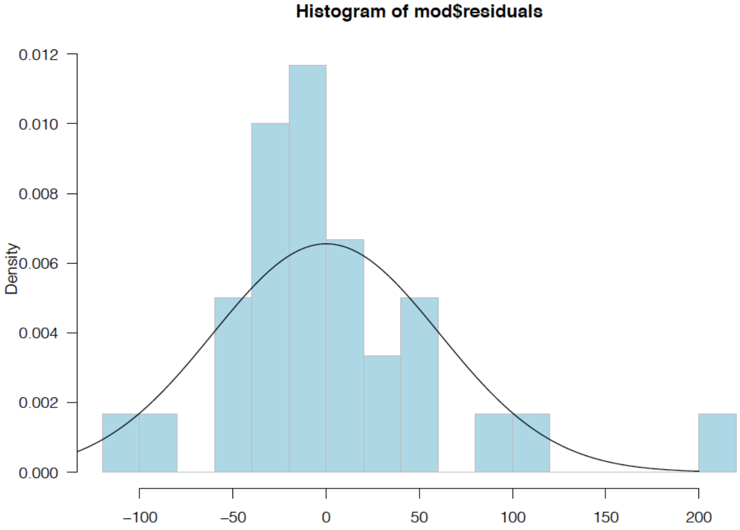
```r
plot(mod$fitted.values, mod$residuals, pch = 16, col = "blue")
abline(h = 0, col = "red")
```



We will revisit this in the end of the lecture

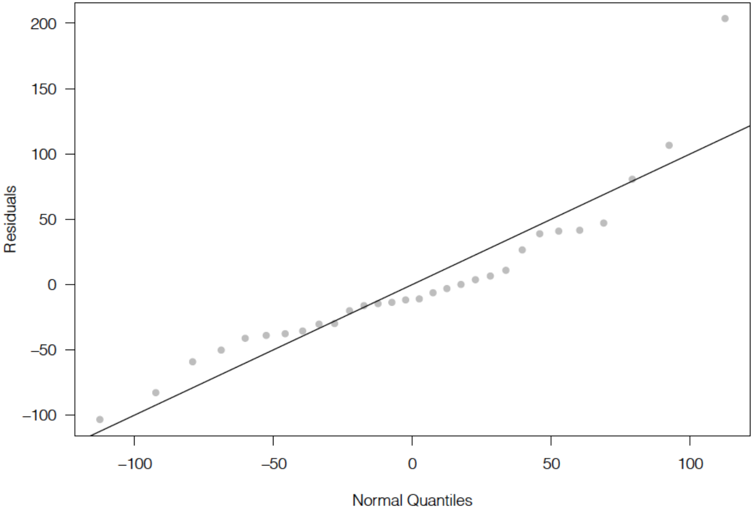
Assessing Normality of Residuals: Histogram

```r
par(las = 1)
hist(mod$residuals, 12, prob = T,
     col = "lightblue", border = "gray")
xg <- seq(-200, 200, 1)
yg <- dnorm(xg, 0, 60.86)
lines(xg, yg)
```



Assessing Normality of Residuals: QQ Plot

```r
plot(qnorm(1:30 / 31, 0, 60.86), sort(mod$residuals), pch = 16,
     col = "gray", xlab = "Normal Quantiles", ylab = "Residuals")
abline(0, 1)
```



Leverage

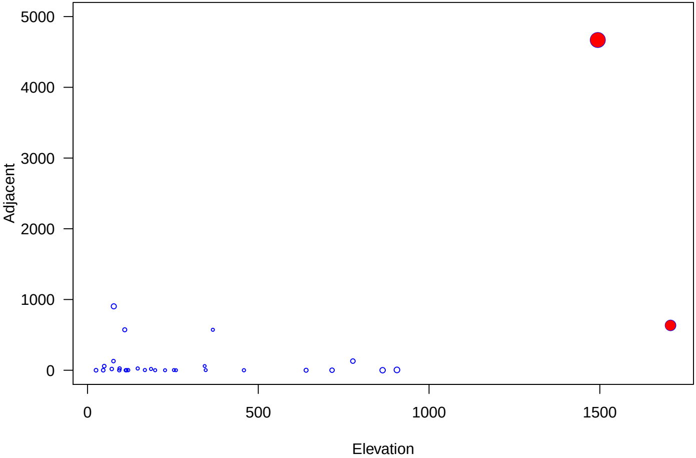
Recall in MLR that $\hat{Y} = \boldsymbol{X}(\boldsymbol{X}^T\boldsymbol{X})^{-1}\boldsymbol{X}^T\boldsymbol{Y} = \boldsymbol{H}\boldsymbol{Y}$ where $\boldsymbol{H}$ is the hat-matrix

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

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

- Can show that Var$(e_i) = \sigma^2(1 - h_i)$, where $e_i = Y_i - \hat{Y}_i$ is the residual for the i_{th} observation
- $\frac{1}{n} \leq h_i \leq 1, \quad 1 \leq i \leq n$ 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`

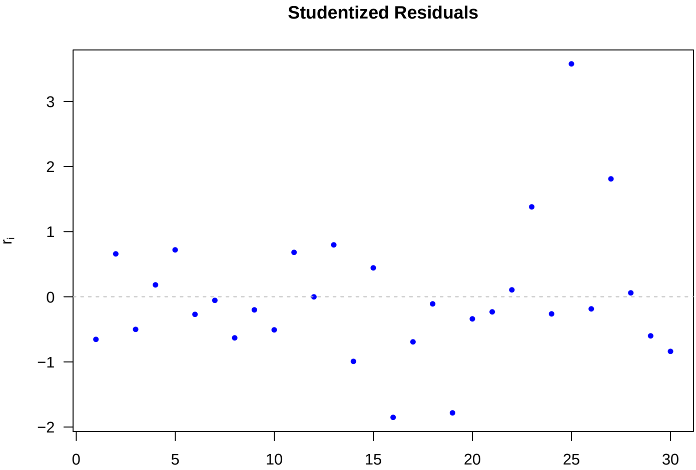


Studentized Residuals

As we have seen Var$(e_i) = \sigma^2(1 - h_i)$, this suggests the use of $r_i = \frac{e_i}{\hat{\sigma}\sqrt{(1-h_i)}}$

- 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 Var$(r_i) = 1$ and Corr(e_i, e_j) tends to be small

Studentized Residuals of `Species` $\sim$ `Elev` + `Adj`



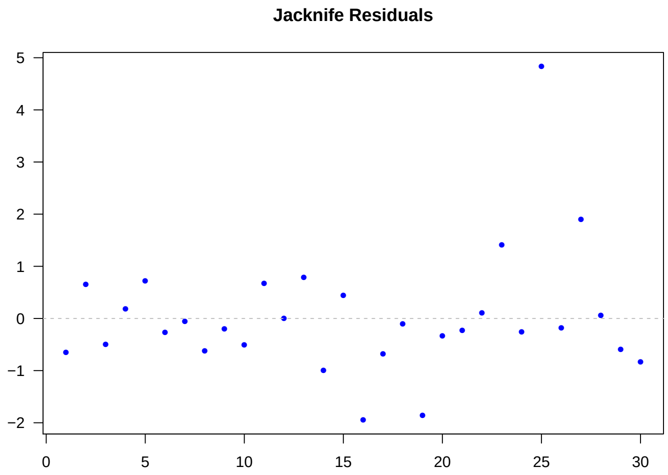
Studentized Deleted Residuals

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

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

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

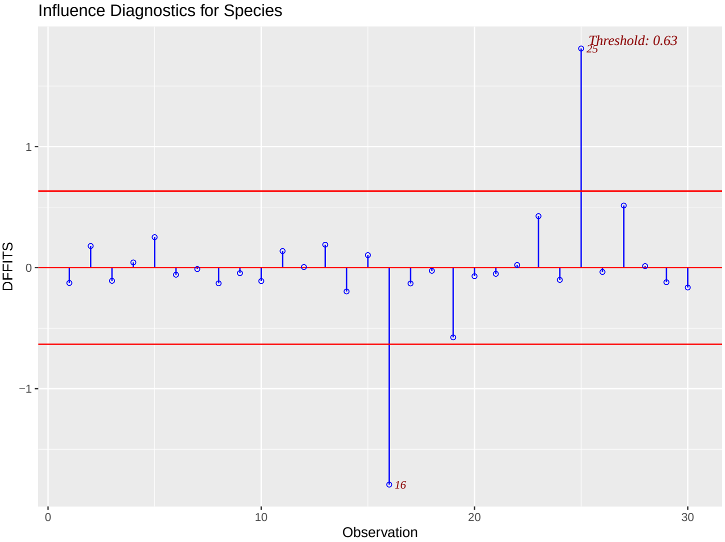
Jackknife Residuals of `Species` $\sim$ `Elev` $+$ `Adj`



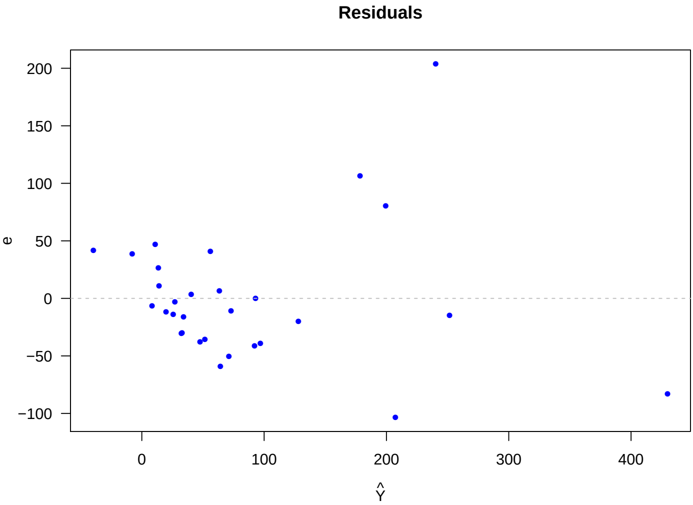
DFFITS

- Difference between the fitted values $\hat{Y}_i$ and the predicted values $\hat{Y}_{i(i)}$
- $\text{DFFITS}_i = \frac{\hat{Y}_i - \hat{Y}_{i(i)}}{\sqrt{\text{MSE}_{(i)} h_i}}$
- 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`



Residual Plot of `Species` $\sim$ `Elev` + `Adj`



Residual Plot After Square Root Transformation