\( a = b_0 + b_1 x \)

\( c = y - \hat{y} \)
Checking for Influential Observations

An observation can be influential because it has an extreme $x$-value, an extreme $y$-value, or both.

Note that the influential observation above are not outliers since their residuals are quite small (even zero).
\[ \hat{y} = \sum_{j=1}^{n} h_{ij} y_j \]

where the \( h_{ij} \) are some functions of the \( x \)'s

\[ H = [h_{ij}] \]

is called the hat matrix

We can think of the \( h_{ij} \) as the leverage exerted by \( y_j \) on the fitted value \( \hat{y}_j \).

If \( \hat{y}_i \) is largely determined by \( y_i \) with very small contribution from the other \( y_j \)'s, then we say that the \( i \)th observations is influential (also called high leverage).
Graph of Anscombe Data

Figure 10.9 Scatter Plot of the Anscombe Data Illustrating an Influential Observation
Checking for Influential Observations

Anscombe Data Illustration:

\[ h_{ii} > \frac{4}{n} = \frac{4}{11} = 0.36 \]

Thus observation 8 is influential

Notice: Observation No. 8 is not an outlier since \( e_8 = y_8 - \hat{y}_8 < 0 \)
How to Deal with Outliers and Influential Observations

- They can give a misleading picture of the relationship between $y$ and $x$.
- Are they erroneous observations? If yes, they should be discarded.
- If the observation are valid then they should be included in the analysis.
- Least Squares is very sensitive to these observations. Do analysis with and without outlier or influential observation.
- Or use robust method: Minimize the sum of the absolute deviations

$$\sum_{i=1}^{n} |y_i - (\beta_0 + \beta_1 x_i)|$$

These estimates are analogous to the sample median. LS estimates are analogous to the sample mean.
Data Transformations

If the functional relationship between $x$ and $y$ is known it may be possible to find a linearizing transformation analytically:

For $y = \alpha x^\beta$ take log on both sides: $\log y = \log \alpha + \beta \log x$

For $y = \alpha e^{\beta x}$ take log on both sides: $\log y = \log \alpha + \beta x$

If we are fitting an empirical model, then a linearizing transformation can be found by trial and error using the scatter plot as a guide.
Scatter Plots and Linearizing Transformations

Figure 10.10 Typical Scatter Plot Shapes and Corresponding Linearizing Transformations
Tire Tread Wear vs. Mileage:
Exponential Model

\[ y = \alpha e^{-\beta x} \]
\[ \Rightarrow \log y = \log \alpha - \beta x \]
Tire Tread Wear vs. Mileage: Exponential Model

\[ WEAR = e^{5.93} \times e^{-0.0298 \text{ MILES}} = 376.15e^{-0.0298 \text{ MILES}} \]
Tire Tread Wear vs. Mileage: Exponential Model

\[ \hat{y} = b_0 + b_1 e^{b_2 x} \]

For linear model: \( R^2 = .953 \)

Residual plot is still curved mainly because observation 1 is an outlier.
Weights and Gas Mileages of 38 1978-79 Model Year Cars

<table>
<thead>
<tr>
<th>Year Cars</th>
<th>WT (1000 lb)</th>
<th>MPG</th>
<th>WT (1000 lb)</th>
<th>MPG</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>4,560</td>
<td>16.0</td>
<td>3,070</td>
<td>16.2</td>
</tr>
<tr>
<td></td>
<td>4,054</td>
<td>15.5</td>
<td>2,505</td>
<td>16.5</td>
</tr>
<tr>
<td></td>
<td>3,605</td>
<td>19.2</td>
<td>2,010</td>
<td>21.9</td>
</tr>
<tr>
<td></td>
<td>3,960</td>
<td>18.5</td>
<td>1,975</td>
<td>14.1</td>
</tr>
<tr>
<td></td>
<td>2,115</td>
<td>30.0</td>
<td>1,985</td>
<td>25.1</td>
</tr>
<tr>
<td></td>
<td>2,590</td>
<td>27.5</td>
<td>2,670</td>
<td>27.4</td>
</tr>
<tr>
<td></td>
<td>3,300</td>
<td>27.3</td>
<td>1,990</td>
<td>34.5</td>
</tr>
<tr>
<td></td>
<td>7,230</td>
<td>20.4</td>
<td>2,131</td>
<td>21.6</td>
</tr>
<tr>
<td></td>
<td>7,690</td>
<td>20.3</td>
<td>2,670</td>
<td>26.4</td>
</tr>
<tr>
<td></td>
<td>3,140</td>
<td>17.0</td>
<td>2,595</td>
<td>26.8</td>
</tr>
<tr>
<td></td>
<td>7,995</td>
<td>21.6</td>
<td>2,760</td>
<td>26.4</td>
</tr>
<tr>
<td></td>
<td>7,410</td>
<td>36.2</td>
<td>2,556</td>
<td>33.5</td>
</tr>
<tr>
<td></td>
<td>3,380</td>
<td>20.6</td>
<td>2,200</td>
<td>34.2</td>
</tr>
<tr>
<td></td>
<td>3,070</td>
<td>20.8</td>
<td>2,010</td>
<td>31.8</td>
</tr>
<tr>
<td></td>
<td>2,620</td>
<td>18.6</td>
<td>2,130</td>
<td>12.3</td>
</tr>
<tr>
<td></td>
<td>3,410</td>
<td>18.1</td>
<td>2,190</td>
<td>30.5</td>
</tr>
<tr>
<td></td>
<td>3,840</td>
<td>17.0</td>
<td>2,615</td>
<td>22.0</td>
</tr>
<tr>
<td></td>
<td>5,725</td>
<td>17.6</td>
<td>2,800</td>
<td>21.5</td>
</tr>
<tr>
<td></td>
<td>5,985</td>
<td>16.5</td>
<td>1,025</td>
<td>31.5</td>
</tr>
</tbody>
</table>
Weights and Gas Mileages of 38
1978-79 Model Year Cars

Using 
Fit Y by X platform

A spline fits the data locally. Our prior curve fits have been global.

Use transformation \( y \to \frac{100}{y} \). The new variable has units of gallons per 100 miles.

This transformation has the advantage of being physically interpretable.
Weights and Gas Mileages of 38 1978-79 Model Year Cars

\[ \text{Variance Stabilizing Transformation} \]

Suppose \( \sigma^2 = \frac{1}{\lambda} (m + x) \)
Variance Stabilizing Transformation

Suppose $\sigma_y \propto \mu^\alpha$

We want to find a transformation of $y$ that yields a constant variance. Suppose the transformation is a power of the original data, say $y^* = y^{\frac{1}{2}}$

<table>
<thead>
<tr>
<th>Relationship Between $\sigma_y$ and $\mu$</th>
<th>$\alpha$</th>
<th>$\lambda = 1 - \alpha$</th>
<th>Transformation</th>
<th>Comment</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\sigma_y \propto \mu^0$</td>
<td>0</td>
<td>1</td>
<td>No transformation</td>
<td></td>
</tr>
<tr>
<td>$\sigma_y \propto \mu^{1/2}$</td>
<td>1/2</td>
<td>1/2</td>
<td>Square root</td>
<td>Poisson (count) data</td>
</tr>
<tr>
<td>$\sigma_y \propto \mu$</td>
<td>1</td>
<td>0</td>
<td>Log</td>
<td></td>
</tr>
<tr>
<td>$\sigma_y \propto \mu^{-1/2}$</td>
<td>3/2</td>
<td>-1/2</td>
<td>Reciprocal square root</td>
<td></td>
</tr>
<tr>
<td>$\sigma_y \propto \mu^{-1}$</td>
<td>2</td>
<td>-1</td>
<td>Reciprocal</td>
<td></td>
</tr>
</tbody>
</table>

For example: Take the square root of Poisson count data since for data so distributed the mean and the variance are the same.
Box-Cox Transformation: 
Empirical Determination of Best Transformation

\[ y^{(\lambda)} = \begin{cases} 
\left( y^2 - 1 \right) / \lambda & \lambda \neq 0 \\
\bar{y} \ln y & \lambda = 0 
\end{cases} \]

where \( \bar{y} = \sqrt[n]{\prod_{i=1}^{n} y_i} \) = geometric mean
Transformation in Tire Data

A better fit than than the polynomial and with one less parameter (3 versus 4)
Weights and Gas Mileages of 38 1978-79 Model Year Cars: Box-Cox Transformation

Reciprocal transformation among suggested transformations
Correlation Analysis

- When it is not clear which is the predictor variable and which is the response variable
- When both variables are random
- Try the bivariate normal distribution as a probability model for the joint distribution of two r.v.’s

Parameters: $\mu_X, \mu_Y, \sigma_X, \sigma_Y$ and

$$\rho = \text{Corr}(X, Y) = \frac{\text{Cov}(X, Y)}{\sqrt{\text{Var}(X)\text{Var}(Y)}}$$

The correlation $\rho$ is a measure of association between the two random variables $X$ and $Y$. Zero correlation corresponds to no association. Correlation of 1 or -1 represents perfect association.
Bivariate Normal Density Function

Figure 18.15 Bell Shaped Surface Defined by the Bivariate Normal Density Function

\[ r = 1 \]
\[ b_1 = 0 \]
\[ \tau = 0 \]

\[ y = b_0 + \alpha x \]

\[ H_0 : b_1 = 0 \]
\[ H_0 : \tau = 0 \]

\[ -1 \leq \tau \leq 1 \]
\[ R^2 = (\rho^2) \]
\[ 0 \leq R^2 \leq 1 \]

Statistical Inference on the Correlation Coefficient
\[ H_0 : b_1 = 0 \]
\[ H_{11} : r = 0 \]

\[-1 \leq r \leq 1 \quad \therefore 0 \leq r^2 \leq 1\]

Statistical Inference on the Correlation Coefficient

\[ R = \frac{\sum_{i=1}^{n} (X_i - \bar{X})(Y_i - \bar{Y})}{\sqrt{\sum_{i=1}^{n} (X_i - \bar{X})^2 \sum_{i=1}^{n} (Y_i - \bar{Y})^2}} \]

\[ H_0 : \rho = 0 \text{ vs } H_1 : \rho \neq 0 \]

Test statistic \( T = \frac{R \sqrt{n-2}}{\sqrt{1-R^2}} \cdot t_{n-2} \)

Equivalent to testing \( H_0 : \beta_1 = 0 \text{ vs. } H_1 : \beta_1 \neq 0 \)

Approximate test statistic available to
- Test nonzero null hypotheses
- Obtain confidence intervals for \( \rho \)

See textbook
Exercise 10.28

The following are the heights and weights of 30 eleven year old girls:\(^\text{21}\)

<table>
<thead>
<tr>
<th>Height (cm)</th>
<th>Weight (kg)</th>
<th>Height (cm)</th>
<th>Weight (kg)</th>
</tr>
</thead>
<tbody>
<tr>
<td>135</td>
<td>26</td>
<td>133</td>
<td>31</td>
</tr>
<tr>
<td>146</td>
<td>33</td>
<td>149</td>
<td>34</td>
</tr>
<tr>
<td>153</td>
<td>55</td>
<td>141</td>
<td>32</td>
</tr>
<tr>
<td>154</td>
<td>50</td>
<td>164</td>
<td>47</td>
</tr>
<tr>
<td>139</td>
<td>32</td>
<td>146</td>
<td>37</td>
</tr>
<tr>
<td>131</td>
<td>25</td>
<td>149</td>
<td>46</td>
</tr>
<tr>
<td>149</td>
<td>44</td>
<td>147</td>
<td>36</td>
</tr>
<tr>
<td>137</td>
<td>31</td>
<td>152</td>
<td>47</td>
</tr>
<tr>
<td>143</td>
<td>36</td>
<td>140</td>
<td>33</td>
</tr>
<tr>
<td>146</td>
<td>35</td>
<td>143</td>
<td>42</td>
</tr>
<tr>
<td>141</td>
<td>28</td>
<td>148</td>
<td>32</td>
</tr>
<tr>
<td>136</td>
<td>28</td>
<td>149</td>
<td>32</td>
</tr>
<tr>
<td>154</td>
<td>26</td>
<td>141</td>
<td>29</td>
</tr>
<tr>
<td>151</td>
<td>48</td>
<td>137</td>
<td>34</td>
</tr>
<tr>
<td>155</td>
<td>36</td>
<td>135</td>
<td>30</td>
</tr>
</tbody>
</table>

(a) Plot weights vs. heights.
(b) Calculate the correlation coefficient. Test if it is significantly greater than 0.7.

\[ H_0: \rho = \rho_0 \]

\[ H_1: \rho \neq \rho_0 \]

\[ \rho = 0.7424 \]

\[ 0.5216 < \rho < 0.8700 \]

\[ \rho 
eq 0.5 \]
Exercise 10.28
Exercise 10.28

By default, a 95% bivariate normal density ellipse is imposed on each scatterplot. If the variables are bivariate normally distributed, this ellipse encloses approximately 95% of the points. The correlation of the variables is seen by the collapsing of the ellipse along the diagonal axis. If the ellipse is fairly round and is not diagonally oriented, the variables are uncorrelated.