Simple Linear Regression

\[ Estimating \sigma^2 : MSE_{\text{Error}} = MS_{\text{Residual}} = S_{LP}^2 = \frac{SS_{\text{Residual}}}{df_{\text{Residual}}} \]

\[ SS_{\text{Residual}} = \sum (\text{Residual})^2 \]

Residual df = n – (# of parameters estimated)
• Estimating Intercept and Slope
• Residual df = n-2

\[ Var(b_1) = \frac{\sigma^2}{(n-1)s_x^2} \]
\[ SE(b_1) = \sqrt{\frac{MS_{\text{Residual}}}{(n-1)s_x^2}} \]

Estimating the mean value of \( y \) for a fixed value of \( x \)

\[ Var(\hat{y}) = Var(\bar{y} + b_1(x - \bar{x})) = Var(\bar{y}) + Var(b_1(x - \bar{x})) \]

= since \( \bar{y} \) and \( b_1 \) are independent

\[ \frac{\sigma^2}{n} + (x - \bar{x})^2 Var(b_1) \]

\[ SE_{\hat{y}} = \sqrt{\frac{MS_{\text{Residual}}}{n} + (x - \bar{x})^2 SE_{b_1}^2} \]
Predicting a new value of y:

Variability in predicting a new y value includes
- Variability in individual observations, $\sigma^2$.
- Variability in estimating $\mu$

$$Var(\hat{y}_{new} - \hat{y}) = \sigma^2 + Var(\hat{y})$$

$$SE(\hat{y}_{new} - \hat{y}) = \sqrt{MS \text{ Residual} + SE_{\hat{y}}^2}$$

As usual
- Confidence Intervals
  $$\text{Estimate} \pm t_{0.975}SE_{\text{Estimate}}$$
- Hypothesis Tests
  $$T = \frac{\text{Estimate} - \#}{SE_{\text{Estimate}}}$$

ANOVA Table

<table>
<thead>
<tr>
<th>Source</th>
<th>df</th>
<th>Sum of Squares</th>
</tr>
</thead>
<tbody>
<tr>
<td>Regression (X)</td>
<td>1</td>
<td>$SS_{\text{Total}} - SS_{\text{Residual}}$</td>
</tr>
<tr>
<td>Residual</td>
<td>n-2</td>
<td>$\sum \text{Residual}^2$</td>
</tr>
<tr>
<td>Total</td>
<td>n-1</td>
<td>$\sum (y_i - \bar{y})^2$</td>
</tr>
</tbody>
</table>

As in Chapter 4:

$SS_{\text{Regression}} = \text{improvement in sum of squared errors by using } X \text{ to predict } Y$
- $SS_{\text{Total}} = \text{sum of squared errors by using } \bar{Y} \text{ predict } Y$
- $SS_{\text{Residual}} = \text{sum of squared errors by using } b_0 + b_1X \text{ to predict } Y$
- $SS_{\text{Regression}} = SS_{\text{Total}} - SS_{\text{Error}}$

$$R^2 = \% \text{ improvement} = \frac{SS_{\text{Total}} - SS_{\text{Residual}}}{SS_{\text{Total}}} \times 100\% = \frac{SS_{\text{Regression}}}{SS_{\text{Total}}} \times 100\%$$