Solution for Homework #3

2.15 a) $X_h = 2: \hat{Y}_h = 10.2 + 4(2) = 18.2 \quad \alpha=.01 \quad t\left(1 - \frac{\alpha}{2}, n - 2\right) = t(.995, 8) = 3.355$

$$s^2\{\hat{Y}_h\} = MSE \left[\frac{1}{n} + \frac{(X_h - \bar{X})^2}{\Sigma(X_i - \bar{X})^2}\right] = 2.2 \ast \left[\frac{1}{10} + \frac{(2 - 1)^2}{10}\right] = .44 \quad s\{\hat{Y}_h\} = .663$$

$$\hat{Y}_h \pm t\left(1 - \frac{\alpha}{2}, n - 2\right)s\{\hat{Y}_h\}$$

18.2 ± (3.355)(.663)
18.2 ± 2.224
(15.976, 20.424) When there are 2 transfers, we expect there to be between 15.976 and 20.424 broken ampules on average, with 99% confidence.

$X_h = 4: \hat{Y}_h = 10.2 + 4(4) = 26.2 \quad \alpha=.01 \quad t\left(1 - \frac{\alpha}{2}, n - 2\right) = t(.995, 8) = 3.355$

$$s^2\{\hat{Y}_h\} = MSE \left[\frac{1}{n} + \frac{(X_h - \bar{X})^2}{\Sigma(X_i - \bar{X})^2}\right] = 2.2 \ast \left[\frac{1}{10} + \frac{(4 - 1)^2}{10}\right] = 2.2 \quad s\{\hat{Y}_h\} = 1.483$$

$$\hat{Y}_h \pm t\left(1 - \frac{\alpha}{2}, n - 2\right)s\{\hat{Y}_h\}$$

26.2 ± (3.355)(1.483)
26.2 ± 4.975
(21.225, 31.175) When there are 4 transfers, we expect there to be between 21.225 and 31.175 broken ampules on average, with 99% confidence.

b) $X_h = 2: \quad s^2\{pred\} = MSE \left[\frac{1}{n} + \frac{(X_h - \bar{X})^2}{\Sigma(X_i - \bar{X})^2}\right] = 2.2 \ast \left[\frac{1}{10} + \frac{(2 - 1)^2}{10}\right] = 2.64$

$$s\{pred\} = 1.625 \quad \hat{Y}_h \pm t\left(1 - \frac{\alpha}{2}, n - 2\right)s\{pred\}$$

18.2 ± (3.355) (1.625)
18.2 ± 5.452
(12.748, 23.652)

We expect with 99% confidence that there will be between 12.748 and 23.652 breakages in the next shipment.
c) $s^2 \{\text{predmean}\} = \text{MSE} \left[ \frac{1}{m} + \frac{1}{n} + \frac{(X_h - \bar{X})^2}{\sum(X_i - \bar{X})^2} \right]^2 = 2.2 \left[ \frac{1}{3} + \frac{1}{10} + \frac{(2-1)^2}{10} \right] = 1.173$

$s\{\text{predmean}\} = 1.083 \quad \hat{Y}_h \pm t \left( 1 - \frac{\alpha}{2}, n-2 \right) s\{\text{predmean}\}$

$18.2 \pm 3.633 \quad 18.2 \pm 3.355 \quad (14.567, 21.833)$

To convert this into a 99% prediction interval for the total number of broken ampules, multiply the boundaries by 3: $(3(14.567), 3(21.833)) = (43.701, 65.499) \text{ or } (44, 65)$.

d) $W^2 = 2F(1-\alpha, 2, n-2) = 2F(.99, 2, 8) = 2(8.649) = 17.298 \quad W = 4.159$

$\hat{Y}_h \pm W \cdot s\{\hat{Y}_h\}$

For $X_h = 2: \quad 18.2 \pm 4.159 \cdot (1.483) \rightarrow (15.443, 20.957)$

For $X_h = 4: \quad 26.2 \pm 4.159 \cdot (1.483) \rightarrow (20.032, 32.368)$

Notice that both of these intervals are wider than those in part a). They should be wider, because the confidence bands encompass the entire regression line, not just one point. Refer to p. 61-62 of the text for more details.

2.25 a)

The REG Procedure
Model: MODEL1
Dependent Variable: amp

Analysis of Variance

<table>
<thead>
<tr>
<th>Source</th>
<th>DF</th>
<th>Sum of Squares</th>
<th>Mean Square</th>
<th>F Value</th>
<th>Pr &gt; F</th>
</tr>
</thead>
<tbody>
<tr>
<td>Model (SSR)</td>
<td>1</td>
<td>160.00000</td>
<td>160.00000</td>
<td>72.73</td>
<td>&lt;.0001</td>
</tr>
<tr>
<td>Error (SSE)</td>
<td>8</td>
<td>17.60000</td>
<td>2.20000</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Corrected Total (SSTO)</td>
<td>9</td>
<td>177.60000</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Root MSE       1.48324 R-Square    0.9009
Dependent Mean 14.20000 Adj R-Sq    0.8885
Coeff Var      10.44535

SSR + SSE = SSTO
df(SSR) + df(SSE) = df(SSTO) \quad \text{Where df(SSR) means degrees of freedom associated with SSR, etc.}

b) $H_0: \beta_1 = 0 \quad H_A: \beta_1 \neq 0 \quad \alpha = .05$

Reject $H_0$ if $F^* > F(1-\alpha, 1, n-2)$, otherwise do not reject $H_0$.

$F^* = \frac{MSR}{MSE} = \frac{160}{2.2} = 72.73 \quad F(.95, 1, 8) = 5.32$
Reject $H_0$. There is a linear association between the number of transfers and the number of broken ampules.

c) Parameter Estimates

<table>
<thead>
<tr>
<th>Variable</th>
<th>DF</th>
<th>Parameter Estimate</th>
<th>Standard Error</th>
<th>t Value</th>
<th>Pr &gt;</th>
<th>t</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intercept</td>
<td>1</td>
<td>10.20000</td>
<td>0.66332</td>
<td>15.38</td>
<td>&lt;.0001</td>
<td></td>
</tr>
<tr>
<td>trans</td>
<td>1</td>
<td>4.00000</td>
<td>0.46904</td>
<td>8.53</td>
<td>&lt;.0001</td>
<td></td>
</tr>
</tbody>
</table>

Recall: $t^* = \frac{b_1}{s(b_1)} = \frac{4.0}{0.469} = 8.529$

The relationship between a t-statistic with n degrees of freedom and an F-statistic with $(1, n)$ degrees of freedom is: $t^2 = F$. So, $(t^*)^2 = 8.529^2 = 72.74$ (difference is due to rounding).

d) $R^2 = \frac{SSR}{SSTO} = \frac{160}{177.6} = 0.9009$ (this appears under the ANOVA table in part a)

$$r = \sqrt{R^2} = \sqrt{0.9009} = 0.9492$$ (r is positive, since $b_1$ is positive)

90.09% of the variation in Y can be explained by introducing X into the regression model.

2.56 a) Using the formulas on p. 68, we have:

$$E\{\text{MSE}\} = \sigma^2 = (0.6)^2 = 0.36$$

$$E\{\text{MSR}\} = \sigma^2 + \beta_1^2 \sum (X_i - \bar{X})^2 \quad \text{Note: } \bar{X} = 8, \sum (X_i - \bar{X})^2 = 114$$

$$= (0.6)^2 + 3^2(114) = 1026.36$$

b) Note, if we have $X = 6, 7, 8, 9$ and 10, we still have $\bar{X} = 8$. Then $\sum (X_i - \bar{X})^2 = 10$. $E\{\text{MSR}\} = (0.6)^2 + 3^2(10) = 90.36$

In an F-test for model fit, larger values of $F = \frac{\text{MSR}}{\text{MSE}}$ will indicate a good fit – the larger, the better.

$$F_a^* = \frac{1026.36}{0.36} = 2851 \quad F_b^* = \frac{90.36}{0.36} = 251$$

Both of these F-values are very large (the largest value in the F(1,3) distribution on p. 1321 is 167) and would lead to a rejection of $H_0$: $\beta_1 = 0$, but $F_a^*$ is larger and thus more supportive of the alternative hypothesis (that a regression relation exists). If the principal purpose were to estimate the mean response for $X = 8$, it would not matter which set of X
values we used, since \((X_h - \bar{X})^2 = 0\) in both cases, which means the formula

\[
s^2 \{ \hat{Y}_h \} = \text{MSE} \left[ \frac{1}{n} + \frac{\left( \sum (X_h - \bar{X})^2 \right)}{\text{MSE}} \right] \]

would have the same value.

2.57 a) \(Y_i - 5X_i = \beta_0 + \epsilon_i\)
\[df_R = n - 1\]

b) \(Y_i - 2 - 5X_i = \epsilon_i\)
\[df_R = n\]

2.61 The formula 2.51 is \(\text{SSR} = b_1^2 \sum (X_i - \bar{X})^2\).

Our starting point (equation 2.72) is \(R^2 = \frac{\text{SSR}}{\text{SSTO}} = \frac{b_1^2 \sum (X_i - \bar{X})^2}{\sum (Y_i - \bar{Y})^2}\)

Taking the square root of both sides gives: \(r = \frac{b_1 \left( \sum (X_i - \bar{X})^2 \right)^{1/2}}{\left( \sum (Y_i - \bar{Y})^2 \right)^{1/2}}\)

Recall from formula 1.10a: \(b_1 = \frac{\sum (X_i - \bar{X})(Y_i - \bar{Y})}{\sum (X_i - \bar{X})^2}\)

Then, \(r = \frac{\sum (X_i - \bar{X})(Y_i - \bar{Y}) \left( \sum (X_i - \bar{X})^2 \right)^{1/2}}{\sum (X_i - \bar{X})^2 \left( \sum (Y_i - \bar{Y})^2 \right)^{1/2}}\)

\[= \frac{\sum (X_i - \bar{X})(Y_i - \bar{Y})}{\left( \sum (X_i - \bar{X})^2 \right)^{1/2} \left( \sum (Y_i - \bar{Y})^2 \right)^{1/2}}\]

\[= \frac{\sum (X_i - \bar{X})(Y_i - \bar{Y})}{\left( \sum (X_i - \bar{X})^2 \sum (Y_i - \bar{Y})^2 \right)^{1/2}}\]

this is equation 2.84 where
\(X_i = Y_{i1}, \bar{X} = \bar{Y}_1, \text{ and } Y_i = Y_{i2}, \bar{Y} = \bar{Y}_2\). From this, we can see that \(r\) is the same whether \(Y_i\) is regressed on \(Y_2\) or \(Y_2\) is regressed on \(Y_i\). Hence, the ratio \(\text{SSR}/\text{SSTO}\) is also the same.