10.1

(a) $\beta_0 = 4.7$
This says that when the U.S. market is flat, the average overseas return will be 4.7%.

(b) $\beta_1 = 0.166$
This number indicates for every 1% increase in the U.S. market, the overseas return will increase 0.166%.

(c) $y_i = 4.7 + 0.166x_i + e_i$, where $e_i$ denotes the variation in overseas returns.
The scatterplot here shows a fairly strong linear relationship that is also positive.

\[ y = 2.666 + 0.627x \]

**Linear Fit**

\[ T\text{-bill} = 2.6662262 + 0.6269356 \text{ Inflation} \]

**Summary of Fit**

<table>
<thead>
<tr>
<th></th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>R\text{Square}</td>
<td>0.448878</td>
</tr>
<tr>
<td>R\text{Square} Adj</td>
<td>0.437631</td>
</tr>
<tr>
<td>Root Mean Square Error</td>
<td>2.18016</td>
</tr>
<tr>
<td>Mean of Response</td>
<td>5.198431</td>
</tr>
<tr>
<td>Observations (or Sum Wgts)</td>
<td>51</td>
</tr>
</tbody>
</table>

**Analysis of Variance**

<table>
<thead>
<tr>
<th>Source</th>
<th>DF</th>
<th>Sum of Squares</th>
<th>Mean Square</th>
<th>F Ratio</th>
<th>Prob&gt;F</th>
</tr>
</thead>
<tbody>
<tr>
<td>Model</td>
<td>1</td>
<td>189.69399</td>
<td>189.694</td>
<td>39.9096</td>
<td>&lt;.0001</td>
</tr>
<tr>
<td>Error</td>
<td>49</td>
<td>232.90189</td>
<td>4.753</td>
<td></td>
<td></td>
</tr>
<tr>
<td>C. Total</td>
<td>50</td>
<td>422.59587</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**Parameter Estimates**

| Term   | Estimate | Std Error | t Ratio | Prob>|t| |
|--------|----------|-----------|---------|------|
| Intercept | 2.6662262 | 0.503848  | 5.29    | <.0001 |
| Influence | 0.6279356 | 0.099239 | 6.32    | <.0001 |

\[ b_1 = 0.627 \]

\[ \text{std error } b_1 = 0.0992 \]

\[ H_0: \beta_1 = 0 \quad \text{vs.} \quad H_a: \beta_1 > 0 \]

\[ t = \frac{0.627}{0.0992} = 6.32 \]

\[ p\text{-value} = 3.75116 \times 10^{-8} \quad \approx 0 \]

There is significant evidence to conclude \( \beta_1 > 0 \).
10.15

(a) If there is no inflation in a particular year, 
\( \beta_0 \) represents the return of Treasuries. In order 
for the government to issue treasury bills, 
there must be a positive return, hence we 
should expect \( \beta_0 > 0 \).

(b) JMP estimates:
\[
\hat{\beta}_0 = 2.6666
\]
Std error \( \hat{\beta}_0 = 0.5038 \)

(c) \( H_0: \beta = 0 \) vs. \( H_a: \beta > 0 \).

\[
t = \frac{2.6666}{0.5038} = 5.29
\]

p-value : \( 1.420606 \times 10^{-6} \approx 0 \)

Yes, there is sufficient evidence to conclude \( \beta_0 > 0 \).

(d) \( \beta_0 \pm t(95,49) \) std error \( \hat{\beta}_0 \)

\[
2.6666 \pm (2.009)(0.5038)
\]

10.21a

(a) \( \hat{y} = 2.6666 + 0.627(3.7) \)

\[
\hat{y} = 4.9859
\]

(b) (cont. on next page)
10.26

(b) Inflation Prediction

<table>
<thead>
<tr>
<th>Lower 95%</th>
<th>Upper 95%</th>
<th>from JMP</th>
</tr>
</thead>
<tbody>
<tr>
<td>3.7</td>
<td>0.56143228</td>
<td>9.41034354</td>
</tr>
</tbody>
</table>

10.30

(a) 90% CI for the Mean from JMP

<table>
<thead>
<tr>
<th>Lower 95%</th>
<th>Upper 95%</th>
</tr>
</thead>
<tbody>
<tr>
<td>3.7</td>
<td>4.47096598</td>
</tr>
<tr>
<td></td>
<td>5.50080984</td>
</tr>
</tbody>
</table>

(b) We would need to find the std error only.

10.35

\[ \sum (x_i - \bar{x})^2 = (n-1)S_x^2 \]
\[ = (50)(3.1068397)^2 \]
\[ = 482.62 \]

\[ SE_{b_1} = \frac{s}{\sqrt{\sum (x_i - \bar{x})^2}} \]
\[ = \frac{2.1801}{\sqrt{482.62}} \]
\[ = .09924 \]

\[ 10.36 \quad S_x^2 = 9.6525 \]

\[ SE_{y^*} = s \sqrt{1 + \frac{1}{n} + \frac{(x^* - \bar{x})^2}{\sum (x_i - \bar{x})^2}} \]
\[ = (2.1801) \sqrt{1 + \frac{1}{51} + \frac{(3.7 - 5.198)^2}{482.62}} \]
\[ = 2.206 \]
10.36 (cont.)

\[
\hat{y} \pm t\cdot SE_{\hat{y}} \\
\hat{y} = 4.9859
\]

\[
4.9859 \pm (2.009)(2.206)
\]

\[
4.9859 \pm 4.4319
\]

\[
(0.554, 9.4178)
\]

10.37

\[H_0: \beta_1 = 0 \text{ vs. } H_a: \beta_1 \neq 0.\]

\[
t = 6.32
\]

\[
F = 39.9096 \quad \text{from earlier JMP output}
\]

\[
t^2 = F
\]

\[
p-value < .0001
\]
Analysis of Variance

<table>
<thead>
<tr>
<th>Source</th>
<th>DF</th>
<th>Sum of Squares</th>
<th>Mean Square</th>
<th>F Ratio</th>
<th>Prob &gt; F</th>
</tr>
</thead>
<tbody>
<tr>
<td>Model</td>
<td>1</td>
<td>189.69399</td>
<td>189.694</td>
<td>39.9096</td>
<td>&lt;.0001</td>
</tr>
<tr>
<td>Error</td>
<td>49</td>
<td>232.90169</td>
<td>4.753</td>
<td></td>
<td></td>
</tr>
<tr>
<td>C. Total</td>
<td>50</td>
<td>422.59568</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

(a) ANOVA Equation

\[ \text{SS}_{\text{Total}} = \text{SS}_{\text{Model}} + \text{SS}_{\text{Error}} \]
\[ = 189.69399 + 232.90169 \]
\[ = 422.59568 \]

(b) Total df = Regression df + Residual df

\[ n-1 = 1 + n-2 \]
\[ 50 = 1 + (51-2) \]
\[ = 1 + 49 \]
\[ = 50 \]

(c) \[ \text{MS}_{\text{Res}} = \frac{232.90169}{49} = 4.753096 \]

\[ \text{MS}_{\text{Reg}} = \frac{189.69399}{1} = 189.69399 \]

(d) \[ F = \frac{\text{MS}_{\text{Reg}}}{\text{MS}_{\text{Res}}} = \frac{189.69399}{4.753096} = 39.9096 \]
10.39

(a) \( R^2 = \frac{SS_{\text{reg}}}{SS_{\text{total}}} = \frac{189.69399}{422.59567} = .4489 \)

(b) \( S = \sqrt{MS_{\text{res}}} = \sqrt{4.753} = 2.1801 \)
2. (a)

Bivariate Fit of $y$ By $x$

\[ y = 2706.9613 + 372.28517x \]

\[ r = +\sqrt{0.637893} = +0.80 \]

\[ y = 3376.7726 + 312.35408x \]

\[ r = +\sqrt{0.281} = +0.53 \]

The correlation drops when two data points with the largest $x$ values are excluded. Note that these two points (circled in the scatterplot) have $x$ value above $\bar{x}$ and $y$ value above $\bar{y}$. So we would expect a higher correlation when the points are included in the analysis.
2. (b) 
\[ y = 2706.968B + 372.28517 (5) \]
\[ = 4568.39 \]

c. There appears to be nonconstant variance. As \( x \) increases, variation in \( y \) tends to increase. But the SLR model must satisfy the assumption of constant variance.
The nonconstant variance appears to be fixed. We observe on data point (circled) with a large residual. Under the change of scale, a SLR model seems quite appropriate.

(e) \[ \hat{y}' = 24.57183 + 16.832139 \times' \]
\[ R^2 = 0.614332 \]
Hence, 61.4% of the raw variation in \( y' \) is accounted for using an equation linear in \( x' \).

(f) \( x = 5 \Rightarrow x' = 2.2361 \)
\[ \hat{y}' = 24.57183 + 16.832139(2.2361) = 62.21 \]
\[ \hat{y} = (\hat{y}')^2 = (62.21)^2 = 3870.10 \]

(g) Std error: 23.88 (from above JMP output)
This measures variation in root reimbursed hospital cost for a fixed length of stay.
\[ df = 33 - 2 = 31 \quad L = \chi^2_{31, 0.975} = 17.539 \quad U = \chi^2_{31, 0.025} = 48.232 \]
95% CI for \( \delta \): \[ \left( \frac{S.E}{\sqrt{n-2}}, \frac{S.E}{\sqrt{n-2}} \right) = (19.14, 31.75) \]
(h) Parameter Estimates

| Term     | Estimate | Std Error | t Ratio | Prob>|t| |
|----------|----------|-----------|---------|-----|
| Intercept| 24.57183 | 9.6986527 | 2.53   | 0.0166 |
| x        | 16.832138| 2.395323  | 7.03   | <0.0001 |

\[ SE_{b_1} = 2.395323 \]

\[ S = 23.88 \text{ from part (g)} \]

\[ \sum (x_i - \bar{x})^2 = (n-1)S^2 = (32)(1.7622)^2 = 99.3712 \]

\[ SE_{b_1} = \frac{S}{\sqrt{\sum (x_i - \bar{x})^2}} = \frac{23.88}{\sqrt{99.3712}} = 2.3955 \]

(i) 95% CI for \( \beta_1 \)

\[ b_1 \pm t_{(31, 0.975)} SE_{b_1} = 16.83 \pm (2.040)(2.395) \]

\[ 14.83 \pm 4.89 \]

\[ (11.94, 21.72) \]

The 95% CI does not contain zero.

Hence reject \( H_0 : \beta_1 = 0 \) and conclude \( \beta_1 \neq 0 \).

(j) 95% CI for \( \mu_{y|x} \)

\[ \hat{y} \pm t_{(31, 0.975)} SE_{\hat{y}} \]

\[ 62.21 \pm (2.040)(23.88) \sqrt{\frac{1}{33} + \frac{(2.2861 - 3.66)^2}{99.3712}} \]

\[ 62.21 \pm 10.97 \]

\[ (51.24, 73.18) \]

95% CI for median reimbursed expense \( x = 5 \) is \( (26.25, 54.5363) \).
95% prediction Interval for \( y' \) when \( x = 5 \):

\[
\hat{y} \pm t(31, 975) \times \sqrt{1 + \frac{1}{n} + \frac{(x^* - \bar{x})^2}{\sum(x_i - \bar{x})^2}}
\]

\[
62.21 \pm (2.040)(23.88)\sqrt{1 + \frac{1}{33} + \frac{(2.2861 - 3.66)^2}{99.8712}}
\]

\[
62.21 \pm 49.94
\]

(12.27, 112.15)  
See output(***)

95% PI for \( y \) when \( x = 5 \)

(150.55, 12577.62)

(\( x = 365 \) is outside the range of the data set.
Hence, this is extrapolation.

PI for \( y \) when \( x = 365 \)

(255.94^2, 436.86^2) = (155505, 190410)

See output (***)

This is not useful from a practical standpoint since the interval is so wide.
<table>
<thead>
<tr>
<th>x</th>
<th>y</th>
<th>x'</th>
<th>y'</th>
<th>Lower 95% Mean y'</th>
<th>Upper 95% Mean y'</th>
<th>Lower 95% Individual y'</th>
<th>Upper 95% Individual y'</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.25</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
</tr>
<tr>
<td>0.50</td>
<td>0.25</td>
<td>0.25</td>
<td>0.25</td>
<td>0.25</td>
<td>0.25</td>
<td>0.25</td>
<td>0.25</td>
</tr>
</tbody>
</table>

**Note:** The values in the table are placeholders for demonstration purposes. Actual values should be used in the analysis.
Bivariate Fit of y By x

Transformed Fit Sqrt to Sqrt

\[ \sqrt{y} = 24.57183 + 16.832138 \sqrt{x} \]

**Summary of Fit**

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>RSquare</td>
<td>0.614332</td>
</tr>
<tr>
<td>RSquare Adj</td>
<td>0.601891</td>
</tr>
<tr>
<td>Root Mean Square Error</td>
<td>23.87776</td>
</tr>
<tr>
<td>Mean of Response</td>
<td>86.14779</td>
</tr>
<tr>
<td>Observations (or Sum Wgts)</td>
<td>33</td>
</tr>
</tbody>
</table>

**Analysis of Variance**

<table>
<thead>
<tr>
<th>Source</th>
<th>DF</th>
<th>Sum of Squares</th>
<th>Mean Square</th>
<th>F Ratio</th>
<th>Prob &gt; F</th>
</tr>
</thead>
<tbody>
<tr>
<td>Model</td>
<td>1</td>
<td>28153.839</td>
<td>28153.8</td>
<td>49.3799</td>
<td></td>
</tr>
<tr>
<td>Error</td>
<td>31</td>
<td>17672.572</td>
<td>570.1</td>
<td>1.1708</td>
<td>&lt;.0001</td>
</tr>
<tr>
<td>C. Total</td>
<td>32</td>
<td>45828.411</td>
<td>1200.3</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**Parameter Estimates**

| Term      | Estimate | Std Error | t Ratio | Prob>|t| |
|-----------|----------|-----------|---------|------|
| Intercept | 24.57183 | 9.898527  | 2.53    | 0.0166 |
| Sqrt(x)   | 16.832138| 2.395323  | 7.03    | <.0001 |

**Fit Measured on Original Scale**

- Sum of Squared Error: 723382931
- Root Mean Square Error: 4830.6245
- RSquare: 0.6354902
- Sum of Residuals: 17674.572

(n) See $x$

I would further investigate since $40,000 is considerably above the prediction limits for $x = 30$. 