9.14

a. 

<table>
<thead>
<tr>
<th>$x_i$</th>
<th>$y_i$</th>
<th>$x_i^2$</th>
<th>$x_iy_i$</th>
</tr>
</thead>
<tbody>
<tr>
<td>7</td>
<td>2</td>
<td>7² = 49</td>
<td>7(2) = 14</td>
</tr>
<tr>
<td>4</td>
<td>4</td>
<td>4² = 16</td>
<td>4(4) = 16</td>
</tr>
<tr>
<td>6</td>
<td>2</td>
<td>6² = 36</td>
<td>6(2) = 12</td>
</tr>
<tr>
<td>2</td>
<td>5</td>
<td>2² = 4</td>
<td>2(5) = 10</td>
</tr>
<tr>
<td>1</td>
<td>7</td>
<td>1² = 1</td>
<td>1(7) = 7</td>
</tr>
<tr>
<td>1</td>
<td>6</td>
<td>1² = 1</td>
<td>1(6) = 6</td>
</tr>
<tr>
<td>3</td>
<td>5</td>
<td>3² = 9</td>
<td>3(5) = 15</td>
</tr>
</tbody>
</table>

Totals: 
\[ \sum x_i = 7 + 4 + 6 + 2 + 1 + 1 + 3 = 24 \]
\[ \sum y_i = 2 + 4 + 2 + 5 + 1 + 6 + 5 = 31 \]
\[ \sum x_i^2 = 49 + 16 + 36 + 4 + 1 + 9 = 116 \]
\[ \sum x_iy_i = 14 + 16 + 12 + 10 + 7 + 6 + 15 = 80 \]

b. 

\[ SS_w = \sum x_iy_i = \frac{(\sum x_i)(\sum y_i)}{n} = 80 - \frac{(24)(31)}{7} = 80 - 106.2857143 = -26.2857143 \]

c. 

\[ SS_w = \sum x_i^2 = \frac{\sum x_i^2}{n} = 116 - \frac{(24)^2}{7} = 82.28571429 = 33.71428571 \]

d. 

\[ \hat{\beta}_1 = \frac{SS_w}{SS_w} = -\frac{-26.2857143}{33.71428571} = -.797770117 = -.7797 \]

e. 

\[ \bar{x} = \frac{\sum x_i}{n} = 3.428571429 \quad \bar{y} = \frac{\sum y_i}{n} = 31 \quad \frac{1}{7} = 4.428571429 \]

f. 

\[ \hat{\beta}_0 = \bar{y} - \hat{\beta}_1 \bar{x} = 4.428571429 - (-.7797)(3.428571429) = 4.428571429 - (-2.673123487) = 7.101694916 \approx 7.102 \]

g. 

The least squares line is \( \hat{y} = \hat{\beta}_0 + \hat{\beta}_1 x = 7.102 - .7797x \).
b. Choose \( y = 1 + x \) since it best describes the relation of \( x \) and \( y \).

c. 

<table>
<thead>
<tr>
<th>( y )</th>
<th>( x )</th>
<th>( y = 1 + x )</th>
<th>( y - \hat{y} )</th>
</tr>
</thead>
<tbody>
<tr>
<td>2</td>
<td>.5</td>
<td>1.5</td>
<td>2 - 1.5 = .5</td>
</tr>
<tr>
<td>1</td>
<td>1.0</td>
<td>2.0</td>
<td>1 - 2.0 = -1.0</td>
</tr>
<tr>
<td>3</td>
<td>1.5</td>
<td>2.5</td>
<td>3 - 2.5 = .5</td>
</tr>
</tbody>
</table>

Sum of errors = 0

<table>
<thead>
<tr>
<th>( y )</th>
<th>( x )</th>
<th>( y = 3 - x )</th>
<th>( y - \hat{y} )</th>
</tr>
</thead>
<tbody>
<tr>
<td>2</td>
<td>.5</td>
<td>3 - .5 = 2.5</td>
<td>2 - 2.5 = -.5</td>
</tr>
<tr>
<td>1</td>
<td>1.0</td>
<td>3 - 1.0 = 2.0</td>
<td>1 - 2.0 = -1.0</td>
</tr>
<tr>
<td>3</td>
<td>1.5</td>
<td>3 - 1.5 = 1.5</td>
<td>3 - 1.5 = 1.5</td>
</tr>
</tbody>
</table>

Sum of errors = 0

d. \( \text{SSE} = \sum (y - \hat{y})^2 \)

SSE for 1st model: \( y = 1 + x \), \( \text{SSE} = (.5)^2 + (-1)^2 + (.5)^2 = 1.5 \)

SSE for 2nd model: \( y = 3 - x \), \( \text{SSE} = (-.5)^2 + (-1)^2 + (1.5)^2 = 3.5 \)

The best fitting straight line is the one that has the smallest least squares. The model \( y = 1 + x \) has a smaller SSE, and therefore it verifies the visual check in part a.

e. Some preliminary calculations are:

\[
\sum x_i = 3 \quad \sum y_i = 6 \quad \sum x_i y_i = 6.5 \quad \sum x_i^2 = 3.5
\]

\[
\text{SS}_w = \sum (x_i - \bar{x})^2 = 3.5 - \frac{(3)(6)}{3} = .5
\]

\[
\text{SS}_x = \frac{\sum x_i^2}{n} - \frac{(\sum x_i)^2}{n} = 3.5 - \frac{(3)^2}{3} = .5
\]

\[
\hat{\beta}_1 = \frac{\sum x_i y_i}{n \sum x_i^2} = \frac{6.5}{3 \cdot 3.5} = .5
\]

\[
\hat{\beta}_0 = \bar{y} - \hat{\beta}_1 \bar{x} = 2 - 1(1) = 1 \Rightarrow \bar{y} - \hat{\beta}_0 \bar{x} = 1 + 1
\]

The least squares line is the same as the second line given.

9.20

a. It appears that there is a positive linear trend. As the year of birth increases, the Z12-note entropy tends to increase.

b. The slope of the line is positive. As the year of birth increases, the Z12-note entropy tends to increase.

c. The line shown is the least squares line – it is the best line through the sample points. We do not know the values of \( \beta_0 \) and \( \beta_1 \) so we do not know the true line of means.
From the printout, the least squares prediction equation is \( \hat{y} = 295.25 - 16.364x \).

b. Using MINITAB, the fitted regression plot and scatterplot are:

Since the data are fairly close the least squares prediction line, the line is a good predictor of annual rainfall.

c. From the printout, the least squares prediction equation is \( \hat{y} = 10.52 + 0.016x \)

Using MINITAB, the fitted regression plot and scatterplot are:

Since the data are not close to the least squares prediction line, the line is not a good predictor of ant species.
9.34 The graph in b would have the smallest \( r^2 \) because the width of the data points is the smallest.

9.36 a. \( r^2 = \frac{\text{SSE}}{n-2} = \frac{.429}{12-2} = .0429 \)

b. \( s = \sqrt{r^2} = \sqrt{.0429} = .2071 \)

c. We would expect most of the observations to be within 2s of the least squares line. This is:
\[ 2s = 2\sqrt{.0429} = .414 \]

9.52 a.

\[ \sum x = 23 \quad \sum x^2 = 111 \quad \sum xy = 81 \]

\[ \sum y = 18 \quad \sum y^2 = 62 \]

\[ \text{SS}_{xy} = \sum xy - \frac{(\sum x)(\sum y)}{n} = 81 - \frac{23(18)}{7} = 21.85714286 \]

\[ \text{SS}_x = \sum x^2 - \frac{[(\sum x)^2]}{n} = 111 - \frac{23^2}{7} = 35.42857143 \]

\[ \text{SS}_y = \sum y^2 - \frac{[(\sum y)^2]}{n} = 62 - \frac{18^2}{7} = 15.71428571 \]

\[ \hat{\beta} = \frac{\text{SS}_{xy}}{\text{SS}_x} = \frac{21.85714286}{35.42857143} = .616935483 = .617 \]

\[ \hat{\alpha} = \bar{y} - \hat{\beta} \bar{x} = \frac{18}{7} - .616935483 \cdot \frac{23}{7} = .544354838 = .544 \]

The least squares line is \( y = .544 + .617x \)

c. The line is plotted on the graph in a.

d. To determine if \( x \) contributes information for the linear prediction of \( y \), we test:
\[ H_0: \ \beta = 0 \]
\[ H_a: \ \beta \neq 0 \]
e. The test statistic is \( t = \frac{\hat{\beta} - 0}{s} = \frac{.617 - 0}{.6678} = 5.50 \)

where \( SSE = SS_{yw} - \hat{\beta}SS_{xt} = 15.71428571 - .616935483(21.85714286) = 2.2298372 \)

\[ s^2 = \frac{SSE}{n-2} = \frac{2.2298372}{7-2} = .44596774 \]

\[ s = \sqrt{.44596774} = .6678 \]

The degrees of freedom are \( n - 2 = 7 - 2 = 5 \).

f. The rejection region requires \( \alpha/2 = .05/2 = .025 \) in each tail of the \( t \) distribution with \( df = 5 \). From Table IV, Appendix A, \( t_{.025} = 2.571 \). The rejection region is \( t < -2.571 \) or \( t > 2.571 \).

Since the observed value of the test statistic falls in the rejection region (\( t = 5.50 > 2.571 \)), \( H_0 \) is rejected. There is sufficient evidence to indicate \( x \) contributes information for the linear prediction of \( y \) at \( \alpha = .05 \).

g. For confidence coefficient .95, \( \alpha = 1 - .95 = .05 \) and \( \alpha/2 = .05/2 = .025 \). From Table IV, Appendix A, with \( df = n - 2 = 7 - 2 = 5, t_{.025} = 2.571 \). The 95% confidence interval is:

\[ \hat{\beta} \pm t_{.025} \frac{s}{\sqrt{SS_{xt}}} \Rightarrow \hat{\beta} \pm t_{.025} \frac{.6678}{\sqrt{35.42857143}} \]

\[ \Rightarrow \hat{\beta} \pm .288 \Rightarrow (.329, .905) \]

9.60 Some preliminary calculations are:

\[ \bar{y} = \frac{\sum y}{n} = \frac{78.8}{16} = 4.925 \]

\[ \bar{x} = \frac{\sum x}{n} = \frac{247}{16} = 15.4375 \]

\[ SS_{xt} = \sum xy - \left( \frac{\sum x \sum y}{n} \right) = 1264.6 - \frac{247(78.8)}{16} = 48.125 \]

\[ SS_{xt} = \sum x^2 - \left( \frac{\sum x \sum y}{n} \right) = 4193 - \frac{247^2}{16} = 379.9375 \]

\[ \hat{\beta} = \frac{SS_{yt}}{SS_{xt}} = \frac{48.125}{379.9375} = .12666557 \]

\[ \hat{\beta}_b = \bar{y} - \hat{\beta}\bar{x} = \frac{78.8}{16} - (.12666557) \left( \frac{247}{16} \right) = 2.969600263 \]

\[ SS_{yw} = \sum y^2 - \left( \frac{\sum y \sum y}{n} \right) = 406.84 - \frac{78.8^2}{16} = 18.75 \]
To determine whether blood lactate level is linearly related to perceived recovery, we test:

\[ H_0: \beta = 0 \]
\[ H_a: \beta \neq 0 \]

The test statistic is

\[ t = \frac{\hat{\beta}_1 - 0}{s / \sqrt{SS_{\epsilon}}} = \frac{12667 - 0}{\sqrt{SS_{\epsilon}}} = 2.597 \]

The rejection region requires \( \alpha^2 = 0.10^2 = 0.05 \) in each tail of the t distribution. From Table IV, Appendix A, with \( df = n - 2 = 16 - 2 = 14 \), \( t_{0.05} = 1.761 \). The rejection region is \( t < -1.761 \) or \( t > 1.761 \).

Since the observed test statistic falls in the rejection region (\( t = 2.597 > 1.761 \)), \( H_0 \) is rejected. There is sufficient evidence to indicate blood lactate level is linearly related to perceived recovery at \( \alpha = 0.10 \).

9.92 a. If a jeweler wants to predict the selling price of a diamond stone based on its size, he would use a prediction interval for \( y \).

b. If a psychologist wants to estimate the average IQ of all patients that have a certain income level, he would use a confidence interval for \( E(y) \).
b. Some preliminary calculations are:

\[
\begin{align*}
\sum x_i &= 28 & \sum x_i^2 &= 140 & \sum x_i y_i &= 196 \\
\sum y_i &= 42 & \sum y_i^2 &= 284
\end{align*}
\]

\[
\begin{align*}
SS_{xy} &= \sum x_i y_i - \frac{\sum x_i \sum y_i}{n} = 196 - \frac{28(42)}{7} = 28 \\
SS_{xx} &= \sum x_i^2 - \left(\frac{\sum x_i}{n}\right)^2 = 140 - \frac{28^2}{7} = 28 \\
SS_{yy} &= \sum y_i^2 - \left(\frac{\sum y_i}{n}\right)^2 = 284 - \frac{42^2}{7} = 32
\end{align*}
\]

\[
\hat{\beta}_0 = \frac{SS_{xy}}{SS_{xx}} = \frac{28}{28} = 1, \quad \hat{\beta}_1 = \frac{\sum (x_i - \bar{x})(y_i - \bar{y})}{\sum (x_i - \bar{x})^2} = 4 \frac{28}{7} = 6 - 4 = 2
\]

The least squares line is \( \hat{y} = 2 + x \).

c. \( SSE = SS_{xy} - \hat{\beta}_1 SS_{xx} = 32 - 1(28) = 4 \)

\[
\hat{s}^2 = \frac{SSE}{n-2} = \frac{4}{5} = .8
\]

d. The form of the confidence interval is \( \hat{y} \pm t_{\alpha/2} \frac{\hat{s}}{\sqrt{n}} \left( \frac{1}{n} + \frac{(x_p - \bar{x})^2}{SS_{xx}} \right) \)

where \( s = \sqrt{\hat{s}^2} = \sqrt{.8} = .8944 \). For \( x_p = 4 \), \( \hat{y} = 2 + (4) = 6 \), and \( \bar{x} = \frac{28}{7} = 4 \).

For confidence coefficient .90, \( \alpha = 1 - .90 = .10 \) and \( \alpha/2 = .10/2 = .05 \). From Table IV, Appendix A, \( t_{.05} = 2.015 \) with \( df = n - 2 = 7 - 2 = 5 \).

The 90% confidence interval is:

\[
6 \pm 2.015(.8944) \sqrt{\frac{1}{7} + \frac{(4-4)^2}{28}} \Rightarrow 6 \pm .681 \Rightarrow (5.319, 6.681)
\]

e. The form of the prediction interval is \( \hat{y} \pm t_{\alpha/2} \hat{s} \sqrt{\frac{1}{n} + \frac{(x_p - \bar{x})^2}{SS_{xx}}} \)

The 90% prediction interval is:

\[
6 \pm 2.015(.8944) \sqrt{\frac{1}{7} + \frac{(4-4)^2}{28}} \Rightarrow 6 \pm 1.927 \Rightarrow (4.073, 7.927)
\]

f. The 95% prediction interval for \( y \) is wider than the 95% confidence interval for the mean value of \( y \) when \( x_p = 4 \).

The error of predicting a particular value of \( y \) will be larger than the error of estimating the mean value of \( y \) for a particular \( x \) value. This is true since the error in estimating the mean value of \( y \) for a given \( x \) value is the distance between the least squares line and the true line of means, while the error in predicting some future value of \( y \) is the sum of two errors—the error of estimating the mean of \( y \) plus the random error that is a component of the value of \( y \) to be predicted.