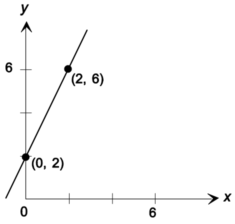
Chapter **3**

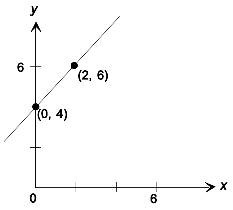
Simple Linear Regression

3.1

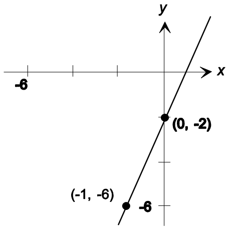
a.



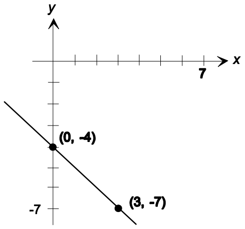
b.



c.



d.



3.2 Since the line passes through the point (0, 1), $1=\beta_0+\beta_1(0)\Rightarrow\beta_0=1$.
Also, since it also passes through the point (2, 3),

$$3=\beta_0+\beta_1(2)\Rightarrow 3=1+2\beta_1\Rightarrow\beta_1=1\Rightarrow y=1+x$$

3.3 a. Using the technique explained in Exercise 3.2:

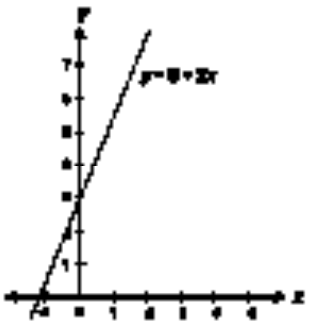
$$\left.\begin{aligned}2&=\beta_0+\beta_1(0)\\ 6&=\beta_0+\beta_1(2)\end{aligned}\right\}\Rightarrow\left.\begin{aligned}\beta_0&=2\\ \beta_1&=2\end{aligned}\right\}\Rightarrow y=2+2x$$

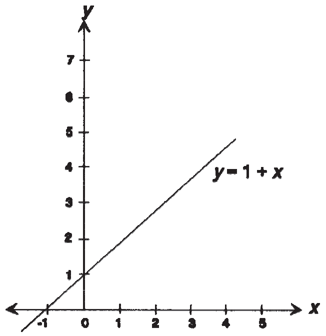
b.
$$\left.\begin{aligned}4&=\beta_0+\beta_1(0)\\ 6&=\beta_0+\beta_1(2)\end{aligned}\right\}\Rightarrow\left.\begin{aligned}\beta_0&=4\\ \beta_1&=1\end{aligned}\right\}\Rightarrow y=4+x$$

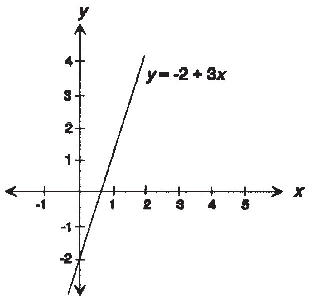
c.
$$\left.\begin{aligned}-2&=\beta_0+\beta_1(0)\\ -6&=\beta_0+\beta_1(-1)\end{aligned}\right\}\Rightarrow\left.\begin{aligned}\beta_0&=-2\\ \beta_1&=4\end{aligned}\right\}\Rightarrow y=-2+4x$$

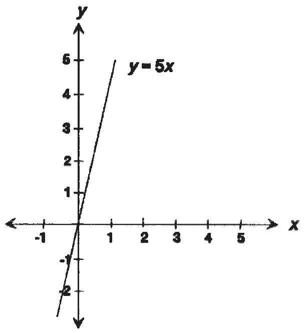
d.
$$\left.\begin{aligned}-4&=\beta_0+\beta_1(0)\\ -7&=\beta_0+\beta_1(3)\end{aligned}\right\}\Rightarrow\left.\begin{aligned}\beta_0&=-4\\ \beta_1&=-1\end{aligned}\right\}\Rightarrow y=-4-x$$

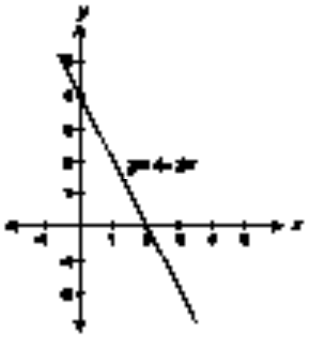
3.4

a. 

b. 

c. 

d. 

e. 

3.5

	Slope (β_1)	***y*-intercept** (β_0)
a.	2	3
b.	1	1
c.	3	−2
d.	5	0
e.	−2	4

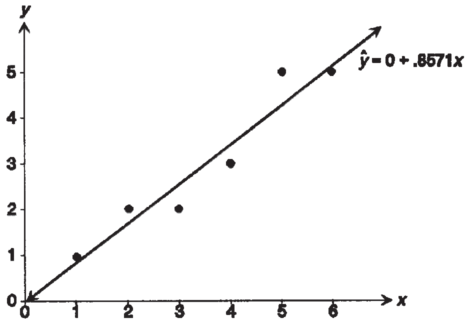
3.6 Some preliminary calculations are:

$$\sum x = 21 \qquad \sum x^2 = 91 \qquad \bar{x} = \frac{21}{6} = 3.5$$

$$\sum y = 18 \qquad \sum y^2 = 68 \qquad \bar{y} = \frac{18}{6} = 3 \qquad \sum xy = 78$$

a.
$$SS_{xx} = \sum x^2 - n\bar{x}^2 = 91 - 6(3.5)^2 = 17.5$$
$$SS_{xy} = \sum xy - n\bar{x}\bar{y} = 78 - 6(3.5)(3) = 15$$
$$SS_{yy} = \sum y^2 - n\bar{y}^2 = 68 - 6(3)^2 = 14$$
$$\hat{\beta}_1 = \frac{SS_{xy}}{SS_{xx}} = \frac{15}{17.5} = 0.8571$$
$$\hat{\beta}_0 = \bar{y} - \hat{\beta}_1\bar{x} = 3 - (0.8571)(3.5) = 0$$

b.



3.7 a. To compute $\hat{\beta}_0$ and $\hat{\beta}_1$, we first construct the following table:

x	y	xy	x^2	y^2
−2	4	−8	4	16
−1	3	−3	1	9
0	3	0	0	9
1	1	1	1	1
2	−1	−2	4	1
$\sum x = 0$	$\sum y = 10$	$\sum xy = -12$	$\sum x^2 = 10$	$\sum y^2 = 36$

Then,

$$SS_{xx} = \sum x^2 - \frac{\left(\sum x\right)^2}{n} = 10 - \frac{(0)^2}{5} = 10$$

$$SS_{xy} = \sum xy - \frac{\left(\sum x\right)\left(\sum y\right)}{n} = -12 - \frac{0(10)}{5} = -12$$

$$SS_{yy} = \sum y^2 - \frac{\left(\sum y\right)^2}{n} = 36 - \frac{(10)^2}{5} = 16$$

$$\bar{y} = \frac{\sum y}{n} = \frac{10}{5} = 2 \qquad \bar{x} = \frac{\sum x}{n} = \frac{0}{5} = 0$$

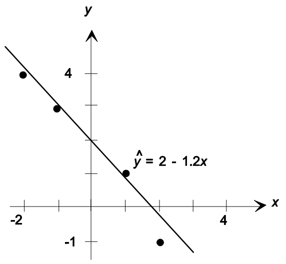
Thus, the least squares estimates of β_0 and β_1 are:

$$\hat{\beta}_1 = \frac{SS_{xy}}{SS_{xx}} = \frac{-12}{10} = -1.2$$

$$\hat{\beta}_0 = \bar{y} - \hat{\beta}_1\bar{x} = 2 - (-1.2)(0) = 2$$

and the equation of the least squares prediction line is $\hat{y} = 2 - 1.2x$.

b.



3.8 a. $y = \beta_0 + \beta_1 x + \varepsilon$

b. Yes, since the data appears to demonstrate a straight-line relationship.

c. $\text{Sales\_Price} = 1.4 + 1.41\,\text{Market\_Val}$

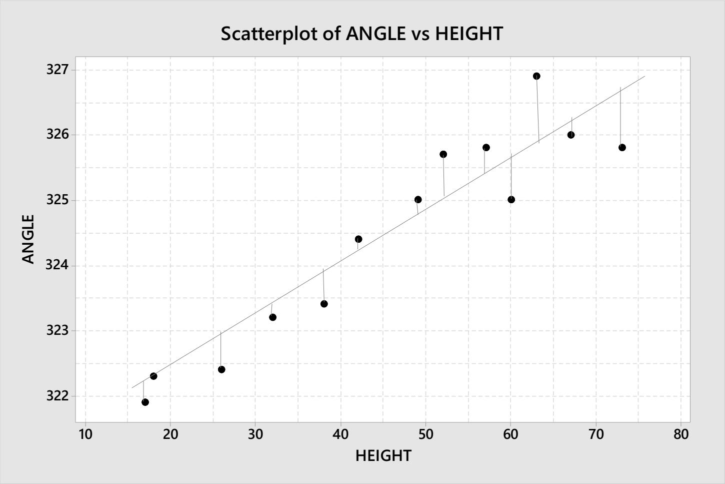
d. $\hat{\beta}_0 = 1.4$, when $x = 0$ (no market value), then the sales price has no practical meaning.

e. Various answers possible. A possible answer for the range on which the slope is $\$100{,}000 < x < \$1{,}000{,}000$.

f. "mean sale price" $= 1.4 + 1.41(\$300{,}000) \approx \$423{,}000$

3.9 a. Yes, there appears to be a positive linear trend. As the height above the horizon increases, the angular size tends to increase.

b & c. A sketch (answers can vary) of the line with lines drawn to the sketch line is:



The estimated deviations and squared deviations are:

ANGLE	HEIGHT	Est Fit	Dev	Sq Dev
321.9	17	322.2	-0.3	0.09
322.3	18	322.3	0.0	0.00
322.4	26	323.0	-0.6	0.36
323.2	32	323.4	-0.2	0.04
323.4	38	323.9	-0.5	0.25
324.4	42	324.2	0.2	0.04
325.0	49	324.8	0.2	0.04
325.7	52	325.0	0.7	0.49
325.8	57	325.4	0.4	0.16
325.0	60	325.7	-0.7	0.49
326.9	63	325.9	1.0	1.00
326.0	67	326.2	-0.2	0.04
325.8	73	326.7	-0.9	0.81
				3.81

The sum of the squared deviations is 3.81.

d. From the sketched line, the y-intercept is about 321 and the slope is about 0.1. These are close to the y-intercept, 320.636, and slope, 0.083, of the regression line.

e. From the printout, the SSE is 3.56465. The sum of squares from the estimated line is 3.81. The SSE from the regression line is smaller.

3.10 a. Using MINITAB, the results are:

Analysis of Variance

Source	DF	Adj SS	Adj MS	F-Value	P-Value
Regression	1	677.45	677.45	24.41	0.003
VO2Max	1	677.45	677.45	24.41	0.003
Error	6	166.55	27.76		
Lack-of-Fit	5	142.05	28.41	1.16	0.604
Pure Error	1	24.50	24.50		
Total	7	844.00			

Coefficients

Term	Coef	SE Coef	T-Value	P-Value	VIF
Constant	-27.2	19.8	-1.38	0.217	
VO2Max	0.558	0.113	4.94	0.003	1.00

Regression Equation

HR% = -27.2 + 0.558 VO2Max

The least squares line is $\hat{y} = -27.2 + 0.558x$.

b. Since 0 is not in the range of observed values of VO2Max, the y-intercept does not have a practical interpretation.

c. $\hat{\beta}_1 = 0.558$ For each unit increase in the value of VO2Max, the mean HR% is estimated to increase by 0.558.

3.11 a. No, there does not appear to any trend for cooperation use versus the average payoff.

b. No, there does not appear to any trend for defective use versus the average payoff.

c. Yes, there appears to be somewhat of a linear relationship for average payoff and punishment use.

d. Negative relationship; the more punishment use, the average payoff decreases.

e. Yes, winners tend to punish less than non-winners.

3.12 a. Using MINITAB, some calculations are:

Analysis of Variance

Source	DF	Adj SS	Adj MS	F-Value	P-Value
Regression	1	6083.84	6083.84	26.35	0.000
Year	1	6083.84	6083.84	26.35	0.000
Error	10	2309.07	230.91		
Lack-of-Fit	9	2301.07	255.67	31.96	0.136
Pure Error	1	8.00	8.00		
Total	11	8392.92			

```
Model Summary
      S     R-sq   R-sq(adj)   R-sq(pred)
15.1956   72.49%     69.74%       54.61%
```

```
Coefficients
Term        Coef   SE Coef   T-Value   P-Value    VIF
Constant   -3675       724     -5.08     0.000
Year       1.870     0.364      5.13     0.000   1.00
```

```
Regression Equation
Cost   =   -3675 + 1.870 Year
```

The least squares line is $\hat{y} = -3675 + 1.870x$.

b. Since 0 is not in the range of observed values of Year, the y-intercept does not have a practical interpretation.

c. $\hat{\beta}_1 = 1.87$ For each unit increase in cost, the mean cost is estimated to increase by 1.87 million dollars.

3.13 a. Some preliminary calculations are:

$$\sum x = 6167 \qquad \sum y = 135.8 \qquad n = 24$$
$$\sum x^2 = 1{,}641{,}115 \qquad \sum y^2 = 769.72 \qquad \sum xy = 34{,}765$$

$$SS_{xy} = \sum xy - \frac{\left(\sum x\right)\left(\sum y\right)}{n} = 34{,}765 - \frac{(6167)(135.8)}{24} = -129.94167$$

$$SS_{xx} = \sum x^2 - \frac{\left(\sum x\right)^2}{n} = 164{,}115 - \frac{(6167)^2}{24} = 56{,}452.958$$

$$\hat{\beta}_1 = \frac{SS_{xy}}{SS_{xx}} = \frac{-129.94167}{56{,}452.958} = -0.002301769 \cong -0.0023$$

$$\hat{\beta}_0 = \bar{y} - \hat{\beta}_1\bar{x} = \frac{135.8}{24} - (-0.002301769)\left(\frac{6167}{24}\right) = 6.249792065 \cong 6.251$$

The least squares line is $\hat{y} = 6.25 - 0.0023x$.

b. $\hat{\beta}_0 = 6.25$ Since $x = 0$ is not in the observed range, $\hat{\beta}_0$ has no interpretation other than being the y-intercept.

$\hat{\beta}_1 = -0.0023$. For each additional increase of 1 part per million of pectin, the mean sweetness index is estimated to decrease by 0.0023.

c. $\hat{y} = 6.25 - 0.0023(300) = 5.56.$

3.14 a. Using MINITAB, some preliminary results are:

Analysis of Variance

Source	DF	Adj SS	Adj MS	F-Value	P-Value
Regression	1	9.08	9.080	0.18	0.676
CDIFF	1	9.08	9.080	0.18	0.676
Error	22	1116.78	50.763		
Lack-of-Fit	21	1026.06	48.860	0.54	0.813
Pure Error	1	90.72	90.720		
Total	23	1125.86			

Coefficients

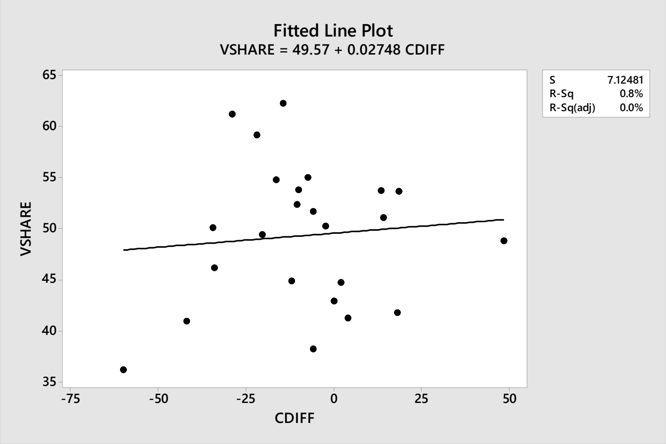
Term	Coef	SE Coef	T-Value	P-Value	VIF
Constant	49.57	1.56	31.76	0.000	
CDIFF	0.0275	0.0650	0.42	0.676	1.00

Regression Equation

VSHARE = 49.57 + 0.0275 CDIFF

The least squares line is $\hat{y} = 49.57 + 0.0275x.$

b. Using MINITAB, the scatterplot is:



There does not appear to be much of a linear relationship between Democratic vote share and charisma difference. There might be a slight positive linear trend.

c. $\hat{\beta}_1 = 0.0275$ For each unit increase in charisma difference, the mean Democratic vote share is estimated to increase by 0.0275 points.

3.15 Some preliminary calculations are:

$$\bar{y} = \frac{\sum x}{n} = \frac{103.07}{144} = 0.71576 \qquad \bar{x} = \frac{\sum y}{n} = \frac{792}{144} = 5.5$$

$$SS_{xy} = \sum xy - \frac{\sum x \sum y}{n} = 586.86 - \frac{792(103.07)}{144} = 19.975$$

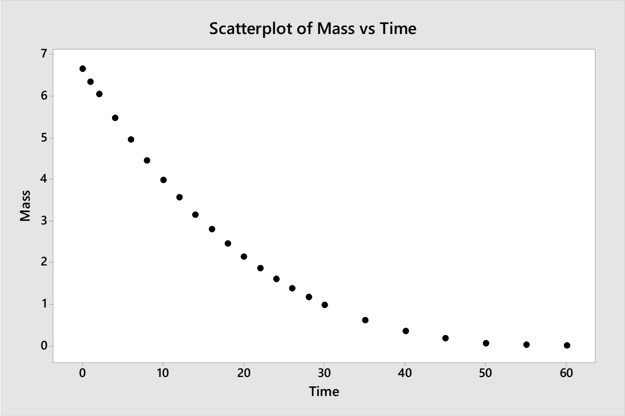
$$SS_{xx} = \sum x^2 - \frac{\left(\sum x\right)^2}{n} = 5{,}112 - \frac{792^2}{144} = 756$$

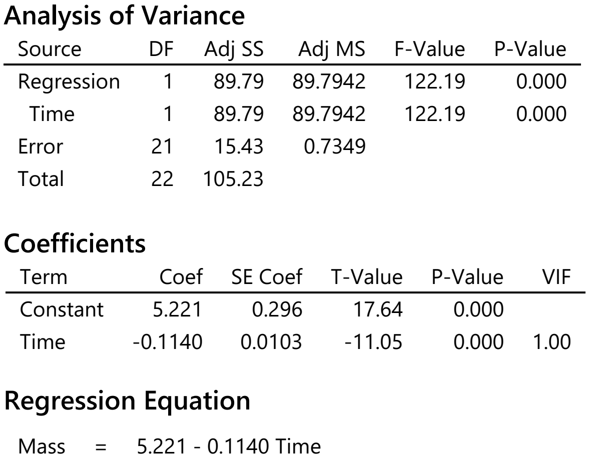
$$\hat{\beta}_1 = \frac{SS_{xy}}{SS_{xx}} = \frac{19.975}{756} = 0.026421957$$

$$\hat{\beta}_o = \overline{y} - \hat{\beta}_1 \overline{x} = \frac{103.07}{144} - (0.026421957)\left(\frac{792}{144}\right) = 0.570443121$$

The estimated regression line is $\hat{y} = 0.5704 + 0.0264x$. Since $x = 0$ is nonsensical, no practical interpretation of $\hat{\beta}_0 = 0.5704$. For each one-position increase in order, estimated recall proportion increases by $\hat{\beta}_1 = 0.0264$.

3.16 The scatterplot in this problem clearly shows a significantly *nonlinear* trend. Therefore, the linear model is not the best to describe the data in this scatter plot.





Analysis of Variance

Source	DF	Adj SS	Adj MS	F-Value	P-Value
Regression	1	89.79	89.7942	122.19	0.000
Time	1	89.79	89.7942	122.19	0.000
Error	21	15.43	0.7349		
Total	22	105.23			

Coefficients

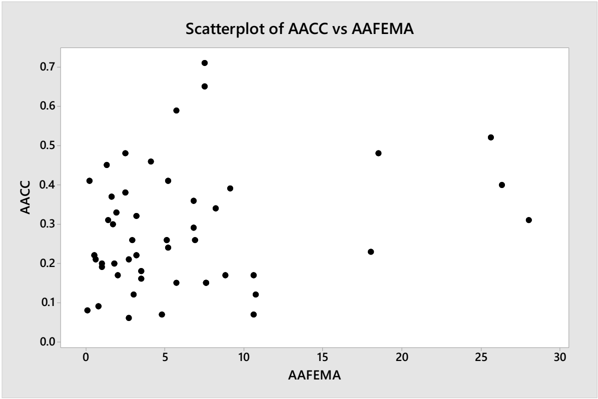
Term	Coef	SE Coef	T-Value	P-Value	VIF
Constant	5.221	0.296	17.64	0.000	
Time	-0.1140	0.0103	-11.05	0.000	1.00

Regression Equation

```
Mass   =   5.221 - 0.1140 Time
```

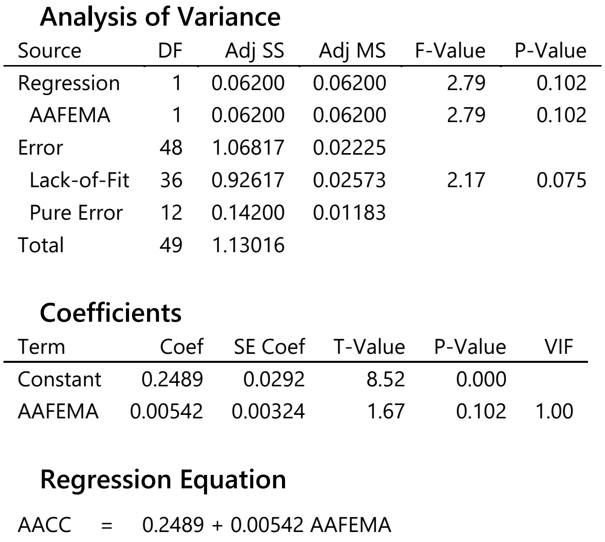
The fitted regression line is $\hat{y} = 5.221 - 0.1140x$. Since the coefficient of time is negative, there is evidence that the mass of the spill tends to decrease as time increases. For each minute increase in time, the mean mass is estimated to diminish by 5.221 pounds.

3.17 a. Using MINITAB, the scatterplot of the data is:



There does not appear to be any apparent trend in the plot.

b. Using MINITAB, the results are:



```
 Analysis of Variance
Source         DF    Adj SS    Adj MS   F-Value   P-Value
Regression      1   0.06200   0.06200      2.79     0.102
  AAFEMA        1   0.06200   0.06200      2.79     0.102
Error          48   1.06817   0.02225
  Lack-of-Fit  36   0.92617   0.02573      2.17     0.075
  Pure Error   12   0.14200   0.01183
Total          49   1.13016

 Coefficients
Term          Coef   SE Coef   T-Value   P-Value    VIF
Constant    0.2489    0.0292      8.52     0.000
AAFEMA     0.00542   0.00324      1.67     0.102   1.00

 Regression Equation
AACC  =  0.2489 + 0.00542 AAFEMA
```

The least squares line is $\hat{y} = 0.2489 + 0.00542x$.

The estimated y-intercept is $\hat{\beta}_0 = 0.2489$ and the estimated slope is $\hat{\beta}_1 = 0.00542$.

c. $\hat{\beta}_0 = 0.2489$ Since 0 is not in the observed range of the average annual FEMA relief, the y-intercept has no practical interpretation.
$\hat{\beta}_1 = 0.00542$ For each unit increase in the average annual FEMA relief, the mean average annual number of public corruption convictions is estimated to increase by 0.00542 per 100,000 residents.

3.18 a. $s^2 = \frac{SSE}{n-2} = \frac{0.219}{9-2} = 0.0313$

b. $s = \sqrt{0.0313} = 0.1769$

3.19 a. Using data from Exercise 3.6,

$SSE = SS_{yy} - \hat{\beta}_1 SS_{xy} = 14 - 0.8751(15) = 1.1435$

$s^2 = \frac{SSE}{n-2} = \frac{1.1435}{6-2} = 0.2859 \qquad s = \sqrt{0.2856} = 0.5347$

b. Using data from Exercise 3.7,

$SSE = SS_{yy} - \hat{\beta}_1 SS_{xy} = 16 - (-1.2)(-12) = 1.6$

$s^2 = \frac{SSE}{n-2} = \frac{1.6}{5-2} = 0.5333 \qquad s = \sqrt{0.5333} = 0.7303$

3.20 a. $s^2 = \frac{SSE}{n-2} = \frac{1.04}{28-2} = 0.04 \qquad s = \sqrt{0.04} = 0.2$

b. We would expect most of the observed value to fall within 2s or $2(0.2) = 0.4$ units of the least squares line.

3.21 a. $y = \beta_0 + \beta_1 x + \varepsilon$

b. The least squares line is $\hat{y} = 120 + 0.3456x$.

c. Assumption 1: The mean of the probability distribution of ε is 0.
Assumption 2: The variance of the probability distribution of ε is constant for all settings of the independent variable x.
Assumption 3: The probability distribution of ε is normal.
Assumption 4: The errors associated with any two different observations are independent.

d. $s = 635.187$

e. $\hat{y} \pm 2s \Rightarrow \hat{y} \pm 2(635.187) \Rightarrow \hat{y} \pm 1270.374$

3.22 a. From Exercise 3.12, $s = 15.1956$.

b. We would expect most of the observed values to fall within 2s or $2(15.1956) = 30.3912$ units of the least squares line.

3.23 a. Using calculations from Exercise 3.13,

$$SS_{yy} = \sum y^2 - \frac{\left(\sum y\right)^2}{n} = 769.72 - \frac{(135.8)^2}{24} = 1.318333$$

$$SSE = SS_{yy} - \hat{\beta}_1 SS_{xy} = 1.318333 - (-0.002301769)(-129.94167) = 1.01924$$

$$s^2 = \frac{SSE}{n-2} = \frac{1.01924}{24-2} = 0.0463 \qquad s = \sqrt{0.0463} = 0.2152$$

b. The units of measure for s^2 are square units. It is very difficult to interpret units such as dollars squared, minutes squared, etc.

c. We would expect most of the observed values to fall within 2s or $2(0.2152) = 0.4304$ units of the least squares line.

3.24 a. The estimate of σ^2 is $s^2 = \frac{SSE}{n-2} = \frac{1.06817}{50-2} = 0.02225$.

b. The estimate of σ is $s = \sqrt{0.02225} = 0.1492$.

c. The estimate of σ can be interpreted practically because it is measured in the same units as the data. The units of measure of σ^2 are square units.

d. We would expect most of the observed values to fall within 2s or $2(0.1492) = 0.2984$ units of the least squares line. In this problem, the units of measure is dollars per capita. However, looking at the scatterplot, the data do not fall close to a straight line. The model will not be very accurate in predicting a state's average annual number of public corruption convictions.

3.25 a. The least squares line with the steepest slope is with the pair AB Magnitude Alert and AB Magnitude No-Tone.

b. The least squares line that produces the largest SSE is with the pair AB Magnitude Alert and AB Magnitude No-Tone.

c. The least squares line that produces the smallest estimate of σ is with the pair AB Magnitude Sim and AB Magnitude Alert.

3.26 a. To determine if β_1 differs from 0, we test:

$$H_0: \beta_1 = 0$$
$$H_a: \beta_1 \neq 0$$

The test statistic is $t = \dfrac{\hat{\beta}_1}{s/\sqrt{SS_{xx}}} = \dfrac{0.8571}{0.5345/\sqrt{17.5}} = 6.71$

The rejection region requires $\alpha / 2 = 0.05 / 2 = 0.025$ in each tail of the t distribution. From Table 2, Appendix D, with $df = n - 2 = 6 - 2 = 4$, $t_{0.025} = 2.776$. The rejection region is $t < -2.776$ or $t > 2.776$.

Since the observed value of the test statistic falls in the rejection region $(t = 6.71 > 2.776)$, H_0 is rejected. There is sufficient evidence to indicate that x contributes information for the prediction of y using a linear model at $\alpha = .05$.

b. To determine if β_1 differs from 0, we test:

$$H_0 : \beta_1 = 0$$
$$H_a : \beta_1 \neq 0$$

The test statistic is $t = \dfrac{\hat{\beta}_1}{s/\sqrt{SS_{xx}}} = \dfrac{-1.2}{0.7303/\sqrt{10}} = -5.20$

The rejection region requires $\alpha / 2 = 0.05 / 2 = 0.025$ in each tail of the t distribution. From Table 2, Appendix D, with $df = n - 2 = 5 - 2 = 3$, $t_{0.025} = 3.182$. The rejection region is $t < -3.182$ or $t > 3.182$.

Since the observed value of the test statistic falls in the rejection region $(t = -5.20 < -3.182)$, H_0 is rejected. There is sufficient evidence to indicate that x contributes information for the prediction of y using a linear model at $\alpha = .05$.

3.27 a. To determine if there is a positive linear relationship between appraised property value and sale price, we test:

$$H_0 : \beta_1 = 0$$
$$H_a : \beta_1 > 0$$

From the printout, the test statistic is $t = 38.132$ and the p-value is $p = 0.000 / 2 = 0.000$. Since the p-value is less than $\alpha\left(p = 0.000 < 0.01\right)$, H_0 is rejected. There is sufficient evidence to indicate there is a positive linear relationship between appraised property value and sale price at $\alpha = 0.01$.

b. From the printout, the 95% confidence interval is $(1.335, 1.482)$. We are 95% confident that for each \$1000 increase in market value, the mean sale price is estimated to increase by from \$1,335 to \$1,482.

c. In order to obtain a narrower confidence interval, one could lower the confidence level (i.e. to 90%) or increase the sample size.

3.28 Some preliminary calculations are:

$$SS_{yy} = \sum y^2 - \frac{\left(\sum y\right)^2}{n} = 769.72 - \frac{(135.8)^2}{24} = 1.3183333$$

$$SSE = SS_{yy} - \hat{\beta}_1 SS_{xy} = 1.3183333 - (-0.002301796)(-129.94167) = 1.019237592$$

$$s^2 = \frac{SSE}{n-2} = \frac{1.019237592}{22} = 0.046329$$

$$s_{\hat{\beta}_1} = \sqrt{\frac{s^2}{SS_{xx}}} = \sqrt{\frac{0.046329}{56452.958}} = 0.000906$$

For confidence level 0.90, $\alpha = 0.10$ and $\alpha / 2 = 0.10/2 = 0.05$. From Table 2, Appendix D with $df = n-2 = 24-2 = 22,\ t_{0.05} = 1.717.$

The confidence interval is:

$$\hat{\beta}_1 \pm t_{0.05} s_{\hat{\beta}_1} \Rightarrow -0.0023 \pm 1.717(0.000906) \Rightarrow (-0.0039, -0.0008)$$

We are 90% confident that the change in the mean sweetness index for each one unit change in the pectin is between −0.0039 and −0.0007.

3.29 a. The equation for the simple linear regression is $y = \beta_0 + \beta_1 x + \varepsilon.$

b. The value of β_0 is probably irrelevant. By definition, β_0 is the mean value of entitlement score for those whose helicopter parent score is 0. We would expect β_1 to be positive. As the helicopter parent score increases, the entitlement score increases.

c. Since the p-value is less than $\alpha\left(p = 0.002 < 0.01\right)$, H_0 is rejected. There is sufficient evidence to indicate there is a positive linear relationship between entitlement scores and helicopter parent score at $\alpha = 0.01$.

3.30 For confidence level 0.95, $\alpha = 0.05$ and $\alpha / 2 = 0.05/2 = 0.025$. From Table 2, Appendix D with $df = n-2 = 50-2 = 48,\ t_{0.025} \approx 2.021.$ The confidence interval is:

$$\hat{\beta}_1 \pm t_{0.025} s_{\hat{\beta}_1} \Rightarrow 0.00542 \pm 2.021(0.00324) \Rightarrow (-0.0011, 0.0120)$$

We are 95% confident that the increase in the mean state's average annual number of public corruption convictions is between -0.0011 and 0.0120 for each unit increase in the state's average annual FEMA relief.

3.31 a. The equation for the simple linear regression is $y = \beta_0 + \beta_1 x + \varepsilon$.

b. The y-intercept does not have any meaning because 0 cannot be in the range of observed beauty index.

c. For each unit increase in the beauty index, the mean relative success is estimated to increase by 22.91 points.

d. To determine if the slope of the line is positive, we test:

$$H_0: \beta_1 = 0$$
$$H_a: \beta_1 > 0$$

The test statistic is $t = \dfrac{\hat{\beta}_1}{s_{\hat{\beta}_1}} = \dfrac{22.91}{3.73} = 6.14$.

The rejection region requires $\alpha = 0.01$ in the upper tail of the t distribution. From Table 2, Appendix D, with $df = n - 2 = 641 - 2 = 639$, $t_{0.01} = 2.326$. The rejection region is $t > 2.326$.

Since the observed value of the test statistic falls in the rejection region $(t = 6.14 > 2.326)$, *H₀* is rejected. There is sufficient evidence to indicate the slope of the line is positive at $\alpha = 0.01$. There is evidence to indicate that as the beauty index increases, the relative success also increases.

3.32 To determine if the simple linear regression model is useful for predicting Democratic vote share, we test:

$$H_0: \beta_1 = 0$$
$$H_a: \beta_1 \neq 0$$

The test statistic is $t = \dfrac{\hat{\beta}_1}{s_{\hat{\beta}_1}} = \dfrac{0.0275}{0.0650} = 0.42$ and the p-value is $p = 0.676$. (From Exercise 3.14)

Since the p-value is not less than $\alpha\,(p = 0.676 \nless 0.10)$, *H₀* is not rejected. There is insufficient evidence to indicate the simple linear regression model is useful for predicting Democratic vote share at $\alpha = 0.10$.

3.33 Using the calculations from Exercise 3.15 and these calculations:

$$SS_{yy} = \sum y^2 - \frac{\left(\sum y\right)^2}{n} = 83.474 - \frac{103.07^2}{144} = 9.70021597$$

$$SSE = SS_{yy} - \hat{\beta}_1\left(SS_{xy}\right) = 9.70021597 - (0.026421957)(19.975) = 9.172437366$$
$$s^2 = \frac{SSE}{n-2} = \frac{9.172437366}{144-2} = 0.064594629$$
$$s = \sqrt{s^2} = \sqrt{0.064594629} = 0.254154735$$

To determine if there is a linear trend between the proportion of names recalled and position, we test:

$$H_0: \beta_1 = 0$$
$$H_a: \beta_1 \neq 0$$

The test statistic is $t = \frac{\hat{\beta}_1 - 0}{s_{\hat{\beta}_1}} = \frac{\hat{\beta}_1}{s/\sqrt{SS_{xx}}} = \frac{0.02642 - 0}{0.25415/\sqrt{756}} = 2.86$

The rejection region requires $\alpha / 2 = 0.01/2 = 0.005$ in each tail of the t distribution. From Table 2, Appendix D, with $df = n - 2 = 144 - 2 = 142$, $t_{0.005} \approx 2.576$. The rejection region is $t < -2.576$ or $t > 2.576$..

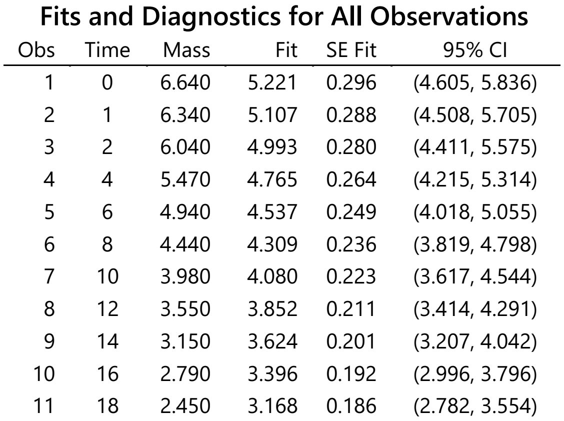
Since the observed test statistic falls in the rejection region $(t = 2.86 > 2.576)$, H_0 is rejected. There is sufficient evidence to indicate the proportion of names recalled is linearly related to position at $\alpha = .01$.

3.34 a. To determine if the spill mass tends to diminish linearly as time increases, we test:

$$H_0: \beta_1 = 0$$
$$H_a: \beta_1 < 0$$

Using information from Exercise 3.16, the test statistic is $t = -11.05$ and the p-value is $p = 0.000/2 = 0.000$. Since the p-value is less than $\alpha\left(p = 0.000 < 0.05\right)$, H_0 is rejected. There is sufficient evidence to indicate the spill mass tends to diminish linearly as time increases at $\alpha = 0.05$.

b. Using MINTAB, the 95% confidence intervals are:



Fits and Diagnostics for All Observations

Obs	Time	Mass	Fit	SE Fit	95% CI
1	0	6.640	5.221	0.296	(4.605, 5.836)
2	1	6.340	5.107	0.288	(4.508, 5.705)
3	2	6.040	4.993	0.280	(4.411, 5.575)
4	4	5.470	4.765	0.264	(4.215, 5.314)
5	6	4.940	4.537	0.249	(4.018, 5.055)
6	8	4.440	4.309	0.236	(3.819, 4.798)
7	10	3.980	4.080	0.223	(3.617, 4.544)
8	12	3.550	3.852	0.211	(3.414, 4.291)
9	14	3.150	3.624	0.201	(3.207, 4.042)
10	16	2.790	3.396	0.192	(2.996, 3.796)
11	18	2.450	3.168	0.186	(2.782, 3.554)

12	20	2.140	2.940	0.181	(2.563, 3.317)
13	22	1.860	2.712	0.179	(2.340, 3.084)
14	24	1.600	2.484	0.179	(2.112, 2.857)
15	26	1.370	2.256	0.182	(1.878, 2.634)
16	28	1.170	2.028	0.186	(1.640, 2.416)
17	30	0.980	1.800	0.193	(1.398, 2.202)
18	35	0.600	1.230	0.218	(0.776, 1.684)
19	40	0.340	0.660	0.251	(0.137, 1.182)
20	45	0.170	0.090	0.290	(-0.513, 0.693)
21	50	0.060	-0.480	0.332	(-1.171, 0.210)
22	55	0.020	-1.051	0.377	(-1.834, -0.267)
23	60	0.000	-1.621	0.423	(-2.500, -0.742)

3.35 a. For each 1% increase in the *ln*(body mass), the mean *ln*(eye mass) is estimated to increase by anywhere from 0.25 to 0.30.

b. For each 1% increase in the *ln*(body mass), the mean *ln*(orbit axis angle) is estimated to decrease by anywhere from 0.14 to 0.50.

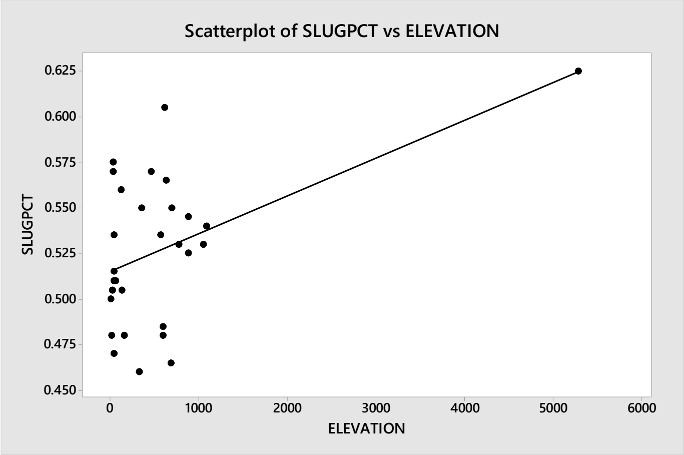
3.36 a. $\hat{\beta}_0 = 0.5151 \quad \hat{\beta}_1 = 0.000021$

b. To determine if there is a positive linear relationship between elevation and slugging percentage, we test:

$$H_0 : \beta_1 = 0$$
$$H_a : \beta_1 > 0$$

From the printout, the test statistic is $t = 2.89$ and the *p*-value is $p = 0.008 / 2 = 0.004$. Since the *p*-value is less than $\alpha\left(p = 0.004 < 0.01\right)$, *H₀* is rejected. There is sufficient evidence to indicate there is a positive linear relationship between elevation and slugging percentage at $\alpha = 0.01$.

c. Using MINITAB, the scatterplot is:



Denver's elevation is much greater than all the others. In addition, if the observation for Denver is deleted, there does not appear to be much of a relationship between elevation and slugging percentage.

d. Using MINITAB, the results are:

```
Analysis of Variance
Source          DF    Adj SS     Adj MS   F-Value   P-Value
Regression       1  0.001389   0.001389      0.98     0.332
  ELEVATION      1  0.001389   0.001389      0.98     0.332
Error           26  0.036922   0.001420
  Lack-of-Fit   22  0.036685   0.001667     28.08     0.003
  Pure Error     4  0.000238   0.000059
Total           27  0.038311

Coefficients
Term            Coef    SE Coef   T-Value   P-Value    VIF
Constant      0.5154     0.0107     48.33     0.000
ELEVATION   0.000020   0.000020      0.99     0.332   1.00

Regression Equation
SLUGPCT   =   0.5154 + 0.000020 ELEVATION
```

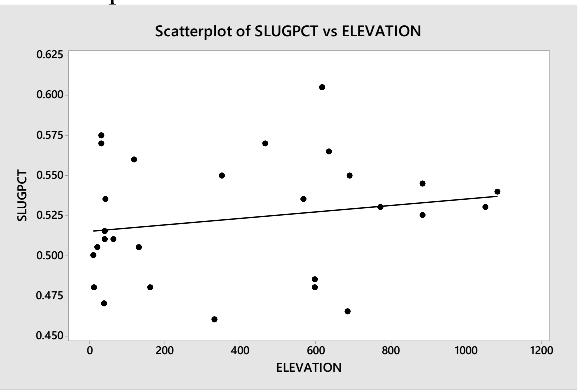
$\hat{\beta}_0 = 0.5154$ $\hat{\beta}_1 = 0.000020$

To determine if there is a positive linear relationship between elevation and slugging percentage, we test:

$$H_0 : \beta_1 = 0$$
$$H_a : \beta_1 > 0$$

From the printout, the test statistic is $t = 0.99$ and the *p*-value is $p = 0.332 / 2 = 0.166$. Since the *p*-value is not less than $\alpha\left(p = 0.166 \not< 0.01\right)$, *H₀* is not rejected. There is insufficient evidence to indicate there is a positive linear relationship between elevation and slugging percentage at $\alpha = 0.01$.

The new plot is:



3.37 a. Years of education and yearly income

b. Number of hours playing video games and GPA

3.38 a. If $r = 0.7$, , there is a positive linear relationship between x and y. As x increases, y tends to increase. The slope is positive.

b. If $r = -0.7$, there is a negative linear relationship between x and y. As x increases, y tends to decrease. The slope is negative.

c. If $r = 0$, , there is a 0 slope. There is no linear relationship between x and y.

d. If $r^2 = 0.64$, then r is either 0.8 or −0.8. The linear relationship between x and y could be either positive or negative.

3.39 a. From Exercise 3.6, $SS_{xx} = 17.5$, $SS_{yy} = 14$ and $SS_{xy} = 15$

$$r = \frac{SS_{xy}}{\sqrt{SS_{xx}SS_{yy}}} = \frac{15}{\sqrt{(17.5)(14)}} = 0.9583$$

From Exercise 3.19, $SSE = 1.1435$

$$r^2 = \frac{SS_{yy} - SSE}{SS_{yy}} = \frac{14 - 1.1435}{14} = 0.9183.$$

There is a strong positive correlation between x and y.
We can explain 91.83% of the variation in the sample y's using the linear model with x.

b. In Exercise 3.7, $SS_{xx} = 10$, $SS_{yy} = 16$ and $SS_{xy} = -12$

$$r = \frac{SS_{xy}}{\sqrt{SS_{xx}SS_{yy}}} = \frac{-12}{\sqrt{10(16)}} = -0.9487.$$

In Exercise 3.7, $SSE = SS_{yy} - \hat{\beta}_1 SS_{xy} = 16 - (-1.2)(-12) = 1.6.$

$$r^2 = \frac{SS_{yy} - SSE}{SS_{yy}} = \frac{16 - 1.6}{16} = 0.90.$$

There is a strong positive linear correlation between x and y.
We can explain 90% of the variation in the sample y's using the linear model with x.

3.40 We would expect the crime rate to increase as U.S. population increases. Therefore, we expect a positive correlation between the variables.

3.41 We would expect the GPA of a college student to be correlated to his/her I.Q. As the I.Q. score increases, we would expect the GPA to increase. Thus, the correlation would be positive.

3.42 a. $r = 0.975$. There is a very strong linear relationship between the sale price of a house and the appraised property market value.

b. $r^2 = 0.9516$. 95.16% of the sample home sale prices is explained by the linear relationship between the appraised value of the house and the final market price.

3.43 a. $r^2 = 0.18$. 18% of the sample number of points scored is explained by the linear relationship between the number of points scored and the number of yards from the opposing goal line.

b. $r = -\sqrt{0.18} = -0.424$. The value of r is negative because the coefficient associated with the number of yards from the opposing goal line in the fitted regression line is negative.

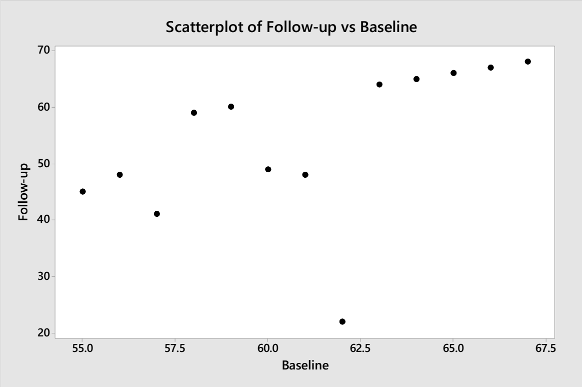
3.44 a. Since the p-value of 0.33 is greater than $\alpha = 0.05$, we cannot conclude that there is a significant linear relationship between cooperation use and average payoff.

b. Since the p-value of 0.66 is greater than $\alpha = 0.05$, we cannot conclude that there is a significant linear relationship between defection use and average payoff.

c. Since the p-value of 0.001 is smaller than $\alpha = 0.05$, we can conclude that there is a significant linear relationship between punishment use and average payoff.

3.45 a. Since the p-value of 0.07 is greater than $\alpha = 0.05$, we cannot conclude that there is a significant linear relationship between baseline and follow-up physical activity for obese young adults; fail to reject $H_0 : \rho = 0$ at $\alpha = .05$.

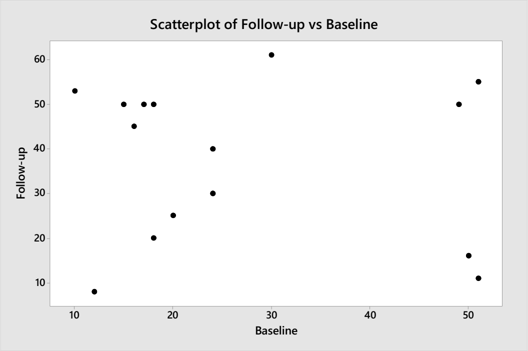
b. A possible scatterplot of the data would be:



c. $r^2 = (.50)^2 = 0.25$, thus 25% of the variability around the sample mean for the total of follow-up number of movements is explained by the linear relationship between the baseline total number of movements for the obese adults and the follow-up total number of movements for the obese adults.

d. Since the correlation value itself is close to zero and the p-value of 0.66 is greater than $\alpha = 0.05$, we cannot conclude that there is a significant linear relationship between baseline and follow-up physical activity for normal weight young adults; fail to reject $H_0 : \rho = 0$ at $\alpha = .05$.

e. A possible scatterplot is:



f. $r^2 = (-.12)^2 = 0.0144$. Thus 1.44% of the variability around the sample mean for the total of follow-up number of movements is explained by the linear relationship between the baseline total number of movements for the normal weight young adults and the total of follow-up number of movements for the normal weight young adults.

3.46 In Exercise 3.13, $SS_{xx} = 56{,}452.958$ and $SS_{xy} = -129.94167$

$$SS_{yy} = \sum y^2 - \frac{\left(\sum y\right)^2}{n} = 769.72 - \frac{135.8^2}{24} = 1.318333$$

$$r = \frac{SS_{xy}}{\sqrt{SS_{xx}SS_{yy}}} = \frac{-129.94167}{\sqrt{56{,}452.958(1.318333)}} = -0.4763.$$

$$SSE = SS_{yy} - \hat{\beta}_1 SS_{xy} = 1.318333 - (-0.002301769)(-129.94167) = 1.01924.$$

$$r^2 = \frac{SS_{yy} - SSE}{SS_{yy}} = \frac{1.318333 - 1.01924}{1.318333} = 0.2269.$$

22.69% of the variability around the sample mean for the sweetness index can be explained by the linear relationship between the sweetness index and the amount of water-soluble pectin.

3.47 a. There is a rather weak negative linear relationship between the numerical value of a last name and the response time.

b. Since the p-value is less than $\alpha\left(p = 0.018 < 0.05\right)$, H$_0$ is rejected. There is sufficient evidence to indicate a negative linear relationship between the numerical value of a last name and the response time.

c. Yes, the analysis supports the researchers' *last name effect* theory. Because the correlation coefficient is negative, as the numerical value of the last name increases, the response time tends to decrease.

3.48 Using the values computed in Exercise 3.15:

$$r=\frac{SS_{xy}}{\sqrt{SS_{xx}SS_{yy}}}=\frac{19.975}{\sqrt{756(9.70031597)}}=0.2333$$

Because *r* is fairly close to 0, there is a very weak positive linear relationship between the proportion of names recalled and position.

$r^2=0.2333^2=0.0544$

5.44% of the sample variance of proportion of names recalled around the sample mean is explained by the linear relationship between proportion of names recalled and position.

3.49 a. To determine if the true population correlation coefficient relating NRMSE and bias is positive, we test:

$$H_0:\rho=0$$
$$H_a:\rho>0$$

The test statistic is $t=\frac{r}{\sqrt{\frac{1-r^2}{n-2}}}=\frac{0.2838}{\sqrt{\frac{1-0.2838^2}{3,600-2}}}=17.753$.

No α value was given, so we will use $\alpha=0.5$. The rejection region requires $\alpha=0.5$ in the upper tail of the *t* distribution. From Table 2, Appendix D, with $df=n-2=3,600-2=3598$, $t_{0.05}=1.645$. The rejection region is $t>1.645$.

Since the observed value of the test statistic falls in the rejection region $(t=17.753>1.645)$, *H₀* is rejected. There is sufficient evidence to indicate the true population correlation coefficient relating NRMSE and bias is positive at $\alpha=0.5$.

b. No, we would not recommend using NRMSE as a linear predictor of bias. The estimated correlation coefficient is $r=0.2838$. This indicates that there is a rather weak positive linear relationship between NRMSE and bias. The sample size was extremely large. The larger the sample size, the easier it is to find statistical significance. In this case, there is statistical significance, but not practical significance.

3.50 a. The sample correlation coefficient between PSI and PHI-F is $r=0.401$. There is a weak positive linear relationship between the perceived sensory intensity and the perceived hedonic intensity for favorite food.

The sample correlation coefficient between PSI and PHI-L is $r=-0.375$. There is a weak negative linear relationship between the perceived sensory intensity and the perceived hedonic intensity for least favorite food.

b. Yes, we agree that those with the greatest taste intensity tend to experience more extreme food likes and dislikes. As the taste intensity increases, the intensity of favorite foods tends to increase. As the taste intensity increases, the intensity of least favorite foods tends to decrease.

3.51 a. $r^2 = 0.948$. 94.8% of the variability around the mean *ln*(eye mass) is explained by the linear relationship between *ln*(eye mass) and *ln*(body mass).

b. From 3.35a, the relationship between *ln* (eye mass) and *ln* (body mass) is positive. Therefore, $r = \sqrt{0.948} = 0.974$. There is a very strong positive linear relationship between *ln* (eye mass) and *ln* (body mass).

c. $r^2 = .375$. 37.5% of the variability around the mean *ln*(orbit axis angle) is explained by the linear relationship between *ln*(orbit axis angle) and *ln*(body mass).

d. From 3.35b, the relationship between *ln*(orbit axis angle) and *ln*(body mass) is negative. Therefore, $r = -\sqrt{0.375} = -0.612$. There is a moderate negative linear relationship between *ln*(orbit axis angle) and *ln*(body mass).

3.52 a. First, examine the formulas for the confidence interval and the prediction interval. The only difference is that the prediction interval has an extra term (a "1") beneath the radical. Thus, the prediction interval must be wider:

$$\sqrt{\frac{1}{n}+\frac{\left(x_p-\overline{x}\right)^2}{SS_{xx}}} < \sqrt{1+\frac{1}{n}+\frac{\left(x_p-\overline{x}\right)^2}{SS_{xx}}}$$

The error in estimating the mean value of y, $E\left(y\right)$, for a given value of x, say x_p, is the distance between the least squares line, $\hat{y} = \hat{\beta}_0 + \hat{\beta}_1 x$, and the true line of means, $E\left(y\right) = \beta_0 + \beta_1 x$. In contrast, the error in predicting some future of y, $\left(\hat{y} - y_p\right)$ is the sum of two errors: the error of estimating the mean of y, $E\left(y\right)$, plus the random error of the actual values of y around its mean. Consequently, 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 value of x.

b. Since the standard error contains the term $\frac{\left(x_p-\overline{x}\right)^2}{SS_{xx}}$, the further x_p is from $\overline{x}$, the larger the standard error. This causes the confidence intervals to be wider for values of x_p further from $\overline{x}$. The implication is our best confidence intervals (narrowest) will be found when $x_p = \overline{x}$.

3.53 a. $\hat{\beta}_1 = \frac{SS_{xy}}{SS_{xx}} = \frac{16.22}{4.77} = 3.400$

$SSE = SS_{yy} - \hat{\beta}_1 SS_{xy} = 59.21 - 3.4\left(16.22\right) = 4.062$

$s^2 = \frac{SSE}{n-2} = \frac{4.062}{20-2} = 0.226.$

b. For $x = 2.5$, $\hat{y} = 2.1 + 3.4\left(2.5\right) = 10.6$

The form of the 95% confidence interval is $\hat{y} \pm t_{\alpha/2} s \sqrt{\frac{1}{n} + \frac{(x - \bar{x})^2}{SS_{xx}}}$.

For confidence coefficient 0.95, $\alpha = 0.05$ and $\alpha / 2 = 0.05 / 2 = 0.025$. From Table 2, Appendix D, with $df = n - 2 = 20 - 2 = 18$, $t_{0.025} = 2.101$.
The 95% confidence interval is:

$$10.6 \pm 2.101\sqrt{0.226}\sqrt{\frac{1}{20} + \frac{(2.5 - 2.5)^2}{4.77}} \Rightarrow 10.6 \pm 0.223 \Rightarrow (10.377, 10.823)$$

We are 95% confident the mean value of y when $x = 2.5$ is between 10.377 and 10.823.

c. For $x = 2.0, \hat{y} = 2.1 + 3.4(2.0) = 8.9$.

The 95% confidence interval is:

$$8.9 \pm 2.101\sqrt{0.226}\sqrt{\frac{1}{20} + \frac{(2.0 - 2.5)^2}{4.77}} \Rightarrow 8.9 \pm 0.320 \Rightarrow (8.580, 9.220)$$

We are 95% confident the mean value of y when $x = 2.0$ is between 8.580 and 9.220.

d. For $x = 3.0$, $\hat{y} = 2.1 + 3.4(3.0) = 12.3$.

The 95% confidence interval is:

$$12.3 \pm 2.101\sqrt{0.226}\sqrt{\frac{1}{20} + \frac{(3.0 - 2.5)^2}{4.77}} \Rightarrow 12.3 \pm 0.320 \Rightarrow (11.980, 12.620)$$

We are 95% confident the mean value of y when $x = 3.0$ is between 11.980 and 12.620.

e. The width of the interval in (b) is $10.823 - 10.377 = 0.446$.
The width of the interval in (c) is $9.220 - 8.580 = 0.640$.
The width of the interval in (d) is $12.620 - 11.980 = 0.640$.

As the value of x moves away from $\bar{x}$ = 2.5, the confidence interval gets wider.

f. The 95% prediction interval is $\hat{y} \pm t_{\alpha/2} s \sqrt{1 + \frac{1}{n} + \frac{(x - \bar{x})^2}{SS_{xx}}}$.

$$12.3 \pm 2.101\sqrt{0.226}\sqrt{1 + \frac{1}{20} + \frac{(3.0 - 2.5)^2}{4.77}} \Rightarrow 12.3 \pm 1.049 \Rightarrow (11.251, 13.349).$$

We are 95% confident that the actual value of y will be between 11.251 and 13.349 when the value of x is 3.

3.54 a. No. We know there is a significant linear relationship between sale price and appraised value. However, the actual sale prices may be scattered quite far from the predicted line.

b. From the printout, the 95% prediction interval for the actual sale price when the appraised value is \$300,000 is $(285.938,\ 561.741)$ or $(\$285{,}938,\ \$561{,}741)$. We are 95% confident that the actual sale price for a home appraised at \$300,000 is between \$285,938 and \$561,741.

c. From the printout, the 95% confidence interval for the mean sale price when the appraised value is \$300,000 is $(408.119,\ 439.560)$ or $(\$408{,}119,\ \$439{,}560)$. We are 95% confident that the mean sale price for a home appraised at \$300,000 is between \$408,119 and \$439,560.

3.55 a. Researchers should use a prediction interval for y with

$$x = 10 \Rightarrow \hat{y} \pm t_{\alpha/2} s\sqrt{1 + \frac{1}{n} + \frac{\left(x_p - \bar{x}\right)^2}{SS_{xx}}} \Rightarrow \hat{y} \pm t_{\alpha/2} s\sqrt{1 + \frac{1}{n} + \frac{\left(10 - \bar{x}\right)^2}{SS_{xx}}}.$$

b. Researchers should use a confidence interval for the mean value of y or $E(y)$, with

$$x = 10 \Rightarrow \hat{y} \pm t_{\alpha/2} s\sqrt{\frac{1}{n} + \frac{\left(x_p - \bar{x}\right)^2}{SS_{xx}}} \Rightarrow \hat{y} \pm t_{\alpha/2} s\sqrt{\frac{1}{n} + \frac{\left(10 - \bar{x}\right)^2}{SS_{xx}}}.$$

3.56 a. We are 95% confident that the actual value of the angular size of the Moon is between 323.502 and 326.108 when the height above the horizon is 50 degrees.

b. We are 95% confident that the mean value of the angular size of the Moon is between 324.448 and 325.163 when the height above the horizon is 50 degrees.

c. No, we would not recommend using the least squares line to predict the angular size of the Moon for a height of 80 degrees because 80 degrees is outside the observed range of data used to construct the least squares line.

3.57 For $x = 300$, the confidence interval for $E(y)$ is $(5.45812,\ 5.65964)$. We are 90% confident that the mean sweetness index is between 5.458 and 5.660 when the amount of pectin is 300.

3.58 a. From Exercises 3.15 and 3.33, $\bar{x} = 5.5$, $SS_{xx} = 756$, $s = 0.25415$, and $\hat{y} = 0.5704 + 0.0264x$. For $x = 5$, $\hat{y} = 0.5704 + 0.0264(5) = 0.7024$.

For confidence coefficient 0.99, $\alpha = 0.01$ and $\alpha/2 = 0.01/2 = 0.005$. From Table 2, Appendix D, with $df = n - 2 = 144 - 2 = 142$, $t_{0.005} \approx 2.576$. The 99% confidence interval is:

$$\hat{y} \pm t_{\alpha/2} s\sqrt{\frac{1}{n} + \frac{\left(x_p - \bar{x}\right)^2}{SS_{xx}}} \Rightarrow 0.7024 \pm 2.576(0.2542)\sqrt{\frac{1}{144} + \frac{(5 - 5.5)^2}{756}}$$

$\Rightarrow 0.7024 \pm 0.0559 \Rightarrow (0.6465,\ 0.7583)$

We are 99% confident that the mean recall of all those in the 5th position is between 0.6465 and 0.7583.

b. For confidence coefficient 0.99, $\alpha = 0.01$ and $\alpha/2 = 0.01/2 = 0.005$. From Table 2, Appendix D, with $df = n-2 = 144-2 = 142$, $t_{0.005} \approx 2.576$. The 99% prediction interval is:

$$\hat{y} \pm t_{\alpha/2} s\sqrt{1+\frac{1}{n}+\frac{(x_p-\bar{x})^2}{SS_{xx}}} \Rightarrow 0.7024 \pm 2.576(0.2542)\sqrt{1+\frac{1}{144}+\frac{(5-5.5)^2}{756}}$$

$\Rightarrow 0.7024 \pm 0.6572 \Rightarrow (0.0452,\ 1.3596)$

We are 99% confident that the actual recall of a person in the 5th position is between 0.0452 and 1.3596. Since the proportion of names recalled cannot be larger than 1, the actual proportion recalled will be between 0.0452 and 1.000.

c. The prediction interval in part b is wider than the confidence interval in part a. The prediction interval will always be wider than the confidence interval. The confidence interval for the mean is an interval for predicting the mean of all observations for a particular value of *x*. The prediction interval is a confidence interval for the actual value of the dependent variable for a particular value of *x*.

3.59 a. From Exercises 3.16 and 3.34, $\bar{x} = 22.87$, $SS_{xx} = 6906.608$, $s = 0.8573$, and $\hat{y} = 5.22 - 0.114x$.

For $x = 15$, $\hat{y} = 5.22 - 0.114(15) = 3.51$.

For confidence coefficient 0.90, $\alpha = 0.10$ and $\alpha = 0.10/2 = 0.05$. From Table 2, Appendix D, with $df = n-2 = 23-2 = 21$, $t_{0.05} = 1.721$. The 90% confidence interval is:

$$\hat{y} \pm t_{\alpha/2} s\sqrt{\frac{1}{n}+\frac{(x_p-\bar{x})^2}{SS_{xx}}} \Rightarrow 3.51 \pm 1.721(0.8573)\sqrt{\frac{1}{23}+\frac{(15-22.87)^2}{6906.608}} \Rightarrow$$

$3.51 \pm 0.34 \Rightarrow (3.17,\ 3.85)$.

We are 90% confident that the mean mass of all spills with an elapsed time of 15 minutes is between 3.17 and 3.85.

b. For confidence coefficient 0.90, $\alpha = 0.10$ and $\alpha/2 = 0.10/2 = 0.05$. From Table 2, Appendix D, with $df = n-2 = 23-2 = 21$, $t_{0.05} = 1.721$. The 90% prediction interval is:

$$\hat{y} \pm t_{\alpha/2} s\sqrt{1+\frac{1}{n}+\frac{(x_p-\bar{x})^2}{SS_{xx}}} \Rightarrow 3.51 \pm 1.721(0.8573)\sqrt{1+\frac{1}{23}+\frac{(15-22.87)^2}{6906.608}} \Rightarrow$$

$3.51 \pm 1.514 \Rightarrow (2.00,\ 5.02)$.

We are 90% confident that the mass of a single spill with an elapsed time of 15 minutes is between 2.00 and 5.02.

3.60 a. To determine if the model is adequate for predicting nitrogen amount, we test:

$$H_0 : \beta_1 = 0$$
$$H_a : \beta_1 \neq 0$$

The test statistic is $t = 32.80$ and the p-value is $p < 0.0001$.

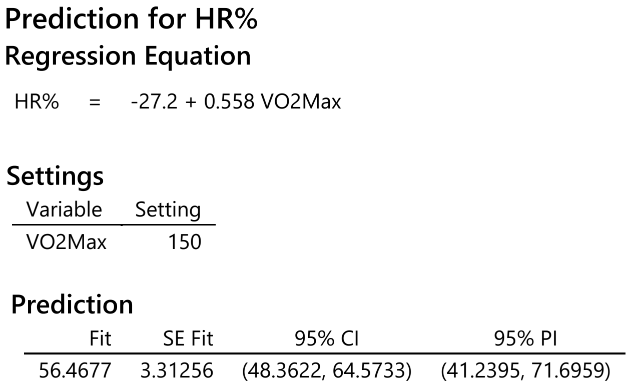
Since the p-value is so small $\left(p < 0.0001\right)$, H_0 is rejected for any reasonable value of α. There is sufficient evidence to indicate that the amount of ammonium contributes information for the prediction of the amount of nitrogen removed using a linear model.

b. From the printout, the 95% prediction interval is $\left(41.8558, 77.8634\right)$. We are 95% confident that the actual amount of nitrogen removed when the amount of ammonium used is 100mg/l will be between 41.8558 and 77.8634 mg/l.

c. The 95% confidence interval for the mean amount of nitrogen removed when the amount of ammonium used is 100 mg/l will be narrower than the prediction interval. This is because the prediction interval for the actual value contains the variability of locating the mean and the variability of the actual values around the mea. The confidence interval for the mean contains only the variability in locating the mean.

3.61 a. The researchers are interested in the confidence interval for $E\left(y\right)$ or the average in-game heart rate of all top-level water polo players who have a maximal oxygen uptake of 150 VO_2max.

b. Using MINITAB, the results are:



Prediction for HR%

Regression Equation

```
HR%   =   -27.2 + 0.558 VO2Max
```

Settings

Variable	Setting
VO2Max	150

Prediction

Fit	SE Fit	95% CI	95% PI
56.4677	3.31256	(48.3622, 64.5733)	(41.2395, 71.6959)

The 95% confidence interval is $\left(48.3622, 64.5733\right)$.

c. We are 95% confident that the average in-game heart rate of all top-level polo players with a VO_2max of 150 is between 48.3622 and 64.5733.

3.62 Step 1. We hypothesize a straight-line probabilistic model: $y = \beta_0 + \beta_1 x + \varepsilon$ where y = the monthly price of recycled colored plastic bottles and x = the monthly price of naphtha.

Step 2: Collect the data. The data have been collected.

Step 3: Estimate the unknown parameters in the proposed model. From the exercise, the least squares estimates of β_0 and β_1 are: $\hat{\beta}_0 = -32.35$ $\hat{\beta}_1 = 4.82$
The least squares line is $\hat{y} = -32.35 + 4.82x$.

The least squares estimate of the slope, $\hat{\beta}_1 = 4.82$, implies that the estimated mean monthly price of recycled colored plastic bottles increases by 4.82 for each additional unit increase in the monthly price of naphtha. This interpretation is valid only over the observed values of the monthly price of naphtha. The estimated y-intercept, $\hat{\beta}_0 = -32.35$, has no practical meaning in this example because 0 will not be within the observed range of values for monthly price of naphtha.

Step 4: Specify the probability distribution of the random error component ε. We assume

(1) $E(\varepsilon) = 0$
(2) $\text{Var}(\varepsilon) = \sigma^2$ is constant for all x-values
(3) ε has a normal distribution
(4) ε's are independent

Step 5: To determine if there is a linear relationship between the monthly price of recycled colored plastic bottles and the monthly price of naphtha, we test:

$$H_0: \beta_1 = 0$$
$$H_a: \beta_1 \neq 0$$

The test statistic is $t = 16.60$.

The rejection region requires $\alpha / 2 = 0.05 / 2 = 0.025$ in each tail of the t distribution. From Table 2, Appendix D, with $df = n - 2 = 120 - 2 = 118$, $t_{0.025} \approx 1.980$. The rejection region is $t < -1.980$ or $t > 1.980$.

Since the observed value of the test statistic falls in the rejection region $(t = 16.60 > 1.980)$, H_0 is rejected. There is sufficient evidence to there is a linear relationship between the monthly price of recycled colored plastic bottles and the monthly price of naphtha at $\alpha = 0.05$.

$r^2 = 0.69$ 69% of the sample variation around the mean monthly price of recycled colored plastic bottles is explained by the linear relationship between the monthly price of recycled colored plastic bottles and the monthly price of naphtha.

3.63 a. Using MINITAB, the results are:

```
Regression Analysis: Corrupt versus GDP
Analysis of Variance
 Source      DF   Adj SS    Adj MS   F-Value   P-Value
 Regression   1   3345.8   3345.76     45.33     0.000
 Error       11    811.9     73.81
 Total       12   4157.7

Model Summary
       S     R-sq   R-sq(adj)
 8.59141   80.47%      78.70%

Coefficients
 Term             Coef    SE Coef   T-Value   P-Value
 Constant        25.89       3.09      8.37     0.000
 GDP          0.000985   0.000146      6.73     0.000

Regression Equation
 Corrupt   =   25.89 + 0.000985 GDP
```

The fitted regression line is $\hat{y} = 25.89 + 0.000985GPD$.

To determine if GDP per capita is a linear predictor of corruption level, we test:

$$H_0 : \beta_1 = 0$$
$$H_a : \beta_1 \neq 0$$

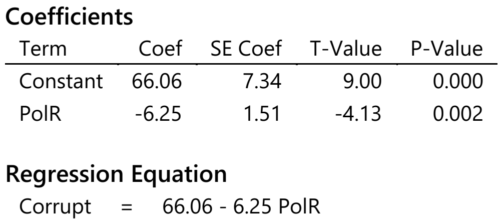
The test statistic is $t = 6.73$ and the p-value is $p = 0.000$. Since the p-value is so small, H_0 is rejected. There is sufficient evidence to indicate GDP per capita is a linear predictor of corruption level for any reasonable value of α.

$r^2 = 0.8047$ This indicates that 80.47% of the variability in the corruption values is explained by the linear relationship between the corruption values and the GDP per capita.

b. Using MINITAB, the results are:

```
Regression Analysis: Corrupt versus PolR
Analysis of Variance
 Source      DF   Adj SS   Adj MS   F-Value   P-Value
 Regression   1     2528   2527.6     17.06     0.002
 Error       11     1630    148.2
 Total       12     4158

Model Summary
       S     R-sq   R-sq(adj)
 12.1732   60.79%      57.23%
```



```
Coefficients
 Term       Coef    SE Coef   T-Value   P-Value
 Constant   66.06      7.34      9.00     0.000
 PolR       -6.25      1.51     -4.13     0.002

Regression Equation
 Corrupt  =  66.06 - 6.25 PolR
```

The fitted regression line is $\hat{y} = 66.06 - 6.25PolR$.

To determine if degree of freedom in political rights is a linear predictor of corruption level, we test:

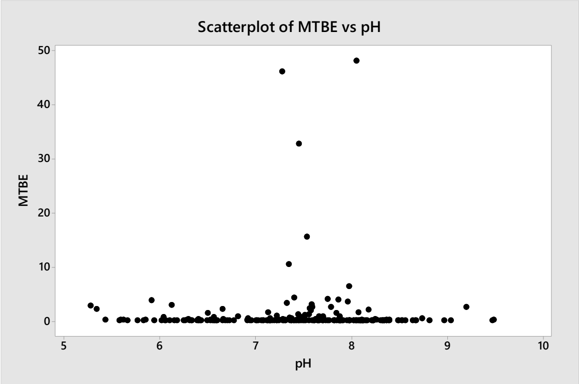
$$H_0 : \beta_1 = 0$$
$$H_a : \beta_1 \neq 0$$

The test statistic is $t = -4.13$ and the p-value is $p = 0.002$. Since the p-value is so small, H_0 is rejected. There is sufficient evidence to indicate GDP per capita is a linear predictor of corruption level for any value of $\alpha > 0.002$.

$r^2 = 0.6079$ This indicates that 60.79% of the variability in the corruption values is explained by the linear relationship between the corruption values and the degree of freedom in political rights.

c. Both variables, GDP per capita and degree of freedom in political rights, are significant predictors of corruption levels. Of the two, GDP per capita is a better predictor because the r^2 value is larger and the p-value for the test is smaller.

3.64 Using MINITAB, a scatterplot of the data is:



From the plot, there does not look like there is a linear relationship between MTBE and pH level.

The proposed linear regression model is $y = \beta_0 + \beta_1 x + \varepsilon$. Using MINITAB, an analysis of the data is:

Analysis of Variance

Source	DF	Adj SS	Adj MS	F-Value	P-Value
Regression	1	2.01	2.008	0.08	0.782
Error	221	5785.93	26.181		
Total	222	5787.94			

Model Summary

S	R-sq	R-sq(adj)	R-sq(pred)
5.11670	0.03%	0.00%	0.00%

Coefficients

Term	Coef	SE Coef	T-Value	P-Value	VIF
Constant	0.35	3.14	0.11	0.911	
pH	0.116	0.420	0.28	0.782	1.00

The parameter estimates of the least squares line are: $\hat{\beta}_0 = 0.35 \quad \hat{\beta}_1 = 0.116$
The least squares line is $\hat{y} = 0.35 + 0.116x$.

The least squares estimate of the slope, $\hat{\beta}_1 = 0.116,$ implies that the estimated MTBE increases by 0.116 for each additional unit increase in the pH level. This interpretation is valid only over the observed values of the pH level which is from 5.28 to 9.48. The estimated y-intercept, $\hat{\beta}_0 = 0.35$ has no practical meaning in this example because 0 will not be within the observed range of the pH levels.

The estimate of σ is $s = 5.1167$. The value of this estimate is very large compared to most of the values of MTBE.

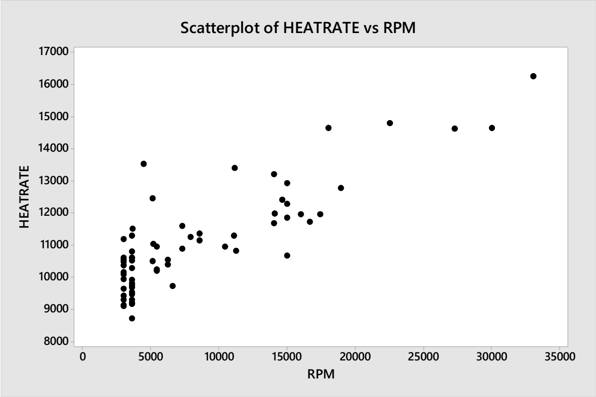
To determine if there is a linear relationship between the MTBE and the pH level, we test:

$$H_0 : \beta_1 = 0$$
$$H_a : \beta_1 \neq 0$$

The test statistic is $t = 0.28$ and the p-value is $p = 0.782$. Since the p-value is so large, H_0 will not be rejected for any reasonable value of α. There is insufficient evidence to indicate there is a linear relationship between the MTBE and the pH level.

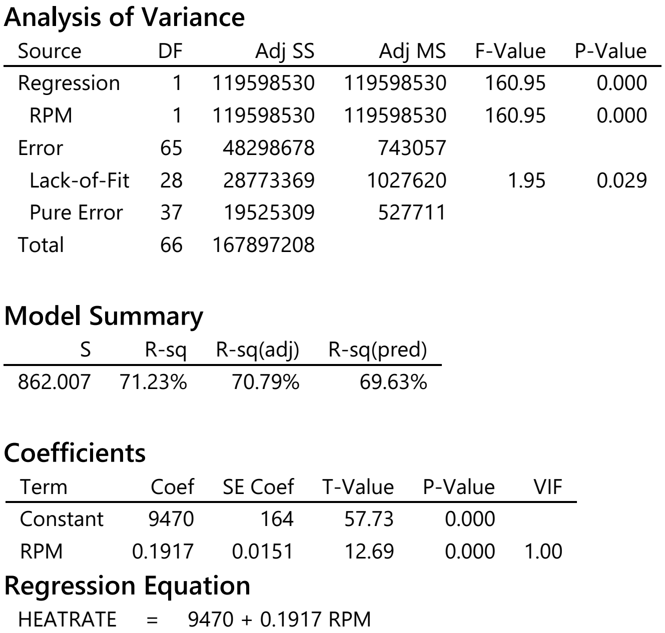
$r^2 = 0.00$ This indicates that 0% of the variability in the MTBE values is explained by the linear relationship between the MTBE values and the pH levels. This would indicate that a linear regression model does not explain the relationship between MTBE and pH.

3.65 Using MINITAB, a scatter plot of the data is:



From the plot, there is evidence to indicate a linear relationship between heat rate and speed.

The proposed linear regression model is $y = \beta_0 + \beta_1 x + \varepsilon$. Using MINITAB, an analysis of the data is:



```
Analysis of Variance
Source          DF     Adj SS     Adj MS   F-Value   P-Value
Regression       1  119598530  119598530    160.95     0.000
  RPM            1  119598530  119598530    160.95     0.000
Error           65   48298678     743057
  Lack-of-Fit   28   28773369    1027620      1.95     0.029
  Pure Error    37   19525309     527711
Total           66  167897208

Model Summary
      S    R-sq  R-sq(adj)  R-sq(pred)
862.007  71.23%     70.79%      69.63%

Coefficients
Term        Coef  SE Coef  T-Value  P-Value   VIF
Constant    9470      164    57.73    0.000
RPM       0.1917   0.0151    12.69    0.000  1.00
Regression Equation
HEATRATE  =  9470 + 0.1917 RPM
```

The parameter estimates of the least squares line are: $\hat{\beta}_0 = 9470 \quad \hat{\beta}_1 = 0.1917$
The least squares line is $\hat{y} = 9470 + 0.1917x$.

The least squares estimate of the slope, $\hat{\beta}_1 = 0.1917$, implies that the estimated heat rate increases by 0.1917 units for each additional unit increase in the speed. This interpretation is valid only over the observed values of the speed level which is from 3,000 to 33,000. The estimated y-intercept, $\hat{\beta}_0 = 9470$ has no practical meaning in this example because 0 will not be within the observed range of the speed levels.

The estimate of σ is $s = 862.007$. We expect most of the observations to fall within $2s = 2(862.007) = 1724.014$ units of their predicted values.

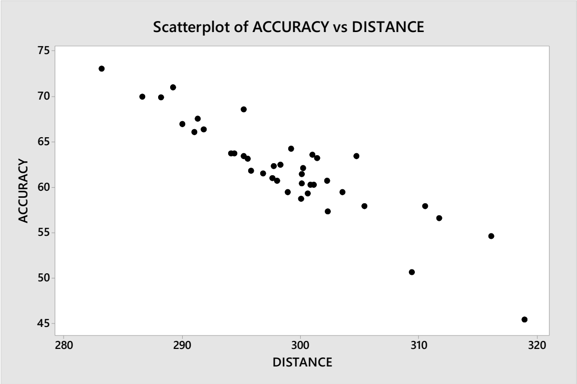
To determine if there is a linear relationship between the heat rate and the speed, we test:

$$H_0 : \beta_1 = 0$$
$$H_a : \beta_1 \neq 0$$

The test statistic is $t = 12.69$ and the p-value is $p = 0.000$. Since the p-value is so small, H_0 will be rejected for any reasonable value of α. There is sufficient evidence to indicate there is a linear relationship between the heat rate and speed.

$r^2 = 0.7173$ This indicates that 71.73% of the variability in the heat rate values is explained by the linear relationship between heat rate and the speed. This indicates that a linear regression line models the relationship between heat rate and speed fairly well.

3.66 Using MINITAB, a scatterplot of the data is:



From the plot, there is evidence to indicate a linear relationship between accuracy and distance.

The proposed linear regression model is $y = \beta_0 + \beta_1 x + \varepsilon$. Using MINITAB, an analysis of the data is:

Analysis of Variance

```
Source          DF    Adj SS    Adj MS   F-Value   P-Value
Regression       1    874.99   874.989    174.95     0.000
  DISTANCE       1    874.99   874.989    174.95     0.000
Error           38    190.06     5.001
  Lack-of-Fit   36    176.55     4.904      0.73     0.735
  Pure Error     2     13.51     6.753
Total           39   1065.04
```

Model Summary

```
      S    R-sq   R-sq(adj)   R-sq(pred)
2.23639  82.16%      81.69%       79.26%
```

```
Coefficients
Term           Coef   SE Coef   T-Value   P-Value    VIF
Constant      250.1      14.2     17.58     0.000
DISTANCE    -0.6294    0.0476    -13.23     0.000   1.00

Regression Equation
ACCURACY   =   250.1 - 0.6294 DISTANCE
```

The parameter estimates of the least squares line are: $\hat{\beta}_0 = 250.1 \quad \hat{\beta}_1 = -0.6294$
The least squares line is $\hat{y} = 250.1 - 0.6294x$.

The least squares estimate of the slope, $\hat{\beta}_1 = -0.6294$, implies that the estimated accuracy decreases by 0.6294 units for each additional yard increase in distance. This interpretation is valid only over the observed values of distance which is from 293.2 to 318.9 yards. The estimated y-intercept, $\hat{\beta}_0 = 250.1$ has no practical meaning in this example because 0 will not be within the observed range of distances.

The estimate of σ is $s = 2.23639$. We expect most of the observations to fall within $2s = 2(2.23639) = 4.473$ units of their predicted values.

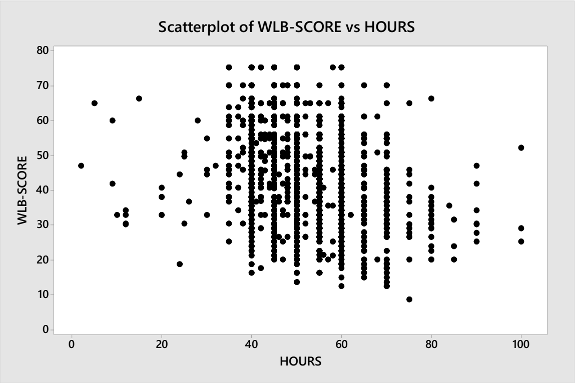
To determine if there is a negative linear relationship between accuracy and distance, we test:

$$H_0 : \beta_1 = 0$$
$$H_a : \beta_1 < 0$$

The test statistic is $t = -13.23$ and the p-value is $p = 0.000 / 2 = 0.000$. Since the p-value is so small, H_0 will be rejected for any reasonable value of α. There is sufficient evidence to indicate there is a negative linear relationship between accuracy and distance.

$r^2 = 0.8216$ This indicates that 82.16% of the variability in the accuracy values is explained by the linear relationship between accuracy and distance. This indicates that a linear regression line models the relationship between accuracy and distance fairly well. The professional golfer has a valid concern.

3.67 Using MINITAB, a scatterplot of the data is:



From the plot, there is evidence to indicate a slight linear relationship between work-life balance scale score and average number of hours worked per week

The proposed linear regression model is $y = \beta_0 + \beta_1 x + \varepsilon$. Using MINITAB, an analysis of the data is:

Analysis of Variance

Source	DF	Adj SS	Adj MS	F-Value	P-Value
Regression	1	23803	23803.1	157.73	0.000
HOURS	1	23803	23803.1	157.73	0.000
Error	2085	314647	150.9		
Lack-of-Fit	42	11939	284.3	1.92	0.000
Pure Error	2043	302708	148.2		
Total	2086	338451			

Model Summary

S	R-sq	R-sq(adj)	R-sq(pred)
12.2845	7.03%	6.99%	6.84%

Coefficients

Term	Coef	SE Coef	T-Value	P-Value	VIF
Constant	62.50	1.41	44.22	0.000	
HOURS	-0.3467	0.0276	-12.56	0.000	1.00

Regression Equation

WLB-SCORE = 62.50 - 0.3467 HOURS

The parameter estimates of the least squares line are: $\hat{\beta}_0 = 62.50 \quad \hat{\beta}_1 = -0.3467$
The least squares line is $\hat{y} = 62.50 - 0.3467x$.

The least squares estimate of the slope, $\hat{\beta}_1 = -0.3467$, implies that the estimated work-life balance scale score decreases by 0.3467 units for each additional average number of hours worked per week. This interpretation is valid only over the observed values of distance which is from 2 to 100 hours. The estimated y-intercept, $\hat{\beta}_0 = 62.50$ has no practical meaning in this example because 0 will not be within the observed range of hours worked.

The estimate of σ is $s = 12.2845$. We expect most of the observations to fall within $2s = 2(12.2845) = 24.569$ units of their predicted values.

To determine if there is a linear relationship between work-life balance scale score and average number of hours worked per week, we test:

$$H_0 : \beta_1 = 0$$
$$H_a : \beta_1 \neq 0$$

The test statistic is $t = -12.56$ and the p-value is $p = 0.000$. Since the p-value is so small, H_0 will be rejected for any reasonable value of α. There is sufficient evidence to indicate there is a linear relationship between work-life balance scale score and average number of hours worked per week.

$r^2 = 0.0703$ This indicates that only 7.03% of the variability in the work-life balance scale scores is explained by the linear relationship between work-life balance scale scores and average number of hours worked per week. This indicates that although there is a significant linear relationship between work-life balance scale score and average number of hours worked per week, the relationship is very weak. Many other factors are influencing work-life balance scale scores.

3.68 Some preliminary calculations are:

$$\sum x = 24 \qquad \sum y = 77 \qquad \sum x^2 = 240 \qquad \sum y^2 = 2403 \qquad \sum xy = 758$$

a. $\hat{\beta}_1 = \dfrac{\sum xy}{\sum x^2} = \dfrac{758}{240} = 3.15833 \approx 3.158$

The fitted model is $\hat{y} = 3.158x$.

b. $SSE = \sum y^2 - \hat{\beta}_1 \sum xy = 2403 - 3.1583333(758) = 8.983359$

$s^2 = \dfrac{SSE}{n-1} = \dfrac{8.983359}{8-1} = 1.283337 \qquad s = \sqrt{1.283337} = 1.1328$

c. To determine if x and y are positively linearly related, we test:

$H_0: \beta_1 = 0$
$H_a: \beta_1 > 0$

The test statistic is $t = \dfrac{\hat{\beta}_1}{s/\sqrt{\sum x^2}} = \dfrac{3.158333}{1.1328/\sqrt{240}} = 43.193$

The rejection region requires $\alpha = 0.05$ in the upper tail of the t distribution. From Table 2, Appendix D, with $df = n - 1 = 8 - 1 = 7$, $t_{0.05} = 1.895$. The rejection region is $t > 1.895$.

Since the observed value of the test statistic falls in the rejection region $(t = 43.193 > 1.895)$, H_0 is rejected. There is sufficient evidence to indicate that x and y are positively linearly related at $\alpha = 0.05$.

d. The form of the confidence interval for β_1 is $\hat{\beta}_1 \pm t_{0.025}\left(\dfrac{s}{\sqrt{\sum x^2}}\right)$.

For confidence coefficient 0.95, $\alpha = 0.05$ and $\alpha / 2 = 0.05 / 2 = 0.025$. From Table 2, Appendix D, with $df = n - 1 = 8 - 1 = 7$, $t_{0.025} = 2.365$. The 95% confidence interval is:

$$\hat{\beta}_1 \pm t_{0.025}\left(\frac{s}{\sqrt{\sum x^2}}\right) \Rightarrow 3.158 \pm 2.365\left(\frac{1.1328}{\sqrt{240}}\right) \Rightarrow 3.158 \pm 0.173 \Rightarrow (2.985,\ 3.331)$$

e. The point estimate for y when $x = 7$ is $\hat{y} = 3.158(7) = 22.106$. The 95% confidence interval for $E(y)$ is:

$$\hat{y} \pm t_{0.025}s\left(\sqrt{\frac{x_p^2}{\sum x^2}}\right) \Rightarrow 22.106 \pm 2.365(1.1328)\left(\sqrt{\frac{7^2}{240}}\right) \Rightarrow 22.106 \pm 1.211 \Rightarrow (20.895,\ 23.317)$$

f. The 95% prediction interval for y is:

$$\hat{y} \pm t_{0.025}s\left(\sqrt{1+\frac{x_p^2}{\sum x^2}}\right) \Rightarrow 22.106 \pm 2.365(1.1328)\left(\sqrt{1+\frac{7^2}{240}}\right)$$

$$\Rightarrow 22.106 \pm 2.940 \Rightarrow (19.166,\ 25.046)$$

3.69 a. The results of the preliminary calculations are provided below:

$$n = 5,\ \sum x^2 = 30,\ \sum xy = -278,\ \sum y^2 = 2589$$

Substituting into the formula for $\hat{\beta}_1$, we have $\hat{\beta}_1 = \frac{\sum xy}{\sum x^2} = \frac{-278}{30} = -9.2667$ and the least squares line is $\hat{y} = -9.2667x$.

b. $SSE = \sum y^2 - \hat{\beta}_1 \sum xy = 2589 - (-9.26666677)(-278) = 12.8667$

$$s^2 = \frac{SSE}{n-1} = \frac{12.8667}{5-1} = 3.2167 \quad s = \sqrt{s^2} = \sqrt{3.2167} = 1.7935$$

c. To determine if x and y are negatively linearly related, we test:

$$H_0: \beta_1 = 0$$
$$H_a: \beta_1 < 0$$

Test statistic is $t = \frac{\hat{\beta}_1}{\frac{s}{\sqrt{\sum x^2}}} = \frac{-9.2667}{\frac{1.7935}{\sqrt{30}}} = -28.30.$

The rejection region requires $\alpha = 0.05$ in the lower tail of the t distribution. From Table 2, Appendix D, with $df = n - 1 = 5 - 1 = 4$, $t_{0.05} = 2.132$. The rejection region is $t < -2.132$.

Since the observed value of the test statistic falls in the rejection region $(t = -28.30 < -2.132)$, H_0 is rejected. There is sufficient evidence to indicate that x and y are negatively linearly related at $\alpha = 0.05$.

d. The form of the confidence interval for β_1 is $\hat{\beta}_1 \pm t_{0.025}\left(\frac{s}{\sqrt{\sum x^2}}\right)$.

For confidence coefficient 0.95, $\alpha = 0.05$ and $\alpha/2 = 0.05/2 = 0.025$. From Table 2, Appendix D, with $df = n - 1 = 5 - 1 = 4$, $t_{0.025} = 2.776$. The 95% confidence interval is:

$$\hat{\beta}_1 \pm t_{0.025}\left(\frac{s}{\sqrt{\sum x^2}}\right) \Rightarrow -9.267 \pm 2.776\left(\frac{1.7935}{\sqrt{30}}\right) \Rightarrow -9.267 \pm 0.909 \Rightarrow (-10.176, -8.358).$$

e. The point estimate for y when $x = 1$ is $\hat{y} = \hat{\beta}_1 x = -9.267(1) = -9.267$. The 95% confidence interval for $E(y)$ is:

$$\hat{y} \pm t_{0.025}s\left(\sqrt{\frac{x_p^2}{\sum x^2}}\right) \Rightarrow -9.267 \pm 2.776(1.7935)\left(\sqrt{\frac{1}{30}}\right) \Rightarrow -9.267 \pm 0.909$$

$$\Rightarrow (-10.176, -8.358).$$

f. The 95% prediction interval for y is:

$$\hat{y} \pm t_{0.025}s\left(\sqrt{1 + \frac{x_p^2}{\sum x^2}}\right) \Rightarrow -9.267 \pm 2.776(1.7935)\left(\sqrt{1 + \frac{1^2}{30}}\right)$$

$$\Rightarrow -9.267 \pm 5.061 \Rightarrow (-14.328, -4.206)$$

3.70 a. The results of the preliminary calculations are provided below:

$$\sum x = 1140 \quad \sum x^2 = 158{,}400 \quad \sum y = 236 \quad \sum xy = 33{,}020 \quad \sum y^2 = 6906$$

Substituting into the formula for $\hat{\beta}_1$, we have $\hat{\beta}_1 = \frac{\sum xy}{\sum x^2} = \frac{33{,}020}{158{,}400} = 0.208459596 \approx 0.2085$ and the least squares line is $\hat{y} = 0.2085x$.

b. $SSE = \sum y^2 - \hat{\beta}_1 \sum xy = 6906 - (0.208459596)(33{,}020) = 22.664$

$$s^2 = \frac{SSE}{n-1} = \frac{22.664}{10-1} = 2.5182 \quad s = \sqrt{s^2} = \sqrt{2.5182} = 1.5869$$

c. To determine if *x* and *y* are positively linearly related, we test:

$$H_0: \beta_1 = 0$$
$$H_a: \beta_1 > 0$$

The test statistic is $t = \dfrac{\hat{\beta}_1}{\dfrac{s}{\sqrt{\sum x^2}}} = \dfrac{0.2085}{\dfrac{1.5869}{\sqrt{158,400}}} = 52.29.$

The rejection region requires $\alpha = 0.05$ in the upper tail of the *t* distribution. From Table 2, Appendix D, with $df = n-1 = 10-1 = 9$, $t_{0.05} = 1.833$. The rejection region is $t > 1.833$.

Since the observed value of the test statistic falls in the rejection region $(t = 52.29 > 1.833)$, H_0 is rejected. There is sufficient evidence to indicate that *x* and *y* are positively linearly related at $\alpha = 0.05$.

d. The form of the confidence interval for β_1 is $\hat{\beta}_1 \pm t_{0.025}\left(\dfrac{s}{\sqrt{\sum x^2}}\right)$.

For confidence coefficient 0.95, $\alpha = 0.05$ and $\alpha/2 = 0.05/2 = 0.025$. From Table 2, Appendix D, with $df = n-1 = 10-1 = 9$, $t_{0.025} = 2.262$. The 95% confidence interval is:

$$\hat{\beta}_1 \pm t_{0.025}\left(\frac{s}{\sqrt{\sum x^2}}\right) \Rightarrow 0.2085 \pm 2.262\left(\frac{1.5869}{\sqrt{158,400}}\right) \Rightarrow 0.2085 \pm 0.0090 \Rightarrow (0.1995,\ 0.2175).$$

e. The point estimate for *y* when $x = 125$ is $\hat{y} = \hat{\beta}_1 x = 0.2085(125) = 26.06$. The 95% confidence interval for $E(y)$ is:

$$\hat{y} \pm t_{0.025} s\left(\sqrt{\frac{x_p^2}{\sum x^2}}\right) \Rightarrow 26.06 \pm 2.262(1.5869)\left(\sqrt{\frac{125^2}{158,400}}\right) \Rightarrow 26.06 \pm 1.13$$
$$\Rightarrow (24.93,\ 27.19).$$

f. The 95% prediction interval for *y* is:

$$\hat{y} \pm t_{0.025} s\left(\sqrt{1 + \frac{x_p^2}{\sum x^2}}\right) \Rightarrow 26.06 \pm 2.262(1.5869)\left(\sqrt{1 + \frac{125^2}{158,400}}\right)$$
$$\Rightarrow 26.06 \pm 3.76 \Rightarrow (22.30,\ 29.82)$$

3.71 a. Some preliminary calculations are:

$n = 8 \quad \sum x^2 = 59.75 \quad \sum xy = 320.5 \quad \sum y^2 = 1738$

Then, $\hat{\beta}_1 = \frac{\sum xy}{\sum x^2} = \frac{320.5}{59.75} = 5.364016736 \approx 5.364$, and the least squares line is $\hat{y} = 5.364x$.

b. To determine if there is a linear relationship between drug dosage and decrease in pulse rate, we test:

$$H_0 : \beta_1 = 0$$
$$H_a : \beta_1 \neq 0$$

The test statistic is $t = \dfrac{\hat{\beta}_1}{\dfrac{s}{\sqrt{\sum x^2}}}$

where $s = \sqrt{s^2} = \sqrt{\dfrac{SSE}{n-1}} = \sqrt{\dfrac{\sum y^2 - \hat{\beta}_1 \sum xy}{n-1}} = \sqrt{\dfrac{1738 - (5.364)(320.5)}{8-1}} = 1.640$

Substituting, we have $t = \dfrac{5.364}{\dfrac{1.640}{\sqrt{59.75}}} = 25.28.$

The rejection region requires $\alpha / 2 = 0.10 / 2 = 0.05$ in each tail of the t distribution. From Table 2, Appendix D, with $df = n - 1 = 8 - 1 = 7$, $t_{0.05} = 1.895$. The rejection region is $t < -1.895$ or $t > 1.895$.

Since the observed value of the test statistic falls in the rejection region $(t = 25.28 > 1.895)$, H_0 is rejected. There is sufficient evidence to indicate that drug dosage and decrease in pulse rate are linearly related at $\alpha = 0.10$.

c. We want to predict the decrease in pulse rate y corresponding to a drug dosage of $x_p = 3.5$ cubic centimeters. First, we obtain the point estimate:

$$\hat{y} = \hat{\beta}_1 x = 5.364(3.5) = 18.774$$

For confidence coefficient 0.99, $\alpha = 0.01$ and $\alpha / 2 = 0.01 / 2 = 0.005$. From Table 2, Appendix D, with $df = n - 1 = 8 - 1 = 7$, $t_{0.005} = 3.499$. The 99% confidence interval is:

$$\hat{y} \pm t_{0.005} s \sqrt{1 + \frac{x_p^2}{\sum x^2}} \Rightarrow 18.774 \pm 3.499(1.640)\sqrt{1 + \frac{(3.5)^2}{59.75}} \Rightarrow 18.774 \pm 6.299$$
$$\Rightarrow (12.475, 25.073).$$

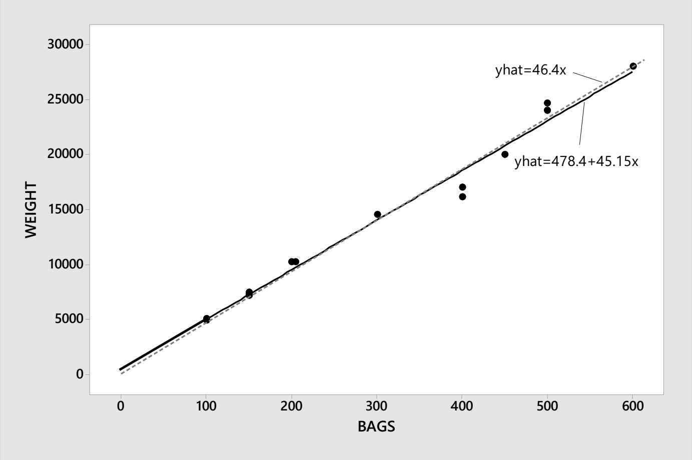
Therefore, we predict the decrease in pulse rate corresponding to a dosage of 3.5cc to fall between 12.475 and 25.073 beats/minute with 99% confidence.

3.72 Some preliminary calculations are:

$\sum x = 4305 \quad \sum x^2 = 1{,}652{,}025 \quad \sum y = 201{,}558 \quad \sum y^2 = 3{,}571{,}211{,}200 \quad \sum xy = 76{,}652{,}695$

a. $\hat{\beta}_1 = \dfrac{\sum xy}{\sum x^2} = \dfrac{76{,}652{,}695}{1{,}652{,}025} = 46.39923427 \approx 46.3992,$ and the least squares line is $\hat{y} = 46.3992x.$

Using MINITAB, the scatterplot of the data with the fitted line is:



b. $SSxx = \sum x^2 - \dfrac{(\sum x)^2}{n} = 1{,}652{,}025 - \dfrac{(4305)^2}{15} = 416{,}490$

$SS_{xy} = \sum xy - \dfrac{(\sum x)(\sum y)}{n} = 76{,}652{,}695 - \dfrac{(4305)(201{,}558)}{15} = 18{,}805{,}549$

$\hat{\beta}_1 = \dfrac{SS_{xy}}{SS_{xx}} = \dfrac{18{,}805{,}549}{416{,}490} = 45.15246224 \approx 45.152$

$\hat{\beta}_0 = \bar{y} - \hat{\beta}_1\bar{x} = \dfrac{201{,}558}{15} - 45.1524622\left(\dfrac{4305}{15}\right) = 478.443$

The fitted line is $\hat{y} = 478.443 + 45.152x.$

c. Since 0 is not contained in the observed range of values of the number of 50-pound bags in the shipment, $\hat{\beta}_0$ has no practical interpretation. Therefore, a value of $\hat{\beta}_0$ that differs from 0 is not unexpected.

d. First, we need to compute s.

$SS_{yy} = \sum y^2 - \dfrac{(\sum y)^2}{n} = 3{,}571{,}211{,}200 - \dfrac{(201{,}558)^2}{15} = 862{,}836{,}042$

$SSE = SS_{yy} - \hat{\beta}_1 SS_{xy} = 862{,}836{,}042 - 45.15246224(18{,}805{,}549) = 13{,}719{,}200.9$

$$s^2 = \frac{SSE}{n-2} = \frac{13,719,200.9}{15-2} = 1,055,323.146 \qquad s = \sqrt{1,055,323.146} = 1027.2892$$

To determine if β_0 should be included in the model, we test:

$$H_0 : \beta_0 = 0$$
$$H_a : \beta_0 \neq 0$$

The test statistic is $t = \dfrac{\hat{\beta}_0}{s\sqrt{\dfrac{1}{n} + \dfrac{\bar{x}^2}{SS_{xx}}}} = \dfrac{478.4}{1027.289\sqrt{\dfrac{1}{15} + \dfrac{287^2}{416,490}}} = 0.906.$

The rejection region requires $\alpha / 2 = 0.10 / 2 = 0.05$ in each tail of the t distribution. From Table 2, Appendix D, with $df = n - 2 = 15 - 2 = 13$, $t_{0.05} = 1.771$. The rejection region is $t < -1.771$ or $t > 1.771$.

Since the observed value of the test statistic does not fall in the rejection region $\left(t = 0.906 \not> 1.771\right)$, H_0 is not rejected. There is insufficient evidence to indicate that β_0 should be included in the model at $\alpha = 0.10$.

3.73 a. Some preliminary calculations are:

$$n = 10 \qquad \sum x^2 = 1{,}933{,}154 \qquad \sum xy = 98{,}946{,}257 \qquad \sum y^2 = 5{,}066{,}358{,}119$$

Then, $\hat{\beta}_1 = \dfrac{\sum xy}{\sum x^2} = \dfrac{98,946,257}{1{,}933{,}154} = 51.18384619 \approx 51.184$, and the least squares prediction equation is $\hat{y} = 51.184x$.

b. To determine if population contributes to the prediction of electricity customers, we test:

$$H_0 : \beta_1 = 0$$
$$H_a : \beta_1 \neq 0$$

The test statistic is $t = \dfrac{\hat{\beta}_1}{\dfrac{s}{\sqrt{\sum x^2}}}$

$$\text{where } s = \sqrt{s^2} = \sqrt{\frac{SSE}{n-1}} = \sqrt{\frac{\left(\sum y^2 - \hat{\beta}_1 \sum xy\right)}{n-1}}$$
$$= \sqrt{\frac{5,066,358,119 - 51.18385(98,946,257)}{10-1}} = 460.4036$$

Substituting, we have $t = \dfrac{51.18}{460.4036/\sqrt{1{,}933{,}154}} = 154.56$

The rejection region requires $\alpha/2 = 0.01/2 = 0.005$ in each tail of the t distribution. From Table 2, Appendix D, with $df = n-1 = 10-1 = 9$, $t_{0.005} = 3.250$. The rejection region is $t < -3.250$ or $t > 3.250$.

Since the observed value of the test statistic falls in the rejection region $(t = 154.56 > 3.250)$, H_0 is rejected. There is sufficient evidence to indicate that population contributes to the prediction of electricity customers at $\alpha = 0.01$.

c. We need the following additional information:

$\sum x = 4286 \quad \sum y = 220{,}297 \quad SS_{xx} = 96{,}174.4 \quad SS_{xy} = 4{,}526{,}962.8 \quad SS_{yy} = 213{,}281{,}298$

$\hat{\beta}_1 = 47.07 \quad \hat{\beta}_0 = 1855.35 \quad SSE = 195{,}568.4 \quad s^2 = 24{,}446.05 \quad s = 156.3523$

The least squares prediction equation is $\hat{y} = 1855.35 + 47.07x$.

To determine if population contributes to the prediction of electricity customers, we test:

$$H_0: \beta_1 = 0$$
$$H_a: \beta_1 \neq 0$$

The test statistic is $t = \dfrac{\hat{\beta}_1}{s/\sqrt{\text{SS}_{xx}}} = \dfrac{47.07}{156.3523/\sqrt{96{,}174.4}} = 93.36$

The rejection region requires $\alpha/2 = 0.01/2 = 0.005$ in each tail of the t distribution. From Table 2, Appendix D, with $df = n-2 = 10-2 = 8$, $t_{0.005} = 3.355$. The rejection region is $t < -3.355$ or $t > 3.355$.

Since the observed value of the test statistic falls in the rejection region $(t = 93.36 > 3.355)$, H_0 is rejected. There is sufficient evidence to indicate that population contributes to the prediction of electricity customers at $\alpha = 0.01$.

d. Without running a formal test, we can compare the two models. The value of s for the model $y = \beta_1 x + \varepsilon$ is 460.4036 while the value of s for the model $y = \beta_0 + \beta_1 x + \varepsilon$ is 156.3523. Since the value of s is much smaller for the second model, it appears that the second model should be used.

For a formal test, refer to part (d) of Exercise 3.66.

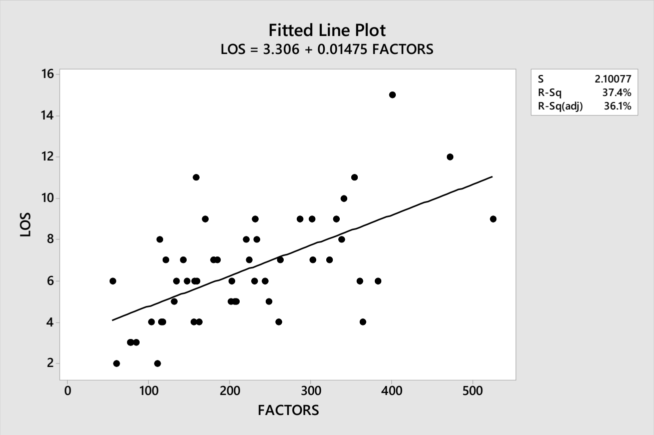
$$H_0: \beta_0 = 0$$
$$H_a: \beta_0 \neq 0$$

The test statistic is $t = \dfrac{\hat{\beta}_0 - 0}{s\sqrt{\dfrac{1}{n} + \dfrac{\overline{x}^2}{SS_{xx}}}} = \dfrac{1855.35}{156.3523\sqrt{\dfrac{1}{10} + \dfrac{428.6^2}{96{,}174.4}}} = 8.37$

The rejection region requires $\alpha / 2 = 0.01/2 = 0.005$ in each tail of the t distribution. From Table 2, Appendix D, with $df = n - 1 = 10 - 1 = 9$, $t_{0.005} = 3.250$. The rejection region is $t < -3.250$ or $t > 3.250$.

Since the observed value of the test statistic falls in the rejection region $(t = 8.37 > 3.250)$, H_0 is rejected. There is sufficient evidence to indicate that β_0 should be included in the model at $\alpha = 0.01$.

3.74 a. Using MINITAB, the scatterplot is:



b. From the printout, the least squares line is $\hat{y} = 3.306 + 0.01475x$.

c. For every one unit increase in the number of factors per patient, we estimate the patient's length of stay to increase 0.01475 days.

d. To determine if the number of factors per patient contributes information for the prediction of the patient’s length of stay, we test:

$$H_0: \beta_1 = 0$$
$$H_a: \beta_1 \neq 0$$

The test statistic is $t = 5.36$ and the p-value is $p < 0.0001$. Since the p-value is less than $\alpha\,(p < 0.0001 < 0.05)$, H_0 is rejected. There is sufficient evidence to indicate that the number of factors per patient contributes information for the prediction of the patient’s length of stay at $\alpha = 0.05$.

e. From the printout, the 95% confidence interval is $(0.00922, 0.02029)$. We are 95% confident that for each additional factor per patient, the patient's length of stay will increase between 0.00917 and 0.02033 days.

f. $r = \sqrt{0.3740} = 0.6116$ There appears to be a moderate positive linear relationship between the number of factors and the length of stay.

g. $r^2 = 0.3740$ 37.4% of the variability around the mean length of stay can be explained by the linear relationship between the number of factors and the length of stay.

h. From the printout, the 95% prediction interval is $(2.44798, 10.98081)$.

i. There is a significant linear relationship between length of stay and the number of factors. However, the value of r^2 is only $r^2 = 0.3740$. Thus, only a little over a third of the variability in the lengths of stays is explained by the model. Many other variables could be affecting the lengths of stay other than the number of factors.

3.75 a. $y = \beta_0 + \beta_1 x + \varepsilon$

b. A value of $r = 0.68$ indicates a moderate positive linear relationship between RMP and SET ratings.

c. The slope is positive since the correlation coefficient is positive.

d. Since the p-value is so small $(p = 0.001)$, H_0 is rejected for any value of $\alpha > 0.001$. This indicates that there is a significant correlation between RMP and SET ratings.

e. $r^2 = (0.68)^2 = 0.4624$ 46.24% of the variability of the sample SET ratings about their mean can be explained by the linear relationship between the SET ratings and the RMP ratings.

3.76 a. Yes. For the men, as the year increases, the winning time tends to decrease. The straight-line model is $y = \beta_0 + \beta_1 x + \varepsilon$. We would expect the slope to be negative.

b. Yes. For the women, as the year increases, the winning time tends to decrease. The straight-line model is $y = \beta_0 + \beta_1 x + \varepsilon$. We would expect the slope to be negative.

c. Since the slope of the women's line is steeper than that for the men, the slope of the women's line will be greater in absolute value.

d. No. The gathered data is from 1880 to 2000. Using this data to predict the time for the year 2020 would be very risky. We have no idea what the relationship between time and year will be outside the observed range. Thus, we would not recommend using this model.

3.77 Using MINITAB, the analyses are:

```
Analysis of Variance
Source        DF   Adj SS   Adj MS   F-Value   P-Value
Regression     1    72.04    72.04      7.11     0.056
  DIAMETER     1    72.04    72.04      7.11     0.056
Error          4    40.55    10.14
Total          5   112.59

Model Summary
      S     R-sq   R-sq(adj)   R-sq(pred)
3.18403   63.98%      54.98%        0.00%

Coefficients
Term        Coef   SE Coef         90% CI        T-Value   P-Value    VIF
Constant    6.35      3.90    (-1.97, 14.68)       1.63     0.179
DIAMETER   0.950     0.356   (0.190, 1.709)        2.67     0.056   1.00

Regression Equation
POROSITY   =    6.35 + 0.950 DIAMETER

Settings
Variable     Setting
DIAMETER          10

Prediction
    Fit      SE Fit          90% CI                 90% PI
15.8501    1.30529   (13.0674, 18.6327)   (8.51395, 23.1862)
```

a. The least squares line is $\hat{y} = 6.35 + 0.950x$.

b. $\hat{\beta}_0 = 6.35$ Since 0 is not in the range of observed values for diameter, $\hat{\beta}_0$ has no meaning.

c. From the printout the 90% confidence interval is $(0.190, 1.709)$. We are 90% confident that for each unit increase in diameter, the mean porosity will increase from 0.190 and 1.709 units.

d. From the printout, the 90% prediction interval is $(8.514, 23.186)$.

3.78 Using MINITAB, the analyses are:

```
Analysis of Variance
Source          DF   Seq SS   Contribution   Adj SS    Adj MS    F-Value   P-Value
Regression       1   0.2330         39.37%   0.2330   0.23300       9.09     0.009
  EMPATHY        1   0.2330         39.37%   0.2330   0.23300       9.09     0.009
Error           14   0.3588         60.63%   0.3588   0.02563
  Lack-of-Fit   10   0.2557         43.20%   0.2557   0.02557       0.99     0.552
  Pure Error     4   0.1031         17.42%   0.1031   0.02578
Total           15   0.5918        100.00%
```

```
Model Summary
        S     R-sq  R-sq(adj)     PRESS  R-sq(pred)
 0.160084   39.37%     35.04%  0.484291      18.16%

Coefficients
Term         Coef  SE Coef            95% CI  T-Value  P-Value   VIF
Constant   -0.392    0.220   (-0.864, 0.079)    -1.79    0.096
EMPATHY    0.0362   0.0120  (0.0104, 0.0619)     3.02    0.009  1.00

Regression Equation
ACTIVITY  =  -0.392 + 0.0362 EMPATHY
```

To determine if people scoring higher in empathy show higher pain-related brain activity, we test:

$$H_0 : \beta_1 = 0$$
$$H_a : \beta_1 > 0$$

The test statistic is $t = 3.02$ and the p-value is $p = 0.009 / 2 = 0.0045$. Since the p-value is very small, H_0 is rejected for any value of $\alpha > 0.0045$. There is sufficient evidence to indicate that people scoring higher in empathy show higher pain-related brain activity at $\alpha > 0.0045$.

3.79 a. Since the p-value for the SG score is $p = 0.739$ and is larger than the significance level of 0.05, then we cannot conclude that ESLR score is linearly related to the SG score.

b. Since the p-value for the SR score is $p = 0.012$ and is smaller than the significance level of 0.05, then we can conclude that ESLR score is linearly related to the SR score.

c. Since the p-value for the ER score is $p = 0.022$ and is smaller than the significance level of 0.05, then we can conclude that ESLR score is linearly related to ER score.

d. $100\left(r^2\right)\%$ of the sample variation in ESLR score can be explained by the linear relationship between ESLR and x (SG, SR, or ER score)

a. 0.2% of the sample variation in ESLR scores around their means can be explained by the linear relationship between ESLR and SG scores.

b. 9.9% of the sample variation in ESLR scores around their means can be explained by the linear relationship between ESLR and SR scores.

c. 7.8% of the sample variation in ESLR scores around their means can be explained by the linear relationship between ESLR and ER scores.

3.80 a. Using MINITAB, the results of the analyses regressing the blood plasma level of 2,3,7,8-TCDD on the fat tissue level of 2,3,7,8-TCDD are:

```
Analysis of Variance
Source        DF   Adj SS    Adj MS   F-Value   P-Value
Regression     1  1105.19   1105.19    132.05     0.000
  FAT          1  1105.19   1105.19    132.05     0.000
Error         18   150.65      8.37
  Lack-of-Fit 15   137.85      9.19      2.15     0.289
  Pure Error   3    12.81      4.27
Total         19  1255.84

Model Summary
      S     R-sq   R-sq(adj)   R-sq(pred)
2.89303   88.00%      87.34%       80.90%

Coefficients
Term        Coef   SE Coef   T-Value   P-Value    VIF
Constant  -0.150     0.841     -0.18     0.860
FAT       0.9009    0.0784     11.49     0.000   1.00

Regression Equation
PLASMA  =  -0.150 + 0.9009 FAT
```

The fitted prediction equation is $\hat{y} = -0.150 + 0.9009x$.

Using MINITAB, the results of the analyses regressing the fat tissue level of 2,3,7,8-TCDD on the blood plasma level of 2,3,7,8-TCDD are:

```
Analysis of Variance
Source        DF   Adj SS    Adj MS   F-Value   P-Value
Regression     1  1198.32   1198.32    132.05     0.000
  PLASMA       1  1198.32   1198.32    132.05     0.000
Error         18   163.35      9.07
  Lack-of-Fit 15   154.56     10.30      3.52     0.164
  Pure Error   3     8.79      2.93
Total         19  1361.67

Model Summary
      S     R-sq   R-sq(adj)   R-sq(pred)
3.01245   88.00%      87.34%       80.90%

Coefficients
Term        Coef   SE Coef   T-Value   P-Value    VIF
Constant   0.970     0.846      1.15     0.267
PLASMA    0.9768    0.0850     11.49     0.000   1.00

Regression Equation
FAT  =  0.970 + 0.9768 PLASMA
```

The fitted prediction equation is $\hat{y} = 0.970 + 0.9768x$.

b. To determine if fat tissue level is a useful predictor of blood plasma level, we test:

$$H_0: \beta_1 = 0$$
$$H_a: \beta_1 \neq 0$$

The test statistic is $t = 11.49$ and the p-value is $p = 0.000$. Since the p-value is less than $\alpha\left(p = 0.000 < 0.05\right)$, H_0 is rejected. There is sufficient evidence to indicate fat tissue level is a useful predictor of blood plasma level at $\alpha = 0.05$.

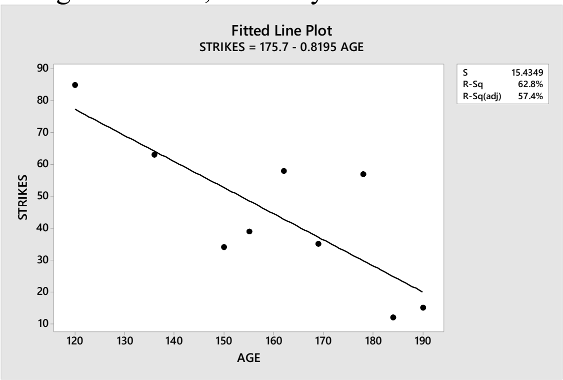
c. To determine if blood plasma level is a useful predictor of fat tissue level, we test:

$$H_0: \beta_1 = 0$$
$$H_a: \beta_1 \neq 0$$

The test statistic is $t = 11.49$ and the p-value is $p = 0.000$. Since the p-value is less than $\alpha\left(p = 0.000 < 0.05\right)$, H_0 is rejected. There is sufficient evidence to indicate blood plasma level is a useful predictor of fat tissue level at $\alpha = 0.05$.

d. If we fit a least squares line through the data, the relationship will be the same regardless of which variable is the dependent variable and which variable is the independent variable. The correlation coefficient and the coefficient of determination will be the same regardless of which variable is the dependent variable and which variable is the independent variable.

3.81 Using MINITAB, the analyses of the data are:



Analysis of Variance

Source	DF	Seq SS	Contribution	Adj SS	Adj MS	F-Value	P-Value
Regression	1	2810	62.76%	2810	2809.9	11.79	0.011
AGE	1	2810	62.76%	2810	2809.9	11.79	0.011
Error	7	1668	37.24%	1668	238.2		
Total	8	4478	100.00%				

Model Summary

S	R-sq	R-sq(adj)	PRESS	R-sq(pred)
15.4349	62.76%	57.43%	2582.04	42.33%

Coefficients

Term	Coef	SE Coef	95% CI	T-Value	P-Value	VIF
Constant	175.7	38.6	(84.4, 267.0)	4.55	0.003	
AGE	-0.819	0.239	(-1.384, -0.255)	-3.43	0.011	1.00

Regression Equation

```
STRIKES  =   175.7 - 0.819 AGE
```

a. The fitted regression line is $\hat{y} = 175.7 - 0.819x$.

b. We see from the plot that there appears to be a moderate negative linear relationship between age and the mean number of strikes.

$\hat{\beta}_0 = 175.7$ Since 0 is not in the observed range of values of age, $\hat{\beta}_0$ has no meaning.

$\hat{\beta}_1 = -0.819$ For each additional day of age for the fish, we estimate that the mean number of strikes will decrease by 0.819 strikes.

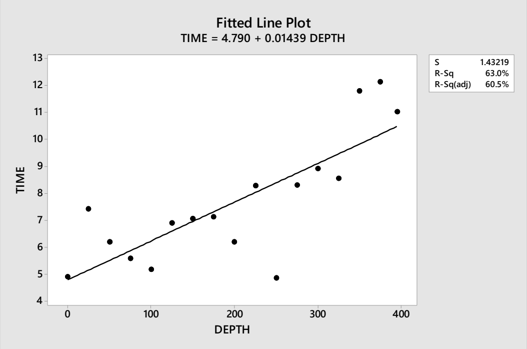
To determine if there is a linear relationship between age of fish and number of strikes, we test:

$$H_0 : \beta_1 = 0$$
$$H_a : \beta_1 \neq 0$$

The test statistic is $t = -3.43$ and the p-value is $p = 0.011$. Since the p-value is less than $\alpha\left(p = 0.011 < 0.05\right)$, H_0 is rejected. There is sufficient evidence to indicate there is a linear relationship between age of fish and number of strikes at $\alpha = 0.05$.

$r^2 = 0.6276$ 62.76% of the variability of the mean number of strikes about their mean is explained by the linear relationship between age and number of strikes.

3.82 Using MINITAB, a scatterplot of the data is:



There appears to be a linear relationship between the time to drill 5 feet and the depth at which drilling begins.

Using MINITAB, the analyses of the data are:

Analysis of Variance

Source	DF	Adj SS	Adj MS	F-Value	P-Value
Regression	1	52.38	52.378	25.54	0.000
DEPTH	1	52.38	52.378	25.54	0.000
Error	15	30.77	2.051		
Total	16	83.15			

```
Model Summary
        S     R-sq  R-sq(adj)  R-sq(pred)
  1.43219   63.00%     60.53%      52.23%

Coefficients
Term          Coef  SE Coef  T-Value  P-Value   VIF
Constant     4.790    0.666     7.19    0.000
DEPTH      0.01439  0.00285     5.05    0.000  1.00

Regression Equation
TIME  =  4.790 + 0.01439 DEPTH
```

The fitted regression line is $\hat{y} = 4.790 + 0.01439x$.

$\hat{\beta}_0 = 4.790$ We estimate the mean time to drill 5 feet when starting at a depth of 0 feet is 4.79 minutes.

$\hat{\beta}_1 = 0.01439$ For each additional foot of depth, we estimate that the mean time to drill 5 feet will increase by 0.0.01439 minutes.

To determine if there is a linear relationship between depth and time, we test:

$$H_0 : \beta_1 = 0$$
$$H_a : \beta_1 \neq 0$$

The test statistic is $t = 5.05$ and the p-value is $p = 0.000$. Since the p-value is less than $\alpha\left(p = 0.000 < 0.05\right)$, H_0 is rejected. There is sufficient evidence to indicate there is a linear relationship between depth and time at $\alpha = 0.05$.

$r^2 = 0.6300$ 63.00% of the variability of the mean time to drill 5 feet about their mean is explained by the linear relationship between time to drill and depth that drilling starts.

3.83 a. To determine if body plus head rotation and active head movement are positively linearly related, we test:

$$H_0 : \beta_1 = 0$$
$$H_a : \beta_1 > 0$$

The test statistic is $t = \dfrac{\hat{\beta}_1}{s_{\hat{\beta}_1}} = \dfrac{0.88 - 0}{0.14} = 6.29.$

The rejection region requires $\alpha = 0.05$ in the upper tail of the t distribution with $df = n - 2 = 39 - 2 = 37$. From Table 2, Appendix D, $t_{0.05} \approx 1.687$. The rejection region is $t > 1.687$.

Since the observed value of the test statistic falls in the rejection region $(t = 6.29 > 1.687)$, H_0 is rejected. There is sufficient evidence to indicate that the two variables are positively linearly related at $\alpha = 0.05$.

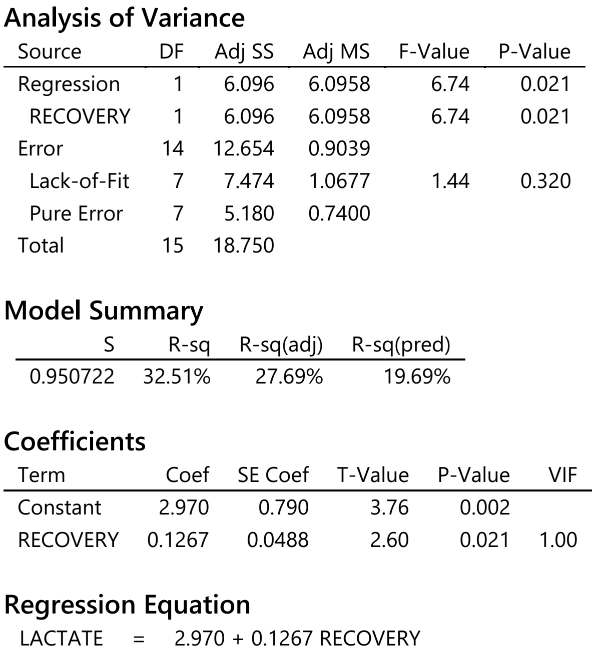
b. For confidence level 0.90, $\alpha = 0.10$ and $\alpha / 2 = 0.10 / 2 = 0.05$. From Table 2, Appendix D, with $df = n - 2 = 39 - 2 = 37$, $t_{0.05} \approx 1.687$. The confidence interval is:

$$\hat{\beta}_1 \pm t_{0.05} s_{\hat{\beta}_1} \Rightarrow 0.88 \pm 1.687(0.14) \Rightarrow 0.88 \pm 0.24 \Rightarrow (0.64, 1.12)$$

We are 90% confident that the true value of β_1 is between 0.64 and 1.12.

c. Because the interval in part b contains the value 1, there is no evidence that the true slope of the line differs from 1.

3.84 Using MINITAB, the analyses of the data are:



```
Analysis of Variance
 Source         DF   Adj SS    Adj MS   F-Value   P-Value
 Regression      1    6.096    6.0958      6.74     0.021
   RECOVERY      1    6.096    6.0958      6.74     0.021
 Error          14   12.654    0.9039
   Lack-of-Fit   7    7.474    1.0677      1.44     0.320
   Pure Error    7    5.180    0.7400
 Total          15   18.750

Model Summary
        S     R-sq   R-sq(adj)   R-sq(pred)
 0.950722   32.51%      27.69%       19.69%

Coefficients
 Term          Coef   SE Coef   T-Value   P-Value    VIF
 Constant     2.970     0.790      3.76     0.002
 RECOVERY    0.1267    0.0488      2.60     0.021   1.00

Regression Equation
 LACTATE  =  2.970 + 0.1267 RECOVERY
```

To determine if blood lactate level is linearly related to perceived recovery, we test:

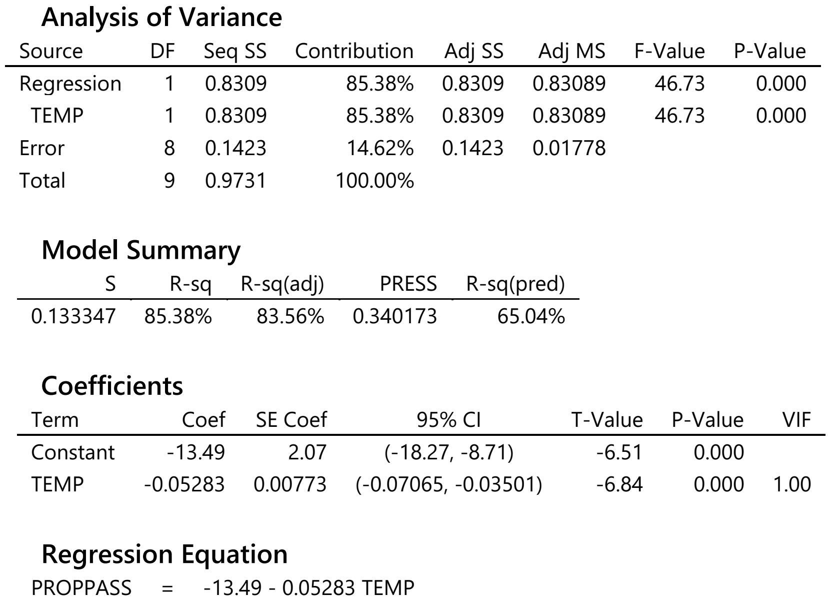
$$H_0 : \beta_1 = 0$$
$$H_a : \beta_1 \neq 0$$

The test statistic is $t = 2.60$ and the p-value is $p = 0.021$. Since the p-value is less than α $(p = 0.021 < 0.10)$, H_0 is rejected. There is sufficient evidence to indicate blood lactate level is linearly related to perceived recovery at $\alpha = 0.10$.

3.85 a. This relationship will have a negative correlation since the researchers claim an "inverse relationship".

b. Solving $t = \frac{r\sqrt{n-2}}{\sqrt{1-r^2}}$ for r using the smallest value of t that leads to a statistically significant result gives: $r^2 = \frac{t^2}{t^2+n-2}$. So if $t = 1.645$ leads to a rejection of $H_0 : \rho = 0$, then $r^2 = \frac{(1.645)^2}{(1.645)^2 + 337 - 2} = .00801$. Thus, $r = -\sqrt{0.00801} = -0.0895$ since r is negative.

3.86 a. Using MINITAB, the results are:



```
Analysis of Variance
Source      DF  Seq SS  Contribution  Adj SS   Adj MS  F-Value  P-Value
Regression   1  0.8309        85.38%  0.8309  0.83089    46.73    0.000
  TEMP       1  0.8309        85.38%  0.8309  0.83089    46.73    0.000
Error        8  0.1423        14.62%  0.1423  0.01778
Total        9  0.9731       100.00%

Model Summary
       S    R-sq  R-sq(adj)     PRESS  R-sq(pred)
0.133347  85.38%     83.56%  0.340173      65.04%

Coefficients
Term          Coef  SE Coef              95% CI  T-Value  P-Value   VIF
Constant    -13.49     2.07      (-18.27, -8.71)    -6.51    0.000
TEMP      -0.05283  0.00773  (-0.07065, -0.03501)   -6.84    0.000  1.00

Regression Equation
PROPPASS  =  -13.49 - 0.05283 TEMP
```

The fitted regression line is $\hat{y} = -13.49 - 0.0528x$.

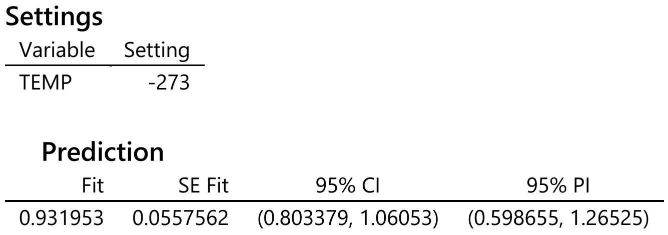
$\hat{\beta}_0 = -13.49$ Since 0 is not within the range of observed value of temperature, $\hat{\beta}_0$ has no meaning.

$\hat{\beta}_1 = -0.0528$ For each degree increase in temperature, the mean proportion of impurity is estimated to decrease by 0.0528.

b. From the printout, the 95% confidence interval for β_1 is $(-0.07065, -0.03501)$. We estimate the mean proportion of impurity will decrease by anywhere from 0.07065 and 0.0351 for each degree increase in temperature. Because 0 is not contained in this interval, there is evidence to indicate that temperature contributes information about the proportions of impurity passing through helium.

c. From the printout, $r^2 = 0.8538$. 85.38% of the variability in the proportion of impurity passing through helium around their means is explained by the linear relationship between the temperature and the proportion of impurity.

d. Using MINITAB, the prediction interval is:



```
Settings
 Variable   Setting
 TEMP          -273

   Prediction
        Fit      SE Fit          95% CI                  95% PI
 0.931953   0.0557562   (0.803379, 1.06053)   (0.598655, 1.26525)
```

The 95% prediction interval is $(0.5987, 1.2653)$. We are 95% confident that the actual proportion of impurities will be between 0.5987 and 1.2653 when the temperature is -273 degrees. Since the proportion cannot be greater than 1, the interval really is $(0.5987, 1.0)$.

e. We have no idea what the relationship between temperature and proportion of impurity looks like outside the observed range.

3.87 a. Piano: $r = 0.447$
Because this value is near 0.5, there is a slight positive linear relationship between recognition exposure time and goodness of view for piano.

Bench: $r = -0.057$
Because this value is extremely close to 0, there is an extremely weak negative linear relationship between recognition exposure time and goodness of view for bench.

Motorbike: $r = 0.619$
Because this value is near 0.5, there is a moderate positive linear relationship between recognition exposure time and goodness of view for motorbike.

Armchair: $r = .294$
Because this value is fairly close to 0, there is a weak positive linear relationship between recognition exposure time and goodness of view for armchair.

Teapot: $r = 0.949$
Because this value is very close to 1, there is a strong positive linear relationship between recognition exposure time and goodness of view for teapot.

b. Piano: $r^2 = (0.447)^2 = 0.1998$
19.98% of the total sample variability around the sample mean recognition exposure time is explained by the linear relationship between the recognition exposure time and the goodness of view for piano.

Bench: $r^2 = (-0.057)^2 = 0.0032$
0.32% of the total sample variability around the sample mean recognition exposure time is explained by the linear relationship between the recognition exposure time and the goodness of view for bench.

Motorbike: $r^2 = (0.619)^2 = 0.3832$

38.32% of the total sample variability around the sample mean recognition exposure time is explained by the linear relationship between the recognition exposure time and the goodness of view for motorbike.

Armchair: $r^2 = (0.294)^2 = 0.0864$

8.64% of the total sample variability around the sample mean recognition exposure time is explained by the linear relationship between the recognition exposure time and the goodness of view for armchair.

Teapot: $r^2 = (0.949)^2 = 0.9006$

90.06% of the total sample variability around the sample mean recognition exposure time is explained by the linear relationship between the recognition exposure time and the goodness of view for teapot.

c. The test is:

$$H_0 : \beta_1 = 0$$
$$H_a : \beta_1 \neq 0$$

Following are the values of α an $t_{\alpha/2}$ that correspond to $df = n - 2 = 25 - 2 = 23$.

α	**0.20**	**0.10**	**0.05**	**0.02**	**0.01**	**0.002**	**0.001**
$t_{\alpha/2}$	1.319	1.714	2.069	2.500	2.807	3.485	3.767

Piano: $t = 2.40$

$2.069 < 2.40 < 2.500 \Rightarrow p \approx 0.025$

For levels of significance greater than $\alpha = 0.025$, H_0 can be rejected. There is sufficient evidence to indicate that there is a linear relationship between goodness of view and recognition exposure time for piano for $\alpha > 0.025$.

Bench: $t = 0.27$

$0.27 < 1.319 \Rightarrow p > 0.2$

H_0 is not rejected. There is insufficient evidence to indicate that there is a linear relationship between goodness of view and recognition exposure time for bench for $\alpha \leq 0.2$.

Motorbike: $t = 3.78$

$3.78 > 3.767 \Rightarrow p < 0.001$

H_0 can be rejected for $\alpha \geq 0.001$. There is sufficient evidence to indicate that there is a linear relationship between goodness of view and recognition exposure time for motorbike for $\alpha \geq 0.001$.

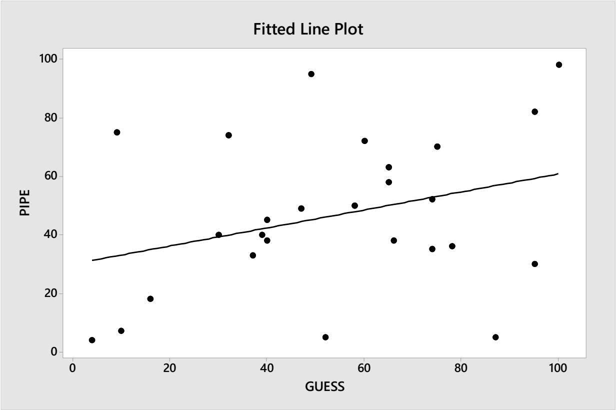
Armchair: $t = 1.47$

$1.319 < 1.47 < 1.717 \Rightarrow p \approx 0.15$

H_0 cannot be rejected for levels of significance $\alpha < 0.15$. There is insufficient evidence to indicate that there is a linear relationship between goodness of view and recognition exposure time for armchair for $\alpha < 0.15$.

Teapot: $t = 14.50$
$14.50 > 3.767 \Rightarrow p < 0.001$
H_0 can be rejected for $\alpha \geq 0.001$. There is sufficient evidence to indicate that there is a linear relationship between goodness of view and recognition exposure time for teapot for $\alpha \geq 0.001$.

3.88 a. Using MINITAB, the scatterplot of the data is:



There is a slight positive linear trend to the data.

b. Using MINITAB, the results are:

Analysis of Variance

Source	DF	Adj SS	Adj MS	F-Value	P-Value
Regression	1	1779	1778.9	2.63	0.118
GUESS	1	1779	1778.9	2.63	0.118
Error	24	16261	677.6		
Lack-of-Fit	20	14728	736.4	1.92	0.278
Pure Error	4	1534	383.4		
Total	25	18040			

Model Summary

S	R-sq	R-sq(adj)	R-sq(pred)
26.0298	9.86%	6.11%	0.00%

Coefficients

Term	Coef	SE Coef	T-Value	P-Value	VIF
Constant	30.1	11.4	2.63	0.015	
GUESS	0.308	0.190	1.62	0.118	1.00

Regression Equation

PIPE = 30.1 + 0.308 GUESS

The fitted regression line is $\hat{y} = 30.1 + 0.308x$.

$\hat{\beta}_0 = 30.1$ Because 0 is not within the observed values of the dowser's guesses, $\hat{\beta}_0$ has no meaning.

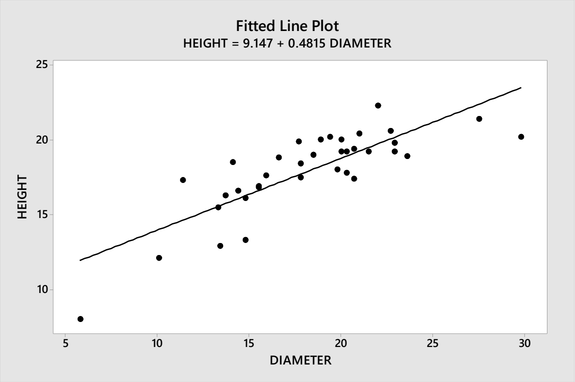
c. To determine if the model is statistically useful for predicting actual pipe location, we test:

$$H_0 : \beta_1 = 0$$
$$H_a : \beta_1 \neq 0$$

The test statistic is $t = 1.62$ and the p-value is $p = 0.118$. Since the p-value is not small, H_0 is not rejected. There is insufficient evidence to indicate the model is statistically useful for predicting actual pipe location at $\alpha < 0.118$.

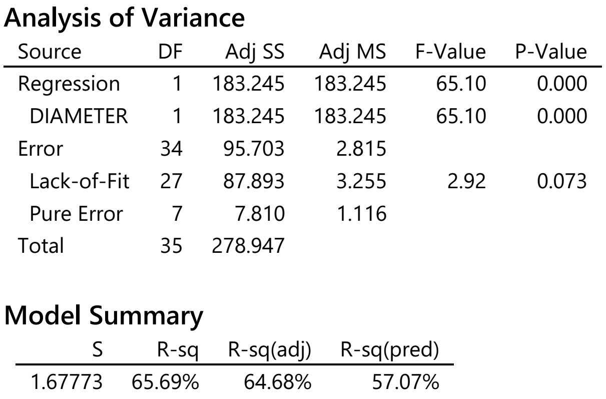
d. Since there is no statistical evidence that there is a linear relationship between the dowsers' guesses and the pipe location, this refutes the conclusion made by the German physicists. In addition, these were the 'best' results of the 'best' dowsers. If there was no relationship between the dowsers' guesses and the pipe location for the 'best' of the 'best', there will not be a relationship between dowsers' guesses and the pipe locations for all of the dowsers.

3.89 a. Using MINITAB, the scatterplot is:



There appears to be a positive linear relationship between breast height diameter and height.

b. Using MINITAB, the results are:



Analysis of Variance

Source	DF	Adj SS	Adj MS	F-Value	P-Value
Regression	1	183.245	183.245	65.10	0.000
DIAMETER	1	183.245	183.245	65.10	0.000
Error	34	95.703	2.815		
Lack-of-Fit	27	87.893	3.255	2.92	0.073
Pure Error	7	7.810	1.116		
Total	35	278.947			

Model Summary

S	R-sq	R-sq(adj)	R-sq(pred)
1.67773	65.69%	64.68%	57.07%

```
Coefficients
Term        Coef    SE Coef  T-Value  P-Value   VIF
Constant    9.15     1.12     8.16     0.000
DIAMETER   0.4815   0.0597    8.07     0.000   1.00

Regression Equation
HEIGHT  =  9.15 + 0.4815 DIAMETER
```

The least squares line is $\hat{y} = 9.15 + 0.4815x$.

$$\hat{\beta}_0 = 9.15$$

$$\hat{\beta}_1 = 0.4815$$

c. The least squares line is printed on the scatterplot in part a.

d. To determine if the breast height diameter contributes information for the prediction of tree height, we test:

$$H_0 : \beta_1 = 0$$
$$H_a : \beta_1 \neq 0$$

The test statistic is $t = 8.07$ and the p-value is $p = 0.000$. Since the p-value is less than $\alpha \left(p = 0.000 < 0.05 \right)$, H_0 is rejected. There is sufficient evidence to indicate the breast height diameter contributes information for the prediction of tree height at $\alpha = 0.05$.

e. Using MINITAB, the results are:

```
Settings
Variable    Setting
DIAMETER       20

Prediction
    Fit      SE Fit          90% CI               90% PI
 18.7763   0.299602   (18.2697, 19.2829)   (15.8945, 21.6581)
```

The 90% confidence interval is $\left(18.2697, 19.2829\right)$. We are 90% confident that the mean height of trees is between 18.2697m and 19.2829m when the breast height diameter is 20cm.