Simple Linear Regression

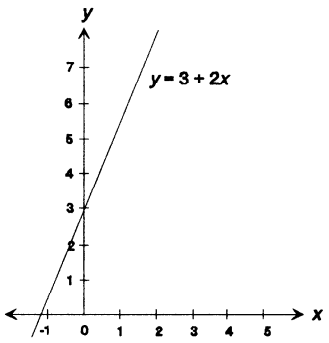
3.2 Since the line passes through the point (0, 1),

$$1 = \beta_0 + \beta_1(0) \Rightarrow \beta_0 = 1$$

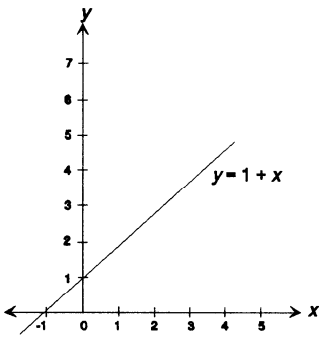
And since it also passes through the point (2, 3),

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

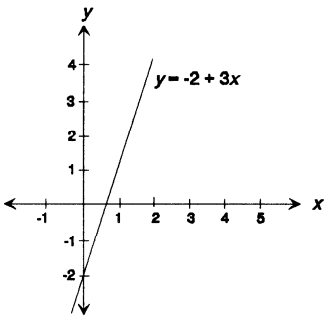
3.4 a.



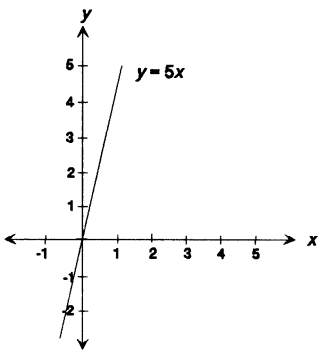
b.



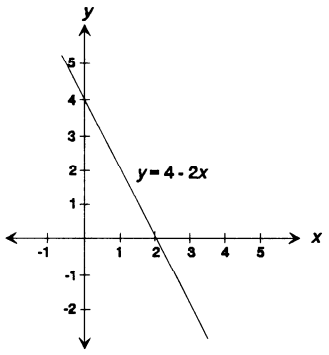
c.



d.



e.



3.6 Summary calculations yield:

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

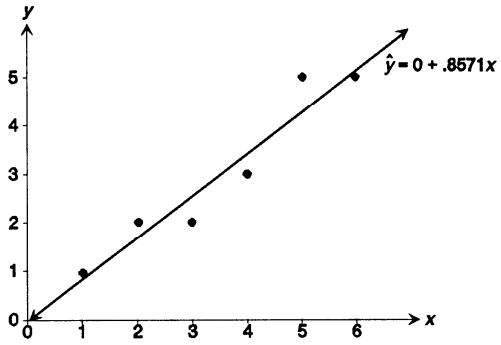
$$\sum y = 18 \qquad \sum y^2 = 68 \qquad \bar{y} = \frac{18}{6} = 3 \qquad \sum xy = 78$$

a. $SS_{xx} = 1 - n(\bar{x})^2 = 91 - 6(3.5)^2 = 91 - 73.5 = 17.5$
$SS_{xy} = 2 - n\bar{x}\bar{y} = 78 - 6(3.5)(3) = 78 - 63 = 15$

$$\hat{\beta}_1 = \frac{SS_{xy}}{SS_{xx}} = \frac{15}{17.5} = .8571$$

$$\hat{\beta}_0 = \bar{y} - \hat{\beta}_1\bar{x} = 3 - (.8571)(3.5) = 0$$

b.



3.8 a. The straight-line model is $y = \beta_0 + \beta_1 x + \varepsilon$.

b. Yes. The data form a rather straight line from the lower left of the plot to the upper right.

c. The fitted model is $\hat{y} = 20.9 + 1.07x$.

d. $\hat{\beta}_0 = 20.9$. The mean sale price when the appraised value is 0 is estimated to be 20.9 or \$20,900. Since $x = 0$ (appraised value = 0) is not in the observed range, this value has no meaning.

e. $\hat{\beta}_1 = 1.07$. For each unit (\$1,000) increase in appraised value, the mean sale price is estimated to increase by 1.07 (\$1,070).

f. \$300,000 $\Rightarrow x = 300$
$\hat{y} = 20.9 + 1.07(300) = 341.9$. The estimated mean sale price for a house appraised at \$300,000 is \$341,900.

3.10 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 that 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.

3.12 a. The equation for the straight-line model is $y = \beta_0 + \beta_1 x + \varepsilon$.

b. Some preliminary calculations are:

$$\bar{y} = \frac{\sum y}{n} = \frac{398}{9} = 44.2222$$

$$\bar{x} = \frac{\sum x}{n} = \frac{1{,}444}{9} = 160.4444$$

$$SS_{xy} = \sum xy - \frac{\sum x \sum y}{n} = 60{,}428 - \frac{1{,}444(398)}{9} = -3{,}428.88889$$

$$SS_{xx} = \sum x^2 - \frac{\left(\sum x\right)^2}{n} = 235{,}866 - \frac{1{,}444^2}{9} = 4{,}184.2222$$

$$\hat{\beta}_1 = \frac{SS_{xy}}{SS_{xx}} = \frac{-3{,}428.88889}{4{,}184.2222} = -.819480593$$

$$\hat{\beta}_o = \bar{y} - \hat{\beta}_1 \bar{x} = \frac{398}{9} - (-.819480593)\left(\frac{1{,}444}{9}\right) = 175.7033307$$

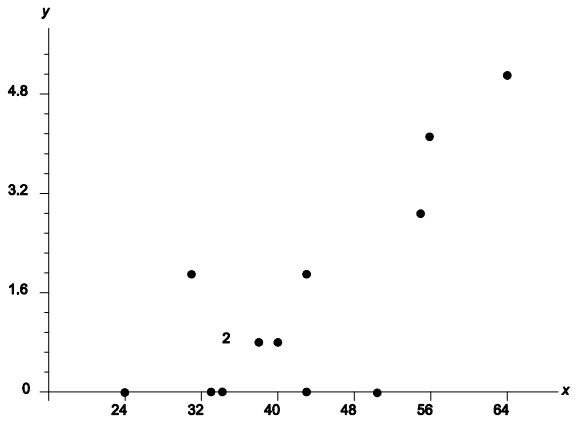
The fitted model is $\hat{y} = 175.7033 - .8195x$.

c. $\hat{\beta}_0 = 175.7033$. Since $x = 0$ (age of fish) is not in the observed range, $\hat{\beta}_0$ has no practical meaning.

d. $\hat{\beta}_1 = -.8195$. For each additional day of age, the mean number of strikes is estimated to decrease by .8195 strikes.

3.14 a. A scattergram of the data is:

From the graph, it appears that as nest box tit occupancy increases the number of flycatchers killed also increases.



b. $\hat{\beta}_0 = -3.04686$. Since $x = 0$ is not in the observed range, $\hat{\beta}_0$ has no interpretation other than the y-intercept.

$\hat{\beta}_1 = 0.10766$. For each additional nest box tit occupancy, the mean number of flycatchers killed is estimated to increase by .10766.

3.16 a. $s^2 = \dfrac{\text{SSE}}{n-2} = \dfrac{.219}{9-2} = .0312857$

b. $s = \sqrt{s^2} = \sqrt{.0312857} = .1769$

We expect most of the observed values of y to fall within $\pm 2s = \pm 2(.1769) = \pm .3538$ of their least squares predicted value.

3.18 $\text{SSE} = \text{SS}_{yy} - \hat{\beta}_1 \text{SS}_{xy}$ $\qquad s^2 = \dfrac{\text{SSE}}{n-2}$, $s = \sqrt{s^2}$

a. From Exercise 3.6, $\text{SS}_{xy} = 15$, $\sum y^2 = 68$, $n = 6$, $\hat{\beta}_1 = .8571$, and $\bar{y} = 3$.

$\text{SS}_{yy} = \sum y^2 - n(\bar{y})^2 = 68 - 6(3)^2 = 68 - 54 = 14$

$\text{SSE} = 14 - (.8571)(15) = 1.1435$

$s^2 = \dfrac{1.1435}{6-2} = .285875$, $s = \sqrt{.285875} = .5347$

We expect most of the sample y-values to fall within $2s = 2(.5347) = 1.0694$ of their least squares predicted values.

b. From Exercise 3.8, SSE = 96,746, $s^2 = 1075$, and $s = 32.79$.
We expect most of the sample sale prices to fall within $2s = (2)(32.79) = 65.58 \Rightarrow$ \$65,580 of their least squares predicted values.

c. From Exercise 3.10, $SS_{yy} = 904.4$, $SS_{xy} = -22.54$, $\hat{\beta}_1 = -7.262$, and $n = 10$.

$SSE = 904.4 - (-7.262)(-22.54) = 740.715$

$s^2 = \dfrac{740.715}{10-2} = 92.59$, $s = \sqrt{92.59} = 9.622$

We expect most of the sample y-values to fall within $2s = 2(9.622) = 19.244$ of their least squares predicted values.

d. From Exercise 3.12, $SS_{yy} = 21.47$, $SS_{xy} = 73.14$, $\hat{\beta}_1 = .2612$, and $n = 15$.

$SSE = 21.47 - (.2612)(73.14) = 2.3655$

$s^2 = \dfrac{2.3655}{15-2} = .18196$, $s = \sqrt{.18196} = .4266$

We expect most of the sample y-values to fall within $2s = 2(.4266) = .8532$ of their least squares predicted values.

3.20 a. Some preliminary calculations are:

$$\sum x = 45.12 \qquad \sum y = 114.6$$

$$\sum x^2 = 88.7788 \qquad \sum y^2 = 575.02 \qquad \sum xy = 225.04$$

$$SS_{xx} = \sum x_i^2 - \frac{\left(\sum x_i\right)}{n} - 88.7788 - \frac{(45.12)^2}{24} = 3.9532$$

$$SS_{xy} = \sum xy - \frac{\left(\sum x\right)\left(\sum y\right)}{n} = 225.04 - \frac{(45.12)(114.6)}{24} = 9.592$$

$$SS_{yy} = \sum y^2 - \frac{\left(\sum y\right)^2}{n} = 575.02 - \frac{(114.6)^2}{24} = 27.805$$

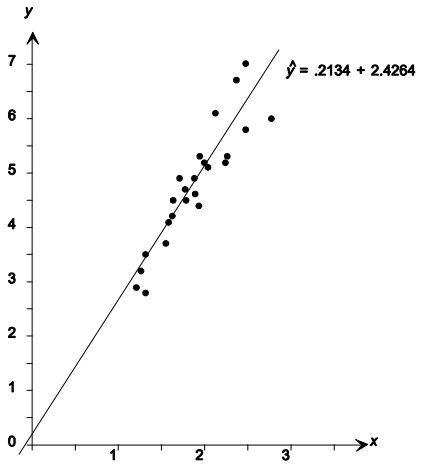
$$\bar{x} = \frac{\sum x}{n} = \frac{45.12}{24} = 1.88 \qquad \bar{y} = \frac{\sum y}{n} = \frac{114.6}{24} = 4.775$$

$$\hat{\beta}_1 = \frac{SS_{xy}}{SS_{xx}} = \frac{9.592}{3.9532} = 2.4264$$

$$\hat{\beta}_0 = \bar{y} - \hat{\beta}_1 \bar{x} = 4.775 - (2.4264)(1.88) = .2134$$

The least squares line is $\hat{y} = .2134 + 2.4264x$.

b.



c. $SSE = SS_{yy} - \hat{\beta} SS_{xy} = 27.805 - (2.4264)(9.592) = 4.531$

$$s^2 = \frac{SSE}{n-2} = \frac{4.531}{24-2} = .206$$

d. $s = \sqrt{s^2} = \sqrt{.206} = .454$

We expect most of the sample y-values to fall within $\pm 2s = \pm 2(.454) = \pm .908$ of their least squares predicted value.

3.22 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 $t = 39.45$. Since this is a one-tailed test, the p-value is half of that on the printout. The p-value is $< .0001/2 = .00005$. Since the p-value is less than α ($p = .00005 < .01$), H_0 is rejected. There is sufficient evidence to indicate a positive linear relationship between appraised property value and sale price at $\alpha = .01$.

b. From the printout, the 95% confidence interval for β_1 is (1.01491, 1.12254). For each unit ($1,000) increase in the appraised value, the mean sale price is estimated to increase anywhere from 1.01491 ($1,014.91) to 1.12254 ($1,122.54).

c. To obtain a narrower confidence interval, we could decrease the level of confidence (i.e. reduce 95% to say 90%).

3.24 a. To determine if y is positively linearly related to x, we test:

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

The test statistic is $t = \dfrac{\hat{\beta}_1 - 0}{s_{\hat{\beta}_1}} = \dfrac{.1077 - 0}{\dfrac{0.2629}{\sqrt{149.9286}}} = 4.004$

The rejection region requires $\alpha = .01$ in the upper tail of the t distribution with df = $n - 2 = 14 - 2 = 12$. From Table 2, Appendix C, $t_{.01} = 2.681$. The rejection region is $t > 2.681$.

Since the observed value of the test statistic falls in the rejection region ($t = 4.004 > 2.681$), H_0 is rejected. There is sufficient evidence to indicate that y is positively linearly related to x at $\alpha = .01$.

b. For confidence level .99, $\alpha = .01$ and $\alpha/2 = .01/2 = .005$. From Table 2, Appendix C, with df = $n - 2 = 14 - 2 = 12$, $t_{.005} = 3.055$. The confidence interval is:

$$\hat{\beta}_1 \pm t_{.005} s_{\hat{\beta}_1} \Rightarrow .1077 \pm 3.055(.0269) \Rightarrow .1077 \pm .0822 \Rightarrow (.0255, .1899)$$

We are 99% confident that the increase in the mean number of flycatchers that are killed for each additional nest box nit increase is between .0255 and .1889. This implies that as the nest box nit occupancy increases, the number of flycatchers killed also increases.

3.26 From 3.20, $\hat{\beta}_1 = 2.4264$ and $SS_{xx} = 3.9532$, $SS_{xy} = 9.592$, $\sum y = 114.6$, $\sum y^2 = 575.02$.

$$SS_{yy} = \sum y^2 - \frac{\left(\sum y\right)^2}{n} = 575.02 - \frac{114.6^2}{24} = 27.805$$

$$SSE = SS_{yy} - \hat{\beta}_1 SS_{xy} = 27.805 - 2.4264(9.592) = 4.531$$

$$s^2 = \frac{SSE}{n-2} = \frac{4.531}{24-2} = .2060 \qquad s = \sqrt{.2060}$$

The confidence interval for β_1 is $\hat{\beta}_1 \pm t_{\alpha/2} s_{\hat{\beta}_1}$ where $s_{\hat{\beta}_1} = \dfrac{s}{\sqrt{SS_{xx}}}$.

For confidence coefficient .95, $\alpha = 1 - .95 = .05$ and $\alpha/2 = .05/2 = .025$. From Table 2 in Appendix C, with df = $n - 2 = 24 - 2 = 22$, $t_{.025} = 2.074$. The 95% confidence interval is:

$$2.4264 \pm 2.074 \frac{.454}{\sqrt{3.9532}} \Rightarrow 2.4264 \pm .4736 \Rightarrow (1.9528, 2.9000)$$

We are 95% confident the mean heat transfer enhancement will increase from between 1.9528 and 2.9000 for each 1 unit increase in unflooded area ratio.

3.28 a. There appears to be a somewhat positive linear relationship.

b. If there was very little snowfall in an area, then the erosion will not be typical. Thus, it seems reasonable to remove these data points.

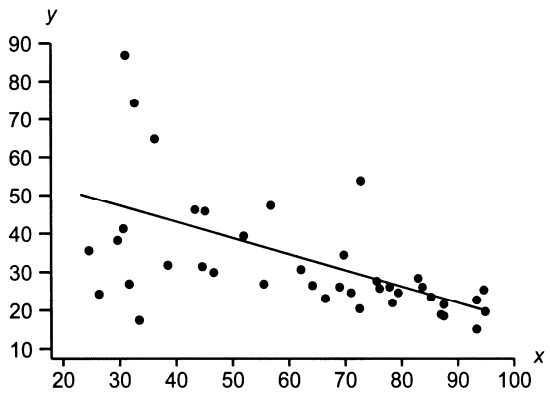
c. For confidence level .90, $\alpha = .10$ and $\alpha/2 = .10/2 = .05$. From Table 2, Appendix C, with df = $n - 2 = 47 - 2 = 45$, $t_{.05} \approx 1.684$. The confidence interval is:

$$\hat{\beta}_1 \pm t_{.05} s_{\hat{\beta}_1} \Rightarrow 1.39 \pm 1.684(.06) \Rightarrow 1.39 \pm .101 \Rightarrow (1.289, 1.491)$$

d. We are 90% confident that the change in the mean McCool winter-adjusted rainfall erosivity index for each one unit change in the once-in-5-year snowmelt runoff amount is between 1.289 and 1.491.

3.30 a. $\hat{\beta}_0 = 57.755,\ \hat{\beta}_1 = -.39961$

b.



c. To determine if country credit risk contributes information for the prediction of market volatility, we test:

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

The test statistic is $t = -4.37$ with a p-value = .000. For any level of significance greater than $\alpha = .000$, H_0 is rejected. There is sufficient evidence to indicate that country credit risk contributes information for the prediction of market volatility at $\alpha > .000$.

d. Answers may vary.
Possible outliers may include the two points (31.8, 87.0) and (32.6, 74.1).

e. Answers may vary.
If the two points in part d are eliminated, the results do not change dramatically. The least squares line is = 47.032 .26333x. The data still provide sufficient evidence to conclude that x contributes information for the prediction of y with a test statistic of $t = -3.76$ and a p-value < .001.

3.32 a. If $r = .7$, there is a positive relationship between x and y. As x increases, y tends to increase. The slope is positive.

b. If $r = -.7$, there is a negative 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 relationship between x and y.

d. If $r^2 = .64$, then r is either .8 or $-.8$. The relationship between x and y could be either positive or negative.

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

3.36 a. From the printout, the coefficient of correlation is $r = .972$. Since this value is close to 1, there is a very strong positive linear relationship between sale price and appraised value.

b. From the printout, the coefficient of determination is R-Sq = 94.5%. This means that 94.5% of the sample variance of the sale prices around the sample mean is explained by the linear relationship between sale price and appraised value.

3.38 a. $r = .14$. Because this value is close to 0, there is a very weak positive linear relationship between math confidence and computer interest for boys.

b. $r = .33$. Because this value is fairly close to 0, there is a weak positive linear relationship between math confidence and computer interest for girls.

3.40 Using the values computed in Exercises 3.15 and 3.29:

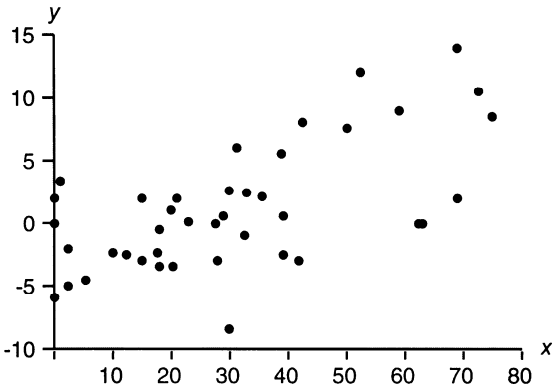
$$r = \frac{SS_{xy}}{\sqrt{SS_{xx}SS_{yy}}} = \frac{19.975}{\sqrt{756(9.700121597)}} = .2333$$

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

$r^2 = .2333^2 = .0544$.

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.

3.42 a.



Yes; There appears to be a positive trend. As digestion efficiency (%) increases, weight change (%) increases.

b. Using MINITAB, $r = .612$. Because this value is near .5, there is a moderate positive linear relationship between the weight change (%) and digestion efficiency (%).

c. To determine if weight change is correlated with digestion efficiency, we test:

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

The test statistic is

$$t = \frac{r}{\sqrt{\frac{1-r^2}{n-2}}} = \frac{.612}{\sqrt{\frac{1-.612^2}{42-2}}} \approx 4.90$$

The rejection region requires $\alpha/2 = .01/2 = .005$ in each tail of the t distribution with df $= n - 2 = 42 - 2 = 40$. From Table 2, Appendix C, $t_{.005} = 2.704$. The rejection region is $t < -2.704$ or $t > 2.704$.

Since the observed values of the test statistic falls in the rejection region ($t = 4.90 > 2.704$), H_0 is rejected. There is sufficient evidence to indicate that weight change is correlated with digestion efficiency at $\alpha = .01$.

d. Using MINITAB, $r = .309$. Because this value if fairly close to 0, there is a weak positive linear relationship between weight change and digestion efficiency.

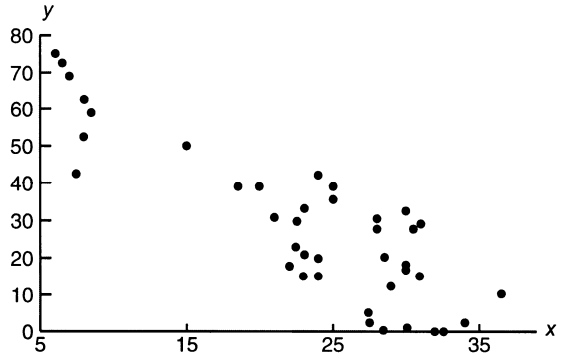
To determine if weight change is correlated with digestion efficiency, we test:

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

The test statistic is $t = \dfrac{r}{\sqrt{\dfrac{1-r^2}{n-2}}} = \dfrac{.309}{\sqrt{\dfrac{1-.309^2}{33-2}}} = 1.81.$

The rejection region requires .005 in each tail of the t distribution with df = 33 − 2 = 31. From Table 2, Appendix C, $t_{.005} \approx 2.75$. The rejection region is $t < -2.75$ or $t > 2.75$. Since the observed value of the test statistic does not fall in the rejection region ($-2.75 < t = 1.81 < 2.75$), H_0 is not rejected. There is insufficient evidence to indicate that weight change is correlated with digestion efficiency at $\alpha = .01$.

e. a.



Yes: There appears to be a negative trend. As acid detergent fibre (%) increases, digestion efficiency (%) decreases.

b. Using MINITAB, $r = .88$. Because this value is near −1, there is a fairly strong negative linear relationship between digestion efficiency (%) and acid-detergent fibre (%).

c. To determine if digestion efficiency is related to acid-detergent fibre, we test:

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

The test statistic is $t = \dfrac{r}{\sqrt{\dfrac{1-r^2}{n-2}}} = \dfrac{-.88}{\sqrt{\dfrac{1-(-.88)^2}{42-2}}} = -11.72$

The rejection region is the same as in part c, $t < -2.704$ or $t > 2.704$. Since the observed value of the test statistic falls in the rejection region ($t = -11.72 < -2.704$), H_0 is rejected. There is sufficient evidence to indicate that digestion efficiency is correlated with acid-detergent fibre at $\alpha = .01$.

d. Using MINITAB, $r = -.646$. Because this value is near −.5, there is a moderate negative linear relationship between digestion efficiency (%) and acid-detergent fibre (%).

To determine if digestion efficiency is related to acid-detergent fibre, we test:

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

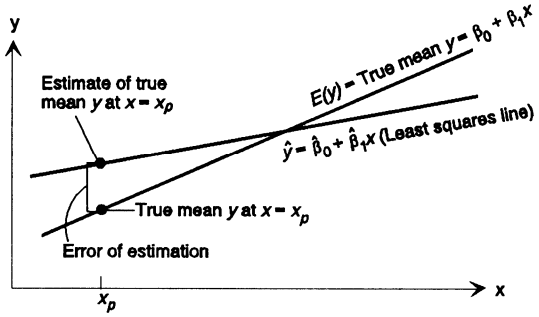
The test statistic is $t = \dfrac{r}{\sqrt{\dfrac{1-r^2}{n-2}}} = \dfrac{-.646}{\sqrt{\dfrac{1-(-.646)^2}{33-2}}} = -4.71.$

The rejection region is the same as in part d, $t < -2.75$ or $t > 2.75$. Since the observed value of the test statistic falls in the rejection region ($t = -4.71 < -2.75$), H_0 is rejected. There is sufficient evidence to indicate that digestion efficiency is correlated with acid-detergent fibre at $\alpha = .01$.

3.44 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:

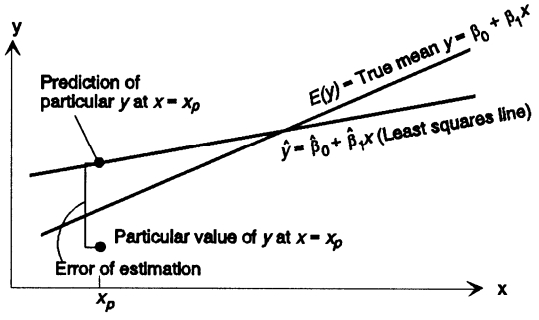
$$\sqrt{\frac{1}{n}+\frac{(x_p-\bar{x})^2}{SS_{xx}}} < \sqrt{1+\frac{1}{n}+\frac{(x_p-\bar{x})^2}{SS_{xx}}}$$

The error in estimating the mean value of y, $E(y)$, for a given value of x, say x_p, is the distance between the least squares line and the true line of means, $E(y) = \beta_0 + \beta_1 x$. This error, [$\hat{y}$ – $E(y)$], is as follows:



In contrast, the error (y_p – $\hat{y}$), in predicting some future of y is the sum of two errors the error of estimating the mean of y, $E(y)$, plus the random error that is a component of the value of y to be predicted (see figure at right).

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

3.46 a. No. We know there is a significant linear relationship between sale price and appraised value. However, the actual sale prices maybe 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 (275.86, 407.26) or (\$275,860, \$407,260). We are 95% confident that the actual sale price for a home appraised at \$300,000 is between \$275,860 and \$407,260.

c. From the printout, the 95% confidence interval for the mean sale price when the appraised value is \$300,000 is (332.95, 350.17) or (\$332,950, \$350,170). We are 95% confident that the mean sale price for a home appraised at \$300,000 is between \$332,950 and \$350,170.

3.48 Answers may vary. One possible answer is:

The 90% confidence interval for $x = 220.00$ is (5.64898, 5.83848). We are 90% confident that the mean sweetness index of all orange juice samples will be between 5.64898 and 5.83848 parts per million when the pectin value is 220.00.

3.50 a. From Exercises 3.15 and 3.29, $\bar{x} = 5.5$, $SS_{xx} = 756$, $s = .25415$, and $\hat{y} = .5704 + .0264x$.

For $x = 5$, $\hat{y} = .5704 + .0264(5) = .7024$

For confidence coefficient .99, $\alpha = .01$ and $\alpha/2 = .01/2 = .005$. From Table 2, Appendix C, with df = $n - 2 = 144 - 2 = 142$, $t_{.005} \approx 2.576$. The 99% confidence interval is:

$$\hat{y} \pm t_{\alpha/2} s \sqrt{\frac{1}{n} + \frac{\left(x_p - \bar{x}\right)^2}{SS_{xx}}} \Rightarrow .7024 \pm 2.576(.2542)\sqrt{\frac{1}{144} + \frac{(5-5.5)^2}{756}}$$

$$\Rightarrow .7024 \pm .0559 \Rightarrow (.6465,\ .7583)$$

We are 99% confident that the mean recall of all those in the 5th position is between .6465 and .7583.

b. For confidence coefficient .99, $\alpha = .01$ and $\alpha/2 = .01/2 = .005$. From Table 2, Appendix C, with df = $n - 2 = 144 - 2 = 142$, $t_{.005} \approx 2.576$. The 99% prediction interval is:

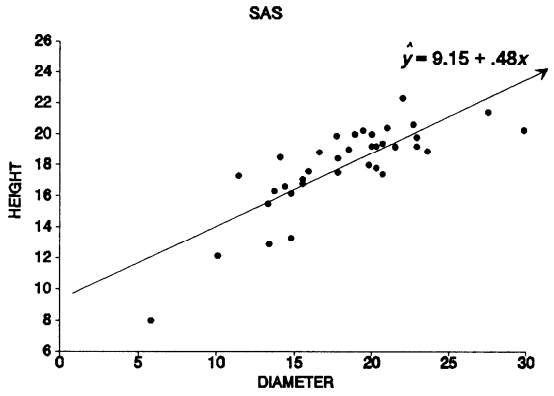
$$\hat{y} \pm t_{\alpha/2} s \sqrt{1 + \frac{1}{n} + \frac{\left(x_p - \bar{x}\right)^2}{SS_{xx}}} \Rightarrow .7024 \pm 2.576(.2542)\sqrt{1 + \frac{1}{144} + \frac{(5-5.5)^2}{756}}$$

$$\Rightarrow .7024 \pm .6572 \Rightarrow (.0452,\ 1.3596)$$

We are 99% confident that the actual recall of a person in the 5th position is between .0452 and 1.3596. Since the proportion of names recalled cannot be larger than 1, the actual proportion recalled will be between .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.

3.52 a.



NOTE: 1 obs had missing values.
1 obs hidden.

b. Fitting a straightline model to the data, the output SAS yields is:

SAS

Model: MODEL1
Dependent Variable: HEIGHT

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Value	Prob>F
Model	1	183.24469	183.24469	65.101	0.0001
Error	34	95.70281	2.81479		
C Total	35	278.94750			

Root MSE	1.67773	R-square	0.6569
Dep Mean	17.90833	Adj R-sq	0.6468
C.V.	9.36845		

Parameter Estimates

Variable	DF	Parameter Estimate	Standard Error	T for H0: Parameter = 0	Prob > |T|
INTERCEP	1	9.146839	1.12131310	8.157	0.0001
RATING	1	0.481474	0.05967333	8.069	0.0001

Obs	Ind Var DIAMETER	Dep Var HEIGHT	Predict Value	Std Err Predict	Lower95% Mean	Upper 95% Mean	Residual
1	20.0	□	18.7763	0.300	18.1675	19.3852	□

The fitted line is $\hat{y} = 9.1468 + .4815x$.

c. The fitted line fits the data well.

d. To determine if the breast height diameter is a useful predictor of tree height, we test:

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

The test statistic is $t = 8.069$ (from printout). The p-value is $p = .0001$ (from printout).

At $\alpha = .05$, there is sufficient evidence to reject H_0. There is sufficient evidence at $\alpha = .05$ to indicate breast height diameter is a useful predictor of tree height.

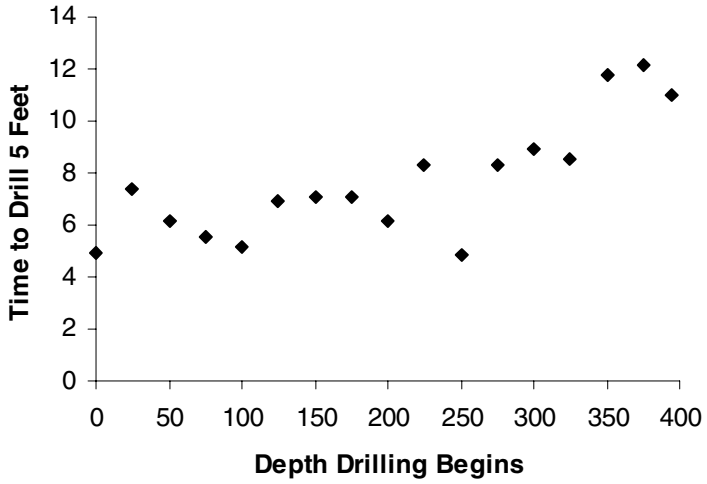
e. The form of the prediction interval is $\hat{y} \pm t_{\alpha/2}\hat{\sigma}_{\hat{y}}$.

For confidence coefficient $.90 = 1 - \alpha \Rightarrow \alpha = 1 - .90 = .10$ and $\alpha/2 = .10/2 = .05$. From Table 2 in Appendix C, with df $= n - (k + 1) = 36 - (1 + 1) = 34$, $t_{.05} \approx 1.697$. The 90% prediction interval is:

$$18.7763 \pm 1.697(.300) \Rightarrow 18.7763 \pm .5091 \Rightarrow (18.2672, 19.2854)$$

We are 90% confident that a future tree with a breast height diameter of 20 cm. will have a tree height in the interval 18.2672 to 19.2854 meters.

3.54 The scattergram of the data is shown below:



It appears that there is a positive relationship between Depth and Time. We hypothesize the following simple linear regression model to predict y = time to drill 5 feet with x = the depth at which drilling began:

$$y = \beta_0 + \beta_1 x + \varepsilon$$

From the MINITAB printout, we find the least squares prediction equation is:

$$\hat{y} = 4.79 + .014x$$

The interpretation of the parameter estimates would be:

$\hat{\beta}_0 = 4.79$: We estimate the mean time to drill 5 feet when drilling starts at 0 feet to be 4.79 minutes.

$\hat{\beta}_1 = .014$: We estimate the mean time to drill 5 feet will increase .014 minutes for each additional foot of starting depth.

To determine if Time and Depth are positively linearly related, we test:

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

The test statistic is: T = 5.05

The *p*-value is: $p = .000/2 = .000$

Since $\alpha = .01 > p = .000$, H_0 is rejected. There is sufficient evidence to indicate that time to drill 5 feet and depth at which drilling begins are positively linearly related.

Two additional interpretations we get from this analysis are:

$R^2 = 60.5$: 60.5% of the variability of the time to drill 5 feet can be explained by the linear relationship between the time to drill 5 feet and the depth at which drilling began.

$s = 1.432$: We expect most of the drilling times to fall within $2s = 2(1.432) = 2.864$ minutes of their respective least squares prediction values.

3.56 Summary calculations yield:

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

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

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

b. $\text{SSE} = \sum y^2 - \hat{\beta}\sum xy = 2403 - (3.158)(758) = 8.983$

$s^2 = \dfrac{\text{SSE}}{n-1} = \dfrac{8.983}{8-1} = 1.283 \qquad s = \sqrt{s^2} = \sqrt{1.283} = 1.133$

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 $t = \dfrac{\hat{\beta}_1}{s/\sqrt{\sum x^2}} = \dfrac{3.158}{1.133/\sqrt{240}} = 43.191$

The rejection region requires $\alpha = .025$ in the upper tail of the *t* distribution with

df = $n - 1 = 8 - 1 = 7$. From Table 2 in Appendix C, the rejection region is $t > 2.365$.

Since the observed value of the test statistic falls in the rejection region ($t = 3.191 > 2.365$), H_0 is rejected. There is sufficient evidence to indicate that x and y are positively linearly related at $\alpha = .025$.

d. The form of the confidence interval for β_1 is:

$$\hat{\beta}_1 \pm t_{\alpha/2}\left(\frac{s}{\sum x^2}\right)$$

$$\Rightarrow \hat{\beta}_1 \pm t_{.025}\left(\frac{s}{\sqrt{\sum x^2}}\right) \Rightarrow 3.158 \pm 2.365\left(\frac{1.133}{\sqrt{240}}\right) \Rightarrow 3.158 \pm .173$$

e. The point estimate when $x = 7$ is $\hat{y} = 3.158(7) = 22.11$.

The 95% confidence interval for $E(y)$ is:

$$\hat{y} \pm t_{.025}s\left(\frac{x_p}{\sqrt{\sum x^2}}\right) \Rightarrow 22.11 \pm 2.365(1.133)\left(\frac{7}{\sqrt{240}}\right) \Rightarrow 22.11 \pm 1.21$$

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

$$\hat{y} \pm t_{.025}s\left(1 + \frac{x_p^2}{\sum x^2}\right) \Rightarrow 22.11 \pm 2.365(1.133)\left(1 + \frac{7^2}{240}\right) \Rightarrow 22.11 \pm 2.94$$

3.58 Summary calculations yield:

$$\sum x = 1140 \qquad \sum x^2 = 158{,}400$$
$$\sum y = 236 \qquad \sum y^2 = 6{,}906 \qquad \sum xy = 33{,}020$$

a. $$\hat{\beta}_1 = \frac{\sum xy}{\sum x^2} = \frac{33{,}020}{158{,}400} = .2085$$

The fitted model is $\hat{y} = .2085x$.

b. $\text{SSE} = \sum y^2 = \hat{\beta}_1 \sum xy = 6{,}906 - (.2085)(33{,}020) = 22.664$

$$s^2 = \frac{\text{SSE}}{n-1} = \frac{22.664}{10-1} = 2.518 \qquad s = \sqrt{s^2} = \sqrt{2.518} = 1.587$$

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 $t = \dfrac{\hat{\beta}_1}{s/\sqrt{\sum x^2}} = \dfrac{.2085}{1.587/\sqrt{158{,}400}} = 52.28$

The rejection region requires $\alpha = .025$ in the upper tail of the t distribution with df = $n - 1 = 10 - 1 = 9$. From Table 2 in Appendix C, $t_{.025} = 2.262$. The rejection region is $t > 2.262$.

Since the observed value of the test statistic falls in the rejection region ($t = 52.28 > 2.262$), H_0 is rejected. There is sufficient evidence to indicate x and y are positively linearly related at $\alpha = .025$.

d. The form for the 95% confidence interval for β_1 is:

$$\hat{\beta}_1 \pm t_{.025}\frac{s}{\sqrt{\sum x^2}} \Rightarrow .2085 \pm 2.262\left(\frac{1.587}{\sqrt{158{,}400}}\right) \Rightarrow .2085 \pm .0900$$

e. The point estimate when $x = 125$ is $\hat{y} = .2085(125) = 26.06$.

The form for the 95% confidence interval for $E(y)$ is:

$$\hat{y} \pm t_{.025}s\left(\frac{x_p}{\sqrt{\sum x^2}}\right) \Rightarrow 26.06 \pm 2.262(1.587)\left(\frac{125}{\sqrt{158{,}000}}\right) \Rightarrow 26.06 \pm 1.13$$

f. The form for the 95% confidence prediction interval for y is:

$$\hat{y} \pm t_{.025}s\left(1 + \frac{x_p^2}{\sum x^2}\right) \Rightarrow 26.06 \pm 2.262(1.587)\left(1 + \frac{125^2}{158{,}400}\right) \Rightarrow 26.06 \pm 3.76$$

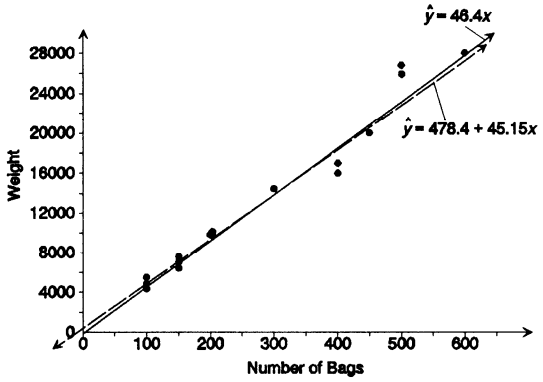
3.60 Summary calculations yield:

$\sum x = 4305$ $\quad \sum x^2 = 1{,}652{,}025$ $\quad \bar{x} = 287$

$\sum y = 201{,}558$ $\quad \sum y^2 = 3{,}571{,}211{,}200$ $\quad \bar{y} = 13{,}437.2$ $\quad \sum xy = 76{,}652{,}695$

a. $\hat{\beta}_1 = \dfrac{\sum xy}{\sum x^2} = \dfrac{76{,}652{,}695}{1{,}652{,}025} = 46.4$

The fitted model, assuming $\beta_0 = 0$, is $\hat{y} = 46.4x$.



b. $SS_{xx} = \sum x^2 - n\bar{x}^2 = 1{,}652{,}025 - 15(287)^2 = 416{,}490$

$SS_{xy} = \sum xy - n\overline{xy} = 76{,}652{,}695 - 15(287)(13{,}437.2) = 18{,}805{,}549$

$$\hat{\beta}_1 = \frac{SS_{xy}}{SS_{xx}} = \frac{18{,}805{,}549}{416{,}490} = 45.15$$

$$\hat{\beta}_0 = \bar{y} - \hat{\beta}_1\bar{x} = 13{,}437.2 - (45.15)(287) = 478.4$$

The fitted line is $\hat{y} = 478.4 + 45.15x$.

c. Since the value $x = 0$ falls outside the range of the sampled values, $\hat{\beta}_0$ has no practical interpretation. Therefore, a value of $\hat{\beta}_0$ that differs from 0 is not unexpected.

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

The test statistic is $t = \dfrac{\hat{\beta}_0 - 0}{s\sqrt{\dfrac{1}{n} + \dfrac{\bar{x}^2}{SS_{xx}}}} = \dfrac{478.4 - 0}{(1027.3)\sqrt{\dfrac{1}{15} + \dfrac{287}{416{,}490}}} = 1.794$

The rejection region requires $\alpha/2 = .10/2 = .05$ in both tails of the t distribution with df $= n - 2 = 15 - 2 = 13$. From Table 2 in Appendix C, $t_{.050} = 1.771$. The rejection region is $t > 1.771$ or $t < -1.771$.

Since the observed value of the test statistic falls in the rejection region ($t = 1.794 > 1.771$), H_0 is rejected. There is sufficient evidence to indicated β_0 should be included in the model at $\alpha = .10$.

3.62 b. The least squares line is $\hat{y} = 3.306 + .01475x$.

c. For every one unit increase in the number of factors per patient, we estimate the patient's length of stay to increase .01475 days.

d. We wish to test:

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

The test statistic is $t = 5.356$ (from printout).

The p-value is $p = .0001$ (from printout).

At $\alpha = .05$, there is sufficient evidence to indicate the number of factors per patient contributes useful information as a predictor of the patient's length of stay.

e. The form of the interval is $\hat{\beta} \pm t_{\alpha/2} \dfrac{s}{\sqrt{SS_{xx}}}$

For confidence coefficient .95 = 1 – $\alpha \Rightarrow \alpha$ = 1 – .95 = .05 and $\alpha/2$ = .05/2 = .025. From Table 2 in Appendix C, with df = n – 2 = 50 – 2 = 48. $t_{.025} \approx 2.021$. The 90% confidence interval is:

$$.01475 \pm 2.021(.00276) \Rightarrow .01475 \pm .00558 \Rightarrow (.0092, .0203)$$

We are 90% confident that for each additional factor per patient, the patient's length of stay will increase .0092 and .0203 days.

f. $r = \sqrt{r^2} = \sqrt{.3740} = .6116$

There appears to be a positive linear relationship between the number of factors and length of stay.

g. r^2 = .3740 (from printout)

37.4% of the sum of squares of deviations of the length of stay values about their mean can be explained using the number of factors as a predictor.

h. The 95% prediction interval is (6.1135, 7.3153) (from printout).

i. There is a lot of variation within the number of factors variable causing it to be not very useful from a practical perspective. Perhaps by classifying these factors differently, the width of the interval might be reduced.

3.64 a. Some preliminary calculations are:

Let x = TCDD levels in plasma and y = TCDD levels in fat tissue.

$$\sum x = 119.8 \qquad \sum x^2 = 1{,}972.82$$

$$\sum y = 137.2 \qquad \sum y^2 = 2{,}302.86 \qquad \sum xy = 2{,}046.06$$

$$SS_{xy} = \sum xy - \frac{\sum x \sum y}{n} = 2{,}046.06 - \frac{119.8(137.2)}{20} = 1{,}224.232$$

$$SS_{xx} = \sum x^2 - \frac{\left(\sum x\right)^2}{n} = 1{,}972.82 - \frac{119.8^2}{20} = 1{,}255.218$$

$$SS_{yy} = \sum y^2 - \frac{\left(\sum y\right)^2}{n} = 2{,}302.86 - \frac{137.2^2}{20} = 1{,}361.668$$

$$\bar{x} = \frac{\sum x}{n} = \frac{119.8}{20} = 5.99 \qquad \bar{y} = \frac{\sum y}{n} = \frac{137.2}{20} = 6.86$$

Using the TCDD level in plasma as the independent variable, the parameter estimates are:

$$\hat{\beta}_1 = \frac{SS_{xy}}{SS_{xx}} = \frac{1{,}224.232}{1{,}255.218} = 0.975314248 \approx 0.9753$$

$$\hat{\beta}_0 = \bar{y} - \hat{\beta}_1\bar{x} = 6.86 - 0.975314248(5.99) = 1.017867653 \approx 1.0179$$

The least squares prediction equation is $\hat{y} = 1.0179 + 0.9753x$.

Using the TCDD level in fat tissue as the independent variable, the parameter estimates are:

$$\hat{\beta}_1 = \frac{SS_{xy}}{SS_{xx}} = \frac{1{,}224.232}{1{,}361.218} = 0.899067907 \approx 0.8991$$

$$\hat{\beta}_0 = \bar{x} - \hat{\beta}_1\bar{y} = 5.99 - 0.899067907(6.86) = -0.177605848 \approx -0.1776$$

The least squares prediction equation is $\bar{x} = -0.1776 + 0.8991y$.

b. If we want to see if fat tissue level (y) is a useful linear predictor of blood plasma level (x), we will use the model:

$$x = \beta_0 + \beta_1 y + \varepsilon$$

We must first calculate SSE, s^2, and s.

$$SSE = SS_{xx} - \hat{\beta}_1 SS_{xy} = 1{,}255.218 - 0.899067907(1{,}224.232) = 154.550298$$

$$s^2 = \frac{SSE}{n-2} = \frac{154.550298}{20-2} = 8.586127667 \quad s = \sqrt{8.586127667} = 2.93021$$

To determine of fat tissue level (y) is a useful linear predictor of blood plasma level (x), we test:

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

The test statistic is $t = \dfrac{\hat{\beta}_1}{s/\sqrt{SS_{yy}}} = \dfrac{.8991}{2.93021/\sqrt{1{,}361.6168}} = 11.32$

The rejection region for this small-sample, two-tailed test requires $\alpha/2 = .05/2 = .025$ in each tail of the t distribution with df = $n - 2 = 20 - 2 = 18$. From Table 2, Appendix C, $t_{.025} = 2.101$. The rejection region is $t < -2.101$ or $t > 2.101$.

Since the observed value of the test statistic falls in the rejection region ($t = 11.32 > 2.101$), H_0 is rejected. There is sufficient evidence to indicate that fat tissue level (y) is a useful predictor of blood plasma level (x) at $\alpha = .05$.

c. If we want to see if blood plasma level (x) is a useful linear predictor of fat tissue level (y), we will use the model:

$$y = \beta_0 + \beta_1 x + \varepsilon$$

We must first calculate SSE, s^2, and s.

$$\text{SSE} = \text{SS}_{yy} - \hat{\beta}_1 \text{SS}_{xy} = 1{,}361.668 - 0.975314248(1{,}224.232) = 167.657088$$

$$s^2 = \frac{\text{SSE}}{n-2} = \frac{167.657088}{20-2} = 9.314282667 \quad s = \sqrt{9.314282667} = 3.05193$$

To determine of blood plasma level (x) is a useful linear predictor of fat tissue level (y), we test:

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

The test statistic is $t = \dfrac{\hat{\beta}_1}{s/\sqrt{\text{SS}_{xx}}} = \dfrac{.9753}{3.05193/\sqrt{1{,}255.218}} = 11.32$

The rejection region is $t < -2.101$ or $t > 2.101$ (from part (b) above).

Since the observed value of the test statistic falls in the rejection region ($t = 11.32 > 2.101$), H_0 is rejected. There is sufficient evidence to indicate that blood plasma level (x) is a useful predictor of fat tissue level (y) at $\alpha = .05$.

3.66 To determine if a positive linear relationship between CEO cash compensation and REIT performance exists, we test:

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

The test statistic is $t = \dfrac{r\sqrt{n-2}}{\sqrt{1-r^2}} = \dfrac{.328\sqrt{16-2}}{\sqrt{1-.328^2}} = 1.299$

The rejection region requires $\alpha = .05$ in the upper tail of the t distribution with df = $n - 2 = 16 - 2 = 14$. From Table 2, Appendix C, $t_{.05} = 1.761$. The rejection region is $t > 1.761$.

Since the observed value of the test statistic is not in the rejection region ($t = 1.299 \not> 1.761$), H_0 cannot be rejected. There is insufficient evidence to indicate that CEO cash compensation and REIT performance are positively linearly related at $\alpha = .05$.

3.68 a. $r = -.50$. This indicates that, for girls, there is a negative linear relationship between the child's weight percentile and the number of cigarettes smoked per day in the child's home. As the number of cigarettes smoked per day in the child's home increases, the child's weight percentile decreases. Since $-.5$ is not very close to -1, this relationship is not very strong.

b. $p = .03$. This indicates that for any $\alpha > .03$, H_0 will be rejected. There is sufficient evidence that, for girls, the child's weight percentile and the number of cigarettes smoked per day in the child's home are linearly correlated.

c. $r = -.12$. This indicates that, for boys, there is a negative linear relationship between the child's weight percentile and the number of cigarettes smoked per day in the child's home. As the number of cigarettes smoked per day in the child's home increases, the child's weight percentile decreases. Since $-.12$ is close to 0, this relationship is very weak.

d. $p = .57$. This indicates that for any $\alpha > .57$, H_0 will be rejected. Since most tests are run with $\alpha \leq .10$, H_0 will not be rejected. There is insufficient evidence that, for boys, the child's weight percentile and the number of cigarettes smoked per day in the child's home are linearly correlated.

3.70 a. Some preliminary calculations are:

$$\sum x_i = 36 \qquad \sum y_i = 629 \qquad n = 8$$
$$\sum x_i^2 = 204 \qquad \sum x_i y_i = 4632$$

$$SS_{xy} = \sum x_i y_i - \frac{\left(\sum x_i\right)\left(\sum y_i\right)}{n}$$
$$= 4632 - \frac{(36)(629)}{8} = 1801.5$$

$$SS_{xx} = \sum x_i^2 - \frac{\left(\sum x_i\right)^2}{n}$$
$$= 204 - \frac{(36)^2}{8} = 42$$

$$\hat{\beta}_1 = \frac{SS_{xy}}{SS_{xx}} = \frac{1801.5}{42} = 42.89285714 \approx 42.89$$

$$\hat{\beta}_0 = \bar{y} - \hat{\beta}_1 \bar{x} = \frac{629}{8} - (42.89285714)\frac{36}{8} - 114.3928571 \approx -114.39$$

The least squares line is $\hat{y} = -114.39 + 42.89x$

b. $\hat{\beta}_0 = -114.39$. Since $x = 0$ is not in the observed range, $\hat{\beta}_0$ has no interpretation other than than being the y-intercept.

$\hat{\beta}_1 = 42.89$. For each additional increase of 1 step, the mean number of unrooted walks increases by an estimated 42.89 (or about 43).

c. Some preliminary calculations are:

$$\sum x_i = 36 \qquad \sum y_i = 9304 \qquad n = 8$$
$$\sum x_i^2 = 204 \qquad \sum x_i y_i = 69168$$

$$SS_{xy} = \sum x_i y_i - \frac{\left(\sum x_i\right)\left(\sum y_i\right)}{n}$$

$$= 69{,}168 - \frac{(36)(9304)}{8} = 27{,}300$$

$$SS_{xx} = \sum x_i^2 - \frac{\left(\sum x_i\right)^2}{n}$$

$$= 204 - \frac{(36)^2}{8} = 42$$

$$\hat{\beta}_1 = \frac{SS_{xy}}{SS_{xx}} = \frac{27{,}300}{42} = 650$$

$$\hat{\beta}_0 = \bar{y} - \hat{\beta}_1 \bar{x} = \frac{9304}{8} - (650)\frac{36}{8} = -1762$$

The least squares line is $\hat{y} = -1762 + 650x$

$\hat{\beta}_0 = -1762$. Since $x = 0$ is not in the observed range, $\hat{\beta}_0$ has no interpretation other than being the y-intercept.

$\hat{\beta}_1 = 650$. For each additional increase of 1 step, the mean number of self-avoiding walks increases by an estimated 650.

d. Some preliminary calculations are:

$$SS_{yy} = \sum y_i^2 - \frac{\left(\sum y_i\right)^2}{n} = 175{,}875 - \frac{(629)^2}{8} = 126{,}419.875$$

$$SSE = SS_{yy} - \hat{\beta}_1 SS_{xy} = 126{,}419.875 - 42.89285714(1801.5) = 49148.39286$$

$$s = \sqrt{s^2} = \sqrt{\frac{SSE}{n-2}} = \sqrt{\frac{49148.39286}{8-2}} \approx 90.506$$

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

For $x_p = 4$, $\hat{y} = -114.39 + 42.89(4) = 57.17$.
For confidence coefficient .99, $\alpha = .01$ and $\alpha/2 = .005$.

From Table 2, Appendix C, with df = 8 − 2 = 6, $t_{.005} = 3.707$. The confidence interval is:

$$57.17 \pm 3.707(90.506)\sqrt{\frac{1}{8} + \frac{(4-4.5)^2}{42}}$$

$$\Rightarrow 57.17 \pm 121.41 \Rightarrow (-64.24,\ 178.58)$$

e. No; Since $x = 15$ is not in the observed range, it is not recommended that simple linear regression be used to predict the number of unrooted walks possible when the walk length is 15 steps. We have to idea if the relationship between walk length and the number of unrooted walks is the same when length is 15 steps.

3.72 a. $\hat{\beta}_1 = .020$. For each additional 1% increase in leaves infected, the mean log of the average number of infections per leaf is estimated to increase by .02.

b. $r^2 = .816$. 81.6% of the total sample variability around the sample mean log of the average number of infections per leaf is explained by the linear relationship between the log of the average number of infections per leaf and the percentage of leaves infected.

c. $s = .288$. We would expect most of the observed values of the log of the average number of infections per leaf to fall within $\pm 2s$ or $\pm 2(.288)$ or .576 units of their predicted values.

d. $r = \sqrt{.816} = .903$. Because this number is close to 1, there is a fairly strong positive linear relationship between the log of the average number of infections per leaf and the percentage of leaves infected.

e. To determine if there is a linear relationship between the log of the average number of infections per leaf and the percentage of leaves infected, we test:

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

The test statistic is $t = \dfrac{r}{\sqrt{(1-r^2)/(n-2)}} = \dfrac{.903}{\sqrt{(1-(.816))/(100-2)}} = 20.83$

The rejection region requires $\alpha/2 = .05/2 = .025$ in each tail of the t distribution with df $= n - 2 = 100 - 2 = 98$. From Table 2, Appendix C, $t_{.025} \approx 1.99$. The rejection region is $t < -1.99$ or $t > 1.99$.

Since the observed value of the test statistic falls in the rejection region ($t = 20.83 > 1.99$), H_0 is rejected. There is sufficient evidence to indicate that there is a linear relationship between the log of the average number of infections per leaf and the percentage of leaves infected at $\alpha = .05$.

f. For $x_p = 80\%$, $\hat{y} = -.939 + .020(80) = .661$. The antilog (base 10) of .661 is 4.58. Thus, when the percentage of leaves infected is 80%, the average number of infections per leaf is predicted to be 4.58.

3.74 a. H_0: $\rho = 0$
H_a: $\rho > 0$

The test statistic is $t = \dfrac{r\sqrt{n-2}}{\sqrt{1-r^2}} = \dfrac{.26\sqrt{130-2}}{\sqrt{1-.26^2}} = 3.05$

The rejection region requires $\alpha = .01$ in the upper tail of the t distribution with df = $n - 2 = 130 - 2 = 128$. From Table 2 in Appendix C, $t_{.01} \approx 2.358$. The rejection region is $t > 2.358$.

Since the observed value of the test statistic falls in the rejection region ($t = 3.05 > 2.358$), H_0 is rejected. We agree that mother and daughter loneliness scores were positively correlated at $\alpha = .01$.

b. Using the rejection region from (a), the rejection region is $t > 2.358$ when testing for positive correlation and $t < -2.358$ when testing for negative correlation.

We will begin by testing the next strongest correlation, $r = -.21$. The test statistic is:

$$t = \frac{r\sqrt{n-2}}{\sqrt{1-r^2}} = \frac{-.21\sqrt{130-2}}{\sqrt{1-(-.21)^2}} = -2.43$$

Since $-2.43 < -2.358$, there is sufficient evidence to indicate mother number of friends and daughter loneliness scores are negatively correlated at $\alpha = .01$.

We next test $r = .19$. The test statistic is:

$$t = \frac{r\sqrt{n-2}}{\sqrt{1-r^2}} = \frac{-.19\sqrt{130-2}}{\sqrt{1-.19^2}} = 2.19$$

Since $2.19 \not> 2.358$, there is insufficient evidence to reject H_0: $\rho = 0$. All other correlation values will lead to the same conclusion at $\alpha = .01$.

c. Both a mother's loneliness and a daughter's loneliness may be caused by other factors such as an absent husband/father, living locations, etc.

d. In part (b), we were testing only for linear correlations. It may be that some of the variables are correlated, but not linearly correlated.