CHAPTER 12

Section 12.1

1.

a. Stem and Leaf display of temp:

17	0	
17	23	stem = tens
17	445	leaf = ones
17	67	
17		
18	0000011	
18	2222	
18	445	
18	6	
18	8	

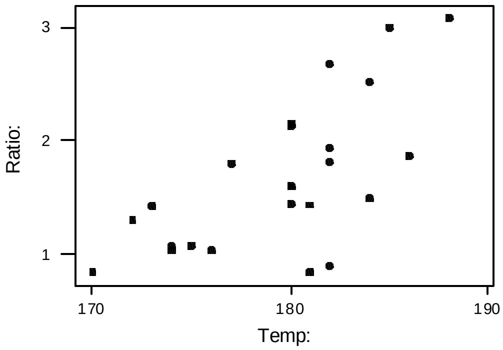
180 appears to be a typical value for this data. The distribution is reasonably symmetric in appearance and somewhat bell-shaped. The variation in the data is fairly small since the range of values (188 – 170 = 18) is fairly small compared to the typical value of 180.

0	889	
1	0000	stem = ones
1	3	leaf = tenths
1	4444	
1	66	
1	8889	
2	11	
2		
2	5	
2	6	
2		
3	00	

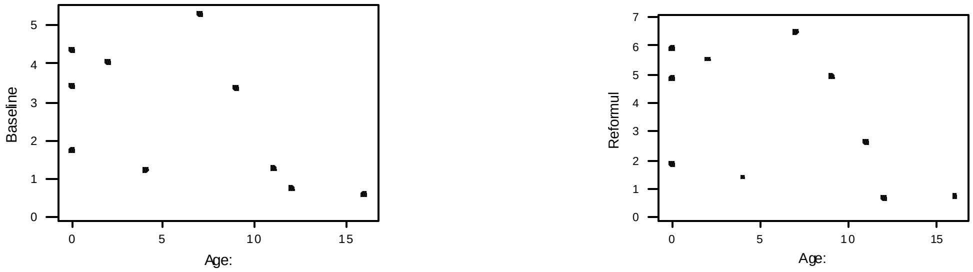
For the ratio data, a typical value is around 1.6 and the distribution appears to be positively skewed. The variation in the data is large since the range of the data (3.08 - .84 = 2.24) is very large compared to the typical value of 1.6. The two largest values could be outliers.

b. The efficiency ratio is not uniquely determined by temperature since there are several instances in the data of equal temperatures associated with different efficiency ratios. For example, the five observations with temperatures of 180 each have different efficiency ratios.

c. A scatter plot of the data appears below. The points exhibit quite a bit of variation and do not appear to fall close to any straight line or simple curve.

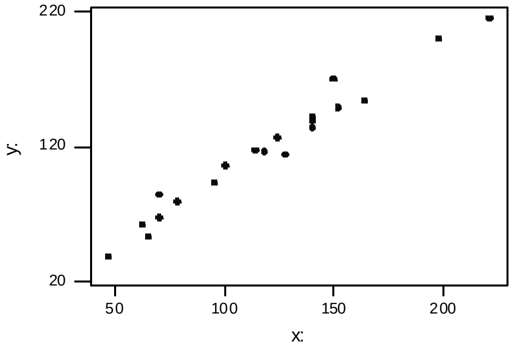


2. Scatter plots for the emissions vs age:



With this data the relationship between the age of the lawn mower and its NO_x emissions seems somewhat dubious. One might have expected to see that as the age of the lawn mower increased the emissions would also increase. We certainly do not see such a pattern. Age does not seem to be a particularly useful predictor of NO_x emission.

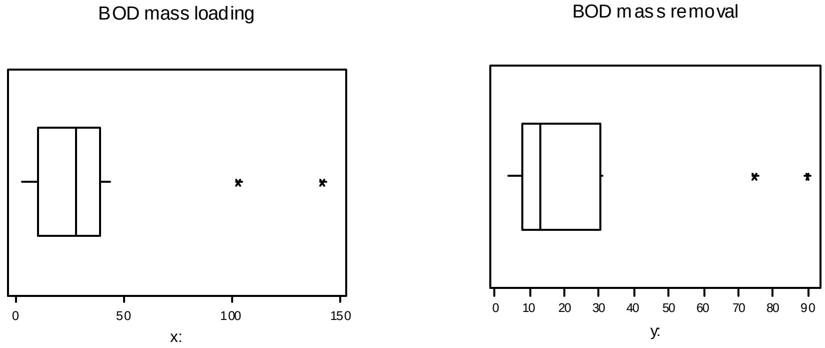
3. A scatter plot of the data appears below. The points fall very close to a straight line with an intercept of approximately 0 and a slbpe of about 1. This suggests that the two methods are producing substantially the same concentration measurements.



4.

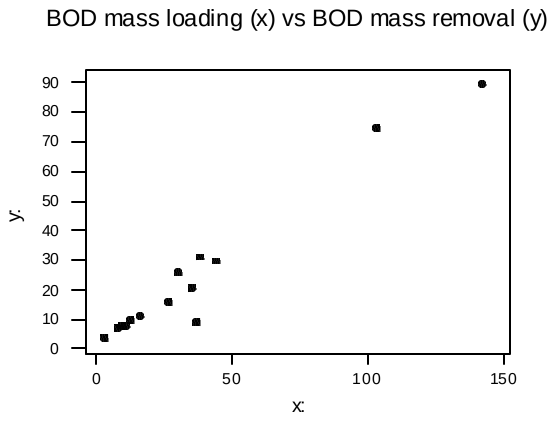
a.

Box plots of both variables:



On both the BOD mass loading boxplot and the BOD mass removal boxplot there are 2 outliers. Both variables are positively skewed.

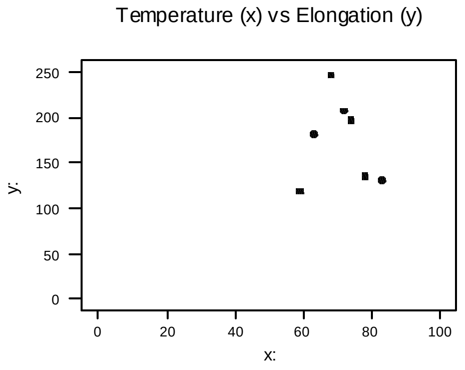
b. Scatter plot of the data:



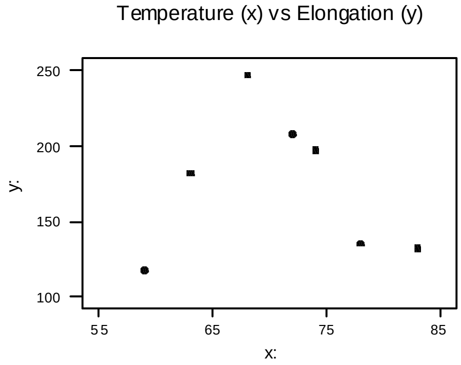
There is a strong linear relationship between BOD mass loading and BOD mass removal. As the loading increases, so does the removal. The two outliers seen on each of the boxplots are seen to be correlated here. There is one observation that appears not to match the liner pattern. This value is (37, 9). One might have expected a larger value for BOD mass removal.

5.

a. The scatter plot with axes intersecting at (0,0) is shown below.



b. The scatter plot with axes intersecting at (55, 100) is shown below.



c. A parabola appears to provide a good fit to both graphs.

6. There appears to be a linear relationship between racket resonance frequency and sum of peak-to-peak acceleration. As the resonance frequency increases the sum of peak-to-peak acceleration tends to decrease. However, there is not a perfect relationship. Variation does exist. One should also notice that there are two tennis rackets that appear to differ from the other 21 rackets. Both have very high resonance frequency values. One might investigate if these rackets differ in other ways as well.

7.

a. $m_{Y\cdot 2500} = 1800 + 1.3(2500) = 5050$

b. expected change = slope = $b_1 = 1.3$

c. expected change = $100b_1 = 130$

d. expected change = $-100b_1 = -130$

8.

a. $m_{Y\cdot 2000} = 1800 + 1.3(2000) = 4400$, and $s = 350$, so $P(Y > 5000)$
$= P\left(Z > \frac{5000 - 4400}{350}\right) = P(Z > 1.71) = .0436$

b. Now E(Y) = 5050, so $P(Y > 5000) = P(Z > .14) = .4443$

c. $E(Y_2 - Y_1) = E(Y_2) - E(Y_1) = 5050 - 4400 = 650$, and
$V(Y_2 - Y_1) = V(Y_2) + V(Y_1) = (350)^2 + (350)^2 = 245{,}000$, so the s.d. of
$Y_2 - Y_1 = 494.97$.
Thus $P(Y_2 - Y_1 > 0) = P\left(z > \frac{100 - 650}{494.97}\right) = P(Z > .71) = .2389$

d. The standard deviation of $Y_2 - Y_1 = 494.97$ (from **c**), and
$E(Y_2 - Y_1) = 1800 + 1.3x_2 - (1800 + 1.3x_1) = 1.3(x_2 - x_1)$. Thus
$P(Y_2 > Y_1) = P(Y_2 - Y_1 > 0) = P\left(z > \frac{-1.3(x_2 - x_1)}{494.97}\right) = .95$ implies that
$-1.645 = \frac{-1.3(x_2 - x_1)}{494.97}$, so $x_2 - x_1 = 626.33$.

9.

a. $b_1 =$ expected change in flow rate (y) associated with a one inch increase in pressure drop (x) = .095.

b. We expect flow rate to decrease by $5b_1 = .475$.

c. $m_{Y\cdot 10} = -.12 + .095(10) = .83,$ and $m_{Y\cdot 15} = -.12 + .095(15) = 1.305$.

d. $P(Y > .835) = P\left(Z > \frac{.835 - .830}{.025}\right) = P(Z > .20) = .4207$
$P(Y > .840) = P\left(Z > \frac{.840 - .830}{.025}\right) = P(Z > .40) = .3446$

e. Let Y_1 and Y_2 denote pressure drops for flow rates of 10 and 11, respectively. Then $m_{Y\cdot 11} = .925,$ so Y_1 - Y_2 has expected value .830 - .925 = -.095, and s.d.
$\sqrt{(.025)^2 + (.025)^2} = .035355$. Thus
$P(Y_1 > Y_2) = P(Y_1 - Y_2 > 0) = P\left(z > \frac{+.095}{.035355}\right) = P(Z > 2.69) = .0036$

10. Y has expected value 14,000 when x = 1000 and 24,000 when x = 2000, so the two probabilities become $P\left(z > \frac{-8500}{s}\right) = .05$ and $P\left(z > \frac{-17,500}{s}\right) = .10$. Thus $\frac{-8500}{s} = -1.645$ and $\frac{-17,500}{s} = -1.28$. This gives two different values for s, a contradiction, so the answer to the question posed is no.

11.

a. $b_1 =$ expected change for a one degree increase = -.01, and $10b_1 = -.1$ is the expected change for a 10 degree increase.

b. $m_{Y \cdot 200} = 5.00 - .01(200) = 3$, and $m_{Y \cdot 250} = 2.5$.

c. The probability that the first observation is between 2.4 and 2.6 is

$$P(2.4 \le Y \le 2.6) = P\left(\frac{2.4 - 2.5}{.075} \le Z \le \frac{2.6 - 2.5}{.075}\right)$$

$= P(-1.33 \le Z \le 1.33) = .8164$. The probability that any particular one of the other four observations is between 2.4 and 2.6 is also .8164, so the probability that all five are between 2.4 and 2.6 is $(.8164)^5 = .3627$.

d. Let Y_1 and Y_2 denote the times at the higher and lower temperatures, respectively. Then $Y_1 - Y_2$ has expected value $5.00 - .01(x+1) - (5.00 - .01x) = -.01$. The standard deviation of $Y_1 - Y_2$ is $\sqrt{(.075)^2 + (.075)^2} = .10607$. Thus

$$P(Y_1 - Y_2 > 0) = P\left(z > \frac{-(-.01)}{.10607}\right) = P(Z > .09) = .4641.$$

Section 12.2

12.

a. $S_{xx} = 39{,}095 - \dfrac{(517)^2}{14} = 20{,}002.929$,

$S_{xy} = 25{,}825 - \dfrac{(517)(346)}{14} = 13047.714$; $\hat{b}_1 = \dfrac{S_{xy}}{S_{xx}} = \dfrac{13{,}047.714}{20{,}002.929} = .652$;

$\hat{b}_0 = \dfrac{\Sigma y - \hat{b}_1 \Sigma x}{n} = \dfrac{346 - (.652)(517)}{14} = .626$, so the equation of the least squares regression line is $y = .626 + .652x$.

b. $\hat{y}_{(35)} = .626 + .652(35) = 23.456$. The residual is $y - \hat{y} = 21 - 23.456 = -2.456$.

c. $S_{yy} = 17{,}454 - \dfrac{(346)^2}{14} = 8902.857$, so

$SSE = 8902.857 - (.652)(13047.714) = 395.747$.

$\hat{s} = \sqrt{\dfrac{SSE}{n-2}} = \sqrt{\dfrac{395.747}{12}} = 5.743$.

d. $SST = S_{yy} = 8902.857$; $r^2 = 1 - \dfrac{SSE}{SST} = 1 - \dfrac{395.747}{8902.857} = .956$.

e. Without the two upper extreme observations, the new summary values are $n = 12, \Sigma x = 272, \Sigma x^2 = 8322, \Sigma y = 181, \Sigma y^2 = 3729, \Sigma xy = 5320$. The new $S_{xx} = 2156.667, S_{yy} = 998.917, S_{xy} = 1217.333$. New $\hat{b}_1 = .56445$ and $\hat{b}_0 = 2.2891$, which yields the new equation $y = 2.2891 + .56445x$. Removing the two values changes the position of the line considerably, and the slope slightly. The new $r^2 = 1 - \dfrac{311.79}{998.917} = .6879$, which is much worse than that of the original set of observations.

13. For this data, n = 4, $\Sigma x_i = 200$, $\Sigma y_i = 5.37$, $\Sigma x_i^2 = 12.000$, $\Sigma y_i^2 = 9.3501$, $\Sigma x_i y_i = 333$. $S_{xx} = 12{,}000 - \frac{(200)^2}{4} = 2000$,

$S_{yy} = 9.3501 - \frac{(5.37)^2}{4} = 2.140875$, and $S_{xy} = 333 - \frac{(200)(5.37)}{4} = 64.5$.

$\hat{b}_1 = \frac{S_{xy}}{S_{xx}} = \frac{64.5}{2000} = .03225$ and $\hat{b}_0 = \frac{5.37}{4} - (.03225)\frac{200}{4} = -.27000$.

$SSE = S_{yy} - \hat{b}_1 S_{xy} = 2.14085 - (.03225)(64.5) = .060750$.

$r^2 = 1 - \frac{SSE}{SST} = 1 - \frac{.060750}{2.14085} = .972$. This is a very high value of r^2, which confirms the authors' claim that there is a strong linear relationship between the two variables.

14.

a. n = 24, $\Sigma x_i = 4308$, $\Sigma y_i = 40.09$, $\Sigma x_i^2 = 773{,}790$, $\Sigma y_i^2 = 76.8823$, $\Sigma x_i y_i = 7{,}243.65$. $S_{xx} = 773{,}790 - \frac{(4308)^2}{24} = 504.0$,

$S_{yy} = 76.8823 - \frac{(40.09)^2}{24} = 9.9153$, and

$S_{xy} = 7{,}243.65 - \frac{(4308)(40.09)}{24} = 45.8246$. $\hat{b}_1 = \frac{S_{xy}}{S_{xx}} = \frac{45.8246}{504} = .09092$ and

$\hat{b}_0 = \frac{40.09}{24} - (.09092)\frac{4308}{24} = -14.6497$. The equation of the estimated regression line is $\hat{y} = -14.6497 + .09092x$.

b. When x = 182, $\hat{y} = -14.6497 + .09092(182) = 1.8997$. So when the tank temperature is 182, we would predict an efficiency ratio of 1.8997.

c. The four observations for which temperature is 182 are: (182, .90), (182, 1.81), (182, 1.94), and (182, 2.68). Their corresponding residuals are: $.90 - 1.8997 = -0.9977$, $1.81 - 1.8997 = -0.0877$, $1.94 - 1.8997 = 0.0423$, $2.68 - 1.8997 = 0.7823$. These residuals do not all have the same sign because in the cases of the first two pairs of observations, the observed efficiency ratios were smaller than the predicted value of 1.8997. Whereas, in the cases of the last two pairs of observations, the observed efficiency ratios were larger than the predicted value.

d. $SSE = S_{yy} - \hat{b}_1 S_{xy} = 9.9153 - (.09092)(45.8246) = 5.7489$.

$r^2 = 1 - \frac{SSE}{SST} = 1 - \frac{5.7489}{9.9153} = .4202$. (42.02% of the observed variation in efficiency ratio can be attributed to the approximate linear relationship between the efficiency ratio and the tank temperature.)

15.

a. The following stem and leaf display shows that: a typical value for this data is a number in the low 40’s. there is some positive skew in the data. There are some potential outliers (79.5 and 80.0), and there is a reasonably large amount of variation in the data (e.g., the spread 80.0-29.8 = 50.2 is large compared with the typical values in the low 40’s).

2	9	
3	33	stem = tens
3	5566677889	leaf = ones
4	1223	
4	56689	
5	1	
5		
6	2	
6	9	
7		
7	9	
8	0	

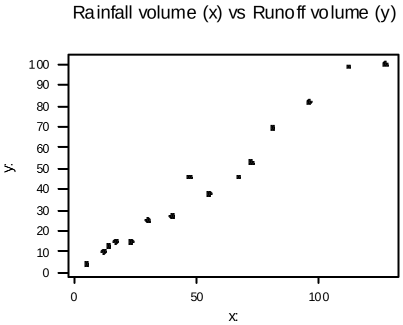
b. No, the strength values are not uniquely determined by the MoE values. For example, note that the two pairs of observations having strength values of 42.8 have different MoE values.

c. The least squares line is $\hat{y} = 3.2925 + .10748x$. For a beam whose modulus of elasticity is x = 40, the predicted strength would be $\hat{y} = 3.2925 + .10748(40) = 7.59$. The value x = 100 isfar beyond the range of the x values in the data, so it would be dangerous (i.e., potentially misleading) to extrapolated the linear relationship that far.

d. From the output, SSE = 18.736, SST = 71.605, and the coefficient of determination is r^2 = .738 (or 73.8%). The r^2 value is large, which suggests that the linear relationship is a useful approximation to the true relationship between these two variables.

16.

a.



Yes, the scatterplot shows a strong linear relationship between rainfall volume and runoff volume, thus it supports the use of the simple linear regression model.

b. $\bar{x} = 53.200$, $\bar{y} = 42.867$, $S_{xx} = 63040 - \dfrac{(798)^2}{15} = 20{,}586.4$,

$S_{yy} = 41{,}999 - \dfrac{(643)^2}{15} = 14{,}435.7$, and

$S_{xy} = 51{,}232 - \dfrac{(798)(643)}{15} = 17{,}024.4$. $\hat{b}_1 = \dfrac{S_{xy}}{S_{xx}} = \dfrac{17{,}024.4}{20{,}586.4} = .82697$ and

$\hat{b}_0 = 42.867 - (.82697)53.2 = -1.1278$.

c. $m_{y\cdot 50} = -1.1278 + .82697(50) = 40.2207$.

d. $SSE = S_{yy} - \hat{b}_1 S_{xy} = 14{,}435.7 - (.82697)(17{,}324.4) = 357.07$.

$s = \hat{s} = \sqrt{\dfrac{SSE}{n-2}} = \sqrt{\dfrac{357.07}{13}} = 5.24$.

e. $r^2 = 1 - \dfrac{SSE}{SST} = 1 - \dfrac{357.07}{14{,}435.7} = .9753$. So 97.53% of the observed variation in runoff volume can be attributed to the simple linear regression relationship between runoff and rainfall.

17. Note: n = 23 in this study.

a. For a one (mg/cm^2) increase in dissolved material, one would expect a .144 (g/l) increase in calcium content. Secondly, 86% of the observed variation in calcium content can be attributed to the simple linear regression relationship between calcium content and dissolved material.

b. $m_{y\cdot 50} = 3.678 + .144(50) = 10.878$

c. $r^2 = .86 = 1 - \frac{SSE}{SST}$, so $SSE = (SST)(1 - .86) = (320.398)(.14) = 44.85572$.

Then $s = \sqrt{\frac{SSE}{n-2}} = \sqrt{\frac{44.85572}{21}} = 1.46$

18.

a. $\hat{b}_1 = \frac{15(987.645) - (1425)(10.68)}{15(139{,}037.25) - (1425)^2} = \frac{-404.3250}{54{,}933.7500} = -.00736023$

$\hat{b}_0 = \frac{10.68 - (-.00736023)(1425)}{15} = 1.41122185$, $y = 1.4112 - .007360x$.

b. $\hat{b}_1 = -.00736023$

c. With x now denoting temperature in oC, $y = \hat{b}_0 + \hat{b}_1\left(\frac{9}{5}x + 32\right)$

$= \left(\hat{b}_0 + 32\hat{b}_1\right) + \frac{9}{5}\hat{b}_1 x = 1.175695 - .0132484x$, so the new $\hat{b}_1$ is -.0132484 and the new $\hat{b}_0 = 1.175695$.

d. Using the equation of **a**, predicted $y = \hat{b}_0 + \hat{b}_1(200) = -.0608$, but the deflection factor cannot be negative.

19. N = 14, $\Sigma x_i = 3300$, $\Sigma y_i = 5010$, $\Sigma x_i^2 = 913{,}750$, $\Sigma y_i^2 = 2{,}207{,}100$, $\Sigma x_i y_i = 1{,}413{,}500$

a. $\hat{b}_1 = \dfrac{3{,}256{,}000}{1{,}902{,}500} = 1.71143233$, $\hat{b}_0 = -45.55190543$, so we use the equation $y = -45.5519 + 1.7114x$.

b. $\hat{m}_{Y\cdot 225} = -45.5519 + 1.7114(225) = 339.51$

c. Estimated expected change $= -50\hat{b}_1 = -85.57$

d. No, the value 500 is outside the range of x values for which observations were available (the danger of extrapolation).

20.

a. $\hat{b}_0 = .3651$, $\hat{b}_1 = .9668$

b. .8485

c. $\hat{s} = .1932$

d. SST = 1.4533, 71.7% of this variation can be explained by the model. Note: $\dfrac{SSR}{SST} = \dfrac{1.0427}{1.4533} = .717$ which matches R-squared on output.

21.

a. The summary statistics can easily be verified using Minitab or Excel, etc.

b. $\hat{b}_1 = \dfrac{491.4}{744.16} = .66034186$, $\hat{b}_0 = -2.18247148$

c. predicted $y = \hat{b}_0 + \hat{b}_1(15) = 7.72$

d. $\hat{m}_{Y\cdot 15} = \hat{b}_0 + \hat{b}_1(15) = 7.72$

22.

a. $\hat{b}_1 = \frac{-404.325}{54.933.75} = -.00736023$, $\hat{b}_0 = 1.41122185$,
$SSE = 7.8518 - (1.41122185)(10.68) - (-.00736023)(987.654) = .049245$,
$s^2 = \frac{.049245}{13} = .003788$, and $\hat{s} = s = .06155$

b. $SST = 7.8518 - \frac{(10.68)^2}{15} = .24764$ so $r^2 = 1 - \frac{.049245}{.24764} = 1 - .199 = .801$

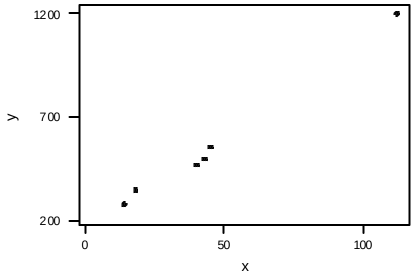
23.

a. Using the $y_i{}'s$ given to one decimal place accuracy is the answer to Exercise 19, $SSE = (150 - 125.6)^2 + ... + (670 - 639.0)^2 = 16,213.64$. The computation formula gives
$SSE = 2,207,100 - (-45.55190543)(5010) - (1.71143233)(1,413,500)$
$= 16,205.45$

b. $SST = 2,207,100 - \frac{(5010)^2}{14} = 414,235.71$ so $r^2 = 1 - \frac{16,205.45}{414,235.71} = .961$.

24.

a.



According to the scatter plot of the data, a simple linear regression model does appear to be plausible.

b. The regression equation is `y = 138 + 9.31 x`

c. The desired value is the coefficient of determination, $r^2 = 99.0\%$.

d. The new equation is `y* = 190 + 7.55 x*.` This new equation appears to differ significantly. If we were to predict a value of y^* for $x^* = 50$, the value would be 567.9, where using the original data, the predicted value for x = 50 would be 603.5.

25. Substitution of $\hat{b}_0 = \frac{\Sigma y_i - \hat{b}_1 \Sigma x_i}{n}$ and $\hat{b}_1$ for b$_o$ and b$_1$ on the left hand side of the normal equations yields $\frac{n\left(\Sigma y_i - \hat{b}_1 \Sigma x_i\right)}{n} + \hat{b}_1 \Sigma x_i = \Sigma y_i$ from the first equation and

$$\frac{\Sigma x_i\left(\Sigma y_i - \hat{b}_1 \Sigma x_i\right)}{n} + \hat{b}_1 \Sigma x_i^2 = \frac{\Sigma x_i \Sigma y_i}{n} + \frac{\hat{b}_1\left(n\Sigma x_i^2 - (\Sigma x_i)^2\right)}{n}$$

$\frac{\Sigma x_i \Sigma y_i}{n} + \frac{n\Sigma x_i y_i}{n} - \frac{\Sigma x_i \Sigma y_i}{n} = \Sigma x_i y_i$ from the second equation.

26. We show that when $\bar{x}$ is substituted for x in $\hat{b}_0 + \hat{b}_1 x$, $\bar{y}$ results, so that $(\bar{x}, \bar{y})$ is on the line $y = \hat{b}_0 + \hat{b}_1 x$: $\hat{b}_0 + \hat{b}_1 \bar{x} = \frac{\Sigma y_i - b_1 \Sigma x_i}{n} + \hat{b}_1 \bar{x} = \bar{y} - \hat{b}_1 \bar{x} + \hat{b}_1 \bar{x} = \bar{y}$.

27. We wish to find b$_1$ to minimize $\Sigma(y_i - b_1 x_i)^2 = f(b_1)$. Equating $f'(b_1)$ to 0 yields $2\Sigma(y_i - b_1 x_i)(-x_i) = 0$ so $\Sigma x_i y_i = b_1 \Sigma x_i^2$ and $b_1 = \frac{\Sigma x_i y_i}{\Sigma x_i^2}$. The least squares estimator of $\hat{b}_1$ is thus $\hat{b}_1 = \frac{\Sigma x_i Y_i}{\Sigma x_i^2}$.

28.

a. Subtracting $\bar{x}$ from each x_i shifts the plot in a rigid fashion $\bar{x}$ units to the left without otherwise altering its character. The last squares line for the new plot will thus have the same slope as the one for the old plot. Since the new line is $\bar{x}$ units to the left of the old one, the new y intercept (height at x = 0) is the height of the old line at x = $\bar{x}$, which is $\hat{b}_0 + \hat{b}_1 \bar{x} = \bar{y}$ (since from exercise 26, $(\bar{x}, \bar{y})$ is on the old line). Thus the new y intercept is $\bar{y}$.

b. We wish b$_0$ and b$_1$ to minimize f(b$_0$, b$_1$) = $\Sigma\left[y_i - \left(b_0 + b_1(x_i - \bar{x})\right)\right]^2$. Equating $\frac{\partial f}{\partial b_0}$ to $\frac{\partial f}{\partial b_1}$ to 0 yields $nb_0 + b_1 \Sigma(x_i - \bar{x}) = \Sigma y_i$, $b_0 \Sigma(x_i - \bar{x}) + b_1 \Sigma(x_i - \bar{x})^2$ $= \Sigma(x_i - \bar{x})^2 = \Sigma(x_i - \bar{x}) y_i$. Since $\Sigma(x_i - \bar{x}) = 0$, $b_0 = \bar{y}$, and since $\Sigma(x_i - \bar{x}) y_i = \Sigma(x_i - \bar{x})(y_i - \bar{y})$ [because $\Sigma(x_i - \bar{x})\bar{y} = \bar{y}\Sigma(x_i - \bar{x})$], $b_1 = \hat{b}_1$. Thus $\hat{b}_0^* = \bar{Y}$ and $\hat{b}_1^* = \hat{b}_1$.

29. For data set #1, $r^2 = .43$ and $\hat{s} = s = 4.03$; whereas these quantities are .99 and 4.03 for #2, and .99 and 1.90 for #3. In general, one hopes for both large r^2 (large % of variation explained) and small s (indicating that observations don't deviate much from the estimated line). Simple linear regression would thus seem to be most effective in the third situation.

Section 12.3

30.

a. $\Sigma(x_i - \overline{x})^2 = 7{,}000{,}000$, so $V(\hat{b}_1) = \dfrac{(350)^2}{7{,}000{,}000} = .0175$ and the standard deviation of $\hat{b}_1$ is $\sqrt{.0175} = .1323$.

b. $P(1.0 \le \hat{b}_1 \le 1.5) = P\left(\dfrac{1.0 - 1.25}{1.323} \le Z \le \dfrac{1.5 - 1.25}{1.323}\right)$
$= P(-1.89 \le Z \le 1.89) = .9412$.

c. Although n = 11 here and n = 7 in **a**, $\Sigma(x_i - \overline{x})^2 = 1{,}100{,}000$ now, which is smaller than in **a**. Because this appears in the denominator of $V(\hat{b}_1)$, the variance is smaller for the choice of x values in **a**.

31.

a. $\hat{b}_1 = -.00736023$, $\hat{b}_0 = 1.41122185$, so
$SSE = 7.8518 - (1.41122185)(10.68) - (-.00736023)(987.645) = .04925$,
$s^2 = .003788$, $s = .06155$. $\hat{s}^2_{\hat{b}_1} = \dfrac{s^2}{\Sigma x_i^2 - (\Sigma x_i)^2 / n} = \dfrac{.003788}{3662.25} = .00000103$,

$\hat{s}_{\hat{b}_1} = s_{\hat{b}_1} =$ estimated s.d. of $\hat{b}_1 = \sqrt{.00000103} = .001017$.

b. $-.00736 \pm (2.160)(.001017) = -.00736 \pm .00220 = (-.00956, -.00516)$

32. Let b_1 denote the true average change in runoff for each 1 m^3 increase in rainfall. To test the hypotheses $H_o : b_1 = 0$ vs. $H_a : b_1 \neq 0$, the calculated t statistic is

$t = \frac{\hat{b}_1}{s_{\hat{b}_1}} = \frac{.82697}{.03652} = 22.64$ which (from the printout) has an associated p-value of P = 0.000. Therefore, since the p-value is so small, H$_o$ is rejected and we conclude that there is a useful linear relationship between runoff and rainfall.

A confidence interval for b_1 is based on n – 2 = 15 – 2 = 13 degrees of freedom. $t_{.025,13} = 2.160$, so the interval estimate is

$\hat{b}_1 \pm t_{.025,13} \cdot s_{\hat{b}_1} = .82697 \pm (2.160)(.03652) = (.748,.906)$. Therefore, we can be confident that the true average change in runoff, for each 1 m^3 increase in rainfall, is somewhere between .748 m^3 and .906 m^3.

33.

a. From the printout in Exercise 15, the error d.f. = n – 2 = 25, $t_{.025,25} = 2.060$. The confidence interval is then

$\hat{b}_1 \pm t_{.025,25} \cdot s_{\hat{b}_1} = .10748 \pm (2.060)(.01280) = (.081,.134)$. Therefore, we estimate with a high degree of confidence that the true average change in strength associated with a 1 Gpa increase in modulus of elasticity is between .081 MPa and .134 MPa.

b. We wish to test $H_o : b_1 = .1$ vs. $H_a : b_1 > .1$. The calculated t statistic is

$t = \frac{\hat{b}_1 - .1}{s_{\hat{b}_1}} = \frac{.10748 - .1}{.01280} = .58$, which yields a p-value of .277. A large p-value such as this would not lead to rejecting H$_o$, so there is not enough evidence to contradict the prior belief.

34.

a. $H_o : b_1 = 0$; $H_a : b_1 \neq 0$

RR: $|t| > t_{a/2,n-2}$ or $|t| > 3.106$

$t = 5.29$: Reject H$_o$. The slope differs significantly from 0, and the model appears to be useful.

b. At the level $a = 0.01$, reject h$_o$ if the p-value is less than 0.01. In this case, the reported p-value was 0.000, therefore reject H$_o$. The conclusion is the same as that of part **a**.

c. $H_o : b_1 = 1.5$; $H_a : b_1 < 1.5$

RR: $t < -t_{a,n-2}$ or $t < -2.718$

$t = \frac{0.9668 - 1.5}{0.1829} = -2.92$: Reject H$_o$. The data contradict the prior belief.

35.

a. We want a 95% CI for $\beta_{1:}$ $\hat{b}_1 \pm t_{.025,15} \cdot s_{\hat{b}_1}$. First, we need our point estimate, $\hat{b}_1$.

Using the given summary statistics, $S_{xx} = 3056.69 - \frac{(222.1)^2}{17} = 155.019$,

$S_{xy} = 2759.6 - \frac{(222.1)(193)}{17} = 238.112$, and $\hat{b}_1 = \frac{S_{xy}}{S_{xx}} = \frac{238.112}{115.019} = 1.536$.

We need $\hat{b}_0 = \frac{193 - (1.536)(222.1)}{17} = -8.715$ to calculate the SSE:

$SSE = 2975 - (-8.715)(193) - (1.536)(2759.6) = 418.2494$. Then

$s = \sqrt{\frac{418.2494}{15}} = 5.28$ and $s_{\hat{b}_1} = \frac{5.28}{\sqrt{155.019}} = .424$. With $t_{.025,15} = 2.131,$ our CI is $1.536 \pm 2.131 \cdot (.424)$ = (.632, 2.440). With 95% confidence, we estimate that the change in reported nausea percentage for every one-unit change in motion sickness dose is between .632 and 2.440.

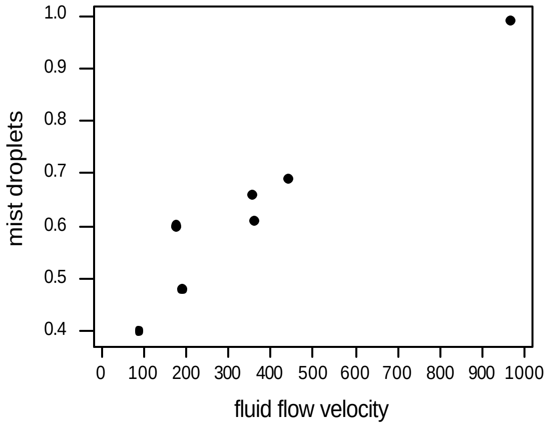
b. We test the hypotheses $H_o : b_1 = 0$ vs $H_a : b_1 \neq 0$, and the test statistic is $t = \frac{1.536}{.424} = 3.6226$. With df=15, the two-tailed p-value = 2P(t > 3.6226) = 2(.001) = .002. With a p-value of .002, we would reject the null hypothesis at most reasonable significance levels. This suggests that there is a useful linear relationship between motion sickness dose and reported nausea.

c. No. A regression model is only useful for estimating values of nausea % when using dosages between 6.0 and 17.6 – the range of values sampled.

d. Removing the point (6.0, 2.50), the new summary stats are: n = 16, ,$\Sigma x_i = 216.1$, $\Sigma y_i = 191.5$, $\Sigma x_i^2 = 3020.69$, $\Sigma y_i^2 = 2968.75$, $\Sigma x_i y_i = 2744.6$, and then $\hat{b}_1 = 1.561$, $\hat{b}_0 = -9.118$, SSE = 430.5264, $s = 5.55$, $s_{\hat{b}_1} = .551$, and the new CI is $1.561 \pm 2.145 \cdot (.551)$, or (.379, 2.743). The interval is a little wider. But removing the one observation did not change it that much. The observation does not seem to be exerting undue influence.

36.

a. A scatter plot, generated by Minitab, supports the decision to use linear regression analysis.



b. We are asked for the coefficient of determination, r^2. From the Minitab output, $r^2 = .931$ (which is close to the hand calculated value, the difference being accounted for by round-off error.)

c. Increasing x from 100 to 1000 means an increase of 900. If, as a result, the average y were to increase by .6, the slope would be $\frac{.6}{900} = .0006667$. We should test the hypotheses $H_o : b_1 = .0006667$ vs. $H_a : b_1 < .0006667$. The test statistic is $t = \frac{.00062108 - .0006667}{.00007579} = -.601$, which is not significant. There is not sufficient evidence that with an increase from 100 to 1000, the true average increase in y is less than .6.

d. We are asked for a confidence interval for b_1. Using the values from the Minitab output, we have $.00062108 \pm 2.776(.00007579) = (.00041069,.00083147)$

37.

a. n = 10, $\Sigma x_i = 2615$, $\Sigma y_i = 39.20$, $\Sigma x_i^2 = 860{,}675$, $\Sigma y_i^2 = 161.94$,

$\Sigma x_i y_i = 11{,}453.5$, so $\hat{b}_1 = \dfrac{12{,}027}{1{,}768{,}525} = .00680058$, $\hat{b}_0 = 2.14164770$, from which SSE = .09696713, s = .11009492 $s = .11009492 \approx .110 = \hat{s}$,

$$\hat{s}_{\hat{b}_1} = \frac{.110}{\sqrt{176{,}852}} = .000262$$

b. We wish to test $H_o : b_1 = .0060$ vs $H_a : b_1 \neq .0060$. At level .10, H_o is rejected if either $t \geq t_{.05,8} = 1.860$ or $t \leq -t_{.05,8} = -1.860$. Since $t = \dfrac{.0068 - .0060}{.000262} = 3.06 \geq 1.1860$, H_o is rejected.

38.

a. From Exercise 23, which also refers to Exercise 19, SSE = 16.205.45, so $s^2 = 1350.454$, $s = 36.75$, and $s_{\hat{b}_1} = \dfrac{36.75}{368.636} = .0997$. Thus $t = \dfrac{1.711}{.0997} = 17.2 > 4.318 = t_{.0005,14}$, so p-value < .001. Because the p-value < .01, $H_o : b_1 = 0$ is rejected at level .01 in favor of the conclusion that the model is useful $(b_1 \neq 0)$.

b. The C.I. for b_1 is $1.711 \pm (2.179)(.0997) = 1.711 \pm .217 = (1.494, 1.928)$. Thus the C.I. for $10b_1$ is $(14.94, 19.28)$.

39. SSE = 124,039.58– (72.958547)(1574.8) – (.04103377)(222657.88) = 7.9679, and SST = 39.828

Source	**df**	**SS**	**MS**	**f**
Regr	1	31.860	31.860	18.0
Error	18	7.968	1.77	
Total	19	39.828		

Let's use α = .001. Then $F_{.001,1,18} = 15.38 < 18.0$, so $H_o : b_1 = 0$ is rejected and the model is judged useful. $s = \sqrt{1.77} = 1.33041347$, $S_{xx} = 18{,}921.8295$, so

$$t = \frac{.04103377}{1.33041347/\sqrt{18{,}921.8295}} = 4.2426, \text{ and } t^2 = (4.2426)^2 = 18.0 = f.$$

40. We use the fact that $\hat{b}_1$ is unbiased for b_1. $E\left(\hat{b}_0\right) = \dfrac{E\left(\Sigma y_i - \hat{b}_1 \Sigma x_i\right)}{n}$

$$= \frac{E(\Sigma y_i)}{n} - E\left(\hat{b}_1\right)\overline{x} = \frac{E(\Sigma Y_i)}{n} - b_1\overline{x}$$

$$= \frac{\Sigma\left(b_0 + b_1 x_i\right)}{n} - b_1\overline{x} = b_0 + b_1\overline{x} - b_1\overline{x} = b_0$$

41.

a. Let $c = n\Sigma x_i^2 - (\Sigma x_i)^2$. Then $E\left(\hat{b}_1\right) = \dfrac{1}{c}E\left[n\Sigma x_i Y_i ... Y_i - (\Sigma x_i)...(\Sigma x_i)(\Sigma Y_i)\right]$

$$= \frac{n}{c}\sum x_i E(Y_i) - \frac{\Sigma x_i}{c}\sum E(Y_i) = \frac{n}{c}\sum x_i\left(b_0 + b_1 x_i\right) - \frac{\Sigma x_i}{c}\sum\left(b_0 + b_1 x_i\right)$$

$$\frac{b_1}{c}\left[n\Sigma x_i^2 - (\Sigma x_i)^2\right] = b_1.$$

b. With $c = \Sigma\left(x_i - \overline{x}\right)^2$, $\hat{b}_1 = \dfrac{1}{c}\Sigma\left(x_i - \overline{x}\right)\left(Y_i - \overline{Y}\right) = \dfrac{1}{c}\Sigma\left(x_i - \overline{x}\right)Y_i$ (since $\Sigma\left(x_i - \overline{x}\right)\overline{Y} = \overline{Y}\Sigma\left(x_i - \overline{x}\right) = \overline{Y}\cdot 0 = 0$), so $V\left(\hat{b}_1\right) = \dfrac{1}{c^2}\Sigma\left(x_i - \overline{x}\right)^2 Var(Y_i)$

$= \dfrac{1}{c^2}\Sigma\left(x_i - \overline{x}\right)^2 \cdot s^2 = \dfrac{s^2}{\Sigma\left(x_i - \overline{x}\right)^2} = \dfrac{s^2}{\Sigma x_i^2 - (\Sigma x_i)^2 / n}$, as desired.

42. $t = \hat{b}_1 \dfrac{\sqrt{\Sigma x_i^2 - (\Sigma x_i)^2 / n}}{s}$. The numerator of $\hat{b}_1$ will be changed by the factor cd (since both $\Sigma x_i y_i$ and $(\Sigma x_i)(\Sigma y_i)$ appear) while the denominator of $\hat{b}_1$ will change by the factor c^2 (since both Σx_i^2 and $(\Sigma x_i)^2$ appear). Thus $\hat{b}_1$ will change by the factor d/c. Because $SSE = \Sigma\left(y_i - \hat{y}_i\right)^2$, SSE will change by the factor d^2, so s will change by the factor d. Since $\sqrt{\bullet}$ in t changes by the factor c, t itself will change by $\dfrac{d}{c}\cdot\dfrac{c}{d} = 1$, or not at all.

43. The numerator of d is |1 – 2| = 1, and the denominator is $\dfrac{4\sqrt{14}}{\sqrt{324.40}} = .831$, so $d = \dfrac{1}{.831} = 1.20$. The approximate power curve is for n – 2 df = 13, and b is read from Table A.17 as approximately .1.

Section 12.4

44.

a. The mean of the x data in Exercise 12.15 is $\bar{x} = 45.11$. Since x = 40 is closer to 45.11 than is x = 60, the quantity $(40 - \bar{x})^2$ must be smaller than $(60 - \bar{x})^2$. Therefore, since these quantities are the only ones that are different in the two $s_{\hat{y}}$ values, the $s_{\hat{y}}$ value for x = 40 must necessarily be smaller than the $s_{\hat{y}}$ for x = 60. Said briefly, the closer x is to $\bar{x}$, the smaller the value of $s_{\hat{y}}$.

b. From the printout in Exercise 12.15, the error degrees of freedom is df = 25. $t_{.025,25} = 2.060$, so the interval estimate when x = 40 is : $7.592 \pm (2.060)(.179)$ $7.592 \pm .369 = (7.223,7.961)$. We estimate, with a high degree of confidence, that the true average strength for all beams whose MoE is 40 GPa is between 7.223 MPa and 7.961 MPa.

c. From the printout in Exercise 12.15, s = .8657, so the 95% prediction interval is $\hat{y} \pm t_{.025,25}\sqrt{s^2 + s_{\hat{y}}^2} = 7.592 \pm (2.060)\sqrt{(.8657)^2 + (.179)^2}$ $= 7.592 \pm 1.821 = (5.771,9.413)$. Note that the prediction interval is almost 5 times as wide as the confidence interval.

d. For two 95% intervals, the simultaneous confidence level is at least 100(1 – 2(.05)) = 90%

45.

a. We wish to find a 90% CI for $m_{y\cdot 125}$: $\hat{y}_{125} = 78.088$, $t_{.05,18} = 1.734$, and

$$s_{\hat{y}} = s\sqrt{\frac{1}{20} + \frac{(125 - 140.895)^2}{18{,}921.8295}} = .3349$$ .Putting it together, we get

$$78.088 \pm 1.734(.3349) = (77.5073,78.6687)$$

b. We want a 90% PI: Only the standard error changes:

$$s_{\hat{y}} = s\sqrt{1 + \frac{1}{20} + \frac{(125 - 140.895)^2}{18{,}921.8295}} = 1.3719$$, so the PI is

$$78.088 \pm 1.734(1.3719) = (75.7091,80.4669)$$

c. Because the x* of 115 is farther away from $\bar{x}$ than the previous value, the term $(x^* - \bar{x})^2$ will be larger, making the standard error larger, and thus the width of the interval is wider.

d. We would be testing to see if the filtration rate were 125 kg-DS/m/h, would the average moisture content of the compressed pellets be less than 80%. The test statistic is $t = \dfrac{78.088 - 80}{.3349} = -5.709$, and with 18 df the p-value is P(t<-5.709) ˜ 0.00. We would reject H_o. There is significant evidence to prove that the true average moisture content when filtration rate is 125 is less than 80%.

46.

a. A 95% CI for $m_{Y\cdot 500}$: $\hat{y}_{(500)} = -.311 + (.00143)(500) = .40$ and

$$s_{\hat{y}_{(500)}} = .131\sqrt{\frac{1}{13} + \frac{(500 - 471.54)^2}{131{,}519.23}} = .03775$$

, so the interval is

$$\hat{y}_{(500)} \pm t_{.025,11} \cdot S_{\hat{y}_{(500)}} = .40 \pm 2.210(.03775) = .40 \pm .08 = (.32,.48)$$

b. The width at x = 400 will be wider than that of x = 500 because x = 400 is farther away from the mean ($\bar{x} = 471.54$).

c. A 95% CI for b_1:

$$\hat{b}_1 \pm t_{.025,11} \cdot s_{\hat{b}_1} = .00143 \pm 2.201(.0003602) = (.000637,.002223)$$

d. We wish to test $H_0 : y_{(400)} = .25$ vs. $H_0 : y_{(400)} \neq .25$. The test statistic is $t = \dfrac{\hat{y}_{(400)} - .25}{s_{\hat{y}_{(400)}}}$, and we reject H_o if $|t| \geq t_{.025,11} = 2.201$.

$$\hat{y}_{(400)} = -.311 + .00143(400) = .2614$$

and

$$s_{\hat{y}_{(400)}} = .131\sqrt{\frac{1}{13} + \frac{(400 - 471.54)^2}{131{,}519.23}} = .0445$$

, so the calculated $t = \dfrac{.2614 - .25}{.0445} = .2561$, which is not ≥ 2.201, so we do not reject H_o. This sample data does not contradict the prior belief.

47.

a. $\hat{y}_{(40)} = -1.128 + .82697(40) = 31.95$, $t_{.025,13} = 2.160$; a 95% PI for runoff is $31.95 \pm 2.160\sqrt{(5.24)^2 + (1.44)^2} = 31.95 \pm 11.74 = (20.21, 43.69)$. No, the resulting interval is very wide, therefore the available information is not very precise.

b. $\Sigma x = 798, \Sigma x^2 = 63{,}040$ which gives $S_{xx} = 20{,}586.4$, which in turn gives $s_{\hat{y}(50)} = 5.24\sqrt{\frac{1}{15} + \frac{(50 - 53.20)^2}{20{,}586.4}} = 1.358$, so the PI for runoff when x = 50 is $40.22 \pm 2.160\sqrt{(5.24)^2 + (1.358)^2} = 40.22 \pm 11.69 = (28.53, 51.92)$. The simultaneous prediction level for the two intervals is at least $100(1 - 2a)\% = 90\%$.

48.

a. $S_{xx} = 18.24 - \frac{(12.6)^2}{9} = .60$, $S_{xy} = 40.968 - \frac{(12.6)(27.68)}{9} = 2.216$; $S_{yy} = 93.3448 - \frac{(27.68)^2}{9} = 8.213$ $\hat{b}_1 = \frac{S_{xy}}{S_{xx}} = \frac{2.216}{.60} = 3.693$; $\hat{b}_0 = \frac{\Sigma y - \hat{b}_1 \Sigma x}{n} = \frac{27.68 - (3.693)(12.6)}{9} = -2.095$, so the point estimate is $\hat{y}_{(1.5)} = -2.095 + 3.693(1.5) = 3.445$. $SSE = 8.213 - 3.693(2.216) = .0293$, which yields $s = \sqrt{\frac{SSE}{n-2}} = \sqrt{\frac{.0293}{7}} = .0647$. Thus $s_{\hat{y}(1.5)} = .0647\sqrt{\frac{1}{9} + \frac{(1.5 - 1.4)^2}{.60}} = .0231$. The 95% CI for $m_{y \cdot 1.5}$ is $3.445 \pm 2.365(.0231) = 3.445 \pm .055 = (3.390, 3.50)$.

b. A 95% PI for y when x = 1.5 is similar: $3.445 \pm 2.365\sqrt{(.0647)^2 + (.0231)^2} = 3.445 \pm .162 = (3.283, 3.607)$. The prediction interval for a future y value is wider than the confidence interval for an average value of y when x is 1.5.

c. A new PI for y when x = 1.2 will be wider since x = 1.2 is farther away from the mean $\bar{x} = 1.4$.

49. 95% CI: (462.1, 597.7); midpoint = 529.9; $t_{.025,8} = 2.306$;

$$529.9 + (2.306)\left(\hat{s}_{\hat{b}_0+\hat{b}_1(15)}\right) = 597.7$$

$$\hat{s}_{\hat{b}_0+\hat{b}_1(15)} = 29.402$$

99% CI: $529.9 \pm (3.355)(29.402) = (431.3,628.5)$

50. $\hat{b}_1 = 18.87349841$, $\hat{b}_0 = -8.77862227$, SSE = 2486.209, s = 16.6206

a. $\hat{b}_0 + \hat{b}_1(18) = 330.94$, $\bar{x} = 20.2909$, $\sqrt{\dfrac{1}{11} + \dfrac{11(18-20.2909)^2}{3834.26}} = .3255$, $t_{.025,9} = 2.262$, so the CI is $330.94 \pm (2.262)(16.6206)(.3255)$ $= 330.94 \pm 12.24 = (318.70,343.18)$

b. $\sqrt{1 + \dfrac{1}{11} + \dfrac{11(18-20.2909)^2}{3834.26}} = 1.0516$, so the P.I. is $330.94 \pm (2.262)(16.6206)(1.0516) = 330.94 \pm 39.54 = (291.40,370.48)$.

c. To obtain simultaneous confidence of at least 97% for the three intervals, we compute each one using confidence level 99%, (with $t_{.005,9} = 3.250$). For x = 15, the interval is $274.32 \pm 22.35 = (251.97, 296.67)$. For x = 18, $330.94 \pm 17.58 = (313.36,348.52)$. For x = 20, $368.69 \pm 0.84 = (367.85,369.53)$.

51.

a. 0.40 is closer to $\bar{x}$.

b. $\hat{b}_0 + \hat{b}_1(0.40) \pm t_{a/2,n-2} \cdot \left(\hat{s}_{\hat{b}_0+\hat{b}_1(0.40)}\right)$ or $0.8104 \pm (2.101)(0.0311)$ $= (0.745,0.876)$

c. $\hat{b}_0 + \hat{b}_1(1.20) \pm t_{a/2,n-2} \cdot \sqrt{s^2 + s^2{}_{\hat{b}_0+\hat{b}_1(1.20)}}$ or

$0.2912 \pm (2.101) \cdot \sqrt{(0.1049)^2 + (0.0352)^2} = (.059,.523)$

52.

a. We wish to test $H_o : b_1 = 0$ vs $H_a : b_1 \neq 0$. The test statistic $t = \dfrac{10.6026}{.9985} = 10.62$ leads to a p-value of < .006 (2P(t > 4.0) from the 7 df row of table A.8), and H_o is rejected since the p-value is smaller than any reasonable a . The data suggests that this model does specify a useful relationship between chlorine flow and etch rate.

b. A 95% confidence interval for b_1: $10.6026 \pm (2.365)(.9985) = (8.24, 12.96)$. We can be highly confident that when the flow rate is increased by 1 SCCM, the associated expected change in etch rate will be between 824 and 1296 A/min.

c. A 95% CI for $m_{Y\cdot 3.0}$: $38.256 \pm 2.365\left(2.546\sqrt{\dfrac{1}{9} + \dfrac{9(3.0-2.667)^2}{58.50}}\right)$

$= 38.256 \pm 2.365(2.546)(.35805) = 38.256 \pm 2.156 = (36.100, 40.412)$, or 3610.0 to 4041.2 A/min.

d. The 95% PI is $38.256 \pm 2.365\left(2.546\sqrt{1 + \dfrac{1}{9} + \dfrac{9(3.0-2.667)^2}{58.50}}\right)$

$= 38.256 \pm 2.365(2.546)(1.06) = 38.256 \pm 6.398 = (31.859, 44.655)$, or 3185.9 to 4465.5 A/min.

e. The intervals for x* = 2.5 will be narrower than those above because 2.5 is closer to the mean than is 3.0.

f. No. a value of 6.0 is not in the range of observed x values, therefore predicting at that point is meaningless.

53. Choice **a** will be the smallest, with d being largest. **a** is less than **b** and **c** (obviously), and **b** and **c** are both smaller than **d**. Nothing can be said about the relationship between **b** and **c**.

54.

a. There is a linear pattern in the scatter plot, although the pot also shows a reasonable amount of variation about any straight line fit to the data. The simple linear regression model provides a sensible starting point for a formal analysis.

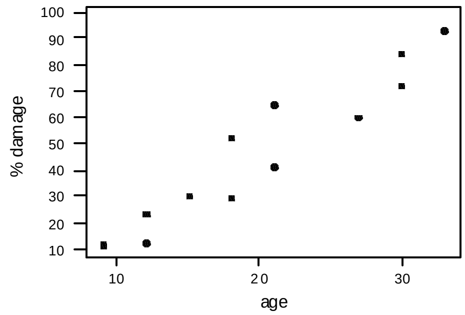
b. n = 141, $\Sigma x_i = 1185, \Sigma x_i^2 = 151{,}825, \Sigma y_i = 5960, \Sigma y_i^2 = 2{,}631{,}200$, and $\Sigma x_i y_i = 449{,}850$, from which $\hat{b}_1 = -1.060132, \hat{b}_0 = 515.446887, SSE = 36{,}036.93,$ $r^2 = .616, s^2 = 3003.08, s = 54.80, s_{\hat{b}_1} = \dfrac{54.80}{\sqrt{51{,}523.21}} = .241$ $H_o : b_1 = 0$ vs $H_a : b_1 \neq 0$, $t = \dfrac{\hat{b}_1}{s_{\hat{b}_1}}$. Reject H$_o$ at level .05 if either $t \geq t_{.025,12} = 2.179$ or $t \leq -2.179$. We calculate $t = \dfrac{-1.060}{.241} = -4.39$. Since $-4.39 \leq -2.179$ H$_o$ is rejected. The simple linear regression model does appear to specify a useful relationship.

c. A confidence interval for $b_0 + b_1(75)$ is requested. The interval is centered at $\hat{b}_0 + \hat{b}_1(75) = 435.9$. $s_{\hat{b}_0 + \hat{b}_1(75)} = s\sqrt{\dfrac{1}{n} + \dfrac{n(75 - \overline{x})^2}{n\Sigma x_i^2 - (\Sigma x_i)^2}} = 14.83$ (using s = 54.803). Thus a 95% CI is $435.9 \pm (2.179)(14.83) = (403.6, 559.7)$.

55.

a. $x_2 = x_3 = 12$, yet $y_2 \neq y_3$

b.



Based on a scatterplot of the data, a simple linear regression model does seem a reasonable way to describe the relationship between the two variables.

c. $\hat{b}_1 = \frac{2296}{699} = 3.284692$, $\hat{b}_0 - 19.669528$, $y = -19.67 + 3.285x$

d. $SSE = 35{,}634 - (-19.669528)(572) - (3.284692)(14{,}022) = 827.0188$,

$s^2 = 82.70188, s = 9.094$. $s_{\hat{b}_0 + \hat{b}_1(20)} = 9.094\sqrt{\frac{1}{12} + \frac{12(20 - 20.5)^2}{8388}} = 2.6308$,

$\hat{b}_0 + \hat{b}_1(20) = 46.03$, $t_{.025,10} = 2.228$. The PI is $46.03 \pm 2.228\sqrt{s^2 + s^2_{\hat{b}_0 + \hat{b}_1(20)}}$

$= 46.03 \pm 21.09 = (24.94, 67.12)$.

56. $\hat{b}_0 + \hat{b}_1 x = \overline{Y} - \hat{b}_1\overline{x} + \hat{b}_1 x = \overline{Y} + (x - \overline{x})\hat{b}_1 = \frac{1}{n}\sum Y_i + \frac{(x - \overline{x})\sum(x_i - \overline{x})Y_i}{n\Sigma x_i^2 - (\Sigma x_i)^2} = \Sigma d_i Y_i$

where $d_i = \frac{1}{n} + \frac{(x - \overline{x})(x_i - \overline{x})}{n\Sigma x_i^2 - (\Sigma x_i)^2}$. Thus $Var(\hat{b}_0 + \hat{b}_1 x) = \sum d_i^2 Var(Y_i) = s^2 \Sigma d_i^2$, which, after some algebra, yields the desired expression.

Section 12.5

57. Most people acquire a license as soon as they become eligible. If, for example, the minimum age for obtaining a license is 16, then the time since acquiring a license, y, is usually related to age by the equation $y \approx x - 16$, which is the equation of a straight line. In other words, the majority of people in a sample will have y values that closely follow the line $y = x - 16$.

58.

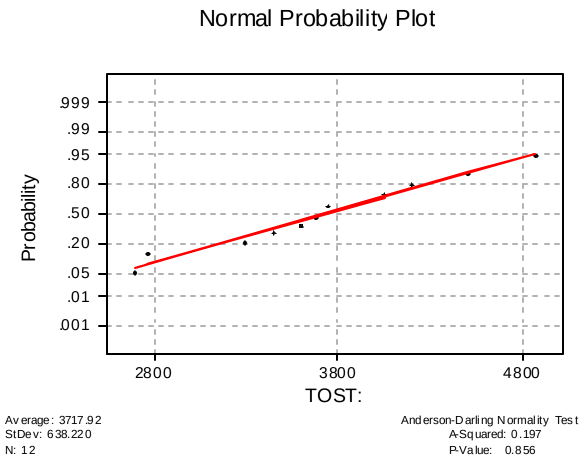
a. Summary values: $\Sigma x = 44{,}615$, $\Sigma x^2 = 170{,}355{,}425$, $\Sigma y = 3{,}860$, $\Sigma y^2 = 1{,}284{,}450$, $\Sigma xy = 14{,}755{,}500$, $n = 12$. Using these values we calculate $S_{xx} = 4{,}480{,}572.92$, $S_{yy} = 42{,}816.67$, and $S_{xy} = 404{,}391.67$. So

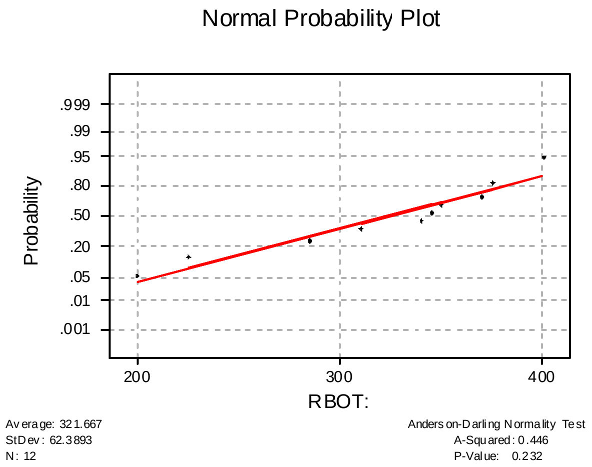
$$r = \frac{S_{xy}}{\sqrt{S_{xx}}\sqrt{S_{yy}}} = .9233.$$

b. The value of r does not depend on which of the two variables is labeled as the x variable. Thus, had we let x = RBOT time and y = TOST time, the value of r would have remained the same.

c. The value of r does no depend on the unit of measure for either variable. Thus, had we expressed RBOT time in hours instead of minutes, the value of r would have remained the same.

d.





Both TOST time and ROBT time appear to have come from normally distributed populations.

e. $H_o : r_1 = 0$ vs $H_a : r \neq 0$. $t = \dfrac{r\sqrt{n-2}}{\sqrt{1-r^2}}$; Reject H_o at level .05 if either $t \geq t_{.025,10} = 2.228$ or $t \leq -2.228$. r = .923, t = 7.58, so H_o should be rejected. The model is useful.

59.

a. $S_{xx} = 251{,}970 - \frac{(1950)^2}{18} = 40{,}720$, $S_{yy} = 130.6074 - \frac{(47.92)^2}{18} = 3.033711$,

and $S_{xy} = 5530.92 - \frac{(1950)(47.92)}{18} = 339.586667$, so

$r = \frac{339.586667}{\sqrt{40{,}720}\sqrt{3.033711}} = .9662$. There is a very strong positive correlation between the two variables.

b. Because the association between the variables is positive, the specimen with the larger shear force will tend to have a larger percent dry fiber weight.

c. Changing the units of measurement on either (or both) variables will have no effect on the calculated value of r, because any change in units will affect both the numerator and denominator of r by exactly the same multiplicative constant.

d. $r^2 = (.966)^2 = .933$

e. $H_o : r = 0$ vs $H_a : r > 0$. $t = \frac{r\sqrt{n-2}}{\sqrt{1-r^2}}$; Reject H_o at level .01 if

$t \ge t_{.01,16} = 2.583$. $t = \frac{.966\sqrt{16}}{\sqrt{1-.966^2}} = 14.94 \ge 2.583$, so H_o should be rejected .

The data indicates a positive linear relationship between the two variables.

60. $H_o : r = 0$ vs $H_a : r \ne 0$. $t = \frac{r\sqrt{n-2}}{\sqrt{1-r^2}}$; Reject H_o at level .01 if either

$t \ge t_{.005,22} = 2.819$ or $t \le -2.819$. r = .5778, t = 3.32, so H_o should be rejected. There appears to be a non-zero correlation in the population.

61.

a. We are testing $H_o : r = 0$ vs $H_a : r > 0$.

$r = \frac{7377.704}{\sqrt{36.9839}\sqrt{2{,}628{,}930.359}} = .7482$, and $t = \frac{.7482\sqrt{12}}{\sqrt{1-.7482^2}} = 3.9066$. We reject H_o since $t = 3.9066 \ge t_{.05,12} = 1.782$. There is evidence that a positive correlation exists between maximum lactate level and muscular endurance.

b. We are looking for r^2, the coefficient of determination. $r^2 = (.7482)^2 = .5598$. It is the same no matter which variable is the predictor.

62.

a. $H_o : r_1 = 0$ vs $H_a : r \neq 0$, Reject H$_o$ if; Reject H$_o$ at level .05 if either $t \geq t_{.025,12} = 2.179$ or $t \leq -2.179$. $t = \dfrac{r\sqrt{n-2}}{\sqrt{1-r^2}} = \dfrac{(.449)\sqrt{12}}{\sqrt{1-(.449)^2}} = 1.74$. Fail to reject H$_o$, the data does not suggest that the population correlation coefficient differs from 0.

b. $(.449)^2 = .20$ so 20 percent of the observed variation in gas porosity can be accounted for by variation in hydrogen content.

63. n = 6, $\Sigma x_i = 111.71, \Sigma x_i^2 = 2{,}724.7643, \Sigma y_i = 2.9, \Sigma y_i^2 = 1.6572$, and $\Sigma x_i y_i = 63.915$.

$r = \dfrac{(6)(63.915) - (111.71)(2.9)}{\sqrt{(6)(2{,}724.7943) - (111.73)^2} \cdot \sqrt{(6)(1.6572) - (2.9)^2}} = .7729$. $H_o : r_1 = 0$ vs $H_a : r \neq 0$; Reject H$_o$ at level .05 if $|t| \geq t_{.025,4} = 2.776$.

$t = \dfrac{(.7729)\sqrt{4}}{\sqrt{1-(.7729)^2}} = 2.436$. Fail to reject H$_o$. The data does not indicate that the population correlation coefficient differs from 0. This result may seem surprising due to the relatively large size of r (.77), however, it can be attributed to a small sample size (6).

64. $r = \dfrac{-757.6423}{\sqrt{(3756.96)(465.34)}} = -.5730$

a. $v = .5\ln\left(\dfrac{.427}{1.573}\right) = -.652$, so (12.11) is $-.652 \pm \dfrac{(1.645)}{\sqrt{26}} = (-.976, -.3290)$, and the desired interval for r is $(-.751, -.318)$.

b. $z = (-.652 + .549)\sqrt{23} = -.49$, so H$_o$ cannot be rejected at any reasonable level.

c. $r^2 = .328$

d. Again, $r^2 = .328$

65.

a. Although the normal probability plot of the x's appears somewhat curved, such a pattern is not terribly unusual when n is small; the test of normality presented in section 14.2 (p. 625) does not reject the hypothesis of population normality. The normal probability plot of the y's is much straighter.

b. $H_o : r_1 = 0$ will be rejected in favor of $H_a : r \neq 0$ at level .01 if $|t| \geq t_{.005,8} = 3.355$. $\Sigma x_i = 864, \Sigma x_i^2 = 78{,}142, \Sigma y_i = 138.0, \Sigma y_i^2 = 1959.1$ and $\Sigma x_i y_i = 12{,}322.4$, so $r = \dfrac{3992}{(186.8796)(23.3880)} = .913$ and $t = \dfrac{.913(2.8284)}{.4080} = 6.33 \geq 3.355$, so reject H$_o$. There does appear to be a linear relationship.

66.

a. We used Minitab to calculate the r$_i$'s: $r_1 = 0.192$, $r_2 = 0.382$, and $r_3 = 0.183$. It appears that the lag 2 correlation is best, but all of them are weak, based on the definitions given in the text.

b. $\dfrac{2}{\sqrt{100}} = .2$. We reject H$_o$ if $|r_i| \geq .2$. For all lags, r$_i$ does not fall in the rejection region, so we cannot reject H$_o$. There is not evidence of theoretical autocorrelation at the first 3 lags.

c. If we want an approximate .05 significance level for the simultaneous hypotheses, we would have to use smaller individual significance level. If the individual confidence levels were .95, then the simultaneous confidence levels would be approximately (.95)(.95)(.95) = .857.

67.

a. Because p-value = .00032 < $\alpha = .001$, H$_o$ should be rejected at this significance level.

b. Not necessarily. For this n, the test statistic *t* has approximately a standard normal distribution when $H_o : r_1 = 0$ is true, and a p-value of .00032 corresponds to $z = 3.60$ (or –3.60). Solving $3.60 = \dfrac{r\sqrt{498}}{\sqrt{1}} - r^2$ for r yields r = .159. This r suggests only a weak linear relationship between x and y, one that would typically have little practical import.

c. $t = 2.20 \geq t_{.025,9998} = 1.96$, so H$_o$ is rejected in favor of H$_a$. The value t = 2.20 is statistically significant -- it cannot be attributed just to sampling variability in the case $r = 0$. But with this n, r = .022 implies $r = .022$, which in turn shows an extremely weak linear relationship.

Supplementary Exercises

68.

a. $n = 8$, $\Sigma x_i = 207, \Sigma x_i^2 = 6799, \Sigma y_i = 621.8, \Sigma y_i^2 = 48{,}363.76$ and $\Sigma x_i y_i = 15{,}896.8$, which gives $\hat{b}_1 = \dfrac{-1538.20}{11{,}543} = -.133258$, $\hat{b}_0 = 81.173051$, and $y = 81.173 - .1333x$ as the equation of the estimated line.

b. We wish to test $H_0 : b_1 = 0$ vs $H_0 : b_1 \neq 0$. At level .01, H_o will be rejected (and the model judged useful) if either $t \geq t_{.005,6} = 3.707$ or $t \leq -3.707$. SSE = 8.732664, s = 1.206, and $t = \dfrac{-.1333}{1.206 / 37.985} = \dfrac{-.1333}{.03175} = -4.2$, which is ≤ -3.707, so we do reject H_o and find the model useful.

c. The larger the value of $\sum (x_i - \overline{x})^2$, the smaller will be $\hat{s}_{\hat{b}_1}$ and the more accurate the estimate will tend to be. For the given x_i's, $\sum (x_i - \overline{x})^2 = 1442.88$, whereas the proposed x values $x_1 = ... = x_4 = 0$, $x_5 = ... = x_8 = 50$, $\sum (x_i - \overline{x})^2 = 5000$. Thus the second set of x values is preferable to the first set. With just 3 observations at x = 0 and 3 at x = 50, $\sum (x_i - \overline{x})^2 = 3750$, which is again preferable to the first set of x_i's.

d. $\hat{b}_0 + \hat{b}_1(25) = 77.84$, and $s_{\hat{b}_0 + \hat{b}_1(25)} = s\sqrt{\dfrac{1}{n} + \dfrac{n(25 - \overline{x})^2}{n\Sigma x_i^2 - (\Sigma x_i)^2}}$

$= 1.206\sqrt{\dfrac{1}{8} + \dfrac{8(25 - 25.875)^2}{11.543}} = .426$, so the 95% CI is

$77.84 \pm (2.447)(.426) = 77.84 \pm 1.04 = (76.80, 78.88)$. The interval is quite narrow, only 2%. This is the case because the predictive value of 25% is very close to the mean of our predictor sample.

69.

a. The test statistic value is $t = \frac{\hat{b}_1 - 1}{s_{\hat{b}_1}}$, and H$_o$ will be rejected if either $t \geq t_{.025,11} = 2.201$ or $t \leq -2.201$. With $\Sigma x_i = 243, \Sigma x_i^2 = 5965, \Sigma y_i = 241, \Sigma y_i^2 = 5731$ and $\Sigma x_i y_i = 5805$, $\hat{b}_1 = .913819$, $\hat{b}_0 = 1.457072$, $SSE = 75.126$, $s = 2.613$, and $s_{\hat{b}_1} = .0693$, $t = \frac{.9138 - 1}{.0693} = -1.24$. Because –1.24 is neither ≤ -2.201 nor ≥ 2.201, H$_o$ cannot be rejected. It is plausible that $b_1 = 1$.

b. $r = \frac{16{,}902}{(136)(128.15)} = .970$

70.

a. sample size = 8

b. $\hat{y} = 326.976038 - (8.403964)x$. When x = 35.5, $\hat{y} = 28.64$.

c. Yes, the model utility test is statistically significant at the level .01.

d. $r = \sqrt{r^2} = \sqrt{0.9134} = 0.9557$

e. First check to see if the value x = 40 falls within the range of x values used to generate the least-squares regression equation. If it does not, this equation should not be used. Furthermore, for this particular model an x value of 40 yields a g value of –9.18, which is an impossible value for y.

71.

a. $r^2 = .5073$

b. $r = +\sqrt{r^2} = \sqrt{.5073} = .7122$ (positive because $\hat{b}_1$ is positive.)

c. We test test $H_0 : b_1 = 0$ vs $H_0 : b_1 \neq 0$. The test statistic t = 3.93 gives p-value = .0013, which is < .01, the given level of significance, therefore we reject H$_o$ and conclude that the model is useful.

d. We use a 95% CI for $m_{Y\cdot 50}$. $\hat{y}_{(50)} = .787218 + .007570(50) = 1.165718$, $t_{.025,15} = 2.131$, s = "Root MSE" = .020308, so

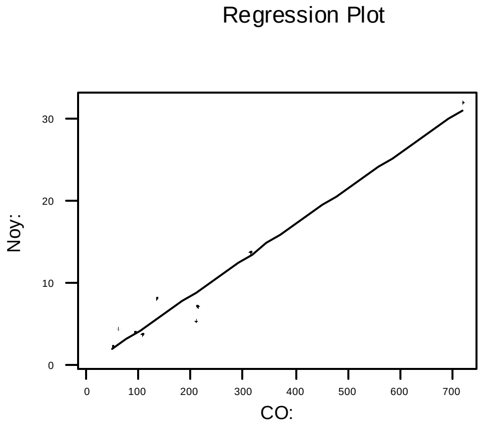
$$s_{\hat{y}(50)} = .20308\sqrt{\frac{1}{17} + \frac{17(50-42.33)^2}{17(41{,}575)-(719.60)^2}} = .051422$$. The interval is , then,

$$1.165718 \pm 2.131(.051422) = 1.165718 \pm .109581 = (1.056137, 1.275299).$$

e. $\hat{y}_{(30)} = .787218 + .007570(30) = 1.0143.$ The residual is $y - \hat{y} = .80 - 1.0143 = -.2143$.

72.

a.



The above analysis was created in Minitab. A simple linear regression model seems to fit the data well. The least squares regression equation is $\hat{y} = -.220 + .0436x$. The model utility test obtained from Minitab produces a t test statistic equal to 12.72. The corresponding p-value is extremely small. So we have sufficient evidence to claim that ΔCO is a good predictor of ΔNO_y.

b. $\hat{y} = -.220 + .0436(400) = 17.228$. A 95% prediction interval produced by Minitab is (11.953, 22.503). Since this interval is so wide, it does not appear that ΔNO_y is accurately predicted.

c. While the large ΔCO value appears to be "near" the least squares regression line, the value has extremely high leverage. The least squares line that is obtained when excluding the value is $\hat{y} = 1.00 + .0346x$. The r^2 value with the value included is 96% and is reduced to 75% when the value is excluded. The value of s with the value included is 2.024, and with the value excluded is 1.96. So the large ΔCO value does appear to effect our analysis in a substantial way.

73.

a. n = 9, $\Sigma x_i = 228, \Sigma x_i^2 = 5958, \Sigma y_i = 93.76, \Sigma y_i^2 = 982.2932$ and $\Sigma x_i y_i = 2348.15$, giving $\hat{b}_1 = \frac{-243.93}{1638} = -.148919$, $\hat{b}_0 = 14.190392$, and the equation $\hat{y} = 14.19 - (.1489)x$.

b. b_1 is the expected increase in load associated with a one-day age increase (so a negative value of b_1 corresponds to a decrease). We wish to test $H_0 : b_1 = -.10$ vs. $H_0 : b_1 < -.10$ (the alternative contradicts prior belief). H$_o$ will be rejected at level .05 if $t = \frac{\hat{b}_1 - (-.10)}{s_{\hat{b}_1}} \le -t_{.05,7} = -1.895$. With SSE = 1.4862, s = .4608, and $s_{\hat{b}_1} = \frac{.4608}{\sqrt{182}} = .0342$. Thus $t = \frac{-.1489 + 1}{.0342} = -1.43$. Because –1.43 is not ≤ -1.895, do not reject H$_o$.

c. $\Sigma x_i = 306, \Sigma x_i^2 = 7946,$ so $\sum (x_i - \overline{x})^2 = 7946 - \frac{(306)^2}{12} = 143$ here, as contrasted with 182 for the given 9 x_i's. Even though the sample size for the proposed x values is larger, the original set of values is preferable.

d. $(t_{.025,7})(s)\sqrt{\frac{1}{9} + \frac{9(28 - 25.33)^2}{1638}} = (2.365)(.4608)(.3877) = .42$, and $\hat{b}_0 + \hat{b}_1(28) = 10.02,$ so the 95% CI is $10.02 \pm .42 = (9.60, 10.44)$.

74.

a. $\hat{b}_1 = \frac{3.5979}{44.713} = .0805$, $\hat{b}_0 = 1.6939$, $\hat{y} = 1.69 + (.0805)x$.

b. $\hat{b}_1 = \frac{3.5979}{.2943} = 12.2254$, $\hat{b}_0 = -20.4046$, $\hat{y} = -20.40 + (12.2254)x$.

c. r = .992, so r^2 = .984 for either regression.

75.

a. The plot suggests a strong linear relationship between x and y.

b. n = 9, $\Sigma x_i = 1797, \Sigma x_i^2 = 4334.41, \Sigma y_i = 7.28, \Sigma y_i^2 = 7.4028$ and $\Sigma x_i y_i = 178.683$, so $\hat{b}_1 = \dfrac{299.931}{6717.6} = .04464854$, $\hat{b}_0 = -.08259353$, and the equation of the estimated line is $\hat{y} = -.08259 - (.044649)x$.

c. $SSE = 7.4028 - (-601281) - 7.977935 = .026146,$

$SST = 7.4028 - \dfrac{(7.28)^2}{9} = .026146, = 1.5141$, and $r^2 = 1 - \dfrac{SSE}{SST} = .983$, so 93.8% of the observed variation is "explained."

d. $\hat{y}_4 = -.08259 - (.044649)(19.1) = .7702$, and
$y_4 - \hat{y}_4 = .68 - .7702 = -.0902$.

e. s = .06112, and $s_{\hat{b}_1} = \dfrac{.06112}{\sqrt{746.4}} = .002237$, so the value of t for testing $H_0 : b_1 = 0$ vs $H_0 : b_1 \neq 0$ is $t = \dfrac{.044649}{.002237} = 19.96$. From Table A.5, $t_{.0005,7} = 5.408$, so $p - value < 2(.0005) = .001$. There is strong evidence for a useful relationship.

f. A 95% CI for b_1 is $.044649 \pm (2.365)(.002237) = .044649 \pm .005291$
$= (.0394, .0499)$.

g. A 95% CI for $b_0 + b_1(20)$ is $.810 \pm (2.365)(.002237)(.3333356)$
$= .810 \pm .048 = (.762, .858)$

76. Substituting x* = 0 gives the CI $\hat{b}_0 \pm t_{a/2,n-2} \cdot s\sqrt{\dfrac{1}{n} + \dfrac{n\bar{x}^2}{n\Sigma x_i^2 - (\Sigma x_i)^2}}$. From Example 12.8, $\hat{b}_0 = 3.621$, SSE = .262453, n = 14, $\Sigma x_i = 890, \bar{x} = 63.5714, \Sigma x_i^2 = 67{,}182,$ so with s = .1479, $t_{.025,12} = 2.179$, the CI is $3.621 \pm 2.179(.1479)\sqrt{\dfrac{1}{12} + \dfrac{56{,}578.52}{148{,}448}}$

$$= 3.621 \pm 2.179(.1479)(.6815) = 3.62 \pm .22 = (3.40, 3.84).$$

77. $SSE = \Sigma y^2 - \hat{b}_0\Sigma y - \hat{b}_1\Sigma xy$. Substituting $\hat{b}_0 = \dfrac{\Sigma y - \hat{b}_1\Sigma x}{n}$, SSE becomes

$$SSE = \Sigma y^2 - \frac{\Sigma y\left(\Sigma y - \hat{b}_1\Sigma x\right)}{n} - \hat{b}_1\Sigma xy = \Sigma y^2 - \frac{(\Sigma y)^2}{n} + \frac{\hat{b}_1\Sigma x\Sigma y}{n} - \hat{b}_1\Sigma xy$$

$$= \left[\Sigma y^2 - \frac{(\Sigma y)^2}{n}\right] - \hat{b}_1\left[\Sigma xy - \frac{\Sigma x\Sigma y}{n}\right] = S_{yy} - \hat{b}_1 S_{xy}$$, as desired.

78. The value of the sample correlation coefficient using the squared y values would not necessarily be approximately 1. If the y values are greater than 1, then the squared y values would differ from each other by more than the y values differ from one another. Hence, the relationship between x and y^2 would be less like a straight line, and the resulting value of the correlation coefficient would decrease.

79.

a. With $s_{xx} = \sum\left(x_i - \overline{x}\right)^2$, $s_{yy} = \sum\left(y_i - \overline{y}\right)^2$, note that $\dfrac{s_y}{s_x} = \sqrt{\dfrac{s_{yy}}{s_{xx}}}$ (since the factor n-1 appears in both the numerator and denominator, so cancels). Thus

$$y = \hat{b}_0 + \hat{b}_1 x = \overline{y} + \hat{b}_1\left(x - \overline{x}\right) = \overline{y} + \frac{s_{xy}}{s_{xx}}\left(x - \overline{x}\right) = \overline{y} + \sqrt{\frac{s_{yy}}{s_{xx}}} \cdot \frac{s_{xy}}{\sqrt{s_{xx}s_{yy}}}\left(x - \overline{x}\right)$$

$$= \overline{y} + \frac{s_y}{s_x} \cdot r \cdot \left(x - \overline{x}\right)$$, as desired.

b. By .573 s.d.'s above, (above, since r < 0) or (since $s_y = 4.3143$) an amount 2.4721 above.

80. With s_{xy} given in the text, $r = \frac{s_{xy}}{\sqrt{s_{xx}s_{yy}}}$ (where e.g. $s_{xx} = \sum(x_i - \overline{x})^2$), and

$\hat{b}_1 = \frac{s_{xy}}{s_{xx}}$. Also, $s = \sqrt{\frac{SSE}{n-2}}$ and $SSE = \Sigma y_i^2 - \hat{b}_0\Sigma y_i - \hat{b}_1\Sigma x_i y_i = s_{yy} - \hat{b}_1 s_{xy}$.

Thus the t statistic for $H_o : \hat{b}_1 = 0$ is

$$t = \frac{\hat{b}_1}{s / \sqrt{\sum(x_i - \overline{x})^2}} = \frac{(s_{xy} / s_{xx}) \cdot \sqrt{s_{xx}}}{\sqrt{(s_{yy} - s_{xy}^2 / s_{xx})/(n-2)}}$$

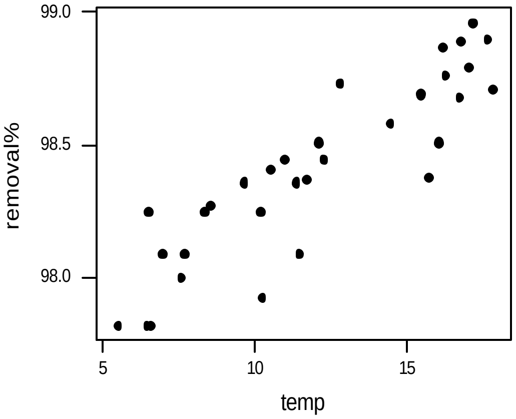
$$= \frac{s_{xy} \cdot \sqrt{n-2}}{\sqrt{(s_{xx}s_{yy} - s_{xy}^2)}} = \frac{(s_{xy} / \sqrt{s_{xx}s_{yy}})\sqrt{n-2}}{\sqrt{1 - s_{xy}^2 / s_{xx}s_{yy}}} = \frac{r\sqrt{n-2}}{\sqrt{1-r^2}}$$ as desired.

81. Using the notation of the exercise above, $SST = s_{yy}$ and $SSE = s_{yy} - \hat{b}_1 s_{xy}$

$$= s_{yy} - \frac{s_{xy}^2}{s_{xx}}, \text{ so } 1 - \frac{SSE}{SST} = 1 - \frac{s_{yy} - \frac{s_{xy}^2}{s_{xx}}}{s_{yy}} = \frac{s_{xy}^2}{s_{xx}s_{yy}} = r^2$$, as desired.

82.

a. A Scatter Plot suggests the linear model is appropriate.



b. Minitab Output:

```
The regression equation is
removal% = 97.5 + 0.0757 temp

Predictor        Coef       StDev          T        P
Constant      97.4986      0.0889    1096.17    0.000
temp         0.075691    0.007046      10.74    0.000

S = 0.1552      R-Sq = 79.4%     R-Sq(adj) = 78.7%

Analysis of Variance

Source            DF          SS          MS         F        P
Regression         1      2.7786      2.7786    115.40    0.000
Residual Error    30      0.7224      0.0241
Total             31      3.5010
```

Minitab will output all the residual information if the option is chosen, from which you can find the point prediction value $\hat{y}_{10.5} = 98.2933$, the observed value y = 98.41, so the residual = .0294.

c. Roughly .1

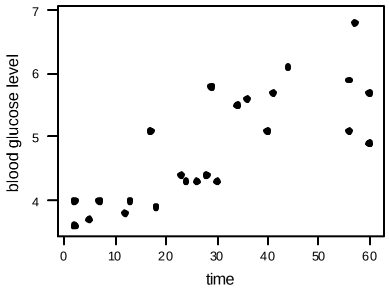
d. R^2 = 79.4

e. A 95% CI for β_1, using $t_{.025,30} = 2.042$:

$$.075691 \pm 2.042(.007046) = (.061303,.090079)$$

f. The slope of the regression line is steeper. The value of s is almost doubled, and the value of R^2 drops to 61.6%.

83. Using Minitab, we create a scatterplot to see if a linear regression model is appropriate.



A linear model is reasonable; although it appears that the variance in y gets larger as x increases. The Minitab output follows:

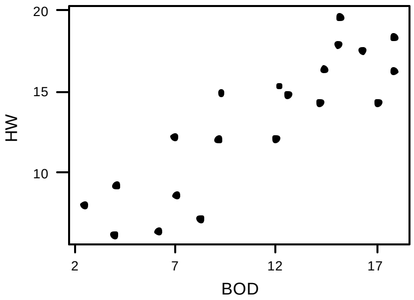
```
The regression equation is
blood glucose level = 3.70 + 0.0379 time

Predictor        Coef       StDev          T        P
Constant       3.6965      0.2159      17.12    0.000
time         0.037895    0.006137       6.17    0.000

S = 0.5525      R-Sq = 63.4%     R-Sq(adj) = 61.7%

Analysis of Variance

Source            DF          SS          MS         F        P
Regression         1      11.638      11.638     38.12    0.000
Residual Error    22       6.716       0.305
Total             23      18.353
```

The coefficient of determination of 63.4% indicates that only a moderate percentage of the variation in y can be explained by the change in x. A test of model utility indicates that time is a significant predictor of blood glucose level. (t = 6.17, p = 0.0). A point estimate for blood glucose level when time = 30 minutes is 4.833%. We would expect the average blood glucose level at 30 minutes to be between 4.599 and 5.067, with 95% confidence.

84.

a. Using the techniques from a previous chapter, we can do a t test for the difference of two means based on paired data. Minitab's paired t test for equality of means gives t = 3.54, with a p value of .002, which suggests that the average bf% reading for the two methods is not the same.

b. Using linear regression to predict HW from BOD POD seems reasonable after looking at the scatterplot, below.



The least squares linear regression equation, as well as the test statistic and p value for a model utility test, can be found in the Minitab output below. We see that we do have significance, and the coefficient of determination shows that about 75% of the variation in HW can be explained by the variation in BOD.

```
The regression equation is
HW = 4.79 + 0.743 BOD

Predictor        Coef       StDev          T        P
Constant        4.788       1.215       3.94    0.001
BOD            0.7432      0.1003       7.41    0.000

S = 2.146       R-Sq = 75.3%     R-Sq(adj) = 73.9%

Analysis of Variance

Source            DF          SS          MS         F        P
Regression         1      252.98      252.98     54.94    0.000
Residual Error    18       82.89        4.60
Total             19      335.87
```

85. For the second boiler, $n = 19$, $\Sigma x_i = 125$, $\Sigma y_i = 472.0$, $\Sigma x_i^2 = 3625$, $\Sigma y_i^2 = 37{,}140.82$, and $\Sigma x_i y_i = 9749.5$, giving $\hat{g}_1 =$ estimated slope $= \dfrac{-503}{6125} = -.0821224$, $\hat{g}_0 = 80.377551$, $SSE_2 = 3.26827$, $SSx_2 = 1020.833$.

For boiler #1, n = 8, $\hat{b}_1 = -.1333$, $SSE_1 = 8.733$, and $SSx_1 = 1442.875$. Thus $\hat{s}^2 = \dfrac{8.733 + 3.286}{10} = 1.2$, $\hat{s} = 1.095$, and $t = \dfrac{-.1333 + .0821}{1.095\sqrt{\frac{1}{1442.875} + \frac{1}{1020.833}}}$

$= \dfrac{-.0512}{.0448} = -1.14$. $t_{.025,10} = 2.228$ and –1.14 is neither ≥ 2.228 nor ≤ -2.228, so H_o is not rejected. It is plausible that $b_1 = g_1$.