CHAPTER 11

Section 11-2

11-1 a) $y_i = \beta_0 + \beta_1 x_i + \varepsilon_i$

$S_{xx} = 157.42 - \frac{43^2}{14} = 25.348571$

$S_{xy} = 1697.80 - \frac{43(572)}{14} = -59.057143$

$\hat{\beta}_1 = \frac{S_{xy}}{S_{xx}} = \frac{-59.057143}{25.348571} = -2.330$

$\hat{\beta}_0 = \bar{y} - \hat{\beta}_1 \bar{x} = \frac{572}{14} - (-2.3298017)(\frac{43}{14}) = 48.013$

$SS_R = \hat{\beta}_1 S_{xy} = -2.3298017(-59.057143)$

$= 137.59$

$SS_E = S_{yy} - SS_R$

$= 159.71429 - 137.59143$

$= 22.123$

$\hat{\sigma}^2 = MS_E = \frac{SS_E}{n-2} = \frac{22.123}{12} = 1.8436$

b) $\hat{y} = \hat{\beta}_0 + \hat{\beta}_1 x$, $\hat{y} = 48.012962 - 2.3298017(4.5)$

$= 37.53$

c) $\hat{y} = 48.012962 - 2.3298017(3.3) = 40.32$

d) $e_i = y_i - \hat{y}_i = 45.6 - 37.53 = 8.07$

11-2 a) $y_i = \beta_0 + \beta_1 x_i + \varepsilon_i$

$S_{xx} = 143215.8 - \frac{1478^2}{20} = 33991.6$

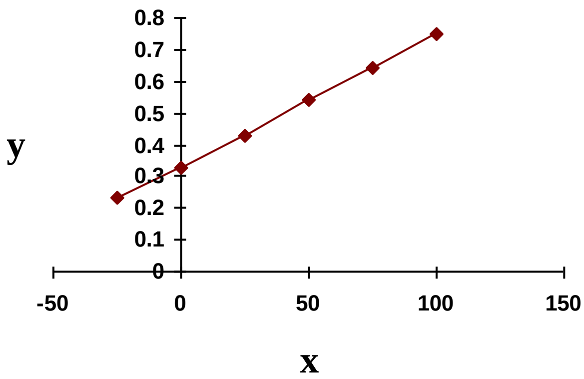
$S_{xy} = 1083.67 - \frac{(1478)(12.75)}{20} = 141.445$

$\hat{\beta}_1 = \frac{S_{xy}}{S_{xx}} = \frac{141.445}{33991.6} = 0.00416$

$\hat{\beta}_0 = \frac{12.75}{20} - (0.0041617512)(\frac{1478}{20}) = 0.32999$

$\hat{y} = 0.32999 + 0.00416x$

$\hat{\sigma}^2 = MS_E = \frac{SS_E}{n-2} = \frac{0.143275}{18} = 0.00796$



b) $\hat{y} = 0.32999 + 0.00416(100) = 0.746$

c) $\hat{y} = 0.32999 + 0.00416(90) = 0.7044$

d) $\hat{\beta}_1 = 0.00416$

11-3 a)

Regression Analysis: Rating Points versus Meters per Att

```
The regression equation is
y = 14.2 + 11.0 x

Predictor      Coef  SE Coef     T      P
Constant     14.185    9.089  1.56  0.129
x            11.039    1.413  7.81  0.000

S = 5.22972   R-Sq = 67.0%   R-Sq(adj) = 65.9%

Analysis of Variance
Source          DF      SS      MS      F      P
Regression       1  1669.0  1669.0  61.02  0.000
Residual Error  30   820.5    27.4
Total           31  2489.5
```

$$y_i = \beta_0 + \beta_1 x_i + \varepsilon_i$$

$$S_{xx} = 1323.648 - \frac{(204.74)^2}{32} = 13.696$$

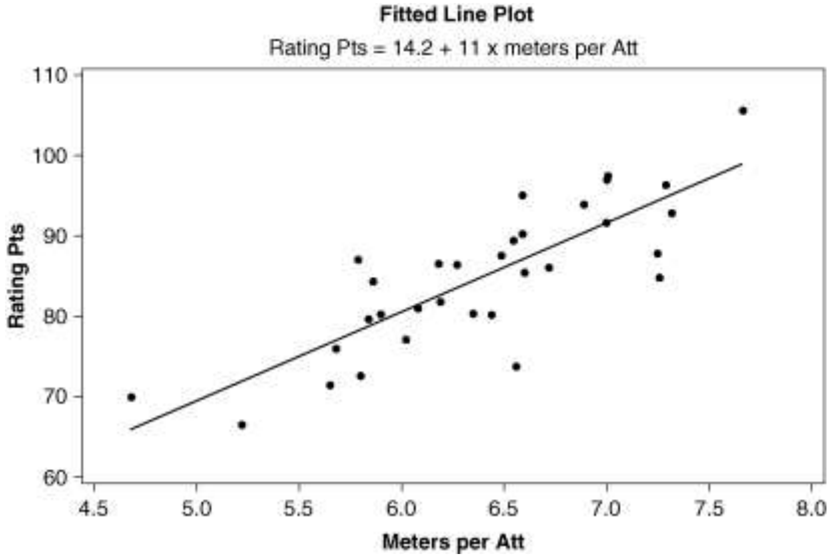
$$S_{xy} = 17516.34 - \frac{(204.74)(2714.1)}{32} = 151.1889$$

$$\hat{\beta}_1 = \frac{S_{xy}}{S_{xx}} = \frac{151.1889}{13.696} = 11.039$$

$$\hat{\beta}_0 = \frac{2714.1}{32} - (11.039)\left(\frac{204.74}{32}\right) = 14.187$$

$$\hat{y} = 14.2 + 11x$$

$$\hat{\sigma}^2 = MS_E = \frac{SS_E}{n-2} = \frac{820.5}{30} = 27.35$$



b) $\hat{y} = 14.2 + 11(6.9) = 90.1$

c) $-\hat{\beta}_1 = -11$

d) $\frac{1}{11} \times 10 = 0.91$

e) $\hat{y} = 14.2 + 11(6.59) = 86.69$

There are two residuals

$$e = y - \hat{y}$$
$$e_1 = 90.2 - 86.69 = 3.51$$
$$e_2 = 95 - 86.69 = 8.31$$

11-4 a)

```
Regression Analysis - Linear model: Y = a+bX
Dependent variable: SalePrice                Independent variable: Taxes
---------------------------------------------------------------------------
                                 Standard            T              Prob.
Parameter          Estimate        Error            Value           Level
Intercept           13.3202       2.57172         5.17948          .00003
Slope               3.32437      0.390276           8.518          .00000
---------------------------------------------------------------------------
```

Analysis of Variance

Source	Sum of Squares	Df	Mean Square	F-Ratio	Prob. Level
Model	636.15569	1	636.15569	72.5563	.00000
Residual	192.89056	22	8.76775		
Total (Corr.)	829.04625	23			

```
Correlation Coefficient = 0.875976          R-squared =  76.73 percent
Stnd. Error of Est. = 2.96104
```

$\hat{\sigma}^2 = 8.76775$

If the calculations were to be done by hand, use Equations (11-7) and (11-8).

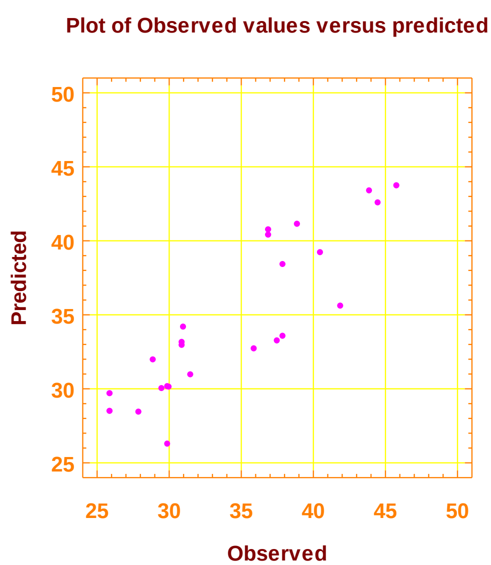
$\hat{y} = 13.3202 + 3.32437x$

b) $\hat{y} = 13.3202 + 3.32437(7.3) = 37.588$

c) $\hat{y} = 13.3202 + 3.32437(5.6039) = 31.9496$

$\hat{y} = 31.9496$
$e = y - \hat{y} = 28.9 - 31.9496 = -3.0496$

d) All the points would lie along a 45 degree line. That is, the regression model would estimate the values exactly. At this point, the graph of observed vs. predicted indicates that the simple linear regression model provides a reasonable fit to the data.



11-5 a)

```
Regression Analysis - Linear model: Y = a+bX
Dependent variable: Usage               Independent variable: Temperature
---------------------------------------------------------------------------
                               Standard
Parameter       Estimate          Error            T              Prob.
Intercept        129.974          0.707         183.80            0.000
Slope            7.59262        0.05798         130.95            0.000
---------------------------------------------------------------------------
                           Analysis of Variance
Source       DF    Sum of Squares    Mean Square          F      Prob. Level
Model         1             57701          57701    7148.85            0.000
Residual     10                34              3
Total        11             57734
---------------------------------------------------------------------------
Stnd. Error of Est. = 1.83431    R-Sq = 99.9%
Correlation Coefficient = 0.9999
```

$\hat{\sigma}^2 = 3$

If the calculations were to be done by hand, use Equations (11-7) and (11-8).

$\hat{y} = 130 + 7.59x$

b) $\hat{y} = -130 + 7.59(13) = 228.67$

c) If monthly temperature increases by 0.5°C, $\hat{y}$ increases by 7.59

d) $\hat{y} = -130 + 7.59(8) = 190.72$

$\hat{y} = 190.72$

$e = y - \hat{y} = 192.70 - 190.72 = 1.98$

11-6 a)

```
The regression equation is
MPG = 39.2 - 0.0402 Engine Displacement

Predictor                Coef   SE Coef      T      P
Constant               39.156     2.006  19.52  0.000
Engine Displacement  -0.040216  0.007671  -5.24  0.000

S = 3.74332   R-Sq = 59.1%   R-Sq(adj) = 57.0%

Analysis of Variance

Source          DF      SS      MS      F      P
Regression       1  385.18  385.18  27.49  0.000
Residual Error  19  266.24   14.01
Total           20  651.41
```

$\hat{\sigma}^2 = 14.01$

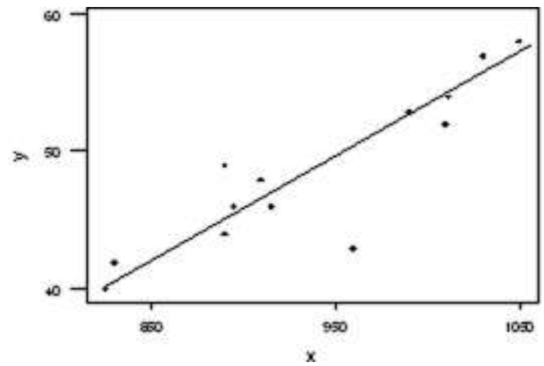
$\hat{y} = 39.2 - 0.0402x$

b) $\hat{y} = 39.2 - 0.0402(175) = 32.165$

c) $\hat{y} = 31.32$

$e = y - \hat{y} = 35.4 - 31.32 = 4.08$

11-7 a)



Predictor	Coef	StDev	T	P
Constant	-16.509	9.843	-1.68	0.122
x	0.06936	0.01045	6.64	0.000

S = 2.706 R-Sq = 80.0% R-Sq(adj) = 78.2%

Analysis of Variance

Source	DF	SS	MS	F	P
Regression	1	322.50	322.50	44.03	0.000
Error	11	80.57	7.32		
Total	12	403.08			

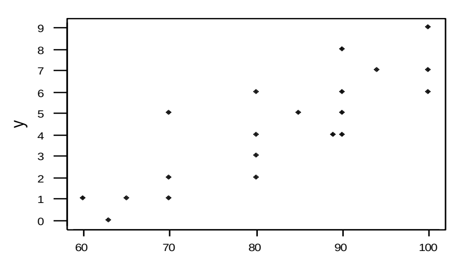
$\hat{\sigma}^2 = 7.3212$

$\hat{y} = -16.5093 + 0.0693554x$

b) $\hat{y} = 46.9509$ $e = 46.9509 - 46 = 0.9509$

c) $\hat{y} = -16.5093 + 0.0693554(960) = 50.07$

11-8 a)



Yes, a linear regression would seem appropriate, but one or two points might be outliers.

Predictor	Coef	SE Coef	T	P
Constant	-10.132	1.995	-5.08	0.000
x	0.17429	0.02383	7.31	0.000

S = 1.318   R-Sq = 74.8%   R-Sq(adj) = 73.4%

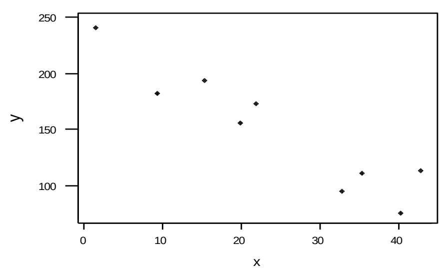
Analysis of Variance

Source	DF	SS	MS	F	P
Regression	1	92.934	92.934	53.50	0.000
Residual Error	18	31.266	1.737		
Total	19	124.200			

b) $\hat{\sigma}^2 = 1.737$ and $\hat{y} = -10.132 + 0.17429x$

c) $\hat{y} = 5.5541$ at $x = 90$

11-9 a)



Yes, a linear regression model appears to be plausible.

Predictor	Coef	StDev	T	P
Constant	234.07	13.75	17.03	0.000
x	-3.5086	0.4911	-7.14	0.000

S = 19.96   R-Sq = 87.9%   R-Sq(adj) = 86.2%

Analysis of Variance

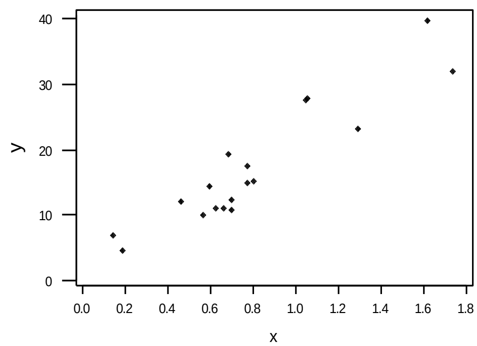
Source	DF	SS	MS	F	P
Regression	1	20329	20329	51.04	0.000
Error	7	2788	398		
Total	8	23117			

b) $\hat{\sigma}^2 = 398.25$ and $\hat{y} = 234.071 - 3.50856x$

c) $\hat{y} = 234.071 - 3.50856(36) = 107.763$

d) $\hat{y} = 163.90 \quad e = -8.90$

11-10 a)



Yes, a simple linear regression model seems appropriate for these data.

Predictor	Coef	StDev	T	P
Constant	0.470	1.936	0.24	0.811
x	20.567	2.142	9.60	0.000

S = 3.716 R-Sq = 85.2% R-Sq(adj) = 84.3%

Analysis of Variance

Source	DF	SS	MS	F	P
Regression	1	1273.5	1273.5	92.22	0.000
Error	16	220.9	13.8		
Total	17	1494.5			

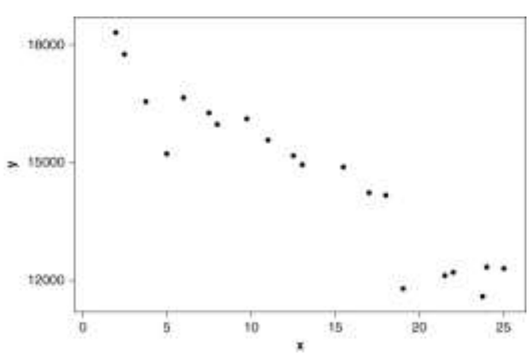
b) $\hat{\sigma}^2 = 13.81$

$\hat{y} = 0.470467 + 20.5673x$

c) $\hat{y} = 0.470467 + 20.5673(1) = 21.038$

d) $\hat{y} = 13.42787 \quad e = 2.5279 \text{ for } x = 0.63$

11-11 a)



Yes, a simple linear regression (straight-line) model seems plausible for this situation.

Predictor	Coef	SE Coef	T	P
Constant	18090.2	310.8	58.20	0.000
x	-254.55	20.34	-12.52	0.000

S = 678.964 R-Sq = 89.7% R-Sq(adj) = 89.1%

Analysis of Variance

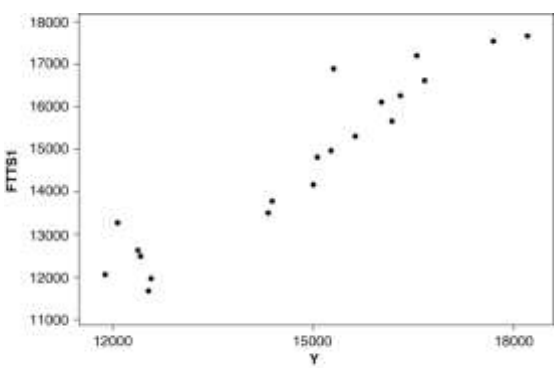
Source	DF	SS	MS	F	P
Regression	1	72222688	72222688	156.67	0.000
Residual Error	18	8297849	460992		
Total	19	80520537			

b) $\hat{\sigma}^2 = 460992$

$\hat{y} = 18090.2 - 254.55x$

c) $\hat{y} = 18090.2 - 254.55(20) = 12999.2$

d) If there were no error, the values would all lie along the 45° line. The plot indicates age is a reasonable regressor variable.



11-12 a)

```
The regression equation is
Porosity = 55.6 - 0.0342 Temperature
```

Predictor	Coef	SE Coef	T	P
Constant	55.63	32.11	1.73	0.144
Temperature	-0.03416	0.02569	-1.33	0.241

S = 8.79376 R-Sq = 26.1% R-Sq(adj) = 11.3%

Analysis of Variance

Source	DF	SS	MS	F	P
Regression	1	136.68	136.68	1.77	0.241

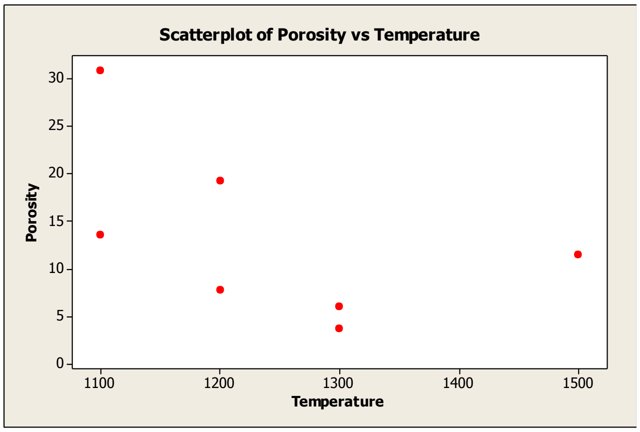
```
Residual Error   5  386.65   77.33
Total            6  523.33
```

b) $\hat{y} = 55.63 - 0.03416x$

$\hat{\sigma}^2 = 77.33$

c) $\hat{y} = 55.63 - 0.03416(1700) = -2.442$

d) $\hat{y} = 14.638 \;\; e = 4.562$



The simple linear regression model doesn't seem appropriate because the scatter plot doesn't indicate a linear relationship.

11-13 a)

```
The regression equation is
BOD = 0.658 + 0.178 Time

Predictor      Coef  SE Coef      T      P
Constant     0.6578   0.1657   3.97  0.003
Time        0.17806  0.01400  12.72  0.000

S = 0.287281   R-Sq = 94.7%   R-Sq(adj) = 94.1%

Analysis of Variance
Source          DF      SS      MS       F      P
Regression       1  13.344  13.344  161.69  0.000
Residual Error   9   0.743   0.083
Total           10  14.087
```

$\hat{y} = 0.658 + 0.178x$

$\hat{\sigma}^2 = 0.083$

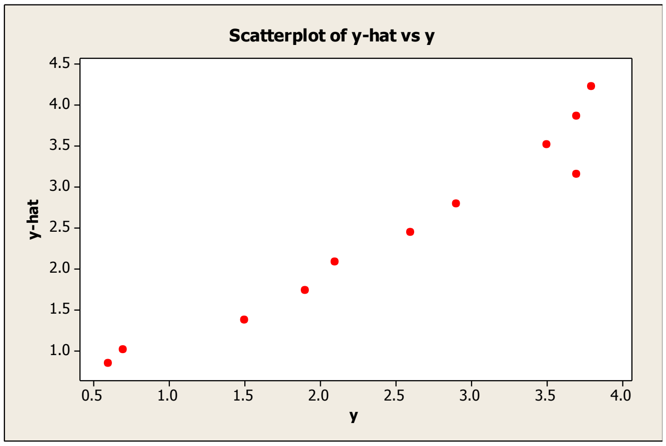
b) $\hat{y} = 0.658 + 0.178(22) = 4.574$

c) 0.178(3) = 0.534

d) $\hat{y} = 0.658 + 0.178(6) = 1.726$

$e = y - \hat{y} = 1.9 - 1.726 = 0.174$

e)

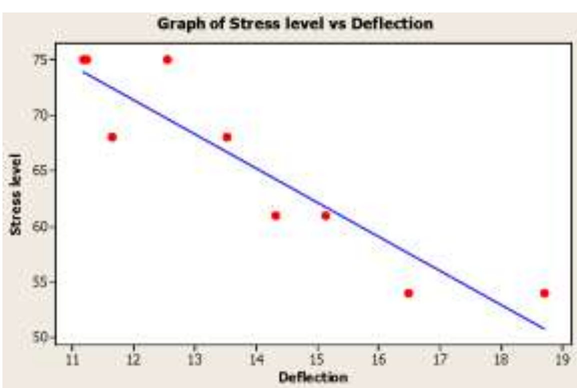


All the points would lie along the 45 degree line $y = \hat{y}$. That is, the regression model would estimate the values exactly. At this point, the graph of observed vs. predicted indicates that the simple linear regression model provides a reasonable fit to the data.

11-14 a)

```
The regression equation is
Deflection = 32.0 - 0.277 Stress level

Predictor          Coef   SE Coef      T      P
Constant         32.049     2.885  11.11  0.000
Stress level   -0.27712   0.04361  -6.35  0.000

S = 1.05743   R-Sq = 85.2%   R-Sq(adj) = 83.1%

Analysis of Variance

Source          DF      SS      MS      F      P
Regression       1  45.154  45.154  40.38  0.000
Residual Error   7   7.827   1.118
Total            8  52.981
```

$\hat{\sigma}^2 = 1.118$



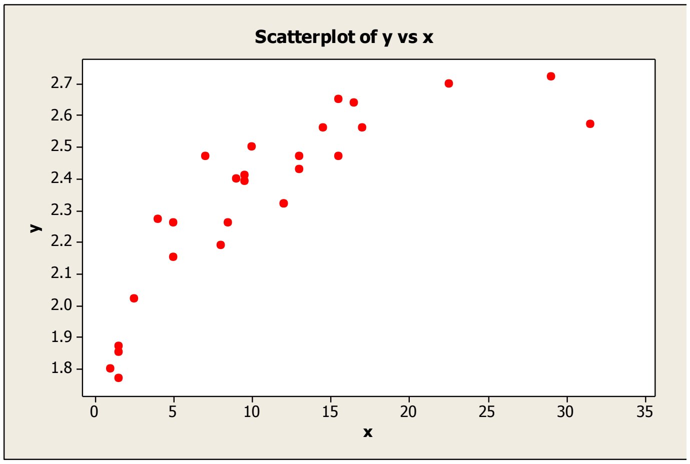
b) $\hat{y} = 32.05 - 0.277(64) = 14.322$

c) $(-0.277)(5) = -1.385$

d) $\dfrac{1}{0.277} = 3.61$

e) $\hat{y} = 32.05 - 0.277(75) = 11.275 \quad e = y - \hat{y} = 12.534 - 11.275 = 1.259$

11-15



It's possible to fit this data with linear model, but it's not a good fit. Curvature is seen on the scatter plot.

a)

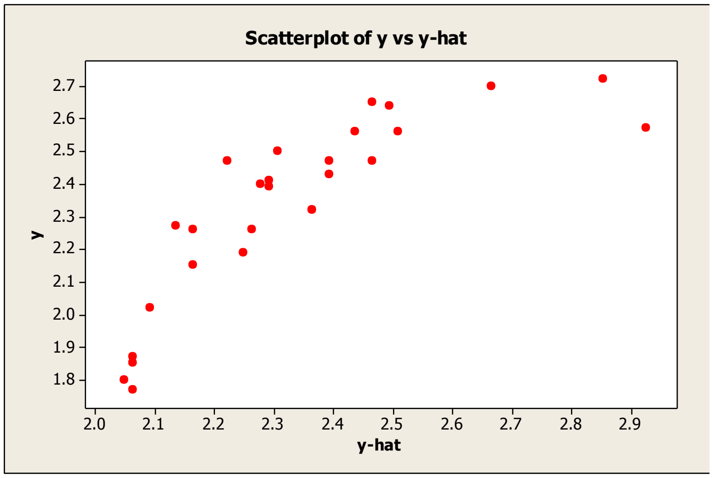
```
The regression equation is
y = 2.02 + 0.0287 x

Predictor       Coef    SE Coef      T      P
Constant     2.01977    0.05313  38.02  0.000
x           0.028718   0.003966   7.24  0.000

S = 0.159159   R-Sq = 67.7%   R-Sq(adj) = 66.4%

Analysis of Variance

Source          DF      SS      MS      F      P
Regression       1  1.3280  1.3280  52.42  0.000
Residual Error  25  0.6333  0.0253
Total           26  1.9613
```

$\hat{y} = 2.02 + 0.0287x$

$\hat{\sigma}^2 = 0.0253$

b) $\hat{y} = 2.02 + 0.0287(16) = 2.4792$

c)



If the relationship between length and age was deterministic, the points would fall on the 45 degree line $y = \hat{y}$. The plot does not indicate a linear relationship. Therefore, age is not a reasonable choice for the regressor variable in this model.

11-16 a) $\hat{y} = 0.3299892 + 0.0041612(\frac{9}{5}x + 32)$

$$\hat{y} = 0.3299892 + 0.0074902x + 0.1331584$$
$$\hat{y} = 0.4631476 + 0.0074902x$$

b) $\hat{\beta}_1 = 0.00749$

11-17 Let x = engine displacement (cm^3) and x_{old} = engine displacement (in^3)

a) The old regression equation is y = 39.2 - 0.0402x_{old}

Because 1 in^3 = 16.387 cm^3, the new regression equation is

$$\hat{y} = 39.2 - 0.0402(x/16.387) = 39.2 - 0.0025x$$

b) $\hat{\beta}_1 = -0.0025$

11-18 $\hat{\beta}_0 + \hat{\beta}_1\bar{x} = (\bar{y} - \hat{\beta}_1\bar{x}) + \hat{\beta}_1\bar{x} = \bar{y}$

11-19 a) The slope and the intercept will be shifted.

b) $\hat{y} = 2132.41 + 36.96z$

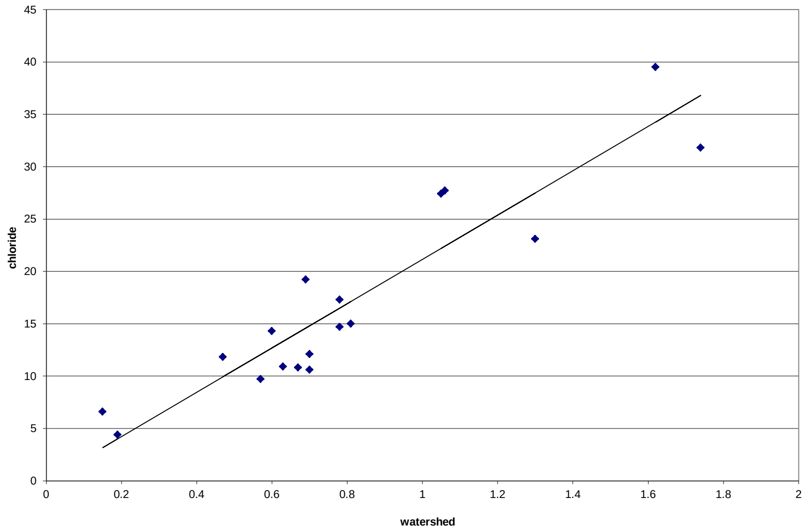
$$\hat{\beta}_0 = 2625.39 \qquad \text{vs.} \qquad \hat{\beta}_0^* = 2132.41$$
$$\hat{\beta}_1 = -36.96 \qquad\qquad \hat{\beta}_1^* = 36.96$$

11-20 a) The least squares estimate minimizes $\sum (y_i - \beta x_i)^2$. Upon setting the derivative equal to zero, we obtain

$$2\sum (y_i - \beta x_i)\ (-x_i) = 2[\sum -y_i x_i + \beta \sum x_i^2] = 0$$

Therefore, $\hat{\beta} = \dfrac{\sum y_i x_i}{\sum x_i^2}$.

b) $\hat{y} = 21.031461x$. The model seems very appropriate—an even better fit.



Section 11-4

11-21 a) $$T_0 = \frac{\hat{\beta}_0 - \beta_0}{se(\beta_0)} = \frac{12.857}{1.032} = 12.4583$$

P-value = 2[P(T_8 > 12.4583)] and P-value < 2(0.0005) = 0.001

$$T_1 = \frac{\hat{\beta}_1 - \beta_1}{se(\beta_1)} = \frac{2.3445}{0.137} = 17.113$$

P-value = 2[P(T_8 > 17.113)] and P-value < 2(0.0005) = 0.001

$$MS_E = \frac{SS_E}{n-2} = \frac{17.55}{8} = 2.1938$$

$$F_0 = \frac{MS_R}{MS_E} = \frac{912.43}{2.1938} = 415.913$$

P-value is near zero

b) Because the P-value of the F-test ≈ 0 is less than α = 0.05, we reject the null hypothesis that $\beta_1 = 0$ at the 0.05 level of significance. This is the same result obtained from the T_1 test. If the assumptions are valid, a useful linear relationship exists.

c) $\hat{\sigma}^2 = MS_E = 2.1938$

11-22 a) $$T_0 = \frac{\hat{\beta}_0 - \beta_0}{se(\beta_0)} = \frac{26.753}{2.373} = 11.2739$$

P-value = 2[P(T_{14} > 11.2739)] and P-value < 2(0.0005) = 0.001

$$T_1 = \frac{\hat{\beta}_1 - \beta_1}{se(\beta_1)} = \frac{1.4756}{0.1063} = 13.8815$$

P-value = 2[P(T_{14} > 13.8815)] and P-value < 2(0.0005) = 0.001

Degrees of freedom of the residual error = 15 – 1 = 14.
Sum of squares regression = Sum of square Total – Sum of square residual error = 1500 – 94.8 = 1405.2

$$MS_{\text{Regression}} = \frac{SS_{\text{Regression}}}{1} = \frac{1405.2}{1} = 1405.2$$

$$F_0 = \frac{MS_R}{MS_E} = \frac{1405.2}{7.3} = 192.4932$$

P-value is near zero

b) Because the P-value of the F-test ≈ 0 is less than α = 0.05, we reject the null hypothesis that $\beta_1 = 0$ at the 0.05 level of significance. This is the same result obtained from the T_1 test. If the assumptions are valid, a useful linear relationship exists.

c) $\hat{\sigma}^2 = MS_E = 7.3$

11-23 a) 1) The parameter of interest is the regressor variable coefficient, β_1

2) $H_0 : \beta_1 = 0$

3) $H_1 : \beta_1 \neq 0$

4) α = 0.01

5) The test statistic is

$$f_0 = \frac{MS_R}{MS_E} = \frac{SS_R / 1}{SS_E / (n-2)}$$

6) Reject H_0 if $f_0 > f_{\alpha,1,12}$ where $f_{0.01,1,12} = 9.33$

7)

$$\begin{aligned} SS_R &= \hat{\beta}_1 S_{xy} = -2.3298017(-59.057143) \\ &= 137.59 \\ SS_E &= S_{yy} - SS_R \\ &= 159.71429 - 137.59143 \\ &= 22.123 \end{aligned}$$

$$f_0 = \frac{137.59}{22.123/12} = 74.63$$

8) Since 74.63 > 9.33 reject H_0 and conclude that compressive strength is significant in predicting intrinsic permeability of concrete at α = 0.01. We can therefore conclude that the model specifies a useful linear relationship between these two variables.

P-value $\cong$ 0.000002

b) $\hat{\sigma}^2 = MS_E = \frac{SS_E}{n-2} = \frac{22.123}{12} = 1.8436$ and $se(\hat{\beta}_1) = \sqrt{\frac{\hat{\sigma}^2}{S_{xx}}} = \sqrt{\frac{1.8436}{25.3486}} = 0.2696$

c) $se(\hat{\beta}_0) = \sqrt{\hat{\sigma}^2 \left[\frac{1}{n} + \frac{\bar{x}^2}{S_{xx}} \right]} = \sqrt{1.8436 \left[\frac{1}{14} + \frac{3.0714^2}{25.3486} \right]} = 0.9043$

11-24 a) 1) The parameter of interest is the regressor variable coefficient, β_1.

2) $H_0 : \beta_1 = 0$

3) $H_1 : \beta_1 \neq 0$

4) $\alpha = 0.01$

5) The test statistic is

$$f_0 = \frac{MS_R}{MS_E} = \frac{SS_R / 1}{SS_E / (n-2)}$$

6) Reject H_0 if $f_0 > f_{\alpha,1,18}$ where $f_{0.01,1,18} = 8.29$

7)

$$\begin{aligned} SS_R &= \hat{\beta}_1 S_{xy} = (0.0041612)(141.445) \\ &= 0.5886 \\ SS_E &= S_{yy} - SS_R \\ &= (8.86 - \tfrac{12.75^2}{20}) - 0.5886 \\ &= 0.143275 \end{aligned}$$

$$f_0 = \frac{0.5886}{0.143275/18} = 73.95$$

8) Since 73.95 > 8.29, reject H_0 and conclude the model specifies a useful relationship at $\alpha = 0.01$.

P-value $\cong 0.000001$

b) $$se(\hat{\beta}_1) = \sqrt{\frac{\hat{\sigma}^2}{S_{xx}}} = \sqrt{\frac{.00796}{33991.6}} = 4.8391 \times 10^{-4}$$

$$se(\hat{\beta}_0) = \sqrt{\hat{\sigma}^2 \left[\frac{1}{n} + \frac{\bar{x}}{S_{xx}} \right]} = \sqrt{.00796 \left[\frac{1}{20} + \frac{73.9^2}{33991.6} \right]} = 0.04091$$

11-25 a)

Regression Analysis: Rating Pts versus Yds per Att

```
The regression equation is
Rating Pts = 14.2 + 10.1 Yds per Att

Predictor      Coef  SE Coef     T      P
Constant     14.195    9.059  1.57  0.128
Yds per Att  10.092    1.288  7.84  0.000

S = 5.21874   R-Sq = 67.2%   R-Sq(adj) = 66.1%

Analysis of Variance
Source          DF      SS      MS      F      P
Regression       1  1672.5  1672.5  61.41  0.000
Residual Error  30   817.1    27.2
Total           31  2489.5
```

Refer to the ANOVA

$$H_0 : \beta_1 = 0$$
$$H_1 : \beta_1 \neq 0$$
$$\alpha = 0.05$$

Because the P-value = 0.000 < α = 0.05, reject H_0. If the assumptions are valid, we conclude that there is a useful linear relationship between these two variables.

b) $\hat{\sigma}^2 = 27.2$

$$se(\hat{\beta}_1) = \sqrt{\frac{\hat{\sigma}^2}{S_{xx}}} = \sqrt{\frac{27.2}{16.422}} = 1.287$$

$$se(\hat{\beta}_0) = \sqrt{\hat{\sigma}^2\left[\frac{1}{n} + \frac{\bar{x}^2}{S_{xx}}\right]} = \sqrt{27.2\left[\frac{1}{32} + \frac{7^2}{16.422}\right]} = 9.056$$

c) 1)The parameter of interest is the regressor variable coefficient β_1.
2) H_0:$\beta_1 = 10$
3) H_1:$\beta_1 \neq 10$
4) α = 0.05
5) The test statistic is $t_0 = \dfrac{\hat{\beta}_1 - \beta_{1,0}}{se(\hat{\beta}_1)}$
6) Reject H_0 if $t_0 < -t_{\alpha/2,n-2}$ where $-t_{0.025,30} = -2.042$ or $t_0 > t_{0.025,30} = 2.042$
7)

$$t_0 = \frac{10.092 - 10}{1.287} = 0.0714$$

8) Because 0.0714 < 2.042, fail to reject H_0. There is not enough evidence to conclude that the slope differs from 10 at α = 0.05.

11-26 Refer to ANOVA for the referenced exercise.
a) 1) The parameter of interest is the regressor variable coefficient, β_1.
2) H_0:$\beta_1 = 0$
3) H_1:$\beta_1 \neq 0$
4) α = 0.01, using t-test
5) The test statistic is $t_0 = \dfrac{\hat{\beta}_1}{se(\hat{\beta}_1)}$
6) Reject H_0 if $t_0 < -t_{\alpha/2,n-2}$ where $-t_{0.005,22} = -2.819$ or $t_0 > t_{0.005,22} = 2.819$
7)

$$t_0 = \frac{3.32437}{0.390276} = 8.518$$

8) Since 8.518 > 2.819 reject H_0 and conclude the model is useful α = 0.01.

b) 1) The parameter of interest is the slope, β_1
2) H_0:$\beta_1 = 0$
3) H_1:$\beta_1 \neq 0$
4) α = 0.01
5) The test statistic is $f_0 = \dfrac{MS_R}{MS_E} = \dfrac{SS_R / 1}{SS_E / (n-2)}$
6) Reject H_0 if $f_0 > f_{\alpha,1,22}$ where $f_{0.01,1,22} = 7.95$

7) Using the results from the referenced exercise

$$f_0 = \frac{636.15569/1}{192.89056/22} = 72.5563$$

8) Because 72.5563 > 7.95, reject H_0 and conclude the model is useful at a significance $\alpha = 0.01$.

The F-statistic is the square of the t-statistic. The F-test is a restricted to a two-sided test, whereas the t-test could be used for one-sided alternative hypotheses.

c) $$se(\hat{\beta}_1) = \sqrt{\frac{\hat{\sigma}^2}{S_{xx}}} = \sqrt{\frac{8.7675}{57.5631}} = .39027$$

$$se(\hat{\beta}_0) = \sqrt{\hat{\sigma}^2\left[\frac{1}{n} + \frac{\bar{x}}{S_{xx}}\right]} = \sqrt{8.7675\left[\frac{1}{24} + \frac{6.4049^2}{57.5631}\right]} = 2.5717$$

d) 1) The parameter of interest is the intercept, β_0.

2) $H_0 : \beta_0 = 0$

3) $H_1 : \beta_0 \neq 0$

4) $\alpha = 0.01$, using t-test

5) The test statistic is $t_0 = \dfrac{\hat{\beta}_0}{se(\hat{\beta}_0)}$

6) Reject H_0 if $t_0 < -t_{\alpha/2,n-2}$ where $-t_{0.005,22} = -2.819$ or $t_0 > t_{0.005,22} = 2.819$

7) Using the results from the referenced exercise

$$t_0 = \frac{13.3201}{2.5717} = 5.179$$

8) Because 5.179 > 2.819 reject H_0 and conclude the intercept is not zero at $\alpha = 0.01$.

11-27 Refer to the ANOVA for the referenced exercise

a) 1) The parameter of interest is the regressor variable coefficient, β_1.

2) $H_0 : \beta_1 = 0$

3) $H_1 : \beta_1 \neq 0$

4) $\alpha = 0.05$

5) The test statistic is $f_0 = \dfrac{MS_R}{MS_E} = \dfrac{SS_R/1}{SS_E/(n-2)}$

6) Reject H_0 if $f_0 > f_{\alpha,1,10}$ where $f_{0.05,1,10} = 10.04$

7) Using the results from Exercise 10-6

$$f_0 = \frac{280583.12/1}{37.746089/10} = 74334.4$$

8) Since 74334.4 > 10.04, reject H_0 and conclude the model is useful $\alpha = 0.05$. P-value < 0.000001

b) $se(\hat{\beta}_1) = 0.0337744$, $se(\hat{\beta}_0) = 1.66765$

c) 1) The parameter of interest is the regressor variable coefficient, β_1.

2) $H_0 : \beta_1 = 10$

3) $H_1 : \beta_1 \neq 10$

4) $\alpha = 0.05$, $\alpha/2 = 0.025$

5) The test statistic is $t_0 = \dfrac{\hat{\beta}_1 - \beta_{1,0}}{se(\hat{\beta}_1)}$

6) Reject H_0 if $t_0 < -t_{\alpha/2,n-2}$ where $-t_{0.025,10} = -2.228$ or $t_0 > t_{0.025,10} = 2.228$

7) Using the results from Exercise 10-6

$$t_0 = \frac{9.21 - 10}{0.0338} = -23.37$$

8) Since $-23.37 < -2.228$ reject H_0 and conclude the slope is not 10 at $\alpha = 0.05$. P-value ≈ 0.

d) H_0: $\beta_0 = 0$ H_1: $\beta_0 \neq 0$

$$t_0 = \frac{-6.3355 - 0}{1.66765} = -3.8$$

P-value < 0.005. Reject H_0 and conclude that the intercept should be included in the model.

11-28 Refer to the ANOVA for the referenced exercise.

$H_0: \beta_1 = 0, H_1: \beta_1 \neq 0$

a)

$$f_0 = \frac{MS_R}{MS_E} = \frac{385.18}{14.01} = 27.49$$

$$f_{0.05,1,19} = 4.38$$

$$f_0 > f_{0.05,1,19}$$

Reject the null hypothesis and conclude that the slope is not zero. The P-value ≈ 0.

b) From the computer output in the referenced exercise

$se(\beta_0) = 2.006$, $se(\beta_1) = 0.007671$

c)

$$H_0: \beta_1 = -0.05; H_1: \beta_1 < -0.05$$

$$t_0 = \frac{\hat{\beta}_1 - \hat{\beta}_{1,0}}{se(\hat{\beta}_1)} = \frac{-0.040216 - (-0.05)}{0.007671} = \frac{0.090216}{0.007671} = 11.76$$

$t_{0.01,19} = 2.539$, since t_0 is not less than $-t_{0.01,19} = -2.539$, do not reject H_0

$P \cong 1.0$

d)

$$H_0: \beta_0 = 0; H_1: \beta_0 \neq 0$$

$$t_0 = \frac{\hat{\beta}_0 - \hat{\beta}_{0,0}}{se(\hat{\beta}_0)} = \frac{39.156 - 0}{2.006} = 19.52$$

$t_{0.005,19} = 2.861$, since $|t_0| > t_{0.005,19}$ reject H_0

$P = 4.95E - 14 \cong 0$

11-29 Refer to the ANOVA for the referenced exercise.

a) $H_0: \beta_1 = 0$

$H_1: \beta_1 \neq 0$

$\alpha = 0.05$

$f_0 = 44.0279$

$f_{0.01,1,11} = 9.65$

$f_0 > f_{0.01,1,11}$

Therefore, reject H_0. P-value ≈ 0

b) $se(\hat{\beta}_1) = 0.0104524$

$se(\hat{\beta}_0) = 9.84346$

c) $H_0: \beta_0 = 0$

$H_1: \beta_0 \neq 0$

$\alpha = 0.05$

$t_0 = -1.67718$

$t_{.025,11} = 2.201$

$|t_0| \not< -t_{\alpha/2,11}$

Therefore, fail to reject H_0. P-value = 0.122

11-30 Refer to the ANOVA for the referenced exercise

a) $H_0: \beta_1 = 0$

$H_1: \beta_1 \neq 0$

$\alpha = 0.05$

$f_0 = 53.50$

$f_{0.05,1,18} = 4.414$

$f_0 > f_{\alpha,1,18}$

Therefore, reject H_0. P-value ≈ 0

b) $se(\hat{\beta}_1) = 0.0256613$

$se(\hat{\beta}_0) = 2.13526$

c) $H_0: \beta_0 = 0$

$H_1: \beta_0 \neq 0$

$\alpha = 0.05$

$t_0 = -5.079$

$t_{.025,18} = 2.101$

$|t_0| > t_{\alpha/2,18}$

Therefore, reject H_0. P-value ≈ 0

11-31 Refer to ANOVA for the referenced exercise

a) $H_0: \beta_1 = 0$

$H_1: \beta_1 \neq 0$

$\alpha = 0.05$

$f_0 = 155.2$

$f_{0.05,1,18} = 4.41$

$f_0 > f_{0.05,1,18}$

Therefore, reject H_0. P-value < 0.00001

b) $se(\hat{\beta}_1) = 45.3468$

$se(\hat{\beta}_0) = 2.96681$

c) $H_0 : \beta_1 = -30$

$H_1 : \beta_1 \neq -30$

$\alpha = 0.01$

$$t_0 = \frac{-36.9618 - (-30)}{2.96681} = -2.3466$$

$t_{.005,18} = 2.878$

$| t_0 | \not> -t_{\alpha/2,18}$

Therefore, fail to reject H_0. P-value = 0.0153(2) = 0.0306

d) $H_0 : \beta_0 = 0$

$H_1 : \beta_0 \neq 0$

$\alpha = 0.01$

$t_0 = 57.8957$

$t_{0.005,18} = 2.878$

$t_0 > t_{\alpha/2,18}$, therefore, reject H_0. P-value < 0.00001

e) $H_0 : \beta_0 = 2500$

$H_1 : \beta_0 > 2500$

$\alpha = 0.01$

$$t_0 = \frac{2625.39 - 2500}{45.3468} = 2.7651$$

$t_{.01,18} = 2.552$

$t_0 > t_{\alpha,18}$, therefore reject H_0. P-value = 0.0064

11-32 Refer to ANOVA for the referenced exercise

a) $H_0 : \beta_1 = 0$

$H_1 : \beta_1 \neq 0$

$\alpha = 0.05$

$f_0 = 92.224$

$f_{0.05,1,16} = 4.49$

$f_0 > f_{\alpha,1,16}$

Therefore, reject H_0.

b) P-value < 0.00001

c) $se(\hat{\beta}_1) = 2.14169$

$se(\hat{\beta}_0) = 1.93591$

d) $H_0 : \beta_0 = 0$

$H_1 : \beta_0 \neq 0$

$\alpha = 0.05$

$t_0 = 0.243$

$t_{0.025,16} = 2.12$

$t_0 \not> t_{\alpha/2,16}$

Therefore, do not reject H_0. There is not sufficient evidence to conclude that the intercept differs from zero. Based on this test result, the intercept could be removed from the model.

11-33 a) Refer to the ANOVA from the referenced exercise.

$$H_0 : \beta_1 = 0$$
$$H_1 : \beta_1 \neq 0$$
$$\alpha = 0.05$$

Because the P-value = 0.000 < α = 0.05, reject H_0. There is evidence of a linear relationship between these two variables.

b) $\hat{\sigma}^2 = 0.083$

The standard errors for the parameters can be obtained from the computer output or calculated as follows.

$$se(\hat{\beta}_1) = \sqrt{\frac{\hat{\sigma}^2}{S_{xx}}} = \sqrt{\frac{0.083}{420.91}} = 0.014$$

$$se(\hat{\beta}_0) = \sqrt{\hat{\sigma}^2\left[\frac{1}{n} + \frac{\bar{x}^2}{S_{xx}}\right]} = \sqrt{0.083\left[\frac{1}{11} + \frac{10.09^2}{420.91}\right]} = 0.1657$$

c)

1) The parameter of interest is the intercept β_0.

2) $H_0 : \beta_0 = 0$

3) $H_1 : \beta_0 \neq 0$

4) $\alpha = 0.05$

5) The test statistic is $t_0 = \dfrac{\beta_0}{se(\beta_0)}$

6) Reject H_0 if $t_0 < -t_{\alpha/2,n-2}$ where $-t_{0.025,9} = -2.262$ or $t_0 > t_{\alpha/2,n-2}$ where $t_{0.025,9} = 2.262$

7) Using the results from the referenced exercise

$$t_0 = \frac{0.6578}{0.1657} = 3.97$$

8) Because t_0 = 3.97 > 2.262 reject H_0 and conclude the intercept is not zero at α = 0.05.

11-34 a) Refer to the ANOVA for the referenced exercise.

$$H_0 : \beta_1 = 0$$
$$H_1 : \beta_1 \neq 0$$
$$\alpha = 0.05$$

Because the P-value = 0.000 < α = 0.05, reject H_0. There is evidence of a linear relationship between these two variables.

b) Yes

c) $\hat{\sigma}^2 = 1.118$

$$se(\hat{\beta}_1) = \sqrt{\frac{\hat{\sigma}^2}{S_{xx}}} = \sqrt{\frac{1.118}{588}} = 0.0436$$

$$se(\hat{\beta}_0) = \sqrt{\hat{\sigma}^2\left[\frac{1}{n} + \frac{\bar{x}^2}{S_{xx}}\right]} = \sqrt{1.118\left[\frac{1}{9} + \frac{65.67^2}{588}\right]} = 2.885$$

11-35 a) $H_0: \beta_1 = 0$

$H_1: \beta_1 \neq 0$

$\alpha = 0.01$

Because the P-value = 0.310 > α = 0.01, fail to reject H_0. There is not sufficient evidence of a linear relationship between these two variables.

```
The regression equation is
BMI = 13.8 + 0.256 Age

Predictor    Coef  SE Coef     T      P
Constant   13.820    9.141  1.51  0.174
Age        0.2558   0.2340  1.09  0.310
S = 5.53982   R-Sq = 14.6%   R-Sq(adj) = 2.4%

Analysis of Variance
Source          DF      SS     MS     F      P
Regression       1   36.68  36.68  1.20  0.310
Residual Error   7  214.83  30.69
Total            8  251.51
```

b) $\hat{\sigma}^2 = 30.69, se(\hat{\beta}_1) = 0.2340, se(\hat{\beta}_0) = 9.141$ from the computer output

c) $se(\hat{\beta}_0) = \sqrt{\hat{\sigma}^2\left[\frac{1}{n} + \frac{\bar{x}^2}{S_{xx}}\right]} = \sqrt{30.69\left[\frac{1}{9} + \frac{38.256^2}{560.342}\right]} = 9.141$

11-36 $t_0 = \frac{\hat{\beta}_1}{\sqrt{\hat{\sigma}^2 / S_{xx}}}$ After the transformation $\hat{\beta}_1^* = \frac{b}{a}\hat{\beta}_1$, $S_{xx}^* = a^2 S_{xx}$, $\bar{x}^* = a\bar{x}$, $\hat{\beta}_0^* = b\hat{\beta}_0$, and

$\hat{\sigma}^* = b\hat{\sigma}$. Therefore, $t_0^* = \frac{b\hat{\beta}_1 / a}{\sqrt{(b\hat{\sigma})^2 / a^2 S_{xx}}} = t_0$.

11-37 $d = \frac{|10-(12.5)|}{5.5\sqrt{31/16.422}} = 0.331$

Assume α = 0.05, from Chart VII*e* and interpolating between the curves for n = 30 and n = 40, $\beta \cong 0.55$

11-38 a) $\frac{\hat{\beta}}{\sqrt{\frac{\hat{\sigma}^2}{\sum x_i^2}}}$ has a t distribution with n – 1 degree of freedom.

b) From Exercise 11-15, $\hat{\beta} = 21.031461, \hat{\sigma} = 3.611768,$ and $\sum x_i^2 = 14.7073$.

The t-statistic in part (a) is 22.3314 and $H_0: \beta_0 = 0$ is rejected at usual α values.

Sections 11-5 and 11-6

11-39 $t_{\alpha/2,n-2} = t_{0.025,12} = 2.179$

a) 95% confidence interval on β_1.

$$\hat{\beta}_1 \pm t_{\alpha/2,n-2} se(\hat{\beta}_1)$$
$$-2.3298 \pm t_{.025,12}(0.2696)$$
$$-2.3298 \pm 2.179(0.2696)$$
$$-2.9173. \le \beta_1 \le -1.7423.$$

b) 95% confidence interval on β_0.

$$\hat{\beta}_0 \pm t_{.025,12} se(\hat{\beta}_0)$$
$$48.0130 \pm 2.179(0.5959)$$
$$46.7145 \le \beta_0 \le 49.3115.$$

c) 95% confidence interval on μ when $x_0 = 2.5$.

$$\hat{\mu}_{Y|x_0} = 48.0130 - 2.3298(2.5) = 42.1885$$
$$\hat{\mu}_{Y|x_0} \pm t_{.025,12}\sqrt{\hat{\sigma}^2\left(\frac{1}{n} + \frac{(x_0-\bar{x})^2}{S_{xx}}\right)}$$
$$42.1885 \pm (2.179)\sqrt{1.844\left(\frac{1}{14} + \frac{(2.5-3.0714)^2}{25.3486}\right)}$$
$$42.1885 \pm 2.179(0.3943)$$
$$41.3293 \le \hat{\mu}_{Y|x_0} \le 43.0477$$

d) 99% on prediction interval when $x_0 = 2.5$.

$$\hat{y}_0 \pm t_{0.005,12}\sqrt{\hat{\sigma}^2(1+\tfrac{1}{n}+\tfrac{(x_0-\bar{x})^2}{S_{xx}})}$$
$$42.1885 \pm 3.055\sqrt{1.844(1+\tfrac{1}{14}+\tfrac{(2.5-3.0714)^2}{25.348571})}$$
$$42.1885 \pm 3.055(1.4056)$$
$$37.8944 \leq y_0 \leq 46.4826$$

It is wider because it depends on both the errors associated with the fitted model and the future observation.

11-40 $t_{\alpha/2,n-2} = t_{0.025,18} = 2.101$

a) $\hat{\beta}_1 \pm (t_{0.025,18})se(\hat{\beta}_1)$
$$0.0041612 \pm (2.101)(0.000484)$$
$$0.0031443 \leq \beta_1 \leq 0.0051781$$

b) $\hat{\beta}_0 \pm (t_{0.025,18})se(\hat{\beta}_0)$
$$0.3299892 \pm (2.101)(0.04095)$$
$$0.24395 \leq \beta_0 \leq 0.41603$$

c) 99% confidence interval on μ when $x_0 = 85^\circ$ F

$t_{\alpha/2,n-2} = t_{0.005,18} = 2.878$
$$\hat{\mu}_{Y|x_0} = 0.683689$$
$$\hat{\mu}_{Y|x_0} \pm t_{.005,18}\sqrt{\hat{\sigma}^2(\tfrac{1}{n}+\tfrac{(x_0-\bar{x})^2}{S_{xx}})}$$
$$0.683689 \pm (2.878)\sqrt{0.00796(\tfrac{1}{20}+\tfrac{(85-73.9)^2}{33991.6})}$$
$$0.683689 \pm 0.0594607$$
$$0.6242283 \leq \hat{\mu}_{Y|x_0} \leq 0.7431497$$

d) 99% prediction interval when $x_0 = 90^\circ F$.
$$\hat{y}_0 = 0.7044949$$
$$\hat{y}_0 \pm t_{.005,18}\sqrt{\hat{\sigma}^2(1+\tfrac{1}{n}+\tfrac{(x_0-\bar{x})^2}{S_{xx}})}$$
$$0.7044949 \pm 2.878\sqrt{0.00796(1+\tfrac{1}{20}+\tfrac{(90-73.9)^2}{33991.6})}$$
$$0.7044949 \pm 0.263567$$
$$0.420122 \leq y_0 \leq 0.947256$$

11-41 $t_{\alpha/2,n-2} = t_{0.025,30} = 2.042$

a) 99% confidence interval on β_1
$$\hat{\beta}_1 \pm t_{\alpha/2,n-2}se(\hat{\beta}_1)$$
$$10.092 \pm t_{0.005,30}(1.287)$$
$$10.092 \pm 2.750(1.287)$$
$$6.553 \leq \beta_1 \leq 13.631$$

b) 99% confidence interval on β_0

$$\hat{\beta}_0 \pm t_{\alpha/2,n-2} se(\hat{\beta}_0)$$
$$14.195 \pm t_{0.005,30}(9.056)$$
$$14.195 \pm 2.75(9.056)$$
$$-10.709 \le \hat{\beta}_0 \le 39.099$$

c) 99% confidence interval for the mean rating when the average yards per attempt is 8.0

$$\hat{\mu} = 14.195 + 10.092(8.0) = 94.931$$

$$\hat{\mu} \pm t_{0.005,30}\sqrt{\hat{\sigma}^2\left(\frac{1}{n} + \frac{(x_0 - \bar{x})^2}{S_{xx}}\right)}$$

$$94.931 \pm 2.75\sqrt{27.2\left(\frac{1}{32} + \frac{(8-7)^2}{16.422}\right)}$$

$$90.577 \le \mu \le 99.285$$

d) 99% prediction interval on $x_0 = 8.0$

$$\hat{y} \pm t_{0.005,30}\sqrt{\hat{\sigma}^2\left(1 + \frac{1}{n} + \frac{(x_0 - \bar{x})^2}{S_{xx}}\right)}$$

$$94.931 \pm 2.75\sqrt{27.2\left(1 + \frac{1}{32} + \frac{(8-7)^2}{16.422}\right)}$$

$$79.943 \le \mu \le 109.919$$

11-42 **Regression Analysis: Price versus Taxes**

```
The regression equation is
Price = 13.3 + 3.32 Taxes


Predictor    Coef  SE Coef     T      P
Constant   13.320    2.572  5.18  0.000
Taxes      3.3244   0.3903  8.52  0.000


S = 2.96104   R-Sq = 76.7%   R-Sq(adj) = 75.7%


Analysis of Variance

Source          DF      SS      MS      F      P
Regression       1  636.16  636.16  72.56  0.000
Residual Error  22  192.89    8.77
Total           23  829.05
```

a) 3.32437 – 2.074(0.3903) = 2.515 ≤ β_1 ≤ 3.32437 + 2.074(0.3903) = 4.134

b) 13.320 – 2.074(0.3903) = 7.985 ≤ β_0 ≤ 13.320 + 2.074(0.39028) = 18.655

c) $39.915 \pm (2.074)\sqrt{8.76775(\frac{1}{24} + \frac{(8.0-6.40492)^2}{57.563139})}$

$$39.915 \pm 1.800$$
$$38.115 \le \hat{\mu}_{Y|x_0} \le 41.715$$

d) $$39.915 \pm (2.074)\sqrt{8.76775\left(1+\frac{1}{24}+\frac{(8.0-6.40492)^2}{57.563139}\right)}$$
$$39.915 \pm 6.399$$
$$33.516 \le y_0 \le 46.314$$

11-43 **Regression Analysis: Usage versus Temperature**

```
The regression equation is
Usage = 130 + 7.59 Temperature

Predictor       Coef  SE Coef       T      P
Constant     129.974    0.707  183.80  0.000
Temperature  7.59262  0.05798  130.95  0.000

S = 1.83431   R-Sq = 99.9%   R-Sq(adj) = 99.9%

Analysis of Variance

Source          DF     SS     MS         F      P
Regression       1  57701  57701  17148.85  0.000
Residual Error  10     34      3
Total           11  57734
```

a) $7.59262 - 2.228(0.05798) = 7.463 \le \beta_1 \le 7.59262 + 2.228(0.05798) = 7.721$

b) $129.974 - 2.228(0.707) = 128.399 \le \beta_0 \le 129.974 + 2.228(0.707) = 131.549$

c) $$228.67 \pm (2.228)\sqrt{3\left(\frac{1}{12}+\frac{(13-8.08)^2}{1000.917}\right)}$$
$$228.67 \pm 1.26536$$
$$227.4046 \le \hat{\mu}_{Y|x_0} \le 229.9354$$

d) $$228.67 \pm (2.228)\sqrt{3\left(1+\frac{1}{12}+\frac{(13-8.08)^2}{1000.917}\right)}$$
$$228.67 \pm 4.061644$$
$$224.6084 \le y_0 \le 232.73164$$

It is wider because the prediction interval includes errors for both the fitted model and for a future observation.

11-44 Refer to the ANOVA for the referenced exercise.

(a) $t_{0.025,\,19} = 2.093$

$$34.96 \le \beta_0 \le 43.36; -0.0563 \le \beta_1 \le -0.0241$$

(b) Descriptive Statistics: x = displacement

Variable	n	Mean	Sum	Sum of Squares
x	21	238.9	5017.0	1436737.0

$$\hat{y} = 33.15 \text{ when } x = 150$$
$$33.15 \pm 2.093\sqrt{14.01\left[\frac{1}{21} + \frac{(150-238.9)^2}{1,436,737.0}\right]}$$
$$33.15 \pm 1.8056$$
$$31.34 \le \mu_{Y|x=150} \le 34.96$$

(c) $\hat{y} = 33.15 \text{ when } x = 150$

$$33.15 \pm 2.093\sqrt{14.01\left[1 + \frac{1}{21} + \frac{(150-238.9)^2}{1,436,737.0}\right]}$$
$$33.15 \pm 8.0394$$
$$25.11 \le Y_0 \le 41.19$$

11-45 a) $0.03689 \le \beta_1 \le 0.10183$
b) $-47.0877 \le \beta_0 \le 14.0691$

c) $43.834 \pm (3.106)\sqrt{7.324951(\frac{1}{13} + \frac{(870-939)^2}{6704597})}$
43.834 ± 3.233
$40.601 \le \mu_{y|x_0} \le 47.067$

d) $43.834 \pm (3.106)\sqrt{7.324951(1 + \frac{1}{13} + \frac{(870-939)^2}{6704597})}$
43.834 ± 9.007
$34.827 \le y_0 \le 52.841$

11-46 a) $0.11756 \le \beta_1 \le 0.22541$
b) $-14.3002 \le \beta_0 \le -5.32598$
c) $5.554 \pm (2.101)\sqrt{1.982231(\frac{1}{20} + \frac{(90-82.3)^2}{30102111})}$
5.554 ± 0.781
$4.773 \le \mu_{y|x_0} \le 6.335$

d) $5.554 \pm (2.101)\sqrt{1.982231(1 + \frac{1}{20} + \frac{(85-82.3)^2}{30102111})}$
5.554 ± 3.059
$2.495 \le y_0 \le 8.613$

11-47 a) $201.552 \le \beta_1 \le 266.590$
b) $-4.67015 \le \beta_0 \le -2.34696$
c) $107.763 \pm (2.365)\sqrt{398.2804(\frac{1}{9} + \frac{(36-24.5)^2}{16514214})}$
128.814 ± 20.638
$108.176 \le \mu_{y|x_0} \le 149.452$

11-48 a) $14.3107 \le \beta_1 \le 26.8239$

b) $-5.18501 \le \beta_0 \le 6.12594$

c) $17.336 \pm (2.921)\sqrt{13.8092(\frac{1}{18} + \frac{(0.82-0.806111)^2}{3.01062})}$

17.336 ± 2.560

$14.776 \le \mu_{y|x_0} \le 19.896$

d) $17.336 \pm (2.921)\sqrt{13.8092(1 + \frac{1}{18} + \frac{(0.82-0.806111)^2}{3.01062})}$

17.336 ± 11.152

$6.184 \le y_0 \le 28.488$

11-49 a) $-313.0885 \le \beta_1 \le -196.0115$

b) $17195.7176 \le \beta_0 \le 18984.6824$

c) $12999.2 \pm 2.878\sqrt{460992(\frac{1}{20} + \frac{(20-13.3375)^2}{1114.6618})}$

12999.2 ± 585.64

$12413.56 \le \mu_{y|x_0} \le 13584.84$

d) $12999.2 \pm 2.878\sqrt{460992(1 + \frac{1}{20} + \frac{(20-13.3375)^2}{1114.6618})}$

12999.2 ± 2039.93

$10959.27 \le y_0 \le 15039.14$

11-50 $t_{\alpha/2,n-2} = t_{0.005,5} = 4.032$

a) 99% confidence interval on $\hat{\beta}_1$

$$\hat{\beta}_1 \pm t_{\alpha/2,n-2} se(\hat{\beta}_1)$$
$$-0.034 \pm t_{0.001,5}(0.026)$$
$$-0.034 \pm 4.032(0.026)$$
$$-0.1388 \le \hat{\beta}_1 \le 0.0708$$

b) 99% confidence interval on β_0

$$\hat{\beta}_0 \pm t_{\alpha/2,n-2} se(\hat{\beta}_0)$$
$$55.63 \pm t_{0.005,5}(32.11)$$
$$55.63 \pm 4.032(32.11)$$
$$-73.86 \le \hat{\beta}_0 \le 185.12$$

c) 99% confidence interval for the mean length when x = 1500:

$$\hat{\mu} = 55.63 - 0.034(1500) = 4.63$$

$$\hat{\mu} \pm t_{0.005,5}\sqrt{\hat{\sigma}^2\left(\frac{1}{n}+\frac{(x_0-\bar{x})^2}{S_{xx}}\right)}$$

$$4.63 \pm 4.032\sqrt{77.33\left(\frac{1}{7}+\frac{(1500-1242.86)^2}{117142.8}\right)}$$

$$4.63 \pm 4.032(7.396)$$

$$-25.19 \le \mu \le 34.45$$

d) 99% prediction interval when $x_0 = 1500$

$$\hat{y} \pm t_{0.005,5}\sqrt{\hat{\sigma}^2\left(1+\frac{1}{n}+\frac{(x_0-\bar{x})^2}{S_{xx}}\right)}$$

$$4.63 \pm 4.032\sqrt{77.33\left(1+\frac{1}{7}+\frac{(1500-1242.86)^2}{117142.8}\right)}$$

$$4.63 \pm 4.032(11.49)$$

$$-41.7 \le y_0 \le 50.96$$

It's wider because it depends on both the error associated with the fitted model as well as that of the future observation.

11-51 Refer to the computer output in the referenced exercise.

$$t_{\alpha/2,n-2} = t_{0.005,9} = 3.250$$

a) 99% confidence interval for $\hat{\beta}_1$

$$\hat{\beta}_1 \pm t_{\alpha/2,n-2}se(\hat{\beta}_1)$$

$$0.178 \pm t_{0.005,9}(0.014)$$

$$0.178 \pm 3.250(0.014)$$

$$0.1325 \le \hat{\beta}_1 \le 0.2235$$

b) 99% confidence interval on β_0

$$\hat{\beta}_0 \pm t_{\alpha/2,n-2}se(\hat{\beta}_0)$$

$$0.6578 \pm t_{0.005,9}(0.1657)$$

$$0.6578 \pm 3.250(0.1657)$$

$$0.119 \le \hat{\beta}_0 \le 1.196$$

c) 95% confidence interval on μ when $x_0 = 10$

$$\hat{\mu}_{y|x_0} = 0.658 + 0.178(10) = 2.438$$

$$\hat{\mu}_{y|x_0} \pm t_{0.025,9}\sqrt{\hat{\sigma}^2\left(\frac{1}{n}+\frac{(x_0-\bar{x})^2}{S_{xx}}\right)}$$

$$2.438 \pm 2.262\sqrt{0.083\left(\frac{1}{11}+\frac{(10-10.09)^2}{420.91}\right)}$$

$$2.241 \le \mu_{y|x_0} \le 2.635$$

Section 11-7

11-52 $$R^2 = \hat{\beta}_1^2 \frac{S_{XX}}{S_{YY}} = (-2.330)^2 \frac{25.35}{159.71} = 0.8617$$

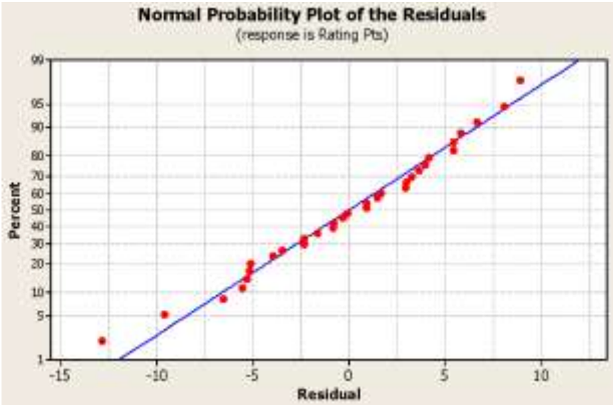
The model accounts for 86.17% of the variability in the data.

11-53 Refer to the Minitab output in the referenced exercise.

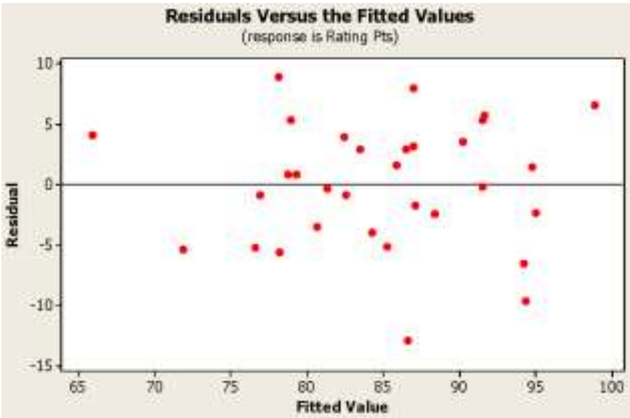
a) $R^2 = 0.672$

The model accounts for 67.2% of the variability in the data.

b) There is no major departure from the normality assumption in the following graph.



c) The assumption of constant variance appears reasonable.

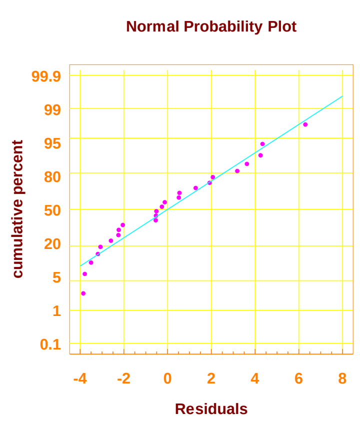


11-54 Use the results from the referenced exercise to answer the following questions.

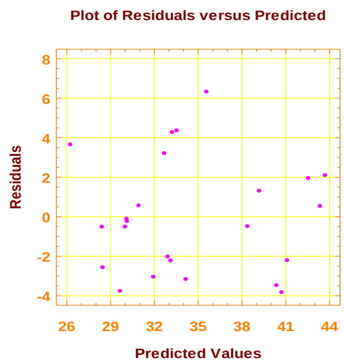
a)

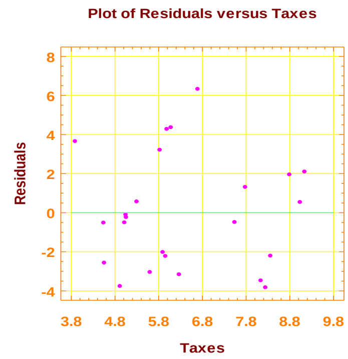
SalePrice	Taxes	Predicted	Residuals
25.9	4.9176	29.6681073	-3.76810726
29.5	5.0208	30.0111824	-0.51118237
27.9	4.5429	28.4224654	-0.52246536
25.9	4.5573	28.4703363	-2.57033630
29.9	5.0597	30.1405004	-0.24050041
29.9	3.8910	26.2553078	3.64469225
30.9	5.8980	32.9273208	-2.02732082
28.9	5.6039	31.9496232	-3.04962324
35.9	5.8282	32.6952797	3.20472030
31.5	5.3003	30.9403441	0.55965587
31.0	6.2712	34.1679762	-3.16797616
30.9	5.9592	33.1307723	-2.23077234
30.0	5.0500	30.1082540	-0.10825401
36.9	8.2464	40.7342742	-3.83427422
41.9	6.6969	35.5831610	6.31683901
40.5	7.7841	39.1974174	1.30258260
43.9	9.0384	43.3671762	0.53282376
37.5	5.9894	33.2311683	4.26883165
37.9	7.5422	38.3932520	-0.49325200
44.5	8.7951	42.5583567	1.94164328
37.9	6.0831	33.5426619	4.35733807
38.9	8.3607	41.1142499	-2.21424985
36.9	8.1400	40.3805611	-3.48056112
45.8	9.1416	43.7102513	2.08974865

b) Assumption of normality does not seem to be violated since the data appear to fall along a straight line.



c) There are no serious departures from the assumption of constant variance. This is evident by the random pattern of the residuals.



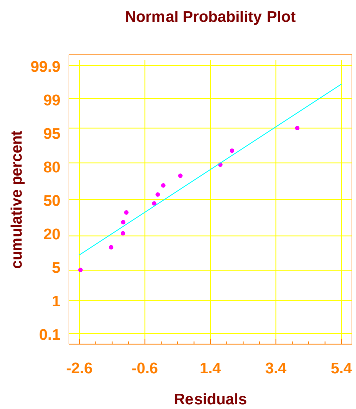


d) $R^2 \equiv 76.73\%$;

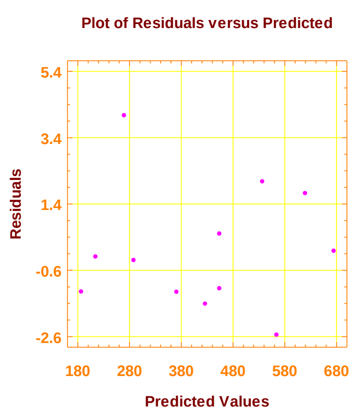
11-55 Use the results of the referenced exercise to answer the following questions

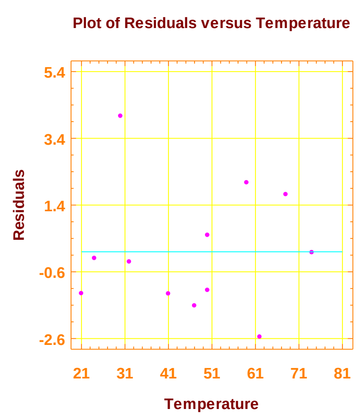
a) $R^2 = 99.986\%$; The proportion of variability explained by the model.

b) Yes, normality seems to be satisfied because the data appear to fall along the straight line.



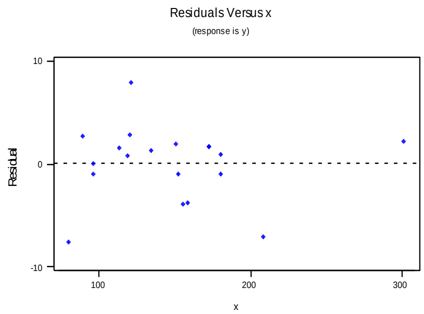
c) There might be lower variance at the middle settings of *x*. However, this data does not indicate a serious departure from the assumptions.

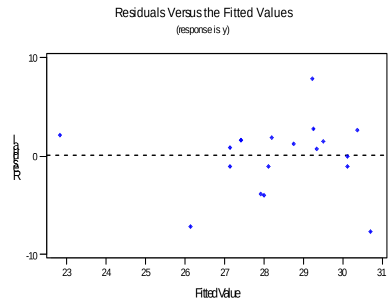




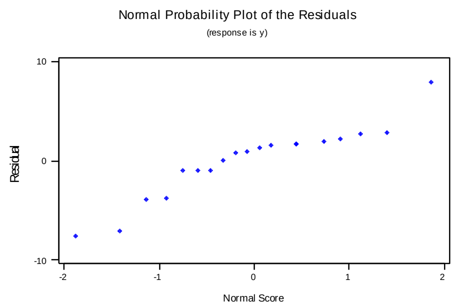
11-56 a) $R^2 = 20.1121\%$

b) These plots might indicate the presence of outliers, but no real problem with assumptions.



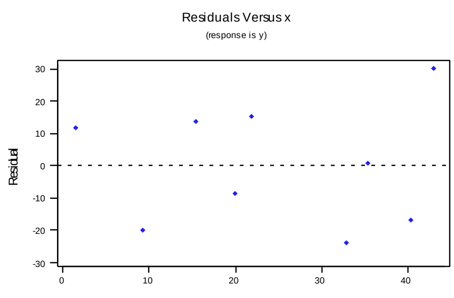


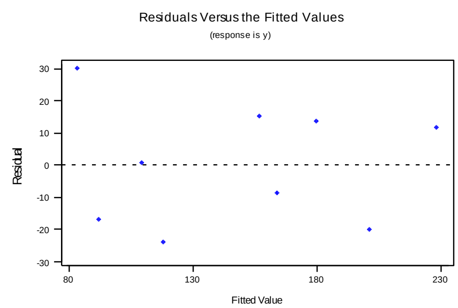
c) The normality assumption appears marginal.



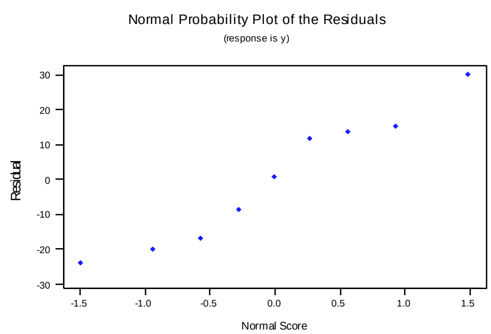
11-57 a) $R^2 = 0.879397$

b) No departures from constant variance are noted.



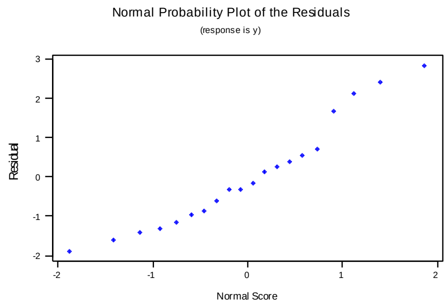


c) Normality assumption appears reasonable.

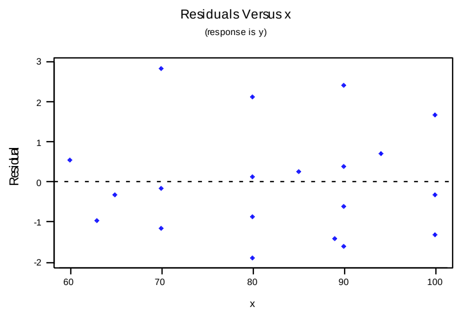


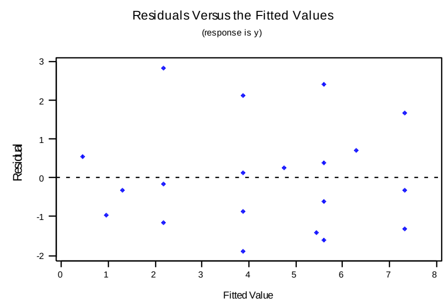
11-58 a) $R^2 = 71.27\%$

b) No major departure from normality assumptions.



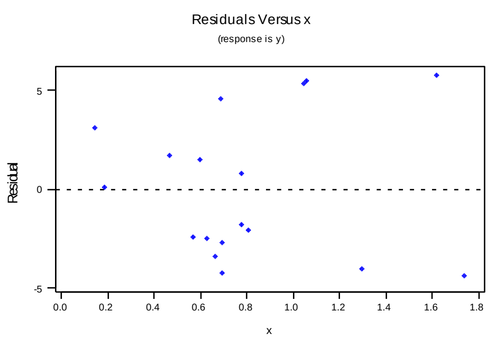
c) Assumption of constant variance appears reasonable.

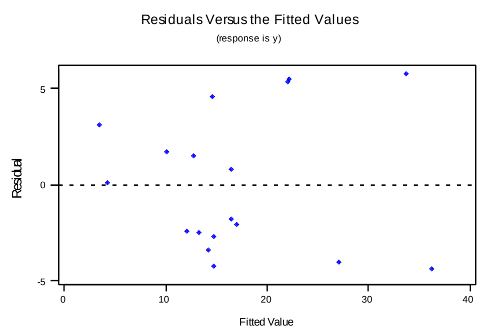




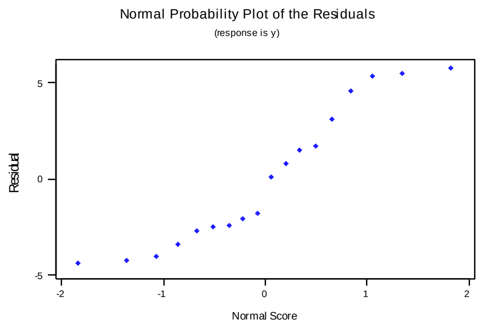
11-59 a) $R^2 = 85.22\%$

b) Assumptions appear reasonable, but there is a suggestion that variability increases slightly with $\hat{y}$.





c) Normality assumption may be questionable. There is some "bending" away from a line in the tails of the normal probability plot.



11-60 a)

```
The regression equation is
Compressive Strength = - 2150 + 185 Density

Predictor      Coef  SE Coef      T      P
Constant    -2149.6    332.5  -6.46  0.000
Density      184.55    11.79  15.66  0.000

S = 339.219   R-Sq = 86.0%   R-Sq(adj) = 85.6%

Analysis of Variance

Source          DF        SS        MS       F      P
Regression       1  28209679  28209679  245.15  0.000
Residual Error  40   4602769    115069
Total           41  32812448
```

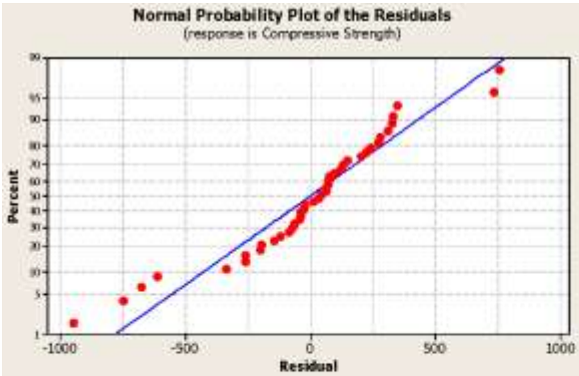
b) Because the P-value = 0.000 < α = 0.01, the model is significant.

c) $\hat{\sigma}^2 = 115069$

d) $R^2 = \dfrac{SS_R}{SS_T} = 1 - \dfrac{SS_E}{SS_T} = \dfrac{28209679}{32812448} = 0.8597 = 85.97\%$

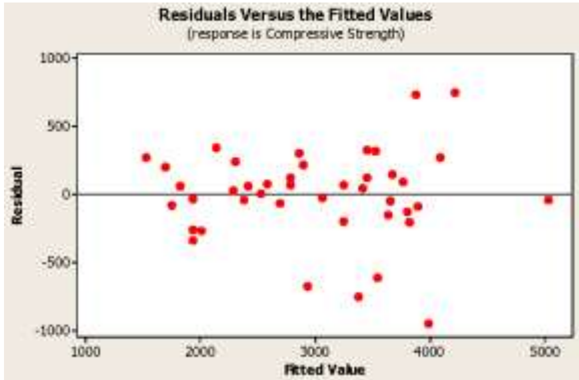
The model accounts for 85.97% of the variability in the data.

e)



No major departure from the normality assumption.

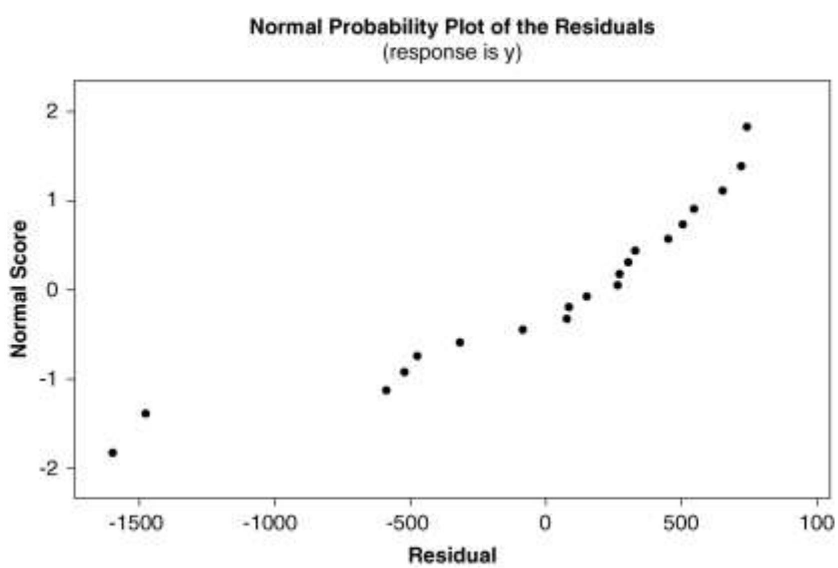
f)



Assumption of constant variance appears reasonable.

11-61 a) $R^2 = 0.896947$ 89% of the variability is explained by the model.

b) Yes, the two points with residuals much larger in magnitude than the others seem unusual.



c) $R^2_{\text{new model}} = 0.8799$

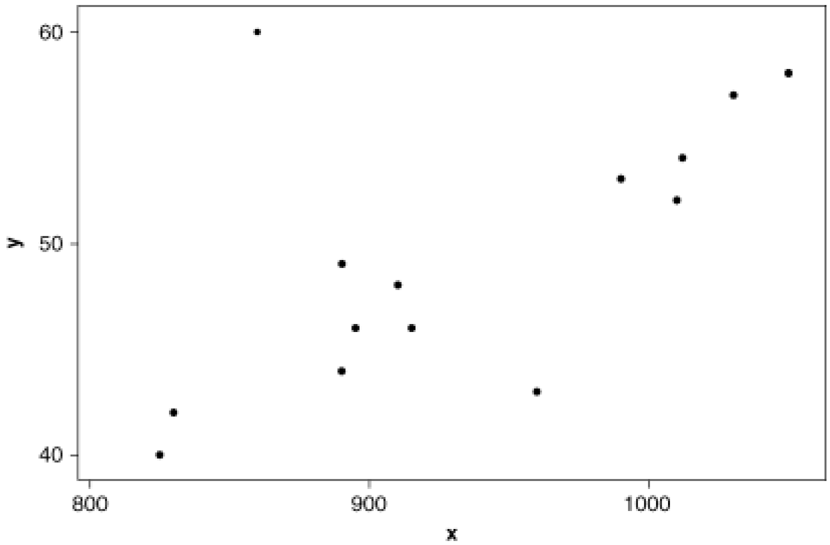
Smaller, because the older model is better able to account for the variability in the data with these two outlying data points removed.

d) $\hat{\sigma}^2_{\text{old model}} = 460992$

$\hat{\sigma}^2_{\text{new model}} = 188474$

Yes, reduced more than 50%, because the two removed points accounted for a large amount of the error.

11-62 a)



$\hat{y} = 0.55 + 0.05237x$

b) $H_0 : \beta_1 = 0 \quad H_1 : \beta_1 \neq 0 \quad \alpha = 0.05$

$f_0 = 7.41$
$f_{.05,1,12} = 4.75$
$f_0 > f_{\alpha,1,12}$

Reject H_0.

c) $\hat{\sigma}^2 = 26.97$

d) $\hat{\sigma}_{orig}{}^2 = 7.502$

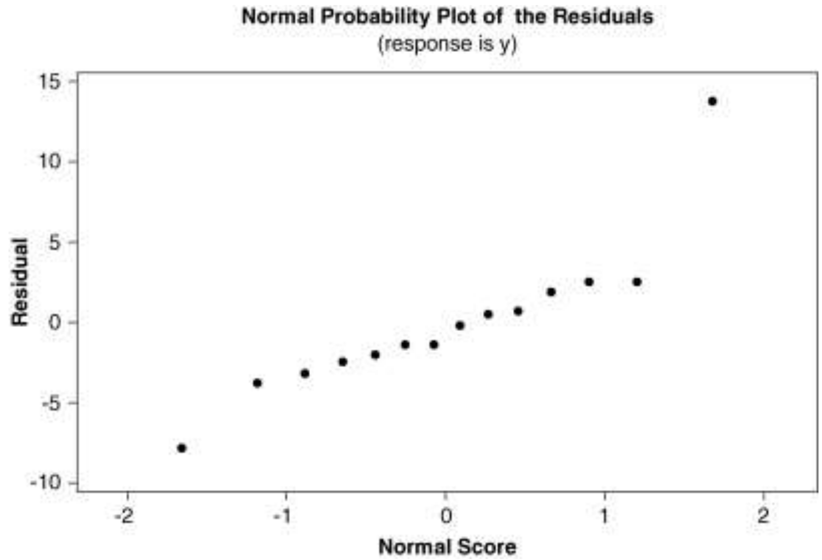
The new estimate is larger because the new point added additional variance that was not accounted for by the model.

e) $\hat{y} = 0.55 + 0.05237(860) = 45.5882$

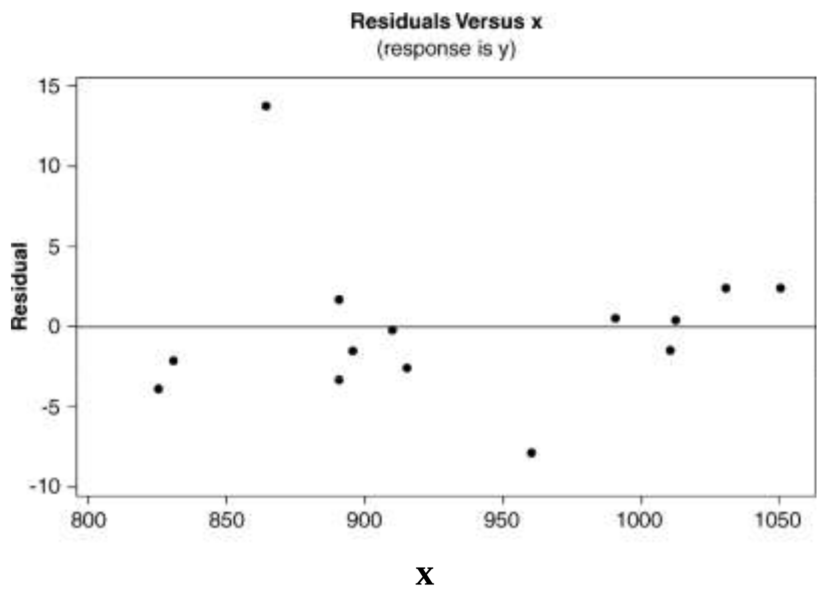
$e = y - \hat{y} = 60 - 45.5882 = 14.4118$

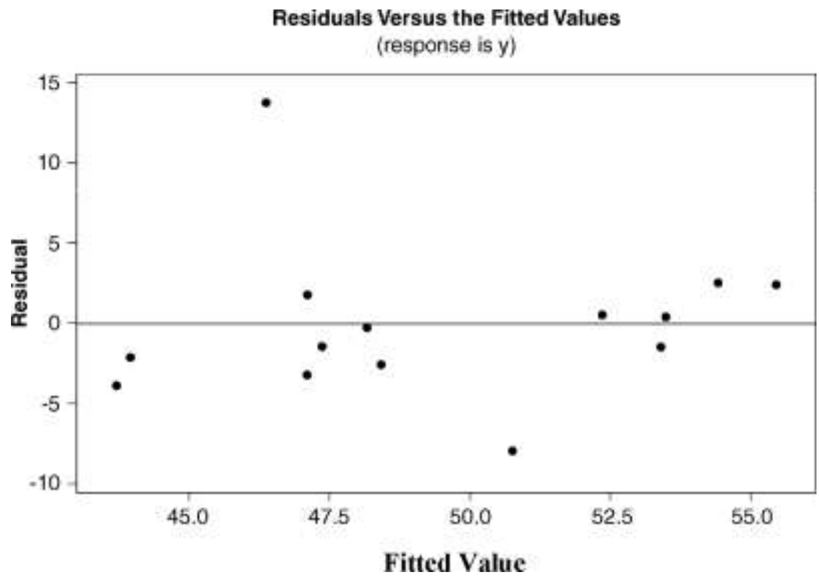
Yes, e_{14} is especially large compared to the other residuals.

f) The one added point is an outlier and the normality assumption is not as valid with the point included.



g) Constant variance assumption appears valid except for the added point.





11-63 Yes, when the residuals are standardized the unusual residuals are easier to identify.

1.11907 –0.75653 –0.13113 0.68314 –2.49705 –2.26424 0.51810

0.48210 0.11676 0.40780 0.22274 –0.93513 0.88167 0.76461

–0.49995 0.99241 0.12989 0.39831 1.15898 –0.82134

11-64 For two random variables X_1 and X_2,

$$V(X_1 + X_2) = V(X_1) + V(X_2) + 2Cov(X_1, X_2)$$

Then,

$$\begin{aligned} V(Y_i - \hat{Y}_i) &= V(Y_i) + V(\hat{Y}_i) - 2Cov(Y_i, \hat{Y}_i) \\ &= \sigma^2 + V(\hat{\beta}_0 + \hat{\beta}_1 x_i) - 2\sigma^2\left[\tfrac{1}{n} + \tfrac{(x_i-\bar{x})^2}{S_{xx}}\right] \\ &= \sigma^2 + \sigma^2\left[\tfrac{1}{n} + \tfrac{(x_i-\bar{x})^2}{S_{xx}}\right] - 2\sigma^2\left[\tfrac{1}{n} + \tfrac{(x_i-\bar{x})^2}{S_{xx}}\right] \\ &= \sigma^2\left[1 - \left(\tfrac{1}{n} + \tfrac{(x_i-\bar{x})^2}{S_{xx}}\right)\right] \end{aligned}$$

a) Because e_i is divided by an estimate of its standard error (when σ^2 is estimated by $\hat{\sigma}^2$), r_i has approximately unit variance.

b) No, the term in brackets in the denominator is necessary.

c) If x_i is near $\bar{x}$ and n is reasonably large, r_i is approximately equal to the standardized residual.

d) If x_i is far from $\bar{x}$, the standard error of e_i is small. Consequently, extreme points are better fit by least squares regression than points near the middle range of x. Because the studentized residual at any point has variance of approximately one, the studentized residuals can be used to compare the fit of points to the regression line over the range of x.

11-65 Using $R^2 = 1 - \frac{SS_E}{S_{yy}}$, $F_0 = \dfrac{(n-2)(1-\frac{SS_E}{S_{yy}})}{\frac{SS_E}{S_{yy}}} = \dfrac{S_{yy} - SS_E}{\frac{SS_E}{n-2}} = \dfrac{S_{yy} - SS_E}{\hat{\sigma}^2}$

Also,

$$SS_E = \sum (y_i - \hat{\beta}_0 - \hat{\beta}_1 x_i)^2$$
$$= \sum (y_i - \bar{y} - \hat{\beta}_1 (x_i - \bar{x}))^2$$
$$= \sum (y_i - \bar{y}) + \hat{\beta}_1^2 \sum (x_i - \bar{x})^2 - 2\hat{\beta}_1 \sum (y_i - \bar{y})(x_i - \bar{x})$$
$$= \sum (y_i - \bar{y})^2 - \hat{\beta}_1^2 \sum (x_i - \bar{x})^2$$
$$S_{yy} - SS_E = \hat{\beta}_1^2 \sum (x_i - \bar{x})^2$$

Therefore, $F_0 = \dfrac{\hat{\beta}_1^2}{\hat{\sigma}^2 / S_{xx}} = t_0^2$

Because the square of a t random variable with $n-2$ degrees of freedom is an F random variable with 1 and $n-2$ degrees of freedom, the usual t-test that compares $|t_0|$ to $t_{\alpha/2,n-2}$ is equivalent to comparing $f_0 = t_0^2$ to $f_{\alpha,1,n-2} = t^2_{\alpha/2,n-2}$.

a) $f_0 = \dfrac{0.9(23)}{1-0.9} = 207$. Reject $H_0: \beta_1 = 0$.

b) Because $f_{0.01,1,23} = 7.88$, H_0 is rejected if $\dfrac{23R^2}{1-R^2} > 7.88$.

That is, H_0 is rejected if

$$23R^2 > 7.88(1-R^2)$$
$$27.28R^2 > 7.88$$
$$R^2 > 0.289$$

Section 11-8

11-66 a) $H_0: \rho = 0$

$H_1: \rho \neq 0$ $\alpha = 0.05$

$$t_0 = \frac{0.8\sqrt{20-2}}{\sqrt{1-0.64}} = 5.657$$
$$t_{0.025,18} = 2.101$$
$$|t_0| > t_{0.025,18}$$

Reject H_0. P-value = (<0.0005)(2) = <0.001

b) $H_0: \rho = 0.5$

$H_1: \rho \neq 0.5$ $\alpha = 0.05$

$$z_0 = (\text{arctanh}\,(0.8) - \text{arctanh}\,(0.5))(17)^{1/2} = 2.265$$
$$z_{.025} = 1.96$$
$$|z_0| > z_{\alpha/2}$$

Reject H_0. P-value = (0.012)(2) = 0.024.

c) $\tanh(\text{arctanh}\ 0.8 - \frac{z_{.025}}{\sqrt{17}}) \le \rho \le \tanh(\text{arctanh}\ 0.8 + \frac{z_{.025}}{\sqrt{17}})$

where $z_{.025} = 1.96$. $0.5534 \le \rho \le 0.9177$.

Because ρ = 0 and ρ = 0.5 are not in the interval, so reject H_0.

11-67 a) $H_0: \rho = 0$

$H_1: \rho > 0$ $\alpha = 0.05$

$$t_0 = \frac{0.75\sqrt{20-2}}{\sqrt{1-0.75^2}} = 4.81$$

$$t_{0.05,18} = 1.734$$

$$t_0 > t_{0.05,18}$$

Reject H_0. P-value < 0.0005

b) $H_0: \rho = 0.1$

$H_1: \rho > 0.1$ $\alpha = 0.05$

$$z_0 = (\text{arctanh}\ (0.75) - \text{arctanh}\ (0.1))(17)^{1/2} = 3.598$$

$$z_{.05} = 1.65$$

$$z_0 > z_\alpha$$

Reject H_0. P-value < 0.0002

c) $\rho \geq \tanh(\text{arctanh}\ 0.75 - \frac{z_{0.05}}{\sqrt{17}})$ where $z_{.05} = 1.65$

$$\rho \geq 0.517$$

Because ρ = 0 and ρ = 0.1 are not in the interval, reject the null hypotheses from parts (a) and (b).

11-68 $n = 30$ $r = 0.83$

a) $\text{H}_0: \rho = 0$

$\text{H}_1: \rho \neq 0$ $\alpha = 0.05$

$$t_0 = \frac{r\sqrt{n-2}}{\sqrt{1-r^2}} = \frac{0.83\sqrt{28}}{\sqrt{1-(0.83)^2}} = 7.874$$

$$t_{.025,28} = 2.048$$

$$t_0 > t_{\alpha/2,28}$$

Reject H_0. P-value = 0.

b) $\tanh(\text{arctanh}\ 0.83 - \frac{z_{.025}}{\sqrt{27}}) \leq \rho \leq \tanh(\text{arctanh}\ 0.83 + \frac{z_{.025}}{\sqrt{27}})$

where $z_{.025} = 1.96$. $0.453 \leq \rho \leq 1.207$.

a) $\text{H}_0: \rho = 0.8$

$\text{H}_1: \rho \neq 0.8$ $\alpha = 0.05$

$$z_0 = (\text{arctanh}\ 0.83 - \text{arctanh}\ 0.8)(27)^{1/2} = 0.4652$$

$$z_{.025} = 1.96$$

$$z_0 \not> z_{\alpha/2}$$

Do not reject H_0. P-value = (0.321)(2) = 0.642.

11-69 n = 50 r = 0.62

a) $H_0: \rho = 0$

$H_1: \rho \neq 0$ $\alpha = 0.01$

$t_0 = \frac{r\sqrt{n-2}}{\sqrt{1-r^2}} = \frac{0.67\sqrt{48}}{\sqrt{1-(0.67)^2}} = 6.253$

$t_{.005,48} = 2.682$

$t_0 > t_{0.005,48}$

Reject H_0. P-value $\cong 0$

b) $\tanh(\text{arctanh } 0.67 - \frac{z_{.005}}{\sqrt{47}}) \le \rho \le \tanh(\text{arctanh } 0.67 + \frac{z_{.005}}{\sqrt{47}})$

where $z_{0.005} = 2.575$.

$0.4096 \le \rho \le 0.8294$

c) Yes.

11-70 a) r = 0.933203.

a) $H_0 : \rho = 0$

$H_1 : \rho \neq 0$ $\alpha = 0.1$

$t_0 = \frac{r\sqrt{n-2}}{\sqrt{1-r^2}} = \frac{0.933203\sqrt{15}}{\sqrt{1-(0.8709)}} = 10.06$

$t_{.05,15} = 1.753$

$t_0 > t_{\alpha/2,15}$

Reject H_0

c) $\hat{y} = 0.72538 + 0.498081x$

$H_0 : \beta_1 = 0$

$H_1 : \beta_1 \neq 0$ $\alpha = 0.1$

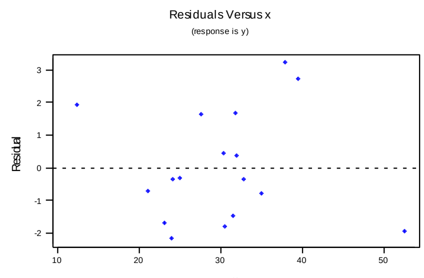
$f_0 = 101.16$

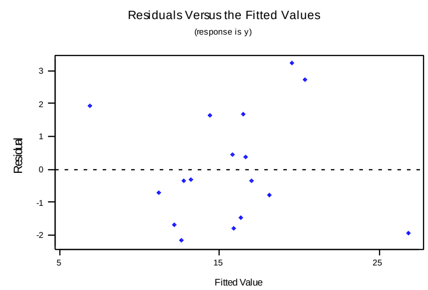
$f_{0.1,1,15} = 3.07$

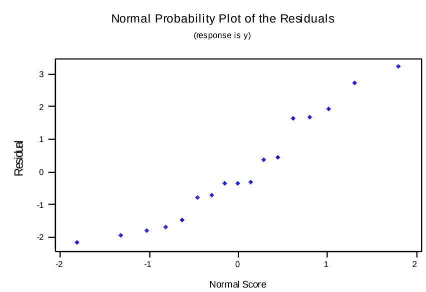
$f_0 >> f_{\alpha,1,15}$

Reject H_0. Conclude that the model is significant at $\alpha = 0.1$. This test and the one in part b) are identical.

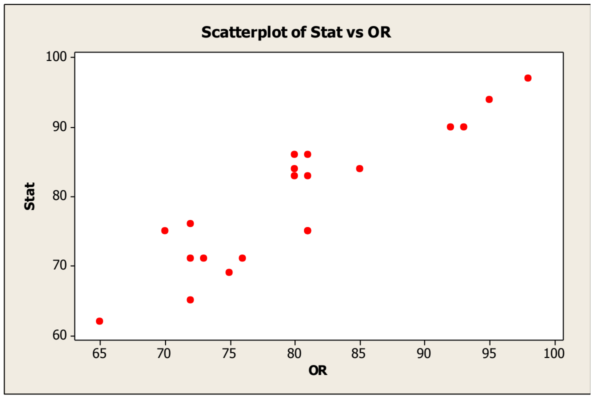
d) No problems with model assumptions are noted.







11-71 a) $\hat{y} = -0.0280411 + 0.990987x$



b) $H_0: \beta_1 = 0$

$H_1: \beta_1 \neq 0$ $\alpha = 0.05$

$f_0 = 79.838$

$f_{0.05,1,18} = 4.41$

$f_0 >> f_{\alpha,1,18}$

Reject H_0

c) $r = \sqrt{0.816} = 0.903$

d) $H_0: \rho = 0$

$H_1: \rho \neq 0$ $\alpha = 0.05$

$$t_0 = \frac{R\sqrt{n-2}}{\sqrt{1-R^2}} = \frac{0.90334\sqrt{18}}{\sqrt{1-0.816}} = 8.9345$$

$t_{0.025,18} = 2.101$

$t_0 > t_{\alpha/2,18}$

Reject H_0

e) $H_0: \rho = 0.8$

$H_1: \rho \neq 0.8$ $\alpha = 0.05$

$z_0 = 1.606$

$z_{.025} = 1.96$

$z_0 < z_{\alpha/2}$

Fail to reject H_0

f) $\tanh(\text{arctanh}\ 0.90334 - \frac{z_{.025}}{\sqrt{17}}) \leq \rho \leq \tanh(\text{arctanh}\ 0.90334 + \frac{z_{.025}}{\sqrt{17}})$ where $z_{0.025} = 1.96$.

$0.7677 \leq \rho \leq 0.9615$

11-72 a) $\hat{y} = 69.1044 + 0.419415x$

b) $H_0: \beta_1 = 0$

$H_1: \beta_1 \neq 0$ $\alpha = 0.05$

$f_0 = 35.744$

$f_{0.05,1,24} = 4.260$

$f_0 > f_{\alpha,1,24}$

Reject H_0

c) $r = 0.77349$

d) $H_0: \rho = 0$

$H_1: \rho \neq 0$ α = 0.05

$t_0 = \frac{0.77349\sqrt{24}}{\sqrt{1-0.5983}} = 5.9787$

$t_{0.025,24} = 2.064$

$t_0 > t_{\alpha/2,24}$

Reject H_0

e) $H_0: \rho = 0.6$

$H_1: \rho \neq 0.6$ α = 0.05

$z_0 = (\text{arctanh } 0.77349 - \text{arctanh } 0.6)(23)^{1/2} = 1.6105$

$z_{.025} = 1.96$

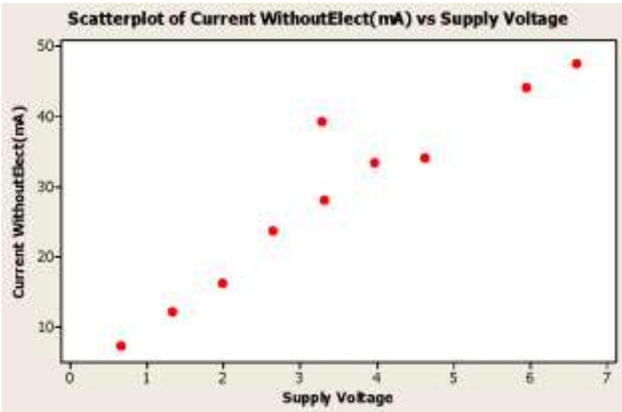
$z_0 \not> z_{\alpha/2}$

Fail to reject H_0

f) $\tanh(\text{arctanh } 0.77349 - \frac{z_{.025}}{\sqrt{23}}) \leq \rho \leq \tanh(\text{arctanh } 0.77349 + \frac{z_{.025}}{\sqrt{23}})$ where $z_{0.025} = 1.96$

$0.5513 \leq \rho \leq 0.8932$

11-73 a)




```
The regression equation is
Current WithoutElect(mA) = 5.50 + 6.73 Supply Voltage

Predictor          Coef  SE Coef     T      P
Constant          5.503    3.104  1.77  0.114
Supply Voltage   6.7342   0.7999  8.42  0.000

S = 4.59061   R-Sq = 89.9%   R-Sq(adj) = 88.6%

Analysis of Variance

Source          DF      SS      MS      F      P
Regression       1  1493.7  1493.7  70.88  0.000
Residual Error   8   168.6    21.1
Total            9  1662.3
```

$\hat{y} = 5.50 + 6.73x$

Yes, because the P-value ≈ 0, the regression model is significant at α = 0.05.

b) $r = \sqrt{0.899} = 0.948$

c) $H_0 : \rho = 0$

$H_1 : \rho \neq 0$

$$t_0 = \frac{r\sqrt{n-2}}{\sqrt{1-r^2}} = \frac{0.948\sqrt{10-2}}{\sqrt{1-0.948^2}} = 8.425$$

$t_{0.025,8} = 2.306$

$t_0 = 8.425 > t_{0.025,8} = 2.306$

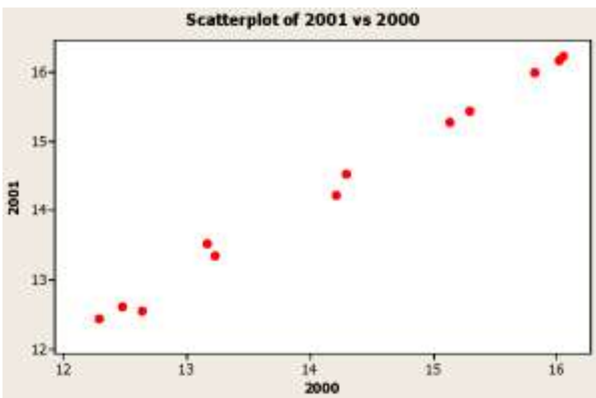
Reject H_0.

d) $$\tanh\left(\text{arctan h} \quad r - \frac{z_{\alpha/2}}{\sqrt{n-3}}\right) \leq \rho \leq \tanh\left(\text{arctan h} \quad r + \frac{z_{\alpha/2}}{\sqrt{n-3}}\right)$$

$$\tanh\left(\text{arctan h} \quad 0.948 - \frac{1.96}{\sqrt{10-3}}\right) \leq \rho \leq \tanh\left(\text{arctan h} \quad 0.948 + \frac{19.6}{\sqrt{10-3}}\right)$$

$0.7898 \leq \rho \leq 0.9879$

11-74 a)




```
The regression equation is
Y2001 = - 0.014 + 1.01 Y2000

Predictor      Coef   SE Coef      T      P
Constant    -0.0144    0.3315  -0.04  0.966
Y2000       1.01127   0.02321  43.56  0.000

S = 0.110372   R-Sq = 99.5%   R-Sq(adj) = 99.4%

Analysis of Variance
```

Source	DF	SS	MS	F	P
Regression	1	23.117	23.117	1897.63	0.000
Residual Error	10	0.122	0.012		
Total	11	23.239			

$\hat{y} = -0.014 + 1.011x$

Yes, because the P-value ≈ 0, the regression model is significant at α = 0.05.

b) $r = \sqrt{0.995} = 0.9975$

c) $H_0 : \rho = 0.9$

$H_1 : \rho \neq 0.9$

$z_0 = (\operatorname{arctan h} \ R - \operatorname{arctan h} \ \rho_0)(n-3)^{1/2}$

$z_0 = (\operatorname{arctan h} \ 0.9975 - \operatorname{arctan h} \ 0.9)(12-3)^{1/2}$

$z_0 = 5.6084$

$z_{\alpha/2} = z_{0.025} = 1.96$

$|z_0| > z_{0.025}$

Reject H_0. P-value = (1 − 1)(2) = 0.000.

d) $\tanh\left(\operatorname{arctan h} \ r - \frac{z_{\alpha/2}}{\sqrt{n-3}}\right) \leq \rho \leq \tanh\left(\operatorname{arctan h} \ r + \frac{z_{\alpha/2}}{\sqrt{n-3}}\right)$

$\tanh\left(\operatorname{arctan h} \ 0.9975 - \frac{1.96}{\sqrt{12-3}}\right) \leq \rho \leq \tanh\left(\operatorname{arctan h} \ 0.9975 + \frac{19.6}{\sqrt{12-3}}\right)$

$0.9908 \leq \rho \leq 0.9993$

11-75 Refer to the computer output in the referenced exercise.

a) r = $\sqrt{0.672}$ = 0.820

b) $H_0 : \rho = 0$

$H_1 : \rho \neq 0$

$$t_0 = \frac{r\sqrt{n-2}}{\sqrt{1-r^2}} = \frac{0.82\sqrt{32-2}}{\sqrt{1-0.82^2}} = 7.847$$

$t_{0.025,30} = 2.042$

$t_0 > t_{0.025,30}$

Reject H_0, P-value < 0.0005

c)

$$\tanh\left(\operatorname{arctan h}(0.082) - \frac{1.96}{\sqrt{32-3}}\right) \leq \rho \leq \tanh\left(\operatorname{arctan h}(0.082) + \frac{19.6}{\sqrt{32-3}}\right)$$

$0.660 \leq \rho \leq 0.909$

d)

$$H_0 : \rho = 0.6$$

$$H_1 : \rho \neq 0.6$$

$$z_0 = (\operatorname{arctanh} R - \operatorname{arctanh} \rho_0)(n-3)^{1/2}$$

$$z_0 = (\operatorname{arctanh} 0.82 - \operatorname{arctanh} 0.6)(32-3)^{1/2}$$

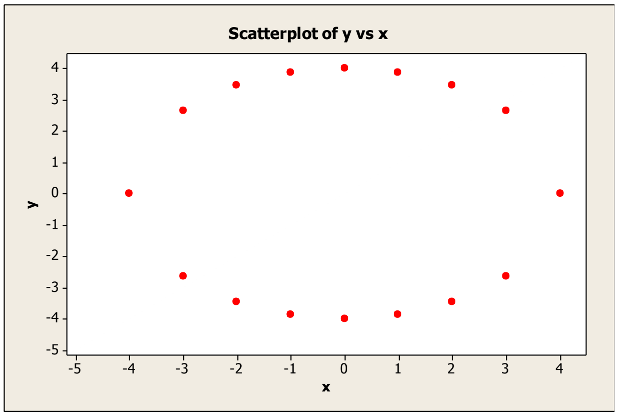
$$z_0 = 2.50$$

$$z_{\alpha/2} = z_{0.025} = 1.96$$

$$|z_0| > z_{0.025}$$

Reject H_0, P-value = 2(0.00621) = 0.0124

11-76



Here r = 0. The correlation coefficient does not detect the relationship between x and y because the relationship is not linear. See the graph above.

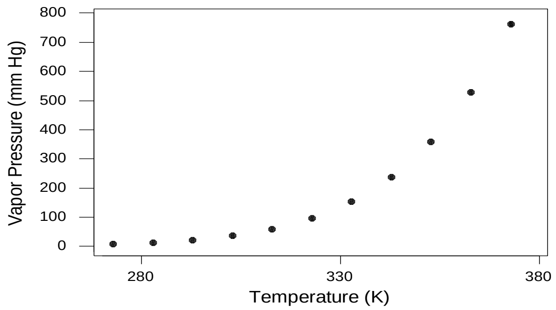
<u>Section 11-9</u>

11-77 a) Yes, $\ln y = \ln \beta_0 + \beta_1 \ln x + \ln \varepsilon$

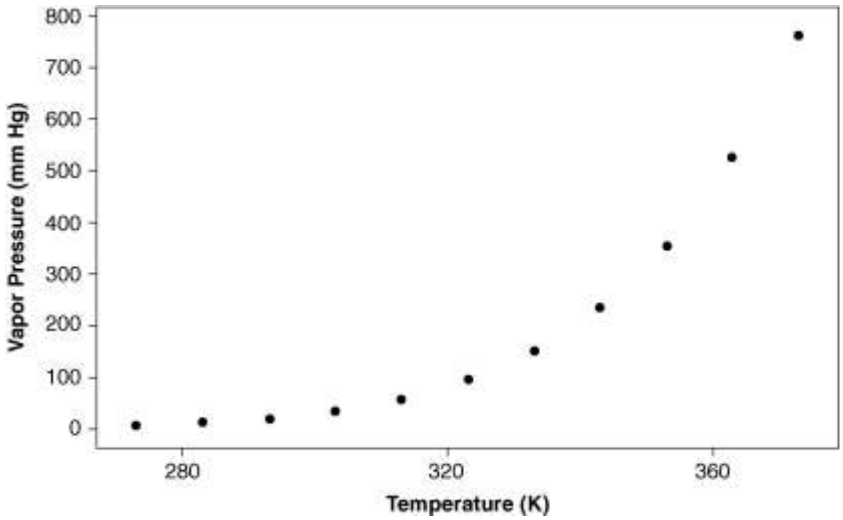
b) No

c) Yes, $\ln y = \ln \beta_0 + x \ln \beta_1 + \ln \varepsilon$

d) Yes, $\dfrac{1}{y} = \beta_0 + \beta_1 \dfrac{1}{x} + \varepsilon$



11-78 a) There is curvature in the data.

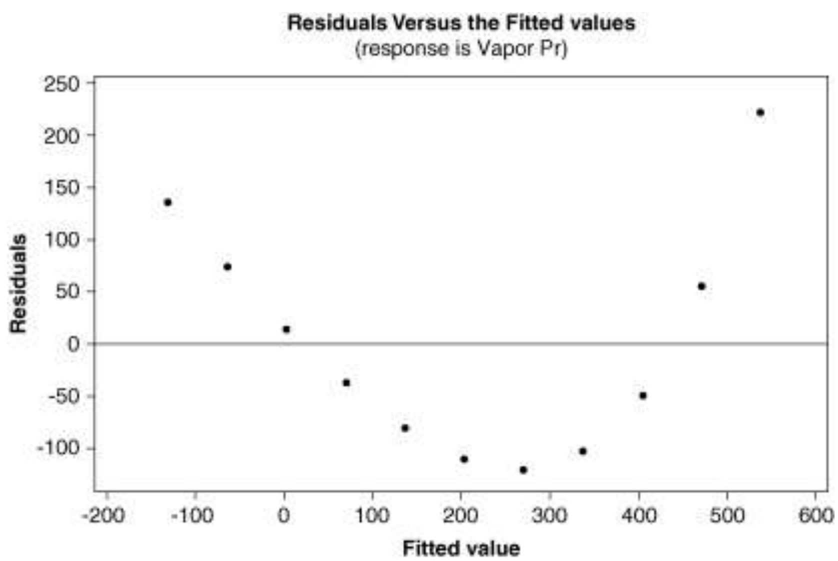


b) $y = -1955.8 + 6.684x$

c)

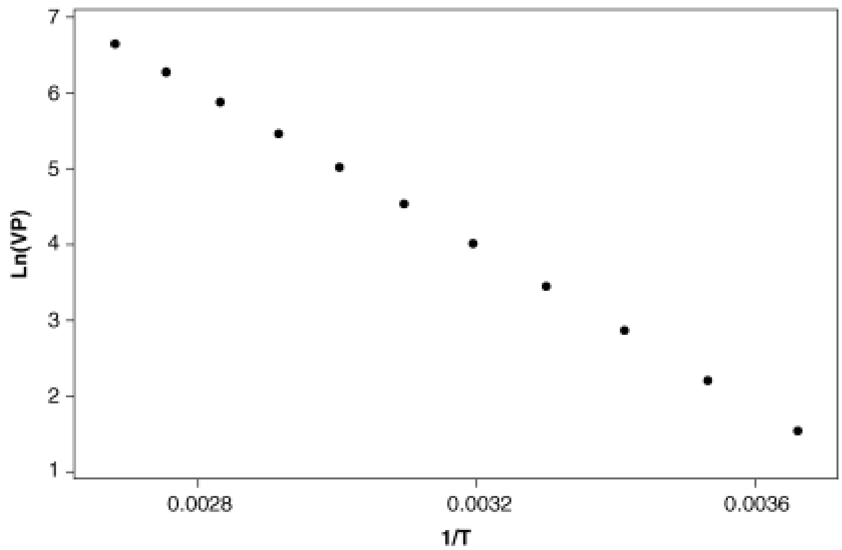
Source	DF	SS	MS	F	P
Regression	1	491448	491448	35.54	0.000
Residual Error	9	124444	13827		
Total	10	615892			

d)



There is a curve in the residuals.

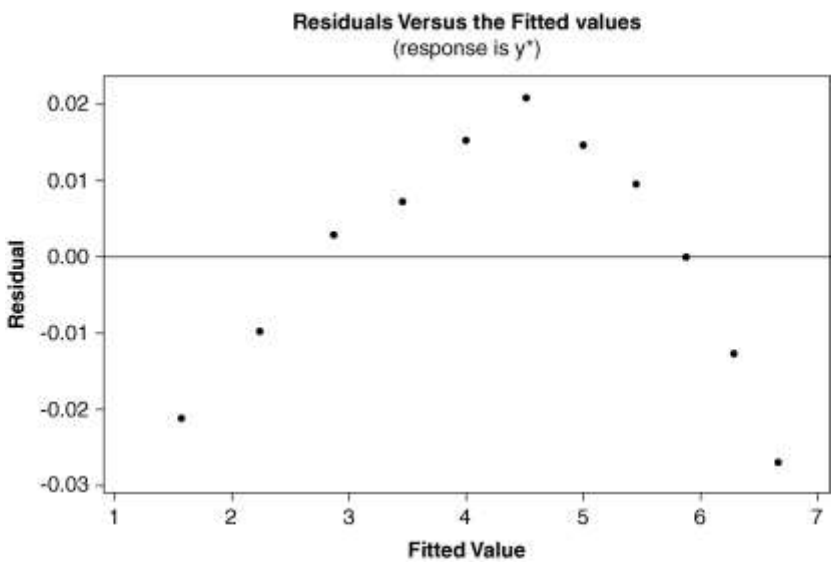
e) The data are linear after the transformation to $y^* = \ln y$ and $x^* = 1/x$.



ln $y = 20.6 - 5185(1/x)$

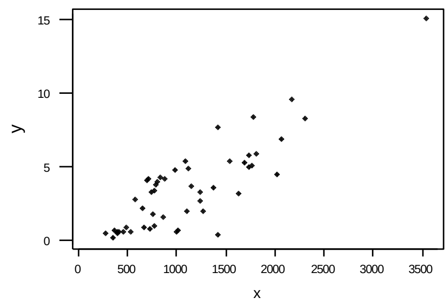
Analysis of Variance

Source	DF	SS	MS	F	P
Regression	1	28.334	28.334	103488.96	0.000
Residual Error	9	0.002	0.000		
Total	10	28.336			



There is still curvature in the data, but now the plot is convex instead of concave.

11-79 a)



b) $\hat{y} = -0.8819 + 0.00385x$

c) $H_0: \beta_1 = 0$

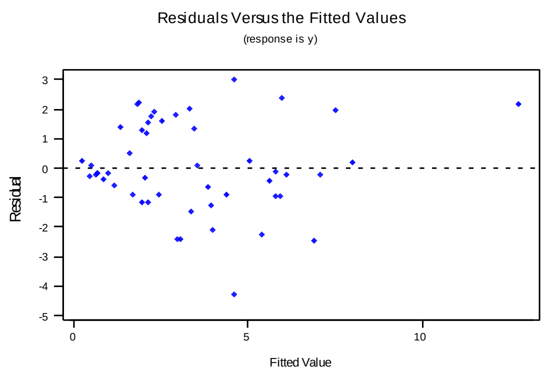
$H_1: \beta_1 \neq 0$ $\alpha = 0.05$

$f_0 = 122.03$

$f_0 > f_{0.05,1,48}$

Reject H_0. Conclude that regression model is significant at $\alpha = 0.05$

d) No, it seems the variance is not constant, there is a funnel shape.



e) $\hat{y}^* = 0.5967 + 0.00097x$. Yes, the transformation stabilizes the variance.

<u>Section 11-10</u>

11-80 a) The fitted logistic regression model is $\hat{y} = \dfrac{1}{1 + \exp[-(-8.84679 - 0.000202x)]}$

The Minitab result is shown below

Binary Logistic Regression: Home Ownership Status versus Income

```
Link Function: Logit
Response Information
Variable                Value    Count
Home Ownership Status   1           11  (Event)
                        0            9
                        Total       20
```

```
Logistic Regression Table
                                                         Odds     95% CI
Predictor        Coef    SE Coef       Z      P  Ratio  Lower  Upper
Constant    -8.84679    4.44559  -1.99  0.047
Income     0.0002027  0.0001004   2.02  0.044   1.00   1.00   1.00

Log-Likelihood = -11.163
Test that all slopes are zero: G = 5.200, DF = 1, P-Value = 0.023
```

b) The P-value for the test of the coefficient of *income* is 0.044 < α = 0.05. Therefore, *income* has a significant effect on home ownership status.

c) The odds ratio is changed by the factor $\exp(\beta_1) = \exp(0.0002027) = 1.000202$ for every unit increase in *income*. More realistically, if income changes by $1000, the odds ratio is changed by the factor $\exp(1000\beta_1) = \exp(0.2027) = 1.225$.

11-81 a) The fitted logistic regression model is $\hat{y} = \dfrac{1}{1+\exp[-(5.33968-0.000224x)]}$

The Minitab result is shown below

Binary Logistic Regression: Number Failing, Sample Size, versus Load (kN/m²)

```
Link Function: Logit
Response Information
Variable           Value    Count
Number Failing     Failure    337
                   Success    353
Sample Size        Total      690

Logistic Regression Table
                                                             Odds     95% CI
Predictor            Coef    SE Coef       Z      P  Ratio  Lower  Upper
Constant          5.33968   0.545698    9.79  0.000
load (kN/m2)   -0.0002246  0.0000228   -9.83 0.000   1.00   1.00   1.00

Log-Likelihood = -421.856
Test that all slopes are zero: G = 112.459, DF = 1, P-Value = 0.000
```

b) The P-value for the test of the coefficient of *load* is near zero. Therefore, *load* has a significant effect on failing performance.

11-82 a) The fitted logistic regression model is $\hat{y} = \dfrac{1}{1+\exp[-(-2.12756+0.113925x)]}$

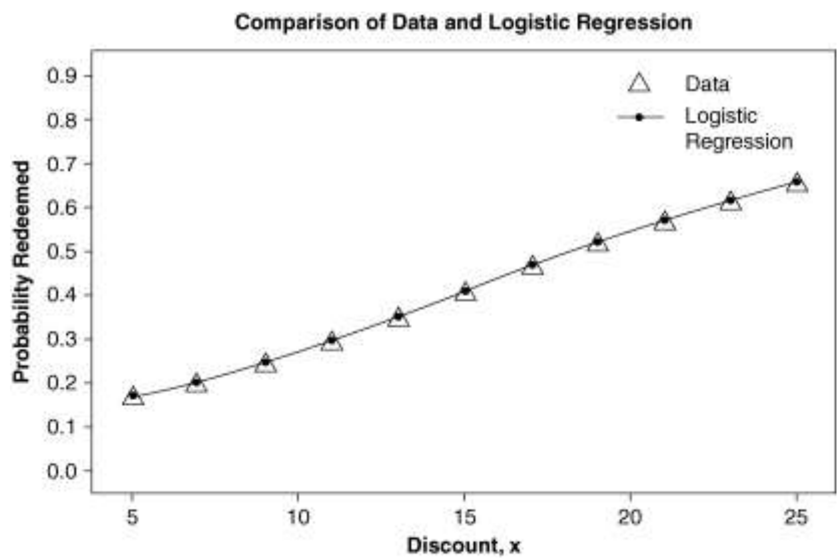
The Minitab results are shown below

Binary Logistic Regression: Number Redee, Sample size, versus Discount, x

```
Link Function: Logit
Response Information
Variable            Value     Count
Number Redeemed     Success    2693
                    Failure    3907
Sample Size         Total      6600

Logistic Regression Table
                                                        Odds     95% CI
Predictor          Coef    SE Coef       Z       P  Ratio  Lower  Upper
Constant     -2.12756  0.0746903  -28.49  0.000
Discount, x  0.113925  0.0044196   25.78  0.000   1.12   1.11   1.13

Log-Likelihood = -4091.801
Test that all slopes are zero: G = 741.361, DF = 1, P-Value = 0.000
```

b) The P-value for the test of the coefficient of *discount* is near zero. Therefore, *discount* has a significant effect on redemption.

c)



d) The P-value of the quadratic term is 0.95 > 0.05, so we fail to reject the null hypothesis of the quadratic coefficient at the 0.05 level of significance. There is no evidence that the quadratic term is required in the model. The Minitab results are shown below

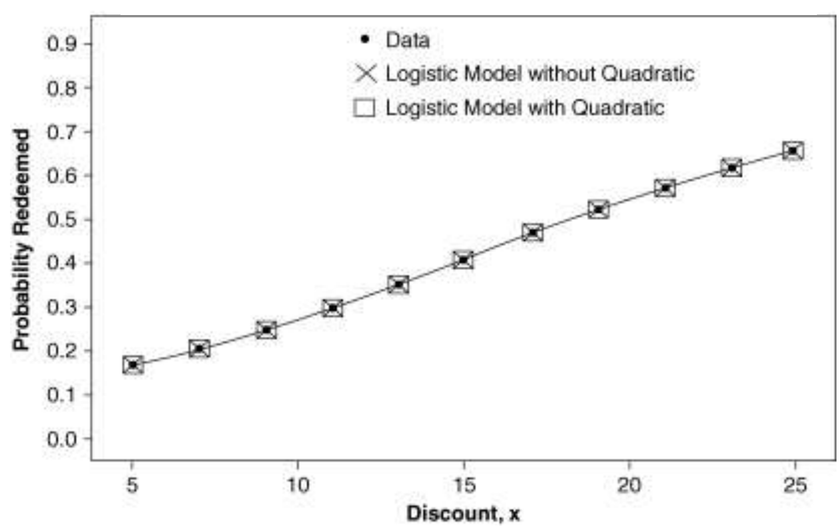
Binary Logistic Regression: Number Redee, Sample size, versus Discount, x

```
Link Function: Logit
```

```
Response Information
Variable            Value      Count
Number Redeemed     Event       2693
                    Non-event   3907
Sample Size         Total       6600

Logistic Regression Table
                                                                      95%
                                                               Odds    CI
Predictor                        Coef    SE Coef        Z       P  Ratio  Lower
Constant                     -2.34947   0.174523   -13.46   0.000
Discount, x                  0.148003  0.0245118     6.04   0.000   1.16   1.11
Discount, x* Discount, x -0.0011084   0.0007827    -1.42   0.157   1.00   1.00

Predictor                 Upper
Constant
Discount, x                1.22
Discount, x*Discount, x    1.00

Log-Likelihood = -4090.796
Test that all slopes are zero: G = 743.372, DF = 2, P-Value = 0.000
```

e) The expanded model does not visually provide a better fit to the data than the original model.

Comparison of Data and two Logistic Regression



11-83 a) The Minitab results are shown below

Binary Logistic Regression: y versus Income x1, Age x2

```
Link Function: Logit
```

```
Response Information
Variable  Value  Count
y         1         10  (Event)
          0         10
          Total     20

Logistic Regression Table
                                                Odds     95% CI
Predictor       Coef    SE Coef       Z       P  Ratio  Lower  Upper
Constant    -7.79891    5.05557  -1.54  0.123
Income x1  0.0000833  0.0000678   1.23  0.220   1.00   1.00   1.00
Age x2       1.06263   0.567664   1.87  0.061   2.89   0.95   8.80

Log-Likelihood = -10.423
Test that all slopes are zero: G = 6.880, DF = 2, P-Value = 0.032
```

b) Because the P-value = 0.032 < α = 0.05 we can conclude that at least one of the coefficients (of *income* and *age*) is not equal to zero at the 0.05 level of significance. The individual z-tests do not generate P-values less than 0.05, but this might be due to correlation between the independent variables. The z-test for a coefficient assumes it is the last variable to enter the model. A model might use either *income* or *age*, but after one variable is in the model, the coefficient z-test for the other variable may not be significant because of their correlation.

c) The odds ratio is changed by the factor $\exp(\beta_1) = \exp(0.0000833) = 1.00008$ for every unit increase in *income* with *age* held constant. Similarly, odds ratio is changed by the factor $\exp(\beta_1) = \exp(1.06263) = 2.894$ for every unit increase in *age* with *income* held constant. More realistically, if income changes by \$1000, the odds ratio is changed by the factor $\exp(1000\beta_1) = \exp(0.0833) = 1.087$ with *age* held constant.

d) At $x_1 = 45000$ and $x_2 = 5$ from part (a)

$$\hat{y} = \frac{1}{1+\exp[-(-7.79891+0.0000833x_1+1.06263x_2)]} = 0.78$$

e) The Minitab results are shown below

Binary Logistic Regression: y versus Income x1, Age x2

```
Link Function: Logit
Response Information
Variable  Value  Count
y         1         10  (Event)
          0         10
          Total     20

Logistic Regression Table
                                                        Odds     95% CI
```

Predictor	Coef	SE Coef	Z	P	Ratio	Lower	Upper
Constant	-0.494471	6.64311	-0.07	0.941			
Income x1	-0.0001314	0.0001411	-0.93	0.352	1.00	1.00	1.00
Age x2	-2.39447	2.07134	-1.16	0.248	0.09	0.00	5.29
Income x1*Age x2	0.0001017	0.0000626	1.62	0.104	1.00	1.00	1.00

```
Log-Likelihood = -8.112
Test that all slopes are zero: G = 11.503, DF = 3, P-Value = 0.009
```

Because the P-value = 0.104 there is no evidence that an interaction term is required in the model.

Supplemental Exercises

11-84 a) $\sum_{i=1}^{n}(y_i - \hat{y}_i) = \sum_{i=1}^{n} y_i - \sum_{i=1}^{n}\hat{y}_i$ and $\sum y_i = n\hat{\beta}_0 + \hat{\beta}_1 \sum x_i$ from the normal equations

Then,

$$(n\hat{\beta}_0 + \hat{\beta}_1 \sum_{i=1}^{n} x_i) - \sum \hat{y}_i$$

$$= n\hat{\beta}_0 + \hat{\beta}_1 \sum_{i=1}^{n} x_i - \sum_{i=1}^{n}(\hat{\beta}_0 + \hat{\beta}_1 x_i)$$

$$= n\hat{\beta}_0 + \hat{\beta}_1 \sum_{i=1}^{n} x_i - n\hat{\beta}_0 - \hat{\beta}_1 \sum_{i=1}^{n} x_i = 0$$

b) $\sum_{i=1}^{n}(y_i - \hat{y}_i)x_i = \sum_{i=1}^{n} y_i x_i - \sum_{i=1}^{n}\hat{y}_i x_i$

and $\sum_{i=1}^{n} y_i x_i = \hat{\beta}_0 \sum_{i=1}^{n} x_i + \hat{\beta}_1 \sum_{i=1}^{n} x_i^2$ from the normal equations. Then,

$$\hat{\beta}_0 \sum_{i=1}^{n} x_i + \hat{\beta}_1 \sum_{i=1}^{n} x_i^2 - \sum_{i=1}^{n}(\hat{\beta}_0 + \hat{\beta}_1 x_i)x_i =$$

$$\hat{\beta}_0 \sum_{i=1}^{n} x_i + \hat{\beta}_1 \sum_{i=1}^{n} x_i^2 - \hat{\beta}_0 \sum_{i=1}^{n} x_i - \hat{\beta}_1 \sum_{i=1}^{n} x_i^2 = 0$$

c) $\frac{1}{n}\sum_{i=1}^{n}\hat{y}_i = \bar{y}$

$$\sum \hat{y} = \sum(\hat{\beta}_0 + \hat{\beta}_1 x)$$

$$\frac{1}{n}\sum_{i=1}^{n}\hat{y}_i = \frac{1}{n}\sum(\hat{\beta}_0 + \hat{\beta}_1 x_i)$$

$$= \frac{1}{n}(n\hat{\beta}_0 + \hat{\beta}_1 \sum x_i)$$

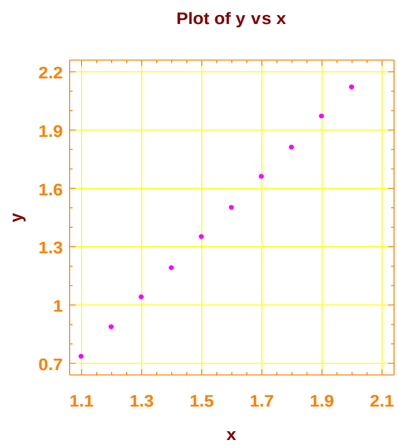
$$= \frac{1}{n}(n(\bar{y} - \hat{\beta}_1 \bar{x}) + \hat{\beta}_1 \sum x_i)$$

$$= \frac{1}{n}(n\bar{y} - n\hat{\beta}_1 \bar{x} + \hat{\beta}_1 \sum x_i)$$

$$= \bar{y} - \hat{\beta}\bar{x} + \hat{\beta}_1 \bar{x}$$

$$= \bar{y}$$

11-85 a)



Yes, a linear relationship seems plausible.

b)

```
                        Model fitting results for: y
Independent variable                coefficient  std. error     t-value
sig.level
CONSTANT                              -0.966824    0.004845   -199.5413
0.0000
x                                      1.543758    0.003074    502.2588
0.0000
--------------------------------------------------------------------------
R-SQ. (ADJ.) = 1.0000  SE = 0.002792  MAE = 0.002063  DurbWat = 2.843
Previously:    0.0000         0.000000            0.000000            0.000
10 observations fitted, forecast(s) computed for 0 missing val. of dep. var.
```

$\hat{y} = -0.966824 + 1.54376x$

c)

```
                   Analysis of Variance for the Full Regression
Source              Sum of Squares      DF     Mean Square      F-Ratio   P-value
Model                      1.96613       1         1.96613      252264.     .0000
Error                 0.0000623515       8   0.00000779394
-----------------------------------------------------------------------------
Total (Corr.)              1.96619       9
R-squared = 0.999968                         Stnd. error of est. = 2.79176E-3
R-squared (Adj. for d.f.) = 0.999964    Durbin-Watson statistic = 2.84309
```

2) $H_0 : \beta_1 = 0$

3) $H_1 : \beta_1 \neq 0$

4) α = 0.05

5) The test statistic is $f_0 = \dfrac{SS_R / k}{SS_E /(n-p)}$

6) Reject H_0 if $f_0 > f_{\alpha,1,8}$ where $f_{0.01,1,8} = 11.26$

7) Using the results from the ANOVA table

$$f_0 = \frac{1.96613/1}{0.0000623515/8} = 252263.9$$

8) Because 252264 > 11.26 reject H_0 and conclude that the regression model is significant at $\alpha = 0.05$.

P-value ≈ 0

d)

```
                 99 percent confidence intervals for coefficient estimates
-----------------------------------------------------------------------------
                     Estimate  Standard error     Lower Limit     Upper Limit
CONSTANT             -0.96682         0.00485        -0.97800        -0.95565
x                     1.54376         0.00307         1.53667         1.55085
-----------------------------------------------------------------------------
```

$1.53667 \le \beta_1 \le 1.55085$

e) 2) $H_0: \beta_0 = 0$

3) $H_1: \beta_0 \ne 0$

4) $\alpha = 0.01$

5) The test statistic is $t_0 = \dfrac{\hat{\beta}_0}{se(\hat{\beta}_0)}$

6) Reject H_0 if $t_0 < -t_{\alpha/2,n-2}$ where $-t_{0.005,8} = -3.355$ or $t_0 > t_{0.005,8} = 3.355$

7) Using the results from the table above

$$t_0 = \frac{-0.96682}{0.00485} = -199.34$$

8) Since $-199.34 < -3.355$ reject H_0 and conclude the intercept is significant at $\alpha = 0.05$.

11-86 a) $\hat{y} = 93.55 + 15.57x$

b) $H_0: \beta_1 = 0$

$H_1: \beta_1 \ne 0$

$\alpha = 0.05$

$f_0 = 12.872$

$f_{.05,1,14} = 4.60$

$f_0 > f_{0.05,1,14}$

Reject H_0. Conclude that $\beta_1 \ne 0$ at $\alpha = 0.05$.

c) $(9.689 \le \beta_1 \le 21.445)$

d) $(79.333 \le \beta_0 \le 107.767)$

e) $\hat{y} = 93.55 + 15.57(2.5) = 132.475$

$$132.475 \pm 2.145\sqrt{136.9\left[\frac{1}{16} + \frac{(2.5-2.325)^2}{7.017}\right]}$$
$$132.475 \pm 6.49$$
$$125.99 \le \hat{\mu}_{Y|x_0=2.5} \le 138.97$$

11-87 $\hat{y}^* = 1.2166 + 0.5086x$ where $y^* = 1/y$. No, the model does not seem reasonable. The residual plots indicate a possible outlier.

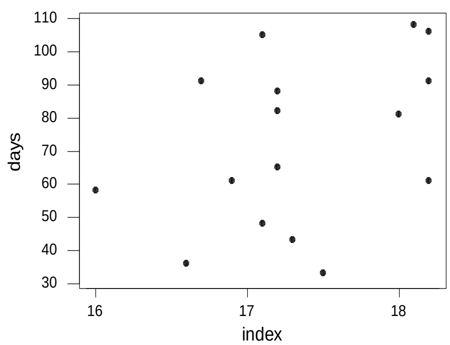
11-88 $\hat{y} = 4.5067 + 2.21517x$, $r = 0.992$, $R^2 = 98.43\%$

The model appears to be an excellent fit. The R^2 is large and both regression coefficients are significant. No, the existence of a strong correlation does not imply a cause and effect relationship.

11-89 $\hat{y} = 0.7916x$

Even though y should be zero when x is zero, because the regressor variable does not usually assume values near zero, a model with an intercept fits this data better. Without an intercept, the residuals plots are not satisfactory.

11-90 a)



b) The regression equation is
$$\hat{y} = -193 + 15.296x$$

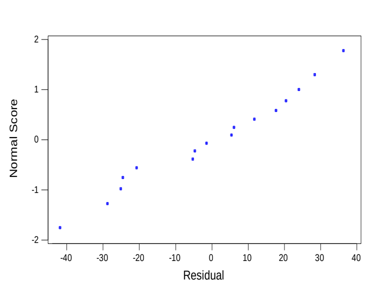
```
Analysis of Variance
Source            DF         SS         MS        F        P
Regression         1     1492.6     1492.6     2.64    0.127
Residual Error    14     7926.8      566.2
Total             15     9419.4
```

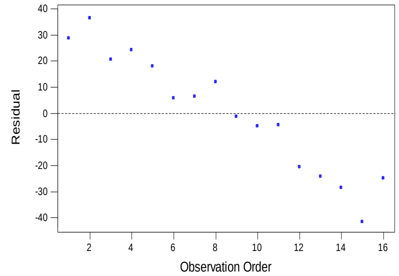
Fail to reject H_0. We do not have evidence of a relationship. Therefore, there is not sufficient evidence to conclude that the seasonal meteorological index (x) is a reliable predictor of the number of days that the ozone level exceeds 0.20 ppm (y).

c) 99% CI on β_1

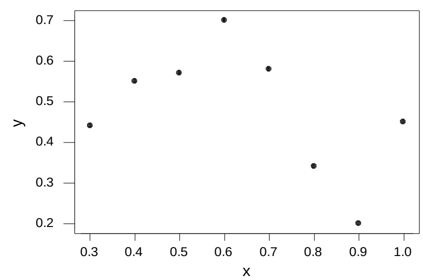
$$\hat{\beta}_1 \pm t_{\alpha/2,n-2}\ se(\hat{\beta}_1)$$
$$-2.3298 \pm t_{.005,12}(0.2697)$$
$$-2.3298 \pm 3.005(0.2697)$$
$$(-3.1402, -1.5194)$$

d) The normality plot of the residuals is satisfactory. However, the plot of residuals versus run order exhibits a strong downward trend. This could indicate that there is another variable should be included in the model and it is one that changes with time.





11-91 a)



b) $\hat{y} = 0.6714 - 2964x$

c)

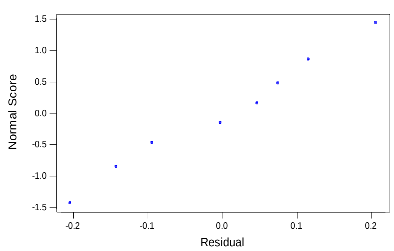
Analysis of Variance

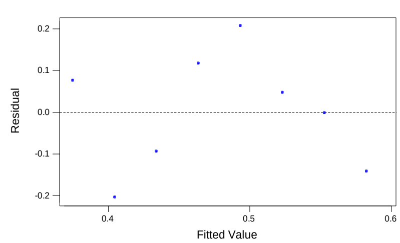
Source	DF	SS	MS	F	P
Regression	1	0.03691	0.03691	1.64	0.248
Residual Error	6	0.13498	0.02250		
Total	7	0.17189			

R^2 = 21.47%

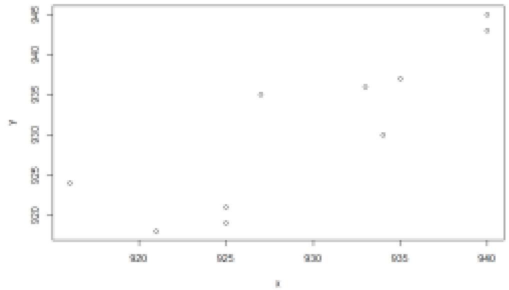
Because the P-value > 0.05, reject the null hypothesis and conclude that the model is significant.

d) There appears to be curvature in the data. There is a dip in the middle of the normal probability plot and the plot of the residuals versus the fitted values shows curvature.





11-92 a)



b) $\hat{y} = -44.61 + 1.05x$

c)

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	1	643.41930	643.41930	21.79	0.0016
Error	8	236.18070	29.52259		
Corrected Total	9	879.60000			

Root MSE	5.43347	R-Square	0.7315
Dependent Mean	930.80000	Adj R-Sq	0.6979
Coeff Var	0.58374		

Parameter Estimates

Variable	DF	Parameter Estimate	Standard Error	t Value	Pr > |t|
Intercept	1	-44.61191	208.94544	-0.21	0.8363
Therm	1	1.04928	0.22476	4.67	0.0016

Reject the null hypothesis and conclude that the model is significant. Here 73.2% of the variability is explained by the model.

d) $H_0 : \beta_1 = 1$

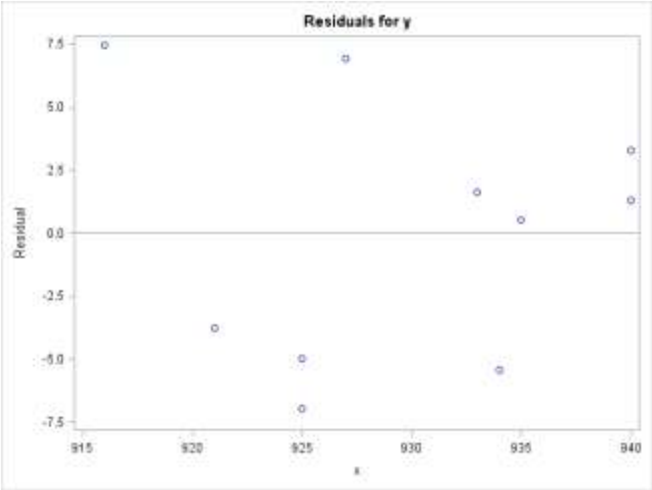
$H_1 : \beta_1 \neq 1$ $\alpha = 0.05$

$$t_0 = \frac{\hat{\beta}_1 - 1}{se(\hat{\beta}_1)} = \frac{1.04928 - 1}{0.22476} = 0.2193$$

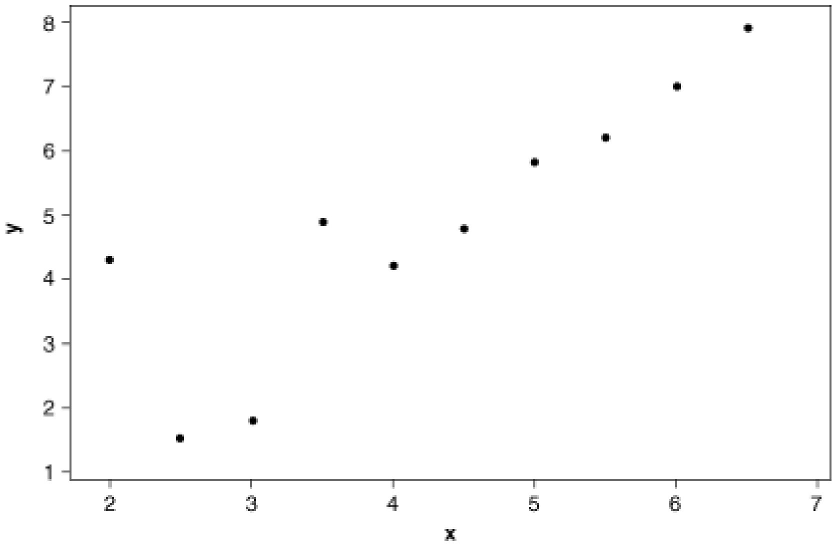
$$t_{a/2,n-2} = t_{.025,8} = 2.306$$

Because $t_0 > -t_{a/2,n-2}$, we fail to reject H_0. There is not enough evidence to reject the claim that the devices produce different temperature measurements.

e) The residual plots to not reveal any major problems.



11-93 a)



b) $\hat{y} = -0.12 + 1.17x$

c)

Source	DF	SS	MS	F	P
Regression	1	28.044	28.044	22.75	0.001
Residual Error	8	9.860	1.233		
Total	9	37.904			

Reject the null hypothesis and conclude that the model is significant.

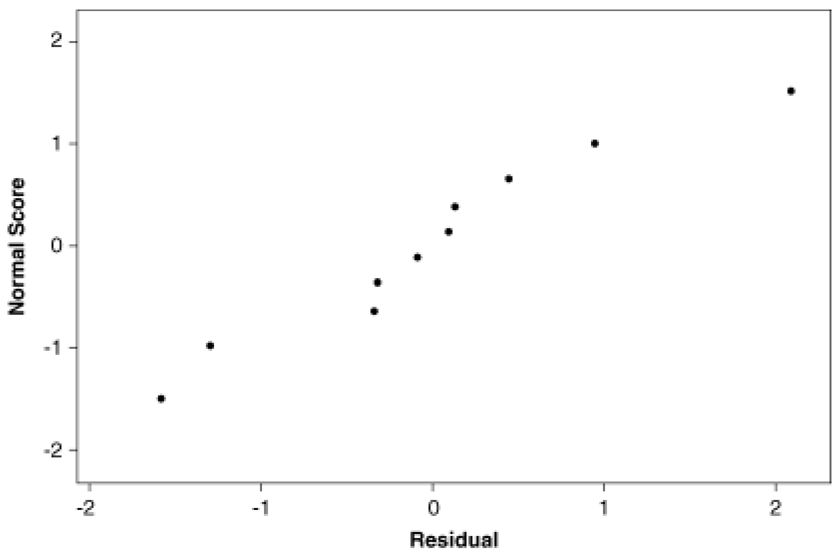
d) $x_0 = 4.25 \quad \hat{\mu}_{y|x_0} = 4.853$

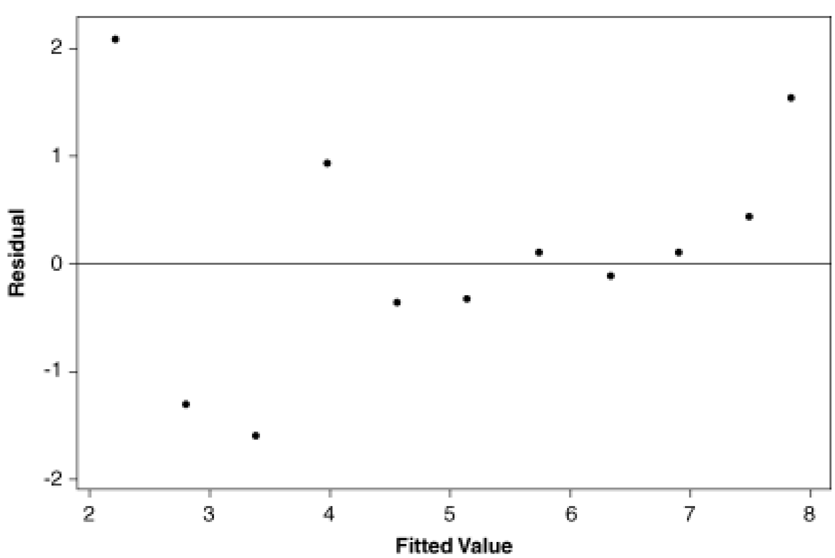
$$4.853 \pm 2.306\sqrt{1.2324\left(\frac{1}{10}+\frac{(4.25-4.25)^2}{20.625}\right)}$$

$$4.853 \pm 2.306(0.35106)$$

$$4.0435 \le \mu_{y|x_0} \le 5.6625$$

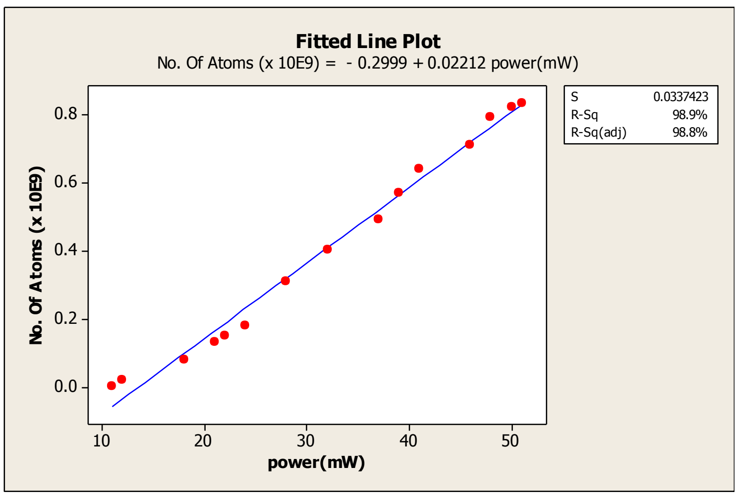
e) The normal probability plot of the residuals appears linear, but there are some large residuals in the lower fitted values. There may be some problems with the model.





11-94 a)

```
The regression equation is
No. Of Atoms (x 10E9) = - 0.300 + 0.0221 power(mW)


Predictor        Coef     SE Coef       T      P
Constant     -0.29989     0.02279  -13.16  0.000
power(mW)   0.0221217   0.0006580   33.62  0.000


S = 0.0337423   R-Sq = 98.9%   R-Sq(adj) = 98.8%


Analysis of Variance

Source          DF      SS      MS        F      P
Regression       1  1.2870  1.2870  1130.43  0.000
Residual Error  13  0.0148  0.0011
Total           14  1.3018
```



b) Yes, there is a significant regression at α = 0.05 because p-value = 0.000 < α.

c) $r = \sqrt{0.989} = 0.994$

d)

$$H_0 : \rho = 0$$

$$H_1 : \rho \neq 0$$

$$t_0 = \frac{r\sqrt{n-2}}{\sqrt{1-r^2}} = \frac{0.994\sqrt{15-2}}{\sqrt{1-.994^2}} = 32.766$$

$$t_{0.05,13} = 1.771$$

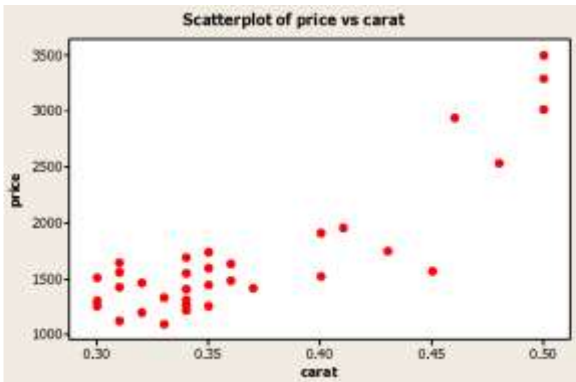
$$t_0 = 32.766 > t_{0.05,13} = 1.771.$$

Reject H_0, P-value ≈ 0.000

e) 99% confidence interval for $\hat{\beta}_1$

$$\hat{\beta}_1 \pm t_{\alpha/2,n-2} se(\hat{\beta}_1)$$
$$0.022 \pm t_{0.05,13}(0.00066)$$
$$0.022 \pm 1.771(0.00066)$$
$$0.0208 \le \hat{\beta}_1 \le 0.0232$$

11-95 a)



The relationship between carat and price is not linear. Yes, there is one outlier, observation number 33.

b) The person obtained a very good price—high carat diamond at low price.

c) All the data

```
The regression equation is
price = - 1696 + 9349 carat

Predictor      Coef  SE Coef      T      P
Constant    -1696.2    298.3  -5.69  0.000
carat        9349.4    794.1  11.77  0.000

S = 331.921   R-Sq = 78.5%   R-Sq(adj) = 77.9%

Analysis of Variance

Source          DF        SS        MS       F      P
Regression       1  15270545  15270545  138.61  0.000
Residual Error  38   4186512    110171
Total           39  19457057
```

$t_{\alpha/2,n-2} = t_{0.005,38} = 2.713$

99% confidence interval on β_1.

$$\hat{\beta}_1 \pm t_{\alpha/2,n-2} se(\hat{\beta}_1)$$
$$9349 \pm t_{0.005,38}(794.1)$$
$$9349 \pm 2.713(794.1)$$
$$7194.6067 \le \beta_1 \le 11503.3933$$

With unusual data omitted

```
The regression equation is
price_1 = - 1841 + 9809 carat_1

Predictor     Coef  SE Coef      T      P
Constant   -1841.2    269.9  -6.82  0.000
carat_1     9809.2    722.5  13.58  0.000

S = 296.218   R-Sq = 83.3%   R-Sq(adj) = 82.8%

Analysis of Variance
Source          DF        SS        MS       F      P
Regression       1  16173949  16173949  184.33  0.000
Residual Error  37   3246568     87745
Total           38  19420517
```

$t_{\alpha/2,n-2} = t_{0.005,37} = 2.718$

99% confidence interval on β_1.

$$\hat{\beta}_1 \pm t_{\alpha/2,n-2} se(\hat{\beta}_1)$$
$$9809 \pm t_{0.005,37}(722.5)$$
$$9809 \pm 2.718(722.5)$$
$$7845.25 \le \beta_1 \le 11772.76$$

The width for the outlier removed is narrower than for the first case.

11-96

```
The regression equation is
Population = 3549143 + 651828 Count

Predictor     Coef  SE Coef      T      P
Constant   3549143   131986  26.89  0.000
Count       651828   262844   2.48  0.029

S = 183802   R-Sq = 33.9%   R-Sq(adj) = 28.4%

Analysis of Variance
Source          DF            SS            MS      F      P
```

```
Regression       1  2.07763E+11  2.07763E+11  6.15  0.029
Residual Error  12  4.05398E+11  33783126799
Total           13  6.13161E+11
```

$\hat{y} = 3549143 + 651828x$

Yes, the regression is significant at $\alpha = 0.01$. Care needs to be taken in making cause and effect statements based on a regression analysis. In this case, it is surely not the case that an increase in the stork count is causing the population to increase, in fact, the opposite is most likely the case. However, unless a designed experiment is performed, cause and effect statements should not be made on regression analysis alone. The existence of a strong correlation does not imply a cause and effect relationship.

Mind-Expanding Exercises

11-97 The correlation coefficient for the *n* pairs of data (x_i, z_i) can be much different from unity. For example, if y = bx and if the x data is symmetric about zero, the correlation coefficient between x and y^2 is zero. In other cases, it can be much less than unity (in absolute value). Over some restricted ranges of x values, the quadratic function y = $(a + bx)^2$ can be approximated by a linear function of x and in these cases the correlation can still be near unity. However, in general, the correlation can be much different from unity and in some cases equal zero. Correlation is a measure of a linear relationship, and if a nonlinear relationship exists between variables, even if it is strong, the correlation coefficient does not usually provide a good measure.

11-98 a) $\hat{\beta}_1 = \dfrac{S_{xY}}{S_{xx}}, \quad \hat{\beta}_0 = \overline{Y} - \hat{\beta}_1 \bar{x}$

$$Cov(\hat{\beta}_0, \hat{\beta}_1) = Cov(\overline{Y}, \hat{\beta}_1) - \bar{x}Cov(\hat{\beta}_1, \hat{\beta}_1)$$

$$Cov(\overline{Y}, \hat{\beta}_1) = \frac{Cov(\overline{Y}, S_{xY})}{S_{xx}} = \frac{Cov(\sum Y_i, \sum Y_i(x_i - \bar{x}))}{nS_{xx}} = \frac{\sum (x_i - \bar{x})\sigma^2}{nS_{xx}} = 0.$$ Therefore,

$$Cov(\hat{\beta}_1, \hat{\beta}_1) = V(\hat{\beta}_1) = \frac{\sigma^2}{S_{xx}}$$

$$Cov(\hat{\beta}_0, \hat{\beta}_1) = \frac{-\bar{x}\sigma^2}{S_{xx}}$$

b) The requested result is shown in part a).

11-99 a) $MS_E = \dfrac{\sum (Y_i - \hat{\beta}_0 - \hat{\beta}_1 x_i)^2}{n-2} = \dfrac{\sum e_i^2}{n-2}$

$$E(e_i) = E(Y_i) - E(\hat{\beta}_0) - E(\hat{\beta}_1)x_i = 0$$

$$V(e_i) = \sigma^2[1 - (\tfrac{1}{n} + \tfrac{(x_i - \bar{x})^2}{S_{xx}})]$$ Therefore,

$$E(MS_E) = \frac{\sum E(e_i^2)}{n-2} = \frac{\sum V(e_i)}{n-2}$$

$$= \frac{\sum \sigma^2[1 - (\frac{1}{n} + \frac{(x_i - \bar{x})^2}{S_{xx}})]}{n-2}$$

$$= \frac{\sigma^2[n-1-1]}{n-2} = \sigma^2$$

b) Using the fact that $SS_R = MS_R$, we obtain

$$E(MS_R) = E(\hat{\beta}_1^2 S_{xx}) = S_{xx}\left\{V(\hat{\beta}_1) + [E(\hat{\beta}_1)]^2\right\}$$
$$= S_{xx}\left(\frac{\sigma^2}{S_{xx}} + \beta_1^2\right) = \sigma^2 + \beta_1^2 S_{xx}$$

11-100 $\hat{\beta}_1 = \dfrac{S_{x_1Y}}{S_{x_1x_1}}$

$$E(\hat{\beta}_1) = \frac{E\left[\sum_{i=1}^{n} Y_i(x_{1i} - \bar{x}_1)\right]}{S_{x_1x_1}} = \frac{\sum_{i=1}^{n}(\beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i})(x_{1i} - \bar{x}_1)}{S_{x_1x_1}}$$
$$= \frac{\beta_1 S_{x_1x_1} + \beta_2 \sum_{i=1}^{n} x_{2i}(x_{1i} - \bar{x}_1)}{S_{x_1x_1}} = \beta_1 + \frac{\beta_2 S_{x_1x_2}}{S_{x_1x_1}}$$

No, $\hat{\beta}_1$ is no longer unbiased.

11-101 $V(\hat{\beta}_1) = \dfrac{\sigma^2}{S_{xx}}$. To minimize $V(\hat{\beta}_1)$, S_{xx} should be maximized.

Because $S_{xx} = \sum_{i=1}^{n}(x_i - \bar{x})^2$, S_{XX} is maximized by choosing approximately half of the observations at each end of the range of *x*. From a practical perspective, this allocation assumes the linear model between *Y* and *x* holds throughout the range of *x* and observing *Y* at only two *x* values prohibits verifying the linearity assumption. It is often preferable to obtain some observations at intermediate values of *x*.

11-102 One might minimize a weighted some of squares $\sum_{i=1}^{n} w_i(y_i - \beta_0 - \beta_1 x_i)^2$ in which a Y_i with small variance (w_i large) receives greater weight in the sum of squares.

$$\frac{\partial}{\beta_0}\sum_{i=1}^{n} w_i(y_i - \beta_0 - \beta_1 x_i)^2 = -2\sum_{i=1}^{n} w_i(y_i - \beta_0 - \beta_1 x_i)$$
$$\frac{\partial}{\beta_1}\sum_{i=1}^{n} w_i(y_i - \beta_0 - \beta_1 x_i)^2 = -2\sum_{i=1}^{n} w_i(y_i - \beta_0 - \beta_1 x_i)x_i$$

Setting these derivatives to zero yields

$$\hat{\beta}_0\sum w_i + \hat{\beta}_1\sum w_i x_i = \sum w_i y_i$$

$$\hat{\beta}_0\sum w_i x_i + \hat{\beta}_1\sum w_i x_i^{\,2} = \sum w_i x_i y_i$$

and these equations are solved as follows

$$\hat{\beta}_1 = \frac{\left(\sum w_i x_i y_i\right)\left(\sum w_i\right) - \sum w_i y_i}{\left(\sum w_i\right)\left(\sum w_i x_i^2\right) - \left(\sum w_i x_i\right)^2}$$

$$\hat{\beta}_0 = \frac{\sum w_i y_i}{\sum w_i} - \frac{\sum w_i x_i}{\sum w_i}\hat{\beta}_1 \quad .$$

11-103 $\hat{y} = \bar{y} + r\,\frac{s_y}{s_x}(x - \bar{x})$

$$= \bar{y} + \frac{S_{xy}\sqrt{\sum (y_i - \bar{y})^2}\,(x - \bar{x})}{\sqrt{S_{xx}S_{yy}}\sqrt{\sum (x_i - \bar{x})^2}}$$

$$= \bar{y} + \frac{S_{xy}}{S_{xx}}(x - \bar{x})$$

$$= \bar{y} + \hat{\beta}_1 x - \hat{\beta}_1 \bar{x} = \hat{\beta}_0 + \hat{\beta}_1 x$$

11-104 a) $\frac{\partial}{\beta_1}\sum_{i=1}^{n}(y_i - \beta_0 - \beta_1 x_i)^2 = -2\sum_{i=1}^{n}(y_i - \beta_0 - \beta_1 x_i)x_i$

Upon setting the derivative to zero, we obtain

$$\beta_0 \sum x_i + \beta_1 \sum x_i^{\,2} = \sum x_i y_i$$

Therefore,

$$\hat{\beta}_1 = \frac{\sum x_i y_i - \beta_0 \sum x_i}{\sum x_i^{\,2}} = \frac{\sum x_i (y_i - \beta_0)}{\sum x_i^{\,2}}$$

b) $V(\hat{\beta}_1) = V\left(\frac{\sum x_i \ (Y_i - \beta_0)}{\sum x_i^{\,2}}\right) = \frac{\sum x_i^{\,2}\sigma^2}{[\sum x_i^{\,2}]^2} = \frac{\sigma^2}{\sum x_i^{\,2}}$

c) $\hat{\beta}_1 \pm t_{\alpha/2, n-1}\sqrt{\frac{\hat{\sigma}^2}{\sum x_i^2}}$

This confidence interval is shorter because $\sum x_i^{\,2} \geq \sum (x_i - \bar{x})^2$. Also, the *t* value based on *n* – 1 degrees of freedom is slightly smaller than the corresponding *t* value based on *n* – 2 degrees of freedom.