CHAPTER 11

Section 11-2

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

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

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

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

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

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

$\hat{y} = 48.012962 - 2.3298017(4.3) = 37.99$

c) $\hat{y} = 48.012962 - 2.3298017(3.7) = 39.39$

d) $e = y - \hat{y} = 46.1 - 39.39 = 6.71$

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

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

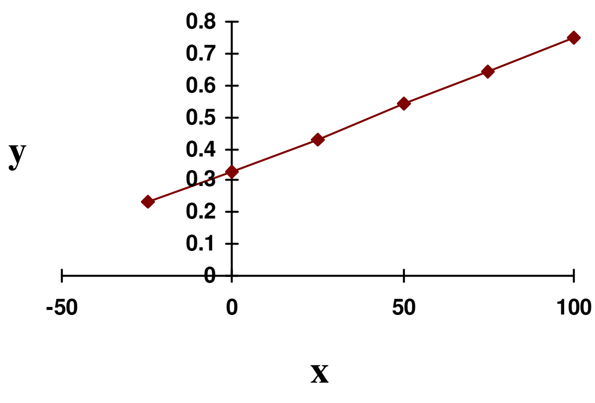
$S_{xy} = 1083.67 - \frac{(1478)(12.75)}{20} = 141.445$

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

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

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

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



b) $\hat{y} = 0.32999 + 0.00416(85) = 0.6836$

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

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

11-3. a) $\hat{y} = 0.3299892 + 0.0041612(\frac{9}{5}x + 32)$

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

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

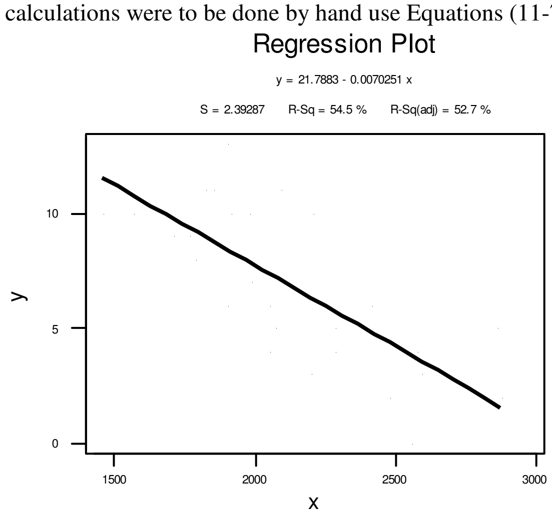
b) $\hat{\beta}_1 = 0.00749$

11-4. a)

```
Regression Analysis - Linear model: Y = a+bX
Dependent variable: Games                    Independent variable: Yards
--------------------------------------------------------------------------------
                               Standard             T              Prob.
Parameter        Estimate        Error            Value            Level
Intercept         21.7883       2.69623           8.081           .00000
Slope          -7.0251E-3     1.25965E-3       -5.57703           .00001
--------------------------------------------------------------------------------
                           Analysis of Variance
Source              Sum of Squares     Df  Mean Square    F-Ratio  Prob. Level
Model                    178.09231      1    178.09231    31.1032       .00001
Residual                 148.87197     26      5.72585
--------------------------------------------------------------------------------
Total (Corr.)            326.96429     27
Correlation Coefficient = -0.738027          R-squared =  54.47 percent
Stnd. Error of Est. = 2.39287
```

$\hat{\sigma}^2 = 5.7258$

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



b) $\hat{y} = 21.7883 - 0.0070251(1800) = 9.143$

c) $-0.0070251(-100) = 0.70251$ games won.

d) $\dfrac{1}{0.0070251} = 142.35$ yds decrease required.

e) $\hat{y} = 21.7883 - 0.0070251(1917) = 8.321$

$e = y - \hat{y}$

$= 10 - 8.321 = 1.679$

11-5. 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).

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

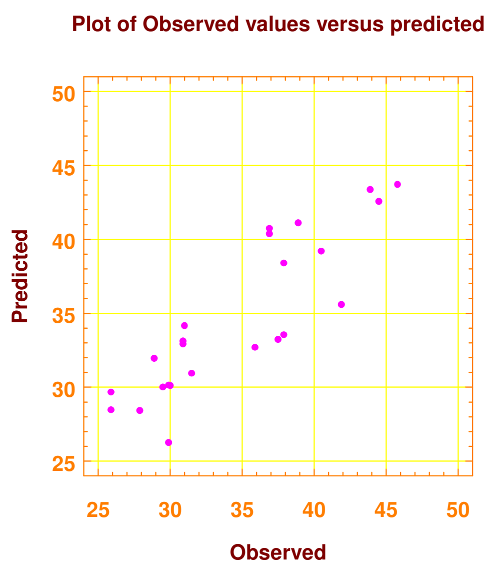
b) $\hat{y} = 13.3202 + 3.32437(7.5) = 38.253$

c) $\hat{y} = 13.3202 + 3.32437(5.8980) = 32.9273$

$$\hat{y} = 32.9273$$

$$e = y - \hat{y} = 30.9 - 32.9273 = -2.0273$$

d) All the points would lie along the 45% axis 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.



11-6. a)

```
Regression Analysis - Linear model: Y = a+bX
Dependent variable: Usage              Independent variable: Temperature
--------------------------------------------------------------------------------
                               Standard            T              Prob.
Parameter       Estimate        Error            Value            Level
Intercept        -6.3355        1.66765        -3.79906          .00349
Slope            9.20836      0.0337744         272.643          .00000
--------------------------------------------------------------------------------
                           Analysis of Variance
Source             Sum of Squares     Df  Mean Square    F-Ratio  Prob. Level
Model                   280583.12      1    280583.12    74334.4       .00000
Residual                37.746089     10     3.774609
--------------------------------------------------------------------------------
Total (Corr.)           280620.87     11
Correlation Coefficient = 0.999933          R-squared =  99.99 percent
Stnd. Error of Est. = 1.94284
```

$$\hat{\sigma}^2 = 3.7746$$

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

$$\hat{y} = -6.3355 + 9.20836x$$

b) $\hat{y} = -6.3355 + 9.20836(55) = 500.124$

c) If monthly temperature increases by 1°F, $\hat{y}$ increases by 9.20836.

d) $\hat{y} = -6.3355 + 9.20836(47) = 426.458$

$\hat{y} = 426.458$

$$e = y - \hat{y} = 424.84 - 426.458 = -1.618$$

11-7. a)

```
Predictor       Coef       StDev          T        P
Constant      33.535       2.614      12.83    0.000
x           -0.03540     0.01663      -2.13    0.047

S = 3.660       R-Sq = 20.1%     R-Sq(adj) = 15.7%

Analysis of Variance
Source       DF          SS          MS         F        P
Regression    1       60.69       60.69      4.53    0.047
Error        18      241.06       13.39
Total        19      301.75
```

$$\hat{\sigma}^2 = 13.392$$

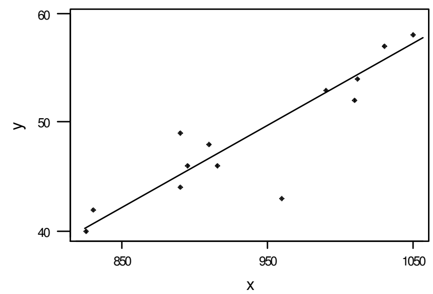
$$\hat{y} = 33.5348 - 0.0353971x$$

b) $\hat{y} = 33.5348 - 0.0353971(150) = 28.226$

c) $\hat{y} = 29.4995$

$$e = y - \hat{y} = 31.0 - 29.4995 = 1.50048$$

11-8. 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			

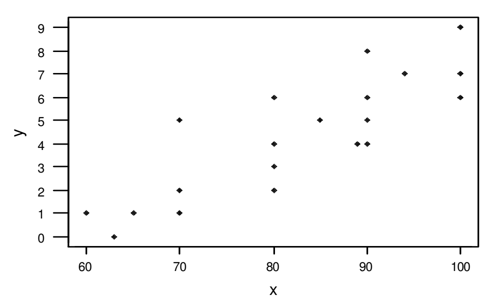
$\hat{\sigma}^2 = 7.3212$

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

b) $\hat{y} = 46.6041$ $\quad e = y - \hat{y} = 1.39592$

c) $\hat{y} = -16.5093 + 0.0693554(950) = 49.38$

11-9. a)



Yes, a linear regression would seem appropriate, but one or two points appear to 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%

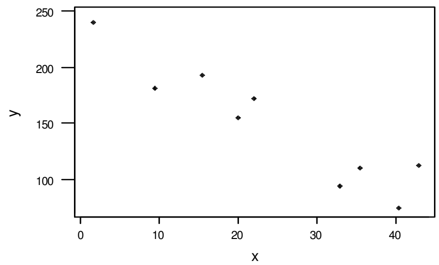
Analysis of Variance

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) $\hat{y} = 4.68265$ at $x = 85$

11-10. a)



Yes, a linear regression model appears to be plausable.

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

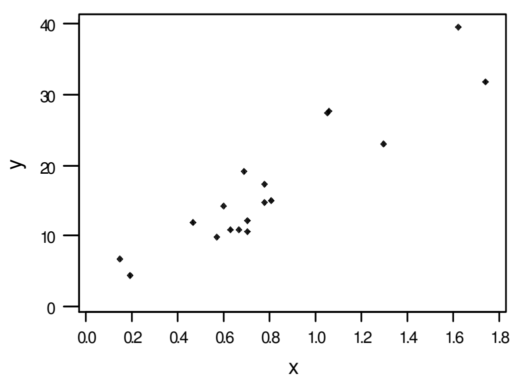
Source	DF	SS	MS	F	P
Regression	1	20329	20329	51.04	0.000
Error	7	2788	398		
Total	8	23117			

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

c) $\hat{y} = 234.071 - 3.50856(30) = 128.814$

d) $\hat{y} = 156.883 \quad e = 15.1175$

11-11. 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			

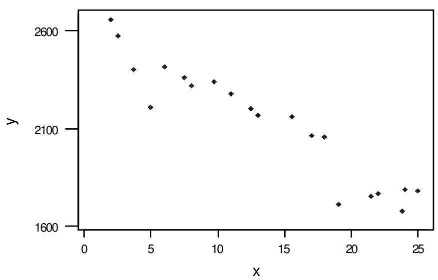
b) $\hat{\sigma}^2 = 13.81$

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

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

d) $\hat{y} = 10.1371 \quad e = 1.6629$

11-12. a)



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

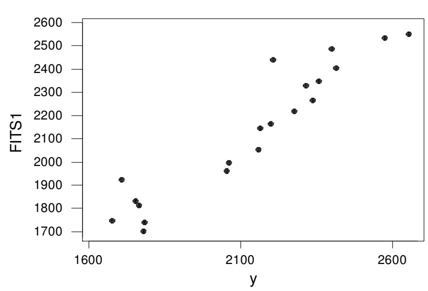
```
Predictor       Coef       StDev          T        P
Constant     2625.39       45.35      57.90    0.000
x            -36.962       2.967     -12.46    0.000

S = 99.05       R-Sq = 89.6%     R-Sq(adj) = 89.0%

Analysis of Variance
Source       DF          SS          MS         F        P
Regression    1     1522819     1522819    155.21    0.000
Error        18      176602        9811
Total        19     1699421
```

b) $\hat{\sigma}^2 = 9811.2$

$\hat{y} = 2625.39 - 36.962x$

c) $\hat{y} = 2625.39 - 36.962(20) = 1886.15$

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



11-13. $\hat{\beta}_0 + \hat{\beta}_1\bar{x} = (\bar{y} - \hat{\beta}_1\bar{x}) + \hat{\beta}_1\bar{x} = \bar{y}$

11-14. a) The slopes of both regression models will be the same, but the intercept will be shifted.

b) $\hat{y} = 2132.41 - 36.9618x$

$$\begin{array}{ccc} \hat{\beta}_0 = 2625.39 & & \hat{\beta}_0^{\ *} = 2132.41 \\ & \text{vs.} & \\ \hat{\beta}_1 = -36.9618 & & \hat{\beta}_1^{\ *} = -36.9618 \end{array}$$

11-15. Let $x_i^{\ *} = x_i - \bar{x}$. Then, the model is $Y_i^* = \beta_0^* + \beta_1^* x_i^* + \varepsilon_i$.

Equations 11-7 and 11-8 can be applied to the new variables using the facts that $\sum_{i=1}^{n} x_i^* = \sum_{i=1}^{n} y_i^* = 0$. Then,

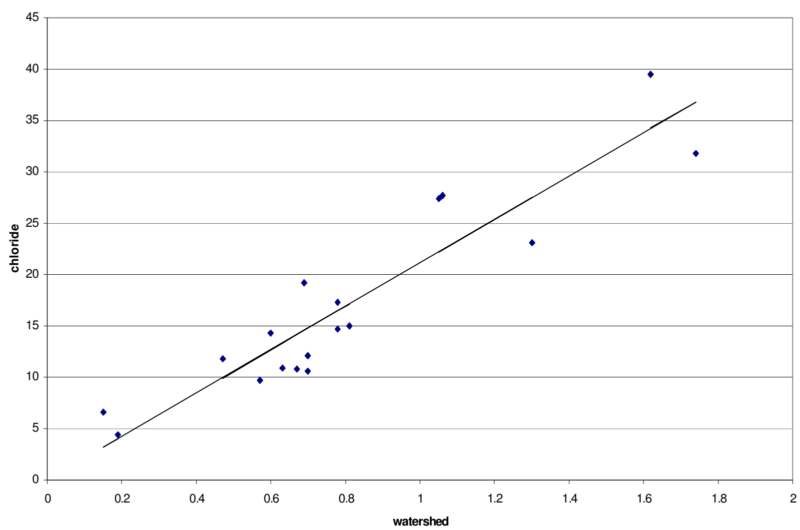
$\hat{\beta}_1^* = \hat{\beta}_1$ and $\hat{\beta}_0^* = 0$.

11-16. 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, $\hat{\beta} = \dfrac{\sum y_i x_i}{\sum x_i^{\ 2}}$.

11-17. $\hat{y} = 21.031461x$. The model seems very appropriate - an even better fit.



Section 11-5

11-18. 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

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

6) Reject H_0 if $f_0 > f_{\alpha,1,12}$ where $f_{0.05,1,12} = 4.75$

7) Using results from Exercise 11-1

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

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

8) Since 74.63 > 4.75 reject H_0 and conclude that compressive strength is significant in predicting intrinsic permeability of concrete at $\alpha = 0.05$. We can therefore conclude model specifies a useful linear relationship between these two variables.

P − value ≅ 0.000002

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

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

11-19. 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

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

6) Reject H_0 if $f_0 > f_{\alpha,1,18}$ where $f_{0.05,1,18} = 4.414$

7) Using the results from Exercise 11-2

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

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

8) Since 73.95 > 4.414, reject H_0 and conclude the model specifies a useful relationship at $\alpha = 0.05$.

$P - \text{value} \cong 0.000001$

b) $se(\hat{\beta}_1) = \sqrt{\frac{\hat{\sigma}^2}{S_{xx}}} = \sqrt{\frac{.00796}{33991.6}} = 4.8391x10^{-4}$

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

11-20. a) Refer to ANOVA table of Exercise 11-4.

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 $f_0 > f_{\alpha,1,26}$ where $f_{0.01,1,26} = 7.721$

7) Using the results of Exercise 10-4

$$f_0 = \frac{MS_R}{MS_E} = 31.1032$$

8) Since $31.1032 > 7.721$ reject H_0 and conclude the model is useful at $\alpha = 0.01$. $P - \text{value} = 0.000007$

b) $se(\hat{\beta}_1) = \sqrt{\frac{\hat{\sigma}^2}{S_{xx}}} = \sqrt{\frac{5.7257}{3608611\ .43}} = .001259$

$$se(\hat{\beta}_0) = \sqrt{\hat{\sigma}^2\left[\frac{1}{n} + \frac{\bar{x}}{S_{xx}}\right]} = \sqrt{5.7257\left[\frac{1}{28} + \frac{2110.13^2}{3608611\ .43}\right]} = 2.6962$$

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

2) $H_0 : \beta_1 = -0.01$

3) $H_1 : \beta_1 \neq -0.01$

4) $\alpha = 0.01$

5) The test statistic is $t_0 = \frac{\hat{\beta}_1 + .01}{se(\hat{\beta}_1)}$

6) Reject H_0 if $t_0 < -t_{\alpha/2,n-2}$ where $-t_{0.005,26} = -2.78$ or $t_0 > t_{0.005,26} = 2.78$

7) Using the results from Exercise 10-4

$$t_0 = \frac{-0.0070251 + .01}{0.00125965} = 2.3618$$

8) Since $2.3618 < 2.78$ do not reject H_0 and conclude the intercept is not zero at $\alpha = 0.01$.

11-21. Refer to ANOVA of Exercise 11-5

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) $\alpha = 0.05$, 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 $-t_{0.025,22} = -2.074$ or $t_0 > t_{0.025,22} = 2.074$

7) Using the results from Exercise 11-5

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

8) Since $8.518 > 2.074$ reject H_0 and conclude the model is useful $\alpha = 0.05$.

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) $\alpha = 0.05$

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 $f_0 > f_{\alpha,1,22}$ where $f_{0.01,1,22} = 4.303$

7) Using the results from Exercise 10-5

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

8) Since $72.5563 > 4.303$, reject H_0 and conclude the model is useful at a significance $\alpha = 0.05$.

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) $\alpha = 0.05$, 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 $-t_{0.025,22} = -2.074$ or $t_0 > t_{0.025,22} = 2.074$

7) Using the results from Exercise 11-5

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

8) Since $5.179 > 2.074$ reject H_0 and conclude the intercept is not zero at $\alpha = 0.05$.

11-22. Refer to ANOVA for Exercise 10-6

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

6) Reject H_0 if $f_0 > f_{\alpha,1,22}$ where $f_{0.01,1,10} = 10.049$

7) Using the results from Exercise 10-6

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

8) Since 74334.4 > 10.049, reject H_0 and conclude the model is useful $\alpha = 0.01$. 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) $\alpha = 0.01$

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 $-t_{0.005,10} = -3.17$ or $t_0 > t_{0.005,10} = 3.17$

7) Using the results from Exercise 10-6

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

8) Since −23.37 < −3.17 reject H_0 and conclude the slope is not 10 at $\alpha = 0.01$. 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.

11-23. Refer to ANOVA table of Exercise 11-7

a) $H_0: \beta_1 = 0$

$H_1: \beta_1 \neq 0$ $\alpha = 0.01$

$$f_0 = 4.53158$$
$$f_{0.01,1,18} = 8.285$$
$$f_0 \not> f_{\alpha,1,18}$$

Therefore, do not reject H_0. P-value = 0.04734. Insufficient evidence to conclude that the model is a useful relationship.

b) $se(\hat{\beta}_1) = 0.0166281$

$se(\hat{\beta}_0) = 2.61396$

c) $H_0 : \beta_1 = -0.05$

$H_1 : \beta_1 < -0.05$

$\alpha = 0.01$

$$t_0 = \frac{-0.0354 - (-0.05)}{0.0166281} = 0.87803$$

$t_{.01,18} = 2.552$

$t_0 \not< -t_{\alpha,18}$

Therefore, do not rejectH_0. P-value = 0.804251. Insufficient evidence to conclude that β_1 is ≥ -0.05.

d) $H_0 : \beta_0 = 0$ $H_1 : \beta_0 \neq 0$ $\alpha = 0.01$

$t_0 = 12.8291$

$t_{.005,18} = 2.878$

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

Therefore, reject H_0. P-value ≅ 0

11-24. Refer to ANOVA of Exercise 11-8

a) $H_0 : \beta_1 = 0$

$H_1 : \beta_1 \neq 0$

$\alpha = 0.05$

$f_0 = 44.0279$

$f_{.05,1,11} = 4.84$

$f_0 > f_{\alpha,1,11}$

Therefore, rejectH_0. P-value = 0.00004.

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, do not rejectH_0. P-value = 0.12166.

11-25. Refer to ANOVA of Exercise 11-9

a) $H_0 : \beta_1 = 0$

$H_1 : \beta_1 \neq 0$

$\alpha = 0.05$

$f_0 = 53.50$

$f_{.05,1,18} = 4.414$

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

Therefore, reject H_0. P-value = 0.000009.

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 = 0.000078.

11-26. Refer to ANOVA of Exercise 11-11

a) $H_0 : \beta_1 = 0$

$H_1 : \beta_1 \neq 0$

$\alpha = 0.01$

$f_0 = 92.224$

$f_{.01,1,16} = 8.531$

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

Therefore, reject H_0.

b) P-value < 0.00001

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

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

d) $H_0 : \beta_0 = 0$

$H_1 : \beta_0 \neq 0$

$\alpha = 0.01$

$t_0 = 0.243$

$t_{.005,16} = 2.921$

$t_0 \ngtr t_{\alpha/2,16}$

Therefore, do not rejectH_0. Conclude, Yes, the intercept should be removed.

11-27. Refer to ANOVA of Exercise 11-12

a) $H_0 : \beta_1 = 0$

$H_1 : \beta_1 \neq 0$

$\alpha = 0.01$

$f_0 = 155.2$

$f_{.01,1,18} = 8.285$

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

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, do not reject H_0. P-value = 0.0153(2) = 0.0306.

d) $H_0 : \beta_0 = 0$

$H_1 : \beta_0 \neq 0$

$\alpha = 0.01$

$t_0 = 57.8957$

$t_{.005,18} = 2.878$

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

11-28. $t_0 = \frac{\hat{\beta}_1}{\sqrt{\hat{\sigma}^2 / S_{xx}}}$ 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, $t_0^* = \frac{b\hat{\beta}_1 / a}{\sqrt{(b\hat{\sigma})^2 / a^2 S_{xx}}} = t_0$.

11-29. 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 11-17, $\hat{\beta} = 21.031461, \hat{\sigma} = 3.611768,$ 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.

11-30. $d = \frac{|-0.01 - (-0.005)|}{2.4\sqrt{\frac{27}{3608611.96}}} = 0.76$, $S_{xx} = 3608611.96$.

Assume α = 0.05, from Chart VI and interpolating between the curves for n = 20 and n = 30, $\beta \cong 0.05$.

Sections 11-6 and 11-7

11-31. $t_{\alpha/2,n-2} = t_{0.025,12} = 2.179$

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 $x_0 = 2.5$.

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

d) 95% on prediction interval when $x_0 = 2.5$.

$$\hat{y}_0 \pm t_{.025,12}\sqrt{\hat{\sigma}^2\left(1 + \tfrac{1}{n} + \tfrac{(x_0-\bar{x})^2}{S_{xx}}\right)}$$
$$42.1885 \pm 2.179\sqrt{1.844\left(1 + \tfrac{1}{14} + \tfrac{(2.5-3.0714)^2}{25.348571}\right)}$$
$$42.1885 \pm 2.179(1.4056)$$
$$39.1257 \le y_0 \le 45.2513$$

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

11-32. $t_{\alpha/2,n-2} = t_{0.005,18} = 2.878$

a) $\hat{\beta}_1 \pm \left(t_{0.005,18}\right) se(\hat{\beta}_1)$.

$0.0041612 \pm (2.878)(0.000484)$

$0.0027682 \leq \beta_1 \leq 0.0055542$

b) $\hat{\beta}_0 \pm \left(t_{0.005,18}\right) se(\hat{\beta}_0)$.

$0.3299892 \pm (2.878)(0.04095)$

$0.212135 \leq \beta_0 \leq 0.447843$

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

$\hat{\mu}_{Y|x_0} = 0.683689$

$\hat{\mu}_{Y|x_0} \pm t_{.005,18}\sqrt{\hat{\sigma}^2(\frac{1}{n} + \frac{(x_0-\bar{x})^2}{S_{xx}})}$

$0.683689 \pm (2.878)\sqrt{0.00796(\frac{1}{20} + \frac{(85-73.9)^2}{33991.6})}$

0.683689 ± 0.0594607

$0.6242283 \leq \hat{\mu}_{Y|x_0} \leq 0.7431497$

d) 99% prediction interval when $x_0 = 90^\circ F$.

$\hat{y}_0 = 0.7044949$

$\hat{y}_0 \pm t_{.005,18}\sqrt{\hat{\sigma}^2(1 + \frac{1}{n} + \frac{(x_0-\bar{x})^2}{S_{xx}})}$

$0.7044949 \pm 2.878\sqrt{0.00796(1 + \frac{1}{20} + \frac{(90-73.9)^2}{33991.6})}$

0.7044949 ± 0.2640665

$0.4404284 \leq y_0 \leq 0.9685614$

Note for Problems 11-33 through 11-35: These computer printouts were obtained from Statgraphics. For Minitab users, the standard errors are obtained from the Regression subroutine.

11-33.

```
              95 percent confidence intervals for coefficient estimates
--------------------------------------------------------------------------------
                         Estimate  Standard error     Lower Limit      Upper Limit
CONSTANT                  21.7883         2.69623         16.2448          27.3318
Yards                    -0.00703         0.00126        -0.00961         -0.00444
--------------------------------------------------------------------------------
```

a) $-0.00961 \leq \beta_1 \leq -0.00444$.

b) $16.2448 \leq \beta_0 \leq 27.3318$.

c) $9.143 \pm (2.056)\sqrt{5.72585(\frac{1}{28} + \frac{(1800-2110.14)^2}{3608325.5})}$

9.143 ± 1.2287

$7.9143 \leq \hat{\mu}_{Y|x_0} \leq 10.3717$

d) $9.143 \pm (2.056)\sqrt{5.72585(1 + \frac{1}{28} + \frac{(1800-2110.14)^2}{3608325.5})}$

9.143 ± 5.0709

$4.0721 \leq y_0 \leq 14.2139$

11-34. 95 percent confidence intervals for coefficient estimates

	Estimate	Standard error	Lower Limit	Upper Limit
CONSTANT	13.3202	2.57172	7.98547	18.6549
Taxes	3.32437	0.39028	2.51479	4.13395

a) $2.51479 \le \beta_1 \le 4.13395$.

b) $7.98547 \le \beta_0 \le 18.6549$.

c) $38.253 \pm (2.074)\sqrt{8.76775(\frac{1}{24} + \frac{(7.5-6.40492)^2}{57.563139})}$

38.253 ± 1.5353

$36.7177 \le \hat{\mu}_{Y|x_0} \le 39.7883$

d) $38.253 \pm (2.074)\sqrt{8.76775(1 + \frac{1}{24} + \frac{(7.5-6.40492)^2}{57.563139})}$

38.253 ± 6.3302

$31.9228 \le y_0 \le 44.5832$

11-35. 99 percent confidence intervals for coefficient estimates

	Estimate	Standard error	Lower Limit	Upper Limit
CONSTANT	-6.33550	1.66765	-11.6219	-1.05011
Temperature	9.20836	0.03377	9.10130	9.93154

a) $9.10130 \le \beta_1 \le 9.31543$

b) $-11.6219 \le \beta_0 \le -1.04911$

c) $500.124 \pm (2.228)\sqrt{3.774609(\frac{1}{12} + \frac{(55-46.5)^2}{3308.9994})}$

500.124 ± 1.4037586

$498.72024 \le \hat{\mu}_{Y|x_0} \le 501.52776$

d) $500.124 \pm (2.228)\sqrt{3.774609(1 + \frac{1}{12} + \frac{(55-46.5)^2}{3308.9994})}$

500.124 ± 4.5505644

$495.57344 \le y_0 \le 504.67456$

It is wider because the prediction interval includes error for both the fitted model and from that associated with the future observation.

11-36. a) $-0.07034 \le \beta_1 \le -0.00045$

b) $28.0417 \le \beta_0 \le 39.027$

c) $28.225 \pm (2.101)\sqrt{13.39232\ (\frac{1}{20} + \frac{(150-149.3)^2}{48436.256})}$

28.225 ± 1.7194236

$26.5406 \le \mu_{y|x_0} \le 29.9794$

d) $28.225 \pm (2.101)\sqrt{13.39232\left(1 + \frac{1}{20} + \frac{(150-149.3)^2}{48436.256}\right)}$
28.225 ± 7.87863
$20.3814 \le y_0 \le 36.1386$

11-37. a) $0.03689 \le \beta_1 \le 0.10183$
b) $-47.0877 \le \beta_0 \le 14.0691$
c) $46.6041 \pm (3.106)\sqrt{7.324951\left(\frac{1}{13} + \frac{(910-939)^2}{67045.97}\right)}$
46.6041 ± 2.514401
$44.0897 \le \mu_{y|x_0} \le 49.1185)$
d) $46.6041 \pm (3.106)\sqrt{7.324951\left(1 + \frac{1}{13} + \frac{(910-939)^2}{67045.97}\right)}$
46.6041 ± 8.779266
$37.8298 \le y_0 \le 55.3784$

11-38. a) $0.11756 \le \beta_1 \le 0.22541$
b) $-14.3002 \le \beta_0 \le -5.32598$
c) $4.76301 \pm (2.101)\sqrt{1.982231\left(\frac{1}{20} + \frac{(85-82.3)^2}{3010.2111}\right)}$
4.76301 ± 0.6772655
$4.0857 \le \mu_{y|x_0} \le 5.4403$
d) $4.76301 \pm (2.101)\sqrt{1.982231\left(1 + \frac{1}{20} + \frac{(85-82.3)^2}{3010.2111}\right)}$
4.76301 ± 3.0345765
$1.7284 \le y_0 \le 7.7976$

11-39. a) $201.552 \le \beta_1 \le 266.590$
b) $-4.67015 \le \beta_0 \le -2.34696$
c) $128.814 \pm (2.365)\sqrt{398.2804\left(\frac{1}{9} + \frac{(30-24.5)^2}{1651.4214}\right)}$
128.814 ± 16.980124
$111.8339 \le \mu_{y|x_0} \le 145.7941$

11-40. a) $14.3107 \le \beta_1 \le 26.8239$
b) $-5.18501 \le \beta_0 \le 6.12594$
c) $21.038 \pm (2.921)\sqrt{13.8092\left(\frac{1}{18} + \frac{(1-0.806111)^2}{3.01062}\right)}$
21.038 ± 2.8314277
$18.2066 \le \mu_{y|x_0} \le 23.8694$
d) $21.038 \pm (2.921)\sqrt{13.8092\left(1 + \frac{1}{18} + \frac{(1-0.806111)^2}{3.01062}\right)}$
21.038 ± 11.217861
$9.8201 \le y_0 \le 32.2559$

11-41. a) $-43.1964 \le \beta_1 \le -30.7272$
b) $2530.09 \le \beta_0 \le 2720.68$
c) $1886.154 \pm (2.101)\sqrt{9811.21\left(\frac{1}{20} + \frac{(20-13.3375)^2}{1114.6618}\right)}$
1886.154 ± 62.370688
$1823.7833 \le \mu_{y|x_0} \le 1948.5247$

d) $1886.154 \pm (2.101)\sqrt{9811.21(1+\frac{1}{20}+\frac{(20-13.3375)^2}{1114.6618})}$

1886.154 ± 217.25275

$1668.9013 \le y_0 \le 2103.4067$

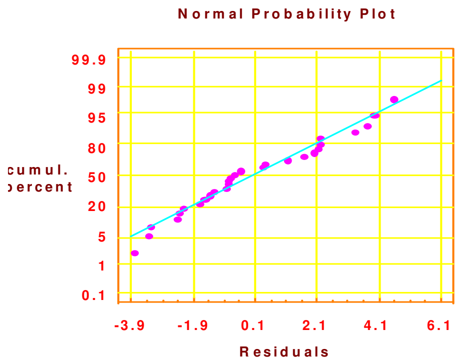
Section 11-7

11-42. Use the results of Exercise 11-4 to answer the following questions.

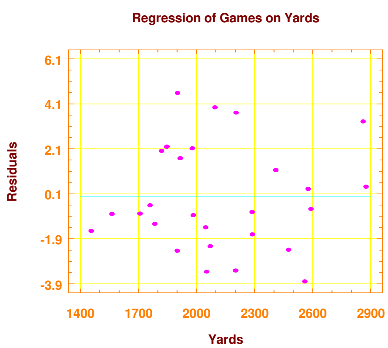
a) $R^2 = 0.544684$; The proportion of variability explained by the model.

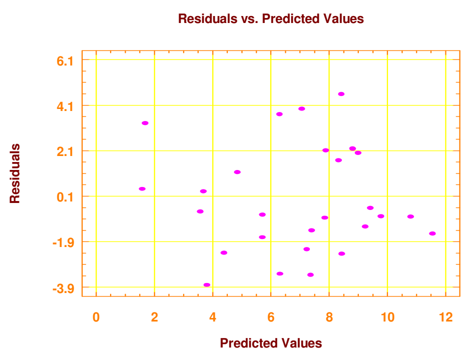
$$R^2_{Adj} = 1 - \frac{148.87197/26}{326.96429/27} = 1 - 0.473 = 0.527$$

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



c) Since the residuals plots appear to be random, the plots do not include any serious model inadequacies.



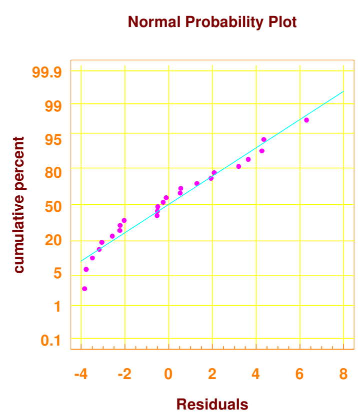


11-43. Use the Results of exercise 11-5 to answer the following questions.

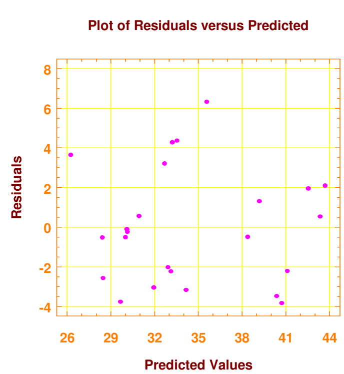
a)

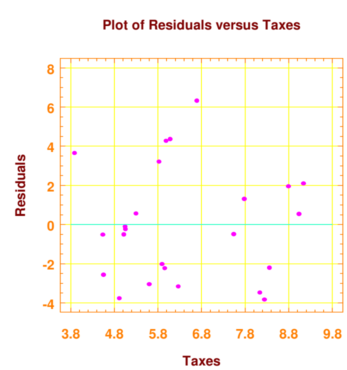
SalePrice	Taxes	Predicted	Residuals
25.9	4.9176	29.6681073	-3.76810726
29.5	5.0208	30.0111824	-0.51118237
27.9	4.5429	28.4224654	-0.52246536
25.9	4.5573	28.4703363	-2.57033630
29.9	5.0597	30.1405004	-0.24050041
29.9	3.8910	26.2553078	3.64469225
30.9	5.8980	32.9273208	-2.02732082
28.9	5.6039	31.9496232	-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.

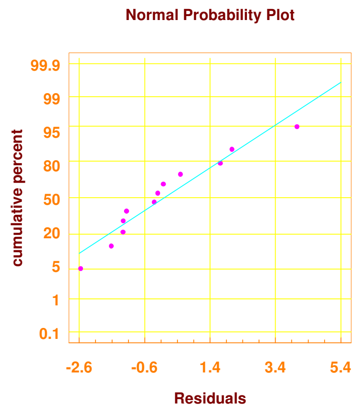




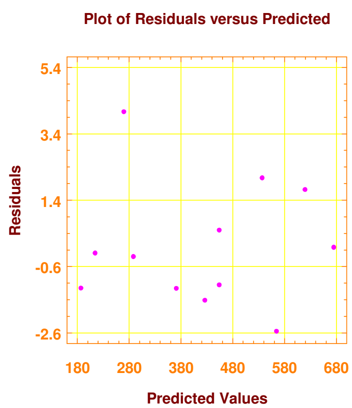
d) $R^2 \equiv 76.73\%$;

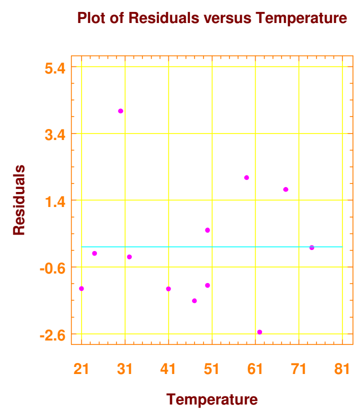
11-44. Use the results of Exercise 11-6 to answer the following questions

a) $R^2 = 99.986\%$; The proportion of variability explained by the model.
b) Yes, normality seems to be satisfied since the data appear to fall along the straight line.

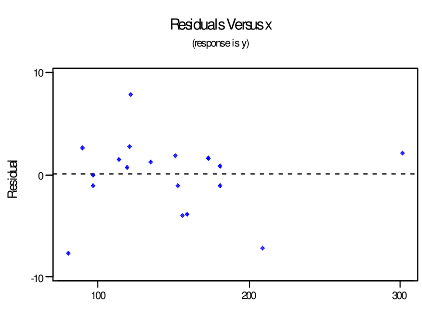


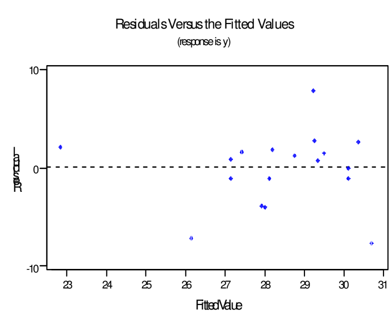
c) There might be lower variance at the middle settings of x. However, this data does not indicate a serious departure from the assumptions.



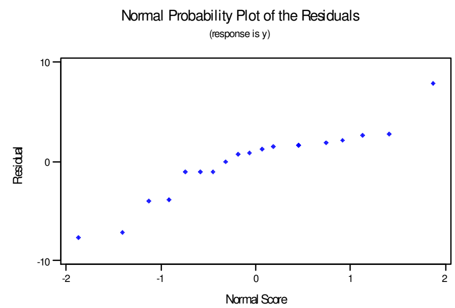


11-45. a) $R^2 = 20.1121\%$
b) These plots indicate presence of outliers, but no real problem with assumptions.

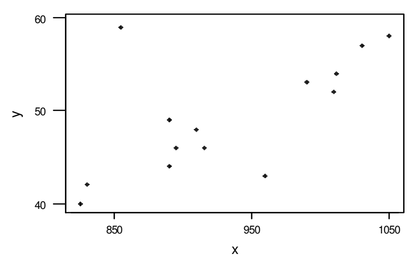




c) The normality assumption appears marginal.



11-46. a)



$$\hat{y} = 0.677559 + 0.0521753x$$

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

$f_0 = 7.9384$

$f_{.05,1,12} = 4.75$

$f_0 > f_{\alpha,1,12}$

Reject H_o.

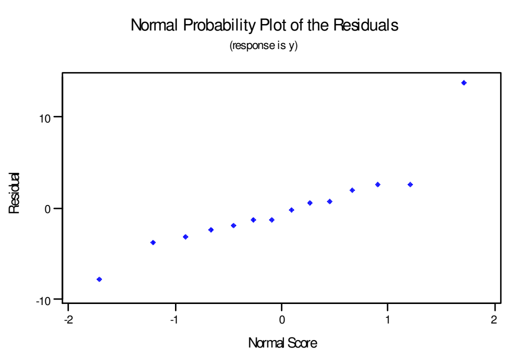
c) $\hat{\sigma}^2 = 25.23842$

d) $\hat{\sigma}_{orig}^{\ 2} = 7.324951$

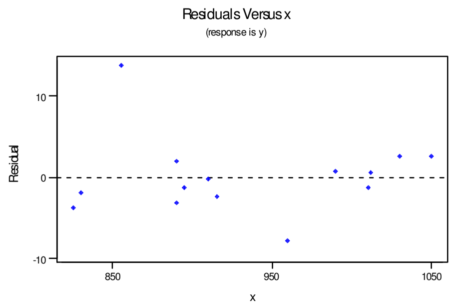
The new estimate is larger because the new point added additional variance not accounted for by the model.

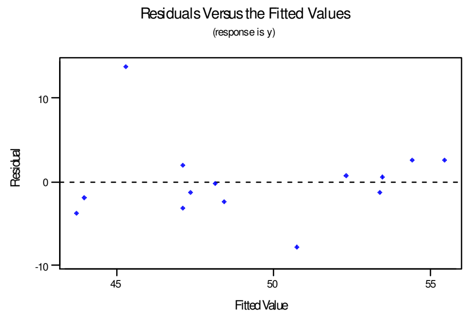
e) Yes, e_{14} is especially large compared to the other residuals.

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



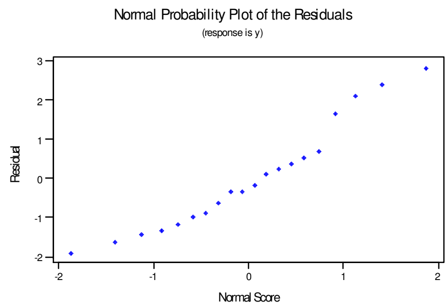
g) Constant variance assumption appears valid except for the added point.



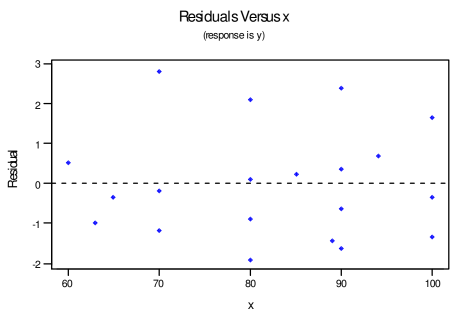


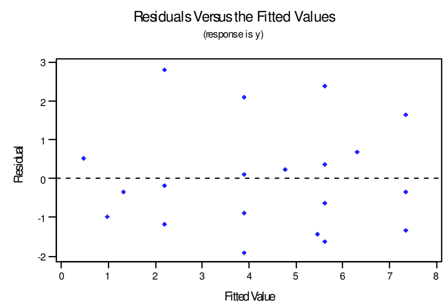
11-47. a) $R^2 = 71.27\%$

b) No major departure from normality assumptions.



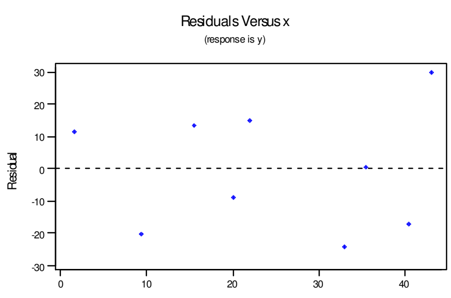
c) Assumption of constant variance appears reasonable.

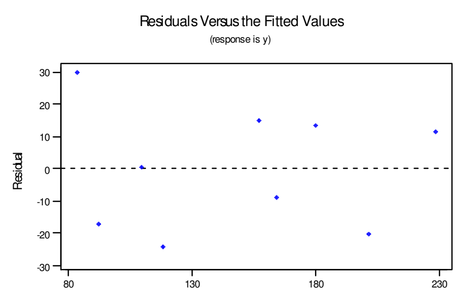




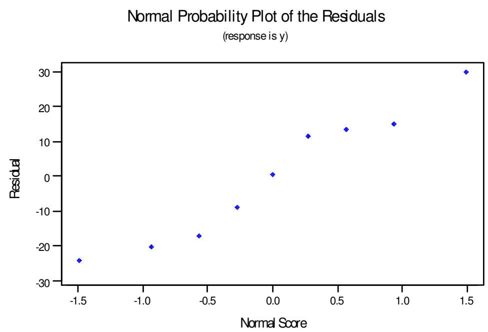
11-48. a) $R^2 = 0.879397$

b) No departures from constant variance are noted.



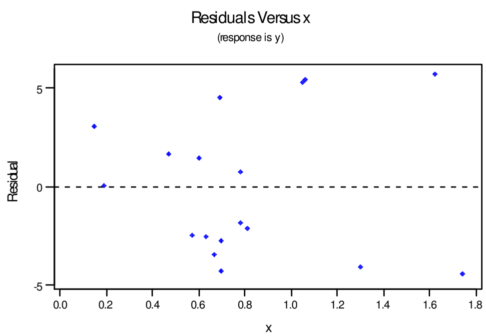


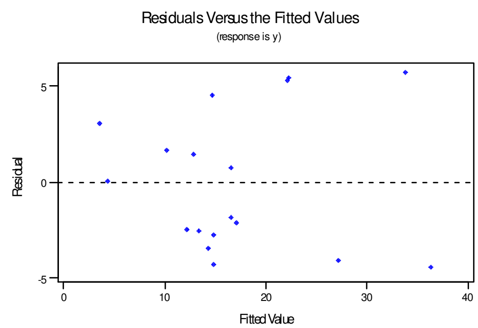
c) Normality assumption appears reasonable.



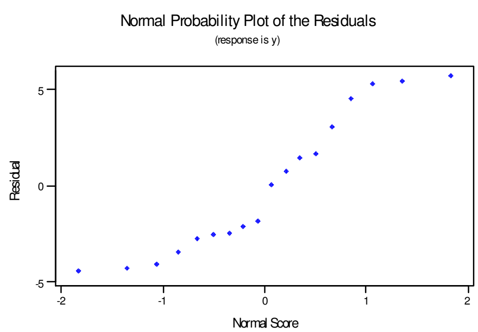
11-49. a) $R^2 = 85.22\%$

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



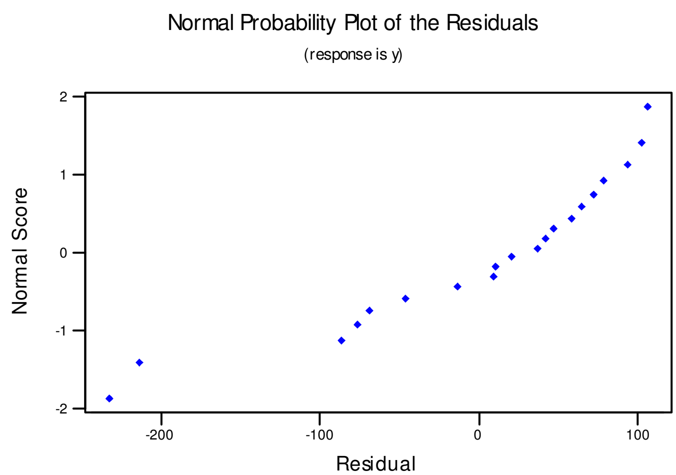


c) Normality assumption may be questionable. There is some " bending" away from a straight line in the tails of the normal probability plot.



11-50. a) $R^2 = 0.896081$ 89% of the variability is explained by the model.

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



c) $R^2_{new\ \ model} = 0.9573$

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

d) $\hat{\sigma}^2_{old\ \ model} = 9811.21$

$\hat{\sigma}^2_{new\ \ model} = 4022.93$

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

11-51. Using $R^2 = 1 - \frac{SS_E}{S_{yy}}$, $\quad F_0 = \dfrac{(n-2)(1-\frac{SS_E}{S_{yy}})}{\frac{SS_E}{S_{yy}}} = \dfrac{S_{yy} - SS_E}{\frac{SS_E}{n-2}} = \dfrac{S_{yy} - SS_E}{\hat{\sigma}^2}$

Also,

$$\begin{aligned} SS_E &= \sum (y_i - \hat{\beta}_0 - \hat{\beta}_1 x_i)^2 \\ &= \sum (y_i - \bar{y} - \hat{\beta}_1 (x_i - \bar{x}))^2 \\ &= \sum (y_i - \bar{y}) + \hat{\beta}_1{}^2 \sum (x_i - \bar{x})^2 - 2\hat{\beta}_1 \sum (y_i - \bar{y})(x_i - \bar{x}) \\ &= \sum (y_i - \bar{y})^2 - \hat{\beta}_1{}^2 \sum (x_i - \bar{x})^2 \end{aligned}$$

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

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 usually t-test that compares $|t_0|$ to $t_{\alpha/2,n-2}$ is equivalent to comparing $f_0 = t_0^2$ to

$f_{\alpha,1,n-2} = t_{\alpha/2,n-2}$.

11-52. a) $f_0 = \dfrac{0.9(23)}{1-0.9} = 207$. Reject $H_0 : \beta_1 = 0$.

b) Because $f_{.05,1,23} = 4.28$, H_0 is rejected if $\dfrac{23R^2}{1-R^2} > 4.28$.

That is, H_0 is rejected if

$$23R^2 > 4.28(1-R^2)$$
$$27.28R^2 > 4.28$$
$$R^2 > 0.157$$

11-53. Yes, the larger residuals are easier to identify.

1.10269	-0.75866	-0.14376	0.66992
-2.49758	-2.25949	0.50867	0.46158
0.10242	0.61161	0.21046	-0.94548
0.87051	0.74766	-0.50425	0.97781
0.11467	0.38479	1.13530	-0.82398

11-54. 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,

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

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 approximate 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.

Section 11-9

11-55. a) $\hat{y} = -0.0280411 + 0.990987x$

b) $H_0 : \beta_1 = 0$

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

$f_0 = 79.838$

$f_{.05,1,18} = 4.41$

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

Reject H_0 .

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

d) $H_0 : \rho = 0$

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

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

$t_{.025,18} = 2.101$

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

Reject H_0.

e) $H_0 : \rho = 0.5$

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

$z_0 = 3.879$

$z_{.025} = 1.96$

$z_0 > z_{\alpha/2}$

Reject H_0.

f) $\tanh(\text{arcta nh } 0.90334 - \frac{z_{.025}}{\sqrt{17}}) \le \rho \le \tanh(\text{arcta nh } 0.90334 + \frac{z_{.025}}{\sqrt{17}})$ where $z_{.025} = 1.96$.

$0.7677 \le \rho \le 0.9615$.

11-56. a) $\hat{y} = 69.1044 + 0.419415\ x$

b) $H_0 : \beta_1 = 0$

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

$f_0 = 35.744$

$f_{.05,1,24} = 4.260$

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

Reject H_0.

c) $r = 0.77349$

d) $H_0 : \rho = 0$

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

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

$t_{.025,24} = 2.064$

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

Reject H_0.

e) $H_0 : \rho = 0.6$

$H_1 : \rho \neq 0.6$ $\alpha = 0.05$

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

$z_{.025} = 1.96$

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

Do not reject H_0.

f) $\tanh(\text{arcta nh } 0.77349 - \frac{z_{.025}}{\sqrt{23}}) \le \rho \le \tanh(\text{arcta nh } 0.77349 + \frac{z_{.025}}{\sqrt{23}})$ where $z_{.025} = 1.96$.

$0.5513 \le \rho \le 0.8932$.

11-57. a) r = -0.738027

b) $H_0 : \rho = 0$

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

$t_0 = \frac{-0.738027\sqrt{26}}{\sqrt{1-0.5447}} = -5.577$

$t_{.025,26} = 2.056$

$|t_0| > t_{\alpha/2,26}$

Reject H_0 . P-value = (3.69E-6)(2) = 7.38E-6

c) $\tanh(\text{arctanh} - 0.738 - \frac{z_{.025}}{\sqrt{25}}) \leq \rho \leq \tanh(\text{arctanh} - 0.738 + \frac{z_{.025}}{\sqrt{25}})$

where $z_{.025} = 1.96$. $-0.871 \leq \rho \leq -0.504$.

d) $H_0 : \rho = -0.7$

$H_1 : \rho \neq -0.7$ $\alpha = 0.05$

$z_0 = (\text{arctanh} -0.738 - \text{arctanh} -0.7)(25)^{1/2} = -0.394$

$z_{.025} = 1.96$

$|z_0| < z_{\alpha/2}$

Do not reject H_0 . P-value = (0.3468)(2) = 0.6936

11-58 $R = \hat{\beta}_1 \left(\frac{S_{xx}}{S_{yy}} \right)^{1/2}$ and $1 - R^2 = \frac{SS_E}{S_{yy}}$.

Therefore, $T_0 = \frac{R\sqrt{n-2}}{\sqrt{1-R^2}} = \frac{\hat{\beta}_1 \left(\frac{S_{xx}}{S_{yy}} \right)^{1/2} \sqrt{n-2}}{\left(\frac{SS_E}{S_{yy}} \right)^{1/2}} = \frac{\hat{\beta}_1}{\sqrt{\frac{\hat{\sigma}^2}{S_{xx}}}}$ where $\hat{\sigma}^2 = \frac{SS_E}{n-2}$.

11-59 n = 50 r = 0.62

a) $H_0 : \rho = 0$

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

$t_0 = \frac{r\sqrt{n-2}}{\sqrt{1-r^2}} = \frac{0.62\sqrt{48}}{\sqrt{1-(0.62)^2}} = 5.475$

$t_{.005,48} = 2.682$

$t_0 > t_{0.005,48}$

Reject H_0 . P-value $\cong 0$

b) $\tanh(\text{arctanh}\, 0.62 - \frac{z_{.005}}{\sqrt{47}}) \leq \rho \leq \tanh(\text{arctanh}\, 0.62 + \frac{z_{.005}}{\sqrt{47}})$

where $z_{.005} = 2.575$. $0.3358 \leq \rho \leq 0.8007$.

c) Yes.

11-60. n = 10000, r = 0.02

a) $H_0 : \rho = 0$

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

$t_0 = \frac{r\sqrt{n-2}}{\sqrt{1-r^2}} = \frac{0.02\sqrt{10000}}{\sqrt{1-(0.02)^2}} = 2.0002$

$t_{.025,9998} = 1.96$

$t_0 > t_{\alpha/2,9998}$

Reject H_0 . P-value = 2(0.02274) = 0.04548

b) Since the sample size is so large, the standard error is very small. Therefore, very small differences are found to be "statistically" significant. However, the practical significance is minimal since r = 0.02 is essentially zero.

11-61. a) r = 0.933203

b) $H_0 : \rho = 0$

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

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

$t_{.025,15} = 2.131$

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

Reject H_0.

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

$H_0 : \beta_1 = 0$

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

$f_0 = 101.16$

$f_{.05,1,15} = 4.543$

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

Reject H_0. Conclude that the model is significant at $\alpha = 0.05$. This test and the one in part b are identical.

d) $H_0 : \beta_0 = 0$

$H_1 : \beta_0 \neq 0$ $\alpha = 0.05$

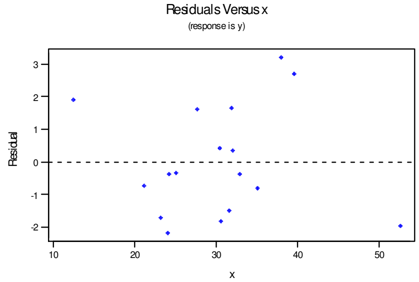
$t_0 = 0.468345$

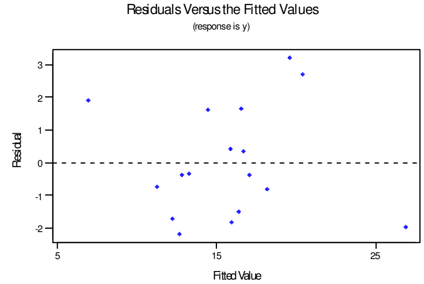
$t_{.025,15} = 2.131$

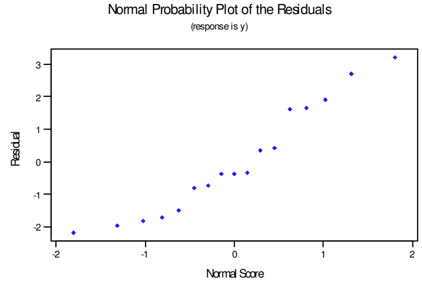
$t_0 \ngtr t_{\alpha/2,15}$

Do not reject H_0. We cannot conclude β_0 is different from zero.

e) No problems with model assumptions are noted.







11-62. n = 25 r = 0.83

a) $H_0 : \rho = 0$

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

$t_0 = \frac{r\sqrt{n-2}}{\sqrt{1-r^2}} = \frac{0.83\sqrt{23}}{\sqrt{1-(0.83)^2}} = 7.137$

$t_{.025,23} = 2.069$

$t_0 > t_{\alpha/2,23}$

Reject H_0. P-value = 0.

b) $\tanh(\text{arctanh } 0.83 - \frac{z_{.025}}{\sqrt{22}}) \le \rho \le \tanh(\text{arctanh } 0.83 + \frac{z_{.025}}{\sqrt{22}})$

where $z_{.025} = 1.96$. $0.6471 \le \rho \le 0.9226$.

c) $H_0 : \rho = 0.8$

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

$z_0 = (\text{arctanh } 0.83 - \text{arctanh } 0.8)(22)^{1/2} = 0.4199$

$z_{.025} = 1.96$

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

Do not reject H_0. P-value = (0.3373)(2) = 0.6746.

Supplemental Exercises

11-63. 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 normal equation

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 normal equations. Then,

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

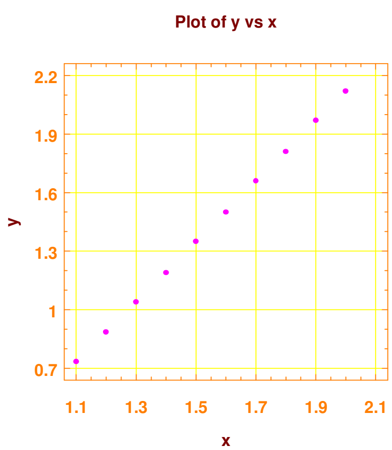
$$\hat{\beta}_0 \sum_{i=1}^{n} x_i + \hat{\beta}_1 \sum_{i=1}^{n} x_i^2 - \hat{\beta}_0 \sum_{i=1}^{n} x_i - \hat{\beta}_1 \sum_{i=1}^{n} x_i^2 = 0$$

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

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

$$\begin{aligned}\frac{1}{n}\sum_{i=1}^{n}\hat{y}_i &= \frac{1}{n}\sum(\hat{\beta}_0 + \hat{\beta}_1 x_i)\\ &= \frac{1}{n}(n\hat{\beta}_0 + \hat{\beta}_1\sum x_i)\\ &= \frac{1}{n}(n(\overline{y} - \hat{\beta}_1\overline{x}) + \hat{\beta}_1\sum x_i)\\ &= \frac{1}{n}(n\overline{y} - n\hat{\beta}_1\overline{x} + \hat{\beta}_1\sum x_i)\\ &= \overline{y} - \hat{\beta}\overline{x} + \hat{\beta}_1\overline{x}\\ &= \overline{y}\end{aligned}$$

11-64. a)



Yes, a straight line 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) $\alpha = 0.05$

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 $f_{0.05,1,8} = 5.32$

7) Using the results from the ANOVA table

$$f_0 = \frac{1.96613/1}{0.0000623515/8} = 255263.9$$

8) Since 2552639 > 5.32 reject H_0 and conclude that the regression model is significant at $\alpha = 0.05$.
P-value < 0.000001

d)

```
              95 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
--------------------------------------------------------------------------------
```

$-0.97800 \leq \beta_0 \leq -0.95565$

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

3) $H_1: \beta_0 \neq 0$

4) $\alpha = 0.05$

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

6) Reject H_0 if $t_0 < -t_{\alpha/2,n-2}$ where $-t_{0.025,8} = -2.306$ or $t_0 > t_{0.025,8} = 2.306$

7) Using the results from the table above

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

8) Since $-199.34 < -2.306$ reject H_0 and conclude the intercept is significant at $\alpha = 0.05$.

11-65. a) $\hat{y} = 93.34 + 15.64x$

b) $H_0 : \beta_1 = 0$

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

$f_0 = 12.872$

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

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

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

c) $(7.961 \leq \beta_1 \leq 23.322)$

d) $(74.758 \leq \beta_0 \leq 111.923)$

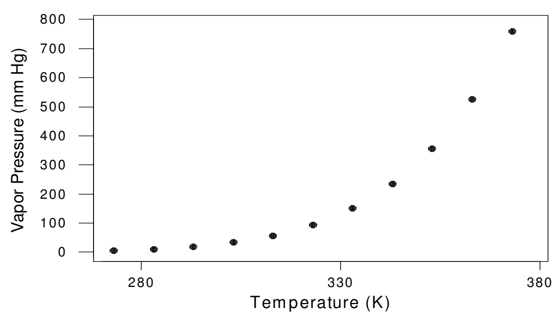
e) $\hat{y} = 93.34 + 15.64(2.5) = 132.44$

$$132.44 \pm 2.145\sqrt{136.27\left[\frac{1}{16} + \frac{(2.5-2.325)^2}{7.017}\right]}$$

$$132.44 \pm 6.47$$

$$125.97 \leq \hat{\mu}_{Y|x_0=2.5} \leq 138.91$$

11-66 a) There is curvature in the data.



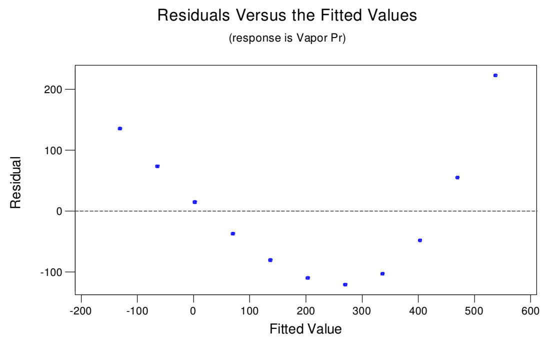
x

b) $y = -1956.3 + 6.686x$

c)

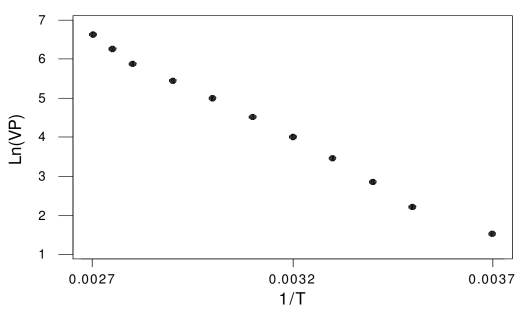
Source	DF	SS	MS	F	P
Regression	1	491662	491662	35.57	0.000
Residual Error	9	124403	13823		
Total	10	616065			

d)

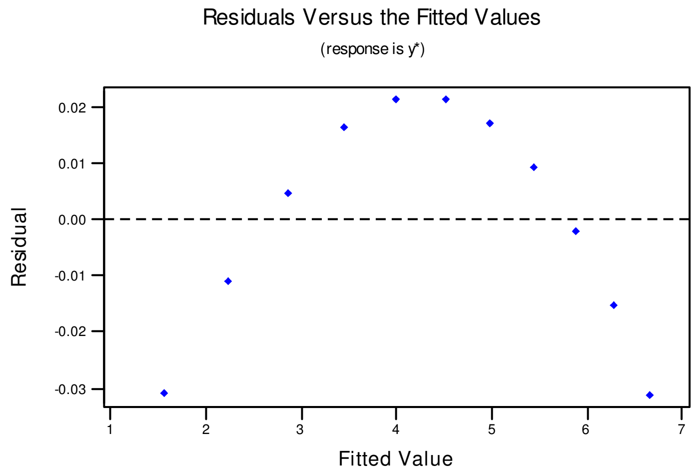


There is a curve in the residuals.

e) The data are linear after the transformation.

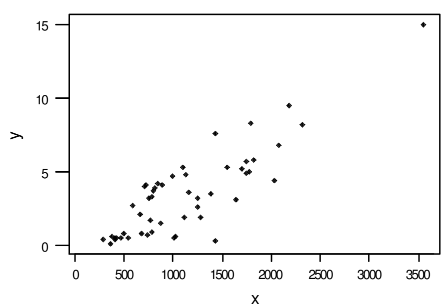



```
lny = 20.6 - 5201 (1/x)

Analysis of Variance
Source            DF         SS         MS         F        P
Regression         1     28.511     28.511  66715.47    0.000
Residual Error     9      0.004      0.000
Total             10     28.515
```



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

11-67. a)



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

c) $H_0 : \beta_1 = 0$

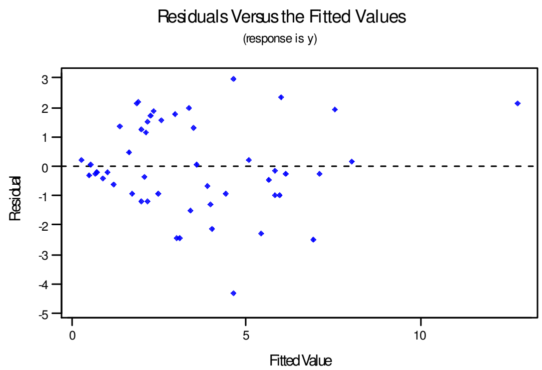
$H_1 : \beta_1 \neq 0$ $\alpha = 0.05$

$f_0 = 122.03$

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

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.

11-68. $\hat{y}^* = 1.2232 + 0.5075x$ where $y^* = \frac{1}{y}$. No, model does not seem reasonable. The residual plots indicate a possible outlier.

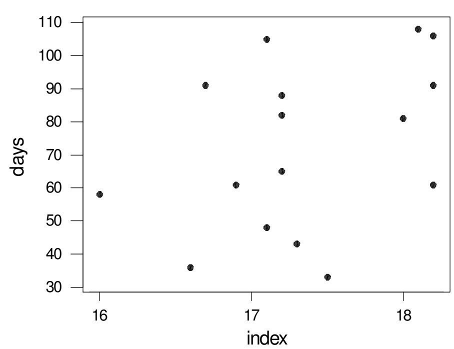
11-69. $\hat{y} = 0.7916x$

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

11-70. $\hat{y} = 4.5755 + 2.2047x$, r= 0.992, $R^2 = 98.40\%$

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

11-71 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			

Cannot reject H_o; therefore we conclude that the model is not significant. Therefore the seasonal meteorological index (x) is not a reliable predictor of the number of days that the ozone level exceeds 0.20 *ppm* (y).

c) 95% CI on β_1

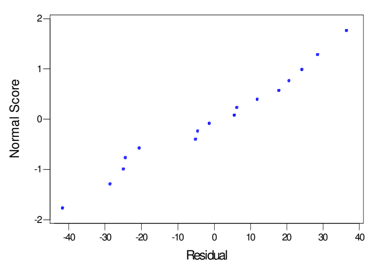
$$\hat{\beta}_1 \pm t_{\alpha/2,n-2} se(\hat{\beta}_1)$$

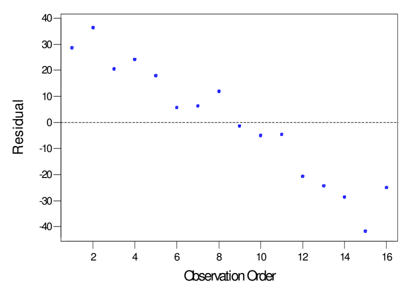
$$15.296 \pm t_{.025,12}(9.421)$$

$$15.296 \pm 2.145(9.421)$$

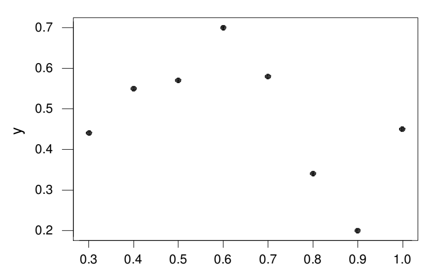
$$-4.912 \leq \beta_1 \leq 35.504$$

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, one that changes with time.





11-72 a)



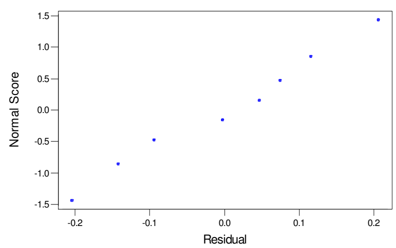
b) $\hat{y} = .6714 - 2964x$

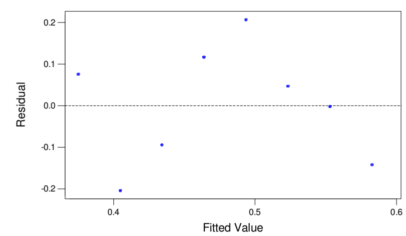
c) Analysis of Variance

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			

$R^2 = 21.47\%$

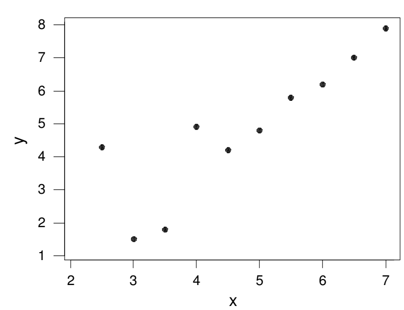
d) There appears to be curvature in the data. There is a dip in the middle of the normal probability plot and the plot of the residuals versus the fitted values shows curvature.





11-73 The correlation coefficient for the n pairs of data (x_i, z_i) will not be near unity. It will be near zero. The data for the pairs (x_i, z_i) where $z_i = y_i^2$ will not fall along the straight line $y_i = x_i$ which has a slope near unity and gives a correlation coefficient near unity. These data will fall on a line $y_i = \sqrt{x_i}$ that has a slope near zero and gives a much smaller correlation coefficient.

11-74 a)



b) $\hat{y} = -0.699 + 1.66x$

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			

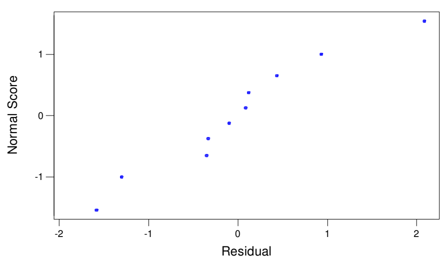
d) x=4.25 $\mu_{y|x_0} = 4.257$

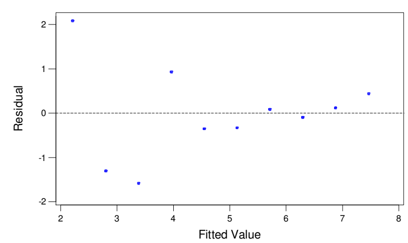
$$4.257 \pm 2.306\sqrt{1.2324\left(\frac{1}{10}+\frac{(4.25-4.75)^2}{20.625}\right)}$$

$$4.257 \pm 2.306(0.3717)$$

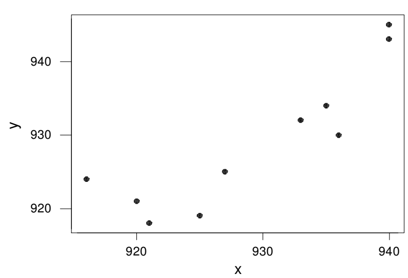
$$3.399 \le \mu_{y|x_o} \le 5.114$$

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





11-75 a)



b) $\hat{y} = 33.3 + 0.9636x$

c)

Predictor	Coef	SE Coef	T	P
Constant	66.0	194.2	0.34	0.743
Therm	0.9299	0.2090	4.45	0.002

S = 5.435 R-Sq = 71.2% R-Sq(adj) = 67.6%

Analysis of Variance

Source	DF	SS	MS	F	P
Regression	1	584.62	584.62	19.79	0.002
Residual Error	8	236.28	29.53		
Total	9	820.90			

Reject the hull hypothesis and conclude that the model is significant. 77.3% of the variability is explained by the model.

d) $H_0 : \beta_1 = 1$

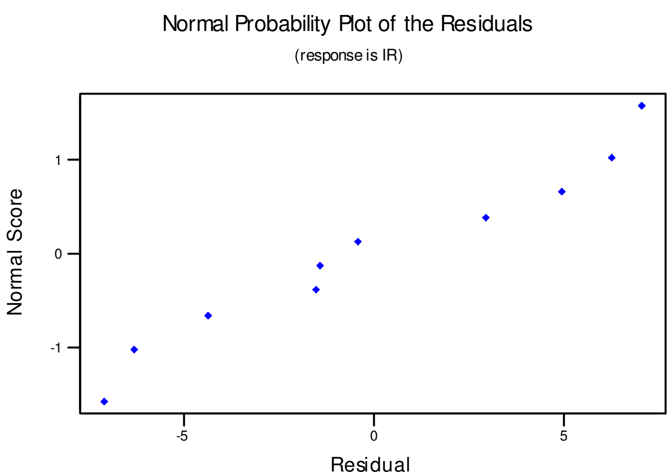
$H_1 : \beta_1 \neq 1$ α=.05

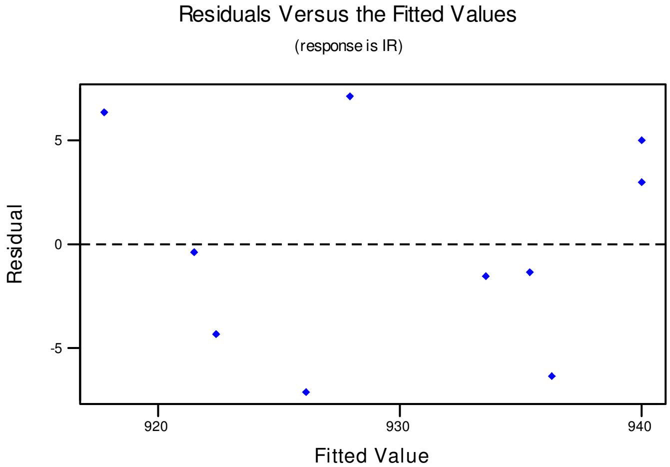
$$t_0 = \frac{\hat{\beta}_1 - 1}{se(\hat{\beta}_1)} = \frac{0.9299 - 1}{0.2090} = -0.3354$$

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

Since $t_0 > -t_{a/2,n-2}$, we cannot reject H_o and we conclude that there is not enough evidence to reject the claim that the devices produce different temperature measurements. Therefore, we assume the devices produce equivalent measurements.

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





Mind-Expanding Exercises

11-76. a) $\hat{\beta}_1 = \dfrac{S_{xY}}{S_{xx}}$, $\quad \hat{\beta}_0 = \overline{Y} - \hat{\beta}_1 \overline{x}$

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

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

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

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

b) The requested result is shown in part a.

11-77. a) $MS_E = \dfrac{\sum (Y_i - \hat{\beta}_0 - \hat{\beta}_1 x_i)^2}{n-2} = \dfrac{\sum e_i^2}{n-2}$

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

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

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

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

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

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

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

11-78. $\hat{\beta}_1 = \dfrac{S_{x_1Y}}{S_{x_1x_1}}$

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

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

11-79. $V(\hat{\beta}_1) = \dfrac{\sigma^2}{S_{xx}}$. 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.

11-80. 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

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

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

as requested.

and

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

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

11-81. $$\hat{y} = \bar{y} + r\frac{s_y}{s_x}(x - \bar{x})$$

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

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

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

11-82. 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.