CHAPTER 12

Sections 12-1

12-1. a) $X'X = \begin{bmatrix} 10 & 223 & 553 \\ 223 & 5200.9 & 12352 \\ 553 & 12352 & 31729 \end{bmatrix}$

$X'y = \begin{bmatrix} 1916.0 \\ 43550.8 \\ 104736.8 \end{bmatrix}$

b) $\hat{\beta} = \begin{bmatrix} 171.055 \\ 3.713 \\ -1.126 \end{bmatrix}$ so $\hat{y} = 171.055 + 3.714x_1 - 1.126x_2$

c) $\hat{y} = 171.055 + 3.714(18) - 1.126(43) = 189.49$

12-2. a) $\hat{\beta} = (X'X)^{-1}X'y$

$\hat{\beta} = \begin{bmatrix} -1.9122 \\ 0.0931 \\ 0.2532 \end{bmatrix}$

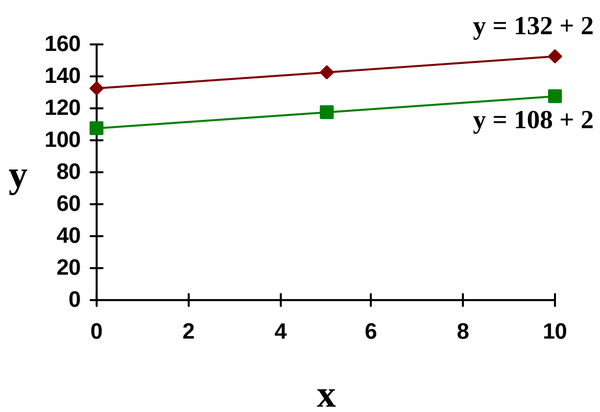
b) $\hat{y} = -1.9122 + 0.0931x_1 + 0.2532x_2$

$\hat{y} = -1.9122 + 0.0931(200) + 0.2532(50) = 29.3678$

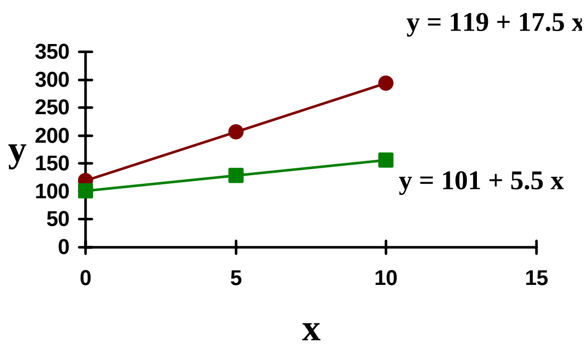
12-3. a)

	Model 1	Model 2
$x_2 = 2$	$\hat{y} = 100 + 2x_1 + 8$	$\hat{y} = 95 + 1.5x_1 + 3(2) + 4x_1$
	$\hat{y} = 108 + 2x_1$	$\hat{y} = 101 + 5.5x_1$
$x_2 = 8$	$\hat{y} = 100 + 2x_1 + 4(8)$	$\hat{y} = 95 + 1.5x_1 + 3(8) + 16x_1$
	$\hat{y} = 132 + 2x_1$	$\hat{y} = 119 + 17.5x_1$

MODEL 1



MODEL 2



The interaction term in model 2 affects the slope of the regression equation. That is, it modifies the amount of change per unit of x_1 on $\hat{y}$.

b) $x_2 = 5$ $\qquad \hat{y} = 100 + 2x_1 + 4(5)$

$\hat{y} = 120 + 2x_1$

Then, 2 is the expected change on $\hat{y}$ per unit of x_1.

NO, it does not depend on the value of x_2, because there is no relationship or interaction between these two variables in model 1.

c)

	$x_2 = 5$	$x_2 = 2$	$x_2 = 8$
	$\hat{y} = 95 + 1.5x_1 + 3(5) + 2x_1(5)$ $\hat{y} = 110 + 11.5x_1$	$\hat{y} = 101 + 5.5x_1$	$\hat{y} = 119 + 17.5x_1$
Change per unit of x_1	11.5	5.5	17.5

Yes, result does depend on the value of x_2, because x_2 interacts with x_1.

12-4

```
The regression equation is
y = - 2.75 + 0.00362 x2 + 0.204 x7 - 0.00467 x8

Predictor        Coef     SE Coef         T        P
Constant       -2.747       7.823     -0.35    0.729
x2          0.0036229   0.0006868      5.28    0.000
x7            0.20355     0.08677      2.35    0.028
x8          -0.004673    0.001272     -3.67    0.001

S = 1.687       R-Sq = 79.1%     R-Sq(adj) = 76.5%
Analysis of Variance
Source            DF          SS          MS         F        P
Regression         3     258.683      86.228     30.31    0.000
Residual Error    24      68.281       2.845
Total             27     326.964
```

a) $\hat{y} = -2.747 + 0.0036x_2 + 0.2035x_7 - 0.0047x_8$

b) $\hat{\sigma}^2 = 2.845$

c) $se(\hat{\beta}_0) = 7.823$, $se(\hat{\beta}_2) = .0006868$, $se(\hat{\beta}_7) = .08677$, and $se(\hat{\beta}_8) = .001272$

d) $\hat{y} = -2.747 + 0.0036229(2000) + 0.20355(60) - 0.004673(1800)$ $\hat{y} = 8.3 \approx 8$ games.

12-5.

Predictor	Coef	StDev	T	P
Constant	33.449	1.576	21.22	0.000
x1	-0.054349	0.006329	-8.59	0.000
x6	1.0782	0.6997	1.54	0.138

S = 2.834   R-Sq = 82.9%   R-Sq(adj) = 81.3%

Analysis of Variance

Source	DF	SS	MS	F	P
Regression	2	856.24	428.12	53.32	0.000
Error	22	176.66	8.03		
Total	24	1032.90			

a) $\hat{y} = 33.4491 - 0.05435x_1 + 1.07822x_6$

b) $\hat{\sigma}^2 = 8.03$

c) $\hat{y} = 33.4491 - 0.05435(300) + 1.07822(2) = 19.30$ mpg.

12-6

Predictor	Coef	SE Coef	T	P
Constant	-123.1	157.3	-0.78	0.459
X1	0.7573	0.2791	2.71	0.030
X2	7.519	4.010	1.87	0.103
X3	2.483	1.809	1.37	0.212
X4	-0.4811	0.5552	-0.87	0.415

S = 11.79   R-Sq = 85.2%   R-Sq(adj) = 76.8%

Analysis of Variance

Source	DF	SS	MS	F	P
Regression	4	5600.5	1400.1	10.08	0.005
Residual Error	7	972.5	138.9		
Total	11	6572.9			

a) $\hat{y} = -123.1 + 0.7573x_1 + 7.519x_2 + 2.483x_3 - 0.4811x_4$

b) $\hat{\sigma}^2 = 139.00$

c) $se(\hat{\beta}_0) = 157.3$, $se(\hat{\beta}_1) = 0.2791$, $se(\hat{\beta}_2) = 4.010$, $se(\hat{\beta}_3) = 1.809$, and $se(\hat{\beta}_4) = 0.5552$

d) $\hat{y} = -123.1 + 0.7573(75) + 7.519(24) + 2.483(90) - 0.4811(98) = 290.476$

12-7.

Predictor	Coef	SE Coef	T	P
Constant	383.80	36.22	10.60	0.002
Xl	-3.6381	0.5665	-6.42	0.008
X2	-0.11168	0.04338	-2.57	0.082

S = 12.35   R-Sq = 98.5%   R-Sq(adj) = 97.5%

Analysis of Variance

Source	DF	SS	MS	F	P
Regression	2	29787	14894	97.59	0.002
Residual Error	3	458	153		
Total	5	30245			

a) $\hat{y} = 383.80 - 3.6381x_1 - 0.1119x_2$

b) $\hat{\sigma}^2 = 153.0$, $se(\hat{\beta}_0) = 36.22$, $se(\hat{\beta}_1) = 0.5665$, and $se(\hat{\beta}_2) = .04338$

c) $\hat{y} = 383.80 - 3.6381(25) - 0.1119(1000) = 180.95$

d)

Predictor	Coef	SE Coef	T	P
Constant	484.0	101.3	4.78	0.041
Xl	-7.656	3.846	-1.99	0.185
X2	-0.2221	0.1129	-1.97	0.188
X1*X2	0.004087	0.003871	1.06	0.402

S = 12.12   R-Sq = 99.0%   R-Sq(adj) = 97.6%

Analysis of Variance

Source	DF	SS	MS	F	P
Regression	3	29951.4	9983.8	67.92	0.015
Residual Error	2	294.0	147.0		
Total	5	30245.3			

$\hat{y} = 484.0 - 7.656x_1 - 0.222x_2 - 0.0041x_{12}$

e) $\hat{\sigma}^2 = 147.0$, $se(\hat{\beta}_0) = 101.3$, $se(\hat{\beta}_1) = 3.846$, $se(\hat{\beta}_2) = 0.113$ and $se(\hat{\beta}_{12}) = 0.0039$

f) $\hat{y} = 484.0 - 7.656(25) - 0.222(1000) - 0.0041(25)(1000) = 173.1$

The predicted value is smaller

12-8.

```
The regression equation is
y = 7.46 - 0.030 x2 + 0.521 x3 - 0.102 x4 - 2.16 x5
Predictor        Coef       StDev          T        P
Constant        7.458       7.226       1.03    0.320
x2            -0.0297      0.2633      -0.11    0.912
x3             0.5205      0.1359       3.83    0.002
x4           -0.10180     0.05339      -1.91    0.077
x5             -2.161       2.395      -0.90    0.382
S = 0.8827      R-Sq = 67.2%     R-Sq(adj) = 57.8%
Analysis of Variance
Source       DF          SS          MS         F        P
Regression    4     22.3119      5.5780      7.16    0.002
Error        14     10.9091      0.7792
Total        18     33.2211
```

a) $\hat{y} = 7.4578 - 0.0297x_2 + 0.5205x_3 - 0.1018x_4 - 2.1606x_5$

b) $\hat{\sigma}^2 = .7792$

c) $se(\hat{\beta}_0) = 7.226$, $se(\hat{\beta}_2) = .2633$, $se(\hat{\beta}_3) = .1359$, $se(\hat{\beta}_4) = .05339$ and $se(\hat{\beta}_5) = 2.395$

d) $\hat{y} = 7.4578 - 0.0297(20) + 0.5205(30) - 0.1018(90) - 2.1606(2.0)$ $\hat{y} = 8.996$

12-9.

```
Predictor        Coef     SE Coef          T        P
Constant        47.82       49.94       0.96    0.353
x1             -9.604       3.723      -2.58    0.020
x2             0.4152      0.2261       1.84    0.085
x3             18.294       1.323      13.82    0.000
--------------------------------------------------------------------------------
```

a) $\hat{y} = 47.82 - 9.604x_1 + 0.44152x_2 + 18.294x_3$

b) $\hat{\sigma}^2 = 12.3$

c) $se(\hat{\beta}_0) = 49.94$, $se(\hat{\beta}_1) = 3.723$, $se(\hat{\beta}_2) = 0.2261$, and $se(\hat{\beta}_3) = 1.323$

d) $\hat{y} = 47.82 - 9.604(14.5) + 0.44152(220) + 18.29(5) = 91.38$

12-10.

```
Predictor         Coef      SE Coef          T        P
Constant      -0.03023      0.06178      -0.49    0.629
temp        0.00002856   0.00003437       0.83    0.414
soaktime     0.0023182    0.0001737      13.35    0.000
soakpct      -0.003029     0.005844      -0.52    0.609
difftime      0.008476     0.001218       6.96    0.000
diffpct      -0.002363     0.008078      -0.29    0.772

S = 0.002296    R-Sq = 96.8%     R-Sq(adj) = 96.2%

Analysis of Variance

Source            DF           SS           MS         F        P
Regression         5   0.00418939   0.00083788    158.92    0.000
Residual Error    26   0.00013708   0.00000527
Total             31   0.00432647
```

a) $\hat{y} = -0.03023 + 0.000029x_1 + 0.002318x_2 - 0.003029x_3 + 0.008476x_4 - 0.002363x_5$

where $x_1 = TEMP$ $x_2 = SOAKTIME$ $x_3 = SOAKPCT$ $x_4 = DFTIME$ $x_5 = DIFFPCT$

b) $\hat{\sigma}^2 = 5.27x10^{-6}$

c) The standard errors are listed under the StDev column above.

d)

$$\hat{y} = -0.03023 + 0.000029(1650) + 0.002318(1) - 0.003029(1.1)$$
$$+ 0.008476(1) - 0.002363(0.80)$$
$$\hat{y} = 0.02247$$

12-11.

Predictor	Coef	StDev	T	P
Constant	-8.01	16.18	-0.50	0.630
Pts	0.49400	0.04806	10.28	0.000
GF	0.001847	0.006424	0.29	0.779
GA	0.00231	0.02091	0.11	0.914
PPG	0.03830	0.05146	0.74	0.472
PPcT	-0.2068	0.2611	-0.79	0.445
SHG	-0.01284	0.02562	-0.50	0.626
PPGA	0.03074	0.03802	0.81	0.436
PKPcT	0.0407	0.1483	0.27	0.789
SHGA	-0.2083	0.1110	-1.88	0.087

S = 1.523 R-Sq = 98.8% R-Sq(adj) = 97.8%

Analysis of Variance

Source	DF	SS	MS	F	P
Regression	9	2125.43	236.16	101.79	0.000
Error	11	25.52	2.32		
Total	20	2150.95			

$$\hat{y} = -8.0119 + 0.494x_1 + 0.0018x_2 + 0.0023x_3 + 0.0383x_4 - 0.2068x_5 -$$
$$0.0128x_6 + 0.030x_7 + 0.0407x_8 - 0.2083x_9$$

where

$x_1 = Pts \quad x_2 = GF \quad x_3 = GA \quad x_4 = PPG \quad x_5 = PPct \quad x_6 = SHG \quad x_7 = PPGA$

$x_8 = PKPcT \quad x_9 = SHGA$

$\hat{\sigma}^2 = 2.32$

The standard errors of the coefficients are listed under the StDev column above.

12-12. a) $f(\beta_0', \beta_1, \beta_2) = \sum [y_i - \beta_0' - \beta_1(x_{i1} - \bar{x}_1) - \beta_2(x_{i2} - \bar{x}_2)]^2$

$$\frac{\partial f}{\partial \beta_0'} = -2\sum [y_i - \beta_0' - \beta_1(x_{i1} - \bar{x}_1) - \beta_2(x_{i2} - \bar{x}_2)]$$

$$\frac{\partial f}{\partial \beta_1} = -2\sum [y_i - \beta_0' - \beta_1(x_{i1} - \bar{x}_1) - \beta_2(x_{i2} - \bar{x}_2)](x_{i1} - \bar{x}_1)$$

$$\frac{\partial f}{\partial \beta_2} = -2\sum [y_i - \beta_0' - \beta_1(x_{i1} - \bar{x}_1) - \beta_2(x_{i2} - \bar{x}_2)](x_{i2} - \bar{x}_2)$$

Setting the derivatives equal to zero yields

$$n\beta_0' = \sum y_i$$

$$n\beta_0' + \beta_1 \sum (x_{i1} - \bar{x}_1)^2 + \beta_2 \sum (x_{i1} - \bar{x}_1)(x_{i2} - \bar{x}_2) = \sum y_i (x_{i1} - \bar{x}_1)$$

$$n\beta_0' + \beta_1 \sum (x_{i1} - \bar{x}_1)(x_{i2} - \bar{x}_2) + \beta_2 \sum (x_{i2} - \bar{x}_2)^2 = \sum y_i (x_{i2} - \bar{x}_2)$$

b) From the first normal equation, $\hat{\beta}_0' = \bar{y}$.

c) Substituting $y_i - \bar{y}$ for y_i in the first normal equation yields $\hat{\beta}_0' = 0$.

Sections 12-2

12-13. a) n = 10, k = 2, p = 3, α = 0.05

$$H_0 : \beta_1 = \beta_2 = \ldots = \beta_k = 0$$

$H_1 : \beta_j \neq 0$ for at least one j

$$S_{yy} = 371595.6 - \frac{(1916)^2}{10} = 4490$$

$$X'y = \begin{bmatrix} \sum y_i \\ \sum x_{i1} y_i \\ \sum x_{i2} y_i \end{bmatrix} = \begin{bmatrix} 1916 \\ 43550.8 \\ 104736.8 \end{bmatrix}$$

$$\hat{\beta}' X'y = \begin{bmatrix} 171.055 & 3.713 & -1.126 \end{bmatrix} \begin{bmatrix} 1916 \\ 43550.8 \\ 104736.8 \end{bmatrix} = 371511.9$$

$$SS_R = 371511.9 - \frac{1916^2}{10} = 4406.3$$

$$SS_E = S_{yy} - SS_R = 4490 - 4406.3 = 83.7$$

$$f_0 = \frac{\frac{SS_R}{k}}{\frac{SS_E}{n-p}} = \frac{4406.3/2}{83.7/7} = 184.25$$

$$f_{0.05,2,7} = 4.74$$

$$f_0 > f_{0.05,2,7}$$

Reject H_0 and conclude that the regression model is significant at α = 0.05.

b) $\hat{\sigma}^2 = MS_E = \dfrac{SS_E}{n-p} = 11.957$

c) $se(\hat{\beta}_1) = \sqrt{\hat{\sigma}^2 c_{11}} = \sqrt{11.957(0.00439)} = 0.229$

$se(\hat{\beta}_2) = \sqrt{\hat{\sigma}^2 c_{22}} = \sqrt{11.957(0.00087)} = 0.10199$

d)

$H_0 : \beta_1 = 0$	$\beta_2 = 0$
$H_1 : \beta_1 \neq 0$	$\beta_2 \neq 0$
$t_0 = \dfrac{\hat{\beta}_1}{se(\hat{\beta}_1)}$	$t_0 = \dfrac{\hat{\beta}_2}{se(\hat{\beta}_2)}$
$= \dfrac{3.713}{0.229} = 16.21$	$= \dfrac{-1.126}{0.10199} = -11.04$

$t_{\alpha/2,7} = t_{0.025,7} = 2.365$

Reject H_0 Reject H_0

Both regression coefficients significant

12-14 $S_{yy} = 742.00$

a) $H_0 : \beta_1 = \beta_2 = 0$

$H_1 : \beta_j \neq 0$ for at least one j

$\alpha = 0.01$

$$SS_R = \hat{\beta}' X' y - \frac{(\sum_{i=1}^{n} y_i)^2}{n}$$

$$= \begin{pmatrix} -1.9122 & 0.0931 & 0.2532 \end{pmatrix} \begin{pmatrix} 220 \\ 36768 \\ 9965 \end{pmatrix} - \frac{220^2}{10}$$

$$= 5525.5548 - 4840$$

$$= 685.55$$

$$SS_E = S_{yy} - SS_R$$

$$= 742 - 685.55$$

$$= 56.45$$

$$f_0 = \frac{\frac{SS_R}{k}}{\frac{SS_E}{n-p}} = \frac{685.55/2}{56.45/7} = 42.51$$

$$f_{0.01,2,7} = 9.55$$

$$f_0 > f_{0.01,2,7}$$

Reject H_0 and conclude that the regression model is significant at $\alpha = 0.01$. P-value = 0.000121

$$\hat{\sigma}^2 = MS_E = \frac{SS_E}{n-p} = \frac{56.45}{7} = 8.0643$$

$$se(\hat{\beta}_1) = \sqrt{8.0643(7.9799E-5)} = 0.0254$$

b) $H_0 : \beta_1 = 0$

$H_1 : \beta_1 \neq 0$

$$t_0 = \frac{\hat{\beta}_1}{se(\hat{\beta}_1)}$$

$$= \frac{0.0931}{0.0254} = 3.67$$

$$t_{0.005,7} = 3.499$$

$$|t_0| > t_{0.005,7}$$

Reject H_0 and conclude that β_1 is significant in the model at $\alpha = 0.01$

P-value = $2(1 - P(t < t_0))$ = 2(1 - 0.996018) = 0.007964

12-15. a) $H_0 : \beta_2 = \beta_7 = \beta_8 = 0$

H_1 for at least one $\beta_j \neq 0$

$\alpha = 0.05$

$f_0 = 30.3085$

$f_{.05,3,24} = 3.0088$

$f_0 > f_{0.05,3,24}$

Reject H_0 and conclude regression model is significant at $\alpha = 0.05$ P-value < 0.000001

b) $\hat{\sigma}^2 = MS_E = 2.845$

$se(\hat{\beta}_2) = \sqrt{\hat{\sigma}^2 c_{jj}} = 0.0006868$

$H_0 : \beta_2 = 0$

$H_1 : \beta_2 \neq 0$

$\alpha = 0.05$

$t_0 = 5.275$

$t_{.025,28-4} = t_{0.025,24} = 2.064$

$|t_0| > t_{\alpha/2,24}$, Reject H_0

$H_0 : \beta_7 = 0$	$H_0 : \beta_8 = 0$
$H_1 : \beta_7 \neq 0$	$H_1 : \beta_8 \neq 0$
$\alpha = 0.05$	$\alpha = 0.05$
$t_0 = 2.346$	$t_0 = -3.673$
Reject H_0	Reject H_0

Conclude the variables are significant at $\alpha = 0.05$

12-16. a) $SS_R(\beta_8 \mid \beta_7, \beta_2, \beta_0) = 38.387$

b) $H_0 : \beta_8 = 0$

$H_1 : \beta_8 \neq 0$

$\alpha = 0.05$

$$f_0 = \frac{SS_R(\beta_8 \mid \beta_7, \beta_2, \beta_0)/1}{MS_E} = \frac{38.387}{2.845} = 13.49$$

$f_{0.05,1,24} = 4.26$

$f_0 > f_{0.05,1,24}$

Reject H_0 and conclude X_8 regressor is significant at $\alpha = 0.05$
P-value = 0.0012

12-17. a) $H_0 : \beta_1 = \beta_6 = 0$

H_1: at least one $\beta \neq 0$

$f_0 = 53.3162$

$f_{\alpha,2,22} = f_{0.05,2,22} = 3.44$

$f_0 > f_{0.05,2,22}$

Reject H_0 and conclude regression model is significant at $\alpha = 0.05$

b) $H_0 : \beta_1 = 0$

$H_1 : \beta_1 \neq 0$

$t_0 = -8.59$

$t_{.025,25-3} = t_{.025,22} = 2.074$

$|t_0| > t_{0.025,22}$,Reject H_0 and conclude β_1 is significant at $\alpha = 0.05$

$H_0 : \beta_6 = 0$

$H_1 : \beta_6 \neq 0$

$\alpha = 0.05$

$t_0 = 1.5411$

$|t_0| \not> t_{0.025,22}$,Do not reject H_0, conclude that evidence is not significant to state β_6 is significant at $\alpha = 0.05$.

No, only x_1 contributes significantly to the regression.

12-18. a) $\beta_1 : t_0 = 4.82$ P-value = 2(4.08 E-5) = 8.16 E-5

$\beta_2 : t_0 = 8.21$ P-value = 291.91 E-8) = 3.82 E-8

$\beta_3 : t_0 = 0.98$ P-value = 2 (0.1689) = 0.3378

b) $H_0 : \beta_3 = 0$

$H_1 : \beta_3 \neq 0$

$\alpha = 0.05$

$t_0 = 0.98$

$t_{0.025,22} = 2.074$

$|t_0| \not> t_{0.025,23}$

Do not reject H_0.

12-19. a) $H_0 : \beta_1 = \beta_2 = \beta_3 = \beta_4 = 0$

H_1 at least one $\beta_j \neq 0$

$\alpha = 0.05$

$f_0 = 10.08$

$f_{0.05,4,7} = 4.12$

$f_0 > f_{0.05,4,7}$

Reject H_0 P-value = 0.005

b) $\alpha = 0.05$

$H_0: \beta_1 = 0 \quad \beta_2 = 0 \quad \beta_3 = 0 \quad \beta_4 = 0$

$H_1: \beta_1 \neq 0 \quad \beta_2 \neq 0 \quad \beta_3 \neq 0 \quad \beta_4 \neq 0$

$t_0 = 2.71 \quad t_0 = 1.87 \quad t_0 = 1.37 \quad t_0 = -0.87$

$t_{\alpha/2, n-p} = t_{0.025,7} = 2.365$

$|t_0| \ngtr t_{0.0257}$ for β_2, β_3 and β_4

Reject H_0 for β_1.

12-20. $\hat{y} = \hat{\beta}_0 + \hat{\beta}_1 x_1 + \hat{\beta}_2 x_2$, Assume no interaction model.

a) $H_0: \beta_1 = \beta_2 = 0$

H_1 at least one $\beta_j \neq 0$

$f_0 = 97.59$

$f_{0.05,2,3} = 9.55$

$f_0 > f_{0.05,2,3}$

Reject H_0 P-value = 0.002

b) $H_0: \beta_1 = 0$

$H_1: \beta_1 \neq 0$

$t_0 = -6.42$

$t_{\alpha/2,3} = t_{0.025,3} = 3.182$

$|t_0| > t_{0.025,3}$

Reject H_0 for regressor β_1.

$H_0: \beta_2 = 0$

$H_1: \beta_2 \neq 0$

$t_0 = -2.57$

$t_{\alpha/2,3} = t_{0.025,3} = 3.182$

$|t_0| \ngtr t_{0.025,3}$

c) $SS_R(\beta_2 | \beta_1, \beta_0) = 1012$

$H_0: \beta_2 = 0$

$H_1: \beta_2 \neq 0$

$\alpha = 0.05$

$f_0 = 6.629$

$f_{\alpha,1,3} = f_{.05,1,3} = 10.13$

$f_0 \ngtr f_{0.05,1,3}$

Do not reject H_0

12-21. a) $H_0 : \beta_1 = \beta_2 = \beta_{12} = 0$

H_1 at least one $\beta_j \neq 0$

$\alpha = 0.05$

$f_0 = 7.714$

$f_{\alpha,3,2} = f_{0.05,3,2} = 19.16$

$f_0 \ngtr f_{0.05,3,2}$

Do not reject H_0

b) $H_0 : \beta_{12} = 0$

$H_1 : \beta_{12} \neq 0$

$\alpha = 0.05$

$SSR(\beta_{12} \mid \beta_1, \beta_2) = 29951.4 - 29787 = 163.9$

$f_0 = \frac{SSR}{MS_E} = \frac{163.9}{147} = 1.11$

$f_{0.05,1,2} = 18.51$

$f_0 \ngtr f_{0.05,1,2}$

Do not reject H_0

c) $\hat{\sigma}^2 = 147.0$

$\hat{\sigma}^2$ (no interaction term) = 153.0

$MS_E(\hat{\sigma}^2)$ was reduced in the interaction term model due to the addition of this term.

12-22. a) $H_0 : \beta_j = 0$ for all j

$H_1 : \beta_j \neq 0$ for at least one j

$f_0 = 7.16$

$f_{.05,4,14} = 3.11$

$f_0 > f_{0.05,4,14}$

Reject H_0 and conclude regression is significant at $\alpha = 0.05$. P-value = 0.0023

b) $\hat{\sigma} = 0.7792$

$\alpha = 0.05$ $\quad t_{\alpha/2,n-p} = t_{.025,14} = 2.145$

$H_0: \beta_2 = 0$	$\beta_3 = 0$	$\beta_4 = 0$	$\beta_5 = 0$
$H_1: \beta_2 \neq 0$	$\beta_3 \neq 0$	$\beta_4 \neq 0$	$\beta_5 \neq 0$
$t_0 = -0.113$	$t_0 = 3.83$	$t_0 = -1.91$	$t_0 = -0.9$
$\|t_0\| \ngtr t_{\alpha/2,14}$	$\|t_0\| > t_{\alpha/2,14}$	$\|t_0\| \ngtr t_{\alpha/2,14}$	$\|t_0\| \ngtr t_{\alpha/2,14}$
Do not reject H_0	Reject H_0	Do not reject H_0	Do not reject H_0

All variables do not contribute to the model.

12-23. a) $H_0 : \beta_1 = \beta_2 = \beta_3 = 0$

$H_1: \beta_j \neq 0$ for at least one j

$f_0 = 828.31$

$f_{.05,3,16} = 3.34$

$f_0 > f_{0.05,3,16}$

Reject H_0 and conclude regression is significant at $\alpha = 0.05$

b) $\hat{\sigma}^2 = 12.2856$

$\alpha = 0.05$ $\quad t_{\alpha/2,n-p} = t_{.025,16} = 2.12$

$H_0: \beta_1 = 0$	$\beta_2 = 0$	$\beta_3 = 0$
$H_1: \beta_1 \neq 0$	$\beta_2 \neq 0$	$\beta_3 \neq 0$
$t_0 = -2.58$	$t_0 = 1.84$	$t_0 = 13.82$
$\lvert t_0 \rvert > t_{0.025,16}$	$\lvert t_0 \rvert \not> t_{0.025,16}$	$\lvert t_0 \rvert > t_{0.025,16}$
Reject H_0	Do not reject H_0	Reject H_0

12-24. a) $H_0: \beta_j = 0$ for all j

$H_1: \beta_j \neq 0$ for at least one j

$f_0 = 158.9902$

$f_{.05,5,26} = 2.59$

$f_0 > f_{\alpha,5,26}$

Reject H_0 and conclude regression is significant at $\alpha = 0.05$

P-value < 0.000001

b) $\alpha = 0.05$ $\quad t_{\alpha/2,n-p} = t_{.025,26} = 2.056$

$H_0: \beta_1 = 0$	$\beta_2 = 0$	$\beta_3 = 0$	$\beta_4 = 0$	$\beta_5 = 0$
$H_1: \beta_1 \neq 0$	$\beta_2 \neq 0$	$\beta_3 \neq 0$	$\beta_4 \neq 0$	$\beta_5 \neq 0$
$t_0 = 0.83$	$t_0 = 12.25$	$t_0 = -0.52$	$t_0 = 6.96$	$t_0 = -0.29$
Do not Reject H_0	Reject H_0	Do not reject H_0	Reject H_0	Do not reject H_0

c) $\hat{y} = 0.010889 + 0.002687x_1 + 0.009325x_2$

d) $H_0: \beta_j = 0$ for all j

$H_1: \beta_j \neq 0$ for at least one j

$f_0 = 308.455$

$f_{.05,2,29} = 3.33$

$f_0 > f_{0.05,2,29}$

Reject H_0 and conclude regression is significant at $\alpha = 0.05$

$\alpha = 0.05$ $\quad t_{\alpha/2,n-p} = t_{.025,29} = 2.045$

$H_0: \beta_1 = 0$	$\beta_2 = 0$
$H_1: \beta_1 \neq 0$	$\beta_2 \neq 0$
$t_0 = 18.31$	$t_0 = 6.37$
$\lvert t_0 \rvert > t_{\alpha/2,29}$	$\lvert t_0 \rvert > t_{\alpha/2,29}$

Reject H_0 for each regressor variable and conclude that both variables are significant at $\alpha = 0.05$

e) $\hat{\sigma}_{part(d)} = 6.7E-6$. Part b) is smaller, suggesting a better model.

12-25. a) $H_0: \beta_j = 0$ for all j

$H_1: \beta_j \neq 0$ for at least one j

$f_0 = 101.79$

$f_{.05,9,11} = 2.90$

$f_0 > f_{0.05,9,11}$

Reject H_0 and conclude regression is significant at $\alpha = 0.05$

b) $H_0 : \beta_j = 0$

$H_1 : \beta_j \neq 0$

$t_{.025,11} = 2.201$

PTS :	$t_0 = 10.28$	Reject H_0
GF :	$t_0 = 0.29$	Do not reject H_0
GA :	$t_0 = 0.11$	Do not reject H_0
PPG :	$t_0 = 0.74$	Do not reject H_0
PPcT :	$t_0 = -0.79$	Do not reject H_0
SHG :	$t_0 = -0.50$	Do not reject H_0
PPGA :	$t_0 = 0.81$	Do not reject H_0
PKPcT :	$t_0 = 0.27$	Do not reject H_0
SHGA :	$t_0 = -1.88$	Do not reject H_0

Nc, only "PTS" (β_1) is significant at $\alpha = 0.05$

c)

$\hat{y} = -5.531 + 0.497x_1 - 0.004x_4$

$f_0 = 510.12$

$f_{.05,2,18} = 3.55$

Reject H_0

$H_0 : \beta_1 = 0$ $\qquad \beta_4 = 0$

$H_1 : \beta_1 \neq 0$ $\qquad \beta_4 \neq 0$

$t_0 = 30.60$ $\qquad t_0 = -0.165$

Reject H_0 $\qquad$ Do not reject H_0

Sections 12-3 and 12-4

12-26. a) $\hat{\beta}_1 \pm t_{\alpha/2,n-p}\sqrt{\hat{\sigma}^2 c_{11}}$

$0.0931 \pm t_{.025,7}\ se(\hat{\beta}_1)$

$0.0931 \pm (2.365)(0.0254)$

0.0931 ± 0.060071

$0.0330 \leq \beta_1 \leq 0.1532$

b) $x_1 = 200$

$x_2 = 50$

$\hat{y}_0 = 29.37$

$X_0'(X'X)^{-1}X_0 = 0.211088$

$29.37 \pm (2.365)\sqrt{8.0643(0.211088)}$

29.37 ± 3.08565

$26.28 \leq \mu_{Y|x_0} \leq 32.46$

c) $\alpha = 0.05$

$$x_1 = 200$$
$$x_2 = 50$$
$$\hat{y}_0 = 29.37$$
$$X_0'(X'X)^{-1}X_0 = 0.211088$$
$$\hat{y} \pm t_{\alpha/2,n-p}\sqrt{\hat{\sigma}^2(1 + X_0'(X'X)^{-1}X_0)}$$
$$29.37 \pm (2.365)\sqrt{8.0643(1.211088)}$$
$$29.37 \pm 7.391$$
$$21.98 \le y_0 \le 36.76$$

12-27. a) $-0.00467 \le \beta_8 \le -0.00127$

b) $\sqrt{\hat{\sigma}^2 X_0'(X'X)^{-1}X_0} = 0.4617 = se(\hat{\mu}_{Y|x_0})$

c) $\hat{y} = -2.747 + 0.003623(2000) + 0.20355(60) - 0.004673(1800) = 8.3$

$$\hat{\mu}_{Y|x_0} \pm t_{.025,24}\ se(\hat{\mu}_{Y|x_0})$$
$$8.30 \pm (2.064)(0.4617)$$
$$8.30 \pm 1.03$$
$$7.35 \le \mu_{Y|x_0} \le 9.25$$

12-28. a) $-0.07219 \le \beta_1 \le -0.03651$

$$-0.8939 \le \beta_6 \le 3.0504$$

b) $\hat{\mu}_{Y|x_0} = 214572$ $\quad$ $se(\hat{\mu}_{Y|x_0}) = 108785$

$$\hat{\mu}_{Y|x_0} \pm t_{.005,22}\ se(\hat{\mu}_{Y|x_0})$$
$$21.4572 \pm (2.819)(1.08785)$$
$$18.391 \le \mu_{Y|x_0} \le 24.524$$

c) $\hat{y} = 32.0223 - 0.064x_1 + 0.01269x_2 + 1.06181x_6 + 0.00069x_{10}$

$$\left.\begin{array}{l} -0.12973 \le \beta_1 \le 0.00174 \\ -0.13236 \le \beta_2 \le 0.15774 \\ -1.76898 \le \beta_6 \le 3.89259 \\ -0.00575 \le \beta_{10} \le 0.00713 \end{array}\right\} \quad \alpha = 0.01$$

d) Intervals of larger model are wider (part (c) intervals are wider). Yes, adding x_2 and x_{10} did not help the significance of the model, the standard errors increased.

12-29. a) 95 % CI on coefficients

$$0.0973 \le \beta_1 \le 1.4172$$
$$-1.9646 \le \beta_2 \le 17.0026$$
$$-1.7953 \le \beta_3 \le 6.7613$$
$$-1.7941 \le \beta_4 \le 0.8319$$

b) $\hat{\mu}_{Y|x_0} = 290.44$ $se(\hat{\mu}_{Y|x_0}) = 7.61$ $t_{.025,7} = 2.365$

$$\hat{\mu}_{Y|x_0} \pm t_{\alpha/2,n-p} \ se(\hat{\mu}_{Y|x_0})$$
$$290.44 \pm (2.365)(7.61)$$
$$272.44 \le \mu_{Y|x_0} \le 308.44$$

c) $\hat{y}_0 \pm t_{\alpha/2,n-p}\sqrt{\hat{\sigma}^2(1+\mathbf{x}_0{}'(\mathbf{X}'\mathbf{X})^{-1}\mathbf{x}_0)}$

$$290.44 \pm 2.365(14.038)$$
$$257.25 \le y_0 \le 323.64$$

12-30. a) $-6.9467 \le \beta_1 \le -0.3295$

$$-0.3651 \le \beta_2 \le 0.1417$$

b) $-45.8276 \le \beta_1 \le 30.5156$

$$-1.3426 \le \beta_2 \le 0.8984$$
$$-0.03433 \le \beta_{12} \le 0.04251$$

These part b. intervals are much wider.
Yes, the addition of this term increased the standard error of the regression coefficient estimators.

12-31. a) $-0.595 \le \beta_2 \le 0.535$

$$0.229 \le \beta_3 \le 0.812$$
$$-0.216 \le \beta_4 \le 0.013$$
$$-7.2968 \le \beta_5 \le 2.9756$$

b) $\hat{\mu}_{Y|x_0} = 8.99568$ $se(\hat{\mu}_{Y|x_0}) = 0.472445$ $t_{.025,14} = 2.145$

$$\hat{\mu}_{Y|x_0} \pm t_{\alpha/2,n-p} \ se(\hat{\mu}_{Y|x_0})$$
$$8.99568 \pm (2.145)(0.472445)$$
$$7.982 \le \mu_{Y|x_0} \le 10.009$$

c) $\hat{y}_0 = 8.99568$ $se(\hat{y}_0) = 100121$

$$8.99568 \pm 2.145(1.00121)$$
$$6.8481 \le y_0 \le 11.143$$

12-32. a) $-20.477 \le \beta_1 \le 1.269$

$-0.245 \le \beta_2 \le 1.076$

$14.428 \le \beta_3 \le 22.159$

b) $\hat{\mu}_{Y|x_0} = 91.372$ $se(\hat{y}_0) = 4.721$ $t_{.005,16} = 2.921$

$91.372 \pm 2.921(4.721)$

$77.582 \le y_0 \le 105.162$

c) $\hat{\mu}_{Y|x_0} = 91.372$ $se(\hat{\mu}_{Y|x_0}) = 3.163$

$91.372 \pm (2.921)(3.163)$

$82.133 \le \mu_{Y|x_0} \le 100.611$

12-33. a) $-0.000042 \le \beta_1 \le 0.000099$

$0.00196 \le \beta_2 \le 0.00268$

$-0.01504 \le \beta_3 \le 0.00898$

$0.00597 \le \beta_4 \le 0.010979$

$-0.01897 \le \beta_5 \le 0.01424$

b) $\hat{\mu}_{Y|x_0} = 0.022466$ $se(\hat{\mu}_{Y|x_0}) = 0.000595$ $t_{.025,26} = 2.056$

$0.0220086 \pm (2.056)(0.000595)$

$0.0212 \le \mu_{Y|x_0} \le 0.0237$

12-34. a) $\hat{\mu}_{Y|x_0} = 0.0171$ $se(\hat{\mu}_{Y|x_0}) = 0.000548$ $t_{.025,29} = 2.045$

$0.0171 \pm (2.045)(0.000548)$

$0.0159 \le \mu_{Y|x_0} \le 0.0183$

b) : width = 2.2 E-3
: width = 2.4 E-3
The interaction model has a shorter confidence interval. Yes, this suggests the interaction model is preferable.

12-35. a) $0.3882 \le \beta_1 \le 0.5998$

b) $\hat{y} = -5.767703 + 0.496501x_1$

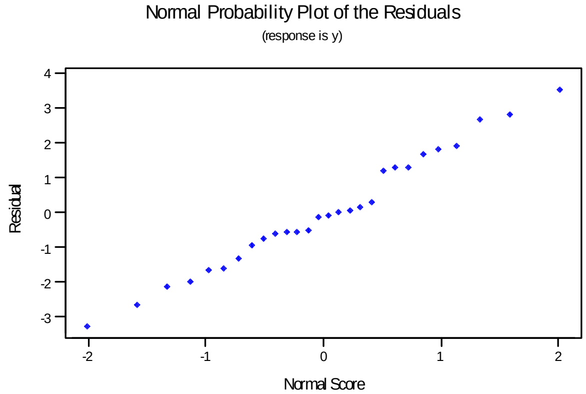
c) $0.4648 \le \beta_1 \le 0.5282$

d) The simple linear regression model has the shorter interval. Yes, the simple linear regression model in this case is preferable.

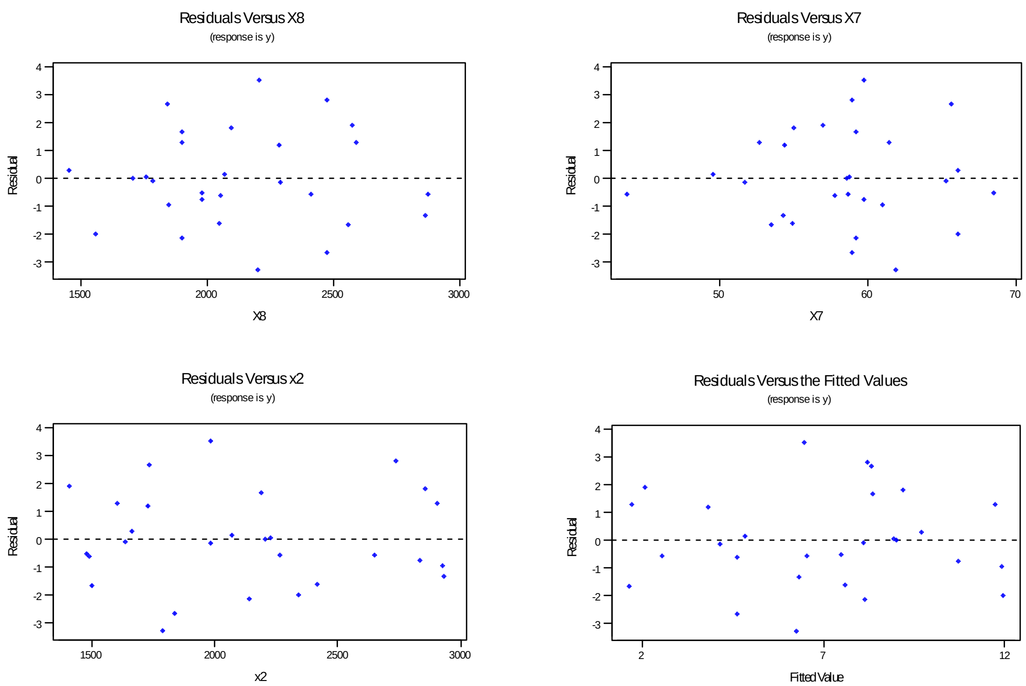
Section 12-5

12-36. a) $R^2 = 79.1$ %

b) Assumption of normality appears adequate.



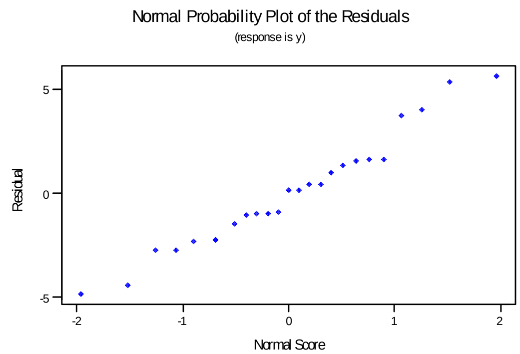
c) Model appears adequate. Some suggestion of nonconstant variance in the plot of x_7.



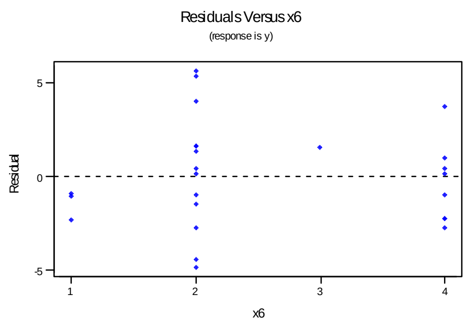
d) No, none of the observations has a Cook's distance greater than 1.

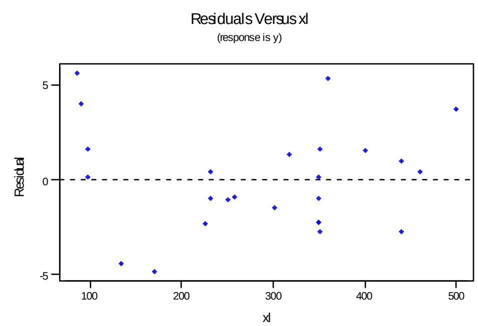
12-37. a) $R^2 = 0.9849$

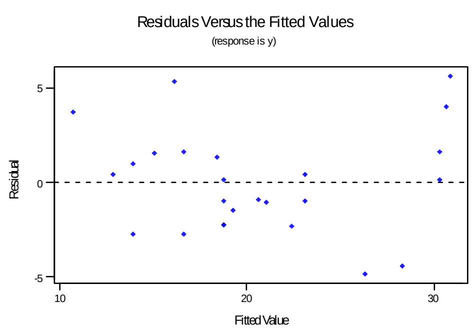
b) Normality assumption appears valid.



c) Assumption of constant variance appears reasonable.







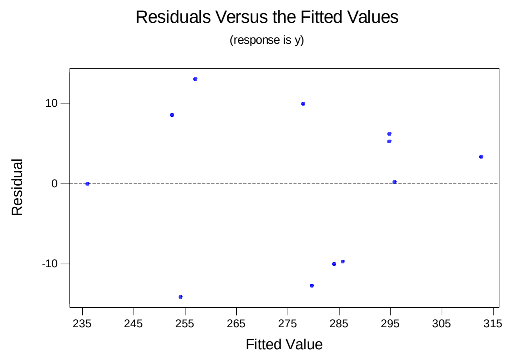
d) No, none of the Cook's distance vales are greater than 1.

COOK1

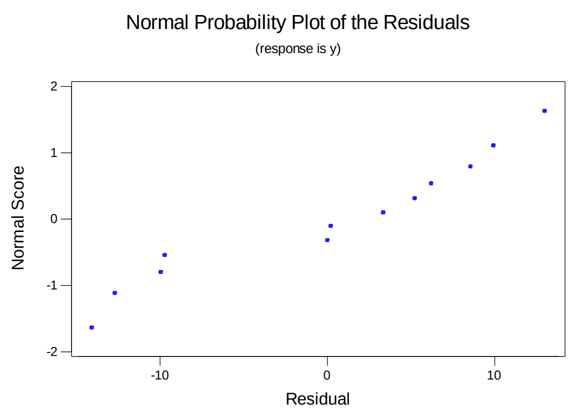
0.000155	0.008206	0.017117	0.039491	0.043651	0.002113	0.160297	0.000034
0.030261	0.329074	0.087966	0.006384	0.006445	0.140511	0.006753	0.005328
0.000583	0.190330	0.001767	0.022620	0.001817	0.116019	0.007782	0.039944
0.030261							

12-38. a) $R^2 = 0.852$

b) The residual plots look reasonable. This is some increase in variability at the middle of the predicted values.



c) Normality assumption is reasonable.

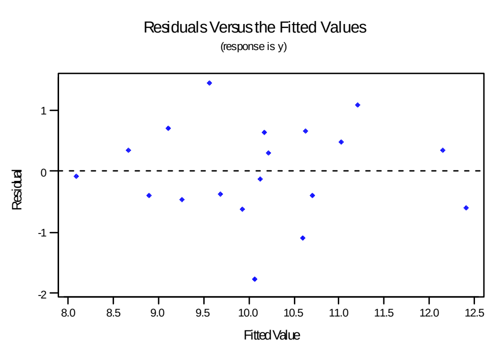


12-39. a) $R^2 = 0.985$

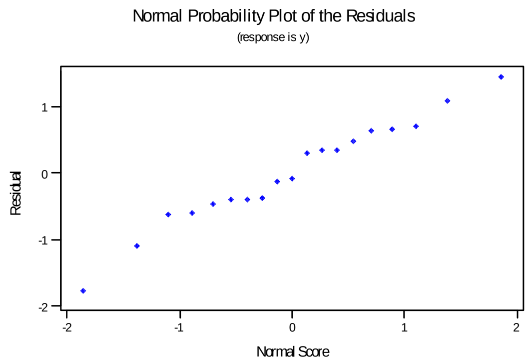
b) $R^2 = 0.99$

R^2 increases with addition of interaction term. No, adding additional regressor will always increase r^2

12-40. a) Some indication of variability increasing with predicted value.

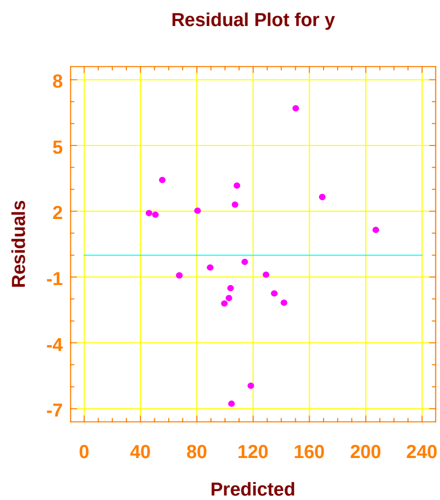


b) No, the normality assumption appears reasonable.



c) No, none of the values are influential. No Cook's distance is greater than one.

12-41. a) There is some indication of nonconstant variance since the residuals appear to "fan out" with increasing values of y.



b)

```
Source               Sum of Squares     DF     Mean Square       F-Ratio    P-value
Model                       30531.5      3         10177.2       840.546      .0000
Error                       193.725     16         12.1078
Total (Corr.)               30725.2     19

R-squared = 0.993695                              Stnd. error of est. = 3.47963
R-squared (Adj. for d.f.) = 0.992513            Durbin-Watson statistic = 1.77758
```

$R^2 = 0.9937$ or 99.37 %;

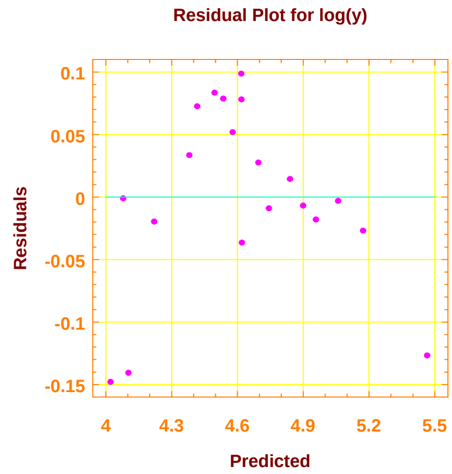
$R^2_{Adj} = 0.9925$ or 99.25%;

c)

```
                     Model fitting results for: log(y)
--------------------------------------------------------------------------------
Independent variable            coefficient  std. error      t-value   sig.level
CONSTANT                            6.22489    1.124522       5.5356      0.0000
x1                                 -0.16647    0.083727      -1.9882      0.0642
x2                                -0.000228    0.005079      -0.0448      0.9648
x3                                 0.157312    0.029752       5.2875      0.0001
--------------------------------------------------------------------------------
R-SQ. (ADJ.) = 0.9574  SE=       0.078919  MAE=       0.053775  DurbWat=  2.031
Previously:    0.0000            0.000000             0.000000            0.000
20 observations fitted, forecast(s) computed for 0 missing val. of dep. var.
```

$\hat{y}^* = 6.22489 - 0.16647x_1 - 0.000228x_2 + 0.157312x_3$

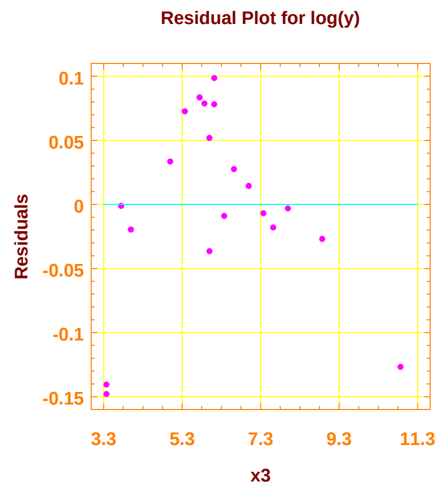
d)



Plot exhibits curvature

There is curvature in the plot. The plot does not give much more information as to which model is preferable.

e)



Plot exhibits curvature

Variance does not appear constant. Curvature is evident.

f)

Model fitting results for: log(y)

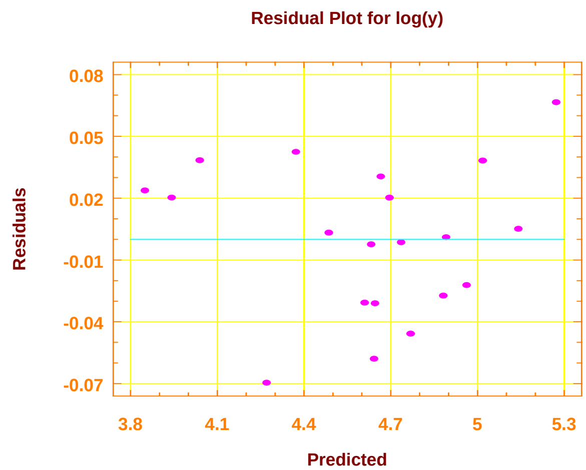
Independent variable	coefficient	std. error	t-value	sig.level
CONSTANT	6.222045	0.547157	11.3716	0.0000
x1	-0.198597	0.034022	-5.8374	0.0000
x2	0.009724	0.001864	5.2180	0.0001
1/x3	-4.436229	0.351293	-12.6283	0.0000

R-SQ. (ADJ.) = 0.9893 SE= 0.039499 MAE= 0.028896 DurbWat= 1.869

Analysis of Variance for the Full Regression

Source	Sum of Squares	DF	Mean Square	F-Ratio	P-value
Model	2.75054	3	0.916847	587.649	.0000
Error	0.0249631	16	0.00156020		
Total (Corr.)	2.77550	19			

R-squared = 0.991006 Stnd. error of est. = 0.0394993
R-squared (Adj. for d.f.) = 0.98932 Durbin-Watson statistic = 1.86891

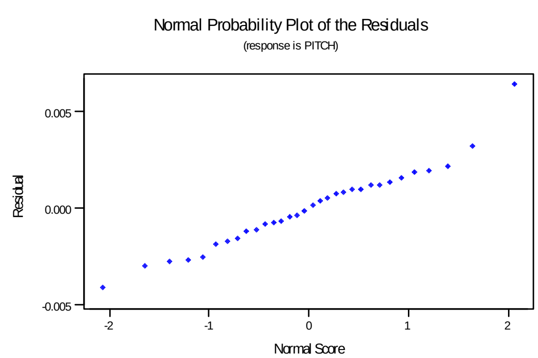


Using 1/x3

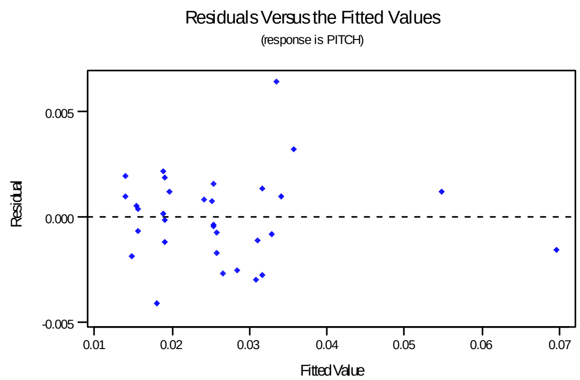
The residual plot indicates better conformance to assumptions.

Curvature is removed when using $1/x_3$ as the regressor instead of x_3 and the log of the response data.

12-42. a) R^2= 0.969
b) Normality is acceptable

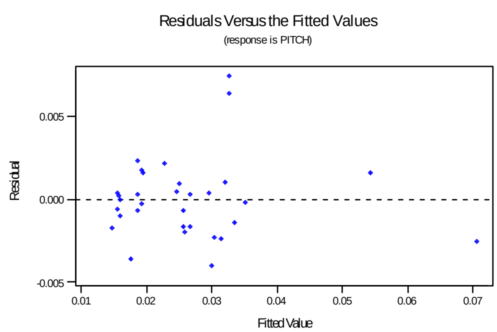


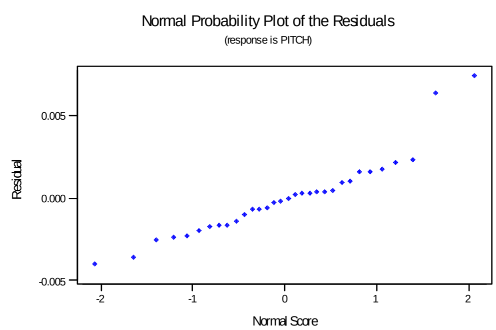
c) Plot is acceptable.



d) Cook's distance values
0.0191 0.0003 0.0026 0.0009 0.0293 0.1112 0.1014 0.0131 0.0076 0.0004 0.0109
0.0000 0.0140 0.0039 0.0002 0.0003 0.0079 0.0022 4.5975 0.0033 0.0058 0.1412
0.0161 0.0268 0.0609 0.0016 0.0029 0.3391 0.3918 0.0134 0.0088 0.5063
The 19th observation is influential

12-43. a) $R^2 = 0.955$. Yes, the R^2 using these two regressors is nearly as large as the R^2 from the model with five regressors.
b) Normality is acceptable, but there is some indication of outliers.

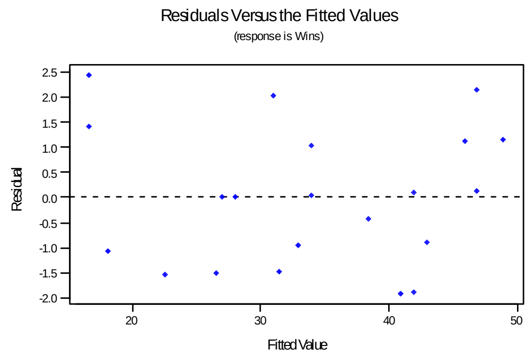




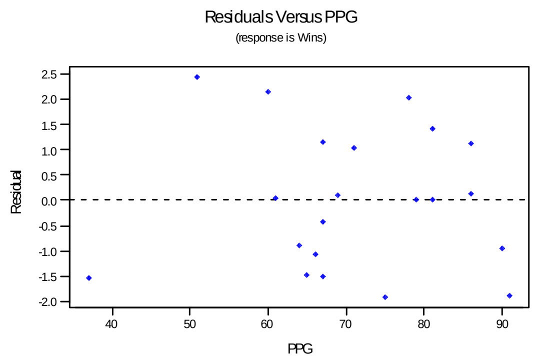
c) Cook's distance values
0.0202 0.0008 0.0021 0.0003 0.0050 0.0000 0.0506 0.0175 0.0015 0.0003 0.0087
0.0001 0.0072 0.0126 0.0004 0.0021 0.0051 0.0007 0.0282 0.0072 0.0004 0.1566
0.0267 0.0006 0.0189 0.0179 0.0055 0.1141 0.1520 0.0001 0.0759 2.3550
The last observation is very influential

12-44. a) $\hat{y} = -5.767703 + 0.496501 x_1$
b) r^2 = 0.982637 or 98.26 %

c) Model appears adequate.



d) No, the residuals do not seem to be related to PPG. If PPG were significant, we would detect some nonrandom pattern in this plot.



12-45. a) $p = k + 1 = 2 + 1 = 3$

Average size = p/n = 3/25 = 0.12

b) Leverage point criteria:

$$h_{ii} > 2(p/n)$$

$$h_{ii} > 2(0.12)$$

$$h_{ii} > 0.24$$

$$h_{17,17} = 0.2593$$

$$h_{18,18} = 0.2929$$

Points 17 and 18 are leverage points.

Sections 12-6

12-46. a) $\hat{y} = -26219.15 + 189.205x - 0.33x^2$

b) $H_0: \beta_j = 0$ for all j

$H_1: \beta_j \neq 0$ for at least one j

$\alpha = 0.05$

$f_0 = 17.2045$

$f_{.05,2,5} = 5.79$

$f_0 > f_{.05,2,5}$

Reject H_0 and conclude that model is significant at $\alpha = 0.05$

c) $H_0: \beta_{11} = 0$
$H_1: \beta_{11} \neq 0$
$\alpha = 0.05$

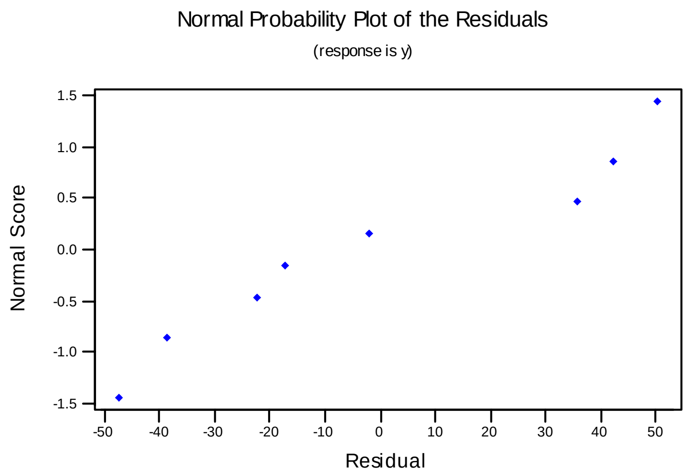
$$t_0 = -2.45$$

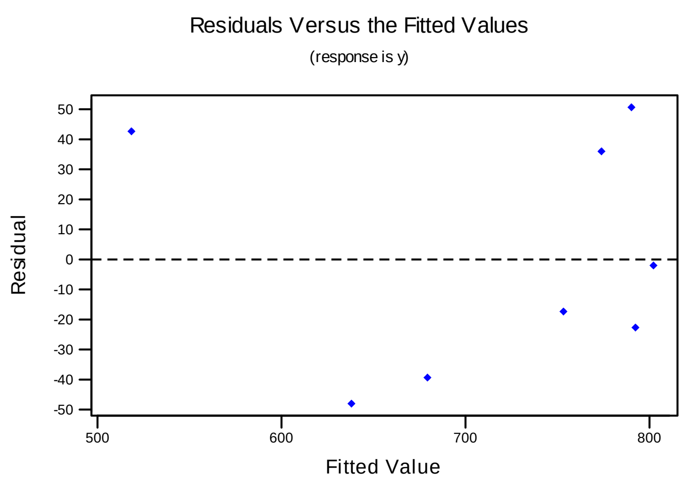
$$t_{\alpha,n-p} = t_{.025,8-3} = t_{.025,5} = 2.571$$

$$|t_0| \ngtr 2.571$$

Do not reject H_0 and conclude insufficient evidence to support value of quadratic term in model at $\alpha = 0.05$

d) One residual is an outlier
Normality assumption appears acceptable
Residuals against fitted values are somewhat unusual, but the impact of the outlier should be considered.



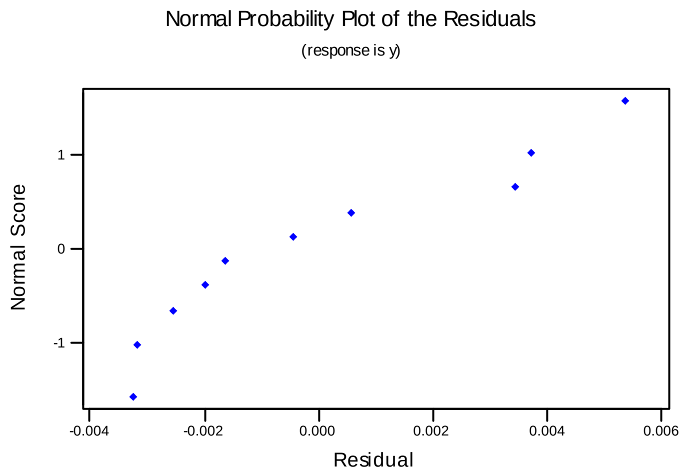


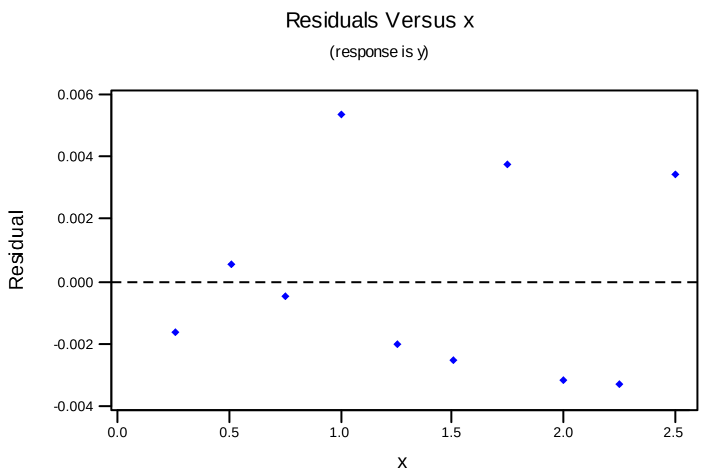
12-47. a) $\hat{y} = -1.633 + 1.232x - 1.495x^2$

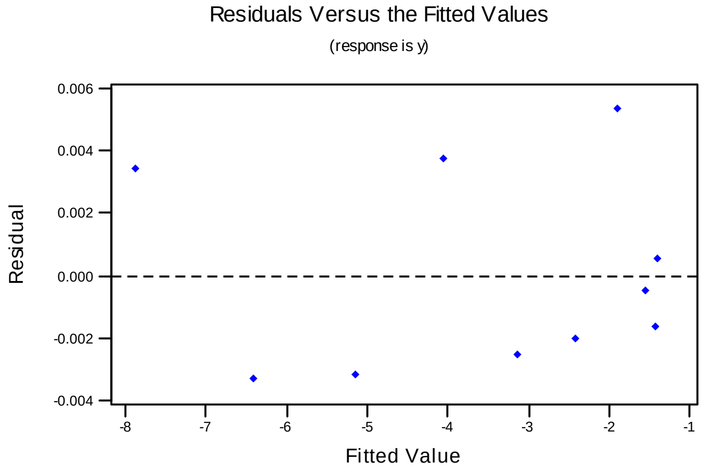
b) f_0 = 1.858613, reject H_0

c) t_0 = –601.64, reject H_0

d) Model is acceptable. Observation number 10 has large leverage.







12-48.

$$\hat{y} = \beta_0^* + \beta_1^* x' + \beta_{11}^* (x')^2$$
$$\hat{y} = 759.395 - 7.607x' - 0.331(x')^2$$
$$\hat{y} = 759.395 - 7.607(x - 297.125) - 0.331(x - 297.125)^2$$
$$\hat{y} = -26202.14 + 189.09x - 0.331x^2$$

12-49. $\hat{y} = 759.395 - 90.783x' - 47.166(x')^2$, where $x' = \frac{x - \bar{x}}{S_x}$

a) At x = 285 $x' = \frac{285 - 297.125}{11.9336} = -1.016$

$$\hat{y} = 759.395 - 90.783(-1.106) - 47.166(-1.106)^2 = 802.943 \text{ psi}$$

b) $\hat{y} = 759.395 - 90.783\left(\frac{x - 297.125}{11.9336}\right) - 47.166\left(\frac{x - 297.125}{11.9336}\right)^2$

$$\hat{y} = 759.395 - 7.607(x - 297.125) - 0.331(x - 297.125)^2$$
$$\hat{y} = -26204.14 + 189.09x - 0.331x^2$$

c) They are the same.

d) $\hat{y}' = 0.385 - 0.847x' - 0.440(x')^2$

where $y' = \frac{y - \bar{y}}{S_y}$ and $x' = \frac{x - \bar{x}}{S_x}$

The "proportion" of total variability explained is the same for both standardized and un-standardized models. Therefore, R^2 is the <u>same</u> for both models.

$y' = \beta_0^* + \beta_1^* x' + \beta_{11}^* (x')^2$ where $y' = \frac{y - \bar{y}}{S_y}$ and $x' = \frac{x - \bar{x}}{S_x}$

$$y' = \beta_0^* + \beta_1^* x' + \beta_{11}^* (x')^2$$

12-50. a) $\hat{y} = -4.46 + 1.38x + 1.47x^2$

b) $H_0 : \beta_j = 0$ for all j

$H_1 : \beta_j \neq 0$ for at least one j

$\alpha = 0.05$

$f_0 = 1044.99$

$f_{.05,2,9} = 4.26$

$f_0 > f_{0.05,2,9}$

Reject H_0 and conclude regression model is significant at $\alpha = 0.05$

c) $H_0 : \beta_{11} = 0$

$H_1 : \beta_{11} \neq 0$ $\alpha = 0.05$

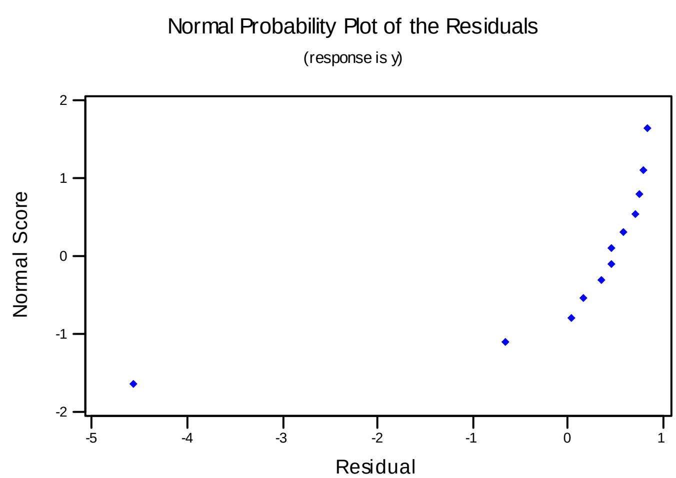
$t_0 = 2.97$

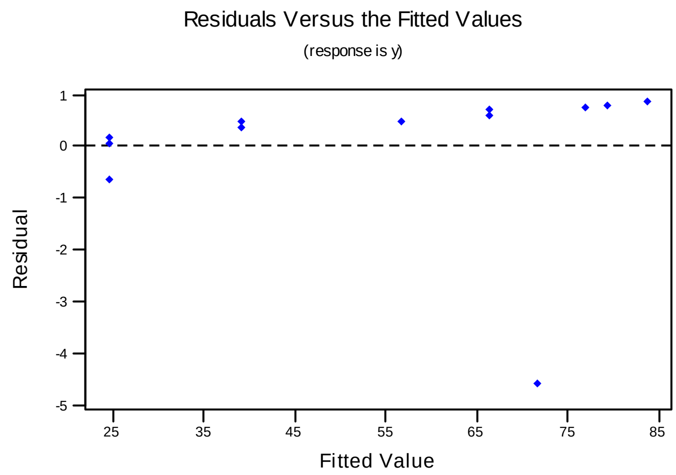
$t_{.025,9} = 2.262$

$|t_0| > t_{0.025,9}$

Reject H_0 and conclude that β_{11} is significant at $\alpha = 0.05$

d) Observation number 9 is an extreme outlier.





e) $\hat{y} = -87.36 + 48.01x - 7.04x^2 + 0.51x^3$

$H_0 : \beta_{33} = 0$

$H_1 : \beta_{33} \neq 0 \quad \alpha = 0.05$

$t_0 = 0.91$

$t_{.025,8} = 2.306$

$|t_0| \ngtr t_{0.025,8}$

Do not reject H_0 and conclude that cubic term is not significant at $\alpha = 0.05$

12-51a)

Predictor	Coef	SE Coef	T	P
Constant	-1.769	1.287	-1.37	0.188
xl	0.4208	0.2942	1.43	0.172
x2	0.2225	0.1307	1.70	0.108
x3	-0.12800	0.07025	-1.82	0.087
x1x2	-0.01988	0.01204	-1.65	0.118
x1x3	0.009151	0.007621	1.20	0.247
x2x3	0.002576	0.007039	0.37	0.719
x1^2	-0.01932	0.01680	-1.15	0.267
x2^2	-0.00745	0.01205	-0.62	0.545
x3^3	0.000824	0.001441	0.57	0.575

S = 0.06092 R-Sq = 91.7% R-Sq(adj) = 87.0%

Analysis of Variance

Source	DF	SS	MS	F	P
Regression	9	0.655671	0.072852	19.63	0.000
Residual Error	16	0.059386	0.003712		
Total	25	0.715057			

$$\hat{y} = -1.769 + 0.421x_1 + 0.222x_2 - 0.128x_3 - 0.02x_{12} + 0.009x_{13} + \\ 0.003x_{23} - 0.019x_1{}^2 - 0.007x_2{}^2 + 0.001x_3{}^2$$

b) H_0 all $\beta_1 = \beta_2 = \beta_3 = \cdots \beta_{23} = 0$

H_1 at least one $\beta_j \neq 0$

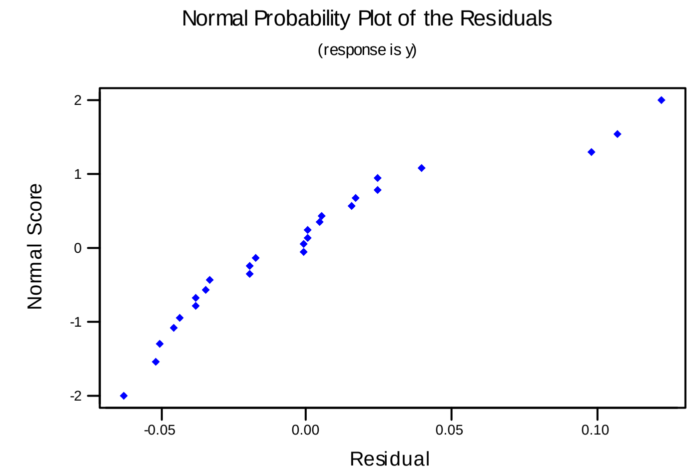
$f_0 = 19.628$

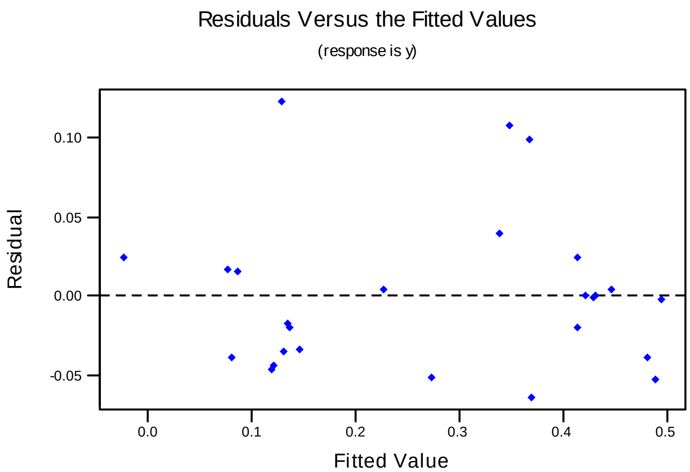
$f_{.05,9,16} = 2.54$

$f_0 > f_{0.05,9,16}$

Reject H_0 and conclude that the model is significant at $\alpha = 0.05$

c) Model is acceptable.





d) $H_0: \beta_{11} = \beta_{22} = \beta_{33} = \beta_{12} = \beta_{13} = \beta_{23} = 0$

H_1 at least one $\beta \neq 0$

$$f_0 = \frac{SS_R(\beta_{11}, \beta_{22}, \beta_{33}, \beta_{12}, \beta_{13}, \beta_{23} \mid \beta_1, \beta_2, \beta_3, \beta_0)/r}{MS_E} = \frac{\frac{0.0359}{6}}{0.003712} = 1.612$$

$f_{.05,6,16} = 2.74$

$f_0 \not> f_{.05,6,16}$

Do not reject H_0

$$\begin{aligned} SS_R(\beta_{11}\beta_{22}\beta_{33}\beta_{12}\beta_{13}\beta_{23} \mid \beta_1\beta_2\beta_3\beta_0) &= SS_R(\beta_{11}\beta_{22}\beta_{33}\beta_{12}\beta_{13}\beta_{23}\beta_1\beta_2\beta_3 \mid \beta_0) - SS_R(\beta_1\beta_2\beta_3 \mid \beta_0) \\ &= 0.65567068 - 0.619763 \\ &= 0.0359 \end{aligned}$$

Reduced Model: $y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3$

12-52. a) Use indicator variable for transmission type

$$x_{11} = \begin{cases} 0 & \text{for manual} \\ 1 & \text{for automatic} \end{cases}$$

b) $\hat{y} = 32.77 - 0.06x_1 + 0.05x_2 + 0.12x_{11}$

c) $H_0: \beta_{11} = 0$

$H_1: \beta_{11} \neq 0 \quad \alpha = 0.05$

$t_0 = 0.05$

$t_{.025,21} = 2.08$

$|t_0| \not> t_{\alpha/2,21}$

Do not reject H_0. No evidence that transmission type affects mileage.

12-53. $y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_{12} x_{12}$

$\hat{y} = 11.503 + 0.153x_1 - 6.094x_2 - 0.031x_{12}$

where $x_2 = \begin{cases} 0 & \text{for tool type } 302 \\ 1 & \text{for tool type } 416 \end{cases}$

Test of different slopes:

$H_0: \beta_{12} = 0$

$H_1: \beta_{12} \neq 0 \quad \alpha = 0.05$

$t_0 = -1.79$

$t_{.025,16} = 2.12$

$|t_0| \not> t_{0.025,16}$

Do not reject H_0. Conclude that data is insufficient to claim that (2) regression models are needed.

Test of different intercepts and slopes using extra sums of squares:

$H_0: \beta_2 = \beta_{12} = 0$

H_1 at least one is not zero

$$SS(\beta_2, \beta_{12} \mid \beta_0) = SS(\beta_1, \beta_2, \beta_{12} \mid \beta_0) - SS(\beta_1 \mid \beta_0)$$
$$= 1013.35995 - 130.60910$$
$$= 882.7508$$
$$f_0 = \frac{SS(\beta_2, \beta_{12} \mid \beta_0)/2}{MS_E} = \frac{882.7508/2}{0.4059} = 1087.40$$

Reject H_0.

12-54. The default settings for F-to-enter and F-to-remove for Minitab were used. Different settings can change the models generated by the method.

a) The min. MS_E equation is: x_2, x_4, x_7, x_8, x_9

$$MS_E = 2.7726$$
$$C_p = 1.3$$
$$\hat{y} = -2.475 + 0.004x_2 + 0.222x_7 - 0.004x_8 - 0.002x_9$$

The min. C_p equation is :x_2, x_7, x_8

$$C_p = 1.3$$
$$MS_E = 2.845$$
$$\hat{y} = -2.747 + 0.0036x_2 + 0.2035x_7 - 0.0047x_8$$

b), c), & d) Same as the minimum C_p model.

e) The minimum C_p model seems the best because stepwise, forward and backward selection all converged to the same model. Also, the model contains fewer variables than the minimum MS_E model.

12-55. The categorical predictor x_{11} is not used. The default settings for F-to-enter and F-to-remove for Minitab were used. Different settings can change the models generated by the method.

a) The min. MS_E equation is: x_1, x_3, x_4, x_5, x_7, x_8, x_{10}

$MS_E = 6.579$ $\quad C_p = 6.1$

The min. C_p equation is: x_5, x_8, x_{10}

$C_p = 5.4$ $\quad MS_E = 7.97$

b) $\hat{y} = 34.434 - 0.048x_1$

$MS_E = 8.82$ $\quad C_p = 6.0$

c) Model is: x_1, x_6, $\boldsymbol{x}_3$ with $C_p = 5.9$ and $MS_E = 8.18$

d) Same as the minimum C_p model: $\hat{y} = 0.341 + 2.862x_5 + 0.246x_8 - 0.010x_{10}$

$C_p = 5.4$ $\quad MS_E = 7.97$

e) Minimum C_p and backward elimination result in the same model. Stepwise and forward selection result in the same model. Because it is much simpler, the stepwise model seems preferable. The minimum C_p model with three predictors is not unreasonable.

12-56. The default settings for F-to-enter and F-to-remove for Minitab were used. Different settings can change the models generated by the method.

a) The min MS_E equation is: x_1, x_2, x_3

$C_p = 3.8$ $\quad MS_E = 134.6$

$$\hat{y} = -162.1 + 0.7487x_1 + 7.691x_2 + 2.343x_3$$

The min C_p equation is: x_1, x_2,

$C_p = 3.4$ $\quad MS_E = 145.7$

$$\hat{y} = 3.92 + 0.5727x_1 + 9.882x_2$$

b) Same as the min C_p equation in part (a)

c) Same as part min MS_E equation in part (a)
d) Same as part min C_p equation in part (a)
e) The minimum MS_E and forward models all are the same. Stepwise and backward regressions generate the minimum C_p model. The minimum C_p model has fewer regressors and it might be preferred, but MS_E has increased.

12-57. The default settings for F-to-enter and F-to-remove for Minitab were used. Different settings can change the models generated by the method.

a) $\hat{y} = 4.656 + 0.511x_3 - 0.124x_4$
b) Same as part a.
c) Same as part a.
d) All models are the same.

12-58. The default settings for F-to-enter and F-to-remove for Minitab were used. Different settings can change the models generated by the method.

a) $\hat{y} = -3.517 + 0.486x_1 - 0.156x_9$
b) Same as part (a).
c) Same as part (a).
d) All models are the same.

12-59 The default settings for F-to-enter and F-to-remove for Minitab were used. Different settings can change the models generated by the method.

a) $\hat{y} = -0.304 + 0.083x_1 - 0.031x_3 + 0.004x_2^2$

$$C_p = 4.04 \quad MS_E = 0.004$$

b) $\hat{y} = -0.256 + 0.078x_1 + 0.022x_2 - 0.042x_3 + 0.0008x_3^2$

$$C_p = 4.66 \quad MS_E = 0.004$$

c) The forward selection model in part (a) is more parsimonious with a lower C_p and equivalent MS_E. Therefore, we prefer the model in part (a).

12-60. The min. C_p equation is: x_3, x_4

$\hat{y} = 4.656 + 0.511x_3 - 0.124x_4$

Adjusted $R^2 = 0.603$. Min. $C_p = 1.65$

Max. adjusted R^2 equation is : x_1, x_3, x_4

$\hat{y} = 2.419 + 0.553x_1 + 0.479x_3 - 0.123x_4$

Adjusted $R^2 = 0.607$, $C_p = 2.61$

No, the maximum adjusted R^2 model has x_1 in addition to x_3 and x_4 in the model. However, the adjusted R^2 values are very close.

12-61. a) The min. C_p model is $\hat{y} = -3.517 + 0.486x_1 - 0.156x_9$ and $C_p = -1.67$

b) Min MS_E model is x_1, x_7, x_9, $MS_E = 1.67$, $C_p = -0.77$

$\hat{y} = -5.964 + 0.495x_1 + 0.025x_7 - 0.163x_9$

c) Maximum adjusted R^2 model is x_1, x_7, x_9, Adj. $R^2 = 0.98448$. Yes, same as minimum the MS_E model.

12-62. n = 30, k = 9, p =9 + 1 = 10 in full model.

a) $\hat{\sigma}^2 = MS_E = 100$ $\qquad R^2 = 0.92$

$$R^2 = \frac{SS_R}{S_{yy}} = 1 - \frac{SS_E}{S_{yy}}$$

$$\begin{aligned} SS_E &= MS_E(n-p) \\ &= 100(30-10) \\ &= 2000 \end{aligned}$$

$$0.92 = 1 - \frac{2000}{S_{yy}}$$

$$25000 = S_{yy}$$

$$SS_R = S_{yy} - SS_E$$

$$= 25000 - 2000 = 23000$$

$$MS_R = \frac{SS_R}{k} = \frac{23000}{9} = 2555.56$$

$$f_0 = \frac{MS_R}{MS_E} = \frac{2555.56}{100} = 25.56$$

$$f_{.05,9,20} = 2.39$$

$$f_0 > f_{\alpha,9,20}$$

Reject H_0 and conclude at least one β_j is significant at α = 0.05.

b) k = 4 p = 5 $SS_E = 2200$

$$MS_E = \frac{SS_E}{n-p} = \frac{2200}{30-5} = 88$$

Yes, MS_E is reduced with new model (k = 4).

c) $C_p = \dfrac{SS_E(p)}{\hat{\sigma}^2} - n + 2p$ $C_p = \dfrac{2200}{100} - 30 + 2(5) = 2$

Yes, C_p is reduced from the full model.

12-63. n = 25 k = 7 p = 8 $MS_{E(\text{full})} = 10$

a) p =4 $SS_E = 300$

$$MS_E = \frac{SS_E}{n-p} = \frac{300}{25-4} = 14.29$$

$$C_p = \frac{SS_E}{MS_{E(full)}} - n + 2p$$

$$= \frac{300}{10} - 25 + 2(4)$$

$$= 5 + 8 = 13$$

YES, $C_p > p$

b) p = 5 $SS_E = 275$

$$MS_E = \frac{SS_E}{n-p} = \frac{275}{30-5} = 11 \qquad C_p = \frac{275}{10} - 25 + 2(5) = 12.5$$

Yes, both MS_E and C_p are reduced.

<u>Supplemental Exercises</u>

12-64. a) $\hat{y} = 4203 - 0.231x_3 + 21.485x_4 + 1.6887x_5$

b) $H_0: \beta_3 = \beta_4 = \beta_5 = 0$

$H_1: \beta_j \neq 0$ for at least one j

α = 0.01 $f_0 = 1651.25$

$f_{.01,3,36} = 4.38$

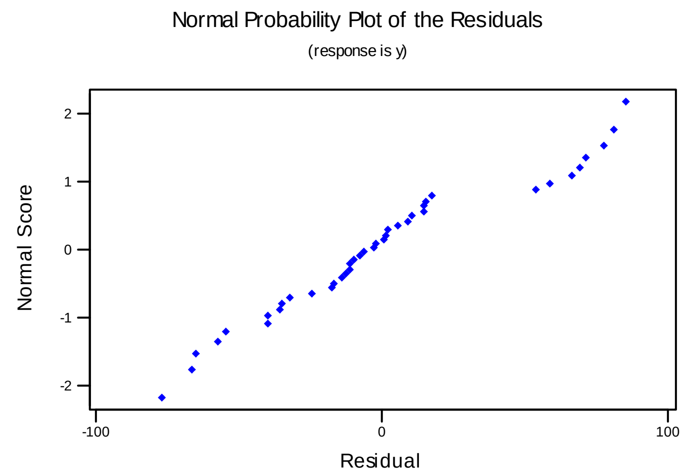
Reject H_0 and conclude that regression is significant. P-value < 0.00001

c) All at $\alpha = 0.01$ $\quad t_{.005,36} = 2.72$

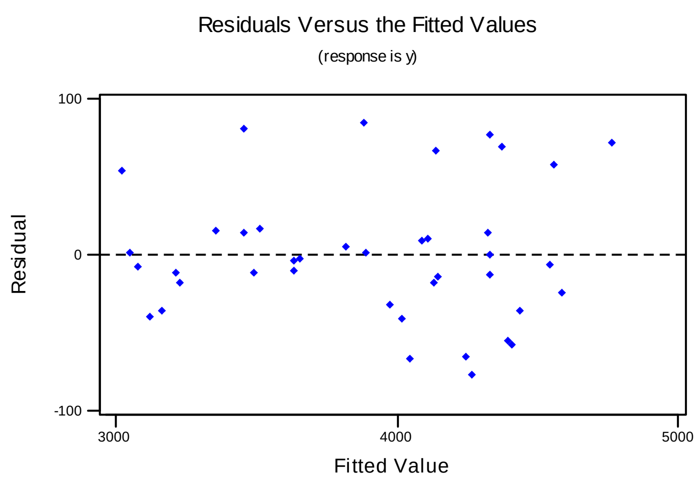
$H_0: \beta_3 = 0$	$H_0: \beta_4 = 0$	$H_0: \beta_5 = 0$
$H_1: \beta_3 \neq 0$	$H_1: \beta_4 \neq 0$	$H_1: \beta_5 \neq 0$
$t_0 = -2.06$	$t_0 = 22.91$	$t_0 = 3.00$
$\lvert t_0 \rvert \not> t_{\alpha/2,36}$	$\lvert t_0 \rvert > t_{\alpha/2,36}$	$\lvert t_0 \rvert > t_{\alpha/2,36}$
Do not reject H_0	Reject H_0	Reject H_0

d) $R^2 = 0.993$ $\qquad$ Adj. $R^2 = 0.9925$

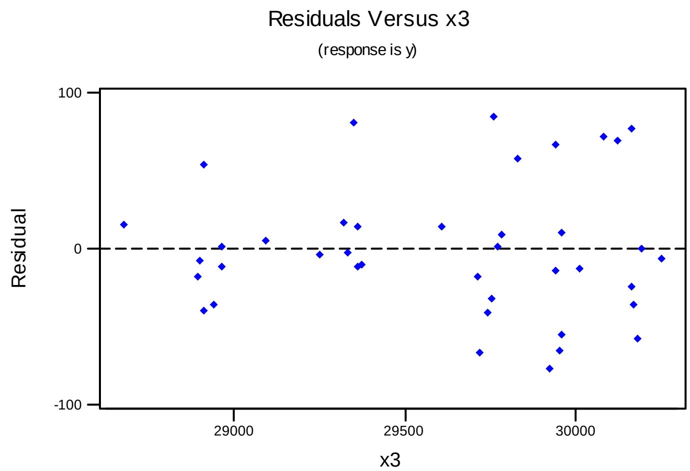
e) Normality assumption appears reasonable. However there is a gap in the line.



f) Plot is satisfactory.



g) Slight indication that variance increases as x_3 increases.



h) $\hat{y} = 4203 - 0.231(1670) + 21.485(170) + 1.6887(1589) = 10153.02$

12-65. a) $H_0\colon \beta_3^* = \beta_4 = \beta_5 = 0$

$H_1\colon \beta_j \neq 0$ for at least one j

$\alpha = 0.01$

$f_0 = 1321.39$

$f_{.01,3,36} = 4.38$

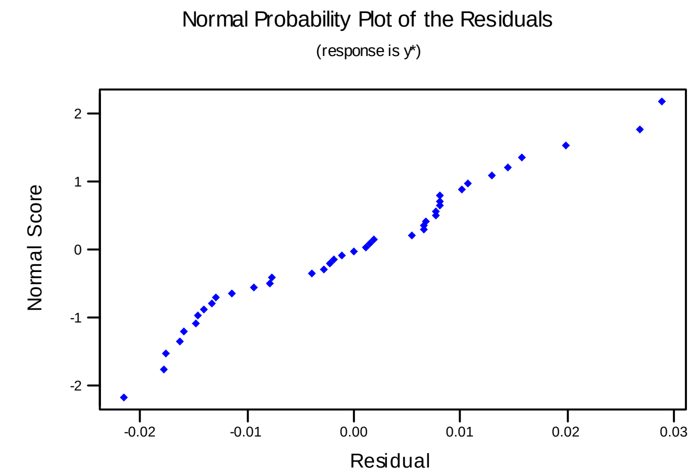
$f_0 >> f_{\alpha,3,36}$

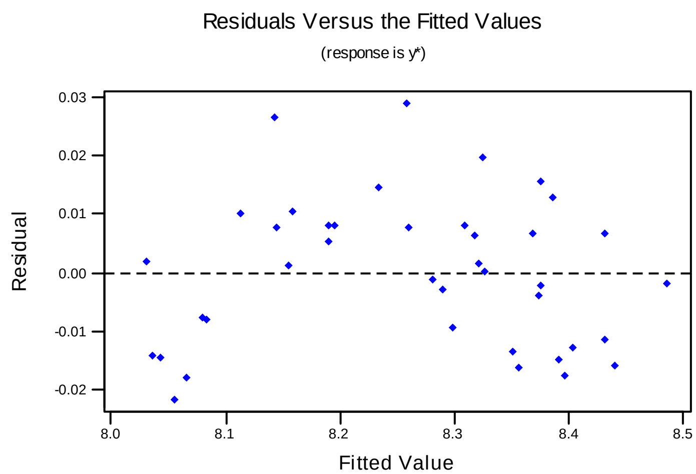
Reject H_0 and conclude that regression is significant. P-value < 0.00001

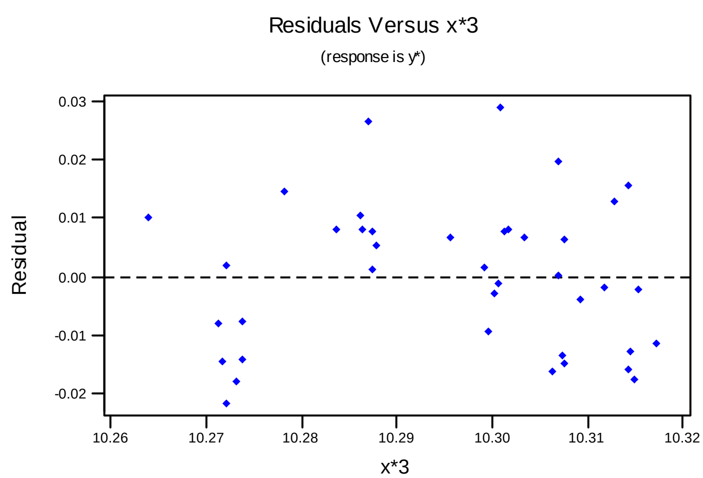
b) $\alpha = 0.01$ $t_{.005,36} = 2.72$

$H_0\colon \beta_3^* = 0$	$H_0\colon \beta_4 = 0$	$H_0\colon \beta_5 = 0$
$H_1\colon \beta_3^* \neq 0$	$H_1\colon \beta_4 \neq 0$	$H_1\colon \beta_5 \neq 0$
$t_0 = -1.45$	$t_0 = 19.95$	$t_0 = 2.53$
$\lvert t_0 \rvert \not> t_{\alpha/2,36}$	$\lvert t_0 \rvert > t_{\alpha/2,36}$	$\lvert t_0 \rvert \not> t_{\alpha/2,36}$
Do not reject H_0	Reject H_0	Do not reject H_0

c) Curvature is evident in the residuals plots from this model.


Normal Probability Plot of the Residuals
(response is y*)
Normal Score
Residual
2
1
0
-1
-2
-0.02
-0.01
0.00
0.01
0.02
0.03


Residuals Versus the Fitted Values
(response is y*)
Residual
Fitted Value
0.03
0.02
0.01
0.00
-0.01
-0.02
8.0
8.1
8.2
8.3
8.4
8.5


Residuals Versus x*3
(response is y*)
Residual
x*3
0.03
0.02
0.01
0.00
-0.01
-0.02
10.26
10.27
10.28
10.29
10.30
10.31
10.32

12-66. Data is on page 464, not in Table 12-20

a) $\hat{y} = 2.630 - 0.285x_1 + 0.207x_2 + 0.456x_3 - 0.568x_4 + 0.005x_5$

$H_0 : \beta_1 = \beta_2 = \beta_3 = \beta_4 = \beta_5 = 0$

H_1 : at least one $\beta_j \neq 0$

$\alpha = 0.01$

$f_0 = 4.63$

$f_{.01,5,19} = 4.17$

$f_0 > f_{\alpha,5,19}$

Reject H_0. *P*-value = 0.006

b) $\alpha = 0.05$ $t_{.025,19} = 2.093$

$H_0:\beta_1 = 0$	$H_0:\beta_2 = 0$	$H_0:\beta_3 = 0$	$H_0:\beta_4 = 0$	$H_0:\beta_5 = 0$
$H_1:\beta_1 \neq 0$	$H_1:\beta_2 \neq 0$	$H_1:\beta_3 \neq 0$	$H_1:\beta_4 \neq 0$	$H_1:\beta_5 \neq 0$
$t_0 = -2.41$	$t_0 = 2.73$	$t_0 = 2.40$	$t_0 = -2.70$	$t_0 = 0.29$
$\lvert t_0\rvert > t_{\alpha/2,19}$	$\lvert t_0\rvert > t_{\alpha/2,19}$	$\lvert t_0\rvert > t_{\alpha/2,19}$	$\lvert t_0\rvert > t_{\alpha/2,19}$	$\lvert t_0\rvert \not> t_{\alpha/2,19}$
Reject H_0	Reject H_0	Reject H_0	Reject H_0	Do not reject H_0

c) $\hat{y} = 2.957 - 0.284x_1 + 0.19996x_2 + 0.457x_3 - 0.584x_4$

$H_0 : \beta_1 = \beta_2 = \beta_3 = \beta_4 = 0$

$H_1 : \beta_j \neq 0$ for at least one j

$\alpha = 0.05$

$f_0 = 6.05$

$f_{.05,4,20} = 2.87$

$f_0 > f_{\alpha,4,20}$

Reject H_0.

$\alpha = 0.05$ $t_{.025,20} = 2.086$

$H_0:\beta_1 = 0$	$H_0:\beta_2 = 0$	$H_0:\beta_3 = 0$	$H_0:\beta_4 = 0$
$H_1:\beta_1 \neq 0$	$H_1:\beta_2 \neq 0$	$H_1:\beta_3 \neq 0$	$H_1:\beta_4 \neq 0$
$t_0 = -2.46$	$t_0 = 2.87$	$t_0 = 2.46$	$t_0 = -2.95$
$\lvert t_0\rvert > t_{\alpha/2,20}$	$\lvert t_0\rvert > t_{\alpha/2,20}$	$\lvert t_0\rvert > t_{\alpha/2,20}$	$\lvert t_0\rvert > t_{\alpha/2,20}$
Reject H_0	Reject H_0	Reject H_0	Reject H_0

d) The addition of the 5th regressor results in a loss of one degree of freedom in the denominator and the reduction in SS_E is not enough to compensate for this loss.

e) Observation 2 is unusually large. Studentized residuals

-0.75442 -3.38260 -0.39956 1.90117 -0.52450 0.62397
-0.42785 2.01822 1.35265 -0.44692 0.73052
-0.29383 -0.32938 1.81797 0.06052 -0.95465
-0.74173 0.30302 0.61565 0.61191 -0.09931
0.77652 -0.71284 -0.81719 -0.48090

f) R^2 for model in part (a): 0.549. R^2 for model in part (c): 0.547. R^2 for model x_1, x_2, x_3, x_4 without obs. #2: 0.802. R^2 increased because observation 2 was not fit well by either of the previous models.

g) $H_0 : \beta_1 = \beta_2 = \beta_3 = \beta_4 = 0$

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

$f_0 = 19.24$

$f_{.05,4,19} = 2.90$

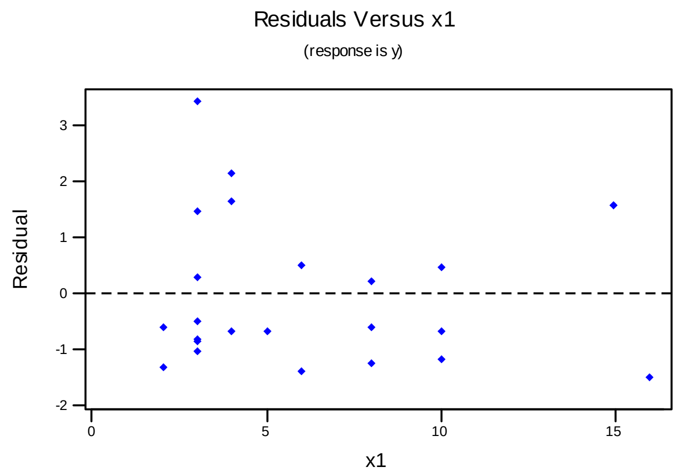
$f_0 > f_{0.05,4,19}$

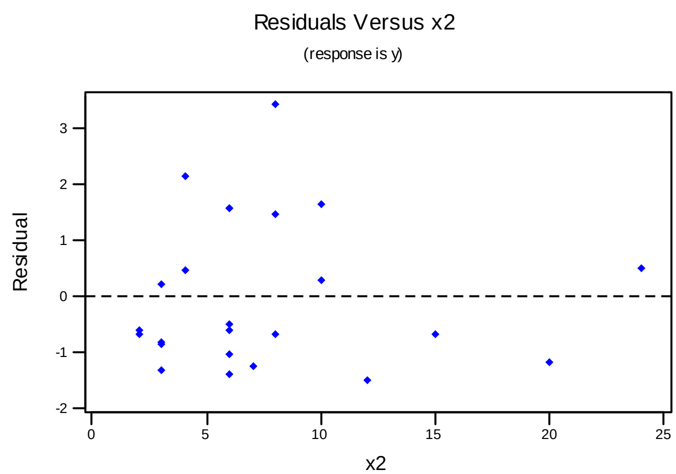
Reject H_0.

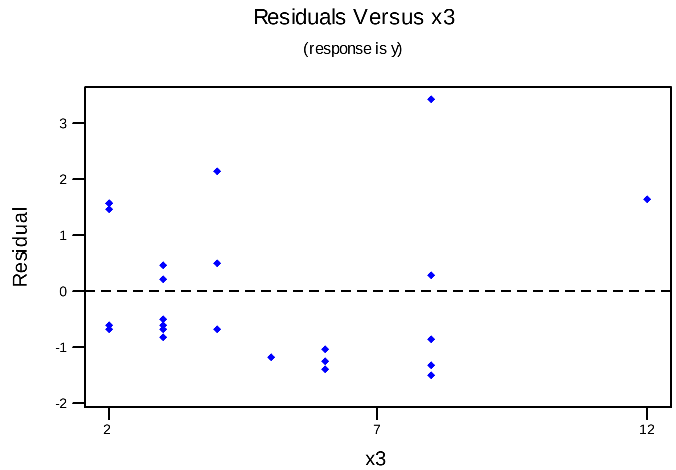
$\alpha = 0.05$ $\quad t_{.025,19} = 2.093$

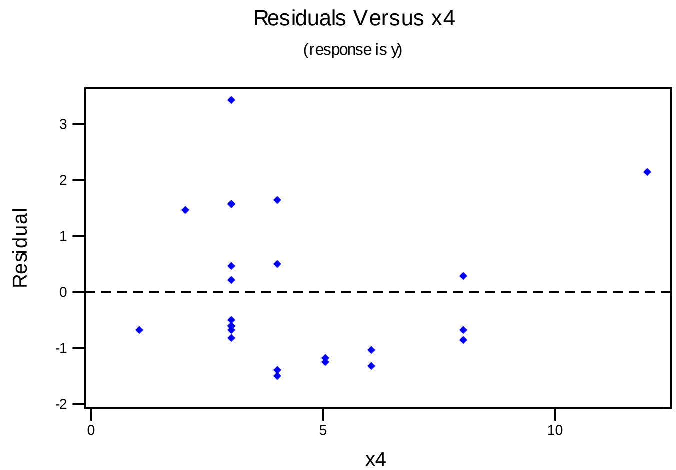
$H_0: \beta_1 = 0$	$H_0: \beta_2 = 0$	$H_0: \beta_3 = 0$	$H_0: \beta_4 = 0$
$H_1: \beta_1 \neq 0$	$H_1: \beta_2 \neq 0$	$H_1: \beta_3 \neq 0$	$H_1: \beta_4 \neq 0$
$t_0 = -3.90$	$t_0 = 6.44$	$t_0 = 3.63$	$t_0 = -3.33$
$\lvert t_0 \rvert > t_{0.025,19}$	$\lvert t_0 \rvert > t_{0.025,19}$	$\lvert t_0 \rvert > t_{0.025,19}$	$\lvert t_0 \rvert > t_{0.025,19}$
Reject H_0	Reject H_0	Reject H_0	Reject H_0

h) There is some indication of curvature.









12-67. Data is on page 464, not in Table 12-20

a) $\hat{y} = -1.060 + 5.509x_1^* + 1.153x_2^* - 3.989x_3^* - 1.102x_4^*$

b) $H_0 : \beta_1 = \beta_2 = \beta_3 = \beta_4 = 0$

H_1 : at least one $\beta_j \neq 0$

$\alpha = 0.05$

$f_0 = 116.21$

$f_{.05,4,19} = 2.90$

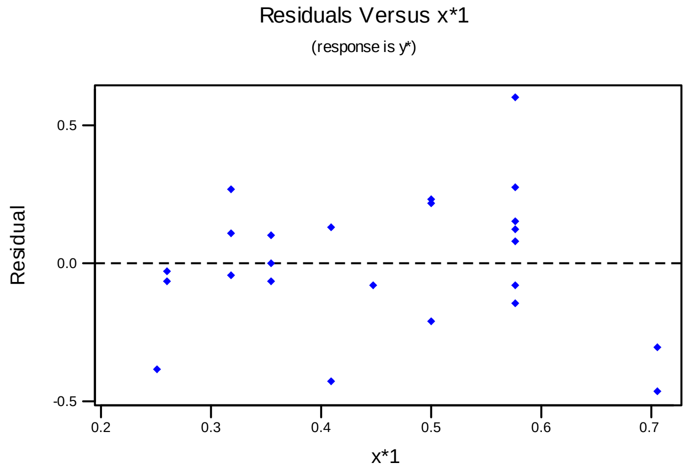
$f_0 >> f_{0.05,4,19}$

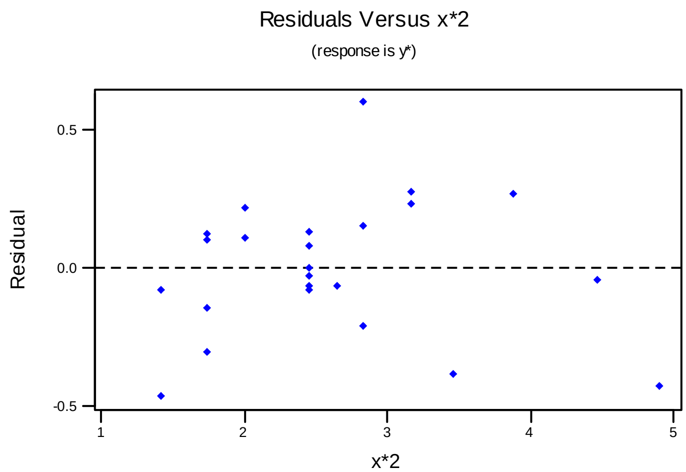
Reject H_0 and conclude that regression is significant at $\alpha = 0.05$.

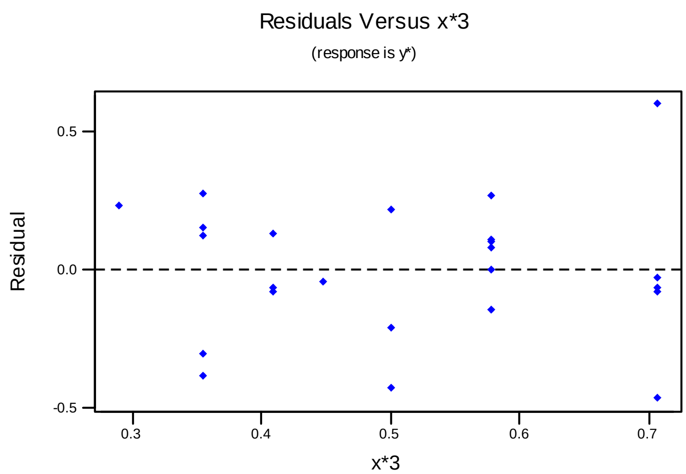
$\alpha = 0.05$ $\quad t_{.025,19} = 2.093$

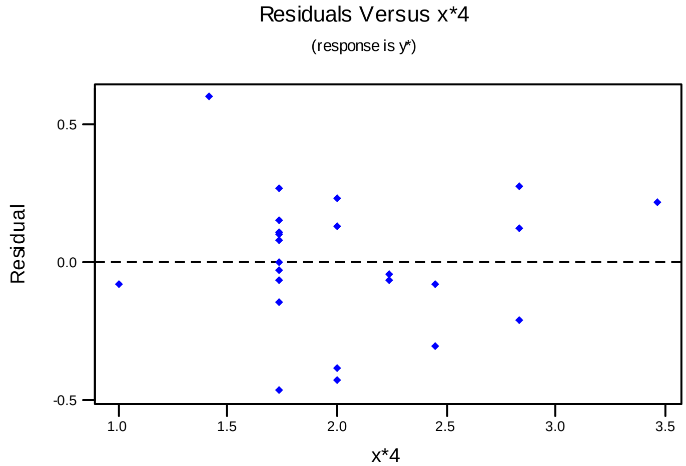
$H_0{:}\beta_1 = 0$	$H_0{:}\beta_2 = 0$	$H_0{:}\beta_3 = 0$	$H_0{:}\beta_4 = 0$
$H_1{:}\beta_1 \neq 0$	$H_1{:}\beta_2 \neq 0$	$H_1{:}\beta_3 \neq 0$	$H_1{:}\beta_4 \neq 0$
$t_0 = 11.67$	$t_0 = 15.46$	$t_0 = -7.28$	$t_0 = -8.43$
$\lvert t_0 \rvert > t_{0.025,19}$	$\lvert t_0 \rvert > t_{0.025,19}$	$\lvert t_0 \rvert > t_{0.025,19}$	$\lvert t_0 \rvert > t_{0.025,19}$
Reject H_0	Reject H_0	Reject H_0	Reject H_0

c) The residual plots are more satisfactory than those in Exercise 12-66.









12-68. a) $\hat{y} = -1709.405 + 2.023x - 0.0006x^2$

b) $H_0 : \beta_1 = \beta_{11} = 0$

H_1 : at least one $\beta_j \neq 0$

$\alpha = 0.05$

$f_0 = 300.11$

$f_{.05,2,7} = 4.74$

$f_0 >> f_{0.05,2,7}$

Reject H_0.

c) $H_0 : \beta_{11} = 0$

$H_1 : \beta_{11} \neq 0$

$\alpha = 0.05$

$$F_0 = \frac{SS_R(\beta_{11} \mid \beta_1)/r}{MS_E} = \frac{2.4276/1}{0.04413}$$

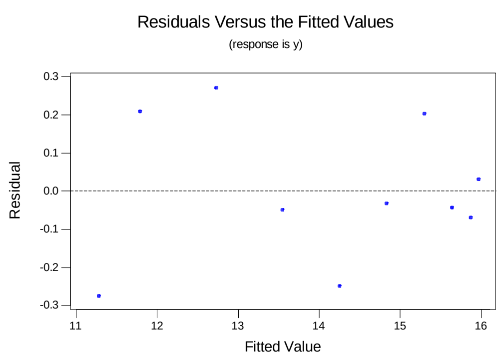
$f_0 = 55.01$

$f_{.05,1,7} = 5.59$

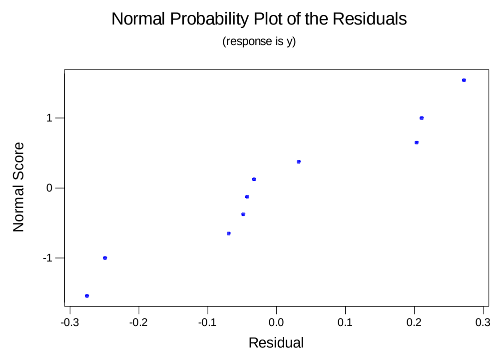
$f_0 >> f_{\alpha,1,7}$

Reject H_0.

d) There is some indication of non-constant variance.



e.) Normality assumption is reasonable.



12-69. a) $\hat{y} = -3950 + 1.0990x_1 + 0.1831x_3 + 3.741x_4 + 0.8375x_5 - 16.346x_6$

$MS_E = 695.90$ $C_p = 5.67$

b) Same as model in part (a)

c) x_1, x_3, x_5, x_6 with $C_p = 5.6$ and $MS_E = 712.89$

d) The model in part (c) is simpler with values for MS_E and C_p similar to those in part (a) and (b). The part (c) model is preferable.

12-70. a) $\hat{y} = -1.060 + 5.509x_1^* + 1.153x_2^* - 3.989x_3^* - 1.102x_4^*$

$MS_E(p) = 0.0750$ Min $C_p = 5.0$

b) Same as part (a)

c) Same as part (a)

d) All models are the same.

12-71. a) $VIF(\hat{\beta}_3^*) = 52.4$

$VIF(\hat{\beta}_4) = 9.3$

$VIF(\hat{\beta}_5) = 29.1$

Yes, VIFs for x_3^* and x_5 exceed 10.

b) Model from Exercise 11-65: $\hat{y} = 19.69 - 1.27x_3^* + 0.005x_4 + 0.0004x_5$

12-72. a) $\hat{y} = 300.0 + 0.85x_1 + 10.4x_2$

$\hat{y} = 300 + 0.85(100) + 10.4(2) = 405.8$

b) $S_{yy} = 1230.5$ $SS_E = 120.3$

$SS_R = S_{yy} - SS_E = 1230.5 - 120.3 = 1110.2$

$$MS_R = \frac{SS_R}{k} = \frac{1110.2}{2} = 555.1$$

$$MS_E = \frac{SS_E}{n-p} = \frac{120.3}{15-3} = 10.025$$

$$f_0 = \frac{MS_R}{MS_E} = \frac{555.1}{10.025} = 55.37$$

$H_0: \beta_1 = \beta_2 = 0$
H_1 : at least one $\beta_j \neq 0$
$\alpha = 0.05$
$f_0 = 55.37$
$f_{.05,2,12} = 3.89$
$f_0 > f_{0.05,2,12}$
Reject H_0 and conclude that the regression model is significant at $\alpha = 0.05$

c) $R^2 = \dfrac{SS_R}{S_{yy}} = \dfrac{1110.2}{1230.5} = 0.9022$ or 90.22%

d) k = 3 p = 4 $SS_E' = 117.20$
$MS_E' = \dfrac{SS_E'}{n-p} = \dfrac{117.2}{11} = 10.65$
No, MS_E increased with the addition of x_3 because the reduction in SS_E was not enough to compensate for the loss in one degree of freedom in the error sum of squares. This is why MS_E can be used as a model selection criterion.

e) $SS_R = S_{yy} - SS_E = 1230.5 - 117.20 = 1113.30$
$$SS_R(\beta_3 \mid \beta_2, \beta_1, \beta_0) = SS_R(\beta_3 \beta_2 \beta_1 \mid \beta_0) - SS_R(\beta_2, \beta_1 \mid \beta_0)$$
$$= 1113.30 - 1110.20$$
$$= 3.1$$
$H_0: \beta_3 = 0$
$H_1: \beta_3 \neq 0$
$\alpha = 0.05$
$$f_0 = \frac{SS_R(\beta_3 \mid \beta_2, \beta_1, \beta_0)/r}{SS_E'/n-p} = \frac{3.1/1}{117.2/11} = 0.291$$
$f_{.05,1,11} = 4.84$
$f_0 \not> f_{0.05,1,11}$
Do not reject H_0.

12-73. a) $R^2 = \dfrac{SS_R}{S_{yy}}$
$SS_R = R^2(S_{yy}) = 0.94(0.50) = 0.47$
$SS_E = S_{yy} - SS_R = 0.5 - 0.47 = 0.03$
$H_0: \beta_1 = \beta_2 = ... = \beta_6 = 0$
H_1 : at least one $\beta_j \neq 0$
$\alpha = 0.05$
$$f_0 = \frac{SS_R/k}{SS_E/n-p} = \frac{0.47/6}{0.03/7} = 18.28$$
$f_{.05,6,7} = 3.87$
$f_0 > f_{0.05,6,7}$
Reject H_0.

b) k = 5   n = 14   p = 6 R^2 = 0.92

$$SS_R' = R^2(S_{yy}) = 0.92(0.50) = 0.46$$
$$SS_E' = S_{yy} - SS_R' = 0.5 - 0.46 = 0.04$$
$$SS_R(\beta\beta_{j_0}, \beta_{i,i=1,2,\ldots,6} \mid \beta_0) = SS_R(full) - SS_R(reduced)$$
$$= 0.47 - 0.46$$
$$= 0.01$$
$$f_0 = \frac{SS_R(\beta_j\ , \beta_{i,i=1,2,\ldots,6} \mid \beta_0)/r}{SS_E'/(n-p)} = \frac{0.01/1}{0.04/8} = 2$$
$$f_{.05,1,8} = 5.32$$
$$f_0 \not> f_{0.05,1,8}$$

Do not reject H_0. There is not sufficient evidence that the removed variable is significant at α = 0.05.

c) $MS_E(reduced) = \dfrac{SS_E}{n-p} = \dfrac{0.04}{8} = 0.005$

$$MS_E(full) = \frac{0.03}{7} = 0.004$$

No, the MS_E is larger for the reduced model, although not by much. Generally, if adding a variable to a model reduces the MS_E it is an indication that the variable may be useful in explaining the response variable. Here the decrease in MS_E is not very great because the added variable had no real explanatory power.

Mind-Expanding Exercises

12-74. Because $R^2 = \dfrac{SS_R}{S_{yy}}$ and $1 - R^2 = \dfrac{SS_E}{S_{yy}}$, $F_0 = \dfrac{SS_R\ /k}{SS_E\ /(n-k-1)}$ and this is the usual F-test for significance of regression. Then, $F_0 = \dfrac{0.90/4}{(1-0.9)/(20-4-1)} = 33.75$ and the critical value is $f_{.05,4,15} = 3.06$. Because 33.75 > 3.06, regression is significant.

12-75. Using n = 20, k = 4, $f_{.05,4,15} = 3.06$. Reject H_0 if

$$\frac{R^2\ /4}{(1-R^2)/15} \ge 3.06$$
$$\frac{R^2}{(1-R^2)} \ge 0.816$$

Then, $R^2 \ge 0.449$ results in a significant regression.

12-76. Because $\hat{\beta} = (X'X)^{-1}X'Y$, $e = Y - X\hat{\beta} = Y - X(X'X)^{-1}X'Y = (I-H)Y$

12-77. From Exercise 12-76, e_i is ith element of $(I-H)Y$. That is,

$$e_i = -h_{i,1}Y_1 - h_{i,2}Y_2 - \ldots - h_{i,i-1}Y_{i-1} + (1-h_{i,i})Y_i - h_{i,i+1}Y_{i+1} - \ldots - h_{i,n}Y_n$$

and

$$V(e_i) = (h_{i,1}^2 + h_{i,2}^2 + \ldots + h_{i,i-1}^2 + (1-h_{i,i}^2) + h_{i,i+1}^2 + \ldots + h_{i,n}^2)\sigma^2$$

The expression in parentheses is recognized to be the *i*th diagonal element of $(I-H)(I-H') = I-H$ by matrix multiplication. Consequently, $V(e_i) = (1-h_{i,i})\sigma^2$. Assume that $i < j$. Now,

$$e_i = -h_{i,1}Y_1 - h_{i,2}Y_2 - \ldots - h_{i,i-1}Y_{i-1} + (1-h_{i,i})Y_i - h_{i,i+1}Y_{i+1} - \ldots - h_{i,n}Y_n$$

$$e_j = -h_{j,1}Y_1 - h_{j,2}Y_2 - \ldots - h_{j,j-1}Y_{j-1} + (1-h_{j,j})Y_j - h_{j,j+1}Y_{j+1} - \ldots - h_{j,n}Y_n$$

Because the y_i's are independent,

$$Cov(e_i, e_j) = (h_{i,1}h_{j,1} + h_{i,2}h_{j,2} + \ldots + h_{i,i-1}h_{j,i-1} + (1-h_{i,i})h_{j,i}$$
$$+ h_{i,i+1}h_{j,i+1} + \ldots + h_{i,j}(1-h_{j,j}) + h_{i,j+1}h_{j,j+1} + \ldots + h_{i,n}h_{j,n})\sigma^2$$

The expression in parentheses is recognized as the *ij*th element of $(I-H)(I-H') = I-H$. Therefore, $Cov(e_i, e_j) = -h_{ij}\sigma^2$.

12-78. $$\hat{\beta} = (X'X)^{-1}X'Y = (X'X)^{-1}X'(X\beta + \varepsilon) = \beta + (X'X)^{-1}X'\varepsilon = \beta + R\varepsilon$$

12-79. a) Min L = $(y - X\beta)'(y - X\beta)$ subject to :$\mathbf{T\beta = c}$

This is equivalent to Min Z = $(\mathbf{y} - \mathbf{X\beta})'(\mathbf{y} - \mathbf{X\beta}) + 2\boldsymbol{\lambda}'(\mathbf{T\beta} - \mathbf{c})$

where $\boldsymbol{\lambda}' = [\lambda_1, \lambda_2, \ldots, \lambda_p]$ is a vector of La Grange multipliers.

$$\frac{\partial \mathbf{Z}}{\partial \boldsymbol{\beta}} = -2\mathbf{X'y} + 2(\mathbf{X'X})\boldsymbol{\beta} + 2\mathbf{T}'\boldsymbol{\lambda}$$

$$\frac{\partial Z}{\partial \lambda} = 2(T\beta - c). \text{ Set } \frac{\partial \mathbf{Z}}{\partial \boldsymbol{\beta}} = 0, \quad \frac{\partial \mathbf{Z}}{\partial \boldsymbol{\lambda}} = 0.$$

Then we get

$$(X'X)\hat{\beta}_c + T'\lambda = X'y$$
$$T\hat{\beta}_c = c$$

where $\hat{\boldsymbol{\beta}}_{\mathbf{c}}$ is the constrained estimator. From the first of these equations,

$$\hat{\boldsymbol{\beta}}_{\mathbf{c}} = (\mathbf{X'X})^{-1}(\mathbf{X'y} - \mathbf{T}'\boldsymbol{\lambda}) = \hat{\boldsymbol{\beta}} - (\mathbf{X'X})^{-1}\mathbf{T}'\boldsymbol{\lambda}$$

From the second, $\mathbf{T}\hat{\boldsymbol{\beta}} - \mathbf{T}(\mathbf{X'X})^{-1}\mathbf{T}'\boldsymbol{\lambda} = \mathbf{c}$ or $\boldsymbol{\lambda} = [\mathrm{T}(\mathrm{X'X})^{-1}\mathrm{T}']^{-1}(\mathbf{T}\hat{\boldsymbol{\beta}} - \mathbf{c})$

Then

$$\hat{\boldsymbol{\beta}}_{\mathbf{c}} = \hat{\boldsymbol{\beta}} - (\mathbf{X'X})^{-1}\mathbf{T}'[\mathbf{T}(\mathbf{X'X})^{-1}\mathbf{T}']^{-1}(\mathbf{T}\hat{\boldsymbol{\beta}} - \mathbf{c})$$
$$= \hat{\boldsymbol{\beta}} + (\mathbf{X'X})^{-1}\mathbf{T}'[\mathbf{T}(\mathbf{X'X})^{-1}\mathbf{T}']^{-1}(\mathbf{c} - \mathbf{T}\hat{\boldsymbol{\beta}})$$

b) This solution would be appropriate in situations where you have tested the hypothesis that $\mathbf{T\beta = c}$ and concluded that this hypothesis cannot be rejected.

12-80. For the piecewise linear function to be continuous at x = x^*, the point-slope formula for a line can be used to show that

$$y = \begin{cases} \beta_0 + \beta_1(x - x^*) & x \le x^* \\ \beta_0 + \beta_2(x - x^*) & x > x^* \end{cases}$$

where $\beta_0, \beta_1, \beta_2$ are arbitrary constants.

Let $z = \begin{cases} 0, & x \le x^* \\ 1, & x > x^* \end{cases}$.

Then, y can be written as $y = \beta_0 + \beta_1(x - x^*) + (\beta_2 - \beta_1)(x - x^*)z$.

Let

$$x_1 = x - x^*$$
$$x_2 = (x - x^*)z$$
$$\beta_0{}^* = \beta_0$$
$$\beta_1{}^* = \beta_1$$
$$\beta_2{}^* = \beta_2 - \beta_1$$

Then, $y = \beta_0 * + \beta_1 * x_1 + \beta_2 * x_2$.

12-81. Should refer to exercise 12-80. If there is a discontinuity at $x = x^*$, then a model that can be used is

$$y = \begin{cases} \beta_0 + \beta_1 x & x \le x^* \\ \alpha_0 + \alpha_1 x & x > x^* \end{cases}$$

Let $z = \begin{cases} 0, & x \le x^* \\ 1, & x > x^* \end{cases}$

Then, y can be written as $y = \beta_0 + \beta_1 x + [(\alpha_0 - \beta_0) + (\alpha_1 - \beta_1)x]z = \beta_0{}^* + \beta_1{}^* x_1 + \beta_2{}^* z + \beta_3{}^* x_2$

where

$$\beta_0{}^* = \beta_0$$
$$\beta_1{}^* = \beta_1$$
$$\beta_2{}^* = \alpha_0 - \beta_0$$
$$\beta_3{}^* = \alpha_1 - \beta_1$$
$$x_1 = x$$
$$x_2 = xz$$

12-82. Should refer to exercise 12-80. One could estimate x* as a parameter in the model. A simple approach is to refit the model in Exercise 12-80 with different choices for x* and to select the value for x* that minimizes the residual sum of squares.