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) $\hat{\beta} = (X'X)^{-1}X'y$

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

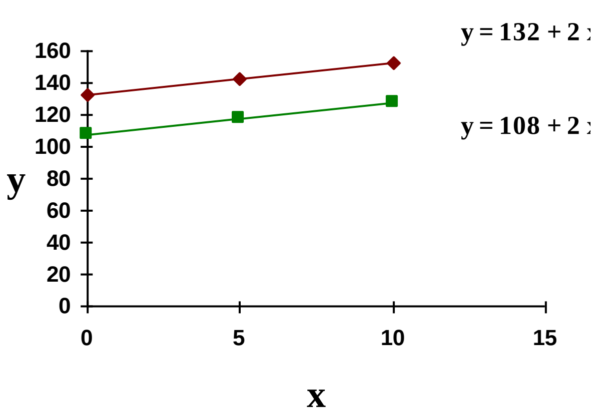
b) $\hat{y} = -1.9122 + 0.0931x_1 + 0.2532x_2$

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

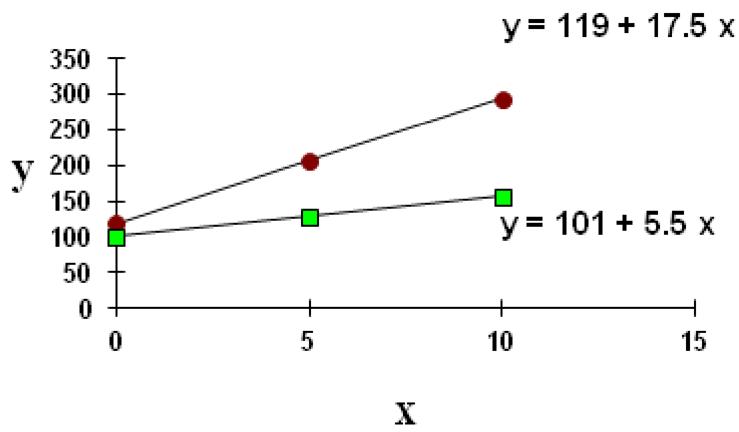
12-3 a)

	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) $x_2 = 6$ $\quad \hat{y} = 100 + 2x_1 + 4(6)$

$$\hat{y} = 124 + 2x_1$$

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)

	$x_2 = 6$	$x_2 = 2$	$x_2 = 8$
	$\hat{y} = 95 + 1.5x_1 + 3(6) + 2x_1(6)$ $\hat{y} = 113 + 13.5x_1$	$\hat{y} = 101 + 5.5x_1$	$\hat{y} = 119 + 17.5x_1$
Change per unit of x_1	13.5	5.5	17.5

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

12-4 a) There are two regressor variables in this model based on the size of the $(X'X)^{-1}$ matrix.

b) The estimate of σ^2 is the $MS_{Residual}$. The $MS_{Residual} = \dfrac{SS_{Residual}}{DF} = \dfrac{307}{14-2} = 25.583$

c) Standard error of $\hat{\beta}_1 = \sqrt{\hat{\sigma}^2 C_{11}} = \sqrt{(25.583)(0.0013329)} = 0.1847$

12-5 a) The results from computer software follow. The model can be expressed as

```
Satisfaction = 144 - 1.11 Age - 0.585 Severity + 1.30 Anxiety
```

b) $\hat{\sigma}^2 = \dfrac{\sum_{t=1}^{n} e_t^2}{n-p} = \dfrac{SS_E}{n-p} = \dfrac{1039.9}{21} = 49.5$

c) $\text{cov}(\hat{\beta}) = \sigma^2 (X'X)^{-1} = \sigma^2 C$, $se(\hat{\beta}) = \sqrt{\hat{\sigma}^2 C_{jj}} = \begin{bmatrix} 5.9 \\ 0.13 \\ 0.13 \\ 1.06 \end{bmatrix}$ from the Minitab output.

d) Because the regression coefficients have different standard errors the parameters estimators do not have similar precision of estimation.

Regression Analysis: Satisfaction versus Age, Severity, Anxiety

```
The regression equation is
Satisfaction = 144 - 1.11 Age - 0.585 Severity + 1.30 Anxiety

Predictor     Coef  SE Coef      T      P
Constant   143.895    5.898  24.40  0.000
Age        -1.1135   0.1326  -8.40  0.000
Severity   -0.5849   0.1320  -4.43  0.000
Anxiety      1.296    1.056   1.23  0.233

S = 7.03710   R-Sq = 90.4%   R-Sq(adj) = 89.0%

Analysis of Variance

Source          DF       SS      MS      F      P
Regression       3   9738.3  3246.1  65.55  0.000
Residual Error  21   1039.9    49.5
Total           24  10778.2

Source    DF  Seq SS
Age        1  8756.7
Severity   1   907.0
Anxiety    1    74.6

Unusual Observations

Obs   Age  Satisfaction    Fit  SE Fit  Residual  St Resid
  9  27.0         75.00  93.28    2.98    -18.28     -2.87R
```

12-6 **Regression Analysis: y versus x1, x2, x3, x4**

```
The regression equation is
y = - 5 + 1.79 x1 + 4.93 x2 + 1.78 x3 - 0.246 x4

Predictor        Coef     SE Coef          T        P
Constant         -4.8       220.8      -0.02    0.983
X1             1.7950      0.6774       2.65    0.033
X2              4.927       9.608       0.51    0.624
X3              1.781       2.425       0.73    0.486
X4            -0.2465      0.9165      -0.27    0.796

S = 16.4907       R-Sq = 71.0%     R-Sq(adj) = 54.5%

Analysis of Variance

Source            DF          SS          MS          F        P
Regression         4      4669.3      1167.3       4.29    0.046
Residual Error     7      1903.6       271.9
Total             11      6572.9
```

a) $\hat{y} = -5 + 1.79x_1 + 4.93x_2 + 1.78x_3 - 0.246x_4$

b) $\hat{\sigma}^2 = 271.9$

c) $se(\hat{\beta}_0) = 220.8$, $se(\hat{\beta}_1) = 0.6774$, $se(\hat{\beta}_2) = 9.608$, $se(\hat{\beta}_3) = 2.425$, and $se(\hat{\beta}_4) = 0.9165$

Because the regression coefficients have different standard errors the parameters estimators do not have similar precision of estimation.

d) $\hat{y} = -5 + 1.79(24) + 4.93(24) + 1.78(90) - 0.246(98) = 292.372$

12-7

```
The regression equation is
mpg = 49.9 - 0.0104 cid - 0.0012 rhp - 0.00324 etw + 0.29 cmp - 3.86 axle
      + 0.190 n/v

Predictor        Coef    SE Coef      T      P
Constant        49.90      19.67   2.54  0.024
cid          -0.01045    0.02338  -0.45  0.662
rhp          -0.00120    0.01631  -0.07  0.942
etw        -0.0032364  0.0009459  -3.42  0.004
cmp             0.292      1.765   0.17  0.871
axle           -3.855      1.329  -2.90  0.012
n/v            0.1897     0.2730   0.69  0.498

S = 2.22830   R-Sq = 89.3%   R-Sq(adj) = 84.8%

Analysis of Variance

Source          DF       SS      MS      F      P
Regression       6  581.898  96.983  19.53  0.000
Residual Error  14   69.514   4.965
Total           20  651.412
```

a) $\hat{y} = 49.90 - 0.01045x_1 - 0.0012x_2 - 0.00324x_3 + 0.292x_4 - 3.855x_5 + 0.1897x_6$

where $x_1 = cid \quad x_2 = rhp \quad x_3 = etw \quad x_4 = cmp \quad x_5 = axle \quad x_6 = n/v$

b) $\hat{\sigma}^2 = 4.965$

$se(\hat{\beta}_0) = 19.67$, $se(\hat{\beta}_1) = 0.02338$, $se(\hat{\beta}_2) = 0.01631$, $se(\hat{\beta}_3) = 0.0009459$,

$se(\hat{\beta}_4) = 1.765$, $se(\hat{\beta}_5) = 1.329$ and $se(\hat{\beta}_6) = 0.273$

c)

$\hat{y} = 49.90 - 0.01045(215) - 0.0012(253) - 0.0032(4500) + 0.292(9.9) - 3.855(3.07) + 0.1897(30.9)$

$= 29.867$

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) $\hat{y} = 7.4578 - 0.0297(22) + 0.5205(31) - 0.1018(92) - 2.1606(2.1) \quad \hat{y} = 9.037$

12-9 **Regression Analysis: Ex12-9y versus Ex12-9x1, Ex12-9x2, Ex12-9x3**

```
The regression equation is
Ex12-9y = 47.8 - 9.60 Ex12-9x1 + 0.415 Ex12-9x2 + 18.3 Ex12-9x3


Predictor    Coef  SE Coef      T      P
Constant    47.82    49.94   0.96  0.353
Ex12-9x1   -9.604    3.723  -2.58  0.020
Ex12-9x2   0.4152   0.2261   1.84  0.085
Ex12-9x3   18.294    1.323  13.82  0.000


S = 3.50508   R-Sq = 99.4%   R-Sq(adj) = 99.2%


Analysis of Variance

Source          DF     SS     MS       F      P
Regression       3  30529  10176  828.31  0.000
Residual Error  16    197     12
Total           19  30725
```

a) y hat = 47.8 - 9.60x1 + 0.415x2 + 18.3x3

b) $\hat{\sigma}^2 = 12$

c) The estimated standard errors of the coefficient estimators are provided in the above table (SE Coef). Because the regression coefficients have different standard errors the parameters estimators do not have similar precision of estimation.

d) y hat = 47.8 - 9.60(15.0) + 0.415(230) + 18.3(7) = 127.35

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)
$$\hat{y} = -0.03023 + 0.000029(1650) + 0.002318(1) - 0.003029(1.1)$$
$$+ 0.008476(1) - 0.002363(0.80)$$
$$\hat{y} = 0.0319$$

12-11

```
The regression equation is
rads = - 440 + 19.1 mAmps + 68.1 exposure time

Predictor          Coef  SE Coef      T      P
Constant        -440.39    94.20  -4.68  0.000
mAmps            19.147    3.460   5.53  0.000
exposure time    68.080    5.241  12.99  0.000

S = 235.718   R-Sq = 84.3%   R-Sq(adj) = 83.5%

Analysis of Variance

Source          DF        SS       MS      F      P
Regression       2  11076473  5538237  99.67  0.000
Residual Error  37   2055837    55563
Total           39  13132310
```

a) $\hat{y} = -440.39 + 19.147x_1 + 68.080x_2$

where $x_1 = mAmps \quad x_2 = ExposureTime$

b) $\hat{\sigma}^2 = 55563$

$se(\hat{\beta}_0) = 94.20$, $se(\hat{\beta}_1) = 3.460$, and $se(\hat{\beta}_2) = 5.241$

c) $\hat{y} = -440.93 + 19.147(20) + 68.080(15) = 963.21$

12-12

```
The regression equation is
ARSNAILS = 0.488 - 0.00077 AGE - 0.0227 DRINKUSE - 0.0415 COOKUSE
           + 13.2 ARSWATER

Predictor        Coef   SE Coef      T      P
Constant       0.4875    0.4272   1.14  0.271
AGE         -0.000767  0.003508  -0.22  0.830
DRINKUSE     -0.02274   0.04747  -0.48  0.638
COOKUSE      -0.04150   0.08408  -0.49  0.628
ARSWATER       13.240     1.679   7.89  0.000

S = 0.236010   R-Sq = 81.2%   R-Sq(adj) = 76.5%

Analysis of Variance

Source          DF       SS       MS      F      P
Regression       4  3.84906  0.96227  17.28  0.000
Residual Error  16  0.89121  0.05570
Total           20  4.74028
```

a) $\hat{y} = 0.4875 - 0.000767x_1 - 0.02274x_2 - 0.04150x_3 + 13.240x_4$

where $x_1 = AGE \quad x_2 = DrinkUse \quad x_3 = CookUse \quad x_4 = ARSWater$

b) $\hat{\sigma}^2 = 0.05570$

$se(\hat{\beta}_0) = 0.4272$, $se(\hat{\beta}_1) = 0.003508$, $se(\hat{\beta}_2) = 0.04747$, $se(\hat{\beta}_3) = 0.08408$, and

$se(\hat{\beta}_4) = 1.679$

c) $\hat{y} = 0.4875 - 0.000767(55) - 0.02274(5) - 0.04150(5) + 13.240(0.625) = 8.399$

12-13
```
The regression equation is
density = - 0.110 + 0.407 dielectric constant + 2.11 loss factor

Predictor                Coef  SE Coef      T      P
Constant              -0.1105   0.2501  -0.44  0.670
dielectric constant    0.4072   0.1682   2.42  0.042
loss factor             2.108    5.834   0.36  0.727

S = 0.00883422   R-Sq = 99.7%   R-Sq(adj) = 99.7%

Analysis of Variance

Source          DF       SS       MS        F      P
Regression       2  0.23563  0.11782  1509.64  0.000
Residual Error   8  0.00062  0.00008
Total           10  0.23626
```

a) $\hat{y} = -0.1105 + 0.4072x_1 + 2.108x_2$

where $x_1 = Dielectric\,Const \quad x_2 = LossFactor$

b) $\hat{\sigma}^2 = 0.00008$

$se(\hat{\beta}_0) = 0.2501$, $se(\hat{\beta}_1) = 0.1682$, and $se(\hat{\beta}_2) = 5.834$

c) $\hat{y} = -0.1105 + 0.4072(2.3) + 2.108(0.032) = 0.8935$

12-14
```
The regression equation is
y = - 171 + 7.03 x1 + 12.7 x2

Predictor     Coef  SE Coef      T      P
Constant   -171.26    28.40  -6.03  0.001
x1           7.029    1.539   4.57  0.004
x2          12.696    1.539   8.25  0.000

S = 3.07827   R-Sq = 93.7%   R-Sq(adj) = 91.6%

Analysis of Variance

Source          DF      SS      MS      F      P
Regression       2  842.37  421.18  44.45  0.000
Residual Error   6   56.85    9.48
Total            8  899.22
```

a) $\hat{y} = -171 + 7.03x_1 + 12.7x_2$

b) $\hat{\sigma}^2 = 9.48$

$se(\hat{\beta}_0) = 28.40$, $se(\hat{\beta}_1) = 1.539$, and $se(\hat{\beta}_2) = 1.539$

c) $\hat{y} = -171 + 7.03(14.5) + 12.7(12.5)$

$= 99.55$

12-15
```
The regression equation is
Useful range (ng) = 239 + 0.334 Brightness (%) - 2.72 Contrast (%)

Predictor          Coef  SE Coef      T      P
Constant         238.56    45.23   5.27  0.002
Brightness (%)   0.3339   0.6763   0.49  0.639
```

```
Contrast (%)    -2.7167   0.6887  -3.94  0.008


S = 36.3493   R-Sq = 75.6%   R-Sq(adj) = 67.4%


Analysis of Variance

Source          DF     SS     MS     F      P
Regression       2  24518  12259  9.28  0.015
Residual Error   6   7928   1321
Total            8  32446
```

a) $\hat{y} = 238.56 + 0.3339x_1 - 2.7167x_2$
where $x_1 = \%Brightness \quad x_2 = \%Contrast$

b) $\hat{\sigma}^2 = 1321$

c) $se(\hat{\beta}_0) = 45.23$, $se(\hat{\beta}_1) = 0.6763$, and $se(\hat{\beta}_2) = 0.6887$

d) $\hat{y} = 238.56 + 0.3339(90) - 2.7167(80) = 51.275$

12-16
```
The regression equation is
Stack Loss(y) = - 39.9 + 0.716 X1 + 1.30 X2 - 0.152 X3

Predictor     Coef  SE Coef      T      P
Constant    -39.92    11.90  -3.36  0.004
X1          0.7156   0.1349   5.31  0.000
X2          1.2953   0.3680   3.52  0.003
X3         -0.1521   0.1563  -0.97  0.344

S = 3.24336   R-Sq = 91.4%   R-Sq(adj) = 89.8%

Analysis of Variance

Source          DF       SS      MS      F      P
Regression       3  1890.41  630.14  59.90  0.000
Residual Error  17   178.83   10.52
Total           20  2069.24
```

a) $\hat{y} = -39.92 + 0.7156x_1 + 1.2953x_2 - 0.1521x_3$

b) $\hat{\sigma}^2 = 10.52$

c) $se(\hat{\beta}_0) = 11.90$, $se(\hat{\beta}_1) = 0.1349$, $se(\hat{\beta}_2) = 0.3680$, and $se(\hat{\beta}_3) = 0.1563$

d) $\hat{y} = -39.92 + 0.7156(65) + 1.2953(28) - 0.1521(90) = 29.173$

12-17 a) The model can be expressed as:
```
Rating Pts = 2.99 + 1.20 Pct Comp + 4.60 Pct TD - 3.81 Pct Int
```

Regression Analysis: Rating Pts versus Pct Comp, Pct TD, Pct Int

```
The regression equation is
Rating Pts = 2.99 + 1.20 Pct Comp + 4.60 Pct TD - 3.81 Pct Int
```

```
Predictor      Coef  SE Coef      T      P
Constant      2.986    5.877   0.51  0.615
Pct Comp    1.19857  0.09743  12.30  0.000
Pct TD       4.5956   0.3848  11.94  0.000
Pct Int     -3.8125   0.4861  -7.84  0.000

S = 2.03479   R-Sq = 95.3%   R-Sq(adj) = 94.8%

Analysis of Variance

Source          DF       SS      MS       F      P
Regression       3  2373.59  791.20  191.09  0.000
Residual Error  28   115.93    4.14
Total           31  2489.52

Source    DF   Seq SS
Pct Comp   1  1614.43
Pct TD     1   504.49
Pct Int    1   254.67
```

b) $\hat{\sigma}^2 = \dfrac{\sum_{t=1}^{n} e_t^2}{n-p} = \dfrac{SS_E}{n-p} = \dfrac{115.93}{28} = 4.14$

c) $\text{cov}(\hat{\beta}) = \sigma^2 (X'X)^{-1} = \sigma^2 C$, $se(\hat{\beta}) = \sqrt{\hat{\sigma}^2 C_{jj}} = \begin{bmatrix} 5.88 \\ 0.097 \\ 0.38 \\ 0.48 \end{bmatrix}$

from the SE Coef column in the computer output.

d) Rating Pts = 2.99 + 1.20*65 + 4.60*5 - 3.81*4 = 88.75

12-18 **Regression Analysis: W versus GF, GA, ...**

```
The regression equation is
W = 512 + 0.164 GF - 0.183 GA - 0.054 ADV + 0.09 PPGF - 0.14 PCTG - 0.163 PEN
    - 0.128 BMI + 13.1 AVG + 0.292 SHT - 1.60 PPGA - 5.54 PKPCT + 0.106 SHGF
    + 0.612 SHGA + 0.005 FG

Predictor       Coef  SE Coef      T      P
Constant       512.2    185.9   2.75  0.015
GF           0.16374  0.03673   4.46  0.000
GA          -0.18329  0.04787  -3.83  0.002
ADV          -0.0540   0.2183  -0.25  0.808
PPGF           0.089    1.126   0.08  0.938
PCTG          -0.142    3.810  -0.04  0.971
PEN          -0.1632   0.3029  -0.54  0.598
BMI          -0.1282   0.2838  -0.45  0.658
AVG            13.09    24.84   0.53  0.606
SHT           0.2924   0.1334   2.19  0.045
PPGA         -1.6018   0.6407  -2.50  0.025
PKPCT         -5.542    2.181  -2.54  0.023
SHGF          0.1057   0.1975   0.54  0.600
SHGA          0.6124   0.2615   2.34  0.033
FG            0.0047   0.1943   0.02  0.981

S = 2.65443   R-Sq = 92.9%   R-Sq(adj) = 86.3%

Analysis of Variance
```

Source	DF	SS	MS	F	P
Regression	14	1390.310	99.308	14.09	0.000
Residual Error	15	105.690	7.046		
Total	29	1496.000			

$$\hat{y} = 512.2 + 0.16374x_1 - 0.18329x_2 - 0.054x_3 + 0.089x_4 - 0.142x_5 - 0.1632x_6 - 0.1282x_7$$

$$+13.09x_8 + 0.2924x_9 - 1.6018x_{10} - 5.542x_{11} + 0.1057x_{12} + 0.6124x_{13} + 0.0047x_{14}$$

where

$$x_1 = GF \quad x_2 = GA \quad x_3 = ADV \quad x_4 = PPGF \quad x_5 = PCTG \quad x_6 = PEN \quad x_7 = BMI$$

$$x_8 = AVG \quad x_9 = SHT \quad x_{10} = PPGA \quad x_{11} = PKPCT \quad x_{12} = SHGF \quad x_{13} = SHGA \quad x_{14} = FG$$

$$\hat{\sigma}^2 = 7.046$$

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

12-19

Predictor	Coef	SE Coef	T	P
Constant	360.81	37.08	9.73	0.002
X1	-3.7518	0.5570	-6.74	0.007
X2	-0.08427	0.04406	-1.91	0.152

S = 12.6074 R-Sq = 98.4% R-Sq(adj) = 97.3%

Analysis of Variance

Source	DF	SS	MS	F	P
Regression	2	28504	14252	89.67	0.002
Residual Error	3	477	159		
Total	5	28981			

a) $\hat{y} = 361 - 3.75x_1 - 0.0843x_2$

b) $\hat{\sigma}^2 = 159$, $se(\hat{\beta}_0) = 37.08$, $se(\hat{\beta}_1) = 0.5570$, and $se(\hat{\beta}_2) = 0.04406$

c) $\hat{y} = 361 - 3.75(25) - 0.0843(1000) = 182.95$

d)

Predictor	Coef	SE Coef	T	P
Constant	476.77	82.68	5.77	0.029
X1	-8.525	3.189	-2.67	0.116
X2	-0.21097	0.09150	-2.31	0.148
X1*X2	0.004787	0.003164	1.51	0.269

S = 10.5443 R-Sq = 99.2% R-Sq(adj) = 98.1%

Analysis of Variance

Source	DF	SS	MS	F	P
Regression	3	28759.0	9586.3	86.22	0.011
Residual Error	2	222.4	111.2		
Total	5	28981.3			

$\hat{y}' = 477 - 8.53x_1 - 0.211x_2 + 0.0048x_1x_2$

e) $\hat{\sigma}^2 = 111.2$, $se(\hat{\beta}_0) = 82.68$, $se(\hat{\beta}_1) = 3.189$, $se(\hat{\beta}_2) = 0.092$ and $se(\hat{\beta}_{12}) = 0.0032$

f) $\hat{y} = 477 - 8.53(25) - 0.211(1000) + 0.0048(25)(1000) = 172.75$

The predicted value is smaller

12-20 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 $\hat{\beta}_0' = 0$.

<u>Sections 12-2</u>

12-21 a) $t_0 = \dfrac{\hat{\beta}_j - \beta_{j0}}{se(\hat{\beta}_j)}$, null hypothesis $\hat{\beta}_j = \beta_{j0}$ is rejected at α level if $|t_0| > t_{\alpha/2, n-p}$

$$F_0 = \frac{SS_R / k}{SS_E / (n-p)} = \frac{MS_R}{MS_E}$$, regression is significant at α level if $f_0 > f_{\alpha, k, n-p}$

The missing quantities are as follows:

Predictor	Coef	SE Coef	T	P	
Constant	253.81	4.781	53.0872	0	
x1	2.7738	0.1846	15.02	0	
x2	-4.6394	0.1526	-30.4024	0	
Source	DF	SS	MS	F	P
Regression	2	22784	11392	445.2899	0
Residual Error	12	307	25.5833		
Total	14	23091			

R-Squared = 22784/23091 = 0.9867

b) From the P-value from the F test (F = 445.2899) for regression is significant.

c) Each individual regressor is significant to the model that contains the other regressors.

12-22 a) $R^2 = \dfrac{SS_R}{SS_T} = \dfrac{1000}{1200} = 0.83$

b) $SS_E = SS_T - SS_R = 1200 - 1000 = 200$

$$R^2_{adj} = 1 - \frac{SS_E/(n-p)}{SS_T/(n-1)} = 1 - \frac{200/(20-3)}{1200/(20-1)} = 0.8137$$

c) $MS_{\text{Regression}} = \dfrac{SS_{\text{Regression}}}{k} = \dfrac{1000}{2} = 500$

$$MS_{\text{Error}} = \frac{SS_E}{n-p} = \frac{1200-1000}{17} = \frac{200}{17} = 11.765$$

$$F = \frac{MS_{\text{Regression}}}{MS_{\text{Error}}} = \frac{500}{11.765} = 42.5$$

The ANOVA table

Source	DF	SS	MS	F	P
Regression	2	1000	500	42.5	< 0.01
Residual Error	17	200	11.765		
Total	19	1200			

For the F test the P-value < 0.0. Therefore the F test rejects the null hypothesis at $\alpha = 0.05$ and also rejects at $\alpha = 0.01$.

d) The ANOVA table after adding a third regressor

Source	DF	SS	MS
Regression	3	785	261.6667
Residual Error	16	415	25.94
Total	19	1200	

$$f = \frac{SS_{\text{Regression}}(\beta_3 \mid \beta_2, \beta_1, \beta_0)/1}{MS_{\text{Error}}} = \frac{1100-1000}{25.94} = 3.855$$

Because $f_{0.05,1,16} = 4.49$, we fail to reject H_0 and conclude that the third regressor does not contribute significantly to the model.

12-23 a) $n = 10$, $k = 2$, $p = 3$, $\alpha = 0.05$

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

$H_1: \beta_j \neq 0$ 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 $\alpha = 0.05$. P-value = 0.000

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

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

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

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

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

Reject H_0, P-value < 0.001   Reject H_0, P-value < 0.001

Both regression coefficients significant

12-24 $S_{yy} = 738.00$

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

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

$\alpha = 0.01$

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

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

$$= 5525.5548 - 4840$$

$$= 685.55$$

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

$$= 738 - 685.55$$

$$= 52.45$$

$$f_0 = \frac{\frac{SS_R}{k}}{\frac{SS_E}{n-p}} = \frac{685.55/2}{52.45/7} = 45.75$$

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

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

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

$$\hat{\sigma}^2 = MS_E = \frac{SS_E}{n-p} = \frac{52.45}{7} = 7.493$$

$$se(\hat{\beta}_1) = \sqrt{7.493(7.9799E-5)} = 0.0245$$

b) $H_0 : \beta_1 = 0$

$H_1 : \beta_1 \neq 0$

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

$$= \frac{0.0931}{0.0245} = 3.8$$

$t_{0.005,7} = 3.499$

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

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 - 0.9966426) = 0.006715

c) x_1 is useful as a regressor in the model.

12-25 a) Degrees of freedom = 20 – 4 = 16

$\beta_1 : t_0 = 4.82$ P-value = 2(9.424 E-5) = 1.88 E-4

$\beta_2 : t_0 = 8.21$ P-value = 2(1.978 E-7) = 3.96 E-7

$\beta_3 : t_0 = 0.98$ P-value = 2 (0.171) = 0.342

b) $H_0 : \beta_3 = 0$

$H_1 : \beta_3 \neq 0$

$\alpha = 0.05$

$t_0 = 0.98$, P-value = 2 (0.171) = 0.342

Because the P-value > α = 0.05, fail to reject H_0. We conclude that x_3 does not contribute significantly to the model.

12-26 a) $H_0 : \beta_1 = \beta_2 = \beta_3 = \beta_4 = 0$

H_1 at least one $\beta_j \neq 0$

$\alpha = 0.1$

$f_0 = 10.08$

$f_{0.1,4,7} = 2.96$

$f_0 > f_{0.1,4,7}$

Reject H_0 P-value = 0.005

b) $\alpha = 0.1$

$H_0 : \beta_1 = 0$	$\beta_2 = 0$	$\beta_3 = 0$	$\beta_4 = 0$
$H_1 : \beta_1 \neq 0$	$\beta_2 \neq 0$	$\beta_3 \neq 0$	$\beta_4 \neq 0$
$t_0 = 2.71$	$t_0 = 1.87$	$t_0 = 1.37$	$t_0 = -0.87$

$t_{\alpha/2,n-p} = t_{0.05,7} = 1.895$

$|t_0| \ngtr t_{0.05,7}$ for β_2, β_3 and β_4

Reject H_0 for β_1.

12-27 a) $H_0 : \beta_1 = \beta_2 = \beta_3 = \beta_4 = \beta_5 = \beta_6 = 0$

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

$f_0 = 19.53$

$f_{\alpha,6,14} = f_{0.1,6,14} = 2.24$

$f_0 > f_{0.1,6,14}$

Reject H_0 and conclude regression model is significant at $\alpha = 0.1$

b) The t-test statistics for β_1 through β_6 are -0.45, -0.07, -3.42, 0.17, -2.90, 0.69. Because $t_{0.05,14} = 1.761$, the regressors that contribute to the model at $\alpha = 0.1$ are *etw* and *axle*.

12-28 a) $H_0: \beta_j = 0$ for all j

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

$f_0 = 7.16$

$f_{0.1,4,14} = 2.39$

$f_0 > f_{0.1,4,14}$

Reject H_0 and conclude that the regression is significant at $\alpha = 0.1$. P-value = 0.0023

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

$\alpha = 0.1$ $t_{\alpha/2,n-p} = t_{.05,14} = 1.761$

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

X_2 and X_5 do not contribute to the model.

12-29 a) $H_0: \beta_1 = \beta_2 = \beta_3 = 0$

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

$f_0 = 828.31$

$f_{0.1,3,16} = 2.46$

$f_0 > f_{0.1,3,16}$

Reject H_0 and conclude regression is significant at $\alpha = 0.1$

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

$\alpha = 0.1$ $t_{\alpha/2,n-p} = t_{.05,16} = 1.746$

$H_0: \beta_1 = 0$	$\beta_2 = 0$	$\beta_3 = 0$
$H_1: \beta_1 \neq 0$	$\beta_2 \neq 0$	$\beta_3 \neq 0$
$t_0 = -2.58$	$t_0 = 1.84$	$t_0 = 13.82$
$\lvert t_0 \rvert > t_{0.05,16}$	$\lvert t_0 \rvert > t_{0.05,16}$	$\lvert t_0 \rvert > t_{0.05,16}$
Reject H_0	Reject H_0	Reject H_0

12-30
```
ARSNAILS = 0.001 + 0.00858 AGE - 0.021 DRINKUSE + 0.010 COOKUSE

  Predictor       Coef   SE Coef      T      P
  Constant      0.0011    0.9067   0.00  0.999
  AGE         0.008581  0.007083   1.21  0.242
  DRINKUSE     -0.0208    0.1018  -0.20  0.841
  COOKUSE       0.0097    0.1798   0.05  0.958
```

```
S = 0.506197   R-Sq = 8.1%   R-Sq(adj) = 0.0%

Analysis of Variance

Source          DF      SS      MS     F      P
Regression       3  0.3843  0.1281  0.50  0.687
Residual Error  17  4.3560  0.2562
Total           20  4.7403
```

a) $H_0: \beta_1 = \beta_2 = \beta_3 = 0$

$H_1: \beta_j \neq 0$ for at least one j; k = 4

$\alpha = 0.01$

$f_0 = 0.50$

$f_{0.01,3,17} = 5.18$

$f_0 < f_{0.01,3,17}$

Do not reject H_0. There is insufficient evidence to conclude that the model is significant at α = 0.01. The P-value = 0.687.

b) $H_0: \beta_1 = 0$

$H_1: \beta_1 \neq 0$

$\alpha = 0.01$

$$t_0 = \frac{\hat{\beta}_1}{se(\hat{\beta}_1)} = \frac{0.008581}{0.007083} = 1.21$$

$t_{0.005,17} = 2.898$

$|t_0| < t_{\alpha/2,17}$. Fail to reject H_0, there is not enough evidence to conclude that β_1 is significant in the model at α = 0.01.

$H_0: \beta_2 = 0$

$H_1: \beta_2 \neq 0$

$\alpha = 0.01$

$$t_0 = \frac{\hat{\beta}_2}{se(\hat{\beta}_2)} = \frac{-0.0208}{0.1018} = -0.2$$

$t_{0.005,17} = 2.898$

$|t_0| < t_{\alpha/2,17}$. Fail to reject H_0, there is not enough evidence to conclude that β_2 is significant in the model at α = 0.01.

$H_0: \beta_3 = 0$

$H_1: \beta_3 \neq 0$

$\alpha = 0.01$

$$t_0 = \frac{\hat{\beta}_3}{se(\hat{\beta}_3)} = \frac{0.0097}{0.1798} = 0.05$$

$t_{0.005,17} = 2.898$

$|t_0| < t_{\alpha/2,17}$. Fail to reject H_0, there is not enough evidence to conclude that β_3 is significant in the model at α = 0.01.

12-31 a) $H_0: \beta_1 = \beta_2 = 0$

$H_0:$ for at least one $\beta_j \neq 0$

α = 0.05

$$f_0 = 99.67$$
$$f_{0.05,2,37} = 3.252$$
$$f_0 > f_{0.05,2,37}$$

```
 The regression equation is
rads = - 440 + 19.1 mAmps + 68.1 exposure time


Predictor          Coef  SE Coef      T      P
Constant        -440.39    94.20  -4.68  0.000
mAmps            19.147    3.460   5.53  0.000
exposure time    68.080    5.241  12.99  0.000


S = 235.718   R-Sq = 84.3%   R-Sq(adj) = 83.5%


Analysis of Variance

Source          DF        SS       MS      F      P
Regression       2  11076473  5538237  99.67  0.000
Residual Error  37   2055837    55563
Total           39  13132310
```

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

b) $\hat{\sigma}^2 = MS_E = 55563$

$$se(\hat{\beta}_1) = \sqrt{\hat{\sigma}^2 c_{jj}} = 3.460$$
$$H_0 : \beta_1 = 0$$
$$H_1 : \beta_1 \neq 0$$
$\alpha = 0.05$
$$t_0 = \frac{\hat{\beta}_1}{se(\hat{\beta}_1)}$$
$$= \frac{19.147}{3.460} = 5.539$$
$$t_{0.025,40-3} = t_{0.025,37} = 2.0262$$
$$|t_0| > t_{\alpha/2,37},$$

Reject H_0 and conclude that β_1 is significant in the model at $\alpha = 0.05$

$$se(\hat{\beta}_2) = \sqrt{\hat{\sigma}^2 c_{jj}} = 5.241$$
$$H_0 : \beta_2 = 0$$
$$H_1 : \beta_2 \neq 0$$
$\alpha = 0.05$
$$t_0 = \frac{\hat{\beta}_2}{se(\hat{\beta}_2)}$$
$$= \frac{68.080}{5.241} = 12.99$$

$$t_{0.025,40-3} = t_{0.025,37} = 2.0262$$
$$|t_0| > t_{\alpha/2,37},$$

Reject H_0 conclude that β_2 is significant in the model at α = 0.05

12-32

```
The regression equation is
y = - 171 + 7.03 x1 + 12.7 x2

Predictor      Coef  SE Coef      T      P
Constant    -171.26    28.40  -6.03  0.001
x1            7.029    1.539   4.57  0.004
x2           12.696    1.539   8.25  0.000


S = 3.07827   R-Sq = 93.7%   R-Sq(adj) = 91.6%

Analysis of Variance

Source          DF      SS      MS      F      P
Regression       2  842.37  421.18  44.45  0.000
Residual Error   6   56.85    9.48
Total            8  899.22
```

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

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

α = 0.05

$$f_0 = \frac{\frac{SS_R}{k}}{\frac{SS_E}{n-p}} = \frac{842.37/2}{56.85/6} = 44.45$$
$$f_{0.05,2,6} = 5.14$$
$$f_0 > f_{0.05,2,6}$$

Reject H_0 and conclude regression model is significant at α = 0.05 P-value ≈ 0

b) $\hat{\sigma}^2 = MS_E = 9.48$

$$se(\hat{\beta}_1) = \sqrt{\hat{\sigma}^2 c_{jj}} = 1.539$$

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

α = 0.05

$$t_0 = \frac{\hat{\beta}_1}{se(\hat{\beta}_1)}$$
$$= \frac{7.03}{1.539} = 4.568$$
$$t_{0.025,9-3} = t_{0.025,6} = 2.447$$
$$|t_0| > t_{\alpha/2,6},$$

Reject H_0, β_1 is significant in the model at α = 0.05

$$se(\hat{\beta}_2) = \sqrt{\hat{\sigma}^2 c_{jj}} = 1.539$$
$$H_0 : \beta_2 = 0$$
$$H_1 : \beta_2 \neq 0$$
$\alpha = 0.05$
$$t_0 = \frac{\hat{\beta}_2}{se(\hat{\beta}_2)}$$
$$= \frac{12.7}{1.539} = 8.252$$
$$t_{0.025,9-3} = t_{0.025,6} = 2.447$$
$$| t_0 | > t_{\alpha/2,6},$$
Reject H_0 conclude that β_2 is significant in the model at $\alpha = 0.05$

c) With a smaller sample size, the difference in the estimate from the hypothesized value needs to be greater to be significant.

12-33
```
Useful range (ng) = 239 + 0.334 Brightness (%) - 2.72 Contrast (%)

Predictor           Coef  SE Coef       T      P
Constant          238.56    45.23    5.27  0.002
Brightness (%)    0.3339   0.6763    0.49  0.639
Contrast (%)     -2.7167   0.6887   -3.94  0.008

S = 36.3493   R-Sq = 75.6%   R-Sq(adj) = 67.4%

Analysis of Variance
Source           DF     SS     MS     F      P
Regression        2  24518  12259  9.28  0.015
Residual Error    6   7928   1321
Total             8  32446
```

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

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

$\alpha = 0.01$
$$f_0 = \frac{\frac{SS_R}{k}}{\frac{SS_E}{n-p}} = \frac{24518/2}{7928/6} = 9.28$$
$$f_{0.01,2,6} = 10.92$$
$$f_0 < f_{0.01,2,6}$$
Fail to reject H_0 and conclude that the regression model is not significant at $\alpha = 0.01$ P-value = 0.015

b) $\hat{\sigma}^2 = MS_E = 1321$
$$se(\hat{\beta}_1) = \sqrt{\hat{\sigma}^2 c_{jj}} = 0.6763$$
$$H_0 : \beta_1 = 0$$
$$H_1 : \beta_1 \neq 0$$
$\alpha = 0.01$

$$t_0 = \frac{\hat{\beta}_1}{se(\hat{\beta}_1)}$$
$$= \frac{0.3339}{0.6763} = 0.49$$
$$t_{0.005,9-3} = t_{0.005,6} = 3.707$$

$|t_0| < t_{\alpha/2,6}$, Fail to reject H_0, there is no enough evidence to conclude that β_1 is significant in the model at $\alpha = 0.01$

$$se(\hat{\beta}_2) = \sqrt{\hat{\sigma}^2 c_{jj}} = 0.6887$$
$$H_0 : \beta_2 = 0$$
$$H_1 : \beta_2 \neq 0$$
$\alpha = 0.01$
$$t_0 = \frac{\hat{\beta}_2}{se(\hat{\beta}_2)}$$
$$= \frac{-2.7167}{0.6887} = -3.94$$
$$t_{0.005,9-3} = t_{0.005,6} = 3.707$$

$|t_0| > t_{\alpha/2,6}$, Reject H_0 conclude that β_2 is significant in the model at $\alpha = 0.01$

12-34
```
The regression equation is
Stack Loss(y) = - 39.9 + 0.716 X1 + 1.30 X2 - 0.152 X3

Predictor     Coef  SE Coef      T      P
Constant    -39.92    11.90  -3.36  0.004
X1          0.7156   0.1349   5.31  0.000
X2          1.2953   0.3680   3.52  0.003
X3         -0.1521   0.1563  -0.97  0.344

S = 3.24336   R-Sq = 91.4%   R-Sq(adj) = 89.8%

Analysis of Variance

Source          DF       SS      MS      F      P
Regression       3  1890.41  630.14  59.90  0.000
Residual Error  17   178.83   10.52
Total           20  2069.24
```

a)
$$H_0 : \beta_1 = \beta_2 = \beta_3 = 0$$
$H_1 : \beta_j \neq 0$ for at least one j

$\alpha = 0.01$
$$f_0 = \frac{\frac{SS_R}{k}}{\frac{SS_E}{n-p}} = \frac{189.41/3}{178.83/17} = 59.90$$
$$f_{0.01,3,17} = 5.18$$
$$f_0 > f_{0.01,3,17}$$

Reject H_0 and conclude that the regression model is significant at $\alpha = 0.01$ P-value < 0.000001

b) $\hat{\sigma}^2 = MS_E = 10.52$
$$se(\hat{\beta}_1) = \sqrt{\hat{\sigma}^2 c_{jj}} = 0.1349$$

$$H_0 : \beta_1 = 0$$
$$H_1 : \beta_1 \neq 0$$
α = 0.01

$$t_0 = \frac{\hat{\beta}_1}{se(\hat{\beta}_1)}$$
$$= \frac{0.7156}{0.1349} = 5.31$$
$$t_{0.005,21-4} = t_{0.005,17} = 2.898$$
$$|t_0| > t_{\alpha/2,17}.$$

Reject H_0 and conclude that β_1 is significant in the model at α = 0.01.

$$se(\hat{\beta}_2) = \sqrt{\hat{\sigma}^2 c_{jj}} = 0.3680$$
$$H_0 : \beta_2 = 0$$
$$H_1 : \beta_2 \neq 0$$
α = 0.01

$$t_0 = \frac{\hat{\beta}_2}{se(\hat{\beta}_2)}$$
$$= \frac{1.2953}{0.3680} = 3.52$$
$$t_{0.005,21-4} = t_{0.005,17} = 2.898$$

$|t_0| > t_{\alpha/2,17}$. Reject H_0 and conclude that β_2 is significant in the model at α = 0.01.

$$se(\hat{\beta}_3) = \sqrt{\hat{\sigma}^2 c_{jj}} = 0.1563$$
$$H_0 : \beta_3 = 0$$
$$H_1 : \beta_3 \neq 0$$
α = 0.01

$$t_0 = \frac{\hat{\beta}_3}{se(\hat{\beta}_3)}$$
$$= \frac{-0.1521}{0.1563} = -0.97$$
$$t_{0.005,21-4} = t_{0.005,17} = 2.898$$
$$|t_0| < t_{\alpha/2,17}.$$

Fail to reject H_0, there is not enough evidence to conclude that β_3 is significant in the model at α = 0.01.

12-35 a) Computer output follows. The test statistic is F = 191.09. Because the P-value is near zero, the regression is significant at $\alpha = 0.01$.

b) $t_0 = \dfrac{\hat{\beta}_j - \beta_{j0}}{se(\hat{\beta}_j)}$, null hypothesis $\hat{\beta}_j = \beta_{j0}$ is rejected at α level if $|t_0| > t_{\alpha/2,n-p}$ or the P-value < α

The P-values of all regressors are less than 0.01. Therefore, all individual variables in the model are significant.

c) The computer output for three regressors is followed by the computer output for two regressors. From the regression sum of squares in each model the F test for x_2 is $F_0 = \frac{SS_R(\beta_1 \mid \beta_2)/r}{MS_E} = \frac{2373.59 - 1782.96}{4.14} = 142.66$

The F-test P-value is near zero. Therefore the regressor (TD percentage) is significant to the model. This is the equivalent to the *t* test on the coefficient of x_2. The F statistic = 142.66 = 11.94^2, except for some round-off error.

Results of regression on three variables and on two variables are shown below.

Regression Analysis: Rating Pts versus Pct Comp, Pct TD, Pct Int

```
The regression equation is
Rating Pts = 2.99 + 1.20 Pct Comp + 4.60 Pct TD - 3.81 Pct Int

Predictor     Coef  SE Coef      T      P
Constant     2.986    5.877   0.51  0.615
Pct Comp   1.19857  0.09743  12.30  0.000
Pct TD      4.5956   0.3848  11.94  0.000
Pct Int    -3.8125   0.4861  -7.84  0.000

S = 2.03479   R-Sq = 95.3%   R-Sq(adj) = 94.8%

Analysis of Variance

Source          DF       SS      MS       F      P
Regression       3  2373.59  791.20  191.09  0.000
Residual Error  28   115.93    4.14
Total           31  2489.52

Source    DF   Seq SS
Pct Comp   1  1614.43
Pct TD     1   504.49
Pct Int    1   254.67


Unusual Observations

Obs  Pct Comp  Rating Pts     Fit  SE Fit  Residual  St Resid
 11      61.1      87.700  83.691   0.371     4.009     2.00R
 18      59.4      84.700  79.668   0.430     5.032     2.53R
 21      65.7      81.000  85.020   1.028    -4.020    -2.29R
 31      59.4      70.000  75.141   0.719    -5.141    -2.70R

R denotes an observation with a large standardized residual.
```

Regression Analysis: Rating Pts versus Pct Comp, Pct Int

```
The regression equation is
Rating Pts = - 9.1 + 1.66 Pct Comp - 3.08 Pct Int

Predictor    Coef  SE Coef      T      P
Constant    -9.11    14.04  -0.65  0.522
Pct Comp   1.6622   0.2168   7.67  0.000
Pct Int    -3.076    1.170  -2.63  0.014

S = 4.93600   R-Sq = 71.6%   R-Sq(adj) = 69.7%

Analysis of Variance

Source          DF       SS      MS      F      P
```

```
Regression       2  1782.96  891.48  36.59  0.000
Residual Error  29   706.56   24.36
Total           31  2489.52
```

12-36 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) $\alpha = 0.05$ $\quad t_{\alpha/2,n-p} = t_{.025,26} = 2.056$

$H_0: \beta_1 = 0$	$\beta_2 = 0$	$\beta_3 = 0$	$\beta_4 = 0$	$\beta_5 = 0$
$H_1: \beta_1 \neq 0$	$\beta_2 \neq 0$	$\beta_3 \neq 0$	$\beta_4 \neq 0$	$\beta_5 \neq 0$
$t_0 = 0.83$	$t_0 = 12.25$	$t_0 = -0.52$	$t_0 = 6.96$	$t_0 = -0.29$
Fail to reject H_0	Reject H_0	Fail to reject H_0	Reject H_0	Fail to 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 c) is smaller, suggesting a better model.

12-37 $\hat{y} = \hat{\beta}_0 + \hat{\beta}_1 x_1 + \hat{\beta}_2 x_2$, Assume no interaction model.

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

H_1 at least one $\beta_j \neq 0$

$f_0 = 97.59$

$f_{0.01,2,3} = 30.82$

$f_0 > f_{0.01,2,3}$

Reject H_0 P-value = 0.002

b) $H_0 : \beta_1 = 0$
$H_1 : \beta_1 \neq 0$
$t_0 = -6.42$
$t_{\alpha/2,3} = t_{0.005,3} = 5.841$
$|t_0| > t_{0.005,3}$
Reject H_0 for regressor β_1.

$H_0 : \beta_2 = 0$
$H_1 : \beta_2 \neq 0$
$t_0 = -2.57$
$t_{\alpha/2,3} = t_{0.005,3} = 5.841$
$|t_0| \not> t_{0.005,3}$
Do not reject H_0 for regressor β_2.

c) $SS_R(\beta_2 \mid \beta_1, \beta_0) = 1012$
$H_0 : \beta_2 = 0$
$H_1 : \beta_2 \neq 0$
$\alpha = 0.01$
$f_0 = 6.629$
$f_{\alpha,1,3} = f_{0.01,1,3} = 34.12$
$f_0 \not> f_{0.05,1,3}$
Do not reject H_0

d) $H_0 : \beta_1 = \beta_2 = \beta_{12} = 0$
H_1 at least one $\beta_j \neq 0$
$\alpha = 0.01$
$f_0 = 7.714$
$f_{\alpha,3,2} = f_{0.01,3,2} = 99.17$
$f_0 \not> f_{0.01,3,2}$
Do not reject H_0

e) $H_0 : \beta_{12} = 0$
$H_1 : \beta_{12} \neq 0$
$\alpha = 0.01$
$SSR(\beta_{12} \mid \beta_1, \beta_2) = 29951.4 - 29787 = 163.9$
$$f_0 = \frac{SSR}{MS_E} = \frac{163.9}{147} = 1.11$$
$f_{0.01,1,2} = 98.50$
$f_0 \not> f_{0.01,1,2}$
Do not reject H_0

f) $\hat{\sigma}^2 = 111.2$
$\hat{\sigma}^2$ (no interaction term) = 159
$MS_E(\hat{\sigma}^2)$ was reduced in the model with the interaction term.

12-38 a) $H_0 : \beta_j = 0$ for all j
$H_1 : \beta_j \neq 0$ for at least one j
From the computer output
$f_0 = 14.09$
$f_{0.01,14,15} = 3.56$
$f_0 > f_{0.01,14,15}$
Reject H_0 and conclude that the regression model is significant at $\alpha = 0.01$

b) $H_0 : \beta_j = 0$

$H_1 : \beta_j \neq 0$

$t_{0.005,15} = 2.947$

GF :	$t_0 = 4.46$	Reject H_0
GA :	$t_0 = -3.83$	Reject H_0
ADV :	$t_0 = -0.25$	Fail to reject H_0
PPGF :	$t_0 = 0.08$	Fail to reject H_0
PCTG :	$t_0 = -0.04$	Fail to reject H_0
PEN :	$t_0 = -0.54$	Fail to reject H_0
BMI :	$t_0 = -0.45$	Fail to reject H_0
AVG :	$t_0 = 0.53$	Fail to reject H_0
SHT :	$t_0 = 2.19$	Fail to reject H_0
PPGA :	$t_0 = -2.50$	Fail to reject H_0
PKPCT :	$t_0 = -2.54$	Fail to reject H_0
SHGF :	$t_0 = 0.54$	Fail to reject H_0
SHGA :	$t_0 = 2.34$	Fail to reject H_0
FG :	$t_0 = 0.02$	Fail to reject H_0

It does not seem that all regressors are important. Only the regressors "GF" (β_1) and "GA" (β_2) are significant at $\alpha = 0.01$

c) The computer result is shown below.

Regression Analysis: W versus GF, PPGF

```
The regression equation is
W = - 8.82 + 0.218 GF - 0.016 PPGF

Predictor      Coef  SE Coef      T      P
Constant     -8.818    9.230  -0.96  0.348
GF          0.21779  0.05467   3.98  0.000
PPGF        -0.0162   0.1134  -0.14  0.888

S = 5.11355   R-Sq = 52.8%   R-Sq(adj) = 49.3%

Analysis of Variance

Source          DF       SS      MS      F      P
Regression       2   789.99  395.00  15.11  0.000
Residual Error  27   706.01   26.15
Total           29  1496.00
```

Because PPGF had a *t* statistic near zero in part (b) there is a concern that it is not an important predictor. We will evaluate its role in the smaller model with GF.

$$\hat{y} = -8.82 + 0.218x_1 - 0.16x_4$$

$$f_0 = 15.11$$

$$f_{0.05,2,27} = 3.35$$

Because $f_0 > f_{0.05,2,27}$, we reject the null hypothesis that the coefficient of GF and PPGF are both zero.

$$H_0 : \beta_1 = 0 \qquad \beta_4 = 0$$
$$H_1 : \beta_1 \neq 0 \qquad \beta_4 \neq 0$$
$$t_0 = 3.98 \qquad t_0 = -0.14$$

Reject H_0 Fail to reject H_0

Based on the t-test, power play goals for (PPGF) is not a logical choice to add to the model that already contains GF.

12-39 a) The computer output follows. The P-value for the F-test is near zero. Therefore, the regression is significant at both $\alpha = 0.05$ or $\alpha = 0.01$

b) $t_0 = \dfrac{\hat{\beta}_j - \beta_{j0}}{se(\hat{\beta}_j)}$. Because the P-values for Age and Severity are < 0.05 both regressors are significant to the model.

Because the P-value for Anxiety is 0.233, it is not significant to the model at level $\alpha = 0.05$.

Regression Analysis: Satisfaction versus Age, Severity, Anxiety

```
The regression equation is
Satisfaction = 144 - 1.11 Age - 0.585 Severity + 1.30 Anxiety

Predictor      Coef  SE Coef       T      P
Constant    143.895    5.898   24.40  0.000
Age         -1.1135   0.1326   -8.40  0.000
Severity    -0.5849   0.1320   -4.43  0.000
Anxiety       1.296    1.056    1.23  0.233

S = 7.03710   R-Sq = 90.4%   R-Sq(adj) = 89.0%

Analysis of Variance

Source          DF       SS      MS      F      P
Regression       3   9738.3  3246.1  65.55  0.000
Residual Error  21   1039.9    49.5
Total           24  10778.2

Source    DF  Seq SS
Age        1  8756.7
Severity   1   907.0
Anxiety    1    74.6
```

12-40 a) **Regression Analysis: Satisfaction versus Age, Severity**

```
The regression equation is
Satisfaction = 143 - 1.03 Age - 0.556 Severity

Predictor      Coef  SE Coef       T      P
Constant    143.472    5.955   24.09  0.000
Age         -1.0311   0.1156   -8.92  0.000
Severity    -0.5560   0.1314   -4.23  0.000

S = 7.11767   R-Sq = 89.7%   R-Sq(adj) = 88.7%

Analysis of Variance

Source          DF       SS      MS      F      P
Regression       2   9663.7  4831.8  95.38  0.000
Residual Error  22   1114.5    50.7
Total           24  10778.2
```

Because the P-value of the F test is less than $\alpha = 0.05$ and $\alpha = 0.01$, we reject the H_0 and conclude that at least one regressor contributes significantly to the model at either α level.

b) Because the P-values from the t-test for both *age* and *severity* regressors are less than $\alpha = 0.05$, we reject the H_0 and conclude that both *age* and *severity* regressors contribute significantly to the model.

c) From $MS_{Residual}$, the estimate of the variance = 50.7. From the computer output below, if the third variable *anxiety* is added to the model, the estimate of the variance is reduced to 49.5. The variance changed very slightly here so it is unlikely that the variable contributes significantly to the model.

Regression Analysis: Satisfaction versus Age, Severity, Anxiety

```
The regression equation is
Satisfaction = 144 - 1.11 Age - 0.585 Severity + 1.30 Anxiety

Predictor     Coef  SE Coef      T      P
Constant   143.895    5.898  24.40  0.000
Age        -1.1135   0.1326  -8.40  0.000
Severity   -0.5849   0.1320  -4.43  0.000
Anxiety      1.296    1.056   1.23  0.233

S = 7.03710   R-Sq = 90.4%   R-Sq(adj) = 89.0%

Analysis of Variance

Source          DF       SS      MS      F      P
Regression       3   9738.3  3246.1  65.55  0.000
Residual Error  21   1039.9    49.5
Total           24  10778.2
```

<u>Sections 12-3 and 12-4</u>

12-41 a) $\hat{\beta}_0 \pm t_{\alpha/2,n-p}\sqrt{\hat{\sigma}^2 c_{00}}$

$$171.055 \pm t_{.025,7}\ se(\hat{\beta}_0)$$
$$171.055 \pm (2.365)(51.217)$$
$$171.055 \pm 121.128$$
$$49.927 \le \beta_0 \le 292.183$$

$$\hat{\beta}_1 \pm t_{\alpha/2,n-p}\sqrt{\hat{\sigma}^2 c_{11}}$$
$$3.713 \pm t_{.025,7}\ se(\hat{\beta}_1)$$
$$3.713 \pm (2.365)(1.556)$$
$$3.713 \pm 3.680$$
$$0.033 \le \beta_1 \le 7.393$$

$$\hat{\beta}_2 \pm t_{\alpha/2,n-p}\sqrt{\hat{\sigma}^2 c_{22}}$$
$$-1.126 \pm t_{.025,7}\ se(\hat{\beta}_0)$$
$$-1.126 \pm (2.365)(0.693)$$
$$-1.126 \pm 1.639$$
$$-2.765 \le \beta_2 \le 0.513$$

b) $x_1 = 18$
$x_2 = 43$
$\hat{y}_0 = 189.471$
$X_0'(X'X)^{-1}X_0 = 0.305065$
$189.471 \pm (2.365)\sqrt{550.7875(0.305065)}$
$158.815 \le \mu_{Y|x_0} \le 220.127$

c) $\alpha = 0.05$
$x_1 = 18$
$x_2 = 43$
$\hat{y}_0 = 189.471$
$X_0'(X'X)^{-1}X_0 = 0.305065$
$\hat{y} \pm t_{\alpha/2,n-p}\sqrt{\hat{\sigma}^2(1 + X_0'(X'X)^{-1}X_0)}$
$189.471 \pm (2.365)\sqrt{550.7875(1.305065)}$
$126.064 \le y_0 \le 252.878$

12-42 a) $\hat{\beta}_0 \pm t_{\alpha/2,n-p}\sqrt{\hat{\sigma}^2 c_{00}}$
$-1.9122 \pm t_{.025,7}\ se(\hat{\beta}_0)$
$-1.9122 \pm (2.365)(10.055)$
-1.9122 ± 23.78
$-25.6922 \le \beta_0 \le 21.8678$
$\hat{\beta}_1 \pm t_{\alpha/2,n-p}\sqrt{\hat{\sigma}^2 c_{11}}$
$0.0931 \pm t_{.025,7}\ se(\hat{\beta}_1)$
$0.0931 \pm (2.365)(0.0827)$
0.0931 ± 0.1956
$-0.1025 \le \beta_1 \le 0.2887$

$\hat{\beta}_2 \pm t_{\alpha/2,n-p}\sqrt{\hat{\sigma}^2 c_{22}}$
$0.2532 \pm t_{.025,7}\ se(\hat{\beta}_0)$
$0.2532 \pm (2.365)(0.1998)$
0.2532 ± 0.4725
$-0.2193 \le \beta_2 \le 0.7257$

b) $x_1 = 200$
$x_2 = 50$
$\hat{y}_0 = 29.37$
$X_0'(X'X)^{-1}X_0 = 0.211088$
$29.37 \pm (2.365)\sqrt{85.694(0.211088)}$
29.37 ± 10.059
$19.311 \le \mu_{Y|x_0} \le 39.429$

c) $\alpha = 0.05$

$$x_1 = 200$$
$$x_2 = 50$$
$$\hat{y}_0 = 29.37$$
$$X_0'(X'X)^{-1}X_0 = 0.211088$$
$$\hat{y} \pm t_{\alpha/2,n-p}\sqrt{\hat{\sigma}^2(1+X_0'(X'X)^{-1}X_0)}$$
$$29.37 \pm (2.365)\sqrt{85.694(1.211088)}$$
$$29.37 \pm 24.093$$
$$5.277 \le y_0 \le 53.463$$

12-43

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	3	30532	10177	840.55	<.0001
Error	16	193.72482	12.10780		
Corrected Total	19	30725			

Root MSE	3.47963	R-Square	0.9937
Dependent Mean	109.22600	Adj R-Sq	0.9925
Coeff Var	3.18571		

Parameter Estimates

Variable	DF	Parameter Estimate	Standard Error	t Value	Pr > |t|	99% Confidence Limits	
Intercept	1	47.17400	49.58148	0.95	0.3555	-97.64267	191.99066
x1	1	-9.73520	3.69162	-2.64	0.0179	-20.51763	1.04723
x2	1	0.42829	0.22393	1.91	0.0739	-0.22577	1.08235
x3	1	18.23745	1.31180	13.90	<.0001	14.40597	22.06894

Output Statistics

Obs	Dependent Variable	Predicted Value	Std Error Mean Predict	99% CL Mean		99% CL Predict		Residual
21	.	101.0957	4.6646	87.4713	114.7201	84.0982	118.0932	.

a) $-20.518 \le \beta_1 \le 1.047$

$$-0.226 \le \beta_2 \le 1.082$$
$$14.406 \le \beta_3 \le 22.069$$

b) $\hat{y}_0 = 101.0957$

$84.098 \le y_0 \le 118.093$

c) $\hat{\mu}_{Y|x_0} = 101.0957$

$87.471 \le \mu_{Y|x_0} \le 114.720$

12-44 a) 95 % CI on coefficients

$0.0973 \le \beta_1 \le 1.4172$

$-3.61373 \le \beta_2 \le 21.4610$

$-4.21947 \le \beta_3 \le 7.09438$

$-1.72211 \le \beta_4 \le 1.74932$

b) $\hat{\mu}_{Y|x_0} = 292.65 \quad se(\hat{\mu}_{Y|x_0}) = 14.49 \qquad t_{.025,7} = 2.365$

$\hat{\mu}_{Y|x_0} \pm t_{\alpha/2,n-p} \quad se(\hat{\mu}_{Y|x_0})$

$292.65 \pm (2.365)(14.49)$

$258.38 \le \mu_{Y|x_0} \le 326.92$

c) $\hat{y}_0 \pm t_{\alpha/2,n-p}\sqrt{\hat{\sigma}^2(1+X_0'(X'X)^{-1}X_0}$

$292.65 \pm 2.365(21.949)$

$240.74 \le y_0 \le 344.56$

12-45 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.

12-46

Parameter Estimates

Variable	DF	Parameter Estimate	Standard Error	t Value	Pr > |t|	95% Confidence Limits	
Intercept	1	7.45781	7.22630	1.03	0.3196	−8.04106	22.95667
x2	1	−0.02970	0.26327	−0.11	0.9118	−0.59436	0.53495
x3	1	0.52051	0.13590	3.83	0.0018	0.22903	0.81199
x4	1	−0.10180	0.05339	−1.91	0.0773	−0.21632	0.01271
x5	1	−2.16058	2.39473	−0.90	0.3822	−7.29677	2.97561

Output Statistics								
Obs	Dependent Variable	Predicted Value	Std Error Mean Predict	95% CL Mean		95% CL Predict		Residual
20	.	4.8086	1.8766	0.7836	8.8336	0.3606	9.2566	.

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) $\hat{\mu}_{Y|x_0} = 4.80868$

$0.7836 \le \mu_{Y|x_0} \le 8.8336$

c) $\hat{y}_0 = 4.8086$

$0.3606 \le y_0 \le 9.2566$

12-47

Parameter Estimates							
Variable	DF	Parameter Estimate	Standard Error	t Value	Pr > |t|	95% Confidence Limits	
Intercept	1	-440.39250	94.19757	-4.68	<.0001	-631.25491	-249.53009
mAmps	1	19.14750	3.46047	5.53	<.0001	12.13593	26.15907
ExposureTime	1	68.08000	5.24107	12.99	<.0001	57.46059	78.69941

Output Statistics								
Obs	Dependent Variable	Predicted Value	Std Error Mean Predict	99% CL Mean		99% CL Predict		Residual
41	.	10.6375	48.1494	-120.1079	141.3829	-642.6513	663.9263	.

a)

$$\hat{\beta}_1 \pm t_{\alpha/2,n-p}\sqrt{\hat{\sigma}^2 c_{11}}$$
$$19.147 \pm t_{.025,37}\ se(\hat{\beta}_1)$$
$$19.147 \pm (2.0262)(3.460)$$
$$19.147 \pm 7.014458$$
$$12.136 \le \beta_1 \le 26.159$$

$$\hat{\beta}_2 \pm t_{\alpha/2,n-p}\sqrt{\hat{\sigma}^2 c_{22}}$$
$$68.080 \pm t_{.025,37}\ se(\hat{\beta}_2)$$
$$68.080 \pm (2.0262)(5.241)$$
$$68.080 \pm 7.014458$$
$$57.461 \le \beta_2 \le 78.700$$

b) $\hat{\mu}_{Y|x_0} = 10.6375$

$$\hat{\mu}_{Y|x_0} \pm t_{\alpha/2,n-p} \quad se(\hat{\mu}_{Y|x_0})$$

$$-120.108 \le \mu_{Y|x_0} \le 141.383$$

c) $\hat{y}_0 = 10.6375$

$$-642.651 \le y_0 \le 663.926$$

12-48

Parameter Estimates

Variable	DF	Parameter Estimate	Standard Error	t Value	Pr > |t|	99% Confidence Limits	
Intercept	1	0.00106	0.90673	0.00	0.9991	−2.62686	2.62898
Age	1	0.00858	0.00708	1.21	0.2423	−0.01195	0.02911
DrinkUse	1	−0.02076	0.10180	−0.20	0.8408	−0.31581	0.27429
CookUse	1	0.00970	0.17981	0.05	0.9576	−0.51142	0.53083

Output Statistics

Obs	Dependent Variable	Predicted Value	Std Error Mean Predict	99% CL Mean		99% CL Predict		Residual
22	.	0.4288	0.1965	-0.1406	0.9981	-1.1449	2.0025	.

a) $t_{0.005,17} = 2.898$

$$-2.627 \le \beta_0 \le 2.629$$

$$-0.012 \le \beta_1 \le 0.029$$

$$-0.316 \le \beta_2 \le 0.274$$

$$-0.511 \le \beta_3 \le 0.531$$

b) $\hat{\mu}_{Y|x_0} = 0.4288$

$$\hat{\mu}_{Y|x_0} \pm t_{\alpha/2,n-p} \quad se(\hat{\mu}_{Y|x_0})$$

$$-0.141 \le \mu_{Y|x_0} \le 0.998$$

c) $\hat{y}_0 = 0.4288$

$$-1.145 \le y_0 \le 2.003$$

12-49 a) $t_{0.05,8} = 1.860$

$-0.576 \le \beta_0 \le 0.355$

$0.0943 \le \beta_1 \le 0.7201$

$-8.743 \le \beta_2 \le 12.959$

b) $\hat{\mu}_{Y|x_0} = 0.8787$ $\quad se(\hat{\mu}_{Y|x_0}) = 0.00926$ $\quad t_{0.005,16} = 1.860$

$$\hat{\mu}_{Y|x_0} \pm t_{\alpha/2,n-p} \;\; se(\hat{\mu}_{Y|x_0})$$
$$0.8787 \pm (1.860)(0.00926)$$
$$0.86148 \le \mu_{Y|x_0} \le 0.89592$$

c) $\hat{y}_0 = 0.8787$ $\qquad se(\hat{y}_0) = 0.0134$
$$0.8787 \pm 1.86(0.0134)$$
$$0.85490 \le y_0 \le 0.90250$$

12-50

```
The regression equation is
y = - 171 + 7.03 x1 + 12.7 x2

Predictor      Coef  SE Coef      T      P
Constant    -171.26    28.40  -6.03  0.001
x1            7.029    1.539   4.57  0.004
x2           12.696    1.539   8.25  0.000

S = 3.07827   R-Sq = 93.7%   R-Sq(adj) = 91.6%

Analysis of Variance

Source          DF      SS      MS      F      P
Regression       2  842.37  421.18  44.45  0.000
Residual Error   6   56.85    9.48
Total            8  899.22
```

a) $\hat{\beta}_1 \pm t_{\alpha/2,n-p}\sqrt{\hat{\sigma}^2 c_{11}}$
$$7.03 \pm t_{0.05,6} \; se(\hat{\beta}_1)$$
$$7.03 \pm (1.943)(1.539)$$
$$7.03 \pm 2.9903$$
$$4.0397 \le \beta_1 \le 10.0203$$

$$\hat{\beta}_2 \pm t_{\alpha/2,n-p}\sqrt{\hat{\sigma}^2 c_{22}}$$
$$12.7 \pm t_{0.05,6} \; se(\hat{\beta}_2)$$
$$12.7 \pm (1.943)(1.539)$$
$$12.7 \pm 2.9903$$
$$9.7097 \le \beta_1 \le 15.6903$$

b)

```
New
Obs     Fit  SE Fit        90% CI               90% PI
  1  140.82    6.65  (127.899, 153.74)  (126.58, 155.06)XX
```

$\hat{\mu}_{Y|x_0} = 140.82 \quad se(\hat{\mu}_{Y|x_0}) = 6.65 \qquad t_{0.025,6} = 2.447$
$$\hat{\mu}_{Y|x_0} \pm t_{\alpha/2,n-p} \;\; se(\hat{\mu}_{Y|x_0})$$
$$140.82 \pm (1.943)(6.65)$$
$$127.899 \le \mu_{Y|x_0} \le 153.74$$

c) $\hat{y}_0 = 140.82$ $\qquad se(\hat{y}_0) = 7.33$
$$140.82 \pm 1.943(7.33)$$
$$126.58 \le y_0 \le 155.06$$

d) The smaller the sample size, the wider the interval

12-51

```
The regression equation is
Useful range (ng) = 239 + 0.334 Brightness (%) - 2.72 Contrast (%)

Predictor           Coef  SE Coef      T      P
Constant          238.56    45.23   5.27  0.002
Brightness (%)    0.3339   0.6763   0.49  0.639
Contrast (%)     -2.7167   0.6887  -3.94  0.008

S = 36.3493   R-Sq = 75.6%   R-Sq(adj) = 67.4%

Analysis of Variance

Source          DF     SS     MS     F      P
Regression       2  24518  12259  9.28  0.015
Residual Error   6   7928   1321
Total            8  32446
```

a) $t_{0.005,6} = 3.707$

$-2.173 \le \beta_1 \le 2.841$

$-5.270 \le \beta_2 \le -0.164$

b) Predicted Values for New Observations

```
New
Obs   Fit  SE Fit       99% CI            99% PI
  1  44.6    21.9  (-36.7, 125.8)  (-112.8, 202.0)

Values of Predictors for New Observations

New                    Contrast
Obs  Brightness (%)         (%)
  1            70.0        80.0
```

$\hat{\mu}_{Y|x_0} = 44.6 \qquad se(\hat{\mu}_{Y|x_0}) = 21.9 \qquad t_{0.005,6} = 3.707$

$\hat{\mu}_{Y|x_0} \pm t_{\alpha/2,n-p} \quad se(\hat{\mu}_{Y|x_0})$

$44.6 \pm (3.707)(21.9)$

$-36.7 \le \mu_{Y|x_0} \le 125.8$

c) $\hat{y}_0 = 44.6 \qquad se(\hat{y}_0) = 42.44$

$44.6 \pm 3.707(42.44)$

$-112.8 \le y_0 \le 202.0$

d) Predicted Values for New Observations

```
New
Obs    Fit  SE Fit       99% CI           99% PI
  1  187.3    21.6  (107.4, 267.2)  (30.7, 344.0)

Values of Predictors for New Observations

New                    Contrast
Obs  Brightness (%)         (%)
  1            50.0        25.0
```

CI: $107.4 \le \mu_{Y|x_0} \le 267.2$

PI: $30.7 \le y_0 \le 344.0$

These intervals are wider because the regressors are set at extreme values in the *x* space and the standard errors are greater.

12-52

Parameter Estimates

Variable	DF	Parameter Estimate	Standard Error	t Value	Pr > |t|	95% Confidence Limits	
Intercept	1	−39.91967	11.89600	−3.36	0.0038	−65.01803	−14.82131
x1	1	0.71564	0.13486	5.31	<.0001	0.43111	1.00017
x2	1	1.29529	0.36802	3.52	0.0026	0.51882	2.07175
x3	1	−0.15212	0.15629	−0.97	0.3440	−0.48187	0.17763

a) $t_{0.025,17} = 2.110$

$$-0.431 \le \beta_1 \le 1.00$$

$$0.519 \le \beta_2 \le 2.072$$

$$-0.482 \le \beta_3 \le 0.178$$

b)

Output Statistics

Obs	Dependent Variable	Predicted Value	Std Error Mean Predict	95% CL Mean		95% CL Predict		Residual
22	.	30.3068	3.0975	23.7716	36.8421	20.8446	39.7691	.

Prediction at $x_1 = 80$, $x_2 = 20$, $x_3 = 85$ is

$\hat{\mu}_{Y|x_0} = 30.307$

$$23.772 \le \mu_{Y|x_0} \le 36.842$$

c) $\hat{y}_0 = 30.307$

$$20.845 \le y_0 \le 39.769$$

d)

Output Statistics

Obs	Dependent Variable	Predicted Value	Std Error Mean Predict	95% CL Mean		95% CL Predict		Residual
22	.	27.7946	3.2062	21.0301	34.5591	18.1726	37.4166	.

Prediction at $x_1 = 80$, $x_2 = 19$, $x_3 = 93$ is $\hat{\mu}_{Y|x_0} = 27.795$

CI: $21.030 \le \mu_{Y|x_0} \le 34.559$

PI: $18.173 \le y_0 \le 37.417$

12-53 a) The computer output follows. The output is used to obtain estimates of the coefficients and standard errors. The confidence intervals for the coefficients are computed from

$$\hat{\beta} - t_{0.025,28} se(\hat{\beta}) \le \beta \le \hat{\beta} + t_{0.025,28} se(\hat{\beta}).$$

Parameter Estimates

Variable	DF	Parameter Estimate	Standard Error	t Value	Pr > |t|	95% Confidence Limits	
Intercept	1	2.98566	5.87686	0.51	0.6154	-9.05254	15.02385
PctComp	1	1.19857	0.09743	12.30	<.0001	0.99899	1.39814
PctTD	1	4.59561	0.38477	11.94	<.0001	3.80744	5.38379
PctInt	1	-3.81251	0.48612	-7.84	<.0001	-4.80827	-2.81674

From the t table, $t_{0.025,\,28} = 2.048$. The confidence intervals for the β's are

$$\begin{bmatrix} -9.052 \\ 0.999 \\ 3.807 \\ -4.808 \end{bmatrix} \le \beta \le \begin{bmatrix} 15.024 \\ 1.398 \\ 5.384 \\ -2.817 \end{bmatrix}$$

b)

Output Statistics

Obs	Dependent Variable	Predicted Value	Std Error Mean Predict	95% CL Mean		95% CL Predict		Residual
33	.	68.2267	1.2895	65.5853	70.8681	63.2921	73.1612	.

From the computer output $\hat{\mu}_{Y|x_0} = 68.23, se(\hat{\mu}_{Y|x_0}) = \sqrt{V(\hat{\mu}_{Y|x_0})} = \sigma\sqrt{x_0'(X'X)^{-1}x_0} = 1.29$

c) $\hat{\mu}_{Y|x_0} - t_{0.025,28} se(\hat{\mu}_{Y|x_0}) \le \mu_{Y|x_0} \le \hat{\mu}_{Y|x_0} + t_{0.025,28} se(\hat{\mu}_{Y|x_0})$,

$$65.59 \le \mu_{Y|x_0} \le 70.87$$

12-54 a) $-0.00003 \le \beta_1 \le 0.000087$

$0.002 \le \beta_2 \le 0.00261$

$-0.012999 \le \beta_3 \le 0.00694$

$0.0064 \le \beta_4 \le 0.01055$

$-0.01614 \le \beta_5 \le 0.01142$

b) $\hat{\mu}_{Y|x_0} = 0.022466$ $\quad se(\hat{\mu}_{Y|x_0}) = 0.000595$ $\quad t_{0.05,26} = 1.706$

$0.0220086 \pm (1.706)(0.000595)$

$0.02099 \le \mu_{Y|x_0} \le 0.0230$

c) $\hat{\mu}_{Y|x_0} = 0.0171$ $\quad se(\hat{\mu}_{Y|x_0}) = 0.000548$ $\quad t_{0.05,29} = 1.699$

$0.0171 \pm (1.699)(0.000548)$

$0.0162 \le \mu_{Y|x_0} \le 0.0180$

d) : width = 0.0018
: width = 0.0020
The interaction model has a shorter confidence interval. Yes, this suggests the interaction model is preferable.

12-55 a) $t_{0.025,14} = 2.145$

$$7.708 \le \beta_0 \le 92.092$$
$$-0.06 \le \beta_2 \le 0.04$$
$$-0.036 \le \beta_3 \le 0.034$$
$$-0.0053 \le \beta_7 \le -0.0012$$
$$-3.494 \le \beta_8 \le 4.078$$
$$-6.706 \le \beta_9 \le -1.004$$
$$-0.567 \le \beta_{10} \le 0.605$$

b) $\hat{\mu}_{Y|x_0} = 29.71 \quad se(\hat{\mu}_{Y|x_0}) = 1.395$

$$\hat{\mu}_{Y|x_0} \pm t_{0.025,14} \; se(\hat{\mu}_{Y|x_0})$$
$$29.71 \pm (2.145)(1.395)$$
$$26.718 \le \mu_{Y|x_0} \le 32.702$$

c) $\hat{y} = 61.001 - 0.0208x_2 - 0.0035x_7 - 3.457x_9$

$$t_{0.025,17} = 2.110$$
$$53.614 \le \beta_0 \le 68.388$$
$$-0.032 \le \beta_2 \le -0.01$$
$$-0.0053 \le \beta_7 \le -0.0017$$
$$-5.662 \le \beta_9 \le -1.252$$

d) The intervals in part c) are narrower. All of the regressors used in part c) are significant, but not all of those used in part a) are significant. The model used in part c) is preferable.

12-56 a) From the Minitab output in Exercise 12-18 the estimate, standard error, t statistic and P-value for the coefficient of GF are:

```
Predictor       Coef  SE Coef      T      P
GF           0.16374  0.03673   4.46  0.000
```

The 95% CI on the regression coefficient β_1 of GF is

$$\hat{\beta}_1 - t_{\alpha/2,n-p} se(\hat{\beta}_1) \le \hat{\beta}_1 \le \hat{\beta}_1 + t_{\alpha/2,n-p} se(\hat{\beta}_1)$$
$$\hat{\beta}_1 - t_{0.005,15} se(\hat{\beta}_1) \le \hat{\beta}_1 \le \hat{\beta}_1 + t_{0.005,15} se(\hat{\beta}_1)$$
$$0.16374 - (2.947)(0.03673) \le \hat{\beta}_1 \le 0.16374 + (2.947)(0.03673)$$
$$0.055497 \le \hat{\beta}_1 \le 0.271983$$

b) The Minitab result is shown below.

Regression Analysis: W versus GF

```
The regression equation is
W = - 8.57 + 0.212 GF

Predictor      Coef  SE Coef      T      P
Constant     -8.574    8.910  -0.96  0.344
GF          0.21228  0.03795   5.59  0.000
```

```
S = 5.02329   R-Sq = 52.8%   R-Sq(adj) = 51.1%

Analysis of Variance

Source          DF       SS      MS      F      P
Regression       1   789.46  789.46  31.29  0.000
Residual Error  28   706.54   25.23
Total           29  1496.00
```

$\hat{y} = -8.57 + 0.212x_1$

c) The 95% CI on the regression coefficient β_1 of GF is

$$\hat{\beta}_1 - t_{\alpha/2,n-p} se(\hat{\beta}_1) \le \hat{\beta}_1 \le \hat{\beta}_1 + t_{\alpha/2,n-p} se(\hat{\beta}_1)$$
$$\hat{\beta}_1 - t_{0.005,28} se(\hat{\beta}_1) \le \hat{\beta}_1 \le \hat{\beta}_1 + t_{0.005,28} se(\hat{\beta}_1)$$
$$0.21228 - (2.763)(0.03795) \le \hat{\beta}_1 \le 0.21228 + (2.763)(0.03795)$$
$$0.104856 \le \hat{\beta}_1 \le 0.317136$$

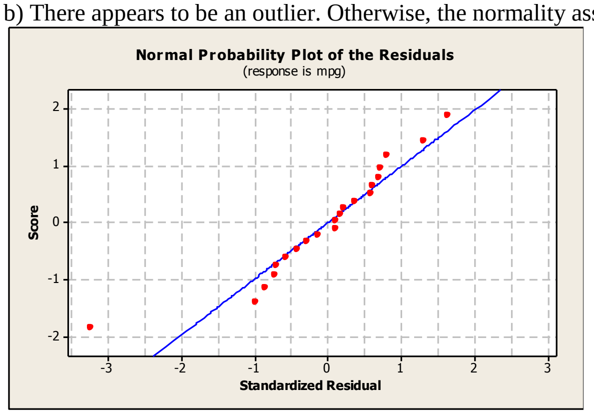
d) The simple linear regression model has the narrower interval. Obviously there are extraneous variables in the model from part a). The shorter interval is an initial indicator that the original model with all variables might be improved. One might expect there are other good predictors in the model from part a), only one of which is included in the model of part b).

Section 12-5

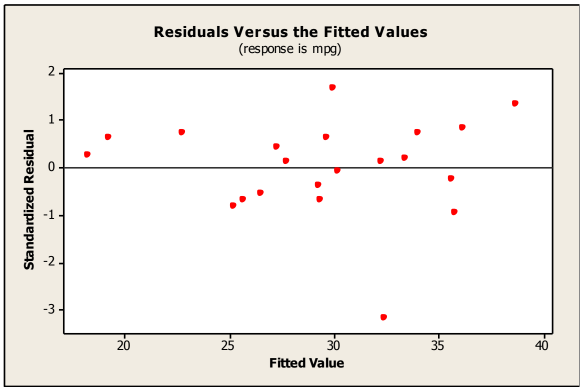
12-57 a)

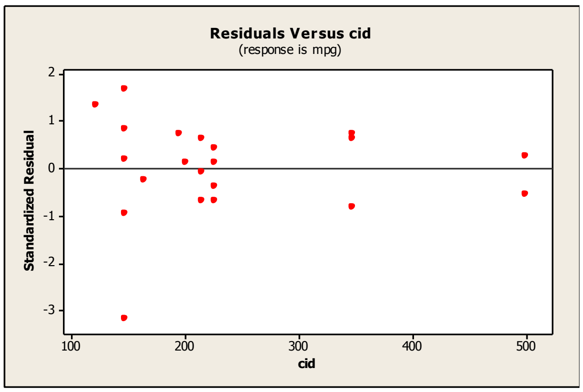
```
The regression equation is
mpg = 49.9 - 0.0104 cid - 0.0012 rhp - 0.00324 etw + 0.29 cmp - 3.86 axle
      + 0.190 n/v

Predictor        Coef    SE Coef      T      P
Constant        49.90      19.67   2.54  0.024
cid          -0.01045    0.02338  -0.45  0.662
rhp          -0.00120    0.01631  -0.07  0.942
etw        -0.0032364  0.0009459  -3.42  0.004
cmp             0.292      1.765   0.17  0.871
axle           -3.855      1.329  -2.90  0.012
n/v            0.1897     0.2730   0.69  0.498

S = 2.22830   R-Sq = 89.3%   R-Sq(adj) = 84.8%
```

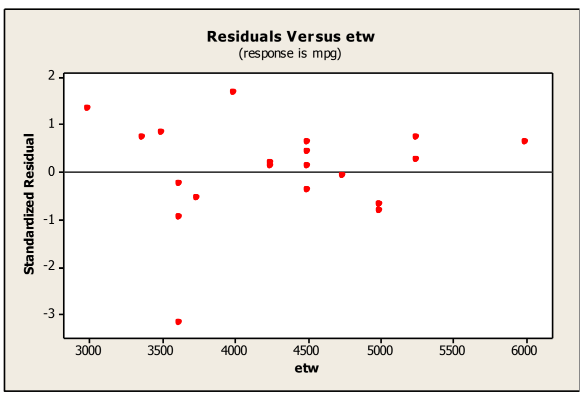
b) There appears to be an outlier. Otherwise, the normality assumption is not violated.

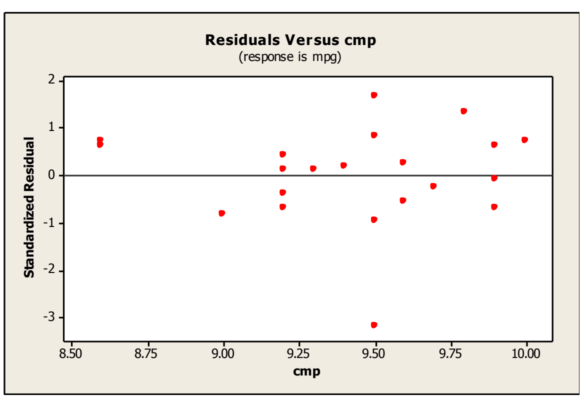


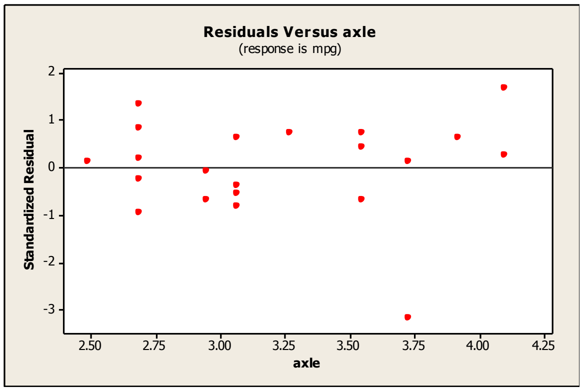
c) The plots do not show any violations of the assumptions.

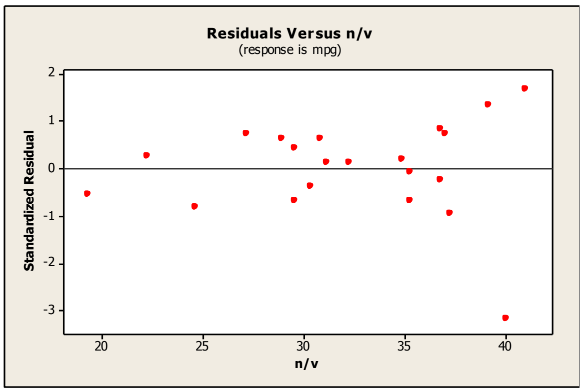

Residuals Versus the Fitted Values
(response is mpg)
Standardized Residual
2
1
0
-1
-2
-3
20
25
30
35
40
Fitted Value


Residuals Versus cid
(response is mpg)
Standardized Residual
2
1
0
-1
-2
-3
100
200
300
400
500
cid


Residuals Versus etw
(response is mpg)
Standardized Residual
2
1
0
-1
-2
-3
3000
3500
4000
4500
5000
5500
6000
etw

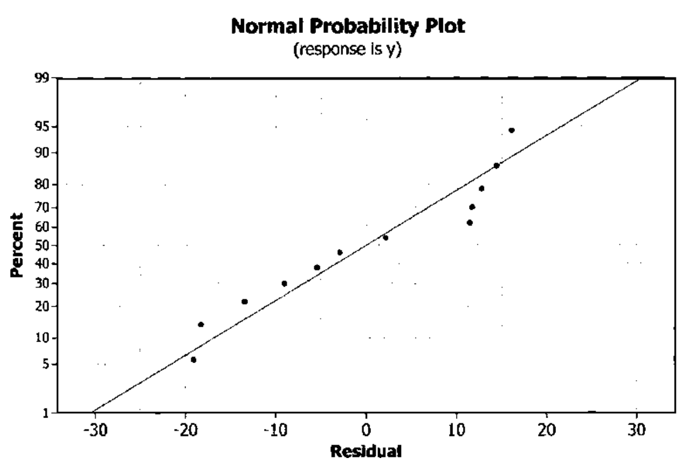

Residuals Versus cmp
(response is mpg)
Standardized Residual
2
1
0
-1
-2
-3
8.50
8.75
9.00
9.25
9.50
9.75
10.00
cmp

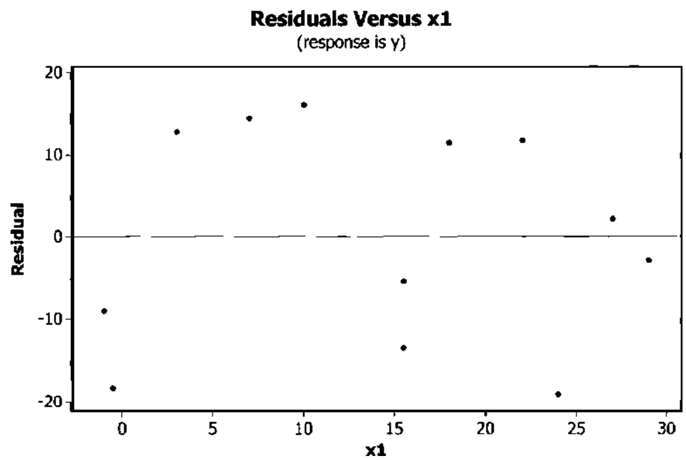


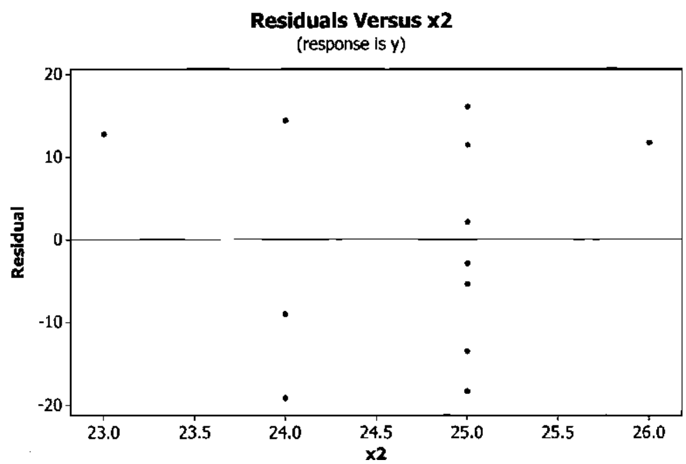


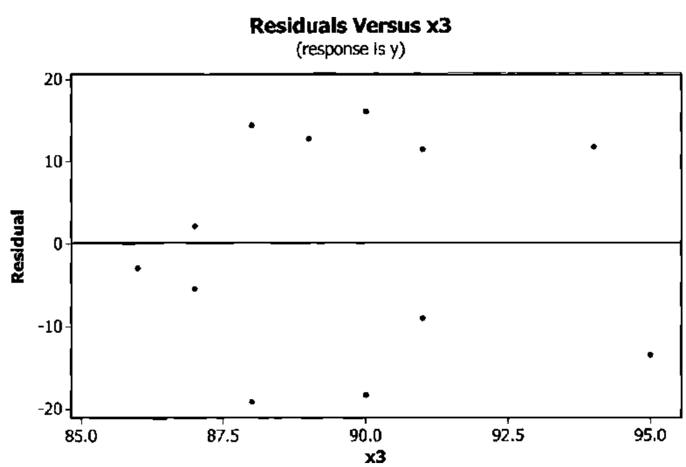
d)
0.036216, 0.000627, 0.041684, 0.008518, 0.026788, 0.040384, 0.003136, 0.196794, 0.267746, 0.000659, 0.075126, 0.000690, 0.041624, 0.070352, 0.008565, 0.051335, 0.001813, 0.019352, 0.000812, 0.098405, 0.574353
None of the values is greater than 1 so none of the observations are influential.

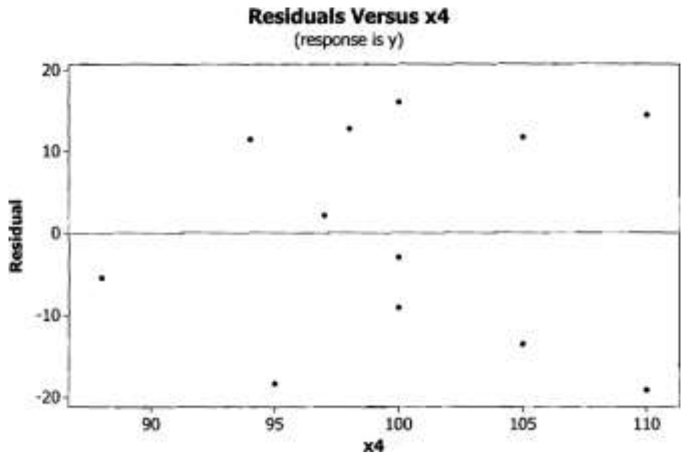
12-58 a) $R^2 = 0.71$
b) The residual plots look reasonable. There is some increase in variability at the middle of the predicted values.
c) Normality assumption is reasonable. The residual plots appear reasonable too.




Residuals Versus x1
(response is y)
Residual
x1
20
10
0
-10
-20
0
5
10
15
20
25
30


Residuals Versus x2
(response is y)
Residual
x2
20
10
0
-10
-20
23.0
23.5
24.0
24.5
25.0
25.5
26.0

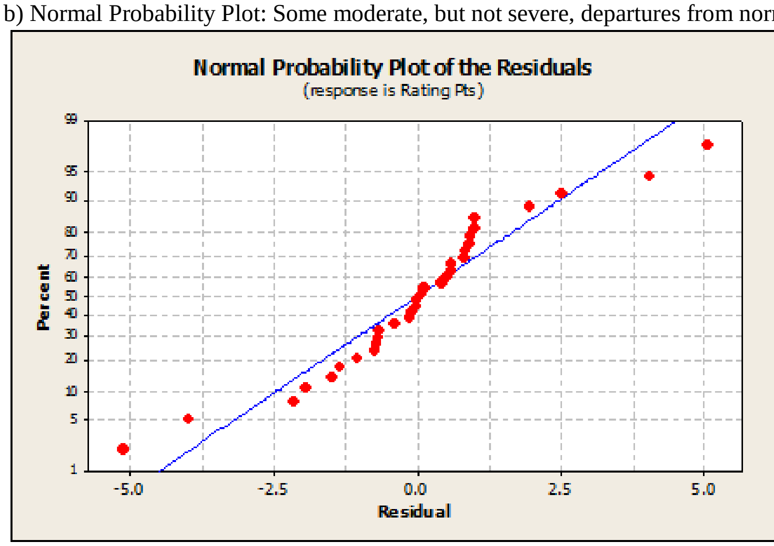

Residuals Versus x3
(response is y)
Residual
x3
20
10
0
-10
-20
85.0
87.5
90.0
92.5
95.0



12-59 a) The computer output follows. The proportion of total variability explained by this model is:

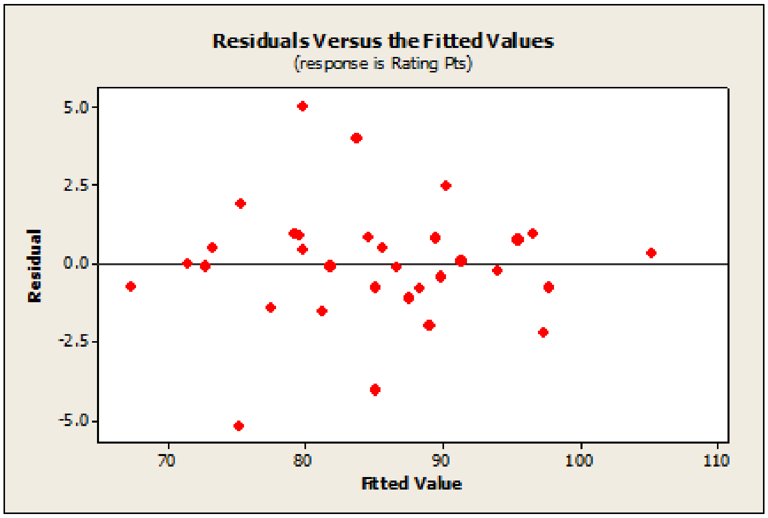
$$R^2 = \frac{SS_R}{SS_T} = \frac{2373.59}{2489.52} = 0.95$$

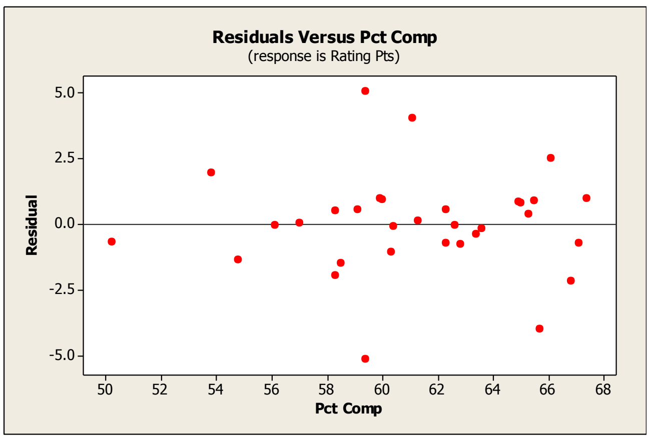
b) Normal Probability Plot: Some moderate, but not severe, departures from normality are indicated.

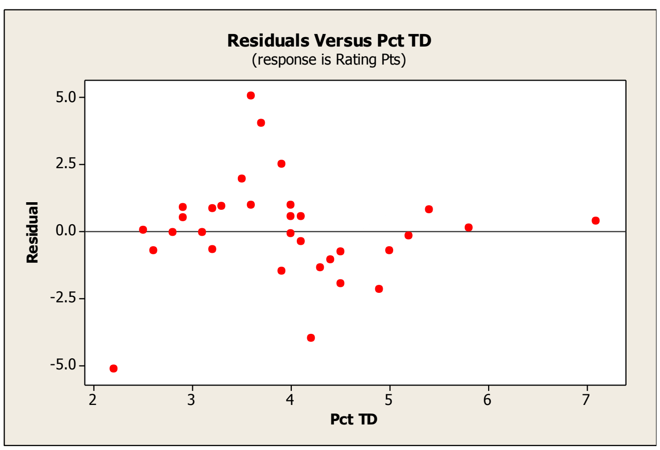


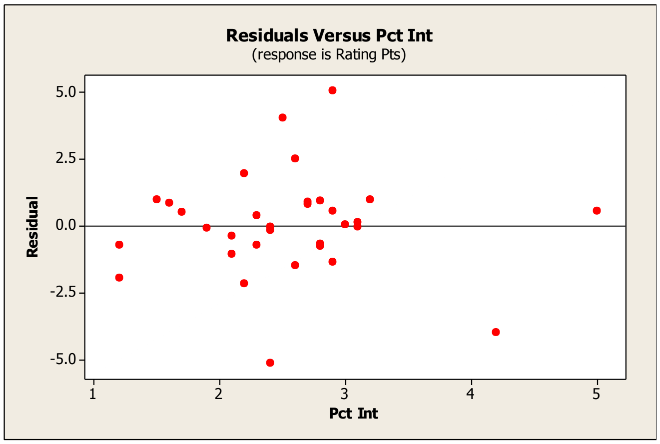
c) Plot the residuals versus fitted value and versus each regressor.

There is no obvious model failure in the plot of fitted values versus residuals. There is a modest increase in variability in the middle range of fitted values. The residual versus PctTD shows some non-random patterns. Possibly a non-linear term would benefit the model.

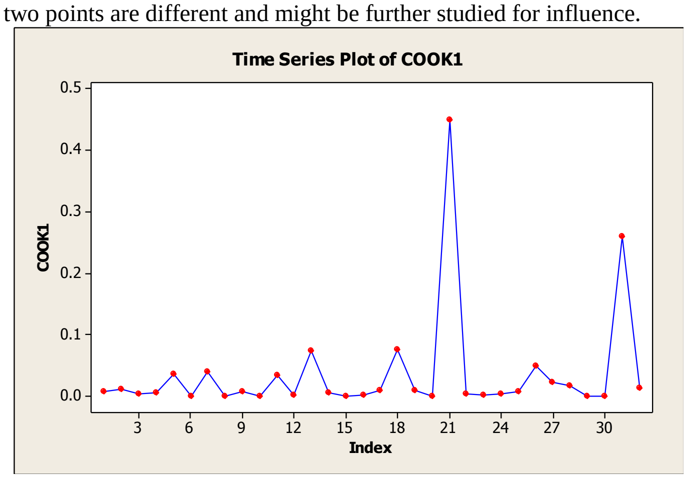

Residuals Versus the Fitted Values
(response is Rating Pts)
Residual
Fitted Value
5.0
2.5
0.0
-2.5
-5.0
70
80
90
100
110


Residuals Versus Pct Comp
(response is Rating Pts)
Residual
Pct Comp
5.0
2.5
0.0
-2.5
-5.0
50
52
54
56
58
60
62
64
66
68


Residuals Versus Pct TD
(response is Rating Pts)
Residual
Pct TD
5.0
2.5
0.0
-2.5
-5.0
2
3
4
5
6
7



d) A plot of Cook's distance measures follows. Although no points exceed the usual criterion of distance greater than 1, two points are different and might be further studied for influence.



Regression Analysis: Rating Pts versus Pct Comp, Pct TD, Pct Int

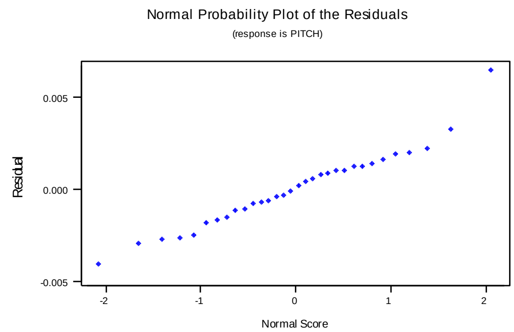
```
The regression equation is
Rating Pts = 2.99 + 1.20 Pct Comp + 4.60 Pct TD - 3.81 Pct Int

Predictor      Coef  SE Coef      T      P
Constant      2.986    5.877   0.51  0.615
Pct Comp    1.19857  0.09743  12.30  0.000
Pct TD       4.5956   0.3848  11.94  0.000
Pct Int     -3.8125   0.4861  -7.84  0.000

S = 2.03479   R-Sq = 95.3%   R-Sq(adj) = 94.8%

Analysis of Variance

Source          DF       SS      MS       F      P
Regression       3  2373.59  791.20  191.09  0.000
Residual Error  28   115.93    4.14
Total           31  2489.52

Source    DF   Seq SS
Pct Comp   1  1614.43
Pct TD     1   504.49
Pct Int    1   254.67

Unusual Observations
```

Obs	Pct Comp	Rating Pts	Fit	SE Fit	Residual	St Resid
11	61.1	87.700	83.691	0.371	4.009	2.00R
18	59.4	84.700	79.668	0.430	5.032	2.53R
21	65.7	81.000	85.020	1.028	-4.020	-2.29R
31	59.4	70.000	75.141	0.719	-5.141	-2.70R

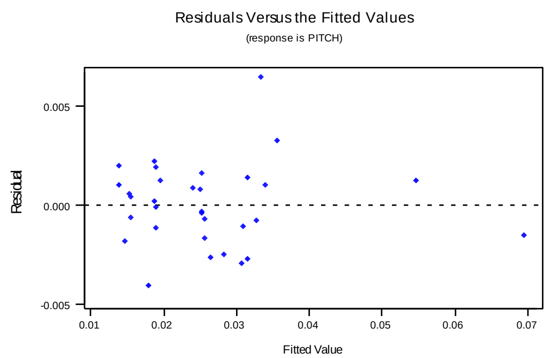
R denotes an observation with a large standardized residual.

12-60 a) $R^2 = 0.969$

b) Normality is acceptable



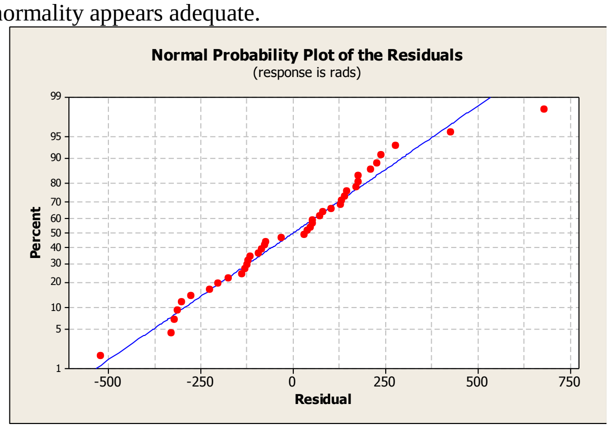
c) Plot is acceptable.



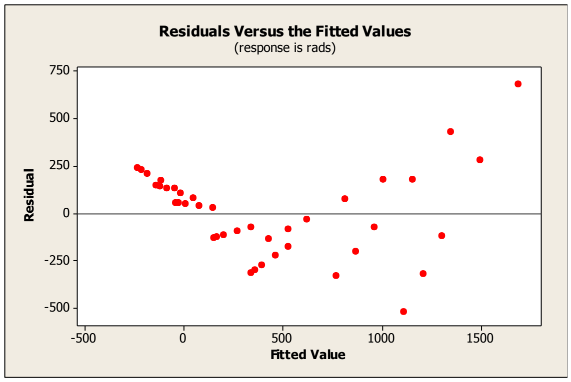
d) Cook's distance values
0.0191 0.0003 0.0026 0.0009 0.0293 0.1112 0.1014 0.0131 0.0076 0.0004 0.0109 0.0000 0.0140 0.0039 0.0002 0.0003 0.0079 0.0022 4.5975* 0.0033 0.0058 0.14120.0161 0.0268 0.0609 0.0016 0.0029 0.3391 0.3918 0.0134 0.0088 0.5063
The 19th observation is influential

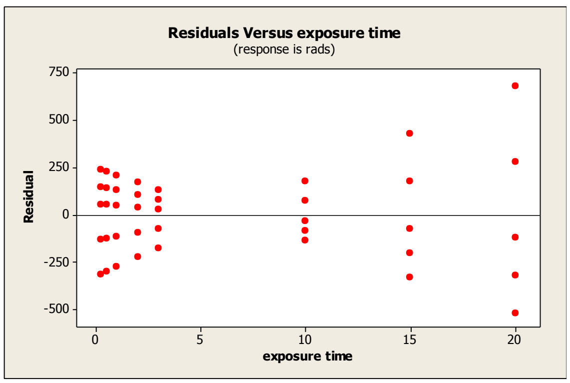
12-61 a) $R^2 = 84.3\%$

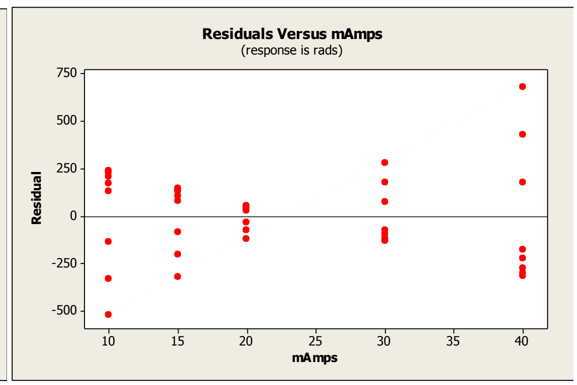
b) Assumption of normality appears adequate.



c) There are funnel shapes in the graphs, so the assumption of constant variance is violated. The model is inadequate.



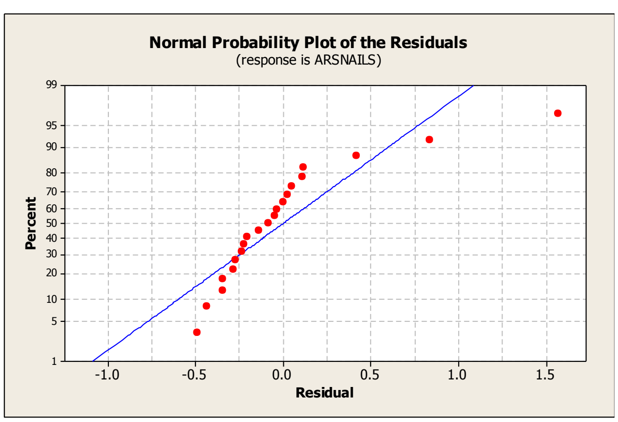




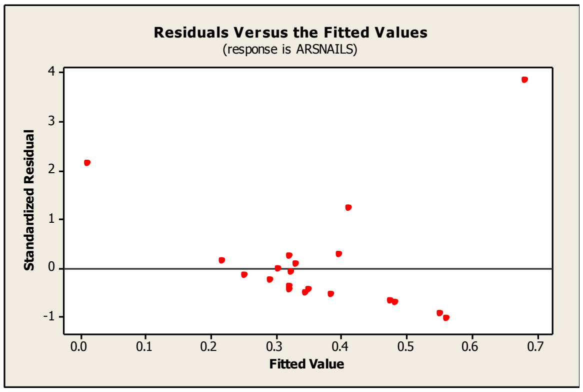
d) Cook's distance values
0.032728 0.029489 0.023724 0.014663 0.008279 0.008611
0.077299 0.3436 0.008489 0.007592 0.006018 0.003612
0.001985 0.002068 0.021386 0.105059 0.000926 0.000823
0.000643 0.000375 0.0002 0.000209 0.002467 0.013062
0.006095 0.005442 0.0043 0.002564 0.0014 0.001459
0.015557 0.077846 0.07828 0.070853 0.057512 0.036157
0.020725 0.021539 0.177299 0.731526
No, none of the observations has a Cook's distance greater than 1.

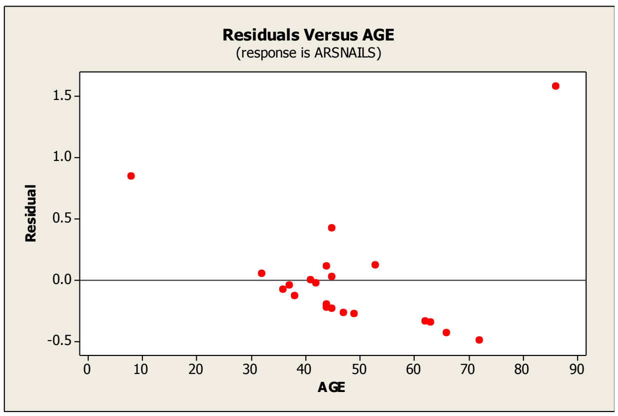
12-62 a) $R^2 = 8.1\%$

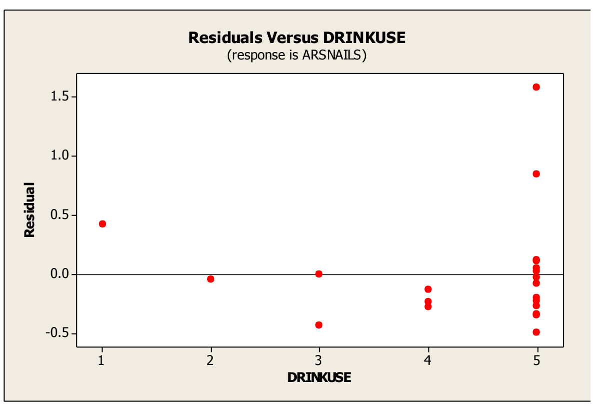
b) Assumption of normality is not adequate.

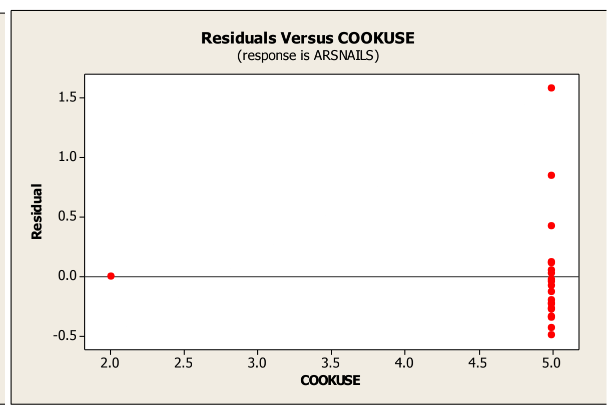


c) The graphs indicate non-constant variance. Therefore, the model is not adequate.









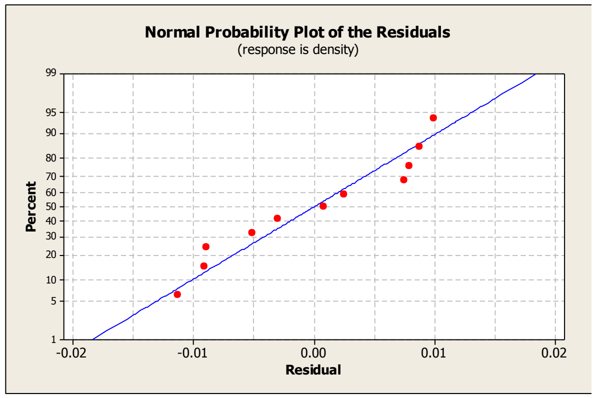
d) Cook's distance values

0.0032	0.0035	0.00386	0.05844	0.00139	0.00005	0.00524	0.00154	*infinity*
0.00496	0.05976	0.37409	0.00105	1.89094	0.68988	0.00035	0.00092	0.0155
0.00008	0.0143	0.00071						

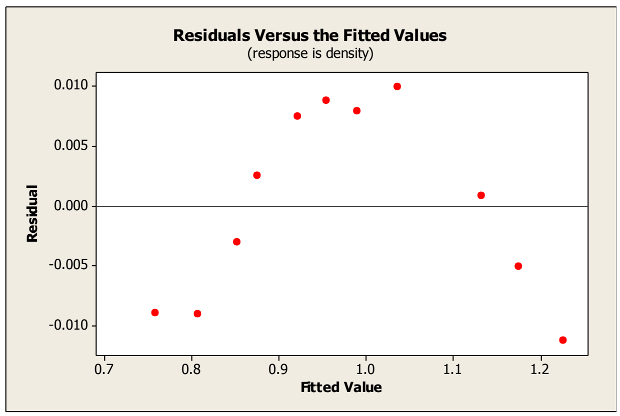
There are two influential points with Cook's distance greater than one. The entry *infinity* in the list above indicate a data point with $h_{ii} = 1$ and an undefined studentized residual.

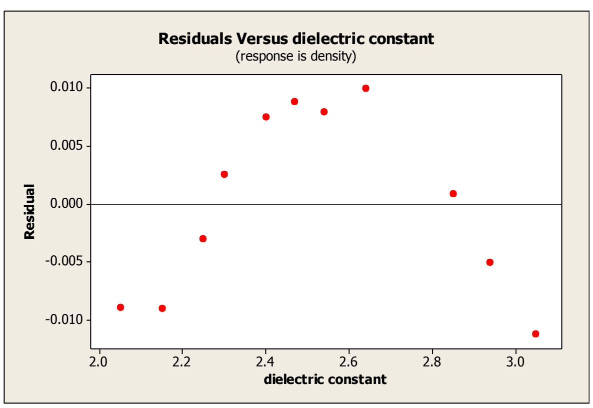
12-63 a) $R^2 = 99.7\%$

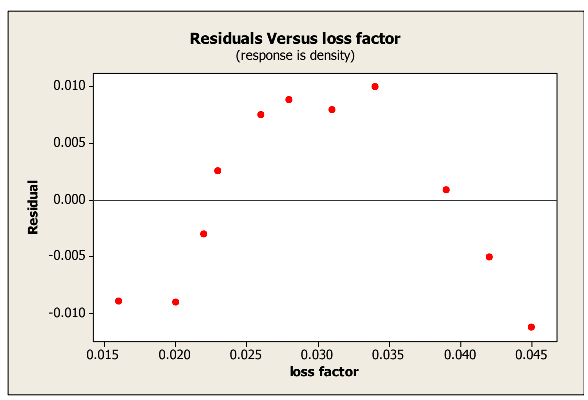
b) Assumption of normality appears adequate.



c) There is a non-constant variance shown in graphs. Therefore, the model is inadequate.



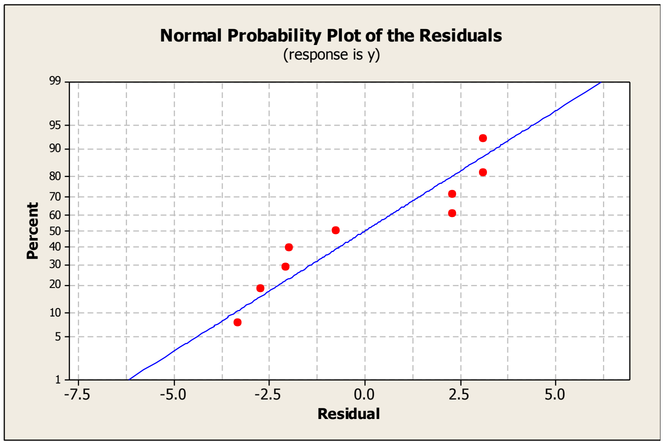




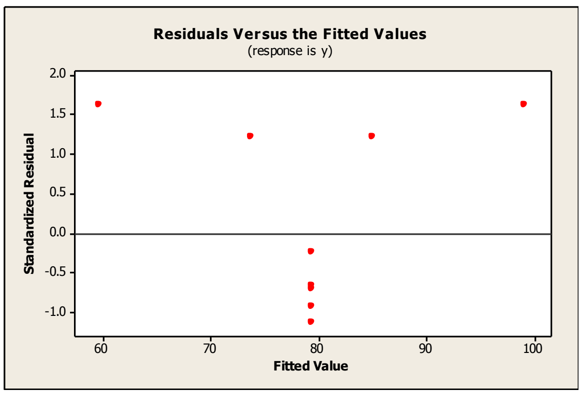
d) Cook's distance values
0.255007 0.692448 0.008618 0.011784 0.058551 0.077203
0.10971 0.287682 0.001337 0.054084 0.485253
No, none of the observations has a Cook's distance greater than 1.

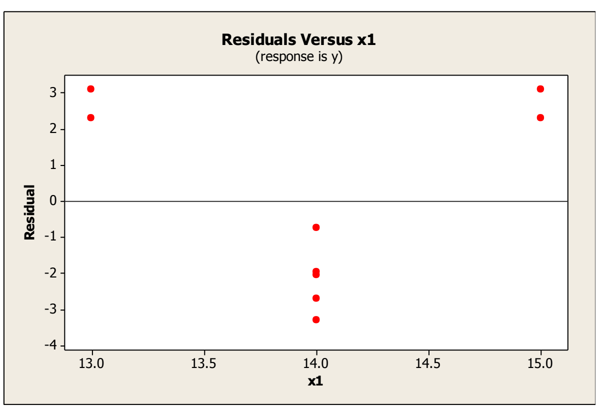
12-64 a) $R^2 = 93.7\,\%$

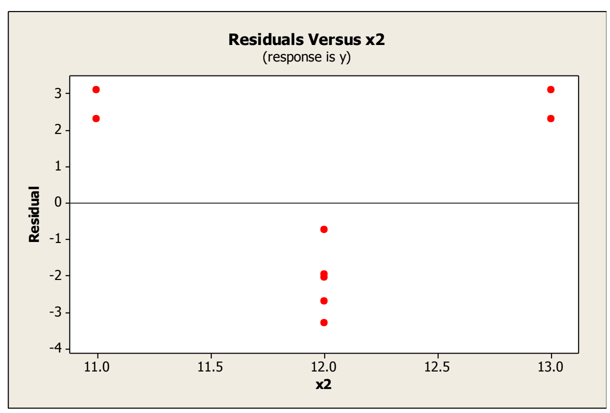
b) The normal assumption appears inadequate



c) The constant variance assumption is not invalid.



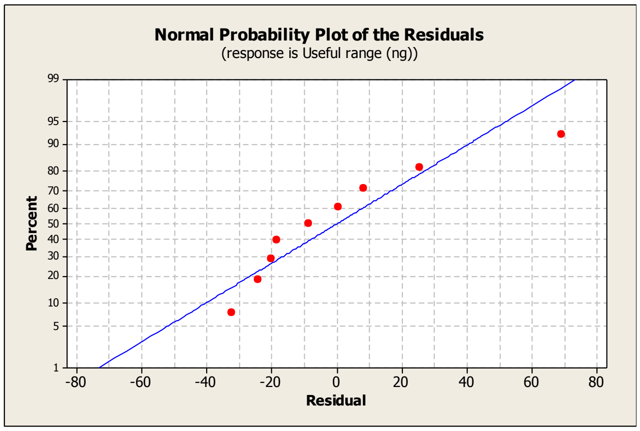




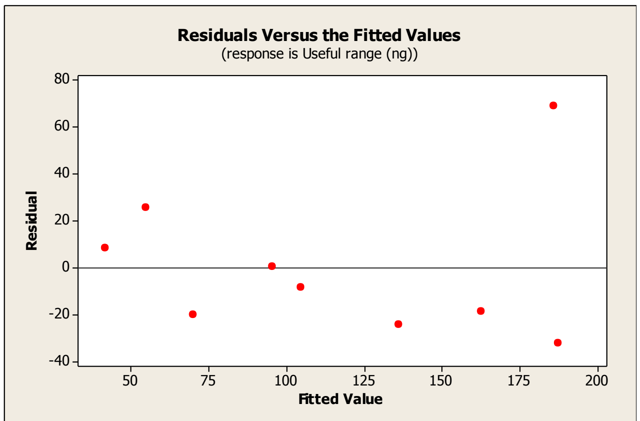
d) Cook's distance values
1.36736 0.7536 0.7536 1.36736 0.0542 0.01917 0.03646 0.02097 0.00282
There are two influential points with Cook's distances greater than 1.

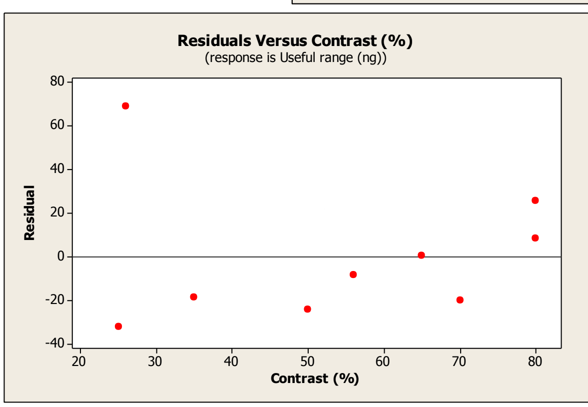
12-65 a) $R^2 = 75.6\%$

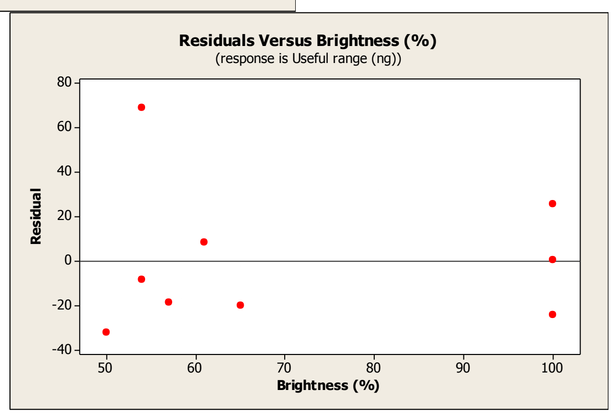
b) Assumption of normality appears adequate.



c) Assumption of constant variance is a possible concern. One point is a concern as a possible outlier.



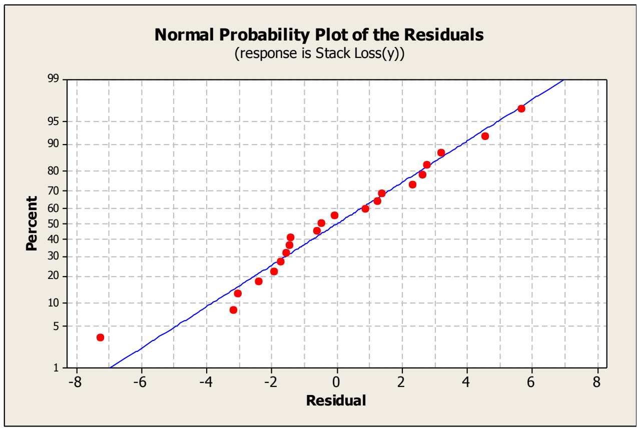




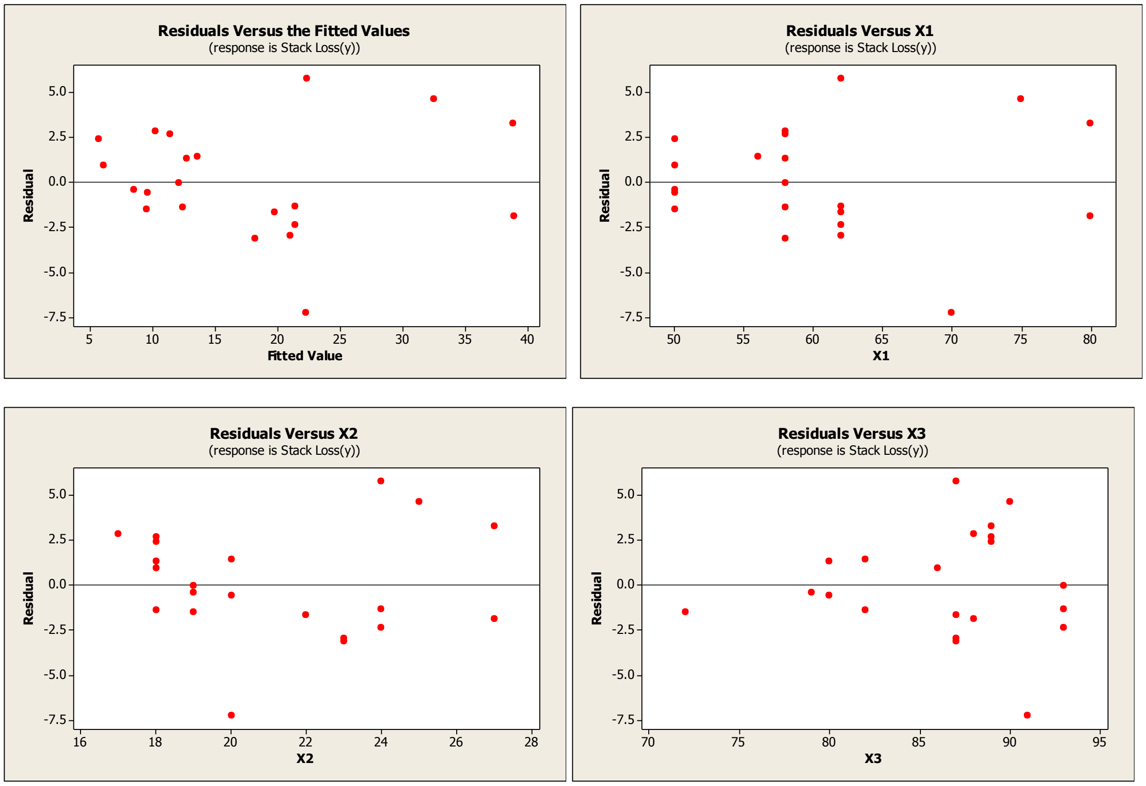
d) Cook's distance values
0.006827 0.032075 0.045342 0.213024 0.000075
0.154825 0.220637 0.030276 0.859916
No, none of the observations has a Cook's distance greater than 1.

12-66 a) $R^2 = 91.4\%$

b) Assumption of normality appears adequate.



c) Assumption of constant variance appears reasonable



d) Cook's distance values
0.15371 0.059683 0.126414 0.130542 0.004048 0.019565
0.048802 0.016502 0.044556 0.01193 0.035866 0.065066
0.010765 0.00002 0.038516 0.003379 0.065473 0.001122
0.002179 0.004492 0.692
No, none of the observations has a Cook's distance greater than 1.

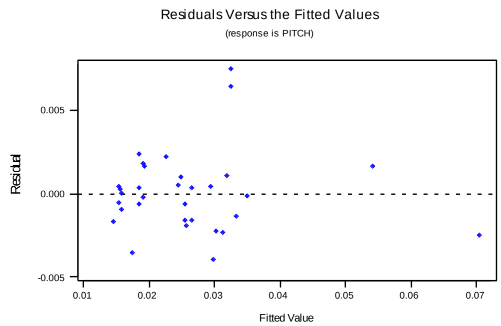
12-67 a) $R^2 = 0.9835$

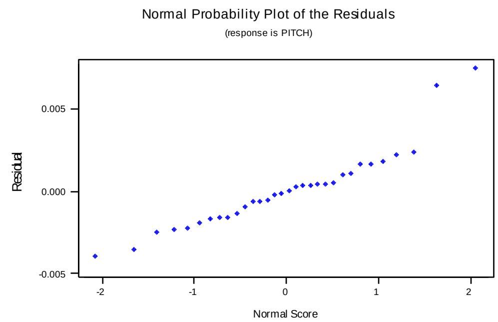
b) $R^2 = 0.992$

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

12-68 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.

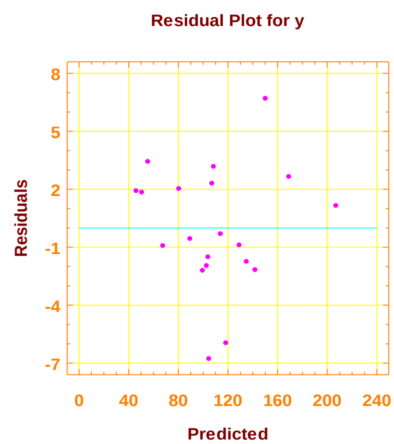
b) Normality is acceptable, but there is some indication of outliers.





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

12-69 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%;

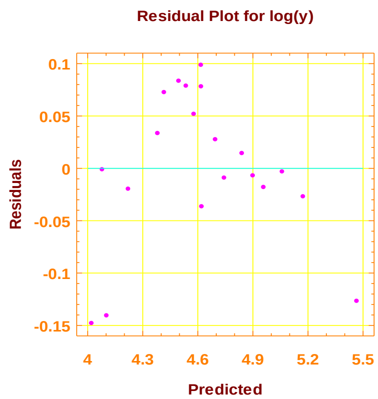
c)

```
                          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.
```

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

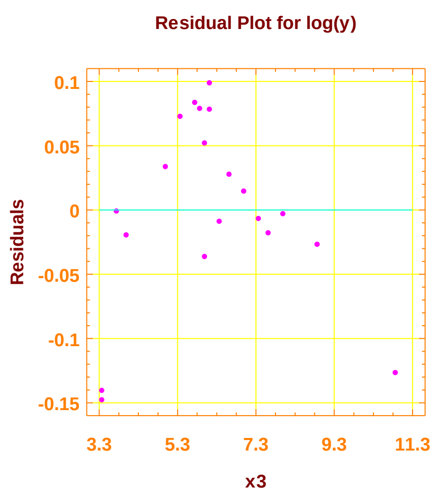
d)



Plot exhibits curvature

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

e)



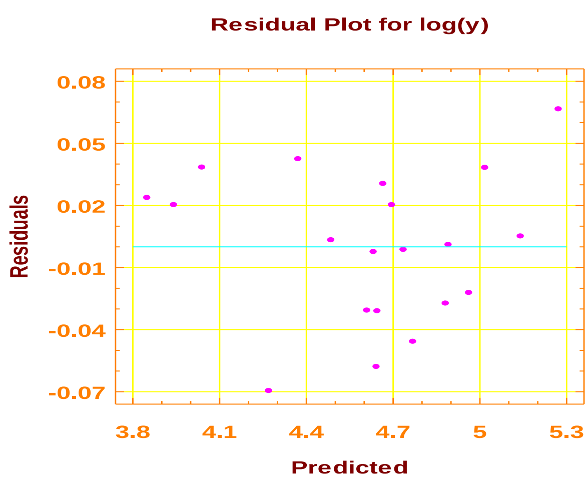
Plot exhibits curvature

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.

12-70 a)

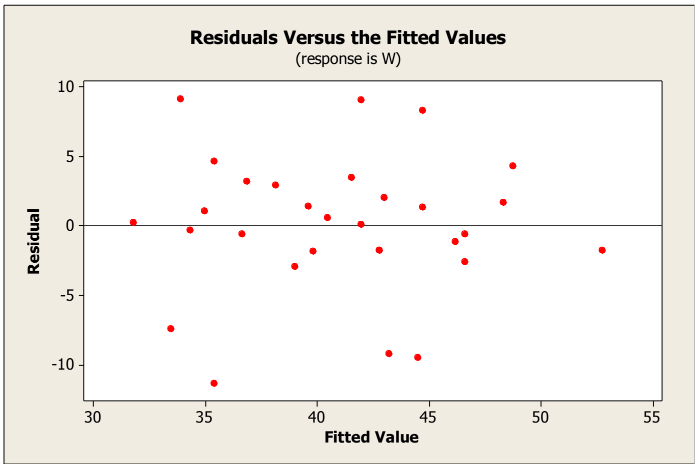
Regression Analysis: W versus GF

```
The regression equation is
W = - 8.57 + 0.212 GF
```

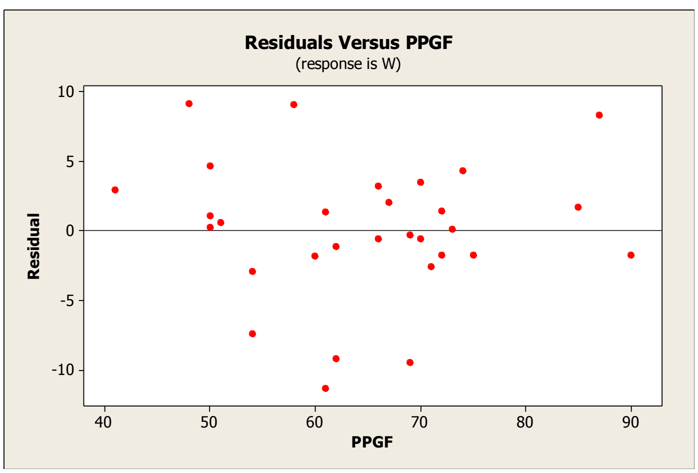
```
Predictor       Coef   SE Coef      T      P
Constant      -8.574     8.910  -0.96  0.344
GF           0.21228   0.03795   5.59  0.000

S = 5.02329   R-Sq = 52.8%   R-Sq(adj) = 51.1%

Analysis of Variance

Source          DF       SS      MS      F      P
Regression       1   789.46  789.46  31.29  0.000
Residual Error  28   706.54   25.23
Total           29  1496.00
```

b) `R-Sq = 52.8%`

c) Model appears adequate.



d) No, the residuals do not seem to be related to PPGF. Because there is no pattern evident in the plot, it does not seem that this variable would contribute significantly to the model.



12-71 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

12-72 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

$\alpha = 0.1$

$$f_0 = 17.20$$
$$f_{0.1,2,5} = 3.78$$
$$f_0 > f_{0.1,2,5}$$

Reject H_0 and conclude that model is significant at $\alpha = 0.1$

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

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

$\alpha = 0.1$

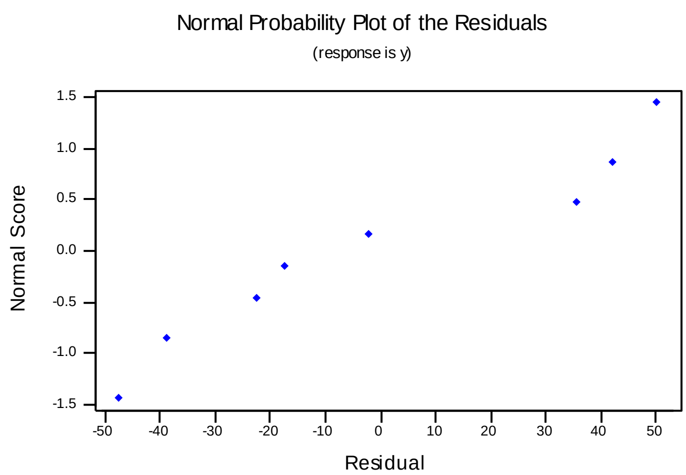
$$t_0 = -2.45$$
$$t_{\alpha/2,n-p} = t_{0.05,8-3} = t_{0.05,5} = 2.015$$
$$|t_0| > 2.015$$

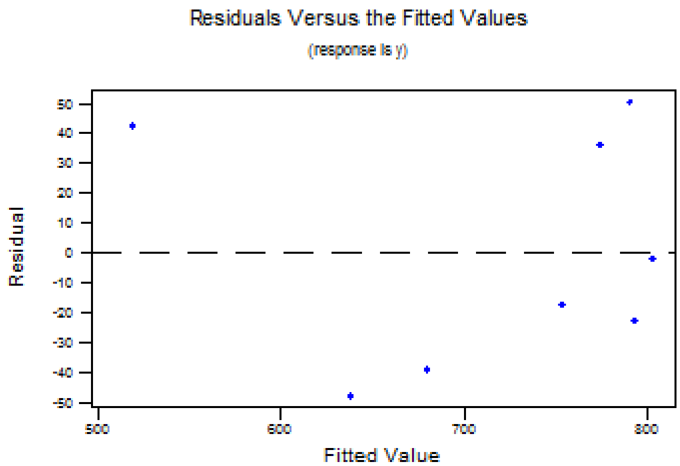
Reject H_0 and conclude sufficient evidence to support value of quadratic term in model at $\alpha = 0.1$.

d) One residual is an outlier

Normality assumption appears acceptable

Residuals against fitted values is somewhat curved, but the impact of the outlier should be considered.





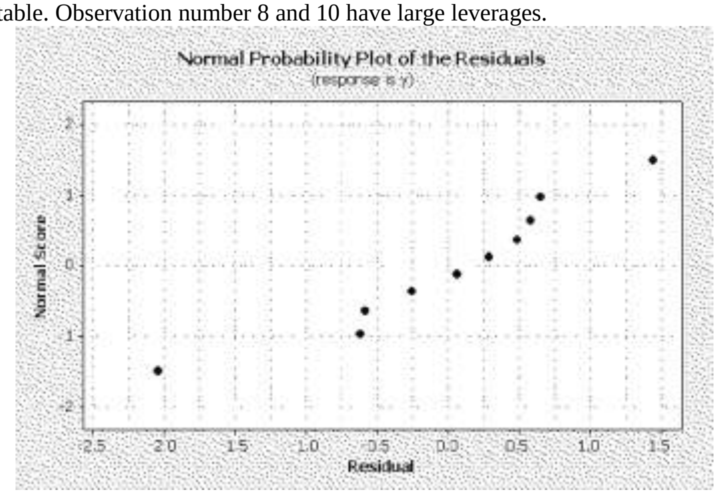
12-73 a) $\hat{y} = 0.97 - 3.75x + 410x^2$

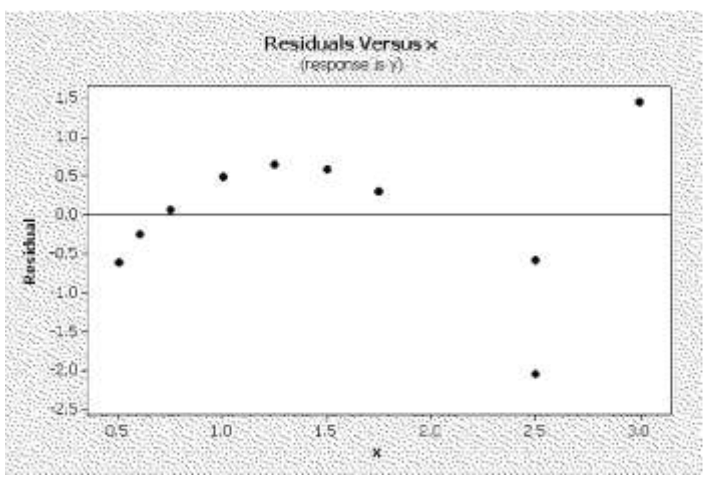
b) $f_0 = 16.86$, $f_0 > f_{0.05,2,7}$, reject H_0 and conclude regression model is significant at $\alpha = 0.05$

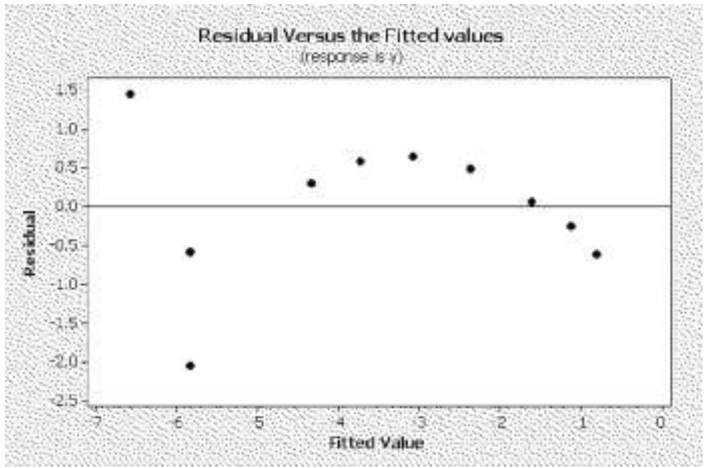
c) $|t_0| < t_{0.025,7}$

$t_0 = 0.66$, fail to reject H_0 and the regression model is not significant at $\alpha = 0.05$

d) Model is not acceptable. Observation number 8 and 10 have large leverages.







12-74 a) $\hat{y} = -10 + 3.64x + 1.25x^2$

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

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

$\alpha = 0.05$

$f_0 = 1010.16$

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

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

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

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

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

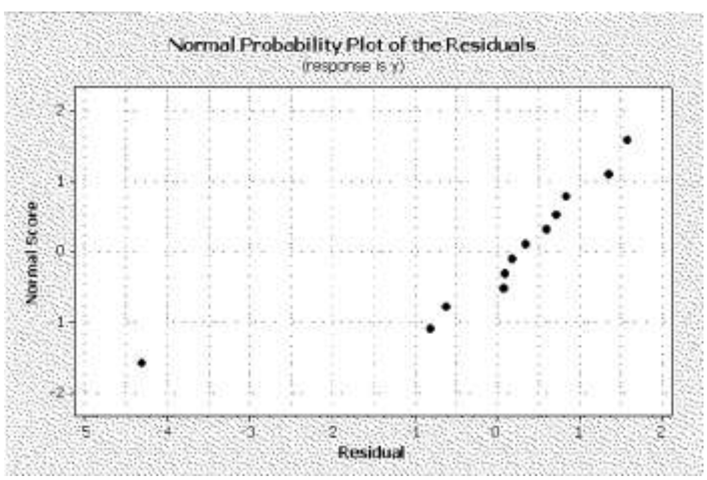
$t_0 = 2.55$

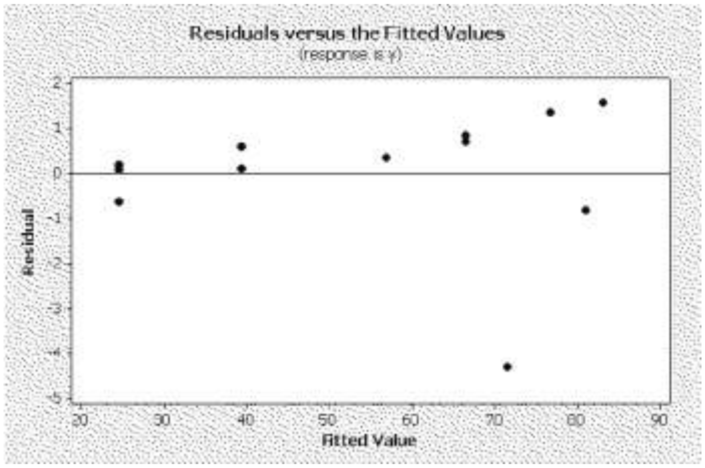
$t_{.025,9} = 2.262$

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

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

d) Observation number 9 is an extreme outlier.





e) $\hat{y} = -80.2 + 43x - 5.92x^2 + 0.425x^3$

$H_0 : \beta_{33} = 0$

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

$t_0 = 0.76$

$t_{.025,8} = 2.306$

$|t_0| \not> t_{0.025,8}$

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

12-75 a)

```
Predictor       Coef      SE Coef        T        P
Constant      -1.769        1.287    -1.37    0.188
x1            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 $\beta_j \neq 0$

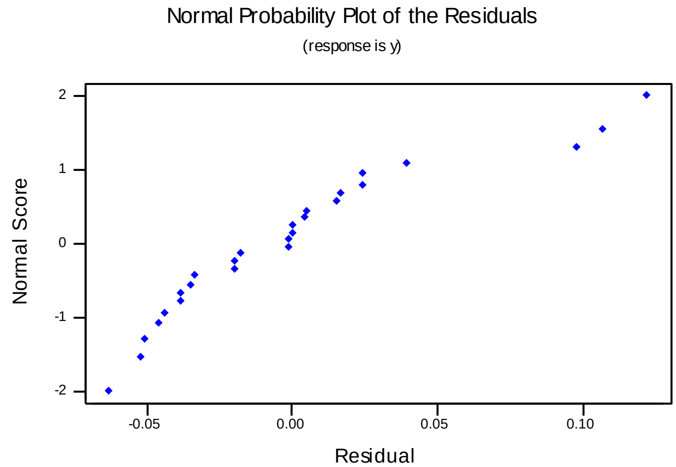
$f_0 = 19.628$

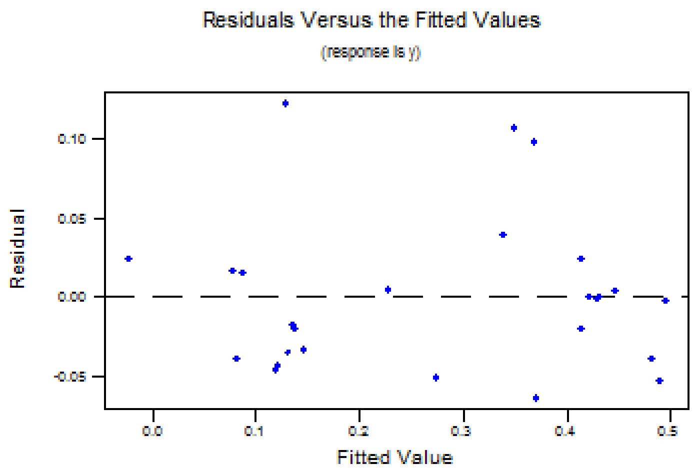
$f_{.01,9,16} = 3.78$

$f_0 > f_{0.01,9,16}$

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

c) Assumptions appear to be reasonable.





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

H_1 at least one $\beta \neq 0$

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

$$f_{.01,6,16} = 4.20$$

$$f_0 \not> f_{.01,6,16}$$

Fail to 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: $y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3$

12-76 a) Create an indicator variable for sex (e.g. 0 for male, 1 for female) and include this variable in the model.

b)
```
The regression equation is
ARSNAILS = - 0.214 - 0.008 DRINKUSE + 0.028 COOKUSE + 0.00794 AGE + 0.167 SEXID

Predictor       Coef    SE Coef       T       P
Constant     -0.2139     0.9708   -0.22   0.828
DRINKUSE     -0.0081     0.1050   -0.08   0.940
COOKUSE       0.0276     0.1844    0.15   0.883
AGE         0.007937   0.007251    1.09   0.290
SEXID         0.1675     0.2398    0.70   0.495

S = 0.514000   R-Sq = 10.8%   R-Sq(adj) = 0.0%

where SEXID = 0 for male and 1 for female
```

c) Because the P-value for testing $H_0: \beta_{sex} = 0$ against $H_1: \beta_{sex} \neq 0$ is 0.495, there is no evidence that the person's sex affects arsenic in the nails.

12-77 a) Use indicator variable for transmission type.
There are three possible transmission types: L4, L5 and M6. So, two indicator variables could be used where x_3=1 if trns=L5, 0 otherwise and x_4=1 if trns=M6, 0 otherwise.

b) $\hat{y} = 56.677 - 0.1457x_1 - 0.00525x_2 - 0.138x_3 - 4.179x_4$

c) The P-value for testing $H_0: \beta_3 = 0$ is 0.919, which is not significant. However, the P-value for testing $H_0: \beta_4 = 0$ is 0.02, which is significant for values of $\alpha > 0.02$. Thus, it appears that whether or not the transmission is manual affects mpg, but there is not a significant difference between the types of automatic transmission.

12-78 $y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_{12} x_{12}$

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

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

Test of different slopes:

$H_0: \beta_{12} = 0$

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

$t_0 = -1.79$

$t_{.025,16} = 2.12$

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

Fail to reject H_0. There is not sufficient evidence to conclude that two 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.

12-79 a) The min C_p model is: x_1, x_2

$$C_p = 3.0 \text{ and } MS_E = 55563.92$$
$$\hat{y} = -440.39 + 19.147x_1 + 68.080x_2$$

The min MS_E model is the same as the min C_p.

b) Same as the model in part (a).
c) Same as the model in part (a).
d) Same as the model in part (a).
e) All methods give the same model with either min C_p or min MS_E.

12-80 The default settings for F-to-enter and F-to-remove, equal to 4, in the computer software were used. Different settings can change the models generated by the method.

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

$C_p = 3.8$ $MS_E = 134.6$

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

The min C_p model is: x_1, x_2,

$C_p = 3.4$ $MS_E = 145.7$

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

b) Same as the min C_p model in part (a)
c) Same as part min MS_E model in part (a)
d) Same as part min C_p model 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.

12-81 a) The min C_p model is: x_1

$$C_p = 1.1 \text{ and } MS_E = 0.0000705$$
$$\hat{y} = -0.20052 + 0.467864x_1$$

The min MS_E model is the same as the min C_p.

b) Same as model in part (a).
c) Same as model in part (a).
d) Same as model in part (a).
e) All methods give the same model with either min C_p or min MS_E.

12-82 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 model is: x_1, x_3, x_4

$C_p = 2.6$ $MS_E = 0.6644$

$$\hat{y} = 2.419 + 0.5530x_1 + 0.4790x_3 - 0.12338x_4$$

The min C_p model is: x_3, x_4

$C_p = 1.6$ $MS_E = 0.7317$

$$\hat{y} = 4.656 + 0.5113x_3 - 0.12418x_4$$

b) Same as the min C_p model in part (a)
c) Same as the min C_p model in part (a)

d) Same as the min C_p model 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.

12-83 a) The min C_p model is: x_2

$$C_p = 1.2 \text{ and } MS_E = 1178.55$$

$$\hat{y} = 253.06 - 2.5453x_2$$

The min MS_E model is the same as the min C_p.
b) Same as model in part (a).
c) Same as model in part (a).
d) Same as model in part (a).
e) All methods give the same model with either min C_p or min MS_E.

12-84 a) The min C_p model is: x_1 ,x_2

$$C_p = 3.0 \text{ and } MS_E = 9.4759$$

$$\hat{y} = -171 + 7.029x_1 + 12.696x_2$$

The min MS_E model is the same as the min C_p.
b) Same as model in part (a).
c) Same as model in part (a).
d) Same as model in part (a).
e) All methods give the same model with either min C_p or min MS_E.

12-85 a) The min C_p model is: x_1, x_2

$$C_p = 2.9 \text{ and } MS_E = 10.49$$

$$\hat{y} = -50.4 + 0.671x_1 + 1.30x_2$$

The min MS_E model is the same as the min C_p.
b) Same as model in part (a).
c) Same as model in part (a).
d) Same as model in part (a).
e) All methods give the same model with either min C_p or min MS_E.
f) There are no observations with a Cook's distance greater than 1 so the results will be the same.

12-86 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)

Best Subsets Regression: W versus GF, GA, ...

Response is W

Vars	R-Sq	R-Sq(adj)	Mallows C-p	S	GF	GA	ADV	PPGF	PCTG	PEN	BMI	AVG	SHT	PPGA	PKPCT	SHGF	SHGA	FG
1	52.8	51.1	74.3	5.0233	X													
1	49.3	47.5	81.6	5.2043														X
2	86.5	85.5	4.7	2.7378	X	X												
2	79.3	77.8	19.9	3.3832	X									X				
3	87.7	86.2	4.2	2.6648	X	X											X	
3	87.3	85.8	5.0	2.7045	X	X							X					
4	88.7	86.9	4.0	2.6011	X	X							X				X	
4	88.2	86.3	5.0	2.6561	X	X						X					X	
5	89.5	87.3	4.4	2.5620	X	X							X	X	X			
5	89.3	87.0	4.8	2.5866	X	X								X	X		X	
6	91.3	89.1	2.4	2.3728	X	X							X	X	X		X	

```
  6  89.9       87.3       5.4  2.5621  X X X             X X   X
  7  92.3       89.8       2.4  2.2894  X X X           X X X   X
  7  91.6       89.0       3.8  2.3874  X X       X     X X X   X
  8  92.6       89.8       3.7  2.2954  X X X     X     X X X   X
  8  92.6       89.8       3.7  2.2967  X X X         X X X X   X
  9  92.7       89.5       5.4  2.3295  X X X     X     X X X X X
  9  92.7       89.5       5.4  2.3309  X X X         X X X X X X
 10  92.8       89.0       7.3  2.3829  X X X X   X     X X X X X
 10  92.8       89.0       7.3  2.3833  X X X   X X     X X X X X
 11  92.8       88.5       9.2  2.4402  X X X X   X   X X X X X X
 11  92.8       88.5       9.2  2.4406  X X X   X X   X X X X X X
 12  92.9       87.9      11.0  2.4936  X X X X   X X X X X X X X
 12  92.9       87.9      11.0  2.4939  X X X   X X X X X X X X X
 13  92.9       87.2      13.0  2.5702  X X X X X X X X X X X X X
 13  92.9       87.2      13.0  2.5703  X X X X   X X X X X X X X X
 14  92.9       86.3      15.0  2.6544  X X X X X X X X X X X X X X
```

From the output the minimum CP model and minimum MSE model are the same.
The regressors are GF, GA, ADV, SHT, PPGA, PKPCT, SHGA. The computer output for this model follows.

Regression Analysis: W versus GF, GA, ADV, SHT, PPGA, PKPCT, SHGA

```
The regression equation is
W = 457 + 0.182 GF - 0.187 GA - 0.0375 ADV + 0.256 SHT - 1.44 PPGA - 4.94 PKPCT
    + 0.489 SHGA

Predictor       Coef  SE Coef      T      P
Constant       457.3    138.5   3.30  0.003
GF           0.18233  0.02018   9.04  0.000
GA          -0.18657  0.03317  -5.62  0.000
ADV         -0.03753  0.02282  -1.65  0.114
SHT          0.25638  0.09826   2.61  0.016
PPGA         -1.4420   0.4986  -2.89  0.008
PKPCT         -4.935    1.679  -2.94  0.008
SHGA          0.4893   0.1785   2.74  0.012

S = 2.28935   R-Sq = 92.3%   R-Sq(adj) = 89.8%

Analysis of Variance

Source          DF       SS      MS      F      P
Regression       7  1380.70  197.24  37.63  0.000
Residual Error  22   115.30    5.24
Total           29  1496.00
```

The model is

$$\hat{y} = 457 + 0.182GF - 0.187GA - 0.0375ADV + 0.256SHT - 1.44PPGA - 4.94PKPCT + 0.489SHGA$$

b)

Stepwise Regression: W versus GF, GA, ...

```
  Alpha-to-Enter: 0.15  Alpha-to-Remove: 0.15

Response is W on 14 predictors, with N = 30

Step              1        2        3        4
Constant     -8.574   40.271   38.311   43.164

GF            0.212    0.182    0.182    0.187
T-Value        5.59     8.68     8.92     9.26
```

```
P-Value       0.000   0.000   0.000   0.000

GA                   -0.179  -0.179  -0.167
T-Value               -8.20   -8.40   -7.51
P-Value               0.000   0.000   0.000

SHGA                           0.27    0.29
T-Value                        1.58    1.76
P-Value                       0.126   0.090

SHT                                  -0.026
T-Value                               -1.51
P-Value                               0.143

S              5.02    2.74    2.66    2.60
R-Sq          52.77   86.47   87.66   88.69
R-Sq(adj)     51.08   85.47   86.23   86.88
Mallows C-p    74.3     4.7     4.2     4.0
```

The selected model from Stepwise Regression has four regressors GF, GA, SHT, SHGA. The computer output for this model follows.

Regression Analysis: W versus GF, GA, SHT, SHGA

```
The regression equation is
W = 43.2 + 0.187 GF - 0.167 GA - 0.0259 SHT + 0.293 SHGA

Predictor      Coef  SE Coef      T      P
Constant     43.164    8.066   5.35  0.000
GF          0.18677  0.02016   9.26  0.000
GA         -0.16683  0.02221  -7.51  0.000
SHT        -0.02587  0.01710  -1.51  0.143
SHGA         0.2926   0.1660   1.76  0.090

S = 2.60115   R-Sq = 88.7%   R-Sq(adj) = 86.9%

Analysis of Variance

Source          DF       SS      MS      F      P
Regression       4  1326.85  331.71  49.03  0.000
Residual Error  25   169.15    6.77
Total           29  1496.00
```

The model is

$$\hat{y} = 43.2 + 0.187GF - 0.167GA - 0.0259SHT + 0.293SHGA$$

c)

Stepwise Regression: W versus GF, GA, ...

```
Forward selection.  Alpha-to-Enter: 0.25

Response is W on 14 predictors, with N = 30

Step              1       2       3       4
Constant     -8.574  40.271  38.311  43.164

GF            0.212   0.182   0.182   0.187
T-Value        5.59    8.68    8.92    9.26
P-Value       0.000   0.000   0.000   0.000

GA                   -0.179  -0.179  -0.167
T-Value               -8.20   -8.40   -7.51
```

```
P-Value              0.000   0.000   0.000

SHGA                         0.27    0.29
T-Value                      1.58    1.76
P-Value                      0.126   0.090

SHT                                 -0.026
T-Value                              -1.51
P-Value                              0.143

S              5.02    2.74    2.66    2.60
R-Sq          52.77   86.47   87.66   88.69
R-Sq(adj)     51.08   85.47   86.23   86.88
Mallows C-p    74.3     4.7     4.2     4.0
```

The model selected by Forward Selection is the same as part (b).

d)

Stepwise Regression: W versus GF, GA, ...

```
Backward elimination.  Alpha-to-Remove: 0.1

Response is W on 14 predictors, with N = 30

Step              1        2        3        4        5        6
Constant      512.2    513.0    511.3    524.0    520.5    507.3

GF            0.164    0.164    0.164    0.166    0.167    0.173
T-Value        4.46     4.97     5.24     5.53     5.66     7.65
P-Value       0.000    0.000    0.000    0.000    0.000    0.000

GA           -0.183   -0.184   -0.184   -0.186   -0.190   -0.191
T-Value       -3.83    -4.58    -4.74    -4.95    -5.43    -5.62
P-Value       0.002    0.000    0.000    0.000    0.000    0.000

ADV          -0.054   -0.054   -0.046   -0.043   -0.040   -0.036
T-Value       -0.25    -0.25    -1.52    -1.48    -1.48    -1.52
P-Value       0.808    0.802    0.147    0.157    0.156    0.145

PPGF          0.089    0.087    0.047    0.031    0.022
T-Value        0.08     0.08     0.59     0.43     0.34
P-Value       0.938    0.937    0.565    0.671    0.739

PCTG           -0.1     -0.1
T-Value       -0.04    -0.04
P-Value       0.971    0.971

PEN         -0.1632  -0.1628  -0.1646  -0.0365  -0.0039  -0.0043
T-Value       -0.54    -0.56    -0.59    -0.38    -0.86    -1.00
P-Value       0.598    0.586    0.564    0.706    0.400    0.330

BMI           -0.13    -0.13    -0.13
T-Value       -0.45    -0.47    -0.49
P-Value       0.658    0.647    0.632

AVG            13.1     13.1     13.2      2.7
T-Value        0.53     0.54     0.57     0.34
P-Value       0.606    0.594    0.574    0.735

SHT            0.29     0.29     0.29     0.31     0.31     0.30
T-Value        2.19     2.30     2.40     2.65     2.71     2.76
P-Value       0.045    0.035    0.028    0.016    0.014    0.012
```

```
PPGA             -1.60    -1.60    -1.61    -1.66    -1.64    -1.60
T-Value          -2.50    -2.61    -2.71    -2.90    -2.96    -3.02
P-Value          0.025    0.019    0.015    0.009    0.008    0.007

PKPCT             -5.5     -5.6     -5.6     -5.7     -5.7     -5.5
T-Value          -2.54    -2.66    -2.77    -2.96    -3.02    -3.09
P-Value          0.023    0.017    0.013    0.008    0.007    0.006

SHGF              0.11     0.11     0.11     0.09     0.09     0.10
T-Value           0.54     0.56     0.62     0.54     0.59     0.62
P-Value          0.600    0.584    0.541    0.593    0.565    0.540

SHGA              0.61     0.61     0.61     0.57     0.54     0.53
T-Value           2.34     2.42     2.59     2.64     2.78     2.85
P-Value          0.033    0.028    0.019    0.016    0.012    0.010

FG                0.00
T-Value           0.02
P-Value          0.981

S                 2.65     2.57     2.49     2.44     2.38     2.33
R-Sq             92.94    92.93    92.93    92.84    92.79    92.75
R-Sq(adj)        86.34    87.19    87.95    88.46    88.99    89.48
Mallows C-p       15.0     13.0     11.0      9.2      7.3      5.4


Step                 7        8        9
Constant         496.5    457.3    417.5

GF               0.178    0.182    0.177
T-Value           8.53     9.04     8.57
P-Value          0.000    0.000    0.000

GA              -0.189   -0.187   -0.187
T-Value          -5.66    -5.62    -5.43
P-Value          0.000    0.000    0.000

ADV             -0.038   -0.038
T-Value          -1.67    -1.65
P-Value          0.109    0.114

PPGF
T-Value
P-Value

PCTG
T-Value
P-Value

PEN            -0.0039
T-Value          -0.94
P-Value          0.358

BMI
T-Value
P-Value

AVG
T-Value
P-Value

SHT              0.298    0.256    0.238
```

```
T-Value         2.76    2.61    2.35
P-Value        0.012   0.016   0.028

PPGA           -1.58   -1.44   -1.34
T-Value        -3.03   -2.89   -2.62
P-Value        0.006   0.008   0.015

PKPCT           -5.4    -4.9    -4.6
T-Value        -3.08   -2.94   -2.65
P-Value        0.006   0.008   0.014

SHGF
T-Value
P-Value

SHGA            0.51    0.49    0.39
T-Value         2.83    2.74    2.23
P-Value        0.010   0.012   0.036

FG
T-Value
P-Value

S               2.30    2.29    2.37
R-Sq           92.60   92.29   91.34
R-Sq(adj)      89.79   89.84   89.09
Mallows C-p      3.7     2.4     2.4
```

The model selected by Backward Selection includes GF, GA, SHT, PPGA, PKPCT, SHGA. The computer output for this model follows.

Regression Analysis: W versus GF, GA, SHT, PPGA, PKPCT, SHGA

```
The regression equation is
W = 418 + 0.177 GF - 0.187 GA + 0.238 SHT - 1.34 PPGA - 4.58 PKPCT + 0.387 SHGA

Predictor       Coef  SE Coef      T      P
Constant       417.5    141.3   2.95  0.007
GF           0.17679  0.02062   8.57  0.000
GA          -0.18677  0.03438  -5.43  0.000
SHT           0.2377   0.1012   2.35  0.028
PPGA         -1.3426   0.5130  -2.62  0.015
PKPCT         -4.578    1.725  -2.65  0.014
SHGA          0.3869   0.1734   2.23  0.036

S = 2.37276   R-Sq = 91.3%   R-Sq(adj) = 89.1%

Analysis of Variance

Source          DF       SS      MS      F      P
Regression       6  1366.51  227.75  40.45  0.000
Residual Error  23   129.49    5.63
Total           29  1496.00
```

The model is

$$\hat{y} = 418 + 0.177GF - 0.187GA + 0.238SHT - 1.34PPGA - 4.58PKPCT + 0.387SHGA$$

e) There are several reasonable choices.
The seven-variable model GF, GA, ADV, SHT, PPGA, PKPCT, SHGA with minimum C_p is a good choice. It has MSE not much larger than the MSE in the full model.

The four-variable model GF, GA, SHT, SHGA from Stepwise Regression(and Forward Selection) is a simpler model with Cp = 4.0 < p =5 and good R-squared.

Even the three-variable model GF, GA, SHGA is reasonable. It is still simpler with a good R-squared. The Cp = 4.2 and this is only slightly greater than p = 4. However, the MSE for this model is somewhat higher than for the six-variable model.

12-87 a) The computer output follows. The first model in the table with seven variables minimizes MS_E and Cp.

Best Subsets Regression: Pts versus Att, Comp, ...

```
PctComp, YdsPerAtt, PctTD, PctInt

Response is Pts

                                                        Y
                                                        d
                                                    P   s
                                                    c   p         P
                                                    t   e   P     c
                                                  C C   r   c     t
                                                A o o Y A   t L I I
                                                t m m d t T T n n n
Vars   R-Sq  R-Sq(adj)  Mallows C-p         S  t p p s t D D g t t
   1   67.2       66.1      38635.3    5.2187          X
   1   64.8       63.7      41381.3    5.4009      X
   2   85.1       84.1      17511.0    3.5748      X       X
   2   83.9       82.8      18943.5    3.7180          X         X
   3   95.3       94.8       5461.8    2.0348      X       X     X
   3   93.1       92.4       8064.5    2.4708      X     X     X
   4  100.0      100.0          5.3   0.14621      X   X   X     X
   4   98.5       98.3       1730.7    1.1713      X   X X     X
   5  100.0      100.0          6.1   0.14558      X X X   X     X
   5  100.0      100.0          6.2   0.14600    X X   X   X     X
   6  100.0      100.0          5.2   0.13990  X   X X X   X     X
   6  100.0      100.0          6.4   0.14362  X X X   X   X     X
   7  100.0      100.0          5.1   0.13632  X   X X X X X     X
   7  100.0      100.0          7.1   0.14277  X   X X X   X   X X
   8  100.0      100.0          7.0   0.13892  X   X X X X X   X X
   8  100.0      100.0          7.1   0.13924  X X X X X X X     X
   9  100.0      100.0          9.0   0.14203  X   X X X X X X X X
   9  100.0      100.0          9.0   0.14204  X X X X X X X   X X
  10  100.0      100.0         11.0   0.14537  X X X X X X X X X X
```

The computer output for this model follows.

Regression Analysis: RatingPts versus Att, PctComp, ...

```
The regression equation is
RatingPts = - 0.69 + 0.00738 Att + 0.827 PctComp - 0.00150 Yds + 4.82 YdsPerAtt
            + 0.0702 TD + 3.04 PctTD - 4.19 PctInt

Predictor        Coef    SE Coef        T      P
Constant       -0.693      1.264    -0.55  0.589
Att          0.007381   0.003258     2.27  0.033
PctComp      0.826817   0.008980    92.07  0.000
Yds        -0.0015027  0.0006166    -2.44  0.023
YdsPerAtt      4.8206     0.2555    18.87  0.000
TD            0.07025    0.04601     1.53  0.140
PctTD          3.0386     0.2013    15.10  0.000
PctInt       -4.19493    0.03545  -118.34  0.000

S = 0.136317   R-Sq = 100.0%   R-Sq(adj) = 100.0%
```

```
Analysis of Variance

Source          DF       SS      MS         F      P
Regression       7  2489.08  355.58  19135.65  0.000
Residual Error  24     0.45    0.02
Total           31  2489.52
```

b) Stepwise regression selects the four-variable model `YdsPerAtt,PctInt,PctTD,PctComp`.

Stepwise Regression: Pts versus Att, Comp, ...

```
  Alpha-to-Enter: 0.15  Alpha-to-Remove: 0.15

Response is Pts on 10 predictors, with N = 32

Step                1        2       3        4
Constant       14.195   24.924  32.751    2.128

YdsperAtt      10.092   10.296   7.432    4.214
T-Value          7.84    11.21    8.78    73.46
P-Value         0.000    0.000   0.000    0.000

PctInt                  -4.786  -4.932   -4.164
T-Value                  -5.49   -7.90  -118.11
P-Value                  0.000   0.000    0.000

PctTD                            3.187    3.310
T-Value                           5.36   101.17
P-Value                          0.000    0.000

PctComp                                  0.8284
T-Value                                   96.04
P-Value                                   0.000

S                5.22     3.72    2.66    0.146
R-Sq            67.18    83.90   92.06    99.98
R-Sq(adj)       66.09    82.79   91.21    99.97
Mallows C-p   38635.3  18943.5  9332.6      5.3
```

The computer output for this model follows.

Regression Analysis: RatingPts versus YdsPerAtt, PctInt, PctTD, PctComp

```
The regression equation is
RatingPts = 2.13 + 4.21 YdsPerAtt - 4.16 PctInt + 3.31 PctTD + 0.828 PctComp

Predictor       Coef   SE Coef        T      P
Constant      2.1277    0.4224     5.04  0.000
YdsPerAtt    4.21364   0.05736    73.46  0.000
PctInt      -4.16391   0.03526  -118.11  0.000
PctTD        3.31029   0.03272   101.17  0.000
PctComp     0.828385  0.008626    96.04  0.000

S = 0.146206   R-Sq = 100.0%   R-Sq(adj) = 100.0%

Analysis of Variance

Source          DF       SS      MS         F      P
Regression       4  2488.95  622.24  29108.69  0.000
Residual Error  27     0.58    0.02
Total           31  2489.52

The model is
```

$$\hat{y} = 2.13 + 4.21 YdsPerAtt - 4.16 PctInt + 3.31 PctTD + 0.828 PctComp$$

c) Forward selection shown below selects the same four-variable model `YdsPerAtt,PctInt,PctTD,PctComp` as in part (b).

Stepwise Regression: Pts versus Att, Comp, ...

```
Forward selection.  Alpha-to-Enter: 0.25

Response is Pts on 10 predictors, with N = 32

Step               1        2        3        4
Constant      14.195   24.924   32.751    2.128

YdsperAtt     10.092   10.296    7.432    4.214
T-Value         7.84    11.21     8.78    73.46
P-Value        0.000    0.000    0.000    0.000

PctInt                 -4.786   -4.932   -4.164
T-Value                 -5.49    -7.90  -118.11
P-Value                 0.000    0.000    0.000

PctTD                             3.187    3.310
T-Value                            5.36   101.17
P-Value                           0.000    0.000

PctComp                                   0.8284
T-Value                                    96.04
P-Value                                    0.000

S               5.22     3.72     2.66    0.146
R-Sq           67.18    83.90    92.06    99.98
R-Sq(adj)      66.09    82.79    91.21    99.97
Mallows C-p  38635.3  18943.5   9332.6      5.3
```

d) Backward elimination shown below selects the six-variable model
`Att, PctComp, Yds, YdsPerAtt, PctTD, PctInt.`
It is similar to the model with minimum MS_E except variable `TD` is excluded.

Stepwise Regression: Pts versus Att, Comp, ...

```
Backward elimination.  Alpha-to-Remove: 0.1


Response is Pts on 10 predictors, with N = 32


Step               1         2         3         4         5
Constant     -0.6871   -0.7140   -0.6751   -0.6928   -0.1704

Att           0.0074    0.0074    0.0074    0.0074    0.0055
T-Value         1.56      2.02      2.22      2.27      1.78
P-Value        0.133     0.056     0.036     0.033     0.088

Comp           0.000
T-Value         0.02
P-Value        0.982

PctComp       0.8253    0.8264    0.8264    0.8268    0.8289
T-Value        16.92     86.99     89.30     92.07     91.04
P-Value        0.000     0.000     0.000     0.000     0.000
```

```
Yds          -0.00159  -0.00158  -0.00156  -0.00150  -0.00083
T-Value         -1.65     -2.20     -2.39     -2.44     -1.88
P-Value         0.115     0.038     0.026     0.023     0.072

YdsperAtt        4.86      4.85      4.85      4.82      4.55
T-Value         11.78     16.28     17.69     18.87     24.46
P-Value         0.000     0.000     0.000     0.000     0.000

TD              0.075     0.075     0.074     0.070
T-Value          1.46      1.50      1.53      1.53
P-Value         0.158     0.148     0.139     0.140

PctTD           3.018     3.019     3.019     3.039     3.341
T-Value         13.47     13.82     14.15     15.10     96.41
P-Value         0.000     0.000     0.000     0.000     0.000

Lng            0.0001    0.0001
T-Value          0.05      0.05
P-Value         0.960     0.961

Int             0.011     0.011     0.010
T-Value          0.31      0.33      0.33
P-Value         0.756     0.747     0.743

PctInt         -4.241    -4.240    -4.238    -4.195    -4.177
T-Value        -27.73    -29.74    -31.63   -118.34   -121.82
P-Value         0.000     0.000     0.000     0.000     0.000

S               0.145     0.142     0.139     0.136     0.140
R-Sq            99.98     99.98     99.98     99.98     99.98
R-Sq(adj)       99.97     99.97     99.98     99.98     99.98
Mallows C-p      11.0       9.0       7.0       5.1       5.2
```

The computer output for this model follows.

Regression Analysis: RatingPts versus Att, PctComp, ...

```
The regression equation is
RatingPts = - 0.17 + 0.00550 Att + 0.829 PctComp - 0.000826 Yds + 4.55
YdsPerAtt
            + 3.34 PctTD - 4.18 PctInt

Predictor        Coef    SE Coef        T      P
Constant       -0.170      1.248    -0.14  0.893
Att          0.005499   0.003095     1.78  0.088
PctComp      0.828933   0.009105    91.04  0.000
Yds        -0.0008259  0.0004398    -1.88  0.072
YdsPerAtt      4.5455     0.1858    24.46  0.000
PctTD         3.34144    0.03466    96.41  0.000
PctInt       -4.17685    0.03429  -121.82  0.000

S = 0.139897   R-Sq = 100.0%   R-Sq(adj) = 100.0%

Analysis of Variance

Source          DF       SS      MS         F      P
Regression       6  2489.03  414.84  21196.51  0.000
Residual Error  25     0.49    0.02
Total           31  2489.52
```

The model is

$$\hat{y} = -0.17 + 0.0055Att + 0.8289PctComp - 0.0008Yds + 4.5455YdsPerAtt + 3.3414PctTd - 4.1769PctInt$$

e) The four variable model (`PctComp, YdsPerAtt, PctTD, PctInt`) has the second minimum Cp and also has small MS_E and large adjusted R-squared. It is also a model with the few regressors so it is preferred.

12-88 a) The min C_p model is: x_1

$$C_p = 0.0 \text{ and } MS_E = 0.2298$$

$$\hat{y} = -0.038 + 0.00850x_1$$

The min MS_E model is the same as the min C_p model

b) The full model that contains all 3 variables

$$\hat{y} = 0.001 - 0.00858x_1 - 0.021x_2 - 0.010x_3$$

where $x_1 = AGE \quad x_2 = DrinkUse \quad x_3 = CookUse$

c) No variables are selected

d) The min C_p model has only the intercept term with $C_p = -0.5$ and $MS_E = 0.2372$

The min MS_E model is the same as the min C_p in part (a).

e) None of the variables seem to be good predictors of arsenic in nails based on the models above (none of the variables are significant).

12-89 This analysis includes the emissions variables hc, co, and co2. It would be reasonable to consider models without these variables as regressors.

Best Subsets Regression: mpg versus cid, rhp, ...

Response is mpg

Vars	R-Sq	R-Sq(adj)	Mallows C-p	S	cid	rhp	etw	cmp	axle	n/v	hc	co	co2
1	66.0	64.2	26.5	3.4137			X						
1	59.1	57.0	35.3	3.7433	X								
2	81.6	79.5	8.6	2.5813	X		X						
2	78.1	75.7	13.0	2.8151			X		X				
3	88.8	86.8	1.3	2.0718	X		X		X				
3	88.8	86.8	1.4	2.0755			X		X	X			
4	90.3	87.8	1.5	1.9917	X		X		X				X
4	89.9	87.3	2.0	2.0302			X		X	X		X	
5	90.7	87.6	2.9	2.0057	X		X	X	X				X
5	90.7	87.6	2.9	2.0064	X		X		X		X		X
6	91.0	87.2	4.5	2.0442	X		X		X		X	X	X
6	91.0	87.1	4.5	2.0487	X		X	X	X		X		X
7	91.3	86.6	6.2	2.0927		X	X		X	X	X	X	X
7	91.2	86.4	6.3	2.1039	X		X	X	X		X	X	X
8	91.4	85.6	8.0	2.1651		X	X	X	X	X	X	X	X
8	91.4	85.6	8.1	2.1654	X	X	X	X	X		X	X	X
9	91.4	84.4	10.0	2.2562	X	X	X	X	X	X	X	X	X

a) The minimum C_p (1.3) model is:

$$\hat{y} = 61.001 - 0.02076x_{cid} - 0.00354x_{etw} - 3.457x_{axle}$$

The minimum MSE (4.0228) model is:

$$\hat{y} = 49.5 - 0.017547x_{cid} - 0.0034252x_{etw} + 1.29x_{cmp} - 3.184x_{axle} - 0.0096x_{c02}$$

b) $\hat{y} = 63.31 - 0.0178x_{cid} - 0.00375x_{etw} - 3.3x_{axle} - 0.0084x_{c02}$

c) Same model as the min MS_E equation in part (a)

d) $\hat{y} = 45.18 - 0.00321x_{etw} - 4.4x_{axle} + 0.385x_{n/v}$

e) The minimum C_p model is preferred because it has a very low MSE as well (4.29)

f) Only one indicator variable is used for transmission to distinguish the automatic from manual types and two indicator variables are used for drv:
x_{trans} = 0 for automatic (L4, L5) and 1 for manual (M6) and

x_{drv1} = 0 if drv = 4 or R and 1 if drv =F; x_{drv2} = 0 if drv = 4 or F and 1 if drv = R.

The minimum C_p (4.0) model is the same as the minimum MSE (2.267) model:

$$\hat{y} = 10 - 0.0038023x_{etw} + 3.936x_{cmp} + 15.216x_{co} - 0.011118x_{c02} - 7.401x_{trans} +$$
$$3.6131x_{drv1} + 2.342x_{drv2}$$

Stepwise:

$$\hat{y} = 39.12 - 0.0044x_{etw} + 0.271x_{n/v} - 4.5x_{trns} + 3.2x_{drv1} + 1.7x_{drv2}$$

Forward selection:

$$\hat{y} = 41.12 - 0.00377x_{etw} + 0.336x_{n/v} - 2.1x_{axle} - 3.4x_{trans} +$$
$$2.1x_{drv1} + 2x_{drv2}$$

Backward selection: same as minimum C_p and minimum MSE.
Prefer the model giving the minimum C_p and minimum MSE.

12-90

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

a) $\hat{y} = 759.395 - 90.783x' - 47.166(x')^2$, where $x' = \dfrac{x - \bar{x}}{S_x}$

b) At x = 285 $x' = \dfrac{285 - 297.125}{11.9336} = -1.016$

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

c) $\hat{y} = 759.395 - 90.783\left(\frac{x - 297.125}{119336}\right) - 47.166\left(\frac{x - 297.125}{119336}\right)^2$

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

d) They are the same.

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

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

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

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

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

12-91 The default settings for F-to-enter and F-to-remove, equal to 4, 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).

12-92 $n = 30$, $k = 9$, $p = 9 + 1 = 10$ in full model.

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

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

$$SS_E = MS_E(n - p)$$

$$= 100(30 - 10)$$

$$= 2000$$

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

$$25000 = S_{yy}$$

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

$$= 25000 - 2000 = 23000$$

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

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

$$f_{0.01,9,20} = 3.46$$

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

Reject H_0 and conclude at least one β_j is significant at $\alpha = 0.01$.

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

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

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

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

Yes, C_p is reduced from the full model.

12-93 n = 30 k = 7 $MS_{E(full)} = 10$

a) p =3 $SS_E = 300$

$$MS_E = \frac{SS_E}{n-p} = \frac{300}{30-3} = 11.11$$

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

$$= \frac{300}{10} - 30 + 2(3)$$

$$= 6$$

YES, $C_p > p$

b) p = 4 $SS_E = 275$

$$MS_E = \frac{SS_E}{n-p} = \frac{275}{30-4} = 10.6 \quad C_p = \frac{275}{10} - 30 + 2(4) = 5.5$$

Yes, both MS_E and C_p are reduced.

Supplemental Exercises

12-94 a) The missing quantities are as follows:

$$T_{Constant} = \frac{Coef}{SE\ Coef} = \frac{517.46}{17.68} = 29.2681$$

From the t table with 16 degrees of freedom, P-value$_{Constant}$ < 2(0.0005), so P-value$_{Constant}$ < 0.001

$$T_{x1} = \frac{Coef}{SE\ Coef}, \text{ SE Coef}_{x1} = \frac{Coef}{T_{x1}} = \frac{11.4720}{36.50} = 0.3143$$

P-value$_{x1}$ < 2(0.0005), so P-value$_{x1}$ < 0.001

$$T_{X2} = \frac{Coef}{SE\ Coef} = \frac{-8.1378}{0.1969} = -41.3296$$

P-value$_{x2}$ < 2(0.0005), so P-value$_{x2}$ < 0.001

$$T_{x3} = \frac{Coef}{SE\ Coef} = \frac{10.8565}{0.6652} = 16.3207$$

P-value$_{x3}$ < 2(0.0005), so P-value$_{x3}$ < 0.001

Regression DF = 19 – 16 = 3

$$SS_{Error} = SS_{Total} - SS_{Regression} = 348983 - 347300 = 1683$$

$$F = \frac{MS_{Regression}}{MS_{Error}} = \frac{115767}{105} = 1102.543$$

P-value < 0.01.

R-Squared = 347300/348983 = 0.995
R-Squared Adjusted = 1 – (1683/16)/(348943/19) = 0.994

b) Because the P-value from the F-test is less than α = 0.05 and less than α = 0.01, we reject the H_0 for either α value and conclude that at least one regressor significantly contributes to the model.

c) Because the P-value from the t-test for the x1, x2, and x3 variables are less than $\alpha = 0.05$, we reject the H_0's and conclude that each individual regressor contributes significantly to the model.

12-95 a) Because the matrix is 3×3 two regressors are in the regression model. The intercept is also in the model.

b) $\text{cov}(\hat{\beta}) = \sigma^2 (X'X)^{-1} = \sigma^2 C$, therefore the variances of the two variables regression coefficients are:

$50(0.0013329) = 0.066645$ and $50(0.0009108) = 0.04554$

c) $se(\hat{\beta}_0) = \sqrt{\hat{\sigma}^2 C_{11}} = \sqrt{50 \times 0.0013329} = 0.258$

12-96

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	3	9863398	3287799	1724.42	<.0001
Error	36	68638	1906.61573		
Corrected Total	39	9932036			

Root MSE	43.66481	R-Square	0.9931
Dependent Mean	3904.00000	Adj R-Sq	0.9925
Coeff Var	1.11846		

Parameter Estimates

Variable	DF	Parameter Estimate	Standard Error	t Value	Pr > |t|	99% Confidence Limits	
Intercept	1	3829.26146	2262.07465	1.69	0.0991	-2322.41579	9980.93871
x3	1	-0.21494	0.10885	-1.97	0.0560	-0.51095	0.08106
x4	1	21.21343	0.90498	23.44	<.0001	18.75235	23.67450
x5	1	1.65659	0.55017	3.01	0.0047	0.16040	3.15279

a) $\hat{y} = 3829 - 0.215x_3 + 21.213x_4 + 1.657x_5$

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

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

$\alpha = 0.01$ $f_0 = 1724.42$

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

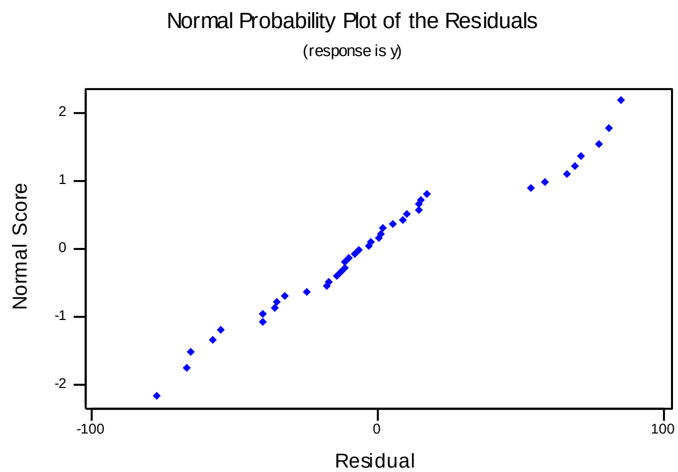
Reject H_0 and conclude that regression is significant. P-value < 0.00001

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

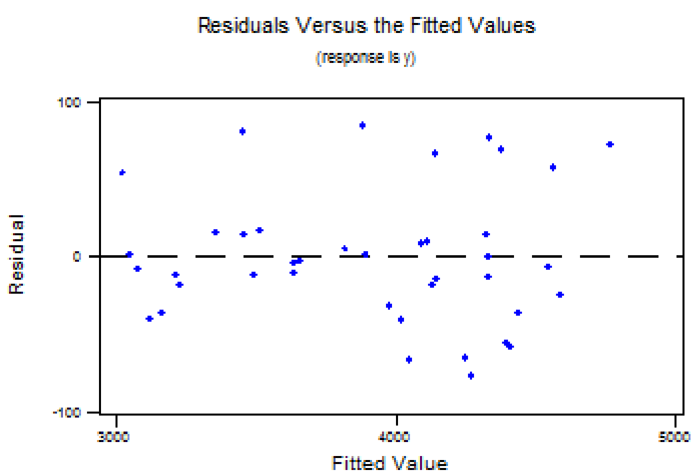
$H_0: \beta_3 = 0$	$H_0: \beta_4 = 0$	$H_0: \beta_5 = 0$
$H_1: \beta_3 \neq 0$	$H_1: \beta_4 \neq 0$	$H_1: \beta_5 \neq 0$
$t_0 = -1.97$	$t_0 = 23.44$	$t_0 = 3.01$
$\|t_0\| \not> t_{\alpha/2,36}$	$\|t_0\| > t_{\alpha/2,36}$	$\|t_0\| > t_{\alpha/2,36}$
Fail to reject H_0	Reject H_0	Reject H_0

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

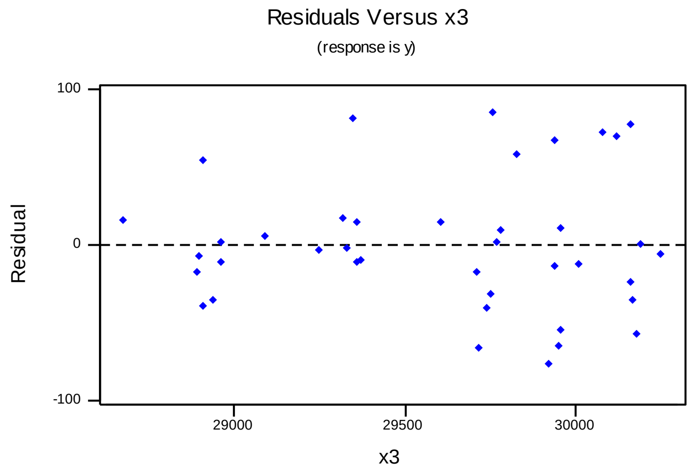
e) Normality assumption appears reasonable. However there is a gap in the line.



f) Plot is satisfactory.



g) Slight indication that variance increases as x_3 increases.



h)

Output Statistics

Obs	Dependent Variable	Predicted Value	Std Error Mean Predict	99% CL Mean		99% CL Predict		Residual
41	.	3743	30.0666	3661	3825	3599	3887	.

$\hat{y} = 3829 - 0.215(29741) + 21.213(170) + 1.657(1630) = 3743$

12-97

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	3	0.68611	0.22870	1323.62	<.0001
Error	36	0.00622	0.00017279		
Corrected Total	39	0.69233			

Root MSE	0.01314	R-Square	0.9910
Dependent Mean	8.26128	Adj R-Sq	0.9903
Coeff Var	0.15911		

Parameter Estimates

Variable	DF	Parameter Estimate	Standard Error	t Value	Pr > |t|	95% Confidence Limits	
Intercept	1	19.69047	9.58709	2.05	0.0473	0.24695	39.13399
x^*3	1	-1.26731	0.95939	-1.32	0.1949	-3.21304	0.67842
x4	1	0.00541	0.00027111	19.97	<.0001	0.00486	0.00596
x5	1	0.00040789	0.00016448	2.48	0.0180	0.00007430	0.00074148

a) $H_0: \beta_3^* = \beta_4 = \beta_5 = 0$

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

$\alpha = 0.05$

$f_0 = 1323.62$

$f_{.05,3,36} = 2.87$

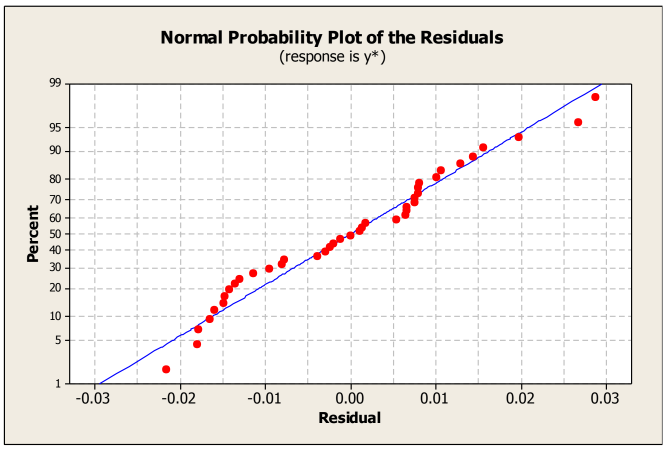
$f_0 >> f_{0.05,3,36}$

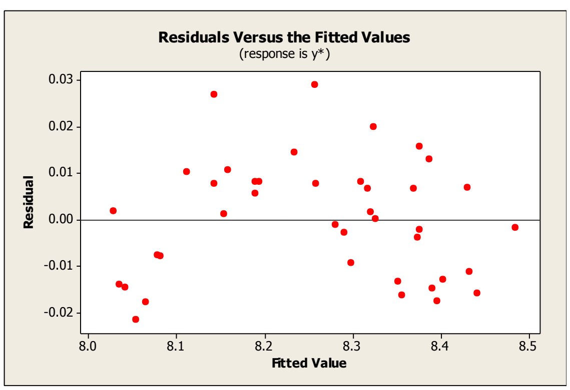
Reject H_0 and conclude that regression is significant. P-value < 0.00001

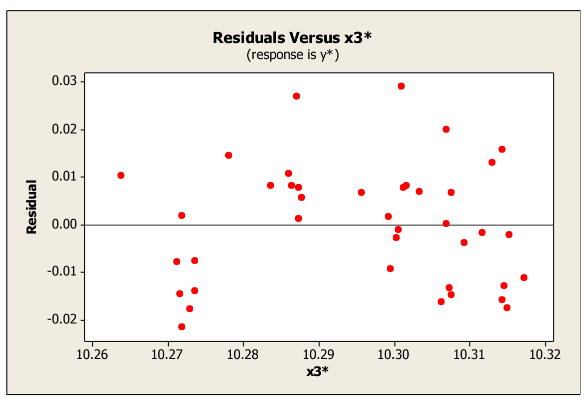
b) $\alpha = 0.05$ $\quad t_{.025,36} = 2.028$

$H_0: \beta_3^* = 0$	$H_0: \beta_4 = 0$	$H_0: \beta_5 = 0$
$H_1: \beta_3^* \neq 0$	$H_1: \beta_4 \neq 0$	$H_1: \beta_5 \neq 0$
$t_0 = -1.32$	$t_0 = 19.97$	$t_0 = 2.48$
$\lvert t_0 \rvert \not> t_{\alpha/2,36}$	$\lvert t_0 \rvert > t_{\alpha/2,36}$	$\lvert t_0 \rvert > t_{\alpha/2,36}$
Fail to reject H_0	Reject H_0	Reject H_0

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







12-98 a) $\hat{y} = 2.86 - 0.291x_1 + 0.206x_2 + 0.454x_3 - 0.594x_4 + 0.0046x_5$

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

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

$\alpha = 0.01$

$f_0 = 4.81$
$f_{.01,5,19} = 4.17$
$f_0 > f_{\alpha,5,19}$
Reject H_0. P-value = 0.005

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

$H_0: \beta_1 = 0$	$H_0: \beta_2 = 0$	$H_0: \beta_3 = 0$	$H_0: \beta_4 = 0$	$H_0: \beta_5 = 0$
$H_1: \beta_1 \neq 0$	$H_1: \beta_2 \neq 0$	$H_1: \beta_3 \neq 0$	$H_1: \beta_4 \neq 0$	$H_1: \beta_5 \neq 0$
$t_0 = 2.48$	$t_0 = 2.74$	$t_0 = 2.42$	$t_0 = -2.80$	$t_0 = 0.25$
$\lvert t_0 \rvert > t_{\alpha/2,19}$	$\lvert t_0 \rvert > t_{\alpha/2,19}$	$\lvert t_0 \rvert > t_{\alpha/2,19}$	$\lvert t_0 \rvert > t_{\alpha/2,19}$	$\lvert t_0 \rvert \not> t_{\alpha/2,19}$
Reject H_0	Reject H_0	Reject H_0	Reject H_0	Do not reject H_0

c) $\hat{y} = 3.15 - 0.290x_1 + 0.199x_2 + 0.455x_3 - 0.609x_4$
$H_0: \beta_1 = \beta_2 = \beta_3 = \beta_4 = 0$
$H_1: \beta_j \neq 0$ for at least one j
$\alpha = 0.05$
$f_0 = 6.29$
$f_{.05,4,20} = 2.87$
$f_0 > f_{\alpha,4,20}$
Reject H_0.
$\alpha = 0.05$ $t_{.025,20} = 2.086$

$H_0: \beta_1 = 0$	$H_0: \beta_2 = 0$	$H_0: \beta_3 = 0$	$H_0: \beta_4 = 0$
$H_1: \beta_1 \neq 0$	$H_1: \beta_2 \neq 0$	$H_1: \beta_3 \neq 0$	$H_1: \beta_4 \neq 0$
$t_0 = -2.53$	$t_0 = 2.89$	$t_0 = 2.49$	$t_0 = -3.06$
$\lvert t_0 \rvert > t_{\alpha/2,20}$	$\lvert t_0 \rvert > t_{\alpha/2,20}$	$\lvert t_0 \rvert > t_{\alpha/2,20}$	$\lvert t_0 \rvert > t_{\alpha/2,20}$
Reject H_0	Reject H_0	Reject H_0	Reject H_0

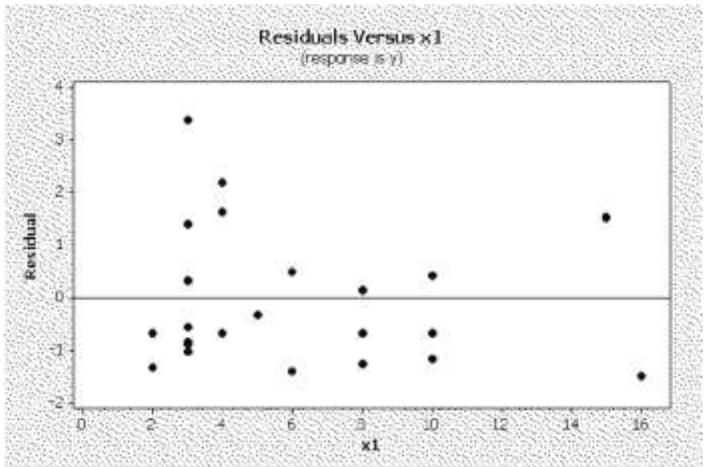
d) The addition of the 5th regressor results in a loss of one degree of freedom in the denominator and the reduction in SS_E is not enough to compensate for this loss.
e) Observation 2 is unusually large. Studentized residuals
–0.80854 –3.37086 –0.40387 2.03560 –0.53056 0.63857 –0.46337 2.02266 1.34194
–0.43964 0.76250 –0.32729 –0.37858 1.80031 0.08162 –0.70885 –0.80306 0.28525
0.61027 0.60649 –0.12561 0.72730 –0.74124 –0.83337 –0.46491
f) R^2 for model in part (a): 0.558. R^2 for model in part (c): 0.557. R^2 for model x_1, x_2, x_3, x_4 without obs. #2: 0.804. R^2 increased because observation 2 was not fit well by either of the previous models.
g) $H_0: \beta_1 = \beta_2 = \beta_3 = \beta_4 = 0$
$H_1: \beta_j \neq 0$ $\alpha = 0.05$
$f_0 = 19.53$
$f_{.05,4,19} = 2.90$
$f_0 > f_{0.05,4,19}$
Reject H_0.

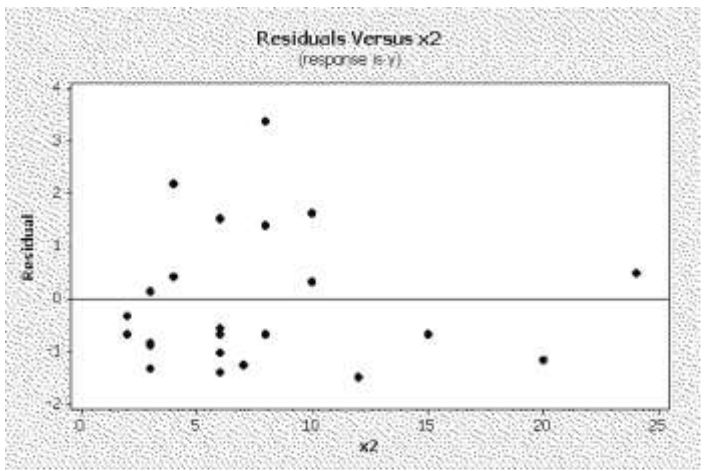
$\alpha = 0.05$ $t_{.025,19} = 2.093$

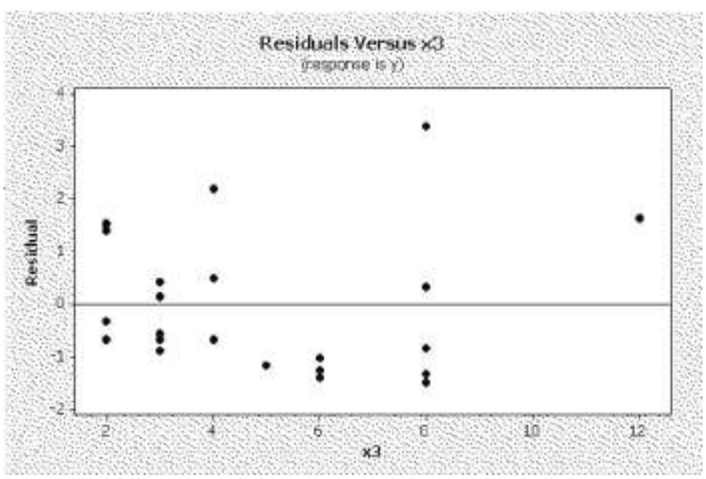
$H_0: \beta_1 = 0$	$H_0: \beta_2 = 0$	$H_0: \beta_3 = 0$	$H_0: \beta_4 = 0$
$H_1: \beta_1 \neq 0$	$H_1: \beta_2 \neq 0$	$H_1: \beta_3 \neq 0$	$H_1: \beta_4 \neq 0$
$t_0 = -3.96$	$t_0 = 6.43$	$t_0 = 3.64$	$t_0 = -3.39$
$\lvert t_0 \rvert > t_{0.025,19}$	$\lvert t_0 \rvert > t_{0.025,19}$	$\lvert t_0 \rvert > t_{0.025,19}$	$\lvert t_0 \rvert > t_{0.025,19}$

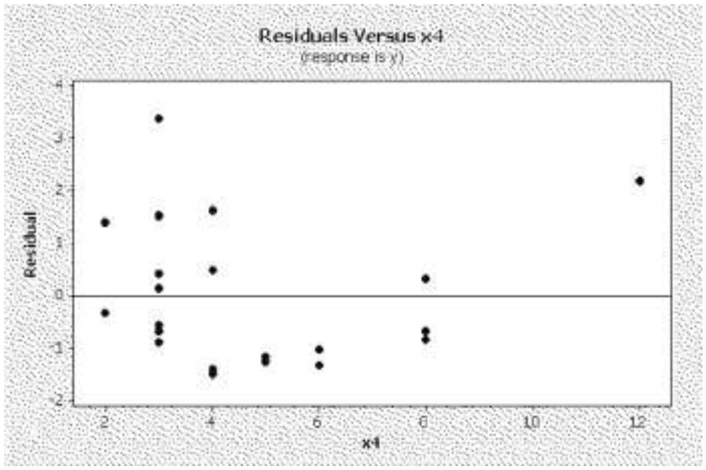
Reject H_0 Reject H_0 Reject H_0 Reject H_0

h) There is some indication of curvature.









12-99 Note that data in row 2 are deleted to follow the instructions in the exercise.

```
a)
The regression equation is
  y* = - 0.908 + 5.48 x1* + 1.13 x2* - 3.92 x3*
             - 1.14 x4*

Predictor       Coef  SE Coef      T      P
Constant     -0.9082   0.6746  -1.35  0.194
  x1*         5.4823   0.4865  11.27  0.000
  x2*        1.12563  0.07714  14.59  0.000
  x3*        -3.9198   0.5619  -6.98  0.000
  x4*        -1.1429   0.1410  -8.11  0.000

S = 0.282333   R-Sq = 95.8%   R-Sq(adj) = 94.9%

Analysis of Variance

Source          DF       SS      MS       F      P
Regression       4  34.7600  8.6900  109.02  0.000
Residual Error  19   1.5145  0.0797
Total           23  36.2745
```

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

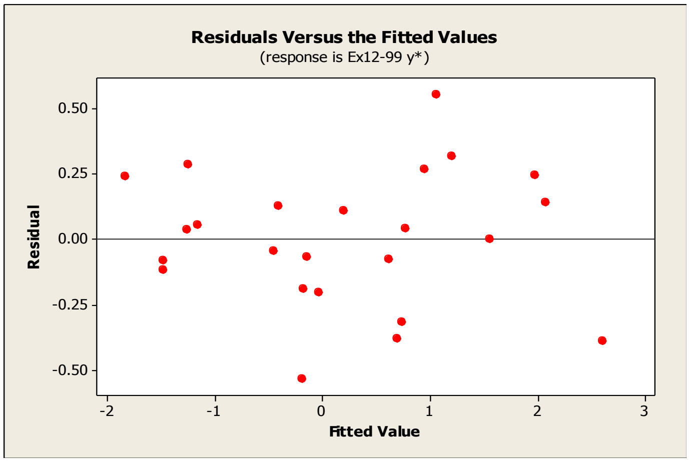
H_1 : at least one $\beta_j \neq 0$

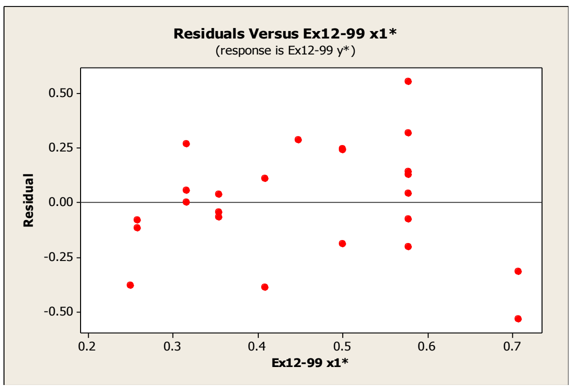
$\alpha = 0.05$

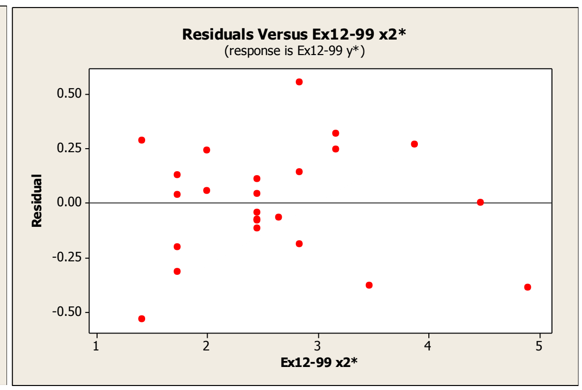
$f_0 = 109.02$, P-value ≈ 0,

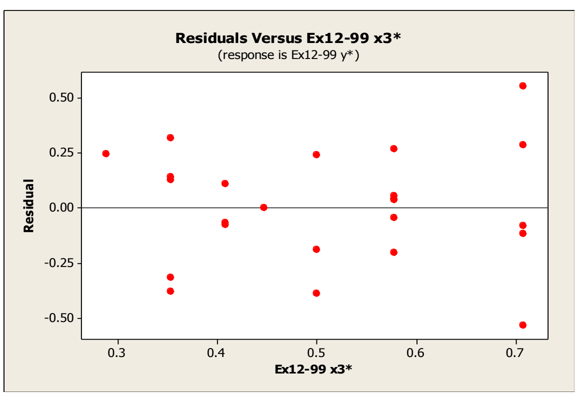
Reject H_0 at $\alpha = 0.05$.
T tests appear in the previous computer output. Because all P-values ≈ 0, all tests reject H_0

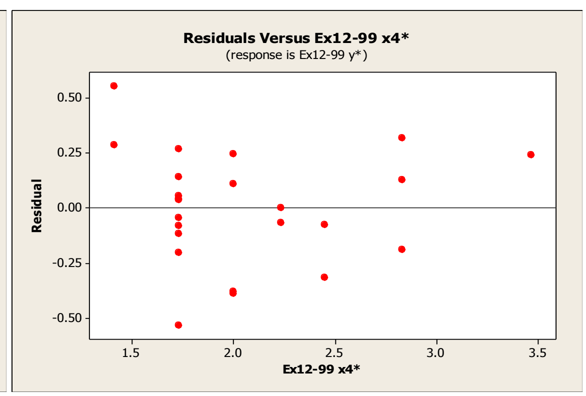
c) The residual plots are more satisfactory than the plots in the previous exercise.











12-100

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	2	26.48710	13.24355	300.11	<.0001
Error	7	0.30890	0.04413		
Corrected Total	9	26.79600			

Root MSE	0.21007	R-Square	0.9885
Dependent Mean	14.12000	Adj R-Sq	0.9852
Coeff Var	1.48773		

Parameter Estimates

Variable	DF	Parameter Estimate	Standard Error	t Value	Pr > |t|	99% Confidence Limits	
Intercept	1	-1709.40539	244.76770	-6.98	0.0002	-2565.96589	-852.84490
x	1	2.02291	0.27980	7.23	0.0002	1.04376	3.00207
x^2	1	-0.00059293	0.00007994	-7.42	0.0001	-0.00087269	-0.00031317

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

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

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

$\alpha = 0.01$

$f_0 = 300.11$

$f_{.01,2,7} = 9.54$

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

Reject H_0.

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

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

$\alpha = 0.01$

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

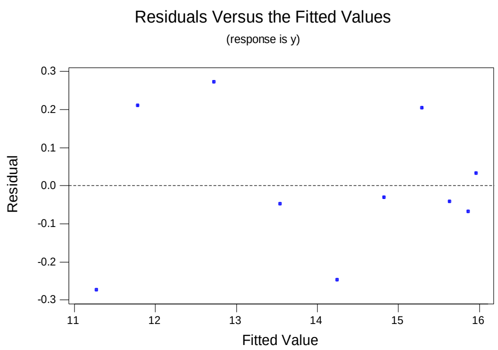
$f_0 = 55.01$

$f_{.01,1,7} = 12.25$

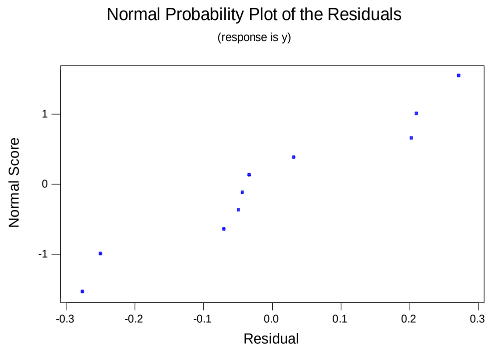
$f_0 >> f_{\alpha,1,7}$

Reject H_0.

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



e) Normality assumption is reasonable.



12-101 a) $\hat{y}^* = 21.068 - 1.404x_3^* + 0.0055x_4 + 0.000418x_5$

$MS_E = 0.013156$

$C_p = 4.0$

b) Same as model in part (a)

c) x_4, x_5 with $C_p = 4.1$ and $MS_E = 0.0134$

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.

e) $VIF(\hat{\beta}_3^*) = 52.4$

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

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

Yes, VIFs for x_3^* and x_5 exceed 10

12-102 a) $\hat{y} = 4.87 + 6.12x_1^* - 6.53x_2^* - 3.56x_3^* - 1.44x_4^*$

$MS_E(p) = 0.41642$

Min $C_p = 5.0$

b) Same as part (a)

c) Same as part (a)

d) All models are the same.

12-103 a) $\hat{y} = 300.0 + 0.85x_1 + 10.4x_2$

$\hat{y} = 300 + 0.85(38) + 10.4(3) = 363.5$

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

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

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

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

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

$H_0: \beta_1 = \beta_2 = 0$

$H_1:$ 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) $R^2 = \frac{SS_R}{S_{yy}} = \frac{1110.2}{1230.5} = 0.9022$ or 90.22%

d) $k = 3 \quad p = 4 \qquad SS_E' = 117.20$

$$MS_E' = \frac{SS_E'}{n-p} = \frac{117.2}{11} = 10.65$$

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) $SS_R = S_{yy} - SS_E = 1230.5 - 117.20 = 1113.30$

$$\begin{aligned} SS_R(\beta_3 \mid \beta_2, \beta_1, \beta_0) &= SS_R(\beta_3\beta_2\beta_1 \mid \beta_0) - SS_R(\beta_2, \beta_1 \mid \beta_0) \\ &= 1113.30 - 1110.20 \\ &= 3.1 \end{aligned}$$

$H_0 : \beta_3 = 0$

$H_1 : \beta_3 \neq 0$

$\alpha = 0.05$

$$f_0 = \frac{SS_R(\beta_3 \mid \beta_2, \beta_1, \beta_0)/r}{SS_E'/n-p} = \frac{3.1/1}{117.2/11} = 0.291$$

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

$f_0 \not> f_{0.05,1,11}$

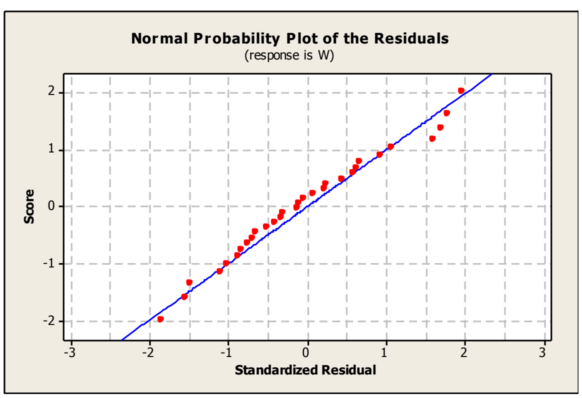
Do not reject H_0.

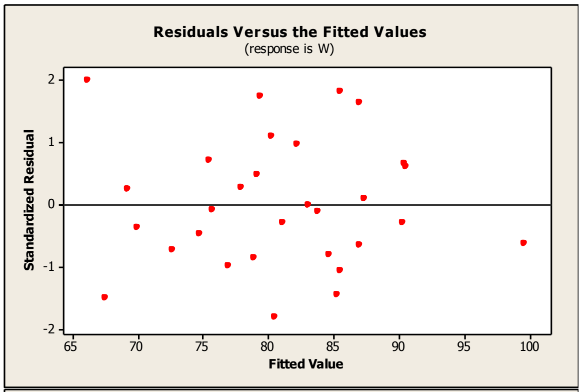
12-104 a) The model with the minimum C_p (-1.3) value is:

```
W = 69.9 + 0.120 HR_B + 0.0737 BB_B - 0.0532 SO_B + 0.0942 SB
```

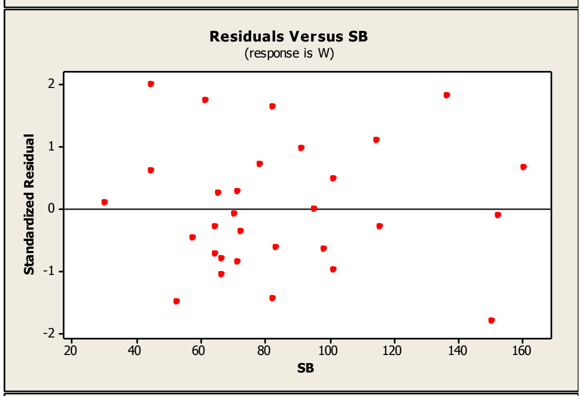
where X_1 = AVG, X_2 = R, X_3 = H, X_4 = 2B, X_5 = 3B, X_6 = HR, X_7 = RBI, X_8 = BB, X_9 = SO, X_{10} = SB, X_{11} = GIDP, X_{12} = LOB and X_{13} = OBP

The model assumptions are not violated based on the following graphs.

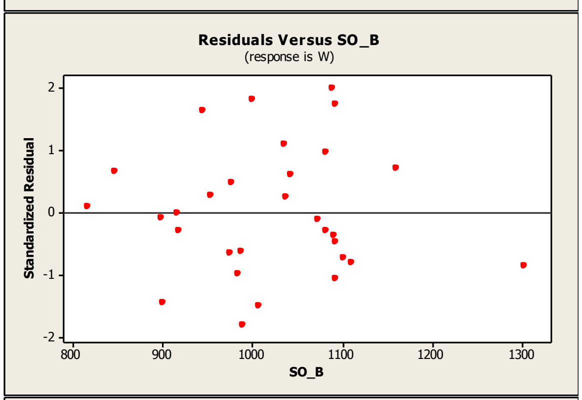




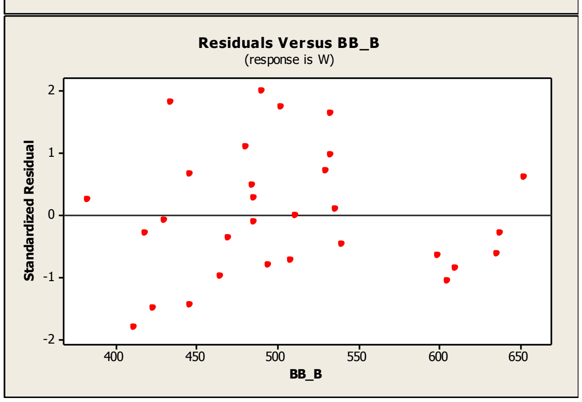
Residuals Versus the Fitted Values
(response is W)
Standardized Residual
Fitted Value



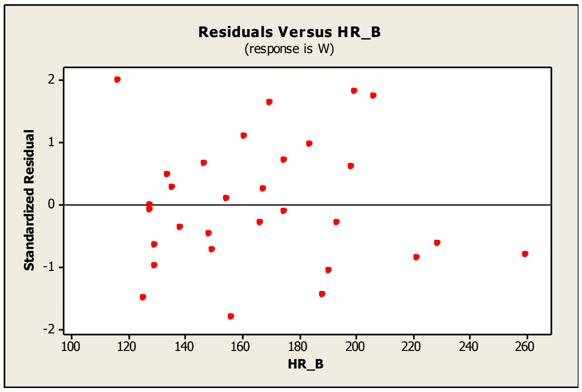
Residuals Versus SB
(response is W)
Standardized Residual
SB



Residuals Versus SO_B
(response is W)
Standardized Residual
SO_B

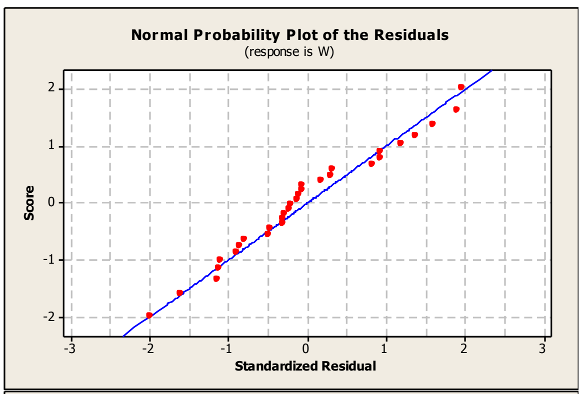


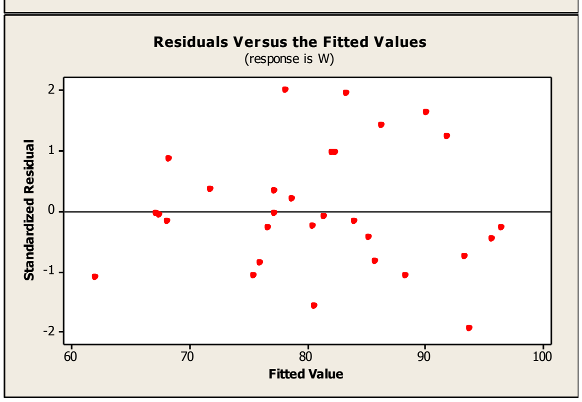
Residuals Versus BB_B
(response is W)
Standardized Residual
BB_B

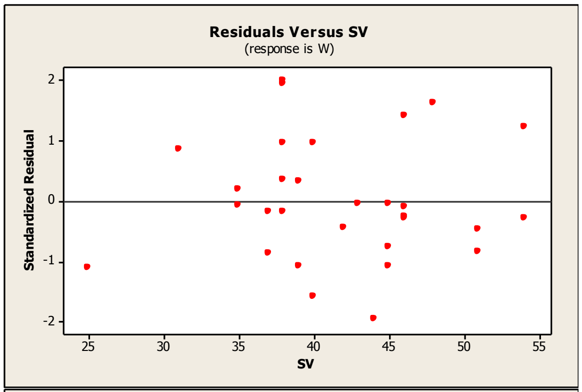


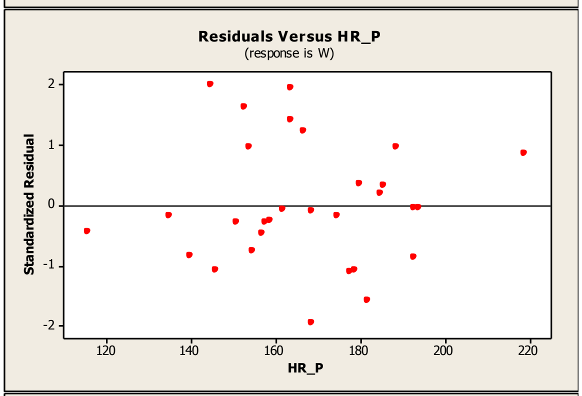
b)
Minimum C_p (1.1) model:
`W = 96.5 + 0.527 SV - 0.125 HR_P - 0.0847 BB_P + 0.0257 SO_P`

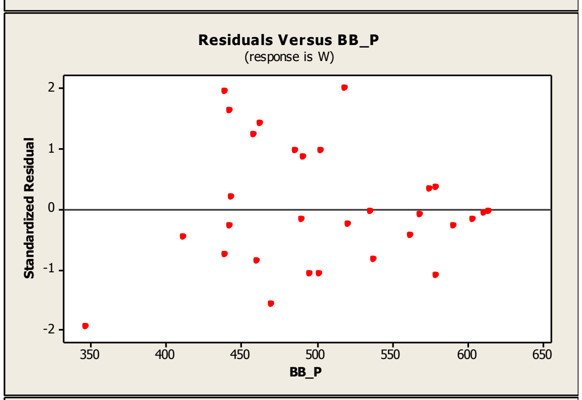
Based on the graphs below, the model assumptions are not violated

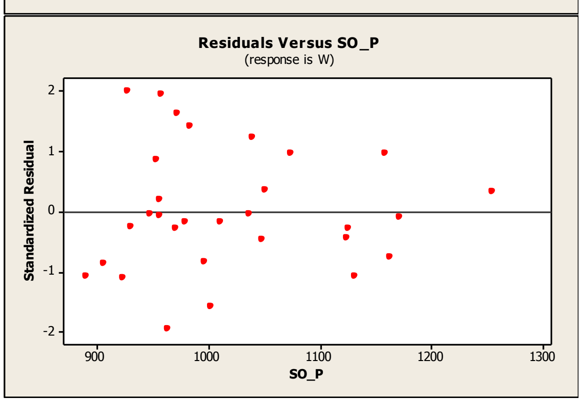








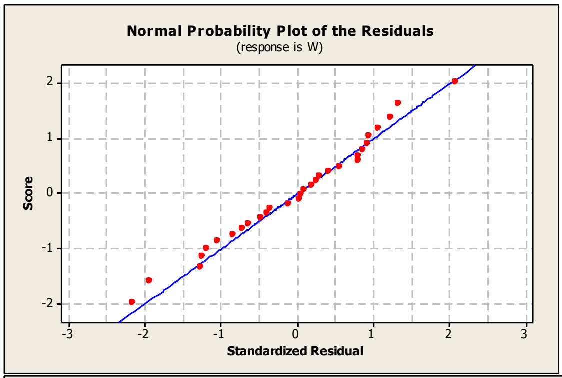


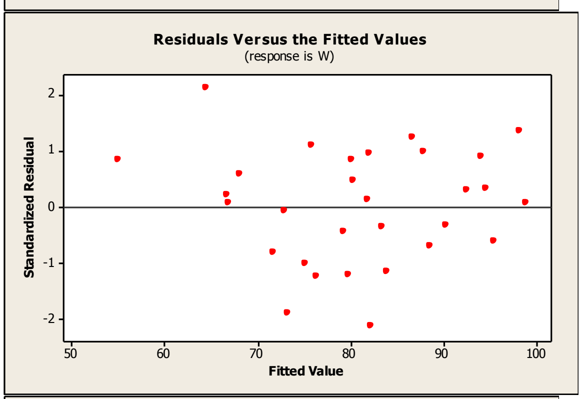


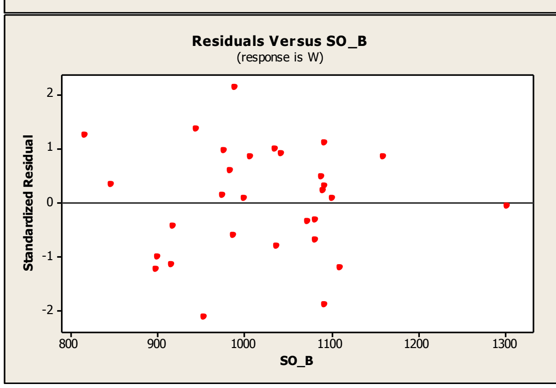
c) Minimum C_p (10.7) model:

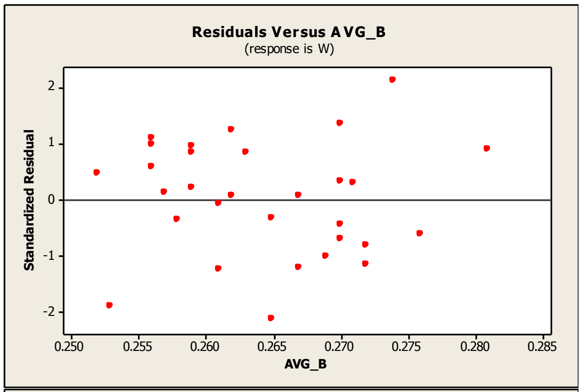
$$\hat{y} = -2.49 + 3277x_{avg\_b} - 0.45303x_{h\_b} - 0.04041x_{2b} - 0.13662x_{3b} + 0.19914x_{rbi}$$
$$- 0.010207x_{so\_b} + 0.07897x_{lob} - 870.2x_{obp} - 134.79x_{era} + 0.81681x_{er} - 0.06698x_{hr\_p}$$
$$- 0.032314x_{bb\_p} + 0.008755x_{so\_p}$$

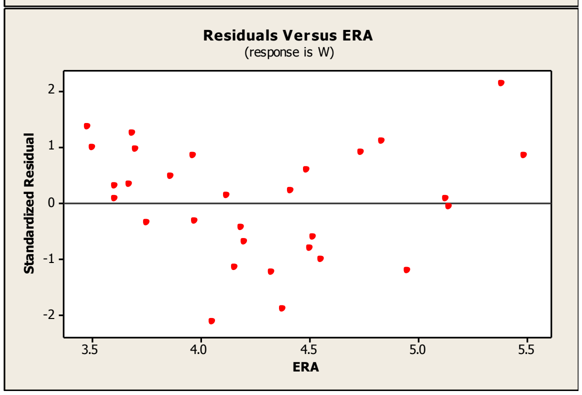
Every variable in the above model is significant at α = 0.10. If α is decreased to 0.05, SO_P is no longer significant. The residual plots do not show any violations of the model assumptions (only a few plots of residuals vs. the regressors are shown).











12-105 a) $R^2 = \dfrac{SS_R}{S_{yy}}$

$$SS_R = R^2(S_{yy}) = 0.94(0.55) = 0.517$$

$$SS_E = S_{yy} - SS_R = 0.55 - 0.517 = 0.033$$

$$H_0 : \beta_1 = \beta_2 = ... = \beta_6 = 0$$

$$H_1 : \text{at least one } \beta_j \neq 0$$

α = 0.05

$$f_0 = \frac{SS_R / k}{SS_E / n - p} = \frac{0.517/6}{0.033/7} = 18.28$$

$$f_{.05,6,7} = 3.87$$

$$f_0 > f_{0.05,6,7}$$

Reject H_0.

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

$$SS_R' = R^2(S_{yy}) = 0.92(0.55) = 0.506$$

$$SS_E' = S_{yy} - SS_R' = 0.55 - 0.506 = 0.044$$

$$\begin{aligned} SS_R(\beta\beta_{j_0}, \beta_{i,i=1,2,\ldots,6} \mid \beta_0) &= SS_R(full) - SS_R(reduced) \\ &= 0.517 - 0.506 \\ &= 0.011 \end{aligned}$$

$$f_0 = \frac{SS_R(\beta_j\ ,\beta_{i,i=1,2,\ldots,6} \mid \beta_0)/r}{SS_E'/(n-p)} = \frac{0.011/1}{0.044/8} = 2$$

$$f_{.05,1,8} = 5.32$$

$$f_0 \not> f_{0.05,1,8}$$

Fail to reject H_0. There is not sufficient evidence that the removed variable is significant at $\alpha = 0.05$.

c) $MS_E(reduced) = \frac{SS_E}{n-p} = \frac{0.044}{8} = 0.0055$

$$MS_E(full) = \frac{0.033}{7} = 0.0047$$

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 large because the added variable had no real explanatory power.

12-106 The Minitab result is shown below. The P-value of the *Surg-Med* indicator variable (third variable) is greater than the alpha level of 0.05, so we fail to reject the H_0 and conclude that *Surg-Med* indicator variable does not contribute significantly to the model. Thus, the surgical and medical service does not impact the reported satisfaction.

Regression Analysis: Satisfaction versus Age, Severity, ...

```
The regression equation is
Satisfaction = 144 - 1.12 Age - 0.586 Severity + 0.41 Surg-Med + 1.31 Anxiety

Predictor      Coef  SE Coef      T      P
Constant    143.867    6.044  23.80  0.000
Age         -1.1172   0.1383  -8.08  0.000
Severity    -0.5862   0.1356  -4.32  0.000
Surg-Med      0.415    3.008   0.14  0.892
Anxiety       1.306    1.084   1.21  0.242

S = 7.20745   R-Sq = 90.4%   R-Sq(adj) = 88.4%

Analysis of Variance

Source          DF       SS      MS      F      P
Regression       4   9739.3  2434.8  46.87  0.000
Residual Error  20   1038.9    51.9
Total           24  10778.2
```

12-107 a) Because the matrix is 4 by 4 there are 3 regressors in the model (plus the intercept).

b) Because $(X'X)^{-1}$ is diagonal each element can be inverted to obtain $(X'X)$ and from the normal equations the (1, 1) element (the top-left element) of $(X'X)$ = n. Therefore, n = 1/0.25 = 4.

c) The original columns are orthogonal to each other.

<u>Mind-Expanding Exercises</u>

12-108 Because $R^2 = \frac{SS_R}{S_{yy}}$ and $1-R^2 = \frac{SS_E}{S_{yy}}$, $F_0 = \frac{SS_R/k}{SS_E/(n-k-1)}$ and this is the usual F-test for significance of regression. Then, $F_0 = \frac{0.95/4}{(1-0.95)/(20-4-1)} = 71.25$ and the critical value is $f_{.05,4,15} = 3.06$. Because 71.25 > 3.06, regression is significant.

12-109 Using n = 20, k = 4, $f_{0.01,4,15} = 4.89$. Reject H_0 if

$$\frac{R^2/4}{(1-R^2)/15} \geq 4.89$$

$$\frac{R^2}{(1-R^2)} \geq 1.304$$

Then, $R^2 \geq 0.566$ results in a significant regression.

12-110 Because $\hat{\beta} = (X'X)^{-1}X'Y$, $e = Y - X\hat{\beta} = Y - X(X'X)^{-1}X'Y = (I-H)Y$

12-111 From the previous exercise, e_i is ith element of $(I\text{-}H)Y$. That is,

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

and

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

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

$$e_i = -h_{i,1}Y_1 - h_{i,2}Y_2 - ... - h_{i,i-1}Y_{i-1} + (1-h_{i,i})Y_i - h_{i,i+1}Y_{i+1} - ... - h_{i,n}Y_n$$
$$e_j = -h_{j,1}Y_1 - h_{j,2}Y_2 - ... - h_{j,j-1}Y_{j-1} + (1-h_{j,j})Y_j - h_{j,j+1}Y_{j+1} - ... - h_{j,n}Y_n$$

Because the y_i's are independent,

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

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

12-112 $$\hat{\beta} = (X'X)^{-1}X'Y = (\mathrm{X'X})^{-1}X'(X\beta + \varepsilon) = \beta + (\mathrm{X'X})^{-1}X'\varepsilon = \beta + R\varepsilon$$

12-113 a) Min L = (**y** – **Xβ**)'(**y** – **Xβ**) subject to **Tβ = c**

This is equivalent to Min Z = (**y** – **Xβ**)'(**y** – **Xβ**) + 2**γ**'(**Tβ** – **c**)
where **γ** = (γ1, γ2, … γp)' is a vector of Lagrange multipliers.

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

$$\frac{\partial Z}{\partial \gamma} = 2(T\beta - c)$$

Set $\frac{\partial Z}{\partial \beta} = 0$ and $\frac{\partial Z}{\partial \gamma} = 0$

Then
$(\mathbf{X'X})\boldsymbol{\beta}_c + \mathbf{T'}\boldsymbol{\gamma} = \mathbf{X'y}$
$\mathbf{T}\boldsymbol{\beta}_c = \mathbf{c}$
where $\boldsymbol{\beta}_\mathbf{c}$ is the constrained estimator.

From the first of these equations
$\boldsymbol{\beta}_\mathbf{c} = (\mathbf{X'X})^{-1}(\mathbf{X'y} - \mathbf{T'}\boldsymbol{\gamma}) = \boldsymbol{\beta} - (\mathbf{X'X})^{-1}\mathbf{T'}\boldsymbol{\gamma}$

From the second
$\mathbf{T}\boldsymbol{\beta} - \mathbf{T}(\mathbf{X'X})^{-1}\mathbf{T'}\boldsymbol{\gamma} = \mathbf{c}$ and $\boldsymbol{\gamma} = [\mathbf{T}(\mathbf{X'X})^{-1}\mathbf{T'}]^{-1}(\mathbf{T}\boldsymbol{\beta} - \mathbf{c})$

Then

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

b) This solution would be appropriate in situations where you know that there are linear relationships between the coefficients.

12-114 a) 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, $y = \beta_0 * + \beta_1 * x_1 + \beta_2 * x_2$.

b) 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

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

c) One could estimate x* as a parameter in the model. A simple approach is to refit the model with different choices for x* and to select the value for x* that minimizes the residual sum of squares.