CHAPTER 6

Note to Instructor: For computer exercises, the procedure 'Regression' under 'Stat' in Minitab can be used for the regression analysis except for computing confidence intervals on the regressor variables.

Sections 6-2

6-1. a)

```
The regression equation is
Thermal = 0.0249 + 0.129 Density

Predictor        Coef       StDev          T        P
Constant     0.024934    0.001786      13.96    0.000
Density      0.128522    0.007738      16.61    0.000
S = 0.0005852   R-Sq = 98.6%     R-Sq(adj) = 98.2%

Analysis of Variance
Source            DF          SS          MS          F        P
Regression         1 0.000094464 0.000094464     275.84    0.000
Residual Error     4 0.000001370 0.000000342
Total              5 0.000095833
```

$\hat{y} = 0.0249 + 0.129x$

b) 0.0005747
-0.0007088
0.0001486
-0.0004799
-0.0000644
0.0005298

c) SSE = 0.000001370 $\hat{\sigma}^2 =$ 0.000000342

d) se($\hat{\beta}_1$) = 0.007738, se($\hat{\beta}_0$) = 0.001786

e) SST = 0.000095833
SSR = 0.000094464, SSE = 0.000001370, and SSR + SSE = 0.000095834
$\therefore$ SST = SSR + SSE

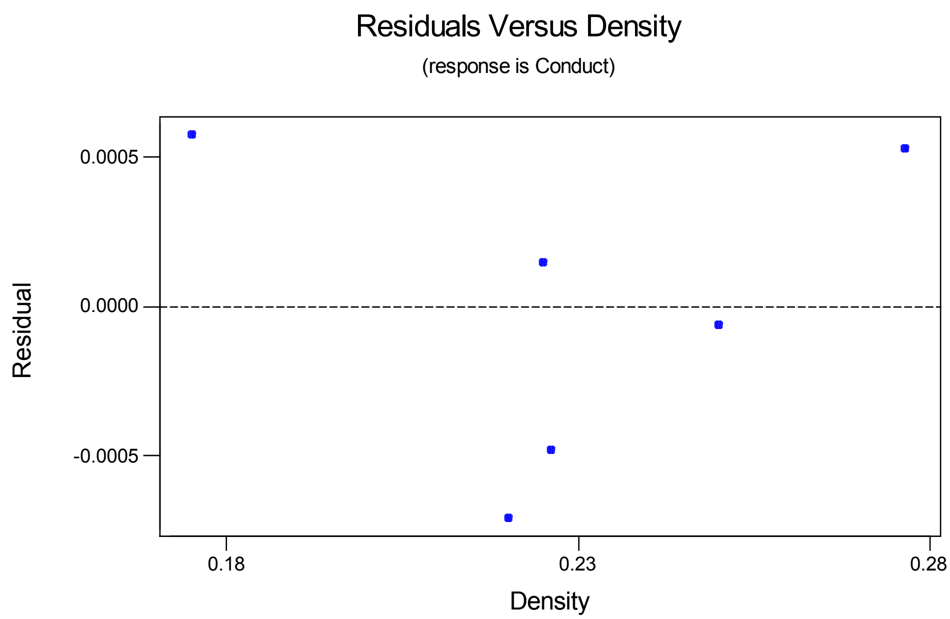
f) R^2 = 98.6%. This is interpreted as 98.6% of the total variability in thermal conductivity can be explained by the fitted regression model.

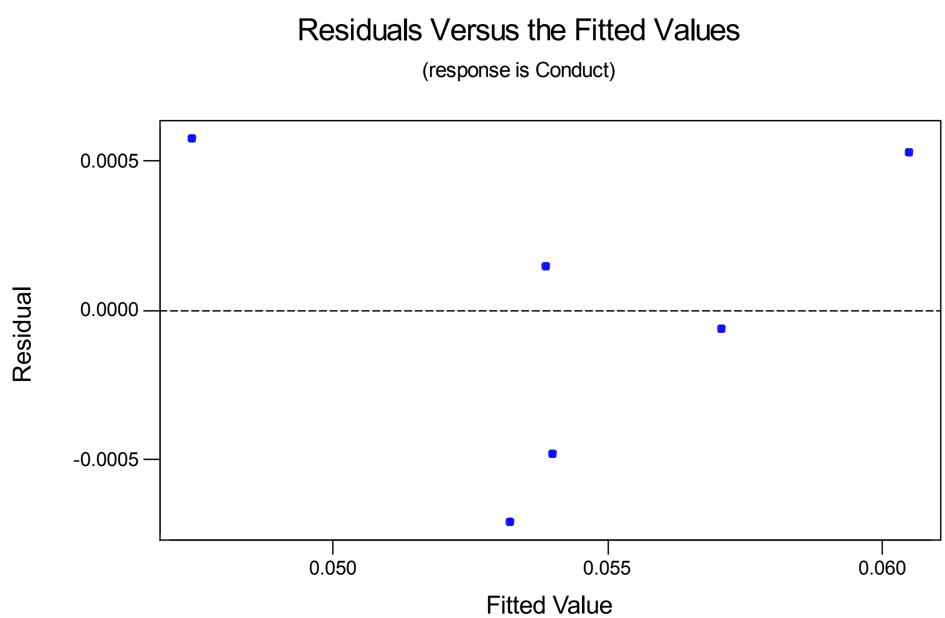
g) See the Minitab output given in part a. Based on the t-tests, we conclude that the slope and intercept are nonzero.

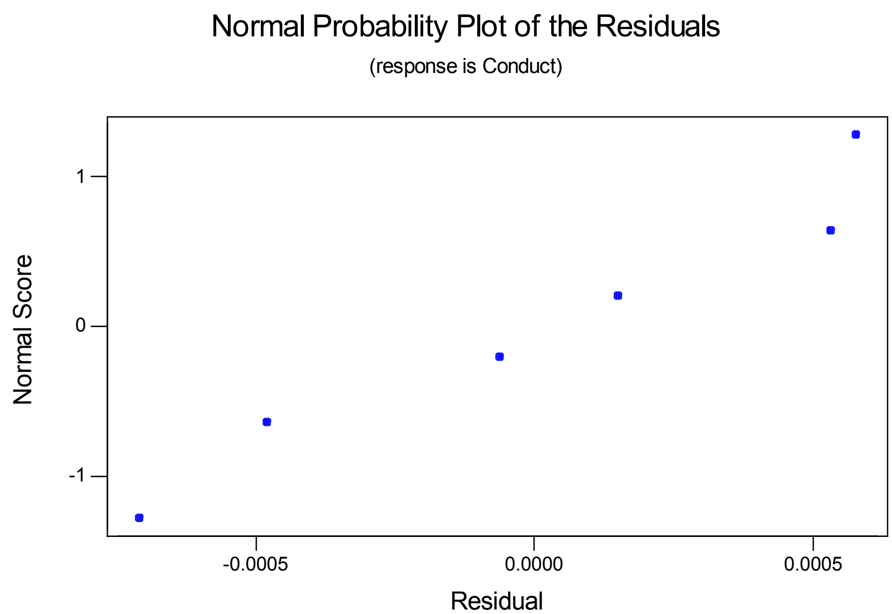
h) See the Minitab output in part a. Based on the analysis of variance, we can reject the null hypothesis and conclude that the regression is significant.

i) β_0: $0.024934 \pm 2.776(0.001786)$; 0.02, 0.03
β_1: $0.128522 \pm 2.776(0.007738)$; 0.107, 0.15

j) Residual Plots


Residuals Versus Density
(response is Conduct)
Residual
0.0005
0.0000
-0.0005
0.18
0.23
0.28
Density


Residuals Versus the Fitted Values
(response is Conduct)
Residual
0.0005
0.0000
-0.0005
0.050
0.055
0.060
Fitted Value



k) r = 0.993, P-value = 0, Therefore, we conclude there is a significant correlation between density and conductivity.

6-2. a)
```
The regression equation is
Usage = - 6.34 + 9.21 Temp

Predictor          Coef        StDev          T        P
Constant         -6.336        1.668      -3.80    0.003
Temp            9.20836      0.03377     272.64    0.000

S = 1.943        R-Sq = 100.0%    R-Sq(adj) = 100.0%

Analysis of Variance

Source             DF          SS          MS          F        P
Regression          1      280583      280583   74334.36    0.000
Residual Error     10          38           4
Total              11      280621
```

$\hat{y} = -6.34 + 9.21x$

b) -1.25010
-0.19519
-0.30208
-1.61751
0.49740
2.07214
1.71688
-0.02329
-2.55294
-1.15260
-1.25734
4.06464

c) SSE = 38 $\hat{\sigma}^2 = 4$

d) se($\hat{\beta}_1$) = 1.668, se($\hat{\beta}_0$) = 0.03377

e) SST = 280621
SSR = 280583, SSE = 38, and SSR + SSE = 280621
$\therefore$ SST = SSR + SSE

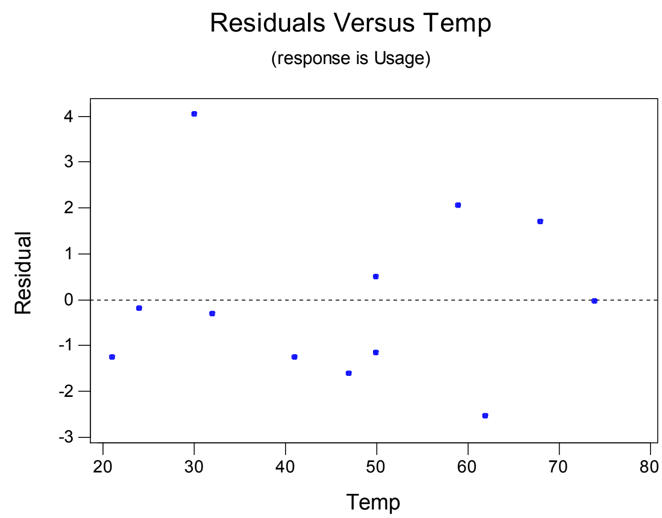
f) R^2 = 100%. This is interpreted as 100% of the total variability in Usage can be explained by the fitted regression model.

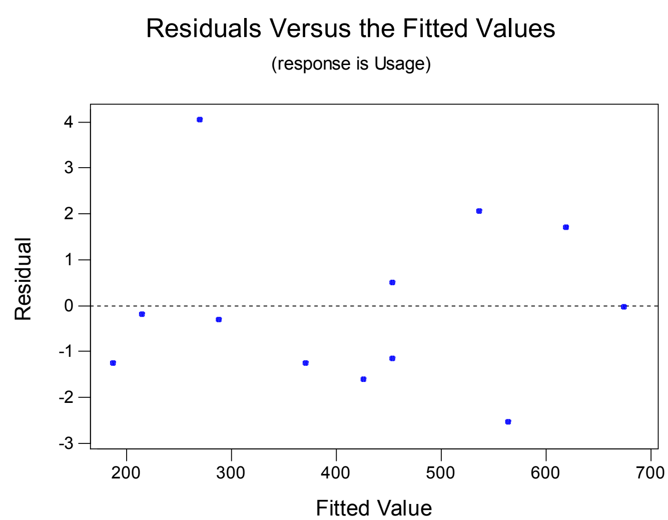
g) See the Minitab output given in part a. Based on the t-tests, we conclude that the slope and intercept are nonzero.

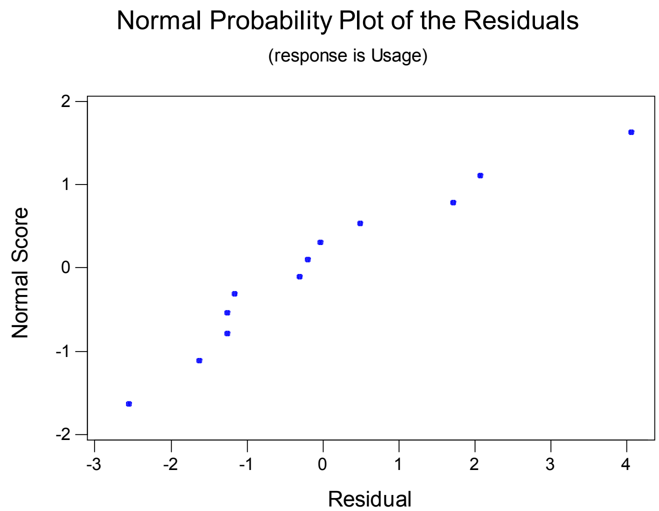
h) See the Minitab output in part a. Based on the analysis of variance, we can reject the null hypothesis and conclude that the regression is significant.

i) β_0: $-6.34 \pm 2.228(1.668)$; 2.62,10.06
β_1: $9.21 \pm 2.228(0.03377)$; 9.13,9.29

j) Residual Plots.







k) r = 1.000, P-value = 0, Therefore, we conclude there is a significant correlation between temperature and usage.

6-3. a)
```
The regression equation is
  Deflect = 0.393 + 0.00333 Temp

Predictor         Coef      SE Coef          T        P
Constant       0.39346      0.04258       9.24    0.000
Temp         0.0033285    0.0005815       5.72    0.000

S = 0.006473     R-Sq = 64.5%      R-Sq(adj) = 62.6%

Analysis of Variance
Source            DF           SS           MS         F        P
Regression         1    0.0013727    0.0013727     32.76    0.000
Residual Error    18    0.0007542    0.0000419
Total             19    0.0021270
```

$\hat{y} = 0.393 + 0.00333x$

b) -0.0054488
0.0072519
0.0065614
-0.0127685
0.0069249
-0.0004269
-0.0027627
-0.0044372
0.0002431
-0.0074196
0.0015643
0.0078738
0.0035833
-0.0077656
-0.0011131
-0.0024386
0.0105570
-0.0054342
-0.0014357
0.0068913

c) SSE = 0.0007542 $\hat{\sigma}^2 =$ 0.0000419

d) se($\hat{\beta}_1$) = 0.0005815, se($\hat{\beta}_0$) = 0.04258

e) SST = 0.0020402

SSR = 0.0013149, SSE = 0.0007253, and SSR + SSE = 0.0020402

$\therefore$ SST = SSR + SSE

f) R^2 = 64.5%. This is interpreted as 64.5% of the total variability in deflection can be explained by the fitted regression model.

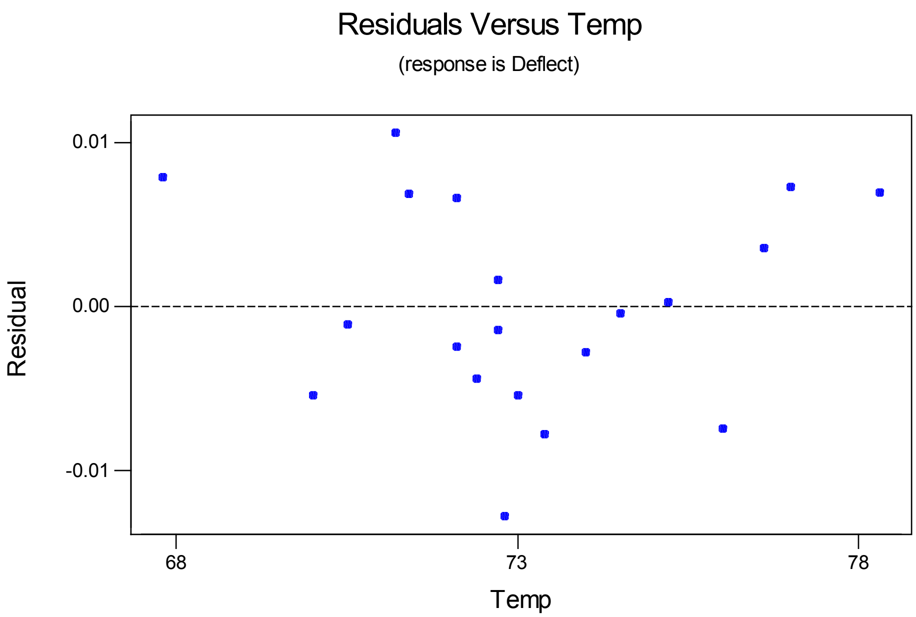
g) See the Minitab output given in part a. Based on the t-tests, we conclude that the slope and intercept are nonzero.

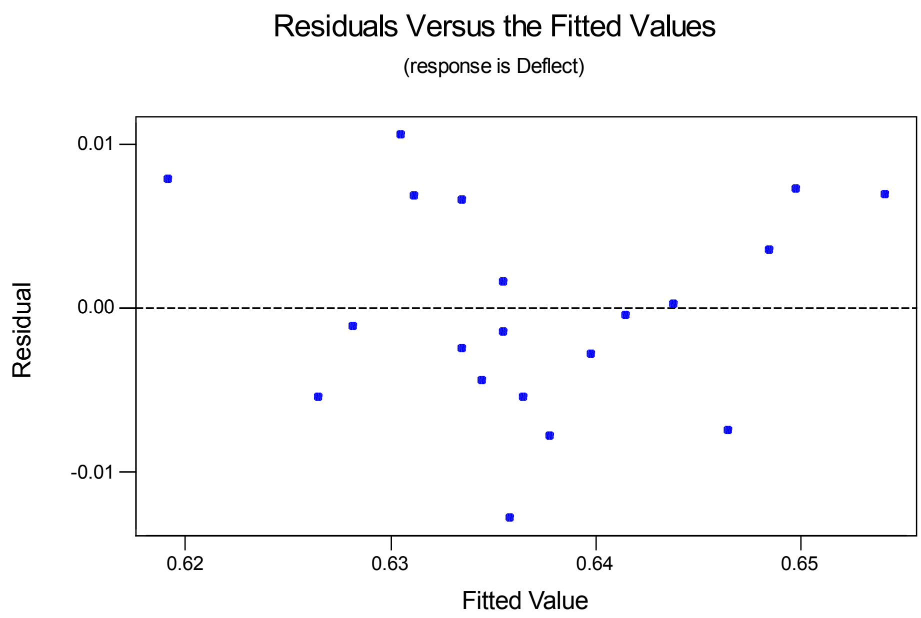
h) See the Minitab output in part a. Based on the analysis of variance, we can reject the null hypothesis and conclude that the regression is significant.

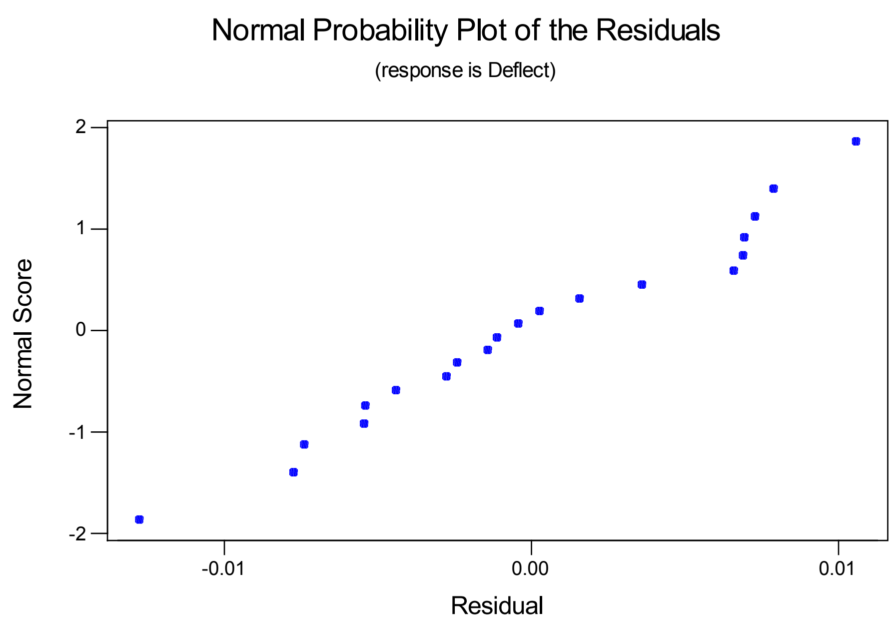
i) β_0: 0.39346 $\pm$ 2.101(0.04258); 0.304, 0.483

β_1: 0.0033285 $\pm$ 2.101(0.0005815); 0.00211, 0.00455

j) Residual Plots







k) r = 0.803, P-value = 0, Therefore, we conclude there is a significant correlation between temperature and deflection.

6-4. a)
```
The regression equation is
Turbidity = - 511 + 26.3 Temperature

Predictor        Coef       StDev          T        P
Constant       -510.7       228.2      -2.24    0.045
Temperat       26.308       9.178       2.87    0.014
```

```
S = 67.68       R-Sq = 40.6%     R-Sq(adj) = 35.7%

Analysis of Variance

Source            DF          SS          MS         F        P
Regression         1       37636       37636      8.22    0.014
Residual Error    12       54963        4580
Total             13       92599
```

$$\hat{y} = -511 + 26.3x$$

b) 33.253
-2.686
11.253
10.622
-2.607
-88.565
-69.041
-75.934
-91.980
-34.116
67.389
110.066
56.435
75.912

c) SSE = 54963 $\hat{\sigma}^2 = 4580$

d) se($\hat{\beta}_1$) = 228.2, se($\hat{\beta}_0$) = 9.178

e) SST = 92599
SSR = 37636, SSE = 54963, and SSR + SSE = 92599
$\therefore$ SST = SSR + SSE

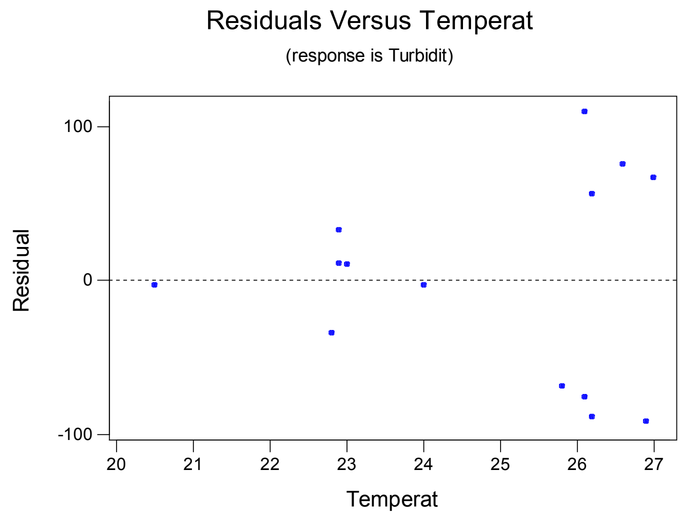
f) $R^2 = 40.6\%$. This is interpreted as 40.6% of the total variability in turbidity can be explained by the fitted regression model.

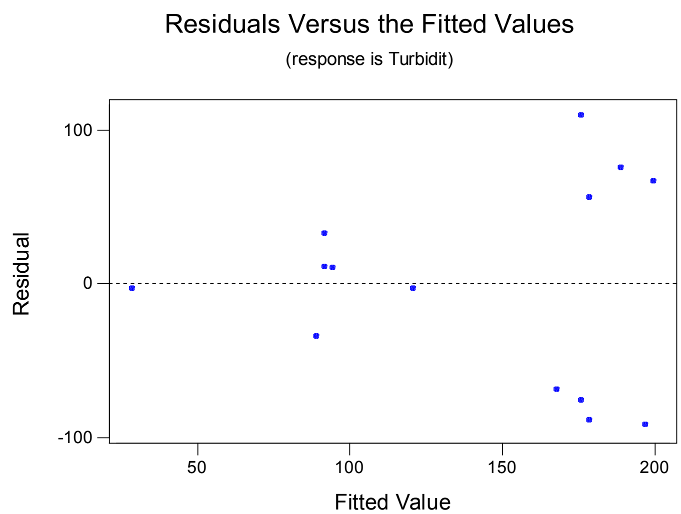
g) See the Minitab output given in part a. Based on the t-tests, we conclude that the slope and intercept are nonzero.

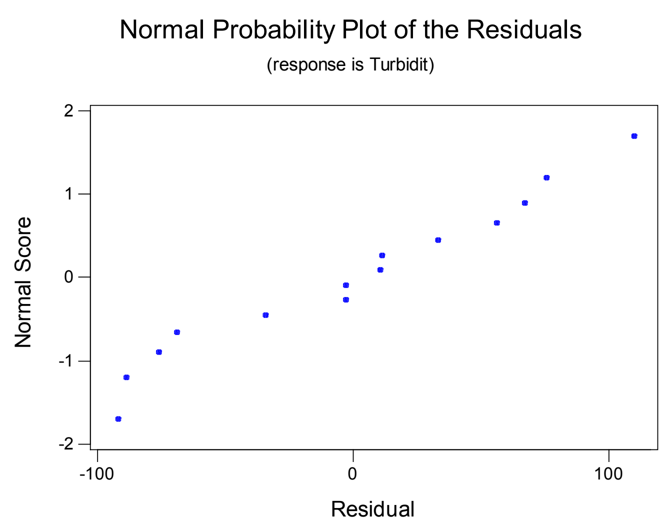
h) See the Minitab output in part a. Based on the analysis of variance, we can reject the null hypothesis and conclude that the regression is significant.

i) β_0: -510.7 ± 2.179(228.2); 13.45,1007.95
β_1: 26.3 ± 2.179(9.178); 6.30,46.30

j) Residual Plots.


Residuals Versus Temperat
(response is Turbidit)
Residual
Temperat
100
0
-100
20
21
22
23
24
25
26
27


Residuals Versus the Fitted Values
(response is Turbidit)
Residual
Fitted Value
100
0
-100
50
100
150
200


Normal Probability Plot of the Residuals
(response is Turbidit)
Normal Score
Residual
2
1
0
-1
-2
-100
0
100

k) r = 0.638, P-value = 0.014, Therefore, we conclude there is a significant correlation between temperature and turbidity.

6-5. a) The regression equation is

```
The regression equation is
permeability = 40.6 - 2.12 strength

Predictor        Coef     SE Coef          T        P
Constant      40.5536      0.7509      54.00    0.000
strength      -2.1232      0.2313      -9.18    0.000

S = 1.038       R-Sq = 86.6%     R-Sq(adj) = 85.6%

Analysis of Variance
Source            DF          SS          MS         F        P
Regression         1      90.759      90.759     84.28    0.000
Residual Error    13      13.999       1.077
Total             14     104.757
```

$\hat{y} = 40.6 - 2.12x$

b) -0.97179
0.00066
1.56517
0.35430
0.21735
0.99417
0.79921
0.83762
0.24199
-1.22252
-0.49351
-0.38411
-0.74715
-2.15947
0.96808

c) SSE = 13.999 $\hat{\sigma}^2 = 1.077$

d) se($\hat{\beta}_1$) = 0.2313, se($\hat{\beta}_0$) = 0.7509

e) SST = 104.757
SSR = 90.759, SSE = 13.999, and SSR + SSE = 104.757
$\therefore$ SST = SSR + SSE

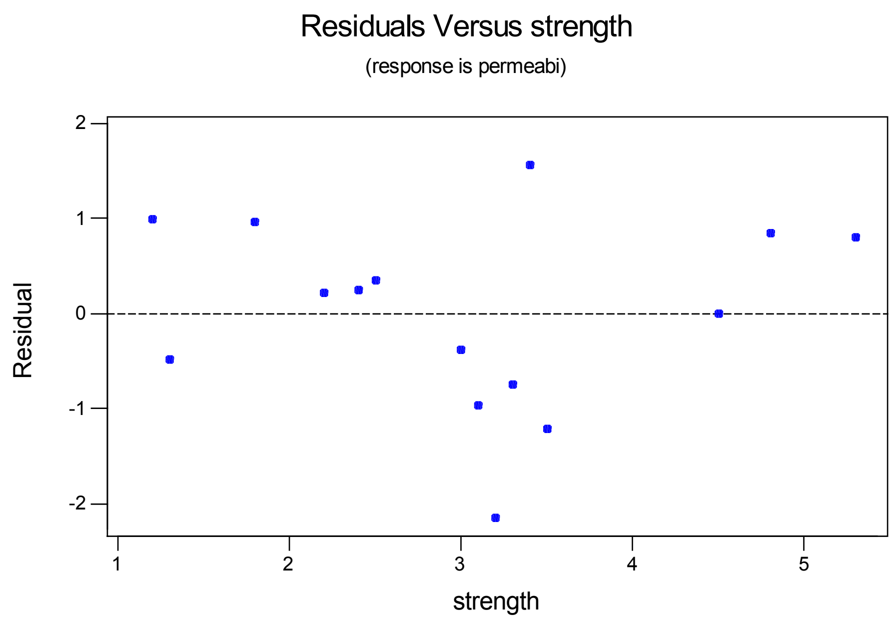
f) R^2 = 86.6%. This is interpreted as 86.6% of the total variability in permeability can be explained by the fitted regression model.

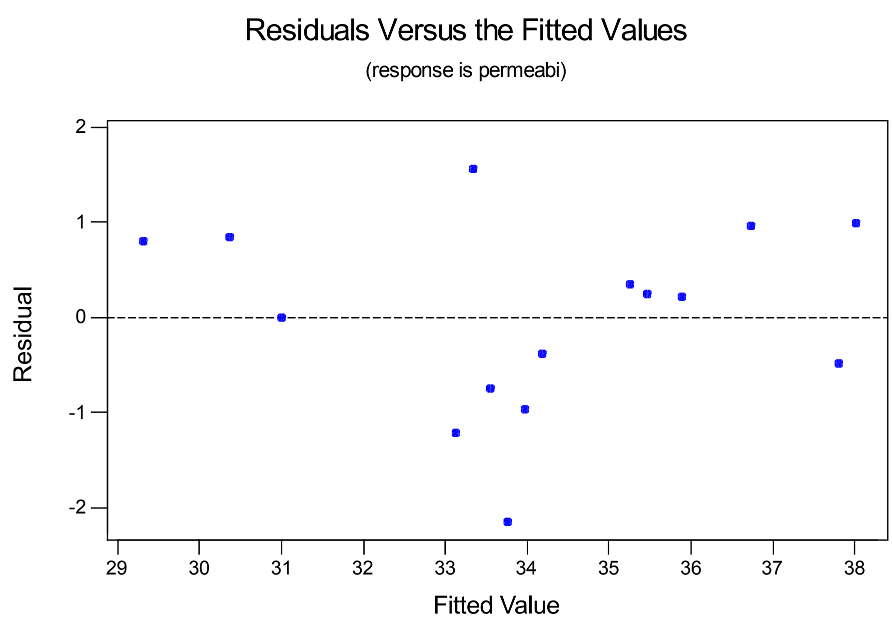
g) See the Minitab output given in part a. Based on the t-tests, we conclude that the slope and intercept are nonzero.

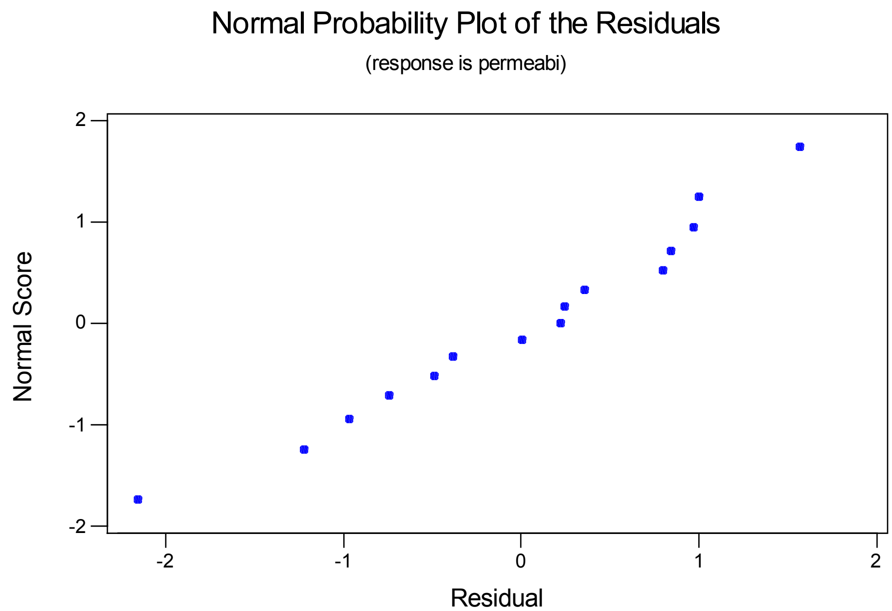
h) See the Minitab output in part a. Based on the analysis of variance, we can reject the null hypothesis and conclude that the regression is significant.

i) β_0: 40.5536 $\pm$ 2.16(0.7509); 38.93, 41.18
β_1: -2.1232 $\pm$ 2.16(0.2313); -2.62, -1.62

j) Residual Plots

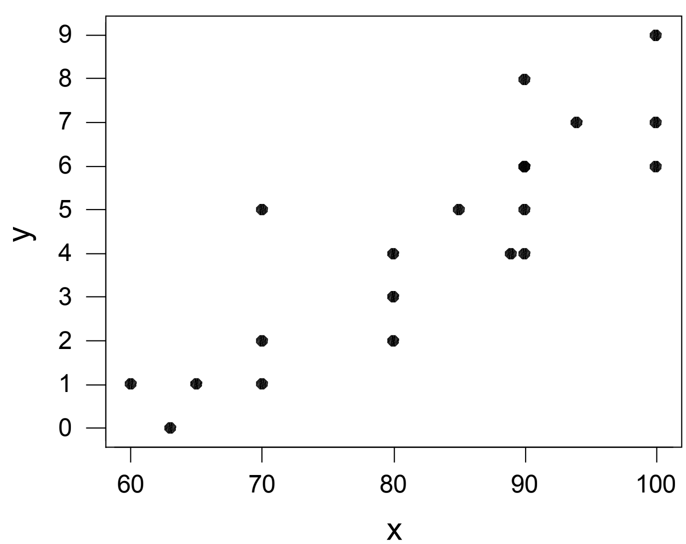

Residuals Versus strength
(response is permeabi)
Residual
strength
2
1
0
-1
-2
1
2
3
4
5


Residuals Versus the Fitted Values
(response is permeabi)
Residual
Fitted Value
2
1
0
-1
-2
29
30
31
32
33
34
35
36
37
38



k) r = -0.931, P-value = 0; Therefore, we conclude there is a significant correlation between strength and temperature.

6-6. a) The plot below implies that a simple linear regression seems reasonable in this situation.



b)
```
The regression equation is
y = - 10.1 + 0.174 x

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

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

Analysis of Variance

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

An estimate of $\hat{\sigma}^2 = 1.737$

c) $\hat{y} = -10.1 + 0.174(85) = 4.69$. The predicted mean rise in blood pressure level associated with a sound pressure level of 85 decibels is 4.69 millimeters of mercury.

6-7. a) 0.055137
b) (0.054460, 0.055813)
c) (0.053376, 0.056897)
d) The prediction interval is wider than the confidence interval.

6-8. a) 472.499
b) (471.183, 473.816)
c) (467.975, 477.024)
d) The prediction interval is wider than the confidence interval.

6-9. a) 36.095
b) (35.059, 37.131)
c) (32.802, 39.388)
d) The prediction interval is wider than the confidence interval.

6-10. a) 4.683
b) (4.055, 5.312)
c) (1.844, 7.523)
d) The prediction interval is wider than the confidence interval.

Section 6-3

6-11. a)
```
The regression equation is
y = 351 - 1.27 x1 - 0.154 x2

Predictor        Coef     SE Coef          T        P       VIF
Constant       350.99       74.75       4.70    0.018
x1             -1.272       1.169      -1.09    0.356       2.6
x2           -0.15390     0.08953      -1.72    0.184       2.6

S = 25.50       R-Sq = 86.2%     R-Sq(adj) = 77.0%

Analysis of Variance

Source            DF          SS          MS         F         P
Regression         2     12161.6      6080.8      9.35     0.051
Residual Error     3      1950.4       650.1
Total              5     14112.0

Source       DF      Seq SS
X1            1     10240.4
x2            1      1921.2
```

b) -24.9866
24.3075
11.8203
-20.4595
12.8296
-3.5113

c) SSE = 1950.4 $\hat{\sigma}^2 =$ 650.1

d) R-Sq = 86.2%, R-Sq(adj) = 77.0%; R-Sq(adj) is less than R-Sq. because the model contains terms that are not contributing significantly to the model. The adjusted R^2 value will penalize the user for adding terms to the model that are not significant.

e) See part a. Based on the p-value from the ANOVA table, the regression model is significant at the 0.10 level of significance.

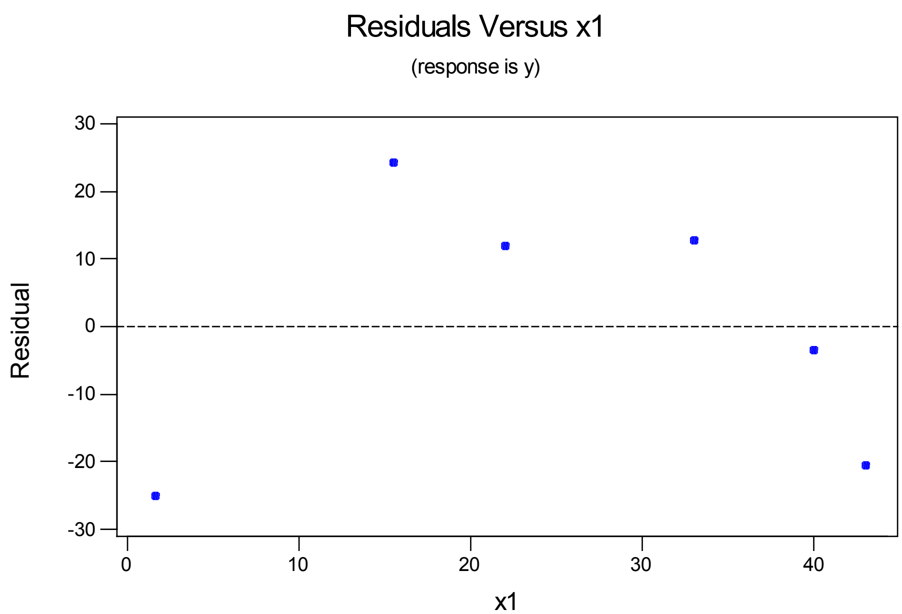
f) se($\hat{\beta}_0$) = 74.75, se($\hat{\beta}_1$) = 1.169, se($\hat{\beta}_2$) = 0.08953

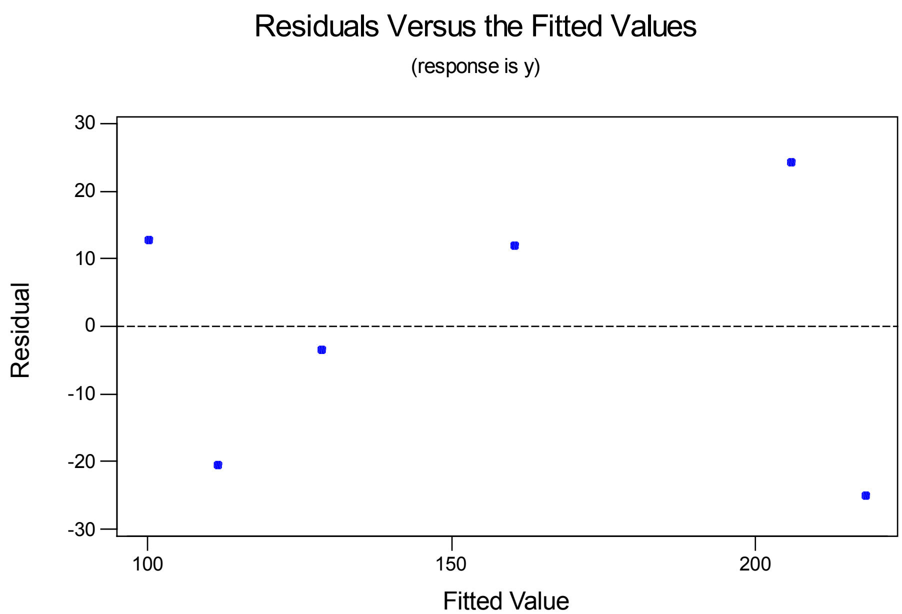
g) See part a. Based on the p-values for each coefficient, the regressors do not appear to be significant at the 0.05 level of significance.

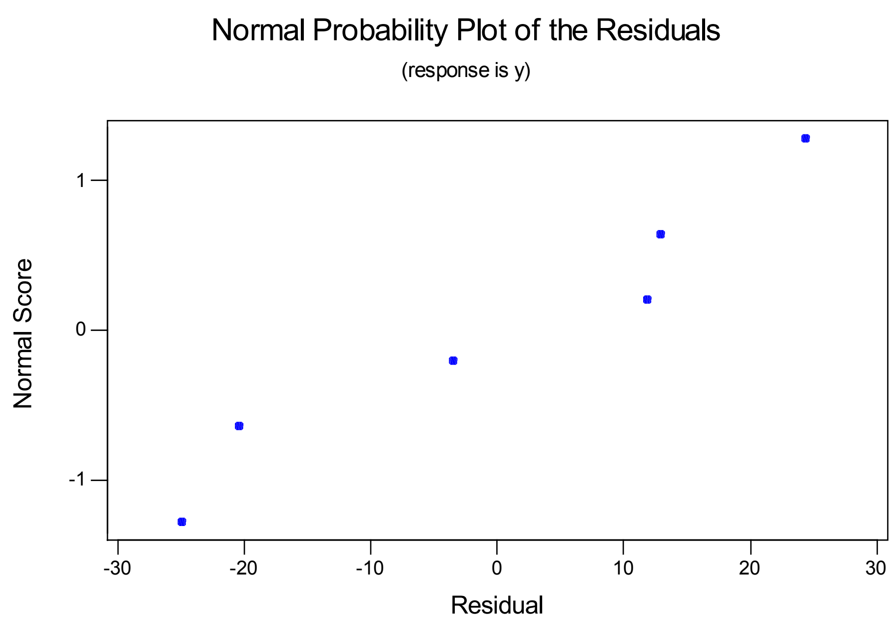
h) β_0: 350.99 ± 3.182(74.74); 113.17, 588.81
β_1: -1.272 ± 3.182(1.169); -4.99, 2.45
β_2 : -0.1539 ± 3.182(0.08953); -0.439, 0.131

i)

Obs	SRES1	COOK1
1	-1.65529	1.69265
2	1.35770	0.63183
3	0.51526	0.02083
4	-1.05590	0.27192
5	1.09375	1.48548
6	-0.18436	0.00898







j) The VIFs are 2.6. There is no indication of a problem with multicollinearity.

6-12. a)
```
The regression equation is
MPG-y = 38.4 - 0.00165 Weight-x1 - 0.0403 Horsepower-x2

Predictor         Coef        StDev          T        P       VIF
Constant        38.387        3.719      10.32    0.000
Weight-x     -0.001648     0.001325      -1.24    0.245       1.8
```

```
Horsepow    -0.040308    0.006299      -6.40    0.000        1.8

S = 2.135       R-Sq = 91.2%     R-Sq(adj) = 89.2%

Analysis of Variance

Source            DF          SS          MS         F        P
Regression         2      423.41      211.70     46.44    0.000
Residual Error     9       41.03        4.56
Total             11      464.44

Source        DF      Seq SS
Weight-x       1      236.70
Horsepow       1      186.71
```

b) 0.16464
-1.41661
2.33925
-1.31445
1.58629
-1.08273
1.12759
-3.77526
1.79269
0.63652
-2.14613
2.08818

c) SSE = 41.03 $\hat{\sigma}^2 =$ 4.56

d) R-Sq = 91.2%, R-Sq(adj) = 89.2% discuss
R-Sq(adj) is less than R-Sq.

e) see part a. Based on the p-value from the ANOVA table, the regression model is significant at the 0.10 level of significance.

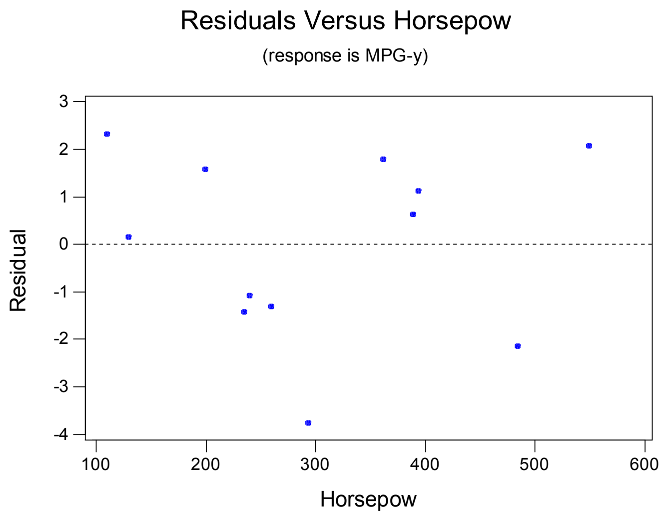
f) se($\hat{\beta}_0$) = 3.719, se($\hat{\beta}_1$) = 0.0013, se($\hat{\beta}_2$) = 0.0063

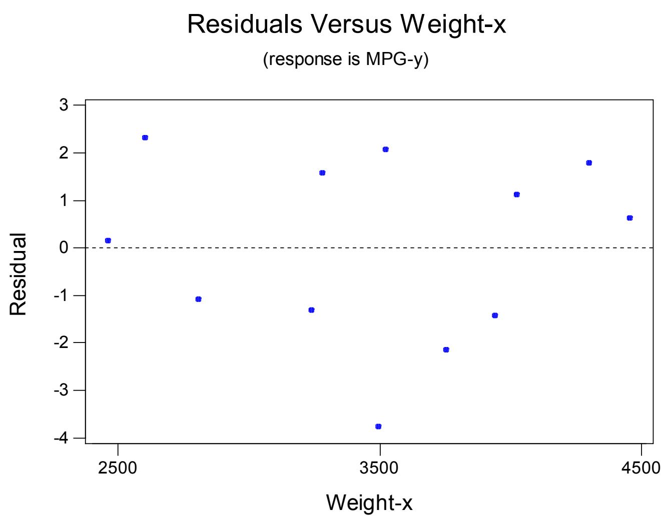
g) See part a. Based on the p-values for each coefficient, only x_1 does not appear to be significant at the 0.05 level of significance.

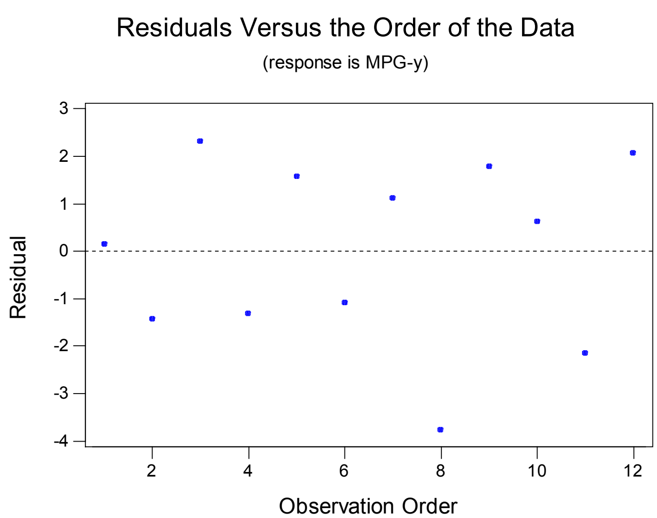
h) β_0: 38.387 ± 2.262(3.791); 29.8118, 46.9622
β_1: -0.0016 ± 2.262(0.0013); -0.0045, 0.0013
β_2 : -0.0403 ± 2.262(0.0063); -0.0546,-0.0260

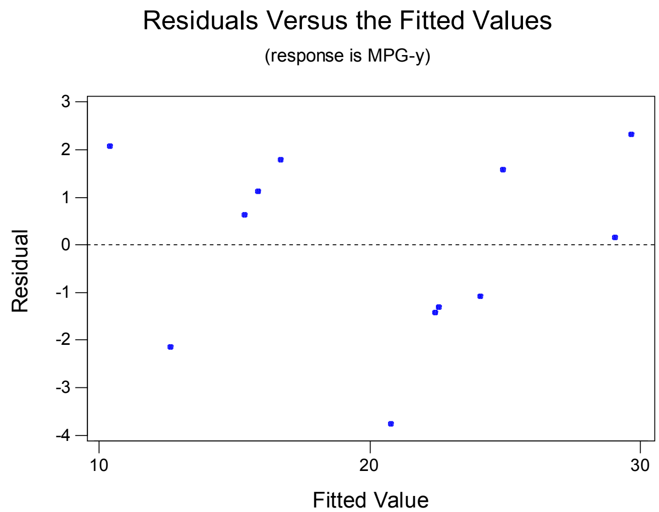
i)

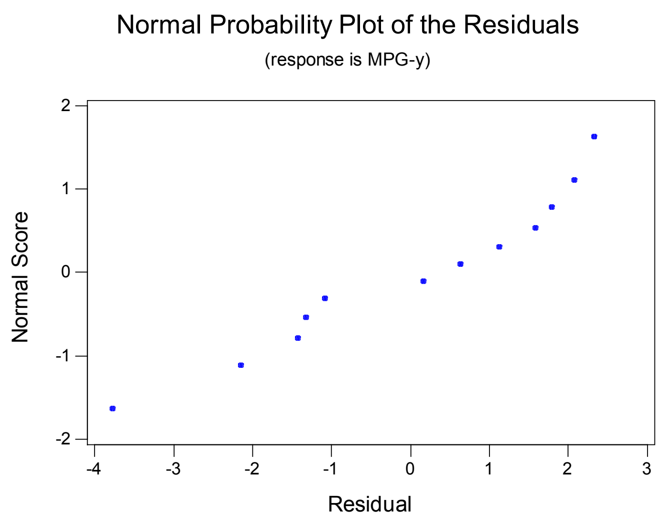
Obs	SRES2	COOK2
1	0.09379	0.00141
2	-0.78074	0.07819
3	1.31029	0.24630
4	-0.64814	0.01519
5	0.80217	0.03558
6	-0.56461	0.02547
7	0.57213	0.01895
8	-1.84772	0.10479
9	0.97075	0.10581
10	0.35819	0.01897
11	-1.18468	0.18208
12	1.53073	1.13238


Residuals Versus Horsepow
(response is MPG-y)
Residual
3
2
1
0
-1
-2
-3
-4
100
200
300
400
500
600
Horsepow


Residuals Versus Weight-x
(response is MPG-y)
Residual
3
2
1
0
-1
-2
-3
-4
2500
3500
4500
Weight-x


Residuals Versus the Order of the Data
(response is MPG-y)
Residual
3
2
1
0
-1
-2
-3
-4
2
4
6
8
10
12
Observation Order





j) The VIFs are 1.8. There is no indication of a problem with multicollinearity.

6-13. a) The regression equation is

```
y = - 103 + 0.605 x1 + 8.92 x2 + 1.44 x3 + 0.014 x4

Predictor          Coef      SE Coef          T        P        VIF
Constant         -102.7        207.9      -0.49    0.636
x1               0.6054       0.3689       1.64    0.145        2.3
x2                8.924        5.301       1.68    0.136        2.2
x3                1.437        2.392       0.60    0.567        1.3
x4               0.0136       0.7338       0.02    0.986        1.0

S = 15.58        R-Sq = 74.5%      R-Sq(adj) = 59.9%

Analysis of Variance
Source              DF          SS          MS         F        P
Regression           4      4957.2      1239.3      5.11    0.030
Residual Error       7      1699.0       242.7
Total               11      6656.3

Source        DF       Seq SS
x1             1       3758.9
```

```
x2              1      1109.4
x3              1        88.9
x4              1         0.1
```

b) -18.7580
1.8862
23.3109
-8.9565
9.1852
6.6436
4.8136
-0.1568
-17.8502
-12.9376
6.6216
6.1980

c) SSE = 1699.0 , $\hat{\sigma}^2 = 242.7$

d) R-Sq = 74.5%   R-Sq(adj) = 59.9%;   R-Sq(adj) is less than R-Sq. since there are terms in the model that are not significant.

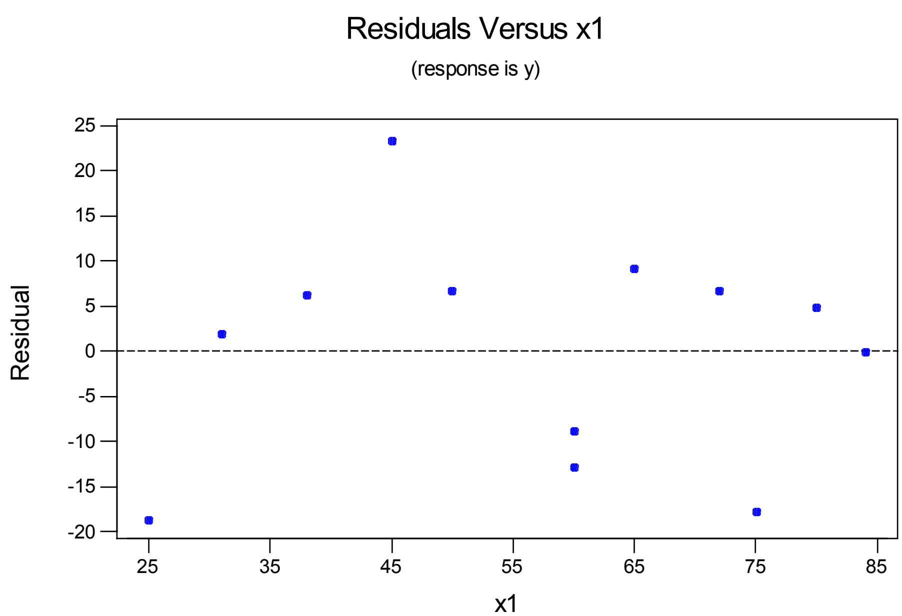
e) see part a. Based on the p-value from the ANOVA table, the regression model is significant at the 0.05 level of significance.

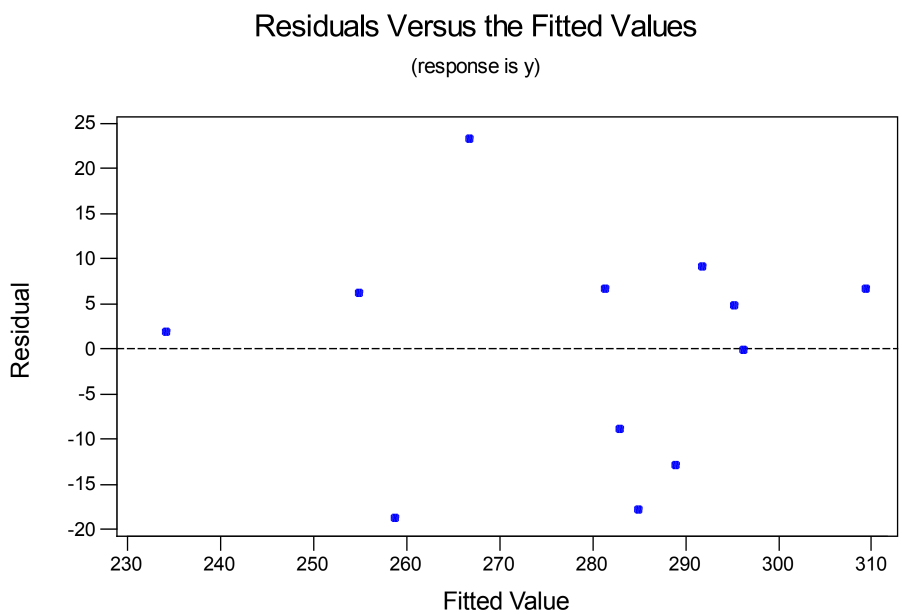
f) se($\hat{\beta}_0$) = 207.9, se($\hat{\beta}_1$) = 0.3689, se($\hat{\beta}_2$) = 5.301, se($\hat{\beta}_3$) = 2.392, se($\hat{\beta}_4$) = 0.7338

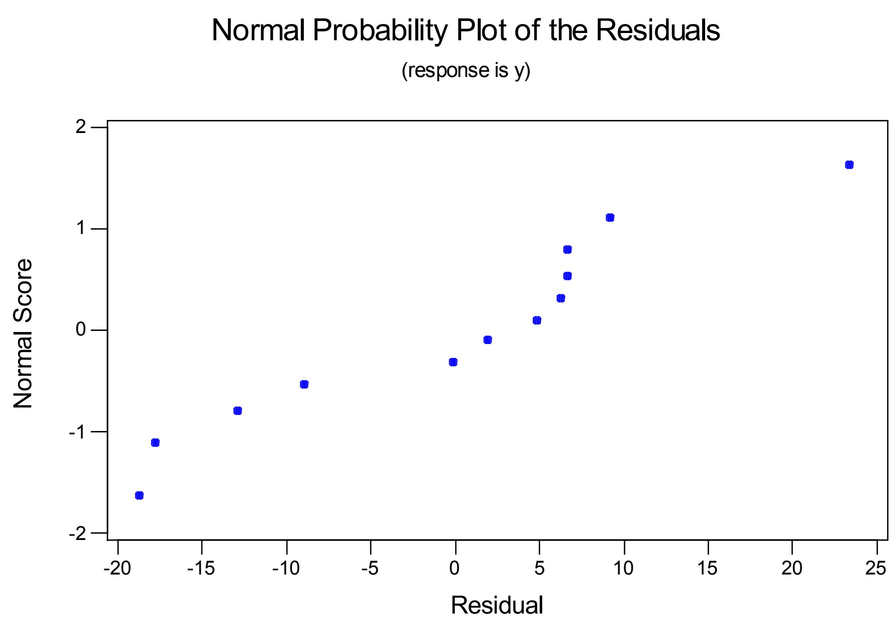
g) See part a. Based on the p-values for each coefficient, the regressors do not appear to be significant.

h) β_0: -102.7 ± 2.365(207.9); -594.38, 388.98
β_1: 0.6054 ± 2.365(0.3689); -0.267, 1.478
β_2 : 8.924 ± 2.365(5.301); -3.613, 21.461
β_3: 1.437 ± 2.365(2.392); -4.22, 7.094
β_4: 0.0136 ± 2.365(0.7338); -1.722, 1.75

i)







k) The VIFs are all less than 10, there is no indication of a problem with multicollinearity.

6-14. a)

```
The regression equation is
HFE = 47.2 - 9.74 Emitter + 0.428 Base + 18.2 EtoB

Predictor          Coef        StDev          T         P        VIF
Constant          47.17        49.58       0.95     0.356
Emitter          -9.735        3.692      -2.64     0.018        6.6
```

```
Base            0.4283      0.2239       1.91    0.074        2.5
EtoB            18.237       1.312      13.90    0.000        9.3

S = 3.480       R-Sq = 99.4%     R-Sq(adj) = 99.3%

Analysis of Variance

Source            DF          SS          MS         F         P
Regression         3       30532       10177    840.55     0.000
Residual Error    16         194          12
Total             19       30725

Source       DF      Seq SS
Emitter       1       23959
Base          1        4233
EtoB          1        2340
```

b) -0.90039
1.83266
-0.31872
-6.78384
-2.18117
-1.51602
1.90876
2.29305
2.01911
-5.96711
-2.21540
3.41999
3.16536
-0.57066
-1.96916
2.64163
-0.93420
6.68822
1.14227
-1.75438

c) SSE = 194 $\hat{\sigma}^2 = 12$

d) R-Sq = 99.4%, R-Sq(adj) = 99.3%. R-Sq(adj) is almost equal to R-Sq.

e) see part a. Based on the p-value from the ANOVA table, the regression model is significant at the 0.10 level of significance.

f) se($\hat{\beta}_0$) = 49.58, se($\hat{\beta}_E$) = 3.692, se($\hat{\beta}_B$) = 0.0039, se($\hat{\beta}_{EtoB}$) = 0.2239

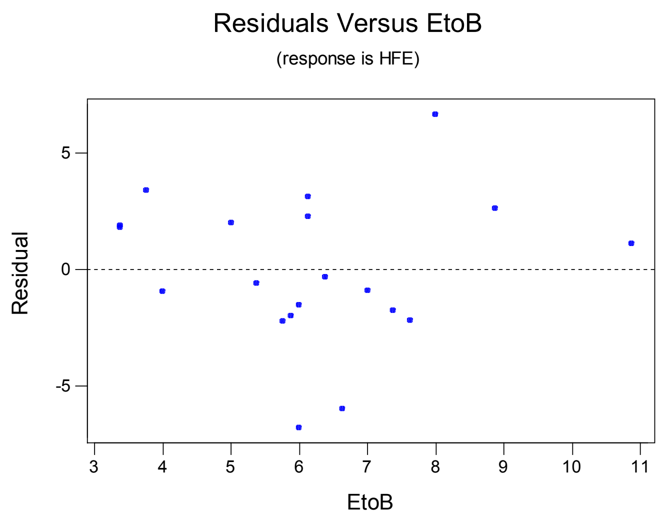
g) See part a. Based on the p-values for each coefficient, only the Emitter and EtoB appear to be significant at the 0.05 level of significance.

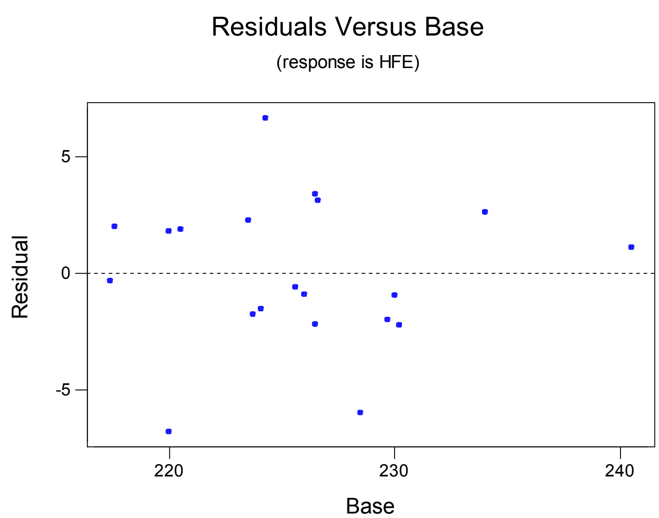
h) β_0: 47.17 ± 2.120(49.58); -57.9396,152.2796
β_E: -9.735 ± 2.120(3.692); -17.5620,-1.9080
β_B : 0.4283 ± 2.120(0.2239); -0.0464,0.9030
β_{EtoB} : 18.237 ± 2.120(1.312); 15.4556,21.0184

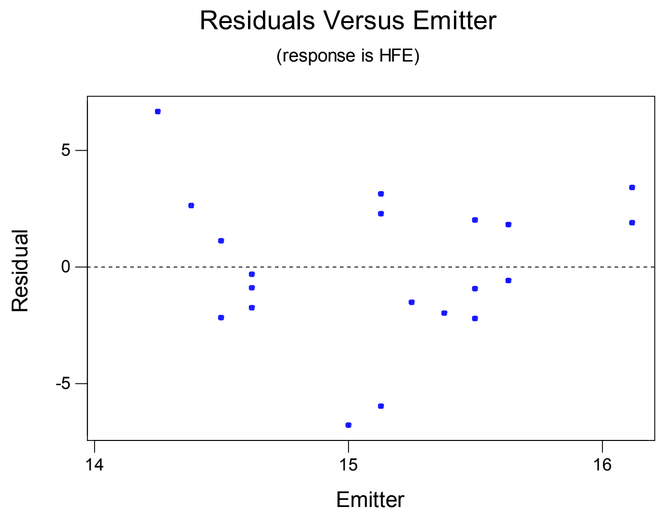
i)

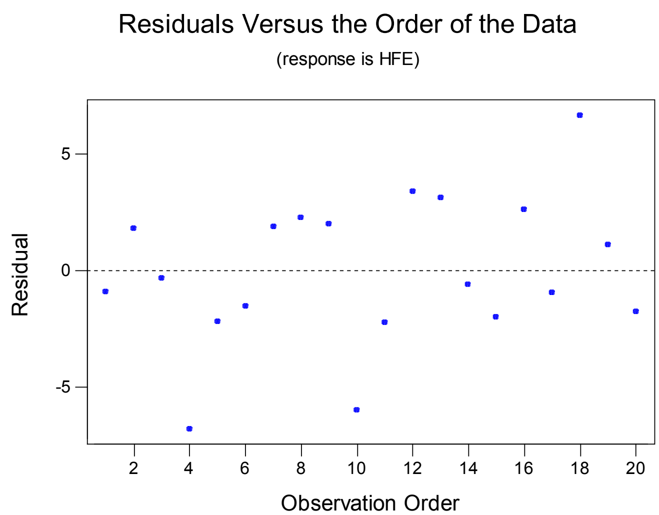
SRES	COOK
-0.27777	0.002938
0.59321	0.023627
-0.10665	0.001012
-2.08750	0.159577
-0.67100	0.016418
-0.45280	0.004106
0.63705	0.035375
0.68266	0.008518
0.67761	0.041745

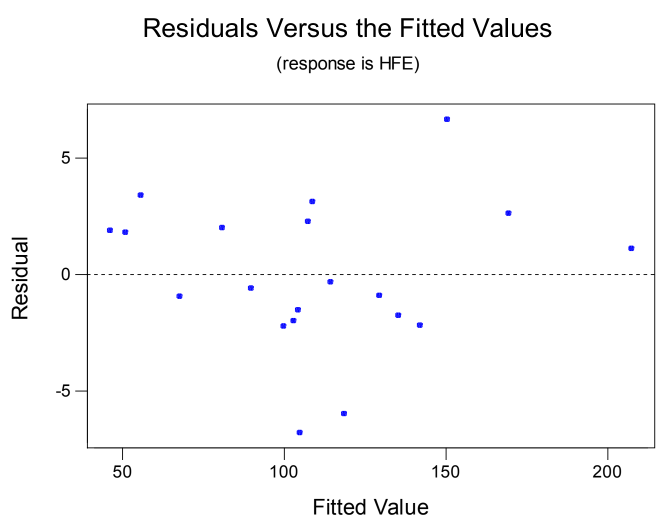
-1.77388	0.055072
-0.68134	0.016855
1.12620	0.099229
0.93545	0.012570
-0.17625	0.001204
-0.59731	0.010172
0.86378	0.054946
-0.40464	0.052052
2.16785	0.319631
0.54024	0.124644
-0.53676	0.009606

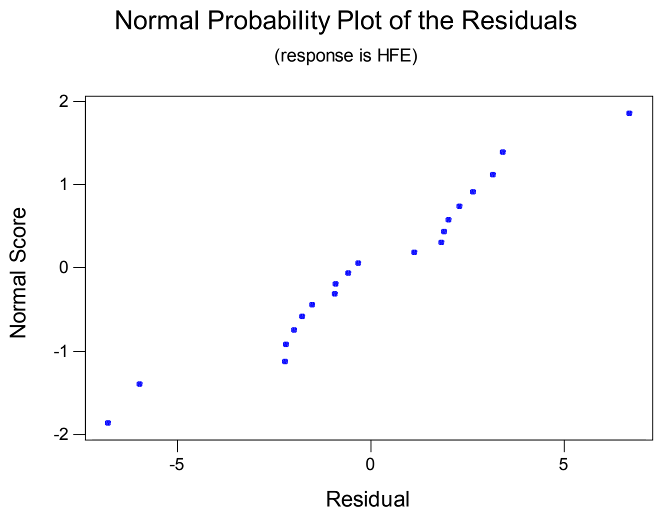





Residuals Versus Emitter
(response is HFE)
Residual
5
0
-5
14
15
16
Emitter


Residuals Versus the Order of the Data
(response is HFE)
Residual
5
0
-5
2
4
6
8
10
12
14
16
18
20
Observation Order


Residuals Versus the Fitted Values
(response is HFE)
Residual
5
0
-5
50
100
150
200
Fitted Value



j) All the VIFs are less than 10. There is no indication of a problem with multicollinearity.

6-15. a) 149.9
b) (85.1, 214.7)
c) (-12.5, 312.3)
d) The prediction interval is wider than the confidence interval.

6-16. a) 29.182
b) (26.631, 31.733)
c) (23.719, 34.644)
d) The prediction interval is wider than the confidence interval.

6-17. a) 287.56
b) (263.77, 311.35)
c) (243.69, 331.44)
d) The prediction interval is wider than the confidence interval.

6-18. a) 91.424. Note also that the values of x's are away from the center, particularly $x_2 = 220$.
b) (85.953, 96.895)
c) (83.249, 99.599)
d) The prediction interval is wider than the confidence interval.

Section 6-4

6-19. a)
```
The regression equation is
 y = 643 + 11.4 x1 - 0.933 x2 - 0.0106 x1x2 - 0.0272 x1^2 +0.000471 x2^2

 Predictor          Coef        StDev          T        P       VIF
 Constant        642.685        0.000          *        *
 x1              11.3862       0.0000          *        *    2675.3
 x2            -0.933346     0.000000          *        *    1283.4
 x1x2         -0.0106334    0.0000000          *        *    8342.1
 x1^2         -0.0271620    0.0000000          *        *     502.4
 x2^2         0.00047076   0.00000000          *        *    3301.5

 S = *
 Analysis of Variance

 Source              DF          SS          MS          F         P
 Regression           5    14112.00     2822.40          *         *
 Residual Error       0           *           *
 Total                5    14112.00
```

Source	DF	Seq SS
x1	1	10240.37
x2	1	1921.21
x1x2	1	827.86
x1^2	1	1056.39
x2^2	1	66.17

b) Because VIF's are much greater than 10, we would suspect that the multicollinearity present in the all second-order model.

c) Because SSE(Full Model) is not available the test statistic can not be computed.

6-20. a) The regression equation is

```
MPG = 53.5 - 0.0074 W - 0.101 HP +0.000001 W*HP +0.000001 W^2 +0.000079 HP^2
```

Predictor	Coef	StDev	T	P	VIF
Constant	53.49	27.17	1.97	0.096	
W	-0.00736	0.02063	-0.36	0.733	496.2
HP	-0.10098	0.08450	-1.19	0.277	368.3
W*HP	0.00000146	0.00002559	0.06	0.956	639.0
W^2	0.00000104	0.00000371	0.28	0.788	767.9
HP^2	0.00007854	0.00005398	1.46	0.196	65.7

```
S = 1.977       R-Sq = 95.0%     R-Sq(adj) = 90.7%
```

Analysis of Variance

Source	DF	SS	MS	F	P
Regression	5	440.988	88.198	22.57	0.001
Residual Error	6	23.449	3.908		
Total	11	464.438			

Source	DF	Seq SS
W	1	236.699
HP	1	186.707
W*HP	1	9.307
W^2	1	0.001
HP^2	1	8.274

b) Because VIF's are much greater than 10, we would suspect that the multicollinearity present in the all second-order model.

c) The regression equation is

```
MPG = 38.4 - 0.00165 W - 0.0403 HP
```

Predictor	Coef	StDev	T	P	VIF
Constant	38.387	3.719	10.32	0.000	
W	-0.001648	0.001325	-1.24	0.245	1.8
HP	-0.040308	0.006299	-6.40	0.000	1.8

```
S = 2.135       R-Sq = 91.2%     R-Sq(adj) = 89.2%
```

Analysis of Variance

Source	DF	SS	MS	F	P
Regression	2	423.41	211.70	46.44	0.000
Residual Error	9	41.03	4.56		
Total	11	464.44			

$$\therefore \; f_0 = \frac{[41.03 - 23.449]/(9-6)}{23.449/6} = \frac{5.8603}{3.908} = 1.4995$$

This results in P-value = P($f_{3,6}$ > 1.4995) = 0.3073. Because P-value is larger than 0.05, we fail to reject H_0 and conclude that the second-order terms do not significantly contribute to the model. Actually we know from the t-tests of the full second-order model in part a).

6-21. a) All possible regression.

```
Response is y

               Adj.                          x x x x
Vars   R-Sq   R-Sq    C-p         s    1 2 3 4

   1   64.5   60.9    1.7    15.381      X
   1   56.5   52.1    3.9    17.022    X
   2   73.1   67.2    1.4    14.095    X X
   2   64.6   56.8    3.7    16.173      X X
   3   74.5   64.9    3.0    14.573    X X X
   3   73.2   63.1    3.4    14.944    X X   X
   4   74.5   59.9    5.0    15.579    X X X X
```

b) Forward selection. Alpha-to-Enter: 0.25

```
 Response is    y     on  4 predictors, with N =   12


    Step          1        2
Constant   -90.1607   0.5287

x2             15.2     10.3
T-Value        4.26     2.36
P-Value       0.002    0.042

x1                      0.50
T-Value                 1.71
P-Value                0.122

S              15.4     14.1
R-Sq          64.46    73.14
R-Sq(adj)     60.90    67.17
C-p             1.7      1.4
```

c) Backward elimination. Alpha-to-Remove: 0.1

```
 Response is    y     on  4 predictors, with N =   12


    Step          1        2        3        4
Constant  -102.7132-101.6100   0.5287 -90.1607

x1             0.61     0.61     0.50
T-Value        1.64     1.76     1.71
P-Value       0.145    0.117    0.122

x2              8.9      8.9     10.3     15.2
T-Value        1.68     1.80     2.36     4.26
P-Value       0.136    0.109    0.042    0.002

x3              1.4      1.4
T-Value        0.60     0.65
P-Value       0.567    0.536

x4             0.01
T-Value        0.02
P-Value       0.986

S              15.6     14.6     14.1     15.4
R-Sq          74.47    74.47    73.14    64.46
R-Sq(adj)     59.89    64.90    67.17    60.90
C-p             5.0      3.0      1.4      1.7
```

d) Model contains only x_1 and x_2 seems to be the "best" among all, in the sense that it has high R-Sq(adj) and small Cp value.

6-22. a) All possible regression.

```
Response is y

                                                x x x
Vars   R-Sq    R-Sq(adj)        C-p         S   1 2 3

   1   99.1         99.0        7.1    3.9403       X
   1   78.0         76.8      542.9    19.389   X
   2   99.2         99.1        5.7    3.7418   X   X
   2   99.1         99.0        9.0    4.0433     X X
   3   99.4         99.3        4.0    3.4796   X X X
```

b) Forward selection. Alpha-to-Enter: 0.25

```
 Response is    y     on  3 predictors, with N =   20


    Step          1        2        3
Constant     -23.62    66.13    47.17

x3            21.51    20.12    18.24
T-Value       44.28    21.59    13.90
P-Value       0.000    0.000    0.000

x1                      -5.4     -9.7
T-Value                -1.72    -2.64
P-Value                0.103    0.018

x2                               0.43
T-Value                          1.91
P-Value                         0.074

S              3.94     3.74     3.48
R-Sq          99.09    99.23    99.37
R-Sq(adj)     99.04    99.13    99.25
C-p             7.1      5.7      4.0
```

c) Backward elimination. Alpha-to-Remove: 0.1

```
Response is    y     on  3 predictors, with N =   20


    Step          1
Constant      47.17

x1             -9.7
T-Value       -2.64
P-Value       0.018

x2             0.43
T-Value        1.91
P-Value       0.074

x3             18.2
T-Value       13.90
P-Value       0.000

S              3.48
R-Sq          99.37
R-Sq(adj)     99.25
C-p             4.0
```

d) Model contains all first-order terms seems to be the "best" among all, in the sense that it has highest R-Sq(adj) and smallest Cp value.

6-23. a) Note that $x_2 = 0$ if using tool type 302 and $x_2 = 1$ if using tool type 416.

```
The regression equation is
y = 14.3 + 0.141 x1 - 13.3 x2

Predictor        Coef     SE Coef          T        P
Constant       14.276       2.091       6.83    0.000
x1           0.141150    0.008833      15.98    0.000
x2           -13.2802      0.3029     -43.85    0.000

S = 0.6771      R-Sq = 99.2%     R-Sq(adj) = 99.1%

Analysis of Variance

Source            DF          SS          MS         F        P
Regression         2     1012.06      506.03   1103.69    0.000
Residual Error    17        7.79        0.46
Total             19     1019.85

Source       DF      Seq SS
x1            1      130.61
x2            1      881.45

Unusual Observations
Obs         x1          y         Fit      SE Fit    Residual    St Resid
 13        248     37.520      36.001       0.244       1.519       2.40R

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

The regression model is significant at 0.01.

b) Regression model for tool type 302 is

```
The regression equation is
y-Tool1 = 11.5 + 0.153 Tool1

Predictor        Coef     SE Coef          T        P
Constant       11.503       1.474       7.81    0.000
Tool1        0.152926    0.006237      24.52    0.000

S = 0.3749      R-Sq = 98.7%     R-Sq(adj) = 98.5%

Analysis of Variance

Source            DF          SS          MS         F        P
Regression         1      84.483      84.483    601.18    0.000
Residual Error     8       1.124       0.141
Total              9      85.607
```

The regression model for tool type 302 is significant at 0.01.

Regression model for tool type 416 is

```
The regression equation is
y-Tool2 = 5.41 + 0.122 Tool2

Predictor        Coef     SE Coef          T        P
Constant        5.409       4.051       1.33    0.219
Tool2         0.12236     0.01722       7.11    0.000

S = 0.8193      R-Sq = 86.3%     R-Sq(adj) = 84.6%

Analysis of Variance

Source            DF          SS          MS         F        P
```

Regression	1	33.889	33.889	50.49	0.000
Residual Error	8	5.370	0.671		
Total	9	39.258			

Unusual Observations

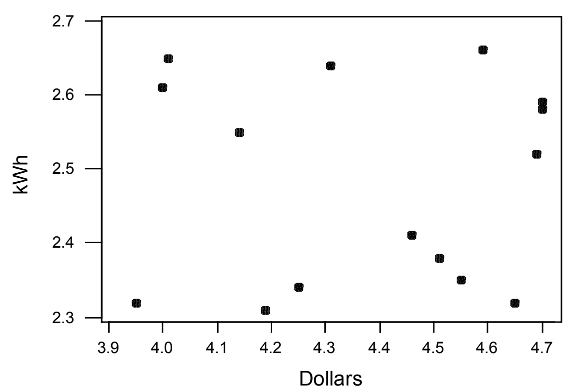
Obs	Tool2	y-Tool2	Fit	SE Fit	Residual	St Resid
3	248	37.520	35.753	0.345	1.767	2.38R

R denotes an observation with a large standardized residual

The regression model for tool type 416 is significant at 0.01.

Supplemental Exercises

6-24. a)



No, a straight line relationship does not seem plausible.

b) The regression equation is kWh = 2.43 + 0.012 Dollars

c) Analysis of Variance

Source	DF	SS	MS	F	P
Regression	1	0.00015	0.00015	0.01	0.933
Residual Error	13	0.26369	0.02028		
Total	14	0.26384			

2) $H_0: \beta_1 = 0$

3) $H_1: \beta_1 \neq 0$

4) $\alpha = 0.05$

5) The test statistic is

$$f_0 = \frac{SS_R / k}{SS_E / (n - p)}$$

6) Reject H_0 if $f_0 > f_{\alpha,1,8}$ where $f_{0.05,1,13} = 4.67$

7) Using the results from the ANOVA table

$$f_0 = \frac{0.00015/1}{0.2637/13} = 0.01$$

8) Since 0.01 < 4.67 do not reject H_0 and conclude that the regression model is not significant at $\alpha = 0.05$.
P-value > 0.10 (from the computer output the P-value is found to be 0.933

d)

Predictor	Coef	StDev	T	P
Constant	2.4299	0.6094	3.99	0.002
Dollars	0.0119	0.1389	0.09	0.933

$$0.0119 - t_{0.025,13}(0.1389) \le \beta_1 \le 0.0119 + t_{0.025,13}(0.1389)$$
$$0.0119 - 2.160(0.1389) \le \beta_1 \le 0.0119 + 2.160(0.1389)$$
$$-0.288 \le \beta_1 \le 0.312$$

e) 2) $H_0 : \beta_1 = 0$
3) $H_1 : \beta_1 \ne 0$
4) $\alpha = 0.05$
5) The test statistic is

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

6) Reject H_0 if $t_0 < -t_{\alpha/2,n-2}$ where $-t_{0.025,13} = -2.160$ or $t_0 > t_{0.025,13} = 2.160$
7) Using the results from the table above

$$t_0 = \frac{0.0119}{0.1389} = 0.0857$$

8) Since $-2.160 < 0.0857 < 2.160$ do not reject H_0 and conclude the slope is practically 0. Dollars is not a significant predictor of electrical usage at $\alpha = 0.05$.

f) 2) $H_0 : \beta_0 = 0$
3) $H_1 : \beta_0 \ne 0$
4) $\alpha = 0.05$
5) The test statistic is

$$t_0 = \frac{\hat{\beta}_0}{se(\hat{\beta}_0)}$$

6) Reject H_0 if $t_0 < -t_{\alpha/2,n-2}$ where $-t_{0.025,13} = -2.160$ or $t_0 > t_{0.025,13} = 2.160$
7) Using the results from the table above

$$t_0 = \frac{2.4299}{0.6094} = 3.987$$

8) Since $3.987 > 2.160$ reject H_0 and conclude the intercept is not zero at $\alpha = 0.05$.

6-25. Using $R^2 = 1 - \frac{SS_E}{S_{yy}}$,

$$F_0 = \frac{(n-2)\left(1 - \frac{SS_E}{S_{yy}}\right)}{\frac{SS_E}{S_{yy}}} = \frac{S_{yy} - SS_E}{\frac{SS_E}{n-2}} = \frac{S_{yy} - SS_E}{\hat{\sigma}^2}$$

Also,

$$SS_E = \sum (y_i - \hat\beta_0 - \hat\beta_1 x_i)^2$$
$$= \sum (y_i - \bar y - \hat\beta_1(x_i - \bar x))^2$$
$$= \sum (y_i - \bar y) + \hat\beta_1{}^2 \sum (x_i - \bar x)^2 - 2\hat\beta_1 \sum (y_i - \bar y)(x_i - \bar x)$$
$$= \sum (y_i - \bar y)^2 - \hat\beta_1{}^2 \sum (x_i - \bar x)^2$$
$$SS_T - SS_E = \hat\beta_1{}^2 \sum (x_i - \bar x)^2$$

Therefore, $F_0 = \dfrac{\hat\beta_1^2}{\hat\sigma^2 / S_{xx}} = t_0^2$

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

6-26. a) From Exercise 6-25, $f_0 = \dfrac{0.9(23)}{1-0.9} = 207$.

Reject $H_0\!:\beta_1 = 0$.

b) Because $f_{0.05,1,23} = 4.28$, H_0 is rejected if $\dfrac{23R^2}{1-R^2} > 4.28$.

That is, H_0 is rejected if

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

6-27. For two random variables X_1 and X_2 ,

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

Then,

$$V(Y_i - \hat Y_i) = V(Y_i) + V(\hat Y_i) - 2Cov(Y_i, \hat Y_i)$$
$$= \sigma^2 + V(\hat\beta_0 + \hat\beta_1 x_i) - 2\sigma^2\left[\frac{1}{n} + \frac{(x_i - \bar x)^2}{S_{xx}}\right]$$
$$= \sigma^2 + \sigma^2\left[\frac{1}{n} + \frac{(x_i - \bar x)^2}{S_{xx}}\right] - 2\sigma^2\left[\frac{1}{n} + \frac{(x_i - \bar x)^2}{S_{xx}}\right]$$
$$= \sigma^2\left[1 - \left(\frac{1}{n} + \frac{(x_i - \bar x)^2}{S_{xx}}\right)\right]$$

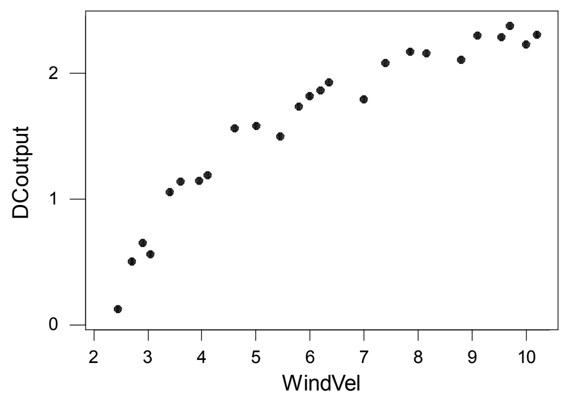
a) Because e_i is divided by an estimate of its standard error (when σ^2 is estimated by $\hat\sigma^2$), r_i has approximate unit standard deviation.

b) No, the term in brackets in the denominator is necessary for the standardized residuals to have unit standard deviation.

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

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

6-28. a)



Scatter diagram shows definite curvature. So, a higher order model may be appropriate or a transformation of variables.

b)
```
The regression equation is
DCoutput = 0.131 + 0.241 WindVel

Predictor        Coef      SE Coef          T        P
Constant       0.1309       0.1260       1.04    0.310
WindVel       0.24115      0.01905      12.66    0.000

S = 0.2361      R-Sq = 87.4%     R-Sq(adj) = 86.9%
```

$\hat{y} = 0.131 + 0.241x$

c)
```
Analysis of Variance
Source            DF          SS          MS         F        P
Regression         1      8.9296      8.9296    160.26    0.000
Residual Error    23      1.2816      0.0557
Total             24     10.2112
```

2) $H_0: \beta_1 = 0$

3) $H_1: \beta_1 \neq 0$

4) $\alpha = 0.05$

5) The test statistic is

$$f_0 = \frac{SS_R / k}{SS_E / (n - p)}$$

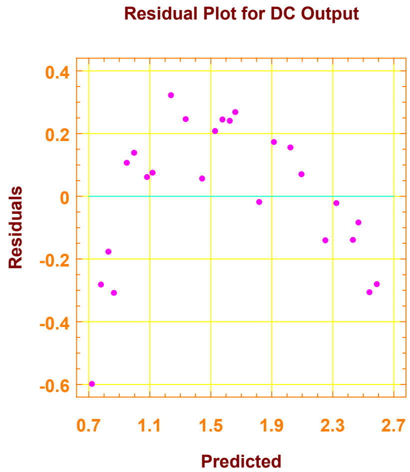
6) Reject H_0 if $f_0 > f_{\alpha,1,23}$ where $f_{0.05,1,23} = 4.28$

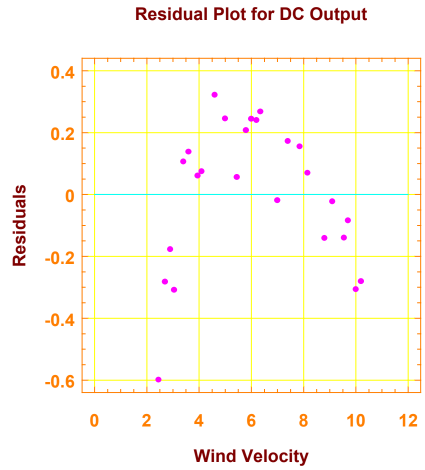
7) Using the results from the ANOVA table

$$f_0 = \frac{8.92961 / 1}{1.28157 / 23} = 160.257$$

8) Since 160.257 > 4.28 reject H_0 and conclude that the regression model is significant at $\alpha = 0.05$.

d)





Conclude plots indicate model inadequacy since the residual plots exhibit nonrandom patterns.

e) Examining the residual plots in part d), a transformation on the x-variable, y-variable, or both would be appropriate. A simple linear regression of y on the transformed variable 1/x may be satisfactory.

f) The following analysis employs the transformed variable, 1/x

```
The regression equation is
DCoutput = 2.98 - 6.93 1/WindVel

Predictor         Coef      SE Coef          T        P
Constant       2.97886      0.04490      66.34    0.000
1/WindVe       -6.9345       0.2064     -33.59    0.000

S = 0.09417      R-Sq = 98.0%      R-Sq(adj) = 97.9%
```

$\hat{y} = 2.98 - 6.93x^{*}$ where x* = 1/x

```
Analysis of Variance

Source            DF          SS          MS        F        P
Regression         1      10.007      10.007  1128.43    0.000
Residual Error    23       0.204       0.009
Total             24      10.211
```

2) $H_0: \beta_1 = 0$

3) $H_1: \beta_1 \neq 0$

4) $\alpha = 0.05$

5) The test statistic is

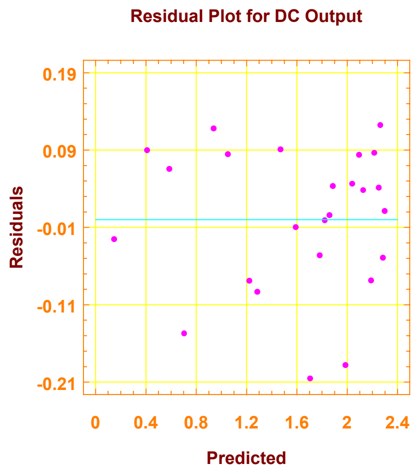
$$f_0 = \frac{SS_R / k}{SS_E / (n-p)}$$

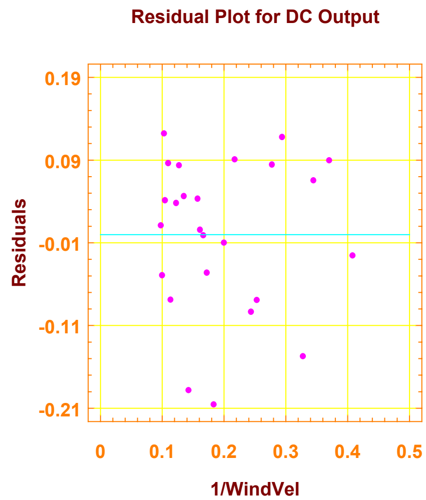
6) Reject H_0 if $f_0 > f_{\alpha,1,23}$ where $f_{0.05,1,23} = 4.28$

7) Using the results from the ANOVA table

$$f_0 = \frac{10.0072 / 1}{0.203970 / 23} = 1128.43$$

8) Since 1128.43 > 4.28 reject H_0 and conclude that the regression model is significant at $\alpha = 0.05$.





Using 1/WindVel

Conclude from the random appearance of the residuals in the plots and significance of regression that the model is adequate. The transformation, 1/(Wind Velocity), appears to be satisfactory as a regressor of Output.

6-29. 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

6-30. a)
```
The regression equation is
y = 3829 - 0.215 x3 + 21.2 x4 + 1.66 x5

Predictor         Coef     SE Coef         T        P
Constant          3829        2262      1.69    0.099
x3             -0.2149      0.1088     -1.97    0.056
x4             21.2134      0.9050     23.44    0.000
x5              1.6566      0.5502      3.01    0.005

S = 43.66       R-Sq = 99.3%     R-Sq(adj)  = 99.3%
```

$\hat{y} = 3829.26 - 0.215x_3 + 21.213x_4 + 1.657x_5$

b)
```
Analysis of Variance

Source            DF        SS         MS        F        P
Regression         3   9863398    3287799  1724.42    0.000
Residual Error    36     68638       1907
Total             39   9932036
```
2) $H_0: \beta_3 = \beta_4 = \beta_5 = 0$
3) $H_1: \beta_j \neq 0$ for at least one j
4) $\alpha = 0.01$
5) The test statistic is

$$f_0 = \frac{SS_R / k}{SS_E / (n-p)}$$

6) Reject H_0 if $f_0 > f_{\alpha,3,36}$ where $f_{0.01,3,36} = 4.38$

7) Using the results from the ANOVA table

$$f_0 = \frac{9863398/3}{68638.2/36} = 1724.42$$

8) Since 1724.42 > 4.38 reject H_0 and conclude that the regression model is significant at $\alpha = 0.01$.

P-value < 0.00001

c) All at $\alpha = 0.01$ $\quad t_{0.005,36} = 2.72$

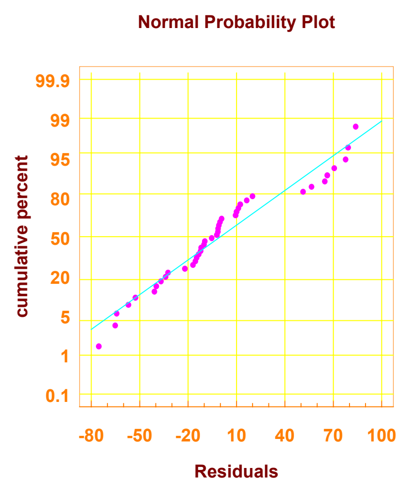
$H_0{:}\beta_3 = 0$	$H_0{:}\beta_4 = 0$	$H_0{:}\beta_5 = 0$
$H_1{:}\beta_3 \neq 0$	$H_1{:}\beta_4 \neq 0$	$H_1{:}\beta_5 \neq 0$
$t_0 = -1.97$	$t_0 = 23.44$	$t_0 = 3.01$
$\lvert t_0 \rvert \not> t_{\alpha/2,36}$	$\lvert t_0 \rvert > t_{\alpha/2,36}$	$\lvert t_0 \rvert > t_{\alpha/2,36}$
Do not reject H_0	Reject H_0	Reject H_0

Do not need x_3 term in the model at $\alpha = 0.01$.

d) $R^2 = 0.993$ $\quad R^2_{adj} = 0.9925$

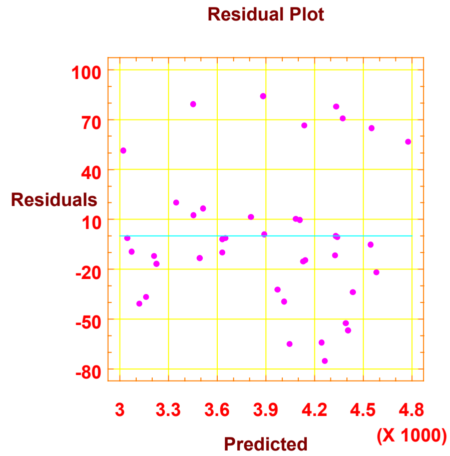
The slight decrease in R^2_{adj} may be reflective of the insignificant x_3 term.

e)



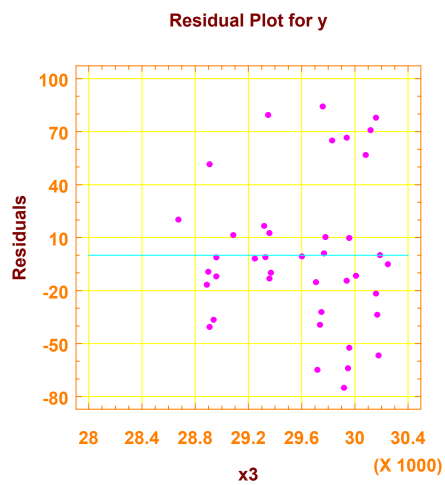
Normality assumption appears reasonable. This is evident by the fact the residuals fall along a straight line.

f)



Plot is satisfactory. There does not appear to be a nonrandom pattern in the residual vs. predicted plot.

g)



Slight indication that variance increases as x_3 increases. This is evident by the "fanning out" appearance of the residuals.

h) Using the equation found in part a):

$$\hat{y} = 3829.26 - 0.215(1670) + 21.213(170) + 1.657(1589) = 9709.39$$

6-31. a) The regression equation is

```
y* = 19.7 - 1.27 x3* + 0.00541 x4 +0.000408 x5

Predictor        Coef       SE Coef          T        P
Constant       19.690         9.587       2.05    0.047
x3*           -1.2673        0.9594      -1.32    0.195
x4          0.0054140     0.0002711      19.97    0.000
x5          0.0004079     0.0001645       2.48    0.018

S = 0.01314     R-Sq = 99.1%     R-Sq(adj) = 99.0%

Analysis of Variance

Source             DF            SS            MS           F         P
```

Regression	3	0.68611	0.22870	1323.62	0.000
Residual Error	36	0.00622	0.00017		
Total	39	0.69233			

2) $H_0: \beta_3^* = \beta_4 = \beta_5 = 0$

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

4) $\alpha = 0.01$

5) The test statistic is

$$f_0 = \frac{SS_R / k}{SS_E / (n-p)}$$

6) Reject H_0 if $f_0 > f_{\alpha,3,36}$ where $f_{0.01,3,36} = 4.38$

7) Using the results from the ANOVA table

$$f_0 = \frac{0.686112 / 3}{0.00622033 / 36} = 1323.62$$

8) Since 1323.62 > 4.38 reject H_0 and conclude that the regression model is significant at $\alpha = 0.01$.

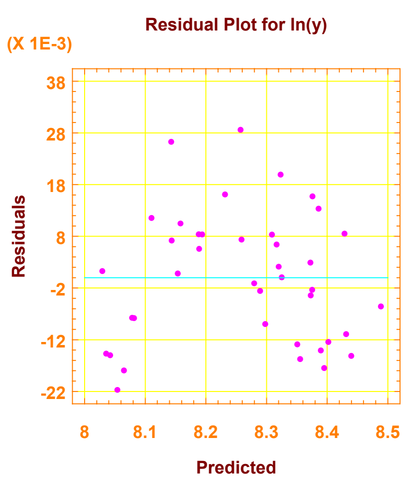
P-value < 0.00001

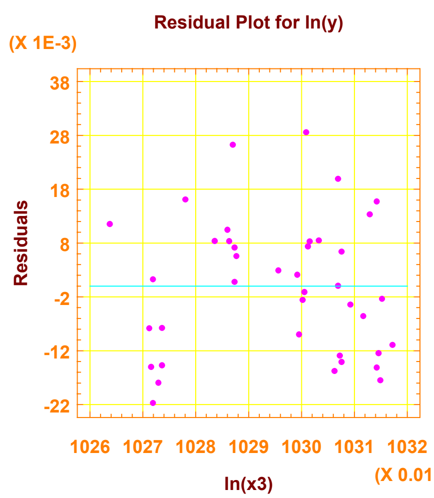
b) $\alpha = 0.01$ $\quad t_{.005,36} = 2.72$

$H_0: \beta_3^* = 0$	$H_0: \beta_4 = 0$	$H_0: \beta_5 = 0$
$H_1: \beta_3^* \neq 0$	$H_1: \beta_4 \neq 0$	$H_1: \beta_5 \neq 0$
$t_0 = -1.32$	$t_0 = 19.97$	$t_0 = 2.48$
$\lvert t_0 \rvert \not> t_{\alpha/2,36}$	$\lvert t_0 \rvert > t_{\alpha/2,36}$	$\lvert t_0 \rvert \not> t_{\alpha/2,36}$
Do not reject H_0	Reject H_0	Do not reject H_0

β_3 : Do not reject H_0 and conclude that $\ln(x_3)$ is not a significant regressor in the model at $\alpha = 0.01$.
β_4 : Reject H_0 and conclude that x_4 is a significant regressor in the model at $\alpha = 0.01$.
β_5 : Do not reject H_0 and conclude that x_5 is not a significant regressor in the model at $\alpha = 0.01$.

c)





Curvature is evident in the residuals plots from this model, whereas non-stable variance was evident in previous model.

6-32. a)
```
The regression equation is
y = - 1709 + 2.02 x -0.000593 x^2

Predictor        Coef     SE Coef          T        P
Constant      -1709.4       244.8      -6.98    0.000
x              2.0229      0.2798       7.23    0.000
x^2       -0.00059293  0.00007994      -7.42    0.000

S = 0.2101      R-Sq = 98.8%     R-Sq(adj) = 98.5%
```

$$\hat{y} = -1709.4054 + 2.0229x - 0.0006x^2$$

b)
```
Analysis of Variance

Source            DF         SS         MS        F        P
Regression         2     26.487     13.244   300.11    0.000
Residual Error     7      0.309      0.044
Total              9     26.796
```

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

3) $H_1: \beta_j \neq 0$

4) $\alpha = 0.05$

5) The test statistic is

$$f_0 = \frac{SS_R / k}{SS_E / (n-p)}$$

6) Reject H_0 if $f_0 > f_{\alpha,3,36}$ where $f_{0.05,2,7} = 4.74$

7) Using the results from the ANOVA table

$$f_0 = \frac{26.4871/2}{0.308899/7} = 300.11$$

8) Since 300.11 > 4.74 reject H_0 and conclude that the regression model is significant at $\alpha = 0.05$.

c) 2) $H_0: \beta_{11} = 0$

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

4) $\alpha = 0.05$

5) The test statistic is

$$t_0 = \frac{\hat{\beta}_1 - \beta_{1,0}}{se(\hat{\beta}_1)}$$

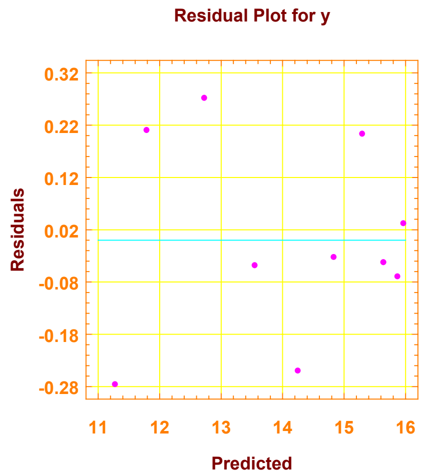
6) Reject H_0 if $t_0 < -t_{\alpha/2,n-2}$ where $-t_{0.025,7} = -2.365$ or $t_0 > t_{0.025,7} = 2.365$

7) Using the results from the table given in part a)

$$t_0 = \frac{-0.000593 - 0}{0.00008} = -7.417$$

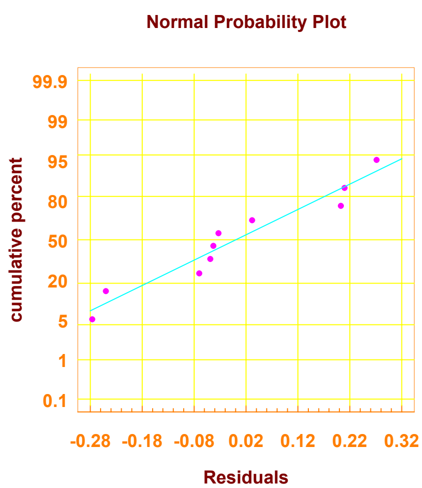
8) Since $-7.417 < -2.365$ reject H_0 and conclude the quadratic term contributes significantly to the model at $\alpha = 0.05$.

d)



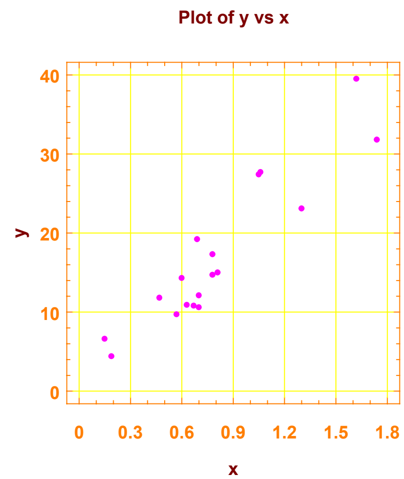
Some indication of nonconstant variance. This is evident by the widening of the residuals as the predicted value increases.

e)



Normality assumption is reasonable. This is evident by the fact that the residuals fall along a straight line.

6- 33. a)



The simple linear regression model seems appropriate. This is evident by the fact that the data fall along a straight line as x and y increase.

b) The regression equation is
y = 0.47 + 20.6 x

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

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

Analysis of Variance

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

$\hat{y} = 0.4705 + 20.5673x$

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

d) $\hat{y} = 0.470467 + 20.5673(0.47) = 10.1371$

$e_i = y_i - \hat{y}_i = 11.8 - 10.1371 = 1.6629$

e) The least squares estimate minimizes $\sum (y_i - \hat{\beta} x_i)^2$.

Upon setting the derivative equal to zero, we obtain

$2\sum (y_i - \hat{\beta} x_i)(-x_i) = 2[\sum y_i x_i - \hat{\beta} \sum x_i{}^2] = 0$

Solving for $\hat{\beta}$,

$\hat{\beta} = \frac{\sum y_i x_i}{\sum x_i{}^2}$.

f) The regression equation is
y = 21.0 x

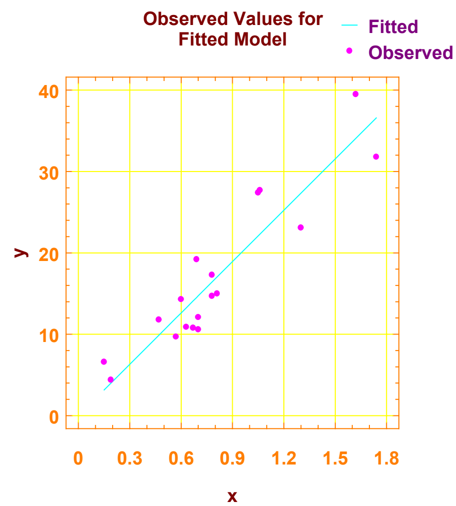
Predictor	Coef	SE Coef	T	P
Noconstant				
x	21.0315	0.9418	22.33	0.000

S = 3.612

```
Analysis of Variance

Source            DF        SS         MS         F         P
Regression         1      6505.4     6505.4    498.69     0.000
Residual Error    17       221.8       13.0
Total             18      6727.1
```

$\hat{y} = 21.031461x$



Examining the plot, the model seems very appropriate - possibly a better fit.

6-34. a) $\dfrac{\hat{\beta}}{\sqrt{\dfrac{\hat{\sigma}^2}{\sum x_i^2}}}$ has a t distribution with n – 1 degrees of freedom.

b) From Exercise 6-33f,

$\hat{\beta} = 21.031461, \quad \hat{\sigma} = 3.611768$, and

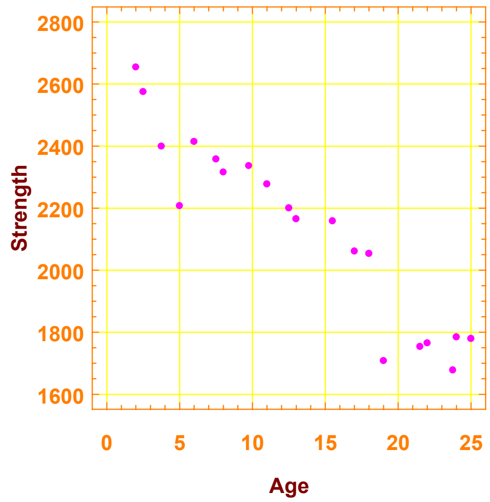
$\sum x_i^2 = 14.7073$.

Therefore,

$t_0 = \dfrac{21.031461}{\sqrt{\dfrac{3.611768}{14.7073}}} = 22.3314$ and $H_0 : \beta = 0$ is rejected at usual α values.

6-35. a)

Plot of Strength vs Age



A straight-line regression model seems appropriate. This is evident by the fact that the data fall along a straight line as age and strength increase.

b) The regression equation is
strength = 2625 - 37.0 age

Predictor	Coef	SE Coef	T	P
Constant	2625.39	45.35	57.90	0.000
age	-36.962	2.967	-12.46	0.000

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

Analysis of Variance

Source	DF	SS	MS	F	P
Regression	1	1522819	1522819	155.21	0.000
Residual Error	18	176602	9811		
Total	19	1699421			

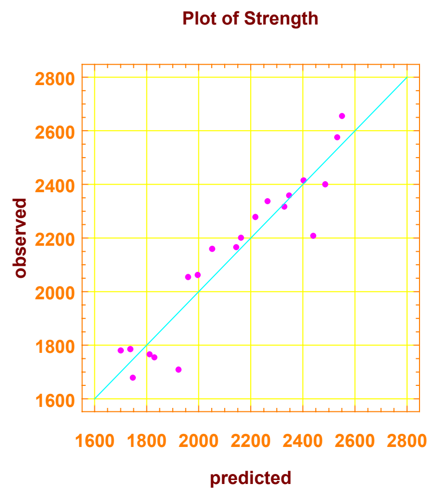
$\hat{y} = 2625.39 - 36.9618x$

c) $\hat{y} = 2625.39 - 36.9618(20) = 1886.154$

d)

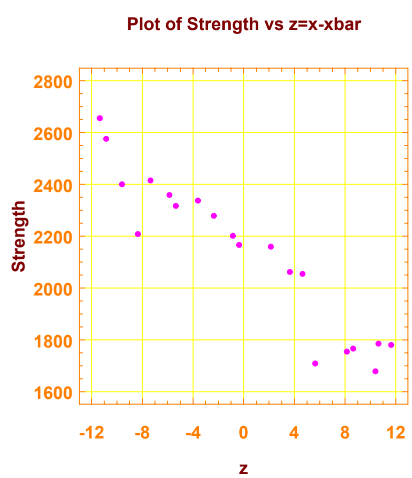
Obs	age	strength	Fit	SE Fit	Residual	St Resid
1	15.5	2158.7	2052.5	23.1	106.2	1.10
2	23.8	1678.2	1747.5	38.0	-69.4	-0.76
3	8.0	2316.0	2329.7	27.2	-13.7	-0.14
4	17.0	2061.3	1997.0	24.7	64.3	0.67
5	5.0	2207.5	2440.6	33.2	-233.1	-2.50R
6	19.0	1708.3	1923.1	27.8	-214.8	-2.26R
7	24.0	1784.7	1738.3	38.6	46.4	0.51
8	2.5	2575.0	2533.0	39.0	42.0	0.46
9	7.5	2357.9	2348.2	28.1	9.7	0.10
10	11.0	2277.7	2218.8	23.2	58.9	0.61
11	13.0	2165.2	2144.9	22.2	20.3	0.21
12	3.8	2399.6	2486.8	36.1	-87.2	-0.95
13	25.0	1779.8	1701.3	41.1	78.5	0.87
14	9.8	2336.8	2265.0	24.6	71.7	0.75
15	22.0	1765.3	1812.2	33.9	-46.9	-0.50
16	18.0	2053.5	1960.1	26.1	93.4	0.98
17	6.0	2414.4	2403.6	31.1	10.8	0.11
18	12.5	2200.5	2163.4	22.3	37.1	0.38

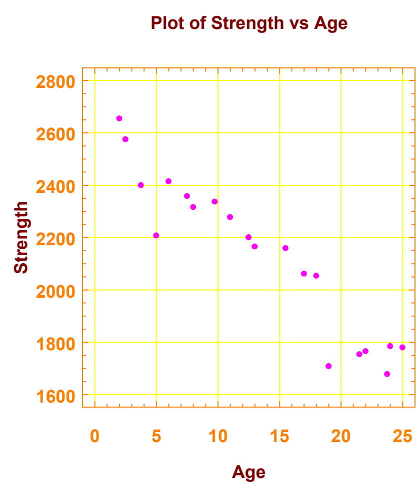
19	2.0	2654.2	2551.5	40.3	102.7	1.14
20	21.5	1753.7	1830.7	32.8	-77.0	-0.82



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

6-36. a)





The slopes of both regression models will be the same, but the intercept will be shifted.

b)
```
The regression equation is
y = 2132 - 37.0 z

Predictor        Coef     SE Coef        T       P
Constant      2132.41       22.15    96.28   0.000
z             -36.962       2.967   -12.46   0.000

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

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

$\hat{\beta}_0 = 2625.39$ vs. $\hat{\beta}_0^* = 2132.41$

$\hat{\beta}_1 = -36.9618$ $\hat{\beta}_1^* = -36.9618$

Since the data is shifted by the average age, the intercept in the model $Y = \beta_0^* + \beta_1^* z + \varepsilon$ is now the average strength.

6-37. $t_0 = \dfrac{\hat{\beta}_1}{\sqrt{\hat{\sigma}^2 / S_{xx}}}$

After the transformation

$\hat{\beta}_1^* = \dfrac{b}{a}\hat{\beta}_1$,

$S_{xx}^* = a^2 S_{xx}$,

$\overline{x}^* = a\overline{x}$, $\hat{\beta}_0^* = b\hat{\beta}_0$,

and $\hat{\sigma}^* = b\hat{\sigma}$.

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

6-38. 2) $H_0: \beta_1 = 10$

3) $H_1: \beta_1 \neq 10$

4) $\alpha = 0.01$

5) The test statistic is

$$t_0 = \frac{\hat{\beta}_1 - \beta_{1,0}}{se(\hat{\beta}_1)}$$

6) Reject H_0 if $t_0 < -t_{\alpha/2,n-2}$ where $-t_{0.005,10} = -3.17$ or $t_0 > t_{0.005,10} = 3.17$

7) Using the results from Exercise 6-2

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

8) Since $-23.37 < -3.17$ reject H_0 and conclude the coefficient of the regressor is significantly different from 10 at $\alpha = 0.01$.

P-value = 0.

6-39. a) All possible regression.

```
Response is y

                                                  x x x x x x
Vars   R-Sq    R-Sq(adj)         C-p         S    1 2 3 4 5 6

   1   99.0         99.0       101.8    50.486          X
   1   99.0         99.0       104.7    51.010    X
   2   99.6         99.5        26.7    33.941    X         X
   2   99.3         99.3        62.9    42.899    X     X
   3   99.7         99.7         7.6    27.791    X   X     X
   3   99.7         99.7         7.9    27.911    X       X X
   4   99.7         99.7         5.6    26.725    X   X   X X
   4   99.7         99.7         6.9    27.205    X     X X X
   5   99.8         99.7         5.6    26.362    X   X X X X
   5   99.8         99.7         7.1    26.916    X X X   X X
   6   99.8         99.7         7.0    26.509    X X X X X X
```

b) Forward selection. Alpha-to-Enter: 0.25

```
Response is    y     on   6 predictors, with N =   40
```

```
    Step            1        2        3        4        5
 Constant      164.99   218.71  1105.93    37.62 -3982.11

 x4             21.43    10.93    -0.06     4.61     3.75
 T-Value        62.12     4.09    -0.02     1.70     1.40
 P-Value        0.000    0.000    0.986    0.098    0.169

 x1                       0.98     1.99     1.24     1.10
 T-Value                  3.95     6.75     4.42     3.87
 P-Value                 0.000    0.000    0.000    0.000

 x6                                -8.1    -13.0    -16.3
 T-Value                          -4.64    -7.54    -6.60
 P-Value                          0.000    0.000    0.000

 x5                                         1.26     0.83
 T-Value                                    4.75     2.39
 P-Value                                   0.000    0.022

 x3                                                  0.18
 T-Value                                             1.81
 P-Value                                            0.079

 S               50.5     42.9     34.4     27.2     26.4
 R-Sq           99.02    99.31    99.57    99.74    99.76
 R-Sq(adj)      99.00    99.28    99.54    99.71    99.73
 C-p            101.8     62.9     28.7      6.9      5.6
```

c) Backward elimination. Alpha-to-Remove: 0.1

```
 Response is    y     on  6 predictors, with N =   40


    Step           1        2        3
 Constant      -4738    -3982    -4280

 x1             1.12     1.10     1.44
 T-Value        3.90     3.87    10.11
 P-Value       0.000    0.000    0.000

 x2           -0.030
 T-Value       -0.79
 P-Value       0.435

 x3             0.23     0.18     0.21
 T-Value        1.95     1.81     2.07
 P-Value       0.059    0.079    0.046

 x4              3.8      3.7
 T-Value        1.43     1.40
 P-Value       0.161    0.169

 x5             0.82     0.83     0.65
 T-Value        2.34     2.39     1.98
 P-Value       0.025    0.022    0.055

 x6            -16.9    -16.3    -17.5
 T-Value       -6.47    -6.60    -7.50
 P-Value       0.000    0.000    0.000

 S              26.5     26.4     26.7
 R-Sq          99.77    99.76    99.75
 R-Sq(adj)     99.72    99.73    99.72
 C-p             7.0      5.6      5.6
```

d) Model contains only x_1, x_3, x_5, and x_6 seems to be the "best" among all, in the sense that it has high R-Sq(adj) and small Cp value with the simple model (only four regressors).

6-40. a) All possible regression.

Response is y*

Vars	R-Sq	R-Sq(adj)	C-p	S	x1	x2	x3*	x4	x5	x6
1	98.8	98.7	54.7	0.015088				X		
1	98.5	98.5	72.0	0.016462	X					
2	99.1	99.0	32.7	0.013115		X		X		
2	99.1	99.0	34.4	0.013277				X	X	
3	99.5	99.4	6.8	0.010145			X	X		X
3	99.4	99.4	10.9	0.010663				X	X	X
4	99.5	99.5	4.6	0.0097127			X	X	X	X
4	99.5	99.4	6.3	0.0099487		X		X	X	X
5	99.5	99.5	5.1	0.0096421		X	X	X	X	X
5	99.5	99.5	6.6	0.0098471	X		X	X	X	X
6	99.5	99.5	7.0	0.0097668	X	X	X	X	X	X

b) Forward selection. Alpha-to-Enter: 0.25

Response is y* on 6 predictors, with N = 40

Step	1	2	3	4	5
Constant	7.275	6.837	6.698	6.728	-15.490
x4	0.00565	0.00430	0.00332	0.00333	0.00280
T-Value	54.80	11.31	6.52	7.98	5.63
P-Value	0.000	0.000	0.000	0.000	0.000
x2		0.00003	0.00006	0.00003	0.00002
T-Value		3.65	4.66	2.52	1.23
P-Value		0.001	0.000	0.016	0.227
x6			-0.00159	-0.00330	-0.00461
T-Value			-2.66	-5.25	-4.86
P-Value			0.011	0.000	0.000
x5				0.00041	0.00026
T-Value				4.33	2.11
P-Value				0.000	0.042
x3*					2.2
T-Value					1.81
P-Value					0.080
S	0.0151	0.0131	0.0122	0.00995	0.00964
R-Sq	98.75	99.08	99.23	99.50	99.54
R-Sq(adj)	98.72	99.03	99.17	99.44	99.48
C-p	54.7	32.7	23.7	6.3	5.1

c) Backward elimination. Alpha-to-Remove: 0.1

Response is y* on 6 predictors, with N = 40

Step	1	2	3
Constant	-16.26	-15.49	-23.51
x1	-0.00004		
T-Value	-0.37		
P-Value	0.713		
x2	0.00002	0.00002	

```
T-Value        1.25     1.23
P-Value       0.220    0.227

x3*             2.3      2.2      3.0
T-Value        1.82     1.81     2.90
P-Value       0.078    0.080    0.006

x4          0.00312  0.00280  0.00296
T-Value        3.14     5.63     6.12
P-Value       0.004    0.000    0.000

x5          0.00027  0.00026  0.00026
T-Value        2.11     2.11     2.07
P-Value       0.042    0.042    0.046

x6         -0.00463 -0.00461 -0.00501
T-Value       -4.81    -4.86    -5.56
P-Value       0.000    0.000    0.000

S           0.00977  0.00964  0.00971
R-Sq          99.55    99.54    99.52
R-Sq(adj)     99.46    99.48    99.47
C-p             7.0      5.1      4.6
```

d) Model contains only x_3^*, x_4, x_5, and x_6 seems to be the "best" among all, in the sense that it has high R-Sq(adj) and small Cp value.