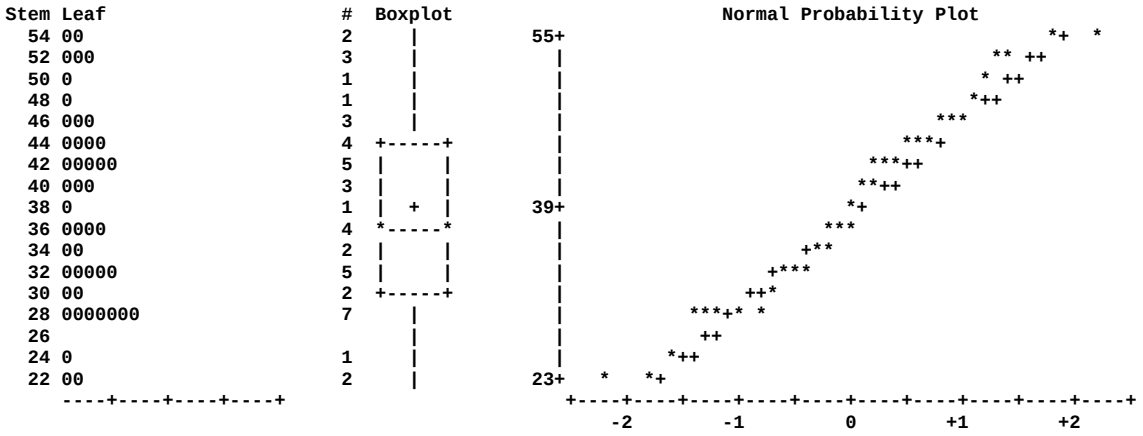
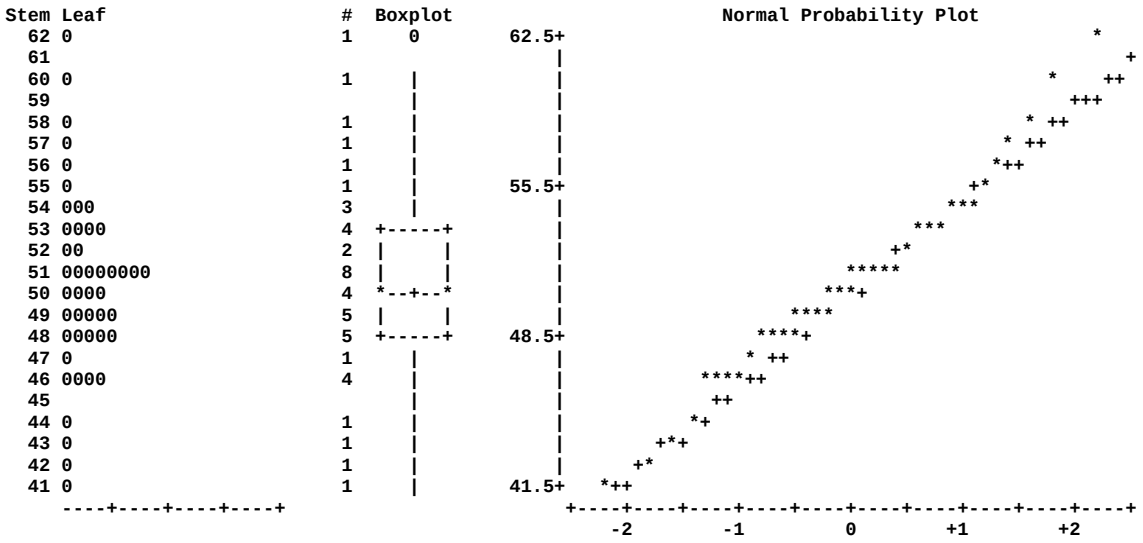
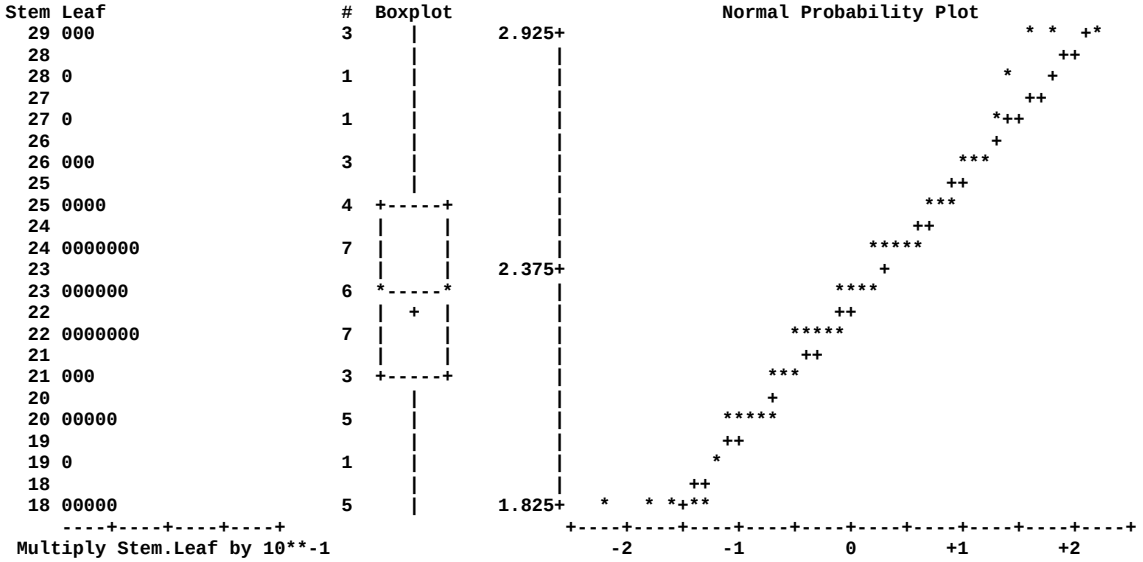
The SAS program I used to obtain the analyses for my answers is given below.

```
************************************************************;
*** EXST7034 Homework Example                              ***;
*** Applied Linear Statistical Models, 5th Edition, 2005   ***;
*** Kutner, Nachtsheim, Neter, and Li, McGraw-Hill/Irwin   ***;
*** Patient satisfaction example (6.15)                    ***;
************************************************************;

dm'log;clear;output;clear';
options nodate nocenter nonumber ps=512 ls=132;
ODS HTML style=minimal rs=none
body='C:\Geaghan\Current\EXST7034\Fall2005\SAS\PatientSatisfaction01.html' ;

TITLE1 'EXST7034 - Assignment 3, KNNL 6.15 : Patient satisfaction';

DATA Satisfaction; INFILE CARDS MISSOVER;
     LABEL X1 = 'Patients age (years)';
     LABEL X2 = 'Severity of illness (an index)';
     LABEL X3 = 'Anxiety level (an index)';
     LABEL Y = 'Patient satisfaction level';
   INPUT Y X1 X2 X3; X0=1;
CARDS; RUN;
;
Data two; set satisfaction; if y eq . then delete; run;

TITLE2 'Univariate analysis';
PROC UNIVARIATE DATA=two NORMAL PLOT; VAR Y X1 X2 X3; RUN;

PROC SORT DATA=two; BY X1 X2 X3 Y; RUN;
PROC PRINT DATA=two; VAR Y X1 X2 X3; TITLE2 'Sorted Raw Data Listing'; RUN;
PROC CORR DATA=two; VAR Y X1 X2 X3; TITLE2 'Correlation Matrix'; RUN;

options nolabel ls=99;
proc reg data=satisfaction; TITLE2 'SLR';
   MODEL  Y=X1 X2 X3;
   OUTPUT OUT=RESIDS PREDICTED=YHAT RESIDUAL=E;
RUN;
data resids; set resids; esquared = e*e; run;

TITLE2 'Breusch- Pagan test of homogeneity';
proc reg data=resids;
   MODEL  esquared=X1 X2 X3;
RUN;

PROC REG DATA=Satisfaction; TITLE2 'Multiple regression with more options';
   MODEL  Y=X1 X2 X3/SS1 SS2 CORRB COVB PCORR1 PCORR2 SCORR1 SCORR2 STB CLB
alpha=0.033333; RUN;
   OUTPUT OUT=RESIDS PREDICTED=YHAT RESIDUAL=E;
   Test1: TEST X1=X2=X3=0;
   Test2: TEST X1=X2=X3;
   Test3: TEST X1=X2;
   Test4: TEST X1=-1;
   Test5: TEST X1=-1, X2=0;
RUN;
PROC UNIVARIATE DATA=RESIDS NORMAL PLOT; VAR E; RUN;
PROC REG DATA=Satisfaction; TITLE2 'Multiple regression with more options';
   MODEL  Y=X1 X2 X3 / alpha=0.1 clm cli;
   TEST X2=X3=0;
RUN;
```

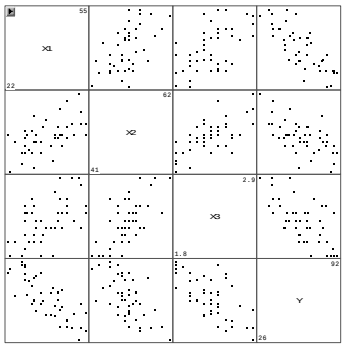
Question 6.15a from KNNL was not requested: Stem and leaf plots for each independent variable. Look for range, spacing & distribution (on ends?) and outliers, not for normality or symmetry.







Question 6.15b from KNNL:



This plot is from SAS interactive

```
EXST7034 - Assignment 3, KNNL 6.15 : Patient satisfaction
Correlation Matrix

The CORR Procedure

   4  Variables:    Y        X1       X2       X3

Simple Statistics
Variable        N          Mean     Std Dev           Sum     Minimum     Maximum    Label
Y              46      61.56522    17.23646          2832    26.00000    92.00000    Patient satisfaction level
X1             46      38.39130     8.91809          1766    22.00000    55.00000    Patients age (years)
X2             46      50.43478     4.31356          2320    41.00000    62.00000    Severity of illness (an index)
X3             46       2.28696     0.29934     105.20000     1.80000     2.90000    Anxiety level (an index)

                    Pearson Correlation Coefficients, N = 46
                           Prob > |r| under H0: Rho=0

                                        Y           X1           X2           X3
Y                                 1.00000     -0.78676     -0.60294     -0.64459
Patient satisfaction level                     <.0001       <.0001       <.0001

X1                               -0.78676      1.00000      0.56795      0.56968
Patients age (years)               <.0001                   <.0001       <.0001

X2                               -0.60294      0.56795      1.00000      0.67053
Severity of illness (an index)     <.0001       <.0001                    <.0001

X3                               -0.64459      0.56968      0.67053      1.00000
Anxiety level (an index)           <.0001       <.0001       <.0001
```

Question 6.15c from KNNL: The regression function is

$$\mathbf{Y_i = 162.875899 - 1.210318X_{1i} - 0.665906X_{2i} - 8.613032X_{3i} + e_i\ .}$$

The patient satisfaction falls -0.665906 units for each unit increase in the index of severity of illness. This result is for severity of illness alone holding the other factors constant. The indication in plain English is that sicker patients are harder to please. Note that this variable was not significant, while older patients may be harder to please as this was a significant variable.

Multiple regression on 3 variables

```
EXST7034 - Assignment 3, KNNL 6.15 : Patient satisfaction
Multiple regression with more options

The REG Procedure
Model: MODEL1
Dependent Variable: Y

Number of Observations Read                          47
Number of Observations Used                          46
Number of Observations with Missing Values            1

Analysis of Variance
                                    Sum of           Mean
Source                      DF      Squares         Square    F Value    Pr > F
Model                        3   9120.46367     3040.15456      30.05    <.0001
Error                       42   4248.84068      101.16287
Corrected Total             45        13369
```

```
Root MSE              10.05798    R-Square      0.6822
Dependent Mean        61.56522    Adj R-Sq      0.6595
Coeff Var             16.33711
```

```
Parameter Estimates
                  Parameter     Standard                                                 Standardized
Variable   DF      Estimate        Error  t Value  Pr > |t|    Type I SS   Type II SS      Estimate
Intercept   1    158.49125     18.12589     8.74    <.0001       174353   7734.51573             0
X1          1     -1.14161      0.21480    -5.31    <.0001   8275.38885   2857.55338      -0.59067
X2          1     -0.44200      0.49197    -0.90    0.3741    480.91529     81.65905      -0.11061
X3          1    -13.47016      7.09966    -1.90    0.0647    364.15952    364.15952      -0.23393
```

```
Parameter Estimates
                    Squared      Squared       Squared       Squared
                 Semi-partial      Partial  Semi-partial       Partial     96.6667% Confidence
Variable   DF   Corr Type I  Corr Type I  Corr Type II  Corr Type II          Limits
Intercept   1             .            .             .             .    118.60755    198.37495
X1          1       0.61898      0.61898       0.21374       0.40211     -1.61425     -0.66897
X2          1       0.03597      0.09441       0.00611       0.01886     -1.52451      0.64050
X3          1       0.02724      0.07894       0.02724       0.07894    -29.09206      2.15173
```

Question 6.15c Supplemental from Geaghan : I requested five additional tests for demonstration purposes. These can all be done by including a "TEST" statement after the PROC REG statements. The first test (H_0: $\beta_1 = \beta_2 = \beta_3 = 0$) is already done by testing the full model (above). Note that the test results are the same. A similar appearing test of hypothesis (H_0: $\beta_1 = \beta_2 = \beta_3$) and the other three requested tests are given below.

1) H_0: $\beta_1 = \beta_2 = \beta_3 = 0$ which is tested by TEST X1=X2=X3=0;

2) H_0: $\beta_1 = \beta_2 = \beta_3$ which is tested by TEST X1=X2=X3;

3) H_0: $\beta_1 = \beta_2$ which is tested by TEST X1=X2;

4) H_0: $\beta_1 = -1$ which is tested by TEST X1=1;

5) H_0: $\beta_1 = -1$, $\beta_2 = 0$ which is tested by TEST X1=-1, X2=0;

The results of these test statements were;

```
Test Test1 Results for Dependent Variable Y
Source              DF    Mean Square    F Value    Pr > F
Numerator            3     3040.15456      30.05    <.0001
Denominator         42      101.16287
```

```
Test Test2 Results for Dependent Variable Y
Source              DF    Mean Square    F Value    Pr > F
Numerator            2      174.89568       1.73    0.1899
Denominator         42      101.16287
```

```
Test Test3 Results for Dependent Variable Y
Source              DF    Mean Square    F Value    Pr > F
Numerator            1      140.41460       1.39    0.2454
Denominator         42      101.16287
```

```
Test Test4 Results for Dependent Variable Y
Source              DF    Mean Square    F Value    Pr > F
Numerator            1       43.97001       0.43    0.5133
Denominator         42      101.16287
```

```
Test Test5 Results for Dependent Variable Y
Source              DF    Mean Square    F Value    Pr > F
Numerator            2       89.40713       0.88    0.4208
Denominator         42      101.16287
```

Question 6.15d from KNNL: Residual plot and analysis of the box plot do not appear to show outliers.

```
EXST7034 - Assignment 3, KNNL 6.15 : Patient satisfaction
Multiple regression with more options

                          Plot of E*YHAT.  Symbol used is 'y'.

  E |
 20 +
    |                                                        y
    |                                            y
    |                                 yy
    |                     y            y                 y  yy
 10 +                                                                      y
    |                y   y                                             y
    |                                                  y     y              y
    |                              y     y       y
    |                                                   y        y
  0 +---------y---------------y-------y---------------------------y----------y----------
    |                                                    y
    |             y y         y                                                    y
    |     y                   y                     y     y
    |                       y
-10 +                                                 yy    y
    |                                       y
    |                             y             yy
    |                                                 y           y
    |                                                                      y
-20 +
    |
    ---+------------+------------+------------+------------+------------+------------+--
      30           40           50           60           70           80           90
                                            YHAT
NOTE: 1 obs had missing values.


EXST7034 - Assignment 3, KNNL 6.15 : Patient satisfaction
Multiple regression with more options

The UNIVARIATE Procedure
Variable:  E


   Stem Leaf  Boxplot
     16 62                       2     |
     14                                |
     12 514                      3     |
     10 10556                    5     |
      8 245                      3  +-----+
      6                             |     |
      4 0467                     4  |     |
      2 0047                     4  |     |
      0 1469                     4  *--+--*
     -0 9                        1  |     |
     -2 968                      3  |     |
     -4 75472                    5  |     |
     -6 7                        1  +-----+
     -8 610                      3     |
    -10 9                        1     |
    -12 78                       2     |
    -14 202                      3     |
    -16 3                        1     |
    -18 4                        1     |
        ----+----+----+----+
```

Question 6.15f from KNNL: Lack of Fit? From the sorted database below, it can be determined that there is only one pair of observation (25, 26) with repeated values on all three independent variables (X1, X2 and X3). A formal test of Lack of Fit is possible, but will have only 1 d.f..

Obs	Y	X1	X2	X3
1	83	22	51	2.0
2	86	23	41	1.8
3	63	25	49	2.0
4	89	28	43	1.8
5	92	28	46	1.8
6	88	29	46	1.9
7	89	29	48	2.4
8	82	29	48	2.5
9	77	29	50	2.1
10	77	29	52	2.3
11	64	30	51	2.4
12	76	31	47	2.0
13	72	32	46	2.6
14	57	32	52	2.4
15	59	33	42	2.0
16	60	33	49	2.1
17	79	33	56	2.5
18	80	34	49	2.2
19	51	34	51	2.3
20	57	36	46	2.3
21	83	36	49	1.8
22	66	36	49	2.0
23	59	37	53	2.1
24	47	38	55	2.2
25	47	40	48	2.2
26	66	40	48	2.2
27	70	41	44	1.8
28	46	42	50	2.2
29	55	42	51	2.7
30	60	43	50	2.3
31	66	43	53	2.3
32	67	43	53	2.4
33	37	44	51	2.6
34	52	44	58	2.9
35	54	45	48	2.4
36	68	45	51	2.2
37	42	47	50	2.6
38	43	47	53	2.5
39	37	47	60	2.4
40	36	49	54	2.9
41	48	50	51	2.3
42	26	52	62	2.9
43	57	53	54	2.2
44	36	53	57	2.8
45	49	55	51	2.4
46	34	55	54	2.5

Question 6.15g from KNNL: Breusch- Pagan test – Regress squared residuals on desired model and get SSRegression (SSR*) = 21356. 2) Calculate $\chi^2_{BP} = \frac{SSR^*}{2} \Big/ \left(\frac{SSE}{n}\right)^2$ where SSE is the error SS from regressing Y on X (SSE = 4248.840681). 3) For our example $\chi^2_{BP} = \frac{21356}{2} \Big/ \left(\frac{4248.84}{46}\right)^2 = 1.252$. 4) With 3 d.f. (the number of predictor variables) the $P(> \chi^2) = 0.750170475$

Question 6.16a, b and c from KNNL : Multiple regression tests of model significance, joint familywise CI on the regression coefficients and calculate R. These are all done on the output below.

Question 6.16a from KNNL : This test of H_0: $\beta_1 = \beta_2 = \beta_3 = 0$ is provided by SAS. The test showed that the P(>F) < 0.0001.

Question 6.16b from KNNL : The joint CI are done by SAS below.

Question 6.16c from KNNL : The value of $R^2 = 0.6822$ so the coefficient of multiple correlation is $\sqrt{0.6822} = 0.82595$.

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	3	9120.46367	3040.15456	30.05	<.0001
Error	42	4248.84068	101.16287		
Corrected Total	45	13369			

Root MSE	10.05798	R-Square	0.6822
Dependent Mean	61.56522	Adj R-Sq	0.6595
Coeff Var	16.33711		

Parameter Estimates

Variable	DF	Parameter Estimate	Standard Error	t Value	Pr > |t|	Type I SS	Type II SS	Standardized Estimate
Intercept	1	158.49125	18.12589	8.74	<.0001	174353	7734.51573	0
X1	1	-1.14161	0.21480	-5.31	<.0001	8275.38885	2857.55338	-0.59067
X2	1	-0.44200	0.49197	-0.90	0.3741	480.91529	81.65905	-0.11061
X3	1	-13.47016	7.09966	-1.90	0.0647	364.15952	364.15952	-0.23393

Parameter Estimates

Variable	DF	Squared Semi-partial Corr Type I	Squared Partial Corr Type I	Squared Semi-partial Corr Type II	Squared Partial Corr Type II	96.6667% Confidence Limits	
Intercept	1	.	.	.	.	118.60755	198.37495
X1	1	0.61898	0.61898	0.21374	0.40211	-1.61425	-0.66897
X2	1	0.03597	0.09441	0.00611	0.01886	-1.52451	0.64050
X3	1	0.02724	0.07894	0.02724	0.07894	-29.09206	2.15173

Question 6.17a from KNNL: Calculation of a 90% CI on a predicted value is given below. Note that an observation (#47) was created with Y=., X=35, X=45 and X=2.2. Since the Y value is missing it will not be used in fitting the model, but the X_i values are complete so it will be used in prediction and confidence interval calculation.

Get 90 percent CI on predicted values using PROC REG

The REG Procedure
Model: MODEL1
Dependent Variable: Y

Output Statistics

Obs	Dependent Variable	Predicted Value	Std Error Mean Predict	90% CL Mean		90% CL Predict		Residual
1	48.0000	47.8871	2.8126	43.1564	52.6177	30.3211	65.4531	0.1129
2	57.0000	66.0797	2.6048	61.6986	70.4607	48.6045	83.5548	-9.0797
3	66.0000	61.9762	1.9392	58.7146	65.2378	44.7476	79.2048	4.0238
4	70.0000	67.9907	3.9420	61.3604	74.6210	49.8207	86.1606	2.0093
5	89.0000	83.2736	3.1283	78.0120	88.5353	65.5572	100.9900	5.7264
6	36.0000	39.6206	3.6065	33.5545	45.6866	21.6488	57.5923	-3.6206
7	46.0000	58.8090	1.8678	55.6675	61.9505	41.6027	76.0152	-12.8090
8	54.0000	53.5741	2.7590	48.9337	58.2146	36.0322	71.1161	0.4259
9	26.0000	32.6597	4.3174	25.3980	39.9214	14.2500	51.0694	-6.6597
10	77.0000	74.9970	2.4218	70.9235	79.0704	57.5964	92.3975	2.0030
11	89.0000	71.8399	2.9768	66.8331	76.8467	54.1975	89.4823	17.1601
12	67.0000	53.6473	1.7673	50.6748	56.6198	36.4711	70.8235	13.3527
13	47.0000	61.1654	3.0228	56.0813	66.2496	43.5009	78.8299	-14.1654
14	51.0000	66.1529	1.8337	63.0687	69.2370	48.9570	83.3487	-15.1529
15	57.0000	44.4832	3.8020	38.0885	50.8780	26.3979	62.5686	12.5168
16	66.0000	68.7947	2.1836	65.1220	72.4674	51.4836	86.1058	-2.7947
17	79.0000	62.3904	3.4775	56.5414	68.2395	44.4908	80.2901	16.6096
18	88.0000	79.4590	2.5128	75.2325	83.6855	62.0220	96.8960	8.5410
19	60.0000	70.8725	1.8411	67.7758	73.9692	53.6744	88.0706	-10.8725
20	49.0000	40.8320	3.6115	34.7577	46.9063	22.8575	58.8065	8.1680
21	77.0000	71.4189	2.8036	66.7035	76.1344	53.8570	88.9808	5.5811
22	52.0000	43.5606	3.7215	37.3013	49.8199	25.5227	61.5985	8.4394
23	60.0000	56.3204	1.8238	53.2528	59.3879	39.1274	73.5133	3.6796
24	86.0000	89.8657	3.7058	83.6327	96.0987	71.8369	107.8945	-3.8657
25	43.0000	47.7339	2.0946	44.2109	51.2568	30.4539	65.0138	-4.7339
26	34.0000	38.1590	3.2271	32.7311	43.5868	20.3925	55.9254	-4.1590
27	63.0000	81.3524	2.9637	76.3677	86.3372	63.7163	98.9886	-18.3524
28	72.0000	66.6051	4.3380	59.3087	73.9014	48.1816	85.0285	5.3949
29	57.0000	66.6471	2.4522	62.5226	70.7716	49.2345	84.0596	-9.6471
30	55.0000	51.6319	3.0171	46.5573	56.7065	33.9701	69.2936	3.3681
31	59.0000	75.3136	3.4419	69.5244	81.1027	57.4334	93.1937	-16.3136
32	83.0000	71.4887	3.3303	65.8874	77.0900	53.6685	89.3090	11.5113
33	76.0000	75.3868	2.1347	71.7963	78.9772	58.0929	92.6806	0.6132
34	47.0000	61.9762	1.9392	58.7146	65.2378	44.7476	79.2048	-14.9762
35	36.0000	35.0751	3.2286	29.6448	40.5055	17.3079	52.8424	0.9249
36	80.0000	68.3839	1.6598	65.5922	71.1756	51.2380	85.5297	11.6161
37	82.0000	70.4929	3.5019	64.6029	76.3829	52.5798	88.4060	11.5071
38	64.0000	69.3723	2.6723	64.8777	73.8669	51.8684	86.8762	-5.3723
39	37.0000	45.9868	4.2786	38.7904	53.1833	27.6028	64.3709	-8.9868
40	42.0000	47.7129	2.9649	42.7260	52.6997	30.0761	65.3496	-5.7129
41	66.0000	54.9943	1.9600	51.6977	58.2910	37.7591	72.2296	11.0057
42	83.0000	83.8932	3.9452	77.2576	90.5289	65.7213	102.0652	-0.8932
43	37.0000	50.6957	2.4845	46.5168	54.8745	33.2702	68.1212	-13.6957
44	68.0000	54.9421	2.2694	51.1251	58.7592	37.5998	72.2845	13.0579
45	59.0000	64.5380	2.7104	59.9793	69.0968	47.0175	82.0585	-5.5380
46	92.0000	81.9476	2.9003	77.0694	86.8258	64.3413	99.5540	10.0524
47	.	69.0103	2.6646	64.5285	73.4920	51.5097	86.5109	.

Sum of Residuals 0
Sum of Squared Residuals 4248.84068
Predicted Residual SS (PRESS) 5057.88627