Chapter Homework for CSC423: 3.16, 3.22, 3.28, 3.46, 3.56
Sarah Cummings

3.16 Spreading rate of spilled liquid: Does mass diminish with time?

Code:

```
*Print first 10 lines of data
title "Sample of Spill Data";
proc print data=perm.liquidspill(obs= 10);
run;
```

Yes. From the table, we can see that the mass of the table tends to diminish as time increases.

Results:

Sample of Spill Data

Obs	TIME	MASS
1	0	6.64
2	1	6.34
3	2	6.04
4	4	5.47
5	6	4.94
6	8	4.44
7	10	3.98
8	12	3.55
9	14	3.15
10	16	2.79

Code:

```
*Create a simple linear regression of spill data
proc reg data=perm.liquidspill;
model mass=time;
run;
```

From our regression parameter estimates, we can conclude the mass reduces by a factor of 0.11402 per minute with p < .001.

Results:

The REG Procedure
Model: MODEL1
Dependent Variable: MASS MASS

Number of Observations Read	23
Number of Observations Used	23

Analysis of Variance					
Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	1	89.79420	89.79420	122.19	<.0001
Error	21	15.43269	0.73489		
Corrected Total	22	105.22689			

Root MSE	0.85726	R-Square	0.8533
Dependent Mean	2.61304	Adj R-Sq	0.8464
Coeff Var	32.80685		

Parameter Estimates						
Variable	Label	DF	Parameter Estimate	Standard Error	t Value	Pr > |t|
Intercept	Intercept	1	5.22070	0.29598	17.64	<.0001
TIME	TIME	1	-0.11402	0.01032	-11.05	<.0001

3.22 Thermal characteristics of fin-tubes:

a) Least squares line to fit the data

Code:

```
*Create a least squares regression for heat data
proc reg data=perm.heat plots=none;
model heat=ratio;
run;
```

Our least squares model is HEAT= 0.21330 + 2.42630RATIO as seen from the parameter estimates in the table to the right.

Results:

Model: MODEL1
Dependent Variable: HEAT

Number of Observations Read	24
Number of Observations Used	24

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	1	23.27392	23.27392	113.00	<.0001
Error	22	4.53108	0.20596		
Corrected Total	23	27.80500			

Root MSE	0.45383	R-Square	0.8370
Dependent Mean	4.77500	Adj R-Sq	0.8296
Coeff Var	9.50421		

Parameter Estimates

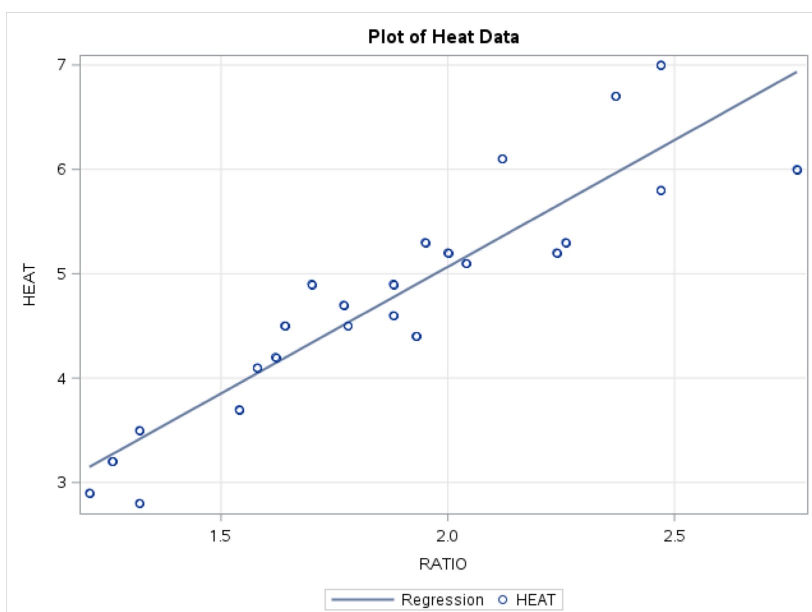
Variable	DF	Parameter Estimate	Standard Error	t Value	Pr > |t|
Intercept	1	0.21339	0.43900	0.49	0.6317
RATIO	1	2.42639	0.22825	10.63	<.0001

b) plot the data with least squares line

Code:

```
*Create scatterplot with regression
ods graphics/reset imagemap;
proc sgplot data=PERM.HEAT;
reg x=RATIO y=HEAT;
scatter x=RATIO y=HEAT;
xaxis grid;
yaxis grid;
run;
ods graphics / reset;
```

Results:



c)Calculate SSE and s squared

SSE from Analysis of Variance table above is **4.53108.** The s squared is calculated by SSE/(n-2), which is 4.521/22=**0.205**

d) calculate s: sqrt(0.205)=**0.453**

3.28 Massage therapy for boxers. Is blood lactate level (y) linearly related to perceived recovery (x)? Use a=.10

Code:

```
*Calculate regression line
proc reg data=perm.boxing2 plots=none;
model lactate=recovery;
plot lactate*recovery;
run;
```

Results:

The REG Procedure
Model: MODEL1
Dependent Variable: LACTATE

Number of Observations Read	16
Number of Observations Used	16

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	1	6.09578	6.09578	6.74	0.0211
Error	14	12.65422	0.90387		
Corrected Total	15	18.75000			

Root MSE	0.95072	R-Square	0.3251
Dependent Mean	4.92500	Adj R-Sq	0.2769
Coeff Var	19.30401		

Parameter Estimates

Variable	DF	Parameter Estimate	Standard Error	t Value	Pr > |t|
Intercept	1	2.96960	0.78959	3.76	0.0021
RECOVERY	1	0.12667	0.04878	2.60	0.0211

Yes. Our slope parameter, 0.12667, corresponds to t=2.60 and p value= 0.02,which is less than alpha. We can thus reject H_0: B_1=0 and conclude the blood lactate level is linearly related to perceived recovery.

3.46 Snow geese feeding trial

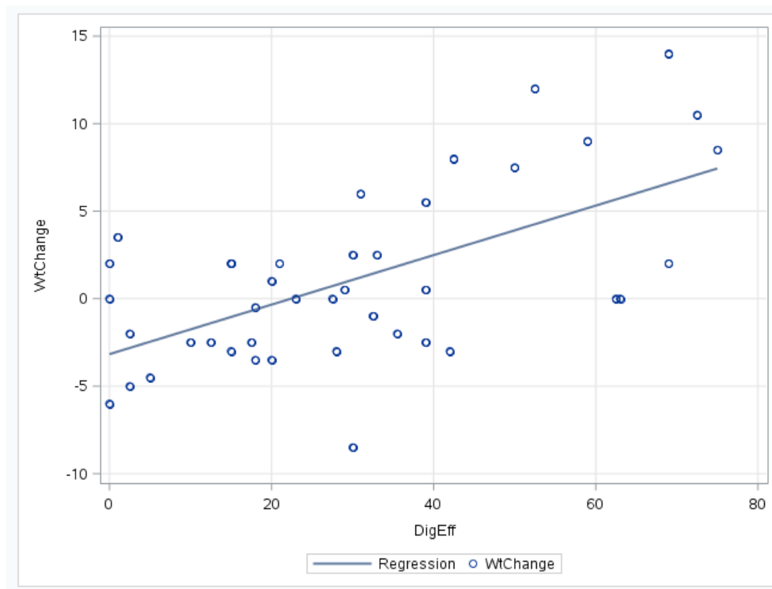
a) Plot weight change (y) versus digestion efficiency (x). Is there a trend?

Code:

```
*Create scatterplot of weight change and efficiency
ods graphics/reset imagemap;
proc sgplot data=PERM.SNOWGEESE;
reg x=DigEff y=WtChange;
scatter x=DigEff y=WtChange;
xaxis grid;
yaxis grid;
run;
ods graphics / reset;
```

Yes, there appears to be a positive correlation between digestive efficiency and weight change.

Results:



b) Find and interpret the correlation coefficient.

Code:

```
*Find Correlation
proc corr data=perm.snowgeese;
var DigEff WtChange;
run;
```

From the table, we find that the Pearson correlation coefficient is 0.61223. This means there is a strong positive relationship between digestive efficiency and weight change.

Results:

The CORR Procedure

2 Variables:	DigEff WtChange

Simple Statistics						
Variable	N	Mean	Std Dev	Sum	Minimum	Maximum
DigEff	42	30.15476	21.63993	1267	0	75.00000
WtChange	42	1.09524	5.00029	46.00000	-8.50000	14.00000

Pearson Correlation Coefficients, N = 42 Prob > |r| under H0: Rho=0		
	DigEff	WtChange
DigEff	1.00000	0.61223 <.0001
WtChange	0.61223 <.0001	1.00000

c) Conduct test to determine if weight change is correlated with digestive efficiency with a=0.01.

Code:

```
*Find Regression Line
proc reg data=perm.snowgeese plots=none;
model WtChange=DigEff;
run;
```

Our B_1 parameter estimate, 0.1415, has a t value of 4.90 which corresponds to a p value of <.001. Because this p value less than alpha, we can confirm there is correlation between digestive efficiency and weight change and reject H_0: B_1=0.

Results:

Model: MODEL1
Dependent Variable: WtChange

Number of Observations Read	42
Number of Observations Used	42

Analysis of Variance					
Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	1	384.24301	384.24301	23.98	<.0001
Error	40	640.87604	16.02190		
Corrected Total	41	1025.11905			

Root MSE	4.00274	R-Square	0.3748
Dependent Mean	1.09524	Adj R-Sq	0.3592
Coeff Var	365.46726		

Parameter Estimates					
Variable	DF	Parameter Estimate	Standard Error	t Value	Pr > |t|
Intercept	1	-3.17067	1.06784	-2.97	0.0050
DigEff	1	0.14147	0.02889	4.90	<.0001

d) Repeat b and c after removing Duck Chow data.

Code:

```
*Remove DuckChow Data
 proc sql noprint;
create table WORK.filter as select * from
PERM.SNOWGEESE where(Trial LT 34);
quit;
*Find Correlation
proc corr data=WORK.filter;
var DigEff WtChange;
run;
```

From the table, we find that the new Pearson correlation coefficient is 0.309. This means there is a moderate positive relationship between digestive efficiency and weight change.

Results:

The CORR Procedure

2 Variables:	DigEff WtChange

Simple Statistics						
Variable	N	Mean	Std Dev	Sum	Minimum	Maximur
DigEff	33	21.25758	13.87056	701.50000	0	50.0000
WtChange	33	-0.54545	3.50973	-18.00000	-8.50000	7.5000

Pearson Correlation Coefficients, N = 33 Prob > |r| under H0: Rho=0		
	DigEff	WtChange
DigEff	1.00000	0.30949 0.0797
WtChange	0.30949 0.0797	1.00000

Code:
```
*Find Regression Line
proc reg data=work.filter plots=none;
model WtChange=DigEff;
run;
```

Results:

The REG Procedure
Model: MODEL1
Dependent Variable: WtChange

Number of Observations Read	33
Number of Observations Used	33

Analysis of Variance					
Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	1	37.75736	37.75736	3.28	0.0797
Error	31	356.42446	11.49756		
Corrected Total	32	394.18182			

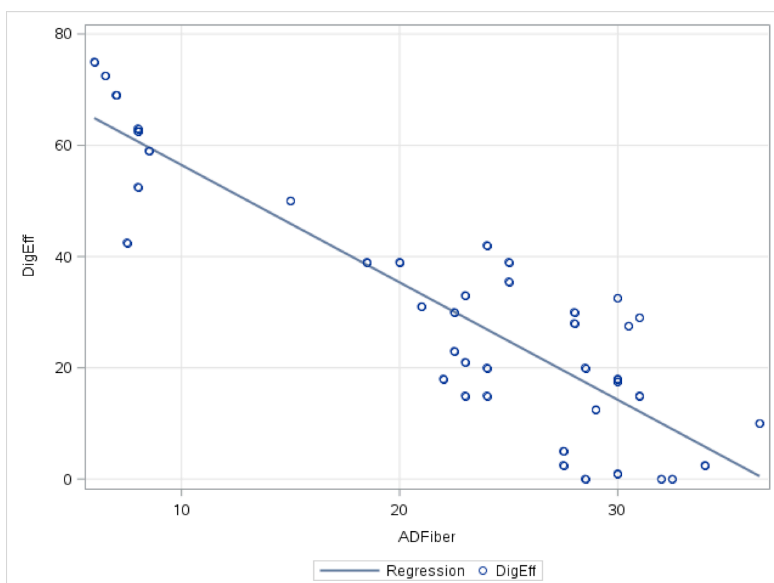
Root MSE	3.39081	R-Square	0.0958
Dependent Mean	-0.54545	Adj R-Sq	0.0666
Coeff Var	-621.64771		

Parameter Estimates					
Variable	DF	Parameter Estimate	Standard Error	t Value	Pr > |t|
Intercept	1	-2.21019	1.09193	-2.02	0.0516
DigEff	1	0.07831	0.04321	1.81	0.0797

Our B_1 parameter estimate, 0.07831, has a t value of 1.81, which corresponds to a p value of 0.0797. Because this p value is larger that a=.01, we fail to reject H_0: B_1=0 and cannot conclude the slope is significant.

e) Repeat a-d with digestive efficiency (y) and acid-detergent fiber (x).
i)Create scatterplot—is there a trend?
Code:
```
*Create Scatterplot
ods graphics/reset imagemap;
proc sgplot data=PERM.SNOWGEESE;
reg x=ADFiber y=DigEff;
scatter x=ADFiber y=DigEff;
xaxis grid;
yaxis grid;
run;
ods graphics / reset;
```

Results:



Yes, there appears to be a negative correlation between acid detergent fiber and digestive efficiency.

ii) Find and interpret correlation coefficient.

Code:

```
*Find correlation coefficient
proc corr data=perm.snowgeese;
var ADFiber DigEff;
run;
```

From the table, we find that the Pearson correlation coefficient is - 0.880. This means there is a strong negative relationship between digestive efficiency and acid detergent fiber.

Results:

The CORR Procedure

2 Variables:	ADFiber DigEff

Simple Statistics						
Variable	N	Mean	Std Dev	Sum	Minimum	Maximum
ADFiber	42	22.46429	9.02327	943.50000	6.00000	36.50000
DigEff	42	30.15476	21.63993	1267	0	75.00000

Pearson Correlation Coefficients, N = 42
Prob > |r| under H0: Rho=0

	ADFiber	DigEff
ADFiber	1.00000	-0.88005 <.0001
DigEff	-0.88005 <.0001	1.00000

iii) Conduct test to determine if acid digestive efficiency is correlated with digestive efficiency with a=0.01.

Code:

```
*Find Regression Line
proc reg data=perm.snowgeese plots=none;
model DigEff=AdFiber;
run;
```

Our B_1 parameter estimate, -2.11058, has a t value of 0.180 which corresponds to a p value of <.0001. Because this p value is less than alpha, we can reject H_0: B_1=0 and confirm there is correlation between digestive efficiency and acid detergent fiber.

Results:

The REG Procedure
Model: MODEL1
Dependent Variable: DigEff

Number of Observations Read	42
Number of Observations Used	42

Analysis of Variance					
Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	1	14870	14870	137.38	<.0001
Error	40	4329.64190	108.24105		
Corrected Total	41	19200			

Root MSE	10.40390	R-Square	0.7745
Dependent Mean	30.15476	Adj R-Sq	0.7689
Coeff Var	34.50167		

Parameter Estimates					
Variable	DF	Parameter Estimate	Standard Error	t Value	Pr > |t|
Intercept	1	77.56735	4.35204	17.82	<.0001
ADFiber	1	-2.11058	0.18007	-11.72	<.0001

iv) Repeat ii-iii with duck chow data removed.

Code:

```
* Remove DuckChow Data
 proc sql noprint;
create table WORK.filter as select * from
PERM.SNOWGEESE where(Trial LT 34);
quit;

*Find Correlation
proc corr data=WORK.filter;
var ADFiber DigEff;
run;
```

Results:

The CORR Procedure

2 Variables:	ADFiber DigEff

Simple Statistics						
Variable	N	Mean	Std Dev	Sum	Minimum	Maximum
ADFiber	33	26.57576	4.77481	877.00000	15.00000	36.50000
DigEff	33	21.25758	13.87056	701.50000	0	50.00000

Pearson Correlation Coefficients, N = 33 Prob > |r| under H0: Rho=0		
	ADFiber	DigEff
ADFiber	1.00000	-0.64591 <.0001
DigEff	-0.64591 <.0001	1.00000

From the table, we find that the new Pearson correlation coefficient is -0.645. This means there is a strong negative relationship between digestive efficiency and acid detergent fiber.

Code:

```
*Find Regression Line
proc reg data=work.filter plots=none
alpha=.01;
model DigEff=ADFiber;
run;
```

Results:

Model: MODEL1
Dependent Variable: DigEff

Number of Observations Read	33
Number of Observations Used	33

Analysis of Variance					
Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	1	2568.49205	2568.49205	22.19	<.0001
Error	31	3588.06856	115.74415		
Corrected Total	32	6156.56061			

Root MSE	10.75845	R-Square	0.4172
Dependent Mean	21.25758	Adj R-Sq	0.3984
Coeff Var	50.60994		

Parameter Estimates					
Variable	DF	Parameter Estimate	Standard Error	t Value	Pr > |t|
Intercept	1	71.12238	10.74973	6.62	<.0001
ADFiber	1	-1.87633	0.39831	-4.71	<.0001

Our B_1 parameter estimate, -1.876, has a t value of -4.71, which corresponds to a p value of <.0001. Because this p value is less than alpha, we can reject H_0: B_1=0 and conclude there is correlation between acid detergent fiber and digestive efficiency.

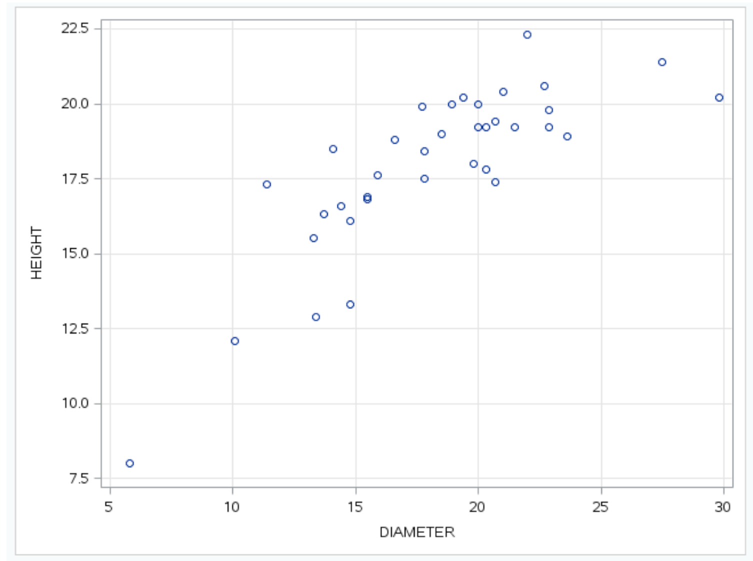
3.56 Predicting heights of spruce trees

a) Make a scatterplot of data

Code:

```
*Create a scatterplot
ods graphics/reset imagemap;
proc sgplot data=PERM.WHITESPRUCE;
scatter x=DIAMETER y=HEIGHT;
xaxis grid;
yaxis grid;
run;
ods graphics / reset;
```

Results:



b) Find estimate for y intercept.

Code:

```
ods graphics / reset;
proc reg data=PERM.WHITESPRUCE
plots=none;
model HEIGHT=DIAMETER;
run;
```

Results:

The REG Procedure
Model: MODEL1
Dependent Variable: HEIGHT

Number of Observations Read	36
Number of Observations Used	36

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	1	183.24469	183.24469	65.10	<.0001
Error	34	95.70281	2.81479		
Corrected Total	35	278.94750			

Root MSE	1.67773	R-Square	0.6569
Dependent Mean	17.90833	Adj R-Sq	0.6468
Coeff Var	9.36845		

Parameter Estimates

Variable	DF	Parameter Estimate	Standard Error	t Value	Pr > |t|
Intercept	1	9.14684	1.12131	8.16	<.0001
DIAMETER	1	0.48147	0.05967	8.07	<.0001

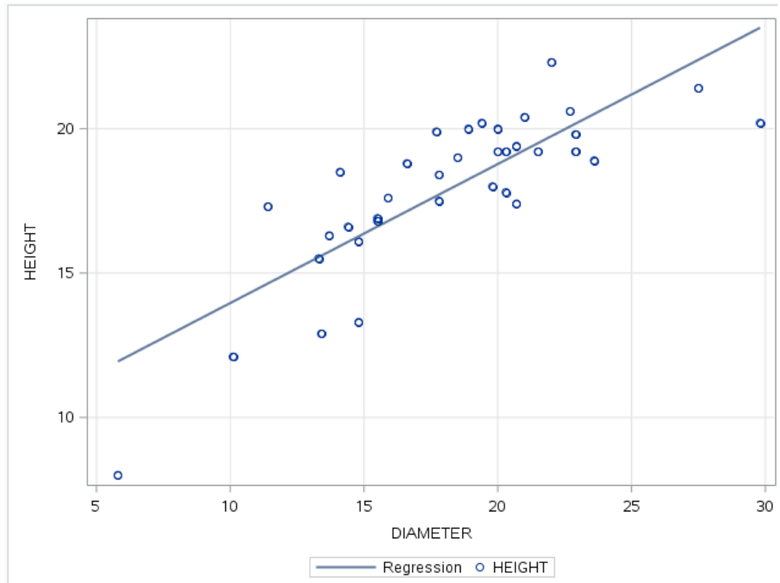
The parameter estimate for the intercept is 9.1468

c) Make a scatterplot with reg. line.

Code:

```
*Create a scatterplot with reression
ods graphics/reset imagemap;
proc sgplot data=PERM.WHITESPRUCE;
reg x=DIAMETER y=HEIGHT;
scatter x=DIAMETER y=HEIGHT;
xaxis grid;
yaxis grid;
run;
ods graphics / reset;
```

Results:



d) Does diameter predict height? Use a=.05.

Code:

```
*Create regression line
proc reg data=perm.WHITESPRUCE plots=none;
model HEIGHT=DIAMETER;
run;
```

Our parameter estimate for slope, .4814, has a t value of 8.07 which corresponds to a p<.0001. Since this p value is less than alpha, we can reject the null hypothesis of slope=0 and confirm that the diameter is a good predictor of height.

Results:

Model: MODEL1
Dependent Variable: HEIGHT

Number of Observations Read	36
Number of Observations Used	36

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	1	183.24469	183.24469	65.10	<.0001
Error	34	95.70281	2.81479		
Corrected Total	35	278.94750			

Root MSE	1.67773	R-Square	0.6569
Dependent Mean	17.90833	Adj R-Sq	0.6468
Coeff Var	9.36845		

Parameter Estimates

Variable	DF	Parameter Estimate	Standard Error	t Value	Pr > |t|
Intercept	1	9.14684	1.12131	8.16	<.0001
DIAMETER	1	0.48147	0.05967	8.07	<.0001

e) Use your LS line to find a 90%CI for average height of a tree with diam=20 cm

Code:

```
proc reg data=perm.WHITESPRUCE
plots=none;
model HEIGHT=DIAMETER/ p cli alpha=0.1;
run;
```

From our results, we obtain CI: (15.3698, 21.1236). Thus, we are 90% confident that the mean height for all white spruce trees with diameter 20 cm is between 15.3698 and 21.1236 meters.

Results:

Model: MODEL1
Dependent Variable: HEIGHT

Output Statistics

Obs	Dependent Variable	Predicted Value	Std Error Mean Predict	90% CL Predict		Residual
1	20.0	18.2467	0.2827	15.3698	21.1236	1.7533
2	16.8	16.6097	0.3226	13.7208	19.4986	0.1903
3	20.2	18.4874	0.2887	15.6088	21.3661	1.7126
4	20.0	18.7763	0.2996	15.8945	21.6581	1.2237
5	20.2	23.4948	0.7467	20.3896	26.6000	-3.2948
6	18.0	18.6800	0.2955	15.7994	21.5606	-0.6800
7	17.8	18.9208	0.3065	16.0369	21.8046	-1.1208
8	19.2	18.7763	0.2996	15.8945	21.6581	0.4237
9	22.3	19.7393	0.3601	16.8377	22.6408	2.5607
10	18.9	20.5096	0.4268	17.5824	23.4369	-1.6096
11	13.3	16.2727	0.3454	13.3763	19.1691	-2.9727
12	20.6	20.0763	0.3878	17.1646	22.9880	0.5237
13	19.0	18.0541	0.2802	15.1779	20.9303	0.9459
14	19.2	19.4985	0.3421	16.6032	22.3938	-0.2985
15	16.1	16.2727	0.3454	13.3763	19.1691	-0.1727
16	19.9	17.6689	0.2812	14.7924	20.5454	2.2311
17	20.4	19.2578	0.3258	16.3679	22.1477	1.1422
18	17.6	16.8023	0.3114	13.9169	19.6877	0.7977
19	18.8	17.1393	0.2954	14.2587	20.0199	1.6607
20	16.9	16.6097	0.3226	13.7208	19.4986	0.2903
21	16.3	15.7430	0.3876	12.8314	18.6547	0.5570
22	21.4	22.3874	0.6216	19.3620	25.4127	-0.9874
23	19.2	18.9208	0.3065	16.0369	21.8046	0.2792
24	19.8	20.1726	0.3962	17.2577	23.0875	-0.3726
25	18.5	15.9356	0.3714	13.0300	18.8412	2.5644
26	12.1	14.0097	0.5583	11.0199	16.9996	-1.9097
27	8.0	11.9394	0.7909	8.8031	15.0757	-3.9394
28	17.4	19.1134	0.3170	16.2262	22.0005	-1.7134
29	18.4	17.7171	0.2806	14.8408	20.5934	0.6829
30	17.3	14.6356	0.4927	11.6789	17.5923	2.6644
31	16.6	16.0801	0.3599	13.1786	18.9815	0.5199
32	12.9	15.5986	0.4002	12.6821	18.5151	-2.6986
33	17.5	17.7171	0.2806	14.8408	20.5934	-0.2171
34	19.4	19.1134	0.3170	16.2262	22.0005	0.2866
35	15.5	15.5504	0.4045	12.6323	18.4686	-0.0504
36	19.2	20.1726	0.3962	17.2577	23.0875	-0.9726

Sum of Residuals	0
Sum of Squared Residuals	95.70281
Predicted Residual SS (PRESS)	119.76223