Statistics for Analytics (BAN 100)

Assignment 5

by : Aaron Gonsalves

(161288196)

PROBLEM 1

CODES

```
proc import datafile='/home/u58712040/BAN100/files/malls.xlsx'
            out=work.malls
            dbms=xlsx
            replace;
            getnames=YES;
run;

title "Malls data";
proc print data=work.malls;
run;
```

Malls data

Obs	Sales	Size	Windows	Competitors	Mall Size	Nearest Competitor
1	4453	3860	39	12	943700	227
2	4770	4150	41	15	532500	142
3	4821	3880	39	15	390500	263
4	4912	4000	39	13	545500	219
5	4774	4140	40	10	329600	232
6	4638	4370	48	14	802600	257
7	4076	3570	37	16	463300	241
8	3967	3870	39	16	855200	220
9	4000	4020	44	21	443000	188
10	4379	3990	38	16	613400	209
11	5761	4930	50	15	420300	220
12	3561	3540	34	15	626700	167
13	4145	3950	36	14	601500	187
14	4406	3770	36	12	593000	199
15	4972	3940	38	11	347100	204
16	4414	3590	35	10	355900	146
17	4363	4090	38	13	490100	206
18	4499	4580	45	16	649200	144
19	3573	3580	35	18	685900	178
20	5287	4380	42	15	106200	149
21	5339	4330	40	10	354900	231
22	4656	4060	37	11	598700	225
23	3943	3380	34	16	381800	163
24	5121	4760	44	17	597900	224
25	4557	3800	36	14	745300	195

a) Multiple Regression model for the data

CODES

```
proc reg data=work.malls;

model sales = size windows competitors mall_size
nearest_competitor;

run;
```

The REG Procedure
Model: MODEL1
Dependent Variable: Sales Sales

Number of Observations Read	25
Number of Observations Used	25

Analysis of Variance

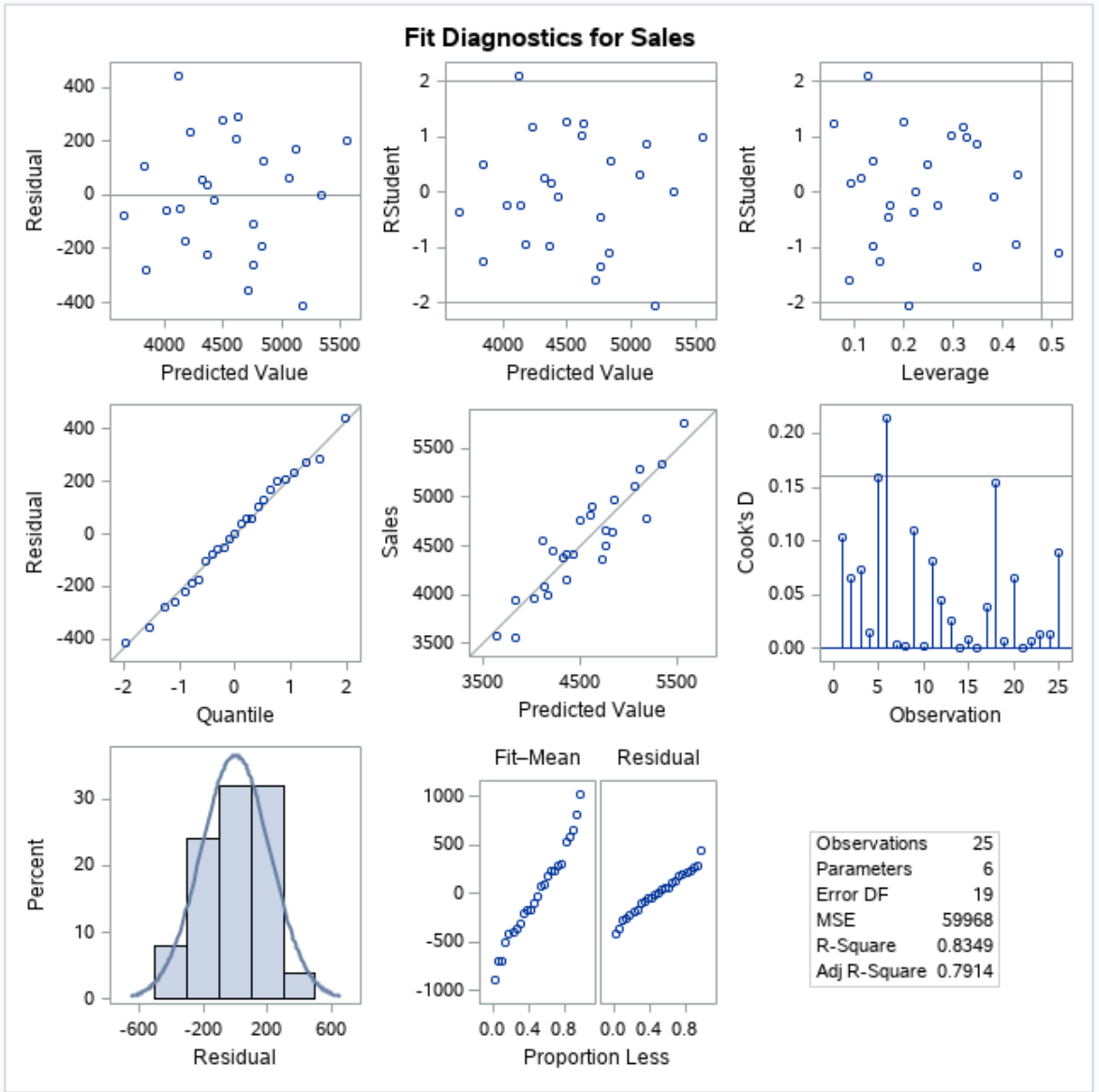
Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	5	5761406	1152281	19.21	<.0001
Error	19	1139390	59968		
Corrected Total	24	6900796			

Root MSE	244.88345	R-Square	0.8349
Dependent Mean	4535.48000	Adj R-Sq	0.7914
Coeff Var	5.39928		

Parameter Estimates

Variable	Label	DF	Parameter Estimate	Standard Error	t Value	Pr > |t|
Intercept	Intercept	1	1506.80179	672.18680	2.24	0.0371
Size	Size	1	0.91937	0.30063	3.06	0.0065
Windows	Windows	1	9.07598	28.82343	0.31	0.7563
Competitors	Competitors	1	-67.68553	21.95288	-3.08	0.0061
Mall_Size	Mall_Size	1	-0.00090285	0.00028062	-3.22	0.0045
Nearest_Competitor	Nearest_Competitor	1	2.09589	1.59443	1.31	0.2043

The REG Procedure
Model: MODEL1
Dependent Variable: Sales Sales



b) Interpret the values of the coefficients in the model.

- Intercept value : 1506.80179
- Regression co-efficient for size : 0.91937
- Regression co-efficient for windows : 9.07598
- Regression co-efficient for competitors : -67.68553
- Regression co-efficient for mall size : -0.00090285
- Regression co-efficient for nearest competitors : 2.09589

c) Test whether the model as a whole is significant. At the 0.05 level of significance, what is your conclusion?

- Here, P –value is <.0001 which is less than 0.05
- So, it can be concluded that the model as a whole is significant.

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	5	5761406	1152281	19.21	<.0001
Error	19	1139390	59968		
Corrected Total	24	6900796			

Root MSE	244.88345	R-Square	0.8349
Dependent Mean	4535.48000	Adj R-Sq	0.7914
Coeff Var	5.39928		

d) Use the model to predict monthly sales for each of the stores in the study.

CODES

```
proc reg data=work.malls;
          model sales = size windows competitors mall_size nearest_competitor/r;
          output out=temp cookd=cook student=studresids;
run;

proc sort data=temp;
          by studresids;
run;

proc print data=temp;
run;
```

Predicting the variable sales from model

- Sales is the dependent variable and its predicted value is shown in the adjacent column.
- Predicted value of sales is mostly similar to actual given value.

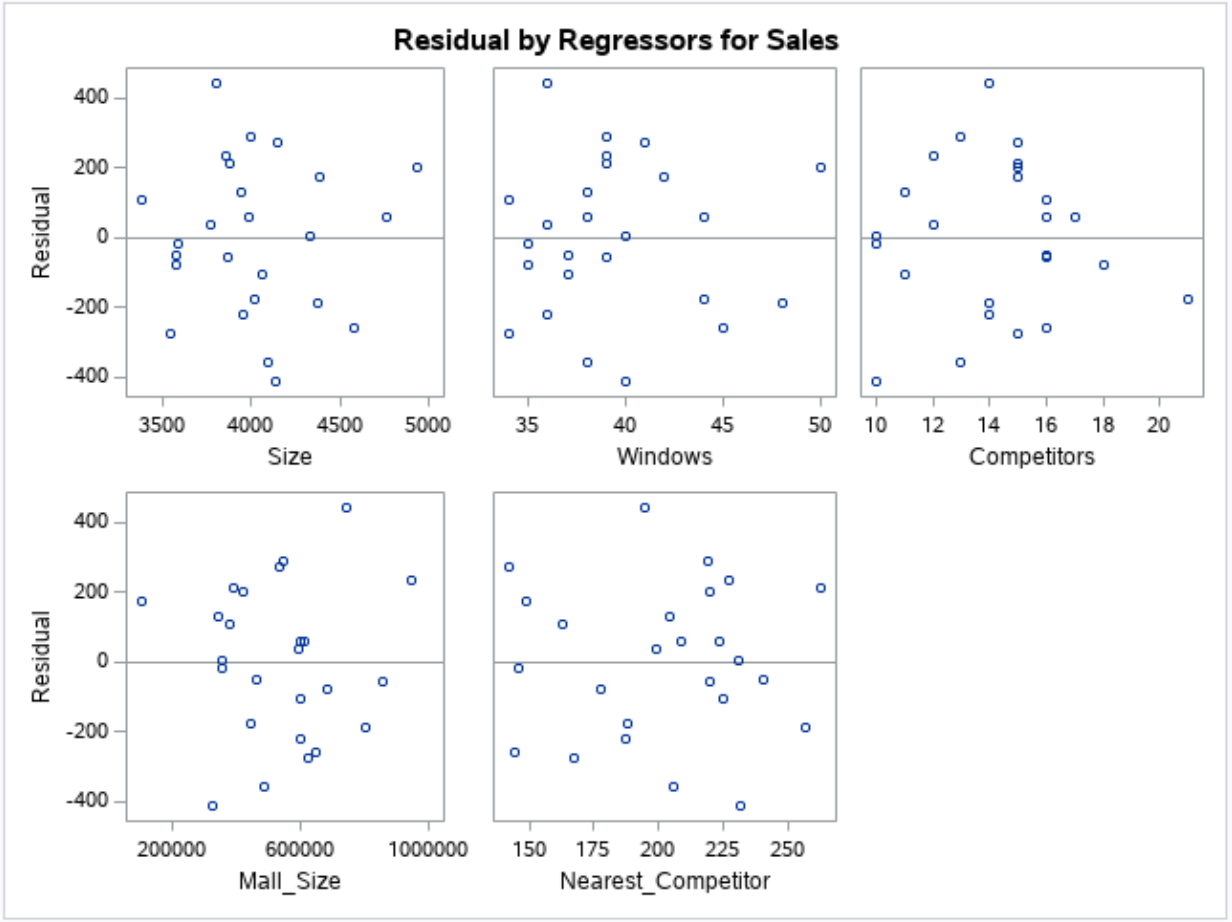
The REG Procedure
Model: MODEL1
Dependent Variable: Sales Sales

Output Statistics							
Obs	Dependent Variable	Predicted Value	Std Error Mean Predict	Residual	Std Error Residual	Student Residual	Cook's D
1	4453	4221	138.4843	231.9397	202.0	1.148	0.103
2	4770	4496	109.7095	274.1233	218.9	1.252	0.066
3	4821	4611	133.3674	209.6974	205.4	1.021	0.073
4	4912	4625	58.2133	287.1634	237.9	1.207	0.015
5	4774	5188	111.9124	-413.8542	217.8	-1.900	0.159
6	4638	4827	175.4666	-188.5227	170.8	-1.104	0.214
7	4076	4129	126.7285	-52.6221	209.5	-0.251	0.004
8	3967	4025	101.6295	-57.7434	222.8	-0.259	0.002
9	4000	4175	160.0566	-174.6896	185.3	-0.943	0.110
10	4379	4321	82.4775	57.7525	230.6	0.250	0.001
11	5761	5559	139.8824	201.5490	201.0	1.003	0.081
12	3561	3839	94.8106	-277.8760	225.8	-1.231	0.045
13	4145	4366	90.8745	-221.3260	227.4	-0.973	0.025
14	4406	4369	74.3130	36.9650	233.3	0.158	0.000
15	4972	4844	90.5548	128.3427	227.5	0.564	0.008
16	4414	4433	151.4788	-18.8276	192.4	-0.098	0.001
17	4363	4721	72.7112	-358.2757	233.8	-1.532	0.038
18	4499	4759	144.7183	-259.6539	197.5	-1.314	0.155
19	3573	3651	114.4841	-78.2762	216.5	-0.362	0.006
20	5287	5116	144.0371	171.0334	198.0	0.864	0.066
21	5339	5338	116.0869	1.4031	215.6	0.007	0.000
22	4656	4762	100.1693	-105.7615	223.5	-0.473	0.008
23	3943	3837	121.4905	106.1836	212.6	0.499	0.014
24	5121	5061	160.8775	59.6326	184.6	0.323	0.013
25	4557	4115	87.0493	441.6432	228.9	1.930	0.090

e) Plot the residuals versus the actual values. Do you think that the model does a good job of predicting monthly sales? Why or why not?

Interpretation

- Based on the residual plot, shown here, there is no systematic pattern, curve or trend for any of the variables.
- It can be concluded that our model is a good fit and does a proper job in predicting monthly sales of each store.



f) Find and interpret the value of R^2 for this model.

- R-Square value : 0.8349 i.e. 83.49%
- As the R-squared value is > 80%, it is an accurate model and can be used for prediction.

Analysis of Variance					
Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	5	5761406	1152281	19.21	<.0001
Error	19	1139390	59968		
Corrected Total	24	6900796			

Root MSE	244.88345	R-Square	0.8349
Dependent Mean	4535.48000	Adj R-Sq	0.7914
Coeff Var	5.39928		

g) Do you think that this model will be useful in helping the planners? Why or why not?

- Based on the shown table, the dependent variable, ‘sales’ and its predicted values are shown.
- These values are almost same.
- For any good model, predicted values should be near to actual values.
- So, it can be concluded that this model will be useful in helping the planners.

Obs	Dependent Variable	Predicted Value
1	4453	4221
2	4770	4496
3	4821	4611
4	4912	4625
5	4774	5188
6	4638	4827
7	4076	4129
8	3967	4025
9	4000	4175
10	4379	4321
11	5761	5559
12	3561	3839
13	4145	4366
14	4406	4369
15	4972	4844
16	4414	4433
17	4363	4721
18	4499	4759
19	3573	3651
20	5287	5116
21	5339	5338
22	4656	4762
23	3943	3837
24	5121	5061
25	4557	4115

h) Test the individual regression coefficients. At the 0.05 level of significance, what are your conclusions?

- Identifying the p-values for each variable from the regression model:

1) Size : 0.0065
2) Windows : 0.7563
3) Competitors : 0.0061
4) Mall_Size : 0.0045
5) Nearest_Competitor : 0.2043

- Variables, 'Windows' and 'Nearest Competitors' have p-value > 0.05, so they are not significant.
- Remaining variables are significant as they have p-value less than 0.05

Parameter Estimates

Variable	Label	DF	Parameter Estimate	Standard Error	t Value	Pr > |t|
Intercept	Intercept	1	1506.80179	672.18680	2.24	0.0371
Size	Size	1	0.91937	0.30063	3.06	0.0065
Windows	Windows	1	9.07598	28.82343	0.31	0.7563
Competitors	Competitors	1	-67.68553	21.95288	-3.08	0.0061
Mall_Size	Mall_Size	1	-0.00090285	0.00028062	-3.22	0.0045
Nearest_Competitor	Nearest_Competitor	1	2.09589	1.59443	1.31	0.2043

i) If you were going to drop just one variable from the model, which one would you choose? Why?

- If any variable is to be dropped by me, I would drop variable, ‘Windows’.
- As from the model, p-value of variable ‘Windows’ is 0.7563(>0.05) which is very high.
- Therefore, it is not significant in this model.

j) Use stepwise regression to find the best model for the data.

CODES

```
title "Forward, Backward and Stepwise Selection Methods
by using default values for SLENTRY and SLSTAY";

proc reg data=work.malls;
          Forward:model sales = size windows competitors mall_size nearest_competitor /
                    selection = forward;
          Backward:model sales = size windows competitors mall_size nearest_competitor /
                    selection = backward;
          Stepwise:model sales = size windows competitors mall_size nearest_competitor /
                    selection = stepwise;
run;
```

Forward selection

Forward, Backward and Stepwise Selection Methods by using default values for SLENTRY and SLSTAY

The REG Procedure
Model: Forward
Dependent Variable: Sales Sales

Number of Observations Read	25
Number of Observations Used	25

Forward Selection: Step 1

Variable Size Entered: R-Square = 0.5814 and C(p) = 27.1707

Analysis of Variance					
Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	1	4012100	4012100	31.94	<.0001
Error	23	2888696	125595		
Corrected Total	24	6900796			

Variable	Parameter Estimate	Standard Error	Type II SS	F Value	Pr > F
Intercept	222.40809	766.39576	10577	0.08	0.7743
Size	1.07258	0.18977	4012100	31.94	<.0001

Bounds on condition number: 1, 1

Forward Selection: Step 2

Variable Competitors Entered: R-Square = 0.7409 and C(p) = 10.8132

Analysis of Variance					
Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	2	5112961	2556481	31.46	<.0001
Error	22	1787835	81265		
Corrected Total	24	6900796			

Variable	Parameter Estimate	Standard Error	Type II SS	F Value	Pr > F
Intercept	1287.75994	681.05078	290546	3.58	0.0719
Size	1.08865	0.15271	4129796	50.82	<.0001
Competitors	-79.57360	21.61997	1100861	13.55	0.0013

Bounds on condition number: 1.0008, 4.0033

Forward Selection: Step 3

Variable Mall_Size Entered: R-Square = 0.8155 and C(p) = 4.2301

Analysis of Variance					
Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	3	5627674	1875891	30.94	<.0001
Error	21	1273122	60625		
Corrected Total	24	6900796			

Variable	Parameter Estimate	Standard Error	Type II SS	F Value	Pr > F
Intercept	1769.60574	611.03962	508470	8.39	0.0086
Size	1.04482	0.13276	3755185	61.94	<.0001
Competitors	-71.03060	18.90237	856069	14.12	0.0012
Mall_Size	-0.00079216	0.00027187	514713	8.49	0.0083

Bounds on condition number: 1.0367, 9.2279

Forward Selection: Step 4

Variable Nearest_Competitor Entered: R-Square = 0.8340 and C(p) = 4.0992

Analysis of Variance					
Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	4	5755460	1438865	25.13	<.0001
Error	20	1145336	57267		
Corrected Total	24	6900796			

Variable	Parameter Estimate	Standard Error	Type II SS	F Value	Pr > F
Intercept	1448.60702	631.55793	301285	5.26	0.0328
Size	1.00396	0.13189	3318012	57.94	<.0001
Competitors	-64.41386	18.89786	665329	11.62	0.0028
Mall_Size	-0.00090157	0.00027419	619136	10.81	0.0037
Nearest_Competitor	2.23558	1.49658	127786	2.23	0.1508

Bounds on condition number: 1.1481, 17.636

No other variable met the 0.5000 significance level for entry into the model.

Summary of Forward Selection							
Step	Variable Entered	Label	Number Vars In	Partial R-Square	Model R-Square	C(p)	F Value	Pr > F
1	Size	Size	1	0.5814	0.5814	27.1707	31.94	<.0001
2	Competitors	Competitors	2	0.1595	0.7409	10.8132	13.55	0.0013
3	Mall_Size	Mall_Size	3	0.0746	0.8155	4.2301	8.49	0.0083
4	Nearest_Competitor	Nearest_Competitor	4	0.0185	0.8340	4.0992	2.23	0.1508

Backward Elimination

Backward Elimination: Step 0

All Variables Entered: R-Square = 0.8349 and C(p) = 6.0000

Analysis of Variance					
Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	5	5761406	1152281	19.21	<.0001
Error	19	1139390	59968		
Corrected Total	24	6900796			

Variable	Parameter Estimate	Standard Error	Type II SS	F Value	Pr > F
Intercept	1506.80179	672.18680	301336	5.02	0.0371
Size	0.91937	0.30063	560848	9.35	0.0065
Windows	9.07598	28.82343	5945.86060	0.10	0.7563
Competitors	-67.68553	21.95288	570069	9.51	0.0061
Mall_Size	-0.00090285	0.00028062	620772	10.35	0.0045
Nearest_Competitor	2.09589	1.59443	103620	1.73	0.2043

Bounds on condition number: 5.8153, 74.153

Backward Elimination: Step 1

Variable Windows Removed: R-Square = 0.8340 and C(p) = 4.0992

Analysis of Variance					
Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	4	5755460	1438865	25.13	<.0001
Error	20	1145336	57267		
Corrected Total	24	6900796			

Variable	Parameter Estimate	Standard Error	Type II SS	F Value	Pr > F
Intercept	1448.60702	631.55793	301285	5.26	0.0328
Size	1.00396	0.13189	3318012	57.94	<.0001
Competitors	-64.41388	18.89786	665329	11.62	0.0028
Mall_Size	-0.00090157	0.00027419	619136	10.81	0.0037
Nearest_Competitor	2.23558	1.49658	127786	2.23	0.1508

Bounds on condition number: 1.1481, 17.636

Backward Elimination: Step 2

Variable Nearest_Competitor Removed: R-Square = 0.8155 and C(p) = 4.2301

Analysis of Variance					
Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	3	5627674	1875891	30.94	<.0001
Error	21	1273122	60625		
Corrected Total	24	6900796			

Variable	Parameter Estimate	Standard Error	Type II SS	F Value	Pr > F
Intercept	1769.60574	611.03962	508470	8.39	0.0086
Size	1.04482	0.13276	3755185	61.94	<.0001
Competitors	-71.03060	18.90237	856069	14.12	0.0012
Mall_Size	-0.00079216	0.00027187	514713	8.49	0.0083

Bounds on condition number: 1.0367, 9.2279

All variables left in the model are significant at the 0.1000 level.

Summary of Backward Elimination								
Step	Variable Removed	Label	Number Vars In	Partial R-Square	Model R-Square	C(p)	F Value	Pr > F
1	Windows	Windows	4	0.0009	0.8340	4.0992	0.10	0.7563
2	Nearest_Competitor	Nearest_Competitor	3	0.0185	0.8155	4.2301	2.23	0.1508

Stepwise selection

Stepwise Selection: Step 1

Variable Size Entered: R-Square = 0.5814 and C(p) = 27.1707

Analysis of Variance					
Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	1	4012100	4012100	31.94	<.0001
Error	23	2888696	125595		
Corrected Total	24	6900796			

Variable	Parameter Estimate	Standard Error	Type II SS	F Value	Pr > F
Intercept	222.40809	766.39576	10577	0.08	0.7743
Size	1.07258	0.18977	4012100	31.94	<.0001

Bounds on condition number: 1, 1

Stepwise Selection: Step 2

Variable Competitors Entered: R-Square = 0.7409 and C(p) = 10.8132

Analysis of Variance					
Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	2	5112961	2556481	31.46	<.0001
Error	22	1787835	81265		
Corrected Total	24	6900796			

Variable	Parameter Estimate	Standard Error	Type II SS	F Value	Pr > F
Intercept	1287.75994	681.05078	290546	3.58	0.0719
Size	1.08865	0.15271	4129796	50.82	<.0001
Competitors	-79.57360	21.61997	1100861	13.55	0.0013

Bounds on condition number: 1.0008, 4.0033

Stepwise Selection: Step 3

Variable Mall_Size Entered: R-Square = 0.8155 and C(p) = 4.2301

Analysis of Variance					
Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	3	5627674	1875891	30.94	<.0001
Error	21	1273122	60625		
Corrected Total	24	6900796			

Variable	Parameter Estimate	Standard Error	Type II SS	F Value	Pr > F
Intercept	1769.80574	611.03962	508470	8.39	0.0086
Size	1.04482	0.13276	3755185	61.94	<.0001
Competitors	-71.03060	18.90237	856069	14.12	0.0012
Mall_Size	-0.00079216	0.00027187	514713	8.49	0.0083

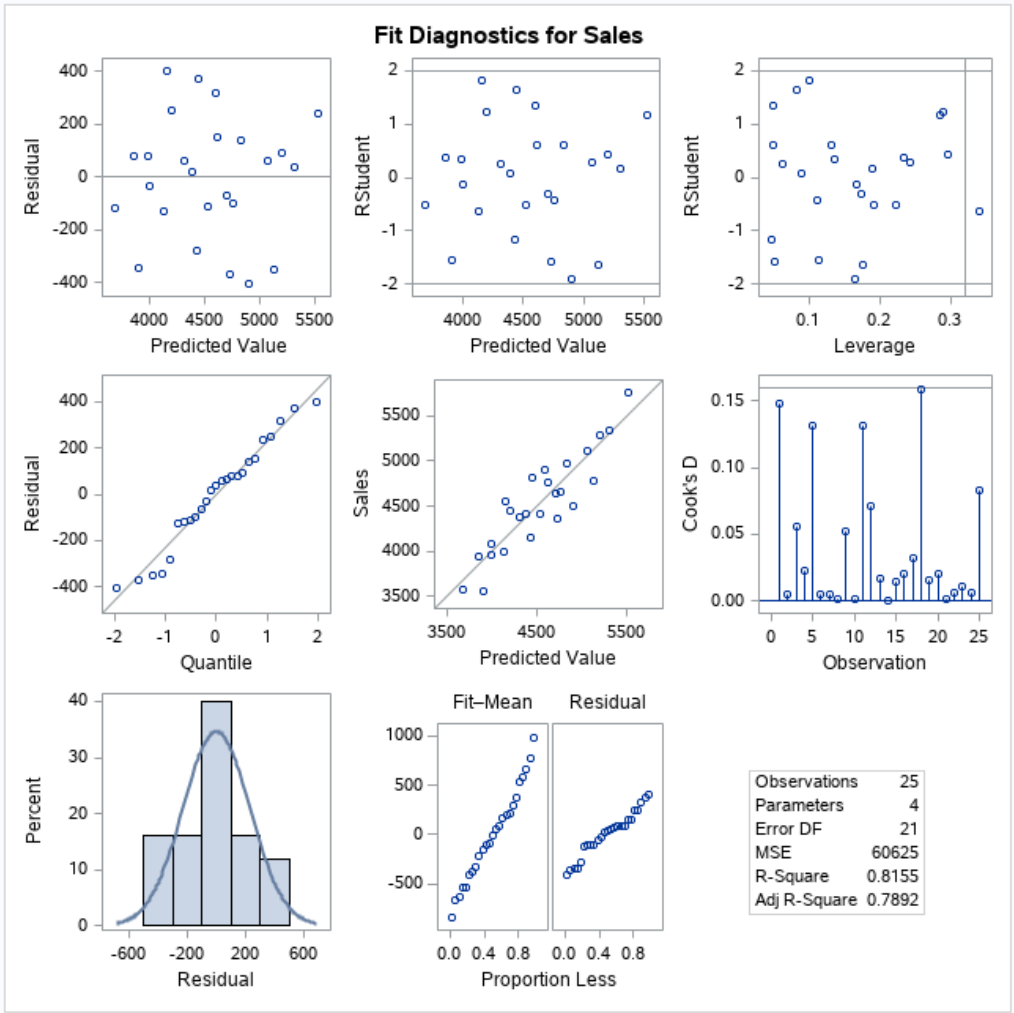
Bounds on condition number: 1.0367, 9.2279

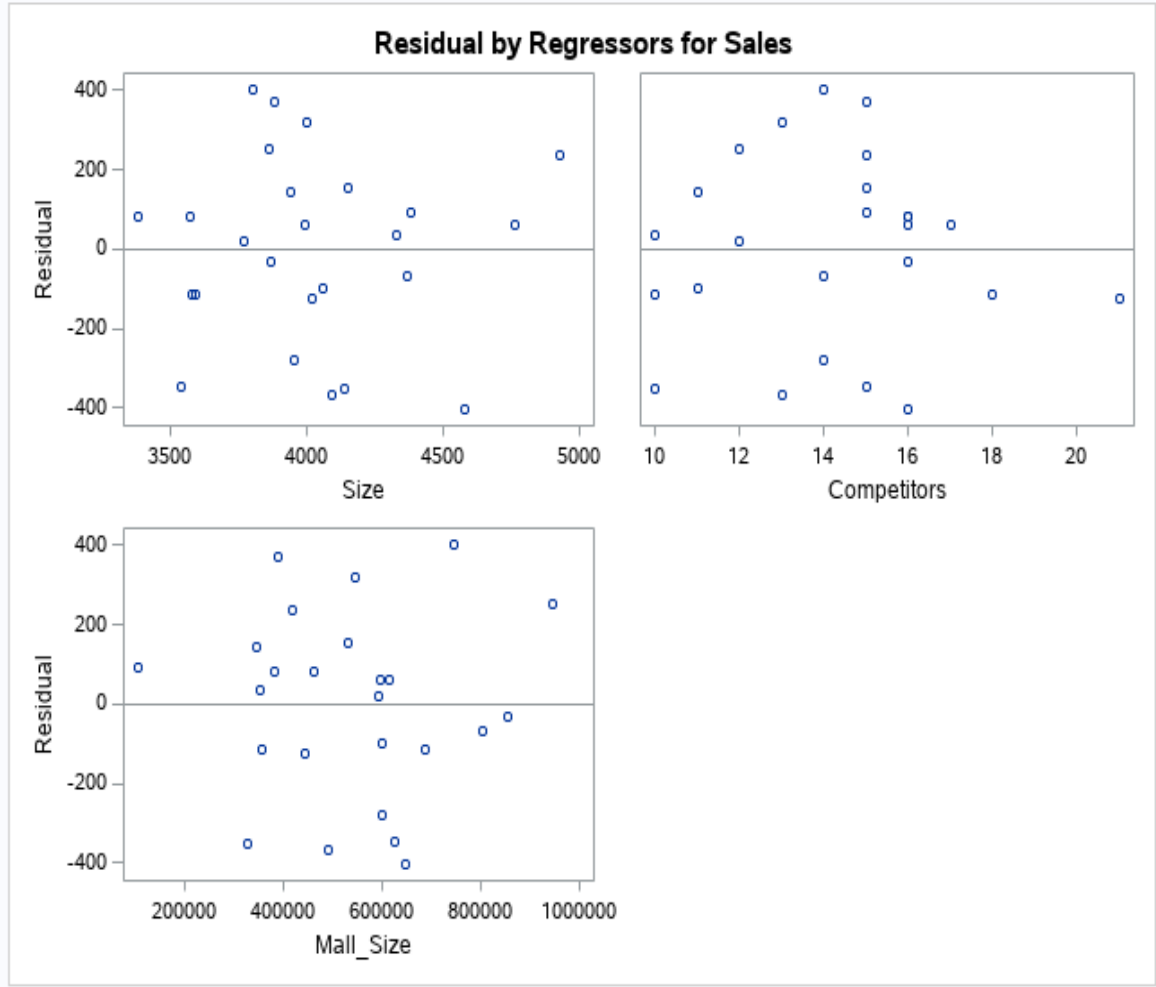
All variables left in the model are significant at the 0.1500 level.

No other variable met the 0.1500 significance level for entry into the model.

Summary of Stepwise Selection									
Step	Variable Entered	Variable Removed	Label	Number Vars In	Partial R-Square	Model R-Square	C(p)	F Value	Pr > F
1	Size		Size	1	0.5814	0.5814	27.1707	31.94	<.0001
2	Competitors		Competitors	2	0.1595	0.7409	10.8132	13.55	0.0013
3	Mall_Size		Mall_Size	3	0.0746	0.8155	4.2301	8.49	0.0083

The REG Procedure
Model: Stepwise
Dependent Variable: Sales Sales





Interpretation

- From all the performed selection methods, the best model for the data is the step wise selection model.

k) Analyze the model you have identified to determine whether it has any problems.

- The model that I have chosen is stepwise selection model, where the final variables in this model are size, competitors and mall_size.
- R-square value : 81.55%
- All the variables that are in the model are significant at 0.1500 level.
- Stepwise regression frequently gives many potential predictor variables but has very little data for estimation of coefficients meaningfully. So adding more data does not contribute.
- Out of 2 predictor variables, only one of them may be in the model if both of them are highly correlated.

l) Write a memo reporting your findings to your boss. Identify the strengths and weaknesses of the model you have chosen.

Findings:

- Out of 3 models utilized here, stepwise selection model is the best.
- From selected model, the variables mall_size, size and competitors have the highest R-squared value, and these variables have the most impact on sales.
- Predicted value of sales is almost same as the actual value of it.
- Regression plots of these variables don’t have any pattern, trend or curve.
- Therefore, this model is the best fit for prediction of sales.

Strengths:

- Can manage large amounts of potential predictor variables by tweaking the model accordingly to choose best predictor variables from available options.
- Faster than other automatic model-selection methods.
- Useful information can be gathered about the quality of predictor variables by observing the order of addition and removal of variables.

Weaknesses:

- Predicted values and confident intervals are very narrow.
- R-squared values are generally very high.
- Excessive collinearity will cause the program to discard predictor variables in the model.
- Coefficients of other variables are very high and regression coefficients are biased.
- If correlation between 2 predictor variables is high, only one of them may get in the model.

PROBLEM 2

CODES

```
proc import datafile='/home/u58712040/BAN100/files/NFLValues.xlsx'
            out=work.nfl
            dbms=xlsx
            replace;
            getnames=YES;
run;

title "NFL values";
proc print data=work.nfl;
run;
```

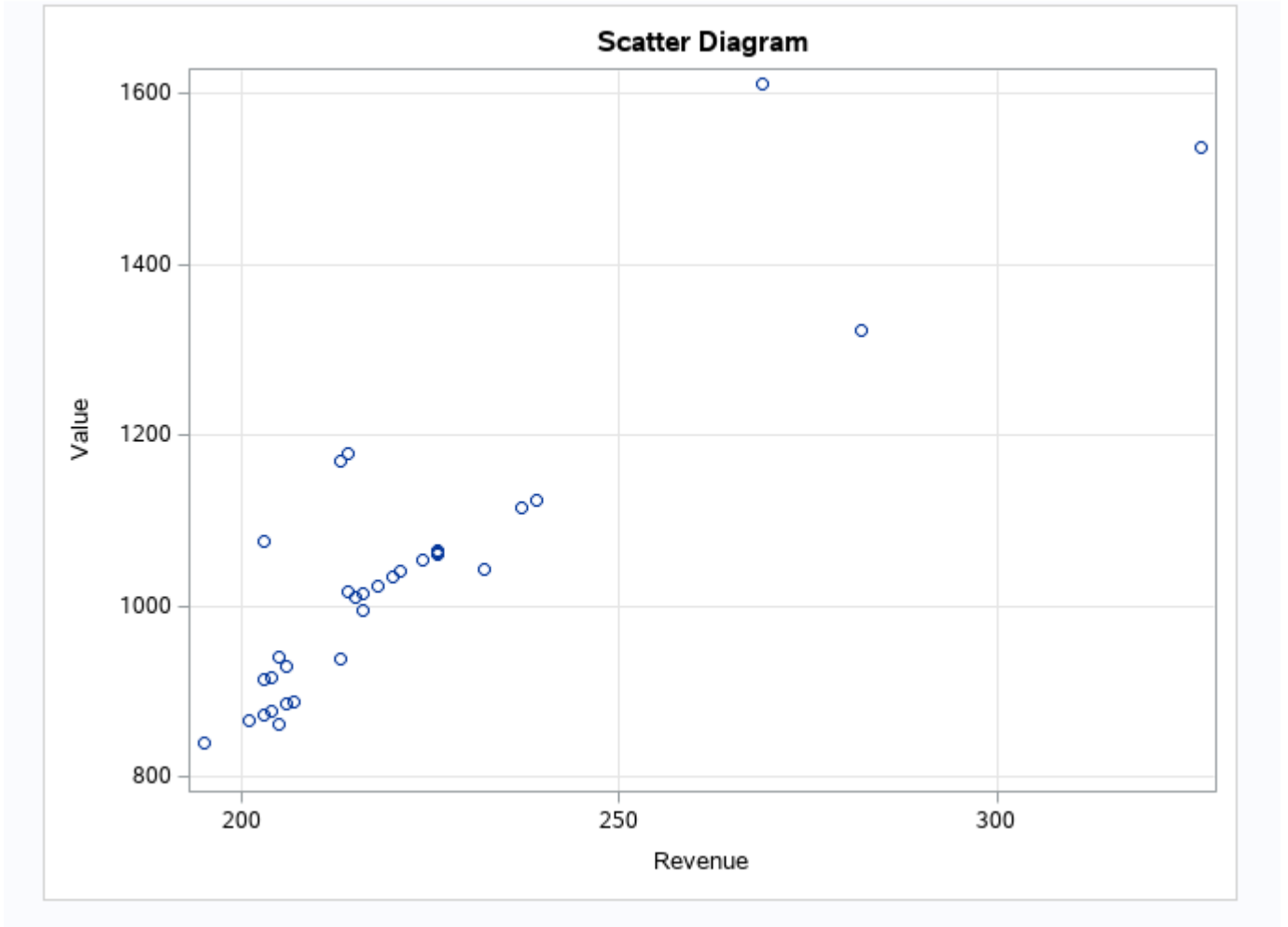
NFL values

Obs	Team	Revenue	Value
1	Arizona Cardinals	203	914
2	Atlanta Falcons	203	872
3	Baltimore Ravens	226	1062
4	Buffalo Bills	206	885
5	Carolina Panthers	221	1040
6	Chicago Bears	226	1064
7	Cincinnati Bengals	205	941
8	Cleveland Browns	220	1035
9	Dallas Cowboys	269	1612
10	Denver Broncos	226	1061
11	Detroit Lions	204	917
12	Green Bay Packers	218	1023
13	Houston Texans	239	1125
14	Indianapolis Colts	203	1076
15	Jacksonville Jaguars	204	876
16	Kansas City Chiefs	214	1016
17	Miami Dolphins	232	1044
18	Minnesota Vikings	195	839
19	New England Patriots	282	1324
20	New Orleans Saints	213	937
21	New York Giants	214	1178
22	New York Jets	213	1170
23	Oakland Raiders	205	861
24	Philadelphia Eagles	237	1116
25	Pittsburgh Steelers	216	1015
26	San Diego Chargers	207	888
27	San Francisco 49ers	201	865
28	Seattle Seahawks	215	1010
29	St Louis Rams	206	929
30	Tampa Bay Buccaneers	224	1053
31	Tennessee Titans	216	994
32	Washington Redskins	327	1538

a) Develop a scatter diagram with Revenue on the horizontal axis and Value on the vertical axis. Does it appear that there are any outliers and/or influential observations in the data?

CODES

```
title "Scatter Diagram";
proc sgplot data=work.nfl;
        scatter y = value x = Revenue;
        xaxis grid;
        yaxis grid;
run;
```



- The scatter plot has outliers and influential observations in the data.
- Some values in the graph are along the trend line, but others are away which are outliers.
- Some outliers are influential observations as they don’t lie along the trend line.
- Outliers are not always considered influential observations.
- Influential observation refers to an observation, upon whose removal the estimation of regression coefficients drastically changes.

b) Develop the estimated regression equation that can be used to predict team value given the value of annual revenue.

CODES

```
proc reg data=work.nfl;
        model value = revenue;
run;
```

The REG Procedure
Model: MODEL1
Dependent Variable: Value Value

Number of Observations Read	32
Number of Observations Used	32

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	1	753008	753008	98.93	<.0001
Error	30	228346	7611.53579		
Corrected Total	31	981354			

Root MSE	87.24412	R-Square	0.7673
Dependent Mean	1040.00000	Adj R-Sq	0.7596
Coeff Var	8.38886		

Parameter Estimates

Variable	Label	DF	Parameter Estimate	Standard Error	t Value	Pr > |t|
Intercept	Intercept	1	-252.07830	130.81712	-1.93	0.0635
Revenue	Revenue	1	5.83167	0.58631	9.95	<.0001

Estimated Regression equation

$$\text{Value} = -252.0783 + 5.83167 * \text{Revenue}$$

c) Use residual analysis to determine whether any outliers and/or influential observations are present. Briefly summarize your findings and conclusions.

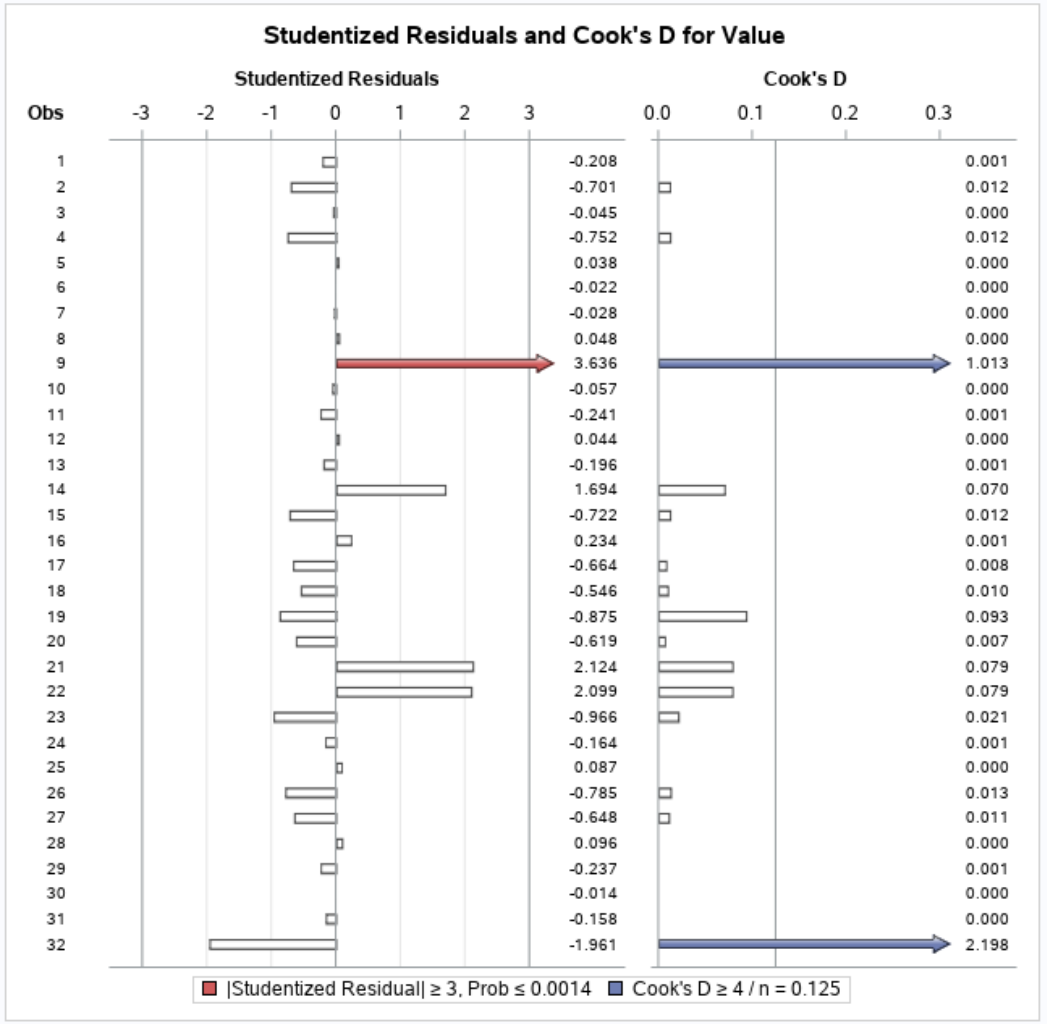
CODES

```
proc reg data=work.nfl;
            model Value = Revenue/r;
            output out = temp cookd=cook student=studresids;
run;

proc sort data=temp;
            by studresids;
run;

proc print data=temp;
run;
```

- The red bar in the Studentized residual shows that it is an outlier, with a value >=3
- The blue bar in Cook's D graph, shows that those values are influential observations in the data.



Sum of Residuals	0
Sum of Squared Residuals	228346
Predicted Residual SS (PRESS)	318993

- The value of studresids for the last observation listed here is 3.63561, which is >3 indicating that it is an outlier.

- This was also shown in the previous output, by a red bar which displayed the outlier.

Obs	Team	Revenue	Value	studresids	cook
1	Washington Redskins	327	1538	-1.96104	2.19753
2	Oakland Raiders	205	861	-0.96594	0.02129
3	New England Patriots	282	1324	-0.87514	0.09348
4	San Diego Chargers	207	888	-0.78503	0.01312
5	Buffalo Bills	206	885	-0.75242	0.01247
6	Jacksonville Jaguars	204	876	-0.72236	0.01235
7	Atlanta Falcons	203	872	-0.70147	0.01208
8	Miami Dolphins	232	1044	-0.66394	0.00827
9	San Francisco 49ers	201	865	-0.64793	0.01113
10	New Orleans Saints	213	937	-0.61904	0.00686
11	Minnesota Vikings	195	839	-0.54587	0.01004
12	Detroit Lions	204	917	-0.24142	0.00138
13	St Louis Rams	206	929	-0.23710	0.00124
14	Arizona Cardinals	203	914	-0.20838	0.00107
15	Houston Texans	239	1125	-0.19575	0.00090
16	Philadelphia Eagles	237	1116	-0.16426	0.00059
17	Tennessee Titans	216	994	-0.15804	0.00042
18	Denver Broncos	226	1061	-0.05683	0.00005
19	Baltimore Ravens	226	1062	-0.04518	0.00003
20	Cincinnati Bengals	205	941	-0.02828	0.00002
21	Chicago Bears	226	1064	-0.02188	0.00001
22	Tampa Bay Buccaneers	224	1053	-0.01415	0.00000
23	Carolina Panthers	221	1040	0.03820	0.00002
24	Green Bay Packers	218	1023	0.04398	0.00003
25	Cleveland Browns	220	1035	0.04789	0.00004
26	Pittsburgh Steelers	216	1015	0.08669	0.00013
27	Seattle Seahawks	215	1010	0.09641	0.00016
28	Kansas City Chiefs	214	1016	0.23441	0.00096
29	Indianapolis Colts	203	1076	1.69352	0.07043
30	New York Jets	213	1170	2.09901	0.07886
31	New York Giants	214	1178	2.12350	0.07895
32	Dallas Cowboys	269	1612	3.63561	1.01277

Findings and Conclusion

- There is only one outlier and a couple of influential observations in the NFLValues dataset.

- Based on the linear regression model, it is concluded that Team value depends on Revenue.

THANK YOU