**DSC423: Data Analysis And Regression / DSC 324: Data Analysis & Statistical Software II**

**Assignment-3** | **Total Points: 25pts for DSC 423 / 20pts for DSC 324**

**Due Date: 04/23/2019 by 11:59 pm**

Note:

* All assignments should be submitted in a **single MS WORD format**, no PDFs or any other file types will be accepted. If you submit any other file type, it will not be graded.
* No extensions will be given unless for a documented reason specified in the syllabus, no late assignments past the due date even a couple of minutes late will be accepted as you have an extra day (8-days) to submit your assignments.
* Submitting work that is not yours is grounds for an automatic ‘F’ for the entire course – this includes taking content and ideas from others or consulting others to complete your deliverables other than your instructor.
* SAS software and virtual server stalls, gets slow and crashes; so start early and keep multiple backups in multiple places/mediums. Late submission or inability to do the assignment due to server and/or software issues will not be accepted. Any issues relating with SAS, contact IS using the phone number provided in the syllabus, I won’t be able to help you with DePaul software related issues.

***Note: For all questions, immaterial if whether the relevant output is asked to be attached or not, make sure to include it. Also, it is important to include the sign (negative/positive or increase/decrease, and units of measurements e.g. $ or $ 99 million,%, etc.) otherwise points will be deducted.***

**Problem 1 [5 pts] – ONLY for GRADUATESre**

A university career center collects information on the job status and starting salary of graduating seniors. Data recently collected over a two-year period included over 900 seniors who had found employment at the time of graduation. The information was used to model starting salary Y as a function of two qualitative independent variables: COLLEGE at four levels {Business, Engineering, Liberal Arts, Nursing} and SEX (male and female).

1. Define the dummy variables to include college (use Business as your baseline) in a regression model for starting salary Y

numCollege1=1 and numCollege2=0 and numCollege3=0 if COLLEGE=Engineering;

numCollege2=0 and numCollege2=1 and numCollege3=0 if COLLEGE=Liberal Arts;

numCollege3=0 and numCollege2=0 and numCollege3=1 if COLLEGE=Nursing;

numCollege1=0 and numCollege2=0 and numCollege3=0 if COLLEGE=Business;

1. Write down the general regression model relating starting salary Y to both college and sex.

B1 = EngineeringSal - BusinessSal

B2 = LiberalArtsSal - BusinessSal

B3= NursingSal – BusinessSal

EngineeringSal = Bo + B1(1) + B2(0) + B3(0) = Bo + B1

LiberalArtsSal = Bo + B1(0) + B2(1) + B3(0) = Bo + B2

NursingSal = Bo + B1(0) + B2(0) + B3(1) = Bo + B3

BusinessSal = Bo + B1(0) + B2(0) + B3(0) = Bo

1. How would your model change if students in Engineering have the same starting salary as students in Business? Show the final regression model.

B1 = EngineeringSal– BusinessSal= 0

Starting Salary = {Bo + B2(numCollege2) + B3(numCollege3)} \* Sex

**Problem 2 [5 pts] – to be answered by everyone**

You will continue the analysis of the banking.txt dataset that was analyzed in Assignment 2 – data file is attached. Answer this question based on your final model from assignment-2.

1. Analyze the residuals of the regression model you found in your previous assignment. Include the residual plots. Discuss your findings.

|  |
| --- |
| 1. print banking dataset |

The REG Procedure

Model: MODEL1

Dependent Variable: Balance

|  |  |
| --- | --- |
| **Number of Observations Read** | 102 |
| **Number of Observations Used** | 102 |

| **Analysis of Variance** | | | | | |
| --- | --- | --- | --- | --- | --- |
| **Source** | **DF** | **Sum of Squares** | **Mean Square** | **F Value** | **Pr > F** |
| **Model** | 3 | 7048628348 | 2349542783 | 388.80 | <.0001 |
| **Error** | 98 | 592215797 | 6043018 |  |  |
| **Corrected Total** | 101 | 7640844145 |  |  |  |

|  |  |  |  |
| --- | --- | --- | --- |
| **Root MSE** | 2458.25514 | **R-Square** | 0.9225 |
| **Dependent Mean** | 24888 | **Adj R-Sq** | 0.9201 |
| **Coeff Var** | 9.87732 |  |  |

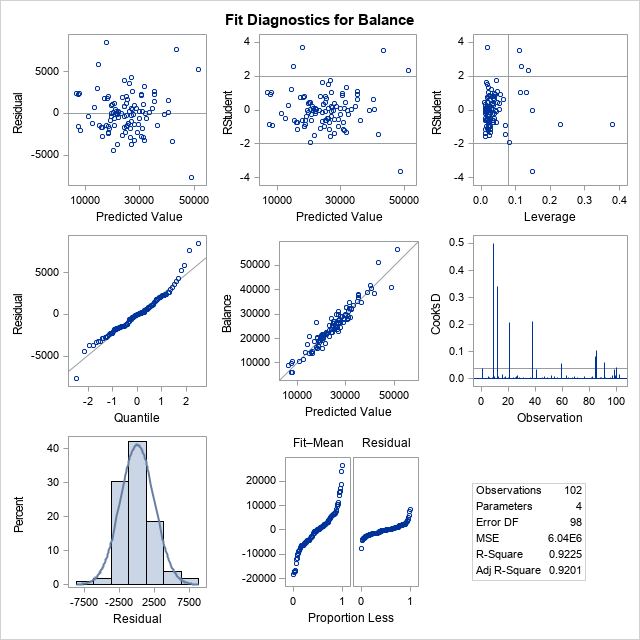
| **Parameter Estimates** | | | | | |
| --- | --- | --- | --- | --- | --- |
| **Variable** | **DF** | **Parameter Estimate** | **Standard Error** | **t Value** | **Pr > |t|** |
| **Intercept** | **1** | -9539.94542 | 4423.05947 | -2.16 | 0.0335 |
| **Age** | **1** | 332.50007 | 72.33549 | 4.60 | <.0001 |
| **Education** | **1** | 288.65551 | 300.53376 | 0.96 | 0.3392 |
| **Income** | **1** | 0.38705 | 0.01748 | 22.14 | <.0001 |

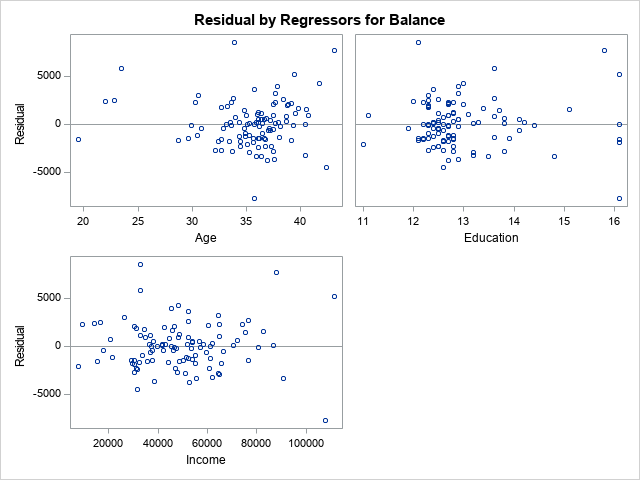
|  |
| --- |
| print banking dataset |

The REG Procedure

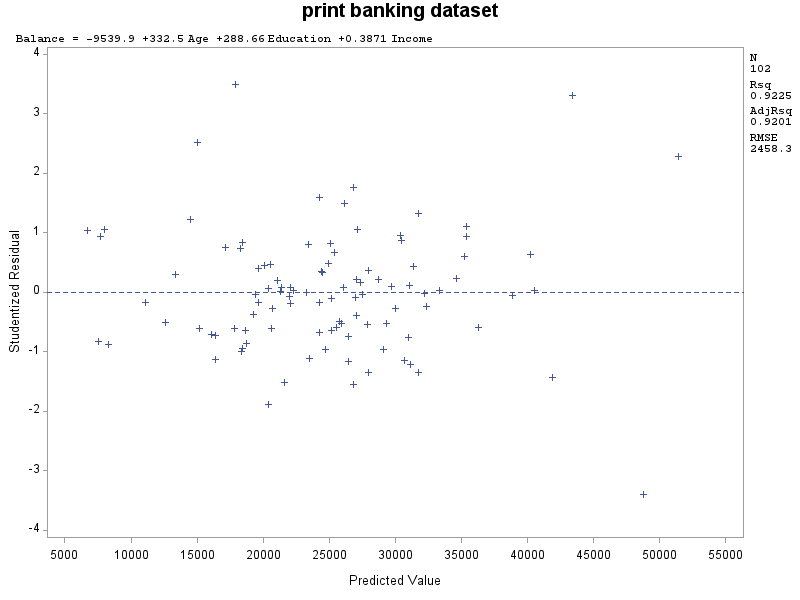
Model: MODEL1

Dependent Variable: Balance

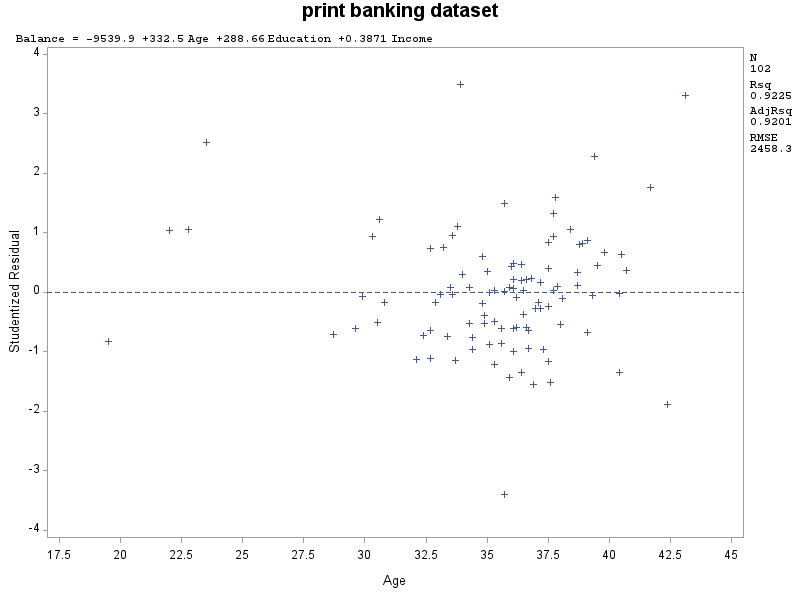




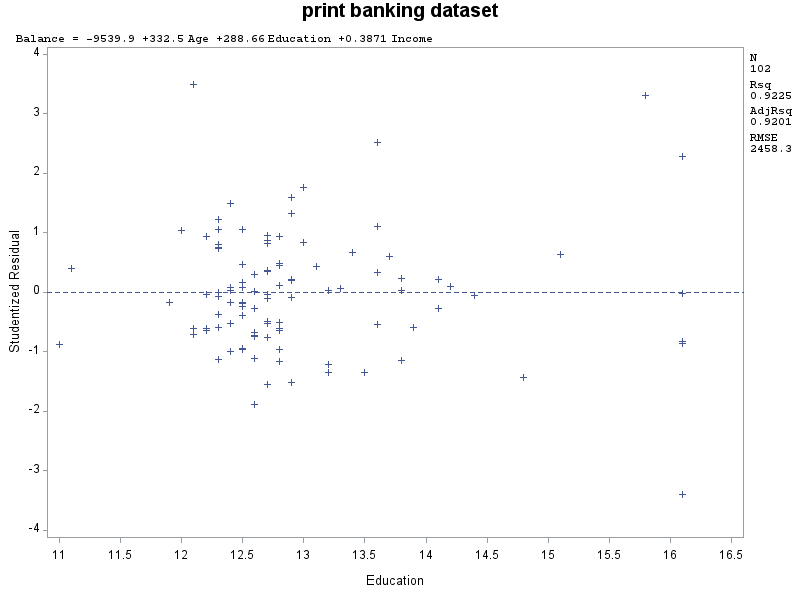
The REG Procedure



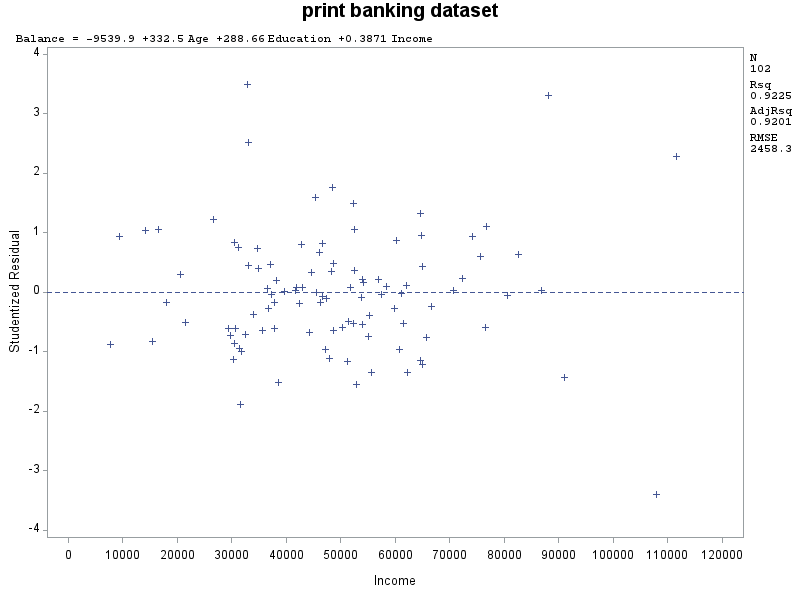
The REG Procedure



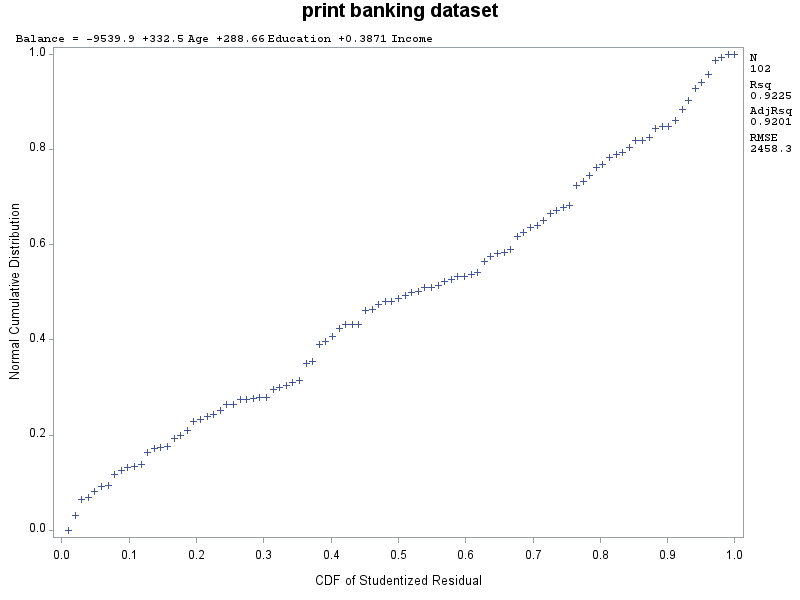
The REG Procedure



The REG Procedure



The REG Procedure



|  |
| --- |
| print banking dataset |

The REG Procedure

Model: MODEL2

Dependent Variable: Balance

|  |  |
| --- | --- |
| **Number of Observations Read** | 102 |
| **Number of Observations Used** | 102 |

| **Analysis of Variance** | | | | | |
| --- | --- | --- | --- | --- | --- |
| **Source** | **DF** | **Sum of Squares** | **Mean Square** | **F Value** | **Pr > F** |
| **Model** | 2 | 7043053576 | 3521526788 | 583.20 | <.0001 |
| **Error** | 99 | 597790568 | 6038289 |  |  |
| **Corrected Total** | 101 | 7640844145 |  |  |  |

|  |  |  |  |
| --- | --- | --- | --- |
| **Root MSE** | 2457.29294 | **R-Square** | 0.9218 |
| **Dependent Mean** | 24888 | **Adj R-Sq** | 0.9202 |
| **Coeff Var** | 9.87345 |  |  |

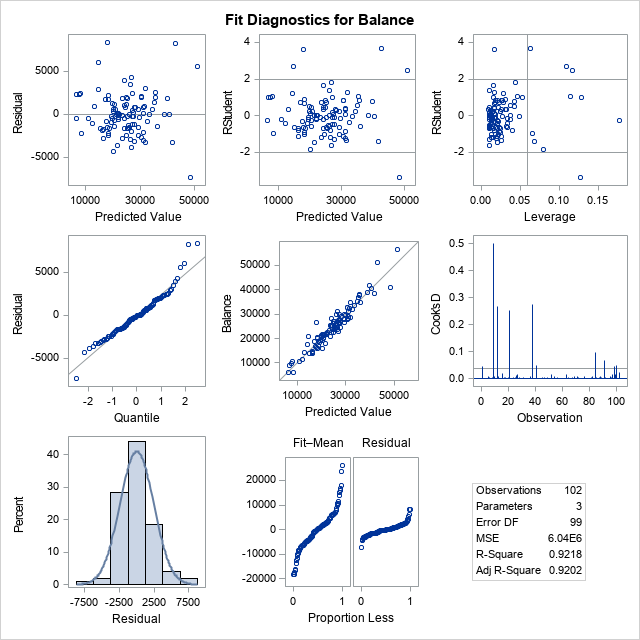
| **Parameter Estimates** | | | | | |
| --- | --- | --- | --- | --- | --- |
| **Variable** | **DF** | **Parameter Estimate** | **Standard Error** | **t Value** | **Pr > |t|** |
| **Intercept** | **1** | -5912.21531 | 2300.75819 | -2.57 | 0.0117 |
| **Age** | **1** | 322.72377 | 71.58774 | 4.51 | <.0001 |
| **Income** | **1** | 0.39661 | 0.01437 | 27.60 | <.0001 |

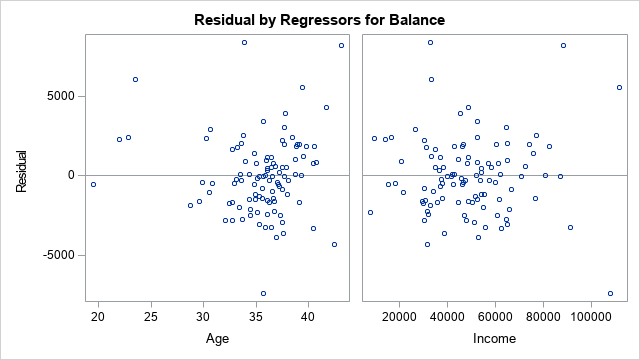
|  |
| --- |
| print banking dataset |

The REG Procedure

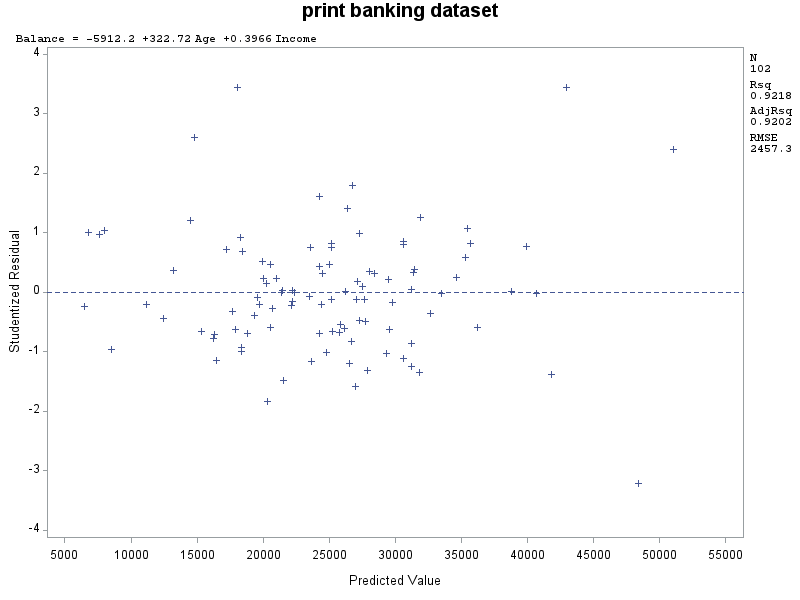
Model: MODEL2

Dependent Variable: Balance

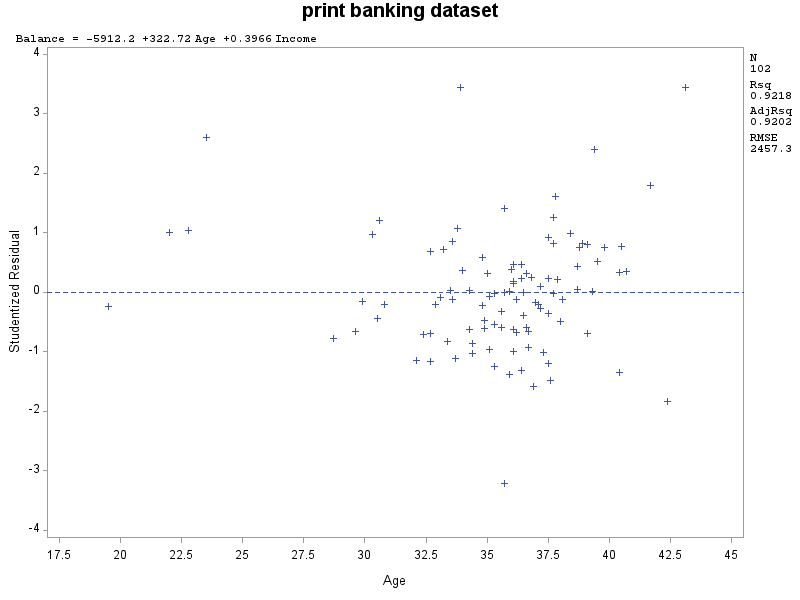




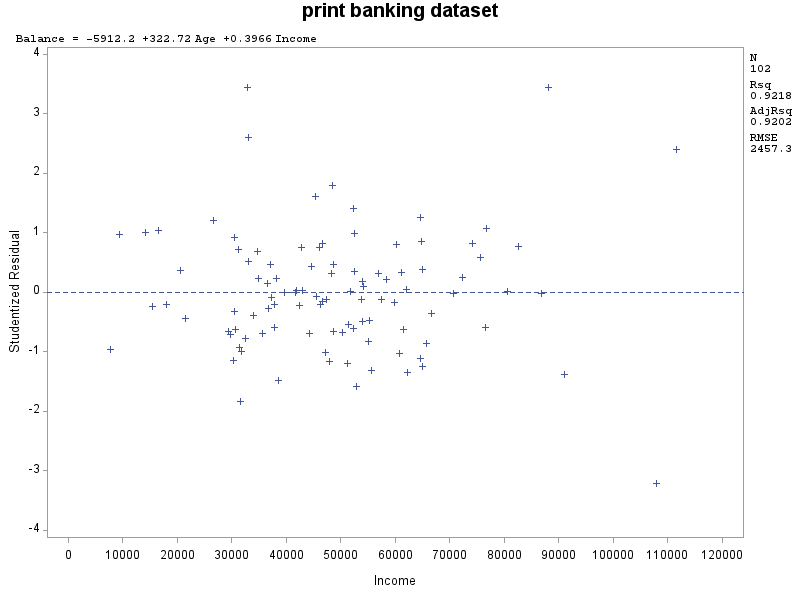
The REG Procedure



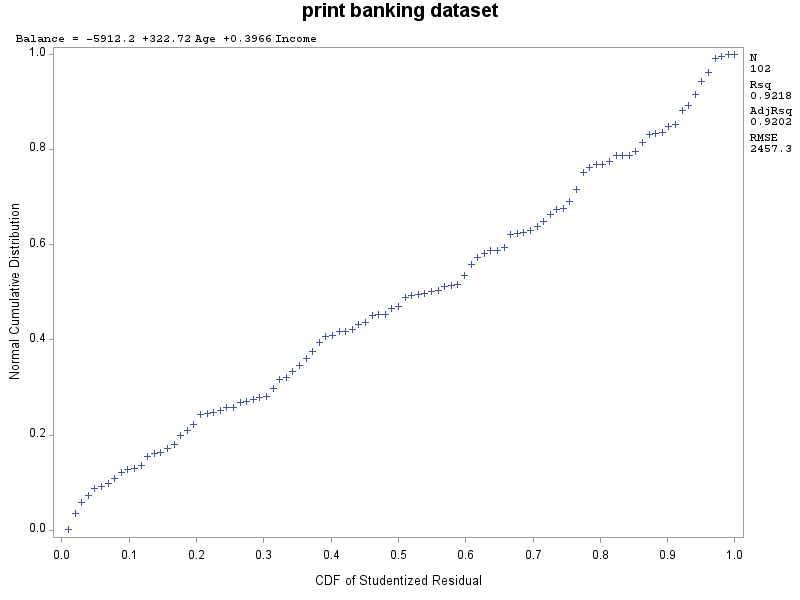
The REG Procedure



The REG Procedure



The REG Procedure



Based on this last graph, it can be seen that there is a linear trend between the variables. Balance and income along with balance and age seem to all fall within the median line, or nearby. There is a high r2 value along with an adjusted r2 so this can be deemed a good model.

1. Conduct a global F-test for overall model adequacy. Write down the test hypotheses and test statistic and discuss conclusions. Include the relevant output.

The Test hypotheses are H0: There is no effect on the dependent variable based on the independent variables.

HA: there is an effect on Y by at least one X variable.

1. Copy and paste your FULL SAS code into the word document along with your answers.

**poc** **import** datafile = "Banking.txt" out= banking replace;

deimiter= '09'x;

getames = yes;

dataow= **2**;

**run**;

title "print banking dataset";

**proc** **print**;

**run**;

title:"descriptives";

**proc** **means** mean std stderr clm min p25 p50 p75 max;

var Age Education Income Balance ;

**run**;

title:"Histogram";

**proc** **univariate** normal;

var Balance ;

\*est= estimate the mean(mu) and the std dev which is sigma);

histogram / normal (mu=est sigma =est);

**run**;

**proc** **gplot** ;

plot balance\*(age education income);

**run**;

**proc** **corr**;

var balance age education income ;

**run**;

**Proc** **reg**;

\*regression model1: Full Model;

Model balance = age education income ;

Plot student.\*predicted.;

Plot student.\* (age education income);

Plot npp.\*student.;

\*regression model2: Full Model;

Model balance = age income;

Plot student.\*predicted.;

Plot student.\*(age income);

Plot npp.\*student.;

**Run**;

**Problem 3 [15pts] – to be answered by everyone**

A national homebuilder builds single-family homes and condominium style townhouses.

The file housesales.txt provides information on the selling price (PRICE), lot cost (COST), type of home (HOME) (SF=single family home or T=condominium style) and region of the country (REGION) (M=Midwest, S=south) for closings during one month.

1. Define the dummy variables for region and home (write them down here), and create them in SAS.

numtype = 1 if type is SF, numtype = 0 if type is T

numregion = 1 if region is M, numregion = 0 if region is S

1. Analyze the association between selling price and each individual attribute (cost, home and region) using appropriate statistics and graphs. Discuss your findings. Include the relevant output.

I was not sure if the question was asking for descriptive statistics like we had in assignment 1 and 2 or for more than that, so I included basic descriptive and also scatterplot and histogram.

|  |
| --- |
| Creating Dummy Variables |

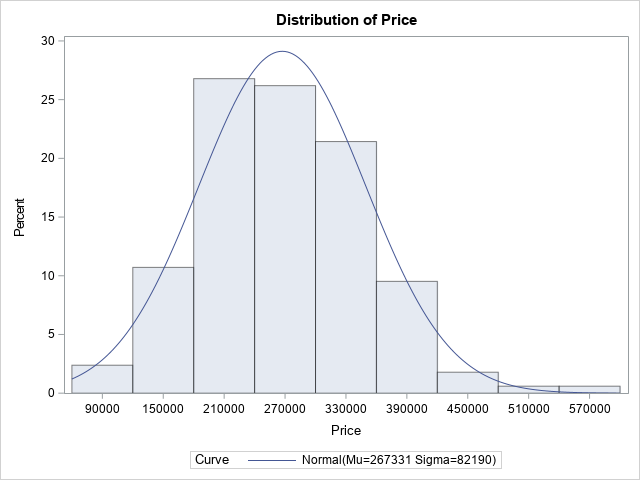
The MEANS Procedure

| **Variable** | **Mean** | **Minimum** | **25th Pctl** | **50th Pctl** | **75th Pctl** | **Maximum** |
| --- | --- | --- | --- | --- | --- | --- |
| |  | | --- | | **Price** | | **Cost** | | **numtype** | | **numregion** | | |  | | --- | | 267330.57 | | 45075.70 | | 0.8095238 | | 0.3273810 | | |  | | --- | | 85145.00 | | 17030.00 | | 0 | | 0 | | |  | | --- | | 204425.00 | | 33417.00 | | 1.0000000 | | 0 | | |  | | --- | | 263974.00 | | 44560.00 | | 1.0000000 | | 0 | | |  | | --- | | 324339.00 | | 50720.50 | | 1.0000000 | | 1.0000000 | | |  | | --- | | 575120.00 | | 85800.00 | | 1.0000000 | | 1.0000000 | |

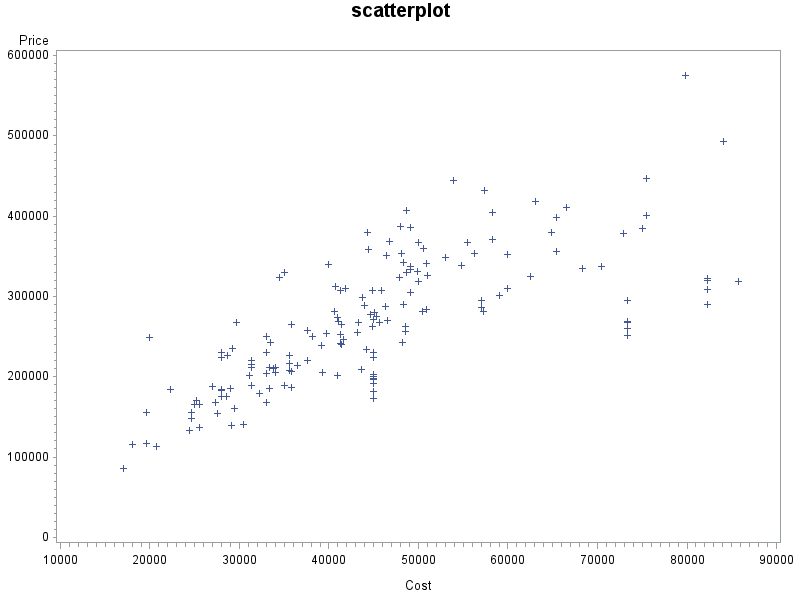
The descriptive for numtype and numregion will seem different because the values are based on 0 and 1. The average selling price was 267,330, and that falls right in the middle of the minimum of 85,145 and the max of 575,120. The inner quartile falls between 204,425 and 324,339, which is a big jump from the min and max. Cost follows a similar trend to Price. Min and Max have a big gap compared to the mean and inner quartile.

|  |
| --- |
| histogram |

The UNIVARIATE Procedure



The model can be said to be right-skewed. There don’t seem to be any apparent outliers, but 57,000 can be said to be a possible outlier.



One can make an argument for the scatterplot showing a linear trend, but the opposite judgement can be made as well that there is not a linear trend once the cost of 60,000 is reached.

1. Fit an adequate regression model for sales price as a function of lot cost, region of country, and type of home. Remove the terms that are not significant. The final model should only contain variables that are significantly associated with sale price. Write down the model equation. Include the relevant output.

Numregion has a p value of .3599 and we can remove this from the original regression model. Once we rerun, the regression model is:

Price = 42579+3.65986 cost + 73847 numtype

Where numtype = 1 if type is SF

| **Parameter Estimates** | | | | | |
| --- | --- | --- | --- | --- | --- |
| **Variable** | **DF** | **Parameter Estimate** | **Standard Error** | **t Value** | **Pr > |t|** |
| **Intercept** | **1** | 42579 | 13396 | 3.18 | 0.0018 |
| **Cost** | **1** | 3.65986 | 0.24597 | 14.88 | <.0001 |
| **numtype** | **1** | 73847 | 9650.41822 | 7.65 | <.0001 |

1. Conduct a global F-test for overall model adequacy. Write down the test hypotheses and test statistic and discuss conclusions. Include the relevant output.

| 1. **Analysis of Variance** | | | | | |
| --- | --- | --- | --- | --- | --- |
| **Source** | **DF** | **Sum of Squares** | **Mean Square** | **F Value** | **Pr > F** |
| **Model** | 3 | 7.367162E11 | 2.455721E11 | 102.89 | <.0001 |
| **Error** | 164 | 3.914126E11 | 2386662063 |  |  |
| **Corrected Total** | 167 | 1.128129E12 |  |  |  |

|  |  |  |  |
| --- | --- | --- | --- |
| **Root MSE** | 48853 | **R-Square** | 0.6530 |
| **Dependent Mean** | 267331 | **Adj R-Sq** | 0.6467 |
| **Coeff Var** | 18.27456 |  |  |

The Test hypotheses are H0: There is no effect on Price regardless of home type and region and HA: there is an effect on price based on type of home and region of home. The F value sits at 102.89. We can say the regression model fits well because SSR goes up while SSE goes down. SSR is closer to the corrected total of 1.12x10^12 than SSE is. The p-value is less than .05 so we can say the null hypothesis of no association between our dependent and independent variable is rejected and the F-test gives strong support to the fitted model.

1. Analyze model residuals to check if assumptions on data are satisfied. Discuss your findings. Include the relevant output.



Based on the residual by regressors for price and cost, we cannot really say there is a strong funnel-shape pattern or inverted-U shape. An argument can be made for the funnel-shape pattern, resulting in not being independent. Linearity and normality are hard to determine as well. It does not seem like there is a constant variance trend above or below the median line.

1. Discuss what the regression model indicates for the relationship between price and home type (i.e. interpret the coefficient values).

|  |
| --- |
| correlation |

The CORR Procedure

|  |  |
| --- | --- |
| **4 Variables:** | Price Cost numtype numregion |

| **Simple Statistics** | | | | | | |
| --- | --- | --- | --- | --- | --- | --- |
| **Variable** | **N** | **Mean** | **Std Dev** | **Sum** | **Minimum** | **Maximum** |
| **Price** | 168 | 267331 | 82190 | 44911536 | 85145 | 575120 |
| **Cost** | 168 | 45076 | 15452 | 7572717 | 17030 | 85800 |
| **numtype** | 168 | 0.80952 | 0.39385 | 136.00000 | 0 | 1.00000 |
| **numregion** | 168 | 0.32738 | 0.47066 | 55.00000 | 0 | 1.00000 |

| **Pearson Correlation Coefficients, N = 168  Prob > |r| under H0: Rho=0** | | | | |
| --- | --- | --- | --- | --- |
|  | **Price** | **Cost** | **numtype** | **numregion** |
| **Price** | |  | | --- | | 1.00000 | |  | | |  | | --- | | 0.72629 | | <.0001 | | |  | | --- | | 0.42818 | | <.0001 | | |  | | --- | | -0.49075 | | <.0001 | |
| **Cost** | |  | | --- | | 0.72629 | | <.0001 | | |  | | --- | | 1.00000 | |  | | |  | | --- | | 0.10799 | | 0.1635 | | |  | | --- | | -0.57161 | | <.0001 | |
| **numtype** | |  | | --- | | 0.42818 | | <.0001 | | |  | | --- | | 0.10799 | | 0.1635 | | |  | | --- | | 1.00000 | |  | | |  | | --- | | -0.17844 | | 0.0207 | |
| **numregion** | |  | | --- | | -0.49075 | | <.0001 | | |  | | --- | | -0.57161 | | <.0001 | | |  | | --- | | -0.17844 | | 0.0207 | | |  | | --- | | 1.00000 | |  | |

The value sits at 0.42818 and that is a low correlation value. If we take a look at the actual regression model:

Price = 42579+3.65986 cost + 73847 numtype

Where numtype = 1 if type is SF

We can say that having a SF house makes the price go up by 73,847. However, we cannot say that this is a very strong correlation.

1. Use the regression analysis to determine whether mean sale prices are different for the two regions? Explain.
2. Copy and paste your FULL SAS code into the word document along with your answers.

Title "Creating Dummy Variables";

**data** housesales;

infile 'housesales.TXT' delimiter='09'x MISSOVER firstobs=**2**;

input Region $ Type $ Price Cost;

numtype=**1**;

if type = 'T' then numtype=**0**;

numregion=**1**;

if region = 'S' then numregion=**0**;

**run**;

**proc** **print** data = housesales;

**run**;

**proc** **means** mean min p25 p50 p75 max;

var Price cost numtype numregion;

**run**;

title "histogram";

**proc** **univariate** normal;

var Price cost numtype numregion;

histogram / normal (mu=est sigma=est);

**run**;

title "scatterplot";

**proc** **GPLOT**;

PLOT Price\*(cost numtype numregion);

**run**;

title "correlation";

**proc** **CORR**;

var price cost numtype numregion;

**run**;

**proc** **reg**;

model price=cost numtype numregion;

**run**;

**proc** **reg**;

model price=cost numtype;

**run**;

|  |
| --- |
| Creating Dummy Variables |

| **Obs** | **Region** | **Type** | **Price** | **Cost** | **numtype** | **numregion** |
| --- | --- | --- | --- | --- | --- | --- |
| **1** | M | SF | 348744 | 53000 | 1 | 1 |
| **2** | M | SF | 274455 | 41000 | 1 | 1 |
| **3** | M | SF | 277720 | 44650 | 1 | 1 |
| **4** | M | SF | 307373 | 41292 | 1 | 1 |
| **5** | M | SF | 271105 | 45000 | 1 | 1 |
| **6** | M | SF | 262740 | 44900 | 1 | 1 |
| **7** | M | SF | 175000 | 28000 | 1 | 1 |
| **8** | M | SF | 201700 | 40940 | 1 | 1 |
| **9** | M | SF | 283440 | 50900 | 1 | 1 |
| **10** | M | SF | 185160 | 29000 | 1 | 1 |
| **11** | M | SF | 323716 | 34500 | 1 | 1 |
| **12** | M | SF | 281487 | 57285 | 1 | 1 |
| **13** | M | SF | 184460 | 22300 | 1 | 1 |
| **14** | M | SF | 289000 | 44000 | 1 | 1 |
| **15** | M | SF | 410810 | 66500 | 1 | 1 |
| **16** | M | SF | 184210 | 28000 | 1 | 1 |
| **17** | M | SF | 223890 | 28000 | 1 | 1 |
| **18** | M | SF | 189120 | 35000 | 1 | 1 |
| **19** | M | SF | 230440 | 33000 | 1 | 1 |
| **20** | M | SF | 330486 | 35000 | 1 | 1 |
| **21** | M | SF | 250005 | 33000 | 1 | 1 |
| **22** | M | SF | 203950 | 33000 | 1 | 1 |
| **23** | M | SF | 230555 | 28000 | 1 | 1 |
| **24** | M | SF | 183370 | 28000 | 1 | 1 |
| **25** | M | T | 112740 | 20700 | 0 | 1 |
| **26** | M | T | 179365 | 32200 | 0 | 1 |
| **27** | M | T | 155870 | 24650 | 0 | 1 |
| **28** | M | T | 155270 | 19600 | 0 | 1 |
| **29** | M | T | 116415 | 19600 | 0 | 1 |
| **30** | M | T | 147905 | 24650 | 0 | 1 |
| **31** | M | T | 139955 | 30400 | 0 | 1 |
| **32** | M | T | 184873 | 33400 | 0 | 1 |
| **33** | M | T | 212079 | 33400 | 0 | 1 |
| **34** | M | T | 265500 | 35800 | 0 | 1 |
| **35** | M | T | 175470 | 28600 | 0 | 1 |
| **36** | M | T | 115350 | 18030 | 0 | 1 |
| **37** | M | T | 85145 | 17030 | 0 | 1 |
| **38** | M | T | 139435 | 29155 | 0 | 1 |
| **39** | M | T | 133070 | 24455 | 0 | 1 |
| **40** | M | SF | 165220 | 25500 | 1 | 1 |
| **41** | M | SF | 136530 | 25500 | 1 | 1 |
| **42** | M | SF | 153845 | 27500 | 1 | 1 |
| **43** | M | SF | 165350 | 25000 | 1 | 1 |
| **44** | M | SF | 168354 | 27316 | 1 | 1 |
| **45** | M | SF | 170000 | 25200 | 1 | 1 |
| **46** | M | SF | 210380 | 33856 | 1 | 1 |
| **47** | M | SF | 268210 | 29700 | 1 | 1 |
| **48** | M | SF | 233900 | 44200 | 1 | 1 |
| **49** | M | SF | 168500 | 33000 | 1 | 1 |
| **50** | M | SF | 248500 | 20000 | 1 | 1 |
| **51** | M | SF | 220257 | 31300 | 1 | 1 |
| **52** | M | SF | 214900 | 31300 | 1 | 1 |
| **53** | M | SF | 211513 | 31300 | 1 | 1 |
| **54** | M | SF | 188603 | 31300 | 1 | 1 |
| **55** | M | T | 187390 | 27000 | 0 | 1 |
| **56** | S | T | 335000 | 68375 | 0 | 0 |
| **57** | S | T | 294450 | 73400 | 0 | 0 |
| **58** | S | T | 267060 | 73400 | 0 | 0 |
| **59** | S | T | 250800 | 73400 | 0 | 0 |
| **60** | S | T | 269410 | 73400 | 0 | 0 |
| **61** | S | T | 267640 | 73400 | 0 | 0 |
| **62** | S | T | 260100 | 73400 | 0 | 0 |
| **63** | S | SF | 301500 | 59000 | 1 | 0 |
| **64** | S | SF | 309075 | 82250 | 1 | 0 |
| **65** | S | SF | 290190 | 82250 | 1 | 0 |
| **66** | S | SF | 322920 | 82250 | 1 | 0 |
| **67** | S | SF | 319602 | 82250 | 1 | 0 |
| **68** | S | SF | 294990 | 57000 | 1 | 0 |
| **69** | S | SF | 286758 | 57000 | 1 | 0 |
| **70** | S | SF | 352781 | 60000 | 1 | 0 |
| **71** | S | SF | 310372 | 60000 | 1 | 0 |
| **72** | S | SF | 400330 | 75510 | 1 | 0 |
| **73** | S | SF | 446507 | 75510 | 1 | 0 |
| **74** | S | T | 198202 | 45025 | 0 | 0 |
| **75** | S | T | 200423 | 45025 | 0 | 0 |
| **76** | S | T | 181916 | 45025 | 0 | 0 |
| **77** | S | T | 203076 | 45025 | 0 | 0 |
| **78** | S | T | 196898 | 45025 | 0 | 0 |
| **79** | S | T | 182237 | 45025 | 0 | 0 |
| **80** | S | T | 224108 | 45025 | 0 | 0 |
| **81** | S | T | 230000 | 45025 | 0 | 0 |
| **82** | S | T | 172749 | 45025 | 0 | 0 |
| **83** | S | SF | 318274 | 85800 | 1 | 0 |
| **84** | S | SF | 191028 | 45000 | 1 | 0 |
| **85** | S | SF | 200119 | 45000 | 1 | 0 |
| **86** | S | SF | 242899 | 48252 | 1 | 0 |
| **87** | S | SF | 387527 | 48000 | 1 | 0 |
| **88** | S | SF | 257040 | 37631 | 1 | 0 |
| **89** | S | SF | 270518 | 46499 | 1 | 0 |
| **90** | S | SF | 265058 | 41404 | 1 | 0 |
| **91** | S | SF | 255000 | 43198 | 1 | 0 |
| **92** | S | SF | 385942 | 49123 | 1 | 0 |
| **93** | S | SF | 354065 | 48115 | 1 | 0 |
| **94** | S | SF | 333158 | 49123 | 1 | 0 |
| **95** | S | SF | 254048 | 39680 | 1 | 0 |
| **96** | S | SF | 246648 | 41600 | 1 | 0 |
| **97** | S | SF | 367600 | 50000 | 1 | 0 |
| **98** | S | SF | 318523 | 50000 | 1 | 0 |
| **99** | S | SF | 359949 | 50591 | 1 | 0 |
| **100** | S | SF | 281824 | 50448 | 1 | 0 |
| **101** | S | SF | 355688 | 65373 | 1 | 0 |
| **102** | S | SF | 305000 | 49067 | 1 | 0 |
| **103** | S | SF | 299096 | 43784 | 1 | 0 |
| **104** | S | SF | 280622 | 45130 | 1 | 0 |
| **105** | S | SF | 404510 | 58225 | 1 | 0 |
| **106** | S | SF | 371152 | 58223 | 1 | 0 |
| **107** | S | SF | 219990 | 37557 | 1 | 0 |
| **108** | S | SF | 432426 | 57422 | 1 | 0 |
| **109** | S | SF | 268000 | 43344 | 1 | 0 |
| **110** | S | SF | 312898 | 40768 | 1 | 0 |
| **111** | S | SF | 267250 | 45676 | 1 | 0 |
| **112** | S | SF | 379000 | 72915 | 1 | 0 |
| **113** | S | SF | 342423 | 48309 | 1 | 0 |
| **114** | S | SF | 337374 | 70399 | 1 | 0 |
| **115** | S | SF | 358162 | 44470 | 1 | 0 |
| **116** | S | SF | 398651 | 65429 | 1 | 0 |
| **117** | S | SF | 280804 | 40667 | 1 | 0 |
| **118** | S | SF | 407076 | 48668 | 1 | 0 |
| **119** | S | SF | 268500 | 41099 | 1 | 0 |
| **120** | S | SF | 444304 | 53938 | 1 | 0 |
| **121** | S | SF | 324266 | 47891 | 1 | 0 |
| **122** | S | SF | 307387 | 45850 | 1 | 0 |
| **123** | S | SF | 369101 | 46773 | 1 | 0 |
| **124** | S | SF | 350702 | 46386 | 1 | 0 |
| **125** | S | SF | 329611 | 48611 | 1 | 0 |
| **126** | S | SF | 242191 | 33434 | 1 | 0 |
| **127** | S | SF | 379424 | 64902 | 1 | 0 |
| **128** | S | SF | 324412 | 62523 | 1 | 0 |
| **129** | S | SF | 340730 | 50850 | 1 | 0 |
| **130** | S | SF | 310100 | 41800 | 1 | 0 |
| **131** | S | SF | 354117 | 56219 | 1 | 0 |
| **132** | S | SF | 330710 | 49920 | 1 | 0 |
| **133** | S | SF | 417790 | 63099 | 1 | 0 |
| **134** | S | SF | 290000 | 48300 | 1 | 0 |
| **135** | S | SF | 274903 | 45345 | 1 | 0 |
| **136** | S | SF | 209400 | 43579 | 1 | 0 |
| **137** | S | SF | 205821 | 39299 | 1 | 0 |
| **138** | S | SF | 287771 | 46300 | 1 | 0 |
| **139** | S | SF | 575120 | 79790 | 1 | 0 |
| **140** | S | SF | 226000 | 35600 | 1 | 0 |
| **141** | S | SF | 216049 | 35600 | 1 | 0 |
| **142** | S | SF | 207345 | 35600 | 1 | 0 |
| **143** | S | SF | 211797 | 34000 | 1 | 0 |
| **144** | S | SF | 204900 | 34000 | 1 | 0 |
| **145** | S | SF | 206400 | 35851 | 1 | 0 |
| **146** | S | SF | 186000 | 35851 | 1 | 0 |
| **147** | S | SF | 249900 | 38200 | 1 | 0 |
| **148** | S | SF | 214205 | 36500 | 1 | 0 |
| **149** | S | SF | 256235 | 48500 | 1 | 0 |
| **150** | S | SF | 262890 | 48500 | 1 | 0 |
| **151** | S | SF | 338065 | 54850 | 1 | 0 |
| **152** | S | SF | 326570 | 51000 | 1 | 0 |
| **153** | S | SF | 239000 | 39169 | 1 | 0 |
| **154** | S | SF | 239870 | 41354 | 1 | 0 |
| **155** | S | SF | 241195 | 41340 | 1 | 0 |
| **156** | S | SF | 252135 | 41341 | 1 | 0 |
| **157** | S | SF | 253055 | 41340 | 1 | 0 |
| **158** | S | SF | 160000 | 29500 | 1 | 0 |
| **159** | S | SF | 337380 | 49150 | 1 | 0 |
| **160** | S | SF | 492820 | 84122 | 1 | 0 |
| **161** | S | SF | 385000 | 75000 | 1 | 0 |
| **162** | S | SF | 340000 | 40000 | 1 | 0 |
| **163** | S | SF | 202000 | 31160 | 1 | 0 |
| **164** | S | SF | 234971 | 29202 | 1 | 0 |
| **165** | S | SF | 225900 | 28618 | 1 | 0 |
| **166** | S | SF | 366990 | 55508 | 1 | 0 |
| **167** | S | SF | 307663 | 44840 | 1 | 0 |
| **168** | S | SF | 379575 | 44294 | 1 | 0 |

|  |
| --- |
| correlation |

The REG Procedure

Model: MODEL1

Dependent Variable: Price

|  |  |
| --- | --- |
| **Number of Observations Read** | 168 |
| **Number of Observations Used** | 168 |

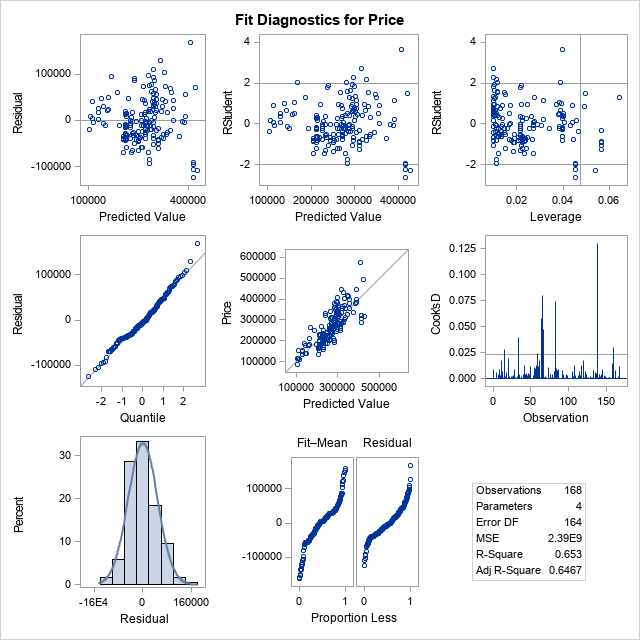
| **Parameter Estimates** | | | | | |
| --- | --- | --- | --- | --- | --- |
| **Variable** | **DF** | **Parameter Estimate** | **Standard Error** | **t Value** | **Pr > |t|** |
| **Intercept** | **1** | 53558 | 17961 | 2.98 | 0.0033 |
| **Cost** | **1** | 3.50527 | 0.29817 | 11.76 | <.0001 |
| **numtype** | **1** | 72566 | 9755.37424 | 7.44 | <.0001 |
| **numregion** | **1** | -9081.58833 | 9890.76247 | -0.92 | 0.3599 |

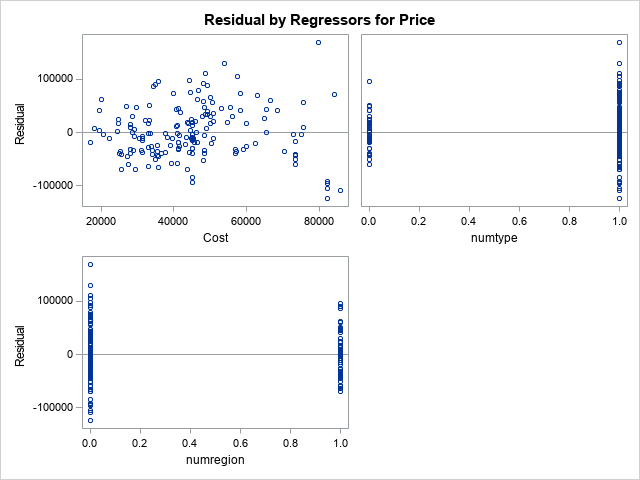
|  |
| --- |
| correlation |

The REG Procedure

Model: MODEL1

Dependent Variable: Price





|  |
| --- |
| correlation |

The REG Procedure

Model: MODEL1

Dependent Variable: Price

|  |  |
| --- | --- |
| **Number of Observations Read** | 168 |
| **Number of Observations Used** | 168 |

| **Analysis of Variance** | | | | | |
| --- | --- | --- | --- | --- | --- |
| **Source** | **DF** | **Sum of Squares** | **Mean Square** | **F Value** | **Pr > F** |
| **Model** | 2 | 7.347041E11 | 3.67352E11 | 154.07 | <.0001 |
| **Error** | 165 | 3.934247E11 | 2384392143 |  |  |
| **Corrected Total** | 167 | 1.128129E12 |  |  |  |

|  |  |  |  |
| --- | --- | --- | --- |
| **Root MSE** | 48830 | **R-Square** | 0.6513 |
| **Dependent Mean** | 267331 | **Adj R-Sq** | 0.6470 |
| **Coeff Var** | 18.26586 |  |  |

Model: MODEL1

Dependent Variable: Price

