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

**Assignment-4** | **Total Points: 26 (DSC 423 and DSC 324)**

**Due Date: 04/30/2019 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.***

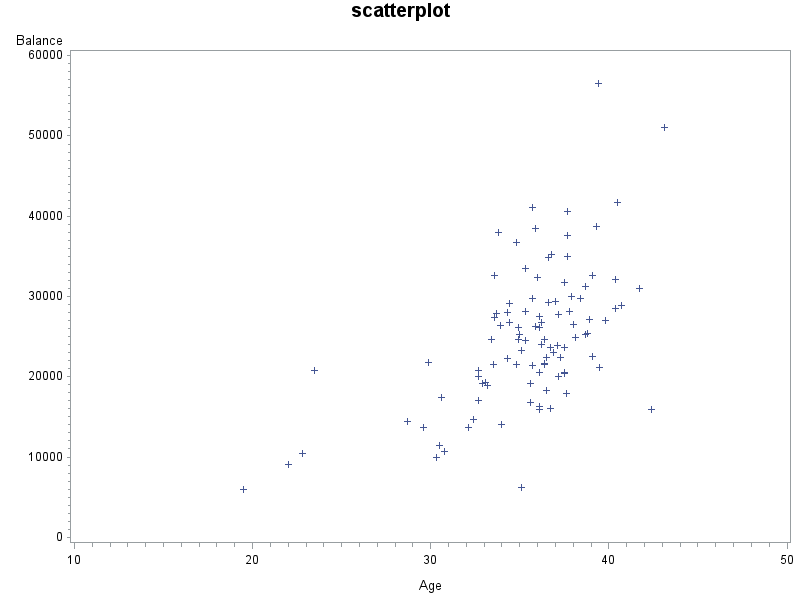
**PROBLEM 1 [16 pts] – to be answered by everyone**

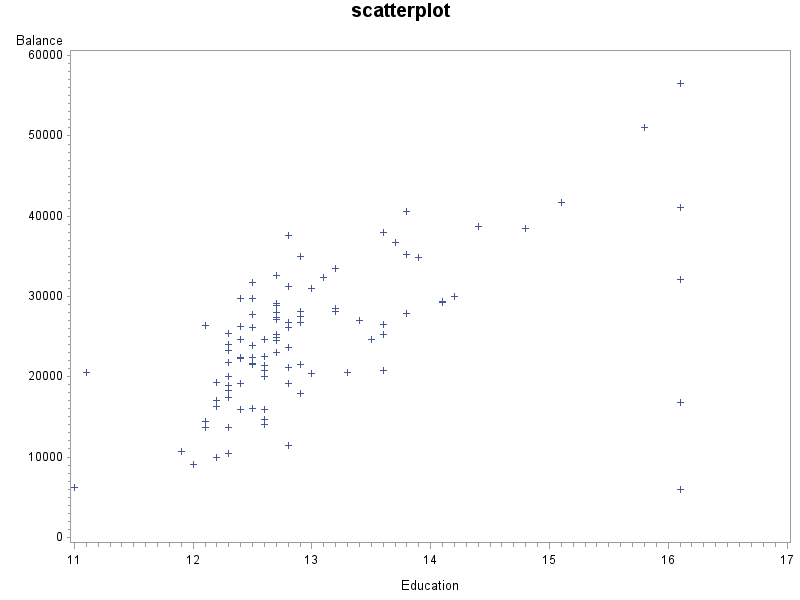
The file bankingfull.txt attached to this assignment contains the full dataset. You analyzed a smaller set for a previous assignment. It provides data acquired from banking and census records for different zip codes in the bank’s current market. Such information can be useful in targeting advertising for new customers or for choosing locations for branch offices. The data show

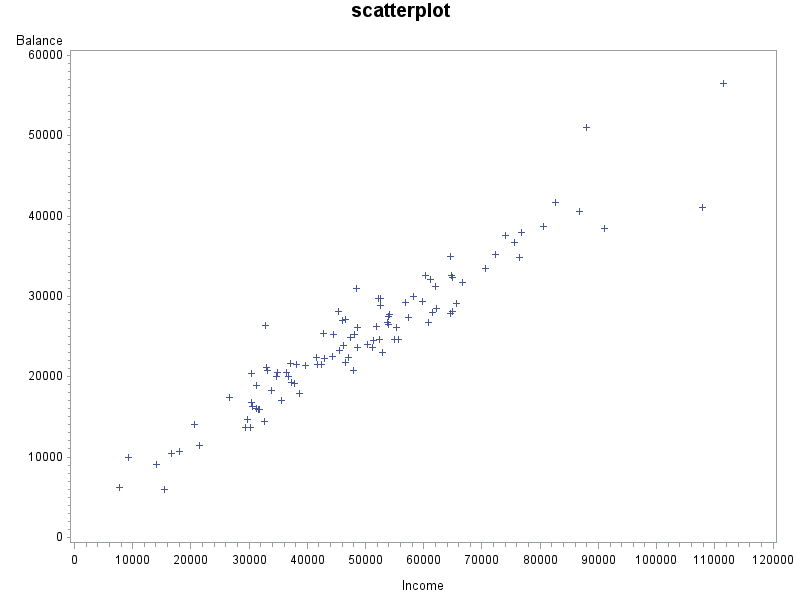
* median age of the population (AGE)
* median years of education (EDUCATION)
* median income (INCOME) in $
* median home value (HOMEVAL) in $
* median household wealth (WEALTH) in $
* average bank balance (BALANCE) in $

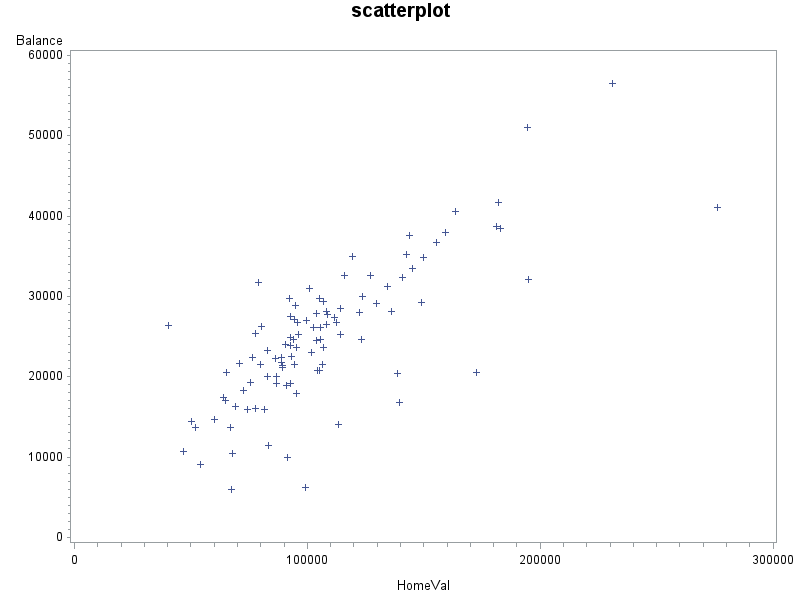
The goal of this exercise is to define a regression model to predict the average bank balance as a function of the other variables.

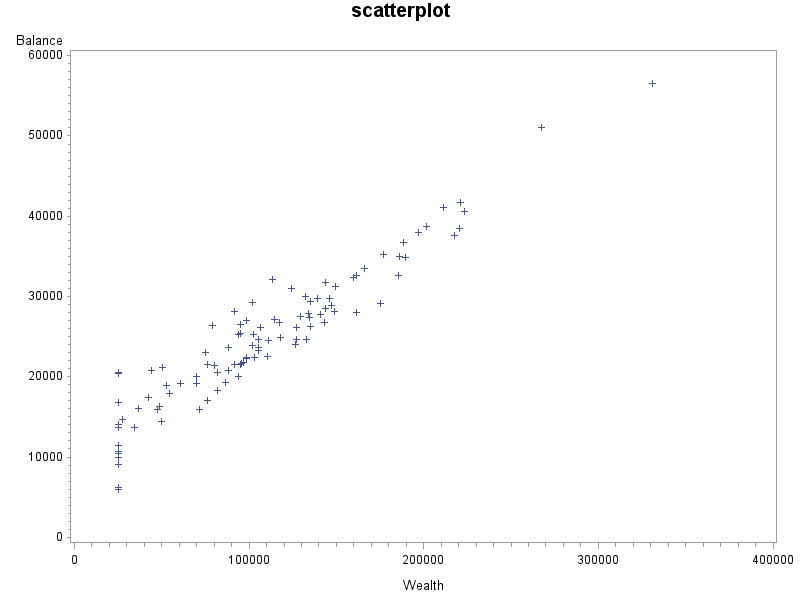
1. Create scatterplots to visualize the associations between bank balance and the other five variables. Include the relevant output. Discuss the patterns displayed by the scatterplot. Also, explain if the associations appear to be linear? (you can create either scatterplots or a matrix plot)









Shown 

Shown above are the scatterplots for bank balance compared with the variables of Age, Education, Income, HomeVal and Wealth. Upon first glance, Balance and Income and Balance and Wealth both seem to show a linear trend. As discussed in the previous assignment, balance and income have a close linear trend. Balance and Wealth seems to start off to have a random trend around the 25,000-35,000 x-value. However, after that, the trend seems to be linear. The scatterplot for Balance and HomeVal don’t seem to have show independence, whereas Balance and age can be said to show independence. These are assumptions at first glance; upon running certain statistical analysis we can make further assumptions.

1. Compute correlation values of bank balance vs the other variables. Include the relevant output. Interpret the correlation values, and discuss which variables appear to be strongly associated.

Balance and income have a value of 0.95168 and balance and wealth have a value of .94781. Income and wealth are also at .94667 and that makes sense because income and wealth are both highly correlated with bank balance as well. These are high values of correlation and go hand in hand with our scatterplots above showing the linear trends. Balance and homeVal are at .76639. This is also a good value, but not as high as wealth and income. Once we take out the insignificant variables and rerun the test, we may see a stronger correlation between balance and homeval.

| **Pearson Correlation Coefficients, N = 102  Prob > |r| under H0: Rho=0** | | | | | | |
| --- | --- | --- | --- | --- | --- | --- |
|  | **Balance** | **Age** | **Education** | **Income** | **HomeVal** | **Wealth** |
| **Balance** | |  | | --- | | 1.00000 | |  | | |  | | --- | | 0.56547 | | <.0001 | | |  | | --- | | 0.55488 | | <.0001 | | |  | | --- | | 0.95168 | | <.0001 | | |  | | --- | | 0.76639 | | <.0001 | | |  | | --- | | 0.94871 | | <.0001 | |
| **Age** | |  | | --- | | 0.56547 | | <.0001 | | |  | | --- | | 1.00000 | |  | | |  | | --- | | 0.17341 | | 0.0813 | | |  | | --- | | 0.47715 | | <.0001 | | |  | | --- | | 0.38649 | | <.0001 | | |  | | --- | | 0.46809 | | <.0001 | |
| **Education** | |  | | --- | | 0.55488 | | <.0001 | | |  | | --- | | 0.17341 | | 0.0813 | | |  | | --- | | 1.00000 | |  | | |  | | --- | | 0.57539 | | <.0001 | | |  | | --- | | 0.75352 | | <.0001 | | |  | | --- | | 0.46941 | | <.0001 | |
| **Income** | |  | | --- | | 0.95168 | | <.0001 | | |  | | --- | | 0.47715 | | <.0001 | | |  | | --- | | 0.57539 | | <.0001 | | |  | | --- | | 1.00000 | |  | | |  | | --- | | 0.79536 | | <.0001 | | |  | | --- | | 0.94667 | | <.0001 | |
| **HomeVal** | |  | | --- | | 0.76639 | | <.0001 | | |  | | --- | | 0.38649 | | <.0001 | | |  | | --- | | 0.75352 | | <.0001 | | |  | | --- | | 0.79536 | | <.0001 | | |  | | --- | | 1.00000 | |  | | |  | | --- | | 0.69848 | | <.0001 | |
| **Wealth** | |  | | --- | | 0.94871 | | <.0001 | | |  | | --- | | 0.46809 | | <.0001 | | |  | | --- | | 0.46941 | | <.0001 | | |  | | --- | | 0.94667 | | <.0001 | | |  | | --- | | 0.69848 | | <.0001 | | |  | | --- | | 1.00000 | |  | |

1. Fit a regression model of balance vs the other five variables (model M1). Compute the VIF statistics for each x-variable and analyze whether there is a problem of multicollinearity and take appropriate action. Include the relevant output. Discuss your answer.

|  |
| --- |
| regression model |

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** | 5 | 7235179873 | 1447035975 | 342.44 | <.0001 |
| **Error** | 96 | 405664272 | 4225669 |  |  |
| **Corrected Total** | 101 | 7640844145 |  |  |  |

|  |  |  |  |
| --- | --- | --- | --- |
| **Root MSE** | 2055.64333 | **R-Square** | 0.9469 |
| **Dependent Mean** | 24888 | **Adj R-Sq** | 0.9441 |
| **Coeff Var** | 8.25962 |  |  |

| **Parameter Estimates** | | | | | |
| --- | --- | --- | --- | --- | --- |
| **Variable** | **DF** | **Parameter Estimate** | **Standard Error** | **t Value** | **Pr > |t|** |
| **Intercept** | **1** | -10711 | 4260.97631 | -2.51 | 0.0136 |
| **Age** | **1** | 318.66496 | 60.98611 | 5.23 | <.0001 |
| **Education** | **1** | 621.86035 | 318.95952 | 1.95 | 0.0541 |
| **Income** | **1** | 0.14632 | 0.04078 | 3.59 | 0.0005 |
| **HomeVal** | **1** | 0.00918 | 0.01104 | 0.83 | 0.4075 |
| **Wealth** | **1** | 0.07433 | 0.01119 | 6.64 | <.0001 |

Before delving too far into these numbers. It is important to look at the p-value’s to see if there are any values that are above .05. Education falls at .0541 and this is above the .05 value. HomeVal is at .4075 and this is way above the p-value as well. Since homeval is so far from the .05 value, we can take this variable out first and re-run the test to see if education is still significant or not.

|  |
| --- |
| regression model take away homeVal |

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** | 4 | 7232255152 | 1808063788 | 429.24 | <.0001 |
| **Error** | 97 | 408588992 | 4212258 |  |  |
| **Corrected Total** | 101 | 7640844145 |  |  |  |

|  |  |  |  |
| --- | --- | --- | --- |
| **Root MSE** | 2052.37854 | **R-Square** | 0.9465 |
| **Dependent Mean** | 24888 | **Adj R-Sq** | 0.9443 |
| **Coeff Var** | 8.24650 |  |  |

| **Parameter Estimates** | | | | | | | |
| --- | --- | --- | --- | --- | --- | --- | --- |
| **Variable** | **DF** | **Parameter Estimate** | **Standard Error** | **t Value** | **Pr > |t|** | **Tolerance** | **Variance Inflation** |
| **Intercept** | **1** | -12432 | 3718.67432 | -3.34 | 0.0012 | . | 0 |
| **Age** | **1** | 325.06528 | 60.40284 | 5.38 | <.0001 | 0.75677 | 1.32140 |
| **Education** | **1** | 773.38004 | 261.43309 | 2.96 | 0.0039 | 0.60397 | 1.65571 |
| **Income** | **1** | 0.15975 | 0.03739 | 4.27 | <.0001 | 0.07956 | 12.56885 |
| **Wealth** | **1** | 0.07299 | 0.01105 | 6.60 | <.0001 | 0.09532 | 10.49135 |

Upon re-running the test, we see that education is now below .05. The r-value actually decreased from .9469 to .9465 but the adjusted r2 value went up from .9441 to .9443. An issue of multicollinearity can be diagnosed between Balance and income and balance and wealth. If more data is taken, this issue can be resolved. VIF >= 10 suggest collinearity. Income and wealth both are above 10, so this suggest collinearity. This goes hand in hand with what was stated in part a and b about income and wealth.

1. Apply your knowledge of regression analysis to define a better model M2. Include the SAS output for both models and answer the following questions :
   1. Analyze the adj-R2 values for both models M1 and M2. Which model has the largest adj-R2 value?

In the original model, the adj-R2 value was .9441 and in model 2 it is .9443. The r value dropped but the adjusted r2 value went up.

* 1. Create residual plots for M2 (standardized residuals vs predicted; standardized residuals vs x-variables; and normal plot of residuals). Analyze the residual plots to check if the regression model assumptions are met by the data. Include the relevant output and discuss your analysis.

Upon looking at the residual plots, it seems that the model assumptions are met by the data. The residual plots are all scattered around the x axis and the quantile and predicted values seem to show a linear trend. The residual by regressors for balance plots show the data scattered along the x axis and based on these plots we can assume the assumptions have been met.

* 1. Analyze if there are any outliers and/or influential points for your M2 model. If so, what actions would you take to address this issue? Make sure to implement any actions you specify here. Include the relevant output.

There seems to be potential outliers when looking at the residual by regressors for balance plots. On the plots for age education and wealth, we can see potential outliers for all three variables. These could be due to data entry errors or measurement errors. However, with the variables we have, there are always chances of potential outliers due to different circumstances for different individuals. These circumstances cannot be accounted for and need to remain random in order to have a proper random sample.

* 1. Compute the standardized coefficients for M2 and discuss which predictor has the strongest influence on balance? Include the relevant output.

We can replace all the variables in the regression with its standardized equivalent and subtract the mean of every observation and divide it by SD. We can see that wealth has the highest value and would have the strongest influence

|  |
| --- |
| regression model take away homeVal |

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** | 3 | 7155379617 | 2385126539 | 481.48 | <.0001 |
| **Error** | 98 | 485464527 | 4953720 |  |  |
| **Corrected Total** | 101 | 7640844145 |  |  |  |

|  |  |  |  |
| --- | --- | --- | --- |
| **Root MSE** | 2225.69532 | **R-Square** | 0.9365 |
| **Dependent Mean** | 24888 | **Adj R-Sq** | 0.9345 |
| **Coeff Var** | 8.94289 |  |  |

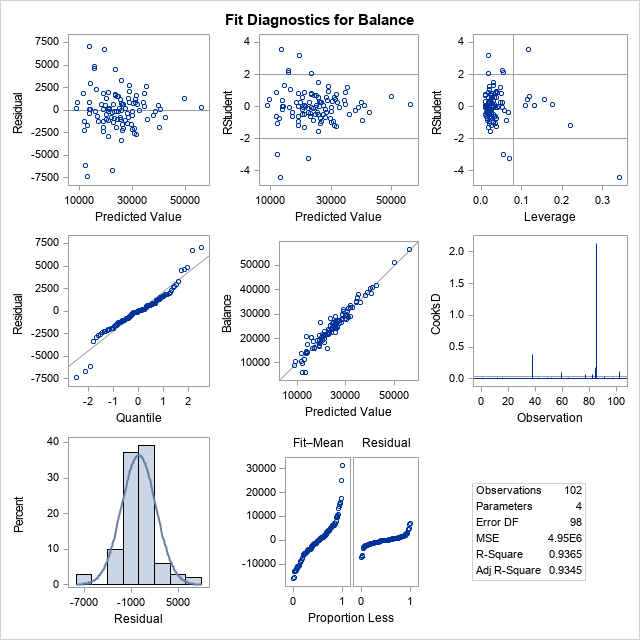
| **Parameter Estimates** | | | | | | |
| --- | --- | --- | --- | --- | --- | --- |
| **Variable** | **DF** | **Parameter Estimate** | **Standard Error** | **t Value** | **Pr > |t|** | **Standardized Estimate** |
| **Intercept** | **1** | -17732 | 3801.66282 | -4.66 | <.0001 | 0 |
| **Age** | **1** | 367.82141 | 64.59824 | 5.69 | <.0001 | 0.16436 |
| **Education** | **1** | 1300.30871 | 249.97314 | 5.20 | <.0001 | 0.15027 |
| **Wealth** | **1** | 0.11647 | 0.00468 | 24.89 | <.0001 | 0.80124 |

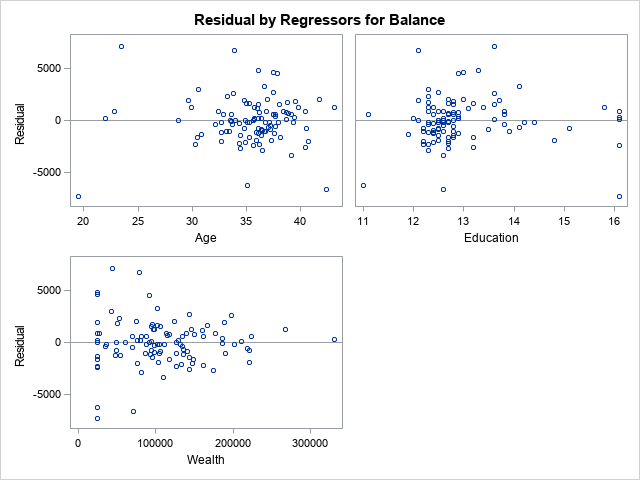
|  |
| --- |
| regression model take away homeVal |

The REG Procedure

Model: MODEL2

Dependent Variable: Balance





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

**proc** **import** datafile = "bankingFull.txt" out= banking replace;

deimiter= '09'x;

getames = yes;

dataow= **2**;

**run**;

title "print bankingFull dataset";

**proc** **print**;

**run**;

title "scatterplot";

**PROC** **GPLOT**;

PLOT Balance\*(Age Education Income HomeVal Wealth);

**RUN**;

title "Correlation";

**Proc** **CORR**;

var Balance Age Education Income HomeVal Wealth;

**run**;

title "regression model";

**PROC** **reg**;

model balance= age Education Income HomeVal Wealth;

**run**;

title "regression model take away homeVal";

**PROC** **reg**;

model balance= age Education Income Wealth /vif tol;

**run**;

**PROC** **reg**;

model balance= age Education Wealth / stb;

**run**;

**Problem 2 [10 pts] – to be answered by everyone**

Analytics is used in many different sports and has become popular with the Money Ball movie. The pgatour2006.csv dataset contains data about 196 tour players in 2006. The variables in the dataset are:

* Player’s name
* PrizeMoney = average prize money per tournament

And a set of metrics that evaluate the quality of a player’s game.

* DrivingAccuracy = percent of times a player is able to hit the fairway with his tee shot
* GIR = percent of time a player was able to hit the green within two or less than par (Greens in Regulation)
* BirdieConversion = percentage of times a player makes a birdie or better after hitting the green in regulation
* PuttingAverage = putting performance on those holes where the green was hit in regulation.
* PuttsPerRound= average number of putts per round (shots played on the green)

You are asked to build a model for PrizeMoney using the remaining predictors, and to evaluate the relative importance of each different aspects of a player’s game on the average prize money.

**Note:** For the non-golfers in the class, you can refer to this page for an explanation of the terms:

<http://en.wikipedia.org/wiki/Glossary_of_golf>

**SAS Code to Import the data**

\*import data from file;

**proc** **import** datafile="pgatour2006.csv" out=MyPga replace;

delimiter=',';

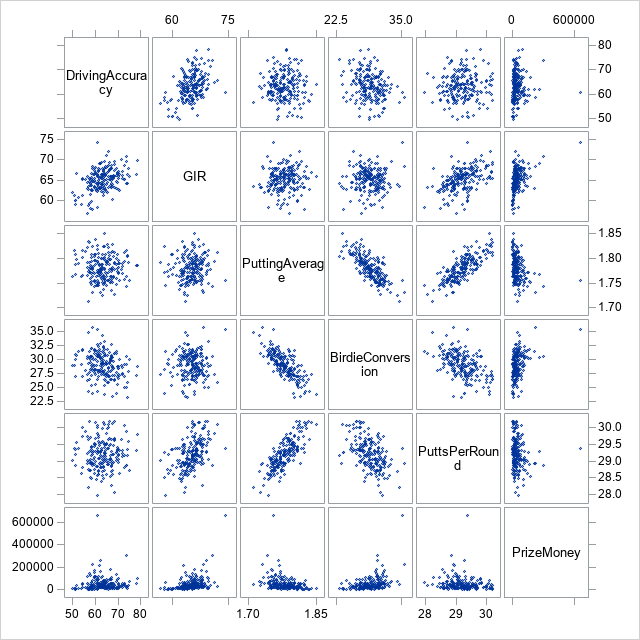
getnames=yes;

**run**;

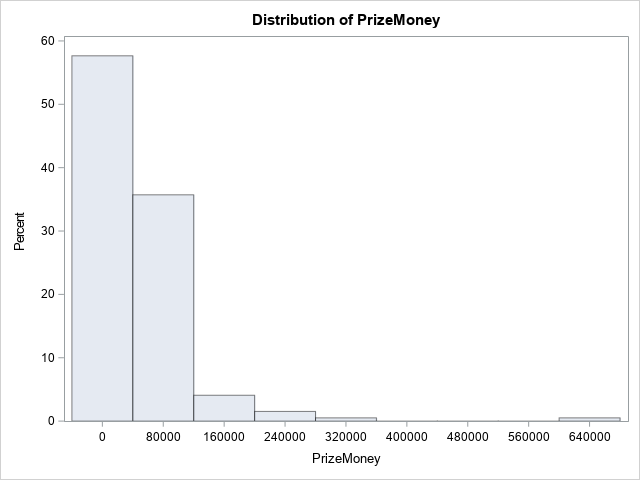
Note:

* The data file is in CSV format
* It is delimitered with a comma
* The SAS dataset it is writing into is MyPga. You can change the name if you like.

1. Create scatterplots to visualize the associations between PrizeMoney and the other 5 variables. Discuss the patterns displayed by the scatterplot. Also, explain if the associations appear to be linear? (you can create scatterplots or a matrix plot). Include the relevant output.



1. Analyze distribution of PrizeMoney, and discuss if the distribution is symmetric or skewed. Include the relevant output.



The graph above shows it is skewed right. There is a potential outlier at 640,000. This makes sense because a lot of the talent in the tournament is from people trying to make it to that large prize money value, and there are some golfers who are experts in their fields and make a lot of money because of it. Also, since there are a select few in the study that make a lot more than everyone else, the distribution will be a little off to accont for this. The inner quartile is a large amount away from the max and this proves this point.

1. Apply a log transformation to PrizeMoney and compute the new variable ln\_Prize=log(PrizeMoney). Analyze distribution of ln\_Prize, and discuss if the distribution is symmetric or skewed. Include the relevant output.

We can see r2 and adj r2 is at .5 and this is not a great value to be at. Putting\_average is also above .05 so we would have to remove putting average and rerun this test again to see if this changes anything.

The REG Procedure

Model: MODEL1

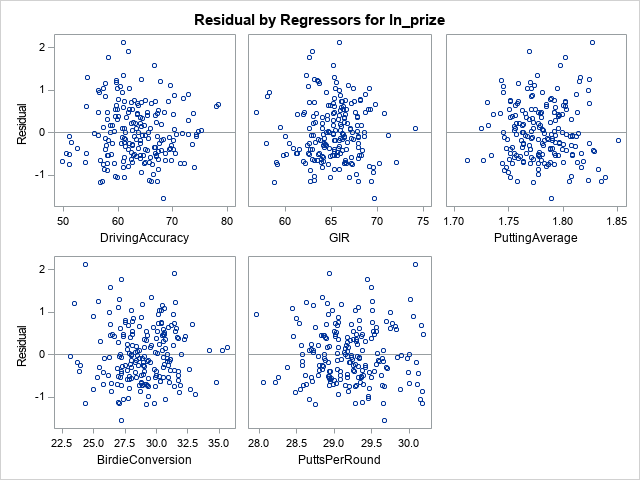
Dependent Variable: ln\_prize

|  |  |
| --- | --- |
| **Number of Observations Read** | 196 |
| **Number of Observations Used** | 196 |

| **Analysis of Variance** | | | | | |
| --- | --- | --- | --- | --- | --- |
| **Source** | **DF** | **Sum of Squares** | **Mean Square** | **F Value** | **Pr > F** |
| **Model** | 5 | 101.43308 | 20.28662 | 44.86 | <.0001 |
| **Error** | 190 | 85.92207 | 0.45222 |  |  |
| **Corrected Total** | 195 | 187.35515 |  |  |  |

|  |  |  |  |
| --- | --- | --- | --- |
| **Root MSE** | 0.67247 | **R-Square** | 0.5414 |
| **Dependent Mean** | 10.37808 | **Adj R-Sq** | 0.5293 |
| **Coeff Var** | 6.47975 |  |  |

| **Parameter Estimates** | | | | | |
| --- | --- | --- | --- | --- | --- |
| **Variable** | **DF** | **Parameter Estimate** | **Standard Error** | **t Value** | **Pr > |t|** |
| **Intercept** | **1** | 8.24102 | 7.16112 | 1.15 | 0.2513 |
| **DrivingAccuracy** | **1** | -0.00075836 | 0.01161 | -0.07 | 0.9480 |
| **GIR** | **1** | 0.26879 | 0.02879 | 9.33 | <.0001 |
| **PuttingAverage** | **1** | 8.74678 | 5.37342 | 1.63 | 0.1052 |
| **BirdieConversion** | **1** | 0.15230 | 0.04083 | 3.73 | 0.0003 |
| **PuttsPerRound** | **1** | -1.20948 | 0.26728 | -4.53 | <.0001 |
|  |  |  |  |  |  |

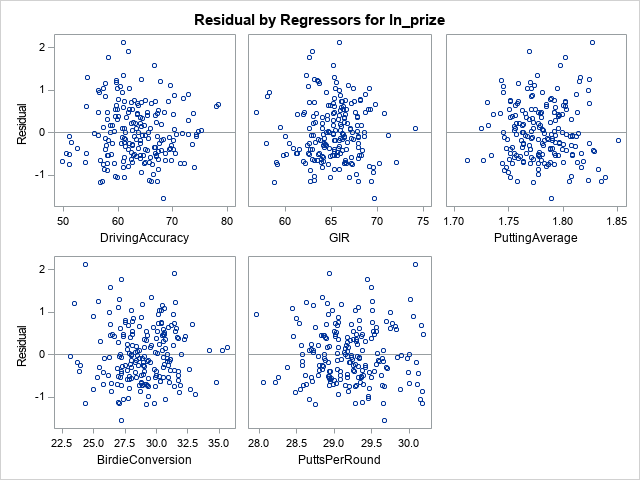


We can see from this that the points are along the x axis and seem to be randomly dispersed. An argument can be made to for this to be unimodal and not skewed any way but we can also make an argument for skewed right.

1. Fit a regression model of ln\_Prize using the remaining predictors in your dataset. Apply your knowledge of regression analysis to define a valid model to predict ln\_Prize. Include the outputs for all the questions below before you analyze them.
   * 1. If necessary remove the non-significant variables. Remember to remove one variable at a time (variable with largest p-value is removed first) and refit the model, until all variables are significant.

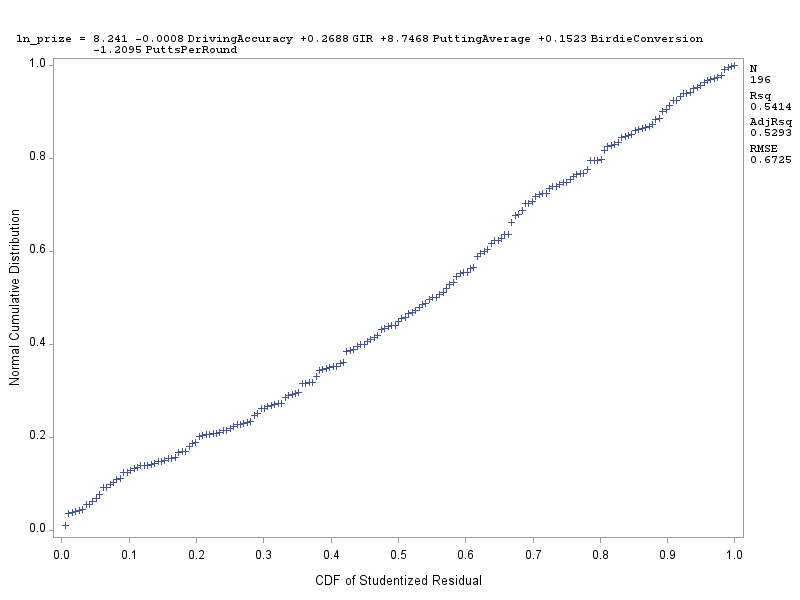
The results for this are shown below the SAS code for this document.

* + 1. Analyze residual plots to check if the regression model is valid for your data. Discuss your analysis.



The residual plots are all scattered around the x axis and the quantile and predicted values seem to show no consistent linear trend. The residual by regressors for balance plots show the data scattered along the x axis and based on these plots we can assume the assumptions have not been met until we continue to take away variables that are not valuable to the final equation model.

* + 1. Analyze if there are any outliers and/or influential points. If there are points in the dataset that need to be investigated, give one or more reason to support each point chosen. Take appropriate action(s) to implement it. Include the relevant outputs. Discuss your answer.



Based on this plot, there are no apparent outliers. If this was made before the final model, there are a lot of potential outliers and this may be because of such variance in skill level in the different golfers.

* + 1. Write down the final model equation. Discuss why this is the best model. Include all relevant statistics/values to substantiate your answer.

Y=.26789gir-1.209048puttsperround is the best model for this. All other variables can be deemed insignificant and can be taken out from the equation.

1. Interpret the regression coefficients in the final model to answer the following question: How does an increase in 1% for GIR affect the average Prize money?

A 1% increase in GIR would result in an additional 1.056% increase in prize money.

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

**proc** **import** datafile = "pgatour2006.csv" out=MyPga replace;

delimiter=',';

getnames=yes;

**run**;

\*part a;

**proc** **sgscatter**;

matrix DrivingAccuracy GIR PuttingAverage BirdieConversion PuttsPerRound PrizeMoney;

**run**;

\*part b;

**proc** **univariate** data=MyPga;

histogram PrizeMoney;

**run**;

\*part c;

**data** new;

set MyPga;

ln\_prize = log(PrizeMoney);

**run**;

**proc** **print**;

**run**;

**proc** **reg** data=new;

model ln\_prize= DrivingAccuracy GIR PuttingAverage BirdieConversion PuttsPerRound;

plot npp.\*student.;

**run**;

Part b:

|  |
| --- |
| The SAS System |

The UNIVARIATE Procedure

Variable: PrizeMoney

| **Moments** | | | |
| --- | --- | --- | --- |
| **N** | 196 | **Sum Weights** | 196 |
| **Mean** | 50891.1684 | **Sum Observations** | 9974669 |
| **Std Deviation** | 63902.9534 | **Variance** | 4083587455 |
| **Skewness** | 5.37634272 | **Kurtosis** | 44.1893165 |
| **Uncorrected SS** | 1.30392E12 | **Corrected SS** | 7.963E11 |
| **Coeff Variation** | 125.567865 | **Std Error Mean** | 4564.49667 |

| **Basic Statistical Measures** | | | |
| --- | --- | --- | --- |
| **Location** | | **Variability** | |
| **Mean** | 50891.17 | **Std Deviation** | 63903 |
| **Median** | 36644.50 | **Variance** | 4083587455 |
| **Mode** | . | **Range** | 660531 |
|  |  | **Interquartile Range** | 40785 |

| **Tests for Location: Mu0=0** | | | | |
| --- | --- | --- | --- | --- |
| **Test** | **Statistic** | | **p Value** | |
| **Student's t** | **t** | 11.14935 | **Pr > |t|** | <.0001 |
| **Sign** | **M** | 98 | **Pr >= |M|** | <.0001 |
| **Signed Rank** | **S** | 9653 | **Pr >= |S|** | <.0001 |

| **Quantiles (Definition 5)** | |
| --- | --- |
| **Level** | **Quantile** |
| **100% Max** | 662771.0 |
| **99%** | 300555.0 |
| **95%** | 145414.0 |
| **90%** | 107294.0 |
| **75% Q3** | 58006.5 |
| **50% Median** | 36644.5 |
| **25% Q1** | 17221.5 |
| **10%** | 9149.0 |
| **5%** | 5285.0 |
| **1%** | 2426.0 |
| **0% Min** | 2240.0 |

| **Extreme Observations** | | | |
| --- | --- | --- | --- |
| **Lowest** | | **Highest** | |
| **Value** | **Obs** | **Value** | **Obs** |
| 2240 | 70 | 217748 | 63 |
| 2426 | 101 | 224027 | 142 |
| 2692 | 97 | 262045 | 2 |
| 3025 | 115 | 300555 | 90 |
| 3635 | 3 | 662771 | 178 |

|  |
| --- |
| The SAS System |

The UNIVARIATE Procedure

Variable: DrivingAccuracy

| **Moments** | | | |
| --- | --- | --- | --- |
| **N** | 196 | **Sum Weights** | 196 |
| **Mean** | 63.380102 | **Sum Observations** | 12422.5 |
| **Std Deviation** | 5.41302276 | **Variance** | 29.3008154 |
| **Skewness** | 0.09568709 | **Kurtosis** | 0.09212821 |
| **Uncorrected SS** | 793052.977 | **Corrected SS** | 5713.659 |
| **Coeff Variation** | 8.54057123 | **Std Error Mean** | 0.38664448 |

| **Basic Statistical Measures** | | | |
| --- | --- | --- | --- |
| **Location** | | **Variability** | |
| **Mean** | 63.38010 | **Std Deviation** | 5.41302 |
| **Median** | 63.24000 | **Variance** | 29.30082 |
| **Mode** | 56.39000 | **Range** | 28.68000 |
|  |  | **Interquartile Range** | 7.22500 |

|  |
| --- |
| **Note: The mode displayed is the smallest of 9 modes with a count of 2.** |

| **Tests for Location: Mu0=0** | | | | |
| --- | --- | --- | --- | --- |
| **Test** | **Statistic** | | **p Value** | |
| **Student's t** | **t** | 163.9235 | **Pr > |t|** | <.0001 |
| **Sign** | **M** | 98 | **Pr >= |M|** | <.0001 |
| **Signed Rank** | **S** | 9653 | **Pr >= |S|** | <.0001 |

| **Quantiles (Definition 5)** | |
| --- | --- |
| **Level** | **Quantile** |
| **100% Max** | 78.430 |
| **99%** | 78.010 |
| **95%** | 72.900 |
| **90%** | 70.190 |
| **75% Q3** | 66.970 |
| **50% Median** | 63.240 |
| **25% Q1** | 59.745 |
| **10%** | 57.250 |
| **5%** | 54.130 |
| **1%** | 50.540 |
| **0% Min** | 49.750 |

| **Extreme Observations** | | | |
| --- | --- | --- | --- |
| **Lowest** | | **Highest** | |
| **Value** | **Obs** | **Value** | **Obs** |
| 49.75 | 91 | 74.33 | 113 |
| 50.54 | 78 | 74.67 | 68 |
| 51.03 | 154 | 75.23 | 159 |
| 51.09 | 48 | 78.01 | 61 |
| 51.12 | 3 | 78.43 | 92 |

|  |
| --- |
| The SAS System |

The UNIVARIATE Procedure

Variable: GIR

| **Moments** | | | |
| --- | --- | --- | --- |
| **N** | 196 | **Sum Weights** | 196 |
| **Mean** | 65.1860714 | **Sum Observations** | 12776.47 |
| **Std Deviation** | 2.72236386 | **Variance** | 7.411265 |
| **Skewness** | -0.2497787 | **Kurtosis** | 0.76380611 |
| **Uncorrected SS** | 834293.083 | **Corrected SS** | 1445.19668 |
| **Coeff Variation** | 4.17629687 | **Std Error Mean** | 0.19445456 |

| **Basic Statistical Measures** | | | |
| --- | --- | --- | --- |
| **Location** | | **Variability** | |
| **Mean** | 65.18607 | **Std Deviation** | 2.72236 |
| **Median** | 65.35500 | **Variance** | 7.41127 |
| **Mode** | 65.46000 | **Range** | 17.28000 |
|  |  | **Interquartile Range** | 3.28500 |

| **Tests for Location: Mu0=0** | | | | |
| --- | --- | --- | --- | --- |
| **Test** | **Statistic** | | **p Value** | |
| **Student's t** | **t** | 335.2252 | **Pr > |t|** | <.0001 |
| **Sign** | **M** | 98 | **Pr >= |M|** | <.0001 |
| **Signed Rank** | **S** | 9653 | **Pr >= |S|** | <.0001 |

| **Quantiles (Definition 5)** | |
| --- | --- |
| **Level** | **Quantile** |
| **100% Max** | 74.150 |
| **99%** | 72.030 |
| **95%** | 69.320 |
| **90%** | 68.290 |
| **75% Q3** | 66.800 |
| **50% Median** | 65.355 |
| **25% Q1** | 63.515 |
| **10%** | 62.070 |
| **5%** | 60.060 |
| **1%** | 57.910 |
| **0% Min** | 56.870 |

| **Extreme Observations** | | | |
| --- | --- | --- | --- |
| **Lowest** | | **Highest** | |
| **Value** | **Obs** | **Value** | **Obs** |
| 56.87 | 184 | 70.20 | 26 |
| 57.91 | 70 | 70.71 | 90 |
| 58.05 | 18 | 71.15 | 100 |
| 58.26 | 1 | 72.03 | 83 |
| 58.85 | 115 | 74.15 | 178 |

|  |
| --- |
| The SAS System |

The UNIVARIATE Procedure

Variable: PuttingAverage

| **Moments** | | | |
| --- | --- | --- | --- |
| **N** | 196 | **Sum Weights** | 196 |
| **Mean** | 1.77985204 | **Sum Observations** | 348.851 |
| **Std Deviation** | 0.02472813 | **Variance** | 0.00061148 |
| **Skewness** | 0.15864479 | **Kurtosis** | -0.1870804 |
| **Uncorrected SS** | 621.022403 | **Corrected SS** | 0.11923871 |
| **Coeff Variation** | 1.38933644 | **Std Error Mean** | 0.0017663 |

| **Basic Statistical Measures** | | | |
| --- | --- | --- | --- |
| **Location** | | **Variability** | |
| **Mean** | 1.779852 | **Std Deviation** | 0.02473 |
| **Median** | 1.778000 | **Variance** | 0.0006115 |
| **Mode** | 1.777000 | **Range** | 0.13900 |
|  |  | **Interquartile Range** | 0.03350 |

| **Tests for Location: Mu0=0** | | | | |
| --- | --- | --- | --- | --- |
| **Test** | **Statistic** | | **p Value** | |
| **Student's t** | **t** | 1007.675 | **Pr > |t|** | <.0001 |
| **Sign** | **M** | 98 | **Pr >= |M|** | <.0001 |
| **Signed Rank** | **S** | 9653 | **Pr >= |S|** | <.0001 |

| **Quantiles (Definition 5)** | |
| --- | --- |
| **Level** | **Quantile** |
| **100% Max** | 1.8510 |
| **99%** | 1.8390 |
| **95%** | 1.8240 |
| **90%** | 1.8130 |
| **75% Q3** | 1.7965 |
| **50% Median** | 1.7780 |
| **25% Q1** | 1.7630 |
| **10%** | 1.7510 |
| **5%** | 1.7440 |
| **1%** | 1.7250 |
| **0% Min** | 1.7120 |

| **Extreme Observations** | | | |
| --- | --- | --- | --- |
| **Lowest** | | **Highest** | |
| **Value** | **Obs** | **Value** | **Obs** |
| 1.712 | 45 | 1.829 | 131 |
| 1.725 | 191 | 1.833 | 151 |
| 1.727 | 25 | 1.835 | 101 |
| 1.730 | 49 | 1.839 | 125 |
| 1.731 | 142 | 1.851 | 177 |

|  |
| --- |
| The SAS System |

The UNIVARIATE Procedure

Variable: BirdieConversion

| **Moments** | | | |
| --- | --- | --- | --- |
| **N** | 196 | **Sum Weights** | 196 |
| **Mean** | 28.9822959 | **Sum Observations** | 5680.53 |
| **Std Deviation** | 2.20655581 | **Variance** | 4.86888855 |
| **Skewness** | -0.0219271 | **Kurtosis** | 0.33102122 |
| **Uncorrected SS** | 165584.235 | **Corrected SS** | 949.433267 |
| **Coeff Variation** | 7.61346105 | **Std Error Mean** | 0.15761113 |

| **Basic Statistical Measures** | | | |
| --- | --- | --- | --- |
| **Location** | | **Variability** | |
| **Mean** | 28.98230 | **Std Deviation** | 2.20656 |
| **Median** | 29.01000 | **Variance** | 4.86889 |
| **Mode** | 27.17000 | **Range** | 12.49000 |
|  |  | **Interquartile Range** | 3.05000 |

|  |
| --- |
| **Note: The mode displayed is the smallest of 3 modes with a count of 3.** |

| **Tests for Location: Mu0=0** | | | | |
| --- | --- | --- | --- | --- |
| **Test** | **Statistic** | | **p Value** | |
| **Student's t** | **t** | 183.8848 | **Pr > |t|** | <.0001 |
| **Sign** | **M** | 98 | **Pr >= |M|** | <.0001 |
| **Signed Rank** | **S** | 9653 | **Pr >= |S|** | <.0001 |

| **Quantiles (Definition 5)** | |
| --- | --- |
| **Level** | **Quantile** |
| **100% Max** | 35.660 |
| **99%** | 35.260 |
| **95%** | 32.230 |
| **90%** | 31.530 |
| **75% Q3** | 30.555 |
| **50% Median** | 29.010 |
| **25% Q1** | 27.505 |
| **10%** | 26.340 |
| **5%** | 25.280 |
| **1%** | 23.430 |
| **0% Min** | 23.170 |

| **Extreme Observations** | | | |
| --- | --- | --- | --- |
| **Lowest** | | **Highest** | |
| **Value** | **Obs** | **Value** | **Obs** |
| 23.17 | 113 | 33.09 | 102 |
| 23.43 | 40 | 34.26 | 191 |
| 23.71 | 177 | 34.75 | 45 |
| 23.86 | 119 | 35.26 | 178 |
| 23.92 | 70 | 35.66 | 142 |

|  |
| --- |
| The SAS System |

The UNIVARIATE Procedure

Variable: PuttsPerRound

| **Moments** | | | |
| --- | --- | --- | --- |
| **N** | 196 | **Sum Weights** | 196 |
| **Mean** | 29.2010714 | **Sum Observations** | 5723.41 |
| **Std Deviation** | 0.44170227 | **Variance** | 0.1951009 |
| **Skewness** | 0.13023768 | **Kurtosis** | -0.0440591 |
| **Uncorrected SS** | 167167.749 | **Corrected SS** | 38.044675 |
| **Coeff Variation** | 1.51262351 | **Std Error Mean** | 0.03155016 |

| **Basic Statistical Measures** | | | |
| --- | --- | --- | --- |
| **Location** | | **Variability** | |
| **Mean** | 29.20107 | **Std Deviation** | 0.44170 |
| **Median** | 29.19000 | **Variance** | 0.19510 |
| **Mode** | 28.93000 | **Range** | 2.23000 |
|  |  | **Interquartile Range** | 0.57500 |

|  |
| --- |
| **Note: The mode displayed is the smallest of 4 modes with a count of 4.** |

| **Tests for Location: Mu0=0** | | | | |
| --- | --- | --- | --- | --- |
| **Test** | **Statistic** | | **p Value** | |
| **Student's t** | **t** | 925.5443 | **Pr > |t|** | <.0001 |
| **Sign** | **M** | 98 | **Pr >= |M|** | <.0001 |
| **Signed Rank** | **S** | 9653 | **Pr >= |S|** | <.0001 |

| **Quantiles (Definition 5)** | |
| --- | --- |
| **Level** | **Quantile** |
| **100% Max** | 30.190 |
| **99%** | 30.180 |
| **95%** | 30.060 |
| **90%** | 29.810 |
| **75% Q3** | 29.485 |
| **50% Median** | 29.190 |
| **25% Q1** | 28.910 |
| **10%** | 28.670 |
| **5%** | 28.540 |
| **1%** | 28.050 |
| **0% Min** | 27.960 |

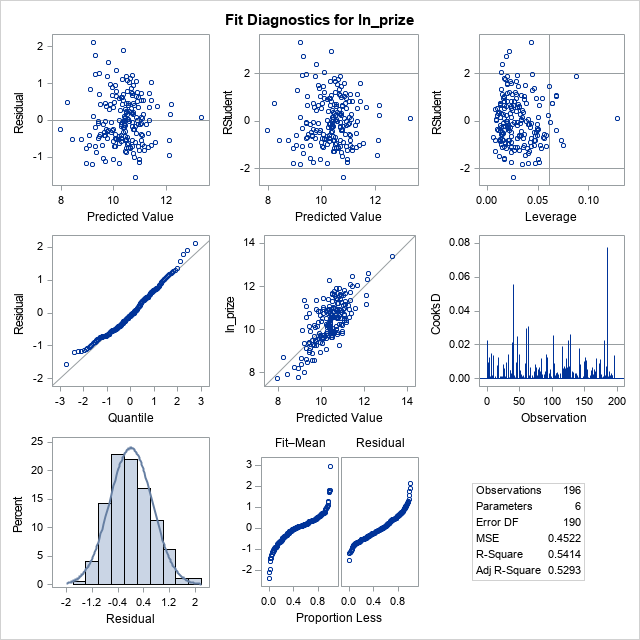
| **Extreme Observations** | | | |
| --- | --- | --- | --- |
| **Lowest** | | **Highest** | |
| **Value** | **Obs** | **Value** | **Obs** |
| 27.96 | 1 | 30.15 | 125 |
| 28.05 | 25 | 30.16 | 123 |
| 28.21 | 45 | 30.18 | 39 |
| 28.26 | 172 | 30.18 | 151 |
| 28.38 | 138 | 30.19 | 108 |

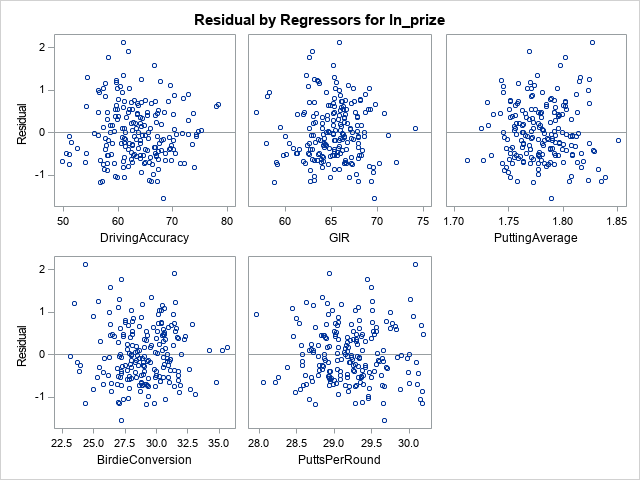
|  |
| --- |
| The SAS System |

The REG Procedure

Model: MODEL1

Dependent Variable: ln\_prize





The REG Procedure

