**Fall 2016**

**BSTATS ASSIGNMENT 2**

**Group No. 2**

**Q1. Find outliers in X, outliers in Y and influential points. You do not need to remove them just list a few of the ones found and indicate how you determine the points were outliers and/or influential.**

We ran Regression on our data to determine if there are any outliers.

Our Training data consists 7 independent(explanatory) variables :

1. **DVSS**
2. **DV1B**
3. **DVDH**
4. **R**
5. **HR**
6. **RBI**
7. **AVG**

And a Dependent variable**:**

1. **Salary**

k=7

n=231

* **Detecting outlying X values: the hat matrix**

Values above 2\*(k+1)/n are considered outliers.

Thus, h salary values>0.0692 were detected as outliers.



* **Detecting outlying Y values: studentized deleted residuals**

Uses a t-test approach. Any value larger than a t-table value with n-k-2 degrees of freedom

t-table value for n-k-2,i.e 222,= 1.9707

Hence, rstudent salary values > 1.9707 and rstudent salary values < -1.9707 will be detected as outliers.



* **Alternative approach to detecting influential observations: Cook’s Distance**

Value above the fiftieth (50) percentile of an F distribution k+1 and n-k-1 degrees of freedom are considered influential.

7+ 1 = 10 and 231 – 7 – 1 = 223 gives us a value of 2.08

Cook’s distance= 0.92078, hence cookd salary values > 0.92078 will be considered as influential observations.



Since, none of the values were greater than Cook’s distance, no influential observations were detected.

The results can be seen in [**Appendix A**](#Appendix_A)

**Q2. Check for assumption violations**

**a. if necessary use transformation to fix any assumptions that you note. Explain why you did or not find any assumption violations.**

**b. check assumptions again after your transformation**

**c. comment on whether the assumption transformations helped or not. If they do not help, then revert back**

1. Linearity: The Average value of the Salary does not have a first order relationship with the independent variables
2. Normality: The values of the Salary are not normally distributed for observations with the same value of the independent variables. But, since the sample size is large (i.e. greater than 30), the data can be considered to be normal.
3. Equal Variation: The variation in the values of the Salary is not equal for the observations with the same value of the independent variables, disregarding the value of the independent variables.

We found that the graphs had assumption violations when observed our original regression after running in SAS. Our results for the same are shown in **Appendix B.**

Dependent variable is violating equal variance assumption. Independent variable is violating linearity assumption. So, we tried using transformations to fix any of the assumptions.

**Transformations**:

Due to the Assumption violations observed above, we did some Transformations on the attributes and observed the plots to determine the changes in the assumptions.

1. Transformed the dependent variable using the Logarithm.

[Appendix C](#Appendix_C) shows the results from the logarithm transformation. The results of the logarithm transformation did not significantly improve the Normality and Equal variance for dependent variable.

1. Transformed the dependent variable using Square root.

We then opted Square root transformation whose results are shown in [**Appendix D**](#Appendix_D). The results of the Square root transformation did not significantly improve the Normality and Equal variance for dependent variable as well.

1. Transformed the Dependent variable using a Reciprocal.

Later we opted Reciprocal transformation whose results are shown in [**Appendix E**](#Appendix_E). The results of the Reciprocal transformation did not significantly improve the Normality and Equal variance for dependent variable here as well.

1. Transformed the independent variables using a Logarithm.

[**Appendix F**](#Appendix_F) shows the results from the logarithm transformation. The results of the logarithm transformation did not significantly improve the Linearity for any of the Independent variables here.

1. Transformed the independent variables using a Square root.

We then opted Square root transformation whose results are shown in [**Appendix G**](#Appendix_G). The results of the Square root transformation did not significantly improve the Linearity for any of the Independent variables here.

1. Transformed the independent variables using a Reciprocal.

Later we opted Reciprocal transformation whose results are shown in [**Appendix H**](#Appendix_H). The results of the Reciprocal transformation did not significantly improve the Linearity for any of the Independent variables here as well.

Since, none of the transformations are satisfying our assumptions we reverted to the original data.

**Q 3. Using all-possible regression, reduce the list of possible models down to three, explaining your reasoning for choosing those three.**

Firstly, we checked the models listed in SAS. Then, we selected the models based on their Mean-Squared-Error, value & number of independent variables.

For the selection of the 3 best models, we compared the MSE which was required to be low and the value to be high, along with less number of independent variables.

The 3 models that we chose out of 127 models are as follows:

|  |  |  |  |
| --- | --- | --- | --- |
| Model Index | Model Number |  | Variables in Model |
| 30 | 3 | 0.2430 | DVDH,HR,RBI |
| 64 | 4 | 0.2508 | DVDH,DVSS,HR,RBI |
| 100 | 5 | 0.2516 | DVDH,DVSS,R,HR,RBI |

For our results refer [**Appendix I**](#Appendix_I)

**Q4. Examine the three models in detail and choose one model. Explain why you chose least squares line for that model.**

**Model 30 : DVDH,HR,RBI**

MSE: 37.49787

F-value = 24.29

P-value: <0.0001

Rsq = 0.2430

Adj Rsq = 0.2330

= 1.09584+9.77332(DVDH)+0.15518(HR)+0.05939(RBI)

**Model 64 : DVSS,DVDH,HR,RBI**

MSE: 37.27513

F-value=18.91

P-value:<0.0001

Rsq=0.2508

Adj Rsq=0.2375

= 1.36614-1.86284(DVSS)+9.64975(DVDH)+0.13339(HR)+0.06453(RBI)

**Model 100 : DVSS,DVDH,R,HR,RBI**

MSE : 37.40239

F-value = 15.13

P-value:<0.0001

Rsq=0.2516

Adj Rsq=0.2349

= 1.47955-1.84036(DVSS)+9.4243(DVDH)-0.0149(R)+0.13101(HR)+0.07786(RBI)

|  |
| --- |
| **The Models are highlighted as shown in** [**Appendix J**](#Appendix_J)  **EXPLANATION:**  The Model 64(MSE=37. 27513) has the least MSE value as compared to the Model 30(MSE=37. 49787) and Model 100(MSE=37.40239).  Although the MSE is least for Model 64, it does not differ much as compared to MSE of Model 30. So, we concluded that it would be better to choose Model 30 as the complexity ( Number of variables) is the least and MSE does not largely vary when compared to the other two models.  The Least Squares line for the selected model (Model 30) is:  **= 1.09584+9.77332(DVDH)+0.15518(HR)+0.05939(RBI)**  **Therefore, from the value, Number of Independent Variables and MSE value, we can infer that Model 30 is the best model amongst the three models.** |

**Q 5.** **Using the least squares equation found in 4 from your training set, predict the values of the dependent variable in your validation data set. Average the absolute differences between the actual value and the predicted value. Discuss the average prediction you would expect in the future using this model.**

* The absolute differences between the actual salary and predicted salary were then calculated in our table and we determined the average prediction.
* The average absolute deviation for our data = 0.266293
* The average predicted salary= 3.802

|  |  |
| --- | --- |
| SE | 0.135151 |
| T table value | 1.970332 |
| MOE | 0.266293 |

The average future predicted salary will be 3.802 and a difference of 0.266293 .

|  |  |  |
| --- | --- | --- |
| Salary | Predicted Salary | Absolute Difference |
| 10.000 | 6.233 | 3.767 |
| 3.200 | 2.370 | 0.830 |
| 0.512 | 1.489 | 0.977 |
| 17.250 | 2.881 | 14.369 |
| 1.500 | 1.987 | 0.487 |
| 11.333 | 11.747 | 0.414 |
| 0.518 | 1.452 | 0.934 |
| 8.500 | 20.577 | 12.077 |
| 0.566 | 8.461 | 7.895 |
| 13.750 | 5.451 | 8.299 |
| 3.750 | 3.415 | 0.335 |
| 1.450 | 3.666 | 2.216 |
| 8.583 | 10.365 | 1.782 |
| 1.500 | 5.501 | 4.001 |
| 12.083 | 7.114 | 4.969 |
| 17.000 | 6.415 | 10.585 |
| 5.100 | 7.141 | 2.041 |
| 0.513 | 4.366 | 3.853 |
| 10.000 | 5.699 | 4.301 |
| 3.500 | 4.735 | 1.235 |
| 0.538 | 7.530 | 6.992 |
| 11.000 | 4.709 | 6.291 |
| 15.000 | 7.137 | 7.863 |
| 6.950 | 9.411 | 2.461 |
| 0.528 | 4.772 | 4.244 |
| 14.325 | 14.760 | 0.435 |
| 24.083 | 12.800 | 11.283 |
| 3.000 | 5.402 | 2.402 |
| 2.100 | 2.772 | 0.672 |
| 7.081 | 4.795 | 2.286 |
| 3.125 | 5.201 | 2.076 |
| 3.900 | 5.649 | 1.749 |
| 1.000 | 2.822 | 1.822 |
| 14.250 | 13.444 | 0.806 |
| 0.509 | 1.155 | 0.646 |
| 8.750 | 5.580 | 3.170 |
| 0.513 | 2.594 | 2.081 |
| 1.050 | 1.644 | 0.594 |
| 0.511 | 4.415 | 3.904 |
| 0.513 | 4.498 | 3.985 |
| 23.778 | 20.257 | 3.521 |
| 8.750 | 7.854 | 0.896 |
| 6.250 | 8.411 | 2.161 |
| 0.530 | 8.365 | 7.835 |
| 0.545 | 4.534 | 3.989 |
| 0.540 | 5.986 | 5.446 |
| 3.125 | 7.342 | 4.217 |
| 22.500 | 10.345 | 12.155 |
| 5.125 | 8.711 | 3.586 |
| 18.556 | 9.685 | 8.871 |
| 0.650 | 6.992 | 6.342 |
| 6.400 | 12.747 | 6.347 |
| 0.521 | 6.629 | 6.108 |
| 10.000 | 4.448 | 5.552 |
| 0.510 | 2.726 | 2.216 |
| 29.200 | 8.401 | 20.799 |
| 0.510 | 1.274 | 0.764 |
| 0.508 | 1.370 | 0.862 |

**Q 6.** **Using the validation data set, see if the model found in 4 is useful.**

**Hypothesis:** Ho: (No Baseball attributes has an effect on the average salary)

H1: At least one has an effect.

**Rejection Region:**

k=3

n=58

n-k-1= 54

F table value: F(3,54)= 0.798717

Reject Ho if F-stat> F(3,54) =0.798717

**Decision:** F-statistic=MSR/MSE= 9.63 >0.798717 (Refer [Appendix K)](#_Appendix_K)

Since the sample means vary too much, you can conclude that not all population means were the same

**Conclusion:** We can say that the changes in at least one of the Baseball attributes (DVDH, HR, RBI) is useful in predicting salary.

**Appendix A**

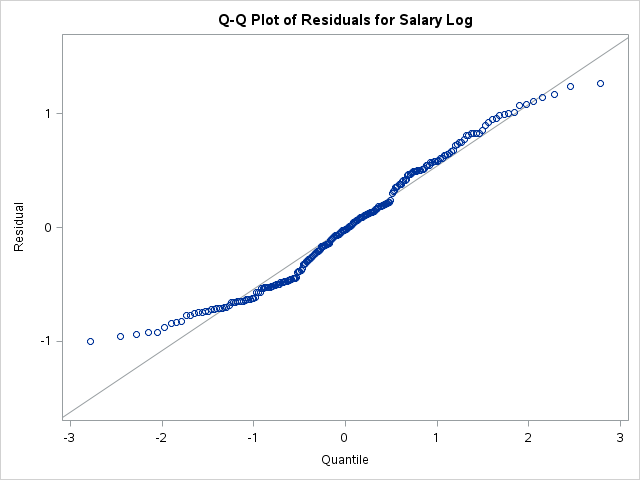
1.**APENDIX A**: ORIGINAL TRAINING SET

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| |  |  | | --- | --- | | |  | | --- | | **Linear Regression Results** | | | |  |  | | --- | --- | | |  | | --- | | **The REG Procedure Model: Linear\_Regression\_Model Dependent Variable: Salary** | | | |  |  |  |  |  | | --- | --- | --- | --- | --- | | | Number of Observations Read | 231 | | --- | --- | | Number of Observations Used | 231 | | | | **Analysis of Variance** | | | | | | | --- | --- | --- | --- | --- | --- | | **Source** | **DF** | **Sum of Squares** | **Mean Square** | **F Value** | **Pr > F** | | Model | 7 | 2863.33066 | 409.04724 | 10.88 | <.0001 | | Error | 223 | 8380.86432 | 37.58235 |  |  | | Corrected Total | 230 | 11244 |  |  |  | | | | Root MSE | 6.13044 | R-Square | 0.2546 | | --- | --- | --- | --- | | Dependent Mean | 5.64273 | Adj R-Sq | 0.2313 | | Coeff Var | 108.64321 |  |  | | | | **Parameter Estimates** | | | | | | | | --- | --- | --- | --- | --- | --- | --- | | **Variable** | **DF** | **Parameter Estimate** | **Standard Error** | **t Value** | **Pr > |t|** | **Variance Inflation** | | Intercept | 1 | 0.26351 | 3.53429 | 0.07 | 0.9406 | 0 | | DV1b | 1 | 1.28802 | 1.41018 | 0.91 | 0.3620 | 1.17967 | | DVSS | 1 | -1.74580 | 1.22539 | -1.42 | 0.1556 | 1.04297 | | DVDH | 1 | 9.92101 | 3.69161 | 2.69 | 0.0077 | 1.07372 | | R | 1 | -0.01082 | 0.03191 | -0.34 | 0.7348 | 4.28848 | | HR | 1 | 0.12611 | 0.09694 | 1.30 | 0.1946 | 5.52310 | | RBI | 1 | 0.06880 | 0.04527 | 1.52 | 0.1300 | 9.40599 | | AVG | 1 | 5.32768 | 15.41792 | 0.35 | 0.7300 | 1.32430 | | | | |  | | --- | | Generated by the SAS System ('SASApp', Linux) on 27 November 2016 at 10:15:51 PM | | | |
| |  |  | | --- | --- | | |  | | --- | | **Linear Regression Results** | | | |  |  | | --- | --- | | |  | | --- | | **The REG Procedure Model: Linear\_Regression\_Model Dependent Variable: Salary** | | | |  | | --- | | img0.png | | img1.png | | img2.png | | | |  | | --- | | Generated by the SAS System ('SASApp', Linux) on 27 November 2016 at 10:15:51 PM | | | |

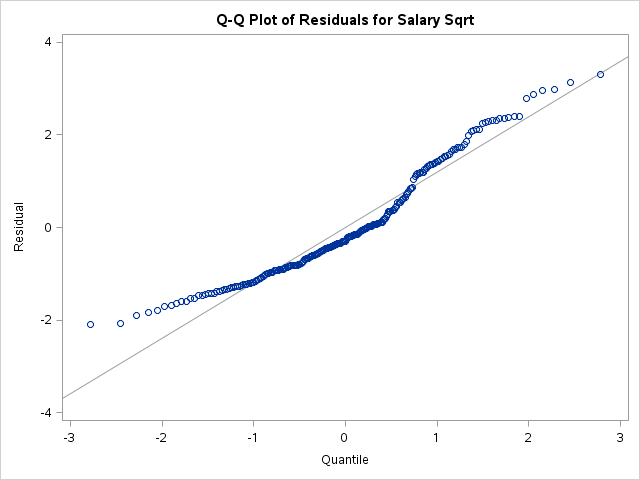
**[Appendix](#Appendix_B) B**

|  |  |
| --- | --- |
| |  | | --- | | **Linear Regression Results** | |
| |  |  | | --- | --- | | |  | | --- | | **The REG Procedure Model: Linear\_Regression\_Model Dependent Variable: Salary** | | | |  | | --- | | img0.png | | Original Regression where we noted that there was an assumption violation of Normality. | | img1.png | | img2.png | | |  | |

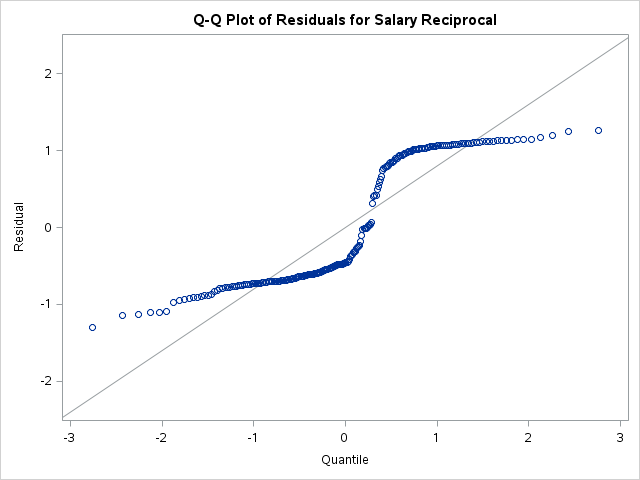
**Appendix C**



**Appendix D**

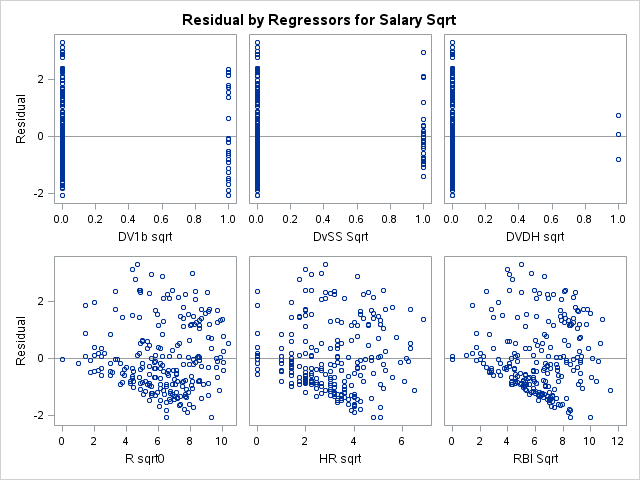


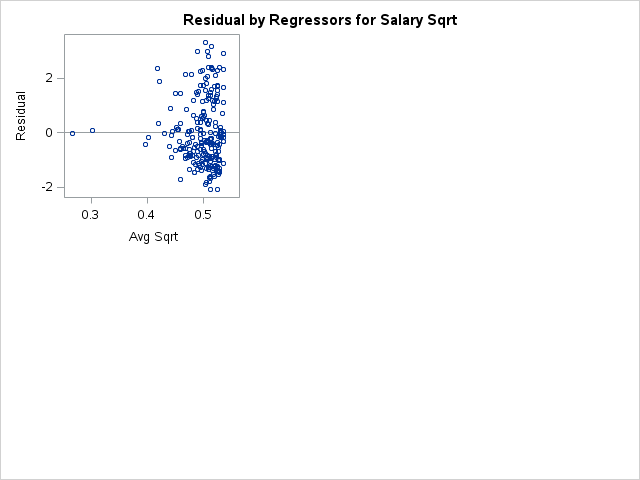
**Appendix E**



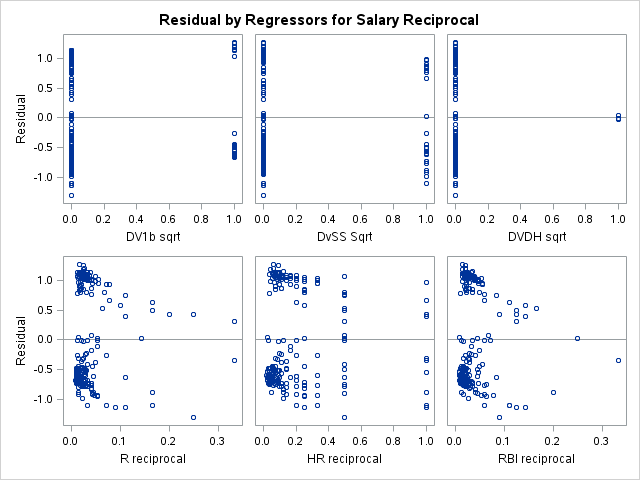
**Appendix F**

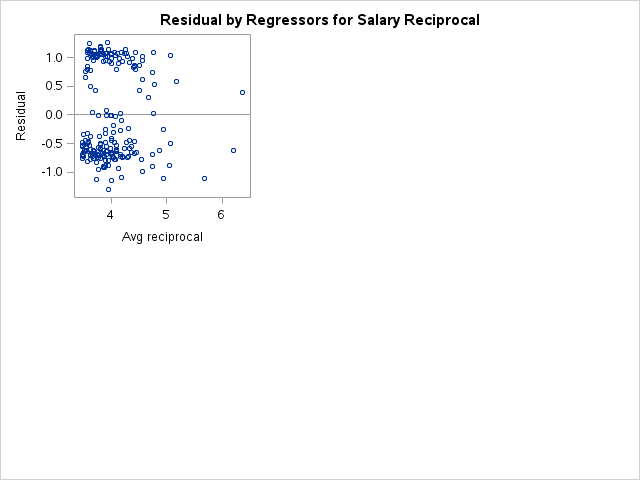
|  |
| --- |
| img1.png |
| img2.png |

**Appendix G**



**Appendix H**





**Appendix I**

|  |  |
| --- | --- |
| |  | | --- | | **The REG Procedure Model: Linear\_Regression\_Model Dependent Variable: Salary   R-Square Selection Method** | |
| |  |  |  |  |  | | --- | --- | --- | --- | --- | | | Number of Observations Read | 231 | | --- | --- | | Number of Observations Used | 231 | | |
| |  |  | | --- | --- | | |  | | --- | |  | | |
| |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  | | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | | |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  | | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | | | **Number in Model** | **R-Square** | **Variables in Model** | | --- | --- | --- | | 1 | 0.2080 | RBI | | 1 | 0.2073 | HR | | 1 | 0.1339 | R | | 1 | 0.0570 | DVDH | | 1 | 0.0388 | AVG | | 1 | 0.0365 | DV1b | | 1 | 0.0211 | DVSS | | 2 | 0.2340 | DVDH RBI | | 2 | 0.2326 | DVDH HR | | 2 | 0.2193 | DVSS RBI | | 2 | 0.2187 | HR RBI | | 2 | 0.2138 | DVSS HR | | 2 | 0.2117 | DV1b RBI | | 2 | 0.2113 | R RBI | | 2 | 0.2099 | DV1b HR | | 2 | 0.2087 | HR AVG | | 2 | 0.2082 | RBI AVG | | 2 | 0.2081 | R HR | | 2 | 0.1771 | DVDH R | | 2 | 0.1515 | DV1b R | | 2 | 0.1501 | DVSS R | | 2 | 0.1350 | R AVG | | 2 | 0.0974 | DV1b DVDH | | 2 | 0.0935 | DVDH AVG | | 2 | 0.0752 | DVSS DVDH | | 2 | 0.0712 | DV1b AVG | | 2 | 0.0605 | DVSS AVG | | 2 | 0.0511 | DV1b DVSS | | 3 | 0.2443 | DVSS DVDH RBI | | 3 | 0.2430 | DVDH HR RBI | | 3 | 0.2398 | DV1b DVDH RBI | | 3 | 0.2387 | DVSS DVDH HR | | 3 | 0.2372 | DV1b DVDH HR | | 3 | 0.2353 | DVDH R RBI | | 3 | 0.2346 | DVDH HR AVG | | 3 | 0.2343 | DVDH R HR | | 3 | 0.2340 | DVDH RBI AVG | | 3 | 0.2271 | DVSS HR RBI | | 3 | 0.2221 | DVSS R RBI | | 3 | 0.2217 | DV1b DVSS RBI | | 3 | 0.2213 | R HR RBI | | 3 | 0.2210 | DV1b HR RBI | | 3 | 0.2194 | DVSS RBI AVG | | 3 | 0.2187 | HR RBI AVG | | 3 | 0.2157 | DVSS HR AVG | | 3 | 0.2157 | DV1b DVSS HR | | 3 | 0.2151 | DVSS R HR | | 3 | 0.2139 | DV1b R RBI | | 3 | 0.2118 | DV1b RBI AVG | | 3 | 0.2115 | DV1b HR AVG | | 3 | 0.2114 | R RBI AVG | | 3 | 0.2110 | DV1b R HR | | 3 | 0.2090 | R HR AVG | | 3 | 0.1978 | DV1b DVDH R | | 3 | 0.1912 | DVSS DVDH R | | 3 | 0.1783 | DVDH R AVG | | 3 | 0.1638 | DV1b DVSS R | | 3 | 0.1527 | DV1b R AVG | | 3 | 0.1515 | DVSS R AVG | | 3 | 0.1296 | DV1b DVDH AVG | | 3 | 0.1123 | DVSS DVDH AVG | | 3 | 0.1091 | DV1b DVSS DVDH | | 3 | 0.0866 | DV1b DVSS AVG | | 4 | 0.2508 | DVSS DVDH HR RBI | | 4 | 0.2485 | DV1b DVSS DVDH RBI | | 4 | 0.2472 | DV1b DVDH HR RBI | | 4 | 0.2453 | DVSS DVDH R RBI | | 4 | 0.2443 | DVSS DVDH RBI AVG | | 4 | 0.2440 | DVDH R HR RBI | | 4 | 0.2431 | DVDH HR RBI AVG | | 4 | 0.2423 | DV1b DVSS DVDH HR | | 4 | 0.2412 | DVSS DVDH HR AVG | | 4 | 0.2410 | DVSS DVDH R HR | | 4 | 0.2403 | DV1b DVDH R RBI | | 4 | 0.2398 | DV1b DVDH RBI AVG | | 4 | 0.2397 | DV1b DVDH R HR | | 4 | 0.2396 | DV1b DVDH HR AVG | | 4 | 0.2354 | DVDH R HR AVG | | 4 | 0.2353 | DVDH R RBI AVG | | 4 | 0.2293 | DVSS R HR RBI | | 4 | 0.2287 | DV1b DVSS HR RBI | | 4 | 0.2271 | DVSS HR RBI AVG | | 4 | 0.2237 | DV1b DVSS R RBI | | 4 | 0.2229 | DV1b R HR RBI | | 4 | 0.2221 | DVSS R RBI AVG | | 4 | 0.2217 | DV1b DVSS RBI AVG | | 4 | 0.2214 | R HR RBI AVG | | 4 | 0.2210 | DV1b HR RBI AVG | | 4 | 0.2178 | DV1b DVSS HR AVG | | 4 | 0.2174 | DV1b DVSS R HR | | 4 | 0.2163 | DVSS R HR AVG | | 4 | 0.2139 | DV1b R RBI AVG | | 4 | 0.2120 | DV1b R HR AVG | | 4 | 0.2080 | DV1b DVSS DVDH R | | 4 | 0.1993 | DV1b DVDH R AVG | | 4 | 0.1928 | DVSS DVDH R AVG | | 4 | 0.1653 | DV1b DVSS R AVG | | 4 | 0.1421 | DV1b DVSS DVDH AVG | | 5 | 0.2539 | DV1b DVSS DVDH HR RBI | | 5 | 0.2516 | DVSS DVDH R HR RBI | | 5 | 0.2510 | DVSS DVDH HR RBI AVG | | 5 | 0.2489 | DV1b DVSS DVDH R RBI | | 5 | 0.2486 | DV1b DVSS DVDH RBI AVG | | 5 | 0.2475 | DV1b DVDH R HR RBI | | 5 | 0.2474 | DV1b DVDH HR RBI AVG | | 5 | 0.2455 | DV1b DVSS DVDH R HR | | 5 | 0.2453 | DVSS DVDH R RBI AVG | | 5 | 0.2452 | DV1b DVSS DVDH HR AVG | | 5 | 0.2442 | DVDH R HR RBI AVG | | 5 | 0.2424 | DVSS DVDH R HR AVG | | 5 | 0.2410 | DV1b DVDH R HR AVG | | 5 | 0.2403 | DV1b DVDH R RBI AVG | | 5 | 0.2303 | DV1b DVSS R HR RBI | | 5 | 0.2295 | DVSS R HR RBI AVG | | 5 | 0.2287 | DV1b DVSS HR RBI AVG | | 5 | 0.2237 | DV1b DVSS R RBI AVG | | 5 | 0.2230 | DV1b R HR RBI AVG | | 5 | 0.2186 | DV1b DVSS R HR AVG | | 5 | 0.2098 | DV1b DVSS DVDH R AVG | | 6 | 0.2543 | DV1b DVSS DVDH HR RBI AVG | | 6 | 0.2543 | DV1b DVSS DVDH R HR RBI | | 6 | 0.2519 | DVSS DVDH R HR RBI AVG | | 6 | 0.2490 | DV1b DVSS DVDH R RBI AVG | | 6 | 0.2479 | DV1b DVDH R HR RBI AVG | | 6 | 0.2469 | DV1b DVSS DVDH R HR AVG | | 6 | 0.2305 | DV1b DVSS R HR RBI AVG | | 7 | 0.2546 | DV1b DVSS DVDH R HR RBI AVG | | | |

**Appendix J**

**Model30:**

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| |  |  | | --- | --- | | |  | | --- | | **Linear Regression Results** | | | |  |  | | --- | --- | | |  | | --- | | **The REG Procedure Model: Linear\_Regression\_Model Dependent Variable: Salary** | | | |  |  |  |  |  | | --- | --- | --- | --- | --- | | | Number of Observations Read | 231 | | --- | --- | | Number of Observations Used | 231 | | | | **Analysis of Variance** | | | | | | | --- | --- | --- | --- | --- | --- | | **Source** | **DF** | **Sum of Squares** | **Mean Square** | **F Value** | **Pr > F** | | Model | 3 | 2732.17833 | 910.72611 | 24.29 | <.0001 | | Error | 227 | 8512.01665 | 37.49787 |  |  | | Corrected Total | 230 | 11244 |  |  |  | | | | Root MSE | 6.12355 | R-Square | 0.2430 | | --- | --- | --- | --- | | Dependent Mean | 5.64273 | Adj R-Sq | 0.2330 | | Coeff Var | 108.52103 |  |  | | | | **Parameter Estimates** | | | | | | | | --- | --- | --- | --- | --- | --- | --- | | **Variable** | **DF** | **Parameter Estimate** | **Standard Error** | **t Value** | **Pr > |t|** | **Variance Inflation** | | Intercept | 1 | 1.09584 | 0.80407 | 1.36 | 0.1743 | 0 | | DVDH | 1 | 9.77332 | 3.61987 | 2.70 | 0.0075 | 1.03472 | | HR | 1 | 0.15518 | 0.09433 | 1.64 | 0.1014 | 5.24191 | | RBI | 1 | 0.05939 | 0.03373 | 1.76 | 0.0796 | 5.23330 | | | | |  | | --- | |  | | | |
| |  |  | | --- | --- | | |  | | --- | | **Linear Regression Results** | | | |  |  | | --- | --- | | |  | | --- | | **The REG Procedure Model: Linear\_Regression\_Model Dependent Variable: Salary** | | | |  | | --- | | img0.png | | img1.png | | | |  | | --- | |  | | | |

**Model64:**

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| |  |  | | --- | --- | | |  | | --- | | **Linear Regression Results** | | | |  |  | | --- | --- | | |  | | --- | | **The REG Procedure Model: Linear\_Regression\_Model Dependent Variable: Salary** | | | |  |  |  |  |  | | --- | --- | --- | --- | --- | | | Number of Observations Read | 231 | | --- | --- | | Number of Observations Used | 231 | | | | **Analysis of Variance** | | | | | | | --- | --- | --- | --- | --- | --- | | **Source** | **DF** | **Sum of Squares** | **Mean Square** | **F Value** | **Pr > F** | | Model | 4 | 2820.01465 | 705.00366 | 18.91 | <.0001 | | Error | 226 | 8424.18033 | 37.27513 |  |  | | Corrected Total | 230 | 11244 |  |  |  | | | | Root MSE | 6.10534 | R-Square | 0.2508 | | --- | --- | --- | --- | | Dependent Mean | 5.64273 | Adj R-Sq | 0.2375 | | Coeff Var | 108.19824 |  |  | | | | **Parameter Estimates** | | | | | | | | --- | --- | --- | --- | --- | --- | --- | | **Variable** | **DF** | **Parameter Estimate** | **Standard Error** | **t Value** | **Pr > |t|** | **Variance Inflation** | | Intercept | 1 | 1.36614 | 0.82079 | 1.66 | 0.0974 | 0 | | DVSS | 1 | -1.86284 | 1.21352 | -1.54 | 0.1262 | 1.03129 | | DVDH | 1 | 9.64975 | 3.61000 | 2.67 | 0.0081 | 1.03523 | | HR | 1 | 0.13339 | 0.09512 | 1.40 | 0.1622 | 5.36132 | | RBI | 1 | 0.06453 | 0.03380 | 1.91 | 0.0575 | 5.28507 | | | | |  | | --- | |  | | | |
| |  |  | | --- | --- | | |  | | --- | | **Linear Regression Results** | | | |  |  | | --- | --- | | |  | | --- | | **The REG Procedure Model: Linear\_Regression\_Model Dependent Variable: Salary** | | | |  | | --- | | img0.png | | img1.png | | |  | | |

**Model100:**

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  | | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | | |  |  | | --- | --- | | |  | | --- | | **Linear Regression Results** | | | |  |  | | --- | --- | | |  | | --- | | **The REG Procedure Model: Linear\_Regression\_Model Dependent Variable: Salary** | | | |  |  |  |  |  | | --- | --- | --- | --- | --- | | | Number of Observations Read | 231 | | --- | --- | | Number of Observations Used | 231 | | | | **Analysis of Variance** | | | | | | | --- | --- | --- | --- | --- | --- | | **Source** | **DF** | **Sum of Squares** | **Mean Square** | **F Value** | **Pr > F** | | Model | 5 | 2828.65682 | 565.73136 | 15.13 | <.0001 | | Error | 225 | 8415.53816 | 37.40239 |  |  | | Corrected Total | 230 | 11244 |  |  |  | | | | Root MSE | 6.11575 | R-Square | 0.2516 | | --- | --- | --- | --- | | Dependent Mean | 5.64273 | Adj R-Sq | 0.2349 | | Coeff Var | 108.38278 |  |  | | | | **Parameter Estimates** | | | | | | | | --- | --- | --- | --- | --- | --- | --- | | **Variable** | **DF** | **Parameter Estimate** | **Standard Error** | **t Value** | **Pr > |t|** | **Variance Inflation** | | Intercept | 1 | 1.47955 | 0.85537 | 1.73 | 0.0851 | 0 | | DVSS | 1 | -1.84036 | 1.21649 | -1.51 | 0.1317 | 1.03282 | | DVDH | 1 | 9.42438 | 3.64642 | 2.58 | 0.0104 | 1.05264 | | R | 1 | -0.01492 | 0.03105 | -0.48 | 0.6312 | 4.07890 | | HR | 1 | 0.13101 | 0.09541 | 1.37 | 0.1711 | 5.37579 | | RBI | 1 | 0.07786 | 0.04376 | 1.78 | 0.0766 | 8.83052 | | | | |  | | --- | | Generated by the SAS System ('SASApp', Linux) on 27 November 2016 at 10:27:28 PM | | | | | |  |  | | --- | --- | | |  | | --- | | **Linear Regression Results** | | | |  |  | | --- | --- | | |  | | --- | | **The REG Procedure Model: Linear\_Regression\_Model Dependent Variable: Salary** | | | |  | | --- | | img0.png | | img1.png | | |  | | | |

# Appendix K

Linear Regression Results

**The REG Procedure  
Model: Linear\_Regression\_Model  
Dependent Variable: Salary**

| **Analysis of Variance** | | | | | |
| --- | --- | --- | --- | --- | --- |
| **Source** | **DF** | **Sum of Squares** | **Mean Square** | **F Value** | **Pr > F** |
| Model | 3 | 1043.87844 | 347.95948 | 9.63 | <.0001 |
| Error | 54 | 1950.65656 | 36.12327 |  |  |
| Corrected Total | 57 | 2994.53501 |  |  |  |
|  |  |  |  |  |  |

| Root MSE | | 6.01026 | R-Square | 0.3486 |
| --- | --- | --- | --- | --- |
| Dependent Mean | | 6.67814 | Adj R-Sq | 0.3124 |
| Coeff Var | | 89.99909 |  |  |
| **Parameter Estimates** | | | | | | | |
| **Variable** | **DF** | **Parameter Estimate** | **Standard Error** | **t Value** | | **Pr > |t|** | **Variance Inflation** |
| Intercept | 1 | 1.24391 | 1.61956 | 0.77 | | 0.4458 | 0 |
| DVDH | 1 | 5.18122 | 4.55031 | 1.14 | | 0.2599 | 1.10684 |
| HR | 1 | 0.30607 | 0.14054 | 2.18 | | 0.0338 | 3.41802 |
| RBI | 1 | 0.02772 | 0.04759 | 0.58 | | 0.5627 | 3.60653 |