**DSC 423 – FINAL PROJECT REPORT**

**ANALYSIS ON THE SUCCESS RATE OF LIVER TRANSPLANT (LTx)**

**NAME:** GOUTHAM SELVAKUMAR

**TABLE OF CONTENTS**

1. **Introduction** ………………………………………………. 3
2. **Analysis & Result** ………………………………………………. 4
3. **Findings** ………………………………………………. 6
4. **Future Work** ………………………………………………. 10
5. **Research References** ………………………………………………. 11
6. **Appendix**  ………………………………………………. 12

1. **INTRODUCTION:**

The objective of this task is to observe the number of patients that are in the specific disposition. Each tolerant no matter what their age is on the edge to get a liver transfer these days however they are not getting one because of the blood bunch that assumes a significant part, in this way, that makes them stand by more extended in the line which prompts passing or pull out. Moreover, with the datasets, we can find the number of patients that received and not received the liver transfer.

The most ideal way to determine those things would be with the assistance of more current innovations that are presented in these ages. Rather than holding up along the line and prompting passing or pulled out, with the assistance of recently presented strategies the holding up rundown can be abbreviated that will prompt the likelihood of getting a fruitful liver transfer. The Perfusion Machine keeps up with the liver in a decent state with the expected measure of blood, supplements, medication and keeping it warm for the patients in the holding up list

Interest in liver transplantation keeps on surpassing contributor organ supply. Contrasting beneficiary endurance with that of equivalent applicants without a transfer can work on comprehension of relocate endurance benefit [1]. A deficiency of wellbeing should be visible as an emergency circumstance, with impacts reaching out to the family members of patients and influencing their personal satisfaction. [1] Liver transplantation (LTx) furnishes the main therapeutic treatment choice with a phenomenal long haul that brings about patients with decompensated cirrhosis of the liver [2].

Endurance benefit expanded with expanding MELD score. Given the serious deficiency of contributor organs, there is a need to recognize which patients get huge endurance benefits from transplantation and which don’t [2]. Enrolled relocate competitors who have effectively entered the transfer pathway establish the best examination gathering to relocate beneficiaries [2]. The capacity of MELD to foresee mortality in patients with ongoing liver infection has added to the advancement of liver portion strategy in the United States. The MELD scoring framework has been executed in the United States for prioritization in 2022, and inside the partnership of Euro relocate November 2006[3]. Patient’s clinical information was gathered from their records and incorporated the length of treatment, Model for End-Stage Liver Disease (MELD) score, and sickness etiology. Liver transfer endurance benefit isn’t equally conveyed across the scope of MELD scores [3]. This is especially helpful in contemplations of organ portion, where relative endurance benefits among a gathering of applicants are still up in the air [2].

Transfers are being performed for certain applicants who have a higher gamble of passing on from the transfer methodology than they have of kicking the bucket from their basic liver sickness. Rather than depending exclusively on the gamble of pre-relocate demise, the endurance benefit among competitors ought to be considered as a part of distribution strategy to guide organs to those probably going to profit from the system [2]. It is important to assess the elements engaged with casual providing care for liver transfer applicants. Among the fundamental psychosocial factors in the signs for a medical procedure, enough culmination of the elements of the essential guardian should be viewed as vital. Purposeless transfers among seriously sick patients are not distinguished under current practice. With 1 year post-relocate follow-up, patients at the lower hazard of pre-relocate passing don’t have a self-evident endurance benefit from liver transfer [1] [2].

1. **ANALYSIS & RESULT:**

I have taken my dataset as Liver\_transplant. The dependent variable is Event trailed by independent variables such as ID, Sex, Age, Abo, Year, and Futime. For observing the investigation of the patients with the achievement pace of liver transplantation, I made 3 dummy variables for Abo (A, B, AB) i.e., Abo1, Abo2, Abo3 with keep going one (O) as the base since it influences the Y variable in it. The variable Sex is likewise transformed into a dummy variable numsex. The dependent variable Event has Ltx, withdraw, censored, and death values in it from which I took Ltx as the fundamental value and made a dummy variable for it as numevent. As Ltx alludes to successful Liver transplantation [1]. The objective is to observe the achievement pace of patients on the waiting list with the blood bunch A, B, and AB.

Since my Dependent variable depends on non-numeric qualities my dataset has been set to Logistic Regression to see as my last model. I involved the STB to track down the standard deviation for independent variables and fitted the last model with the selection method. The selection method that I utilized was Stepwise and Backward rsquare. In the wake of, utilizing it I was left with the last model which is required for additional investigation of Multicollinearity.

In the wake of fitting it with the last model variables, I utilized descriptive like Boxplot among the dependent and independent variables to observe the connection between them which further aids in understanding the information totally [3]. The Scatterplots are utilized alongside it which shows the straight connections among them. Along these lines, that could provide us with a knowledge of how the information relates to dependent and independent.

For checking in the event that there is collinearity present among the independent variables we need to utilize Variance Inflation Factor (VIF) and Tolerance (tol). At the point when these two are utilized, we can dispose of the independent variables which are non-significant and can furnish it with a superior model. Whenever the better model is given subsequent stage is actually looking at them with the Residual plots and Influential points [2].

The following is to find the relationship between the independent variables and the dependent variables from the given table. From that table, we can find assuming there is any connection among the factors or they are unequivocally connected with the dependent variable. The relationship takes esteem from the final model that we got from the Logistic Regression Technique.

The Influential Points and Residuals are don’t observe the outliers that are available in the studentized residual. Subsequent to checking with the CDF of studentized residual, in the event that the pictorial portrayal isn’t sufficiently legitimate, then, at that point, we need to find for points that are incurring the standard deviation in the diagram for the dataset Liver\_transplant [2][3]. There can be at least one influential point present in them which we need to kill and find the result once more. Thus, after the end of that multitude of points, we need to review our studentized residuals once again.

After fulfillment of the last model with the Logistic Selection method, observing the Multicollinearity, Influential Points, and Residuals, we need to do the Validation technique. The Dataset Liver\_transplant must be tried with Training and Test Dataset respectively. The Training and Test set procedure follows the stepwise method in the Logistic. After the approval is finished with the expected value that we got from the classification table, the final model has been identified, as the non-significant variables being wiped out from the Dataset.

To finish our Dataset totally, we would have to review the information for Collinearity present among the variables i.e., VIF and tol. Once more whenever it is done, the Influential points and residuals are done to track down the outliers and kill them consistently in a grouping. On the off chance that there aren’t any outliers or VIF not being higher in value, the datasets are set to be finished. Along these lines, this will carry us to the final model dataset.

1. **FINDINGS:**

log(numevent=1/numevent=0)=-263.8-262.9-257.8+0.3497\*numsex+ 0.0238\*age + 0.1299\*year + 0.0031\*futime

This is the Final Model Equation for Logistic Regression using the Selection method of Stepwise and Backward rsquare.

**NUMSEX** – If the changes occur in the numsex the odds for numevent will increase by 41.9% and with a 95% Confidence Interval of between 0.2 % and 100.9%. So, there will be an increase in odds against the outcome of the numevent.

**AGE** - If there is a change in the addition of age, the outcome odds of numevent are incremented by 2.4 % and with 95% Confidence Intervals of 0.7% and 4.2%. Moreover, the outcome of the numevent will have increase in odds.

**YEAR** – If the year is affected, the odds of the numevent are 13.9% and a 95% Confidence Interval of 5.5% and 22.9%. So, the numevent will face an increase in the outcome.

**FUTIME** – If there is a change in the increase of futime the odds of numevent will rose up by 0.3% and 95% Confidence Interval between 0.2% and 0. So, the odds of numevent will get affected by the increase in outcome. **(*Appendix A)***

There is no linear association present between numevent and the rest of the independent variables (numsex, age, year, and futime). The relationship between age and futime most likely shows the people around the age 40 to 60 have higher time of entry to final. The people in the year of 1996 are shown to be larger in the entry of final disposition of futime. ***(Appendix B)***

The Boxplots show the Pictorial Representation of the dependent variable (numevent) vs the independent variables (year) and (age). So, from the Boxplots we can analyze that the median age for numevent 0 is 1996 which is from 1995 to 1997 i.e., the first quartile to the third quartile respectively. Moreover, this numevent 0 tells us that only a minimal amount of people underwent Ltx and got a successful Liver transplant. The numevent 1 depicts that from the early year of 1992 people were added to the waiting list for liver transplant, the median is of 1994 which is from the first quartile to third quartile i.e., 1993 to 1996 consecutively. So, the numevent of this year 1 shows that there were huge factors of people added to the waiting list and got Ltx. The other Boxplot shows the pictorial representation between numevent and the age. The numevent 0 shows the median of 50 for the first quartile and third quartile from 45 to 58 of age. Moreover, the q1, q2, and the q3 follow the same addition into the numevent. So, it won’t affect the final disposition that much and will not result in a drastic change [3][2]. The change in age addition doesn’t add that much to the final disposition. **(*Appendix C)***

The correlation table values show that there is multicollinearity present as the variable’s year and futime are moderately least correlated. The numevent is least linear correlated with numsex and age. There is a moderate linear relationship between the numevent and the two predictors such as year and futime. The coefficient value for numsex is low as -0.04 followed by age as -0.05[1]. The year and futime are moderately least correlated with -0.28 and -0.37 respectively.

***(Appendix D)***

The parameter estimates are given for numsex are -0.02241, -0.00190 for the age followed by a year at -0.02205 and futime at -0.00051250. The table gives the VIF to prove that there is no multicollinearity present and no need for a new model. So, the values are normal, and this is said to be the final regression model. ***(Appendix E)***

The Studentized Residual Plot after finding the final regression model shows us that the normality of data points in it is not similar to it as the points are randomly scattered and irrelevant. The graph also depicts that there are outliers present in it and have constant variance and independence [3]. The residual plot of age has outliers present on the left side of the graph, these residuals also have constant variance and independence present in them. ***(Appendix F)***

The Influential Plots have some insignificant values that are to be deleted one by one. The data point 259 is deleted first in order to run the studentized residual, followed by data point 674 which gives us the correct influential and residual plot for the final regression model. Then, the studentized residual is done for the final model to find the residual plots. The normality graph is perfect and there is no reason as this normality plot is not satisfied. ***(Appendix G)***

The Validation Technique circles back to the training and test dataset. The data is parted into the training and test sets with 75% of observations to be haphazardly chosen for the training set. The Frequency is done to check in the event that the split has been done accurately for training and the test set. Another variable train\_y is made for the training set and NA for the testing set. Presently, the selection method is done on the training set to see as the final model and compute the predicted value on a training set and acquire the cut-off value for p [3]. The cut-off value of 0.6 has been found from the classification table by adding the percentage section of sensitivity and specificity along with the biggest number that connects with prob level. At long last, the classification matrix is done between the numevent and pred\_y. The 15 observations for the test set have been evaluated. (***Appendix H)***

True Positive = 160 False Positive = 18

True Negative = 19 False Negative = 6

Sensitivity (or) Recall of the dataset is 1.0375, the accuracy of the validation model is 0.8817, precision is 1.1125, with Specificity followed by 1.9457 and the F-metric is to be 1.0734.

After the validation technique has been done for training and test dataset. We need to find if there is any multicollinearity present among the final model (year, futime). When the VIF and tol have been computed for the final model it shows that there is no collinearity among the variables as the final model value depicts it as normal. The parameter estimate for age is -0.02468 followed by futime of -0.00050248. ***(Appendix I)***

The Studentized Residual plot has been evaluated for the final model with collinearity as VIF and tol check done. The outliers present in the influential points are eliminated in order for the final model to produce the residual plots without any errors in the data points. The number 259 is removed as it corresponds to the outliers for the final model. ***(Appendix J)***

The final model is now finally checked for residual plots second time after finding the training and test with VIF and tol done. The residual plots are good with the normality of the graph being normal and not depicting any kind of random data points being irregular***. (Appendix K)***

1. **FUTURE WORKS:**

The extra road that merits investigation is that the first human liver transplantation (LTx) was performed by T.E. Staezl in 1963 [1]. Non-illness explicit insignificant posting rules for Ltx are a Child-Pugh score more prominent than 7. Most Ltx is performed orthotopically, heterotopic transplantations are appointed to cases like intense liver transplant and metabolic sickness. The deficiency of giver organs has prompted the improvement of methods like transplantation of decreased size organs [2]. Primary graft non-work happens in 1-8% of cases and is an intense complexity. Bleeding happens all the more regularly in instances of graft dysfunction and hypothermia.

The likelihood of liver transplant (LTx) and death for applicants shifts by donation administration region (DSA), making it challenging for the candidates to get the logical results explicit to their area [3]. A liver transplant is the last choice for relieving patients with seriously harmed livers that can never again satisfy their metabolic capacity. It is important to assess the variables associated with informal caregiving for liver transplant candidates [2]. An individual encountering an intense liver transplant might be put close to the first spot on the list on the grounds that their risk of death could be more approaching contrasted to somebody with a constant condition.

The MELD score was viewed as better than the CTP score in predicting the 3-months mortality and thusly the MELD score was carried out in 2022 in the United States for the prioritization of LTx beneficiaries [1] [3]. In nations inside the collusion of Eurotransplant, the MELD score for prioritization of patients anticipating LTx was started in November 2006 and at present little data is accessible concerning the prognostic capacity of this designation framework contrasted with the past framework, which depended on CTP score and holding uptime.

1. **RESEARCH REFERENCES:**
2. Robert M. Merion, Douglas E. Schaubel, Dawn M. Dykstra, Richard B. Freeman, Friedrich K. Port, Robert A. Wolfe. (2004). *The Survival Benefit of Liver Transplantation.* American Society of TRANSPLANTATION.

<https://onlinelibrary.wiley.com/doi/full/10.1111/j.1600-6143.2004.00703.x>

1. Eliane Tiemi Miyazki, Randolfo dos Santos Jr, M. Cristina Miyazaki, Neide M. Domingos, Hellen C. Felicio, Marcia F. Rocha, Paulo C. Arroyo Jr, William J. Duca, Renato F. Silva, Rita C.M.A Silva. (2010). *Patients on the waiting list for liver transplantation: Caregiver burden and stress.* AMERICAN ASSOCIATION FOR THE STUDY OF LIVER DISEASES.

<https://aasldpubs.onlinelibrary.wiley.com/doi/full/10.1002/lt.22130>

1. Daniel Gotthardt, Karl Heinz Weiss, Melanie Baumgartner, Alexandra Zahn, Wolfgang Stremmel, Jan Schmidt, Thomas Bruckner & Peter Sauer. (2009). *Limitations of the MELD score in predicting mortality or need for removal from waiting list in patients awaiting liver transplantation.* BMC Gastroenterol.

<https://link.springer.com/article/10.1186/1471-230X-9-72#article-info>

* **APPENDIX A**

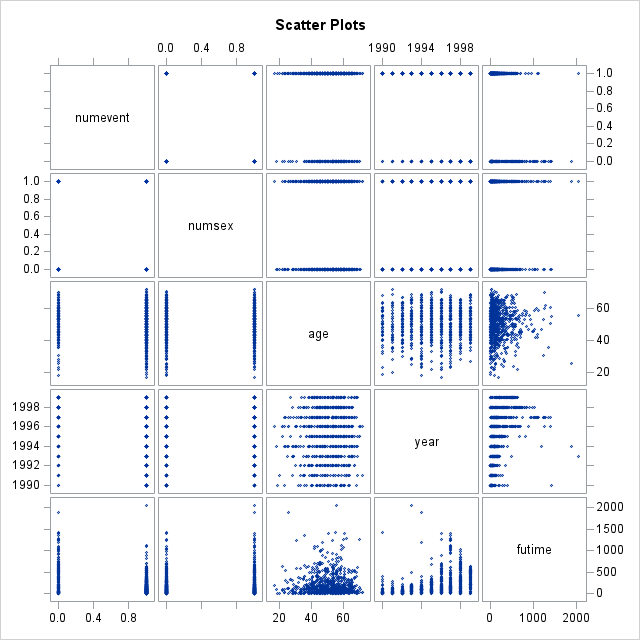
Using the Stepwise and Backward Selection Method

| **Analysis of Maximum Likelihood Estimates** | | | | | | |
| --- | --- | --- | --- | --- | --- | --- |
| **Parameter** |  | **DF** | **Estimate** | **Standard Error** | **Wald Chi-Square** | **Pr > ChiSq** |
| **Intercept** | **censored** | 1 | -263.8 | 78.0473 | 11.4235 | 0.0007 |
| **Intercept** | **death** | 1 | -262.9 | 78.0469 | 11.3444 | 0.0008 |
| **Intercept** | **ltx** | 1 | -257.8 | 77.9840 | 10.9263 | 0.0009 |
| **numsex** |  | 1 | 0.3497 | 0.1776 | 3.8772 | 0.0489 |
| **age** |  | 1 | 0.0238 | 0.00879 | 7.3115 | 0.0069 |
| **year** |  | 1 | 0.1299 | 0.0391 | 11.0121 | 0.0009 |
| **futime** |  | 1 | 0.00313 | 0.000361 | 75.1547 | <.0001 |

| **Odds Ratio Estimates** | | | |
| --- | --- | --- | --- |
| **Effect** | **Point Estimate** | **95% Wald Confidence Limits** | |
| **numsex** | 1.419 | 1.002 | 2.009 |
| **age** | 1.024 | 1.007 | 1.042 |
| **year** | 1.139 | 1.055 | 1.229 |
| **futime** | 1.003 | 1.002 | 1 |

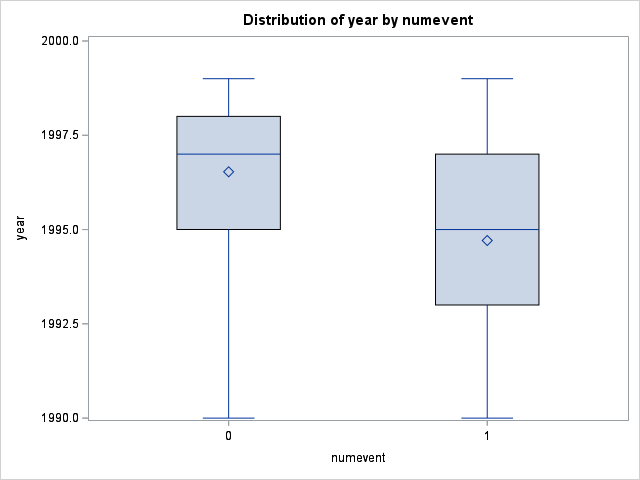
* **APPENDIX B**

Pictorial Representation of Scatter Plots

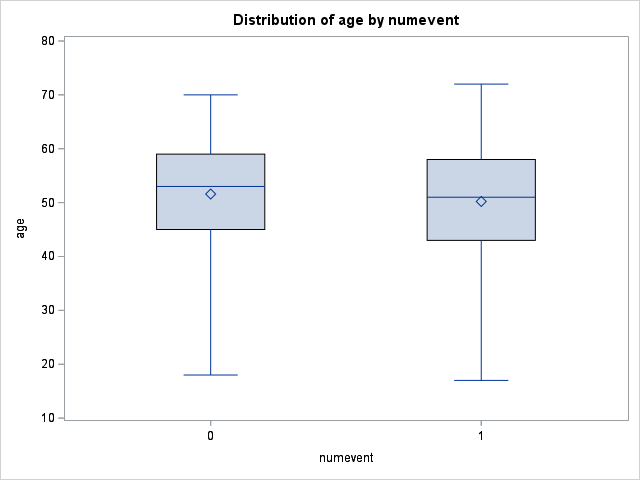


* **APPENDIX C**

Boxplots between numevent vs year



Boxplots between numevent vs year



* **APPENDIX D**

Correlation Coefficients

| **Pearson Correlation Coefficients  Prob > |r| under H0: Rho=0  Number of Observations** | | | | | |
| --- | --- | --- | --- | --- | --- |
|  | **numevent** | **numsex** | **age** | **year** | **futime** |
| **numevent** | |  | | --- | | 1.00000 | |  | | 815 | | |  | | --- | | -0.04660 | | 0.1838 | | 815 | | |  | | --- | | -0.05598 | | 0.1143 | | 797 | | |  | | --- | | -0.28223 | | <.0001 | | 815 | | |  | | --- | | -0.37180 | | <.0001 | | 815 | |
| **numsex** | |  | | --- | | -0.04660 | | 0.1838 | | 815 | | |  | | --- | | 1.00000 | |  | | 815 | | |  | | --- | | -0.02964 | | 0.4033 | | 797 | | |  | | --- | | 0.10075 | | 0.0040 | | 815 | | |  | | --- | | 0.01717 | | 0.6245 | | 815 | |
| **age** | |  | | --- | | -0.05598 | | 0.1143 | | 797 | | |  | | --- | | -0.02964 | | 0.4033 | | 797 | | |  | | --- | | 1.00000 | |  | | 797 | | |  | | --- | | 0.05991 | | 0.0910 | | 797 | | |  | | --- | | 0.00513 | | 0.8851 | | 797 | |
| **year** | |  | | --- | | -0.28223 | | <.0001 | | 815 | | |  | | --- | | 0.10075 | | 0.0040 | | 815 | | |  | | --- | | 0.05991 | | 0.0910 | | 797 | | |  | | --- | | 1.00000 | |  | | 815 | | |  | | --- | | 0.39877 | | <.0001 | | 815 | |
| **futime** | |  | | --- | | -0.37180 | | <.0001 | | 815 | | |  | | --- | | 0.01717 | | 0.6245 | | 815 | | |  | | --- | | 0.00513 | | 0.8851 | | 797 | | |  | | --- | | 0.39877 | | <.0001 | | 815 | | |  | | --- | | 1.00000 | |  | | 815 | |

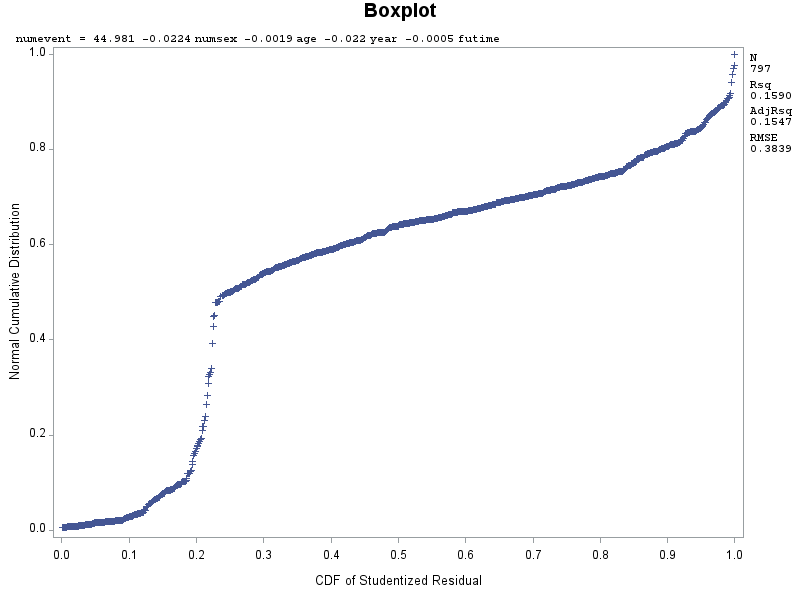
* **APPENDIX E**

Finding Multicollinearity

| **Parameter Estimates** | | | | | | | |
| --- | --- | --- | --- | --- | --- | --- | --- |
| **Variable** | **DF** | **Parameter Estimate** | **Standard Error** | **t Value** | **Pr > |t|** | **Tolerance** | **Variance Inflation** |
| **Intercept** | **1** | 44.98115 | 11.34580 | 3.96 | <.0001 | . | 0 |
| **numsex** | **1** | -0.02241 | 0.02750 | -0.82 | 0.4152 | 0.98808 | 1.01207 |
| **age** | **1** | -0.00190 | 0.00132 | -1.43 | 0.1522 | 0.99465 | 1.00538 |
| **year** | **1** | -0.02205 | 0.00569 | -3.87 | 0.0001 | 0.82241 | 1.21594 |
| **futime** | **1** | -0.00051250 | 0.00005810 | -8.82 | <.0001 | 0.83377 | 1.19937 |

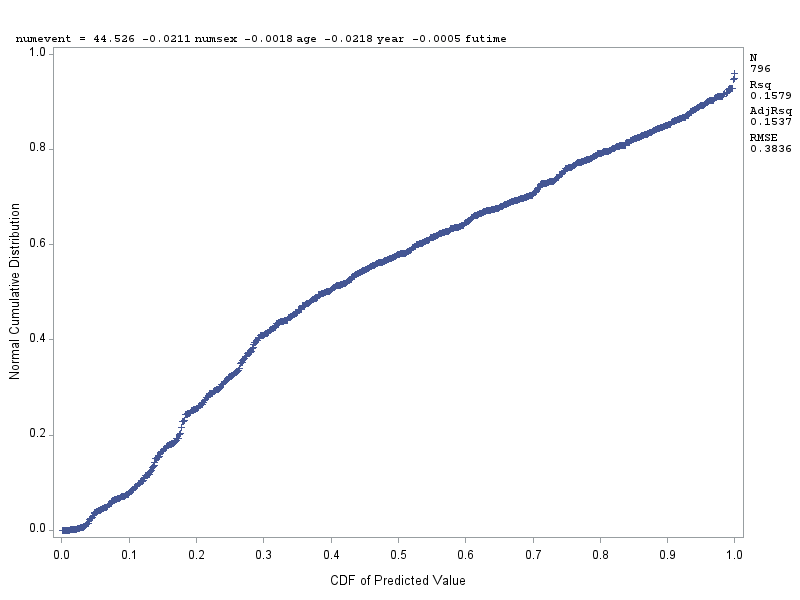
* **APPENDIX F**

Residual Plot after finding VIF and tol



* **APPENDIX G**

Residual Plot after deleting outliers in influential points



* **APPENDIX H**

Training and Test Set

| **Classification Table** | | | | | | | | | |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **Prob Level** | **Correct** | | **Incorrect** | | **Percentages** | | | | |
| **Event** | **Non- Event** | **Event** | **Non- Event** | **Correct** | **Sensi- tivity** | **Speci- ficity** | **Pos Pred** | **Neg Pred** |
| **0.100** | 469 | 0 | 142 | 1 | 76.6 | 99.8 | 0.0 | 76.8 | 0.0 |
| **0.150** | 469 | 0 | 142 | 1 | 76.6 | 99.8 | 0.0 | 76.8 | 0.0 |
| **0.200** | 469 | 0 | 142 | 1 | 76.6 | 99.8 | 0.0 | 76.8 | 0.0 |
| **0.250** | 469 | 1 | 141 | 1 | 76.8 | 99.8 | 0.7 | 76.9 | 50.0 |
| **0.300** | 468 | 4 | 138 | 2 | 77.1 | 99.6 | 2.8 | 77.2 | 66.7 |
| **0.350** | 467 | 8 | 134 | 3 | 77.6 | 99.4 | 5.6 | 77.7 | 72.7 |
| **0.400** | 466 | 13 | 129 | 4 | 78.3 | 99.1 | 9.2 | 78.3 | 76.5 |
| **0.450** | 465 | 22 | 120 | 5 | 79.6 | 98.9 | 15.5 | 79.5 | 81.5 |
| **0.500** | 454 | 35 | 107 | 16 | 79.9 | 96.6 | 24.6 | 80.9 | 68.6 |
| **0.550** | 446 | 48 | 94 | 24 | 80.7 | 94.9 | 33.8 | 82.6 | 66.7 |
| **0.600** | 436 | 58 | 84 | 34 | 80.7 | 92.8 | 40.8 | 83.8 | 63.0 |
| **0.650** | 404 | 63 | 79 | 66 | 76.3 | 86.0 | 44.4 | 83.6 | 48.8 |
| **0.700** | 381 | 69 | 73 | 89 | 73.5 | 81.1 | 48.6 | 83.9 | 43.7 |
| **0.750** | 344 | 81 | 61 | 126 | 69.4 | 73.2 | 57.0 | 84.9 | 39.1 |
| **0.800** | 274 | 104 | 38 | 196 | 61.8 | 58.3 | 73.2 | 87.8 | 34.7 |

| **Obs** | **Selected** | **ID** | **age** | **sex** | **abo** | **year** | **futime** | **event** | **Abo1** | **Abo2** | **Abo3** | **numsex** | **numevent** | **train\_y** | **\_LEVEL\_** | **phat** | **lcl** | **ucl** | **pred\_y** |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **1** | 0 | 5 | 70 | m | O | 1996 | 1271 | censored | 0 | 0 | 0 | 1 | 0 | . | 1 | 0.29184 | 0.15449 | 0.48172 | 0 |
| **2** | 0 | 13 | 38 | f | O | 1994 | 135 | ltx | 0 | 0 | 0 | 0 | 1 | . | 1 | 0.85997 | 0.82218 | 0.89079 | 1 |
| **3** | 0 | 18 | 34 | f | A | 1997 | 97 | ltx | 1 | 0 | 0 | 0 | 1 | . | 1 | 0.77394 | 0.72230 | 0.81839 | 1 |
| **4** | 0 | 22 | 62 | f | O | 1995 | 223 | ltx | 0 | 0 | 0 | 0 | 1 | . | 1 | 0.80531 | 0.76714 | 0.83854 | 1 |
| **5** | 0 | 28 | 50 | m | O | 1990 | 129 | ltx | 0 | 0 | 0 | 1 | 1 | . | 1 | 0.93744 | 0.89289 | 0.96420 | 1 |
| **6** | 0 | 34 | 43 | f | A | 1999 | 310 | ltx | 1 | 0 | 0 | 0 | 1 | . | 1 | 0.59064 | 0.50992 | 0.66676 | 0 |
| **7** | 0 | 37 | 38 | f | O | 1992 | 29 | ltx | 0 | 0 | 0 | 0 | 1 | . | 1 | 0.92173 | 0.88476 | 0.94754 | 1 |
| **8** | 0 | 40 | 47 | m | A | 1995 | 199 | ltx | 1 | 0 | 0 | 1 | 1 | . | 1 | 0.81270 | 0.77527 | 0.84514 | 1 |
| **9** | 0 | 51 | 54 | m | O | 1991 | 1 | ltx | 0 | 0 | 0 | 1 | 1 | . | 1 | 0.93945 | 0.90350 | 0.96256 | 1 |
| **10** | 0 | 56 | 62 | f | B | 1991 | 18 | death | 0 | 1 | 0 | 0 | 0 | . | 1 | 0.93749 | 0.90065 | 0.96126 | 1 |
| **11** | 0 | 58 | 51 | f | O | 1992 | 34 | ltx | 0 | 0 | 0 | 0 | 1 | . | 1 | 0.92101 | 0.88381 | 0.94701 | 1 |
| **12** | 0 | 59 | 46 | m | AB | 1998 | 346 | ltx | 0 | 0 | 1 | 1 | 1 | . | 1 | 0.62597 | 0.56387 | 0.68418 | 1 |
| **13** | 0 | 62 | 67 | f | O | 1993 | 166 | ltx | 0 | 0 | 0 | 0 | 1 | . | 1 | 0.87797 | 0.83576 | 0.91049 | 1 |
| **14** | 0 | 68 | 64 | m | O | 1994 | 104 | ltx | 0 | 0 | 0 | 1 | 1 | . | 1 | 0.86724 | 0.83026 | 0.89715 | 1 |
| **15** | 0 | 71 | 31 | f | B | 1996 | 168 | ltx | 0 | 1 | 0 | 0 | 1 | . | 1 | 0.78740 | 0.74830 | 0.82187 | 1 |

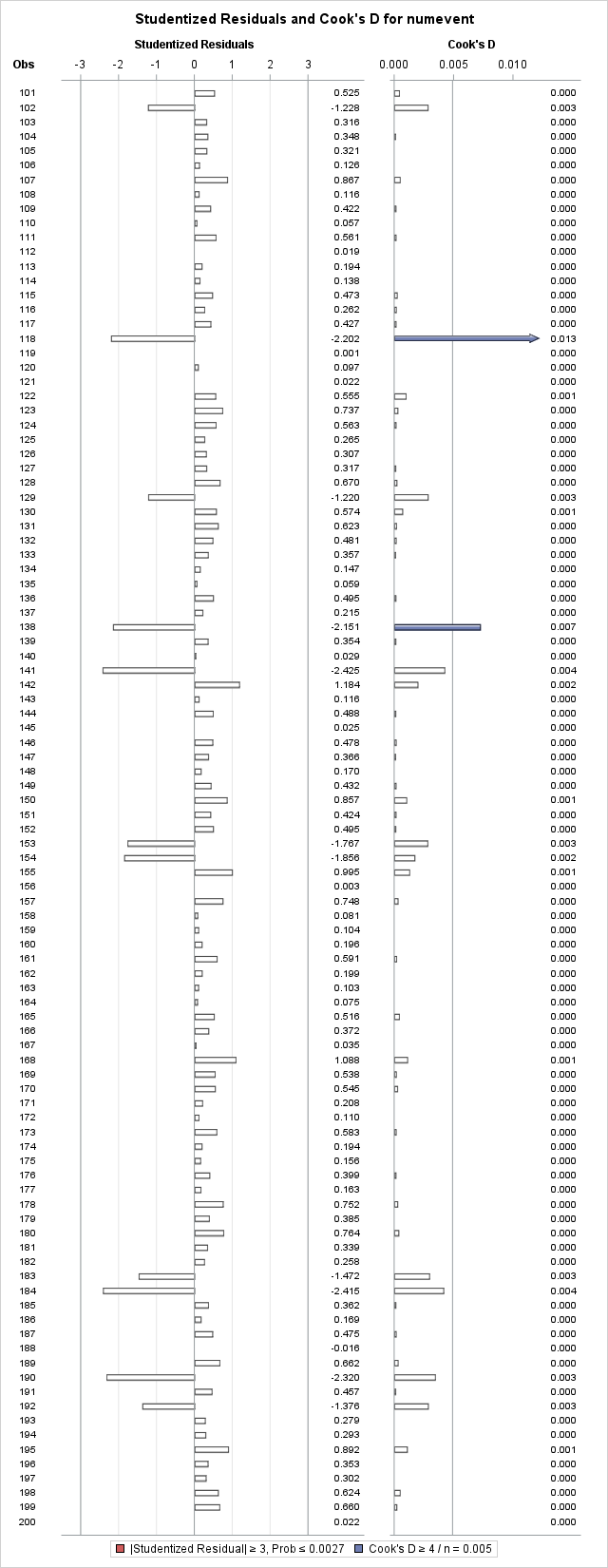
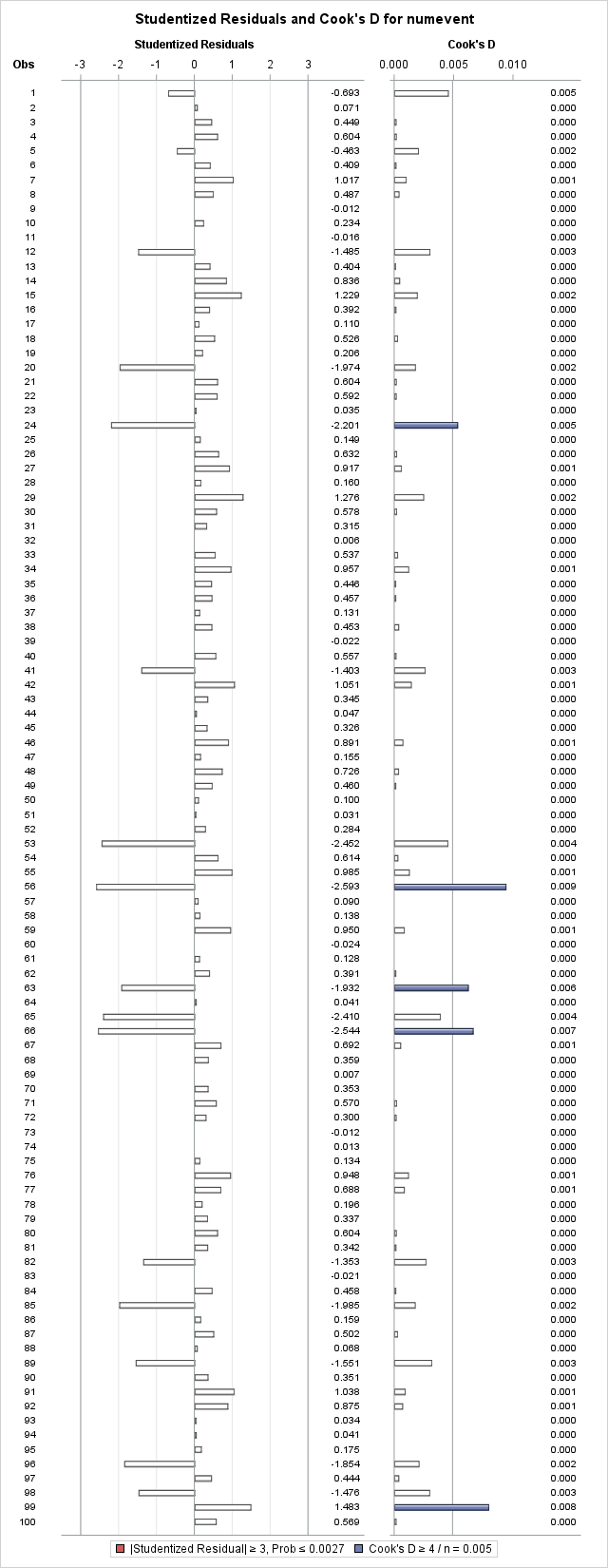
* **APPENDIX I**

Finding Multicollinearity for Final Model

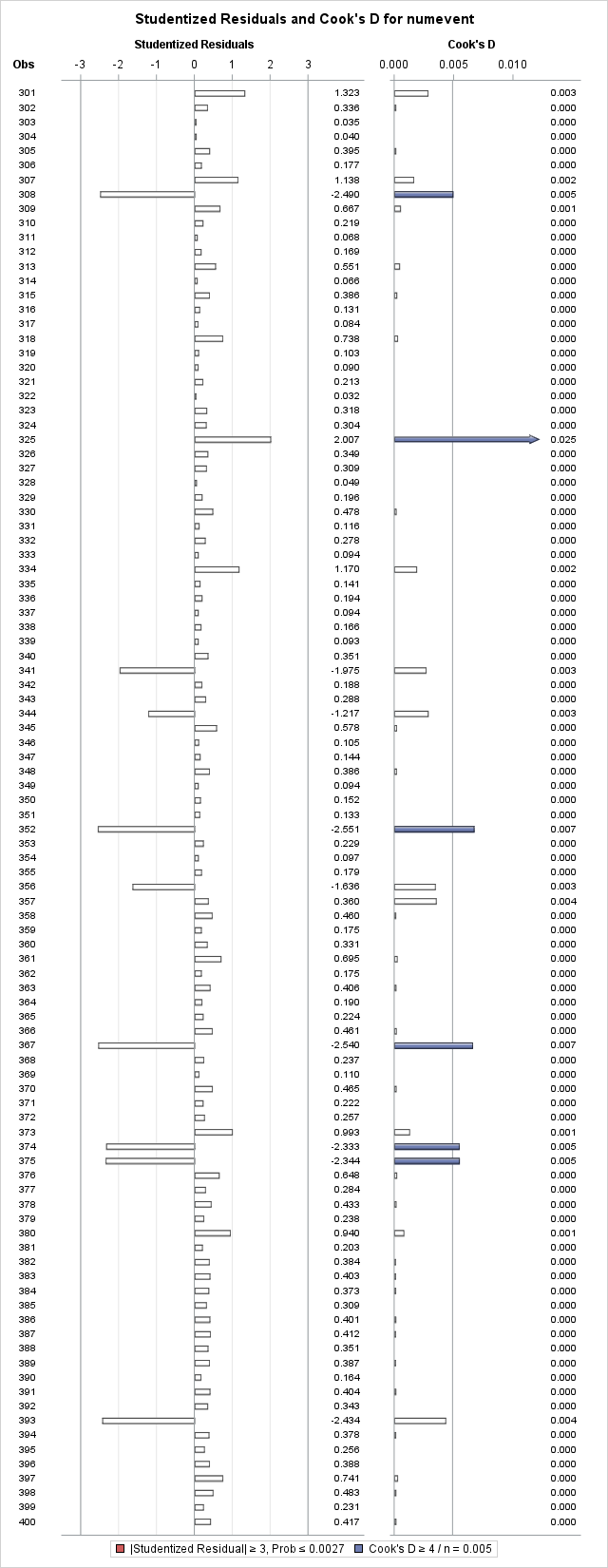
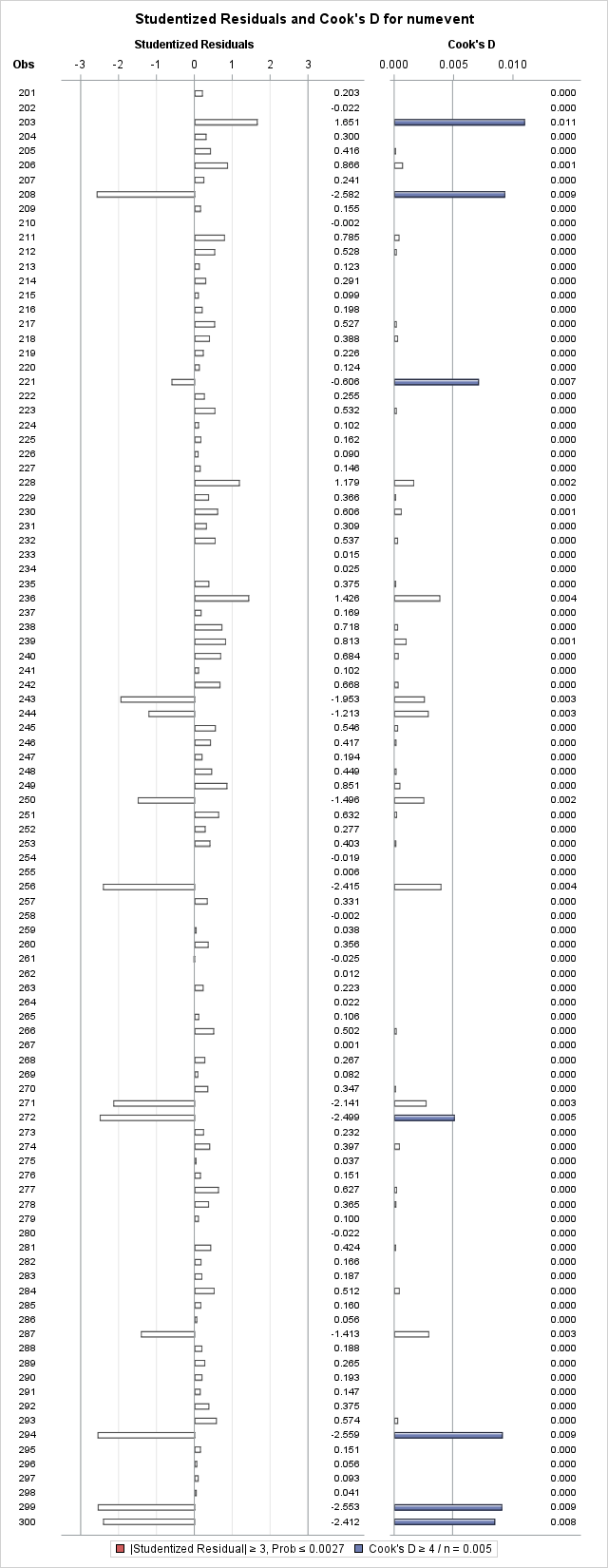
| **Parameter Estimates** | | | | | | | |
| --- | --- | --- | --- | --- | --- | --- | --- |
| **Variable** | **DF** | **Parameter Estimate** | **Standard Error** | **t Value** | **Pr > |t|** | **Tolerance** | **Variance Inflation** |
| **Intercept** | **1** | 50.12613 | 10.83902 | 4.62 | <.0001 | . | 0 |
| **year** | **1** | -0.02468 | 0.00544 | -4.54 | <.0001 | 0.84099 | 1.18908 |
| **futime** | **1** | -0.00050248 | 0.00005718 | -8.79 | <.0001 | 0.84099 | 1.18908 |

* **APPENDIX J**

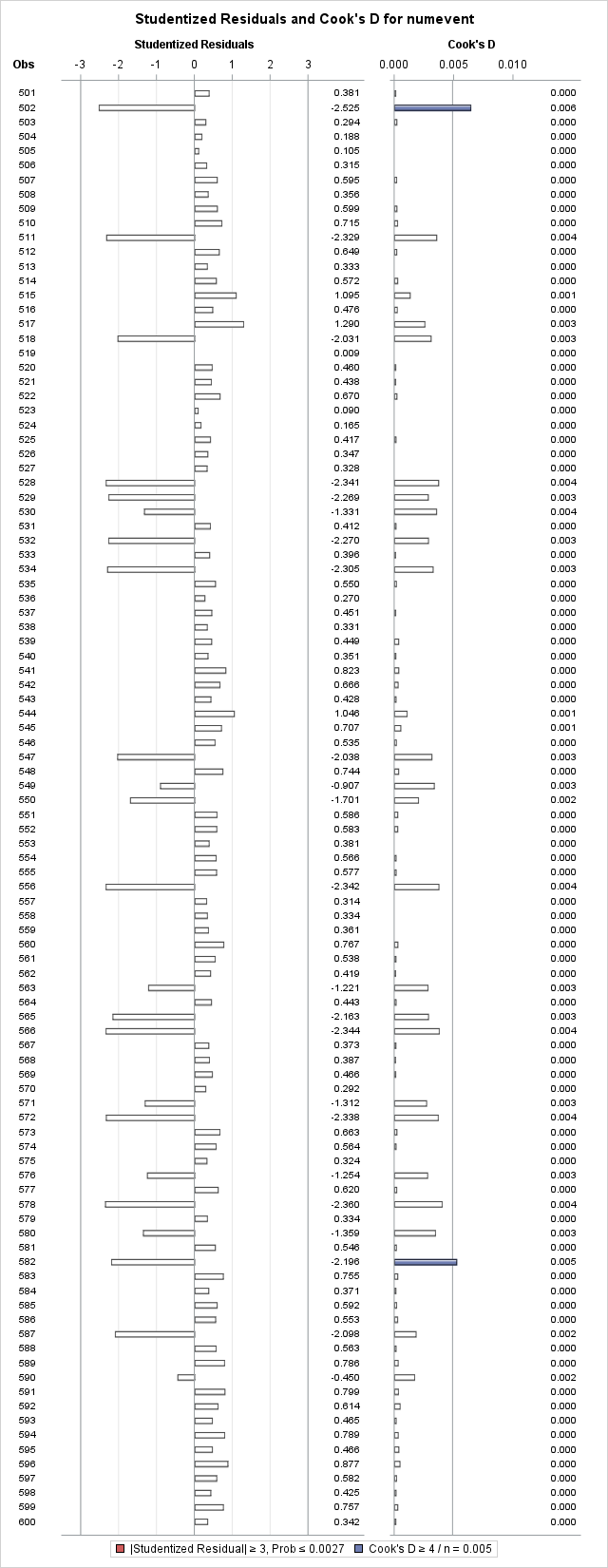
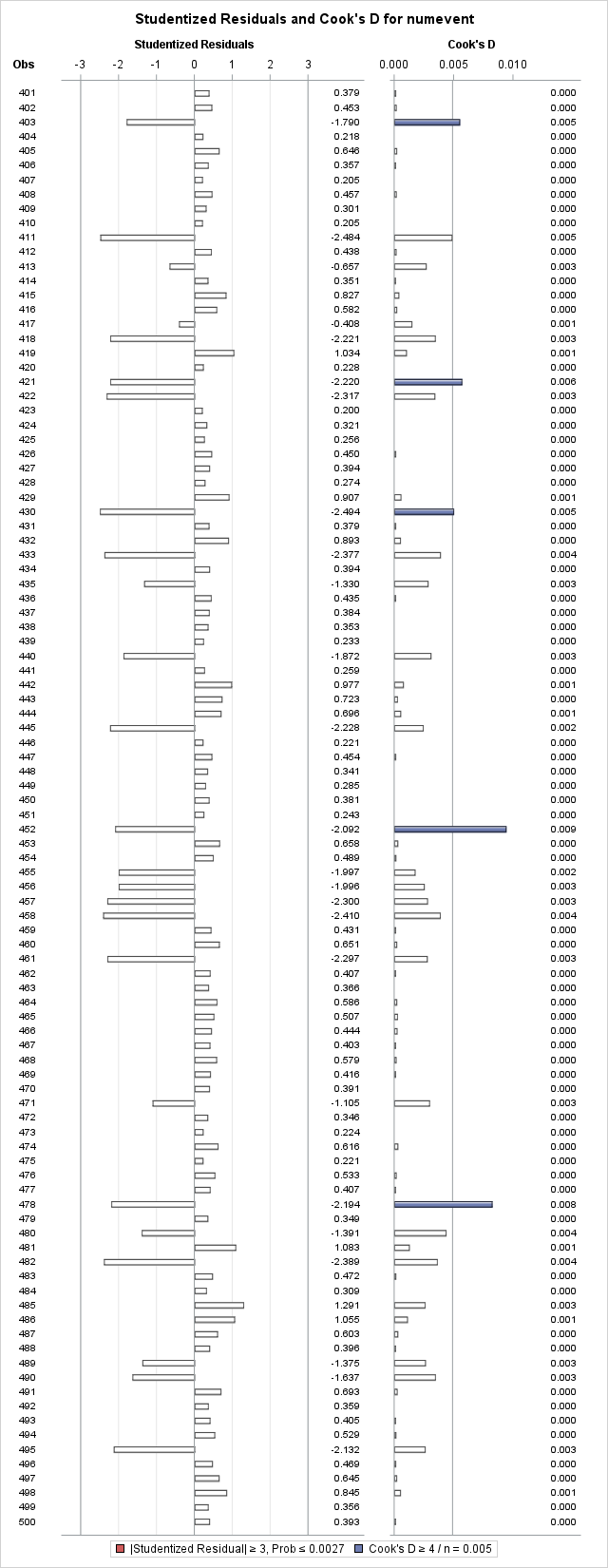
Finding and eliminating the outliers present in the Influential Points



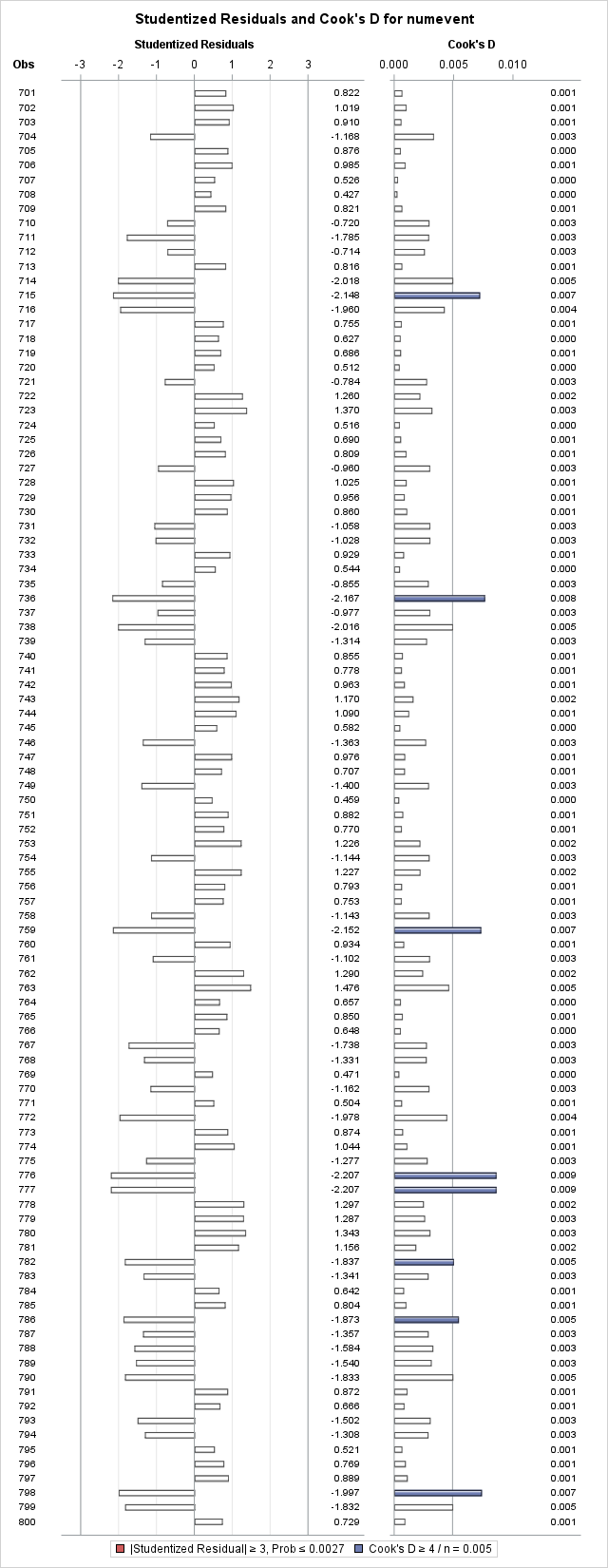
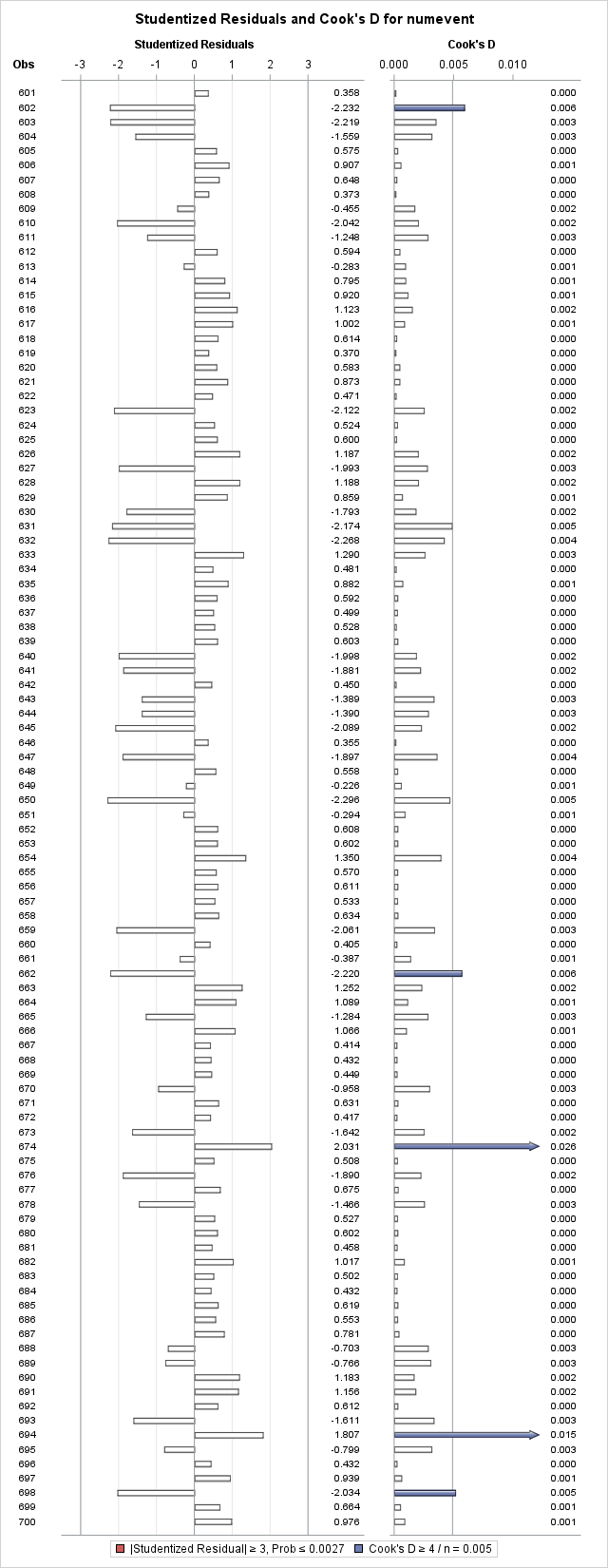
From Observation 1 - 200



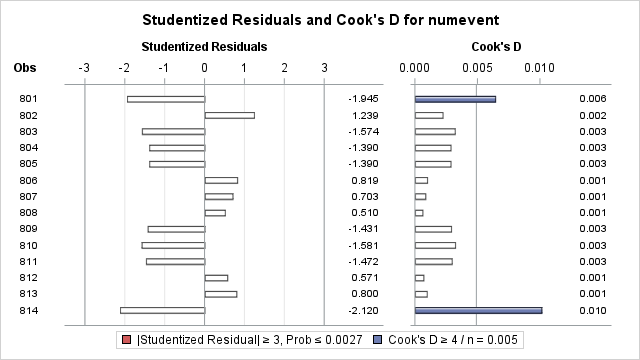
From Observation 201 - 400



From Observation 401 - 600



From Observation 601 - 800



From Observation 801 - 814

* **APPENDIX K**

Studentized Residual Plot and Normality for the Final Model

