**Predicting the survival of patients based on age, ejection fraction, serum creatine level, age and follow-up period.**

By Annmay Sharma

annmays2@illinois.edu

University of Illinois at Urbana-Champaign

**Introduction**

Cardiovascular diseases are an important cause of death worldwide, with nearly 17 million people per year. These diseases include coronary heart disease (heart attacks), cerebrovascular diseases (strokes), heart failure and others. Through our statistical analysis we want to understand more about the survival of patients suffering from such heart diseases. For this, we will conduct an initial analysis on the data to understand the distribution of the various continuous and categorical predictors. Thereafter, we will try to conduct regression on the target variable that is whether the patient deceased during the follow-up period. Based off the model and the statistical significance of various predictors, we will try to come up with a regression model that is able to predict whether the patient will survive or not during the follow up period given a set of predictors.

**Background**

**Dataset**

The dataset we will be working on is a dataset of heart failure clinical records obtained from the UCI machine learning repository. This dataset contains the medical record of 299 heart failure patients collected at the Faisalabad Institute of Cardiology and at the Allied Hospital in Faisalabad (Punjab, Pakistan), during April-December 2015. The patients consisted of 105 women and 194 men, and their ages range between 40 and 95 years old. All 299 patients had left ventricular systolic dysfunction and had previous heart failures that put them in classes III and IV of New York Heart Association (NYHA) classification of the stages of heart failure.

We have 13 features in this dataset that report body, and lifestyle information. For the sake of statistical analysis, we can split the features into two categories – continuous and categorical. The features that are quantitative or continuous are age (in years), level of the creatine phosphokinase enzyme in the blood, percentage of blood leaving the heart at each contraction, platelets in the blood, level of serum creatine in the blood, level of serum sodium in the blood, and the follow-up period (in days). The features that are categorical include whether the patient has anaemia, if the patient has hypertension, whether the patient is diabetic, the sex of the person (0 for woman, 1 for man), and whether the patient is a smoker.

**Initial Data Analysis**

We conduct an initial analysis on the various variables in the dataset to better understand the distribution of these features before we use them as predictors in our regression analysis. This includes checking the percentage of our binary variables, checking the normality in our continuous variables, and checking certain scatter plots and relationships between variables.

The first analysis is to check the range of the various variables and the distribution, starting with the categorical variables. Below, we can see the frequency of the binary categorical variables and see whether they are primarily true or false. Most of the Boolean variables (anaemia, hbp and smoking) are more skewed towards being false.

| **anaemia** | **Frequency** | **Percent** | **Cumulative Frequency** | **Cumulative Percent** |
| --- | --- | --- | --- | --- |
| **0** | 170 | 56.86 | 170 | 56.86 |
| **1** | 129 | 43.14 | 299 | 100.00 |

| **diab** | **Frequency** | **Percent** | **Cumulative Frequency** | **Cumulative Percent** |
| --- | --- | --- | --- | --- |
| **0** | 174 | 58.19 | 174 | 58.19 |
| **1** | 125 | 41.81 | 299 | 100.00 |

| **hbp** | **Frequency** | **Percent** | **Cumulative Frequency** | **Cumulative Percent** |
| --- | --- | --- | --- | --- |
| **0** | 194 | 64.88 | 194 | 64.88 |
| **1** | 105 | 35.12 | 299 | 100.00 |

| **sex** | **Frequency** | **Percent** | **Cumulative Frequency** | **Cumulative Percent** |
| --- | --- | --- | --- | --- |
| **0** | 105 | 35.12 | 105 | 35.12 |
| **1** | 194 | 64.88 | 299 | 100.00 |

| **smk** | **Frequency** | **Percent** | **Cumulative Frequency** | **Cumulative Percent** |
| --- | --- | --- | --- | --- |
| **0** | 203 | 67.89 | 203 | 67.89 |
| **1** | 96 | 32.11 | 299 | 100.00 |

| **Death\_E** | **Frequency** | **Percent** | **Cumulative Frequency** | **Cumulative Percent** |
| --- | --- | --- | --- | --- |
| **0** | 203 | 67.89 | 203 | 67.89 |
| **1** | 96 | 32.11 | 299 | 100.00 |

Thereafter, we will conduct an analysis on the distribution of the continuous variables. In order to check the distribution, we will analyze the mean, standard deviation, and the range of these features. We can see that some of the variables like age and level of serum sodium have a lower relative standard deviation compared to our other variables. This might factor into checking whether these features also have a normal distribution or not.

| **Variable** | **N** | **Mean** | **Std Dev** | **Minimum** | **Maximum** |
| --- | --- | --- | --- | --- | --- |
| age creatinine\_p eject\_f plat serum\_c serum\_s time | 299 299 299 299 299 299 299 | 60.8338930 581.8394649 38.0836120 263358.03 1.3938796 136.6254181 130.2608696 | 11.8948091 970.2878807 11.8348407 97804.24 1.0345101 4.4124773 77.6142080 | 40.0000000 23.0000000 14.0000000 25100.00 0.5000000 113.0000000 4.0000000 | 95.0000000 7861.00 80.0000000 850000.00 9.4000000 148.0000000 285.0000000 |

The above claim of assuming normal distributions for age and serum sodium are supported with graphs against normal distribution and goodness-of-fit tests for normal distributions. For example, these are the results for the goodness-of-fit tests for age with supporting histograms.

| **Tests for Normality** | | | | |
| --- | --- | --- | --- | --- |
| **Test** | **Statistic** | | **p Value** | |
| **Shapiro-Wilk** | **W** | 0.97547 | **Pr < W** | <0.0001 |
| **Kolmogorov-Smirnov** | **D** | 0.069751 | **Pr > D** | <0.0100 |
| **Cramer-von Mises** | **W-Sq** | 0.235874 | **Pr > W-Sq** | <0.0050 |
| **Anderson-Darling** | **A-Sq** | 1.642448 | **Pr > A-Sq** | <0.0050 |



We can see that we get very low p-values for all the goodness-of-fit tests that means that we can reject the null that the distribution of these features is not normal. Similarly, for the goodness-of-fit tests for serum sodium we get a very low p-value supporting that this feature is normally distributed.

These analyses on the variables are also seen in the probability plots of these features that show a clear normal distribution. An example of the same is this scatter plot for the platelet count of patients.



The distribution shows that this is normally distributed and gives further evidence towards the normality of the feature.

**Relationship analysis of target variable and predictors**

Upon the initial analysis of the features independently, we will shift our focus to the target variable that is the whether the patient survives or not during the follow-up period and the various continuous and categorical variables. In order to see the correlation between the continuous variables and whether the patient deceased during the follow-up period, I used pearson correlation to see how high the correlation values would be and how low the p-values would be for the various features.

| **Pearson Correlation Coefficients, N = 299 Prob > |r| under H0: Rho=0** | | | | | | | |
| --- | --- | --- | --- | --- | --- | --- | --- |
|  | **creatinine\_p** | **plat** | **serum\_c** | **serum\_s** | **eject\_f** | **time** | **age** |
| **Death\_E** | 0.06273 0.2796 | -0.04914 0.3972 | 0.29428 <.0001 | -0.19520 0.0007 | -0.26860 <.0001 | -0.52696 <.0001 | 0.25373 <.0001 |

The above pearson correlation factors show us some strong correlations between the target variable and age, time, level of serum creatinine and the ejection fraction. The p-value for these variables is also very low showing strong statistical significance. While variables like time and ejection fraction are negatively correlated to target variable (meaning that they negatively contribute to the death of the patient), variables such as age contributes towards the death of patients. This makes sense as an older patient is more likely to die than a younger patient.

Now, let us have a look at the categorical variables and their relationship with the target variable. For this we will construct frequency boxes to see if there is direct correlation between them i.e is there a particular diagonal in the box heavily skewed that implies that the variable is negatively or positively correlated.

While the analysis is conducted on all categorical variables, we will just look at the results of the correlation between death and smoking. The remaining analyses tables will be attached in the appendix.

| **Table of smk by Death\_E** | | | |
| --- | --- | --- | --- |
| **smk** | **Death\_E** | | |
| **Frequency Percent** | **0** | **1** | **Total** |
| **0** | 137 45.82 | 66 22.07 | 203 67.89 |
| **1** | 66 22.07 | 30 10.03 | 96 32.11 |
| **Total** | 203 67.89 | 96 32.11 | 299 100.00 |

| **Statistic** | **DF** | **Value** | **Prob** |
| --- | --- | --- | --- |
| **Chi-Square** | 1 | 0.0476 | 0.8272 |
| **Likelihood Ratio Chi-Square** | 1 | 0.0478 | 0.8270 |
| **Continuity Adj. Chi-Square** | 1 | 0.0073 | 0.9318 |
| **Mantel-Haenszel Chi-Square** | 1 | 0.0475 | 0.8275 |
| **Phi Coefficient** |  | -0.0126 |  |
| **Contingency Coefficient** |  | 0.0126 |  |
| **Cramer's V** |  | -0.0126 |  |

| **Fisher's Exact Test** | |
| --- | --- |
| **Cell (1,1) Frequency (F)** | 137 |
| **Left-sided Pr <= F** | 0.4682 |
| **Right-sided Pr >= F** | 0.6352 |
|  |  |
| **Table Probability (P)** | 0.1034 |
| **Two-sided Pr <= P** | 0.8947 |

From the above analyses, we can see that there is no real serial correlation between being a smoker and dying during the follow-up period. The initial frequency box clearly shows a distributed spread of the frequencies of the various combinations which shows us that there is no clear correlation between the two.

The Chi-square value is really low at 0.0476 and the p-value is really high at 0.8272 which means that we cannot reject the null hypothesis that there is no serial correlation between the patient being a smoker and the occurrence of death of the patient during the follow-up period. Similarly, looking at the F vaue of the 1, 1 cell – it is a realtively lower value. The P-value of the one-sided test is really high giving furtherr evidence that the two variables are not serially correlated.

Hence, we can conclude that including the categorical variables in the regression would not give us optimal results as they do not statistically contribute towards predicting if a patient is more or less likely to die during the follow-up period.

**Logistic Regression Analysis**

Upon, the initial analysis on the data and eliminating the insignificant values of the dataset we can now look into the model itself. Let us perform a logistic regression with the death\_E as the target variable and the quantitative, statistically significant variables as the predictors.

| **Model Information** | |
| --- | --- |
| **Data Set** | WORK.HEARTF |
| **Response Variable** | Death\_E |
| **Number of Response Levels** | 2 |
| **Model** | binary logit |
| **Optimization Technique** | Fisher's scoring |

| **Model Fit Statistics** | | |
| --- | --- | --- |
| **Criterion** | **Intercept Only** | **Intercept and Covariates** |
| **AIC** | 377.349 | 235.486 |
| **SC** | 381.049 | 257.689 |
| **-2 Log L** | 375.349 | 223.486 |

| **Testing Global Null Hypothesis: BETA=0** | | | |
| --- | --- | --- | --- |
| **Test** | **Chi-Square** | **DF** | **Pr > ChiSq** |
| **Likelihood Ratio** | 151.8625 | 5 | <.0001 |
| **Score** | 121.7373 | 5 | <.0001 |
| **Wald** | 71.5715 | 5 | <.0001 |

| **Analysis of Maximum Likelihood Estimates** | | | | | |
| --- | --- | --- | --- | --- | --- |
| **Parameter** | **DF** | **Estimate** | **Standard Error** | **Wald Chi-Square** | **Pr > ChiSq** |
| **Intercept** | 1 | -9.4930 | 5.4058 | 3.0839 | 0.0791 |
| **serum\_c** | 1 | -0.6860 | 0.1740 | 15.5352 | <.0001 |
| **serum\_s** | 1 | 0.0646 | 0.0384 | 2.8297 | 0.0925 |
| **eject\_f** | 1 | 0.0734 | 0.0158 | 21.6394 | <.0001 |
| **time** | 1 | 0.0209 | 0.00292 | 51.3474 | <.0001 |
| **age** | 1 | -0.0425 | 0.0150 | 7.9834 | 0.0047 |

| **Odds Ratio Estimates** | | | |
| --- | --- | --- | --- |
| **Effect** | **Point Estimate** | **95% Wald Confidence Limits** | |
| **serum\_c** | 0.504 | 0.358 | 0.708 |
| **serum\_s** | 1.067 | 0.989 | 1.150 |
| **eject\_f** | 1.076 | 1.043 | 1.110 |
| **time** | 1.021 | 1.015 | 1.027 |
| **age** | 0.958 | 0.931 | 0.987 |

The above logistic regression is statistically significant but can still be improved upon. While the AIC values are lower realtive to similar models, it can be reduced to a smalller value. Upon seeing the significance of individual predictors we can say that serum osdium is not as significant as we initially thought it was. To produce our best model, we will now conduct backward model selection and see which model is the best to determine the death effect of the patient.

Forward and stepwise selection are also good ways to choose a model, but I chose backward selection since I already had narrowed the model to a few features and wanted to eliminate unnecessary features. Furthermore, the forward selection and stepwise selection would take a longer computing time given that I had 13 features and it could be more baised.

Upon backward selection, we see that the regresssion selected is with the same predictors as above except serum sodium as expected. However, we see a negligible change in the AIC of the model. But, the predictors of the model are now clearly statistically significant at even a 99% confidence interval. Hence, we can go ahead and try to use a different approach to this log model to get the best results.

The results of the best model according to backward selection are below.

| **Model Fit Statistics** | | |
| --- | --- | --- |
| **Criterion** | **Intercept Only** | **Intercept and Covariates** |
| **AIC** | 377.349 | 236.301 |
| **SC** | 381.049 | 254.804 |
| **-2 Log L** | 375.349 | 226.301 |

| **Testing Global Null Hypothesis: BETA=0** | | | |
| --- | --- | --- | --- |
| **Test** | **Chi-Square** | **DF** | **Pr > ChiSq** |
| **Likelihood Ratio** | 149.0474 | 4 | <.0001 |
| **Score** | 120.3402 | 4 | <.0001 |
| **Wald** | 72.3323 | 4 | <.0001 |

| **Analysis of Maximum Likelihood Estimates** | | | | | |
| --- | --- | --- | --- | --- | --- |
| **Parameter** | **DF** | **Estimate** | **Standard Error** | **Wald Chi-Square** | **Pr > ChiSq** |
| **Intercept** | 1 | -0.6045 | 1.0362 | 0.3403 | 0.5596 |
| **serum\_c** | 1 | -0.7198 | 0.1746 | 16.9936 | <.0001 |
| **eject\_f** | 1 | 0.0748 | 0.0156 | 23.1232 | <.0001 |
| **time** | 1 | 0.0206 | 0.00288 | 51.1545 | <.0001 |
| **age** | 1 | -0.0433 | 0.0149 | 8.4862 | 0.0036 |

| **Odds Ratio Estimates** | | | |
| --- | --- | --- | --- |
| **Effect** | **Point Estimate** | **95% Wald Confidence Limits** | |
| **serum\_c** | 0.487 | 0.346 | 0.686 |
| **eject\_f** | 1.078 | 1.045 | 1.111 |
| **time** | 1.021 | 1.015 | 1.027 |
| **age** | 0.958 | 0.930 | 0.986 |

Now, in order to get the best fit of the model, I had to to use the GENMOD procedure to obtain a lower AIC. Since the regression is a log regresssion, I used the log it link function and a binomila distribution to fit the model. Along with this, the use of only the statistically significant predictors that we have found will enable us to fit the best model.

| **Model Information** | |
| --- | --- |
| **Data Set** | WORK.HEARTF |
| **Distribution** | Binomial |
| **Link Function** | Logit |
| **Dependent Variable** | Death\_E |

| **Criteria For Assessing Goodness Of Fit** | | | |
| --- | --- | --- | --- |
| **Criterion** | **DF** | **Value** | **Value/DF** |
| **Log Likelihood** |  | -113.1507 |  |
| **Full Log Likelihood** |  | -113.1507 |  |
| **AIC (smaller is better)** |  | 236.3014 |  |
| **AICC (smaller is better)** |  | 236.5062 |  |
| **BIC (smaller is better)** |  | 254.8036 |  |

| **LR Statistics For Type 1 Analysis** | | | | |
| --- | --- | --- | --- | --- |
| **Source** | **Deviance** | **DF** | **Chi-Square** | **Pr > ChiSq** |
| **Intercept** | 375.3488 |  |  |  |
| **serum\_c** | 347.2521 | 1 | 28.10 | <.0001 |
| **eject\_f** | 324.3211 | 1 | 22.93 | <.0001 |
| **time** | 235.4141 | 1 | 88.91 | <.0001 |
| **age** | 226.3014 | 1 | 9.11 | 0.0025 |

| **LR Statistics For Type 3 Analysis** | | | |
| --- | --- | --- | --- |
| **Source** | **DF** | **Chi-Square** | **Pr > ChiSq** |
| **serum\_c** | 1 | 18.21 | <.0001 |
| **eject\_f** | 1 | 28.32 | <.0001 |
| **time** | 1 | 78.98 | <.0001 |
| **age** | 1 | 9.11 | 0.0025 |

We can see a reduced AIC upon fitting the model using the GENMOD procedure. This is because of the binomial fit we use for the target variable. The likelihood ratios from type 1 analysis is reallty high which shows us that the features are statistically significant in this fit. Similarly, all features are also statistically significant when we conduct type 3 analysis.

Now, let us take this model that we have arrived at and analyse the predictions. We are going to take the predicted frequencies and compare them with the observed frequencies of the dataset (0 if patient survived or 1 if patient is deceased)

| **Table of Death\_E by \_INTO\_** | | | |
| --- | --- | --- | --- |
| **Death\_E** | **\_INTO\_(Formatted Value of the Predicted Response)** | | |
| **Frequency Percent** | **0** | **1** | **Total** |
| **0** | 183 61.20 | 20 6.69 | 203 67.89 |
| **1** | 30 10.03 | 66 22.07 | 96 32.11 |
| **Total** | 213 71.24 | 86 28.76 | 299 100.00 |

From the above predictions, we can see that there is a misclassification error rate close to 17%. This means that out of the 299 observations, the model was able to predict 249 of the observations correctly. This means that there is room for improvement in the model and we could work to find more predictors that show a stronger correlation to the probability of the patient dying during the follow-up period.

**Conclusion**

Upon statistical analysis to determine which features provide the highest contribution towards the decease of patients during the follow-up period, we concluded that the main predictors were ejection fraction, time of follow-up period, age, and serum creatinine. The final model to predict the log odds of the patient dying during the follow-up period is –

Log(odds(patient dying) = 0.6045 + 0.7198 \* serum creatinine – 0.0748 \* ejection fraction – 0.0206 \* time + 0.0433 \* age.

This is our best model upon analysing correlations of various features, the individual features independently, and checking the statistical significance of the features in regresssion models. The dataset is relatively small and has a lot of insignificant features which can lead to overfitting issues. If we had more information about the physical features, the hereditery issues and the past medical history of the patients, we could have conducted a more accurate logistic regression for predicting the probability of the patient dying duirng the follow-up period.

**Appendix**













| **Tests for Normality** | | | | |
| --- | --- | --- | --- | --- |
| **Test** | **Statistic** | | **p Value** | |
| **Shapiro-Wilk** | **W** | 0.911509 | **Pr < W** | <0.0001 |
| **Kolmogorov-Smirnov** | **D** | 0.116068 | **Pr > D** | <0.0100 |
| **Cramer-von Mises** | **W-Sq** | 0.924776 | **Pr > W-Sq** | <0.0050 |
| **Anderson-Darling** | **A-Sq** | 4.989043 | **Pr > A-Sq** | <0.0050 |



| **Goodness-of-Fit Tests for Normal Distribution** | | | | |
| --- | --- | --- | --- | --- |
| **Test** | **Statistic** | | **p Value** | |
| **Kolmogorov-Smirnov** | **D** | 0.11606814 | **Pr > D** | <0.010 |
| **Cramer-von Mises** | **W-Sq** | 0.92477561 | **Pr > W-Sq** | <0.005 |
| **Anderson-Darling** | **A-Sq** | 4.98904312 | **Pr > A-Sq** | <0.005 |

| **Tests for Normality** | | | | |
| --- | --- | --- | --- | --- |
| **Test** | **Statistic** | | **p Value** | |
| **Shapiro-Wilk** | **W** | 0.514263 | **Pr < W** | <0.0001 |
| **Kolmogorov-Smirnov** | **D** | 0.286765 | **Pr > D** | <0.0100 |
| **Cramer-von Mises** | **W-Sq** | 8.096309 | **Pr > W-Sq** | <0.0050 |
| **Anderson-Darling** | **A-Sq** | 41.90613 | **Pr > A-Sq** | <0.0050 |



| **Goodness-of-Fit Tests for Normal Distribution** | | | | |
| --- | --- | --- | --- | --- |
| **Test** | **Statistic** | | **p Value** | |
| **Kolmogorov-Smirnov** | **D** | 0.2867647 | **Pr > D** | <0.010 |
| **Cramer-von Mises** | **W-Sq** | 8.0963088 | **Pr > W-Sq** | <0.005 |
| **Anderson-Darling** | **A-Sq** | 41.9061265 | **Pr > A-Sq** | <0.005 |

|  |  |
| --- | --- |
| **1 With Variables:** | Death\_E |
| **7 Variables:** | creatinine\_p plat serum\_c serum\_s eject\_f time age |

| **Simple Statistics** | | | | | | |
| --- | --- | --- | --- | --- | --- | --- |
| **Variable** | **N** | **Mean** | **Std Dev** | **Sum** | **Minimum** | **Maximum** |
| **Death\_E** | 299 | 0.32107 | 0.46767 | 96.00000 | 0 | 1.00000 |
| **creatinine\_p** | 299 | 581.83946 | 970.28788 | 173970 | 23.00000 | 7861 |
| **plat** | 299 | 263358 | 97804 | 78744051 | 25100 | 850000 |
| **serum\_c** | 299 | 1.39388 | 1.03451 | 416.77000 | 0.50000 | 9.40000 |
| **serum\_s** | 299 | 136.62542 | 4.41248 | 40851 | 113.00000 | 148.00000 |
| **eject\_f** | 299 | 38.08361 | 11.83484 | 11387 | 14.00000 | 80.00000 |
| **time** | 299 | 130.26087 | 77.61421 | 38948 | 4.00000 | 285.00000 |
| **age** | 299 | 60.83389 | 11.89481 | 18189 | 40.00000 | 95.00000 |

| **Pearson Correlation Coefficients, N = 299 Prob > |r| under H0: Rho=0** | | | | | | | |
| --- | --- | --- | --- | --- | --- | --- | --- |
|  | **creatinine\_p** | **plat** | **serum\_c** | **serum\_s** | **eject\_f** | **time** | **age** |
| **Death\_E** | 0.06273 0.2796 | -0.04914 0.3972 | 0.29428 <.0001 | -0.19520 0.0007 | -0.26860 <.0001 | -0.52696 <.0001 | 0.25373 <.0001 |

| **Table of smk by Death\_E** | | | |
| --- | --- | --- | --- |
| **smk** | **Death\_E** | | |
| **Frequency Percent** | **0** | **1** | **Total** |
| **0** | 137 45.82 | 66 22.07 | 203 67.89 |
| **1** | 66 22.07 | 30 10.03 | 96 32.11 |
| **Total** | 203 67.89 | 96 32.11 | 299 100.00 |

|  |
| --- |
| ***Statistics for Table of smk by Death\_E*** |

| **Statistic** | **DF** | **Value** | **Prob** |
| --- | --- | --- | --- |
| **Chi-Square** | 1 | 0.0476 | 0.8272 |
| **Likelihood Ratio Chi-Square** | 1 | 0.0478 | 0.8270 |
| **Continuity Adj. Chi-Square** | 1 | 0.0073 | 0.9318 |
| **Mantel-Haenszel Chi-Square** | 1 | 0.0475 | 0.8275 |
| **Phi Coefficient** |  | -0.0126 |  |
| **Contingency Coefficient** |  | 0.0126 |  |
| **Cramer's V** |  | -0.0126 |  |

| **Fisher's Exact Test** | |
| --- | --- |
| **Cell (1,1) Frequency (F)** | 137 |
| **Left-sided Pr <= F** | 0.4682 |
| **Right-sided Pr >= F** | 0.6352 |
|  |  |
| **Table Probability (P)** | 0.1034 |
| **Two-sided Pr <= P** | 0.8947 |

|  |
| --- |
| ***Sample Size = 299*** |

| **Table of anaemia by Death\_E** | | | |
| --- | --- | --- | --- |
| **anaemia** | **Death\_E** | | |
| **Frequency Percent** | **0** | **1** | **Total** |
| **0** | 120 40.13 | 50 16.72 | 170 56.86 |
| **1** | 83 27.76 | 46 15.38 | 129 43.14 |
| **Total** | 203 67.89 | 96 32.11 | 299 100.00 |

|  |
| --- |
| ***Statistics for Table of anaemia by Death\_E*** |

| **Statistic** | **DF** | **Value** | **Prob** |
| --- | --- | --- | --- |
| **Chi-Square** | 1 | 1.3131 | 0.2518 |
| **Likelihood Ratio Chi-Square** | 1 | 1.3086 | 0.2527 |
| **Continuity Adj. Chi-Square** | 1 | 1.0422 | 0.3073 |
| **Mantel-Haenszel Chi-Square** | 1 | 1.3087 | 0.2526 |
| **Phi Coefficient** |  | 0.0663 |  |
| **Contingency Coefficient** |  | 0.0661 |  |
| **Cramer's V** |  | 0.0663 |  |

| **Fisher's Exact Test** | |
| --- | --- |
| **Cell (1,1) Frequency (F)** | 120 |
| **Left-sided Pr <= F** | 0.8980 |
| **Right-sided Pr >= F** | 0.1537 |
|  |  |
| **Table Probability (P)** | 0.0516 |
| **Two-sided Pr <= P** | 0.2628 |

|  |
| --- |
| ***Sample Size = 299*** |

| **Table of Death\_E by diab** | | | |
| --- | --- | --- | --- |
| **Death\_E** | **diab** | | |
| **Frequency Percent** | **0** | **1** | **Total** |
| **0** | 118 39.46 | 85 28.43 | 203 67.89 |
| **1** | 56 18.73 | 40 13.38 | 96 32.11 |
| **Total** | 174 58.19 | 125 41.81 | 299 100.00 |

|  |
| --- |
| ***Statistics for Table of Death\_E by diab*** |

| **Statistic** | **DF** | **Value** | **Prob** |
| --- | --- | --- | --- |
| **Chi-Square** | 1 | 0.0011 | 0.9732 |
| **Likelihood Ratio Chi-Square** | 1 | 0.0011 | 0.9732 |
| **Continuity Adj. Chi-Square** | 1 | 0.0000 | 1.0000 |
| **Mantel-Haenszel Chi-Square** | 1 | 0.0011 | 0.9732 |
| **Phi Coefficient** |  | -0.0019 |  |
| **Contingency Coefficient** |  | 0.0019 |  |
| **Cramer's V** |  | -0.0019 |  |

| **Fisher's Exact Test** | |
| --- | --- |
| **Cell (1,1) Frequency (F)** | 118 |
| **Left-sided Pr <= F** | 0.5376 |
| **Right-sided Pr >= F** | 0.5622 |
|  |  |
| **Table Probability (P)** | 0.0998 |
| **Two-sided Pr <= P** | 1.0000 |

|  |
| --- |
| ***Sample Size = 299*** |

| **Table of hbp by Death\_E** | | | |
| --- | --- | --- | --- |
| **hbp** | **Death\_E** | | |
| **Frequency Percent** | **0** | **1** | **Total** |
| **0** | 137 45.82 | 57 19.06 | 194 64.88 |
| **1** | 66 22.07 | 39 13.04 | 105 35.12 |
| **Total** | 203 67.89 | 96 32.11 | 299 100.00 |

|  |
| --- |
| ***Statistics for Table of hbp by Death\_E*** |

| **Statistic** | **DF** | **Value** | **Prob** |
| --- | --- | --- | --- |
| **Chi-Square** | 1 | 1.8827 | 0.1700 |
| **Likelihood Ratio Chi-Square** | 1 | 1.8630 | 0.1723 |
| **Continuity Adj. Chi-Square** | 1 | 1.5435 | 0.2141 |
| **Mantel-Haenszel Chi-Square** | 1 | 1.8764 | 0.1707 |
| **Phi Coefficient** |  | 0.0794 |  |
| **Contingency Coefficient** |  | 0.0791 |  |
| **Cramer's V** |  | 0.0794 |  |

| **Fisher's Exact Test** | |
| --- | --- |
| **Cell (1,1) Frequency (F)** | 137 |
| **Left-sided Pr <= F** | 0.9328 |
| **Right-sided Pr >= F** | 0.1074 |
|  |  |
| **Table Probability (P)** | 0.0403 |
| **Two-sided Pr <= P** | 0.1948 |

|  |
| --- |
| ***Sample Size = 299*** |

| **Table of sex by Death\_E** | | | |
| --- | --- | --- | --- |
| **sex** | **Death\_E** | | |
| **Frequency Percent** | **0** | **1** | **Total** |
| **0** | 71 23.75 | 34 11.37 | 105 35.12 |
| **1** | 132 44.15 | 62 20.74 | 194 64.88 |
| **Total** | 203 67.89 | 96 32.11 | 299 100.00 |

|  |
| --- |
| ***Statistics for Table of sex by Death\_E*** |

| **Statistic** | **DF** | **Value** | **Prob** |
| --- | --- | --- | --- |
| **Chi-Square** | 1 | 0.0056 | 0.9405 |
| **Likelihood Ratio Chi-Square** | 1 | 0.0056 | 0.9405 |
| **Continuity Adj. Chi-Square** | 1 | 0.0000 | 1.0000 |
| **Mantel-Haenszel Chi-Square** | 1 | 0.0056 | 0.9406 |
| **Phi Coefficient** |  | -0.0043 |  |
| **Contingency Coefficient** |  | 0.0043 |  |
| **Cramer's V** |  | -0.0043 |  |

| **Fisher's Exact Test** | |
| --- | --- |
| **Cell (1,1) Frequency (F)** | 71 |
| **Left-sided Pr <= F** | 0.5201 |
| **Right-sided Pr >= F** | 0.5827 |
|  |  |
| **Table Probability (P)** | 0.1028 |
| **Two-sided Pr <= P** | 1.0000 |

|  |
| --- |
| ***Sample Size = 299*** |

| **Death\_E** | **N Obs** | **Variable** | **N** | **Mean** | **Std Dev** | **t Value** | **Pr > |t|** |
| --- | --- | --- | --- | --- | --- | --- | --- |
| 0 | 203 | creatinine\_p plat serum\_c serum\_s eject\_f time age | 203 203 203 203 203 203 203 | 540.05 266657.49 1.18 137.22 40.27 158.34 58.76 | 753.80 97531.20 0.65 3.98 10.86 67.74 10.64 | 10.21 38.95 25.81 490.85 52.83 33.30 78.70 | <.0001 <.0001 <.0001 <.0001 <.0001 <.0001 <.0001 |
| 1 | 96 | creatinine\_p plat serum\_c serum\_s eject\_f time age | 96 96 96 96 96 96 96 | 670.20 256381.04 1.84 135.38 33.47 70.89 65.22 | 1316.58 98525.68 1.47 5.00 12.53 62.38 13.21 | 4.99 25.50 12.25 265.20 26.18 11.13 48.35 | <.0001 <.0001 <.0001 <.0001 <.0001 <.0001 <.0001 |

| **Model Information** | |
| --- | --- |
| **Data Set** | WORK.HEARTF |
| **Response Variable** | Death\_E |
| **Number of Response Levels** | 2 |
| **Model** | binary logit |
| **Optimization Technique** | Fisher's scoring |

| **Model Fit Statistics** | | |
| --- | --- | --- |
| **Criterion** | **Intercept Only** | **Intercept and Covariates** |
| **AIC** | 377.349 | 235.486 |
| **SC** | 381.049 | 257.689 |
| **-2 Log L** | 375.349 | 223.486 |

| **Testing Global Null Hypothesis: BETA=0** | | | |
| --- | --- | --- | --- |
| **Test** | **Chi-Square** | **DF** | **Pr > ChiSq** |
| **Likelihood Ratio** | 151.8625 | 5 | <.0001 |
| **Score** | 121.7373 | 5 | <.0001 |
| **Wald** | 71.5715 | 5 | <.0001 |

| **Analysis of Maximum Likelihood Estimates** | | | | | |
| --- | --- | --- | --- | --- | --- |
| **Parameter** | **DF** | **Estimate** | **Standard Error** | **Wald Chi-Square** | **Pr > ChiSq** |
| **Intercept** | 1 | -9.4930 | 5.4058 | 3.0839 | 0.0791 |
| **serum\_c** | 1 | -0.6860 | 0.1740 | 15.5352 | <.0001 |
| **serum\_s** | 1 | 0.0646 | 0.0384 | 2.8297 | 0.0925 |
| **eject\_f** | 1 | 0.0734 | 0.0158 | 21.6394 | <.0001 |
| **time** | 1 | 0.0209 | 0.00292 | 51.3474 | <.0001 |
| **age** | 1 | -0.0425 | 0.0150 | 7.9834 | 0.0047 |

| **Odds Ratio Estimates** | | | |
| --- | --- | --- | --- |
| **Effect** | **Point Estimate** | **95% Wald Confidence Limits** | |
| **serum\_c** | 0.504 | 0.358 | 0.708 |
| **serum\_s** | 1.067 | 0.989 | 1.150 |
| **eject\_f** | 1.076 | 1.043 | 1.110 |
| **time** | 1.021 | 1.015 | 1.027 |
| **age** | 0.958 | 0.931 | 0.987 |

| **Model Information** | |
| --- | --- |
| **Data Set** | WORK.HEARTF |
| **Response Variable** | Death\_E |
| **Number of Response Levels** | 2 |
| **Model** | binary logit |
| **Optimization Technique** | Fisher's scoring |

| **Model Fit Statistics** | | |
| --- | --- | --- |
| **Criterion** | **Intercept Only** | **Intercept and Covariates** |
| **AIC** | 377.349 | 235.486 |
| **SC** | 381.049 | 257.689 |
| **-2 Log L** | 375.349 | 223.486 |

| **Testing Global Null Hypothesis: BETA=0** | | | |
| --- | --- | --- | --- |
| **Test** | **Chi-Square** | **DF** | **Pr > ChiSq** |
| **Likelihood Ratio** | 151.8625 | 5 | <.0001 |
| **Score** | 121.7373 | 5 | <.0001 |
| **Wald** | 71.5715 | 5 | <.0001 |

| **Model Fit Statistics** | | |
| --- | --- | --- |
| **Criterion** | **Intercept Only** | **Intercept and Covariates** |
| **AIC** | 377.349 | 236.301 |
| **SC** | 381.049 | 254.804 |
| **-2 Log L** | 375.349 | 226.301 |

| **Testing Global Null Hypothesis: BETA=0** | | | |
| --- | --- | --- | --- |
| **Test** | **Chi-Square** | **DF** | **Pr > ChiSq** |
| **Likelihood Ratio** | 149.0474 | 4 | <.0001 |
| **Score** | 120.3402 | 4 | <.0001 |
| **Wald** | 72.3323 | 4 | <.0001 |

| **Analysis of Maximum Likelihood Estimates** | | | | | |
| --- | --- | --- | --- | --- | --- |
| **Parameter** | **DF** | **Estimate** | **Standard Error** | **Wald Chi-Square** | **Pr > ChiSq** |
| **Intercept** | 1 | -0.6045 | 1.0362 | 0.3403 | 0.5596 |
| **serum\_c** | 1 | -0.7198 | 0.1746 | 16.9936 | <.0001 |
| **eject\_f** | 1 | 0.0748 | 0.0156 | 23.1232 | <.0001 |
| **time** | 1 | 0.0206 | 0.00288 | 51.1545 | <.0001 |
| **age** | 1 | -0.0433 | 0.0149 | 8.4862 | 0.0036 |

| **Odds Ratio Estimates** | | | |
| --- | --- | --- | --- |
| **Effect** | **Point Estimate** | **95% Wald Confidence Limits** | |
| **serum\_c** | 0.487 | 0.346 | 0.686 |
| **eject\_f** | 1.078 | 1.045 | 1.111 |
| **time** | 1.021 | 1.015 | 1.027 |
| **age** | 0.958 | 0.930 | 0.986 |

| **Model Information** | |
| --- | --- |
| **Data Set** | WORK.HEARTF |
| **Distribution** | Binomial |
| **Link Function** | Logit |
| **Dependent Variable** | Death\_E |

| **Criteria For Assessing Goodness Of Fit** | | | |
| --- | --- | --- | --- |
| **Criterion** | **DF** | **Value** | **Value/DF** |
| **Log Likelihood** |  | -113.1507 |  |
| **Full Log Likelihood** |  | -113.1507 |  |
| **AIC (smaller is better)** |  | 236.3014 |  |
| **AICC (smaller is better)** |  | 236.5062 |  |
| **BIC (smaller is better)** |  | 254.8036 |  |

| **Analysis Of Maximum Likelihood Parameter Estimates** | | | | | | | |
| --- | --- | --- | --- | --- | --- | --- | --- |
| **Parameter** | **DF** | **Estimate** | **Standard Error** | **Wald 95% Confidence Limits** | | **Wald Chi-Square** | **Pr > ChiSq** |
| **Intercept** | 1 | -0.6045 | 1.0362 | -2.6353 | 1.4264 | 0.34 | 0.5596 |
| **serum\_c** | 1 | -0.7198 | 0.1746 | -1.0620 | -0.3776 | 16.99 | <.0001 |
| **eject\_f** | 1 | 0.0748 | 0.0156 | 0.0443 | 0.1053 | 23.12 | <.0001 |
| **time** | 1 | 0.0206 | 0.0029 | 0.0150 | 0.0263 | 51.15 | <.0001 |
| **age** | 1 | -0.0433 | 0.0149 | -0.0725 | -0.0142 | 8.49 | 0.0036 |
| **Scale** | 0 | 1.0000 | 0.0000 | 1.0000 | 1.0000 |  |  |

|  |  |
| --- | --- |
| **Note:** | The scale parameter was held fixed. |

| **LR Statistics For Type 1 Analysis** | | | | |
| --- | --- | --- | --- | --- |
| **Source** | **Deviance** | **DF** | **Chi-Square** | **Pr > ChiSq** |
| **Intercept** | 375.3488 |  |  |  |
| **serum\_c** | 347.2521 | 1 | 28.10 | <.0001 |
| **eject\_f** | 324.3211 | 1 | 22.93 | <.0001 |
| **time** | 235.4141 | 1 | 88.91 | <.0001 |
| **age** | 226.3014 | 1 | 9.11 | 0.0025 |

| **LR Statistics For Type 3 Analysis** | | | |
| --- | --- | --- | --- |
| **Source** | **DF** | **Chi-Square** | **Pr > ChiSq** |
| **serum\_c** | 1 | 18.21 | <.0001 |
| **eject\_f** | 1 | 28.32 | <.0001 |
| **time** | 1 | 78.98 | <.0001 |
| **age** | 1 | 9.11 | 0.0025 |

| **Model Information** | |
| --- | --- |
| **Data Set** | WORK.HEARTF |
| **Response Variable** | Death\_E |
| **Number of Response Levels** | 2 |
| **Model** | binary logit |
| **Optimization Technique** | Fisher's scoring |

| **Model Fit Statistics** | | |
| --- | --- | --- |
| **Criterion** | **Intercept Only** | **Intercept and Covariates** |
| **AIC** | 377.349 | 236.301 |
| **SC** | 381.049 | 254.804 |
| **-2 Log L** | 375.349 | 226.301 |

| **Testing Global Null Hypothesis: BETA=0** | | | |
| --- | --- | --- | --- |
| **Test** | **Chi-Square** | **DF** | **Pr > ChiSq** |
| **Likelihood Ratio** | 149.0474 | 4 | <.0001 |
| **Score** | 120.3402 | 4 | <.0001 |
| **Wald** | 72.3323 | 4 | <.0001 |

| **Analysis of Maximum Likelihood Estimates** | | | | | |
| --- | --- | --- | --- | --- | --- |
| **Parameter** | **DF** | **Estimate** | **Standard Error** | **Wald Chi-Square** | **Pr > ChiSq** |
| **Intercept** | 1 | 0.6045 | 1.0362 | 0.3403 | 0.5596 |
| **serum\_c** | 1 | 0.7198 | 0.1746 | 16.9936 | <.0001 |
| **eject\_f** | 1 | -0.0748 | 0.0156 | 23.1232 | <.0001 |
| **time** | 1 | -0.0206 | 0.00288 | 51.1545 | <.0001 |
| **age** | 1 | 0.0433 | 0.0149 | 8.4862 | 0.0036 |

| **Odds Ratio Estimates** | | | |
| --- | --- | --- | --- |
| **Effect** | **Point Estimate** | **95% Wald Confidence Limits** | |
| **serum\_c** | 2.054 | 1.459 | 2.892 |
| **eject\_f** | 0.928 | 0.900 | 0.957 |
| **time** | 0.980 | 0.974 | 0.985 |