Report for the Statistical Analysis of the Dobutamine Drug Test

Eltone Mabodo

2022-10-06

# 1. Data Exploration

## 1.1 Introduction and Background

The risk of experiencing a cardiac event is typically tested by “stress echocardiography”. In this test, the patient’s heart rate is raised through exercise whilst at the same time the various measurements of the heart are taken. However, this test may be problematic if the person under the study is not willing or unable to manage the stress induced by the hard exercise. This typically happens to the elderly people.

In this study, the researcher used data from a study into a drug called “dobutamine”. This drug is used as an alternative means of putting the heart under stress, rather than putting the people to a physical exercise. The research questions to be answered are listed in the following section.

## 1.2 Research Questions

This study aimes at answering the following research questions:

* Is stress echocardiography test still effective at predicting a cardiac event when the stress on the heart is induced by dobutamine rather than through physical exercises?
* What are the factors that may be useful in predicting a cardiac event?

## 1.3 Summary Statistics of the data

This section will provide brief descriptive statistics of the data from the drug test mentioned above. These include the mean, standard deviation and the median (for continous variables). For categorical variables (or factors), only the frequencies or absolute figures will be presented.

### 1.3.1 Descriptive Statistics for the continous variables.

Basically, the data analysed had **558** individuals who were included in the sample. Of these individuals, **89** had experienced the cardic event, whilst **469** did not experience the cardiac event. The descriptive statistics for the continous variables used in this study are given by the table below.

Table 1: Descritpive Statistics for the Continous Variables

| Variables | Sample | Mean | SD | Median | Minimum | Maximum | Skew | Kurtosis | SE |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Age | 558 | 67.34 | 12.05 | 69 | 26 | 93 | -0.58 | 0.41 | 0.51 |
| Base EF | 558 | 55.60 | 10.32 | 57 | 20 | 83 | -1.28 | 1.74 | 0.44 |
| Dob EF | 558 | 65.24 | 11.76 | 67 | 23 | 94 | -1.18 | 1.46 | 0.50 |
| Source: Author's Calculations |

From Table 1 above, the average age in the sample was **67.34** years, whilst the baseline cardiac ejection fraction average was **55.6%**. On the other hand, the average ejection fraction with patient on the dobutamine drug was **65.24%**. A normal ejection fraction should be in the range of 50% to 70%, which implies that on average, the sample which was selected had a normal ejection fraction both at baseline and after the patients were given the dobutamine drug.

More so the summary statistics table above shows that there were no missing values for all the continous variables it shows that the sample size used to calculate the descriptive statistics for all the continous variables was **558**, which is equivalent to **558**, the total number of the individuals whose heart tests were taken.

The table also shows the median for the variables, which were **69**, **57**, and **67** for the age, baseline cardiac ejection fraction and ejection fraction with patient on dobutamine, respectively. Another summary statistic which was calculated for all the continous variables in **Table 1** above was their skewness. It calculates how far a particular random variable’s distribution deviates from a symmetric distribution like the normal distribution. All the continous variables were negatively skewed as shown by their coefficient of skewenes in **Table 1** above. The skewness coefficient for age was **-0.58** whilst that of the baseline cardiac ejection fraction was **-1.28**. On the other hand, ejection fraction for patients on dobutamine drug was **-1.18**. What these skenesses signify is that the median values for all the continuous variables were greater than the mean values.

More so, the table shows the standard deviations of all the variables. The standard deviations measures the extent to which the values of the specific variable deviates from the mean of that variable. The less this variability the better. From **Table 1** above the standard deviation for the age was **12.05** whilst that of baseline EF was **10.32**. Dobutamine induced EF had a standard deviation of **10.32**. These statistics indicates that there was less variability in baseline EF compared to dobutamine-induced EF.

The population mean’s likelihood to differ from a sample mean is indicated by the standard error of the mean, or simply standard error. It reveals how much the sample mean would change if a study were to be repeated with fresh samples drawn from a single population. Hence if a fresh sample was to be used in this study, the average age is likely to change by **0.51** years, whilst the baseline EF is likely to change by **0.44%**. The dobutamine-induced EF will likely change by **0.5%**. Given these figures, one can safely say that the means for these variables will not change much if a new sample was selected.

### 1.3.2 Descriptives for factors

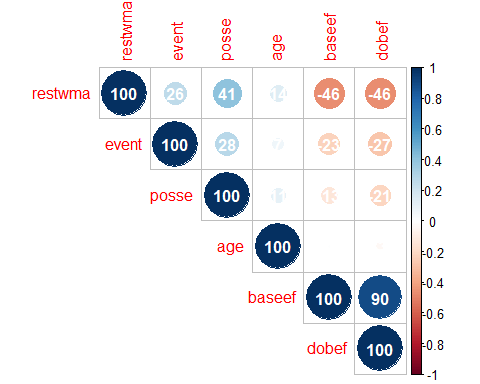
From the dataset, the number of people who experienced the cardiac event were **89**, whilst those who did not experience the cardiac event were **469**. Those who recorded wall motion anomally on ecgocardiogram were **301** whilst those who did not were **257**. More so, those who have shown positive stress echocardiography test were **136** whilst those who did not were **136**.

## 1.4 Relationships amoung Variables

This section will present (i) the relationship between the outcome variable (event) and the other variables and (ii) the relationship between dobef and baseef.

### 1.4.1 The Relationship between the outcome variable and the other variables

The graph below shows the relationship amoung all the variables in the dataset, including the the relationship between the dependent variaable and the other variables.



From the graph above, there is a positive relationship between the cardiac event happening and a positive stress echocardiography test. This is shown by the correlation coefficient of **0.28**. More so, the graph shows that there is close to no relationship between the event and the age of the patient, correlation coefficient of **0.07**. The graph also shows that there is a negative relationship between the event and baseline cardiac ejection fraction and dobutamine-induced ejection fraction. The correlation coefficient of the event and baseline ejection fraction is **-0.23** and that of the event and dobutamine-induced ejection fraction is **-0.27**. There is also a positive relatioship between the event and wall motion anomaly as indicated by the correlation coefficient of **0.26**

### 1.4.2 The relationship between dobrf and baseef

The relationship between dobef and baseef in the data is **0.9**. This is the correlation coefficient between these two variables and shows that there is a very strong and positive relationship between these two variables. In regression analysis, one of these variables must be dropped in order to avoid the problem of multicolinearity.

# 2. Logistic Regression Analysis

To estimate the regression model, we start by the basic or naive equation where the cardiac event is being regressed on positive stress echocardiography test variable. After this equation, we will add the other two risk factors, which are wall motion anomaly on echocardiogram and the ejection fraction with patient on dobutamine. If we are treating the age of the patient and the baseline ejection fraction as confounders, it means that these variables affect both the probability of the cardiac event happening which is the dependent variable and the main explanatory variable, which is positive stress echocardiography test. If these variables are true confounders of this relationship, then adding these variables or dropping them from the final model will fundamentally change the relationship between the dependent and the main explanatory variables. The table below represents the results of the model building process.

Table 1: Logistic Regression Modelstrue1false

Dependent Variable: Cardiac Event Happening

|  | Model 1 | Model 2 | Model 3 | Model 4 | Model 5 |
| --- | --- | --- | --- | --- | --- |
| (Intercept) | -2.176 | -0.350 | -0.419 | -0.823 | -0.851 |
|  | (0.000) | (0.646) | (0.586) | (0.432) | (0.417) |
| posseYes | 1.505 | 1.043 | 1.003 | 1.033 | 0.997 |
|  | (0.000) | (0.000) | (0.000) | (0.000) | (0.000) |
| restwmaYes |  | 0.843 | 0.876 | 0.825 | 0.857 |
|  |  | (0.015) | (0.012) | (0.017) | (0.015) |
| dobef |  | -0.036 | -0.050 | -0.036 | -0.049 |
|  |  | (0.000) | (0.034) | (0.000) | (0.038) |
| baseef |  |  | 0.018 |  | 0.016 |
|  |  |  | (0.500) |  | (0.534) |
| age |  |  |  | 0.007 | 0.007 |
|  |  |  |  | (0.509) | (0.543) |
| Num.Obs. | 558 | 558 | 558 | 558 | 558 |
| AIC | 455.9 | 429.7 | 431.2 | 431.3 | 432.9 |
| BIC | 464.6 | 447.0 | 452.9 | 452.9 | 458.8 |
| Log.Lik. | -225.955 | -210.849 | -210.620 | -210.628 | -210.433 |
| F | 38.568 | 19.290 | 14.510 | 14.505 | 11.634 |
| RMSE | 0.35 | 0.34 | 0.34 | 0.34 | 0.34 |
| *1*Figures in parenthesis are p-values |
| Source: Author's Calculations |

## 2.1 Explanation of the model building process.

In this study, the researcher used a forward stepwise regression to select the model of best fit. In this process, the basic and naive logistic regression was fitted with only one dependent variable (posse) and event as an outcome variable. However, because there are also other variables which are likely to influence the cardic event happening other than only through using the results from the echocardiography test. This basic model will suffer from model misspecification which ultimately affect the interpretation and the coefficients. Hence, in **Model 2**, other explanatory variables such as wall motion anomally and the ejection fraction with the patient on dobutamine were included. In addition, since we are taking the age of the patient and the baseline ejection fraction as **counfounders**, these variables were included as explanatory variables, one-by-one inititially and finally both at once, and the results are shown from **Model 3 to Model 5**. The reason was to see how they affect the coefficient of posse in **Model 2**. If these variables are confounders, then the addition or ommission of these variables should greatly change the coefficient of the main variable in the model, which is positive echocardiography test. Hence from **Table 2** above, adding only age to **Model 2** has reduced the coefficient of posse from 1.043 to 1.033, which is not a big difference. On the other hand, adding only baseline ejection fraction has led to the coefficient of posse from 1.043 to 1.003 as shown by **Model 3**. Adding both confounders to **Model 2** at once shows that gives us **Model 5** and shows that the coefficient of posse decrease from 1.043 to 0.997. Hence, the conclusion at this stage of the model selection may be that both age and baseline ejection fraction are confounders of posse and should be included in the finale model. However, due to the fact that there is a strong correlation between baseef and dobef, the final model should include only one of these variable. This led to the variable baseef being dropped from the model and the resultant model was **Model 4**. However, in **Model 4**, age is not statistically significant and the addition of this variable to **Model 2** did not change the coefficient of posse much. Hence we can safely drop it from the model. Hence the final selected model is **Model 2**

More so, we can use the Anova test to see which model performs better among the models which were estimated. Thelower the p-value, the better the model. This test reveal that **Model 2** is the most appropriate model as shown by the table below. Using the Akaike Information Criterion (AIC) method of model evaluation, **Model 2** was also chosen since it is the one with the least AIC value.

Table 1: Model Selection Test

Test carried out using the Chi-Square Test

| Models | Resid. Df | Resid. Dev | Df | Deviance | Pr(>Chi) |
| --- | --- | --- | --- | --- | --- |
| Model 1 | 556 | 451.91 | NA | NA | NA |
| Model 2 | 554 | 421.70 | 2 | 30.21 | 0.00 |
| Model 3 | 553 | 421.24 | 1 | 0.46 | 0.50 |
| Model 4 | 553 | 421.26 | 0 | -0.02 | NA |
| Model 5 | 552 | 420.87 | 1 | 0.39 | 0.53 |
| Source: Author's Calculations |

## 2.2 Intepretation of the results in **Model 2**

The odds of experiencing the cardiac event for people who have recorded a positive stress echocardiography test are **2.84** higher compared to that of people who did not. This is statistically significant at 1% level of significance as shown by the p-value in the table above. More so, compared to the patients without a wall motion anomaly, the odds of experiencing a cardiac event for the people with wall motion anomaly are **2.32** higher. This is true at 5% level of significance. In addition, as the ejection fraction for people on dobutamine increases, the odds of the patient experiencing the cardiac event decreases by **0.96**. This assertion is true at 1% level of significance.

# 3. Conclusion

In conclusion, the study has shown that stress echocardiography is still effective at predicting a cardiac event when the stress on the heart is induced by dobutamine rather than through exercising. More so the studyb revealed that the other variables which can predict a cardiac event are wall motion anomally and ejection fraction with patient on dobutamine.