

## Biostatistics, Assignment 1

---

Petersen Trentin (PTRTRE004)

April 9, 2024

# Introduction

Cardiovascular health is always a great concern and with the numerous different ways there are to put your cardiovascular health at risk, it necessitates the need for methods of diagnosing cardiovascular related diseases and defects that might put one at risk of experiencing a cardiac event. Stress echocardiography is a method of determining how well the heart is functions under stress.

Traditionally to put the heart under stress the patient does some form of exercise to raise their heart rate, however this could pose risks to older patients whose hearts cannot endure the stress induced by physical exertion. An alternative method of inducing stress on the heart is through the use of the drug, Dobutamine. The aim of the research is to evaluate the predictive performance of a stress echocardiography test at predicting a cardiac event if Dobutamine was used to induce stress on the heart.

The data was derived from a study that compared the predictive performance of a stress echocardiography test at predicting a cardiac event with different methods (Dobutamine vs exercise) of inducing stress on the heart. This subset only includes data about the cases where dobutamine was used to induce stress on the heart. There are 558 observations in the data, of which 350 will be used as training data and the remaning 208 observations will be used as validation data. The data included the following variables:

- Cardiac event experienced in the following 12 months (0=no, 1=yes): *event* (response)
- Patient's age in years: *age*
- Baseline Cardiac Ejection Fraction: *baseef*
- Cardiac Ejection Fraction on Dobutamine: *dobef*
- Was the Stress Echocardiography test positive (0=yes, 1=no): *posse*
- Was there a Wall Motion Anomaly detected on the echocardiogram (0=yes, 1=no): *restwma*

For ease of interpretation, the binary representations for the *posse* and *restwma* variables were changed from (0=yes, 1=no) to (1=yes, 0=no) such that they use the same encoding as the response variable, *event*.

## Statistical Methods

Data Exploration: In the data exploration both univariate analyses and bivariate analyses were conducted. The univariate analyses was conducted to understand each variables distribution and basic statistics such as the mean, median and standard deviation. The bivariate analyses focus on the relationships between pairs of independent variables. Age as a confounder was also explored.

Logistic Regression: The dependent variable *event* is a binary variable and therefore we use logistic regression to map a linear combination of predictor variables to a probability of *event* occurring which is then encoded into a 1 or a 0 depending on the decision rule/classification scheme.

Model Building: A backwards approach was taken to build the logistic regression model to find the optimal model for the training data.

Model Diagnostics: A thorough diagnostic analysis of the final model was conducted to ensure goodness of fit and that there aren't any highly influential points distorting the values of the regression coefficient estimates. This step is crucial for ensuring that the model is reliable and robust.

Model Interpretation: A summary of the final model and its regression coefficient estimates were given as well as interpretations for each regression coefficient estimate. The interpretation allows for understanding of the underlying relationships in the data which can be used to inform clinical decisions.

Classification Scheme Derivation: The final logistic regression model was used to determine a classification scheme that aligned with the context of the problem at hand. The final results for the Sensitivity, Specificity and Likelihood ratios are also provided as a measure of the final model + classification schemes performance.

# Data Exploration

## Univariate Analyses

To begin the data exploration, a table of descriptive statistics for the continuous variables: *age*, *baseef* and *dobef* was created. These statistics are displayed in Table 1. below.

Table 1: Summary statistics for *age*, *baseef* and *dobef* variables

	Min	1st Qu.	Median	Mean	3rd Qu.	Max	Standard Deviation	Range
<i>age</i>	26	59.25	68	66.994	75	93	12.460	67
<i>baseef</i>	20	53.00	57	55.897	62	79	10.254	59
<i>dobef</i>	23	63.00	68	65.651	74	94	11.631	71

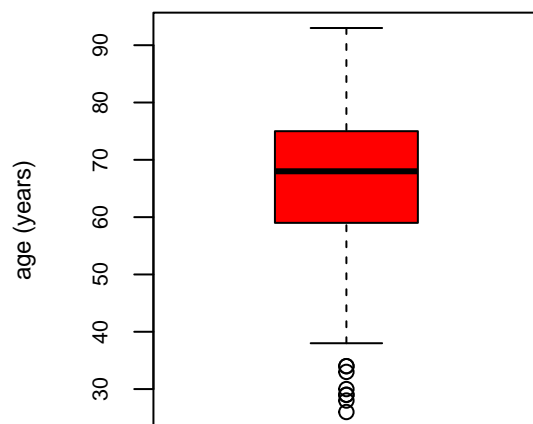


Figure 1: Age Distribution

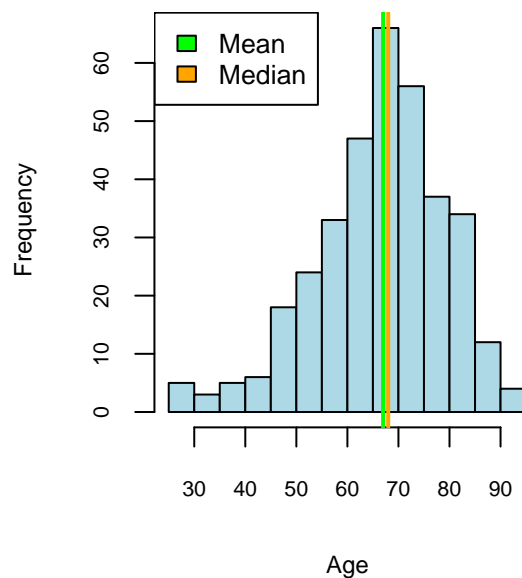


Figure 2: Distribution of Ages

*age*:

In the training data the minimum and maximum *age* over all cases was 26 and 93 respectively. The interquartile range was 15.75, which tells us that 50% of the cases in the training data are spread over 15.75 years of age, from 59.25 years of age (1st Qu.) to 75 years of age (3rd Qu.). Although the range is 67 which might indicate there is a good spread of ages, the values of the first and third quartiles show us that majority of cases in the study are observations of elder people. The standard deviation of ages is 12.46 years, therefore more than 50% of the data is contained within one standard deviation from both the mean and the meadian. There are also some outliers as indicated by Figure 1.

The mean *age* of 67 is fairly close to the median *age* of 68 which suggests that the distribution of the *age* variable is fairly symmetric around the center. If we look at the histogram of ages depicted in Figure 2. we can see that the ages are approximately normally distributed but slightly skewed to the left.

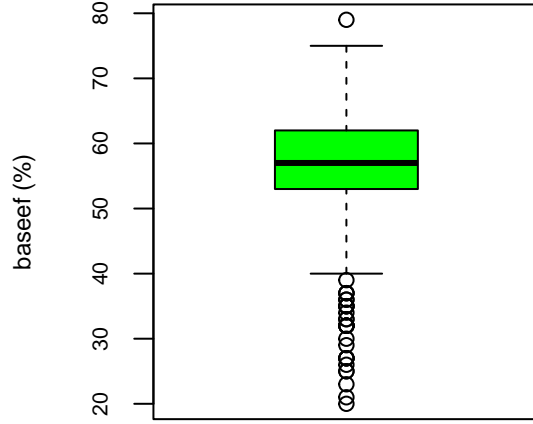


Figure 3: Baseline EF

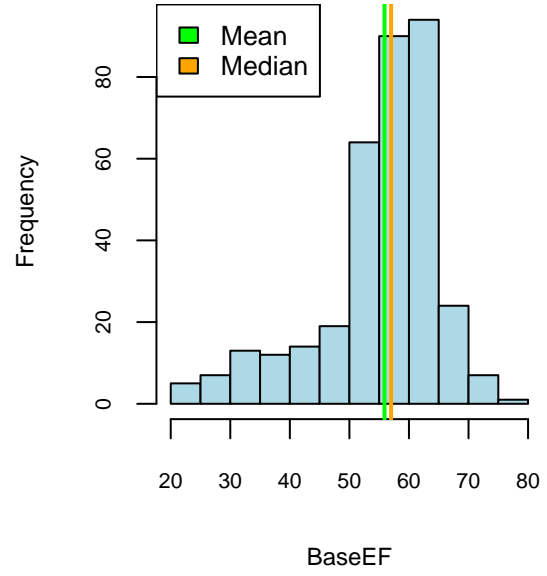


Figure 4: Distribution of baseef Values

*baseef*:

The minimum and maximum *baseef* measurements over all the cases in the training data were 20% and 79% respectively. 25% of the cases in the training data had a *baseef* value of 53% or lower as indicated by value of the 1st Quartile. 75% of the cases had a *baseef* value of 62% or lower as indicated by the value of the 3rd Quartile. The standard deviation of the *baseef* variable is 10.254%.

The mean *baseef* was 55.9% which close to the median value of 57% and thus we can see in the histogram below that the distribution of *baseef* values is very roughly symmetrical. The mean value of 55.9% is on the lower end of normal heart function. With the a normal heart function being defined by a ejection fraction of 55% to 70%. There are a number of reasons, most of which are common in older adults, which can lead to a decrease in *baseef*. Therefore we might be observing this fairly low mean *baseef* value due to the distribution of ages in the data.

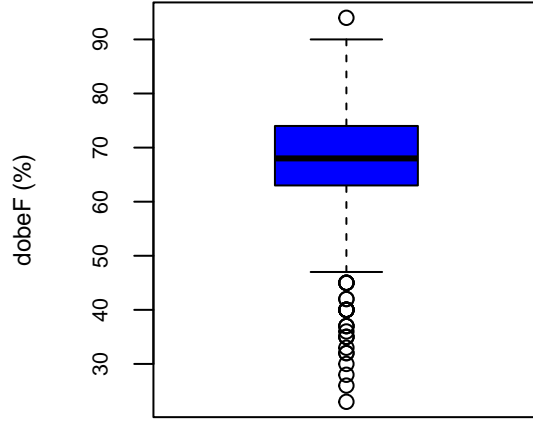


Figure 5: EF on Dobutamine

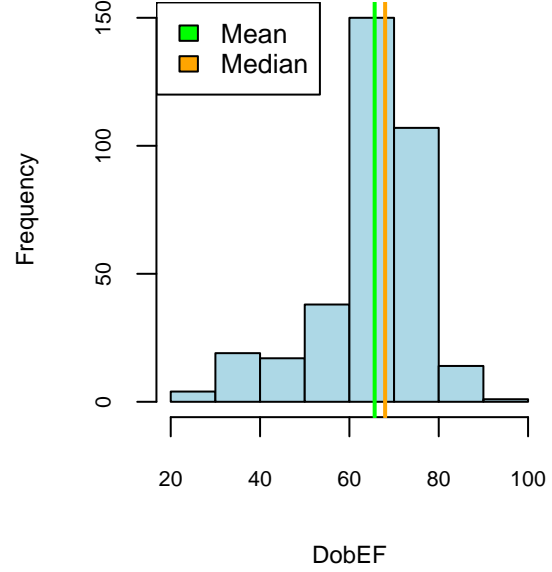


Figure 6: Distribution of dobef Values

*dobef*:

The minimum and maximum *dobef* measurements over all the cases in the training data were 23% and 94% respectively. The 1st Quartile has a value of 63%, this means that a quarter of all the cases had a *dobef* value of 63% or below. Similarly the 3rd Quartile has a value of 74% which indicates that 75% of the cases have a *dobef* value of 74% or below.

The mean *dobef* value is 65.65% which is close to the median value of *dobef* which was 68%. The standard deviation of the *dobef* variable is 11.631%.

Figures 3&5. as well as Figures 4&6. show that the distributions of the *baseef* and *dobef* variables are very similar, however, the values of *dobef* are higher than the *baseef* values on average. Figures 3&5. also display a large number of outliers for both the *baseef* and *dobef*, the effects of these outliers will be determined in the model diagnostics section.

Table 2: Summary statistics for Posse, Restwma and Event

	Frequency - Yes	Frequency - No	(%) - Yes	(%) - No	Mode
Stress Echocardiography was positive (posse)	80	270	22.9	77.1	No (0)
Wall motion anamoly on echocardiogram (restwma)	191	159	54.6	45.4	Yes (1)
Cardiac event was experienced (event)	52	298	14.9	85.1	No (0)

*posse:*

In 22.9% of the cases the Stress Echocardiography test was positive i.e.  $posse = 1$  and in the other 77.1% of the cases the Stress Echocardiography test was negative i.e.  $posse = 0$ . The modal class was “No”-(0).

*restwma:*

In 54.6% of the cases, a wall motion anomaly was detected on the echocardiogram i.e.  $restwma = 1$  and in the other 45.4% of the cases, no wall motion anomaly was detected on the echocardiogram i.e.  $restwma = 0$ . The modal class was “Yes”-(1).

*event:*

In 14.9% of the cases, the patient experienced a cardiac event in the following 12 months i.e.  $event = 1$  and in the other 85.1% of the cases, no cardiac event was experienced by the patient in the following 12 months. The modal class was “No”-(0).

These statistics are shown graphically in Figures 7-9. below.

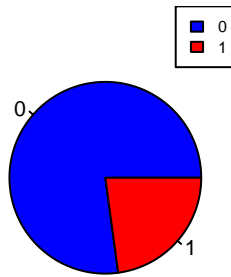


Figure 7: posse Distribution

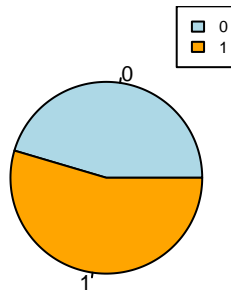


Figure 8: restwma Distribution

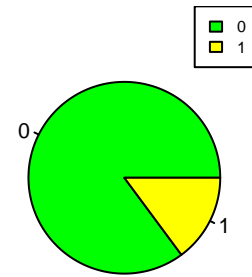


Figure 9: event Distribution

## Bivariate Analyses

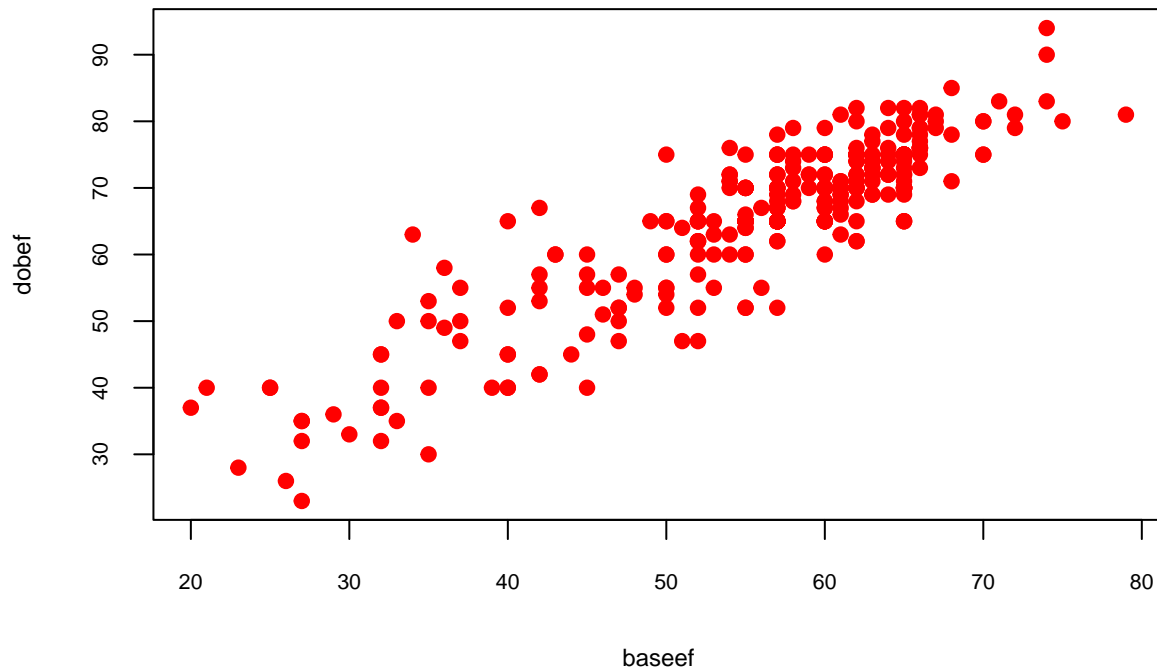


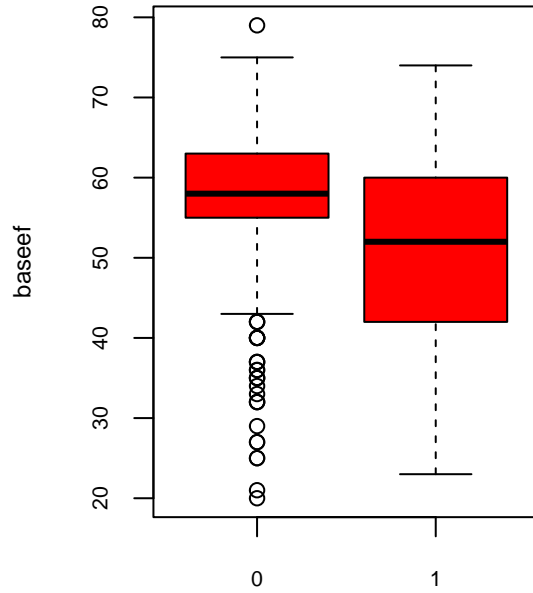
Figure 10: dobeef vs baseef

### Relationship between *baseef* and *dobeef*:

Figure 10 shows a strong positive linear relationship,  $\rho = 0.892$ , between the *baseef* and *dobeef* variables. As *baseef* increases so does *dobeef*. This means that if both of these variables were to be included in the logistic regression model that there would be multicollinearity. Multicollinearity poses problems in model fitting as it can lead to misrepresentation of the regression coefficient estimates, making them unreliable and difficult to interpret.

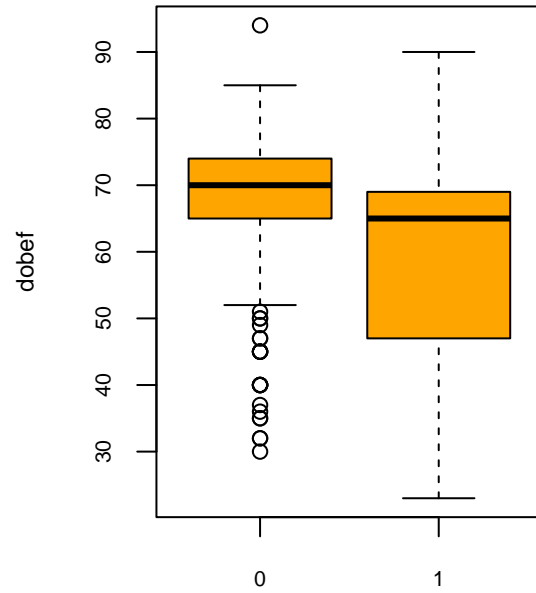
When multicollinearity is present it is common practice to either simply remove one of the correlated predictors from the model if it is sensible to do so in the context of the problem, or one could use regularization (either L1 or L2 or a combination of the two, an elastic-net) to mitigate the effect of multicollinearity.

The clinical interpretation of this relationship would be that a patient's baseline cardiac ejection fraction is predictive of how their ejection fraction will change due to the use of dobutamine.



Event (Yes=1) (No=0)

Figure 11: baseef vs event



Event (Yes=1) (No=0)

Figure 12: dobef vs event

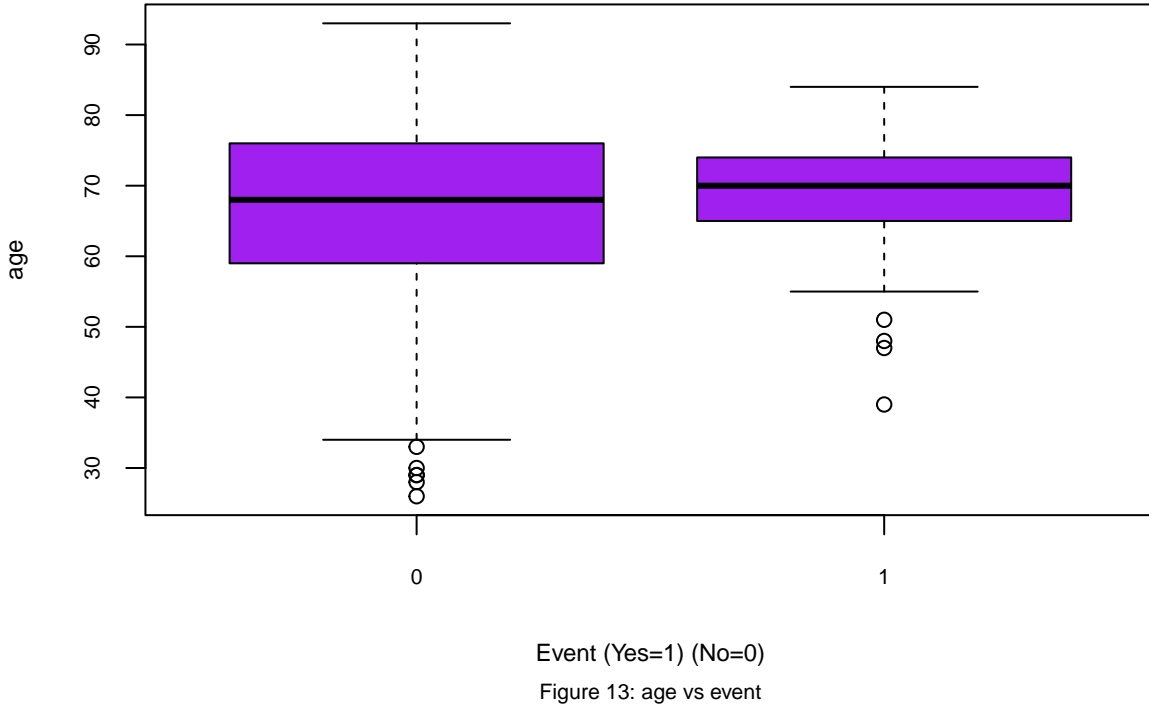
### Relationship between *baseef*, *dobef* and the *event* outcome:

Figure 11 shows the distribution of *baseef* for patients, grouped by whether or not they experienced a cardiac event within the following 12 months. Patients who didn't experience a cardiac event ( $event = 0$ ) have a smaller spread of *baseef* values around the median, meaning less variability of *baseef* values. Patients where  $event = 1$  had a distribution with higher interquartile range (IQR) and thus variability of *baseef* values. The median value of *baseef* was higher amongst the group of patients where  $event = 0$  than the group of patients where  $event = 1$ . The depicted distribution suggest that there is a possible link between lower *baseef* values and a higher risk of experiencing a cardiac event in the following 12 months. The opposite applies for higher *baseef* values.

Figure 12 shows the distribution of *dobef* for patients, grouped by whether or not they experienced a cardiac event within the following 12 months. The distributions look very similar to those in Figure 11, which is not unexpected due to the relationship between *dobef* and *baseef* explored previously. Therefore we can make the same observations, there is a possibility that lower *dobef* values are linked with a higher risk of experiencing a cardiac event in the following 12 months and vice versa. This aligns with the clinical aspect as a lower ejection fraction indicates that the heart is not functioning as well as it should therefore compromising the health of the heart which could lead to a cardiac event.

These relationships will be further explored in the model fitting section where logistic regression will be used to determine whether the effects of the *baseef* and *dobef* variables on the response is statistically significant.





#### Relationship between *age* and the *event* outcome:

Figure 13 shows the distribution of *age* for patients, grouped by whether or not they experienced a cardiac event within the following 12 months. The boxplot shows that the median *age* value in both the group of patients where *event* = 1 and where *event* = 0 is nearly identical. The IQR, representing the middle 50% of the data, is larger for the group of patients where *event* = 0, suggesting greater variability of *age* in this group.

These distributions suggest that there is a probability that the outcome of *event* is not strongly associated with *age*, however, this may be linked to the distribution of ages in the cases as 50% of patients were in the age range 59 to 75. This concentration of older patients could make the relationship between the age of the patient and the occurrence of a cardiac event difficult to interpret/observe, as there is less variability in age among the younger patients (because there are fewer of them). The effect of age on the occurrence of a cardiac event will be further investigated in the model fitting section.

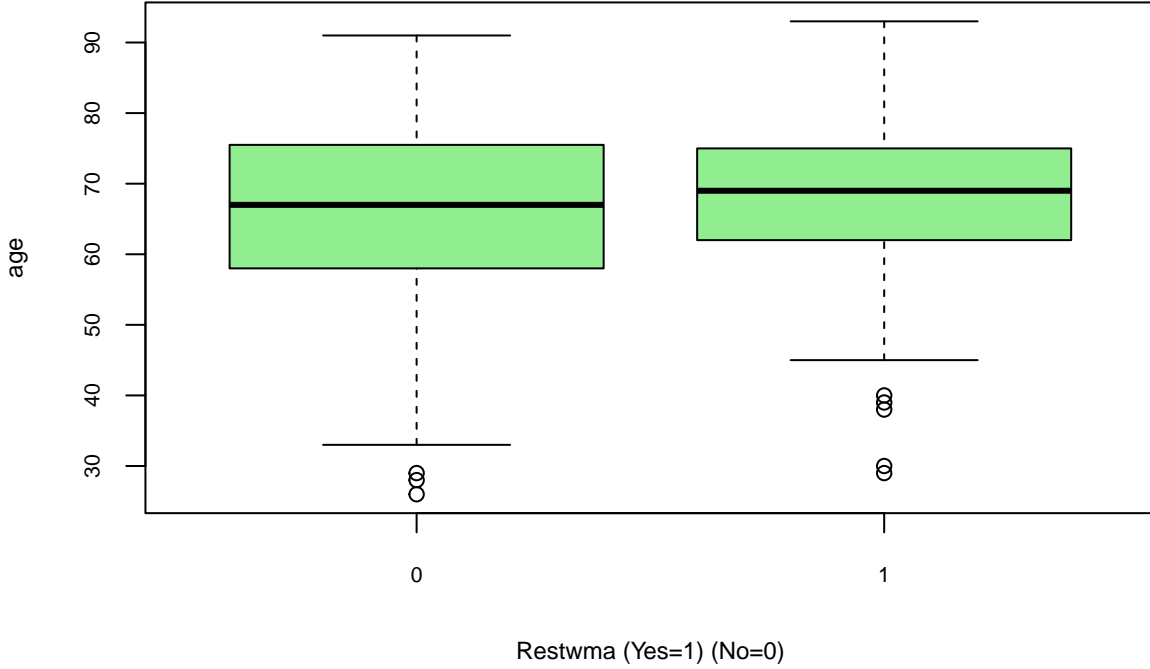


Figure 14: age vs restwma

#### **age as a confounder:**

For age to be a confounder it has to be associated with the outcome (*event*) as well as the exposure (*baseef*, *dobef*, *posse*, *restwma*), but it should not be a consequence of the exposure.

Clinically it makes sense that age would be associated with the probability of having a cardiac event. However Figure 13 suggests that there may be very little/no association between age and the event. The same reasoning can be used as mentioned before, the skewed distribution of *age* is distorting the true association between *age* and *event*. Therefore, we will conclude that *age* is associated with *event* based on clinical evidence.

Then we examine the association between age and if a wall motion anomaly was detected on the echocardiogram as clinically it is sound to say that wall motion anomalies are more common in older individuals. However, the distribution shown in Figure 14 looks much like the distribution in Figure 13 which we will conclude happens for the same reason as with the association between *age* and *event*.

Finally, *age* is not a consequence of having a certain value of *baseef* or *dobef*, therefore it is not a consequence of the exposure.

*age* will be treated as a confounder and any possible interactions it has with other variables will be explored.

### Relationship between *event* and *posse*:

The relationship between *posse* and *event* was examined by creating a contingency table and calculating the odds and risk ratios.

Table 3: Contingency table for *posse* and *event*

	event=1	event=0
posse=1	22	58
posse=0	30	240

Risk Ratio = 2.48 with a 95% CI (1.52;4.04). In the group of patients where the stress echocardiography test was positive, *posse* = 1, the risk of experiencing a cardiac event in the following 12 months was 2.48 times higher than in the group of patients where the stress echocardiography test was negative, *posse* = 0.

Odds Ratio = 3.03 with a 95% CI (1.63;5.64). The effect of the stress echocardiography test being positive on the odds of experiencing a cardiac event in the following 12 months is 3.03. In other words, those whose stress echocardiography tests are positive have an odds of having a cardiac event in the following 12 months that is 3.03 times the odds of those whose tests were negative.

These ratios (and the fact that their CI's don't include 1) indicate a statistically significant relationship between the predictor *posse* and the response *event*

### Relationship between *event* and *restwma*:

Similarly to *posse*, the relationship between *restwma* and *event* was examined by creating a contingency table and calculating the odds and risk ratios.

Table 4: Contingency table for *restwma* and *event*

	event=1	event=0
restwma=1	40	151
restwma=0	12	147

Risk Ratio = 2.77 with a 95% CI (1.51;5.11). In the group of patients where a wall motion anomaly was detected, *restwma* = 1, the risk of experiencing a cardiac event in the following 12 months was 2.77 times higher than in the group of patients where no wall motion anomaly was detected, *restwma* = 0.

Odds Ratio = 3.25 with a 95% CI (1.64;6.43). The effect of having detected a wall motion anomaly, *restwma* = 1, on the odds of experiencing a cardiac event in the following 12 months is 3.25. In other words, patients where a wall motion anomaly was detected in their echocardiogram have an odds of having a cardiac event in the following 12 months that is 3.25 times the odds of patients where no wall motion anomaly was detected in their echocardiogram.

These ratios (and the fact that their CI's don't include 1) indicate a statistically significant relationship between the predictor *restwma* and the response *event*.

# Logistic Regression

## Model Building:

A backwards approach was selected to build the logistic regression model.

Step 1 is to identify which predictors should be included in the saturated model from which we will then remove insignificant predictors. To do so, univariate models are fitted for each predictor. Their significances can be summarised by Table 5 below:

Table 5: Predictors and their significance when fitted in a univariate model

Predictor	P-Value
<i>age</i>	0.3574045
<i>baseef</i>	0.0000774
<i>dobef</i>	0.0000042
<i>posse</i>	0.0004534
<i>restwma</i>	0.0007423

Therefore for the initial saturated model, Model A, we will include all predictors but *age* i.e. event ~ baseef+dobef+posse+restwma.

Thereafter, the process of selecting the best model (steps 2-5 of backwards approach) can then be summarised by Table 6 below:

Table 6: Backwards model building approach steps 2-5 summarised

Step	Model	Variable/s Removed	Variable/s Added	Current Variable/s in Model	LRT P-value	LRT vs.	AIC
1	A	-	-	baseef+dobef+posse+restwma	NA	NA	275.089
2	B	baseef	-	dobef+posse+restwma	0.745	vs A	273.195
3	C	restwma	-	dobef+posse	0.358	vs B	272.04
4	D	-	age	age+dobef+posse	0.436	vs C	273.433
5	E	-	age*baseef	age*baseef+dobef+posse	0.966	vs D	277.363
6	F	age*baseef	age*dobef	age*dobef+posse	0.647	vs D	275.223
7	G	age*dobef	age*restwma	age*restwma+dobef+posse	0.067	vs D	272.033
8	H	age*restwma	age*posse	age*posse+dobef+posse	0.368	vs D	274.623

The steps above can be summarized as follows:

1. A model with all the retained risk factors/variables is fitted, Model A.
2. Steps 2 and 3, the least significant variable is removed. After step 3 all our variables were significant.
3. Step 4, *age* is added into the model as it needs to be adjusted for as it's a confounder.
4. Steps 5-8, interactions between age and the 4 other predictor variables are explored.
5. Model G is chosen.

The final model that was chosen was Model G for two reasons. Firstly it had the lowest AIC amongst all the models that included *age*. Secondly the likelihood ratio test between model G and model D produced a p-value of 0.067, this means that there is sufficient evidence (at a significance level of  $\alpha = 0.07$ ) against the null hypothesis.

The null hypothesis of a likelihood ratio test is that the additional predictors in the more complex model don't improve the models fit. Therefore we can reject the null hypothesis (at a significance level of  $\alpha = 0.07$ ) and conclude that Model G, the more complex model that includes the interaction term between *age* and *restwma* is a better fit to the model than Model D, the simpler model.

The final model, Model G, includes the following predictor variables and the response, *event*:

- *age* - Age of patient in years
- *restwma* - Wall motion anomaly on echocardiogram (1=yes, 0=no)
- *dobef* - Ejection fraction on dobutamine
- *posse* - Stress echocardiography test was positive (1=yes, 0=no)
- *age : restwma* - Interaction term between *age* and *restwma*

All the included predictors are significant at a significance level of  $\alpha = 0.05$  as can be seen in the table of coefficients, Table 7, below:

Table 7: Final logistic regression model

	Estimate	Std. Error	z value	Pr(> z )
(Intercept)	-3.397	2.290	-1.483	0.138
age	0.058	0.029	1.997	0.046
restwma1	5.004	2.386	2.097	0.036
dobef	-0.045	0.013	-3.360	0.001
posse1	0.749	0.351	2.134	0.033
age:restwma1	-0.067	0.033	-2.031	0.042

The final model, Model G, can be expressed as follows:

$$\text{logit}(p) = -3.39655 + (0.05787 * \text{age}) + (5.00355 * \text{restwma}) - (0.04487 * \text{dobef}) + (0.74868 * \text{posse}) - (0.06693 * \text{age} * \text{restwma})$$

## Model Interpretation

When all predictor variables are set to 0 the odds of experiencing a cardiac event are  $\exp(-3.397) = 0.03$ . The 95% confidence interval for the intercept coefficient on the odds scale is:  $(4 \times 10^{-4}; 2.9778)$

The main effect of *age* (*restwma* = 0) can be interpreted as follows: whilst keeping everything else constant, a one unit increase in age will result in the odds of experiencing a cardiac event, changing by a factor of  $\exp(0.05787) = 1.06$  (this is the same as a 6% increase in the odds of experiencing a cardiac event). The 95% confidence interval for the *age* coefficient on the odds scale is: (1.0011; 1.1215)

The interaction term *age* : *restwma* ( $0.05787 - 0.06693 = -0.00906$ ) defines by how much the main effect of *age* changes when a wall motion anomaly is detected *restwma* = 1. The odds change by a factor of  $\exp(-0.00906) = 0.99$  (this is the same as a 1% decrease in the odds). The 95% confidence interval for the *age* : *restwma* coefficient on the odds scale is: (0.8767; 0.9977)

The *restwma* term can't be interpreted on its own due to the presence of the interaction term and setting *age* = 0 doesn't make clinical sense.

Holding everything else constant, a one unit increase in the *dobef* value will result in the odds of experiencing a cardiac event changing by a factor of  $\exp(-0.045) = 0.96$  (this is the same as a 4% decrease in the odds of experiencing a cardiac event). The 95% confidence interval for the *dobef* coefficient on the odds scale is: (0.9314; 0.9815)

Holding everything else constant, if *posse* = 1 i.e. the stress echocardiography test was positive, the odds of experiencing a cardiac event will change by a factor of  $\exp(0.749) = 2.11$ . The 95% confidence interval for the *posse* coefficient on the odds scale is: (1.0631; 4.2047)

## Model Diagnostics

### Goodness-of-Fit test:

Table 8: Hosmer-Lemeshow Test for overall model fit

	X-squared	DoF	P-value
Model G	10.5011	8	0.2316

The overall goodness of fit of a model can be assessed by using a Hosmer-Lemeshow test. The null hypothesis of this test is that the model is a good fit to the data. The alternate hypothesis is that the model is not a good fit to the data.

As can be seen in Table 8, when the Hosmer-Lemeshow test was applied to Model G it produced a p-value of 0.2316, which indicates insufficient evidence against the null hypothesis and thus, it can be said that Model G is a good fit to the data.

### Form of the linear predictor:

Firstly we can plot the deviance residuals against the linear predictor. This allows for us to examine the form of the linear predictor which we can use to determine whether the linear component of the model is adequate or not.

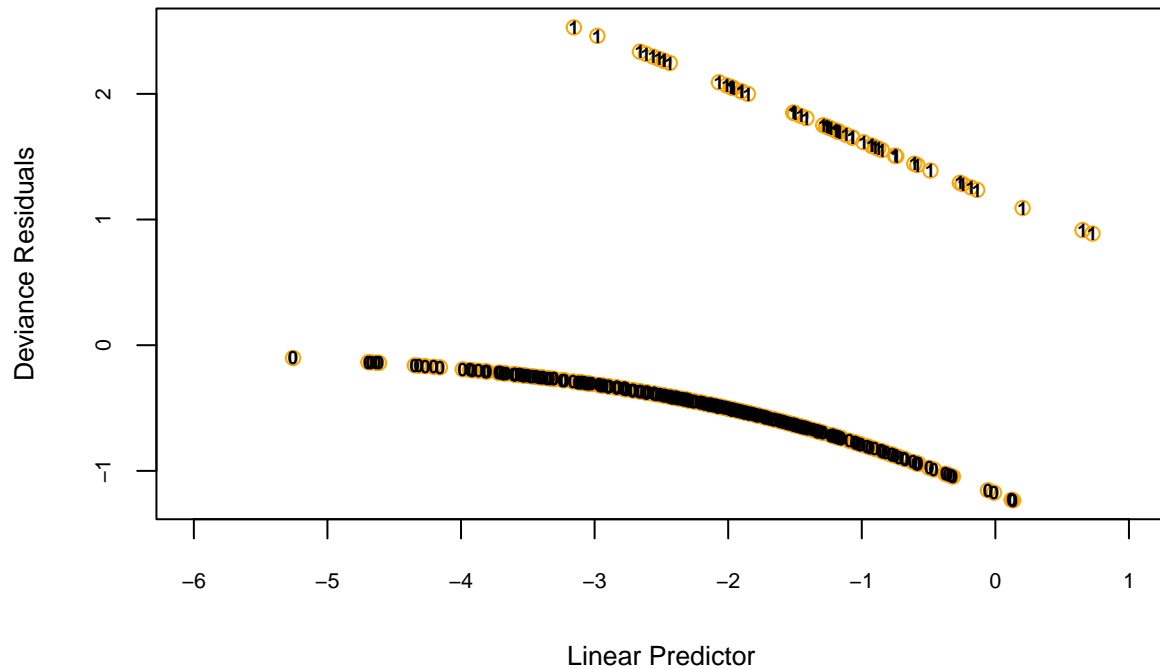


Figure 15: Checking the form of the linear predictor

Figure 15 exhibits a curvilinear band of points with positive deviance residuals for the observations in the data where  $event = 1$  and a second band of points with negative deviance residual values for observations in the data where  $event = 0$ .

This pattern is indicative of the chosen model behaving as expected: as the linear predictors increase, so do the predicted probabilities for  $event = 1$ , resulting in smaller residuals for these cases. On the other hand, the absolute values of the residuals for cases where  $event = 0$  become larger with higher predictor values.

## Outliers:

To determine which observations in the data are classified as outliers, we can look for points for which the standardised residuals are  $> |2|$ . Outliers can be identified graphically as shown in the Figure 16. below by all the points which lie on or above the red line. Outliers can also be determined statistically by examining/manipulating the data.

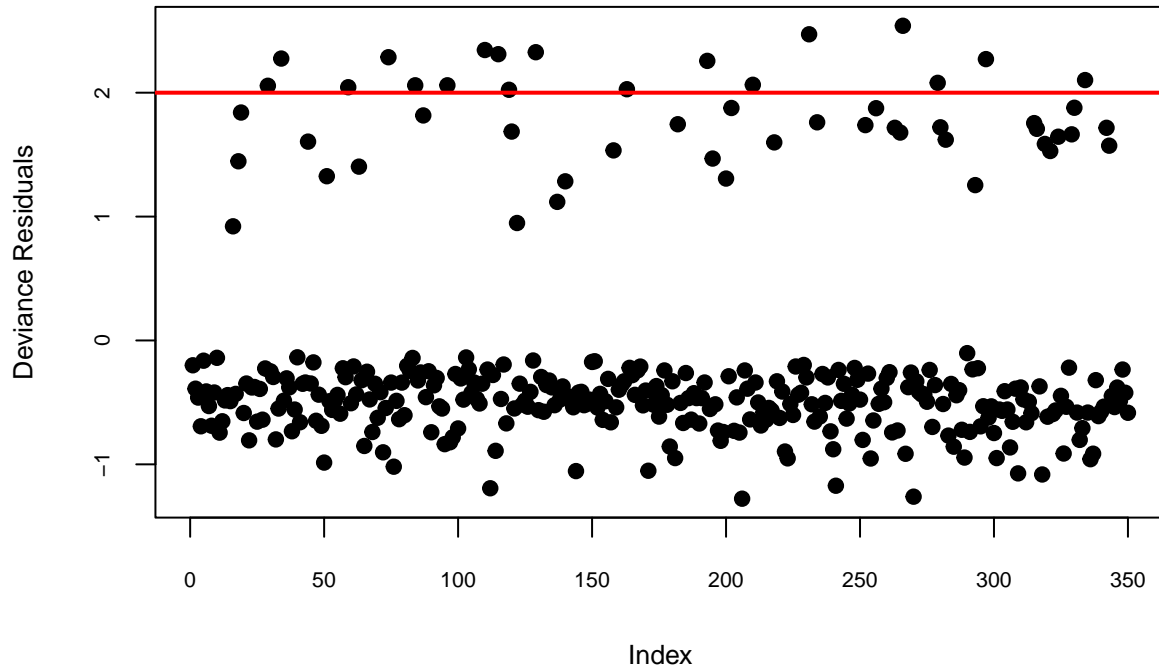


Figure 16: Identifying outliers

Table 9 below shows the cases that were classified as outliers. It is shown that all of these cases were assigned low probabilities of experiencing a cardiac event in the following 12 months,  $event = 1$ , yet the cases still experienced the event. The next step in the model diagnostics will be to check if any of these outlier cases are influential.

Table 9: Cases identified as outliers

	age	baseef	dobef	restwma	posse	event	probabilites
557	70	66	77	1	0	1	0.0770949
480	71	60	72	0	0	1	0.0745581
154	67	61	70	0	0	1	0.0653497
155	59	61	81	1	0	1	0.0716059
517	67	65	69	0	0	1	0.0681444
253	70	64	76	1	0	1	0.0803488
433	77	74	90	0	0	1	0.0483758
489	60	64	72	0	0	1	0.0408856
341	74	57	75	0	0	1	0.0772933



## Influential Points/Cases:

To determine which cases are influential the following plot shown in Figures 17 and 18 can be used.

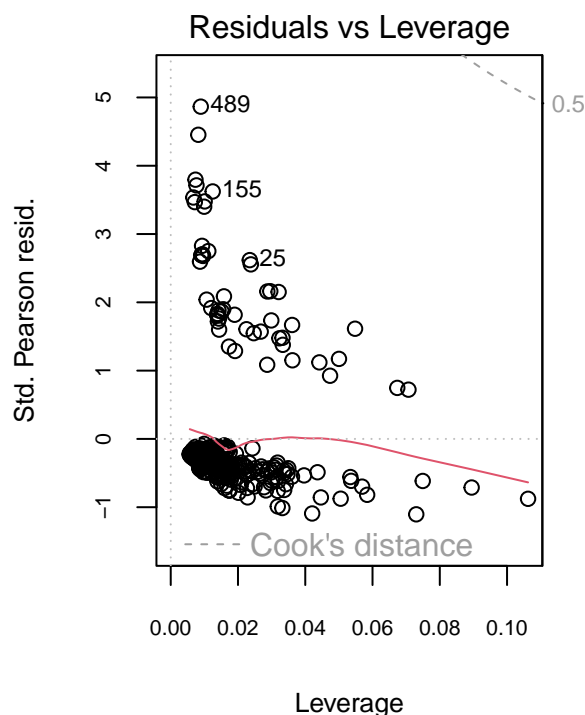


Figure 17: Identifying influential points

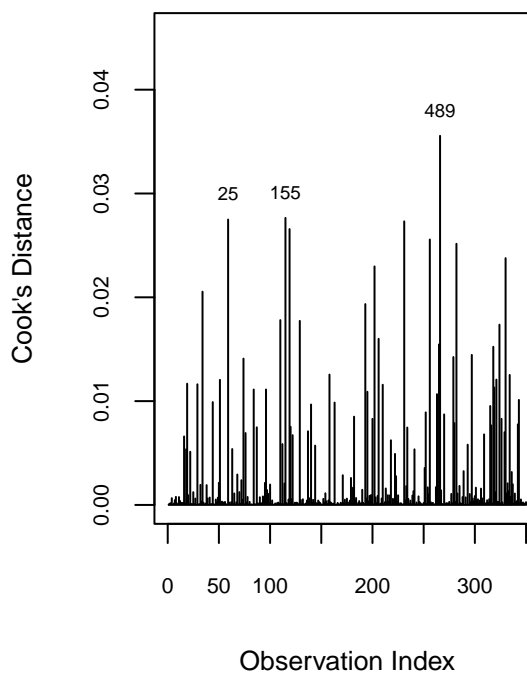


Figure 18: Cook's Distances

Figure 17 shows some of the cases that are clear outliers such as cases 489 and 155 however, neither of these cases nor any of the others depicted in the plot are influential. This can be concluded as none of Cook's distances for these cases are over the common threshold of 1 which identifies a point/case as influential (which is depicted by the dashed line in the top right corner of figure 17). A clearer depiction of the Cook's distances is shown in Figure 18.

Figure 17 shows case 259 as having the maximum leverage with a value of 0.106, however the case is clearly not influential as shown in Figure 17.

We will then observe the dfbeta values for the top 3 most influential cases which are tabulated as follows:

Table 10: dfbeta values for the top 3 most influential cases

	(Intercept)	age	restwma1	dobef	posse1	age:restwma1
489	0.4171389	-0.0053278	-0.4292026	0.0002445	-0.0049313	0.0053342
155	-0.1732120	-0.0001421	0.1399598	0.0026580	-0.0392115	-0.0010360
25	-0.4490573	0.0066574	0.4274492	0.0004585	-0.0119560	-0.0066413

The effects of the observation/case on the regression coefficient estimates (dfbetas) can be interpreted as follows:

Case 489 has a relatively large influence on the *restwma* and intercept coefficient estimate and the sign of the *dfbeta* value suggests that removing this case would increase the *restwma* coefficient estimate and decrease the intercept coefficient estimate. Case 489's *dfbeta* values for the other regression coefficients are relatively small, suggesting that removing case 489 would have little to no significant effect on the coefficient estimate.

Case 155 has a relatively large influence on the intercept and the sign of the *dfbeta* value suggests that removing case 155 would increase the intercept coefficient estimate. *restwma*'s coefficient estimate is also moderately influenced by case 155, removing case 155 would decrease *restwma*'s coefficient estimate. Case 155's influence on the other coefficient estimates is insignificant and removing case 155 would not change the estimates.

The interpretation of case 25 is similar to that of case 155 with the difference being that the effect of removing case 25 has a larger effect on the intercept and *restwma* coefficient estimates.

After a thorough diagnostic analysis, including checks for the goodness of fit of the model, the form of the linear predictor, outliers and influential points, the model exhibits robust performance no significant deviations from what is expected in the diagnostics.

The residuals behave as expected and display no patterns that suggest violations of the underlying assumptions. The influence diagnostics indicate that there are no individual cases significantly impact the model's coefficients.

Overall, the diagnostic tests support the claim that the model is a good fit for the data, thus the next step to be taken is to derive a classification scheme.

## Classification Scheme

A logistic regression model can be used to predict the probability that the response,  $event = 1$ . Therefore to transform these probabilities into a binary outcome we need to apply a classification scheme which makes use of a threshold that determines whether the prediction is encoded into a 1 or a 0.

For example,  $\hat{p} > \pi_0$  then encoded with a 1 and if  $\hat{p} < \pi_0$  then encoded with a 0. Changing  $\pi_0$  can effect the accuracy, sensitivity, specificity, false positive rate, false negative rate and a number of other performance metrics. Unfortunately it is not as simple as finding the best  $\pi_0$  value as there are crucial trade-offs such as: decreasing the threshold increases sensitivity but decreases specificity and vice versa.

In the context of the problem that we are faced with, it might be more worthwhile to strive for higher sensitivity at the cost of lower specificity.

This can be argued as increased sensitivity means the models ability to detect positive cases is enhanced, which is vital in the case of predicting cardiac events where the stakes are high. The trade-off involves an increase in false positives leading to situations where precautions might be taken unnecessarily. In the context of cardiac events missing a true positive could have far more severe repercussions than false alarms which might just involve in taking precautions against a cardiac event even though the prediction was a false positive.

The pros and cons of higher sensitivity vs specificity can be debated eternally but ultimately it was chosen to slightly favour sensitivity over specificity in the context of this problem with the choice being driven by wanting to minimize the risk of overlooking a cardiac event.

The decision on the value for the threshold,  $\pi_0$  was, made based off the ROC curve depicted by Figure 19.

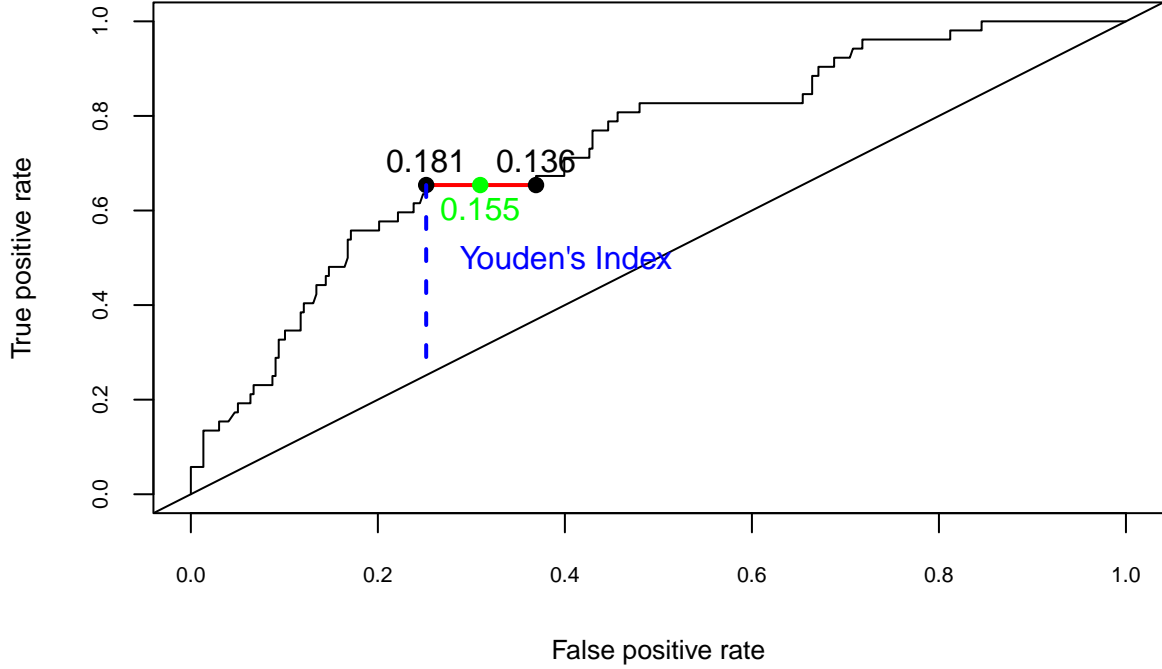


Figure 19: ROC curve for the chosen model

Initially the Youdin's index value was calculated for each threshold value on the ROC curve. The threshold value that produced the highest Youdin's Index value was used as a starting point,  $\pi_0 = 0.181$ . Youdin's index assumes that False Positives and False Negatives are equally desirable, which as mentioned before is not desired, we want slightly more false positives.

Therefore the threshold was selected as the mean of the thresholds that produced the same sensitivity (all the points on the red line) as the threshold that maximised the Youdin's index. The thinking behind this being that even though this minor change in threshold had no effect on the sensitivity in the training data, it may have an effect in the test data.

The final threshold was chosen to be  $\pi_0 = 0.155$  as depicted by the green point in Figure 19.

The final model, Model G and its classification threshold defined as:

$$\text{logit}(p) = -3.39655 + (0.05787 * \text{age}) + (5.00355 * \text{restwma}) - (0.04487 * \text{dobe}) + (0.74868 * \text{posse}) - (0.06693 * \text{age} * \text{restwma})$$

$$\pi_0 = 0.155$$

were then used to make predictions using the validation data set. This yielded the following results:

Table 11: Final metrics from testing Model G on validation data

	Metrics
Sensitivity	0.8108
Specificity	0.6959
Accuracy	0.7163

Table 12: Likelihood Ratio Estimates and their 95% CIs

	Likelihood Ratio Estimate	95% Confidence Interval
LR+	2.666	2.025 ; 3.51
LR-	0.272	0.139 ; 0.534

## Conclusions

The logistic regression analysis identified *age*, *dobef* (ejection fraction on dobutamine), *posse* (stress echocardiography test result), *restwma* (wall motion anomaly), and the interaction between *age* and *restwma* as significant predictors of a cardiac event within the following 12 months.

From this we can provide a conclusion to the primary objective and state that the stress echocardiography test, when using dobutamine to induce stress, is effective in predicting cardiac events. Therefore Dobutamine could prove to be a reliable alternative to the traditional methods of raising a persons heart rate for the purpose of performing a stress echocardiography test which could expand the population of people for which the test is applicable.

In order to truly compare the use of Dobutamine and exercise in inducing stress on the heart for a stress echocardiography test, a model would have to be built using data given on people who induced stress with exercise. The performance of Dobutamine could then be quantified through a comparison of the logistic model built using Dobutamine data and the logistic model built using exercise data.