Survival Analysis of Post-Myocardial Infarction Patients

Alvein, Orr, Pham

5/22/2020

Abstract

** Two sentences each, at most!

[Brief Intro]

[Objective + methods]

Using survival and hazard analysis tools, we see to better understand the nature of survival for patients that have survived the infarction. As rate of infarction cases increase, so will the rate of patients that will require monitoring.We found that

Results We have found that…

[Conclusion] It's clear that survival rates is not the same after a heart attack.??

Introduction

Myocardial infarctions are becoming largely common among U.S. populations. The number of myocardial infarctions are remarkedly increased from [start date] to [end date] by [x] amount (add citation). In 2015, approximately 23% of all fatalities in the United States was related to some degree of heart disease (cdc. cite please). Unsurprisingly, clinical studies have shown harmful symptoms in post-infarction survival patients. Our obtained dataset to examine the tangible difference in survability rates from the course of year following an infarction episode. We examine record

[Insert more kaggle context]

```
##      speed           dist       
##  Min.   : 4.0   Min.   :  2.00  
##  1st Qu.:12.0   1st Qu.: 26.00  
##  Median :15.0   Median : 36.00  
##  Mean   :15.4   Mean   : 42.98  
##  3rd Qu.:19.0   3rd Qu.: 56.00  
##  Max.   :25.0   Max.   :120.00  
```

We have obtained a data set examining the the length and survival of post-myocardial patients over the course of one year.

Objective

We seek to apply survival analysis techniques to this given dats set to examine the characteristics of survival rates among post-infarction patients. Additionally, we hope to achieve better understanding of survival rates as stratified among several classifications.

Dataset

We have obtained our data set from Kaggle. The data set contains 133 of observations and records 9 variables. Please refer to the data set below for more details.

Survival is the the number of months patient survived (has survived, if patient is still alive). Because all the patients had their heart attacks at different times, it is possible that some patients have survived less than one year but they are still alive. Check the second variable to confirm this. Such patients cannot be used for the prediction task mentioned above.

Alive.end is a binary variable if a patient is alive by the end of the survival variable, 0 means the patient has died at the end of the time period while 1 means that patient is still alive.

p.age is a patient's age at the time of the infarction.

If pericardial-effusion is present around the heart, it is recorded as a 1. If the effusion is not present around the heart, it is recorded as a 0.

Fractional shortening, the measure of contracility around the heart, is recorded as continuous numbers. Lower numbers can be concidered as increasingly abnormal.

E-point septal separation (epss), another measure of contractility, is also recorded with continuous numbers. Larger numbers are increasingly abnormal.

Left ventricular end-diastolic dimension is a measure of the size of the heart at end-diastole. Large hearts tend to be sick hearts.

Wall-motion-index is a wall-motion-score divided by number of segments seen. Usually 12-13 segments are seen in an echocardiogram.

Alive-at-1: 0 means patient was either dead after 1 year or had been followed for less than 1 year. 1 means patient was alive at 1 year.

[add table summarizing the headers]

```
##      speed           dist       
##  Min.   : 4.0   Min.   :  2.00  
##  1st Qu.:12.0   1st Qu.: 26.00  
##  Median :15.0   Median : 36.00  
##  Mean   :15.4   Mean   : 42.98  
##  3rd Qu.:19.0   3rd Qu.: 56.00  
##  Max.   :25.0   Max.   :120.00  
```

[add the data set??]


```
##      speed           dist       
##  Min.   : 4.0   Min.   :  2.00  
##  1st Qu.:12.0   1st Qu.: 26.00  
##  Median :15.0   Median : 36.00  
##  Mean   :15.4   Mean   : 42.98  
##  3rd Qu.:19.0   3rd Qu.: 56.00  
##  Max.   :25.0   Max.   :120.00  
```


[add details from kaggle]

Mathematical Definitions

Censoring

Left Censored: Because all the patients had their heart attacks at different times, it is possible that some patients have survived less than one year but they are still alive. Check the second variable to confirm this. Such patients cannot be used for the prediction task mentioned above.

Curves

equations

[latex]

Methodology

Summary statistics

Kaplan-Meier

Weibull Fits

Cox PH

Results

Table of Summary Statistics

```
##      speed           dist       
##  Min.   : 4.0   Min.   :  2.00  
##  1st Qu.:12.0   1st Qu.: 26.00  
##  Median :15.0   Median : 36.00  
##  Mean   :15.4   Mean   : 42.98  
##  3rd Qu.:19.0   3rd Qu.: 56.00  
##  Max.   :25.0   Max.   :120.00
```

KM Curve

```
##      speed           dist       
##  Min.   : 4.0   Min.   :  2.00  
##  1st Qu.:12.0   1st Qu.: 26.00  
##  Median :15.0   Median : 36.00  
##  Mean   :15.4   Mean   : 42.98  
##  3rd Qu.:19.0   3rd Qu.: 56.00  
##  Max.   :25.0   Max.   :120.00
```

Weibull Curve

```
##      speed           dist       
##  Min.   : 4.0   Min.   :  2.00  
##  1st Qu.:12.0   1st Qu.: 26.00  
##  Median :15.0   Median : 36.00  
##  Mean   :15.4   Mean   : 42.98  
##  3rd Qu.:19.0   3rd Qu.: 56.00  
##  Max.   :25.0   Max.   :120.00
```

Cox Proportional Hazard

```
##      speed           dist       
##  Min.   : 4.0   Min.   :  2.00  
##  1st Qu.:12.0   1st Qu.: 26.00  
##  Median :15.0   Median : 36.00  
##  Mean   :15.4   Mean   : 42.98  
##  3rd Qu.:19.0   3rd Qu.: 56.00
```

```
##  Max.   :25.0   Max.   :120.00
```

Model Diagnostics

AIC, BIC, and Confidence Intervals

```
##      speed           dist       
##  Min.   : 4.0   Min.   :  2.00  
##  1st Qu.:12.0   1st Qu.: 26.00  
##  Median :15.0   Median : 36.00  
##  Mean   :15.4   Mean   : 42.98  
##  3rd Qu.:19.0   3rd Qu.: 56.00  
##  Max.   :25.0   Max.   :120.00
```

Residual Analysis/QQ Plot

```
##      speed           dist       
##  Min.   : 4.0   Min.   :  2.00  
##  1st Qu.:12.0   1st Qu.: 26.00  
##  Median :15.0   Median : 36.00  
##  Mean   :15.4   Mean   : 42.98  
##  3rd Qu.:19.0   3rd Qu.: 56.00  
##  Max.   :25.0   Max.   :120.00
```

```
##      speed           dist       
##  Min.   : 4.0   Min.   :  2.00  
##  1st Qu.:12.0   1st Qu.: 26.00  
##  Median :15.0   Median : 36.00  
##  Mean   :15.4   Mean   : 42.98  
##  3rd Qu.:19.0   3rd Qu.: 56.00  
##  Max.   :25.0   Max.   :120.00
```

Discussion

References

Salzberg, S. (1988). Exemplar-based learning: Theory and implementation (Technical Report TR-10-88). Harvard University, Center for Research in Computing Technology, Aiken Computation Laboratory (33 Oxford Street; Cambridge, MA 02138).

Kan, G., Visser, C., Kooler, J., & Dunning, A. (1986). Short and long term predictive value of wall motion score in acute myocardial infarction. British Heart Journal, 56, 422-427.

CDC Heart Dieasease