# Topic: Survival Analytics

**Instructions:**

Please share your answers filled in-line in the word document. Submit code separately wherever applicable.

Please ensure you update all the details:

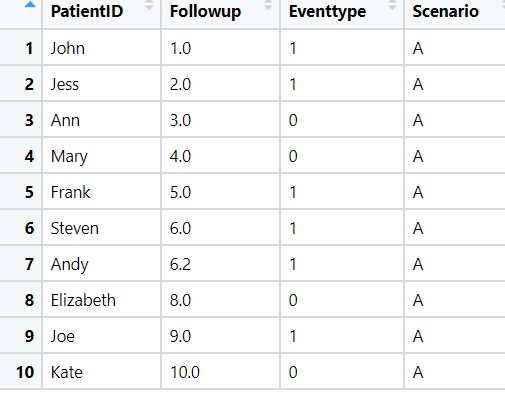
**Name: \_Prajay B. Urkude**

**Batch Id: 16092021**

**Topic: Survival Analytics**

**Problem Statement:**

The following dataset contains patient ID, follow up, event type, and scenarios. Build a survival analysis model on the given data.



Ans: Business Objectives:

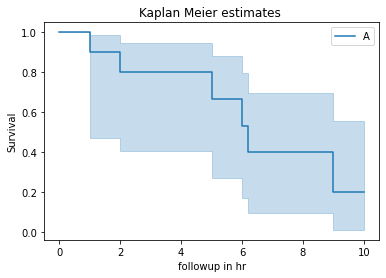
To perform the survival analysis on the above datasets to find the the time within which the patient should be cure.

|  |  |  |  |
| --- | --- | --- | --- |
| **Name of Feature** | **Description** | **Type** | **Relevance** |
| Patient Id | Id of patient | Qualitative, Nominal | Irrelevant |
| Followup | Followup of to the patient | Quantitative, Ratio | Relevant |
| Event type | Event alive (death = 1, survive = 0) | Quantitative, Nominal | Relevant |
| Scenario | Scenario of the of the patient | Qualitative, Nominal | Relevant |

**Steps For the Survival Analysis:-**

* Import the required libraries. Pandas, numpy, matplotlib, lifelines

From lifeline Library import KaplanMeierFitter function for survival analysis

* Load the data
* Doing the univariate analysis and Exploratory data analysis.
* Checking the head i.e., top 5 rows of the datasets
* Checking the columns names of the datasets
* Checking the null values if any available in dataset.
* Checking the duplicate values in the datasets
* Checking the information i.e., datatypes of the datasets
* Exploratory data analysis. mean, median, mode, count, min, max etc.
* Check the distribution of the data.
* Dropping the unwanted column which is not useful for the analysis.
* Converting the nonnumerical data into numerical data by using one hot encoding or Label Encoder or pandas get\_dummies function as per the requirement
* Converting the continuous data into discrete form if necessary.
* **Model Building**
* Define The KaplaneMeierFitter function
* Fit the function in the data by taking survival time and the event as a eventtype.
* Plot the relationship on the graph by using the plt function.

From the above graph we can say that as the follow up time increases the chances of survival decreases and at 1 the chances of death is zero and as the time lapse the chance of survival is decreses so we have to do immediate follow-up after the accident happen so that the patient can survive.

**Problem Statement: -**

A large room

Description automatically generatedECG of different age groups of people has been recorded. The survival time in hours after the operation is given and the event type is denoted by 1 (if dead) and 0 (if alive). Perform survival analysis on the dataset given below and provide your insights in the documentation.

Ans: **Business Objective:**

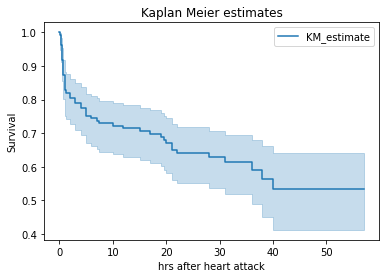
To do the survival analysis on the the above datasets and to find out what are the factors which can cause early death of patients after heart attack.

|  |  |  |  |
| --- | --- | --- | --- |
| **Name of Feature** | **Description** | **Type** | **Relevance** |
| survival\_time\_hr | Survival time of patient in hours | Quantitative, ratio | Relevant |
| alive | Whether the patient death = 1 or alive = 0 | Quantitative, ratio | Relevant |
| age | Age of the patient | Quantitative, ratio | Relevant |
| pericardialeffusion | beat to beat variation in both the amplitude and the axis of the QRS complex | Quantitative, ratio | Relevant |
| fractionalshortening | percentage change in left ventricular diameter during systole | Quantitative, ratio | Relevant |
| epss | E point to septal seperation | Quantitative, ratio | Relevant |
| lvdd | Left ventricular diastolic dysfunction. | Quantitative, ratio | Relevant |
| wallmotion-score | regional abnormalities in contractile function | Quantitative, ratio | Relevant |
| wallmotion-index | Each segment is given a score based on its systolic function | Quantitative, ratio | Relevant |
| multi\_sensor | The electrocardiogram id multisensored | Quantitative, ratio | Relevant |
| name | Name of the patient | Qualitative, Nominal | Irrelevant |
| group | Groups of the patient depends on its condition | Quantitative, ratio | Relevant |

**Steps For the Survival Analysis:-**

* Import the required libraries. Pandas, numpy, matplotlib, lifelines

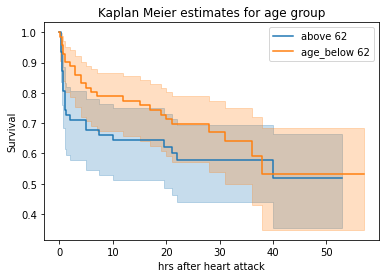
From lifeline Library import KaplanMeierFitter function for survival analysis

* Load the data
* Doing the univariate analysis and Exploratory data analysis.
* Checking the head i.e., top 5 rows of the datasets
* Checking the columns names of the datasets
* Checking the null values if any available in dataset.
* Checking the duplicate values in the datasets
* Checking the information i.e., datatypes of the datasets
* Exploratory data analysis. mean, median, mode, count, min, max etc.
* Check the distribution of the data.
* Dropping the unwanted column which is not useful for the analysis.
* Converting the nonnumerical data into numerical data by using one hot encoding or Label Encoder or pandas get\_dummies function as per the requirement
* Converting the continuous data into discrete form if necessary.
* **Model Building**
* Define The KaplaneMeierFitter function
* Fit the function in the data by taking survival time and the event as a eventtype.
* Plot the relationship on the graph by using the plt function.
* 

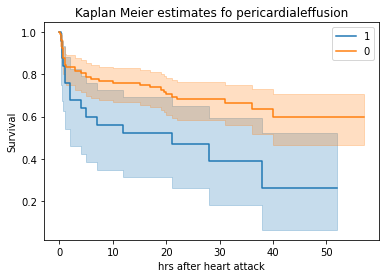
From the above graph we can see that survival rate decreases as the time in hrs increases

The kaplan estimates 1 for intial hours following the treatment and it decreses to 0.65 after 40 hours.'

* Now we will see how the other factors affect the survival rate of the patient.

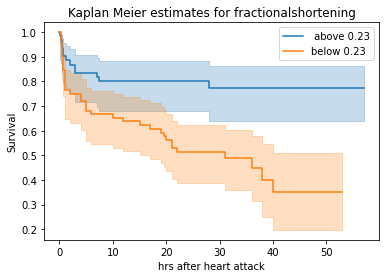
1) Age : Applying kaplanmeierfitter model on survival time and alive Events for the age above/below mean age

I from the above graph we can say that the patient has the more than 62 age there is less chances of hazard at the early stage and if he has the age below 62 the rate of hazard is less compared to age above 62.

2) pericardialeffusion: Applying kaplanmeierfitter model on survival time and alive Events for the pericardialeffusion' = 1 and pericardialeffusion' = 0

If the patient has the pericardialenfussion issue then the rate of hazards happening is high compare to the patient

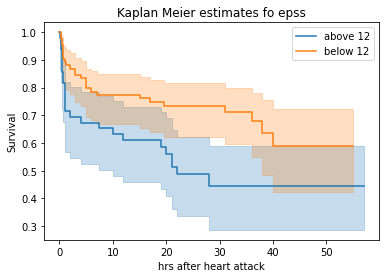
Which don't have this issue.

3) fractionalshortening: Applying kaplanmeierfitter model on survival time and alive Events for fractionalshortening above/below mean value

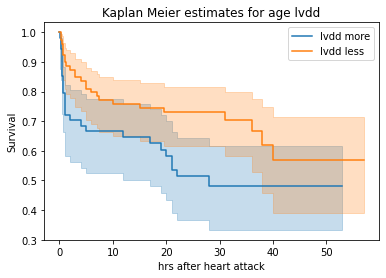
As we cans see from graph as the value is less for fractionalshortening there is less chances of survival

And as the value is more for fractionalshortening there is more chances of survival.

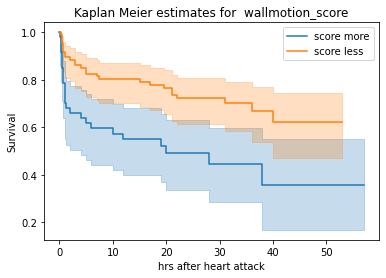
4) epss: Applying kaplanmeierfitter model on survival time and alive Events for epss above/ below mean value



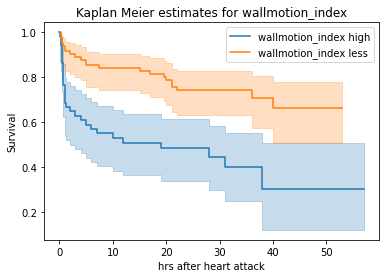
From the graph we can say that the as the epss is high the rate of dander is high and the epss is less then the rate of dander is low.

5) lvdd : Applying KaplanMeierFitter model on survival time and alive Events for lvdd above/below mean value

As lvdd is less then the rate of getting into hazard is low and if lvdd is more then the rate of getting into hazard is less

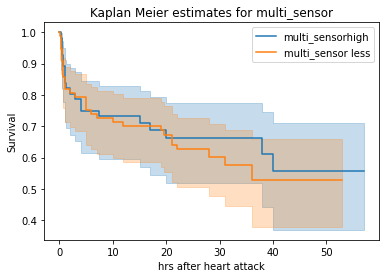
5) wallmotion\_score : Applying KaplanMeierFitter model on survival time and alive Events for wallmotion\_score above/below mean value

As we can see from the graph wallmotion score is less then more chances of survive and if the score is more there is less chances of survival

6) wallmotion\_score : Applying KaplanMeierFitter model on survival time and alive Events for wallmotion\_index above/below mean value

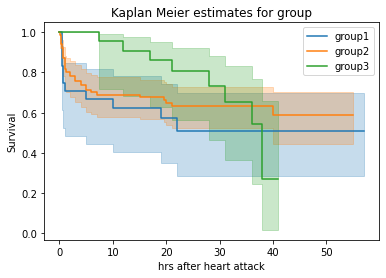
from the graph we can say that as wallmotion index is high then the chances of survival is less with time and if it is iess then then the chances is more with time.

6) multi\_sensor: Applying KaplanMeierFitter model on survival time and alive Events for multi\_sensor above/below meanvalue



If multisensor is less then chance of survival is less and if it is high chance of survival is slightly more

6) Group :- Applying KaplanMeierFitter model on survival time and alive Events for group =1, 2 and 3



Group1 patient has less chances of survival initially and as the as the time lapse and the gives the treatments its chances of survive increses it constat at 50% probability.

Group2 has the less chances of survival initially and as the patient gets treatment its chances of survival increases and it id greater than group1 patient with almost 65 % probability

Group3 has more probability upto 100% of survival upto 10hrs and after 10hrs the probability of survival decreases drastically and at 40th hour patients survival probability reaches to 20% only.

\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*