**Survival Analytics**

**Name: Upadhyay Sachin Naresh**

**Batch ID: Data Science\_08032021**

**Problem 1:**

**1.Business Problem:**

**Objective:**

Analyze the information given in the following ’patients dataset’ to survival analytics.

**2. Work on each feature of the dataset to create a data dictionary as displayed in the below image:**

|  |  |  |  |
| --- | --- | --- | --- |
| NAME OF FEATURE | Description | Type | Relevance |
| Index | Type | quantitative | Irrelevant, it doesn’t have any info |
| PatientID | PatientID | quantitative | Relevant |
| Followup | Followup | quantitative | Relevant |
| Eventtype | Eventtype | quantitative | Relevant |
| Scenario | Scenario | quantitative | Relevant |

**3. Survival Analytics:**

**R code:**

Step 1 install the required packages and import the library

install.packages('survminer')

install.packages("survival")

library(survminer)

library(survival)

library(readr)

step 2 load the data det and check the values

step 3 structure of the data

str(Patient)

'data.frame': 10 obs. of 4 variables:

$ PatientID: chr "John" "Jess" "Ann" "Mary" ...

$ Followup : num 1 2 3 4 5 6 6.2 8 9 10

$ Eventtype: int 1 1 0 0 1 1 1 0 1 0

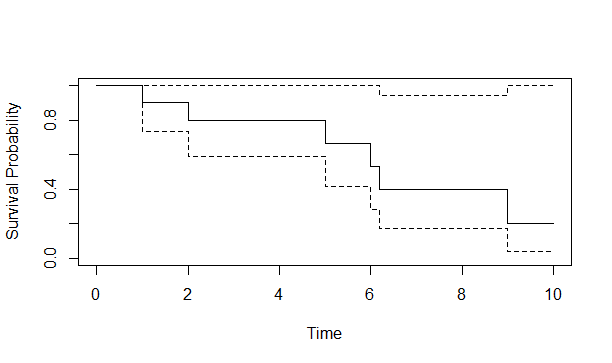
$ Scenario : chr "A" "A" "A" "A" ...

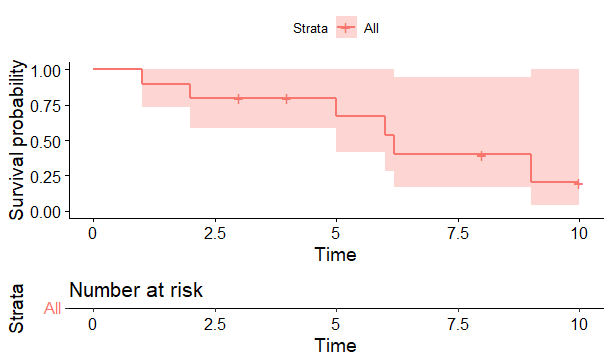
Step 4 removing the id column and describing the variables

Step 5 The group coulumn having the only one category.

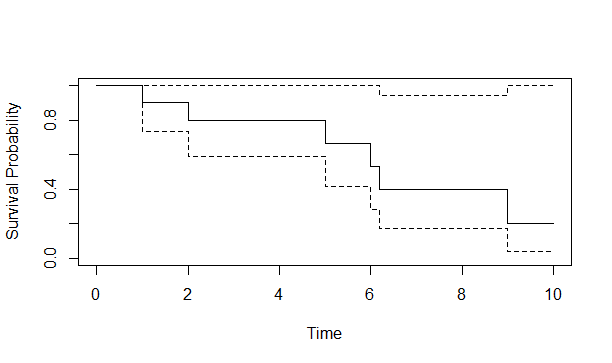
Step 6 Kaplan-Meier non-parametric analysis

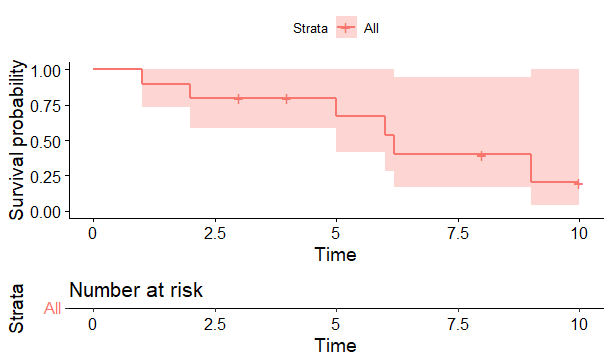
Hear the group has only one category so, we build the through the 1 or group





Step 7 Kaplan-Meier non-parametric analysis by group





**Python code:**

Step 1 load the data and check the dataset

Step 2 describe the dataset

Followup Eventtype

count 10.000000 10.000000

mean 5.420000 0.600000

std 2.993994 0.516398

min 1.000000 0.000000

25% 3.250000 0.000000

50% 5.500000 1.000000

75% 7.550000 1.000000

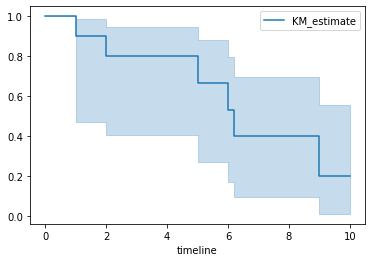
max 10.000000 1.000000

step 3 Importing the KaplanMeierFitter model to fit the survival analysis

step 4 Importing the KaplanMeierFitter model to fit the survival analysis

step 5 Fitting KaplanMeierFitter model on Time and Events for death

step 6 Time-line estimations plot



**4. Conclusion:**

Hence this the survival analytics of the model**.**

**Problem 2:**

**1.Business Problem:**

**Objective:**

Analyze the information given in the following ’ECG\_Surv dataset’ to survival analytics.

Different age group of people has been recorded, the survival time in hours after operation and the event(death) occurred is denoted by 1 and 0 represent still alive

**2. Work on each feature of the dataset to create a data dictionary as displayed in the below image:**

|  |  |  |  |
| --- | --- | --- | --- |
| NAME OF FEATURE | Description | Type | Relevance |
| Index | Type | quantitative | Irrelevant, it doesn’t have any info |
| Survival\_time\_hr | Survival\_time\_hr | quantitative | Relevant |
| alive | alive | quantitative | Relevant |
| Age | Age | quantitative | Relevant |
| Pericardialeffusion | Pericardialeffusion | quantitative | Relevant |
| Fractionalshortening | Fractionalshortening | quantitative | Relevant |
| epss | epss | quantitative | Relevant |
| ivdd | ivdd | quantitative | Relevant |
| Wallmotion-score | Wallmotion-score | quantitative | Relevant |
| Wallmotion- index | Wallmotion- index | quantitative | Relevant |
| Multi-sensor | Multi-sensor | quantitative | Relevant |
| Name | Name | quantitative | Relevant |
| group | group | quantitative | Relevant |

**3. Survival Analytics:**

**R code:**

Step 1 load the library

library(survminer)

library(survival)

library(readxl)

step 2 load the dataset and check the dataset

step 3 str of the ecg data

str(ecg)

tibble[,12] [133 x 12] (S3: tbl\_df/tbl/data.frame)

$ survival\_time\_hr : num [1:133] 11 19 16 57 19 26 13 50 19 25 ...

$ alive : num [1:133] 0 0 0 0 1 0 0 0 0 0 ...

$ age : num [1:133] 71 72 55 60 57 68 62 60 46 54 ...

$ pericardialeffusion : num [1:133] 0 0 0 0 0 0 0 0 0 0 ...

$ fractionalshortening: num [1:133] 0.26 0.38 0.26 0.253 0.16 0.26 0.23 0.33 0.34 0.14 ...

$ epss : chr [1:133] "9" "6" "4" "12.061999999999999" ...

$ lvdd : chr [1:133] "4.5999999999999996" "4.0999999999999996" "3.42" "4.6029999999999998" ...

$ wallmotion-score : chr [1:133] "14" "14" "14" "16" ...

$ wallmotion-index : chr [1:133] "1" "1.7" "1" "1.45" ...

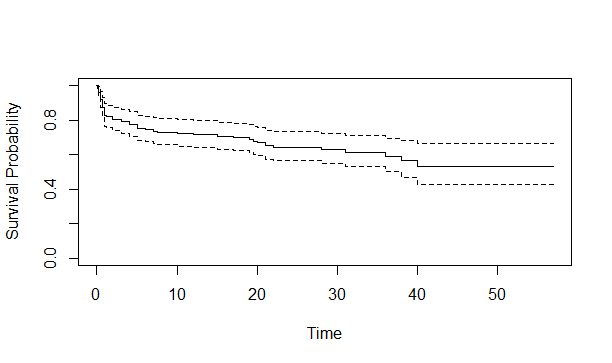
$ multi\_sensor : chr [1:133] "1" "0.58799999999999997" "1" "0.78800000000000003" ...

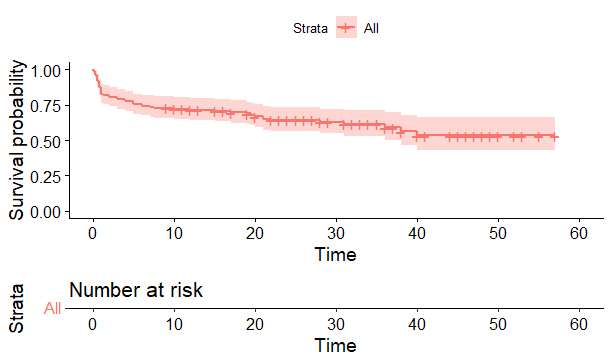
$ name : chr [1:133] "name" "name" "name" "name" ...

$ group : num [1:133] 1 1 1 1 1 1 1 1 1 1 ...

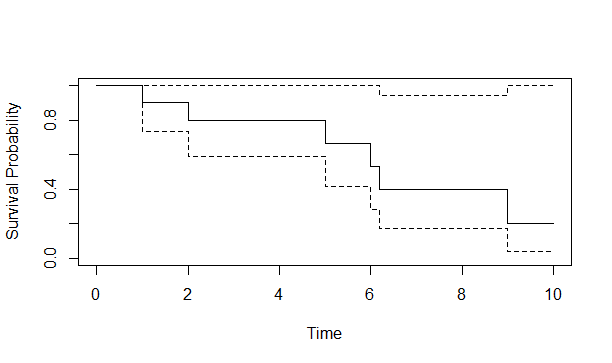
Step 4 defining the variables

Step 5 Kaplan-Meier non-parametric analysis





Step 6 Kaplan-Meier non-parametric analysis by group



**Python code:**

Step 1 load the data and check the data

Step 2 describing the data set

**survival\_time\_hr alive ... multi\_sensor group**

**count 133.000000 133.000000 ... 129.000000 133.000000**

**mean 21.795338 0.383459 ... 0.786202 1.984962**

**std 15.885313 0.488067 ... 0.225661 0.590133**

**min 0.030000 0.000000 ... 0.140000 1.000000**

**25% 6.000000 0.000000 ... 0.714000 2.000000**

**50% 22.000000 0.000000 ... 0.786000 2.000000**

**75% 33.000000 1.000000 ... 0.857000 2.000000**

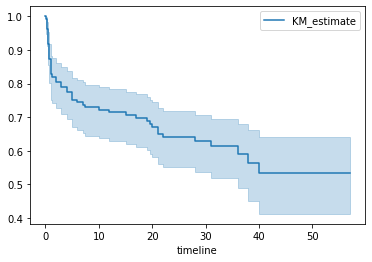
**max 57.000000 1.000000 ... 2.000000 3.000000**

**[8 rows x 11 columns]**

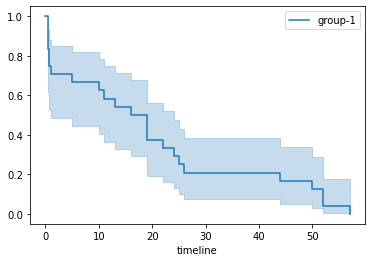
Step 3 Importing the KaplanMeierFitter model to fit the survival analysis

Initiating the KaplanMeierFitter model

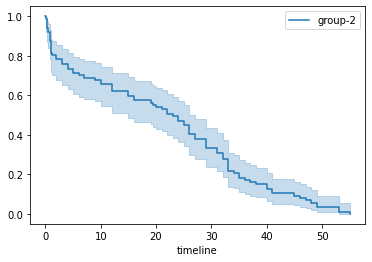
Step 4 Fitting KaplanMeierFitter model on Time and groups for death



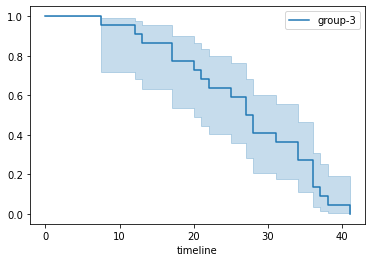
Step 5 Applying KaplanMeierFitter model on Time and groups for the group "1"



Step 6 Applying KaplanMeierFitter model on Time and groups for the group "2"



Step 7 Applying KaplanMeierFitter model on Time and groups for the group "3"



**4. Conclusion:**

Hence the survival analytics is done for the ECG data.