

Topic: Survival Analytics

Perform Kaplan Meier analysis for the given data and get the life table

	PatientID	Followup	Eventtype	Scenario
1	John	1.0	1	A
2	Jess	2.0	1	A
3	Ann	3.0	0	A
4	Mary	4.0	0	A
5	Frank	5.0	1	A
6	Steven	6.0	1	A
7	Andy	6.2	1	A
8	Elizabeth	8.0	0	A
9	Joe	9.0	1	A
10	Kate	10.0	0	A

Solution:

- 1) Loading the dataset.
- 2) The dataset consists of 10 observations and 4 variables.
- 3) Define the variable names for Follow up and eventtype, where event takes the 2 values 0's and 1's which mean weather the patient have the disease or with no disease.

Kaplan – Meier Non Parametric Analysis:

- 4) This type of analysis, the time to a specific event, such as death or disease recurrence, is of interest and two (or more) groups of patients are compared with respect to this time.
- 5) A non-parametric statistic is not based on the assumption of an underlying probability distribution, which makes sense since survival data has a skewed distribution.
- 6) These statistics gives the probability that an individual patient will survive past a particular time t . At $t = 0$, the Kaplan-Meier estimator is 1 and

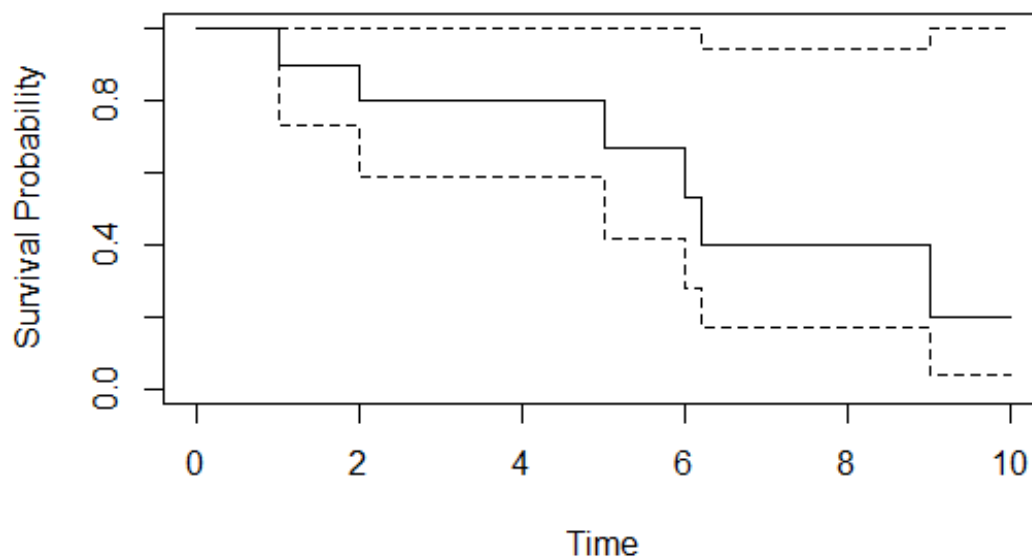
with t going to infinity, the estimator goes to 0.

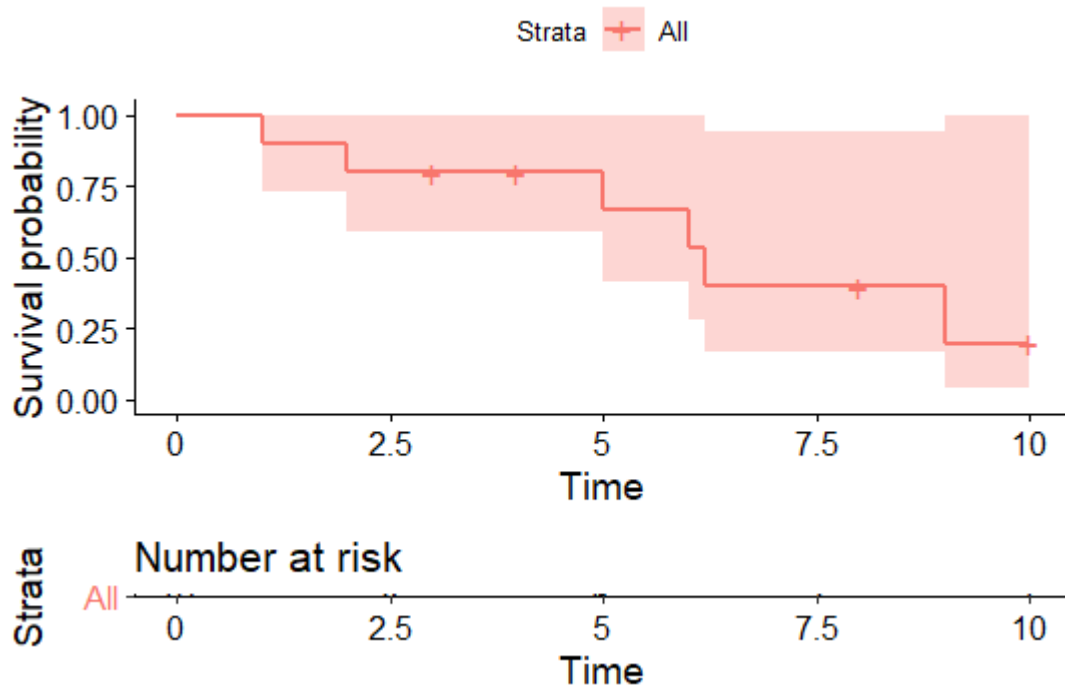
```
time n.risk n.event survival std.err lower 95% CI upper 95% CI
1.0    10      1    0.900  0.0949    0.7320    1.000
2.0     9      1    0.800  0.1265    0.5868    1.000
5.0     6      1    0.667  0.1610    0.4153    1.000
6.0     5      1    0.533  0.1755    0.2798    1.000
6.2     4      1    0.400  0.1751    0.1696    0.943
9.0     2      1    0.200  0.1663    0.0392    1.000
> |
```

- ✓ From the above summary, as the follow up increases, the risk of the patient will be decreased.

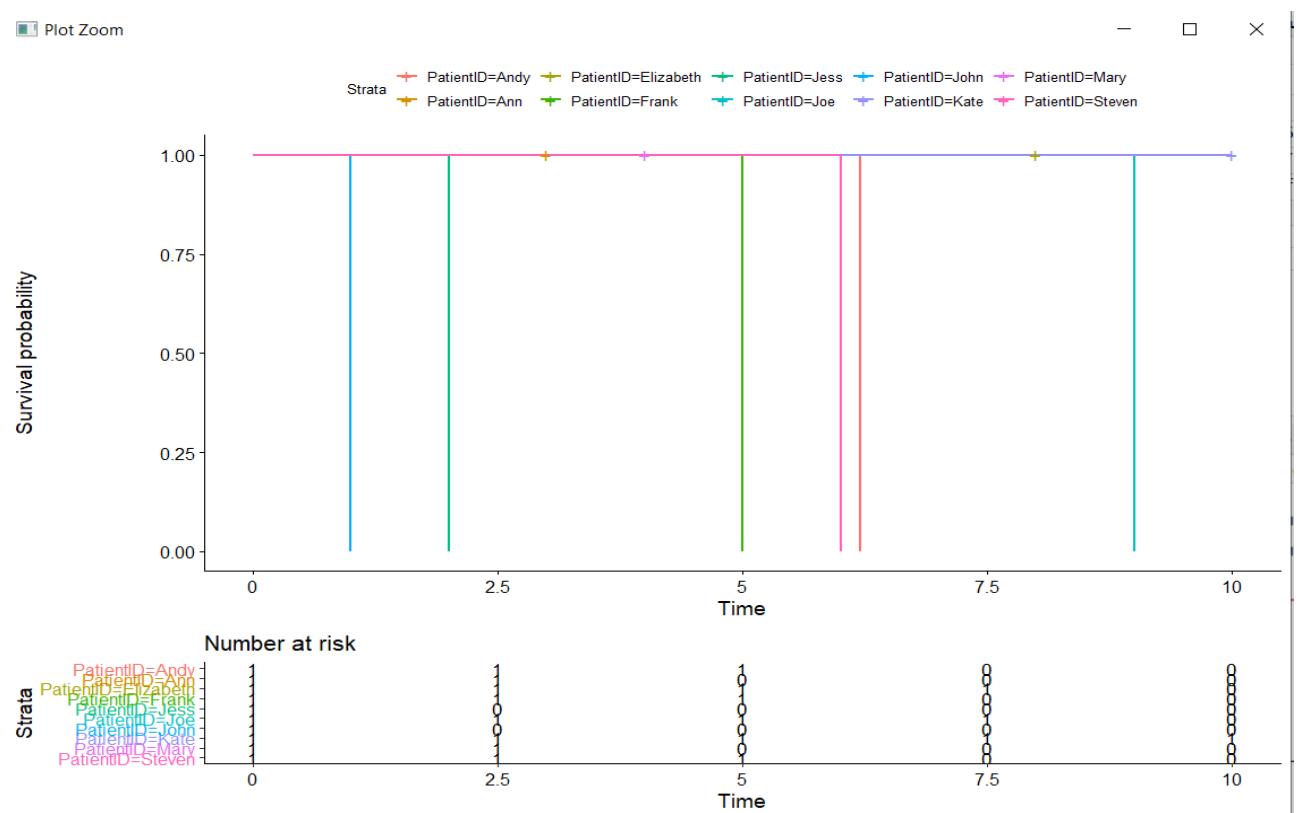
Graphical Representation:

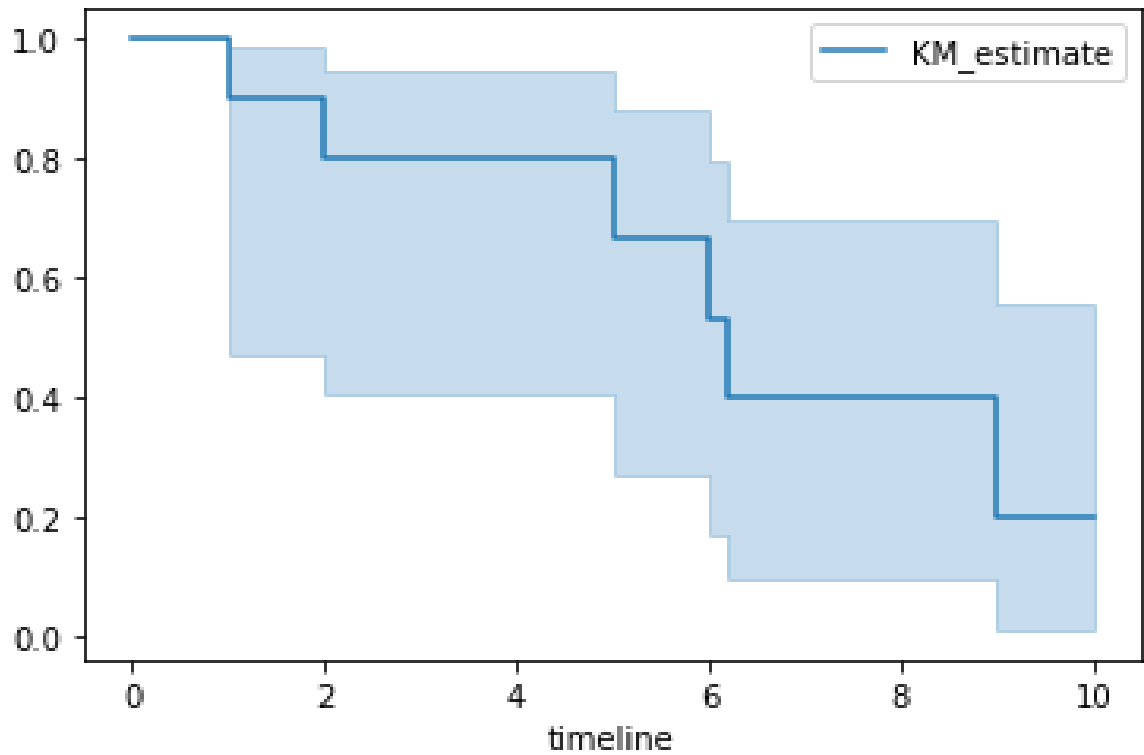
- ✓ Below graph shows the relationship between time and event for the patients with the probability values and the risk range forming with group 1





✓ Below graph represent the relation between the event, time and patients.





Inferences:

Using the survival function shows that, if the patient increases the follow up then the risk of the patient getting disease also decreases. Which means that if the patient follow ups on a regular intervals of time ,then his survival becomes inclusive.

Hints:

1. Business Problem
 - 1.1. Objective
 - 1.2. Constraints (if any)
2. Data Pre-processing
 - 2.1 Data cleaning, Feature Engineering, EDA etc.
3. Model Building
 - 3.1 Partition the dataset
 - 3.2 Model(s) - Reasons to choose any algorithm
 - 3.3 Model(s) Improvement steps
 - 3.4 Model Evaluation
 - 3.5 Python and R codes
4. Result Share the benefits/impact of the solution - how or in what way the business (client) gets benefit from the solution provided.

Note:

1. For each assignment the solution should be submitted in the format
2. Research and Perform all possible steps for improving the model(s) accuracy & reduce the RMSE (also evaluate errors like MAPE, MAE etc.)
3. All the codes (executable programs) are running without errors
4. Documentation of the module should be submitted along with R & Python codes, elaborating on every step mentioned here