

## Task 5 – Rohit Garg

### 1. Objective

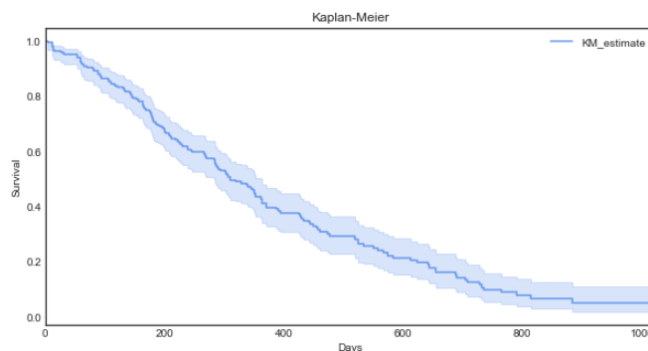
The current analysis is done to understand the survival in patients with advanced lung cancer from the North Central Cancer Treatment Group.

- Determine the survival curve through the Kaplan Meyer Estimator
- Understand differences between Males and Females
- Driver Analysis with Cox Proportional Hazard

### 2. Survival Analysis

Survival Analysis is very common for Subscription type businesses.

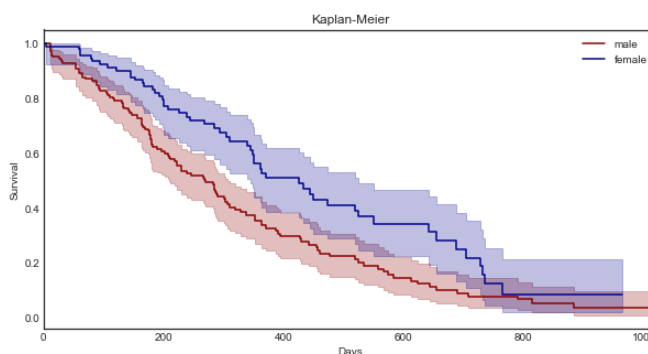
- **Kaplan Meier Estimator** is non-parametric statistic used to estimate the survival function (probability of a person surviving) from the lifetime data. In medical research, it is often used to measure the fraction of patients living for a specific time after treatment or diagnosis.
- **Right Censoring** is done. The subject under observation is still alive. In this case, we can not have our timing when our event of interest (death) occurs.



*At time 0 days 100% of patients survived.  
After 6 months 75% of patients survived  
After 9 months 50% of patients survived  
After 15 months 25% of patients survived  
After 3 years 0% of patients survived*

Log Rank Test is done to test if there are statistical differences in the survival distribution of 2 groups

- **Null Hypothesis** is there is no difference between both groups
- If **p value > 0.05** then we accept the null hypothesis



*test\_statistic: 10.33  
p-value: <0.005  
-log2(p): 9.57*

*Hence, we reject the null hypothesis and claim that the survival rate for male and female are very different*

### 3. Cox Proportional Hazard Regression

Survival Analysis does not allow other predictors. Thus, Cox Proportional Hazard regression helps to determine the relationship between the survival time of a subject and one or more predictor variables

It is observed that **sex** and **ph.ecog** are the key predictors.

	coef	exp(coef)	se(coef)	z	p	-log2(p)
<b>sex</b>	-0.51	0.60	0.20	-2.59	0.01	6.71
<b>ph.ecog</b>	0.48	1.62	0.13	3.65	<0.005	11.88

Concordance	0.64
Partial AIC	1000.75
log-likelihood ratio test	19.48 on 2 df
-log2(p) of ll-ratio test	14.05

