Contents

[Useful Links 1](#_Toc15116547)

[Censored data 2](#_Toc15116548)

[Point censoring 2](#_Toc15116549)

[interval censoring 2](#_Toc15116550)

[Hazard function 3](#_Toc15116551)

[Survival Function 3](#_Toc15116552)

[1. Non-parametric solution (Kaplan Meier Product Limit Method) 3](#_Toc15116553)

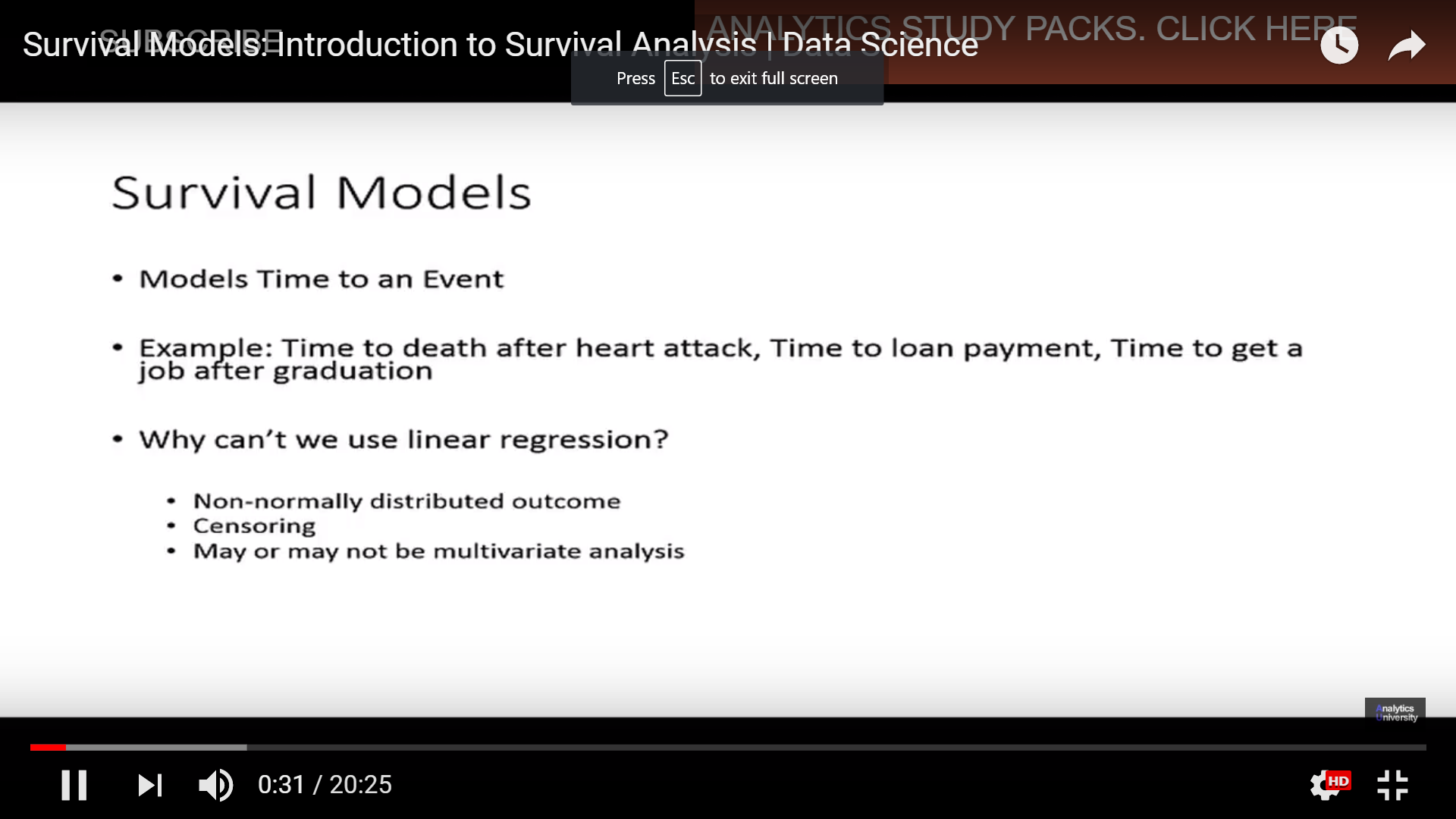
[2. Semi – Parametric solution( Cox-Proportional Hazard Method ) 3](#_Toc15116554)

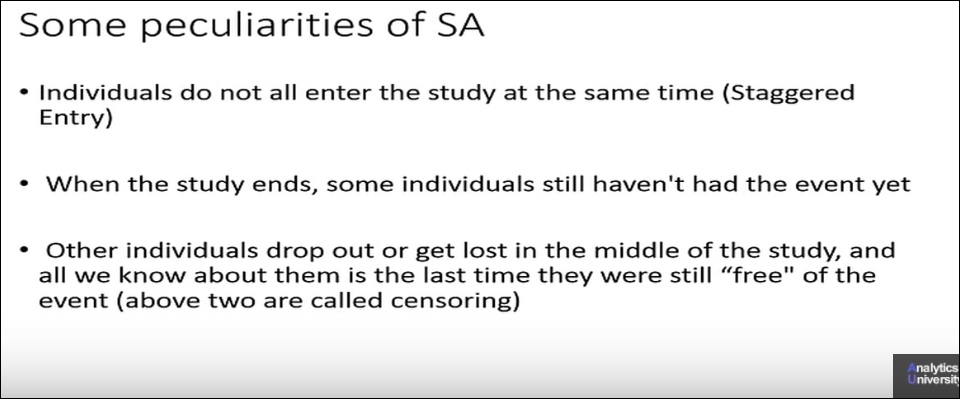
[3. Parametric solution ( Weibull and Exponential Methods) - 3](#_Toc15116555)

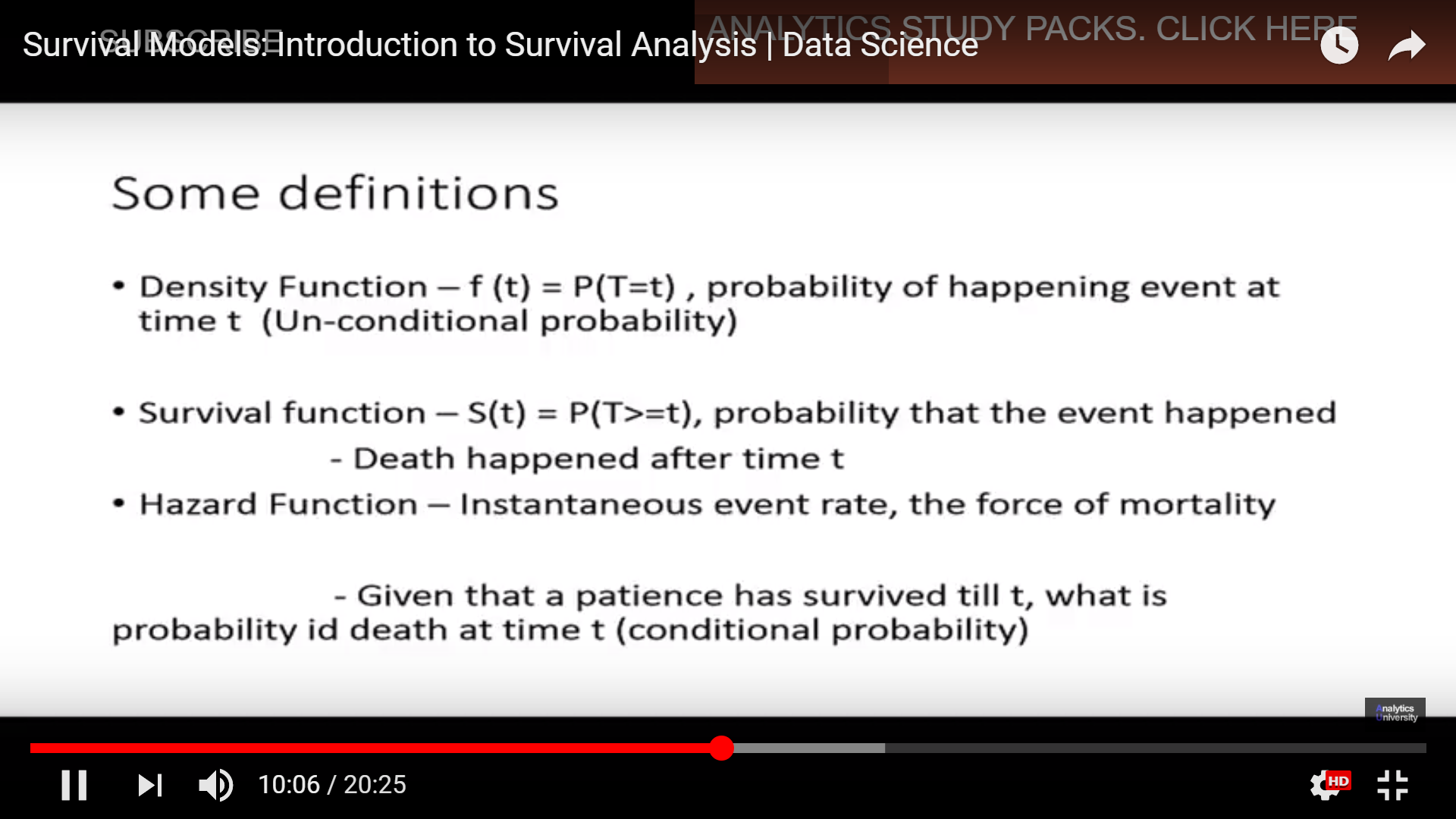
[Kaplan-Meier Estimators 3](#_Toc15116556)

# Useful Links

<https://github.com/anurag-code/Survival-Analysis-Lifelines/blob/master/Survival%20Analysis%20-%20Quick%20Implementation.ipynb?source=post_page--------------------------->







# Censored data

* Data for which the result is not known

## Point censoring

**Right censoring-** *event does not occur within the study duration*

Observations are called censored when the information about their survival time is incomplete; the most commonly encountered form is right censoring. Suppose patients are followed in a study for 20 weeks. A patient who does not experience the event of interest for the duration of the study is said to be right censored.

**Left censoring -***An individual is said to be left censored if the patient had been on risk for disease for a* period before entering the study.

## interval censoring

when time to event may be known only up to a time interval

Hazard function (instantaneous rate of occurrence of the event)

The hazard function gives the potential that the event will occur, per time unit, given that an individual has survived up to the specified time

x

# Survival Function

**Probability of surviving beyond a point**

The dependent variable in survival analysis is composed of two parts: one is the time to event and the other is the event status, which records if the event of interest occurred or not

Following are the three ways of estimation :

# 1. Non-parametric solution (Kaplan Meier Product Limit Method)

Nonparametric tests are also called distribution-free tests because they don't assume that your data follow a specific distribution.

Simplest solution which can be used for **descriptive** analysis, but cannot be extrapolated to find out the survival of censored data with high time span.

A nonparametric estimator of the survival function, the Kaplan Meier method is widely used to estimate and graph survival probabilities as a function of time. It can be used to obtain univariate descriptive statistics for survival data, including the median survival time, and compare the survival experience for two or more groups of subjects

# 2. Semi – Parametric solution( Cox-Proportional Hazard Method )

(Widely used in the industry. Will be discussed in detail in this article.

# 3. Parametric solution ( Weibull and Exponential Methods) -

We will not touch up on this route of estimation. The reason being the parameters found by different software have different signs. We will cover this in one of the coming articles.

# Kaplan-Meier Estimators

Although the math behind Kaplan-Meier estimators is extremely simple ([Wikipedia link](http://en.wikipedia.org/wiki/Kaplan%E2%80%93Meier_estimator), for those interested), we won’t go into it here. Instead, suffice it to say that Kaplan-Meier estimators predict survival probabilities over a given period of time for “right-censored” data. “Right-censored” just means that some of the observations in the data weren’t observed for as long as the period the researcher is interested in analyzing. (For example, we want to look at a year of churn, but some of our customers signed up a month ago).

Kaplan-Meier estimators reliably incorporate all available data at each individual time interval to estimate how many observations are still “surviving” at that time.

