VPM's B.N BANDODKAR COLLEGE OF SCIENCE
DEPARTMENT OF STATISTICS
T.Y.B.Sc SEMESTER V
SURVIVAL ANALYSIS

INTRODUCTION:

In many studies, the outcome of interest is related to the timing of the occurrence of an event. In a clinical setting, one may be interested in measuring how long a chronically ill patient survive after receiving a certain treatment. In another scenario, one may be interested in determining which of three drugs, compared to placebo, provides symptom relief most rapidly.

In many area's such a medicine, biology, epidemiology, demography, economics, engineering, actuarial science and other fields the statistical analysis of what are variously referred to as lifetime, survival time, failure time, time to event or duration data plays an important role in survival analysis.

The objectives of modelling and statistical analysis of lifetime data are to estimate the probability of an event for average lifetime of an object under study or to compare the lifetime distribution of experimental objects(may be animal, human beings or machine).There may be some factors which affect survival time of an object, to identify such risk factors is also an important objective of survival analysis.

Survival analysis deals with the analysis of time until the occurrence of a well-defined event. Event may be death of an individual or occurrence of disease or complication in a disease. In industrial application it may be time to failure of a unit or some component in a system. Within economics it can be the time to acceptance of a job offer for an unemployed person. In demography an event can be getting married.

Following examples will illustrate the way in which lifetime data arise:-

1) Manufactured items with mechanical or electronic components are often subjected to life test in order to obtain information on their durability. This involves putting items in operation often in a laboratory setting and observing them until they fail. So that lifetime of unit is a time from installation to the time it seizes operating satisfactorily.

2) Demographer's and social scientists are interested in the duration of certain life states for humans. For eg : marriage: the lifetime of a marriage is time from day of marriage to its end. End of marriage may be due to divorce or death etc.

A standard experiment in the investigation of carcinogenic substances is one in which laboratory animals are subjected to doses of the substances and then observed to see if they develop tumors. Here variable of interest is the time of appearance of tumor measured from when the dose is administered.

TIME TO EVENT DATA AND CENSORING:

For each subject enrolled in a study, the researcher records the amount of time elapsing between the point at which each subject entered into the study until he or she experiences one of the three possible events i.e. the event occurs, the event does not occur, or the subject is lost to follow up. The total amount of time between the initial enrolment in the study and the occurrence of one of the three outcomes is known as the research subject's survival time, or time to event. Hence the information gathered on each subject is often referred to as survival data or time to event data.

DEFINITION OF TIME TO EVENT DATA:
Survival data, or time to event data are measurements of elapsed time between the initial enrolment in a study and the final disposition of the study subject. This elapsed time could be represented by the time of initial diagnosis or it could be represented by the point in time when one enters the study.

DEFINITION OF CENSORED DATA:
Censored data are represented by measurements for which we have some information about survival time, but the exact survival time is not known.

In singly censored data, a fixed number of subjects enter into a study at the same time. Once in the study, some of the subjects will not experience the event. Their survival time is known to be some length of time greater than the length of the study. This is known as Type I censoring.
It could also be that for research or ethical reasons the study is ended after a certain proportion of the subjects experience the condition of interest, with the remaining proportion having not experienced the event when the study is ended. This is called Type II censoring. Another type of censoring that may occur is known as progressively censored data in which the period of study is fixed, but the subject may enter the experiment at different times. Patients may then either experience or not experience the event of interest, with those not experiencing the event having unknown survival times. This is called Type III censoring. Data for which exact endpoints are not known, either because the subject dropped out of the study, was withdrawn from the study, or survived beyond the termination of the study are called right censored data because the survival times extend beyond the right tail of the distribution of survival times. Conversely we could have data for which exact beginning points are not known. For eg: if a subject with the condition enters the study, but it is not known exactly when the condition developed in the patient then these data are known as left censored data because their survival times are truncated on the left side of the distribution of the survival time distribution, causing the difference in time between diagnosis and entering into the study to be unknown.

Generally for purposes of analysis, a dichotomous variable is used to distinguish survival times of those subjects who experience the event of interest and those that do not because of the censoring mechanisms. This variable is called s status variable with zero indicating that event do not occur and hence the survival time is censored, and a 1 indicating that the event of interest did occur.

STATISTICAL DISTRIBUTION FUNCTIONS:

SURVIVAL RANDOM VARIABLE:-

A r.v is a survival time r.v if an observed outcome lies in the interval $[0,\infty)$, usually survival time r.v is denoted by T.

Like other r.v's lifetime r.v is described or characterized by different function such as cumulative distribution function, pdf, survival function, hazard function. All this function's are mathematically equivalent in the sense that if one of them is known others can be obtained.

CUMULATIVE DISTRIBUTION FUNCTION (CDF):

CDF or simply df of a lifetime r.v T denoted by F(t) and is defined by $F(t) = P(T \leq t)\ \forall\ t \geq 0$

Characteristic properties of distribution function are,

(i) $0 \leq F(t) \leq 1$.

(ii) F(t) is non decreasing function.

(iii) F(t) is right continuous.

(iv) $\lim_{t \to 0} F(t) = 0 \quad ; \ \lim_{t \to \infty} F(t) = 1$

i.e $F(0) = 0$; $F(\infty) = 1$

Probability distribution function (pdf) :

A real valued function f(t) is said to be pdf of lifetime r.v T if it satisfies the following properties

(i) $f(t) \geq 0$ (ii) $\int_0^\infty f(t)dt = 1$

These two properties are known as characteristic properties of pdf.

SURVIVAL FUNCTION:

The survival function of lifetime r.v T is denoted by S(t) or $\bar{F}(t)$ and is defined as,

$\bar{F}(t) = P(T > t)\ ; t\ \geq 0$

Characteristic property of survival function

(i) $0 \leq \bar{F}(t) \leq 1$

(ii) $\bar{F}(t)$ is non – increasing function of T

(iii) $\bar{F}(t)$ is right continuous.

(iv)) $\lim_{t \to 0} \bar{F}(t) = 1 \quad ; \ \lim_{t \to \infty} \bar{F}(t) = 0$

$\bar{F}(t) = \ P(T > t) = \ 1 - P(T\ \leq t) = \ 1 - F(t)$=S(t)

The CDF represents the probability that an event time is less than or equal to some specified measurement time. F(t) is an increasing function that runs from a value of zero (it is assumed theoretically that no events have occurred at the initiation of the study), to a value of 1 (it is assumed theoretically that all events have occurred at the conclusion of the study).

In Survival analysis, a closely related function that is more commonly used than F(t) is a function that runs from a value of 1 (it is assumed that all subject at the initiation of the study have survived to that point) to a value 0 (it is assumed theoretically that none of the subjects have survived when the study ends, though some subjects may be censored).

It is known as survival time distribution and defined as S(t) = 1 – F(t)

HAZARD FUNCTION:

Hazard function of a lifetime r.v T is denoted by h(t) and is defined as the limiting value of the ratio of the conditional probability that component will fail within the time interval t to t+δt given that component already survived upto time t, upto δt where δt is very small.

$$h(t) = \lim_{\delta t \to 0} \frac{P(t<T \leq t+\delta t / T>t)}{\delta t}$$

Properties : -

(i) $h(t) \geq 0$ for all $t \geq 0$

(ii) $\lim_{t \to 0} h(t) = 0$

Hazard function is commonly used for describing the behaviour of failure. Hazard function provides rate of failures within a small interval of time given that component is alive at the beginning of interval whereas survival function gives the probability that component will survive beyond certain time t. Hazard function is also known as hazard rate, force of mortality, failure rate etc.