1.1 General Statistics Concepts
What is Statistics?

There are three kinds of lies: lies, damned lies, and Statistics.

- attributed by Mark Twain to Benjamin Disraeli
A statistician can have his head on an oven and his feet in ice and he will say that on the average he feels fine.

- by ?!? -
It is easy to lie with Statistics, but it is easier to lie without them.

- Frederick Mosteller
Science of data which involves collecting, classifying, summarizing, organizing, analyzing and interpreting numerical information.

Science of making decision
Review

- Probability
- Random variable
- Expectation, variance, covariance and correlation
- Distribution function
- Sample mean and median
- Sampling distribution
- Estimation
- Hypothesis testing
- Comparison of mean and distribution
- Regression analysis
- Parametric/ semiparametric / nonparametric inference
What is Biostatistics?

Referring to applications of statistical methods in the biological and medical science. Specific areas of biological research in which statisticians often play an important role are varied - agriculture, forestry, ecology and experimental biology. In medical research, statisticians are involved in the design and analysis of cross-sectional studies such as estimating disease prevalence or cohort studies of modelling disease progression and prospective randomized clinical trials in evaluate experimental therapeutic interventions.

- Survival analysis
- Even history data analysis (Longitudinal analysis)
- Sequential analysis
What is Survival Analysis?

It represents one of the principal tools of the modern biostatistician and these methods that were developed to capture the information hidden in the length of time from the beginning of an investigation to an occurrence of the event of interest. Especially, survival analysis focuses on understanding how the time from origin to occurrence of an event varies in target population or whether concurrently measured explanatory variable (covariates) such as gender, age, treatment received, year or place of manufacture, exposure to suspected toxic substances, etc are associated with the pattern of variation.
Unique feature of Survival Analysis: Censoring and truncation
- Due to the limitation of a study design, the outcome of primary interest may not always be observed.
1.2 Basic Survival Concepts
Cumulative Density Function (CDF)

$T$: Time until a specific event happen where $0 \leq T \leq \infty$. Positive random variable.

CDF of $T$ is defined by

$$F(t) = P(T \leq t).$$

Properties

- $F(0) = 0$.
- $\lim_{x \to \infty} F(x) = 1$.
- $F(x)$ is nondecreasing function. For all $x_1 \leq x_2$, $F(x_1) \leq F(x_2)$.
- $F$ is right continuous and has left limit.
Probability Density Function (PDF) If \( P(t \in A) = \int_{t \in A} f(t)dt \) for any subset of \( \mathbb{R} \), \( f \) is called the pdf of \( T \). In this case, \( T \) is continuous.

Probability Mass Function (PMF) If \( P(t \in A) = \sum_{t \in A} f(t) \) for any subset of \( \mathbb{R} \), \( f \) is called the pmf of \( T \). In this case, \( T \) is discrete.
Survival Function

\[ S(t) = P(T \geq t). \]

Properties

- \( S(t) = 1 - F(t) \)
- \( S(0) = 1 \)
- \( \lim_{x \to \infty} S(x) = 0. \)
- \( S \) is nonincreasing.
- \( S \) is right continuous, has left limit.
- \( f(t) = \frac{d}{dt} F(t) = -\frac{d}{dt} S(t). \)