Inferential Statistics

The procedures applied to find the probability of an event are known as inferential statistics. It involves observing a sample from a population and using analytical tools to draw conclusions from those observations.

Sampling

Since it is impossible to observe a complete population, it is a common practice to select a sub-group from the population using a certain logic based on the nature of the study; this procedure is known as sampling.

Types of Sampling

1. ***Simple Random Sampling***: In this technique, all the data points have equal probability of selection. The samples are drawn at random. It can be further divided into two types:
    a. Sampling with replacement: where data points have a probability of redundancy.
    b. Sampling without replacement: where data points can be drawn only once.
2. ***Stratified Sampling***: In this technique, the data points are first divided into sub-groups based on some criteria and then random samples are taken from each subgroup.
3. ***Systematic Sampling***: In this technique, the first sample is selected at random but the following samples are selected at a particular interval.

Central Limit Theorem (CLT)

According to CLT, the frequency distribution of any sample is normal distribution given that the sample is large enough (n>30), regardless of the frequency distribution of a population. Furthermore, the mean of the sample means is close to that of the mean of the population and the standard error = $\sigma/\sqrt{n}$ where σ = standard deviation of the population and n = sample size.

Estimation

Inferential Statistics allows us to make an estimation of the parameter of a population by calculating the statistic of the sample

Sampling Error: It tells us how much the calculated value differs from that of the stated value.

Sampling error = Population parameter – Sample statistic

Types of estimate

***Point estimate*:** makes estimation about the mean of a population. The drawback of point estimate is that it is impossible to calculate the mean of a population and hence, errors are inevitable.

Interval estimate: To overcome drawback of point estimate, an interval range is selected with point estimate as the mid-point. The mean can lie between the lower limit and the upper limit of the range.

Types of Inferential Statistics

Hypothesis Testing

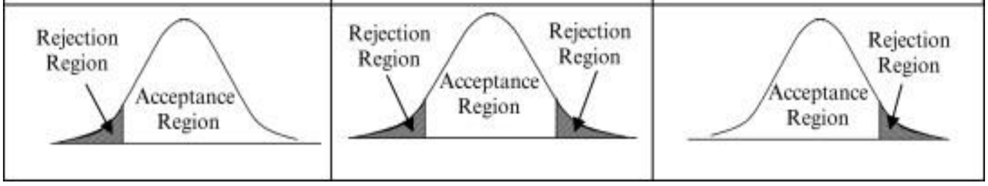
Hypothesis is an assumption made about the parameter of a population. It is of two types:

***Null Hypothesis* (H0):** This states that no relationship exists between the variables being studied.

***Alternative Hypothesis* (Hα):** This states that there is a relationship between the variables being studied.

Testing Methods:

One-Tailed Test (Left Tail)	Two-Tailed Test	One-Tailed Test (Right Tail)
$H_0 : \mu_X = \mu_0$ $H_1 : \mu_X < \mu_0$	$H_0 : \mu_X = \mu_0$ $H_1 : \mu_X \neq \mu_0$	$H_0 : \mu_X = \mu_0$ $H_1 : \mu_X > \mu_0$



Tailed Test Types

***Critical value approach*: It involves the following steps:**

a) State H0 and Hα
b) Define alpha. Compute test statics = (sample mean – population mean)/standard error.
c) Compute critical value.
    a. Negative for a right-tailed test.
    b. Positive for a left-tailed test.
    c. Either positive or negative for a two-tailed test depending on if the test static is positive or negative.
d) Compare test static with critical value.
    a. If test static > critical value for positive class, reject H0.
    b. If test static > critical value for negative class, accept Hα.

P-value approach: It involves the following steps:

a) State H0 and Hα.
b) Define alpha. Compute test statics = (sample mean – population mean)/standard error.
c) Compute p-value.
d) Compare p-value and alpha value.

    If p-value<alpha reject H0.

Confidence interval approach: It involves the following steps:

a) State H0 and Hα.

b) Define alpha. Compute test statics = (sample mean – population mean)/standard error.
c) Compute the population parameter confidence level.
d) If the μ lies in the interval, then accept H0, else, reject H0.

Testing types:

***Z Test*:** It is applied on the data that follows a normal distribution and has sample size > 30. It is used to test if the sample mean and population mean are equal given the population variance.

$$z = \frac{\bar{X} - \mu}{\frac{\sigma}{\sqrt{n}}}$$

If z statistic > z critical value, reject H0.

***T Test*:** It is applied to the data that follows a T-distribution and has a sample size < 30 and the population standard deviation is unknown.

$$t = \frac{\bar{X} - \mu}{\frac{s}{\sqrt{n}}}$$

If t statistic > t critical value, reject H0.

***F Test*:** It is applied to the data in order to check if there is difference between the variances of two samples.

For a right tailed f test:

H0: ${\sigma_1}^2 = {\sigma_2}^2$, where ${\sigma_1}^2 and\ {\sigma_2}^2$ is the variance of first and second population respectively.

Hα: ${\sigma_1}^2 > {\sigma_2}^2$

$f = \frac{{\sigma_1}^2}{{\sigma_2}^2}$

If f statistic > f critical value, reject H0.

.