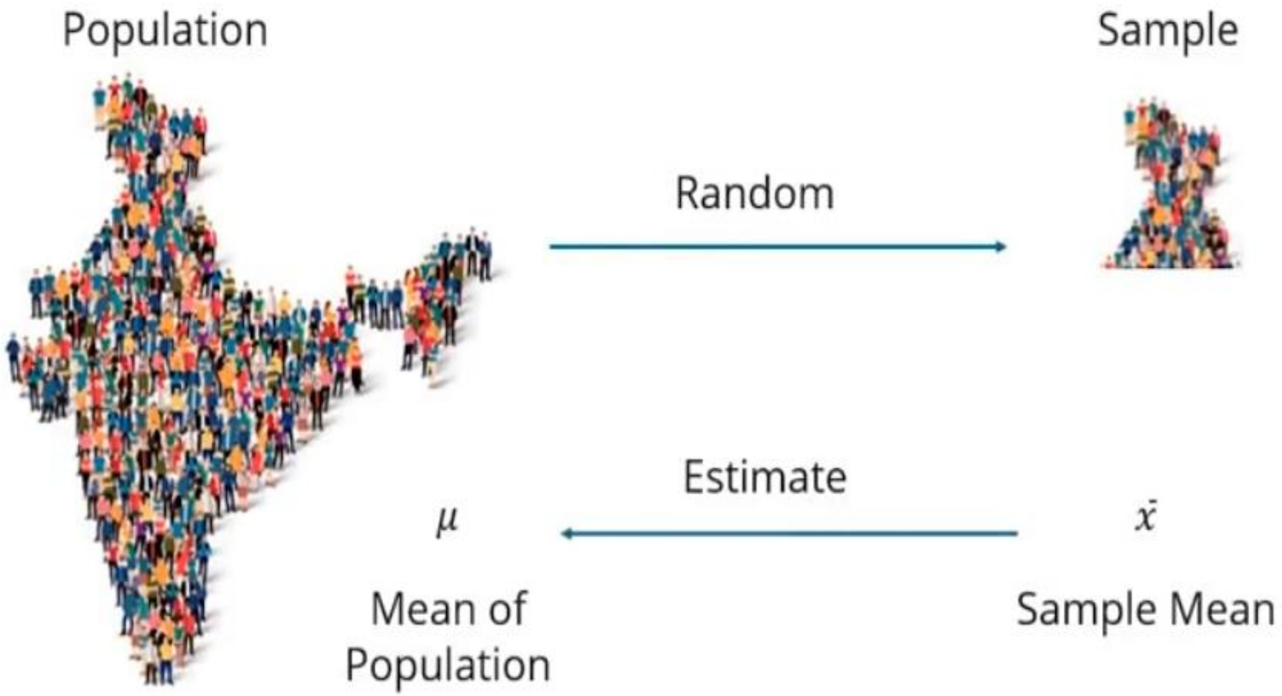
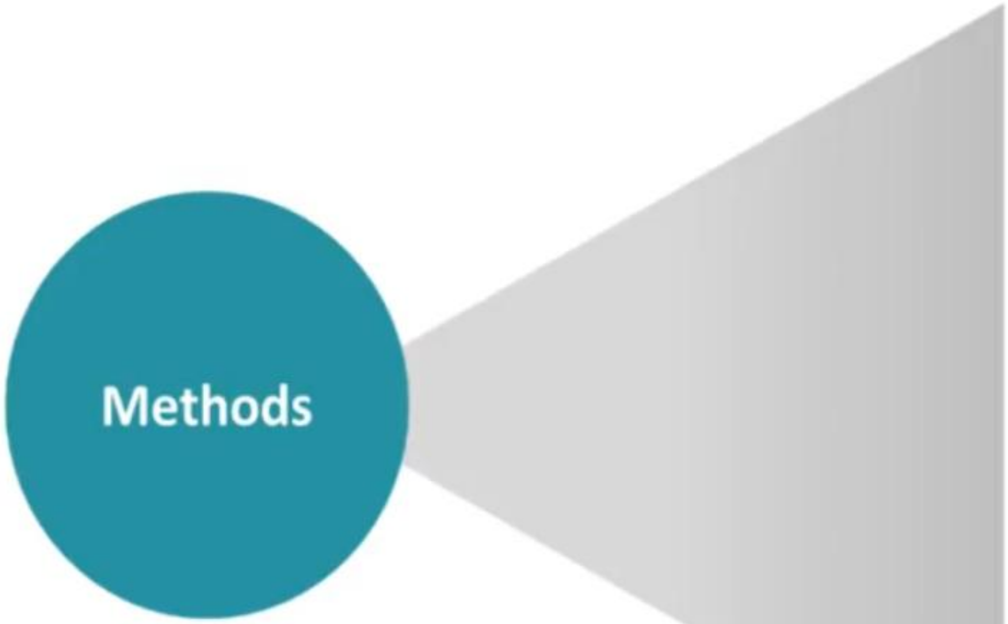
Inferential Statistics

Point Estimation

Point Estimation is concerned with the use of the sample data to measure a single value which serves as an approximate value or the best estimate of an unknown population parameter.



Finding the Estimates



Method of Moments
Estimates are found out by equating the first k sample moments to the corresponding k population moments

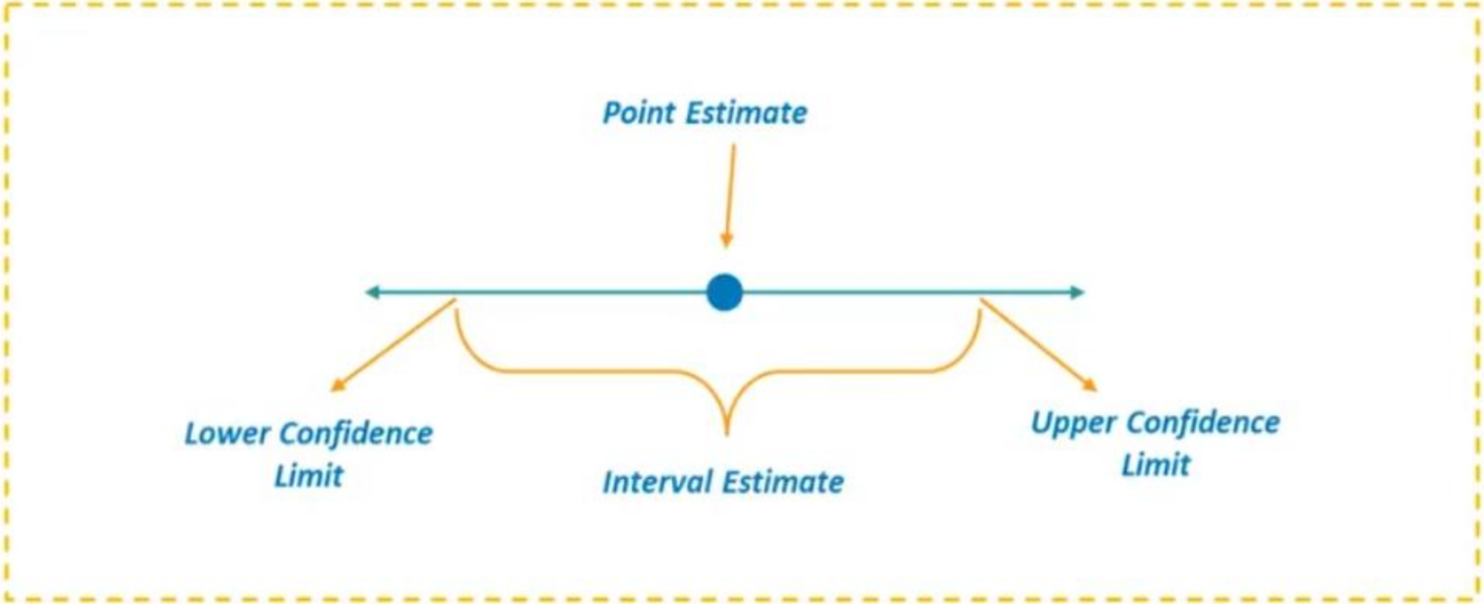
Maximum of Likelihood
Uses a model and the values in the model to maximize a likelihood function. This results in the most likely parameter for the inputs selected

Bayes' Estimators
Minimizes the average risk (an expectation of random variables)

Best Unbiased Estimators
Several unbiased estimators can be used to approximate a parameter (which one is "best" depends on what parameter you are trying to find)

Interval Estimate

An Interval, or range of values, used to estimate a population parameter is called Interval Estimate.



Confidence Interval

01 Confidence Interval is the measure of your confidence, that the interval estimate contains the population mean, μ

02 Statisticians use a confidence interval to describe the amount of uncertainty associated with a sample estimate of a population parameter

03 Technically, a range of values so constructed that there is a specified probability of including the true value of a parameter within it

Margin of error

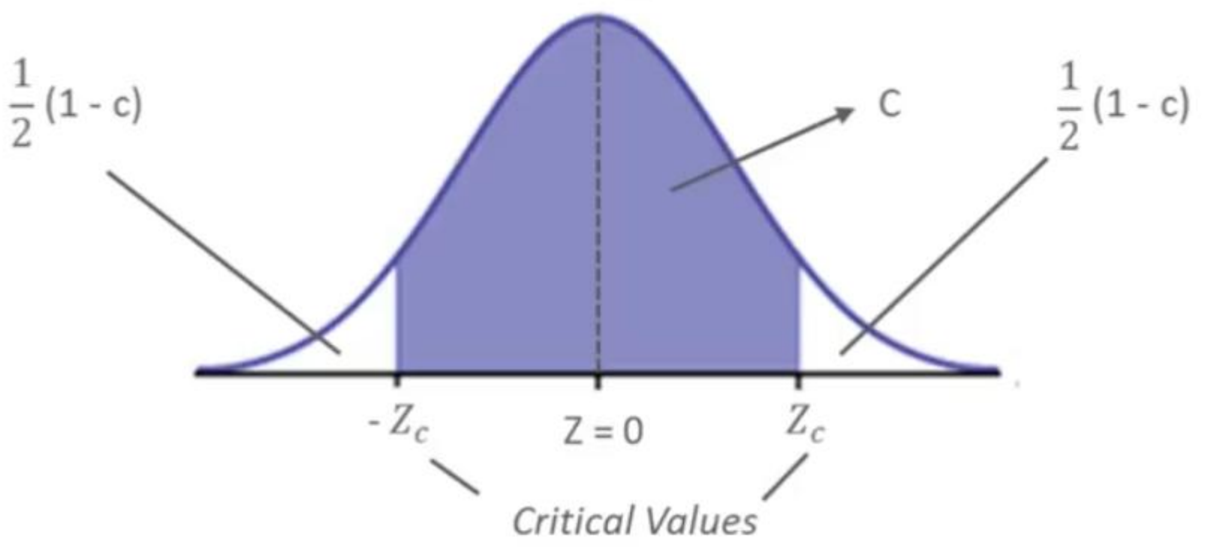
- Difference between the point estimate and the actual population parameter value is called the **Sampling Error**
- When μ is estimated, the sampling error is the difference $\mu - \bar{x}$

Margin of Error E, for a given level of confidence is the greatest possible distance between the point estimate and the value of the parameter it is estimating

$$\boxed{E = Z_c \frac{\sigma}{\sqrt{n}}}$$

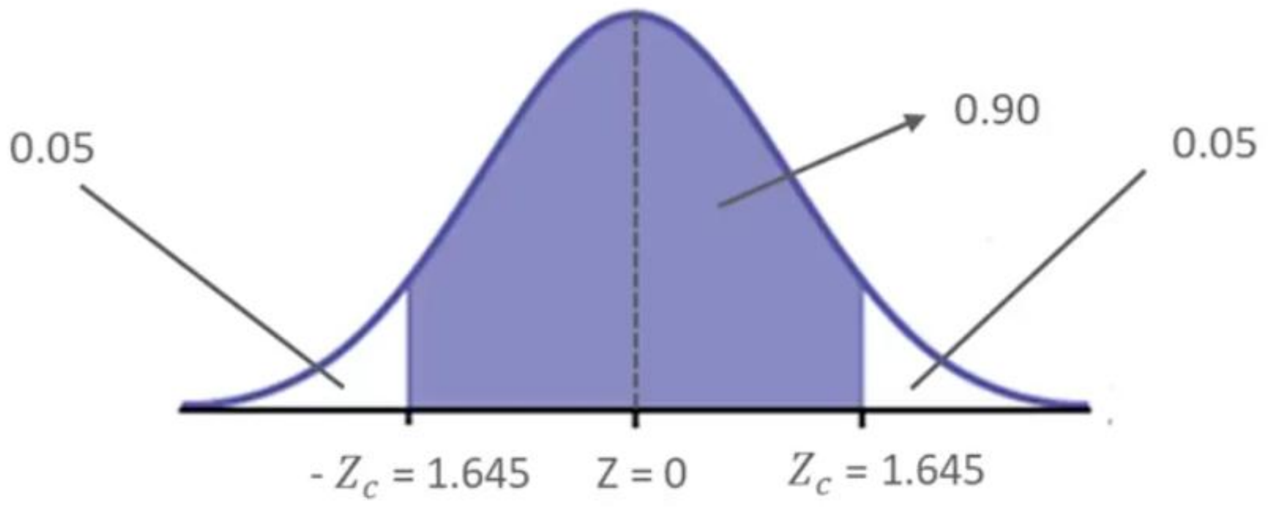
Estimating Level of Confidence

The level of confidence c, is the probability that the interval estimate contains the population parameter.



C is the area beneath the normal curve between the critical values
Corresponding Z score can be calculated using the standard normal table

If the level of confidence is 90%, this means that you are 90% confident that the interval contains the population mean, μ.



The Corresponding Z – scores are ± 1.645

Margin of Error-Use Case

A random sample of 32 textbook prices is taken from a local college bookstore. The mean of the sample is $\bar{x}$ = 74.22, and the sample standard deviation is S = 23.44. Use a 95% confidence level and find the margin of error for the mean price of all textbooks in the bookstore

You know by formula,

$$\boxed{E = Z_c \frac{\sigma}{\sqrt{n}}}$$

E = 1.96 * (23.44/√32) ≈ 8.12

Hypothesis Testing

Statisticians use hypothesis testing to formally check whether the hypothesis is accepted or rejected.

Hypothesis testing is conducted in the following manner:

- ❖ **State the Hypotheses –** This stage involves stating the null and alternative hypotheses.
- ❖ **Formulate an Analysis Plan –** This stage involves the construction of an analysis plan.
- ❖ **Analyse Sample Data –** This stage involves the calculation and interpretation of the test statistic as described in the analysis plan.
- ❖ **Interpret Results –** This stage involves the application of the decision rule described in the analysis plan.

Example-Hypothesis Testing

Assume the event is free of bias.

So, what is the probability of John not cheating?

P(John not picked for a day) = $\frac{3}{4}$

P(John not picked for 3 days) = $\frac{3}{4} \times \frac{3}{4} \times \frac{3}{4}$ = 0.42 (approx)

P(John not picked for 12 days) = $(\frac{3}{4})^{12}$ = $\mathbf{0.032 < 0.05}$

Null Hypothesis (H_0) : Result is no different from assumption.

Alternate Hypothesis (H_a) : Result disproves the assumption.

Probability of Event < 0.05 (5%)