Frequently Asked Questions

1. What do you mean by statistical inference?

Answer:
One of the prime techniques of Statistics is sampling whose object is to study the features of the population on the basis of the sample observations. A careful selection of the sample is expected to reveal these features and hence we shall infer about the population from a statistical analysis of the sample. This process is known as statistical Inference. Hence, by statistical inference we mean that certain conclusions drawn about certain parameters of the population under consideration based on the study of sample observations.

2. What are the two types of problems in Statistical Inference?

Answer:
- Estimation
- Test of hypothesis

Firstly, we may have no information at all about some characteristics of the population, especially the values of the parameters in the distribution and it is required to obtain the estimates of these parameters. The true parameter will be unknown and one objective of sampling could be to estimate its value.This is a problem of Estimation.

Secondly, some information or hypothetical values of the parameters may be available and it is required to test how far the hypothesis is tenable in the light of information provided by the sample. This is the problem of test of hypothesis or Tests of Significance Given point estimator(s) from samples, we may wish to infer about the reproducibility of results, or if any statistical differences exist.

3. Explain two different ways of estimation in statistical Inference.

Answer:
The estimation is done in two different ways:

- Point Estimation
- Interval Estimation

In Point Estimation the estimated value is given by a single quantity which is the function of a sample observations (that is a statistic). This function is called the estimator of the parameter and the value of the estimator is called an estimate.

In Interval Estimation, an interval within which the parameter is expected to lie is given by using two quantities based on the sample values. This is known as a confidence Interval.

4. Explain briefly point estimate and point estimation.

Answer:
A number of functions or statistic (such as mean, median, minimum of observations, etc) can be defined based on the random samples drawn from a population which involves an

unknown parameter θ and any one of them can be taken as an estimator of the unknown parameter θ.

Each value of each functions is used in estimating the parameter is the point estimate of θ . In general an estimate of an unknown parameter is called a point estimate. Estimator of a population parameter is a function of the sample information that yields a single number. The corresponding realization is called the point estimate of the parameter. That means when we provide a single numerical estimate of the population parameter (θ) based on the sample information it is known as point estimation.

5. Give one example for point estimate.

Answer:

Statisticians use sample statistics to estimate population parameters. For example, sample means are used to estimate population means; sample proportions, to estimate population proportions.

As an example of a point estimate, assume we wanted to estimate the mean time it takes 12-year-olds to run 100 yards. The mean running time of a random sample of 12-year-olds would be an estimate of the mean running time for all 12-year-olds. Thus, the sample mean, $\bar{x}$, would be a point estimate of the population mean, μ.

6. Name the criteria that have to be satisfied by the point estimator to be called as a best estimator.

Answer:

Statisticians have developed some criteria, which they find useful in deciding which estimator might be the more advantageous in specific circumstances.

According to R. A. Fisher the criteria for a best estimator are:

- Unbiasedness
- Consistency
- Efficiency
- Sufficiency

7. What are the different methods of point estimation?

Answer:

Point estimation refers to the process of estimating a parameter from a probability distribution, based on observed data from the distribution. There are many methods available for the estimation of the population parameters using Point Estimation method. Some of them are:

- Method of maximum Likelihood
- Method of Moments
- Method of minimum variance
- Method of Chi-square
- Bayesian's estimators etc

8. What do you mean by consistent and sufficient point estimators?

Answer:

Consistent estimators

An estimator is a sample statistic. i.e. it is a function of the sample elements. For many estimators variance of the sampling distribution of an estimator decreases as sample size increases. We would like that estimator stays as close as possible to the parameter it estimates as sample size increases. Such an estimator is said to be a consistent estimator.

The property of consistency is a limiting property.

Sufficient estimators

An estimator is sufficient if it makes so much use of information in the sample that no other estimator could extract from the sample additional information about the population parameter being estimated.

9. What is the role of a sample mean in statistical estimation?

Answer:

In general, we use statistics as a means of characterizing the nature of some sample based on a few key indicators.

The first indicator is known as The Sample Mean:

- The mean quantity in some sample represents the average value or the most probable value in the sample.
    - A sample mean is calculated by summing up the individual measurements and dividing by the number of measurements, usually denoted as N.
    - All samples can be characterized by a mean value regardless of the shape of the distribution.

Also according to Central Limit Theorem: The distribution of means of random sample taken from a population having mean μ and finite variance σ^2 approaches Normal distribution with mean μ and variance σ^2/n as n goes to infinity.

10. Briefly explain unbiased and efficient point estimators.

Answer:

Unbiased estimators

If an estimator *tn* estimates θ then difference between them

(tn-θ) is called the estimation error. Bias of the estimator is defined as the expectation value of this difference

$B(\theta) = E(tn-\theta) = E(tn) - \theta$.If the bias is equal to zero then the estimation is called unbiased. For example sample mean is an unbiased estimator:

Efficient estimators

Estimators whose sampling distributions have smaller variances are considered superior. The idea here is that different random samples will give different values of the estimator, consistent with the sampling distribution. An unbiased point estimator with least variance is said to be an efficient estimator.

11. List some point estimators that are used in estimation of the parameters of the population?

Answer:

Some of the potential point estimators:

	population parameters (θ)	**point estimators ($\hat{\theta}$)**
mean	μ	$\bar{x}, \bar{x}$
variance	σ^2	s^2
standard deviation	σ	s
proportion	Π(P)	p
correlation coefficient	ρ	r
linear regression	β_0, β_1	b_0, b_1
difference between two means	$\mu_2 - \mu_1$	$\bar{x}_2 - \bar{x}_1$
difference between two proportions	$\pi_2 - \pi_1$	$p_2 - p_1$

12. Explain the different types of point estimators used for various types of variables.

Answer:
The point estimate we choose depends on the "nature" of the outcome of interest. Some of the point estimates suggested for different types of variables are:

- Continuous Variables
  - Examples: change in tumour volume or tumour diameter
  - Commonly used point estimates: mean, median
- Binary Variables
  - Examples: response, progression, > 50% reduction in tumour size
  - Commonly used point estimate: proportion, relative risk, odds ratio

- Time-to-Event (Survival) Variables
  - Examples: time to progression, time to death, time to relapse
  - Commonly used point estimates: median survival, *k*-year survival, hazard ratio
- Other types of variables: nominal categorical, ordinal categorical
  - Mode for nominal and median for ordinal categorical variables

13. Explain the applications of point estimators.

Answer:

Often point estimates are used as parts of other statistical calculations. For example, a point estimate of the standard deviation is used in the calculation of a confidence interval for μ. Point estimates of parameters are often used in the formulas for significance testing.

Point estimates are not usually as informative as confidence intervals. Their importance lies in the fact that many statistical formulas are based on them.

Nevertheless, the convenience of point estimates outweighs their deficiencies in some instances. Two examples are:

i. Interval estimates of certain fundamental physical constants would be very difficult or inconvenient to work with in calculations. Thus, although quantities such as the gravitational acceleration, g; Avogadro's number, N; and so forth are numbers, which are experimentally determined and thus subject to sampling errors of one sort or another, we normally use point estimates of them rather than interval estimates. Of course, many of these fundamental constants have been estimated with high precision so that errors in their estimates are not significant for many applications.

ii. As you'll see shortly, when we carry out various procedures of statistical inference focusing on one population parameter of greatest interest, the formulas that result may involve the values of other population parameters. In such situations, we can usually obtain adequately accurate results by using point estimates for the parameters of secondary interest in order to derive formulas for an interval estimate of the parameter of greatest interest.
For example, in deriving formulas for interval estimates of the population mean, μ, we require the value of the population standard deviation, σ. Since μ is unknown, it is very unlikely that we will know the value of σ (though in some instances we might). Rather than backing up one more step and determining an interval estimate for σ, it is more usual to use the available value of s as a point estimate of σ in the formula for the interval estimate of μ.

14. What are the limitations of point estimation?

Answer:

A point estimate is often insufficient because it is either right or wrong. If we are told only that our point estimator of any parameter is wrong and we don't know how wrong it is and we cannot be certain of the estimate's reliability. Therefore, a point estimate is much more useful if it is accompanied by an estimate of the error that might be involved.

It is obvious that a point estimate is normally different from the actual value of the parameter. The reason is that the point estimate is derived from the random sample whose value vary

from a sample to sample where $\bar{x}$ is the estimator of the mean μ , the values of $\bar{x}$ will vary in a manner that most of them are spread about closely on both sides of μ.

As the parameter to be estimated is unknown, neither the error in the point estimate evaluated noted nor it is accurately measured. This greatly reduces he practical utility of point estimation. Accuracy can be expressed in probabilistic terms by stating how likely or probable it is that a particular value of an estimator $\hat{\theta}$ is equal to the parameter θ to be estimated This necessitates that an estimator should be expressed in the form of an interval rather than a single numerical value

In practice, confidence interval estimates are used more commonly by far than point estimates. Nevertheless, since point estimates are used in certain important ways in statistics, and carry with them some important concepts and terms, we need to look at them briefly

Some people refer to point estimates as a "best guess" value. The term "guess" is a bit pessimistic, but it does give you the sense that there is a degree of uncertainty in relying on point estimates.

15. What are the two important points to be considered to get a best estimator of the population parameter?

Answer:

- The sampling distribution of the point estimator should be centered over the true value of the parameter to be estimated.
- The spread (as measured by the variance) of the sampling distribution should be as small as possible.