**Statistical Inference**

* Statistical inference is concerned with decision making based on observed data, using probability concepts to deal with uncertainty.
* Given a specific probability distribution, we can calculate the probabilities of various events. For example, knowing that ~ , we can calculate . Similarly, knowing that ~ , we can calculate .
* Roughly speaking, statistical inference is concerned with the opposite sort of problem. For example, knowing that ~ , where the value of is unknown, and having observed (say = 32), what can we say about ? Similarly, knowing that ~ , where the value of is unknown, and having observed (say = 105), what can we say about ?

# **Types of Statistical Inference**

* **Estimation:** Parameter(s) of a population may be unknown and it may be necessary to make a guess about them on the basis of a sample. This type of problem is known as estimation problem. The theory in this connection is known as theory of estimation.
* ***Point estimation:*** Point estimation concerns with choosing a single number/value calculated from the sample observations, which will be representative to the unknown parameter. In other words, in case of point estimation we try to find a function of the sample observations, which may be taken as a measure of the estimate of the population parameter.
* ***Interval estimation:*** Interval estimation concerns with choosing an interval of values calculated from the sample observations, within which the unknown parameter will lie with certain confidence level. In other words, in case of interval estimation we try to find two functions of the sample observations such that it isvery likely that the interval contained between the two functions will include the parameter.
* **Test of Hypothesis:** A procedure, based on sample evidence and probability theory, used to determine whether the hypothesis is a reasonable statement and should not be rejected, or is unreasonable and hence should be rejected. The theory in this connection is called as theory of testing of hypothesis.

**Hypothesis:** A statement about the value of a population parameter developed for the purpose of testing.

Examples of hypotheses made about a population parameter are:

* The mean monthly income for systems analysts is Rs. 45000.
* The mean weight of new born baby is 3.3 kg.
* Twenty percent of all juvenile offenders are caught and sentenced to prison.

***Illustratioin***:

An industrial plant emits SO2 into the atmosphere. To know its contribution to polluting the atmosphere, daily emission of SO2 (in tons) for 70 consecutive days were recorded. The data is given below:

|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| 13.2 | 19.4 | 18.0 | 23.9 | 26.1 | 15.5 | 21.6 | 17.9 | 14.5 | 14.4 |
| 23.7 | 11.8 | 22.9 | 17.5 | 20.9 | 19.2 | 16.2 | 17.3 | 18.1 | 8.3 |
| 10.7 | 26.4 | 15.9 | 11.0 | 21.4 | 27.5 | 22.7 | 6.2 | 31.8 | 25.9 |
| 19.0 | 9.8 | 7.7 | 20.4 | 22.3 | 13.5 | 26.8 | 26.6 | 28.5 | 18.1 |
| 24.6 | 22.7 | 11.2 | 17.0 | 9.4 | 13.9 | 19.1 | 20.1 | 13.3 | 23.5 |
| 24.3 | 15.2 | 14.7 | 19.3 | 18.7 | 28.6 | 21.5 | 22.5 | 24.1 | 23.0 |
| 15.8 | 12.3 | 20.5 | 24.8 | 12.8 | 19.4 | 16.9 | 16.7 | 18.5 | 29.6 |

State the types of inference you are to make in answering the following three questions:

1. What is the average daily emission of SO2 from the plant and what is the s.d. of daily emission of SO2 from the plant?
2. What is the interval of values within which the average daily emission of SO2 will lie?
3. Is the average daily emission of SO2 equals to 20.1 tons, which was found to be the average two years ago?

# **Few important terms**

**Sample**: A sample is a subset of a population. Typically, the population is very large; making a census or a complete enumeration of all the values in the population is impractical or impossible. The sample represents a subset of manageable size.

**Statistic**: A statistic (singular) is any quantity computed from values in a [sample](https://en.wikipedia.org/wiki/Sample_(statistics)) which is considered for a statistical purpose. Statistical purposes include *estimating a*[*population*](https://en.wikipedia.org/wiki/Statistical_population)*parameter*, *describing a sample*, or *evaluating a hypothesis*. For examples, the [average (or mean)](https://en.wikipedia.org/wiki/Arithmetic_mean) of sample values is a statistic.

When a statistic is being used for a specific purpose, it may be referred to by a name indicating its purpose. For examples, a [*descriptive statistic*](https://en.wikipedia.org/wiki/Descriptive_statistic) is used to summarize the sample data and a [*test statistic*](https://en.wikipedia.org/wiki/Test_statistic) is used in [statistical hypothesis testing](https://en.wikipedia.org/wiki/Statistical_hypothesis_testing). Again, a single statistic can be used for multiple purposes, e.g. the sample mean can be used to estimate the population mean, to describe a sample data set, or to test a hypothesis.

* A statistic is calculated by applying a function to the values of the items comprising the sample. Thus, in statistical theory, a **statistic** is defined as ***a function*** of the random sample***.***
* For examples, , etc. are statistic. , etc. summarize the information contained in the sample observations/data and therefore, in essence, a statistic is a summarization of data.
* The term *statistic is used both* ***for the function and for the value of the function*** on a given sample.

**Estimator**: An estimator is a statistic (that is, a function of the sample observations) that is used to infer about the value of an unknown parameter of the population from which the sample is collected. The sample mean, can be an estimator of the population mean *µ*, and the sample proportion can be used as an estimator of the population proportion.

**Estimate**: The computed specific numerical value of an estimator, obtained based on a particular set of sample observations, is known as the estimate of the parameter. The estimate of a parameter is denoted by.

###### Characteristics of a good Point Estimator

* Unbiased
* Minimum-Variance and Efficiency
* Consistency
* Sufficiency

**Unbiased Estimator:**

An estimator (i.e. statistic) is called an unbiased estimator of a parameter if . In other words, if and only if the mean of the sampling distribution of is equal to .

Bias of an estimator is the difference between the estimator's expected value and the true value of the parameter being estimated. Thus, an estimator with zero bias is called unbiased. Otherwise the estimator is said to be biased.

* **Let is a random sample from . and are selected as the estimators of and respectively. Are they unbiased estimator?**

This implies is unbiased estimator of.

This implies that is a biased estimator of .

* **Is unbiased estimator of ?**

This implies that is unbiased estimator of .

* **Suppose from a normal population a sample of size 2 is collected. Say, and are two estimators of the population mean (). Are they unbiased estimator of ? Which one would you prefer to use as an estimator of?**

This implies that both the estimators are unbiased. However, to answer the second question we need to know the variances of the two estimators.

Now,

This implies that is less than . Therefore, is preferable as an estimator of.

**Minimum Variance Unbiased (MVU) Estimator**

A statistic will be called a minimum variance unbiased (MVU) estimator of if it is unbiased and has the smallest variance among all unbiased estimators of. Thus is an MVU estimator of if

, and

, whatever the true value of may be

where is any other unbiased estimator of .

Cramer has used the term ‘efficient estimator’ to mean a minimum-variance unbiased estimator. All other unbiased estimators may then be called inefficient estimators. The efficiency of such an inefficient estimator is then defined as the ratio of the variance of an MVU estimator () of to variance of the unbiased estimator . Thus,

Efficiency of ,

Obviously, , the value unity occurring only if is also of minimum variance. An efficient estimator has less variability and so we are more likely to make an estimate close to the true parameter value.

* **In estimating the mean of a normal population on the basis of a random sample of size 2*n*+1, what is the asymptotic efficiency of median with respect to the mean?**

Let and be respectively the mean and median of the random sample. Then,

and

and , for large

Relative efficiency of median, is

and the asymptotic efficiency is

.

Thus, for large sample, the sample median is only 64% as efficient as the sample mean. Or, take the inverse (1/0.64) and we can state that the sample mean is 56% more efficient than the sample median.

**NOTE: An MVU estimator is unique, in the sense that if both *T*0 and *T*1 are MVU estimators, then *T*0 = *T*1 almost everywhere.**

An efficient estimator has high degree of concentration around and it converges to most rapidly. No universally satisfactory measure for the speed of stochastic convergence is available. If, however, we confine ourselves to the estimators that are asymptotically normally distributed, then the speed of stochastic convergence may be very well measured **by the reciprocal of the variance** of the asymptotic distribution. In such a case, the estimator **having least sampling variance** is called the most efficient estimator.

***Note that***

* *An estimator may be directly examined for biasness*
* *However, it is not immediately apparent how to satisfy oneself that an estimator has the smallest variance among all unbiased estimators.*
* *The most widely used method to achieve this task is based on the use of Cramer-Rao inequality, which provides the lower bound of the variance of all the estimators of under some conditions, called regularity conditions.*

**Regularity conditions for the Cramer-Rao Inequality**

Let is the p.d.f of and is the single unknown parameter. Suppose, are the random sample of size collected from the population and suppose be an unbiased estimator of . Then the regularity conditions for the Cramer-Rao inequality are stated as follows:

* exist
* , (i.e. range of does not involve )
* , (i.e. does not involve )
* exists and is positive

Any situation where the conditions hold is called a regular estimation case.

**Cramer-Rao Inequality Theorem**

In any regular estimation case, the variance of an unbiased estimator for , which is assumed to be differentiable function of, satisfies the following inequality:

where , the joint p.d.f. of

, when are iid

Therefore,

### An alternative form of Cramer-Rao Inequality is

### 

**Note that**

* The denominator in the C-R Lower Bound, i.e. or is **denoted by and is known as Fisher Information**, which is a measure of the amount of information that the random sample gives about the unknown parameter .
* If is an unbiased estimator of parameter (not a function of ), then

So, in such a case C-R Lower Bound is simply .This explains why is called “information”: the larger the value of , smaller is the C-R Lower Bound and consequently smaller is the variance, and therefore, we would be more certain about location of the unknown parameter value.

* In caseof distributions with more than one parameter, we will have Hessian, which is the matrix of second derivatives of the log likelihood function with respect to the parameters in the denominator of C-R inequality. So in such cases, Fisher information takes the form of a matrix and is known as Fisher Information Matrix.

**Minimum variance bound (MVB) estimator**

An unbiased estimator *T* of , for which the Cramer-Rao Lower Bound is attained may be called a minimum variance bound (MVB) estimator, i.e. if C-R Lower Bound.

It is important to note that the MVB estimator is not the same as MVU estimator or efficient estimators as used by Cramer, *because the least attainable variance of a MVU estimator may be greater than the Cramer-Rao Lower Bound.*

* **Suppose be a random sample from , known. Show that is an MVB estimator for.**

Here, , the parameter for which MVB is to be estimated. We know that . Therefore, is an unbiased estimator of , i.e. .

Now,

, where  is a constant

Hence,

And

Also, . Therefore, C-R Lower Bound for the variance of the estimator of is .

Again, we know that , which coincides with the C-R Lower Bound for the estimator of the parameter .

Hence, is an MVB estimator for .

* **Suppose be a random sample from, known. Show that is an MVB estimator of .**

Here, , the parameter for which MVB is to be estimated. We know that . Therefore, is an unbiased estimator of , i.e. .

Now, , since

, where  is a constant

Hence,

Therefore, C-R Lower Bound for the variance of the estimator of is

Now, and , i.e.

Also it is known that and

Thus, is an unbiased estimate of and its variance coincides with the C-R Lower Bound for, i.e. . Hence, is an MVB estimator of .

#### **Consistency of estimators**

An estimator is said to be consistent if its accuracy (probability of estimates close to the value of the population parameter) increases with the increase in the sample size. This means that the probability

for any given , should be an increasing function of . For a consistent estimator, the sampling distribution becomes concentrated on the value of the parameter it is intended to estimate as the sample size approaches infinity.

**Sufficient conditions for consistency**

Let a sequence of estimators, say , is generated from the original estimator by varying . Let the sequence be such that for each , the expectation and variance of exist and

, and

as

Then is consistent estimator for.

**Note: If be a consistent estimator for and a continuous function of. Then is a consistent estimator of .**

# **Sufficiency of estimator**

An estimator is said to be sufficient (i.e. sufficient statistic) for a parameter, if it contains all the information in the sample regarding the parameter. More precisely, a statistic is said to be sufficient for estimating , if the conditional distribution of the sample for given is independent of.

**Sufficient conditions for sufficiency**

A necessary and sufficient condition for to be a sufficient statistic for is that the joint density function for can be expressed in the following form:

where is a function of . The function contains and in the form of only, while the function does not contain. This is known as Neyman’s Factorization Principle.

The above expression can be written alternatively as

or,

It may be noted that in order that may have as an MVB estimator, must be a sufficient statistic for.

Therefore, when a sufficient statistic exists,one has to find out a suitable estimator from the functions of the sufficient statistic only.

**Note:** Let the statistic is sufficient for the parameter and let be a function with *one-to-one correspondence* with . Then, is also sufficient for.

* **Let be a random sample of size from a Poisson population. Is a sufficient statistic for estimation of?**

Thus, has been broken into two factors – one contains in the form of and parameter , whereas the other is free from the parameter . Therefore, is a sufficient statistic for estimating.

* **Show that, if is known in a random sample from a normal population, is a sufficient estimator for ; but is not a sufficient estimator of , if  is known. Suggest a sufficient estimator for .**

Here,

[Since,

]

Thus, has been broken into two factors – one contains in the form of and parameter , whereas the other is free from the parameter . Therefore, is sufficient estimator of, when is known.

On the other hand, if is known then while the second factor contains both (the estimator) and (the parameter); but the first factor too contains (the parameter). So, is not a sufficient estimator of .

However, if we define , then we find that

where, and

Here, the first term contains both (the estimator) and (the parameter) while the second term is independent of . So, as defined above is a sufficient estimator of . 

* **Show that sample proportion is sufficient estimator of the population parameter in a Point Binomial distribution.**

The joint pmf of in a set of Bernoulli trial,

(a function of sample proportion and population proportion only)

This implies that the necessary and sufficient conditions for sufficiency are satisfied. Therefore, sample proportion is a sufficient estimator for the population parameter .

* **Show that , where is i.i.d. , is unbiased, consistent, efficient and sufficient estimator of µ.**

is unbiased estimator of.

* as

as

is consistent estimator of .

or,

or,

C-R lower bound (since )

Now, , which is equal to C-R lower bound.

is an efficient estimator for .

where, and

is sufficient statistic for .