International Journal for Multidisciplinary Research (IJFMR)
E-ISSN: 2582-2160 ● Website: www.ijfmr.com ● Email: editor@ijfmr.com

A Comparative Study of Theory of Estimation and Testing of Hypothesis in Statistical Inference

Dr. Mohinder Pal

Associate Professor, Govt. Degree College Udhampur

Abstract

Statistical inference is an important branch of statistics concerned with drawing conclusions about a population on the basis of sample information. Two fundamental components of statistical inference are the Theory of Estimation and the Testing of Hypothesis. Estimation is concerned with determining unknown population parameters from sample data, whereas hypothesis testing is concerned with making decisions about population characteristics by assessing evidence against a stated assumption. This paper presents a comparative study of estimation and hypothesis testing, including their concepts, methods, important properties, advantages, disadvantages and relationship within statistical inference.

Keywords: Statistical inference, estimation, point estimation, hypothesis testing, null hypothesis, alternative hypothesis, significance level, statistical decision.

1. INTRODUCTION:

The Statistical inference may be termed as the logic of drawing statistically valid conclusions about the population parameters on the basis of sample drawn from it in a scientific manner. In this paper, we shall develop the technique which enables us to generalise the results of the sample to the population; to find how far these generalisations are valid and also to estimate the population parameters along with degree of confidence. In statistical terminology, this learning is termed as statistical inference which is of two kinds; estimator and hypothesis testing. Both types of statistical inference utilise the information provided by the sample, for drawing some conclusions about the parameters of the population; yet each type of inference uses this information in different ways. The theory of estimation is divided into two parts: point estimation and interval estimation. The aim of point estimation is obtain a single value which is the best guess of the parameter interest. In interval estimation the object is to obtain interval within which the true value of the parameter may be said to lie with some given level of probability which expresses the confidence we have that the value lies within the stipulated range. In this paper, in testing of hypothesis we have to learn how to use samples to decide whether a population possesses a particular characteristics and to understand the basis of testing procedure, its types, one tailed and two tailed test and also the concept of critical region and P-values.

2. POINT ESTIMATION—CRITERIA FOR GOOD ESTIMATORS

There are various methods with which we may obtain point estimation or point estimates of the parameters of the phenomena under study. There is naturally a problem of choosing the one which gives us the best estimate. Also, how are we to decide whether any estimate is the best or whether it is good or better than another obtained by a different method? That is; we need to devise a criterion to call an

estimator a best one. We, therefore, have to do two things: (a) to specify various properties of an estimator that go to make it a best estimator and, (b) to devise different methods that could give rise to estimators that possess at least some of these desirable properties. The desirable properties or the main criteria for a good estimator obtained from small samples according to R.A. Fisher ;

1. **Unbiasedness,**
2. **Consistency,**
3. **Efficiency,**
4. **Sufficiency**.

2.1 Unbiasedness:

A statistic t is said to be an Unbiased Estimator of a parameter θ, if the expected value of t is.

$E(t)= \theta$

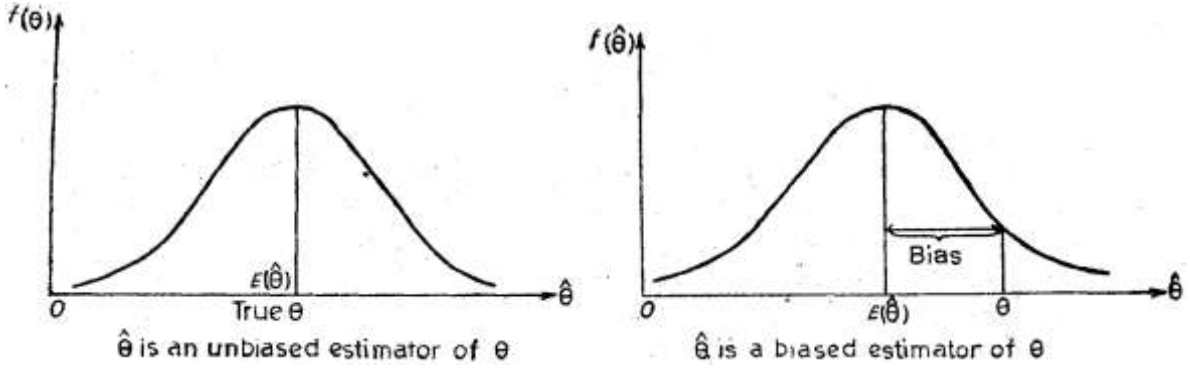
Otherwise, the estimator is said to be 'biased'. The bias of an estimator is defined as the difference between its expected value and the true value of the parameter. Mathematically, the bias of a statistic in estimating θ is given as

Bias= $E(t) - \theta$

If $E(t) - \theta > 0$ t is said to be positively biased.

$E(t) - \theta < 0$ t is said to be negatively biased

When the bias is positive, that is, when the mean value of the distribution is larger than its parameter, then the estimator is said to be upward biased. Conversely, when the bias is negative, the estimator is biased downwards.



Since the distribution is assumed to be a symmetric one, the mean is shown at the centre of the distribution, and it is equal to the true value of the parameter in Figure on the left hand side; but is not equal to the true value of the parameter in Figure on the left hand side

2.2 Consistency:

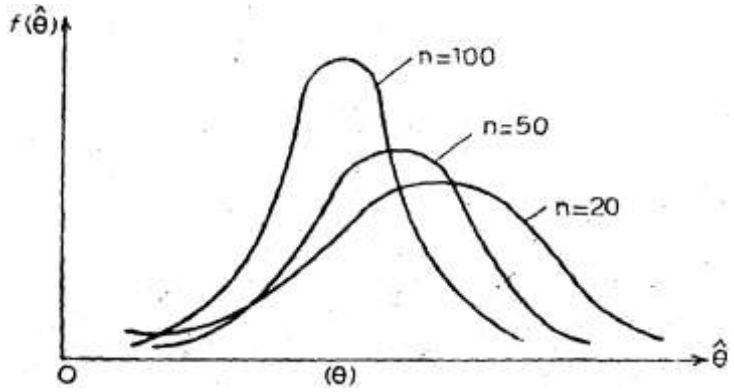
A statistic t computed from a sample of n observations is said to be a Consistent Estimator of a parameter θ, if it converges in probability to 0 as n tends to infinity. This means that the larger the sample size (n), the less is the chance that the difference between t, and θ will exceed any fixed value. Given any arbitrary small positive quantity $\in$,

$$\underset{n\to\infty}{Lt}\ P\{|t_n-\theta|>\in\}=0$$

$$If\ E[t_n]\to\theta \quad and \quad Var[t_n]\to 0\ as\ n\to\infty\ then\ t_n\ will\ be\ a\ consistent\ estomator\ of\ \theta$$

To see whether an estimator is consistent, we should therefore examine its bias and variance as sample size is increased. If both bias and variance decrease as n becomes larger, and at the limit (as $n\to\infty$)

both become zero, then estimator is assumed to possess the property of consistency. This is illustrated in Figure which is given below, which shows that as the sample size increases from 20 to 100 observations both bias of and its variance decrease.



Since sum of squared bias and variance is equal to the MSE, the disappearance of the bias and variance as $n \to \infty$ is equivalent to the disappearance of the MSE, so that we can also say;

A statistic t is said to be a consistent estimator of a parameter θ, if $\underset{n\to\infty}{Lt}\ MSE[t] = 0$ i.e. in sampling from a Normal population $N(\mu, \sigma^2)$ both the sample mean and the sample median are consistent estimators.

2.3 Efficiency

This criterion based upon the variances of the sampling distributions of the estimators which enable us to choose between the estimators with the common property of consistency usually known as **efficiency.** Of two consistent estimators for the same parameter, the statistic with the smaller sampling variance is said to be "more efficient". Thus if t and t' are both consistent estimators of θ, and

$$Var(y) < Var(t')$$

then t is 'more efficient' than t' in estimating θ,

If a consistent estimator exists whose sampling variance is less than that of any other consistent estimator, it is said to be **"most efficient"**; and it provides a standard for the measurement of 'efficiency' of a statistic. If V_0 be the variance of the most efficient estimator and V be the variance of any other estimator, then the **efficiency** of the estimator is defined as

$$Efficiency = \frac{V_o}{V}$$

Obviously, the measure of efficiency cannot exceed 1.

2.4 **Sufficiency**

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

A statistic is said to be a 'sufficient estimator' of a parameter θ, if it contains all information in the sample about θ If a statistic t exists such that the joint distribution of the sample is expressible as the product of two factors, one of which is the sampling distribution of t and contains θ, but the other factor is independent of θ, then t will be a sufficient estimator of θ

Thus if $x_1, x_2, \ldots\ldots x_n$ is a random sample from a population whose probability function is $f(x.\theta)$ an d t is sufficient statistic for the estimation of θ, we can write

$f(x_1.\theta), f(x_2.\theta), f(x_3.\theta)\ldots\ldots\ldots\ldots, f(x_n.\theta) = g(t.\theta), h(x_1, x_2.x_3\ldots\ldots\ldots, x_n)$

Where $g(t.\theta)$, is the sampling distribution of t and contains only θ, but $h(x_1, x_2.x_3\ldots\ldots\ldots, x_n)$ is independent of θ, since parameter θ occurring in the joint distribution of all the sample observations can be contained the distribution of statistic t, it is said that t alone can provide all the information regarding θ, therefore sufficient for θ.

3. Method of Point Estimation

So far we have been discussing the requirements of a good estimator. There are various methods of estimation which lead us to estimators that possess different properties. These estimators are known by the names that indicate the nature of the technique used in deriving the formula. The method of the maximum likelihood, method of moments, least squares method; all three methods lead to estimators which are known by the names of these techniques.

3.1 Method of Maximum Likelihood

This method was initially formulated by C.F. Gauss but as a general method of estimation was introduced by Prof. R.A. Fisher.

Let x_1, x_2, ... x_n, be a random sample from a population with p.m.f. (for discrete case) or p.d.f. (for continuous case)f(x, θ), where θ is the parameter. Then the joint distribution of the sample observations viz.

$$L = f(x_1.\theta), f(x_2.\theta), f(x_3.\theta)\ldots\ldots\ldots\ldots, f(x_n.\theta) = \prod_{i-1}^{n} f(x_i, \theta)$$

is called the Likelihood Function of the sample.

The Method of Maximum Likelihood consists in choosing as an estimator of θ that statistic, which when substituted for θ, maximizes the likelihood function L. Such a statistic is called a maximum likelihood estimator (MLE) denoted by θ_0

Since log L is maximum when L is maximum, in practice the m.l.e. of θ is obtained by maximizing log L. This is achieved by differentiating log L partially with respect to θ, and using the two relations

$$\left[\frac{\partial}{\partial\theta}\log L\right]_{\theta=\theta_0} = 0, \quad \left[\frac{\partial^2}{\partial\theta^2}\log L\right]_{\theta=\theta_0} < 0 \qquad \ldots\ldots\ldots\ldots\ldots(1)$$

Here L>0 and Log L are non decreasing function of L. Eq (1) can be rewritten by

$$\frac{1}{L}\frac{\partial L}{\partial\theta} = 0 \qquad \Rightarrow \frac{1}{L}\frac{\partial \log L}{\partial\theta} = 0 \qquad \ldots\ldots\ldots\ldots\ldots(2)$$

Here (2) is termed as likelihood equation for estimating θ.

3.2 Properties

1. A MLE is not necessarily unique and unbiased.
2. A MLE may not be consistent and uniformly minimum variance unbiased estimators (UMVUE)
3. Under very general conditions, maximum likelihood estimators are consistent. Huzurbazar in 1984 showed that under regularity conditions as sample size n tends to infinity, there exists a unique consistent maximum likelihood estimator.
4. Under certain conditions, it has been observed that maximum likelihood estimator is consistent but it is generally unbiased.

5. When maximum likelihood estimator exists, it is most efficient for large samples under the assumption but the distribution of the estimator tends to normal.

3.3 Method of Moments

The underlying principle in this method is that the sample moments reflect the population characteristics in the sense that the expected values of the sample moments are equal to the population moments.

It was first put forward by Karl Pearson in 1894. The method of moments consists of equating the sample moments to the corresponding moments of distribution, which are the functions of the unknown parameters. Here, we equate as many sample moments as there are unknown parameters. Solving these equations simultaneously we get the estimates of the moments of the population in terms of sample variates.

Here we equate the moments of the population with the corresponding moments of the sample, i.e. setting

$$\mu_r' = m_r'$$

Where $\mu_r' = E(x^r)$ and $m_r' = \frac{1}{n}\sum_{i=1}^{n} x_i^r$ Also $\mu_1' = E(x) = \mu$

These relations when solved for the parameters give the estimates by the method of moments. This method is applicable only when the population moments exist. The method is generally applied for fitting theoretical distributions to observed data.

3.4 Properties

i. Under fairly general conditions, the estimates obtained by the method of moments will have asymptotically normal distribution for large n.

ii. The mean of the distribution of estimate will differ from the true value of the parameter by a quantity of order 1/n

iii. The variance of the distribution of estimate will be of the type c^2/n.

iv. In general, the deviation estimators obtained by the method of moments are less efficient than the maximum likelihood estimators. In particular cases, they are equivalent.

3.5 Method of Least Squares estimation

The idea of least square estimation emerges from the method of maximum likelihood itself. Consider the ML estimation of μ on the basis of a sample of size n from a normal population $N(\mu, \sigma^2)$ The p.d.f of Normal distribution is

$$f(x, \mu, \sigma^2) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{1}{2\sigma^2}[x-\mu]^2\right) \qquad ;(-\infty < x < \infty)$$

And its likelihood function is

$$L = \prod_{i-1}^{n} f(x_i, \mu, \sigma^2) = \left(\frac{1}{\sigma\sqrt{2\Pi}}\right)^n \exp\left(-\frac{1}{2\sigma^2}\sum_{i=1}^{n}(x_i-\mu)^2\right)$$

The logarithm of likelihood function L is

$$\log L = \sum_{i=1}^{n} \log(x_i, \mu, \sigma^2) = \Sigma\left[-\log\sigma - \frac{1}{2}\log(2\pi) - \frac{(x_i-\mu)^2}{2\sigma^2}\right] = -\frac{n}{2}\log\sigma^2 - \frac{n}{2}\log(2\pi) - \frac{\Sigma(x_i-\mu)^2}{2\sigma^2}$$

Maximising log L implies that $\sum(yi-\mu)^2$ must be minimised, i.e. sum of squares, $\sum(yi-\mu)^2$ must be least square. The method of least square estimation is mostly used in estimating the parameters of linear

function. This very idea can be translated by considering μ itself a linear function of certain parameters β_j (j = 1,2,….k) i.e.

$\mu = \sum x_j \beta_j$ where j=1,2,…..k

where x_j's are some known coefficients of $\boldsymbol{\beta_j's}$ forming a linear function of $\boldsymbol{\beta_j}$.

3.6 Properties

1. Least squares estimators are unbiased in case of linear models.
2. The estimators of β_0 and β_1 are the uniformly minimum variance estimators.
3. The least squares estimators β_0 and $\beta_1$do not possess the asymptotically normal property.
4. The least square method is a device to obtain best linear unbiased estimators under linear set up.

4. Testing of Hypothesis

Hypothesis testing begins with an assumption; called hypothesis that we make about a population parameter. To test the validity of our assumption, we gather sample data and determine the difference between the hypothesized value and the actual value of the sample mean. Then we judge whether the difference is significant. The smaller the difference, the greater the likelihood that our hypothesized value for the mean is correct. Unfortunately, the difference between the hypothesized population parameter and the actual statistic is more often neither so large that we automatically reject our hypothesis nor so small that we just as quickly accept it. So in hypothesis testing, as in most significant real-life decisions, clear-cut solutions are the exception, not the rule.

4.1. TYPES OF HYPOTHESIS

The procedure which enables us to decide on the basis of a sample, whether a hypothesis is true or not, is called Test of Hypothesis or Test of Significance. There are two hypotheses:

- Null Hypothesis
- Alternative Hypothesis.

4.2 NULL HYPOTHESIS

A hypothesis which is to be actually tested for acceptance or rejection is termed as null hypothesis. The decision maker should adopt a null or neutral attitude regarding the outcome of the test. It is denoted by H_0.In the words of prof. R.A Fisher "Null hypothesis is the hypothesis which is tested for possible rejection under the assumption that it is true"

In hypothesis testing, a statistician or decision-maker should not be motivated by prospects of profit or loss resulting from the acceptance or rejection of the hypothesis. This means that the difference is just due to the fluctuations of sampling. e.g. $H_0 : \mu = \mu_0$ where μ_0 is some specified value of μ.

4.3 ALTERNATIVE HYPOTHESIS

Any hypothesis which contradicts the null hypothesis H_0 is called an Alternative Hypothesis and is denoted by the symbol H_1. If null hypothesis is $H_0 : \mu = \mu_0$ where μ_0 is some specified value of μ than alternative could be

(1) $H_1 : \mu \neq \mu_0$ i.e,. $\mu < \mu_0$ or $\mu > \mu_0$ (Two tailed alternative)

(2) $H_1 : \mu < \mu_0$ left tail (one tailed alternative)

(3) $H_1 : \mu > \mu_0$ right tail (one tailed alternative)

4.4 STATISTICAL HYPOTHESIS

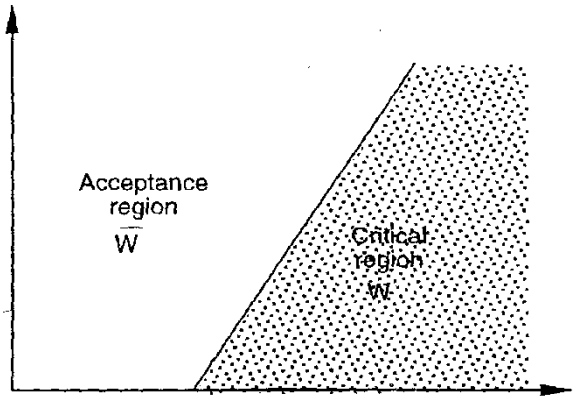
If statistical hypothesis specifies the population completely then it is termed as simple. If statistical hy

pothesis does not specifies the population completely then it is termed as Composite.

4.5 TEST OF SIGNIFICANCE

A research worker or an experimenter has always some fixed ideas about certain population parameter(s) based on prior experiments, surveys or experience. Sometimes these ideas might have been fixed in the mind. There is a need to ascertain whether these ideas or claims are correct or not by collecting information in the form of data. In this way, we come across two types of problems, first is to draw inferences about the population on the basis of sample data and the other is to decide whether our sample observations have come from a postulated population or not. By hypothesis we mean to give postulated or stipulated value(s) of a parameter. Also, instead of giving values, some relationship between parameters is postulated in the case of two or more populations. On the basis of observational data, a test is performed to decide whether the postulated hypothesis be accepted or not. This involves certain amount of risk. This amount of risk is termed as a level of significance. When the hypothesis is accepted, we consider it a non significant result and if the reverse situation occurs, it is called a significant result.

4.6 CRITICAL REGION (C.R.)



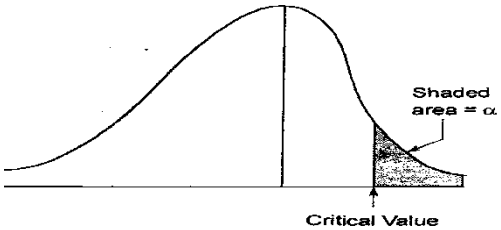
A statistic is used to test the hypothesis H_0. The test statistic follows some known distribution. In a test, the area under the probability density curve is divided into two regions, viz., the region of acceptance and the region of rejection. The region of rejection is the region in which H_0 is rejected. It means that if the value of test statistics lies in this region, H_0 (null hypothesis) will be rejected.

The region of rejection is called a **critical region**. Moreover, the area of the critical region is equal to the level of significance α.

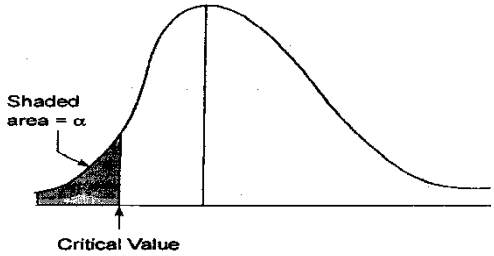
4.7 ONE AND TWO-TAILED TESTS

If the alternative hypothesis, H_1 is of the type $\mu > \mu_0$ or $\mu < \mu_0$ etc., the critical region lies on only one tail of the probability density curve. In this situation the test is called **one-tailed test**.

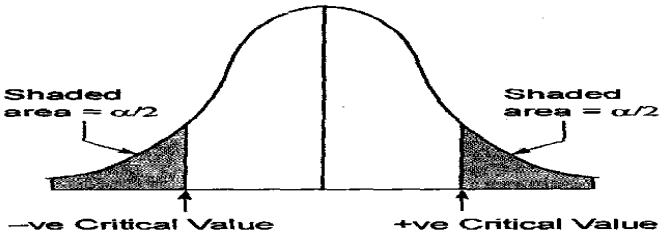
If H_1 is of the type $H_1 : \mu > \mu_0$ the critical region is towards the right tail as shown below



On contrary to this, if the alternative hypothesis, H_1 is of the type $\mu < \mu_0$ the critical region lies on only one tail (left tail) of the probability density curve. In this situation ($H_1 : \mu > \mu_0$) the critical region is towards the left tail



If the test is two-tailed, i.e., it is of the type $H_1 : \mu \neq \mu_0$ then the test is called two-tailed test and in such a case the critical region lies in both the right and left tails of the sampling distribution of the test statistic, with total area equal to the level of significance as shown in diagram.



If the alternative hypothesis is of the type $H_1 : \mu \neq \mu_0$ i.e,. $\mu < \mu_0$ or $\mu > \mu_0$ the critical region lies at the both the tails , in this situation test is called two tailed test and an area equal to $\alpha/2$ lies at the both the tails.

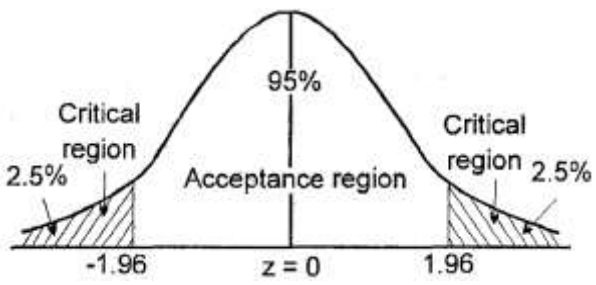
4.8 SIZE (LEVEL OF SIGNIFICANCE) AND POWER OF A TEST

In testing a given hypothesis the minimum probability with which we would be willing to risk a type one error is called as level of significance. The size of a test is the probability of rejecting the null hypothesis when it is true, and is usually denoted by α. The level of, significance and size are synonymous in a practical sense. Therefore.

$P[rejetH_0 / H_0] = \alpha$

Suppose, that under a given hypothesis the sampling distribution of a statistic θ is approximately a normal distribution with mean E(θ) and standard deviation σ_θ .

Then $z = \dfrac{\text{Observed value} - \text{Expected value}}{\text{Standard Error of } \theta}$ is called the standardized normal variable or z-score, and its distribution is the standardized normal distribution with mean 0 and standard deviation 1, the graph of which is shown below.



From the above figure, we see that if the test statistic z of a sample statistic θ lies between —1.96 and 1.96, then we are 95% confident that the hypothesis is true. But if for a simple random sample we find that the test statistic (or z-score) z lies outside the range —1.96 to 1.96, i.e. if z > 1.96, we would say that such an event could happen with probability of only 0.05 (total shaded area in the above figure if the given hypothesis were true). In this case, we say that z-score differed significantly from the value expected under the hypothesis and hence, the hypothesis is to be rejected at 5% (or 0.05) level of significance. Here the total shaded area 0.05 in the above figure represents the probability of being wrong in rejecting the hypothesis. Thus if z > 1.96, we say that the hypothesis is rejected at a 5% level of significance.

4.9 POWER OF A TEST :

The power of a test is defined as the probability of rejecting the null hypothesis when it is actually false, i.e., when H_1 is true. It is ability of a statistical test to detect the alternative hypothesis when it is true .

$$\text{Power} = P[\text{rejet}H_0 / H_1] = 1 - P[\text{accept}H_0 / H_1]$$

$$= 1 - \text{Prob.of type} - \text{II error} = 1 - \beta$$

where β is the probability of type II error.

4.10 DEGREES OF FREEDOM: is the number of independent observations in a set.

4.11 P-VALUES; It may be defined as the smallest level of α at which H_0 is rejected. In this situation , it is not inferred whether H_0 is accepted or rejected at level 0.05, 0.01 or any other value, but the statistician only give the smallest level of α at which H_0 is rejected.

5. Comparative Study of Estimation and Hypothesis of Testing

Although estimation and hypothesis testing are related, they differ in their objectives and interpretation.

Basis	**Estimation**	**Hypothesis Testing**
Objective	To estimate an unknown population parameter	To estimate a claim about a population parameter
Output	Point estimate or interval	Statistical decision or conclusion
Main concepts	Estimator, estimate, confidence interval	Null hypothesis, alternative hypothesis, test statistic and p-value
Uncertainty	Stated through standard errors	Estimated through significance level and p-value
Decision	Provides an estimate	Leads to rejection or non-rejection of H_0
Example	Estimate average population income	Test whether average population income is ₹40,000
Common measures	Mean, variance, proportion, confidence interval	T, F, Chi square and related tests
Interpretation	To measure an unknown parameter	Evaluates evidence concerning a hypothesis

6. Advantages and Disadvantages of Estimation

6.1 Advantages;

1. It provides numerical information about unknown population parameters.
2. Confidence intervals communicate uncertainty in estimates.
3. Estimation is useful for scientific measurement and prediction.
4. It allows researchers to compare estimated effects across populations.
5. It often provides more information about effect magnitude than a simple significance decision.

6.2 Disadvantages

1. The quality of an estimate depends on the quality and representativeness of the sample.
2. Different estimators may have different properties.
3. Confidence intervals depend on assumptions and the sampling model.
4. A point estimate alone does not adequately describe uncertainty.
5. Small samples may produce imprecise estimates.

7. Advantages and Disadvantages of 'Testing of Hypothesis'

7.1 Advantages

1. It provides a systematic framework for statistical decision-making and to evaluate specific claims.
2. It can be used to compare groups or assess relationships.
3. It is widely applicable in scientific, economic, medical, agricultural and social research.
4. It provides a formal method for controlling the probability of Type I error.

7.2 Disadvantages

1. Statistical significance does not necessarily imply practical or scientific importance.
2. Results can be strongly influenced by sample size and incorrect assumptions can lead to misleading conclusions.
3. A failure to reject H_0 does not prove that H_0 is true and the p-value does not directly measure the size or importance of an effect.
4. Testing several hypotheses without appropriate adjustment can increase the risk of false-positive findings.

References:

1. Casella, G., & Berger, R. L. *Statistical Inference*. Duxbury Press.
2. Hogg, R. V., McKean, J. W., & Craig, A. T. *Introduction to Mathematical Statistics*. Pearson.
3. Mood, A. M., Graybill, F. A., & Boes, D. C. *Introduction to the Theory of Statistics*. McGraw-Hill.
4. Rao, C. R. *Linear Statistical Inference and Its Applications*. Wiley.
5. Gupta, S. C., & Kapoor, V. K. *Fundamentals of Mathematical Statistics*. Sultan Chand & Sons.
6. Walpole, R. E., Myers, R. H., Myers, S. L., & Ye, K. *Probability and Statistics for Engineers and Scientists*. Pearson.