STATISTICAL INFERENCE I

BY

TANUJIT CHAKRABORTY

Indian Statistical Institute

Mail : tanujitisi@gmail.com

STATISTICAL INFERENCE

TOPICS :—

i) Elements of Estimation

ii) Elements of Testing of Hypothesis.

STATISTICAL INFERENCE 1

QUES:- Distinguish between Point & Interval estimation. [C.U.] (5)

Point Estimation	Interval Estimation
In statistics, point estimation involves the use of sample data to calculate a single value (known as a statistic) which is to serve as a "best estimate" of an unknown (fixed or random) population parameter. Let $(X_1, X_2, \ldots, X_n)$ is a random sample drawn from a population having distribution function F_θ, $\theta \in \Theta$, where the functional form of F is known except the parameter θ. If we are to guess a specific feature of the parent distribution, it can be explicitly written as a function of θ. Suppose we are to guess $\gamma(\theta)$ a real valued function of θ. The statistic $T(X_1, X_2, \ldots, X_n)$ is said to be an estimator of $\gamma(\theta)$, if we guess $\gamma(\theta)$ by $T(X_1, X_2, \ldots, X_n)$ given $(X_1, X_2, \ldots, X_n) = (x_1, x_2, \ldots, x_n)$, $T(X_1, X_2, \ldots, X_n)$ is said to be an estimate of $\gamma(\theta)$.	In statistics, interval estimation is the use of sample data to calculate an interval of possible (or probable) values of an unknown population parameter, in contrast to point estimation, which is a single number parameter. Neyman (1937) identified interval estimation ("estimation by interval") as distinct from point estimation ("estimation by unique estimate"). In doing so, he recognized that then-recent work quoting results in the form of an estimate plus-or-minus a standard deviation indicated that interval estimation was actually the problem statisticians really had in mind. An interval estimate of a real-values parameter θ is any pair of functions, $L(x_1, x_2, \ldots, x_n)$ and $U(x_1, x_2, \ldots, x_n)$, of a sample that satisfy $L(\mathbf{x}) \le U(\mathbf{x})$ for all $\mathbf{x} \in \mathbf{X}$. If $\mathbf{X} = \mathbf{x}$ is observed, the inference $L(\mathbf{x}) \le \theta \le U(\mathbf{x})$ is made. The random interval $[L(\mathbf{X}), U(\mathbf{X})]$ is called an interval estimator. Although in the majority of cases we will work with finite values for L and U, there is sometimes interest in one-sided interval estimates. For instance, if $L(\mathbf{x}) = -\infty$, then we have the one-sided interval $(-\infty, U(\mathbf{x})]$ and the assertion is that $\theta \le U(\mathbf{x})$. We could similarly take $U(\mathbf{x}) = \infty$ and have a one-sided interval $[L(\mathbf{x}), \infty)$. Although the definition mentions a closed interval $[L(\mathbf{x}), U(\mathbf{x})]$, it will sometimes be more natural to use an open interval $([L(\mathbf{x}), U(\mathbf{x}))$ or even a half-open and half-closed interval. We will use whichever seems most appropriate for the particular problem at hand. **The most prevalent forms of interval estimation are:** Confidence intervals (a frequentist method) and Credible intervals (a Bayesian method). Other common approaches to interval estimation, which are encompassed by statistical theory, are: Tolerance & Prediction intervals (used mainly in Regression Analysis). • Credible intervals can readily deal with prior information, while confidence intervals cannot. • Confidence intervals are more flexible and can be used practically in more situations than credible intervals: one area where credible intervals suffer in comparison is in dealing with non-parametric models.

STATISTICAL INFERENCE

- Parametric Inference
  (When the population distribution is known before, we apply parametric inference)
  - Point Estimation
  - Testing of Hypothesis
  - Interval Estimation
- Non-parametric Inference
  (When we have no information regarding the population distribution, we apply non-parametric inference)

STATISTICAL INFERENCE

INTRODUCTION:-

A progress in science is often ascribed to experimentation. The research workers perform an experiment and obtain some data. On the basis of the data, certain conclusions are drawn. The conclusions usually go beyond the materials and operations of the particular experiment. In other words, the scientist may generalise from a particular experiment to the class of all similar experiments. The theory is however designed to form a model of a certain group of phenomena in physical world, and the abstract objects and propositions of the theory have their counter-parts in certain observable things and relation between things. If the model is to be practically useful, there must be some kind of general agreement between their theoretical proposition and empirical counter-part. When a certain proposition has its counter-part in some directly observable relation, we must require that our observations should show that this relation holds. If in repeated tests, an agreement of these character has been found and if we regard this agreement as sufficiently accurate and permanent, the theory may be accepted for practical use.

Inductive Inference is well known to be a hazardous process. In fact, it is a theorem of logic that in inductive inference, uncertainity is present. One can't make absolutely certain generalization. However, uncertain inferences can be made and the degree of uncertainty can be measured if the experiment has been performed in accordance with certain principles. One function of statistics is the provition of techniques for making inductive inferences and for measuring the degree of uncertainty of such inferences.

INTRODUCTION OF ESTIMATION: → Suppose we are given a random sample from a population, the distribution of which has a known mathematical form but involves a certain number of unknown parameters. In general, we can construct an infinite number of function of sample values that might claim to estimate the parameters. According to Prof. R.A. Fisher, An estimator is said to be the best if it is, —

i) Unbiased
ii) Consistent
iii) efficient
iv) sufficient.

These are the criteria for the 'best' estimates. It is noted that a parameters has a meaningful physical interpretation.

THEORY OF POINT ESTIMATION :→

The national behind point estimation is quite simple. Let $(X_1, \ldots, X_n)$ is a random sample drawn from a population having distribution function F_θ, where $\theta \in \Theta$, the parameter space and here the functional form of F is known except the parameter θ (θ levels the parent distribution and it may be partially known, otherwise there is nothing to infer, moreover the labelling parameter θ may be vector valued). Now, we are interested to guess the unknown population parameter θ, as the knowledge of θ implies the knowledge about the the entire population. If we can guess a specific feature of the parent distribution, it can be explicitly written as a function of θ. Suppose, we are to guess $\gamma(\theta)$, a real valued function of θ,

The statistic $T(X_1, X_2, \ldots, X_n)$ is said to be an estimator of $\gamma(\theta)$, if we guess $\gamma(\theta)$ by $T(X_1, \ldots, X_n)$ given $(X_1, \ldots, X_n) = (x_1, x_2, \ldots, x_n)$, the realised value of the random sample, $T(x_1, \ldots, x_n)$ is said to be an estimate of $\gamma(\theta)$.

This procedure is called parametric Point Estimation.

Definition: Point estimator

A point estimator is any function of $T(X_1, X_2, \ldots, X_n)$, where $(X_1, \ldots, X_n)$ is a random sample, i.e. a point estimator is a function of random sample independent of unknown population parameter, i.e. estimator is a statistic.

EXAMPLE:– Suppose we are given the lifes (in hours) of dry cells of a specific capacity and of a certain brand. A dry cell survives if it can be burnt of a stretch of 500 hours. Let the objective to guess the population projection of dry cells that could survives, i.e. we are to guess the chance of a survival. Here it would be reasonable to assume that the life distribution is exponential with mean θ. It denotes the sample of lifes by $(X_1, \ldots, X_n)$ then the parametric function of interest be $\gamma(\theta) = P_\theta[X_1 > 500]$

$$= e^{-500/\theta}$$

If we guess θ, the population mean by the sample mean, $\bar{X}$, then an estimator of $\gamma(\theta)$ will be $e^{-500/\bar{X}}$, where as if the sample mean is observed as 612, then an estimate of $\gamma(\theta)$ will be, $e^{-500/612}$.

PROPERTIES OF POINT ESTIMATOR :—

Closeness :— Let $(X_1, X_2, \ldots, X_n)$ be a random sample drawn from a population $F_\theta(\cdot)$, $\theta \in \Theta$ (the parameter may be vector valued).

Consider the problems of estimating a real valued parametric function $\gamma(\theta)$ (often $\gamma(\theta)$ will be θ itself). Let $T(\underset{\sim}{X})$ is an estimator of $\gamma(\theta)$, but this is not possible except in trivial cases, one of which is as follows:

e.g.

The sample is drawn from the population with distribution function

$$f_\theta(x) = \begin{cases} 1 & \text{if } \theta - \frac{1}{2} < x < \theta + \frac{1}{2} \\ 0 & \text{OW} \end{cases}$$

where θ is an integer and $\theta \in \Theta$. The parameter space (Θ) consists of all integers. Consider estimating θ based on a single observation X_1 and the statistic $T(X_1)$ is defined such that $T(X_1) = X_1$, then $T(X_1)$ will always correctly estimate θ. The problem posed in this e.g. is really non statistical, since one knows the value of θ after taking one observation.

In general, we are not able to find any statistic that will estimate any parametric function $\gamma(\theta)$, correctly. For this reason, we look for an estimator $T(\underset{\sim}{X})$ that is "close" to $\gamma(\theta)$. The term "close" can be interpreted in many ways. The statistic $T(\underset{\sim}{X})$ has its distribution or a family of distribution depending on θ. So, we look for those values of $T(\underset{\sim}{X})$ which is concentrated near $\gamma(\theta)$. We know that mean and variance of a distribution measure its location and spread, respectively, So we might require of an estimator whose mean is near or equal to $\gamma(\theta)$ and have small variance. These two notions are the primary concepts of unbiasedness and minimum variance.

<u>Mathematical argument :—</u> Let $T(\underset{\sim}{X})$ be a close estimator of θ, i.e. all the realised values of $T(\underset{\sim}{X})$, say $T_1, T_2, \ldots, T_k$ fall close to θ, i.e. $|T_i - \theta| < \epsilon \quad \forall \epsilon > 0 \quad \forall i = 1(1)k$.

Now, $E(T) = \frac{1}{k}\sum_{i=1}^{k} T_i$

$$\therefore |E(T) - \theta| = \left|\frac{1}{k}\sum_{i=1}^{k} T_i - \theta\right| = \frac{1}{k}\left|(T_1-\theta) + (T_2-\theta) + \cdots + (T_k - \theta)\right|$$
$$\leq \frac{1}{k}\left[|T_1-\theta| + |T_2-\theta| + \cdots + |T_k-\theta|\right]$$
$$\leq \frac{1}{k}\cdot k\epsilon = \epsilon \qquad \text{[By Triangle inequality]}$$

$$\therefore |E(T) - \theta| < \epsilon$$

$\therefore$ If T is close then this implies that the values of $E(T)$ falls near to θ. This conditions lead us to the notion of unbiasedness.

Definition : More Concentrated estimator

Let $T(\underset{\sim}{X})$ and $T'(\underset{\sim}{X})$ be two estimators of a real valued parametric function $\gamma(\theta)$. T is said to be more concentrated estimator of $\gamma(\theta)$ than T' if and only if

$$P_\theta\left[\gamma(\theta)-\lambda < T < \gamma(\theta)+\lambda\right] \geq P_\theta\left[\gamma(\theta)-\lambda < T' < \gamma(\theta)+\lambda\right]$$

$\forall\ \lambda > 0$ and for each θ in Θ

Remark :— The above is an ideal condition in the sense of closeness of the estimate and the parametric function. The condition implies that, realised T is expected to be more close to $\gamma(\theta)$ compared to the rival estimate of T', i.e. realisation on T is more concentrated arround $\gamma(\theta)$ compared to the realisation on T'.

Definition : Pitman-closer

Let $T(\underset{\sim}{X})$ and $T'(\underset{\sim}{X})$ be two estimators of a real valued parametric function $\gamma(\theta)$. T is said to be a Pitman closer estimator of $\gamma(\theta)$ than T' if and only if

$$P_\theta\left[\,|T-\gamma(\theta)| < |T'-\gamma(\theta)|\,\right] \geq \frac{1}{2} \quad \forall\ \theta \in \Theta$$

Remark :— The above is an ideal condition of closeness between an estimate and the parametric function but the mathematical handling of the conditions are too difficult. Hence, they can't be employed as criteria in choosing an estimation of a parametric function from a class of estimators.

▨ The property of Pitman-closer is a desirable property of most concentrated estimator, yet rarely there will exist a Pitman-closer estimator. Both Pitman closer and more concentrated estimator are intuitively attractive properties to be used to compare estimators, yet they are not always useful. Given two estimators T and T', one does not have to be more concentrated or Pitman-closer than the other. What often happens is that one, say T', is Pitman-closer or more concentrated for some θ in Θ, and the other T is Pitman closer or more concentrated for some other θ in Θ; Since θ is unknown, we can't say which estimator is preferred. Therefore, we need some simple criteria and such a criteria is given in terms of Mean square error.

Mean-Squared Error (MSE) :— A useful, though perhaps crude, measure of goodness or closeness of an estimator $T(\underset{\sim}{X})$ of $\gamma(\theta)$ is what is called the mean-squared error of the estimator.

Definition : Mean-squared error

Let $T(X_1, X_2, \ldots, X_n)$ be an estimator of $\gamma(\theta)$. Here mean-squared error (MSE) of T while estimating $\gamma(\theta)$ is given by,

$$MSE_\theta(T) = E_\theta(T - \gamma(\theta))^2.$$

It is basically a second order risk function.

[as the quantity $E_\theta(T-\gamma(\theta))^2$ being the average of the squared errors $\{T(x_1, x_2, \ldots, x_n) - \gamma(\theta)\}^2$, it is termed as Mean-squared error.]

$$MSE_\theta(T) = \sum_{\underset{\sim}{x}} \{T(x_1, x_2, \ldots, x_n) - \gamma(\theta)\}^2 \, P_{X_1 \ldots X_n}(x_1, \ldots, x_n),$$

when T is discrete.

$$= \int_{\mathbb{R}^n} \{T(x_1, \ldots, x_n) - \gamma(\theta)\}^2 f_{\underset{\sim}{X}}(x_1, \ldots, x_n)\, dx_1 \ldots dx_n,$$

when T is continuous.

It is noted that the minimization of MSE of T, ensures that the realisations on T fall close to $\gamma(\theta)$. Therefore, with analogy to the ideal condition, we prefer an estimator T to T' while estimating $\gamma(\theta)$ if

$$MSE_\theta(T) \leq MSE_\theta(T') \quad \forall\ \theta \in \Theta$$

i.e. $E_\theta(T - \gamma(\theta))^2 \leq E_\theta(T' - \gamma(\theta))^2 \ \forall\ \theta \in \Theta$

We can note that $E_\theta(T - \gamma(\theta))^2$ is a measure of goodness as well as a measure of spread of T values about $\gamma(\theta)$, just as the variance of a random variable is a measure of its spread about its mean. If we are to compare estimators by looking at their respective MSE, naturally we would prefer one with small or smallest MSE, but such estimator rarely exist. To overcome this difficulty we confine ourselves to some restricted classes obtained by imposing some optimality criteria of point estimation and search for the best estimator within the restricted class.

Note that, $MSE_\theta(T) = E_\theta(T - \gamma(\theta))^2$

$$= E_\theta\left[(T - E(T)) + (E(T) - \gamma(\theta))\right]^2$$

$$= E_\theta(T - E(T))^2 + (E(T) - \gamma(\theta))^2$$

[since the product term vanishes]

Idea of Unbiasedness and minimum variance :—

$$MSE_\theta(T) = E_\theta(T-\gamma(\theta))^2$$
$$= E_\theta(T-E_\theta(T))^2 + \{E_\theta(T)-\gamma(\theta)\}^2$$
$$= V_\theta(T) + b^2_\theta(T)$$

where, $V_\theta(T)$ = variance
$b_\theta(T)$ = bias

Definition: Bias of an Estimator

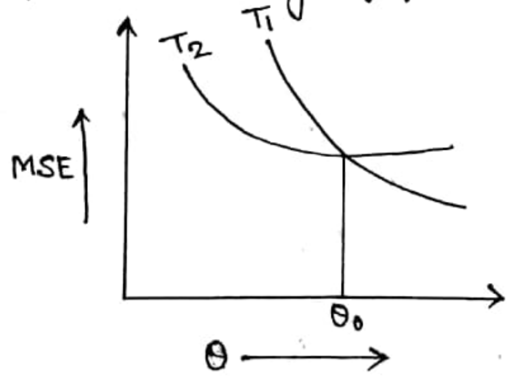
Let T be an estimator of $\gamma(\theta)$, then the quantity $\{E_\theta(T)-\gamma(\theta)\}$ is termed as the bias of an estimator.

We already have,

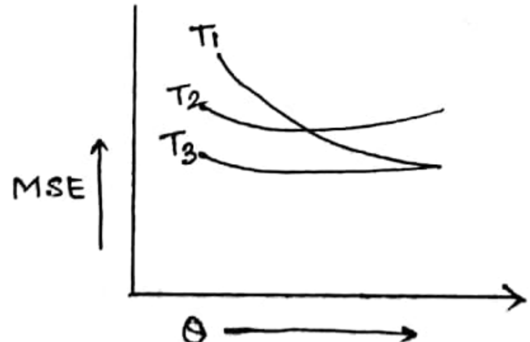
$$MSE_\theta(T) = bias^2_\theta(T) + Var_\theta(T)$$

Therefore, minimization of MSE is equivalent to the simultaneous minimization of both variance and bias. This leads us to the concept of unbiasedness and minimum variance.

Let θ be a scalar, suppose the MSE of the estimator T_1 and T_2 while estimating $\gamma(\theta)$ be as follows :



Here a choice between T_1 and T_2 can not made uniformly, i.e. for each θ in Θ, but can be made locally. Namely, if $\theta < \theta_0$, it prefer the estimator T_2 where as we prefer T_1 when $\theta > \theta_0$.



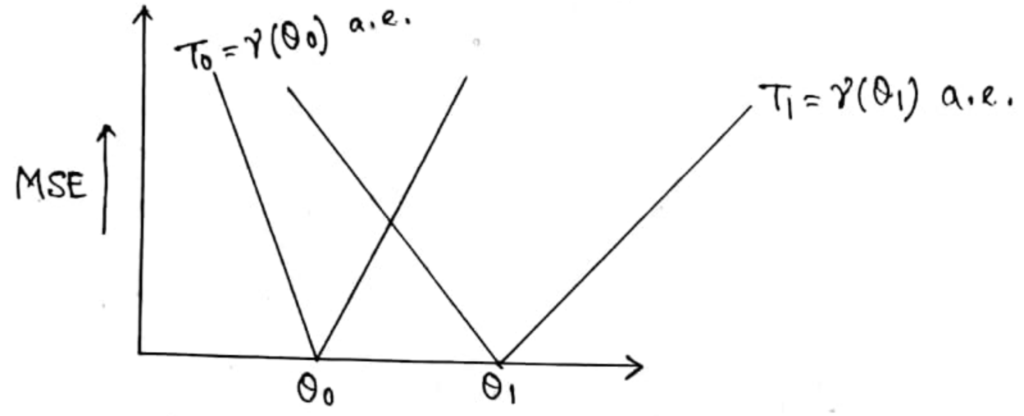
But in case of the following situation, the estimator T_3 is uniquely preferable compared to T_1, T_2.

In fact, there does not exist the best estimator of $\gamma(\theta)$ within the class of all estimators in the sense at least MSE for all values of θ. Since for each $\theta \in \Theta$. We can define a trivial estimator of $\gamma(\theta)$ $\ni$ its MSE vanishes for that specific choice of θ.

i.e. if $T_0 = \gamma(\theta_0)$ a.e.

then $MSE_{\theta_0}(T_0) = 0$

But it does not imply that MSE of T_0 will be small for other values of θ. In fact, those should be significantly large.



Thus in order to find a good estimator of $\gamma(\theta)$. We confine ourselves to some restricted classes obtain by imposing same optimality criteria of the point estimation. As for example, if we consider the class of unbiased estimators of $\gamma(\theta)$. We would like to choose the one which has uniformly the least variance, called the uniformly minimum variance unbiased estimators, or, if we consider a class of consistent estimators of $\gamma(\theta)$ [all converges in probability to $\gamma(\theta)$].

We would like to choose the one which converges more rapidly to $\gamma(\theta)$, i.e. efficient for $\gamma(\theta)$, etc. Thus choice of an estimator depends on a set of such criteria, namely, unbiasedness, minimum variance, consistency, efficiency, robustness, etc. and the choice of such criteria depends on the purpose of estimation.

Definition :— Unbiased estimator [C, U]

$T(X_1, X_2, \ldots, X_n)$ is said to be an unbiased estimator of a real valued parametric function $\gamma(\theta)$ if the mean of the sampling distribution of T is $\gamma(\theta)$ for each θ in Θ, the parameter space.

i.e. $E_\theta(T) = \gamma(\theta) \quad \forall\ \theta \in \Theta$

Otherwise, if $E_\theta(T) \neq \gamma(\theta)$ for some θ, T is said to be biased for $\gamma(\theta)$.

with the bias $b_\theta(T) = E_\theta(T) - \gamma(\theta)$, average of the difference of realised T's from $\gamma(\theta)$.

Note: If the a.m. of the sampling distribution of T is $\gamma(\theta)$, T is said to be unbiased in mean. If the median of the sampling distribution of T is $\gamma(\theta)$, T is said to be unbiased in median and if the mode of the sampling distribution of T is $\gamma(\theta)$, T is said to be unbiased in mode.

Remark 1. Unbiased estimator does not always exist.

example:

i) Let us consider a random variable, $X \sim bin(1, p)$. Suppose we want to estimate the parametric function $\gamma(p) = p^2$. Now, for a statistic $T(X)$ to be unbiased for $\gamma(p)$, one must require,

$$E_p(T(X)) = p^2 \quad , 0<p<1$$

i.e. $p^2 = pT(1) + (1-p)T(0)$

$$\Rightarrow p^2 + p[T(0) - T(1)] - T(0) = 0$$

But the LHS of the above expression is a power series (with at least one co-efficient non-zero), which vanishes $\forall\ p \in (0,1)$, which is impossible. Therefore, we can't have an unbiased estimator for p^2.

ii) Suppose $X \sim Bin(n, p)$, where, n is specified. Here, no unbiased estimator of $\frac{1}{p}$ exists based on X. If possible let,

$T(x)$ is unbiased for $\frac{1}{p}$.

$$\therefore E_p[T(X)] = \frac{1}{p} \ \forall\ p \in (0,1)$$

$$\Rightarrow \sum_j T(j)\binom{n}{j}p^j(1-p)^{n-j} = \frac{1}{p}$$

Note that, LHS $\leq \sum_j |T(j)|\binom{n}{j}p^j(1-p)^{n-j}$ [a finite quantity]

But RHS $\rightarrow \infty$ as $p \rightarrow 0$, i.e. a contradiction occurs.

iii) Catch-recatch Problem :— Let there be θ fishes in a tank of which M are caught, tagged and released. Thereafter n fishes are caught again of which x are found to be tagged then there does not exist any unbiased estimator of θ based on x.

Note that,

$$P_\theta[X=x] = \frac{\binom{M}{x}\binom{\theta-M}{n-x}}{\binom{\theta}{n}}$$

Given the sample the parameter space is

$$\theta \in \{(M+n-x), (M+n-x+1), \ldots\ldots\ldots\}$$

i.e. the parameter space is not bounded above. If possible let, $T(X)$ be unbiased for θ,

Define, $a = \min\{T(0), T(1), \ldots\ldots, T(n)\}$
$b = \max\{T(0), T(1), \ldots\ldots, T(n)\}$

evidently,

$$a \le E_\theta\{T(X)\} \le b$$

$$\Rightarrow a \le \theta \le b$$

Hence, the contradiction, since the parameter space is not bounded.

Remark 2. Unbiased estimator may sometimes be absurd.

example:

i) Let us consider the random variable, $X \sim P(\lambda)$.
Let us define a statistic $T(X) = (-2)^X$ for estimating the parametric function $\gamma(\lambda) = e^{-3\lambda}$

$$\therefore E_\lambda[T(X)] = \sum_{x=0}^{\infty}(-2)^x \frac{e^{-\lambda}\lambda^x}{x!} = \sum_{x=0}^{\infty} e^{-\lambda}.\frac{(-2\lambda)^x}{x!}$$
$$= e^{-\lambda}.e^{-2\lambda}$$
$$= e^{-3\lambda}$$

$\therefore T(X) = (-2)^X$ is an unbiased estimator for $\gamma(\lambda) = e^{-3\lambda}$

$$\text{but } (-2)^X = \begin{cases} +ve & \text{if } X \text{ is even} \\ -ve & \text{if } X \text{ is odd} \end{cases}$$

i.e. if x is odd, the $(-2)^X$ is negative, and it is absurd to have a negative estimator of a positive parametric function.

ii) Let $X_1, \ldots\ldots, X_n$ be a random sample drawn from a $N(\mu, 1)$ population. We know $\overline{X} \sim N(\mu, \frac{1}{n})$. Here $\overline{X}^2 - \frac{1}{n}$ unbiasedly estimate μ^2 which is positive for $\mu \neq 0$, where as an unbiased estimate may occasionally be negative.

Remark 3. There may exist infinitively many unbiased estimator.

example: Let us consider $X_1, X_2, \ldots, X_n \sim$ i.i.d. $P(\lambda)$

Then $E_\lambda(\bar{X}) = \frac{1}{n}\sum_{i=1}^{n} E_\lambda(X_i) = \frac{1}{n} \cdot n\lambda = \lambda$

and $E_\lambda\left[\sum_{i=1}^{n}(X_i - \bar{X})^2\right] = (n-1)\lambda$

i.e. $E_\lambda(s^2) = \lambda$, where $s^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \bar{X})^2$

Let us define, $T_\alpha(\underset{\sim}{X}) = \alpha\bar{X} + (1-\alpha)s^2$, $0 \le \alpha \le 1$

$$E_\lambda\{T_\alpha(\underset{\sim}{X})\} = \alpha E(\bar{X}) + (1-\alpha)E(s^2)$$
$$= \alpha\lambda + \lambda(1-\alpha)$$
$$= \lambda$$

$\therefore$ For each $\alpha \in [0,1]$, $T_\alpha(\underset{\sim}{X})$ is unbiased for λ.
Hence this completes the proof.

Remark 4. Unbiasedness alone does not make any sense.

Justification:- There are situations where unbiasedness ensures poor estimation. Suppose T is an unbiased estimator of θ. Further assume that the sampling distribution of T is extremely positively skewed, i.e. θ lies on the right tail of the sampling distribution. If we regard an observed T that is an estimate to be likely then the estimate should fall close to the mode of the distribution and hence it should not be close to θ. This situation is quite natural since minimisation of MSE ensures the simultaneous minimisation of the bias and variance of the sampling distribution of the statistic.

Remark 5. Pooling of Information

If there exists a number of unbiased estimators of a parameter θ then whatever the precision the individual estimator, the pooled estimate θ based on all the estimators be a precise one.

Suppose $T_1, T_2, \ldots, T_K$ are all independently distributed and unbiased for θ. Here θ can be unbiasedly estimated by the polled estimate,

$\overline{T_K} = \frac{1}{K}\sum_{i=1}^{K} T_i$. If, moreover, the variances of the estimators are uniformly bounded, then in the long run, i.e. for large K, $\overline{T_K}$ converges in probability to θ.

$$V(\overline{T}_k) = \frac{1}{k^2}\sum_{i=1}^{k} V(T_i)$$ {covariance term vanishes due to independence}

$$\leq \frac{1}{k^2}\sum_{i=1}^{k} c = \frac{c}{k}$$ { Variance are uniformly bounded $\Rightarrow V(T_i) < c$, a finite positive quantity $\forall i$ }

$$= \frac{c}{k} \to 0 \text{ as } k \to \infty$$

$\therefore E(\overline{T}_k) = \theta$, $V(\overline{T}_k) \to 0$

$$\therefore \overline{T}_k \xrightarrow{P} \theta$$

On the other hand, if T_i's are all biased with common bias β then the pooled estimate $\overline{T}_k$ approaches to $\theta + \beta$ in the long run instead of θ. Thus, it is advisable not to combine the biased estimators of θ, even if the bias is negligable.

<u>Remark 6.</u> Let T_n be an estimator of θ based on a couple of size n. T_n is said to be asymptotically unbiased for θ. If $E(T_n) \to \theta$, whenever $n \to \infty$.

Let $X_1, X_2, \ldots, X_n$ be a random sample drawn from $R(0, \theta)$ population.

Here, $X_{(n)}$ is a biased estimator of θ. It can be shown that, $E(X_{(n)}) = \frac{n}{n+1} \cdot \theta$, but the bias vanishes in long run, since, $E(X_{(n)}) \to \theta$ as $n \to \infty$.

$\therefore X_{(n)}$ is asymptotically unbiased towards θ.

Finding an unbiased estimator is a primary step forward, towards finding a good estimator. After finding the class of unbiased estimators, we search for that estimator in that class in order to have minimum MSE. Next, we could introduce the concept of minimum variance unbiased estimator.

Minimum Variance Unbiased Estimator (MVUE) :—

From the prior discussion, we know that for an estimator $(T(\underset{\sim}{X}))$, for estimating a parametric function $\gamma(\theta)$, $\theta \in \Theta$, the mean squared error is given by,

$$MSE_\theta(T) = bias^2_\theta(T) + Var_\theta(T).$$

Since estimators with uniformly minimum mean-squared error rarely exists, i.e. in order to have minimum MSE we find that class of estimators for which bias is zero and the variance is minimum for the estimator. From this condition the concept of MVUE is introduced.

Definition:— Uniformly Minimum variance Unbiased Estimator (UMVUE)

Let $(X_1, X_2, \ldots, X_n)$ be a random sample from a population F_θ, $\theta \in \Theta$, the parameter space. Then an estimator T for $\gamma(\theta)$ is said to be a UMVUE of $\gamma(\theta)$ if

i) $E_\theta(T) = \gamma(\theta) \quad \forall \quad \theta \in \Theta$

ii) $E_\theta(T^2) < \infty$ and $Var_\theta(T) \leq Var_\theta(T')$, where T' being an another estimator of $\gamma(\theta)$ satisfying $E_\theta(T') = \gamma(\theta)$.

Definition:— Best linear Unbiased Estimator (BLUE)

Let $(X_1, X_2, \ldots, X_n)$ be a random sample from a population F_θ, $\theta \in \Theta$, the parametric space. Then an estimator T for $\gamma(\theta)$ is said to be BLUE of $\gamma(\theta)$ if

i) T is linear in X_i's, i.e., the class of estimators are linear function of the random variables X_i's.

ii) $E_\theta(T) = \gamma(\theta) \ \forall \ \theta \in \Theta$

iii) $E_\theta(T^2) < \infty$, and $Var_\theta(T) \leq Var_\theta(T')$, where T' being an another estimator of $\gamma(\theta)$ satisfying $E_\theta(T') = \gamma(\theta)$.

Note:— 1. Here 'best' refers to minimum variance.

2. $V_\theta(T) \leq V_\theta(T')$, here equality holds when $T = T'$ almost everywhere.

Result :- Sample mean is the Best linear Unbiased estimator (BLUE) of the population mean. (4) [C.U.]

Proof :- Let $(x_1, x_2, \ldots, x_n)$ be a random sample from a population following a distribution with mean μ and variance σ^2.

$\therefore$ Sample mean, $\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i$

Let, $T = l_0 + \sum_{i=1}^{n} l_i x_i$ be the BLUE of the mean and l_i's are suitable constants.

From the definition of BLUE

i) $E(T) = \mu$

ii) $V(T)$ is minimum among the variances of all other linear unbiased estimators of μ.

Now, $E(T) = \mu$

$\Rightarrow l_0 + \sum_{i=1}^{n} l_i E(x_i) = \mu$

$\therefore l_0 = 0$ and $\sum_{i=1}^{n} l_i = 1$, since the above is an identity.

$V(T) = 0 + \sigma^2 \sum_{i=1}^{n} l_i^2$

$= \sigma^2 \sum_{i=1}^{n} l_i^2$

If $\sum_{i=1}^{n} l_i^2$ is minimum then $V(T)$ is minimum.

$\therefore \sum_{i=1}^{n} l_i^2$ is minimum subject to $\sum_{i=1}^{n} l_i = 1$.

Now let us construct the function

$f(l_1, l_2, \ldots, l_n) = \sum_{i=1}^{n} l_i^2 + \lambda\left(\sum_{i=1}^{n} l_i - 1\right)$, where λ is Lagrange multiplier.

Now, $\frac{\partial f}{\partial l_i} = 0$

$\Rightarrow 2l_i + \lambda = 0 \Rightarrow l_i = -\frac{\lambda}{2}$

again, $\sum_{i=1}^{n} l_i = 1$

$\Rightarrow -\frac{n\lambda}{2} = 1$

$\Rightarrow \lambda = -\frac{2}{n}$.

$\therefore l_i = \frac{1}{n}$

$\therefore T = \frac{1}{n}\sum_{i=1}^{n} x_i = \bar{x}$

$\therefore$ Sample mean is a BLUE of population mean.

Method of finding estimator :— [C.U.] (4)

Method of moments :→

A very important method of finding estimators is method of moments, proposed by Karl Pearson. Let $(X_1, X_2, \ldots, X_n)$ be a random sample from the population F_{θ_i}, $i = 1(1)k$ $\forall\, \theta_i \in \Theta$. By method of moments, estimators are found by equating the first k sample moments to the corresponding k population moments and then solving the resulting system of simultaneous equations.

More precisely, define,

$m_1' = \frac{1}{n}\sum x_i$, $\mu_1' = E(X)$

$m_2' = \frac{1}{n}\sum x_i^2$, $\mu_2' = E(X^2)$

$\vdots$

$m_k' = \frac{1}{n}\sum x_i^k$, $\mu_k' = E(X^k)$

The population moment μ_j' will be typically a function of $(\theta_1, \theta_2, \ldots, \theta_k)$, say $\mu_j'(\theta_1, \theta_2, \ldots, \theta_k)$.

The method of moment estimator $(\hat{\theta}_1, \hat{\theta}_2, \ldots, \hat{\theta}_k)$ is obtained by solving the simultaneous system of equations.

$$m_1' = \mu_1'(\theta_1, \theta_2, \ldots, \theta_k)$$
$$m_2' = \mu_2'(\theta_1, \theta_2, \ldots, \theta_k)$$
$$\vdots$$
$$m_k' = \mu_k'(\theta_1, \theta_2, \ldots, \theta_k)$$

Now, we illustrate the method by an example:

For example, suppose $X_1, X_2, \ldots, X_n$ be a random sample drawn from $N(\mu, \sigma^2)$.

The parameters are $\theta_1 = \mu$, $\theta_2 = \sigma^2$

We have, $m_1' = \frac{1}{n}\sum_{i=1}^{n} X_i = \overline{X}$

$$m_2' = \frac{1}{n}\sum_{i=1}^{n} X_i^2$$

$\mu_1' = E(X) = \mu = \theta_1$

$\mu_2' = E(X^2) = V(X) + E^2(X) = \sigma^2 + \mu^2$
$= \theta_1^2 + \theta_2$

$\hat{\theta}_1 = \hat{\mu} = \overline{X}$ = sample mean.

i.e. θ_1 is estimated by $\overline{x}$.

and $\hat{\theta}_2 = \hat{\sigma}^2 = \frac{1}{n}\sum_{i=1}^{n} X_i^2 - \hat{\mu}^2$

$$= \frac{1}{n}\sum_{i=1}^{n} X_i^2 - \bar{X}^2$$

$$= \frac{1}{n}\sum_{i=1}^{n} (X_i - \bar{X})^2 = \text{sample variance.}$$

$\therefore$ θ_2 is estimated by $\hat{\sigma}^2$.

- Method of moments: EXAMPLE : $\longrightarrow$

Example 1. Let $X_1, X_2, \ldots, X_n$ be a random sample from a Poisson distribution with parameter λ. As there is only one parameter, hence only one equation, which is

$$M_1' = \mu_1' = \mu_1'(\lambda) = \lambda.$$

Hence the method-of-moments estimator of λ is $M_1' = \bar{X}$, which says estimate the population mean λ with the sample mean $\bar{x}$.

Example 2. Let $X_1, X_2, \ldots, X_n$ be a random sample from the negative exponential density $f(x;\theta) = \theta e^{-\theta x} I_{(0,\infty)}(x)$. To estimate θ, The method-of-moments equation is

$$M_1' = \mu_1' = \mu_1'(\theta) = \frac{1}{\theta};$$

Hence the method-of-moments estimator of θ is $1/M_1' = \frac{1}{\bar{X}}$.

Example 3. Let $X_1, X_2, \ldots, X_n$ be a random sample from a uniform distribution on $(\mu - \sqrt{3}\sigma, \mu + \sqrt{3}\sigma)$. Here the unknown parameters are two, namely μ and σ, which are the population mean and standard deviation. The method-of-moments equations are

$$M_1' = \mu_1' = \mu_1'(\mu, \sigma) = \mu$$

and

$$M_2' = \mu_2' = \mu_2'(\mu, \sigma) = \sigma^2 + \mu^2;$$

Hence the method-of-moments estimators are $\bar{X}$ for μ and

$$\sqrt{\frac{1}{n}\sum X_i^2 - \bar{X}^2} = \sqrt{\frac{1}{n}\sum (X_i - \bar{X})^2} \text{ for } \sigma.$$

Remark:- Method-of-moments estimators are not uniquely defined.

Problem :— Distinguish between population and sample.

ANS :—

population	sample
1. It is defined as a total of the items under consideration.	1. It is defined as a proportion of the population selected.
2. The characteristics of a population are called as parameters.	2. The characteristics of a sample are known as statistics.
3. The population parameters are ~~known~~ generally denoted by Greek letters. For example, μ = Population mean, σ = Population standard deviation	3. The sample statistics are generally denoted by italic letters. For example, $\bar{X}$ = sample mean, S = Sample standard deviation

Problem :— Let X be a single observation from $P(\lambda)$. Is $\frac{1}{\lambda}$ unbiasedly estimable based on X ?

ANS :— X be a single observation from $P(\lambda)$. PMF of X is given by, $f(x) = \dfrac{e^{-\lambda}.\lambda^x}{x!}$, $x = 0, 1, 2, \ldots,$ $\lambda > 0$

Now,
$$E(X) = \sum_{x=0}^{\infty} x . \frac{e^{-\lambda}.\lambda^x}{x!}$$
$$= \lambda \sum_{x-1=0}^{\infty} \frac{e^{-\lambda}.\lambda^{x-1}}{(x-1)!}$$
$$= \lambda . e^{-\lambda}. e^{\lambda}$$
$$= \lambda .$$

$\therefore \frac{1}{\lambda}$ is not unbiasedly estimable based on X, where $X \sim P(\lambda)$.

Problem :— Explain the concept of "unbiasedness" and "minimum variance" in inference. [C.U.]

ANS :—

Unbiasedness : $\rightarrow$ The mean squared error (MSE) of an estimator $T(X_1, X_2, \ldots, X_n)$ of a parameter θ is the function of θ defined by $E_\theta(T-\theta)^2$.

Now, $E_\theta(T-\theta)^2 = E_\theta(T^2) - 2E_\theta(T\theta) + E_\theta(\theta^2)$

$$= E_\theta(T^2) - E_\theta^2(T) + E_\theta^2(T) - 2E_\theta(T\theta) + \theta^2$$

$$= Var_\theta(T) + [E_\theta(T) - \theta]^2 \quad \text{———} \langle 1 \rangle$$

Now, we define, the bias of an estimator.
The bias of a point estimator T of a parameter θ is the difference between the expected value of T and θ.

$\therefore \; Bias_\theta(T) = E_\theta(T) - \theta$

$\therefore$ From $\langle 1 \rangle$, $E_\theta(T-\theta)^2 = Var_\theta(T) + Bias_\theta(T)$

Now T is said to be an unbiased estimator of θ if $Bias_\theta(T) = 0$
i.e. $E_\theta(T) = \theta \;\forall\; \theta \in \Theta$

Now, there may exist a biased estimator with negligible bias such that the MSE of the estimator less than the variance of that of unbiased estimator while estimating the same parametric function since from $\langle 1 \rangle$, for an unbiased estimator

$$MSE = Var_\theta(T).$$

Minimum Variance : $\longrightarrow$ Let $(X_1, X_2, \ldots, X_n)$ be a random sample drawn from F_θ, $\theta \in \Theta$. Consider the following class of estimators of a real valued parametric function $\gamma(\theta)$ from the class of estimators

$$\mathcal{U} = \{ T(X_1, \ldots, X_n) : E_\theta(T) = \gamma(\theta) \;\forall\; \theta,\; E_\theta(T^2) < \infty \;\forall\; \theta$$

$T_0 \in \mathcal{U}$ is said to be the minimum variance estimator of $\gamma(\theta)$ if $V_\theta(T_0) \le V_\theta(T) \;\forall\; \theta \in \Theta$ and $T \in \mathcal{U}$, with equality holds iff $T_0 = T$ almost everywhere.

BLUE :-

If the above class of Unbiased estimator $\mathcal{U}$ becomes such that

$$\mathcal{U} = \{ T(X_1, \ldots, X_n) : T \text{ is linear in } X_i\text{'s},\; E_\theta(T) = \gamma(\theta) \;\forall\; \theta;\; E(T^2) < \infty, \;\forall\; \theta \}$$

and $T_0 \in \mathcal{U}$ is said to be BLUE of $\gamma(\theta)$ if $V_\theta(T_0) \le V_\theta(T)$ $\forall\; T \in \mathcal{U} \;\forall\; \theta$, with equality holds iff $T = T_0$ a.e.

Problem :—

1. $\{X_1, X_2, \ldots, X_n\}$ come from $B(\pi)$. Find the unbiased estimators of i) π, ii) π^2, iii) $\pi(1-\pi)$, iv) $(1+\pi)^k$, $k \in \mathbb{N}$ based on all X_i's.

Solution:— Define, $T = \sum_{i=1}^{n} X_i$, here T is sufficient for π. Thus we obtain the unbiased estimators of the parametric function based on T.

i) In order to have an unbiased estimator of π, we begin with

$$E(T) = E\left[\sum_{i=1}^{n} X_i\right] = \sum_{i=1}^{n} \pi = n\pi.$$

$$\therefore \pi = E\left(\frac{T}{n}\right)$$

$\therefore$ π is unbiasedly estimated by $\frac{T}{n} = \bar{X}$.

ii) To have an unbiased estimator of π^2, first we determine,

$$E(T^2) = V(T) + E^2(T) \qquad [\because T \sim Bin(n, \pi)]$$

$$\Rightarrow E(T^2) = n\pi(1-\pi) + n^2\pi^2$$

$$\Rightarrow E(T^2) - n\pi = n(n-1)\pi^2$$

$$\Rightarrow E(T^2) - E(T) = n(n-1)\pi^2$$

$$\Rightarrow E\left(\frac{T(T-1)}{n(n-1)}\right) = \pi^2.$$

$\therefore$ π^2 is unbiasedly estimated by $\frac{T(T-1)}{n(n-1)} = \frac{\bar{X}(n\bar{X}-1)}{(n-1)}$.

iii) To have an unbiased estimator of $\pi - \pi^2$, we will start from (i) and (ii)

$$\pi - \pi^2 = \pi(1-\pi) = E\left[\frac{T}{n} - \frac{T(T-1)}{n(n-1)}\right]$$

$$= E\left[\bar{X}\left(1 - \frac{n\bar{X}-1}{n-1}\right)\right]$$

$$= E\left[\bar{X}\left(\frac{n-1-n\bar{X}+1}{n-1}\right)\right]$$

$$= E\left[\frac{n\bar{X}}{n-1}(1-\bar{X})\right]$$

$\therefore$ $\pi(1-\pi)$ is unbiasedly estimated by $\frac{n\bar{X}}{n-1}(1-\bar{X})$.

iv) $$(1+\pi)^k = \sum_x \binom{k}{x} \pi^x = \sum_x \binom{k}{x} \frac{(T)_x}{(n)_x}$$

$$\pi \triangleq \frac{T}{n}$$

$$\pi^2 \triangleq \frac{(T)_2}{(n)_2}$$

$$\vdots$$

$$\pi^x \triangleq \frac{(T)_x}{(n)_x}$$

$\therefore$ $(1+\pi)^k$ is unbiasedly estimated by

$$\sum_x \binom{k}{x} \frac{(T)_x}{(n)_x}.$$

2. Let $X_1, X_2, \ldots, X_n$ be an i.i.d. random sample drawn from $R(0,\theta)$ population. Suggest unbiased estimators of θ based on $X_{(1)}$, $X_{(n)}$ and $\bar{X}$. Compare the estimators.

ANS:-

i) $$P[X_{(n)} \le x] = P[X_1, X_2, \ldots, X_n \le x]$$
$$= \prod_{i=1}^{n} P(X_i \le x) \quad \text{[due to independence]}$$
$$= \{P[X_1 \le x]\}^n \quad [\because X_i\text{'s are iid}]$$
$$= \left(\frac{x}{\theta}\right)^n I_{0<x<\theta} + I_{x \ge \theta}$$

PDF of $X_{(n)}$ is,

$$f_{X_{(n)}}(x) = \frac{n x^{n-1}}{\theta^n} I_{0<x<\theta}$$

$$E(X_{(n)}) = \frac{n}{\theta^n}\int_0^\theta x . x^{n-1} dx = \frac{n}{n+1}\theta$$

$$E\left[\frac{n+1}{n} X_{(n)}\right] = \theta$$

$\therefore$ θ is unbiasedly estimated by, $T_1 = \frac{n+1}{n} X_{(n)}$.

ii) $$P[X_{(1)} \le x] = 1 - P[X_{(1)} > x]$$
$$= 1 - P[X_1, \ldots, X_n > x]$$
$$= 1 - \prod_{i=1}^{n} P[X_i > x] \quad \text{[due to independence]}$$
$$= \left\{1 - \left(1 - \frac{x}{\theta}\right)\right\}^n I_{0<x<\theta} + I_{x \ge \theta}$$

PDF of $X_{(1)}$ is,

$$f_{X_{(1)}}(x) = \frac{n}{\theta}\left(1 - \frac{x}{\theta}\right)^{n-1} I_{0<x<\theta}$$

$$E(X_{(1)}) = \int_0^\theta x . \frac{n}{\theta}\left(1 - \frac{x}{\theta}\right)^{n-1} dx$$
$$= n\theta \int_0^1 y(1-y)^{n-1} dy$$
$$= n\theta\, \beta(2, n)$$
$$= \frac{\theta}{n+1}.$$

$\therefore$ $E[(n+1) X_{(1)}] = \theta$

$\therefore$ θ is unbiasedly estimated by, $T_2 = (n+1) X_{(1)}$.

iii) $E(\bar{X}) = \frac{1}{n}\sum_{i=1}^{n} E(X_i) = \frac{1}{n}\cdot\sum_{i=1}^{n}\frac{\theta}{2} = \frac{\theta}{2}$

$E(2\bar{X}) = \theta$

$\therefore$ θ is unbiasedly estimated by $T_3 = 2\bar{X}$.

Now, we compare the estimators from the point of view of MSE, as the estimates are unbiased, it is enough to have the sampling variance of the unbiased estimators.

i) $V(T_1) = \left(\frac{n+1}{n}\right)^2 V_\theta(X_{(n)})$

$E(X_{(n)}^2) = \frac{n}{\theta^n}\int_0^\theta x^{n+1}dx$

$= \theta^2 \cdot \frac{n}{n+2}$.

$V(T_1) = \left(\frac{n+1}{n}\right)^2 \left[\theta^2 \cdot \frac{n}{n+2} - \left(\frac{n\theta}{n+1}\right)^2\right]$

$= \frac{\theta^2}{n(n+2)}$.

ii) $V(T_2) = V((n+1)X_{(1)})$

$= (n+1)^2 V_\theta(X_{(1)})$

$E(X_{(1)}^2) = \int_0^\theta x^2 \cdot \frac{n}{\theta}\left(1-\frac{x}{\theta}\right)^{n-1}dx$

$= n\theta^2\int_0^1 t^2(1-t)^{n-1}dt$

$= n\theta^2 \beta(3,n)$

$= \frac{2\theta^2}{(n+1)(n+2)}$

$\therefore$ $V(T_2) = (n+1)^2\left[\frac{2\theta^2}{(n+1)(n+2)} - \frac{\theta^2}{(n+1)^2}\right]$

$= \frac{\theta^2 n}{n+2}$.

iii) $V(T_3) = V(2\bar{X}) = 4V(\bar{X}) = \frac{4}{n}V(X_i) = \frac{\theta^2}{3n}$

$\left[\because V(X_i) = \frac{\theta^2}{12}\right]$

— x —

Statistical Inference

- Statistical Inference
  - Estimation
    - Point Estimation
    - Interval Estimation
  - Testing of Hypothesis

INTRODUCTION :- A sample from the distribution of a population is useful in making inferences about the population characteristics. The process of going from known sample to the unknown popln. has been called statistical inference.

(1) Estimation :— Some features of the popln. in which an investigator is interested, may be knownto him and he may want to make a guess about this features, on the basis of a random sample drawn from the popln. This type of problem is called problem of estimation.

(2) Testing of Hypothesis :— Some tentative information on a feature of the population may be available to the investigator and he may want to see whether the information is tenable in the light of the random sample taken from the population. This type of problem is called the problem of testing of hypothesis.

(1) Concept of Estimation :— The problem of estimation is loosely defined as: assume that some characteristics of the elements of the popln. can be represented by a r.v. X whose PMF or PDF is $f(x, \theta)$ where the functional form of the PMF or PDF is known except the parameter θ, $\theta \in \Omega$. The set Ω is called the parameter space. Let $(x_1, x_2, \dots, x_n)$ be an observed random sample from $f(x, \theta)$. On the basis of the observed random sample, it is desired to estimate the value of the parameter θ. This estimation is done in two ways.

(a) Point Estimation :- The problem of point estimation is to pick or select a statistic $T(\underset{\sim}{x}) = T$ that best estimates the parameters.

The numerical value of $T(\underset{\sim}{x})$ when an observed value of $\underset{\sim}{X}$ is $\underset{\sim}{x}$ is called an estimate of θ while such a statistic $T(\underset{\sim}{x})$ is called an estimator of θ. Let (X_1, X_2, X_3) be a random sample from $f(x, \theta)$. Then $\bar{X} = \dfrac{X_1 + X_2 + X_3}{3}$ is an estimator of θ. If the observed sample is $(-1, 1, 3)$, then the sample mean, $\bar{x} = 1$ is an estimate of θ.

(b) Interval Estimation :— The problem of interval estimation is to define 2 statistic $T_1(\underset{\sim}{x})$ and $T_2(\underset{\sim}{x})$ such that (T_1+T_2) constitutes an interval for which the probability can be determined that it contains the parameter θ.

(a) Point Estimation :— It is clear that if any given problem of estimation, we may have a large, often an infinite no. of estimators, we may choose from.

Requirement of good estimator / Measures of quality of the estimator

Clearly we could like the estimator $T(\underset{\sim}{x}) = T$ to be close to θ. Since T is a R.V., the usual measures of closeness $|T-\theta|$ is also a R.V. Example of such measure of closeness are

Part : 1 : $P[\,|T-\theta| < \epsilon\,] \;\; \forall \; \epsilon > 0$ ——— (1)

Part : 2 : $E\left[\,|T-\theta|^r\,\right]$, for some $r>0$ ——— (2)

$$\left[P[\,|T-\theta| < \epsilon\,] > 1 - \frac{E[\,|T-\theta|^r\,]}{\epsilon^r} \right]$$

We want to be large (1) but to be small (2).

Mean Square Error (MSE) :— A useful, though perhaps a crude measure of closeness of an estimator T of θ is $E(T-\theta)^2$, which is obtained from (2) by putting $r=2$.

Definition :— Let T is an estimator of θ. The quantity $E(T-\theta)^2$ is defined to be the MSE of estimator T.

Notation :— $MSE_\theta(T) = E[T-\theta]^2$.

Note that, $E[T-\theta]^2$ is a measure of spread of the values of T about the parameter θ. If we are to compare estimators by looking at there respective MSE's, naturally we would prefer (1) with small or smallest MSE.

Here the requirement is to choose T_0 such that $MSE_\theta(T_0) \leq MSE_\theta(T)$. for all T, for $\theta \in \Omega$. But such estimator rearly exists.

Note that,
$$\begin{aligned} MSE_\theta(T) &= E(T-\theta)^2 \\ &= E\left[T - E(T) + E(T) - \theta\right]^2 \\ &= E\{T - E(T)\}^2 + \{E(T) - \theta\}^2 \\ &\quad + 2E\{T-E(T)\}\{E(T)-\theta\} \\ &= var(T) + \{b(\theta,T)\}^2 \end{aligned}$$

Hence, to control MSE, we need to control both Var(T) and $\{b(\theta,T)\}^2$, the quantity $b(\theta,T) = E(T) - \theta$, is called the bias of T in estimating θ.

One approach is to restrict attention to those estimator which have zero bias, i.e. $E(T)=\theta \ \forall \ \theta \in \Omega$.

If $b(\theta,T)=0$, then T is called an unbiased estimator of θ and

$$MSE_{\theta}(T)=Var(T).$$

Now, it is required to find an estimator with uniformly minimum MSE among all unbiased estimator, which is equivalent to finding an estimator with uniformly minimum variance among all unbiased estimator. This is the concept of unbiasedness and minimum variance.

Unbiasedness :-

Definition :- An estimator T is defined to be an unbiased estimator (UE) of θ if $E(T)=\theta \ \forall \ \theta \in \Omega$.

Unbiasedness of T says that T has no systematic error, it neither overestimates nor underestimates θ on an average.

Biasedness :-

Definition :- An estimator T is said to be biased for the parameter θ if $E(T) \neq \theta$ for some $\theta \in \Omega$.

Ex. 1. Unbiased Estimator of population moments :—

Let $X_1, X_2, \ldots, X_n$ be a r.s from a popln. with finite kth order moment $\mu_k' = E(X_1^k)$. Nothing else is known about the popln. distribution. Find an unbiased estimator of μ_r', $1 \le r \le k$.

Solution :- Define $m_r' = \frac{1}{n}\sum_{i=1}^{n} X_i^r$

Then,
$$E(m_r') = \frac{1}{n}\sum_{i=1}^{n} E(X_i^r)$$
$$= \frac{1}{n}\cdot n \cdot E(X_1^r), \text{ as } X_i\text{'s are i.i.d.} \Leftrightarrow X_i^r\text{'s are i.i.d.}$$
$$= E(X_1^r)$$
$$= \mu_r', \quad 1 \le r \le k.$$

Hence, the sample rth order raw moment is an unbiased estimator (UE) of μ_r', $r=1(1)k$.

Ex. 2. Let $X_1, X_2, \ldots, X_n$ be the random sample from an infinite population with mean μ and variance $\sigma^2 (<\infty)$. Show that $s^2 = \frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})^2$ is a biased estimator of σ^2. Hence, find an UE of σ^2.

Solution:—

$$E[s^2] = \frac{1}{n} E\left[\sum_{i=1}^{n}(x_i - \mu - \bar{x} + \mu)^2\right]$$

$$= \frac{1}{n} E\left[\sum_{i=1}^{n}(x_i - \mu)^2 - n(\bar{x} - \mu)^2\right]$$

$$= \frac{1}{n}\left\{\sum_{i=1}^{n} Var(x_i) - n\, Var(\bar{x})\right\}$$

$$= \frac{1}{n}\cdot\left\{n\sigma^2 - \frac{n.\sigma^2}{n}\right\} = \frac{n-1}{n}\sigma^2$$

[Here, $E(x_i) = \mu$, $Var(x_i) = \sigma^2$
$E(\bar{x}) = \mu$, $Var(\bar{x}) = \frac{1}{n^2}\sum_{i=1}^{n} V(x_i) = \frac{\sigma^2}{n}$.]

Here, $E(s^2) = \frac{n-1}{n}\cdot\sigma^2 \neq \sigma^2$ $\therefore$ Bias $= E(s^2) - \sigma^2$
$= -\sigma^2/n$
$\longrightarrow 0$ as $n \to \infty$.

$$\Rightarrow E\left(\frac{ns^2}{n-1}\right) = \sigma^2.$$

Hence $s'^2 = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{x})^2$ is a UE of σ^2.

$\therefore$ Bias$(\sigma^2, s^2) = E(s^2) - \sigma^2$
$= -\frac{1}{n}\sigma^2 \longrightarrow 0$ as $n \to \infty$.

Ex. 3. Let $X_1, X_2, \ldots, X_n$ be a r.s. from $P(\lambda)$ distn. S.T. for $0 \le \alpha \le 1$, $T_\alpha = \alpha\bar{x} + (1-\alpha)s^2$ is an UE of λ and comment.

Solution:— We know that $\bar{x}$ and s^2 are UEs of the popln. mean and variance, respectively. Since for $P(\lambda)$ distn., $\bar{X} = s^2 = \lambda$.

Hence, $E(T_\alpha) = \alpha E(\bar{x}) + (1-\alpha)E(s^2)$
$= \alpha.\lambda + (1-\alpha)\lambda$
$= \lambda$, $\alpha \in [0,1]$

For each $\alpha \in [0,1]$, T_α is an UE of λ. Hence there are infinitely many UEs of λ of the form

$$T_\alpha = \alpha\bar{x} + (1-\alpha)s^2.$$

Uniformly Minimum Variance Unbiased Estimator (UMVUE):-

[WBSU'11]

Let T_1 and T_2 be two different UEs of θ. Then there exists an infinitely many UEs of θ of the form:

$$T_\alpha = \alpha . T_1 + (1-\alpha) T_2 \quad , \quad 0 \le \alpha \le 1.$$

Which of these should we choose?

Here comes the concept of UMVUE.

Definition:—

(a) An estimator T^* is defined to be UMVUE of θ iff

i) $E(T^*) = \theta \quad \forall \ \theta \in \Omega$.

ii) $Var_\theta(T^*) \le Var_\theta(T) \quad \forall \ \theta \in \Omega$,

for any estimator T which satisfies $E(T) = \theta \ \forall \ \theta \in \Omega$.

(b) An UE is said to be UMVUE of θ if it has minimum variance among all UEs of θ.

Ex. 1. Let $X_1, X_2, \ldots, X_n$ be a r.s. from $U(0,\theta)$. Find two UEs of θ, one based on $\bar{X}$ and other based on $X_{(n)}$. Which one is better?

Solution:—

$$E(\bar{X}) = E(X_1) = \frac{\theta}{2}$$

$$\Rightarrow E(2\bar{X}) = \theta.$$

Hence $T_1 = 2\bar{X}_\theta$ is an UE of θ.

$$E[X_{(n)}] = \int_0^\theta x . \frac{n x^{n-1}}{\theta^n} dx \qquad \left[\because f_{X_{(n)}}(x) = \begin{cases} \frac{n x^{n-1}}{\theta^n} & , 0<x<\theta \\ 0 & , OW \end{cases} \right]$$

$$= \frac{n}{\theta^n} \int_0^\theta x^n dx = \frac{n\theta}{n+1}$$

$$\Rightarrow E\left\{ \frac{n+1}{n} X_{(n)} \right\} = \theta$$

Hence, $T_2 = \frac{n+1}{n} X_{(n)}$ is an UE of θ.

Now, $Var(T_1) = 4 . V(\bar{X}) = 4 . \frac{V(X_1)}{n} = 4 . \frac{\theta^2}{12n} = \frac{\theta^2}{3n}$

and $Var(T_2) = \left(\frac{n+1}{n}\right)^2 E(X_{(n)}^2) - E^2\left(\frac{n+1}{n} X_{(n)}\right)$

$$= \left(\frac{n+1}{n}\right)^2 . \int_0^\theta x^2 . \frac{n x^{n-1}}{\theta^n} dx - \theta^2$$

$$= \left(\frac{n+1}{n}\right)^2 . \frac{n}{\theta^n} . \frac{\theta^{n+2}}{n+2} - \theta^2$$

$$= \left\{ \frac{(n+1)^2}{n(n+2)} - 1 \right\} \theta^2 = \frac{\theta^2}{n(n+2)}$$

For $X_{(1)}$:-

$P[X_{(1)} \le x] = 1 - P[X_{(1)} > x]$

$= 1 - P[X_1, \ldots, X_n > x]$

$= 1 - \prod_{i=1}^n P[X_i > x]$ [due to indep.]

$= \{1 - (1 - \frac{x}{\theta})\}^n, \ 0<x<\theta$

$\therefore f_{X_{(1)}}(x) = \frac{n}{\theta}\left(1-\frac{x}{\theta}\right)^{n-1}$

$\therefore E(X_{(1)}) = \frac{\theta}{n+1}$

$\therefore E((n+1) X_{(1)}) = \theta$

$\therefore \theta$ is unbiasedly estimated by $(n+1) X_{(1)}$.

Note that $\frac{V(T_1)}{V(T_2)} = \frac{n+2}{3} \ge 1, \ n \in \mathbb{N}$

For $n > 1$, $V(T_1) > V(T_2)$ and T_2 has smaller variance than T_1. Hence, $T_2 = \frac{n+1}{n} X_{(n)}$ is better estimator in estimating θ.

Theorem: – The UMVUE of a parameter, if exists, is unique.

Proof: – If possible, let T_1 and T_2 be two UMVUEs of θ.

Then $V(T_1) = V(T_2) = \nu$, say.

Clearly $T = \frac{T_1+T_2}{2}$ is also an UE of θ.

Hence, $var(T) \geq \nu$

$\Rightarrow var\left(\frac{T_1+T_2}{2}\right) \geq \nu$

$\Rightarrow \frac{1}{4}\left[V(T_1) + V(T_2) + 2cov(T_1, T_2)\right] \geq \nu$

$\Rightarrow \frac{1}{4}\left[\nu + \nu + 2\rho\nu\right] \geq \nu$ $\left[\because cov(T_1, T_2) = \rho\sqrt{V(T_1)V(T_2)} = \rho\nu\right]$

$\Rightarrow \rho \geq 1$, but $|\rho| \leq 1$.

Hence, $\rho = 1$.

$\Rightarrow T_1 = a + bT_2$, $b > 0$ with prob. 1.

Now, $E(T_1) = a + b.E(T_2)$

$\Rightarrow \theta = a + b\theta \;\; \forall \theta$

$\Rightarrow a = 0, b = 1$, equating the coefficients of constant term and θ.

$\left[V(T_1) = b^2 V(T_2) \Rightarrow b^2 = 1, b > 0 \Rightarrow b = 1.\right.$ and

$\left. E(T_1) = a + bE(T_2) \Rightarrow \theta = a + 1.\theta \Rightarrow a = 0 \right]$

Hence $T_1 = T_2$ with prob. 1.

i.e. UMVUE, if exists, is unique.

Ex. 2. Let T_1, T_2 be two UEs with common variance $\alpha\sigma^2$, where σ^2 is the variance of the UMVUE. Show that,

$$\rho_{T_1, T_2} \geq \frac{2-\alpha}{\alpha}.$$

Solution: –

Note that, $T = \frac{T_1+T_2}{2}$ is an UE of the parameter.

Clearly, $V(T) \geq \sigma^2$

$\Rightarrow V\left(\frac{T_1+T_2}{2}\right) \geq \sigma^2$

$\Rightarrow \frac{1}{4}\left[V(T_1) + V(T_2) + 2cov(T_1, T_2)\right] \geq \sigma^2$

$\Rightarrow \frac{1}{4}\left[2\alpha\sigma^2 + 2\rho_{T_1, T_2} \cdot \alpha\sigma^2\right] \geq \sigma^2$

$\Rightarrow \frac{\alpha}{2}\left\{1 + \rho_{T_1, T_2}\right\} \geq 1$

$\Rightarrow \rho_{T_1, T_2} \geq \frac{2}{\alpha} - 1 = \frac{2-\alpha}{\alpha}.$

FURTHER PROBLEMS:—

Ex.1. Estimating p^2 for Bernoulli distribution

(a) Let $X_1, X_2, \ldots, X_n$ be a r.s. from $B(1,p)$, $0<p<1$, $n \geq 2$. Can we estimate p^2 unbiasedly based on $X_1, \ldots, X_n$? If so, how?

(b) Let X be a single observation from $B(1,p)$. Can you estimate p^2 unbiasedly based on X?

Solution:-

(a) Let $T = \sum_{i=1}^{n} X_i$. Then T denotes the no. of successes in n independent Bernoulli trials.

Hence, $T \sim Bin(n,p)$.

$\left[\because E\left[(T)_r\right] = (n)_r \cdot p^r,\ r \leq n\right]$

We have, $E\{T(T-1)\} = n(n-1)p^2$

$$\Rightarrow E\left\{\frac{T(T-1)}{n(n-1)}\right\} = p^2$$

Hence $h(T) = \frac{T(T-1)}{n(n-1)}$ is an UE of p^2.

(b) If possible, let $T(x)$ be an UE of p^2.

Then by definition,

$$E(T(X)) = p^2 \quad \forall\, p \in (0,1)$$

$$\Rightarrow \sum_{x=0}^{1} T(x)\, P[X=x] = p^2$$

$$\Rightarrow T(0)\cdot(1-p) + T(1)p = p^2$$

$$\Rightarrow p^2 + \{T(0) - T(1)\}p - T_0 = 0 \quad \forall\, p \in (0,1) \qquad \text{(i)}$$

Clearly, (i) is an identity in p.

Equating the coefficients of p^2, p and constant term, we get,

$1 = 0 \rightarrow$ absurd

and $T(0) - T(1) = 0$

Hence, there exists no $T(X)$ which will satisfy "$E[T(X)] = p^2$" $\forall\, p \in (0,1)$.

Hence, there is no UE of p^2 based on a single observation X from $Bin(1,p)$.

Ex. (2). Let X be a single observation from $P(\lambda)$. Does there exist an UE of $\frac{1}{\lambda}$?

Solution:- If possible, let $T(X)$ be an UE of $\frac{1}{\lambda}$.

Then $E(T(X)) = \frac{1}{\lambda} \;\forall\; \lambda > 0$

$$\Rightarrow \sum_{x=0}^{\infty} T(x)\, e^{-\lambda} \cdot \frac{\lambda^x}{x!} = \frac{1}{\lambda} \quad \forall\, \lambda > 0$$

$$\Rightarrow \sum_{x=0}^{\infty} T(x) \cdot \frac{\lambda^{x+1}}{x!} = e^{\lambda}$$

$$\Rightarrow \sum_{x=0}^{\infty} T(x) \cdot \frac{\lambda^{x+1}}{x!} = \sum_{x=0}^{\infty} \frac{\lambda^x}{x!}, \; \lambda > 0$$

$$\Rightarrow 1 + \left\{\frac{1}{1!} - \frac{T(0)}{0!}\right\}\lambda + \left\{\frac{1}{2!} + \frac{T(1)}{1!}\right\}\lambda^2 + \cdots\cdots = 0 \;\forall\, \lambda > 0$$

By uniqueness of powers series, we have

$$1 = 0 \text{ (absurd)}$$

$$\frac{1}{1!} - \frac{T(0)}{0!} = 0, \quad \frac{1}{2!} - \frac{T(1)}{1!} = 0, \cdots\cdots$$

Hence, there exists no UE of $\frac{1}{\lambda}$ based on X.

Ex. 3.

(a) Starting from the equation $\sigma^2 = E(X^2) - \mu^2$, we get $\mu^2 = E(X^2 - \sigma^2)$ and $(X^2 - \sigma^2)$ is an UE of μ^2, what is its principal defects?

Solution:-

Hints:- (a) If σ is unknown, then $(X^2 - \sigma^2)$ is not a ~~statistic~~ and not measurable or observable. Then, $(X^2 - \sigma^2)$ can not be used as an estimator of μ^2.

(b) Show that if $\hat{\theta}$ is an UE of θ and $Var(\hat{\theta}) \neq 0$, $\hat{\theta}^2$ is not an UE of θ^2.

Hints:-
$$0 < Var(\hat{\theta}) = E(\hat{\theta}^2) - E^2(\hat{\theta})$$
$$= E(\hat{\theta}^2) - \theta^2$$
$$\Rightarrow E(\hat{\theta}^2) > \theta^2.$$

Ex. 4. Let $X_1, X_2, \ldots, X_n$ be a r.s. from $N(0, \sigma^2)$ distn. Suggest an UE of σ based on $\sum_{i=1}^{n} |X_i|$ and also an alternative UE based on $\sum_{i=1}^{n} X_i^2$.

Solution:- Note that, $E\left(\sum_{i=1}^{n} |X_i|\right) = \sum_{i=1}^{n} E|X_i| = \sum_{i=1}^{n} \sigma \sqrt{\frac{2}{\pi}}$

$$= \sigma . n . \sqrt{\frac{2}{\pi}}$$

$$\Rightarrow E\left\{\sqrt{\frac{\pi}{2}} \cdot \frac{1}{n} \sum_{i=1}^{n} |X_i|\right\} = \sigma$$

$$\Rightarrow T_1 = \sqrt{\frac{\pi}{2}} \cdot \left(\frac{1}{n} \sum_{i=1}^{n} |X_i|\right) \text{ is an UE of } \sigma^2.$$

Now, $\chi^2 = \frac{\sum_{i=1}^{n} X_i^2}{\sigma^2} \sim \chi_n^2$.

$$\left[\; E(\chi^2) = n \Rightarrow E\left(\frac{1}{n} \sum_{i=1}^{n} X_i^2\right) = \sigma^2 \right.$$

$$\left. \Rightarrow \frac{1}{n} \sum_{i=1}^{n} X_i^2 \text{ is an UE of } \sigma^2 \right]$$

Now, $E\left[\sqrt{\chi^2}\right] = \int_0^\infty \sqrt{x} \cdot \frac{1}{2^{n/2} \Gamma(n/2)} \cdot e^{-x/2} x^{\frac{n}{2}-1} dx$

$$= \frac{2^{\frac{n+1}{2}} \Gamma\left(\frac{n+1}{2}\right)}{2^{n/2} \Gamma(n/2)} = \frac{\sqrt{2}\, \Gamma\left(\frac{n+1}{2}\right)}{\Gamma\left(\frac{n}{2}\right)} = c_n, \text{ say}$$

$$\Rightarrow E\left(\frac{\sum_{i=1}^{n} X_i^2}{\sigma^2}\right)^{1/2} = c_n \Rightarrow E\left(\frac{1}{c_n} \cdot \sqrt{\sum_{i=1}^{n} X_i^2}\right) = \sigma.$$

$$\Rightarrow T_2 = \frac{1}{c_n} \cdot \sqrt{\sum_{i=1}^{n} X_i^2} \text{ is an UE of } \sigma.$$

Ex. 5. Let $X_1, X_2, \ldots, X_n$ be a r.s. from $N(\mu, 1)$. Find an UE of μ^2.

Solution:- $V(\bar{X}) = \frac{1}{n}$

$$\Rightarrow E(\bar{X}^2) - E^2(\bar{X}) = \frac{1}{n}$$

$$\Rightarrow E\left(\bar{X}^2 - \frac{1}{n}\right) = \mu^2.$$

Note that, the estimator $\left(\bar{X}^2 - \frac{1}{n}\right)$ can take negative values in estimating a positive parameter μ^2 and $\left(\bar{X}^2 - \frac{1}{n}\right)$ is not so sensitive.

Ex.6. Let $X_1, X_2, \ldots, X_n$ be a r.s. from $N(\mu, \mu)$, $\mu > 0$. Find an UE of μ^2 based on both $\bar{X}$ and S^2.

Solution:-

Here $\bar{X}$ is an UE of population mean $E(X_1) = \mu$ and S^2 is UE of popl^n. variance $V(X_1) = \mu$.

Hence, $E(\bar{X} \cdot S^2) = E(\bar{X}) \cdot E(S^2) = \mu^2$.

[For a normal sample, $\bar{X}$ and S^2 are independently distributed]

N.T. $\alpha \bar{X} + (1-\alpha) S^2$ is an UE of μ, $0 \le \alpha \le 1$.

Ex.7. Let $X_1, X_2, \ldots, X_n$ be a r.s. from the PDF

$$f(x) = \begin{cases} \theta x^{\theta - 1}, & 0 < x < 1 \\ 0, & \text{OW} \end{cases}, \text{ where } \theta > 0.$$

Find an UE of (i) $\frac{1}{\theta}$, (ii) θ.

Solution:- i) Let $Z_i = -2\theta \ln X_i$, then $X_i = e^{-\frac{Z_i}{2\theta}}$

The PDF of Z_i is,

$$f_{Z_i}(z_i) = \begin{cases} \theta \left(e^{-\frac{z_i}{2\theta}}\right)^{\theta - 1} \left| \frac{d}{dz_i}\left(e^{-\frac{z_i}{2\theta}}\right)\right|, & \text{if } 0 < z_i < \infty \\ 0, & \text{OW} \end{cases}$$

$$= \begin{cases} \frac{1}{2} e^{-z_i/2}, & 0 < z_i < \infty \\ 0, & \text{OW} \end{cases}$$

$$\Rightarrow Z_i \overset{iid}{\sim} \chi^2_2 \quad \forall\, i = 1(1)n.$$

$$\Rightarrow \sum_{i=1}^{n} Z_i \sim \chi^2_{2n}$$

i.e. $$Y_i = \sum_{i=1}^{n} (-2\theta \ln X_i) \sim \chi^2_{2n}$$

Now, $$E\left(\sum_{i=1}^{n} -2\theta \ln X_i\right) = 2n$$

$$\Rightarrow E\left(-\frac{1}{n}\sum_{i=1}^{n} \ln X_i\right) = \frac{1}{\theta}$$

$$\Rightarrow T_1 = \frac{1}{n}\sum_{i=1}^{n} -\ln X_i \text{ is an UE of } \frac{1}{\theta}.$$

ii) Now, $$E\left(\frac{1}{Y}\right) = E\left(\frac{1}{\chi^2_{2n}}\right) = 2^{-1} \frac{\Gamma\left(\frac{2n}{2} - 1\right)}{\Gamma\left(\frac{2n}{2}\right)} \text{ if } n > 1$$

$$= \frac{1}{2} \cdot \frac{\Gamma(n-1)}{\Gamma(n)} = \frac{1}{2(n-1)}, \; n > 1.$$

$$\Rightarrow E\left(\frac{1}{\sum_{i=1}^{n} -2\theta \ln X_i}\right) = \frac{1}{2(n-1)}, \; n > 1$$

$$\Rightarrow E\left(\frac{n-1}{\sum_{i=1}^{n} -\ln X_i}\right) = \theta, \; n > 1.$$

$$\Rightarrow T_2 = \frac{n-1}{\sum_{i=1}^{n} -\ln X_i} \text{ is an UE of } \theta.$$

Ex. 8. Unbiased estimator may sometimes be absurd.

Give an example of Absurd Unbiased estimator.

Solution:- Let X be a single observation of $P(\lambda)$. If possible, let, $T(X)$ be an UE of $e^{-3\lambda}$.

Then $E[T(X)] = e^{-3\lambda}$, $\forall \lambda > 0$

$$\Rightarrow \sum_{x=0}^{\infty} T(x) \cdot e^{-\lambda} \cdot \frac{\lambda^x}{x!} = e^{-3\lambda}$$

$$\Rightarrow \sum_{x=0}^{\infty} T(x) \cdot \frac{\lambda^x}{x!} = e^{-2\lambda} = \sum_{x=0}^{\infty} \frac{(-2\lambda)^x}{x!}, \quad \lambda > 0$$

By uniqueness of Power series, we have

$$\frac{T(x)}{x!} = \frac{(-2)^x}{x!} \quad \forall \; x = 0, 1, 2, \ldots$$

$$\Rightarrow T(x) = (-2)^x \quad \forall \; x = 0, 1, 2, \ldots$$

Hence, $T(x) = (-2)^x$ is the unique UE of $e^{-3\lambda}$.

N.T. $T(x) = (-2)^x = \begin{cases} 2^x, & x = 0, 2, 4, \ldots \\ -2^x, & x = 1, 3, 5, \ldots \end{cases}$

Hence, $T(X)$ is UE but it takes negative values in estimating a positive parameter $e^{-3\lambda}$. This is an example of absurd UE.

Remark:- (1) Here $T(X) = (-2)^X$ is the only or unique UE of $e^{-3\lambda}$. Hence, $T(X) = (-2)^X$ is the UMVUE of $e^{-3\lambda}$.

(2) For $X \sim P(\lambda)$, $P_X(t) = e^{\lambda(t-1)}$, $t \in R$

$$\Rightarrow E[t^X] = e^{\lambda(t-1)}, \quad t \in R$$

Put, $t = -2$,

$$E[(-2)^X] = e^{-3\lambda}.$$

Ex. 9. If $X \sim Bin(n, p)$, then show that only polynomial in p of degree $\leq n$ are unbiasedly estimable.

Solution:- [A parametric function $\psi(\theta)$ is unbiasedly estimable if $E\{T(X)\} = \psi(\theta)$, for some $T(X)$, $\forall \theta \in \Omega$.]

Let $\psi(p)$ be an unbiasedly estimable parametric function.

Then $\exists$ a statistic $T(X) \ni$

$$\psi(p) = E(T(X)) \quad \forall \; p \in (0,1)$$

$$= \sum_{x=0}^{n} T(x) \binom{n}{x} p^x (1-p)^{n-x}$$

$$= \sum_{x=0}^{n} T(x) \cdot \binom{n}{x} p^x \left\{ \sum_{k=0}^{n-x} \binom{n-x}{k} (-p)^k \right\}$$

$$= \sum_{x=0}^{n} \sum_{k=0}^{n-x} (-1)^k T(x) \binom{n}{x} \binom{n-x}{k} p^{x+k},$$ which is a polynomial in p of degree $\leq n$.

Remark:- N.T. (i) $\sqrt{p}$, (ii) $\frac{1}{p}$, (iii) e^p, (iv) $\log p$ are not polynomials and hence not unbiasedly estimable. If $X \sim B(1, p)$, then only linear function in p are unbiasedly estimable. Hence, p^2, a 2nd degree polynomial is not unbiasedly estimable.

Best Linear Unbiased Estimator (BLUE) :—

Let $X_1, X_2, \ldots, X_n$ be a r.s. from a population with mean μ and variance $\sigma^2 (< \infty)$. Then an estimator $T = \sum_{i=1}^{n} a_i X_i$ is called a linear estimator. A linear estimator $T = \sum_{i=1}^{n} a_i X_i$ is unbiased for μ

iff $E(T) = \mu \;\forall \mu$

iff $\left(\sum_{i=1}^{n} a_i\right)\mu = \mu \;\forall \mu$

iff $\sum_{i=1}^{n} a_i = 1$.

[The estimator ~~T~~, $T = \sum_{i=1}^{n} a_i e^{X_i}$ is not linear estimator; also, $T_3 = \bar{X}^2$, $T_4 = s^2$ are linear estimators.]

Definition:— A linear unbiased estimator $T = \sum_{i=1}^{n} a_i X_i$ with $\sum_{i=1}^{n} a_i = 1$ of μ that has the minimum variance among all linear unbiased estimators of μ, is called the BLUE of μ.

Theorem:— If $X_1, X_2, \ldots, X_n$ be a r.s. from a population with mean μ and variance σ^2, show that the sample mean $\bar{X}$ is the BLUE of μ. [WBSU'11]

Proof:— BLUE of μ is the estimator which has the minimum variance in the class $\mathcal{C} = \left\{T : T = \sum_{i=1}^{n} a_i X_i, \sum_{i=1}^{n} a_i = 1\right\}$ of all linear UEs of μ.

Note that, $Var(T) = \left(\sum_{i=1}^{n} a_i^2\right)\sigma^2$, as X_i's are iid and $\sum_{i=1}^{n} a_i = 1$.

To minimize $Var(T) = \sigma^2\left(\sum_{i=1}^{n} a_i^2\right)$ subject to $\sum_{i=1}^{n} a_i = 1$,

By C-S inequality,

$$\left(\sum_{i=1}^{n} a_i \cdot 1\right)^2 \leq \left(\sum_{i=1}^{n} a_i^2\right)\left(\sum_{i=1}^{n} 1^2\right)$$

$$\Rightarrow \sum_{i=1}^{n} a_i^2 \geq \frac{1}{n} \quad \text{as} \quad \sum_{i=1}^{n} a_i = 1.$$

N.T. with $\sum_{i=1}^{n} a_i = 1$, $\sum_{i=1}^{n} a_i^2$ attains its minimum

iff '=' holds in C-S inequality.

iff $a_i \propto 1 \;\forall i = 1(1)n$

iff $a_i = k \;\forall i = 1(1)n$

iff $a_i = \frac{1}{n} \;\forall i$ as $1 = \sum_{i=1}^{n} a_i = nk$

Hence, $T = \frac{1}{n}\sum_{i=1}^{n} X_i$ has the minimum variance among all linear UEs of μ.

$\Leftrightarrow T = \bar{X}$ is the BLUE of μ.

Ex.1. Let $X_1, X_2, \ldots, X_n$ be n independent variables with common mean μ and variances $\sigma_i^2 = V(X_i)$, $i=1(1)n$. Find the BLUE of μ.

Solution:- To find an estimator T such that it has the minimum variance in the class $\mathcal{C} = \{T : T = \sum_{i=1}^n a_i X_i, \sum_{i=1}^n a_i = 1\}$ of all UEs of μ.

Note that $Var(T) = \sum_{i=1}^n a_i^2 \sigma_i^2$ where $\sum_{i=1}^n a_i = 1$.

By C-S inequality,

$$\left(\sum_{i=1}^n a_i \cdot \sigma_i \cdot \frac{1}{\sigma_i}\right)^n \le \left(\sum_{i=1}^n a_i^2 \sigma_i^2\right)\left(\sum_{i=1}^n \frac{1}{\sigma_i^2}\right)$$

$$\Rightarrow \sum_{i=1}^n a_i^2 \sigma_i^2 \ge \frac{1}{\sum_{i=1}^n \frac{1}{\sigma_i^2}}, \text{ as } \sum_{i=1}^n a_i = 1$$

Now, $\sum_{i=1}^n a_i = 1$, $\sum_{i=1}^n a_i^2 \sigma_i^2$ attains its minimum value iff '=' holds in cauchy-schwartz inequality,

iff $a_i \sigma_i \propto \frac{1}{\sigma_i}$

iff $a_i = \frac{k}{\sigma_i^2}$ $\forall i$

iff $a_i = \left(\frac{1}{\sigma_i^2}\right) \Big/ \left(\sum_{i=1}^n \frac{1}{\sigma_i^2}\right)$ $\quad \left[\because 1 = \sum_{i=1}^n a_i = k \cdot \sum_{i=1}^n \frac{1}{\sigma_i^2} \Rightarrow k = \frac{1}{\sum_{i=1}^n \frac{1}{\sigma_i^2}}\right]$

Hence, $T = \dfrac{\sum_{i=1}^n \frac{1}{\sigma_i^2} \cdot X_i}{\sum_{i=1}^n \frac{1}{\sigma_i^2}}$ is the BLUE of μ.

Ex.2. Let $X_1, X_2, \ldots, X_n$ be a r.s. from a popl^n with mean μ and variance σ^2. Suggest two UEs based on all X_i's and compare their performances.

Solution:- Note that any weighted average of X_i's is an UE of μ based on all X_i's.

$$T = \frac{\sum_{i=1}^n w_i X_i}{\sum_{i=1}^n w_i} \text{ is an UE of } \mu.$$

(i) $T_1 = \frac{1}{n}\sum_{i=1}^n X_i$, $\quad$ (ii) $T_2 = \dfrac{\sum_{i=1}^n i X_i}{\frac{n(n+1)}{2}}$

Now, $V(T_1) = \frac{\sigma^2}{n}$, and $Var(T_2) = \frac{4}{\{n(n+1)\}^2} \cdot \sum_{i=1}^n i^2 \sigma^2$

$$= \frac{4\sigma^2}{\{n(n+1)\}^2} \cdot \frac{n(n+1)(2n+1)}{6}$$

$$= \frac{2(2n+1)}{3n(n+1)} \cdot \sigma^2 = \frac{\sigma^2}{n}\left(\frac{4n+2}{3n+3}\right) > \frac{\sigma^2}{n}$$

$\therefore Var(T_2) > Var(T_1)$

Hence, T_1 has smaller variance than T_2 and T_1 is better than T_2. Infact $T_1 = \overline{X}$ is the **BLUE** of μ.

Method of finding Estimators :—

(I) Method of Moments : [The substitution Principle] (Due to Karl Pearson)

One of the oldest and simplest method of estimation is the method of moments or the substitution principle. Let $f(x, \theta_1, \theta_2, \ldots, \theta_k)$ be the PDF or PMF of the given popln., whose moments μ_r', $r = 1(1)k$, exists. Then, in general, μ_r' will be the function of $\theta_1, \theta_2, \ldots, \theta_k$. Let $X_1, X_2, \ldots, X_n$ be a r.s. from the given popln..
Define, $m_r' = \frac{1}{n}\sum_{i=1}^{n} X_i^r$ as the r^{th} order sample raw moment.

The method of moments consists in equating the k sample moments m_r', with the corresponding population moments μ_r' and solving k equations for k unknowns

$$\mu_r'(\theta_1, \theta_2, \ldots, \theta_k) = m_r', \quad r = 1(1)k.$$

$$\Rightarrow \theta_i = h(m_1', m_2', \ldots, m_k'), \quad i = 1(1)k.$$

Then, by method of moments,

$\hat{\theta}_i = h_i(m_1', \ldots, m_k')$ is the required estimator θ_i, $i = 1(1)k$.

This method is quite reasonable if the sample is a good representation of the population.

Rational behind the Method of Moments :—

Note that X_i's are iid RVs.

$\Leftrightarrow X_i^r$'s are iid RVs.

Hence, by Khinchin's WLLN,

$$\frac{1}{n}\sum_{i=1}^{n} X_i^r \xrightarrow{P} E(X_1^r), \text{ provided } \mu_r' = E(X_1^r) \text{ exists.}$$

$$\Leftrightarrow m_r' \xrightarrow{P} \mu_r', \text{ provided } \mu_r' \text{ exists.}$$

Again, $E(m_r') = \mu_r'$

$\Rightarrow m_r'$ is an UE of μ_r'.

It can be shown that, under general conditions, m_r' are asymptotically normal. Based on the above facts, we can equate m_r' to μ_r', quite reasonable.

Remark :- Method of moments may lead to absurd estimators. If we are asked to compute estimators of θ in $N(\theta, \theta)$ or, $N(\theta, \theta^2)$ by the method of moments, then we can verify this assertion.

Example :- Let $X_1, X_2, \ldots, X_n$ be a r.s. from $P(\lambda)$.

Note that, $E(X_i) = \lambda = V(X_i)$

By method of moments,

$$\mu_1' = m_1' \;;\; \mu_2' = m_2'$$

$$\therefore \mu_2' - \mu_1' = m_2' - m_1'$$

$$\Leftrightarrow \lambda = \overline{X} \text{ and } \lambda = m_2 \text{ or } s^2$$

The method of moments leads to using either $\overline{X}$ or s^2, as an estimator of λ.

To avoid ambiguity, we take the estimator involving the lowest order sample moments.

Ex.1. Let X_i's be the r.s from Geometric (p) $\forall\, i = 1(1)n$. Find an MME of the parameter. Comment on the quality of estimator.

Solution:-

By Method of moments,

$$\mu_1' = \overline{X} \Rightarrow \frac{1}{p} = \overline{X}$$

An MME of p is $\hat{p} = \frac{1}{\overline{X}}$

Note that, $0 < \hat{p} = \frac{1}{\overline{X}} \leq 1$

$$\Rightarrow \hat{p} = \frac{1}{\overline{X}} \in \Omega = (0,1)$$

and $E(\hat{p}) = E\left(\frac{1}{\overline{X}}\right) > \frac{1}{E(\overline{X})} = \frac{1}{1/p} = p$.

$\Rightarrow \hat{p}$ is the unbiased estimator.

Ex.2. Let X_i's $(i = 1(1)n)$ be a r.s. from $B(\alpha, \alpha)$ of 1st kind. Find an MME of α and comment on the quality of the estimator.

Ex.3. Find the estimator for λ by the method of moments in the exponential distribution [WBSU'11]

$$f(x,\lambda) = \frac{1}{\lambda} e^{-x/\lambda}, \quad \lambda > 0, x > 0$$
$$= 0, \quad \text{otherwise}$$

Solution:-

For exponential distribution,

$$\mu_1' = E(X) = \int_0^\infty x \cdot \frac{1}{\lambda} e^{-x/\lambda}\, dx$$
$$= \lambda$$

Now, the sample moment m_1' is given by

$$m_1' = \frac{1}{n} \sum X_i = \overline{X}$$

Equating μ_1' and m_1', we get

$$\hat{\lambda} = \overline{X}$$

(II) Method of Least Squares:- Let $y = f(x, \theta_1, \theta_2, \ldots, \theta_k)$ be the approximate regression equations of Y on X, which is assumed to be linear in parameters $\theta_1, \theta_2, \ldots, \theta_k$.

Let (x_i, y_i), $i = 1(1)n$, be an observed data on (X, Y). Define,

$e_i = y_i - f(x_i, \theta_1, \theta_2, \ldots, \theta_k)$ as the error in the prediction.

For a r.s. (X_i, Y_i), $i = 1(1)n$, we assume that

$$E_i = Y_i - f(x_i, \theta_1, \ldots, \theta_k)$$
$$\sim N(0, \sigma^2), \text{ where } \sigma^2 \text{ is constant.}$$

Then the likelihood of the observed errors $e_1, e_2, \ldots, e_n$ is $L(e_1, \ldots, e_n : \theta_1, \ldots, \theta_k) = \left(\frac{1}{\sigma\sqrt{2\pi}}\right)^n e^{-\frac{1}{2\sigma^2}\sum_{i=1}^{n} e_i^2}$

The observed sample $\{(x_i, y_i) : i = 1(1)n\}$ may be regarded as the most likely or most probable.

Hence the observed error $(e_1, e_2, \ldots, e_n)$ is also most probable.

Hence, we shall maximize the likelihood L w.r.t. $\theta_1, \theta_2, \ldots, \theta_k$

Now, maximizing L is equivalent to minimizing $\sum_{i=1}^{n} e_i^2$

$$= \sum_{i=1}^{n} \{y_i - f(x_i, \theta_1, \theta_2, \ldots, \theta_k)\}^2.$$

Hence, the principle of least squares consist in minimizing the sum of squares of errors w.r.t. the parameters $\theta_1, \theta_2, \ldots, \theta_k$.

It can be shown that the least squares estimators are the solutions of $\frac{\partial}{\partial \theta_i} \sum_{i=1}^{n} \{Y_i - f(x_i, \theta_1, \ldots, \theta_k)\}^2 = 0 \ \forall\ i = 1(1)k.$

Ex.1. If $Y \sim N\left(\beta x_i, \frac{\sigma^2}{x_i}\right)$ when $X = x_i$, $i = 1(1)n$, find the LSE of β based on the r.s. (x_i, Y_i).

Solution:- Here $Y/X = x_i \sim N\left(\beta x_i, \frac{\sigma^2}{x_i}\right)$ ~~when $X = x_i$~~.

$$\Rightarrow E(Y/X = x_i) = \beta x_i \ \forall\ i = 1(1)n.$$

Note, $e_i = Y_i - \beta x_i \sim N\left(0, \frac{\sigma^2}{x_i}\right)$, when $X = x_i$.

$$\Rightarrow e_i\sqrt{x_i} \sim N(0, \sigma^2)$$

To maximize $L = \frac{1}{\left(2\pi\frac{\sigma^2}{x_i}\right)^{n/2}} \cdot e^{-\frac{1}{2}\frac{\Sigma e_i^2}{\sigma^2/x_i}}$

i.e. to minimize $\sum_{i=1}^{n} e_i^2 x_i$.

Normal equation is : $\frac{\partial}{\partial\beta}\left\{\sum_{i=1}^{n}(y_i-\beta x_i)^2 x_i\right\}=0$

$$\Rightarrow 2\sum_{i=1}^{n}(y_i-\beta x_i)(-x_i^2)=0$$

$$\Rightarrow \Sigma x_i^2 y_i = \beta\Sigma x_i^3$$

$$\Rightarrow \beta = \frac{\Sigma x_i^2 y_i}{\Sigma x_i^3}.$$

Ex.2 When $X=x_i$, then $E(Y_i)=\beta x_i$ and $Var(Y_i)=\sigma^2$ $\forall$ $i=1(1)n$. Define $\hat{\beta}=\frac{\Sigma x_i Y_i}{\Sigma x_i}$, show that $\sum_{i=1}^{n}(Y_i-\beta x_i)^2 \geq \sum_{i=1}^{n}(Y_i-\hat{\beta}x_i)^2$. Also, show that $E(\hat{\beta})=\beta$ and $Var(\hat{\beta})=\frac{\sigma^2}{\Sigma x_i^2}$. If each Y_i follows normal distribution, S.T. $\hat{\beta}$ is a normal variable.

Solution:– Here $E(Y/X=x_i)=\beta x_i$. Then $e_i=Y_i-\beta x_i$ $\forall$ $i=1(1)n$.
By method of least squares, to minimize

$$\sum_{i=1}^{n}e_i^2=\sum_{i=1}^{n}(Y_i-\beta x_i)^2 \text{ w.r.t. } \beta,$$

Normal equation is : $\frac{\partial}{\partial\beta}\sum_{i=1}^{n}(Y_i-\beta x_i)^2=0$

$$\Rightarrow \beta=\frac{\Sigma x_i Y_i}{\Sigma x_i^2}=\hat{\beta}$$

Hence, $\Sigma(Y_i-\beta x_i)^2$ is minimum when $\beta=\hat{\beta}$.

$$\Rightarrow \Sigma(Y_i-\beta x_i)^2 \geq \Sigma(Y_i-\hat{\beta}x_i)^2$$

$$E(\hat{\beta})=E\left(\frac{\Sigma x_i y_i}{\Sigma x_i^2}\right)=\frac{\Sigma x_i E(Y_i)}{\Sigma x_i^2}=\frac{\Sigma x_i\cdot\beta x_i}{\Sigma x_i^2}=\beta.$$

$$\text{and } V(\hat{\beta})=V\left(\frac{\Sigma x_i y_i}{\Sigma x_i^2}\right)=\frac{\Sigma x_i^2 V(Y_i)}{(\Sigma x_i^2)^2}=\frac{\sigma^2}{\Sigma x_i^2}$$

Note that, $\hat{\beta}=\sum_{i=1}^{n}\left(\frac{x_i}{\Sigma x_i^2}\right)Y_i$ is a linear combination of normal variables Y_i, $i=1(1)n$.

Hence, $\hat{\beta}\sim N\left(E(\hat{\beta}),V(\hat{\beta})\right) \Rightarrow \hat{\beta}\sim N\left(\beta,\frac{\sigma^2}{\Sigma x_i^2}\right)$

[Q.E.D.]

INTERVAL ESTIMATION

Introduction: In the theory of point estimation we have tried to estimate the value of the unknown parameter pointwise. But in the theory of point estimation it is far from our intention to point out an estimate of θ, the unknown parameter. Here we try to find an interval in which the parameter value is contained with certain degree of confidence. This interval contains the parameter value with a certain probability which is related to the level of significance of the testing problem whose acceptance region may have a relationship with that interval. This interval is termed as confidence interval and the probability for containing the parameter value is termed as the confidence coefficient.

Level of significance and confidence interval : →

Let us consider the following testing problem,
$H_0 : \theta = \theta_0$ vs $H_1 : \theta \neq \theta_0$. Here $x_1, x_2, \ldots, x_n$ be the random sample and W be the critical region.
Let the level of significance is assigned to be α

$$P[\text{Rejecting true } H_0] \leq \alpha$$

$$\Rightarrow P[\underset{\sim}{X} \in W / H_0] \leq \alpha$$

$$\Rightarrow 1 - P[\underset{\sim}{X} \in W / H_0] \geq 1 - \alpha$$

$$\Rightarrow P[\text{Accepting the null when it is true}] \geq 1 - \alpha$$

$$\Rightarrow P[\text{Containing the true value of the parameter}] \geq 1 - \alpha$$

Here the confidence coefficient is $100(1-\alpha)\%$.

Fundamental notation of confidence estimation : → So far we have considered a random variable or some function of it as the basis observable quantity. Let X be a random variable and a, b be two given positive real numbers then,

$$P(a < X < b) = P(a < X \text{ and } X < b) = P\left(b < \frac{bX}{a} \text{ and } X < b\right) = P\left(X < b < \frac{bX}{a}\right)$$

as if we know the distribution of X and the quantities a and b, then we can determine the probability $P(a < X < b)$. Consider the interval $I(X) = \left(X, \frac{bX}{a}\right)$. This is an interval with a random variable in the end points and hence it takes the values $\left(x, \frac{bx}{a}\right)$ whenever the random variable X takes the values of x. Thus $I(X)$ is a random quantity and is an example of a random interval. Note that $I(X)$ includes the value b with a certain fixed probability. In general, larger the length of the interval, the larger the coverage probability.

Confidence Interval :- Let $x_1, x_2, \ldots, x_n$ be a random sample of size n on a random variable X having distribution belonging to the family

$$\mathcal{H} = \{ f_\theta(x) : \theta \in \Theta \}$$

if $\underline{\theta}(\underline{x})$ and $\bar{\theta}(\underline{x})$ be two statistics $\ni$

$$P_\theta[\underline{\theta}(x) < \theta < \bar{\theta}(x)] \geq 1-\alpha .$$ then $(\underline{\theta}(\underline{x}), \bar{\theta}(\underline{x}))$ is called a confidence interval with confidence coefficient $(1-\alpha)$. Confidence interval means the region where true value of the parametric function lies.

- **Example:-** $X \sim N(\theta, \sigma^2)$; σ^2 is known.

Find a confidence interval of θ with confidence coefficient $(1-\alpha)$.

Ans:-

$$P_\theta\left(\left| \frac{\sqrt{n}(\bar{X}-\theta)}{\sigma} \right| > \tau_{\alpha/2} \right) = \alpha$$

$$\Rightarrow P_\theta\left[-\tau_{\alpha/2}\frac{\sigma}{\sqrt{n}} < \bar{x}-\theta < \tau_{\alpha/2}\frac{\sigma}{\sqrt{n}} \right] = 1-\alpha$$

$$\Rightarrow P_\theta\left[\bar{x} - \tau_{\alpha/2}\frac{\sigma}{\sqrt{n}} < \theta < \bar{x} + \tau_{\alpha/2}\frac{\sigma}{\sqrt{n}} \right] = 1-\alpha .$$

$$\therefore \underline{\theta}(\underline{x}) = \bar{x} - \tau_{\alpha/2}\frac{\sigma}{\sqrt{n}} \text{ and } \bar{\theta}(\bar{x}) = \bar{x} + \tau_{\alpha/2}\frac{\sigma}{\sqrt{n}} .$$

Hence $100(1-\alpha)\%$ confidence interval of θ is given by

$$\left[\bar{x} - \tau_{\alpha/2}\cdot\frac{\sigma}{\sqrt{n}} , \bar{x} + \tau_{\alpha/2}\cdot\frac{\sigma}{\sqrt{n}} \right] .$$

Step to find out Confidence Interval: —

i) Give the critical region of the both tailed test at level α.

ii) Reverse the inequality sign and hence the RHS will be $(1-\alpha)$.

iii) From the inequality under probability solve for θ.

Problems Relating Confidence Interval

Ex. 1. Confidence Interval for the mean when the variance of normal distribution is known: —

→ Let us assume that we have a r.s. from Normal popln. with mean μ and variance σ^2. As we know ~~[illegible]~~ that the most efficient point estimator for the population mean μ is the sample mean $\bar{X}$, we can find a C.I. for μ by considering the sampling distribution of $\bar{X}$.

$$\bar{X} \sim N(\mu, \sigma^2/n)$$

and $$Z = \frac{\bar{X}-\mu}{\sigma/\sqrt{n}} \sim N(0,1)$$

so, that, $$f(z) = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}z^2}, \quad z \in \mathbb{R}$$

Now, let us assume that $z_{\alpha/2}$ be the value of Z such that

$$P(Z \geq z_{\alpha/2}) = \int_{z_{\alpha/2}}^{\infty} f(z)\,dz = \int_{z_{\alpha/2}}^{\infty} \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}z^2} dz = \alpha/2$$

and $z_{1-\alpha/2} = -z_{\alpha/2}$ the value of Z such that

$$P(Z \leq -z_{\alpha/2}) = \int_{-\infty}^{-z_{\alpha/2}} f(z)\,dz = \int_{-\infty}^{-z_{\alpha/2}} \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}z^2} dz = \alpha/2$$

Thus, clearly $P(-z_{\alpha/2} \leq Z \leq z_{\alpha/2}) = 1-\alpha$

$$\text{or, } P\left(-z_{\alpha/2} \leq \frac{\bar{X}-\mu}{\sigma/\sqrt{n}} \leq z_{\alpha/2}\right) = 1-\alpha$$

$$\text{or, } P\left(-z_{\alpha/2}\cdot\frac{\sigma}{\sqrt{n}} \leq \bar{X}-\mu \leq z_{\alpha/2}\cdot\frac{\sigma}{\sqrt{n}}\right) = 1-\alpha$$

$$\text{or, } P\left(\bar{X} - z_{\alpha/2}\cdot\frac{\sigma}{\sqrt{n}} \leq \mu \leq \bar{X} + z_{\alpha/2}\cdot\frac{\sigma}{\sqrt{n}}\right) = 1-\alpha$$

Thus, the $(1-\alpha)\%$ confidence interval for μ in $N(\mu, \sigma^2)$ is

$$\boxed{\bar{X} - z_{\alpha/2}\cdot\frac{\sigma}{\sqrt{n}} \leq \mu \leq \bar{X} + z_{\alpha/2}\cdot\frac{\sigma}{\sqrt{n}}}$$

Ex. 2. Confidence interval for the mean when variance of the normal population is not known:

→ If variance is not known, then σ is replaced by s, where $s^2 = \frac{1}{n-1}\sum(x_i-\bar{x})^2$

In this case, we use the t-statistic defined as

$$t = \frac{\bar{x}-\mu}{s/\sqrt{n}} \sim t_{n-1}$$

Now,
$$P(-t_{\alpha/2} \le t \le t_{\alpha/2}) = \int_{-\alpha/2}^{\alpha/2} f(t, n-1)\,dt = 1-\alpha$$

$$\Rightarrow P\left(-t_{\alpha/2} \le \frac{\bar{x}-\mu}{s/\sqrt{n}} \le t_{\alpha/2}\right) = 1-\alpha$$

$$\text{or, } P\left(\bar{x} - t_{\alpha/2}\cdot\frac{s}{\sqrt{n}} \le \mu \le \bar{x} + t_{\alpha/2}\cdot\frac{s}{\sqrt{n}}\right) = 1-\alpha$$

Thus, $(1-\alpha)100\%$ confidence interval for μ is

$$\boxed{\bar{x} - t_{\alpha/2}\cdot\frac{s}{\sqrt{n}} \le \mu \le \bar{x} + t_{\alpha/2}\cdot\frac{s}{\sqrt{n}}}$$

Example:- 1. Obtain 95% confidence intervals for mean of a normal distn with known variance σ^2.

Ans:-
$$\bar{x} \sim N(\mu, \sigma^2/n)$$

thus
$$Z = \frac{\bar{x}-\mu}{\sigma/\sqrt{n}} \sim N(0,1)$$

Also,
$$P(-z_{\alpha/2} \le Z \le z_{\alpha/2}) = P\left(\bar{x} - z_{\alpha/2}\frac{\sigma}{\sqrt{n}} \le \mu \le \bar{x} + z_{\alpha/2}\frac{\sigma}{\sqrt{n}}\right) = 1-\alpha \text{ (say)}$$

for $\alpha = 0.05$, $z_{\alpha/2} = 1.96$

Then we have,

$$P\left(\bar{x} - 1.96\frac{\sigma}{\sqrt{n}} \le \mu \le \mu + 1.96\frac{\sigma}{\sqrt{n}}\right) = 0.95$$

and thus,

$\left(\bar{x} - 1.96\frac{\sigma}{\sqrt{n}}, \bar{x} + 1.96\frac{\sigma}{\sqrt{n}}\right)$ is a confidence interval for μ, with a confidence coefficient 0.95.

Ex. 2. Find 95% confidence interval for exponential distribution with p.d.f. $f(x) = \theta e^{-\theta x}$, $0 \le x \le \infty$, $\theta > 0$,

Ans:- $E(X) = \frac{1}{\theta}$, $V(X) = \frac{1}{\theta^2}$

thus, $E(\bar{X}) = \frac{1}{\theta}$,

$$V(\bar{X}) = \frac{1}{n^2} V(X_1 + X_2 + \cdots + X_n)$$
$$= \frac{1}{n^2} \cdot n V(X)$$
$$= \frac{1}{n\theta^2}$$

Using CLT for large n, we have

$$Z = \frac{\hat{\theta} - E(\hat{\theta})}{\sqrt{V(\hat{\theta})}} \sim N(0,1)$$

i.e. $$Z = n\left(\frac{1}{\theta} - \bar{X}\right) \sim N(0,1)$$

$$\Rightarrow \sqrt{n}(1 - \theta\bar{X}) \sim N(0,1)$$

Hence, 95% confidence limits for θ are given by

$$P\left[-1.96 \le \sqrt{n}(1-\theta\bar{X}) \le 1.96\right] = 0.95$$

Now, $\sqrt{n}(1-\theta\bar{X}) \le 1.96$

$$\Rightarrow \left(1 - \frac{1.96}{\sqrt{n}}\right)\frac{1}{\bar{X}} \le \theta \quad \text{——— (1)}$$

and

$$-1.96 \le \sqrt{n}(1-\theta\bar{X})$$

$$\Rightarrow \theta \le \left(1 + \frac{1.96}{\sqrt{n}}\right)\frac{1}{\bar{X}} \quad \text{——— (2)}$$

Hence from (1) & (2), the 95% C.I. is given by

$$\theta = \left(1 \pm \frac{1.96}{\sqrt{n}}\right) \cdot \frac{1}{\bar{X}}$$

STATISTICAL HYPOTHESIS TESTING

Introduction:- Now we begin the study of statistical problem that forms the problem of hypothesis testing. As the term suggests, one wishes to decide whether or not some hypothesis that has been formulated is correct. The choice here lies between only two decission: accepting or rejecting the hypothesis. A decision procedure for such a problem is called a test of hypothesis.

In experimental research, our object is sometimes merely to estimate parameters. Thus one may wish to estimate the yield of a new hybrid line of corn. But more often the ultimate purpose will involve some use of the estimate. One may wish, for example, to compare the yield of the new line with that of the standard line and perhaps recommend that the newline replace the standard line if it appears superior. This is a common situation in research. The notion of hypotheses testing has been developed from these phenomenons.

The testing of hypotheses is seen to be closely related to the problem of estimation. It will be instructive, however to develop the theory of testing independently of the theory of estimation, at least in the begining.

DEFINITION: Statistical hypothesis

A statistical hypothesis is an assertion or conjecture about the distribution of one or more random variables. 'Accepting or bejecting the hypthesis' are the two decisions in our hand, so here we define a procedure to be termed as test of a statistical hypothesis.

Statistical hypothesis are classified as follows:

if Θ_0 is a single point in Θ, then the hypothesis is said to be simple hypothesis, otherwise it is called composite.

- **Testing of Hypothesis :—** In parametric testing of hypothesis, we start with a family of distribution.

$$\mathcal{H} = \{ f_\theta(x) : \theta \in \Theta \}$$

which is known except the parameter θ and here our object is to verify whether the value of θ lies in a specified subset Θ_0 of Θ. Here Θ is called the parameter space, to perform the test we are guided by a random sample $(X_1, X_2, \ldots, X_n)$ of some fixed size n. Here $\underset{\sim}{x} = (x_1, x_2, \ldots, x_n)$ is called a sample point and the all possible sample points together constitute a set, called sample space, denoted by $\mathcal{X}$.

To test for a hypothesis $H_0 : \theta \in \Theta_0$ agains $H : \theta \notin \Theta_0$, we divide the sample space $\mathcal{X}$ into two disjoint parts :— one is called the critical region (or, rejection region), denoted by W and the other is called the acceptance region, denoted by A (or, W^c).

If the observed sample point $\underset{\sim}{x} = (x_1, x_2, \ldots, x_n) \in W$ then we reject the hypothesis $H_0 : \theta \in \Theta_0$ against $H : \theta \notin \Theta_0$

Now, in the process of developing the test rule, we can commit two types of errors :

(i) Rejection of a true hypothesis, called the type-I error.

(ii) Acceptance of a wrong hypothesis, called the type-II error.

While constructing a critical region, care should be taken, so that both the types of errors stated above remains undercontrol. But unfortunately it is not possible to minimize both the kinds of error simultaneously. So, the usual practice is to minimize the 2nd kind of error for a fixed level of the first kind.

Now, in order to talk about the testing of hypothesis, we need to introduce some notations and definitions.

Some notations & definitions:—

Definition of Null hypothesis:— Let $\underset{\sim}{X} = (X_1, \dots, X_n)$ be a random observable sample drawn from same members of the family of distribution

$$\mathcal{F} = \{ f_\theta(x) : \theta \in \Theta \}$$

A null hypothesis is a statement about the unknown parameter θ, which is framed from our existing belief, or, past experience.
A null hypothesis is usually denoted by, $H_0 : \theta \in \Theta_0 \subset \Theta$.

Any hypothesis neglects (deny) the null hypothesis is called alternative hypothesis, such hypothesis is denoted by,

$$H_A : \theta \in \Theta_1 \; [\neq \Theta_0]$$

Probability of TYPE-I and TYPE-II error:— For a family of distribution $\mathcal{F} = \{ F_\theta(x) : \theta \in \Theta \}$, while testing the null hypothesis, $H_0 : \theta \in \Theta_0$ against the alternative $H_A : \theta \in \Theta$.
We can commit the following two errors:—

i) We can wrongly reject a true null hypothesis which is called error of type-I, denoted by E_I.

ii) We can wrongly accept a false null hypothesis which is called error of type-II, denoted by E_{II}.

Now, $P[E_I]$ = Probability of error I is denoted by α, and
$P[E_{II}]$ = Probability of error II is denoted by β.

<u>Note</u>:— $P(E_I) = P(\underset{\sim}{X} \in W / \theta \in \Theta_0) = \alpha$, and
$P(E_{II}) = \beta$

So, while constructing a critical region W, we shall try to minimize β or, maximize $(1-\beta) = P[\underset{\sim}{X} \in W / \theta \in \Theta]$, called the power of critical region for a preassigned value of α.

Level of significance and size of a critical region:— Let us consider the family of distribution,

$$\mathcal{F} = \{ F_\theta(x) : \theta \in \Theta \}$$

and let $\underset{\sim}{X} = (X_1 \dots X_n)$ be a random sample drawn from a member of this family. For testing the null hypothesis

$$H_0 : \theta \in \Theta_0 \quad vs \quad H : \theta \in \Theta$$

a critical region W_0 is said to be of level α if,

$$P_\theta(\underset{\sim}{X} \in W_0) \leq \alpha \;\; \forall \; \theta \in \Theta_0$$

In this case, $\sup_{\theta \in \Theta_0} P(\underset{\sim}{X} \in W_0)$ is called the size of the test.

Power of a critical region :— Let us consider $\underset{\sim}{X} = (x_1, \cdots, x_n)$ be a random sample on a random variable having a distribution belonging to the family

$$\mathcal{F} = \{F_\theta : \theta \in \Theta_0\}$$

then for testing $H_0 : \theta \in \Theta_0$ Vs $H : \theta \in \Theta_1$

$P_\theta(W_0) = P_\theta(\underset{\sim}{X} \in W_0)$ for $\theta \in \Theta_1$ is called the critical region of the test at the point θ.

While constructing a critical region, care should be taken so, that $P_\theta(W_0)$ attains its maximum possible value for all $\theta \in \Theta$ and such a critical region if exists is called uniformly most powerful (UMP) critical region.

Uniformly most powerful critical region :— Let $\underset{\sim}{X} = (X_1, \cdots, X_n)$ be a random sample on an random variable X having a distribution belonging to the family

$$\mathcal{F} = \{F_\theta : \theta \in \Theta\}$$

then for testing $H_0 : \theta \in \Theta_0$ Vs $H : \theta \in \Theta$

a critical region W_0 is said to be UMP among the class of level α critical region if

$$P_\theta(W_0) \leq \alpha \ \forall \ \theta \in \Theta_0 \quad \cdots\cdots\cdots (1)$$

and $$P_\theta(W_0) \geq P_\theta(W) \ \forall \ \theta \in \Theta_1 \quad \cdots\cdots\cdots (2)$$

where, W is any other critical region satisfying (1).

Most Powerful (MP) critical region:— Let $\underset{\sim}{X} = (X_1, \dots, X_n)$ be a random sample on a random variable X having distribution belonging to the family,

$$\mathcal{F} = \{ F_\theta : \theta \in \Theta \}$$

Then for testing a simple null hypothesis,

$H_0 : \theta = \theta_0$ against a simple alternative hypothesis

$H_1 : \theta = \theta_1$

A critical region W_0 is said to be most powerful (MP) level α critical region if,

$$P_{\theta_0}(W_0) = \alpha \text{ and } \cdots\cdots (1)$$

$$P_{\theta_1}(W_0) \geq P_{\theta_1}(W) \cdots\cdots (2)$$

for any other critical region W satisfying (1).

Construction of Most Powerful critical region:— Let $X_1, X_2, \dots, X_n$ be jointly distributed random variables with joint PDF or PMF $f(\underset{\sim}{x})$ for testing,

$H_0 : \theta = \theta_0$ Vs

$H_1 : \theta = \theta_1$

$$W_0 = \left\{ (x_1, x_2, \dots, x_n) : \frac{f_{\theta_1}(\underset{\sim}{x})}{f_{\theta_0}(\underset{\sim}{x})} \geq k \right\}$$ is most powerful of its size.

Proof: (For continuous case only)

$$P_{\theta_1}(W_0) = P_{\theta_1}[\underset{\sim}{X} \in W_0] = \int_{W_0} f_{\theta_1}(\underset{\sim}{x})\, d\underset{\sim}{x} \quad [\text{its a multiple integral}]$$

$$= \int_{W_0 \cap W^c} f_{\theta_1}(\underset{\sim}{x})\, d\underset{\sim}{x} + \int_{W_0 \cap W} f_{\theta_1}(\underset{\sim}{x})\, d\underset{\sim}{x} \cdots\cdots (1)$$

and, $$P_{\theta_1}(W) = P_{\theta_1}[\underset{\sim}{X} \in W]$$

$$= \int_W f_{\theta_1}(\underset{\sim}{x})\, d\underset{\sim}{x} = \int_{W \cap W_0} f_{\theta_1}(\underset{\sim}{x})\, d\underset{\sim}{x} + \int_{W \cap W_0^c} f_{\theta_1}(\underset{\sim}{x})\, d\underset{\sim}{x} \cdots\cdots (2)$$

From (1) and (2), (1) − (2) gives —

$$P_{\theta_1}(W_0) - P_{\theta_1}(W) = \int_{W_0 \cap W^c} f_{\theta_1}(\underset{\sim}{x})\, d\underset{\sim}{x} - \int_{W \cap W_0^c} f_{\theta_1}(\underset{\sim}{x})\, d\underset{\sim}{x}$$

$$\geq k \int_{W_0 \cap W^c} f_{\theta_0}(\underset{\sim}{x})\, d\underset{\sim}{x} - k \int_{W \cap W_0^c} f_{\theta_0}(\underset{\sim}{x})\, d\underset{\sim}{x}$$

$$= k \left[\int_{W_0 \cap W^c} f_{\theta_0}(\underset{\sim}{x})\, d\underset{\sim}{x} + \int_{W_0 \cap W} f_{\theta_0}(\underset{\sim}{x})\, d\underset{\sim}{x} - \int_{W \cap W_0^c} f_{\theta_0}(\underset{\sim}{x})\, d\underset{\sim}{x} - \int_{W \cap W_0} f_{\theta_0}(\underset{\sim}{x})\, d\underset{\sim}{x} \right]$$

$$= k \left[\int_{W_0} f_{\theta_0}(\underset{\sim}{x})\, d\underset{\sim}{x} - \int_W f_{\theta_0}(\underset{\sim}{x})\, d\underset{\sim}{x} \right]$$

$$= k \left[P_{\theta_0}(W_0) - P_{\theta_0}(W) \right]$$

$$= 0 \quad [\because W \text{ and } W_0 \text{ are of same size}]$$

Example:- 1. Let, $X \sim N(\theta, \sigma^2)$, σ^2 is known. On the basis of a random sample $\underset{\sim}{X} = (X_1, \ldots, X_n)$, from the distribution of $\underset{\sim}{X}$. Find the most powerful critical region for testing, $H_0: \theta = \theta_0$ vs $H: \theta = \theta_1 (> \theta_0)$

Solution:- The joint PDF of $(X_1, \ldots, X_n)$ is,

$$f_\theta(\underset{\sim}{x}) = \left(\frac{1}{2\pi\sigma^2}\right)^{n/2} \exp\left[-\frac{1}{2\sigma^2}\sum_{i=1}^{n}(x_i - \theta)^2\right], \; x_i \in \mathbb{R} \; \forall\, i = 1(1)n.$$

By Neyman-Pearson lemma,
the most powerful (MP) critical region is given by,

$$W_0 = \left\{\underset{\sim}{x} : \frac{f_{\theta_1}(\underset{\sim}{x})}{f_{\theta_0}(\underset{\sim}{x})} \geq k\right\}$$

$$\frac{f_{\theta_1}(\underset{\sim}{x})}{f_{\theta_0}(\underset{\sim}{x})} = \exp\left[-\frac{1}{2\sigma^2}\left\{\sum_{i=1}^{n}(x_i - \theta_1)^2 - \sum_{i=1}^{n}(x_i - \theta_0)^2\right\}\right] > k$$

$$\Leftrightarrow -\frac{n}{2\sigma^2}\left[(\bar{x} - \theta_1)^2 + (\bar{x} - \theta_0)^2\right] > \ln k$$

$$\Leftrightarrow -\frac{n}{2\sigma^2}\left[(\theta_1^2 - \theta_0^2) - 2\bar{x}(\theta_1 - \theta_0)\right] > \ln k$$

$$\Leftrightarrow (\theta_1 - \theta_0)\frac{n\bar{x}}{2\sigma^2} > \ln k + \frac{n(\theta_1^2 - \theta_0^2)}{2\sigma^2}$$

$$\Leftrightarrow \bar{x} > c \quad [\because \theta_1 > \theta_0]$$

So, the MP critical region is given by,

$$W_0 = \{\underset{\sim}{x} : \bar{x} > c\}$$

where, constant c is such that,

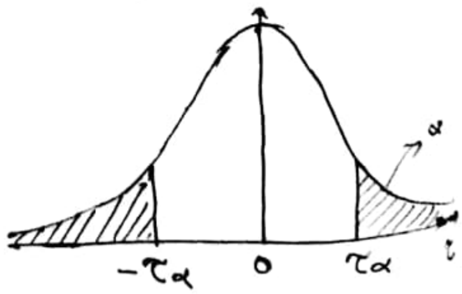
i.e. $P_{\theta_0}(\bar{x} > c) = \alpha$

i.e. $P_{\theta_0}\left\{\frac{\sqrt{n}(\bar{X} - \theta_0)}{\sigma} > \frac{\sqrt{n}(c - \theta_0)}{\sigma}\right\} = \alpha$

when, H_0 is true, i.e. when $X \sim N(\theta_0, \sigma^2)$

$$\bar{X} \sim N\left(\theta_0, \frac{\sigma^2}{n}\right)$$

i.e. $\tau = \frac{\sqrt{n}(\bar{X} - \theta_0)}{\sigma} \sim N(0,1)$



$$\Rightarrow P_\theta\left(\tau > \frac{\sqrt{n}(c - \theta_0)}{\sigma}\right) = \alpha$$

$$\Rightarrow \frac{\sqrt{n}(c - \theta_0)}{\sigma} = \tau_\alpha \longrightarrow \text{the upper } \alpha \text{ point of standard normal distribution.}$$

$$\Rightarrow c = \theta_0 + \frac{\sigma}{\sqrt{n}}\tau_\alpha$$

So, the most powerful region for testing
$H_0 : \theta = \theta_0$ Vs $H : \theta = \theta_1 (>\theta_0)$ is given by,

$$W_0 = \left\{ \underset{\sim}{X} : \bar{X} > \theta_0 + \frac{\sigma}{\sqrt{n}} \tau_\alpha \right\}$$

Note that, this critical region does not depend in any way on the value of θ_1 except for the fact that $\theta_1 > \theta_0$, so this critical region is actually uniformly most powerful (UMP) for testing $H_0 : \theta = \theta_0$ Vs $H : \theta > \theta_0$.

2. Let $X \sim N(\mu, \theta)$; μ is known on the basis of a random sample $\underset{\sim}{X} = (X_1, \ldots, X_n)$ from the distribution of $\underset{\sim}{X}$. Find the MP critical region for testing

$$H_0 : \theta = \theta_0 \text{ Vs}$$
$$H : \theta = \theta_1 (>\theta_0)$$

<u>Solution:-</u> The joint PDF of $(X_1, \ldots, X_n)$ is

$$f_\theta(\underset{\sim}{x}) = \left(\frac{1}{2\pi\theta}\right)^{n/2} \exp\left[-\frac{1}{2\theta}\sum_{i=1}^{n}(x_i-\mu)^2\right], \quad x_i \in \mathbb{R} \ \forall\, i = 1(1)n.$$

By Neyman-Pearson lemma the MP critical region is given by

$$W_0 = \left\{ \underset{\sim}{x} : \frac{f_{\theta_1}(\underset{\sim}{x})}{f_{\theta_0}(\underset{\sim}{x})} \geq k \right\}$$

$$\therefore \frac{f_{\theta_1}(\underset{\sim}{x})}{f_{\theta_0}(\underset{\sim}{x})} = \exp\left[-\frac{1}{2}\left(\frac{1}{\theta_1} - \frac{1}{\theta_0}\right)\sum_{i=1}^{n}(x_i-\mu)^2\right] \cdot \left(\frac{\theta_0}{\theta_1}\right)^{n/2} > k$$

$$\Leftrightarrow \frac{n}{2}\ln\left(\frac{\theta_0}{\theta_1}\right) - \frac{1}{2}\left(\frac{1}{\theta_1} - \frac{1}{\theta_0}\right)\sum_{i=1}^{n}(x_i-\mu)^2 > \ln k$$

$$\Leftrightarrow \left(\frac{1}{\theta_0} - \frac{1}{\theta_1}\right)\sum_{i=1}^{n}(x_i-\mu)^2 > k'$$

$$\Leftrightarrow \sum_{i=1}^{n}(x_i-\mu)^2 > \frac{k'}{\left(\frac{1}{\theta_0} - \frac{1}{\theta_1}\right)} = c \text{ (say)} \qquad \left[\because \theta_1 > \theta_0 \ \therefore \frac{1}{\theta_0} - \frac{1}{\theta_1} > 0\right]$$

So, the MP critical region is given by,

$$W_0 = \left\{ \underset{\sim}{x} : \sum_{i=1}^{n}(x_i-\mu)^2 > c \right\}$$, where, c is determined $\ni$

$$P_{\theta_0}(W_0) = \alpha$$

i.e. $P_{\theta_0}\left\{\sum_{i=1}^{n}(X_i-\mu)^2 > c\right\} = \alpha$

i.e. $P_{\theta_0}\left\{\frac{\sum_{i=1}^{n}(X_i-\mu)^2}{\theta_0} > \frac{c}{\theta_0}\right\} = \alpha$ ——— (*)

Now, under $H_0 : \theta = \theta_0$

$$\frac{\sum_{i=1}^{n}(X_i-\mu)^2}{\theta_0} \sim \chi^2_n$$

so from (*), $P_{\theta_0}\left(\chi^2_n > \frac{c}{\theta_0}\right) = \alpha$

$$\therefore \frac{c}{\theta_0} = \chi^2_{\alpha;n} \longrightarrow$$ the upper α point of χ^2 distn. with n degree of freedom.

So, $W_0 = \left\{ \underset{\sim}{x} : \sum_{i=1}^{n}(x_i-\mu)^2 > \theta_0 \chi^2_{\alpha;n} \right\}$

TEST OF SIGNIFICANCE FOR NORMAL DISTN

Problem:- Let $\underset{\sim}{X} = (X_1, \ldots, X_n)$ be a random sample on a random variable X having distribution $N(\theta, \sigma^2)$; θ and σ^2 unknown then for these testing problems:

i) $H_0 : \theta = \theta_0$ vs
$H : \theta > \theta_0$

ii) $H_0 : \theta = \theta_0$ vs
$H : \theta < \theta_0$

iii) $H_0 : \theta = \theta_0$ vs
$H : \theta \neq \theta_0$

Find the critical region.

Solution:- $\boxed{\theta\text{ - test}}$:- $\boxed{\sigma^2 \text{ unknown}}$:

i) At level of significance α, we reject H_0 against H if $\bar{X} > c$, i.e. if where c is such that $P_{\theta_0}(\bar{X} > c) = \alpha$

$$\text{or, } P_{\theta_0}\left(\frac{\sqrt{n}(\bar{X} - \theta_0)}{s} > \frac{\sqrt{n}(c - \theta_0)}{s}\right) = \alpha$$

where, $s^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \bar{X})^2$

Under $H_0 : \theta = \theta_0$.

$$T = \frac{\sqrt{n}(\bar{X} - \theta_0)}{s} = \frac{\sqrt{n}(\bar{X} - \theta_0)}{\sigma} \Big/ \sqrt{\frac{(n-1)s^2}{\sigma^2} \Big/ (n-1)}$$

$$= \frac{\tau}{\sqrt{\chi^2/(n-1)}}, \text{ where } \tau = \frac{\sqrt{n}(\bar{X} - \theta_0)}{\sigma} \sim N(0,1)$$

[Under H_0]

and $\frac{(n-1)s^2}{\sigma^2} \sim \chi^2_{n-1}$.

Also, τ is independent of χ^2 as $\bar{X}$ is independent of s^2. So, under $H_0 : \theta = \theta_0$

$$T \sim t_{n-1}$$

Thus, $P_{\theta_0}(T > c) = \alpha$

$$\Rightarrow P_{\theta_0}\left\{t_{n-1} > \frac{\sqrt{n}(c - \theta_0)}{s}\right\} = \alpha$$

$$\Rightarrow \frac{\sqrt{n}(c - \theta_0)}{s} = t_{\alpha; n-1}$$ → the upper α part of t distn with $(n-1)$ d.f.

So, $c = \theta_0 + \frac{s}{\sqrt{n}} t_{\alpha; n-1}$

i.e. at $100\alpha\%$ level we reject H_0 if

$$\bar{X} > \theta_0 + \frac{s}{\sqrt{n}} t_{\alpha; n-1}$$

(ii) We reject H_0 against H if $x < c'$ where c' is such that

$$P_{\theta_0}(\bar{X} < c') = \alpha$$

or, $$P_{\theta_0}\left(t_{n-1} < \frac{\sqrt{n}(c'-\theta_0)}{s}\right) = \alpha$$

i.e. $$\frac{\sqrt{n}(c'-\theta_0)}{s} = t_{1-\alpha, n-1} = -t_{\alpha, n-1}$$ [By symmetry of t distribution]

i.e. $$c' = \theta_0 - \frac{s}{\sqrt{n}} t_{\alpha; n-1}.$$

We reject H_0 if $\bar{x} < c' = \theta_0 - \frac{s}{\sqrt{n}} t_{\alpha; n-1}$.

(iii) We reject H_0 if $\bar{x} > c$ or $\bar{x} < c'$

where c & c' are determined $\ni$

$$P_{\theta_0}(\bar{x} < c' \text{ or } \bar{x} > c) = \alpha$$

i.e. $$P_{\theta_0}(c' < \bar{x} < c) = 1-\alpha$$

i.e. $$P_{\theta_0}\left(\frac{\sqrt{n}(c'-\theta_0)}{s} < t_{n-1} < \frac{\sqrt{n}(c-\theta_0)}{s}\right) = 1-\alpha$$

Solving the above equation for c and c' we can get infinitely many solution. However for convenience we choose c & c' $\ni$

$$P_{\theta_0}\left(t_{n-1} \leq \frac{\sqrt{n}(c'-\theta_0)}{s}\right) = \frac{\alpha}{2}$$

& $$P_{\theta_0}\left(t_{n-1} \geq \frac{\sqrt{n}(c-\theta_0)}{s}\right) = \frac{\alpha}{2}$$

i.e. $$\frac{\sqrt{n}(c'-\theta_0)}{s} = -t_{\alpha/2, n-1} \Rightarrow c' = \theta_0 - t_{\alpha/2, n-1}\frac{s}{\sqrt{n}}$$

and $$\frac{\sqrt{n}(c-\theta_0)}{s} = t_{\alpha/2, n-1} \Rightarrow c = \theta_0 + t_{\alpha/2, n-1}\frac{s}{\sqrt{n}}$$

We reject H_0 if $|\bar{x} - \theta_0| > t_{\alpha/2; n-1}\frac{s}{\sqrt{n}}$.

i.e. $|\bar{x}| > \theta_0 + t_{\alpha/2; n-1} \cdot \frac{s}{\sqrt{n}}$.

Note:- For the above test the test statistic is called student's 't' statistic and this test is called <u>student's 't' test.</u>

Problem:- Let $\underset{\sim}{X} = (X_1, \dots, X_n)$ be an r.s. on a RV X having distn. $N(\mu, \theta)$; μ and θ unknown. Then for these testing problems:-

i) $H_0: \theta = \theta_0$
Vs. $H: \theta > \theta_0$

ii) $H_0: \theta = \theta_0$
vs. $H: \theta < \theta_0$

iii) $H_0: \theta = \theta_0$
vs. $H: \theta \neq \theta_0$

Find the critical region for each cases.

ANS:- $\boxed{\text{variance test :-}}$ $\boxed{\mu \text{ unknown}}$:

i) At level of significance α, we reject H_0 against H if

$$\sum_{i=1}^{n}(X_i-\mu)^2 > c$$

i.e. if where c is $\ni$ $P_{\theta_0}\left[\sum_{i=1}^{n}(X_i-\mu)^2 > c\right] = \alpha$

or, $P_{\theta_0}\left[\sum_{i=1}^{n}(X_i-\bar{X})^2 > c\right] = \alpha$

where $\bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$, $\bar{X}$ is an unbiased estimator of μ.

$$\therefore P_{\theta_0}\left[\sum_{i=1}^{n}(X_i-\bar{X})^2 > c\right] = \alpha$$

$$\text{or, } P_{\theta_0}\left[(n-1)s^2 > c\right] = \alpha$$

$$\text{or, } P_{\theta_0}\left[\frac{(n-1)s^2}{\theta_0} > \frac{c}{\theta_0}\right] = \alpha$$

$$\text{or, } P_{\theta_0}\left[\chi^2_{n-1} > \frac{c}{\theta_0}\right] = \alpha$$

$$\therefore \frac{c}{\theta_0} = \chi^2_{\alpha;n-1} \longrightarrow \text{upper } \alpha \text{ point of } \chi^2_{n-1}.$$

$$\therefore c = \theta_0 \chi^2_{\alpha;n-1}$$

$\therefore$ We reject H_0 against H if $\sum_{i=1}^{n}(X_i-\bar{X})^2 > \theta_0 \chi^2_{\alpha;n-1}$.

ii) at level of significance α, we reject H_0 against H if,

$$\sum_{i=1}^{n}(X_i-\mu)^2 < c$$

i.e. if where c is such that,

$$P_{\theta_0}\left[\sum_{i=1}^{n}(X_i-\mu)^2 < c\right] = \alpha$$

$$\text{or, } P_{\theta_0}\left[\sum_{i=1}^{n}(X_i-\bar{X})^2 < c\right] = \alpha$$

$$\text{or, } P_{\theta_0}\left[\frac{(n-1)s^2}{\theta_0} < \frac{c}{\theta_0}\right] = \alpha \quad \text{or, } P_{\theta_0}\left[\chi^2_{n-1} < \frac{c}{\theta_0}\right] = \alpha$$

$\therefore \frac{c}{\theta_0} = \chi^2_{1-\alpha;n-1}$

or, $c = \theta_0 \chi^2_{1-\alpha;n-1}$

$\therefore$ We reject Ho if $\sum_{i=1}^{n}(x_i - \bar{x})^2 < \theta_0 \chi^2_{1-\alpha;n-1}$.

iii) At level of significance α, we reject Ho against H if,

$\sum_{i=1}^{n}(x_i - \mu)^2 < c_1$ or, $\sum_{i=1}^{n}(x_i - \mu)^2 > c_2$

i.e. where, c_1 and c_2 are such that,

$P_{\theta_0}\left[\sum_{i=1}^{n}(x_i - \mu)^2 < c_1 \text{ or, } \sum_{i=1}^{n}(x_i - \mu)^2 > c_2\right] = \alpha$

or, $P_{\theta_0}\left[\sum_{i=1}^{n}(x_i - \bar{x})^2 < c_1 \text{ or, } \sum_{i=1}^{n}(x_i - \bar{x})^2 > c_2\right] = \alpha$

there are infinitely many choice of c_1 and c_2 but we choose c_1 and c_2 such that,

$P_{\theta_0}\left[\sum_{i=1}^{n}(x_i - \bar{x})^2 < c_1\right] = \alpha/2$

or, $c_1 = \theta_0 \chi^2_{1-\alpha/2;n-1}$

$P_{\theta_0}\left[\sum_{i=1}^{n}(x_i - \bar{x})^2 > c_2\right] = \alpha/2$

or, $c_2 = \theta_0 \chi^2_{\alpha/2;n-1}$.

$\therefore$ We reject Ho if

$\sum_{i=1}^{n}(x_i - \bar{x})^2 < \theta_0 \chi^2_{1-\alpha/2;n-1}$ or, $\sum_{i=1}^{n}(x_i - \bar{x})^2 > \theta_0 \chi^2_{\alpha/2;n-1}$.

Two Sample Problem:

Ex:- Let X and Y be two independent normal variables such that $X \sim N(\mu_1, \sigma^2)$, $Y \sim N(\mu_2, \sigma^2)$ on the basis of 2 independent random sample,
$(X_1, X_2, \ldots, X_{n_1})$ and $(Y_1, Y_2, \ldots, Y_{n_2})$ of size n_1 and n_2 drawn from the distn. of X and Y respectively.
Test $H_0: \mu_1 = \mu_2$ against different alternatives.

Solution:- Let $\bar{X} = \frac{1}{n}\sum_{i=1}^{n_1} X_i$, and $\bar{Y} = \frac{1}{n_2}\sum_{i=1}^{n_2} Y_i$

Now, $\bar{X} \sim N(\mu_1, \sigma^2/n_1)$

$\bar{Y} \sim N(\mu_2, \sigma^2/n_2)$

Also, $\bar{X}$ is independent of $\bar{Y}$ ($\because$ X and Y are independent).

So, $\bar{X} - \bar{Y} \sim N\left(\mu_1 - \mu_2, \sigma^2\left(\frac{1}{n_1} + \frac{1}{n_2}\right)\right)$

To test for $H_0: \mu_1 - \mu_2 = 0$ against
$H_1: \mu_1 - \mu_2 > 0$
we reject H_0 if $\bar{X} - \bar{Y} > c$

Vs. $H_2: \mu_1 - \mu_2 < 0$.
we reject H_0 if $\bar{X} - \bar{Y} < c'$

Vs $H_3: \mu_1 - \mu_2 \neq 0$ we reject H_0 if
$\bar{X} - \bar{Y} < k_1$ or, $\bar{X} - \bar{Y} > k_2$.

The constants c, c', k_1 & k_2 are to be determined from the size condition of the test.

The critical region for $H_0: \mu_1 - \mu_2 = 0$ Vs.
$H_1: \mu_1 - \mu_2 > 0$
is $W_{01} = \{(X_1, X_2, \ldots, X_{n_1}, Y_1, Y_2, \ldots, Y_{n_2}) : \bar{X} - \bar{Y} > c\}$

$P_{H_0}(W_0) = \alpha \Rightarrow P_{H_0}(\bar{X} - \bar{Y} > c) = \alpha$

$$\Rightarrow P_{H_0}\left[\frac{\bar{X} - \bar{Y}}{\sigma\left(\frac{1}{n_1} + \frac{1}{n_2}\right)^{1/2}} > \frac{c}{\sigma\left(\frac{1}{n_1} + \frac{1}{n_2}\right)^{1/2}}\right] = \alpha$$

$$\Rightarrow P_{H_0}\left[\tau > \frac{c}{\sigma\left(\frac{1}{n_1} + \frac{1}{n_2}\right)^{1/2}}\right] = \alpha$$

$$\because \tau = \frac{\bar{X} - \bar{Y}}{\sigma\left(\frac{1}{n_1} + \frac{1}{n_2}\right)^{1/2}} \sim N(0,1), \text{ under } H_0.$$

so, $\frac{c}{\sigma\left(\frac{1}{n_1} + \frac{1}{n_2}\right)^{1/2}} = \tau_\alpha \rightarrow$ the upper α point of $N(0,1)$

$\therefore c = \tau_\alpha \sigma\left(\frac{1}{n_1} + \frac{1}{n_2}\right)^{1/2}$

So, $W_{01} = \left\{(\underset{\sim}{X}, \underset{\sim}{Y}) : \bar{X} - \bar{Y} > \tau_\alpha \sigma\left(\frac{1}{n_1} + \frac{1}{n_2}\right)^{1/2}\right\}$

Similarly,
the critical region for, $H_0 : \mu_1 - \mu_2 = 0$ Vs.
$H_2 : \mu_1 - \mu_2 < 0$ is

$$W_{02} = \left\{ (\underset{\sim}{X}, \underset{\sim}{Y}) : \overline{X} - \overline{Y} < \tau_{1-\alpha}\, \sigma \left(\frac{1}{n_1} + \frac{1}{n_2}\right)^{1/2} \right\}$$

$$= \left\{ (\underset{\sim}{X}, \underset{\sim}{Y}) : \overline{X} - \overline{Y} < -\tau_{\alpha}\, \sigma \left(\frac{1}{n_1} + \frac{1}{n_2}\right)^{1/2} \right\}$$

and
the critical region for, $H_0 : \mu_1 - \mu_2 = 0$ Vs.
$H_3 : \mu_1 - \mu_2 \neq 0$

$$W_{03} = \{ (\underset{\sim}{X}, \underset{\sim}{Y}) : \overline{X} - \overline{Y} < K_1 \text{ or } > K_2 \}$$

$$P_{H_0}(W_0) = \alpha$$

$$\Rightarrow P_{H_0}(\overline{X} - \overline{Y} < K_1 \text{ or, } \overline{X} - \overline{Y} > K_2) = \alpha \quad \text{———} (*)$$

How, under $H_0 : \mu_1 - \mu_2 = 0$

$$\tau = \frac{\overline{X} - \overline{Y}}{\sigma \left(\frac{1}{n_1} + \frac{1}{n_2}\right)^{1/2}} \sim N(0,1)$$

So, (*) will be $P_{H_0}\left[\tau < \dfrac{K_1}{\sigma\left(\frac{1}{n_1}+\frac{1}{n_2}\right)^{1/2}} \text{ or, } \tau > \dfrac{K_2}{\sigma\left(\frac{1}{n_1}+\frac{1}{n_2}\right)^{1/2}} \right] = \alpha$

So, satisfying (*) we can choose

$$\frac{K_1}{\sigma\left(\frac{1}{n_1} + \frac{1}{n_2}\right)^{1/2}} = -\tau_{\alpha/2}$$

and $$\frac{K_2}{\sigma\left(\frac{1}{n_1} + \frac{1}{n_2}\right)^{1/2}} = \tau_{\alpha/2}.$$

i.e. $$K_1 = -\tau_{\alpha/2}\, \sigma\left(\frac{1}{n_1} + \frac{1}{n_2}\right)^{1/2}$$

$$K_2 = \tau_{\alpha/2}\, \sigma\left(\frac{1}{n_1} + \frac{1}{n_2}\right)^{1/2}$$

So, the critical region is given by,

$$W_{03} = \left\{ (\underset{\sim}{X}, \underset{\sim}{Y}) : |\overline{X} - \overline{Y}| > \tau_{\alpha/2}\, \sigma \left(\frac{1}{n_1} + \frac{1}{n_2}\right)^{1/2} \right\}$$

Find W_{01}, W_{02}, W_{03} when σ is unknown

$$W_{01} = \{(\underset{\sim}{X}, \underset{\sim}{Y}) : \bar{X} - \bar{Y} > c\}$$

Now, $P_{H_0}[\bar{X} - \bar{Y} > c] = \alpha$

or, $$P_{H_0}\left[\frac{\bar{X} - \bar{Y}}{s\left(\frac{1}{n_1} + \frac{1}{n_2}\right)^{1/2}} > \frac{c}{s\left(\frac{1}{n_1} + \frac{1}{n_2}\right)^{1/2}}\right] = \alpha$$

Now, $$t = \frac{\bar{X} - \bar{Y}}{s\left(\frac{1}{n_1} + \frac{1}{n_2}\right)^{1/2}} = \frac{\dfrac{\bar{X} - \bar{Y}}{\sigma\left(\frac{1}{n_1} + \frac{1}{n_2}\right)^{1/2}}}{\sqrt{\dfrac{(n_1+n_2-2)s^2}{\sigma^2} \Big/ (n_1+n_2-2)}}$$

$$\frac{\bar{X} - \bar{Y}}{\sigma\left(\frac{1}{n_1} + \frac{1}{n_2}\right)^{1/2}} \sim N(0,1) \quad \left[s^2 = \frac{(n_1-1)s_1^2 + (n_2-1)s_2^2}{n_1+n_2-2}\right]$$

$$\frac{(n_1+n_2-2)s^2}{\sigma^2} \sim \chi^2_{n_1+n_2-2}$$

So, t may be expressed in the form,

$$\frac{\tau}{\sqrt{\chi^2 / n_1+n_2-2}}, \text{ where } \tau \text{ is independent of } \chi^2$$

So, $t \sim t_{n_1+n_2-2}$ under H_0

$\therefore P_{H_0}[\bar{X} - \bar{Y} > c] = \alpha$

$$\Rightarrow P_{H_0}\left[t > \frac{c}{s\left(\frac{1}{n_1} + \frac{1}{n_2}\right)^{1/2}}\right] = \alpha$$

$$\therefore \frac{c}{s\left(\frac{1}{n_1} + \frac{1}{n_2}\right)^{1/2}} = t_{\alpha;\, n_1+n_2-2}$$

$\longrightarrow$ the upper α point of t-distn. with n_1+n_2-2 degree of freedom.

$$\Rightarrow c = s\left(\frac{1}{n_1} + \frac{1}{n_2}\right)^{1/2} t_{\alpha;\, n_1+n_2-2}$$

Eg. Let X and Y be two independent normal variables $\ni$ $X \sim N(\mu_1, \sigma_1^2)$ and $Y \sim N(\mu_2, \sigma_2^2)$ on the basis of two independent samples

$$\underset{\sim}{X} = (X_1, \ldots, X_{n_1})$$

$$\underset{\sim}{Y} = (Y_1, \ldots, Y_{n_2})$$

drawn from the distn of X and Y respectively.

Test for $H_0: \frac{\sigma_1^2}{\sigma_2^2} = 1$ against different alternatives.

<u>Solution</u>:-

$$s_1^2 = \frac{1}{n_1-1}\sum_{i=1}^{n_1}(X_i - \bar{X})^2, \quad s_2^2 = \frac{1}{n_2-1}\sum_{i=1}^{n_2}(Y_i - \bar{Y})^2$$

Now, testing $H_0: \sigma_1^2/\sigma_2^2 = 1$ Vs

$H_1: \sigma_1^2/\sigma_2^2 > 1$.

We reject Ho if $\frac{s_1^2}{s_2^2} > c$, where c is determined such that,

$$P_{H_0}\left(\frac{s_1^2}{s_2^2} > c\right) = \alpha$$

Now, under Ho,

$$\frac{\dfrac{(n_1-1)s_1^2}{\sigma^2(n_1-1)}}{\dfrac{(n_2-1)s_2^2}{\sigma^2(n_2-1)}} = \frac{\chi_1^2/(n_1-1)}{\chi_2^2/(n_2-1)}, \text{ say}$$

σ_1^2 and σ_2^2 are same under Ho.

$\therefore \chi_1^2 \sim \chi^2_{n_1-1}$ & $\chi_2^2 \sim \chi^2_{n_2-1}$

and χ_1^2 is independent of χ_2^2.

thus under $H_0: \sigma_1^2/\sigma_2^2 = 1$.

$$\therefore \frac{s_1^2}{s_2^2} \sim F_{(n_1-1)(n_2-1)}$$

$\therefore c = F_{\alpha;\, \overline{n_1-1},\, \overline{n_2-1}}$ the upper α point of F distn with (n_1-1) & (n_2-1) d.f.

Paired sample Test :-

Ex:- Let $(X, Y) \sim N_2(\mu_1, \mu_2, \sigma_1^2, \sigma_2^2, \rho)$ on the basis of a paired sample,

$$\{(x_i, y_i) : i = 1(1)n\}.$$

Find a test for $H_0 : \mu_1 = \mu_2$ against different alternatives.

Solution:- Let $z_i = x_i - y_i \quad \forall\, i = 1(1)n$

then $\mu_Z = E(z_i)$

$$= E(x_i - y_i) = \mu_1 - \mu_2$$

$$\sigma_Z^2 = V(z_i) = V(x_i - y_i) = V(x_i) + V(y_i) - 2\,\mathrm{cov}(x_i, y_i)$$

$$= \sigma_1^2 + \sigma_2^2 - 2\rho\sigma_1\sigma_2$$

Also, $(z_1, z_2, \ldots, z_n)$ may be looked upon as independent observations from a normal popln $N(\mu_Z, \sigma_Z^2)$.

Testing of $H_0 : \mu_1 = \mu_2$ is equivalent to test $H_0' : \mu_Z = 0$ when a sample is drawn from a normal population $N(\mu_Z, \sigma_Z^2)$; σ_Z^2 is being unknown. thus as in the one sample situation, we reject $H_0' : \mu_Z = 0$ against $H_1 : \mu_Z > 0$ if $\bar{z} > c$.

where c is $\ni$ $P_{H_0}(\bar{z} > c) = \alpha$

$$\Rightarrow P_{H_0}\left[\frac{\sqrt{n}\,\bar{z}}{s_z} > \frac{\sqrt{n}\,c}{s_z}\right] = \alpha \quad \text{——— (1)}$$

Now, $\dfrac{\sqrt{n}\,\bar{z}}{s_z} \sim t_{n-1}$ under H_0

Hence from (1), $\dfrac{\sqrt{n}\,c}{s_z} = t_{\alpha; n-1} \longrightarrow$ the upper α point of t-distn with $(n-1)$ d.f.

$$\text{So, } c = \frac{s_z}{\sqrt{n}}\, t_{\alpha; n-1}$$

i.e. at the $100\alpha\%$ level of the critical region is given by,

$$W_0 = \left\{ \bar{z} > \frac{s_z}{\sqrt{n}}\, t_{\alpha; n-1} \right\}$$

Ex. $(X, Y) \sim N_2(\mu_1, \mu_2, \sigma_1^2, \sigma_2^2, \rho)$ on the basis of a paired sample $\{(X_i, Y_i) : i = 1(1)n\}$.

Find a test for $H_0 : \rho = 0$ against $H : \rho \neq 0$.

Solution:- From the paired sample, we calculate the sample correlation coefficient

$$r = \frac{\sum_{i=1}^{n}(X_i - \bar{X})(Y_i - \bar{Y})}{\sqrt{\sum_{i=1}^{n}(X_i - \bar{X})^2}\sqrt{\sum_{i=1}^{n}(Y_i - \bar{Y})^2}}$$

If this sample correlation coefficient which is a sample analogue of ρ is either too small or too large, then we can predict that $\rho \neq 0$ so at $100\alpha\%$ level, we reject $H_0 : \rho = 0$ against $H : \rho \neq 0$ if $|r| > c$, here c is determined such that,

$$P_{H_0}(|r| > c) = \alpha$$

let us consider the function

$$\psi(|r|) = \frac{|r|\sqrt{n-2}}{\sqrt{1 - |r|^2}}$$

this is an increasing function of $|r| > c \Leftrightarrow \psi(|r|) > k$.

So, $P_{H_0}(|r| > c) = \alpha$

$$\Leftrightarrow P_{H_0}\left(\frac{|r|\sqrt{n-2}}{\sqrt{1-r^2}} > k\right) = \alpha$$

but we know that under

$H_0 : \rho = 0$, $\dfrac{r\sqrt{n-2}}{\sqrt{1-r^2}} \sim t_{n-2}$.

so, given

$$P_{H_0}\left(\left|\frac{r\sqrt{n-2}}{\sqrt{1-r^2}}\right| > k\right) = \alpha$$

$$\Rightarrow P_{H_0}(|t_{n-2}| > k) = \alpha$$

$$\Rightarrow P_{H_0}(t_{n-2} < -k \text{ or, } t_{n-2} > k) = \alpha$$

i.e. $k = t_{\alpha/2; n-2} \longrightarrow$ the upper $\alpha/2$ point of t-distribution with $(n-2)$ d.f.

$\therefore$ The critical region of the testing problem is:

$$W = \left\{ r : \left|\frac{r\sqrt{n-2}}{\sqrt{1-r^2}}\right| > t_{\alpha/2; n-2} \right\}$$

Remark:- The above testing is also valid if the distn. of Y given $X = x$ is normal with mean as a linear function at x, i.e. $Y | X = x \sim N(\alpha + \beta x, \sigma^2)$; (say).

Ex. $(X, Y) \sim BN(\mu_1, \mu_2, \sigma_1^2, \sigma_2^2, \rho)$

Test $H_0: \sigma_1^2 = \sigma_2^2$ against different alternatives.

Solution:- To test H_0 let us define,

$$U = X - Y \text{ and } V = X + Y$$

then $Cov(U, V) = \sigma_1^2 - \sigma_2^2$

Hence, under H_0, $Cov(U, V) = 0$

Hence, testing of $H_0: \sigma_1^2 = \sigma_2^2$ is equivalent to testing.

$H_0': \rho_{UV} = 0$

So, in the case, we compute the sample correlation

$$r_{UV} = \frac{\sum_{i=1}^{n}(U_i - \bar{U})(V_i - \bar{V})}{\sqrt{\sum_{i=1}^{n}(U_i - \bar{U})^2}\sqrt{\sum_{i=1}^{n}(V_i - \bar{V})^2}}$$

and at $100\alpha\%$ level of significance, we reject H_0 or equivalently H_0' against $H: \sigma_1^2 \neq \sigma_2^2$ (or, equivalently $H': \rho_{UV} \neq 0$)

if $$\left|\frac{r_{UV}\sqrt{n-2}}{\sqrt{1 - r_{UV}^2}}\right| > t_{\alpha/2; n-2}$$

Test for Regression Coefficient:—

$x_1, x_2, \dots, x_n$ be a set of n fixed values assumed by a non-stochastic variable X and let $Y_1, Y_2, \dots, Y_n$ be independently distributed normal variables such that

$$Y_i/X \sim N(\alpha + \beta x_i, \sigma^2), \text{ where } \sigma \text{ is not known.}$$

Test for $H_0: \beta = \beta_0$ against different alternatives.

$\Rightarrow$ Consider the statistic,

$$b = \frac{\sum(x_i - \bar{x})(Y_i - \bar{Y})}{\sqrt{\sum(x_i - \bar{x})^2}}$$

then $b \sim N(\beta, \sigma^2/s_{xx})$

To test for $H_0: \beta = \beta_0$ against $\beta > \beta_0$

We reject H_0 if

$b > c$, where c is such that

$$P_{H_0}(b > c) = \alpha$$

or, $$P_{H_0}\left\{\frac{(b-\beta_0)\sqrt{S_{xx}}}{s_{y.x}} > \frac{(c-\beta_0)\sqrt{S_{xx}}}{s_{y.x}}\right\} = \alpha \quad \cdots\cdots (*)$$

where, $$s^2_{y.x} = \sum_{i=1}^{n} \frac{1}{n-2}\left[y_i - \bar{y} - b(x_i - \bar{x})\right]^2$$

Under $H_0: \beta = \beta_0$

$$\frac{(b-\beta)\sqrt{S_{xx}}}{\sigma} \sim N(0,1).$$

and it is known that, $\dfrac{(n-2)s^2_{y.x}}{\sigma^2} \sim \chi^2_{n-2}$

also, $s^2_{y.x}$ is independent of b.

So, $$\frac{(b-\beta_0)\sqrt{S_{xx}}}{s^2_{y.x}} = \frac{\dfrac{(b-\beta_0)\sqrt{S_{xx}}}{\sigma}}{\sqrt{\dfrac{(n-2)s^2_{y.x}/\sigma^2}{(n-2)}}} \sim t_{n-2},$$ under H_0.

Thus for $(*)$,

$$\frac{(c-\beta_0)\sqrt{S_{xx}}}{s_{y.x}} = t_{\alpha;n-2} \longrightarrow$$ upper α point of t-distn. with $(n-2)$ d.f.

$$\therefore c = \beta + \frac{s_{y.x}}{\sqrt{S_{xx}}} \cdot t_{\alpha;n-2}$$

so at $100\alpha\%$ level we reject H_0 if

$$b > \beta + \frac{s_{y.x}}{\sqrt{S_{xx}}} t_{\alpha;n-2}.$$

☒ (Fisher's t - test) Test for difference of Two population means:-

Two sample problem:- Let X and Y be two independent normal variables such that $X_1 \sim N(\mu_1, \sigma^2)$ and $X_2 \sim N(\mu_2, \sigma^2)$ on the basis of two independent random sample of size n_1 & n_2. Test $H_0: \delta = \mu_1 - \mu_2 = \delta_0$ (say) against different alternatives.

Solution:- Here, $\bar{X}_1 \sim N(\mu_1, \sigma^2/n_1)$

and $\bar{X}_2 \sim N(\mu_2, \sigma^2/n_2)$

$$X_1 - X_2 \sim N\left(\mu_1 - \mu_2, \frac{\sigma^2}{n_1} + \frac{\sigma^2}{n_2}\right)$$

Now, under H_0, define

$$Z = \frac{\bar{x}_1 - \bar{x}_2 - \delta_0}{\sigma\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}} \sim N(0,1) \quad [\because \delta_0 = \mu_1 - \mu_2]$$

Also, it should be noted that

$$\frac{(n_1 - 1)s_1^2}{\sigma^2} \sim \chi^2_{n_1 - 1}, \text{ and,}$$

$$\frac{(n_2 - 1)s_2^2}{\sigma^2} \sim \chi^2_{n_2 - 1}$$

Hence by additive property of χ^2- variates

$$\frac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2}{\sigma^2} \sim \chi^2_{n_1 + n_2 - 2},$$

here $$s^2 = \frac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2}{n_1 + n_2 - 2}$$

Under H_0, $$t = \frac{Z}{\sqrt{\chi^2_{n_1+n_2-2}/n_1+n_2-2}} = \frac{\bar{X}_1 + \bar{X}_2 - \delta_0}{\sigma\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}} \cdot \frac{1}{\sqrt{s^2/\sigma^2}}$$

Thus, $$t = \frac{\bar{X}_1 - \bar{X}_2 - \delta_0}{s\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}} \sim t_{n_1+n_2-2}$$

The test criteria for various alternative hypothesis is as follows.

H_1	Name of Test	Reject H_0 at level α if
$\delta > \delta_0$	Right tail test	$t > t_{n_1+n_2-2}(\alpha)$
$\delta \neq \delta_0$	Two tail test	$\lvert t\rvert > t_{n_1+n_2-2}(\alpha/2)$

If our H_1 is $\delta < \delta_0$, the roles of two populations are interchanged i.e., the suffixes 1 and 2 are interchanged and the right tail test given above is used.

Also in case $\delta_0 = 0$, the above test reduces to testing the equality of two population means.

It is pertinent to note that before applying the above test, we should test the assumption of equality of population variances using 'F' test. If the population variances are not equal, the above test is not valid.

Test of significance of an Observed Sample correlation Ratio η_{YX} :-

Here, Ho is given by Ho : Population correlation Ratio is zero.

The test statistic is

$$F = \frac{\eta^2}{1-\eta^2} \cdot \frac{N-h}{h-1} \sim F(h-1, N-h)$$

where, N is the sample size from a bivariate normal population arranged in h-arrays.

Test of Significance for Linearity of Regression :-

For testing the hypothesis of linearity of regression, our test statistic is

$$F = \frac{\eta^2 - r^2}{1-\eta^2} \cdot \frac{N-h}{h-2} \sim F(h-2, N-h)$$

Note:- For all the above tests the decisions can be made by comparing the tabulated values of F with calculated values with the given degrees of freedom.

Scanned by CamScanner

Test of Significance of Difference of Means :—

Let us assume that $\bar{X}_1$ and $\bar{X}_2$ are respectively the means of two random sample of size n_1 and n_2. Let us also assume that $\mu_i (i=1,2)$ and σ_i^2 $(i=1,2)$ be two means and variances of two populations. Then for large sample size,

$$\bar{X}_1 \sim N(\mu_1, \sigma_1^2/n_1)$$

and

$$\bar{X}_2 \sim N(\mu_2, \sigma_2^2/n_2)$$

Since the difference of two independent normal variables is also a normal variate.

Thus,
$$Z = \frac{(\bar{X}_1 - \bar{X}_2) - E(\bar{X}_1 - \bar{X}_2)}{S.E.(\bar{x}_1 - \bar{x}_2)} \sim N(0,1)$$

Here, $H_0 : \mu_1 = \mu_2$

Thus,
$$E(\bar{X}_1 - \bar{X}_2) = \mu_1 - \mu_2 = 0$$

$$V(\bar{X}_1 - \bar{X}_2) = V(X_1) + V(X_2)$$

$$= \frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}$$ [∵ covariance term is zero due to independence]

Hence, under H_0,

$$Z = \frac{\bar{X}_1 - \bar{X}_2}{\sqrt{\sigma_1^2/n_1 + \sigma_2^2/n_2}} \sim N(0,1)$$

In case, $\sigma_1^2 = \sigma_2^2 = \sigma^2$, then

$$Z = \frac{\bar{X}_1 - \bar{X}_2}{\sigma\left[\frac{1}{n_1} + \frac{1}{n_2}\right)} \sim N(0,1)$$

When the population variances σ_1^2 and σ_2^2 are unknown, we estimate them by their corresponding sample variances, as $\hat{\sigma}_1^2 = s_1^2$ and $\hat{\sigma}_2^2 = s_2^2$

In case $\sigma_1^2 = \sigma_2^2$, then we use the pooled estimate as

$$\hat{\sigma}_1^2 = \hat{\sigma}_2^2 = \hat{\sigma}^2 = \frac{n_1 s_1^2 + n_2 s_2^2}{n_1 + n_2}.$$

—— × ——

Test of Significance related to Two Univariate Normal Populations [Uncorrelated Case]

Suppose that the distribution of the study variable X if each of the two populations be normal and uncorrelated. Suppose that the distn. of two population be $N(\mu_1, \sigma_1^2)$ and $N(\mu_2, \sigma_2^2)$, respectively. Suppose, further that, $X_{11}, X_{12}, \ldots, X_{1n_1}$ be an r.s. from $N(\mu_1, \sigma_1^2)$ and that $X_{21}, X_{22}, \ldots, X_{2n_2}$ be an another r.s. from $N(\mu_2, \sigma_2^2)$. The first set of observation is also supposed to be drawn independently of the second set. Then $\bar{X}_1 = \frac{1}{n_1}\sum_{j=1}^{n_1} X_{1j}$ and $(n_1-1)s_1^2 = \sum_{j=1}^{n_1}(X_{1j}-\bar{X}_1)^2$;

$\bar{X}_2 = \frac{1}{n_2}\sum_{j=1}^{n_2} X_{2j}$ and $(n_2-1)s_2^2 = \sum_{j=1}^{n_2}(X_{2j}-\bar{X}_2)^2$

(a) $\boxed{\text{To test } H_0: \mu_1-\mu_2 = \xi_0 \text{ (known) against } H_1: \mu_1-\mu_2 \neq \xi_0}$

Case-I [σ_1, σ_2 known]

Under $H_0: \mu_1-\mu_2 = \xi_0$, we can expect that the observed value $(\bar{x}_1-\bar{x}_2-\xi_0)$ is small. If the observed value $(\bar{x}_1-\bar{x}_2-\xi_0)$ is quite large in magnitude, we shall suspect H_0 and give support to H_1.

Here, $\left.\begin{array}{l}\bar{X}_1 \sim N\left(\mu_1, \frac{\sigma_1^2}{n_1}\right) \\ \frac{(n_1-1)s_1^2}{\sigma_1^2} \sim \chi^2_{n_1-1}\end{array}\right\}$ independently.

and $\left.\begin{array}{l}\bar{X}_2 \sim N\left(\mu_2, \frac{\sigma_2^2}{n_2}\right) \\ \frac{(n_2-1)s_2^2}{\sigma_2^2} \sim \chi^2_{n_2-1}\end{array}\right\}$ independently.

Both are independent.

Then $\bar{X}_1 - \bar{X}_2 \sim N\left(\mu_1-\mu_2, \frac{\sigma_1^2}{n_1}+\frac{\sigma_2^2}{n_2}\right)$

$\Rightarrow \bar{X}_1 - \bar{X}_2 \sim N\left(\xi_0, \frac{\sigma_1^2}{n_1}+\frac{\sigma_2^2}{n_2}\right)$, under H_0.

$\Rightarrow (\bar{X}_1 - \bar{X}_2 - \xi_0) \sim N\left(0, \frac{\sigma_1^2}{n_1}+\frac{\sigma_2^2}{n_2}\right)$, under H_0.

The distribution of $(\bar{X}_1 - \bar{X}_2 - \xi_0)$ is symmetric about zero with

$$S.E. = \sqrt{\frac{\sigma_1^2}{n_1}+\frac{\sigma_2^2}{n_2}}.$$

Comparing the deviation $(\bar{x}_1-\bar{x}_2-\xi_0)$ w.r.t. its $S.E. = \sqrt{\frac{\sigma_1^2}{n_1}+\frac{\sigma_2^2}{n_2}}$; if $\left|\frac{\bar{x}_1-\bar{x}_2-\xi_0}{\sqrt{\frac{\sigma_1^2}{n_1}+\frac{\sigma_2^2}{n_2}}}\right| > c$, we reject H_0 in favour of H_1, where c is so chosen that $P_{H_0}\left[\left|\frac{\bar{X}_1-\bar{X}_2-\xi_0}{\sqrt{\frac{\sigma_1^2}{n_1}+\frac{\sigma_2^2}{n_2}}}\right| > c\right] = \alpha$.

$\Rightarrow c = \tau_{\alpha/2}$, as $\frac{\bar{X}_1-\bar{X}_2-\xi_0}{\sqrt{\frac{\sigma_1^2}{n_1}+\frac{\sigma_2^2}{n_2}}} \sim N(0,1)$, under H_0.

Hence; we reject $H_0: \mu_1-\mu_2=\xi_0$ against $H_1: \mu_1-\mu_2 \neq \xi_0$ at level α iff the observed value $\left|\dfrac{\bar{x}_1-\bar{x}_2-\xi_0}{\sqrt{\frac{\sigma_1^2}{n_1}+\frac{\sigma_2^2}{n_2}}}\right| > \tau_{\alpha/2}$.

Remark:- p-value of the above testing problem:—

Here; $T = \dfrac{\bar{X}_1-\bar{X}_2-\xi_0}{\sqrt{\frac{\sigma_1^2}{n_1}+\frac{\sigma_2^2}{n_2}}}$ is the test statistic and let 't_0' be the observed value of T.

For the alternative $H_1: \mu_1-\mu_2 \neq \xi_0$, the p-value is

$$p = P_{H_0}\left[\,|T| \geqslant |t_0|\,\right]$$
$$= 2P_{H_0}\left[T \geqslant |t_0|\right]$$
$$= 2\left\{1-\Phi(|t_0|)\right\}, \text{ as } T \sim N(0,1), \text{ under } H_0.$$

Case-II $[\sigma_1, \sigma_2$ unknown but equal$]$

[Fisher's t-test]

For the sake of simplicity we assume that unknown S.D.'s are equal. Assume that $\sigma_1=\sigma_2=\sigma$ (unknown).
Under $H_0: \mu_1-\mu_2=\xi_0$, we can expect that $(\bar{x}_1-\bar{x}_2-\xi_0)$ is small.
If the observed value of $(\bar{x}_1-\bar{x}_2-\xi_0)$ is quite large in magnitude, we shall suspect H_0 and give support to H_1.
Here, $\bar{X}_1-\bar{X}_2-\xi_0 \sim N\left(0, \sigma^2\left(\frac{1}{n_1}+\frac{1}{n_2}\right)\right)$, under the assumption $\sigma_1=\sigma_2=\sigma$.
$\Rightarrow (\bar{X}_1-\bar{X}_2-\xi_0)$ is symmetrically distributed about zero with $S.E. = \sigma\sqrt{\frac{1}{n_1}+\frac{1}{n_2}}$ which is known.

Define, $s^2 = \dfrac{(n_1-1)s_1^2+(n_2-1)s_2^2}{n_1+n_2-2}$ as the pooled sample variance.
Here, s^2 is an U.E. of σ^2.
Clearly, s is an estimate of σ.
$\hat{S.E.}(\bar{X}_1-\bar{X}_2-\xi_0) = s\sqrt{\frac{1}{n_1}+\frac{1}{n_2}}$; comparing the deviation $(\bar{x}_1-\bar{x}_2-\xi_0)$ w.r.t. an estimate of its $S.E. = s\sqrt{\frac{1}{n_1}+\frac{1}{n_2}}$; if

$$\left|\frac{\bar{x}_1-\bar{x}_2-\xi_0}{s\sqrt{\frac{1}{n_1}+\frac{1}{n_2}}}\right| > c, \text{ we reject } H_0 \text{ in favour of } H_1, \text{ where}$$

c is so chosen that

$$P_{H_0}\left[\,\left|\frac{\bar{X}_1-\bar{X}_2-\xi_0}{s\sqrt{\frac{1}{n_1}+\frac{1}{n_2}}}\right| > c\right] = \alpha \quad \cdots\cdots\cdots (*)$$

Distribution of Test Statistic :-

Here $T = \dfrac{\bar{X}_1 - \bar{X}_2 - \xi_0}{s\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}}$ is the test statistic.

Note that, $\bar{X}_1 \sim N\left(\mu_1, \frac{\sigma^2}{n_1}\right)$
and $\dfrac{(n_1-1)s_1^2}{\sigma^2} \sim \chi^2_{n_1-1}$ independently

$\bar{X}_2 \sim N\left(\mu_2, \frac{\sigma^2}{n_2}\right)$
and $\dfrac{(n_2-1)s_2^2}{\sigma^2} \sim \chi^2_{n_2-1}$ independently

Since they are independent.

Now, $\bar{X}_1 - \bar{X}_2 \sim N\left(\mu_1 - \mu_2, \sigma^2\left(\frac{1}{n_1} + \frac{1}{n_2}\right)\right)$
and $\dfrac{(n_1-1)s_1^2 + (n_2-1)s_2^2}{\sigma^2} \sim \chi^2_{n_1+n_2-2}$. independently.

By definition of t-distribution :

$$\frac{\bar{X}_1 - \bar{X}_2 - (\mu_1 - \mu_2)}{\sigma\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}} \Bigg/ \sqrt{\frac{(n_1-1)s_1^2 + (n_2-1)s_2^2}{\sigma^2} \Big/ (n_1+n_2-2)}$$

$$\sim t_{n_1+n_2-2}.$$

$\Rightarrow$ Under H_0, $T = \dfrac{\bar{X}_1 - \bar{X}_2 - \xi_0}{s\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}} \sim t_{n_1+n_2-2}$.

From (*), $c = t_{\alpha/2, n_1+n_2-2}$

Hence, we reject $H_0: \mu_1 - \mu_2 = \xi_0$ against $H_1: \mu_1 - \mu_2 \neq \xi_0$ at level α if the observed value

$$\left| \frac{\bar{x}_1 - \bar{x}_2 - \xi_0}{s\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}} \right| > t_{\alpha/2, n_1+n_2-2}.$$

Remark:- Consider the testing problem of $H_0: \frac{\mu_1}{\mu_2} = \xi_0$ against $H_1: \frac{\mu_1}{\mu_2} \neq \xi_0$.

- **Define**, $T = \bar{x}_1 - \xi_0 \bar{x}_2 \sim N\left(\mu_1 - \xi_0 \mu_2, \frac{\sigma_1^2}{n_1} + \xi_0^2 \cdot \frac{\sigma_2^2}{n_2}\right)$

$\Rightarrow T \sim N\left(0, \frac{\sigma_1^2}{n_1} + \xi_0^2 \cdot \frac{\sigma_2^2}{n_2}\right)$, under H_0.

Case:-I: $\rightarrow$ [σ_1, σ_2 known]

Test Statistic:-

$$\frac{\bar{x}_1 - \bar{x}_2 \xi_0}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}\xi_0^2}} \sim N(0,1), \text{ under } H_0.$$

Case:-II: $\rightarrow$ [$\sigma_1 = \sigma_2 = \sigma$ (unknown)]

Test Statistic:-

$$\frac{\bar{x}_1 - \bar{x}_2 \xi_0}{S\sqrt{\frac{1}{n_1} + \xi_0^2 \frac{1}{n_2}}} \sim t_{n_1+n_2-2}, \text{ under } H_0.$$

(b) $\boxed{\text{To test } H_0: \sigma_1/\sigma_2 = \xi_0 \text{ against } H_1: \sigma_1/\sigma_2 \neq \xi_0}$:—

Case-I:- [μ_1, μ_2 known]

Here, $s_{i0}^2 = \frac{1}{n_i}\sum_{j=1}^{n_i}(x_{ij} - \mu_i)^2$ is a measure of population variance σ_i^2, $i = 1,2$.

Under $H_0: \frac{\sigma_1}{\sigma_2} = \xi_0$, we can expect that

$$\frac{s_{10}^2}{s_{20}^2} \simeq \xi_0^2 \Leftrightarrow \frac{s_{10}^2}{s_{20}^2} \cdot \frac{1}{\xi_0^2} \simeq 1$$

If the observed $\frac{s_{10}^2}{s_{20}^2} \cdot \frac{1}{\xi_0^2} > c_1 (> 1)$ or $< c_2 (< 1)$, we shall reject H_0 in favour of H_1, where c_1 and c_2 are so chosen that,

$$P_{H_0}\left[\frac{s_{10}^2}{s_{20}^2} \cdot \frac{1}{\xi_0^2} > c_1 \text{ or } < c_2\right] = \alpha, \text{ the level of significance.}$$

Note that,

$$\left.\begin{aligned} \frac{n_1 s_{10}^2}{\sigma_1^2} &\sim \chi^2_{n_1} \\ \frac{n_2 s_{20}^2}{\sigma_2^2} &\sim \chi^2_{n_2} \end{aligned}\right\rangle$$

$$\Rightarrow \frac{\frac{n_1 s_{10}^2}{\sigma_1^2}\Big/ n_1}{\frac{n_2 s_{20}^2}{\sigma_2^2}\Big/ n_2} \sim F_{n_1, n_2}$$

$$\Rightarrow \frac{s_{10}^2}{s_{20}^2} \cdot \frac{1}{\xi_0^2} \sim F_{n_1, n_2}, \text{ under } H_0: \frac{\sigma_1}{\sigma_2} = \xi_0.$$

Now, we assign equal error probability to both the tails,

$$P_{H_0}\left[\frac{s_{10}^2}{s_{20}^2}\cdot\frac{1}{\xi_0^2} > c_1\right] = \frac{\alpha}{2} = P_{H_0}\left[\frac{s_{10}^2}{s_{20}^2}\cdot\frac{1}{\xi_0^2} < c_2\right]$$

$$\Rightarrow c_1 = F_{\alpha/2;n_1,n_2} \text{ and } c_2 = F_{1-\alpha/2;n_1,n_2}$$
$$= F_{\alpha/2;n_1,n_2}.$$

The level α test of $H_0: \frac{\sigma_1}{\sigma_2} = \xi_0$ against $H_1: \frac{\sigma_1}{\sigma_2} \neq \xi_0$ is given by:

Reject H_0 iff $\frac{s_{10}^2}{s_{20}^2}\cdot\frac{1}{\xi_0^2} \in \left[F_{\alpha/2;n_1,n_2}, F_{\alpha/2;n_1,n_2}\right]$

Case-II :→ $[\mu_1, \mu_2$ unknown$]$

Here, $s_i^2 = \frac{1}{n_i-1}\sum_{j=1}^{n_i}(x_{ij}-\bar{x}_i)^2$ is a measure of the population variance σ_i^2, $i=1,2$.

Under $H_0: \frac{\sigma_1}{\sigma_2} = \xi_0$, we can expect that $\frac{s_1^2}{s_2^2}\cdot\frac{1}{\xi_0^2} \simeq 1$.

If $\frac{s_1^2}{s_2^2}\cdot\frac{1}{\xi_0^2} > c_1\,(>1)$ or $< c_2\,(<1)$, we reject H_0 in favour of H_1, where c_1 and c_2 are so chosen that,

$$P_{H_0}\left[\frac{s_1^2}{s_2^2}\cdot\frac{1}{\xi_0^2} > c_1 \text{ or } < c_2\right] = \alpha.$$

Note that, $\frac{(n_1-1)s_1^2}{\sigma_1^2} \sim \chi^2_{n_1-1}$, $\frac{(n_2-1)s_2^2}{\sigma_2^2} \sim \chi^2_{n_2-1}$ independently.

$$\Rightarrow \frac{s_1^2}{s_2^2}\cdot\frac{1}{\xi_0^2} \sim F_{n_1-1,n_2-1}; \text{ under } H_0.$$

Assuming;

$$P_{H_0}\left[\frac{s_1^2}{s_2^2}\cdot\frac{1}{\xi_0^2} > c_1\right] = \frac{\alpha}{2} = P_{H_0}\left[\frac{s_1^2}{s_2^2}\cdot\frac{1}{\xi_0^2} < c_2\right]$$

$$\Rightarrow c_1 = F_{\alpha/2;n_1-1,n_2-1}$$
$$c_2 = F_{1-\alpha/2;n_1-1,n_2-1} = F_{\alpha/2;n_1-1,n_2-1}$$

The critical region at level α

$$\frac{s_1^2}{s_2^2}\cdot\frac{1}{\xi_0^2} \in \left[F_{\alpha/2;n_2-1,n_1-1}, F_{\alpha/2;n_1-1,n_2-1}\right]$$

Ex. 1. Let $X_{i1}, X_{i2}, \ldots, X_{in}$ be a r.s. from a popln. following $N(\mu_i, \sigma^2)$ $\forall$ $i = 1(1)3$. Find the test procedure from testing $H_0: \mu_1 - 2\mu_2 + \mu_3 = 0$.

Solution:- To test $H_0: \mu_1 - 2\mu_2 + \mu_3 = 0$ against
$H_1: \mu_1 - 2\mu_2 + \mu_3 \neq 0$

$\boxed{\sigma\text{ - Known :}}$ – Under $H_0: \mu_1 - 2\mu_2 + \mu_3 = 0$; we can expect that $\bar{x}_1 - 2\bar{x}_2 + \bar{x}_3 \simeq 0$. If the magnitude $\bar{x}_1 - 2\bar{x}_2 + \bar{x}_3$ is quite large then we reject H_0 in favour of H_1.

Now, $\bar{X}_i \sim N\left(\mu_i, \frac{\sigma^2}{n}\right)$ independently $\forall$ $i = 1(1)3$.

$$\bar{X}_1 - 2\bar{X}_2 + \bar{X}_3 \sim N\left(\mu_1 - 2\mu_2 + \mu_3, \frac{6\sigma^2}{n}\right)$$

$$\bar{X}_1 - 2\bar{X}_2 + \bar{X}_3 \sim N\left(0, \frac{6\sigma^2}{n}\right), \text{ under } H_0.$$

To compute the deviation w.r.t. the S.E. $= \sqrt{\frac{6\sigma^2}{n}}$; we reject H_0 iff

$$\left|\frac{\bar{x}_1 - 2\bar{x}_2 + \bar{x}_3}{\frac{\sqrt{6}\sigma}{\sqrt{n}}}\right| > c \text{ ; where, } c \text{ is so chosen that}$$

$$P_{H_0}\left[\frac{|\bar{X}_1 - 2\bar{X}_2 + \bar{X}_3|}{\sqrt{6}\sigma/\sqrt{n}} > c\right] = \alpha$$

$$\therefore c = \tau_{\alpha/2}.$$

$\left[\text{Here, } T = \frac{\bar{X}_1 - 2\bar{X}_2 + \bar{X}_3}{\sqrt{6}\sigma/\sqrt{n}}, \text{ is the test statistic} \sim N(0,1)\right]$

We reject H_0 against H_1, at the 'α' level of significance iff

$$\frac{|\bar{x}_1 - 2\bar{x}_2 + \bar{x}_3|}{\frac{\sqrt{6}\sigma}{\sqrt{n}}} \sim \tau_{\alpha/2}.$$

$\boxed{\sigma\text{ unknown : –}}$ Under H_0, we can expect that the observed value $\bar{x}_1 - 2\bar{x}_2 + \bar{x}_3 \simeq 0$. If the observed value $\bar{x}_1 - 2\bar{x}_2 + \bar{x}_3$ is quite large in magnitude then we reject H_0.

Now, $\bar{X}_i \sim N\left(\mu_i, \frac{\sigma^2}{n}\right)$

Under H_0: $\bar{X}_1 - 2\bar{X}_2 + \bar{X}_3 \sim N\left(0, 6\sigma^2/n\right)$

and $S.E.(\bar{X}_1 - 2\bar{X}_2 + \bar{X}_3) = \sqrt{\frac{6}{n}}.\sigma$

Define, $s^2 = \frac{(n_1-1)s_1^2 + (n_1-1)s_2^2 + (n_1-1)s_3^2}{(3n-3)}$, as a pooled sample variance of the 3 samples.

Clearly, s is an estimate of σ and

$$\hat{S.E}(\bar{X}_1 - 2\bar{X}_2 + \bar{X}_3) = \sqrt{\frac{6}{n}}.s$$

Now, comparing the deviation $(\bar{x}_1 - 2\bar{x}_2 + \bar{x}_3)$ w.r.t. an estimate of its S.E., i.e., $\hat{S.E.}(\bar{X}_1 - 2\bar{X}_2 + \bar{X}_3) = \sqrt{\frac{6}{n}} \cdot s$

If the observed value $\left| \frac{\bar{x}_1 - 2\bar{x}_2 + \bar{x}_3}{\sqrt{\frac{6}{n}} \cdot s} \right| > c$, then we reject H_0 against H_1, where, $c = t_{\alpha/2; 3n-3}$. (Do yourself).

Ex.(2): Let $X_1, \dots, X_n$ be a r.s. from $N(\mu_1, \sigma_1^2)$ and $Y_1, \dots, Y_m$ be a r.s. from $N(\mu_2, \sigma_2^2)$, where μ_1 is known and the others are unknown. Find the test procedure of $H_0: \sigma_1 = \sigma_2$.

Solution:- To test the hypothesis $H_0: \sigma_1 = \sigma_2 = \sigma$ (say)

$\boxed{\mu_1 \text{ is known}}$:- Here $s_{10}^2 = \frac{1}{n}\sum_{i=1}^{n}(X_i - \mu_1)^2$ is a measure of σ_1^2 and

$$s_{20}^2 = \frac{1}{m-1}\sum_{i=1}^{m}(Y_i - \bar{y})^2,$$ where $\bar{y}$ is the measure of μ_2

Under $H_0: \sigma_1/\sigma_2 = 1$, we can expect that $\frac{s_{10}^2}{s_{20}^2} \simeq 1$. If the observed value $\frac{s_{10}^2}{s_{20}^2} > c_1 (>1)$ or $< c_2 (<1)$, we reject H_0 in favour of H_1, where c_1 and c_2 are so chosen such that

$$P_{H_0}\left[\frac{s_{10}^2}{s_{20}^2} > c_1 \text{ or } < c_2\right] = \alpha,$$ the level of significance.

Note that, $\frac{n s_{10}^2}{\sigma_1^2} \sim \chi_n^2$
$\frac{(m-1)s_{20}^2}{\sigma_2^2} \sim \chi^2_{m-1}$ independently.

$$\therefore \frac{\frac{n s_{10}^2}{\sigma_1^2}\Big/n}{\frac{(m-1)s_{20}^2}{\sigma_2^2}\Big/(m-1)} \sim F_{n;m-1}$$

$$\Rightarrow \frac{s_{10}^2}{s_{20}^2} \sim F_{n;m-1},$$ under $H_0: \sigma_1 = \sigma_2 = \sigma$ (say).

Now, we assigns equal error probability to both the tail,

$$\therefore P_{H_0}\left[\frac{s_{10}^2}{s_{20}^2} > c_1\right] = \frac{\alpha}{2} = P_{H_0}\left[\frac{s_{10}^2}{s_{20}^2} < c_2\right]$$

$$\Rightarrow c_1 = F_{\alpha/2; n, m-1} \text{ and } c_2 = F_{1-\alpha/2; n, m-1} = F_{\alpha/2; m-1, n}.$$

$\therefore$ The level α test of $H_0: \sigma_1 = \sigma_2$ is given by:

Reject H_0 iff $\frac{s_{10}^2}{s_{20}^2} \in \left[F_{\alpha/2; n, m-1}, F_{\alpha/2; m-1, n}\right]$

Testing Relating to a Bivariate normal Distribution

Suppose in a given population, the variables X and Y are distributed according to a $BN(\mu_X, \mu_Y, \sigma_x^2, \sigma_y^2, \rho)$ law. Let $(x_1, y_1), (x_2, y_2), \ldots, (x_n, y_n)$ be a given r.s. from this population. Here, $X \sim N(\mu_x, \sigma_x^2)$ and $Y \sim N(\mu_y, \sigma_y^2)$ with correlation coefficient ρ. (Correlated case).

(a) $\boxed{\text{To test } H_0: \mu_x - \mu_y = \xi_0 \text{ (correlated case)}}$:-

(Paired t-test)

When two variables X and Y are measured in the same unit, then we may be interested in testing $H_0: \mu_x - \mu_y = \xi_0$.

Define, $D = X - Y$

Clearly, $D \sim N(\mu_D, \sigma_D^2)$; where $\mu_D = \mu_x - \mu_y$ and $\sigma_D^2 = \sigma_x^2 + \sigma_y^2 - 2\rho\sigma_x\sigma_y$

Then $d_i = x_i - y_i \; \forall i = 1(1)n$ can be considered as an observed sample from the univariate normal, i.e. $N(\mu_D, \sigma_D^2)$ population.

Therefore, the H_0 reduces to $H_0: \mu_D = \xi_0$

Under $H_0: \mu_D = \xi_0$; we can expect that $(\bar{d} - \xi_0)$ is small.

If $(\bar{d} - \xi_0)$ is quite large in magnitude. We suspect H_0 and give support to $H_1: \mu_D \neq \xi_0$.

Note that, $\bar{D} \sim N\left(\xi_0, \frac{\sigma_D^2}{n}\right)$; under $H_0: \mu_D = \xi_0$.

$$\Rightarrow (\bar{D} - \xi_0) \sim N\left(0, \frac{\sigma_D^2}{n}\right) \text{ which is symmetric about '0'.}$$

Under H_0, with $S.E. = \frac{\sigma_D}{\sqrt{n}}$ (unknown)

Now, $\hat{S.E.} \doteq \frac{\hat{\sigma}_D}{\sqrt{n}} = \frac{s_D}{\sqrt{n}}$, where $s_D = \frac{1}{n-1}\sum_{i=1}^{n}(d_i - \bar{d})^2$

Comparing the deviation $(\bar{d} - \xi_0)$ w.r.t. an estimate of the S.E., i.e. w.r.t. $\frac{s_D}{\sqrt{n}}$, if the observed $\left|\frac{(\bar{d} - \xi_0)}{s_D/\sqrt{n}}\right| > c$; we reject H_0 in favour of $H_1: \mu_D \neq \xi_0$, at level α, where c is so chosen that

$$P_{H_0}\left[\frac{\sqrt{n}(\bar{D} - \xi_0)}{s_D} > c\right] = \alpha$$

$$\Rightarrow c = t_{\alpha/2; n-1}.$$

[Here, $\frac{\sqrt{n}(\bar{D}-\xi_0)}{\sigma_D} \sim N(0,1)$

$\frac{(n-1)s_D^2}{\sigma_D^2} \sim \chi^2_{n-1}$

independently, under Ho.

Here,
$$\therefore \frac{\frac{\sqrt{n}(\bar{D}-\xi_0)}{\sigma_D}}{\sqrt{\frac{(n-1).s_D^2}{\sigma_D^2}\Big/(n-1)}} \sim t_{n-1}, \text{ under Ho.}$$

$$\Rightarrow \frac{\sqrt{n}(\bar{D}-\xi_0)}{s_D^2} \sim t_{n-1}; \text{ under Ho.} \;]$$

Hence, we reject $H_0: \mu_x - \mu_y = \xi_0$ against $H_1: \mu_x-\mu_y \neq \xi_0$ at level α, if the observed value $\left|\frac{\sqrt{n}(\bar{d}-\xi_0)}{s_D}\right| > t_{\alpha/2;n-1}$.

Remark:— Sometimes, we may be interested in testing $H_0: \eta = \frac{\mu_x}{\mu_y} = \eta_0$.

$\Leftrightarrow H_0: \mu_x - \eta_0 \mu_y = 0$

Define, $Z = X - \eta_0 . Y$

Then, $Z \sim N(\mu_Z, \sigma_Z^2)$

where, $\mu_Z = \mu_x - \eta_0 . \mu_y$

$\sigma_Z^2 = \sigma_x^2 + \eta_0^2 \sigma_y^2 - 2\eta_0 . \rho \sigma_x \sigma_y$

Then $Z_i = X_i - \eta_0 . Y_i \;\; \forall i = 1(1)n$ is a r.s. from $N(\mu_Z, \sigma_Z^2)$.

Here, $\bar{Z} \sim N\left(\mu_Z, \frac{\sigma_Z^2}{n}\right)$

$\frac{(n-1)s_Z^2}{\sigma_Z^2} \sim \chi^2_{n-1}$

independently.

$$\Rightarrow \frac{\sqrt{n}(\bar{Z}-0)}{s_Z} \sim t_{n-1}; \text{ under } H_0: \mu_Z = 0.$$

The test statistic is, $T = \frac{\sqrt{n}\bar{Z}}{s_Z} \sim t_{n-1}$, under Ho.

Hence, we reject $H_0: \frac{\mu_x}{\mu_y} = \eta_0$ against $H_1: \frac{\mu_x}{\mu_y} \neq \eta_0$ at level α, if the observed value $\left|\frac{\sqrt{n}\bar{Z}}{s_Z}\right| > t_{\alpha/2;n-1}$.

(b) $\boxed{\text{To test } H_0: \rho = 0 \text{ (uncorrelated case):}}$

Here, $r = \dfrac{\sum (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum (x_i - \bar{x})^2}\sqrt{\sum (y_i - \bar{y})^2}}$ is an estimate of ρ.

If $|r|$ is quite larger than zero, then we suspect $H_0: \rho = 0$ and give support to $H_1: \rho \neq 0$.

Under, $H_0: \rho = 0$

$$\frac{r\sqrt{n-2}}{\sqrt{1-r^2}} \sim t_{n-2}$$

If $|r| > c$, we suspect H_0 in favour of $H_1: \rho \neq 0$, where c is such that $P_{H_0}[|r| > c] = \alpha$

Now, $|r| > c$

$$\Leftrightarrow \frac{|r|}{\sqrt{1-r^2}} > \frac{c}{\sqrt{1-c^2}}$$

$$\Leftrightarrow \frac{|r|\sqrt{n-2}}{\sqrt{1-r^2}} > \frac{c\sqrt{n-2}}{\sqrt{1-c^2}} = c'$$

Hence if $\left|\frac{r\sqrt{n-2}}{\sqrt{1-r^2}}\right| > c'$; we reject H_0 in favour of $H_1: \rho \neq 0$, where c' is such that $P\left[\left|\frac{r\sqrt{n-2}}{\sqrt{1-r^2}}\right| > c'\right] = \alpha$

$$\Rightarrow c' = t_{\alpha/2;n-2}.$$

The critical region of the testing problem is;

$$W = \left\{ r : \left|\frac{r\sqrt{n-2}}{\sqrt{1-r^2}}\right| > t_{\alpha/2;n-2} \right\}$$

<u>Remark:-</u> The above testing is also valid if the distn of Y given $X = x$ is normal with mean as a linear function of x, i.e. $Y|X = x \sim N(\alpha + \beta x, \sigma^2)$; (say).

(c) $\boxed{\text{To test } H_0: \xi = \frac{\sigma_x}{\sigma_y} = \xi_0 \text{ (correlated case)}}$:—

When two ~~univariate~~ variables are measured in the same units, then we may be interested in testing $H_0: \xi = \frac{\sigma_x}{\sigma_y} = \xi_0$

Define, $U = X + \xi_0 . Y$

$V = X - \xi_0 . Y$

Then (U, V) follows BN with $\operatorname{cov}(U,V) = \sigma_x^2 - \xi_0^2 \sigma_y^2$.

Hence, $H_0: \frac{\sigma_x}{\sigma_y} = \xi_0 \Leftrightarrow H_0: \rho_{U,V} = 0$.

Hence, based on the bivariate data (U_i, V_i), $i = 1(1)n$; the test of $H_0: \rho_{U,V} = 0$.

Here, $$r_{U,V} = \frac{\sum (U_i - \bar{U})(V_i - \bar{V})}{\sqrt{\sum (U_i - \bar{U})^2} \sqrt{\sum (V_i - \bar{V})^2}}$$ is an estimate of $\rho_{U,V}$.

Under $H_0: \rho_{U,V} = 0$

$$\frac{r_{U,V}\sqrt{n-2}}{\sqrt{1 - r_{U,V}^2}} \sim t_{n-2}.$$

If $$|r_{U,V}| > c \Leftrightarrow \frac{|r_{U,V}|}{\sqrt{1 - r_{U,V}^2}} > \frac{c}{\sqrt{1-c^2}}.$$

$$\Leftrightarrow \frac{|r_{UV}|\sqrt{n-2}}{\sqrt{1 - r_{U,V}^2}} > \frac{c\sqrt{n-2}}{\sqrt{1-c^2}} = c'$$

Hence, if $\left| \frac{r_{U,V}\sqrt{n-2}}{\sqrt{1 - r_{U,V}^2}} \right| > c'$, we reject H_0 in favour of $H_1: \rho_{UV} \neq 0$,

where, c is such that

$$P\left[\left| \frac{r_{U,V}\sqrt{n-2}}{\sqrt{1 - r_{U,V}^2}} \right| > c' \right] = \alpha \Rightarrow c' = t_{\alpha/2, n-2}.$$

The critical region of the testing problem is:

$$W = \left\{ r_{U,V} : \left| \frac{r_{U,V}\sqrt{n-2}}{\sqrt{1 - r_{U,V}^2}} \right| > t_{\alpha/2; n-2} \right\}$$

Therefore, we reject $H_0: \frac{\sigma_x}{\sigma_y} = \xi_0$ against $H_1: \frac{\sigma_x}{\sigma_y} \neq \xi_0$

$\Leftrightarrow H_0: \rho_{U,V} = 0$ against $H_1: \rho_{U,V} \neq 0$ at level α, if the observed value $\left| \frac{r_{U,V}\sqrt{n-2}}{\sqrt{1 - r_{U,V}^2}} \right| > t_{\alpha/2; n-2}$.

———— o ————

TESTING OF HYPOTHESIS

INTRODUCTION :- A statistical hypothesis will be a hypothesis about the distn. of the popln. As the term suggests, one wishes to decide whether or not some hypothesis that has been formulated is correct. The choice here lies between two decisions: 'accepting or rejecting the hypothesis'.

A decision procedure for such a problem is called a test of the hypothesis. A problem of testing hypothesis is posed as follows: the decision is to be based on the value of a certain r.v. X, the distn. of which is known to being to a class of distn.. Given a r.v. $X_1, \ldots, X_n$ from $f(x,\theta)$, to test whether the data support $\theta \in \Omega$, where $\Omega_0 \cup \Omega_1 = \Omega$.

Definition:- Simple & composite hypothesis

A statistical hypothesis is an assertain or conjecture about the distribution of the population. If the statistical hypothesis specifies the distn. of the population completely, then it is called simple hypothesis. If the statistical hypothesis does not specify the distn. of the popln completely, it is called composite hypothesis.

Examples:- Let $X_1, X_2, \ldots, X_n$ be a r.s. from $N(\theta, 5^2)$. Now the assumption that " the mean of the distn. is ≤ 17", is denoted by $H : \theta \leq 17$ and is a statistical hypothesis.

The hypothesis $H : \theta = 17$ is a simple hypothesis, since it completely specifies the distn.. On the other hand, the hypothesis $H : \theta \leq 17$ is a composite hypothesis, since it does not specify the distn. completely.

Test of statistical Hypothesis:—

Definition:- A test of statistical hypothesis H is a rule or procedure for deciding whether to reject or to accept H on the basis of the given random sample from the population.

Example:- Let $x_1, x_2, \ldots, x_n$ be an observed r.s. from $N(\theta, 5^2)$. Consider a hypothesis $H : \theta = 17$, one possible test is as follows:

$$\text{Reject } H \text{ iff } \bar{x} < 17 - \frac{5}{\sqrt{n}} \quad \text{or} \quad \bar{x} > 17 + \frac{5}{\sqrt{n}}$$

<u>Critical Region and Test</u> :— Let $\mathscr{X}$ denotes the collection of all possible samples of size n i.e. $\mathscr{X} = \{(x_1, \cdots, x_n)$: $(x_1, x_2, \cdots, x_n)$ is a possible value of $(X_1, X_2, \cdots, X_n)\}$. Here $\mathscr{X}$ is called the sample space or potential data set.

A test procedure assigns to each possible value of the data $\underset{\sim}{x}$ of $\underset{\sim}{X}$. One of the two decisions: accept H or reject H, and thereby divides the sample space $\mathscr{X}$ into two complementary regions $\mathscr{X}_0$ and $\mathscr{X}_1$ such that if $\underset{\sim}{x}$ falls in $\mathscr{X}_0$, the hypothesis H is accepted, otherwise it is rejected.

The set $\mathscr{X}_0$ is called the region of acceptance and $\mathscr{X}_1$ is called the region of rejection or critical region of H.

<u>The choice of Null Hypothesis</u> :— In any testing problem, to test whether the data supports the hypothesis $H_1 : \theta \in \Omega_0$ or the hypothesis $H_2 : \theta \in \Omega_1$.

In the formulation of testing problem, the roles of H_1 and H_2 are not symmetric. In order to decide which one of the two hypothesis should be taken as null hypothesis H_0, the difference between the roles and the implications of this two term should be clearly understood. In testing hypothesis, a statistician should be completely impartial and should have no brief for any party or company, nor should he allows his personal views to influence the decission. Let us suppose that, the bulbs manufactured on the standard process have an averages life μ hours and μ_0 is the mean life of bulb manfacturer by the new process, and three hypothesis may be setup in this approach —

i) $\mu_0 > \mu$, ii) $\mu_0 < \mu$, iii) $\mu_0 = \mu$.

The first two statements appeared to be biased since they reflect a preferential attitude to one or the other of the two processes. Hence the best course is to addopt the hypothesis of no difference assteted in (iii). This suggests that the statistician should take up the nutral or null attitude regarding the outcome of the test.

This neutral or non-committal attitude of the statistician before the sample values are taken as key of the choice of the null hypothesis. Keeping in mind the potential losses due to the wrong decision, the decision make of some what conservative in holding the null-hypothesis as true unless their is the strong evidence that is false and to him the consequences of wrongly rejecting a null hypothesis since to be more serious than those of wrongly accepting it. Hence, we denote by H_0 that hypothesis among H_1 and H_2, the false rejection of which is regarded as more serious and call it the null hypothesis. the other hypothesis will be denoted by H_A or H_1 and call it the alternative hypothesis.

Performance of a Test :- While performing a test one may arrive at the correct ~~performin~~ decision or may commit one of the two errors:

(i) Rejecting the null hypothesis H_0 when it is true,

(ii) Accepting H_0 when it is false.

True State	Decission from sample	
	Reject H_0	Accept H_0
H_0 is true	Wrong [Type I error]	Correct
H_0 is false (or, H_A is true)	Correct	Wrong [Type II error]

Definition :- Rejection of null hypothesis H_0 when it is true, is called Type-I error.

and, Acceptance of H_0 when it is false is called Type-II error

The probability of rejecting H_0 when it is true $= P[\underset{\sim}{X} \in W \mid \theta], \theta \in \Omega_0$, where W is the critical region of the test of $H_0 : \theta \in \Omega_0$ against $H_1 : \theta \in \Omega_1$, is called the probability of Type-I error.

The probability of accepting H_0 when it is false $= P[\underset{\sim}{X} \in W^c / \theta], \theta \in \Omega_1$

$$= 1 - P[\underset{\sim}{X} \in W / \theta], \theta \in \Omega_1,$$

is called the Probability of type-II error.

It is desirable to carry out the test in a manner which keeps the probabilities of two types of error to a minimum level. Unfortunately, for a given sample size n, both the error probabilities can't be controlled simultaneously.

[Let C and D be two critical regions such that $C \subset D$.
Then $P_\theta[C] \leq P_\theta[D]$, $\theta \in \Omega_0$
and $1 - P_\theta[C] > 1 - P_\theta[D]$, $\theta \in \Omega_1$.

Thus by shrinking or enlarging a given critical region, we can decrease one type of error probability of the cost of increase is the error probability of other type.]

Level of Significance, Size and Power of a test : —

The usual procedure of finding a test is to restrict the probability of type-I error and then to minimize the probability of type-II error. Note that here we control that error which is more serious or important that is type-I error. Thus, one selects a number $\alpha \in (0,1)$ and impose the condition that $P[\underset{\sim}{X} \in W / \theta] \leq \alpha$, $\forall\, \theta \in \Omega_0$, then the quantity α is called the level of significance of the testing problem.

The choice of level of significance, of course, depends on the experimental himself. If he thinks that, the rejection of null hypothesis, when actually it is true, will be a serious error, he will choose a rather small value of α, say $0{\cdot}01, 0{\cdot}001$. On the other hand, if he thinks that this error is not so serious, he will not mind taking a value as high as $\cdot 05$, $\cdot 10$.

The number $\sup_{\theta \in \Omega_0} P[\underset{\sim}{X} \in W / \theta]$ is called the size of the test is given by the critical region W. The size of a test gives the maximum possible probability of committing the type-I error and it provides the quality of the test.

Now, subject to the condition, " $P[\underset{\sim}{X} \in W / \theta] \leq \alpha$, $\forall\, \theta \in \Omega_0$ "

$$\Leftrightarrow \sup_{\theta \in \Omega_0} P[\underset{\sim}{X} \in W / \theta] \leq \alpha$$

$$\Leftrightarrow \{\text{sizes of the critical region } W\} \leq \alpha,$$

it is desired to minimize $P[\underset{\sim}{X} \in W^c / \theta]$, $\theta \in \Omega_1$.

$\Leftrightarrow$ to maximize $P[\underset{\sim}{X} \in W / \theta]$, $\theta \in \Omega_1$,

to get the best test.

The probability of rejection of the null hypothesis H_0, when it is false, i.e., $P[\underset{\sim}{X} \in W/\theta]$ evaluated for a given $\theta = \theta_1 \in \Omega_1$, is called the Power of the test given by the critical region W of H_0 against H_1 at $\theta = \theta_1$.

Considering as a function of θ_1, $\theta \in \Omega$, the parameter space; $P[\underset{\sim}{X} \in W/\theta]$ is called <u>the power function</u> of the test given by the critical region W and it is denoted by $\beta_W(\theta)$.

Note that, for $\theta \in \Omega_0$,

$$\beta_W(\theta) = P[\underset{\sim}{X} \in W/\theta] = \text{Probability of type-I error.}$$

and for $\theta \in \Omega_1$,

$$\beta_W(\theta) = P[\underset{\sim}{X} \in W/\theta]$$
$$= 1 - P[\underset{\sim}{X} \in W^c/\theta]$$
$$= 1 - [\text{Probability of type II error}]$$

<u>Remark:-</u> <u>Size α and level α tests:</u>

If $P[\underset{\sim}{X} \in W/\theta] \leq \alpha$, $\forall\ \theta \in \Omega_0$

$$\Leftrightarrow \sup_{\theta \in \Omega_0} P[\underset{\sim}{X} \in W/\theta] \leq \alpha$$

$$\Leftrightarrow \{\text{size of the test } W\} \leq \alpha,$$

the test given by the critical region W is called <u>a level α test.</u>

If $\sup_{\theta \in \Omega_0} P[\underset{\sim}{X} \in W/\theta] = \alpha_1$, say,

the test given by the critical region W is a <u>size α_1 test.</u>

Hence all the test whose sizes are less that or equal to α, the level of significance of the testing problem, ~~are~~ known as <u>level α tests.</u>

- <u>Why $\{1 - P(\text{Type II error})\}$ is called the power of a test?</u>

<u>ANS:-</u> $1 - P(\text{Type II error}) = \text{Prob}[\text{Rejecting a false hypothesis}]$ which is disirable and the more the probability the more powerful will be the test for testing H_0 Vs H_1.

That is why $\{1 - P(\text{Type II error})\}$ is said to be the Power of a test.

Ex.1. A sample of size one is taken from Exp (mean $\frac{1}{\theta}$). To test $H_0: \theta = 2$ against $H_1: \theta = 1$. Consider the critical region W: reject H_0 iff $x \geq 1$. Find the probability of type I error & type-II error.

Solution:-

The critical region, $W = \{x : x \geq 1\}$

The power function of W is $\beta(\theta) = P[X \in W / \theta]$

$$= P[X \geq 1 / \theta]$$

$$= \int_1^{\infty} \theta e^{-\theta x} dx$$

$$= \lim_{t \to \infty} \int_1^t \theta e^{-\theta x} dx$$

$$= \lim_{t \to \infty} \left[-e^{-\theta x} \right]_1^t$$

$$= \lim_{t \to \infty} \left[e^{-\theta} - e^{-\theta t} \right]$$

$$= e^{-\theta}.$$

or, $P[X \geq 1/\theta]$

$$= 1 - P[X < 1/\theta]$$

$$= 1 - \int_0^1 \theta e^{-\theta x} dx$$

$$= e^{-\theta}.$$

Probability of type-I error: $= P[X \in W / H_0]$

$$= P[X \in W / \theta = 2]$$

$$= \beta(\theta = 2).$$

$$= e^{-2}$$

Probability of type-II error: $= P[\underset{\sim}{x} \in W^c / H_1]$

$$= 1 - P[\underset{\sim}{x} \in W / \theta = 1]$$

$$= 1 - \beta(1)$$

$$= 1 - e^{-1}.$$

Ex. 2. Let $X_1, \ldots, X_n$ be a r.s. from $N(\theta, 5^2)$. To test $H_0: \theta \leq 17$ against $H_1: \theta > 17$. Find the size and the power function of the test: reject H_0 iff $\bar{X} > 17 + \frac{5}{\sqrt{n}}$.

Solution:- Critical region: $W = \left\{ \underset{\sim}{x} : \bar{x} > 17 + \frac{5}{\sqrt{n}} \right\}$

The power function of the test is

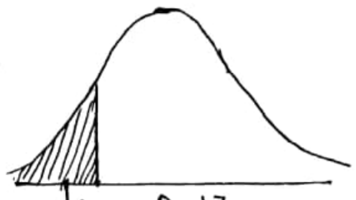
$$\beta(\theta) = P[\underset{\sim}{x} \in W / \theta] = P_\theta\left[\bar{X} > 17 + \frac{5}{\sqrt{n}}\right]$$

$$= P_\theta\left[\frac{\bar{X} - \theta}{5/\sqrt{n}} > \frac{17 - \theta}{5/\sqrt{n}} + 1\right]$$

$$= 1 - \Phi\left(\frac{17 - \theta}{5/\sqrt{n}} + 1\right) \quad \left[\because \text{as } \bar{X} \sim N\left(\theta, \frac{5^2}{n}\right)\right)$$

$$= \Phi\left(-1 + \frac{\theta - 17}{5/\sqrt{n}}\right) \quad [\because \Phi(\alpha) + \Phi(1-\alpha) = 1]$$

$$\text{Size of the test} = \sup_{\theta \in \Omega_0} P[\underset{\sim}{X} \in W | \theta]$$
$$= \sup_{\theta \in \Omega_0} \beta(\theta)$$
$$= \sup_{\theta \le 17} \Phi\left(-1 + \frac{\theta - 17}{5/\sqrt{n}}\right)$$
$$= \Phi\left(-1 + \frac{17-17}{5/\sqrt{n}}\right)$$
$$= \Phi(-1)$$

[

$-1 + \frac{\theta - 17}{5/\sqrt{n}}$

& $\Phi\left(-1 + \frac{\theta - 17}{5/\sqrt{n}}\right)$ is an increasing function of θ.]

Ex.3. Let X_1, X_2 be a random sample from $R(0, \theta)$. To test $H_0: \theta = \theta_0$ against $H_1: \theta \ne \theta_0$. Find the probabilities of type I and type II error of the test: reject H_0 if $\max(x_1, x_2) > \theta_0$ or $< \theta_0\sqrt{\alpha}$

[C.U. 2005]

Solution:— Let $M = \max\{x_1, x_2\}$

Critical region: $W = \{(x_1, x_2) : \max(x_1, x_2) > \theta_0 \text{ or } < \theta_0\sqrt{\alpha}\}$

The power function of the test:

$$\beta(\theta) = P[(X_1, X_2) \in W / \theta]$$
$$= P_\theta[M > \theta_0 \text{ or } < \theta_0\sqrt{\alpha}]$$
$$= 1 - P_\theta[\theta_0\sqrt{\alpha} \le M \le \theta_0]$$
$$= 1 - \{F_M(\theta_0) - F_M(\theta_0\sqrt{\alpha})\}$$
$$= 1 - \left\{\left(\frac{\theta_0}{\theta}\right)^2 - \left(\frac{\theta_0\sqrt{\alpha}}{\theta}\right)^2\right\}, \quad \text{where } F_M(m) = \left(\frac{m}{\theta}\right)^2, \ 0 < m < \theta$$
$$= 1 - \left(\frac{\theta_0}{\theta}\right)^2 (1 - \alpha)$$

$$\text{Prob. of type I error} = P[(X_1, X_2) \in W / \theta = \theta_0]$$
$$= \beta(\theta_0)$$
$$= 1 - (1 - \alpha) = \alpha$$

$$\text{Prob. of type II error} = P[(X_1, X_2) \in W^c / \theta], \quad \theta \ne \theta_0$$
$$= 1 - \beta(\theta), \quad \theta \ne \theta_0$$
$$= \left(\frac{\theta_0}{\theta}\right)^2 (1 - \alpha), \quad \theta \ne \theta_0.$$

Ex. 4. Let X_1, X_2 be a r.s. from an exponential distribution with mean θ. To test $H_0: \theta = 2$ against $H_1: \theta = 4$. Consider the critical region

$$C = \{(x_1, x_2): x_1 + x_2 \geq 9.5\}.$$ Find the size and power of the test provided by the critical region C.

Hints:- $\beta(\theta) = P_\theta[X_1 + X_2 \geq 9.5]$

$$= P_\theta[Y \geq 9.5], \text{ where } Y = X_1 + X_2 \sim \text{Gamma}(\theta, n=2)$$

$$= \int_{9.5}^{\infty} \frac{1}{\Gamma(2)\,\theta^2} e^{-x/\theta} \cdot x^{2-1} dx$$

$$= \int_{9.5/\theta}^{\infty} e^{-z} z\, dz, \quad \text{where } x/\theta = z$$

$$= \left[-ze^{-z} - \int(-e^{-z})dz\right]_{9.5/\theta}^{\infty}$$

$$= \left[-(z+1)e^{-z}\right]_{9.5/\theta}^{\infty}$$

$$= \left(1 + \frac{19}{2\theta}\right)e^{-19/2\theta}$$

Ex. 5. An urn contains 10 marbles of which M are white and $10-M$ are black. To test $H_0: M = 5$ against $H_1: M = 6$, one draws 3 marbles from the urn WOR. The null hypothesis is rejected if the sample contains 2 or 3 white marbles; O.W. it is accepted. Find the size and the power of the test.

Hints:- Let X be the no. of white marbles in a sample of size 3 drawn WOR.

Therefore, $$P_M[X = x] = \begin{cases} \dfrac{\binom{M}{x}\binom{10-M}{3-x}}{\binom{10}{3}}, & x = 0, 1, 2, 3. \\ 0, & \text{o.w.} \end{cases}$$

$$W = \{x : x = 2, 3\}$$

Power function, $$\beta(M) = P_M(X \in W) = P_M(X = 2, 3)$$

$$= \frac{\binom{M}{2}\binom{10-M}{1}}{\binom{10}{3}} + \frac{\binom{M}{3}}{\binom{10}{3}}$$

Ex. 6. A man has 6 dice out of which, an unknown no. m, is known to be bias, so that when tossed these always, show 6, the rest are all unbiased. To test $H_0: m=2$ against alternative $H_1: m=1$. The following rule is suggested: toss all the dice and reject H_0 if the no. of sixes is 3 or less. Find the probability of type-I and type II error. [C.U.]

Solution:— Define, $x_i = \begin{cases} 1 & \text{if the ith die results in 'six'} \\ 0 & \text{OW}, \quad i=1(1)6 \end{cases}$

$$W = \left\{ (x_1, \ldots, x_6) : \sum_{i=1}^{6} x_i \le 3 \right\}$$

Power function: $\beta(m) = P(\underset{\sim}{X} \in W / m)$

$$= P_m \left[\sum_{i=1}^{6} X_i \le 3 \right]$$

$$= P \left[\sum_{i=1}^{6-m} X_i \le 3-m \right]$$

[WLG, let the last m dice are so biased that they always show six, i.e. $x_{6-m+1} = \cdots\cdots = x_6 = 1$.

$$\sum_{i=1}^{6} x_i \le 3 \Leftrightarrow \sum_{i=1}^{6-m} x_i \le 3-m \Big]$$

$$\therefore \beta(m) = \sum \binom{6-m}{x} \left(\frac{1}{6}\right)^x \left(\frac{5}{6}\right)^{6-m-x}$$

$$= \frac{1}{6^{6-m}} \left\{ \sum_{x=0}^{3-m} \binom{6-x}{x} 5^{6-m-x} \right\}$$

Probability of type I error $= \beta(m=2)$

$$= \sum_{x=0}^{1} \binom{4}{x} \frac{5^{4-x}}{6^4}$$

Probability of type II error $= 1 - \beta(1)$

$$= 1 - \sum_{x=0}^{2} \binom{5}{x} \frac{5^{5-x}}{6^5}.$$

Test Problem:— For testing $H_0: \theta \in \Theta_0$ Vs. $H_1: \theta \in \Theta - \Theta_0$ at $100 \times \alpha$ % level of significance the triplet $(\alpha, \Theta_0, \Theta)$ is called the test problem.

Test of Significance :— Suppose that $X \sim N(\mu, \sigma^2)$, where σ is known but μ is unknown.

We wish to test $H_0 : \mu = \mu_0$ against $H_1 : \mu \neq \mu_0$. In order to test H_0, let us assume, to begin with, that H_0 is true. Let $x_1, x_2, \dots, x_n$ be a given r.s. from $N(\mu, \sigma^2)$.

[N.T. $\bar{X} \sim N\left(\mu, \frac{\sigma^2}{n}\right) \Rightarrow \bar{X} \sim N\left(\mu_0, \frac{\sigma^2}{n}\right)$, under $H_0 : \mu = \mu_0$

$$\Rightarrow \frac{\bar{X} - \mu_0}{\sigma/\sqrt{n}} \sim N(0,1), \text{ under } H_0 \Big]$$

For the given r.s. $x_1, x_2, \dots, x_n$ if $H_0 : \mu = \mu_0$ is true, that we can except that $\bar{x} \simeq \mu_0$, i.e. $(\bar{x} - \mu_0)$ is small. If $|\bar{x} - \mu_0|$ is quite large (positive) quantity, then we suspect H_0. Comparing the deviation $(\bar{x} - \mu_0)$ w.r.t. its S.E., i.e. $S.E.(\bar{X} - \mu_0) = \sqrt{V(\bar{X} - \mu_0)} = \sqrt{V(\bar{X})} = \frac{\sigma}{\sqrt{n}}$,

if $\left| \frac{\bar{x} - \mu_0}{\sigma/\sqrt{n}} \right| > c$, where c is the quantity which is sufficiently large, then we reject $H_0 : \mu = \mu_0$ in favour of $H_1 : \mu \neq \mu_0$.

If we assign the probability of false rejection of H_0 as a small quantity α, then 'c' is so chosen that

$$P_{H_0}\left[\left| \frac{\bar{X} - \mu_0}{\sigma/\sqrt{n}} \right| > c \right] = \alpha.$$

[In particular, $P_{H_0}\left[\left| \frac{\sqrt{n}(\bar{X} - \mu_0)}{\sigma} \right| > 2.576 \right] = 0.01$,

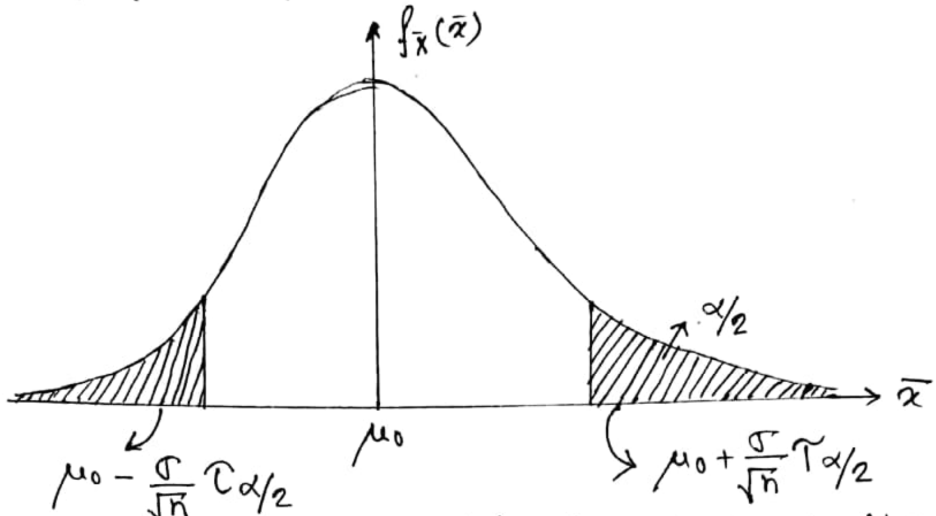
i.e. in repeated sampling from the population under $H_0 : \mu = \mu_0$ in only one out of hundred samples, the value $\left| \frac{\sqrt{n}(\bar{X} - \mu_0)}{\sigma} \right|$ is expected to exceeds 2.576. If in an observed sample $\left| \frac{\sqrt{n}(\bar{x} - \mu_0)}{\sigma} \right|$ exceeds 2.576, then it means that the value has been obtained which is very improbable under H_0 and as the sample is regarded as most likely samples, we shall suspect H_0. Hence, the fact " $P_{H_0}\left[\left| \frac{\sqrt{n}(\bar{x} - \mu_0)}{\sigma} \right| > 2.576 \right] = 0.01$ " provides a test for H_0 against H_1.]

Therefore, we reject $H_0: \mu = \mu_0$ against $H_1: \mu \neq \mu_0$ at level of significance α iff the observed value

$$\left| \frac{\sqrt{n}(\bar{x} - \mu_0)}{\sigma} \right| > c \text{ , where } c \text{ is so chosen that}$$

$$P_{H_0}\left[\left| \frac{\sqrt{n}(\bar{x} - \mu_0)}{\sigma} \right| > c \right] = \alpha .$$

Hence, a test of this kind is called <u>a test of significance</u>.



<u>Test of Significance related to a Univariate Normal Distribution :-</u>

Let $X_1, X_2, \ldots, X_n$ be an observed random sample from $N(\mu, \sigma^2)$ distn.

Define, $\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i$ and $s^2 = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{x})^2$

(a) $\boxed{\text{To test } H_0: \mu = \mu_0 \text{ against } H_1: \mu \neq \mu_0}$ [WBSU'11]

<u>Case I</u> : <u>σ known</u>

Under $H_0: \mu = \mu_0$, we can except that $(\bar{x} - \mu_0)$ is small. If the observed value of $|\bar{x} - \mu_0|$ is quite large in magnitude, then we shall suspect $H_0: \mu = \mu_0$ and support $H_1: \mu \neq \mu_0$. Now comparing the deviation $(\bar{x} - \mu_0)$ w.r.t. its S.E., i.e. S.E.$(\bar{X} - \mu_0) = \frac{\sigma}{\sqrt{n}}$, if the observed value $\left| \frac{\sqrt{n}(\bar{x} - \mu_0)}{\sigma} \right| > c$, where c is sufficiently large, then we reject H_0 infavour of $H_1: \mu \neq \mu_0$.

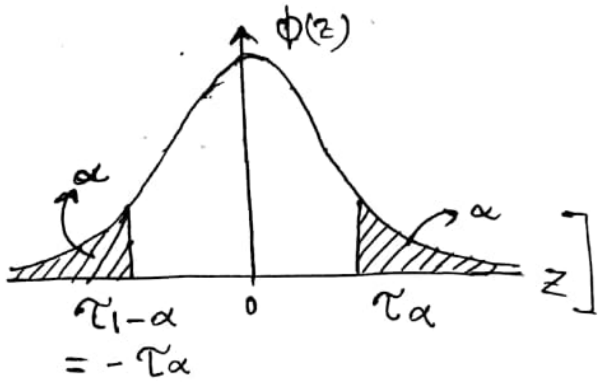
Let, the level of significance be α.

Here, 'c' is so chosen that

$$P_{H_0}\left[\left| \frac{\sqrt{n}(\bar{x} - \mu_0)}{\sigma} \right| > c \right] = \alpha \text{ , Note that, under } H_0 \text{,}$$

$$\bar{X} \sim N\left(\mu_0, \frac{\sigma^2}{n}\right) \Rightarrow \frac{\sqrt{n}(\bar{X} - \mu_0)}{\sigma} \sim N(0,1) .$$

[If $Z \sim N(0,1)$, then $P[Z > \tau_\alpha] = \alpha$ and $P[Z > \tau_{1-\alpha}] = 1-\alpha$, clearly, $\tau_{1-\alpha} = -\tau_\alpha$. τ_α is known as the upper-α point of standard normal deviate.



]

Hence, $\alpha = P_{H_0}[|Z| > c]$

$$= 2P[Z > c]$$

$$\Rightarrow P[Z > c] = \alpha/2$$

$$\Rightarrow c = \tau_{\alpha/2}.$$

Hence, we reject $H_0 : \mu = \mu_0$ against $H_1 : \mu \neq \mu_0$ at level of significance α iff the observed value

$$\left| \frac{\sqrt{n}(\bar{x} - \mu_0)}{\sigma} \right| > \tau_{\alpha/2}.$$

Remark :-

(1)

Null Hypothesis	Case	Alternative Hypothesis	Critical region
$H_0 : \mu = \mu_0$	σ known	i) $H_1 : \mu > \mu_0$ ii) $H_1 : \mu < \mu_0$	i) $\frac{\sqrt{n}(\bar{x}-\mu_0)}{\sigma} > \tau_\alpha$ ii) $\frac{\sqrt{n}(\bar{x}-\mu_0)}{\sigma} < -\tau_\alpha$.

(2)

By acceptance of a hypothesis, we don't mean that it is proved to be true. All that is implied is that so far as the given sample is concern, we find no reason to question the validity of the hypothesis. Nor does rejection of H_0 mean a disprove of H_0. It means simple that, in the light of the given sample, H_0 does not seem to be a plausible hypothesis.

Case II : σ unknown

[Student's t-test]

Under $H_0 : \mu = \mu_0$ (we have, $E(\bar{X}) = \mu_0$), we can expect that $(\bar{x} - \mu_0)$ is small. If the observed value $(\bar{x} - \mu_0)$ is large in magnitude, we suspect H_0 and indicates support to $H_1 : \mu \neq \mu_0$. Now, comparing the deviation $(\bar{x} - \mu_0)$ w.r.t. an estimate of its S.E., i.e. $\hat{S.E}(\bar{x} - \mu_0) = \frac{\hat{\sigma}}{\sqrt{n}} = \frac{s}{\sqrt{n}}$, if the observed value

$$\left| \frac{\sqrt{n}(\bar{x} - \mu_0)}{s} \right| > c$$

, c is sufficiently large, we reject H_0 in favour of $H_1 : \mu \neq \mu_0$, where c is so chosen that

$$P_{H_0}\left[\left| \frac{\sqrt{n}(\bar{X} - \mu_0)}{S} \right| > c \right] = \alpha$$

, the level of significance. —— (*)

Now, under $H_0 : \mu = \mu_0$,

$$\left.\begin{array}{l} \bar{X} \sim N\left(\mu_0, \frac{\sigma^2}{n}\right) \\ \frac{(n-1)S^2}{\sigma^2} \sim \chi^2_{n-1} \end{array}\right\} \text{independently.}$$

$$\Rightarrow \left\{\begin{array}{l} \frac{\sqrt{n}(\bar{X} - \mu_0)}{\sigma} \sim N(0,1) \\ \frac{(n-1)S^2}{\sigma^2} \sim \chi^2_{n-1} \end{array}\right\} \text{independently.}$$

By definition of t-distribution,

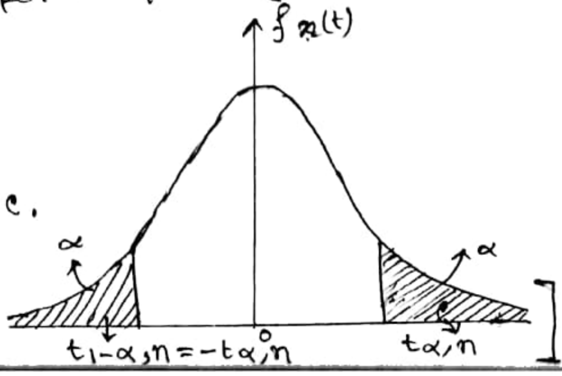
$$\frac{\frac{\sqrt{n}(\bar{X} - \mu_0)}{\sigma}}{\sqrt{\frac{(n-1)S^2}{\sigma^2} \Big/ (n-1)}} \sim t_{n-1}, \text{ under } H_0$$

$$\Rightarrow \frac{\sqrt{n}(\bar{X} - \mu_0)}{S} \sim t_{n-1}, \text{ under } H_0.$$

[If $t \sim t_n$ - distn, then $P[t > t_{\alpha,n}] = \alpha$ and
$P[t > t_{1-\alpha,n}] = 1 - \alpha$.

Clearly, $t_{1-\alpha,n} = -t_{\alpha,n}$.
Here $t_{\alpha,n}$ is known as the upper-α point of the t-distn with n degree of freedom.

If $Z \sim N(0,1)$, then
$P[|Z| > c] < P[|t| > c]$, for large c.



]

From (*), $\alpha = P_{H_0}[|t|>c]$, where $t \sim t_{n-1}$, under H_0.

$$= 2P_{H_0}[t>c]$$

$$\Rightarrow P_{H_0}[t>c] = \alpha/2$$

$$\Rightarrow c = t_{\alpha/2, \overline{n-1}}.$$

Hence, we reject $H_0: \mu = \mu_0$ against $H_1: \mu \neq \mu_0$ at α level of significance iff the observed value

$$\left| \frac{\sqrt{n}(\bar{x}-\mu_0)}{s} \right| > t_{\alpha/2, \overline{n-1}}.$$

Remark:-

Null hypothesis	Case	Alternative Hypothesis	Critical region
$H_0: \mu = \mu_0$	σ unknown	i) $H_1: \mu > \mu_0$ ii) $H_1: \mu < \mu_0$	i) $\frac{\sqrt{n}(\bar{x}-\mu_0)}{s} \sim t_{\alpha, n-1}$ ii) $\frac{\sqrt{n}(\bar{x}-\mu_0)}{s} \sim -t_{\alpha, n-1}$

(b) $\boxed{\text{To test } H_0: \sigma = \sigma_0 \text{ against } H_1: \sigma \neq \sigma_0}$

Case I: μ known

Note that $s_0^2 = \frac{1}{n-1}\sum_{i=1}^{n}(x_i-\mu)^2$ is a measure of population variability σ^2. Under $H_0: \sigma = \sigma_0$, we can except that $s_0^2 \simeq \sigma_0^2$, i.e. $\frac{s_0^2}{\sigma_0^2} \simeq 1$. If the observed value $\frac{s_0^2}{\sigma_0^2}$ is either quite smaller or quite larger than unity, then we shall suspect $H_0: \sigma = \sigma_0$ and give support to $H_1: \sigma \neq \sigma_0$. If the observed value $\frac{s_0^2}{\sigma_0^2} < c_1$ (<1) or $\frac{s_0^2}{\sigma_0^2} > c_2$ (>1), c_1 is sufficiently smaller and c_2 is sufficiently larger than '1', then we reject $H_0: \sigma = \sigma_0$ in favour of $H_1: \sigma \neq \sigma_0$. If we assign the level of significance α, then c_1 and c_2 are so chosen that $P_{H_0}\left[\frac{s_0^2}{\sigma_0^2} < c_1 \text{ or } > c_2\right] = \alpha$.

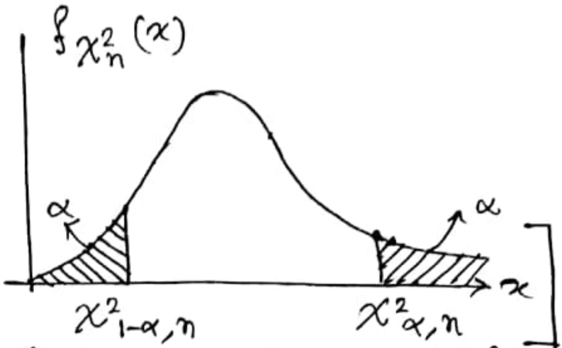
Note that, under H_0,

$$\frac{s_0^2}{\sigma_0^2} = \frac{\sum_{i=1}^{n}(x_i-\mu)^2}{n\sigma_0^2}, \text{ where } \frac{\sum_{i=1}^{n}(x_i-\mu)^2}{\sigma_0^2} \sim \chi^2_n.$$

[Here, $P[\chi^2_n > \chi^2_{\alpha,n}] = \alpha$

and $P[\chi^2_n > \chi^2_{1-\alpha,n}] = 1-\alpha$

$\chi^2_{\alpha,n}$ is known as the upper-α point of χ^2-distn with n d.f.



]

We assign equal probability to tails of the sampling distn. of $\dfrac{\Sigma(x_i-\mu)^2}{\sigma_0^2}$.

Hence, $P_{H_0}\left[\dfrac{s_0^2}{\sigma_0^2} < c_1\right] = \dfrac{\alpha}{2} = P_{H_0}\left[\dfrac{s_0^2}{\sigma_0^2} > c_2\right]$

$$\Rightarrow P_{H_0}\left[\frac{n s_0^2}{\sigma_0^2} < n c_1\right] = \frac{\alpha}{2} = P_{H_0}\left[\frac{n s_0^2}{\sigma_0^2} > n c_2\right]$$

$$\Rightarrow P_{H_0}\left[\chi^2_n < n c_1\right] = \frac{\alpha}{2} = P_{H_0}\left[\chi^2_n > n c_2\right]$$

$$\Rightarrow n c_1 = \chi^2_{1-\alpha/2,n} \; ; \; n c_2 = \chi^2_{\alpha/2,n}.$$

Hence, we reject $H_0: \sigma = \sigma_0$ against $H_1: \sigma \neq \sigma_0$ at α level of significance iff the observed value

$$\frac{\Sigma(x_i-\mu)^2}{\sigma_0^2} < \chi^2_{1-\frac{\alpha}{2},n}, \text{ or,}$$

$$\frac{\Sigma(x_i-\mu)^2}{\sigma_0^2} > \chi^2_{\frac{\alpha}{2},n}.$$

$\left[\dfrac{s_0^2}{\sigma_0^2} < c_1 \Rightarrow \dfrac{\Sigma(x_i-\mu)^2}{\sigma_0^2} < n c_1 = \chi^2_{1-\frac{\alpha}{2},n}\right]$

In a testing problem, depending on the nature of the alternative hypothesis, if the left tail / right tail / both the tails of the curve of the sampling distn. of the test statistic is used for defining the critical region, then the test is called the left tailed / right tailed / two-tailed test.

In testing $H_0: \mu = \mu_0$ against $H_1: \mu > \mu_0$ for a $N(\mu, \sigma^2)$ popln., critical region: $\dfrac{\sqrt{n}(\bar{x}-\mu_0)}{s} > t_{\alpha,n-1}$. The value $t_{\alpha,n-1}$ is known as the critical value.

The critical value for a given level of significance (α) in the boundary of the acceptance region of a test of a testing problem.

(***)

Case II : μ unknown

Here $s^2 = \frac{1}{n-1}\sum_{i=1}^{n}(x_i-\mu)^2$ is a measure of a population variance (σ^2). Under $H_0 : \sigma = \sigma_0$, we can expect that

$\frac{s^2}{\sigma_0^2} \simeq 1$. If the observed $\frac{s^2}{\sigma_0^2}$ is quite large or small relative to 1, we suspect H_0 and give support to $H_1 : \sigma \neq \sigma_0$.

If the observed value of $\frac{s^2}{\sigma_0^2} > c_1 \,(>1)$ or $\frac{s^2}{\sigma_0^2} < c_2 \,(<1)$, then we reject H_0 in favour of $H_1 : \sigma \neq \sigma_0$, where c_1 and c_2 are so chosen that

$$P_{H_0}\left[\frac{s^2}{\sigma_0^2} > c_1 \text{ or } < c_2\right] = \alpha$$

Now, $\frac{(n-1)s^2}{\sigma_0^2} \sim \chi^2_{n-1}$, under H_0.

We assign equal error probabilities to both the tails of the distribution:

$$P_{H_0}\left[\frac{s^2}{\sigma_0^2} > c_1\right] = \frac{\alpha}{2} = P_{H_0}\left[\frac{s^2}{\sigma_0^2} < c_2\right]$$

Now, $$\frac{\alpha}{2} = P_{H_0}\left[\frac{(n-1)s^2}{\sigma_0^2} > (n-1)c_1\right] = P_{H_0}\left[\chi^2_{n-1} > (n-1)c_1\right]$$

$$\Rightarrow (n-1)c_1 = \chi^2_{\alpha/2;\, n-1}$$

Similarly; $(n-2)c_2 = \chi^2_{1-\alpha/2;\, n-1}$

Hence we reject $H_0 : \sigma = \sigma_0$ against $H_1 : \sigma \neq \sigma_0$ at α-level of significance if the obseroverved

$$\frac{(n-1)s^2}{\sigma_0^2} > \chi^2_{\alpha/2;\, n-1} \text{ or } < \chi^2_{1-\alpha/2;\, n-1}$$

Remark:-

ll. Hyp.	Case	Alt. hyp.	Critical region.
$H_0 : \sigma = \sigma_0$	μ known	i) $H_1 : \sigma > \sigma_0$ ii) $H_1 : \sigma < \sigma_0$	i) $\frac{\sum(x_i-\mu)^2}{\sigma_0^2} > \chi^2_{\alpha,n}$ ii) $\frac{\sum(x_i-\mu)^2}{\sigma_0^2} < \chi^2_{1-\alpha,n}$
	μ unknown	i) $H_1 : \sigma > \sigma_0$ ii) $H_1 : \sigma < \sigma_0$	i) $\frac{(n-1)s^2}{\sigma_0^2} > \chi^2_{\alpha,n-1}$ ii) $\frac{(n-1)s^2}{\sigma_0^2} < \chi^2_{1-\alpha,n-1}$

Test Statistic : In a testing problem, a test procedure or critical region is defined in terms of a statistic. Then the statistic is called the test statistic of the test of the testing problem.

In the testing problem of testing $H_0 : \mu = \mu_0$ in $N(\mu, \sigma^2)$ population, the test statistic is

$$T = \frac{\sqrt{n}(\bar{x} - \mu_0)}{s}$$

Critical value or point : —

In a testing problem, the boundary point (value) of the acceptance region of a test is called the critical point (value) of the test.

. In a test : reject H_0 iff $\frac{\sqrt{n}(\bar{x} - \mu_0)}{s} > t_{\alpha, n-1}$ of $H_0 : \mu = \mu_0$ against $H_1 : \mu > \mu_0$, the point/value $t_{\alpha, n-1}$ is the critical value of the test.

Left tailed, Right tailed, both tailed tests: In a testing problem, it is a test procedure or critical region uses the left/ right/ both tails of the curve of the sampling distribution of the test statistic is defining values that lead to the rejection of null hypothesis..

p-values/probability values : The choice of a specific α is completely arbitary and is determined by non-statistical considerations such as the possible consequences of rejecting H_0 falsely and the economic and practical implications of the decision to reject H_0. There is another value associated with a statistical test, it is called the probability value or the p-value.

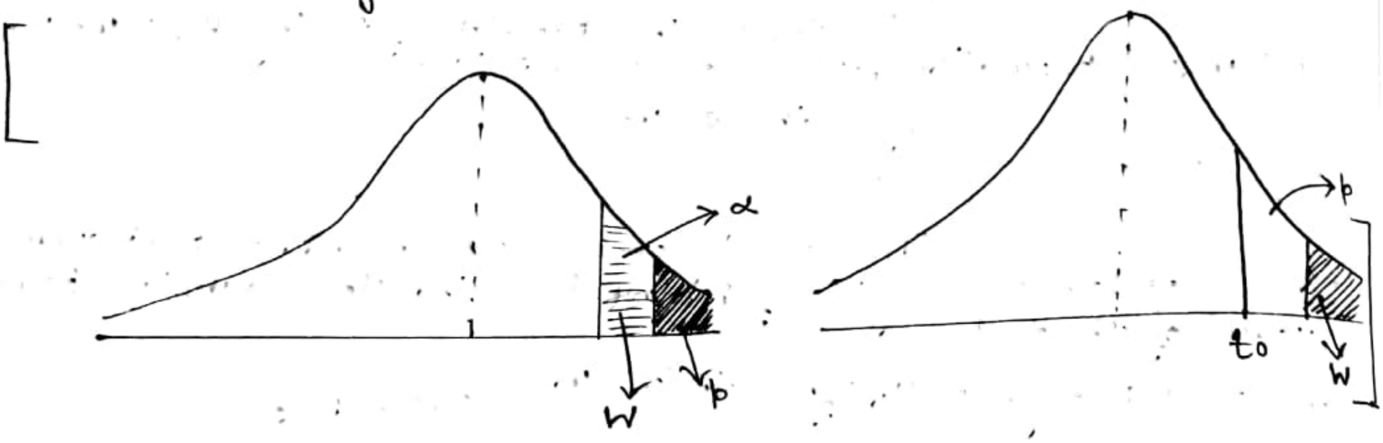
Definition :- The value associated with a test, is a probability that we obtains the observed value of the test statistic or a value that is more extreme in the direction given by the alternative hypothesis when H_0 is true.

[For example, let $X \sim N(\mu, \sigma^2)$. To test $H_0 : \mu = 4$ against $H_1 : \mu > 4$. If we take a r.s. of size $n = 9$ and we are given $\bar{x} = 4{\cdot}3$, $s = 1{\cdot}2$, then the observed value of the test statistic T is $t_0 = \frac{\sqrt{n}(\bar{x} - 4)}{s} = \frac{\sqrt{9}(4{\cdot}3 - 4)}{1{\cdot}2} = 0{\cdot}75$.

Then the p-value is
$$\begin{aligned} p &= P_{H_0}[T \geq t_0] \\ &= P_{H_0}[t_8 \geq 0{\cdot}75] \\ &= 0{\cdot}24. \end{aligned}$$
]

The smaller the p-value, the more extreme the outcome the stronger the evidence against H_0.

If α is the given level of significance, we reject H_0 if $p \leq \alpha$ and we don't reject H_0 if $p > \alpha$.



Rather than selecting the critical region is advance with a particular level α and giving the conclusion, the p-value of a test can be reported and reader ultimately makes a decision for any level α.

For both sided alternative, the p-value $= P_{H_0}\left[|T| \geq |t_0|\right]$

$$= 2P_{H_0}\left[T \geq |t_0|\right],$$

where, the distribution of T is symmetric about zero. If the distribution of T is not symmetric, then the p-value is not well defined for both sided alternative, we define the p-value as

$$p = 2\left\{\text{smaller of the two one-sided p-values}\right\}$$

$$= 2 \cdot \text{minimum}\left\{P_{H_0}\left[T \geq t_0\right], P_{H_0}\left[T \leq t_0\right]\right\}$$

—— x ——

Test related to Population Proportion :— (For Binomial Distn.)

(1) Single Proportion :— Let 'p' be the proportion of individuals possessing a character 'A', and p is unknown, in an infinite popln.

Consider a r.s. of size n drawn from the popln. Let X be the no. of members in the sample possessing the character A.

To test $H_0 : p = p_0$.

Here $X \sim Bin(n, p)$, Under H_0, $X \sim Bin(n, p_0)$.
Let x_0 be the observed value of X, in a given sample.

(a) $\boxed{H_1 : p > p_0}$

If $p > p_0$, we can expect that $x > x_0$, as the success probability increases we can expect larger numbers of successes in a sample.

Here, the p-value $= P[X \geq x_0 / H_0]$

$$= \sum_{x \geq x_0} \binom{n}{x} p_0^x (1-p_0)^{n-x}.$$

If the p-value $\leq \alpha$, the level of significance, we shall consider x_0 to be an unlikely value under H_0 and reject H_0.
If p value $> \alpha$, H_0 is accepted at α level of significance.

(b) $\boxed{H_1 : p < p_0}$ The p-value $= P[X \leq x_0 / H_0]$

$$= \sum_{x=0}^{x_0} \binom{n}{x} p_0^x (1-p_0)^{n-x}$$

If p-value $\leq \alpha$, reject H_0 and p-value $> \alpha$, accept H_0 at 'α'-level.

(c) $\boxed{H_1 : p \neq p_0}$

Then p-value $= 2 \cdot \min \{ P_{H_0}[X \geq x_0], P_{H_0}[X \leq x_0] \}$

If p-value $\leq \alpha$, reject H_0 and p-value $> \alpha$, accept H_0 at 'α' level.

(2) Two proportions :— Let p_1 and p_2 be the proportions of individuals having a characteristics A in two infinite popln.. Let X_1 and X_2 denote the nos. of members having the characteristics A in the random samples of size n_1 and n_2 drawn independently from the two populations. To test $H_0: p_1 = p_2$, let $X = X_1 + X_2$. Under $H_0: p_1 = p_2 = p$, say. $X = X_1 + X_2 \sim Bin(n_1+n_2, p)$, where $X_1 \sim Bin(n_1, p)$, $X_2 \sim Bin(n_2, p)$, independently.

Under Ho, the conditional distn. of X_1 given that $X_1 + X_2 = x$ is given by the PMF :

$$P[X_1 = x_1 / X_1 + X_2 = x] = \frac{\binom{n_1}{x_1}\binom{n_2}{n - x_1}}{\binom{n_1+n_2}{x}}, \quad x_1 = 0, 1, \ldots, n_1,$$ which is independent of 'p'.

If for given r.s's, the observed value of X_1 is x_{10} and that of X is x_0, then we have $P_{H_0}[X_1 = x_1 / X_1 + X_2 = x_0]$

$$= \frac{\binom{n_1}{x_1}\binom{n_2}{x_0 - x_1}}{\binom{n_1+n_2}{x_0}}, \quad x_0 = 0(1)n_1.$$

(a) $\boxed{H_1: p_1 > p_2}$ The p-value $= P_{H_0}[X_1 \geq x_{10} | X_1 + X_2 = x_0]$

$$= \sum_{x_1 \geq x_0} \frac{\binom{n_1}{x_1}\binom{n_2}{x_0 - x_1}}{\binom{n_1+n_2}{x_0}}$$

[N.T. if $p_1 > p_2$, we can expect large value of X_1 given the total $X_1 + X_2 = x_0$]

If the p-value $\leq \alpha$, reject Ho and if the p-value $> \alpha$, accept Ho at 'α' level of significance.

(b) $\boxed{H_1 : p_1 < p_2}$ The p-value $= P_{H_0}[X_1 \leq x_{10} | X_1 + X_2 = x_0]$

$$= \sum_{x_1 \leq x_{10}} \frac{\binom{n_1}{x_1}\binom{n_2}{x_0 - x_1}}{\binom{n_1+n_2}{x_0}}$$

[N.T. if $p_1 < p_2$, we can except large value of X_1 given the total $X_1 + X_2 = x_0$]

If the p-value $\leq \alpha$, reject Ho and if the p-value $> \alpha$, accept Ho at 'α' level of significance.

(c) $\boxed{H_1 : p_1 \neq p_2}$

The p-value $= 2 \min \{ P_{H_0}[X_1 \geq x_0 / X_1 + X_2 = x_0], P_{H_0}[X_1 \leq x_0 / X_1 + X_2 = x_0] \}$

If p-value $\leq \alpha$, we reject Ho and if the p-value $> \alpha$, accept Ho, at α-velvel of significance.

Tests Related to Poisson Distribution:—

(1) Single Populations:— Let $X_1, X_2, \dots, X_n$ be a r.s. from a $P(\lambda)$ popln, λ unknown. To test $H_0: \lambda = \lambda_0$.

Note that, $Y = \sum_{i=1}^{n} X_i \sim P(n\lambda_0)$.

For a given r.s. $x_1, x_2, \dots, x_n$, let y_0 be the observed value of Y.

(a) $\boxed{H_1: \lambda > \lambda_0}$

If $\lambda > \lambda_0$, we can expect $y > y_0$

the p-value $= P_{H_0}[Y \geq y_0]$

$$= \sum_{y=y_0}^{\infty} e^{-n\lambda_0} \cdot \frac{(n\lambda_0)^y}{y!} = p, \text{ say}.$$

If $p \leq \alpha$, we reject H_0 and if $p > \alpha$, accept H_0 at 'α' level.

(b) $\boxed{H_1: \lambda < \lambda_0}$

If $\lambda < \lambda_0$, we can expect $y < y_0$

the p-value $= P_{H_0}[Y \leq y_0]$

$$= \sum_{y=0}^{y_0} e^{-n\lambda_0} \cdot \frac{(n\lambda_0)^y}{y!} = p, \text{ say}$$

If $p \leq \alpha$, reject H_0 and if $p > \alpha$, accept H_0 at 'α' level.

(c) $\boxed{H_1: \lambda \neq \lambda_0}$

p-value $= 2 \min\{P_{H_0}[Y \geq y_0], P_{H_0}[Y \leq y_0]\}$

If $p \leq \alpha$, reject H_0 and if $p > \alpha$, accept H_0 at α-level of significance.

(2) Two populations :- Let $X_{11}, X_{12}, \ldots, X_{1n_1}$ be a r.s. from $P(\lambda_1)$

& $X_{21}, X_{22}, \ldots, X_{2n_2}$ " " " " $P(\lambda_2)$

drawn independently.

Here, $Y_1 = \sum_{i=1}^{n_1} X_{1i} \sim P(n_1 \lambda_1)$
$Y_2 = \sum_{i=1}^{n_2} X_{2i} \sim P(n_2 \lambda_2)$ independently,

Then $Y = Y_1 + Y_2 \sim P(n_1\lambda_1 + n_2\lambda_2)$ under Ho. and, Condition distn. of Y_1 given $Y=y$ is $Bin\left(y, \frac{n_1}{n_1+n_2}\right)$.

To test $H_0 : \lambda_1 = \lambda_2$:-

Under Ho, $P[Y_1 = y_1 / Y_1 + Y_2 = y]$

$$= \binom{y}{y_1} \left(\frac{n_1}{n_1+n_2}\right)^{y_1} \left(\frac{n_2}{n_1+n_2}\right)^{y-y_1}, \text{ where } \lambda_1 = \lambda_2 = \lambda \text{ (say)}$$

Let, for given r.s.'s, the observed value of Y and Y_1 are y_0 and y_{10}, respectively.

Here, test will be based on the statistic Y_1 given $Y = y_0$, whose distn. is free from λ, under $H_0 : \lambda_1 = \lambda_2 = \lambda$.

(a) $\boxed{H_1 : \lambda_1 > \lambda_2}$

The p-value $= P_{H_0}[Y_1 \geq y_{10} / Y = y_0]$

$$= \sum_{y_1 \geq y_{10}} \binom{y_0}{y_1} \left(\frac{n_1}{n_1+n_2}\right)^{y_1} \left(\frac{n_2}{n_1+n_2}\right)^{y_0-y_1} = p, \text{ say}.$$

If $p \leq \alpha$, reject Ho and if $p > \alpha$, accept Ho at 'α'-level.

(b) $\boxed{H_1 : \lambda_1 < \lambda_2}$

The p-value,

$$p = P_{H_0}[Y_1 \leq y_{10} / Y = y_0]$$
$$= \sum_{y_1 \leq y_{10}} \binom{y_0}{y_1} \left(\frac{n_1}{n_1+n_2}\right)^{y_1} \left(\frac{n_2}{n_1+n_2}\right)^{y_0-y_1}$$

(c) $\boxed{H_1 : \lambda_1 \neq \lambda_2}$

The p-value,

$$p = 2 \min \left\{ P_{H_0}[Y_1 \leq y_{10} | Y = y_0], P_{H_0}[Y \geq y_{10} / Y = y_0] \right\}$$

If $p \leq \alpha$, reject Ho and if $p > \alpha$, accept Ho at α level of significance.

C.U. 2009

Ex. Let $X_1 \sim Bin(n_1, p_1)$ and $X_2 \sim Bin(n_2, p_2)$ independently. Two test are to be performed (i) $H_{01}: p_1 = \frac{1}{2}$ against $H_{11}: p_1 \neq \frac{1}{2}$,

(ii) $H_{02}: p_1 = p_2$ against $H_{12}: p_1 \neq p_2$.

(a) Are the null hypothesis H_{01} and H_{02} simple or composit?

(b) Describe the test procedure in (i).

(c) Like the test in (i), why cannot a test based on binomial distn. be constructed for (ii). Describe how this can be performed.

(d) Suppose H_{02} is accepted and we can assume $p_1 = p_2$. How will you test $H_{03}: p_1 = p_2 = \frac{1}{2}$ against $H_{13}: p_1 = p_2 \neq \frac{1}{2}$.

Hints:-

(a)

(i) $\rightarrow$ simple hypothesis as the value of p_1 is known.

(ii) $\rightarrow$ composite hypothesis, as " " " $p_1 = p_2$ is not known / unknown.

(b) To test $H_{01}: p_1 = \frac{1}{2}$ against $H_{11}: p_1 \neq \frac{1}{2}$

Here $X_1 \sim Bin(n_1, p_1)$ and $X_2 \sim Bin(n_2, p_2)$.

Under H_{01}.

Let the observed value of X_1 is x_{10},

Note that, under H_{01}, the distn. of X_1, i.e. $Bin(n, \frac{1}{2})$ is symmetric.

Under H_{01}, we can expect $x_{10} \simeq \frac{n_1}{2}$.

i.e. $(x_{10} - \frac{n_1}{2})$ is small, infact $E(X_1) = \frac{n_1}{2}$.

If $p_1 \neq \frac{1}{2}$, then we can expect $\left|x_{10} - \frac{n_1}{2}\right|$ is a large quantity.

Here, $\left|X_1 - \frac{n_1}{2}\right|$ is the test - statistic.

Now, the p-value, $= P_{H_0}\left[\left|X_1 - \frac{n_1}{2}\right| \geq \left|x_{10} - \frac{n_1}{2}\right| = d_0\right]$,

as the distn. X_1 is symmetric, under H_{01}.

$$= \sum_{\left|x_1 - \frac{n_1}{2}\right| \geq d_0} \binom{n_1}{x} \cdot \frac{1}{2^{n_1}}$$

$$= \frac{1}{2^{n_1}} \left\{ \sum_{x_1 \leq \frac{n_1}{2} - d_0} \binom{n_1}{x_1} + \sum_{x_1 \geq \frac{n_1}{2} + d_0} \binom{n_1}{x_1} \right\}$$

(c) To test $H_{11}: p_1 = p_2$ against $H_{12}: p_1 \neq p_2$.
N.T. $X_1 + X_2 \sim Bin(n_1+n_2, p)$, under $H_{11}: p_1 = p_2 = p$ (unknown)
Here, H_{11} is composite, the distn. of $X_1 + X_2$ is not completely known. Hence, a test, like the test (i), can't be performed based on $X_1+X_2 \sim Bin(n_1+n_2, p)$, as the p-value cannot be obtainable, as they depend on unknown p. Therefore, the test of (ii) based on (X_1+X_2) can't be performed as the test of (i) based on X_1.

(d). To test $H_{03}: p = 1/2$ against $H_{13}: p \neq 1/2$

If H_{02} is accepted, then $p_1 = p_2 = p$ (say)
Here, $X = X_1+X_2 \sim Bin(n_1+n_2, p)$.
[The testing procedure is same as part (b)]

C.U. 2008

Ex. Suppose X_1 and X_2 are two independently poisson r.v.s with $E(X_k) = \mu_k$, $k=1,2$. Find the regression coefficient (β) of X_1 on X_1+X_2. Carry out a suitable exact test for $H_0: \beta = \frac{1}{2}$ against $H_1: \beta \neq \frac{1}{2}$.

Hints:-
$$X_1 / X_1+X_2 = x \sim Bin\left(x, \frac{\mu_1}{\mu_1+\mu_2}\right)$$

The regression of X_1 on (X_1+X_2) is:
$$E(X_1 / X_1+X_2 = x) = x \cdot \frac{\mu_1}{\mu_1+\mu_2}$$, which is linear in x.

Clearly, $\beta = \frac{\mu_1}{\mu_1+\mu_2}$. Hence $H_0: \beta = \frac{1}{2} \Leftrightarrow H_0: \mu_1 = \mu_2$

To test $H_0: \mu_1 = \mu_2$ against $H_1: \mu_1 \neq \mu_2$

☆ Ex.1. Distinguish between i) simple hypothesis and composite hypothesis, ii) confidence interval and acceptance region.

Ans:- i) A simple hypothesis is defined as the hypothesis which completely specifies the random vector together with the basic assumption. On the other hand a composite hypothesis is defined as the hypothesis which does not specify completely the distribution of the random vector together with the basic assumption.

Let H be a hypothesis, $H : \{F_\theta : \theta \in \Theta_0\}$, $\Theta_0 \subset \Theta$. Now, if Θ_0 is a singleton set then H is a simple hypothesis. On the other hand, if Θ_0 consists more than one point then the hypothesis H is composite.

For e.g., for $N(\mu, 1)$ popl^n.

$H : \mu = 1$ Vs $H_1 : \mu < 1$.

Here H is a simple hypothesis and H_1 is a composite hypothesis.

ii) When we define a test i.e. we either accept or reject the null hypothesis, we consequently partition the sample space w.r.t. a critical value of the statistic obtained from the sample. The partition of the sample space for which the value of the statistic is such that we accept the null hypothesis, is called the acceptance region.

On the other hand, confidence interval means a region in which the true value of the parametric function lies, i.e. the formation of confidence interval is far from the concept of point estimation of the parametric function.

So, it is necessary for a hypothesis testing problem that if the given parametric function lies on its confidence interval on $100(1-\alpha)\%$ confidence interval, then the null hypothesis is accepted, i.e. the realized value of the statistic lies on the acceptance region at $100\alpha\%$ level of significance.

☆ Ex.2. Explain the concept of test of significance. Discuss the notions of two types of errors and their relations with the level of significance and power of a test is testing statistical hypothesis.

ANS:-

Test of significance: The test of significance is a rule obtained on the basis of sample observations by which we accept or reject a null hypothesis. Note that to define a test is equivalent to partition the sample space into two disjoint sets. Let us consider the problem of testing the mean of a normal distribution vanishes against it is unity based on a sample of size 4 given that the population s.d. is unity.

Here we reject the null hypothesis if the sample mean $\bar{x}$ exceeds $\cdot 823$, o/w we accept it. Clearly, this decision rule is the test and we can always define a set W, $W = \{\underset{\sim}{x} : \bar{x} > \cdot 823\}$, So, the sample space $\mathscr{X}$ is partitioned into W and $\mathscr{X} - W$. Clearly we reject the null hypothesis if $\underset{\sim}{x} \in W$ and accept it if $\underset{\sim}{x} \in \mathscr{X} - W$. The region W is reffered to as the critical region or rejection region and the $\mathscr{X} - W$ is termed as acceptance region.

Type-I and Type-II error:

Since the decision rule regarding the rejection or acceptance of null hypothesis solely depends on the realised value of the random vector, one may commit two types of errors. The first kind of error is rejecting the null hypothesis, even when the hypothesis is true. This error is termed as type-I error. The second kind of error is accepting the null hypothesis, even when the hypothesis is false. This kind of error is called type-II error.

Decision taken / True situation	Accept null	Reject null
Null is true	✓	Type I error
Null is false	Type II error	✓

Relationship with level of significance: Let us consider the following test procedure,

$H_0: \theta = \theta_0$ Vs. $H_1: \theta = \theta_1$

We reject or accept the null hypothesis at $100\alpha\%$ level of significance, i.e. if we repeat the 100 times then atmost α times the true null hypothesis will be rejected.

$$\therefore P[\text{Type I error}] \leq \alpha \quad \text{———} \langle i \rangle$$

Subject to the condition $\langle i \rangle$ we choose that test for which probability of Type II error is least.

We introduce the concept of level of significance for the reason that we can't minimize the probability of type I error and type II error simultaneously. That is why we set an upper bound to the probability of type I error which is termed as the level of significance.

Relationship with power of test: Let us consider the following test procedure, $H_0: \theta = \theta_0$ Vs $H_1: \theta = \theta_1$

Here $\mathscr{X}$ is the sample space of the statistic and we partition $\mathscr{X}$ into W and $\mathscr{X} - W$, where W is the critical region and $\mathscr{X} - W$ is the acceptance region.

$$\therefore P[\text{Type I error}] = P_{\theta_0}[\underset{\sim}{x} \in W]$$

$$\therefore P[\text{Type II error}] = 1 - P_{\theta_1}[\underset{\sim}{x} \in W]$$

Let we accept or reject the null hypothesis at $100\alpha\%$ level of significance

$$P_{\theta_0}[\underset{\sim}{x} \in W] \leq \alpha \quad \text{———} \langle ii \rangle$$

subject to the condition $\langle ii \rangle$ we have to find that test for which $P[\text{type II error}]$ is least, i.e. $1 - P_{\theta_1}[\underset{\sim}{x} \in W]$ is least,

i.e. $P_{\theta_1}[\underset{\sim}{x} \in W]$ is greatest.

The probability $P_{\theta_1}[\underset{\sim}{x} \in W]$ is termed as the power of the test, i.e. the greater the power of test the less the $P[\text{Type II error}]$, i.e. the test is more powerful under equal level of significance.

$\therefore$ Power of a test $= 1 - P[\text{Type II error}]$.

This is the required relation.

☆ **Ex. 3.** To test whether a coin is perfect is tossed five times and the number (X) of heads is noted. If X is 2, 3 or 4, the coin is taken to be perfect. Find the probability of type I error and the power function. ~~Complete~~ Compute the power of the test when probability of getting a head from a single toss of the coin 0·6. Suggest a critical function with higher power.

ANS:- Here X be the RV representing the number of head.

We are to test

$H_0 : p = 1/2$ Vs. $H_1 : p \neq 1/2$, where, p is the probability for getting a head.

The coin is tossed 5 times.

$\therefore$ Under H_0, $X \sim bin(5, 1/2)$

The critical region is given by, $W = \{x : x = 0, 1, 5\}$

$$\therefore P[\text{Type I error}] = P_{H_0}(\underset{\sim}{x} \in W)$$

$$= \binom{5}{0}\left(\frac{1}{2}\right)^5 + \binom{5}{1}\left(\frac{1}{2}\right)^5 + \binom{5}{5}\left(\frac{1}{2}\right)^5$$

$$= \frac{7}{2^5}.$$

For a general $p(\neq 1/2)$, we get the power function of the test,

$$\beta_\theta = P_p(\underset{\sim}{x} \in W)$$

$$= \binom{5}{0}(1-p)^5 + \binom{5}{1}p(1-p)^4 + \binom{5}{5}p^5$$

$$= (1-p)^5 + 5p(1-p)^4 + p^5.$$

Now, let us reset the testing problem as follows,

$H_0 : p = 1/2$ Vs. $H_1 : p = 3/5$

$X \sim bin(5, 3/5)$, under H_1.

Power of the test $= P_{H_1}(\underset{\sim}{x} \in W)$

$$= \binom{5}{0}\left(\frac{2}{5}\right)^5 + \binom{5}{1}\left(\frac{3}{5}\right)\left(\frac{2}{5}\right)^4 + \binom{5}{5}\left(\frac{3}{5}\right)^5$$

$$= \left(\frac{2}{5}\right)^5 + 3\left(\frac{2}{5}\right)^4 + \left(\frac{3}{5}\right)^5.$$