Statistical Inference - I

Code : 18K5S07

UNIT : 3

Methods of Estimation:-

We shall briefly ouline some of the important Methods for obtaining such estimators.

(i) Methods of Maximum Likelihood Estimation.
(ii) Method of Minimum Variance
(iii) Methods of Moments.
(iv) Methods of least square
(v) Method of Minimum Chi-square
(vi) Methods of Inverse probability.

Method of Maximum Likelihood Estimation:-

The most general Method of Estimation known is the Method of Maximum likelihood estimators (MLE).

Liklihood function:-

Let $x_1, x_2, \cdots, x_n$ be a random sample of size n from a population with density function $f(x, \theta)$. Then the Likelihood function of the sample value $x_1, x_2, \cdots, x_n$ usually denoted by $L = L(\theta)$ is their joint density function

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

L gives the relative likelihood that the random variable assume a particular set of values $x_1\, x_2 \ldots x_n$ For a given sample $x_1, x_2, \ldots, x_n$ becomes a function of the variable θ, the parameter.

The principle of maximum likelihood consists in finding an estimator for the unknown parameter $\theta = (\theta_1, \theta_2 \cdots \theta_k)$, say, which maximum the Likelihood function $L(\theta)$ For variations in parameter. ie., We wish to find $\hat{\theta} = (\hat{\theta}_1, \hat{\theta}_2 \cdots \hat{\theta}_k)$ Sothat $L(\hat{\theta}) > L(\theta)\ \forall\ \theta \in \Theta$ ie., $L(\hat{\theta}) = \sup L(\theta)\ \forall\ \theta \in \Theta$

Thus if there exists a function $\hat{\theta} = \hat{\theta}(x_1, x_2 \cdots x_n)$ of the sample value which maximises L for variations in θ, Then $\hat{\theta}$ is to be taken as an estimator of θ. $\hat{\theta}$ is usually called maximum Likelihood Estimator.

$$\frac{\partial L}{\partial \theta} = 0 \text{ and } \frac{\partial^2 L}{\partial \theta^2} < 0$$

Since $L > 0$, and $\log L$ is a non-decreasing function of L; L and $\log L$ attain their extreme values at the same value of $\hat{\theta}$. The first of the two equations can be written as

$$\frac{1}{L} \cdot \frac{\partial L}{\partial \theta} = 0 \Rightarrow \frac{\partial \log L}{\partial \theta} = 0.$$

$$\frac{\partial}{\partial \theta_i} \log L = \frac{\partial}{\partial \theta_i} \log L(\theta_1, \theta_2, \ldots \theta_k) = 0$$
$$i = 1, 2, \ldots k$$

The equations are usually referred to as the likelihood Equations for estimating the parameters.

Properties of Maximum Likelihood Estimators:-

We Make the following assumptions, known as the regularity conditions.

(i) The first and second order derivatives viz., $\frac{\partial \log L}{\partial \theta}$ and $\frac{\partial^2 \log L}{\partial \theta^2}$ exist and are continuous functions of θ in a range R.

for every θ in R $\left| \frac{\partial}{\partial \theta} \log L \right| < f_1(x)$ and $\left| \frac{\partial^2}{\partial \theta^2} \log L \right| < f_2(x)$ Where $f_1(x)$ and $f_2(x)$ in integrable functions over $(-\infty, \infty)$

(ii) The third order derivative $\frac{\partial^3}{\partial \theta^3} \log L$ exists such that $\left| \frac{\partial^3}{\partial \theta^3} \cdot \log L \right| < M(x)$ where $E[M(x)] < k$, a positive quantity.

(iii) for every θ in R.

$$E\left(-\frac{\partial^2}{\partial \theta^2} \log L\right) = \int_{-\infty}^{\infty} \left(-\frac{\partial^2}{\partial \theta^2} \log L\right) L \, dx = I(\theta)$$

is finite and non-zero

(iv) The range of integration is independent of θ. But if the range of integration depend on θ. Then $f(x, \theta)$ Vanishes at the extremes depending in θ.

Example : 1

In random sampling from normal Population $N(\mu, \sigma^2)$, find the maximum Likelihood estimators for

(i) μ when σ^2 is known (ii) σ^2 when μ is known and (iii) the simultaneous estimation of μ and σ^2

Solution :-

$X \sim N(\mu, \sigma^2)$, then

$$L = \prod_{i=1}^{n} \left[\frac{1}{\sigma\sqrt{2\pi}} \exp\left\{ -\frac{1}{2\sigma^2}(x_i-\mu)^2 \right\} \right]$$

$$= \left(\frac{1}{\sigma\sqrt{2\pi}} \right)^n \exp\left\{ -\sum_{i=1}^{n}(x_i-\mu)^2 / 2\sigma^2 \right\}$$

$$\log L = -\frac{n}{2}\log(2\pi) - \frac{n}{2}\log\sigma^2 - \frac{1}{2\sigma^2}\sum_{i=1}^{n}(x_i-\mu)^2$$

Case (i) When σ^2 is known, the MLE equations for estimating μ is

$$\frac{\partial}{\partial\mu}\log L = 0 \Rightarrow \frac{1}{2\sigma^2}\sum_{i=1}^{n} 2(x_i-\mu)(-1) = 0$$

$$\Rightarrow \sum_{i=1}^{n}(x_i-\mu) = 0 \Rightarrow \hat{\mu} = \frac{1}{n}\sum_{i=1}^{n} x_i = \bar{x}$$

Hence MLE for μ is the Sample mean $\bar{x}$.

Cas (ii) When μ is known, the Likelihood equation for estimating σ^2 is :

$$\frac{\partial}{\partial\theta^2}\log L = 0 \Rightarrow -\frac{n}{2}\times\frac{1}{\sigma^2} + \frac{1}{2\sigma^4}\sum_{i=1}^{n}(x_i-\mu)^2 = 0$$

$$\Rightarrow n - \frac{1}{\sigma^2}\sum_{i=1}^{n}(x_i-\mu)^2 = 0 \Rightarrow \hat{\sigma}^2 = \frac{1}{n}\sum_{i=1}^{n}(x_i-\mu)^2$$

Case (iii) The Likelihood equation for simultaneous estimation of μ and σ^2 are

$\frac{\partial}{\partial \mu} \log L = 0$ and $\frac{\partial}{\partial^2} \log L = 0$ thus giving $\hat{\mu} = \bar{x}$

and $\hat{\sigma_2} = 1/n \sum_{i=1}^{n} (x_i - \hat{\mu})^2 = 1/2 \sum_{i=1}^{n} (x_i \bar{x})^2 = s^2$

The sample variance

Example : 2

(a) Find the maximum likelihood estimate for the parameter λ of a poisson distribution on the basis of a sample of size n. Also find its variance.

(b) Show that the sample mean $\bar{x}$, is sufficient for estimating the parameter λ of the poisson distribution.

Solution :-

The probability function of the poisson distribution with parameter λ is given by

$$P(X = x) = f(x, \lambda) = \frac{e^{-\lambda} \lambda^x}{x!} ; x = 0, 1, 2 \ldots$$

Likelihood function of random sample $x_1, x_2, \ldots x_n$ of a observations from this population is.

$$L = \prod_{i=1}^{n} f(x_i, \lambda) = \frac{e^{n\lambda} \lambda^{\sum_{i=1}^{n} x_i}}{x_1! x_2! \ldots x_n!}$$

$$\therefore \operatorname{Log} L = n\lambda + n\bar{x} \log \lambda - \sum_{i=1}^{n} \log (x_i!)$$

The Likelihood equation for estimating λ is

$$\frac{\partial}{\partial \lambda} \log L = 0 \Rightarrow -n + \frac{n\bar{x}}{\lambda} = 0 \Rightarrow \lambda = \bar{x}.$$

Thus the MLE for λ is the sample mean $\bar{x}$.
The variance of estimate is given by

$$\frac{1}{V(\hat{\lambda})} = E\left\{\frac{\partial^2}{\partial \lambda^2}(\log L)\right\}$$

$$\therefore \quad V(\hat{\lambda}) = \frac{\lambda}{n}.$$

(b) For the poisson distribution with parameter λ, we have

$$\frac{\partial}{\partial \lambda}\log L = -n + \frac{n\bar{x}}{\lambda} = n\left(\frac{\bar{x}}{\lambda} - 1\right) = \psi(\bar{x}, \lambda),$$

a function of $\bar{x}$ and λ only.
Hence $\bar{x}$ is sufficient for estimating λ.

Example : 3

State the as precisely as possible the properties of the MLE. Obtain the MLEs of α and β for a random sample from the exponential population.

$$f(x;\alpha,\beta) = y_0 e^{-\beta(x-\alpha)}, \quad \alpha \le x \le \infty, \beta > 0$$

and y_0 being a constant.

Solution:-

Here first of all we shall determine the constant y_0 from the consideration that the total area under a probability curve is unity.

$$\therefore y_0 \int_\alpha^\beta \exp[-\beta(x-\alpha)]\,dx \Rightarrow y_0 \left.\frac{e^{-\beta(x-\alpha)}}{-\beta}\right|_\alpha^\beta = 1$$

$$\Rightarrow \frac{y_0}{\beta}(0-1) = 1 \Rightarrow y_0 = \beta$$

$$f(x;\alpha,\beta)=\beta e^{-\beta(x-\alpha)}, \quad \alpha \le x \le \infty$$ (7)

If $x_1, x_2, \dots, x_n$ is a random sample of n observations from this population, then

$$L=\prod_{i=1}^{n} f(x_i;\alpha,\beta)=\beta^n \exp\left\{-\beta\sum_{i=1}^{n}(x_i-\alpha)\right\}$$

$$=\beta^n \exp[-n\beta(\bar{x}-\alpha)]$$

$\therefore \log L = n\log\beta - n\beta(\bar{x}-\alpha)$

The likelihood equations for estimating α and β gives $\frac{\partial}{\partial\alpha}\log L=0=n\beta$ and $\frac{\partial}{\partial\beta}\log L=0=\frac{n}{\beta}-n(\bar{x}-\alpha)$

Thus the Likelihood equations fail to gives as valid estimates of α and β and we try to locate MLE's for α and β by Maximising L directly.

L is maximum $\Rightarrow \log L$ is maximum

If $\beta(\bar{x}-\alpha)$ is minimum, which is so if α is maximum.

Consequently equations gives $1/\beta=\bar{x}-\hat{\alpha}=\bar{x}-x_{(1)}$

$$\Rightarrow \hat{\beta}=\frac{1}{\bar{x}-x_{(1)}}$$

Hence MLE's for α and β are given by

$$\hat{\alpha}=x_{(1)} \text{ and } \hat{\beta}=\frac{1}{\bar{x}-x_{(1)}}$$

Method of Moments :-

This Method was discovered and studied in detail by Karl Pearson.

Let $f(x; \theta_1, \theta_2, \cdots, \theta_k)$ be the density function of the parent population with k parameters $\theta_1, \theta_2, \cdots \theta_k$. If μ_r' denotes the r^{th} moment about origin, then

$$\mu_r' = \int_{-\infty}^{\infty} x^r f(x; \theta_1, \theta_2, \cdots \theta_k)\,dx \qquad (r = 1, 2, \cdots k)$$

In general $\mu_1', \mu_2', \cdots \mu_k'$ will be function of the parameters $\theta_1, \theta_2, \cdots \theta_k$.

Let x_i, $i = 1, 2, \ldots, n$ be a random sample of size n from the given population the methods of Moments consists in solving the k-equation for $\theta_1, \theta_2, \cdots \theta_k$ in terms of $\mu_1', \mu_2', \cdots \mu_k'$ and then replacing these moments μ_r' : $r = 1, 2 \ldots k$ by the sample moments

e.g: $\hat{\theta}_i = \theta_i(\hat{\mu}_1', \hat{\mu}_2' \ldots \hat{\mu}_k') = \theta_1(m_1', m_2' \cdots m_k')$ $\quad i = 1, 2 \ldots k$

Were m_i' the i th moment origin in the sample.

Then by the method of moments $\hat{\theta}_1, \hat{\theta}_2, \cdots \hat{\theta}_k$ are the required estimators of the $\theta_1, \theta_2, \cdots \theta_k$ respectively.

RAO-BLACKWELL THEOREM :-

Let X and Y be r.v. such that $E(Y) = \mu$ and $Var(Y) = \sigma_Y^2 > 0$

Let $E(Y/X=x) = \phi(x)$. Then (i) $E[\phi(X)] = \mu$ and

(ii) $Var[\phi(X)] \leq Var(Y)$

Proof :-

Let $f_{XY}(x,y)$ be the joint p.d.f of r.v. X and Y, $f_1(\cdot)$ and $f_2(\cdot)$ for given $X=x$ such that $h(y/x) = f(x,y)/f_1(x)$

$$E(Y/X=x) = \int_{-\infty}^{\infty} y \cdot h(y/x)\,dy = \int_{-\infty}^{\infty} y \cdot \frac{f(x,y)}{f_1(x)}\,dy$$

$$= \frac{1}{f_1(x)} \int_{-\infty}^{\infty} y f(x,y)\,dy = \phi(x), \text{ say}$$

$$\Rightarrow \int_{-\infty}^{\infty} y f(x,y)\,dy = \phi(x) \cdot f_1(x).$$

We observe that the conditional distribution of Y given $X=x$ does not depend on the parameter μ. Hence X is sufficient statistic for μ. Also

$$E[\phi(x)] = E[E(Y/X)] = E(Y) = \mu.$$

Now $var(Y) = E[Y - E(Y)]^2 = E(Y-\mu)^2$

$$= E(Y - \phi(x) + \phi(x) - \mu)^2$$

$$= E[Y-\phi(x)]^2 + E\{[\phi(x)-\mu)^2 + 2 \cdot E[Y-\phi(x) \cdot \phi(x)-\mu)]]\}$$

The product term gives

$$E\left[\left(y-\phi(x)\right)\left(\phi(x)-\mu\right)\right] = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \{y-\phi(x)\}\{\phi(x)-\mu\}\cdot f(x,y)\,dx\,dy$$

$$= \int_{-\infty}^{\infty} \{y-\phi(x)\}\, h(y/x)\, dy = 0$$

$$\therefore\ E\{y-\phi(x)\}\{\phi(x)-\mu\} = 0$$

$$Var(Y) = E(Y-\phi(x))^2 + Var\{\phi(x)\}$$

$$Var\ Y \geq var(\phi(x))$$

$$Var(\phi(x)) \leq var(Y).$$

Hence the proof.

Confidence Interval and confidence Limits :-

Let x_i, $(i=1,2,\dots,n)$ be a random sample of n observations from a population involving a single unknown parameter θ, (say). Let $f(x,\theta)$ be the probability function of the parent distribution from which the sample is drawn and let us suppose that this distribution is continuous. Let $t = t(x_1, x_2, \dots, x_n)$ a function of the sample value of an estimate of the population parameter θ. With the sampling distribution given by $g(t,\theta)$.

The technique of confidence interval due to Neyman and is obtain below. we choose once for all some small value of a (5% or 1%) and then determine two constants c_1 and c_2 such that.

$$P(c_1 < \theta < c_2 \mid t) = 1-\alpha$$

The quantities c_1 and c_2, so determind are known as the confidence limit (or) fiducial limits and the interval (c_1, c_2) within which the unknown value of the population parameter is expected to lie, is called confidence coefficient

Example:-

If we take a large sample from a normal population with mean μ and standard deviation σ then.

(a) The statistic $t = \frac{\bar{x}-\theta}{s/\sqrt{n}}$ follows student's t-distribution with $(n-1)$ degrees of freedom. Hence $100(1-\alpha)\%$ confidence limits for θ are given by.

$$p(|t| \leq t_\alpha) = 1-\alpha \Rightarrow p\left[|\bar{x}-\theta| \leq \frac{s}{\sqrt{n}} t_\alpha\right] = 1-\alpha$$

$$\therefore p\left(\bar{x} - t_\alpha \cdot \frac{s}{\sqrt{n}} \leq \theta \leq \bar{x} + t_\alpha \cdot \frac{s}{\sqrt{n}}\right) = 1-\alpha$$

Where t_α is the tabulated value of t for $(n-1)$ d.f at significance level 'α'. Hence the required confidence interval for θ is.

$$\left[\bar{x} - t_\alpha \frac{s}{\sqrt{n}}, \quad \bar{x} + t_\alpha \frac{s}{\sqrt{n}}\right]$$

(b) Case (i) θ is known and equal to μ (say)

Then $\frac{\Sigma(x_i-\mu)^2}{\sigma} = \frac{ns^2}{\sigma^2} \sim \chi^2_{(n)}$.

If we define χ^2_α as the value of χ^2 such that

$$P(\chi^2 > \chi^2_\alpha) = \int_{\chi^2_\alpha}^{\infty} p(\chi^2) d\chi^2 = \alpha$$

Where $p(\chi^2)$ is the p.d.f of χ^2 distribution with n d.f then the required confidence interval is given by $P\left\{\chi^2_{(1-\alpha/2)} \leq \chi^2 \leq \chi^2_{\alpha/2}\right\} = 1-\alpha$

Hence equation gives $P\left\{\frac{ns^2}{\chi^2_{\alpha/2}} \leq \sigma^2 \leq \frac{ns^2}{\chi^2_{(1-\alpha/2)}}\right\} = 1-\alpha$.

Where $\chi^2_{\alpha/2}$ and $\chi^2_{1-(\alpha/2)}$ are obtained using p.d.f

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

and $P(-1.96 \le Z \le 1.96) = 0.95.$

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

$$\Rightarrow P\left[\bar{x} - 1.96\, \sigma/\sqrt{n} \le \mu \le \bar{x} + 1.96 \frac{\sigma}{\sqrt{n}}\right] = 0.96$$

thus $\bar{x} \pm 1.96\, \sigma/\sqrt{n}$ are 95% confidence limits for the unknown parameter(μ) then the population mean and the interval.

$\bar{x} - 1.96\, \sigma/\sqrt{n}$, $\bar{x} + 1.96\, \sigma/\sqrt{n}$ is called the 95% confidence interval.

Hence 99% confidence limits for μ are $\bar{x} \pm 2.58\, \sigma/\sqrt{n}$ and 99% confidence interval for μ is $\left(\bar{x} - 2.58 \frac{\sigma}{\sqrt{n}} ; \bar{x} + 2.58 \frac{\sigma}{\sqrt{n}}\right)$

Example: 1

Obtain $100(1-\alpha)\%$ confidence interval for the parameter (a) θ and (b) σ^2, of the normal distribution

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

Solution:-

Let X_i $(i=1,2,\ldots,n)$ be a random sample of size n from the density $f(x;\theta,\sigma)$ and

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

Thus eq. 95% confidence interval. In this case the statistic

$$P\left[\frac{ns^2}{\chi^2_{0.05}} \le \sigma^2 \le \frac{ns^2}{\chi^2_{0.975}}\right] = 0.95.$$

Case (ii) θ is unknown

$$\frac{\Sigma(x_i-\bar{x})^2}{\sigma^2} = \frac{ns^2}{\sigma^2} \sim \chi^2_{(n-1)} \text{ d.f}$$

Test of significance for single proportion:-

If x is the number of successes in n independent the trials with constant probability P of success for each trial.

$E(x) = nP$ and $Var(x) = nPQ$ and where $Q = 1-P$, is the probability of failure

It has been prove that for large n, the binomial distribution tends to normal distribution

Hence $$Z = \frac{x - E(x)}{\sqrt{Var(x)}} = \frac{x-nP}{\sqrt{nPQ}} \sim N(0,1).$$

Hence the proof.

Example: 1

A random sample of 500 apples was taken from a large consignment and 60 were found to be bad. obtain the 98% confidence limits for the percentange of bad apples in the consignment.

Solution:-

We have p = proportion of bad apples in the sample. $= \frac{60}{500} = 0.12$

Since significant value of z at 98% confidence coefficient is 2.33, 98% confidence limits for population proportion.

$$p \pm 2.33\sqrt{\frac{pq}{n}} = 0.12 \pm 2.33\sqrt{\frac{0.12 \times 0.88}{500}}$$

$$= 0.12 \pm 2.33 \times 0.01453$$

$$(0.08615,\ 0.15385).$$

Test of Significance for Difference Proportions:-

Let X_1, X_2 be the number of persons possessing the given attribut. A in random Sample of Size n_1 and n_2 from the two proportions respectively

If P_1 and P_2 are population proportions, then

$E(p_1) = P_1$, $E(p_2) = P_2$ and $V(P_1) = \frac{P_1Q_1}{n_1}$ and

$V(P_2) = \frac{P_2Q_2}{n_2}$; Since for large samples. p_1 and p_2 Sample are independently and asymptotically normally distributed.

$$z = \frac{(p_1 - p_2) - E(p_1 - p_2)}{\sqrt{V(p_1 - p_2)}} \sim N(0,1)$$

Under the null hypothesis $H_0: P_1 - P_2$ ie., there is no significan difference between the sample problem-

Hence under $H_0: P_1 - P_2$ the test statistic for the difference proportion are independent.

$$z = \frac{P_1 - P_2}{\sqrt{PQ\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}} \sim N(0,1).$$

Under $H_0: p_1 - p_2 = p$ an unbiased estimate of the population proportion.

$$\hat{p} = \frac{n_1 p_1 + n_2 p_2}{n_1 + n_2} = \frac{x_1 + x_2}{n_1 + n_2}$$

The estimate is unbiased. Since

$$E(\hat{p}) = \frac{1}{n_1 + n_2} E(n_1 p_1 - n_2 p_2)$$
$$= \frac{1}{n_1 + n_2} n_1 p_1 + n_2 p_2 = p$$

Test of Significance for Single mean:-

We have proved that if $x_i, (i = 1, 2, \cdots n)$ is a random sample of size n from a normal population with mean μ and variance σ^2, then the sample mean is distributed normally with mean μ and variance σ^2/n, ie $\bar{x} \sim N(\mu, \sigma^2/n)$.

The standard normal variate corresponding to $\bar{x}$ is

$$z = \frac{\bar{x} - \mu}{\sigma / \sqrt{n}}$$

Under the null hypothesis H_0: the sample ~~has~~ has been drawn from a population with mean μ and variance σ^2

i.e., There is no significant different between the sample mean ($\bar{x}$) and population Mean (μ).

The test statistic $z = \dfrac{\bar{x} - \mu}{\sigma / \sqrt{n}} \sim N(0, 1)$

Example :-

A Sample of 900 members has a mean 3.4 Cms. and S.D 261 Cms. Is the sample from a large population of mean 3.25 Cms and S.D 2.61 Cms.

If the population is normal and its mean is unknown, find the 95% and 98% fiductual with limits of true mean.

Solution :-

Null Hypothesis (H_0):

The Sample has been drawn from the Population with mean $\mu = 3.25$ Cms. and S.D $\sigma = 261$ cms

H_1: $\mu \neq 3.25$ (Two tailed)

Test Statistic $z = \dfrac{\bar{x} - \mu}{\sigma/\sqrt{n}} \sim N(0,1)$

$$z = \frac{3.40 - 3.25}{2.61/\sqrt{900}} = 1.73.$$

Since $|z| < 1.96$, we conclude that the data don't Provide be accepted at 5% Level of Significance.

95% fiducial Limits for the population mean μ are

$$\bar{x} \pm 1.96\left(\sigma/\sqrt{n}\right) = 3.40 \pm 1.96\left(2.61/\sqrt{900}\right)$$
$$= 3.40 \pm 0.1705$$

98% fiducial limits for μ are given by

$$\bar{x} \pm 2.33\frac{\sigma}{\sqrt{n}} = 3.40 \pm 2.33 \times \frac{2.61}{30}$$
$$= 3.40 \pm 0.2027$$

Test of significance for difference of means:-

Let $\bar{x}_1$ be the mean of a sample of size n_1 from a population with Mean μ_1 and Variance σ_1^2 and Let $\bar{x}_2$ be the mean of an independent random sample of size n_2 from another population with mean μ_2 and variance σ_2^2. Then since Sample sizes are large

$$\bar{x}_1 \sim N(\mu_1, \sigma_1^2/n_1) \text{ and } \bar{x}_2 \sim N(\mu_2, \sigma_2^2/n_2)$$

$\bar{x}_1 - \bar{x}_2$ being the difference of two independent normal variates is also a normal Variate.

$$z = \frac{(\bar{x}_1 - \bar{x}_2) - E(\bar{x}_1 - \bar{x}_2)}{S.E(\bar{x}_1 - \bar{x}_2)} \sim N(0,1).$$

Under the Null Hypothesis $H_0: \mu_1 = \mu_2$ ie., there is no significant difference between the sample means, we get.

Under $H_1: \mu_1 \neq \mu_2$ the test statistic

$$z = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{(\sigma_1^2/n_1) + (\sigma_2^2/n_2)}} \sim N(0,1)$$

Example:-

In a certain factory there are two independent processes manufacturing the same item. The average weight in a sample of 250 items produced from one process is found to be 120 ozs. with a standard deviation of 12 ozs. while the corresponding figures in a sample of 400 items from the other process are 124 and 14.

obtain the Standard error of difference between the two sample means. Is this difference significant? Also find the 99% confidence limits for the difference in the average of items produced by the two processes respectively.

Solution :-

Null hypothesis $H_0 : \mu_1 = \mu_2$ ie., the Sample means do not differ significantly.

Alternative Hypothesis. $H_1 : \mu_1 \neq \mu_2$

$$z = \frac{\bar{x}_1 - \bar{x}_2}{S.E(\bar{x}_1 - \bar{x}_2)} = \frac{120-124}{1.034} = 3.87.$$

$\therefore z = 3.87$

Conclusion :-

Since $|z| > 3$, The null Hypothesis is rejected and we conclude that there is significant difference between the sample means.

$$|\bar{x}_1 - \bar{x}_2| \pm 2.58 \; S.E(\bar{x}_1 - \bar{x}_2)$$

$$4 \pm 2.58 \times 1.034 = 4 \pm 2.67$$

$$= 6.67 \text{ and } 1.33.$$

$$1.33 < |\mu_1 - \mu_2| < 6.67.$$

UNIT – V

t-test for single mean:-

Suppose we want to test

(i) If a random sample x_i $(i=1,2,\ldots n)$ of size n has been drawn from a normal population with a specified mean, say μ_0 (or)

(ii) If the sample mean differs significantly from the hypothetical value μ_0 of the population mean.

Under Null Hypothesis H_0:

(i) The sample has been drawn from the population with mean μ_0 (or)

(ii) There is no significant difference between the sample mean $\bar{x}$ and the population mean μ_0.

$$t = \frac{\bar{x}-\mu_0}{s/\sqrt{n}} \quad \text{where } \bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i$$

and $s^2 = \frac{1}{n-1}\sum_{i=1}^{n}(x_i-\bar{x})^2$

follows student's t-distribution with $(n-1)$ d.f.

We now compare the calculated value of t with the tabulated value at certain level of significance.

If calculated $|t| >$ tabulated t, reject the null Hypothesis. If calculated $|t| <$ tabulated t, H_0 accept the null Hypothesis.

Example :-

A random Sample of 10 boys had the following I.Q.'s: 70, 120, 110, 101, 88, 83, 95, 98, 107, 100. Do these data support the assumption of a population mean I.Q of 100?. Find a reasonable range in which most of the mean I.Q value of Samples of 10 boys lie.

Solution :-

Null Hypothesis H_0: The data are consistent with the assumption of a mean I.Q. of 100 in the Population ie., $\mu = 100$

Alternative Hypothesis H_1: $\mu \neq 100$

Test statistic $t = \dfrac{\bar{x} - \mu}{\sqrt{S^2/n}} \sim t_{(n-1)}$ d.f

Where $\bar{x}$ and s^2 are to be computed from the sample value of I.Q.'s

x	$(x - \bar{x})$	$(x - \bar{x})^2$
70	-27.2	739.84
120	22.8	519.84
110	12.8	163.84
101	3.8	14.44
88	-9.2	84.44
83	-14.2	201.64
95	-2.2	4.84
98	0.8	0.64
107	9.8	96.04
100	2.8	7.84
972		1833.60

Here $n = 10$, $\bar{x} = \frac{972}{10} = 97.2$, $S^2 = \frac{1833.60}{9} = 203.73$

$\therefore |t| = \frac{97.2 - 100}{\sqrt{203.73/10}} = 0.62$

Tabulated value for $(10-1)$ d.f ie., 9 d.f for two tailed test is 2.262

H_0 : accepted the null hypothesis. We may conclude that the data are consistent with the assumption of mean I.Q of 100 in the Population.

The 95% confidence limits within the mean I.Q. value of samples of 10 boys will lie are given by.

$$\bar{x} \pm t_{0.05} \, S/\sqrt{n} = 97.2 \pm 2.262 \times 4.514$$
$$= 97.2 \pm 10.21 = 107.41 \text{ and } 86.99$$

Hence the required 95% confidence interval is $(86.99, 107.41)$

t-test for difference of Means:-

Suppose we want to test if two independent samples $x_i \,(i = 1, 2, \ldots n_1)$ and $y_j \,(j = 1, 2, \ldots, n_2)$ of size n_1 and n_2 have been drawn from two normal Population with means μ_x and μ_y respectively.

Under the Null Hypothesis (H_0) that the samples have been drawn from the normal Population variance are equal ie $\sigma_x^2 = \sigma_y^2 = \sigma^2$

$$t = \frac{(\bar{x}-\bar{y})-(\mu_x-\mu_y)}{S\Big/\sqrt{\frac{1}{n_1}+\frac{1}{n_2}}} \sim t_{(n_1+n_2-2)}.$$

Where $\bar{x} = \frac{1}{n_1}\sum_{i=1}^{n_1} x_i$; $\bar{y} = \frac{1}{n_2}\sum_{j=1}^{n_2} y_i$ and

$$S^2 = \frac{1}{n_1+n_2-2}\left[\Sigma(x_i-\bar{x})^2 + \Sigma(y_j-\bar{y})^2\right]$$

is an unbiased estimate of the common Population Variance σ^2, follows Student's t-distribution with (n_1+n_2-2) d.f.

Example :-

Samples of two types of electric light bulbs were tested for Length of Life and following data were obtained.

	Type I	Type - II
Sample No.	$n_1 = 8$	$n_2 = 7$
Sample Mean	$\bar{x}_1 = 1,234$ hrs	$\bar{x}_2 = 1,036$ Hrs
Sample S.D	$S_1 = 36$ Hrs	$S_2 = 40$ Hrs.

Is the difference in the mean sufficient to warrant that type I is superior to type - II

Solution:-

Null Hypothesis $H_0: \mu_x = \mu_y$ ie., the two types of electric bulbs are identical.

Alternative Hypothesis $H_1: \mu_x \neq \mu_y$ or $\mu_x > \mu_y$

The test statistic $t = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{s^2\left(\frac{1}{n_1}+\frac{1}{n_2}\right)}} \sim t_{(n_1+n_2-2)}$ d.f

where $s^2 = \frac{1}{n_1+n_2-2}\left[\Sigma(x_1-\bar{x}_1)^2 + \Sigma(x_2-\bar{x}_2)^2\right]$

$$t = \frac{1234-1036}{\sqrt{1659.08\left(\frac{1}{8}+\frac{1}{7}\right)}} = 9.36.$$

Conclusion

Since calculated 't' is much greater than tabulated 't' it is highly significant and Ho is rejected. Hence the two types of electric bulbs differ significantly.

F-test for Equality of two population Variances:-

We want to test (i) whether two independent samples $x_i (i=1,2,\dots,n_1)$ and $y_j (j=1,2,\dots n_2)$ have been drawn from the normal populations with the same variance σ^2 (or) (ii) whether the two independent estimates of the population variance are homogeneous or not.

Under Null Hypothesis (Ho) (i) $\sigma_x^2 = \sigma_y^2 = \sigma^2$

ie., the population variances are equal, (or)
(ii) Two independent estimates of the population variance are homogeneous. the statistic F is given by

$$F = \frac{S_x^2}{S_y^2} \text{ where } S_x^2 = \frac{1}{n_1-1}\sum_{i=1}^{n_1}(x_i-\bar{x})^2 \text{ and}$$

$S_y^2 = \frac{1}{n_2-1}\Sigma(y_i-\bar{y})^2$

The common population variance σ^2 obtained from two independent samples and it follows Snedecor's F-distribution with (v_1, v_2) d.f where $v_1 = n_1 - 1$ and $v_2 = n_2 - 1$

Proof:-

$$F = \frac{S_x^2}{S_y^2} = \frac{n_1}{n_1 - 1} s_x^2 \Big/ \frac{n_2}{n_2 - 1} s_y^2$$

Since $\Rightarrow \dfrac{n_1 s_x^2}{\sigma_x^2}$ and $\dfrac{n_2 s_y^2}{\sigma_y^2}$ are independent chi-square variance with $(n_1 - 1)$ and $(n_2 - 1)$ d.f respectively.

Example:-

In one sample of 8 observations, the sum of the square of deviations of the sample values from the sample mean was 84.4 and in the other sample of 10 observation it was 102.6. Test whether this difference is significant at 5% level.

Solution:-

Here $n_1 = 8$, $n_2 = 10$ and $\Sigma(x - \bar{x})^2 = 84.4$,

$$\Sigma(y - \bar{y})^2 = 102.6$$

$$S_x^2 = \frac{1}{n_1 - 1} \Sigma(x - \bar{x})^2 = \frac{84.4}{7} = 12.057$$

$$S_y^2 = \frac{1}{n_2 - 1} \Sigma(y - \bar{y})^2 = \frac{102.6}{9} = 11.4$$

Under $H_0 : \sigma_x^2 = \sigma_y^2 = \sigma^2$ ie., the estimates of σ^2 is given by the samples are homogeneous.

$$F = \frac{S_x^2}{S_y^2} = \frac{12.057}{11.4} = 1.057.$$

Tabulated $F_{0.05}$ for $(7, 9)$ d.f is 3.29

H_0: may be accepted at 5% level

Inference about a Population Variance in chi-Square distribution :-

We want to test if a random sample x_i $(i = 1, 2, \dots, n)$ has been drawn from a Normal Population with a specified Variances $\sigma^2 = \sigma_0^2$

Under the Null Hypothesis that the population Variances is $\sigma^2 = \sigma_0^2$.

test statistic $$\chi^2 = \sum_{i=1}^{n}\left[\frac{(x_i - \bar{x})^2}{\sigma_0^2}\right]$$

$$= \frac{1}{\sigma_0^2}\left[\sum_{i=1}^{n} x_i^2 - \frac{(\Sigma x_i)^2}{n}\right]$$

$$= \frac{ns^2}{\sigma_0^2}$$ follows chi-square distribution with $(n-1)$ d.f