Unit-III

Methods OF estimation :-

So For we have been discussing requisites of a good estimator now we shall briefly outline some of importants methods for obtaining such estimators.

commonly used methods are

i) Method OF maximum Likihood estimation.

ii) Method of minimum variance

iii) method of moments

iv) method of inverse probability

v) method of minimum chi-square.

Maximum Likelihood estimations :

The method of maximum Likilihood estimation is from theoritical. point of view, the most general method of estimation known is the method of maximum Likihood estimators (MLE) which was initially Formulated by CF Grauss but as a general method OF estimation was First introduced by prof R.A Fisher and Later on developed by him in a series of papers before introducing the method of we will First define Likihood function.

Likelihood Function:

Definition:

Let $x_1, x_2, \cdots x_n$ be a random sample of size n From a population with density Function of the Sample values $x_1, x_2, \cdots x_n$ usually denoted by $L = L(\theta)$ is their joint density Function, given by

$$L = F(x_i, \theta)\; F(x_2, \theta), \cdots F(x_n, \theta) = \prod_{i=1}^{n} F(x_i, \theta) \to ①$$

L gives the relative Likelihood that the random variables assume a particular set of values, $x_1, x_2, \cdots x_n$ fora given Sample $x_1, x_2, \cdots x_n$ L become a function of the variable θ, the parameter.

$$l = F(x_i, \theta)\; f(x_2, \theta), \cdots F(x_n, \theta) = \prod_{i=1}^{n} F(x_i, \theta) \to ①$$

It is allso known as a Likihood function.

The principle of maximum Likelihood consists in finding an estimator for the unknown parameter. $\theta = (\theta_1, \theta_2, \cdots \theta_t)$ Say, which maximises the Likilihood Function $L(\theta)$ for variance in parameter (i.e) we wish to find $\hat{\theta} = (\hat{\theta}, \hat{\theta}, \cdots \hat{\theta}_k)$ so that

$$L(\hat{\theta}) > L(\theta) \ \forall \theta \in \Theta \text{ (i.e)}$$

$$L(\hat{\theta}) = \sup L(\theta) \sqrt{\ } \theta \in \Theta$$

This is there exists a function $\hat{\theta} = \hat{\theta}(x_1, x_2, \cdots x_n)$ of the Sample values which maximises L for variations in θ, then $\hat{\theta}$ is to be takes as an estimator of θ, $\hat{\theta}$ is usually called maximum Likilihood estimation (M.L.E)

Thus $\hat{\theta}$ is the solution if any of

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

Since $L > 0$, and $\log L$ is a non-decreasing function of L and $\log L$ attain their extreme values (maxima or minima) of the Same value of $\hat{\theta}$; The first of the two equations can be

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

$$\theta = \frac{\partial \log L}{\partial \theta} = 0$$

a from which is much more convinient from particular point of view.

If θ is vector valued paramitor then $\hat{\theta} = (\hat{\theta_1}, \hat{\theta_2}, \dots \hat{\theta}.L)$ is given by the solution of simultaneous equation.

$$\partial/_{\partial\theta}{}_i \log L = \partial/_{\partial\theta}{}_i \log L(\theta_1, \theta_2; \theta k) = 0 \quad i = 1, 2, \dots k \rightarrow ①$$

The above equation ③ and ④ are usually referred to as the Likelihood equations for estimating the parametors.

Proporties of maximum Likelihood :

we make the following assumptions known as the Regularity conditions.

From theoritical point of view, the most general method of estimation known is the method of maximum Likilihood Estimators (MLE) which was initially formulated by C.F.Gauss

i) The first and second order derivatives.

viz ; $\frac{\partial \operatorname{Log} L}{\partial \theta}$ and $\frac{\partial^2 \operatorname{Log} L}{\partial \theta}$ exist,

and are continuous function of θ in a range R

including. the true value θ_0 of the parameter for almost all x.

For every θ in R, $\left| \partial/\partial\theta \operatorname{Log} L \right| < F(x)$ and

$\left| \frac{\partial^2}{\partial \theta^2} \operatorname{Log} L \right| < F_2(x)$ where $F_1(x)$ and $F_2(x)$ are integrable

Functions over $(-\infty, \infty)$

ii) The third order derivative $\frac{\theta^3}{\theta\theta^3} \operatorname{Log} L$

exists such that $\left| \frac{\partial^3}{\partial \theta^3} \operatorname{Log} L \right| < M(x)$.

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 depends on θ,

then $f(x,\theta)$ Vanishes at the extroms depending on θ [This assumption is two make the difforonce between under the integral Sign is Valid.

under the above assumption MLE possess the number of important properties. which will be stated in the form of the theorem.

Method of moments :-

This method was discovered and studied in detail by Karl pearson

Let $f(x;\theta_1, \theta_2, \dots \theta_k)$ be the density function of the parent population with k parameter $\theta_1, \theta_2, \dots \theta_k$. If μ_r denotes rth Moment about origin. Then

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

$r = 1, 2, \dots k$

In genaral $\mu r', \mu_2, \cdots \mu_K'$ will be function of the parametors $\theta_1, \theta_2, \cdots \theta_K$

Let x_i $i=1,2,\cdots n$ be a random sample of size n from the given population. The method of moments consist is solving the k-equation for $\theta_1, \theta_2, \cdots \theta_n$ in terms of $\mu_1, \mu_2', \cdots \mu_K'$ and

Then roplacing those moments, $\mu r'_i$ $r=1,2,\cdots k$ by the sample moments.

oxample :-

$$\hat{\theta}_i = \theta_i(\hat{\mu}_1', \hat{\mu}_2', \cdots \mu_K') = \theta_i(\mu_1' \mu_2' \cdots \mu_K')$$

$$i = 1,2,\cdots k$$

where μ_r' is the ith moment about origin in the sample.

Then by the method of moments $\hat{\theta}_1, \hat{\theta}_2, \cdots \hat{\theta}_K$ are the roquired astimators of $\theta_1, \theta_2, \cdots \theta_K$ rospectivoly

1. a) find the maximum Likilihood estimate for the parameter λ of a poisson distribution on the basis of a sample size n allso find its variance.

b) show that the sample mean $\bar{x}$, is sufficiont for estimating the paramotor λ of the poisson distribution.

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

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

Likilihood Function of random Sample $x_1, x_2, \dots x_n$ of n vibrations from this population is,

$$l = \prod_{i=1}^{n} f(x_i,\lambda) = \frac{e^{-n\lambda}\,\lambda^{\sum_{i=1}^{n} x_i}}{x_i!\,x_2!\dots x_n!}$$

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

The Likelihood equation for estimating λ is $\partial/\partial\lambda$

$$\log L = 0 \Rightarrow -n + n\bar{x}/\lambda = 0$$

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

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

$$= E\left\{\frac{-\partial}{\partial\lambda}(-n + n\bar{x}/\lambda)\right\}$$

$$= E\left\{-\left(-n\bar{x}/\lambda^2\right)\right\} = n/\lambda^2 \; E(\bar{x})$$

$$= n/\lambda \; \left[\because E(\bar{x}) = n/\lambda\right]$$

$$v(\hat{\lambda}) = \lambda/n$$

b) For the poisson distribution with parameter λ we have

$$\frac{\partial}{\partial\lambda} \log L = -n + n\bar{x}/\lambda$$

$$= n(\bar{x}/\lambda^{-1})$$

$$= \phi(\bar{x}, \lambda)$$

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

3. state as precisely as possible the properties of the MLE obtain the MLE's OF α and β For a random sample from the exponential population.

$$f(x: \alpha, \beta) = y_0 e^{-\beta(x - \infty)}, \quad a \leq x < \infty$$

$$\beta > 0$$

and you being a constant.

The probability curve is unity

$$\therefore y_0 \int_{\alpha}^{\infty} \exp\left[-\beta(x-\alpha)\right] dx$$

$$y_0 \left| \frac{e^{-\beta(x-\infty)}}{-\beta} \right|_{\alpha}^{\infty} = 1$$

$-y_0/_\beta \; (0-1) = 1 \qquad y_0 = \beta$

$F(x; \alpha, \beta) = \beta\, e^{-\beta(x-\alpha)} \qquad \alpha \leq x < \infty$

if $x_1, x_2, \ldots x_n$ is a random Sample of n observation from this population.

$L = \prod_{i=1}^{n} F(x_i; \alpha, \beta) = \beta \exp\left\{-\beta \sum_{i=1}^{n} (x_i - \alpha)\right\} \Rightarrow \beta^n \exp(-n\beta \; (\bar{x} - \alpha)\,]$

$\log L = n \log B - n\beta \; (\bar{x} - \alpha) \rightarrow ①$

The equation α and β give,

$\frac{\partial}{\partial \alpha} \log L = 0 = n\beta \rightarrow ②$ and

$\frac{\partial}{\partial \beta} \log L = 0 = n/_\beta - n(\bar{x} - \alpha) \rightarrow ③$

L is maximum $\Rightarrow$ Log L is maximum from① Log L is maximum.

i.e $\hat{\alpha} = x(1)$ consequently ③ gives $1/_\beta = \bar{x} - \hat{\alpha} = \bar{x} - x(1)$

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

Hence, MLE's for α and β are given by,

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

Unit - IV

Interval ostimation :

interval ostimation is the uso of sample data to calculate an interval of possible values of an unknown population parameter; this is in contrast to point ostimation, which gives a single value. Confidence interval and crodible intervals

The Forms of interval ostimation :

* confidenco interval [a frequentist method]
* crediblo intervals [a Bayesian method]
* likelihood method intervals (a Likilihoodist method)
* Fiducial interval (a fiducial method

Other forms of statistical Intervals, which do not ostimate parameters, include :

* tolerance intervals (an ostimates of population
* prediction intervals [used mainly in regression analysis)

Some other types of outcome are point estimate and decisions.

Confidence Level

The probability that the value of parameter falls with in a specified range of values

definition:-

A confidence level refers to the percentage of all possible samples that can be expected to include the true population parameter

For example:-

Suppose all possible samples were selected from the same population, and a confidence interval were computed for each sample.

a 95% confidence level implies that 95% of the confidence intervals would include the true population parameter. $CI = \bar{x} \pm z \frac{s}{\sqrt{n}}$

Confidence co-efficient :-

The confidence co-efficient is the confidence level stated as a proportion rather than as a percentage.

eg:-

if you had a confidence level of 99%, the confidence co-efficient would be .99. In general, the higher the coefficient, the more certain you are that your results are accurate.

Confidenu interval:

In statistics, a confidence Interval is a type of estimate computed from the statistics of the obsorad data. This proposes a range of palausible values for an unknown parameter. The interval has an associated confidence Level that the true parameter is in the proposed range.

$$CI = \bar{x} \pm z \frac{s}{\sqrt{n}}$$

properties (or) characteristics of confidence interval:

The confidence interval (CI) is a range of values. It is expressed as a percentage and is expected to contain the best estimate of a statistical parameter.

A confidence interval of 95% mean, it is 95% certain that our population parameter lies in between this confidence interval provides a range of plausible values for a parameter.

* developed from a sample
* for a given confidence Level

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. 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 be an estimate of the population parameter θ.

the sampling distribution given by $g(t,\theta)$ we choose once for all some value of

α (5% or 1%) and then

determined two constans say c_1 and c_2 such that

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

The quantities c_1 and c_2. so determined are known as the confidence Limits or fiducial limits and the interval $[c_1, c_2]$ with in which the unknown value of the population parameter is expected to Lie, is called the confidence interval and $(1-\alpha)$ confidence co-efficient.

Thus if we take $\alpha = 0.05$ or 0.01 we shall get 95% or 99%. confidence limits.

How to Find c_1 and c_2 let T_1 and T_2 be two statistics Such that,

$P(T_1 > \theta) = \alpha_1$ --> ①

and $P(T_2 < \theta) = \alpha_2$ --> ②

where α_1 and α_2 are constants independent of θ ① and ② can be combined to give.

$P(T_1 < \theta < T_2) = 1-\alpha$

where $\alpha = \alpha_1 + \alpha_2$. Statistics T_1 and T_2 defined in ① and ② may be taken as c_1 and c_2 defined in for example, if we take a Large sample from a normal population with mean μ and standard deviation σ, then $z = \frac{\bar{x}-\mu}{\sigma/\sqrt{n}} \sim N(0,1)$

confidence interval

$P[-z_\alpha \le z \le z_\alpha] = 0.95$

and $P[-1.96 \le z \le 1.96] = 0.95$ (from Normal probability tables)

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

Thus,

$\therefore \left(x \pm 1.96\frac{\sigma}{\sqrt{n}}\right)$ are 95% confidence limits for the unknown parameter μ, 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

Also

$P(-2.58 \leq z \leq 2.58) = 0.99$ or

$P(-2.58 \leq \frac{\bar{x}-\mu}{\sigma/\sqrt{n}} \leq 2.58) = 0.99$

$P(\bar{x} - 2.58 \frac{\sigma}{\sqrt{n}} \leq \mu \leq \bar{x} + 2.58\, \sigma/\sqrt{n}) = 0.99$

Hence 99% confidence limits for μ are ;

$\bar{x} \pm 2.58\, \sigma/\sqrt{n}$ and 99%

confidence interval for μ is

$(\bar{x} - 2.58\, \sigma/\sqrt{n}, \bar{x} + 2.58\, \sigma/\sqrt{n})$

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 that is $\bar{x}$ follows normal $\mu, \sigma^2/n$

that is $\bar{x} \sim N(\mu, \sigma^2/n)$ even in random sampling from non normal population provided the sample size n is large (c.f) thus for large samples the S.D normal variate is corresponding to $\bar{x}$ is $z = \frac{\bar{x}-\mu}{\sigma/\sqrt{n}}$

under the null hypothesis Ho. that the sample has been drawn From the population with mean μ and variance σ^2 that is no significant difference and the sample mean $(\bar{x})$ and population mean (μ) The test statistics for Large Sample is $z = \frac{\bar{x}-\mu}{\sigma}, \sim N(0,1)$

Remarks:

i) if the population s.D σ is unkown then we use in estimate provide by the sample variance given by,

$\hat{\sigma^2} = s^2, \hat{\sigma} = s$ (for Large Samples)

ii) Confidence limits for μ 1.95%. Confidence interval for μ is given by.

$|z| \leq 1.96$, that $\left| \frac{\bar{x}-\mu}{\sigma/\sqrt{n}} \right| \leq 1.96$ which

(i.e)

$\bar{x} - 1.96 (\sigma/\sqrt{n}) \leq \mu \leq \bar{x} + 1.96 (\sigma/\sqrt{n}$

iii) The Confidence Limits for any parameter $(p, \mu ..)$ are allso Known as its Fiducial Limits problem

1. A sample of 900 membei has a mean 3·4 cm and s·D 2·61 cm is the sample From a Large population of mean 3·25 cm and s-D 2·61 cm if the population is normal and its mean is unknown, find the 95%. and 98%. Fiducial limits of true mean.

Null hypothesis :

$H_0 : \mu = 3 \cdot 25$

Alltornative hypothosis (H1) : $\mu \neq 3 \cdot 25$

(two tailod) tost statistics under H0 . the tost statistics is

Test Statistics :

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

$\bar{x} = 3 \cdot 4$ c.m $n = 900$ $\mu = 3 \cdot 25$ c.m $\sigma = 2 \cdot 61$

$$z = \frac{3 \cdot 4 - 3 \cdot 25}{\frac{2 \cdot 61}{\sqrt{900}}}$$

$$= 0 \cdot 15 \times 30 / 2 \cdot 61$$

$$|z| = 1 \cdot 73$$

Conclusion :

since $|z| < 1 \cdot 96$ we conclude that the data don't provide is any evidence against the null hypothesis (H0) which may.

Therefore be accepted at 5%. Level of significance 95%. Fiducial limits for the population

Test of significance difference of mean :

Let $\bar{x}$ be the mean of a sample of size n , from a population with mean μ , and variance σ^2, and Let $\bar{x_2}$ be the mean of and independent random sample of sizes n_2 from another population with μ_2 and variance σ_2^2. then since samplo sizes are Large,

$$\bar{x}_1 \sim N\left(\mu_1 \frac{\sigma^2}{n_1}\right) \text{ and } \bar{x_2} \sim N\left(\mu_2, \frac{\sigma_2^2}{n_2}\right)$$

also $\bar{x_1} - \bar{x_2}$, being the difference of two independent normal variates is allso a normal variate . The value of z (S.N.V) corresspondidng $\bar{x_1} - \bar{x_2}$ is given by,

$$z = \frac{(\bar{x_1} - \bar{x_2}) - E(\bar{x_1} - \bar{x_2})}{V(\bar{x_1} - \bar{x_2})} \sim N(0,1)$$

under the null hypothesis . $H_0 : \mu_1 = \mu_2$, that there is no significant difference between the sample means, we get

$$E(\bar{x_1} - \bar{x_2}) = E(\bar{x_1}) - E(\bar{x_2})$$

$$= \mu_1 - \mu_2 = 0$$

$$V(\bar{x_1} - \bar{x_2}) = V(\bar{x_1}) - V(\bar{x_2})$$

The co-variance term Vanishes, since the Sample means $\bar{x_1}$ and $\bar{x_2}$ are independent there under $H_0: \mu_1 = \mu_2$. The test statistics becomes (for Large Samples)

$$z = \frac{(\bar{x_1} - \bar{x_2})}{\sqrt{\sigma_1^2/n_1 + \frac{\sigma_2^2}{n_2}}} \sim N(0,1)$$

$$z = \frac{\bar{x_1} - \bar{x_2}}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}} \sim N(0,1)$$

Remarks:-

1. If $\sigma_1^2 = \sigma_2^2 = \sigma^2$ that if the samples have been drawn from the population with common S.D σ then under $H_0: \mu_1 = \mu_2$

$$z = \frac{\bar{x_1} - \bar{x_2}}{\sigma\sqrt{1/n_1 + 1/n_2}} \sim N(0,1)$$

2. If in $z = \frac{\bar{x_1} - \bar{x_2}}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}} \sim N(0,1)$. σ is not known it is based on Sample variance.

Unit-V

Students 't' distribution:

Let x_i $(i=1,2,...n)$ be a random sample of size n from a normal population with mean μ and variance σ^2 then students t is defined by the statistics.

$$\boxed{t = \frac{\bar{x}-\mu}{s/\sqrt{n}} \rightarrow \textcircled{1}}$$

where $\bar{x} = 1/n \sum_{i=1}^{n} x_i$ is the sample mean and $s^2 = \frac{1}{n-1} \sum_{i=1}^{n} (x_i-\bar{x})^2$ is an unbiased estimate of the population variance σ^2 and it follows student t distribution with $v=(n-1)$ d.f with probability density function.

$$f(t) = \frac{1}{\sqrt{v}B(1/2, v/2)} \cdot \frac{1}{\left(1+\frac{t^2}{v}\right)\left(\frac{v+1}{2}\right)} \quad -\infty < t < \infty \rightarrow \textcircled{2}$$

Derivation of Students t distribution:

The expression ① can be re written as

$$t^2 = \frac{n(\bar{x}-\mu)^2}{s^2}$$

$$= \frac{n(\bar{x}-\mu)^2}{ns^2/(n-1)}$$

$$\Rightarrow \frac{t^2}{(n-1)} = \frac{(\bar{x}-\mu)^2}{\sigma^2/n} \cdot \frac{1}{ns^2/\sigma^2}$$

$$= \frac{(\bar{x}-\mu)^2 \; (\sigma^2/n)}{ns^2/\sigma^2}$$

since $x_i (i = 1, 2, \ldots n)$ is a random sample, mean μ Variance σ^2

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

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

Hence, $\frac{(\bar{x}-\mu)^2}{\sigma^2/n}$ being the square

of a standard normal variate is a chi-square variate with 1 d.f

$\frac{ns^2}{\sigma^2}$ is a χ^2 variate with $(n-1)$ d.f

$$dF(t) = \frac{1}{B(1/2, V/2)} \cdot \frac{(t^2/v)^{1/2-1}}{\left(1+\frac{t^2}{v}\right)^{(v+1)/2}} \, d\left(t^2/v\right) \quad 0 < t^2 < \infty$$

$\therefore$ where $V = (n-1)$]

$$= \frac{1}{\sqrt{v}\,B\left(\frac{1}{2} \cdot \frac{1}{2}\right)} \cdot \frac{1}{\left(1+\frac{t^2}{v}\right)\left(\frac{v+1}{2}\right)} \, dt \quad -\infty < t < \infty$$

the factor 2 disappearing since the integral from $-\infty$ to ∞ must be unity.

Application of t distribution ∞

Than t distribution has a wide number of application in statistics some of which are annumeratad below,

i) To tast if the sample mean ($\bar{x}$) diffors, significantly from the hypothetical value μ of the population mean.

ii) To tast the significance of the difforonce betwoon two sample means.

iii) To tast the significance of an observed sample correlation co-officiant and sample ragrassion co-officiant.

In the following sactions we will discuss those application in detail one by one.

t - Test for single mean :

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

ii) if the sample mean diffors significantly from the hypothetical value μ_0.

under the Null hypothesis : Ho

Test statistics :

$$t = \frac{\bar{x}-\mu_0}{s/\sqrt{n}}$$

where $\bar{x} = 1/n \sum_{i=1}^{n} x_i$ and

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

Follows the studunts t - distribution with (n-1) d.f we now Compare the calculate value of t with the tabulated value of Certain Level of significance.

IF calculated $|t| >$ tabulated t. null hypothesis is rejected and if calculated $|t| <$ tabulated t. Ho may be accepted at the level of significance adopted.

t- Test for difference of mean :

suppose we want to test if two indopondent samples $x_i (i=1,2,\dots n)$ and $y_j (j=1,2,\dots n)$ of sizes n_1 and n_2 have been drawn from two normal populations with mean μx and μy respectively.

under the null hypothesis Ho, that the Samples have been drawn from the normal population with mean μx and μy and under the assumption that the population variance are equal i.e

$$\sigma x^2 = \sigma y^2 = \sigma^2 \text{ say}$$

$$t = \frac{(\bar{x}-\bar{y})-(\mu x-\mu y)}{S\sqrt{\frac{1}{n_1}+\frac{1}{n_2}}}$$

where $\bar{x} = \frac{1}{n_1} \sum_{i=1}^{n} xi$

$$\bar{y} = \frac{1}{n_2} \sum_{j=1}^{n} yi$$

$$S^2 = \frac{1}{n_1+n_2-2} \sum_{i} \left[(xi-\bar{x})^2 + \sum_{j} (yj-\bar{y})^2 \right]$$

is an unbiased estimate of the common population variance σ^2

Follow student's t-distribution with (n_1+n_2-2) d.f

proof:

$$z = \frac{(\bar{x}-\bar{y})-E(\bar{x}-\bar{y})}{\sqrt{V(\bar{x}-\bar{y})}} \sim N(0,1)$$

$$E(\bar{x}-\bar{y}) = E(\bar{x})-E(\bar{y})$$

$$V(\bar{x}-\bar{y}) = V(\bar{x})+V(\bar{y})$$

$$= \frac{\sigma x^2}{n_1} + \frac{\sigma y^2}{n_2} => \sigma^2 \left(\frac{1}{n_1} + \frac{1}{n_2} \right)$$

$$= \frac{(\bar{x}-\bar{y})-(\mu_x-\mu_y)}{\sqrt{\sigma^2 \left(1/n_1 + 1/n_2 \right)}} \sim N(0,1) \to \textcircled{1}$$

let

$$\chi^2 = \frac{1}{\sigma^2} \left[\sum_{i=1}^{n} (xi-\bar{x})^2 + \sum_{j=1}^{n_2} (yj-\bar{y})^2 \right]$$

$$= \left[\sum_{i} (xi-\bar{x})^2/\sigma^2 \right] + \left[\sum_{j} (yj-\bar{y})^2/\sigma^2 \right]$$

$$= \frac{n_1 s x^2}{\sigma^2} + \frac{n_2 s y^2}{\sigma^2} \to \textcircled{2}$$

since,

$$\frac{n_1 s_x^2}{\sigma^2} \text{ and } \frac{n_2 s_x^2}{\sigma^2}$$

variables with (n_1-1) and (n_2-1) d.f respectively by the additive property of chi-square defined in ② is a χ^2 variate with $(n_1-1)+(n_2-1)$ i.e n_1+n_2-2 d.f

Further since sample mean and Sample Variance are independently distributed ξ and χ^2 are independent random variable.

$$t = \frac{\xi}{\sqrt{\frac{\chi^2}{n_1+n_2-2}}}$$

$$= \frac{(\bar{x}-\bar{y})-(\mu_x-\mu_y)}{\sigma^2\sqrt{1/n_1+1/n_2}} \times \frac{1}{\frac{1}{n_1+n_2-2}\left[\left\{\sum_i (x_i-\bar{x})^2+\sum_j (y_i-\bar{y})^2\right\}/\sigma^2\right.}$$

$$= \frac{(\bar{x}-\bar{y})-(\mu_x-\mu_y)}{S\sqrt{1/n_1+1/n_2}}$$

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

and it follows Students t distribution with (n_1+n_2-2) d.f

F-distribution :

Dafinition :

If x and y are two indipendent chi-square variates with v_1 and v_2 d.f respectively them.

F statistics is defined by

$$F = \frac{X/v_1}{Y/v_2}$$

F test for Equality of Two population variances :

Suppose we want to test ii) whether two independent samples. x_i $(i=1,2,\dots n)$, y_j $(j=1,2,\dots n)$ have been drawn from the normal populations with same variance σ^2 (say) or

ii) whether the two independent estimates of the population variance are homogenious are not.

null hypothesis: Ho

i) $\sigma x^2 = \sigma y^2 = \sigma^2$

$$F = \frac{Sx^2}{Sy^2}$$

where.

$$Sx^2 = \frac{1}{n_1-1}\sum_{1=j}^{n}(xi-\bar{x})^2$$ and

$$sy^2 = \frac{1}{n_2-1} \sum_{j=1}^{n_2} (y_j-\bar{y})^2$$

proof:

$$F = \frac{Sx^2}{Sy^2} \left[\frac{n_1}{n_1-1} Sx^2\right] \Big/ \left[\frac{n_2}{n_2-1} Sy^2\right]$$

$$= \left[\frac{n_1 sx^2}{\sigma x^2} \cdot \frac{1}{(n_1-1)}\right] \Big/ \left[\frac{n_2 sy^2}{\sigma y^2} \cdot \frac{1}{n_2-1}\right]$$

$[\because \sigma x^2 = \sigma y^2 = \sigma^2]$ under H_0 ... $n_2]$

Since $\frac{n_1 sx^2}{\sigma x^2}$ and $\frac{n_2 sy^2}{\sigma y^2}$ are independent chi-square with (n_1-1) and (n_2-1) df respectively, F follows Snedecor's F distribution with $(n_1-1), (n_2-1)$ d.f.

constants of F distribution:-

μ_r' (about origin) $= E(F^r)$

$$= \int_0^\infty F^r f(F)\, dF$$

$$= \frac{(v_1/v_2)^{v_1/2}}{B(v_1/2, v_2/2)} \int_0^\infty F^r \frac{F^{(v_1/2)-1}}{(1+v_1/v_2 F)^{v_1+v_2/2}}\, dF \rightarrow ①$$

To ovalaate the integrals put = v_1/v_2 f=y so that df = v_2/v_2 dy

$$\mu_r' = \frac{[v_1/v_2]^{v_1/2}}{B\left(v_1/2,\, v_2/2\right)} \int_0^\infty \frac{\left(v_2/v_1\, y\right)^{r+(v_1/2)-1}}{(1+y)^{(v_1+v_2)/2}} \cdot \left[\frac{v_2}{v_1}\right] y$$

$$= \frac{\left(\frac{v_2}{v_1}\right)^r}{B\left(v_1/2,\, v_2/2\right.} \int_0^\infty \frac{y^{r+(v_1/2)-1}}{(1+y)^{[(v_1/2)+r+(v_2/2)-r}} \Big\} dy$$

$$\Rightarrow \left[\frac{v_2}{v_1}\right]^r, \frac{1}{B\left(v_1/2,\, v_2/2\right)} \quad B\left[r+v_1/2,\, v_2/2-r\right]$$