02-12-21

antithm

standard error $= \frac{\sigma}{\sqrt{n}}$; $\hat{H}$

The process that enables a decision maker to ~~daw~~ draw an interface about population characteristics by analyzing the difference between the value obtained from sample and the hypothesized value of param

$\boxed{\square}$ Any statement about any phenomenon is termed as hypothesis.

$\boxed{\square}$ Statical hypothesis is a statement ~~that~~ about population characteristics that can be tested on the basis of sample data.

$\boxed{\square}$ The hypothesis that is fomulated for its possible rejection using sample data is called null hypothesis : $H_0 \longrightarrow$ Denoted by

$\boxed{\square}$ The hypothesis which is true if null hypothesis is false is called alternative hypothesis. H_A or $H_1 \longrightarrow$ denoted by.

⊞ The hypothesis which completely specifies all the parameters of the related population, is called simple hypothesis.

⊞ The hypothesis which does not ~~always~~ completely specify ~~the~~ the parameters related to population, is called a composite hypothesis.

⊞ **Type I Error:**

Definition: Rejecting the null hypothesis (H_0) when it is actually true.

The probability of making type I error is represented by α. (e.g. 0.05 or 5%)

⊞ **Type II Error:**

Definition: Failing to reject null hypothesis when it is actually false.

The probability of making type II error is Represented by β.

Decision	State of nature	
	Ho is true	Ho is false
Reject Ho	Type-I Ennon	Connect Decision
Fail to Reject Ho	Connect Decision	Type-II Ennon

The probability of committing type-1 ennon is called level of sigificance.

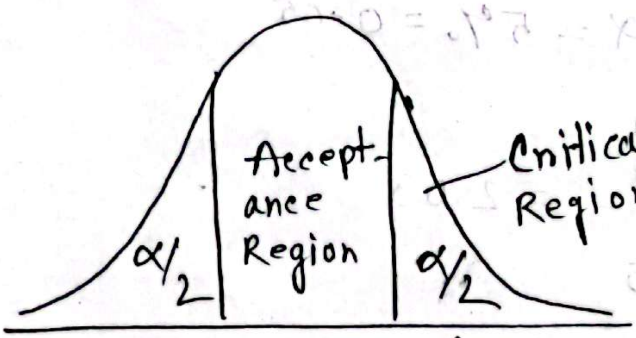
↳ Area Under critical region

symbolically,

$$\alpha = P(\text{reject } H_0 \mid H_0 \text{ is true})$$

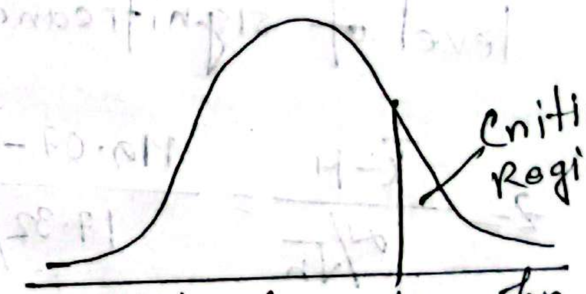
The complement of the probability of type II ennon is called the powerr of a test.

power of the test $1-\beta = 1 - P(\text{Accept } H_0 \mid H_0 \text{ is False})$



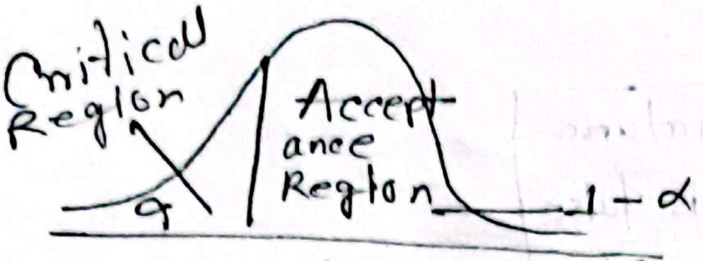
Critical Region for two tailed test size α

$H_0: \mu=0$, $H_1: \mu \neq 0$



Critical region for right tailed test of size α

$H_0: \mu=0$, $H_1: \mu>0$



Critical Region for left-tailed test of size α

$H_0 : \mu = 0$; $H_1 : \mu < 0$

P value or observed significance

~~te~~ observed level of significance level of a statical test is the smallest value at which the null hypothesis can be rejected.

$\bar{x} = 109.33$ ~~→~~ 119.07

$\sigma^2 = 300$

$\Rightarrow \sigma = 17.32$

Null hypothesis $H_0 : \mu = 110$

Alternative " $H_1 : \mu \neq 0$

level of significance $\alpha = 5\% = 0.05$

$$z = \frac{\bar{x} - \mu}{\sigma/\sqrt{n}} = \frac{119.07 - 110}{17.32/\sqrt{15}} = 2.03$$

$\mu = 880\,kg$ $\quad \sigma = 21 \quad n = 50 \quad |z| > 1.96$

$\bar{x} = 871$

[illegible]

$H_0 : \mu = 880$ and $H_1 : \mu \neq 880$

$$z = \frac{\bar{x} - \mu}{\sigma / \sqrt{n}} = \frac{871 - 880}{21/\sqrt{50}} = -3.03$$

$$z = p - \pi$$

Setting up hypothesis

↓

set up suitable level of significance

↓

Computation of test statistic

↓

Determine the critical region or rejection region

↓

Calculating test statistic

↓

taking decision

Reject H_0 and Accept H_1 $\qquad$ Accept H_0 and reject H_1

Nonparametric statistics is a method of statical analysis that does not assume a normal distribution.

Differences between parametric and non parametric test

Parametric test	Non parametric test
1. Population parameter is known	1. Parameter is unknown
2. Quantitative study	2. Qualititative study
3. Data is normally distributed	3. Data is not normally distributed
4. Large sample size	4. small sample size
5. Central tendency is mean is used	5. central tendency median is used
6. probability sampling	6. Non probability sampling
7. More powerful	7. less powerful.
8. Measurement scale is interval or ratio	8. Measurement scale is nominal or ordinal.

Advantages of non-parametric Methods:—

1. They can be used to test population parameters when the variables are not normally distributed

2. They can be used when data are nominal or ordinal

3. They can be used to test hypothesis that do not involve population parameters.

4. In some cases, the computations are easier than those parametric counterparts

5. Easy to understand.

Disadvantages:

1. They are less sensitive than their parametric counterparts, when assumption of the parametric methods are met.

2. They tend to use less information than the parametric tests.

3. They are less efficient than their parametric counterparts when the assumptions of the parametric methods are met.

(i) A sign test is a statical method that compares two related samples to determine if there is a difference between the medians of two samples.

(ii) A convience store owner hypothesizes that the median number of ~~sh~~ snow cones she sells per day 40. A random sample of 20 days yields the following data for the number of snow cones sold each day.

18	43	40	16	22
30	29	32	37	36
39	34	39	45	28
36	40	34	39	52

At $\alpha = 0.05$ test the owner's hypothesis

Solution:

H_0: median = 40 (claim) and

H_1: median $\neq$ 40

	B	A	
A	3	2	+
B	0	1	−
C	5	4	+
D	4	0	+
E	2	1	+
F	4	3	+
G	3	1	+
H	4	3	+
I	2	2	0
J	2	5	−

− + 0 − −

− − − − −

− − − + −

− 0 − − +

n = 18

$\alpha = 0.05$

test value = 3

4 > 3

H_0 : MD = 36.45

H_0 : MD ≠ 36

$\alpha = 0.05$

n = 50

n = 9 test value = 1

$\alpha = 0.05$

$$z = \frac{(X+0.5) - n/2}{\sqrt{n}/2}$$

$$= \frac{(21+0.5) - 50/2}{\sqrt{50}/2} = -0.99$$

1 < 2 rejected

$$z = \frac{R - \mu_R}{\sigma_R}$$

$$\mu_R = \frac{n_1(n_1+n_2+1)}{2}$$

$$\sigma_R = \sqrt{\frac{n_1 n_2 (n_1+n_2+1)}{12}}$$

R = sum of Ranks of smaller sample

n_1 = smaller of sample

n_2 = larger of sample

critical value = 1

$n_1 \geq 10$ $n_2 \geq 10$

Formula for z test value in the sign test when $n \geq 26$.

$$z = \frac{(X+0.5) - n/2}{\sqrt{n}/2}$$

where,

X = smaller number of sample size

n = sample size

Formula for wilcoxon Rank sum test when samples are independent

$$z = \frac{R - \mu_R}{\sigma_R}$$

$$\mu_R = n_1(n_1+n_2+1)/2$$

$$\sigma_R = \sqrt{\frac{n_1 n_2 (n_1+n_2+1)}{12}}$$

R = sum of Rank of smaller sample size n_1

n_1 = smaller sample size

n_2 = larger sample size

$(n_1 \geq 10)$ & $n_2 \geq 10$

☐ Sign test is the simple non-parametric test is used ~~for~~ to test value of a median for a specific sample.

H_0 = There is no difference

H_1 = There is a difference

8	9	10	11	12	13	14	15	15
M	M	M	M	M	A	M	~~A~~	M
16	16	17	17	18	18	19	19	
A	M	A	M	A	M	M	A	
21	22	24	25	26	28			
A								

$\Sigma R = 1 + 2 + 3 + 4 + 5 + 7 + 11.5 + 10.5 + 14.5 + 16 + 2$

$= 9.3$

$$\mu_R = \frac{12(12+12+1)}{2} = 122$$

$$\sigma_R = \sqrt{\frac{1.12(11+12+1)}{12}}$$

$$\frac{R - \mu_R}{\sigma_R} = -2.41$$

$$f(F) = \frac{\left(\frac{n_1}{n_2}\right)^{n_1/2} \cdot \left(\frac{n_1}{n_2} F\right)^{n_1/2}}{\beta\left(\frac{n_1}{2}, \frac{n_2}{2}\right)\left(1 + \frac{n_1}{n_2} F\right)^{\frac{n_1+n_2}{2}}} ; F > 0$$

FFF = 1 M

$\frac{FFF}{1}$ $\frac{MM}{2}$ $\frac{FFFF}{3}$ $\frac{M}{4}$ $\frac{F}{5}$ $\frac{MMM}{6}$ $\frac{FFFF}{7}$ $\frac{MM}{8}$ $\frac{FFF}{9}$ $\frac{MM}{10}$

$\alpha = 0.05$

$$f(x^2)\,dx^2 = \int_0^\infty \frac{1}{2^{n/2}\sqrt{n/2}} (x^2)^{n/2-1} e^{-x^2/2}\, dx^2$$

$n_1(F) = 15 = e$

$n_2(M) = 10 \longrightarrow \int_0^\infty \frac{1}{2^{n/2}\sqrt{n/2}} (x^2)^{n/2-1} e^{-x^2/2}\, dx^2$

F or 18 $\int_0^\infty \frac{1}{2^{n/2}\sqrt{n/2}} (x^2)^{n/2-1} e^{-x^2/2}\, dx^2$

x

x

$x\ x\ x$ $\quad \frac{\sqrt{n}}{\alpha^n} = \int_0^\infty \frac{1}{2^{n/2}\sqrt{n/2}} (x^2)^{n/2-1} e^{-x^2/2}\, dx^2$

$\frac{\sqrt{n}}{\alpha^n} = \int_0^\infty e^{-\alpha x} x^{n-1}$

$$\Rightarrow \frac{1}{2^{n/2}\cancel{\sqrt{n/2}}} \cdot \frac{\cancel{\sqrt{n/2}}}{\left(\frac{1}{2}\right)^{n/2}}$$

$\frac{3}{2}-1$

$\Rightarrow \frac{1}{2}\sqrt{1/2}$ $\quad \Rightarrow \frac{1}{\cancel{2^{n/2}}\ \cancel{\sqrt{n/2}}}\ \cancel{2^{n/2}}$

$\Rightarrow 1$

Question:

(b) Show that the total probability of chi-square χ^2 distribution is unity.

proof:
$$\int_0^\infty f(\chi^2)\, d\chi^2 = 1$$

The pdf of χ^2 distribution with n-degree of freedom is given by

$$f(\chi^2) = \frac{1}{2^{n/2}\Gamma\frac{n}{2}} (\chi^2)^{n/2-1} e^{-\chi^2/2} ; \quad 0 < \chi^2 < \infty$$

Thus,
$$\int_0^\infty f(\chi^2)\, d\chi^2 = \int_0^\infty \frac{1}{2^{n/2}\Gamma\frac{n}{2}} (\chi^2)^{n/2-1} e^{-\chi^2/2}\, d\chi^2$$

$$= \frac{1}{2^{n/2}\Gamma\frac{n}{2}} \int_0^\infty e^{-\frac{1}{2}\chi^2} (\chi^2)^{n/2-1}\, d\chi^2$$

$$\left[\because \frac{\Gamma n}{\alpha^n} = \int_0^\infty e^{-\alpha x} x^{n-1}\, dx\right]$$

$$= \frac{1}{2^{n/2}\cancel{\Gamma\frac{n}{2}}} \frac{\cancel{\Gamma\frac{n}{2}}}{\left(\frac{1}{2}\right)^{n/2}}$$

$$\Rightarrow \frac{1}{2^{n/2}} \cdot 2^{n/2}$$

$$= 1 \text{ (showed)}$$

Q^n: Find Mean and variance of chi-square (χ^2) distribution.

Soln: The pdf of χ^2-distribution with n degree of freedom is given by -

$$f(\chi^2) = \frac{1}{2^{n/2}\,\Gamma n/2}(\chi^2)^{n/2-1} e^{-\chi^2/2}; \quad 0 < \chi^2 < \infty$$

Mean:

$$E(\chi^2) = \int_0^\infty \frac{1}{2^{n/2}\,\Gamma n/2}(\chi^2)^{n/2-1} e^{-\chi^2/2}\, d\chi^2$$

$$\Rightarrow \frac{1}{2^{n/2}\,\Gamma n/2}\int_0^\infty (\chi^2)^{n/2-1} e^{-\chi^2/2}\, d\chi^2$$

=

$\chi \; \chi \; \chi \; t$

Question: find mean and variance of chi square (χ^2) distribution.

Soln: The pdf of χ^2 distribution with n degree of freedom is given by:

$$f(\chi^2) = \frac{1}{2^{n/2}\Gamma n/2}(\chi^2)^{n/2-1} e^{-\chi^2/2} d(\chi^2)$$

Mean:

$$E(\chi^2) = \int_0^\infty \chi^2 f(\chi^2)\, d\chi^2$$

$$= \int_0^\infty \chi^2 \frac{1}{2^{n/2}\Gamma n/2}(\chi^2)^{n/2-1} e^{-\chi^2/2} d(\chi^2)$$

$$\Rightarrow \frac{1}{2^{n/2}\Gamma n/2}\int_0^\infty \chi^{(n/2+1)-1} e^{-\chi^2/2} d\chi^2$$

$$\Rightarrow \frac{1}{2^{n/2}\Gamma n/2} \cdot \frac{\Gamma n/2+1}{(1/2)^{n/2+1}} \quad \left[\because \frac{\Gamma n}{\alpha^n} = \int_0^\infty x^{n-1} e^{-\alpha x} dx\right]$$

$$= \frac{1}{\cancel{2^{n/2}}\cancel{\Gamma n/2}} \cancel{n/2}\, \cancel{\Gamma n/2} \cdot \cancel{2} \cdot \cancel{2^{n/2}} \qquad \Gamma n+1 = n\Gamma n$$

$$= n$$

$E[\chi^2] = n$

$(E[\chi^2])^2 = n^2$

$$E[(\chi^2)^2] = \int_0^\infty (\chi^2)^2 f(\chi^2)\, d\chi^2$$

$$= \int_0^\infty (\chi^2)^2 \frac{1}{2^{n/2}\Gamma n/2} \cdot e^{-\frac{\chi^2}{2}} (\chi^2)^{n/2 - 1} d\chi^2$$

$$\Rightarrow \frac{1}{2^{n/2}\Gamma n/2} \int_0^\infty (\chi^2)^{(n/2+2)-1} e^{-\chi^2/2} \cdot d\chi^2$$

$$\Rightarrow \frac{1}{2^{n/2}\Gamma n/2} \cdot \frac{\Gamma n/2+2}{(1/2)^{n/2+2}}$$

$$\Rightarrow \frac{1}{\cancel{2^{n/2}}\,\cancel{\Gamma n/2}} (n/2+1)\, n/2 \cdot \cancel{\Gamma n/2} \cdot 2^2 \cdot \cancel{2^{n/2}}$$

$$\Rightarrow \left(\frac{n^2}{4} + \frac{n}{2}\right) 4 = n^2 + 2n$$

$$n^2 + 2n - n^2 = 2n$$

$$f(t) = \frac{1}{\sqrt{n}\, \beta(1/2, n/2) \left(1 + \frac{t^2}{n}\right)^{\frac{n+1}{2}}}, \; -\infty < t < \infty$$

$$-\frac{1}{2} + 1$$

$$\Rightarrow \frac{-1+2}{2}$$

$$3/2 - \frac{1}{2} = \frac{n}{2} - 1$$

$$-\frac{1}{2} + 1 - 1 + 1$$

$$\Rightarrow \frac{3-1}{2} = 1 \quad \Rightarrow \frac{n-2}{2}$$

$$-\frac{1}{2} + 1 - 1 + 1$$

$$\Rightarrow -\frac{1}{2} + 2 - 1$$

$$\Rightarrow \frac{-1+4}{2} - 1$$

$$\Rightarrow \frac{3}{2} - 1$$

$$\Rightarrow$$

☐ Sampling distribution is a distribution of statistic obtained ~~from~~ through large number of samples drawn from a specific population.

for example. χ^2, F, t distribution

☐ The probability distribution of a parameter is called parent distribution.

for example: Normal, binomial distributions.

☐ Sum of squares of n independent standard normal variates is call chi (χ^2)-square variate with n degrees of freedom.

Let, $z_1, z_2, z_3 \ldots z_n$ be n independent standard normal variates, then chi-square denoted by χ^2 is defined as

$$\chi^2_n = \sum_{i=1}^{n} z_i^2$$

☐ However if $x_1, x_2, x_3, \ldots x_n$ are n independently and ~~indel~~ indentically distributed ~~normal~~ random variable each of which is normally distributed with mean μ and variance σ^2. Then

$$\chi^2_n = \sum_{i=1}^{n}\left(\frac{x_i-\mu}{\sigma}\right)^2$$

is a chi-square (χ^2_n) variable with n degree of freedom.

properties of χ^2 distribution: –

(i) χ^2 is a continuous type of distribution and its range is 0 to ∞ i.e $0 < \chi^2 < \infty$

(ii) The distribution contains only one parameter which is degree of freedom of distribution.

(iii) The mean and variance of χ^2 – distribution for n.d.f is n and 2n respectively.

(iv) The mode of χ^2 distribution for n.d.f is (n−2)

(v) ~~The~~ χ^2 – distribution tends to normal distribution for large degree of freedom

(vi) It is positively skewed distribution for smaller values of n.

(vii) The distribution become symmetrical as n tends to infinity $(n \to \infty)$

Application / uses of chi-square (χ^2) distribution:

(i) To test if the hypothetical value of the population variance is $\sigma^2 = \sigma_0^2$ (say).

(ii) To test the goodness of fit.

(iii) To test the independence of attributes.

(iv) To test the homogeneity of independent estimates of the population variance.

(v) To test the homogeneity of independent estimate of the population correlation coefficient.

⊡ Show that, the total probability of chi-square (χ^2) - distribution is unity.

i.e, $\int_0^\infty f(\chi^2)\, d\chi^2 = 1$

$$\Rightarrow \int_0^\infty \frac{1}{2^{n/2}\,\Gamma n/2}\, e^{-\chi^2/2}\, (\chi^2)^{n/2-1}\, d\chi^2$$

$$\Rightarrow \frac{1}{2^{n/2}\,\Gamma n/2} \int_0^\infty e^{-\chi^2/2}\, (\chi^2)^{n/2-1}\, d\chi^2$$

$$\Rightarrow \frac{1}{2^{n/2}\,\cancel{\Gamma n/2}}\, \frac{\cancel{\Gamma n/2}}{(1/2)^{n/2}} \left[\int_0^\infty e^{-ax} x^{n-1}\, dx = \frac{\Gamma n}{a^n}\right]$$

$$\Rightarrow 1$$

Find mean and variance of chi-square (χ^2) distribution.

Solution:

The p.d.f of χ^2 distribution with n degree of freedom is given by

$$f(\chi^2) = \frac{1}{2^{n/2}\Gamma n/2} (\chi^2)^{n/2 - 1} e^{-\chi^2/2} d\chi^2$$

$$E(\chi^2) = \int_0^\infty \chi^2 f(\chi^2)\, d\chi^2$$

$$= \int_0^\infty \chi^2 \frac{1}{2^{n/2}\Gamma n/2} (\chi^2)^{n/2 - 1} e^{-\chi^2/2} d\chi^2$$

$$\Rightarrow \frac{1}{2^{n/2}\Gamma n/2} \int_0^\infty e^{-\chi^2/2} (\chi^2)^{(n/2+1)-1} \cancel{e^{-\chi^2/2}}\, d\chi^2$$

$$\Rightarrow \frac{1}{2^{n/2}\Gamma n/2} \frac{\Gamma n/2 + 1}{\left(\frac{1}{2}\right)^{n/2+1}}$$

$$\Rightarrow \frac{\cancel{n/2}\,\cancel{\Gamma n/2}\; \cancel{2}\cdot\cancel{2^{n/2}}}{\cancel{2^{n/2}}\,\cancel{\Gamma n/2}}$$

$$= n \quad \text{(Ans)}$$

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

$$= \frac{1}{2^{n/2}\Gamma n/2} \int_0^\infty e^{-x^2/2} (x^2)^{(n/2+2)-1} dx^2$$

$$\Rightarrow \frac{1}{2^{n/2}\Gamma n/2} \frac{\Gamma n/2+2}{\left(\frac{1}{2}\right)^{\Gamma n/2+2}}$$

$$= \frac{1}{\cancel{2^{n/2}}\cancel{\Gamma n/2}} 2^2 \cdot \cancel{2^{\Gamma n/2}} \cdot \left(\frac{n}{2}+1\right) \cdot n/2 \cdot \cancel{\Gamma n/2}$$

$$\Rightarrow 2^2 \cdot \left(\frac{n}{2}+1\right) \cdot n/2$$

$$\Rightarrow \cancel{2n+4} \quad \cancel{2n \; n^2+} \quad n^2+2n$$

$$(E[x^2])^2 = (n)^2 = n^2$$

$$\text{Variance} = \cancel{n^2} + 2n - \cancel{n^2}$$
$$= 2n$$

Mean $= n$

uestion:

let u be a N(0,1) varriale and v be a chi-square ~~variabl~~ variate with n degree of freedom. Also u and v are independent.

Define $t = \frac{u}{\sqrt{v/n}}$, then t ~~will~~ will follow t distribution with n degree of freedom.

The form of t distribution with n degree of freedom is given below:-

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

Properties: of t distribution:

(I) t-distribution is an even function.

(II) t-distribution is symmetric about $t = 0$

(III) Mean, Media, Mode = 0

(IV) Variance of the distribution is $\frac{n}{n-2}$, $n>2$

(V) total probability of t-distribution is equal to 1.

i.e $\int_{-\infty}^{\infty} f(t) \cdot dt = 1$

(VI) For large n t-distribution reduces to standard normal distribution.

(vii) Since, $\beta_1 = 0$ and $\beta_2 = 3 + \frac{6}{n-4} > 3$ therefore, the distribution is symmetric ($\beta_1 = 0$) and leptokurtic ($\beta_2 > 3$)

(viii) If is a continuous type of distribution and its range extends from $-\infty$ to ∞. i.e $-\infty < t < \infty$

(ix) Mgf of t-distribution doest not exist.

Application or uses of t distribution:-

(i) To test if sample mean $(\bar{x})$ differ ~~finical~~ significantly from hypothetical value of μ of the population mean.

(ii) To test the significance of the difference between two sample mean.

(iii) To test the single population mean

(iv) To test the significance of an observed partial correlation coefficient

(v) To test the significance of observed sample correlation coefficient and sample regression coefficient.

Distinguish between t and normal distribution:

t-distribution	Normal distribution
(I) The pdf of t-distribution is $f(t) = \frac{1}{\sqrt{n}\, \beta(1/2, n/2)\left(1+\frac{t^2}{n}\right)^{\frac{n+1}{2}}}$, $-\infty < t < \infty$	(I) The pdf of normal dist^n is $f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-1/2\left(\frac{x-\mu}{\sigma}\right)^2}$, $-\infty < x < \infty$
(II) Mean, median, mode of this distribution conside at zero	(II) Mean, Median, mode of this distribution are not zero
(III) It is an exact sampling distribution	(III) It is parent distribution
(IV) The distribution is symmetric and leptokurtic since $\beta_1 = 0$ and $\beta_2 > 3$	(IV) The distribution is symmetric and mesokurtic since, $\beta_1 = 0$ and $\beta_2 = 3$

Question:

Show that, the total probability of t- density is equal to 1.

i.e $\int_{-\infty}^{\infty} f(t)\,dt = 1$

- **proof:**

$$f(t) = \frac{1}{\sqrt{n}\,\beta(1/2, n/2)\left(1+\frac{t^2}{n}\right)^{\frac{n+1}{2}}}$$

$$\int_{-\infty}^{\infty} f(t)\,dt = \int_{-\infty}^{\infty} \frac{1}{\sqrt{n}\,\beta(1/2, n/2)\left(1+\frac{t^2}{n}\right)^{\frac{n+1}{2}}}$$

$$= \frac{1}{\sqrt{n}\,\beta(1/2, n/2)} \int_{-\infty}^{\infty} \frac{dt}{\left(1+\frac{t^2}{n}\right)^{\frac{n+1}{2}}}$$

let,

$\frac{t^2}{n} = w \Rightarrow t = \sqrt{nw}$

$\Rightarrow 2t \cdot dt = dw\, n$

$\Rightarrow dt = \frac{n}{2t}\,dw$

$= \frac{n}{2\sqrt{nw}}\,dw$

$= \frac{\sqrt{n}}{2\sqrt{w}}\,dw$

$$\int_{-\infty}^{\infty} f(t)\,dt = \cancel{2}\int_0^{\infty} \frac{1}{\cancel{\sqrt{n}}\,\beta\left(\frac{1}{2},\frac{n}{2}\right)(1+w)^{\frac{n+1}{2}}}\,\frac{\cancel{\sqrt{n}}}{\cancel{2}\sqrt{w}}\,dw$$

$$\Rightarrow 2 \int \quad \frac{1}{\cancel{2}\,\beta\left(\frac{1}{2},\frac{n}{2}\right)}\int_0^{\infty} \frac{w^{-1/2}}{(1+w)^{\frac{n+1}{2}}}\,dw$$

$$= \frac{1}{\cancel{2}\,\beta\left(\frac{1}{2},\frac{n}{2}\right)}\int_0^{\infty} \frac{w^{1/2-1}}{(1+w)^{1/2+n/2}}\,dw$$

$$\left[\because \beta(l,m) = \int_0^{\infty} \frac{x^{l-1}}{(1+x)^{l+m}}\,dx\right]$$

$$= \frac{1}{\cancel{\beta\left(\frac{1}{2},\frac{n}{2}\right)}}\cdot \cancel{\beta\left(\frac{1}{2},\frac{n}{2}\right)}$$

$$= 1$$

$$\therefore \int_{-\infty}^{\infty} f(t)\,dt = 1$$

Therefore the total probability of t-density is equal to 1.

(showed)

Question:

Find mean, variance of t- distribution.

Answer:

Mean:

We know that, the pdf of t-distribution is

$$f = \frac{1}{\sqrt{n}\ \beta\left(\frac{1}{2}, \frac{n}{2}\right)\left(1+\frac{t^2}{n}\right)^{\frac{n+1}{2}}} \ ; \ -\infty < t < \infty$$

We know,

$$E(t) = \int_{-\infty}^{\infty} t \cdot f(t)\, dt$$

$$= \int_{-\infty}^{\infty} t \cdot \frac{1}{\sqrt{n}\ \beta\left(\frac{1}{2}, \frac{n}{2}\right)\left(1+\frac{t^2}{n}\right)^{\frac{n+1}{2}}}\, dt$$

$$\Rightarrow E(t) = \frac{1}{\sqrt{n}\ \beta\left(\frac{1}{2}, \frac{n}{2}\right)} \int_{-\infty}^{\infty} \frac{t}{\left(1+\frac{t^2}{n}\right)^{\frac{n+1}{2}}}\, dt$$

$$= \frac{1}{\sqrt{n} \cdot \beta\left(\frac{1}{2}, \frac{n}{2}\right)} \cdot 0$$

$= 0$ [since the integrated is an odd function of t]

$\therefore$ Mean $= E(t) = 0$

Now,

$$E(t^n) = \int_{-\infty}^{\infty} t^n f(t) \cdot dt$$

$$= \int_{-\infty}^{\infty} \frac{t^n}{\sqrt{n}\, \beta(1/2, n/2) \left(1 + t^2/n\right)^{\frac{n+1}{2}}} dt$$

let, $w = t^2/n \Rightarrow t^2 = nw \Rightarrow t = \sqrt{nw}$

$\Rightarrow n\,dw = 2t \cdot dt$

$\Rightarrow dt = \dfrac{n\,dw}{2t}$

$= \dfrac{n}{2\sqrt{nw}} dw$

$\Rightarrow \dfrac{\sqrt{n}}{2\sqrt{w}} dw$

$$\beta(l, m) = \int_0^{\infty} \frac{x^{l-1}}{(1+x)^{l+m}} dx$$

when, $t = -\infty$ ~~t~~ then $w = -\infty$

" $t = \infty$ then $w = \infty$

$$E(t^2) = \int_{-\infty}^{\infty} \frac{nw}{\cancel{\sqrt{n}}\, \beta(1/2, n/2) (1+w)^{\frac{n+1}{2}}} \frac{\cancel{\sqrt{n}}}{2\sqrt{w}} dw$$

$$= \frac{n}{\beta(1/2, n/2)} \cdot \cancel{2} \cdot \int_0^{\infty} \frac{w^{-1/2+1}}{(1+w)^{1/2+n/2}} \cdot \frac{dw}{\cancel{2}}$$

$$\Rightarrow \frac{n}{\beta(1/2, n/2)} \cdot \int_0^{\infty} \frac{w^{3/2-1}}{(1+w)^{3/2+\frac{n-2}{2}}} dw$$

$$\Rightarrow \frac{n}{\beta(1/2, n/2)} \cdot \beta\left(3/2, \frac{n-2}{2}\right)$$

$$E(t^2) = n \frac{\Gamma 3/2 \cdot \Gamma \frac{n-2}{2}}{\Gamma \frac{3}{2} + \frac{n-2}{2}} \Bigg/ \frac{\Gamma 1/2 \cdot \Gamma n/2}{\Gamma 1/2 + n/2}$$

$$= \frac{n \cdot \frac{\cancel{1/2 \Gamma 1/2}}{\cancel{\Gamma \frac{n+1}{2}}}}{} \Bigg/ \frac{\cancel{\Gamma \frac{n+1}{2}}}{\Gamma 1/2 \cdot \Gamma n/2}$$

$$\Rightarrow \cancel{n \cdot \frac{1}{2} \Gamma 1/2}$$

$$\Rightarrow \frac{n \cdot \frac{1}{2} \Gamma 1/2 \cdot \Gamma \frac{n-2}{2}}{\Gamma \frac{n+1}{2}} \cdot \frac{\Gamma \frac{n+1}{2}}{\Gamma 1/2 \cdot \Gamma n/2}$$

$$\Rightarrow \frac{\frac{n}{2} \cdot \Gamma \frac{n-2}{2}}{\left(\frac{n}{2} - 1\right) \Gamma \frac{n-2}{2}}$$

$$\Rightarrow \frac{n}{2} \cdot \frac{2}{n-2}$$

$$= \frac{n}{n-2}$$

$$E(t^2) = \frac{n}{n-2}$$

$$V(t) = E(t^2) - \left(E(t)\right)^2$$

$$= \frac{n}{n-2} - 0 = \frac{n}{n-2}$$

Question:

Show that, mean, median, mode of t - distribution are identical or equal and hence its zero.

i.e. mean = median = mode = 0

Answer:

Mean:

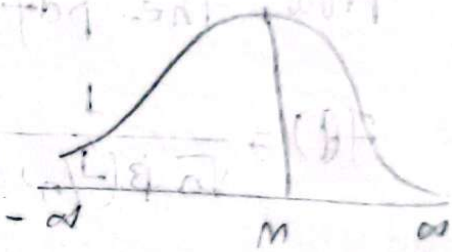
We have already got it in the last question.

i.e.

$E(t) = 0$

Mean = 0

Median:



Let, M be the median of the distribution.

$$\therefore \int_{-\infty}^{M} f(t)\,dt = \frac{1}{2} = \int_{M}^{\infty} f(t)\,dt$$

Now,

$$M = \int_{M}^{\infty} f(t)\,dt = 1/2 \quad \text{---} (I)$$

i.e. $\int_{-\infty}^{\infty} f(t)\,dt = 1$

$$\Rightarrow 2\int_{0}^{\infty} f(t)\,dt = 1 \Rightarrow \int_{0}^{\infty} f(t)\,dt = 1/2 \quad \text{---} (II)$$

Comparing equⁿ with equⁿ ②

$M = 0$

Median $= 0$

Mode of t-distribution

Mode will be obtained by the solution of the equation.

$$\frac{d \log f(t)}{dt} = 0 \; ; \text{ provided } \frac{d^2 \log f t}{dt^2} < 0$$

Now, the pdf of t distribution is

$$f(t) = \frac{1}{\sqrt{n}\, \beta(1/2, n/2) (1 + t^2/n)^{\frac{n+1}{2}}}$$

$$\Rightarrow \log f(t) = \log \frac{1}{\sqrt{n}\, \beta(1/2, n/2)} + \log \frac{1}{(1+t^2/n)^{\frac{n+1}{2}}}$$

$$\Rightarrow \frac{d \log f(t)}{dt} = 0 + \frac{d}{dt}\left(\log (1 + t^2/n)^{-(n+1)/2}\right)$$

$$= -\frac{n+1}{\cancel{2}} \cdot \frac{1}{(1 + t^2/n)} \cdot \frac{\cancel{2}t}{n}$$

$$= -\frac{(n+1)t}{n(1 + t^2/n)}$$

$$\frac{d \log f(t)}{dt} = 0$$

$$-\frac{(n+1)t}{n(1+t^2/n)} = 0$$

$$\Rightarrow \quad -t(n+1) = 0$$

$\therefore t = 0$

It is easy to verify that $\frac{d^2 \log f(t)}{dt^2} < 0$ at

$t = 0$

Hence, ~~t~~ $t = 0$ is the mode of the distribution.

Mode $= 0$

Hence, mean $=$ median $=$ mode $= 0$

<u>Question</u>: Establish the relationship between t-distribution and ~~chau~~ cauchy distribution.

<u>Answer:</u>

The ~~Relation~~ relationship between t-distribution and cauchy distribution are given follows:

we know that, the pdf of t distribution is

$$f(t) = \frac{1}{\sqrt{n}\, B(1/2, n/2)\, (1+t^2/n)^{\frac{n-1}{2}}}$$

If $n=1$, then we get the form of above equation.

$$f(t) = \frac{1}{\sqrt{n}\,\beta(1/2, n/2)\,(1+t^2)^{\frac{n+1}{2}}}$$

$$f(t) = \frac{1}{\beta(1/2, 1/2)(1+t^2)^1} = \frac{1}{\frac{\Gamma 1/2\,\Gamma 1/2}{\Gamma 1/2 + 1/2}\,(1+t^2)}$$

$$= \frac{1}{\sqrt{\pi}\sqrt{\pi}\,(1+t^2)} \qquad [\because \Gamma 1/2 = \sqrt{\pi}] \quad [\Gamma 1 = 1]$$

$$= \frac{1}{\pi(1+t^2)}\ ; -\infty < t < \infty$$

which is the pdf of standard cauchy distribution which is the relationship between t-distribution and cauchy distribution.

"F-distribution"

F-distribution: The F-distribution is the distribution of the ratio of two independent chi-square (χ^2) random variables devided by their respective degrees of freedom.

If χ_1^2 and χ_2^2 are two independent chi-square variates having n_1 and n_2 degrees of freedom respectively, then the statistic is given as —

$$F = \frac{\chi_1^2/n_1}{\chi_2^2/n_2}$$

has the F-distribution with n_1 and n_2 degrees of freedom.

In matematically, $F \sim F(n_1, n_2)$

The density function of F is

$$f(F) = \frac{\frac{n_1}{n_2}\left(\frac{n_1}{n_2}F\right)^{\frac{n_1}{n_2}-1}}{\beta\left(\frac{n_1}{2}, \frac{n_2}{2}\right)\left(1+\frac{n_1}{n_2}F\right)^{\frac{n_1+n_2}{2}}} ; F>0$$

$$f(F) = \frac{\left(\frac{n_1}{n}\right)^{\frac{n_1}{2}}\left(\frac{n_1}{n_2}F\right)^{n_1/n_2-1}}{\beta\left(\frac{n_1}{2}, \frac{n_2}{n}\right)\left(1+\frac{n_1}{n_2}F\right)^{\frac{n_1+n_2}{2}}} \quad F>0$$

Prob: Properties of F-distribution:

(i) F-distribution is a continuous type of distribution and its range 0 to ∞ i.e. $0 < F < \infty$

(ii) It is an exact sampling distribution.

(iii) It is derived from chi-square (χ^2) distribution.

(iv) The mode of F-distribution is $\frac{n_2(n_1-2)}{n_1(n_2+2)}$

(v) The distribution is positively skewed

(vi) If n_1 and n_2 are to very large, the F-distribution tends to normal distribution.

(vii) If $F \sim F(n_1, n_2)$, then $1/F \sim F(n_1, n_2)$

(viii) If $F \sim F(n_1, n_2)$, then $\frac{n_1}{n_2} F \sim \beta_1\left(\frac{n_1}{2}, \frac{n_2}{2}\right)$

(ix) If $F \sim F(n_1, n_2)$, then $\frac{1}{1+\frac{n_1}{n_2}F} \sim B_1\left(\frac{n_1}{2}, \frac{n_2}{2}\right)$

Application on uses of F distribution:

(i) F distribution is used to test the equality population variance.

(ii) It is used for testing the linearity of regression.

(iii) F distribution is used to test the equality of several means.

(iv) It is used for testing the significance of and observed multiple correlation coefficient and sample correlation ratio.

Show that, the total probability of F-density is equal to 1.

$$\text{i.e.} \int_0^\infty f(t)\, dt = 1$$

proof:

we know that, the pdf of F-distribution is as.

$$f(F) = \frac{\left(\frac{n_1}{n_2}\right)^{n_1/2} \left(\frac{n_1}{n} F\right)^{n/2 - 1}}{\beta\left(\frac{n_1}{2}, \frac{n_2}{2}\right)\left(1 + \frac{n_1}{n_2} F\right)^{\frac{n_1+n_2}{2}}} \; ; \; 0 < F < \infty$$
$(F > \infty)$

$$\text{Now,} \int_0^\infty f(F)\, dF = \int_0^\infty \frac{\frac{n_1}{n_2}\left(\frac{n_1}{n_2} F\right)^{\frac{n_1}{2} - 1}}{\beta\left(\frac{n_1}{2}, \frac{n_2}{2}\right)\left(1 + \frac{n_1}{n_2} F\right)^{\frac{n_1+n_2}{2}}} dF$$

Let, $w = \frac{n_1}{n_2} F$

$$\Rightarrow F = \frac{n_2}{n_1} w$$

$$\Rightarrow dF = \frac{n_2}{n_1} dw$$

when, $F = 0$ then $w = 0$

$F = \infty$, then $w = \infty$

$$\Rightarrow \int_0^\infty f(F)\cdot dF = \frac{\frac{n_1}{n_2}\cdot\left(\frac{n_1}{n_2}\right)^{\frac{n_1}{2}-1}}{\beta\left(\frac{n_1}{2},\frac{n_2}{2}\right)}\int_0^\infty \frac{\left(\frac{n_2}{n_1}w\right)^{n_1/2-1}}{(1+w)^{\frac{n_1+n_2}{2}}}\cdot\frac{n_2}{n_1}\,dw$$

$$= \frac{\left(\frac{n_1}{n_2}\right)^{\frac{n_1}{2}-1}\cdot\left(\frac{n_2}{n_1}\right)^{n_1/2-1}}{\beta\left(\frac{n_1}{2},\frac{n_2}{2}\right)}\int_0^\infty \frac{w^{n_1/2-1}}{(1+w)^{n_1/2+n_2/2}}\,dw$$

$$\Rightarrow \frac{1}{\beta\left(\frac{n_1}{2},\frac{n_2}{2}\right)}\cdot\beta\left(\frac{n_1}{2},\frac{n_2}{2}\right)$$

$$= 1$$

$$\int_0^\infty f(F)\,dF = 1$$

Find mean and variance of F-distribution.

Answer:

We know that, the p.d.f of F-distribution is,

$$f(F) \quad \frac{\left(\frac{n_1}{n_2}\right)\cdot\left(\frac{n_1}{n_2}F\right)^{n_1/2-1}}{B\left(\frac{n_1}{2},\frac{n_2}{2}\right)\left(1+\frac{n_1}{n_2}F\right)^{\frac{n_1+n_2}{2}}}$$

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

$$= \int_0^\infty \frac{\left(\frac{n_1}{n_2}\right)\left(\frac{n_1}{n_2}F\right)^{n_1/2-1}}{B\left(\frac{n_1}{2},\frac{n_2}{2}\right)\left(1+\frac{n_1}{n_2}\right)F}\,dF\cdot F$$

let,

$$\frac{n_1}{n_2}F = w$$

$F=0$, $w=0$

$$\Rightarrow F = \frac{n_2}{n_1}w$$

$F=\infty$, $w=\infty$

$$\Rightarrow dF = \frac{n_2}{n_1}\cdot dw$$

$$E(F) = \int_0^\infty \frac{\left(\frac{n_2}{n_1}\cdot w\right)\left(\frac{n_1}{n_2}\right)\cdot\left(\frac{\cancel{n_1}}{\cancel{n_2}}\cdot\frac{\cancel{n_2}}{\cancel{n_1}}w\right)^{\frac{n_1}{2}-1}}{B\left(\frac{n_1}{2},\frac{n_2}{2}\right)(1+w)^{\frac{n_1+n_2}{2}}}\,\frac{n_2}{n_1}\,dw$$

$$= \frac{n_2}{n_1} \int_0^\infty \frac{w^{\left(\frac{n_1}{2}+1\right)-1}}{(1+w)^{\frac{n_1+n_2}{2}}} \, dF \cdot {}^{-1}\!\Big/ \beta\left(\frac{n_1}{2}, \frac{n_2}{2}\right)$$

$$= \frac{n_2}{n_1} \int_0^\infty \frac{w^{\left(\frac{n_1}{2}+1\right)-1}}{(1+w)^{\left(\frac{n_1}{2}+1\right)+\left(\frac{n}{2}-1\right)}} \, dF \cdot {}^{-1}\!\Big/ \beta\left(\frac{n_1}{2}, \frac{n_2}{2}\right)$$

$$\Rightarrow \frac{n_2}{n_1} \, \beta\left(\left(\frac{n_1}{2}+1\right), \left(\frac{n}{2}-1\right)\right) \cdot \beta\left(\frac{n_1}{2}, \frac{n_2}{2}\right)$$

$$\Rightarrow \frac{n_2}{n_1} \; \frac{\Gamma\left(\frac{n_1}{2}+1\right) \Gamma\left(\frac{n_2}{2}-1\right) \Big/ \Gamma\left(\frac{n_1}{2}+\frac{n_2}{2}\right)}{\dfrac{\Gamma\frac{n_1}{2} \, \Gamma\frac{n_2}{2}}{\Gamma\left(\frac{n_1}{2}+\frac{n_2}{2}\right)}}$$

$$\Rightarrow \frac{n_2}{\cancel{n_1}} \cdot \frac{\frac{\cancel{n_1}}{2} \, \cancel{\Gamma\frac{n_1}{2}} \cdot \frac{n_2}{2} \cdot \cancel{\Gamma\frac{n_2}{2}-1}}{\cancel{\Gamma\frac{n_1}{2}} \cdot \cancel{\beta}\frac{n_2}{2}-1 \; \cancel{\Gamma\frac{n_2}{2}-1}}$$

$$\Rightarrow \frac{n_2}{2\left(\frac{n_2}{2}-1\right)}$$

$$\Rightarrow \frac{n_2}{n_2-2}$$

$$E(F^r) = \int_0^\infty F^r \frac{\frac{n_1}{n_2}\left(\frac{n_1}{n_2}F\right)^{\frac{n_1}{2}-1}}{B\left(\frac{n_1}{2},\frac{n_2}{2}\right)\left(1+\frac{n_1}{n_2}F\right)^{\frac{n_1+n_2}{2}}}$$

$$\Rightarrow \int_0^\infty \frac{\left(\frac{n_2}{n_1}w\right)^r \cdot \frac{n_1}{n_2}\left(\frac{n_1}{n_2}\cdot w\,\frac{n_2}{n_1}\right)^{\frac{n_1}{2}-1}}{B\left(\frac{n_1}{2},\frac{n_2}{2}\right)\left(1+\cancel{\frac{n_1}{n_2}}\,w\right)^{\frac{n_1+n_2}{2}}}\,\frac{n_2}{n_1}\,dw$$

$$\Rightarrow \frac{\left(\frac{n_2}{n_1}\right)^r}{B\left(\frac{n_1}{2},\frac{n_2}{2}\right)} \int_0^\infty \frac{(w)^{(n_1/2+2)-1}}{(1+w)^{\frac{n_1}{2}+\left(\frac{n_2}{2}\right)}}$$

$$\Rightarrow \frac{\left(\frac{n_2}{n_1}\right)^r}{B\left(\frac{n_1}{2},\frac{n_2}{2}\right)} \int_0^\infty \frac{w^{(n_1/2+2)-1}}{(1+w)^{\left(\frac{n_1}{2}+2\right)+\left(\frac{n_2}{2}-2\right)}}$$

$$\Rightarrow \frac{\left(\frac{n_2}{n_1}\right)^r}{B\left(\frac{n_1}{2},\frac{n_2}{2}\right)} \cdot B\left(\frac{n_1}{2}+2,\ \frac{n_2}{2}-2\right)$$

$$\Rightarrow \left(\frac{n_2}{n_1}\right)^r \frac{\Gamma\left(\frac{n_1}{2}+2\right)\Gamma\left(\frac{n_2}{2}-2\right)}{\cancel{\Gamma\left(\frac{n_1}{2}+\frac{n_2}{2}\right)}} \cdot \frac{\cancel{\Gamma\left(\frac{n_1}{2}+\frac{n_2}{2}\right)}}{\Gamma\frac{n_1}{2}\cdot\Gamma\frac{n_2}{2}}$$

$$\Rightarrow \left(\frac{n_2}{n_1}\right)^r \frac{\left(\frac{n_1}{2}+1\right)\left(\frac{n_1}{2}\right)\cancel{\Gamma\frac{n_1}{2}}\ \cancel{\Gamma\left(\frac{n_2}{2}-2\right)}}{\cancel{\Gamma\frac{n_1}{2}}\cdot\left(\frac{n_2}{2}-1\right)\left(\frac{n_2}{2}-2\right)\cancel{\Gamma\left(\frac{n_2}{2}-2\right)}}$$

$$\Rightarrow \frac{\frac{n_2^2}{n_1^2}\left(\frac{n_1+2}{2}\right)\cdot\frac{n_1}{2}}{\left(\frac{n_2-2}{2}\right)\left(\frac{n_2-4}{2}\right)}$$

$$\Rightarrow \frac{n_2^2\,(n_1+2)}{n_1\,(n_2-2)(n_2-4)}$$

$$V(F) = E(F^2) - [E(F)]^2$$

$$= \frac{n_2^2\,(n_1+2)}{n_1(n_2-2)(n_2-4)} - \frac{n_2^2}{(n_2-2)^2}$$

$$= \frac{n_2^2(n_1+2)(n_2-2) - n_2^2\, n_1(n_2-4)}{n_1(n_2-2)^2(n_2-4)}$$

$$= \frac{2n_2^2(n_2+n_1-2)}{n_1(n_2-2)^2(n_2-4)}$$

Question:
Establish relationship between t and F distribution.

Sol^n:
We know,
The pdf of F distribution with n_1 and n_2 degrees of freedom is —

$$f(F) = \frac{\left(\frac{n_1}{n_2}\right)\left(\frac{n_1}{n_2}F\right)^{\frac{n_1}{2}-1}}{\beta\left(\frac{n_1}{2}, \frac{n_2}{2}\right)\left(1+\frac{n_1}{n_2}F\right)^{\frac{n_1+n_2}{2}}}$$

Now putting $n_1 = 1$ and $n_2 = n$ we get

$$f(F) = \frac{\frac{1}{n}\left(\frac{1}{n}F\right)^{1/2-1}}{\beta\left(\frac{1}{2}, \frac{n}{2}\right)\left(1+\frac{1}{n}F\right)^{\frac{1+n}{2}}}$$

$$= \frac{F^{1/2-1}}{\sqrt{n}\,\beta\left(\frac{1}{2}, \frac{n}{2}\right)\left(1+\frac{1}{n}F\right)^{\frac{1+n}{2}}}$$

Let, $t^2 = F$

$\Rightarrow 2t = \frac{dF}{dt} = J$

$\Rightarrow dF = 2t\,dt$

$$\frac{\frac{1}{2}+\frac{n}{2}}{\Gamma n/2}$$

$$\frac{1}{\sqrt{2n}}$$

hence, pdf of t distribution is,

$$f(t) = \frac{(t^2)^{-1/2}}{\sqrt{n}\ \beta\left(\frac{1}{2}, \frac{n}{2}\right)\left(1 + t^2/n\right)^{\frac{n+1}{2}}} |J|$$

$$= \frac{2t \cdot t^{-1}}{\sqrt{n}\ \beta\left(\frac{1}{2}, \frac{n}{2}\right)\left(1 + \frac{t^2}{n}\right)^{\frac{n+1}{2}}}$$

$$\therefore f(t) = \frac{2}{\sqrt{n}\ \beta\left(\frac{1}{2}, \frac{n}{2}\right)\left(1 + \frac{t^2}{n}\right)^{\frac{n+1}{2}}}$$

this function is not one to one the the function is an even function.

Therefore the pdf of t distribution is

$$f(t) = \frac{1}{\sqrt{n}\ \beta\left(\frac{1}{2}, n/2\right)\left(1 + t^2/n\right)^{\frac{n+1}{2}}} \ ; -\infty < t < \infty$$

which is the pdf of t distribution with n degrees of freedom.

Hence $t^2 \sim F(1, n)$ (showed).

This is the relationship between t_n and $f_{(1,n)}$ distribution

Question: Show that large degree of freedom t - distribution tends to normal distribution.

proof:
We know that, the pdf of t - distribution is,

$$f(t) = \frac{1}{\sqrt{n}\,\beta\left(\frac{1}{2}, \frac{n}{2}\right)\left(1+\frac{t^2}{n}\right)^{\frac{n+1}{2}}} \; ; \; -\infty < t < \infty$$

$$\therefore \; f(t) = \frac{1}{\sqrt{n}\,\beta\left(\frac{1}{2}, \frac{n}{2}\right)} \cdot \frac{1}{\left(1+t^2/n\right)^{\frac{n+1}{2}}}$$

taking limits on both side we have,

$$\lim_{n\to\infty} f(t) = \lim_{n\to\infty} \frac{1}{\sqrt{n}\,\beta\left(\frac{1}{2}, \frac{n}{2}\right)} \cdot \lim_{n\to\infty} \frac{1}{\left(1+t^2/n\right)^{\frac{n+1}{2}}}$$

Now, $$\lim_{n\to\infty} \frac{1}{\sqrt{n}\,\beta\left(\frac{1}{2}, \frac{n}{2}\right)} = \lim_{n\to\infty} \frac{\Gamma\left(\frac{1}{2}+\frac{n}{2}\right)}{\sqrt{n}\,\Gamma\frac{1}{2}\cdot\Gamma\frac{n}{2}}$$

$$= \lim_{n\to\infty} \frac{\Gamma\left(\frac{1}{2}+\frac{n}{2}\right)}{\sqrt{n}\sqrt{\pi}\,\Gamma\frac{n}{2}}$$

$$\Rightarrow \frac{1}{\sqrt{n}\sqrt{\pi}} \lim_{n\to\infty} \frac{\Gamma\left(\frac{1}{2}+\frac{n}{2}\right)}{\Gamma\frac{n}{2}}$$

$$= \frac{1}{\sqrt{n}\sqrt{\pi}} \left(\frac{n}{2}\right)^{1/2} = \frac{1}{\sqrt{2\pi}}$$

CamScanner

Again,

$$\lim_{n\to\infty}\left\{\frac{1}{\sqrt{n}\,B\left(\frac{1}{2},\frac{n}{2}\right)}\right\}=\frac{1}{\sqrt{2\pi}}$$

Also,

$$\lim_{n\to\infty}\left\{\left(1+\frac{t^2}{n}\right)^{-\left(\frac{n+1}{2}\right)}\right\}$$

$$=\lim_{n\to\infty}\left(1+\frac{t^2}{n}\right)^{-\frac{n}{2}}\cdot\lim_{n\to\infty}\left(1+\frac{t^2}{n}\right)^{-1/2}$$

$$=e^{-t^2/2}\cdot 1$$

Now,

$$\lim_{n\to\infty}\left(1+\frac{t^2}{n}\right)^{-\frac{n+1}{2}}=e^{-t^2/2}$$

Hence,

$$\lim_{n\to\infty}f(t)=\frac{1}{\sqrt{2\pi}}e^{-t^2/2};\ -\infty<t<\infty$$

which is pdf of normal distribution. Therefore, for large degree of freedom t-distribution tends to normal distn.

Hypothesis: Any statement about any phenomenon is termed as hypothesis.

Statical hypothesis: Statical hypothesis is a statement about population characteric that can be tested on the basis of sample data.

Null hypothesis: The hypothesis that is formulated for its possible rejection using sample data is called null hypothesis.

Alternative hypothesis: The hypothesis which is true if null hypothesis is false is called alternative hypothesis. Alternative hypothesis indicates the type of test (left, right or two trailed.

Simple hypothesis: The hypothesis, which completely specifies all the parameters of related population is called simple hypothesis.

Composite hypothesis: The hypothesis which does not completely specify all the parameteres of a related population is called composite hypothesis.

Errors in decision making:

Decision	State of nature	
	H_0 is true	H_0 is false
Reject H_0	Type 1 error	Correct desn.
Fail to Reject H_0	Correct decision	Type II error

Type I error:

previous —

Interpretation of level of significance:

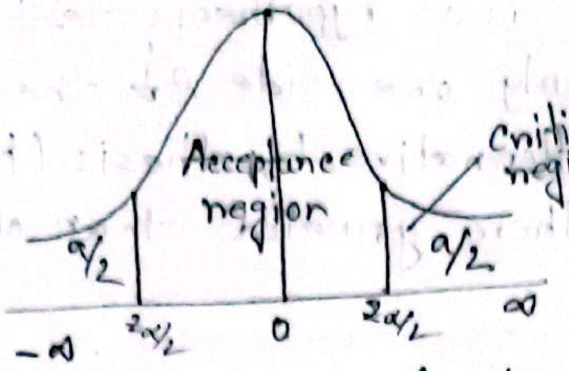
Generally, a significance of level of 0.05 or 0.10 is considered, although other values are also used. Thus if 0.05 is the level of signf, it will mean that there are about 5 samples out of 100 that should direct reject the hypothesis when it should be actually accepted. So $(1-0.05) = 0.95$ is the probability of accepting null hypothesis when it is true, i.e there are 95% confidence in taking right decision.

One-tailed test: One tailed test is a hypothesis test where the rejection region is only one side of the distribution. It is used when alternative hypothesis (H_1) states that a parameter is either greater than or less than a specific value.

A left tailed test: When the rejection region is in the left of tail of the distribution of the test statistice, the test is called left-tailed test. If null hypothesis is $H_0: \mu_0 = 0$, then the alternative hypothesis will be $H_1: \mu < 0$.

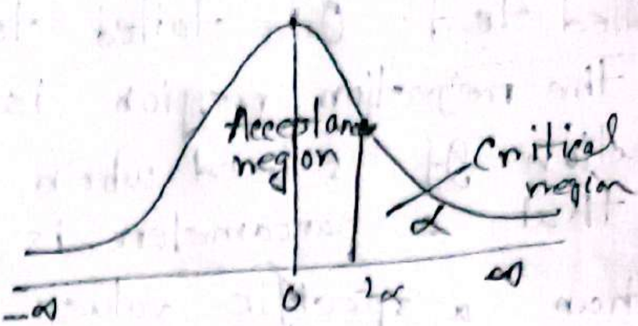
A right tailed test: When the rejection region is in the right tail of the distribution of the test statistic, the test is called a right tailed test. If the null hypothesis is $H_0: \mu_0 = 0$ then the alternative hypothesis will be $H_1: \mu > 0$

A two tailed test: When the rejection region is equally devided in the left and right tails of the distribution of the test statistic, the test is called a two tailed test. If the null hypothesis is $H_0: \mu_0 = 0$ then the two-sided alternative hypothesis is defind by $H_1: \mu \neq 0$.



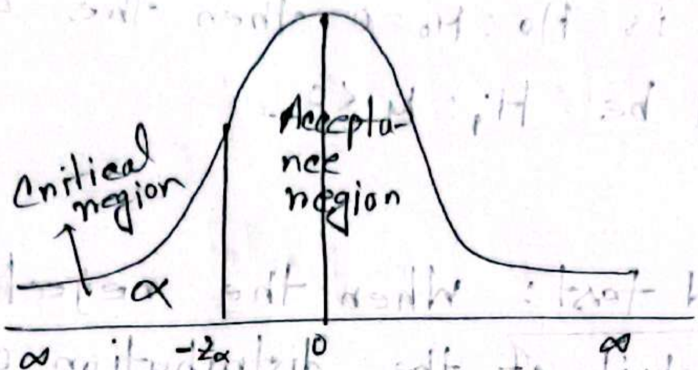
Critical region for two tailed test size α

$H_0: \mu_0 = 0$, $H_1: \mu \neq 0$



Critical region for right tailed test of size α

$H_0: \mu = 0$, $H_1: \mu > 0$



Critical region for left-tailed test of size α

$H_0: \mu_0 = 0$, $H_1: \mu < 0$

Critical region: Critical region is the range of values for the test statistic that leads to rejecting null hypothesis.

Acceptance region: Acceptance region the range of values for test statistic ~~that leads to~~ where null hypothesis is not rejected.

Test statistic: The statistic which is used to provide the ~~ed~~ evidence about rejection or acceptance of null hypothesis, is called test statistic.

Critical value: The value of test sample statistic that separates acceptance region and rejection region is called critical value.

Critical values for ~~test~~ z statistic for 5% level of significance:

right-tailed test : 1.645
left-tailed test : -1.645
two-tailed test : ± 2.33

Test of hypothesis: Hypothesis testing is a statical method used to provide evidence about rejection or acceptance of null hypothesis. (Also called test of signifance.

□ steps of hyothesis testing:–

step –1: set up suitable level of significance

↓

Computation of text statiste

↓

Determine the critical region on rejection region

↓

Calculate text statistic.

↓

taking decision

↙ Reject H_0 and accept H_1

↘ Accept H_0 and reject H_1.

□ if computed value < critical value then accept H_0

" " " > " " " reject H_0

Some important test of significance :-

(i) Normal test / z-test

(ii) t-test

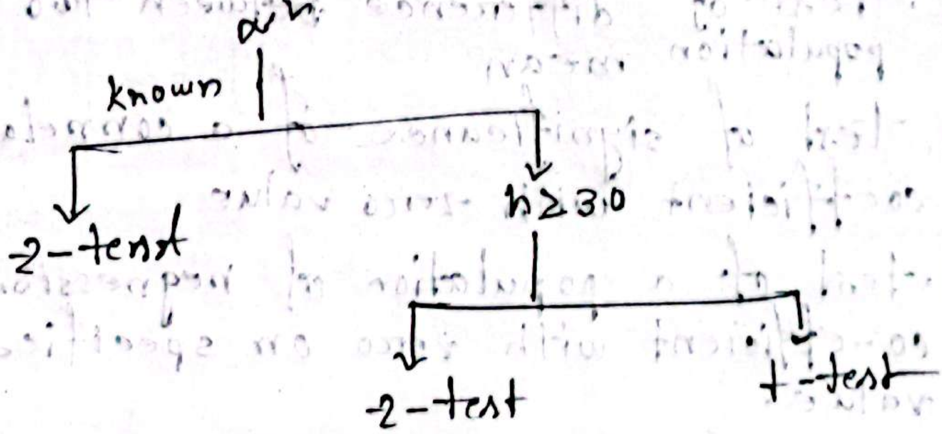
(iii) χ^2 test

(iv) F-test

<u>z-test</u> : z-test is a statistical test used to determine there is a significant difference between sample and population means or between two sample means, when the population variance is known and the sample size is large (typically, $n \geq 30$)

(a) σ^2 is known,
$$z = \frac{\bar{x}-\mu}{\sigma/\sqrt{n}}$$

$\sigma^2 \rightarrow$ population variance.



(ii) σ^2 is not known $(n \geq 30)$

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

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

$$s^2 = \frac{1}{n-1} \sum (x - \bar{x})^2$$

Applications of z-statistic:

1. test of single population mean
2. test of equality of two population means
3. test of single population proportion
4. test for difference between two population proportions.
5. test of specified correlation coefficient.
6. test of equality of two correlation coefficient.

Applications of t-test:

1. test of single population mean
2. test of difference between two population mean
3. test of significance of a correlation coefficient with zero value
4. test of a population of regression co-efficient with zero or specified value.
5. test of difference between two population regression coefficient.

Applications of χ^2 distribution:

(I) The Test of population varriance with specific value.

(II) Test of equality of several variance.

(III) Test of equality of several correlation coefficient

(IV) test of equality of several population proportions

(V) test of independence of attributes

(VI) test of goodness of fit.

Application of F-statistic:—

(I) ~~Th~~ Test of significance of diffence between two population variance

(II) Test of significance of several population means

(III) Test of significance of two or more regression coefficient.

Differences between type-I error and type-II error.

(I) Type-I error occures when the null hypothesis H_0 is rejected though it is true	(I) Type-II error occurs when the null hypothesis is not rejected although it is false
(II) Probability of type-(I) error is denoted by α	(II) Prob. — by β

(III) P(~~type-I~~ reject error | Ho is true) = α

(III) ~~P(type-II error | Ho is false~~

~~(III) reject~~

(III) P(reject Ho | Ho is true)

(IV) probability of type-I error is called level of significance

(V) Determines the critical value of any ~~test~~ statistic

(IV) P(accept Ho | Ho is false) = β

(IV) not called

(V) not used

Computed value < Critical value; accept null hypothesis

Computed value > Critical value; reject

Null hypothesis V Alternative Hypothesis

Null hypothesis	Alternative hypothesis
(i) The hypothesis that is formulated for its possible rejection is called null hypothesis	(i) The hypothesis which is true if null hypothesis is false is called alternative hypothesis
(ii) It is denoted by H_0	(ii) It is denoted by H_1
(iii) It is the complement of Alternative hypothesis	(iii) It is the complement of null hypothesis
(iv) No difference between a sample estimate and true population value	(iv) there is difference.

Problem -

A sample of 100 manager is found to have a mean age of 25 years. Can it be reasonably regarded as a ~~Can it~~ a sample from a large population of mean as 26.8 years and standard deviation of 1.5 years.

Solⁿ:

Given,

$n = 400$ | null hypothesis, $H_0 : \mu = 26.8$

$s = 1.5$ | Alternative " , $H_1 : \mu \neq 26.8$

$\mu = 26.8$

$\bar{x} = 25$

$$z = \frac{\bar{x} - \mu}{s/\sqrt{n}} = \frac{25 - 26.8}{1.5/\sqrt{400}} = -24$$

The calculated value $|z| = 24$

The critical value of z at $\alpha = 0.05$ level of significance is 1.96, which is less than calculated value, So, the null hypothesis is rejected.

Problem: 2

$n = 300$

$\bar{x} = 16.0$

$\mu = 16.8$

$s = 5.2$

null hypothesis,

$H_0 : \mu = 16.8$

$H : \mu_1 \neq 16.8$

$$z = \frac{\bar{x} - \mu}{s/\sqrt{n}} = \frac{16.0 - 16.8}{5.2/\sqrt{300}} = -2.665$$

calculated value of z is $|z| = 2.665$

The critical value of z at $\alpha = 0.05$ is 1.96. As $2.665 > 1.96$ Hence ~~H_0~~ null hypothesis is rejected.

4 A sample of 100 students gave a mean weight of 58kg. with S.D of 4 kg. Test the hypothesis that the mean weight in the population is 60 kg.

Soln:

$H_0 : \mu_0 = 60\,kg$

$H_1 : \mu_1 \neq 60\,kg$

$S.D = 4$

$n = 100$

$\bar{x} = 58$

$\mu = 60$

$$z = \frac{\bar{x} - \mu}{S.D/\sqrt{n}} = \frac{58 - 60}{4/\sqrt{100}} = -5$$

$|z| = 5 > 1.96$

5% — critical value — 1.96

So, null hypothesis is rejected.

Test of hypothesis concerning two population Means:

Suppose $x_{11}, x_{12}, x_{13} \ldots x_{1n_1}$ be random sample of size n_1 drawn from normal distribution population with mean μ_1 and variance σ_1^2 and $x_{21}, x_{22}, x_{23} \ldots x_{2n_2}$ be another sample of size n_2 drawn from normal population with mean μ_2 and variance σ_2^2. Suppose, the observed sample means are $\bar{x}_1$ and $\bar{x}_2$. In earlier chapter, it is mentioned that the distribution of the difference between two sample means follows normal distribution when variance are known for all possible sample sizes, that means,

$$z = \frac{(\bar{x}_1 - \bar{x}_2) - (\mu_1 - \mu_2)}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}} \sim (N(0,1))$$

Situations:

(1) Independent samples with known population variances, sample sizes are large or small.

$$z = \frac{(\bar{x}_1 - \bar{x}_2) - (\mu_1 - \mu_2)}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}}$$

(2) Independent samples with unknown population variances, sample sizes are large.

$$s_1^2 = \frac{\sum (x_1 - \bar{x})^2}{n_1 - 1}$$

$$z = \frac{(\bar{x}_1 - \bar{x}_2) - (\mu_1 - \mu_2)}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}}$$

$$= \frac{1}{n_1 - 1}\left(\sum x_i^2 + \frac{(\sum x_1)^2}{n}\right)$$

Similary for s_2^2

(iii) Independent populations for small sample sizes (≤30) with unknown but equal variances.

$$t = \frac{(\bar{x}_1 - \bar{x}_2) - (\mu_1 - \mu_2)}{s\sqrt{1/n_1 + 1/n_2}}$$

⊡ It is wanted to investigate that if male and female typists earn comparable wages. The sample data for daily wages for male and female is given below:-

	male	female
Sample size	60	60
Mean wage	158.50 tk	141.60 tk
SD (population)	18.20 tk	20.60 tk

Test weather the mean wages of male and female typists at 5% and 1% level of significance.

Soln:

Let the wages of male and female are independently distributed with means μ_1 & μ_2 and known variances σ_1^2 and σ_2^2 respectively. It is one-tailed test, so we consider the following hypothesis.

$H_0: \mu_1 = \mu_2$ against $H_1: \mu_1 > \mu_2$

As the sample sizes are large, under null hypothesis, the value of test statistic is given by

$$Z = \frac{(\bar{x}_1 - \bar{x}_2)}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}}$$

For $\alpha = 0.05$, the critical region $Z > 1.645$

" $\alpha = 0.01$, " " " $Z > 2.33$

$n_1 = 60$, $n_2 = 60$

$\bar{x}_1 = 158.50$, $\bar{x}_2 = 141.60$

$s_1^2 = 18.20$, $s_2 = 20.60$

Thus for value of z is, $z = \dfrac{(158.50 - 141.60)}{\sqrt{\frac{(18.20)^2}{60} + \frac{(20.60)^2}{60}}}$

$= 4.76$

Conclusion: The computed value of z is greater than critical values at both level of significance. Hence null hypothesis is rejected.

[5] A potential by buyerer of electric bulbs bought 100 bulbs each of two famous brands A & B. Upon testing both these sample, he found that breand A had mean life of 1500 hours with standared daviation 50 hours whereas brand B had average life of 1580 hours with standared daviation of 60 hours. Can it be concluded at 5% level of significance that the bulbs of two breands significantly in quality.

Solution:

~~We assume that the parrent~~ population

$H_0: \mu_1 = \mu_2$ against,

$H_1: \mu_1 \neq \mu_2$

Since sample sizes are large, so under the null hypothesis, the value of test statistic is

$$z = \frac{(\bar{x}_1 - \bar{x}_2)}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}}$$

$$= \frac{1500 - 1530}{\sqrt{\frac{50^2}{100} + \frac{60^2}{100}}} = -3.841$$

at $\alpha = 0.05$ $|z| > 1.96$. Hence rejects null hypothesis.

Test of hypothesis about a population proportion:-

Hypothesis testing for a population proportion is a statical method used to determine wether the proportion of certain characteristic in the population is equal to specific value based on sample data.

$$z = \frac{p - \pi_0}{\sqrt{\frac{\pi_0(1-\pi_0)}{n}}}$$

p – sample proportion
n → " size

Example:

For the following questions carry out the test of significance of population at 5% level of significance.

(I) $H_0: \pi = 0.25$, $H_1: \pi \neq 0.25$, $n = 100$, $x = 40$

(II) $H_0: \pi = 0.40$, $H_1: \pi > 0.40$, $n = 200$, $x = 100$

Soln:

The statistic to be used for testing the given hypothesis is given by,

$$z = \frac{p - \pi_0}{\sqrt{\frac{\pi_0(1-\pi_0)}{n}}}$$

(I) $p = \frac{x}{n} = \frac{10}{100} = 0.1$

$H_0: \pi = 0.25$, $H_1: \pi \neq 0.25$

$n = 100$

$\pi_0 = 0.25$

$$z = \frac{0.1 - 0.25}{\sqrt{\frac{0.25(1-0.25)}{100}}}$$

$= 3.46 > 1.96$

Hence will reject null hypothesis.

(II) $H_0: \pi = 0.40$, $H_1: \pi \neq 0.40$

$n = 200$

$p = \frac{100}{200} = 0.5$

$$z = \frac{0.5 - 0.4}{\sqrt{\frac{0.4(1-0.4)}{200}}}$$

$= 2.89 > 1.645$

Hence rejects null hypothesis.

CamScanner

Parametric test:

A parametric test is a statical test that makes certain assumptions about the population parameters (like mean, standard daviation etc) and the distribution of the data. Example: z-test, t-test.

Non-parametric test:

A non parametric test is a statical test that doesn't not assume any specific distribution of the population and is often used when data doesn't meet the assumption of parametric tests.
Example: chi-square test, ~~the~~ sign test etc.

Advantages:

1. They can be used to test population parameters when the variable is not normally distributed.

2. They can be used when data are nominal or ordinal.

3. They are easy to understand

4. They can be used to test hypothesis that do not involve population parameters.

5. In some cases, the computations are easier than those for the parametric counterparts.

CamScanner

disadvantages:

There are three disadvantages:—

1. They are less sensitive than their parametric counterparts when the assumptions of the parametric methods met.

2. They tend to use less information than the parametric test.

3. They are less efficient than ...... met.

Sign test:

The sign test is a nonparametric test used to determine whether there is significant difference between the median of a sample and a hypothesized value, or to compare the medians of two related samples.

Example:

A convenience store owner hypothesizes that the median number of ~~show~~ snow cones she sells perday is 40. A random sample of 20 days yields the following data for the number of snow cones sold ~~each~~ each day.

18	13	40	16	22
30	29	32	37	36
39	34	39	45	28
36	40	31	39	52

at $\alpha = 0.05$, test the owner's hypothesis.

Sol^n:

step-1: state the hypothesis and identify the claim.

H_0: median $= 40$ (claim)

H_1: median $\neq 40$

step-2: Find the critical value. Compare each value of the data with the median. If the value is greater than the median, replace the value with a plus sign. If it is less than the median, replace it with minus sign. And if it is equal to the median, replace it with a 0. The completed table follows:-

−	+	0	−	
−	−	−	−	
−	−	−	+	−
−	0	−	−	+

here, $n = 18$ and $\alpha = 0.05$. for the two tailed test, the critical value is 4.

CS CamScanner

step-3: Compare the test value. Count the numbers of plus and minus signs obtained in step-2, and use smaller value as test value. Since there are three 3 plus signs and 15 minus signs. 3 is the test value

step-4: Make the decision. Compare the test value 3 with the critical value 4. If the test value is less than or equal to critical value, the null hypothesis is rejected. In this case, the null hypothesis is rejected since $3<4$

step-5: summarize the results. There is enough evidence to reject the claim that the median number of snow cones sold per day is 40.

Formula for z test value in the sign test when $n \geq 26$.

$$z = \frac{(x+0.5) - n/2}{\sqrt{n}/2}$$

Based on information from the U.S census Bureau, the median age of foreign-born U.S. residents is ~~his area and~~ finds 36.4 years. A researcher selects a sample of 50 foreign-born U.S. residents in his area and finds that 21 are older than 36.4 years. At $\alpha = 0.05$, test the claim that the median age of the residents is at least 36.4 y.

Soln:

Step-1: state the hypotheses and identify the claim.

$H_0 : MD = 36.4$ (claim) and $H_1 : MD < 36.4$

Step-2: Find the critical value. Since $\alpha = 0.05$ and $n = 50$, and since this is a left tailed test, the critical value is -1.65, obtained from table E.

Step-3: Compute the test value:

$$z = \frac{(x+0.5) - n/2}{\sqrt{n}/2} = \frac{(21+0.5) - 50/2}{\sqrt{50}/2}$$

$$= -3.5/3.5355$$

$$= -0.99 > -1.65$$

Step-4: Make decision. do not reject null hypotheses.

Step-5: There is enough evidence to accept null hypotheses.

CS CamScanner

A medical researcher believed the number of ear infections in swimmers can be reduced if the swimmers use earplugs. A sample of 10 people was selected and the number of infections for a four month period was recorded. During the first two months, the swimmers did not use the earplugs; during the second two months they did.

$\alpha = 0.05$

Number of ear infections

Swimmer	Before X_B	After X_B
A	3	2
B	0	1
C	5	4
D	4	0
E	2	1
F	4	3
G	3	1
H	5	3
I	2	2
J	1	3

Sol^n:

Step-1: State the hypotheses and identify the claim.

~~step-2~~ H_0: The number of ear infections will not be reduced.

H_1: It will be reduced.

Step-2: Find the critical value. & Subtract the after values of X_A from the before values of X_B and indicate the difference by a positive or negative sign or 0, according to the value, as shown in the table.

Swimmers	Before X_B	After X_A	Sign of diff.
A	3	2	+
B	0	1	−
C	5	4	+
D	4	0	+
E	2	1	+
F	4	3	+
G	3	1	+
H	5	3	+
I	2	2	0
J	1	3	−

Here $n = 9$ & $\alpha = 0.05$ (one-tailed). Hence critical value = 1.

step-3: Compute the test value. Count the number of possible possitive and negative signs found in step-2, and use the smaller values as the test value. There are 2 negative signs, so the H-test value is 2.

step-4: Make decisions. There are 2 negative signs. The decision is not to reject null hypo. ~~As~~ since 2>1.

step-5: There is not enough evidence to support the claim that the use of earplugs reduced the number of infections.

The wilcoxon rank sum test:

The Wilcoxon Rank Sum test is a non-parametric statical test used to compare two independent groups to determine weather there is a significant difference between their distributions. If the null hypotheses is true, means there is no difference in the population distributions.

Formula for the Wilcoxon Rank Sum Test when samples are independent:

$$z = \frac{R - \mu_R}{\sigma_R}$$

where,

$$\mu_R = \frac{n_1(n_1 + n_2 + 1)}{2}$$

$$\sigma_R = \sqrt{\frac{n_1 n_2 (n_1 + n_2 + 1)}{12}}$$

R = sum of ranks for smaller sample size (n_1)

n_1 = smaller of sample sizes

n_2 = ~~smaller~~ larger . . . sizes

$n_1 \geq 10$ & $n_2 \geq 10$

Example:

Two independent samples of army and recuits are selected, and the time in minutes it takes each recuit to complete an obstacle course is recorde, as shown in the table. At $\alpha = 0.05$, is there a difference in the times it takes the recuits to complete the course.

Mean = 19.67

Army	15	18	16	17	13	22	24	17	19	21	26	28
Marines	14	9	16	19	10	12	11	8	15	18	25	

Mean = 14.27

step-1:
state the null hypothesis.

H_0: There is no differrence in the times it takes the recruits to complete the obstacle cource.

H_1: There is a differrence

step-2: Find the critical value. Since $\alpha = 0.05$ and this test is two tailed test, use the values of $+1.96$ to -1.96

step-3: Compute the test value. ~~Since $\alpha=0.05$ and this test is a two tailed test,~~

a. Combine the data from the two samples, arrange the combined data in order, and rank each value. Be sure to indicate the group.

Time :

Time	Group	Rank
8	M	1
9	M	2
10	M	3
11	M	4
12	M	5
13	A	6
14	M	7
15	A	$8 \rightarrow 8.5$
15	M	$9 \rightarrow 8.5$
16	A	10.5
16	M	10.5
17	A	12.5
17	A	12.5
18	M	14.5
18	A	14.5
19	A	16.5
19	M	16.5

time	Group	Rank
21	A	~~18~~
22	A	~~20~~ 19
24	A	20
25	M	21
26	A	22
28	A	23

b. sum of the Ranks of the group with the smaller size.

$$R = 1 + 2 + 3 + 4 + 5 + 7 + 8.5 + 10.5 + 14.5 + 16.5 + 21 = 93$$

c. Substitute in the formulas.

$$\mu_R = \frac{n_1(n_1 + n_2 + 1)}{2} = \frac{11(11 + 12 + 1)}{2} = 132$$

$$\sigma_R = \sqrt{\frac{n_1 n_2 (n_1 + n_2 + 1)}{12}} = \sqrt{\frac{(11)(12)(11 + 12 + 1)}{12}}$$

$$= \sqrt{264} = 16.2$$

$$z = \frac{R - \mu_R}{\sigma_R} = \frac{93 - 132}{16.2} = -2.41$$

step-4: Make decision. The decision is to reject null hypothesis, since $-2.41 < -1.96$.

step-5: summarize the results. There is enough evidence to support the claim that there is a difference in the times it takes the recruits to complete the course.

The run test:

The run test is a non-parametric statical test used to check whether a sequence of data points is random or if there is any pattern or trend present.

run:

A run is a succession of identical letters preceded or followed by a different letter or no letter at all, such as the beginning or end of the succession.

[7] On a commuter train, the conductor wishes to see whether the passengers enter the train at random. He observes the first 25 people, with the following sequence.

FFFMMFFFFMFMMFFFFMMFFFMM

Test for randomness at $\alpha = 0.05$.

Solution:

step-1: State the hypothesis and identify the claim.

H_0: The passengers board the train at random, according to their gender.

H_1: null hypothesis is not true.

step-2: Find the number of runs. Arrange the letters according to runs of males and females, as shown:

Run	Gender
1	FFF
2	MM
3	FFFF
4	M
5	F
6	MM

Run	Gender
7	F F F F
8	M M
9	F F F
10	M M

There are 15 female and 10 male.

so, $n_1 = 15$ & n_2 10

step-3: Find the critical value. ~~Sin~~ As $n_1 = 15$, $n_2 = 10$ ~~so~~ and $\alpha = 0.05$ so, critical values are 7 and 18

step-4: Make decision. As 10 is between 7 and 18, so it do not reject null hypothesis.

step-5: Summarize the results: There is enough evidence to reject the hypothesis that the passengers board the train at random according to their gender.

Population:

The collection of all units of a specific type in a given region at a particular time is termed as a population or universe.

Example: A population of Rajshahi University students, a population of books in a library, a population of tree in a certain region or country.

Question:

Explain the concept of estimation with example

Answer:

Estimation is a process of finding an estimation or approximation, which is a value that is usable for some purpose even if input data maybe incomplete, ~~ur~~ uncertain or unusable.

Let x be a random variable which represent some characteristics of the elements in a population whose density function is assumed $f(x|\theta)$ θ is unknown parameter.

Again let the values $x_1, x_2, \ldots x_n$ of a random sample $X_1, X_2 \ldots X_n$ from $f(x|\theta)$ can be observed sample values $x_1, x_2, \ldots x_n$. it is desired to estimate the value of unknown parameter θ on $\tau(\theta)$ (sample function of θ)

Estimation can be devided into two types:—

(i) point estimation

(ii) Interval estimation

for example if $f(x|\theta)$ is the normal density function that is

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

where the parameter θ is (μ, σ^2) and if it desired to estimate the mean that is $\tau(\theta) = \mu$ then the statistic $\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i$ is the possible point estimator on of $\tau(\theta) = \mu$.

Estimator:

Any function of a random sample $x_1, x_2, \ldots, x_n$ that are being used / observed say $T_n(x_1, x_2, \ldots, x_n)$ is called a statistic. If it is used to estimate the unknown parameter θ of the distribution is called an estimator.

Example: sample mean $\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i$, sample variance $s^2 = (n-1)^{-1}\sum_{i=1}^{n}(x_i - \bar{x})$ are estimator of the population μ and population variance σ^2 respectively

Estimate:

A particular value of a estimation is called estimate.

Example, the sample mean $\bar{x} = 5.6$ (say) is an estimate value of the estimaton.

point estimation:

A point estimate is a single number that is used to estimate an unknown population parameter.

Another definition, suppose $(x_1, x_2 \dots x_n)$ is a sample from ~~normal~~ a density $f(x|\theta)$ where θ is unknown fixed value, which can assume any value in one dimentional real parameter space Ω. Let t be a function of $x_1, x_2 \dots x_n$ so that t is a statistic and hence a random variable. If t is used to estimate θ then t is called point estimaton of θ. If the realized value of t from a sample is used for θ then t is called a point estimate of θ.

for examples if $f(x|\theta)$ is a normal density function that is $f(x|\theta) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left[-\frac{1}{2}\left(x-\mu/\sigma\right)^2\right]$ where the parameter θ is (μ, σ^2) and it ~~is called~~ desired to estimate the mean, that is $\tau(\theta)=\mu$ then the statistic $\bar{x} = n^{-1}\sum_{i=1}^{n} x_i$ is the possible point estimator of $\tau(\theta) = \mu$.

Interval Estimation: The interval estimation is to define two statistic say $t_1(x_1, x_2 \dots x_n)$ and $t_2(x_1, x_2 \dots x_n)$ so that $\{t_1(x_1, x_2 \dots x_n), t_2(x_1, x_2 \dots x_n)\}$ constitutes an interval for which the probability can be determined that is contains the unknown $\tau(\theta)$.

For example, if $f(x|\theta)$ is the normal density that is $f(x) = (\sigma\sqrt{2\pi})^{-1} \exp(-\frac{1}{2}(x-\mu/\sigma)^2]$ where the parameter θ is (μ, σ^2) and if it is desired to ~~determine~~ estimate the mean is $\tau(\theta) = \mu$. Then the statistic $\bar{x} = n^{-1}\sum_{i=1}^{n} x_i$ is a possible point estimation of $\tau(\theta) = \mu$ and $(\bar{x} - 2\sqrt{\sigma^2/n}, \bar{x} + 2\sqrt{s^2/n})$ is a possible interval estimator $\tau(\theta) = \mu$ where $s^2 = (n-1)^{-1}\sum_{i=1}^{n}(x_i - \bar{x})^2$

Properties of good estimator:

(i) Unbiasedness
(ii) Consistency
(iii) Efficiency
(iv) Sufficiency

Unbiasedness:

Any statistic where mathematical expectation is equal to a parameter θ is called an unbiased estimator of the parameter θ. Otherwise the statistic is called biased.

Let t_n be a statistic calculated from a sample $(x_1, x_2 \ldots x_n)$ of size n from density $f(x|\theta)$. If for all n and θ $E(t_n) = \theta$, then t_n is called called an unbiased estimator of θ.

In case t_n be a biased estimator the difference $E(t) - \theta$ is the amount of bias and $E\{t_n - \theta\}^2$ is called is called mean square error. Mean square error of t_n = variance of t_n + bias^2

For example, if a random sample $(x_1, x_2 \ldots x_n)$ of size n is drawn from a normal distribution on population with mean θ and variance σ^2 then,

$$E(\bar{x}) = \frac{1}{n} E(x_1 + x_2 + x_3 \ldots x_n)$$

$$= \frac{1}{n} \{E(x_1) + E(x_2) \ldots E(x_n)\}$$

$$\Rightarrow \frac{1}{n} n\theta$$

$$\Rightarrow \theta$$

And, $$E(s^2) = \frac{1}{n-1} E\left[\sum (x_i - \bar{x})^2\right]$$

$$= \frac{\sigma^2}{n-1} \cdot E\left[\frac{\sum (x_i - \bar{x})^2}{\sigma^2}\right]$$

$$E(s^2) = \frac{\sigma^2}{n-1} E\left[\sum (x_i^2)_{n-1}\right]$$

$$= \sigma^2/_{n-1} (n-1)$$

$\Rightarrow E(s^2) = \sigma^2$

Thus $\bar{x}$ and s^2 are an unbiased estimator.

Consistency:

Let t_n be a statistic calculated from a sample $(x_1, x_2 \dots x_n)$ of size n from density $f(x|\theta)$.

If $\lim_{n\to\infty} P[|t_n - \theta| < \varepsilon] = 1 - \gamma$

where ε and γ are arbitary small positive numbers then t_n is called consistent estimator of θ.

Consistency is a large sample property. It is not defined for a small sample. A statistic is said to be consistent estimator of the population parameter if it approaches the parameter as the sample size increases

For example, if $x_1, x_2 \dots x_n$ is a random sample from a population with finite mean $E(x_1) = \mu < \infty$.

Now we have,

$$\bar{x} = n^{-1} \sum_{i=1}^{n} x_i$$

$$E(\bar{x}) = \mu \quad \text{as} \quad n \longrightarrow \infty$$

Hence sample mean is always consistent estimator of the population mean μ.

Efficiency:

If $(x_1, x_2 \dots x_n)$ be a sample from density $f(x|\theta)$ and t be an unbiased consistent estimator of θ and further no other estimators have variance less than that t, then t is said to be the most efficient estimator of θ.

Let t^* be any other unbiased statistic. The efficiency of t^* is the ratio of reciprocal of the variance of t^* to the amount of information in the data. Actually the efficiency of t^* is measured by,

$$e(t^*) = \frac{V(t)}{V(t^*)} \quad \text{--- (1)}$$

The efficiency of t^* ~~measures~~ represents fraction of the relavent information available actually utilized by t^*. since $V(t) \leq V(t^*)$ Since $V(t) \leq V(t^*)$ the efficiency of any statistic between 0 to 1.

For example, ~~to~~ let $x \sim N(\mu, \sigma^2)$ and $x_1, x_2 \dots x_3$ be random sample then

$$T_1 = \frac{x_1 + x_2 + x_3}{3} \sim N(\mu, \sigma^2/3)$$

$$T_2 = \frac{1}{2}(x_1 + x_2) \sim N(\mu, \sigma^2/2)$$

Here both T_1 and T_2 are unbiased estimators of μ. But $var(T_1) < var(T_2)$ implies that T_1 is more efficient than T_2.

sufficiency:

An estimator is said to be sufficient for a parameter if it contains all the information in the sample regarding the parameter.

If $T = f(x_1, x_2 \dots x_n)$ is an estimator of a parameter θ, based on a sample $x_1, x_2 \dots x_n$ of size n from the population with density $f(x|\theta)$ such that the conditional distribution of $x_1, x_2 \dots x_n$ given T, is independent of θ, then T is sufficient estimator of θ.

For example, $x \sim B(n, \theta)$

sample : $x_1, x_2 \dots x_n$

$\therefore f(x) = \binom{n}{x} \theta^x (1-\theta)^{n-x}$

Now, $P(x) = \theta^{\Sigma x_i} (1-\theta)^{Nn - \Sigma x_i}$ [Bernoulli distn]

$\therefore p(\Sigma x_i) = \binom{n}{\Sigma x_i} \theta^{\Sigma x_i} (1-\theta)^{Nn - \Sigma x_i}$

Now, $\frac{P(x)}{p(\Sigma x_i)} = 1/\binom{n}{\Sigma x_i}$ which is independent of θ.

So, Σx_i is a sufficient estimator of θ.

Question: What do you mean by maximum likelihood func^n

Ans:

Let $x_1, x_2 \dots x_n$ be a random sample of size n from a population with density func^n $f(x|\theta)$. If the joint pdf may be regarded as a function of θ. It is called the likelihood function denoted by $L(\theta)$, defind as

$$L = f(x_1;\theta)\, f(x_2;\theta_2) \dots f(x_n;\theta)$$

$$= \prod_{i=1}^{n} f(x_i;\theta)$$

Example: let $x_1, x_2 \dots x_n$ be a random sample from a distribution with p.d.f

$$f(x;\theta) = \theta^x (1-\theta)^{1-x} \quad ; \quad x = 0, 1; \quad 0 < \theta < 1$$

The joint pdf is

$$f(x_1, x_2 \dots x_n;\theta) = \theta^{\sum_{i=1}^{n} x_i} (1-\theta)^{n-\sum_{i=1}^{n} x_i}$$

If the joint pdf is a function of θ, then it is called Likelihood function, denoted by $L(\theta)$, defind as,

$$L(\theta) = \theta^{\sum_{i=1}^{n} x_i} (1-\theta)^{n-\sum_{i=1}^{n} x_i}$$

Question:
Define maximum likelihood estimator with an example.

Ans:
Maximum likelihood estimator: Let $x_1, x_2 \dots x_n$; $L(\theta) = \prod_{i=1}^{n} f(x_i;\theta)$ be the likelihood function for the random variable $x_1, x_2, \dots, x_n$. The the value of θ ~~wil~~ which maximize the likelihood function that the value of θ is called maximum likelihood estimator or MLE of θ. It is usually denoted by $\hat{\theta}$.

The MLE of θ is the solution of likelihood equation

$$\frac{\delta L(\theta)}{\delta \theta} = 0 \quad ; \quad \frac{\delta \log L(\theta)}{\delta \theta} = 0$$

If $\hat{\theta}$ is the MLE of θ. Then $\frac{\delta L(\hat{\theta})}{\delta \hat{\theta}} = 0$

and $\frac{\delta^2 \log L(\theta)}{\delta \theta^2} < 0$

Example:
Let $x_1, x_2 \dots x_n$ be random sample drawn from a normal distribution with μ and variance 1.

Then the pdf,

$$f(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}(x-\mu)^2}$$

Now the likelihood function is,

$$L(\theta) = \left(\frac{1}{2\pi}\right)^{n/2} e^{-\frac{1}{2}\sum(x_i-\theta)^2}$$

$$\Rightarrow \log L(\theta) = n/2 \log(1/2\pi) - 1/2 \sum(x_i-\theta)^2$$

Now,

$$\frac{d\log L(\theta)}{d\theta} = 0$$

$$\Rightarrow \sum(x_i-\theta) = 0$$

$$\Rightarrow \hat{\theta} = \bar{x}$$

which maximize the likelihood function.
Hence $\hat{\theta} = \bar{x}$ is the MLE of θ.

Question:

Describe the principal of maximum likelihood estimation.

Answer:

The principal of maximum likelihood estimation in finding an estimation for the unknown parameter $\theta = (\theta_1, \theta_2 \cdots \theta_n)$ say which maximizes the likelihood function $L(\theta)$ for variations in parameter i.e we wish to find $\hat{\theta} = (\hat{\theta}_1, \hat{\theta}_2, \cdots, \hat{\theta}_n)$ so that

$$L(\hat{\theta}) = \sup (L(\theta)) \quad \forall \theta \in \Omega.$$

Let $x_1, x_2, \ldots x_n$ be a random sample from the density $f(x;\theta)$ for a given sample, this can be treated as a function of θ. This function is called the likelihood function of θ. It is usually denoted by $L(\theta)$ defind as

$$L(\theta) = \prod_{i=1}^{n} f(x_i;\theta)$$

The estimator of θ which maximize the likelihood function is called the MLE of θ. Therefore the solution of MLE is

$$\frac{\delta \log L}{\delta \theta} \leq 0 \text{ given the MLE of } \theta \text{ func}^n.$$

$$\frac{\delta^2 \log L}{\delta \theta^2} < 0$$

☐ Write down properties of MLE:

→ MLE are not generally unbiased but asymptotically unbiased.

→ MLE of θ is consistent under some conditions.

→ MLE of θ is ~~asymp~~ asymtotically efficient.

→ MLE of θ is invariant under functional transformation.

→ MLE of θ is a function of sufficient statistic if exist.

→ If a MLE estimator exist, then the MLE is a MLB estimator.

→ MLE is asymptotically normally distributed as

$$N\left(\theta_0, \frac{1}{E\left(-\frac{\partial^2 \log L(\theta)}{\partial \theta^2}\right)}\right)$$

where θ_0 is the true value of θ.

Question: What are the advantages of MLE method over other method of point estimation.

Answer:

→ MLE methods obtained consistent estimator.

→ MLE methods obtained most efficient estimator.

→ If sufficient statistic exists. MLE method obtained the sufficient estimator

→ MLE has invariant property.

Question:

Is MLE always unbiased?

Ans:

No, MLE is not always unbiased. Because,

Let $X \sim N(\theta_1, \theta_2)$ and $x_1, x_2 \ldots x_n$ be a random sample. Then we know the MLE of θ_1 and θ_2 is

$\theta_1 = \bar{x}$ and ——— (i)

$\theta_2 = \frac{1}{n}\sum_{i=1}^{n}(x_i-\bar{x})^2$ ——— (ii)

Taking expection of (i) on both sides, we have.

$$E(\hat{\theta}_1) = E(\bar{x})$$
$$= E(1/n \sum x_i)$$
$$\Rightarrow 1/n\, E(\sum x_i)$$
$$\Rightarrow 1/n \cdot n\theta_1$$
$$\Rightarrow \theta_1$$
$$\therefore E(\hat{\theta}_1) = \theta_1$$

So, $\hat{\theta}_1$ is an unbiased estimaton of θ_1

Again taking expectation of (ii) on both sides, we have,

$$E(\hat{\theta}_2) = 1/n\, E(\sum(x_i-\bar{x})^2$$
$$= \frac{\sigma^2}{n} E\left(\frac{\sum_{i=1}^{n} x_i-\bar{x}}{\sigma^2}\right)^2$$
$$\Rightarrow \frac{\sigma^2}{n} E(\chi^2_{n-1})$$
$$\Rightarrow \frac{\sigma^2}{n}(n-1)$$
$$E(\hat{\theta}_2) = \frac{n-1}{n}\sigma^2 \neq \sigma^2$$

This means, $\hat{\theta}_2$ is a biased estimate of θ_2.
Hence we say that MLE is not always unbiased.

⊞ Distingush between joint density function & likelihood function.

Joint density $func^n$	Likelihood $func^n$
1. The joint density $func^n$ is a function of sample observations.	1. The likelihood $func^n$ is a $func^n$ of parameters of the $dist^n$.
2. If $(x_1, x_2 \dots x_n)$ be a random sample then the joint density $func^n$ is given by $f(x_1, x_2 \dots x_n)$	2. If $x_1, x_2 \dots x_n$ be a random sample then the likelihood function is given by $L(x_1, x_2 \dots x_n \mid \theta)$ which is $func^n$ of θ
3. The joint density satisfies the property as, $f(x_1, x_2 \dots x_n) \geq 0$; $\int\int \dots \int f(x_1, x_2 \dots x_n)\, dx_1, dx_2 \dots dx_n = 1$	3. It may or may not satisfy the condition $\int_\theta L(\theta \mid x)\, d\theta = 1$
4. It can not be used to estimate MLE	4. It is used to estimate MLE.

CamScanner

problem: obtain the MLE of θ_1 & θ_2 from $N(\theta_1, \theta_2)$ based on random sample $x_1, x_2 \ldots x_n$. also check the unbiasedness of the MLEs.

Solution:

Given that,

$X \sim N(\theta, \theta)$

Then the p.d.f of x is

$$f(x : \theta_1, \theta_2) = \frac{1}{\sqrt{2\pi\theta_2}} e^{-\frac{1}{2\theta_2}\sum_{i=1}^{n}(x_i - \theta)^2} \quad -\infty < x < \infty, \; -\infty < \theta_1 < \infty, \; \theta_2 \geq 0$$

The likelihood function is,

$$L = \prod_{i=1}^{n} f(x_i ; \theta_1, \theta_2)$$

$$= \left(\frac{1}{2\pi\theta_2}\right)^{n/2} e^{-\frac{1}{2\theta_2}\sum_{i=1}^{n}(x_i - \theta_1)^2}$$

$$= \left(\frac{1}{2\pi}\right)^{n/2} (\theta_2)^{-n/2} e^{-1/2\theta_2 \sum_{i=1}^{n}(x_i - \theta_1)^2}$$

Taking log on both sides, we have:

$$\log L = n/2 \log 1/2\pi - n/2 \log \theta_2 - 1/2\theta_2 \sum_{i=1}^{n}(x_i - \theta_1)^2 \quad \text{--- (1)}$$

first derivative of (1) w.r.t. θ_1 and set equal to zero i.e

$$\frac{\delta \log L}{\delta \theta_1} = 0$$

$$\Rightarrow -\frac{1}{2\theta_2}\cdot 2\sum(x_i-\theta_1)(-1)=0$$

$$\Rightarrow \sum(x_i-\theta_1)=0$$

$$\Rightarrow \sum x_i - n\theta_1 = 0$$

$$\Rightarrow n\bar{x}-n\theta_1=0$$

$$\Rightarrow \theta_1=\bar{x}$$

$$\therefore \theta_1=\bar{x}$$

$$\frac{d^2\log L}{d\theta_1^2} = -n/\theta_2$$

$$-E(\ ) = n/\theta_2$$

$$V(\hat{\theta}_1) = \theta_2/n$$

$$\lim_{n\to\infty} V(\hat{\theta}_1) = \theta/n$$

which is the MLE of θ_1

Again, the first derivitive of ① w.r.t. θ_2 we have,

$$\frac{d\log L}{d\theta_2}=0$$

$$\Rightarrow -\frac{n}{2}\cdot\frac{1}{\theta_2}+\frac{1}{2\theta_2^2}\sum(x_i-\hat{\theta}_1)^2=0$$

$$\Rightarrow -n\theta_2+\sum(x_i-\hat{\theta}_1)^2=0$$

$$\Rightarrow n\theta_2=\sum(x_i-\hat{\theta}_1)^2$$

$$\Rightarrow \theta_2=\frac{1}{n}\sum_{n}(x_i-\hat{\theta}_1)^2$$

$$\therefore \hat{\theta}_2=\frac{1}{n}\sum_{i=1}^{n}(x_i-\hat{\theta}_1)^2$$

which is the MLE of θ_2.

Unbiasedness — আগেই আছে - - -

Consistency:

We have,

$$\frac{d\log L}{d\theta} = \frac{n}{\theta} - \Sigma x_i$$

$$\Rightarrow \frac{d^2\log L}{d\theta^2} = -\frac{n}{\theta^2}$$

Now,

$$E\left(-\frac{d^2\log L}{d\theta^2}\right) = \frac{n}{\theta^2}$$

we know,

$$V(\hat{\theta}) = \frac{1}{E\left(-\frac{d^2\log L}{d\theta^2}\right)}$$

$$= \theta^2/n$$

Now, $\lim_{n\to\infty} V(\hat{\theta}) = \lim_{n\to\infty} \theta^2/n$

$$= 0$$

Hence, $\bar{x}$ is also a consistent estimator of θ.

CamScanner

Problem: Let $x_1, x_2 \ldots x_n$ be a random sample with p.d.f is

$$f(x;\theta) = \frac{1}{\theta} e^{-x/\theta} \; ; \; x \geq 0 \; ; \; \theta > 0$$

(I) Find the MLE of θ.

(II) Show that MLE of θ sufficient, unbiased and consistent estimator.

Solution:

(I) Given that;

$$f(x;\theta) = \frac{1}{\theta} e^{-x/\theta} \; ; \; x \geq 0, \; \theta > 0$$

The likelihood function is,

$$L = \prod_{i=1}^{n} f(x_i;\theta)$$

$$= \frac{1}{\theta^n} e^{-\sum x_i/\theta}$$

Taking log on both sides, we have

$$\log L = -n\log\theta - \frac{1}{\theta}\sum x_i \quad \ldots \quad (1)$$

Now differentiating (1) w.r.t θ we have and set equal to zero,

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

$$\Rightarrow -\frac{n}{\theta} + \frac{1}{\theta^2}\sum x_i = 0$$

$$\Rightarrow -\frac{n}{\theta} = \frac{n\bar{x}}{\theta^2} \quad = \frac{\bar{x}}{\theta} = 1$$

$$\Rightarrow \bar{x} = \theta \Rightarrow \hat{\theta} = \bar{x}$$

And,

$$\frac{\partial^2 \log L}{\partial \theta^2 \theta} = -\frac{n}{\theta^2} - \frac{2\Sigma x_i}{\theta^3} < 0$$

which is negative definite.

Hence, $\hat{\theta} = \bar{x}$ is the MLE of θ.

Problem:

let $x_1, x_2 \ldots x_n$ be a random sample with p.d.f

$$f(x;\theta) = \theta^x (1-\theta)^{1-x} \quad ; \; x = 0,1$$

The likelihood function is

$$L = \prod_{i=1}^{n} f(x_i;\theta)$$

$$\Rightarrow \theta^{\Sigma x_i} (1-\theta)^{n-\Sigma x_i}$$

taking log on both sides, we have,

$$\log L = \Sigma x_i \log\theta + (n - \Sigma x_i) \log(1-\theta)$$

Now differentiating w.r.t θ and set equal to zero.

i.e

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

$$\Rightarrow \frac{\Sigma x_i}{\theta} + \frac{n - \Sigma x_i}{(1-\theta)}(-1) = 0$$

$$\Rightarrow \frac{\Sigma x_i}{\theta} - \frac{n - \Sigma x_i}{1-\theta} = 0$$

$$\Rightarrow \frac{\Sigma x_i}{\theta} = \frac{n - \Sigma x_i}{1-\theta}$$

$$\Rightarrow \Sigma x_i - \Sigma x_i \theta = n\theta - \theta \Sigma x_i$$

$$\Rightarrow n\theta - \Sigma x_i = 0$$

$$\Rightarrow n\theta - n\bar{x} = 0$$

$$\Rightarrow \theta = \bar{x} \quad \Rightarrow \hat{\theta} = \bar{x}$$

which is MLE of θ

And,

$$\frac{d^2 \log L}{d\theta^2}\Big|_{\theta=\hat\theta} = -\frac{\sum x_i}{\theta^2} + \frac{n-\sum x_i}{(1-\theta)^2}(-1)$$

$$\Rightarrow -\frac{\sum x_i}{\theta^2} - \frac{n-\sum x_i}{(1-\theta)^2} < 0$$

Hence, $\hat\theta = \bar{x}$ is the MLE of θ.

Unbiasedness:

We have,

$$\hat\theta = \bar{x}$$

$$= \frac{1}{n}\sum x_i$$

$$\Rightarrow E(\hat\theta) = \frac{1}{n} E(\sum x_i)$$

$$= \frac{1}{n} n\theta$$

$$\therefore E(\hat\theta) = \theta$$

Hence, $\hat\theta$ is an unbiased estimator of θ.

Problem:

Let $x_1, x_2, \ldots x_n$ be a random sample with p.d.f.

$$f(x;\theta) = \frac{e^{-\theta}\theta^{x}}{x!} \; ; \; x \geq 0 \; ; \; \theta > 0$$

(I) Find the MLE of θ.

(II) Show that MLE of θ is unbiased.

Solⁿ:

Given,

$$f(x;\theta) = \frac{e^{-\theta}\theta^{x}}{x!}$$

The likelihood function is

$$L = \prod_{i=1}^{n} f(x_i;\theta)$$

$$= \prod_{i=1}^{n} \frac{e^{-\theta}\theta^{x}}{x!}$$

$$\Rightarrow \frac{e^{-\theta n} \cdot \theta^{\Sigma x_i}}{\prod_{i=1}^{n}(x_i)!}$$

taking log on both side

$$\log L = -\theta n + \Sigma x_i \log\theta + \log\left[\frac{1}{\prod_{i=1}^{n}(x_i)}\right] \quad \text{--- (1)}$$

differentiating (1) with respect to θ and set the equn to 0.

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

$$= -n + \frac{\sum x_i}{\theta} \cdot 1 0 = 0$$

$$\Rightarrow -n\theta + \sum x_i = 0$$

$$\Rightarrow n\theta = n\bar{x}$$

$$\Rightarrow \theta = \bar{x}$$

$$\therefore \hat{\theta} = \bar{x}$$

$$\frac{d^2 \log L}{d\theta} = \frac{-\sum x_i}{\theta^2} < 0$$

Hence $\hat{\theta} = \bar{x}$ is the MLE of θ

<u>Unbiasness:</u>

we have

$$\hat{\theta} = \bar{x}$$

$$= \frac{1}{n} \sum x_i$$

$$E(\hat{\theta}) = \frac{1}{n} E(\sum x_i)$$

$$= \frac{1}{n} n\theta$$

$$E(\hat{\theta}) \Rightarrow \theta$$

Hence $\hat{\theta}$ is an unbiased estimator

$\Rightarrow$ Let $x_1, x_2 \ldots x_n$ be a random sample with pdf is,

$$f(x; n, \theta) = \binom{n}{x} \theta^{x} (1-\theta)^{n-x};$$

(i) Find MLE of θ.

(ii) Check unbiasedness.

Sol^n:

Given that,

$$f(x; n, \theta) = \binom{n}{x} \theta^{x} (1-\theta)^{n-x}; \quad x = 0, 1 \ldots \alpha$$

The likelihood function is

$$L = \prod_{i=1}^{N} f(x_i; n\theta)$$

$$= \prod_{i=1}^{N} \binom{n}{x} \theta^{\Sigma x_i} (1-\theta)^{Nn - \Sigma x_i}$$

taking log on both sides and we have,

$$\log L = \log \prod_{i=1}^{N} \binom{n}{x} + \Sigma x_i \log \theta + nN - \Sigma x_i \,(1-\theta) \quad \text{—— (1)}$$

differentaing (1) & setting as 0

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

$$\Rightarrow \frac{\Sigma x_i}{\theta} + \frac{nN - \Sigma x_i}{(1-\theta)}(-1) = 0$$

$$\Rightarrow \Sigma x_i - \theta \cancel{\Sigma x_i} = nN\theta - \cancel{\theta \Sigma x_i}$$

$$\Rightarrow \cancel{n}N\theta = \cancel{n}\bar{x}$$

$$\Rightarrow \theta = \bar{x}/n$$

$$\therefore \hat{\theta} = \frac{\bar{x}}{n}$$

And,

$$\frac{d^2 \log L}{d\theta^2} = -\frac{\Sigma x_i}{\theta^2} - \frac{nN - \Sigma x_i}{(1-\theta)^2} < 0$$

which is negative definite.

Hence, $\hat{\theta} = \frac{\bar{x}}{n}$ is the MLE of θ

Now,

$$\hat{\theta} = \frac{\bar{x}}{n}$$

$$= \frac{1}{nN}\Sigma x_i$$

$$E(\hat{\theta}) = \frac{1}{nN} E(\Sigma x_i)$$

$$\Rightarrow \frac{1}{nN} \cancel{N(n\theta)}\, Nn\theta$$

$$\Rightarrow \frac{\cancel{1}}{\cancel{n}}\theta$$ (Hence unbiased)

$$E(\hat{\theta}) = \theta$$