σ^2 (varience) $= \dfrac{\sum (x_i - \mu)^2}{N}$

s^2 (sample varience) $= \dfrac{\sum x_i - \bar{x})^2}{n-1}$

proportion $= \dfrac{x}{n}$

x = Number of Successness

n = Total number of observation.

<u>some formula:</u>

1) $B(l, m) = \int_0^1 x^{l-1} (1-x)^{m-1} \, dx$

2) $B(l, m) = \int_0^\infty \dfrac{x^{l-1}}{(1+x)^{l+m}} \, dx$.

3) $B(l, m) = \dfrac{\Gamma l \, \Gamma m}{\Gamma l+m}$

4) $\dfrac{\Gamma n}{\alpha^n} = \int_0^\infty x^{n-1} \cdot e^{-\alpha x} \, d.x$

5) $\Gamma n = (n-1) \Gamma(n-1)$

6) $\Gamma n = \int_0^\infty x^{n-1} e^{-x}$

Test of hypothesis

Hypothesis: Any statement about the population is called hypothesis.

STATISTICAL HYPOTHESIS: Statistical hypothesis is a statement about population characteristics that can be tested on basis of sample data.

Null Hypothesis: Null hypothesis is the hypothesis which is to be tested for possible rejection under the assumption is true. It's denoted by H_0. Ex: $H_0 : \mu = \mu_0$

Alternative Hypothesis: The hypothesis, which is true if the null hypothesis is false is called alternative hypothesis. It's denoted by H_A or H_1

Ex: $H_1 : \mu \neq \mu_0$

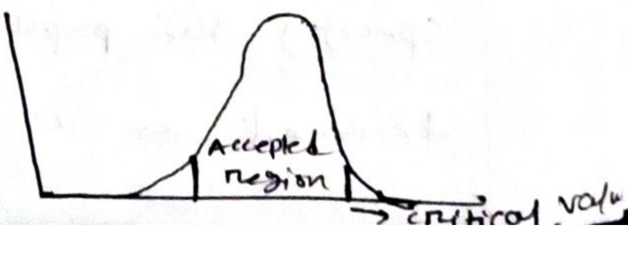
simple Hypothesis: If the hypothesis doesnot specify the population completely then it is termed as a simple hypothesis.

Composite Hypothesis: If the hypothesis doesnot specify the population completely then it is termed as a composite hypothesis.

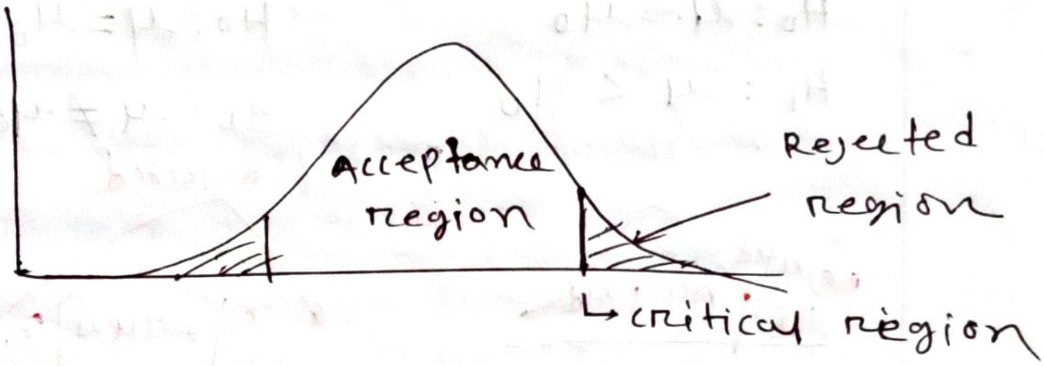
Parametric Hypothesis: when the hypothesis concerning the parameters of the distribution ~~with specified~~ or unspecified parameter provided from the distribution is called parametric hypothesis.

Non parametric hypothesis: A nonparmetric hypothesis test is a statistical methode that doesnot make assumptions about the distribution of data. Ex:- $H_0 : F_x(x) = F_y(y)$

Critical Region Or Rejection Region:
A critical region is a set of values for a test statistic that leads to the rejection of a null hypothesis.



Acceptance Region : A set of values for the test statistics for which the null hypothesis is accepted.



Test of significance : Test of significance is a statistical formal procedure that compares data with a claim or hypothesis. Its helps decides whether to accept or reject a hypothesis by compareing the observed data with expected results. The aim of test of significance is to reject the null hypothesis.

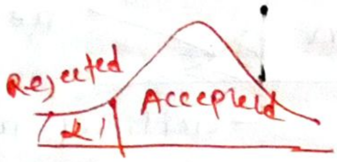
One Tailed Test : A test of any statistical hypothesis where the alternative hypothesis is one tailed (right or left tailed).

Two Taid Test : ... The alternative hypothesis is two tailed.

One Taided Test (Left)

$H_0 : \mu = \mu_0$

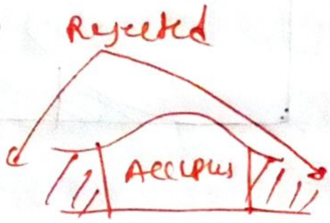
$H_1 : \mu < \mu_0$



Two Tailed Test

$H_0 : \mu = \mu_0$

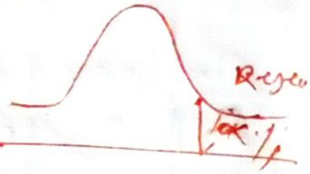
$H_1 : \mu \neq \mu_0$



One Tailded Test (Right)

$H_0 : \mu = \mu_0$

$H_1 : \mu > \mu_0$



Error

Type 1 Error: If the null hypothesis is actually true but on the basis of sample we reject Ho (accepting H_1). This type of error is known as type-1 error.

Type 2 Error: If the null hypothesis is false but on the basis of sample we accept Ho (H_1 is true) this type of Error is known as type-II error.

Q: which kinds of error is more serius and why?

I think that type-I-error is more serious than type-II error.

Suppose: In a murder case,

H_0: The man is not guilty.

H_1: The man is guilty.

In type 1 error, accept ~~H_0~~ H_1. so he ~~can~~ should be punished. but if H_0 is accepted, which is type-II error. he should be acquitted. If a guilty man ~~is~~ be acquitted then an error of type-II is commited whereas if an innocent may be convicted an error of type I is commited.

From the example type-II error is more serius.

<u>Level of significance</u>: the probability of rejecting a true null hypothesis is called level of significance. It's denoted by α. and represents the chance of making a type-1- error.

$\alpha = 0.05 = 5\%$

$\alpha = 0.01 = 1\%$

$\alpha = 0.1 = 10\%$

Two tailed	One tailed
$\lvert Z \rvert > 1.96$	$Z \le -1.64$ or $Z \geq 1.64$ or
$\lvert Z \rvert > 2.58$	$Z \le -2.3$ or ...
$\lvert Z \rvert >$	

Level of..	one..	two.
10%	1.28	± 1.64
5%	1.64	± 1.96
1%	2.33	± 2.58

<u>Power of a test</u>: the complement of the probability of type II error is called the power of a test.

$$P = 1 - \beta \text{ (accept } H_0 \text{ / } H_0 \text{ is false)}$$

Degree of freedom: Degree of freedom refers to the number of independent values that can be vary in a sample, or the number of values in a calculation that are free to vary.

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

p-value: is actual risk of committing a type-I error. The p-value measures the strength of evidence against H_0.

Procedure of Testing a hypothesis:

Step-1: Setting up of hypothesis

Step-2: set up a suitable level of significance

Step-3: computation of test statistic

Step-4: Determine critical region

Step-5: calculating test statistic

Step-6: Taking Decision.

□ Some important test of significance:

1. Normal test (z-test) $n \geq 30$
2. t-test $n < 30$
3. χ^2-test (chi square test)
4. F-test.

z-Test: (Normal Test)

A z-test is a hypothesis test for data that follow a normal distribution.

If $n > 30$, then we can Apply z-test.

$$z = \frac{U - E(U)}{\sigma(U)}$$ $\longrightarrow$ standard error.

$$z = \frac{\bar{x} - \mu}{\sigma/\sqrt{n}}$$ $\boxed{n \geq 30}$ $\boxed{\text{varience known}}$ | $$t = \frac{\bar{x} - \mu}{s/\sqrt{n}}$$

$\bar{x}$ = sample mean

σ = standard deviation of population

μ = population mean to be tested.

$\frac{\sigma}{\sqrt{n}}$ = standart error of ~~standard~~ sample mean.

Application of z-test

1) Test of a single population mean.

2) Test of equality of two population means.

3) Test of a single population proportion.

4) Test ~~of a~~ for difference between two populatio proportions.

5) Test of a specified correlation co-efficient.

6) Test of equality of two-population correlation co-efficient.

Application of t-statistic:

1) Test of a single population mean,

2) Test of difference between two population mean

3) Test of significance of a correlation co-efficient with zero value.

~~4) Test of equality of several population proportion.~~

~~5) Test of independence of attributes.~~

~~6) Test of goodness of fit.~~

4) population of regression co-efficient with zero.

5) difference between two population regression co-efficient;

Application of x^2- statistic:

1) Test of a population varience with specific value.

2) ,, equality of several varience.

3) equality of several correlation co-efficient

4) Test of equality of several population proportions.

5) Independence of attributes.

6) Test of goodness of fit.

Application of F- statistic:

1) Test significance of difference between two population variences.

2) Test of significance of several population means.

3) Test of significance of two or more regression co-efficient.

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

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

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

	10%	5%	2.5%	1%	0.5%	0.2%
One tailed	1.28	1.64	1.96	2.33	2.58	2.88
Two tailed	1.64	1.96	2.33	2.58	2.88	3.08

16.6.1

the managing director of a firm claims that his firm product 110 items on average daily.

A random sample of 15 days gives the following data set;

110, 118, 130, 140, 142, 146, 112, 100, 95, 98, 96, 122, 123, 124, 130

It normal distribution and varience $\sigma = 300$

can we conclude at 5% level of significance that the average daily production of items of that firm.

a) 110 items. b) more than 110 items.

c) Less than 110 items.

a)

$H_0 : \mu = 110$

$H_1 : \mu_1 \neq 110$

$$z = \frac{\bar{x} - \mu_0}{\sigma/\sqrt{n}} = \frac{119.07 - 110}{17.32/\sqrt{15}}$$

$= 2.03.$

$\bar{x} =$ ~~110~~ sample mean

$= \frac{\Sigma x}{n} = \frac{1786}{15}$

$= 119.07$

$\mu_0 = 110$

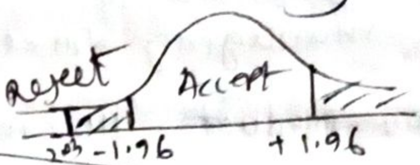
$\sigma^2 = 300$

$\sigma = \sqrt{300} = 17.32$

5% level of significance, two tailed test

$\alpha = 1.96$

$z > \alpha$ so it reject null hypothesis.



~~ⓐ~~

p-value:

$\boxed{P(Z > 2.03) = 0.0212}$

not needed, It also have p-table ☺

ⓑ

$H_0 : \mu = 110$

$H_1 : \mu > 110$ [Right tailed test]

$z = 2.03$

in Right tailed test the critical value is ~~1.64~~ $\boxed{1.64}$

Hence It Reject null hypothesis.

ⓒ

$H_0 : \mu = 110$

$H_1 : \mu < 110$ (Left tailed test)

critical value is -1.64. $z = 2.03$.

so accept the null hypothesis.

16.8.3

$H_0: \mu = 1500$

$H_1: \mu \neq 1500$

$\bar{x} = \cancel{115}\ 1510$

$\sigma = 45$

$n = 100$.

$$Z = \frac{\bar{x} - \mu_0}{\sigma/\sqrt{n}} = \frac{1510-1500}{45/\sqrt{100}} = 2.22$$

In two tailed test the critical value of 5% is 1.96. Hence Failed to accept null hypothesis.

16.7.1 $H_0: \mu_1 = \mu_2$; $H_1: \mu_1 > \mu_2$ [Right tailed test]

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

$n_1 = n_2 = 60$

$\bar{x}_1 = 158.5$

$\bar{x}_2 = 141.60$

$\sigma_1 = 18.20$

$\sigma_2 = 20.60$

$$= \frac{(158.5 - 141.6)}{\sqrt{\frac{18.20^2}{60} + \frac{20.60^2}{60}}}$$

$= 4.96$

5% level of significance, critical value is 1.96 1.64. so accept null hypothesis.

16.7.2

100 bulbs bought

	A
$\bar{x}$	1300
sd	50

$H_0 : \mu_1 = \mu_2$

$H_1 : \mu_1 \neq \mu_2$

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

$$= \frac{1300 - 1330}{\sqrt{\frac{50^2}{100} + \frac{6}{10}}}$$

$= \cancel{7.81}\ -3.8$

In two tailed tes

is $|z| > 1.96$ (s

$|z| > 1.96$ so nu

16.8.12

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

Forecasts of corporate earning per share made on a regular basis by many financial analysts. In random of 600 forecasts, it was found 382 these forecasts extended the actual outcome for ernings. Test agains a two tailed alternative the null hypothesis that the population population proportion of forecasts that are higher than actual outcomes is 0.50 at 5% level of significance.

$H_0: \pi = 0.5$

$H_1: \pi \neq 0.5$

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

$$p = \frac{382}{600} = 0.637$$

$$= \frac{0.637 - 0.5}{\sqrt{\frac{0.5(1-0.5)}{600}}}$$

$= 6.73$ $\boxed{\cancel{> 1.96}}$.

since 6.73 is much bigger than 1.96.

so ~~t~~ null hypothesis is rejected.

— x —

~~Ex~~ 16.8.1

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

$p = \frac{x}{n} = \frac{40}{100}$

$= 0.4$

$\pi_0 = 0.25$

$n = 100$

$$= \frac{0.4 - 0.25}{\sqrt{\frac{0.25(1-0.25)}{100}}}$$

$$= 3.46 > 1.96$$

Two tailed test, the critical value is 1.96 since null hypothesis is rejected.

Non parametric test

A non-parametric test is a statistical method that doesn't assume a specific probability distribution for the data. It used when data doesnot meet the assumptions of parametric test such as normality or varience.

Advantages:

1. Works when the varience isn't normally distributed.
2. useful for nominal or ordinal data.
3. Test hypothesis without population parameter.
4. Easier computation in some cases.
5. simple to understand.

Disadvantages:

1. Less sensitive than parametric methods.
2. Uses less information, ~~&~~ reducing precision.
3. Less efficient, requiring larger sample sizes.

Parametric test vs non parametric test:

Parametric test:

- population parameter is known
- Quantitive study.
- Applicable for Quantive/continous variable.
- Data normally distributed.
- measurement scale is interval or ratio.
- Large sample size.
- central tendency is mean.
- probability sampling.
- More powerful.

Non parametric test:

- population parameter is unknown.
- Qualitive study.
- Applicable for contionous and discrete.
- No such distribution.
- Measurenment scale is nominal or ordinal.
- small sample size.
- central tendency is median.
- non-prabability sampling.
- Less powerful.

Sign test: The test value is smaller number of plus or minus.

18	43	40	16	22		−	+	0	−	−
30	29	32	37	36		−	−	−	−	−
39	34	39	45	28		−	−	−	+	−
36	40	34	39	~~45~~ 52		−	0	−	−	+

$\boxed{\text{Median} = 40}$

at $\alpha = 0.05$

step: 1

H_0 : median $= 40$

H_1 : median $\neq 40$

Step: 2

$n = 18$; ~~test value~~ critical value $= 4$

Step: 3

test value $= 3$ (smaller value $= 3$)

Step: 4

$3 < 4$, the null hypothesis is rejected.

Step: 5

Reject null hypothesis.

when sample size is $n \geq 26$

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

x = smaller number of (+) or (−)

n = sample size.

13.2

H_0: MD = 36.4

H_1: MD < 36.4

$\alpha = 0.05$; $n = 50$; left tailed test

critical value is −1.65

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

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

Accept the null hypothesis.

Pair simple sign test:

13.2 Ear Infection in swimers.

Swimmer	Before X_B	After X_A	Sign
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	−

$n = 9$;

test value = 2

$\alpha = 0.05$ (one-tailed) at most ~~one~~ 1 negative

Sign. to reject null hypothesis.

$2 > 1$ so Reject null hypothesis.

The wilcoxon Rank test: used to compare two independent sample to determine it they come from the same distribution.

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

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

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

R = Sum of Ranks for smaller sample size

n_1 = smaller smaller size

n_2 = larger sample size

$n_1 \geq 10$ and $n_2 \geq 11$

Time to complete an obstacle course:
Two independent Army and marine recruits are selected. and the times it take each recruits to complete an obstakel council. At $\alpha = 0.05$, is there is a diffenence in time?

Ho: There is no different in times.

H1: There is a differente in times.

													Mean
A	15	18	16	17	13	22	24	17	19	21	26	28	19.67
M	14	9	16	19	10	12	11	8	15	18	25		14.27

$n_1 = 11$

$n_2 = 12$

Compute the test value;

arrange the combined data in order and rank each value.

Times	8	9	10	11	12	13	14	15	15	16	16	17	17	18
Group	M	M	M	M	M	M	A	M	A	M	A	A	M	A
Rank	1	2	3	4	5	6	7	8.5	8.5	10.5	10.5	12.5	12.5	14.5

	18	19	19	21	22	24	25	26	28
	M	A	M	A	A	A	M	A	A
	14.5	15.5	15.5	18	19	20	21	22	23

R = sum of the Ranks of smaller size (M)

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

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

$$\sigma_R = \sqrt{\frac{n_1 n_2 (n_1+n_2+1)}{12}} = 16.2$$

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

$$= \frac{93 - 132}{16.2}$$

$$= -2.41 .$$

at $\alpha = 0.05$ the critical value is 1.96 (two tailed)

$-2.41 < 1.96$, so, Reject the null hypothesis. There is a difference in time.

<u>Run Test</u> used to check randomness in a sequence of data.

Test randomness at $\alpha = 0.05$

FFF MM FFFF M F M MM FF FF MM FFF MM

Step 1.

H_0 : The passenger board the train at random

H_1 : Not Random.

Run	Gender
1	FFF
2	MM
3	FFFF
4	M
5	F
6	MMM
7	FFFF
8	MM
9	FFF
10	MM

There is 15 Female (n_1)
10 Male (n_2)

Step 3: Find critical value (7, and 18)

~~Step 4:~~ ~~The~~ make the decision,

the number of runs 10 is between 7 and 18, donot reject null hypothesis

Step 5: not enough evidence to reject null hypothesis.

Passenger board the train at random according to gender.

"Sampling Distribution"

A sampling distribution is a probability distribution of a statistics that obtained ~~from a~~ through repeated sampling of a specific population.

There are 3 types of sampling distribution are :—

1. χ^2 distribution
2. t- distribution
3. F- distribution.

Parent Distribution: The probability distribution of parameter is called parent distribution.

For example: Normal, Bionomial distribution are parent distribution.

* Distinguish between sampling and parent distribution.

Parent Distribution	Sample Distribution
1. whole Population	1. A small part of population
2. Large or infinite size	2. small or limited size.
3. True population parameter (μ, σ^2)	3. Estimates the population parameters ($\bar{x}, s^2$)
4. Represent true underlying distribution	4. Helps guess the parent distribution.
5. Doesnot change.	5. [illegible]

χ^2 (chi-square) Distribution.

The sum of squares of n independent standard normal variable is called chi-square (χ^2) variate with n degree of freedom.

Let $z_1, z_2, z_3 \dots z_n$ be n independent standard normal variates, then

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

However if $x_1, x_2, x_3 \dots x_n$ are n independently and identically distributed random variables each of which is normally distributed with mean μ and varience σ^2 Then

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

Properties of χ^2 distribution:

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

d.f (degree of freedom)

II) The distribution contains only one parameter which is the degree of freedom of the distribution.

III) The mean and varience of x^2-distribution for n d.f is n and $2n$ respectively.

IV) The mode of x^2-distribution for n d.f is $n-2$.

V) The moment generatting function of x^2-distribution for n d.f is $(1-2t)^{-n/2}$

VI) x^2-distribution trends to normal distribution for large degree of freedom.

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

VII) The distribution becomes symmetrical as n tends to infinity $(n \to \infty)$

Application of χ^2-distribution:

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

ii) To test goodness of fit.

iii) To test the independence of attributes.

iv) To test the homogenecity of independent estimate of the population varience.

v) to test the homogenecity of independent estimates of population correlation coefficient.

vi) To combine various probabilitys obtained from independent experiments to give a single test of significance.

Question: show that, the total probability of χ^2-distribution is unity.

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

pdf = probability density function

Proof:

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

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

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

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

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

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

$$= 1$$

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

Thus the total probability of χ^2-distribution is unity (unity)

Question: Find the mean and varienee of χ^2-distribution.

Solution:

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

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

Mean:

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

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

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

$$= \frac{1}{2^{n/2}\cdot\Gamma n/2} \frac{\Gamma n/2+1}{\left(\frac{1}{2}\right)^{n/2+1}} \qquad \boxed{\because \frac{\Gamma n}{\alpha^n} = \int_0^\infty x^{n-1} e^{-\alpha x}\, dx}$$

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

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

$$\therefore E(\chi^2) = n$$

$\therefore$ mean $= n$

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

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

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

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

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

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

$$= \cancel{2}n\left(\frac{n+2}{\cancel{2}}\right)$$

$$= n^2 + 2n$$

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

$$\therefore V(\chi^2) = E[(\chi^2)^2] - [E(\chi^2)]^2$$

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

$$= 2n$$

Therefore, the mean and varience of χ^2-distribution is n and $2n$ respectively.

t-distribution

Let u be a $N(0,1)$ variate and v be a chi-square (χ^2) variate with n degree of freedom. Also u and v are independent.

Define $t = \dfrac{u}{\sqrt{v/n}}$ Then t will follow t distribution with n df.

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

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

Properties of t-distribution:

i) t-distribution is an even function.

ii) t-distribution is symmetric about $t=0$

iii) mean = median = mode = 0

iv) varience of the distribution is $\frac{n}{n-2}$; $n>2$

v) The total probability of t-density 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) All large n t-distribution~~

vii) All odd order row moments are zero i.e. $\mu_{2r+1} = 0$

viii) Even order row moments are found by the relation:

$$\mu'_{2r} = \frac{n^r \Gamma(r+\frac{1}{2}) \Gamma(\frac{n}{2}-r)}{\Gamma\frac{1}{2} \Gamma\frac{n}{2}} \; ; r=1,2,3 \ldots$$

Mgf = Moment generating function.

ix) since $\beta = 0$ and $\beta_2 = 3 + \frac{6}{n-4} > 3$, therefore, the distribution is symmetric

$\beta_1 = 0$ and $(\beta_2 > 3)$ leptokurtic.

x) It is a continous type of distribution and its range extends form $-\infty$ to ∞ ie $-\infty < t < \infty$.

xi) Mgf of t-distribution doesnot exist,

Application of t-distribution:

i) To test if the sample mean $(\bar{x})$ differs significiantly from the hypothetical value of μ of the population mean.

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

iii) To test the significance of observed sample correlation co-efficient and sample regression co-efficient.

iv) To test the significance of an observed partial correlation co-efficient.

v) To test single population mean.

*Distinguish between t and normal distribution:-

t-distribution	Normal distribution
① The pdf. of t-distribution is: $f(t) = \frac{1}{\sqrt{n}\, \beta(\frac{1}{2}, \frac{n}{2}) (1+\frac{t^2}{n})^{\frac{n+1}{2}}}$; $-\infty < t < \infty$	① The pdf of normal distn is: $f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}$; $-\infty < x < \infty$
② Mean = median = mode = 0	② Mean, median, mode are not zero.
③ It is an exact sampling distribution	③ It is a parent distribution.
④ The distribution is symmetric and leptokurtic since $\beta_1 = 0$ and $\beta_2 > 3$	④ The distribution is symmetric and Mesokurtic since $\beta_1 = 0$, $\beta_2 = 3$.

Question: show that, the total probability of t-density is equal to 1.

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

proof:

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

Let, $w = \frac{t^2}{n}$ $\therefore t = \sqrt{nw}$

$\Rightarrow t^2 = nw$

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

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

$\Rightarrow dt = \frac{n}{2\sqrt{nw}} dw$

$\Rightarrow dt = \frac{\sqrt{n}}{2\sqrt{w}} dw$

$\therefore \int_{-\infty}^{\infty} f(t)\, dt = 2\int_{0}^{\infty} \frac{1}{\cancel{\sqrt{n}}\, \beta(\frac{1}{2}, \frac{n}{2}) (1+w)^{\frac{n+1}{2}}} \cdot \frac{\cancel{\sqrt{n}}}{2\sqrt{w}} dw$

[since The integral is an even function of t]

$$\Rightarrow \int_{-\infty}^{\infty} f(t)\, dt = \int_{0}^{\infty} \frac{1}{\beta(\frac{1}{2},\frac{n}{2})} \frac{w^{-\frac{1}{2}}}{(1+w)^{\frac{n+1}{2}}}\, dw$$

$$= \frac{1}{\beta(\frac{1}{2},\frac{n}{2})} \int_{0}^{\infty} \frac{w^{\frac{1}{2}-1}}{(1+w)^{\frac{1}{2}+\frac{n}{2}}}\, dw$$

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

$$= 1$$

$\therefore$ the total probability of t-density is equal to 1. (Showed)

Question: Find mean, varience of t-distribution.

Answer:

we know that, the pdf of t-distribution is,

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

we know,

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

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

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

$$= \frac{1}{\sqrt{n}\, \beta\left(\frac{1}{2} \cdot \frac{n}{2}\right)} \cdot 0$$ [since the integrand is an odd fuction of t]

$= 0$

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

Now,

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

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

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

$\Rightarrow t^2 = nw$

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

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

$\Rightarrow dt = \frac{n}{2\sqrt{nw}} dw$

$\Rightarrow dt = \frac{\sqrt{n}}{2\sqrt{w}} dw$

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

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

$$= 2 \cdot \frac{1}{2} \int_{0}^{\infty} \frac{nw}{\beta\left(\frac{1}{2}, \frac{n}{2}\right)} \cdot \frac{w^{-\frac{1}{2}}}{(1+w)^{\frac{n+1}{2}}} dw$$

[since the integral is an even function of w

$$\Rightarrow E(t^2) = \boxed{\cancel{2\int_0}\ \frac{n}{\beta\left(\frac{1}{2},\frac{n}{2}\right)}\int_0^\infty \frac{w^{3/2-1}}{(1+w)^{3/2+\frac{n-2}{2}}}\,dw}$$

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

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

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

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

$$= \frac{\dfrac{n\cdot\Gamma\left(\frac{3}{2}\right)\cdot\Gamma\left(\frac{n-2}{2}\right)}{\Gamma\left(\frac{3}{2}+\frac{n-2}{2}\right)}}{\dfrac{\Gamma\left(\frac{1}{2}\right)\,\Gamma\left(\frac{n}{2}\right)}{\Gamma\left(\frac{n+1}{2}\right)}} \qquad \cancel{n\cdot\Gamma\left(\frac{3}{2}\right)}$$

$$= \frac{n \cdot \Gamma\frac{n}{2}\, \Gamma\frac{n-2}{2} \Big/ \cancel{\Gamma\frac{n+1}{2}}}{\Gamma\frac{1}{2}\, \Gamma\frac{n}{2} \Big/ \cancel{\Gamma\frac{n+1}{2}}}$$

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

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

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

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

$$\Rightarrow V(t) = E(t^2) - [E(t)]^2$$

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

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

Question:

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

i.e. mean = mode = median = 0

Meam:

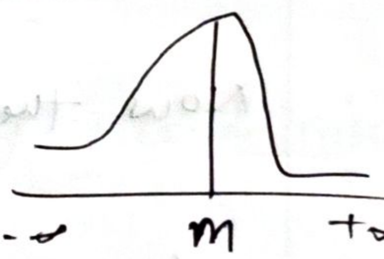
$\therefore E(t) = 0$

$\therefore$ meam = 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,

$$M = \int_{M}^{\infty} f(t) \cdot dt = \frac{1}{2} \quad \text{---(i)}$$

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

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

$$\Rightarrow 2\int_{0}^{\infty} f(t) \cdot dt = 1$$

$$\therefore \int_{0}^{\infty} f(t) \cdot dt = \frac{1}{2} \quad \text{— (ii)}$$

Comparing (i) and (ii) we get,

$\therefore M = 0$

$\therefore$ median $= 0$

Hence the median of t- distribution is zero.

<u>Mode of t-distribution:</u>

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 —

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

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

Now,

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

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

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

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

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

$$\Rightarrow t = 0$$

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

at $t=0$,

Hence $t=0$ is the mode of the distribution

$\Rightarrow$ Mode $= 0$.

Hence Mean = Median = Mode = 0 (Showed).

Question: Establish the relationship between t- distribution and cauchy distribution.

Answer:

We know, the pdf of t distribution, is as

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

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

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

$$= \frac{1}{\frac{\Gamma\frac{1}{2}\ \Gamma\frac{1}{2}}{\Gamma\left(\frac{1}{2}+\frac{1}{2}\right)}(1+t^2)}$$

$$= \frac{1}{\sqrt{\pi}\cdot\sqrt{\pi}\cdot(1+t^2)} \qquad \Gamma 1 = 1$$

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

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

Proof:

we know that, the pdf of t- distribution is as

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

taking limit on both sides, we have —

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

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

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

$$= \lim_{n\to\infty} \frac{\Gamma\left(\frac{1}{2} + \frac{n}{2}\right)}{\sqrt{n} \cdot \sqrt{\pi} \cdot \Gamma\frac{n}{2}} \quad \left[\because \Gamma\tfrac{1}{2} = \sqrt{\pi}\right]$$

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

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

$$= \frac{1}{\sqrt{\pi}.\cancel{(n^{1/2})}} . \cancel{(n^{1/2})} . 2^{-\frac{1}{2}}$$

$$= \frac{1}{\sqrt{2\pi}}$$

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

Also

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

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

$$= e^{-t^2/2} . 1$$

Now

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

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

which is the pdf standard normal distribution

Therefore ~~t~~ distribution tends to normal distribution.

F - distribution

The F- distribution is the distribution of the ratio of two independent chi - square (χ^2) random

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 mathimatically, $F \sim F(n_1, n_2)$

The density function of F is —

$$f_1(F) = \frac{\frac{n_1}{n_2}\left(\frac{n_1}{n_2}F\right)^{\frac{n_1}{2}-1}}{\beta\left(\frac{n_1}{2}, \frac{n_2}{2}\right)\left(1+\frac{n_1}{n_2}F\right)^{\frac{n_1+n_2}{2}}} \quad ; F>0$$

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

Properties of F-distribution:

i) F-distribution is continous type of distribution and its range is 0 to ∞. ie $0<F<\infty$

ii) It is an exact sampling distribution.

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

iv) If $F \sim F(n_1, n_2)$ the the mean and varience is $\frac{n_2}{n_2-2}$ and varience $\frac{2n_2^2(n_1+n_2-2)}{n_1(n_2-2)^2(n_2-4)}$ respectively.

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

vi) If $F \sim F(n_1, n_2)$, then $\frac{1}{F} \sim F(n_2, n_1)$

vii) If $F \sim F(n_1, n_2)$ then $\frac{n_1}{n_2} F \sim \beta_2\left(\frac{n_1}{2}, \frac{n_2}{2}\right)$

viii) . then $\frac{1}{1+\frac{n_1}{n_2}F} \sim \beta_1\left(\frac{n_1}{2}, \frac{n_2}{2}\right)$

ix) If n_1 and n_2 are vary large, then F-distribution tends to normal distribution.

x) The distribution is positively skewed.

Application of F-distribution:

i) F-distribution is used to test equality of population varience.

ii) used for testing the significance of an observed multiple correlation coefficient and sample correlation ratio.

iii) Used to testing the linearity of regression.

iv) to test the equality of several means.

Question:

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

i.e. $$\int_0^\infty f(F)\cdot dF = 1$$

we know the pdf of F-distribution is as

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

NOW,

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

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 $w=0$

$F=\infty$ " $w=\infty$

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

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

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

$= 1$ (Proved)

Question:

Find mean and varience of F-distribution.

Answer:

The pdf of F-distribution,

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

Mean:

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

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

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

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

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

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

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

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

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

$$= \frac{n_2}{n_1} \cdot \frac{\Gamma\left(\frac{n_1}{2}+1\right) \cdot \Gamma\left(\frac{n_2}{2}-1\right)}{\Gamma\left(\frac{n_1}{2}\cancel{+1}+\frac{n_2}{2}\cancel{-1}\right)} \times \frac{\cancel{\Gamma\left(\frac{n_1}{2}+\frac{n_2}{2}\right)}}{\Gamma\frac{n_1}{2} \cdot \Gamma\frac{n_2}{2}}$$

$$= \frac{n_2}{\cancel{n_1}} \cdot \frac{\frac{\cancel{n_1}}{2}\cancel{\Gamma\frac{n_1}{2}} \cdot \cancel{\Gamma\left(\frac{n_2}{2}-1\right)}}{\cancel{\Gamma\frac{n_1}{2}} \cdot \left(\frac{n_2}{2}-1\right) \cancel{\Gamma\left(\frac{n_2}{2}-1\right)}}$$

$$= \frac{n_2}{\cancel{2}} \times \frac{\cancel{2}}{n_2-2}$$

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

$$\therefore E(F) = \frac{n_2}{n_2+2} = \text{mean.}$$

Now,

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

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

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

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

$\cancel{E(F)}$

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

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

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

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

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

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

$$\therefore \mu_2' = E(F^2) = \frac{n_2^2 (n_1+2)}{n_1 (n_2-2)(n_2-4)}$$

Now variance,

$$V(F) = E(F^2) - \{E(F)\}^2$$

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

$$\boxed{= \frac{\cancel{n_1 n_2^3} + 2}{n_1 (n_2-2)^2 (n_2-4)}}$$

$$= \frac{(n_2^3 + 2n_2^2)(n_1+2) - n_1 n_2^2 (n_2-4)}{n(n_2-2)^2(n_2-4)}$$

$$= \frac{\cancel{n_1 n_2^3} + 2n_2^3 - 2n_1 n_2^2 - 4n_2^2 - \cancel{n_1 n_2^3} - 4n_1 n_2^2}{n(n_2-2)^2(n_2-4)}$$

$$= \frac{2n_2^3 + 2n_1 n_2^2 - 4n_2^2}{n(n_2-2)^2(n_2-4)}$$

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

$$\therefore \text{Vanience} = \frac{2n^2(n_2+n_1-2)}{n_1(n_2-2)^2(n_2-4)}$$

ESTIMATION

population: population is a set of similar items or events which is of interest for some experiment.

sample: A sample is a representative part of a population that is taken and considered for study.

parameter: population parameter is a number that describes a characteristic of an entire group or population.

Ex: population mean (μ), varience (σ^2) proportion (π) etc.

statistic: Any function of sample observation of items is called statistic.

Ex: $\bar{x}$, σ^2, p etc.

Random sample: is a subset of a statistical population.

Question: Explain the concept of estimation with the example.

Answer:

Estimation: Estimation is a process of finding an estimate or approximation, which is a value that is usable for some purpose even if input data may be incomplete, uncertain or unstable.

The 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

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

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 a possible point estimate on of $\tau(\theta) = \mu$.

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

Ex: sample mean, sample varience is the estimator of population mean or population varience –

Estimate: A particular value of estimator is called an estimator. Ex: $\bar{x} = 5.33$

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

Another defination, suppose $(x_1, x_2, x_3 \cdots x_n)$ is a sample. from 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 has a random variable. If t is used to estimate θ then t is called a point estimator of θ. If the realized value of t from from a sample is used for θ then t is called a point estimate of θ.

For example: if $f(x|\theta)$ is the normal density function that is $f(x|\theta) = (1/\sigma\sqrt{2\pi}) \exp[-\frac{1}{2}(x-\mu/\sigma)]$ where the parameter θ is (μ, σ^2) and it is 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 a possible point estimator of $\tau(\theta) = \mu$.

Interval extimation: 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 $\{t_1(x_1, x_2 \dots x_n), t_2(x_1, x_2 \dots x_n)\}$ constitates an interval for which the probability can be determined that is contains the unknown $\tau(\theta)$.

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

Properties of a good estimator:

(i) Unbaisedness.

(ii) Consististency

(iii) Efficiency

(iv) Sufficiency.

Unbaisedness: Any statistic whose mathmatical expectation is equal to a parameter θ is called an unbaised estimator of the parameter θ. Otherwise the statistic is said to be baised.

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

In case t_n be a biased estimator the difference $E(t_n) - \theta$ is the amount of bias and $E(t_n - \theta)^2$ is called meansqure error meansquare error of t_n = Variance of t_n + bias^2.

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

$$E(\bar{x}) = \frac{1}{n} E(x_1 + x_2 \cdots + x_n)$$
$$= \frac{1}{n}\left[E(x_1) + E(x_2) + \cdots E(x_n)\right]$$
$$= \frac{1}{n} \cdot n\theta$$
$$= \theta$$

$\therefore E(\bar{x}) = \theta$

And

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

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

$$= \frac{\sigma^2}{n-1} E\,(\chi^2_{n-1})$$

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

$$= \sigma^2$$

$\therefore E(\sigma^2) = \sigma^2$

Thus $\bar{x}$ and s^2 are an unbaised estimator of θ and σ^2 respectively.

Consistency:

Let t_n be a statistic calculated from a sample $(x_1, x_2, x_3 \cdots x_n)$ of size n from density $f(x|\theta)$

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

where ε and δ are arbitary small positive number then t_n is called a 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 $x_1, x_2 \cdots x_n$ is the random sample from a population with finite mean $E(x_i) = \mu < \infty$

Now we have

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

$$\therefore E(\bar{x}) = \mu$$

and $E(\bar{x}) = \mu$ as $n \to \infty$

Hence sample mean $\bar{x}_n$ is always a consistent estimator of the population mean μ.

Efficiency:

If $(x_1, x_2 \ldots x_n)$ be a sample from density $f(x|\theta)$ and t be an unbaised consistent estimator of θ and for no other estimators have varience less than that t, then t is said to be the most efficient estimator of θ (also simply called efficient estimator of θ)

Let t^* be any other unbaised statistic. the efficiency of t^* is the ration of reciprocal of the varience of t^* to the amount of information in the data. Actally the efficiency of t^* measured by

$$e(t^*) = \frac{v(t)}{v(t^*)} \quad \cdots (1)$$

the efficiency of t^* represents the fraction of the relevant information available actually utilized by t^* since $v(t) \leq v(t^*)$ the efficiency of any statistic vary between 0 to 1.

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

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

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

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 that T_2.

Sufficienty:

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

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

For example; $x \sim B(n \cdot \theta)$

Sample : $x_1, x_2 \cdots x_n$

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

Now,

$$P(x) = \theta^{\sum x_i} (1-\theta)^{n n - \sum x_i}$$

Now $\frac{P(x)}{P(\sum x_i)} = 1 \Big/ \binom{n}{\sum x_i}$

which is independent of θ. So $\sum x_i$ is a sufficient estimator of θ.

Theorem: Cramer- Rao Lower Bound (RLB):

suppose

(i) $x_1, x_2 \cdots x_n$ are independent random variables each with density $f(x|\theta)$, $\theta \in \Omega$ an open interval on the real line:

(ii) t is an estimator of parameter θ.

(iii) $E(t) = \theta + b(\theta)$, where $b(\theta)$ is the bias of t and is a differential function of θ.

(iv) The following regularity condition hold

(a) for almost all x, $\frac{\partial L}{\partial \theta}$ exist θ.

(b) $\frac{\delta}{\delta\theta}\int \cdots \int L = \int \cdots \int \frac{\partial L}{\partial \theta}$ which is possible when limit of integration are indepent of θ.

(c) $E\left[\frac{\partial \log L}{\partial \theta}\right]^2 > 0$ for $\theta \in \Omega$

(d) $\frac{\delta}{\delta\theta} \int \cdots \int t L$

$= \int \cdots \int t \frac{\partial L}{\partial \theta}$

Then for all ~~$\theta \in \Omega$~~ $\theta \in \Omega$

$$V(t) \geq \frac{[1 + b'(\theta)]^2}{n E\left[\frac{\partial \log t}{\partial \theta}\right]^2}$$