Continuous P.d - Normal Distribution -

* It is most widely used in continuous probability.
* Used to study the behaviour of continuous random variable like weight, height, etc.
* Normal Distribution is also known as Gaussian distribution.
* The graph of Normal Distribution is called Normal Curve or ~~Bell~~ Bell-shaped curve.
* The normal distribution is a probability function that describes how the values of the variable are distributed.
* Normal Distribution is defined by and given by the following probability function

$$\boxed{P(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}}$$

* A standard Normal Distribution is that normal distribution whose mean = 0 & S.D = 1.
* Standard Normal Distribution formula -

$$\boxed{Z = \frac{x-\mu}{\sigma}}$$

Z is the z-score (standard sum)
x is the value to be standardized
μ is mean
σ is S.D.

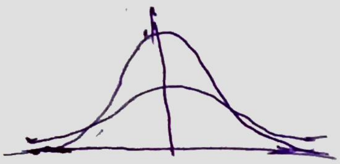
$$\boxed{P(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}z^2}}$$

* The graph of the Normal Distribution depends on two factors - mean & S.D.

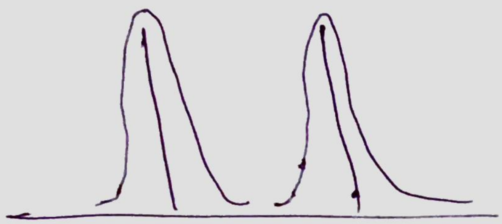
* The mean of a distribution determines the location of the centre of the graph. and S.D determines the width of the graph.

small S.D

— Bigger S.D.



Mean same
S.D different



mean diff
S.D same

Properties of Normal Distribution –

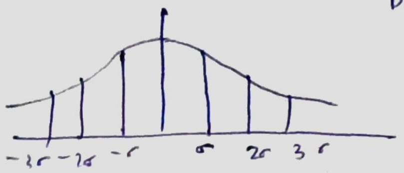
* Continuos Probability Distribution. takes into a count continuous variable like height, weight, time etc.
* Perfectly symmetrical & bell shaped. It is divided into two equal halves with mean at centre.
* It has only one modal distribution.
* Equality of mean, median, mode.
* The normal curve has the tendency to touch the x-axis but it never touches it. (Asymptomatic behaviour).
* The normal curve extends to infinity i.e. from -ve infinity to +ve ∞.
* The total area under the normal curve is 1.
* In a normal distribution, Mean deviation is $\frac{4}{5} \times S.D$.

* ~~Normal Distribution has two~~.

Area under the curve b/w $\mu - 1\sigma$ & $\mu + 1\sigma$ is 68.26%

b/w $\mu - 2\sigma$ & $\mu + 2\sigma$ is 95.45%

b/w $\mu - 3\sigma$ & $\mu + 3\sigma$ is 99.73%

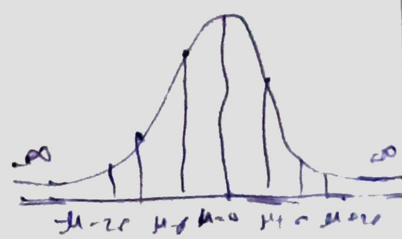


Binomial Distribution (when $p = q$)

Q. A sample of 100 dry battery cells t
The length of life produced the following results -
$\bar{x} = 12$ hrs, $\sigma = 3$ hrs. Assuming the data is normally distributed. What percentage of Battery cells are expected to have life i) More than 15 hrs
ii) Less than 6 hrs iii) B/w 10 and 14 hrs.

$$P(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}$$

$$S.N.D = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}z^2}$$



$$z = \frac{x-\mu}{\sigma}, \quad \mu = 0 \ \& \ \sigma = 1$$

$$z = x$$

i) $\mu = 12$, $\sigma = 3$ $\quad z = \frac{x-\mu}{\sigma} = \frac{15-12}{3} = 1$

$$P(z>1) = P(0<z<\infty) - P(0<z<1)$$
$$= 0.5 - 0.341 = \cancel{159}\ 0.159$$

ii) $z = \frac{6-12}{3} = -2$

$$P(z<-2) = P(z>2) = P(0<z<\infty) - P(0<z<2)$$
$$= 0.5 - 0.47 = 0.023$$

iii) $z = \frac{10-12}{3} = -\frac{2}{3}$

$z = \frac{14-12}{3} = \frac{2}{3}$

$$P\left(-\frac{2}{3}<z<\frac{2}{3}\right) = P(-0.66<\cancel{z}<0.66)$$
$$= 2P(0<z<0.66)$$
$$= 2 \times 0.2454 = 0.5508$$

Q. In a sample of 1000 cases, the mean of a certain test is 14 and S.D is 2.5. Assuming the distribution to be normal find i) how many students score b/w 12 and 15?

ii) How many score above 18?

iii) How many score below 8?

iv) How many score 16?

$$Z = \frac{12-14}{2.5} = -\frac{2}{2.5} = -0.8$$

$$Z = \frac{15-14}{2.5} = \frac{1}{2.5} = 0.4$$

i) $P(-0.8<z<0.4) = P(0<z<0.8) + P(0<z<0.4)$
$$= 0.288 + 0.1554$$
$$= 0.4435$$

ii) $Z = \frac{18-14}{2.5} = \frac{4}{2.5} = 1.6$

$$P(z>1.6) = P(0<z<\infty) - P(0<z<1.6)$$
$$= 0.5 - 0.4452 = 0.0548$$

iii) $Z = \frac{8-14}{2.5} = \frac{-6}{2.5} = -2.4$

$$P(z<-2.4) = P(z>2.4)$$
$$= 0.5 - 0.4918$$
$$= 0.0082$$

iv) $Z = \frac{16-14}{2.5} = \frac{2}{2.5} = 0.8$

$P(Z=0.8) = 0.2881$

Q. The income of a group of 10,000 persons was found to be normally distributed with mean ₹ 750 and S.D of ₹ 50. Show that of this group about 95% had income exceeding ₹ 668 and only 5% had income exceeding ₹ 832. Also find the lowest income among the richest 100.

$Z = \frac{668-750}{50} = \frac{-82}{50} = -1.64$

$P(Z > -1.64) = P(Z < 1.64)$

$= 0.5 + 0.4495$

$= 0.9495$

$Z = \frac{832-750}{50} = 1.64$

$P(Z > 1.64) = 0.5 - 0.4495 = 0.0505$

Unit - 5

Hypothesis
- Null hypothesis (H_0)
- Alternate hypothesis (H_1 or HA)

Null hypothesis (H_0) → Hypothesis → Statistical tests
- Parametric test
- Non- Parametric tests

Hypothesis- It is educated guess about the population or a prediction of the relationship b/w two or more variables.

Two types of Hypothesis :-

a) Null Hypothesis	Alternate Hypothesis
* A statement which states that there is no relationship b/w variables. Ex - ↑ in no. of cancer patient is not due to ↑ in pollution	A statement which states there is relationship b/w the variables. Ex - ↑ in cancer due to ↑ in pollution.
* It is denoted by H_0.	It is denoted by H_1 or HA.
* It is an exact opposite of what an investigator predicts or expects.	* It is an exactly what an investigator predicts or expects.
* It is the null hypothesis which is being tested i.e. statistical test are performed on the null hypothesis which	Statistical test ~~ie.~~ ~~non parametric test~~ are ~~performed on~~ not used in alternate hypothesis.

is parametric test.

Parametric Test - * These test are applied under the circumstances where the population is normally distributed or is assumed to be normally distributed.

* Parameters like mean, S.D etc. are used.
* These are applied where the data is quantitative.
* These are applied where the scale of measurement is either an interval or a ratio scale.

* **Non-Parametric Test**

* These test are applied under the circumstances where the population is not normally distributed (skewed distribution or uniform distribution).
* These test are also called as Distribution free tests.
* Parameters like mean, S.D are not used.

* t test, z test, annova test.	Chi square test Spearman Rank test

*

* These are applied where data is qualitative.
* These are applied where the scale of measurement is either ordinal or are nominal scale.

Basic Concept of Sampling & types of Hypothesis

Populati·   individuals under study is said to be population.

Finite population - A population containing finite no. of individuals is called finite population. For ex - No. of students in college.

Infinite population - A population containing infinite no. of individuals. Ex - Bacteria in curd, Particles in cement bag.

* Sample - A Sample is subset of entire ~~com~~ population
To study population, we select random sample and it is common method. Each member of population has equal chance to be include in sample.

Parameters of Statistics

Two main parameters in Statistic which is μ and σ.
Mean for population denoted by μ, S.D by σ.
For sample, mean is $\bar{x}$ and S.D by s.

Standard Error -
Standard deviation of sampling distribution of statistics is known as standard error.

It play a important role in large sample size and it forms basis of testing hypothesis.

Hypothesis- ~~a~~ Usually it is required to make decision about population on the basis of sample. To take decisions, it is necessary to make assumptions about population. These assumptions are called Hypothesis.

Null Hypothesis - It is definite statement about ~~pop~~ population and denoted by H_0. ~~A~~ It is hypothesis which test for possible rejection under assumption that it is true.

Alternate Hypothesis - Hypothesis which is complementary of Null Hypothesis denoted by H_1 or H_a, where H_0 is rejected, H_1 is accepted.

Let μ_0 be average marks of maths IV in sessional exams.

~~to~~ $H_0 : \mu = \mu_0$ (which is mean of population).

H_1 will be

$H_1 : \mu \neq \mu_0$ (two tailed)

$H_1 : \mu > \mu_0$ (~~one~~ single tail or right tail)

$H_1 : \mu < \mu_0$ (single tail or left tail)

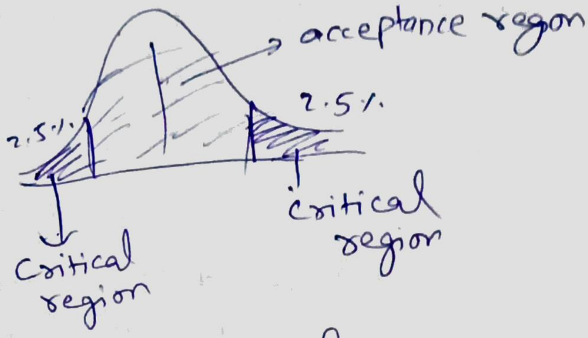
By the use of this, we know wheather the test is two tail or one tail.

Test of Significance-

A procedure which enable us to decide wheather the hypothesis is accepted or rejected on the basis of sample information or to determine observed sample statistic differ significantly from expected result.

Level of Significance -

It is probability of rejection of null hypothesis when it is true. It is also known as α.



Probability of critical region is known as level of significance denoted by α.

Level of significance is testing of hypothesis are 5% and 1%.

Error in Sampling -

Type 1 Error - The probability of rejection of null hypothesis H_0 when it is true is called type I error.

$$P(\text{reject } H_0, \text{ when it is true})$$
$$= P(\text{reject } H_0 / H_0) = \alpha$$

Type 2 Error - Probability of acceptance of null hypothesis when it is false.

$$P(\text{accept } H_0, \text{ when it is false})$$
$$= P(\text{accept } H_0 / H_1) = \beta$$

when H_1 is true,

NOTE: $(1-\beta)$ is known as Power of Test.

Test of Significance -

* Student t-test -

i) When the sample size is less than 30
ii) μ is given.
iii) S.D is unknown.

$$t = \frac{\bar{x} - \mu}{S / \sqrt{n}} \quad , \quad S = \sqrt{\frac{\Sigma (x - \bar{x})^2}{n-1}}$$

$\bar{x}$ is sample mean, μ is population mean,
n is sample size, S is population S.D.

for one mean

* and $t_{cal} = \frac{\bar{x} - \mu}{s / \sqrt{n-1}}$, s = S.D of sample

* Dof = $n-1$

For two mean $\bar{x}$ & $\bar{y}$, $t = \frac{\bar{x} - \bar{y}}{S \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}}$

$$S = \frac{n_1 s_1^2 + n_2 s_2^2}{n_1 + n_2 - 2}$$

$$= \frac{\Sigma (x_1 - \bar{x}_1)^2 + (x_2 - \bar{x}_2)^2}{n_1 + n_2 - 2}$$

Q. The lifetime of electric bulb pr a random sample of 10 from a large consignment w the following data

Items	1	2	3	4	5	6	7	8	9	10
lives in 000's hours	4.2	4.6	3.9	4.1	5.2	3.8	3.9	4.3	4.4	5.6
$(x - \bar{x})$	-0.2	0.2	-0.5	-0.3	+0.8	-0.6	-0.5	-0.1	0	1.2

Can we accept the hypothesis that the average lifetime of bulb is 4000 hrs.

Sol^n. Null Hypothesis $H_0: \mu = 4000$ hrs i.e. there is no significance difference b/w population mean and sample mean.

$$t = \frac{\bar{x} - \mu}{s/\sqrt{n}} \quad , n = 10 \qquad s = \sqrt{\frac{\Sigma(x-\bar{x})^2}{n-1}}$$

$$\bar{x} = \frac{\Sigma x_i}{n} = 4.4$$

$$t_{cal} = \frac{4.4\cancel{0} - 4000}{\cancel{0}.589/\sqrt{10}} \qquad = \sqrt{\frac{3.12}{9}} = 0.58$$

$$= \frac{400}{\cancel{0.183}\ 186.25}$$

$$= \cancel{2.185} = 2.147$$

dof' $t_{cal} < t_{tab}$ i.e. $2.18 < 2.26$ => null hypothesis could be

i.e. the average lifetime of bulb 4000 hrs.

Q. A sample of 18 items has a mean 24 units and S.D 3 units. Test the hypothesis that it is a random sample from a normal population with mean 27 units.

Null Hypothesis $H_0: \mu = 27$ i.e. there is no significance difference b/w population mean & sample mean

$n = 18$, $\bar{x} = 24$, $s = 3$

Testing of hypothesis:

$$t_{cal} = \frac{\bar{x} - \mu}{s/\sqrt{n-1}} = \frac{24 - 27}{3/\sqrt{18-1}} = \frac{-\cancel{3}}{\cancel{3}/\sqrt{17}} = \frac{-1}{\sqrt{17}}$$

$$= -4.12$$

$t_{cal} < t_{tab}$

Q. Two sample of sodium vapour bulbs ~~for~~ tested for length of life and the following results were brought:

	Size	Mean	Sample S.D
Type I	8	1234 hrs	36 hrs
Type II	7	1036 hrs	40 hrs

$$S = \frac{n_1 s_1^2 + n_2 s_2^2}{n_1 + n_2 - 2}$$

$$= \frac{8 \times 36^2 + 7 \times 40^2}{13}$$

$$= 1659.07$$

$$t_{cal} = \frac{\bar{x} - \bar{y}}{S\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}} = \frac{1234 - 1036}{1659 \times \sqrt{\frac{1}{8} + \frac{1}{7}}}$$

$$= \frac{198}{444.37} = \cancel{0.445}\ 0.234$$

Null hypothesis, $\mu_1 = \mu_2$ there is no significance difference b/w population mean & sample mean.

dof = 13 $\qquad t_{tab} = 2.16$

$t_{cal} < t_{tab}$. i.e. two sample mean are equal.

F-test

$$F = \frac{S_1^2}{S_2^2} = \frac{\sigma_1^2}{\sigma_2^2}, \quad \sigma_1^2 > \sigma_2^2$$

$$S_1^2 > S_2^2$$

n_1	n_2
s_1^2	s_2^2

$$S_1 = \frac{n_1 s_1^2}{n_1 - 1} \quad \& \quad S_2 = \frac{n_2 s_2^2}{n_2 - 1}$$

$\downarrow$

dof $= \nu_1 = n-1$ & $\nu_2 = n_2 - 1$

① $H_0 : \sigma_1^2 = \sigma_2^2$

② Fcal

The assumptions on which Ftest is based are :-

1) The population for each sample must be normally distributed.

ii) The sample must be random & independent.

iii) The ratio of σ_1^2 to σ_2^2 should be equal to 1 or greater than 1.

Applications -

F test is used to test -

i) wheather two independent samples have been drawn from the normal populations with the same variance σ^2.

ii) Wheather the two independent estimates of the population variance are homogenous or not.

Q. Two random samples drawn from two normal populations are as follows:

x_1	A	17	27	18	25	27	29	13	17
x_2	B	16	16	20	27	26	25	21	

Test whether the sample are drawn from same normal population.

Soln. — i) $H_0 : \sigma_1^2 = \sigma_2^2$ i.e. there is no significance difference b/w population variance.

ii) Testing of Hypothesis —

F-test → $F = \frac{\sigma_1^2}{\sigma_2^2} = \frac{S_1^2}{S_2^2}$

$S_1 = \sqrt{\frac{\Sigma(x_1 - \bar{x}_1)^2}{n_1 - 1}}$

$S_2 = \sqrt{\frac{\Sigma(x_2 - \bar{x}_2)^2}{n_2 - 1}}$

$\bar{x}_1 = 21.625$,

$x_1 - \bar{x}_1 \rightarrow -4.625 \quad 5.375 \quad -3.625 \quad 3.375 \quad 5.375 \quad 7.375 \quad -8.625 \quad -4.625$

$\Sigma(x_1 - \bar{x}_1)^2 = 253.875$

$\bar{x}_2 = 21.571$

$x_2 - \bar{x}_2 \rightarrow -5.571 \quad -5.571 \quad -1.571 \quad 5.429 \quad 4.429 \quad 3.429 \quad -0.571$

$\Sigma(x_2 - \bar{x}_2)^2 = 125.714$

$S_1 = \sqrt{\frac{253.875}{7}} = \sqrt{36.267} \Rightarrow \boxed{S_1^2 = 36.267}$

$S_2 = \sqrt{\frac{125.714}{6}} = \sqrt{20.952} \Rightarrow \boxed{S_2^2 = 20.952}$

$F_{cal} = \frac{S_1^2}{S_2^2} = 1.73$, $F_{tab\,(v_1, v_2)} = 4.21$

$v_1 = 8 - 1 = 7$

$v_2 = 7 - 1 = 6$

Conclusion: $F_{cal} < F_{tab}$

$\Rightarrow$ There is no significance difference b/w population variance i.e. $\sigma_1^2 = \sigma_2^2$.

Q. Two independent samples of size 7 and 6 had the following values:-

A	28	30	32	33	31	29	34
B	29	30	30	24	27	28	

Examine wheather the samples have been drawn from normal population having the same variance.

Soln:- $H_0 : \sigma_1^2 = \sigma_2^2$.

χ^2 - test [Chi-square test]

$$\chi^2 = \Sigma\left[\frac{(O_i - E_i)^2}{E_i}\right]$$, where O_i - Observed frequency
E_i - Expected frequency

For simple data, $E_i = \frac{\Sigma O_i}{n}$

Dof, $\nu = n-1$

NOTE:- i) In case of Binomial Distribution,
dof, $\nu = n-1$

ii) In Case of Poisson Distribution,
$\nu = n-2$

iii) In Case of Normal Distribution,
$\nu = n-3$

Q. The following table gives the no. of accidents that took place in an industry during various ways B. Test if accidents are uniformally distribution over the B.

Days	Mon	Tue	Wed	Thurs	Fri	Sat
No. of Accidents	14	18	12	11	15	14

Soln:- H_0: Accidents are uniformally distributed.

$E_i = \frac{\Sigma O_i}{n} = \frac{84}{6} = 14$

O_i	14	18	12	11	15	14
E_i	14	14	14	14	14	14
$O_i - E_i$	0	4	-2	-3	1	0

$\Sigma\frac{(O_i - E_i)^2}{E_i} = \cancel{16+4+9+1} = \cancel{30}$ 2.142

$\cancel{\chi}^2 = \frac{30}{14} = 2.142$

$\chi^2 = 2.142$

$\nu = 5$

$\chi^2(5\%, 5) = 11.07$

Conclusion: $\chi^2_{cal} < \chi^2_{tab}$

Hence, Accidents are uniformally distributed.

Q. A die is thrown 276 times and results of these throws are given below:-

No. appeared on die	1	2	3	4	5	6
Frequency	40	32	29	59	57	59

Test wheather the die is biased or not.

H_0: Dia is unbiased.

Testing of Hypothesis:-

χ-Test- $\chi^2 = \Sigma \left[\frac{(O_i - E_i)^2}{E_i} \right]$

$E_i = \frac{\Sigma O_i}{n} = \frac{276}{6} = 46$

O_i	40	32	29	59	57	59
E_i	46	46	46	46	46	46
$O_i - E_i$	-6	-14	-17	13	11	13
$(O_i - E_i)^2 =$	36	196	289	169	121	169

$\frac{(O_i - E_i)^2}{E_i} =$ 0.782 4.26 6.28 3.67 2.63 3.673

$\chi^2 = 21.30$

$\chi^2_{(5\%, 5)} = 11.07$

$\chi^2_{cal} > \chi^2_{tab}$

Dia is Biased.

Q. A survey of 320 families with 5 children shows the following distribution.

No. of boys and girls	5 B 0 G	4 B 1 G	3 B 2 G	2 B 3 G	1 B 4 G	0 B 5 G	
No. of families	18	56	110	88	40	8	Total = 320

Test the hypothesis that the distribution is ~~normally~~ Binomial distributed and male / female are equally probable.

H_0 :- Male and Female births are equally probable.

Testing of Hypothesis :-

$$P(X=r) = {}^nC_r p^r q^{n-r}, \quad p = q = \frac{1}{2}$$

$r=0$, $P(X=0) = {}^5C_0 \left(\frac{1}{2}\right)^0 \left(\frac{1}{2}\right)^5 = \frac{1}{32} \times 320 = 10$

$r=1$, $P(X=1) = {}^5C_1 \left(\frac{1}{2}\right)^1 \left(\frac{1}{2}\right)^4 = \frac{5}{32} \times 320 = 50$

$P(X=2) = {}^5C_2 \left(\frac{1}{2}\right)^2 \left(\frac{1}{2}\right)^3 = \frac{10}{32} \times 320 = 100$

$P(X=3) = {}^5C_3 \left(\frac{1}{2}\right)^3 \left(\frac{1}{2}\right)^2 = \frac{10}{32} \times 320 = 100$

$P(X=4) = {}^5C_4 \left(\frac{1}{2}\right)^4 \left(\frac{1}{2}\right)^1 = \frac{5}{32} \times 320 = 50$

$P(X=5) = {}^5C_5 \left(\frac{1}{2}\right)^0 \left(\frac{1}{2}\right)^5 = \frac{1}{32} \times 320 = 10$

E_i	10	50	100	100	50	10
O_i	18	56	110	88	40	8
$O_i - E_i$	8	6	10	−12	−10	−2
$(O_i - E_i)^2$	64	36	100	144	100	4
$\frac{(O_i - E_i)^2}{E_i}$	6.4	0.72	1	1.44	2	0.4

$$\boxed{\chi^2 = 11.96}$$

$\nu = n - 1 = 6 - 1 = 5$

level of significance at 5%,

$\chi^2_{tab} = 11.070$,

$\because \chi^2_{cal} > \chi^2_{tab}$ Null Hypothesis is rejected.

χ^2 test as a Test of Independence:-

Observed Frequency :-

			Total
	a O_1	b O_2	$a+b$
	c O_3	d O_4	$c+d$
Total	$a+c$	$b+d$	$N = a+b+c+d$

$$O_1, \quad E_1 = \frac{(a+b)(a+c)}{N}$$

$$O_2, \quad E_2 = \frac{(a+b)(b+d)}{N}$$

$$O_3, \quad E_3 = \frac{(c+d)(a+c)}{N}$$

$$O_4, \quad E_4 = \frac{(c+d)(b+d)}{N}$$

Q. A survey of 320 families ...

Date __/__/__

Page No.

Contigency Table -

$\frac{(a+b)(a+c)}{N}$	$\frac{(a+b)(b+d)}{N}$
$\frac{(c+d)(a+b)}{N}$	$\frac{(c+d)(b+d)}{N}$

2×2

$\downarrow \quad \downarrow$

$p \quad q$

$$D.o.f,\ V = (p-1)(q-1) = 1$$

Q. From the following table regarding the colour of eyes of father and son. Test if the colour of son's eye is associated with that of father

Eye colour of father (B) / Eye colour of son (A)

	Light	Not light
Light	471	51
Not light	148	230

Soln:

B \ A	Light	Not light	Total
Light	359.02	162.98	522
Not light	259.98	118.02	378
	619	281	900

$\chi^2 = 266.338$

$\chi^2_{tab} = 3.841$

$\chi^2_{cal} > \chi^2_{tab} \Rightarrow$ Ho is rejected.

ANOVA (Analysis of Variance)

* ANOVA is statistical tool that can be used for comparison among more than two groups.
* It is used for testing the hypothesis of equality of more than two normal population means.
* $$F\text{-Statistics} = \frac{\text{Variance b/w the samples}}{\text{Variance within the sample}}$$
* Variance b/w the sample > Variance within the sample.

One way ANOVA -

Used to analyse the effects of one independent factor on dependent variables.

Assumptions:- i) Each sample is randomly drawn.

ii) Population from which the samples are drawn are normally distributed.

iii) Each sample is independent of other sample.

iv) Each of population has same variance and identical means.

Procedure -

X_1	X_1^2	X_2	X_2^2	X_3	X_3^2
0					
0					
0					
0					
ΣX_1	ΣX_1^2	ΣX_2	ΣX_2^2	ΣX_3	ΣX_3^2

i) GT (Grand Total) $= \Sigma X_1 + \Sigma X_2 + \Sigma X_3$

ii) Correction Factor, $C.F = \frac{(G.T)^2}{n}$; $n \rightarrow$ Total no. of items

iii) Sum of squares b/w samples (SSC) =

$$\frac{(\Sigma X_1)^2}{n_1} + \frac{(\Sigma X_2)^2}{n_2} + \frac{(\Sigma X_3)^2}{n_3} - C.F$$

dof $= c-1$; columns Here, $c = 3$

iv) Total Sum of Squares (SST)

$$= \Sigma X_1^2 + \Sigma X_2^2 + \Sigma X_3^2 - C.F$$

dof $= n-1$; n - Total no of items Here ($n=12$)

v) Sum of squares within the sample (SSE) → Error

$\Rightarrow$ SST = SSC + SSE

$$\boxed{SSE = SST - SSC}$$

dof $= n-c$

vi) Mean sum of squares b/w the samples (Variance)

$$MSC = \frac{SSC}{df} = \frac{SSC}{c-1}$$

vii) Mean Sum of square within sample (Variance)

$\Rightarrow MSE = \frac{SSE}{n-c}$

~~F-Stat~~

$\boxed{\text{F-Statistics} = \frac{MSC}{MSE}}$, $MSC > MSE$

Level of Significance

* Aanova Table -

Source of Variation	~~Source~~ Sum of squares	df	Mean Sum of squares	F
B/w the sample	SSC	$c-1$	$MSC = \frac{SSC}{c-1}$	$F = \frac{MSC}{MSE}$; $MSC > MSE$
within the sample	SSE	$n-c$	$MSE = \frac{SSE}{n-c}$	
Total	SST	$n-1$		

Q. It is desired to compare three hospitals with regards to the number of deaths per month. A sample of death records of each hospital and the number of death was as given below - From these data suggest ~~the~~ a difference in the no. of deaths per months among three hospitals.

5% Level of Significance

A	B	C
3	6	7
4	3	3
3	3	4
5	4	6
0	4	5

Null Hypothesis (H_0):. There is no significance diff b/w no of death per month among three hospital

Alternate Hypothesis (H_1): There is significance diff in the no. of death per month among three hospital.

Soln:-

X_A	X_A^2	X_B	X_B^2	X_C	X_C^2
3	9	6	36	7	49
4	16	3	9	3	9
3	9	3	9	4	16
5	25	4	16	6	36
0	0	4	16	5	25

$\Sigma X_A = 15$ $\Sigma X_A^2 = 59$ $\Sigma X_B = 20$ $\Sigma X_B^2 = 86$ $\Sigma X_C = 25$ $\Sigma X_C^2 = 135$

(1) $G.T = \Sigma X_A + \Sigma X_B + \Sigma X_C$

$= 15 + 20 + 25 = 60$

(2) $C.F = \frac{(G.T)^2}{n} = \frac{3600}{15} = 240$

③ ~~Sum~~ $SSC = \frac{(\Sigma x_A)^2}{n_1} + \frac{(\Sigma x_2)^2}{n_2} + \frac{(\Sigma x_3)^2}{n_3} - C.F$

$= \frac{225}{5} + \frac{400}{5} + \frac{625}{5} - 240$

$= (45 + 80 + 125) - 240$

$= 250 - 240 = 10$ $df = c - 1 = 2$

④ $SST = \Sigma x_1^2 + \Sigma x_2^2 + \Sigma x_3^2 - CF$

$= (59 + 86 + 135) - 240$

$= 40$

$df = n - 1 = 15 - 1 = 14$

⑤ $SSE = SST - SSC$

$= \cancel{240}\ 40 - 10 = \cancel{230}\ 30$

$df = n - c = 15 - 3 = 12$

⑥ $MSC = \frac{SSC}{c-1} = \frac{10}{2} = 5$

⑦ $MSE = \frac{SSE}{n-c} = \frac{30}{12} = 2.5$

$F\text{-statistics} = \frac{MSC}{MSE}$

$= \frac{5}{2.5} = 2$

$F_{cal} = 2$

DATE.
PAGE.

Anova Table -

Source of Samples	Sum of squares	df	Mean sum of squares	F
B/w the sample	$SSC = 10$	$c-1 = 2 = \nu_1$	$MSC = \frac{SSC}{c-1} = 5$	$F = \frac{MSC}{MSE} = 2$
Within the sample	$SSE = 30$	$12 = \nu_2$	$MSE = \frac{SSE}{n-1} = 2.5$	
Total	$SST = 40$	$n-1 = 14$		

$F_{tab} = 3.89$

$\because F_{cal} < F_{tab}$

$\Rightarrow$ There is no significance difference b/w no. of deaths per month among three hospital.

DATE.
PAGE.

Moment Generating Function (MGF):-

Let x be a random variable then MGF is represented by

$$M_x(t) = \sum_x e^{tx} P(x) \quad , x \rightarrow \text{Discrete r.v}$$

$$M_x(t) = \int_x e^{tx} P(x)\, dx \quad ; x \rightarrow \text{continuous r.v}$$

* Expected value of x.

$E(x)$ = first moment about origin $= V_1$
$E(x^2)$ = Second moment about origin $= V_2$
$E(x^3)$ = Third moment about origin $= V_3$

$$E(x) = V_1 = \frac{d\, M_x(t)}{dt}\Big]_{t=0}$$

$$E(x^2) = V_2 = \frac{d^2\, M_x(t)}{dt^2}\Big]_{t=0}$$

$$E(x^3) = V_3 = \frac{d^3\, M_x(t)}{dt^3}\Big]_{t=0}$$

$$E(x^r) = V_r = \frac{d^r\, M_x(t)}{dt^r}\Big]_{t=0}$$

* $V_1 = \bar{x}$ * μ_2 = variance

Q. Find the MGF of distribution $f(x) = P(x) = \frac{1}{c} e^{-x/c}$, where $0 < x < \infty$, $c > 0$.
Hence find mean and variance.

Soln:

$$M_x(t) = \int_0^\infty e^{tx} \frac{1}{c} e^{-x/c} \, dx$$

$$= \frac{1}{c} \int_0^\infty e^{tx} \cdot e^{-x/c} \, dx$$

$$= \frac{1}{c} \int_0^\infty e^{\left(t - \frac{1}{c}\right)x} \, dx$$

$$= \frac{1}{1-ct} = (1-ct)^{-1}$$

$$= 1 + ct + (ct)^2 + (ct)^3 + \cdots$$

$$V_1 = \frac{d}{dt} M_x(t) \Big]_{t=0} \qquad V_2 = \frac{d^2}{dt^2} M_x(t) \Big]_{t=0}$$

$$= c \qquad\qquad = 2c^2$$

Mean $= c$

$$\mu_2 = V_2 - \bar{x}^2$$

$$= 2c^2 - c^2 = c^2$$

Variance $= c^2$

DATE :

PAGE :

Q. Find the MGF of Binomial Distribution given by $P(x) = {}^nC_x p^x q^{n-x}$. Also find the first and second moment about the mean.

$$M_x(t) = \sum_{x=0}^{n} e^{tx}\, {}^nC_x p^x q^{n-x}$$

$$= \sum_{x=0}^{n} {}^nC_x (e^t p)^x q^{n-x}$$

$$= {}^nC_0 (e^t p)^0 q^n + {}^nC_1 (e^t p)^1 q^{n-1} + \cdots$$

$$= (a+b)^n$$

$$= (q + e^t p)^n$$

$$M_x(t) = (q + e^t p)^n$$

$$v_1 = \frac{d}{dt} M_x(t) \Big]_{t=0} = \frac{d}{dt}(q + pe^t)^n$$

$$= n(q+pe^t)^{n-1} \cdot pe^t \Big]_{t=0}$$

$$= n(q+p)^{n-1} \cdot p \qquad [\because p+q=1]$$

$$= np$$

$$\boxed{v_1 = np = \text{mean}}$$

$$v_2 = \frac{d}{dt}\left[n(q+pe^t)^{n-1} \cdot pe^t\right]$$

$$= np\left[(q+pe^t)^{n-1} \cdot e^t + e^t (q+pe^t)\right]$$

$$= \cancel{np\left[e^t(q+pe^t)\right]\{1+} = np(np+1-p)$$

$$np\{(q+p)^{\cancel{n-1}} + \cancel{(q+p)}\}$$

$$\mu_2 = v_2 - (\bar{x})^2$$

$$= np(np+1-p) - (np)^2$$

$$= np(1-p) = npq$$