PAGE NO :
DATE : / /

Unit - IV

Sampling and Statistical Inference.

Sampling is a technique of selecting a subset of a population to make statistical inference from it and estimate characteristics of the entire population.

Eg:- exit poll, vaccine ~~polio~~ trials.

→ Defns.

1) Population:

It is the set of all aggregate objects under study. (denoted by N)

2) Sample:

It is the subset of the population that is considered to study the behaviour of the population. (sample size denoted by n)

If $n \geq 30$, it is considered to be large sample.

* The statistical constants of popln such as mean, variance etc are called parameters.

* The statistical constants of samples which are used to estimate the parameters are called statistics. Eg: $\bar{x}$, s^2

Sampling can be done in diff. ways but it is important to have a random

Sample.

i) With replacement
ii) Without replacement

→ Sampling Distribut"

It is the frequency distribut" of a statistic over many random samples from a single population.

If we have diff. samples of size n from a population of size N, for each of these samples we can compute mean, variance etc. They will not be same. If we grp these acc. to their freqs. then the freq distribut" so generated is called the sampling distribut". In particular, we can have sampling distribut" of mean, var. etc. The std. deviation of the sampling distribut" is called the std. error.

→ Questions:

1) A popl" consists of nos {1, 2, 3} form a sampling distribut" of the mean for random sample of size 2 with replacement.

$N = 3$, $\mu = 2$, $\sigma^2 = \frac{1}{N}\Sigma(x-\mu)^2$

$\sigma^2 = \frac{1}{3}\{1 + 1\}$

$\sigma^2 = 2/3$

$S = \{(1,1), (1,2), (1,3), (2,1), (2,2), (2,3), (3,1), (3,2), (3,3)\}$

$Mean_s = \{1, 1.5, 2, 1.5, 2, 2.5, 2, 2.5, 3\}$

x_i	$f(x_i)$	$x_i f_i$	$x - \bar{x}$	$(x-\bar{x})^2$	$f(x-\bar{x})^2$
1	1	1	−1	1	1
1.5	2	3	−0.5	0.25	0.5
2	3	6	0	0	0
2.5	2	5	0.5	0.25	0.5
3	1	3	1	1	1
	9	18			3

Mean, $\bar{x} = \frac{\Sigma x_i f_i}{\Sigma f_i} = \frac{18}{9} = 2$

Variance → $s^2 = \frac{1}{\Sigma f_i}\Sigma f_i (x_i - \bar{x})^2$

$s^2 = \frac{1}{9} \times 3 = 1/3$.

$s^2 = 0.33$

$s^2 = \frac{\sigma^2}{n} = \frac{2}{3} \times \frac{1}{2}$

2) Central Limit Theorem

In a popl" N with mean μ & variance σ^2, if we take sufficiently large random samples with replacement, then the dist" of the sample means will be normally distributed.

→ Defⁿˢ.

1) Hypothesis: The assumptⁿ that we make regarding the parameter of the populatⁿ.

2) test of significance/hypothesis
An imp aspect of sampling theory is to assert that the parameter of the popⁿ is the same as the statistic obtained of the random process. The process that decides whether to accept the hypothesis or not is called test of significance

In order to arrive at a decision, we have to make certain assumptions known as hypotheses. The hypo is accepted or rejected based on some statistical test. We decide whether the sample statistic is significantly different from the parameter at a desired level of significance. Hence, these tests are called, tests of Sign

There are 2 types of hypothesis.

1) Null hypothesis
The hypo which assumes that there is no significant differenc b/w sample statistic and the parameter of the populatⁿ It is denoted by H_0

2) Alternate hypo
Under this hypo, we assume that there is a significant differenc b/w sample & populatⁿ statistics

→ Type 1 and Type 2 errors.

Type 1: When a true hypo is rejected, it is called type 1 error.

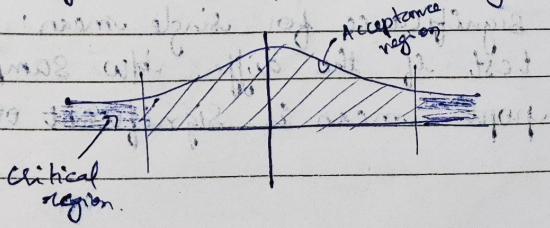
type 2: When a false hypo is accepted, it is called type 2 error.

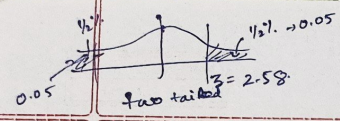
Hypothesis	Decision	
	Accepted	Rejected
True	✓	type 1
False	type 2	✓

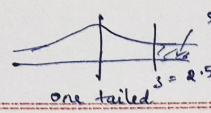
→ Level of Significance. (los)
The prob level below which we reject a hypo is called the level of significance.
It can be 1%, 2%, 5%, 10% etc.
In general, '1%' or 5% los is considered

→ Critical region & Acceptance region.

C-region is the region which corresponds to the reject" of hypo
A-region → acceptance of hypo



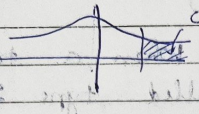




PAGE NO :
DATE : / /

→ One tailed and two tailed tests

* test on statistical hypotheses where the altesnate hypo is one sided is called one tailed test.



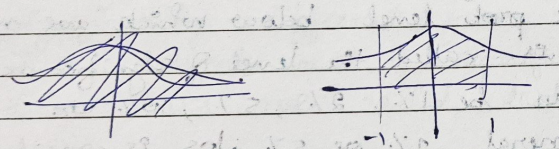
→ Confidence Interval.

Let μ_s and σ_s be the mean & S.D of the sampling distribut^n of a test statistic S. If the sampling dist. of S is normal then, S lies in $(\mu_s - \sigma_s, \mu_s + \sigma_s)$ what % of times.

$$z = \frac{x - \mu}{\sigma}$$

$$z_1 = \frac{\mu - \sigma - \mu}{\sigma} \qquad z_2 = \frac{\mu + \sigma - \mu}{\sigma}$$

$z_1 = -1$ $\qquad z_2 = 1$ $\qquad \sigma_s$.



$A(3) = 0.3413 + 0.3143$

$= 0.6826$

$= 68.26\%$

→ Test For large Samples. (z-test).

• Test of significance for single mean:
(To test if the diff b/w sample mean and popl^n mean is significant or not)

PAGE NO :
DATE : / /

Suppose we have a popl^n of size N with mean μ and variance σ^2. Let us draw a sample of size n from it. Let $\bar{x}$ be the sample mean. The std. error of mean of a random sample of size n is given by,

$$\text{S.E of mean} = \frac{\sigma}{\sqrt{n}}$$

• Under the null hypothesis, there is no diff. b/w std mean & sample mean.

$H_0 : \bar{x} = \mu$

• The test statistic is $z = \frac{\bar{x} - \mu}{\sigma/\sqrt{n}}$

• If σ is not known, the test statistic,

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

where s is the S.D. of the sample.

• If l.o.s is α and z_α is the critical value, then

$$-z_\alpha < |z| = \left| \frac{\bar{x} - \mu}{\sigma/\sqrt{n}} \right| < z_\alpha$$

• The limits of popl^n mean, μ are given by

$$\bar{x} - z_\alpha \frac{\sigma}{\sqrt{n}} < \mu < \bar{x} + z_\alpha \frac{\sigma}{\sqrt{n}}$$

- At 5% l.o.s, 95% confidence limits are

$$\bar{x} - 1.96\frac{\sigma}{\sqrt{n}} < \mu < \bar{x} + 1.96\frac{\sigma}{\sqrt{n}}$$

- At 1% l.o.s, 99% confidence limits are

$$\bar{x} - 2.58\frac{\sigma}{\sqrt{n}} < \mu < \bar{x} + 2.58\frac{\sigma}{\sqrt{n}}$$

→ Problems:-

1. A normal popn has a mean of 6.8 and S.D of 1.5. A sample of 400 members gave a mean of 6.75. Is the difference significant.

Null hypothesis, $H_0 : \bar{x} = \mu$

$\mu = 6.8$, $\sigma = 1.5$, $n = 400$, $\bar{x} = 6.75$

$$z = \frac{\bar{x} - \mu}{\sigma/\sqrt{n}} = \frac{6.75 - 6.8}{1.5/20}$$

$$z = -0.66667 \quad (2.666)$$

$|z| < z_\alpha$ for 5% & 1% los

⇒ H_0 is accepted

(2) if $\bar{x} = 6.95$, $z = 2$ $\quad 2 > 1.96$

$z > z_\alpha$ for 5% los

H_0 is rejected.

and as $z_\alpha < z_\alpha$ for 1% l.o.s $\quad (2 < 2.58)$

H_0 is accepted.

2. The mean weight for a random sample of size 100 is 64 g. The S.D of the weight dis of popn is 3g. Can we say that the mean weight of the popln is 66 g is at 5% los? Also set up 95% and 99% confidence interval for the mean weight of the population.

$n = 100$, $\bar{x} = 64$, $\sigma = 3$, $\mu = 66$

$$z = \frac{\bar{x} - \mu}{\sigma/\sqrt{n}} = \frac{64 - 66}{3/10}$$

$$z = -6.6667$$

$|z| > z_\alpha$ for 5% & 1% los.

⇒ H_0 is not accepted.

for 95% confidence interval, $z < 1.96$

$$1.96 > \frac{64 - \mu}{0.3}$$

$$0.588 > 64 - \mu$$

$$\mu > 64 - 0.588$$

$$\mu > 63.412$$

95% confidence limits are

$64 - 1.96 \times 0.3 < \mu < 64 + 1.96 \times 0.3$

$63.412 < \mu < 64.588$

99% confidence limits are

$64 - 2.58 \times 0.3 < \mu < 64 + 2.58 \times 0.3$

$63.22 < \mu < 64.774$

8. $n = 500$, $\bar{x} = 3.4$, $\mu = 3.25$

$\sigma = 1.71$

$$z = \frac{\bar{x} - \mu}{\sigma/\sqrt{n}} = \frac{3.4 - 3.25}{1.71/\sqrt{500}}$$

$$= \frac{0.15 \times 22.3607}{1.71}$$

$$1.96 < = 1.9614 < 2.58$$

5% 1%

H_0 : accepted at 1% los.
rejected at 5% los

→ Test for Difference b/w two means

Let $\bar{x}_1, \bar{x}_2$ be the means of 2 samples of sizes n_1 and n_2, from populations with means μ_1 and μ_2 and S.D is σ_1 and σ_2. As the two samples are independent for large values of n_1 and n_2. $\bar{x}_1 - \bar{x}_2$ is approximately normally distributed and then,

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

Under the Null Hypothesis:

$\mu_1 = \mu_2$, then $z_E = \dfrac{\bar{x}_1 - \bar{x}_2}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}}$

Note:

1) If σ_1^2 and σ_2^2 are unknown, then we can use the sample sd's at

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

2) If the 2 populations have a common variance σ^2, then

$$z = \frac{\bar{x}_1 - \bar{x}_2}{\sigma\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}}$$

→ Questions.

1) Two random samples of 100 students from two schools A and B were drawn. The CGPA of students from A had mean 2.82 with sd. 0.63 and from

school B had mean 2.43 with s.d. 0.65. Does the data indicate any difference in the mean CGPA of students from the schools?

→ $n_1 = n_2 = 100$ $\quad \bar{x}_1 = 2.82, \quad s_1 = 0.63$

$\bar{x}_2 = 2.43, \quad s_2 = 0.65$

$$z = \frac{2.82 - 2.43}{\sqrt{\frac{0.63^2}{100} + \frac{0.65^2}{100}}}$$

$$z = \frac{0.39 \times 10}{\sqrt{0.3969 + 0.4225}}$$

$$z = \frac{39}{0.9052} = 4.3084$$

$z > z_\alpha$ for 1% & 5%

$\Rightarrow H_0$ is rejected.

2) The mean heights in a large samples of size 1000 and 2000 are 67.5 inches and 68 inch resp. Can the samples be regarded as drawn from normal popn with s.d 2.5 inches.

$n_1 = 1000 \quad \bar{x}_1 = 67.5 \quad \sigma = 2.5$

$n_2 = 2000 \quad \bar{x}_2 = 68.0$

$H_0 : \mu_1 = \mu_2$

$$z = \frac{\bar{x}_1 - \bar{x}_2}{\sigma\sqrt{\frac{1}{n_1} - \frac{1}{n_2}}} = \frac{67.5 - 68}{2.5\sqrt{\frac{1}{1000} - \frac{1}{2000}}}$$

$$= \frac{-0.5}{2.5\sqrt{\frac{1000}{2\times10^6}}}$$

$$= -5.163 > z_\alpha \text{ for } 1\% \text{ \& } 5\%.$$

$\therefore H_0$ is Rejected.

3) $n_1 = n_2 = 100 \quad \bar{x}_1 = 1400$ hrs $\quad s_1 = 200$ hrs

$\bar{x}_2 = 1200$ hrs. $\quad s_2 = 100$ hrs.

$$z = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}} = \frac{1400 - 1200}{\sqrt{\frac{200^2}{100} + \frac{100^2}{100}}}$$

$$= \frac{200}{\sqrt{400 + 100}} = \frac{200}{\sqrt{500}} = \frac{200}{22.3607}$$

$$= 8.9443 > z_\alpha \text{ for } 1\% \text{ \& } 5\%.$$

$\therefore H_0$ is Rejected.

→ Test of Significance for Proportion.

- For Single Proportion.

$$z = \left(\frac{x-\mu}{\sigma}\right) = \frac{x - np}{\sqrt{npq}}$$

where,
$x \rightarrow$ no. of successes in n independent trials
$p \rightarrow$ prob. of success in each trial
$q = 1 - p$.

Questions

1) A die is thrown 9000 times and a throw of 3 or 4 is observed 3240 times. Can the die be regarded as unbiased.

H_0: die is unbiased.

$n = 9000 \quad p = \frac{2}{6} = \frac{1}{3} \quad q = \frac{2}{3}$

$x = 3240$

$$z = \frac{3240 - 9000(1/3)}{\sqrt{9000 \times 1/3 \times 2/3}}$$

$$= \frac{240}{\sqrt{2000}}$$

$$= \frac{240}{44.7213} = 5.59 > z_\alpha \text{ of } 1\%$$

H_0 is rejected. Die is biased.

2) In a sample of 1000 people in a state, 540 are rice eaters & the rest are wheat eaters. Can at 1% & 5% l.o.s that both rice & wheat are equally popular in the state.

$p = \frac{1}{2}, \quad q = \frac{1}{2} \quad n = 1000$

for $x = 540$

H_0: rice & wheat are equally popular.

$$z = \frac{540 - 1000 \times 1/2}{\sqrt{1000 \times \frac{1}{2} \times \frac{1}{2}}}$$

$$= \frac{40}{\sqrt{250}} = \frac{40}{15.8113}$$

$$z = 2.5298$$

$z > z_\alpha$ for 5% los $\therefore H_0$ is rejected.

$z < z_\alpha$ for 1% los

$\therefore H_0$ can be accepted.

3. A coin was tossed 400 times & heads turned up 225. Test the hypothesis that the coin is unbiased

$n = 400, \quad x = 225 \quad p = 1/2, \quad q = 1/2$

4: A manufactures claims that only 4% of his products are defected. A random sample of 600 sample contains 36 defectives. Test the claim of the manufacturer.

$n = 600$, $x = 36$, $p = 0.04$, $q = 0.96$

$$z = \frac{36 - 600(0.04)}{\sqrt{600 \times 0.04 \times 0.96}}$$

$$z = \frac{36 - 24}{\sqrt{23.04}}$$

$$z = \frac{36 - 24}{4.8} = \frac{12}{4.8}$$

$z = 2.5 > z_c$ for 5% LOS

$2.5 < z_c$ for 1% LOS.

∴ The claim of manufacturer is rejected for 5% LOS & accepted at 1% LOS.

5. A new medicine was given to 50 patients and found to be effective in 37 patient. Test the hypothesis that the medicine is effective in 80% patients.

$n = 50$, $x = 37$, $p = 0.8$, $q = 0.2$

$$z = \frac{37 - 50(0.8)}{\sqrt{50 \times 0.8 \times 0.2}}$$

$$z = \frac{37 - 40}{\sqrt{8}}$$

$$z = \frac{-3}{2.8284}$$

$z = -1.0606$

$z < z_c$ for both 5% & 1% LOS

∴ The medicine is affective in 80% patients

→ Test of significance for small samples ($n < 30$)

∴ Degree of freedom:

It is the no. of independent values that a statistical analysis can estimate.

i.e. d.f is the no of values that are free to vary as we estimate parameters.

For sample size less than 30, we do not assume that the sampling distribut" is normal and the values given by the sample are sufficiently close to the populat" values.

→ Student's t-test

(Test of significance of the mean of a small sample).

Let x_i $(i = 1, 2 \ldots n)$ be a random sample of size n from a normal population with mean μ and variance σ^2.

Then $t = \frac{\bar{x} - \mu}{s/\sqrt{n}}$ where $s^2 = \frac{1}{n-1} \sum_{i=1}^{n} (x_i - \bar{x})^2$

s^2 is an unbiased estimate of the population variance σ^2 and it follows t-distribution with $(n-1)$ d.f.

→ Questions:

1) A random sample of size 16 has a mean 53. The sum of squares of the deviation from mean is 135. Can the sample be regarded as taken from a population of mean 56? Also obtain 95% and 99% confidence limits for the mean of the mean of population.

H_0: populat^n of mean is 56., $\bar{x} = \mu$.

$n = 16 \quad \bar{x} = 53 \quad \mu = 56 \quad \sum (x_i - \bar{x})^2 = 135$

$s^2 = \frac{1}{15} \times 135$ $\qquad s^2 = \frac{1}{n-1} (x_i - \bar{x})^2$

$s = 3$

$t = \frac{\bar{x} - \mu}{s/\sqrt{n}} = \frac{53 - 56}{3/4}$

$= -4$

$t_c (5\%, 15\,df) = 2.131$

$t_c (1\%, 15\,df) = 2.947$

$|t| = 4 > t_c$ (for 1% & 5% los).

$\therefore H_0$ is rejected for both 1% & 5%.

i.e. $\bar{x} \neq \mu$.

→ Confidence Intervals

:) 95% CI → $-2.131 < t < 2.131$

$-2.131 < \frac{\bar{x} - \mu}{s/\sqrt{n}} < 2.131$

$-\frac{s}{\sqrt{n}}(2.131) < \bar{x} - \mu < 2.131 \frac{s}{\sqrt{n}}$

$\bar{x} + 2.131 \frac{s}{\sqrt{n}} > \mu > \bar{x} - 2.131 \frac{s}{\sqrt{n}}$

$\mu \in \left(\bar{x} \pm 2.131 \frac{s}{\sqrt{n}}\right)$

$\mu \in \left(53 \pm 2.131 \times \frac{3}{4}\right)$

$\mu \in (51.9017, 54.5932)$

:) 99% CI →

$\mu \in \left(\bar{x} \pm 2.947 \frac{s}{\sqrt{n}}\right)$

$\mu \in \left(53 \pm 2.947 \times \frac{3}{4}\right)$

$\mu \in (50.7897, 55.2102)$

2) A sample of 20 items has mean 42 units and s.d 5 units. Test the hypothesis that it is a random sample from a normal population with mean 45 units.

$H_0 : \bar{x} = \mu$

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

$n = 20, \quad \bar{x} = 4, \quad \mu = 45, \quad S = 5 \Rightarrow S^2 = 25$

$S^2 = \frac{1}{n} \Sigma (x_i - \bar{x})^2, \quad s^2 = \frac{1}{n-1}(x_i - x^2)$

$$s^2 = \frac{n}{n-1} S^2$$

$s^2 = \frac{20}{19} \times 25 = 26.3158$

$s = 5.1299$

$t = \frac{\bar{x} - \mu}{s/\sqrt{n}} = \frac{4 - 45}{5.1299/\sqrt{20}}$

$t = -2.6153.$

$t > t_c$ for 5% l.o.s.
$t < t_c$ for 1% l.o.s.

$\therefore H_0$ is rejected at 5% los & accepted at 1% l.o.s.

95% CI $\rightarrow \quad \mu \in \left(\bar{x} \pm 2.131 \frac{s}{\sqrt{n}}\right)$

$\mu \in \left(4 \pm 2.131\left(\frac{5.1299}{\sqrt{20}}\right)\right)$

$\mu \in (4 \pm 2.444)$

$\mu \in (1.556, 6.444)$

99% CI $\rightarrow \quad \mu \in \left(\bar{x} \pm 2.947\left(\frac{5.1299}{\sqrt{20}}\right)\right)$

$\mu \in (4 \pm 3.3804)$

$\mu \in (0.6196, 7.3804)$

3. 9 items of a sample have the values 45, 47, 50, 52, 48, 47, 49, 53, 51. Does it differ significantly from the assumed mean of 47.5.

$H_0 : \mu = \bar{x}$

$n = 9, \quad \mu = 47.5 \quad \bar{x} = 49.1111$

$s^2 = \frac{1}{n-1} \Sigma (x_i - x^2)^2 = \frac{1}{8} \times 54.8765$

$s^2 = 6.8596$

$s = 2.619$

$t = \frac{\bar{x} - \mu}{s/\sqrt{n}} = \frac{49.1111 - 47.5}{2.619/\sqrt{9}} = \frac{1.6111}{0.873}$

$t = 1.845 < t_c$ for 5% & 1% los

$\therefore H_0$ is accepted for both 5% & 1% los

$t_c(5\%, 8\ d.f) = 2.306$

$t_c(1\%, 8\ d.f) = 3.355$

4. The heights of 10 boys in a class is 70, 67, 62, 68, 61, 68, 70, 64, 64, 66 inches. Is it reasonable to believe that the avg height of boys is 64 inches

5. 10 students from a school have IQ 70, 120, 110, 101, 88, 83, 95, 98, 107, 100 ~~does~~ this data support the assumption of avg ~~school~~ school Iq to be 100?

$n = 10, \quad \mu = 100, \quad \bar{x} = 97.2$

$$s^2 = \frac{1}{n-1}\sum (x_i - \bar{x})^2 = \frac{1}{9}\left(\sum x_i - x^2\right)$$

$$s = 14.2735$$

$$t = \frac{\bar{x} - \mu}{s/\sqrt{n}} = \frac{97.2 - 100}{14.2735/\sqrt{10}} = \frac{-2.8}{4.5136}$$

$$t = 0.6203 < t_c\ (5\%\ 10s)$$

$t_c(5\%, 9\ df) = 2.26$

$H_0 : \bar{x} = \mu$ is accepted)

→ Two Sample t-test

For 2 samples of size n_1 & n_2

$$t = \frac{(\bar{x} - \bar{y}) - (\mu_1 - \mu_2)}{\sqrt{s_x^2\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}}$$

$$s_x^2 = \frac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2}{(n_1 + n_2 - 2)}$$

$$= \frac{1}{n_1 + n_2 - 2}\left\{\sum (x_i - \bar{x})^2 + \sum (y_i - \bar{y})^2\right\}$$

$H_0 : \mu_1 = \mu_2$

→ Questions :-

1) Two samples give the foll^n data

Sample A :

44 44 56 46 47 47 58 53 49 55 46 57 51

Sample B :

35 38 35 29 40 39 32 41 42 30 31 39

Test if they come from populations with the same mean.

→ $H_0 : \mu_1 = \mu_2$; $n_1 = 13, \quad n_2 = 12$

$\bar{x} = 50.2307 \quad \bar{y} = 35.9166$

$$s_x^2 = \frac{1}{n_1 + n_2 - 2}\left\{306.307 + 226.915\right\} = 23.17$$

$t \in (n_1 + n_2 - 2)$ df

$$t = \frac{(\bar{x} - \bar{y}) - (\mu_1 - \mu_2)}{\sqrt{S_x^2 \left(\frac{1}{n_1} + \frac{1}{n_2}\right)}}$$

$$t = \frac{50.25 - 35.92}{\sqrt{23.17\left(\frac{1}{13} + \frac{1}{12}\right)}} = 7.41$$

$t_c(5\%, n_1 + n_2 - 2 \text{ df}) = t_c(5\%, 23 \text{ df}) = 2.069.$

$t > t_c \Rightarrow H_0$ is rejected

2. Samples of 2 types of electric bulbs were tested for length of life & the foll" data was obtained.

	Type I	Type II
Sample no.	$n_1 = 8$	$n_2 = 7$
Mean	$\bar{x} = 1234$	$\bar{y} = 1036$
sd	$s_1 = 36$	$s_2 = 40$

From data, can you infer that type 1 is better than type 2?

$$\Sigma(x_i - \bar{x})^2 = n_1 s_1^2 = 8 \times 36^2 = 10368$$

$$\Sigma(y_i - \bar{y}) = n_2 s_2^2 = 7 \times 40^2 = 11200$$

$$s_x^2 = \frac{1}{8+7-2}\{10368 + 11200\}.$$

$s_x^2 = 1659.077$

$$t = \frac{198}{\sqrt{1659.077\left(\frac{1}{8} + \frac{1}{7}\right)}}$$

$t = 9.39$

$t_c = (5\%, 13 \text{ df}) = 2.16$

$t > t_c$ H_0 is rejected.

3. A grp of 10 children were fed on diet A & another grp of 8 children were fed on diet B for a period of 6 months. The foll" increase in weights was recorded.

A:	5	6	8	1	12	4	3	9	6	10
B:	2	3	6	8	10	1	2	8		

Test whether diets A & B differ significantly regarding their effect on increase in weight.

$\bar{x} = 6.4$ $\bar{y} = 5$ $n_1 = 10$ $n_2 = 8$

$\Sigma(x_i - \bar{x})^2 = 102.4$

$\Sigma(y_i - \bar{y})^2 = 82$

$$s_x^2 = \frac{1}{10+8-2}\{102.4 + 82\}.$$

$$s_x^2 = \frac{1}{16} \times 184.4 = 11.525$$

$$t = \frac{1.4}{\sqrt{11.525\left(\frac{1}{10} + \frac{1}{8}\right)}} = \frac{1.4}{1.6103} = 0.8694$$

$t_c(5\%, 16 \text{ df}) = 2.12$ $\mu_A = \mu_B$

$t < t_c$ $\therefore$ H_0 is accepted.

4. A new process for producing synthetic diamond is viable only if the avg weight of the diamond > 0.5 carat. The weight of 6 diamond thus generated are 0.46, 0.51, 0.52, 0.48, 0.57, & 0.54 carat. Test the viability of the process.

→ $\mu > 0.5$, $\bar{x} = 0.5133$ (avg of the data), $n = 6$

$\Sigma (x_i - \bar{x})^2 = 7.933 \times 10^{-3}$

$s^2 = \frac{1}{5} \times 7.933 \times 10^{-3} \Rightarrow s = \sqrt{1.5866 \times 10^{-3}} = 0.03983$

$t = \frac{0.5133 - 0.5}{\frac{0.0398}{\sqrt{6}}} = 0.8179$

$H_0 : \bar{x} = \mu$

t (5% los, 5 df) = 2.015 (one sided data)

$t < t_c$; H_0 is accepted for 5% & 1% los.

→ F-test.

It is used to calculate the ratio of variances of normally distributed statistics. Let $X = \{x_1, x_2, x_3 \ldots x_n\}$ & $Y = \{y_1, y_2, y_3\}$ be the values of 2 samples drawn from the same normal population. Then $F = \frac{s_1^2}{s_2^2}$ where

$s_1^2 = \frac{1}{n-1} \sum_{i=1}^{n} (x_i - \bar{x})^2$ & $s_2^2 = \frac{1}{n-1} \sum_{i=1}^{m} (y_i - \bar{y})^2$

∴ Larger value to be taken in the numerator (always).

→ F-test for comparing variance :-

1) Two samples of size 9 and 8 gives sum of squares of deviatn from their resp. means as 160 inches^2 & 91 inches^2. Can these be regarded as drawn from the normal popn with the same var?

→ $n = 9$, $m = 8$, $\sum (x_i - \bar{x})^2 = 160$, $\sum (y_i - \bar{y})^2 = 91$

$s_1^2 = \frac{1}{8}(160) = 20$, $s_2^2 = \frac{1}{7}(91) = \frac{91}{7} = 13$

$F = \frac{20}{13} = 1.5384$, F_c (5%, 8, 7 df) = 3.73

H_0 is accepted as $F < F_c$. Thus the two popn variances are equal.

2) Two independent samples have the foll^n values

A: 28 30 32 33 33 29 34 (7)

B: 29 28 27 24 23 28 (6)

Examine if they have been drawn from the normal popn with the same variance.

→ $\bar{x} = 31.2857$, $\bar{y} = 26.5$, $\Sigma (x_i - \bar{x})^2 = 31.4285$

$\Sigma (y_i - \bar{y})^2 = 29.5$.

$s_1^2 = \frac{1}{n-1} \Sigma (x_i - \bar{x})^2 = \frac{1}{6}(31.4285) = 5.238$

$s_2^2 = \frac{1}{m-1} \Sigma (y_i - \bar{y})^2 = \frac{1}{5}(29.5) = 5.9$

$F = \frac{5.9}{5.238} = 1.1263$ F_c (5%, 5, 6 df) = 4.39.

$F < F_c \Rightarrow H_0$ accepted.

3) If 2 samples of size 10 & 14 have std dev^n 1.5 & 1.2, can we say that they are drawn from normal popl^n with the same variance?

→ $n_1 = 10$, $n_2 = 14$, $S_1 = 1.5$, $S_2 = 1.2$

$$s_1^2 = \frac{n_1}{n_1 - 1} S_1^2 = \frac{10}{9} \times (1.5)^2 = 2.5$$

$$s_2^2 = \frac{n_2}{n_2 - 1} S_2^2 = \frac{14}{13} (1.2)^2 = 1.5507$$

$$F = \frac{2.5}{1.5507} = 1.6121 \quad / \quad F_c(5\%, 9, 13) = 2.7$$

$F < F_c \Rightarrow H_0$ accepted!

4) A teacher in 2 classes A and B, the same subject, class A has 16 students & 25 students. In an exam, although there was no big diff in mean gradis, class A has a std dev of 9 & B has 12. Can we conclude that variability of B > A.

→ Chi Square test

It is used for

i) Independence of attributes

ii) Goodness of fit.

(*) Most of the tests were based on the assumption that the sample was drawn from a normal popl^n. But there are situations when the data doesn't follow normal dist and the observat^n are recorded only as presence of absence of an attribute. In such a situation, we use non-paramet-ric tests of significance.

$$\chi_i^2 = \frac{(O_i - E_i)^2}{E_i}$$

O_i → Observed frequency

E_i → Expected "

$$\chi^2 = \sum_i \chi_i^2$$

when null hypothesis is true, χ^2 follows chi-square dist with $(r-1)(c-1)$ d.f.

where, r = no. of rows

c = no. of columns.

→ Questions:

1) In a health survey, 400 individuals were asked if they were vaccinated against flu and if they subsequently had flu. The foll^n data was

obtained. Test if vaccination prevents flu.

	Vaccinated	Not va	
Got flu	60	85	145
Not got flu	190	65	255
	250	150	400

H_0 : vaccination & catching flu are independent

Attribute	O_i	E_i	$\chi_i^2 = (O_i - E_i)^2 / E_i$
(vacc & got flu)	60	$\frac{145 \times 250}{400} = 90.625$	10.349
(Not vacc & got flu)	85	$\frac{150 \times 145}{400} = 54.375$	17.248
(vacc & not got flu)	190	$\frac{250 \times 255}{400} = 159.375$	5.8848
(Not vacc & not got flu)	65	$\frac{150 \times 255}{400} = 95.625$	9.808
			43.29

$\chi^2 = \sum_i \chi_i^2 = 43.29$

$df = (2-1)(2-1) = 1$

$\chi^2_{0.95}(1\ d.f) = 3.84$

$\therefore H_0$ is rejected

Q) Given the data below, perform χ^2 test to check if hair colour and eye colour are dependent.

Eye colour \ Hair colour	Black	Grey	
Black	40	20	60
Blue	20	30	50
Brown	60	30	90
	120	80	200

H_0 : Hair colour & Eye colour are independent

Attribute	O_i	E_i	$\chi_i^2 = (O_i - E_i)^2 / E_i$
Black / Black	40	$\frac{120 \times 60}{200} = 36$	0.4444
~~grey / Black~~	~~20~~		
grey / Black	20	$\frac{60 \times 80}{200} = 24$	0.6667
Black / Blue	20	$\frac{50 \times 120}{200} = 30$	3.3333
grey / Blue	30	$\frac{50 \times 80}{200} = 20$	5.0
Black / Brown	60	$\frac{120 \times 90}{200} = 54$	0.6667
grey / Brown	30	$\frac{80 \times 90}{200} = 36$	1.0
			11.1111

$\chi^2 = \sum_i \chi_i^2 = 11.1111$

$\chi^2_{(0.95)}\{(3-1)(2-1)df\} = \chi^2_{(0.95)}(2\ df) = 5.99$

$\chi^2 > \chi^2_{(0.95)}(2\ df)$

$\therefore H_0$ is rejected.

a) The no. of demands for a particular spare part in a shop is found to vary day by day. In a sample study, the foll^{n} data is obtained.

Day :	Mon	Tue	Wed	Thu	Fri	Sat
No. of parts demanded	124	125	110	120	126	115

d.f = (r-1)(c-1) = (2-1)(c-1) = 5

χ_c^2 (95%, 5 d.f) = 11.1

$\chi^2 = \sum_i \chi_i^2 =$ 1.6832.

Attribute	O_i	E_i	χ_i^2
1	124	120	0.133
2	125	120	0.2083
3	110	120	0.8333
4	120	120	0
5	126	120	0.3
6	115	120	0.2083
			1.6832

$\chi^2 < \chi_c^2$

$\therefore H_0$ is accepted.

Q) A sample analysis of exam results of 500 students was made. It is as follows. 220 failed, 170 had III class, 90 had II class, 20 had I class. Do the follows the ratio 4:3:2:1?

Total = 500, 4:3:2:1

	O_i	E_i	χ_i^2
Fail	220	200	
III	170	150	
II	90	100	
I	20	50	

χ_c^2 (0.95, 3 d.f) = 7.81

Q) A iservey of 800 families with 4 children each gave the follⁿ data

No. of boys	0	1	2	3	4
No. of girls	4	3	2	1	0
families	32	178	290	236	64

Is the result consistent with the hypothesis that a girl or a boy is equally likely to be born?

$\therefore p = 1/2, q = 1/2$

B	G	O_i	E_i
0	4	32	$^4C_0 (1/2)^0 (1/2)^{4-0} = 1/16 \times 800 = 50$
1	3	178	$^4C_1 (1/2)^1 (1/2)^{4-1} = 4/16 \times 800 = 200$
2	2	290	$^4C_2 (1/2)^2 (1/2)^2 = 6/16 \times 800 = 300$
3	1	236	$^4C_3 (1/2)^3 (1/2)^1 = 4/16 \times 800 = 200$
4	0	64	$^4C_4 (1/2)^4 (1/2)^0 = 1/16 \times 800 = 50$

χ_i^2 : 6.48 2.42 0.3333 6.48 3.92

$\sum \chi_i^2 = \chi^2 =$ 19.633

χ_c^2 (0.95, (5-1)(2-1) d.f) = χ_c^2 (0.95, 4 d.f) = 9.49

H_0 is rejected.

Q) A die is thrown 276 and the results of these throws are shown below.

No. appeared	O_i	E_i	χ_i^2
1	44	46	0.0869
2	52	46	0.7826
3	42	46	0.3478
4	50	46	0.34878
5	49	46	0.1956
6	39	46	1.0652
			2.4781

$\chi_c^2 (0.95, 5\ df) = 11.1$

$\Sigma \chi_i^2 < \chi_c^2$

$\therefore H_0$ is accepted.

Q. Test the given data for goodness of fit for Poisson dist.

x	0	1	2	3	4
f	122	60	15	2	1

$\mu = \frac{\Sigma x_i f_i}{\Sigma f_i} = \frac{100}{200} = 0.5$

$P(x) = \frac{e^{-\mu} \mu^x}{x!}$

$P(x=0) = \frac{e^{-0.5} (0.5)^0}{0!} = 0.6065$

$P(x=1) = \frac{e^{-0.5} (0.5)}{1!} = 0.3033$

$P(x=2) = \frac{e^{-0.5} (0.5)^2}{2!} = 0.0758$

$P(x=3) = \frac{e^{-0.5} (0.5)^3}{3!} = 0.0126$

$P(x=4) = \frac{e^{-0.5} (0.5)^4}{4!} = 0.00158$

$P(x=5) = \frac{e^{-0.5} (0.5)^5}{5!} = 0.00158$

x	O_i	E_i	χ_i^2
0	122	121.30	0.0044
1	60	60.66	0.0072
2	15	15.16	0.0017
3	2	2.52	0.1073
4	1	0.316	1.4805
			1.6011

We lose 1 d.f in calculating μ.

$\therefore d.f = (5-1)(2-1) - 1 = 3\ df.$

$\chi_c^2 = (0.95, 3\ df) = 7.81$

$\Sigma \chi_i^2 < \chi_c^2$

$\therefore H_0$ is accepted.

Q) The table gives the no. of good and bad parts produced by each of the 3 shifts in a factory. Test if the prod^n of bad parts is independent of the shifts.

shifts	good	Bad		
1	960	40	1000	40
2	940	50	990	50
3	950	60	1100	
	2850	150	3000	

shifts	Oi	Ei	χ_i^2
1	960	$\frac{1000 \times 2850}{3000} = 950$	0.1053
2	940	$\frac{990 \times 2850}{3000} = 940.5$	0
3	950	$\frac{1100 \times 2850}{3000} = 1045$	

$\chi_c^2(0.95, 2\,df) = 5.99$

$\chi_i^2 < \chi_c^2$

$\therefore H_0$ is accepted.

Q. Fit a Poisson distribut^n for the foll^n data and test the goodness of fit.

x :	0	1	2	3	4	5	6	Total
f :	273	70	30	7	7	2	1	390

$\mu = \frac{\Sigma fx}{\Sigma f} = \frac{195}{390} = 0.5$

$P(x=0) = \frac{e^{-0.5}(0.5)^0}{0!} = 0.6065 \times 390 = 236.535$

$P(x=1) = \frac{e^{-0.5}(0.5)^1}{1!} \times 390 = 118.287$

$P(x=2) = 29.562$

$P(x=3) = 4.914$

$P(x=4) = 0.6162$

$P(x=5) = 0.06162$

$P(x=6) = 0.00513.$

x	Oi	Ei	χ_i^2
0	273	236.535	5.6016
1	70	118.287	19.7116
2	30	29.562	0.0065
3	7	4.914	0.8851
4	7	0.6162	66.1358
5	2	0.06162	60.9756
6	1	0.00513	192.9369

$\chi_c^2(0.95, (6-1)\,df) = 11.1$

$\Sigma \chi_i^2 >> 11.1$

$\therefore H_0$ is Rejected.

Q)

Face appeared on dice	Oi	Ei	χ^2
1	55	$\frac{1}{6} \times 300 = 50$	0.5
2 or 3	108	$\frac{1}{3} \times 300 = 100$	0.64
4 or 5	94	$\frac{1}{3} \times 300 = 100$	0.36
6	43	$\frac{1}{6} \times 300 = 50$	0.98
	300		2.48

$\chi_c^2(0.95, 3\,df) = 7.81$

$\chi_i^2 < \chi_c^2 = \therefore H_0$ is accepted.

Q.

No. of heads	Oi	Ei	χ_i^2
0	15	${}^5C_0(1/2)^5(1/2)^0 = 10$	2.5
1	45	${}^5C_1(1/2)^5 = 50$	0.5
2	85	${}^5C_2(1/2)^5 = 100$	2.25
3	95	${}^5C_3(1/2)^5 = 100$	0.25
4	60	${}^5C_4(1/2)^5 = 50$	2
5	20	${}^5C_5(1/2)^5 = 10$	10
	320		17.5

$\chi_c^2(0.95, 5df) = 11.1$

$\chi_i^2 > \chi_c^2$ H_0 is rejected.

Q.

Attribute	Oi	Ei	χ_i^2
Red	34	$9/16 \times 64 = 36$	0.1111
Black	10	$3/16 \times 64 = 12$	0.3333
white	20	$4/16 \times 64 = 16$	1
			1.4444

$\chi_c^2(0.95, 2df) = 5.99$

H_0 is accepted.