Sampling Theory.

* A large collection of individuals or attributes or numerical data can be understood as a population or universe.

* A finite subset of the universe is called a sample.

* The no. of individuals in a Sample is called a sample size. If the sample size (n) is less than or equal to 30, the Sample is said to be small, otherwise it is a large sample.

* The process of selecting a sample from the population is called a sampling.

* The selection of an individual or item from the population in such a way that each has the same chance of being selected is called as random sampling.

Testing of Hypothesis :-

* In order to arrive at a decision regarding the population through a sample of the population, we have to make certain assumption referred to as hypothesis which may or may not be true.

* The hypothesis formulated for the purpose of its rejection under the assumption that it is true is called the Null Hypothesis, denoted by H_0.

* Any Hypothesis which is complimentary to the null hypothesis is called Alternative Hypothesis denoted by H_1.

Ex :- To test whether a process B is better than a process A, we can formulate the hypothesis as "there is no difference b/w the process A and B".

* Significance level :-

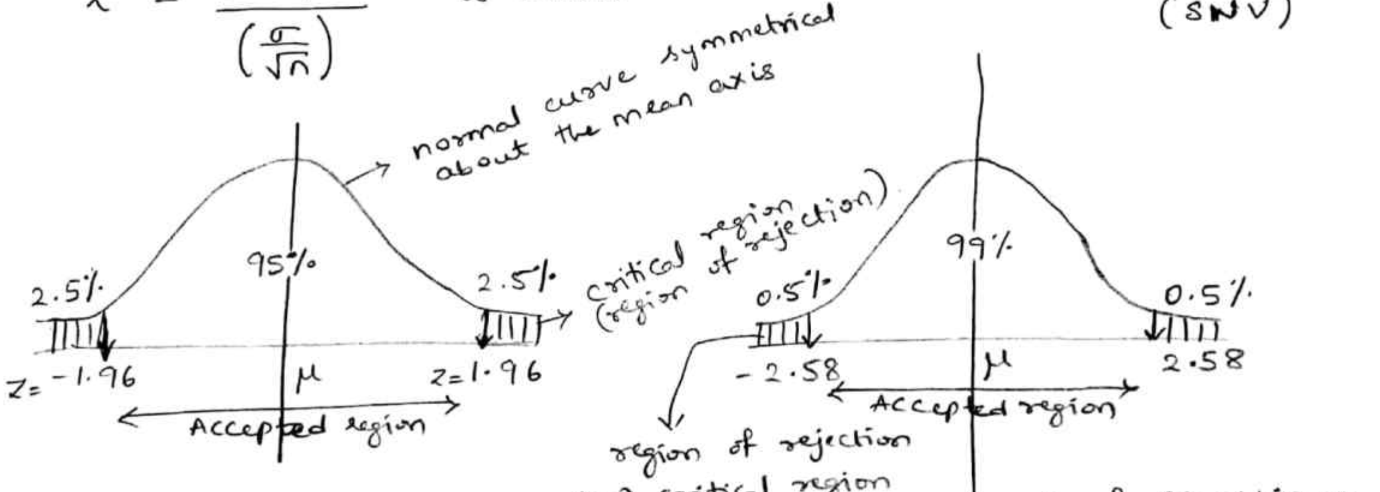
The Prob. level, below which leads to the rejection of the hypothesis is known as the significance level. This probability is conventionally fixed at 0.05 or 0.01 i.e., 5% or 1%. These are called significance levels.

* Tests of significance and Confidence intervals :-

→ The process which helps us to decide about the acceptance or rejection of the hypothesis is called the test of significance.

→ Let us suppose that, we have a normal population with mean μ and S.D σ. If $\bar{x}$ is the sample mean of a random sample of size n, the quantity z defined by

$$z = \frac{\bar{x} - \mu}{\left(\frac{\sigma}{\sqrt{n}}\right)}$$ is called the Standard Normal Variate (SNV)



5% Level of significance
fig (1)

1% level of Significance.
fig (2)

→ In 5% level of significance, 95% of confidence that we can accept the Ho.

→ In 1% level of significance, 99% of confidence that we can accept the Ho.

Ex :- 1) $z = -1.2$ falls in accepted region. So we accept Ho.
2) $z = 2.0$ falls in critical region. So we reject Ho
} in 5% level of significance.

→ These 2 significance levels are two tailed test b'coz we can calculate the z-value at both ends (+ve & −ve end)

→ In 2 tailed test, critical values of z are constant.

for 5% level → $z = -1.96$ and $z = 1.96$

for 1% level → $z = -2.58$ and 2.58.

→ In one tailed test, we calculate z-value at any one of the end either +ve or −ve and these values are constant.

For 5% level → $z = -1.645$ or 1.645

for 1% level → $z = -2.33$ or 2.33

→ In z-test, 5% level of significance is denoted by $z_{0.05}$ and 1% level of significance is denoted by $z_{0.01}$.

→ If calculated value of a test is less than table value, then H_0 is accepted.

→ If calculated value of a test is more than table value, then H_0 is rejected.

Test	Critical values of z	
	5% level	1% level
one-tailed test	−1.645 (or) 1.645	−2.33 (or) 2.33
two-tailed test	−1.96 and 1.96	−2.58 and 2.58

Note :-

n – sample size

1) for a small sample i.e $n \leq 30$, apply t-test.
2) for a large sample i.e $n > 30$, apply z-test.

If we calculate the values at <u>both the ends</u>, then it is 2-tailed test, otherwise 1-tailed test

+ve & −ve end.

P.T.O

* Errors :- In a test process, there can be 4 possible situations of which 2 of the situations leads to 2 types of errors, given below:

	Accepting the hypothesis	Rejecting the hypothesis
Hypothesis true	correct decision	wrong decision (Type I error)
Hypothesis false	wrong decision (Type II error)	correct decision

* Confidence intervals :-

from fig(1), we find that 95% of the area lies b/w $z = -1.96$ and $z = +1.96$.

ie with 95% confidence, we can say that z lies b/w -1.96 and $+1.96$.

$$\Rightarrow \quad -1.96 \le z \le +1.96.$$

$$\Rightarrow \quad -1.96 \le \left(\frac{\bar{x}-\mu}{\sigma/\sqrt{n}}\right) \le +1.96$$

$$\Rightarrow \quad -\frac{1.96\,\sigma}{\sqrt{n}} \le (\bar{x}-\mu) \le +\frac{1.96\,\sigma}{\sqrt{n}}$$

$$\Rightarrow \quad \mu \le \bar{x} + \frac{1.96\,\sigma}{\sqrt{n}} \quad \text{and} \quad \bar{x} - \frac{1.96\,\sigma}{\sqrt{n}} \le \mu.$$

combining, $$\bar{x} - \frac{1.96\,\sigma}{\sqrt{n}} \le \mu \le \bar{x} + \frac{1.96\,\sigma}{\sqrt{n}}$$

This expression is 95% confidence interval.

|||ly $$\bar{x} - \frac{2.58\,\sigma}{\sqrt{n}} \le \mu \le \bar{x} + \frac{2.58\,\sigma}{\sqrt{n}}.$$

This expression is 99% confidence interval.

Test of Significance of Proportions.

Standard Normal Variate, $z = \frac{x - np}{\sqrt{npq}}$ where

$n \to$ sample size., $x \to$ observered no. of successes.

$p \to$ prob. of success, $q \to$ prob. of failure.

Note :-

1) $p \pm 2.58\sqrt{\frac{pq}{n}}$ are the probable limits at 1% level of significance.

2) $p \pm 1.96\sqrt{\frac{pq}{n}}$ are the probable limits at 5% level of significance. where $\sqrt{\frac{pq}{n}}$ is the S.D or standard error proportion of successes.

Problems :-

1) A coin is tossed 1000 times and head turns up 540 times. Decide on the hypothesis that the coin is unbiased.

Soln :- Let us suppose that the coin is unbiased.

Prob. of getting a head in one toss, $p = 1/2$. $\Rightarrow q = 1/2$.

Sample size, $n = 1000$.

Observed no. of successes, $x = 540$.

$$\therefore \quad z = \frac{x - np}{\sqrt{npq}} = \frac{540 - (1000)(0.5)}{\sqrt{1000 \times (0.5)^2}} = 2.53$$

$$\therefore \quad z = 2.53 \begin{cases} > z_{0.05} = 1.96 \\ < z_{0.01} = 2.58 \end{cases}$$

Thus the hyp. is accepted at 1% & rejected at 5% level of significance.

i.e the coin is unbiased at 1% level of significance.

P.T.O.

2) In 324 throws of a six faced 'die', an odd no. turned up 181 times. Is it reasonable to think that the 'die' is an unbiased one?

soln :- Let us assume that the die is unbiased.

prob. of turn up of an odd no. in a single throw is $p = \frac{3}{6} = \frac{1}{2}$.

$\Rightarrow q = 1/2$

sample size, $n = 324$.

observed no. of successes $= 181 = x$.

$\therefore z = \frac{x - np}{\sqrt{npq}} = \frac{181 - (324)(0.5)}{\sqrt{(324)(0.5)^2}}$

$z = 2.11 < 2.58$

Thus the die is unbiased at 1% level of significance.

3) A die is thrown 9000 times and a throw of 3 or 4 was observed 3240 times. Show that the die cannot be regarded as an unbiased one.

soln :. prob. of getting 3 or 4 in a single throw is

$p = \frac{2}{6} = \frac{1}{3} \Rightarrow q = 1 - \frac{1}{3} = \frac{2}{3}$.

Sample size, $n = 9000$.

Observed no. of successes, $x = 3240$.

$\therefore z = \frac{x - np}{\sqrt{npq}} = \frac{3240 - (9000)(\frac{1}{3})}{\sqrt{9000 \times \frac{1}{3} \times \frac{2}{3}}}$

$z = 5.37 > 2.58$.

Thus the die is biased.

4) A survey was conducted in a slum locality of 2000 families by selecting a sample of size 800. It was revealed that 180 families were illiterates. Find the probable limits of the illiterate families in the population of 2000.

Sol^n: prob. of illeterate families, $p = \frac{180}{800} = 0.225$

$\Rightarrow q = 0.775$

Probable limits of illeterate families

$= p \pm 2.58\sqrt{\frac{pq}{n}}$

$= 0.225 \pm 2.58\sqrt{\frac{(0.225)(0.775)}{800}}$

$= 0.187$ and 0.263

for a population of 2000, the probable limits of illiterate families are 2000 × 0.187 and 0.263 × 2000

i.e 374 to 526 are probable itteterate families.

5) A sample of 900 days was taken in a coastal town and it was found that on 100 days, the weather was very very hot. obtain the probable limits of the % of very hot weather.

Sol^n :- prob. of very hot weather, $p = \frac{100}{900} = \frac{1}{9}$

$\Rightarrow q = 8/9$.

probable limits $= p \pm 2.58\sqrt{\frac{pq}{n}}$

$= \frac{1}{9} \pm (2.58)\sqrt{\frac{\frac{1}{9}\times\frac{8}{9}}{900}}$

$= 0.084$ and 0.138

Prob. limits of very hot weather is 8.4% to 13.8%

6) In a Sample of 500 men, it was found that 60% of them had over weight. what can we infer about the propor-tion of people having over weight in the population?

Sol^n :- Prob. of persons having over weight, $p = 60\% = 0.6$.

$q = 0.4$

$$\text{Probable limits} = p \pm 2.58 \sqrt{\frac{pq}{n}}$$

$$= 0.6 \pm 2.58 \sqrt{\frac{(0.6)(0.4)}{500}}$$

$$= 0.5435 \text{ and } 0.6565$$

Thus the probable limits of people having over weight is 54.35% to 65.65%.

7) A manufacturer claimed that atleast 95% of the equipment
QP which he supplied to a factory conformed to specifications. An examination of a sample of 200 pieces of equipment revealed that 18 of them were faulty. Test his claim at a significance level of 1% and 5%.

Soln :- 95% of equipment conformed to specifications

$\Rightarrow$ 5% are not conformed to specifications (may be faulty)

$\Rightarrow$ $p = 5\% = 0.05$ and $q = 0.95$.

Let x denote the observed no. of successes ie no. of faulty items.

$x = 18$. $n = 200$.

$$\therefore \quad z = \frac{x - np}{\sqrt{npq}} = \frac{18 - 200 \times 0.05}{\sqrt{(200)(0.05)(0.95)}}$$

$$z = 2.5955 > z_{0.05} = 1.96$$

$$z_{0.01} = 2.58$$

Thus his claim is not supported.

Scanned by CamScanner

Test of Significance for Single Mean :-

S.N.V , $z = \dfrac{\bar{x} - \mu}{\sigma/\sqrt{n}}$ where σ is the S.D of the population.

(If σ is not known, then we use $z = \dfrac{\bar{x} - \mu}{s/\sqrt{n}}$; where s is the S.D of the sample).

Note :- 1) This z is to test whether the difference b/w the sample mean $\bar{x}$ & population Mean μ is significant or not.

2) 95% confidence limits are $\bar{x} - 1.96\dfrac{\sigma}{\sqrt{n}} < \mu < \bar{x} + 1.96\dfrac{\sigma}{\sqrt{n}}$

3) 99% Confidence limits are $\bar{x} - 2.58\dfrac{\sigma}{\sqrt{n}} < \mu < \bar{x} + 2.58\dfrac{\sigma}{\sqrt{n}}$

Problems:-

1) A random sample of 400 items chosen from an infinite population is found to have a mean of 82 and a S.D of 18. Find the 95% confidence limits for the mean of the population from which the sample is drawn.

QP

Sol^n:- By data, $n = 400$, $\bar{x} = 82$, $\sigma = 18$.

95% Confidence limits for μ are

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

$$82 - \frac{1.96 \times 18}{20} < \mu < 82 + \frac{1.96 \times 18}{20}$$

$$80.236 < \mu < 83.764$$

95% confidence limits are 80.236 to 83.764

2) The life of certain computer is approximately normally distributed with mean 800 hrs & S.D of 40 hrs. If a random sample of 30 computers has an average life of 788 hours, test the hyp - that $\mu = 800$ hrs against alternate hyp. $\mu \neq 800$ hrs at
(i) 5 % and (ii) 1 % level of significance.

Soln :. $\mu = 800$, $\sigma = 40$, $n = 30$, $\bar{x} = 788$

$$\therefore \quad z = \frac{\bar{x} - \mu}{\sigma}\sqrt{n} = \frac{788 - 800}{40}\sqrt{30}$$

$$z = -1.6432$$

$$|z| = 1.64 \begin{cases} < 1.96 = z_{0.05} \\ < 2.58 = z_{0.01} \end{cases}$$

$\therefore \mu = 800$ is accepted at both levels of significance.

3) A sample of 100 tyres is taken from a lot. The mean life of tyres is found to be 39350 km with a S.D of 3260. Can it be considered as a true random sample from a population with mean life of 40000 kms ? Use 0.05 log. Establish 99% confidence limits within which the mean life of tyres expected to lie.

Soln :. $n = 100$, $\bar{x} = 39350$, $s = 3260$, $\mu = 40000$.

$$\therefore \quad z = \frac{\bar{x} - \mu}{s}\sqrt{n} = \frac{39350 - 40000}{3260}\sqrt{100}$$

$$|z| = 1.9939 > z_{0.05} = 1.96$$

Thus it cannot be considered as a true random sample from a population of $\mu = 40000$ kms.

Scanned by CamScanner

99% Confidence limits are given by

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

$$39350 - \frac{2.58 \times 3260}{\sqrt{100}} < \mu < 39350 - \frac{2.58 \times 3260}{\sqrt{100}}$$

$$38509 < \mu < 40191$$

Hean, μ is expected to lie b/w 38509 to 40191.

4) A sugar factory is expected to sell sugar in 100 kg bags.
A sample of 144 bags taken from a day's output shows the average & S.D of weights of these bags as 99 & 4 kg resp. Can we conclude that the factory is working as per standards ($z = 1.96$ at 5% log)

Soln :- By data, $\mu = 100$, $n = 144$, $\bar{x} = 99$, $s = 4$.

$$z = \frac{\bar{x} - \mu}{s}\sqrt{n} = \frac{99 - 100}{4}\sqrt{144}$$

$$|z| = 3 > 1.96 = z_{0.05}$$

Thus the factory is not working as per standards.

5) The mean and S.D of marks scored by a sample of 100 students are 67.45 and 2.92. find (i) 95% and (ii) 99% confidence intervals for estimating the mean marks of the student population.

Soln :- By data, $n = 100$, $\bar{x} = 67.45$, $s = 2.92$

i) 95% confidence intervals are

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

$$67.45 - \frac{1.96(2.92)}{\sqrt{100}} < \mu < 67.45 + \frac{1.96(2.92)}{\sqrt{100}}$$

$$\Rightarrow \quad 66.88 < \mu < 68.02$$

μ lies b/w 66.88 to 68.02.

(ii) 99% confidence intervals are

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

$$\Rightarrow 67.45 - \frac{2.58\,(2.92)}{10} < \mu < 67.45 + \frac{2.58\,(2.92)}{10}$$

$$66.70 < \mu < 68.20.$$

μ lies b/w 66.70 to 68.20.

Test of significance of difference b/w means :

Standard Normal variate, $z = \dfrac{(\bar{x}_1 - \bar{x}_2)}{\sqrt{\dfrac{\sigma_1^2}{n_1} + \dfrac{\sigma_2^2}{n_2}}}$ where

$(\bar{x}_1, \sigma_1)$ and $(\bar{x}_2, \sigma_2)$ are the mean and S.D of 2 large samples of size n_1 and n_2 resp.

Note :- 1) If the samples are drawn from the same population, then $\sigma_1 = \sigma_2 = \sigma$, so that $z = \dfrac{(\bar{x}_1 - \bar{x}_2)}{\sigma\sqrt{\dfrac{1}{n_1} + \dfrac{1}{n_2}}}$

2) Confidence limits for the difference of means of the population are $(\bar{x}_1 - \bar{x}_2) \pm z_c \sqrt{\dfrac{\sigma_1^2}{n_1} + \dfrac{\sigma_2^2}{n_2}}$. where z_c are the critical values.

Problems :

1) Intelligent tests were given to 2 groups of boys & girls.

	Mean	S.D	Size
Boys	75	8	60
Girls	73	10	100

Find out if the 2 means significantly differ at 5% level of significance.

Soln :- Let H_0 : There is no significant difference b/w the mean scores. $(\cancel{\bar{x}_1 = \bar{x}_2})$.

$$z = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{\dfrac{\sigma_1^2}{n_1} + \dfrac{\sigma_2^2}{n_2}}} = \frac{75-73}{\sqrt{\dfrac{64}{60} + \dfrac{100}{100}}}$$

$$z = 1.39 < 1.96$$

$\therefore$ H_0 is accepted.

Thus there is no significant difference b/w mean scores.

2) A sample of 100 bulbs produced by a company A showed a mean life of 1190 hours and a S.D of 90 hrs. Also a sample of 75 bulbs produced by a company B showed a mean life of 1230 and a S.D of 120 hrs. Is there a difference b/w the mean life time of the bulbs produced by the 2 companies at (i) 5% level of significance (ii) 1% level of significance.

Soln :- By data, $n_1 = 100$, $\bar{x}_1 = 1190$, $\sigma_1 = 90$ (company A)
$n_2 = 75$, $\bar{x}_2 = 1230$, $\sigma_2 = 120$ (company B)

Let Ho : There is no difference b/w the mean lifetime of bulbs.

$$\therefore \quad z = \frac{(\bar{x}_1 - \bar{x}_2)}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}} = \frac{(1190 - 1230)}{\sqrt{\frac{90^2}{100} + \frac{120^2}{75}}}$$

$$|z| = 2.42 \begin{cases} > z_{0.05} = 1.96 \\ < z_{0.01} = 2.58 \end{cases}$$

Thus Ho is accepted at 1% level of significance and rejected at 5% level of significance.

3) A random sample for 1000 workers in a company has mean wage of Rs. 50 per day and S.D of Rs. 15. Another sample of 1500 workers from another company has mean wage of Rs. 45 per day and S.D of Rs. 20. Does the mean rate of wages varies b/w the 2 companies? Find the 95% and 99% confidence limits for the difference of the mean wages of the population of the 2 companies.

Soln :- Company - 1 : $\bar{x}_1 = 50$, $\sigma_1 = 15$, $n_1 = 1000$
company - 2 : $\bar{x}_2 = 45$, $\sigma_2 = 20$, $n_2 = 1500$.

Let Ho : There is no significant difference b/w the mean wages of the 2 companies.

$$\therefore \quad z = \frac{(\bar{x}_1 - \bar{x}_2)}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}} = \frac{5}{\sqrt{\frac{15^2}{1000} + \frac{20^2}{1500}}}$$

$$z = 7.13 \begin{cases} > z_{0.05} = 1.96 \\ > z_{0.01} = 2.58 \end{cases}$$

Thus H_0 is rejected at both levels of significance. ie there is a significant difference b/w the mean wages.

4) The mean of 2 large samples of 1000 and 2000 members are 168.75 cms and 170 cms resp. Can the samples be regarded as drawn from the same population of S.D 6.25 cms?

soln :- By data, $\bar{x}_1 = 168.75$, $\bar{x}_2 = 170$

$n_1 = 1000$, $n_2 = 2000$.

Let us assume that the samples are drawn from the same population of S.D 6.25 cms.

ie $\sigma = 6.25$

$$\therefore \quad z = \frac{\bar{x}_2 - \bar{x}_1}{\sigma\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}} = \frac{1.25}{6.25\sqrt{\frac{1}{1000} + \frac{1}{2000}}}$$

$$z = 5.16 \begin{cases} > 1.96 \\ > 2.58 \end{cases}$$

Thus the hypothesis is rejected. ie the samples cannot be regarded as drawn from the same population.

Test of significance for difference of proportions (attributes) for 2 samples :- (9)

S.N.V, $z = \dfrac{p_1 - p_2}{\sqrt{pq\left(\frac{1}{n_1}+\frac{1}{n_2}\right)}}$; $p = \dfrac{n_1 p_1 + n_2 p_2}{n_1 + n_2}$

where p_1, p_2 are the sample proportions in respect of an attribute corresponding to 2 large samples of size n_1, n_2 drawn from 2 populations.

Problems :-

1) QP In a city A, 20% of a random sample of 900 school boys had a certain slight physical defect. In another city B, 18.5% of a random sample of 1600 school boys had the same defect. Is the difference b/w the proportions significant ?

Sol :- By data, $n_1 = 900$, $n_2 = 1600$

$p_1 = 0.20$, $p_2 = 0.185$

Let Ho : There is no significant difference b/w the two proportions (~~ie $p_1 = p_2 = p$ (say)~~)

$\therefore \; z = \dfrac{p_1 - p_2}{\sqrt{pq\left(\frac{1}{n_1}+\frac{1}{n_2}\right)}}$ where $p = \dfrac{n_1 p_1 + n_2 p_2}{n_1 + n_2}$

$\boxed{p = 0.19} \Rightarrow \boxed{q = 0.81.}$

$\Rightarrow z = \dfrac{0.2 - 0.185}{\sqrt{(0.19)(0.81)\left(\frac{1}{900}+\frac{1}{1600}\right)}}$

$z = 0.9202 \begin{cases} < z_{0.01} = 2.58 \\ < z_{0.05} = 1.96 \end{cases}$

$\therefore$ Ho is accepted at both levels of significance.

ie there is no significant difference b/w the proportions.

2) One type of aircraft is found to develop engine trouble in 5 flights out of a total of 100 and another type in 7 flights out of a total of 200 flights. Is there a significant difference in the 2 types of aircrafts so far as engine defects are concerned?

Soln :- Let p_1 & p_2 be the proportion of defects in the 2 types of aircrafts.

$$\therefore \quad p_1 = \frac{5}{100} = 0.05, \qquad p_2 = \frac{7}{200} = 0.035$$

$$n_1 = 100, \quad n_2 = 200.$$

Let H_0 : there is no significant difference b/w the 2 types of aircrafts.

$$z = \frac{p_1 - p_2}{\sqrt{pq\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}} \qquad \text{where } p = \frac{n_1 p_1 + n_2 p_2}{n_1 + n_2}$$

$$\Rightarrow \boxed{p = 0.04} \Rightarrow \boxed{q = 0.96}$$

$$\therefore \quad z = \frac{0.05 - 0.035}{\sqrt{(0.04)(0.96)\left(\frac{1}{100} + \frac{1}{200}\right)}}$$

$$z = 0.625 \begin{cases} < z_{0.05} = 1.96 \\ < z_{0.01} = 2.58 \end{cases}$$

H_0 is accepted at both levels of significance.

3) Random sample of 1000 engineering students from a city A and 800 from city B were taken. It was found that 400 students in each of the sample were from payment quota. Does the data reveal a significant difference b/w the 2 cities in respect of payment quota students?

P.T.O.

Soln :- By data, $n_1 = 1000$, $n_2 = 800$.

$$p_1 = \frac{400}{1000} = 0.4, \qquad p_2 = \frac{400}{800} = 0.5$$

$$\therefore \; p = \frac{n_1 p_1 + n_2 p_2}{n_1 + n_2} = \boxed{\frac{4}{9} = p} \Rightarrow \boxed{q = 5/9.}$$

Let Ho : there is no significant difference b/w the 2 cities.

$$z = \frac{p_2 - p_1}{\sqrt{pq\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}} = \frac{0.1}{\sqrt{\left(\frac{4}{9}\right)\left(\frac{5}{9}\right)\cdot\left(\frac{1}{1000} + \frac{1}{800}\right)}}$$

$$z = 4.243 \begin{cases} > z_{0.05} = 1.96 \\ > z_{0.01} = 2.58 \end{cases}$$

Thus Ho is rejected at both levels of significance.

4) A company has the head office at Kolkata and a branch at Mumbai. The personnel director wanted to know if the workers at the 2 places would like the introduction of a new plan of work and a survey was conducted for this purpose. Out of sample of 500 workers at Kolkata, 62% favoured the new plan. At Mumbai, out of sample of 400 workers, 41% were against the new plan. Is there any significant difference b/w the 2 groups in their attitude towards the new plan at 5% level.

Soln :- Let p_1 & p_2 be the sample proportion of workers favouring the new plan at Kolkata & Mumbai resp.

By data, $n_1 = 500$, $p_1 = 0.62$

$n_2 = 400$, $p_2 = 0.59$ (since given $q_1 = 0.41$)

Let Ho : there is no significant difference b/w the 2 groups in their attitude towards the new plan.

Scanned by CamScanner

$$p = \frac{n_1 p_1 + n_2 p_2}{n_1 + n_2} \Rightarrow \boxed{p = 0.607}$$

$$\Rightarrow \boxed{q = 0.393}$$

$$\therefore \quad z = \frac{p_1 - p_2}{\sqrt{pq\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}} = \frac{0.62 - 0.59}{\sqrt{(0.607)(0.393)\left[\frac{1}{500} + \frac{1}{400}\right)}}$$

$$z = 0.916 \begin{cases} < 1.96 \\ < 2.58 \end{cases}$$

Thus H_0 is accepted at 5% level of significance.

5) It is required to test whether the proportion of smokers among students is less than that among the lectures. Among 60 randomly picked students, 2 were smokers. Among 17 randomly picked lecturers, 5 were smokers. What would be your conclusion?

QP

Soln:- By data, $n_1 = 60$, $p_1 = 2/60 = 1/30$.

$n_2 = 17$, $p_2 = 5/17$

$$\therefore \quad p = \frac{n_1 p_1 + n_2 p_2}{n_1 + n_2} = 0.0909 \text{ (or) } \boxed{p = 1/11}$$

$$\Rightarrow \boxed{q = 10/11}$$

$$\therefore \quad z = \frac{p_1 - p_2}{\sqrt{pq\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}} = \frac{\frac{1}{30} - \frac{5}{17}}{\sqrt{\left(\frac{1}{11}\right)\left(\frac{10}{11}\right)\left|\frac{1}{60} + \frac{1}{17}\right.}}$$

$$|z| = 3.30 > \begin{cases} z_{0.05} = 1.96 \\ z_{0.01} = 2.58 \end{cases}$$

The required test is rejected.

Student's 't' distribution

I Student's 't' test for a sample mean

The statistic 't' is defined as follows:

$$t = \frac{\bar{x} - \mu}{(S/\sqrt{n})} = \frac{\bar{x} - \mu}{s}\sqrt{n}.$$

where μ is the mean of the universe (population)

n is the sample size.

$\bar{x} = \frac{1}{n}\sum x$ is the sample mean.

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

the sample variance.

Note:- 1) Degrees of freedom (d.f) is the no. of values generated by a sample of small size for estimating a population parameter.

2) for one sample, d.f $= \nu = n-1$.

3) for a small sample, say $n \le 30$, we apply t-test.

4) To test the hypothesis, whether the sample mean $\bar{x}$ differs significantly from the population mean μ, we compute student's 't'.

5) If $|t| > t_{0.05}$ (where $t_{0.05}$ is the table value of student's 't'), then the difference b/w $\bar{x}$ and μ is significant at 5% level of significance and the hypothesis is rejected.

imp (or) considerable

agreeable

6) If $|t| < t_{0.05}$, then the data is said to be consistent with the hypothesis that μ is the mean of the population and the hypothesis is accepted.

7) 95% confidence limits for μ are $\bar{x} \pm \frac{s}{\sqrt{n}} t_{0.05}$

8) 99% " ——— " ——— $\bar{x} \pm \frac{s}{\sqrt{n}} t_{0.01}$

II Test of significance of difference b/w sample means

Consider 2 independent samples x_i $(i=1,2,\dots n_1)$ and y_j $(j=1,2,\dots n_2)$ drawn from a normal population.

Student 't' is given by

$$t = \frac{\bar{x}-\bar{y}}{s\sqrt{\frac{1}{n_1}+\frac{1}{n_2}}} \quad \text{where}$$

$\bar{x}, \bar{y}$ are the mean of the 2 samples.

$$s^2 = \frac{1}{n_1+n_2-2}\left[\sum(x-\bar{x})^2 + \sum(y-\bar{y})^2\right].$$

(or) $s^2 = \dfrac{n_1 s_1^2 + n_2 s_2^2}{n_1+n_2-2}$ (where s_1, s_2 are S.D of 2 samples)

Note :- 1) For 2 samples, d.f, $\nu = n_1+n_2-2$.

2) We assume the null hypothesis $H_0 : \mu_x = \mu_y$ ie the samples have been drawn from the normal populations with the same means. ie samples means $\bar{x}$ and $\bar{y}$ do not differ significantly.

Problems :-

1) Find the student's 't' for the foll variable values in a sample of eight : $-4, -2, -2, 0, 2, 2, 3, 3$, taking the mean of the universe to be zero.

Soln :- By data, $\mu = 0$, $n = 8$.

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

$$\bar{x} = \frac{1}{n}\Sigma x = 0.25.$$

$$s^2 = \frac{1}{n-1}\left[\Sigma x^2 - \frac{1}{n}(\Sigma x)^2\right] = \frac{1}{7}\left[50 - \frac{1}{8}(2)^2\right]$$

$$s^2 = 7.0714 \Rightarrow s = 2.6592$$

$$① \Rightarrow t = \frac{0.25 - 0}{2.6592}\sqrt{8} \Rightarrow \boxed{t = 0.2659}$$

2) A sample of 10 measurements of the diameter of a ~~Sphere~~ sphere gave a mean of 12 cm and a s.D of 0.15 cm. find the 95% confidence limits for the actual diameter.

Soln :- By data, $n = 10$, $\bar{x} = 12$, $s = 0.15$

For 9 d.f, $t_{0.05} = 2.262$

Confidence limits for the actual diameter is given by

$$\bar{x} \pm \frac{s}{\sqrt{n}} t_{0.05} = 12 \pm \frac{0.15}{\sqrt{10}}(2.262)$$

$$= 12 \pm 0.1073$$

$$= 11.893, \; 12.107$$

Thus 11.893 cm to 12.107 cm is the confidence limits for the actual diameter.

3) A machine is expected to produce nails of length 3 inches. A random sample of 25 nails gave an average length of 3.1 inch with S.D 0.3. Can it be said that the machine is producing nails as per specification? ($t_{0.05}$ for 24 d.f is 2.064)

soln :- By data, $\mu = 3$, $n = 25$, $\bar{x} = 3.1$, $S = 0.3$

$$t = \frac{\bar{x} - \mu}{S}\sqrt{n} = \frac{3.1 - 3}{0.3}\sqrt{25}$$

$$\boxed{t = 1.67}$$

for single sample, d.f is $\nu = n - 1 = 24$.

Let Ho : There is no difference b/w the 2 means $\bar{x}$ & μ.
i.e the machine is producing nails as per specification

Thus $t = 1.67 < 2.064$

we can accept the Ho.

4) Ten individuals are chosen at random from a population and their heights in inches are found to be 63, 63, 66, 67, 68, 69, 70, 70, 71, 71. Test the hypothesis that the mean height of the universe is 66 inches. ($t_{0.05} = 2.262$ for 9 d.f)

Q

soln :- By data, $\mu = 66$, $n = 10$.

$$t = \frac{\bar{x} - \mu}{S}\sqrt{n} \rightarrow (1)$$

$$\bar{x} = \frac{1}{n}\sum x_i = \frac{1}{10}\left[63 + 63 + 66 + 67 + 68 + 69 + 70 + 70 + 71 + 71\right]$$

$$\boxed{\bar{x} = 67.8}$$

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

$\therefore S^2 = 9.067 \Rightarrow \boxed{S = 3.011}$

3

① [redacted] $t = \frac{67.8 - 66}{3.011}\sqrt{10}$ $\Rightarrow$ $\boxed{t = 1.89 < 2.262}$

Thus the hypothesis is accepted at 5% level of significance.

→ a thing that evokes a specific reaction in an organ/tissue.

5) Qs A certain stimulus administered to each of the 12 patients resulted in the foll change in blood pressure — 5, 2, 8, −1, 3, 0, 6, −2, 1, 5, 0, 4. Can it be concluded that the stimulus will increase the blood pressure? ($t_{0.05}$ for 11 d.f = 2.201)

soln :- Let Ho = The stimulus administration is not accompanied with increase in the blood pressure, then we can take $\mu = 0$.

By data, $n = 12$. $\mu = 0$ (assumption)

$$t = \frac{\bar{x} - \mu}{s}\sqrt{n} \longrightarrow ①$$

where $\bar{x} = \frac{1}{n}\sum x_i = \frac{1}{12}[5+2+8-1+3+6-2+1+5+4]$

$$\boxed{\bar{x} = 2.5833}$$

$s^2 = \frac{1}{n-1}\left[\sum x^2 - \frac{1}{n}(\sum x)^2\right] \Rightarrow s^2 = 9.538$

$$\boxed{S = 3.088}$$

$\therefore$ $t = 2.8979 \simeq 2.9 > 2.201$

$\therefore$ we reject Ho at 5% level of significance.

ie the stimulus will increase the blood pressure.

6) Qs The nine items of a sample have the foll values 45, 47, 50, 52, 48, 47, 49, 53, 51. Does the mean of these differ significantly from the assumed mean of 47.5.

soln :- Let Ho : $\mu = 47.5$.

ie there is no significant difference b/w the sample mean & population mean.

Scanned by CamScanner

By data, $n = 9$.

$$t = \frac{\bar{x} - \mu}{s}\sqrt{n} = \frac{\bar{x} - 47.5}{s}\sqrt{9} \rightarrow (1)$$

WKT $\bar{x} = \frac{\Sigma x}{n} = 49.11$

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

$$= \frac{1}{8}\left[21762 - \frac{1}{9}(442)^2\right].$$

$$= 6.8611$$

$$\boxed{s = 2.6194}$$

$$\therefore \quad t = \frac{49.11 - 47.5}{2.6194}\sqrt{9}$$

$$\boxed{t = 1.8439} < t_{0.05} = 2.306.$$

Thus the hypothesis H_0 is accepted.

ie there is no significant difference b/w $\bar{x}$ and μ.

7) A mechanist is making engine parts with axle diameter of 0.7 inch. A random sample of 10 parts shows mean diameter 0.742 inch with a standard deviation of 0.04 inch. On the basis of this sample, would you say that the work is inferior? → low quality.

Sol^n :- By data, $n = 10$, $\bar{x} = 0.742$, $s = 0.04$, $\mu = 0.7$

Let H_0 : The product is not inferior.

ie there is no significant difference b/w $\bar{x}$ & μ

$$t = \frac{\bar{x} - \mu}{s}\sqrt{n} = \frac{(0.742) - 0.7}{0.04}\sqrt{10}.$$

$$t = 3.32$$

For 9 d.f, $t_{0.05} = 2.262$

$\therefore \quad t = 3.32 > 2.262$

$\therefore$ Ho is rejected.

$\Rightarrow$ there is a significant difference b/w $\bar{x}$ and μ.

$\therefore$ the work is inferior.

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

Soln :- Let Ho : There is no significant difference b/w the sample mean and the population mean.

ie Ho : $\mu = 56$.

By data, $n = 16$, $\bar{x} = 53$, $\mu = 56$.

and $\sum (x - \bar{x})^2 = 135$

$\therefore \quad s^2 = \frac{1}{n-1} \sum (x - \bar{x})^2 = \frac{1}{15}(135) = 9 \Rightarrow \boxed{s = 3}$

Thus $t = \frac{\bar{x} - \mu}{s} \sqrt{n} = \frac{53 - 56}{3} \sqrt{16} = -4$

$\therefore \quad |t| = 4 > t_{0.05} = 2.131$ (for 15 d.f)

The Hypothesis Ho is rejected.

ie the sample mean has not come from a population having 56 as mean.

95% confidence limits for μ are $\bar{x} \pm \frac{s}{\sqrt{n}} t_{0.05}$

$= 53 \pm \frac{3}{\sqrt{16}}(2.131) = 53 \pm 1.5983$

$= 54.5983, 51.4017$

99% confidence limits for μ are $\bar{x} \pm \frac{s}{\sqrt{n}} t_{0.01}$

$= 53 \pm \frac{3}{4}(2.947) = 53 \pm 2.2103$

$= 55.2103, 50.7897$.

Problems on Student's 't' for 2 samples.

1) A group of boys and girls were given an intelligence test. The mean score, S.D score and numbers in each group are as follows:

	Boys	Girls
Mean	74	70
SD	8	10
n	12	10

Is the difference b/w the means of the 2 groups significant at 5% level of significance? ($t_{0.05} = 2.086$ for 20 d.f)

Soln :- By data, $\bar{x} = 74$, $s_1 = 8$, $n_1 = 12$ (Boys)

$\bar{y} = 70$, $s_2 = 10$, $n_2 = 10$ (Girls)

we have $t = \dfrac{\bar{x} - \bar{y}}{s\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}}$

where $s^2 = \dfrac{n_1 s_1^2 + n_2 s_2^2}{n_1 + n_2 - 2} = 88.4$

$\Rightarrow \boxed{s = 9.4}$

$\therefore \; t = \dfrac{74 - 70}{9.4\sqrt{\frac{1}{12} + \frac{1}{10}}} \Rightarrow t = 0.994 < t_{0.05} = 2.086.$

Let Ho : There is no significant difference b/w the means of the 2 groups.

Thus Ho is accepted at 5% level of significance.

2) Two types of batteries are tested for their length of life and the foll. results were obtained.

Battery A : $n_1 = 10$, $\bar{x}_1 = 500$ hrs, $s_1^2 = 100$

Battery B ; $n_2 = 10$, $\bar{x}_2 = 560$ hrs, $s_2^2 = 121$

compute student's 't' and test whether there is a significant difference in the 2 means.

Soln :- $t = \dfrac{\bar{x}_2 - \bar{x}_1}{s\sqrt{\frac{1}{n_1}+\frac{1}{n_2}}}$ where $s^2 = \dfrac{n_1 s_1^2 + n_2 s_2^2}{n_1+n_2-2}$

$s^2 = 122.78$

$\Rightarrow \boxed{s = 11.0805}$

For $\nu = n_1+n_2-2 = 18$ d.f, $t_{0.05} = 2.101$.

$\therefore t = \dfrac{560-500}{11.0805\sqrt{\frac{1}{10}+\frac{1}{10}}} = 12.1081 > t_{0.05} = 2.101.$

Let Ho : There is no significant difference in the 2 means.

Thus Ho is rejected.

3) A group of 10 boys fed on a diet A and another group of 8 boys fed on a different diet B for a period of 6 months recorded the foll. increase in weights (in pounds)

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

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

Soln :- Let Ho : The 2 diets do not differ significantly regarding their effect on increase in weight.

Let x correspond to diet A and y to diet B.

$\bar{x} = \dfrac{\Sigma x}{n_1} = \dfrac{64}{10} = 6.4$; $\bar{y} = \dfrac{\Sigma y}{n_2} = \dfrac{40}{8} = 5.$

$\Sigma(x-\bar{x})^2 = 102.4$; $\Sigma(y-\bar{y})^2 = 82.$

$\therefore s^2 = \dfrac{1}{n_1+n_2-2}\left[\sum_1^{n_1}(x-\bar{x})^2 + \sum_1^{n_2}(y-\bar{y})^2\right]$

$s^2 = \frac{1}{16}(102.4+82) = 11.525 \Rightarrow \boxed{s = 3.395}$

$\therefore t = \dfrac{\bar{x}-\bar{y}}{s\sqrt{\frac{1}{n_1}+\frac{1}{n_2}}} = \dfrac{6.4-5}{3.395\sqrt{\frac{1}{10}+\frac{1}{8}}} = 0.8694 < t_{0.05} = 2.120$ (for 16 d.f)

Thus Ho is accepted.

4) Eleven students were given a test in Statistics. They were (6)
QP given a month's further tution & second test of equal difficu-
-lty was held at the end of it. Do the marks give
evidence that the students have benefited by extra coaching.

Boys	1	2	3	4	5	6	7	8	9	10	11
Marks I test x	23	20	19	21	18	20	18	17	23	16	19
Marks II test y	24	19	22	18	20	22	20	20	23	20	17

Soln :- Let H_0 : There is no significant difference b/w the means of 2 test Marks.

ie the students have not benefited by extra coaching.

$$\bar{x} = \frac{\Sigma x}{n_1} = \frac{214}{11} = 19.45 \quad ; \quad \bar{y} = \frac{\Sigma y}{n_2} = \frac{225}{11} = 20.45$$

$$\Sigma (x-\bar{x})^2 = 50.7275 \quad ; \quad \Sigma (y-\bar{y})^2 = 44.7275$$

$$\therefore \quad s^2 = \frac{1}{n_1+n_2-2} \left[\Sigma (x-\bar{x})^2 + \Sigma (y-\bar{y})^2 \right]$$

$$= \frac{1}{20} \left[50.7275 + 44.7275 \right]$$

$$= 4.7728$$

$$\boxed{s = 2.1847}$$

$$\therefore \quad t = \frac{\bar{x}-\bar{y}}{s\sqrt{\frac{1}{n_1}+\frac{1}{n_2}}} = \frac{19.45-20.45}{(2.1847)\sqrt{\frac{1}{11}+\frac{1}{11}}} = -1.0735 \quad \cancel{< t_{0.05} = 2.08 \text{ (for 20 d.f)}}$$

$$\therefore \quad |t| = 1.0735 < t_{0.05} = 2.086 \text{ (for 20 d.f)}$$

Thus H_0 is accepted at 5% level of significance.
ie the students have not benefited by extra coaching.

Chi-square distribution

$$\chi^2 = \sum_{i=1}^{n} \frac{(O_i - E_i)^2}{E_i}$$ where O_i and E_i are respectively the observed and estimated (theoritical) frequencies. $(i = 1, 2, 3 \ldots n)$

Note :- 1) As a test of goodness of fit, the value of chi-square is used to study the correspondence b/w observed and the theoritical frequencies.

2) d.f = n−1 if the data is given in a series of 'n' no. (PTO) for note ⑦

3) If the expected frequencies are less than 5, we group them suitably for computing the value of chi-square.

4) $\Sigma O_i = \Sigma E_i = n$ = total frequency for n−1 d.f.

P.T.O for note 5 & 6.

Problems

1) A die is thrown 264 times and the number appearing on the face (x) follows the foll. frequency distribution

QP

{ unbiased → fair coin whose prob = 1/2
biased → unfair coin }

x	1	2	3	4	5	6
f	40	30	26	56	52	60

Calculate the value of χ^2 and test the hypothesis that the die is unbiased given that $\chi^2_{0.05}(5) = 11.07$ and $\chi^2_{0.01}(5) = 15.09$

Sol^n :- The frequencies in the given data are the observed frequencies. Assuming the hypothesis that the dice is unbiased, the expected no. of frequencies for the no's 1, 2, 3, 4, 5, 6 to appear on the face is $\frac{264}{6} = 44$ each. Thus we have

No. on the dice	1	2	3	4	5	6
observed frequency (O_i)	40	30	26	56	52	60
Expected frequency (E_i)	44	44	44	44	44	44

$$\chi^2 = \frac{\Sigma (O_i - E_i)^2}{E_i} = \frac{(40-44)^2}{44} + \frac{(30-44)^2}{44} + \cdots + \frac{(60-44)^2}{44}$$

$\chi^2 = 22.73 > \chi^2_{0.05}$ and $> \chi^2_{0.01}$

∴ hyp is rejected ⇒ the die is biased.

Scanned by CamScanner

2) The following table gives the no. of aircraft accidents that occured during the various days of the week. find whether the accidents are uniformly distributed over the week. Given $\psi^2_{0.05} = 12.59$ for (6 d.f).

Days	Sun	Mon	Tue	wed	Thur	fri	Sat	Total
No. of accidents	14	16	8	12	11	9	14	84

(Let There is no significant difference b/w O_i and E_i)

Sol^n :- Let us take the hyp that the accidents are uniformly distributed over the week.

The No. of accidents on different days of the week are the observed frequencies. The expected frequencies assuming the accidents are uniformly distributed is $\frac{84}{7} = 12$ each.

we have

Days	Sun	Mon	Tue	wed	Thur	fri	sat
O_i	14	16	8	12	11	9	14
E_i	12	12	12	12	12	12	12

$$\therefore \psi^2 = \sum \frac{(O_i - E_i)^2}{E_i} = \left[(14-12)^2 + (16-12)^2 + \cdots + (14-12)^2\right] \times \frac{1}{12}$$

$$\psi^2 = \frac{41}{12} = 3.42 < \psi^2_{0.05} = 12.59 \quad (\text{for } 6 \text{ d.f}).$$

Thus H_0 is accepted at 5% level of significance.

Note :- 5) If the calculated value of χ^2 is less than the corresponding tabulated value, then we accept the null hypothesis and conclude that there is a good correspondence b/w theory & experiment.

6) If $\psi^2 > \psi^2_{0.05}$, then we reject the null hypothesis and conclude that the experiment does not support the theory.

7) For B.D, d.f = n−1, for P.D, d.f = n−2
for N.D, d.f = n−3.

3) A sample analysis of examination results of 500 students was made. It was found that 220 students had failed, 170 had secured third class, 90 had secured second class, and 20 had secured first class. DO these figures support the general examination result which is in the ratio 4:3:2:1 for the respective categories ($\chi^2_{0.05} = 7.81$ for 3 d.f) (2)

Soln :- Let us take the hypothesis that these figures support to the general result in the ratio 4:3:2:1

∴ The expected frequencies in the respective category are

$\frac{4}{10} \times 500$, $\frac{3}{10} \times 500$, $\frac{2}{10} \times 500$, $\frac{1}{10} \times 500$

ie 200, 150, 100, 50

O_i	220	170	90	20
E_i	200	150	100	50

$$\therefore \quad \chi^2 = \frac{\sum (O_i - E_i)^2}{E_i}$$

$$= \frac{(220-200)^2}{200} + \frac{(170-150)^2}{150} + \frac{(90-100)^2}{100} + \frac{(20-50)^2}{50}$$

$$= 23.67 > \chi^2_{0.05} = 7.81$$

Thus the hypothesis is rejected at 5% level of significance.

4) QP Genetic theory states that children having one parent of blood type M and other of blood type N will always be one of the three types M, MN, N and that the proportions of these types will be on an average 1:2:1. A report says that out of 300 children having one M parent and one N parent, 30% were found to be of type M, 45% of Type MN and the remaining of type N. Test the theory of χ^2 test. Given $\chi^2_{0.05} = 5.99$ for 2 d.f.

Soln :- Let us assume the hypothesis that there is a good correspondence b/w observed and the theoretical frequencies.

By data, Observed frequencies are

M	MN	N	Total
$\frac{30}{100} \times 300 = 90$	$\frac{45}{100} \times 300 = 135$	$\frac{25}{100} \times 300 = 75$	300

Proportions of these types are 1 : 2 : 1 (theory)

$\therefore$ The corresponding Expected frequencies are

$\frac{1}{4} \times 300$, $\frac{2}{4} \times 300$, $\frac{1}{4} \times 300$

ie 75 , 150 , 75.

Thus we have

	M	MN	N
O_i	90	135	75
E_i	75	150	75

$$\therefore \quad \chi^2 = \frac{\sum (O_i - E_i)^2}{E_i} = \frac{(90-75)^2}{75} + \frac{(135-150)^2}{150} + \frac{(75-75)^2}{75}$$

$$\chi^2 = 4.5 < \chi^2_{0.05} = 5.99 \quad (\text{for } 2 \text{ d.f})$$

$\therefore$ the hypothesis is accepted.

5) fit a Poisson distribution for the foll. data and test the goodness of fit given that $\chi^2_{0.05} = 7.815$ for 3 d.f.

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

Soln :- Let us assume the hypothesis that the fitness is good.

TO find theoretical frequencies :- (E_i)

$$\text{Mean } \mu = \frac{\sum fx}{\sum f} = \frac{0 + 60 + 30 + 6 + 4}{200} = \boxed{0.5 = m}$$ for

Poisson distribution.

WKT $P(x) = \frac{m^x e^{-m}}{x!}$ and Let $f(x) = 200 \; P(x)$.

Scanned by CamScanner

$$\therefore \quad f(x) = 200 \frac{(0.5)^x e^{-0.5}}{x!}$$

$$\boxed{f(x) = 121.3 \frac{(0.5)^x}{x!}}$$

Putting $x = 0, 1, 2, 3, 4$ in $f(x)$, we get E_i's.

Thus we have

x	0	1	2	3	4
O_i	122	60	15	2	1
E_i	121	61	15	3	0

Since the last of the expected frequency is 0, we club it with the previous one and write

O_i	122	60	$15+2+1 = 18$	~~2+1=3~~
E_i	121	61	$15+3+0 = 18$	~~3+0=3~~

$$\therefore \quad \chi^2 = \sum \frac{(O_i - E_i)^2}{E_i} = \frac{(122-121)^2}{121} + \frac{(60-61)^2}{61} + 0 + 0.$$

$$= 0.025 < \chi^2_{0.05} = 7.815$$

$\therefore$ The fitness is considered good $\Rightarrow$ hyp. is accepted.

6) The no. of accidents per day (x) as recorded in a textile
HW industry over a period of 400 days is given below. Test the goodness of fit in respect of Poisson distribution for the foll. data: (given $\chi^2_{0.05} = 9.49$ for 4 d.f)

x :	0	1	2	3	4	5
f :	173	168	37	18	3	1

Soln :- let us assume the hypothesis that the fitness is good.

Mean, $\mu = m$ (for P.D) $= \frac{\sum fx}{\sum f} = 0.7825$

$$P(x) = \frac{m^x e^{-m}}{x!} \quad \text{and let} \quad f(x) = 400\, P(x)$$

$$\therefore \quad f(x) = 400 \frac{(0.7825)^x e^{-0.7825}}{x!}$$

$$\Rightarrow \quad f(x) = 182.9 \, \frac{(0.7825)^x}{x!}$$

Putting $x = 0, 1, 2, 3, 4, 5$, we get E_i's.

x	0	1	2	3	4	5	
O_i	173	168	37	18	3	1	22 ~~3+1=4~~
E_i	183	143	56	15	3	0	18 ~~3+0=3~~

we club last 2.

$$\therefore \quad \chi^2 = \sum \frac{(O_i - E_i)^2}{E_i} = 12.252 > \chi^2_{0.05} = 9.49.$$

Thus the hypothesis is rejected.

$\Rightarrow$ fitness is not considered good.

7) 4 coins are tossed 100 times and the foll. results were obtained. Fit a binomial distribution for the data and test the goodness of fit ($\chi^2_{0.05} = 9.49$ for 4 d.f)

No. of heads	0	1	2	3	4
Frequency	5	29	36	25	5

soln:- Let x denote the no. of heads & f the corresponding frequency. Assume the hyp. that the fitness is good.

$$\text{Mean}, \ \mu = \frac{\sum fx}{\sum f} = \frac{196}{100} = 1.96.$$

But for binomial distribution, Mean, $\mu = np$.

$$1.96 = 4p$$

$$\Rightarrow p = 0.49 \Rightarrow q = 0.51.$$

we have $P(x) = {}^nC_x \, p^x q^{n-x} = {}^4C_x (0.49)^x (0.51)^{4-x}$

Let $f(x) = 100 \times P(x) = 100 \times {}^4C_x (0.49)^x (0.51)^{4-x}$

Putting $x = 0, 1, 2, 3, 4$, we get the E_i's frequencies.

(4)

x	0	1	2	3	4
O_i	5	29	36	25	5
E_i	7	26	37	24	6

$$\therefore \quad \chi^2 = \sum \frac{(O_i - E_i)^2}{E_i} = \frac{(5-7)^2}{7} + \frac{(29-26)^2}{26} + \dots + \frac{(5-6)^2}{6}$$

$$= 1.15 < \chi^2_{0.05} = 9.49$$

Thus the hyp. that the fitness is good is accepted.

8) Five dice were thrown 96 times and the no.'s 1, 2 or 3 appearing on the face of the dice follows the frequency distribution as below

No. of dice showing 1, 2 or 3 (x)	5	4	3	2	1	0
Frequency (f)	7	19	35	24	8	3

Test the hyp. that the data follows a Binomial distribution.
($\chi^2_{0.05} = 11.07$ for 5 d.f) PTO.

Soln :- Let us assume the hyp. that the data follows a Binomial dist.

Mean, $\mu = \frac{\Sigma fx}{\Sigma f} = \frac{272}{96} = 2.83$

WKT $\mu = np$.

$2.83 = 5p \implies p = 0.566 \approx 0.57 \implies q = 0.43$.

$\therefore f(x) = 96 \times P(x)$

$= 96 \times {}^nC_x \, p^x q^{n-x}$

$= 96 \times {}^5C_x (0.57)^x (0.43)^{5-x}$

O_i	7	19	35	24	8	3
E_i	6	22	33	25	9	1

club $8+3=11$

$9+1=10$

$\therefore \chi^2 = \sum \frac{(O_i - E_i)^2}{E_i} = \frac{(7-6)^2}{6} + \frac{(19-22)^2}{22} + \frac{(35-33)^2}{33} + \frac{(24-25)^2}{25} + \frac{(11-10)^2}{10}$

$\chi^2 = 0.837 < \chi^2_{0.05} = 11.07$

Thus the hyp. is accepted.