UNIT- III

TESTING OF HYPOTHESIS (sampling)

* population : The collection of individuals (or) attributes is called population.

* Sample : It is finite subset of population.

* Sample Size : The number of elements is called sample size and it is denoted by 'n'.

* SMALL SAMPLE : $n \leq 30$

* LARGE SAMPLE : $n > 30$

* Parameter : The statistical values like mean, variance, mode... of the population are called parameter

* Statistic : The statistical values like mean, variance, mode... of the sample are called statistic.

* Sampling : It is a process of selection of sample from the population.

Notation :

Attributes	Population parameter	Sample statistic
Size	N	n
Mean	μ	$\bar{x}$
SD	σ	s
proportion	P	ϕ

* NULL HYPOTHESIS : H_0

It is a definite statement about the population parameter

* Alternative Hypothesis : H_1

It is a compliment of Null hypothesis.

* Level Of Significance (LOS) : α

It is the level of probability in which the null hypothesis is rejected.

The common L.O.S used are 5%, 1%, 10%.

* Critical region (or) rejection region:

The region under which the null hypothesis is rejected in a normal distribution is called critical region

* Acceptance : It is the region where null hypothesis accepted

→ Errors Of Hypothesis :

Type 1 : Rejecting null hypothesis when it is true. This is also called as producer risk.

Type II error : Accepting null hypothesis which it is false. This is called consumer risk.

→ PROCEDURE FOR TESTING HYPOTHESIS :

Step 1 : Fix Null hypothesis

Step 2 : Write Alternative Hypothesis

Step 3 : Fix L.O.S

Step 4 : Test Statistic (formula)

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

Step 5 : Compare calculated 'z' with tabulated 'z'.

if $z_{cal} < z_{tab}$, Accept H_0

if $z_{cal} > z_{tab}$, Reject H_0

Standard Error: It is the standard deviation of sample.

Sample

Small sample — Large sample

$n \leq 30$ — $n > 30$

$\downarrow$ — $\downarrow$

't' - test (Mean) — Normal distribution

F - test (variance)

χ^2 - test (Attribute)

Small sample:

→ standard's 't' - test

Application:

1. It is used to test for single mean
2. It is used to test the difference between mean
3. It is used for finding correlation of samples

Test 1 — Test for single mean

The test statistic is

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

where, $\bar{x}$ – sample mean

μ – population mean

s = Sample S.D

n – sample size

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

degree of freedom $= n-1$

(Q) A sample of 26 bulbs gives a mean life of 990 hrs, with SD 20hr. The manufacturer claims that, the mean life of bulb is 1000 hrs. Is the sample upto the standard.

$$n = 26 < 30$$

Given sample is small sample

Step 1: Null Hypothesis H_0:
$$\mu = 1000\text{hrs}$$

Step 2: Alternative Hypothesis H_1:
$$\mu \neq 1000\text{hrs}$$

Step 3:
LOS: 5%

Test static: $t = \dfrac{\bar{x} - \mu}{s/\sqrt{n-1}}$

$\bar{x} = 990$, $\mu = 1000$, $s = 20$, $n = 26$

$$t = \frac{990 - 1000}{20/\sqrt{25}} = \frac{-25}{10} = -2.5$$

$|t| = 2.5$

$t_{cal} = 2.5$, degree of freedom $(n-1) = 25$

At 5% LOS, for 25 dof, $t_{tab} = 2.06$

Conclusion: $t_{cal} > t_{tab}$, Reject $H_0 \Rightarrow \mu \neq 1000\text{hrs}$

Test 2: Testing difference of mean

Test Statistic is $t = \dfrac{\bar{x}_1 - \bar{x}_2}{s\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}}$

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

(or)

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

dof is $n_1 + n_2 - 2$

(Q) A group of 10 rats, fed on diet A and another group of 8 rats, fed on deit B and weight was recorded. Test diet A superior over diet B.

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

Diet B : 2 3 6 8 10 1 2 8 – –

Sol: Null hypothesis H_0: There is no significant difference b/w diet A and diet B. $\mu_1 = \mu_2$

Alternative hypothesis H_1: $\mu_1 > \mu_2$ (one tailed-test)

L.O.S : 5%

d.o.f : $n_1+n_2-2 = 10+8-2 = 16$

The test statistic is

$$t = \frac{\bar{x}_1 - \bar{x}_2}{s\sqrt{\frac{1}{n_1}+\frac{1}{n_2}}} \qquad \bar{x}_1 = \frac{\Sigma x_1}{n_1}, \quad \bar{x}_2 = \frac{\Sigma x_2}{n_2}$$

$$s_1^2 = \frac{\Sigma(x_1-\bar{x}_1)^2}{n_1}, \quad s_2^2 = \frac{\Sigma(x_2-\bar{x}_2)^2}{n_2}$$

x_1	x_2	$x_1-\bar{x}_1$ $(x_1-6.4)$	$x_2-\bar{x}_2$ (x_2-5)	$(x_1-\bar{x}_1)^2$	$(x_2-\bar{x}_2)^2$
5	2	-1.4	-3	1.96	9
6	3	-0.4	-2	0.16	4
8	6	1.6	1	2.56	1
1	8	-5.4	3	20.16	9
12	10	5.6	5	31.16	25
4	1	-2.4	-4	3.76	16
3	2	-3.4	-3	11.56	9
9	8	2.6	3	6.76	9
6	-	-0.4	-	0.16	-
10	-	3.6	-	10.96	-
				102.21	82

$\bar{x} = 5, \bar{x}_1 = 5, \; s_1^2 = 102.21, \; s_2^2 = 10.25, \; n_1 = 10, \; n_2 = 8$

$$s^2 = \frac{n_1 s_1^2 + n_2 s_2^2}{n_1+n_2-2} = 11.5 \Rightarrow s = 3.34$$

$$t = \frac{6.4-5}{3.34\sqrt{\frac{1}{10}+\frac{1}{8}}} = 0.869$$

$$|t|_{cal} = 0.869$$

At 5%, LOS, for 16 d.o.f, $t_{tab} = 2.12$

$t_{cal} < t_{tab}$

H_0 is accepted

Conclusion: There is no significant difference b/w the two groups

(Q) The IQ of 16 students of 1 area of a city gives mean 107 to the SD = 10, while the IQ of 14 students from another area of city gives the mean 112, SD = 8. Is there any significant difference IQ's of 2 groups at (i) 1% L.O.S (ii) 5% LOS

Sol:
$\bar{x}_1 = 107$ $\bar{x}_2 = 112$
$s_1 = 10$ $s_2 = 8$
$n_1 = 16$ $n_2 = 14$

$df = n_1+n_2-2 = 16+14-2 = 28$

NH- H_0: $\mu_1 = \mu_2$. There is no significant difference b/w IQ's of two groups.

AH H_1: $\mu_1 \neq \mu_2$

$$s^2 = \frac{n_1 s_1^2 + n_2 s_2^2}{n_1+n_2-2} = \frac{(1600)+896}{28} = 89.14 \Rightarrow S = 9.44$$

Test statistic: $t = \frac{\bar{x}_1 - \bar{x}_2}{s\sqrt{\frac{1}{n_1}+\frac{1}{n_2}}} = \frac{107-112}{9.44\sqrt{\frac{1}{16}+\frac{1}{4}}} \Rightarrow t = 1.5$

$|t| = 1.5$

(i) 1% LOS for 28 df, $t_{tab} = 2.76 \Rightarrow t_{cal} < t_{tab}$ accept H_0

(ii) 5% LOS for 28 df, $t_{tab} = 2.65$

$t_{cal} < t_{tab}$, Accept H_0

There is no significant difference b/w IQ's of two groups

(Q) Two horses A and B were tested according to time (in sec) to run a particular track to the following results

Horse : A 28 30 32 33 33 29 34

Horse B 29 30 30 24 27 27 –

Test whether you can differentiate with the horses use the fact that 5% value of B for 11 d.o.f is 2.2

Null Hypothesis H_0 :

Two horses are equally efficient

$\mu_1 = \mu_2$

Alternative Hypothesis H_1 :

$\mu_1 \neq \mu_2$

L.O.S = 5%

$d.o.f = n_1 + n_2 - 2$

$= 7 + 6 - 2$

$= 11$

Test Statistic :

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

$$\bar{x}_1 = \frac{\Sigma x_1}{n_1}, \quad \bar{x}_2 = \frac{\Sigma x_2}{n_2}$$

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

x_1	x_2	$x_1 - \bar{x}_1$ $(x_1 - 31.28)$	$x_2 - \bar{x}_2$ $(x_2 - 27.83)$	$(x_1 - \bar{x}_1)^2$	$(x_2 - \bar{x}_2)^2$
28	29	-3.28	1.17	10.75	1.36
30	30	-1.28	2.17	1.63	4.70
32	30	0.72	2.17	0.518	4.70
33	24	1.72	-3.83	2.95	14.66
33	27	1.72	-0.83	2.95	0.68
29	27	-2.28	-0.83	5.19	0.65
34	–	2.72	–	7.39	–
				31.37	26.84

Scanned by CamScanner

$\Sigma x_1 = 219 \quad \Sigma x_2 = 167$

$\bar{x}_1 = \frac{219}{7} = 31.285$

$\bar{x}_2 = \frac{167}{6} = 27.83$

$s^2 = \frac{(31.37) + (26.84)}{11}$

$s^2 = 5.29$

$s = \sqrt{5.29} = 2.3$

$t = \frac{31.28 - 27.83}{2.3\sqrt{\frac{1}{7} + 1/6}} = \frac{3.45}{2.3\sqrt{13/42}}$

$= 2.69$

$|t| = 2.69$

$t_{cal} = 2.69$

At 5% LOS, 11 d.o.f, $t_{tab} = 2.2$

$t_{cal} > t_{tab}$

Conclusion: Two horses are not equally efficient

H_0 is rejected. //

Test 3: Testing for paired observation (i.e correlated (dependent))

paired 't' - test:

The test statistic is $t = \frac{\bar{d}}{s/\sqrt{n-1}}$

where, $\bar{d} = \frac{\Sigma d}{n}$, $d = x_1 - x_2$

$s^2 = \frac{\Sigma d^2}{n} - \bar{d}^2$, $df = n-1$

(Q) 11 students were given in test in mathematics. They were given 1month stution and 2nd test was held. At the end of tution do the marks provide the students are benifited by extra coaching

Test 1 :	23	20	19	21	18	20	18	17	23	16	19
Test 2 :	24	19	22	18	20	22	20	20	23	20	18

Sol: NH H_0: There is no signified difference b/w the marks in Test 1 & Test 2

AH H_1: $\mu_1 \neq \mu_2$

LOS : 5%

d.o.f : $n-1 = 11-1$
$= 10$

Test statistics :

$$t = \frac{\bar{d}}{s/\sqrt{n-1}} \quad , \bar{d} = \frac{\Sigma d}{n}$$

$$s^2 = \frac{\Sigma d^2}{n} - \bar{d}^2$$

x_1	x_2	$d = x_1 - x_2$	d^2
23	24	-1	1
20	19	1	1
19	22	-3	9
21	18	3	9
18	20	-2	4
20	22	-2	4
18	20	-2	4
17	20	-3	9
23	23	0	0
16	20	-4	16
19	18	1	1

$\Sigma d^2 = 58$

$\Sigma d = -12$

$s^2 = \frac{58}{11} - (-1.09)^2$

$= 4.08$

$s = 2.02$

$t = \frac{-1.09}{2.02\sqrt{10}} = -1.71$

$|t| = 1.71$

$t_{cal} = 1.71$

At 5% LOS, 10 df, $t_{tab} = 2.23$

$t_{cal} < t_{tab}$

H_0 is accepted

Conclusion: The students are not benifited by extra Coaching.

(2) A company is testing 2 machines, the random sample of 8 emplyees are selected. each employee was given each machine for 1hr. The no of components produced

Machine employee :	1	2	3	4	5	6	7	8
Machine 1 :	96	107	84	99	102	87	93	101
Machine 2 :	99	112	90	97	108	97	94	98

Test whether there is a difference of Mean no of components produced

Sol NH H_0: $\mu_1 = \mu_2$

AH H_1: $\mu_1 \neq \mu_2$

Test statistics: $t = \frac{\bar{d}}{s/\sqrt{n-1}}$ $n = 8$

$s^2 = \frac{\Sigma d^2}{n} - \bar{d}^2$

x_1	x_2	d	d^2
96	99	-3	9
107	112	-5	25
84	90	-6	36
99	97	2	4
102	108	-6	36
87	97	-10	100
93	94	1	1
101	98	3	9

LOS 5%

$df = n-1 = 7$

$\bar{d} = \frac{-26}{8} = -3.25$

$s^2 = \frac{220}{8} - (-3.25)^2 = 16.94$

$s = 4.115$

$t = \frac{3.25}{(4.115)(\sqrt{7})} = -2.09$

$|t| = 2.09$

$t_{cal} = 2.09$

$t_{tab} = 2.36$

at 5% LOS, 7df

$t_{cal} < t_{tab}$

Accepted H_0

$\Rightarrow \mu_1 = \mu_2$ //

Snedecor's F-test:

The test statistic is

$$F = \frac{S_1^2}{S_2^2}, \text{ if } S_1^2 > S_2^2$$

(or)

$$F = \frac{S_2^2}{S_1^2} \text{ if } S_2^2 > S_1^2$$

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

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

df is $(n_1 - 1, n_2 - 1)$ (or) $(n_2 - 1, n_1 - 1)$

Application:

1) Testing significant difference b/w population variances

Type-2: Testing whether the samples are drawn from the same population.

(*C)

Test 1 : Testing difference b/w population Variances

(a) In a sample of 10 observations the sum of squares of deviation of sample value from the mean was 120. Another sample of 12 observations it was 314. Test whether there is a significant difference b/w the sample at 5%.

$n_1 = 10$, $n_2 = 12$

$\Sigma (x_1 - \bar{x}_1)^2 = 120$

$\Sigma (x_2 - \bar{x}_2)^2 = 314$

$S_1^2 = \frac{120}{10} = 12$

$S_2^2 = \frac{314}{12} = 26.16$

NH H_0 : $\sigma_1^2 = \sigma_2^2$

AH H_1 : $\sigma_1^2 \neq \sigma_2^2$

Test Statistic :

$F = \frac{26.16}{12} = 2.18$

LOS ; 5%

df : $(n_2 - 1, n_1 - 1) = (11, 9)$

$F_{cal} = 2.18$

$F_{tab} = 3.15$

$F_{cal} < F_{tab}$

H_0 is accepted

$\sigma_1^2 = \sigma_2^2$ //

2) The random samples from a normal population given the following result

Sample I : 20 16 26 27 23 22 —

Sample II : 27 33 42 35 32 34 38

Test whether the samples are drawn from same population

Sol: To test for samples came for same population, we have to apply

(1) t - test

(11) F - test

NH H_0 : $\mu_1 = \mu_2$, $\sigma_1^2 = \sigma_2^2$

The samples are drawn from same population

AH H_1 : $\mu_1 \neq \mu_2$, $\sigma_1^2 \neq \sigma_2^2$

t - test:

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

x_1	x_2	$x_1-\bar{x}_1$	$x_2-\bar{x}_2$	$(x_1-\bar{x}_1)^2$	$(x_2-\bar{x}_2)^2$
20	27	-2.33	-7.43	5.42	55.20
16	33	-6.33	-1.43	40.06	
26	42	3.67	7.57	13.96	
27	35	4.67	0.57	21.8	
23	32	0.67	-2.43	0.44	
22	34	-0.33	-0.43	0.1	
-	38	-	4.43	-	
134	241			81.504	133.68

$$\bar{x}_1 = \frac{\Sigma x_1}{n_1} = \frac{134}{6} = 22.3$$

$$\bar{x}_2 = \frac{\Sigma x_2}{n_2} = \frac{241}{6} = 34.43$$

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

$$= \frac{81.504 + 133.68}{6+7-2}$$

$$= 19.56$$

$$\boxed{s = 4.42}$$

$$t = \frac{22.3 - 34.43}{4.42\sqrt{1/6+1/7}}$$

$$\boxed{t = 4.95}$$

At 5% LOS, 11df, $t_{tab} = 2.2$

$t_{cal} = 4.95$

$t_{cal} > t_{tab}$

F-test:

$$S_1^2 = \frac{\Sigma(x_1-\bar{x}_1)^2}{n_1} \quad , \quad S_2^2 = \frac{\Sigma(x_2-\bar{x}_2)^2}{n_2}$$

$$= 13.384 \qquad\qquad = 19.11$$

$$\cancel{F} = \frac{13.\cancel{3}84}{19.11} = \cancel{0}.7\cancel{1}0$$

$$F = \frac{19.11}{13.384} = 1.388$$

$F_{cal} = 1.388$

At (6,5) df, $F_{tab} = 4.95$, $F_{cal} < F_{tab} \Rightarrow \sigma_1^2 = \sigma_2^2$

Since $\mu_1 \neq \mu_2$

Samples are not drawn from same population

1) The nicotin contained in two random samples of tobaco are given below.

Sample I	21	24	25	26	27	
Sample II	22	27	28	30	31	36

Can you say that the two samples came from same population

Sol: $n_1 = 5$, $n_2 = 6$

$$\bar{x}_1 = \frac{\Sigma x_1}{n_1}, \quad \bar{x}_2 = \frac{\Sigma x_2}{n_2}$$

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

$$S_1^2 = \frac{\Sigma(x_1-\bar{x}_1)^2}{n_1} \qquad S_2^2 = \frac{\Sigma(x_2-\bar{x}_2)^2}{n_2}$$

x_1	x_2	$x_1-\bar{x}_1$ $(x_1-24.6)$	$x_2-\bar{x}_2$ (x_2-29)	$(x_1-\bar{x}_1)^2$	$(x_2-\bar{x}_2)^2$
21	22	-3.6	-7	12.96	49
24	27	-0.6	-2	0.36	4
25	28	0.4	-1	0.16	1
26	30	1.4	1	1.96	1
27	31	2.4	2	5.76	4
—	36	—	7	—	49
123	174			23.6	108

$$\bar{x}_1 = \frac{\Sigma x_1}{n_1} = \frac{123}{5} = 24.6$$

$$\bar{x}_2 = \frac{\Sigma x_2}{n_2} = \frac{174}{6} = 29$$

$$s^2 = \frac{\Sigma(x_1-\bar{x}_1)^2 + \Sigma(x_2-x_1)^2}{n_1+n_2-2} = \frac{23.6+108}{5+6-2} = \frac{344}{9}$$

$$= 14.62$$

$$s = \sqrt{14.62} = 3.82$$

$$S_1^2 = \frac{\Sigma(x_1-\bar{x}_1)^2}{n_1} = \frac{23.6}{5} = 4.7$$

$$S_2^2 = \frac{\Sigma(x_2-\bar{x}_2)^2}{n_2} = \frac{108}{6} = 18$$

NH H_0: The samples are drawn from same population

$$\mu_1 = \mu_2 \ , \ \sigma_1^2 = \sigma_2^2$$

AH H_1: $\mu_1 \neq \mu_2 \ , \ \sigma_1^2 \neq \sigma_2^2$

t-test: The test statistic

$$t = \frac{\bar{x}_1 - \bar{x}_2}{S\sqrt{\frac{1}{n_1}+\frac{1}{n_2}}} = \frac{24.6 - 29}{3.82\sqrt{\frac{1}{5}+\frac{1}{6}}}$$

$$\Rightarrow \frac{-4.4}{(3.82)(0.60)} = \frac{-4.4}{2.29}$$

$$= -1.91$$

$$|t| = 1.91$$

$$t_{cal} = 1.91$$

At 5% LOS, 9df, $t_{tab} = 3.26$

$$t_{cal} < t_{tab}$$

$$\Rightarrow \mu_1 = \mu_2$$

F-test: The test statistic is

$$F = \frac{S_2^2}{S_1^2} = \frac{18}{4.7} = 3.67$$

$$F_{cal} = 3.67$$

for $(n_2-1, n_1-1) = (5, 4)$ df

at 5% LOS

$$F_{tab} = 6.26$$

$F_{cal} < F_{tab} \Rightarrow \sigma_1^2 = \sigma_2^2$, H_0 is accepted

Both Samples came from same population //

* Chi- Square test (χ^2- test):

The test statistic is

$$\boxed{\chi^2 = \sum_i \frac{(O_i - E_i)^2}{E_i}}$$

Where, O_i - Observed frequency

E_i - Expected or Theoretical frequency

Application:

(1) Testing difference b/w observed and expected frequency

(2) Testing goodness of fit

(3) Testing Independance of Attributes

Condition for χ^2- test

(1) $4 \le n \le 16$, 'n' is no of observations

(2) Individual observed frequency should be ≥ 10

(3) If any $O_i < 10$, the frequency can be combined with neighbouring frequency.

Test I : Testing difference b/w observed and expected frequencies:

1) The following table gives no of accidents that occur during the various days of the week. Test whether the accidents are uniformly distributed over the week

Day :	Mon	Tue	Wed	Thu	Fri	Sat
No of accidents :	15	19	13	12	16	15

Sol: Expected frequency $= \frac{15+19+13+12+16+15}{6}$

$= \frac{90}{6} = 15$

NH H_0: The no of accidents are uniformly distributed over a week

$O_i = E_i$

AH H_1: $O_i \neq E_i$

Test statistic:

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

LOS = 5%

$df = n-1 = 6-1$
$= 5$

O_i	E_i	$O_i - E_i$	$(O_i - E_i)^2$	$\frac{(O_i-E_i)^2}{E_i}$
15	15	0	0	0
19	15	4	16	1.06
13	15	-2	4	0.26
12	15	-3	9	0.6
16	15	1	1	0.06
15	15	0	0	0
				1.98

$$\boxed{\chi^2 = 1.98}$$

At LOS 5%, 5df, $\chi^2_{tab} = 11.07$

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

H_0 is accepted

$O_i = E_i$

The no of accidents are distributed uniformly over a week //

Scanned by CamScanner

Q) Theorey predicts that the proportion of beans in 4 groups A, B, C & D are in the ratio 9 : 3 : 3 : 1. In an experiment among 1600 beans. the numbers in four groups where 882, ~~338~~ 313, 287, 118 ~~thes~~ the experiment support the theorey

Expected frequency :

$$E(A) = \frac{9}{16} \times 1600 = 900$$

$$E(B) = \frac{3}{16} \times 1600 = 300$$

$$E(C) = \frac{3}{16} \times 1600 = 300$$

$$E(D) = \frac{1}{16} \times 1600 = 100$$

Null Hypothesis : H_0

The experiment results support the Efficiency

$$O_i = E_i$$

AH H_1 : There is a significant difference b/w expected & Observed frequencies

Test statistic : $\chi^2 = \Sigma \frac{(O_i - E_i)^2}{E_i}$

$$\sum_i O_i = \sum_i E_i$$

O_i	E_i	$O_i - E_i$	$(O_i - E_i)^2$	$\frac{(O_i - E_i)^2}{E_i}$
882	900	-18	324	0.36
313	300	13	169	0.56
287	300	-13	169	0.56
118	100	18	324	3.24
1600	1600			$\chi^2 = 4.72$

$$\chi^2_{cal} = 4.72$$

$df = n-1 = 4-1 = 3$, At 5% LOS , 3 d f, $\chi_{tab} = 7.82$

Scanned by CamScanner

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

→ H_0 is accepted

→ The experimental result support the theorey.

Test 2: Testing Of goodness Of fit

1) Fit a binomial distribution for the following data

x:	0	1	2	3	4	5	6	Total
f:	5	15	28	12	7	8	5	80

find the goodness of fit.

Sol: The pmf of BD is

$$P(X=x) = nC_x p^x q^{n-x}, \quad x = 0,1,2..n$$

$n = 6$, Mean $= np$, Variance $= npq$

$$\text{Mean} = np = \frac{\sum fx}{\sum f}$$

$$= \frac{(0+15+56+36+28+40+30)}{80}$$

$$= \frac{205}{80} = 2.563$$

$np = 2.563$,

$$p = \frac{2.563}{6} = 0.427$$

$$q = 1-p = 1-0.427$$
$$= 0.573$$

$\Rightarrow n=6,\ p=0.427,\ q=0.573$

$$P(X=x) = 6C_x (0.427)^x (0.573)^{6-x},$$
$x = 0,1,2,..6$

Expected frequencies

$f(x) = N\,P(x)$ $\quad N = \sum f$

$= 80\,(P(x))$

Scanned by CamScanner

x	$P(X=x) = 6C_x (0.427)^x (0.573)^{6-x}$	Expected frequency $NP(x)$
0	$6C_0 (0.427)^0 (0.573)^6 = 0.035$	2.8
1	$6C_1 (0.427)(0.573)^5 = 0.158$	12.64
2	$6C_2 (0.427)^2 (0.573)^4 = 0.294$	23.5
3	$6C_3 (0.427)^3 (0.573)^3 = 0.29$	23.4
4	$6C_4 (0.427)^4 (0.573)^2 = 0.16$	13.09
5	$6C_5 (0.427)^5 (0.573)^1 = 0.048$	3.9
6	$6C_6 (0.427)^6 (0.573)^0 = 0.00606$	0.48

$\Sigma f_i = 79.81$
$\simeq 80$

NH H_0: The BD is the best fit

$O_i = E_i$

Alt H_1: $O_i \neq E_i$

Test Statistic: $\chi^2 = \Sigma \frac{(O_i - E_i)^2}{E_i}$

$\Sigma f_i = \Sigma O_i$

LOS 5%, Df for BD = n-1

χ^2 table:

O_i	E_i	$O_i - E_i$	$(O_i - E_i)^2$	$\frac{(O_i-E_i)^2}{E_i}$
5, 15 } 20	15.44	4.56	20.79	1.346
28 = 28	23.5	4.5	20.25	0.861
12 = 12	23.4	-11.4	129.9	5.551
~~7~~, 8, 5 } + 20	17.47	2.53	6.40	0.366
				8.124

$\chi^2_{cal} = 8.124$

At 5% LOS, df = 3

$\chi^2_{tab} = 7.82$

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

Ho is rejected

$\Rightarrow$ BD is not the best fit //

2) Fit a Poisson distribution to the foll. data & Test the goodness of fit

x	0	1	2	3	4	5	Total
f	142	156	69	27	5	1	400

Note: df for BD = n-1

df for PD = n-2

df for ND = n-3

Sol: The pmf of PD: $P(X=x) = \frac{e^{-\lambda}\lambda^x}{x!}$, $x = 0, 1, 2 \ldots$

$$\text{Mean} = \lambda = \frac{\Sigma fx}{\Sigma fi} = \frac{0+156+138+81+20+5}{400}$$

$$= \frac{400}{400}$$

$\boxed{\lambda = 1}$

$$P(X=x) = \frac{e^{-1}\lambda^x}{x!} = \frac{e^{-1}(1)^x}{x!} = \frac{e^{-1}}{x!}, \; x = 0, 1, 2 \ldots$$

Fitting PD

x	$P(X=x)$	$N \cdot P(x) = 400 \cdot P(x)$
0	$e^{-1}/0! = 0.3678$	147.12
1	$e^{-1}/1! = 0.3678$	147.12
2	$e^{-1}/2 = 0.183$	73.2
3	$e^{-1}/6 = 0.06$	24
4	$e^{-1}/24 = 0.015$	6
5	$e^{-1}/120 = 0.003$	1

$\Sigma E_i = 398.64$

Scanned by CamScanner

χ^2 table :

O_i	E_i	(O_i-E_i)	$(O_i-E_i)^2$	$\frac{(O_i-E_i)^2}{E_i}$
142	147	-5	25	0.170
156	147	9	81	0.55
69	73	-4	16	0.217
33	31	2	4	0.12
				$\chi^2 = 1.05$

NH H_0: The PD is the best fit

$O_i = E_i$

NH H_1: $O_i \neq E_i$

Test statistic: $\chi^2 = \Sigma \frac{(O_i - E_i)^2}{E_i}$

$\chi^2_{cal} = 1.05$

for PD, $df = n-2$
$= 4-2$
$= 2$

At 5% LOS, 2 df

$\chi^2_{tab} = 5.99$

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

H_0 is accepted

The PD is the best fit //

Scanned by CamScanner

Test 3: Testing Independance Of Attributes:

Let A & B be two charecteristics (attributes) then

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

i.e 2×2 Contingency table

Y \ X	A_1	A_2	Total
B_1	a	b	a+b
B_2	c	d	c+d
Total	a+c	b+d	N = a+b+c+d

$$\rightarrow \chi^2 = \frac{N(ad-bc)^2}{(a+c)(b+d)(a+b)(c+d)}$$

$$\chi^2 = \frac{N(ad-bc)^2}{(a+c)(b+d)(a+b)(c+d)}$$

Expected frequencies

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

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

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

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

1) The following data are collection on two charecters.

Attributes	Smokers	Non-Smokers	Total
Literates	83	57	140
Illiterates	45	68	113
Total	128	125	253

Based on this, can you say that there is a relation b/w smoking and literacy?

Scanned by CamScanner

NH H_0: There is no relation b/w smoking & literacy.

AH H_1: There is a relation b/w smoking & literacy.

LOS: 5%

Df: $(r-1)(s-1)$,

$r=2, S=2$

$\therefore Df = (2-1)(2-1) = 1$

Test Statistic:

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

Expected frequency

$$E(83) = \frac{140 \times 128}{253} = 70.83 \simeq 71$$

$$E(57) = \frac{140 \times 125}{253} = 69.17 \simeq 69$$

$$E(45) = \frac{113 \times 128}{253} = 57.16 \simeq 57$$

$$E(68) = \frac{113 \times 125}{253} = 55.83 \simeq 56$$

O_i	E_i	$(O_i - E_i)$	$(O_i - E_i)^2$	$\frac{(O_i-E_i)^2}{E_i}$
83	71	12	144	2.02
57	69	-12	144	2.08
45	57	-12	144	2.52
68	56	12	144	2.57

$\chi^2 = 9.19$

$\chi^2_{cal} = 9.19$

$\chi^2_{tab} = 3.84$ at 5% LOS 1df

$\chi^2_{cal} > \chi^2_{tab} \Rightarrow H_0$ is rejected

Conclusion: There is no relation b/w smoking & literacy

2nd method (or)

$$\chi^2 = \frac{N(ad-bc)^2}{(a+b)(a+c)(b+c)(b+d)}$$

$$= \frac{253\,((83)(68)-(45)(56))^2}{(140)(113)(128)(125)}$$

$= 947$

2) The company has to choose 3 pension plans and the opinions of the following random sample of 500 employees are showed

Job classification	Pension Plans 1	2	3
Salaried workers	160	140	40
Hourly workers	40	60	60

Determine whether the preference of plan is independent of the classification.

				Total
	160	140	40	340
	40	60	60	160
Total	200	200	100	500

Expected frequency

$E(160) = \frac{340 \times 200}{500} = 136$

$E(140) = \frac{340 \times 200}{500} = 136$

$E(40) = \frac{340 \times 100}{500} = 68$

$E(40) = \frac{160 \times 200}{500} = 64$

$E(60) = \frac{160 \times 200}{500} = 64$

$E(60) = \frac{160 \times 100}{500} = 32$

NH H_0 : There is no preference of plan over job classification

NH H_1 : There is a preference of plan over job classification

Test statistic : $\chi^2 = \sum_i \frac{(O_i - E_i)^2}{E_i}$

LOS 5%

DF $(r-1)(s-1)$

$r = 2,\ s = 3$

$(2-1)(3-1) = 2$

$df = 2$

O_i	E_i	$O_i - E_i$	$(O_i - E_i)^2$	$\frac{(O_i - E_i)^2}{E_i}$
160	136	24	576.	4.23
140	136	4	16	0.11
40	68	-28	784	11.52
40	64	-24	576	9
60	64	-4	16	0.25
60	32	28	784	24.5
				$\chi^2 = 49.61$

$\chi^2_{cal} = 49.61$

At 5% LOS, 1df, $\chi^2_{tab} = 3.43$

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

H_0 is rejected

Large Samples: $(n \geq 30)$

1. Testing for Single Mean
2. Testing difference b/w mean
3. Testing for single proportion
4. Testing for difference b/w two proportions

Test 1 (single mean)

Test statistic:

$$Z = \frac{\bar{x} - \mu}{\sigma / \sqrt{n}}, \quad \begin{array}{l} \bar{x} - \text{sample mean} \\ \mu - \text{population mean} \\ \sigma - \text{population SD} \\ n - \text{sample SD} \end{array}$$

Test 2 (Difference b/w mean)

$$Z = \frac{\bar{x}_1 - \bar{x}_2}{\sigma \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}} \quad (\text{or}) \quad \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{\frac{\sigma^2}{n_1} + \frac{\sigma^2}{n_2}}}, \quad \sigma^2 = \frac{n_1 s_1^2 + n_2 s_2^2}{n_1 + n_2}$$

Critical Values for Large scale (Z-test):

Nature of test	5% LOS	1% LOS
Two tailed test $(\mu_1 \neq \mu_2)$	1.96	2.58
One tailed test (Right tailed $\mu_1 > \mu_2$, Left tailed $\mu_1 < \mu_2$)	1.65	2.33

Promblems on Test 1 (single Mean):

1) A sample of 100 students is taken from a large population with mean height is 160cm. Can it be reasonably regarding that the poupulation the mean height is 165cm with SD = 10cm

Sol $n = 100 > 30$, the given sample is large sample.

NH H_0: $\mu = 165$ cm

AH H_1: $\mu \neq 165$ cm

Test statistic:

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

$\bar{x} = 160$, $\mu = 165$

$\sigma = 10$, $n = 100$

$$= \frac{160 - 165}{10/10}$$

$$z = -5$$

$$|z|_{cal} = 5$$

LOS 5%

$z_{tab} = 1.96$

$z_{cal} > z_{tab} \Rightarrow H_0$ is rejected

$\mu \neq 165$ cm.

2) A random sample of 200 tins of coconut oil, gave an average weight 4.95 kg. SD = 0.21 kg. do you accept the hypothesis of new weight 5kg per tin at 5% LOS

$n = 200 > 30$, the given sample is large

NH H_0: $\mu = 5$ kg

NH H_1: $\mu \neq 5$ kg

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

$$z = \frac{4.95 - 5}{0.21/\sqrt{200}} = \frac{-0.5}{0.21/14.14} = -3.367$$

$|Z|_{cal} = 3.367$

At 5% LOS, $Z_{tab} = 1.96$

$Z_{cal} > Z_{tab}$

Reject H_0

$\mu \neq 5$ kg

Test 2: (problems)

1) Test the significance of difference b/w the mean of samples drawn from two normal population with same SD from following data

	size	mean	SD
Sample I	100	61	4
Sample II	200	63	6

$n_1 = 100$, $n_2 = 200$

$\bar{x}_1 = 61$, $\bar{x}_2 = 63$

$S_1 = 4$, $S_2 = 6$

$$\sigma^2 = \frac{n_1 S_1^2 + n_2 S_2^2}{n_1 + n_2}$$

$$= \frac{(100)(16) + (36)(200)}{300}$$

$$= \frac{1600 + 7200}{300}$$

$$= \frac{8800}{300}$$

$\sigma^2 = 29.33$

$\sigma = 5.41$

NH H_0: $\mu_1 \neq \mu_2$

There is no significant difference b/w mean

AH H_1: $\mu_1 \neq \mu_2$

Test statistic: $Z = \dfrac{\bar{x}_1 - \bar{x}_2}{\sigma\sqrt{\frac{1}{n_1}+\frac{1}{n_2}}}$

$$= \frac{61-63}{5.41\sqrt{\frac{1}{200}+\frac{1}{100}}}$$

$$|Z| = +2.13$$

At 5% LOS, $Z_{tab} = 1.96$

$$Z_{cal} > Z_{tab}$$

H_0 is rejected

$$\mu_1 \neq \mu_2$$

At 1% LOS, $Z_{tab} = 2.58$

$$Z_{cal} < Z_{tab}$$

H_0 is accepted

$$\mu_1 = \mu_2 \;//$$

X

Q) A ~~sample~~ sample of 6400 English men have a mean height 170cm, the SD = 6.4cm. while a simple sample of 1600 Americans has a mean height 172cm, SD = 6.3cm do the data indicate that the Americans are taller than english men

Sol:

$n_1 = 6400$

$n_2 = 1600$

NH H_0: $\mu_1 = \mu_2$

There is no significant difference b/w mean hieght of english men & American men.

Alt H1: $\mu_1 < \mu_2$

Test statistic:

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

$\bar{x}_1 = 170 \quad \bar{x}_2 = 172$

$n_1 = 6400, \quad n_2 = 1600$

$$\sigma^2 = \frac{(6400)(6.4)^2 + (1600)(6.3)^2}{6400 + 1600}$$

$$= \frac{(6400)(40.96) + (1600)(39.69)}{8000}$$

$$= \frac{26214.4 + 63504}{8000}$$

$\sigma^2 = 40.706$

$\sigma = 6.34$

$$Z = \frac{170 - 172}{6.4\sqrt{\frac{1}{6400} + \frac{1}{1600}}} = -11.2$$

$|Z|_{cal} = 11.2$

At 5% LOS, for one tailed test, $Z_{tab} = 1.63$

$Z_{cal} > Z_{tab}$

Reject H_0

Accept H_1

$\mu_1 < \mu_2$

Americans are taller than English men. //

Test 3: Test for Single proportion

Test statistic is

$$Z = \frac{p-P}{\sqrt{\frac{PQ}{n}}}$$

where- p- sample proportion

P - population proportion

$Q = 1-P$, $q = 1-p$

NOTE: The Confidence(or) Fiducal Limits (95%) for single proportion at 5% LOS is

$$P \pm Z_{0.05}\left(\sqrt{\frac{PQ}{n}}\right)$$

99% confidence limits are

$$P \pm Z_{0.01}\sqrt{\frac{PQ}{n}}$$

2. Standard error is the standard deviation of respective test statistic.

Problem: A coin tossed of 400 times and the H appeard 216 times test the hypothesis that the coin is unbiased

NH H_0: The coin is Unbiased

$$P = \frac{1}{2} = 0.5$$

NH H_1: The Coin is biased

$$P \neq 0.5$$

Test Statistic:

$$Z = \frac{p-P}{\sqrt{\frac{PQ}{n}}}$$

$P = 0.5$, $Q = 1-P = 0.5$

Scanned by CamScanner

$$P = \frac{216}{400} = 0.54$$

$$Z = \frac{0.54 - 0.5}{\sqrt{\frac{(0.5)(0.5)}{400}}} = 1.6$$

$$|Z| = 1.6$$

At 5% LOS, $Z_{tab} = 1.96$

$Z_{cal} < Z_{tab}$

H_0 is accepted

The coin is unbiased.

2) A producer confesses that 22% of items manufactured by this company will be defective. To test his claim a random sample of 80 items are selected and known to contain 13 defectives. Test the validity of producers claim at 1% LOS.

NH H_0 : $P = \frac{22}{100} = 0.22$

AH H_1 : $P \neq 0.22$

Test statistic : $p = \frac{13}{80} = 0.16$

$P = 0.22$, $Q = 1 - 0.22$, $n = 80$
$= 0.78$

$$Z = \frac{0.16 - 0.22}{\sqrt{\frac{(0.78)(0.22)}{80}}} = -1.3$$

$$|Z| = 1.3$$

At 1% LOS, $Z_{tab} = 2.33$

$Z_{cal} < Z_{tab} \Rightarrow H_0$ is accepted

$P = 0.22$

Producer's claim is valid for 1% LOS

Test 4: Testing for difference of Mean

Test statistic $Z = \frac{P_1 - P_2}{\sqrt{PQ\left(\frac{1}{n_1}+\frac{1}{n_2}\right)}}$ where

$$P = \frac{n_1 P_1 + n_2 P_2}{n_1 + n_2}, \quad Q = 1 - P$$

(population proportion are not given)

(or)

when population proportions are given, then

$$Z = \frac{P_1 - P_2}{\sqrt{\frac{P_1 Q_1}{n_1} + \frac{P_2 Q_2}{n_2}}}, \quad Q_1 = 1 - P_1, \quad Q_2 = 1 - P_2$$

Q) Random Samples of 400 men and 600 woman where asked whether they would like to have a flyover near there Residence. 200 men and 325 women where in favour of proposal. Test the hypothesis that proportions of men and woman in favour of proposal are Same (i) at 5% LOS (ii) At 1% LOS

$n_1 = 400, \quad n_2 = 600$

NH H_0:

$P_1 = \frac{200}{400} = 0.5 \qquad P_2 = \frac{325}{600} = 0.541$

The men and woman are equal in favour of proposal

$P_1 = P_2$

AH H_1: $P_1 \neq P_2$

Test statistic:

$$P = \frac{(400)(0.5) + (0.541)(600)}{1000}$$

$$= \frac{200 + 326.4}{1000}$$

$$= 0.525$$

Scanned by CamScanner

$P = 0.525 \quad Q = 0.475$

$$Z = \frac{0.5 - 0.541}{\sqrt{(0.24)(0.0025 + 0.0016)}}$$

$$= \frac{-0.04}{0.064}$$

$$= -0.625$$

$$|Z|_{cal} = 0.625$$

At 5% LOS $\quad Z_{0.05} = 1.96, \quad Z_{0.01} = 2.58$

$Z_{cal} < Z_{tab} \Rightarrow H_0$ is accepted

$P_1 = P_2$

Q) In a large city A 20% of random sample of 900 school boys had a slide physical defect in other large city B 18.5% of random Sample of 1600 boys had the same defect. Is the difference b/w the proportions significant.

NH₀ :

$n_1 = 900, \quad n_2 = 1600$

$P_1 = 0.2 \quad P_2 = 0.185$

$$P = \frac{(900)(0.2) + (1600)(0.185)}{2500}$$

$$= \frac{180 + 296}{2500}$$

$P = 0.1904 \approx 0.19$

$Q = 1 - 0.19$

$= 0.81$

$Z = 0.915$, At 5% LOS, $Z_{tab} = 1.96$

$Z_{cal} < Z_{tab}$

H_0 is Accepted //