(1) Test of Significance (or) Test of Hypothesis

→ A statistical hypothesis or simply hypothesis is an assumption about the parameters of the distribution, and sometimes it is also concerns the type and nature of distribution.

→ Hypothesis testing enables us to make statement about population parameters.

→ Many experiments are carried out with the intention of testing a hypothesis.

→ Hypothesis is an assumption or statement which may or may not be true.

→ Hypothesis is a statement about a population parameter. It is an assumption made in order to arrive at a decision regarding population through a sample of population.

eg. 1) The avg. height of soldiers in the army is 165cm

2) The given machine has an effective life of 20years

→ These entire hypothesis may be verified on the basis of certain sample test.

→ Procedures or tests which enable us to decide whether to accept or reject the hypothesis is called test of hypothesis or test of significance.

→ To verify our assumption which is based on sample study we collect data & find out the difference between the sample value & the population value.

→ If there is a difference or the ~~scr~~ difference is very small then our hypothesized value is correct.

→ In general there are two types of hypothesis namely (i) Null hypothesis (H_0)
(ii) Alternative hypothesis(H_1)

□ Null Hypothesis

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

□ Alternative Hypothesis

→ The negation of Null hypothesis is called Alternative hypothesis. It is denoted by H_1.

Degree of freedom (dof)

$$\boxed{V = n-1}$$

→ Degree of freedom is denoted by V and is defined by $\boxed{V = n-1}$. where n = no. of observations

Chi - Square Test (χ^2)

→ Chi - Square test is defined as

$$\chi^2 = \sum_{i=1}^{n} \frac{(O_i - E_i)^2}{E_i}$$

where O_i is observed frequency

E_i is expected frequency

$$\boxed{\chi^2 = \sum_{i=1}^{n} \frac{(O_i - E_i)^2}{E_i}}$$

→ Chi square test is used to check the significance of discrepancy between theory and experiment. It enables us to find the deviation of experiment to theory is just by chance or it is really due to inadequacy of the theory to fit the observed data.

→ Conditions under which Chi Square test is used

1) The total number of observation used in the test must be less than or equal to 30.

2) The test is wholly dependent on degrees of freedom.

3) each of the observations making up the sample for chi square test should be independent of each other

4) The observation collected for chi-square test must be based on method of random sampling.

Q Find the value of Chi-Square for the following data and test for goodness of fit (5% level of significance)

observed frequency	14	15	18	20	15	10
expected frequency	17	10	15	25	10	15

$$\boxed{\chi^2 = \sum_{i=1}^{n} \frac{(O_i - E_i)^2}{E_i}}$$

$$\chi^2 = \frac{(14-17)^2}{17} + \frac{(15-10)^2}{10} + \frac{(18-5)^2}{15} + \frac{(20-25)^2}{25} + \frac{(15-10)^2}{10} + \frac{(10-15)^2}{15}$$

$$\chi^2 = \frac{9}{17} + \frac{25}{10} + \frac{9}{15} + \frac{25}{25} + \frac{25}{10} + \frac{25}{15}$$

$$\chi^2_{Cal} = 8.7960$$

Scanned with CamScanner

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

Since χ^2_{cal} ~~calculated~~ $< \chi^2_{tab}$ $\therefore$ we accept the null hypothesis and the fit is good

(2)

observed frequency	14	18	12	11	15	14
Expected frequency	14	14	14	14	14	14

$$\chi^2 = \sum_{i=1}^{n} \frac{(O_i - E_i)^2}{E_i}$$

$$\chi^2 = \frac{(14-14)^2}{14} + \frac{(18-14)^2}{14} + \frac{(12-14)^2}{14} + \frac{(11-14)^2}{14} + \frac{(15-14)^2}{14}$$

$$+ \frac{(14-14)^2}{14}$$

$$= 0 + \frac{16}{14} + \frac{4}{14} + \frac{9}{14} + \frac{1}{14}$$

$$= \frac{30}{14}$$

$$\chi^2_{cal} = 2.1428$$

$$\chi_{tab} = 11.070$$

$\therefore \chi^2_{cal} < \chi^2_{tab}$ we accept null hypothesis fit is good.

T-Test

When the population of std. deviation is not known and size of the sample is less than or equal to 30 we use t test

→ To test significance of mean sample the test statistics is given by

$$t = \frac{\bar{x} - \mu}{\left(\frac{s}{\sqrt{n}}\right)}$$

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

n = size of sample

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

Q1 → Find the student t test for the follow variable values in a sample of eight −4, −2, −2, 0, 2, 2, 3, 3 using the mean universe to be zero ($\mu = 0$)

$$t = \frac{\bar{x} - \mu}{\left(\frac{s}{\sqrt{n}}\right)}$$

$$\bar{x} = \frac{\Sigma x_i}{n} = \frac{-4 - 2 - 2 + 0 + 2 + 2 + 3 + 3}{8}$$

$$= \frac{2}{8} = 0.25$$

x_i	$x_i - x_0$	$(x_i - x)^2$
-4	-4.25	18.0625
-2	-2.25	5.0625
-2	-2.25	5.0625
0	-0.25	0.0625
2	1.75	3.0625
2	1.75	3.0625
3	2.75	7.5625
3	2.75	7.5625

$\Sigma x = 2$ $\quad$ $\Sigma x - \bar{x} = 0$ $\quad$ $\Sigma (x - \bar{x})^2 = 49.5$

$$S = \sqrt{\frac{49.5}{7}}$$

$$S = 2.6592$$

$$t = \frac{\bar{x} - 0}{\left(\frac{S}{\sqrt{n}}\right)}$$

$$t = \frac{0.25 - 0}{\frac{2.6592}{\sqrt{8}}}$$

$$t_{cal} = 0.2659$$

$$t_{tab} = 1.895$$

$t_{cal} < t_{tab}$ we accept null hypothesis

② A Random sample of 10 boys had IQ
70, 120, 110, 101, 88, 83, 95, 98, 107, 100
Do these data support the assumption of population mean of 160 (5%.)

$$\bar{x} = \frac{\Sigma x}{n} = \frac{972}{10} = 97.2$$

x	$x-\bar{x}$	$(x-\bar{x})^2$
70	-27.2	739.84
120	22.8	519.84
110	12.8	163.84
101	3.8	14.44
88	-9.2	84.64
83	-14.2	201.64
95	-2.2	4.84
98	0.8	0.64
107	9.8	96.04
100	2.8	7.84

$\Sigma x = 972$ $\quad$ $\Sigma x - \bar{x} = 0$ $\quad$ $\Sigma(x-\bar{x})^2 = 1833.6$

$$S = \sqrt{\frac{1833.6}{9}}$$

$$S = 14.2735$$

$$t = \frac{97.2 - 160}{\left(\frac{14.2735}{\sqrt{10}}\right)}$$

$$t_{cal} = -13.91$$

~~tcal~~ as IQ cannot be negetive

$t_{cal} = 13.91$

$t_{tab} = 1.833$

~~too~~

$t_{cal} > t_{tab}$. alternative hypothesis

Data does not support the assumption of population mean of 160. ~~(150)~~

F - Test

$$F = \frac{S_1^2}{S_2^2} \quad (S_1^2 > S_2^2)$$

$$V_1 = n_1 - 1$$

$$V_2 = n_2 - 1$$

$$F = \frac{S_2^2}{S_1^2} \quad (S_2^2 > S_1^2)$$

$$S_1^2 = \frac{\Sigma(x-\bar{x})^2}{n_1 - 1} \qquad S_2^2 = \frac{\Sigma(y-\bar{y})^2}{n_2 - 1}$$

→ The objective of F test is to find whether two independent estimates of population variance differs significantly or whether the two samples may be regarded as drawn from normal population having the same variance.

→

Q → Two samples are drawn from two normal populations, test whether the two samples have the same variance at 5% level of significance

Sample 1	60	65	71	74	76	82	85	87		
Sample 2	61	66	67	85	78	63	85	86	88	91

$$\bar{x} = \frac{\sum x}{n_1} = \frac{600}{8} = 75$$

$$\bar{y} = \frac{\sum y}{n_2} = \frac{770}{10} = 77$$

$n_1 = 8$

$n_2 = 10$

x	$x-\bar{x}$	$(x-\bar{x})^2$	y	$(y-\bar{y})$	$(y-\bar{y})^2$
60	-15	225	61	-16	256
65	-10	100	66	-11	121
71	-4	16	67	-10	100
74	-1	1	85	8	64
76	1	1	78	1	1
82	7	49	63	-14	196
85	10	100	85	8	64
87	12	144	86	9	81
			88	11	121
			91	14	196
$\sum x = 600$	$\sum x-\bar{x} = 0$	$\sum (x-\bar{x})^2 = 636$	$\sum y = 770$	$\sum y-\bar{y} = 0$	$\sum (y-\bar{y})^2 = 1200$

$$S_1^2 = \frac{\sum (x-\bar{x})^2}{n_1-1} \Rightarrow \frac{636}{7} = 90.8571$$

$$S_2^2 = \frac{\sum (y-\bar{y})^2}{n_2-1} = \frac{1200}{9} = 133.33$$

$$F_{cal} = \frac{S_2^2}{S_1^2} = \frac{133.33}{90.85} = 1.4675 \quad F_{tab} = 3.68$$

Scanned with CamScanner

$\therefore$ $F_{cal} < F_{tab}$ Ho is accepted

the two samples have same variance at 5% of significance.

Q → Two random samples drawn from two normal population having the variable values as below

Sample 1	19	17	16	28	22	23	19	24	26			
Sample 2	28	32	40	37	30	35	40	28	41	45	30	36

x	$x-\bar{x}$	$(x-\bar{x})^2$	y	$y-\bar{y}$	$(y-\bar{y})^2$
19	-2.5555	6.5305	28	-7.1666	51.3601
17	-4.5555	20.7525	32	-3.1666	10.0273
16	-5.5555	30.8635	40	4.8334	23.3617
28	6.4445	41.5315	37	1.8334	3.3613
22	0.4445	0.1975	30	-5.1666	26.6937
23	1.4445	2.0865	35	-0.1666	0.0277
19	-2.5555	6.5305	40	4.8334	23.3617
24	2.4445	5.9755	28	-7.1666	51.3601
26	4.4445	19.7535	41	5.8334	34.0285
			45	9.8334	96.6957
			30	-5.1666	26.6937
			36	0.8334	0.6945

$\Sigma(x-\bar{x})^2 = 134.2215$

$\Sigma(y-\bar{y})^2 = 347.666$

$$\bar{x} = \frac{\Sigma x}{9} = \frac{194}{9} = 21.5555$$

$$\bar{y} = \frac{\Sigma y}{12} = \frac{422}{12} = 35.1666$$

$$S_1^2 = \frac{\Sigma(x-\bar{x})^2}{n_1-1} = \frac{134.2215}{8} = 16.7776$$

$$S_2^2 = \frac{\Sigma(y-\bar{y})^2}{n_2-1} = \frac{347.66}{11} = 31.6054$$

$$F = \frac{S_2^2}{S_1^2} = \frac{31.6054}{16.7776} = \underline{\underline{1.8837}}$$

$F_{cal} = 1.8837$

$F_{tab} = 3.35$

$\therefore$ $F_{cal} < F_{tab}$ H_0 is accepted

March - April 2022

Q Two samples of size 9 and 8 give the sum of squares of deviation from their respective means equal to 1160 and 91 respectively. Can they be regarded as drawn from two normal populations with same variance.

$\cancel{S_1^2 = \frac{116}{8} = 14.5}$

$\cancel{S_2^2 = \frac{91}{7} = 13}$

$\cancel{F = \frac{S_1^2}{S_2^2} = 1.1153}$

$S_1^2 = 1160$

$S_2^2 = 91$

$F_{cal} = \frac{S_1^2}{S_2^2} = 1.27582$

$F_{tab} = 3.73$

$F_{cal} < F_{tab}$ null hypothesis is accepted

two normal populations with same variance

March-April 21

Q → Two random samples of sizes of 9 and 7 give the sum of the squares of deviation from their respective means 175 and 95 respectively. Can they be regarded as drawn from normal population with same variance

$S_1^2 = 175$ $n_1 = 9$

$S_2^2 = 95$ $n_2 = \cancel{8}7$

$F_{cal} = \frac{S_1^2}{S_2^2} = 1.8421$

$F_{tab} = 4.15$

$F_{cal} < F_{tab}$ H_0 is accepted

two normal populations with same variance

Large Samples Test of Significance (Z-test)

→ If the sample is large ie ≥ 30 then we consider such samples as large sample

→ The test of significance used in large sample is diffrent from those used in small samples because small samples fail to satisfy the assumption under which large sample analysis is done.

→ Under the large sample test we will see four important test of significance

(i) Testing of significance for single proportion

(ii) Testing of significance for difference of proporti

(iii) Testing of significance for single mean

(iv) Testing of significance for diffrence of mean and difference of std deviation.

One tailed Test :- If we have to test whether the population mean (μ) has specified value (μ_0) then null hypothesis is

$$H_0 : \mu = \mu_0$$

→ The alternative hypothesis may be

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

ie (i) $\mu > \mu_0$

(ii) $\mu < \mu_0$ is known as two tailed

(right & left tailed)

→ $\mu > \mu_0$ [right tailed test]

→ $\mu < \mu_0$ (left tailed test)

Errors of Sampling

① Type I Error

→ If the null hypothesis (H_0) is true but it is rejected by test procedure then the error made is called type I error or α error.

② Type II error

→ If the null hypothesis (H_0) is false but it is accepted by test procedure then the error commited is called Type II error beta error.

08/11/2022

Test of significance of single mean

→ Let a random sample of size $n \geq 30$ has the sample mean $\bar{x}$ and μ be the population mean, also the population mean μ has a specified value μ_0

★ **Working Rule**

1. H_0
2. H_1
3. level of significan
4. Test statistic (formula) $Z = \dfrac{\bar{x} - \mu}{\frac{\sigma}{\sqrt{n}}}$

Q A sample of 64 students have an mean weight 70 kgs. Can this be regarded as sample from a population with mean weigh 56 kgs and standard deviation 25 kgs.

Sol given $\bar{x} = 70$ kgs

$\mu = 56$ kgs.

$\sigma = 25$ kgs

$n = 64$

(i) H_0: A sample of 64 Students have a mean weight 70 kgs. Can this be regarded as sample from a propulation with mean weight 56 kgs & std. deviation 25 kgs

(ii) H_1 : Sample cannot be regarded

(iii) Level of significance : 0.05 (assumption)

(iv) Test statistic $= Z = \dfrac{\bar{x} - \mu}{\frac{\sigma}{\sqrt{n}}} = \dfrac{70 - 56}{\frac{25}{\sqrt{64}}} = 4.48$

$\therefore$ The null hypothesis is rejected since $|Z| > 1.645$

Q In a random sample of 60 workers, the avg. time taken by them to get to work is 33.8 mins with a standard deviation 6.1 mins. Can we ~~ge~~ reject the null hypothesis, $\mu = 32.6$ mins in favor of alternative hypothesis $\mu > 32.6$ mins (test at 1% level of significance.

Sol given $\bar{x} = 33.8$ mins

$\mu = 32.6$ mins

$\sigma = 6.1$ mins

$n = 60$

① H_0 : ~~Sample of 60 workers~~ $\mu = 32.6$

② $H_1 : \mu > 32.6$

③ level of significance $= 0.01$

④ test statistic $= Z = \dfrac{\bar{x} - \mu}{\frac{\sigma}{\sqrt{n}}} = \dfrac{33.8 - 32.6}{\frac{6.1}{\sqrt{60}}}$

$= 1.5237$

Test of significance of difference of mean.

Let $\bar{x}_1$ & $\bar{x}_2$ be the sample means of two independent large random samples of sizes n_1 & n_2 drawn from two populations having means μ_1, μ_2 and S.Ds σ_1 & σ_2

→ To test whether the two population means are equal

1. Let Null Hypothesis be $H_0 : \mu_1 = \mu_2$
2. The Alternative hypothesis is $H_1 : \mu_1 \neq \mu_2$
3. Level of significance.
4. Test statistic is $Z = \dfrac{\bar{x}_1 - \bar{x}_2}{\sqrt{\dfrac{\sigma_1^2}{n_1} + \dfrac{\sigma_2^2}{n_2}}}$

Rejection Rule for $H_0 : \mu_1 = \mu_2$

1) A sample of height of 6400 englishmen has an mean of 67.85 inches and a S.D of 2.56 inches with a sample of heights of 1600 Australians has a mean of 68.55 inches and S.D of 2.52 inches. Do the data indicate the Australians are on avg. taller than englishmen ? (use α as 0.05)

Sol given

n_1 = size of 1st sample = 6400

n_2 = size of 2nd sample = 1600

$\bar{x}_1$ = mean of 1st sample = 67.85

$\bar{x}_2$ = mean of 2nd sample = 65.88

σ_1 = S.D of 1st sample = 2.56

σ_2 = S.D of 2nd sample = 2.52

1) Null Hypothesis be $H_0 : \mu_1 > \mu_2$

2) Alternative hypothesis $H_1 : \mu_1 < \mu_2$

3) Level of significance = 0.05

1.97

0.070

4) $$z = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}} = \frac{67.85 - 68.55}{\sqrt{\frac{(2.56)^2}{6400} + \frac{(2.52)^2}{1600}}}$$

Q → The mean life of a sample of 10 elect bulbs was found to be 1456 hrs with S.D of 423 hrs A second sample of 17 bulbs chosen from a diffrent batch showed a mean life of 1280 hrs with SD of 398 hrs Is there a significant diffrence between the means of two batches.

Sol $n_1 = 10$

$n_2 = 17$

$\bar{x}_1 = 1456$

$\bar{x}_2 = 1280$

$\sigma_1 = 423$

$\sigma_2 = 398$

1) Null hypothesis be $H_0: \mu_1 = \mu_2$

2) Alternative hypothesis be $H_1 = \mu_1 \neq \mu_2$

3) Level of significance $= 0.05$

4) $$|Z| = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}} = \frac{1456 - 1280}{\sqrt{\frac{423^2}{10} + \frac{398^2}{17}}} = 1.067$$

$|Z| = 1.067$

Since $|Z| = 1.067 < 1.96$

Hence we accept H_0 at 5% level of signifi & conclude that there is no diff. b/w the mean life of electric bulbs of two batches

3) A researcher wants to know the intelligence of students in a college. He selected two groups of students. In the first group there are 150 students having a mean IQ of 75 with S.D of 15. in the second group of 250 students having mean IQ of 70 with S.D of 20 (use α as 0.01)

$n_1 = 150$

$n_2 = 250$

$\bar{x}_1 = 75$

$\bar{x}_2 = 70$

$\sigma_1 = 15$

$\sigma_2 = 20$

1) $H_0 : \mu_1 = \mu_2$

2) $H_1 : \mu_1 \neq \mu_2$

3) Level of significance $= 0.01$

4) $$|Z| = \frac{75 - 70}{\sqrt{\frac{15^2}{150} + \frac{20^2}{250}}} = 2.839$$

$|Z| = 2.839 > 2.58$ we reject H_0 & conclude t gp has been not taken from same populati

Test of Significance of single proportion

large Sample.

→ Suppose a large random sample of size n has a sample proportion p of members possessing a certain attribute (ie proportion of success)

→ To test the hypothesis that the proportion P in the population has a specific value P_0.

1) H_0

2) H_1

3) Level of significance

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

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

Sol given sample size = 200

No. of pieces confirming to specification = 200 - 18

= 182

p = proportion of pieces confirming to specification

$= \frac{182}{200} = 0.91$

P = population proportion $= \frac{95}{100} = 0.95$

$Q = 1 - 0.95$

$Q = 0.05$

(i) $H_0 : P = 95\%$

(ii) $H_1 : P < 95\%$

(3) $|Z| = \frac{p-P}{\sqrt{\frac{PQ}{n}}} = \frac{0.91-0.95}{\sqrt{\frac{0.95 \times 0.05}{200}}} = -2.59$

$|Z| = 2.59 > 1.96$ hence H_0 is rejected & manufacturers claim is rejected

Q → In a sample of 1000 people in Karna. 540 are rice eaters & rest are wheat eaters. Can we assume both Rice & wheat eaters are equally popular in the State at 10% of level of significance

$n = 1000$

$\cancel{p = \frac{1000 - 540}{1000}} = \frac{540}{1000} = 0.54$

$P = 0.5$

$Q = 1 - 0.5 = 0.5$

$$|Z| = \frac{0.54 - 0.5}{\sqrt{\frac{0.5 \times 0.5}{1000}}} = 2.529$$

Since $|Z| = 2.529 < 2.58$ we accept H_0 & conclude that both rice & wheat are equally popular in the state.

10/11/2022

Test of ~~sig~~ significance of difference of proportions.

→ Let P_1 & P_2 be the sample proportions in 2 large samples of sizes n_1 & n_2 drawn from two populations having proportion P_1 & P_2

→ Test statistic (formula), $Z = \dfrac{P_1 - P_2}{\sqrt{pq\left(\frac{1}{n_1}+\frac{1}{n_2}\right)}}$

$$\text{where } P = \frac{n_1 P_1 + n_2 P_2}{n_1 + n_2}$$

$$p + q = 1$$

$$q = 1 - p$$

Q A random sample of 400 men & 600 women were asked whether they would like to have a flyover near their residence. 200 men & 325 women were in favor of proposal. Test the hypothesis that proportion of men & women in favor of proposal are same at 5% level of significance.

Sol (i) $H_0 : P_1 = P_2$ assume that there is no significant difference b/w men & women as far as proposal of flyover is concerned

(ii) $H_1 : P_1 \neq P_2$

(iii) level of significance $= 0.05$

given

$n_1 = 400$

$n_2 = 600$

$p_1 = \frac{200}{400} = 0.5$

$p_2 = \frac{325}{600} = 0.5416$

$$P = \frac{n_1 p_1 + n_2 p_2}{n_1 + n_2} = \frac{400(0.5) + (600)(0.5416)}{1000}$$

$$= 0.52496$$

$$q = 1 - P = 1 - 0.52496 = 0.47504$$

$$Z = \frac{p_1 - p_2}{\sqrt{Pq\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}} = \frac{0.5 - 0.5416}{(0.52496)(0.47504)\left[\frac{1}{400} + \frac{1}{600}\right]}$$

$$\frac{0.5 - 0.5416}{\sqrt{(0.52496)(0.47504)(0.0041)}}$$

$= -1.271$

$|Z| = 1.271 < 1.96$

we accept H_0 and conclude that there is no significant diffrence between the option of men & women as far as proposal of flyover is concerned.

Q → A manufacturer of electronics equipment subjects samples of 2 brands of transistors to an accelerated performance test if 45 of 180 transistors of the first kind & 34 of 120 of 2^{nd} kind fail the test what can we conclude at the level of significance of 5% about the diffrence b/w corresponding sample proportion.

Sol

$n_1 = 180$

$n_2 = 120$

$p_1 = \frac{45}{180} = 0.25$

$p_2 = 0.28333$

(i) $H_0 : P_1 = P_2$

(ii) $H_1 : P_1 \neq P_2$

(iii) level of significance = 5%.

Scanned with CamScanner

$$P = \frac{P_1 n_1 + P_2 n_2}{n_1 + n_2} = \frac{(0.25)(180) + (0.283)(120)}{180 + 120}$$

$$P = 0.2632$$

$$q = 1 - P = 0.7368$$

$$z = \frac{P_1 \mp P_2}{\sqrt{Pq\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}}$$

$$z = -0.6358$$

$$|z| = 0.6358 < 1.96$$

we accept H_0