Parametric tests

S.No	Type of Test		Objectives	Formulae
1	Z test	a	To find the significance of difference between mean of population and mean of sample when the sample size is large [> 30 units]	$Z = \frac{\bar{x} - \mu}{\sigma / \sqrt{n}}$
		b	To find the significance of difference between the means of two large samples	$Z = \frac{\overline{x_1} - \overline{x_2}}{\sqrt{\frac{\sigma x_1^2}{n_1} + \frac{\sigma x_2^2}{n_2}}}$
		c	To find the significance of difference between population proportion and sample proportion	$Z = \frac{\hat{p} - p}{\sqrt{\frac{p.q}{n}}}$
		d	To find the significance of difference between two sample proportions [In case of percentages, they are converted into proportion. *eg.* 60% as 0.6]	$Z = \frac{\hat{p}_1 - \hat{p}_2}{\sqrt{\frac{\hat{p}_1 . \hat{q}_1}{n_1} + \frac{\hat{p}_2 . \hat{q}_2}{n_2}}}$
2	**'t' – test**	a	To find the significance of difference	$t = \frac{\bar{x} - \mu}{s / \sqrt{n}}$

			between mean of population and mean of sample when the same size is small [< 30 units] and population variance is not known	
		b	To find the significance of difference between means of two small samples[or atleast one is a small sample]	$$t=\frac{\bar{x}-\bar{y}}{s\sqrt{\frac{1}{n_1}+\frac{1}{n_2}}}$$ Where $$s=\sqrt{\frac{(n_1-1)s_x^2+(n_2-1)s_y^2}{n_1+n_2-2}}$$
3	**Chi-Square test as a parametric test**		To find out the significance of difference between population variance and sample variance	$$\chi^2=\frac{s^2(n-1)}{\sigma^2}$$

Types of commonly employed parametric tests, their objectives and the formulae

Notes :

N = Size of population [No. of units in the population]
n = Size of sample [No. of units in the sample]
n_1 = Size of sample 1
n_2 = Size of sample 2

μ = Mean of population
$\bar{x}$ = Mean of sample
$\bar{x}_1$ = Mean of sample 1
$\bar{x}_2$ = Mean of sample 2
σ^2 = Variance of population
σ = Standard Deviation of population
s^2 = Variance of sample
s= Standard Deviation of sample or combined S.D s of two samples or S.D of differences
s_1^2 = Variance of sample 1
s_2^2 = Variance of sample 2
s_1 = Standard Deviation of sample 1
s_2 = Standard Deviation of sample 2
d = Difference of value in the same sample unit [before and after treatment]
$\bar{d}$ = Mean of differences
p = Proportion in population
$q = 1 - p$
$q_1 = 1 - p_1$ ($p_1 =$ proportion in population 1)
$q_2 = 1 - p_2$ ($p_2 =$ proportion in population 2)
$\hat{p}$ = Proportion in sample [$\hat{p}$ is read as p hat]

$\hat{q} = 1 - \hat{p}$ [$\hat{q}$ is read as q hat]
$\hat{p}_1$ = Proportion in sample 1
$\hat{p}_2$ = Proportion in sample 2
$\hat{q}_1 = 1 - \hat{p}_1$
$\hat{q}_2 = 1 - \hat{p}_2$
r = Correlation coefficient

Z – test

Z- test, a parametric test, is employed to find out the significance of difference between

i. sample mean and population mean
ii. two sample means
iii. proportion in sample and proportion in population and
iv. two sample proportions.

Z test is used when the sample size is large [≥30 subjects) and population variable is known. If population variance is not given, the sample variance can be taken as population variance as the sample is large. In case the sample size is small (≤ 30) but normal population variance is known, Z test can be used. To find out the critical Z value (table value) Z-distribution is used. As finding out the table Z value is a little time consuming the table Z-values for two-tailed test and one-tailed tests for commonly used levels of significance are presented in the table 9.2.

Type of Test	Significance level				
	10%	5%	2.5%	2%	1%
Two – tailed	1.65	1.96	2.24	2.33	2.58
One-tailed [left]	-1.28	-1.65	-1.96	-2.05	-2.33
One –tailed [right]	+ 1.28	+ 1.65	+1.96	+2.05	+2.33

Critical (table) Z values

To test the significance of difference between population mean and sample mean [when sample size is large and population variance is known]

Example 1: A particular variety of wheat plants has shown a mean height of 82.63cm and standard deviation of 3.89cm. From this population of plants, 50 plants are selected at random and each plant is inoculated with a chemical which claims to increase the height of plants. After inoculation the mean of the height of 50 plants is found to be 83.66cm. On the basis of this evidence can it be now concluded that the chemical has a beneficial effect on the growth of plants at 5% level of significance?

There is a numerical difference of 1.03cm (83.66 – 82.63cm) in the means. Now the question is, “is the difference significant?” $\left(\mu = \bar{x}\right)$

Solution

Null hypothesis (Ho): The chemical has no effect on the growth of wheat plants. That is, the mean height of the sample 83.66cm does not differ significantly from the population mean of 82.63cm. The increase in height of 1.03cm is only due to chance factor and not due to the chemical.

Alternative Hypothesis (Ha): The chemical has a positive effect on the growth of wheat plants and the mean height of the sample 83.66cm significantly differs from the population mean of 82.63cm. The increase in height of 1.03cm is due to the effect of chemical and not due to chance factor.

Given Data: μ = 82.63cm, $\bar{x}$ = 83.66cm, n = 50, S.D = 3.89

Test criterion: Here, the sample mean is to be compared with population mean. As S.D of population is known and the sample size is large (>30), Z test is used.

Calculated Z value:

$$Z = \frac{\bar{x} - \mu}{\sigma / \sqrt{n}} \quad = \frac{83.66 - 82.63}{3.89 / \sqrt{50}} = 1.87$$

Table value : Table Z value at 5% level of significance is 1.96 [As finding the table Z value at different levels of significance is a little difficult it is better to remember the table Z values for the commonly used levels of significance. Table Z value at 10% level is 1.65; at 5% level 1.96 and at 1% level 2.58.

Interpretation: As the calculated Z value (1.87) is less than the table Z value of 1.96, the null hypothesis is accepted.

Inference: Though there is an increase in the height of wheat plants consequent to the application of the chemical, the increase is not statistically significant. That is, the observed increase of 1.03 cm is mostly due to chance factor and not entirely due to the chemical. Therefore, it is inferred that the chemical has no effect to increase the height of wheat plants.

Example 2: The overweight ladies taking brisk walks in the Race Course in Coimbatore, had a mean weight of 87kg two years ago. Now, a sample of 40 ladies is taken at random and the mean weight is found to be 72kg. Variance is found to be 25kg. Test at 1% level of significance whether brisk walk is effective in reducing weight.

Solution

Null hypothesis (Ho): Brisk walk has no effect in reducing weight $\left(\mu = \bar{x}\right)$ where μ is the mean weight of ladies 2 years ago and $\bar{x}$ is the mean weight of 40 ladies after 2 years of walking.

Alternative hypothesis (Ha): Brisk walk reduces body weight:

Given data: μ= 87 kg. $\bar{x}$: 72 kg, σ^2 = 25kg, S.D = $\sqrt{\sigma^2} = \sqrt{25} = 5$; n = 40

Test criterion: Here, the sample mean is to be compared with the population mean. As S.D of population is known and the sample size is large (>30), Z test is used

Calculated Z value:

$$. Z = \frac{\mu - \bar{x}}{\sigma / \sqrt{n}} = \frac{87 - 72}{5/\sqrt{40}} = \frac{15}{0.79} = 18.97$$

[In test values the sign is ignored]

Table value: The table Z value at 1% level of significance is 2.58

Interpretation: As the calculated Z value of 18.97 is greater than the table value of 2.58, the null hypothesis is rejected.

Inference: The mean reduction in weight of 15kg consequent to brisk walk is not a chance factor. It is inferred that continuous brisk walk helps in reducing body weight.

To test the significance of difference between means of two large samples.

Example 1: A social worker desires to find out whether there is any significant difference in daily wages between the plantation workers in Valparai [plantation Zone in Coimbatore dt. of Tamilnadu) and Idukki in Kerala. The data collected are given below:

Plantation Zone	Sample size	Sample mean (Rs)	Population variance (Rs)
Valparai	80	82	125
Idukki	70	85	136

Differences in daily wages in plantations (Example)

Test the difference at 5% level of significance.

Solution

Null Hypothesis (Ho): There is no significant difference in daily wages between the workers in Valparai and that of Idukki *i.e* the wages are equal ($\mu_1 = \mu_2$). As the samples are large (>30), it is assumed $\mu_1 = \overline{x_1}$ and $\mu_2 = \overline{x_2}$.

Alternative hypothesis (Ha): There is difference in daily wages *i.e.* they are not equal ($\mu_1 \neq \mu_2$).

Given data : $\bar{x}_1 = 82$, $\bar{x}_2 = 85$, $n_1 = 80$ $n_2 = 70$, $\sigma_1^2 = 125$, $\sigma_2^2 = 136$

Test criterion : Here, two sample means are compared for significance of difference. As the standard deviations of the populations are known and the samples are large, Z test is used.

$$Z = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{\frac{\sigma x_1^2}{n_1} + \frac{\sigma x_2^2}{n_2}}} = \frac{82 - 85}{\sqrt{\frac{125}{80} + \frac{136}{70}}} = \frac{-3}{\sqrt{1.56 + 1.94}} = \frac{-3}{1.87} = -1.60 = 1.60 \text{ (sign is ignored)}$$

Table Z value at 5% level of significance is 1.96

Interpretation: As the calculated Z value of 1.60 is less than the table value of 1.96, the N.H. is accepted.

Inference: Though the average daily wage of Idukki workers is Rs.85 and that of Valparai workers is Rs.82, statistically there is no significant difference. The numerical difference is only a chance factor.

Example 2: Intelligence test given to one group of girls and another group of boys showed the following results :

Gender	Number of Students Tested	Standard Deviation	Mean Intelligence Score
Girls	50	10	75
Boys	100	12	70

Scores of intelligence tests of students (example)

Is the difference in the mean scores statistically different at 1% level of significance?

Solution

N.H : There is no difference in the intelligence test scores of girls and boys *i.e.* they are equal ($\mu_1 = \mu_2$)

A.H : There is difference *i.e* they are not equal ($\mu_1 \neq \mu_2$)

Given data : $\bar{x}_1 = 75$ $\bar{x}_2 = 70$; $n_1 = 50$ $n_2 = 100$, $\sigma_1 = 10$, $\sigma_2 = 12$.

Test criterion : Here, two sample means are compared for significance of difference. As the S.D s of the populations are known and the samples are large in size (>30)

Z test is used.

$$Z = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{\frac{\sigma x_1^2}{n_1} + \frac{\sigma x_2^2}{n_2}}} = \frac{75 - 70}{\sqrt{\frac{(10)^2}{50} + \frac{(12)^2}{100}}} = \frac{5}{\sqrt{3.44}} = \frac{5}{1.855} = 2.695$$

Table value : The table Z value at 1% level of significance is 2.58

Interpretation : As the calculated Z value (2.695) is greater than the table Z value of 2.58, the N.H is rejected and alternative hypothesis is accepted

Inference : There is significant difference between the intelligence scores of girls and boys. The girls, in the particular test, appear to be more intelligent than boys.

To test the significance of difference between population proportion and sample proportion:

Example 1: The public complained that the VIP free passes at Palani temple rope – way

is 20%, but the officials contested that this figure was high and conducted a survey to verify the generalization. In the survey, they found that out of 500 persons checked, 85 were holding VIP passes. Test at 5% level of significance, whether the officials are right in claiming that 20% VIP pass traveling is on the higher side.

Solution

N.H : There is no difference in the population proportion (VIP pass traveling)

ie H_o: p = 20% or 0.2

A.H : There is difference in the population proportion (Ha : $p \neq 0.20$).

Given data : p = 0.20 , [therefore q = (1 – p) = 0.8] , n = 500;

$\hat{p}$ (sample proportion) is $\frac{85}{500} = 0.17$

Test Criterion : Here, as the population proportion is to be compared with the sample proportion Z test is used

Calculated Z value :

$$Z = \frac{\hat{p} - p}{\sqrt{\frac{p.q}{n}}} = \frac{0.17 - 0.20}{\sqrt{\frac{0.2x0.8}{500}}} = \frac{-0.03}{0.0179} = 1.68$$

Table Z value at 5% level of significance is 1.96

Interpretation : As the calculated Z value of 1.68 is less than the table Z value of 1.96 at 5% level of significance, the N.H is accepted.

Inference: Though the officials reported a proportion of 0.17 (85 / 500) of VIP pass traveling, it is not significantly different from the generalized proportion of 0.20 of VIP pass traveling. The difference observed was due to chance factor.

Example 2: A coin is tossed 100 times under identical conditions independently yielding 30 heads and 70 tails. Test at 1% level of significance whether or not the coin is unbiased.

Solution

N.H : The coin is unbiased (i.e proportions of heads and tails while tossing will be equal i.e p = q = $\frac{50}{100} = 0.5$

A.H : The coin is biased *ie* p≠0.5

Given Data : p = 0.5 ; $\hat{p} = \frac{30}{100} = 0.3; n = 100$

Test Criterion : Here, as the population proportion is to be compared with sample proportion, Z test is used

Calculated Z Value :

$$Z = \frac{\hat{p} - p}{\sqrt{\frac{p.q}{n}}} = \frac{0.3 - 0.5}{\sqrt{\frac{0.5 \, x \, 0.5}{100}}} = -\frac{0.2}{0.05} = -4 = 4$$

Interpretation : As the calculated Z value of 4 is greater than the table Z value of 2.58 at 1% level, the N.H is rejected and A.H is accepted.

Inference : If the coin is not biased head should turn up about 50% of times tossed (a proportion of 0.5). In the trial head turned only 30 times out of 100 times tossed (*i.e* a proportion 0.3) The difference in proportions (0.5 and 0.3) is significantly different . That is, the coin is biased against the head.

To test the significance of difference between two sample proportions.

Example 1: In a study it is found out that 252 among 450 girl students go in for latest dress other than churidhars and 242 among 620 boys go after ultra – modern dress. Find out at 1% level, who is more inclined towards modern dress, the boys or the girls?

Solution

N.H : There is no difference between girls and boys in going in for modern dress *i.e.* the groups are equally inclined towards modern dress $p_1=p_2$ ($\hat{p}_1=\hat{p}_2$).

A.H : There is difference *i.e.* $p_1 \neq p_2$ ($\hat{p}_1 \neq \hat{p}_2$).

Given Data : $\hat{p}_1=\frac{252}{450}=0.56$ $\hat{p}_2=\frac{242}{620}=0.39$

n_1(sample size of girls) = 450 ; n_2 (sample size of boys) = 620:

$\hat{p}_1$ = Proportion of girls in the sample going in for latest dress

$\hat{p}_2$ =Proportion of boys in the sample going in for ultra –modern dress

Test criterion : As the proportions of two samples are to be compared Z test is used

Calculated Z value :

$$Z=\frac{\hat{p}_1-\hat{p}_2}{\sqrt{\frac{\hat{p}_1 . \hat{q}_1}{n_1}+\frac{\hat{p}_2 . \hat{q}_2}{n_2}}}=\frac{0.56-0.39}{\sqrt{\frac{(0.56)(0.44)}{450}+\frac{(0.39)(0.61)}{620}}}=\frac{0.17}{\sqrt{0.00055+0.00038}}=\frac{0.17}{0.03}=5.67$$

Interpretation : As the calculated Z value of 5. 67 is greater than the table value of 2.58 at 1% level, the N.H is rejected and A.H is accepted.

Inference : There is significance difference between girls and boys in going in for modern dress. Girls appear to be more inclined towards modern dress than boys.

Example 2: 500 units from a factory are inspected and 12 are found to be defective and out of 800 units in another factory the same no. (12) of units is found to be defective. Can it be concluded at 5% level of significance that production at the second factory is better than the first factory?

Solution

N.H: There is no difference in the proportion of defective units between the first and second factory i. e $p_1 = p_2 = 0.5$

A.H : There is difference ($p_1 \neq p_2$)

Given data:

$\hat{p}_1$ = (Proportion of defects in the sample in the 1st factory)=12 / 500 = 0.024

$\hat{p}_2$ = (Proportion of defects in the sample in the 2nd factory)=12/ 800 = 0.015

n_1 (sample size of the 1st factory)= 500 and n_2(sample size of the 2nd factory) = 800

Test criterion : As two sample proportions are to be compared here Z test is used

Calculated Z value :

$$Z = \frac{\hat{p}_1 - \hat{p}_2}{\sqrt{\frac{\hat{p}_1 \cdot \hat{q}_1}{n_1} + \frac{\hat{p}_2 \cdot \hat{q}_2}{n_2}}} = \frac{0.024 - 0.015}{\sqrt{\frac{(0.024)(0.976)}{500} + \frac{(0.015)(0.985)}{800}}} = \frac{0.009}{\sqrt{0.0000468 + 0.0000184}}$$

$$\frac{0.009}{0.0080746} = 1.12$$

Interpretation : As the calculated Z value of 1.12 is less than the table Z value of 1.96 at 5% level, the N.H is accepted.

Inference : There is no significant difference in the proportions of defective items between the two factories, though it appears that proportion of defects is less in second factory.

‘t’ test

Student’s t [‘t’] test, a parametric test, is employed to find out the significance of difference between

(i) sample mean and population means and (ii) two sample means when the sample size is small (<30 subjects) and population variance is not known. To find out the critical (table) t value “t” distribution is used. Tables “t” values are given in all standard statistics books both for two – tailed and one-tailed tests for commonly used levels of significance.

To test the significance of difference between population mean and sample mean when sample size is small and population variance is not known.

Example 1: Prices of shares (in rupees) of a company on different days in a month were found to be 66, 65,69,70,69, 71, 70,63, 64 and 68. Test the mean price of the shares in the month is 65.

Solution

N.H : The monthly mean price of the shares (Rs. 65) does not differ from the 10 day average *i.e.* $\mu = \bar{x}$

A.H : There is difference $\mu \neq \bar{x}$

Given data : μ = Rs. 65; $\bar{x} = \dfrac{66+65+69+70+69+71+70+63+64+68}{10} = 67.5$

n = 10

Test Criterion : As population mean is to be compared with sample mean and the sample happens to be small(<30) t test is used

Calculated t value : $t = \dfrac{\bar{x} - \mu}{s/\sqrt{n}} = \dfrac{67.5 - 65}{2.8/\sqrt{10}} = 2.81$

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

$$= \sqrt{\frac{(66-67.5)^2 + (65-67.5)^2 + (69-67.5)^2 + (70-67.5)^2 + (69-67.5)^2 + (71-67.5)^2 + (70-67.5)^2 + (63-67.5)^2 + (64-67.5)^2 + (68-67.5)^2}{10-1}}$$

= 2.8

Table value : To find out table 't' values two aspects are to be confirmed

(i). Degrees of freedom (df) : In the case of single sample, the df = (n-1) and in the case of two samples, df = $(n_1 + n_2) - 2$

(ii). One-tailed or two-tailed: Whether one has to look for the 't'–value under one – tailed or two-tailed depends on the alternative hypothesis

- If the A.H is $\mu_1 \neq \mu_2$ (there is difference) values under two – tailed are seen.

- If the A.H is $\mu_1 < \mu_2$ or $\mu_1 > \mu_2$ (greater than or less than), values under one tailed are seen.

In the present case df (n-1) is = 9. Table "t" value for 9 df at 5% level of significance is 2.262 [Two-tailed test as the A.H is "not equal to"]

Interpretation : As the calculated value is less than the table value N.H is accepted.

Inference : Though there is numerical difference, the difference is not significant. That is, the mean price of shares is Rs.65/- only.

Example 2*:* The cool drinks stall at Singanallur in Coimbatore had a mean sale of 510 bottles of cool drinks per day . After the inauguration of the new bus stand the sale per day for the first seven days were 501, 519, 512, 513, 505, 526 and 508. Test at 10% level of significance whether there was any increase in the sale of cool drinks in the stall after inauguration of the bus stand.

Solution

N.H : There is no difference in the mean sale of cool drinks before or after opening up of the bus stand i.e $\mu_1 = \mu_2 (\mu = \bar{x})$

A.H : There is an increase in sales after opening up of the bus stand.($\bar{x} > \mu$)

Given data : $\mu = 510$, n = 7

$$\bar{x} = \frac{501+519+512+513+505+526+508}{7} = 512$$

Test criteria: As population mean is to be compared with a sample mean (where the sample is small (i.e <30) "t" test is used :

Calculated "t" value : $t = \dfrac{\bar{x}-\mu}{s/\sqrt{n}}$

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

$$== \sqrt{\frac{(501-512)^2+(519-512)^2+(512-512)^2+(513-512)^2+(505-512)^2+(526-512)^2+(508-512)^2}{7-1}}$$

$= 8.49$

$$t = \frac{512-510}{8.49/\sqrt{7}} = \frac{2}{8.49/2.65} = \frac{2}{3.2} = 0.625$$

Table value: Table "t" value for 6 df (n-1) at 10% level of significance is 1.44 (value under one-tailed is noted as the A.H is $\overline{x} > \mu$ *i.e* greater than context)

Interpretation: As the calculated "t" value (0.625) is less than the table 't' value (1.44)

The N.H is accepted.

Inference: There is no significant increase in the average sale of cool drinks after opening up of the bus stand. Though there is an increase of sales for one week, the difference is not significant.

To test the significance of difference between two small samples.

Example 1: An agricultural scientist tested the effect of two fertilizers, A and B on the yield of tomato. He applied fertilizer A in 5 plots and B in another 5 plots. The following were the yields from the 10 plots:

Fertilizer	Yield in kg / plot					
	1st plot	2nd Plot	3rd Plot	4th Plot	5th Plot	Mean
Fertilizer A	9	10	13	11	7	10
Fertilizer B	15	10	14	15	11	13

Table 9.5: Difference in the effects of fertilizers on tomato (example)

Test at 5% level of significance whether there is any difference in the effects of the two fertilizers

Solution

N.H: There is no difference in the effects of the fertizers A and B on the yield of tomato i.e $\mu_1 = \mu_2$ ($\overline{x}_1 = \overline{x}_2$)

A.H: There is difference

Given data: $n_1=5$; $n_2=5$ and yield in all the 10 plots

Test criterion: As the means of two small samples are to be compared "t" test is used.

Calculated t value:

$$t = \frac{\overline{x} - \overline{y}}{s\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}}$$

where $s=\sqrt{\dfrac{(n_1-1)s_1^2+(n_2-1)s_2^2}{n_1+n_2-2}}$

$$s_1^2=\frac{\sum(x-\bar{x}_1)^2}{n-1}=\frac{(9-10)^2+(10-10)^2+(13-10)^2+(11-10)^2+(7-10)^2}{5-1}=5$$

$$s_2^2=\frac{\sum(x-\bar{x}_2)^2}{n-1}=\frac{(15-13)^2+(10-13)^2+(14-13)^2+(15-13)^2+(11-13)^2}{5-1}=5.5$$

$$s=\sqrt{\frac{(4\times5)+(4\times5.5)}{(5+5)-2}}=2.29\ ; \qquad t=\frac{10-13}{2.29\sqrt{\frac{1}{5}+\frac{1}{5}}}=\frac{3}{1.45}=2.07$$

Table value: Table "t" value for 8 df (n_1 + n_2 – 2) at 5% level of significance is 2.306 [value under two-tailed test as A.H is $\mu_1 \neq \mu_2$]

Interpretation: As the calculated absolute value is less than table 't' value (2.306) the N.H is accepted

Inference: There is no significant difference in the effects of the two fertilizers. The difference observed is due to extraneous factors.

Example 2: Two salesmen A and B were working in a district. From a sample survey conducted by the Head Office, the following results were obtained

Salesmen	A	B
No. of sales made	20	18
Average sales (in thousand Rs.)	170	205
Standard Deviation (in thousand Rs.)	20	25

Differences in sales by salesmen (example)

Test at 1% level of significance whether the sales closed by salesman B is higher than that of salesman A.

Solution

N.H : There is no difference in the sales closed by both the salesmen *i.e* $\mu_1 = \mu_2 (\bar{x}_1 = \bar{x}_2)$

A.H : Sales made by salesman B is higher than that of salesman A i.e $\mu_2 > \mu_1$ $(\bar{x}_2 > \bar{x}_1)$

Given data : $n_1 = 20$; $n_2 = 18$ $\bar{x}_1 = 170$ $\bar{x}_2 = 205$ $s_1 = 20$ and $s_2 = 25$

Test criterion : As the means of two small samples are to be compared "t" test is used.

Calculated t value : $t = \dfrac{\bar{x}_1 - \bar{x}_2}{s\sqrt{\dfrac{1}{n_1} + \dfrac{1}{n_2}}}$

$$s = \sqrt{\frac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2}{n_1 + n_2 - 2}} = \sqrt{\frac{19(20)^2 + 17(25)^2}{(20 + 18) - 2}} = 22.5$$

$$t = \frac{170 - 205}{22.5\sqrt{\frac{1}{20} + \frac{1}{18}}} = -\frac{35}{7.31} = -4.79 \ = 4.79 \text{ (sign is ignored)}$$

Table value: Table "t" Value for 36 ($n_1 + n_2 - 2$) df at 1% level of significance lies between 2.423 and 2.457 (values under one-tailed probability level as the A.H. is $\bar{x}_1 < \bar{x}_2$)

Interpretation: As the calculated value (4.79) is higher than the table value (2.423 – 2.457) the N.H is rejected and A.H is accepted.

Inference: The sales made by salesman B is significantly higher than the sales made by salesman A.

Chi-Square as a Parametric test

Chi-square can be used as a parametric as well as a non-parametric test. As a parametric teat, it is used to find out the significance of difference between population variance and sample variance.

To find out the significance of difference between population variance and sample variance :

Example 1: The purchase manager in a star hotel, while placing order for apples for a party, stipulates that the size of the apples must be uniform and the standard deviation should not vary more than 15g. The supplier delivers 1000 apples. To verify whether the apples are uniform in weight he takes a sample of 10 apples and finds the S.D is 18g. Test at 10% level of significance whether the apples are delivered as per the stipulation.

Solution

N.H : There is no difference in population variance and sample variance i.e. variance of population is equal to sample variance [$\sigma^2 = s^2$]

A.H : There is difference.

Given data : Population S.D (σ) = 15g, S.D(s) of sample = 18g [Therefore σ^2 = 225 and

s^2 = 324] n = 10 , df = (n-1) = 9

Test criterion : As population variance is to be compared with sample variance χ^2 test as a parametric test is used.

Calculated χ^2 *value* : $\chi^2 = \dfrac{s^2(n-1)}{\sigma^2} = \dfrac{324(10-1)}{225} = 12.96$

Table value : Table χ^2 for 9 df at 10% level of significance is 14.68

Interpretation : As the calculated χ^2 value (12.96) is less than the table χ^2 value (14.68) the N.H is accepted.

Inference : There is no significant difference in the variances between the population and sample . The manager can accept the lot as difference in variances between his expectation and the sample is only due to chance factor.

Example 2: For the Republic Day Parade, NCC cadets of almost uniform heights are required. The authorities stipulate that the S.D should not be more than 3.0 cm. They took a sample of 10 cadets from a college and their heights are shown below:

Cadet no. :	1	2	3	4	5	6	7	8	9	10
Height (cm) :	173	165	169	173	166	164	176	172	168	174

Will the authorities accept the cadets from the college? Test at 5% level of significance.

Solution

N.H : There is no difference in the desired variance of population and sample variance

A.H : There is difference

Given data : σ = 3 cm *i.e.* σ^2 = 9; n = 10 ; heights of 10 units in the sample.

Test criterion : As population variance is to be compared with sample variance, χ^2 as a parametric test is used.

Calculated χ^2 value :

$$\chi^2 = \frac{s^2(n-1)}{\sigma^2}$$

$$\overline{x} = \frac{173+165+169+173+166+164+176+172+168+174}{10} = 170$$

$$s^2 = \frac{\sum(x-\overline{x})^2}{n-1}$$

$$= \frac{(173-170)^2+(165-170)^2+(169-170)^2+(173-170)^2+(166-170)^2+(164-170)^2+(176-170)^2+(172-170)^2+(168-170)^2+(174-170)^2}{9}$$

$= \frac{156}{9} = 17.33$

Calculated $\chi^2 = \frac{17.33(10-1)}{9} = 17.33$

Table value : Table χ^2 for 9 df at 5% is 16.919

Interpretation : As the calculated value (17.33) is greater than the table value (16.919), the N.H is rejected and A.H is accepted.

Inference : The variance of the sample is greater than the desired population variance. Hence, the authorities will not accept the cadets from the particular college.

Non – Parametric tests:

- In nonparametric test there is no need to make assumption that a population is distributed in the shape of a normal curve or another specific shape.
- Generally non-parametric tests are easy to carry out and understand.
- Sometimes even formal ranking is not required.
- These tests can be used when the measurements are not as accurate as is necessary for parametric tests.

Nevertheless, nonparametric tests are not so powerful as parametric tests.

Summary of non-parametric tests

Sl.No	Type of test	Objectives	Formulae
1.	Chi-square test as goodness of fit test	To find out the significance of difference between observed and expected values	$\chi^2 = \sum \frac{(O-E)^2}{E}$
2.	Chi-square test as test of independence	To find out whether there is any association between two sets of variables	$\chi^2 = \sum \frac{(O-E)^2}{E}$

Chi-square as goodness of fit :

As a non-parametric test, Chi-Square test is one of the simplest and most widely used tests. No assumptions are made about the population being sampled. Chi-square (pronounced as Kighsquare) values (χ^2) describe the magnitude of difference between theory and observation (*i.e* difference between observed values and expected values). Using χ^2 test as "test of goodness of fit" one can know whether a given difference between theory (expected values) and observation (observed values) can be attributed to chance factor or there is a real difference between expected and observed values. If χ^2 value is zero, it means that observed and expected values completely coincide (*i.e* the theory or expectation holds good). Higher the χ^2 value, greater is the discrepancy between observed and expected values. The formula for χ^2 is $\chi^2 = \sum \frac{(\mathrm{O}-\mathrm{E})^2}{\mathrm{E}}$ where O = Observed value and E = Expected or theoretical value.

For interpreting the discrepancy, the calculated χ^2 value is compared with the table value of χ^2 for a given df at specified level of significance; df is n-1, where 'n' is the no. of observations.

Example 1: An agricultural scientist desired to find out whether Mendelian 'law of inheritance' for shape (Round and Wrinkle) for peas is in accordance with the ratio of 3:1. To verify the theory he randomly selected a sample of 960 peas. Of these 960 peas, 704 were round and 256 were wrinkled. Do these observations conform to be expected ratio of 3:1 at 5% level of significance?

Solution

Based on the law, it is expected that the round peas should be 720 (3/4 x 960) and the wrinkled peas 240(1/4 x 960). But the observed values are 704 round peas and 256 wrinkled peas.

N.H : There is no significant difference between the expected values (theoretical values) and the observed values. That is, the law of inheritance holds good.

A.H : There is significant difference, that is, the law is incorrect.

	Round	**Wrinkled**	**Total**
Expected	720	240	960
Observed	704	256	960

Table 9.25: Observed and Expected values for peas (example)

Test criterion : As the observed values are to be compared with the expected values chi-square as a test of 'goodness of fit' is used.

Calculated χ^2 *Value :*

$$\chi^2 = \sum \frac{(O-E)^2}{E} = \frac{(704-720)^2}{720} + \frac{(256-240)^2}{240} = 1.42$$

Table value : Table χ^2 value for 1 df (df = no. of observations -1 = 2-1) at 5% level of significance is 3.84.

Interpretation : As the calculated χ^2 value (1.43) is less than the table value (3.84), the N.H is accepted.

Inference : It is inferred that the observation of 704 round peas and 256 wrinkled peas is as per the expected values of the Mendelian law (3:1 ratio of round peas and wrinkled peas). The difference seen is due to chance factor. Thus observations conform to the law.

Example 2: The number of parts manufactured for a particular spare part in a factory was found to vary from day to day. In a sample study the following information was obtained.

Day	**Mon**	**Tue**	**Wed**	**Thurs**	**Fri**	**Sat**	**Total**
No. of parts manufactured	1124	1125	1110	1120	1126	1115	6720

Production of spare parts (day-wise)

Test at 5% level of significance whether the production depends on the day of the week. (Delhi University, MBA,2000)

Solution

N.H : The production does not vary with the day of the week, it is same on all days.

A.H : Production varies from day to day.

Given data : Production in a day

Day	Mon	Tue	Wed	Thurs	Fri	Sat	Total
Observed	1124	1125	1110	1120	1126	1115	6720
Expected	1120	1120	1120	1120	1120	1120	6720

Observed vs Expected values for production (day-wise)

As per the N.H production does not depend on the day *i.e.* Production is expected to be uniform on all days *i.e.*1120 per day. (Total production in a week (6720) divided by the no. of working days (6)).

Test criterion : As the observed values are to be compared with the expected values χ^2 test is used.

Calculated χ^2 *value :*

$$\chi^2 = \sum \frac{(O-E)^2}{E} = \frac{(1124-1120)^2}{1120} + \frac{(1125-1120)^2}{1120} + \frac{(1110-1120)^2}{1120} + \frac{(1120-1120)^2}{1120} + \frac{(1126-1120)^2}{1120} + \frac{(1115-1120)^2}{1120} = 0.179$$

Table χ^2 *value :* Table χ^2 value at 5 df (6 observations – 1) at 5% level of significance is 11.07

Interpretation : As the calculated χ^2 value (0.179) is less than the table χ^2 value (11.07) the N.H is accepted.

Inference : It is inferred that there is no significant difference in the day-to-day production. The difference seen is only due to chance factor.

χ^2 test as a test of independence

Chi-square is also used to test independence of two or more variables. When individuals/items are classified simultaneously on two or more variables, the resulting table of cell frequencies is called a contingency table. χ^2 test is applied to contingency table to find out if the two or more variables are independent or associated

Example 1: In a survey of fertilizer practices in India, each of the 323 cotton – growing fields selected for the survey was classified on the twin criteria of irrigation practice (irrigated and not-irrigated) and the practice of manuring (maured and not-manured) resulting in the following contingency table.

	Irrigated	**Not irrigated**	**Total**
Manured	75	35	110
Not-manured	115	98	213
Total	190	133	323

Irrigation and Manuring (Observed values)

Test at 5% level of significance whether irrigation and manuring are independent of each other or depend on each other.

Solution

N.H : Irrigation and manuring are independent

A.H: Irrigation and manuring depend on each other

Test criterion : As independence of two variables is to be tested, χ^2 as a test of independence is used .

Test values : The expected values are calculated and presented in brackets along with the observed values.

	Irrigated	**Not irrigated**	**Total**
Manured	75(64.7)	35(45.3)	110
Not – manured	115(125.3)	98(87.7)	213
Total	190	133	323

Irrigation and Manuring (Observed & Expected values)

In a contingency table expected value in each cell is obtained by multiplying the column total with the row total and dividing the product by the grand total. For the observed value of 75 the expected value is

$$\frac{(190 \text{ X } 110)}{323} = 64.7$$

Calculated χ^2 *value :*

$$\chi^2 = \sum \frac{(O-E)^2}{E} = \frac{(75-64.7)^2}{64.7} + \frac{(35-45.3)^2}{45.3} + \frac{(115-125.3)^2}{125.3} + \frac{(98-87.7)^2}{87.7} = 6.04$$

Table value : (In a contingency table df = (c-1)(r -1) where c is the no. of columns and r is the no. of rows). Table χ^2 value for 1 df [(2-1)(2-1) at 5% level of significance is 3.84

Interpretation: As the calculated value (6.04) is greater than the table value (3.84), the N.H is rejected.

Inference : Irrigation and manuring are mutually related. i.e Generally irrigated fields are manured or manured fields are irrigated.

Example 2: A sample of 400 regular TV viewers was drawn from a city through convenience sampling and their preference of TV programme was enquired. The data are given below:

Category of TV viewers	Movies	News	Serials	Total
Teenagers	120	30	50	200
Adults	10	75	15	100
Old people	10	30	60	100
Total	140	135	125	400

TV viewers and programs (Observed values) (example)

Test at 1% level of significance whether there is any association between the different categories of people and types of TV programmes preferred.

Solution

N.H: The different groups of people and the TV programmes viewed are independent

A.H : There is association

Test criterion : As independence of variables is to be tested, χ^2 test is used.

Expected values are calculated and presented (in brackets) along with the observed values

Categories Of TV viewers	Types of TV programs			Total
	Movies	News	Serials	
Teenagers	120(70	30(67.50)	50(62.50)	200
Adults	10(35)	75(33.75)	15(31.25)	100
Old People	10(35)	30(33.75)	60(31.25)	100
Total	140	135	125	400

TV viewers and programs (Observed and Expected values) (example)

Calculated χ^2 *value:*

$$\chi^2 = \sum \frac{(O-E)^2}{E}$$

$$= \frac{(120-70)^2}{70} + \frac{(10-35)^2}{35} + \frac{(10-35)^2}{35} + \frac{(30-67.5)^2}{67.5} + \frac{(75-33.75)^2}{33.75} + \frac{(30-33.75)^2}{33.75}$$

$$+ \frac{(50-62.50)^2}{62.50} + \frac{(15-31.25)^2}{31.25} + \frac{(60-31.25)^2}{31.25}$$

$$= 180.45$$

df = (c-1)(r-1) = (3-1)(3-1) = 4

Table value : Table χ^2 for 4 df at 1% level of significance is 13.28

Interpretation : As the calculated value (180.45) is greater than the table value (13.28), the N.H is rejected.

Inference : There is significant association between the different categories of TV viewers and the types of TV programme. That is, teenagers prefer movies, adults news and old people serials.