# DEPARTMENT OF MATHEMATICS

**NAME OF THE SUBJECT : STATISTICS & NUMERICAL METHODS**

**SUBJECT CODE : MA8452**

**REGULATION : 2017**



## UNIT – I : TESTING OF HYPOTHESIS

# Unit-I Testing of Hypothesis

**Population:**

A population in statistics means a set of object. The population is finite or infinite according to the number of elements of the set is finites or infinite.

**Sampling:**

A sample is a finite subset of the population. The number of elements in the sample is called size of the sample.

**Large and small sample:**

The number of elements in a sample is greater than or equal to 30 then the sample is called a large sample and if it is less than 30, then the sample is called a small sample.

**Parameters:**

Statistical constant like mean μ, variance $\sigma^2$, etc., computed from a population are called parameters of the population.

**Statistics:**

Statistical constants like $\bar{x}$, variance $S^2$, etc., computed from a sample are called samlple staticts or statistics.

| POPULATION (PARAMETER) | SAMPLE (STATISTICS) |
|---|---|
| Population size=N | Sample size=n |
| Population mean= μ | Sample mean= $x$ |
| Population s.d.=σ | Sample s.d.=S |
| Population proportion= P | Sample proportion= p |



**<u>Tests of significance or Hypothesis Testing:</u>**

**Statistical Hypothesis:**

In making statistical decision, we make assumption, which may be true or false are called Statistical Hypothesis.

**Null Hypothesis( $H_0$ ):**

For applying the test of significance, we first setup a hypothesis which is a statement about the population parameter. This statement is usually a hypothesis of no true difference between sample statistics and population parameter under consideration and so it is called null hypothesis and is denoted by $H_0$.

**Alternative Hypothesis ( $H_1$ ):**

Suppose the null hypothesis is false, then something else must be true. This is called an alternative hypothesis and is denoted by $H_1$.

Eg. If $H_0$ is population mean μ=300, then $H_1$ is $\mu \neq 300$ $(ie. \mu < 300 \ or \ \mu > 300)$ or $H_1 \ is \ \mu > 300 \ or \ H_1 \ is \ \mu < 300$. So any of these may be taken as alternative hypothesis.

**<u>Error in sampling:</u>**

After applying a test of significance a decision is to be taken to accept or reject the null hypothesis $H_0$.

**Type I error:** The rejection of the null hypothesis $H_0$ when it is true is called type I error.

**Type II error:** The acceptance of the null hypothesis $H_0$ when it is false is called type II error.

**Level of significance:**

The probability of type I error is called level of significance of the test and it is denoted by α. We usually take either α=5% or α=1%.

**One tailed and Two tailed test:**

If $\theta_0$ is a population parameter and $\theta$ is the corresponding sample statistics and if we setup the null hypothesis $H_0 : \theta = \theta_0$, then the alternative hypothesis which is complementary to $H_0$ can be anyone of the following:

(i) $H_1 : \theta \neq \theta_0$ $(\theta < \theta_0 \;\; or \;\; \theta > \theta_0)$ (ii) $H_1 : \theta < \theta_0$ (iii) $H_1 : \theta > \theta_0$

Alternative hypotheis, whereas $H_1$ given in (ii) is called a left-tailed test. And (iii) is called a right tailed test.

**Level of significance:**

The probability of Type I error is called the level of significance of the test and is denoted by $\alpha$ .

**Critical region:**

For a test statistic, the area under the probability curve, which is normal is divided into two region namely the region of acceptance of $H_0$ and the region of rejection of $H_0$. The region in which $H_0$ is rejected is called critical region. The region in which $H_0$ is accepted is called acceptance region.



**Procedure of Testing of Hypothesis**:

(i) State the null hypothesis $H_0$

(ii) Decide the alternative hypothesis $H_1$ (ie, one tailed or two tailed)

(iii) Choose the level of significance α (α=5% or α=1%).

(iv) Determine a suitable test statistic.

$$\text{Test statistic} = \frac{t - E(t)}{S.E \; of \; (t)}$$

(v) Compute the computed value of $|z|$ with the table value of z and decide the acceptane or the rejection of $H_0$.

If $|z|$<1.96, $H_0$ may be accepted at 5% level of significance. If $|z|$>1.96, $H_0$ may be rejection at 5% level of significance.

If $|z|$<2.58, $H_0$ may be accepted at 1% level of significance. If $|z|$>2.58, $H_0$ may be rejection at 1% level of significance.

For a single tail test(right tail or left tail) we compare the computed value of $|z|$ with 1.645(at 5% level) and 2.33(at 1% level) and accept or reject $H_0$ accordingly.

**Test of significance of small sample:**

When the size of the sample (n) is less than 30, then that sample is called a small sample. The following are some important tests for small sample,

(I) students t test

(II) F-test

(III) $\chi^2$ -test

I **Student t test**

(i). Test of significance of the difference between sample mean and population mean
(ii). Test of significance of the difference between means of two small samples

(i) **Test of significance of the difference between sample mean and population mean:**

The studemts 't' is defined by the statistic $t = \frac{\bar{x} - \mu}{S/\sqrt{n}}$ where $\bar{x}$ =sample mean, μ=population mean, S=standard deviation of sample,
n= sample size.

**Note:**

If standard deviation of sample is not given directly then, the static is given by $t = \frac{\bar{x} - \mu}{S/\sqrt{n}}$, where

$$\bar{x} = \frac{\sum_{i=1} x_i}{n}, S^2 = \frac{\sum_{i=1} (x_i - \bar{x})^2}{n-1}$$



**Confident Interval:**

The confident interval for the population mean for small sample is $\bar{x} \mp t_\alpha \frac{s}{\sqrt{n}}$

$$\Rightarrow \left( \bar{x} - t_\alpha \frac{s}{\sqrt{n}}, \bar{x} + t_\alpha \frac{s}{\sqrt{n}} \right)$$

**Working Rule:**

(i) Let $H_0 : \mu = \bar{x}$ (there is no significant difference between sample mean and population mean)

$H_1 : \mu \neq \bar{x}$ (there is no significant difference between sample mean and population mean)(Two tailed test)

Find $t = \frac{\bar{x} - \mu}{S/\sqrt{n-1}}$.

Let $t_\alpha$ be the table value of t with v=n-1 degrees of freedom at $\alpha$ % level of significance.

**Conclusion:**

If $|t| < t_\alpha$, $H_0$ is accepted at $\alpha$ % level of significance.

If $|t| > t_\alpha$, $H_0$ is rejected at $\alpha$ % level of significance.

**Problem:**

1. **The mean lifetime of a sample of 25 bulbs is found as 1550h, with standard deviation of 120h. The company manufacturing the bulbs claims that the average life of their bulbs is 1600h. Is the claim acceptable at 5% level of significance?**

**Solution:**

Given sample size n=25, mean $\bar{x}$=1550, S.D.(S)=120, population mean μ=1600

Let $H_0 : \mu = 1600$ ( the claim is acceptable)

$H_1 : \mu \neq 1600$ ($\mu \neq \bar{x}$)(two tailed test)

Under $H_0$ , the test statistic is $t = \dfrac{\bar{x}-\mu}{S/\sqrt{n}} = \dfrac{1550-1600}{120/\sqrt{25}} = -2.0833$

$\therefore |t| = 2.0833$

From the table, for v=24, $t_{0.05}$=2.064. Since $|t| > t_{0.05}$

$\therefore H_0$ is rejected

Conclusion: The claim is not acceptable.

2. **Test made on the breaking strength of 10 pieces of a metal gave the following results: 578,572,570,568,572,570,570,572,596, and 584kg. Test if the mean breaking strength of the wire can be assumed as 577kg.**

**Solution:**

let us first compute sample mean $\bar{x}$ and sample standard deviation S and then test if $\bar{x}$ differs significantly from the population mean μ=577.

| x | $x-\bar{x}$ | $(x-\bar{x})^2$ |
|---|---|---|
| 578 | 2.8 | 7.84 |
| 572 | -3.2 | 10.24 |
| 570 | -5.2 | 27.04 |
| 568 | -7.2 | 51.84 |
| 572 | -3.2 | 10.24 |
| 570 | -5.2 | 27.04 |
| 570 | -5.2 | 27.04 |
| 572 | -3.2 | 10.24 |
| 596 | 20.8 | 432.64 |
| 584 | 8.8 | 77.44 |
| **5752** | **0** | **681.6** |



Where

$$\bar{x} = \frac{\sum_{i=1}^{n} x_i}{n} = \frac{5752}{10} = 575.2,$$

$$S^2 = \frac{\sum_{i=1}^{n}\left(x_i - \bar{x}\right)^2}{n-1} = \frac{681.6}{9} = 75.733$$

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

$H_1 : \mu \neq \bar{x}$

Under $H_0$ , the test statistic is $t = \frac{\bar{x} - \mu}{S/\sqrt{n}} = \frac{572.2 - 577}{\sqrt{75.733}/\sqrt{10}} = -1.74$

$\therefore |t| = 1.74$

Tabulated value of t for v=9 degrees of freedom $t_{0.05}$ =2.262

Since $|t| < t_{0.05}$ . $\therefore H_0$ is accepted

Conclusion:

$\therefore$ The mean breaking strength of the wire can be assumed as 577kg at 5% level of significance.

**3. A random sample of 10 boys had the following I.Q's: 70, 120, 110, 101, 88, 83, 95, 98,107,100. Do these data support the assumption of a population mean I.Q of 100 ? Find a reasonable range in which most of the mean I.Q. values of samples of 10 boys lie.**

**Solution:**

**Given** $\mu = 100,\ n = 10$

**Null Hypothesis:**

$H_0 : \mu = 100$ i.e., The data are consist with the assumption of men IQ of 100 in the population

**Alternate Hypothesis:**

$H_1 : \mu \neq 100$ i.e., The data are consist with the assumption of men IQ of 100 in the population



**Level of Significance :** $\alpha = 5\% \Rightarrow \alpha = 0.05$

**Test Statistic :**

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

where $S^2 = \frac{1}{n-1}\sum (x - \bar{x})^2$

$$\bar{x} = \frac{\sum x}{n} = \frac{70+120+110+101+88+83+95+98+107+100}{10} = \frac{972}{10} = 97.2$$

$$S^2 = \frac{1}{10-1}\begin{bmatrix}(70-97.2)^2 + (120-97.2)^2 + (110-97.2)^2 + (101-97.2)^2 + (88-97.2)^2 \\ +(83-97.2)^2 + (95-97.2)^2 + (98-97.2)^2 + (107-97.2)^2 + (100-97.2)^2\end{bmatrix}$$

$$S^2 = \frac{1}{9}(1833.6) = 203.73 \Rightarrow S = 14.2734$$

$$t = \frac{97.2 - 100}{14.2734/\sqrt{10}} = \frac{2.8}{4.5136} = 0.6203$$

**Table value :** $t_{\alpha,n-1} = t_{5\%,10-1} = t_{0.05,9} = 2.262$ **(Two –tailed test)**

**Conclusion :**

Here $t > t_\alpha$

i.e., The table value >calculated value,

$\therefore$ we accept the null hypothesis and conclude that the data are consistent with the assumption of mean I.Q of 100 in the population.

To find the confidence limit:

$$\left(\bar{x} \mp t_\alpha \frac{S}{\sqrt{n}}\right) = \left(97.2 \mp 2.262 \times \frac{14.2734}{\sqrt{10}}\right) = \left(97.2 \mp (2.262)(4.514)\right) = (86.99, 107.41)$$

A reasonable range in which most of the mean I.Q. values of samples of 10 boys lies (86.99,107.41)

4. **A random sample of 16 values from a normal population showed a mean of 41.5 inches and the sum of squares of deviations from this mean equal to 135 square inches. Show that the assumption of a mean of 43.5 inches for the population is not reasonable. Obtain 95 percent and 99 percent confidence limits for the same.**

**Solution:**

Given $\bar{x} = 41.5, \ \mu = 43.5, \ n = 16$

Sum of squares of deviations from mean= $\sum\left(x - \bar{x}\right)^2 = 135$

The parameter of interest is $\mu$.

**Null Hypothesis $H_0$:** $\mu$ =43.5 i.e., the assumption of a mean of 43.5 inches for the population is reasonable.

**Alternative Hypothesis $H_1$:** $\mu \neq 43.5$ i.e., the assumption of a mean of 43.5 inches for the population is not reasonable.

Level of significance: (i) $\alpha$ =5% = 0.05, degrees of freedom = 16–1=15

(ii) $\alpha$ =1% =0.01, degrees of freedom = 16–1=15

Test Statistic : $t = \dfrac{\bar{x} - \mu}{S/\sqrt{n}}$

where $S^2 = \dfrac{1}{n-1}\sum(x - \bar{x})^2 = \dfrac{1}{16-1} 135 = 9 \Rightarrow S = 9$

$$t = \frac{41.5 - 43.5}{3/\sqrt{16}} = \frac{-8}{3} = -2.667 \quad \Rightarrow |t| = 2.667$$

**Conclusion:**

(i) Since $|t|$ =2.667 > 2.131 so we reject $H_0$ at 5% level of significance.

So we conclude that the assumption of mean of 43.5 inches for the population is not reasonable.

(ii) Since $|t|$ =2.667 < 2.947 so we accept $H_0$ at 1% level of significance.

So we conclude that the assumption is reasonable.

95% confidence limits:

$$\left(\bar{x} \mp t_\alpha \frac{S}{\sqrt{n}}\right) = \left(41.5 \mp \left(2.947 \times \frac{3}{4}\right)\right) = (41.5 \mp 1.5983) = (39.9, 43.09)$$

$\therefore 39.902 < \mu < 43.098$

99% confidence limits:

$$\left(\bar{x} \mp t_\alpha \frac{S}{\sqrt{n}}\right) = \left(41.5 \mp \left(2.947 \times \frac{3}{4}\right)\right) = (41.5 \mp 2.2101) = (39.29, 43.71)$$

$\therefore 39.29 < \mu < 43.71$

5. **Ten oil tins are taken at random from an automatic filling machine. The mean weight of the tins is 15.8 kg and standard deviation is 0.5 kg. Does the sample mean differ significantly from the intended weight of 16 kg?**

**Solution:**

Given $\bar{x} = 15.8,\ \mu = 16,\ s = 0.50,\ n = 10$

**Null Hypothesis $H_0$**: $\mu = 16$ the sample mean weight is not different from the intended weight.

**Alternative Hypothesis $H_1$**: $\mu \neq 16$ i.e., the sample mean weight is not different from the intended weight.

**Level of significance:** $\alpha$ =5%= 0.05, degrees of freedom = 10-1=9

**Test Statistic :** $t = \dfrac{\bar{x} - \mu}{\frac{s}{\sqrt{n}}}$

$$t = \frac{15.8 - 16}{\frac{0.50}{\sqrt{10}}} = \frac{-0.2}{0.1581} = -1.27 \Rightarrow |t| = 1.27$$

**Critical value :** The critical value of $t$ at 5% level of significance with degrees of freedom 9 is 2.26

**Conclusion:**

Here calculated value < table value.

so we accept $H_0$ at 5% level of significance.

Hence the sample mean weight is not different from the intended weight.

(ii) **Test of significance of the difference between means of two small samples:**

To test the significance of the difference between the means $\bar{x}_1$ and $\bar{x}_2$ of sample of size $n_1$ and $n_2$.

Under $H_0$, the test statistic is $t = \dfrac{\bar{x}_1 - \bar{x}_2}{S\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}}$,

*where* $S = \sqrt{\dfrac{n_1 s_1^2 + n_2 s_2^2}{n_1 + n_2 - 2}}$ *or* $S^2 = \dfrac{\sum\left(x_1 - \bar{x}_1\right)^2 + \sum\left(x_2 - \bar{x}_2\right)^2}{n_1 + n_2 - 2}$ (*if* $s_1, s_2$ *is not given directly*)

Degrees of freedom(df) v = $n_1 + n_2$ -2

<u>Note:</u>

If $n_1 = n_2$ =n and if the pairs of values $x_1$ and $x_2$ are associated in some way (or correlated).

Then we use the statistic is $t = \frac{\bar{d}}{S/\sqrt{n-1}}$, where $\bar{d} = \frac{\sum d}{n}$ and $S^2 = \frac{\sum(d-\bar{d})^2}{n}$

Degrees of freedom v = n-1

**Confident Interval:**

The confident interval for difference between two population means for small sample is $(\bar{x}_1 - \bar{x}_2) \mp t_\alpha S\sqrt{\frac{1}{n_1}+\frac{1}{n_2}}$

**Problem:**

1. **Samples of two types of electric bulbs were tested for length of life and the following data were obtainded.**

| Sample | Size | Mean | S.D |
|---|---|---|---|
| I | 8 | 1234h | 36h |
| II | 7 | 1036h | 40h |

**Is the difference in the means sufficient to warrant that type I bulbs are superior type II bulbs?**

Solution:

Here $\bar{x}_1$ =1234, $\bar{x}_2$ =1036, $n_1$ =8, $n_2$ =7, $s_1$ =36, $s_2 = 40$

Let $H_0 : \bar{x}_1 = \bar{x}_2$,

$H_1 : \bar{x}_1 > \bar{x}_2$ (ie. Type I bulbs are superior to type II bulbs) (one tail test)

Under $H_0$, the test statistic is $t = \frac{\bar{x}_1 - \bar{x}_2}{S\sqrt{\frac{1}{n_1}+\frac{1}{n_2}}}$,

$$where\ S = \sqrt{\frac{n_1 s_1^2 + n_2 s_2^2}{n_1 + n_2 - 2}} = 40.7317$$

$$\therefore t = \frac{1234 - 1036}{40.7317\sqrt{\frac{1}{8}+\frac{1}{7}}} = 9.39$$

Degrees of freedom v = $n_1 + n_2$ -2=13

Tabulated value of t for 13 d.f. at 5% level of significance is $t_{0.05}$ =1.77

Since $|t| > t_{0.05}$. $\therefore H_0$ is rejected. $H_1$ is accepted.

Conclusion:

Type I bulbs may be regarded superior to type II bulbs at 5% level of significance.

**2. Two independent sample of size 8 and 7 contained the following value:**

| | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| **Sample I** | **19** | **17** | **15** | **21** | **16** | **18** | **16** | **14** |
| **Sample II** | **15** | **14** | **15** | **19** | **15** | **18** | **16** | |

**Is the difference between the sample means significant?**

**Solution:**

| $x_1$ | $x_1-\overline{x_1}$ | $\left(x_1-\overline{x}_1\right)^2$ | $x_2$ | $x_2-\overline{x_2}$ | $\left(x_2-\overline{x}_2\right)^2$ |
|---|---|---|---|---|---|
| 19 | 2 | 4 | 15 | -1 | 1 |
| 17 | 0 | 0 | 14 | -2 | 4 |
| 15 | -2 | 4 | 15 | -1 | 1 |
| 21 | 4 | 16 | 19 | 3 | 9 |
| 16 | -1 | 1 | 15 | -1 | 1 |
| 18 | 1 | 1 | 18 | 2 | 4 |
| 16 | -1 | 1 | 16 | 0 | 0 |
| 14 | -3 | 9 | | | |
| **136** | **0** | **36** | **112** | **0** | **20** |

$$\overline{x}_1 = \frac{\sum x_1}{n} = \frac{136}{8} = 17, \overline{x}_2 = \frac{\sum x_2}{n} = \frac{112}{7} = 16$$

$$S^2 = \frac{\sum\left(x_1-\overline{x}_1\right)^2+\sum\left(x_2-\overline{x}_2\right)^2}{n_1+n_2-2} = \frac{36+20}{8+7-2} = 4.3076 \Rightarrow S = 2.0754$$

Let $H_0 : \overline{x_1} = \overline{x_2}$,

$H_1 : \overline{x_1} \neq \overline{x_2}$ (Two tailed test)

Under $H_0$, the test statistic is $t = \dfrac{\overline{x_1}-\overline{x_2}}{S\sqrt{\dfrac{1}{n_1}+\dfrac{1}{n_2}}} = \dfrac{17-16}{2.0754\sqrt{\dfrac{1}{8}+\dfrac{1}{7}}} = 0.9309$

$|t| = 0.9309$

Degrees of freedom v = v = $n_1+n_2$-2=13

From the 't' table, v = 13 degrees freedom at 5% level of significance is $t_{0.05}$=2.16

Since $|t| < t_{0.05}$ $\therefore$ $H_0$ is accepted

<u>Conclusion:</u>

The two sample mean do not differ significantly at 5% level of significance.

**3. The following data represent the biological values of protein from cow's milk and buffalo's milk:**

| | | | | | | |
|---|---|---|---|---|---|---|
| **Cow's milk** | **1.82** | **2.02** | **1.88** | **1.61** | **1.81** | **1.54** |
| **Buffalo's milk** | **2.00** | **1.83** | **1.86** | **2.03** | **2.19** | **1.88** |

**Examine whether the average values of protein in the two samples significantly differ at 5% level.**

**Solution:**

Given $n_1 = n_2 = 6$

$H_0: \ \mu_1 = \mu_2$ There is no significant difference between the means of the two samples.

$H_1: \ \mu_1 \neq \mu_2$ There is a significant difference between the means of the two samples.

Test Statistic: $t = \dfrac{\bar{x} - \bar{y}}{S\sqrt{\dfrac{1}{n_1} + \dfrac{1}{n_2}}}$

| $x$ | $y$ | $x - \bar{x}$<br>$x - 1.78$ | $(x - \bar{x})^2$ | $y - \bar{y}$<br>$y - 1.965$ | $(y - \bar{y})^2$ |
|---|---|---|---|---|---|
| 1.82 | 2 | 0.04 | 0.0016 | 0.035 | 0.00123 |
| 2.02 | 1.83 | 0.24 | 0.0576 | -0.135 | 0.01823 |
| 1.88 | 1.86 | 0.1 | 0.01 | -0.105 | 0.01102 |
| 1.61 | 2.03 | -0.17 | 0.0289 | 0.065 | 0.00425 |
| 1.81 | 2.19 | 0.03 | 0.0009 | 0.225 | 0.0506 |
| 1.54 | 1.88 | -0.24 | 0.0576 | -0.085 | 0.00723 |
| Total<br>10.68 | 11.79 | | 0.1566 | | 0.09256 |



$\bar{x} = \dfrac{\sum x}{n_1} = \dfrac{10.68}{6} = 1.78; \ \bar{y} = \dfrac{\sum y}{n_2} = \dfrac{11.79}{6} = 1.965$

$S^2 = \dfrac{1}{6+6-2}\left[0.1566 + 0.09256\right] = (0.1)(0.2492) = 0.0249 \quad \Rightarrow S = 0.1578$

$t = \dfrac{1.78 - 1.956}{(0.1578)\sqrt{\dfrac{1}{6} + \dfrac{1}{6}}} = \dfrac{-0.176}{(0.1578)(0.5774)} = \dfrac{-0.176}{0.0911} = 1.9319$

**Critical value:** The critical value of t at 5% level of significance with degrees of freedom 10 is 2.228

Here calculated value < table value, we accept $H_0$

(i.e,) The difference between the mean protein values of the two varieties of milk is not significant at 5% level.

**4. The following data relate to the marks obtaind by 11 students in 2 test, one held at the beginning of a year and the other at the end of the year intensive coaching.**

| | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| **Test 1** | **19** | **23** | **16** | **24** | **17** | **18** | **20** | **18** | **21** | **19** | **20** |
| **Test 2** | **17** | **24** | **20** | **24** | **20** | **22** | **20** | **20** | **18** | **22** | **19** |

**Do the data indicate that the students have benefited by coaching?**

**Solution:**

The given data relate to the marks obtained in 2 tests by the same set of students. Hence the marks in the 2 set can be regarded as correlated.

We use t-test for paired values.

Let $H_0 : \overline{x_1} = \overline{x_2}$ ,

$H_1 : \overline{x_1} < \overline{x_2}$ (one tailed test)

| $x_1$ | $x_2$ | d = $x_1$ - $x_2$ | $d^2 = \left(\overline{x}_1 - \overline{x}_2\right)^2$ | d - $d$ | $\left(d - \bar{d}\right)^2$ |
|---|---|---|---|---|---|
| 19 | 17 | 2 | 4 | 3 | 9 |
| 23 | 24 | -1 | 1 | 0 | 0 |
| 16 | 20 | -4 | 16 | -3 | 9 |
| 24 | 24 | 0 | 0 | 1 | 1 |
| 17 | 20 | -3 | 9 | -2 | 4 |
| 18 | 22 | -4 | 16 | -3 | 9 |
| 20 | 20 | 0 | 0 | 1 | 1 |
| 18 | 20 | -2 | 4 | -1 | 1 |
| 21 | 18 | 3 | 9 | 4 | 16 |
| 19 | 22 | -3 | 9 | -2 | 4 |
| 20 | 19 | 1 | 1 | 2 | 4 |
| | | **-11** | | | **58** |



$$\bar{d} = \frac{\sum d}{n} = \frac{-11}{11} = -1 \quad S^2 = \frac{\sum\left(d - \bar{d}\right)^2}{n} = \frac{58}{11} = 5.272$$

the test statistic is $t = \dfrac{\bar{d}}{S/\sqrt{n-1}} = \dfrac{-1}{\sqrt{5.272}/\sqrt{10}} = -1.377 \Rightarrow |t| = 1.377$

from the table, v = n-1 = 10 (d.f.), $t_{0.05}$ =1.812

Since $|t| < t_{0.05}$ $\therefore$ $H_0$ is accepted

<u>Conclusion:</u>

The students have not benefitted by coaching.

**5.** **Ten Persons were appointed in the officer cadre in an office. Their performance was noted by giving a test and the marks were recorded out of 100.**

| Employee | A | B | C | D | E | F | G | H | I | J |
|---|---|---|---|---|---|---|---|---|---|---|
| Before training | 80 | 76 | 92 | 60 | 70 | 56 | 74 | 56 | 70 | 56 |
| After training | 84 | 70 | 96 | 80 | 70 | 52 | 84 | 72 | 72 | 50 |

**By applying the t-test, can it be concluded that the employees have been benefited by the training?**

**Solution:**

**Null Hypothesis $H_0$:** $\mu_1 = \mu_2$ i.e., the employees have not been benefited by the training.

**Alternative Hypothesis $H_1$:** $\mu_1 \neq \mu_2$ i.e., the employees have been benefited by the training.

**Level of significance:** $\alpha = 5\% = 0.05$ (one tailed test)

**Test Statistic :** $t = \dfrac{\bar{d}}{\frac{S}{\sqrt{n}}}$

where $S^2 = \dfrac{1}{n-1}\sum (d-\bar{d})^2$ & $\bar{d} = \dfrac{\sum d}{n}$

| Employees | Before | After | d | $(d-\bar{d})^2$ |
|---|---|---|---|---|
| A | 80 | 84 | -4 | 0 |
| B | 76 | 70 | 6 | 100 |
| C | 92 | 96 | -4 | 0 |
| D | 60 | 80 | -20 | 256 |
| E | 70 | 70 | 0 | 16 |
| F | 56 | 52 | 4 | 64 |
| G | 74 | 84 | -10 | 36 |
| H | 56 | 72 | -16 | 144 |
| I | 70 | 72 | -2 | 4 |
| J | 50 | 50 | 6 | 100 |
| Total | | | 44 | 44.4 |



$$\bar{d} = \frac{\sum d}{n} = \frac{-40}{10} = -4$$

$$S^2 = \frac{1}{n-1}\sum (d-\bar{d})^2 = \frac{1}{9}(720) = 80$$

$$t = \frac{\bar{d}}{\frac{S}{\sqrt{n}}} = \frac{-4}{8.94/\sqrt{10}} = -1.414 \Rightarrow |t| = 1.414$$

**Critical value :** The critical value of tat 5% level of significance with degrees of freedom 9 is 1.83

**Conclusion:**
Here calculated value < table value.
so we accept $H_0$
Hence the employees have not been benefited by the training.

6. **The weight gains in pounds under two systems of feeding of calves of 10 pairs of identical twins is given below.**

| **Twin pair** | **1** | **2** | **3** | **4** | **5** | **6** | **7** | **8** | **9** | **10** |
|---|---|---|---|---|---|---|---|---|---|---|
| **Weight gains under System A** | **43** | **39** | **39** | **42** | **46** | **43** | **38** | **44** | **51** | **43** |
| **Sytem B** | **37** | **35** | **34** | **41** | **39** | **37** | **37** | **40** | **48** | **36** |

**Discuss whether the difference between the two systems of feeding is significant.**

**Solution:**

**Null Hypothesis $H_0$:** $\mu_1 = \mu_2$ i.e., there is no significance difference between the two system of feedings

**Alternative Hypothesis $H_1$:** $\mu_1 \neq \mu_2$ i.e., there is significance difference between the two systemsof feedings.

**Level of significance:** $\alpha = 5\% = 0.05$ ( Two tailed test)

**Test Statistic :** $t = \frac{\bar{d}}{\frac{S}{\sqrt{n}}}$

where $S^2 = \frac{1}{n-1}\sum(d-\bar{d})^2$ & $\bar{d} = \frac{\sum d}{n}$



| Twin Pair | System A x | System B y | $d = x - y$ | $(d-\bar{d})^2$ |
|---|---|---|---|---|
| 1 | 43 | 37 | 6 | 2.56 |
| 2 | 39 | 35 | 4 | 0.16 |
| 3 | 39 | 34 | 5 | 0.36 |
| 4 | 42 | 41 | 1 | 11.56 |
| 5 | 46 | 39 | 7 | 6.76 |
| 6 | 43 | 37 | 6 | 2.56 |
| 7 | 38 | 37 | 1 | 11.56 |
| 8 | 44 | 40 | 4 | 0.16 |
| 9 | 51 | 48 | 3 | 1.96 |
| 10 | 43 | 36 | 7 | 6.76 |
| Total | | | 44 | 44.4 |

$$\bar{d} = \frac{\sum d}{n} = \frac{44}{10} = 4.4$$

$$S^2 = \frac{1}{n-1}\sum(d-\bar{d})^2 = \frac{1}{9}(44.4) = 4.93 \quad \Rightarrow S = 2.08$$

$$t = \frac{\bar{d}}{\frac{S}{\sqrt{n}}} = \frac{4.4}{2.08/\sqrt{10}} = 6.68$$

**Critical value :** The critical value of tat 5% level of significance with degrees of freedom 9 is 2.62

**Conclusion:**

Here calculated value < table value.

so we accept $H_0$

Hence there is no significance difference between the two systems of feedings.

II **F-test**

(i) To test whether if there is any significant difference between two estimates of population variance

(ii) To test if the two sample have come from the same population.

We use F-test:

The test statistic is given by $F = \frac{S_1^2}{S_2^2}$, if $S_1^2 > S_2^2$

Where $S_1^2 = \frac{n_1 s_1^2}{n_1 - 1}$ [ $n_1$ is the first sample size] and $S_2^2 = \frac{n_2 s_2^2}{n_2 - 1}$ [ $n_2$ is the second sample size]

The degrees of freedom $(v_1, v_2) = (n_1 - 1\ n_2 - 1)$

**Note :**

1. If $S_1^2 < S_2^2$ then $F = \frac{S_2^2}{S_1^2}$ (always F > 1)

2. To test whether two independent samples have been drawn from the same normal population, we have to test

i) Equality of population means using t-test or z-test, according to sample size.

ii) Equality of population variances using F-test



**Problem:**

1. **A sample of size 13 gave an estimated population variance of 3.0, while another sample of size 15 gave an estimate of 2.5. Could both sample be from population with the same variance?**

**Solution:**

Given $n_1 = 13$, $n_2 = 15$, $S_1^2 = 3.0$, $S_2^2 = 2.5$

Let $H_0 : S_1^2 = S_2^2$ (the two samples have been drawn from populations with same variance}

$H_1 : S_1^2 \neq S_2^2$

The test statistics is $F = \frac{S_1^2}{S_2^2} = \frac{3}{2.5} = 1.2$

From the table, with degrees of freedom v = $(n_1 - 1\ n_2 - 1)$ = (12, 14)

$F_{0.05} = 2.53$ Since $F < F_{0.05}$ $\therefore H_0$ is accepted

Conclusion:

The two sample could have come from two normal population with the same variance.

2. **Two sample of size 9 and 8 give the sums of squares of deviations from their respective means equal to 160 and 91 respectively. Could both samples be from populations with the same variance?**

**Solution:**

Given $n_1=9,\ n_2=8,\ \sum(x-\bar{x})^2=160,\ \sum(y-\bar{y})^2=91$

$$S_1^2=\frac{\sum(x-\bar{x})^2}{n_1-1}=\frac{160}{8}=20,\quad S_2^2=\frac{\sum(y-\bar{y})^2}{n_2-1}=\frac{91}{7}=13$$

Let $H_0:\sigma_1^2=\sigma_2^2$ (the two normal populations have the same variance}

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

The test statistics is $F=\frac{S_1^2}{S_2^2}=\frac{20}{13}=1.538$

From the table, with degrees of freedom v = $(n_1-1\ n_2-1)$ = (8,7)

$F_{0.05}=3.73$ Since $F<F_{0.05}\ \therefore H_0$ is accepted

Conclusion:

The two sample could have come from two populations with the same variance.

3. **Two random samples gave the following data:**

| Sample | Size | Mean | Variance |
|---|---|---|---|
| I | 8 | 9.6 | 1.2 |
| II | 11 | 16.5 | 2.5 |

**Cane we conclude that the two samples have been drawn from the same normal population?**



**Solution:**

The two samples have been drawn from the same normal population we have to check

(i) the variance of the population do not differ significantly by F-test.

(ii) the sample means do not differ significantly by t-test.

(i) F-test:

Given $n_1=8,\ n_2=11,\ s_1^2=1.2,\ s_2^2=2.5,\ \bar{x}_1=9.6,\ \bar{x}_2=16.5$

$$S_1^2=\frac{n_1s_1^2}{n_1-1}=\frac{8(1.2)}{7}=1.37\qquad S_2^2=\frac{n_2s_2^2}{n_2-1}=\frac{11(2.5)}{10}=2.75$$

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

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

The test statistics is $F=\frac{S_2^2}{S_1^2}$ (since $S_1^2<S_2^2$)

$$=\frac{2.75}{1.37}=2.007$$

From the table, $F_{0.05}(n_2-1,n_1-1)=F_{0.05}(10,7)=3.63$

Since $F<F_{0.05}\ \therefore H_0$ is accepted

(ii) t-test:(Equality of means)

Let $H_0:\mu_1=\mu_2$

$H_1:\mu_1\neq\mu_2$

Under $H_0$ , the test statistic is $t = \dfrac{\bar{x}_1 - \bar{x}_2}{S\sqrt{\dfrac{1}{n_1} + \dfrac{1}{n_2}}}$,

$$where\ S = \sqrt{\frac{n_1 s_1^2 + n_2 s_2^2}{n_1 + n_2 - 2}} = \sqrt{\frac{8(1.2) + 11(2.5)}{8 + 11 - 2}} = 1.4772$$

$$t = \frac{9.6 - 16.5}{1.4772\sqrt{\frac{1}{8} + \frac{1}{11}}} = -10.0525 \Rightarrow |t| = 10.0525$$

From the table ,with degrees of freedom $n_1 + n_2$ -2=17, $t_{0.05}$ =2.110

$\sin ce\ |t| > t_{0.05}$ $\therefore H_0$ is rejected ie. $\mu_1 \neq \mu_2$

Conclusion:

$\therefore$ The two samples could not have been drawn from the same normal population.

**4. Two independent samples of 5 and 6 items respectively had the following values of the following values of the variable:**

| **Sameple1:** | **21** | **24** | **25** | **26** | **27** | |
|---|---|---|---|---|---|---|
| **Sameple2:** | **22** | **27** | **28** | **30** | **31** | **36** |

**Can you say that the two samples came from the same population?**

**Solution:**

Let $H_0 : \sigma_1^2 = \sigma_2^2$ and $\mu_1 = \mu_2$ ( the two samples have been drawn from the same population)

$H_1 : \sigma_1^2 \neq \sigma_2^2$ and $\mu_1 \neq \mu_2$

(i) F-test : (Equality of variance)

| $x_1$ | $x_1 - \bar{x}_1$ | $(x_1 - \bar{x}_1)^2$ | $x_2$ | $x_2 - \bar{x}_2$ | $(x_2 - \bar{x}_2)^2$ |
|---|---|---|---|---|---|
| 21 | -3.6 | 12.96 | 22 | -7 | 49 |
| 24 | -0.6 | 0.36 | 27 | -2 | 4 |
| 25 | 0.4 | 0.16 | 28 | -1 | 1 |
| 26 | 1.4 | 1.96 | 30 | 1 | 1 |
| 27 | 2.4 | 5.76 | 31 | 2 | 4 |
| | | | 36 | 7 | 49 |
| 123 | | 21.2 | 174 | | 108 |

$$\bar{x}_1 = \frac{\sum x_1}{n} = \frac{123}{5} = 24.6, \bar{x}_2 = \frac{\sum x_2}{n_2} = \frac{174}{6} = 29$$

$$s_1^2 = \frac{\sum (x_1 - \bar{x}_1)^2}{n_1 - 1} = \frac{21.2}{4} = 5.3,\ s_2^2 = \frac{\sum (x_2 - \bar{x}_2)^2}{n_2 - 1} = \frac{108}{5} = 21.6$$

$$S_1^2 = \frac{n_1 s_1^2}{n_1 - 1} = \frac{5(5.3)}{4} = 6.625 \quad S_2^2 = \frac{n_2 s_2^2}{n_2 - 1} = \frac{6(21.6)}{5} = 25.92$$

The test statistics is $F = \dfrac{S_2^2}{S_1^2}$ $(\sin ce\ S_1^2 < S_2^2)$

$= \frac{25.92}{6.625} = 3.912$

From the table, $F_{0.05}(n_2 - 1, n_1 - 1) = F_{0.05}(5,4) = 6.26$

Since $F < F_{0.05}$ $\therefore H_0$ is accepted

(ii) t-test:(Equality of means)

Under $H_0$ , the test statistic is $t = \frac{\bar{x}_1 - \bar{x}_2}{S\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}}$,

$where\ S = \sqrt{\frac{n_1 s_1^2 + n_2 s_2^2}{n_1 + n_2 - 2}} = \sqrt{\frac{5(5.3) + 6(21.6)}{5 + 6 - 2}} = 4.164$

$t = \frac{24.6 - 29}{4.16\sqrt{\frac{1}{5} + \frac{1}{6}}} = -1.746 \Rightarrow |t| = 1.746$

From the table ,with degrees of freedom $n_1 + n_2$ -2=9, $t_{0.05}$ =2.262

$\sin ce\ |t| < t_{0.05}$ $\therefore H_0$ is accepted ie. $\mu_1 \neq \mu_2$

Conclusion: $\therefore$ The two samples could have been drawn from the same normal population.

**5. Two random samples gave the following results:**

| Sample | Size | Sample mean | Sum of squares of deviations from the mean |
|---|---|---|---|
| 1 | 10 | 15 | 90 |
| 2 | 12 | 14 | 108 |

**Test whether the samples come from the same normal population at 5% level of significance.**



**Solution:**

A normal population has 2 parameters namely mean µ and variance $\sigma^2$. To test if independent samples have been drawn from the same normal population, we have to test

1) Equality of population means using t-test or z-test, according to sample size.

2) Equality of population variances using F-test.

Given $\bar{x} = 15$, $\bar{y} = 14$, $n_1 = 10$, $n_2 = 12$, $\sum(x - \bar{x})^2 = 90$, $\sum(y - \bar{y})^2 = 108$

**i) t-test to test equality of population means:**

Null hypothesis $H_0$: $\mu_1 = \mu_2$ there is no difference between the two population means.

Alternate Hypothesis $H_1$: $\mu_1 \neq \mu_2$ there is difference between the two population means.

**Level of Significance** : $\alpha = 5\% = 0.05$ (Two tailed test )

**Test statistic:** $t = \frac{\bar{x} - \bar{y}}{S\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}}$

Where $S^2 = \frac{1}{n_1 + n_2 - 2}\left[\sum(x - \bar{x})^2 + \sum(y - \bar{y})^2\right] = \frac{1}{10 + 12 - 2}(90 + 108) = 9.9$

$S = \sqrt{9.9}$ =3.146

$$t = \frac{15-14}{3.146\sqrt{\frac{1}{10}+\frac{1}{12}}} = 0.742$$

Critical value: The critical value of t at 5% level of significance with degrees of freedom $n_1 + n_2 - 2 = 10 + 12 - 2 = 20$ is 2.086

Conclusion: calculated value < table value

$H_0$ is Accepted.

**ii) F-test to test equality of populations variances:**

**Null Hypothesis H₀:** $\sigma_1^2 = \sigma_2^2$ The population Variances are equal

**Alternative Hypothesis H₁:** $\sigma_1^2 \neq \sigma_2^2$ The population Variances are not equal

**Level of significance:** $\alpha = 5\%$

**Test Statistics:**

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

Where $S_1^2 = \frac{1}{n_1 - 1}\sum (x-\bar{x})^2 = \frac{1}{10-1}(90) = 10$

$S_1^2 = \frac{1}{n_1 - 1}\sum (y-\bar{y})^2 = \frac{1}{12-1}(108) = 9.818$

Here $S_1^2 > S_2^2$ $\therefore F = \frac{S_1^2}{S_2^2} = \frac{10}{9.818} = 1.02$

**Critical value:** The critical value of $F$ at 5% level of significance with degrees of freedom $(n_1 - 1, n_2 - 1) = (9, 11)$ is 2.90



Here calculated value < table value, we accept $H_0$

**Conclusion:** Both null hypothesis $\mu_1 \neq \mu_2$ and $\sigma_1^2 = \sigma_2^2$ are accepted.

Hence we may conclude the two samples are drawn from same normal population.

III $\chi^2$**-test:**

(i). $\chi^2$-Test for a specified population variance

(ii). $\chi^2$-test is used to test whether differences between observed and expected frequencies are significant (goodness of fit).

(iii). $\chi^2$-test is used to test the independence of attributes.

$\chi^2$ **-Test for a specified population variance:**

The test statistics $\chi^2 = \frac{ns^2}{\sigma^2}$

Which follows $\chi^2$ - distribution with (n – 1) degrees of freedom

**<u>Problem:</u>**

**1. The lapping process is used to grind certain silicon wafers to the proper thickness is acceptable only $\sigma$, the population S.D. of the thickness of dice cut from the wafers, is at most 0.5mil. Use the 0.05 level of significance to test the null hypothesis $\sigma$ =0.5 against the alternative hypothesis $\sigma$ >0.5, if the thickness of 15 dice cut from such wafers have S.D of 0.64mil.**

**Solution:**

Given $n = 15$, s=0.64, $\sigma$ =0.5

$H_0 : \sigma = 0.5$, $H_1 : \sigma > 0.5$

Under $H_0$, The test statistics $\chi^2 = \frac{ns^2}{\sigma^2} = \frac{15\,(0.64)^2}{(0.5)^2} = 24.576$

From $\chi^2$ table, with degrees of freedom = 14, $\chi^2_{0.05} = 23.625$

$\therefore \chi^2 > \chi^2_{0.05}$ $H_0$ is rejected. Hence $\sigma > 0.5$

**$\chi^2$-test is used to test whether differences between observed and expected frequencies are significant (goodness of fit):**

The test statistics $\chi^2 = \sum_i \left[ \frac{(O_i - E_i)^2}{O_i} \right]$

Where $O_i$ is observed frequency, and $E_i$ is the expected frequency.

If the data given in a series of n number, then degree of freedom = n - 1 .

<u>**Note**</u>: In case of binomial distribution d.f = n – 1, poisson distribution d.f = n – 2, normal distribution d.f = n – 3.

<u>**Problem:**</u>

1. **The following data give the number of aircraft accident that occurred during the various days of a week:**

| **Days :** | **Mon** | **Tue** | **Wed** | **Thu** | **Fri** | **Sat** |
|---|---|---|---|---|---|---|
| **No of accidents:** | **15** | **19** | **13** | **12** | **16** | **15** |

**Test the whether the accident are uniformly distributed over the week.**

**Solution:**

The expected number of accident on any day = $\frac{90}{6} = 15$

Let $H_0$: Accidents occur uniformly over the week

$H_1$: Accidents not occur uniformly over the week

| **Days** | **Observed Freqency** $(O_i)$ | **Expected Frequency** $(E_i)$ | $(O_i - E_i)$ | $\frac{(O_i - E_i)^2}{E_i}$ |
|---|---|---|---|---|
| Mon | 15 | 15 | 0 | 0 |
| Tue | 19 | 15 | 4 | 1.066 |
| Wed | 13 | 15 | -2 | 0.266 |
| Thu | 12 | 15 | -3 | 0.6 |
| Fri | 16 | 15 | 1 | 0.066 |
| Sat | 15 | 15 | 0 | 0 |
| | | 90 | | 1.998 |

Now, $\chi^2 = \sum_i \left[ \frac{(O_i - E_i)^2}{O_i} \right] = 1.998$

Here 6 observations are given, degrees of freedom = n – 1= 6 – 1 = 5

From $\chi^2$ table, with degrees of freedom = 5, $\chi^2_{0.05} = 11.07$

$\therefore \chi^2 < \chi^2_{0.05}$ $H_0$ is accepted.

Conclusion: $\therefore$ Accidents occur uniformly over the week

**2. A survey of 320 families with 5 children each revealed the following distribution:**

| **No. of Boys:** | **5** | **4** | **3** | **2** | **1** | **0** |
|---|---|---|---|---|---|---|
| **No. of Girls:** | **0** | **1** | **2** | **3** | **4** | **5** |
| **No. of families:** | **14** | **56** | **110** | **88** | **40** | **12** |

**Is the result consistent with the hypothesis that male and female births are equally probable?**

**Solution:**

Let $H_0$: Male and female births are equally probable

$H_1$: Male and female births are not equally probable

Probability of male birth $= p = \frac{1}{2}$, Probability of female birth $= q = \frac{1}{2}$

The probability of x male births in a family of 5 is $p(x) = 5C\ p^x q^{5-x}, x = 0,1,2...5$

Expected number of families with x male births $= 320 \times 5C_x p^x q^{5-x}, x = 0,1,2...5$

$$= 320 \times 5C_x \left(\frac{1}{2}\right)^x \left(\frac{1}{2}\right)^{5-x}$$

$$= 320 \times 5C_x \left(\frac{1}{2}\right)^5 = 10 \times 5C_x$$

The $\chi^2$ is calculated using the following table:

| No. of Boys | Observed Freqency $(O_i)$ | Expected Frequency $E_i = 10 \times 5C_x$ | $(O_i - E_i)$ | $\frac{(O_i - E_i)^2}{E_i}$ |
|---|---|---|---|---|
| 5 | 14 | 10 | 4 | 1.6 |
| 4 | 56 | 50 | 6 | 0.72 |
| 3 | 110 | 100 | 10 | 1 |
| 2 | 88 | 100 | -12 | 1.44 |
| 1 | 40 | 50 | -10 | 2 |
| 0 | 12 | 10 | 2 | 0.4 |
| Total | 320 | 320 | | 7.16 |

$$\therefore \chi^2 = 7.16$$

The tabulated value of $\chi^2$ for n – 1 = 6 – 1 =5 degrees of freedom at 5% level of significance $= \chi^2_{0.05} = 11.07$

Since $\chi^2 < \chi^2_{0.05}$. So we accepted $H_0$.

Conclusion: $\therefore$ The male and female births are equally probable.

**3. Fit a poisson distribution to the following data and test the goodness of fit.**

| **x:** | **0** | **1** | **2** | **3** | **4** | **5** | **6** |
|---|---|---|---|---|---|---|---|
| **f(x):** | **275** | **72** | **30** | **7** | **5** | **2** | **1** |

**Solution:**

Mean of the given distribution $= \bar{x} = \dfrac{\sum f_i x_i}{\sum f_i} = \dfrac{189}{392} = 0.482$

To fit a poisson distribution to the given data:
We take the parameter of the poisson distribution equal to the mean of the given distribution. $= \lambda = \bar{x} = 0.482$

The poisson distribution is given by $P(X = x) = \dfrac{e^{-\lambda} \lambda^x}{x!}$; $x = 0,1,2...\infty$

and the expected frequencies are obtained by $f(x) = \left(\sum f_i\right) \times \dfrac{e^{-\lambda} \lambda^x}{x!} = 392 \times \dfrac{e^{-0.482} (0.482)^x}{x!}$

we get $f(0) = 392 \times \dfrac{e^{-0.482} (0.482)^0}{0!} = 242.1$, $f(1) = 392 \times \dfrac{e^{-0.482} (0.482)^1}{1!} = 116.69$

$f(3) = 4.518$, $f(4) = 0.544$, $f(5) = 0.052 \approx 0.1$, $f(6) = 0.004 \approx 0$

| x: | 0 | 1 | 2 | 3 | 4 | 5 | 6 | Total |
|---|---|---|---|---|---|---|---|---|
| Expected Frequency: | 242.1 | 116.69 | 28.12 | 4.518 | 0.544 | 0.052 | 0.004 | 392 |

$H_0$: The poisson distribution fit well into the data.

$H_1$: The poisson distribution does not fit well into the data.

The $\chi^2$ is calculated using the following table:

| **x** | **Observed Freqency** $(O_i)$ | **Expected Frequency** $(E_i)$ | $\dfrac{(O_i - E_i)^2}{E_i}$ |
|---|---|---|---|
| 0 | 275 | 242.1 | 4.471 |
| 1 | 72 | 116.7 | 17.122 |
| 2 | 30 | 28.1 | 0.128 |
| 3 | **7** | **4.5** | |
| 4 | **5** **15** | **0.5** **5.1** | **19.218** |
| 5 | **2** | **0.1** | |
| 6 | **1** | **0** | |
| Total | 392 | 392 | 40.939 |

$\therefore \chi^2 = 40.939$

The tabulated value of $\chi^2$ for = 7 – 1 –1– 3 =2 degrees of freedom at 5% level of significance $= \chi^2_{0.05} = 5.991$

Since $\chi^2 > \chi^2_{0.05}$. So we rejected $H_0$.

Conclusion: $\therefore$
The Poisson distribution is not a good fit to the given data.

**$\chi^2$-test is used to test the independence of attributes:**

An attributes means a equality or characteristic. $\chi^2$- test is used to test whether the two attributes are associated or independent. Let us consider two attributes A and B. A is divided into three classes and B is divided into three classes.

| | | Attribute B | | | |
|---|---|---|---|---|---|
| | | $B_1$ | $B_2$ | $B_3$ | Total |
| Attribute A | $A_1$ | $a_{11}$ | $a_{12}$ | $a_{13}$ | $R_1$ |
| | $A_2$ | $a_{21}$ | $a_{22}$ | $a_{23}$ | $R_2$ |
| | $A_3$ | $a_{31}$ | $a_{32}$ | $a_{33}$ | $R_3$ |
| | Total | $C_1$ | $C_2$ | $C_3$ | **N** |

Now, under the null hypothesis $H_0$: The attributes A and B are independent and we calculate the expected frequency $E_{ij}$ for varies cells using the following formula.

$$E_{ij} = \frac{R_i \times C_j}{N}, \; i = 1, 2, \ldots r, \; j = 1, 2, \ldots s$$

| | | | |
|---|---|---|---|
| $E(a_{11}) = \frac{R_1 \times C_1}{N}$ | $E(a_{12}) = \frac{R_1 \times C_2}{N}$ | $E(a_{13}) = \frac{R_1 \times C_3}{N}$ | $R_1$ |
| $E(a_{21}) = \frac{R_2 \times C_1}{N}$ | $E(a_{22}) = \frac{R_2 \times C_2}{N}$ | $E(a_{23}) = \frac{R_2 \times C_3}{N}$ | $R_2$ |
| $E(a_{31}) = \frac{R_3 \times C_1}{N}$ | $E(a_{32}) = \frac{R_3 \times C_2}{N}$ | $E(a_{33}) = \frac{R_3 \times C_3}{N}$ | $R_3$ |
| $C_1$ | $C_2$ | $C_3$ | **N** |

and we compute $\chi^2 = \sum_{i=1}^{r} \sum_{j=1}^{s} \frac{(O_{ij} - E_{ij})^2}{E_{ij}}$



Which follows $\chi^2$ distribution with n = (r-1) (s-1) degrees of freedom at 5% or 1% level of significance.

**1. Calculate the expected frequencies for the following data presuming two attributes viz., conditions of home and condition of child as independent.**

| | Condition of home | | |
|---|---|---|---|
| | | **Clean** | **Dirty** |
| **Condition of Child** | **Clean** | **70** | **50** |
| | **Fair** | **80** | **20** |
| | **Dirty** | **35** | **45** |

**Use Chi-Square test at 5% level of significance to state whether the two attributes are independent.**

**Solution:**

**Null hypothesis** $H_0$: Conditions of home and conditions of child are independent.

**Alternate hypothesis** $H_1$: Conditions of home and conditions of child are not independent.

**Level of significance**: $\alpha = 0.05$

**The test statistics:** $\chi^2 = \sum_{i=1}^{r} \sum_{i=1}^{s} \frac{(O_{ij} - E_{ij})^2}{E_{ij}}$

**Analysis:**

| | Condition of home | | Total |
|---|---|---|---|
| | | Clean | Dirty | |
| Condition of Child | Clean | 70 | 50 | 120 |
| | Fair | 80 | 20 | 100 |
| | Dirty | 35 | 45 | 80 |
| Total | | 185 | 115 | 300 |

Expected Frequency $= \frac{\text{Corresponding row total} \times \text{Column total}}{\text{Grand Total}}$

Expected Frequency for 70 $= \frac{120 \times 185}{300} = 74$, Expected Frequency for 80 $= \frac{100 \times 185}{300} = 61.67$,

Expected Frequency for 35 $= \frac{80 \times 185}{300} = 49.33$, Expected Frequency for 50 $= \frac{120 \times 115}{300} = 46$,

Expected Frequency for 20 $= \frac{100 \times 115}{300} = 38.33$, Expected Frequency for 45 $= \frac{80 \times 115}{300} = 30.67$



| $O_{ij}$ | $E_{ij}$ | $O_{ij} - E_{ij}$ | $(O_{ij} - E_{ij})^2$ | $\frac{(O_{ij} - E_{ij})^2}{E_{ij}}$ |
|---|---|---|---|---|
| 70 | 74 | -4 | 16 | $\frac{16}{74} = 0.216$ |
| 50 | 46 | 4 | 16 | 0.348 |
| 80 | 61.67 | 18.33 | 335.99 | 5.448 |
| 20 | 38.33 | -18.33 | 335.99 | 8.766 |
| 35 | 49.33 | -14.33 | 205.35 | 4.163 |
| 45 | 30.67 | 14.33 | 205.35 | 6.695 |
| Total | | | | 25.636 |

$\therefore \chi^2 = 25.636$

$\alpha = 0.05$ Degrees of freedom $= (r-1)(c-1) = (3-1)(2-1) = 2 \quad \therefore \chi^2_\alpha = 5.991$

**Conclusion:**

Since $\chi^2 > \chi^2_\alpha$, we Reject our Null Hypothesis $H_0$. Hence, Conditions of home and conditions of child are not independent.

**2. The following contingency table presents the reactions of legislators to a tax plan according to party affiliation. Test whether party affiliation influences the reaction to the tax plan at 0.01 level of signification.**

| **Reaction** | | | | |
|---|---|---|---|---|
| **Party** | **In favour** | **Neutral** | **Opposed** | **Total** |
| **Party A** | **120** | **20** | **20** | **160** |
| **Party B** | **50** | **30** | **60** | **140** |
| **Party C** | **50** | **10** | **40** | **100** |
| **Total** | **220** | **60** | **120** | **400** |

**Solution:**

**Null hypothesis** $H_0$ : Party affiliation and tax plan are independent.

**Alternate hypothesis** $H_1$ : Party affiliation and tax plan are not independent.

**Level of significance**: $\alpha = 0.05$

**The test statistic**: $\chi^2 = \sum_{i=1}^{r} \sum_{i=1}^{s} \frac{(O_{ij} - E_{ij})^2}{E_{ij}}$

**Analysis:**



| | | **Reaction** | | |
|---|---|---|---|---|
| **Party** | **Infavour** | **Neutral** | **Opposed** | **Total** |
| **Party A** | 120 | 20 | 20 | 160 |
| **Party B** | 50 | 30 | 60 | 140 |
| **Party C** | 50 | 10 | 40 | 100 |
| **Total** | 220 | 60 | 120 | **400** |

E(120)= $\frac{160 \times 220}{400} = 88$; E(20)= $\frac{160 \times 60}{400} = 24$; E(20)= $\frac{160 \times 120}{400} = 48$

E(50)= $\frac{140 \times 220}{400} = 77$; E(30)= $\frac{140 \times 60}{400} = 21$; E(60)= $\frac{140 \times 120}{400} = 42$

E(50)= $\frac{100 \times 220}{400} = 55$; E(10)= $\frac{100 \times 60}{400} = 15$; E(40)= $\frac{120 \times 100}{400} = 30$

| $O_{ij}$ | $E_{ij}$ | $O_{ij}$ - $E_{ij}$ | $(O_{ij} - E_{ij})^2$ | $\frac{(O_{ij} - E_{ij})^2}{E_{ij}}$ |
|---|---|---|---|---|
| 120 | 88 | 32 | 1024 | 11.64 |
| 20 | 24 | -4 | 16 | 0.67 |
| 20 | 48 | -28 | 784 | 16.33 |
| 50 | 77 | -27 | 729 | 9.47 |
| 30 | 21 | 9 | 81 | 3.86 |
| 60 | 42 | 18 | 324 | 7.71 |
| 50 | 55 | -5 | 25 | 0.45 |
| 10 | 15 | -5 | 25 | 1.67 |
| 40 | 30 | 10 | 100 | 3.33 |
| Total | | | | 55.13 |

$\therefore \chi^2 = 55.13$

$\alpha = 0.05$ Degrees of freedom = $(r-1)(s-1) = (3-1)(3-1) = 4 \quad \therefore \chi^2_{0.05} = 13.28$

**Conclusion:** Since $\chi^2 > \chi^2_\alpha$, we Reject our Null Hypothesis $H_0$

Hence, the Party Affiliation and tax plan are dependent.



3. **From a poll of 800 television viewers, the following data have been accumulated as to, their levels of education and their preference of television stations. We are interested in determining if the selection of a TV station is independent of the level of education**

| Educational Level | | | | |
|---|---|---|---|---|
| **Public** | **High School** | **Bachelor** | **Graduate** | **Total** |
| **Broadcasting** | **50** | **150** | **80** | **280** |
| **Commercial Stations** | **150** | **250** | **120** | **520** |
| **Total** | **200** | **400** | **200** | **800** |

**(i) State the null and alternative hypotheses.**

**(ii) Show the contingency table of the expected frequencies. (iii) Compute the test statistic.**

**(iv) The null hypothesis is to be tested at 95% confidence. Determine the critical value for this test.**

**Solution:**

(i) **Null Hypothesis**: Selection of TV station is independent of level of education

**Alternative Hypothesis**: Selection of TV station is not independent of level of education

(ii) **Level of significance**: $\alpha = 0.05$

| Educational Level | | | | |
|---|---|---|---|---|
| **Public** | **High School** | **Bachelor** | **Graduate** | **Total** |
| **Broadcasting** | **50** | **150** | **80** | **280** |
| **Commercial Stations** | **150** | **250** | **120** | **520** |
| **Total** | **200** | **400** | **200** | **800** |

**To Find Expected frequency:**

$$\text{Expected Frequency} = \frac{\text{Corresponding row total} \times \text{Column total}}{\text{Grand Total}}$$

$$\text{Expected Frequency for 50} = \frac{280 \times 200}{800} = 70, \text{Expected Frequency for 150} = \frac{280 \times 400}{800} = 140$$

$$\text{Expected Frequency for 80} = \frac{280 \times 200}{800} = 70, \text{Expected Frequency for 150} = \frac{520 \times 200}{800} = 130$$

$$\text{Expected Frequency for 250} = \frac{520 \times 400}{800} = 260, \text{Expected Frequency for 120} = \frac{520 \times 200}{800} = 130$$

**The test statistic**: $\chi^2 = \sum_{i=1}^{r} \sum_{i=1}^{s} \frac{(O_{ij} - E_{ij})^2}{E_{ij}}$

**Analysis:**

| $O_{ij}$ | $E_{ij}$ | $O_{ij}$ - $E_{ij}$ | $(O_{ij} - E_{ij})^2$ | $\frac{(O_{ij} - E_{ij})^2}{E_{ij}}$ |
|---|---|---|---|---|
| 50 | 70 | -20 | 400 | 5.714 |
| 150 | 140 | 10 | 100 | 0.174 |
| 80 | 70 | 10 | 100 | 1.428 |
| 150 | 130 | 20 | 400 | 3.076 |
| 250 | 260 | -10 | 100 | 0.385 |
| 120 | 130 | -10 | 100 | 0.769 |
| TOTAL | | | | 11.546 |

(iii) Test statistic = 11.546

(iv) Critical Chi-Square = 5.991,

**Conclusion:** Calculated value > table value

Hence, we reject Null Hypothesis.

**Large sample:**

If the size of the sample n>30, then that samplw is said to be large sample. There are four important test to test the significance of large samples.

(i). Test of significance for single mean.
(ii). Test of significance for difference of two means.
(iii). Test of significance for single proportion
(iv). Test of significance for difference of two proportions.

**Note:**
(i). The sampling distribution of a static is approximately normal, irrespective of whether the distribution of the population is normal or not.
(ii). The sample statistics are sufficiently close to the corresponding population parameters and hence may be used to calculate the standard errors of the sampling distribution.
(iii). **Critical values for some standard LOS's (For Large Samples)**

| **Nature of test** | **1% (0.01) (99%)** | **2% (0.02) (98%)** | **5% (0.05) (95%)** | **10% (0.1) (90%)** |
|---|---|---|---|---|
| Two Tailed Test | $\|z_\alpha\| = 2.58$ | $\|z_\alpha\| = 2.33$ | $\|z_\alpha\| = 1.96$ | $\|z_\alpha\| = 1.645$ |
| One Tailed Test (Right tailed Test) | $z_\alpha = 2.33$ | $z_\alpha = 2.055$ | $z_\alpha = 1.645$ | $z_\alpha = 1.28$ |
| One Tailed Test (Left tailed Test) | $z_\alpha = -2.33$ | $z_\alpha = -2.055$ | $z_\alpha = -1.645$ | $z_\alpha = -1.28$ |



**Problem based on Test of significance for single mean:**

The test statistic $z = \frac{\bar{x} - \mu}{\sigma / \sqrt{n}}$ where $\bar{x}$ =sample mean, μ=population mean, $\sigma$ = standard deviation of population, n= sample size.

**Note:**

If standard deviation of population is not known then the static is $z = \frac{\bar{x} - \mu}{S / \sqrt{n}}$,

where S = standard deviation of sample.

**Confident Interval:**

The confident interval for μ when $\sigma$ is known and sampling is done from a normal population or with a large sample is $\bar{x} \mp z_\alpha \frac{\sigma}{\sqrt{n}}$

$$\Rightarrow \left( \bar{x} - z_{\alpha} \frac{\sigma}{\sqrt{n}}, \bar{x} + z_{\alpha} \frac{\sigma}{\sqrt{n}} \right)$$

If s is known ($\sigma$ is not known): $\bar{x} \pm z_{\alpha} \frac{s}{\sqrt{n}}$

1. **A sample of 100 students is taken from a large population, the mean height in the sample is 160cm. Can it be reasonable regarded that in the population the mean height is 165cm, and s.d. is 10cm. and find confident limit. Use an level of significance at 1%**

   **Solution:**

   Given n = 100, $\bar{x}$ =160cm, μ=165cm, $\sigma$ =10cm

   Let $H_0 : \mu = 165$

   $H_1 : \mu \neq 165$ (two tailed test)

   Under $H_0$, the test statistic is $z = \frac{\bar{x} - \mu}{\sigma/\sqrt{n}} = \frac{160 - 165}{10/\sqrt{100}} = -5$

   $\therefore |z| = -5$

   From the table, $z_{0.01}$ =2.58. Since $|z| > z_{0.01}$ $\therefore H_0$ is rejected. hence $\mu \neq 165$.

   Confident Interval:

   $$\left( \bar{x} - z_{\alpha} \frac{\sigma}{\sqrt{n}}, \bar{x} + z_{\alpha} \frac{\sigma}{\sqrt{n}} \right) = \left( 160 - 2.58 \frac{10}{\sqrt{100}}, 160 + 2.58 \frac{10}{\sqrt{100}} \right) = (157.42, 162.58)$$

2. **The mean breaking strength of the cables supplied by a manufacture is 1800 with a S.D of 100. By a new techniques in the manufacturing process, it it claimed that the breaking strength of the cable has increased. In order to test this claim, a sample of 50 cables is tested and it is found that the mean breaking strength is 1850. Can we support the claim at 1% level of significance?**

   

   **Solu:**

   Given n = 50, $\bar{x}$ =1850, μ=1800, $\sigma$ =100

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

   $H_1 : \bar{x} > \mu$ (one tailed test)

   Under $H_0$, the test statistic is $z = \frac{\bar{x} - \mu}{\sigma/\sqrt{n}} = \frac{1850 - 1800}{100/\sqrt{50}} = 3.535$

   $\therefore |z| = 3.535$

   From the table, $z_{0.01}$ =2.33. Since $|z| > z_{0.01}$ $\therefore H_0$ is rejected. hence $\bar{x} > \mu$.

3. **A sample of 900 members has a mean of 3.4 cms and s.d is 2.61 cms. Is the sample from a large population of mean 3.25cm and s.d is 2.61 cms. If the population is normal and its mean is unknown find the 95% confidence limits of true mean.**

   **Solution:**

   Given $n = 900$, $\mu = 3.25$, $\bar{x} = 3.4cm$, $\sigma = 2.61$, $s = 2.61$

   **Null Hypothesis $H_0$ :** Assume that there is no significant difference between sample mean and population mean. (i.e) $\mu = 3.25$

   **Alternative Hypothesis $H_1$ :** Assume that there is a significant difference between sample mean and population mean. (i.e) $\mu \neq 3.25$

**Level of significance** : $\alpha = 5\%$

**Test Statistic** :

$$z = \frac{\bar{x}-\mu}{\frac{s}{\sqrt{n}}} = \frac{3.4-3.25}{\frac{2.61}{\sqrt{900}}} = 1.724$$

**Critical value:** The critical value of $z$ for two tailed test at 5% level of significance is 1.96

**Conclusion:**

*i.e.,* $z = 1.724 < 1.96 \Rightarrow$ calculated value < tabulated value

Therefore We accept the null hypothesis $H_0$.

i.e., The sample has been drawn from the population with mean $\mu = 3.25$

**To find confidence limit:**

95% confidence limits are

$$\bar{x} \pm 1.96 \frac{\sigma}{\sqrt{n}} = 3.4 \pm 1.96\left(\frac{2.61}{\sqrt{900}}\right) = 3.4 \pm 0.1705 = (3.57, 3.2295)$$

4. **A lathe is set to cut bars of steel into lengths of 6 centimeters. The lathe is considered to be in perfect adjustment if the average length of the bars it cuts is 6 centimeters. A sample of 121 bars is selected randomly and measured. It is determined that the average length of the bars in the sample is 6.08 centimeters with a standard deviation of 0.44 centimeters.**
**(i) Formulate the hypotheses to determine whether or not the lathe is in perfect adjustment.**
**(ii) Compute the test statistic.**
**(iii) What is your conclusion?**



**Solution:**

Given $n = 121$, $\bar{x} = 6.08$, $\mu = 6$, $S = 0.44$

**Null Hypothesis $H_0$**: $\mu = 6$ i.e., Assume that the lathe is in perfect adjustment

**Alternative Hypothesis $H_1$**: $\mu \neq 6$ i.e., Assume that the lathe is not in perfect adjustment.

**Level of Significance :** $\alpha = 0.05$

**ii) Test Statistic :**

$$z = \frac{\bar{x}-\mu}{\frac{S}{\sqrt{n}}} = \frac{6.08-6}{\frac{0.44}{\sqrt{121}}} = \frac{0.08}{0.04} = 2$$

Table value: Table value at 5% level of significance is 1.96

**iii) Conclusion:**

Here calculated value > tabulated value

Hence we reject $H_0$.

5. **The mean life time of a sample of 100 light tubes produced by a company is found to be 1580 hours with standard deviation of 90 hours. Test the hypothesis that the mean lifetime of the tubes produced by the company is 1600 hours.**

**Solution:**

Given $n = 100$, $\bar{x} = 1580$, $\mu = 1600$, $S = 90$

**Null Hypothesis $H_0$**: $\mu = 1600$ i.e., There is no significance difference between the sample mean and population mean

**Alternative Hypothesis H1**: $\mu \neq 1600$ i.e., There is a significance difference between the sample mean and population mean

**Level of Significance :** $\alpha = 5\% = 0.05$

**Test Statistic :**

$$z = \frac{\bar{x} - \mu}{\frac{S}{\sqrt{n}}} = \frac{1580 - 1600}{\frac{90}{\sqrt{100}}} = \frac{-20}{9} = -2.22$$

$|z| = 2.22$

Table value: Table value at 5% level of significance is 1.96 (two tailed test)

**Conclusion:**

Here calculated value > tabulated value

Hence we reject $H_0$.

Hence the mean life time of the tubes produced by the company may not be 1600 hrs.

**Problem based on Test of significance for difference of two means:**

The test statistic $z = \dfrac{\bar{x}_1 - \bar{x}_2}{\sqrt{\dfrac{\sigma_1^2}{n_1} + \dfrac{\sigma_2^2}{n_2}}}$ where $\sigma_1, \sigma_2$ are S.D. of populations.

Test Statistic:

i) $Z = \dfrac{\bar{x}_1 - \bar{x}_2}{\sigma\sqrt{\dfrac{1}{n_1} + \dfrac{1}{n_2}}}$ If $\sigma$ is known and $\sigma_1 = \sigma_2$

ii)ii) $Z = \dfrac{\bar{x} - \bar{y}}{\sqrt{\dfrac{S_1^2}{n_1} + \dfrac{S_2^2}{n_2}}}$ If $\sigma$ is not known and $\sigma_1 \neq \sigma_2$, $S_1^2, S_2^2$ are known.



**Confident Interval:**

The confident interval for difference between two population mean for large sample,

(1) when $\sigma(\sigma_1, \sigma_2)$ is known is $(\bar{x}_1 - \bar{x}_2) \pm z_\alpha \sqrt{\dfrac{\sigma_1^2}{n_1} + \dfrac{\sigma_2^2}{n_2}}$

(2). when s $(s_1, s_2)$ is known is $(\bar{x}_1 - \bar{x}_2) \pm z_\alpha \sqrt{\dfrac{s_1^2}{n_1} + \dfrac{s_2^2}{n_2}}$

1. **In a random sample of size 500, the mean is found to be 20. In another independent sample of size 400, the mean is 15. Could the samples have been drawn from the same population with S.D 4?**

**Solution:**

Given $\bar{x}_1 = 20$, $\bar{x}_2 = 15$, $n_1 = 500$, $n_2 = 400$, $\sigma = 4$

Null hypothesis $H_0$: $\mu_1 = \mu_2$ The samples have been drawn from the same population.

Alternate Hypothesis $H_1$: $\mu_1 < \mu_2$ The samples could not have been drawn from same population.

**Level of Significance** : $\alpha = 5\% = 0.05$ (Two tailed test )

**Test statistic:** $z = \dfrac{\overline{x_1} - \overline{x_2}}{\sigma\sqrt{\dfrac{1}{n_2} + \dfrac{1}{n_1}}} = \dfrac{20 - 15}{4\sqrt{\dfrac{1}{500} + \dfrac{1}{400}}} = 18.6$

Critical value: The critical value of t at 1% level of significance is 2.58

Conclusion: calculated value > table value

$H_0$ is rejected

The samples could not have been drawn from same population.

2. **Test significance of the difference between the means of the samples, drawn from two normal populations with the same SD using the following data:**

| | Size | Mean | Standard Deviation |
|---|---|---|---|
| **Sample I** | **100** | **61** | **4** |
| **Sample II** | **200** | **63** | **6** |

**Solution:**

Given $\bar{x}_1 = 60,\ \bar{x}_2 = 63, s_1 = 4,\ s_2 = 6, n_1 = 100,\ n_2 = 200$

Null hypothesis $H_0:\ \mu_1 = \mu_2$ there is no significance difference between the means of the samples.

Alternate Hypothesis $H_1:\ \mu_1 \neq \mu_2$ there is a significance difference between the means of the samples.

**Level of Significance** : $\alpha = 5\% = 0.05$ (two tailed test )

**Test statistic:** $z = \dfrac{\overline{x_1} - \overline{x_2}}{\sqrt{\dfrac{s_1^{\ 2}}{n_2} + \dfrac{s_2^{\ 2}}{n_1}}} = \dfrac{61 - 63}{\sqrt{\dfrac{4^2}{200} + \dfrac{6^2}{100}}} = -3.02 \Rightarrow |z| = 3.02$

Critical value: The critical value of t at 5% level of significance is 1.96

Conclusion: calculated value > table value

$H_0$ is rejected .Therefore the two normal populations, from which the samples are drawn, may not have the same mean though they may have the same S.D.

3. **A sample of heights of 6400 Englishmen has a mean of 170cm and a S.D of 6.4cm, while a simple sample of heights of 1600 Americans has a mean of 172cm and a S.D of 6.3cm. D the data indicate that Americans are on the average, taller than Englishmen?**

**Solution:**

Given $\bar{x}_1 = 170,\ \bar{x}_2 = 172, s_1 = 6.4,\ s_2 = 6.3,\ n_1 = 6400,\ n_2 = 1600$

Null hypothesis $H_0:\ \mu_1 = \mu_2$ there is no significance difference between the heights of Americans and Englishmen.

Alternate Hypothesis $H_1: \mu_1 < \mu_2$ Americans are on the average, taller than Englishmen

**Level of Significance** : $\alpha = 5\% = 0.05$ (one tailed test )

**Test statistic:** $z = \dfrac{\overline{x_1} - \overline{x_2}}{\sqrt{\dfrac{s_1^{\ 2}}{n_2} + \dfrac{s_2^{\ 2}}{n_1}}} = \dfrac{170 - 172}{\sqrt{\dfrac{6.4^2}{6400} + \dfrac{6.3^2}{1600}}} = -11.32 \Rightarrow |z| = 11.32$

Critical value: The critical value of t at 5% level of significance is 1.645

Conclusion: calculated value > table value

$H_0$ is rejected. We conclude that the data indicate that Americans are on the average, taller than Englishmen.

4. **The aveage marks scored by 32 boys is 72 with a S.D of 8, while that for 36 girls is 70 with a S.D of 6. Test at 1%level of significance whether the boys perform beter than girls.**

**Solution:**

Given $\bar{x}_1 = 72, \bar{x}_2 = 70, s_1 = 8, s_2 = 6, n_1 = 32, n_2 = 36$

$H_0: \mu_1 = \mu_2$ (Both perfom are equal)

$H_0: \mu_1 > \mu_2$ (Boys are better than girls) (one tailed test)

**The test statistic:** $z = \frac{\overline{x_1} - \overline{x_2}}{\sqrt{\frac{s_1^2}{n_2} + \frac{s_2^2}{n_1}}} = \frac{72 - 70}{\sqrt{\frac{8^2}{32} + \frac{6^2}{36}}} = 1.15$

Critical value: The critical value of t at 1% level of significance is 2.33

**Conclusion:** calculated value < table value

$H_0$ is accepted. Hence both are equal.

**Problem based on Test of significance for single proportion:**

To test the significant difference between the sample proportion p and the population proportion P, then we use the test statistic

$z = \frac{p - P}{\sqrt{\frac{PQ}{n}}}$, where Q = 1 – P



**Confident Interval:**

The confident interval for population proportion for large sample is $p \; z_{\alpha}\sqrt{\frac{PQ}{n}}$

1. **In a big city 325 men out of 600 men were found to be smokers. Does this information support the conclusion that the majority of men in this city are smokers?**

**Solution:**

Given n=600 , Number of smokers=325

p = sample proportion of smokers $\Rightarrow p = 325/600 = 0.5417$

P= Population proportion of smokers in the city $= 1/2 = 0.5 \Rightarrow Q = 0.5$

**Null Hypothesis $H_0$:** The number of smokers and non-smokers are equal in the city.

**Alternative Hypothesis $H_1$:** P > 0.5 (Right Tailed)

**Test Statistic:**

$$z = \frac{p - P}{\sqrt{\frac{PQ}{n}}} = \frac{0.5417 - 0.5}{\sqrt{\frac{0.5*0.5}{600}}} = 2.04$$

**Critical value:**
Tabulated value of z at 5% level of significance for right tail test is 1.645.

**Conclusion:**

Since Calculated value of z > tabulated value of z.

We reject the null hypothesis. The majority of men in the city are smokers.

**2. 40 people were attacked by a disease and only 36 survived. Will you reject the hypothesis that the survival rate, if attacked by this disease, is 85% at 5% level of significance?**

**Solution:**

Given

The Sample proportion, $p = \frac{36}{40} = 0.90$

Population proportion $P = 0.85 \Rightarrow Q = 1 - P = 1 - 0.85 = 0.15$

**Null Hypothesis H₀**: $P = 0.85$ i.e., There is no significance difference in survival rate

**Alternative Hypothesis H₁**: $P \neq 0.85$

i.e., There is a significance difference in survival rate.

**Level of Significance :** $\alpha = 0.05$

**Test Statistic :**

$$z = \frac{p - P}{\sqrt{\frac{PQ}{n}}} = \frac{0.90 - 0.85}{\sqrt{\frac{0.85 \times 0.15}{40}}} = 0.886$$

**Table value:** Tabulated value of z at 5% level of significance is 1.96



**Conclusion :** The table value >calculated value

Hence we accept the null hypothesis

Conclude that the survival rate may be taken as 85%.

**3. A Manufacturer of light bulbs claims that an average 2% of the bulbs manufactured by his firm are defective. A random sample of 400 bulbs contained 13 defective bulbs. On the basis of this sample, can you support the manufacturer's claim at 5% level of significance?**

**Solution:**

Given $n = 400$

$p =$ Sample proportion of defectives $= \frac{X}{n} = \frac{13}{400} = 0.0325$

**Null Hypothesis H₀**: $P = 2\% = 0.02$ i.e., Assume that 2% bulbs are defective.

**Alternative Hypothesis H₁**: $P \neq 2\% \neq 0.02$ i.e., Assume that 2% bulbs are non-defective.

**Level of significance:** $\alpha = 5\% = 0.05$

**Test Statistic :** $z = \frac{p - P}{\sqrt{\frac{PQ}{n}}}$

$$z = \frac{0.0325 - 0.02}{\sqrt{\frac{0.02 \times 0.98}{400}}} = \frac{0.0125}{0.0007} = 1.7857$$

**Critical value :** The critical value of tat 5% level of significance is 1.645 (one tailed test)

**Conclusion:**

Here calculated value > table value.

So we accept $H_0$. Hence the manufacturers claim cannot be supported.

4. **A salesman in a departmental store claims that at most 60 percent of the shoppers entering the store leave without making a purchase. A random sample of 50 shoppers should that 35 out of them left without making a purchase. Are these sample reults consistent with the claim of the salesman? Use an LOS of 0.05.**

**Solution:**

Let p = Sample proportion of shoppers not making a purchase = $\frac{35}{50} = 0.7$

P = Population proportion of shoppers not making a purchase = $60\% = \frac{60}{100} = 0.6$,

and Q = 1 – P = 0.4

**$H_0$**: $P = 0.6$ i.e., The claim is accepted

**$H_1$**: $P \neq 0.6$ (two tailed test)

The test Statistic is $z = \frac{p-P}{\sqrt{\frac{PQ}{n}}} = \frac{0.7-0.6}{\sqrt{\frac{0.6\times0.4}{50}}} = 1.445$

From the table, $z_{0.05}$ =1.96. Since $|z| < z_{0.05}$ $\therefore H_0$ is accepted

**Conclusion:**

The sample reults are consistent with the claim of the salesman.

**Problem based on Test of significance for Two proportion:**

To test the significant difference between the sample proportion $p_1$ and $p_2$ and the population proportion P, then we use the test statistic

$$z = \frac{p_1 - p_2}{\sqrt{PQ\left(\frac{1}{n_1}+\frac{1}{n_2}\right)}}$$, where Q = 1 – P

If P is not known, then $P = \frac{n_1 p_1 + n_2 p_2}{n_1 + n_2}$

**Confident Interval:**

The confident interval for difference between two population proportion for large sample is $(p_1 - p_2)\,\square_{\alpha}\sqrt{PQ\left(\frac{1}{n_1}+\frac{1}{n_2}\right)}$

1. **Before an increase in excise duty on tea, 800 people out of a sample of 1000 were consumers of tea. After the increase in duty, 800 people were consumers of tea in a sample of 1200 persons. Find whether there is significant decrease in the consumption of tea after the increase in duty. Also find confident limit.**

**Solution:**

Given $n_1$ =1000, $n_2$ =1200

$p_1$ = proportion of tea drinkers before increase inexcise duty= $\frac{800}{1000} = 0.8$

$p_2$ = proportion of tea drinkers before increase inexcise duty= $\frac{800}{1200} = 0.6667$

Null hypothesis: $H_0 : P_1 = P_2$ there is no significance difference in the consumption of tea before after increase in excise duty

Alternate hypothesis: $H_1 : P_1 \neq P_2$ there is a significance difference in the consumption of tea before after increase in excise duty

Level of significance: $\alpha = 5\% = 0.05$

**Test Statistic:** $z = \dfrac{p_1 - p_2}{\sqrt{PQ\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}}$

Where

$P = \dfrac{n_1 p_1 + n_2 p_2}{n_1 + n_2} = \dfrac{(0.8)(1000) + (0.67)(1200)}{1000 + 1200} = 0.7273 \quad \Rightarrow Q = 1 - P = 1 - 0.7273 = 0.2727$

$$z = \frac{0.8 - 0.6667}{\sqrt{(0.7273)(0.2727)\left(\frac{1}{1000} + \frac{1}{1200}\right)}} = \frac{0.1333}{0.01907} = 6.99$$

**Critical value**: the critical value of $z$ at 5% level of significance is 1.645

**Conclusion:**

Here calculated value > table value

$\therefore$ We reject $H_0$

Hence there is no significance difference in the consumption of tea before after increase in excise duty.



**Confident Interval:**

The confident interval for difference between two population proportion for large sample is

$$(p_1 - p_2)\ z_{\alpha}\sqrt{PQ\left(\frac{1}{n_1} + \frac{1}{n_2}\right)} = (0.8 - 0.667)\ 1.645\sqrt{0.7273 \times 0.2727\left(\frac{1}{1000} + \frac{1}{1200}\right)}$$

$$= (0.1016, 0.1644)$$

2. **Random samples of 400 men and 600 women asked whether they would like to have a flyover near their residence.200 men and 325 women were in favor of the proposal. Test the hypothesis that proportions of men and women in favor of the proposal are same against that they are not, at 5% level.**

**Solution:**

Given $n_1 = 400,\ n_2 = 600$

$p_1$ = proportion of men = $\frac{200}{400} = 0.5$

$p_2$ = proportion of women = $\frac{325}{600} = 0.541$

Null hypothesis: $H_0 : P_1 = P_2$ Assume that there is no significant difference between the option of men and women as far as proposal of flyover is concerned.

Alternate hypothesis: $H_1 : P_1 \neq P_2$ Assume that there is significant difference between the option of men and women as far as proposal of flyover is concerned

Level of significance: $\alpha = 5\% = 0.05$ (two tailed)

Test Statistic: $z = \dfrac{p_1 - p_2}{\sqrt{PQ\left(\dfrac{1}{n_1} + \dfrac{1}{n_2}\right)}}$

Where $P = \dfrac{n_1 p_1 + n_2 p_2}{n_1 + n_2} = \dfrac{(400)(0.5) + (600)(0.541)}{400 + 600} = 0.525 \quad \Rightarrow Q = 1 - P = 1 - 0.525 = 0.475$

$z = \dfrac{0.5 - 0.541}{\sqrt{(0.525)(0.475)\left(\dfrac{1}{400} + \dfrac{1}{600}\right)}} = \dfrac{-0.041}{0.032} = -1.34 \quad \Rightarrow |z| = 1.34$

Critical value: the critical value of z at 5% level of significance is 1.96

Conclusion:

Here calculated value < table value

$\therefore$ We accept $H_0$ at 5% level of significance.

Hence There is no difference between the option of men and women as far as proposal of flyover are concerned.

3. **A machine puts out 16 imperfect articles in a sample of 500. After the machine is overhauled, it puts out 3 imperfect articles in a batch of 100. Has the machine improved?**

**Solution:**

**Hypothesis:**

$H_0 : P_1 = P_2$

$H_1 : P_1 > P_2$

**Level of Significance :** $\alpha = 0.05$



**Test Statistic :** $Z = \dfrac{p_1 - p_2}{\sqrt{PQ\left(\dfrac{1}{n_1} + \dfrac{1}{n_2}\right)}}$

**Analysis:**

The Sample proportion,

$p_1 = \dfrac{16}{500} = 0.032, \quad p_2 = \dfrac{3}{100} = 0.03, \quad P = \dfrac{n_1 p_1 + n_2 p_2}{n_1 + n_2} = 0.032 \ \ \& \ Q = 1 - P = 0.968$

$Z = \dfrac{p_1 - p_2}{\sqrt{PQ\left(\dfrac{1}{n_1} + \dfrac{1}{n_2}\right)}} = \dfrac{0.032 - 0.03}{\sqrt{0.032 \times 0.968\left(\dfrac{1}{500} + \dfrac{1}{100}\right)}} = 0.1037$

**Table value :** $Z_\alpha = 1.645$

**Conclusion:**

Calculated value < table value

Hence we accept the null hypothesis and conclude that the machine has not improved after overhauling.