Lecture note on Test of Hypothesis

Course Teacher:
Md. Shahajada Mia
Assistant Professor
Department of Statistics
Pabna University of Science and Technology

- **Test Regarding Mean**

Let $x_1, x_2, \cdots\cdots\cdots, x_n$ be n sample observations drawn independently from a population with mean μ and variance σ^2. Let us assume that the sample observations follow normal distribution, i.e. $x \sim N(\mu, \sigma^2)$. The problem is to set the null hypothesis

$$H_0: \mu = \mu_0$$

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

The assumptions for the test are

(i) σ^2 is unknown and n is small $(n < 30)$
(ii) σ^2 is unknown and n is large $(n \geq 30)$
(iii) σ^2 is known (n is small or large) and its value is ${\sigma_0}^2$ (say).

The test statistic to test the significance of H_0 is

$$t = \frac{\bar{x}-\mu_0}{\frac{s}{\sqrt{n}}} \text{, under assumption (i)}$$

This 't' follows student's 't' distribution with (n-1) d.f.

The test statistic is

$$z = \frac{\bar{x}-\mu_0}{\frac{s}{\sqrt{n}}} \sim N\ (0,1) \text{, under assumption (ii)}$$

The test statistic is

$$z = \frac{\bar{x}-\mu_0}{\frac{\sigma_0}{\sqrt{n}}} \sim N\ (0,1) \text{, under assumption (iii)}$$

Where, $\bar{x} = \frac{\sum x_i}{n}$ and $s^2 = \frac{1}{n-1} \sum (x_i - \bar{x})^2$.

Comment: If the calculated value of test statistic is greater than the critical value of test statistic then the null hypothesis (H_0) is rejected otherwise it is accepted.

or if p-value is less than the level of significance then the null hypothesis (H_0) is rejected otherwise it is accepted.

- **Test statistics for testing a population mean** (μ)

1) $z_{calculated} = \frac{\bar{x} - \mu_0}{\frac{\sigma}{\sqrt{n}}}$ $\quad \left[When,\ \sigma\ is\ known\ and\ n\ is\ large \right]$

2) $z_{calculated} = \frac{\bar{x} - \mu_0}{\frac{s}{\sqrt{n}}}$ $\quad \left[When,\ \sigma\ is\ unknown\ and\ n\ is\ large \right]$

3) $z_{calculated} = \dfrac{\bar{x} - \mu_0}{\dfrac{\sigma}{\sqrt{n}}}$ $\left[When,\ \sigma\ is\ known\ and\ n\ is\ small \right]$

4) $t_{calculated} = \dfrac{\bar{x} - \mu_0}{\dfrac{s}{\sqrt{n}}}$ $\left[When,\ \sigma\ is\ unknown\ and\ n\ is\ small \right]$

❖ **Lower or left tail test about a population mean** (μ)

$$H_0: \mu \geq \mu_0 \qquad\qquad H_A: \mu < \mu_0$$

Reject H_0 if

$$z_{calculated} < -z_\alpha \quad \left(For\ 1, 2, 3 \right)$$

$$t_{calculated} < -t_\alpha \quad \left(For\ 4 \right)$$

❖ **Upper or right tail test about a population mean** (μ)

$$H_0: \mu \leq \mu_0 \qquad\qquad H_A: \mu > \mu_0$$

Reject H_0 if

$$z_{calculated} > z_\alpha \quad \left(For\ 1, 2, 3 \right)$$

$$t_{calculated} > t_\alpha \quad \left(For\ 4 \right)$$

❖ **Two or both tail test about a population mean** (μ)

$$H_0: \mu = \mu_0 \qquad\qquad H_A: \mu \neq \mu_0$$

Reject H_0 if

$$z_{calculated} > z_{\frac{\alpha}{2}} \quad or \quad z_{calculated} < -z_{\frac{\alpha}{2}}$$

$$t_{calculated} > t_{\frac{\alpha}{2}} \quad or \quad t_{calculated} < -t_{\frac{\alpha}{2}}$$

❖ **Upper confidence limit for lower or left tail test about a population mean** (μ)

$$H_0: \mu \geq \mu_0 \qquad\qquad H_A: \mu < \mu_0$$

Accept H_0 if

$$z_{calculated} > -z_\alpha \quad \left(For\ 1, 2, 3 \right)$$

$$\Rightarrow \frac{\bar{x} - \mu}{\dfrac{\sigma}{\sqrt{n}}} > -z_\alpha \qquad \Rightarrow \bar{x} - \mu > -z_\alpha \frac{\sigma}{\sqrt{n}}$$

$$\Rightarrow -\mu > -\bar{x} - z_\alpha \frac{\sigma}{\sqrt{n}} \qquad \Rightarrow \mu < \bar{x} + z_\alpha \frac{\sigma}{\sqrt{n}}$$

Again, accept H_0 if

$$t_{calculated} > -t_\alpha \quad \left(For\ 4 \right)$$

$$\Rightarrow \frac{\bar{x}-\mu}{\frac{s}{\sqrt{n}}} > -t_\alpha \qquad \Rightarrow \bar{x}-\mu > -t_\alpha \frac{s}{\sqrt{n}}$$

$$\Rightarrow -\mu > -\bar{x} - s_\alpha \frac{s}{\sqrt{n}} \qquad \Rightarrow \mu < \bar{x} + t_\alpha \frac{s}{\sqrt{n}}$$

- **Lower confidence limit for upper or right tail test about a population mean** (μ)

$$H_0: \mu \le \mu_0 \qquad H_A: \mu > \mu_0$$

Accept H_0 if

$$z_{calculated} < z_\alpha \quad \left(For\ 1, 2, 3 \right)$$

$$\Rightarrow \frac{\bar{x}-\mu}{\frac{\sigma}{\sqrt{n}}} < z_\alpha \qquad \Rightarrow \bar{x}-\mu < z_\alpha \frac{\sigma}{\sqrt{n}}$$

$$\Rightarrow -\mu < -\bar{x} + z_\alpha \frac{\sigma}{\sqrt{n}} \qquad \Rightarrow \mu > \bar{x} - z_\alpha \frac{\sigma}{\sqrt{n}}$$

Again, accept H_0 if

$$z_{calculated} < t_\alpha \quad \left(For\ 4 \right)$$

$$\Rightarrow \frac{\bar{x}-\mu}{\frac{s}{\sqrt{n}}} < t_\alpha \qquad \Rightarrow \bar{x}-\mu < t_\alpha \frac{s}{\sqrt{n}}$$

$$\Rightarrow -\mu < -\bar{x} + t_\alpha \frac{s}{\sqrt{n}} \qquad \Rightarrow \mu > \bar{x} - t_\alpha \frac{s}{\sqrt{n}}$$

- **Problem-1**

Assume that in a follow up study involving a sample of 145 Bangladeshi internet users, the sample mean was 10.8 hours per month and the sample standard deviation was 9.2 hours.

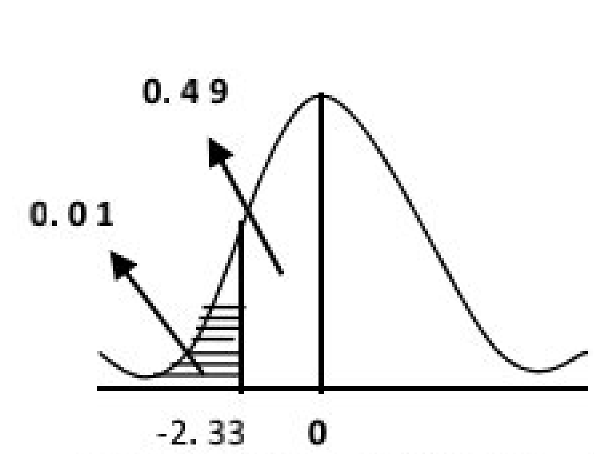
Formulate the null and alternative hypothesis that can be used to determine whether the sample data support the conclusion that Bangladeshi internet users have a population mean less than 13 hours per month. At 1% level of significance what is your conclusion?

Solution:

The null and alternative hypotheses are given as follows:

$$H_0: \mu \ge 13$$

$$H_A: \mu < 13$$



Here, we have that

$\alpha = 0.01$, $\bar{x} = 10.8$ **Hours per month**

$n = 145$ $\left(\textbf{Large}\right)$, $s = 9.2$ **Hours** $\left(\sigma \textbf{ Unknown}\right)$

So, the critical value or the tabulated value is given by:

$$Z_{tabulated} = Z_{\alpha} = Z_{0.01} = \pm 2.33$$

That is, the null hypothesis will be rejected if

$Z_{calculated} < -2.33$ or $Z_{calculated} > 2.33$

Under the H_0, the test statistic is given by:

$$Z_{\text{calculated}} = \frac{\bar{x} - \mu}{\frac{s}{\sqrt{n}}} = \frac{10.8 - 13}{\frac{9.2}{\sqrt{145}}} = -2.88$$

Comment: Since $Z_{\text{calculated}} < -2.33$. So, the null hypothesis is rejected. That means, the sample data support the conclusion that Bangladeshi internet users have a population mean less than 13 hours per month.

Or for $Z_{calculted} = -2.88$ the p-value is 0.002. Since p-value is less than 0.01, so the null hypothesis is rejected. That means, the sample data support the conclusion that Bangladeshi internet users have a population mean less than 13 hours per month.

❖ **Problem-2**

A research company charges to a client based on the assumption that a survey can be completed in a mean time of 15 minutes or less. If a longer mean time is necessary, a premium rate is charged. Suppose a sample of 35 surveys shows a sample mean of 17 minutes and a sample standard deviation of 4 minutes.

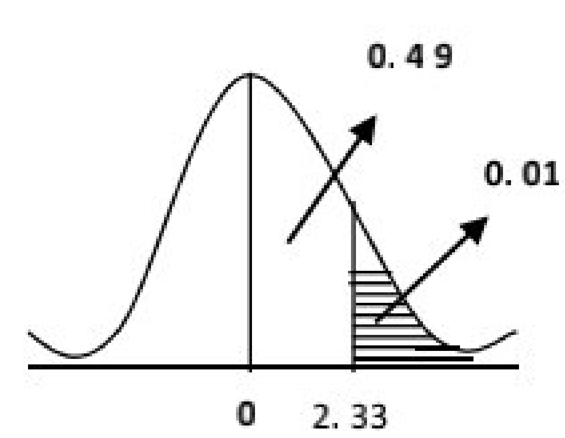
Formulate the null and alternative hypothesis such that the rejection of H_0 will support the charge o f a premium rate. At 1% level of significance what is your conclusion?

Solution:

The null and alternative hypotheses are given as follows:

$H_0: \mu \leq 15$

$H_A: \mu > 15$



Here, we have that

$\alpha = 0.01$, $\bar{x} = 17$ **minutes**

$n = 35$ $\left(\textbf{Large}\right)$, $s = 4$ **minutes** $\left(\sigma \textbf{ Unknown}\right)$

So, the critical value or the tabulated value is given by:

$$Z_{tabulated} = Z_{\alpha} = Z_{0.01} = \pm 2.33$$

That is, the null hypothesis will be rejected if

$Z_{calculated} < -2.33$ or $Z_{calculated} > 2.33$

Under the H_0, the test statistic is given by:

$$Z_{\text{calculated}} = \frac{\bar{x} - \mu}{\frac{s}{\sqrt{n}}} = \frac{17 - 15}{\frac{4}{\sqrt{35}}} = 2.96$$

Comment: Since $Z_{\text{calculated}} > 2.33$. So, the null hypothesis is rejected. That means, the premium rate is charged.

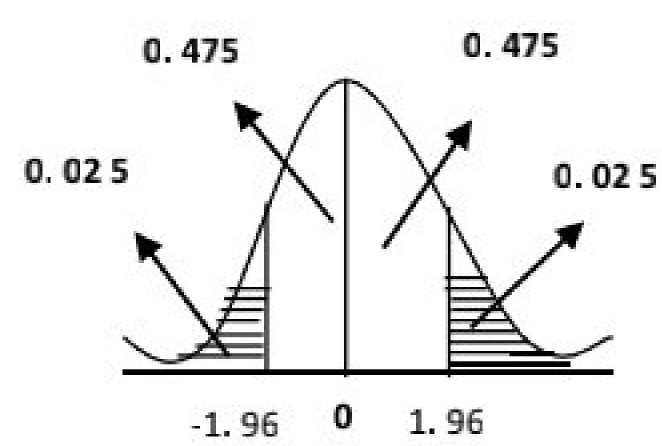
❖ **Problem-3**

A survey found that the mean charitable contribution on the tax returns was 1075 Tk. Assume a sample of October 2001 tax returns will be used to conduct a hypothesis test designed to determine whether any change has occurred in the mean charitable contributions. Assume that a sample of 200 tax returns has a sample mean of 11610 Tk. and a sample standard deviation of 840 Tk.

Formulate the appropriate null and alternative hypothesis and test the hypothesis at 5% level of significance.

Solution:

The null and alternative hypotheses are given as follows:



$$H_0: \mu = 1075$$

$$H_A: \mu \neq 1075$$

Here, we have that

$\alpha = 0.05$, $\bar{x} = 1160$ **Tk.**

$n = 200$ (**Large**) , $s = 840$ **Tk.** (σ **Unknown**)

So, the critical value or the tabulated value is given by:

$$Z_{\text{tabulated}} = Z_{\frac{\alpha}{2}} = Z_{0.025} = \pm 1.96$$

That is, the null hypothesis will be rejected if

$$Z_{\text{calculated}} > 1.96 \quad or \quad Z_{\text{calculated}} < -1.96$$

Under the H_0, the test statistic is given by:

$$Z_{\text{calculated}} = \frac{\bar{x} - \mu}{\frac{s}{\sqrt{n}}} = \frac{1160 - 1075}{\frac{840}{\sqrt{200}}} = 1.43$$

Comment: Since $-1.96 < Z_{\text{calculated}} < 1.96$. So, the null hypothesis is not rejected. That means, no change has occurred in the mean charitable contributions.

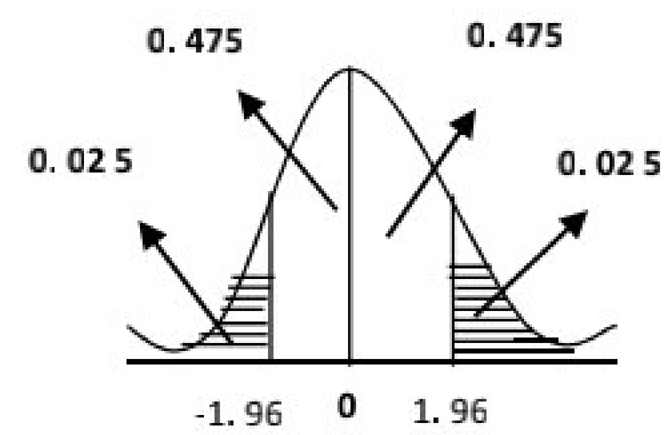
❖ Problem-4

A survey reported that the mean wage of the population of Rampura is 26,133 Tk. per month. A sample of 550 people showed a sample mean wage of 25,457 Tk. per month and a sample standard deviation of 7,600 Tk.

Formulate the appropriate null and alternative hypothesis that can be used to determine whether the sample data support the conclusion that the population mean wage per month differs from the mean wage of 26,133 Tk. per month. At 5% level of significance what is your conclusion?

Solution:

The null and alternative hypotheses are given as follows:



$$H_0 : \mu = 26{,}133$$

$$H_A : \mu \neq 26{,}133$$

Here, we have that

$\alpha = 0.05$, $\bar{x} = 25{,}457$ **Tk.**

$n = 550 \left(\textbf{Large}\right)$, $s = 7{,}600$ **Tk.** $\left(\sigma \textbf{ Unknown}\right)$

So, the critical value or the tabulated value is given by:

$$Z_{\text{tabulated}} = Z_{\frac{\alpha}{2}} = Z_{0.025} = \pm 1.96$$

That is, the null hypothesis will be rejected if

$$Z_{\text{calculated}} > 1.96 \quad \textit{or} \quad Z_{\text{calculated}} < -1.96$$

Under the H_0, the test statistic is given by:

$$Z_{\text{calculated}} = \frac{\bar{x} - \mu}{\frac{s}{\sqrt{n}}} = \frac{25.457 - 26{,}133}{\frac{7{,}600}{\sqrt{550}}} = -2.09$$

Comment: Since $Z_{\text{calculated}} < -1.96$. So, the null hypothesis is rejected. That means, the sample data support the conclusion that the population mean wage per month differs from the mean wage of 26,133 Tk. per month.

❖ Problem-5

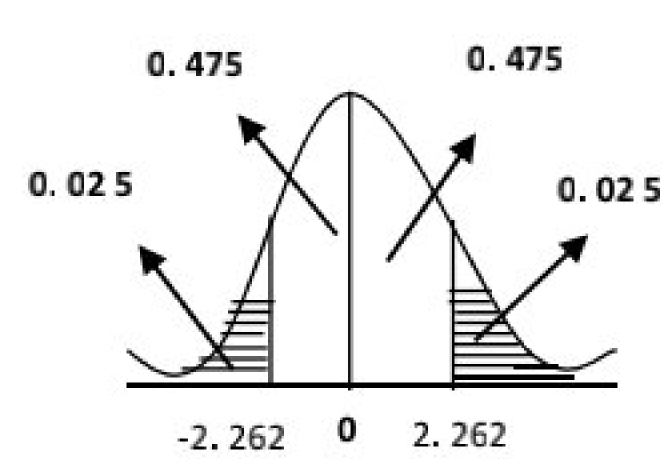
A sample of 10 financial services corporations provided the following earnings per share data:

1.92 , 2.16 , 3.63 , 3.16 , 4.02 , 3.14 , 2.20 , 2.34 , 3.05 , 2.38

Formulate the null and alternative hypothesis that can be used to determine whether the populations mean earnings per share differ from TK.3. At 5% level of significance, what is your conclusion?

Solution:

The null and alternative hypotheses are given as follows:



$$H_0 : \mu = 3$$

$$H_A : \mu \neq 3$$

Here, we have that

$$\alpha = 0.05 \quad , \quad \bar{x} = 2.80 \ \textbf{Tk.}$$

$$n = 10 \ (\textbf{Small}) \quad , \quad s = 0.70 \ \textbf{Tk.} \ (\sigma \ \textbf{Unknown})$$

So, the critical value or the tabulated value is given by:

$$t_{\text{tabulated}}(9) = t_{\frac{\alpha}{2}}(9) = t_{0.025}(9) = \pm 2.262$$

That is, the null hypothesis will be rejected if

$$t_{\text{calculated}} > 2.262 \quad or \quad t_{\text{calculated}} < -2.262$$

Under the H_0, the test statistic is given by:

$$t_{\text{calculated}} = \frac{\bar{x} - \mu}{\frac{s}{\sqrt{n}}} = \frac{2.80 - 3}{\frac{0.70}{\sqrt{10}}} = -0.90$$

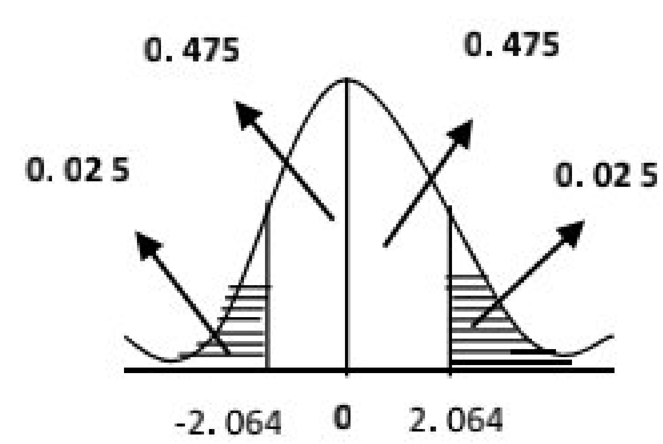
Comment: Since $-2.262 < t_{\text{calculated}} < 2.262$. So, the null hypothesis is not rejected. That means, the populations mean earnings per share does not differ from TK.3.

❖ Problem-6

Assume a sample of 25 households showed a sample mean daily expenditure of 84.50 Tk. with a sample standard deviation of 14.50 Tk. At 5% level of significance, test $H_0: \mu = 90$ *vs* $H_A: \mu \neq 90$. What is your conclusion?

Solution:

The null and alternative hypotheses are given as follows:



$$H_0: \mu = 90$$

$$H_A: \mu \neq 90$$

Here, we have that

$$\alpha = 0.05 \quad , \quad \bar{x} = 84.5 \ \textbf{Tk.}$$

$$n = 25 \ (\textbf{Small}) \quad , \quad s = 14.5 \ \textbf{Tk.} \ (\sigma \ \textbf{Unknown})$$

So, the critical value or the tabulated value is given by:

$$t_{\text{tabulated}}(24) = t_{\frac{\alpha}{2}}(24) = t_{0.025}(24) = \pm 2.064$$

That is, the null hypothesis will be rejected if

$$t_{\text{calculated}} > 2.064 \quad or \quad t_{\text{calculated}} < -2.064$$

Under the H_0, the test statistic is given by:

$$t_{\text{calculated}} = \frac{\bar{x} - \mu}{\frac{s}{\sqrt{n}}} = \frac{84.5 - 90}{\frac{14.5}{\sqrt{25}}} = -1.90$$

Comment: Since $-2.064 < t_{\text{calculated}} < 2.064$. So, the null hypothesis is not rejected.

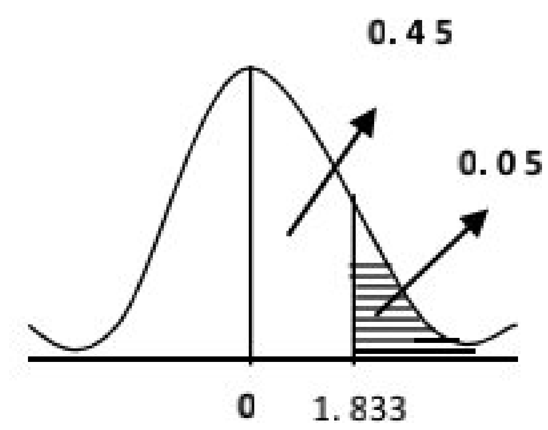
❖ **Problem-7**

For cost estimating purposes, managers use 2 hours of labor time for the planting of a medium size. Actual times (in hours) from a sample of 10 planting during the past month are as follows:

1.9 , 1.7 , 2.8 , 2.4 , 2.6 , 2.5 , 2.8 , 3.2 , 1.6 , 2.5

Formulate the null and alternative hypothesis that can be used to determine whether the mean tree-planting time exceeds 2 hours. At 5% level of significance, what is your conclusion?

Solution:



The null and alternative hypotheses are given as follows:

$$H_0: \mu \leq 2$$

$$H_A: \mu > 2$$

Here, we have that

$\alpha = 0.05$, $\bar{x} = 2.4$ **hours**

$n = 10$ (**Small**) , $s = 0.52$ **hours** (σ **Unknown**)

So, the critical value or the tabulated value is given by:

$$t_{\text{tabulated}}(9) = t_{\alpha}(9) = t_{0.05}(9) = 1.833$$

That is, the null hypothesis will be rejected if

$$t_{\text{calculated}} > 1.833$$

Under the H_0, the test statistic is given by:

$$t_{\text{calculated}} = \frac{\bar{x} - \mu}{\frac{s}{\sqrt{n}}} = \frac{2.4 - 2}{\frac{0.52}{\sqrt{10}}} = 2.43$$

Comment: Since $t_{\text{calculated}} > 1.833$. So, the null hypothesis is rejected. That means, the mean tree-planting time exceeds 2 hours.

❖ **Test of Equality of Two Means**

Let $x_{11}, x_{12}, \cdots\cdots\cdots, x_{1n_1}$ be n_1 sample observations drawn independently from a normal population $x_1 \sim N\ (\mu_1, {\sigma_1}^2)$. Another independent sample observations are $x_{21}, x_{22}, \cdots\cdots\cdots, x_{2n_2}$ which are drawn from a population $x_2 \sim N\ (\mu_2, {\sigma_2}^2)$. The objective is to test the hypothesis

$$H_0: \mu_1 = \mu_2$$

$$H_1: \mu_1 \neq \mu_2$$

Assumptions:

(i) Two samples are independent.
(ii) ${\sigma_1}^2$ and ${\sigma_2}^2$ are known values ; n_1 and n_2 may be small or large
(iii) ${\sigma_1}^2 = {\sigma_2}^2 = \sigma^2$ are unknown values ; n_1 and n_2 are small (<30)
(iv) ${\sigma_1}^2$ and ${\sigma_2}^2$ are unknown values ; both n_1 and n_2 are large (≥ 30)

The test statistic to test the significance of null hypothesis is

$$z = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{\frac{{\sigma_1}^2}{n_1} + \frac{{\sigma_2}^2}{n_2}}} \sim N(0,1)$$, under assumption (ii)

The test statistic is

$$t = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{s^2(\frac{1}{n_1} + \frac{1}{n_2})}} \sim N(0,1) \text{ , under assumption (iii)}$$

This 't' is distributed as Student's 't' with $(n_1 + n_2 - 2)$ d.f.

The test statistic is

$$z = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}} \sim N(0,1) \text{ , under assumption (iv)}$$

Where, $\bar{x}_1 = \frac{\sum x_{1i}}{n_1}, \bar{x}_2 = \frac{\sum x_{2i}}{n_2}, s_1^2 = \frac{1}{n_1 - 1}\sum(x_{1i} - \bar{x}_1)^2, s_2^2 = \frac{1}{n_2 - 1}\sum(x_{2i} - \bar{x}_2)^2$

and $s^2 = \frac{(n_1-1)s_1^2 + (n_2-1)s_2^2}{n_1 + n_2 - 2}$

Decision: If the calculated value of test statistic is greater than the critical value of test statistic then the null hypothesis (H_0) is rejected otherwise it is accepted.

❖ **Problem-8**

The number of computer scientists coming out from two different universities A and B are employed in different organizatins to do job related to computer. The numbers are given for different years as follows

University	Number of graduates employed in computer related job
A	18, 16, 15, 20, 18, 15, 12
B	20, 14, 12, 22, 16, 14, 15, 20, 12, 18, 10

Do you think that employment facility for the both universities are similar?

Solution:

Let x_1 and x_2 be the number of graduates of university A and B respectively. Assume that $x_1 \sim N\ (\mu_1, \sigma_1^2)$ and $x_2 \sim N\ (\mu_2, \sigma_2^2)$. Also assume that $\sigma_1^2 = \sigma_2^2 = \sigma^2$. We need to test the hypothesis

$$H_0: \mu_1 = \mu_2$$

$$H_1: \mu_1 \neq \mu_2$$

Since $n_1 = 7\ (< 30)$ and $n_2 = 11\ (< 30)$ and σ^2 is not known, the test statistic is

$$t = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{s^2(\frac{1}{n_1} + \frac{1}{n_2})}}$$

Where,

$$\bar{x}_1 = \frac{\sum x_{1i}}{n_1} = \frac{114}{7} = 16.29\ ,\ \bar{x}_2 = \frac{\sum x_{2i}}{n_2} = \frac{163}{11} = 14.28,$$

$$s_1^2 = \frac{1}{n_1-1}\sum(x_{1i} - \bar{x}_1)^2 = \frac{1}{n_1-1}\left[\sum x_{1i}^2 - \frac{(\sum x_{1i})^2}{n_1}\right] = \frac{1}{7-1}\left[1898 - \frac{(114)^2}{7}\right] = 6.905$$

$$s_2^2 = \frac{1}{n_2-1}\sum(x_{2i} - \bar{x}_2)^2 = \frac{1}{n_2-1}\left[\sum x_{2i}^2 - \frac{(\sum x_{2i})^2}{n_2}\right] = \frac{1}{11-1}\left[2569 - \frac{(163)^2}{7}\right] = 15.364$$

and $s^2 = \frac{(n_1-1)s_1^2+(n_2-1)s_2^2}{n_1+n_2-2} = \frac{(7-1)\times 6.905+(11-1)\times 15.364}{7+11-} = 12.192$

The statistic 't' is distributed as Student's 't' with $(n_1 + n_2 - 2) = (7 + 11 - 2) = 16$ d.f.

Here

$$t = \frac{16.29 - 14.28}{\sqrt{12.192 \times (\frac{1}{7} + \frac{1}{11})}} = \frac{1.47}{1.688} = 0.67$$

At 5% level of significance the critical value of t with 16 d.f. is $t_{0.025,16} = 2.12$.

Comment: Since $|t| = 0.67 < 2.12$ therefore we can say that H_0 is accepted i.e. the employment facility for the students of both universities is similar.

- **Problem-9**

The daily temperature (in degree Celsius) of two months during summer season are shown below:

Months	Daily temperature (in degree Celsius)
1	32, 34, 31, 33, 35, 36, 34, 34, 34, 35, 32, 33, 33, 33, 32, 32, 34, 33, 32, 34, 32, 31, 33, 34, 35, 34, 33, 33, 33, 34, 34
2	34, 34, 35, 35, 35, 35, 35, 35, 36, 37, 34, 33, 34, 35, 34, 34, 36, 34, 33, 34, 32, 33, 34, 36, 35, 35, 35, 34, 35, 34

Do you think that the temperature of both months are similar?

Solution:

Let x_1 and x_2 be the temperature of month-1 and month-2 respectively. Assume that $x_1 \sim N\ (\mu_1, \sigma_1^2)$ and $x_2 \sim N\ (\mu_2, \sigma_2^2)$.

We need to test the hypothesis

$$H_0: \mu_1 = \mu_2$$

$$H_1: \mu_1 \neq \mu_2$$

Since $n_1 = 31\ (> 30)$ and $n_2 = 30\ (\geq 30)$ and σ^2 is not known, the test statistic is

$$z = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}}} \sim N(0,1)$$

Here,

$\bar{x}_1 = \frac{\sum x_{1i}}{n_1} = \frac{1032}{31} = 33.29$, $\bar{x}_2 = \frac{\sum x_{2i}}{n_2} = \frac{1035}{30} = 34.5$,

$$s_1^2 = \frac{1}{n_1-1}\sum(x_{1i} - \bar{x}_1)^2 = \frac{1}{n_1-1}\left[\sum x_{1i}^2 - \frac{(\sum x_{1i})^2}{n_1}\right] = \frac{1}{31-1}\left[34398 - \frac{(1032)^2}{31}\right] = 1.41$$

$$s_2^2 = \frac{1}{n_2-1}\sum(x_{2i} - \bar{x}_2)^2 = \frac{1}{n_2-1}\left[\sum x_{2i}^2 - \frac{(\sum x_{2i})^2}{n_2}\right] = \frac{1}{30-1}\left[35739 - \frac{(1035)^2}{30}\right] = 1.09$$

$$\therefore z = \frac{33.29 - 34.5}{\sqrt{\frac{1.41}{31} + \frac{1.09}{30}}} = -4.23$$

Comment: Since $|z| > 1.96$ therefore we can say that H_0 is rejected at 5% level of significance i.e. the temperatures in two months are not similar.

- **Problem-10**

The average salaries of 15 female workers of industry-1 is 1125.00 taka and standard deviation is 75.00 taka. The average salaries of 20 female workers of industry-2 is 1325.00 taka and standard deviation is 225.00 taka. Are the female workers of two different industries similarly paid?

Solution:

Let x_1 and x_2 be the monthly salaries of female workers in industry-1 and industry-2 respectively. Assume that $x_1 \sim N\ (\mu_1, {\sigma_1}^2)$ and $x_2 \sim N\ (\mu_2, {\sigma_2}^2)$. We need to test the hypothesis

$$H_0: \mu_1 = \mu_2$$

$$H_1: \mu_1 \neq \mu_2$$

The test statistic is

$$t = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{s^2(\frac{1}{n_1} + \frac{1}{n_2})}}$$

Where,

$\bar{x}_1 = 1125.00\,,\ \bar{x}_2 = 1325.00,\ {s_1}^2 = (75.00)^2 = 5625.00,\ {s_2}^2 = (225.00)^2 = 50625.00$

and $s^2 = \frac{(n_1-1){s_1}^2+(n_2-1){s_2}^2}{n_1+n_2-2} = \frac{(15-1)\times 5625.00\ \ (20-1)\times 50625.00}{15+20} = 31534.09$

The statistic 't' is distributed as Student's 't' with $(n_1 + n_2 - 2) = (15 + 20 - 2) = 33$ d.f.

$$t = \frac{1125.00 - 1325.00}{\sqrt{31534.09 \times (\frac{1}{15} + \frac{1}{20})}} = \frac{-200}{60.65} = -3.30$$

The critical value of t at 5% level of significance with 33 d.f. is $t_{0.025,33} = 2.035$.

Comment: Since $|t| > 2.035$ therefore we can say that H_0 is rejected. The salaries of female workers in two industries are not similar.

- **Test of Equality of Two Correlated Means**

Let $x_1, x_2, \cdots\cdots, x_n$ be n observations recorded from n sample points in an occasion. Let $y_1, y_2, \cdots\cdots, y_n$ be another n observations recorded from n sample points in a different occasion. Here it is assumed that $(x_i, y_i; i = 1,2, \dots \dots, n)$ is the i-th pair of observation from i-th sample point. Assume that $x_i \sim N\ (\mu_1, \sigma^2)$ and $y_i \sim N\ (\mu_2, \sigma^2)$. The problem is to set the hypothesis

$$H_0: \mu_1 = \mu_2$$
$$H_1: \mu_1 \neq \mu_2$$

The hypothesis implies that $H_0\colon\ \mu_1 - \mu_2 = 0$. To test the hypothesis, we can consider the differences of pairs of observations, where the difference of ith pair of observation is

$$d_i = x_i - y_i; i = 1,2, \dots \dots, n$$

The mean and variance of d_i is

$$\bar{d} = \frac{\sum d_i}{n} \text{ and } s_d{}^2 = \frac{1}{n-1}\sum(d_i - \bar{d})^2$$

The test statistic for the hypothesis is

$$t = \frac{\bar{d}}{s_d/\sqrt{n}}$$

This 't' statistic follows Student's 't' distribution with (n-1) d.f.

Comment: If the calculated value of test statistic (t) is greater than the critical value of test statistic then the null hypothesis (H_0) is rejected otherwise it is accepted.

- **Problem-11**

The following data represent the blood sugar of a group of patients before (B) and after (A) of a specific treatment. Is the treatment successful?

Blood sugar [(B), x]	142, 146, 156, 120, 138, 160, 150, 155, 180, 210
Blood sugar [(A), y]	120, 130, 120, 115, 100, 160, 150, 138, 160, 180

Solution:

Let $x_i \sim N\,(\mu_1, \sigma^2)$ and $y_i \sim N\,(\mu_2, \sigma^2)$. We need to test the hypothesis

$$H_0: \mu_1 = \mu_2$$

$$H_1: \mu_1 > \mu_2$$

Here, $d_i = x_i - y_i$; 22, 16, 36, 5, 38, 0, 0, 17, 20, 30

$$\bar{d} = \frac{\sum d_i}{n} = \frac{184}{10} = 18.4$$

and $s_d{}^2 = \frac{1}{n-1}\sum(d_i - \bar{d})^2 = \frac{1}{n-1}\left[\sum d_i{}^2 - \frac{(\sum d_i)^2}{n}\right] = \frac{1}{10-1} \times \left[5094 - \frac{(184)^2}{10}\right] = 189.82$

$$s_d = \sqrt{189.82} = 13.78$$

The test statistic is

$$t = \frac{\bar{d}}{s_d/\sqrt{n}} = \frac{18.4}{13.78/\sqrt{10}} = 4.22$$

The critical value of test statistic at 5% level significance with 9 d.f. is $t_{0.05,9} = 1.83$.

Comment: Since $|t| > 1.83$ therefore we can say that H_0 is rejected. The treatment is not effective.

- **Test Regarding Proportion**

If the variable under study is qualitative in nature, the parameter equivalent to the mean of the variable is population proportion. For example, let $X_1, X_2, \cdots\cdots, X_N$ be the values of the variable X observed in N population units, where $X_i = 1$ if ith unit possesses some characteristic under study otherwise $X_i = 0$. (The characteristics are, for example, residential status, occupation, color of hair, family planning adoption behavior etc.)

Let,

$$\sum_{i=1}^{N} X_i = A = A \ of \ the \ units \ possesses \ a \ particular \ characteristic.$$

Then, $P = \frac{1}{N}\sum_{i=1}^{N} X_i = \frac{A}{N} = Proportion \ of \ the \ units \ possesses \ a \ characteristic$

Again let,

$$\sum_{i=1}^{n} x_i = a = a \ of \ the \ sample \ units \ possesses \ a \ particular \ characteristic.$$

Then

$p = \frac{1}{n}\sum_{i=1}^{n} x_i = \frac{a}{n} = Sample \ proportion \ of \ the \ units \ possesses \ a \ characteristic$

Where n is the sample size.

The problem is to test the hypothesis

$$H_0: P = P_0 \text{ (a known value)}$$

$$H_1: P \neq P_0$$

The test statistic is

$$Z = \frac{p - P_0}{\sqrt{\frac{P_0 Q_0}{n}}} \sim N(0,1) \text{ , where } Q_0 = 1 - P_0$$

Comment: If $|Z| > 1.96$ then H_0 is rejected otherwise it is accepted at 5% level of significance. Sometimes two independent samples are drawn from two populations, where $P_1 \ and \ P_2$ are the proportions of individuals possessing the characteristic under study in population-1 and population-2 respectively.

Where,

$P_1 = \frac{1}{N_1}\sum_{i=1}^{N_1} X_{1i} = \frac{A_1}{N_1}$ and $P_2 = \frac{1}{N_2}\sum_{i=1}^{N_2} X_{2i} = \frac{A_2}{N_2}$

The corresponding sample proportion of $P_1 \ and \ P_2$ are

$p_1 = \frac{1}{n_1}\sum_{i=1}^{n_1} x_{1i} = \frac{a_1}{n_1}$ and $p_2 = \frac{1}{n_2}\sum_{i=1}^{n_2} x_{2i} = \frac{a_2}{n_2}$

The problem is to test the hypothesis

$$H_0: P_1 = P_2$$

$$H_1: P_1 \neq P_2$$

The test statistic is

$$Z = \frac{p_1 - p_2}{\sqrt{PQ(\frac{1}{n_1} + \frac{1}{n_2})}} \sim N(0,1) \text{ , where } P = \frac{a_1 + a_2}{n_1 + n_2} \text{ and } Q = 1 - P$$

Comment: If $|Z| > 1.96$ then H_0 is rejected otherwise it is accepted at 5% level of significance.

- **Problem-12**

A sample of 15 students are selected from a group of 100 students and their grade in S.S.C examination is recorded as follows:

Students	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
Grade	B	C	A	D	B	C	D	A	B	C	D	B	C	C	D

Do you think that 10% students get grade A?

Solution:

Since we are interested in studying the characteristic grade A, the students who got grade A, for them let us assign value 1, otherwise 0. Then we have $n = 15, a = 2$ and $= \frac{a}{n} = \frac{2}{15} = 0.13$, $P_0 = 0.10$ and $Q_0 = 1 - 0.10 = 0.90$

We need to test the hypothesis

$$H_0: P = P_0$$

$$H_1: P \neq P_0$$

The test statistic is

$$Z = \frac{p - P_0}{\sqrt{\frac{P_0 Q_0}{n}}} = \frac{0.13 - 0.10}{\sqrt{\frac{0.10 \times 0.90}{15}}} = 0.39$$

Comment: Since $|Z| < 1.96$ therefore we can say that H_0 is accepted i.e. it can be considered that 10% students got grade A.

- **Problem-13**

Five percent workers of an industry are usually injured every year. In a year 50 workers are investigated and found that 10 of them are injured during work in the industry. Does the sample provide similar information about the overall proportion of injured workers?

Solution:

Here given that

$$n = 50, a = 10, p = \frac{10}{50} = 0.20, P_0 = 0.05, Q_0 = 1 - 0.05 = 0.95$$

We need to test the hypothesis

$$H_0: P = P_0$$

$H_1: P \neq P_0$The test statistic is

$$Z = \frac{p - P_0}{\sqrt{\frac{P_0 Q_0}{n}}} = \frac{0.20 - 0.05}{\sqrt{\frac{0.05 \times 0.95}{50}}} = 4.87$$

Comment: Since $|Z| > 1.96$ therefore we can say that H_0 is rejected at 5% level of significance i.e. the sample proportion of injured workers is not similar to assumed proportion of injured workers.

❖ **Problem-14**

In an industry 100 workers are working, 25 of them are skilled. In another industry there are 18 skilled workers out of 125 workers. Are the skilled workers similar in both industries?

Solution:

We need to test the hypothesis

$$H_0: P_1 = P_2$$

$$H_1: P_1 \neq P_2$$

The test statistic is

$$Z = \frac{p_1 - p_2}{\sqrt{PQ(\frac{1}{n_1} + \frac{1}{n_2})}}$$

$$= \frac{0.25 - 0.144}{\sqrt{0.19 \times 0.81 \times (\frac{1}{100} + \frac{1}{125})}} = 2.01$$

Here,

$$p_1 = \frac{a_1}{n_1} = \frac{25}{100} = 0.25$$

$$p_2 = \frac{a_2}{n_2} = \frac{18}{125} = 0.144$$

$$P = \frac{a_1 + a_2}{n_1 + n_2} = \frac{25 + 18}{100 + 125} = 0.19$$

and $Q = 1 - P = 1 - 0.19 = 0.81$

Comment: Since $|Z| > 1.96$ therefore we can say that H_0 is rejected at 5% level of significance i.e. the proportion of skilled workers in two industries are not similar.

❖ **Problem-15**

Among 100 literate couples 60 adopted family planning and among 500 illiterate couples 100 adopted family planning. Are the proportions of literate and illiterate couples similar?

Solution:

We need to test the hypothesis

$$H_0: P_1 = P_2$$

$$H_1: P_1 \neq P_2$$

The test statistic is

$$Z = \frac{p_1 - p_2}{\sqrt{PQ(\frac{1}{n_1} + \frac{1}{n_2})}}$$

$$= \frac{0.60 - 0.25}{\sqrt{0.31 \times 0.69 \times (\frac{1}{100} + \frac{1}{500})}} = 6.91$$

Here,

$$p_1 = \frac{a_1}{n_1} = \frac{60}{100} = 0.60$$

$$p_2 = \frac{a_2}{n_2} = \frac{125}{500} = 0.25$$

$$P = \frac{a_1 + a_2}{n_1 + n_2} = \frac{60 + 125}{100 + 500} = 0.31$$

and $Q = 1 - P = 1 - 0.31 = 0.69$

Comment: Since $|Z| > 1.96$ therefore we can say that H_0 is rejected at 5% level of significance i.e. the proportion of adopted couples among literate and illiterate couples are not similar.

❖ **Test Regarding Variance**

Let $x_1, x_2, \cdots\cdots\cdots, x_n$ be a random sample drawn from $N(\mu, \sigma^2)$, where μ is unknown.

The objective is to test the hypothesis is

$$H_0: \sigma^2 = \sigma_0^2 (\text{say})$$

(i) $H_1: \sigma^2 \neq \sigma_0^2$ Or (ii) $H_1: \sigma^2 > \sigma_0^2$ or (iii) $H_1: \sigma^2 < \sigma_0^2$

Since μ is unknown, it is estimated by $\bar{x} = \frac{\sum x_i}{n}$ and the estimate of σ^2 is $s^2 = \frac{1}{n-1}\sum(x_i - \bar{x})^2$.

Then the test statistic is

$$\chi^2 = \frac{(n-1)s^2}{\sigma_0^2}$$

The χ^2 is distributed as chi-square distribution with $(n-1)$ d.f.

Comment:

For alternative hypothesis (i) $H_1: \sigma^2 \neq \sigma_0^2$, it is two sided alternative hypothesis.

Hence, if ${\chi^2}_{cal} \le {\chi^2}_{1-\frac{\alpha}{2}, n-1}$ or ${\chi^2}_{cal} \ge {\chi^2}_{\frac{\alpha}{2}, n-1}$ then H_0 is rejected.

For alternative hypothesis (ii) $H_1: \sigma^2 > \sigma_0^2$, it is a right sided test and the conclusion is made as follows

If ${\chi^2}_{cal} \ge {\chi^2}_{\alpha, n-1}$ then H_0 is rejected.

For alternative hypothesis (ii) $H_1: \sigma^2 < \sigma_0^2$, it is a left sided test and the conclusion is made as follows

If ${\chi^2}_{cal} \le {\chi^2}_{1-\alpha, n-1}$ then H_0 is rejected.

❖ **Test of Equality of Two Variances**

Let $x_{11}, x_{12}, \cdots\cdots\cdots, x_{1n_1}$ be n_1 sample observations drawn independently from a normal population $x_1 \sim N(\mu_1, \sigma_1^2)$. Another independent sample observations are $x_{21}, x_{22}, \cdots\cdots\cdots, x_{2n_2}$ which are drawn from a population $x_2 \sim N(\mu_2, \sigma_2^2)$. Where μ_1 and μ_2 are unknown. The estimate of μ_1 is $\bar{x}_1 = \frac{\sum x_{1i}}{n_1}$ and the estimate of μ_2 is $\bar{x}_2 = \frac{\sum x_{2i}}{n_2}$.

The objective is to test the hypothesis

$$H_0: \sigma_1^2 = \sigma_2^2$$
$$H_1: \sigma_1^2 \neq \sigma_2^2$$

The test statistic is

$$F = \frac{s_1^2}{s_2^2}\ ;\ s_1^2 < s_2^2$$

Where, $s_1^2 = \frac{1}{n_1 - 1}\sum(x_{1i} - \bar{x}_1)^2$, $s_2^2 = \frac{1}{n_2 - 1}\sum(x_{2i} - \bar{x}_2)^2$

This F is distributed as variance ratio with $(n_1 - 1)$ and $(n_2 - 1)$ d.f.

Comment: If $F_{cal} \ge F_{0.05;\, n_1-1, n_2-1}$ then H_0 is rejected at 5% level of significance otherwise it is accepted.

❖ **Problem-16**

The following observations represent the systolic blood pressure (x, mm of Hg) of some patients:

$$x: 70, 85, 92, 90, 95, 79, 80, 85, 90, 85, 95$$

Do you think that the sample is drawn from a population $N(\mu, 50)$.

Solution:

We need to test the hypothesis

$$H_0: \sigma^2 = \sigma_0{}^2 = 50$$

$$H_1: \sigma^2 > \sigma_0{}^2$$

$\bar{x} = \frac{\sum x_i}{n} = \frac{946}{11} = 86$ and

$s^2 = \frac{1}{n-1}\sum(x_i - \bar{x})^2 = \frac{1}{n-1}\left[\sum x_i{}^2 - \frac{(\sum x_i)^2}{n}\right] = \frac{1}{11-1}\left[81930 - \frac{(946)^2}{11}\right] = 57.4$

The test statistic is

$$\chi^2 = \frac{(n-1)s^2}{\sigma_0{}^2} = \frac{(11-1) \times 57.4}{50} = 11.48$$

The χ^2 is distributed as chi-square distribution with $(n-1) = (11-1) = 10$ d.f.

The critical value of χ^2 at 5% level of significance with 10 d.f. is $\chi^2{}_{0.05,10} = 18.307$.

Comment: Since calculated value of $\chi^2 < 18.307$ therefore we can say that H_0 is accepted.

❖ **Problem-17**

The number of telephone calls received by the emergency word during office hours in different days are as follows:

160, 172, 121, 144, 100, 108, 175, 200, 105, 95,102

Do you think that the sample follows $N(\mu, 1500)$.

Solution:

We need to test the hypothesis

$$H_0: \sigma^2 = \sigma_0{}^2 = 1500$$

$$H_1: \sigma^2 < \sigma_0{}^2$$

$\bar{x} = \frac{\sum x_i}{n} = \frac{1482}{11} = 134.73$ and

$s^2 = \frac{1}{n-1}\sum(x_i - \bar{x})^2 = \frac{1}{n-1}\left[\sum x_i{}^2 - \frac{(\sum x_i)^2}{n}\right] = \frac{1}{11-1}\left[231305 - \frac{(1482)^2}{11}\right] = 1363.82$

The test statistic is

$$\chi^2 = \frac{(n-1)s^2}{\sigma_0{}^2} = \frac{(11-1) \times 1363.82}{1500} = 9.09$$

The χ^2 is distributed as chi-square distribution with $(n-1) = (11-1) = 10$ d.f.

The critical value of χ^2at 5% level of significance with 10 d.f. is ${\chi^2}_{0.95,10} = 3.94$.

Comment: Since calculated value of $\chi^2 > 3.94$ therefore we can say that H_0 is rejected.

- **Problem-18**

The number of students admitted in two private universities in different years are as follows:

University-1; x_{1i}	155	165	170	190	220	250	250	
University-2; x_{2i}	300	355	360	360	360	400	400	400

Are the variation in admission of students in different years in two universities are same?

Solution:

Let $x_{1i} \sim N\ (\mu_1, {\sigma_1}^2)$ and $x_{2i} \sim N\ (\mu_2, {\sigma_2}^2)$.

We need to test the hypothesis

$$H_0: {\sigma_1}^2 = {\sigma_2}^2$$

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

We have

$${s_1}^2 = \frac{1}{n_1-1}\sum(x_{1i}-\bar{x}_1)^2 = \frac{1}{n_1-1}[\sum {x_{1i}}^2 - \frac{(\sum x_{1i})^2}{n_1}] = \frac{1}{7-1}[289650 - \frac{(1400)^2}{7}]$$

$$= 1608.33$$

$${s_2}^2 = \frac{1}{n_2-1}\Sigma(x_{2i}-\bar{x}_2)^2 = \frac{1}{n_2-1}[\Sigma {x_{2i}}^2 - \frac{(\Sigma x_{2i})^2}{n_2}] = \frac{1}{8-1}[1084825 - \frac{(2935)^2}{7}] = 1149.55$$

The test statistic is

$$F = \frac{{s_2}^2}{{s_1}^2} = \frac{1149.55}{1608.33} = 0.71$$

The critical value of F at 5% level of significance is $F_{0.05,7,6} = 3.51$

Comment: Since calculated value of $F < 3.51$ therefore we can say tht H_0 is accepted.