7

Test of Hypothesis (1)

❖ Statistical Hypothesis

Mitra

A statistical hypothesis is some statement or assertion about a population or equivalently about the probability distribution characterizing a population which we want to verify on the basis of information available from a sample.

Example: A few examples of statistical hypothesis that relate to our real life are as follows:

- A physiscian may hypothesize that the recommended drug is effective in 90 percent cases.
- A nutritionist claims that at most 75 percent of the per school children in a certain country have protein deficient diets.
- An administrator of business farm claims that the average work efficiency of his workers is at least 90 percent.
- A sewing machine company claims that their new machine is superior to the one availabe in the market.
- The court assumes that the indicated person is innocent.

❖ Null Hypothesis

According to Prof. R. A. Fisher, the hypothesis which we are going to test for possible rejection under the assumption that it is true is called the null hypothesis. Usually it is denoted by H_0.

Example: If $X \sim N(\mu, \sigma^2)$, where μ and σ^2 are parameters. Let us consider μ has a specified value μ_0(say). Then the null hypothesis is $H_0: \mu = \mu_0$

Some examples of null hypothesis relate to real life are as follows:

- There is no difference in the incidence of malnutrition between vaccinated and non-vaccinated.
- Males do not smoke more than females.
- There is no association between level of education and knowledge of child nutrition among women.
- Two teaching methods A and B are equally effective.

❖ Alternative Hypothesis

Any hypothesis which is complementary to the null hypothesis is called an alternative hypothesis. Usually it is denoted by H_1.

Example: If we want to test the null hypothesis that the population has a specified mean μ_0 (say) i.e. $H_0: \mu = \mu_0$ then the alternative hypothesis could be $H_1: \mu \neq \mu_0$ or $H_1: \mu > \mu_0$ or $H_1: \mu < \mu_0$.

A few more examples of alternative hypothesis corresponding to null hypothesis to relate real life are as follows:

- There has been a difference in the incidence of malnutrition between vaccinated and non-vaccinated.
- Males smoke more than females do.
- There is association between level of education and knowledge of child nutrition among women.

- Two teaching methods A and B are different.

❖ Simple Hypothesis

If the hypothesis specifies the population completely then it is termed as a simple hypothesis.

Example: If $x_1, x_2, \ldots \ldots, x_n$ is a random sample of size n from a normal population with mean μ and variance σ^2, then the hypothesis $H_0: \mu = \mu_0, \sigma^2 = \sigma_0^2$ is a simple hypothesis.

❖ Composite Hypothesis

If the hypothesis does not specify the population completely then it is termed as a composite hypothesis.

Example: If $x_1, x_2, \ldots \ldots, x_n$ is a random sample of size n from a normal population with mean μ and variance σ^2 then each of the following hypothesis is a composite hypothesis.

→ $\mu = \mu_0$	→ $\sigma^2 = \sigma_0^2$
→ $\mu < \mu_0,\ \sigma^2 = \sigma_0^2$	→ $\mu > \mu_0,\ \sigma^2 = \sigma_0^2$
→ $\mu = \mu_0,\ \sigma^2 < \sigma_0^2$	→ $\mu = \mu_0,\ \sigma^2 > \sigma_0^2$
→ $\mu < \mu_0,\ \sigma^2 > \sigma_0^2$	→ $\mu > \mu_0,\ \sigma^2 < \sigma_0^2$

Note: A hypothesis which does not specify completely r parameters of a population is termed as a composite hypothesis with r degrees of freedom.

❖ Parametric Hypothesis

When the hypothesis concerning the parameters of the distribution, provided the form of the distribution is called parametric hypothesis.

Example: If $X \sim N(\mu, \sigma^2)$ then $H_0: \mu = \mu_0$ is a parametric hypothesis.

❖ Non-parametric Hypothesis

While the hypothesis regarding the form of the distribution with specified or unspecified parameters is called non-parametric hypothesis.

Example: $H_0: F_X(x) = F_Y(y)$ is a non-parametric hypothesis, where $F_X(x)$ and $F_Y(y)$ are the distribution function of two population.

❖ Test of Significance

Test of significance is a statistical procedure to arrive at a conclusion or decision on the basis of samples and to test whether the formulated hypothesis can be accepted or rejected in probability sense. The aim of test of significance is to reject the null hypothesis.

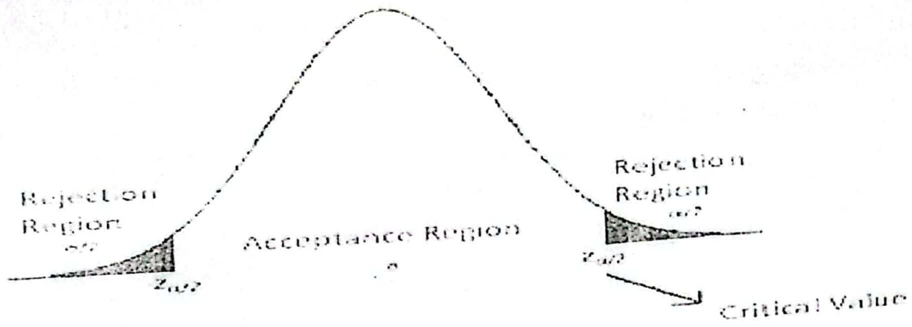
❖ Critical Region Or Rejection Region

Let $x_1, x_2, \ldots \ldots, x_n$ be a sample point designated by X in a n-dimensional sample space. If X falls in the region for which we reject H_0 when it is true then the region is called critical region. It is usually denoted by w.

In other wards, if the value of a test statistic falls in the region for which we reject H_0 when it is true then the region is known as critical region or rejection region. Usually it is denoted by w.

❖ **Acceptance Region**

If the value of a test statistic falls in the region for which we accept H_0 when it is true then the region is called acceptance region. It is denoted by w'.



❖ **One Tailed Test**

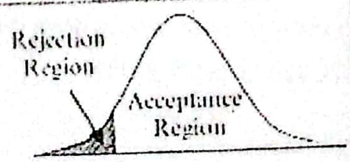
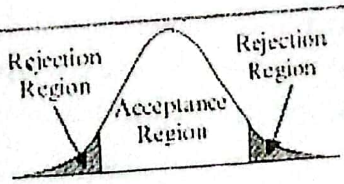
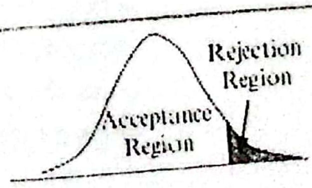
A test of any statistical hypothesis where the alternative hypothesis is one tailed (right or left tailed) is called a one tailed test. In such a case the critical region is given by the portion of the area lying in the first or last tails of the probability curve of the test statistic.

Example: A test for testing the mean of a population $H_0: \mu = \mu_0$ against the alternative hypothesis $H_1: \mu > \mu_0$ (*Right tailed*) or $H_1: \mu < \mu_0$ (*left tailed*) is a one tailed test. In the right tailed test $H_1: \mu > \mu_0$, the critical region lies entirely in the right tail of the sampling distribution of $\bar{x}$, while for the left tail test $H_1: \mu < \mu_0$, the critical region is entirely in the left tail of the distribution.

❖ **Two Tailed Test:**

A test of statistical hypothesis where the alternative hypothesis is two tailed as called two tailed test. In such a case the critical region is given by the portion of the area lying in both the tails of the probability curve of the test statistic.

Example: Suppose that there are two population brands of bulbs, one manufactured by standard process (with mean life μ_1) and the other manufactured by some new technique (with mean life μ_2). If we want to test of the bulbs differ significantly then our null hypothesis is $H_0: \mu_1 = \mu_2$ and the alternative will be $H_1: \mu_1 \neq \mu_2$ this giving us a two tailed test.

One-Tailed Test (Left Tail)	Two-Tailed Test	One-Tailed Test (Right Tail)
$H_0: \mu_x = \mu_0$ $H_1: \mu_x < \mu_0$	$H_0: \mu_X = \mu_0$ $H_1: \mu_x \neq \mu_0$	$H_0: \mu_x = \mu_0$ $H_1: \mu_X > \mu_0$
		

❖ **Type I Error**

If the null hypothesis is actually true but on the basis of sample we reject H_0 (accepting H_1), this type of error is known as type I error.

The probability of type I error is denoted by α. Thus

$$\alpha = \text{Probability of type I error}$$
$$= \text{Probability of rejecting } H_0 \text{ when } H_0 \text{ is true}$$

Symbolically, $P(x \in w \mid H_0) = \alpha \quad ; \quad where\ x = (x_1, x_2, \cdots, x_n)$

Example: In a murder case suppose that

$$H_0 : The\ man\ is\ not\ guilty$$
$$H_1 : The\ man\ is\ guilty$$

On the basis of sample if we reject the H_0 but it is true this type of error is known as type I error.

❖ **Type II Error**

If the null hypothesis is false but on the basis of sample we accept H_0 (H_1 is true), this type of error is known as type II error.

The probability of type II error is denoted by β. Thus

$$\beta = \text{Probability of type II error}$$
$$= \text{Probability of accepting } H_0 \text{ when } H_0 \text{ is false}$$

Symbolically, $P(x \in \overline{w} \mid H_1) = \beta \quad ; \quad where\ x = (x_1, x_2, \cdots, x_n)$

Example: In a murder case suppose that

$$H_0 : The\ man\ is\ not\ guilty$$
$$H_1 : The\ man\ is\ guilty$$

On the basis of sample if we accept the H_0 but it is false this type of error is known as type II error.

Question 1: Which kinds of error is more serius and why?

Solution:

I think that type I error is more serious than type II error

To support my idea an example are given below

$$H_0 : The\ man\ is\ not\ guilty$$
$$H_1 : The\ man\ is\ guilty$$

Suppose a judge is to give verdict an a criminal case. The accused may be guilty or not guilty. The judgement should be such that if the accused be guilty he should be punished while if he is not guilty he should be acequitted. If a guilty man be acquitted then an error of type II is commited whereas if an innocent may be convicted an error of type I is commited.

From the above example it is clear that type I error is more serious than type II error because in type I error an innocent man be convicted though he is not guilty.

❖ **Level of Significance**

The probability of type I error is known as the level of significance of the test. It is also called the size of the critical region. Usually it is denoted by α. Mathematically it can be written as

$P(x \in w \mid H_0) = \alpha$ or we can say

α = Probability of type I error

= Probability of rejecting H_0 when H_0 is true

❖ **Power of the Test**

The power of the test is the probability of rejecting the null hypothesis when in fact it is false and should be rejected. It is the probability of correctly decision. It can be expressed by $1-\beta$.

Mathematically it can be written as $P(x \in w \mid H_1) = 1-\beta$

Example 16·1. *Given the frequency function :*

$$f(x, \theta) = \frac{1}{\theta}, \quad 0 \le x \le \theta$$
$$= 0, \text{ elsewhere}$$

and that you are testing the null hypothesis H_0: $\theta = 1$ against H_1 : $\theta = 2$, by means of a single observed value of x. What would be the sizes of the type I and type II errors, if you choose the interval (i) $0{\cdot}5 \le x$, (ii) $1 \le x \le 1{\cdot}5$ as the critical regions ? Also obtain the power function of the test.

[Gauhati Univ, B.Sc. 1993; Calcutta Univ. B.Sc. (Maths Hons.), 1987]

Solution. Here we want to test

$H_0 : \theta = 1$, against $H_1 : \theta = 2$.

(*i*) Here $W = \{x : 0{\cdot}5 \le x\} = \{x : x \ge 0{\cdot}5\}$

and $\bar{W} = \{x : x \le 0{\cdot}5\}$

$$\alpha = P\{x \in W \mid H_0\} = P\{x \ge 0{\cdot}5 \mid \theta = 1\}$$
$$= P\{0{\cdot}5 \le x \le \theta \mid \theta = 1\} = P\{0{\cdot}5 \le x \le 1 \mid \theta = 1\}$$
$$= \int_{0.5}^{1} [f(x, \theta)]_{\theta=1}\, dx = \int_{0.5}^{1} 1.\, dx = 0{\cdot}5$$

Similarly,

$$\beta = P\{x \in \bar{W} \mid H_1\} = P\{x \le 0{\cdot}5 \mid \theta = 2\}$$
$$= \int_{0}^{0.5} [f(x, \theta)]_{\theta=2}\, dx = \int_{0}^{0.5} \tfrac{1}{2}\, dx = 0{\cdot}25$$

Thus the sizes of type I and type II errors are respectively

$\alpha = 0{\cdot}5$ and $\beta = 0{\cdot}25$

and power function of the test $= 1 - \beta = 0{\cdot}75$

(*ii*) $W = \{x : 1 \le x \le 1{\cdot}5\}$

$$\alpha = P\{x \in W \mid \theta = 1\} = \int_{1}^{1.5} [f(x, \theta)]_{\theta=1}\, dx = 0,$$

since under $H_0 : \theta = 1, f(x, \theta) = 0$, for $1 \le x \le 1{\cdot}5$.

$$\beta = P\{x \in \bar{W} \mid \theta = 2\} = 1 - P\{x \in W \mid \theta = 2\}$$

SA-1
7/4/20

Test of Hypothesis (2)

❖ Steps in Solving Testing of Hypothesis Problem

The major steps involved in the solution of a "testing of hypothesis" problem may be outlined as follows:

- Explicit knowledge of the nature of the population distribution and the parameter(s) of interest, i.e., the parameter(s) about which the hypothesis are set up.
- Setting up of the null hypothesis H_0 and the alternative hypothesis H_1 in terms of the range of the parameter values each one embodies.
- Choose the appropriate level of significance (α) depending on the reliability of the estimates and permissible risk.
- Choose the suitable test statistic and compute the test statistic under the null hypothesis.
- We compare the computed value of test statistic with the significant value (tabulated value) at the given level of significance α.
    - If the computed value of test statistic (in modulus value) is greater than the significant value (tabulated value) then we reject the null hypothesis at level of significance α and we say that it is significant.
    - If the computed value of test statistic is less than the tabulated value then we accept the null hypothesis at level of significance α and we say that it is not significant.

❖ The p-Value Approach to Hypothesis Testing:

The $p-value$ is the probability of obtaining a test statistic equal to or more extreme than the result obtained from the sample data, given that the null hypothesis H_0 is really true.

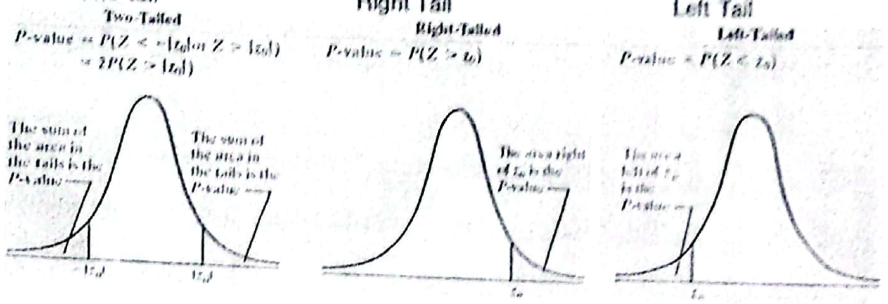
The $p-value$ is often referred to as the observed level of significance, which is the smallest level at which H_0 can be rejected for a given set of data. The decision rule for rejecting H_0 in the $p-value$ approach is:

- If the $p-value$ greater than or equal to α, the null hypothesis is not rejected.
- If the $p-value$ is smaller than α, the null hypothesis is rejected.

P-Value Approach

Assume that the null hypothesis is true.
The P-Value is the probability of observing a sample mean that is as or more extreme than the observed.
How to compute the P-Value for each type of test:
Step 1: Compute the test statistic $z_0 = \dfrac{\bar{x} - \mu_0}{\sigma / \sqrt{n}}$

Two-tail Right Tail Left Tail



❖ **Steps in Determining the** $p-value$

- State the null hypothesis H_0
- State the alternative hypothesis H_1.
- Choose the level of significance α.
- Choose the sample size n.
- Determine the appropriate statistical technique and corresponding test statistic to use.
- Collect the data and compute the sample value of the appropriate test statistic.
- Calculate the $p-value$ based on the test statistic. This involves
  - → Sketching the distribution under the null hypothesis H_0.
  - → Placing the test statistic on the horizontal axis.
  - → Shading in the appropriate area under the curve, on the basis of the alternative hypothesis H_1.
- Compare the $p-value$ to α.
- Make the statistical decision. If the $p-value$ is greater than or equal to α, the null hypothesis is not rejected. If the $p-value$ is smaller than α, the null hypothesis is rejected.

- Express the statistical decision in terms of the particular situation.

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

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

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

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

❖ **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)

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

Reject H_0 if

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

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

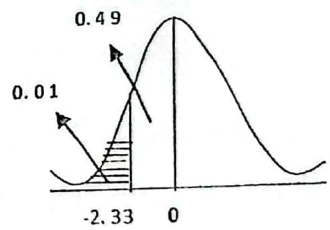
❖ Problem-1

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:



$$H_0: \mu \geq 13$$

$$H_A: \mu < 13$$

Here, we have that

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

$n = 145$ (**Large**) , $s = 9.2$ **Hours** (σ **Unknown**)

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

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

Under the H_0, the test statistic is given by:

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

Comment: Since $Z_{calculated} < -2.33$. 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.

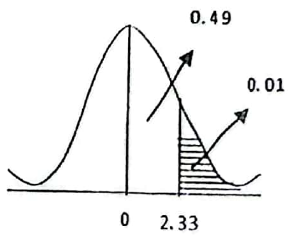
Formulate the null and alternative hypothesis such that the rejection of H_0 will support the charge of a premium rate. At 1% level of significance what is your conclusion?

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

$n = 35$ (Large) , $s = 4$ minutes (σ Unknown)

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

$$Z_{calculated} < -2.33 \quad \text{or} \quad Z_{calculated} > 2.33$$

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 study reported that the teens spend a mean of 5.72 Tk. per visit to fast food restaurants. In a follow up study, a sample of 102 teens visit to fast food restaurants in Dhaka City found a sample mean of 5.98 Tk. and a sample standard deviation of 1.24 Tk.

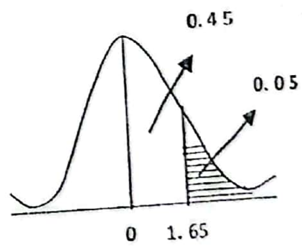
Formulate the null and alternative hypothesis that can be used to determine whether the sample data support the conclusion that the teens in Dhaka City have a population mean expenditure more than 5.72 Tk. per visit to fast food restaurants. At 5% level of significance what is your conclusion?

Solution:

The null and alternative hypotheses are given as follows:

$$H_0: \mu \leq 5.72$$

$$H_A: \mu > 5.72$$



Here, we have that

$\alpha = 0.05$, $\bar{x} = 5.98$ Tk.

$n = 102$ (Large) , $s = 1.24$ Tk. (σ Unknown)

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

$$Z_{tabulated} = Z_{\alpha} = Z_{0.01} = \pm 1.65$$

That is, the null hypothesis will be rejected if

$$Z_{calculated} < -1.65 \quad \text{or} \quad Z_{calculated} > 1.65$$

Under the H_0, the test statistic is given by:

$$Z_{calculated} = \frac{\bar{x} - \mu}{\frac{s}{\sqrt{n}}} = \frac{5.98 - 5.72}{\frac{1.24}{\sqrt{102}}} = 2.12$$

Comment: Since $Z_{calculated} > 1.65$. So, the null hypothesis is rejected. That means, the sample data support the conclusion that the teens in Dhaka City have a population mean expenditure more than 5.72 Tk. per visit to fast food restaurants.

❖ **Problem-4**

A production line operates with a mean filling weight of 16 ounces per container. Over-filling or under-filling is a serious problem and the production line should be shut down if either occurs. From past data, standard deviation is assumed to be known 0.8 ounces. A quality control inspector samples 30 items every two ours and at that time makes the decision of whether to shut down for adjustment.

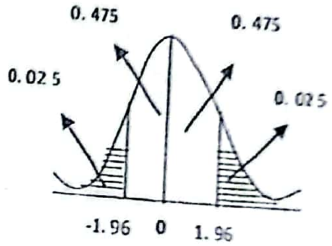
Formulate the appropriate null and alternative hypothesis that can be used to determine whether the sample data support the conclusion that the production line should be shut down. If a sample mean of 16.32 ounces were found, what action would you recommend at 5% level of significance?

Solution:

The null and alternative hypotheses are given as follows:

$$H_0: \mu = 16$$

$$H_A: \mu \neq 16$$



Here, we have that

$\alpha = 0.05$, $\bar{x} =$ 16.32 ounces

$n = 30$ (Large) , $\sigma =$ 0.8 ounces (σ known)

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

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

That is, the null hypothesis will be rejected if

$Z_{calculated} > 1.96$ *or* $Z_{calculated} < -1.96$

Under the H_0, the test statistic is given by:

$$Z_{calculated} = \frac{\bar{x} - \mu}{\frac{\sigma}{\sqrt{n}}} = \frac{16.32 - 16}{\frac{0.8}{\sqrt{30}}} = 2.19$$

Comment: Since $Z_{calculated} > 1.96$. So, the null hypothesis is rejected. That means, the sample data support the conclusion that the production line should be shut down.

❖ Problem-5

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.

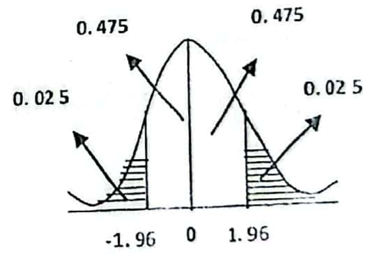
Formulate the appropriate null and alternative hypothesis and test the hypothesis at 5% level of significance.

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:

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

That is, the null hypothesis will be rejected if

$Z_{calculated} > 1.96$ *or* $Z_{calculated} < -1.96$

Under the H_0, the test statistic is given by:

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

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

❖ **Problem-6**

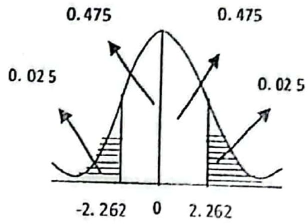
A sample of 10 financial services corporations provided the following earnings per share data:

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

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

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

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

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

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

That is, the null hypothesis will be rejected if

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

Under the H_0, the test statistic is given by:

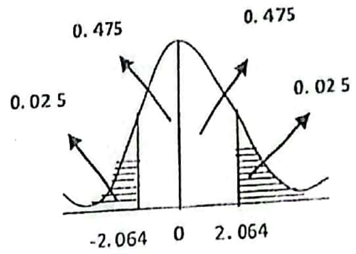
$$t_{calculated} = \frac{\bar{x} - \mu}{\frac{s}{\sqrt{n}}} = \frac{2.80 - 3}{\frac{0.70}{\sqrt{10}}} = -0.90$$

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

❖ **Problem-7**

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

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

$n = 25$ (Small) , $s = 14.5$ Tk. (σ Unknown)

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

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

That is, the null hypothesis will be rejected if

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

Under the H_0, the test statistic is given by:

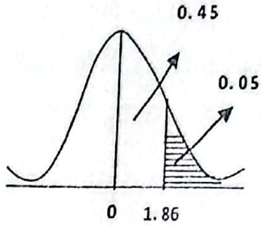
$$t_{calculated} = \frac{\bar{x} - \mu}{\frac{s}{\sqrt{n}}} = \frac{84.5 - 90}{\frac{14.5}{\sqrt{25}}} = -1.90$$

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

❖ **Problem-8**

Using nine test drives, the mean driving distance by a driver was 286.9 yards with a sample standard deviation of 10 yards. Formulate the null and alternative hypothesis that can be used to determine whether the driver has populations mean driving distance greater than 280 yards. At 5% level of significance, what is your conclusion?

Solution:



The null and alternative hypotheses are given as follows:

$$H_0: \mu \leq 280$$

$$H_A: \mu > 280$$

Here, we have that

$\alpha = 0.05$, $\bar{x} = 286.9$ yards

$n = 9$ (Small) , $s = 10$ yards (σ Unknown)

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

$$t_{\text{tabulated}}(8) = t_{\alpha}(8) = t_{0.05}(8) = 1.86$$

That is, the null hypothesis will be rejected if

$$t_{\text{calculated}} > 1.86$$

Under the H_0, the test statistic is given by:

$$t_{\text{calculated}} = \frac{\bar{x} - \mu}{\frac{s}{\sqrt{n}}} = \frac{286.9 - 280}{\frac{10}{\sqrt{9}}} = 2.07$$

Comment: Since $t_{\text{calculated}} > 1.86$. So, the null hypothesis is rejected.

❖ **Test of Equality of Two Means**

Let $x_{11}, x_{12}, \cdots\cdots, x_{1n_1}$ 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 $x_{21}, x_{22}, \cdots\cdots, x_{2n_2}$ 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

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

The test statistic is

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

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-9**

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

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-10**

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-11**

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

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 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,

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

The corresponding sample proportion of P_1 *and* P_2 are

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

The problem is to test the hypothesis

$$H_0: P_1 = P_2$$

$$H_1: P_1 \neq P_2$$

The test statistic is

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

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 $Q_0 = 1 - 0.10 = 0.90$

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

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

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

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 ${\chi^2}_{cal} \leq {\chi^2}_{1-\frac{\alpha}{2}, n-1}$ or ${\chi^2}_{cal} \geq {\chi^2}_{\frac{\alpha}{2}, n-1}$ 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 ${\chi^2}_{cal} \geq {\chi^2}_{\alpha, n-1}$ 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 $\chi^2_{cal} \leq \chi^2_{1-\alpha, n-1}$ 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 $F_{cal} \geq F_{0.05;\, n_1-1, n_2-1}$ 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

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

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

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

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

Here,

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

$\bar{x} = \frac{\sum x_i}{n} = \frac{946}{11} = 86$ 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^2 at 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

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

$$= 1608.33$$

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

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.