9.11. ▶ CENTRAL LIMIT THEOREM

The normal distribution which we have studied in previous chapter plays an important role in statistics and probability. When sampling is done from a normal population, the means of the samples drawn from this population are themselves normally distributed. But when sampling is not done from a normal population, then the size of the sample plays a critical role. For small sample size, the shape of sampling distribution will depend upon the shape of population from which it is taken. But as the sample size gets larger and larger, the shape of the sampling distribution will become more and more like a normal distribution irrespective of the shape of the population from which it is taken. The **central limit theorem** explains this sort of relationship between the shape of the population distribution and sampling distribution.

Central limit theorem states that if from a population having mean μ and standard deviation σ, a large number of random samples are drawn, then the distribution of sample means will be approximately normally distributed.

9.12. ▶ HYPOTHESIS AND ITS TESTING

Hypothesis : A statistical hypothesis is an assumption or statement about the value of a population parameter which may be true or false. Such assumptions are made before any final decision is taken about the population on the basis of sample information.

According to Prof. Morris Hamburg, "*A hypothesis in statistics is simply a quantitative statement about the population*".

Testing of Hypothesis. The method or procedure which enables us to decide whether a hypothesis is true or false (on the basis of sample results) is known as *Test of Hypothesis* or *Test of Significance*. The main purpose of sampling theory is to study these tests. By testing the hypothesis, we can find out whether hypothesis deserves acceptance or rejection.

Procedure of Hypothesis Testing. We first set up a hypothesis. The hypothesis is a set of inference that is drawn regarding the parameter of the population. For example, we may set a hypothesis that the mean weight of students in a school is 60 kgs. This is an assumption about the parameter (mean) of the population and is called the *hypothesized value of the parameter*.

To find out whether our hypothesis is valid or not, we shall randomly select some students of the school and find the sample mean weight, which we call sample statistics. We then find the difference between the actual value of the sample statistic and the assumed or hypothesized value of the parameter and judge whether the difference is significant. The smaller the difference, the greater the likelihood that our hypothesized value is correct *i.e.*, our hypothesis is valid.

Generally, a single hypothesis is not set. In fact, two complimentary hypothesis are set at the same time. If one of the hypothesis is accepted, then the other is automatically rejected and vice-versa.

9.13. ▶ TYPES OF HYPOTHESIS

There are two types of hypothesis :

(*a*) Null Hypothesis

(*b*) Alternative Hypothesis

9.13.1. Null Hypothesis

The concept of null hypothesis is very important in testing the significance of difference. Null hypothesis suggests that there is no significant difference between sample statistic and the corresponding population parameter value or between two sample statistics. Null hypothesis is also known as *hypothesis of no difference*.

A null hypothesis is denoted by H_0 and usually it is suggested at the first step of any testing. For example, if we want to test whether the sample with mean μ_0 may be regarded as drawn from the population having mean μ, then we set a null hypothesis (H_0) that difference between the population mean and sample mean is negligible or not significant. Thus, our null hypothesis is $H_0 : \mu = \mu_0$ (*i.e.* the mean of the population = mean of the sample).

The rejection of the null hypothesis indicates that there is a significant difference between the means and acceptance of null hypothesis indicates that the difference between the means is not significant.

9.13.2. Alternative Hypothesis

Any hypothesis which contradicts the null hypothesis H_0 is called alternative hypothesis. It is denoted by H_a.

The two hypothesis H_0 and H_a are such that if one is accepted, the other is rejected and vice-versa.

For example, if null hypothesis H_0 is : $\mu = \mu_0$, then alternative hypothesis may be

(*a*) $H_a : \mu \neq \mu_0$ (*i.e.*, $\mu > \mu_0$ or $\mu < \mu_0$)

(*b*) $H_a : \mu > \mu_0$

(*c*) $H_a : \mu < \mu_0$.

9.14. ▶ ERRORS IN TESTING OF A HYPOTHESIS

As we have learnt that null hypothesis (H_0) and alternative hypothesis (H_a) are complementary to each other. Thus either H_0 is true or H_a is true. Both H_0 and H_a cannot be true. Ideally, the hypothesis testing procedure should lead to the acceptance of H_0 when H_0 is true and rejection of H_0 when H_0 is false (*i.e.*, H_a is true). However, the correct conclusions are not always possible and we come across certain errors which are called *errors in hypothesis*.

There are primarily two kinds of errors.

1. Type I Error
2. Type II Error

Type I errors are made when we reject the null hypothesis in testing a procedure, although it is true. The probability of making such type of errors is usually denoted by Greek letter α.

Type II errors are made when we accept the null hypothesis in testing a procedure, although it is false. The probability of making such type of errors is usually denoted by Greek letter β.

The following table will help in understanding these types of errors :

	H_0 Accepted	H_0 Rejected
H_0 is true	No error (correct decision)	Type I Error
H_0 is false	Type II Error	No error (correct decision)

9.15. ▶ LEVEL OF SIGNIFICANCE

In testing a given hypothesis, the *level of significance is defined as the degree of significance with which we reject or accept the given hypothesis*. In a hypothesis test, the significance level is taken as α, which denotes the probability of making a Type I error *i.e.*, probability of rejecting the null hypothesis, although it is true.

Since 100% accuracy is not possible in taking decision over acceptance or rejection of hypothesis, therefore level of significance is the maximum possible probability with which we are willing to commit an error in rejecting the null hypothesis H_0 even when it is true.

The level of significance is usually mentioned before a test procedure so that the results obtained may not influence our decision. Generally, we take 5% or 1% level of significance.

A significance level of 5% indicates that the probability of rejecting the null hypothesis, though it is true is 0·05 *i.e.*, we run the risk of rejecting a true hypothesis 5 out of every 100 times. By accepting the null hypothesis at 5% we run the risk of making a wrong decision about 5% of the time *i.e.*, our decision will be correct upto the extent of 95%.

By testing at 1% we seek to reduce the chance of making a wrong judgement as the element of risk is reduced to 1% in this case and so our decision will be correct upto the extent of 99%.

Remark : A significance level of 5% implies that if the same sampling method is used for different samples, then 5% of the samples drawn will not include the population parameter.

9.15.1. Confidence Level

In statistics, confidence level refers to the percentage of all possible samples that can be expected to include the true population parameter. For example, a 95% confidence level indicates that 95% of the samples drawn will include the population parameter while the remaining 5% of the samples drawn will not include the population parameter.

Thus, if the confidence level is 95%, then the significance level is 5%.

9.15.2. Confidence Interval

It is range of values that contain the population parameter.

A confidence interval $[a, b]$ for the population parameter with a 95% confidence level indicates that in 95% of samples drawn, the population parameter falls within the confidence interval.

9.16. ▶ TAILED TESTS OF HYPOTHESIS

One Tailed and Two-Tailed Tests :

As discussed earlier, if we have to check whether the population mean (μ) has a specified value μ_0, then the null hypothesis is $H_0 : \mu = \mu_0$ and the alternative hypothesis may be

(*i*) $H_a : \mu \neq \mu_0$ (*i.e.*, $\mu > \mu_0$ or $\mu < \mu_0$) or (*ii*) $H_a : \mu > \mu_0$ or (*iii*) $H_a : \mu < \mu_0$.

The alternative hypothesis in (*i*) *viz.*, $H_a : \mu \neq \mu_0$ (*i.e.*, $\mu > \mu_0$ or $\mu < \mu_0$) is called a *two tailed alternative* (both right tail and left tail)

The alternative hypothesis in (*ii*) *viz.*, $H_a : \mu > \mu_0$ is called a *right tailed alternative*.

The alternative hypothesis in (*iii*) *viz.*, $H_a : \mu < \mu_0$ is called a *left tailed alternative*.

The right tailed or left tailed alternatives are called *one tailed alternatives*.

If the alternative hypothesis H_a in a test of a statistical hypothesis is two-tailed (*i.e.* both right-tailed and left-tailed) then the corresponding test is called a **two-tailed test** and if the alternative hypothesis H_a in a test of a statistical hypothesis is one-tailed (*i.e.* either right-tailed or left-tailed but not both) then the corresponding test is called a **one-tailed test.** Thus, the *alternative hypothesis plays a very important role in deciding whether one tailed test (right or left) or a two tailed test is to be conducted.* A two tailed test is appropriate when the null hypothesis is some specified value and the alternative hypothesis is a value not equal to the specified value of null hypothesis.

9.16.1. Critical Region or Rejection Region

The set of values for the test statistic Z at which the null hypothesis H_0 is rejected is known as the *critical region* or *rejection region*.

(*i*) In a two tailed test, there are two critical regions (rejection regions) at α level of significance and these lie equally in both the tails of the sampling distribution of means.

Thus, in a two tailed test, if the significance level is 5%, then the rejection region will be equal to 0·05 of area, equally splitted under both tails of the curve as 0·025 of area, and the acceptance region will be equal to 0·95 of the area, equally splitted on each side of μH_0 as 0·475 of area [See fig. 9.1]. From the table of areas under the normal curve, we observe that area of 0·475 on the left of μH_0 corresponds to $-1.96 \leq Z \leq 0$ and area of 0·475 on the right of μH_0 corresponds to $0 \leq Z \leq 1.96$. Thus the acceptance region for two tailed test is $-1.96 \leq Z \leq 1.96$ *i.e.*, $|Z| \leq 1.96$ and the rejection region is $Z < -1.96$ and $Z > 1.96$ *i.e.*, $|Z| > 1.96$. The value ± 1.96 is called the critical value of Z at 5% level of significance for a two tailed test.

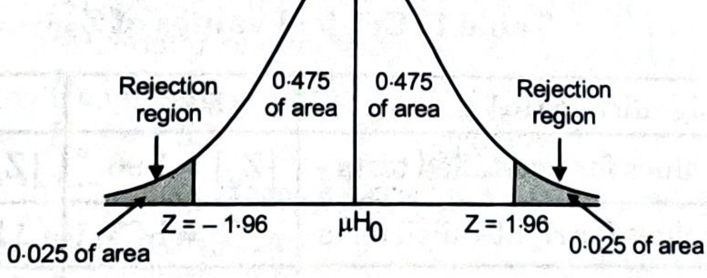

Fig. 9.1 (Two Tailed)

(*ii*) In a right tailed test, the critical region (rejection region) at α level of significance lies entirely in the right tail of the sampling distribution of means.

Thus, in a right tailed test, if the significance level is 5%, then the rejection region will be equal to 0·05 of area under the right tail and the acceptance region will be 0·95 of the area, of which 0·45 of the area is on the right side of μH_0 and 0·50 of the area is on the left side of μH_0 [See fig. 9.2]. From the table of areas under the normal curve, we observe that area of 0·45 on the right side of μH_0 corresponds to $0 \le Z \le 1{\cdot}65$. Thus the acceptance region for right tailed test is $Z \le 1{\cdot}65$ and the rejection region is $Z > 1{\cdot}65$. The value 1·65 is called the critical value of Z at 5% level of significance for a right tailed test.

(*iii*) In a left tailed test, the critical region (rejection region) at α level of significance lies entirely in the left tail of the sampling distribution of means.

Thus, in a left tailed test, if the significance level is 5%, then the rejection region will be equal to 0·05 of the area under the left tail and the acceptance region will be 0·95 of the area, of which 0·50 of area is on the right side of the μH_0 and 0·45 of the area is on the left side of μH_0 [See fig. 9.3]. From the table of areas under the normal curve, we observe that area of 0·45 on the left side of μH_0 corresponds to $-1{\cdot}65 \le Z \le 0$. Thus the acceptance region for left tailed test is $Z \ge -1{\cdot}65$ and the rejection region is $Z < -1{\cdot}65$. The value $-1{\cdot}65$ is called the critical value of Z at 5% level of significance for a left tailed test.

The following figures show the acceptance and rejection regions for one tailed test at 5% level of significance :

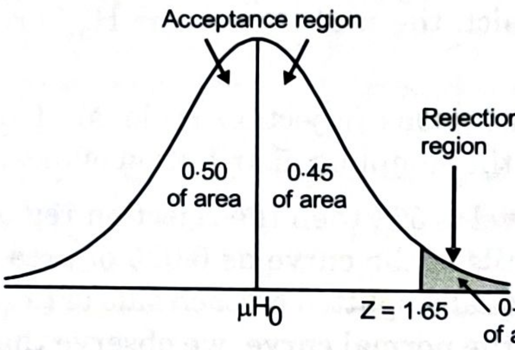

Fig. 9.2 (Right Tailed)

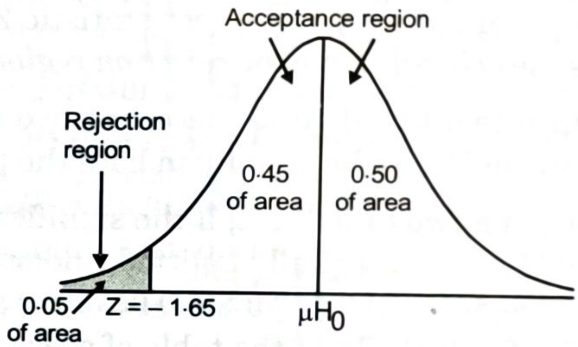

Fig. 9.3 (Left Tailed)

Proceeding in the similar manner, we can find the rejection and acceptance regions at 1% level of significance.

The critical values of Z at α level of significance are denoted by Z_α. The critical values for two-tailed and one-tailed (right tailed and left tailed) tests at 5% and 1% level of significance are given in the following table :

Table 1 : Critical values of Z

Level of significance (α)	5%	1%
Critical values for two-tailed tests	$\lvert Z_\alpha \rvert = 1{\cdot}96$	$\lvert Z_\alpha \rvert = 2{\cdot}58$
Critical values for right-tailed tests	$Z_\alpha = 1{\cdot}65$	$Z_\alpha = 2{\cdot}33$
Critical values for left-tailed tests	$Z_\alpha = -1{\cdot}65$	$Z_\alpha = -2{\cdot}33$

9.17. ▶ Testing of Hypothesis about Population Mean (μ) When Population Variance (σ^2) is known Using Critical Values (Z-Test)

This test is based on the normal probability distribution and is specifically used for testing significance of various statistical measures, particularly mean.

Z-test is useful for comparing the mean of a sample with the hypothesized value of mean of the normal population in case of large samples, when the population variance is known.

The procedure of testing a hypothesis using this test is as given below :

(*i*) Set up the null hypothesis H_0 about the population mean (μ).

(*ii*) Set up an alternative hypothesis H_a which contradicts H_0. From the nature of H_a decide whether one-tailed (right or left tail) test or two-tailed test is to be applied.

(*iii*) Choose a suitable level of significance α, if it is not given in a problem. (If not given, then we generally choose 5% level of significance).

(*iv*) Calculate the value of the test statistic Z by using :

$$Z = \frac{\bar{x} - \mu_0}{\frac{\sigma}{\sqrt{n}}}$$

where $\bar{x}$ = sample mean

μ_0 = hypothetised value of population mean

σ = population standard deviation

n = size of the sample.

(*v*) Compare the calculated value of the test statistic Z with Z_α *i.e.*, the critical value of Z at α level of significance.

The rejection or acceptance of null hypothesis can be determined as follows :

(a) One tailed (right tailed) test :

(*i*) If $Z > Z_\alpha$, then reject the null hypothesis H_0 at α level of significance.

(*ii*) If $Z < Z_\alpha$, then accept the null hypothesis H_0 at α level of significance.

(b) One tailed (left tailed) test :

(*i*) If $Z < Z_\alpha$, then reject the null hypothesis H_0 at α level of significance.

(*ii*) If $Z > Z_\alpha$, then accept the null hypothesis H_0 at α level of significance.

(c) Two tailed test :

(*i*) If $|Z| > Z_\alpha$, then reject the null hypothesis H_0 at α level of significance.

(*ii*) If $|Z| < Z_\alpha$, then accept the null hypothesis H_0 at α level of significance.

SOLVED EXAMPLES

EXAMPLE 1. *Consider the following hypothesis :*

$$H_0 : \mu = 35$$
$$H_a : \mu \neq 35$$

A sample of 81 items is taken whose mean is 37·5 and the population standard deviation is 5. Test the hypothesis at 5% level of significance.

Solution. Here $\bar{x} = 37{\cdot}5$, $\mu_0 = 35$, $n = 81$ and $\sigma = 5$

From the nature of the alternative hypothesis H_a, we observe that two tailed test is suitable.

The test statistic Z is given by $Z = \dfrac{\bar{x} - \mu_0}{\dfrac{\sigma}{\sqrt{n}}}$

$$\Rightarrow \quad Z = \frac{37 \cdot 5 - 35}{\dfrac{5}{\sqrt{81}}} = \frac{2 \cdot 5}{5} \times 9 = 4 \cdot 5$$

The critical value of Z for a two tailed test at 5% level of significance (*i.e.*, Z_α) is 1·96. Clearly, $4{\cdot}5 > 1{\cdot}96$ *i.e.*, $|Z| > Z_\alpha$.

Hence, the null hypothesis H_0 is rejected.

EXAMPLE 2. *Test the following hypothesis at 1% level of significance :*

$$H : \mu_0 = 60$$
$$H_a : \mu < 60$$

A sample of 100 is taken with $\bar{x} = 58$ and the population standard deviation is 16.

Solution. Here $\bar{x} = 58$, $\mu_0 = 60$, $n = 100$ and $\sigma = 16$

From the nature of the alternative hypothesis H_a, we observe that left tailed test is suitable.

The test statistic Z is given by $Z = \dfrac{\bar{x} - \mu_0}{\dfrac{\sigma}{\sqrt{n}}}$

$$\Rightarrow \quad Z = \frac{58 - 60}{\dfrac{16}{\sqrt{100}}} = \frac{-2}{16} \times 10 = -\frac{20}{16} = -1{\cdot}25$$

The critical value of Z for a left tailed test at 1% level of significance (*i.e.*, Z_α) is $-2{\cdot}33$. Clearly, $-1{\cdot}25 > -2{\cdot}33$ *i.e.*, $Z > Z_\alpha$.

Hence, we accept the null hypothesis H_0.

EXAMPLE 3. *The surveyers claimed that in the state of Delhi, there has been increase in the average size of the plot since the past five years. The previous study reported a mean size of 175 sq. yards with a population standard deviation of 130 sq. yards. A sample of 25 plots is taken and a sample mean of 250 sq. yards is obtained. Is the evidence enough to support the claim of the surveyor at 5% level of significance.*

Solution. Here, $\bar{x} = 250$, $\mu_0 = 175$, $n = 25$ and $\sigma = 130$

Let us define null hypothesis H_0 : Mean size of the plot = 175 sq. yards

and the alternative hypothesis H_a : Mean size of the plot > 175 sq. yards

[∵ *It is claimed that there is an increase in the average size of plots*]

From the nature of the alternative hypothesis H_a we observe that right tailed test is suitable.

The test statistic Z is given by $Z = \dfrac{\bar{x} - \mu_0}{\dfrac{\sigma}{\sqrt{n}}}$

$$\Rightarrow \quad Z = \frac{250 - 175}{\frac{130}{\sqrt{25}}} = \frac{75 \times 5}{130} = 2{\cdot}88$$

The critical value of Z for a right tailed test at 5% level of significance (*i.e.*, Z_α) is 1·65.

Clearly, 2·88 > 1·65 *i.e.*, $Z > Z_\alpha$.

Hence, we reject the null hypothesis H_0 and accept the alternative hypothesis H_a that mean size of the plot is greater than 175 sq. yards which supports the claim of the surveyor.

EXERCISE 9.1

1. Consider the following hypothesis :

$$H_0 : \mu = 46$$
$$H_a : \mu \neq 46$$

A sample of 49 items is taken whose mean is 48·6 and the population standard deviation is 8. Test the hypothesis at 5% level of significance.

2. Test the following hypothesis at 5% and 1% level of significance :

$$H : \mu_0 = 72$$
$$H_a : \mu > 72$$

A sample of 144 is taken with $\overline{x} = 74$ and the population standard deviation is 14.

3. A sample of 400 male students is found to have a mean height of 67·47 inches. Can it be reasonably regarded as a sample from a large population with mean height of 67·39 inches and standard deviation of 1·30 inches. Test at 5% level of significance.

4. The mean of a certain production process is known to be 50 with a standard deviation of 2·5. The production manager has claimed a decrease in the production process this year. A sample of 12 items is taken, which gives a mean value of 48·5. Is the manager's claim valid. Test the manager's claim at 1% level of significance.

ANSWERS

1. Z = 2·275; Null hypothesis H_0 is rejected
2. Z = 1·71; Null hypothesis H_0 is rejected at 5% level of significance and accepted at 1% level of significance
3. Z = 1·231; Null hypothesis H_0 is accepted
4. Z = – 2·0784; Null hypothesis H_0 is rejected.

9.18. ▶ STUDENT'S *t*–TEST

The sampling theory for large samples is not applicable in small samples because when samples are small, we cannot assume that the sampling distribution is normal. So we require a different technique for testing small samples, particularly when population parameters are unknown.

In the year 1905, Sir William Gosset developed a test called ***t*-test** for testing small samples and published his discovery under the pen name Student. Thus this test is also called Student's *t*-test. This test was based on *t*-distribution and so it is also known as Student's *t*-distribution. It is used when the sample size is less than 30 and the population standard deviation is unknown.

If $\{x_1, x_2, \ldots\ldots, x_n\}$ is any random sample of size n drawn from a normal population with mean μ and variance σ^2, then the test statistic t is defined by

$$t = \frac{\bar{x} - \mu}{\frac{S}{\sqrt{n}}}, \qquad \ldots(1)$$

where $\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i$ is the sample mean

and $S^2 = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{x})^2$ is the unbiased estimate of population variance σ^2.

If s^2 is the sample variance, then $s^2 = \frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})^2$

$$\Rightarrow \quad \frac{S^2}{s^2} = \frac{n}{n-1} \Rightarrow \frac{S^2}{n} = \frac{s^2}{n-1}$$

$$\Rightarrow \quad \frac{S}{\sqrt{n}} = \frac{s}{\sqrt{n-1}} \qquad \ldots(2)$$

Thus, the test statistic $t = \dfrac{\bar{x} - \mu}{\frac{S}{\sqrt{n}}}$ can also be defined as $t = \dfrac{\bar{x} - \mu}{\frac{s}{\sqrt{n-1}}}$ [*Using* (2)]

where s is the sample standard deviation.

The test statistic t follows Student's t-distribution with probability density function $f(t)$ given by

$$f(t) = C\left(1 + \frac{t^2}{\nu}\right)^{-\frac{\nu+1}{2}}, \; -\infty < t < \infty, \text{ where}$$

C = constant required to make the area under curve equal to unity

and $\nu = n - 1$ is the number of degrees of freedom.

The t-distribution has been derived mathematically under the assumption of normal distribution.

9.18.1. Degrees of Freedom

The degrees of freedom refers to the number of values which have the freedom to vary. It is denoted by ν and it is an integral part of inferential statistics. For example, if 30 students are to be seated in a classroom containing 30 seats, the first 29 students have a choice to sit at any of the 30 seats, but the 30th student is left with only one seat *i.e.*, the first 29 students are free to vary as regards their sitting position but the 30th student cannot vary. Hence the degrees of freedom = (30 – 1) = 29.

Thus, the degrees of freedom refers to number of logically independent values which have a freedom to vary in computing a statistic.

Remark : For a sample with n observations $x_1, x_2, x_3, \ldots\ldots, x_n$, the sample mean $\bar{x}$ has n degrees of freedom and the sample standard deviation has $n - 1$ degrees of freedom (since sample standard deviation depends on mean)

9.19. ▶ PROPERTIES OF t–DISTRIBUTION

1. The probability density function of a t-distribution is given by

$$f(t) = C\left(1+\frac{t^2}{\nu}\right)^{-\frac{\nu+1}{2}}$$

We observe that $f(t) = -f(t)$ and therefore the t-distribution curve is symmetrical about the line $t = 0$.

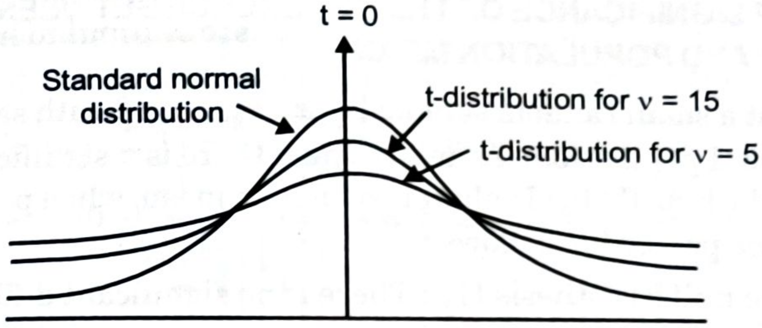

Fig. 9.4

2. The variable t of t-distribution lies between $-\infty$ to $+\infty$.
3. As the number of degrees of freedom increases, the t-distribution curve moves closer and closer to the standard normal distribution curve as shown in fig. 9.4.
4. The mean of the t-distribution is zero.
5. The variance of the t-distribution is greater than 1, but approaches 1 as the degrees of freedom and the sample size becomes large. Thus, as the sample size increases, the variance of the t-distribution approaches the variance of the standard normal distribution.
6. If degrees of freedom is infinite *i.e.*, $\nu = \infty$, then the t-distribution curve and normal distribution curve are exactly the same.
7. The t-distribution has proportionately greater area in its tails than the normal distribution.

9.19.1. The t-table

A t-table is a table showing probabilities (areas) under the probability density function of the t-distribution for different degrees of freedom. The t-distribution values or t-values have been tabulated for various levels of significance α *viz.*, $\alpha = 0{\cdot}10, 0{\cdot}05, 0{\cdot}025, 0{\cdot}01, 0{\cdot}005$ and degrees of freedom $(\nu) = 1, 2, 3, \ldots\ldots, 29$. This table is appended at the end of this book.

From the t-table we can obtain the values of t with ν degrees of freedom at α level of significance. This value is denoted by $t_\nu(\alpha)$, where $t_\nu(\alpha)$ is such that $P(|t| > t_\nu(\alpha)) = \alpha$.

The t-distribution has different values for each degree of freedom. When degrees of freedom is 30 or more, then probabilities related to t-distribution are approximated using the normal distribution table.

Reading of t-table : If we have to find the value of t at 5% level of significance (*i.e.*, $\alpha = 0{\cdot}05$) with 9 degrees of freedom *i.e.*, we have to find $t_9(0{\cdot}05)$, then we locate the value which is at the intersection of the column headed by 0·05 and the row of 9 degrees of freedom (d.f.). From the t-table, we observe that $t_9(0{\cdot}05) = 1{\cdot}833$.

9.20. ▶ APPLICATIONS OF t-DISTRIBUTION

In this section we shall discuss the following two applications of t-distribution :

1. To test the significance of the mean of a random sample *i.e.*, to test if the sample mean ($\bar{x}$) differs significantly from the hypothetical value of the population mean (μ). This is also called **one-sample t-test.**
2. To test the significance of the difference between means of two independent samples. This is also called **two-sample t-test.**

9.21. ▶ TEST OF SIGNIFICANCE OF THE DIFFERENCE BETWEEN MEAN OF A RANDOM SAMPLE AND POPULATION MEAN

Consider that a small random sample $\{x_1, x_2, x_3, \ldots, x_n\}$ with sample size n and mean $\bar{x}$ is drawn from a normal population. To determine if there is a significant difference between the sample mean and the hypothetical value of population mean, when population standard deviation (σ) is not known, we proceed as follows :

1. Define the null hypothesis H_0 : There is no significant difference between the sample mean ($\bar{x}$) and the population mean (μ).

and the alternative hypothesis H_a : There is a significant difference between the sample mean ($\bar{x}$) and the population mean (μ).

2. Calculate the test statistic t by using $t = \dfrac{\bar{x} - \mu}{\dfrac{s}{\sqrt{n-1}}}$,

where $\bar{x}$ = sample mean, μ = population mean (or hypothetical mean of population)

s = sample standard deviation, n = sample size

3. Find the value of $|t|$ and compare this value with the table value of t with $\nu = (n-1)$ degrees of freedom at α level of significance *i.e.*, $t_\nu(\alpha)$.

(*i*) If the calculated value of $|t| <$ table value of $t_\nu(\alpha)$, then accept the null hypothesis H_0 (*i.e.*, difference is not significant) at α level of significance.

(*ii*) If the calculated value of $|t| >$ table value of $t_\nu(\alpha)$, then reject the null hypothesis H_0 and accept the alternative hypothesis H_a (*i.e.*, the difference is significant) at α level of significance.

The following figure exhibits the acceptance and rejection regions at α level of significance.

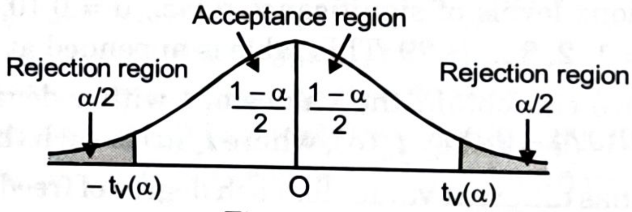

Fig. 9.5

Remark : For a two-tailed t-test, the table value of $\frac{\alpha}{2}$ is taken for α. Thus at 5% significance level *i.e.*, for $\alpha = 0{\cdot}05$ and $\nu = 9$, the table value of $t_9(0{\cdot}025)$ should be taken. This is so because at α level of significance, $\frac{\alpha}{2}$ of area is located in each tail of the curve (Ref. fig. 9.5).

9.21.1. Confidence Intervals for Population Mean

The confidence interval for population mean (μ) with 95% confidence level (*i.e.*, at 5% level of significance) is given by

$$\left[\bar{x} \pm \frac{s}{\sqrt{n-1}} \,.\, t_{\nu}(0\cdot05)\right],\ \text{where } s \text{ is the sample standard deviation.}$$

$\bar{x} - \frac{s}{\sqrt{n-1}} t_{\nu}(0\cdot05)$ and $\bar{x} + \frac{s}{\sqrt{n-1}} t_{\nu}(0\cdot05)$ are called 95% confidence limits or confidence limits at 5% level of significance.

The confidence interval for population mean (μ) with 99% confidence level (*i.e.*, at 1% level of significance) is given by

$$\left[\bar{x} \pm \frac{s}{\sqrt{n-1}} \,.\, t_{\nu}(0\cdot01)\right],\ \text{where } s \text{ is the sample standard deviation.}$$

$\bar{x} - \frac{s}{\sqrt{n-1}} t_{\nu}(0\cdot01)$ and $\bar{x} + \frac{s}{\sqrt{n-1}} t_{\nu}(0\cdot01)$ are called 99% confidence limits or confidence limits at 1% level of significance.

SOLVED EXAMPLES

EXAMPLE 1. *An industry produces iron rods with mean thickness 0·025 cm. A random sample of 10 rods was found to have mean thickness of 0·024 cm with a standard deviation of 0·002 cm. Test the significance of difference between the mean thickness at 5% level of significance.*

*[Use $t_9(0{\cdot}05) = 2{\cdot}262$]** *[CBSE Sample Paper 2024]*

Solution. Here, population mean (μ) = 0·025 cm, sample mean ($\bar{x}$) = 0·024 cm, sample standard deviation (s) = 0·002 cm and sample size (n) = 10.

Define the null hypothesis H_0 : There is no significant difference between the sample mean ($\bar{x}$) and the population mean (μ)

and the alternative hypothesis H_a : There is a significant difference between the sample mean ($\bar{x}$) and the population mean (μ).

The test statistic t is given by $t = \dfrac{\bar{x} - \mu}{\dfrac{s}{\sqrt{n-1}}}$

$$\Rightarrow \quad t = \frac{0\cdot024 - 0\cdot025}{\dfrac{0\cdot002}{\sqrt{10-1}}} = -\frac{0\cdot001 \times 3}{0\cdot002} = -1\cdot5$$

$$\Rightarrow \quad |t| = 1\cdot5$$

* **Note:** For a two tailed test, table value of α/2 is taken for α. Here the given value 2·262 infact corresponds to t_9 (0·025) in the t-table.

The test statistic t follows Student's t-distribution with $\nu = n - 1 = 10 - 1$ *i.e.*, 9 degrees of freedom.

We now compare the calculated value of $|t|$ with the table value of t for 9 degrees of freedom at 5% level of significance *i.e.*, $t_9(0.05)$, which is given to be 2·262.

Clearly, $1.5 < 2.262$ *i.e.*, calculated value of $|t|$ < table value of $t_9(0.05)$.

Thus the null hypothesis H_0 is accepted at 5% level of significance.

Hence, there is no significant difference between the population mean and sample mean at 5% level of significance.

EXAMPLE 2. *A factory produces PVC pipes with mean inner diameter of 4 cm. A sample of 17 pipes gives a mean inner diameter of 4·02 cm and a standard deviation of 0·09 cm. Is the difference in the value of means significant ? Test at 5% level of significance.* [*Use* $t_{16}(0.05) = 2.120$]

Solution. Here, population mean $(\mu) = 4$ cm, sample mean $(\bar{x}) = 4.02$ cm

sample standard deviation $(s) = 0.09$ cm and sample size $(n) = 17$

Define the null hypothesis H_0 : There is no significant difference between the sample mean $(\bar{x})$ and the population mean (μ)

and the alternative hypothesis H_a : There is a significant difference between the sample mean $(\bar{x})$ and the population mean (μ).

The test statistic t is given by $t = \dfrac{\bar{x} - \mu}{\dfrac{s}{\sqrt{n-1}}}$

$$\Rightarrow \qquad t = \frac{4.02 - 4}{\dfrac{0.09}{\sqrt{17-1}}} = \frac{0.02 \times 4}{0.09} = 0.889$$

The test statistic t follows Student's t-distribution with $\nu = n - 1 = 17 - 1$ *i.e.*, 16 degrees of freedom.

We now compare the calculated value of $|t|$ with the table value of t for 16 degrees of freedom at 5% level of significance *i.e.*, $t_{16}(0.05)$, which is given to be 2·120.

Clearly, $0.889 < 2.120$ *i.e.*, calculated value of $|t|$ < table value of $t_{16}(0.05)$

Thus the null hypothesis H_0 is accepted at 5% level of significance.

Hence, there is no significant difference between the sample mean and the population mean at 5% level of significance.

EXAMPLE 3. *A publishing company launched a new book and distributed it to the retail shops for selling it. Before an advertisement campaign, the mean sales per week per shop was 180 books. After the campaign, a sample of 26 shops was taken and mean sales per week was found to be 192 books with standard deviation 18. Can you consider the advertisement campaign effective at 5% level of significance ?* [*Use* $t_{25}(0.05) = 2.060$]

Solution. Here, population mean $(\mu) = 180$, sample mean $(\bar{x}) = 192$,

sample standard deviation $(s) = 18$ and sample size $(n) = 26$

Define the null hypothesis H_0: There is no significant difference between the sample mean $(\bar{x})$ and the population mean (μ)

and the alternative hypothesis H_a : There is a significant difference between the sample mean $(\bar{x})$ and the population mean (μ).

The test statistic t is given by $t = \dfrac{\bar{x}-\mu}{\dfrac{s}{\sqrt{n-1}}}$

$$\Rightarrow \quad t = \frac{192-180}{\dfrac{18}{\sqrt{26-1}}} = \frac{12 \times 5}{18} = 3{\cdot}333$$

The test statistic t follows Student's t-distribution with $\nu = n - 1 = 26 - 1$ *i.e.*, 25 degrees of freedom.

We now compare the calculated value of $|t|$ with the table value of t for 25 degrees of freedom at 5% level of significance *i.e.*, t_{25} (0·05), which is given to be 2·060.

Clearly, 3·333 > 2·060 *i.e.*, calculated value of $|t|$ > table value of t_{25} (0·05).

Thus the null hypothesis H_0 is rejected and the alternative hypothesis H_a is accepted at 5% level of significance.

Hence, there is a significant difference between the population mean and the sample mean and therefore we can conclude that the advertisement campaign is effective.

EXAMPLE 4. *A random sample of size 16 has mean 65 and the sum of squares of the deviation taken from mean is 256. Can this sample be regarded as taken from the population having 68 as mean. Test at 5% level of significance.* *[Use $t_{15}(0{\cdot}05) = 2{\cdot}131$]*

Solution. Here, population mean $(\mu) = 68$, sample mean $(\bar{x}) = 65$, $\Sigma(x_i - \bar{x})^2 = 256$ and sample size $(n) = 16$

We have $\quad s^2 = \dfrac{1}{n}\Sigma(x_i - \bar{x})^2 = \dfrac{256}{16} \quad \Rightarrow \quad s = \sqrt{\dfrac{256}{16}} = 4$

Define the null hypothesis H_0: There is no significant difference between the sample mean $(\bar{x})$ and the population mean (μ)

and the alternative hypothesis H_a : There is a significant difference between the sample mean $(\bar{x})$ and the population mean (μ).

The test statistic t is given by $t = \dfrac{\bar{x}-\mu}{\dfrac{s}{\sqrt{n-1}}}$

$$\Rightarrow \quad t = \frac{65-68}{\dfrac{4}{\sqrt{16-1}}} = -\frac{3}{4} \times \sqrt{15} = -\frac{3}{4} \times 3{\cdot}87 = -2{\cdot}90$$

$$\Rightarrow \quad |t| = 2{\cdot}90$$

The test statistic t follows Student's t-distribution with $\nu = n - 1 = 16 - 1$ *i.e.*, 15 degrees of freedom.

We now compare the calculated value of $|t|$ with the table value of t for 15 degrees of freedom at 5% level of significance *i.e.*, t_{15} (0·05), which is given to be 2·131

Clearly, 2·90 > 2·131 *i.e.*, calculated value of $|t|$ > table value of $t_{15}(0{\cdot}05)$

Thus the null hypothesis H_0 is rejected and alternative hypothesis H_a is accepted at 5% level of significance.

Hence, there is a significant difference between the population mean and the sample mean and therefore we can conclude that the sample is not taken from the population having 68 as mean.

EXAMPLE 5. *The manufacturer of electrical items makes electric tubelights and claims that his tubelights have a mean life of 25 months. A random sample of 6 such tubelights gave the following values :*

Life in months : 24, 26, 30, 20, 20, 18.

Can you regard the manufacturers claim to be valid at 1% level of significance ?

[Use $t_5\,(0{\cdot}01) = 4{\cdot}032$]

Solution. Here, population mean $(\mu) = 25$

$$\text{sample mean } (\bar{x}) = \frac{\Sigma x_i}{n} = \frac{138}{6} = 23 \text{ and sample size } (n) = 6$$

Define the null hypothesis H_0: There is no significant difference between the sample mean $(\bar{x})$ and the population mean (μ)

and the alternative hypothesis H_a: There is a significant difference between the sample mean $(\bar{x})$ and the population mean (μ).

Calculation of s :

x_i	$x_i - \bar{x}$ $(\bar{x} = 23)$	$(x_i - \bar{x})^2$
24	1	1
26	3	9
30	7	49
20	–3	9
20	–3	9
18	–5	25
$\Sigma x_i = 138$		$\Sigma (x_i - \bar{x})^2 = 102$

$$\therefore \quad s = \sqrt{\frac{\Sigma (x_i - \bar{x})^2}{n}} = \sqrt{\frac{102}{6}} = \sqrt{17} = 4{\cdot}123$$

The test statistic t is given by $t = \dfrac{\bar{x} - \mu}{\dfrac{s}{\sqrt{n-1}}}$

$$\Rightarrow \quad t = \frac{23 - 25}{\dfrac{4{\cdot}123}{\sqrt{6-1}}} = \frac{-2}{4{\cdot}123} \times 2{\cdot}236 = -1{\cdot}085 \quad \Rightarrow \quad |t| = 1{\cdot}085$$

The test statistic t follows Student's t-distribution with $\nu = n - 1 = 6 - 1$ *i.e.*, 5 degrees of freedom.

We now compare the calculated value of $|t|$ with the table value of t for 5 degrees of freedom at a 1% level of significance *i.e.*, $t_5(0.01)$, which is given to be 4.032.

Clearly, $1.085 < 4.032$ *i.e.*, calculated value of $|t|$ < table value of $t_5(0.01)$.

Thus, the null hypothesis H_0 is accepted at 1% level of significance.

Hence, there is no significant difference between the sample mean and population mean, and therefore the manufacturer's claim is valid at 1% level of significance.

EXAMPLE 6. *The lifetime of electric bulbs for a random sample of 10 bulbs from a large consignment gave the following data :*

Bulb :	1	2	3	4	5	6	7	8	9	10
Lifetime in thousand hours :	4·2	4·6	3·9	4·1	5·2	3·8	3·9	4·3	4·4	5·6

Determine whether we can accept the hypothesis that the average lifetime of bulbs is 4 thousands hours. Test at 5% level of significance. [*Use* $t_9(0.05) = 2.262$]

Solution. Here, population mean (μ) = 4, sample mean $(\bar{x}) = \frac{\Sigma x_i}{10} = \frac{44}{10} = 4.4$

and sample size (n) = 10

Define the null hypothesis H_0 : There is no significant difference between the sample mean ($\bar{x}$) and the population mean (μ)

and the alternative hypothesis H_a : There is a significant difference between the sample mean ($\bar{x}$) and the population mean (μ).

Calculation of s :

x_i	$x_i - \bar{x}$ ($\bar{x} = 4.4$)	$(x_i - \bar{x})^2$
4·2	– 0·2	0·04
4·6	0·2	0·04
3·9	– 0·5	0·25
4·1	– 0·3	0·09
5·2	0·8	0·64
3·8	– 0·6	0·36
3·9	– 0·5	0·25
4·3	– 0·1	0·01
4·4	0	0
5·6	1·2	1·44
$\Sigma x_i = 44$		$\Sigma (x_i - \bar{x})^2 = 3.12$

$$\therefore \quad s = \sqrt{\frac{\Sigma (x_i - \bar{x})^2}{n}} = \sqrt{\frac{3\cdot12}{10}} = \sqrt{0\cdot312} = 0\cdot56$$

The test statistic t is given by $t = \dfrac{\bar{x} - \mu}{\dfrac{s}{\sqrt{n-1}}}$

$$\Rightarrow \quad t = \frac{4\cdot4 - 4}{\dfrac{0\cdot56}{\sqrt{9}}} = \frac{0\cdot4 \times 3}{0\cdot56} = 2\cdot143$$

The test statistic t follows Student's t-distribution with $\nu = n - 1 = 10 - 1$ *i.e.*, 9 degrees of freedom.

We now compare the calculated value of $|t|$ with the table value of t for 9 degrees of freedom at 5% level of significance *i.e.*, $t_9(0\cdot05)$, which is given to be 2·262.

Clearly, $2\cdot143 < 2\cdot262$ *i.e.*, calculated value of $|t|$ < table value of $t_9(0\cdot05)$.

Thus the null hypothesis H_0 is accepted at 5% level of significance.

Hence, there is no significant difference between the sample mean and population mean, and therefore we can accept that the average life time of the bulbs is 4 thousand hours.

EXAMPLE 7. *A random sample of 26 values taken from a normal population has a mean of 160 cm and the sum of squares of deviation from the mean is 2450 cm^2. Is the assumption that the population mean is 165 cm valid or not? Test at 5% and 1% levels of significance. Also, obtain the 95% and 99% confidence limits.* [*Use* $t_{25}(0\cdot05) = 2\cdot060$ *and* $t_{25}(0\cdot01) = 2\cdot787$]

Solution. Here, population mean $(\mu) = 165$, sample mean $(\bar{x}) = 160$, sample size $(n) = 26$

and $\quad \sum_{i=1}^{26} (x_i - \bar{x})^2 = 2450$

$$\therefore \quad s^2 = \frac{1}{n}\sum_{i=1}^{n} (x_i - \bar{x})^2 = \frac{2450}{26} = 94\cdot23$$

$$\Rightarrow \quad s = \sqrt{94\cdot23} = 9\cdot707$$

Define the null hypothesis H_0 : There is no significant difference between the sample mean $(\bar{x})$ and the population mean (μ)

and the alternative hypothesis H_a : There is a significant difference between the sample mean $(\bar{x})$ and the population mean (μ).

The test statistic t is given by $t = \dfrac{\bar{x} - \mu}{\dfrac{s}{\sqrt{n-1}}}$

$$\Rightarrow \quad t = \frac{160 - 165}{\dfrac{9\cdot707}{\sqrt{26-1}}} = -\frac{5 \times 5}{9\cdot707} = -2\cdot575$$

$$\Rightarrow \quad |t| = 2{\cdot}575$$

The test statistic t follows Student's t-distribution with $\nu = (n-1) = 26-1$ *i.e.*, 25 degrees of freedom.

We now compare the calculated value of $|t|$ with the table value of t for 25 degrees of freedom at 5% and 1% levels of significance.

At 5% level of significance : We have $t_{25}(0{\cdot}05) = 2{\cdot}060$. [*Given*]

Clearly $2{\cdot}575 > 2{\cdot}060$ *i.e.*, calculated value of $|t| >$ table value of $t_{25}(0{\cdot}05)$

Thus the null hypothesis H_0 is rejected and the alternative hypothesis is accepted at 5% level of significance.

Hence, there is a significant difference between the sample mean and population mean and therefore our assumption that population mean is 165 cm is not valid.

The 95% confidence limits (*i.e.*, at 5% level of significance) are :

$$\bar{x} - \frac{s}{\sqrt{n-1}}\, t_{25}(0 \cdot 05) \quad \text{and} \quad \bar{x} + \frac{s}{\sqrt{n-1}}\, t_{25}(0 \cdot 05)$$

i.e., $$160 - \frac{9{\cdot}707}{5} \times 2{\cdot}060 \quad \text{and} \quad 160 + \frac{9{\cdot}707}{5} \times 2{\cdot}060$$

i.e., $$160 - 3{\cdot}999 = 156{\cdot}001 \text{ and } 160 + 3{\cdot}999 = 163{\cdot}999.$$

At 1% level of significance : We have $t_{25}(0{\cdot}01) = 2{\cdot}787$. [*Given*]

Clearly, $2{\cdot}575 < 2{\cdot}787$ *i.e.*, calculated value of $|t| <$ table value of $t_{25}(0{\cdot}01)$

Thus the null hypothesis H_0 is accepted at 1% level of significance.

Hence, there is no significant difference between the sample mean and population mean and therefore our assumption that population mean is 165 cm is valid.

The 99% confidence limits (*i.e.*, at 1% level of significance) are :

$$\bar{x} - \frac{s}{\sqrt{n-1}}\, t_{25}(0 \cdot 01) \quad \text{and} \quad \bar{x} + \frac{s}{\sqrt{n-1}}\, t_{25}(0 \cdot 01)$$

i.e., $$160 - \frac{9{\cdot}707}{5} \times 2{\cdot}787 \quad \text{and} \quad 160 + \frac{9{\cdot}707}{5} \times 2{\cdot}787$$

i.e., $$160 - 5{\cdot}411 = \mathbf{154{\cdot}589} \quad \text{and} \quad 160 + 5{\cdot}411 = \mathbf{165{\cdot}411}.$$

EXERCISE 9.2

1. Consider the following hypothesis :

$$H_0 : \mu = 200$$
$$H_a : \mu \neq 200$$

A sample of 17 items gave a mean of 210 and standard deviation 50. Test the hypothesis at 5% as well as 1% level of significance. [Use $t_{16}(0{\cdot}05) = 2{\cdot}120$, $t_{15}(0{\cdot}01) = 2{\cdot}921$]

2. An auto industry is making a bearings of engine with mean diameter of 0·7 inch. A random sample of 10 parts gives a mean diameter 0·742 inch with a standard deviation of 0·04 inch. Is the difference in the values of mean diameters significant. Test at 5% level of significance. [Use $t_9(0{\cdot}05) = 2{\cdot}262$]

3. A machine is designed to produce insulating washers for electrical devices with mean thickness 0·080 cm. A random sample of 26 washers was found to have a mean thickness of 0·075 cm with a standard deviation of 0·005 cm. Test the significance of the difference between the means at 5% level of significance. [Use $t_{25}(0{\cdot}05) = 2{\cdot}060$].

4. A machine produces washers of thickness 0.50 mm. To determine whether the machine is in proper working order, a sample of 10 washers is chosen for which the mean thickness is 0.53 mm and the standard deviation is 0.03 mm. Test the hypothesis at 5% level of significance whether the machine is working in proper order. [Given $t_9(0{\cdot}05) = 2{\cdot}262$] [*CBSE Sample Paper 2024*]

5. For a random sample of size 10 from a normal population, the mean is 12·1 and the standard deviation is 3·2. Is it reasonable to suppose that the mean of the population is 14·57 ? Test at 1% significance level. [Use $t_9(0{\cdot}01) = 3{\cdot}250$]

6. The average annual labour turnover of industrial units in a large industrial area is 320 employees. A sample of 9 industrial units taken at random gives the average annual labour turnover of 300 employees with a standard deviation of 30 employees. Examine the significance of difference between the average turnover at 5% level of significance. [Use $t_8(0{\cdot}05) = 2{\cdot}306$]

7. A random sample of size 16 has 53 as mean. The sum of the squares of the deviations taken from mean is 135. Can this sample be regarded as taken from the population having 56 as mean ? Test at 5% level of significance. [Use $t_{15}(0{\cdot}05) = 2{\cdot}13$]

8. A sample of 10 copper wires drawn from a large lot have the following breaking strength (in kg weight).

578, 572, 570, 568, 572, 578, 570, 572, 596, 544

Is the assumption that the mean breaking strength of the lot is 578 kg valid. Test at 5% level of significance. [Use $t_9(0{\cdot}05) = 2{\cdot}262$]

9. The I.Q. scores of a random sample of 14 boys taken from a population were found to be :

70, 120, 110, 101, 88, 83, 95, 98, 107, 100, 128, 90, 82, 86

Is the assumption that the mean I.Q. of the population is 100, valid or not. Test at 5% level of significance. [Use $t_{13}(0{\cdot}05) = 2{\cdot}160$]

10. Ten students are selected at random from a college and their heights are found to be 100, 104, 108, 110, 118, 120, 122, 124, 126 and 128 cms. Can you claim that the mean height of the students of college is 110 cm. Test at 5% level of significance. [Use $t_9(0{\cdot}05) = 2{\cdot}262$]

11. A random sample of the heights of 10 students from a large number of students in a school have a mean height of 5·6 ft. with a standard deviation of 0·75 ft. Find (*i*) 95% and (*ii*) 99% confidence limits for the average height of all the students of the school. [Use $t_9(0{\cdot}05) = 2{\cdot}262$; $t_9(0{\cdot}01) = 3{\cdot}250$]

ANSWERS

1. $|t| = 0{\cdot}8$; No significant difference at both 5% and 1% level of significance
2. $|t| = 3{\cdot}15$; difference is significant
3. $|t| = 5$; difference is significant
4. $|t| = 3$; No
5. $|t| = 2{\cdot}32$; Yes
6. $|t| = 1{\cdot}885$; no significant difference
7. $|t| = 4$; No
8. $|t| = 1{\cdot}488$; Valid
9. $|t| = 0{\cdot}716$; Valid
10. $|t| = 1{\cdot}94$; Yes
11. (*i*) 5·035 ; 6·165 (*ii*) 4·788 ; 6·413

9.22. ▶ TEST OF SIGNIFICANCE OF THE DIFFERENCE BETWEEN THE MEANS OF TWO INDEPENDENT RANDOM SAMPLES

Let the two independent random samples $\{x_1, x_2, x_3, \ldots, x_{n_1}\}$ and $\{y_1, y_2, y_3, \ldots, y_{n_2}\}$ of sizes n_1 and n_2 with means $\bar{x}$ and $\bar{y}$ and standard deviation s_1 and s_2 respectively, be drawn from a same normal population. To determine if there is a significant difference between the two sample means $\bar{x}$ and $\bar{y}$, we proceed as follows :

1. Define the null hypothesis H_0 : There is no significant difference between the means of two samples

and the alternative hypothesis H_a : There is a significant difference between the means of two samples.

2. Calculate the test statistic t by using $t = \dfrac{\bar{x} - \bar{y}}{S\sqrt{\dfrac{1}{n_1} + \dfrac{1}{n_2}}}$

or $t = \dfrac{\bar{x} - \bar{y}}{S} \times \sqrt{\dfrac{n_1 n_2}{n_1 + n_2}}$, where S (combined standard deviation) $= \sqrt{\dfrac{n_1 s_1^2 + n_2 s_2^2}{n_1 + n_2 - 2}}$

Here, the test statistic t follows Student's t-distribution with $\nu = n_1 + n_2 - 2$ degrees of freedom.

3. Find the value of $|t|$ and compare this value with the table value of t for $\nu = (n_1 + n_2 - 2)$ degrees of freedom at α level of significance *i.e.*, $t_\nu(\alpha)$.

(*i*) If the calculated value of $|t| <$ table value of $t_\nu(\alpha)$, then accept the null hypothesis H_0 (*i.e.*, the difference between the mean of two samples is not significant) at α level of significance.

(*ii*) If the calculated value of $|t| >$ table value of $t_\nu(\alpha)$, then reject the null hypothesis H_0 and accept the alternative hypothesis H_a (*i.e.*, the difference between the mean of two samples is significant) at α level of significance.

Remark : We have $s_1^2 = \dfrac{1}{n_1}\sum_{i=1}^{n_1} (x_i - \bar{x})^2$ and $s_2^2 = \dfrac{1}{n_2}\sum_{i=1}^{n_2} (y_i - \bar{y})^2$

$$\therefore \qquad S = \sqrt{\frac{n_1 s_1^2 + n_2 s_2^2}{n_1 + n_2 - 2}} = \sqrt{\frac{\sum_{i=1}^{n_1} (x_i - \bar{x})^2 + \sum_{i=1}^{n_2} (y_i - \bar{y})^2}{n_1 + n_2 - 2}}$$

SOLVED EXAMPLES

EXAMPLE 1. *For the following data, examine if the means of two random samples differ significantly at 5% level of significance :* [*Use* $t_{25}(0.05) = 2.060$]

	Size	*Mean*	*Standard deviation*
Sample I :	10	200	10
Sample II :	17	220	8

Solution. Here $n_1 = 10,\ \bar{x} = 200,\ s_1 = 10$ and $n_2 = 17,\ \bar{y} = 220,\ s_2 = 8$

$$\therefore \qquad S = \sqrt{\frac{n_1 s_1^{\,2} + n_2 s_2^{\,2}}{n_1 + n_2 - 2}} = \sqrt{\frac{10 \times (10)^2 + 17 \times (8)^2}{10 + 17 - 2}}$$

$$= \sqrt{\frac{1000 + 1088}{25}} = \sqrt{\frac{2088}{25}} = \sqrt{83 \cdot 52} = 9 \cdot 14$$

Define the null hypothesis H_0 : There is no significant difference between the means of two samples

and the alternative hypothesis H_a : There is a significant difference between the means of two samples.

The test statistic t is given by $t = \dfrac{\bar{x} - \bar{y}}{S} \times \sqrt{\dfrac{n_1 n_2}{n_1 + n_2}}$

$$\Rightarrow \qquad t = \frac{200 - 220}{9 \cdot 14} \times \sqrt{\frac{10 \times 17}{10 + 17}} = -\frac{20}{9 \cdot 14} \times \sqrt{\frac{170}{27}} = -\frac{20 \times 2 \cdot 51}{9 \cdot 14} = -5 \cdot 492$$

$\Rightarrow |t| = 5 \cdot 942$

The test statistic t follows Student's t-distribution with $\nu = (n_1 + n_2 - 2) = (10 + 17 - 2)$ *i.e.*, 25 degrees of freedom.

We now compare the calculated value of $|t|$ with the table value of t for 25 degrees of freedom at 5% level of significance *i.e.*, $t_{25}(0 \cdot 05)$, which is given to be $2 \cdot 060$.

Clearly, $5 \cdot 942 > 2 \cdot 060$ *i.e.*, calculated value of $|t|$ > table value of $t_{25}\ (0 \cdot 05)$.

Thus, the null hypothesis H_0 is rejected and alternative hypothesis H_a is accepted at 5% level of significance.

Hence, there is a significant difference between the sample means of two random samples.

EXAMPLE 2. *The mean of I.Q.'s of a group of 18 students from one area of a city was found to be 95 with a standard deviation of 10, while the mean of I.Q.'s of a group of 12 students from another area of a city was found to be 100 with a standard deviation of 8. Test whether there is a significant difference between the I.Q.'s of two groups of students at*

(i) 1% level of significance *(ii) 5% level of significance.*

[Use $t_{28}\ (0 \cdot 01) = 2 \cdot 763,\ t_{28}\ (0 \cdot 05) = 2 \cdot 048$]

Solution. Here $n_1 = 18,\ \bar{x} = 95,\ s_1 = 10$ and $n_2 = 12,\ \bar{y} = 100,\ s_2 = 8$

$$\therefore \qquad S = \sqrt{\frac{n_1 s_1^{\,2} + n_2 s_2^{\,2}}{n_1 + n_2 - 2}} = \sqrt{\frac{18 \times (10)^2 + 12 \times (8)^2}{18 + 12 - 2}}$$

$$= \sqrt{\frac{1800 + 768}{28}} = \sqrt{\frac{2568}{28}} = 9 \cdot 58$$

Define the null hypothesis H_0 : There is no significant difference between I.Q.'s of two groups of students.

and the alternative hypothesis H_a : There is a significant difference between I.Q.'s of two groups of students.

The test statistic t is given by $t = \dfrac{\bar{x} - \bar{y}}{S} \times \sqrt{\dfrac{n_1 n_2}{n_1 + n_2}}$

$$\Rightarrow \quad t = \frac{95 - 100}{9 \cdot 58} \times \sqrt{\frac{18 \times 12}{18 + 12}} = -\frac{5}{9 \cdot 58} \times \sqrt{\frac{216}{30}} = -\frac{5}{9 \cdot 58} \times 2 \cdot 68 = -1 \cdot 398$$

$$\Rightarrow \quad |t| = 1 \cdot 398$$

The test statistic t follows Student's t-distribution with $\nu = 18 + 12 - 2$ *i.e.*, 28 degrees of freedom.

We now compare the calculated value of $|t|$ with the table value of t for 28 degrees of freedom at 1% and 5% level of significance.

At 1% level of significance : We have $t_{28}(0 \cdot 01) = 2 \cdot 763$. *[Given]*

Clearly, $1 \cdot 398 < 2 \cdot 763$ *i.e.*, calculated value of $|t|$ < table value of $t_{28}(0 \cdot 01)$.

Thus, the null hypothesis H_0 is accepted at 1% level of significance.

Hence, there is no significant difference between I.Q.'s of two groups of students at 1% level of significance.

At 5% level of significance : We have $t_{28}(0 \cdot 05) = 2 \cdot 048$. *[Given]*

Clearly, $1 \cdot 398 < 2 \cdot 048$ *i.e.*, calculated value of $|t|$ < table value of $t_{28}(0 \cdot 05)$.

Thus, the null hypothesis H_0 is accepted at 5% level of significance.

Hence, there is no significant difference between I.Q.'s of two groups of students at 5% level of significance.

EXAMPLE 3. *The following information was obtained related to the sales of a product in two shops in Delhi and Mumbai :*

Delhi :	$\bar{x} = 3 \cdot 5$	$\Sigma x_i = 40$	$\Sigma x_i^2 = 275$	$n_1 = 12$
Mumbai :	$\bar{y} = 5$	$\Sigma y_i = 48$	$\Sigma y_i^2 = 235$	$n_2 = 10$

Is there a evidence of significant difference in the sales in Delhi and Mumbai. Test at 5% level of significance. *[Use $t_{20}(0 \cdot 05) = 2 \cdot 086$]*

Solution. We have $S^2 = \dfrac{n_1 s_1^2 + n_2 s_2^2}{n_1 + n_2 - 2}$

Now, $s_1^2 = \dfrac{1}{n_1} \Sigma(x_i - \bar{x})^2$ and $s_2^2 = \dfrac{1}{n_2} \Sigma(y_i - \bar{y})^2$

$$\Rightarrow \quad S^2 = \frac{\Sigma(x_i - \bar{x})^2 + \Sigma(y_i - \bar{y})^2}{n_1 + n_2 - 2}$$

$$\therefore \quad S^2 = \frac{(\Sigma x_i^2 - n_1 \bar{x}^2) + (\Sigma y_i^2 - n_2 \bar{y}^2)}{n_1 + n_2 - 2}$$

$$\left[\begin{aligned} \because \Sigma(x_i - \bar{x})^2 &= \Sigma x_i^2 - 2\bar{x}\, \Sigma x_i + \Sigma \bar{x}^2 \\ &= \Sigma x_i^2 - 2\bar{x} \,.\, n_1 \bar{x} + n_1 \bar{x}^2 \\ &= \Sigma x_i^2 - 2n_1 \bar{x}^2 + n_1 \bar{x}^2 \\ &= \Sigma x_i^2 - n_1 \bar{x}^2 \end{aligned} \right]$$

$$= \frac{[275 - 12 \times (3 \cdot 5)^2] + [235 - 10 \times (5)^2]}{12 + 10 - 2}$$

$$= \frac{275 - 147 + 235 - 250}{20} = \frac{113}{20} = 5 \cdot 65$$

$$\Rightarrow \quad S = \sqrt{5 \cdot 65} = 2 \cdot 377$$

Define the null hypothesis H_0 : There is no significant difference between the sales in Delhi and Mumbai.

and the alternative hypothesis H_a : There is a significant difference between the sales in Delhi and Mumbai.

The test statistic t is given by $t = \frac{\bar{x} - \bar{y}}{S} \times \sqrt{\frac{n_1 n_2}{n_1 + n_2}}$

$$\Rightarrow \quad t = \frac{3 \cdot 5 - 5}{2 \cdot 377} \times \sqrt{\frac{12 \times 10}{12 + 10}}$$

$$\Rightarrow \quad t = -\frac{1 \cdot 5}{2 \cdot 377} \times \sqrt{\frac{120}{22}} = -\frac{1 \cdot 5}{2 \cdot 377} \times 2 \cdot 34 = -1 \cdot 477$$

$$\Rightarrow \quad |t| = 1 \cdot 477$$

The test statistic t follows Student's t-distribution with $\nu = 12 + 10 - 2$ *i.e.*, 20 degrees of freedom.

We now compare the calculated value of $|t|$ with the table value of t for 20 degrees of freedom at 5% level of significance *i.e.*, $t_{20}(0 \cdot 05)$, which is given to be $2 \cdot 086$.

Clearly, $1 \cdot 477 < 2 \cdot 086$ *i.e.*, calculated value of $|t|$ < table value of $t_{20}(0 \cdot 05)$.

Thus, the null hypothesis H_0 is accepted at 5% level of significance.

Hence, there is no evidence of significant difference in the sales in Delhi and Mumbai.

EXAMPLE 4. *Two laboratories A and B carry out independent estimates of protein content in a mixture made by a firm. A sample is taken from each batch, halved, and the separated halves are sent to the two laboratories. The protein content obtained by the laboratories is recorded below :*

Sample No. :	1	2	3	4	5	6	7	8	9	10
Lab. A :	5	6	5	3	7	8	10	4	4	8
Lab. B :	10	6	6	7	8	7	8	9	3	6

Test whether there is a significant difference between the mean protein content obtained by the two laboratories A and B, at 5% level of significance. [*Use* $t_{18}(0 \cdot 05) = 2 \cdot 101$]

Solution. Define the null hypothesis H_0 : There is no significant difference between the mean protein content obtained by the two laboratories A and B.

and the alternative hypothesis H_a : There is a significant difference between the mean protein content obtained by the two laboratories A and B.

Here $\bar{x} = \frac{\Sigma x_i}{n_1} = \frac{60}{10} = 6$ and $\bar{y} = \frac{\Sigma y_i}{n_2} = \frac{70}{10} = 7$

We now compute the values of $\Sigma (x_i - \bar{x})^2$ and $\Sigma (y_i - \bar{y})^2$ from the following tables :

x_i	$x_i - \bar{x}$ $(\bar{x} = 6)$	$(x_i - \bar{x})^2$
5	−1	1
6	0	0
5	−1	1
3	−3	9
7	1	1
8	2	4
10	4	16
4	−2	4
4	−2	4
8	2	4
$\Sigma x_i = 60$		$\Sigma (x_i - \bar{x})^2 = 44$

y_i	$y_i - \bar{y}$ $(\bar{y} = 7)$	$(y_i - \bar{y})^2$
10	3	9
6	−1	1
6	−1	1
7	0	0
8	1	1
7	0	0
8	1	1
9	2	4
3	−4	16
6	−1	1
$\Sigma y_i = 70$		$\Sigma (y_i - \bar{y})^2 = 34$

Now, $$S^2 = \frac{\Sigma(x_i - \bar{x})^2 + \Sigma(y_i - \bar{y})^2}{n_1 + n_2 - 2}$$

$$\Rightarrow \quad S^2 = \frac{44 + 34}{10 + 10 - 2} = \frac{78}{18} \quad \Rightarrow \quad S = \sqrt{\frac{78}{18}} = 2{\cdot}08$$

The test statistic t is given by $t = \dfrac{\bar{x} - \bar{y}}{S} \times \sqrt{\dfrac{n_1 n_2}{n_1 + n_2}}$

$$\Rightarrow \quad t = \frac{6 - 7}{2{\cdot}08}\sqrt{\frac{10 \times 10}{10 + 10}} = -\frac{1}{2{\cdot}08}\sqrt{\frac{100}{20}} = -\frac{1}{2{\cdot}08} \times 2{\cdot}24 = -1{\cdot}077$$

$$\Rightarrow \quad |t| = 1{\cdot}077$$

The test statistic t follows Student's t-distribution with $\nu = 10 + 10 - 2$ *i.e.*, 18 degrees of freedom.

We now compare the calculated value of $|t|$ with the table value of t for 18 degrees of freedom at 5% level of significance *i.e.*, $t_{18}(0{\cdot}05)$, which is given to be $2{\cdot}101$.

Clearly, $1{\cdot}077 < 2{\cdot}101$ *i.e.*, calculated value of $|t|$ < table value of $t_{18}(0{\cdot}05)$.

Thus the null hypothesis H_0 is accepted at 5% level of significance.

Hence, there is no significant difference between the mean protein content obtained by the two laboratories A and B.

EXAMPLE 5. *The marks obtained by two groups of students in an examination are as follows :*

Group A :	*31*	*38*	*26*	*34*	*32*	*18*	*24*	*26*	*30*	*31*	*21*	*25*			
Group B :	*44*	*24*	*33*	*10*	*16*	*37*	*40*	*37*	*26*	*43*	*40*	*21*	*23*	*19*	*22*

Examine the significance of difference between the mean marks of students of the two groups at 5% level of significance. [*Use* $t_{25}(0{\cdot}05) = 2{\cdot}060$]

Solution. Define the null hypothesis H_0 : There is no significant difference between the mean marks of students of two groups

and the alternative hypothesis H_a : There is a significant difference between the mean marks of students of two groups.

Here, $n_1 = 12$, $n_2 = 15$, $\bar{x} = \frac{\Sigma x_i}{n_1} = \frac{336}{12} = 28$ and $\bar{y} = \frac{\Sigma y_i}{n_2} = \frac{435}{15} = 29$

We now compute the values of $\Sigma (x_i - \bar{x})^2$ and $\Sigma (y_i - \bar{y})^2$ from the following tables :

x_i	$x_i - \bar{x}$ ($\bar{x} = 28$)	$(x_i - \bar{x})^2$
31	3	9
38	10	100
26	−2	4
34	6	36
32	4	16
18	−10	100
24	−4	16
26	−2	4
30	2	4
31	3	9
21	−7	49
25	−3	9
$\Sigma x_i = 336$		$\Sigma (x_i - \bar{x})^2 = 356$

y_i	$y_i - \bar{y}$ ($\bar{y} = 29$)	$(y_i - \bar{y})^2$
44	15	225
24	−5	25
33	4	16
10	−19	361
16	−13	169
37	8	64
40	11	121
37	8	64
26	−3	9
43	14	196
40	11	121
21	−8	64
23	−6	36
19	−10	100
22	−7	49
$\Sigma y_i = 435$		$\Sigma (y_i - \bar{y})^2 = 1620$

Now, $$S^2 = \frac{\Sigma(x_i - \bar{x})^2 + \Sigma(y_i - \bar{y})^2}{n_1 + n_2 - 2} = \frac{356 + 1620}{12 + 15 - 2} = \frac{1976}{25}$$

$$\Rightarrow \quad S = \sqrt{\frac{1976}{25}} = 8{\cdot}89$$

The test statistic t is given by $t = \frac{\bar{x} - \bar{y}}{S} \times \sqrt{\frac{n_1 n_2}{n_1 + n_2}}$

$$\Rightarrow \quad t = \frac{28 - 29}{8{\cdot}89}\sqrt{\frac{12 \times 15}{12 + 15}} = -\frac{1}{8{\cdot}89}\sqrt{\frac{180}{27}} = -\frac{1}{8{\cdot}89} \times 2{\cdot}58 = -0{\cdot}290$$

$$\Rightarrow \quad |t| = 0{\cdot}290$$

The test statistic t follows Student's t-distribution with $\nu = 15 + 12 - 2$ *i.e.*, 25 degrees of freedom.

We now compare the calculated value of $|t|$ with the table value of t for 25 degrees of freedom at 5% level of significance *i.e.*, $t_{25}(0.05)$, which is given to be 2.060.

Clearly, $0.290 < 2.060$ *i.e.*, calculated value of $|t|$ < table value of $t_{25}(0.05)$.

Thus the null hypothesis H_0 is accepted at 5% level of significance.

Hence, there is no significant difference between the mean marks of two group of students.

EXERCISE 9.3

1. For the following data examine if the means of two random samples differ significantly at 5% level of significance.

	Size	Mean	Standard deviation
Sample I :	6	40	8
Sample II :	5	50	10

[Use t_9 (0.05) = 2.262]

2. The sales in shops of two towns for a new product gave the following result :

Town	Mean Sales	Variance	Sample Size
A :	$\bar{x} = 57$	$s_1^2 = 5.3$	$n_1 = 5$
B :	$\bar{y} = 61$	$s_2^2 = 4.8$	$n_2 = 7$

Is there a significant difference between the sales in shops of town A and B ? Test the hypothesis at 5% level of significance.
[Use t_{10} (0.05) = 2.228]

3. The mean produce of wheat samples of 10 fields is 145 kg per acre with a standard deviation of 42 kgs. The mean produce of another samples of 17 fields gives a mean of 128 kg per hectare with a standard deviation of 39 kgs. Is there a significant difference between the mean of two samples. Test at 5% level of significance.
[Use t_{25} (0.05) = 2.060]

4. In a test attempted by two groups of students, the following marks were obtained :

First group :	18	20	36	50	49	36	34	49	41
Second group :	29	28	26	35	30	44	46		

Test the significance of difference between the mean of the marks secured by the students of the above two groups at 5% level of significance.
[Use t_{14}(0.05) = 2.145]

5. The heights (in inches) of 6 randomly chosen sailors are : 63, 67, 69, 72, 75, 74.

The heights (in inches) of 10 randomly chosen soldiers are : 61, 62, 65, 66, 69, 71, 70, 71, 72, 73.

Examine if there is a significant difference between the average heights of the sailors and soldiers. Test at 5% level of significance.
[Use t_{14} (0.05) = 2.145]

6. The weights (in kg) of workers in the factories of two cities is as given below :

City A :	59	48	49	46	66	69	74	42	68	56	72				
City B :	55	50	49	66	62	68	71	74	64	75	78	72	70	71	65

Is there a significant difference between the mean weights of the workers in the factories of two cities ? Test at 5% level of significance.
[Use t_{25} (0.01) = 2.787]

7. Two types of drugs were used on 5 and 7 patients for reducing their weight. The decrease in the weight (in pounds) after using the drugs for six months was as follows :

Drug A :	10	12	13	11	14		
Drug B :	8	9	12	14	15	10	9

Is there a significant difference in the efficacy of the two drugs? Test at 5% level of significance.
[Use t_{10} (0·05) = 2·223]

8. The increase in weights (in pounds) of 7 persons fed on diet A and 5 persons fed on diet B is as follows :

Diet A :	12	15	11	16	14	14	16
Diet B :	8	10	14	10	13		

Test whether diets A and B differ significantly as regards their effect on increase in weight at 5% level of significance. [Use t_{10} (0·05) = 2·228]

ANSWERS

1. $|t| = 1{\cdot}66$; No significant difference
2. $|t| = 2{\cdot}792$; Yes
3. $|t| = 1{\cdot}023$; No significant difference
4. $|t| = 0{\cdot}571$; No significant difference
5. $|t| = 0{\cdot}89$; Yes
6. $|t| = 1{\cdot}768$; No significant difference
7. $|t| = 0{\cdot}735$; No significant difference
8. $|t| = 2{\cdot}39$; No significant difference

COMPETENCY BASED QUESTIONS

[MCQ's–Fill in the Blanks–Very Short Answer Type–Assertion-Reason Based–Case Study Based Questions]

Type I : Multiple Choice Questions

1. A random sample of mean $\bar{x}$ is chosen from a population with mean μ and variance σ^2. Which of the following is a statistic :

(a) μ (b) $\bar{x}$ (c) σ^2 (d) none of these

[CBSE Sample Paper 2023]

2. Standard deviation of a sample from a population is called a

(a) Standard error (b) Parameter
(c) Statistic (d) Central limit

3. In a school, a random sample of 145 students is taken to check whether a student's average calory intake is 1500 or not. The collected data of average calories intake of sample students is presented in a frequency distribution, which is called a :

(a) Statistics (b) Sampling distribution
(c) Parameter (d) Population sampling

[CBSE Sample Paper 2023]

4. If $H_0 : \mu = \mu_0$, then for a right tailed test, H_a will be

(a) $\mu \neq \mu_0$ (b) $\mu < \mu_0$ (c) $\mu > \mu_0$ (d) none of these

5. The degree of significance with which we reject the given hypothesis is known as

(a) confidence limit (b) confidence interval
(c) level of significance (d) none of these

6. The critical value of Z for a right tailed test at 5% level of significance is

(a) 2·33 (b) 1·65 (c) 1·96 (d) – 1·65

7. If the sample size is large, then the shape of sampling distribution is approximately equal to

(a) normal distribution (b) *t*-distribution
(c) binomial distribution (d) none of these

8. If $\bar{x}$ is the mean of a sample of size *n* taken from a population with mean μ, then the estimation for population mean is given by

(a) $\dfrac{\bar{x}-\mu}{\frac{S}{\sqrt{n}}}$ (b) $\dfrac{\bar{x}-\mu}{\frac{S}{n}}$ (c) $\dfrac{\bar{x}-\mu}{\frac{S^2}{n}}$ (d) none of these

[CBSE Sample Paper 2024]

9. For the purpose of *t*-test of significance, a random sample of size 34 is drawn from a normal population. Then the degree of freedom (ν) is

(a) $\dfrac{1}{34}$ (b) 33 (c) 34 (d) 35

[CBSE Sample Paper 2023]

10. For a student *t*-test, the test statistic *t* is given by

(*a*) $t = \bar{x} - \mu$ (*b*) $t = \frac{\bar{x}}{s}$ (*c*) $t = \frac{\bar{x} - \mu}{s}$ (*d*) $t = \frac{\bar{x} - \mu}{\frac{s}{\sqrt{n-1}}}$

11. The range of variable *t* of the *t*-distribution is

(*a*) (0, 1) (*b*) $(-\infty, \infty)$ (*c*) (– 1, 1) (*d*) (1, 2)

12. The mean of *t*-distribution is

(*a*) 0 (*b*) 1 (*c*) 2 (*d*) not defined

13. If the calculated value of $|t| < t_v(\alpha)$, then the null hypothesis is

(*a*) rejected (*b*) accepted

(*c*) cannot be determined (*d*) neither accepted nor rejected

[*CBSE 2023*]

14. If the null hypothesis is $H_0 : \mu = \mu_0$, then the alternative hypothesis H_a for a left tailed test is

(*a*) $\mu < \mu_0$ (*b*) $\mu > \mu_0$ (*c*) $\mu \neq \mu_0$ (*d*) $\mu = \mu_0$

15. If the sample size (*n*) is 28, then the degrees of freedom (ν) is equal to

(*a*) 28 (*b*) 25 (*c*) 24 (*d*) 27

16. For testing the significance of difference between the means of two independent samples, the degrees of freedom (ν) is taken as

(*a*) $n_1 - n_2 + 2$ (*b*) $n_1 - n_2 - 2$ (*c*) $n_1 + n_2 - 2$ (*d*) $n_1 + n_2 + 2$

[*CBSE 2023*]

17. The test statistic *t* for testing the significance of difference between the means of two independent samples is given by

(*a*) $t = \frac{\bar{x} - \bar{y}}{\sqrt{S}}$ (*b*) $t = \frac{\bar{x} - \bar{y}}{S\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}}$ (*c*) $t = \frac{\bar{x} - \bar{y}}{\frac{S}{\sqrt{n-1}}}$ (*d*) $t = \frac{\bar{x} + \bar{y}}{S\sqrt{\frac{1}{n_1} - \frac{1}{n_2}}}$

18. A machine makes car wheels and in a random sample of 26 wheels, the test statistic (*t*) is found to be 3.07. As per the *t*-distribution test (at 5% level of significance), what can you say about the quality of wheels produced by the machine? (Use $t_{25}(0.05) = 2.06$)

(*a*) superior quality (*b*) inferior quality

(*c*) same quality (*d*) cannot say

[*CBSE Sample Paper 2023*]

Type II. Fill in the Blanks:

19. A collection of objects, experiments, events etc. is called __________.

20. Any finite set of statistical individuals drawn from a population for investigation is called __________.

21. The number of statistical individuals contained in a sample is called __________.

22. A measurable characteristic of population is called __________.

23. A measurable characteristic of sample is called __________.

24. In a hypothesis test, the significance level α denotes the probability of making __________ error.

25. A hypothesis which suggests that there is no significant difference between the sample statistic and the population parameter is called ________.
26. A hypothesis which contradicts the null hypothesis is called ________.
27. The sampling error can be reduced by increasing the ________.
28. The critical region is also known as ________ region.
29. The number of independent values which have the freedom to vary in computing a statistic is called ________.

Type III : Very Short Answer Questions :

30. Give the difference between population and sample.
31. Briefly explain the various types of sampling.
32. Define parameter and statistic.
33. State Central limit theorem.
34. Briefly explain the errors in testing of hypothesis ?
35. What do you understand by level of significance ?
36. Give four properties of t-distribution.
37. Write the confidence interval for the population mean with 95% confidence level.

Type IV - Assertion-Reason Based Questions :

In the following questions a Statement of Assertion (A) is followed by a Statement of Reason (R). Choose the correct option from the following options :

(*a*) Both Assertion (A) and Reason (R) are true and Reason (R) is the correct explanation of Assertion (A).

(*b*) Both Assertion (A) and Reason (R) are true but Reason (R) is not the correct explanation of Assertion (A).

(*c*) Assertion (A) is true but Reason (R) is false.

(*d*) Assertion (A) is false but Reason (R) is true.

1. **Assertion (A) :** The variance of a population consisting of four observations 1, 3, 5, 7 is 5.

   Reason (R) : Variance $= \dfrac{\Sigma(x_i - \bar{x})^2}{n}$.

2. **Assertion (A) :** Consider the following hypothesis test : $H_0 : \mu \leq 25$; $H_a : \mu > 25$
   If a sample of 40 items is taken whose mean is 26.4 and the population standard deviation is 6, then the value of test statistic Z is 1.48.

   Reason (R) : The value of the test statistic Z is given by $Z = \dfrac{\bar{x} - \mu_0}{\dfrac{\sigma}{\sqrt{n}}}$ where $\bar{x}$ = sample mean,

   μ_0 = hypothetised value of population mean, σ = population standard deviation, n = size of the sample.

3. **Assertion (A) :** Critical values of Z for two-tailed tests at 5% and 1% level of significance respectively are $|Z_\alpha| = 2{\cdot}58$ and $|Z_\alpha| = 1{\cdot}960$.

   Reason (R) : The set of values for the test statistic Z at which the null hypothesis H_0 is rejected is known as the critical region or rejection region.

4 **Assertion (A) :** If null hypothesis H_0 is $\mu = \mu_0$, then the alternative hypothesis may be $H_a : \mu \neq \mu_0$ (*i.e.*, $\mu > \mu_0$ or $\mu < \mu_0$)

Reason (R) : Any hypothesis which contradicts the null hypothesis H_0 is called alternative hypothesis (denoted by H_a).

5. **Assertion (A) :** From a population having mean 0.7, a sample of size 10 is taken whose mean is 0.742 and standard deviation is 0.04. Then the value of test statistic t is 3.15.

Reason (R) : The test statistic t is given by $t = \dfrac{\bar{x} - \mu}{\dfrac{s}{\sqrt{n-1}}}$

6. **Assertion (A) :** For the following data, the combined standard deviation is 17.92.

	Size	Mean	Standard deviation
Sample I :	20	400	20
Sample II :	34	440	16

Reason (R) : Combined standard deviation = $\sqrt{\dfrac{n_1 s_1^2 + n_2 s_2^2}{n_1 + n_2}}$, where n_1, n_2 and s_1, s_2 are the sizes and standard deviations of two samples.

7. **Assertion (A) :** The mean of t distribution is 0.

Reason (R) : The t distribution curve is symmetrical about the line $t = 0$.

ANSWERS

Type I - Multiple Choice Questions :

1. (*b*) 2. (*c*) 3. (*b*) 4. (*c*) 5. (*c*)
6. (*b*) 7. (*a*) 8. (*a*) 9. (*b*) 10. (*d*)
11. (*b*) 12. (*a*) 13. (*b*) 14. (*a*) 15. (*d*)
16. (*c*) 17. (*b*) 18. (*b*)

Type II - Fill in the Blanks :

19. population 20. sample 21. sample size 22. parameter
23. statistic 24. Type I 25. null hypothesis 26. alternative hypothesis
27. sample size 28. rejection 29. degrees of freedom

Type III - Very Short Answer Questions :

37. $\left[\bar{x} \pm \dfrac{s}{\sqrt{n-1}} t_v(0 \cdot 05)\right]$

Type IV - Assertion-Reason Based Questions :

1. (*a*) 2. (*a*) 3. (*d*) 4. (*a*)
5. (*a*) 6. (*c*) 7. (*b*)