Chapter 6(1) Test of Hypothesis

6(1).01 Introduction

Hypothesis tests are widely used in business and industry for making decisions. It is here that probability and sampling theory plays an ever increasing role in constructing the criteria on which business decisions are made. For example, in order to increase consumer awareness of a product or service, it might be necessary to compare the effectiveness of different types of advertising campaigns. Or, in order to offer more profitable investment opportunities to its customers, an investment firm might wish to compare the profitability of different types of investment of portfolios.

If a sample contains 30 or more ($n \geq 30$) observations, then it is called large sample. A study of test statistic for large samples is called large sample statistic. In this chapter we study only normal test (z -test) for large sample ($n \geq 30$) and also three important test statistic such as students t-test, F-test and χ^2-test for small sample ($n < 30$). In most introductory statistics books, n=30 is treated as the break point between large and small samples.

6(1).02 Some Necessary Definitions:

Hypothesis:

Any statement about the population is called hypothesis. In general word, hypothesis is a pre-assumption.

Hypothesis is a statement about a population parameter developed for the purpose of testing. In other words, a hypothesis is a conclusion which is tentatively drawn on logical basis.

Parametric Hypothesis :

Any hypothesis about the parameter of a population distribution is known as parametric hypothesis.

For example : Suppose, the average age of all students in a college is 20 years. The hypothesis about their mean value is called parametric hypothesis.

Non-Parametric Hypothesis :

Any hypothesis about a population distribution is called a non-parametric hypothesis.

For example, a company's owner claim that the weight of his produced commodity follows the normal distribution. The hypothesis about the weight of his produced commodity is called non-parametric hypothesis.

Test of significance Or, Hypothesis testing Or, Test of hypothesis:

Test of hypothesis is a statistical procedure that is used to provide evidence in favour of some statement (called a hypothesis). The purpose of hypothesis testing is to determine whether a claimed (hypothesized) value for a population parameter. Hypothesis testing permits the scientist to mak generalizations about populations from sample data.

The test of hypothesis discloses the fact whether the difference between the computed statistic and the hypothetical parameter is significant or otherwise. Hence, the test of hypothesis is also known as the test of significance.

Hypothesis tests are widely used in business and industry for making decisions. It is here that probability and sampling theory plays an ever increasing role in constructing the criteria on which business decisions are made. For example, in order to increase consumer awareness of a product or service, it might be necessary to compare the effectiveness of different types of advertising campaigns. Or, in order to offer more profitable investment opportunities to its customers, an investment firm might wish to compare the profitability of different types of investment of portfolios.

Statistical hypothesis:

Statistical hypothesis is some assumption or statement, which may or may not be true, about a population or about the probability distribution characterising the given population, which we want to test on the basis of the evidence from a random sample. A statistical hypothesis is a statement about the probability distribution of a set of random variables.

There are two types of statistical hypothesis :

(i) Null hypothesis

(ii) Alternative hypothesis

(i) Null hypothesis : The statistical hypothesis that is set up for testing a hypothesis is known as null hypothesis. The null hypothesis is set up in testing a statistical hypothesis only to decide whether to accept or reject the null hypothesis.

The hypothesis that we assume is said to be null hypothesis. A null hypothesis sates that there is no difference between a sample estimate and the true population value. The null hypothesis is generally denoted by Ho.

According to professor R. A. fisher "Null hypothesis is the hypothesis which is to be tested for possible rejection under the assumption it is true."

The steps of the null hypothesis : The following steps must be taken into consideration while setting up a null hypothesis.

(i) In order to test the significance of the difference between a sample statistic and the population parameter or between the two different sample statistical we set up the null hypothesis to that the difference is not significant. There are may be some difference but that is solely due to sampling fluctuations.

(ii) In order to test any statement about the population, we hypothesis that is true.

(ii) Alternative hypothesis : The negative of null hypothesis is called the alternative hypothesis. In other words, any hypothesis which is not a null hypothesis is called an alternative hypothesis.

A statistical hypothesis that disagrees with the null hypothesis or that is simply opposite of the null hypothesis is said to be alternative hypothesis. It is denoted by H_1 or Ha.

For example (i) If $H_0 : \mu = 0$ then the alternative hypothesis may be

$$H_1 : \mu \neq 0,\ H_1 : \mu > 0 \text{ or } H_1 : \mu < 0$$

(ii) If $H_0 : \mu_1 = \mu_2$, then the alternative hypothesis may be

$$H_1 : \mu_1 \neq \mu_2,\ H_1 : \mu_1 > \mu_2 \text{ or } H_1 : \mu_1 < \mu_2 \text{ etc.}$$

Type I Error :
In a test of hypothesis, the type I error occurs when the null hypothesis H_0 is rejected although it was true. The probability of a type I error is denoted by α, α is also known as the level of significance.

Type II error :
In a test of hypothesis, the type II error occurs when the null hypothesis H_0 is not rejected (i.e. accepted) although it was false. The probability of type II error is denoted by β

Table for Decision from Sample

Decision	State of Nature	
	H_0 true	H_0 false
Reject H_0	Type I error	Correct decision
Do not reject H_0	Correct decision	Type II error

Level of Significance :
The level of significance is the maximum probability of making a type I error. Generally, it is specified in a test; if there rests no specification then we consider the level of significance $\alpha = 0.05$. ~~Thus~~ It implies that we are 95% (or 0.95) confident about the significance of our decision and we have 5% (or 0.05) chance to occur a type I error.

Power of a test :
Null hypothesis is not really truth but the probability of the rejected null hypothesis on the basis of sample data, is called power of a test.
If the probability of type II error is β in the test of significance, then $(1-\beta)$ is called power of a test.
Mathematically, if P (Ho is not truth | H_1 is truth) = β , $\beta = P(H_0$ accepted | H_0 is false)
then the power of a test = P (Ho rejected / H_1 truth) = $1 - \beta$ = P(rejecting H_0 | H_0 is false)

Test statistic :
Any mathematical function of the elements belong to a sample is called as statistic of that sample. Suppose, a sample $(x_1, x_2, \text{------}\ x_n)$ is taken from a population $(X_1, X_2 \text{------} X_N)$. Then any function of the sample $(x_1, x_2, \text{-------}, x_n)$ such as, $T = f(x_1, x_2, \text{-------}, x_n)$ or $\bar{x} = \frac{\Sigma x_i}{n}$ etc. are called statistic of the sample $(x_1, x_2, \text{-----}, x_n)$. We can find the estimated value of the parameter by statistic.
In a test of hypothesis, the test statistic is a random variable whose value is computed from the sample data and used to reject or do not reject the null hypothesis.

Rejection region (or Critical region) : A rejection (or critical) region is a set of possible values of the test statistic (in a test of hypothesis) that leads the null hypothesis to be rejected.

Acceptance Region : An acceptance region is a set of possible values of the test statistic (in a test of hypothesis) that leads the null hypothesis to be accepted.

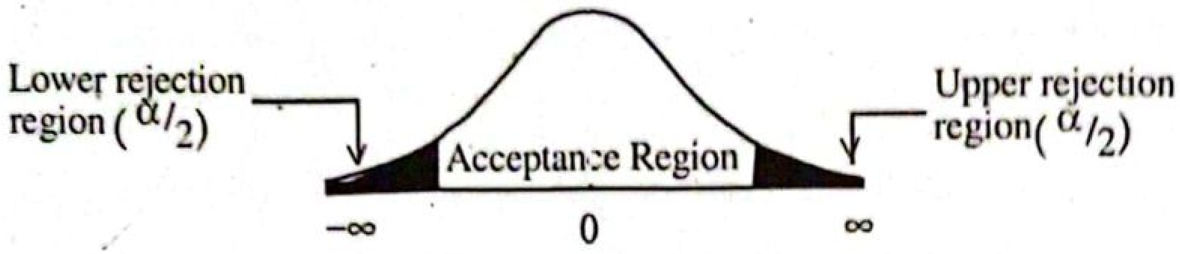


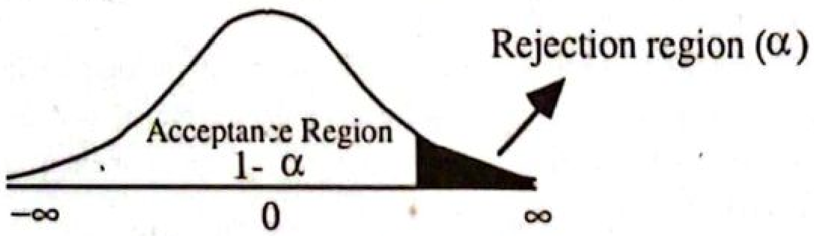
Fig. : Acceptance and rejection region.

One-tailed test (One-sided test) :

An one-tailed test is such a hypothesis test for which the rejection region consists in only one-side (or direction).

There are two types of one tailed test, such as :

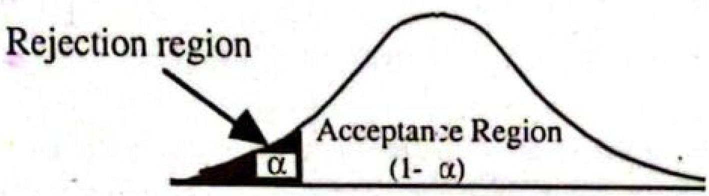
(i) Right-tailed test : A test of any statistical hypothesis where the alternative is right -tailed such as $H_0 : \mu = \mu_o$ vs $H_1 : \mu > \mu_o$ is called a right-tailed test.



The normal curve for right tailed test.

(ii) Left-tailed test : A test of any statistical hypothesis where the alternative is left - sided such as -

$H_o : \mu = \mu_o$ vs $H_1 : \mu < \mu_o$ is called a left-tailed test.



The normal curve for left-tailed test.

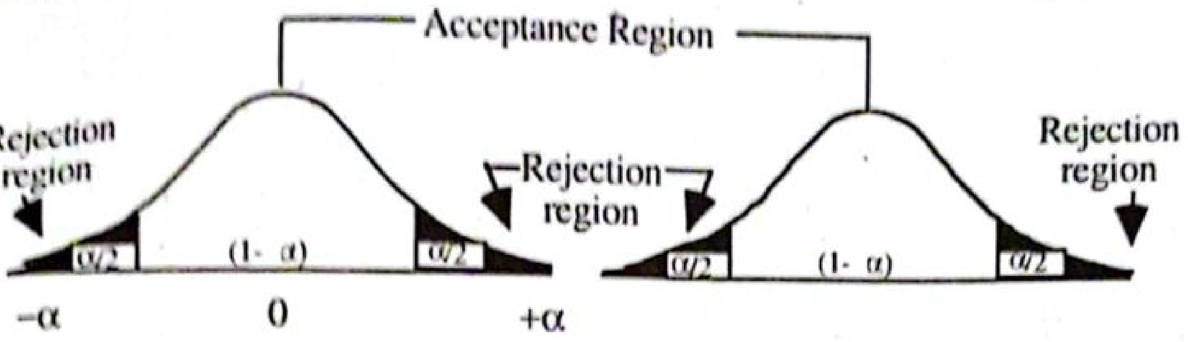
Two-tailed test (Two sided Test) :

A two-tailed test is such a hypothesis test for which the rejection region consists in both the side or direction.

Definition : A test of any statistical hypothesis where the alternative is two-sided such as -

$H_o : \mu = \mu_o$

$H_1 : \mu \neq \mu_o$ ($\mu < \mu_o$ or $\mu > \mu_o$)



The normal curve for two-tailed test.

Degrees of Freedom (df.) : By degrees of freedom we mean the number of cases to which the values can be assigned arbitrarily without violating the restrictions placed. For example, if we are to choose 5 numbers with their total 120. Then the number of restriction imposed is 1 and that is, their total 120.' Here, although we are to choose 5 numbers but, actually we can choose only 4 numbers arbitrarily. Here, the degrees of freedom, $v = n - r = 5 - 1 = 4$, where r refers to the number of independent constraints.

In general, when we fit a binomial distribution, the number of degrees of freedom is 1 less than the number of classes (cases). When we fit a poisson distribution, the degrees of freedom is 2 less than the number of classes (or cases) because we use here two restrictions (the total frequency and arithmetic mean). And when we use a normal distribution, the degrees of freedom is lessened by 3 because we use here 3 restrictions (total frequency, mean, standard deviation).

6(1).03 Procedure of Testing a Hypothesis Details :

The following are the steps involved in general procedure for hypothesis testing :

Step 1 : Setting up of hypothesis :

There are two type of hypothesis. Such that:

(i) Null hypothesis and (ii) Alternative hypothesis

(i) Null hypothesis : The statistical hypothesis that is set up for testing a hypothesis is known as null hypothesis. The null hypothesis is set up in testing a statistical hypothesis only to decide whether to accept or reject the null hypothesis.

The hypothesis that we assume is said to be null hypothesis. A null hypothesis sates that there is no difference between a sample estimate and the true population value. The null hypothesis is generally denoted by Ho. According to professor R. A. fisher "Null hypothesis is the hypothesis which is to be tested for possible rejection under the assumption it is true."

(ii) Alternative hypothesis : The negative of null hypothesis is called the alternative hypothesis. In other words, any hypothesis which is not a null hypothesis is called an alternative hypothesis.

A statistical hypothesis that disagrees with the null hypothesis or that is simply opposite of the null hypothesis is said to be alternative hypothesis. It is denoted by H_1 or Ha.

For example (i) If $H_0 : \mu = 0$ then the alternative hypothesis may be

$H_1 : \mu \neq 0, H_1 : \mu > 0$ or $H_1 : \mu < 0$

(ii) If $H_0 : \mu_1 = \mu_2$, then the alternative hypothesis may be

$H_1 : \mu_1 \neq \mu_2, H_1 : \mu_1 > \mu_2$ or $H_1 : \mu_1 < \mu_2$ etc.

Step 2 : Set up a suitable level of significance :
The level of significance usually denoted by α is the maximum probability o making type I error. The commonly used levels of significance are 5% (0.05) and 1% (0.01).

When we take 5% level of significance then there are 5 chances out of 100 that we would reject the null hypothesis when it should be accepted, i.e. we are about 95% confident that we have made the right decision. When the null hypothesis is rejected at $\alpha = 0.05$, the test result is said to be 'significant'.

Step 3 : Computation of test statistic :
After setting up the null hypothesis and alternative hypothesis, we compute the test statistic. The test statistic is a statistic based on appropriate probability distribution. It is used to test whether the null hypothesis set up should be accepted or rejected.

Many of the test statistic that we shall encounter will be of the following form :

$$\text{Test statistic} = \frac{\text{Sample Statistic} - \text{Hypothetical Population parameter}}{\text{Standard error of sample statistic}}$$

The following are the most commonly used test statistic.

The normal test or the z – statistic is defined as $z = \dfrac{u-E(u)}{S.E(u)}$

Here, u = statistic, E(u) =The expected value of the statistic u, S.E(u)= Standard error of the statistic u.
In this case, the test statistic z is normally distributed with mean zero and variance one.
If the sample size is small (n < 30), then we used the student's t- statistic is defined as.

$$t = \frac{u-E(u)}{\hat{\sigma}(u)} \; ; \quad \text{with d.f. } v = n - 1$$

In briefly, (a) If the population standard deviation is known, then we use z - test.
(b) If the population standard deviation is not known and sample size is large ($n \geq 30$), then we use also z - test.

(c) If the population standard deviation is not known and sample size is small (n < 30), then we use the t- test.

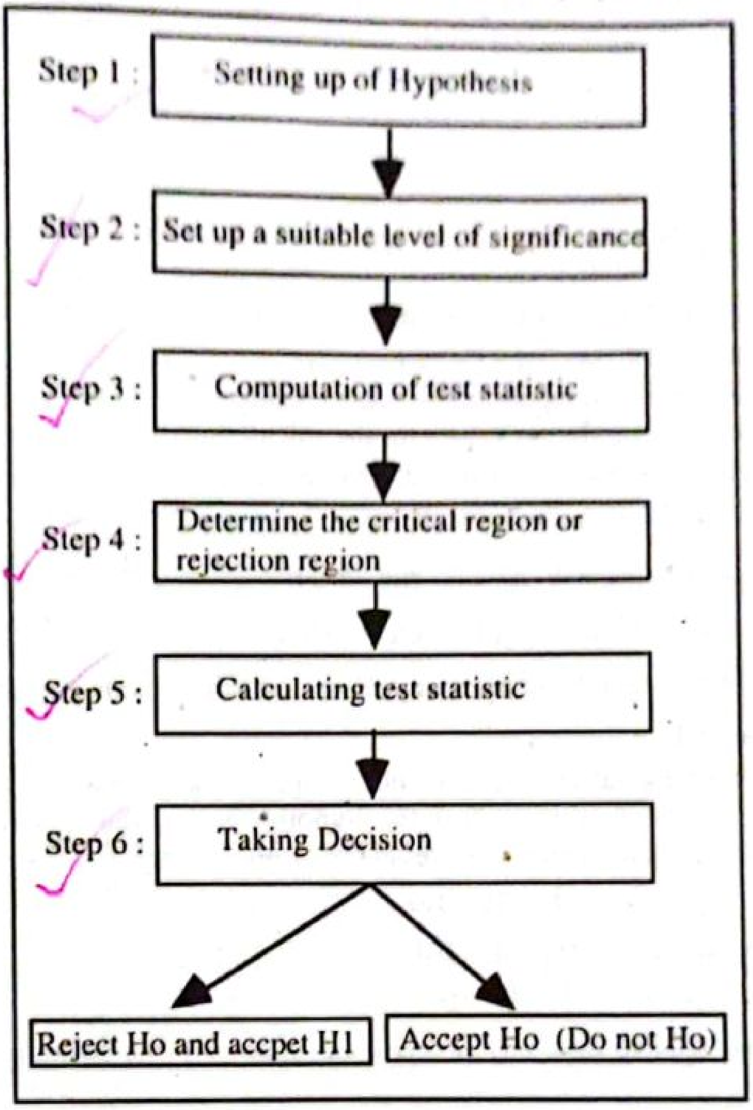
Step 4 : Determine the critical region or rejection region :
The rejection region or critical region is the region of the standard normal curve corresponding to a predetermined level of significance α (which is fixed for knowing the probability of making a type I error. i.e. rejecting the null hypothesis Ho when it is true).
The region under the normal curve which is not covered by the rejection region is known as acceptance region. The value of α, the level of significance, indicate the important that one attaches to the consequences associated with incorrectly rejecting Ho. In general, one uses a level of significance of α =5%, = 0.05, indicating that one is willing to accept a 5% chance of being wrong to reject Ho.

Step 5 : Calculating test statistic :
In this step hypothesis testing is the performance of various computations from a sample of size n, the necessary for the test statistic obtained in step 3. Then, we need to see whether sample result falls in the critical region or in the acceptance regions.

Step 6 :Taking decision : If the computed value of the test statistic is less than the critical value, then the computed value of the test statistic falls in the acceptance region and the null hypothesis is accepted. If the computed value of the test statistic is greater than the critical value, then the computed value of the test statistic falls in the rejection region and the null hypothesis is rejected.

computed value < critical value ; accepted Ho.
computed value > critical value ; rejected Ho.

In brief, we may present the general procedure of hypothesis testing by flow chart in the following way



6(1).04 Procedure of Testing a Hypothesis in Brief :

We may perform a test of hypothesis very easily by the following necessary seven steps below :

Step 1: State the available information, statistical assumption and population parameter.

Step 2 : Specify the null and alternative hypothesis.

Step 3 : Specify the significance level α.

Step 4 : Select the test statistic and state it's sampling distribution under Ho.

Step 5 : Formulate the acceptance and rejection region.

Step 6 : Compute the test statistic from the data.

Step 7 : Draw a conclusion (or decision), that is, whether the null hypothesis has been rejected or not and then the necessary comment.

6(1).05 Some Important Test of Significance :

The important tests of significance in statistics might be perfofmed by the following tests :

(a) Normal test

(b) t- test

(c) χ^2 - test

(d) F- test

6(1).06 Normal Test : Normal test is widely used in testing hypothesis regarding means, proportions, coefficient of correlation etc.

Assumptions for Normal Test : The necessary assumptions for normal test is as follows :

(i) The random sampling distribution of a statistic is approximately normal.

(ii) The values provided by the sample data are sufficiently close to the population values. And the values can be used instead of population values for the calculation of the standard error of the estimate.

By Chebyshev's Inequality we know that if a statistic be u, it's expected value E(u) and estimated or known standard error S. E. (u) then the test statistic

$z = \frac{u-E(u)}{S.E(u)}$ is normally distributed with mean 0 and variance 1. i.e., $z \sim N(0, 1)$. When the sample size n is large ($n \geq 30$) this approximation is quite satisfactory.

Usually, normal tests are two-tailed test but many times one-tailed test is also appropriate.

A guide to decision rule :

A statistical test may be attempted at several levels of significance, the most common levels being 1% and 5%. For example, suppose want to test the hypothesis that the population mean μ equals a specified value μ_o against μ is not equal to or less than or greater than μ_o. Stated symbolically,

(i) $H_0 : \mu = \mu o$ against $H_1 : \mu \neq \mu o$

(ii) $H_0 : \mu = \mu o$ against $H_1 : \mu < \mu o$

(iii) $H_0 : \mu = \mu o$ against $H_1 : \mu > \mu o$

at two different levels of significance, viz α =0.05 and α = 0.01. We provide below a table to guide you to take decision.

	Decision rule		
	$\alpha = 0.05$	$\alpha = 0.01$	P-value
H_1	Reject Ho if	Reject Ho if	
(i) $\mu \neq \mu_o$ (For left or right tailed test)	z >1.96 or z<-1.96 (i.e., | z | > 1.96)	z>2.58 or z<- 2.58 (i.e., | z | > 2.58)	Under the standard normal curve to the right of | z |
(ii) $\mu < \mu_o$ (For left tailed test)	$z \leq - 1.64$	$z \leq - 2.33$	The area under the standard normal curve to the left of z
(iii) $\mu > \mu_o$ (For right tailed test)	$z \geq 1.64$	$z \geq 2.33$	The area under the standard normal curve to the right of z.

■ **Uses of normal test or z-test**

Normal test is very popular and widely used method in statistical test of hypothesis. In general, the normal test is used in the following important cases :

(i) Normal test is used in testing hypothesis regarding the single population mean with its specified value from sample.

(ii) It is used to test the difference between the two independent population means.

(iii) It is used to test of significance about the population proportion.

(iv) It is used to test the equality of the two population proportions.

(v) It is used to test the correlation coefficient.

(vi) It is used to test the equality of the two population correlation coefficients.

Example 1 : A sample of 400 items is taken from a population whose standard deviation is 1.5. The mean of the sample is 25. Test whether the sample has come from a population with mean 26.8.

Solution : We want to test the null hypothesis, $H_0 : \mu = 26.8$

against $H_1 : \mu \neq 26.8$

Since the sample size, n > 30. So, we may use normal test.

We know, the test statistic,

$$z = \frac{\overline{x} - \mu}{\sigma/\sqrt{n}} \sim N(0,1)$$

cal > tab , Ho rejected

Where, $\overline{x}$ = Sample mean = 25, μ = Population mean = 26.8

σ = Population standard deviation = 1.5, n = Sample size = 400

$$\therefore\ z = \frac{25 - 26.8}{1.5/\sqrt{400}} = \frac{-1.8}{1.5/20} = \frac{-1.8}{0.075} = -24$$

$\therefore\ |z| = 24$

The critical value of z at $\alpha = 5\% = 0.05$ level of significance $|z_{0.025}| = 1.96$ (For two tailed test)

Since, $|z| > |z_{0.025}|$. so, the null hypothesis is rejected and the test is significant.

Comment : The sample has not come from a population with mean 26.8.

6(1).07 Tests of Hypothesis Concerning Large Samples:

If a sample contains 30 or more ($n \geq 30$) observations, then it is called large sample. A study of test statistic for large samples is called large sample statistic. In this chapter we study only normal test (z -test) for large sample ($n \geq 30$).

Tests of hypothesis involving large samples are based on the following assumptions:

(i) The random sampling distribution of a statistic is approximately normal.

(ii) The values provided by the sample data are sufficiently close to the population values. And the values can be used instead of population values for the calculation of the standard error of the estimate.

In this case, the normal distribution plays a vital role in tests of hypothesis based on large samples (Central limit theorem).

Let us consider that $\hat{\mu}$ is an unbiased estimate of the population parameter μ. On the basis of $\hat{\mu}$, taken from sample observations, it is to test the hypothesis whether the sample is drawn from a population whose parameter μ.

We want to test, null hypothesis. Ho : $\mu = \hat{\mu}$.

If sampling distribution of μ is normal, then $z = \dfrac{\hat{\mu} - \mu}{\sigma_{\hat{\mu}}} \sim N(0,1)$.

Let us consider test the hypothesis at $100\alpha\%$ level of significance.
From the tables of area under the standard normal curve corresponding to given α, we can determine an ordinate z_α such that

$$P[|z_\alpha| > z_\alpha] = \alpha$$

$$\Rightarrow P[-z_\alpha \le z \le z_\alpha] = 1 - \alpha$$

If $\alpha = 0.01$, then $z_\alpha = 2.58$ and if $\alpha = 0.05$, then $z_\alpha = 1.96$ and so on.

If the difference between $\hat{\mu}$ and μ is more than z_α times, the standard error of $\hat{\mu}$, the difference is regarded significant and Ho is rejected at $100\alpha\%$ level of significance.

Again if the difference between $\hat{\mu}$ and μ is less than or equal to z_α times, the standard error of $\hat{\mu}$, the difference is regarded insignificant and Ho is accepted at $100\alpha\%$ level of significance.

6(1).08 Testing Hypothesis about Population Mean:

Let, a random sample having n observed values $x_1, x_2, \text{-----} x_n$ respectively, which is selected from a normal distribution $N(\mu, \sigma^2)$. We have to test this sample is selected from that normal distribution with mean μ_0.

That is, we want to test, null hypothesis. Ho : $\mu = \mu_0$ (the mean value of the given population)

Where, alternative hypothesis, $H_1 : \mu \neq \mu_0$ (For two tailed test)

$H_1 : \mu < \mu_0$ (For left tailed test)

$H_1 : \mu > \mu_0$ (For right tailed test)

Suppose, sample variance, S^2, which is the estimated value of the population variance σ^2.

i. e. $s^2 = \dfrac{\Sigma(x_i - \overline{x})^2}{n-1} = \dfrac{1}{n-1}\{\Sigma x_i^2 - \dfrac{(\Sigma x_i)^2}{n}\}$

Where, sample mean $\overline{x} = \dfrac{\Sigma x_i}{n}$

In this case, for testing the null hypothesis the selection of the test statistic z-test is given below :

(a) If the population variance (σ^2) is known, then the test statistic, $z = \dfrac{\overline{x} - \mu}{\sigma/\sqrt{n}} \sim N(0,1)$

(b) If the population variance (σ^2) is not known and the sample size is large ($n \ge 30$), then the test statistic, $z = \dfrac{\overline{x} - \mu}{s/\sqrt{n}} \sim N(0, 1)$

Decision :

(a) For testing the two tailed, that is, $H_1 : \mu \neq \mu_0$.

We may reject the null hypothesis if $z < -z_{\alpha/2}$ or, $z > z_{\alpha/2}$ that is, $|z| > z_{\alpha/2}$

Otherwise, we accept the null hypothesis

(b) For testing the left tailed, that is, $H_1 : \mu < \mu_0$.

If $z < -z_\alpha$, then the null hypothesis is rejected. Otherwise, the null hypothesis is accepted.

(c) In the case of the right tailed test, i.e. $\mu > \mu_0$.

If $z > z_\alpha$, then the null hypothesis is rejected. Otherwise, the null hypothesis is accepted.

Example- 2 : A sample of 400 items is taken from a population whose standard deviation is 1.5. The mean of the sample is 25. Test whether the sample has come from a population with mean 26.8.

Solution : Given, $\overline{x} = 25$, $\mu = 26.8$, $\sigma = 1.5$, $n = 400$

We want to test the null hypothesis, $H_0 : \mu = 26.8$

against $H_1 : \mu \neq 26.8$

Since, the sample size, $n > 30$. So, we may use normal test.

$$\therefore \text{Test statistic, } z = \frac{\overline{x} - \mu}{\sigma/\sqrt{n}} \sim N(0,1)$$

$$= \frac{25 - 26.8}{1.5/\sqrt{400}} = \frac{-1.8}{1.5/20} = \frac{-1.8}{0.075} = -24$$

$$\therefore |z| = 24$$

The critical value of z at $\alpha = 5\% = 0.05$ level of significance $|z_{0.025}| = 1.96$ (For two tailed test)

Since, $|z| > |z_{0.025}|$. so, the null hypothesis is rejected and the test is significant.

Comment : The sample has not come from a population with mean 26.8.

6(1).09 Testing Hypothesis about the difference between two means:

Let, the two random samples having n_1 and n_2 observations $x_{11}, x_{12}, \text{-----------}, x_{1n_1}$ and $x_{21}, x_{22}, \text{--------}, x_{2n_2}$ are selected from the two independent normal distribution with their population means are $\mu_1 . \mu_2$ and variances are σ_1^2, σ_2^2 respectively.

We want to test, the population means are equal. That is, Ho : $\mu_1 = \mu_2$.

Suppose the estimated two sample variance s_1^2 and s_2^2 of the population variance σ_1^2 and σ_2^2 respectively.

That is, $$s_1^2 = \frac{\Sigma (x_{1i} - \overline{x}_1)^2}{n_1 - 1} = \frac{1}{n_1 - 1} \left\{ \Sigma x_{1i}^2 - \frac{(\Sigma x_{1i})^2}{n_1} \right\} ,$$

$$s_2^2 = \frac{\Sigma (x_{2i} - \bar{x}_2)^2}{n_2 - 1} = \frac{1}{n_2 - 1}\left\{\Sigma x_{2i}^2 - \frac{(\Sigma x_{2i})^2}{n_2}\right\}$$

Where, the two sample means are $\bar{x}_1$ and $\bar{x}_2$ respectively.

In order to test the null hypothesis, we can used the any one of the following test statistic;

(a) When the population variances (σ_1^2 and σ_2^2) are known, then $z = \dfrac{\bar{x}_1 - \bar{x}_2}{\sqrt{\dfrac{\sigma_1^2}{n_1} + \dfrac{\sigma_2^2}{n_2}}}$

(b) When the population variances (σ_1^2 and σ_2^2) are unknown and sample size large ($n_1 \geq 30$ and $n_2 \geq 30$), then test statistic, $z = \dfrac{\bar{x}_1 - \bar{x}_2}{\sqrt{\dfrac{s_1^2}{n_1} + \dfrac{s_2^2}{n_2}}} \sim N(0,1)$.

Decision : For test statistic z , if $z < -z_{\alpha/2}$ or, $z > z_{\alpha/2}$ that is, $|z| > z_{\alpha/2}$, then the null hypothesis rejected. Otherwise, the null hypothesis is accepted.

Problem- 3 : A man buys 50 electric bulbs of 'philips' and 50 electric bulbs of 'Osram'. He finds that 'philips' bulbs give an average life of 1500 hours with a standard deviation of 60 hours and 'Osram' bulbs give an average life of 1512 hours with a standard deviation of 80 hours. Is there a significant difference in the mean life of the two makes of bulbs?

Solution : We want to test the null hypothesis, $H_0 : \mu_1 = \mu_2$

i.e., there is no significant difference in the mean life of the two makes of bulbs.

Against $H_1 : \mu_1 \neq \mu_2$

We know, the test statistic, $z = \dfrac{\bar{x}_1 - \bar{x}_2}{\sqrt{\dfrac{s_1^2}{n_1} + \dfrac{s_2^2}{n_2}}} \sim N(0,1)$

Here, $n_1 = 50$, $\bar{x}_1 = 1500$, $s_1 = 60$

$n_2 = 50$, $\bar{x}_2 = 1512$, $s_2 = 80$

$$\therefore \quad z = \frac{1500 - 1512}{\sqrt{\dfrac{(60)^2}{50} + \dfrac{(80)^2}{50}}} = \frac{-12}{\sqrt{72 + 128}} = \frac{-12}{14.142} = -0.849$$

$\therefore$ The calculated value of z is $|z| = 0.849$

The critical value of z at 5% level of significance is $|z_{0.025}| = 1.96$

Since $|z| < |z_{0.025}|$ So, the null hypothesis is not rejected and the test is insignificant. Hence there is no significant difference in the mean life of the two makes of bulbs.

6(1).10 Tests of Hypothesis concerning Attributes

An attribute is a characteristic of an individual which cannot be measured numerically. For example, marital status, beauty, efficiency, honesty etc. are attributes. In the same way, in a garments factory a certain item, the defectives of a unit is an attribute. As distinguished from variables where quantitative measurement of a phenomenon is possible in case of attributes we can only find out the presence or absence of a certain characteristic. For example, in the study of attribute 'married' a sample may be taken people classified as married and unmarried. With such data, the binomial type of problem may be formed. The selection of an individual on sampling may be called 'event', the appearance of an attribute "A" may taken as "success" and its non - appearance, as "failure".

The sampling distribution of the number of successes, being a binomial model with mean $\mu = np$ and standard deviation $\sigma = \sqrt{npq}$.

Then the test statistic, $z = \dfrac{x - np}{\sqrt{npq}} \sim N(0,1)$

Example-4: In 600 throws of a die, even numbers appeared 360 times. At 5% level of significance, you can say that the die is an unbiased or fair.

Solution : We want to test the null hypothesis, H_0: The die is an unbiased or fair.

From the given information, we have, $p = q = \frac{1}{2}$, n= 600

Mean, $\mu = np = 600 \times \frac{1}{2} = 300$ and standard deviation $\sigma = \sqrt{npq} = = \sqrt{600 \times \frac{1}{2} \times \frac{1}{2}} = 12.25$

Test statistic, $z = \dfrac{x - np}{\sqrt{npq}} = \dfrac{360 - 300}{12.25} = 4.9$

At 5% level of significance, the table value of z=1.96.

Since, the calculated value of z is greater than the table value of z. So, the null hypothesis is rejected. Hence, the die is not an unbiased or fair.

6(1).11 Testing Hypothesis about the population proportion:

Suppose, in a population with size N there are A observations having certain characteristics and a sample with size n is drawn from the population where there are 'm' observations having the specific (or certain) characteristics. Then the sample proportion 'p' be defined as $p = \frac{m}{n}$.

If unbiased estimator of the population proportion π is the sample proportion p, then we have to test,

Null hypothesis, Ho : $\pi = \pi_0$ (the certain proportion of the given population)

Alternative hypothesis, $H_1 : \pi \neq \pi_0$ (For two tailed test)

$H_1 : \pi < \pi_0$ (For left tailed test)

$H_1 : \pi > \pi_0$ (For right tailed test)

Test statistic, $z = \dfrac{p - \pi}{\sqrt{\dfrac{\pi(1-\pi)}{n}}} \sim N(0,1)$

Decision :

(a) For testing the two tailed, that is, $H_1 : \pi \neq \pi_0$

(i) We may reject the null hypothesis if $z < -z_{\alpha/2}$ or, $z > z_{\alpha/2}$ that is, $|z| > z_{\alpha/2}$

Otherwise, we accept the null hypothesis.

(b) For testing the left tailed, that is, $H_1 : \pi < \pi_0$

(i) If $z < -z_{\alpha}$ then the null hypothesis is rejected. Otherwise, the null hypothesis is accepted.

(c) In the case of the right tailed test, i.e. $H_1 : \pi > \pi_0$

(i) If $z > z_{\alpha}$, then the null hypothesis is rejected. Otherwise, the null hypothesis is accepted.

Example-5: In a random sample of 800 items manufactured by a factory, number of defective parts was found to be 60. The factory claims that only 7% items defective. Applying the appropriate test, verify the factory's claim.

Solution : We want to test the null hypothesis, H_0: There is no significant difference beteen the sample proportion and the population proportion.

That is, Ho : $\pi = p = 7\% = 0.07$

From the given information,

Sample proportion, $p = \dfrac{60}{800} = 0.075$

and population proportion, $\pi = 0.07$

Test statistic, $z = \dfrac{p - \pi}{\sqrt{\dfrac{\pi(1-\pi)}{n}}} = \dfrac{0.075 - 0.07}{\sqrt{\dfrac{0.07(1-0.07)}{800}}} = 0.556$

At 5% level of significance, the table value of z=1.96.

Since, the calculated value of z is less than the table value of z. So, the null hypothesis is acccepted. Hence, there is no significant difference beteen the sample proportion and the population proportion. That is, the factory's claim is justified.

6(1).12 Testing Hypothesis about the difference between two population proportions:

Let, the two independent large sample sizes n_1 and n_2 ($n_1 \geq 30$ and $n_2 \geq 30$) are selected from the two binomial population. Where there are m_1 and m_2 observations having the specific (or certain) characteristics. Suppose, the two sample proportions p_1 and p_2 observations having the specific (or certain) characteristics and the two population proportions are π_1 and π_2 respectively.

We want to test, $H_o : \pi_1 = \pi_2$

$H_1 : \pi_1 \neq \pi_2$ (For two tailed test)

$H_1 : \pi_1 < \pi_2$ (For left tailed test)

$H_1 : \pi_1 > \pi_2$ (For right tailed test)

Test statistic, $z = \dfrac{p_1 - p_2}{\sqrt{p(1-p)\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}} \sim N(0, 1)$

Where, $p_1 = \dfrac{m_1}{n_1}$, $p_2 = \dfrac{m_2}{n_2}$ and $p = \dfrac{m_1+m_2}{n_1+n_2}$

Decision :

(i) For two tailed test) (That is, $H_1 : \pi_1 \neq \pi_2$), If $z < -z_{\alpha/2}$ or $z > z_{\alpha/2}$ that is $|z| \geq z_{\alpha/2}$ then the null hypothesis is rejected. Otherwise, the null hypothesis is accepted.

(ii) For left tailed test (That is, $H_1 : \pi_1 < \pi_2$), If $z \leq -z_{\alpha}$ then the null hypothesis is rejected. Otherwise, the null hypothesis is accepted.

(iii) For right tailed test (That is, $H_1 : \pi_1 > \pi_2$), If $z \geq z_{\alpha}$ then the null hypothesis is rejected. Otherwise, the null hypothesis is accepted.

Example-6: Before an increase in excise duty on tea, 400 people out of 500 people were found to be tea drinkers. After an increase in duty, 400 people were tea drinkers in a sample of 600 people. State whether there is a significant decrease in the consumption of tea. [BBA (Hons), DU, 2014]

Solution : We want to test, $H_o : \pi_1 = \pi_2$.

That is, there is no sinificant difference in the consumption of tea after increase in duty.

Against , $H_1 : \pi_1 \neq \pi_2$

We know, the test statistic, $z = \dfrac{p_1 - p_2}{\sqrt{p(1-p)\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}} \sim N(0,1)$

Here, $n_1 = 500$, $n_2 = 600$, $m_1 = 400$, $m_2 = 400$

$\therefore \quad p_1 = \dfrac{m_1}{n_1} = \dfrac{400}{500} = 0.8, \qquad p_2 = \dfrac{m_2}{n_2} = \dfrac{400}{600} = 0.667$

$p = \dfrac{m_1 + m_2}{n_1 + n_2} = \dfrac{400+400}{500+600} = 0.73$

$\therefore \quad z = \dfrac{0.8 - 0.667}{\sqrt{0.73\,(1-0.73)\left(\frac{1}{500} + \frac{1}{600}\right)}} = \dfrac{0.133}{0.027} = 4.93$

The critical value of z at 5% level of significance is $z_{0.025} = 1.96$

Since, $z > z_{0.025}$. So, the null hypothesis is rejected. Hence, there is sinificant difference in the consumption of tea after increase in duty.

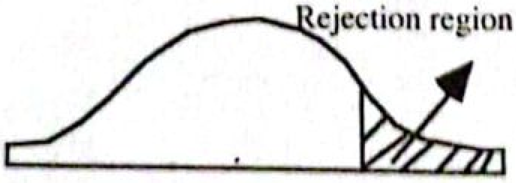
6(1).13 Distinguish between Type I error and Type II error:

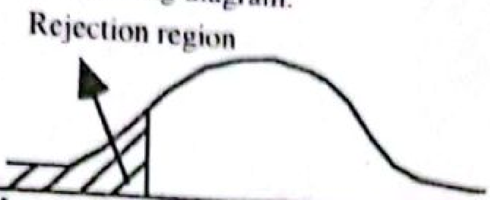
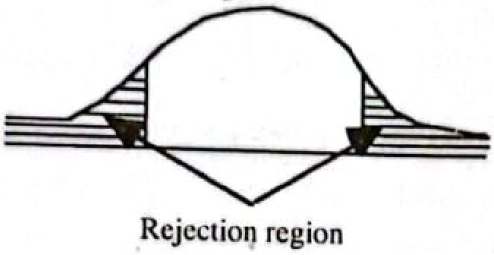
There are the following difference between Type I error and Type II error as below :

Points of difference	Type I error	Type II error
(i) Definition	In a test of hypothesis, the type I error occurs when the null hypothesis Ho is rejected although it was true.	In a test of hypothesis, the type II error occurs when the null hypothesis Ho is not rejected (i. e., accepted) although it was false.
(ii) Symbol	The probability of type I error is denoted by α.	The probability of type II error is denoted by β.
(iii) Mathematically	P(Type I error) =P (Ho rejected / Ho true) = α.	P (Type II error) = P (Ho accepted /Ho is false)=β.
(iv) Level of significance	The probability of type I error is called level of significance.	The probability of type II is not called level of significance.
(v) Uses	In order to determine the critical value of any test statistic, type I error can be used.	In order to determine the critical value of any test statistic, type II error can not be used.

6(1).15 Distinguish between rejection region and acceptance region:

There are the important difference between rejection region and acceptance region as follows :

Points of difference	Rejection Region	Acceptance region
(i) Definition	A rejection region is a set of possible values of the test statistic that leads the null hypothesis to be rejected.	An acceptance region is a set of possible values of the test statistic that leads the null hypothesis to be accepted.
(ii) Right-tailed test	In Right-tailed test, the rejection region lies on the right side of a distribution. The rejection region is shown below by the following diagram. 	In right-tailed test, the acceptance region lies on the left side of a distribution. The acceptance region is shown below by the following diagram. 

(iii) Left -tailed test	In left-tailed test, the rejection region lies on the left side of a distribution. The rejection region is shown below by the following diagram. 	In left-tailed test, the acceptance region lies on the right side of a distribution. The acceptance region is shown below by the following diagram. 
(iv) Two -tailed test	In two-tailed test, the rejection region lies on the two side of a distribution. The rejection region is shown below by the following diagram. 	In two-tailed test, the acceptance region lies on the between the two rejection region (i.e., middle position) of a distribution. The acceptance region is shown below by the following diagram. 

6(1).16 Distinguish between the null hypothesis and alternative hypothesis:

There are the following difference between null hypothesis and alternative hypothesis.

Points of difference	Null hypothesis	Alternative hypothesis
(i) Definition	The statistical hypothesis that is set up for testing a hypothesis is known as null hypothesis.	The negative of null hypothesis is called the alternative hypothesis.
(ii) Symbol	The null hypothesis is denoted by Ho.	The null hypothesis is denoted by H1 or HA or Ha.
(iii) Difference	A null hypothesis states that there is no difference between a sample estimate and the true population value.	An alternative hypothesis states that there is difference between a sample estimate and the true population value.
(iv) Complementary	Null hypothesis is not called the complementary hypothesis of the alternative hypothesis.	Alternative hypothesis is called the complementary hypothesis of the null hypothesis.
(v) Negative concept	Null hypothesis is the negative concept on behalf of hypothesis.	Alternative hypothesis is the negative concept on against of hypothesis.

6(1).17 Distinguish between parametric test and non-parametric test:

There are the following difference between parametric test and non-parametric test.

Points of difference	Parametric test	Non-parametric test
(i) Definition	The test about the parameter of a population distribution is known as parametric test.	The test about the population distribution is known as non-parametric test.
(ii) Concepts	Parametric test is a pre-conception about the population.	Non-parametric is not a pre-conception about the population.
(iii) Methods	In order to test the parametric test, the method of the normal test (z-test), t-test, F-test, χ^2 test many be used.	For testing sign test, Run test, Median test, Will Coxon test etc. may be used.
(iv) Decision	The acquired decision is more effective in the method of parametric test.	The acquired decision is less effective in the method of non-parametric test.

Mathematical Problems

The mathematical problems of this chapter can be divided into the following several formats for easy solve.

Format (1) : Tests about single population mean
or, Test for a specified mean or population mean
Format (2) : Tests about the difference between two population means.
Format (3) : Test for single population proportion
or, Test for specified proportion or pululation proportion
Format (4) : Test about the difference between two population proportions.

Format (1) : Tests about single population mean.

For population information,	For sample information,
μ = population mean,	$\overline{X}$ = Sample mean
σ = population S.D.,	s = sample S.D
N = population size,	n = sample size

Tips to intoduce the values of parameters (μ, σ) and statistic ($\overline{X}$, s)

(i) If a sample of 420 items is taken from a population (e.g., all students, all characteristics) whose mean and standard deviation are 26.8 and 1.5 respectively.

Then n = 420, μ = 26.8, σ = 1.5.

(ii) The mean and standard deviation of a random sample of 40 items are 66 and 3 respectively in a population.

Then, n = 40, $\overline{X}$ = 66, s = 3.

Selection of hypothesis :

$H_0 : \mu = *$ (Given value)

Against, $H_1 : \mu < *$ (If the given condition is less than / decrease/ not more than/ below in the question)

$H_1 : \mu > *$ (If the given condition is more than/ greater than/ increase in the question.)

$H_1 : \mu \neq *$ (There is no direction of Difference in the question.)

Selection of test statistic :

(i) $z = \dfrac{\overline{x} - \mu}{\sigma/\sqrt{n}} \sim N(0,1)$ [If the value of σ is given in the question.]

(ii) $z = \dfrac{\overline{x} - \mu}{s/\sqrt{n}} \sim N(0,1)$ [If the value of s is given in the question and $n \geq 30$]

Decision :

(i) **For left tailed test:** if (a) Calculated value < critical value that is, $z < z_\alpha$, then H_o is rejected .

(b) Calculated value > critical value, then H_0 is accepted.

(ii) **For right tailed test:** if (a) Calculated value > critical value that is, $z > z_\alpha$, then H_o is rejected .

(b) Calculated value < critical value, then H_0 is accepted.

(iii) **For two tailed test:** if (a) | Calculated value | > |critical value | i.e., $|z| > |z_{\alpha/2}|$, then H_o is rejected.

(b) |Calculated value | < |critical value | , then H_0 is accepted.

[**N.B. The confidence interval of μ is given in Chapter 10 (Large and Small Sample Estimation]**

Tips to find the critical value or table value of z:

For one tailed test,	For two tailed test
$Z_\alpha = Z_{0.10} = 1.282$ (For $\alpha = 10\%$)	$Z_{\alpha/2} = Z_{0.05} = 1.645$ (For $\alpha = 10\%$)
$Z_\alpha = Z_{0.05} = 1.645$ (For $\alpha = 5\%$)	$Z_{\alpha/2} = Z_{0.025} = 1.96$ (For $\alpha = 5\%$)
$Z_\alpha = Z_{0.025} = 1.96$ (For $\alpha = 2\frac{1}{2}\%$)	$Z_{\alpha/2} = Z_{\alpha/2} = Z_{0.01} = 2.326$ (For $\alpha = 2\%$)
$Z_\alpha = Z_{0.010} = 2.326$ (For $\alpha = 1\%$)	$Z_{\alpha/2} = Z_{0.005} = 2.576$ (For $\alpha = 1\%$ 1)
$Z_\alpha = Z_{0.005} = 2.576$ (For $\alpha = \frac{1}{2}\%$)	$Z_{\alpha/2} = Z_{0.0005} = 3.291$ (For $\alpha = 0.1\%$)

For example:

(a) The critical value of z at $\alpha = 5\% = 0.05$ level of significance is

$Z_{cri.}$ or $Z_{tab.}$ or $Z\alpha$ or $Z_{0.05} = 1.645$ (For one tailed test)

[**N.B.** If the calculated value of z is negative, then we write,

The calcualted value is $|z|$ or $|Z_{cal.}|$ or | Calculeted value of z | = *

and The critical value of z is | or $|Z_{cri.}|$ or $|Z_{0.05}|$ or | Critical value of z | = *]

(b) The critical value of z at $\alpha = 5\% = 0.05$ level of significance is

$Z_{cri.}$ or $Z_{tab.}$ or $Z_{\alpha/2}$ or $Z_{0.025} = 1.96$ (For two tailed test)

[**N.B.** If the calculated value of z is negative, then we write,

The calcualted value is $|z|$ or $|Z_{cal.}|$ or | Calculeted value of z | = *

and The critical value of z is | or $|Z_{cri.}|$ or $|Z_{0.025}|$ or | Critical value of z | = *]

Problem - (1): A sample of 400 managers is found to have a mean age of 25 years. Can it be reasonably regarded as a sample from a large population of mean age 26.8 years and standard deviation of 1.5 year?

Solution : Given, $n = 400$, $\overline{x} = 25$, $\mu = 26.8$, $\sigma = 1.5$.

We want to test the null hypothesis, $H_0 : \mu = 26.8$

Against the alternative hypothesis, $H_1 : \mu \neq 26.8$

$\therefore$ Test statistic, $z = \dfrac{\overline{x} - \mu}{\sigma/\sqrt{n}} = \dfrac{25 - 26.8}{1.5/\sqrt{400}} = \dfrac{-1.8}{1.5/20} = \dfrac{-1.8}{0.075} = -24$

$\therefore$ The calculated value of z is $|z| = 24$.

The critical value of z at $\alpha = 0.05$ level of significance is 1.96, which is less than the calculated value of z (24). So, the null hypothesis (H_0) is rejected. Hence, the given sample can not be regarded as a random sample from the large population whose mean age is 26.8 years.

Problem- (2) : Is it likely that a mean sample of 300 items, with mean 16.0 is a random sample from a large population whose mean is 16.8 and standard deviation 5.2?

Solution : Given, $\overline{x} = 16.0$, $\mu = 16.8$, $\sigma = 5.2$, $n = 300$

We want to test the null hypothesis, $H_0 : \mu = 16.8$, i.e., There is no significant difference between the sample mean and the population mean.

Against the alternative hypothesis, $H_1 : \mu \neq 16.8$.

$\therefore$ Test statistic, $z = \dfrac{\overline{x} - \mu}{\sigma/\sqrt{n}} = \dfrac{16.0 - 16.8}{5.2/\sqrt{300}} = \dfrac{-0.8}{5.2/17.321} = \dfrac{-0.8}{0.3002} = -2.665$

$\therefore$ The calculated value of z is $|z| = 2.665$.

The critical value of z at $\alpha = 0.01$ level of significance is 2.58, , which is less than the calculated value of z (2.665). So, the null hypothesis (H_0) is rejected and the test is significant. Hence, the given sample can not be regarded as a random sample from the large population whose mean is 16.8.

Problem - 3 : An analysis of the mid-term results of 50 students of 8th batch section B in statistics revealed that the average marks obtained by the students was 25 with a standard deviation 7. Can it be regarded as a sample of all the students of 8th batch section B having mean mark 20. [BBA, third semester, N. U. 2010]

Solution : From the given information, we have, $n = 50$, $\bar{x} = 25$, $s = 7$, $\mu = 20$

We want to test, $H_0 : \mu = 20$

Vs $H_1 : \mu \neq 20$

Since $n > 30$. So, we can use the normal test.

The test statistic, $z = \dfrac{\bar{x} - \mu}{s/\sqrt{n}} = \dfrac{25 - 20}{7/\sqrt{50}} = 5.05$

The calculated value of z is $z = 5.05$

The critical value of z at 5% level of significance is 1.96, which is less than the calculated value of z (3.03). So, the null hypothesis (H_0) is rejected. Hence, it can be regarded as a sample of all the students of 8th batch section B not having mean mark 20.

Problem-(4): A random sample of 100 students gave a mean weight of 58 kg with a S.D. of 4 kg. Test the hypothesis that the mean weight in the population is 60 kg.

Solution : Given, $\overline{x} = 58$, $\mu = 60$, $s = 4$ and $n = 100$

We want to test $H_0 : \mu = 60$ against $H_1 : \mu \neq 60$.

$$\therefore \text{ The test statistic, } z = \frac{\overline{x} - \mu}{s/\sqrt{n}} = \frac{58 - 60}{4/\sqrt{100}} = \frac{-2}{4/10} = \frac{-2}{0.4} = -5$$

$\therefore |z_{cal.}| = 5$

The critical value of z at 5% level of significance is $|z_{0.025}| = 1.96$

Since, $|z_{cal.}| > |z_{0.025}|$. So, the null hypothesis is rejected and the test is significant. Hence, the mean weight in the population could not be 60 kg.

Problem-(5): The standard deviation of the weight of 100 gm bread made by a certain bakery is 1 gm. On a certain day the owner doubted that the production is out of control. To check whether its products is under control, employees select a random sample of 25 breads and find that their mean weight is 99.5 gm. Test the doubt of the owner at 5% level of significance.

[B.B.S. (Accounting/ Management/Finance/Marketing),Part-3,N.U. 2004]

Solution: Let the population mean weight of breads is μ.

We want to test the null hypothesis, $H_0 : \mu = 100$

aganist, $H_1 : \mu \neq 100$

Given, $n = 25$, $\overline{x} = 99.5$, $\mu = 100$, $s = 1$

$$\therefore \text{Test statistic } z = \frac{\overline{x} - \mu}{s/\sqrt{n}} = \frac{99.5 - 100}{1/\sqrt{25}} = -2.5$$

$\therefore |z| = 2.5$

The critical value of z at 5% level of significance is $|z_{0.025}| = 1.96$.

Since $|z| > |z_{0.025}|$. So, the null hypothesis is rejected and the test is significant.

Hence, the population mean weight of bread is not 100 gm.

Problem- (6) : The mean life time of 100 florescent light bulbs produced by a company is computed to be 1570 hours with a standard deviation of 120 hours. If μ is the mean life time of all the bulbs produced by the company , test the null hypothesis $\mu = 1600$ hours against the alternative hypothesis $\mu \neq 1600$ hours using a level of significance of (i) 0.05 (ii) 0.01

Solution : Given, $\overline{X} = 1570$, $\mu = 1600$, $s = 120$, $n = 100$

We want to test $H_0 : \mu = 1600$.

$\therefore$ Test statistic, $z = \dfrac{\overline{X} - \mu}{s/\sqrt{n}} = \dfrac{1570 - 1600}{120/\sqrt{100}} = \dfrac{-30}{12} = -2.5$

$\therefore$ The calcualted value is $|z| = 2.5$

(i) The critical value of z at 5% level of significance (0.05) is $z_{0.025} = 1.96$

Since $|z_{cal.}| > |z_{0.025}|$ the null hypothesis is rejected and the test is significant.

(ii) The critical value of z at 1% level of significance (0.01) is $|z_{0.005}| = 2.58$

Since $|z| < |z_{0.005}|$ the null hypothesis is not rejected and the test is insignificant.

Problem- (7) : The mean yield (population value) for variety A is known to be 20 tons per acre. The average yield for variety B as estimated from a random sample of 64 plots is 18 tons/ acre and standard deviation is 4.8 per plot. Do the data indicate that the mean yield (population value) for variety B is also 20 tons/ acre?

Solution : Given, $\overline{X} = 18$, $\mu = 20$, $s = 4.8$, $n = 64$

We want to test, $H_0 : \mu = 20$ against $H_1 : \mu \neq 20$

We know, The test statistic, $z = \dfrac{\overline{X} - \mu}{s/\sqrt{n}} \sim N(0,1)$

$\therefore$ Test statistic, $z = \dfrac{18 - 20}{4.8/\sqrt{64}} = \dfrac{-2}{4.8/8} = \dfrac{-2}{0.6} = -3.33$

$\therefore |z| = 3.33$

The critical value of z at 5% level of signifieance is, $|z_{0.025}| = 1.96$

Since $|z| > |z_{0.025}|$. Therefore, the null hypothesis is rejected and hence the test is significant. Hence, the mean yield for variety B is not 20 tons/acre.

Problem (8) : A sample of 56 antique dealers in Bangladesh revealed the following sales last year.

Sales (in thousand Tk.)	54-56	56-58	58-60	60-62	62-64	64-66
No. of firms	10	12	20	8	4	2

Is it likely that the sample has come from the population with an average yearly sales of Tk. 55,000.

Solution : We want to test, $H_0 : \mu = 55$ (in thousand Tk.)

against $H_1 : \mu \neq 55$ (in thousand Tk.)

We know, The test statistic, $z = \dfrac{\overline{X} - \mu}{s/\sqrt{n}} \sim N(0,1)$

Here, mean, $\overline{X} = \frac{\sum fx}{N}$ and standard deviation, $s = \sqrt{\frac{\Sigma fx^2}{N} - (\frac{\Sigma fx}{N})^2}$

Table for calculation of mean and standard deviation

Sales (in thousand Tk.)	No. of firms (f)	Mid values (x)	fx	fx^2
54-56	10	55	550	30250
56-68	12	57	684	38988
58-60	20	59	1180	69620
60-62	8	61	488	29768
62-64	4	63	252	15876
64-66	2	65	130	8450
	N = 56		$\Sigma fx = 3284$	$\Sigma fx^2 = 192952$

Mean, $\overline{X} = \frac{3284}{56} = 58.64$

Standard deviation, $s = \sqrt{\frac{192952}{56} - (\frac{3284}{56})^2}$

$= \sqrt{3445.5714 - (58.6429)^2} = \sqrt{3445.5714 - 3438.9847}$

$= \sqrt{6.5867} = 2.5664$

$\therefore$ Test statistic, $z = \frac{58.64 - 55}{2.5664/\sqrt{56}} = \frac{3.64}{0.34} = 10.71$

$\therefore$ The calculated value of z = 10.71

The critical value of z at 5% level of significance is = 1.96 .

Since the calculated value of z is greater than the critical value. So, the null hypothesis is rejected. Hence, there is a significant difference between the sample average and the population average yearly sales.

Problem (9) : The mean annual turnover rate of the 200-count bottle of Bayer Aspirin is 6.0 with a standard deviation of 0.50 (This indicates that the stock of Bayer turns over on the pharmacy shelves an average of 6 times per year). It is suspected that the mean turnover has changed and it not 6.0. Use the 0.05 significance level.

(i) State the null and alternative hypothesis.

(ii) What is the probability of a type I error?

(iii) Give the formula for the test statistic.

(iv) State the decision rule.

(v) A random sample of 64 bottles of the 200-count size Bayer Aspirin showed a mean of 5.84. Shall we reject the hypothesis that the population mean is 6.0? Interpret the result.

Solution : (i) Here, the null hypothesis, H_0 : $\mu = 6.0$
and the alternative hypothesis, H_1 : $\mu \neq 6.0$.

(ii) We know, probability of type I error
= Probability of rejecting Ho (when Ho is true) = α = level of significance

For this problem, level of significance, $\alpha = 0.05$ (given)
$\therefore$ Probability of type I error = 0.05.

(iii) Since the sample size, n = 200 which is large enough. Therefore, we may use the normal test

Here the z-statistic, $z = \dfrac{\bar{X} - \mu}{\sigma/\sqrt{n}} \sim N(0,1)$

(iv) Decision Rule : We should not reject the null hypothesis if calculated z-value lies between ± 1.96 (i.e., between - 1.96 to + 1.96).
[Note : To state the decision we have to calculate the test-statistic at first, then have to compare the calculated value with the critical value at specific level of significance.]

(v) Here, calculated z-value, $z = \dfrac{\bar{X} - \mu}{\sigma/\sqrt{n}} = \dfrac{5.84 - 6.0}{0.50/\sqrt{64}}$ [From the given information]
$= - 2.56$

Since calculated z-value does not lie between $\pm$ 1.96 [the critical value at 0.05 level of significance], we reject the null hypothesis. That is, H_1 might be accepted.

Interpretation of the Result : As H_1 is accepted, the mean turnover rate is not equal to 6.0.

Format (2) : Tests about the difference between two means.

Working structure :

Given, $n_1 = *$, $\bar{X}_1 = *$, $\sigma_1^2 = *$ Or, $s_1^2 = *$
$n_2 = *$ $\bar{X}_2 = *$ $\sigma_2^2 = *$ Or, $s_2^2 = *$

We want to test the null hypothesis $H_0 : \mu_1 = \mu_2$
against $H_1 : \mu_1 \neq \mu_2$ Or, $H_1 : \mu_1 > \mu_2$ Or, $H_1 : \mu_1 < \mu_2$

(a) When the population variances (σ_1^2 and σ_2^2) are known,

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

(b) When the sample variances (s_1^2 and s_2^2) are given and sample size large ($n_1 > 30$ and $n_2 > 30$)

Test statistic , $z = \dfrac{\bar{x}_1 - \bar{x}_2}{\sqrt{\dfrac{s_1^2}{n_1} + \dfrac{s_2^2}{n_2}}} \sim N(0,1)$.

Decision :

If the calculated value of z is greater than the tabulated value of z , then the null hypothesis rejected. Otherwise, null hypothesis is accepted.

The Confidence Interval of the difference between two population means $(\mu_1 - \mu_2)$:

(i) $(\overline{x}_1 - \overline{x}_2) - z_{\alpha/2}\sqrt{\frac{\sigma_1^2}{n_1}+\frac{\sigma_2^2}{n_2}} \le (\mu_1 - \mu_2) \le (\overline{x}_1 - \overline{x}_2) + z_{\alpha/2}\sqrt{\frac{\sigma_1^2}{n_1}+\frac{\sigma_2^2}{n_2}}$

(ii) $(\overline{x}_1 - \overline{x}_2) - z_{\alpha/2}\sqrt{\frac{s_1^2}{n_1}+\frac{s_2^2}{n_2}} \le (\mu_1 - \mu_2) \le (\overline{x}_1 - \overline{x}_2) + z_{\alpha/2}\sqrt{\frac{s_1^2}{n_1}+\frac{s_2^2}{n_2}}$

Problem-10: Random samples drawn from two places gave the following data relating to the heights of adults males:

	Dhaka	Rajshahi
No. of adults males in sample	1200	1500
Average height (in inches)	68.50	68.58
Standard deviation (inches)	2.5	3.0

Test at 5% level of significance, that the mean height is the same for adults males in two places.

Solution : Given, $n_1 = 1200$, $\overline{x}_1 = 68.50$, $s_1 = 2.5$

$n_2 = 1500$, $\overline{x}_2 = 68.58$, $s_2 = 3.0$

Let, μ_1 and μ_2 denotes the population mean of height for adults males in two places.

We want to test, $H_0 : \mu_1 = \mu_2$ against $H_1 : \mu_1 \ne \mu_2$

We know, Test statistic, $$z = \frac{\overline{x}_1 - \overline{x}_2}{\sqrt{\frac{s_1^2}{n_1}+\frac{s_2^2}{n_2}}}$$

$$= \frac{68.50 - 68.58}{\sqrt{\frac{(2.5)^2}{1200}+\frac{(3.0)^2}{1500}}}$$

$$= \frac{-0.08}{\sqrt{0.0052 + 0.006}} = \frac{-0.08}{\sqrt{0.0112}} = \frac{-0.08}{0.1058} = -0.756$$

$\therefore$ The calculated value of z is $|z| = 0.756$

The critical value of z at 5% level of significance is $|z_{0.025}| = 1.96$.

Since, the calculated value of z is less than the table value of z. So, the null hypothesis is accepted. Hence, the mean height is the same for adults males in two places.

Problem-(11): Intelligence test on two groups of boys and girls gave the following results:

	Mean	S.D.	Numbers
Girls	83	12	60
Boys	82	14	110

Is there significant difference in the mean scores of boys and girls?

[B.B.S. (Accounting/ Management/Finance/Marketing),Part-3,N.U. 2005]

Solution : Let the population mean scores of boys and girls are μ_1 and μ_2 respectively.

We want to test, $H_o : \mu_1 = \mu_2$

Vs $H_1 : \mu_1 \neq \mu_2$

Given, $\bar{x}_1 = 82$, $s_1 = 14$, $n_1 = 110$

$\bar{x}_2 = 83$, $s_2 = 12$, $n_2 = 60$

$$\text{Test statistic, } z = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}} = \frac{82 - 83}{\sqrt{\frac{(14)^2}{110} + \frac{(12)^2}{60}}} \quad \frac{-1}{2.0449} = -0.49$$

$\therefore |z| = 0.49$

The critical value of z at 5% level of significance is $|z_{0.025}| = 1.96$

Since $|z| < |z_{0.025}|$. So, the null hypothesis is accepted.

Hence, there is no significant difference in the mean scores of boys and girls.

Problem-(12): The mean breaking strength of the cables supplied by a manufacturer is 1500 with standard deviation 60. By a new technique in the manufacturing process it is claimed that the breaking strength of the cables has increased. In this claim a sample of 50 cales is tested. It is found that the mean breaking strength is 1512. Can we support the claim at 5% level? Calculate the 95% confidence interval of the difference between breaking strength of the two types of cables.

Solution : Given, $n_1 = 50$, $\bar{x}_1 = 1500$, $s_1 = 60$

$n_2 = 50$, $\bar{x}_2 = 1512$, $s_2 = 80$

We want test, Ho : $\mu_1 = \mu_2$ against $H_1 : \mu_1 \neq \mu_2$.

$$\therefore \text{Test statistic, } z = \frac{\bar{x}_1 - \bar{x}_1}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}} = \frac{1500 - 1512}{\sqrt{\frac{(60)^2}{50} + \frac{(80)^2}{50}}} = \frac{-12}{\sqrt{72+128}} = \frac{-12}{14.142} = -0.849$$

$\therefore$ The calculated value of z is $|z| = 0.849$

Conclusion: The critical value of z at 5% level of significance is $z_{0.025} = 1.96$.

Since $z_{cal.} < z_{0.025}$. So, the null hypothesis is accepted. Hence we may conclude that the breaking strength of the cables has inreased.

The 95% confidence interval of the difference between breaking strength of the two types of cables is

$$(\bar{x}_1 - \bar{x}_2) - z_{\alpha/2} \sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}} \leq (\mu_1 - \mu_2) \leq (\bar{x}_1 - \bar{x}_2) + z_{\alpha/2} \sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}$$

$$\Rightarrow (1500-1512) - z_{0.025}\sqrt{\frac{(60)^2}{50}+\frac{(80)^2}{50}} \le (\mu_1-\mu_2) \le (1500\text{-}1512)+z_{0.025}\sqrt{\frac{(60)^2}{50}+\frac{(80)^2}{50}}$$

$$\Rightarrow -12 - 1.96\times 14.142 \le (\mu_1\text{-}\mu_2) \le -12 + 1.96\times 14.142$$

$$\Rightarrow -12 - 27.718 \le (\mu_1\text{-}\mu_2) \le -12 + 27.718$$

$$\Rightarrow -39.718 \le (\mu_1-\mu_2) \le 15.718$$

Problem (13) : Suppose a random sample of 10 women from tribal population and 14 women from non-tribal population were drawn and their number of children recorded. The average number of children born to these two groups of women were respectively 4 and 3. The population variances in the number of children computed from the census data were 1.5 and 1 respectively. Use 5 percent level of significance to see if the sample data reflect any difference in the mean number of children born in the population.

Solution : Given, $n_1 = 10$, $\bar{x}_1 = 4$, $\sigma_1^2 = 1.5$

$n_2 = 14$, $\bar{x}_2 = 3$, $\sigma_2^2 = 1$

We want test, Ho : $\mu_1 = \mu_2$ against $H_1 : \mu_1 \neq \mu_2$.

We know, Test statistic, $z = \dfrac{\bar{x}_1-\bar{x}_2}{\sqrt{\dfrac{\sigma_1^2}{n_1}+\dfrac{\sigma_2^2}{n_2}}} \sim N(0,1)$

$$\therefore z = \frac{4-3}{\sqrt{\frac{1.5}{10}+\frac{1}{14}}} = \frac{1}{\sqrt{0.15+0.0714}} = 2.13$$

$\therefore$ The calculated value of z is $z = 2.13$

Conclusion: The critical value of z at 5% level of significance is $z_{0.025}=1.96$. Since $z_{cal} > z_{0.025}$. So, the null hypothesis is rejected. Hence we may conclude that these is a significant difference between the true average number of children born to tribal women and non-tribal women.

Problem-(14) : Two types of new cars produced in Bangladesh are tested for petrol mileage. One group consisting of 36 cars averaged 14 kms. per letre. While the other group consisting of 72 cars averaged 12.5 kms per letre.

(i) What test statistic is appropriate, if $\sigma_1^2 = 1.5$ and $\sigma_2^2 = 2.0$?

(ii) Test, whether there exists a significant difference in the petrol consumption of these two types of cars (use $\alpha = 0.01$)

Solution : Given, $n_1=36$, $\bar{x}_1=14$, $\sigma_1^2 = 1.5$

$n_2 = 72$, $\bar{x}_2=12.5$, $\sigma_2^2 = 2.0$

(i) The appropriate test statistic is z test. Because the population variances of the two groups are known.

(ii) Let, μ_1 and μ_2 be the population mean of 1st and 2nd group respectively.

We want to test the null hypothesis, $H_0 : \mu_1 = \mu_2$ against $H_1 : \mu_1 \neq \mu_2$

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

$$= \frac{14 - 12.5}{\sqrt{\frac{1.5}{36} + \frac{2.0}{72}}} = \frac{1.5}{0.264} = 5.68$$

$\therefore z = 5.68$

The critical value of z at 1% level of significance ($\alpha = 0.01$) is $z_{0.005} = 2.58$

Since $z > z_{0.005}$. So, the null hypothesis is rejected and the test is significant. Hence, there is a significant difference in the petrol consumption of the two types of cars.

Problem-(15): Two types of new cars produced by a company are tested for petrol mileage. One group consisting of 36 cars averaged 14 Kms. per litre, while the other consisting of 72 cars averaged 12 Kms. per litre.

(i) What test statistic is appropriate if $s_1^2 = 1.5$ and $s_2^2 = 2.0$?

(ii) Test whether there exists a significant difference in the petorl consumption of these two types of cars (use a = 0.01).

[B.B.S. (Accounting/ Management/Finance/Marketing),Part-3,N.U. 2005]

Solution: Given, $n_1 = 36$, $\bar{x}_1 = 14$ $s_1^2 = 1.5$

$n_2 = 72$, $\bar{x}_2 = 12$ $s_2^2 = 2.0$

(i) Since the population variances of the two groups are known. So, the appropriate test statistic is z-test.

(ii) Let, μ_1 and μ_2 be the population mean of the petrol consumption the two types of cars.

We want to test, $H_0 : \mu_1 = \mu_2$

Vs $H_1 : \mu_1 \neq \mu_2$

$$\text{Test statistic, } z = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}} = \frac{14 - 12}{\sqrt{\frac{(1.5)^2}{36} + \frac{(2)^2}{72}}} = \frac{2}{0.3436} = 5.82$$

The critical value of z at a = 0.01 level of significance is $z_{0.005} = 2.58$

Since $z < z_{0.005}$. So, the null hypothesis is rejected.

Hence, there is exists a significant difference in the petrol consumption of the two types of cars.

Problem- (16) : The average monthly income of 80 teachers in Dhaka City College is Tk. 20 thousand with standard deviation Tk. 3 thousand per month. While for 60 teachers in Dhaka Commerce College the corresponding quantities are Tk. 22.5 and 5 thousands respectively. Do the above data indicate any real difference between the average monthly income of teachers in the two college?

Solution : Given, $n_1 = 80$, $\bar{x}_1 = 20$, $s_1^2 = (3)^2 = 9$,
$n_2 = 60$, $\bar{x}_2 = 22.5$, $s_2^2 = (5)^2 = 25$

Let, μ_1 and μ_2 denotes the population mean of monthly income of teachers in Dhaka City College and Dhaka Commerce College respectively.

We want to test, $H_0 : \mu_1 = \mu_2$ against $H_1 : \mu_1 \neq \mu_2$

We know, Test statistic, $$z = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}} \sim N(0,1)$$

$$\therefore z = \frac{20 - 22.5}{\sqrt{\frac{9}{80} + \frac{25}{60}}} = \frac{-2.5}{\sqrt{0.1125 + 0.41667}} = \frac{-2.5}{\sqrt{0.52917}} = \frac{-2.5}{0.72744} = -3.44$$

$\therefore$ The calculated value of z is $|z| = 3.44$

The critical value of z at 5% level of significance is $|z_{0.025}| = 1.96$.

Since, $|z| > |Z_{0.025}|$. So, the null hypothesis is rejected and the test is significant. Hence, there is real difference between the average monthly income of teachers in the two college.

Problem- (17) : A product is produced in two ways. A pilot test on 64 times from each method indicates that the product of method 1 has sample mean tensile strength 106 Ibs and a standard devation 12 Ibs, whereas in method 2 the corresponding values of mean and standard deviation are 100 Ibs and 10 Ibs respectively. Greater tensile strength in the product is preferable. Use an appropriate large sample test of 5% level of significance to test whether or not method 1 is better for processing the product. State clearly the null hypothesis.

[MBA, D.U-2003]

Solution : Given, $n_1 = 64$, $\bar{x}_1 = 106$, $s_1 = 12$
$n_2 = 64$, $\bar{x}_2 = 100$, $s_2 = 10$

Let, μ_1 and μ_2 be the population mean tensile strength of method 1 and method 2 respectively.

We want to test , $H_o : \mu_1 = \mu_2$ against, $H_1 : \mu_1 > \mu_2$

$$\therefore \text{Test statistic, } z = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}} = \frac{106-100}{\sqrt{\frac{(12)^2}{64} + \frac{(10)^2}{64}}} = \frac{6\times 8}{\sqrt{244}} = 3.07$$

$\therefore z = 3.07$

The critical value of z at 5% level of significance is $z_{0.05} = 1.65$ (for right tailed test)

Since $z > z_{0.05}$. So, the null hypothesis is rejected and the test is significant. Hence method 1 is better than method 2.

Problem- (18) : A simple sample of the height of 6,400 Indians has a mean of 67.85 inches and a standard deviation of 2.56 inches while a simple sample of heights of 1600 Bangladeshis has a mean of 68.55 inches and standard deviation of 2.52 inches. Do the data indicate that the Bangladeshis are on the average taller than the Indians? Give reasons for your answer.

Solution : Given, $n_1 = 6400$, $\overline{x}_1 = 67.85$, $s_1 = 2.56$

$n_2 = 1600$, $\overline{x}_2 = 68.55$, $s_2 = 2.52$

Let, μ_1 and μ_2 be the population mean-height of the Indians and the Bangladeshis respectively.

We want to test the null hypothesis $H_0 : \mu_1 = \mu_2$ i.e, There is no significant difference between the mean height of Indians and Bangladeshis. Against, $H_1 : \mu_1 < \mu_2$

We know the test statistic, $$z = \frac{\overline{x}_1 - \overline{x}_2}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}} \sim N(0,1)$$

$$\therefore \quad z = \frac{67.85 - 68.55}{\sqrt{\frac{(2.56)^2}{6400} + \frac{(2.52)^2}{1600}}} = \frac{-0.7}{\sqrt{0.001024 + 0.003969}} = \frac{-0.7}{0.07066} = -9.907$$

$\therefore$ The calculated value is $z = -9.907$

The critical value of z at 5% level of significance, $z_{0.05} = -1.65$ (for one tailed test)

Since $z < z_{0.05}$. The null hypothesis is rejected and the test is significant. Hence, the data indicates that the Bangladeshis are on the average taller than the Indians.

Reasons : From the given data we observe that Bangladeshis has mean height higher and standard deviation smaller than that of the Indians.

Problem - (19) : In a survey of buying habits, 400 women shoppers are 8 chosen at random in super market A. Their average weekly food expenditure is Tk. 250 with a standard deviation of Tk. 40. For another group of 400 women shoppers chosen at random in super market B located in another are of the city, the average weekly food expenditure is Tk. 220 with a standard deviation of Tk. 55. Test at 1% level of significance, whether the average weekly food expenditure of the population of women shoppers are equal.

[BBA, Third Semester, N.U. 2009]

Solution : From the given information, we have,

$n_1 = 400$, $\overline{x}_1 = 250$, $s_1 = 40$

$n_2 = 400$, $\overline{x}_2 = 220$, $s_2 = 55$

We want to test, $H_0 : \mu_1 = \mu_2$ against, $H_1 : \mu_1 \neq \mu_2$

$$\therefore \text{ Test statsitic, } z = \frac{\overline{x}_1 - \overline{x}_2}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}} = \frac{250-220}{\sqrt{\frac{(40)^2}{400} + \frac{(55)^2}{400}}} = \frac{30}{3.4004} = 8.822$$

$\therefore$ The calculated value of $z = 8.822$

The critical value of z at 1% level of significance is = 2.58 .

Since the calculated value of z is greater than the critical value. So, the null hypothesis is rejected. Hence, the average weekly food expenditure of the population of women shoppers are not equal.

Problem-20: You are working as a purchase manager for a company. The following information has been supplied to you by two manufacturers of electric bulbs :

	Company A	Company B
Mean life	1300 hours	1288 hours
Standard deviation	82 hours	93 hours
Sample Size	80	80

Which brand of bulbs are you going to purchase if you desire to take a risk of 5%?

[BBA Third Semester N. U. 2006]

Solution : Given, $n_1 = 80$, $\bar{x}_1 = 1300$, $s_1 = 82$

$n_2 = 80$, $\bar{x}_2 = 1288$, $s_2 = 93$

We want to test, $Ho : \mu_1 = \mu_2$

(i. e. there is no significant difference in the quality of the two brands of bulbs.)

$$\therefore \text{ The test statistic } z = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}} = \frac{1300-1288}{\sqrt{\frac{(82)^2}{80} + \frac{(93)^2}{80}}} = \frac{12}{\sqrt{84.05+108.11}} = \frac{12}{13.86} = 0.87$$

Since the calculated value of z = 0.87 is less than the critical value of z = 1.96 (5% level of significance). So, the null hypothesis is accepted.

Hence, there is no significant difference in the quality of the two brands of bulbs and we can purchase any brand of bulbs we desire to take a risk of 5%.

Format (3) : Test for specified proportion or pululation proportion

Working Structure: Let, the population proportion $= \pi$.

We want to test, $Ho : \pi =$ Given value

Test statistic, $z = \dfrac{p-\pi}{\sqrt{\dfrac{\pi-(1-\pi)}{n}}}$ [Here, sample proportion, $p = \frac{m}{n}$]

Problem- (21) : A random sample of 500 bulbs was taken from a large consignment for examination and 40 were found to be defective. Test the suppliers claim that the proportion of defcetive bulbs in the consignment is 0.03?

Solution : We want to test the null hypothesis, $H_0 : \pi = 0.03$. Where, $H_1 : \pi \neq 0.03$

We know, the test statistic, $z = \dfrac{p - \pi}{\sqrt{\dfrac{\pi(1-\pi)}{n}}} \sim N(0,1)$

Where, $p =$ Sample proportion of defective bulbs $= \frac{m}{n} = \frac{40}{500} = 0.08$

$\pi =$ Population proportion of defective bulbs $= 0.03$

$$\therefore \quad Z = \frac{0.08 - 0.03}{\sqrt{\frac{0.03(1-0.03)}{500}}} = \frac{0.05}{\sqrt{\frac{0.03 \times 0.97}{500}}} = \frac{0.05}{0.0076} = 6.5789$$

The critical value of z at 5% level of significance is $Z_{0.025.} = 1.96$

Since $Z > Z_{0.025}$ the null hypothesis is rejected. Hence, the supplier's significantly claim that the proportion of defective bulbs in the consignment is 0.03.

Problem - 22: A random sample of 300 admitted students is found to be 50 students are mentally disordered in a college. Test the mentally disordered of all students in that college is more that 10%.

Solution : Let, the population proportion $= \pi$.

We want to test, $H_0 : \pi = 0.1$ (That is, the mentally disordered of all students is 10%)

$H_1 : \pi > 0.1$ (That is, the mentally disordered of all students is more than 10%)

Test statistic, $z = \dfrac{p-\pi}{\sqrt{\dfrac{\pi\text{-}(1-\pi)}{n}}} \sim N(0, 1)$

Here, $n = 300$, sample proportion, $p = \dfrac{m}{n} = \dfrac{50}{300} = 0.17$

$\therefore\ z = \dfrac{0.17 - 0.1}{\sqrt{\dfrac{0.1(1-0.1)}{300}}} = \dfrac{0.07}{0.0173} = 4.046$

Since the calculated value of z = 4.046 is greater than the critical value of z = 1.645 (5% level of significance). So, the null hypothesis is rejected.

Hence, the mentally disordered of all students is more than 10%.

Format (4) : Test about the difference between two population proportions.

Working Structure:

We want to test, $H_0 : \pi_1 = \pi_2$. i.e., there is no ....

We know the test statistic, $z = \dfrac{p_1 - p_2}{\sqrt{p\,(1-p)\,(\dfrac{1}{n_1} + \dfrac{1}{n_2})}} \sim N(0,1)$

Here, n_1 = 1st sample size , n_2 =2nd sample size

m_1 = Number of the given characteristic of the 1st sample

m_2 = Number of the given characteristic of the 2nd sample

$\therefore$ 1st sample proportion, $p_1 = \dfrac{m_1}{n_1}$, 2nd sample proportion, $p_2 = \dfrac{m_2}{n_2}$

and $p = \dfrac{m_1 + m_2}{n_1 + n_2}$

A 95% C.I. for $\pi_1 - \pi_2$ is $[(p_1 - p_2) \pm z_{\alpha/2}\sqrt{\dfrac{P_1(1-p_1)}{n_1} + \dfrac{P_2(1-p_2)}{n_2}}]$

Problem (23) : In a random sample of 1,000 persons from Dhaka City, 400 are found to be consumers of wheat. In a sample of 800 from Khulna City, 400 are found to be consumers of wheat. Do the data reveal a significant difference between Dhaka City and Khulna, so far as the proportion of wheat consumers concerned?

Solution : We want to test, $H_o : \pi_1 = \pi_2$. That is, there is no difference between Dhaka City and Khulna City, so far as the proportion of wheat consumers concerned. Here, $H_1 : \pi_1 \neq \pi_2$

We know, the test statistic, $z = \dfrac{p_1 - p_2}{\sqrt{p(1-p)\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}} \sim N(0,1)$

Where, $n_1 = 1000$, $n_2 = 800$, $m_1 = 400$, $m_2 = 400$

$$\therefore \quad p_1 = \frac{m_1}{n_1} = \frac{400}{1000} = 0.4, \qquad p_2 = \frac{m_2}{n_2} = \frac{400}{800} = 0.5$$

$$p = \frac{m_1 + m_2}{n_1 + n_2} = \frac{400 + 400}{1000 + 800} = \frac{800}{1800} = \frac{4}{9}$$

$$\therefore z = \frac{0.4 - 0.5}{\sqrt{\frac{4}{9}\left(1 - \frac{4}{9}\right)\left(\frac{1}{1000} + \frac{1}{800}\right)}} = \frac{-0.1}{\sqrt{\frac{4}{9} \times \frac{5}{9}\left(\frac{4+5}{4000}\right)}}$$

$$= \frac{-0.1}{\sqrt{\frac{4}{9} \times \frac{5}{9} \times \frac{9}{4000}}} = \frac{-0.1}{\sqrt{0.00056}} = \frac{-0.1}{0.02366} = -4.23$$

$\therefore |z| = 4.23$

The critical value of z at 5% level of significance is, $|z_{0.025}| = 1.96$

Since, $|z| > |z_{0.025}|$. Therefore, the null hypothesis is rejected and the test is significant. We can conclude that there is a significant difference between Dhaka and Khulna City so far as the proportion of wheat consumers concerned.

Problem- (24) : In a simple random sample of 600 men taken from a Pabna city 400 are found to be smokers. In another simple random sample of 900 men taken from Natore city 450 are smokers. Do the data indicate that there is a significant difference in the habit of smoking in the two cities?

Solution : We want to test, $H_o : \pi_1 = \pi_2$. That is, there is no significant difference in the habit of smoking in the two cities. Where the alternative hypothesis, $H_1 : \pi_1 \neq \pi_2$

We know, the test statistic, $z = \dfrac{p_1 - p_2}{\sqrt{p(1-p)\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}} \sim N(0,1)$

Here, $n_1 = 600$, $n_2 = 900$, $m_1 = 400$, $m_2 = 450$

$$\therefore \quad p_1 = \frac{m_1}{n_1} = \frac{400}{600} = 0.667, \qquad p_2 = \frac{m_2}{n_2} = \frac{450}{900} = 0.5$$

$$p = \frac{m_1 + m_2}{n_1 + n_2} = \frac{400 + 450}{600 + 900} = \frac{850}{1500} = 0.5667$$

$$\therefore \quad z = \frac{0.66 - 0.5}{\sqrt{0.5667\,(1 - 0.5667)\,(\frac{1}{600} + \frac{1}{900})}} = \frac{0.16}{\sqrt{0.24555\,(0.001667 + 0.001111)}}$$

$$= \frac{0.16}{\sqrt{0.24555 \times 0.00278}} = \frac{0.16}{0.0261177} = 6.126$$

The critical value of z at 5% level of significance is, $z_{0.025.} = 1.96$

Since $z > z_{0.025.}$ So, the null hypothesis is rejected and the test is significant. Hence, there is a significant difference in the habit of smoking of the two cities.

Problem- (25) : A machine produced 20 defected articles in a batch of 500 products. After overhauling it produced 3 defectives in a batch of 100 products. Has the machine improved?

Solution : We want to test, $H_o : \pi_1 = \pi_2$. That is, the machine has not improved after overhauling.

Against , $H_1 : \pi_1 \neq \pi_2$

We know, the test statistic, $z = \dfrac{p_1 - p_2}{\sqrt{p\,(1-p)\,(\frac{1}{n_1} + \frac{1}{n_2})}} \sim N(0,1)$

Here, $n_1 = 500$, $n_2 = 100$, $m_1 = 20$, $m_2 = 3$

$$\therefore \quad p_1 = \frac{m_1}{n_1} = \frac{20}{500} = 0.04, \qquad p_2 = \frac{m_2}{n_2} = \frac{3}{100} = 0.03$$

$$p = \frac{m_1 + m_2}{n_1 + n_2} = \frac{20 + 3}{500 + 100} = \frac{23}{600} = 0.0383$$

$$\therefore \quad z = \frac{0.04 - 0.03}{\sqrt{0.0383\,(1 - 0.0383)\,(\frac{1}{500} + \frac{1}{100})}} = \frac{0.01}{\sqrt{0.03683 \times \frac{6}{500}}} = \frac{0.01}{0.021023} = 0.4757$$

The critical value of z at 5% level of significance is $z_{0.025.} = 1.96$

Since, $z < z_{0.025}$.. So, the null hypothesis is accepted and the test is insignificant. Hence, the machine has not improved after overhauling.

Problem- (26) : Before an increase in excise duty on tea 400 pepole out of a sample of 500 persons were found to be tea drinkers. After an increase in the duty, 400 persons were known to be tea drinkers in a sample of 600 people. Do you think that there has been a significant decrease in the consumption of tea after the increase in the excise duty.

Solution : We want to test, $H_o : \pi_1 = \pi_2$. That is, there is no significant decrease in the consumption of tea after increase in the excise duty. Here, $H_1 : \pi_1 > \pi_2$

We know the test statistic, $z = \dfrac{p_1 - p_2}{\sqrt{p(1-p)\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}} \sim N(0,1)$

Here, $n_1 = 500$, $n_2 = 600$, $m_1 = 400$, $m_2 = 400$

$\therefore \quad p_1 = \dfrac{m_1}{n_1} = \dfrac{400}{500} = 0.8, \qquad p_2 = \dfrac{m_2}{n_2} = \dfrac{400}{600} = 0.667$

$$p = \frac{m_1 + m_2}{n_1 + n_2} = \frac{400 + 400}{500 + 600} = \frac{800}{1100} = 0.727$$

$$\therefore z = \frac{0.8 - 0.667}{\sqrt{0.727\,(1 - 0.727)\left(\frac{1}{500} + \frac{1}{600}\right)}} = \frac{0.133}{\sqrt{0.19847 \times (0.002 + 0.00167)}}$$

$$= \frac{0.133}{0.19847 \times 0.00367} = \frac{0.133}{0.0269} = 4.944$$

The critical value of z at 1% level of significance is $z_{0.01} = 2.33$ (for one tailed test)

Since, $z > z_{0.01}$, the null hypothesis is rejected and the test is significant. Hence, there is a significant decrease in the consumption of tea after the increase in the excise duty.

Problem - 27 : In a recent interview on banking profession, 123 out of 200 male banking professionals rated the banking career as "Excellent" and 87 out of 150 female banking professionals rated it as "Excellent:. Using the 10% level of significance, can we conclude that there is a significant difference in the proportion of male versus female banking professionals who rate the banking career as Exchellent?

(a) What is the null hypothesis?

(b) Should the null hypothesis be rejected or accepted?

(c) What is your conclusion based on the hypothesis? [BBA, N. U. 2007]

Solution : (a) Let, π_1 and π_2 be the population proportion of male and female respectively.

The required null hypothesis is $H_0 : \pi_1 = \pi_2$. That is, there is no significant difference in the population proportion of male and female banking professionals rated the banking carrer as "Excellent".

(b) Given, $n_1 = 200$, $m_1 = 123$, $n_2 = 150$, $m_2 = 87$

Sample proportion of male, $p_1 = \dfrac{123}{200} = 0.615$ and sample proportion of female, $p_2 = \dfrac{87}{150} = 0.58$

$\therefore$ Pooled proportion, $p = \dfrac{123+87}{200+150} = \dfrac{210}{350} = 0.6$

$\therefore$ The test statistic, $z = \dfrac{p_1 - p_2}{\sqrt{p(1-p)\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}} = \dfrac{0.615 - 0.58}{\sqrt{0.6(1-0.6)\left(\frac{1}{200} + \frac{1}{150}\right)}} = \dfrac{0.035}{0.0529} = 0.6616$

The critical value of z at 10% level of significance is $z_{cri} = 1.645$

Since, the calculated value z is less than the critical value. So, the null hypothesis (Ho) should be accepted.

(c) Conclusion : On the based the null hypothesis, we can say that there is no significant difference in the population proportion of male and female banking professionals rated the banking carrer as "Excellent."

Exercise-6(1)

Part- A: Brief questions and answer

1. What is hypothesis?

Ans. Any statement about the population is called hypothesis.

2. What is test of hypothesis?

Ans. Test of hypothesis is a statistical procedure that is used to provide evidence in favour of some statement (called a hypothesis).

3. What do you meant by statistical hypothesis?

Ans. Statistical hypothesis is some assumption or statement, which may or may not be true, about a population or about the probability distribution characterising the given population, which we want to test on the basis of the evidence from a random sample.

4. What are the types of statistical hypothesis?

Ans. There are two types of statistical hypothesis : (i) Null hypothesis (ii) Alternative hypothesis

5. What is null hypothesis?

Ans. A null hypothesis sates that there is no difference between a sample estimate and the true population value.

6. Define (i) alternative hypothesis (ii) Parametric hypothesis (iii) Non-parametric hypothesis.

Ans. (i) Alternative hypothesis : The negative of null hypothesis is called the alternative hypothesis.

(ii) Parametric Hypothesis : Any hypothesis about the parametric of a population distribution is known as parametric hypothesis.

(iii) Non-Parametric Hypothesis : Any hypothesis about a population distribution is called a non-parametric hypothesis.

7. What is type I error?

Ans. Type I Error : In a test of hypothesis the type I error occurs when the null hypothesis H_0 is rejected although it was true. The probability of a type I error

8. What is type II error?

Ans. Type II error : In a test of hypothesis the type II error occurs when the null hypothesis H_0 is not rejected (i.e., accepted) although it was false.

9. Which symbol is used to denote the level of significance?

Ans. Level of Significance is denoted by α.

10. Which type of error is more unpredictable?

Ans. Type I Error is more unpredictable?

11. Null hypothesis is usually denoted by ----------.

Ans. Null hypothesis is usually denoted by Ho

12. Alternative hypothesis is denoted by -----------.

Ans. Alternative hypothesis is denoted by H_1 or Ha

13. The probability of type I error is denoted by ------------.

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

14. The probability of type II error is denoted by ---------.

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

15. What is level of significance?

Ans. The level of significance is the maximum probability of committing a type I error.

16. Different between type I error and type II error?

Ans. The probability of type I error is called level of significance where as the probability of type II is not called level of significance.

17. What are the critical values at 1% and 5%.

Ans. The critical values at 1% implies that we are 99% (or 0.99) confident about the significance of our decision and we have 1% (or 0.01) chance to occur a type I error.

The critical values at 5% implies that we are 95% (or 0.95) confident about the significance of our decision and we have 5% (or 0.05) chance to occur a type I error.

18. ------------ is the maximum probability of making a type I error.

Ans. The level of significance is the maximum probability of committing a type I error.

19. What is acceptance region?

Ans. An acceptance region is a set of possible values of the test statistic (in a test of hypothesis) that leads the null hypothesis to be accepted.

20. What is rejection region?

Ans. A rejection (or critical) region is a set of possible values of the test statistic (in a test of hypothesis) that leads the null hypothesis to be rejected.

21. What do you understand by test statistic?

Ans. Any mathematical function of the elements belong to a sample is called as statistic of that small.

22. What is one-tailed test or one-sided test?

Ans. An one-tailed test is such a hypothesis test for which the rejection region consists in only one-side (or direction).

23. What are the types of one tailed test?

Ans. There are two types of one tailed test, such as : (i) Right-tailed test (ii) Left-tailed test .

24. What is right-tailed test?

Ans. Right-tailed test : A test of any statistical hypothesis where the alternative is right -tailed such as $H_0 : \mu = \mu_0$ vs $H_1 : \mu > \mu_0$ is called a right-tailed test.

25. What is left-tailed test?

Ans. Left-tailed test : A test of any statistical hypothesis where the alternative is left - sided such as $H_0 : \mu = \mu_0$ vs $H_1 : \mu < \mu_0$ is called a left-tailed test.

26. What do you understand by two-tailed test or two sided test?

Ans. Two-tailed test (Two sided Test) : A two-tailed test is such a hypothesis test for which the rejection region consists in both the side or direction.

27. What is the critical value?

Ans. Critical value : A value that is the dividing point between the region where the null hypothesis is not rejected and the region where it is rejected.

28. What do you understand by degrees of freedom?

Ans. By degrees of freedom we mean the number of cases to which the values can be assigned arbitrarily without violating the restrictions placed.

29. What is the probability of type I error?

Ans. The probability of type I error is the level of significance.

30. What are the assumptions for normal test or z - test?

Ans. The necessary assumptions for normal test is as follows :

(i) The random sampling distribution of a statistic is approximately normal.

(ii) The values provided by the sample data are sufficiently close to the population values.

31. Normal test is widely used in testing hypothesis regarding ....

Ans. Normal test is widely used in testing hypothesis regarding means, proportions, coefficient of correlation etc.

32. What is the test statistic of the normal test?

Ans The test statistic of the normal test, $z = \frac{u - E(u)}{S.E(u)}$ is normally distributed with mean 0 and variance1.

33. If the null hypothesis Ho : $\mu = \mu_0$, then state (i) the alternative hypothesis in two tailed test. (ii) the alternative hypothesis in left tailed test (iii) the alternative hypothesis in right tailed test.

Ans. (i) $H_1 : \mu \neq \mu o$ (ii) $H_1 : \mu < \mu o$ (iii), $H_1 : \mu > \mu o$

34. Which test is known as small test?

Ans. Many times t-test is called as a small sample test.

35. Write the formula for difference of two means in case of large sample test?

Ans. When the population variances (σ_1^2 and σ_2^2) are unknown and sample size large ($n_1 \geq 30$ and $n_2 \geq 30$) Test statistic , $z = \frac{\overline{x}_1 - \overline{x}_2}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}} \sim N(0,1)$.

36. Give the formula for testing hypothesis of single population mean in case of population variance is known.

Ans. If the population variance (σ^2) is known, then the test statistic, $z = \frac{\overline{x} - \mu}{\sigma/\sqrt{n}} \sim N(0,1)$

37. Give the formulae for test hypothesis of single population mean in case of population variance is unknown.

Ans. If the population variance (σ^2) is not known and the sample size is large ($n \geq 30$), then the test statistic, $z = \frac{\overline{x} - \mu}{s/\sqrt{n}} \sim N(0, 1)$

38. Write down the formula of hypothesis testing for single population proportion.

Ans. Test statistics $z = \dfrac{p - \pi}{\sqrt{\dfrac{\pi(1-\pi)}{n}}} \sim N(0,1)$

39. Write down the formula of hypothesis testing for difference between two population proportions.

Ans. Test statistic, $z = \dfrac{p_1 - p_2}{\sqrt{p(1-p)\left(\dfrac{1}{n_1} + \dfrac{1}{n_2}\right)}} \sim N(0, 1)$

Part- B & C (Theoretical Questions & Mathematical Problems)

THEORITICAL QUESTIONS

1. Define hypothesis, parametric hypothesis, non-parametric hypothesis and test of significance.
2. What do you mean by statistical hypothesis? What are the various types of statistical hypothesis? Explain each hypothesis with examples.

   Or, What do you mean by null hypothesis and alternative hypothesis?

   Or, What do you mean by test of hypothesis?
3. What do you mean by type I error and type II error?

   Or, Define type I error and type II error.

   Or, What do you mean by type I error and type II error?
4. What do you mean by level of significance, power of test and test statistic?
5. What do you mean by critical region or rejection region and acceptance region?
6. What do you mean by one tailed and two tailed test?

   Or, Discuss the one tailed and two tailed test.
7. What is degree of freedom?
8. Discuss the various steps of test of significance.

   Or, Explain the procedure generally followed in testing a hypothesis.

   Or, Discuss the procedure or steps of the test of hypothesis.

   Or, Discuss the procedure followed in testing hypothesis.
9. Discuss Hypothesis testing for single population mean.

   Or, Test the specified mean of the population.

   Or, Describe the tests of the difference the specified value of the population mean.

   Or, Compare the sample mean with a specified value of the population mean.

   Or, Discuss the methods of the test of hypothesis about population mean.
10. Discuss Hypothesis testing for difference between two population means.

    Or, What tests can be applied to test the difference between two means? Write the formula of that tests and explain.

    Or, Discuss the difference test of means.

    Or, Describe the test of hypothesis about the difference between two means.

Or, How would you test the significance of the difference between two means?

Or, Test the equality the two population means. Or, Ho : $\mu_1 = \mu_2$

11. Describe Hypothesis Testing for single population proportion.
12. Discuss the Hypothesis Testing for two population proportion.

    Or, How can you test the equality of the two population proportions?
13. Distinguish between Type I error and Type II error.
14. Distinguish between rejection region and acceptance region.
15. Distinguish between the null hypothesis and alternative hypothesis.
16. Distinguish between parametric test and non-parametric test.

MATHEMATICAL PROBLEM :

Format- 1 : Test about population mean / single population mean.

● **1.** A random sample of 100 students gave a mean weight of 58 kg with the standard deviation 4 kg. Test the hypothesis that the mean weight in the population is 60 kg.

[$z_{cal.} = 5$, No.]

● **2.** A sample of 400 managers is found to have a mean height of 171.38 cms. Can it be reasonobly regarded as a sample from a large population of mean height 171.17 cms and standard deviation of 3.30 cms? [Ans. z = 1.31] [BBA, DU. 2004]

● **3.** The standard deviation of the weight of 100 gm bread made by a certain bakery is 1 gm. On a certain day the owner doubted that the production is out of control. To check whether its products is under control, employees select a random sample of 25 breads and find that their mean weight is 99.5 gm. Test the doubt of the owner at 5% level of significance.

[Hints : $\sigma = 1$, $\mu = 100$, $\overline{X} = 99.5$, $n = 25$, $z = -2.5$, H_0 reject.] [BBS (H), Part-III, NU-2004]

● **4.** A sample of 100 tyres was taken from a big stock of particular manufactures and the average life of these types was found to be 40000 km with a standard deviation of 5000 km. Could the sample come from a population with a mean life of 39500 km? [$z_{cal.} = 1$, NO.]

● **5.** An ambulance service claims that it takes, on the average 8.9 minutes to reach its destination in emergency calls. To check on this claim, the agency which licenses ambulance services has been timed on 50 emergency calls, getting a mean of 9.3 minutes with a standard deviation of 1.8 minutes. At the level of significance of 0.05, does this constitute evidence that the figure is to law?

[BBA DU. 2005]

[**Ans.** z = 16, there is no significance difference between the average figure observed and the average figure claimed.]

● **6.** Is it likely that a sample of 300 items, with mean 16.0 is a random sample from a large population whose mean is 16.8 and standard deviation 5.2? Calculate the 98% limits of the mean of such samples. [$z_{cal.} = 2.67$, No, 15.3 and 16.69]

● **7.** A sample of 100 tyres is taken from lot. The mean life of tyres is found to be 39,350 km, with a standard deviation of 3260. Could the sample come from a population with mean life of 40,000 km? Establish 99% confidence limits within which the mean life of tyres is expected to lie.

[$z_{cal.} = 1.994$, No (5% level), 38508.92 to 40191.08]

● **8.** A sample of 400 items is drawn from a normal population whose mean is 5 and variance 4. The sample mean is 4.45. Can the sample be regarded as true random sample drawn from the population?

[$|z_{cal.}| = 5.5$, z_{cri} (1%). (1%) = 2.58, sig]

● 9. The average life time of a sample of 120 bulbs produced by a company is found to be 1,600 hours with standard deviation of 100 hours. Test the hypothesis that the average life time of the bulbs produced by the company is 1,700 hours.] [$|z_{cal.}|$ = 10.95, NO.]

● 10. The mean life time of 150 light tubes made by a company gave mean life time of 1440 hours with a standard deviation of 40 hours. Is it likely that the sample has come from a population that has come from a population with a mean life time of 1550 hours?
[$|z_{cal.}|$ = 33.68, No.]

● 11. A sample of 420 items is taken from a population whose standard deviation is 2.4. The mean of the sample is 28. Test whether the sample has come from a population with mean 30.
[$|z_{cal.}|$ = 17.09]

● 12. The height obtained from a random sample of size 100 is 64 inches. The standard deviation of the distribution of height of the population is known to be 3 inches. Test the statement that the mean height of the population is 67 inches at 5% level of significance. Also set up 99% limits of the mean height of the population [$z_{cal.}$ = 10, $z_{cri.}$ = 1.96 (5% level.), σ = 3, limits 63.2 to 64.8]

● 13. A sample of 64 farm laboures engaged in paddy harvesting operations shows an average monthly wage rate of Tk. 200 with a standard deviation of Tk. 9. Using 5% level of significance, verify if the sample result indicates that their current average monthly wage rate is greater than Tk. 198.
[Ho: μ = 198, H_1: μ > 198, z = 1.78.]

● 14. Record of several years of applicants for admission at FMS showed their mean score is 31.5. An administrator is increased in knowing whether the caliber of recent applicants has changed. For testing this hypothesis the scores of a sample of 100 applicants from the scores of recent applicants is obtained from admission office. The mean for this turned out to be 328. The sample standard deviation is 38, which may also be assumed for the population. Test the hypothesis using 5% level of significance. [BBA(Hons), Department of Accounting, DU,-2014]
[Hints.: σ = 38, μ = 328, $\overline{X}$ = 315, n = 100, $z_{cal.}$ = 3.421, $z_{cri.}$ = 1.96 (5% level.)]

● 15. By using the following information :
H_o : μ = 50, H_1 : $\mu \neq 50$
The sample mean is 49 and the sample size is 36. The population standard deviation is 5. Use the 0.05 significance level, answer the following question.
(i) Is this a one or two - tailed test?
(ii) What is the dicision rule?
(iii) What is the value of the test statistic?
(iv) What is your decision regarding H_o?
(v) What is the p-value? interpret it.
[Ans. (i) two (ii) reject Ho and accept H_1 when the the calculated value of z does not fall in the region from - 1.96 and 1.96. (iii) z = -1.2 (iv) not reject Ho (v) p=2(.5 - .3849)= .2302 chance of finding a z value this is large when H_o is true.]

Format (2) : Tests about the difference between two population means.

● 16. Intelligence test on two groups of boys and girls gave the following results :

	Mean	Standard deviation	Numbers
Boys	83	12	60
Girls	82	14	110

Is there significant difference in the mean scores of boys and girls? [z = 0.49]

● 17. In a study to test whether there is difference between the average heights of adult females born in two different countries, random samples yielded the following results:

$n_1 = 120$, $\overline{x}_1 = 62.7$, $s_1 = 2.50$

$n_2 = 150$, $\overline{x}_2 = 61.8$, $s_2 = 2.62$

[Ans. $z = 11.69$] [BBA, Part-II, NU-2010]

● 18. You are working as a purchase manager for a company. The following information has been supplied to you by two manufacturers of electric bulbs :

	Company A	Company B
Mean life	1300 hours	1288 hours
Standard deviation	82 hours	93 hours
Sample / Size	80	80

Which brand of bulbs are you going to purchase if you desire to take a risk of 5%?

[BBA, Part-I, NU-2006]

● 19. Two types of new cars produced by a company are tested for petrol mileage. One group consisting of 36 cars averaged 14 Kms. per litre, while the other consisting of 72 cars averaged 12 Kms. per litre.

(i) What test statistic is appropriate if $\sigma_1^2 = 1.5$ and $\sigma_2^2 = 2.0$?

(ii) Test whether there exists a significant difference in the petorl consumption of these two types of cars (use $\alpha = 0.01$). [BBS (H), Part-III, NU-2005]

[**Ans.** (i) The appropriate test statistic to be used is the test of difference between two means.

(ii) $z = 5.68$, H_0: reject.]

● 20. The study identified a random sample of 562 female and 852 male students who had achieved the same high score on the mathematics portion of the test. That is, the female and male students were viewed as having similarly high abilities in mathematics. The verbal scores for the two samples are as given :

Female students : $\overline{X}_1 = 547$, $S_1 = 83$

Male students : $\overline{X}_2 = 525$, $S_2 = 78$

Do the data support the conclusion that given a population of female students and a population of male students similarly high mathematics abilities, the female students will have a significanlty higher verbal ability? Test at a 5% level of significance. What is your conclusion? [MBA, DU. 2003]

[Hints : $H_0 : \mu_1 \geq \mu_2$ vs $H_1 : \mu_1 < \mu_2$, $z = 4.995$, H_0 rejected]

● 21. A random sample of 100 mill workers at Natore showed their mean wages to be Tk. 350 per day with standard deviation Tk. 28. A sample of 150 mill workers in Gazipur should the mean wage to be Tk. 390 per day with standard deviation Tk. 40. On the basis of the data would you conclude the mean wages of mill workers in Gazipur are higher than those at Natore? [$z_{cal.} = -9.29$]

● 22. A random sample from 200 villages was taken from Pabna district and the average population per village was found to be 250 with standard deviation 40. Another random sample of 200 villages from the same district gave an average population of 480 per village with standard deviation 60. Is the difference between the averages of the two samples statistically significant?

● **23.** In a survey of buying habits, 400 female shoppers are chosen at random in Agora. Their average weekly food expenditure is Tk. 250 with a standard deviation of Tk. 40. For another group of 400 female shoppers chosen at random in Metro Plaza located in another area of Dhaka City. The average weekly food expenditure is Tk. 220 with a standard deviation of Tk. 55. Test at 1% level of significance, whether the average weekly food expenditures of the populations of female shoppers are equal.

[Ans. $z = 8.822$]

● **24.** A sample of 65 observations is selected from one population. The sample mean is 2.67 and the sample standard deviation is 0.75. A sample of 50 observations is selected from a second population. The sample mean is 2.59 and the sample standard deviation is 0.66. Conduct the following test of hypothesis using the 0.05 significance level. $H_0 : \mu_1 \leq \mu_2$ vs $H_1 : \mu_1 > \mu_2$

(i) Is this a one-tailed or a two-tailed test?

(ii) State the decision rule.

(iii) Compute the value of the test statistic.

(iv) What is your decision regarding?

(v) What is the p - value?

● **25.** The selection of a store location depends on many factors, one of which is the level of family income areas around the proposed site. Suppose that a large department-store chain is trying to decide whether to build a new store in a Dhaka City or in the nearby Gazipur City. building costs are lower in Gazipur City and the company decides it will build there unless average family income is higher in Dhaka City than Gazipur City. A survey of 100 residences in each of the cities found that the mean annual family income was Tk. 2,99,800 in Dhaka City and Tk. 2,86,500 in Gazipur City. From other sources, it is known that the population standard deviations of annual family income are Tk. 47,400 in Dhaka City and Tk. 53,650 in Gazipur City. At the 5% level of significance can it be concluded that the mean family income in Dhaka City exceeds that of Gazipur City.

[Hints : $H_0 : \mu_1 \geq \mu_2$ vs $H_1 : \mu_1 < \mu_2$, $z = 1.8578$, H_0 rejected]

Format- 3: Test for single population proportion
or, Test for specified proportion or pululation proportion

● **26.** A random sample of 460 bulbs was taken from a large consignment for examination and 40 were found to be defective. Test the suppliers claim that the proportion of defective bulbs in the consignment is 0.04?

[$z = 5.14$]

● **27.** A coin is tossed 200 times under identical conditions independently yielding 60 heads 140 tails. Test at 1% level of significance, whether or not the coin is unbiased. State clearly the null hypothesis and the alternative hypothesis.

[$H_0 : \pi = 0.5$, $H_1 : \pi \neq 0.5$, $p = \frac{60}{200} = 0.3$, the coin is biased]

● **28.** A sales clerk in the departmental store claims that 60% of the shoppers entering the store leave without making a purchase. A random sample of 50 shoppers showed that 35 of then left without buying anything. Are these sample results consistent with the claim of the sales clerk? Use a level of significance of 0.05. [Ans. $z = 1.45$] [BBA, Part (II), N.U. 2006]

● **29.** A random sample of 400 house wives was selected to know their individual wives as to whether they prefer brand A detergent or brand B. Brand A was favoured by 180 house wives, while brand B was favoured by the rest. Do these data provide sufficient evidence to indicate a difference for the two brands detergents? [Ho : $\pi = 0.5$, $\rho = \frac{180}{400} = 0.45$, $z = -2.0$]

Format-4: Test about the difference between two population proportions.

● **30.** In a random sample of 2000 farmers selected from the Rajshahi devision in the year 1986, 50 percent farmers stated that the level of rainfall during the paddy season was satisfactory in that year. In the year 1987, 60 percent out of a random sample of 2500 farmers observed the same for that year. Test the hypothesis whether the average level of rainfull during the paddy season in Rajshahi was in the both years. [$z_{cal.} = -6.67$]

● **31:** In a random sample of 880 males at old Dhaka, 440 were found to be smokers which in another random sample of 1000 males at new Dhaka 480 were found to be smokers. Discuss the question whether the data reveal a significance difference at old Dhaka and new Dhaka so far as the proportion of smokers is concerned. [$z = 0.8658$]

● **32.** In a random sample of 100 persons taken from village A, 60 are found to be consuming tea. In another sample of 200 persons taken from village B, 100 persons are found to be consuming tea. Do the data reveal. Significant difference between the two villages so far as the habit of taking tea is concerned? [MBA, DU. 1999]

● **33.** A tax firm is interested in comparing the quality of work at two of its regional offices. By randomly selected samples of tax retuurns prepared at each office and verifying the sample returns accuracy, the firm will be able esimate the proportion of erroneous returns prepared at each office. Indepenent random samples from the two offices provide the following information:

	Sample size	Number of returns with errors
Office I	250	35
Office II	300	27

COnduct a hypothesis test to determine whether the error proportions differ between the two offices at 1% level of significance. $z=1.85$ [BBA(Hons), Department of Marketing, DU,-2011]

● **34.** In a public opinion survey, 60 out a sample of 100 high-income voters and 40 out of a sample of 75 income voters supported a decrease in sales tax. Can we concluded at the 5% level of significance that the proportion of voters favoring a sales tax decrease differs between high and low income voters?

[BBA(Hons), Department of Management, DU,-2015]

[$z_{cal.} = 0.927$, $z_{0.025} = 1.96$]

● **35.** 500 units from factory 'A' are inspected and 12 are found to be defective, 800 units from factory 'B' are inspected and 12 are found to be defective. Can it be concluded at 5% level of significance that production at factory 'B' is better than in factory 'A'?

[BBA(Hons), Department of marketing, RU,-2015]

[$z_{cal.} = 1.184$, $z_{0.025} = 1.96$]