If the form of alternative hypothesis has > sign, the entire critical region is taken on the right side of the distribution.

Step 5 : Making Decision

We compute the value of the appropriate statistic and ascertain whether the computed value falls in acceptance or rejection region depending on the specified Level of significance. In finding the acceptance or rejection region we have to use critical values given in Statistical Tables. By comparing the computed value and the critical value decision is taken for accepting or rejecting H_0. If the computed value < critical value, we accept H_0, otherwise we reject H_0.

[JNTU (H) Dec. 2009 (Set No. 2)]

8.4 ERRORS OF SAMPLING

The main objective in sampling theory is to draw valid inferences about the population parameters on the basis of the sample results. In practice we decide to accept or to reject the lot after examining a sample from it. As such we have two types of errors.

(i) Type I error : Reject H_0 when it is true.

[JNTU (A) Dec. 2009 , (H) Dec. 2014, May 2017]

It is the error of rejecting Null hypothesis H_0, when it is true. When a null hypothesis is true, but the difference (of means) is significant and the hypothesis is rejected, then a Type I Error is made. The probability of making a Type I error is denoted by α , the level of significance. The probability of making a correct decision is then $(1-\alpha)$.

(ii) Type II error : Accept H_0 when it is wrong *i.e.*, accept H_0 when H_1 is true.

It is the error of accepting the null hypothesis H_0 when it is false.

In other words, if the Null Hypothesis is false but it is accepted by test, the error committed is called Type II error or β error.

If we write

P(Reject H_0 when it is true) = P(Type I error) = α

and P(Accept H_0 when it is wrong) = P(Type II error) = β

then α and β are called sizes of Type I and Type II errors respectively

i.e., $\alpha = P$ (Rejecting a good lot)

$\beta = P$ (Accepting a bad lot)

The sizes of Type I and Type II errors are also known as *producer's risk* and *consumer's risk* respectively.

The statistical testing of hypothesis aims at limiting the Type I error to a preassigned value (say : 1% or 5%) and to minimize the type II error. The only way to reduce both types of errors is to increase the sample size, if possible.

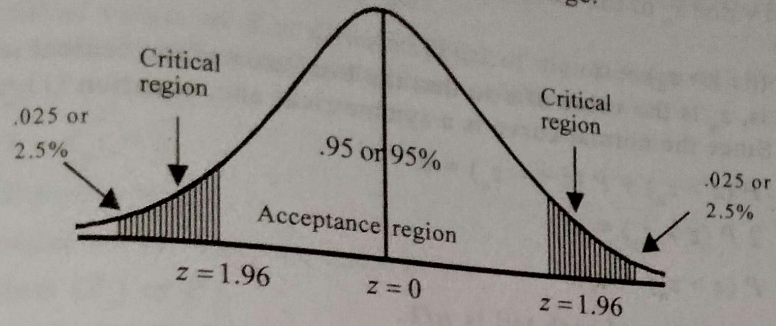
1. Critical Region

[JNTU (K) March 2014 (Set No. 1)]

Under a given hypothesis let the sampling distribution of a statistic t is approximately a normal distribution with mean $E(t)$ and S. D. $=\sigma_t =$ S. E. of t.

$$z = \frac{t - E(t)}{\text{S.E. of } t} = \frac{\text{Observed value} - \text{Expected value}}{\text{S.E. of } t}$$ is called the standardised

normal variate or test statistic or z - score and its distribution is the standard normal distribution with mean 0 and S.D. 1 when the sample is large.



From above figure, we see that if z lies between -1.96 and 1.96 then we are 95% confident that the hypothesis is true, since the area under the standard normal curve between $z = -1.96$ and $z = 1.96$ is 0.95 *i.e.*, 95% of the total area *i.e.* $P(-1.96 \leq z \leq 1.96) = 0.95$. But if for a simple random sample we find that z lies outside the range -1.96 to 1.96 *i.e.*, if $|z| > 1.96$, then we say that such an event occurs with probability of only 0.05 when the given hypothesis is true. In this case, we say that z - score differ significantly from the value expected under the hypothesis and hence the hypothesis H_0 is to be rejected at 5% level of significance. Thus if $|z| > 1.96$, the hypothesis H_0 is rejected at 5% level of significance. The set of z - scores outside the range -1.96 to 1.96 *i.e.*, $|z| > 1.96$ constitutes the *critical region* or *region of rejection* of the hypothesis or the *region of significance*. On the otherhand, the set of z - scores inside the range -1.96 to 1.96 is called the region of acceptance of the hypothesis. The values -1.96 and 1.96 are called *critical values* at 5% level of significance.

Similarly, we can define critical region at any other level of significance.

Definition. A region corresponding to a statistic 't', in the sample space S which leads to the rejection of H_0 is called *Critical Region* or *Rejection Region*. Those region which lead to the acceptance of H_0 give us a region called *Acceptance Region*.

In other words, the region in which a sample value falling is rejected is known as the critical region. In general we take two critical regions which cover 5% and 1 % areas of the normal curve.

[JNTU 2004S]

2. Critical Values or Significant Values :

The value of the test statistic, which separates the critical region (or rejection region) and the acceptance region is called the *critical value* or *significant value*. This value is dependent on

(*i*) the level of significance used, and

(*ii*) the alternative hypothesis, whether it is one-tailed or two-tailed.

For larger samples, corresponding to the statistic t, the variable $z = \dfrac{t - E(t)}{\text{S.E of } t}$ is normally distributed with mean 0 and variance 1.

The value of z given above under the null hypothesis is known as test statis... The critical value z_α of the test statistic at level of significance α for a two-tailed te... is given by

$$P(|z| > z_\alpha) = \alpha \qquad \ldots(1)$$

That is, z_α is the value of z so that the total area of the critical region on bo... tails is α. Since the normal curve is a symmetrical one, equation (1) implies,

$$P(z > z_\alpha) + P(z < -z_\alpha) = \alpha$$

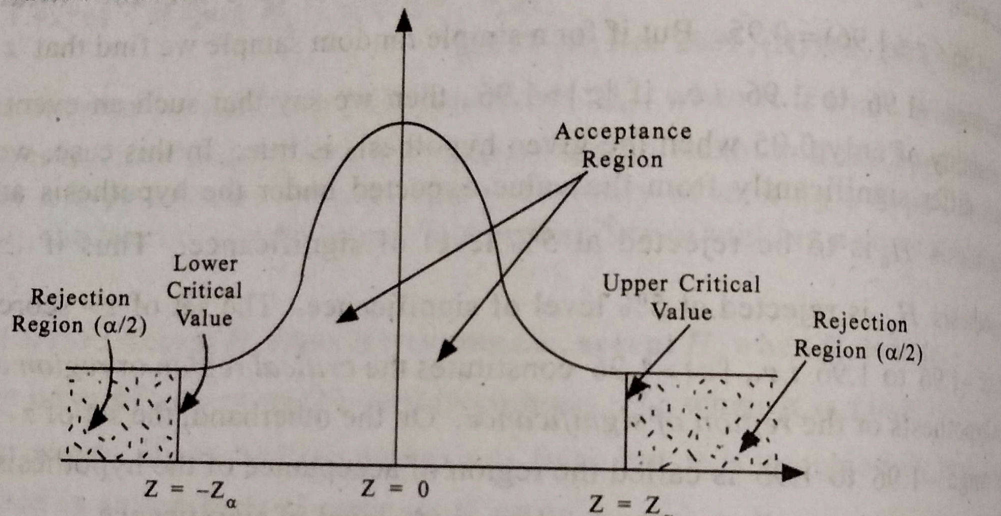
i.e. $\quad 2P(z > z_\alpha) = \alpha$

or $\quad P(z > z_\alpha) = \alpha/2$

That is, the area of each tail is $\alpha/2$.

The critical value z_α is that value such that the area to the right of z_α is $\alpha/2$ an... the area to the left of $-z_\alpha$ is $\alpha/2$. (Refer figure).

3. Two-Tailed test at level of significance 'α' :



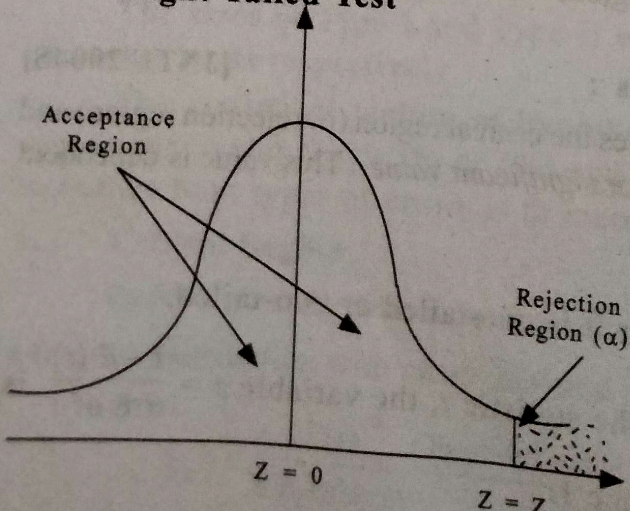
In the case of one-tailed alternative,

$P(Z > Z_\alpha) = \alpha$ if it is one-tailed (right)

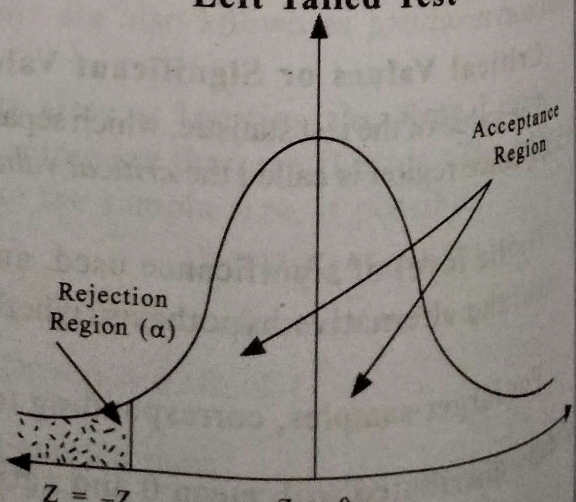
$P(Z < -Z_\alpha) = \alpha$ if it is one-tailed (left)

For Level of Significance 'α'

Right Tailed Test



Left Tailed Test



From the above figures, it is clear, that the critical value of Z for a single-tailed test (right or left) at level of significance 'α' is same as the critical value of Z for two-tailed test at level of significance '2α'.

The critical values of Z at different level of significance (α) for both single tailed and two-tailed tests are calculated from equations

$$P(|Z| > Z_\alpha) = \alpha$$

$$P(Z > Z_\alpha) = \alpha$$

$$P(Z < -Z_\alpha) = \alpha$$

using the normal tables. They are listed below.

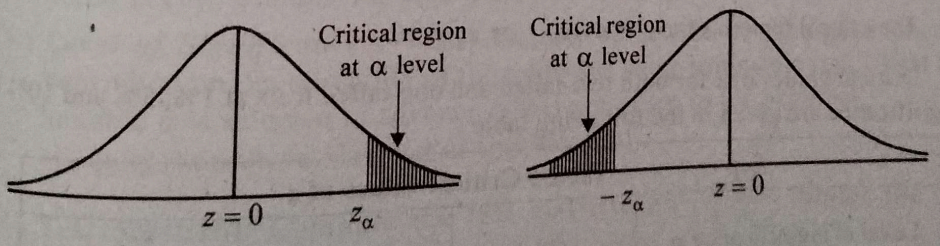
Critical values (Z_α) of Z :

	Level of Significance		
	1% (.01)	**5% (.05)**	**10%(.1)**
Two-tailed test	$\lvert Z_\alpha \rvert = 2.58$	μ_0 $\lvert Z_\alpha \rvert = 1.96$	$\lvert Z_\alpha \rvert = 1.645$
Right-tailed test	$Z_\alpha = 2.33$	$Z_\alpha = 1.645$	$Z_\alpha = 1.28$
Left-tailed test	$Z_\alpha = -2.33$	$Z_\alpha = -1.645$	$Z_\alpha = -1.28$

8.5 ONE - TAILED AND TWO - TAILED TESTS

[JNTU (A) Dec. 2009, (K) Nov. 2011, (H) May 2012 (Set No. 1), May 2017, Sept. 2017]

If we have to test whether the population mean μ has a specified value μ_0, then the Null Hypothesis is $H_0 : \mu = \mu_0$ and the Alternative Hypothesis may be (*i*) $H_1 : \mu \neq \mu_0$ (*i.e.*, $\mu > \mu_0$ or $\mu < \mu_0$) or (*ii*) $H_1 : \mu > \mu_0$ or (*iii*) $H_1 : \mu < \mu_0$. The Alternative Hypothesis in (*i*) is known as a two-tailed (*i.e.*, both right and left tail) alternatives and the alternative hypothesis in (*ii*) and (*iii*) are known as right - tailed and left - tailed alternatives respectively.



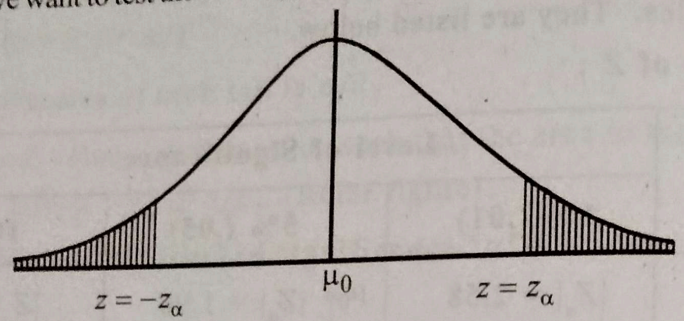
One-Tailed Test :

If the Alternative Hypothesis H_1 in a test of a statistical hypothesis be one - tailed (*i.e.*, either right - tailed or left - tailed but not both), then the test is called a *one-tailed test*. For example, to test whether the population mean $\mu = \mu_0$, we have $H_0 : \mu = \mu_0$ against the alternative hypothesis H_1 given by

(*i*) $H_1 : \mu > \mu_0$ (right tailed) or (*ii*) $H_1 : \mu < \mu_0$ (left - tailed) and the corresponding test is a single - tailed or one-tailed or one sided. In the right - tail test $H_1 : \mu > \mu_0$, the critical region (or rejection region) $z > z_\alpha$ lies entirely in the right tail of the sampling distribution of sample mean $\bar{x}$ with area equal to the level of significance α (see figure). Similarly, in the left - tailed test $(H_1 : \mu < \mu_0)$, the critical region $z < -z_\alpha$ lies entirely in the left tail of the sampling distribution of the sample mean $\bar{x}$ with area equal to the level of significance α (see figure).

Two-Tailed Test :

Suppose we want to test the Null Hypothesis $H_0 : \mu = \mu_0$ against the Alternative Hypothesis $H_1 : \mu \neq \mu_0$.



Since H_1 is two - tailed alternative hypothesis, the critical region under the curve is equally distributed on both sides of the mean.

Thus, the critical area under the right - tail = The critical area under the left - tail

$$= \text{Half of the total area}$$

$$= \frac{1}{2} \text{ probability of rejection}$$

$$= \frac{\alpha}{2}$$

with critical statistic $Z_{\alpha/2}$, where α is the level of significance.

The critical region is then, $z \leq -z_{\alpha/2}$ or $z_{\alpha/2} \leq z$.

Critical values of x for both two-tailed and one-tailed tests at 1%, 5% and 10% level of significance are given in the following table :

Table : Critical values of z

Level of siginificance α	1%	5%	10%
Critical values for two-tailed test	$\lvert z_\alpha \rvert = 2.58$	$\lvert z_\alpha \rvert = 1.96$	$\lvert z_\alpha \rvert = 1.645$
Critical values for Right-tailed test	$z_\alpha = 2.33$	$z_\alpha = 1.645$	$z_\alpha = 1.28$
Critical values for Left-tailed test	$z_\alpha = -2.33$	$z_\alpha = -1.645$	$z_\alpha = -1.28$

Applying one-tailed or two-tailed test for a particular problem depends entirely on the nature of the Alternative Hypothesis. If the alternative test is two-tailed we apply two-tailed test and if Alternative Hypothesis is one-tailed, we apply one-tailed test.

For example, to test whether a coin is biased or not, two - tailed test should be used, since a biased coin gives either more number of heads than tails (which corresponds to right tail) or more number of tails than heads (which corresponds to left tail)

Example : Consider two population brands of bulbs one manufactured by routine process (mean μ_1) and the other manufactured by new technique (mean μ_2). If we want to test if the bulbs differ significantly then the hypothesis is $H_0 : \mu_1 = \mu_2$ and the alternative hypothesis will be $H_0 : \mu_1 \neq \mu_2$. This gives us a two-tailed test. Suppose if we want to test if the bulbs produced by new process (μ_2) have higher average life than those produced by standard process (μ_1), then we have

$$H_0 : \mu_1 = \mu_2 \text{ and } H_1 : \mu_1 < \mu_2$$

In this case we have to adopt a left-tailed test.

If we want to test whether the product of new process (μ_2) is inferior to that of standard process (μ_1), then we have

$$H_0 : \mu_1 = \mu_2 \text{ and } H_1 : \mu_1 > \mu_2$$

which gives a right-tailed test.

Hence the decision about applying a two-tail test or a single-tail (left or right) test will mainly depend on the problem under study.

PROCEDURE FOR TESTING OF HYPOTHESIS :

[JNTU (H) Nov. 2009 (Set No.2,4), (K) Nov. 2011, Dec. 2013 (Set No. 3)]

Various steps involved in testing of Hypothesis are given below : Infact, the same steps are followed for conducting all tests of significance.

Step 1 : ***Null Hypothesis :*** Define or set up a Null Hypothesis H_0 taking into consideration the nature of the problem and data involved.

Step 2 : ***Alternative Hypothesis :*** Set up the Alternative Hypothesis H_1 so that we could decide whether we should use one-tailed or two-tailed test.

Step 3 : ***Level of Significance :*** Select the appropriate level of significance(α) depending on the reliability of the estimates and permissible risk. That is, a suitable α is selected in advance if it is not given in the problem. (Usually we choose 5% level of significance)

Step 4 : ***Test Statistic:*** Compute the test statistic $Z = \frac{t - E(t)}{\text{S.E of } t}$ under the null hypothesis.

Here t is a sample statistic and S.E. is the standard error of t.

Step 5 : ***Conclusion :*** We compare the computed value of the test statistic Z with the critical value Z_α at given level of significance (α).

If $|Z| < Z_\alpha$, (that is, if the absolute value of the calculated value of Z is less than the critical value Z_α) we conclude that it is not significant. We accept the null hypothesis.

If $|Z| > Z_\alpha$ then the difference is significant and hence the null hypothesis is rejected at the level of significance α.

Clearly,

For two-tailed test :

If $|Z| < 1.96$ accept H_0 at 5% level of significance.

If $|Z| > 1.96$ reject H_0 at 5% level of significance.

If $|Z| < 2.58$, accept H_0 at 1% level of significance.

If $|Z| > 2.58$ reject H_0 at 1% level of significance.

For single-tailed (right or left) test :

If $|Z| < 1.645$, accept H_0 at 5% level of significance.

If $|Z| > 1.645$, reject H_0 at 5% level of significance.

If $|Z| < 2.33$ accept H_0 at 1% level of significance.

If $|Z| > 2.33$ reject H_0 at 1% level of significance.

8.6. TEST OF SIGNIFICANCE FOR LARGE SAMPLES

If the sample size, $n > 30$, then we consider such samples as large samples. The tests of significance used in large samples are different from those used in small samples because small samples fail to satisfy the assumptions under which large sample analysis is done. If n is large, the distributions, such as Binomial, Poisson, Chi-square etc. are closely approximated by normal distributions. Therefore, for large samples, the sampling distribution of a statistic is approximately a normal distribution.

Suppose we wish to test the hypothesis that the probability of success in such trial is p. Assuming it to be true, the mean μ and the standard deviation σ of the sampling distribution of number of successes are np and $\sqrt{npq}$ respectively.

If x be the observed number of successes in the sample and Z is the standard normal variate then $Z = \dfrac{x-\mu}{\sigma}$.

Thus we have the following test of significance :

(*i*) If $|Z| < 1.96$, the difference between the observed and expected number of successes is not significant.

(*ii*) If $|Z| > 1.96$, the difference is significant at 5% level of significance.

(*iii*) If $|Z| > 2.58$, the difference is significant at 1% level of significance.

Assumptions for Large Samples :

The following are the assumptions under which significance tests are applied.

1. The random sampling distribution of statistic has the properties of the normal curve. This may not hold good in case of small samples.
2. Values (*i.e.*, statistic) given by the samples are sufficiently close to the population values (*i.e.*, parameters) and can be used in its place for calculating the standard error (S. E.) of the estimate.

For example, for computing the S. E., if the S. D. of the universe is not known, it can be replaced by the S. D. of the sample. In case of small samples, this is not possible.

SOLVED EXAMPLES

Example 1 : A coin was tossed 960 times and returned heads 183 times. Test the hypothesis that the coin is unbiased. Use a 0.05 level of significance.

Solution : Here $n = 960$, $p =$ Probability of getting head $= 1/2$

$$\therefore\ q = 1 - p = \frac{1}{2}; \quad \mu = np = 960\left(\frac{1}{2}\right) = 480$$

$$\sigma = \sqrt{npq} = \sqrt{(np)\,q} = \sqrt{480 \times \frac{1}{2}} = \sqrt{240} = 15.49$$

$x =$ number of successes $= 183$

1. **Null Hypothesis** H_0 : The coin is unbiased
2. **Alternative Hypothesis** H_1 : The coin is biased
3. **Level of significance :** $\alpha = 0.05$
4. **The test statistic is** $Z = \dfrac{x-\mu}{\sigma} = \dfrac{183-480}{15.49} = \dfrac{-297}{15.49} = -19.17$

$$\therefore\ |Z| = 19.17$$

As $|Z| > 1.96$, the null hypothesis H_0 has to be rejected at 5% level of significance and we conclude that the coin is biased.

Example 2 : A coin was tossed 400 times and returned heads 216 times. Test the hypothesis that the coin is unbiased. Use a 0.05 Level of significance.

[JNTU (K) Nov. 2009 (Set No. 3)]

Solution : Here $n = 400$

$$p = \text{Probability of getting head} = \frac{1}{2}$$

$$q = 1 - p = 1 - \frac{1}{2} = \frac{1}{2}, \quad \mu = np = 400\left(\frac{1}{2}\right) = 200$$

$$\sigma = \sqrt{npq} = \sqrt{np(q)} = \sqrt{200 \times \frac{1}{2}} = \sqrt{100} = 10$$

$x =$ number of successes $= 216$

1. **Null Hypothesis** H_0 : The coin is unbiased
2. **Alternative Hypothesis** H_1 : The coin is biased
3. **Level of significance :** $\alpha = 0.05$
4. **The test statistic is** $z = \dfrac{x-\mu}{\sigma}$

$$\therefore\ z = \frac{216-200}{10} = \frac{16}{10} = 1.6$$

As $|z|<1.96$, the null hypothesis H_0 has to be accepted and we conclude that the coin is unbiased.

Example 3 : A die is tossed 960 times and it falls with 5 upwards 184 times. Is the die unbiased at a level of significance of 0.01 ?

Solution : Here $n=960$, $p=$ The probability of throwing 5 with one die $=1/6$

$$\therefore q=1-\frac{1}{6}=\frac{5}{6};\quad \mu=np=960\left(\frac{1}{6}\right)=160$$

$$\sigma=\sqrt{npq}=\sqrt{160\times\frac{5}{6}}=11.55$$

$x=$ number of successes $=184$

1. **Null Hypothesis** H_0 : The die is unbiased
2. **Alternative Hypothesis** H_1 : The die is biased
3. **Level of significance :** $\alpha=0.01$
4. **The test statistic is** $Z=\dfrac{x-\mu}{\sigma}=\dfrac{184-160}{11.55}=\dfrac{24}{11.55}=2.08$

As $|Z|<2.58$, the null hypothesis H_0 has to be accepted at 1 % level of significance and we conclude that the die is unbiased.

Note : As $|Z|>1.96$, the null hypothesis H_0 has to be rejected at 5% level of significance and we conclude that the die is biased.

Example 4 : A die is tossed 256 times and it turns up with an even digit 150 times. Is the die biased ?

Solution : Here $n=256$, $p=$ The probability of getting an even digit (2 or 4 or 6) $=\frac{3}{6}=\frac{1}{2}$

$$q=1-p=\frac{1}{2};\quad \mu=np=256\left(\frac{1}{2}\right)=128$$

$$\sigma=\sqrt{npq}=\sqrt{np\,(q)}=\sqrt{128\times\frac{1}{2}}=\sqrt{64}=8$$

$x=$ number of successes $=150$

1. **Null Hypothesis** H_0 **:** The die is unbiased
2. **Alternative Hypothesis** H_1**:** The die is biased
3. **Level of significance :** $\alpha=0.05$
4. **The test statistic is** $Z=\dfrac{x-\mu}{\sigma}$ *i.e.*, $Z=\dfrac{150-128}{8}=\dfrac{22}{8}=2.75$

As $|Z|>1.96$, the null hypothesis H_0 has to be rejected at 5% level of significance and we conclude that the die is biased.

Example 5 : Mean of population $= 0.700$, mean of the sample $= 0.742$, standard deviation of the sample $= 0.040$, sample size $= 10$. Test the null hypothesis for population mean $= 0.700$

[JNTU (H) III yr. Nov. 2015]

Solution : Given $\mu =$ Mean of population $= 0.7$

$\overline{x} =$ Mean of sample $= 0.742$

$\sigma =$ Standard deviation of the sample $= 0.040$

and $n =$ Sample size $= 10$

The test statistic is $z = \dfrac{\overline{x} - \mu}{\sigma / \sqrt{n}}$

i.e., $z = \dfrac{0.742 - 0.7}{0.04 / \sqrt{10}} = 3.32$

Since $|z| > 1.96$, the sample is not from the population whose mean is 0.7.

Under large sample tests, we will see four important tests to test the significance.

1. Testing of significance for single proportion.
2. Testing of significance for difference of proportions.
3. Testing of significance for single mean.
4. Testing of significance for difference of means.

8.7 TEST OF SIGNIFICANCE OF A SINGLE MEAN - LARGE SAMPLES

Let a random sample of size $n\,(n > 30)$ has the sample mean $\overline{x}$, and population mean μ. Also the population mean μ has a specified value μ_0.

Working Rule :

1. **The Null Hypothesis** is $H_0 : \overline{x} = \mu$ *i.e.,* "there is no significance difference between the sample mean and population mean" or the sample has been drawn from the parent population.

2. **The Alternative Hypothesis** is ***(i)*** $H_1 : \overline{x} \neq \mu\ (\mu \neq \mu_0)$ or

   (ii) $H_1 : \overline{x} > \mu\ (\mu > \mu_0)$ or ***(iii)*** $H_1 : \overline{x} < \mu\ (\mu < \mu_0)$

   Since n is large, the sampling distribution of $\overline{x}$ is approximately normal.

3. **Level of Significance.** Set the level of significance α.

4. **The Test Statistic :**

   We have the following two cases.

   Case I : When the standard deviation σ of population is known.

   In this case, standard Error of Mean, S. E. $(\overline{x}) = \dfrac{\sigma}{\sqrt{n}}$, where $n =$ sample size,

   $\sigma =$ s.d. of population.

$\therefore$ The test statistic is

$$z = \frac{\bar{x}-\mu}{\text{S.E.}(\bar{x})} = \frac{\bar{x}-\mu}{\sigma/\sqrt{n}}, \text{ where } \mu \text{ is the sample mean.}$$

Case II : When the standard deviation σ of population is not known.

In this case, we take s, the s.d. of sample to compute the S. E. of mean

$$\therefore \quad \text{S.E.}(\bar{x}) = \frac{s}{\sqrt{n}}$$

Hence the test statistic is

$$z = \frac{\bar{x}-\mu}{\text{S.E.}(\bar{x})} = \frac{\bar{x}-\mu}{s/\sqrt{n}}$$

5. Find the critical value z_α of z at the level of significance α from the normal table.

6. **Decision :**

   (*a*) If $|z| < z_\alpha$, we accept the Null Hypothesis H_0.

   (*b*) If $|z| > z_\alpha$, we reject the Null Hypothesis H_0.

   The rejection rule for $H_0 : \bar{x} = \mu$ (or $\mu = \mu_0$) is given below :

Table : Critical values of z

Level of siginificance α	1%	5%	10%
Critical region for $\mu \neq \mu_0$	$\|z\| > 2.58$	$\|z\| > 1.96$	$\|z\| > 1.645$
Critical region for $\mu > \mu_0$	$z > 2.33$	$z > 1.645$	$z > 1.28$
Critical region for $\mu < \mu_0$	$z < -2.33$	$z < -1.645$	$z < -1.28$

Note: 1. We reject Null Hypothesis H_0 when $|z| > 3$ without mentioning any level of significance.

The test statistic is, $z = \dfrac{\bar{x}-\mu}{\sigma/\sqrt{n}}$ where σ is the *S.D* of the population.

If the population *S.D* is not known, then use the statistic

$$z = \frac{\bar{x}-\mu}{S/\sqrt{n}} \text{ where } S \text{ is the sample } S.D.$$

2. The values $\bar{x} \pm 1.96\dfrac{\sigma}{\sqrt{n}}$ are called 95% fiducial limits or confidence limits for the mean of the population corresponding to the given sample.

Similarly, $\bar{x} \pm 2.58\dfrac{\sigma}{\sqrt{n}}$ (or) $[\bar{x} - 2.58\,(\text{S.E. of } \bar{x}), \bar{x} + 2.58\,(\text{S.E. of } \bar{x})]$ are called 99% confidence limits and $\bar{x} \pm 2.33\dfrac{\sigma}{\sqrt{n}}$ are called 98% confidence limits.

SOLVED EXAMPLES

Example 1 : According to the norms established for a mechanical (or an electrical) aptitude test, persons who are 18 years old have an average height of 73.2 with a standard deviation of 8.6. If 4 randomly selected persons of that age averaged 76.7, test the hypothesis $\mu = 73.2$ against the alternative hypothesis $\mu > 73.2$ at the 0.01 level of significance.

[JNTU 2004, 2005S (Set No. 1), (H) Nov. 2012, Sept. 2017]

Solution : Given $n = 4$, $\mu = 73.2$, $\bar{x}$ = mean of the sample = 76.7 and σ = S.D of population = 8.6

1. **Null Hypothesis** $H_0 : \mu = 73.2$
2. **Alternative Hypothesis** $H_1 : \mu > 73.2$ (Right-tailed test)
3. **Level of significance :** $\alpha = 99\%$ (or probability is 0.01)
4. **The test statistic is** $z = \dfrac{\bar{x}-\mu}{\dfrac{\sigma}{\sqrt{n}}} = \dfrac{76.7-73.2}{\dfrac{8.6}{\sqrt{4}}} = \dfrac{3.5}{4.3} = 0.814$

Tabulated value of z at 99% level of significance is 2.33.
Hence calculated z < tabulated z

$\therefore$ The null hypothesis H_0 is accepted. That is, $\bar{x}$ and μ do not differ significantly.

Example 2 : A sample of 64 students have a mean weight of 70 kgs. Can this be regarded as a sample from a population with mean weight 56 kgs and standard deviation 25 kgs.

[JNTU 2006, (A) Nov. 2010 (Set No. 2)]

Solution :

Given $\bar{x}$ = mean of the sample = 70 kgs

μ = mean of the population = 56 kgs

σ = S.D of population = 25 kgs

and n = sample size = 64

1. **Null Hypothesis** H_0**:** A sample of 64 students with mean weight of 70 kgs can be regarded as a sample from a population with mean weight 56 kgs and standard deviation 25 kgs.
2. **Alternative Hypothesis** H_1 **:** Sample cannot be regarded as one coming from the population.
3. **Level of significance :** $\alpha = 0.05$ (assumption)
4. **The test statistic is** $z = \dfrac{\bar{x}-\mu}{\dfrac{\sigma}{\sqrt{n}}} = \dfrac{70-56}{\left(\dfrac{25}{\sqrt{64}}\right)} = 4.48$

5. The null hypothesis H_0 is rejected, since $|Z| > 1.645$

Note : The null hypothesis can be rejected even at 1% level of significance.

Example 3 : An oceanographer wants to check whether the depth of the ocean in a certain region is 57.4 fathoms, as had previously been recorded. What can he conclude at the 0.05 level of significance, if readings taken at 40 random locations in the given region yielded a mean of 59.1 fathoms with a standard deviation of 5.2 fathoms.

[JNTU 2003, 2003 S, 2004, (H) Sept. 2017]

<u>Solution</u> : Given $n = 40$, $\bar{x} = 59.1$ and $\sigma = 5.2$

1. **Null Hypothesis** $H_0 : \mu = 57.4$
2. **Alternative hypothesis** $H_1 : \mu \neq 57.4$
3. **Level of significance :** $\alpha = 0.05$
4. **The test statistic** is $Z = \dfrac{\bar{x}-\mu}{\frac{\sigma}{\sqrt{n}}} = \dfrac{59.1-57.4}{5.2/\sqrt{40}} = 2.067$

Tabulated value of Z at 5% level of significance is 1.96

Hence calculated $Z >$ tabulated Z.

$\therefore$ The null hypothesis H_0 is rejected.

Example 4 : In a random sample of 60 workers, the average time taken by them to get to work is 33.8 minutes with a standard deviation of 6.1 minutes. Can we reject the null hypothesis $\mu = 32.6$ minutes in favour of alternative null hypothesis $\mu > 32.6$ at $\alpha = 0.025$ level of significance.

[JNTU 2005 (Set No.1)]

<u>Solution</u> : Given $n = 60$, $\bar{x} = 33.8$, $\mu = 32.6$ and $\sigma = 6.1$

1. **Null Hypothesis** $H_0 : \mu = 32.6$
2. **Alternative Hypothesis** $H_1 : \mu > 32.6$
3. **Level of significance :** $\alpha = 0.025$
4. **The test statistic** is $Z = \dfrac{\bar{x}-\mu}{\frac{\sigma}{\sqrt{n}}} = \dfrac{33.8-32.6}{6.1/\sqrt{60}} = \dfrac{1.2}{0.7875} = 1.5238$

Tabulated value of Z at 0.025 level of significance is 2.58.

Hence calculated $Z <$ tabulated Z

$\therefore$ The null hypothesis H_0 is accepted.

Example 5 : A sample of 900 members has a mean of 3.4 cms and S.D 2.61 cms this sample has been taken from a large population of mean 3.25 cm and S.D 2.61 cms. If the population is normal and its mean is unknown find the 95% fiducial limits of true mean.

[JNTU (H) May 2011 (Set No. 1), (K) Nov. 2011 (Set No.

Solution :

Given $n = 900$, $\mu = 3.25$,

$\bar{x} = 3.4$ cm, $\sigma = 2.61$,

and $S = 2.61$

1. **Null Hypothesis H_0 :** Assume that the sample has been drawn from the population with mean $\mu = 3.25$.

2. **Alternative Hypothesis H_1 : $\mu \neq 3.25$.**

3. The test statistic is, $Z = \dfrac{\bar{x} - \mu}{\dfrac{\sigma}{\sqrt{n}}} = \dfrac{3.4 - 3.25}{2.61/\sqrt{900}} = 1.724$

   i.e., $Z = 1.724 < 1.96$

   $\therefore$ We accept the null hypothesis H_0.

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

   95% confidence limits are given by

$$\bar{x} \pm 1.96 \frac{\sigma}{\sqrt{n}} = 3.4 \pm 1.96 \times \frac{2.61}{\sqrt{900}} = 3.4 \pm 0.1705$$

   i.e., 3.57 and 3.2295

Example 6 : A sample of 400 items is taken from a population whose standard deviation is 10. The mean of the sample is 40. Test whether the sample has come from a population with mean 38. Also calculate 95% confidence interval for the population.

[JNTU 2005, (H) Dec. 2011 (Set No. 1)]

Solution : Given $n = 400$, $\bar{x} = 40$, $\mu = 38$ and $\sigma = 10$

1. **Null Hypothesis H_0 : $\mu = 38$**

2. **Alternative Hypothesis H_1 : $\mu \neq 38$**

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

4. **The test statistic is,** $Z = \dfrac{\bar{x} - \mu}{\dfrac{\sigma}{\sqrt{n}}} = \dfrac{40 - 38}{10/\sqrt{400}} = 4$

   i.e., $Z = 4 > 1.96$

   $\therefore$ We reject the null hypothesis H_0.

   i.e., The sample is not from the population whose mean is 38.

   95% confidence interval is $\left(\bar{x} - 1.96.\dfrac{\sigma}{\sqrt{n}}, \bar{x} + 1.96.\dfrac{\sigma}{\sqrt{n}}\right)$

   i.e., $\left(40 - \dfrac{1.96\,(10)}{\sqrt{400}}, 40 + \dfrac{1.96\,(10)}{\sqrt{400}}\right)$

or $\left(40-\frac{19.6}{20}, 40+\frac{19.6}{20}\right)$ or $(40 - 0.98,\ 40 + 0.98)$

i.e., (39.02, 40.98)

Example 7 : An ambulance service claims that it takes on the average less than 10 minutes to reach its destination in emergency calls. A sample of 36 calls has a mean of 11 minutes and the variance of 16 minutes. Test the claim at 0.05 level significance.

[JNTU 2005, (H) May 2012 (Set No.4)]

Solution : Given $n = 36$, $\bar{x} = 11$, $\mu = 10$ and $\sigma = \sqrt{16} = 4$

1. **Null Hypothesis** $H_0 : \mu = 10$
2. **Alternative Hypothesis** $H_1 : \mu < 10$
3. **Level of significance :** $\alpha = 0.05$
4. **The test statistic** is, $Z = \dfrac{\bar{x}-\mu}{\frac{\sigma}{\sqrt{n}}} = \dfrac{11-10}{4/\sqrt{36}} = \dfrac{6}{4} = 1.5$

Tabulated value of Z at 5% level of significance is 1.645.

Hence calculated $Z <$ tabulated Z

$\therefore$ We accept the null hypothesis H_0.

Example 8 : It is claimed that a random sample of 49 tyres has a mean life of 15200 km. This sample was drawn from a population whose mean is 15150 kms and a standard deviation of 1200 km. Test the significance at 0.05 level.

[JNTU 2005, 2006S, (K) May 2013 (Set No.2)]

Solution : Given $n = 49$, $\bar{x} = 15200$, $\mu = 15150$ and $\sigma = 1200$

1. **Null Hypothesis** $H_0 : \mu = 15150$
2. **Alternative Hypothesis** $H_1 : \mu \neq 15150$
3. **Level of significance :** $\alpha = 0.05$
4. **Critical region :** Accept the null hypothesis if $-1.96 < Z < 1.96$
5. **The test statistic** is, $Z = \dfrac{\bar{x}-\mu}{\frac{\sigma}{\sqrt{n}}} = \dfrac{15200-15150}{1200/\sqrt{49}} = 0.2917$

Since $|Z| < 1.96$ therefore, we accept the null hypothesis.

Example 9 : An insurance agent has claimed that the average age of policy holders who issue through him is less than the average for all agents which is 30.5 years. A random sample of 100 policy holders who had issued through him gave the following age distribution.

Age	16-20	21-25	26-30	31-35	36-40
No. of persons	12	22	20	30	16

Calculate the Arithmetic mean and Standard deviation of this distribution and use these values to test his claim at 5% level of significance.

Solution : Take $A = 28,\ d_i = x_i - A$

$$\therefore \text{ A.M.} = \bar{x} = A + \frac{h\sum f_i d_i}{N} = 28 + \frac{5\times 16}{100} = 28.8$$

$$S.D : S = h\sqrt{\frac{\sum fd^2}{N} - \left(\frac{\sum fd}{N}\right)^2} = 5\cdot\sqrt{\frac{164}{100} - \left(\frac{16}{100}\right)^2} \quad [\because h = 5]$$

$$= 6.35$$

1. **Null Hypothesis H_0 :** The sample is drawn from a population with mean μ i.e. $\bar{x}$ and μ do not differ significantly where $\mu = 30.5$ years.
2. **Alternative Hypothesis H_1 :** $\mu < 30.5$ years (left tail test)

Now, $\bar{x} = 28.8$, $S = 6.35$, $\mu = 30.5$ years and $n = 100$

3. **The test statistic** is, $Z = \dfrac{\bar{x}-\mu}{S/\sqrt{n}} = \dfrac{28.8-30.5}{6.35/\sqrt{100}} = -2.677$

$\therefore |Z| = 2.68.$

Tabulated value of Z at 5% level of significance is 1.645 (left tail test).

Here calculated $Z >$ tabulated Z.

$\therefore$ The Null hypothesis H_0 is rejected.

i.e., $\bar{x}$ and μ differ significantly.

i.e., The sample is not drawn from a population with mean $\mu = 30.5$ years

Example 10 : The mean life time of a sample of 100 light tubes produced by a company is found to be 1560 hrs with a population S.D of 90 hrs. Test the hypothesis for $\alpha = 0.05$ that the mean life time of the tubes produced by the company is 1580 hrs.

[JNTU (A) Dec. 2009 (Set No. 1)]

Solution : Given $\bar{x}$ = Mean of the sample = 1560 hrs

μ = Mean of the population = 1580 hrs

n = Sample size = 100

σ = Standard deviation = 90 hrs

1. **Null Hypothesis** $H_0 : \mu = 1580$
2. **Alternative Hypothesis** $H_1 : \mu \neq 1580$
3. **Level of significance :** $\alpha = 0.05$
4. **The test statistic is** $z = \dfrac{\bar{x}-\mu}{\sigma/\sqrt{n}} = \dfrac{1560-1580}{90/\sqrt{100}} = \dfrac{-20}{9}$

$\therefore\ |z| = \frac{20}{9} = 2.22$

$z_{\alpha/2}$ for $\alpha = 0.05$ is 1.96. Since $|z| > 1.96$, the null Hypothesis H_0 is rejected.

$\therefore\ \mu \neq 1580$.

Note: $z_{\alpha/2}$ for $\alpha = 0.01$ is 2.58

Since $|z| < 2.58$, the null Hypothesis H_0 can be accepted at 1% level of significance.

Example 11 : The length of life X of certain computers is approximately normally distributed with mean 800 hours and standard deviation 40 hours. If a random sample of 30 computers has an average life of 788 hours, test the null hypothesis that μ=800 hours against the alternative that $\mu \neq 800$ hours at (*i*) 0.5% (*ii*) 1% (*iii*) 4% (*iv*) 5% (*v*) 10% (*vi*) 15% level.

<u>**Solution**</u> : The Null Hypothesis is $H_0 = 800$ hours.

The Alternative Hypothesis is $H_1 : \mu \neq 800$ hours.

Since H_1 is two-tailed (*i.e.* two sided), we are to use two -tailed test.

Let us assume that H_0 is true.

The test statistic is $z = \frac{\bar{x} - \mu}{\sigma / \sqrt{n}}$

Here $\bar{x}$ = sample mean = 788 hours

n = sample size = 30

σ = standard deviation = 40

$\therefore\ z = \frac{788 - 800}{40/\sqrt{30}} = -1.643$ and $|z| = 1.643$

(*i*) Since $|z| < 2.81$, we accept the Null Hypothesis H_0 at 0.5% level of significance.

(*ii*) Since $|z| < 2.58$, we accept the Null Hypothesis H_0 at 1% level of significance.

(*iii*) Since, $|z| < 2.06$, we accept H_0 at 4% level of significance.

(*iv*) Since, $|z| < 1.96$, we accept H_0 at 5% level of significance.

(*v*) Since, $|z| < 1.645$, we accept H_0 at 10% level of significance.

(*vi*) Since, $|z| > 1.44$, we reject H_0 at 15% level of significance.

Example 12 : In 64 randomly selected hours of production, the mean and the standard deviation of the number of acceptance pieces produced by an automatic stamping machine are $x = 1.038$ and $\sigma = .146$

At the .05 level of significance does this enable us to reject the null hypothesis $\mu = 1.000$ against the alternative hypothesis $\mu > 1.000$?

[JNTU (H) Nov. 2010 (Set No. 4)]

Solution : Let the Null Hypothesis be $H_0 : \mu = 1.000$

Then the Alternative Hypothesis is $H_1 : \mu > 1.000$

Here $\bar{x}$ = Mean of the sample $= 1.038$

μ = Mean of the population $= 1.000$

σ = S.D. of the population $= 0.146$

and n = Sample size $= 64$

$\therefore$ The test statistic is $z = \dfrac{\bar{x} - \mu}{\sigma / \sqrt{n}}$

$$= \frac{1.038 - 1.000}{0.146 / \sqrt{64}} = \frac{0.038}{0.146 / 8}$$

$$= 2.082$$

Thus we see that $z = 2.082 > 1.645$

Hence, we reject the Null Hypothesis H_0 at 5% level of significance and conclude that the mean of the population $\mu > 1.000$

Example 13 : A trucking rm suspects the claim that average life of certain tyres is atleast 28,000 miles. To check the claim the rm puts 40 of these tyres on its trucks and gets a mean life time of 27463 miles with a standard deviation of 1348 miles. Can the claim be true?

[JNTU (H) Apr. 2012 (Set No. 2)]

Solution : The Null Hypothesis is $H_0 : \mu = 28{,}000$ miles

The Alternative Hypothesis is $H_1 : \mu \neq 28{,}000$

Since H_1 is two - tailed (*i.e.*, two - sided), we are to use two - tail test.

Let us assume that H_0 is true.

The test statistic is $z = \dfrac{\bar{x} - \mu}{\text{S.E. of } \bar{x}} = \dfrac{\bar{x} - \mu}{\sigma / \sqrt{n}}$

Here $\bar{x}$ = sample mean $= 27463$ miles, $n = 40$ and $\sigma = s = 1348$ miles

[$\because$ population S. D. σ is not known]

$$\therefore z = \frac{27463 - 28000}{1348 / \sqrt{40}} = \frac{-537(\sqrt{40})}{1348} = -2.52$$

and $|z| = |-2.52| = 2.52 > 1.96$

Hence, the Null Hypothesis H_0 is rejected at 5% level of significance and we conclude that the mean life of tyres, cannot be taken as 28000 miles.

8.8 TEST FOR EQUALITY OF TWO MEANS - LARGE SAMPLES

(Test of significance for difference of means of two large samples)

Let $\bar{x}_1$ and $\bar{x}_2$ be the sample means of two independent large random samples sizes n_1 and n_2 drawn from two populations having means μ_1 and μ_2 and standard deviations σ_1 and σ_2. To test whether the two population means are equal.

Let the Null Hypothesis be $H_0 : \mu_1 = \mu_2$

Then the Alternative Hypothesis is $H_1 : \mu_1 \neq \mu_2$

S. E. of $(\bar{x}_1 - \bar{x}_2) = \sqrt{\dfrac{\sigma_1^2}{n_1} + \dfrac{\sigma_2^2}{n_2}}$, where σ_1 and σ_2 are the S. D. of the two populations

To test whether there is any significant difference between $\bar{x}_1$ and $\bar{x}_2$, we have to use the statistic

$$z = \frac{(\bar{x}_1 - \bar{x}_2) - \delta}{\text{S.E. of } (\bar{x}_1 - \bar{x}_2)} = \frac{(\bar{x}_1 - \bar{x}_2) - \delta}{\sqrt{\dfrac{\sigma_1^2}{n_1} + \dfrac{\sigma_2^2}{n_2}}}$$

where $\delta = \mu_1 - \mu_2$ (= given constant.)

If $\delta = 0$, the two populations have the same means.
If $\delta \neq 0$ the two populations are different.

Under $H_0 : \mu_1 = \mu_2$, the test statistic becomes $z = \dfrac{\bar{x}_1 - \bar{x}_2}{\sqrt{\dfrac{\sigma_1^2}{n_1} + \dfrac{\sigma_2^2}{n_2}}}$

is approximately normally distributed with mean 0 and S. D. 1.

Note : If the samples have been drawn from the population with common S. D. σ, then $\sigma_1^2 = \sigma_2^2 = \sigma^2$

$$\text{Hence } z = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{\dfrac{\sigma^2}{n_1} + \dfrac{\sigma^2}{n_2}}} = \frac{\bar{x}_1 - \bar{x}_2}{\sigma \sqrt{\dfrac{1}{n_1^2} + \dfrac{1}{n_2^2}}}$$

is normally distributed with mean zero and standard deviation one.

If σ is not known we can use an estimate of σ^2 given by $\sigma^2 = \dfrac{n_1 S_1^2 + n_2 S_2^2}{n_1 + n_2}$

Rejection Rule for $H_0 : \mu_1 = \mu_2$

(*i*) If $|z| > 1.96$, reject H_0 at 5% level of significance.

(*ii*) If $|z| > 2.58$, reject H_0 at 1% level of significance.

(*iii*) If $|z| > 1.645$, reject H_0 at 10% level of significance.

(*iv*) If $z > 3$ then either the samples have not been drawn from the same population or the sampling is not simple.

Otherwise accept H_0.

Note : If the two samples are drawn from two populations with unknown Standard deviations σ_1^2 and σ_2^2, then σ_1^2 and σ_2^2 can be replaced by sample variances S_1^2 and S_2^2 provided both the samples n_1 and n_2 are large.

In this case, the test statistic is $z = \dfrac{(\bar{x}_1 - \bar{x}_2) - \delta}{\sqrt{\dfrac{S_1^2}{n_1} + \dfrac{S_2^2}{n_2}}}$

Write the formula for testing the hypothesis concerning "Two Means".

[JNTU Nov. 2008 (Set No.1)]

Solution : Let $\bar{x}_1$ be the mean of a random sample of size n_1 drawn from a population with mean μ_1 and variance σ_1^2. Let $\bar{x}_2$ be the mean of an independent random sample of size n_2 drawn from another population with mean μ_2 and variance σ_2^2. To test the hypothesis for difference of means, the statistic is given by

$$z = \frac{|\bar{x}_1 - \bar{x}_2|}{\sqrt{\dfrac{\sigma_1^2}{n_1} + \dfrac{\sigma_2^2}{n_2}}}$$

If $|z| < 1.96$, it is accepted that the samples have come from two populations with the same mean, at 5% level of significance. Otherwise, at this level of significance, we claim that the difference in means is significant.

SOLVED EXAMPLES

Example 1 : The means of two large samples of sizes 1000 and 2000 members are 67.5 inches and 68.0 inches respectively. Can the samples be regarded as drawn from the same population of S.D 2.5 inches. **[JNTU (A) Nov. 2010 (Set No. 1)]**

Solution : Let μ_1 and μ_2 be the means of the two populations.

Given $n_1 = 1000$, $n_2 = 2000$ and $\bar{x}_1 = 67.5$ inches, $\bar{x}_2 = 68$ inches

Population $S.D$, $\sigma = 2.5$ inches

1. **Null Hypothesis H_0 :** The samples have been drawn from the same population of $S.D$ 2.5 inches

   i.e., $\mu_1 = \mu_2$ and $\sigma = 2.5$ inches

2. **Alternative Hypothesis H_1 :** $\mu_1 \neq \mu_2$

3. **The test statistic is,** $z = \dfrac{\bar{x}_1 - \bar{x}_2}{\sqrt{\dfrac{\sigma^2}{n_1} + \dfrac{\sigma^2}{n_2}}} = \dfrac{67.5 - 68}{\sqrt{(2.5)^2\left(\dfrac{1}{1000} + \dfrac{1}{2000}\right)}}$

$\Rightarrow z = \dfrac{-0.5}{0.0968} = -5.16$

$\therefore |z| = 5.16 \geq 1.96$ *i.e.*, the calculated value of $z >$ the table value of z.

Hence the null hypothesis H_0 is rejected at 5% level of significance and we conclude that the samples are not drawn from the same population of $S.D.$ 2.5 inches.

Example 2 : The mean yield of wheat from a district A was 210 pounds with S.D. 10 pounds per acre from a sample of 100 plots. In another district the mean yield was 200 pounds with S.D. 12 pounds from a sample of 150 plots. Assuming that the S.D of yield in the entire state was 11 pounds, test whether there is any significant difference between the mean yield of crops in the two districts.

Solution : Let μ_1 and μ_2 be the means of the two populations.

Given $\bar{x}_1 = 210$, $\bar{x}_2 = 200$, and $n_1 = 100$, $n_2 = 150$,

Population S.D., $\sigma = 11$

1. **Null Hypothesis** $H_0 : \mu_1 = \mu_2$ *i.e.* there is no difference between μ_1 and μ_2
2. **Alternative Hypothesis** $H_1 : \mu_1 \neq \mu_2$
3. **The test statistic is** $z = \dfrac{\bar{x}_1 - \bar{x}_2}{\sqrt{\dfrac{\sigma^2}{n_1} + \dfrac{\sigma^2}{n_2}}} = \dfrac{210 - 200}{\sqrt{11^2\left(\dfrac{1}{100} + \dfrac{1}{150}\right)}} = 7.04178$

$\therefore \ |z| = 7.041 > 1.96$ *i.e.* the calculated value of $z >$ the table value of z.

Hence we reject the Null Hypothesis H_0 at 5% level of significance and conclude that there is a significant difference between the mean yield of crops in the two districts.

Example 3 : In a survey of buying habits, 400 women shoppers are chosen at random in super market 'A' located in a certain section of the city. Their average weekly food expenditure is ₹ 250 with a S.D. of ₹ 40. For 400 women shoppers chosen at random in super market 'B' in another section of the city, the average weekly food expenditure is ₹ 220 with a S.D. of ₹ 55. Test at 1% level of significance whether the average weekly food expenditure of the two populations of shoppers are equal.

Solution : Let μ_1 and μ_2 be the means of the two populations.

Given $n_1 = 400$, $\bar{x}_1 = ₹\,250$, $S_1 = ₹\,40$

$n_2 = 400$, $\bar{x}_2 = ₹\,220$, $S_2 = ₹\,55$

1. **Null Hypothesis** H_0 **:** Assume that the average weekly food expenditure of the two populations of shoppers are equal *i.e.*, $H_0 : \mu_1 = \mu_2$
2. **Alternative Hypothesis** $H_1 : \mu_1 \neq \mu_2$
3. **Test statistic is,** $z = \dfrac{\bar{x}_1 - \bar{x}_2}{\sqrt{\dfrac{S_1^2}{n_1} + \dfrac{S_2^2}{n_2}}} = \dfrac{250 - 220}{\sqrt{\dfrac{(40)^2}{400} + \dfrac{(55)^2}{400}}} \doteq \dfrac{30}{3.4} = 8.82$

i.e. $z = 8.82 > 2.58$

Hence we reject the Null Hypothesis H_0 at 1% level of significance and conclude that the average weekly food expenditure of the two populations of shoppers are not equal.

Example 4 : Samples of students were drawn from two universities and from their weights in kilograms, mean and standard deviations are calculated and shown below. Make a large sample test to test the significance of the difference between the means.

	Mean	S.D	Size of the sample
University A	55	10	400
University B	57	15	100

[JNTU 2005 (Set No. 3)]

Solution :

Given $\bar{x}_1 = 55$, $\bar{x}_2 = 57$, $n_1 = 400$, $n_2 = 100$, $S_1 = 10$ and $S_2 = 15$

1. **Null Hypothesis** $H_0 : \bar{x}_1 = \bar{x}_2$ *i.e.*, there is no difference
2. **Alternative Hypothesis** $H_1 : \bar{x}_1 \neq \bar{x}_2$
3. **Level of significance :** $\alpha = 0.05$ (assumed)
4. **Critical region :** Accept H_0 if $-1.96 < Z < 1.96$
5. **The test statistic** is $Z = \dfrac{\bar{x}_1 - \bar{x}_2}{\sqrt{\dfrac{S_1^2}{n_1} + \dfrac{S_2^2}{n_2}}} = \dfrac{55-57}{\sqrt{\dfrac{100}{400} + \dfrac{225}{100}}} = \dfrac{-2}{\sqrt{\dfrac{1}{4} + \dfrac{9}{4}}} = -1.26$

$\therefore |Z| = 1.26 < 1.96$

Hence, we accept the Null Hypothesis H_0 at 5% level of significance and conclude that there is no significant difference between the means.

Example 5 : The research investigator is interested in studying whether there is a significant difference in the salaries of MBA grades in two metropolitan cities. A random sample of size 100 from Mumbai yields on average income of Rs. 20,150. Another random sample of 60 from Chennai results in an average income of Rs. 20,250. If the variances of both the populations are given as σ_1^2 = Rs. 40,000 and σ_2^2 = Rs. 32,400 respectively.

[JNTU Nov. 2008 (Set No. 1)]

Solution : Given $n_1 = 100$, $\bar{x}_1 = 20{,}150$ and $n_2 = 60$, $\bar{x}_2 = 20250$

and $\sigma_1^2 = 40{,}000$, $\sigma_2^2 = 32400$

Let the Null Hypothesis be $H_0 : \mu_1 = \mu_2$. That is, the difference of means is significant.

Then the Alternative Hypothesis is $H_1 : \mu_1 \neq \mu_2$

The test statistic is $z = \dfrac{\bar{x}_1 - \bar{x}_2}{\sqrt{\dfrac{\sigma_1^2}{n_1} + \dfrac{\sigma_2^2}{n_2}}} = \dfrac{20150 - 20250}{\sqrt{\dfrac{40000}{100} + \dfrac{32400}{60}}}$

$= \dfrac{100}{\sqrt{400+540}} = \dfrac{100}{30.66} = 3.26$

Since $Z = 3.26 > 1.96 = Z_{0.05}$, we reject the Null Hypothesis at 5% level of significance and conclude that there is a significant difference in the salaries of MBA grades in two cities.

Example 6 : A researcher wants to know the intelligence of students in a school. He selected two groups of students. In the first group there 150 students having mean IQ of 75 with a S.D. of 15 in the second group there are 250 students having mean IQ of 70 with S.D of 20. Is there a significant difference between the means of two groups ?

[JNTU Nov. 2008 (Set No. 4), (H) Nov. 2015]

Solution : Given $n_1 = 150$, $\bar{x}_1 = 75$, $\sigma_1 = 15$

and $n_2 = 250$, $\bar{x}_2 = 70$, $\sigma_2 = 20$

1. **Null Hypothesis H_0 :** The groups have been came from the same population

   i.e., $\mu_1 = \mu_2$

2. **Alternative Hypothes** is $H_1 : \mu_1 \neq \mu_2$

3. **The test statistic is,** $z = \dfrac{\bar{x}_1 - \bar{x}_2}{\sqrt{\dfrac{\sigma_1^2}{n_1} + \dfrac{\sigma_2^2}{n_2}}}$

$$= \frac{75 - 70}{\sqrt{\dfrac{225}{150} + \dfrac{400}{250}}} = \frac{5}{\sqrt{\dfrac{9}{5} + \dfrac{8}{5}}} = \frac{5\sqrt{5}}{\sqrt{17}} = 2.7116$$

Tabulated value of z at 1% level of significance is 2.33

$\therefore$ Calculated $z >$ tabulated z.

Hence we reject the Null Hypothesis H_0 at 1% level of significance and conclude that the groups have not been taken from the same population.

Example 7 : Two types of new cars produced in U.S.A. are tested for petrol mileage, one sample is consisting of 42 cars averaged 15 kmpl while the other sample consisting of 80 cars averaged 11.5 kmpl with population variances as $\sigma_1^2 = 2.0$ and $\sigma_2^2 = 1.5$ respectively. Test whether there is any significance difference in the petrol consumption of these two types of cars. (use $\alpha = 0.01$).

[JNTU Apr. 2009 (Set No. 1)]

Solution : Let the types of the cars be named as A and B.

Number of cars of type A = 42

Average mileage for A = $\bar{x}_1 = 15$, Variance = $\sigma_1^2 = 2.0$

Number of cars of type B = 80

Average mileage for B = $\bar{x}_2 = 11.5$

Variance = $\sigma_2^2 = 1.5$

1. **Null Hypothesis H_0 :** $\mu_1 = \mu_2$
2. **Alternative Hypothes** is $H_1 : \mu_1 \neq \mu_2$
3. **The test statistic** is $z = \dfrac{|\bar{x}_1 - \bar{x}_2|}{\sqrt{\dfrac{\sigma_1^2}{n_1} + \dfrac{\sigma_2^2}{n_2}}} = \dfrac{|15 - 11.5|}{\sqrt{\dfrac{2}{42} + \dfrac{1.5}{80}}}$

$$= \frac{3.5}{\sqrt{0.0476+0.01875}} = \frac{3.5}{\sqrt{0.06635}} = 13.587$$

Tabulated value of z at 1% significance level is 2.58 (Two-tailed)

Since $(z)_{calculated} > 2.58$ (z table), we reject Null Hypothesis H_0 at 1% level of significance and conclude that there is a significant difference in petrol consumption.

Example 8 : A simple sample of the height of 6400 Englishmen has a mean of 67.85 inches and a S. D of 2.56 inches while a simple sample of heights of 1600 Australians has a mean of 68.55 inches and S. D. of 2.52 inches. Do the data indicate the Australians are on the average taller than the Englishmen? (Use α as 0.01). **[JNTU (A) 2009 (Set No. 2)]**

__Solution__ : We are given

n_1 = Size of the first sample = 6400

n_2 = Size of the second sample = 1600

$\bar{x}_1$ = Mean of the first sample = 67.85

$\bar{x}_2$ = Mean of the second sample = 68.55

σ_1 = Standard deviation of the first sample = 2.56

σ_2 = Standard deviation of the second sample = 2.52

1. **Null Hypothesis** $H_0 : \mu_1 = \mu_2$
2. **Alternative Hypothesis** $H_1 : \mu_1 < \mu_2$
3. **Level of significance :** $\alpha = 0.05$
4. **The test statistic** is

$$z = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{\left(\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}\right)}} = \frac{67.85-68.55}{\sqrt{\frac{(2.56)^2}{6400} + \frac{(2.52)^2}{1600}}} = \frac{-0.7}{\sqrt{\frac{6.5536}{6400} + \frac{6.35}{1600}}}$$

$$= \frac{-0.7}{\sqrt{0.001+0.004}} = \frac{-0.7}{0.0707} = -9.9$$

$\therefore \ |z| = 9.9 > 1.96$

Hence, we reject the Null Hypothesis H_0 at 5% level of significance and conclude that the Australians are taller than Englishmen.

Example 9 : The mean life of a sample of 10 electric bulbs (or motors) was found to be 1456 hours with S.D. of 423 hours. A second sample of 17 bulbs (motors) chosen from a different batch showed a mean life of 1280 hours with S.D. of 398 hours. Is there a significant difference between the means of two batches? **[JNTU (K) 2009, Nov. 2012 (Set No. 1)]**

__Solution :__ It is given that

n_1 = Sample size of first batch = 10

n_2 = Sample size of second batch = 17

$\bar{x}_1$ = Mean life of first batch = 1456

$\bar{x}_2$ = Mean life of second batch = 1280

σ_1 = Standard deviation of first batch = 423

σ_2 = Standard deviation of second batch = 398

1. **Null Hypothesis** $H_0 : \mu_1 = \mu_2$
2. **Alternative Hypothesis** $H_1 : \mu_1 \neq \mu_2$
3. **Level of significance :** $\alpha = 0.05$
4. **The test statistic is** $z = \dfrac{\bar{x}_1 - \bar{x}_2}{\sqrt{\dfrac{\sigma_1^2}{n_1} + \dfrac{\sigma_2^2}{n_2}}} = \dfrac{1456 - 1280}{\sqrt{\dfrac{(423)^2}{10} + \dfrac{(398)^2}{17}}}$

$$= \frac{176}{\sqrt{17892.9 + 9317.88}} = \frac{176}{164.96} = 1.067$$

Since $z < z_\alpha = 1.96$, we accept the null hypothesis H_0 *i.e.*, there is no difference between the mean life of electric bulbs of two batches.

Example 10 : The average marks scored by 32 boys is 72 with a S.D. of 8. While that for 36 girls is 70 with a S.D. of 6. Does this indicate that the boys perform better than girls at level of significance 0.05?

Solution : Let μ_1 and μ_2 be the means of the two populations

Let the Null Hypothesis be $H_0 : \mu_1 = \mu_2$

Then the Alternative Hypothesis is $H_1 : \mu_1 > \mu_2$

Let us assume that H_0 is true, *i.e.*, there is no difference between μ_1 and μ_2

Since the sample sizes are large we use the test statistic

$$z = \frac{\bar{x} - \bar{y}}{\sqrt{\dfrac{\sigma_1^2}{n_1} + \dfrac{\sigma_2^2}{n_2}}}$$

Here $\bar{x} = 72, \bar{y} = 70, \sigma_1 = 8, \sigma_2 = 6, n_1 = 32, n_2 = 36$

$$\therefore\ z = \frac{72 - 70}{\sqrt{\dfrac{64}{32} + \dfrac{36}{36}}} = \frac{2}{\sqrt{2+1}} = \frac{2}{\sqrt{3}} = 1.1547 < 1.96$$

Since the computed value of z is less than the table value, we cannot reject the Null Hypothesis at 5% level and conclude that the performance of boys and girls is the same.

Example 11 : At a certain large university a sociologist speculates that male students spend considerably more money on junk food than do female students. To test her hypothesis, the sociologist randomly selects from the registrar's records the names of 200 students. Of these, 125 are men and 75 are women. The sample mean of the average amount spent on junk food per week by the men is Rs. 400 and standard deviation is 100. For the women the sample mean is Rs. 450 and he sample standard deviation is Rs.150. Test the difference between the means at .05 level.

[JNTU (H) Nov. 2010 (Set No. 2)]

Solution : Let μ_1 and μ_2 be the means of the two populations.

Let the Null Hypothesis be $H_0 : \mu_1 = \mu_2$

Then the Alternative Hypothesis is $H_1 : \mu_1 \neq \mu_2$

Let us assume that H_0 is true *i.e.*, there is no difference between μ_1 and μ_2.

We are given

n_1 = Number of men = 125

n_2 = Number of women = 75

$\bar{x}_1$ = Mean of men = 400

$\bar{x}_2$ = Mean of women = 450

σ_1 = S.D. of men = 100

σ_2 = S.D. of women = 150

Level of significance, $\alpha = 0.05$

The test statistic is $z = \dfrac{\bar{x}_1 - \bar{x}_2}{\sqrt{\dfrac{\sigma_1^2}{n_1} + \dfrac{\sigma_1^2}{n_2}}} = \dfrac{400 - 450}{\sqrt{\dfrac{(100)^2}{125} + \dfrac{(150)^2}{75}}}$

$$= \frac{-50}{\sqrt{80 + 300}} = \frac{-50}{\sqrt{380}} = \frac{-50}{19.49} = -2.5654$$

$\therefore |z| = |-2.5654| = 2.5654 > 1.96$ *i.e.*, the difference is highly significant.

Hence, we reject the Null Hypothesis at 5% level of significance and conclude that the two population means are not equal.

Example 12 : A company claims that its bulbs are superior to those of its main competitor. If a study showed that a sample of 40 of its bulbs have a mean life time of 647 hrs of continuous use with a S.D of 27 hrs. While a sample of 40 bulbs made by its main competitor had a mean life time of 638 hrs of continuous use with a S.D of 31 hrs. Test the significance between the difference of two means at 5% level. (**OR** Does this substantiate the claim at 0.05 level of significance) **[JNTU (H) Nov. 2010 (Set No. 4), III yr. Nov. 2015]**

Solution : Let μ_1 and μ_2 be the means of the two populations.

Let the Null Hypothesis be $H_0 : \mu_1 = \mu_2$

Then the Alternative Hypothesis is $H_1 : \mu_1 > \mu_2$

Since the sample sizes are large, the test statistic

$$z = \frac{\bar{x} - \bar{y}}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}}$$

is approximately normally distributed with mean 0 and S.D. 1.

Here $\bar{x} = 647, \bar{y} = 638, \sigma_1 = 27, \sigma_2 = 31, n_1 = n_2 = 40$

$$\therefore z = \frac{647 - 638}{\sqrt{\frac{(27)^2}{40} + \frac{(31)^2}{40}}} = \frac{9}{\sqrt{\frac{729 + 961}{40}}} = \frac{9}{6.5} = 1.38$$

$\therefore z = 1.38 < 1.645$

Since the computed value of z is less than the table value, we cannot reject the Null Hypothesis at 5% level and conclude that the difference between the two sample means is not significant.

Example 13 : Studying the flow of traffic at two busy intersections between 4 p.m and 6 p.m to determine the possible need for turn signals. It was found that on 40 week days there were on the average 247.3 cars approaching the first intersection from the south which made left turn, while on 30 week days there were on the average 254.1 cars approaching the first intersection from the south made left turns. The corresponding sample standard deviations are 15.2 and 12. Test the significance between the difference of two means at 5% level.

[JNTU (H) Dec. 2011 (Set No.1)]

Solution : Let the average cars in two places be μ_1 and μ_2 respectively.

Let the Null Hypothesis be $H_0 : \mu_1 = \mu_2$

Then the Alternative Hypothesis is $H_1 : \mu_1 \neq \mu_2$

Let us assume that H_0 is true *i.e.*,there is no significant difference between μ_1 and μ_2.

Since the sample sizes are large, the test statistic is

$$z = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}}$$

Here $n_1 = 40, n_2 = 30, \bar{x}_1 = 247.3, \bar{x}_2 = 254.1, \sigma_1 = s_1 = 15.2$ and $\sigma_2 = s_2 = 12$.

$$\therefore z = \frac{247.3 - 254.1}{\sqrt{\frac{(15.2)^2}{40} + \frac{(12)^2}{30}}} = \frac{-6.8}{\sqrt{5.776 + 4.8}}$$

$$= \frac{-6.8}{\sqrt{10.576}} = \frac{-6.8}{3.2521} = -2.091$$

$$\therefore\ |z| = 2.091 > 1.96$$

Since the computed value of z is greater than the table value, we reject the Null Hypothesis at 5% level and conclude that the two average cars are significantly different *i.e.,* they are not the same in the two busy intersections.

Example 14 : In a certain factory there are two independent processes for manufacturing the same item. The average weight in a sample of 700 items produced from one process is found to be 250 gms with a standard deviation of 30 gms while the corresponding gures in a sample of 300 items from the other process are 300 and 40. Is there signicant dierence between the mean at 1% level.

[JNTU (H) Apr. 2012 (Set No.1)]

<u>**Solution**</u> : Let the average weight in the two independent processes be μ_1 and μ_2 respectively.

Let the Null Hypothesis be $H_0 : \mu_1 = \mu_2$

Then the Alternative Hypothesis is $H_1 : \mu_1 \neq \mu_2$

Let us assume that H_0 is true *i.e.,* there is no significant difference between μ_1 and μ_2.

The test statistic is $z = \dfrac{\bar{x}_1 - \bar{x}_2}{\sqrt{\dfrac{\sigma_1^2}{n_1} + \dfrac{\sigma_2^2}{n_2}}}$

Here $\bar{x}_1 = 250, \bar{x}_2 = 300, \sigma_1 = 30, \sigma_2 = 40$ and $n_1 = 700, n_2 = 300$

$$\text{Now } z = \frac{250-300}{\sqrt{\frac{900}{700} + \frac{1600}{300}}} = \frac{-50}{\sqrt{\frac{9}{7} + \frac{16}{3}}} = -19.43$$

1.92.

Thus $|z| = 19.43 > 2.58$

Since the computed value of z is greater than the table value, we reject the Null Hypothesis at 1% level and conclude that there is a significant difference between the means.

Example 15 : The mean height of 50 male students who participated in sports is 68.2 inches with a S.D of 2.5. The mean height of 50 male students who have not participated in sport is 67.2 inches with a S.D of 2.8. Test the hypothesis that the height of students who participated in sports is more than the students who have not participated in sports.

[JNTU (H) Apr. 2012 (Set No. 2)]

<u>**Solution**</u> : Let the mean height in the two cases be μ_1 and μ_2 respectively.

Let the **Null Hypothesis** be $H_0 : \mu_1 = \mu_2$

Then the **Alternative Hypothesis** is $H_1 : \mu_1 \neq \mu_2$

Let us assume that H_0 is true, *i.e.*, there is no significant difference between μ_1 and μ_2

The test statistic is $z = \dfrac{\bar{x}_1 - \bar{x}_2}{\sqrt{\dfrac{\sigma_1^2}{n_1} + \dfrac{\sigma_2^2}{n_2}}}$

Here $\bar{x}_1 = 68.2, \bar{x}_2 = 67.2, \sigma_1 = 2.5, \sigma_2 = 2.8, n_1 = 50, n_2 = 50$

$$\text{Now } z = \frac{68.2 - 67.2}{\sqrt{\dfrac{(2.5)^2}{50} + \dfrac{(2.8)^2}{50}}} = \frac{1}{\sqrt{\dfrac{6.25 + 7.84}{50}}} = 1.88$$

$\therefore |z| = 1.88 < 1.96$

Since the computed value of z is less than the tabulated value of z, we accept the Null Hypothesis at 5% level of significance and conclude that there is no significant difference in the heights.

Example 16 : The nicotent in milligrams of two samples of tobacco were found to be as follows. Find the standard error and confidential limits for the difference between the means at 0.05 level.

Sample A	24	27	26	23	25	
Sample B	29	30	30	31	24	36

[JNTU (H) Apr. 2012 (Set No. 4)]

Solution :

Calculations for means and s_1^2, s_2^2					
Sample A			Sample B		
x_1	$x_1 - \bar{x}_1 = x_1 - 25$	$(x_1 - \bar{x}_1)^2$	x_2	$x_2 - \bar{x}_2 = x_2 - 30$	$(x_2 - \bar{x}_2)^2$
24	−1	1	29	−1	1
27	2	4	30	0	0
26	1	1	30	0	0
23	−2	4	31	1	1
25	0	0	24	−6	36
			36	6	36
$125 = \sum x_1$		10	$180 = \sum x_2$		74

$$\bar{x}_1 = \frac{\sum x_1}{n_1} = \frac{125}{5} = 25, \ \bar{x}_2 = \frac{\sum x_2}{n_2} = \frac{180}{6} = 30$$

Hence variance, $s_1^2 = \frac{\sum(x_1 - \bar{x}_1)^2}{n_1 - 1} = \frac{10}{4} = 2.5$

and variance, $s_2^2 = \frac{\sum(x_2 - \bar{x}_2)^2}{n_2 - 1} = \frac{74}{5} = 14.8$

$$\therefore \text{S. E. of } (\bar{x}_1 - \bar{x}_2) = \sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}} = \sqrt{\frac{25}{5} + \frac{14.8}{6}} = 1.72$$

Hence the 95% confidence limits are $(\bar{x}_1 - \bar{x}_2) \pm 1.96$ (S. E. of $\bar{x}_1 - \bar{x}_2$)

i.e., $(25 - 30) \pm 1.96(1.72)$ *i.e.*, -5 ± 3.37 or $(-8.37, -1.63)$

REVIEW QUESTIONS

1. (*a*) Define statistical Hypothesis

   (*b*) Define (*i*) Test of statistical hypothesis

   (*ii*) Type - I and Type II error

2. (*a*) Write about (*i*) Null hypothesis (*ii*) Type I and Type II errors (*iii*) Alternative hypothesis. **[JNTU 2008S]**

   (OR) Explain briefly the following : (*i*) Type I error (*ii*) Type II error (*iii*) Critical Region

   (*b*) Explain the types of errors in sampling.

3. What is meant by level of significance. **[JNTU (H) Nov. 2009 (Set No. 3)]**

4. (*a*) Write about

   (*i*) Critical region (*ii*) Left tailed test (*iii*) Right tailed test

   (*iv*) Two tailed test. (OR) Write about one tailed and two tailed tests. **[JNTU 2004, (H) May 2012 (Set No. 1)]**

   (*b*) What is meant by Level of significance, one tailed and two tailed tests ?

   (*c*) Write about (*i*) Null hypothesis (*ii*) Critical Region

   (*iii*) Level of Significance

5. Write about null hypothesis and testing of null hypothesis. **[JNTU 2005 (Set No. 4)]**

6. Explain the procedure generally followed in testing of hypothesis. **[JNTU (H) Nov. 2009 (Set No. 2,4)]**

   (OR) Explain working rule for testing of Hypothesis.

7. Write shor notes on type I and type II Error. **[JNTU (H) Nov. 2009 (Set No. 2)]** **[JNTU (A) Nov. 2010 (Set No. 4)]**

   (OR) Explain the types of errors in sampling

   (OR) Explain type I and type II of errors in testing of hypothesis

8. Differentiate two-tailed test of hypothesis from one-tailed test.

8.9 TEST OF SIGNIFICANCE FOR SINGLE PROPORTION - LARGE SAMPLES

[JNTU (H) Nov. 2009 (Set-4)]

Suppose a large random sample of size n has a sample proportion p of members possessing a certian attribute (*i.e.*, proportion of successes). To test the hypothesis that the proportion P in the population has a specified value P_0.

Let us set the Null Hypothesis be $H_0 : P = P_0$ (P_0 is a particular value of P)

Then the Alternative Hypothesis is

(*i*) $H_1 : P \neq P_0$ (*i.e.*, $P > P_0$ or $P < P_0$)

(*ii*) $H_1 : P > P_0$ or

(*iii*) $H_1 : P < P_0$

Since n is large, the sampling distribution of p is approximately normal.

$\therefore$ If H_0 is true, the test statistic

$$z = \frac{p - P_0}{\text{S.E. of } p}$$

or $z = \dfrac{p - P}{\sqrt{\dfrac{PQ}{n}}}$, where p is the sample proportion

is approximately normally distributed.

The critical region for z depending on the nature of H_1 and level of significance α is given in the following table :

Table : Rejection Rule for $H_0 : P = P_0$

Level of siginificance	1%	5%	10%
Critical region for $P \neq P_0$ (Two-tailed test)	$\lvert z \rvert > 2.58$	$\lvert z \rvert > 1.96$	$\lvert z \rvert > 1.645$
Critical region for $P > P_0$ (Right-tailed test)	$z > 2.33$	$z > 1.645$	$z > 1.28$
Critical region for $P < P_0$ (Left-tailed test)	$z < -2.33$	$z < -1.645$	$z < -1.28$

Note 1 : Without any reference to the level of significance, we may reject the Null Hypothesis H_0 when $|z| > 3$.

Note 2 : (*i*) Limits for population proportion P are given by $p \pm 3\sqrt{\frac{pq}{n}}$ where $q = 1 - p$.

(*ii*) Confidence interval for proportion P for large sample at α level of significance is

$$P - z_{\alpha/2} \cdot \sqrt{\frac{PQ}{n}} < P < P + z_{\alpha/2} \cdot \sqrt{\frac{PQ}{n}} \text{ where } Q = 1 - P \text{ and}$$

$z_{\alpha/2} = 1.96$ (for 95%), $z_{\alpha/2} = 2.33$ (for 98%) and

$z_{\alpha/2} = 2.58$ (for 99%).

SOLVED EXAMPLES

Example 1 : A manufacturer claimed that atleast 95% of the equipment which he supplied to a factory conformed to specifications. An examination of a sample of 200 pieces of equipment revealed that 18 were faulty. Test his claim at 5% level of significance.

[JNTU (K), (H) Nov. 2009, (H) Dec. 2011 (Set No. 1)]

Solution ; Given sample size, $n = 200$

Number of pieces confirming to specification = 200 − 18 = 182

$\therefore \quad p$ = Proportion of pieces confirming to specifications

$$= \frac{182}{200} = 0.91$$

$$P = \text{Population proportion} = \frac{95}{100} = 0.95$$

1. **Null Hypothesis H_0 :** The proportion of pieces confirming to specifications *i.e.* $P = 95\%$.
2. **Alternative Hypothesis H_1 :** $P < 0.95$ (left - tail test)
3. **The test statistic** is $z = \dfrac{p-P}{\sqrt{\dfrac{PQ}{n}}} = \dfrac{0.91-0.95}{\sqrt{\dfrac{0.95\times 0.05}{200}}} = \dfrac{-0.04}{0.0154} = -2.59.$

Since alternative hypothesis is left tailed, the tabulated value of Z at 5% level of significance is 1.645.

Since calculated value of $|z| = 2.6$ is greater than 1.645, we reject the Null Hypothesis H_0 at 5% level of significance and conclude that the manufacture's claim is rejected.

Example 2 : In a sample of 1000 people in Karnataka 540 are rice eaters and the rest are wheat eaters. Can we assume that both rice and wheat are equally popular in this state at 1% level of significance ? **[JNTU (K), (H) Nov. 2009, (A) Nov. 2010 (Set No. 3), (H) Sept. 2017]**

Solution : Given

$n = 1000$

$$p = \text{Sample proportion of rice eaters} = \frac{540}{1000} = 0.54$$

$$P = \text{Population proportion of rice eaters} = \frac{1}{2} = 0.5$$

$\therefore\ Q = 0.5$

Null Hypothesis H_0 : Both rice and wheat are equally popular in the state.

Alternative Hypothesis H_1 : $P \neq 0.5$ (two - tailed alternative)

$$\textbf{Test statistic is } z = \frac{p-P}{\sqrt{\frac{PQ}{n}}} = \frac{0.54-0.5}{\sqrt{\frac{0.5\times 0.5}{1000}}} = 2.532$$

The calculated value of $z = 2.532$

The tabulated value of z at 1% level of significance for two-tailed test is 2.58.

Since calculated $z <$ tabulated z, we accept the Null Hypothesis H_0 at 1% level of significance and conclude that both rice and wheat are equally popular in the state.

Example 3 : In a big city 325 men out of 600 men were found to be smokers. Does this information support the conclusion that the majority of men in this city are smokers ?

Solution : Given $n = 600$

Number of smokers = 325

$$p = \text{Sample proportion of smokers} = \frac{325}{600} = 0.5417$$

$$P = \text{Population proportion of smokers in the city} = \frac{1}{2} = 0.5$$

$Q = 1 - P = 1 - 0.5 = 0.5$

1. **Null Hypothesis H_0 :** The number of smokers and non-smokers are equal in the city.
2. **Alternative Hypothesis :** $P > 0.5$ (Right tailed)
3. **The Test statistic** is $z = \dfrac{p-P}{\sqrt{\frac{PQ}{n}}} = \dfrac{0.5417-0.5}{\sqrt{\frac{0.5\times 0.5}{600}}} = 2.04$

$\therefore$ Calculated value of $Z = 2.04$

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

Since, calculated value of $z >$ tabulated value of z, we reject the Null Hypothesis and conclude that the majority of men in the city are smokers.

Example 4 : A die was thrown 9000 times and of these 3220 yielded a 3 or 4. Is this consistent with the hypothesis that the die was unbiased? **[JNTU (K) Nov. 2009 (Set No. 3)]**

Solution : Given $n = 9000$

p = Proportion of successes of getting 3 or 4 in 9000 throws $= \frac{3220}{9000} = 0.3578$

P = Population proportion of successes

$= P(\text{getting a 3 or 4}) = \frac{1}{6} + \frac{1}{6} = \frac{2}{6} = \frac{1}{3} = 0.3333$

$\therefore\ Q = 1 - P = 1 - \frac{1}{3} = \frac{2}{3} = 0.6667$

1. **Null Hypothesis H_0 :** The die is unbiased.
2. **Alternative Hypothesis $H_1 : P \neq \frac{1}{3}$** (two-tailed)
3. **The test statistic** is $z = \dfrac{p-P}{\sqrt{\frac{PQ}{n}}} = \dfrac{0.3578 - 0.3333}{\sqrt{\frac{0.3333 \times 0.6667}{9000}}} = 4.94$

Calculated $z = 4.94$

As $z > 2.58$, the null hypothesis H_0 has to be rejected at 1% level of significance and we conclude that the die is biased.

Example 5 : A random sample of 500 pineapples was taken from a large consignment and 65 were found to be bad. Find the percentage of bad pineapples in the consignment.

Solution : Given

$n = 500$

p = Proportion of bad pineapples in the sample $= \frac{65}{500} = 0.13$

$q = 1 - p = 0.87$

We know that the limits for population proportion P are given by

$$p \pm 3\sqrt{\frac{pq}{n}} = 0.13 \pm 3\sqrt{\frac{0.13 \times 0.87}{500}} = 0.13 \pm 0.045 = (0.085,\ 0.175)$$

$\therefore$ The percentage of bad pineapples in the consignment lies between 8.5 and 17.5.

Example 6 : In a random sample of 125 cool drinkers, 68 said they prefer thumsup to pepsi. Test the null hypothesis P = 0.5 against the alternative hypothesis P > 0.5. **[JNTU 2006S (Set No. 2)]**

Solution : We have $n = 125$, $x = 68$ and $p = \frac{x}{n} = \frac{68}{125} = 0.544$

1. **Null Hypothesis H_0 :** $P = 0.5$
2. **Alternative Hypothesis H_1 :** $P > 0.5$

3. **Level of significance :** $\alpha = 0.05$ (assumed)

4. **The test statistic is** $z = \dfrac{p-P}{\sqrt{\dfrac{PQ}{n}}} = \dfrac{0.544-0.5}{\sqrt{\dfrac{(0.5)(0.5)}{125}}}$

$$= \frac{0.044}{0.045} = 0.9839$$

Since calculated value of $|Z|$ is less than 1.645, we accept the Null Hypothesis H_0 at 5% level of significance.

Example 7 : Experience had shown that 20% of a manufactured product is of the top quality. In one day's production of 400 articles only 50 are of top quality. Test the hypothesis at 0.05 level. **[JNTU 2005 S, 2008 S (Set No.2)]**

Solution : We have $n = 400$, $x = 50$ and $p = \dfrac{x}{n} = \dfrac{50}{400} = 0.125$

1. **Null Hypothesis** $H_0 : P = 0.2$
2. **Alternative Hypothesis** $H_1 : P \neq 0.2$
3. **Level of significance :** $\alpha = 0.05$
4. **The test statistic is** $z = \dfrac{p-P}{\sqrt{\dfrac{PQ}{n}}}$

i.e., $z = \dfrac{0.125-0.2}{\sqrt{\dfrac{(0.2\,(0.8)}{400}}} = \dfrac{-0.075}{0.02} = -3.75$

Since $|z| = 3.75 > 1.96$, we reject the Null Hypothesis H_0 at 5% level of significance.

i.e., $P = 20\%$ is not correct.

Example 8 : A social worker believes that fewer than 25% of the couples in a certain area have ever used any form of birth control. A random sample of 120 couples was contacted. Twenty of them said that they have used. Test the belief of the social worker at 0.05 level.

Solution : We have **[JNTU 2005 (Set No.1)]**

$$n = 120,\ x = 20,\ p = \frac{x}{n} = \frac{20}{120} = \frac{1}{6} \text{ and } P = 0.25,\ Q = 1 - P = 0.75$$

1. **Null Hypothesis** $H_0 : P = 0.25$
2. **Alternative Hypothesis** $H_1 : P < 0.25$ (left tailed test)
3. **Level of significance,** $\alpha = 0.05$
4. **The test statistic is**

$$z = \frac{p-P}{\sqrt{\frac{PQ}{n}}} = \frac{\frac{1}{6}-0.25}{\sqrt{\frac{(0.25(0.75)}{120}}} = \frac{-0.0833}{0.0395} = -2.107$$

Since $|z| = 2.107 < 2.33 = z_{0.05}$, we accept the null hypothesis H_0. That is, the claim or belief of social worker is true.

Example 9 : A manufacturer claims that only 4% of his products are defective. A random sample of 500 were taken among which 100 were defective. Test the hypothesis at 0.05 level. **[JNTU 2005 (Set No. 2)]**

Solution : We have

$x = 100$, $n = 500$, $p = \frac{x}{n} = 0.2$ and $P = 0.04$, $Q = 1 - P = 0.96$

1. **Null Hypothesis H_0 :** $P = 0.04$
2. **Alternative Hypothesis H_1 :** $P > 0.04$ (Right tailed test)
3. **Level of significance :** $\alpha = 0.05$
4. **The test statistic** is

$$z = \frac{p-P}{\sqrt{\frac{PQ}{n}}} = \frac{0.2-0.04}{\sqrt{\frac{(0.04)(0.96)}{500}}} = \frac{-0.16}{0.00876} = -18.26$$

Since $|z| = 18.26 > 1.645 = z_{0.05}$, we reject the null hypothesis H_0.

Example 10 : In a sample of 500 from a village in Rajasthan, 280 are found to be wheat eaters and the rest rice eaters. Can we assume that the both articles are equally popular. **[JNTU (K) 2009, May 2012 (Set No. 4)]**

Solution : Given $n = 500$

p = sample proportion of wheat eaters $= \frac{280}{500} = 0.56$

P = population proportion of wheat eaters $= \frac{1}{2} = 0.5$

$\therefore Q = 1 - P = 1 - 0.5 = 0.5$.

1. **Null Hypothesis** H_0 : Both rice and wheat are equally popular in Rajasthan.
2. **Alternative Hypothesis** $H_1 : P \neq 0.5$
3. **Level of significance :** $\alpha = 0.01$
4. **The test statistic** is $z = \frac{p-P}{\sqrt{PQ/n}} = \frac{0.56-0.5}{\sqrt{(0.5)(0.5)/500}} = \frac{0.06}{0.022} = 2.7272$

$\therefore$ The calculated value of $z = 2.7272$

The tabulated value of z at 1% level of significance for two-tailed test is 2.58.

Since calculated value of $z >$ tabulated z, we reject the null hypothesis H_0. *i.e.*, Both rice and wheat are not equally popular in Rajasthan at 1% level of significance.

Example 11 : Among 900 people in a state 90 are found to be chapati eaters. Construct 99% confidence interval for the true proportion. **[JNTU 2005 S, (K) May 2010S (Set No. 2)]**

Solution : Given $x = 90,\ n = 900$

$$\therefore\ P = \frac{x}{n} = \frac{90}{900} = \frac{1}{10} = 0.1 \text{ and } Q = 1 - P = 0.9$$

Now $\sqrt{\frac{PQ}{n}} = \sqrt{\frac{(0.1)(0.9)}{900}} = 0.01$

Confidence interval is

$$\left(P - z_{\alpha/2}\cdot\sqrt{\frac{PQ}{n}},\ P + z_{\alpha/2}\cdot\sqrt{\frac{PQ}{n}}\right) \text{ or } \left(P - 3\sqrt{\frac{PQ}{n}},\ P + 3.\sqrt{\frac{PQ}{n}}\right)$$

i.e., $(0.1 - 0.03,\ 0.1 + 0.03)$

i.e., $(0.07,\ 0.13)$.

Example 12 : In a random sample of 160 workers exposed to a certain amount of radiation, 24 experienced some ill effects. Construct a 99% confidence interval for the corresponding true percentage. **[JNTU 2004, 2006S (Set No. 1)]**

Solution : We have $x = 24,\ n = 160$ and $P = \frac{24}{160} = 0.15,\ Q = 1 - P = 0.85$

Now $\sqrt{\frac{PQ}{n}} = \sqrt{\frac{0.15 \times 0.85}{160}} = 0.028$

Confidence interval at 99% level of significance is

$$\left(P - 3\sqrt{\frac{PQ}{n}},\ P + 3\sqrt{\frac{PQ}{n}}\right)$$

i.e., $(0.15 - 3 \times 0.028,\ 0.15 + 3 \times 0.028)$

i.e., $(0.066,\ 0.234)$.

Example 13 : In a random sample of 400 industrial accidents, it was found that 231 were due atleast partially to unsafe working conditions. Construct a 99% confidence interval for the corresponding true proportion. **[JNTU 2006S, (K) May 2010S (Set No. 2)]**

Solution : We have $x = 231,\ n = 400$ and $P = \frac{x}{n} = \frac{231}{400} = 0.5775$

$Q = 1 - P = 1 - 0.577 = 0.4225$

Now $\sqrt{\frac{PQ}{n}} = \sqrt{\frac{0.5775 \times 0.4225}{400}} = 0.0247$

Confidence interval is

$$\left(P-3.\sqrt{\frac{PQ}{n}},\ P+3.\sqrt{\frac{PQ}{n}}\right)$$

i.e., $(0.5775 - 3 \times 0.0247,\ 0.5775 + 3 \times 0.0247)$

i.e., $(0.5034,\ 0.6516)$.

Example 14 : 20 people were attacked by a disease and only 18 survived. Will you reject the hypothesis that the survival rate if attacked by this disease is 85% in favour of the hypothesis that is more at 5% level. **[JNTU 2008S, (K) May 2010S, (H) April 2012 (Set No. 1)]**

Solution :

n = Sample size = 20

X = Number of survived people = 18

p = Proportion of survived people $= \frac{X}{n} = \frac{18}{20} = 0.9$

$P = 0.85$

$\therefore Q = 1 - P = 1 - 0.85 = 0.15$

1. **Null Hypothesis H_0 : $P = 0.85$**
2. **Alternative Hypothesis H_1 : $P > 0.85$** (Right tailed test)
3. **Level of significance : $\alpha = 0.5$**
4. **The test statistic** is

$$z = \frac{p-P}{\sqrt{\frac{PQ}{n}}} = \frac{0.9-0.85}{\sqrt{\frac{0.85\times0.15}{20}}} = \frac{0.05}{0.08} = 0.625$$

$\therefore$ Calculated $z = 0.625$

Tabulated z at 5% level of significance, $z_\alpha = 1.645$

Since calculated $z <$ tabulated z, we accept the Null Hypothesis H_0

i.e., The proportion of the survived people is 0.85.

Example 15 : If 80 patients are treated with an antibiotic 59 got cured. Find a 99% confidence limits to the true population of cure. **[JNTU 2004 S (Set No. 2)]**

Solution : $n = 80$, $x = 59$ and $P = \frac{x}{n} = \frac{59}{80} = 0.7375$

$\therefore Q = 1 - P = 1 - 0.7375 = 0.2625$

Now $\sqrt{\frac{PQ}{n}} = \sqrt{\frac{0.7375\times0.2625}{80}} = 0.049$

Confidence interval is $\left(P - z_{\alpha/2} \cdot \sqrt{\frac{PQ}{n}}, P + z_{\alpha/2} \cdot \sqrt{\frac{PQ}{n}}\right)$ where $z_{\alpha/2} = 2.58$

or $\left(P - 3.\sqrt{\frac{PQ}{n}}, P + 3.\sqrt{\frac{PQ}{n}}\right)$

i.e., $(0.7375 - 3 \times 0.049, 0.7375 + 3 \times 0.049)$ *i.e.*, $(0.59, 0.88)$

Example 16 : In a random sample of 100 packages shipped by air freight 13 had some damage. Construct 95% confidence interval for the true proportion of damage package.

[JNTU (H) Nov. 2010 (Set No. 2)]

Solution : Here p = Sample proportion of damage packages $= \frac{13}{100} = 0.13$

$\therefore q = 1 - p = 1 - 0.13 = 0.87$

S.E. of $p = \sqrt{\frac{PQ}{n}} = \sqrt{\frac{pq}{n}}$ ($\because$ P is not known, we take p for P)

$$= \sqrt{\frac{0.13 \times 0.87}{100}} = 0.034$$

$\therefore$ 95% confidence limits for the population proportion of P of damage packages are

$p \pm 1.96\,(\text{S.E. of } p) = 0.13 \pm 1.96\,(0.034) = 0.13 \pm 0.067$

i.e., the confidence limits are 0.063 and 0.197

Hence the 95% confidence interval for the true proportion of damage packages is $(0.063, 0.197)$.

Example 17 : In a hospital 480 females and 520 male babies were born in a week. Do these figures confirm the hypothesis that males and females are born in equal number?

Solution : Null Hypothesis H_0 : The probability of equal proportion *i.e.* $P = 0.5$

Alternative Hypothesis H_1 : The proportion is not equal *i.e.*, $P \neq 0.5$

n = Total number of births $= 480 + 520 = 1000$

p = proportion of females born $= \frac{480}{1000} = 0.48$

Population proportion $= 0.5 = P, Q = 1 - P = 1 - 0.5 = 0.5$

The test statistic is

$$z = \frac{p - P}{\sqrt{PQ/n}} = \frac{0.48 - 0.5}{\sqrt{\frac{0.5 \times 0.5}{1000}}} = \frac{-0.02}{0.0158} = -1.265$$

$\therefore |z| = 1.265$. We use two-tailed test.

Since $|z| = 1.265 < 1.96$, we accept the Null Hypothesis H_0 at 5% level of significance and conclude that the males and females are born in equal proportions.

Example 18 : A manufacturer claimed that at least 95% of the equipment which he supplied to a factory confomed to specifications. An examination of a sample of 200 pieces of equipments received 80 were faulty. Test the claim at 0.05 level.

[JNTU (H) Nov. 2010 (Set No. 3)]

Solution : We have $P = 95\% = 0.95, n = 200$

$\therefore Q = 1 - P = 1 - 0.95 = 0.05$

$p =$ observed proportion of faulty pieces $= \dfrac{80}{200} = 0.4$

Let the Null Hypothesis be $H_0 : P = 0.95$

Then the Alternative Hypothesis is $H_1 : P < 0.95$

Since H_1 is one-tailed (left-tailed), we use one-tailed (left-tailed) test.

Let us assume that H_0 is true.

The test statistic is $z = \dfrac{p - P}{\sqrt{\dfrac{PQ}{n}}} = \dfrac{0.4 - 0.95}{\sqrt{\dfrac{0.95 \times 0.05}{200}}}$

$$= \frac{-0.55}{0.0154} = -0.0085$$

Thus we see that $z = -0.0085 > -1.645$

Hence, we accept the Null Hypothesis H_0 at 5% level of significance and conclude that the claim of the manufacturer is justified.

Example 19 : A random sample of 500 Apples was taken from a large consignment of 60 were found to be bad, obtain the 98% confidence limits for the percentage number of bad apples in the consignment.

[JNTU (H) Dec. 2011 (Set No. 2)]

Solution : We are given

$n =$ Sample size $= 500$

$x =$ Number of bad apples $= 60$

$\therefore p =$ Proportion of bad apples $= \dfrac{x}{n} = \dfrac{60}{500} = 0.12$

Hence $q = 1 - p = 1 - 0.12 = 0.88$

$\therefore$ Standard error of $p = \sqrt{\dfrac{pq}{n}} = \sqrt{\dfrac{0.12 \times 0.88}{500}} = 0.014$

The 98% confidence limits for p are

$$\left(p - z_{\alpha/2}\cdot\sqrt{\frac{pq}{n}},\, p + z_{\alpha/2}\cdot\sqrt{\frac{pq}{n}}\right) \text{ where } z_{\alpha/2} = 2.33$$

$$= (0.12 - 2.33 \times 0.014, 0.12 + 2.33 \times 0.014)$$

$$= (0.087, 0.153)$$

Thus the confidence limits for percentage of bad apples

$$= (0.087 \times 100, 0.153 \times 100) = (8.7, 15.3)$$

Example 20 : In a study designed to investigate whether certain detonators used with explosives in coal mining meet the requirement that at least 90% will ignite the explosive when charged. It is found that 174 of 200 detonators function properly. Test the null hypothesis P = 0.9 against the alternative hypothesis P 0.9 at the 0.05 level of signicance.

[JNTU (H) Apr. 2012 (Set No. 2)]

<u>**Solution**</u> : Let the Null Hypothesis be $H_0 : P = 0.9$

Then the Alternative Hypothesis is $H_1 : P \neq 0.9$

Since H_1 is two - tailed, we use two - tailed test.

Let us assume that H_0 is true.

The test statistic is $z = \dfrac{p - P}{\sqrt{\dfrac{PQ}{n}}}$

If p is the observed proportion of detonators functioning properly, then $p = \dfrac{174}{200} = 0.87$

$Q = 1 - P = 1 - 0.9 = 0.1$ and $n = 200$

$$\therefore\ z = \frac{0.87 - 0.9}{\sqrt{\dfrac{0.9 \times 0.1}{200}}} = \frac{-0.03}{\sqrt{\dfrac{0.09}{200}}} = -1.41$$

$$\therefore\ |z| = 1.41 < 1.645$$

Hence, we accept the Null Hypothesis at 5% level of significance and conclude that there is no significant evidence to say that the given kind of detonators fails to meet the required standard.

8.10 TEST FOR EQUALITY OF TWO PROPORTIONS (OR TEST OF SIGNIFICANCE OF DIFFERENCE BETWEEN TWO SAMPLE PROPORTIONS - LARGE SAMPLES)

Let p_1 and p_2 be the sample proportions in two large random samples of sizes n_1 and n_2 drawn from two populations having proportions P_1 and P_2.

To test whether the two samples have been drawn from the same population.

1. **The Null Hypothesis** $H_0 : P_1 = P_2$
2. **The Alternative Hypothesis** $H_1 : P_1 \neq P_2$
3. **The Test Statistic :** There are two ways of computing a test statistic z.

(a) **When the population proportions P_1 and P_2 are known.**

In this case, $Q_1 = 1 - P_1$ and $Q_2 = 1 - P_2$ and p_1, p_2 are sample proportions.

$\therefore$ Standard Error of Difference = S. E. $(p_1 - p_2) = \sqrt{\dfrac{P_1Q_1}{n_1} + \dfrac{P_2Q_2}{n_2}}$

Hence the test statistic is

$$z = \frac{p_1 - p_2}{\text{S.E.}(p_1 - p_2)} = \frac{p_1 - p_2}{\sqrt{\dfrac{P_1Q_1}{n_1} + \dfrac{P_2Q_2}{n_2}}}$$

(b) **When the population proportions P_1 and P_2 are not known but sample proportions p_1 and p_2 are known.**

In this case we have two methods to estimate P_1 and P_2.

(i) **Method of Substitution :** In this method, sample proportions p_1 and p_2 are substituted for P_1 and P_2.

$$\therefore \text{S. E.}(p_1 - p_2) = \sqrt{\frac{p_1q_1}{n_1} + \frac{p_2q_2}{n_2}}$$

Hence the test statistic is

$$z = \frac{p - p_2}{\text{S. E.}(p_1 - p_2)} = \frac{p_1 - p_2}{\sqrt{\dfrac{p_1q_1}{n_1} + \dfrac{p_2q_2}{n_2}}}$$

(ii) **Method of Pooling :** In this method, the estimated value for the two population proportions is obtained by pooling the two sample proportions p_1 and p_2 into a single proportion p by the formula given below.

Sample proportion of two samples or estimated value of p is given by

$$p = \frac{n_1p_1 + n_2p_2}{n_1 + n_2}, \text{ so that } q = 1 - p$$

$$\therefore \text{S. E.}(p_1 - p_2) = \sqrt{\frac{pq}{n_1} + \frac{pq}{n_2}} = \sqrt{pq\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}$$

Hence the test statistic is

$$z = \frac{p_1 - p_2}{S.E.(p_1 - p_2)} = \frac{p_1 - p_2}{\sqrt{pq\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}}$$

Rejection Rule for $H_0 : P_1 = P_2$.

(*i*) If $|z| > 1.96$, then reject H_0 at 5% level of significance.

(*ii*) If $|z| > 2.58$, then reject H_0 at 1% level of significance.

(*iii*) If $|z| > 1.645$, then reject H_0 at 10% level of significance.

SOLVED EXAMPLES

Example 1 : Random samples of 400 men and 600 women were asked whether they would like to have a flyover near their residence. 200 men and 325 women were in favour of the proposal. Test the hypothesis that proportions of men and women in favour of the proposal are same, at 5% level. **[JNTU (H) May 2011 (Set No.2)]**

Solution : Given sample sizes, $n_1 = 400$, $n_2 = 600$

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

Proportion of women, $p_2 = \frac{325}{600} = 0.5416$

1. **Null Hypothesis H_0 :** Assume that there is no significant difference between the option of men and women as far as proposal of flyover is concerned.
   i.e., $H_0 : p_1 = p_2 = p$
2. **Alternative Hypothesis $H_1 : p_1 \neq p_2$** (two tailed)
3. **The test statistic** is $z = \frac{p_1 - p_2}{\sqrt{pq\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}}$ [Method of pooling]

where $p = \frac{n_1 p_1 + n_2 p_2}{n_1 + n_2}$

$$= \frac{400 \times \frac{200}{400} + 600 \times \frac{325}{600}}{400 + 600} = \frac{525}{1000} = 0.525$$

and $q = 1 - p = 1 - 0.525 = 0.475$

$$\therefore \quad z = \frac{0.5 - 0.5416}{\sqrt{0.525 \times 0.475\left(\frac{1}{400} + \frac{1}{600}\right)}} = \frac{-0.0416}{0.032} = -1.3$$

Thus $|z| = 1.3$

Since $|z| < 1.96$, we accept the null hypothesis H_0 at 5% level of significance.

i.e., there is no difference of opinion between men and women as far as proposal of flyover is concerned.

Example 2 : Random samples of $\overset{n_1}{400}$ men and $\overset{n_2}{200}$ women in a locality were asked whether they would like to have a bus stop near their residence. $\overset{x_1}{200}$ men and $\overset{x_2}{40}$ women in favour of the proposal. Test the significance between the differience of two proportions at 5% level.

[JNTU (H) Dec. 2011 (Set No.1)]

Solution : Let P_1 and P_2 be the population proportions in a locality who favour the bus stop.

Let the Null Hypothesis be $H_0 : P_1 = P_2$

Then the AlternativeHypothesis is $H_1 : P_1 \neq P_2$

Here n_1 = Number of men in I sample = 400

n_2 = Number of women in II sample = 200

x_1 = Number of men in favour of the proposal = 200

x_2 = Number of women in favour of the proposal = 40

$$\therefore\ p_1 = \frac{x_1}{n_1} = \frac{200}{400} = \frac{1}{2} \text{ and } p_2 = \frac{x_2}{n_2} = \frac{40}{200} = \frac{1}{5}$$

We have $p = \dfrac{n_1 p_1 + n_2 p_2}{n_1 + n_2} = \dfrac{x_1 + x_2}{n_1 + n_2} = \dfrac{200 + 40}{400 + 200} = \dfrac{240}{600} = \dfrac{2}{5}$

$$\therefore\ q = 1 - p = 1 - \frac{2}{5} = \frac{3}{5}$$

Assuming that H_0 is true, the test statistic is

$$z = \frac{p_1 - p_2}{\sqrt{pq\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}} = \frac{\frac{1}{2} - \frac{1}{5}}{\sqrt{\frac{2}{5} \times \frac{3}{5}\left(\frac{1}{400} + \frac{1}{200}\right)}}$$ 0.5·1

$$= \frac{0.3}{0.0424} = 7.07 > 1.96$$

Since $z > 1.96$, we reject the Null Hypothesis at 5% level of significance and conclude that the difference between the two proportions is highly significant *i.e.*, there is difference between the men and women in their attitude towards the bus stop near their residence.

Example 3 : A manufacturer of electronic equipment subjects samples of two completing brands of transistors to an accelerated performance test. If 45 of 180 transistors of the first kind and 34 of 120 transistors of the second kind fail the test, what can he conclude at the level of significance a = 0.05 about the difference between the corresponding sample proportions?

[JNTU 2003S, 2004 (Set No. 1)]

__Solution :__ We have $n_1 = 180,\ x_1 = 45,\ x_2 = 34,\ n_2 = 120$

and $p_1 = \dfrac{x_1}{n_1} = \dfrac{45}{180} = 0.25,\ p_2 = \dfrac{x_2}{n_2} = \dfrac{34}{120} = 0.283$

$$\therefore \quad p = \frac{n_1 p_1 + n_2 p_2}{n_1 + n_2} = \frac{x_1 + x_2}{n_1 + n_2} = \frac{45+34}{180+120} = \frac{79}{300} = 0.263$$

$$q = 1 - p = 1 - 0.263 = 0.737$$

1. **Null Hypothesis** $H_0 : p_1 = p_2$ *i..e,* there is no difference
2. **Alternative Hypothesis** $H_1 : p_1 \neq p_2$ *i.e.,* there is a difference
3. **Level of significance :** $\alpha = 0.05$
4. **The test statistic** is

$$z = \frac{p_1 - p_2}{\sqrt{pq\left(\dfrac{1}{n_1} + \dfrac{1}{n_2}\right)}} \quad \text{[Method of pooling]}$$

$$= \frac{0.25 - 0.283}{\sqrt{(0.263)(0.737)\left(\dfrac{1}{180} + \dfrac{1}{120}\right)}}$$

$$= \frac{-0.033}{\sqrt{(0.194)(0.01388)}} = \frac{-0.033}{0.0519} = -0.6358$$

$\therefore \quad |z| = 0.6358$

Since $|z| < 1.96$, we accept the null hypothesis H_0 at 5% level of significance *i.e.*, the difference between the proportions is not significant.

Example 4 : On the basis of their total scores, 200 candidates of a civil service examination are divided into two groups, the upper 30% and the remaining 70%. Consider the first question of the examination. Among the first group, 40 had the correct answer, whereas among the second group, 80 had the correct answer. On the basis of these results, can one conclude that the first question is not good at discriminating ability of the type being examined here ?

[JNTU 2003S, 2004, JNTU (K) Nov. 2009 (Set No. 2)]

__Solution :__ We have $n_1 = 60,\ n_2 = 140,\ x_1 = 40,\ x_2 = 80$

and $p_1 = \dfrac{x_1}{n_1} = \dfrac{40}{60} = \dfrac{2}{3} = 0.667,\ p_2 = \dfrac{x_2}{n_2} = \dfrac{80}{140} = 0.571$

$$\therefore p = \frac{n_1 p_1 + n_2 p_2}{n_1 + n_2} = \frac{x_1 + x_2}{n_1 + n_2} = \frac{40+80}{60+140} = \frac{120}{200} = 0.6$$

and $q = 1 - p = 1 - 0.6 = 0.4$

1. **Null Hypothesis** $H_0 : p_1 = p_2$
2. **Alternative Hypothesis** $H_1 : p_1 \neq p_2$
3. **Level of significance :** $\alpha = 0.05$ (assumed)
4. **The test statistic** is

$$z = \frac{p_1 - p_2}{\sqrt{pq\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}} \quad \text{[Method of pooling]}$$

$$= \frac{0.667 - 0.571}{\sqrt{(0.6)(0.4)\left(\frac{1}{60} + \frac{1}{40}\right)}} = \frac{0.096}{0.1} = 0.96$$

Since $|z| < 0.96$, we accept the null hypothesis H_0 at 5% level of significance i.e., the first question is good enough in discriminating the ability of the students of both groups.

Note : Suppose the population proportions P_1 and P_2 are given and $P_1 \neq P_2$. If we want to test the hypothesis that the difference $(P_1 - P_2)$ in population proportions is likely to be hidden in simple samples of sizes n_1 and n_2 from the two populations respectively, then

$$z = \frac{(p_1 - p_2) - (P_1 - P_2)}{\sqrt{\left(\frac{P_1 Q_1}{n_1} + \frac{P_2 Q_2}{n_2}\right)}}$$

If the sample proportions are not known, then we use

$$|z| = \frac{|P_1 - P_2|}{\sqrt{\frac{P_1 Q_1}{n_1} + \frac{P_2 Q_2}{n_2}}}$$

Example 5 : A cigarette manufacturing firm claims that its brand A line of cigarettes outsells its brand B by 8%. If it is found that 42 out of a sample of 200 smokers prefer brand A and 18 out of another sample of 100 smokers prefer brand B, test whether the 8% difference is a valid claim.

Solution : Here $n_1 = 200$, $n_2 = 100$,

$$p_1 = \frac{42}{200} = 0.21, \quad p_2 = \frac{18}{100} = 0.18$$

and $P_1 - P_2 = 8\% = \frac{8}{100} = 0.08$

1. **Null Hypothesis** H_0 **:** Assume that 8% difference in the sale of two brands of cigarettes is a valid claim i.e.,

$$H_0 : P_1 - P_2 = 0.08$$

2. **Alternative Hypothesis** $H_1 : P_1 - P_2 \neq 0.08$ (Two tail test)

3. **The test statistic is,** $z = \dfrac{(p_1 - p_2) - (P_1 - P_2)}{\sqrt{pq\left(\dfrac{1}{n_1} + \dfrac{1}{n_2}\right)}}$ [Method of pooling]

Now $p = \dfrac{n_1 p_1 + n_2 p_2}{n_1 + n_2} = \dfrac{200(0.21) + 100(0.18)}{200 + 100} = \dfrac{42 + 18}{300} = 0.2$

and $q = 1 - 0.2 = 0.8$

$$\therefore \quad z = \frac{0.03 - 0.08}{\sqrt{0.2 \times 0.8\left(\dfrac{1}{200} + \dfrac{1}{100}\right)}} = \frac{-0.05}{0.0489} = -1.02$$

$\therefore |z| = 1.02$

Since $|z| < 1.96$, we accept the null hypothesis H_0 at 5% level of significance. *i.e.*, 8% difference in the sale of two brands of cigarettes is a valid claim.

Example 6 : In two large populations, there are 30%, and 25% respectively of fair haired people. Is this difference likely to be hidden in samples of 1200 and 900 respectively from the two populations. **[JNTU (A) Nov. 2010]**

Solution :

Given $n_1 = 1200$, $n_2 = 900$

P_1 = Proportion of fair haired people in the first population $= \dfrac{30}{100} = 0.3$

P_2 = Proportion of fair haired people in the second population $= \dfrac{25}{100} = 0.25$

1. **Null Hypothesis** H_0 : Assume that the sample proportions are equal *i.e.*, the difference in population proportions is likely to be hidden in sampling *i.e.*, $H_0 : P_1 = P_2$

2. **Alternative Hypothesis** $H_1 : P_1 \neq P_2$

3. **The test statistic is,** $z = \dfrac{P_1 - P_2}{\sqrt{\dfrac{P_1 Q_1}{n_1} + \dfrac{P_2 Q_2}{n_2}}}$

where $Q_1 = 1 - P_1 = 1 - 0.3 = 0.7$; $Q_2 = 1 - P_2 = 1 - 0.25 = 0.75$

$$\therefore \quad z = \frac{0.3 - 0.25}{\sqrt{\dfrac{0.3 \times 0.7}{1200} + \dfrac{0.25 \times 0.75}{900}}}$$

$$= \frac{0.05}{0.0195} = 2.56$$

i.e., $z = 2.56$

Since $z > 1.96$, therefore we reject the Null Hypothesis H_0 at 5% level of significance (Two-tailed test) *i.e.*, the sample proportions are not equal. Thus we conclude that the difference in population proportions is unlikely that the real difference will be hidden.

Example 7 : A company wanted to introduce a new plan of work and a survey was conducted for this purpose. Out of sample of 500 workers in one group 62% favoured the new plan and another group of sample of 400 workers 41% were against the new plan. Is there any significant difference between the two groups in their attitude towards the new plan at 5% level of significance?

[JNTU (A) Dec. 2009 (Set No. 1)]

Solution : Given $n_1 = 500, n_2 = 400, p_1 = 0.62$ and $p_2 = 1 - 0.41 = 0.59$.

$$\therefore\ p = \frac{n_1 p_1 + n_2 p_2}{n_1 + n_2} = \frac{500(0.62) + 400(0.59)}{500 + 400} = \frac{310 + 236}{900} = \frac{546}{900} = 0.607$$

and $q = 1 - p = 1 - 0.607 = 0.393$.

1. **Null Hypothesis** $H_0 : p_1 = p_2$ *i.e.*, there is no significant difference between the two groups in their attitude towards the new plan.
2. **Alternative Hypothesis** $H_1 : p_1 \neq p_2$ (two tailed test)
3. **Level of significance :** $\alpha = 0.05$
4. **The test statistic** is $z = \dfrac{p_1 - p_2}{\sqrt{pq\left(\dfrac{1}{n_1} + \dfrac{1}{n_2}\right)}} = \dfrac{0.62 - 0.59}{\sqrt{(0.607)(0.393)\left(\dfrac{1}{500} + \dfrac{1}{400}\right)}}$

$$= \frac{0.03}{\sqrt{0.2385 \times 0.0045}} = \frac{0.03}{0.3276} = 0.0916$$

Since $|z| < 1.96$, the null hypothesis is accepted at 5% level of significance. Hence, we conclude that there is no significant difference between the two groups in their attitude towards the new plan.

Example 8 : In a city A, 20% of a random sample of 900 school boys has a certain slight physical defect. In another city B, 18.5% of a random sample of 1600 school boys has the same defect. Is the difference between the proportions significant at 0.05 level of significance.

[JNTU (K) Nov. 2009 (Set No. 1), (H) Nov. 2015]

Solution : We are given

n_1 = First sample size = 900

n_2 = Second sample size = 1600

x_1 = 20% of 900 = 180

x_2 = 18.5% of 1600 = 296

$$\therefore p_1 = \frac{x_1}{n_1} = \frac{180}{900} = 0.2, \quad p_2 = \frac{x_2}{n_2} = \frac{296}{1600} = 0.185$$

Thus $p = \dfrac{n_1 p_1 + n_2 p_2}{n_1 + n_2} = \dfrac{x_1 + x_2}{n_1 + n_2} = \dfrac{180+296}{900+1600} = \dfrac{476}{2500} = 0.19$

Now $q = 1 - p = 1 - 0.19 = 0.81$

1. **Null Hypothesis** $H_0 : p_1 = p_2$
2. **Alternative Hypotheses** $H_1 : p_1 \neq p_2$
3. **Level of significance :** $\alpha = 0.05$
4. **The test statistic is** $z = \dfrac{p_1 - p_2}{\sqrt{pq\left(\dfrac{1}{n_1} + \dfrac{1}{n_2}\right)}}$

i.e., $z = \dfrac{0.2 - 0.185}{\sqrt{(0.19)(0.81)\left(\dfrac{1}{900} + \dfrac{1}{1600}\right)}} = \dfrac{-0.015}{\sqrt{(0.1539)(0.001736)}}$

$= \dfrac{-0.015}{0.01634} = -0.918$

$\therefore |z| = 0.918$.

Since $|z| < 1.96$, we accept the null hypothesis H_0 at 5% level of significance *i.e.*, there is no significant difference between the proportions.

Example 9 : In a random sample of 1000 persons from town A, 400 are found to be consumers of wheat. In a sample of 800 from town B, 400 are found to be consumers of wheat. Do these data reveal a significant difference between town A and town B, so far as the proportion of wheat consumers is concerned? **[JNTU (H) Nov. 2009, (K) May 2012 (Set No. 2)]**

Solution : We have

n_1 = sample size of town A = 1000

n_2 = sample size of town B = 800

x_1 = Number of consumers of wheat from town A = 400

x_2 = Number of consumers of wheat from town B = 400

p_1 = Proportion of consumers of wheat in town A $= \dfrac{x_1}{n_1} = \dfrac{400}{1000} = 0.4$

p_2 = Proportion of consumers of wheat in town B $= \dfrac{x_2}{n_2} = \dfrac{400}{800} = 0.5$

$\therefore p = \frac{x_1 + x_2}{n_1 + n_2} = \frac{400+400}{1000+800} = \frac{800}{1800} = \frac{8}{18} = \frac{4}{9}$ and $q = 1 - p = 1 - \frac{4}{9} = \frac{5}{9}$.

1. **Null Hypothesis** $H_0 : p_1 = p_2$ *i.e.*, there is no difference.
2. **Alternative Hypothesis** $H_1 : p_1 \neq p_2$ *i.e.*, there is a difference.
3. **Level of significance :** $\alpha = 0.05$
4. **The test statistic is** $z = \frac{p_1 - p_2}{\sqrt{pq\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}} = \frac{0.4 - 0.5}{\sqrt{\frac{4}{9} \times \frac{5}{9}\left(\frac{1}{1000} + \frac{1}{800}\right)}}$

$$= \frac{-0.1}{\sqrt{\frac{20}{81} \times \frac{1800}{(1000)(800)}}} = -4.242$$

$\therefore |z| = 4.2424$

Also $z_\alpha = 1.96$

Since $|z| > z_\alpha$, we reject the null hypothesis H_0 at 5% level of significance.

$\therefore$ There is significant difference between town A and town B, as the proportion of wheat consumers is concerned.

Example 10 : In a sample of 600 students of a certain college 400 are found to use ball pens. In another college from a sample of 900 students 450 were found to use ball pens. Test whether 2 colleges are significantly different with respect to the habit of using ball pens.

[JNTU (H) Nov. 2010 (Set No. 1)]

Solution : Let P_1 and P_2 be the population proportions of students who use ball pens in two colleges.

Let the Null Hypothesis be $H_0 : P_1 = P_2$

Then the Alternative Hypothesis is $H_1 : P_1 \neq P_2$

Given x_1 = Number of students who use ball pens in first college = 400

n_1 = Sample size of first sample = 600

x_2 = Number of students who use ball pens in second college = 450

n_2 = Sample size of second sample = 900

$\therefore p_1$ = Proportion of students who use ball pens in first college $= \frac{x_1}{n_1} = \frac{400}{600} = \frac{2}{3}$

p_2 = Proportion of students who use ball pens in second college $= \frac{x_2}{n_2} = \frac{450}{900} = \frac{1}{2}$

$$\therefore\ p=\frac{n_1p_1+n_2p_2}{n_1+n_2}=\frac{x_1+x_2}{n_1+n_2}=\frac{400+450}{600+900}=\frac{850}{1500}=\frac{17}{30}=0.57$$

$$q=1-p=1-\frac{17}{30}=\frac{13}{30}=0.43$$

Assuming that H_0 is true, the test statistic is

$$z=\frac{p_1-p_2}{\sqrt{pq\left(\frac{1}{n_1}+\frac{1}{n_2}\right)}}=\frac{\frac{2}{3}-\frac{1}{2}}{\sqrt{(0.57)(0.43)\left(\frac{1}{600}+\frac{1}{900}\right)}}$$

$$=\frac{1}{6.\sqrt{0.2451(0.0027)}}=6.48$$

$\therefore\ z=6.48>1.96$

Hence, we reject the Null Hypothesis at 5% level of significance and conclude that there is a significant difference between the two colleges as far as using ball pens habit is concerned.

Example 11 : 100 articles from a factory are examined and 10 are found to be defective. 500 similar articles from a second factory are found to be 15 defective. Test the significance between the difference of two proportions at 5% level. **[JNTU (H) Nov. 2010 (Set No. 2)]**

Solution : Let P_1 and P_2 be the proportions of defective articles in the population of articles manufactured by the two factories.

Let the Null Hypothesis be $H_0 : P_1 = P_2$

Then the Alternative Hypothesis is $H_1 : P_1 \neq P_2$

Here $n_1=100$, $n_2=500$, $p_1=\frac{10}{100}=0.1$ and $p_2=\frac{15}{500}=0.03$

We have $p=\frac{n_1p_1+n_2p_2}{n_1+n_2}=\frac{100(0.1)+500(0.03)}{100+500}=\frac{10+15}{600}=\frac{25}{600}=\frac{1}{24}=0.042$

$\therefore\ q=1-p=1-0.042=0.958$

and S.E. of $(p_1-p_2)=\sqrt{pq\left(\frac{1}{n_1}+\frac{1}{n_2}\right)}=\sqrt{(0.042)(0.958)\left(\frac{1}{100}+\frac{1}{500}\right)}=0.022$

Assuming that H_0 is true, the test statistic is

$$z=\frac{p_1-p_2}{\text{S.E. of }(p_1-p_2)}=\frac{0.1-0.03}{0.022}=\frac{0.07}{0.022}=3.18$$

$\therefore\ z=3.18>1.96$

Hence, we reject the Null Hypothesis at 5% level of significance and conclude that there is a significant difference between the two proportions.

Example 12 : A machine produced 20 defective articles in a batch of 400. After overhauling it produced 10 defectives in a batch of 300. Has the machine being improved after overhauling? **[JNTU(K) Nov.2011 (Set No.4)]**

Solution : If p_1 and p_2 be the sample proportions of imperfect items put out by the machine before and after the overhauling and P_1 and P_2 be the corresponding population proportions.

Here $n_1 = 400, x_1 = 20, n_2 = 300$ and $x_2 = 10$.

Now $p_1 = \frac{20}{400} = \frac{1}{20}, p_2 = \frac{10}{300} = \frac{1}{30}$

Thus $p = \frac{n_1 p_1 + n_2 p_2}{n_1 + n_2} = \frac{x_1 + x_2}{n_1 + n_2} = \frac{30}{700} = \frac{3}{70}$

$\therefore q = 1 - p = 1 - \frac{3}{70} = \frac{67}{70}$

1. **Null Hypothesis:** $H_0 : p_1 = p_2$
2. **Alternative Hypothesis:** $H_1 : p_1 > p_2 \Rightarrow$ It leads to a one-tailed test.
3. **Level of significance:** $\alpha = 0.05$
4. Test Statistic is $z = \frac{p_1 - p_2}{\sqrt{pq\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}} = \frac{\frac{1}{20} - \frac{1}{30}}{\sqrt{\frac{3}{70} \times \frac{67}{70}\left(\frac{1}{400} + \frac{1}{300}\right)}}$

$$= \frac{0.0167}{\sqrt{0.041(0.0058)}} = \frac{0.0167}{0.0154} = 1.0844.$$

5. **Critical value:** The critical value of z at $\alpha = 0.05$ is $z_{0.05} = 1.645$ for one tailed test
6. **Decision.** Since $|z|(=1.0844) < z_{0.05}(=1.645)$

$\Rightarrow$ The Null Hypothesis is accepted

$\Rightarrow$ There is no significant difference.

Hence, the machine has not improved after overhauling.

Example 13 : A machine puts out 9 imperfect articles in a sample of 200 articles. After the Machine is overhauled it puts out 5 imperfect articles in a sample of 700 articles. Test at 5% level whether the Machine is improved? **[JNTU (H) Nov. 2010 (Set No. 3)]**

Solution : Let P_1 and P_2 be the proportions of imperfect articles in the population of articles manufactured by the machine before and after overhauling, respectively.

Let the Null Hypothesis be $H_0 : P_1 = P_2$

The Alternative Hypothesis is $H_1 : P_1 > P_2$

We have $p=\dfrac{n_1p_1+n_2p_2}{n_1+n_2}$ and S.E. of $(p_1-p_2)=\sqrt{pq\left(\dfrac{1}{n_1}+\dfrac{1}{n_2}\right)}$

Here $n_1=200, p_1=\dfrac{9}{200}=0.045, n_2=700, p_2=\dfrac{5}{700}=0.00714$

$$\therefore\ p=\frac{0.045\times 200+0.00714\times 700}{200+700}=\frac{9+5}{900}=\frac{14}{900}=0.0156$$

and $q=1-p=1-0.0156=0.9844$

$$\therefore\ \text{S.E. of } (p_1-p_2)=\sqrt{0.0156\times 0.9844\left(\frac{1}{200}+\frac{1}{700}\right)}=0.0099$$

Assuming that H_0 is true, the test satistic is

$$z=\frac{p_1-p_2}{\text{S.E. of } (p_1-p_2)}=\frac{0.045-0.00714}{0.0099}=\frac{0.03786}{0.0099}=3.824$$

Since $z=3.824>1.96$, we reject the Null Hypothesis at 5% level of significance and conclude that the machine is improved.

Example 14 : A machine puts out 16 imperfect articles in a sample of 500 articles. After the Machine is overhauled it puts out 3 imperfect articles in a sample of 100 articles. Has the Machine improved? **[JNTU (H) Dec. 2011 (Set No.3)]**

Solution : Let P_1 and P_2 be the proportions of imperfect articles in the population of articles manufactured by the machine before and after overhauling, respectively.

Let the Null Hypothesis be $H_0 : P_1 = P_2$

Then the Alternative Hypothesis is $H_1 : P_1 > P_2$

Here

n_1 = Sample size before the machine overhauling = 500

x_1 = No. of imperfect articles before overhauling = 16

n_2 = Sample size after the machine overhauling = 100

x_2 = No. of imperfect articles after overhauling = 3

$$\therefore\ p_1=\frac{x_1}{n_1}=\frac{16}{500}=0.032 \text{ and } p_2=\frac{x_2}{n_2}=\frac{3}{100}=0.03 \text{ and } p_1>p_2$$

We have $p=\dfrac{n_1p_1+n_2p_2}{n_1+n_2}=\dfrac{x_1+x_2}{n_1+n_2}=\dfrac{16+3}{500+100}=\dfrac{19}{600}=0.032$

and $q = 1 - p = 0.968$

Since H_1 is one - tailed (right - tailed), we use one-tailed test (right-tailed test).

Let us assume that H_0 is true. The test statistic is

$$z = \frac{p_1 - p_2}{\sqrt{pq\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}} = \frac{0.032 - 0.03}{\sqrt{0.032 \times 0.968\left(\frac{1}{500} + \frac{1}{100}\right)}}$$

$$= \frac{0.002}{0.019} = 0.104$$

Thus we see that $z = 0.104 < 1.645$. Hence we accept the Null Hypothesis H_0 at 5% level of significance and conclude that the machine has improved.

Example 15 : In an investigation on the machine performance the following results are obtained.

	No.of units inspected	No.of defectives
Machine 1	375	17
Machine 2	450	22

Test whether there is any significant performance of two machines at $\alpha = 0.05$

[JNTU (H) Nov. 2010 (Set No. 4)]

Solution : Let P_1 and P_2 be the proportions of defective units in the population of units inspected in Machine 1 and Machine 2 respectively.

Let the Null Hypothesis be $H_0 : P_1 = P_2$

The Alternative Hypothesis is $H_1 : P_1 > P_2$

$p_1 = \frac{17}{375} = 0.045$, $p_2 = \frac{22}{450} = 0.049$ and $p_1 > p_2$

We have $p = \frac{n_1 p_1 + n_2 p_2}{n_1 + n_2}$ and S.E. of $(p_1 - p_2) = \sqrt{pq\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}$

Here $n_1 = 375, n_2 = 450, p_1 = 0.045, p_2 = 0.049$

$$\therefore p = \frac{375 \times 0.045 + 450 \times 0.049}{375 + 450} = \frac{17 + 22}{375 + 450} = \frac{39}{825} = 0.047$$

and $q = 1 - p = 1 - 0.047 = 0.953$

$$\therefore \text{S.E. of } (p_1 - p_2) = \sqrt{0.047 \times 0.953\left(\frac{1}{375} + \frac{1}{450}\right)} = 0.015$$

Assuming that H_0 is true, the test statistic is

$$z = \frac{p_1 - p_2}{\text{S.E. of } (p_1 - p_2)} = \frac{0.045 - 0.049}{0.015} = -0.267$$

$\therefore |z| = 0.267 < 1.96$

Hence, we accept the Null Hypothesis at 5% level of significance and conclude that there is no significant difference in performance of two machines.

Example 16 : Among the items produced by a factory out of 500, 15 were defective, in another sample out of 400, 20 were defective. Test the significance between the differences of two proportions at 5% level. **[JNTU (H) Dec. 2011 (Set No. 2)]**

Solution : Let P_1 and P_2 be the proportions of defective items in the population of two sample items produced by the factory.

Let the Null Hypothesis be $H_0 : P_1 = P_2$

Then the Alternative Hypothesis is $H_1 : P_1 \neq P_2$.

Here $n_1 = 500, n_2 = 400, x_1 = 15$ and $x_2 = 20$

$$\therefore p_1 = \frac{x_1}{n_1} = \frac{15}{500} = \frac{3}{100} = 0.03 \text{ and } p_2 = \frac{x_2}{n_2} = \frac{20}{400} = \frac{1}{20} = 0.05$$

We have $p = \dfrac{n_1 p_1 + n_2 p_2}{n_1 + n_2} = \dfrac{x_1 + x_2}{n_1 + n_2} = \dfrac{15 + 20}{500 + 400} = \dfrac{35}{900} = 0.039$

Now $q = 1 - p = 1 - 0.039 = 0.961$

Assuming that H_0 is true, the test statistic is

$$z = \frac{p_1 - p_2}{\sqrt{pq\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}} = \frac{0.03 - 0.05}{\sqrt{0.039 \times 0.961\left(\frac{1}{500} + \frac{1}{400}\right)}}$$

$$= \frac{-0.02}{0.013} = -1.54$$

$\therefore |z| = 1.54 < 1.96$

Hence we accept the Null Hypothesis at 5% level of significance and conclude that there is no significant difference between the two proportions.

Example 17 : Among the items produced by a factory out of 800, 65 were defective in another sample out of 300, 40 were defective. Test the significance between the difference of two proportions at 1% level.

[JNTU (H) Apr. 2012 (Set No. 2)]

Solution : Let P_1 and P_2 be the proportions of defective items in the population of items produced by a factory in two samples respectively.

Let the Null Hypothesis be $H_0 : P_1 = P_2$

The Alternative Hypothesis is $H_1 : P_1 < P_2$

Since H_1 is one - tailed (left - tailed), we use one - tailed test (left - tailed test)

Here $n_1 = 800, x_1 = 65, p_1 = \frac{65}{800} = 0.081, n_2 = 300, x_2 = 40, p_2 = \frac{40}{300} = 0.1333$

and $p_1 < p_2$

$$\therefore\ p = \frac{n_1 p_1 + n_2 p_2}{n_1 + n_2} = \frac{x_1 + x_2}{n_1 + n_2} = \frac{65 + 40}{800 + 300} = \frac{105}{1100} = 0.095$$

and $q = 1 - p = 1 - 0.095 = 0.905$

Assuming that H_0 is true, the test statistic is

$$z = \frac{p_1 - p_2}{\sqrt{pq\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}} = \frac{0.081 - 0.1333}{\sqrt{0.095 \times 0.905\left(\frac{1}{800} + \frac{1}{300}\right)}}$$

$$= \frac{-0.0523}{\sqrt{0.086 \times 0.00459}} = \frac{-0.0523}{0.01986} = -2.6334$$

Thus we see that $z = -2.6334 < -2.33$

Hence, we reject the Null Hypothesis H_0 at 1% level of significance and conclude that there is a significant difference between the two proportions.

Example 18 : Before an increase on excise duty on tea 500 people out of a sample of 900 found to have the habit of having tea. After an increase on excise duty 250 are have the habit of having tea among 1100. Is there any decrease in the consumption of tea. Test at 5% level. **[JNTU (H) Nov. 2010 (Set No. 4)]**

Solution : Let P_1 and P_2 be the population proportions of people who have the habit of tea before and after an increase on excise duty on tea respectively.

Let the Null Hypothesis be $H_0 : P_1 = P_2$

The Alternative Hypothesis is $H_1 : P_1 > P_2$

Now $p_1 = \frac{500}{900} = 0.556,\ p_2 = \frac{250}{1100} = 0.227$

We have $p = \frac{n_1 p_1 + n_2 p_2}{n_1 + n_2}$ and S.E. of $(p_1 - p_2) = \sqrt{pq\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}$

Here $n_1 = 900, n_2 = 1100, p_1 = 0.556, p_2 = 0.227$

$$\therefore\ p=\frac{900\times0.556+1100\times0.227}{900+1100}=0.375$$

$$q=1-p=1-0.375=0.625$$

$$\therefore\ \text{S.E. of } (p_1-p_2)=\sqrt{0.375\times0.625\left(\frac{1}{900}+\frac{1}{1100}\right)}=0.022$$

Assuming that H_0 is true, the test statistic is

$$z=\frac{p_1-p_2}{\text{S.E. of } (p_1-p_2)}=\frac{0.556-0.227}{0.022}=14.95$$

Hence, we reject the Null Hypothesis at 5% level of significance and conclude that there is a significant difference in the consumption of tea before and after an increase on excise duty on tea.

Example 19 : During a country wide investigation the incidence of tuberculosis was found to be 1%. In a college of 400 students 3 reported to be affected, where as in another college of 1200 students 10 were affected. Does this indicate any significant difference?

[JNTU (H) Apr. 2012 (Set No.3)]

Solution : Let P_1 and P_2 be the population proportions of students who are affected tuberculosis in two colleges respectively.

Let the Null Hypothesis be $H_0 : P_1 = P_2$

Then the Alternative Hypothesis is $H_1 : P_1 \neq P_2$

Here $n_1 = 400, x_1 = 3, n_2 = 1200, x_2 = 10$

$$\therefore\ p_1=\frac{x_1}{n_1}=\frac{3}{400}=0.0075 \text{ and } p_2=\frac{x_2}{n_2}=\frac{10}{1200}=0.0083$$

Given $p=0.01$ and $q=1-P=0.99$

Assuming that H_0 is true, the test statistic is $z=\dfrac{p_1-p_2}{\sqrt{pq\left(\frac{1}{n_1}+\frac{1}{n_2}\right)}}$

$$\therefore\ z=\frac{0.0075-0.0083}{\sqrt{0.01\times0.99\left(\frac{1}{400}+\frac{1}{1200}\right)}}=\frac{-0.0008}{0.0057}=-0.14$$

$$\therefore\ |z|=0.14<1.96$$

Hence, we accept the Null Hypothesis at 5% level of significance and conclude that there is no significant difference between the two colleges as far as the incidence of tuberculosis concerned.

Example 20 : A random sample of 300 shoppers at a supermarket includes 204 who regularly use cents off coupons. In another sample of 500 shoppers at a supermarket includes 75 who regularly use cents off coupons

(a) Construct confidence interval for the probability that any one shopper at the supermarket, selected at random, will regularly use cents off coupons.

(b) Test the significance between the difference of two proportions at 2% level.

[JNTU (H) Apr. 2012 (Set No. 4)]

Solution : Here we are given

$n_1 = 300, x_1 = 204, n_2 = 500$ and $x_2 = 75$

$p_1 =$ Proportion of shoppers who use cents off coupons in the first sample $= \frac{204}{300} = 0.68$

$p_2 =$ Proportion of shoppers who use cents off coupons in the second sample

$= \frac{75}{500} = 0.15$

(a) The 98% confidence interval for the probability that any one shopper in sample selected at random is

$$\left(p_1 - z_{\alpha/2} \cdot \sqrt{\frac{p_1 q_1}{n_1}}, p_1 + z_{\alpha/2} \cdot \sqrt{\frac{p_1 q_1}{n_1}}\right) \text{ where } q_1 = 1 - p_1 = 1 - 0.68 = 0.32$$

i.e, $\left(0.68 - (2.33)\sqrt{\frac{0.68 \times 0.32}{300}}, 0.68 + (2.33)\sqrt{\frac{0.68 \times 0.32}{300}}\right)$

i.e., $(0.68 - 0.063, 0.68 + 0.063)$ or $(0.62, 0.74)$

(b) Here we set up the Null Hypothesis H_0 that $p_1 = p_2$

i.e., the sample proportions are equal

i.e., there is no difference in the two proportions.

Then the Alternative Hypothesis is $H_1 : p_1 \neq p_2$

Now $p = \frac{n_1 p_1 + n_2 p_2}{n_1 + n_2} = \frac{x_1 + x_2}{n_1 + n_2} = \frac{204 + 75}{300 + 500} = \frac{279}{800} = 0.35$

$\therefore q = 1 - p = 1 - 0.35 = 0.65$

Hence the test statistic is

$$z = \frac{p_1 - p_2}{\sqrt{pq\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}}$$

$$= \frac{0.68-0.15}{\sqrt{0.35\times0.65\left(\frac{1}{300}+\frac{1}{500}\right)}} = \frac{0.53}{0.035} = 15.14$$

Since $z = 15.14 > 2.58$, we reject the Null Hypothesis at 1% level of significance, and conclude that there is a significant difference in the two proportions.

REVIEW QUESTIONS

1. (*i*) Define Null Hypothesis. [JNTU (H) III yr. Nov. 2015]

   (*ii*) Explain the terms : Null Hypothesis and Alternate hypothesis [JNTU (H) Dec. 2014]

2. Explain the terms Type I and Type II errors. [JNTU (H) Dec. 2014]

   (or) Write about type I error and type II error. [JNTU (H) May 2017]

3. Explain the terms (*i*) one-tailed and (*ii*) two-tailed tests [JNTU (H) May, Sept. 2017]

4. Write the Test Statistic :

   (*i*) in the test of significance for single mean.

   (*ii*) for the test of significance for single proportion and

   (*iii*) for the test of significance for difference of proportion.

EXERCISE

1. In a random sample of 400 persons from a large population, 120 are females. Can it be said that males and females are in the ratio 5 : 3 in the population ? use 1% level of significance.

2. (*i*) A coin is tossed 900 times and heads appear 490 times. Does this result support the hypothesis that the coin is unbiased.

   (*ii*) A coin is tossed 10,000 times and it turns up head 5195 times. Test the hypothesis that the coin is unbiased. Use a 0.01 level of significance.

   (*iii*) A die is tossed 9000 times and it falls with 5 or 6 upwards 3240 times. Test the hypothesis that the die is unbiased. Use a 0.01 level of significance.

3. A wholesaler in apples claims that only 4% of the apples supplied by him are defective. A random sample of 600 apples contained 36 defective apples. Test the claim of the wholesaler.

4. In a sample of 500 people in Tamil Nadu 280 are tea drinkers and the rest are coffee drinkers. Can we assume that both coffee and tea are equally popular in this state at 1% level of significance.

5. A manufacturer claimed that atleast 98% of the steel pipes which he supplied to a factory conformed to specifications. An examination of a sample of 500 pieces of pipes revealed that 30 were defective. Test his claim at a significance level 5%.

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

7. In a sample of 600 students of a certain college 400 are found to use ball pens. In another college, from a sample of 900 students 450 were found to use ball pens. Test whether the two colleges are significantly different with respect to the habit of using ball pens. **[JNTU (H) Nov. 2010, Nov. 2012 (Set No. 1)]**

8. In a certain district A, 450 persons were considered regular consumers of tea out of a sample of 1000 persons. In another district B, 400 were regular consumers of tea out of a sample of 800 persons. Do these facts reveal a significant difference between the two districts as far as tea drinking habit is concerned ?

9. (*i*) A machine produced 10 defective articles in a batch of 200. After overhauling it produced 4 defectives in a batch of 100. Has the machine been improved.

   (*ii*) A machine puts out 21 defective articles in a sample of 500 articles. Another machine gives 3 defective articles in a sample of 100. Are the two machines significantly different in their performance ?

   (*iii*) A machine puts out 10 imperfect articles in a sample of 200. After the machine is overhauled it puts out 4 imperfect articles in a batch of 100. Has the machine been improved ?

10. A random sample of 100 articles selected from a batch of 2,000 articles shows that the average diameter of the articles = 0.354 with a *S.D.* = 0.048. Find 95% confidence interval for the average of this batch of 2,000 articles.

11. A sample of 100 iron bars is said to be drawn from a large number of bars whose lengths are normally distributed with mean 4 feet and *S.D.* 0.6 feet. If the sample mean is 4.2 feet, can the sample be regarded as a truly random sample.

12. The mean life of a sample of 100 electric bulbs produced by a company is found to be 1570 hrs with a *S.D.* of 120 hrs. If μ is the mean life time of all the bulbs produced by the company, test the hypothesis $\mu = 1600$ hrs against the alternative hypothesis $\mu \neq 1600$ hrs at 5% level of significance.

13. A sample of 400 male students is found to have a mean height of 171.38 cm. Can it be reasonably regarded as a sample from a large population with mean height of 171.17 cm and *S.D.* 3.30 cm.

14. A sample of 100 workers in a large plant gave a mean assembly time of 294 seconds with a *S.D.* of 12 seconds in a time and motion study. Find a 95% confidence interval for the mean assembly time for all the workers in the plant.

15. The mean breaking strength of the cables supplied by a manufacturer is 1800 with a *S.D.* 100. By a new technique in the manufacturing process, it is claimed that the breaking strength of the cables have increased. In order to test this claim, a sample of 50 cables is tested. It is found that the mean breaking strength is 1850. Can we support the claim at 1% level of significance.

16. An investigation of the relative merits of two kinds of flash light batteries showed that a random sample of 100 batteries of brand A tested on the average 36.5 hrs with a $S.D.$ of 1.8 hrs while a random sample of 80 batteries of brand B tested on the average 36.8 hrs with a $S.D.$ of 1.5 hrs. Use a level of significance of 0.05 to test whether the observed difference between the average life times is significant.

17. Given the following information relating to two places A and B. Test whether there is any significant difference between their mean wages :

	A	B
Mean wages (Rs.)	47	49
***S.D.* (Rs.)**	28	40
Number of workers	1000	1500

18. The mean consumption of food grains among 400 sampled middle class consumers is 380 gms per day per person with a $S.D.$ of 120 gms. A similar sample survey of 600 working class consumers gave a mean of 410 gms with a $S.D.$ of 80 gms. Are we justified in saying that the two classes consume the same quantity of food grains. Use 5% level of significance.

19. In a city 250 men out of 750 were found to be smokers. Does this information support the conclusion that the majority of men in this city are smokers.

[JNTU 2005S, 2006S, 2008S (Set No. 1)]

20. Test the significance of the difference between the means of the sample from the following data :

	Size of sample	Mean	S.D
Sample A	100	61	4
Sample B	200	63	6

(**Hint :** Refer Solved Example 4)

21. In a random sample of 100 tube lights produced by companny A, the mean life time of tube light is 1190 hours with standard deviation of 90 hours. Also in a random sample of 75 tube lights from company B the mean life time is 1230 hours with standard deviation of 120 hours. Is there a difference between the mean life times of the two brands of tube lights at a significance level of 0.05?

22. 500 articles from a factory are examined and found to be 2% defective. 800 similar articles from a second factory are found to haave only 1.5 % defective. Can it reasonably be concluded that the products of the first factory are inferior to those of second ?

23. In a referendum submitted to the student's body at a university 850 men and 566 women voted. 530 of the men and 304 of the women voted in favor of a matter. Does this indicate a significant difference of the opinion on the matter at least 1% level, between men and women students ?

ANSWERS

1. $Z = -3.125$; The males and females in the population are not in the ratio 5 : 3.
2. (*i*) $H_0 : P = \frac{1}{2}$ (coin is unbiased), $Z = 2.39$, rejected at 5% level.

   (*ii*) $H_0 : P = \frac{1}{2}$ (coin is unbiased), $Z = 3.9$, rejected at 1% level.

   (*iii*) $H_0 : P = \frac{1}{3}$ (die is unbiased), $Z = 5.4$, rejected at 1% level.
3. $H_0 : P = 0.04$; $H_1 : P > 0.04$, $Z = 10$, highly significant.
4. $H_0 : P = \frac{1}{2}$, $H_1 : P \neq \frac{1}{2}$, $Z = 2.68$, significant at 1% level.
5. $H_0 : P = 0.98$, $H_1 : P < 0.98$ (left-tail test), $Z = 6.38$, H_0 rejected.
6. $H_0 : P_1 = P_2$, $H_1 : P_2 < P_1$, $Z = 1.04$, not significant.
7. $H_0 : P_1 = P_2$; $H_1 : P_1 \neq P_2$; $|Z| = 6.48$, significant.
8. $H_0 : P_1 = P_2$, $H_1 : P_1 \neq P_2$, $|Z| = 2.08$, significant at 5% level.
9. (*i*) $H_0 : P_1 = P_2$, $H_1 : P_1 > P_2$, $|Z| = 0.3846$, not significant.

   (*i.e.*, not improved the performance of the machine)
10. 0.34, 0.36.
11. $H_0 : \mu = 4$ ft, $H_1 : \mu \neq 4$ ft, $Z = 3.33$ significant.
12. $|Z| = 2.5$, significant at 5% level of significance.
14. 291.64, 296.35.
15. $H_0 : \mu = 1800$, $\sigma = 100$, $H_1 : \mu_1 > 1800$, $Z = 3.53$, Accept H_0
16. $H_0 : \bar{x}_1 = \bar{x}_2$; $|Z| = 1.21$, Accept H_0
17. $H_0 : \bar{x}_1 = \bar{x}_2$, $|Z| < 1.96$, Accept H_0
18. $H_0 : \bar{x}_1 = \bar{x}_2$, $H_1 : \bar{x}_1 \neq \bar{x}_2$, $|Z| = 4.39$ significant.
19. $H_0 : p = \frac{1}{3}$, $H_1 : P < 0.5$; $z = -0.128$, H_0 rejected.
21. $|z| = 2.42$

OBJECTIVE TYPE QUESTIONS

1. A hypothesis is true, but is rejected, this is an error of type [ ]
   (a) I (b) II
   (c) I and II (d) None
2. A hypothesis is false, but accepted, this is an error of type [ ]
   (a) I (b) II
   (c) I and II (d) None
3. A random sample of 400 products contains 52 defective items. Standard error of proportion is [ ]
   (a) 0.168 (b) 0.0168
   (c) 0.0016 (d) 1.68
4. 500 eggs are taken from a large consignment and 50 are found to be bad. Standard error of proportion is [ ]
   (a) 1.3 (b) 0.13
   (c) 0.013 (d) None
5. Among 900 people in a state 90 are found to be chapati eaters. The 99% confidence interval for the true proportion is [ ]
   (a) (0.08, 0.12) (b) (0.8, 1.2)
   (c) (0.07, 0.13) (d) None
6. In a random sample of 160 workers exposed to a certain amount of radiation, 24 experienced some ill effects. The 99% confidence interval for the true proportion is [ ]
   (a) (0.065, 0.234) (b) (0.65, 0.23)
   (c) (0.56, 0.023) (d) None
7. A manufacturer of electric bulbs claims that the percentage defectives in his product does not exceed 6. A sample of 40 bulbs is found to contain 5 defectives. The test statistic $|Z| =$ [ ]
   (a) 0.24 (b) 1.24
   (c) 0.024 (d) 1.12
8. In a city 250 men out of 750 men were found to be smokers. The test statistic $|Z| =$ [ ]
   (a) 2.25 (b) 2.5
   (c) 5.25 (d) 5.50
9. A die is thrown 256 times. An even digit turns up 150 times. Then the die is [ ]
   (a) biased (b) unbiased
   (c) not determined (d) None

10. A single - tailed test is used when __________

11. Null hypothesis is defined as __________

ANSWERS

1. *a* 2. *b* 3. *b* 4. *c*
5. *c* 6. *a* 7. *b* 8. *c*
9. *b*
10. We are testing the hypothesis that one process is better than another.
11. The hypothesis formulated for the sake of rejecting it,under the assumption that it is true.

PREVIOUS GATE QUESTIONS

1. Type II error in hypothesis testing is [GATE 2016(CE)]
   (A) acceptance of the null hypothesis when it is false and should be rejected.
   (B) rejection of the null hypothesis when it is true and should be accepted.
   (C) rejection of the null hypothesis when it is false and should be rejected.
   (D) acceptance of the null hypothesis when it is true and should be accepted.
2. For accepting or rejecting a null hypothesis, which one of the following is NOT used as a significance test method in Statistics ? [GATE 2018(AG)]
   (A) z- test
   (B) Student's t-test
   (C) Pearson's correlation
   (D) Relative standard deviation
3. Standard error is [GATE 2018(BT)]
   (A) the probability of a type I error in a statistical test
   (B) the error in estimating a sample standard deviation
   (C) the standard deviation of a variable that follows standard normal distribution.
   (D) the standard deviation of distribution of sample means

ANSWERS

1. A 2. D 3. D