LESSON 5

TESTING OF HYPOTHESIS

Introduction:-

In a statistical investigation the interest usually lies in the assessment of the general magnitude and the study of variation with respect to one or more characteristics relating to individuals belonging to a group. This group of individuals under study is called POPULATION (or) UNIVERSE.

Professor Morris Hamburg states that 'a hypothesis in statistics in simply a quantitative statement about a population 'it is an assumption that is made about parameter values and then its validity is tested by statistical technique which ultimately tell us whether the hypothesis is correct and is sustained or whether it is false and is to be rejected.

Sample and sample size

A finite subset of statistical individuals in a population is called a sample and the number of individuals in a sample is called the sample size.

Sampling Error:-

For determining population characteristics in stead of enumerating the entire population the individuals in the sample only are observed. Then the sample characteristics are utilized to approximately determine or estimate the population. The error involved in such approximation is known as sampling error and is inherent and unavoidable in every sampling scheme.

Parameter and statistic:-

The statistical constants of the population. viz.,mean(μ) variance σ^2 are usually referred to as parameters.

Population mean is denoted by μ

Population variance is denoted by σ^2

The statistical measures computed from the sample observations alone are called statistics.

Sample mean is denoted by $\bar{x}$.

Sample variance is denoted by s^2.

Unbiased estimate:

A statistic $t = t(x_1,x_2,\ldots,x_n)$, a function of the sample values x_1, x_2 , ...,x_n is an unbiased estimate of the population parameter θ, if $E(t) = \theta$.

That is if E(statistic) = Parameter

Then statistic is said to be an unbiased estimate of the parameter.

Standard Error:-

The standard deviation of the sampling distribution of a statistic is known as its STANDARD ERROR abbreviated as standard error.

The standard errors of some of the well-known statistics, are given below,

where n is the sample size,

σ^2 is the population variance and

P is the population parameter and

Q=1-P; n_1 and n_2 represent the size of two independent random samples respectively drawn from the given population (s)

S.No.	Statistic	Standard Error
1	Sample Mean $\bar{x}$	$\sigma / \sqrt{n}$
2.	Observed sample proportion 'p'	$\sqrt{PQ/n}$
3	Sample s.d: s	$\sqrt{(\sigma^2/2n)}$
4	Sample variance:s2	$\sigma^2\sqrt{(2/n)}$
5	Sample quartiles	$1.36263\sigma/\sqrt{n}$
6.	Sample median	$1.25331\ \sigma/\sqrt{n}$
7	Sample correlation coefficient (r)	$(1-\rho^2)/\sqrt{n}$
8	Sample moment:μ_3	$\sigma^3\sqrt{96/n}$
9	Sample moment: μ_4	$\sigma^4\sqrt{96/n}$
10	Sample coefficient of variation (v)	$(v/\sqrt{2n})\ \sqrt{(1+(2v^3)/10^4)} = v/\sqrt{(2n)}$
11	Difference of two sample means $(\bar{x}_1 - \bar{x}_2)$	$\sqrt{(\sigma_1^2/n_1) + (\sigma_2^2/n_2)}$
12	Difference of two sample s.d's: $(s_1 - s_2)$	$\sqrt{(\sigma_1^2/2n_1) + (\sigma_2^2/2n_2)}$
13	Difference of two sample proportions: $(p_1 - p_2)$	$\sqrt{(P_1Q_1/n_1) + (P_2Q_2/2n_2)}$

Standard Error enables us to determine the probable limits within which the population parameter may be expected to lie.

Tests of Significance:-

The study of the tests of significance, enables us to decide on the basis of the sample results, if

i) the deviation between the observed sample statistic and the hypothetical parameter value.

ii) the deviation between two independent sample statistics;

Null Hypothesis:-

For applying the test of significance we first set up a hypothesis – a definite statement about the population parameter. Such a hypothesis, which is a hypothesis of no difference is called null hypothesis and is usually denoted by H_0. This null hypothesis presumes that there is no difference between sample statistic and the parameter value.

Having set up the null hypothesis, we compute the probability that the deviation between the observed sample statistic and the hypothetical parameter value might have occurred due to fluctuations of sampling. If the deviation comes out to be significant null hypothesis is rejected at the particular level of significance adopted. If the deviation is not significant, null hypothesis may be retained at that level.

Alternative hypothesis :- Any hypothesis which is complementary to the null hypothesis is called an alternative hypothesis, usually denoted by H_1.

For example : If we want to test the null hypothesis that the population has a specified mean μ_0., (i.e) $H_0: \mu = \mu_0$ then the alternative hypothesis could be

(i) $H_1: \mu \neq \mu_0$ that is $\mu > \mu_0$ or $\mu < \mu_0$

(ii) $H_1: \mu > \mu_0$

(iii) $H_1: \mu < \mu_0$

The alternative hypothesis in (i) is known as Two-tailed alternative and the alternatives in (ii) and (iii) are known as Right tailed and left tailed alternatives respectively.

Errors in Sampling :- There are two types of errors in sampling

Type I error : Reject H_0 when it is true.

Type II error: Accept H_0 when it is wrong.

That is, accept H_0 when H_1 is true.

If P{ Reject H_0 when it is true} = p {Reject H_0/H_0} = α

And p{Accept H_0 when it is wrong} = P {Accept $H_0/H1$} = β

Then α and β are called the sizes of type I error and type II error respectively.

Critical Region and level of significance:-

A region in the sample space S which amounts to rejection of H_0 is termed as **CRITICAL REGION OF REJECTION**.

If ω is the critical region and if $t = t(x_1, x_2, \ldots, x_n)$ is the value of the statistic based on a random sample of size n, then

$$P(t \in \omega / H_0) = \alpha, \; P(t \in \omega / H_0) = \beta$$

where $\bar{\omega}$, the complementary of ω, is calle the *acceptance region.*

The probability that a random value of the statistic t belongs to the critical region is known as the *level of significance.*

The level of significane usually employed in testing of hypothesis are 5% and 1%. If the sample size is less than or equal to 30 then it is called *small sample* and if the sample size is greater than 30 then it is called *large sample*.

Tests of significance for large samples:-

For large values of n, we apply the normal test. which is based upon the following fundamental property of the normal probability curve.

If $X \sim N(\mu, \sigma^2)$ then $Z = \dfrac{X - \mu}{\sigma} = \dfrac{X - E(X)}{\sqrt{V(X)}} \sim N(0, 1)$

The significant values of Z at 5% and 1% levels of significance for two-tailed tests are 1.96 and 2.58 respectively. In all probability we should expect a standard normal variate to lie between ± 3

Compute the test statistic Z under H_0. Thus for a two-tailed test if $|Z| > 1.96$ then H_0 is rejected at 5% level of significance. Similarly if $|Z| > 2.58$, H_0 is contradicted at 1% level of significane and if $|Z| \leq 2.58$, H_0 may be accepted at 1% level of significance.

For a single-tailed test (Right tail or left tail) we compare the computed value of $|Z|$ with 1.645 (at 5% level) and 2.33 (at 1% level) and accept or reject H_0 accordingly.

Procedure for Testing of Hypothesis

1) Null hypothesis: Set up the Null hypothesis H_0

2) Alternative hypothesis: Setup the Alternative hypothesis H_1. This will enable us to decide

whether we ... ...tailed (right or left) test or two-tailed test.

3) Level of significance: Choose the appropriate level of significance (α) depending on the reliability of the estimates.

4) Test Statistic : Compute the test statistic

$$Z = \frac{t - E(t)}{S.E(t)} \quad \text{under } H0$$

5) Conclusion: We compare the computed value of Z in step 4 with the significant value(tabulated value) Z_α at the given level of significane 'α'.

If $|Z| < Z_\alpha$ then the null hypothesis may be accepted

If $|Z| > Z_\alpha$ then the null hypothesis is rejected at the level of significance α

Test of significance for single proportion:

If X is the number of successes in 'n' trials with constant probability P of success for each trial, then $E(X) = nP$ and $V(X) = nPQ$ where $Q = 1-P$ is the probability of failure.

For large n, $X \sim N(nP, nPQ)$

That is, $$Z = \frac{X - E(X)}{\sqrt{V(X)}} = \frac{X - nP}{\sqrt{nPQ}} \sim N(0,1)$$

and we can apply the normal test.

Problem 1:- In a sample of 1,000 people in Maharashtra, 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?

Solution Given, $n=1{,}000$

$X=$ number of rice eaters $= 540$

$p=$ sample proportion of rice eaters

$$= \frac{X}{n} = \frac{540}{1000} = 0.54$$

Null hypothesis: H_0: Both rice and wheat are equally popular in the state so that

P = population proportion of rice eaters in Maharashtra

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

Alternative hypotheis: H_1: $P \neq 0.5$ (two-tailed test)

Test Statistic : under H0, the test statistic is

$$Z = \frac{p-P}{\sqrt{PQ/n}} \sim N(0,1) \quad \text{(since n is large)}$$

$$Z = \frac{0.54 - 0.50}{\sqrt{0.5 \times 0.5/1000}} = 2.532$$

Conclusion: The significane or critical value of Z at 1% level of significance for two tailed test is 2.58.

Since computed value of $Z = 2.532$ is less than 2.58, it is not significant at 1% level of significance. Hence the null hypothesis is accepted and we may conclude that rice and wheat are equally popular in Maharashtra state.

Test of significance for Difference of proportions:

Suppose we want to compare two distinct populations with respect to the prevalence of a certain attribute, say A, among their members. Let X_1, X_2 be the number of persons possessing the given attribute A in random samples of sizes n_1 and n_2 from the two populations respectively. Then sample proportions are given by

$$p_1 = \frac{X_1}{n_1} \quad \text{and} \quad p_2 = \frac{X_2}{n_2}$$

If P_1 and P_2 are population proportions, then

$$E(p_1) = P_1, \; E(p_2) = P_2, \; V(p_1) = \frac{P_1Q_1}{n_1}, \; V(p_2) = \frac{P_2Q_2}{n_2}$$

$$Z = \frac{(p_1 - p_2) - E(p_1 - p_2)}{\sqrt{V(p_1 - p_2)}} \sim N(0,1)$$

Under the null hypothesis H_0: $P_1 = P_2$,

The test statistic for the difference of proportions is

$$Z = \frac{p_1 - p_2}{\sqrt{PQ(1/n_1 + 1/n_2)}} \sim N(0,1)$$

Under H_0: $P_1 = P_2 = P$, an unbiased estimate of the population proportion P, based on both the samples is given by

$$\hat{P} = \frac{n_1p_1 + n_2p_2}{n_1 + n_2} = \frac{X_1 + X_2}{n_1 + n_2}$$

Problem 2:- Random samples of 400 men and 600 women were asked whether they would like to have a flyover near their residence. 200men and 325 women were in favour of the proposal. Test the hypothesis that propositions of men and women in favour of the proposal, are same against that they are not at 5% level.

Solution:-

Null hypothesis: H_0: $P_1 = P_2 = P$ (say)

There is no significant difference between the opinions of men and women as far as proposal of flyover is concerned.

Alternative hypothesis: H_1: $P_1 \neq P_2$ (two-tailed)

Given $n_1=400$, X_1= number of men favouring the proposal = 200

$n_2=600$, X_2= number of women favouring the proposal =325

Therefore,

p_1 = proportion of men favouring the proposal in the sample

$$\frac{X_1}{n_1} = \frac{200}{400} = 0.5$$

p_2 = proportion of men favouring the proposal in the sample

$$\frac{X_2}{n_2} = \frac{325}{600} = 0.54$$

Test statistic: Since samples are large, the test statistic under the null hypothese, H0 is:

$$Z = \frac{p_1 - p_2}{\sqrt{\hat{P}\hat{Q}(1/n_1+1/n_2)}} \sim N(0,1)$$

$$\hat{P} = \frac{n_1p_1+n_2p_2}{n_1+n_2} = \frac{X_1+X_2}{n_1+n_2} = \frac{200+325}{400+600} = 0.525$$

$$\hat{Q} = 1-\hat{P} = 0.475$$

$$Z = \frac{0.500 - 0.541}{\sqrt{0.525 \times 0.475 \times (1/400 + 1/600)}} = -1.269$$

Conclusion: Since $|Z| = .269$ which is less than 1.96, it is not significant at 5% level of significane. Hence H_0 may be accepted at 5% level of significane and we may conclude that men and women do not differ significantly as regards proposal of flyover is concerned.

Test of significance for single mean.

If x_i, $(i=1,2,\ldots,n)$ is a random sample of size n from a normal population with mean μ and variance σ^2, thane the sample mean is distributed normally with mean μ and

variance σ^2/n, i.e., $\bar{x} \sim N(\mu, \sigma^2/n)$.

Thus for large samples , the standard normal variate

corresponding to $\bar{x}$ is:

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

under the *null hypothesis* Ho, that the sample has been drawn from a population with mean μ and variance σ^2. i.e., there is no significant difference between the sample

mean $(\bar{x})$ and population mean (μ) the test statistic (for large samples), is:

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

Problem 3: A sample of 900 members has a mean 3.4 cms and s.d 2.61 cms. Is the sample from a large population of mean 3.25cms and s.d. 2.61 cms? If the population is normal and its mean is unknown. Find the 95% and 98% fiducial limits of true mean.

Solution: *Null Hypothesis*(H_0) : The sample has been drawn from the population with mean μ=3.25 cms and S.D. σ = 2.61 cms.

Alternative Hypothesis: H_1: $\mu \neq 3.25$ (Two-tailed)

Test Statistic: Under H_0, the test statistic is $Z = \frac{\bar{x} - \mu}{\sigma/\sqrt{n}} \sim N(0,1)$

Here we are given : $\bar{x}$=3.4 cms, n=900 cms, μ = 3.25 cms and σ = 2.61 cms

Therefore

$$Z = \frac{3.40 - 3.25}{2.61 / \sqrt{900}} = \frac{0.15 \times 30}{2.61} = 1.73$$

Since | Z | < 1.96, we conclude that the data don't provide us any evidence against the null hypothesis (H_0) which may, therefore, be accepted at 5% level of significance.

95% fiducial limits for the population mean are:

$\bar{x} \pm 1.96\,(\sigma/\sqrt{n})$ = 3.40 ± 0.1705 i.e.. 3.5705 and 3.2295

98% fiducial limits for μ are given by:

$\bar{x} \pm 2.33\,(\sigma/\sqrt{n})$ = 3.40 ± 0.2027 i.e., 3.6027 and 3.1973

Problem 4: The guaranteed average life a certain type of electric light bulbs is 1,000 hours with a standard deviation of 125 hours. It is decided to sample the output so as to ensure that 90 per cent of the bulbs do not fall short of the guaranteed average by more than 2.5 per cent. What must be the minimum size of the sample?

Solution. Here μ = 1,000 hours, σ = 125 hours.

Since we do not want the sample mean to be less than the guaranteed average mean(μ = 1,000) by more than 2.5% we should have

$$\bar{x} > 1{,}000 - 2.5\% \text{ of } 1{,}000 \Rightarrow \bar{x} > 1{,}000 - 25 = 975$$

Let n be the given sample size. Then

$$Z = \frac{\bar{x} - \mu}{\sigma/\sqrt{n}} \sim N(0,1), \text{ since sample is large.}$$

We want $$Z = \frac{\bar{x} - \mu}{\sigma/\sqrt{n}} > \frac{975 - 1{,}000}{125/\sqrt{n}} = -\sqrt{n}/5$$

According to the given condition : $P(Z > -\sqrt{n}/5) = 0.90 \Rightarrow P(0<Z<\sqrt{n}/5) = 0.40$

$$\Rightarrow \sqrt{n}/5 = 1.28$$

$$n = 25 \times (1.28)^2 = 41 \text{ (approx).}$$

Test of Significance for Difference of Means: $\bar{x}_1$ be the mean of a sample of size n_1

from a population with mean μ_1 and variance σ_1^2 and let $\bar{x}_2$ be the mean of an independent random sample of size n_2 from another population with mean μ_2 and variance σ_2^2 Then, since sample sizes are large,

$$\bar{x}_1 \sim N(\mu_1, \sigma_1^2 / n_1) \text{ and } \bar{x}_2 \sim N(\mu_2, \sigma_2^2 / n_2)$$

Also $\bar{x}_1 - \bar{x}_2$, being the difference of two independent normal variates is also a

normal variate. The value of Z (S.N.V) corresponding to $\bar{x}_1 - \bar{x}_2$ i. given by:

$$Z = \frac{(\bar{x}_1 - \bar{x}_2) - E(\bar{x}_1 - \bar{x}_2)}{S.E.(\bar{x}_1 - \bar{x}_2)} \sim N(0,1)$$

Under the null hypothesis, $H_0 : \mu_1 = \mu_2$, i.e., there is no significant difference between the sample means, we get:

$E(\bar{x}_1 - \bar{x}_2) = E(\bar{x}_1) - E(\bar{x}_2) = \mu_1 - \mu_2 = 0$; $V(\bar{x}_1 - \bar{x}_2) = V(\bar{x}_1) + V(\bar{x}_2)$

$$= \sigma_1^2 / n_1 + \sigma_2^2 / n_2$$

the covariance term vanishes, since the sample means $\bar{x}_1 - \bar{x}_2$ are independent.

Thus under $H_0 : \mu_1 = \mu_2$, the test statistic becomes (for large samples),

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

Problem 5: The means of two single large samples of 1,000 and 2,000 members are 67.5 inches and 68.0 inches respectively. Can the samples be regarded as drawn from the same population of standard deviation 2.5 inches? (Test at 5% level of signifinance.)

Solution. In usual notations we are given.

n_1= 1,000, n2=2,000, $\bar{x}_1$= 67.5 inches $\bar{x}_2$= 68.0 inches.

Null hypothesis, Ho : $\mu_1 = \mu_2$ and σ = 2.5 inches, i.e., the samples thave been drawn from the same population of standard deviation 2.5 inches.

Alternative Hypothesis, $H_1 : \mu_1 \neq \mu_2$ (Two – tailed)

Test Statistic. Under H_0, the test statistic is:

$$Z = \frac{\bar{x} \qquad \bar{x}_2}{\sqrt{\{\sigma 2(1/n1+1/n2)\}}} \sim N(0,1)$$

$$\text{Now } Z = \frac{67.5 - 68.0}{2.5 \times \sqrt{(1/1000 + 1/2000)}} = -5.1$$

Conclusion:- Since | Z | > 3, the value is highly significant and we reject the null hypothesis and conclude that samples are certainly not from the same population with standard deviation 2.5

Problem 6:- 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 Rs. 250 with a standard deviation of Rs. 40. For 400 women shoppers chosen at random in super market 'B' in another section of the city, the average weekly food expenditure is Rs. 220with a standard deviation of Rs. 55. Test at1% level of significance whether the average weekly food expenditure of the two populations of shoppers are equal.

Solution: In the usual notations, we are given that

2.

$n_1 = 400$, $\bar{x}_1 =$ Rs. 250, $s_1 =$ Rs. 40

$n_2 = 400$, $\bar{x}_2 =$ Rs. 220, $s_2 =$ Rs. 55

Null hypothesis: $H_0: \mu_1 = \mu_2$, i.e., the average weekly food expenditures of the two populations of shoppers are equal.

Alternative hypothesis: $H_1: \mu_1 \neq \mu_2$. (Two tailed)

Test Statistic:- Since samples are large, under H_0, the test statistic is :

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

Since σ_1 and ~~σ_2, the~~ population standard deviations are not known, we can take for large

samples $\sigma_1^2 = s_1^2$ and $\sigma_2^2 = s_2^2$ and then Z is given by :

$$Z = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{(s_1^2/n_1 + s_2^2/n_2)}} = \frac{250 - 220}{\sqrt{\{(40)^2/400 + (55)^2/400\}}} = 8.82$$

Conclusion: Since $|Z|$ is much greater than 2.58, the null hypothesis is rejected at 1% level of significance and we conclude that the average weekly expenditures of two populations of shoppers in markets A and B differ significantly.

Test of Significance for the Difference of Standard deviations:

If s_1 and s_2 are the standard deviations of two independent samples, then under null hypothesis,

$H_0: \sigma_1 = \sigma_2$, i.e., sample standard deviations don't differ significantly, the statistic:

$$Z = \frac{s_1 - s_2}{S.E(s_1 - s_2)} \sim N(0,1) \text{ for large samples}$$

But in case of large samples, the S.E. of the differences of the sample standard deviations is given by :

$$S.E.(s_1-s_2) = \sqrt{(\sigma_1^2/2n_1+\sigma_2^2/2n_2)}$$

Therefore $Z = \dfrac{s_1 - s_2}{\sqrt{(\sigma_1^2/2n_1+\sigma_2^2/2n_2)}} \sim N(0,1)$

σ_1^2 and σ_2^2 are usually unknown and for large samples, we use their estimates given by the corresponding sample variances. Hence the test statistic reduces to

$$Z = \frac{s_1 - s_2}{\sqrt{(s_1^2/2n_1+s_2^2/2n_2)}} \sim N(0,1)$$

Problem :- 7

Random samples drawn from two countries give the following data relating to the heights of adult males:

	Country A	Country B
$\bar{x}$ Mean height (in inches) 5	67.42	67.25
s1, s2 Standard deviation (in inches)	2.58	2.50
n1, n2 Number in samples	1,000	1,200

(i) Is the difference between the means significant?

(ii) Is the difference between the standard deviations significant?

Solution:-

In usual notations, we are given:

$n_1 = 1,000$ $\bar{x}_1 = 67.42$ inches $s_1 = 2.58$ inches

$n_2 = 1,200$ $\bar{x}_2 = 67.25$ inches $s_2 = 2.50$ inches

$\hat{\sigma}_1 = s_1 = 2.58,$ $\hat{\sigma}_2 = s_2 = 2.50$

i) $H_0: \mu_1 = \mu_2$, i.e., the sample means do not differ significantly

$H_1: \mu_1 \neq \mu_2$ (Two tailed)

Under the Null hypothesis: H_0, the test statistic is

$$\boxed{Z = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{(s_1^2/n_1 + s_2^2/n_2)}}} \sim N(0,1)$$

0.17

$$Z = \frac{67.42 - 67.25}{\sqrt{\{(2.58)^2/1000 + (2.50)^2/1200\}}} = 1.56$$

Conclusion: Since $|Z| < 1.96$, null hypothesis may be accepted at 5% level of significance and we may conclude that there is no significant difference between the sample means.

ii) Under H_0: that there is no significant difference between sample standard deviations,

$$\boxed{Z = \frac{s_1 - s_2}{S.E(s_1 - s_2)}} \sim N(0,1) \text{ for large samples}$$

$$\boxed{S.E.(s_1 - s_2) = \sqrt{(\sigma_1^2/2n_1 + \sigma_2^2/2n_2)} = \sqrt{(s_1^2/2n_1 + s_2^2/2n_2)}}$$

If σ_1 and σ_2 are not known and $\sigma_1 = s_1$, $\sigma_2 = s_2$.

$S.E.(s_1 - s_2) = \sqrt{\{(2.58)^2/(2 \times 1000) + (2.50)^2/(2 \times 1200)\}}$
$= 0.07746$

Hence $Z = \dfrac{2.58 - 2.50}{0.07746} = 1.03$

Conclusion: Since $|Z| < 1.96$, the data don't provide us any evidence against the null hypothesis which may be accepted at 5% level of significance. Hence the sample standard deviations do not differ significantly

Chi-square Test:-

In general if X_i, $(i = 1,2,\ldots,n)$ are n independent normal variates with means μl and variances σ_i^2, $(i = 1,2,\ldots,n)$, then

$$\chi^2 = \sum_{i=1}^{n}\left(\frac{x_i - \mu_i}{\sigma_i}\right)$$

$$\chi^2 = \sum_{i=1}^{n}\left(\frac{X_i - \mu_i}{\sigma_i}\right)^2,$$ is a chi-square variate with n degrees of freedom.

The number of independent variates which make up the up the statistic(χ^2) is known as degrees of freedom (d.f) and is usually denoted by υ ν

APPLICATIONS OF CHI-SQUARE DISTRIBUTION

χ^2 – distribution has a large number of applications in statistics, some of which are enumerated below:

(i) To test if the hypothetical value of the population variance is $\sigma^2 = \sigma_0^2$ (say).

(ii) To test the 'goodness of fit'.

(iii) To test the independence of attributes.

(iv)To test the homogeneity of independent estimates of the population variance.

(v) To combine various probabilities obtained from independent experiments to

give a single test of significance.

(vi) To test the homogeneity of independent estimates of the population

correlation coefficient.

Inferences About a Population Variance.

$\bar{x} = \frac{x}{n}$

Suppose we want to test if a random sample x_i, ($i = 1,2,3,\ldots,n$) has been drawn from a normal population with a specified variance $\sigma^2 = \sigma_0^2$ (say)

Under the null hypothesis that the population variance is $\sigma^2 = \sigma_0^2$, the statistic

$$\chi^2 = \sum\left[\frac{(x_i - \bar{x})^2}{\sigma_0^2}\right] = \frac{1}{\sigma_0^2}\left[\sum x_i^2 - \frac{(\sum x_i)^2}{n}\right] = \frac{ns^2}{\sigma_0^2}$$

follows chi-square distribution with (n-1) d.f

By comparing the calculated value with the tabulated value of $\chi 2$ for (n-1) d.f at certain level of significance (usually 5%), we may retain or reject the null hypothesis.

Problem 8 :-

It is belived that the precision (as measured by the variance) of an instrument is no more than 0.16. Write down null and alternative hypothesis for testing this belief. Carryout the test at 1% level given 11 measurements of the same subject on the instrument:

2.5, 2.3, 2.4, 2.3, 2.5, 2.7, 2.5, 2.6, 2.6, 2.7, 2.5.

Solution.

Null Hypothesis H_0: $\sigma^2 = 0.16$

Alternative Hypothesis H_1 : $\sigma^2 > 0.16$

COMPUTATION OF SAMPLE VARIANCE

$\bar{x} = \frac{27.6}{11}$

$\bar{x} = 2.51$

X	$\bar{x} = 2.51$ $X - \bar{X}$	$(X - \bar{X})^2$
2.5	-0.01	0.0001
2.3	-0.21	0.0441
2.4	-0.11	0.0121
2.3	-0.21	0.0441
2.5	-0.01	0.0001
2.7	+0.19	0.0361
2.5	-0.01	0.0001
2.6	+0.09	0.0081
2.6	+0.09	0.0081
2.7	+0.19	0.0361
2.5	-0.01	0.0001
$\bar{X} = 2.51$	$\Sigma(X - \bar{X})^2 = 0.1891$	

Under the null hypothesis, H_0: $\sigma^2 = 0.16$, the test statistic is

$$\chi^2 = \frac{\Sigma(X - \bar{X})^2}{\sigma^2} = \frac{0.1891}{0.16} = 1.182$$

which follows χ^2 distribution with d.f. $n - 1 = 11 - 1 = 10$.

Since the calculated value of χ^2 is less than the tabulated value 23.2 of χ^2 for 10 d.f. at 1% level of significance, it is not significant. Hence H_0 may be accepted and we conclude that the data are consistent with the hypothesis that the precision of the instrument is 0.16.

Problem 9:-

Test the hypothesis that $\sigma = 10$, given that $s = 15$ for a random sample of size 50 from a normal population.

ution

Null hypothesis, H_0: $\sigma = 10$

We are given $n = 50$, $s = 15$

Therefore $\chi^2 = \frac{ns^2}{\sigma^2} = \frac{50 \times 225}{100} = 112.5$

$\boxed{\text{the test statistic is } Z = \sqrt{2\chi^2} - \sqrt{2n-1} \sim N(0, 1)}$

$$= \sqrt{225} - \sqrt{99}$$
$$= 15 - 9.95$$
$$= 5.05.$$

Since $|Z| > 3$, it is significant at all levels of significance and hence H_0 is rejected and we conclude that $\sigma \neq 10$.

Goodness of Fit Test:

A very powerful test for testing the significance of the discrepancy between theory and experiment was given by Prof. Karl Pearson in 1900 and is known as "*Chi-square test of goodness of fit*". It enables us to find if the deivation of the experiment from theory is just by chance or is it really due to the inadequacy of the theory to fit the observed data.

If f_i $(i=1,2,\ldots,n)$ is a set of observed frequencies and e_i $(i=1,2,\ldots,n)$ is the corresponding set of expected frequencies, then Karl Pearson's chi-square is given by:

$$\chi^2 = \sum_{i=1}^{n} \left[\frac{(f_i - e_i)^2}{e_i}\right], \qquad \sum_{i=1}^{n} f_i = \sum_{i=1}^{n} e_i \qquad \ldots\ldots(*)$$

follows chi-square distribution with $(n-1)$ d.f.

... rule:

Accept H_0 if $\chi^2 \leq \chi^2_\alpha(n-1)$ and reject H0 if $\chi^2 > \chi^2_\alpha(n-1)$, where χ^2 is the calculated value of chi-square obtained on using (*) and χ^2_α (n-1) is the tabulated value of chi-square for (n-1) d.f and level of significance α.

Problem 10: The demand for a particular spare part in a factory was found to vary from day-to-day. In a sample study the following information was obtained:

Days:	Mon	Tue	Wed	Thu	Fri	Sat
No.of parts Demanded:	1124	1125	1110	1120	1126	1115

Test the hypothesis that the number of parts demanded does not depend on the day, of the week. (given: the values of chi-square significance at 5,6,7 d.f respectively 11.07, 12.59, 14.07 at the 5% level of significance)

Solution:

Null Hypothesis, H_0 is the number of parts demanded does not depend on the day of the week. Under the null hypothesis, the expected frequencies of the spare part demanded on each of the six days would be:

(1/6)(1124 + 1125 + 1110 + 1120 + 1126 + 1115) = 1120

CALCULATIONS FOR χ^2

Days	Frequency		$(f_i - e_i)^2$	$\frac{(f_i - e_i)^2}{e_i}$
	Observed (f_i)	Expected (e_i)		
Mon	1124	1120	16	0.014

Tues	1125	1120	25	0.022
Wed	1110	1120	100	0.089
Thu	1120	1120	0	0
Fri	1126	1120	36	0.032
Sat	1115	1120	25	0.022
Total	6720	6720		0.179

$$\chi 2 = \frac{\sum(f_i - e_i)^2}{e_i} = 0.179$$

The number of degrees of freedom = 6 – 1 = 5

The tabulated $\chi 2_{0.05}$ for 5 d. f. = 11.07

Since calculated value of $\chi 2$ is less than the tabulated value, it is not significant and the null hypothesis may be accepted at 5% level of significance. Hence we conclude that the number of parts demanded are same over the 6 – day period.

Problem 11:-

The following figure show the distribution of digits in numbers chosen at random from a telephone directory:

Digits :	0	1	2	3	4	5	6	7	8	9	Total
Frequency:	1026	1107	997	966	1075	933	1107	972	964	853	10,000

Test whether the digits may be taken to occur equally frequently in the directory.

Solution:

Null hypothesis : digits occur equally frequently in the directory.

Under the null hypothesis, the expected frequency for each of the digits 0,1,2,....9 is 10,000/10 = 1,000. The value of $\chi 2$ is computed as follows.

CALCULATIONS FOR χ^2

Digits	Frequency		$(f_i - e_i)^2$	$\frac{(f_i - e_i)^2}{e_i}$
	Observed (f_i)	Expected (e_i)		
0	1026	1000	676	0.676
1	1107	1000	11449	11.449
2	997	1000	9	0.009
3	966	1000	1156	1.156
4	1075	1000	5625	5.625
5	933	1000	4489	4.489
6	1107	1000	11149	11.449
7	972	1000	784	0.784
8	964	1000	1296	1.296
9	853	1000	21609	21.609
Total	10,000	10,000		58.542

$$\chi 2 = \frac{\sum (f_i - e_i)^2}{e_i} = 58.542$$

The number of degrees of freedom = 10 – 1 = 9

The tabulated $\chi 2_{0.05}$ for 9 d. f. = 16.919

Since calculated value of $\chi 2$ is much greater than the tabulated value, it is highly significant and we reject the null hypothesis.Thus we conclude that the digits are not uniformly distributed in the directory.

Exercise

1) The following table gives the number of aircraft accidents that occurred during the seven days of the week. Find whether the accidents are uniformly distributed over the week.

Days :	mon	tue	wed		thu	frisat	total
No. of Accidents:	14	18	12	11	15	14	84

2) A die is thrown 60 times with the following results

Face :	1	2	3	4	5	6
Frequency:	8	7	12	8	14	11

Test at 5% level of significance if the die is unbiased. assuming that $P(\chi 2 > 11.1) = 0.05$ with 5 d.f.

3) 200 digits are chosen at random from a set of tables The frequencies of the digits were:

Digits:	0	1	2	3	4	5	6	7	8	9
Frequency:	18	19	23	21	16	25	22	20	21	15

Use $\chi 2$ test to assess the correctness of hypothesis that the digits were distributed in equal numbers in the table. given that the values of $\chi 2$ are respectively 16.9, 18.3 and 19.7 for 9, 10 and 11 degrees of freedom at 5% level of significance.

6

4) A sample of 15 values shows that the s.d is 6.4. Is this compatible with the hypothesis that the sample is from a normal population with s.d 5? 5

5) Test the hypothesis $\sigma = 8$, given that $s=10$, for a random sample of size 51 from a normal population

...: H_1: $\mu \neq 100$

Test Statistic: Under H0, the test statistic is

$$\frac{\bar{x} - \mu}{\sqrt{s^2/(n-1)}} \sim t_{(n-1)}$$

where $\bar{x}$ and S^2 are to be computed from the sample values of I.Q.'s

Calculation for sample mean and s.d.

X	$(x-\bar{x})$	$(x-\bar{x})^2$
70	-27.2	739.84
120	22.8	519.84
110	12.8	163.84
101	3.8	14.44
88	-9.2	84.64
83	-14.2	201.64
95	-2.2	4.84
98	0.8	0.64
107	9.8	96.04
100	2.8	7.84
Total 972		1833.60

Hence n= 10, $\bar{x} = 972 / 10 = 97.2$ and $S^2 = 1833.60/9 = 203.73$

$$| t | = \frac{|97.2 - 100|}{\sqrt{203.73/10}} = 0.62$$

Tabulated $t_{0.05}$ for (10-1) i.e., 9 d.f for two-tailed test is 2.262

Conclusion: Since calculated t is less than tabulated $t_{0.05}$ for 9 d.f., H0 may be accepted at 5% level of significance and we may conclude that the data are consistent with the assumption of mean I.Q. of 100 in the population.

The 95% confidence limits within which the mean I.Q. values of samples of 10 boys will lie are given by:

$\bar{x} \pm t_{0.05}\ S/\sqrt{n} = 97.2 \pm 2.262 \times 4.514 = 107.41$ and 86.99

Hence the required 95% confidence interval is [86.99, 107.41]

Problem 3: The heights of 10 males of a given locality are found to be 70, 67, 62, 68,61, 68,70, 64, 64,66 inches. Is it reasonable to believe that the average height is greater than 64 inches? Test at 5% significance level assuming that for 9 degrees of freedom $P(t > 1.83) = 0.05$

Solution: Null hypothesis, H_0: $\mu = 64$ inches.

Alternative hypothesis: H1: $\mu > 64$ inches

Calculations for sample mean and s.d.

X	70	67	62	68	61	68	70	64	64	66	Total
$(x-\bar{x})$	4	1	-4	2	-5	2	4	-2	-2	0	0
$(x-\bar{x})^2$	16	1	16	4	25	4	16	4	4	0	90

$$\bar{x} = \sum x / n = 660/10 = 66$$

$$S2 = \frac{1}{(n-1)} \sum(x - \bar{x})^2 = 90/9 = 10$$

Test Statistic: Under H_0, the test statistic is:

$$t = \frac{\bar{x} - \mu}{\sqrt{S^2/n}} = \frac{66 - 64}{\sqrt{10/10}} = 2$$

which follows student's t-distribution with 10-1 = 9.d.f.

Tabulated value of t for 9 d.f. at 5% level of significance for single(right) tailed test is 1.833.

Conclusion: Since calculated value of t is greater than the tabulated value , it is significant. Hence H0 is rejected at 5% level of significance and we conclude that the average height is greater than 60 inches.

t-test for Difference of Means:

Suppose we want to test if two independent samples x_i $(i=1,2,...n_1)$ and y_j, $(j=1,2,... n_2)$ of sizes n_1 and n_2 have been drawn from two normal populations with means μ_X and μ_Y respectively.

Under the null hypothesis (H0) that the samples have been drawn from the normal populations with means μ_X and μ_Y and under the assumption that the population variance are equal, i.e., $\sigma_X^2 = \sigma_Y^2 = \sigma^2$ (say), the statistic

$$t = \frac{(\bar{x} - \bar{y}) - (\mu_X - \mu_Y)}{S\sqrt{(1/n_1 + 1/n_2)}}$$

Where $\bar{x} = (1/n_1) \sum_{j=1}^{n} x_i$

$$\bar{y} = (1/n_2) \sum_{j=1}^{n} y_j$$

$$S^2 = \frac{1}{n_1+n_2-2} \left[\sum_i (x_i - \bar{x})^2 + \sum_j (y_j - \bar{y})^2 \right]$$

is an unbiased estimate of the common population variance σ^2, follows Student's distribution with (n_1+n_2-2) d.f.

Paired t-test for Difference of Means:

Let us now consider the case when i) the sample size are equal, i.e, $n_1=n_2=n$ and ii) the two samples are not independent but the sample observations are paired together, .e., the pair of observations (x_i, y_i) $(i=1,2,\ldots n)$ corresponds to the same (ith) sample unit. The problem is to test if the sample means differ significantly or not.

Here we consider the increments, $d_i = x_i - y_i$ $(i=1,2,\ldots n)$

Under the null hypothesis, H_0 that increments are due to fluctuations of sampling:, the statistic:

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

Where

$$\bar{d} = (1/n) \sum_i d_i \quad \text{and} \quad S^2 = \frac{1}{(n-1)} \sum_i (d_i - \bar{d})^2$$

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

Problem 4:-

Below are given the gain in weights (in kgs) of pigs fed on two diets A and B

Gain in weight

Diet A : 25, 32, 30, 34, 24, 14, 32, 24, 30, 31, 35, 25
Diet B: 44, 34, 22, 10, 47, 31, 40, 30, 32, 35, 18, 21, 35, 29, 22

Test, if the two diets differ significantly as regards their effect on increase in weight.

Solution:

Null hypothesis, H_0: $\mu_X = \mu_Y$, i.e., there is no significant difference between the mean increase in weight due to diets A and B.

Alternative hypothesis, H_1: $\mu_X \neq \mu_Y$ (two tailed)

Diet A			Diet B		
X	$(x-\bar{x})$	$(x-\bar{x})^2$	y	$(y-\bar{y})$	$(y-\bar{y})^2$
25	-3	9	44	14	196
32	4	16	34	4	16
30	2	4	22	-8	64
34	6	36	10	-20	400
24	-4	16	47	17	289
14	-14	196	31	1	1
32	4	16	40	10	100
24	-4	16	30	0	0
30	2	4	32	2	4
31	3	9	35	5	25
35	7	49	18	-12	144
25	-3	9	21	-9	81
			35	5	25
			29	-1	1
			22	-8	64

$\sum x=336$	$\sum(x-\bar{x})=0$	$\sum(x-\bar{x})^2=380$	$\sum y=450$	$\sum(y-\bar{y})=0$	$\sum(y-\bar{y})^2=1410$

$\bar{x} = (336/12) = 28,$ $\bar{y} = (450/15) = 30$

$$S^2 = \frac{1}{n_1+n_2-2}\left[\sum_i (x_i - \bar{x})^2 + \sum_j (y_j - \bar{y})^2\right]$$

$$= 71.6$$

$n_1=12$, $n_2=15$

Under null hypothesis(H_0):

$$t = \frac{(\bar{x} - \bar{y})}{S\sqrt{(1/n_1 + 1/n_2)}} \sim t_{n1+n2-2}$$

$$T = \frac{28 - 30}{\sqrt{71.6(1/12 + 1/15)}} = -0.609$$

Tabulated $t_{0.05}$ for $(12 + 15 - 2) = 25$ d.f. is 2.06

Conclusion : Since calculated $|t|$ is less than tabulated H0 may be accepted at 5% leve' of significance and w may conclude that the two diets do not differ significantl as regards their effect on increase in weight.

Problem 5: - Samples of two types of electric bulbs were tested for length of life and following data were obtained:

	Type I	Type II
Sample No.	$n_1=8$	$n_2=7$
Sample means	$\bar{x}_1=1234$ hrs.	$\bar{x}_2=1036$ hrs.

Sample s.d $\qquad s_1=36$ hrs $\qquad s_2=40$ hrs.

Is the difference in the means sufficient to warrant that type I is superior to type II regarding length of life?

Solution:

Null hypothesis. H_0: $\mu_X = \mu_Y$, i.e., the two types I and II of electric bulbs are identical.

Alternative hypothesis, H_1: $\mu_X > \mu_Y$, i.e., type I is superior to type II

Test Statistic: Under H_0, the test statistic is:

$$t = \frac{(\bar{x} - \bar{y})}{S\sqrt{(1/n_1 + 1/n_2)}} \sim t_{n1+n2-2} = t_{13}$$

where

$$S^2 = \frac{1}{n_1+n_2-2}\left[\sum_i (x_i - \bar{x})^2 + \sum_j (y_j - \bar{y})^2\right]$$

$$= \frac{1}{13}\left[8 \times (36)^2 + 7 \times (40)^2\right]$$

$$=1659.08$$

$$t= \frac{1234 - 1036}{\sqrt{1659.08(1/8 + 1/7)}} = 9.39$$

Tabulated value of t for 13 d.f. at 5% level of significance for right (single)-tailed test is 1.77.

Conclusion: Since calculated 't' is much greater than tabulated 't', it is highly significant and H_0 is rejected. Hence the two types of electric bulbs differ significantly. Further, since $\bar{x}_1$ is much greater than $\bar{x}_2$, we conclude that type I is definitely superior to type II.

t-test for testing the significance of an observed sample correlation coefficient

If r is the observed correlation coefficient in a sample of n pairs of observations from a bivariate normal population, then Prof. Fisher proved that under the null hypothesis, H_0: $\rho = 0$, i.e., *population correlation coefficient is zero*, the statistic:

$$t = \frac{r}{\sqrt{(1-r^2)}} \sqrt{(n-2)}$$

follows Student's t-distribution with (n-2) d.f

If the value of t comes out to be significant, we reject H_0 at the level of significance adopted and conclude that $\rho \neq 0$, i.e., 'r' is significant of correlation in the population.

If t comes out to be non-significant, then H_0 may be accepted and we conclude that variables may be regarded as uncorrelated in the population.

Problem 6: a) A random sample of 27 pairs of observations from a normal population gave a correlation coefficient of 0.6. Is this significant of correlation in the population?

b) Find the least value of r in a sample of 18 pairs of observations from a bi-variate normal population, significant at 5% level of significance.

Solution: a)

We set up the null hypothesis, H_0: $\rho = 0$, i.e., the observed sample correlation coefficient is not significant of any correlation in the population.

Under H_0: $$t = \frac{r}{\sqrt{(1-r^2)}} \sqrt{(n-2)} \sim t_{(n-2)}$$

$$t = \frac{0.6\sqrt{27-2}}{\sqrt{(1-0.36)}} = 3.75$$

Tabulated $t_{0.05}$ for $(27-2) = 25$ d.f. is 2.06

Conclusion: Since calculated t is much greater than the tabulated t, it is significant and hence H_0 is discredited at 5% level of significance. Thus we conclude that the variables are correlated in the population.

b) Here n=18. From the tables $t_{0.05}$ for $(18-2) = 16$ d.f. is 2.12.

Under H_0: $\rho = 0$,

$$t = \frac{r}{\sqrt{(1-r^2)}} \sqrt{(n-2)} \sim t_{(n-2)} = t_{16}$$

Inorder that the calculated value of t is significant at 5% level of significance, we should have

$$\left| \frac{r\sqrt{(n-2)}}{\sqrt{(1-r^2)}} \right| > t0.05 \quad \Rightarrow \quad \left| \frac{r\sqrt{16}}{\sqrt{(1-r^2)}} \right| > 2.12$$

$$r^2 > (4.493/20.493) = 0.2192$$

Hence $|r| > 0.4682$

$\frac{r^2}{1-r^2} = (2.12)$

Problem 7:- A coefficient of correlation of 0.2 is derived from a random sample of 625 pairs of observations. I) Is this value of r significant ? II) what are the 95% and 99% confidence limits to the correlation coefficient in the population?

Solution: Under the null hypothesis H_0: $\rho = 0$, i.e., the value of $r = 0.2$ is not significant,

The test statistic is :

$$t = \frac{r}{\sqrt{(1-r^2)}} \sqrt{(n-2)} \sim t_{(n-2)}$$

$$t = \frac{0.2 \times \sqrt{(625-2)}}{\sqrt{(1-0.04)}} = 5.09$$

Since d.f.=625 – 2 = 623, the significant values of t are same as in the case of normal distribution, viz., $t_{0.05} = 1.96$ and $t_{0.01}=2.58$. Since calculated t is much greater than these values; it is highly significant. Hence H_0: $\rho = 0$ is rejected and we conclude that the sample correlation is significant of correlation in the population.

95% Confidence Limits for ρ (population correlation coefficient) are :

$$r \pm 1.96\ S.E.(r) = r \pm 1.96\ (1-r^2) / \sqrt{n}.$$

$$= 0.2 \pm (1.96 \times 0.96 / \sqrt{625})$$
$$= (0.125, 0.275)$$

99% Confidence Limits for ρ are:

$$0.2 \pm 2.58 \times 0.0384 = 0.2 \pm 0.099$$
$$= (0.101, 0.299)$$

F-test for Equality of Two Population variances:

Suppose we want to test i) whether two independent samples x_i, $(i=1,2,\ldots,n_1)$ and y_j, $(j=1,2,\ldots,n_2)$ have been drawn from the normal populations with the same variance σ^2 or ii) whether the two independent estimates of the population variance are homogeneous or not.

Under the null hypothesis (H_0) that

i) $\sigma_X^2 = \sigma_Y^2 = \sigma^2$, i.e., the population variances are equal, or

ii) Two independent estimates of the population variance are homogeneous, the statistic F is given by:

$$\boxed{F = \frac{S_X^2}{S_Y^2}}$$

Where $$\boxed{S_X^2 = \frac{1}{(n_1-1)} \sum_{i=1}^{n_1} (x_i - \bar{x})^2}$$

$$\boxed{S_Y^2 = \frac{1}{(n_2-1)} \sum_{j=1}^{n_2} (y_j - \bar{y})^2}$$

are unbiased estimates of the common population variance σ^2 obtained from the independent samples and it follows Snedecor's F-distribution with (υ_1, υ_2) d.f. where $\upsilon_1 = n_1 - 1$ and $\upsilon_2 = n_2 - 1$

Problem 8: Two random samples gave the following results:

	Size	Sample mean	Sum of squares of Deviations from the mean
1	10	15	90 $\Sigma(x_1-\bar{x}_1)^2$
2	12	14	108 $\Sigma(x_2-\bar{x}_2)^2$

Test whether the samples come from the same normal population at 5% level of significance.

[Given : $F_{0.05}(9,11) = 2.90$, $F_{0.05} = (11, 9) = 3.10$ and $t_{0.05}(20) = 2.086$,

$t_{0.05}(22) = 2.07$]

Solution:

A normal population has two parameters, viz., mean μ and variance σ^2. To test if two independent samples have been drawn from the same normal population, we have to test (i) the equality of population mean, and (ii) the equality of population variances.

Null hypothesis: The two samples have been drawn form the same normal populton , i.e.,

$$H_0: \mu_1=\mu_2 \text{ and } \sigma_1^2 = \sigma_2^2$$

Equality of means will be tested by applying t-test and equality of variances will be tested by applying F-test. Since t-test assumes $\sigma_1^2 = \sigma_2^2$, we first apply F-test and then t-test.

In usual notations, we are given:

$n_1=10$, $n_2=12$

$x_1=15$, $x_2=14$

$\Sigma(x_1-\bar{x}_1)^2 = 90$, $\Sigma(x_2-\bar{x}_2)^2 = 108$

F-test.

$$S_1^2 = \frac{1}{(n_1-1)} \Sigma (x_1 - \bar{x}_1)^2 = 90/9 = 10$$

$$S_2^2 = \frac{1}{(n_2-1)} \Sigma (x_2 - \bar{x}_2)^2 = 108/11 = 9.82$$

Since $S_1^2 > S_2^2$, under H_0: $\sigma_1^2 = \sigma_2^2$, the test statistic is

$$F = \frac{S_1^2}{S_2^2} \sim F(n1-1, n2-1) = F(9,11)$$

$$F = \frac{S_1^2}{S_2^2} = 10/9.82 = 1.018$$

Tabulated $F_{0.05}$ (9,11) = 2.90 since calculated F is less than tabulated F, it is not significant. Hence null hypothesis of equality of population variances may be accepted.

Since $\sigma_1^2 = \sigma_2^2$, we can apply t test for testing H_0: $\mu_1 = \mu_2$

t-test: Under H_0' : $\mu_1 = \mu_2$, against alternative hypothesis, H_1' : $\mu_1 \neq \mu_2$, the test statistic is :

$$\boxed{t = \frac{(\bar{x}_1 - \bar{x}_2)}{\sqrt{S^2(1/n_1 + 1/n_2)}}} \sim t_{n1+n2-2} = t_{20}$$

where

$$\boxed{S^2 = \frac{1}{n_1+n_2-2} \left[\Sigma (x_1 - \bar{x}_1)^2 + \Sigma (x_2 - \bar{x}_2)^2 \right]}$$

$$= (1/20)\,[90+108] = 9.9$$

$$t = \frac{15 - 14}{\sqrt{9.9(1/10+1/12)}}$$

$$= 0.742$$

Since tabulated value t0.05 for 20 d.f. = 2.086. Since | t | <t0.05,. it is not significant. Hence the hypothesis H' : $\mu_1 = \mu_2$ may be accepted. Since both the hypothesis, i.e., H_0' : $\mu_1 = \mu_2$ and H_0: $\sigma_1^2 = \sigma_2^2$ are accepted; we may regard that the given samples have been-drawn from the same normal population.

r

Problem 9: A correlation coefficient of 0.72 is obtained from a sample of 29 pairs of observations.

n

i) Can the sample be regarded as drawn from a bivariate normal population in which true correlation coefficient is 0.8? ρ

ii) Obtain 95% confidence limits for ρ in the light of the information provided by the sample.

Solution: i) H0: There is no significant difference between r=0.72; and ρ=0.80, i.e., the sample can be regarded as drawn from the bivariate normal population with ρ=0.8. Here

$$\boxed{Z = (1/2) \log_e [(1+r)/(1-r)]} = 1.1513 \log_{10} 6.14 = 0.907$$

Seek

$$\xi = (1/2) \log_e [(1+\rho) / (1-\rho)] = 1.513 \times 0.9541 = 1.1$$

$$\text{S.E. } (Z) = \frac{1}{\sqrt{(n-3)}} = \frac{1}{\sqrt{26}} = 0.196$$

Under H0, the test statistic is :

$$U = \frac{Z - \xi}{1/\sqrt{(n-3)}} \sim N(0,1)$$

$$U = \frac{(0.907 - 1.100)}{0.196} = -0.985$$

Since $|U| < 1.96$, it is not significant at 5% level of significance and H_0 may be accepted. Hence the sample may be regarded as coming from a bivariate normal population with $\rho = 0.8$

ii)95% confidence limits for p on the basis of the information supplied by the sample are given by:

$(Z-\xi/(1/\sqrt{n-3})) \leq 1.96$

$|U| \leq 1.96 \Rightarrow |Z - \xi| \leq 1.96 \times (1/\sqrt{(n-3)}) = 1.96 \times 0.196$

$\rightarrow |0.907 - \xi| \leq 0.384$ or

$0.907 - 0.384 \leq \xi \leq 0.907 + 0.384$

$\rightarrow 0.523 \leq \xi \leq 1.291$ or ξ formula

$0.523 \leq (1/2)\boxed{\log_e[(1+p)/(1-p)]} \leq 1.291$

$$\frac{0.523}{1.1513} \leq \log_{10}[(1+p)/(1-p)] \leq \frac{1.291}{1.1513}$$

$$0.4543 \leq \log_{10}[(1+p)/(1-p)] \leq 1.1213 \quad ---- (*)$$

$\log_{10}[(1+p)/(1-p)] = 0.4543 \quad \log_{10}[(1+p)/(1-p)] = 1.1213$

1-3 =

$(1+\rho)/(1-\rho) = \text{Antilog}(0.4543) = 2.8464$

$[(1+\rho)/(1-\rho)] = \text{Antilog}(0.4543) = 2.846$ $[(1+\rho)/(1-\rho)] = \text{Antilog}(0.4543) = 2.846$

$\rho = 0.4799$ $(1+\rho)/(1-\rho) = \text{Antilog}(1.1213)$

$\rho = 0.86$ $= 13.22$

Hence, substituting in (*) we get

$$0.48 \leq \rho \leq 0.86$$

Exercise

1) Ten individuals are chosen at random from a normal population and their heights are found to be 63, 63, 66, 67,68,69,70,70,71,71 inches. Test if the sample belongs to the population whose mean heights is 66". [Given $t_{0.05} = 2.62$ for 9 d.f.]

2. A random sample of 8 cigarettes of a certain brand has an average nicotine content of 18.6 milligrams and a standard deviation of 2.4 milligrams. Is this in line with the manufacturer's claim that average nicotine content does not exceed 17.5 milligrams? Use a 0.01 level of significance and assume the distribution of nicotine contents to be normal.

3)The average length of time for students to register for summer classes at a certain college has been 50 minutes with s.d of 10 minutes. A new registration procedure using modern computing machines is being tried. If a random sample of 12 students had an average registration time of 42 minutes with s.d. of 11.9 minutes under the new system, test the hypothesis that the population mean has not changed , using 0.05 as level of significance

... how you would use Student's t test to decide whether the two sets of observations [: . .27.18.25,27,29,27,23,17] and [16.16,20,16,20,17,15,21] indicate samples drawn from the same universe.

5) Two independent samples of 8 and 7 items respectively had the following values of the variables:.

Sample I :.	9	11	13	11	15	9	12	14
Sample II:	10	12	10	14	9	8	10	

Do the estimates of population variance differ significantly?

6) Five measurements of the output of two units have given the following results (in kilograms of material per one hour of operation)

Unit A :	14.1	10.1	14.7	13.7	14.0
Unit B:	14.0	14.5	13.7	12.7	14.1

Assuming that both samples have been obtained from normal population, test at 10% significance level if two populations have the same variance, it being given $F_{0.95}(4,4) = 6.39$

n_1.

7. In one sample of 10 observations from a normal population, the sum of the squares of the deviations of the sample values from the sample mean is 102.4 $(x_1-\bar{x}_1)^2$ and in another sample of 12 observations n_2 from another normal population, the sum of the squares of the deviations of the sample values from the sample mean is 120.5 $(x_2-\bar{x}_2)^2$. Examine whether the two normal populations have the same variance.

Transport n/w & Trees Application.