Kunthavai Naachiyaar Govt. Arts College (W) Autonomous, Thanjavur

B.Stat Major **18K5SO8**

Core Course -VIII - Statistical Inference - II **Hrs:5**

Credit:5

Unit – I

Testing of hypothesis – definition, Simple and composite hypothesis, power of a test, most powerful test. Test of significance based on Normal distribution – Test of Significance for single mean and difference of means, Test for single proportion and difference of proportions Test for single S.D & difference of S.D – Simple problems.

Unit – II

Small sample test based on 't' distribution assumptions application – Test of significance for single mean and difference of means, Paired 't' test, Test of significance of correlation coefficient- Partial correlation and Regression coefficient - simple problems.

Unit – III

F- Test - Application- Test for equality of population variances , Test for observed multiple correlation coefficient, observed sample correlation ratio, linearity of Regression.

Unit – IV

Chi-square test –Application- Test of significance based on population variance, test for goodness of fit and test for independence of attributes – simple problems.

Unit – V

Non-parametric tests – Definition, advantages and disadvantages – Run test, Median test, Sign test and Mann Witney U-test (One sample and two samples) – Simple problems.

Books for Study :

1. Fundamentals of mathematical statistics – S.C.Gupta & V.K.Kapoor, Sultan chand & sons, New Delhi, 11th thoroughly revised edition.
2. Statistical Methods- S.P. Gupta, Sultan chand & sons, New Delhi, 35th revised edition – 2007.

30 greater than

2m

Simple and Composite hypothesis:.

(7) A statistical hypothesis is some statement are ascertation about a population or equalently about the probability distribution characterising a population which we want to Verify on the basis of information available from a sample. If this statistical hypothesis specifies the population completely then it is termed as a **simple statistical hypothesis**, otherwise it is called a **Composite Statistical hypothesis**.

Test of Statistical Hypothesis:

Null hypothesis :- $H_0 : \mu = \mu_0$

2m

(8) The technique of randomisation used for the selection of sample units makes the test of significance valid for us, for applying the test of significance we first set up a hypothesis a definite statement about the population parameter such a hypothesis which is usually a hypothesis of no difference is called

Null hypothesis, and is usually a hypothesis denoted by H_0. According to prof. R.A. fishere null hypothesis is the hypothesis which is tested for possible rejection under the assumption that is true.

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 (say), i.e., $H_0 ; \mu = \mu_0$ then the alternative hypothesis could be,

i) $H_1 : \mu \neq \mu_0$ (i.e., $\mu > \mu_0$ or $\mu < \mu_0$)

ii) $H_1 : \mu > \mu_0$ iii) $H_1 : \mu < \mu_0$

The alternative hypothesis is (i) known as a two-tailed alternative and the alternative in (ii) are known as right-tailed and the left-tailed alternatives respectively.

Critical region :-

A region (corresponding to a statistic t) in the sample space S which amount to rejection of H_0 is termed as **critical region** of rejection. If w is the critical region and if $t = t(x_1, x_2, \dots x_n)$ is the value of the statistic based on a random sample of size n, then

value 5% or 1% level

$$P(\underset{\text{variable}}{t} \in \underset{\text{critical region}}{w} / H_0) = \alpha, \quad P(t \in \bar{w} / H_1) = \beta$$

where $\bar{w}$, the Complementary set of w, is called the acceptance region we have

$$w \cup \bar{w} = S \text{ and } w \cap \bar{w} = \phi.$$

Power of a Test :-

$1-\beta$ defined as $\int_w L_1\, dx = 1 - \int_{\bar{w}} L_1\, dx$ and $P(x \in W/H_1) = 1-\beta$ is called the power function of the test hypothesis H_0 against the alternative hypothesis H_1 the value of the power function and a parameter point is called the power of the test at that point.

Most Powerful test :-

Let us consider the problem of testing a simple hypothesis $H_0, \theta = \theta_0$ against a simple alternative hypothesis $H_1, \theta = \theta_1$.

Definition :-

The critical region ω is the most powerful critical region of size α. For testing $H_0 : \theta = \theta_0$ against $H_1 : \theta = \theta_1$ if

$P(x \in \omega / H_0) = \int_\omega L_0 \, dx = \alpha \rightarrow$ ① and

$P(x \in \omega / H_1) \geq P(x \in \omega_1 / H_1)$ for every other critical region ω_1 satisfying equation ①.

Procedure for Testing of hypothesis :-

Null hypothesis :-

Set up the Null hypothesis H_0.

Alternative hypothesis :-

Set up the alternative hypothesis H_1. This will enable us to decide whether we have to use a single-tailed (right or left) test or two-tailed test.

Level of Significance :-

Choose the appropriate level of significance (α) depending on the reliability of the estimation and permissible risk. This is to be decided before sample is drawn, i.e., α is fixed in advance.

Test Statistic (or test criterion):-

Compute the test statistic.

$$z = \frac{t - E(t)}{S.E(t)}, \text{ under } H_0.$$

Conclusion:-

We compare the computed value of z in step 4 with the significant value (tabulated value) z_α at the given level of significance 'α'.

If $|z| < z_\alpha$, i.e., if the calculated value of z (in modulus value) is less than z_α we say it is not significant. By this we mean that the difference $t - E(t)$ is just due to fluctuations of sampling and the sample data do not provide us sufficient evidence against the null hypothesis which may, therefore, be accepted.

If $|z| > z_\alpha$, i.e., if the computed value of test statistic is greater than the critical or significant value, then we say that it is significant and the null hypothesis is rejected at level of significant α, i.e., with confidence co-efficient $(1-\alpha)$.

Test of Significance of 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/n, i.e., $\bar{x} \sim N(\mu, \sigma^2/n)$.
However, this result holds, i.e., $\bar{x} \sim N(\mu, \sigma^2/n)$ even in random sampling from a non-normal population provided the sample size n is large. [C.L.T. Thus for large sample, the standard normal variate corresponding to $\bar{x}$ is z.

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

Under the null hypothesis H_0, that the sample has been drawn from a population with mean μ and variance σ^2, there ~~the~~ is ~~sample mean is distributed normally~~ no significant difference between the sample mean $(\bar{x})$ and population mean (μ) the test statistic (~~for~~) is

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

Remark:-

1. If the population s.d. σ is unknown, then we use its estimate provided by the

sample variance given by

$\hat{\sigma}^2 = s^2 \Rightarrow \hat{\sigma} = s$ (for large samples)

2. Confidence limits for μ. 95% confidence interval for μ is given by,

$|Z| \leq 1.96$, i.e., $\left|\frac{\bar{x}-\mu}{\sigma/\sqrt{n}}\right| \leq 1.96$.

$\Rightarrow \bar{x} - 1.96(\sigma/\sqrt{n}) \leq \mu \leq \bar{x} + 1.96(\sigma/\sqrt{n})$

and $\bar{x} \pm 1.96\,\sigma/\sqrt{n}$ are known as 95% confidence limits for μ. Similarly, 99% confidence limits for μ are $\bar{x} \pm 2.58\,\sigma/\sqrt{n}$ and 98% confidence limits for μ are $\bar{x} + 2.33\sigma/\sqrt{n}$.

However, in sampling from a finite population of size N, the corresponding 95% and 99% confidence limits for μ are respectively,

$\bar{x} \pm 1.96 \frac{\sigma}{\sqrt{n}} \sqrt{\frac{N-n}{N-1}}$ and $\bar{x} \pm 2.58 \frac{\sigma}{\sqrt{n}} \sqrt{\frac{N-n}{N-1}}$.

3. The confidence limits for any parameter (p, μ, etc.) are also known as its fiducial limits.

Problems :-

(1) 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 3.25 cms and s.d. 2.61 cms?

If the population is normal and its mean is unknown, find true mean.

Null hypothesis :- H_0 :

The sample has been drawn from the population with mean $\mu = 3.25$ cms and S.D. $\sigma = 2.61$ cms.

Alternative hypothesis : H_1 :-

$\mu \neq 3.25$ (Two-tailed).

Test statistic :-

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

$$= \frac{3.4 - 3.25}{2.61/\sqrt{900}}$$

$$= \frac{0.15}{0.087} = 1.7241.$$

Conclusion :-

Since $|Z| < 1.96$, we conclude that the data is null hypothesis H_0 is accepted at 5% level of significance.

(2) An insurance agent has claimed that the average age of policy hold as who insure though him less than the average of for all agent which is 30.5 years a random sample of 100 policy holds who had insure though him gave the following age distribution.

Age of last birthday : 16-20 21-25 26-30 31-35 36-40

Number of persons : 12 22 20 30 16

Calculate the arithmetic mean and standard deviation of this distribution and use these value to test this claim and has 5% of significance.

Class Interval	m	f	$d = \frac{m-A}{c}$	fd	d^2	fd^2
15.5–20.5	18	12	-2	-24	4	48
20.5–25.5	23	22	-1	-22	1	22
25.5–30.5	28	20	0	0	0	0
30.5–35.5	33	30	1	30	1	30
35.5–40.5	38	16	2	32	4	64
				16		164

$$\bar{x}, \mu = A + \frac{\Sigma fd}{N} \times c$$

$$= 28 + \frac{16}{100} \times 5$$

$$= 28 + 0.16 \times 5$$

$$= 28 + 0.8 = 28.8.$$

$$\therefore \sigma = \sqrt{\frac{\Sigma fd^2}{N} - \left(\frac{\Sigma fd}{N}\right)^2} \times C. = \sqrt{\frac{164}{100} - \left(\frac{16}{100}\right)^2}$$

$$= \sqrt{1.64 - \frac{256}{10000}} \times 5$$

$$= \sqrt{1.64 - 0.0256} = \sqrt{1.6144} \times 5$$

$$= 1.2705 \times 5$$

$$= 6.3529.$$

Null hypothesis :-

$\mu = 30.5$ and $\sigma = 6.3529$, no difference.

Alternative hypothesis :-

$\mu \neq 30.5$ years (one tail - left tailed test).

Test statistic :-

$$z = \frac{\bar{x} - \mu}{\sqrt{s^2/n}}$$

$$= \frac{28.8 - 30.5}{\sqrt{(6.35)^2/100}} = \frac{-1.7}{0.635}$$

$$= -2.68.$$

Conclusion :-

The insurance claim that the average age of policy holders who insure though him is less than the average for all agents is valid.

Test of Significance for difference of mean :-

Let $\bar{x}_1$ be the mean of a sample of size n, form 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 $\bar{x}_1 - \bar{x}_2$ is 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.

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

$$V(\bar{x}_1 - \bar{x}_2) = V(\bar{x}_1) + V(\bar{x}_2) = \frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}$$

the covariance curve vanishes,. hence the sample means $\bar{x}_1$ and $\bar{x}_2$ are independent.

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

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

Remarks :-

If $\sigma_1^2 = \sigma_2^2 = \sigma^2$; (i.e) if the samples have been drawn from the populations with common S.D, σ, then under $H_0: \mu_1 = \mu_2$,

$$Z = \frac{\bar{x}_1 - \bar{x}_2}{\sigma\sqrt{1/n_1 + 1/n_2}} \sim N(0,1) \ldots \text{①}$$

If Eqn ①, σ is not known, then its estimate based on the sample variance is used. If the sample sizes are not sufficient is large, then as unbiased estimate of σ^2 is given by,

$$\hat{\sigma}_2 = \frac{(n_1-1)S_1^2 + (n_2-1)S_2^2}{(n_1+n_2-2)}, \text{ Since}$$

$$E(\hat{\sigma^2}) = \frac{1}{n_1+n_2-2}\left\{(n_1-1)E(S_1^2) + (n_2-1)E(S_2^2)\right\}$$

$$= \frac{1}{n_1+n_2-2}\left[(n_1-1)\sigma^2 + (n-1)\sigma^2\right] = \sigma^2$$

But since sample sizes are large,

$S_1^2 \simeq s_1^2$, $S_2^2 \simeq s_2^2$, $n_1-1 \simeq n$, $n_2-1 \simeq n_2$.

Therefore in practice, for large samples, the following estimate of σ^2 without any serious error is used.

$$\hat{\sigma_2} = \frac{n_1S_1^2 + n_2 S_2^2}{n_1+n_2} \ldots$$

However, if sample sizes are small, then an exact sample test, t-test for difference of means (c.f) is to be used.

If $\sigma_1^2 \neq \sigma_2^2$ and σ_1 and σ_2 are not known, then they are estimated from sample values. This results in some error, which is practically immaterial if samples are large, these estimates for large samples are given by, $\hat{\sigma_1}^2 = S_1^2 \simeq s_1^2$ and $\hat{\sigma}_2^2 = S_2^2$

(Since samples are large).

In this case, gives

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

Sample problems :-

(1) The means of two single large samples of 1000 and 2000 members are 67.5 inches and 68 inches respectively. Can the samples be regarded as drawn from the same population of S.D. 2.5 inches?

↓ two tail test

(Test at 5% level of significance)

Null hypothesis :-

$\mu_1 = \mu_2$. The sample has been drawn from the same population.

Alternative hypothesis :-

$\mu_1 \neq \mu_2$ (two tailed test).

Test statistic :-

(1) of remark. Print

$$z = \frac{\bar{x}_1 - \bar{x}_2}{\sigma\sqrt{1/n_1 + 1/n_2}} \sim N(0,1)$$

$$= \frac{67.5 - 68}{2.5\sqrt{1/1000 + 1/2000}} = \frac{-0.5}{25\sqrt{0.001 + 0.0005}}$$

$$= \frac{-0.5}{2.5\sqrt{0.0015}}$$

$$= \frac{-0.5}{2.5 \times 0.0387} = \frac{-0.5}{0.0968}$$

$$= -5.163$$

Conclusion :-

The null hypothesis rejected that the samples are certainly non-same from the S.D 2.5.

Calculated value is greater than the tabulated value. So the null hypothesis is rejected.

2. In a survey of buying habit 400 women shopers are choosen at random in supermarket A located in a certain section of the city. The average weekly food expenditure is Rs. $\bar{x}_1$ 250 with a standard deviation of Rs. 40, σ_1 for 400 women shopers choosen at random in super market B in another certain section of the city. The average weekly food expenditure is Rs. 220 $\bar{x}_2$ with S.D is Rs. 35. σ_2 Test at 1% level of significance whether the average weekly food expenditure of the two population of shopers are equal.

Null Hypothesis :-

$\mu_1 = \mu_2$. i.e the average weekly food expenditure of the two populations of shopers are equal.

Alternative hypothesis :-

$\mu_1 \neq \mu_2$. (Two tailed test)

Test statistic :-

S.D nu kudutha S, use pannu

$$Z = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{s_1^2/n_1 + s_2^2/n_2}}$$

variance nu kudutha σ^2 use pannu

$$= \frac{250 - 220}{\sqrt{\frac{40^2}{400} + \frac{55^2}{400}}}$$

$$= \frac{30}{\sqrt{\frac{1600}{400} + \frac{3025}{400}}} = \frac{30}{\sqrt{4 + 7.5625}}$$

$$= \frac{30}{\sqrt{11.5625}} = \frac{30}{3.40036}$$

$= 8.8225$ ⇒ tabulate. 1% = 2.58

Conclusion :- $|Z| > Z_\alpha$

Calculated value is greater than tabulated value. So null hypothesis H_0 is rejected.

That the averages weekly expenditures of two population of shopers in markets A and B differ significantly.

(3) The average hourly wage of a sample of 150 workers in a plant A was Rs. 2.56 with a standard deviation of Rs. 1.08. The average hourly wage of a sample of 200 workers in a plant B was Rs. 2.87 with a standard deviation of Rs. 1.28. Can an applicant safely assure that the hourly wages paid by plant B are higher than those paid by plant A?

One tailed (right).

Null hypothesis :-

$\mu_2 = \mu_1$. There is no significant difference between the mean level of wages of workers in plant A and B.

Alternative hypothesis :-

$\mu_2 > \mu_1$ (one tailed → right tailed test)

Test statistic :-

$$Z = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{s_1^2/n_1 + s_2^2/n_2}}$$

$$= \frac{2.56 - 2.87}{\sqrt{\frac{(1.08)^2}{150} + \frac{(1.28)^2}{200}}}$$

That the averages weekly expenditures of two population of shopers in markets A and B differ significantly.

③ The average hourly wage of a sample of 150 workers in a plant A was Rs. 2.56 with a standard deviation of Rs. 1.08. The average hourly wage of a sample of 200 workers in a plant B was Rs. 2.87 with a standard deviation of Rs. 1.28. Can an applicant safely assure that the hourly wages paid by plant B are higher than those paid by plant A?

One tailed (right).

Null hypothesis :-

$\mu_2 = \mu_1$. There is no significant difference between the mean level of wages of workers in plant A and B.

Alternative hypothesis :-

$\mu_2 > \mu_1$ (one tailed → Right tailed test)

Test statistic :-

$$Z = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{s_1^2/n_1 + s_2^2/n_2}}$$

$$= \frac{2.56 - 2.87}{\sqrt{\frac{(1.08)^2}{150} + \frac{(1.28)^2}{200}}}$$

$$= \frac{-0.31}{\sqrt{\frac{1.1664}{180} + \frac{1.6384}{200}}}$$

$$= \frac{-0.31}{\sqrt{0.0077 + 0.0081}}$$

$$= \frac{-0.31}{\sqrt{0.0158}} = \frac{-0.31}{0.1256}$$

$$= -2.468.$$

Conclusion :- $|z| > z_\alpha$.

Calulated value is greater than the tabulated value. So null hypothesis H_0 is rejected.

Test of Significance of Single proportion.

If x is the number of success in n independent trials with constant probability P of (success) for each trial.-

$E(x) = nP$ and $V(x) = nPQ$. where $Q = 1-P$, is the probability of failure. It has been proved that for large n, $X \sim N(nP, nPQ)$ i.e the binomial distribution tends to normal distribution.

Hence for large n;

$x \sim N(nP, nPQ)$ i.e,

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

Remark!-

In a sample of size n, Let x be the number of persons possessing the given attributes. Then,

Observed proportion of success $= X/n$
$= P_1$ (say)

$\therefore E(P) = E(X/n) = 1/n\, E(x) = 1/n\, nP = P.$

Thus, the sample proportion p gives

an unbiased estimate of the population proportion (P).

Also,

$$v(P) = V(x/n) = 1/n^2 V(x) = 1/n^2 \, nPQ = \frac{PQ}{n}$$

$$\Rightarrow S.E(P) = \sqrt{PQ/n}$$

Since, x and consequently x/n is asymptotically normal for large n, the normal test for the proportion of success becomes

$$z = \frac{p - E(p)}{S.E(p)} = \frac{p - P}{\sqrt{PQ/n}} \sim N(0,1).$$

2. If we have sampling from a finite population of size N, then

$$S.E(P) = \sqrt{\left(\frac{N-n}{N-1}\right)\frac{PQ}{n}}$$

3. Since the probable limits for a normal variate X are $E(x) \pm 3\sqrt{V(x)}$, the probable limits for the observed proportion of success are:-

$$E(p) \pm 3S.E.(p), \text{ i.e., } P \pm 3\sqrt{PQ/n}.$$

If P is not known then taking p (the sample proportion) as a estimate of p, the probable limits for the proportion in the population are: $p \pm 3\sqrt{pq/n}$.

However, the limits for p at level of significance α are given by:

$$p \pm z_\alpha \sqrt{pq/n}$$

where z_α is the significant value of z at level of significance α.

In particular: 95% confidence (two tail test) limits for P (capital P) are given by; $P \pm 1.96\sqrt{PQ/n}$ and: 99% of confidence limits for P are given by; $P \pm 2.58\sqrt{PQ/n}$.

Simple problems:-

① In a sample of thousand (n) people in Maharashtra 540 (X) are rice eaters and rest are wheat eaters. Can be assume that both rice and wheat are equally (Prob = $1/2 = .5$) popular in this state, at 1% level of significance?

$$X = 540, \; n = 1000$$

$$p = X/N = \frac{540}{1000} = 0.54.$$

(small p)

X = number of rice eaters

$n = 1000$; p = sample proportions of rice eaters.

Null hypothesis:.

$H_0, P = 0.5$

Both rice and wheat are equally popular in this state. So that P = population proportion of rice eaters in Maharashtra = 0.5.

$Q = 1 - P = 1 - 0.5$

$Q = 0.5.$

Alternative hypothesis:-

$P \neq 0.5$ (two tail test)

test statistic:-

$$z = \frac{p - P}{\sqrt{PQ/n}} \sim N(0,1)$$

$$= \frac{0.54 - 0.5}{\sqrt{\frac{0.5 \times 0.5}{1000}}}$$

$$= \frac{0.04}{\sqrt{\frac{0.25}{1000}}} = \frac{0.04}{\sqrt{0.0002}}$$

$$= \frac{0.04}{0.0158} \simeq 2.529.$$

$2.53 < 2.58.$

Conclusion :-

$|Z| < Z\alpha$.

∴ Calculated value is less than the tabulated value. So null hypothesis is accepted.

∴ Rice and wheat are equally popular in Maharashtra state.

② A random sample of 500 (n) apples was taken from a large consignment and 60 (x) were found to be bad. Obtain the 98 % confidence limits for the percentage of bad apples in the consignment.

$P \pm 2.3\sqrt{PQ/n}$

$x = 60,\ n = 500$

∴ $p = x/n = \frac{60}{500} = 0.12$.

$q = 1 - p = 1 - 0.12$

$q = 0.88$.

Null hypothesis :-

$H_0 : P = 0.12$.

Alternative hypothesis :-

$H_1 : P \neq 0.12$.

(Two tailed test)

Test Statistic :-

$$z = P \pm 2.33\sqrt{PQ/n}$$

$$= 0.12 \pm 2.33\sqrt{\frac{(0.12)(0.88)}{500}}$$

$$= 0.12 \pm 2.33(0.01453)$$

$$= 0.12 \pm 0.03386$$

$$= 0.1538 \text{ (or) } + 0.086$$

$$= 15.38 \text{ (or) } 8.6 .$$

Conclusion :- $Z\alpha = 2.33$, $|Z| > Z\alpha$.

Calulated value is greater than the tabulated vallue.

So the null hypothesis H_0 is ~~accepte~~ rejected.

③ 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 it is more at 5% level.

$$p = x/n = 18/20 = 0.9, \quad q = 1-p = 1-0.9 = 0.15$$

small

$$P = 85\% = 85/100 = 0.85$$

Null hypothesis:-

$H_0 : P = 0.85$, the proportion of person survived after attack by a disease is 85%.

Alternative hypothesis:-

$H_1 : P > 0.85$ (one tail → right tailed test).

Test Statistic:-

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

$$= \frac{0.9 - 0.85}{\sqrt{\frac{(0.85) \times (0.15)}{20}}}$$

$$= \frac{0.05}{\sqrt{\frac{0.1275}{20}}} = \frac{0.05}{\sqrt{0.006375}}$$

$$= \frac{0.05}{0.07984}$$

$$= 0.6262$$

$5\% = z_\alpha = 1.645$

Conclusion :-

Calculated value is less than the tabulated value. So null hypothesis is 5% of significance value is accepted.

Test of significance for Difference of proportion :-

Suppose we want to compare two distinct populations with respect to the prevalence of a certain attributes, 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 :

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

(small p_1, small p_2)

If P_1 and P_2 are population proportions then

(capital P_1, capital P_2)

$$E(p_1) = P_1, \quad E(p_2) = P_2$$

and $V(p_1) = \frac{P_1 Q_1}{n_1}$ and

$$v(P_2) = \frac{P_2 Q_2}{n_2}.$$

Since for large samples P_1 and P_2 are independently and asymptotically normally distributed, ($P_1 - P_2$) is also normally distributed. Then the standard variable corresponding to the difference $(P_1 - P_2)$ is given by:

small

$$Z = \frac{(P_1 - P_2) - E(P_1 - P_2)}{\sqrt{v(P_1 - P_2)}} \sim N(0, 1).$$

Under the null hypothesis $H_0: P_1 = P_2$ i.e., there is no significant difference between the sample proportions, we have

capital P

$$E(P_1 - P_2) = E(P_1) - E(P_2) = P_1 - P_2 = 0.$$

[under H_0]

Also $v(P_1 - P_2) = v(P_1) + v(P_2)$,

the co-variance term $cov(P_1, P_2)$ vanishes, since sample proportions are independent.

$$\Rightarrow v(P_1 - P_2) = \frac{P_1 Q_1}{n_1} + \frac{P_2 Q_2}{n_2}$$

$$= PQ(1/n_1 + 1/n_2)$$

[under H_0

Hence, under $H_0: P_1 = P_2$ the test statistic for the difference of proportions becomes:-

$$Z = \frac{p_1 - p_2}{\sqrt{PQ\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}} \sim N(0,1) \quad \text{(sample proportion)}$$

In general we do not have any information as to the proportion of A's in the populations from which the samples have been taken. Under $H_0: P_1 = P_2 = P$ (say), an unbiased estimate of the population proportion P (capital P), based on both the samples is given by

$$\hat{P} = \frac{n_1 p_1 + n_2 p_2}{n_1 + n_2} = \frac{X_1 + X_2}{n_1 + n_2}.$$

The estimate is unbiased, since

$$E(\hat{P}) = \frac{1}{n_1 + n_2} E\left[n_1 p_1 + n_2 p_2\right]$$

$$= \frac{1}{n_1 + n_2}\left[n_1 E(p_1) + n_2 E(p_2)\right]$$

$$= \frac{1}{n_1 + n_2}(n_1 P_1 + n_2 P_2) = P.$$

$[\because P_1 = P_2 = P$, under $H_0]$

Remark :-

Suppose we want to test the significance of the difference between P_1 and P_2 where $P = \frac{n_1 P_1 + n_2 P_2}{n_1 + n_2}$

gives pooled estimate of the population proportion on the basis of both the samples we have,

$$V(P_1 - P_2) = V(P_1) + V(P_2) - 2\,cov(P_1, P_2).$$

Since P_1 and P_2 are not independent $Cov(P_1, P_2) \neq P_2$.

$$cov(P_1, P_2) = E[\{P_1 - E(P_1)\}\{P_2 - E(P_2)\}]$$

$$= E[\{P_1 - E(P_1)\}\{1/n_1+n_2 \{n_1P_1 + n_2P_2 - E(n_1P_1 + n_2P_2)\}].$$

$$= \frac{1}{n_1+n_2} E[\{P_1 - E(P_1)\}\{n_1(P_1 - E(P_1)) + n_2(P_2 - E(P_2))\}]$$

$$= \frac{1}{n_1+n_2}[n_1 E\{P_1 - E(P_1)\}^2 + n_2 E\{(P_1 - E(P_1)\}\{P_2 - E(P_2)\}]$$

$$= \frac{1}{n_1+n_2}[n_1 V(P_1) + n_2\, cov(P_1, P_2)]$$

$$= \frac{1}{n_1+n_2} n_1 V(P_1). \qquad [\because Cov(P_1, P_2) = 0]$$

$$= \frac{n_1}{n_1+n_2} \cdot \frac{Pq}{n_1} = \frac{Pq}{n_1+n_2} \quad \left[\because \hat{P} = P, \hat{Q} = q\right]$$

$$var(p) = var\left[\frac{n_1 p_1 + n_2 p_2}{n_1+n_2}\right] = \frac{1}{(n_1+n_2)^2} var(n_1 p_1 + n_2 p_2)$$

$$= \frac{1}{(n_1+n_2)^2}\left[n_1^2 \, var(p_1) + n_2^2 \, var(p_2)\right]$$

Cov. variance term vanishes since P_1 and P_2 are independent.

$$\therefore var(\hat{p}) = \frac{1}{(n_1+n_2)^2}\left(n_1^2 \cdot \frac{Pq}{n_1} + n_2^2 \frac{Pq}{n_2}\right)$$

$$Var(p) = \frac{Pq}{n_1+n_2}.$$

Substituting (4) and simplifying, we shall get,

$$V(P_1 - P_2) = \frac{Pq}{n_1} + \frac{Pq}{n_1+n_2} - 2\frac{Pq}{n_1+n_2}.$$

$$= Pq\left[\frac{n_2}{n(n_1+n_2)}\right].$$

Also, $E(P_1 - P_2)$.

$$= E(P_1) - E(P_2) = P - P = 0$$

Thus the test statistic in this case becomes :-

$$Z = \frac{(P_1 - P_2) - E(P_1 - P_2)}{S.E(P_1 - P_2)}.$$

$$= \frac{P_1 - P_2}{\sqrt{\left\{\frac{n_2}{(n_1+n_2)} \cdot \frac{PQ}{n_1}\right\}}} \sim N(0,1).$$

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

We have seen that in the usually notations:

$$Z = \frac{(P_1 - P_2) - E(P_1 - P_2)}{S.E(P_1 - P_2)}.$$

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

Here sample proportions are not given if we set up the null hypothesis $H_0: P_1 = P_2$ (i.e) the sample will not reveal the difference in the population proportion or in other words, the difference in population proportions is likely to the hidden in sampling.

the test statistic becomes,

$$Z = \frac{|p_1 - p_2|}{\sqrt{\left(\frac{P_1 Q_1}{n_1} + \frac{P_2 Q_2}{n_2}\right)}} \sim N(0,1)$$

Simple Problems :-

1) Random samples of 400 men and 600 women where asked whether they would like to have a flyover near the residence. 200 men and 325 women where in favour of the proposal. Test the hypothesis that proportions of men and women in favour of the proposal are same, against that they are not at 5 % level. $x_1 = 200$, $x_2 = 325$, $n_1 = 400$, $n_2 = 600$

$$p_1 = x_1/n_1 = \frac{200}{400}$$

$$p_1 = 0.5$$

$$p_2 = x_2/n_2 = \frac{325}{600} = 0.54.$$

x_1 = number of men favour the proposal.
x_2 = number of women favour the proposal.

Null hypothesis :

$H_0 : P_1 = P_2$, there is no significant difference between the opinion of men and women as per as proposal of flyover is consistent.

Alternative hypothesis :

$P_1 \neq P_2$. (Two tail test.

x_1 = number of men favour the proposal = 200

x_2 = number of women favour the proposal = 325.

~~p_1 =~~

~~p_2 =~~

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

$$\hat{P} = \frac{x_1 + x_2}{n_1+n_2}$$

$$= \frac{200+325}{1000} = 0.525$$

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

$$= 1 - 0.525 = 0.475$$

$$z = \frac{0.5 - 0.54}{\sqrt{0.525 \times 0.475\left(\frac{1}{400}+\frac{1}{600}\right)}} = \frac{-0.04}{\sqrt{0.2493(0.025 + 0.0166)}}$$

$$= \frac{-0.04}{\sqrt{0.2493(0.0416)}}$$

$$= \frac{-0.04}{\sqrt{0.01037}} = \frac{-0.04}{0.0319}$$

$$= -1.24 . \quad z\alpha = 1.96$$

Conclusion:-

$|x| < |z_\alpha|$.

Calculated value is less than the tabulated value. Null hypothesis Ho accepted.

Hence Ho may be accepted at 5% level of significance and we may conclude that men and women do not differ significantly as regards proposal of flyover & concern.

2) In a large city A, 20% of a random sample of 900 school children ~~20%~~ had a defective eye sign in other large city B 15% of random sample of 1600 children had the same defect. Is this difference between the ~~s~~ two proportion significant?

Null hypothesis:-

$$P_1 = P_2.$$

Alternative hypothesis:-

$$P_1 \neq P_2 \text{ (two tailed test)}.$$

Test Statistic:-

$$P_1 = \frac{20}{100} = 0.2 \quad P_2 = \frac{15}{100} = 0.15$$

$$n_1 = 900, \quad n_2 = 1600$$

$$\hat{P} = \frac{n_1 P_1 + n_2 P_2}{n_1 + n_2}$$

$$= \frac{900(0.2) + 1600(0.15)}{900 + 1600}$$

$$= \frac{180 + 240}{2500}$$

$$= \frac{420}{2500}$$

$$\hat{P} = 0.168.$$

Conclusion :-

$z < z_\alpha$,

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

$$= 1 - 0.168$$

$$\hat{Q} = 0.832.$$

$$Z = \frac{P_1 - P_2}{\sqrt{\hat{P}\hat{Q}\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}}$$

$$= \frac{(0.2) - (0.15)}{\sqrt{(0.168)(0.832)\left(\frac{1}{900} + \frac{1}{1600}\right)}}$$

$$= \frac{0.05}{\sqrt{0.1397(0.0011 + 0.000625)}}$$

$$= \frac{0.05}{\sqrt{0.1397\,(0.001735)}}$$

$$= \frac{0.05}{\sqrt{0.0002425}}$$

$$= \frac{0.05}{0.01556} = 3.211.$$

$z = 3.211$ $\quad z_\alpha = 1.96 \rightarrow$ 5 % level of Significance.

Conclusion :- $z > z_\alpha$

Calulated value of z is greater than the tabulated value z_α. So the null hypothesis H_0 is rejected at 5 % level of sig. Difference between the two proportion is significant.

UNIT - II

Student's 't' distribution :- (Small Sample)

Let x_i, $i = 1, 2, \ldots n$ be a random sample of size n from a normal population with mean μ and variance σ^2 then students 't' defined by the statistics.

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

wher $\overline{x} = \frac{1}{n}\sum_{i=1}^{n} x_i$ is the sample mean and $S^2 = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \overline{x})^2$ is on unbiased estimates of the population variance σ^2 and it is follows student's 't' distribution with $\nu = (n-1)$ d.f with probability density function.

Applications of 't' distribution :-

To test if the sample mean $(\overline{x})$ differs significantly from the hypothetical value μ of the population mean.

To test the significance of the difference between two sample means.

To test the significance of an observed sample correlation co-efficient and sample regression co-efficient.

To test the significance of observed partial correlation co-efficient.

't' test for single mean:

* If a random sample x_i, $i=1,2,\dots,n$ of size n has been drawn from a normal population with a specified mean says μ_0 or if the sample mean differs significantly from the hypothetical value μ_0 of the population mean.

* Under the null-hypothesis H_0:

* The sample has been drawn from the population with mean μ_0 (or)...

* There is no significant difference between the sample mean $\bar{x}$ and the population mean μ_0.

$$t = \frac{\bar{x} - \mu}{S^2/\sqrt{n}} \sim N(0,1)$$

where $\bar{x} = 1/n \sum_{i=1}^{n} x_i$ is the sample mean and $S^2 = 1/n-1 \sum_{i=1}^{n} (x_i - \bar{x})$;

follows students 't' distribution with $(n-1)$ d.f.

Conclusion:

We know compare the calculated value of 't' with the tabulated value at certain level of significance if calculated (modulus of) $|t|$ greater than tabulated 't', null hypothesis is rejected and if calculated ~~modulus~~ of $|t| <$ tabulated 't' null hypothesis may be accepted the level of significance is adopted.

Sample ~~mean~~ formula

1) A mechanics is making in engine parts with ~~axel~~ diameters of 0.700 inches a random sample of 10 parts shows a mean diameters of ~~d max~~ 0.742 inches with a standard deviation of 0.040. Compute the statistics you would use to test whether the work is meeting the specifications also state how you would proceed further.

Null hypothesis:

$\mu = 0.700$. That is the product is conforming to specification.

Alternative hypothesis:-

H_1: $\mu \neq 0.700$ (Two tailed test)

Test Statistic:-

$$t = \frac{\bar{x} - \mu}{\sqrt{s^2/n-1}} \sim N(0,1)$$

$\bar{x} = 0.742$, $\mu = 0.7$, $s = 0.040$.

$$t = \frac{0.742 - 0.7}{\sqrt{\frac{(0.040)^2}{10-1}}}$$

$$= \frac{0.042}{\sqrt{0.00016}}$$

$$= \frac{0.042}{0.01264}$$

$\nu = (n-1)$ df
$= 9$ df.

$t = 3.315$ $5\% = 2.26$.

$3.315 > 2.26$.

Conclusion:-

Calculated value is greater than the tabulated value, so null hypothesis 5% d.f is rejected. That the product is not meeting the specification.

Remark 't' test.

(3) Assumptions for Student's 't' dis

The following assumptions are made 't' test.

The parent population from which the sample is drawn normal.

The sample observations are independent (i.e) the sample is random.

The population standard deviation σ is unknown.

2) The mean weekly sales of soap bars in department stores was 146.3 bars per store after an advertising combine the mean weekly sales in 22 stores for a typical week increased to 153.7 and showed a standard deviation of 17.2. Was the advertisement compaing successful.

Null hypothesis:-

$H_0 : \mu = 146.3$, the advertising comping is not successful.

Alternative hypothesis:-

$H_1 : \mu > 146.3$. (One tail → right tail)

Test Statistic :-

$$t = \frac{\bar{X} - \mu}{\sqrt{S/_{n-1}}}$$

$$= \frac{153.7 - 146.3}{\sqrt{(17.2)^2/_{22-1}}}$$

$$= \frac{7.4}{\sqrt{295.84/_{21}}}$$

$$= \frac{7.4}{\sqrt{14.0876}}$$

$$= \frac{7.4}{3.7533}$$

$$t = 1.970$$

$\nu = (n-1)$ d.f
$= (22-1)$
$= 21$ df
$t = 2.08$

Conclusion :-

Calulated value is less than the tabulated value. So null hypothesis is H_0 is accepted.

t-test for difference of mean :-

Suppose we want to test if two independent samples x_i $(i = 1, 2, \ldots n_1)$ and y_j $(j = 1, 2, \ldots n_2)$ of sizes n_1 and n_2 have been drawn from two normal populations with mean μ_x and μ_y respectively. Under the null hypothesis (H_0) that the samples have been drawn from the normal populations with means μ_x and μ_y respectively and under the assumption that the population variance are equal, (i.e) $\sigma_x^2 = \sigma_y^2 = \sigma^2$ (say), the statistic

$$t = \frac{(\bar{x} - \bar{y}) - (\mu_x - \mu_y)}{S\sqrt{\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}}$$

where,

$$\bar{x} = \frac{1}{n_1}\sum_{i=1}^{n_1} x_i, \quad \bar{y} = \frac{1}{n_2}\sum_{j=1}^{n_2} y_j$$

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

is an unbiased estimate of the common population variance σ^2, follows Student's 't' distribution with ~~$\left(\frac{n_1 + n_2}{2}\right)$ d.f~~ $(n_1 + n_2 - 2)$ d.f.

Remarks :-

S^2, defined in equation ① is an unbiased estimate of the common population variance σ^2, since

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

$$= \frac{1}{n_1+n_2-2} E\left[(n_1-1)S_x^2 + (n_2-1)S_y^2\right]$$

$$= \frac{1}{n_1+n_2-2} \left[(n_1-1)E(S_x^2) + (n_2-1)E(S_y^2)\right]$$

$$= \frac{1}{n_1+n_2-2} \left[(n_1-1)\sigma^2 + (n_2-1)\sigma^2\right] = \sigma^2.$$

An important deduction which is of much practical utility is discussed below.

Suppose we want to test if,

a) Two independent samples x_i ($i=1,2,\dots n_1$) and y_j ($j=1,2,\dots n_2$) have been drawn from the populations with same means, or

(b) the two sample means $\bar{x}$ and $\bar{y}$ differ significantly or not.

Under the null hypothesis H_0,

where

a) That sample have been drawn from the populations with same means, (i.e) $\mu_x = \mu_y$ (or) (b) the sample means $\bar{x}$ and $\bar{y}$ do not differ significantly, the statistic

$$t = \frac{\bar{x} - \bar{y}}{S\sqrt{\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}} \quad \left[\because \mu_x = \mu_y \text{ under } H_0\right]$$

Where symbols are defined in equation (1), follows student's 't' distribution with $(n_1 + n_2 - 2)$ d.f.

On the assumption of 't' test for difference of means. Here we make the following three fundamental assumptions,

* Parent populations, from which the sample have been drawn are normally distributed.

* The population variances are equal and unknown, (i.e) $\sigma_x^2 = \sigma_y^2 = \sigma^2$ (say), where σ^2 is unknown.

* The two samples are random and independent of each other.

Thus before applying 't' test for testing the equality of means it is theoretically desirable to test the.

test the equality of population variances by applying ~~f~~ 'F' test. If the variances do not come out to be equal then 't' test becomes invalid. and in that case behrens 't' test of fiducial interval is thus used. For practical problems. However, the assumptions (1) (2) are taken for granted.

Simple problems ::

1) Samples of two types of electric light bulb where tested for length of life and following data were obtained..

	TYPE I	TYPE II
Sample Number	$n_1 = 8$	$n_2 = 7$.
Sample Mean	$\bar{X}_1 = 1234$ hrs	$\bar{X}_2 = 1036$ hrs
Sample Standard deviation	$S_1 = 36$ hrs	$S_2 = 40$ hrs

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

Null hypothesis :-

$H_0 : \mu_1 = \mu_2$.

Alternative hypothesis :-

$H_1 : \mu_1 > \mu_2$. (one tail → right tailed test).

Test statistic :-

$$t = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{s^2(1/n_1 + 1/n_2)}} \sim t_{(n_1+n_2-2)} \text{ d.f.}$$

$$S^2 = \frac{1}{n_1+n_2-2}\left[n_1 s_1^2 + n_2 s_2^2\right]$$

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

$$= \frac{1}{13}\left[10368 + 11200\right]$$

$$= 1/13\left[21568\right]$$

$$S^2 = 1659.076$$

$$t = \frac{1234 - 1036}{\sqrt{1659(1/8 + 1/7)}}$$

$$= \frac{198}{\sqrt{1659 \cdot (0.125 + 0.1428)}}$$

$$= \frac{198}{\sqrt{1659(0.2678)}}$$

$$= \frac{198}{\sqrt{444.298}}$$

$$= \frac{198}{21.078}$$

$$\therefore t = 9.3936.$$

d.f $\Rightarrow$

$= n_1+n_2-2$

$= 8+7-2$

$= 13$ df

$= 1.77$

Conclusion :-

Calculated value is greater than the tabulated value. So the null hypothesis is Rejected at 5% level of Significance.

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

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

Test if the two diets differ significantly as regards there affect on increasing in weight.

$$\bar{X}_1 = \frac{\Sigma x}{n} = \frac{336}{12} = 28.$$

$$\bar{X}_2 = \frac{\Sigma y}{n} = \frac{450}{15} = 30.$$

Null hypothesis:

$$H_0 : \mu_1 = \mu_2.$$

Alternative hypothesis:

$$H_1 : \mu_1 > \mu_2$$ (One tailed - right tail test)

Test Statistic:

X	Y	$(X-\bar{X})$	$(Y-\bar{Y})$	$(X-\bar{X})^2$	$(Y-\bar{Y})^2$
25	44	-3	14	9	196
32	34	4	4	16	16
30	22	2	-8	4	64
34	10	6	-20	36	400
24	47	-4	17	16	289
14	31	-14	1	196	1
32	40	4	10	16	100
24	30	-4	0	16	0
30	32	2	2	4	4
31	35	3	5	9	25
35	18	7	-12	49	144
25	21	-3	-9	9	81
	35		-5		25
	29		-1	380	1
	22		-8		64

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

$$= \frac{1}{12+15-2}\left[380 + 1410\right]$$

$$= \frac{1}{25}\left[1790\right]$$

$$S^2 = 71.6 .$$

$$t = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{S^2\left(1/n_1 + 1/n_2\right)}} \sim t(n_1+n_2-2)$$

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

$$= \frac{-2}{\sqrt{71.6(0.15)}} = \frac{-2}{\sqrt{10.74}}$$

$$= \frac{-2}{3.2771} = -0.61026$$

d.f = 2.06

Conclusion :-

Calculated value is less than the tabulated value. So the null hypothesis H_0 is accepted at level of significance.

Paired 't' test for difference of means:

Let us now consider the case when
1. the sample sizes are equal. (i.e) $n_1 = n_2 = n$ (say), and
2. the two samples are not independent but the sample observations are paired together, (i.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 are not. for
3. For example, suppose we want to test the efficacy of a particular drug (say), for inducing sleep.
4. Let x_i and y_i $(i = 1, 2, \ldots n)$ be the readings, in hours of sleep, on the ith individual, before and after the drug is given respectively.
5. Here instead of applying the difference of the means test discussed in 't' test, we apply the paired 't' test given below.

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. (i.e) the drug is not responsible for these increments, the statistic t

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

where $\bar{d} = 1/n \sum_{i=1}^{n} d_i$ and

$$S^2 = \frac{1}{n-2} \sum_{i=1}^{n} (d_i - \bar{d})^2.$$

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

Simple problems :-

(1) A certain stimulus administered to each of the 12 patients resulted in the following increase of blood pressure 5, 2, 8, -1, 3, 0, -2, 1, 5, 0, 4 and 6 can it be concluded that the stimulus will in general, be accompined by an increase in blood pressure?

Alternate

Null hypothesis:

$$H_0 : \mu_x = \mu_y .$$

Alternative hypothesis:

$$H_1 : \mu_x < \mu_y .$$

Test Statistic:

d	d^2
5	25
2	4
8	64
-1	1
3	9
0	0
-2	4
1	1
5	25
0	0
4	16
6	36
31	185

$$\bar{d} = \frac{\Sigma d}{n} = \frac{31}{12} = 2.583 .$$

$$s^2 = \frac{1}{n-1}\left[\Sigma d^2 - \frac{(\Sigma d)^2}{n}\right].$$

2 In a certain experiment to compare two types of animal food's A and B. The following result of increase weights were observed in animals.

~~Food~~

Food A	49	53	51	52	47	50	52	53	407
Food B	52	55	52	53	50	54	54	53	423

i) Assuming that the two samples of animals are independent. Can we conclude that food B is better than food A.

ii) Also examining ~~in~~ these case ~~of~~ when the same set of 8 animals were used in both the foods.

Null hypothesis:-

$$H_0 : \mu_x = \mu_y$$

Alternative hypothesis:-

$$H_1 : \mu_x < \mu_y .$$

Test Statistic:-

$$t = \frac{\bar{d}}{\sqrt{s^2/n}} .$$

x	y	d = (x - y)	d^2
49	52	-3	9
53	55	-2	4
51	52	-1	1
52	53	-1	1
47	50	-3	9
50	54	-4	16
52	54	-2	4
53	53	0	0
		-16	44.

$$\bar{d} = \frac{\Sigma d}{n} = \frac{-16}{8} = -2.$$

$$S^2 = \frac{1}{n-1}\left[\Sigma d^2 - \frac{(\Sigma d)^2}{n}\right]$$

$$= \frac{1}{8-1}\left[44 - \frac{(16)^2}{8}\right]$$

$$= \frac{1}{7}\left[44 - \frac{256}{8}\right] = \frac{1}{7}\left[44 - 32\right]$$

$$= \frac{1}{7}\left[12\right]$$

$$S^2 = 1.714.$$

$$t = \frac{\bar{d}}{\sqrt{s^2/n}}$$

$$= \frac{|-2|}{\sqrt{1.714/8}} = \frac{2}{\sqrt{1.714/8}}$$

$$= \frac{2}{\sqrt{0.2142}} = \frac{2}{0.4628}$$

$$t = 4.3208.$$

Conclusion:-

Calculated value is greater than the tabulated value. So the null hypothesis H_0 is rejected at $(n-1)$ df.

"t" test for testing the significance of observed sample correlation co-efficient:-

If r is the observed correlation co-efficient 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 co-efficient is zero.

the Statistic,

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

follows Studentnts 't' distribution with $(n-2)$ df.

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

If the 't' comes to be non-significant then H_0 may be accepted and we conclude that variables may be regarded as uncorrelated in the population.

Simple problems :-

1) 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 ~~greatest~~ latest last value of r in a sample of 18 pairs of observation from a bivariate normal population significant at 5% level of significance.

a) Null hypothesis :-

$H_0 : \rho = 0$, (i.e) the observed sam. correlation co-efficient is not significant of any correlation in the population.

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

$$= \frac{0.6\sqrt{(27-2)}}{\sqrt{1-(0.6)^2}}$$

$$= \frac{0.6\sqrt{25}}{\sqrt{1-0.36}}$$

$$= \frac{0.6 \times 5}{\sqrt{0.64}} = \frac{0.6 \times 5}{0.8}$$

$$t = 3/0.8 = 3.75.$$

b) $H_0 : \rho = 0$, (i.e)

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

$2.21 =$

2) A coefficient of correlation of 0.2 is derived from a random sample of 625 pairs of observations. Is this value of r significant?

Null hypothesis:-

$$\rho = 0$$

Alternative hypothesis:-

$$\rho \neq 0$$

Test Statistic:-

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

$$t = \frac{0.2}{\sqrt{1-(0.2)^2}} \sqrt{625-2}$$

$$= \frac{0.2}{\sqrt{1-0.04}} \sqrt{623}$$

$$= \frac{0.2}{\sqrt{0.96}} \times 24.95$$

$$= \frac{0.2 \times 24.95}{0.9797}$$

$$= \frac{4.99}{0.9797} = 5.09.$$

Conclusion :-

calulated value is greater than tabulated value. So Ho is rejected.