Sem IV Unit-IV

Non - Parametric tests

(1) Def :- A statistical test which does not depend on the particular form of frequency function from which the samples are drawn and not related with estimating the parameter is called Non - parametric test.

Non - Parametric tests does not make any assumption regarding the form of the population.

Assumptions of Non Parametric tests :-

1) Sample observations are independent.
2) The variable under study is continuous.
3) P.d.f is continious.
4) Lower order moments exists.

Advantage of Non - Parametric tests :-

1) NP methods are very simple and easy to apply.
2) NP methods do not require much complicated sampling theory.
3) No assumption is made about the form of the frequency function of the parent population from which the samples have drawn.
4) Generally the socio - economic data are not normally distributed, NP methods have found

applications in psychometry, sociology and educational statistics.

5) If the data is given in ranks or signs or grades (like A, A^+, B, B^+ etc) then NP methods are more useful and easy to apply. But parametric methods cannot apply.

6) No parametric technique will apply to the data which are measured in nominal scale, while NP methods exist to deal with such data.

Disadvantage of Non-Parametric tests:—

1) If the measurements are nominal or ordinal, we can apply NP methods. Even in this case, if a parametric test exists it is more powerful than NP test.

2) NP tests are designed to test statistical hypothesis only and not for estimating parameters.

3) So far, no NP method exist for testing interactions in ANOVA unless special assumptions about the additivity of the model are made.

2. Parametric tests Vs Non Parametric tests :-

Parametric test.	Non parametric test
1) The sample observations are independent.	1) The sample observations are independent.
2) Variable in the study need not be continious. (may be discrete or continious)	2) variable in the study is continious.
3) It depends on the form of the frequency function from which the samples have drawn.	3) It does not depend on any particular form of frequency function from which the samples are drawn.
4) It is based on the assumption that parent population is assumed to be known (or normal).	4) Non parametric tests do not have any assumption regarding the parent population.
5) In this method, we can estimate the population parameters which are not known.	5) In this method, we can not estimate the population parameters.
6) If the measurements are nominal or ordinal, we cannot apply parametric methods.	6) If the measurements are nominal or ordinal, Non parametric methods can be applied.
7) This method can not be applied for data in ranks and grades.	7) This method can be applied for data in ranks and grades.

Parametric tests	Non parametric tests
8) Mean and variance are more general statistical measures used in this method	8) In this method, median is usual measure for testing the hypothesis.
9) To test two independent samples we use t-test.	9) To Test two independent sample we use run or Mann-whitney u-tes
10) To test dependent samples or two samples of same size we apply t-test.	10) To test the paired samples (with equal size) we can apply sign test or Wilcoxon signed rank test
11) These methods are more efficient or greater power.	11) These methods are less efficient and lest power.
12) These methods are strong.	12) These methods are weak.
13) In this method, we can draw more conclusions.	13) In this method, we can not draw more conclusions.

(3) Measurment scales :-

Measurement is the process of mapping or assigning numbers to objects or observations in such a way that the numbers are amenable to analyses by manipulation or operation according to certain rules. S.S. Stevens (1946) defines measurement as the assignment of numerals to objects or events according to rules, and points out that different rules leads to rules and points out that different kinds of scales and different kinds of measurements.

There are 4 types of measurement scales. They are

1) Nominal scale
2) Ordinal scale
3) Interval scale.
4) Ratio scale.

Nominal scale :- Nominal scale is the simplest measurement scale in which objects are labeled with the help of numericals for their identification. It is a least powerful measurement scale which specify order or distance relationship of an object.

In this type of measurement scale associates objects with catagories and then describes differences between them. Hence nominal scale is also called as counted data.

Eg:- Roll numbers assigned to student for their identification.

Ordinal scale :- ordinal scale is the lowest level of measurement which specify rank order of a particular event. This type of scaling is mostly used in research which are related to qualitative phenomena

Eg:- Ranks of student in a class.

Interval scale :- Interval scale is a measurement scale which is constructed by adjusting equal interval of units based on some standard rules.

Eg:- Thermometer for measuring temperature on Fahrenheit.

Ratio scale :- Ratio scale is the measurment scale that gives actual amount of variables interms of specific units. Statistical calculations and real number calculations can be done with the help of ratio scale.

Eg:- Ratio scale is commonly used for measuring physical dimensions like weight, distance and height etc.

Non - Parametric tests

- One sample
  - Sign test
  - run test
- Two sample
  - Independent
    - wald-wolfowitz run test
    - Mann-Whitney U-test
    - Median test
  - dependent.
    - wilcoxon signed rank test
    - Sign test

Sign test :- Let $x_1\, x_2 \cdots x_n$ be a random sample drawn from a population of size n and let M be the median of the given distribution (or) for the given observations.

Null hypothesis $H_0: M = M_0$ i.e the median divides the distribution equally.

To test H_0 subtract M_0 from each observation i.e calculate $x_1 - M_0,\ x_2 - M_0 \cdots x_n - M_0$. If we observe the signs of the above expressions n terms, we get some +ve signs, −ve signs and some values may be zeros.

If any value of $x_i - M_0$ is zero, leave or omit that observation and do the remaining calculations for the reduced sample size.

Since M_0 is median we have

$$P[X > M_0] = P[X < M_0] = 1/2$$

Let us consider k be the number of positive signs then k follows binomial distribution with probability of getting a success.

$$\therefore \quad P(k) = {}^nC_k \left(1/2\right)^k \left(1/2\right)^{n-k} \quad k = 0, 1 2 \cdots n$$

$$\text{i.e} \quad P(k) = {}^nC_k \left(1/2\right)^k, \quad k = 0, 1, 2 \cdots n$$

From the cumulative binomial tables the values of P as follows.

$$P = \Sigma \, {}^nC_k \left(1/2\right)^k$$

Calculate $2P$ and compare it with level of significance α.

If $2P > \alpha$, we may accept H_0

Otherwise we may reject H_0.

→ If sample size n is large (generally $n \geq 25$) then under H_0 statistic k follows normal distribution with mean $E(k) = \frac{n}{2}$ & $V(k) = \frac{n}{k}$

$$E(k) = nP = n \cdot 1/2 = n/2 \qquad V(k) = npq = n \cdot 1/2 \cdot 1/2 = \frac{n}{4}$$

$$\therefore \text{ Test statistic} \quad Z = \frac{k - n/2}{\sqrt{n/4}} \sim N(0,1)$$

If the calculated value of Z is greater than the table value Z at $\alpha\%$ level of significance then we reject H_0 otherwise we accept H_0.

Sign test (Two samples) :-

Let (x_i, y_i), $i = 1 2 \cdots n$ be a paired sample observations of size n drawn from two different populations with probability density functions $f_1(x)$ and $f_2(y)$ respectively.

Then setup the null hypothesis $H_0 : f(x) = f(y)$

Alternative hypothesis $H_1 : f(x) \neq f(y)$

calculate $d_i = x_i - y_i$ $\quad i = 1 2 \cdots n$

If we observe the signs, we will get some +ve signs, some negative signs and some values may be zero. If any value of d_i is zero, remove that pair of observations from the sample.

under H_0 we know that

$$P[x_i - y_i > 0] = \frac{1}{2} = P[x_i - y_i < 0]$$

Let us consider a statistic K as the number of +ve signs.

We get the critical values x_1 and x_2 from the tables of binomial distribution to the corresponding values of n at $\alpha/2$ and $1 - \alpha/2$

If $k \leq x_1$ or $k \geq x_2$ i.e $x_2 \leq k \leq x_1$

then we accept Ho otherwise we may reject Ho.

For large values of n i.e $n \geq 25$, we follow normal distribution.

i.e $E(k) = np = n \cdot \frac{1}{2} = \frac{n}{2}$

$v(k) = npq = n \cdot \frac{1}{2} \cdot \frac{1}{2} = \frac{n}{4}$

Test statistic $Z = \dfrac{k - n/2}{\sqrt{n/4}} \sim N(0,1)$

where k is no. of +ve signs.

If the calculated value of Z is greater than the table value of Z then we accept Ho otherwise we reject Ho.

Wald - wolfowitz run test (one sample) :-

Def of run :- A run is defined as a sequence of letters of one kind surrounded by a sequence of letters of another kind. The number of elements in a run is usually refered as the length of a run.

Eg:-

$$\begin{array}{ccccccccc} A & A & B & A & B & B & A & B & A \\ 1 & & 2 & 3 & 4 & & 5 & 6 & 7 \end{array}$$

Total no. of runs $= 7$.

Let $x_1, x_2 \cdots x_n$ be a random sample of size n from a population.

Setup the null hypothesis H_0 : Sample is drawn at random

Alternative hypothesis H_1 : Sample is not drawn at random.

To test H_0 first arrange the sample observations in ascending order and obtain the median.

Let M be the median of the given observations.

If $x_i \leq M$ denote A and if $x_i > M$ denote B.

Then we will get some arrangement of A, B's.

Find the total number of runs denoted by r.

Find the no. of letters in A denoted by n_1, and the no. of letters in B denoted by n_2.

If $n_1 \leq 20$, $n_2 \leq 20$ consider it as a small sample corresponding to the values of n_1, n_2 at α% level of significances we get the critical values r_1 and r_2 from the run tables.

If the value of r lies between r_1 and r_2, we accept H_0 otherwise we reject H_0.

If anyone of n_1 and n_2 is greater than 20 then the distribution of r follows normal and test statistic as follows.

$$Z = \frac{r - \left[\frac{2n_1 n_2}{n_1+n_2} + 1\right]}{\sqrt{\frac{2n_1 n_2 (2n_1 n_2 - n_1 - n_2)}{(n_1+n_2)^2 (n_1+n_2-1)}}} \sim N(0,1)$$

If the calculated value of Z is greater than the table value of Z at α % level of significance then we reject H_0 otherwise we accept H_0.

Wald – Wolfowitz Run test (Two samples) :-

Let $X_1\ X_2 \dots X_{n_1}$ and $Y_1\ Y_2 \dots Y_{n_2}$ be independent random samples from two different populations with density function $f_1(X)$ and $f_2(Y)$ respectively.

Null hypothesis $H_0 : f_1(X) = f_2(Y)$

Alternative hypothesis $H_1 : f_1(X) \neq f_2(Y)$ (T.T.T)

First combine two samples and arrange the sample observations in the order of their magnitude. Now total no. of elements are $n = n_1 + n_2$.

In this arrangement denote A if the element belongs to first sample and denote B if the element

Wilcoxon signed rank test (Two samples) :-

Let $x_1\ x_2 \dots x_n$ and $y_1\ y_2 \dots y_n$ be two random samples of sizes n from two different populations with probability density functions $f_1(x)$ and $f_2(y)$ respectively.

In sign test we consider signs only. If we consider ranks of the differences then it is most powerful test than the sign test. If we consider signs & ranks also then the test is called Wilcoxon signed rank test.

Setup the null hypothesis $H_0 : f_1(x) = f_2(y)$

i.e the two random samples are coming from the same distribution.

First calculate $d_i = x_i - y_i \quad i = 1\ 2 \dots n$.

Any of $d_i = 0$, remove that pair (x_i, y_i) from the sample and calculates ranks for $|d_i|$.

And then Calculate sum of ranks of +ve signs and −ve signs separately. Obtain T as minimum of these sums.

i.e $T = \min\{$ sum of ranks of +ve signs, sum of ranks of −ve signs $\}$.

We will get the critical value of T for the corresponding value of n at a specified level of significance

If $T_{cal} > T_{tab}$ then accept H_0 otherwise reject H_0.

belongs to second sample. Then we will get some arrangements of A, B's.

obtain the total no. of runs and denoted by r.

Denote runs in A by r_A and runs in B by r_B.

If $n_1 \leq 20$, $n_2 \leq 20$ corresponding to the values of n_1 and n_2 with α level of significance. We will get the critical values r_1 and r_2 from the run tables. If $r_1 < r < r_2$ then we accept Ho otherwise we reject Ho.

If any of n_1 and n_2 greater than 20 then we apply normal test. r follows normal distribution

with $$E(r) = \frac{2n_1 n_2}{n_1+n_2} + 1$$

and $$V(r) = \frac{2n_1 n_2 (2n_1 n_2 - n_1 - n_2)}{(n_1+n_2)^2 (n_1+n_2-1)}$$

$\therefore$ Test statistic

$$Z = \frac{r - E(r)}{S.E\ of\ (r)} \sim N(0,1)$$

If the calculated value of Z is greater than the table value of Z at α % level of significance then we reject Ho otherwise we accept Ho.

$\therefore$ If $n \geq 25$ we use normal test.

Hence T follows normal distribution with mean and variance as

$$\text{Mean } E(T) = \frac{n(n+1)}{4}$$

$$\text{and variance } v(T) = \frac{n(n+1)(2n+1)}{24}$$

$\therefore$ Test statistic

$$Z = \frac{T - E(T)}{S \cdot E \text{ of } (T)} \sim N(0,1)$$

If the calculated value of Z is greater than the table value of Z at $\alpha\%$ level of significance then we reject H_0 otherwise we accept H_0.

Mann - Whitney U- test (Two samples) :-

Let $X_1\ X_2 \cdots X_{n_1}$ and $Y_1\ Y_2 \cdots Y_{n_2}$ be two independent random samples of sizes n_1 and n_2 from two populations with probability density functions $f_1(x)$ and $f_2(y)$ respectively.

Null hypothesis $H_0 : f_1(x) = f_2(y)$

i.e Two samples are drawn from same distribution.

Alternative hypothesis $H_1 : f_1(x) \neq f_2(y)$

i.e Two samples are not drawn from same distribution.

First combine two samples and let T denote the sum of the ranks of the second series, let us say Y series in the combined ordered sample.

Test statistic is defined as

$$u = n_1 n_2 + \frac{n_2(n_2+1)}{2} - T$$

To find the distribution of T is very difficult. But Mann and whitney have obtained mean and variance for T, and established U is assymptotically normal distributed with mean $E(u) = \frac{n_1 n_2}{2}$

and variance $v(u) = \frac{n_1 n_2 (n_1+n_2+1)}{12}$

$\therefore$ Test statistic

$$Z = \frac{u - E(u)}{SE \text{ of } (u)} \sim N(0,1)$$

If the calculated value of Z is greater than the table value of Z at α% level of significance then we reject Ho otherwise we accept Ho.

Median test (Two samples) :-

Signtest and wilcoxon signed rank tests are useful if the sample sizes are equal. If the sizes of two samples are different, these tests cannot be applied. In such a case the median test is relevant to test the hypothesis of two samples have been drawn from the same population.

Let $X_1, X_2 \ldots X_{n_1}$ and $Y_1, Y_2 \ldots Y_{n_2}$ be two independent random samples from two different populations with probability density functions $f_1(x)$ and $f_2(y)$ respectively.

Null hypothesis $H_0 : f_1(x) = f_2(y)$

i.e two samples have drawn from the same population

Alternative hypothesis $H_1 : f_1(x) \neq f_2(y)$

i.e two samples have not been drawn from the same population.

To test H_0, first combine two samples in assending order. Let M be the median of arrangement. Denote number of values greater than or equal to median M in the first sample by m_1 and in the second sample by m_2. The number of values less than M in the first sample are $n_1 - m_1$ and in the second sample are $n_2 - m_2$. This can be represented in the following 2×2 contingency table.

	Sample I	Sample II	Total
No. of observations $\geq M$	m_1	m_2	$m_1 + m_2$
No. of observations $< M$	$n_1 - m_1$	$n_2 - m_2$	$(n_1+n_2)-(m_1+m_2)$
Total	n_1	n_2	$n_1 + n_2$

If either n_1 or n_2 or both is less than 10 then.
we apply hyper geometric distribution.

$$P(m_1) = \frac{{}^{n_1}C_{m_1} \cdot {}^{n_2}C_{m_2}}{{}^{n_1+n_2}C_{m+m_2}}$$

We calculate P, if $P > \alpha$ then we accept Ho otherwise we reject Ho.

If both n_1 & n_2 greater than 10, we apply chi-square distribution, we calculate χ^2 using 2x2 contingency table

$$\chi^2 = \frac{N(ad-bc)^2}{(a+c)(b+d)(a+b)(c+d)}$$

where $N = a+b+c+d$.

If the calculated value of χ^2 is greater than the table value of χ^2 at α % level for 1 df then we reject Ho otherwise we accept Ho.

If any one of the frequency is less than 5 then we apply Yate's correction for chi-square statist

$$\chi^2 = \frac{N\left[|ad-bc| - N/2\right]^2}{(a+c)(b+d)(a+b)(c+d)}$$

UNIT - 2

LARGE SAMPLE TESTS

SAQS

under unit-1

- Defination of Large sample & Large Sample Test.
- Role of Normal distribution in Large sample.
- Defination of critical or significant value.
- One tailed test.
  - Right-tailed test
  - Left-tailed test.
- Two tailed test
- steps in testing of Hypothesis
- Large sample test for Means.
  - Single Mean
  - Difference of Mean
- Large sample test for standard deviations.
  - Single standard deviation
  - Difference of Standard deviation
- ☆ Large Sample test for proportions.
  - Single proportion
  - Difference of proportion.
- ☆ Large sample test for correlation co-efficient.
  - Single correlation
  - Difference of correlation.

Large Sample and Large Sample Test

The entire statistical theory is used/based on sample size (n) for conducting testing of Hypothesis. If the sample size (n) $n \geq 30$ then it is called Large sample and the test depending on the large sample is called Large Sample test.

Role of Normal Distribution in Large Sample.

We have already studied that for large value of 'n'. almost all distributions.

Example :-

Binomial, Poisson, Negative binomial, Exponential etc. follows normal distribution.

In this case we already normal distribution to entire large sample test from the area of property of normal distribution we have.

If $X \sim N(N, \sigma^2)$ then $z = \frac{X-\mu}{\sigma} \sim N(0,1)$

usually,

$$z = \frac{X - E(X)}{S.E \text{ of } (X)} \sim N(0,1)$$

From the standard normal tables we have

$$P[-3 < z < 3] = 0.9973$$

$$P[|z| < 3] = 0.9973$$

∴ we have,

$$P[|z| \geq 3] = 1 - 0.9973$$
$$= 0.0027$$

We expect the standard normal variate (z) lies between ± 3 i.e, $|z| < 3$ then we accept H_0 otherwise $|z|$ is ≥ 3 $(|z| > 3)$ then we reject H_0.

critical or Significant or Table Value

The value of test statistic which seperates rejection on critical region and the acceptance region is called critical or significant or Table value.

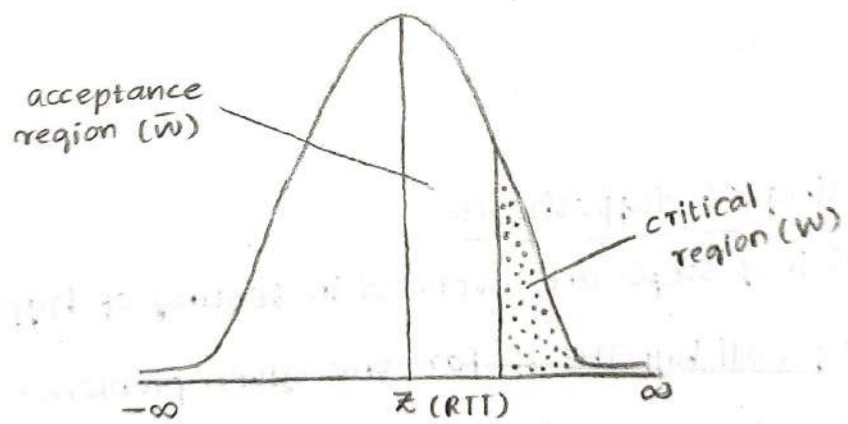
Critical value depends on level of significance (α) and Alternative hypothesis (H_1) decides whether the test is two tailed or one-tailed test.

One-Tailed test

Right tailed and Left tailed tests are called one-Tailed test.

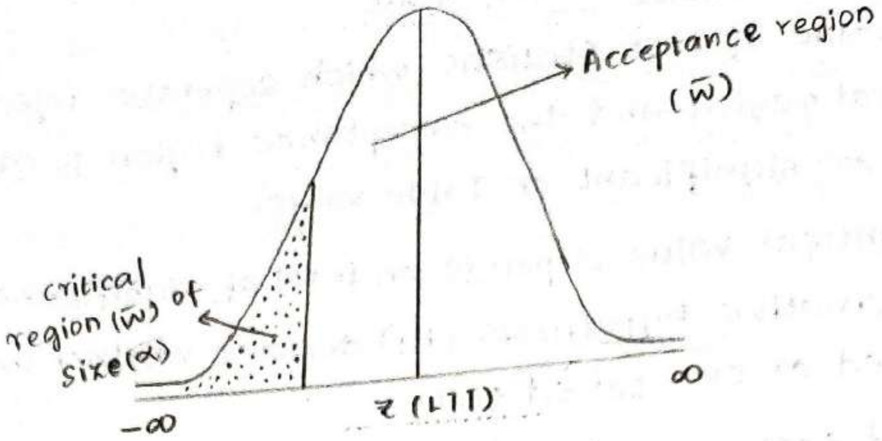
Right-Tailed test

A statistical Hypothesis in which alternative hypothesis (H_1) contains $>$ symbol then the critical region (w) lies on the right side of the distribution is called Right-tailed test and is shown in the following diagram.



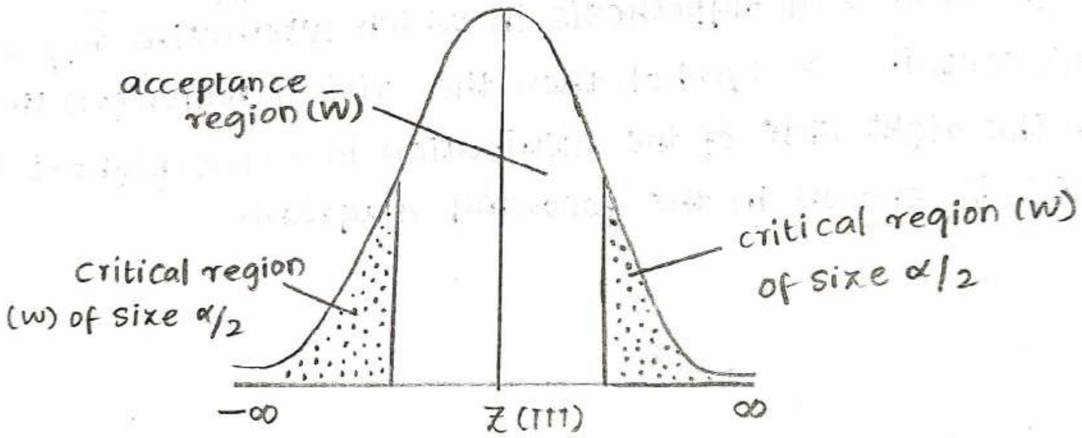
Left-Tailed Test

A statistical Hypothesis in which alternative hypothesis (H_1) contains $<$ symbol then the critical region (w) lies on the Left side of the distribution is called Left tailed test. and is shown in the following diagram.



Two-Tailed Test

A statistical hypothesis in which alternative hypothesis (H_1) contains not equal to ($\neq$) symbol then the critical region (w) lies on both sides of the distribution is called Two-Tailed Test and as shown in the following diagram.



Steps in the Testing of Hypothesis

The following steps are involved in testing of hypothesis.

1. Step:- specify null hypothesis for the given problem.

2. Step :- Specify alternative hypothesis based on null hypothesis.

3. Step:- Fix the Level of Significance (α)

α' generally takes 5% or 1% or 10%.

and z_α is the table value.

4. Step:- under Ho, test Statistics is given by

$$Z = \frac{t - E(t)}{S.E\ of\ (t)} \sim N(0,1)$$

calculate $|z|$ if neccessary.

5. Step :- If the calculated value of z is greater than ($>$) the table value of z at α % level of significance then we get, reject H_0 otherwise we accept H_0.

Notations

- N - represents population size
- n - represents sample size
- μ - represents population mean
- $\bar{x}$ - represents sample mean
- P - represents population proportion
- p - represents sample proportion
- σ - represents population standard deviation.
- s - Sample standard deviation.
- ρ - Population correlation coefficient.
- r - sample correlation coefficient.

Large sample test for single mean.

Let us consider a random sample $x_1, x_2, \dots x_n$ of size 'n' drawn from a normal population with mean 'μ' and variance 'σ^2' then the sample mean $\bar{x}$ follows Normal distribution with mean 'μ' and variance '$\frac{\sigma^2}{n}$'. i.e;

$$\boxed{\begin{array}{l} X \sim N(\mu, \sigma^2) \text{ then,} \\ \bar{x} \sim N\left(\mu, \frac{\sigma^2}{n}\right) \end{array}}$$

Step - 1

Null hypothesis $H_0 : \mu = \mu_0$ i.e, the sample is drawn from the population mean. [OR] there is no significant difference between the sample and population mean.

Step-2

Alternative Hypothesis $H_1: \mu \neq \mu_0$ (T.T.T) or $\mu > \mu_0$ (R.T.T) or $\mu < \mu_0$ (L.T.T) i.e; the sample mean is not drawn from the population mean [OR] there is a significant difference between the sample and population mean.

Step-3

'α' is called the level of significance usually takes 5% or 1% or 10% and Z_α is the table value.

Step-4

The test statistic is given by

$$Z = \frac{t - E(t)}{S.E \text{ of } (t)} \sim N(0,1)$$

$$Z = \frac{\bar{X} - E(\bar{X})}{S.E \text{ of } (\bar{X})} \sim N(0,1)$$

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

$\therefore$ under H_0

$$\boxed{Z = \frac{\bar{X} - \mu_0}{\sigma_0/\sqrt{n}} \sim N(0,1)}$$

calculate $|Z|$ if necessary.

Step-5

If the calculated value of Z is greater than the table value of Z at α% level of significance then we reject (H_0) otherwise, we accep (H_0).

Large Sample Test for difference means

Let us consider, A random sample $x_1, x_2 \ldots x_{n_1}$ of size n_1 drawn from a normal population with mean μ_1 and variance σ_1^2.

consider, Another random sample $x_1, x_2 \ldots x_{n_2}$ of size n_2 drawn from a normal population with mean μ_2 and variance σ_2^2 then the sample means $\bar{X}, \bar{Y}$ follows normal distribution with mean (μ_1) and variance $\frac{\sigma_1^2}{n_1}$ and mean (μ_2) and variance $\frac{\sigma_2^2}{n_2^2}$ respectively. i.e;

$\bar{X} \sim N\left[\mu_1, \frac{\sigma_1^2}{n_1}\right]$ and $\bar{Y} \sim N\left[\mu_2, \frac{\sigma_2^2}{n_2}\right]$ Then we have.

$$\boxed{\bar{X} - \bar{Y} \sim N\left[\mu_1 - \mu_2, \frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}\right]}$$

Step-1

Null Hypothesis $H_0: \mu_1 = \mu_2$. i.e. two population means are equal or there is no significant difference between the population means.

Step-2

Alternative Hypothesis $H_1: \mu_1 \neq \mu_2$ (T.T.T) or $\mu_1 > \mu_2$ (R.T.T) or $\mu_1 < \mu_2$ (L.T.T) i.e. ~~{the sample}~~ two population means are not equal or there is significant difference between the population means.

Step-3

'α' is called the level of significance usually takes 5% or 1% or 10% and Z_α is the table value.

Step-4

The test statistic is given by

$$z = \frac{t - E(t)}{S.E \text{ of } (t)} \sim N(0,1)$$

$$z = \frac{(\bar{X}-\bar{Y}) - E(\bar{X}-\bar{Y})}{S.E \text{ of } (\bar{X}-\bar{Y})} \sim N(0,1)$$

$$= \frac{(\bar{X}-\bar{Y}) - E(\bar{X}-\bar{Y})}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}} \sim N(0,1)$$

$\therefore$ Under H_0

$$\boxed{z = \frac{(\bar{X}-\bar{Y})}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}} \sim N(0,1)}$$

calculate $|z|$ if neccessary.

Step-5

If the calculated value of z is greater than table value of z at $\alpha\%$ level of significance then we reject (H_0) otherwise, we accept (H_0).

NOTE

1. If population standard deviations are not given i.e. σ_1 and σ_2 is not given then they can be estimated by using.

$$\sigma_1 = s_1 \text{ and } \sigma_2 = s_2$$

$$\boxed{\therefore z = \frac{\bar{X}-\bar{Y}}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}} \sim N(0,1)}$$

2. If the both the population standard deviations are equal i.e, $\sigma_1 = \sigma_2 = \sigma$ then we have.

$$\boxed{Z = \frac{\bar{X} - \bar{Y}}{\sigma\sqrt{\frac{1}{n_1}+\frac{1}{n_2}}} \sim N(0,1)}$$

Large Sample Test for Single Standard deviation.

Let us consider, A Random sample $x_1, x_2, \ldots x_n$ of sixe n drawn from a normal population with mean(μ) and variance σ^2 then the sample standard deviation (s) follows normal distribution with mean (σ) and variance $\frac{\sigma^2}{2n}$ i.e;

$$\boxed{X \sim N(\mu, \sigma^2) \text{ then } s \sim N\left(\sigma, \frac{\sigma^2}{2n}\right)}$$

Step-1

Null Hypothesis $H_0 : \sigma = \sigma_0$ i.e; the sample is drawn from the population standard deviation (or) there is no significant difference between the sample & population standard deviation.

Step-2

Alternative Hypothesis $H_1 : \sigma \neq \sigma_0$ (T.T.T) or $\sigma > \sigma_0$ (R.T.T) or $\sigma < \sigma_0$ (L.T.T) i.e; the sample is not drawn from the population standard deviation [OR] There is a significant difference between population & sample standard deviation.

Step-3

'α' is called the level of signifance usually takes 5% or 1% or 10% and Z_α is the table value.

Step-4

The test stastictic is given by

$$Z = \frac{t - E(t)}{S.E \text{ of } (t)} \sim N(0,1)$$

$$z = \frac{s - E(s)}{S.E \text{ of } (s)} \sim N(0,1)$$

$$z = \frac{s - \sigma}{\sigma / \sqrt{2n}} \sim N(0,1)$$

$\therefore$ under Ho

$$\boxed{z = \frac{s - \sigma_0}{\sigma_0 / \sqrt{2n}} \sim N(0,1)}$$

calculate $|z|$ if necessary.

Step-5

If the calculated value of z is greater than table value of z at α% level of significance then we reject Ho otherwise we accept Ho.

Large Sample test for difference of standard deviations.

Let us consider,

A random sample $x_1, x_2, \ldots x_n$ of size n' drawn from a normal population with mean μ_1, and variance σ_1^2.

consider,

A random sample $y_1, y_2, \ldots y_{n_2}$ of size n_2 drawn from a normal population with mean μ_2 and variance σ_2^2. Then sample standard deviations s_1 & s_2 follows normal distribution with it mean σ_1 & σ_2 and variances $\frac{\sigma_1^2}{2n_1}$ & $\frac{\sigma_2^2}{2\sigma_2}$ respectively. i.e;

$$s_1 \sim N\left[\sigma_1, \frac{\sigma_1^2}{2n_1}\right] \ \& \ s_2 \sim N\left[\sigma_2, \frac{\sigma_2^2}{2n_2}\right]$$

Then we have.

$$\boxed{s_1 - s_2 \sim N\left[\sigma_1 - \sigma_2, \frac{\sigma_1^2}{2n_1} + \frac{\sigma_2^2}{2n_2}\right]}$$

Step-1 :

Null Hypothesis Ho : $\sigma_1 = \sigma_2$ that is population standard deviations are equal ~~th~~ or there is no significant difference between the population standard deviations.

step-2

Alternative Hypothesis $H_1 : \sigma_1 \neq \sigma_2$ (T.T.T) or $\sigma_1 > \sigma_2$ (R.T.T) or $\sigma_1 < \sigma_2$ (L.T.T) i.e, two population standard deviations are not equal [or] there is a significant difference between the two population standard deviations.

step-3

'α' is called the level of significance usually takes 5% or 1% or 10% and Z_α is the table value.

Step-4

The test statistic is given by.

$$Z = \frac{t - E(t)}{S.E \text{ of } (t)} \sim N(0,1)$$

$$Z = \frac{(S_1 - S_2) - E(S_1 - S_2)}{S.E \text{ of } (S_1 - S_2)} \sim N(0,1)$$

$$\boxed{Z = \frac{(S_1 - S_2) - (\sigma_1 - \sigma_2)}{\sqrt{\frac{\sigma_1^2}{2n_1} + \frac{\sigma_2^2}{2n_2}}}}$$

under Ho:

$$\boxed{Z = \frac{S_1 - S_2}{\sqrt{\frac{\sigma_1^2}{2n_1} + \frac{\sigma_2^2}{2n_2}}} \sim N(0,1)}$$

calculate $|z|$ if necessary.

step-5

If the calculated value of z is greater than the table value of z at α% level of significance then we reject (Ho) otherwise we accept (Ho).

NOTE :-

1. If the population standard deviations are not given i.e; σ_1 & σ_2 are not given then they can be estimated by using $\sigma_1 = s_1$ and $\sigma_2 = s_2$.

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

2. If both the population standard deviations are equal. i.e; $\sigma_1 = \sigma_2 = \sigma$; then we have

$$\boxed{\therefore z = \frac{s_1 - s_2}{\sigma\sqrt{\frac{1}{2n_1} + \frac{1}{2n_2}}} \sim N(0,1)}$$

Large Sample test for single Proportion

Let 'x' represents number of success out of n trails and P be the probability of success, which follows binomial distribution.

we know that,

Mean and variance of binomal distribution is given by.

$$E(x) = nP \quad \text{and} \quad V(x) = nPQ$$

let,

The sample proportion be denoted by 'p' and is given by

$$\boxed{p = \frac{x}{n}}$$

$$E(p) = E\left[\frac{x}{n}\right] = \frac{1}{n}E(x) = \frac{1}{\cancel{n}}\cancel{n}P = P$$

$$\boxed{E(p) = P}$$

$$V(p) = V\left(\frac{x}{n}\right) = \frac{1}{n^2}V(x) = \frac{1}{n^2}\cancel{n}PQ = \frac{PQ}{n}$$

$$\boxed{V(p) = \frac{PQ}{n}}$$

$$\boxed{\therefore\ p \sim N\left[P, \frac{PQ}{n}\right]}$$ [For large value of 'n' binomial follows normal distn]

Step-1

Null Hypothesis $H_0 : P = P_0$ i.e; sample is drawn from the population (or) there is no significant difference between the sample of population proportion.

Step-2

Alternative Hypothesis $H_1 : P \neq P_0$ (T.T.T) or $H_1 : P > P_0$ (R.T.T) or $P < P_0$ (L.T.T) i.e; the sample is not drawn from the population or there is a significant difference between Sample & population proportion.

Step-3

'α' is called the level of significance usually takes 5%. or 1%. or 10%. and Z_α is the table value.

Step-4

The test statistic is given by

$$Z = \frac{t - E(t)}{S.E \text{ of } (t)} \sim N(0,1)$$

$$Z = \frac{p - E(p)}{S.E \text{ of } (p)} \sim N(0,1)$$

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

under H_0

$$\boxed{Z = \frac{p - P_0}{\sqrt{\frac{P_0 Q_0}{n}}} \sim N(0,1)}$$

calculate $|Z|$ if necessary.

Step-5

If the calculated value of z is greater than the table value of z at $\alpha\%$ level of significance then we reject (H_0) otherwise we accept (H_0).

Large Sample Test for difference of proportions.

Let,

X_1 represents no. of success out of n_1 trials and P_1 be the probability of success for first sample and Let X_2 represents no. of success out of n_2 trials and P_2 be the probability of success for second sample. then both samples follows binomial distribution.

Then we have,

$$E(X_1) = n_1 P_1 \quad \& \quad E(X_2) = n_2 P_2$$

$$V(X_1) = n_1 P_1 Q_1 \qquad V(X_2) = n_2 P_2 Q_2.$$

Let,

The sample proportions are denoted by 'p_1' and 'p_2'

Then we have,

$$\boxed{p_1 = \frac{X_1}{n_1}} \quad \& \quad \boxed{p_2 = \frac{X_2}{n_2}}$$

$$\boxed{p_1 \sim N\left[P_1, \frac{P_1 Q_1}{n_1}\right]} \quad \& \quad \boxed{p_2 \sim N\left[P_2, \frac{P_2 Q_2}{n_2}\right]}$$

Then we have,

$$\boxed{p_1 - p_2 \sim N\left[P_1 - P_2, \frac{P_1 Q_1}{n_1} + \frac{P_2 Q_2}{n_2}\right]}$$

Step-1

Null Hypothesis $H_0 : P_1 = P_2$ i.e; two population proportions are equal or there is no significant difference between the population proportion.

Step-2

Alternative Hypothesis $H_1: P_1 \neq P_2$ (T.T.T) or $P_1 > P_2$ (R.T.T) or $P_1 < P_2$ (L.T.T) i.e; two population proportions are not equal or there is no significant difference between the population proportions.

Step-3

'α' is called the level of significance usually takes 5% or 1% or 10% and Z_α is the table value.

Step-4

The test statistic is given by

$$Z = \frac{t - E(t)}{S.E \text{ of } (t)} \sim N(0,1)$$

$$Z = \frac{(p_1 - p_2) - E(p_1 - p_2)}{S.E \text{ of } (p_1 - p_2)} \sim N(0,1)$$

$$\boxed{Z = \frac{(p_1 - p_2) - (P_1 - P_2)}{\sqrt{\frac{P_1Q_1}{n_1} + \frac{P_2Q_2}{n_2}}} \sim N(0,1)}$$

under Ho.

$$\boxed{Z = \frac{p_1 - p_2}{\sqrt{\frac{P_1Q_1}{n_1} + \frac{P_2Q_2}{n_2}}} \sim N(0,1)}$$

calculate $|Z|$ if neccessary.

Step-5

If the calculated value of Z is greater than the table value of Z at α% level of significance then we reject (Ho) otherwise we accept (Ho).

NOTE

1. If $P_1 = P_2 = P$, $Q_1 = Q_2 = Q$ then we have,

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

2. If P & Q are not given then it can be estimated by using.

$$\boxed{P = \frac{x_1 + x_2}{n_1 + n_2}} \quad (or) \quad \boxed{P = \frac{n_1 p_1 + n_2 p_2}{n_1 + n_2}}$$

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

Large sample test for Single correlation coefficient.

Let us consider,

A Bivariate data $(x_1, y_1), (x_2, y_2) \ldots (x_n, y_n)$ of 'n' pairs of observations drawn from a normal population

Let, r be the sample correlation coefficient and 'ρ' be the population correlation coefficient.

Step-1

Null Hypothesis $H_0 : \rho = \rho_0$ i.e; the sample is drawn from the population or there is no significant difference between the sample and the population correlation coefficient

Step-2

Alternative Hypothesis $H_1 : \rho \neq \rho_0$ (T.T.T) or $\rho > \rho_0$ (R.T.T) or $\rho < \rho_0$ (LTT) i.e; the sample is not drawn from the population or there is a significant difference between the sample and the population correlation coefficient.

Step-3

'α' is called the level of significance usually takes the values 5% or 1% or 10% & Z_α is the table value.

Step-4

CASE - I

[when ρ is small i.e; ρ tends to 0] then for large value of 'n' the distribution of sample correlation coefficient (r) follows normal distribution with mean ['ρ'] and variance $\left[\frac{(1-\rho^2)^2}{n}\right]$. i.e;

$$\boxed{r \sim N\left[\rho, \frac{(1-\rho^2)^2}{n}\right]}$$

The test statistic is given by

$$Z = \frac{t - E(t)}{S.E \text{ of } (t)} \sim N(0,1)$$

$$Z = \frac{r - E(r)}{S.E \text{ of } (r)} \sim N(0,1)$$

$$\boxed{Z = \frac{r-\rho}{\frac{1-\rho^2}{\sqrt{n}}} \sim N(0,1)}$$

under H_0:

$$\boxed{Z = \frac{r-\rho_0}{\frac{1-\rho_0^2}{\sqrt{n}}} \sim N(0,1)}$$

calculate $|Z|$ if necessary.

Step-5

If the calculated value of Z is greater than the table value of Z at α% level of significance then we reject (H_0) otherwise we accept (H_0).

CASE - II

[ρ tends to 1] i.e;

ρ is more than ± 0.7, then we use Fisher's z transformation.

$\therefore$ we have,

$$\boxed{z = \frac{1}{2}\log_e\left(\frac{1+r}{1-r}\right)}$$

mean & variance of 'z' are given by

$$\boxed{E(z) = \frac{1}{2}\log_e\left(\frac{1+\rho}{1-\rho}\right)} \quad \text{and}$$

$$\boxed{V(z) = \frac{1}{n-3}}$$

$\therefore$ The test statistic is given by

$$V = \frac{t - E(t)}{S.E \text{ of } (t)} \sim N(0,1)$$

$$V = \frac{z - E(z)}{S.E \text{ of } (z)} \sim N(0,1)$$

$$\boxed{V = \frac{\frac{1}{2}\log_e\left(\frac{1+r}{1-r}\right) - \frac{1}{2}\log_e\left(\frac{1+\rho}{1-\rho}\right)}{\frac{1}{\sqrt{n-3}}} \sim N(0,1)}$$

under H_0:

$$\boxed{V = \frac{\frac{1}{2}\log_e\left(\frac{1+r}{1-r}\right) - \frac{1}{2}\log_e\left(\frac{1+\rho_0}{1-\rho_0}\right)}{\frac{1}{\sqrt{n-3}}} \sim N(0,1)}$$

Calculate $|z|$ if necessary.

Step-5

If the calculated value of v is greater than the table values of z at α% level of significance th[...] (H_0)

otherwise we accept (H_0).

Large Sample test for difference of correlation coefficients.

Let us consider,

A bi-variate random sample $(x_1, y_1), (x_2, y_2) \ldots (x_{n_1}, y_{n_1})$ of size 'n_1', drawn from a normal population.

Let us consider,

Another random bi-variate sample $(x_1, y_1) \ldots (x_{n_2}, y_{n_2})$ of size 'n_2' drawn from another normal population.

Let the sample correlation coefficients are denoted by r_1 and r_2 and population correlation coefficient are denoted by ρ_1, ρ_2.

step-1

Null Hypothesis $H_0: \rho_1 = \rho_2$. i.e; two population correlation coefficients are equal (or) there is no sigificance difference between the population correlation coefficients.

step-2

Alternative Hypothesis H_1, $\rho_1 \neq \rho_2$ (T.T.T) $\rho_1 > \rho_2$ (R.T.T) $\rho_1 < \rho_2$ (L.T.T) ie; two population correlation coefficients are not equal or there is a significant difference between the population correlation coefficients.

step-3

'α' is called the level of significance at 5% or 1% or 10% and z_α is the table value.

step-4

case-I

[when ρ_1 & ρ_2 are small] i.e; ρ_1 & ρ_2 tends to zero then for large value n_1 & n_2 we have.

$$r_1 \sim N\left[\rho_1, \frac{(1-\rho_1^2)^2}{n_1}\right].$$

$$r_2 \sim N\left[\rho_2 , \frac{(1-\rho_2^2)^2}{n_2}\right]$$

$$\boxed{\therefore \; r_1 - r_2 \sim N\left[\rho_1-\rho_2 , \frac{(1-\rho_1^2)^2}{n_1} + \frac{(1-\rho_2^2)^2}{n_2}\right]}$$

The test statistic is given by

$$Z = \frac{t - E(t)}{S.E \text{ of } (t)} \sim N(0,1)$$

$$Z = \frac{(r_1-r_2) - E(r_1-r_2)}{S.E \text{ of } (r_1-r_2)} \sim N(0,1)$$

$$\boxed{Z = \frac{(r_1-r_2)-(\rho_1-\rho_2)}{\sqrt{\frac{(1-\rho_1^2)^2}{n_1}+\frac{(1-\rho_2^2)^2}{n_2}}} \sim N(0,1)}$$

under H_0:

$$\boxed{Z = \frac{r_1 - r_2}{\sqrt{\frac{(1-\rho_1^2)^2}{n_1}+\frac{(1-\rho_2^2)^2}{n_2}}} \sim N(0,1)}$$

Calculate $|z|$ if necessary.

Step-5

If the calculated value of z is greater than the table value of z and α% level of significance the we reject (H_0), otherwise we accept (H_0).

Case-II

[If ρ_1 & ρ_2 tends to one] i.e; ρ_1 & ρ_2 are more than ± 0.7 then we use Fisher's z transformation.

we have,

$$Z_1 = \frac{1}{2}\log_e\left(\frac{1+r_1}{1-r_2}\right) \;\&$$

$$z_2 = \frac{1}{2} \log_e \left(\frac{1+r_2}{1-r_2} \right)$$

Mean & variance of z are given by

$$\Rightarrow E(z_1) = \frac{1}{2} \log_e \left(\frac{1+\rho_1}{1-\rho_1} \right) \;\&$$

$$\Rightarrow E(z_2) = \frac{1}{2} \log_e \left(\frac{1+\rho_2}{1-\rho_2} \right)$$

$$\Rightarrow V(z_1) = \frac{1}{n_1 - 3} \;\&\; V(z_2) = \frac{1}{n_2 - 3}$$

The test statistic is given by

$$V = \frac{t - E(t)}{S.E \text{ of } (t)} \sim N(0,1)$$

$$V = \frac{(z_1 - z_1) - E(z_1 - z_2)}{S.E \text{ of } (z_1 - z_2)} \sim N(0,1)$$

$$V = \left[\frac{1}{2} \log_e \left(\frac{1+r_1}{1-r_2} \right) - \frac{1}{2} \log_e \left(\frac{1+r_2}{1-r_2} \right) \right]$$

$$\boxed{V = \frac{\left[\frac{1}{2} \log_e \left(\frac{1+\rho_1}{1-\rho_1} \right) - \frac{1}{2} \log_e \left(\frac{1-\rho_2}{1-\rho_2} \right) \right]}{\sqrt{\frac{1}{n_1 - 3} + \frac{1}{n_2 - 3}}}}$$

under H_0:

$$\boxed{V = \frac{\left[\frac{1}{2} \log_e \left(\frac{1+r_1}{1-r_1} \right) - \frac{1}{2} \log_e \left(\frac{1+r_2}{1-r_2} \right) \right]}{\sqrt{\frac{1}{n_1 - 3} + \frac{1}{n_2 - 3}}} \sim N(0,1)}$$

calculated $|z|$ if necessary.

Step-5

If the calculated value of V is greater than the table value of z at $a\%$ level of significance then we reject (H_0) otherwise we ~~g~~ accept (H_0).

UNIT - III

SMALL SAMPLE TESTS

- Defination of small samaple test
- t - test
  - t-test for single mean
  - t-test for difference of means
  - paired t-test
  - t-test for correlation coefficient
- F-test
  - F-test for equality of variances (population)
- Chi-square test
  - chi-square test for population variance
  - Chi-square test for goodness of fit.
  - chi-square test for independence of attributes.
- order of statistics. {UNIT - 2}

Small Sample and Small Sample test:

If the sample size $n \leq 29$ (or) $n < 30$ then it is called a small sample and the test depending on small samples are called small sample tests.

The entire small sample theory is based on Exact sampling distributions. [χ^2, t and F].

Assumptions of t-test

1. The sample is drawn from normal population.
2. All the samples are independent.
3. 'σ' is not known / given.

t-test for single mean.

Let us consider,

A random sample $x_1, x_2, \ldots x_n$ of size 'n' drawn from a normal population with mean 'μ' and variance 'σ^2', where 'σ^2' is not given.

Step-1

Null Hypothesis $H_0: \mu = \mu_0$ i.e, the sample is drawn from the population mean (or) there is no significant difference between the sample & population.

Step-2

Alternative Hypothesis $H_1: \mu \neq \mu_0$ (T.T.T) or $\mu > \mu_0$ (R.T.T) or $\mu < \mu_0$ (L.T.T) i.e; the sample is not drawn from the population mean (or) there is a significant difference between the sample and population mean.

Step-3

α is called the level of significance usually takes 5% or 1% or 10% and t_α is the table value.

Step-4

The test statistic under Ho is given by

$$\boxed{t = \frac{\bar{x} - \mu_0}{S/\sqrt{n}} \sim t_{n-1}}$$

(or)

$$\boxed{t = \frac{\bar{x} - \mu_0}{s/\sqrt{n-1}} \sim t_{n-1}}$$

where,

$$\boxed{\bar{x} = \frac{\sum x}{n}} \quad \boxed{S^2 = \frac{1}{n-1}\sum (x-\bar{x})^2} \; ; \; \boxed{s^2 = \frac{1}{n}\sum (x-\bar{x})^2}$$

calculate |t| if necessary.

Step-5

If the calculated value of t is greater than the table value of t at α% level of significance then we reject (Ho), otherwise we accept (Ho).

t-test for difference of means.

Let us consider,

Two independent random samples $[x_1, x_2, \ldots, x_{n_1}]$ & $[y_1, y_2, \ldots, y_{n_2}]$ of sizes n_1 & n_2 drawn from a normal population with means μ_1, μ_2 and variances σ_1^2, σ_2^2 respectively. where σ_1 & σ_2 are not given.

Step-1

Null Hypothesis $H_0: \mu_1 = \mu_2$ i.e; two population means are equal (or) there is no significant difference between the population means.

Step-2

Alternative Hypothesis $H_1: \mu_1 \neq \mu_2$ (T.T.T) or $\mu_1 < \mu_2$ (L.T.T) or $\mu_1 > \mu_2$ (R.T.T) i.e; two population means are not equal (or) there is a significant difference between the population means.

step-3

'α' is called the level of significance usually takes 5% or 1% or 10% and t_α is the table value.

step-4

under H_0:

$$\boxed{t = \frac{\bar{X} - \bar{Y}}{S\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}} \sim t_{n_1+n_2-2}}$$

where,

$$\boxed{\bar{X} = \frac{1}{n_1}\sum_{i=1}^{n_1} x_i} \quad \text{and} \quad \boxed{\bar{Y} = \frac{1}{n_2}\sum_{j=1}^{n_2} Y_j}$$

$$\boxed{S^2 = \frac{1}{n_1+n_2-2}\left[\sum_{i=1}^{n_1}(x-\bar{x})^2 + \sum_{j=1}^{n_2}(Y-\bar{Y})^2\right]} \text{ (or) } \boxed{S^2 = \frac{n_1 s_1^2 + n_2 s_2^2}{n_1+n_2-2}}$$

calculate $|t|$ if necessary.

Step-5

If the calculated value of t is greater than the table value of t at α% level of significance then we reject (H_0) otherwise we accept (H_0).

Paired t-test

Paired t-test is used under the following conditions -

1. Sample sizes are equal. i.e;

$$n_1 = n_2 = n$$

2. Both the samples are dependent.

Let us condider,

'n' pairs of observations $(x_1, Y_1) (x_2, Y_2) \ldots (x_n, Y_n)$ drawn from a normal population with means μ_1, μ_2 and variances σ_1^2, σ_2^2 where σ_1 and σ_2 are not given.

Step-1

Null Hypothesis $H_0: \mu_1 = \mu_2$ i.e; there is no significant between the samples before and after [increase or decrease].

Step-2

$H_1: \mu_1 > \mu_2$ (R.T.T)

Alternative Hypothesis $H_1: \mu_1 \neq \mu_2$ (T.T.T) or $\mu_1 < \mu_2$ (L.T.T)

i.e; there is a significant difference between the samples before and after [increase or decrease].

Step-3

'α' is called the level of significance usually takes 5% or 1% or 10% and t_α is the table value.

Step-4

under H_0,

The statistic is given by.

$$\boxed{t = \frac{\bar{d}}{S/\sqrt{n}} \sim t_{n-1}}$$

(or)

$$\boxed{t = \frac{\bar{d}}{s/\sqrt{n-1}} \sim t_{n-1}}$$

where,

$$\boxed{d = x_i - y_i} \text{ and } \boxed{\bar{d} = \frac{\Sigma d}{n}}$$

$$\boxed{S^2 = \frac{1}{n-1} \Sigma (d-\bar{d})^2} \quad \boxed{s^2 = \frac{1}{n} \Sigma (d-\bar{d})^2}$$

calculate $|t|$ if necessary.

Step-5

If the calculated value of 't' is greater than the table value of t at α% level of significance then we reject (H_0), otherwise we accept (H_0).

t-test for correlation coefficient.

Let us consider,

'n' pairs of observations (x_1, y_1) (x_2, y_2) . . . (x_n, y_n) drawn from a normal population. Let, r be the sample correlation coefficient and 'ρ' be the population correlation coefficient.

Step-1

Null Hypothesis $H_0: \rho = 0$ i.e; both the random variables x and y are uncorrelated.

Step-2

Alternative Hypothesis $H_1: \rho \neq 0$ (T.T.T) i.e; both the random variables x and y are correlation.

Step-3

'α' is called the level of significance usually takes 5% or 1% or 10% and t_α is the table value.

Step-4

The test statistic is given by

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

under H_0:

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

calculate $|t|$ if necessary.

Step-5

If the calculated value of t is greater than the table value of t at α% level of significance then we reject (H_0) otherwise we accept (H_0).

F-test for equality of population variances.

Let us consider,

Two independent random samples $x_1, x_2, \dots x_{n_1}$ and $Y_1, Y_2, \dots, Y_{n_2}$ of sizes n_1 and n_2 drawn from a normal population with variances σ_1^2 and σ_2^2 respectively.

Step-1

Null Hypothesis $H_0: \sigma_1^2 = \sigma_2^2$ i.e; two population variances are equal (or) there is no significant difference between the population variances.

Step-2

Alternative Hypothesis $H_1: \sigma_1^2 \neq \sigma_2^2$ (T.T.T) or $\sigma_1^2 > \sigma_2^2$ (R.T.T) or $\sigma_1^2 < \sigma_2^2$ (L.T.T) i.e; two population variances are not equal (or) there is a significant difference between the population variances.

Step-3

'α' is called the level of significance usually takes 5% or 1% or 10% and F_α is the table value.

Step-4

under H_0,

The test statistic is given by

$$F = \frac{S_X^2}{S_Y^2} \sim F(n_1-1, n_2-1)$$

(or)

$$F = \frac{S_Y^2}{S_X^2} \sim F(n_2-1, n_1-1)$$

where,

$$\left[S_X^2 = \frac{1}{n_1-1} \sum_{i=1}^{n_1} (X-\bar{X})^2\right]$$

$$\left[S_Y^2 = \frac{1}{n_2-1} \sum_{i=1}^{n_2} (Y-\bar{Y})^2\right]$$

step-5

If the calculated value of F is greater than the table value F at α % level of significance then we reject (H_0). otherwise we accept (H_0).

χ^2-test for population variance

Let us consider,

A random sample $x_1, x_2, \dots x_n$ of size 'n' drawn from a normal population with variance 'σ^2'.

Step-1

Null Hypothesis $H_0 : \sigma^2 = \sigma_0^2$ i.e; the sample is drawn from the population variance (or) there is no significant difference between the sample and population.

step-2

Alternative Hypothesis $H_1 : \sigma^2 \neq \sigma_0^2$ (T.T.T) or $\sigma^2 > \sigma_0^2$ (R.T.T) or $\sigma^2 < \sigma_0^2$ (L.T.T) i.e; the sample is not drawn from the population variance (or) there is a significant difference between the sample and the population.

step-3

'α' is called the level of significance usually takes 5% or 1% or 10% and χ^2_α is the table value.

Step-4

under H_0,

The test statistic is given by.

$$\boxed{\chi^2 = \frac{ns^2}{\sigma^2} \sim \chi^2_{n-1}}$$

where,

$$\boxed{s^2 = \frac{1}{n} \Sigma (x - \bar{x})^2}$$

(or)

$$\boxed{\chi^2 = \frac{\Sigma (x-\bar{x})^2}{\sigma^2} \sim \chi^2_{n-1}}$$

Step-5

If the calculated value of χ^2 is greater than the table value of χ^2 at α% level of significance then we reject (H_0), otherwise we accept (H_0).

χ^2-test for goodness of fit

This is the most powerful test for testing the significance of observed and expected values. It is discovered by Karl pearson. Let 'O_i' $[i = 1, 2, \ldots n]$ be the observed frequencies and E_i $[i = 1, 2, \ldots, n]$ be the expected frequencies.

Step-1

Null Hypothesis H_0: the given distribution is to best fit.

Step-2

Alternative Hypothesis H_1: the given distribution is not best fit.

Step-3

α is called the level of significance usually takes 5% or 1% or 10% and χ^2_α is the table value.

Step-4

under H_0,

The test statistic is given by

$$\boxed{\chi^2 = \sum_{i=1}^{n} \frac{(O_i - E_i)^2}{E_i} \sim \chi^2_{n-1}}$$

Step-5

If the calculated value of χ^2 is greater than the table value of χ^2 at α% level of significance then we reject (H_0) otherwise we accept (H_0).

χ^2-test for independent / independence of attributes.

Let us consider,

Two attributes A and B. 'A' is divided into 'r' classes and 'B' is divided into 's' classes [i.e; 'i' is running from $1 \to r$ and j is running from $1 \to s$], and the classification table of $r \times s$ as follows.

	B_1	B_2	......	B_j	......	B_s	
A_1	(A_1B_1)	(A_1B_2)	....	(A_1B_j)	.....	(A_1B_s)	(A_1)
A_2	(A_2B_1)	(A_2B_2)	.....	(A_2B_j)	.....	(A_2B_s)	(A_2)
$\vdots$	$\vdots$						$\vdots$
A_i	(A_iB_1)	(A_iB_2)	.....	(A_iB_j)	.....	(A_iB_s)	(A_i)
$\vdots$	$\vdots$						$\vdots$
A_r	(A_rB_1)	(A_rB_2)	....	(A_rB_j)	....	(A_rB_s)	(A_r)
	(B_1)	(B_2)	.....	(B_j)	.....	(B_s)	N

Let $[O_{ij}]$ be the observed frequency and $E[ij]$ be the expected frequency [i is running from $1 \to r$ and j is running from $1 \to s$].

Step-1

Null Hypothesis H_0: The given attributes A and B are ~~not~~ independent ~~[i.e; dependent]~~.

Step-2

Alternative Hypothesis H_1: The give attributes A and B are not independent [i.e; dependent].

Step-3

'α' is the level of significance usually takes 5% or 1% or 10% and χ^2_α is the table value.

Step-4

under Ho,

The test statistic is given by

$$\boxed{\chi^2 = \sum_{i=1}^{r}\sum_{j=1}^{s} \frac{[O_{ij} - E_{ij}]^2}{E_{ij}} \sim \chi^2_{(r-1)(s-1)}}$$

where,

$$\left[E_{ij} = \frac{\text{Row total} \times \text{column Total}}{\text{Grand Total } (N)} \right]$$

Step-5

If the calculated value of χ^2 is greater than the table value of χ^2, at α% level of significance, then we reject (Ho), otherwise we accept (Ho).

TABLE VALUES.

critical values	5%	1%
Two-Tailed Test	1.96	2.58
Right Tailed Test	1.645	2.33
Left Tailed Test.	-1.645	-2.33