

6) Draw conclusions

If calculated value < Table value then null hypothesis is accepted

If calculated value > Table value then null hypothesis is rejected

Table 1: Important *notations* or *symbols*

Parameter	Sample	Population
Mean	\bar{x}	μ
Standard Deviation	s	σ

2. Steps of Research Methodology

The main steps in research process are:

- Defining the research problem
- Review of literature
- Formulating the Hypothesis
- Preparing the research design
- Data collection
- Data analysis
- Interpretation and Report writing [6]

3. Tools available for testing Hypothesis and Decision Making

3.1 Large Sample tests

When the sample size is $n \geq 30$, then apply large sample tests.

3.1.1. Testing of significance for single proportion:-

Applications: To find significant difference between *proportion* of sample and population

$$Z = \frac{p - P}{\sqrt{PQ/n}}$$

3.1.2 Testing of significance for difference of proportions:-

Applications: To find significant difference between two sample proportions p_1 and p_2

$$Z = \frac{p_1 - p_2}{\sqrt{PQ(\frac{1}{n_1} + \frac{1}{n_2})}}$$

$$P = \frac{n_1 p_1 + n_2 p_2}{n_1 + n_2}, Q = 1 - P$$

3.1.3 Testing of significance for single mean:-

Applications: To find significant difference between *mean* of sample and population

$$Z = \frac{\bar{x} - \mu}{\sigma / \sqrt{n}} \text{ When population S.D. is known}$$

$$Z = \frac{\bar{x} - \mu}{s / \sqrt{n}} \text{ When population S.D. is not known}$$

3.1.4 Testing of significance for difference of means:-

Applications: To find significant difference between two sample means \bar{x}_1 and \bar{x}_2

$$Z = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}} \text{ When population S.D. is known}$$

$$Z = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}} \text{ When population S.D. is not known}$$

3.1.5. Testing of significance for difference of Standard Deviations:-

Applications: To find significant difference between two sample S. D. s_1 and s_2

$$Z = \frac{s_1 - s_2}{\sqrt{\frac{\sigma_1^2}{2n_1} + \frac{\sigma_2^2}{2n_2}}} \text{ When population S.D. is known.}$$

$$Z = \frac{s_1 - s_2}{\sqrt{\frac{s_1^2}{2n_1} + \frac{s_2^2}{2n_2}}} \text{ When population S.D. is not known.}$$

Table 2: Large Sample Tests

Unknown Parameter	One Sample	Two Samples
Proportion	$z = \frac{p - P}{\sqrt{PQ/n}}$	$z = \frac{p_1 - p_2}{\sqrt{PQ(\frac{1}{n_1} + \frac{1}{n_2})}}$, $P = \frac{n_1 p_1 + n_2 p_2}{n_1 + n_2}, Q = 1 - P$
Mean	$z = \frac{\bar{x} - \mu}{\sigma / \sqrt{n}}$ Or $z = \frac{\bar{x} - \mu}{s / \sqrt{n}}$	$z = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}}$ Or $z = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}}$
S.D.	—	$z = \frac{s_1 - s_2}{\sqrt{\frac{\sigma_1^2}{2n_1} + \frac{\sigma_2^2}{2n_2}}}$ Or $z = \frac{s_1 - s_2}{\sqrt{\frac{s_1^2}{2n_1} + \frac{s_2^2}{2n_2}}}$

3.1.6 Chi Square Test

It is an important test amongst various tests of significance and was developed by Karl Pearson in 1900. It is based on frequencies and not on the parameters like mean, S.D. etc.

Applications: Chi Square test is used to compare observed and expected frequencies objectively. It can be used (i) as a test of goodness of fit and (ii) as a test of independence of attributes.

Conditions for applying χ^2 Test:-

- The total number of items N must be at least 50.
- No expected cell frequency should be smaller than 10. If this type of problem occurs then difficulty is overcome by grouping two or more classes before calculating (O-E).

3.1.6. (a) Chi Square Test As a test of goodness of fit:-

Chi square test enables us to see how well does the assumed theoretical distribution (such as Binomial distribution, Poisson distribution or Normal distribution) fit to the observed data.

Formula: $\chi^2 = \sum \frac{(O_{ij} - E_{ij})^2}{E_{ij}}$

O_{ij} = observed frequency of the cell in ith row and jth column.

E_{ij} = expected frequency of the cell in ith row and jth column.

Degree of freedom=
 n-1 (For Binomial Distribution)
 n-2 (For Poisson Distribution)
 n-3 (For Normal Distribution)

Where n= total no. of terms in a series

Formula: $t = \frac{\bar{x} - \mu}{s / \sqrt{n}}$ when s.d. is given

When S.D is not given then find s by using the formula $s^2 =$

$$\frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})^2$$

Degree of freedom= n-1

3.2.1.(b) t-test for difference of means of two samples:-

Applications: It is used to compare the mean of two samples of size n_1 and n_2 when population variances are equal. [9].

Formula: $t = \frac{\bar{x} - \bar{y}}{s \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}}$ when s.d. of two samples is known

And $s^2 = \frac{n_1 s_1^2 + n_2 s_2^2}{n_1 + n_2 - 2}$

When S.D is not given then find s by using the formula $s^2 =$

$$\frac{1}{n_1 + n_2 - 2} \sum_{i=1}^n (x_i - \bar{x})^2$$

Degree of freedom= $n_1 + n_2 - 2$

3.1.6. (b) Chi Square Test As a test of independence of attributes:-

χ^2 test enables us to explain whether two attributes are associated or not.

e.g. It may help in finding whether a new drug is effective in curing a disease or not

Formula: $\chi^2 = \sum (O - E)^2 / E$

Where 'O' represents the observed frequency. E is the expected frequency under the null hypothesis and computed by

$$E = \frac{\text{row total} \times \text{column total}}{\text{sample size}}$$

Here degree of freedom= (r-1) (s-1)

3.2.1.(c) Paired t-test:-

Applications:- A paired samples t-test is used when size of two samples is equal and is used to compare two related means. [2]

Formula:- $t = \frac{\bar{d}}{s / \sqrt{n-1}}$

Where \bar{d} = Mean of paired differences

and $s^2 = \frac{\sum d^2}{n} - (\bar{d})^2$ where $d = x_i - y_i$

3.1.6.(c) Difference in test of goodness of fit and independence of attributes

For the goodness-of-fit test, a theoretical relationship is used to calculate the expected frequencies. For the test of independence, only the observed frequencies are used to calculate the expected frequencies. [4]

3.2.2. Snedecor's Variance Ratio Test or F- Test

The name was coined by George W. Snedecor, in honour of Sir Ronald A. Fisher. Fisher initially developed the statistic as the variance ratio in the 1920s.

3.2 Small Sample tests

When the sample size is $n < 30$ then apply small sample tests.

Applications:

i) F-test is used to test whether the two samples are from the same normal population with equal variance or from two normal populations with equal variances.

ii) To test whether the two independent samples have been drawn from the same population, we test:

- A) Equality of means by t-test
- B) Equality of population variance by F-test

But as t-test assumes that the sample variances are equal, we first apply F-test and then t-test. [10].

3.2.1. Student's t-test

t statistic was developed by William S. Gossett and was published under the pseudonym Student.

Formula:- $F = \frac{\sigma_1^2}{\sigma_2^2}$ where numerator is greater than denominator

When variance is known

$$\sigma_1^2 = \frac{n_1 s_1^2}{n_1 - 1}, s_1^2 = \text{Variance of sample 1}$$

$$\sigma_2^2 = \frac{n_2 s_2^2}{n_2 - 1}, s_2^2 = \text{Variance of sample 2}$$

3.2.1.(a) t-test for the mean of a random sample:-

Applications:- It is used to test whether the mean of a sample deviates significantly from a stated value when variance of population is unknown.

When variance is not known

$$\sigma_1^2 = \frac{\sum(x_1 - \bar{x}_1)^2}{n_1 - 1} \quad \sigma_2^2 = \frac{\sum(x_2 - \bar{x}_2)^2}{n_2 - 1}$$

3.3 Analysis of Variance (ANOVA)

ANOVA was developed in the 1920's by R.A. Fisher.

Applications: This technique is used when multiple sample cases are there. The significant difference between the means of two samples can be tested through t-test, but the difficulty arises when we are to find the significant difference amongst more than two sample means at the same time. [3]

Table 3: One way classification ANOVA

Source of variation	Degree of freedom(d.f.)	Sum of squares (s.s.)	Mean squares (m.s.)	F
Between rows	$v_1 = h - 1$	Calculate as in (i) below	s.s./d.f.	$F_1 = \frac{\text{m.s. between rows}}{\text{Error m.s.}}$
Between columns	$v_2 = k - 1$	Calculate as in (ii) below	s.s./d.f.	$F_2 = \frac{\text{m.s. between columns}}{\text{Error m.s.}}$
Error	$v = (h-1)(k-1)$	Calculate as in (iii) below	s.s./d.f.	_____

Here (i) Sum of squares between samples = $\frac{(\sum X_1)^2}{N} + \frac{(\sum X_2)^2}{N} + \frac{(\sum X_3)^2}{N} + \frac{(\sum X_4)^2}{N} - \frac{T^2}{N}$

If X_1, X_2, X_3, X_4 are samples.

And correction factor = $\frac{T^2}{N} - \frac{(\sum X_1 + \sum X_2 + \sum X_3 + \sum X_4)^2}{N}$

(ii) Sum of squares with in samples = Total sum of squares - Sum of squares between samples

Where total sum of squares = $\sum X_1^2 + \sum X_2^2 + \sum X_3^2 + \sum X_4^2 - \frac{T^2}{N}$

Table 4: Two way classification ANOVA

Source of variation	Degree of freedom (d.f.)	Sum of squares (s.s.)	Mean squares (m.s.)	F
Between Samples	$v_1 = h - 1$	Calculate as in (i) below	s.s./d.f.	$F = \frac{\text{m.s. between samples}}{\text{m.s. with in samples}}$
Within samples	$v_2 = N - h$	Calculate as in (ii) below	s.s./d.f.	Where Num. > Denom.

(i) Sum of squares between rows = $\frac{1}{k} \sum T_i^2 - \frac{T^2}{N}$ where $\sum T_i^2 =$ sum of squares of rows total

(ii) Sum of squares between columns = $\frac{1}{h} \sum T_j^2 - \frac{T^2}{N}$ where $\sum T_j^2 =$ sum of squares of columns total

(iii) Error sum of squares = Total sum of squares - (Sum of squares between rows) - (Sum of squares between columns)

Where total sum of squares = $\sum X_1^2 + \sum X_2^2 + \sum X_3^2 + \sum X_4^2 - \frac{T^2}{N}$

4. Conclusion

With the correct use of above discussed tests, valid results can be found. So precaution should be taken while selecting the tests of hypothesis for large and small sample tests otherwise one get invalid results. That is why selection of a correct statistical test is much important.

References

- [1] Rinoj P.K., 2014: Role of Research Methodology in Academic projects of Management Students, Volume-3, Issue-7, ISSN No 2277 – 8160
- [2] Evie McCrum-Gardner, 2007: Which is the correct statistical test to use?, British Journal of Oral and Maxillofacial Surgery 46 (2008) 38–41
- [3] Kothari C.R., 2004: Research Methodology Methods and Techniques, 2nd Revised Edition .New Delhi, New Age International Publishers.
- [4] Minhaz Fahim Zibran , CHI-Squared Test of Independence, Department of Computer Science , University of Calgary, Alberta, Canada
- [5] Bhattacharya, Dipak Kumar Research Methodology, New Delhi, Excell books.
- [6] Pannerselvam, R. 2014: Research Methodology, New Delhi. Prentice Hall of India Pvt Ltd.
- [7] Bali N.P., Gupta P.N., Gandhi C.P., 2008: Quantitative Techniques, New Delhi, and University Science Press.
- [8] Cochran W.G., 1963: Sampling Techniques, New York, John Wiley & Sons.
- [9] Chance, William A. 1975: Statistical Methods for Decision Making, Bombay, D.B. Taraporevala sons & Co. Pvt. Ltd.
- [10] Chaturvedi, J.C., 1953: Mathematical Statistics, Agra, Nok Jhonk Karyalaya.

Author Profile



Joginder Kaur is working as a Lecturer (Mathematics) in a college affiliated to Punjab Technical University, Punjab. She received her M.Sc. in Mathematics in 2007 from Punjab University, Chandigarh. Her interested areas of research are Probability and Statistics, Complex Numbers and Numerical Methods.