LOVELY PROFESSIONAL UNIVERSITY

EDUCATIONAL STATISTICS

Edited By
Dinesh Kumar

Printed by
LAXMI PUBLICATIONS (P) LTD.
113, Golden House, Daryaganj,
New Delhi-110002
for
Lovely Professional University
Phagwara

SYLLABUS

Educational Statistics

Objectives

- To understand the basic concepts of research, research processes and designs.
- To understand the techniques of selecting different types of research problems.
- To acquire the skills of preparing the research proposal.

Sr. No.	Content
1	Statistical methods: frequency distribution and graphical representation, Measurement of central tendency -Mean, Median, Mode, Measures of variability SD, MD, QD. -
2	Hypothesis–concept, characteristics, sources, Formulation of hypotheses, testing of hypotheses, Quantitative data collection, scoring, analysis and interpretation, Qualitative data – collection, scoring, analysis and interpretation, Preparation of research report and dissemination of findings
3	Probability: Normal Probability Curve and its Uses.
4	Correlation-rank difference, product moment method
5	Chi-square test, T-test, F-test.

Content

Unit 1: Statistical Methods 1
Dinesh Kumar, Lovely Professional University

Unit 2: Measurement of Central Tendency—M, Md, Mo 7
Kulwinder Pal, Lovely Professional University

Unit 3: Hypothesis : Concepts, Characteristics and Sources 23
Kulwinder Pal, Lovely Professional University

Unit 4: Formulation and Testing of Hypothesis 33
Dinesh Kumar, Lovely Professional University

Unit 5: Qualitative and Quantitative Data 38
Dinesh Kumar, Lovely Professional University

Unit 6: Preparation of Research Report 46
Kulwinder Pal, Lovely Professional University

Unit 7: Probability : Normal Probability Curve and its Uses 64
Dinesh Kumar, Lovely Professional University

Unit 8: Measurement of Dispersion QD, MD, SD 81
Dinesh Kumar, Lovely Professional University

Unit 9: Correlation : Rank Difference Method, Product Moment Method 97
Kulwinder Pal, Lovely Professional University

Unit 10: Chi-Square Test 113
Dinesh Kumar, Lovely Professional University

Unit 11: T-Test 123
Dinesh Kumar, Lovely Professional University

Unit 12: F-Test 137
Kulwinder Pal, Lovely Professional University

Notes

Unit 1: Statistical Methods

CONTENTS

Objectives

Introduction

1.1 Development, Meaning and Definition of Statistics

1.2 Statistical Terms

1.3 InitialStep in Statistics

1.4 Frequency- distribution Table

1.5 Summary

1.6 Keywords

1.7 Review Questions

1.8 Further Readings

Objectives

After studying this unit, students will be able to:

- Understand the development, meaning and definition of statistics
- Understand statistical terms and initial steps
- Understand frequency – distribution table.

Introduction

Sankhyiki in English is called STATISTICS. It seems that statistics is derived from Latin language's term STATUS or from Italian STATISTA. **Yule and Kendall** has said that origin of this word is STATUS. The meaning of both STATUS and STATISTA had been political STATE in ancient times. Use of these words had been done in the context of political situation. In ancient times king had to keep records of different types of informations regarding his state as his own power, population of state, birth rate, death rate, state's revenue and expenditure etc., this types of State's description called statistics gradually. In this way science of statistics was incepted.

1.1 Development, Meaning and Definition of Statistics

With the induction of statistics word this matter is clear that the use of this science in the beginning numbers of birth-death, income-expenditure numbers, and state's population description's record had been done. But later on as the state's political, social and economic condition become more complicated.

In the course of time statistics not only used as a political subjects but also it spread upto state's economic and social problems and conditions. In this way statistics term modern meaning seems to be

Notes

not more than 100 years old. In seventeenth centuries including insurance scheme the use of statistics was increased to a fair degree also and again it will be used in the all fields of knowledge. Now a days statistics is uses in Science, Mathematics, Education Psychology, Economics etc. in all subjects.

Scientist has given definition from a number of types. **W. G. Sutcliffe** in his book, 'Elementary Statistical Methods' had been written the definition-

"Statistics comprises the collection, tabulation, presentation and analysis of an aggregate of facts, Collected in a methodical manner without bias and related to predetermined purpose."

Notes According to **A. L. Bowley,** "Statistics may be called as the science of counting.

According to Boddington-"Statistics is the science of estimates and probabilities"

According to Lovitt – " Statistics is the science which deals with the collection, classification and tabulation of numerical facts as the insist for explanation, description and comparision of phenomena"

According to **Ferguson** "Statistics deals with collection, Classification description and interpretation of data obtained by the conduct of surveys and experiments.

Self Assessment

Fill in the blanks:

1. According to Boddington, Statistics estimation and probabilities is .....
2. Scattered number or numbers origin is called .....
3. In one class interval insert number of digits class interval is called .....

1.2 Statistical Terms

Some important terms which is used in statistics are following-

Data – Scattered digits or numbers are called basic data. Basic data from a types from statistical point of view are understood as compulsory, but on its basis we can not draw any conclusion when it could not classified or analyzed properly.

Example Basic data are the numeral form of our Physical or Psychological observations as a intelligence test, examination marks, height, weight etc.

Ungrouped Data – When basic data is obtained on a short sample then at that condition classification can not considered as essential and calculation can do directly, as a 10, 12, 20, 13, 17, 15, 8, 6, 16, 14 (N = 10) This type of arrangement of data are called un-classified data (N = 30).

Grouped Data – When data is obtained on a big sample (N > 30), then is such a condition classification is essential and this type of management of data is called grouped data.

Range – The difference get from the highest number to lowest number, is called range or expansion e.g.,

Range = Maximum score – Minimum Score

Notes

Class-Interval – From range, distribution of data's correct information cannot finds.

Therefore this range again divided into sub range and this sub range is called class – interval.

Number of C. I. – Internal can be calculated from following formula

$$\text{No. of C.I.} = \frac{\text{Range}}{\text{Size of C.I.}} + 1$$

Size of C. I. - In a class interval inherited number of digit is called length of class interval of limit of class interval. Actual digits can be calculated by inclusive and exclusive series 'basis'.

Mid-point of C.I. – The mid point of class interval is find out by adding maximum and minimum score and it divided by 2.

Did You Know?

$$\text{Mid Point of C.I.} = \frac{\text{Max. Score + Mini. Score}}{2}$$

Exact Limits of C.I. – Statistical scale is differ from mathematical scale. In mathematical scale absolute zero exists and in mathematics the meaning of 5 is only 5 where as in statistical scale due to absence of absolute zero the meaning of 5 is not only 5 rather. Then it varies from 4.5 to till 5.5 is. In this way in statistics to know the exact limit of any number .5 is deducted from lower limit and .5 added to upper limit. This is the reason that statistical scale is consider less reliable than mathematical scale.

Therefore to find out the exact limit of a class interval, .5 is subtracted to lower limit and in upper limit .5 is added , for example class interval 3-5 exact limit will be 2.5-5.5.

Frequency- Which numbers repeats how many times or any number is repeated how many times, that is called their frequency. For example, 10, 12, 15, 10, 18, 22, 10, 25 in number 10 is repeated thrice, hence frequency of 10 will be called 3. It is expressed by letter *f* or *F*.

Frequency Distribution- On the basis of size and magnitude of numbers, arrangement process is called frequency distribution.

Tallies or Tally Marks – Tally mark infact in an specific class interval lies, shows the actival numbers.

After completion of counting up to 4 to get 5 first four is crossed by a transversal line as ~~||||~~ (show 5 counting)

Series-The systematic form of class interval is called series. There are two types of series – exclusive and inclusive. In exclusive series the upper limit of a class interval is only formality, which we not counted in that, clouted in next class interval. In other words in exclusive series upper limit is exclusive series upper limit excluded where as in inclusive series both limits of class intervals is assumes under that class intervals. For example,

Table 24.1

Exclusive series	Inclusive series
40-50	40-49
30-40	30-39
20-30	20-29
10-20-	10-19
0-10	0-9

Notes

As far as the question of arrangement of series is concerned we can write it in two orders ascending and descending. In ascending order we write as 1, 2, 3 5 in descending order we write as 6, 5, 4, 3, 2, 1,

Use of statistics in education—Importance of statistics in education is due to following courses-

1. Exact description
2. Summarization- simple and meaningful
3. To make data simple and conceivable
4. Helpful in drawing simple conclusion
5. Helpful in forecasting
6. To get knowledge about relation
7. Helpful in comparative study
8. To find out course effect
9. Helpful in observation formulation
10. Helpful in sorting the problems.

Task Explain the importance of statistics is education.

Self Assessment

Multiple Choices Questions:

4. Differencing getting by lowest number from highest number is called _____.

   (a) Range (b) Expansion (c) Class (d) Limit

5. Statistical scale is differ from _____ scale.

   (a) Class interval (b) Frequency (c) Mathematical (d) Fixed

6. The process of systemization of numbers according to their size and nature is called _____ distribution.

   (a) Series (b) Counting (c) Recurring (d) Removal

7. The Systematic form of class interval is called _____.

   (a) Series (b) Discrete (c) Marks (d) Ascending

1.3 Initial Step in Statistics

Preparing a frequency distribution Table—Illustration—Prepare a frequency distribution table from following data. Keep class interval range 5.

90, 93, 95, 98, 96, 95, 97, 101, 100, 99, 104, 102, 103, 102, 104, 103, 105, 107, 106, 105, 106, 107, 105, 107, 106, 110, 109, 108, 110, 102 (N = 30)

1.4 Frequency–distribution Table

Solution– Minimum marks = 90

Maximum marks = 120

. Range = 120–90 = 30

$$\text{Number of class interval} = \left(\frac{\text{Range}}{\text{Size}} + 1\right) = \frac{30}{5} + 1 = 6 + 1 = 7$$

Table 24.2

S. No	*C.I.*	*Tallies*	*Frequencies*
1.	90-94	||	2
2.	95-99	卌 |	6
3.	100-104	卌 |||	8
4.	105-109	卌 卌 |	11
5.	110-114	||	2
6.	115-119	0	0
7.	120-124	|	1
	Total		N = 30 (CHECK)

Steps Involved in Preparing Frequency Distribution Table-

To prepare frequency distribution table should be determined following steps-

1. First of all select minimum and maximum marks from given data. Then encircled these marks separately.
2. Get range (by the help of formula) e.g. subtracted minimum marks from maximum marks.
3. If class interval range is not given, then determine arbitrarily. Take care so that rage is such type that taking it number of class interval did not exceed ten (approx view). From convience point of keep class interval between 5-10.
4. Determine number of class.
5. It is desirable to write class in table (exclusive or inclusive) by ascending or descending order. First class will be begun from minimum marks and in that fix limit will be included.
6. Now take one by one marks. Marks belongs to that class before that a vertical pi (tally marks) drawn. After completion of four, for fifth counting drawn cross transversal line over that completed block.

Caution Continued tally marks till maximum marks table prepared.

7. In frequency column write tally marks after counting.
8. Make total of all frequencies.
9. Sum total of all frequencies must be equal to number of data. This is test of suitability of table.

Notes

1.5 Summary

- In English shankhiki is called statistics. It seems that statistics term derived from STATUS of Latin language or STATISTA from Italian prior this word had been used in political contest.
- Statistics comprises the collection, tabulation, presentation and analysis of an aggregate of facts, collected in a methodical manner without bias and related to predetermined purpose.
- Scattered marks or numbers are called fundamental data. Basic data are considered to be a compulsory material from statistical process point of view.
- In a class interval included number is called range of class interval or limit of class interval.
- Determine actual number of marks on inclusive or exclusive series basis.

1.6 Keywords

1. **Frequency**—In given data the number that comes regularly, which has more frequency.
2. **Exclusive**—To eliminate, remover.

1.7 Review Questions

1. Prepared frequency distribution table from following data. Keep range of class interval 10.
   135, 139, 138, 110, 116, 119, 122, 135, 134, 126, 127, 128, 131, 139, 140, 150, 145, 184, 148, 155, 154, 160, 158, 165, 170, 175, 180, 179, 190, 160, 170, 155, 148, 149, 123, 186, 117, 118, 119, 120 (N=40).
2. Prepared frequency distribution table from following table. Keep class interval range 3.
   18, 24, 10, 12, 20, 28, 20, 14, 18, 114, 18, 14, 18, 20, 24, 14, 18, 20,20, 12 (N = 20).
3. Having taken examination of high school class of 30 students whore marks are following with the help of this prepared distribution frequency of tables keep class 10.
   38, 36, 80, 82, 85, 28, 30, 20, 90, 72, 75, 58, 42, 36, 75, 72, 68, 65, 70, 66, 48, 50, 10, 63, 57, 55, 62, 61, 50, 62. (N=30).
4. With the help of following data prepared distribution frequency table in which keep range 8.
   16, 18, 18, 11, 15, 5, 20, 18, 11, 25, 30, 15, 17, 11, 16, 22, 25, 20, 22, 15, 30, 32, 18, 7, 20, 25, 22, 22, 35, 37, 40, 20, 11, 25, 29, 10, 15, 35, 50, 10. (N=40).

Answers: Self Assessment

1. Science	2. data	3. Limit	4. (a) Range
5. (c) Mathematical	6. (b) Frequency	7. (a) Series	

1.8 Further Readings

Books

1. Educational Techniques and Evaluation—*Dr Rampal Singh, Bhatt Brothers.*
2. Education Technique—*R. A. Sharma, Bhatt Brothers.*
3. Education Technique—*S. S. mathur, Bhatt Brother.*
4. Educational Research Methods—*Shareen and Shashi kala, Vinod Pustak Mandir.*

Notes

Unit 2: Measurement of Central Tendency—M, Md, Mo

CONTENTS

Objectives

Introduction

2.1 Meaning and Definition of Central Tendency

2.2 Mean : M

2.3 Median : Md

2.4 Mode : Mo

2.5 Assumptions Underlying M, Md and Mo

2.6 Summary

2.7 Keywords

2.8 Review Questions

2.9 Further Readings

Objectives

After studying this unit, students will be able to:

- Understand the meaning and definition of central tendency
- Understand mean, median and mode
- Understand assumptions underlying of central tendency.

Introduction

Having collected data in process of research from which we can not reach on any conclusion. For that classification tabulation of data is necessary. From the series of facts we have to find out a number which can be done suitable representative of that series. That value which express ability of group precisely is one number, is called measure of central tendency.

2.1 Meaning and Definition of Central Tendency

During studies of data researcher interested in that matter that he has to taken help of only one measure to know the tendency of his group. This type of assumption is known as **central tendency**. To clarify the meaning of central tendency suppose, five students marks are 10, 11, 9, 12 and 13. By observation these marks certainly differ from one another, rather than is this a type of central tendency to be seemed. The meaning to say is that amount five marks one marks is such that around which rest four marks are cantered.

Notes

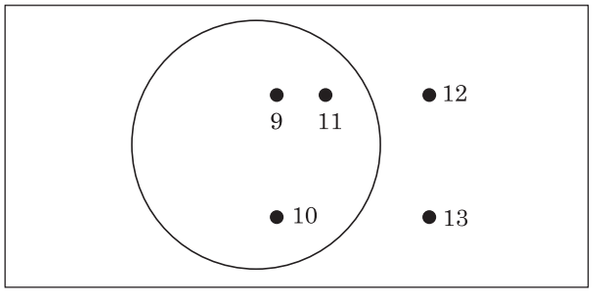

Fig. 2.1

In the above figure if we set marks 11 in the centre of circle and rest four marks around that then it seems that these four marks close to the marks situated in the centre or four marks come towards the marks situated in the centre. In simple language we can say that some marks are less than 11 and some are greater then 11. In statistical language this tendency of data is called central tendency.

Ross has written clearly definition of central tendency

"It is the tendency of the scores to bunch or concentrate somewhere mean the centre. It is that value which typifies or best represents the whole distribution" – ***Ross, C.C.***

Measures of Central Tendency—Garrett has written regarding necessity and importance of central tendency measures-

"The value of a measure of central tendency is two fold. Firstly it is a average which represents all of the scores made by the group and as such gives a concise description of the performance of the group as a whole and secondly it enables us to compare two or more groups in terms of typical performance."

—Garrett

There are three measures of central tendency.

1. Mean
2. Median
3. Mode

2.2 Mean: M

Definition- What is called mean in statistics, in normal mathematics it is called average. Symbolically mean is represent by sign, M. The Simplest method to calculate mean is sum of number is divided by their numbers. Some main definitions are as follows:

Notes Ferguison— "Mean is the sum of a set of measurements divided by the numbers of measurements of the set."

John Griffin - "In statistical usage since average is a single values used to represent all the values of a set of observations, an average may by thought of as a measures of central value".

F. C. Mills - The arithmetic mean is the centre of gravity of a "distribution."

Computation—Related to mean two types of data may be-

1. Ungrouped data
2. Grouped data

Notes

Ungrouped Data

Example: Compute mean from following data

7, 10, 8, 13, 11, 14, 9, 9, 13, 15

Solution - In ungrouped data following formula to get mean

$$M = \frac{\Sigma X}{N}$$

Where M = Mean

ΣX = Sum of datas

N = Number of datas

According to question $\Sigma X = 7 + 10 + 8 + 13 + 11 + 14 + 9 + 9 + 13 + 15 = 109$

N = 10

Hence, $M = \frac{\Sigma X}{N} = \frac{109}{10} = 10.9$ **Ans.**

Important Stage

1. Calculate sum of all marks
2. Calculate total number of marks
3. Put the value in formula and calculate answer.

Grouped Data: Questions related to grouped data can be solved by two methods'

1. Long method
2. Short method

(1) Long Method – Under this method following formula is uses:-

$$M = \frac{\Sigma fx}{N}$$

Where, M = mean

Σfx = Product of marks and frequencies.

N = Sum of Frequencies

Example: By long method calculate mean of following data.

Table 2.1

S. No.	*C.I.*	*Mid Point (X)*	*f*	*fx*
1.	55–59	57	1	57
2.	50–54	52	1	52
3.	45–49	47	3	141
4.	40–44	42	4	168
5.	35–39	37	6	222
6.	30–34	32	7	224
7.	25–29	27	12	324
8.	20–24	22	6	132
9.	15–19	17	8	136
10.	10–14	12	2	24
Total			N=50	Σfx =1480

Notes

Putting the value in formula,

$$M = \frac{\Sigma fx}{N} = \frac{1480}{5} = 29.60 \textbf{ Ans.}$$

Important Stage

1. Find the mid value (X) of each class interval.
2. Multiply mid point (X) to frequencies (*f*)
3. In this way find sum of all *fX*.
4. To obtained N sum all frequencies.
5. Putting the obtained value in formula calculate mean.

(2) Short–cut Method or Assumed Mean Method—In this method following formula is used.

$$M = A.M. + \frac{\Sigma fd}{N} \times C.I.$$

Where, M = Actual mean

A.M. = Assumed mean

Σfd = Sum of total frequencies

C.I. = Range of class interval

Illustration: Shortcut method to find mean

Table: 2.2

S. No.	*C.I.*	*f*	**d**	*fd*
1.	90–99	1	+3	3
2.	80–89	4	+ 2	8
3.	70–79	7	+ 1	7
4.	60–69	10	0	0
5.	50–59	9	–1	–9
6.	40–49	9	–2	–18
7.	30–39	4	–3	–12
8.	20–29	3	–4	–12
9.	10–19	3	–5	–15
Total		N=50		$\Sigma fd = -48$

Putting the value in formula

$$M = A.M. + \frac{\Sigma fd}{N} \times C.I..$$

$$= 64.5 + \left(\frac{-48}{50}\right) \times 10$$

$$= 64.5 - \frac{48}{5}$$

= 64.5 – 9.6

= 54.9 **Ans.**

Important Stage

1. Prepared frequency distribution table.
2. Select assumed mean.

Caution That class interval is selected for assumed mean which lies exactly in the middle or which class interval frequency is most.

(This class interval should be lies in middle also.)

3. Just before the assumed mean group put 0 (Zero) in deviation column and greater than this group is given positive value (+1, +2, +3… and less group is given negative value (-1, -2, -3…) respectively.
4. Find the product of frequencies and deviations.
5. In this way find the sum of all *fd*, pay attention to sign also.
6. Range of class, obtain the value of A.M. and N, putting this value in formula calculate mean.

Self Assessment

Fill in the blanks:

1. Mean is that value which obtained from the sum of numbers ..... by numbers.
2. Central tendency have three measures mean, median, and .....
3. ..... related can be two data-grouped and ungrouped datas.

2.3 Median: Md

Definition

Prof. Bowley: "The median is the point that divides a ranked distribution of measurements into two equal parts."

Lindquist: "Median is that point on the scale of scores below which one half of the scored (50%) lie above which one half of the scores (50%) lie."

Computation: Median is calculated from two types of datas.

1. Ungrouped Data
2. Grouped Data

Ungrouped datas: Although, to find out the median of an ungrouped data, no need of any formula, though, from convenient point of view $\left(\frac{N+1}{2}\right)$ th term can be used.

Case I: When numbers of measures is odd.

Example: Find out the median of following data.

Notes

23, 23,22, 20, 19, 17, 16, 15, 15, 25, 28, 13, 19

Solution: Re-arranging the numbers in ascending order again we get

13, 15, 15, 16, 17, 18, 19, 20, 22, 23, 23, 25, 28

By formula, Median $= \left(\frac{N+1}{2}\right)$ th term

$= \left(\frac{13+1}{2}\right)$ th term

$= \left(\frac{14}{2}\right)$ th term

$= 7^{th}$ term

= 19 (To get seventh term count from left to right or right to left)

Hence Median = 19 **Ans.**

When numbers of measure are even.

Example : Find out median of following datas.

34, 37, 46, 42, 33, 44, 30, 40, 35, 33

Solution in ascending order of the numbers, we get

30, 33, 33, 34, 35, 37, 40, 42, 44, 46

Now Median $= \left(\frac{N+2}{2}\right)$ th term

$= \left(\frac{10+1}{2}\right)$ th term

$= \left(\frac{11}{2}\right)$ term

$= 5.5^{th}$ term

$= \left(\frac{\text{fifth term} + \text{sixth term}}{2}\right)$

$= \left(\frac{35+37}{2}\right)$

$= \left(\frac{72}{2}\right)$

= 36

Hence Md = 36 **Ans.**

Important stages:

1. Arranged the ascending or descending order of given data.
2. Find out the total number (N) of data.
3. Calculate median by applying formula.
4. To find out the median of even numbers taken two successive term and do half the sum of that.

Notes

Grouped Data: To find out the median of grouped data related questions following formula is used:-

Did You Know?

$$Md = L + \left(\frac{\frac{N}{2} - f_b}{f}\right) x C.I.$$

Where Md = Median Value

L = Lower limit of that class in which our median value lies.

N/2 = Half of all frequencies.

f_b = The sum of all classes frequencies which lies below median group.

f = Frequency of median group.

C.I. = Limit of class interval

Example: From following table calculate median

Table 2.3

S. No.	*C.I.*	*f*	*cf*	
1.	90–99	2	60	
2.	80–89	6	58	
3.	70–79	5	52	
4.	60–69	11	47	
5.	50–59	20	36	← **Median group**
6.	40–49	12	16	
7.	30–39	3	4	
8.	20–29	1	1	
Total		N=60		

N/2 = 30

Now, $$Md = L + \left(\frac{\frac{N}{2} - f_b}{f}\right) \times C.I.$$

$$= 49.5 + \left(\frac{30-16}{20}\right) \times 10$$

$$= = 49.5 + \left(\frac{14}{20}\right) \times 10$$

= 49.5+7

= 56.5

Hence, Md = 56.5 **Ans.**

Notes

Test- Our answer must be lies in median group. Here 56.5 which lies in 50-59 that is median group. Hence, our answer is correct.

Important Stages:

1. Arranged frequency distribution table.
2. Calculate cumulative frequency. There are two methods of H. first methods is that the frequency before lowest class interval that is written in of column as it is adding to its next continuously. Second method is that the frequency of highest group is written in of column as it is deducting continuously. Paid a attention that the sum of all frequencies comes in highest class interval, by whatever methods we calculate *cf*.
3. Do half of frequencies and observed that in column where *cf* comes. This will be your median group and encircled it.
4. Find out actual lower limit of median group interval.
5. Please set systemically. Find out class interval limit.
6. Now, all these values, e.g. f, f_b, $\frac{N}{2}$, L and C.I.
7. Check the verification of correctness of answer.

Important Case I: When there are gaps in distribution and N/2 lies in more than one C.I.

At that condition applicable formula is following.

$$Md = \frac{Md_1 + Md_2}{2}$$

Where Md = Median value

Md_1 =That class interval when for observation of N/2, counting from top to below.

Md_2 = *That class* interval when +we to find out N/2 from below to top.

Example: From following table calculate median.

Table 2.4

S. No.	*C.I*	*f*	*cf*	
1.	90-99	2	10	
2.	80-89	1	8	
3.	70-79	0	7	
4.	60-69	0	7	
5.	50-59	2	7	← Md_1
6.	40-49	0	5	
7.	30-39	0	5	
8.	20-29	3	5	← Md_2
9.	10-19	2	2	
Total		N=10		

$$N/2 = 5$$

$$Md = \frac{Md_1 + Md_2}{2}$$

Notes

Now, Md_1 = 50-59

Md_2 = 50-29

Now combining these two-class interval one class interval. One class interval will be made which is 29-59 then it changed into actual class interval. Subtract 0.5 from lower limit and add 0.5 in upper limit. In this way our class interval limits will be 19.5-59.5.

Hence, Median $= \frac{19.5 + 59.5}{2}$

$= \frac{79.0}{2}$

= 39.5 **Ans.**

Important stages

1. Prepare frequency distribution table.
2. Complete the commutative frequency column.
3. Observe that N/2 lies in which class. If it lies more then one class then calculate Md_1 and Md_2.
4. To get Md_1 in frequency column form top to bottom calculate N/2 and in that class it falls, encircled it.
5. Similarly find Md_2 counting form bottom to top.
6. Now, combining these class make one large class and write it in actual limitations.
7. Putting these values in formula computed median value.

Important case II: When N/2 lies in one C. I. only, but gaps are still there.

Step: At that condition, having combined class to one another less their number and again basic formula can be used. In the following example rather (N/2) e.g. C. I. lies in 40-44, but some class as a 35-39 and 30-34 before zero frequencies still remain. Hence, here including two class make one class. Clearly to do so frequencies will be added and zero frequencies will be removed.

After following this process, basic formula is used.

Example : Calculate median from following data:

Table 2.5

S No.	C.I.	*f*	*cf*	
1.	45-49	5	16	
2.	40-44	4	11	← Median
3.	35-39	0	7	
4.	30-34	0	7	
5.	25-29	3	7	
6.	20-24	2	4	
7.	15-19	2	2	
Total		N = 16		

In this question, since N/2 lies in one C. I, still before some class zero are remain. Hence to eliminate zero having combined classes and make less their number. In this way the transformed form of table is following.

Notes

Table 2.6					
S.No.	*C.I.*		*f*	*cf*	
1.	45-54		5	16	
2.	35-44	4	10	11	← Median
3.	25-34	3	4	7	
4.	15-24	4	4	4	
Total			N=16		

N/2 = 8

Now, $Md = L + \frac{N/2 - f_b}{f} \times C.I.$

$= 34.5 + \frac{8-7}{4} \times 10$

$= 34.5 + \frac{1}{4} \times 10$

= 34.5 + 2.5

= 37.0

Hence, Md. = 37 **Ans.**

Self Assessment

State whether the following statements are True or False:

4. Median does not divide distribution in equal parts.
5. Median is written as Md also.
6. C.I. = Range of class interval is their length.
7. Mean is the centre of gravity of any distribution.

2.4 Mode : Mo

According to **Crow** and **Crow**: "The score in a given series of data that appears most frequently is called the mode and is represented by the symbol Mo."

Kenny: "Mode is the most frequently occurring value in the series."

Guilford: - "The mode is strictly defined as the point on the scale of measurement with maximum frequency in a distribution.'

Computation: Mode value can be calculated two types of data also-

1. Ungrouped data
2. Grouped data

Ungrouped Data

Example : Calculate mode from following data.

8, 9, 9, 13, 14, 13, 17, 16, 17, 16, 18, 20, 17

Notes

Solution: Here, 17 is evidently the most often recurring measures and is therefore the crude or empirical mode.

Here, 17 is repeated thrice, hence it is the mode value of distribution.

Grouped Data - To find out the mode of classified data following formula is used.

$$Mo = L + \frac{f_a}{f_a + f_b} \times C.I.$$

Where, Mo = Mode value

L = Lower limit of highest frequency class

f_a = Frequency of that class which lies above modal group.

f_b = Frequency of that class which lies just below the modal group.

C.I. = Range of class interval.

Task Write a short note on mode.

Example: Calculate the mode from the following data.

Table 2.7

S. No.	*C.I.*	*f*	
1.	123-127	6	
2.	118-122	10	
3.	113-117	20	
4.	108-112	20	
5.	103-107	15	
6.	98-102	30	← Mode Group
7.	93-97	5	
8.	88-92	8	
9.	83-87	6	
10.	78-82	4	
		N=120	

Now, $$Mo = L + \frac{f_a}{f_a + f_b} \times C.I.$$

$$= 97.5 + \left[\frac{15}{15+5}\right] \times 5$$

$$= 97.5 + 15/20 \times 5$$

$$= 97.5 + 15/4$$

Notes

= 97.5 + 3.75

= 101.25

Hence, Mo = 101.25 **Ans**.

Note :

1. From one table to find mean, median and mode, to calculate mean and median formula is same (above mentioned formula), but to calculate mode formula differs.

Example Mode = 3 Median – 2 Mean

(Mo = 3Md-2M)

2. If any question such a type of two class whose frequency is highest and equal also then (a) If those class are some distant then that question mode value will be and (b) If those class interval are to gather then mode value will be mean of mid point of those class intervals.

2.5 Assumptions Underlying M, Md and Mo

1. Mean (M)

1. When distribution is normal.
2. When each marks of distribution is given weightage.
3. When find out the centre of gravity precisely.
4. When almost reliability is desirable.
5. When to find out true mode.
6. When to calculate standard deviation, correlation, and standard error.

2. Median (Md)

1. When distribution of marks is not general.
2. When central tendency is to find out more accurately.
3. When more accuracy is not needed.
4. When distribution is skewed.
5. When distribution is incomplete e.g. some class have zero frequency.
6. When to find out the exact mid-point of group.
7. When in the end marks tendency is much.

3. Mode (Mo)

1. When to find out guess of central tendency quickly.
2. When distribution of marks is not normal.
3. When distribution is incomplete.
4. When to find out special measures of central tendency.
5. When to calculate central tendency on by observations.
6. When a rough estimate of central value will do.

Notes

7. when is to find out need such a special marks of central tendency which frequency is much more. **Garrett** has written correctly " When we describe the style of dress or shoes worn by the ' average woman for instance, the modal or most popular fashion is usually meant. In like manner, in speaking of the average wage in a certain industry, we often mean the modal wage under specified conditions."

2.6 Summary

- Having studied at data reader interested in that matter that he has to take help of only one measure to know the tendency of his group such a type of assumption is know as central tendency.
- Mr. Ross has written to define control tendency clearly. " It is the tendency of the scores to bunch or concentrate some where near the centre … It is that value which type files or best represents the whole distribution."
- There are three measures of central tendency-(1). Mean, (2) Median and (3). Mode.
- According to Prof. Bowley " Mean is that value of a series which divides distribution in two equal parts.

2.7 Keywords

1. **Frequency –** Happening again and again repetition.
2. **Mean –** In general mathematics which is called average in statistics it is called mean.

2.8 Review Questions

1. What do you mean by measures of central tendency? Describe its assumed values.
2. Calculate mean, median, and mode of following data.

Table		
C.I.	*f*	
200-219	6	
180-199	8	
160-179	13	
140-159	10	
120-139	6	
100-119	4	
80-99	8	**Ans**. M-148.12
60-79	2	Md = 155.5
40-59	1	M0=170.26

3. Compute median from following data.

Notes

Table		
C. I.	***f***	
0-10	2	
10-20	18	
20-30	30	
30-40	40	
40-50	35	
50-60	20	
60-70	6	
70-80	3	**Ans**. M = 37.15

4. Find Mean from following data.

   10, 16, 17, 22, 26, 24, 26, 11 — **Ans**. M = 19

5. Find Median from following data.

   5,7,9,12,10,8,7,15,21 — **Ans**. Md = 9

6. Find Median and Mode from following data.

   10, 6,7,10,9,8,7,5,10 — **Ans**. Md = 8
   Mo = 10

7. Find Median from following data.

   20,18,16,15,11,29,39,40,15,20 — **Ans**. Md = 19

8. Find Mean, Median, and Mode from following data.

   12,16,18,25,35,40,42,12,17,19,12 — **Ans**. M = 22.5
   Md = 18
   Mo = 12

9. Find Median, and Mode from following data.

   62,80,72,74,72,78,76,92,88,84,80,72 — **Ans**. Md = 77
   Mo = 72

10. Find Median from following data.

Table		
C.I.	***f***	
90–99	2	
80–89	1	
70–79	0	
60–69	0	
50–59	2	
40–49	0	
30–39	0	
20–29	3	
10–19	2	Ans. Md = 39.5

Notes

11. Find Median from following data.

Table		
C.I.	*f*	
45–49	5	
40–44	5	
35–39	10	
30–34	20	
25–29	0	
20–24	15	
15–19	10	
10–14	8	
5–9	4	
0–4	3	**Ans.** Md = 27.0

12. Find Median from following data.

Table		
C.I.	*f*	
20–21	2	
18–19	1	
16–17	0	
14–15	0	
12–13	2	
10–11	0	
8–9	0	
6–7	2	
4–5	1	
2–3	1	
0–1	1	**Ans.** Md = 9.5

13. Find Median from following data.

Table		
C.I.	*f*	
1–3	6	
3–5	53	
5–7	85	
7–9	56	
9–11	21	
11–13	16	
13–15	4	
15–17	4	**Ans**. Md = 65

Notes

14. Find Mode from following data.

Table		
C.I.	*f*	
45–49	2	
40–44	3	
35–39	7	
30–34	13	
25–29	8	
20–24	4	
15–19	3	**Ans**. Mo = 31.83

15. Find Mean, Median, and Mode from following data.

Table		
C.I.	*f*	
150–159	2	
140–149	2	
130–139	4	
120–129	1	
110–119	5	
100–109	5	
90–99	12	
80–89	10	
70–79	12	
60–69	10	
50–59	1	**Ans.** M = 92.62
		Md = 88.50
		Mo = 80.26

Answers: Self Assessment

1. Part 2. Mode 3. Mean 4. False

5. True 6. True 7. True

2.9 Further Readings

Books

1. Educational Techniques and Assessment—*Dr Rampal Singh, Bhatt Brothers.*
2. Education Techniques—*R. A. Sharma, Bhatt Brothers.*
3. Educational Research's Procedures—*L. kaul, Vikas Publishing.*
4. Educational Techniques Management and Assessment—*J. C. Agrawal, Bhatt Brothers.*

Unit 3: Hypothesis: Concepts, Characteristics and Sources

Notes

CONTENTS

Objectives
Introduction
3.1 Meaning and Structure of Hypothesis
3.2 Types of Hypothesis
3.3 Importance or Purpose of Hypothesis
3.4 Source or Elements of Hypothesis
3.5 Evaluation of Hypothesis
3.6 Summary
3.7 Keywords
3.8 Review Questions
3.9 Further Readings

Objectives

After studying this unit, students will be able to:

- Understand the meaning and structure of hypothesis
- Understand types of hypothesis, purpose and importance
- Understand source or element of hypothesis
- Understand evaluation of hypothesis.

Introduction

After the final decision of research problem their solution process begins means data collection is started, but before starting to collection of data it is necessary to definite for in which direction proceed. Formula is vested in that hypothese which is formulated by researcher on the basis of their knowledge by study, imagination and creativity.

In absence of hypothesis he will have to wander hither and thither by that his time and power will be consumed. Hence all researchers admit that as far as possible, the beginning of research should be done from hypothesis because in **Wann Dalen** words "Hypothesis works as a lighthouse in path of research."

3.1 Meaning and Structure of Hypothesis

When before any person any problem arises then he thinks about its solution. As a result, which measures comes in his mind that is probable solution of problems. It is another matter that later

Notes

on it proves true or false. For example a student is declared failed again and again in test. What is this reason, to know that for research he is submitted to a psychologist. For the solution of problem psychologists his causes of failure's do hypothesis. It may be he is of lack of intelligence, he is weak in study already, may he become sick in time of test, may be he is not interested in study etc., a number of hypothesis may be possible. Among these which are true and which are false. If solution is present into a question format (as a why student fails again and again) then these hypothesis can be possible answer of that question, but this is a starting point of research process.

Notes In **Townsands** words "Hypothesis is the projected answer of problem".

"Parikalpna" is Hindi translation of English word hypothesis, which means such a thesis which is hypo. In **Hodnet** words, Hypothesis is the eyes of researcher by which he peeps into unsystematic facts and discovers their solution. **Waan Dalen's** according; hypothesis researcher's solution of problem or about the answer of problematic questions an intelligent guess. He assumes hypothesis such a solution of problem which is only in the form of suggestion. **Karligar** assumes that hypothesis is an imaginary statement about the relation between two or more than two variables. As a "intelligence affects boys" academic achievement". This is a hypothesis. In this between achievement and intelligence, a special type of relation has been imagined.

When any problem occurred before any person and he tries to find out the solution, then first he imagines on the basis of his knowledge, experience, studies etc that what would be the solution of that. Alike as he tries to answer of any question, then first assumes that what would be their possible answer. He presents these probable solution, possible answer in ' format of generalizations and after that it is tested that how for they are time. There generalizations are called hypothesis. Similarly hypothesis in **Best** (1977) Words, "Such an inference which is formulated for the explanation of objectivity, incidences and circumstances are done temporarily, and which helps to further proceed of research work means later on if it is verified then solution of problem is and position is cleared". In our daily life occurred problem's solutions (e.g. formation of hypothesis and later their test) is done similarly.

Lighting bulb in the room all of sudden put out, immediate person asks to himself " what happened" a problem is occurred. The solution of a problem to get the answer of this he formulates one by one hypothesis and tested them. In the final status is cleared first do guess means formulates hypothesis "Probabilly electricity is cut out from sources". Having gone outside watches rest of the houses and finds that all house electric supply is continuing. Then this hypothesis "Fuse of my own house has not gone out of order". Having unplucked cut out he tests it and find that there is no disorder in it. This hypothesis to proves false. Then he do an another third hypothesis "Perhaps bulb had not fused. After the testing of bulb he finds that bulb is fused. This hypothesis proves true that bulb has fused and the whole situation become clear to sort out all the problems. Some process is followed and in that hypothesis (inferences, probable answers and solutions) plays on important role. Hypothesis too is a special and scientific process to sort out of problems. Hence in research two hypothesis has an important role.

Which is known, regarding that 'what is' 'why is' regarding this inferences is the hypothesis.

3.2 Types of Hypothesis

Applied scientist user in his research process a number of types of hypothesis. Various writers have described it by different types. It has been explained in details ahead.

Good and Hatt's classification

Notes

Good and **Hatt** (1952) have mentioned three types of hypothesis. He classified it on this basis that in hypothesis expressed, what is the level of obstruction? If the level of research is high then hypothesis' level also will be high. This depends upon that how abstracter extensive the conclusion and generalization of research are. By this approach hypothesis are of three types-

(i) lowest level

(ii) Higher level

(iii) Highest level

(i) Lowest level hypothesis: This type of hypothesis is found only in lowest level and they explain only some empirical uniformities. They are in the form of ordinary guess and statement. As an eleven years old boys' are very sensitive, average level of intelligence of students of class eighth between 90-100, India's backward classless socio-economic condition is not satisfactory etc. These are of low level's research hypothesis.

(ii) Higher level of hypothesis: From the aforesaid, higher level hypothesis concepts are based upon mutual relation. In these the matched level. These are vested in matched express any abstract concepts. The purpose of research of these types based hypothesis is extensive study of incidences.

Example As a due to rapid physical changing an eleven years old boys are most sensitive.

In this hypothesis the purpose of research is not only to know that these boys are sensitive or not, but also further to know that is their sensitivity cause within him rapid physical changing. Hence this hypothesis is no longer of low level become some complicated.

(iii) Highest level of hypothesis: In this level of hypothesis abstract thinking's highest level is vested. Thus this level of hypothesis is the most complicated. First and second level's hypothesis's related from, "what is" where as third (highest) level's hypothesis related with "why is". On the basis of highest level's theories and laws are evaluated. For example to guess that why and how Physical changing become the cause of sensitivity, will give occurrence of highest level of hypothesis physical changing and sensitivity related empirical uniformities mutually connecting will research that link which is of another chain's link. In this case level of thinking will be most abstract. Research complicatedness will be too much. The formulation of this type of hypothesis will be depend upon research's high level creative capacity and imagination power.

How the statement of hypothesis has been done from this point of view they are of two types. These are— (i) Question form and (ii) Declared form

(i) Question form hypothesis is in the format of question, as a – Does home environment effects psychological development of boys'? Some researcher likes hypothesis' question format.

(ii) Declared hypothesis is in the form of assertion. In that one facts guessed declaration. This facts is in the form of any events true position or about variables' mutual relation. This is very simple statement as home environment affects the psychological development of boys. Mostly researcher presents their statement in this format.

There is no differences between these two forms. Only difference is to write and to statement.

From an another approach hypothesis have been divided into two categories- (i) Experimental hypothesis and (ii) Statistical hypothesis

(i) Experimental hypothesis are called substantive hypothesis, scientific hypothesis, research hypothesis, operational hypothesis general hypothesis, empirical or alternative. It can be expressed in two forms- Question form and declared form consequently it can be of any level – low, higher, highest.

Notes

(ii) Statistical hypothesis is called null hypothesis, elimination hypotheses or zero hypothesis. The testing of experimental hypothesis could not be done on oneself basis. Therefore it has to change in statistical form do their testing indirectly. Statistical hypothesis is tested by statistical method. In this testing it has rejected experimental hypothesis is accepted. For example "Home environment affects self concept". This experimental hypothesis changing into statistical hypothesis it will be expressed as it " Home environment does not affects on boys' self concept".

From an another approach statistical and experimental hypothesis are of two types—(i) Directional and (ii) Non-directional

Directional hypothesis is called one-tailed hypothesis also. In there comparable groups differences direction is determined means one groups mean second groups mean's up and down taking which side of randomized curve, it is directed. In this only one side of normal probability is used as Boys' academic achievement is better than that of girls' achievement set it first – Two change it is null hypothesis "Boys' academic achievement is not much better than girls' academic achievement." For the testing of this hypothesis girl's academic achievement's mean will be blocked middle of normal probability curve and boys mean's on the right end. Difference will be only one side of curve.

Non-directional hypothesis is called two-tailed hypotheses also. In this group's mean's differences will be taken only one end of curve, such is not mentioned. This differences lies on either side of the curve means direction of differences is not mentioned. As a "There are a differences between boys' and girls' intelligence". (Statistical "there is no differences) In this hypothesis this is not mentioned that any intelligence mean is less or more from another groups' mean only to observe that difference is or not. If boys' intelligence's mean put on the centre of normal probability curve, then girls' mean from curve's centre on either side. If boys' are less than mean then left side and if more than right side. Whiever side will be, differences will be. This is to be testing. Hence this type of hypothesis is called two-tailed or non-directional hypothesis. Only to write the hypothesis' statement that it is directional or non-directional. In non-directional "Difference is" or "difference is not" user these words but in directional "is less", "is much", "is superior", "is good", "is bad" etc. such words are used.

Did You Know? Word 'Parikalpana' is Hindi translation of word 'Hypothesis'.

3.3 Importance or Purpose of Hypothesis

In research hypothesis' has important roles, be carve a number of purpose satisfies by the medium of it.

Researcher's view's is that where it is possible there research should be taken assistance of hypothesis, but it means not that without hypothesis research cannot be done. According to **Karligar's** without hypothesis also research is possible. According to him particularly in exploratory research, hypothesis is not inevitable, but in modern science field it is not possible. **Waan Dalen's** view is that in those researches in which relation of cause-effect is aim, hypothesis is particularly required. If the aim is to describe any event's present status then, there is much need of it, although there too researcher's may have for the discovery of suitable research-scope hypothesis may needed. In fact in the initial stage of research whenever problem is not finally at exploratory stage. Hallway's statement is too in such a research whose purpose is to data collection only, being hypothesis is not necessary. Historical research, descriptive research, legal research, documentary research, bibliographical research etc possibility of use of hypothesis is nominal, rather than in some cases these research too such can be, in which use of hypothesis proves useful. These field's studies generally based upon inductive logic

Notes

system. Hence in that research does not begins from hypothesis rather finished by hypothesis. **Beacon** had given challenge to compulsion of hypothesis in research whose statement was that hypothesis is produced bias and prejudice in the researcher, by that objectivity of research becomes less, but it seems no justification in it.

In fact hypothesis has great importance in research. There are number of advantage of its.

1. **Mouli**, "According to hypothesis clarify the direction of research. As a result researcher has not wander hither and thither. He can avoid collection of unnecessary facts and information. He does not kill his time in study of fruitless literature because hypothesis determines avoid that which types of facts to be collected and what is to be studied in which book." Thus, according to **Waan Dalen** hypothesis is worked like a pool between research problem and sort out of solution. It clarifies the limitations of problems what is significant and what is not, it can too determines by hypothesis. Consequently which facts and proofs are collected how show they classified, it is clarified also hypothesis wine their composition and analysis is possible to do. Which variables are vested in hypothesis only to know about that. Thus research's saves a lot time money and power.
2. By research hypothesis it can be clarified also that how research methods should be used for research design. Which types of relation assumes in hypothesis for the testing of that accordingly design is selected. A number of design of collection of data, and their analysis as a pre and post testing, controlled groups design are available. Which of them will be more suitable it depends upon that what is hypothesis. Likewise a number of methods data's depends upon selection of hypothesis too. Hypothesis indicates that data will be collected by which persons, which place and which type of Psychological test. From this point of view hypothesis has a great importance in the field of research.
3. Hypothesis presents that framework also, in which as a result of sort out of problems conclusion is established. For example. It is to be studied boys' problematic behavior then selected to hypothesized cause factor, what tested that and at that framework. Problematic behavior will be explained. There is no question arise to go beyond out of that. Hence to explain the problem in objective form, hypothesis is very helpful in the presentation of research's result significantly gets much help from that.

Example When we watch a figure the whole image prevails in the brain in the same way conclusion based on hypothesis exported collectively problems form.

4. There are a number of aspects of problem whose study is essential and important but all of them cannot be studied together. Which aspect is to be studied or going on, it is cleared from hypothesis. Thus hypothesis is a path which indicates that where to arrive. **Karligar** compared research to scientific game and said that like player, research "First set chance and after that throw dice. First throw dice then take change, not such happen. After the collection of data, chance cannot be changed.
5. Hypothesis is not the final point of creation of knowledge. Testing of a new hypothesis gives birth to another, second testing is third. Like the waves of a ocean then push to each other ahead as a result of that knowledge range is extensive. Each research's hypothesis indicates towards next hypothesis.

Self Assessment

Fill in the blanks:

1. According to Waan Dalen, hypothesis in the path of …… is worked.

Notes

2. Research too is special of sort out of problem…… process.
3. Low hypothesis …… founds in research level.
4. Higher hypothesis …… based upon mutual relation.

3.4 Sources or Elements of Hypothesis

Hypothesis's formulation is as tough as selection of problem and source of elements of hypothesis is same as selection of problem. There is no fixed method or formula for formulation of hypothesis like selection of problem, but some essential condition, circumstances and base is which is helpful in this work. These are called soure of formulation of hypothesis or elements or helping factor.

1. Researcher's study knowledge and Experience: These three are much helpful in the formation of hypothesis. Not only the study of own subjects, but also other subject's study is much beneficial. From interdisciplinary studies develops new ideas, increased knowledge as a result of that problem based variable's.

New relation seems, which sometimes give birth to a suitable hypothesis on the basis of intensive and extensive study only researcher can find out that how much and what types of study for research determined subject have been done so far, how they have been tested under hypothesis, which types of error are, which are new direction, towards which researcher's attention has not been gone so far. In hypothesis based variables which are theoritical conceptions are, after the study of that their operational definition's knowledge is acquired. On the basis of that variables' related, formulation of hypothesis get help. In the framework of research subject researcher's will as much knowledge of that as much their concepts of that. He will have to able to formulation of hypothesis as logical and important.

Sometimes apart from own subjects, study of other subjects is much beneficial regarding this. In education, Psychology and in Sociology's field a great many concepts are which have been taken from other fields and on that basis in this field completely new type formulation of hypothesis doing they have been tested. For example **Kurt Lewin** for the explanation of human behavior formulation of that hypothesis and testing them that was based upon Physic's concept forced. Like user controlling of behavior **Feed-Back** rule had been taken from engineering thermo state. Thus this type of completely new insight is developed through extensive and intensive study. Thus researcher's background of knowledge is an important source of suitable and superior hypothesis.

Researcher's own experience role is not less important regarding this. Events happening in the community, conditions of life, behavior of a person, works of institutions, purpose, process of work and going on changing is political, social and economical's as extensive experience researcher's will have as able he will have precise, significant and suitable formulation of hypothesis. These experience of their in the clarification of problem based variables many and different types of relation helps reseacher's. Among these new and fundamental hypothesis prevails.

If a researcher's study knowledge and experiences range is very intensive then there is not possible for him to discovery of new relation of problem based variables and formulation of fundamental hypothesis.

2. Mental Caliber of Researcher: These are too a source of formulation of higher level hypothesis. Study, knowledge and experience in itselves do not able to formulation of hypothesis, but at mental level their manipulation or at mental level their explanation, composition rationalisiation, etc., are important. The birth of a for said hypothesis by such mental exercise. Researcher's creative ability a great role in this creative thinking mutual relations of problem based variables presents into a new dimension. Then formulation of higher level hypothesis becomes simple and possible. Researcher's imagination is a powerful chain of process of formulation of hypothesis. Science had done miracle taking the help of imagination of human being. Scientist **Alberrt Einestine** assumes that maris

imagination is much more powerful than his knowledge. A high level imagination researcher's far go beyond to conventional problem-solution, pre answer of questions. Hence research's far does not admit it. He thinks over that seriously "Why' "What", "How" enters tough valleys and is discovered such a new facts, perspectives relevancies, factors which presents, before then problem into a new and wonderful approach and give birth to very new hypothesis. **Thorndyce, Pavloveas'** and **Skinner**, behind the evolution of induction-theory such a types of hypothesis will haves.

3. Reasoning by analogy: alike conditions and event's logical analysis is also helpful in the formulation of hypothesis. Some times are subject or one field events, ideas and theories much alike to other subject, events of region, By comparing of that or doing analysis after matched, sometimes researcher's finds a new ideas which is helpful in the formulation of fundamental hypothesis. It is called reasoning by analogy. For example a research studying effectiveness of academic institutions, reads in the book of sociology that lack of controlling in society give birth to social disorder and corruption. He is compared to that condition of academic institutions condition, analyze that then this idea born that if lack of controlling within academic institution then it can be spoiled the environment of academic – institution and its adverse effective non of institution.

4. Some other suggestions: In the formulation of above hypothesis time has its own importance. On the problem of research upto a long period of time thinking should be done seriously. Thought over that should be a different point of view.

Caution Researcher's should not do hurry in the formulation of hypothesis.

At times for some time problem should be forgotten. To do so in that thinking alesence period research's subconscious mind will have thinking over different aspects of problems, worked a little aloof from conscious mind thinking link. Hence unexpected but very useful hypothesis prevails.

Researcher guide, experts, administrator, officer sitting with colleagues and discussion with him seriously over hypothesis is useful in this context. To give speech on any aspect of under research process problem (if get a chance), write essay on that, presents problem in seminar for discussion etc. are beneficial.

Self Assessment

Multiple Choice Questions:

5. Declared hypothesis is _____ in the form of statement.
   (a) Indefinite (b) Definite (c) Changing (d) Neutral
6. Question form hypothesis _____ is the form of _____ .
   (a) Question (b) Answer (c) Alternative (d) Incidence
7. Directional hypothesis is called _____ hypothesis also.
   (a) Two tailed (b) One Tailed (c) Parallel (d) directed
8. In research hypothesis have _____ role.
   (a) Negligible (b) Worthless (c) Important (d) None of these

3.5 Evaluation of Hypothesis

The evaluation of hypothesis means that to which hypothesis is formulated how far they are suitable mean how far those characteristics exists which should be in a good hypothesis. Experts has described

Notes

good hypothesis, accepted imagination some merits. There are formulated criterion of required hypothesis. The characteristics of this criterion are following-

1. Hypothesis should be Plausible: Its meaning is that which solution possibility express in those, that seems logical and true. In other words which relation is imagined among variables, from any point of view they are logical. If that relation from any point of view cannot be reliable about their to be true then that hypothesis cannot be understood suitable. This hypothesis is that "available number of bench and desk in school affects students' academic achievement" totally beyond reliability, because between the number of bench and desk and students academic achievement from any point of view seems no relation. Which relation has been imagined that is neither relevant nor logically possible. There is no relation at all between number of benches and desks and students' academic achievement- neither any experience basis nor theory basis or previous experiment basis. Hence this hypothesis is not plausible. It cannot be accepted thus a suitable hypothesis is that presents a logical, reasonable, reliable, facts based possible explanation. Consequently it should not be impossible from implementation point of view. As a "This hypothesis that six hours study in a day in place of it eight hour study doing students' academic level can be gifted. From implication point of view it is impossible. Due to a number of factor increased the study hour could not possible. Beyond this inherited relation in hypothesis or solution not go against ongoing rules and theories "To rise the academic levels of student fear and hard discipline is necessary". In this hypothesis probable solution "fear and hard discipline "is against the theory of Psychology. Therefore this hypothesis is not acceptable.

2. Hypothesis should be testable: Its meaning is that which possible solution or relation is imagined means which explanation of solution presents, that is such a type that it can be tested As a – " intelligent but underachievers is suffered from low feelings both are such a characteristics that their measures means hypothesis testing is possible. Now we can say that "intelligences students' failure cause their pre-birth rites", then this hypothesis cannot be accepted because" "pre-birth rites's measures is not possible. Hence hypothesis cannot be tested. If inherited variables of hypothesis result's practically observations, testing and measures cannot be done, then testing of hypothesis cannot be done also. At that it is considered to be, worthless. Similarly if hypothesis used variables analysis any technique or method is not available then their testing, measuring will not possible. If problematic variables' nature is such that in their measuring will take many years then that hypothesis is not understood as good hypothesis, although that is not a hard and fast rule.

If in hypothesis such variables are included that cannot be defined then their testing cannot be done. As a hypothesis "modern age tendencies has been adverse effect on life style of true religious person" in "modern age" "tendencies" "true religious" "life style" all these words or variables are such that which cannot be defined objectively . Therefore, it cannot be measured also. In this case, hypothesis can not be tested also.

This from any reason or approaches hypothesis used variables and measures cannot be done then it does not consider suitable, because their testing cannot be done.

3. Hypothesis should be extensive: Hypothesis as much as extensive, as much as considers to be beneficial and valuable. Its meaning is that, the hypothesis which "problem related all or much more facts explained, that is considered to be extensive and useful. That hypothesis from this point of view is very limited, that has no utility. Hence to make hypothesis useful, try to make its field more extensive. For example, in education Psychology for the explanation of induction many hypothesis are available but none of one do not explain some, second other facts. Not a single hypothesis field is so much extensive that it had explained all those facts and events whose others do different hypothesis. Suppose that a hypothesis is " achievement motivation the most important". If this hypothesis explains only eighth class' students' academic progress or does not explain any other circumstances, facts or incidence then it will be considered to very intensive and their utility will not be too much . Now if it explains all levels students' academic progress then their extensiveness will increase. Similarly, if achievement motivation's importance besides students, teacher's progress their learning, teaching progress etc., founds, then their extensiveness' field and will be much extensive. Similarly it explains

Notes

industrialists, traders, lawyers employers states and nation's progress, then it will be considered to be extensive and useful. Hence, if any one hypothesis's basis if any generalization is available that explains or presents different fields' in different circumstances happened events and problems, then that hypothesis will be considered to be more extensive, useful and beneficial. Theory of gravitation, atomic theory, theory of evolution such a hypothesis's example, whose explanatory power and problem solution are much extensible.

Task Give your thoughts on hypothesis testing.

4. Hypothesis should be theory based: Hypothesis in related field should be based upon using theories. In every hypothesis between variables a relation is imagined, but this imagination pro-theory means a solid base should be for such thinking. These bases are pre-researched theories, rule, generalization etc., in the form, available. Those hypothesis not opposite to it rather accordingly theories that is considered to be suitable; As the hypothesis is that "Students' positive self-concept is the factor of their tested higher success' can be considered as theory-based. Success in test is a special type of behavior which a number of other behaviors such as –present in the class regularly, read attentively, practice regularly, try to eliminate difficulties etc., depends on that. Behavior is affected by persons' personality qualities. **Karl Rodger** and by many others Psychologists postulated "Self concept theory" assumption is that positive self-concept give birth to constructive, organized and integrated. For higher level testing Success these types of behavior is desirable. Hence positive self concept should be a factor of tests success. This hypothesis founds to be attached with use theory. Pre determined imagination, ideas, guess relation's imagination theory originated mean theory based is. Accepted theories appose or against that hypothesis is not considered to be suitable. Such hypothesis' that present theories short comings, their errors, to remove their inconsistency formulated can be accepted but they do no match with familiar theories or not supported by any theory then it cannot accepted.

5. Hypothesis should be purposeful: Its meaning is that hypothesis should be related with hypothesis, means be it should present the solution of problem. Any problem from many point of view a number of solution can be. As a result from different approach a number of hypothesis can be, but all those hypothesis should be purposeful necessarily. Rather than differ from each other which is the purpose of problem solution from that point of view hypothesis should be related to that purpose. For example, students' failure in test in cause finding from many point of view hypothesis can be formulated. From Psychological point of view "Lack of motivation and interest in student" from sociological approach it can be hypothesis "Students' economic social environment not be favorable", Similarly educational approach hypothesis can be "Teaching method not be suitable" But all these hypothesis rather to differ each related to problem's solution's acceptable forms, hence all are acceptable. Difference means that oppose to each but rather they should support to one another. Each hypothesis is formulated for a definite purpose. So, to fullfil in that purpose should be supportive, only then it is considered to be suitable and acceptable.

6. Hypothesis should be simple: Its meaning is that , to present the solution of problem or their explanation simply means presented solution is not based upon such a complicated solution is not based upon such a complicated theory, assumption, conception that it is hard to understood. If possible the hypothesis present in simplest form is considered to be more suitable. Thus hypothesis presented in simple precise and in consistent language is required. Use of unnecessary words, use of tough word, impulsive or value based expressions for hypothesis not appropriate. For example this hypothesis that "Students who are at the top of success, who are set forth example stars are rich of intellectual talents." From aforesaid point of view it is not to be said acceptable. Directly it is to say more suitable that "to achieve good success in examination students' are more talented."

Notes

3.6 Summary

- Parikalpana is the Hindi translation of English word 'hypothesis' that means such a thesis which is hypo.
- In **Hodnet** words hypothesis are the eyes of researcher's by which he peeps in problematic in consistency (unsystematic facts) and in them sort out of problems.
- Research too is a specific and scientific process of sort out of problem. So, in research also hypothesis has an important role.
- Hypothesis has an important role in research because various purposes fulfils by that medium. Karilar's according without hypothesis research is impossible. According to him particularly in exploratory research hypothesis not inevitable, but in the field of modern science such is not possible.
- **Waan dalen's** view is that those research in which cause-effect relation discovery is aim, hypothesis is required particularly.

3.7 Keywords

1. **Creativity -** Creative or making new things
2. **Null –** Worthless, zero

3.8 Review Questions

1. Give the definition of hypothesis and explain its meaning.
2. Doing classification of hypothesis, mention its type.
3. Clarify the importance and purpose of hypothesis.
4. Mention source or elements of hypothesis.
5. Evaluate hypothesis.

Answers: Self Assessment

1. Light house	2. Scientific	3. General	4. Concepts
5. (b) Definite	6. (a) Question	7. One tailed	8 (c) Important

3.9 Further Readings

Books

1. Educational Techniques—*R. A. Sharma, Bhatt Brothers.*
2. Educational Research's procedure—*L. Kaul, Vikas publishing.*
3. Educational Technique – *S.S. Mathur, Bhatt Brother.*
4. Educational Technique and Assesment—*Dr. Rampal Singh, Bhatt Brother.*

Unit 4: Formulation and Testing of Hypothesis

Notes

CONTENTS

Objectives

Introduction

4.1 Formulation of Hypothesis

4.2 Fundamental Basis of Hypothesis

4.3 Formal Conditions for Testing Hypothesis

4.4 Testing of Hypothesis

4.5 Summary

4.6 Keywords

4.7 Review Questions

4.8 Further Readings

Objectives

After studying this unit, students will be able to:

- Know the formulation and fundamental basis of hypothesis
- Know formal conditions for testing hypothesis
- Know testing of hypothesis.

Introduction

There are more than one formulation of hypothesis to solve any problem by statement. When some problem arises in front of a person and he tries to find out the solution, then he imagines what will be the probable solution on the basis of his knowledge and experience. He presents hypothesis of probable solutions and probable answers and then he examines to what extents this hypothesis is correct.

4.1 Formulation of Hypothesis

There are more than one formulation of hypothesis to solve any problem by statement. Here are three formulation of hypothesis of one problem that will decide the solution of problems-

(a) Prize increases the achievement of study.

(b) Prize decreases the achievement of study.

(c) Prize does not affect the achievement of study.

First two hypotheses are directional hypothesis and third hypothesis in non-directional hypothesis. Principle of excession can be used for class excession. Second hypothesis cannot be said as good

Notes

hypothesis because it is not right to use it and there is no principle foundation for this hypothesis.

There is no need of any base of principle for **third type** of hypothesis, so this can also be used, but only one hypothesis can be used out of first and third hypothesis. Both hypotheses cannot be used because these are contradictory to each other.

Entry and exit both are used for thought process. Many psychologists and principle of education use both types of processes in which formal behavior is used in the format work.

Notes For research, entry and exit both are used for thought process in making hypothesis.

Thought process of entry and exit can be used for the research of formulation of hypothesis conclusions derived from the special condition of hypothesis are generalized. Research process used to get conclusions from research the basis of hypothesis.

Hypothesis is a probable assumption on the basis of knowledge and principle. It directs the work of researchers. The most important and difficult work of a researcher is to formulate hypothesis. There are three characteristics of a good hypothesis:-

(a) Enough data are collected with patience.

(b) Conclusions can be derived and generalized.

(c) To get result discussion are made with the quantitative data.

Hypothesis plays an important role in the process central point of research. Hypothesis is the entire research process which helps in the selection of new method, measurement of process formation of research, collection of data and discussion with data.

According to a study, analysis and confirmation of main hypothesis is done but with this analysis secondary hypothesis is also confirmed. When many hypothesis are used then each hypothesis is formulated in a separate form and it is presented on which basic principal of hypothesis is made data are analysed for the confirmation of each hypothesis so that it can be accepted or rejected. According to **H.H. Mackson**, purpose of research is not only formulation of hypothesis and its confirmation but also to discover new facts. Formulation of hypothesis and its confirmation is not made in all research works research is also done without any hypothesis. Hypothesis is meanly used in the research.

Self Assessment

Multiple Choice Questions:

1. There are _____ formulation of hypothesis for solution of any problem by speech.

   (a) One (b) Some (c) More than one (d) None of these

2. Both entry and exit are included for _____ process.

   (a) Study (b) Thought (c) Attraction (d) support

3. Researchers are related to realities at two levels – operational and _____ .

   (a) Unprofessional (b) conceptual (c Social (d) Psychological

4. Hypothesis is a probable _____ which is based upon knowledge and principle.

   (a) Assumptions (b) Reflection (c) Criticism (d) Study

4.2 Fundamental Basis of Hypothesis

Notes

Researches are related to facts at two levels.

(a) Operational and (b) conceptual

(a) Appropriate words should be used to define the analysis of events and facts at **operational level** so they actual facts can be used in the research.

(b) At **conceptual level** researchers should define events and facts in cause, affect form to get the knowledge of that events. At this level researchers define events in such a form in which he tries to move from specific event to normal event. It gives the knowledge of relationship between variables and its capacity. Normally there is a movement from operational to conceptual in the formulation of hypothesis. Researcher generalizes from event and provides its benefit universally. It is necessary to define events and facts at both levels for research because at operational level definition confirms hypothesis. At conception level definition shifts research from conceptive to operational.

Did You Know? Aim of research is not only formulation of hypothesis and its confirmation but also to discover new facts.

Difficulties in the formulation of useful hypothesis: Generally these are the problems in the formulation of hypothesis as mentioned below:-

1. Lack of clear theoretical context.
2. Lack of capability and ability to use theoretical context.
3. Lack of complete knowledge of available experimental equipment and its process.
4. Lack of ability to formulate hypothesis in appropriate words.

Task Express your news about the difficulties in the formulation of hypothesis.

4.3 Formal Conditions for Testing Hypothesis

There are two types of formal hypothesis.

1. Research hypothesis and
2. Quantitative hypothesis and zero hypotheses.

Researchers formulate their research in two forms-Directional hypothesis and non-directional hypothesis. Available data are analyzed and lasted by statistical processes. Zero hypotheses is used in this technique. Some formal conditions for testing hypothesis are necessary such as:-

1. Exit process must be used in the statement of hypothesis so that it can be generalized.

Caution Hypothesis must be written in formal and clear words.

2. The structure of hypothesis must be in such a form which can be accepted or rejected

Notes

3. Report of hypothesis must be special and in tested form so that formal side of hypothesis can be examined.

4.4 Testing of Hypothesis

Development of hypothesis is not a solution of problem. Hypothesis is only an expected solution of it. It is a solution of problem which is understood when it is accepted. Acceptance and rejection of hypothesis is made on the basis of external appearance. So, researchers analyse the hypothesis after its development. It has been decided on the basis of analysis. Whether it represents solution of problem or not. Formulation of hypothesis depends on three things- (i) Fixation of results of hypothesis, (ii) To decide whether results are really available or not so that its measurement, analysis, selection and analysis of verification of tests and the method of analysis, (iii) To collect data (related research and its analysis and testing and to conclude whether the hypothesis under consideration is acceptable or not acceptable.

1. Fixation of Formulation of hypothesis

Under this it is decided if hypothesis is true then in which form and where its result will be available. Some hypothesis are in such a form whose results of external events, situation and facts are available directly of appropriate situation, place and form. Example, of hypothesis "Road accidents happen more at night ", then the result of hypothesis is true means number of road accidents at night will be more than day and its number can be worked out. Result of some hypothesis are not available apparently. So, their indirect forms and presentation have to be discovered.

Example If the hypothesis is "It is difficult to accept mentally-retarded children socially."

If hypothesis is true then its results cannot be seen apparently. Results of expression can be found in apparent form. There results can be fight among mentally retarded children, low in the member of their friends, their more sentimental, to live separately, not liked by their friends. It is not easy to find these types of invisible results. It is very necessary to conclude that selection of results should be compels only related to the hypothesis. It these are not related to hypothesis then their testing is useless. Along this their presentation must be in clear and measurable forms and in the form of situation and events.

2. Selection of methods of testing and development of testing

Development of hypothesis is the probable solution of problem. That probable solution can be true or false. So, the next step of research is to discover methods and medium of testing of hypothesis. To present evidence on the basis of which it is decided whether the hypothesis is true or not. There data are available by physiological medium and testing. There are deduced consequences of movable research data which means these are the information on the basis of which it is tested by hypothesis whether it is true or not. So development and selection of methods must be valid and error-less otherwise in the absence of valid information unavailability of correct evidence it cannot be concluded whether hypothesis is correct or not.

Self Assessment

Fill in the blanks:

5. After development of hypothesis……. tests the hypothesis.
6. Development of …… is the probable form of solution of problem.
7. There are mainly three ……. of testing of hypothesis.

Notes

3. Hypothesis–Testing: Last chain

On the basis of selected methods and information from the medium of testing, data is collected and tested and its main aim is to select or reject the hypothesis. The first thing to be kept in mind that hypothesis can be selected or rejected. It can not be proved and verified. The words "To prove" and "To Verify" show that these words are "Final truth", "Undoubtely true". It is true which actual situations are different. The truth is that testing of hypothesis is verified in the probable from. So the use of words" Acceptance "or "Rejection" are considered correct and are presented in the probable form of conclusion of acceptance. These things are discussed in details.

4.5 Summary

- Statement of any problem and to solve it more than one hypotheses are formulated.
- Hypothesis is important in research. Hypothesis is centre of research which helps in the selection of new ideology, methods of measurement, form of research, collection of data and analysis of quantitative methods.
- Statement of H.H. Mackson is that aim of research is not only formulation of hypothesis and its confirmation but also to discover new facts.
- Researchers test the hypothesis after its formulation. It is concluded on the basis of testing whether this hypotheses presents solution of problem or not.

4.6 Keywords

1. **Methods**—Defined methods of special activities to work in technical fields.
2. **Exit**—To conclude from less general to more general form for estimation.

4.7 Review Questions

1. What do you mean by formulation of hypothesis?
2. Explain testing of hypothesis.
3. Explain formal conditions for testing of hypothesis.

Answers: Self Assessment

1. (a) More than one	2. (b) Thought	3. (b) conceptual
4. (a) assumptions	5. researchers	6. hypothesis
7. basis		

4.8 Further Readings

Books

1. Educational Technique—*S. S. Mathur, Bhatt Brothers.*
2. Methods of Educational Research—*Shareen and Shashikala, Vinod Pustak Mandir.*
3. Management and Valuation of Educational Techinique—*J. C. Agarwal Bhatt Brothers.*
4. Educational Technique and Valuation—*Dr. Rampal Singh, Bhatt Brothers.*

Notes

Unit 5: Qualitative and Quantitative Data

CONTENTS

Objectives
Introduction
5.1 Qualitative Analysis
5.2 Quantitative Analysis
5.3 Conclusion
5.4 Concept of Generalization
5.5 Summary
5.6 Keywords
5.7 Review Questions
5.8 Further Readings

Objectives

After studying this unit, students will be able to:

- Understand qualitative analysis
- Understand quantitative analysis
- Understand the concept of generalization and conclusion.

Introduction

Data or documentary evidence whose types and collection methods have been mentioned previously are only a means. They have no importance in themselves. Ends are for fulfillment of research purpose i.e. drawing conclusion related to the problem or to get the answer of true problematic questions. Documentary evidence i.e. data is only means to achieve these objectives on the basis of their analysis for reaching at destination.

For the analysis or treatment or processing of these data a number of methods are used. Some writers distinguish between processing and analysis, for example **Gultong** (1967) under processing only collects data at one place, keeps them in different categories and keeps other types of required arrangement. By analysis he means in the context of hypotheses end true problematic questions on the basis of evidences establishment of conclusion but many other writers do not differentiate between them. They understand all those activates components of data analysis. If it consider as components of analysis or keep it separately, it does not means any difference in fact data's collection or processing, analysis and interpretation three activates are separately in reality.

There is not a single method or techniques of analysis and interpretation of data is much different to one another also. Therefore for each type of analysis of data any one method may be, it is not possible. Which types of data is, they are analyze by that method, which is appropriate for that nonetheless generally analysis of data can be divided in two groups –(1) Qualitative analysis (II) Quantitative analysis

5.1 Qualitative Analysis

Notes

Historical studies, some descriptive studies and ethnographical studies' data are quantitative and descriptive. It cannot be expressed in numbers and digits. Therefore its analysis can not be done by numerals method also. Hence, for its analysis, qualitative method is followed.

Under this on equality basis data's can be classified e.g. on the basis of hypotheses (which data is associated with which hypotheses) grouping can be done and on that basis conclusion is established. If hypotheses had not formutated before then on study time which characteristics is to be found on that basis classification can be done. For example if "informal education's status in division (Jan Pad)' has been done and related to that many types data have collected then all that data on equality basis included in some groups, on that basis that education's status' on many aspects throw light can be done. Finally, combines all those aspects collectively it can be stated that what is the status of informal education. Like that in observational studies, which material is available after the observation of behavior, phenomenas and circumstances, that is on equality basis having classification on that group in which seems characteristics, on that basis persons, phenomones etc. can be described. Mostly survey's purpose is only description of institutions, communities, society, caste, professions, processes etc. of present status. These studies' data is mostly descriptive e. g. literal. On that classification basis any conclusion is established.

In this way whatever material is available in ethnographical studies that is synthesised in the context of important questions. On that basis an important sketch is prepared it is presented only in descriptive style. The base of this whole analysis is logical and theoretical. Statistical methods are not used in the analysis of this data.

Contrary to this in the analysis of this data attention is paid that there is a logical and meaningful relation among, purpose of studies, Processes, methods, results etc., by increment of knowledge. The quality of study depends upon this. On the basis of analysis of whole data and facts, having kept it mutually associated and organized, a complete perspective is formulated. Under perspective "why", "what", "how"'s sceneries' are given stress in descriptive and reflective style expression. Conclusions are based upon researcher's own logical decision.

Like this is historical studies whatever evidences are collected, their internal and external criticism is done. After that in descriptive style only conclusion is established. The base of analysis and criticism mainly researcher's logic, analogy, classification of facts vested relation etc are. In historical studies for mutation of hypotheses is done tentatively. Having collected facts related to those, which are in the form of as per word facts on the basis of their logical analysis hypotheses, is tested. This whole analysis is quantitative in which are, researcher's thinking, his insight, his judgment, his imagination, logic and intelligence. Play an important role. The main example of quantitative analysis are **Piyaje's** cognitive development theory, **Friod's** unconscious mind theory, **Arrald Geesal**'s child related **Margret Meed studi**. In the field of Comparative studies, purposes of different counties education, their basic philosophy, texts, developments' comparative study, are the example of qualitative analysis.

In qualitative analysis that mental process help is taken, that are *(i)* inductive - deductive reasoning, *(ii)* reasoning by analogy, *(iii)* thinking and imagination, *(iv)* analysis of similarities, *(v)* analysis of difference, *(vi)* establishment of relation *(viiI)* relation explanation by theories approach *(viii)* finally description of results on the basis of all synthesis.

Self Assessment

Fill in the blanks:

1. Historical, descriptive and ethnographical studies' data ……. and is descriptive.
2. In historical studies' collected materials' internal and …….… is criticized.
3. Numerical analysis is possible where data in …….. is available.

Notes

5.2 Quantitative Analysis

Notes Numerical analysis is possible only that condition where data is available in the form of numbers and digits.

In applied science these data are the marks of Psychological examination. Observational data is also available in numerical forms for several times. The analysis of these types of data is done by the basis of statistical methods.

A number of statistical methods of this type has developed. Basically these methods have been divided into two parts- (i) Descriptive statistical methods and (ii) Inferential statistics, descriptive statistical methods has been used there, where collected data's having presented in precise form on that basis any population, any group or any specific class characteristics is to be described. For example, a teacher wants to know what is level of mental ability of his class students. He gives a intelligence test to the pupils on the basis of which their marks are available. Having tabulized them, their mean is calculated on that basis. He can describe the intelligence level of students of class. Besides mean a number of statistical value which is calculated. In different conditions, different statistical value are calculated. Which statistical methods will be used for analysis of data, it is depends upon research's purpose and hypotheses.

Descriptive Methods of Statistical Analysis

Under this heading all those methods come which only describe the characteristics of any group or population or presents data any summarily.

Regarding any speciality, what is the status of any group or population, get the answer of these questions by these methods analysis basis.

1. Mean
2. Median
3. Mode
4. Standard Deviation
5. Correlation
6. Graphical representation of data

Meaning of these methods, use, statistical calculation, etc., has been described is detail in further chapters of this book.

Inferential Methods of Statistical Analysis

Caution These methods are used, where on the basis of a sample larger population's characteristics are described.,

Under this a number of statistical methods are included, but these methods are based upon only one concept. This concept is of standard error. Aforesaid all statistical values (mean, median, standard

Notes

deviation, correlation) having known its measurements error, on that basis it is stated that any conclusion is how far true about whole population. These concepts and all those methods have been mentioned in the second part of book. Main methods which have been mentioned are-

1. S.E.m. e.g. Standard error of mean.
2. S.E.r. e.g. Standard error of correlation.
3. S. E. σ. e.g. Standard error of standard deviation.

Cause- effect Relation's Statistical Analysis' Methods

In applied sciences, a number of circumstances of researches are such that in which phenomenas, behaviours and course of results is determined.

Example It determines that "Is intelligence course of students' educational achievement?"

"Having combined both intelligence and initiative how it effects educational achievements" etc. These types of studies comes under experimental research and its process a number of times is much complicated. These studies have special strategy that is called experimental design. From more complicated to most complicated designs have been developed by experts on the basis of these designs whatever data is available, their analysis have special statistical methods also. These are called statistical designs. It has been mentioned in details in further chapter. These main methods are-

1. T-test
2. F-test (ANOVA, Analysis of variance)
3. ANCOVA, Analysis of covariance
4. Factorial Design
5. Regression Analysis

Relation's Analysis of Statistical Methods

Any studies's purpose is only to know the relationship between two variables. Who is whose cause, to know it is not purpose. Such studies' data is analyzed by correlation methods, these methods are of many types, for example –

1. Rank order (rho) product moment correlation: When only correlation between two variables is to find out then these methods are used.
2. Biserial *r*, tetrachoric *r*, Phicoefficent of correlation and contingency coefficient of correlation: These are the specific methods to find out relation between two variables also.
3. Partial *r*: where any one variables effect has been cancelled, to find out the relation between two variables then this method is used. For example, if intelligence and inspiration collectively affects educational achievement then to know it having controlled initiative effects, how much relation of intelligence with academic achievement.
4. Multiple correlation: Where it is to know that more than one variables together how much effects a dependent variables, these methods are used.

What is the meaning of these methods how it can be used and in these how statistical values are calculated, all those matter have been mentioned elsewhere.

Notes

Statistical Analysis' Non-parametric Test Methods

Further T-Test, F-test, ANOVA, ANCOVA etc., methods have been mentioned. To use these methods some preconditions are there. These conditions have been mentioned in further chapters. If these conditions are not fulfilled then these methods can not be used which are called parametric tests, but in that conditions it cannot be used, in those its non – parametric tests alternatives are used a number of non-parametric tests are among those some mostly in used methods one following-

1. Chi-Square Test
2. Median Test
3. Runs Test
4. Kolmogorov- Smirnov Test (K. M. Test)
5. Mann Whitney 'U' Test
6. The Walsh Test
7. Wilcoxon – Wallis one-way variation analysis (K. W. one way ANOVA)
8. Wilcoxon matched pairs Singed Rank Test.
9. Friendman's two-way ANOVA by Ranks

Self Assessment

State whether the following statements are True or False:

4. The main example of quantitative analysis is Piyaze's mental development theory.
5. In quantitative analysis help of mental processes is taken.
6. Chi-square tests is non–parametric methods of statistical analysis.
7. The special strategies of studies are called statistical design.

5.3 Conclusion

The analysis of data is an important chain related to fulfillment of purpose of research, but on the basis of the result achieved do not become relevant as long as till their explanation is not done in the context of problems. Therefore, it is quite pertinent to add with purpose of problems getting results by analysis. For example, suppose that the "Analysis of academic achievement of mathematics problem was of students of class eighth. Having chosen a sample to sort out the problems, under that comes students' given them a test of mathematics to find out the mean of marks. Suppose it comes 60. This is the result of our analysis, but this result is worthless. Mean 60 is not meaningful. The purpose of problem was not to find out mean. This result was not ends, This is only a mean which we had been selected to sort out the problem. This means under some fix limit number of boys obtain marks or can had been to get percentiles. Like this way if problem is " Boys and girls mathematics' marks obtaining any differences is or not and we having marks obtained by them, taking difference of their mean marks and having their tested, states that " difference is worthful" Which is only a result of analysis of data, then it has no meaning. Our problems were to know that their "difference in achievement or not". It is clear from above result that research datas' analysis got available statistical result, not got avails sort out problems. Sort out of problems these results' in the context of purpose of problems having explanation are available. Like that on the basis of analysis of descriptive studies data which facts, characteristics, cause, relation etc prevails, they are only unreasonable, meaning less, uncoherant and figures. It becomes worthful only when it is added to problems and its statement is taken. In the

form of raw material by the analysis of data gets results with context to problems to explain and to narration them is called establishment of conclusion.

Establishment of conclusion and Process

Did You Know? Conclusions of research are the tested solution of problems.

Their propagation is done by analysis of data on result basis. By this approach result of analysis and conclusion establishment are two different stages of research process. For conclusion establishment getting result by data analysis is to add results with problems e.g. in the context of purposes of problems having done their explanation that is restatement. Almost in all research hypotheses is the gateway to purposes. Therefore these hypotheses presents that context under the limit of that establishment of conclusion and do statement is essential. These hypotheses context makes conclusion relevant and conceivable. If we say " Average educational achievement is 60" then it has no meaning. To make it worthful hypotheses " Academic achievement is satisfactory" its explanation will be done in that context and will be made sure that " average achievement 60" is satisfactory or not. For this available measures or available evidences anywhere, preedited research result, having compared with education application recommendation critically result will be established and as such statement factory. In the whole process of it researcher's mental ability, knowledge of subject and imagination power have great importance.

Conclusion establishment is a specific activities. That form depends upon method of research. For example, historical studies' forms of conclusion historical facts, their hypotheses, and depends upon review of historical facts. In experimental review of historical facts. In experimental research this applies characteristics of Psychological test, types of measures and research design is depends upon.

Conclusion's establishment's level researcher having gone in the depth of true results and reviews in the background of that theories and practibility.

In the interval of results vested and invisible relation and indicates towards process. From which point of view gets result and conclusion seems to be true, explains that. By the medium of this explanation importance of conclusion is exposed. For example, result "In unmarried persons suicide tendency is found much more" that this explanation that "between loneliness and suicide may be relation, by this approach may be important that one theory (Durkhim's theory) supports **Durkhim's** theory has got support from this conclusion. In this way explanation of result is increased the importance of conclusion. Explanation of result throws light on that matter that from which point of view result is important. Under the explanation of result researcher considers it that what is his purpose, what is their relevancy, what answer they presents of basic problems etc.'

For the establishment of conclusion on the basis of data analysis obtained result from statistical point of view explanation is needed. Under this relevancy of result is examined. These methods are use only those studies whose data is quantitative and their analysis is done by statistical methods.

The base of result in these studies, selection a sample through a definite process, but conclusion are applicable on whole population. This process is called the process of generalization. On the basis of it wideness of conclusion is to found out, e. g. it is to know that what is possibility of true to be of that conclusion? In simple language, its meaning will be how far that conclusion true about total population. The base of generalization is to be sample's population's representative. If sample is full representative of population, only then this type probability statement, e.g. validity of generalization determination and mention possible. At that condition the use of that methods are possible, on that basis relevancy is tested.

Notes

These methods are-

1. Standard error of mean (S.E.m)
2. Standard error of difference of two mean (S. E. D)
3. Standard error of correlation coefficient (S.E.r.)
4. Chi-square
5. F-ratio

5.4 Concept of Generalization

Concept of generalization's use is done in the testing of significance of conclusions. Sample based result population's (whose sample representative) how far indicator of that characteristics, to get the answer of this question available on the basis of process of generalization, but such a statement is possible only then when sample did representative of population and selection has been done randomization, but generally it is not possible that sample did representative of population cent-percent.

Their possibility to be true (e.g. their possibility of some, which in that condition when research is done on whole units of population) is not cont-percent. How much this is possible, to do statement of these is to be the test of conclusion's worthfulness. For this levels of confidence is to be used.

Task Express your view on generalization rule.

Laws of Generalization

1. Generalization will be as much as useful and valid as sample will be representative of population. If sample is not representative of population then their result cannot be true about population i.e. at that condition generalization is not possible.
2. In which conditions research has been done, those condition will be as far as alike to condition of population, as far as generalization possibility to be true. It is too controlled condition research has been done under laboratory, then its result will be more accurate, but their generalization possibility will be less because outside laboratory such a controlled condition will be not.
3. Having done the experiment on a few sample, possibility of generalization of results is less.
4. Insufficient or incomplete data narrow down the validity of generalization. Therefore conclusion should be based on adequate accurate data.
5. Generalization should not be against predetermined facts, theories and nature's laws.
6. That should be in the form of either in solution of problems or in the form of answer of problem's question.
7. At the time generalization of conclusion, limitation of research design, the inadequateness, demerits etc., should be paid due attention. For example if correlation design has been used then it is keep in mind that generalization can not be done about cause-effect relationship. It is weakness of this design. Like these every research have its own demerits. It should be keep in mind.

Notes

5.5 Summary

- There are various methods of research and their data is also different from each other.
- There is not a single method or technique of analysis and interpretation of data.
- Generally analysis of data can be divided in two group-(1) Qualitative analysis and (2) quantitative analysis.
- Historical studies, some descriptive studies and life sketch studies' data are qualitative and descriptive.
- In different conditions different statistical values are calculated. Which statistical method will be used for the analysis of data, it depends upon purpose and hypotheses of research.

5.6 Keywords

1. **Generalization** – Generalization to make general data from research point of view so that sample can be representative of population.
2. **Observational**- Able to observe.

5.7 Review Questions

1. What do you mean by qualitative analysis?
2. What do you mean by quantitative analysis? Describe the methods of quantitative analysis?
3. What do you mean by conclusion establishment? Describe its form and process.

Answers: Self Assessment

1. Qualitative	2. External	3. Numbers	4. False
5. True	6. True	7. False	

5.8 Further Readings

Books

1. Educational Research Methods—*Shareen and Shashi kala, Vinod Pustak mandir*
2. Educational Research's Procedures—*L. Kaul, Vikas publishing*
3. Educational Techniques—*S. S. Mathur, Bhatt Bothers.*
4. Educational Techniques—*R. A. Sharma, Bhatt Brothers.*

Unit 6: Preparation of Research Report

Notes

CONTENTS

Objectives
Introduction
6.1 Format of the Thesis
6.2 Style of Writing
6.3 Construction of Tables and Figures
6.4 Presentation of Reference Materials
6.5 Presentation of an Appendix
6.6 Typing of Report
6.7 Evaluation of Report
6.8 Summary
6.9 Keywords
6.10 Review Questions
6.11 Further Readings

Objectives

After studying this unit, students will be able to:

- Understand format of the thesis and style of writing
- Understand construction of tables and figures
- Understand reference material and presentation of on appendix
- Understand typing of report and evaluation.

Introduction

After the completion of research work whole description is presented in the format of thesis. How thesis should be prepared, regarding this at international level with a little change has followed a universal law. For researches knowledge of there rules and methods is essential. According to there rules and methods he should have prepared his thesis. It has been described in this chapter whole materials of chapter has divided in following parts.

(a) Format of the Thesis
(b) Style of writing
(c) Constructions of the Tables and Figures
(d) Presentation of Reference Material
(e) Presentation of Appendix

Notes

(f) Typing of Report

(g) Evaluation of Report

6.1 Format of the Thesis

Under this, this matter has been explained that after the completion of research, how should be presented their whole description in written format mean, how should have been thesis's written format. From this point of view whole thesis is divided in following

1. Preliminary Needs

(i) Title Page

(ii) Letter of Approval

(iii) Acknowledgement

(iv) Preface

(v) Contents

(vi) List of Tables

(vii) List of Figures

2. Main Contents

(i) Introduction

(a) Statements of the problem

(b) Purpose of the problem

(c) Analysis of related Researches

(d) Basic Concepts of hypothesis

(e) Statement of Hypothesis

(f) Describe of Technical Terms

(ii) Survey of Related Literature

(iii) Process and Method of Study

(iv) Result of Study

(v) Explanation of Result of study

(vi) Summary (whole description's)

3. Reference Materials (Bibliographic, Appendix, etc.)

1. Primary Needs

In each thesis prior to describe of research, according to rule and tradition of universities some pages are added due to formality. In these title page, letter of approval, acknowledgement, preface, content's (catalogue of chapters) list of tables and list of figures pages are included.

I. Title page

In this page which information is given- that is as follows- (i) heading of the problem, (ii) Name of the researched (including titles), (iii) Faculty and name of that university where thesis is to be presented,

Notes

(iv) Name of the title for which thesis is presented, (v) the year in which it is presented. How these information should be typed on the pages. Regarding this special rule is as follows-

(i) Heading of problem is typed in capital letters. Between the page both side left equal margin from top 1½″ space is left. If title is more than one line then it is printed in double space and reverse pyramid format. See specimen on next page. This is in English but in Hindi it should be appeared in some format.

(ii) This page's important term's first letter of other all information's is typed in capital letters. Saw specimen given on this page.

There is no question of capital letter in Hindi. Hence problem heading in a little large size and rest all information can be typed a little less size. In electrotyping such an arrangements is given. If there is not such arrangement then problem heading can be under lined.

II. Letter of Approval

If according to rule of university letter of approval attachment is necessary then done their photocopy attached in thesis as per as.

A STUDY OF THE RELATIONSHIP BETWEEN PERSONALITY FACTORS AND ADMINISTRATIVE EFFECTIVENESS OF THE PRINCIPALS IN ORISSA

By
NARAYAN TRIPATHY
M.A. (Education), M. Phil, (Education),
Lecturer (Education)
Bhadrak College, Orissa

A THESIS
Submitted to the University of Meerut (U.P)
in Fulfillment of the Requirement for the
Degree of Ph. D. in Education.

Supervised by

DR. R.P. BHATNAGAR
Retd. Professor and Head,
Department of Education,
Dean, Faculty for Education,
Meerut University, Meerut.

DEPARTMENT OF EDUCATION
Institute of Advanced Studies
Meerut University (U.P.)
1984

TABLE OF CONTENTS

APPROVAL SHEET	II
ACKNOWLEDGEMENT	III
LIST OF TABLES	V
LIST OF FIGURES	VII

Notes

Chapter

1. INTRODUCTION 1

Background of the Study 3

Statement of the Problem 7

Objectives of the Study 8

Hypothesis of the Study 9

Significance of the Study 10

Definition of Terms 12

Scope of the Study 14

II. REVIEW OF RELATED LITERATURE 15

(Subdivisions as in the First Chapter)

III. METHOD AND PROCEDURE 40

IV. THE DATA AND THE RESULTS 82

V. INTERPRETATION AND DISCUSSION OF RESULTS 102

VI. IMPLICATIONS OF FINDINGS AND SUGGESTIONS FOR FURTHER RESEARCH 114

VII. SUMMARY AND CONCLUSION 135

BIBLIOGRAPHY 156

APPENDIX-A 177

APPENDIX-B 181

LIST OF TABLES

Table

1. Mean and Standard Deviation of Group-A 30
2. Mean and Standard Deviation of Group-B 36
3. *t*- Values on All Variables between Group-A and Group-B 40

LIST OF ILLUSTRATIONS

Figure **Page**

1. I.Q. Distribution Curve of Boys 38
2. Comparative Bar-diagrams Showing Achievements of Boys and Girls 46
3. I.Q. Distribution Curve of Girls 58
4. Comparative Bar-diagram Showing I.Q. Distributions of Boys and Girls 60

III. Acknowledgment Page and Preface

In each thesis after letter of approval one on two pages are acknowledgement and preface. Preface deals in very short purpose and scope of research is to be mentioned. After that from that persons get important help is research work, express grateful to all of that, but gratefulness should not be exaggerated. This should not be flirting. Should be objective. Whenever it is not necessary gratefulness and acknowledgement do not keeps separately rather it should be written together.

Notes

IV. Contents

After that in each thesis those are the pages in which it is shown that what is in thesis very first page of it space leaving 1" from top and in middle of page "Vishay-Suchi" (if in Hindi) or "Table of Contents" in English is typed. From that, "leave space left side in capital letter, approval sheet, acknowledgement and preface etc. chapters and bibliography etc. heading given double space between each are typed. If thesis is in English then from where chapter begins, above that chapter heading is typed whose only first letter is capital. Each heading's right side there page number is given from where page number is starts, above ½" that heading page is typed whose only first letter is capital. There is no question of capital letter arise in Hindi.

From letter of approval upto table of figures number of pages should be in Roman digit (as a I, III, IV etc.) and after that all page are numbered in Arabic digit. Bibliography and appendix's page number 100 in order of thesis's page number goes on ahead.

Chapter headings is in capital, but their sections heading in small letters (first letter in capital) are typed each titles page no. as per given in thesis and term of heading also as per thesis.

As far as possible heading should be in one line and syntax should be same. Contents specimen has been given on the previous page.

V. List of Tables

After content those pages are on which tables and figures list in presented list of tables and figures is typed on separate page. Under this which information are presented, that is (i) order Number of tables and figures (as per thesis) which is types in Arabic digits, (ii) heading of tables (iii) Number of that page on which their available in thesis on the previous page its specimen is given, see that.

2. Main Contents

Notes In the Main contents of thesis, research work is described. This part is mainly readable and of permanent importance.

Under this basically six chapters are included. (*i*) Introduction (*ii*) Survey of related literature (*iii*) Design of the study (*iv*) Analysis of data, result and their interpretation (*v*) Utility of conclusion and suggestion for next research (*vi*) Summary

I. Introduction

Under this heading researcher presents whole introduction means he throws light on that how problem is originated, how it is differ from related research, what is their extensive meaning, what are their variables, what is their meaning, what is the purpose of problem, what are their hypothesis, what is their scope, what is their importance, etc.

II. Survey of Related Literature

Under this related to problem whatever literature is available, their to be studied if it seems necessary at suitable place it is mentioned. Some educationists its mention in whole thesis where consider to necessary and significant, thereby should be used, in the favour and for that a separate chapter made don't consider appropriate. Regarding this it is appropriate that or he does not think so, research should do accordingly his own thesis guidelines. If a chapter is prepared then it is divided into many perspective for fulfillment of different purpose it should be mentioned. Having studied of previous research in the field of problems or have got answered of questions, next more how much, why and

Notes

in what direction need of research is. What is the importance of research that whole introduction, hypothesis of problems' how related literature presents a logical base, how has defined in related literature variables of hypothesis, how they have measured etc a number of aspect are in the explanation of that related literature and survey of informations can be much beneficiary. In this chapter what out to be and how whole matter should be composed regarding this not any format can be this is depends upon to a far extent types of problem.

III. Methods of study and research process

In the third chapter of management whole process of research is described. Researcher writes correctly and in details that how research has been done. Under this definition of population, method of selection of sample, definition of variables, description of test to use their measure, information's analysis method etc are described in details what will be written, this depends upon to large extent that what types of study is. Specially paid attention is that not any important information should be left. In reader's mind on reading time this question should not arise that how it had been done.

IV. Analysis of information and result

Generally in the fourth chapter of thesis data and information's analysis and research results are presented. This is an important part of whole research process. Data can be of quantitative or qualitative or anyone of them or both types. Regarding their presentation it will not be appropriate to mention any fix rule or process become this is depends upon that what types of problem is and what types of related data is. A general suggestion is that presentation of data and their analysis accordingly hypothesis. Having done the statement of hypothesis it should be explained that for the testing of that which types of informations and data have collected it should be presented in the form of table properly after that how their analysis have been done what result got, it should be mentioned in that order result's explanation should be given. Explanation means evolution of significancy of result is a tough job. In this researcher's method, short comings of process itself error occurred in the collocation of data, available result of previously done research etc keeping in mind and having mentioned 1+ it is to statement that how far logical the obtained result, probable and reliable are. If other research's result is different from any result have been come then it is essential to mention why it is so.

V. Suggestion for next research and utility of results

Not any research is comple in itself Hence researcher feels for a number of times during this work that if it is done by another types, another types of information is collected, big sample is selected or taking different population research is done analysis any other method is used then it will be much better. Similarly during research work. Some new research problems prevails in his mind, But he himself now could neither research on another problem and nor can change in their process.

Hence thesis's in the IV chapter or prepared a separates chapter-V presented in that. In this sequence obtained result of that chapter how can they useful in education's field mentions that general for this to keep a separate for this is considered to be more suitable.

VI. Summary and conclusion

This is final is final chapter of thesis. In this chapter researcher is presented whole thesis's work procedure and all chapters (except chapter-II) Summary. This is the chapter which is read much. Examined reads this chapter at first. In chapter-II survey of related literature is done. It is not necessary to present their summary separately it is to write in short with statement of problem. It is written in 15 to 20 pages sufficient.

Some researcher is submitted this chapter as an extract. In fact obstruct is differrent from "Summary and conclusion" of Chapter. It is a rule to write obstruct in approx 600 words. According to rule of university its 4 or 5 copies demanded separately. In somewhere it is tradition to add it in the begning. Its purpose to know reader should reat it completely or not.

Notes

3. Reference Material

After the completion of six chapters in the some order reference material is included in the thesis. Under this firstly presented bibliography.

After that in some order appendix is presented. How it can be written and presented, it has been mentioned in details ahead.

Self Assessment

Fill in the blanks:

1. After completion of research work…….. to be presented in the format.
2. Problem's heading in capital letters ……is done.
3. In each thesis after letter of approval acknowledgement and …… are.
4. From letter of approval to list of figures page number…… marked up in digit.

6.2 Style of Writing

In the previous section-9 it has been mentioned that what, where is written in thesis. In this section it try to explain how all those should be written.

Report of research or thesis is a scientific and technical report. That is read by experts that know it very well that in that where, what and how it should be written, Hence keeping in mind expert's approach the thesis should be written. At international level approved indices (Compbell, 1954 Dugdell; 196 Turabian, 1960) can be found introduction there approaches. Each researcher before to write thesis having read any one of them that instruction should be followed from begging to end some general suggestion are presented below:

1. Style of writing should be fact based objective, logical and simple. To say matter indirectly, that is not based on facts; being impulsive and sentimental, to say his matter being very defensive, try to affect reader unnecessarily express in style-there are considered to be the demerits of style of writing.
2. In writing style researcher's schoolship, the in intensive technical study, art of writing and mastery over language should be seems clearly.
3. **Collection of description--** The collection of description of each chapters should be in order and symmetrically means firstly it should be divided in main parts. Main parts heading in the centre of page and subsection's headings should be written left in the margin under each section and subsection related matter should be captured each section and subsection collection of thoughts should be symmetric means which thoughts prior to that or should be captured later that can written in that order. Something is written somewhere, some any another place, it should not be done. Related thoughts should be come at one place and in same order at which order they are related to one another. This coherence is most required aspect of style of writing. Every section and subsection have a theme. Keeping that in centre all thought around that should be collected systematically. In each section and subsection descriptions' thought' facts, logics as required should be made a series consequently it should be so much collected that their one link exit from between then series broken append at once.
4. **Language--**The purpose of research report is to give reader objective information of research. Language is an powerful medium of this information. Hence it should be such a type that it can succeeded in their purpose because p[purpose of report of research is to give objective information hence figure of speech language, use of tough language, is use of hard word,

Notes

rhythmic language use, use such a word which is not in use now a days should not use long sentences. Such a type of language creates difficulties to understand thoughts and facts. Language should be simple and clear so that study related matter understood quickly and avoid misunderstood. If any words or concepts use is inevitable about that it seems to be difficult to understand then and there clarify it two or three sentences.

5. **Balance and specification-** It is to be necessary in whole report's descriptive material presentation balance is must means description's different sections, Subsection etc write to need in as details, so will have written. Any me chapter very extensively and second unnecessarily in short write creates imbalance in report. This balance must be maintained in each chapter within also. Any important aspect is having written 8-10 lines and to finished it, and to finished it, and of general importance aspect is written in many pages is an enplane of imbalance. In report same point of thoughts or words or statement is such that which is specific. It needs stress on them so that they attracts reader. Hence it is to write in underlined, in capital letters or a different front style in this specification process balance is necessary also. Where is to stress necessary, only there is given. Worthless thoughts, statements, words to give on those is not appropriate.
6. **Unity and clearness of thoughts-** In the whole description of report the unity of thoughts and material is much essenmtial. Research's any one aspect related all thought should be appeared in one place. If pective is of formulation of one test then its whole description should be appeared in chapter there in one place. At different place or in that chapter or in other chapters related to that material should not be scattered. This unity of report presentation increased the clearness of report. Clearness is most important charactertices of report writing.
7. **Some other suggestion-**In the writing of report researcher should keep following matter in mind-

   (i) In writing of report third person is used, e.g. "researcher experienced, I experienced" not.

   (ii) In the text of report abbreviations is not used. Their use is prohibited in footnote, bibliography, appendix and table.

   (iii) Percentage is written in words, not% like this

   (iv) In the begning of sentence number is written in words, not in digit

   (v) That number which is less the 100, written in words, not in digit

   (vi) Between paragraph thoughts is given number then that serial number is written in brackets, as (3) or (c)

   (vii) Whatever done in past by researcher or by some other persons, at the time of their mention uses past tense as a "Selection of Sample" had been done by randomization and this was based upon city's 110 schools

   (viii) But when given table, given r4esult of research and pre established and universal statements, rules, theories is mentioned then use present tense. As a "It is clear frm this table' "Skiner's refroze theory indicates towards this' one the basis of analysis this conclusion is drawn' etc.

6.3 Construction of Tables and Figures

In each research report tables and figures from subject clarification point of view has great importance. That matter which takes many paragraphs to understand it clear from only one table study. Some

Notes

is to say about figure also Hence tables and figures how can be types, to know this for researcher is essential.

1. **Rules Regarding Tables –** Regarding construction of table some rule is as follows-

   (i) In a well constructed table a number of mutually related facts is adjusted such a way that he can express any one main though is as a – a samples many characteristics (as a total number of units, Number of School, number of boys and girls, all their different test scores etc) can be shown in one table. If it will be explained then to write a paragraph

   (ii) Table should be prepared in such a way that each matter understood one self means it should be self explanatory.

   (iii) General and simple numerical facts under text should be written reflective. There is no advantage to make table for that e. g. Total 120 school, 40 out of them had selected and in this 40 from each school 10 percent units had selected to show such a simple truth there is no need to prepares a table.

   (iv) Some general types statistical facts in the order of text on informal table (in which heading of table, column and rows are not) form can be presented writing one sentence after that insert: (colon) below that presented as a total number of boys and girls that appered in there three test- as follows was.

	Boys	Girls	Total
Intelligence test	170	180	250
Interest	190	80	270
Personality test	150	100	250
	510	360	770

   Such a type of table is not mentioned in content.

   (v) In that sentence as a context indicates towards table, after that sentence table is presented, Never before that it presented.

   (vi) If table is to large that after that sentence it cannot captured on some page then after ending the final paragraph should be presented to next page.

   (vii) If table is greater than half page then it should be presented in one whole page centre.

   (viii) Which tables that is great than one page, it should be presented in appendix.

   (ix) In whole report from beginning to end (including index also) all table number continuously.

   (x) There is two methods to write order number of table. In one on the top most written table ahead of it in Arabic written their table number and below that leaving two space their title is written. In the end of heading not insert any punctuation marks. If table in English then (TABLE) and below their heading both typed in capital letter. If heading is longer than one sentence then it is typed in single space in the form opposite pyramid. In the second method order number of table keep separate in one line written in the beginning of heading line, **— Table 40- group a intelligence test scores.**

   (xi) Table's columns and rows' heading should be written in very short and their order number in bracket should be written below that if heading is big then it be broad side typed.

   (xii) Table related footnote, just below table leaving a ½ space typed, not in end of page.

   (xiii) Column's heading above one double or darked horizontal rule and below that single line is drawn. In the end of table one single horizontal line is drawn.

Notes

2. **Rules related to figures-** Regarding presentation of figures following rules are-

   (i) Order number of figure in Arabic number below term figure- fig. puts ahead and their heading either to but word figure or written in that line or just up of figure or under the figure an where can be written in whole report from beginning to end all figures order number is putting continuously.

   (ii) Heading of figure in capital letter and if more than one lines then presented in the form of opposite pyramid, There is no punctuation marks puts in end of headings.

   (iii) Figure is presented after just that lines in which there caption has come.

6.4 Presentation of Reference Materials

In report each researcher is to present a number of authors, researcher and expert's view, statements etc. context here and there. In each chapter such references are included. All of those separately symbol number within the text where context is come to be presented and in the end in the bibliography their full detail is to given. In anywhere complete description in the below page in form of footnote is given. Footnotes and catalogue of context books presented to complete description a little difference is occurred. It will be clarified later.

There are a number of methods to denote symbol marks of context within the text. Most general method is that which context appears where, there in bracket author's name and after that putting come the publication year of that book is inserted as a (Bhatnagar, 1993). All there context book prepare a catalogue alphabetically added the end of theirs order second method in bracket only write that order humber which appears in the end of book in catalogue is as a (60). Its meaning will be full description of context books in catalogue order number is given at 60. A third method is based on footnote. Under this context appears within the text in super of context footnote order number putting in Arabic digit whose meaning is that their full description bottom of page is given in the footnotes.

Presentation of quotation in report, some rules also which are as follows:-

First of all it should be remembered that it is not considered to be good to use much quotations in the reports. It is quite inevitable then quotation should be used other rules regarding this is given below.

1. Quotations

Regarding quotations some rules are as follows.

(i) Quotation should not be too long. Small and under stance ended quotations is considered to be good. It should be closed with quotation

(ii) Under quotation if appeared another quotations then it should be enclosed by single quote.

(iii) Last word of quotation up half space typed context in Arabic digit if sentences is typed just after punctuation marks. This symbol number should be same as given in footnotes with their details

(iv) Long quotations (four or more than four typed lines) in a separate paragraph in single space in the centre of page without quotation marks is typed. In this long quotation if any another quotations comes then it should be closed by double quotations marks. The quotation which is abnormally long that is presented in appendix.

(v) If any footnote is presented by in the format of footnote then in the form of paragraph in single space and kept under quotation marks.

Notes

(vi) Kf first word of quotation from grammatically in sentence aforesaid which has been said related to that then it is written in small rather than capitals, whether in original sentence it may be capital, such a question arises only in English, in Hindi there is no such question arises.

(vii) After last quotation marks only one punctuation marks is used. If full stop or coma then that is under quotation marks, colon or semicolon are then outside quotations marks if question mark or exclamation mark keeps under quotation marks. But question mark is not a part of quotation, part of a sentence then it is kept outside quotation marks.

2. Footnote

Regarding footnote rule are as follows:

(i) Footnotes in the bottom of page below text typed given three space from left to right one space Below text one two inch long line from left to right is typed. From two space below of that footnote is typed.

(ii) Footnote is typed in single space, but between two footnote are given two space

(iii) First line of footnotes like Para line typed from left to right leaving three space and their first word's half space above their citation number which comes in text, that is typed.

(iv) The number of footnote given either by page wise or in whole report from begin to end in a series. It is more convincible to put down pagewise

(v) If cited material is any formula of table then on that not put quotation symbol number rather it is shown by any asterisk or any other symbol mark.

(vi) In thesis published essays and published books and in report which method is followed that is differ from thesis method

(vii) If in a context booked footnote, some pages and some footnotes appears again then putting their context symbol number only that is (lbid) is written. If context material is some, but page number is differ then some page 8 (lbid, p.s) as follows is written

(viii) If any footnotes references has been comes before after that another footnotes appears after that again some context comes, then in footnote author's name, after that semicolon and after that abbreviations (op.sit) and their page number is written. If any writers'. More then one books citation presented in thesis then related to context specific books' name is also given each times but the page number is some then 'loc.cit' is used.

These three abbreviation's uses specimen has been given below-

1. P. Bhatnagar, In applied sciences research's experimental design, Loyal Book Depot. Merrut 1993 page 220
2. lbid
3. lbid 214
4. Mauli, Jeorge J. The Science of Educational Research, Ureshiya Publishing House, Pvt. Ltd. New Delhi 1964, page 308
5. Bhatnagar, loc-cit (means Bhatnagar's book and same page)
6. Waan Dalen, D.B. Understanding Educational Research, Macgraw Hill Inc. New York, 1966, Page 360
7. Bhatnagar op. Cit. 260
8. R.P Bhattnagar, Education Administration, Loyal Book Depot. Meerut 1989, Page 52
9. Bhatnagar "Design of Research" Op. cit page 240

Notes

Some other mostly used abbreviates are given below. It is generally uses in books and thesis

Ch., chap, chaps.	Chapter (s)
e.g.	for example
ed., edd.	edition (s)
ed., eds.	editor (s); edited by
et. al.	and other (Bhatnagar et al.)
i.e.	that is
MS, MSS	manuscript (s)
mimeo.	mimeographed
n., nn.	footnote (s) *e.g.n.* 10 or nn. 2.4
p.pp.	page (s)
passim	here and there
Pt. Pts.	part or parts
sec., secs.	section (s)
trans.	translated by
rev.	revised or revision

3. Bibliography

(a) Bibliography is presented in the end of thesis in chapters "summary and conclusion" It is included under appendix. The presentation of footnote and bibliography appears a little different. In footnotes author's first name comes first and later on last sub name e.g., R.P. Bhatnagar. But in reference book his surname comes first and initial name comes later, e.g; Bhatnagar, R.P. Second difference is that footnote is types as paragraph means their first line from left, leaving three space from margin is typed so that other lines of below appears to be left side but is context catalogue first line starts from margin and other lines starts leaving 3-4 space leaving from margin and typed. Thus is context catalogue first (top) appears in the left side where as other lines shifted within it will cleared from below given example.

Footnote – (For books)

R.P. Bhatnagar; **In applied Sciences research's design**, merrut: Loyal book depot Second No. 1993

Bibliography (For book)

Bhatnagar, R. P. **in applied science research design,** Merrut: Loyal book depot; Second No. 1993

(b) **In bibliography whole order of description is as follows-** (i) Author's sub name (ii) Actual name (iii) Title of books which is underlined in typing thesis (iv) edition (v) valume if more than one (vi) If more than one (vii) Publication place (Viii) Name of publisher (iX) year of publication of copyright year.

(c) **In any journal printed essay's context description is as follows-** (i)Sub name of author (ii) Actual name (iii) Heading of essay (iv) Name of journal which is underlined (v) volume number (month, day , year (vi) Page.

(d) **In daily news paper printed context material description is presented by following types-** (i) News paper is underlined, (ii) Month, day, year (ii) Page if authors name is also then like books comes before news papers name.

(e) **Unpublished material context description is given as follows-** (i) author sub name (ii) Actual name (iii) Heading of material which is underlined (iv) Types of thesis e.g. Ph. D thesis, or report or any paper presented in press conference etc. (v) Place where presented or where available. In bibliography context description regarding some rules are as follows-

Notes

(i) If one writer more than one bibliography appears then there is no need to write authors name each time. After first context some writers appears second context then rather to write the name of author a six space long rule is drawn. After that putting come rest description is presented after first context will be appeared context presented alphabetically.

(ii) If any bibliography two or three writers are then first writers sub name comes first then actual name after that, but later both actual name comes first and sub name after that if number of author more than three then only given first writer's name after that putting semicolon at or and other is written.

(iii) If there is no one author of book rather than editor writing their name after that in bracket (ed) is written.

Example It is sufficient to write city name in place of publication place, but place is such that it is little known then their full address is written necessary.

Self Assessment

State whether the following statements is True or False:

5. In the third chapter of thesis, research process is mentioned.
6. The purpose of research report is to give reader's objective information of research
7. In report the number less than 100, is written in digits.
8. No research is perfect in it self.

6.5 Presentation of an Appendix

In every thesis there are three types of appendix added- (i0 Bibliography, (ii0 Psychological tests and other tools which is used in collection of data and (iii0 statistical material etc, e.g., different tests scores long table, etc. which had not been presented in the chapters of thesis. Among those first of all, placed bibliography. The one page before first page on fly sheet in the middle appendix (a) bibliography is typed from next page list begins. At that page in the top heading "Bibliography" is typed pagination is done in the order last page of last chapter. Flysheet is also included in pagination. First page of bibliography and page number of fly sheet like other chapters heading page in the bottom of page putting in centre with all other appendix flysheet in also appears. That is also included in pagination. In appendix (b) Psychological tests and other instruments, its flysheet appendix (c) Tests and tools is typed. Similarly in appendix (d) statistical and other data appears. On that flysheet appendix (e) data is typed. There is no need to give calculation of result under this. But raw data must be given without that thesis is considered to be incomplete.

6.6 Typing of Report

Did You Know? For typing of report white 'Brand' paper is used.

Notes

Generally $8\frac{1}{4}$" x 11" Size paper is used for typing that weight should be 20 pounds. For second and third copy light paper can also be used.

Report should be typed by an experienced typist has typed some report prior and that it knows how tables, figures, footnote, bibliography are typed. After that researcher himself should guide him. Typing pages should be read side by side and mistakes should be removed. It is a tradition in typing to leaving space $1\frac{1}{4}$" top and bottom $1\frac{1}{2}$" in the left side and 1" right hand margin. It is re commended that to use PICA type typewriter, but electronic typewritten can be used also whatever typewriter is used from beginning to end that typewriter should be used. There is no need of carbon copy any more, because all researcher will have used photocopy. But if to do carbon copy then a black carbon paper and medium range ribbon should be used. Before giving hand-written manuscript read it thoroughly. Generally in the pagination of footnotes, bibliography, examples, tables and figures mistakes have been found. Having repeat a many times should be checked.

In the each page of typed two or three corrections are valid, but more than this cutting, increase number of words, add sentences in the middle is not approved such a page should be typed again.

On each page of report page number is printed. Preliminaries needed pages are given roman number and it is printed at centre of bottom. On titl page do not put any page number, but it counted as (I). Hence next to this page number II is put on approval letter neither page number is put on nor counted it. Rest on all pages in Arabic number top of the page right hand side number is put on, but each chapter's first page number put on centre of bottom.

6.7 Evaluation of Report

After the completion of report writing before giving for printing the researcher should evaluate it himself. As there is not universal base of evaluation, no any fixed criterion of that, but some writers regarding this give some suggestions. That can help researcher. **Waan Dalen** has presented a questionnaire in his book. This question of that questionnaire helping in to know how much report prepared by him is of reasonable level and what shortcomings have been left in them. **Waan Dalen** has included all aspects related question of thesis in this with some correction that questionnaire has been presented here.

1. Headings of problems

(i) Does heading explain scope of problem correctly in a definitely format?

(ii) Has it been written clearly, precise and descriptive format

(iii) Have unnecessary words and phrases eliminated from that?

2. Primary Needs

(i) Have tile page of report, letter of approval, acknowledgement, table of contents, table list and figure list all systematically and according that rules which mentioned already?

(ii) Has all those paginated correctly?

3. Problem statement and their description

(i) Have all those aspects of problem analyzed correctly, which is important as a previous have done research's relation with problem their variables definition, importance of problem, their purpose, etc?

Notes

(ii) Which relation has been supposed between problem based variables, has their probability logical verification been done?

(iii) Has problems to be fundamental clarification done?

(iv) Has problem's scopes clarification done?

4. Related literature and survey of informations

(i) Has related to problem based variables previously did research's completely surveyed been done?

(ii) Has that survey been done from criticism point of view?

(iii) Has from that survey the conclusion is drawn that on present research on problems was to be done necessary?

(iv) Does survey gives as the critical base of problem?

5. Formation of hypothesis and statement

(i) Is there any theoretical base of hypothesis?

(ii) Can hypothesis be tested?

(iii) Are deduced consequences logical?

(iv) Has hypothesis been written clearly and in short?

6. Method and process

(i) Has population been defined corrected?

(ii) Is sample representative of population?

(iii) Have for the collection of data suitable instruments, reliable and valid tests been used?

(iv) Are those instrument and tests suitable from another point of view?

(v) Have these instruments and tests been used properly?

(vi) Have necessary material and information etc been collected?

(vii) The method which has been used has their satisfactory reason been presented?

(viii) Has previously done research used method clearly been critically explain?

(ix) Have it been mentioned that from where and how data collected?

(x) Has in use of statistical methods their underlying assumptions been paid attention?

(xi) Has research's internal and external validity's effecting elements been tried to control?

7. Presentation and analysis of thesis

(i) Has collected data's presentation according to rule been done?

(ii) Has each hypothesis's test been done independently?

(iii) Has conclusion's objectively previously done research' introduction been explained?

(iv) Is result is based upon research data?

8. Summary and conclusion

(i) Is summary short and clear?

Notes

(ii) Is summary based upon research work?

(iii) Are given suggestion for next research solid and logical?

(iv) Has result's utility been explained?

9. Bibliography list and appendix

(i) Has bibliography, style and construction according to rule done?

(ii) Have all bibliography been included?

(iii) Have all required materials been presented in appendix?

10. Report's format and style

(i) Have chapters' composed accordingly rule?

(ii) Has pagination been done correctly?

(iii) Is language and style according to rule?

(iv) Has in each chapter material description been presented by balanced manner?

(v) Have in the printing all those precautions been taken which are mentioned already?

6.8 Summary

- After the completion of research work whole description is presented in thesis format in which format thesis should be prepared, regarding this at international level with a little changing general accepted rules are followed.
- In each thesis, before the description of research according to the rules and tradition of university some pages are added due to formality reason in these title page, letter of approval, acknowledgement, preamble, contents (chapter list) table list and figure list pages are included.
- In each type of thesis three types of appendix are added – i. Bibliography ii. Psychological test and iii. Statistical data
- After the completion of report writing and before giving to printing, researcher should do assessment himself.

6.9 Keywords

1. **Approval**—Support, Taking approval before starting any research work.
2. **Appendix**—In the end of thesis writing given supplementary material.

6.10 Review Questions

1. Describe the format of description presentations?
2. What do you mean by style of writing and how it should be written?
3. How tables and figures are drawn?
4. How cited material is presented?
5. What do you mean by assessment of report? Describe it

Answers: Self Assessment

Notes

1. Thesis	2. Printed	3. Preamble	4. Roman
5. True	6. True	7. False	8. False

6.11 Further Readings

Books

1. Academic Research work procedure—*L. Kaul, Vikas Publishing.*
2. Academic Research Methods—*Sharin and Shashikala, Vinod Pustak Mandir.*
3. Academic Technique and Assessment—*Dr Rampal Singh, Bhatt Brothers.*
4. Academic Technique Management and Assessment—*J.C. Agrawal, Bhatt Brothers.*

Notes

Unit 7: Probability : Normal Probability Curve and its Uses

CONTENTS

Objectives
Introduction
7.1 Characteristics of Normal Probability
7.2 Normal Probability Curve
7.3 Uses of Normal Distribution
7.4 Uses of Normal Probability Distribution Diagram
7.5 Measuring Divergence from Normality
7.6 Summary
7.7 Keywords
7.8 Review Questions
7.9 Further Readings

Objectives

After studying this unit, students will be able to:

- Understand the characteristics of normal probability
- Understand normal probability curve
- Understand use of normal probability and use of diagram
- Understand the measuring divergence from normality.

Introduction

When mean of any diagram's both sides data lies in such a way that if we show them by a curve diagram and made following type bell – shape then we call that type of distribution as normal distribution.

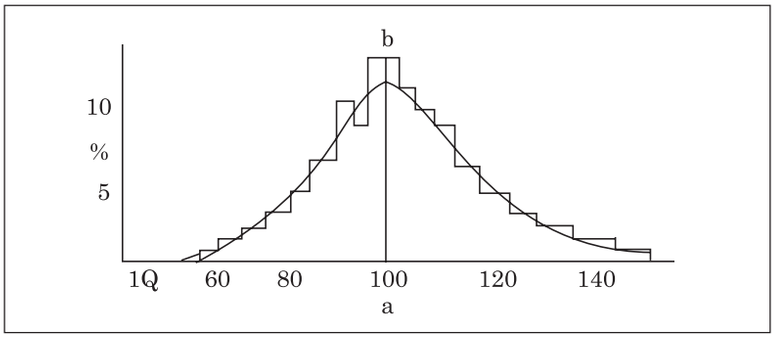

Fig. 7.1

Notes

In this type of distribution maximum number of data lies in the mid and as we precede both side from mid point then score of data will become less. In the end of both sides minimal score remain. Second matter is that both side left and right of mid-point in the position of distribution equality.

If we divided in two equal part at point *a* of above figure then *a* will be the mid-point. If we draw perpendicular a b then both side of line a b divided domain will be equal in shape and area. This is the **main** characteristics of normal probability.

Secondly we observe that mostly scores lies closely to mid-point and as remove from beginning both sides, scores is least in equal ratio and in the end of both side minimal scores remain.

In this figure if at point a mean 100 lies then we observe that in both side on equal data are. In one side less than 20 then in second side more than 20. In one side between 80-100, 20 score lies then in second side also between 100-200, 20 score lies. Again in one side between 60-80 below 80 only 20 score are removed. And in second side also as it is. In this way both sides have complete similarity. As either side of a. Both sides of scores has removed equi-distance. In one side as much as less than 100 in other side as much as more than 100. This is the meaning of normal probability.

In statistics normal probability has great importance. In some places use of normal probability is of great worthful. On the basis of normal probability our mean's utility, relevancy and justification is determined. This is clear from ahead instance.

1. 55, 50, 54, 52, 50, 50, 60, 70, 75, 84 mean = $\frac{600}{10}$ =60

2. 60, 70, 60, 50, 68, 66, 62, 56, 54, 52 mean= $\frac{600}{10}$ =60

Both above distribution mean is 60. But is their relevancy a great difference. One is even and other is odd. In the first condition range of data is from 50 to 80; in the second it only in 50 to 70. Both mean is 60. In this way we observes that in the previous example in one side below 60, minimal number is 50 and maximal 84 E. G. in one side range is 10 and in second side 24. In other words above 60 range has been gone upto 24 and in the below has been gone upto only 10. In this way in either side of mean skewedness has been occurred. If we shown it by a graph then this type of shape will be made.

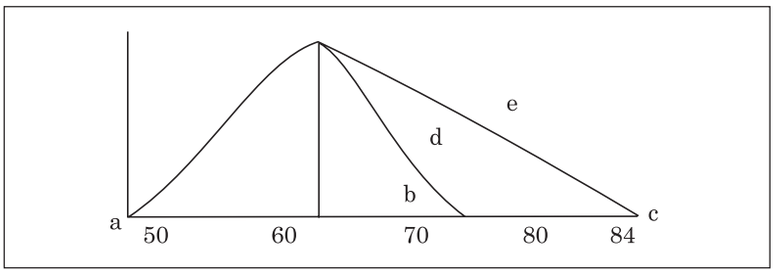

Fig. 7.2

In this figure's in b position neither scores exists around mean 60 nor exists equal-distance. It is clear form that more than 60 is much and less than 60 is least. The meaning of it will be that mean 60 is not proves to be more useful. He has not the capability at scores correctly. By meaning of mean 60 almost whole data should be equal to 60, but we observe in above condition some data is equal to 84 also. A great number is greater than 60. In the other hand minimal scores is equal to 50. Rather it is in some degree around 60. Therefore both sides of mean distribution has no equality. This is called skewed distribution.

Notes

In the second condition a mean is only 60 also, but at that condition range of distribution is upto 50 to 70. e.g., in one side less than 60 is 10 and on the second side more than 60, range is upto 10. It is clear that as score less than 60 as much more than 60. Both side distribution is same. Therefore is this distribution mean 60 express data more correctly. This is called normal distribution.

7.1 Characteristics of Normal Probability

Normal distribution have following characteristics:

1. If N is 100 and their distribution is even then in this position up and down of mean, e.g. 50 percent scores will be. If we do not assumes normal distribution range 100, assumes only 1 then up and down of mean range will be .5.
2. In this way both sides of mean upto limit of ±1σ, aggregately 2/3 part covers. If total range area 100 then upto +1.0 σ to -10 σ between approx 2/3 e.g. , 68.26 scores exists. If total range area assumes 1 then between mean +1.0 and (M)-1.0, .6826 part will be covered.
3. In the same way between the limit of M +2.0 σ and (M) -2.O σ approx 95.44 percent part covers e.g. if total range area is of 100 data then approx 95 data between +2.0 σ and (M) –2.0 σ will be lies. If we assumes total range 1 then between M± 2.0 σ scores will be .9544.

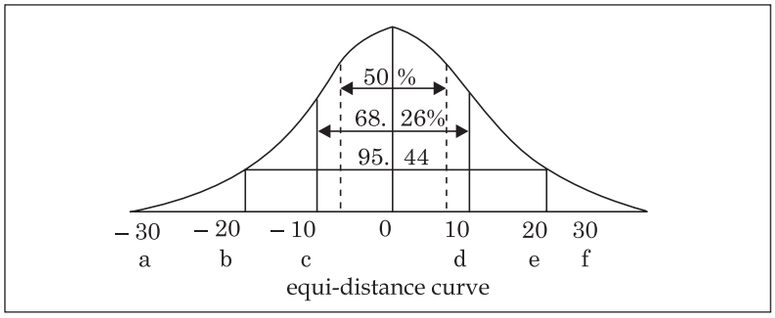

Fig. 7.3

In the someway between the limit of M ± 3 σ 99.74 percent data will be lies.

From the above figure it will be become more clear.

In this figure between point a and b limit explains limit of (M) ±1 σ. Between (M) + 1 σ is 1/3 part and between (M) - 1 σ 1/3 part. Between (M) - 1 σ 1/3 scores is below (M). e.g. lower. In the gone way between (M) + 1 σ one third data is greater than mean. In this way both sides from -1 σ to till + 1 σ total scores are two third.

In the same way between c and d 95.44 percent scores exists point c from mean point Lt. situated at distance -2 σ and d at distance +2 σ between these points 95.44 data are exists.

This means that small data below (M) [between (M) -2 σ] = $\frac{95.44}{2}$ =47.72 percent data will be exist and above (M) e.g. between (M) + 2 σ scores will be some.

Some way distance between e and f explains limit of (M) + 3 σ. The total number of scores between these is 99.72.

Example

Suppose 600 student I. Q's is given whose mean is 107.5 and (S.D) 8.6 and score is even.

Then normal probability's according to above characteristics among 600, 68.26% (409.56) student's I. Q's 107.5 + x 8.6(M ± 1 σ) e.g. will be 98.9-116.1%.

Notes

Similarly 95.44 percent's 107.5I ± 2 x 8.6 (M $\pm 2\ \sigma$) e.g. 107.5 ± 17.2 =90.3-124.7's will be. In other words in the condition of 95 percent, possibility is that their I. Q's will be between 124.7 and 90.3 only in 5 percent their I. Q's possibility beyond limit. If we taken limit of $\pm 3\ \sigma$ then it can be said that 99.72 percent's I. Q's 107.5 $\pm 3\ \sigma$ e.g. 106.4 ± 3 x 8.6 = 107.5 $\pm$ 25.8 = 81.7-133.3 (between) will be only 1 percent boy's I. Q's possibility beyond this limit e.g. in normal probability condition among 600 students only 6 students I. Q' will beyond limit of 81.7-133.3 and 594 will be under this limit.

(i) In this way percentage number vested in normal probability and by limit we know total data's status. This is the great characteristics of normal probability and it is their great utility.

(ii) Second characteristics is that in normal probability mean, medium and mode all lies at the mid-point of distribution and their value is equal also.

(iii) Third characteristics is that in normal probability condition inter quartile is of. 6745 stander deviation (S.D.), e.g. more than S.D. is 40% inter quartile (Q) is called Probable Error (P.E.) also. Therefore this relation can be expressed in this way also.

$$P.E. = 0.6745\ \sigma \text{ or } \sigma = 1.4826\ P.E.$$

Between Q and M from below e.g., less than that 25 percent data lies. In this way in normal probability Q's range 'middle 50 percent' data are and since in Q and P.E. normal probability is same, hence $\pm$ P.E. e.g. between $\pm$ 0.6745 σ (because P.E. = 0.6754 σ) above and below mean 25% data exists.

In this way to know the situation of data instead of $\pm 1\ \sigma$, $\pm 2\ \sigma$, $\pm 3\ \sigma$ we can use $\pm$PE, $\pm$2PE, $\pm$3PE.

7.2 Normal Probability Curve

This characteristic of normal probability based upon theory of probability. Theory of probability has great importance not only in statistics but also in all sciences. Therefore it is consider to be complete. If anyone incidence occurrence is of 'a' type and 'b' is such a type is which they can not occurred then probability of that incidence will be a/a +b. Similarly not probability will be b/a + b.

In this way out of 52 cards we draw 1 card and try to know it that what is the possibility of that card to be a king, then according to theory of probability we can say that its possibility will be $\frac{4}{52}$, because in whole cards king of paan is only one.

Since card are 52, therefore drawn cards of kings possibility or not possibility's total probability are 52 out of these four kings, Therefore in these by 4 types that cards possibility can be done and in 48 type its kings not possibility will be.

Therefore types of occurrence = 4, not occurrence types = 48.
Drawn cards kings probability (above according to law of a/a+b) = $\frac{4}{4+48} = \frac{4}{52}$.

Similarly according to rule of b/a + b not occurrence probability = $\frac{48}{4+48} = \frac{48}{52}$.

Example One coin on which one side Figure and other side year has printed toss. What would be probability that coins falling on earth Figure will be on up. There are two types of incidence happenings either Figure or year will be come. Therefore probability of Figure (P) $\frac{1}{2}$ will be and not happenings probability will be $\frac{1}{2}$ also.

Notes

If we added probability and non-probability, in every condition result will be 1. Now, if we toss same two coins then by following methods fall on earth probability will be:

F-F, F-Y, Y-F, Y-Y

Total 4 probability are, either both coins have Figure or both have year, and in one year and in one Figure or in one Figure or year F–Y or Y-F is something. Hence Y- F probability will be twice e.g. twice such will be happen that on one coin either Figure or year will come. In this way

1. Y- F probability (P) = $\frac{1}{4}$.
2. Y- F probability (P) = – $\frac{2}{4}$ = $\frac{1}{2}$.
3. Y- F probability (P)= $\frac{1}{4}$.

Adding left hand side and right hand side F–F + Y F+YY or $(Y^2+2\ F\ Y+Y^2 = (F+Y)^2 = 1$

Similarly tossing 3 coins following 8 combination will be get.

(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
F F F	F F Y	F Y F	Y F F	F Y Y	Y F Y	Y Y F	Y Y Y

But combination (2), (3), (4) are some. Hence, F, F, Y probability three times e.g. $\frac{1}{8}+\frac{1}{8}+\frac{1}{8}=\frac{3}{8}$ will be, since total 8 combination are and each possibility is $\frac{1}{8}$ e.g. $3F^2$ probability will be $\frac{3}{8}$ similarly F Y Y (5), (6), (7) combination's possibility will be 3/8 e.g. $3Y^2$ F probability (P) will be $\frac{3}{8}$ also. In other words on two coins sun and on one coins statue together come, combination is 3—It means it can happen thrice only.

Besides these probability possibility of FFF will be (P) $\frac{3}{8}$ YYY will be $\frac{1}{8}$ also.

Since,

1. FFF's probability (P) = $\frac{1}{8}$
2. $3F^2$, Y Probability (P) = $\frac{3}{8}$
3. $3Y^2$ F Probability (P) = $\frac{3}{8}$
4. YYY's probability (P) = $\frac{1}{8}$

Hence, $F^3+3F^2Y + 3Y^2F + Y^3 = 1$

(Adding probability of left hand side and right hand side)

or $(F + Y)^3 = 1$

Notes

Similarly we toss 10 coins then FY's (Figure and Year) Combination and their probability $(F + S)^{10} = 1$'s expansion, means

$F^{10} + 10F9\ S + 45\ F^8 S^2 + 120\ F^7 S^3 + 210\ F^6 S^4 + 252\ F^5 S^5 + 210\ F^4 S^6 + 120\ F^3 S^7 + 45\ F^2 S^8 + 10\ F\ S^9 + S^{10} = 1$

i.e. total occurrence of incidences will be

1 + 10 + 45 + 120 + 210 + 252 + 210 + 120 + 45 + 10 + 1 = 1024, out of these

1. F^{10} probability will be (Probability of 10 statue all together) $\frac{1}{1024}$
2. F^9 Y probability will be (Figure on 9 and sun on 1 probability =10) $\frac{10}{1024}$
3. $F^9 Y^2$ probability will be (Figure on 8, sun on 2 = 45) $\frac{45}{1024}$
4. $F^7 Y^3$ probability will be (Figure on 7, sun on 3 = 120) $\frac{120}{1024}$
5. $F^6 Y^4$ probability will be (Figure on 6, sun on 4 = 210) $\frac{210}{1024}$
6. $F^5 Y^5$ probability will be (Similarly) $\frac{252}{1024}$
7. $F^4 Y^6$ probability will be (Similarly) $\frac{210}{1024}$
8. $F^3 Y^7$ probability will be= (similarly) $\frac{120}{1024}$
9. $F^2 Y^8$ probability will be (Similarly) $\frac{45}{1024}$
10. $F^1 Y^9$ probability will be (Similarly) $\frac{10}{1024}$
11. Y^{10} probability will be (Similarly) $\frac{1}{1024}$

In this way on binomial theorm basis probability of incidences is determined.

Now if we prepared a graph from the combination of probabilities (1, 10, 45, 120, 210, 252, 210, 120, 45, 10,1) then it will be called normal probability histograms. For this we taken on x axis F^{10} Y, F^9 Y ….. etc. combinations and on y-axis their probabilities in this way following histogram will be, prepared:

Notes

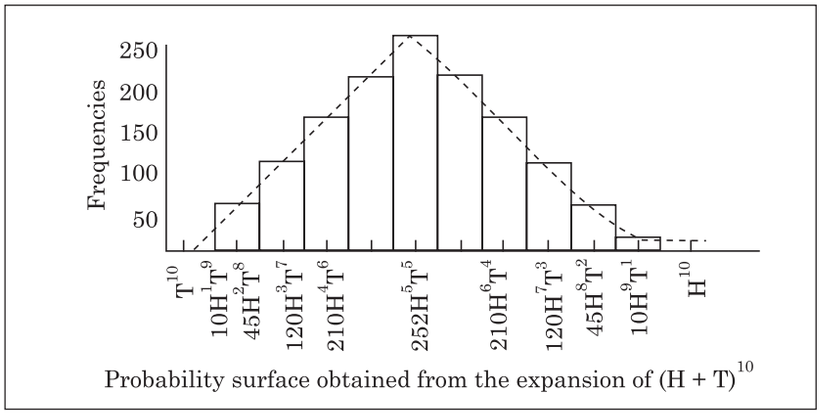

Fig. 7.4

From above histogram one thing is complete clear. Between 252 the greatest probability line on either side of it distribution of data is equal. On one side 210, 120, 45, 10, 1 is and on other side is 210, 120, 45, 10,1 the also. In this way in this histogram on either side of 252 area is equal and figure of both side some also.

In above histogram $(F\ S)^{10}$ express only the expansion of 10 coins statue and sun addition. Therefore in that 10 angle are seems. If the number of coins (N) upto infinite then in his to gram infinite angle will be made but they will not visible, and joining their points a smooth curve will be made that mid-point's either side of area like histogram will be some and either side of figure's part will be some also therefore this curve is called normal probability curve. This is called probability curve due to it is prepared on the basis of probability combinations. Secondly for this reason it is called that if having taken the mid point of this curve, divided it in two parts then both parts will be equal in area and figure. In the middle probability is most and again both side of its distribution is same, as is clear from above histogram.

On the basis of that normal distribution is explained and it is applicable. All characteristics of normal distribution is taken from this basis.

Self Assessment

Fill in the blanks:

1. In statistics…… has great importance.
2. On the basis of normal distribution…….. determines their utility, relevancy and justification.
3. In normal distribution mean, median and mode value normal distribution's…….. lies on point.

7.3 Uses of Normal Distribution

By the next two to three example equal distribution's use will be cleared.

Example 1.

The mean of an equal distribution data M is = 125 (S.D.) = 11 then (a) between 114 and 136 how many data lies among whole data? (b) how many data will be lies more than 140?

Notes

M = 125 S.D. = 11

To be find out number lies between 114–136

(a) M 114 is 11 less than mean 125 or -1 σ below. Similarly, 136 is 11 more than its mean 125 or +1 σ is distant or above. Therefore below 114–136 same number will be which will be from M between ± 1 σ limitations. Since in the condition of normal distribution between M ± 1 σ 68.26 percent data lies, therefore between 114-136 two third (2/3) of whole data or 68.26% data will be.

(b) From mean 125, 140 whole 15 is greater or from 125 to 140 distance is 15 scores. S. D. divided by 11, it will be said that from 125 to 140's number lies a distance of 1.36 σ. Therefore mean +1.36 σ (e.g. above mean one side) 41.31% data will be laid (by table). Hence above 140 whole data 50.0 – 41.31 = 8.69% or 9% will be.

Example 2.

A Normal distribution data's mean is 16 and S.D is σ 4 then between 75% data, in which limit will be lies. Between the half of 75% data will be lies in either side of mean. By observation of table (A) we find out that between M+1.15 σ 37.5% data lies. Likewise other 37.5% will be lies mean and 1.15 σ . Hence total 75% will be laid between . Since σ =4 therefore 75% data will be laid between is therefore. They will be laid between. Hence this limit 20.6–11.4 have. Between 20.6 and 11.4, 75% data will be laid.

7.4 Uses of Normal Probability Distribution Diagram

In any distribution mean and on x-axis between any S.D. how much percent data lies, to know this a table is being used.

Observe table A in the end of book. Above first line of that x/σ 00.02……..09 and up to down in the column σ number has been written. In the first column towards up to down distance of SD(σ) has been given which upto one digit of decimal. If we have to observe the data between mean and 1.2 σ

.00 .02 ...09 then we read first $\frac{x}{\sigma}$ in column digit 1.2 and having read above's first line come down in the same column exact before 1.2 where both between data of mean– +1.2 σ .

In the same way between 1.38 σ will be to observe then first draw a line before 1.3 of first column and again in the first line of above having seen .08 come down to that column. Where before 1.3 line it intersect that is the number of percentage.

Second thing is to remember is that that numbers has been given in the table they are given in 10,000. Hence to see percentage then place decimal after two digit from left.

6. Extra example question

1. M = 14.4

S.D. (σ) = 2.5 then in whole distribution between marks 12-16 per thousands how many student will lies?

Distance between M and marks 12 σ – = $\frac{12-14.4}{2.5} = \frac{-2.4}{2.5} = \frac{2.4}{2.5}$ =–.96 σ

The percentage of number between mean and – 0.96 σ (from table) = 33.15 σ distance between mean and marks 16 = $\frac{16-14.4}{2.5}$ = + 0.64 σ **Ans**

Notes

Between mean and +.64 number of percentage of data (from table) =23.89, hence between both marks 12-16 total percentage = 33.15 + 23.89 = 57.04

In 1000 = 57.04 x 10 = 570.4 e.g. 570

2. M = 14.4 σ = 2.5

If frequency distribution is even than what is probability of students' in 1000 to get more than 18 marks?

$$\sigma - \text{Distance, between mean and marks 18} = \frac{18 - 14.4}{2.5} = \frac{+3.6}{2.5} = +1.44\ \sigma$$

Between mean and 1.44σ number of percentage students (from table)=42.51

Hence, number of percentage more than 18 = 50.0 – 42.51 = 7.49 (Since in normal distribution above mean each half part of below is 50%.

Hence, in 1000, taken more than 18 marks, number of answer getting = 74.9 or 75.

But taken upper limit of marks 18 as 18.5 the difference will be came in answer 25. If asked upper taken integer rather then upper limit and where is asked data below any marks, there marks lower limit is taken. As a –

3. In the above question to find out the number getting marks below 8.

$$\text{Distance between 8 and } \sigma = \frac{7.7 - 14.4}{2.5} = \frac{-6.9}{2.5} = -2.76\ \sigma$$

Percentage student between mean and –2.76 σ = 49.71(table)

Taken 7.5 rather than 8 σ distance = 2.76 and percentage data = 49.71

Therefore, in 1000 respectively = 497

Below 7.5 = 500-497 = 3

4. If normal distribution of marks

Mean (M) =14.4

S.D. σ = 2.5 then what is the probability to get more than 15 marks obtaining student. Mean and 15 marks e.g. $\left(\frac{15.5 - 14.4}{2.5}\right)$ -.44 σ between of distance.

Number of data =17.0% (By table (A).

Number of more than 15 marks getting = 15 – 17 = 33%

Hence probability of 1 among 3 that their will be more than 15.

5. How many data will be laid between mean and 1.54 σ

Observe in table (A)'s first column below $\frac{\sigma}{x}$ 1.5 and before that draw straight line.

Again in topmost line below column .04 move downwards and just before 1.5 line drawn intersect it at that row which number drawn, observe that. This number is 43.83. e.g. in out of 100, it is 43.83 if whole distribution assumes 10,000 then it will be 4383. If assume 1 then 4383 will be and assumes 10 then 4.383 will be.

6. Similarly between mean and –1.73 σ data will be 45.82%. The meaning of negative (-) sign in its that these data less value than mean e.g., will be laid left side.

7. Similarly if we have to calculate total data between 1.54 σ and – 1.73 then add data between mean and 1.54, between mean and 1.75.

Notes

Therefore total data =43.83 +45.82 = 89.65 percent will be. In normal distribution curve that position as follows.

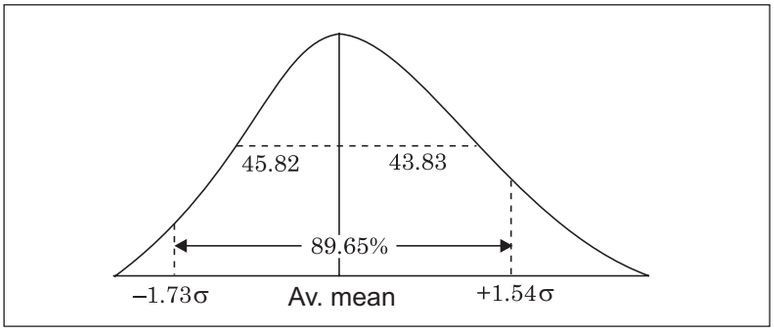

Fig. 7.5

8. Similarly if it is to calculate that more than +1.54σ marks obtaining number how much then to show that position by in this way.

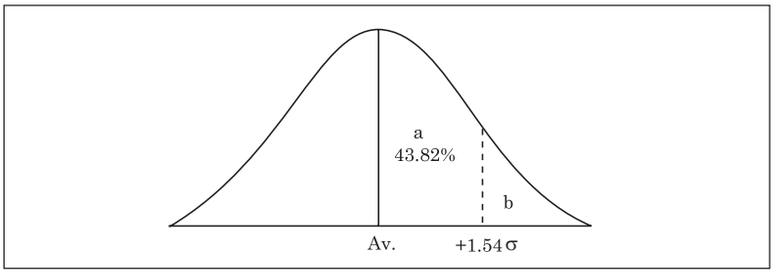

Fig. 7.6

In this figure whole area in the condition of normal distribution consider to be one unit which can be divided by mean into two equal parts. Therefore both a+b will be1/2 of whole. If 100 data lies in whole part then in half e.g. in a+b will be 50.0 + 1.54σ average's right hand that area in which 43.82% data has been covered, therefore in b's part will be 50.0 – 43.82 = 6.18%. If whole part considers to be 1000 then in part a+b 500 data will be, and by table in part a 4382 will be. Hence in part b 61.8 will be.

Similarly (taken as x) left side to x number of data can be seen also.

PROBABLE ERROR

Similarly in place of σ P.E. table is used also. By observation that scores can be find out or P.E. can be changed in scores can be find out because P.E. = 0.6745 σ's is. If it is to be calculated how much data will be between mean and -2.7 P.E. then we will converted P.E. into σ.

1P.E. = .6745 σ

–2.7P.E = 2.7 x 67.45σ = – 1.82σ

In table σ mean exists 1.82σ between 46.56. Hence, between mean and -2.7 P.E. 46.56% data had.

9. Now class 8's students' marks in history in mean to 40.0 and mean of marks of mathematics 62.0. Having after some months taken examination mean of history 52, and mathematics 72.0 respectively, then student in which subject progressed and much more progressed to others. S.D. of history (σ) 3.6 and S. D. of math σ =9.60 is.

Notes

	At this time	After Some month	Progress
History's mean	40.0	50.0	12.0
Mathematics mean	62.0	72.0	10.8
Progress in distance	(1) In history = 12/3.60 = 3.33σ		
'	(2) In mathematics 10.8/9.6 = 1.12σ		

Having changed σ in distance both subjects progress, they have been mutually comparable. Hence it is clear that in history has been more progressed. 3.36/1.12 = approx. thrice more have been progressed.

10. Four questions a,b,c,d in which a's 50%, b's 60%, c's 70% and d's80%, students do correctly. What is difficulties between a and b compare with the difficulties of c and d. If distribution is normal then difficulties S.D. (σ) will be following types.

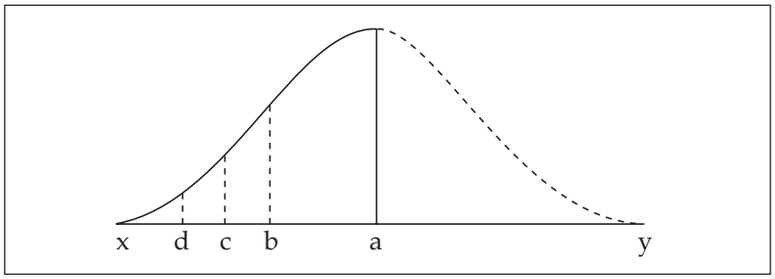

Fig. 7.7

(i) 50% students can do a correctly e.g. a on scale σ_ (on line x, y) lies on point at a because 50% will be laid on either side. Therefore it lies exactly on that point where mean falls. Therefore distance between mean and that point will be zero.

(ii) 60% does 'b' correctly and 40% can not does e.g. 60% student up and 40% lies below. y upto a becomes 50%. To keep 60% above below a point 10% taken more upto b point. Therefore b's position mean (a point) as distance below σ. as between 10% data observing table (A) it comes - .25σ.

Distance of a = 0.0σ b's distance = – 25σ.

Difference = + .25σ, hence a from b +.25 much more tough.

Similarly point C from mean 70.0 – 50 = 20% at a distance of 20% data comes under .52σ distance. Hence c from a .52σ less difficult.

Similarly 'd' from a (80-50) =30% e.g.-.84σ is less tough.

Difference of difficulties c and d = 0-.84σ -.52σ= -.32σ e.g. C from d -.32σ is more tough.

Hence 'a' difficulties with respect to 'b' that difficulties which is less than c with respect to d is 1.28 times $\left(\frac{-.32}{-.25}\right)$ = 1.28.

11. In a class of student has give a+, a, a-j, b+, b, b-j, c +, c, c- and d grade. If related qualification is normal distribution then out of 500 how much students will come in each grade?

In this question whole range of normal distribution has been divided into 10 categories. Whole range of normal distribution both side of mean is upto 3σ. Hence total range=6σ.

The range of 6σ has been divided into 10 category.

Notes

Hence each category's range area = $\frac{6}{10}\sigma$ = .60σ will be. In the following figure there position will be as follows.

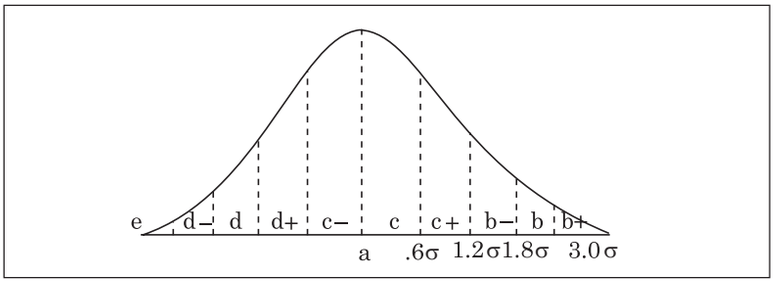

Fig. 7.8

Mean will be just laid in at point a from a to b is far away 6σ From b to b+ far away .6σ. Hence b's distance from a (.6+.6)=1.2σ. Similarly c's distance from a is 1.8σ a's 1.8+.6 =2.4σ and a+'s 2.4σ +.6σ =3.0σ.

1. Mean e.g. from point a and .6σ, number of data between them (from table A)=22.57%. This is number of students which comes in grade b. This number will be 500.

$$=22.57 \times \frac{500}{100} = 112.85 \text{ or } 113.$$

   Since on both side of mean destruction is same, therefore getting of b grade number will be 113 also.

2. Similarly distance mean and b+ 1.2σ under this scores from table 38.49% but between mean and b comes 22.57%. Hence only getting b grade's number=38.49-22.57=15.92 or 16. This is in 500=80 some will be number of getting grade c+.
3. Similarly distance between mean and a- is % 8σ and data between this = 46.41% is , hence number of getting a grade 46.41-38.49 = 7.92 or 8%, hence in 500, 8 × 5 = 40. Some number will be left to mean getting c grade.
4. Percentage number of getting a and c = Distance between mean and a or c (2.40σ) will be the percentage of that number. The scores between mean and c- = 49.18 - 46.41 = 2.77 this in 500 = 2.77 × 5 = 13.85 = 14
5. Similarly number of getting grade d and a+ = 49.865 (data comes upto 3.0 percentage) -49.865(data comes under 2.4σ)=.685% in 500 = 5 × .685 = 3.425 or 3. Similarly out of 500 student in each grade category number of student will be:

d	c-	c	c+	b-	b	b+	a-	a	a+
3	14	40	80	113	113	80	40	13	3

12. Suppose I.Q's distribution of 500 students are normal and their mean is 72, S.D(σ) is 10. In this 10%, 30%, 40, 15% grade is given to a,b,c,d respectively. Then find each two category dividing I.Q.

Mean does just two equal parts of normal distribution. There mean will be lies just in the middle of curve line. Upper 10% part under a and 30% under b. From mean to upto point c's whole part is 50%. From a to c it will be (30+10) = 40% comes under a and b. Hence between mean and point a 50-40=10% will be remain.

Notes

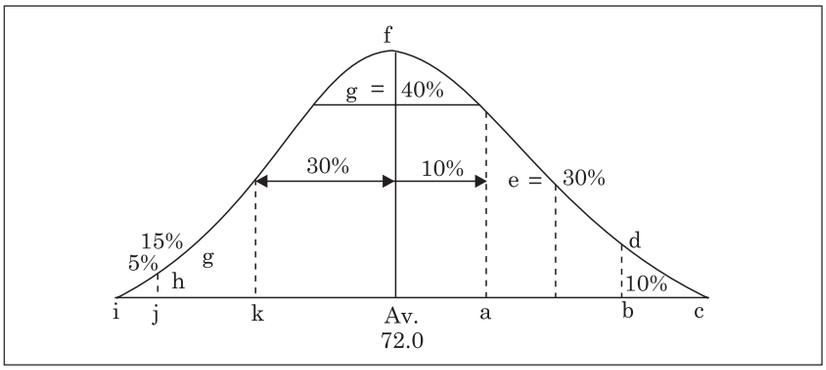

Fig. 7.9

Similarly from Q to R under S and P to Q 15% comes under d. Total parts under S and d = 5 + 15 =20/%. From mean to point R half part =50%, Hence between mean and P's part =50-20=30%. Both side of mean P to a till part is remain in which left side 30% and right had side 10% covers.

Each two category divider I. Q. b,a, p and Q points will be laid.

1. a, from mean lies such a distance, between that 10% data covers. From table 10% s σ distance = .25σ, but σ = 10 . Hence at that point which I. Q. will be .25 x 10 = 2.5 I.Q. up e.g. 72 + 2.5= 74.5 or 75.
2. Similarly b from mean as distance σ as between (10%+30%) = 40% data exists. From table this distance is 1.28σ. Since and mean= 72 therefore at this point situated I.Q.will be = 72 + 1.28 x 10 = 84.8 or 85.
3. Similarly at point P will be I. Q. left from mean from a distance.84σ because 30% scores lies between mean and .84σ. Therefore at this point will be situated I.Q. = 727.84 x 10 = 72-8.4 = 63.6 or 64
4. 'd' point from mean (30+15) =45% range area's distance. This distance in A 1.645σ is Therefore at this point will be situated I.Q. = 72 - (1.645 x 10) = 72 - 16.45 = 55.55 or 56.

Therefore at four division point situated I.Q. will be 56, 64, 75, 85 e.g. above 85 I.Q. in the a category, above 75 I.Q' in b category, above 64 in c category above 56 in d category and below 56 in S category.

Self Assessment

State whether the following statements are True or False:

4. In normal distribution half part of both side of curve is equal.
5. In positive skew mostly data lies right to mean.
6. There are four types of skews.
7. Kurtosis indicate towards the flatness and verticalness of the curve.

7.5 Measuring Divergence from Normality

Notes In normal distribution mean, medium and mode all the three lies only at one point and half part of both side of curve is same.

Notes

In that symmetry presents as already has mentioned under characteristics of normal distribution. If distribution curve differ from this position then it is called abnormal distribution. Abnormal distribution can be of two types.

(a) Skew

(b) Kurtosis

(a) Skew

In skew abnormalness both value mean and median, having not coincided but lies differently and centre of gravity is remove either right or left. In normal distribution mean and median, both has some value and abnormalness is zero. As they close to each other as skew quantity will be least and as much as normal will be lesast. Skew is of two types also :

1. Positive skew
2. Negative skew

1. Positive skew: In positive skew mostly data is lies left side of mean and in right side a few data and is scattered as shown in the below figure e.g. less scores than mean is much more in number and more value data from that are also, but they are scattered to a far distance. In second positive skew median lies left to mean.

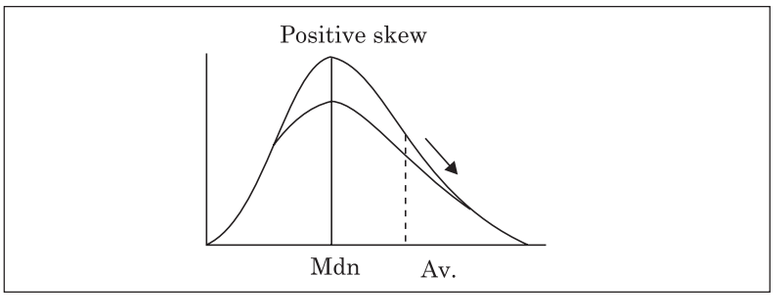

Fig. 7.10

2. Negative skew: In that condition mostly data are collected to the right of mean e.g. greater value data from mean have much in number and they are close to one another. Secondly median value is lies right to median e.g. median value is greater than mean. It is clear from below figure.

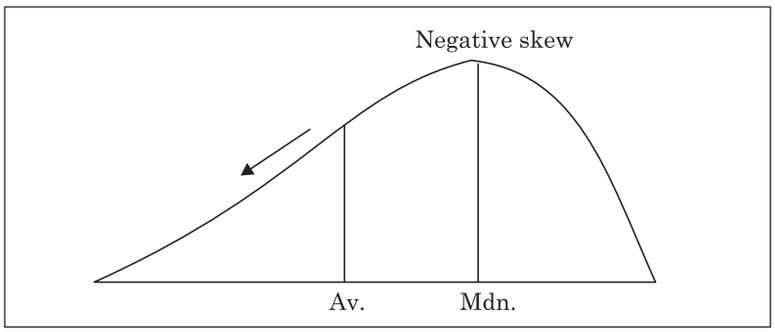

Fig. 7.11

Measures of skewness:-Skew abnormalities is calculated by following formula.

Notes

$$SK=\frac{3(\text{mean-median})}{\sigma} \quad ...(1)$$

Second formula by percentile is of calculation.

$$SK=\frac{P_{90}P_{10}}{2} - P_{50} \quad ...(2)$$

Having calculated by both formulae as much as skewness in quantity, as much as difference will be. Therefore they are not comparable.

(b) Kurtosis

Kurtosis indicates towards curve's flat and peaked position. Normal distribution curve should as much as flat and height, if more than that flatness and peakedress is found then it comes under kurtosis.

Normal distribution curve called mesokurtic also. That kurtosis's height is more than normal curve height that is called leptokurtic. That kurtosis's flatness more than normal curve flatness that is called platy kurtic. In the following figure these three position has been shown.

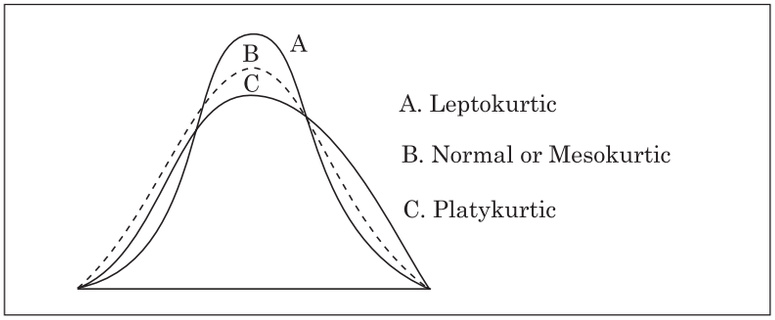

Fig. 7.12

Measures of kurtosis

Following is the formula of calculation of kurtosis: $Ku = \frac{Q}{P_{90} - P_{10}}$

The kurtosis of normal distribution curve is .7.3. If any curve is more than it then distribution is called leptokurtic.

Task What do you mean by kurtosis?

(c) Reason for abnormal frequency distribution

Why any frequency distribution having not normal have non normal, there are a number of reasons behind it. In psychological and educational test it is essential to know it. In the following conditions skew or kurtosis position is accrued:

Notes

(1) Sample:

(1) When we measures any features or characteristics then we take only some limit number of that properly holder or characteristics holder and population one or two sample. Having measured that properties that result applies to whole population. As is does so that whole population measure is impossible.

Did You Know? Non-normal distribution is of two types skew and kurtosis.

Our result's precision and exactness depends upon this samples own characteristics. When this selection is not right or any important fault comes in this then measures of distribution is not normal too. For example in our sample mostly boys will be brilliant or lees intelligent then theirs data of distribution of intelligence result will not be some. Therefore it is clear that sample is such a type that in which measures related property or characteristics' distribution is normal, e.g. neither many of that property holder nor very few e.g. of all types are.

(2) Not only objects or persons' characteristics quantity distribution but also their number effects on their measures on distribution. If sample is too low e.g. in that very few things or persons are measuring, only then distribution is normal.

(3) A number of external effects which is related to measures related characteristics, caused occurred oddness in distribution of measures as a much older age, social status, language skill etc. as such a factor which is related to intelligence. If sample is too high or objective, did not pay attention on selection time and any one primness comes in it then measures in distribution occurs skewness or kurtosis.

2. Wrong and un-appropriate test

Second reason is uses of that tests which is not appropriate for that sample and whose formulation is not properly.

(1) Un-appropriate test neither very tough nor simple. As a result measures must be either in left or right. Therefore distribution will not be normal. If a test of 12 aged is given to 8 or 9 year boys then measures distribution will not be normal in first condition that is positive skew or in second condition it will be negative skew.

Caution If test construction has not right then it will not reliable and exact. Therefore which measures will comes, may be their distribution is not normal.

(2) Similarly if administration of test has been not properly rather then distribution skewness is occurred.

7.6 Summary

- In statistics normal distribution has great importance. In several places use of normal distribution is of much worthful on the basis of normal distribution ours mean's utility, exactness and justification determines.

Notes

- In normal distribution by included percentage number and limitations we know total datas status. This is a great characteristic of normal distribution and it has great utility.
- In normal distribution value of mean, median and mode lies at midpoint of distribution and their value is equal also.
- This characteristics of normal distribution based upon theory of probability. Theory of probability has not only in statistics but also in all sciences great important.

7.7 Keywords

1. **Kurtosis:** Curve's flat and peaked position is called kurtosis.
2. **Table:** Table, name list, Tabulation.

7.8 Review Questions

1. What do you mean by normal distribution? Describe its characteristics.
2. What is called normal probability curve. Explain with examples.
3. How normal distribution table is used?
4. What do you mean by skewness of frequency distribution.

Answers: Self Assessment

1. Normal distribution	2. Mean	3. Middle	4. True
5. False	6. False	7. True.	

7.9 Further Readings

Books

1. Educational research methods—*Shareen and shashikala, vinod pustak mandir*
2. Educational techniques—*S.S. Mathur, Bhatt Brothers.*
3. Education Research Work Procedure—*L. Kaul, Vikash Publishing.*
4. Education Technique And Appraisal—*Dr Ram Pal Singh, Bhatt Brothers.*

Unit 8: Measurement of Dispersion QD, MD, SD

Notes

CONTENTS

Objectives
Introduction
8.1 Meaning and Definition of Dispersion
8.2 Kinds of Dispersion Measures
8.3 Quartile Deviation: QD
8.4 Mean Deviation: MD
8.5 Standard Deviation: SD
8.6 Uses of Standard Deviation
8.7 Summary
8.8 Keywords
8.9 Review Questions
8.10 Further Readings

Objectives

After studying this unit, students will be able to:

- Understand the meaning, definition and kinds of dispersion
- Understand quartile and mean deviation
- Understand standard deviation.

Introduction

Dispersion is that property of measurement by which it is calculated up to how much the value of terms is deviated from its mean values. The central tendency of whole series is found by mean values (mean, median and mode) but by this the complete and definite knowledge of different term values of series cannot be determined in the case when there is large difference in values of mean and other terms values. So in such situations there needs of measures which provide more information of term values i.e., how much deviation is of term values is mean values. Such kind of spread is called the measures of deviation or dispersion.

8.1 Meaning and Definition of Dispersion

Normally, the measures of scatter or spread of the separates scores around their central tendency is called the measure of variability or dispersion.
Some specific definition are as follows:-

1. **Crow and Crow,** "The extent to which cases tend to gather around the average or central tendency or the extent to which they disperse themselves is called their variability or deviation."

Notes

2. **Boring, Langfeld and Weld,** "Measures of variability tell us how widely the data scatter about their mean."

8.2 Kinds of Dispersion Measures

To get spreadness or dispersion in statistics mainly four measures are used.

1. Range
2. Quartile deviation
3. Mean deviation
4. Standard deviation

1. Range

Range is the simplest measure of dispersion. By this it can be explained without any calculation at a glance which of the two groups is homogeneous or which is heterogeneous.

R is the symbol of range and its formula is-

R = Maximum score – Minimum score.

Example Suppose in mathematics ten students marks are as follows - 55, 52, 50, 48, 90, 45, 60, 40, 43, 44, then its range will be = 90 – 40 = 50.

There are some limitations of this measures also, therefore range should be used very limited. The greatest limitations of this measures is that only ends numbers has given importance not other scores. Therefore this measures of deviations does not give any information of group's internal scatterness.

Secondly, when number of N is lees or much gaps in frequency distribution, then in this condition range do not do exact measure or deviations. Therefore in such cases we should not use this measures.

Caution When number of two samples is different then there should not use range to compare the deviations because there is a strong possibility of increasing range with the increase in number N also.

8.3 Quartile Deviation: QD

Quartile deviation is an improvement over range deviation. Under range measure only highest and lowest marks at the ends is considered and no importance is given to the middle scores of distribution, so there is no correct information obtained about the homoginity of groups. To remove this shortcomings in quartile deviation 50% scores of the middle has been given importance. 25% scores of above and below distribution is given no importance.

Q is its symbol and following is the formula:-

$$Q = \frac{Q_3 - Q_1}{2} = \frac{P_{75} - P_{25}}{2}$$

Notes

With the help of figure there is an attempt of more clarification of it.

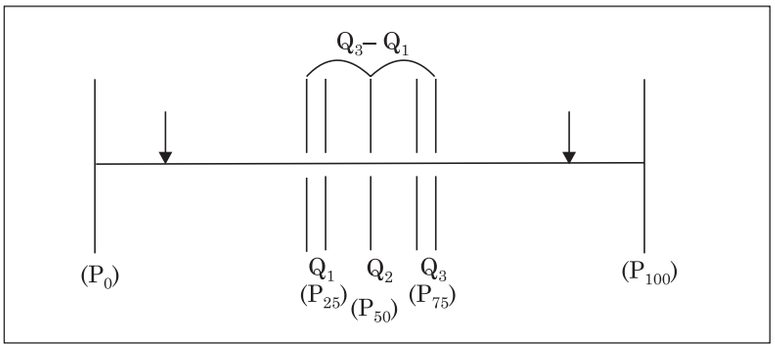

Fig. 8.1

Quartile range is also called semi-inter quartile Range. Some scholar has tried to define it as follows.

Notes Skinner, " Quartiles are the three points that divide a distribution into four equal parts."

Odell, "The quartile deviation or semi-inter quartile Range is one-half of the distance between the first and third quartiles."

Calculation of Quartile Deviation: QD

Example: Calculate quartile from the following data—

Table 8.1

S. No	*C.L.*	*f*	*Cf*	
1.	40-42	5	50	
2.	37-39	8	45	← Q_3
3.	34-36	9	37	
4.	31-32	10	28	
5.	28-30	7	18	← Q_1
6.	25-27	5	11	
7.	22-24	4	6	
8.	19-21	2	2	
		N=50		

Now, $$Q = \frac{Q_3 - Q_1}{2}$$

and $$Q_1 = L + \left(\frac{\frac{N}{4} - f_b}{f_a}\right) \times \text{C.I. and } Q_3$$

Notes

$$Q_3 = L + \left(\frac{\frac{3N}{4} - f_b}{f_a}\right) \times C.I.$$

Where Q_1 and Q_3 are the first and third quartile of group

L= Lower limit of the class interval in which Q_1 (P_{25}) and Q_3 (P_{75}) lie.

f_b= the sum of frequencies of all those classes which lie below the class of Q_1 and Q_3.

f_a= Frequency of Q_1 and Q_3 class

C.I.= Limit of class Interval.

From above example Q_1, i.e. the value P_{25} is as follows.

$$Q_1 = 27.5 + \frac{(12.5 - 11)}{7} \times 3$$

$$= 27.5 + \frac{(1.5)}{7} \times 3$$

$$= 27.5 + \frac{(4.5)}{7}$$

$$= 27.5 + .64$$

$$= 28.14 \text{ (Note } \frac{N}{4} = \frac{50}{4} = 12.5)$$

Similarly, $$Q_3 = 36.5 + \frac{(37.5 - 37)}{8} \times 3$$

$$= 36.5 + \frac{.5}{8} \times 3$$

$$= 36.5 + \frac{1.5}{8}$$

$$= 36.5 + .19$$

$$= 36.69$$

Hence, $$Q = \frac{Q_3 - Q_1}{2} = \frac{36.69 - 28.14}{2}$$

$$= \frac{8.55}{2}$$

$$= 4.28 \textbf{ Ans}$$

Self Assessment

Fill in the blanks:

1. The range of measure of separates scores from the central values to that group is called…….. .
2. To find out deviation mainly four measures are used in …… .
3. Quartile deviation is a improved form over …….. .

8.4 Mean Deviation: MD

Notes

Symbol = AD or MD

Garrett writes to clarify the meaning of mean deviation "The average deviation or mean deviation (AD or MD) is the mean of a deviations of all the separate scores in a series taken from their mean (occasionally from the median or mode). In averaging deviations to find the A.D., no account is taken of signs and all deviations whether plus or minus are treated as positive.

— Garrett 1958

Mean deviation is the mean of deviations of different scores from mean. To find mean deviation we find differences from mean to scores. We neither deduct mean to actual scores nor scores from mean rather than we find difference between mean and scores. This is the reason we do not take care of sign. In fact our aim is to find only deviation of means from scores, distance and range. To ignore plus (+) and minus (-) sign is the special feature in finding mean deviation therefore in statistics it is rarely used.

1. Ungrouped data
2. Grouped data

1. Mean deviation from ungrouped data: Following are the formula to calculate mean deviation of ungrouped data:

$$M.D = \frac{\Sigma |d|}{N}$$

Where, M.D. = Mean Deviation

D = Difference of any scores from mean.

|d| = Sum of deviation without sign

N = Total numbers of scores:

Illustration: Calculate mean deviation from following ungrouped data.

Table 8.2

S. No.	X marks obtained	d	=	*(X – M)*
1.	10	10–18	=	– 8
2.	15	15–18	=	– 3
3.	10	10–18	=	– 8
4.	20	20–18	=	2
5.	25	25–18	=	7
6.	15	15–18	=	– 3
7.	25	25–18	=	7
8.	20	20–18	=	2
9.	17	17–18	=	– 1
10.	23	23–18	=	5
		Σ | d | = 46		

$$M = \frac{\Sigma X}{N} = \frac{180}{10} = 18$$

Notes

Hence, M.D. $= \frac{\Sigma|d|}{N} = \frac{46}{10} = 4.6$ **Ans**

2. M.D. of grouped data: Following are the formula to use in finding out the mean deviation of grouped data.

$$\textbf{M.D.} = \frac{\Sigma|fd|}{N}$$

Illustration: Calculate mean deviation from following grouped data:

Table 8.3

S.No.	*C.I.*	*f*	*X*	*d*	*fd*
1.	55–59	1	57	27.4	27.4
2.	50–54	1	52	22.4	22.4
3.	45–49	3	47	17.4	52.2
4.	40–44	4	42	12.4	49.6
5.	35–39	6	37	7.4	44.4
6.	30–34	7	32	2.4	16.8
7.	25–29	12	27	–2.6	–31.2
8.	20–24	6	22	–7.6	–45.6
9.	15–19	8	17	–12.6	–100.8
10.	10–14	2	12	–17.6	–35.2
		N =50	M = 29.6		$\Sigma\|fd\| = 425.6$

$$\textbf{M.D.} = \frac{\Sigma|fd|}{N} = \frac{425.6}{50}$$

$= 8.51$ **Ans.**

Important steps:

1. Find the mean of group's scores.
2. Find mid point of each class.
3. By deducting mean from mid-point, calculate deviations (X-M).
4. Find fd by multiplying of each deviation and their related frequency.
5. Find product of total $\Sigma|fd|$ of obtained products.
6. Calculate mean deviation by putting all these values in formula.

8.5 Standard Deviation: SD

During calculating mean deviation we do care of sign of deviations and it is the special feature of mean deviation. Hence to make it error free, mean deviation or standard deviation is calculated. In other word, we can say that standard deviation is an improved form of mean deviation.

To ward off the problems of signs under mean deviation squares of deviation are taken under standard deviations. So that minus sign is converted into plus signs. After that by finding average of these squared deviations its square root have been taken. This is called the mean deviation or standard deviation.

Notes

According to **Reichmann "The standard deviation is also known as root mean square deviation. It is the square root of the mean value of all the deviation squared taken from the distributions mean**.'

In general words, mean deviation is defined in this way-

Very generally, the standard deviation is a measure of how the scores cluster or disperse around the mean and is conventionality represented by Greek letter sigma.

$$\text{Symbolically, } \sigma = \sqrt{\frac{\Sigma fd^2}{N}}$$

Calculation: For the calculation of standard deviation, following three forms of questions have been used.

Case I : When only ungrouped data has been given.

Case II : When frequencies with ungrouped data has also been given.

Case III : When classified data has been given.

Case I : When only ungrouped data has been given.

Such type of questions can be solved by two methods – Actual mean method and Assumed mean method. In these methods following formula are used—

$$\text{S.D.} = \sqrt{\frac{\Sigma d^2}{N}} \qquad \text{(Actual Mean Method)}$$

And

$$\text{S.D.} = \sqrt{\frac{fd^2}{N} - \left(\frac{\Sigma d^2}{N}\right)^2} \qquad \text{(Assumed Mean Method)}$$

Illustration: Calculate standard deviation of following data by actual mean method

Table 8.4

S.No.	*X*	*d = (X – M)*	d^2
1.	10	10-18 = – 8	64
2.	15	15–18 = –3	9
3.	10	10–18 = –8	64
4.	20	20–18 = 2	4
5.	25	25–18 = 7	49
6.	15	15–18 = –3	9
7.	25	25–18 = 7	49
8.	20	20–18 = 2	4
9.	17	17–18 = –1	1
10.	23	23–18 = 5	25
N = 10	ΣX = 180		Σd^2 = 278

$$M = \frac{\Sigma X}{N} = \frac{180}{10} = 18$$

$$\sigma = \frac{\Sigma X}{N} = \frac{180}{10} = 18$$

$$= \sqrt{\frac{\Sigma d^2}{N}} = \sqrt{\frac{278}{10}}$$

$$= \sqrt{27.8} = 5.27 \textbf{ Ans.}$$

Table 85			
S.No.	*X*	*d = (X – M)*	d^2
1.	10	10–17 = –7	49
2.	15	15–17 = –2	4
3.	10	10–17 = –7	49
4.	20	20–17 = 3	9
5.	25	25–17 = 8	64
6.	15	15–17 = –2	4
7.	25	25–17 = 8	64
8.	20	20–17 = 3	9
9.	17	17–17 = 0	0
10.	23	23–17 = 6	36
N = 10	ΣX = 10		$\Sigma d^2 = 278$

Important stages

1. Systematized distributed scores.
2. Calculate actual mean from distributed scores.
3. Calculate deviation by using formula d=X–M.
4. Find the square of deviations d^2.
5. Find Σd^2 by adding deviations squares.
6. Calculate S.D. by putting all these values in formula.

Case I: When frequency with their ungrouped data is given.

Example: Calculate standard deviation of following data by assumed mean method

Table : 8.6					
S. No.	*X*	*f*	*d = (X – M)*	*fd*	fd^2
1.	65	3	65–75 = –10	–30	300
2.	68	4	68–75 = –7	–28	196
3.	70	4	70–75 = –5	–20	100
4.	72	5	72–75 = –3	–15	45
5.	74	6	74–75 = –1	–6	6
6.	75	7	75–75 = 0	0	0
7.	78	6	78–75 = 3	18	54
8.	80	5	80–75 = 5	25	125
9.	82	4	82–75 = 7	28	196
10.	84	6	84–75 = 9	54	486
N = 50				$\Sigma fd = 26$	$\Sigma fd^2 = 1508$

Notes

Now, S.D. or $\sigma = \sqrt{\frac{\Sigma d^2}{N} - \left(\frac{\Sigma fd}{N}\right)}$

$= \sqrt{\frac{1508}{50} - \left(\frac{26}{50}\right)}$

$= \sqrt{30.16 - 0.52}$

$= \sqrt{30.16 - 0.27}$

$= \sqrt{29.89}$

= 5.47 Ans.

Note

1. If score is in the range of 10-20, then actual mean method is convincible.
2. If actual mean comes in decimal then use assumed mean method.
3. The selection of assumed mean data should be done carefully.

Case II- When classified data has been given.

Standard deviation of related question of grouped data is found by two methods- Long method and short-cut method- Following are the used formula in these methods:

$$\sigma = \sqrt{\frac{\Sigma fd^2}{N}} \text{ (Long Method)}$$

$$\sigma = \sqrt{\frac{\Sigma fd^2}{N} - \left(\frac{\Sigma fd}{N}\right)^2} \times \text{C.I. (Short Method)}$$

Example: Calculate standard deviation of following data by long method.

Table 8.7

S. No.	*C.I.*	*f*	*X*	*fX*	*d(X – M)*	*fd*	*fd²*
1.	40–44	3	42	126	14.3	42.9	613.47
2.	35–39	4	37	148	9.3	37.2	345.96
3.	30–34	4	32	128	4.3	17.2	73.96
4.	25–29	7	27	189	0.7	4.9	3.43
5.	20–24	5	22	110	5.7	28.5	162.45
6.	15–19	3	17	51	10.7	32.1	343.47
7.	10–15	2	12	24	15.7	31.4	492.98
		N = 28	ΣfX = 776			Σfd² = 2035.72	

$$\mathbf{M} = \frac{\Sigma fX}{N} = \frac{776}{28} = 27.7$$

$$\mathbf{S,D,} = \sqrt{\frac{\Sigma fd^2}{N}}$$

$$= \sqrt{\frac{2035.72}{28}}$$

$$= \sqrt{72.71}$$

$$= 8.52 \textbf{ Ans.}$$

Did You Know? Standard deviation is also called the square root of average deviation.

Important steps:

1. Find the mid-point (x) of each class interval related formula is—

$$X = \frac{\text{Lower limit} + \text{Upper limit}}{2}$$

2. Find *fx* by multiplying mid-point of all class interval with their respective frequencies.
3. Find sum of all *fx*.
4. Calculate mean by using formula (M = Σ*fx/N*.
5. Calculate deviation (d) by formula d = X – M.
6. Calculate fd by multiplying each class frequencies to their deviation.
7. Find fd^2 multiplying fd and d.
8. Find summation of all fd^2.
9. Divide Σfd^2 by N.
10. Obtain square root of obtained quotient. This will be the required standard deviation.

Example: Calculate standard deviations of following data by short–cut method.

Table 8 .8

S. No.	*C.I.*	*f*	*d*	*fd*		fd^2
1.	90–99	1	+4	4		16
2.	80–89	4	+3	12	+ 40	36
3.	70–79	7	+2	14		28
4.	60–69	10	+1	10		10
5.	50–59	9	0	0		0
6.	40–49	9	–1	–9		9
7.	30–39	4	–2	–8	– 38	16
8.	20–29	3	–3	–9		27
9.	10–19	3	–4	–12		48
		N = 50		Σfd = 2		$\Sigma fd^2 = 190$

Now, S.D or σ $= \sqrt{\frac{\Sigma fd^2}{N} - \left(\frac{\Sigma fd}{N}\right)^2} \times \text{C.I.}$

Notes

Putting obtained values in the formula we get

$$\sigma = \sqrt{\frac{190}{50} - \left(\frac{2}{50}\right)^2} \times 10 = \sqrt{3.8 - (.04)^2} \times 10$$

$$= \sqrt{3.8 - .016} \times 10$$

$$= \sqrt{3.7984} \times 10$$

$$= 1.95 \times 10$$

$$= 19.50 \textbf{ Ans.}$$

Important Steps:

1. Assume appropriate class interval or the class-interval of the most frequency as assumed mean.
2. Assume the deviation of selected class interval as zero.
3. Find out deviation of all classes. For this above class (greater value) of assumed C. I. is given positive integral values as +1, +2, +3 and lower class intervals is given negative integral values as -1, -2, -3…….
4. Find out fd multiplying by each class interval's frequency and deviations.
5. Find out sum of fd keeping in mind their sign. This is called Σfd.
6. Calculate fd^2 multiplying each class intervals fd to d.
7. Find out summation of fd^2. This is called Σfd^2.
8. Calculate S.D. by putting obtained value in formula.

Self Assessment

Multiple Choice Questions:

4. Range is the _____ measure of deviation.

   a. Simplest b. Complicated c. Expensive d. long term
5. Symbol of quartile is _____.

   a. R b. Q c. N d. M
6. _____ Deviation is mean of deviations of different scores from means.

   a. Range b. Mean c. Median d. Standard
7. Standard deviation is also called _____ of average deviation.

   a. Equal b. Half c. Square root d. Full

8.6 Uses of Standard Deviation

Range

1. When standard deviation is to find out quickly.
2. When to get information of highest and lowest scores.

Notes

Quartile Deviation (Q.D.)

1. When the value of central tendency is shown by mean.
2. When distribution is incomplete.
3. When the value of some scores is either very high or very low.

Mean Deviation (M.D.)

1. When distribution is almost normal.
2. When deviation is dispersed excessively.
3. When normal reliability is needed.

Task When do the deviation values used?

Standard Deviation (S.D.)

1. When more reliability is needed.
2. When to find correlation coefficient etc.
3. When distribution is normal.
4. When the value of central tendency is shown by the middle value.

8.7 Summary

- In normal scenes the measures of scattered or spread separate scores around their central tendency is called the measures of variability or dispersion.
- According to boring, Langfield and weld " Measures of variability tell us how widely the data is scattered about their mean."
- To find out deviation or expansion in statistics mainly four measures uses—
  1. Expansion (Range) 2. Quartile Deviation 3. Mean Deviation 4. Standard Deviation

8.8 Keywords

1. **Quartile:** That point which divides the scores of distributions into four equal part.
2. **Deviation:** Difference of central value of that group from separate scores of any group.

8.9 Review Questions

1. With the help of following scores calculate range-15, 7, 19, 18, 21, 27, 18, 12, 29, 13 Ans 22
2. Find out quartile deviation from following table.

Notes

Table		
C.I.	f	
40–44	3	
35–39	3	
30–34	5	
25–29	7	
20–24	5	
15–19	3	
10–14	2	**Ans**. Q = 6.0

3. Find out quartile deviation from following data.

Table		
C.I.	f	
55–59	1	
50–54	1	
45–49	3	
40–44	4	
35–39	6	
30–34	7	
25–29	12	
20–24	6	
15–19	8	
10–14	2	**Ans**. Q = 7.5

4. Find out quartile deviation from following table.

Table		
C.I.	f	
80–89	5	
70–79	2	
60–69	3	
50–59	2	
40–49	5	
30–39	4	
20–29	3	
10–19	2	
0–9	4	**Ans.** Q = 21.67

5. With the help of following data find mean deviation.

   4, 6, 8, 10, 12, 16, 18, 20 **Ans.** M.D. = 4.44

6. With the help of following scores find mean deviation.

   6, 8, 10, 12, 14 **Ans.** M.D. = 24

Notes

7. With the help of following scores find standard deviation

   (a) 12, 18, 17, 13, 14, 16 **Ans.** $\sigma = 2.10$

   (b) 24, 23, 12, 14, 17, 18, 16, 20 **Ans.** $\sigma = 3.90$

   (c) 4, 6, 8, 10, 12, 14, 16, 18, 20 **Ans.** $\sigma = 5.16$

   (d) 16, 13, 18, 17, 20, 19, 21, 17, 22, 12
   16, 16, 18, 15, 18, 19, 16, 13, 21, 17 **Ans.** $\sigma = 2.7$

8. With the help of discrete series find standard deviation.

   X: 65, 68, 60, 72, 74, 71, 73, 80, 82,83

   f : 3, 4, 4, 5, 6, 9, 6, 5, 2, 6 **Ans.**$\sigma = 6.36$

9. Calculate mean deviation with the help of following data.

Table		
C.I.	f	
40–45	3	
35–40	4	
30–35	4	
25–30	7	
20–25	5	
15–20	3	
10–15	2	**Ans.** M.D. = 6.9

10. Find out mean deviation from following data.

Table		
C.I.	f	
29–31	2	
26–28	3	
23–25	4	
20–22	3	
17–19	2	
14–16	1	**Ans.** M.D. = 3.52

11. Define mean deviation, at the same time with the help of following data calculate S.D.

Table		
C.I.	f	
100–104	3	
95–99	4	
90–94	4	
85–89	11	
80–84	3	
75–79	4	
70–74	2	**Ans.** $\sigma = 8.15$

Notes

12. Find out standard deviation from following data.

Table		
C.I.	f	
40–42	5	
37–39	8	
34–36	9	
31–33	10	
28–30	7	
25–27	5	
22–24	4	
19–21	2	**Ans.** $\sigma = 5.70$

13. Find out standard deviations from following data.

Table		
C.I.	f	
70–74	3	
65–69	6	
60–64	7	
55–59	8	
50–54	15	
45–49	10	
40–44	7	
35–39	3	
30–34	1	**Ans.** $\sigma = 9.45$

14. From following data calculate quartile mean and standard deviation.

Table		
C.I.	f	
60–65	2	
54–59	2	
48–53	6	
42–47	3	
36–41	8	
30–35	8	
24–29	4	
18–23	6	
12–17	10	
6–11	6	
0–5	5	**Ans.** Q = 12.88 M.D. = 14.35 S.D. = 16.62

Notes

15. From following data find out quartile mean and standard deviations.

Table		
C.I.	f	
77–83	3	
70–76	13	
63–69	16	
56–62	13	
49–55	6	
42–48	22	
35–41	18	
28–34	6	
21–27	1	
14–20	2	**Ans.** Q = 12.42 M.D. = 13.06 S.D. = 14.94

Answers: Self Assessment

1. Deviation
2. Statistics
3. Expansion or Range
4. (a) Simple
5. (b) Q
6. (b) Mean
7. (c) Square root

8.10 Further Readings

Books

1. Educational Technique—*S.S. mathur, Bhatt Brohters.*
2. Educational Techniques, Management and Evaluation—*J.C. Agrawal, Bhatt Brothers.*
3. Education Techniques—*R.A. Sharma Bhatt Brothers.*
4. Educational Research Methods—*Shareen and Shashikala, Vinod pustak mandir.*

Notes

Unit 9: Correlation: Rank Difference Method, Product Moment Method

CONTENTS

Objectives
Introduction
9.1 Definition of Correlation
9.2 Kinds of Correlation
9.3 Coefficient of Correlation
9.4 Spearman's Rank Difference Method
9.5 Product Moment Method
9.6 Summary
9.7 Keywords
9.8 Review Questions
9.9 Further Readings

Objectives

After studying this unit, students will be able to:

- Understand the definition and kinds of correlation
- Understand coefficient of correlation
- Understand rank difference and product moment method.

Introduction

Generally, we say that those students who opt science are interested in engineering and medical field, commerce should be opted by those students who are interested in business, Hindi, English and Sanskrit should be opted by those who are interested in literature, music and art should be opted by those who have aesthetic value or rather than we say that persons of short-height are more talented than the tall persons, persons with oval forehead get more honours in society, ball headed man is wealthy etc. Behind these feelings have a long experience or special related facts. Along with our assertions or prediction of these kinds are not based on the observations of short duration only but plenty of logics can be gathered to fulfill it. For example, if we want to assure in the field of psychology how much an individuals achievements in one field have related to or affected to the other field, then in the language of Pschology, we will symbolyse this relationship in terms of variable by establishing relations among different variables. We can easily predict that the individual it can be selected for the particular profession or not .Along with the behaviour of the individual can be controlled according. The mutual relation among variables are called correlation. With the help of following examples the meaning of correlation can be done more clearly.

Notes

1. Effects of socio-economic conditions of the parents on child's performance.
2. Relationship between I.Q and Achievement.
3. Effects of Motivation on Learning.
4. Effects of Home Environments on child's social activities.
5. Role of Material Aids in increasing teaching-effectiveness.
6. Effects of Treatment-environment on child's morale.
7. Relationship between student morale and academic achievement.

9.1 Definition of Correlation

Some main definitions are as follows.

When two variable are related with each other in such a fashion that the increase in one variable causes increase or decrease in another variables or decrease in one variable causes decrease or increase in another variable, *i.e. vice-versa,* then the two variables are said to be correlated.

Thus, correlation is the relationship between two variables or two different sets of data.

Correlation deals with the relationship between two or more different groups of data in order to ascertain whether any relation exists between them or not and to obtain a numerical expression of the degree of such a relationship. ***—Bayliss***

Notes Correlation is a statistical measures of the degree of association between two variables.

"The coefficient of correlation is almost as important to the Psychological tester as is the balance to the chemist."

—W.L Sumner, 'Statistics in school', Oxford Basic Blackwell, 1950

9.2 Kinds of Correlation

Correlation have mainly three types.

1. Positive Correlation
2. Negative Correlation
3. Zero Correlation

Positive Correlation

Positive Correlation occurs in that condition when inresent in one variables causes increase in other e.g., the students which gets maximum marks in mathematics the same students gets maximum marks in English or Vice-Versa, decrease of one variable causes decrease in other variable e.g, the students which gets minimum marks in art, the same student gets minimum marks in Hindi also, e.g., I.Q and academic achievement's relationship.

Notes

Example

High I.Q. → High Percentage of marks

Low I.Q. → Low Percentage of marks

According to above example the student whose I.Q. level is superior, his percentage of marks should be high also. In other words, a student who gets maximum marks in examination, his I.Q. level should be superior also. Right-now the student whose I.Q. level is low, his percentage of marks should be minimal also. In other words, a students who get-less sores in test, their I.Q. should also be less. This condition is of positive correlation.

Unsystematic graph shows correlation +1.00

One proportional line added it to a graph.

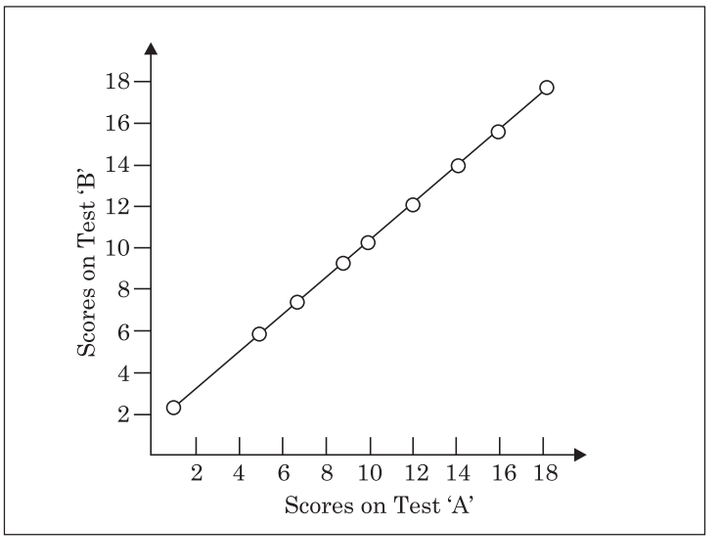

Fig. 9.1

Negative Correlation

Negative Correlation occurs at that condition when increase in one variable causes decrease in other variable e.g., the student who stands first in an examination, the same student gets last position in another examination and similarly all students' absurdities just reverse or its opposite, we say that when decrease in one variable causes increase in another variable, e.g., the student gets final position in one test the same student stands first in another test. In these condition, negative or negative perfect correlation occurs such type of perfect conditions are generally found less.

Example

More extrovert students → Low pass percentage

Less extrovert students → High pass percentage

According to above example, those student who is more extrovert they pass less or we can say that the cause of failure of students are their extrovert personality. Just then that student who is less extrovert e.g., they are introvert. They succeed in examination more or we can say that their cause of success is their introvert personality.

Notes

Unsystematic graph shows correlation -1.00.

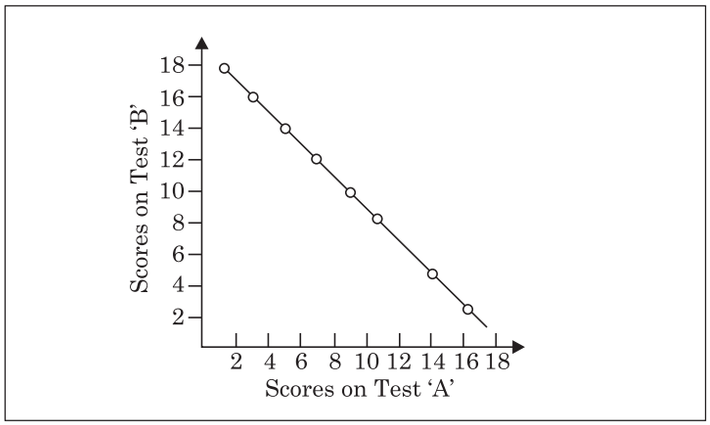

Fig. 9.2

Zero Correlation

Zero Correlation occurs at that condition when increases or decrease of one variables has no effect on other variable e.g. on the basis of student's one subject obtained marks we can not guess about his other subjects marks.

Example

Children from good social status → Pass percentage 80

Children from poor social status → Pass percentage 80

According to above example there is no correlation to children's social status to their academic achievement e.g, fail or pass. Children comes from either aristocrate or poor family theirs percentage of passing will be the same. In this way a good musician may be a good artist or maybe not. Unrelated graph shows correlation 0.

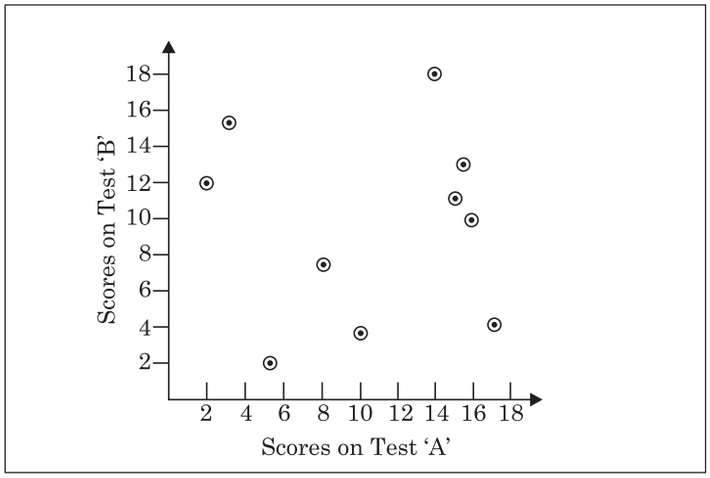

Fig. 9.3

Self Assessment

Notes

Fill in the blanks:

1. Correlation shows relationship ……..two variables or groups
2. Correlations are mainly …… in types.
3. Those students who are more……. are less pass in examination.

9.3 Coefficient of Correlation

In educational field mostly two series of marks are needed to determine the amount of relationship. It can be expressed as coefficient of correlation.

On the basis of above discussion conclusion is drawn from it that in positive correlation increase in one variables causes increase in another also and decrease in one cause decrease in other also. In negative correlation decrease in one variables relates with another variables with more accuracy. Relationship will be considered as much close as the quantity of coefficient will be more. In Zero correlation on the basis of one subject's achievements cannot guess about other subject's achievement. All these three relationship numerically express as +1.00, -1.00 and .00 respectively. Generally these types of perfect conditions does not exist, hence correlation of this quantity are considered between +1.00 and-1.00.

In fact, correlation shows us what the nature of relationship of two variables is from this point of view correlation is that number which indicate us that how much two variables is related to each other e.g., how much the changes in one variable cause change in second variable. This correlation coefficient lies between +1.00 and -1.00 (±1) and symbolically represented by (ρ) or rho.

"The coefficient of correlation (r) is a convenient quantitative index of relationship, i.e. what is the degree of relationship? or, to what extent two variables are related with each other? The coefficient of correlation is almost as important to the Psychological tester as is the balance to chemist." ***W.L. Sumner, 'Stat, in Schools'*** ***—Lindquist and Blummers.***

Interpretation of Correlation Coefficient or Results

To explain correlation coefficient or conclusions we take notice of two things ±0 sign and magnitude of results.

If correlation coefficients are positive then its meaning will be that, in a given distribution increase in one variables cause increase in another variables also or decrease in one variable cause decrease in another also. For example, if children's I.Q. and their subjective achievement find correlation then in this condition correlation coefficient will be positive and having analyzed of this mathematical fact it will be said that as increase in children I.Q. as will be increase in their subjective achievement.

Similarly, just opposite of it, if correlation are negative then its meaning will be that in a given distribution if one variables will decrease it causes increase another. For example, correlation is found between men's age and their memory then in this condition correlation coefficient will be negative and explaining to this mathematical fact it will be said that as men's age increases their memory will decrease.

For the explanation of correlation coefficient or drawn conclusion **Guilford's Table used**, which is following-

Quantity of Correlation Coefficient	Explanation
.00 to ±.20	Negligible
±.21 to ±.40	Low

Notes

±.41 to ±.70	Moderate
±.71 to ±.90	High
+.91 to ±.99	Very High
±1.00	Perfect

Methods of Computing Coefficient of Correlation

There are three main methods to computation of coefficient of correlation

1. Spearman's Rank Difference Method
2. Pearson's Product-Moment Method
3. Scatter-Diagram Method

9.4 Spearman's Rank Difference Method

We call rank difference method as series order method also. This method was postulated by Charles spearman and this method is called Spearman's correlation method after his name. Correlation coefficient obtained by this method are called Spearman correlation coefficient or called Spearman correlation coefficient or spearman's ranking correlation coefficient also. By this method obtained correlation coefficient are express by Latin language letter ρ (rho).

Where there are relatively few measures in the series, the Rank Difference method of correlation is employed generally. As the name indicates the size of the correlation coefficient (ρ) depends upon the sum of the differences in ranks between each pair of measures (a pair means one individual's measures in the two variables).

In this method both the scores of this variables are ranked separately on the basis of their size. After separating the scores of both variables calculate the difference of rank assigned on two variables and taking square of them, which is known as DV. After calculating the value ΣD^2 by the help of following formula calculate the value of correlation coefficient.

$$\rho \text{ or rho} = 1 - \frac{6\Sigma D^2}{N(N^2 - 1)}$$

where, ρ = Correlation coefficient

ΣD^2 = Sum of squares of rank difference

N = Total no. of students

Illustration 1: Marks obtained by 12 students in two subjects mathematics and economics 12 students marks are given below. By Rank difference method calculate its coefficients coefficient.

Economics:	60	34	40	50	45	41	22	43	42	66	64	46
Math:	75	32	34	40	45	33	12	30	36	72	41	57

Solution: Find out Correlation Coefficient

Table 9.1

Student	Economics	Math	R1	R^2	D	D^2
A	60	75	3	1	2	4
B	34	32	11	10	1	1

Notes

C	40	34	10	8	2	4
D	50	40	4	6	2	4
E	45	45	6	4	2	4
F	41	63	9	9	0	0
G	22	12	12	12	0	0
H	43	30	7	11	4	16
I	42	36	8	7	1	1
J	66	72	1	2	1	1
K	64	41	2	5	3	9
L	46	57	5	3	2	4
N = 12						$\Sigma d^2 = 48$

$$\rho = 1 - \frac{6\Sigma D^2}{N(N^2-1)}$$

$$= 1 - \frac{6 \times 48}{12\,(12^2-1)}$$

$$= 1 - \frac{288}{12\,(144-1)}$$

$$= 1 - \frac{288}{12 \times 143}$$

$$= 1 - \frac{288}{1716}$$

$$= \frac{1716-288}{1716}$$

$$= \frac{1428}{1716}$$

= + 0.83 **Ans**.

Explanation: The results of marks obtained by 12 students in both two subjects economics and mathematics show the high positive correlation between them.

Example: By following data find correlation coefficient between father and son height's

Height of father in inch	20	25	32	28	30	27	35	60	80
Height of son in inch	21	27	30	24	28	22	37	64	82

Solution: Calculate correlation coefficient between father and son height's.

Table 9.2

Father's height	**Son's height**	**R^1**	**R^2**	**D**	**D^2**
20	21	9	9	0	0
25	27	8	6	2	4
32	30	4	4	0	0

28	24	6	7	1	1
30	28	5	5	0	0
27	22	7	8	1	1
35	37	3	3	0	0
60	64	2	2	0	0
80	82	1	1	0	0
N = 9					$\Sigma D^2 = 6$

$$\rho = 1 - \frac{6\Sigma D^2}{N(N^2-1)}$$

$$= 1 - \frac{6 \times 6}{9\,(9^2-1)}$$

$$= 1 - \frac{36}{9(81-1)}$$

$$= 1 - \frac{36}{9 \times 80}$$

$$= 1 - \frac{36}{720}$$

$$= 1 - \frac{1}{20}$$

$$= 20 - \frac{1}{20}$$

$$= \frac{19}{20}$$

= +0.95 **Ans.**

Did You Know? Rank Difference Method is developed by Charles Spearman.

Illustration: 3 By following data calculate rank correlation coefficient

Student	A	B	C	D	E	F	G	H	I	J	K	L
Science	50	54	56	59	60	62	61	65	67	71	71	74
Math	22	25	34	28	26	30	32	30	28	34	36	40

Notes

Solution:

Table 9.3						
Student	Science marks	Math Marks	R_1	R_2	D	D^2
A	50	22	12	12	0	0
B	54	25	11	11	0	0
C	56	34	10	3.5	6.5	42.25
D	59	28	9	8.5	.5	.25
E	60	26	8	10	2	4
F	62	30	6	6.5	.5	.25
G	61	32	7	5	2	4
H	65	30	5	6.5	1.5	2.25
I	67	28	4	8.5	4.5	20.25
J	71	34	2.5	3.5	1	1
K	71	36	2.5	2	.5	.25
L	74	40	1	1	0	0
N = 12						$\Sigma D^2 = 74.50$

$$= 1 - \frac{6\Sigma D^2}{N(N^2 - 1)}$$

$$= 1 - \frac{6 \times 74.50}{12(12^2 - 1)}$$

$$= 1 - \frac{447}{12(144 - 1)}$$

$$= 1 - \frac{447}{12 \times 143}$$

$$= 1 - \frac{447}{1716}$$

$$= \frac{1716 - 447}{1716}$$

$$= \frac{1269}{1716}$$

= +.739 **Ans**.

Explanation: The marks obtained in both subjects mathematics and science by 12 students show correlation is highly positive between them.

Interpretation: The value of rank correlation between science and mathematics is +.739. it is clear that correlation coefficient is positive and of high degree. Hence on the basis of this correlation coefficient it is said that the student obtained high marks in science they show tendency of obtaining high marks in mathematics also and the student obtained less marks in science, they show tendency of getting less marks in mathematics also. Along with most of the students have done so.

Notes

Example 4: By rank difference method calculate correlation coefficient of following two sets of data

Set A	80	80	80	75	75	62	74	66	71	59	67	67
Set B	75	75	69	73	62	58	70	61	61	61	63	72

Table 28.4

Set A	Set B	R_1	R_2	D	D^2
80	75	2	1.5	.5	.25
80	75	2	1.5	.5	.25
80	69	2	6	4	16
75	73	4.5	3	1.5	2.25
75	62	4.5	8	3.5	12.25
62	58	11	12	1	1
74	70	6	5	1	1
66	61	10	10	0	0
71	61	7	10	3	9
59	61	12	10	2	4
67	63	8.5	7	1.5	2.25
67	72	8.5	4	4.5	20.25
N = 12					ΣD2 = 68.50

$$\rho = 1 - \frac{6\Sigma D^2}{N(N^2 - 1)}$$

$$= 1 - \frac{6x68.50}{12(12^2 - 1)}$$

$$= 1 - \frac{411}{12(144 - 1)}$$

$$= 1 - \frac{411}{12x143}$$

$$= 1 - \frac{411}{1716}$$

$$= \frac{1716 - 411}{1716}$$

$$= \frac{1305}{1716}$$

= 0.76 **Ans.**

Explanation: This result show high positive correlation between two sets of data.

Limitations of Rank-Method: Rank difference method is not more reliable. In this method care is taken only on ranks, their own value has no significance. For example A, B, C's marks in mathematics 30, 40, 50 respectively and in English 70, 80, 90, then in first and second condition rank will be only

3,2,1. Now if English's marks is done 100, 110, 120 then their rank will be same. Now in this method in every condition answer will be some. Obviously, this method is not more, correct.

Utility of Correlation Coefficient: The main utility of correlation coefficient are as follows-

1. Helpful in Prediction
2. To determine reliability of test
3. To know the estimation of validity test
4. Useful in diagnosis
5. To get the information of heredity and effects of environment.
6. Elimination of false beliefs
7. Use in factor-analysis
8. In establishing the cause and effect relationship

Besides aforesaid points utility of correlation coefficient can be cleared as follows-

1. If a student has scored highest in home test then on the basis of that we predict that the student must stand first in board examination.
2. To know reliability either we apply Test-Retest Method or Split Half Method or any others we must find out the value of correlation coefficient.
3. In validity determination we correlate out of test to any criterion and on the basis of correlation coefficient validity is known.
4. If we had proved that in science and mathematics exists positive correlation by given test on a big group and if we do not see directly in class then in this condition, correlation solves our problems
5. **Francis Galton, Karl Pearson** and **Thorndice** have done many experiments to observe the effects of heredity and environment on peoples.

Example Having experimented on 5000 boys Karl Pearson proved that there is no relation of face and physical appearance to their mental ability.

7. To understand the form of intelligence Spearman was the first to take help of factor-analysis-technique.
8. What is the impact of environment on children's achievement and motivation, this type of relationship is seen also.

Task Express your views on the application of correlation coefficient.

Self Assessment

State whether the following statements are True or False.

1. Correlation is a measurement at the approximate level between two variables.
2. Positive correlation occurs when increase of one variable causes decrease in another variable

Notes

3. Rank difference method is also called series rank method.
4. There is no relation between children motivation and their academic achievement.

9.5 Product Moment Method

This method is called Pearson method.

Two variables x and y which express some features (time ,place etc.) of the values of infinite pairs like (x_1, y_1), $(x_2\ y_2)$.... (x_n, y_n) respectively. In present studies use of Karl Pearson method is used between two variables in linear correlation.

Karl Pearson's correlation coefficient is defined by *r* (x,y) following—

$$r(x,y) = \frac{\sum_{i=1}^{n}(x_i - \bar{x})(\bar{y}_i - \bar{y})}{\sqrt{\sum_{i=1}^{n}(x_i - \bar{x})^2}\sqrt{\sum_{i=1}^{n}(\bar{y}_i - \bar{y})^2}}$$

$$\text{Or generally } r = \frac{\Sigma(x - \bar{x})(y - \bar{y})}{\sqrt{\Sigma(x - \bar{x})^2}\sqrt{\Sigma(y - y)^2}}$$

Where $\bar{x}$ and $\bar{y}$ are AM's of x and y series respectively.

This method is called the direct method of Karl Pearson calculation.

If there is no possibility of confusion in it then we write r (x, y) as only *r*.

By mathematical method it can be proved as $-1 \le r \le 1$.

If correlation between two variables is linear then the explanation of correlation coefficient of Karl Pearson will be done by following way.

Value of 'r'	Status of correlation linear between variables
$r = +1$	• Perfect clear Positive Correlation
$0.75 \le r < 1$	• High level Positive Correlation
$0.50 \le r < 0.75$	• Normal Positive Correlation
$0 < r < 0.50$	• Low level Positive Correlation
$r = 0$	• Absence of correlation
$-0.50 < r < 0$	• Low level negative correlation
$-0.75 < r \le -0.50$	• Normal negative correlation
$-1 < r \le -0.75$	• High level of negative correlation
$r = -1$	• Perfect clear negative correlation

Remarks 1. Karl Pearson's correlation coefficient can be described as product moment correlation coefficient or Karl Pearson moment correlation coefficient also.

Remarks 2. In Karl Pearson's correlation coefficient r can be indicated by ρ (x, *y*) or only by ρ also.

Letter ρ is Greek alphabet 'rho'

Remarks 3. Karl Pearson's geometrical correlation coefficient is called coefficient of calculation (determination) also.

Notes

For example

If $r = 0.753$ then determination of coefficient will be $(0.783)^2 = 0.567$.

Coefficients of determination always exist between 0 and 1 (in which both are included)

Remarks 4.

$$r = \frac{\Sigma(x-\bar{x})(y-\bar{y})}{\sqrt{\Sigma(x-\bar{x})^2}\sqrt{\Sigma(y-\bar{y})^2}}$$ which is described.

$$r = \frac{\Sigma(x-\bar{x})(y-\bar{y})}{\sqrt[n]{\frac{\Sigma(x-\bar{x})^2}{n}}\sqrt{\sqrt{\frac{\Sigma(y-\bar{y})^2}{n}}}}$$

$$r = \frac{\Sigma(x-\bar{x})\ (y-\bar{y})}{n\sigma_n\sigma_y}$$

Example: With the help of following chart calculate correlation coefficient

X	10	12	15	8	5	16
Y	15	16	8	10	11	12

According to that above question will be solve likewise.

Table 9.5

X	***Y***	***x***	***y***	x^2	y^2	***xy***
10	15	-1	+3	1	9	-3
12	16	+1	+4	1	16	+4
15	8	+4	-4	16	16	-16
8	10	-3	-2	9	4	+6
5	11	-6	-1	36	1	+6
16	12	+5	-0	25	0	0
66	72	0	0	$\Sigma x^2 = 88$	$\Sigma y^2 = 46$	$\Sigma xy = -3$

Mean at $x = 11$ and $y = 12$ respectively,

1. First we calculate mean of both series separately.
2. Then we will calculate difference from mean of both series. From each number of first series we deduct mean value. Thus we will get mean value (11) and difference from number *x*. Similarly we will deduct mean value 12 from each number of second series and in this way we get Y column.
3. Now in the next column we square *x* and *y* column's number. Thus we get x^2 and y^2 column.
4. Again we multiply the number of *x* and *y* column in the next column Thus we get *xy* column.
5. Again we add three column viz x^2, y^2 and xy separately. This will be written as Σx, Σy and *xy*.

Notes

After doing all these calculation now we will use formula. By product moment method, correlation coefficient will be calculated by following formula.

$$x = \frac{\Sigma xy}{\sqrt{\Sigma x^2 \times \Sigma y^2}}$$

In the above question

$$\Sigma xy = -3$$

$$\Sigma x^2 = 88$$

$$\Sigma y^2 = 46$$

Hence,

$$r = \frac{3}{\sqrt{(88)46}}$$

$$= \frac{-3}{\sqrt{4048}}$$

$$= \frac{-3}{63.62}$$

$$= -0.047.$$

6. According to a third method we do not calculate actual mean rather than we use assumed mean. Below the same question has been solved by assumed mean.

Table 9.6

X	*Y*	*x*	*y*	x^2	y^2	*xy*
10	15	0	5	0	25	0
12	16	2	6	4	36	12
15	8	5	-2	25	4	-10
8	10	-2	0	4	0	0
5	11	-5	+1	25	1	-5
16	12	+6	+2	36	4	+12
66	72	0	0	94	70	+9

In the above question actual mean are $x = 11$, $y = 12$

Whereas assumed mean x = 10 and y = 10

$$C_x = 1, C_y = 2$$

C_x (first difference of actual mean and assumed mean = 11 – 10 = 1

C_y (Second's difference of actual mean and assumed mean) = 12 – 10 = 2

$$C_x^2 = (1)^2 = 1$$

$$C_y^2 = (2)^2 = 4$$

$$\sigma_x = \sqrt{\frac{\Sigma x^2}{x1} - C_x^2} = \sqrt{\frac{94}{6}} - 1 = 3.93$$

$$\sigma_y = \sqrt{\frac{\Sigma y^2}{N} - C_y^2}$$

Notes

$$= \sqrt{\frac{70}{6} - 4} = 2.76$$

According to formula

$$r_{xy} = \frac{\frac{\Sigma xy}{N} - C_x \cdot C_y}{\sigma_x \cdot \sigma_y}$$

$$= \frac{\frac{9}{6} - 1 \times 2}{3.83 \times 2.76}$$

$$= \frac{-.5}{10.57} = (-0.047).$$

This method is used in such a condition which is convenient and useful in which actual mean comes in decimals examples 104.3 or 98.6 etc because in such a condition deduction of such mean from true scores and to make their squares is very tough. Hence in place of actual mean we take mean of perfect integer around scores. This is called assumed mean. By doing so difficulties of x and y and making their squares is removed.

In fact, the use of rank system (rho)'s is more appropriate when to calculate correlation coefficient of less scores. Up to a range of 30 scores rank method can be used. Calculate correlation coefficient by product moment method of more than this number of scores.

Correlation coefficient (rho) computed by rank method almost equal to calculated by product moment method. **J. P. Gilford** statement is that in any condition correlation coefficient calculated by both methods is not more difference than 0.018 and in each condition correlation coefficient ρ (rho) calculated by product moment method (*r*) is greater than the correlation coefficient calculated by Rank system (only left to condition Zero and + 1.0 correlation coefficient). But this statement is not correct through practical approach. By previous example it is clear. By Rank Method result ρ is + 0.085 and by Product—Moment result is –0.047. Difference of both is approx 0.038. Second, ρ (rho) is greater then *r*. Product moment correlation coefficient is calculated only in those condition where following conditions is fulfilled:

1. When sample is large.
2. When distribution of scores is even.
3. When measurement is available in continuously.

9.6 Summary

- When two variables are inter-related in such a way that increase in one causes increase in second also or vice-versa, then in that condition we can say that correlation exists between these two quantities.
- Correlation is of three types—1. Positive correlation, 2. Negative correlation, 3. Zero correlation
- In educational field mostly the quantity of two series need to determine. This relationship can be expressed as coefficient of correlation.
- Rank method is called series rank method also. This method had been postulated by Karl Spearman and name of we call this method Spearman correlation method after his name and we called the correlation coefficient obtained by this method as Spearman's correlation coefficient or Spearman's series rank correlation coefficient also.

Notes

- Rank method is not more reliable. In this method only ranks of numbers has been considered their own value has of no importance.
- This is called direct method of computing of Karl Pearson's correlation coefficient.

9.7 Keywords

1. **Rotation**—rotating, revolving.
2. **Scattering**—Scatter, Spread.

9.8 Review Questions

1. What is meant by correlation? Does it always reflect cause and effect relationship between two variables?
2. What are main characteristics of the Karl Pearson's correlation coefficient. Upon which concepts this formula is based?
3. What do you mean by correlation coefficient? Explain its general rules.
4. By Spearman's Rank Difference Method calculate correlation coefficient.

Student	A	B	C	D	E	F	G	H	I	J
Math	78	36	98	25	75	80	25	62	36	40
Hindi	84	54	36	60	54	92	36	62	36	98
								(Ans. 0.46)		

5. By following data calculate Spearman's correlation coefficient-

X	15	25	38	25	15	12	40	25
Y	40	60	20	25	30	50	20	30
							(Ans. 0.64)	

Answers: Self Assessment

1. between 2. Three 3. Extrovert 4. True .
5. False 6. True 7. False

9.9 Further Readings

Books

1. Educational Research's work procedure—*L. Kaul, Vikas Publishing*
2. Educational Techniques and Assessment—*Dr. Rampal Singh, Bhatt Brothers.*
3. Educational Technique—*S. S. Mathur, Bhatt Brothers.*
4. Educational Technique Management and Evaluation—J. *C. Agrawal, Bhatt Brothers*

Unit 10: Chi-Square Test

Notes

CONTENTS

Objectives
Introduction
10.1 Meaning and Structure of Non-Parametrical Test
10.2 Chi-Square Test
10.3 Summary
10.4 Keywords
10.5 Review Questions
10.6 Further Readings

Objectives

After studying this unit, students will be able to:

- Understand the meaning and structure of non-parametrical test
- Understand Chi-Square Test.

Introduction

It has been already said that the use of parametric test like - T- test and F – Test's of depends upon several conditions. It can be used till these conditions is not fulfilled. These two equal distribution and homogeneity of variance are the most important conditions. These does not fulfils many conditions of research. Hence in that these test cannot be used. In this condition non-parametric test is used. Each paramedic test has a number of non-paramedic options are available.

10.1 Meaning and Structure of Non-Parametrical Test

Use of Parametrical test is based upon some concepts regarding population. Their applications depends upon the characteristics of these parameters. Hence these are called parametrical test. Just its opposite non-parametrical test does not impose any condition of anyone regarding parametrical population, therefore these are called non-parametrical tests. In some non-parametrical test some concepts are vested, but they are least and of general types. In the universe, there is no condition regarding distribution of variables. Therefore, generally it is called is distribution free statistical method. In non-parametrical statistical techniques groups are compared on the basis of distribution of variables in place of prameters such as mean, variance etc.

Advantage of Non-Parametrical Tests

These tests have following advantages.

1. In the condition of very low sample (N=6) only these tests can be used.

Notes

2. This test can be used where the samples have been categorized in different population groups
3. Whatever may be the shape of distribution of variables in the whole population these tests have the capacity of correct statement of probability of available results.
4. When thesis is available in the form of categories and ranks then the application of these tests is possible.
5. Understanding these tests, learning their application and computation regarding these etc. are very simple.
6. It can be used in those circumstances in which parametrical test is applicable.
7. In any condition of parametrical test where their conditions do not fulfils, non-parametrical tests can be used.

Demerits of Non-parametrical Tests

These tests have some own limitations also which is mentioned below-

1. In the condition of any one parametrical test as much power of the test on sample of definite size is available as to get the same power at test in the condition of non-parametrical test sample's size will be made to increase In other words, there tests power is less.
2. There is none tests in which interaction affects of variables can be studies. In spite of these difficulties non-parametrical tests have great importance because research condition is such a way in that parametrical conditions is fulfiled and alternative are found out.

Some specific and more used non-parametrical tests have been described in this chapter.

Non-parametrical Independent Design Group

Under this heading those non-parametrical tests has been dealt which is used in comparing of independent groups.

Self Assessment

Fill in the blanks:

1. Chi-square design is …….of mutual relationship of two variables.
2. In the case of single sample to find out expected frequencies……hypotheses can be used.
3. Chi-square is an alternative of non-parametrical test of ……...

10.2 Chi-Square Test

Chi-square (χ^2) is such a non-parametrical statistical method in which it is tested that a frequency distribution which is obtained by research is significantly differ or not on the basis of expected frequency distribution. If they are not different then both is considered to be associated. Thus Chi-Square is a test method of mutual relationship of two variables or two frequency distribution. The help of this analytical method is taken to know whether the difference between obtained and expected frequencies is only a chance or a realistic. There are a number of types of this design. It is used mainly in four different conditions. These are:

Notes

1. One way status
2. Two independent group's status
3. Contingency-analysis
4. More than two independent groups' status

1. One way Status

This method is used where there is acquired frequency distribution is compared to such a distribution which is a standard distribution and which is expected from a definite hypotheses as a normal probability hypotheses or normal distribution's hypothesis. It will be clear from following given example.

Example 1

One researcher try to know whether the educated females like to do jobs. To know this he selected 48 educated females and asked her if she likes doing job. She was told to choose one out of three alternatives ('yes', 'no' 'neutral'). He got each answer's frequencies like this-24, 12, 12. On the basis of normal distribution he tested if obtained distribution (24, 12, 12) is significantly differ from normal distribution (16, 16, 16). He is assumption was that if really females likes doing job then their distribution should be differ from expected distribution (16, 16, 16) i.e, if they are not serious about doing jobs or hot then 48 women's distribution answer under three alternate will be same (16, 16, 16). He tested this distribution by Chi-Square method to know whether the obtained distribution is significantly different or not. To do this he prepared a table and on the basis of that he calculated the value of Chi-square. The whole method of this calculation has been given in the following table.

Notes Chi-square design is a method of two variables or two frequencies mutual relationship.

Table 10.1: One sample Chi-Square Test

	Preferred	Not-preferred	Indifferent	Total
Fo	24	12	12	48
Fe	16	16	16	48
Fo-Fe	8	4	4	
$(Fo-Fe)^2$	64	16	16	
$\frac{(Fo-Fe)^2}{Fe}$	4	1	1	$6 = x^2$ value

In the above table Fo's meaning is required frequencies, Fe's meaning expected frequencies. This one-way sample is an example of Chi-square because in this data has been collected from only one sample (48 females) i.e, only one sample has been used.

Formula of Calculation of Chi-square is

$$x^2 = \Sigma \frac{(Fo - Fe)^2}{Fe}$$

Notes

In which meaning of Fe is expected frequencies and Fo's meaning is required frequencies.

By above formula table-1 each category Chi-square value has been calculated. After that the values of all categories are sum up. This whole Chi-Square has got 6.0.

Example

As in T-test the value of T is tested for significance, in the same way, in Chi-Square test, the significance of value of Chi-square is tested on the basis of Chi-Square in table-D.

If obtained Chi-value is equal to or more than the given value in the table-D, then it is considered as significant. Rest of the problem process and argument just is as T-test.

Calculation of Dependent Factors (*df*)

In the use of table-D concepts of *df* is used. The formula of getting it is (r-1) x (k-1) i.e, deducting one from each number of row and column, remaining is multiplied mutually. In the Previous example there are 3 Rows and Columns. Thus *df* = (3-1) x (2-1) = 2.

Relevancy- In table D there are 2*df* and at 0.5 level Chi-value is 5.991. Obtained value is 6.0 which is more than table value.

Hence hypotheses H_o that "there is no difference between both distribution" reject it, this conclusion will be drawn that there is significant difference in distribution. In other words it will be to say that females like doing job because "yes" answer frequencies is maximum. If "no" answers frequencies may maximum then conclusion is that females do not likes doing job. Hence, in fact what will be the conclusion, this depends upon data frequencies observations.

If test is to be done from one side then the value of Chi in table D at 0.05 levels observes in column. 01 and for .01 observe in column 0.02

Hypotheses of Expected frequencies

In the condition of oneway sample to know expected frequencies, the concept of "normal probability" can be used as it has been cleared in previous given example. Besides this "normal distribution" concept it can be used also. It depends upon condition of research which method will be used and where. For example if we try to know that any class scores distribution is normal or not then expected frequency will be calculated on the basis of normal distribution.

Limitations of Chi-Square

The application of Chi-Square is prohibited in following conditions.

1. When df only one e.g., K (group) = 2 and any expected frequency is less than 5.
2. When df is more than 1 but more than 20% frequency is less than 5 or any expected frequency is less than 1.

2. Two Independent Group Design

In this condition two independent groups are compared on the basis of Chi-Square, exactly as in T-test, by finding the value of T. The difference is so much that by finding the valve of Chi-square in place

Notes

of T their relevancy is tested and conclusion is drawn whether the both groups are different from any characteristics or are dependent variables to each others Both groups are not different to each other, this hypotheses (Ho) is the base of the test. Hence Chi-Square parametrical test is an alternate of non-parametrical. In case the conditions of T-test is not fulfilled it can not be used in that ase Chi-Square can be used. For example suppose a researcher wants to know "If there is any difference in the desires of doing a job between male and female." He asked one question to each of male and female and puts three alternative of answers "yes", "no" and "natural's" and he gets data. In this condition Chi-Square test can be used.

Method of Use

There will be the same method as in Chi-square Test which has already been described in the case of one way group but there is a difference in this condition in which two groups is used in place of one group. Preparing the table the each column's Chi-Square will be computed and in the end summing all those will give Chi-Square.

The formula to find out Chi-Square is afore e.g.,

$$\text{Chi-square} = \Sigma \frac{(Fo - Fe)^2}{Fe}$$

By following illustration method of computation has been clarified.

Example 2

Suppose that the above given question, researcher has asked to 90 males and 100 females and get answer in the form of 'yes', 'no' and 'neutral'. Having prepared the frequency distribution table and test of this hypotheses that "There is no any wrathful difference between male and female about doing desired job". Following table shows data and computation method.

Table. 10.2: Relation between Necessity and Sexuality.

		Response		Mode	
		Yes	No	Indifferent	Total
Male	Fo	14.0	66.0	10.010.0	90
	Fe	196.42	62.52	8.05	
	Fo-Fe	5.42	30.48	1.95	
	$\frac{(Fo - Fe)^2}{Fe}$	1.51	.20	0.47	
	Fo	27.0	66	7.0	100
Female	Fe	21.57	64.44	9.0	
	Fo-Fe	5.43	3.49	2.0	
	$\frac{(Fo - Fe)^2}{Fe}$	1.30	3.18	0.49	
Total			41	17132	190

Notes

To solve out this problem, each column frequencies will be calculated in the value of Chi-Square and finally all values will be solved by putting following procedure.

1. *Fe* (1) *i.e.*, cell-1= $\frac{41}{190} x \frac{90}{1} - 19.42$ (out of 190 total *Fe* for 'yes' response = 41; how many out of 90, total of first row).
2. *Fe* (2) *i.e.* for cell -2 = $\frac{132}{190} x \frac{90}{1} = 16.52$ (out of 190 *Fe* for 'No' response=132; how many out of 90).
3. *Fe* (3) *i.e.* for cell-3 = $\frac{17}{190} x \frac{90}{1} = 8.05$
4. (out of 190 Fe for cell -3, indiff. Response=17; how many out of 90)
5. *Fe* (4) = $\frac{11}{190} x \frac{100}{1} =$ = 21.57 (same argument).
6. *Fe* (5) = $\frac{132}{190} x \frac{100}{1}$ = 64.44 (same argument)
7. *Fe* (6) = $\frac{17}{190} x \frac{100}{1} =$ = 9.00 (same argument)

After that we find out (*Fo-Fe*) for each cell. These values are also given in the table for each cell. and divided by *Fe* convert it into Chi- Square. There are respectively 1.51 + 2.0 + 0.47 + 1.30 + 0.18 + 0.49. The sum of all these i.e, $\Sigma\chi^2 = 4.51$.

Relevancy of Chi-Square

In table D (2 x 1 = 2*df*) at level 0.05 Chi value is 4.991 but obtained value is 4.15 which is less than table value. Hence at level 0.05 H_o *e.g.*, "there is no difference" hypotheses can not be rejected. As a result of this "there is no difference" will have to accept because the value of Chi-Square is not significant. It can be happened by incidentally also.

Conclusion

Therefore the conclusion is drawn that there is no relationship of sex to desire of doing job.

3. Contingency Analysis

If distribution of frequencies is in the form of 2 x 2 table, then there is another method to calculate Chi-square which is very simple and used in most conditions. In this rows (*r*)'s number is 2 and columns (*k*) number is 2 also. Hence in this *df* = (r-1) (*k*-1) e.g., (2-1) (2-)-1. To find out Chi-Square following formula is

$$x^2 = \frac{H\left[|AD - BC| - \frac{N}{2}\right]^2}{(A+B)(C+D)(A-C)(B+C)}$$

Notes

In which $|AD\text{-}BC|$'s meaning is that deduct *AD* form *BC* or *vice-versa*

In the above formula deduct N/2 from (AD-BC) which is called (Yates Correction)

The reason of this obtained Chi values' have to make to their actual e.g., theory based and expected Chi-Square distribution. According to **Moron**e (1954) tabulated distribution is bionomical distribution is which is not continuous whereas Chi-Square distribution is continuous. Therefore to make similar binomial distribution into Chi-Square distribution this correction is to be done.

Example-3

A researcher try to know if students' socio-economic status affects their academic achievement. He selected 48 of low level Socio-economic students and divided their academic achievements into two category-success and failure. He arranged this data into 2 x 2 table calculate Chi-Square value on contingency formula and tested the hypotheses. Following table and based calculation are given below:-

Table. 10.3 : Examination Result

	Fail	Pass	Total
SET	36(A)	12(B)	48
SES	32(c)	16(D)	48
Total	68	28	96

To know the answer of this question Chi-Square test can be followed. In this data 2 x 2's contingence table is prepared and putting the value in formula.

$$x^2 = \frac{H\left[|AD-BC|-\frac{N}{2}\right]^2}{(A+B)(C+D)(A-C)(B+D)}$$

$$= \frac{\left[96|(36)(16)-(12)(32)|-96/2\right]^2}{48\times 48\times 68\times 28}$$

$$= = \frac{96\left[576-384-48\right]^2}{4386816}$$

$$= \frac{1990656}{4386816}$$

$$= 0.4537 \text{ for } df$$

Relevancy and Conclusion

At 1 df and 0.05 level in table D, value of Chi-Square is 3.841. Obtained Chi-Square value is 0.4537 which is very low than table value. Hence H_o hypotheses "both groups (HSES & LSES) there is no difference" cannot rejected. Obtained Chi-Square is not wrathful e.g., it can be happened incidentally also. Therefore final conclusion will be that students' Socio-economic level does not affect their academic achievement. It seems that there is no relationship between socio-economic level and academic achievements.

Notes

2 × 2 Contingence Methods limitations.

Did You Know? Contingency method is used only when expected frequencies is fairly large.

If this condition is not fulfilled then by combining nearly frequencies by reducing number of cells frequency of 2×2 table can be increased. More than that following points is considered necessary.

1. When the size of *N* is greater than 40 then for continuity use of application is must for correction.
2. When *N* lies between 20-40 then use above formula only when, all expected frequency of cells is 5 or greater than 5.

Caution When *N* is less than 20 then Chi-Square should not be used

4. More than two independent group status

If there are two or more than two Independent Group to be compared then Chi-square method can be used. In this situation frequencies are arranged in *K* x *r* table. Hypothesis (H_o) is remained as previous. *k* is not different from themselves. χ^2 test is done in the same way as it is explained earlier. But in this method calculation of frequencies of all cells is different. It will be explained with suitable examples later on.

It is the regular option of one-way analysis of variance. If any research is in more than two group but if the condition of *F*-test is not satisfied then it will be used.

Illustration

A researcher measures intelligence of 100 students of a class to know that there is any relationship between intelligence and academic achievement and categorized it into three groups-extraordinary, reasonably intelligence and less intelligence. Thus he divides academic achievements into four category as a first, second, third and fail. He prepares a table 4 x 3 by taking the students falling in the group whose size and Chi-Square calculation is given below.

Table 10.4: Academic Achievement

Wisdom		Fail	III Division	II Division	I Division	Total
Maximum	O	5	6	9	8	28
Wise	E	5.88	8.40	7.84	5.88	
	O-E	.88	2.40	1.16	3.88	
Sample	O	5	16	13	8	
Wise	E	8.82	12.00	11.76	8.82	42
	O-E	3.82	3.40	3..40	.82	
Minimum	O	11	8	6	5	30
Wise	E	6.30	9.0	8.40	6.30	
	O-E	4.70	1.0	2.40	1.30	
Total		21	30	28	21	100

Notes

In table 27.4 before obtained frequencies of each cell has given and before E expected frequencies is given. To calculate expected frequencies multiply the sum of each cell and each column then divide it by sum of frequencies (*N*). Behind this there is an argument. Take cell one. In this all persons are those which lie in first line and in first column. The probability of any one person in this line will be obtained by dividing the sum of all number (N) by 100 e.g., that will be 28/100. Similarly the probability of the person in first column will be 21/100. Hence his probability in first line and in first column will be $\frac{28}{100} \times \frac{21}{100}$. This is the probability which occurred in one . total frequency is 100. Hence out of 100 how many will be the probability $\left(\frac{28}{100} \times \frac{21}{100}\right)$multiply it by 100. Thus total expected frequencies of first cell will be $\frac{28}{100} \times \frac{21}{100} \times \frac{100}{1}$ =5.88

Similarly by multiplying all the sums of expected frequencies of the cells and their columns will be divided by 100. All other cells expected frequencies can be found. It is numerated in table 4 on the basis of that by previous formula each cells Chi-Square can be find out.

Task Throw light on Chi-Square having more, than two independent groups.

$\chi = \frac{(O-E)^2}{E}$ i.e.

$(.88)^2 + 5.88 / (2.40)^2 + 8.40 / (1.16)^2 + 7.84 / (3.88)^2 + 5.88 / (3.82)^2 + 8.82 / (3.40)^2 + 12.60 / (1.24)^2 + 11.76 / (.82)^2 + 8.82 / (4.70)^2 + 6.30 / (1.0)^2 + 9.0 / (2.40)^2 + 8.40 / (1.30)^2 + 6.30$

Having calculation of above value Chi-Square =

.774 + .686 + .172 + 2.560 + 1.654 + .917 + 1.538 + 0.76 + .506 + .111 + .686 + .268=12.948

Relevancy and Conclusion

In this example *df* = (4-1) (3-1) = 6.

In the table D at *df* 6 and .05 level Chi-Square is 12.59. Obtained value of this example is 12.948 which is more than table value. Hence "intelligence does not affects academic achievements" of Ho. Or "there is no relationship of intelligence with academic achievement". It is rejected and this conclusion is formulated that intelligence affect academic achievements or they are related to each other.

Self Assessment

State whether the following statements are True or False:

4. Chi-Square distribution is continuous.
5. Contingency method is used where there is expected frequencies of fairly very small.
6. In Chi-Square test means each cell expected frequencies calculation method is different
7. When to compare less than two independent groups then Chi-Square can be used

Notes

10.3 Summary

Parametrical test's based upon some assumption. used regarding population characteristics Just its opposite non-parametrical test does not impose any condition regarding. Population regarding parametrical Therefore it is called non-parametrical test.

- Chi-Square is such a parametrical statistical method in which it is tested that any frequency distribution which is obtained frm any research that on the basis of any hypotheses from expected frequencies distribution differ significantly or not
- There are a number of types of there design mainly it is used in four different condition. 1. One way condition 2. Two independent groups condition 3. Contingency analysis 4. More than two independent groups condition

10.4 Keywords

1. **Theory based:** on the basis of theory.
2. **Negative:** underscores that numbers which is negative value is called negative.

10.5 Review Questions

1. What do you mean by non-parametrical test? Mention its merits and demerits.
2. What do you mean by Chi-Square design? Mention its types.
3. What do you mean by two independent groups design? Describe it.
4. What is contingency analysis ? Describe.

Answers: Self Assessment

1. Test method	2. Equal or normal distribution		3. T-Test
4. True	5. False	6. True	7. False

10.6 Further Readings

Books

1. Academic Research Methods – *Shareen and Shashikala, Vinod Pustak Mondir*
2. Academic Techniques – *S.S. Mathur, Bhatt Brothers*
3. Education Technique – *R.A. Sharma, Bhatt Brothers*
4. Academic Technique and Assessment – *Dr Ram Pal Singh, Bhatt Brothers*

Unit 11: T-Test

Notes

CONTENTS

Objectives
Introduction
11.1 Main Basis of Statistical Analysis
11.2 Parametrical Test: T-Test
11.3 T-Test : Special Remark
11.4 Single-end and Double-end Test
11.5 Summary
11.6 Keywords
11.7 Review Questions
11.8 Further Readings

Objectives

After studying this unit, students will be able to:

- Understand main basis of statistical analysis
- Understand parametrical test: T-Test
- Understand single-end and double-end test.

Introduction

In the last few chapters strategy of experimental design, their types and rules had been explained. In the coming chapters some specific methods of statistical analysis of data have been described. Researchers collected various types of data in different conditions. All types of data analysis by any one statistical method is not possible. Various types of data analysis is done by a numerous methods. There are number of statistical methods which can be used. Hence researcher has to select a method which is more suitable according to problem's nature, purpose and in context of collected material. Therefore, for a researcher it is quite pertinent to know, which are the main methods, which method to be used in which condition and how it can be used. In coming chapters different statistical methods have been described in this context.

11.1 Main Basis of Statistical Analysis

In the present chapter some matters related to statistical analysis process have been explained. It seems to be important to do so these are the main basis of statistical analysis.

1. Basic Procedure
2. Null Hypothesis

Notes

3. Relevancy Test
4. Types of error
5. Statistical Test Sample
6. Parametrical Test
7. Non-Parametrical Test

1. Basic Procedure

In every experimental research of statistical analysis process, following steps are followed.

1. Mention of statistical hypotheses or null hypotheses.
2. For the rejection of null hypotheses one suitable statistical test (As a T-Test or F-Test or any other test) selection
3. For the rejection of null hypotheses, selection of relevancy level. (as a 0.01 or 0.05) and their mention.
4. Under step-2 fixed value of statistical test's computation
5. Under Step-4 compare obtained statistical value to that related stabilized Statistical value and on that basis to take decision that null hypothesis should be rejected or not.
6. Conclusion drawing regarding research analysis.

2. Null Hypothesis

The first gateway of experimental data analysis from where to enter is null hypothesis. It was invented by late scientist, mathematician Ronald Fisher. Research hypothesis like scientific hypothesis or also called substantive hypothesis which is only on operational statement and in which between two variables any relationship specially vested, changes into null hypothesis because it is null hypothesis whose statistical test is possible. On the basis of that rejection regarding research hypothesis something to say or draw conclusion is possible. There is not any statiscal method on the basis of which basis research hypothesis can be accepted or rejected directly and drawn conclusion can be drawn regarding that. Only null hypothesis rejection or non-rejection basis formulation of conclusion regarding research hypothesis is done. Due to this reason research hypothesis is called alternative hypothesis H_1 also. If null hypothesis (H_0) is rejected then alternative hypothesis (H_1) i.e., research hypothesis is accepted. Similarly if null hypothesis (H_0) can not be rejected then research hypothesis i.e., alternative hypothesis (H_1) can be rejected. In null hypothesis there is an inverse relationship between H_0 and H_1. In construction of research design it is necessary to mention two types hypothesis, in the begining.

Null-hypothesis is rejected at that condition when obtained statistical value. (T.F, Chi-Squares, etc.) pre-determined relevancy level is equal to given value of table or more than that. In this way H_0 is the base of testing data.

3. Relevancy-Test

The purpose of relevancy test is to know that the results obtained on statistical test basis, that is not due to incidence e.g., that is due to that variables, that treatment on independent variables whose effects is studying. When H_0 is rejected at any level of relevancy then it is accepted that obtained result occurred due to independent variables or treatment. This is not purely co-incidental. In other words. obtained result is real e.g., occurred as a result of independent variables assumed as. Thus ahead logical order vested in relevancy test.

Notes

H_0 Rejected → worthful result → real result e.g., under experiment treatment's effect's cause, not by incidentally

H_0 - not rejected → result worthless → result by incidentally e.g., under treatment effect's absence.

Regarding this context it is quite necessary to understand the concept of relevancy level. It is known as Alfa (α) a Greek Letter. Its meaning is that " That concept which researcher select to reject of null hypothesis" for example .05 level chore for that purpose then its meaning will be that obtained statistical value (T, F, χ, etc.) probability of occurrence is equal or less than .05 then H_0 will be rejected. Hence is relevancy level is determined at that time it is fix also when null hypothesis (H_0) will be rejected. Thus relevancy level is that criterion, that base which is use to reject in H_0 Alfa or relevancy level is such a dividing point on normal probability curve below that all probability exists that rejected H_0 and above that all those which does not accept it. For the rejection of null hypothesis generally .05 and .01 level is used. But it selects arbiter ally. As per rule selection of relevancy level should be scientific and logical.

Consequently its selection should be done prior to collection of data because sample size is such on element which to pay attention in decision making. Hence researcher keep in mind problems nature on the basis at result's importance and required accuracy Alpha should be selected. To take decision of it got answer of such types of questions

1. How important are the results?
2. How much risk is there in incorrect result?
3. How much accuracy is required ?
4. According to conditions What is the possibility of change error occurrence? Etc

4. Types of Error

Whatever may be the type of research design are and whichever alpha level may be, null hypothesis (H_0) rejected or not rejected very natural of being error. As relevancy level determined these error is determined also. Null hypotheses rejected or not rejected causes two types of error, it is called type-I, type- II, errors. The meaning of type – I error is H_0's wrong or incorrect rejection e.g., having it true be that is rejected. In other words, its meaning is that draw conclusion that independent variables or treatment has effected on dependent variables whereas really there is no effect because in dependent variables occurred or available changing happened only by incidence. Second type error's meaning Ho incorrect or wrong acceptance e.g. accept Ho where as it should be rejected.

In other words conclusion drawn from it that there is no effect of independent variables on dependent variables whereas it do, no relationship but relationship is. Type- I depends upon error's probable relevancy level. If relevancy level is large (as in place of .01, .01) then H_0 rejection as incorrect form possibility increase e.g. type –I error vested in relevancy level and it is called Alpha (α) Type – II error is called Beta (β). The limitations of these two types of error depends upon sample result. Therefore it is said that alpha (α) and N (Sample result) regarding these two take decision begning of research. **Segal's statement** is that "If we want to less there two types of error then we should increase result of sample (N)."

There is a special type of relationship between two types of error. If we try to less one type of error then other types of error increased. Therefore researcher should compromise between there two types of error and try to maximize balance. The meaning of statistical test is "H' s correctly rejection". Which can be express by following formula.

$$\text{Power} = 1 - \beta \text{ (type – II error)}$$

Since type – II error probability increase in sample N causes less, it can be understood clearly that having greater the result of sample statistical test power is increased.

Notes

5. Statistical Test Sample

To know relevancy of hypothesis' test any one statistical test (T-Test or F-Test) uses.

Every Statistical method or test out of these depends upon any specific statistical model. As a T-Test's base in T distribution model. Similarly F-Test's base is F distribution. Some types of many statistical model or tests are which uses in various condones for the test of hypothesis. Therefore this problem arise that in any one condition which statistical method used. Availability of a number of alternatives, selection any one of that must be a sound base and logic.

One the most important rule is that Statistical method should be used which is most powerful in that condition e.g. in which null hypothesis' incorrect rejection possibility less and right rejection possibility is more this is called statistical test power. Hence in the selection of Statistical model one matter keep in mind that is his power. Besides this some other matter pay attention also, as a- method of sample selection, characteristics of population, types of measures scale to be used for measuring etc. regarding all those each statistical test have some conditions. Fulfilling these conditions that test uses. Some statistical tests are such that whose conditions are most complicated and therefore they are most powerful. Secondly some tests' conditions do not exist if it exists then very ordinary nature. Therefore there tests is not more powerful. The result getting on these basis they are not much accurate or valid and is of general natures but it should be memorized that if powerful tests conditions and their basic concepts do not fulfill then in that condition their result may be least accurate with respect to less powerful tests result. Therefore, important is to observe that in the given condition which test is more powerful. Not any statistical method are such that which being more powerful in all circumstances.

Power and basic assumptions point at view two types of statistical model which uses. That is

(i) Parametric

(ii) Non-parametric

Under parametric test, T-Test, F-Test and their various forms comes. There tests are more powerful, although its conditions are very tough and hard therefore, they are powerful only when their conditions are fulfilled On the other hand, non-parametric tests whose nominal conditions, in some parameter related no conditions terms and conditions and basic concepts of each has been mention ahead.

6. Parametrical Test

According to mostly writers major parametric tests conditions are following-

1. The selection of sample from that population is which distribution of variables are normal. Regarding this Nunlee's statement is that without fulfilling this condition, whenever any big problem does not arise, distribution of modules is not much different from normal distribution (as a J- curve). Nunlee added further that if this rule is broken, it does not cause great loss because T-Test and F-Test are considered to be more powerful and "concepts of normal distribution not fulfilling affect result to very low degree.
2. Second assumption is homogeneity of distribution. The meaning of it is that all comparable groups must be alike to inside group distribution. In the condition of F- Test use it consider to be more powerful. If does not so, F-Test's result relevancy not comes prevail whereas in really between groups significant difference will be.

   Regarding above two conceptions **Karligar Bonu** (1960), **Anderson** (1961), **Games** & **Lukas** 1966 etc. Statement is that "There is given more stress than to need on normal distribution and homogeneity".
3. A third concept is that available measures on variables should be of continuous and equal interval so that mathematical mechanism editing will be possible.

Notes

4. A fourth concept that is mentioned by **Segal** is it that distribution of units in groups and their selection from population should be done by normal probability method e.g., observations, measuring etc. Should be independent.

If these conditions or concepts do not fulfill in any research process then regarding parametric test this statement that they are not more powerful, is not appropriate. Perhaps, it is difficult to say what is power of any test. At that condition regarding accuracy of research result guess is tough also consequently in that condition probability of hypothesis test mention is not meaningful.

7. Non-Parametric Test

Regarding non-parametrical test measurements' distribution like normal distribution does not bound any concepts. From this approach they are called distribution free tests. Similarly continuity of measures and their equal interval being condition does not impose. The uses of these tests can be done in very small sample condition. Hence it is called small sample test. In condition of very small sample parametrical test cannot be used.

The use of non-parametrical test is done in following condition.

(i) When data or measures (e.g. – I, II, III) or in form of ranks.

(ii) When distribution of scores is too skewed or how will be the distribution, it is not possible to guess.

(iii) When the magnitude of sample is very small.

(iv) When data is not in discrete only available in tendencies.

It is clear frm above description that at the time of selection of statistical test a number of things to be considered. Specially

(i) Power of test

(ii) Basic concepts of test

(iii) Form of available scores, etc

In the next few chapters important parametrical and non-parametrical test has been described. Where and how they uses, mention it in detailed.

Self Assessment

Fill in the blanks:

1. Null hypothesis was invented by scientist mathematician ……..
2. Research hypothesis is also called ……hypothesis (H_1)
3. Base of T-Test ……..is model.
4. The meaning of statistical test power accuracy (H_0)……..

11.2 Parametrical Test: T-Test

Parametrical test deals with T- Test and F- Test. In this chapter only T-Test has been described. If groups are constructed only independent variables e.g. manipulation of independent variables has been done in two categories and only these two groups compares on dependent variables is to be done then at that condition T-Test uses. T-Test is a parametrical test and in this only one independent variables which is manipulated by two or more than two groups. These types of research design are called single variable designs.

Notes

From this point of view T-Test is of two types-

(i) Independent group based

(ii) Correlated or remeasurable based.

Independent groups' condition both two groups unit differs. In the condition of correlation groups one unit of a group keeps under two treatment. If unit matched distributed in two groups rather groups' unit differ they are consider to be correlated. T-Test is used in both conditions but their formula is different.

From an another print of view both types of T-Test have two forms

(i) Small –Sample based

(ii) Large-Sample based

In both these conditions separate formula is usesd. Thus T-Test has total four types-

(i) Large Sample independent group

(ii) Large Sample correlated group

(iii) Small Sample independent group

(iv) Small Sample correlated group

Each is described separately ahead.

1. Large Sample Independent Group T-Test

This test is done where number of groups are two, group is independent, sample is large and fulfills the condition of parametrical test. In this comparison is done on the basis of mean difference. Hypothesis which is to be tested that is as follows " There is no significant difference between mean's of groups". For this, difference of mean is converted into T-Value or (C.R. Value) obtained, T or C.R. Value is compared with at pre-selected significant level (.05 or .01) value given in T-Table on the observations of value of T in table degrees of freedom is keep in mind. If obtained T-Value at that significant level and at that degree of freedom equal to table value or more than that null hypothesis (H_0) is rejected and accepted alternative hypothesis (H_1). If this obtained value is less than table value then in place of H_0, H_1 is rejected. Thus indirectly or H_0 is not rejected drawn conclusion regarding research hypothesis. It will be cleared from given below example.

Example

Hypothesis- An eleven years old boys' achievement-motivation level and his academic achievement has no difference.

Analytical Design: - high achievement motivation and low achievement motivation an eleven years old boys two independent group constructed, calculate their mean and deviation value separately of their academic achievement (marks in the test).

In the table of following example data has been presented.

Table 11.1: Relationship between achievement Motivation and Academic Achievement

	Group I Low motivation	**Group II High motivation**
Mean	88.80	90.5
Deviation	7.81	11.56
Numbers	83	97

Notes

This is an example of large sample independent group. From this C.R. calculation process is as follows.

Step 1: Calculate S.E. of each mean its formula is

$$S.E. = \frac{S.D.}{\sqrt{N}}$$

Hence, $$S.E_1 = \frac{7.81}{\sqrt{83}} = .857$$

$$S.E_2 = \frac{11.56}{\sqrt{95}} = 1.22$$

Step 2: Now on this basis calculate $S.E._D$ (S.E. of difference).

Its formula is

$$S.E._D = \sqrt{\left[(S.E._1)^2 + (S.E._2)^2\right]} \qquad \text{Formula (1)}$$

Accordingly this formula $S.E._D = \sqrt{\left[(.857)^2 + (1.22)^2\right]}$

$= 1.493$

Step 3. Now calculate value of T. Its formula is –

$$T = \frac{D}{S.E._D} = \frac{1.70}{1.493} = 1.138$$

Step 4

Test of Significance- For this 196 df and at level .05 in table (B) given t value (1.97)is now compared with obtained t value. It is clear that obtained t value 1.138 is less than 1.97 from table t value. Hence this difference at level .05 is not relevant. If not relevant at that level then it will not be relevant at level .01 also. Hence H_o cannot be rejected. As a result H_0 (having difference) is rejected.

Step 5

Conclusion- Therefore conclusion is drawn that both groups' mean has no real difference, e.g., probably there is no relationship between achievement motivation and academic achievement.

Large Sample Correlated Group T-Test

In example-1 given method is used there, where, groups are independent, but when groups are correlated or matched then following method is used. In this to calculate standard error (S.E) formula is same as example- 1 but to calculate $S.E._D$. formula is following

$$S.E._D = (S.E._1^2 + S.E._2^2 - 2r_{12}S.E._1 S.E._2) \qquad \text{Formula (2)}$$

$$= \sqrt{\left[\frac{1}{N}\left(S.D._2^2 - 2r_{12}\right)\left(S.D._1 \times S.D._2\right)\right]}$$

Similarly to calculate value of t, formula is same as before e.g.

$$t = \frac{D}{S.E._D}$$

It has cleared by below given example.

Notes

Illustration 2. Eleven years old 625 boys were given a test of personality balance At that their mean & standard deviation (S.D.) 96.7 and 8.0 respectively found. After that they have been given collective consultation. After two months again he had been given some test. This time mean of their equilibrium was 106.7 and S.D. 10.0. the correlation at both times test was found 0.52. Was collective consultation was impactful?

To get the answer of this question both times equilibrium test's mean's difference relevancy will be needed to test, because both groups are correlated. Therefore to calculate their mean difference $S.E._D$ formula-2 will be used. Whole process will be as follows.

1. $S.E._1$ (first time mean) $\frac{S.D.}{\sqrt{N}} = \frac{8}{\sqrt{625}} = \frac{8}{25} = 32$

2. $S.E._2$ (Second time mean) $\frac{S.D.}{\sqrt{N}} = \frac{10}{625} = \frac{10}{25} = .40$

3. $S.E._D = \sqrt{(S.E._1^2 + S.E._2^2 - 2n_2 S.E._1 S.E._2}$

$$= \sqrt{\left[(.32)^2 + (.40)^2 - 2 \times .52 \times .32 \times .40\right]}$$

$$= .3596$$

After the use of second alternate formula $S.E._D$ value will be same

$$t = \frac{D}{S.E._D}$$

$$= \frac{106.7 - 96.7}{.3596}$$

$$= \frac{10}{.3596}$$

$$= 27.80$$

This t value at each level of table-B 624 *df* (N-1) is much more than given value. Hence H_o is rejected and conclusion is drawn that collective counsel treatment effects equilibrium definitely.

Computation of Degrees of Freedom

It is necessary to memorize in the condition of independent groups computation of degrees of freedom done on the basis of formula $n_1 + n_2 - 2$. The reason of this is that in the condition of correlated and matched group one person tested twice. Hence correlation is formulated with their measurement. Second times obtained measure of test freedom get less because due to group remain some second time achievement cannot be complete differ and independent.

Hence measures comes in first times that cannot be complete free from second times. Therefore one person twice measure do not give two *df* but one *df*. In this way correlated group condition to calculate *df* N – 1 formula uses.

2. Small Sample Independent Group T- Test

If the sum of both two independent groups units is less than 30 then it is considered as small sample. In this condition S.D and $S.E._D$ calculation formula are different. Rest process is same as before. To

Notes

calculate S.D. and $S.E._D$ formula is as follows.

1. $$\sqrt{\left(\frac{\Sigma x^2 + \Sigma y^2}{N_1 + N_2 - 2}\right)}$$

In which Σx^2 and y^2 are the sum squares of differences from their mean of both respective groups.

2. $$S.E._D = S.D. \sqrt{\left(\frac{N_1 + N_2}{N_1 \times N_2}\right)}$$

Clarification of computation of whole process is done through ahead given example.

Table 11.2					
Group X			**Group Y**		
X	**x**	**x²**	**Y**	**y**	**y²**
110	4.5	20.25	115	2.0	4.0
112	6.5	42.25	112	-1.0	1.0
95	-10.5	110.25	109	-4.0	16.0
105	-0.5	0.25	112	-1.0	1.0
111	5.5	30.25	117	+4.0	16.0
97	-8.5	72.25			
112	6.5	42.25			
102	-3.5	12.25			
Sun = 844 N_1=8 Mean =105.5		330= Σx^2	565 N_2=5 Mean =113.0		38 Σy^2

1. Pooled S.D. = $\sqrt{\left(\frac{\Sigma x^2 + \Sigma y^2}{N_1 + N_2 - 2}\right)} = \sqrt{\left(\frac{330 + 35}{8 + 5 - 2}\right)} = 5.78$

2. S.E. = Pooled S.D. × $\left(\frac{N_1 + N_2}{N_1 + N_2}\right) = 5.78 \times \sqrt{\left(\frac{13}{40}\right)} = 3.302$

3. C.R. (t) = $\frac{\text{Diff.}}{3.302} = \frac{7.5}{3.302} = 2.271$ for *df* 11

Task How degree of freedom is computed.

4. **Conclusion:** In table – B, 11*df* ($n_1 + n_2 - 2$) and at .05 level value of *t* is 2.20 but obtained value of + is 2.271. Hence difference of both two mean's in real. Therefore at this level H can be rejected. At .01 level, it is not relevant because at that level value of *t* is 3.11 and obtained value is less than that.

Notes

This analytical method is very important in the context of studies done in laboratory because in these studies result of sample is less but in this conation groups homogeneity or conformity test is necessary only homogeneity of variance exists then this method uses.

4. Small Sample Correlated Group T- Test

In this condition $S.E._D$ calculation formulas some which is use in large sample correlated group e.g., in this condition

$$S.E._D = \sqrt{S.E._1^2 + S.E._2^2 - 2r_{12} \times S.E._1 \times S.E._2}$$

But to obtain there is a second formula also. Its base is both groups measurement difference description. Calculate mean and standard deviation (S.D.) of differences. It is named as D and S_D. On this basis $S.E._D$ is calculated that formula is—

$$S.E._D = \frac{S_D}{\sqrt{N}}$$

This $S.E._D$'s $\frac{D}{S.E._D}$ by formula converted into T Value. It will be cleared from following example.

Table 11.3

X	Y	D	D²
100	105	-5.0	25
90	95	-5.0	25
106	100	+6.0	36
95	105	-10.0	100
102	99	+3.0	9
104	98	+6.0	36
90	110	-20.0	400
105	106	-1.0	1
99	100	-1.0	1
		ΣD = - 27.0	ΣD²= 633

1. $\bar{D} = \frac{-27}{9} = -3.0$

Let d = (D – $\bar{D}$), a deviation score in D. Then:

2. $\Sigma d^2 = \frac{\Sigma D^2}{n} = 633 - \frac{(-27)^2}{9} = 552.00$

3. $\sqrt{\left(\frac{\Sigma d^2}{n-1}\right)} = \sqrt{\left(\frac{552.0}{8}\right)} = 8.306$

4. $\frac{S_D}{\sqrt{N}} = \frac{8.306}{\sqrt{9}} = 2.768$

5. $Z(t) \frac{\bar{D}}{S.E._D} = \frac{-3.0}{2.768} = -1.08$

Notes

If having used formula 2 calculated value of T will be same as aforesaid method. In table B, 8*df* (N-1) and at .05 t′ s value is 2.31. Obtained T value is 1.08 which is less than table value. Hence H_0 can be rejected. Therefore research hypothesis (H_1) e.g., " difference is" is rejected and established this conclusion that between measures of X and Y there is a significant difference.

This method can be used in case of large sample correlated groups also. From computation point of view, this method is very simple, but on that basis we get less information. By this, both groups correlation between scores knowledge is not acquired and neither by its groups' means and standard deviation (S.D) is to be found.

But it should be remaindered that large and small sample correlated groups in both conditions calculation of value of T, method is some but if group has been matched equal on the basis of their mean and S.D. Then to calculate S.E.D formula is folling.

$$\sqrt{\left[S.E._1^2 + S.E._2^2 \times (1 - r_{xy}^2)\right]}$$

Rest Process in this condition is same as before also.

Did You Know? Each statistical method or test based upon a specific statistical model.

Self Assessment

Multiple Choice Questions:

5. To test the hypothesis any one …… test is used.
   (a) Mathematical (b) Statistical (c) Political (d) None of there
6. From basic assumptions point of view statistical models of …… types.
   (a) One (b) Two (c) Four (d) Five
7. T-Test is a……. test.
   (a) Parametrical (b) Non-Parametrical (c) Sample (d) Homogeneous
8. This type of research (T-Test) design …….is called.
   (a) Bi-Variables design (b) Variable design
   (c) Single Variable designs (d) All three

11.3 T-Test: Special Remark

Regarding T-Test's uses some concepts have been mentioned. Some of there are- (i) In population distribution of variables being normal (ii) Sample's population's variance being homogeneous. It can be tested by statistical methods also. It has been mentioned in the higher level books of Statistics as **Hedge's** (1973) book etc. Again it is important thing that T-Test is a powerful statistical test and its result is not much affected if above concepts are broken, but different size independent groups condition is its exception. In this condition if homogeneity of variance does not fulfill conditions then result will be affected to a great extent. Therefore in this condition, concepts must be tested. If this condition does not fulfill then T-Test should not be used. In large sample condition rather not fulfilling there condition does not make special difference, but in small sample condition lack of conformity

Notes

of variance affects result. But in this condition too, if result of sample (Ns) is equal then problem can be reduced some degree.

Caution In the condition of small sample before use of T-Test homogeneity of variance must be tested.

11.4 Single-end and Double–end Test

The explanation of this concept has done in chapter.10 under experimental design. Here only to explain that in case of single-end test how to observe table's T value. For single end test in table at level .05, .10 column and at level .01 in .02 column value of *t* is observed. If on double – end basis obtained *t* value at any significant level not comes significant then to use of single –end test sometimes occurred significant. But to decide which test *t* (Single – end or double- end) will be used, taken prior to collection of data at planning level should be done.

To increase senility of T- Test

Notes The meaning of sensitivity of independent variables or treatment's effect on dependent variable presents their capability or their work efficiency.

Several such measures by that T-Test's this senility can be increased. It has been mentioned ahead.

1. **To Decrease the level of noise**: In this respect the meaning of noise is effects of uncontrolled variables. There effects present in many forms. Experimental design's rules and under different types it has been mentioned already. Mainly its sources are-
   - *(i)* Direction given to persons under research.
   - *(ii)* Incomplete control of post studies variables
   - *(iii)* Management of study and no control over circumstances
   - *(iv)* Researcher's bias and
   - *(v)* persons under research mental set.

Hence due to occurred "noise" (effects), to control it do every possible measures so that experimental effects which is occurred due to independent variables or treatment, explored come before.

2. **To increase size of sample**- To increase the sensitivity of T- Test it is the second measure. To do so random error minimizes and type – I and type – II errors' are decreased. In this way ,effects of independent variables is more cleared.
3. **To remove the effects of "Floor" and "Roof" –** Its meaning is that to measure the variables' used tests should be neither very simple then mostly persons mark will too high, may be cent-percent. At that condition some of them to go high in the tests possibility is decreased. As a result difference will be less due to occurred by treatment. This situation is called "Collide form roof move down towards floor incidence". Consequently if test is very hard then a

Notes

great may number of persons (from marks approach) do not rise above floor. This is called "Floor effect". In this condition treatment effects does not be cleared. Therefore test should be neither very tough nor very simple, then T-Test's sensitivity increases.

4. **To Increase the reliability of Measures**: To do so increase the reliability of Psychological test do one rather many times measures and taken their average and can be done to control properly measuring conditions and management.
5. **Selection of a more Suitable Design:** In study – 10 a number of experimental designs have mentioned. Under this included independent group design, matched group design, correlated repeated group design. T-Test's Sensitivity much in repeated measures design, less than that in matched group design, and least in independent group design. It so happen due to more and less controlling of studied variables. In independent group design. It so happen due to more and less controlling of studied variables. In independent group design this control is least. Hence to increase the sensitivity of T- Test; one measures is that to choose that design in which T-Test sensitivity is most, but it is not possible always. At that condition T- Test sensitivity too less than other two. Therefore it is necessary to keep in mind each merits and demerits and their suitability. Although if in the condition at research it is possible to select such a design which covers most sensitivity of T-Test.

11.5 Summary

- The first gateway of research data's analysis, through which enter is null hypothesis. It had been invented by Late Scientist, mathematician Ronald Fisher.
- If on independent variables only group has constructed or manipulation of independent variables have divided into two groups and comparison is do only these two groups with dependent variables, then at that condition T- Test can be used.
- T- Test is a parametrical test and in this only one independent variable is there which can be manipulated in two or more than two categories.
- T-Test is of two types – (i) Independent group based and (ii) Correlated or remeasurable based.
- Small sample correlated group test can be used in case of large sample correlated group. From computation point of view this method is very simple, but on that basis very few information can be obtained.

11.6 Keywords

1. **Null Hypothesis –** The first stage of experimental data analysis, from where research starts, called null hypothesis.
2. **Equal Interval –** having equal distance, having equal interval.

11.7 Review Questions

1. Mention main basis of Statistical analysis.
2. What do you mean by Parametrical T- Test? Mention its types.
3. Describe single- end and double – end test.
4. Throw light on large sample correlated group T- Test

Notes

Answers: Self Assessment

1. Ronald Fisher	2. Alternative	3. T- Distribution	4. Rejection
5. (b) Statistical	6. (b) Two	7. (a) Parametrical	8. (c) Single variables design

11.8 Further Readings

Books

1. Academic Techniques Management and Assessment—*J.C. Agrawal, Bhatt Brothers*
2. Educational Research Work Procedure—*L. Kaul, Vikas Publishing*
3. Education Technique—*R. A. Sharma, Bhatt Brothers*

Unit 12: F-Test

Notes

CONTENTS

Objectives
Introduction
12.1 One-way Analysis- variance
12.2 Normal Probability Group Design
12.3 One–way Anova
12.4 Two-way Analysis of Variance (Independent Group)
12.5 Two-way Analysis of Variance (Correlated Group)
12.6 Randomized Block Design
12.7 Co-variance Analysis
12.8 L.S.D. (Latin Square Design)
12.9 Summary
12.10 Keywords
12.11 Review Questions
12.12 Further Readings

Objectives

After studying this unit, students will be able to:

- Understand one-way analysis-variance and normal probability group design
- Understand one-way anova
- Understand two-way analysis
- Understand randomized block design and Co- variance analysis.

Introduction

In previous chapter statistical analysis' T- Test had described. The use of T- Test can be done only in that condition in which only two groups are to be compared.

In its one independent variable give only two treatment or divide at two levels, but condition of research is not always same that they are limited to independent variables of two levels or two treatments. In a number of condition independent variables' division in many treatment or many level understood to be more required. At that condition. more than two groups are mutually compared. For example at motivation level what impact on students academic achievement, to studies there researcher determines three (much more motivated reasonably motivated, very low motivated) or more than three level, all students three or more than three groups divided and compare it to academic achievement than

Notes

studies. In this condition T- Test cannot uses. In this condition that method, that test's uses, that is called analysis of variance. It had invented by **Ronald Fisher** in short it is called Anova.

In analysis of variance in place of two mean's difference, total variance of groups total measures, and their different parts or section is the base of analysis. Behind the use of it some concepts are followed which is mentioned in context of T- Test. Like T- Test this test which is called F- Test, is an parametrical test. Hence this test uses when parametrical test concept fulfills. This is a very versatile test which has many transformed form and which can be used in many research conditions. The most general form of its simple analysis of variance and completely randomized groups design or one-way analysis of variance.

Second complicated form is two–way Anova and factor analysis. Firstly, simple variance analysis method will be mentioned.

12.1 One–way Analysis- variance

The literally meaning of 'variance' is difference, but in statistical language its meaning is 'Average of Squares of difference '. This difference is taken from score and mean.

Basically variances' formula is:-

$$\text{Variance} = \frac{\Sigma(x - \bar{x})^2}{N}$$

or and in more impure form $= \frac{\Sigma(x - \bar{x})^2}{N-1}$ or $\frac{\Sigma x^2}{N-1}$

1. Parts of Total Variance

Notes Total variance e.g., from sum of score's mean's taken differences' squares can be divided into many sections or components.

At least two parts or section are always there. These are: between group variance and within groups variance. If these two are added then it will be equal to total variance. By given below example it will be clarified:

Table 12.1 Data of three groups

A-1			A-2			A-3		
X_1	x_1	x_3^2	X_2	x_2	x_3^2	X_3	x_3	x_3^2
6	2	4	4	1	1	6	+1	1
2	–2	4	2	–1	1	5	+0	0
3	–1	1	5	2	4	4	–1	1
5	1	1	1	–2	4	4	–1	1
4	0	0	3	0	0	6	1	1
$\Sigma X_1 = 20$			$\Sigma X_2 = 15$			$\Sigma X_3 = 25$		
$\bar{X}_1 = 4$			$\bar{X}_2 = 3$			$\bar{X}_3 = 5$		
Σx^2		10			10			04
$\Sigma X_1 = 60$								

Notes

Suppose that above table shows three independent randomized groups' (high motivation, normal motivation, low motivation level) boys present academic achievement's scores. We can say these groups as A-1, A-2, A-3. On the basis at analysis of this table it is to be tested that "Does motivation effects on academic achievements".

On the basis of this table firstly calculate total variance which is presented in example -1

Example – 1

Computation of total variance			
X	x	x^2	
6	2	4	
2	–2	4	$e(V_1) = \frac{\Sigma x^2}{N-1} = \frac{34}{14}$
3	–1	1	Total variance = 2.428
5	+1	1	
4	0	0	
4	0	0	
2	–2	4	
5	+1	1	Using formula $\frac{\Sigma x^2}{N}$ value comes to
1	–3	9	$\frac{34}{15}$ = 2.266
3	–1	1	
6	2	1	
5	1	1	
4	0	0	
4	0	0	
6	2	4	
X = 60 $\overline{X}$ = 4.0 ΣX^2 =34			

This total variance is supposed to be occurred due to two reasons. Firstly due to that difference which is vested at three levels of motivation. Secondly due to that difference which lies in each group of boys. Due to these two reasons occurred, variance is called – (i) Intergroup variance and (ii) within group variance.

1. **Inter group variance** due to treatments and within groups variance sampling error or chance error which is called randomized error and that after matched all groups remain something suppose to be occurred due to those. Within groups variance in every experimental researches known as error.
2. **Between Groups Variance** Below given example -2 table's basis in which three groups' mean has been arranged as measures between groups variance is calculated. By this formula $\frac{\Sigma X^2}{N=1}$ we get 1.0 and by formula. $\frac{\Sigma X^2}{N-1}$ we get .666

Notes

Example 2

Calculation of variance			
X	x	x^2	
4	0	0	
3	–1	1	
5	+1	1	$(V_b) = \frac{\Sigma x^2}{N-1} = \frac{2}{2} = 1.00$
ΣX=12			by using formula $\frac{\Sigma x^2}{N}$, it
M = 4.0			results $\frac{2}{3}$ = .666
Σx^2 =2			

3. **Within Groups variance:** Table – 29.1's each group's (A-1, A-2, A-3) within group variance calculation formula is same as given previously. To find it calculate three groups variance separately and then their average taken. Is Each group's Σx^2 given in table -1. This is 10, 10, 4 respectively. In each N – 1 = 4 Hence within groups variance is $V_b = \frac{\frac{10}{4}+\frac{10}{4}+\frac{10}{4}}{3} = \frac{6}{3} = 2.0$

(By formula $\frac{\Sigma x^2}{N-1}$) within group variance $V_b = \frac{\frac{10}{5}+\frac{10}{5}+\frac{10}{5}}{3}$ =1.60 (By formula $\frac{\Sigma x^2}{N-1}$)

By both formula above both types of variance are following.

(i) By formula $\frac{\Sigma x^2}{N-1}$ (a). V_b = 2.0 (b) V_w =1.0

Hence, accordingly this formula

$$V_b + V_w = (2.0+1.0) = 3.00$$

(ii) By formula $\frac{\Sigma x^2}{N}$ (a). V_b = 1.60 (b) V_w = .666

Then accordingly this formula

$$V_b + V_w = 1.60 + .666 = 2.266$$

(iii) According to formula $\frac{\Sigma x^2}{N-1}$

Total variance (V_b) = 2.428

According to formula $\frac{\Sigma x^2}{N}$, V_b = 2.266

According to formula $\frac{\Sigma x^2}{N-1}$

Notes

$$V+(2.428)= V_b + V_w \ (3.0)$$

According to formula $\frac{\Sigma x^2}{N}$

$$V + (2.266) = V_b \ (1.60 + .666) = 2.266$$

On the basis of above calculation it can be seen clearly that using formula $\frac{\Sigma x^2}{N}$ total.

Variance (V_t) which is 2.266 inter group variance (V_b) and within group variance (V_w) which is 1.60 as .666 respectively, is exactly equal to sum. Therefore, it is proved that total variance is equal to sum of intergroup and within group variance. It can be expressed by formula as following.

$$V_t = V_b + V_w$$

But using formula $\frac{\Sigma x^2}{N-1}$ It is not found so.

There remains a little difference. In this condition V_t = 2.48 and $V_b + V_w$= 1+2= 3.0. This difference occurs due to that in formula $\frac{\Sigma x^2}{N-1}$ in place of denominator degrees of freedom (N-1) concept is used. Accordingly mathematics rule to do so is more scientific and satisfactory, because on the basis of this formula available value lies value of population's very close. Therefore this is the formula which is always used always use to get value. Here formula $\frac{\Sigma x^2}{N}$ is used only to clarified that V_t can be divided in two parts always into (V_b and V_w). V_b as SS_A and V_w as SS_w and V_t as SS_T can be written. Hence $SS_T = SS_A + SS_w$. Always this formula should be remembered.

4. **Between groups variance's (SS_A) explanation :** Between groups variance that variance which gets by each groups mean and three groups (if more than three, then all those) collective mean's differences squares divided by N-1. From example -2 it is clear. The occurrence of this variance's cause consider to be independent variables (their at different levels taken difference or various treatments)

Notes Between groups variance is that variance in which subvariable or vested independent variance affects and for the research of that experiment is done.

This variance dose not completely occur due to subvariable (independent variables). At least randomized error (by choice error), error lies. If variance is more than that error then it can be accepted due to treatment.

Within groups variance's (SS_W) explanation

This is that variance which occurred due to in separate groups lies persons or between units remain difference. This is correct that group's by randomized post studied variable controlled, try to make matched, rather than it no happen difference between units completely eliminated. Due to this reason within groups variance not complete Zero. This remain variance is called SS_W and it is considered to

Notes

be by chance. The characteristics is that it can be estimated, but cannot eliminate completely. Therefore ,it is named as "error".

F-Ratio

Example

On the basis of SS_A and SS_W the decision is taken over this matter that experimental effect e.g., independent variables or treatments are affected dependent variables or not.

In Anova the significance of difference of groups, their mutual comparison basis is this ratio. This ratio means SS_A/SS_w is called F-ratio. **Sir Ronald Fisher's** one pupil **Snedecor** as mark of their respect, named this ratio **F**.

If parametric tests concept fulfils and taken many randomized sample calculate F-ratio then their distribution will be normal, as in the assumption of mathematician. On this basis **Snedecor** has prepared F-value's a table taken different degrees of freedom which is called **F-Table**. This is present is the end of this book under appendix table-C. In this table given F-value is compared with obtained F-Value then conclusion is drawn. Behind this assessment the logic is as types. If make difference of treatment does not affect on dependent variable then the value of SS_A must be equal to SS_w. In that condition F's value will be 1.0

Thus if independent variables have effect's dependent variables really means that is cause of dependent variables the F's value should be greater than 1.0 but how greater is 1.0 then it can be said that independent variables or treatment's effect is significant or real?

The solution of this problem solved by given value of T- Table. Process is this type that on the basis of data firstly calculated F' Value. Then this F- value is compared with given value of table. If on the basis of data obtained f value table's at any level of 0.05 or 0.01 determines equal to the given value of decrees of freedom or more than that then null hypothesis is rejected and conclusion is drawn that treatment is effected on dependent variables really. Regarding F- value some other facts should be understand properly also. There are—

1. The rule to calculate F value is that between SS_A and SS_w which is greater that is put on place numerator and second on place of denominator anyone of both can be greater. Hence either of them can be put up and down, but in variance analysis condition SS_A Should be kept above always. If in any condition SS_A is less SS_w is more then there is no need to calculate F' value because then at once it is cleared that there is no effect of treatment on result mean F is not significant.
2. F – Test inevitably one – tailed test because regarding this it is to be considered that Did SS_A Value from SS_w as much as higher that it is called significant? For that SS_w is kept on the middle point (0.0) normal distribution curve and it is observed that how much above is SS_A Above curve's one end expand in one direction. Therefore it is an single test.
3. The concept of degrees of freedom in F- Test is most important. Its use is to select the value of table based value with which compares obtained F – value. For this F- ratio table C's understand through and through essential. In this table at the top in horizontal direction SS_A degrees of freedom (*df*) and left side in column vertically SS_w's df observed from both side move where meet at an point at that place two value available. In these one is light black colour and other dark black colour. Dark black colour holding means greater value level of 0.01 and light colour holding means small value level of 0.05 is. All these matters will be cleared by ahead given examples.

12.2 Normal Probability Group Design

Notes

Single-way Anova

In the previous chapters SS_t, SS_A, SS_w calculation had been shown but by which method they had been calculated actually they are not used. Actually which method is used that step is at follows:

1. Calculate Correction terms value (c)
2. Calculate SS_t
3. Calculate SS_A
4. Calculate SS_w
5. Convert SS_A and SS_w into mean
6. Calculate F – ratio
7. Compares it with table value
8. On the basis of significant test drawn conclusion

Example 3

Table 12.2: Single Way Answer Variance Analysis

X_1	X_2	X_3
6	4	6
2	2	5
3	5	4
5	1	4
4	3	6
20	15	25 =ΣX=60

(Having obtained measures on dependent variable for all the subjects in the different groups a table like the above is prepared. In this X is our independent variable which has three treatment conditions or levels, X_1, X_2, and X_3. The analysis of variance will follow the following steps):

1. Correction Term (C) = $(\Sigma X)^2/N = (60)^2/15 = 240.0$
2. SS_T = In Table-2 squares of total data's sum has been given
   $Ex^2 - c = (6^2 + 2^2 + \ldots 4^2 + 6^2) - 240.0 = 34.0$
3. $SS_A = \frac{(SX_1)^2}{N_1} + \frac{(SX_2)^2}{N_2} + \frac{(SX_3)^2}{N_3} = C$

If N is equal in all groups then aforesaid formula will be written as $\frac{\Sigma x_1^2 + \Sigma x_2^2 + \Sigma x_3^2}{N} - C$; Putting all value

$$SS_A = \frac{(20)^2 + (15)^2 + (25)^2}{5} - 240 = 10.0$$

4. $SS_W = SS_T - SS_A = 34.0 - 10.0 = 24.00$
5. Mean $SS_W = \frac{10}{2}$ = 5.0 as here three group are and *df* is (3 – 1) =2

Notes

6. Mean $SS_W = \frac{24}{12} = 2.00$ *df* will be 12.

7. F-ration $\frac{MSS_A}{MSS_W} = \frac{5}{2} = 2.50$

8. **Significance –** F value is not significant because at 2, 12 df and at level .05 table – C's F value is 3.88 Hence H_0 cannot be rejected Result is not significant.

9. **Treatment is not differ to one –** another and each have equal effects on depend variables. In other words independent variable X does not affect dependent variables.

Remarks- 1

Calculation of SS_w's degrees of freedom can be done by two types:

(i) From each group's N deduct 1 and adding all remainders as it is done in previous example (5 – 1) + (5 – 1) + (5 – 1) = 12 is done.

(ii) Second method is that from whole *df* group's *df* number deducted, as in previous example total *df* number is 15 – 1 = 14 and group's *df*'s number is 3 – 1 = 2. Hence *df* of SS_w = 14 – 2 means 12.

(iii) SS_A's degrees of freedom number acquired from deduction of 1 from total number of groups.

Remarks -2

It observed also that to calculate SS_t and SS_A Method is same which had clarified in example -1 and 2. Only it is difference that in that example this method is in form of mean and each's deviation whereas in this example is in based upon raw score's use. In deviation form that's formula was $\Sigma x^2/N_1$ but in fundamental numerical form this formula is $\Sigma x^2 - \frac{(\Sigma x)^2}{N}$

In which $\Sigma x^2/N_1$ is correction term (C). This C to calculate SS_A and SS_t remains same. Hence it is calculated once in the beginning because in this method fundamental digits are used. Hence the, meaning of SS_r is sum of square of total scores separately. Like wise the meaning of SS_A is sum of columns separately squares total.

Did You Know? SS_A/SS_w is called F- ratio also.

12.3 One–way Anova

In previous example taken independent groups one-way Anova's method had been explained. Here some method is explained in context of repeated measures design. In this condition only one group of persons is kept under treatment. As in pre-test and post-test design groups of a persons firstly and later is tested twice. Thus persons one or two variables make equal and after that distributed in groups then their measures. In these two types of management Anova (one-way) repeated measures design is used. Under this method which is calculated that is a little different to previous independent groups method. Repeated measures design's special utility vested in this matter by that between groups' unit lies occurred variance due to difference of that and that is called error of variance, it becomes less

Notes

and in this way research's accuracy and validity is increased. Therefore it is apowerful medium of controlling of heterogeneity of subjects, but it is not understood from that these difference, variances are completely ended. Some still remain, but the must be less, because in this units of one group is worked like controlled groups units.

A main demerits of this design is that in this units pre –test is effected on post –test. After pre test there is great change comes in units and they are not representative of whole population in which they are selected. Due to there two causes precision of experiments and external validity both has undesirable effected. Being there difficulties this design is considered to be superior to independent group design. It is called within group design or correlated group design. In this design a new sources of variance added which is called between subjects variance or SS_s which is eliminated from $SS_T = SS_A + SS_s + SS_w$ is called SS_{error} also. It will be cleared from ahead example and how these variance are calculated and how the conclusion is drawn.

Example 4

On recall remembering order's study of effects 8 students have been given 10 pair of words and after two days he has to recall it. After that another 10 pair of words is given for remember and after two days he has to say recalled it. Thus remembering, recall this work goes on for 6 schedule. On this basis which hypothesis had been to test that had "on recall there is no effects of remembering order". As is prejudice that after first schedule recall is the most and after that it becomes less gradually.

For the clarification/of data/ of test method data has been presented in following tale. For statistical computation and analysis including a little changing some steps have been followed which has been already mentioned in independent group case. This is as follows.

Table 12.3: Serial Number of Table

Subject	L-1	L-2	L-3	L-4	L-5	L-6	Total
S_1	7	3	2	2	1	1	16
S_2	4	8	3	8	1	2	26
S_3	7	6	3	1	5	4	26
S_4	8	6	1	0	2	0	17
S_5	7	2	3	0	1	3	16
S_6	6	3	3	1	1	1	15
S_7	4	2	0	0	0	0	6
S_8	6	7	5	1	3	2	24
Total Σ =	49	37	20	13	14	13	146

In this case for the fulfill of analysis we will follow the previous example of independent group design.

1. C $= (146)^2/48 = 444.08$
2. SST $= (7^2 + 4^2 .... (0)^2 + 2^2) - 444.08 = 740.0 - 444.08 = 295.92$
3. $SS_A = \dfrac{(49)^2 + (37)^2 + (20)^2 + (13)^2 + (14)^2 + (13)^2}{8} - 444.08$

$$= \frac{4704}{8} - 444.08 = 143.92$$

Notes

4. $SS_s = \frac{16^2 + 26^2 + 26^2 + 17^2 + 16^2 + 15^2 + 6^2 + 24^2}{6} - 444.08$

$= \frac{2990}{6} - 444.08 = 54.25$

5. $SS_{e \times r} = SS_T - (SS_A + SS_s)$

$= 295.92 - (143.92 + 54.25)$

$= 295.92 - 198.17 = 97.75$

In the a for said the meaning of $SS_{e \times r}$ is occurred variance by interaction between columns and rows. It is called interaction.

6. Mean SS_A = 14.92/5=28.78
7. Mean $SS_{e \times r}$ = 97.75/35=2.79
8. F-ratio = $\frac{28.78}{2.79} = 10.32$
9. **Conclusion:** In table C 3, 35 df and at level 0.01 F's value is 3.59 Obtained F-value 10.32 is which much greater than table value. Hence to avoid Ho the conclusion is drawn that memorizing order definitely effects on recall after first catalogue recall was at most. After that it become less gradually.

Computational Explanation

In the case of repeated measures design method of calculation is a litter different to that of independent groups design case. In this case each person is measured, second, third, fourth etc means more than one measure therefore between the scores of them correlation is occurred. In other words, between the pair of subjects also a new type of variance is occurred. Although in each group same persons are, but after a test a number of differences occurred in them and they are not same, as it was prior due to this cause a new type of variance is a added. Hence SST have three parts $SS_W + SS_A + SS_S$. Among these SS_W is understood as that variance which occurred due to persons one groups keeping different measuring conditions. It is called between rows variance SS_r or SS_a also Like SS_A is also a systematic variance SS_S. Between the scores of group as much as correlation exists, systematic variance is as much as also and variance of error is also less. Therefore SS_t or SS_r rows is computed. It has been cleared under example's Step -4.

Between independent group design and repeated measures design from statistical analysis point of view second difference is occurred in selection of error term means SS_w. In case of independent group design, this is SS_w, but in repeated measures design, it is residual variance which is called interaction variance e.g., $SS_{c \times r}$ also. In fact in case of repeated measures design calculation of SS_w is not possible, but keep in mind that both SS_w and SS_{res} are presents estimate of error, rather SS_{res} is considered to be a much better estimate. Hence in case of repeated measures design rather SS_w, SS_{res}'s calculation of F value uses. Its calculation as previous example's in step has shown from SS_r sum of and SSe is done after deduction of it therefore it is called residual or remainder variance.

12.4 Two-way Analysis of Variance (Independent Group)

The use of aforesaid motioned method as is explained in illustration, uses there, where only one group of unit measured on dependent variables again and again or used matched group, but it can be used

Notes

in that case also where to control any one variable. Its one example is given below. This method comes under the category of two–way analysis of variance.

Example 5

A researcher want to study about an eleven years old boy's home environment effect on his psychological development. He collects scores of five schools' boys' home environment and his psychological developments on home environment he prepares school wise boys' two groups (desirable and non-desirable home environment). School wise their on dependent variables (developments) scores in the form of mean he having arranged in the following table, by F- Test he want to analyze it. But he is doubled that various schools environment much differ to one another causes effects of school. Can stop to prevail of effects of home environment. Therefore it should be controlled. Hence he uses previous method by following type so that effects of school should be ended.

Table 12..4 : Home Environment and Psychology Relationship between development (N = 9 in each Cell)

School	Home environment		Total
	Low groups	High groups	
1	56	35	91
2	59	38	97
3	62	32	94
4	57	31	88
5	52	45	97
Means	286	181	467
	6.36	4.02	

It can be done exactly as it had been done in previous example, in that due to differences between subjects variance had been ended (differences between lines) observed this whole process as it is.

1. C (Correction term) = $(467)^2/10 = 21808.9$
   (Taken each groups mean scores as individual
2. $SS_T = (56^2 + 59^2 \ldots\ldots\ldots 31^2 + 45^2) - C =$ 1204.1
3. (Home environment) = $\frac{286^2 + 181^2}{5} - C$ =1102.5
4. SS_{row} (school) = $\frac{91^2 + 97^2 \ldots\ldots\ldots 88^2 + 97^2}{2} - C$
5. $SS_{res} = SS_T - (+ SS_{schools})$
   $= 1284.1 - (1102.5 + 30.6)$ = 151.0
6. Mean $SS_{(A)}$ = 1102.5/1, *df* being 1 = 1102.5

Notes

7. Mean SS_{res} = Mean $SS_{H\times S}$ = $\frac{151-0}{4}$ =37.75

8. F = $\frac{1102.5}{37.75}$ = 29.20.

9. **Conclusion –** in table C,1 4 *df* and at level .01 value of F is 21.27 where as researcher's obtained value is 29.20 which is greater than table value. Hence H_0 can be rejected and conclusion is drawn that home environment effects children's psychological development's

In this like previous written method SS_T is divided into three parts – $SS_A + SS_{school} + SS_{res}$. Due to differences of school occurred variances $(SS_s)_{row}$'s eliminated from SS_T. It is to keep in mind if the number of boys of school's is some then mean of scores assumes one by one scores can be applied in table. From computation point of view its result is not affected. This whole process is some as, as mentioned in example –4.

Error- Variance (SS_{res})

Under example -4 & example -5 which error of variance has been mentioned, it has many a name as a –remainder or residual variance and interaction variance, similarly it can be written as many types as a SS_{res}, SS_{exv} etc. this is that variance which due to treatments, remains not any difference exist and which is by by chance or randomized method selection of unit and distribution in groups remain differences between them occurred as assumed. Consequently this is that variance which form total variance, both treatments' (SS_A) and school's variance (SS_s) remain after deduction of those therefore it is considered to be more accurate than SS_w. If effects of differences of school's environment, effects of studied variables and effects of treatments eliminated completely then SS_{cxr}'s value will be equal to SS_w otherwise. Value of SS_{cxr} always less than SS_w. If it is less than that, then it should be understood as by chance and in that case SS_w should be used for F- value.

Self Assessment

Fill in the blanks:

1. F – Test is a ……. test.
2. The meaning of variance in statistical language ……
3. F- Test is definitely …… test.

12.5 Two – way Analysis of Variance (Correlated Group)

Two- way (independent group) variance analysis example has been presented in previous page (example 5), but such a case of research may be in which groups are correlated or matched. In this case also Anova can be used but its method is rather a little change with respect to previous method. Two –way analysis of variance's (Correlated group) this method's example has been given ahead.

Example -6

A researcher want to study a six-year old children effects of practice on learning. He controlled sex effects. Hence he choose 5 boys and six girls randomly and then practice trial's taken under. After every trial measures their improvements. Thus which data available at that basis ahead table has prepared and calculate necessary variances. Whole calculation has been shown below table.

Notes

(Two – way – variance)
Correlated measures Design
Process of Learning

Table 12 .5							
	S_S	I	II	III	IV	V	Total
	1	4	7	3	4	6	24
	2	3	4	6	9	7	29
Boys	3	2	2	8	5	8	25
	4	5	5	5	8	5	28
	5	4	6	7	6	9	32
Ex		18	24	29	32	35	138 (Group-I)
	1	6	3	4	8	10	31
	2	8	5	5	7	8	33
Girls	3	4	5	5	5	9	28
	4	3	3	6	6	10	28
	5	6	6	8	9	10	39
Ex		27	22	28	35	47	159 (Group-II)

In this analysis following steps are included-

1. C = total number's (is divided in) group I and group II sum of squares

$$= \frac{297^2}{50} = 1764.18$$

2. SS_t = 1991 – 1764.18 = 226.82
3. SS_t = Sum of squares the total of all the ten groups divided by n minus the correction

$$= 18^2 + 24^2 + \ldots\ldots\ldots + 35^2 + 47^2/5 - 1764.18$$

$$= \frac{9421}{5} - 1764.18 = 1889.2 - 1764.18 = 120.02$$

4. SS_t = Sum of squares of the totals of all the rows divided by n minus the correction term

$$= 242 + 29^2 + \ldots\ldots\ldots\ldots - 28^2 + 39^2/5 - 1764.18$$

$$= \frac{8989}{5} - 1764.18 = 1797.8 - 1764.18$$

$$= 33.62$$

5. SS_{Trials} = Sum of squares of the totals of all the trials (column sum of scores for boys and girls both for each trial divided by n minus correction).

$$= \frac{45^2 + 46^2 + 57^2 + 67^2 + 82^2}{10} - 1764.18$$

$$= 18603/10 - 1764.18$$

$$= 1860.3 - 1764.18 = 96.12$$

Notes

6. SS_G = Sum of squares between Groups of boys and girls

$$= \frac{138^2 + 159^2}{25} - 1764.18$$

$$= \frac{44325}{25} - 1764.18 = 1773 - 1764.18 = 8.82$$

7. SS_W = Within groups sum of squares.

This is obtained by subtracting between-groups variance from SS_8 (sum of squares among subjects).

= 33.62 – 8.82 = 24.8.

8. $SS_{S \times Trials}$ = This can be obtained by subtracting SS_s and SS_{Trials} from SS_t (total sum of squares).

226.82 – (33.62 + 96.12) = 97.08.

9. $SS_{Groups \times Trials}$

This we can get by subtracting SS_{Groups} (Sum of squares between groups) and SS_{Trials} from SS_{rr} as follows:

120.02 – (8.82 + 96.12) = 15.08.

10. Pooled $SS_{S \times Trials}$

This we can get by subtracting $SS_{Groups \times Trials}$ from $SS_{S \times Trials}$. It is used as the error term for determining the significance of treatments.

= 97.08 - 15.08 = 82.0

Table: 12.6				
	SS	*df*	*Mean SS*	*F*
Groups (Sex)	8.82	1	8.82	3.44 Sig. at .05 level
Within groups (error)	24.80	8	3.10	
Trials	96.12	4	24.03	9.386**
Groups × Trials	15.08	4	3.77	1.472 Not Sig.
Pooled $SS_{s\times Trials}$	82.00	32	2.56	
		49		

**Sig. at .01 level

Conclusion – In F table –C 4, 32 *df* and at level 0.01 F's value is 4.02 *wf* whereas exercise obtains F – value is 9.386 which is much greater than table value. Hence H_0 is rejected and conclusion is drawn practice necessarily effects on improvements. Sex's F-ratio (3.44) is significant at level .05. Hence it will be accepted that sex's effects on improvements means difference between boys and girls improvements, but interaction variance (1.472) is not significant on that basis it will be to say practice and sex when interaction with each other then there is no significant effects on improvements.

Two–way Anova's (correlated group) is used there where two independent variables are there and one group of units has been measured on dependent variables again and again. It can also be used where controlled effects of one variables to study of others effects.

12.6 Randomized Block Design

Notes

In the beginning of this chapter randomized group design had been described. In that units it was distributed in group by randomized method and it assumed that group was matched from true study point of view, but in fact they do not matched. In their units rather remain difference due to that accuracy of research becomes less units, specially among men a number of differences exists. In groups their randomized distribution is not able to eliminate completely these differences and diversities. As a result effects of treatment does not seem. Hence to less these differences and to increase accuracy of research and internal validity purpose researches is used on another design which is called "randomized block – design. In short it is written as RB–*k* design which means number group based randomized block design. By this design which data gets their variance analysis too accordingly given method of example – 5 in which in place of error do not use SS_w rather used SS_{exr}. In the given example ahead its whole process has been mentioned.

Example – 7

A researcher wanted to stay an eleven years old boy's learning two techniques of behaviour-control (admire and punishment) effect consequently wanted to control intelligence because intelligence is considered to be main factor of learning. So he formed blocks of 3-3 boys of equal. In one block boys I.Q were 119, 120, 121 respectively, in second I.Q. was 110, 111, 112 third 99, 100, 101, in fourth 90, 91, 92 in forth 81, 82, 83. From each block one boy's randomizely he has been distributed in three group (admired group, punished group, controlled group). Experiment goes on for one month. After that his learning has measured. Whole data had been adjusted in the table by following types. After that variance analysis has been done. Whole computation process is given in the table below.

Table 12.7 Randomized Block Design

Class	Prized Group	Punished Group	Controlled Group	Total
1	14	13	10	37
2	13	14	11	38
3	15	10	12	37
4	18	14	13	45
5	10	11	14	35
Sum	70	62	60	192

Now, the ANOVA was carried out. The following values were computed stepwise.

1. $C = (192)^2 / 15 = 2457.6$
2. $SS^T = 14^2 + 13^2 .... 13^2 + 14^2) - 2457.6$

   $= 2526 - 2457.6 = 68.4$

3. $SS_A = \dfrac{70^2 + 62^2 + 60^2}{5} - 2457.6 = 2468.8 - 2457.6 = 11.2$

4. $SS_{Blocks} = \dfrac{37^2 + 38^2 + 37^2 + 45^2 + 35^2}{3} - 2457.6$

   $= 2477.3 - 2457.6 = 19.7$

Notes

5. $SS_{exr} = SS_T - (SS_A + SS_{Blocks})$ = 68.4 – (11.2 + 19.73)

   = 68.4 - 30.93 = 37.47

6. Mean $SS_A = \frac{11.2}{2}$ = 5.6, *df* being 2

7. Mean $SS_{c \times r} = \frac{37.47.2}{8}$ = 4.68 *df* being 8(2 x 4, i.e., 2 for columns and 4 for rows)

8. F-ratio = $\frac{Mss_{(A)}}{Mss_{(C \times r)}} = \frac{5.6}{4.68}$ = 1.96.

9. **Conclusion –** obtained F-ratio (1.196) in table C 2, 8 *df* and at level 0.05 given value is less than 4.46. hence H_0 cannot be rejected. The conclusion is drawn that there is no difference in behavior controlling methods. He has no effects on learning.

Simply it can be known $SS_{black\ x\ treatment}$. Or remainder variance can never be greater then SS_w. Its size SS_{blocks} depends upon its magnitude. If it is greater than $SS_{c\ x\ r}$ result will be least.

Utility of RB–*k* design or advantage of this design

1. In this difference becomes least between units because made it matched and distributed in the block. Due to this reason error variance SSw is also less and experiment internal validity is increased.
2. Due to difference in block which variance occurred it is eliminated also. Hence due to this error become least.
3. This design is superior to independent group randomized design. Its efficiency can be calculated by formula $\frac{MeanSS_w}{Mean\ SS_{c \times r}} \times 100$. It is clear from that this design is superior to randomized design.
4. This design is very flexible. Taken a number of levels of treatments and blocks, it can be analyzed.
5. The simplicity of analysis is its additional characteristics.

Demerits → One demerit of this design is its inconvenience. When treatments are more in number then in due to larger number in block it is tough to make it similar. In one block such a larger number exists makes more unconveniciable and if it matched more than one variables then it becomes more and more tough.

12.7 Co-variance Analysis

In experimental research controlling of true study variables is considered to be most important, because in absence of that result can not be more accurate and valid. To attain this purpose a number of measures and methods has been mentioned. Among these one is analysis of covariance. In short it is called ANCOVA. This is an extended form of ANOVA. ANCOVA is a statistical method of controlling of variables. Due to this research's result' accuracy is increased. In previous example randomized block design had been described in which on intelligence matched units under treatment it had been distributed in this way intelligence had been controlled. In co-variance analysis does

Notes

not do so their measures can be done and by statistical analysis their effects can be eliminated. In variance of treatments (SS_{TT}) having corrected by statistical method on covariate intelligence, due to lies differences between units occurred effects can be eliminated. Later measures' regression effects on initial measures which is called between groups regression, is possible by medium of that. Thus co-variance analysis is such a statistical method in which various treatments' among distrinution test of significancy of differences is done as long as, when from that, that variance is eliminated which occurred among them in the begning due to true study variables or variables. In other words from between groups variance in covariates by initials differences occurred variances eliminated means after correction this analysis is done. Hence from $SS_{T(y)}$, $SS_{A(y)}$ and $SS_{w(y)}$ that variance is eliminated which occurred due to X (covariate) and Y (sub variable) correlation. Thus final analysis is done on the basis of these corrected variances. It will be cleared from next example.

In the following table suppose X variable is intelligence and Y boys' learning achievement. Three treatments groups (applaused, punished and controlled). Researcher wants to know that on boys' learning achievement these treatment effects significantly separately, but he wants to control intelligence effects whole data has been presented in the table and steps of analysis process has been given below.

Example 8

Table 12.8: Co-variance Analysis

Admired group			*Punished group*			*Controlled group*		
X_1	Y_1	X_1Y_1	X_2	Y_2	X_2Y_2	X_3	Y_3	X_3Y_3
33	18	594	34	31	1054	34	15	510
42	34	1428	55	45	2475	4	8	32
40	22	880	9	1	9	12	18	216
31	24	744	50	33	1650	16	15	240
Sum 146	98	3646	148	110	5188	66	56	998

1. First of all, we have to find out three correction terms for X, Y and XY:
   (a) C_x (correction term for X) $= (146+148+66)^2/12$
   $=(360)^2/12=10800.0$
   (b) C_y(correction term for Y) $=(98+110+56)^2/12 = 5808.00$
   (c) C_{xy} (correction term for XY) $= (\Sigma x)(\Sigma Y)/N$
   $= \frac{360 \times 264}{12} = 7920.00$
2. Now, total sum of squares has to be calculated for X, Y and XY:
   (a) $SS_{T(x)} = (33^2 + 42^2 .... 12^2 + 16^2) - C$
   $= 13748.0–10800.0$
   $= 2948.0$
   (b) $SS_{T(Y)} = (18_2 + 34_2 ..... 18_2 + 15_2) - C$
   $= 7454.0–5808$
   $= 1646.0$
   (c) $SS_{T(Xy)}$= Sum of XY total for each treatment group-C
   $= (3646+5188+998)–C$

Notes $= 9832 - 7920$

$= 1912.0$

3. Now, we have to calculate between –groups sum of squares for X, Y and XY as follows:

(a) $SS_{A(X)} = \frac{146^2 + 148^2 + 66^2}{4} - C$

$= 11894.0 - 10800.0 = 1094.0$

(b) $SS_{A(Y)} = \frac{98^2 + 110^2 + 56^2}{4} - C$

$= 6210.0 - 5808.0 = 402.0$

(c) $SS_{A(xy)} = \frac{X_1Y_1 + X_2Y_2 + X_3Y_3}{4} - C$

$= \frac{146x98 + 148x110 + 66x56}{4} - C$

$= 8571\text{-}7920$

$= 651.0$

4. Now, SS_W for X, Y and XY can be calculated by just subtracting SS_A from SS_r

a. $SS_{W(x)} = 2948.0 - 1094.0 = 1850.0$

b. $SS_{W(Y)} = 1646.0 - 402 = 1244.0$

c. $SS_{A(xy)} = 1912.0 - 651.0 = 1261.0$

Upto this point the calculations were done exactly in the same way as they were done in case of simple analysis of variance. After this, now adjusted values of $SS_{T(y)}$ and $SS_{w(y)}$ have to be calculated. It is here stage that, adjustment for initial differences on the covariate X has to be made. The procedure is as follows:

5. (a) Adj. $SS_{T(y)} = SS_{T(y)} - \frac{SS_{T(xy)}}{SS_{T(x)}} = 1646.0 - \frac{(1912)^2}{2948} = 405.93$

(b) Adj. $SS_{W(y)} = SS_{W(y} - \frac{SS_{W(xy)}}{SS_{W(x)}} = 1244.0 - \frac{(1261)^2}{1854} = 386.33$

(c) Adj $SS_{A(y)} = 405.93 - 386.33 = 19.6$

Adj. $SS_{A(y)}$ is calculated by subtracting Adj. $SS_{W(y)}$ from Adj. $SS_{T(y)}$

6. Next step is to convert Adj. $SS_{A(y)}$ and $SS_{W(y)}$ into mean squares by dividing them by their degrees of freedom as follow:

(a) Mean Adj. $SS_{A(y)} = \frac{19.6}{2} = 9.8$, *df* being 2.

(b) Mean $SS_{W(y)} = \frac{386.33}{8} = 48.29$, *df* being 8 as one more *df* is lost in case to ANCOVA as compared to ANOVA.

Notes

7. F-ratio = $\frac{\text{MeanAdj.SS}_{A(y)}}{\text{MeanAdj.SS}_{W(y)}} = \frac{9.8}{48.29}$ Not Sig.

8. **Conclusion**—H_0 cannot be rejected value of Mean Adj. $SS_{.A(y)}$ is less than mean Adj $SS_{w(y).}$ Hence there is no need to calculate. F- ratio. It will be less than that therefore it can not be significant. Therefore conclusion is drawn that after the controlling of intelligence no effects of treatments on boys' learning achievement.

ANCOVA Vested Conceptions

The use of ANCOVA is based upon those conditions and conceptions. That is.

Caution Treatment groups' should be selected from one population by randomization

1. Within groups corrected measures' distribution should be normal.
2. Treatment groups should be homogeneous.
3. Y measures' (Later) on X measures (initials) regression should be equal in each treatment group.
4. Regression should be linear.

Utility of ANCOVA

In the following conditions ANCOVA has much utility.

1. Where intact groups are used, as a in education field to use on any school's total class because it is not possible to exit separately only some boys out of class not possible to experiment on them. In that case it is not possible also unit by randomizedly distributed in groups, but from many point of view much differ to one another therefore differences between them affected result adversely. To eliminate it a couple of important variables can be controlled by medium of ANCOVA.
2. In the beginning if any variables between units being differences is not seem and later on it is necessary to eliminate their effects then use of ANCOVA proves to be useful-

Task Express your views on the utility of ANCOVA.

Multiple Co-variance Analysis

Above one covariate's example has been presented, but co-variance analysis can be done in that case also where more than one covariates are to be controlled. Upto two covariates calculation is simple, but more than one covariates are needed to be controlled then calculation is much long and tough, but there are no changing in rules. This method can be used with randomized block design also with

Notes

Latin groups design and factorial design can be used also. **Kirk** (1968) and **Lindquist** (1970) has explained it in more detail.

Self Assessment

Multiple Choice Questions:

4. In repeated measures designs only one persons ……..is kept in all treatments.

   (a) organization (b) Caste (c) groups (d) structure

5. By randomized block design ……is obtained

   (a) Rule (b) data (c) interactions (d) generalization

6. In experimental research true study variables ……is most essential.

   (a) Controlling (b) measuring (c) accuracy (d) result

7. ……..Design is superior to randomized block design

   (a) factorial design (b) Latin square (c) ANCOVA (d) ANOVA

12.8 L.S.D. (Latin Square Design)

This design is superior to randomize block design. In there two true study variables can be controlled. Like block design to make similar groups in these also, blocks are used, but in this on two true study variables make equal blocks. Thus each units of blocks at least equal on those two variables. The different levels of these true study variables are distributed in rows and columns. Thus in L.S.D gets information about effects at true study variables.

Concepts

The use of this design is only possible:

When blocking variables are equal to the number of treatments. Treatments are distributed in rows and columns by randomly. At the same time it keeps in mind too that each treatment should become in each column and row once. In each cell of row and column should become only one treatment.

An another condition of this design is also that the effects of treatments should be additive and they should not do interactions with rows and columns. Consequently rows and columns should not have inter action to each other. Below by on example its process of use has been explained.

Suppose that a researcher want to study the general activities of mice effects of one medicine five dose but he thinks that age of mice is also on component that affects their activites. Hence for the controlling of age he makes five block of equal age of mice–one in 5-6 months, second in 7-8 months, third in 9-10 months fourth in 11-12 months and fifth in 13-14 months. In each block keeps 5 mice. If experiment is such that in one day only 5 mice can be tested then he thinks that day to-day occurred changing effects result then he try to control it also keeping all those in minds to analyze of data he prepares a following types table. In this days above as headings of columns and blocks of ages left side from up to down wardes. After that taken on from each blocks of mice by randomly in five days prepares distribution separately. Aforesaid table takes form of this type:

Example 9

Notes

Table 12.9: Latin Square Design					
Age in Months	Mon.	Tues.	Wed.	Thur.	Fri.
13-14	1	3	4	2	5
11-12	5	4	3	1	2
9-10	2	5	1	4	3
7-8	4	2	5	3	1
5-6	3	1	2	5	4

After that treatment's (medicine) different doses' by randomization is imposed in each row and column, but the condition is that each treatment come only once in row and column. Thus distributed treatments. Which order prepares that is presented to ahead given table.

Table 12.9 (A): Treatment Facility					
Age in Months	Mon.	Tues.	Wed.	Thur.	Fri.
13-14	A	E	C	B	D
11-12	D	C	B	A	E
9-10	E	A	D	C	B
7-8	B	D	A	E	C
5-6	C	B	E	D	A

These A, B, C, D and e letters denote five levels of drug in which.

A = 10 mg, B = 15 mg,

C = 20 mg, D = 25 mg,

E = 30 mg,

Suppose the experiment was conducted and the researcher obtained following scores on the dependent variable.

Table 12.9 (B): Latin Square design data						
Age in Months	Mon.	Tues.	Wed.	Thur.	Fri.	Total
13–14	8	10	4	6	10	38
11–12	8	2	4	14	15	43
9–10	12	10	6	5	9	42
7–8	5	4	12	14	3	38
5–6	3	7	12	7	10	39
	36	33	38	46	47	200

The analysis of the data cast in this table is done as follows:

1. $C = \frac{\Sigma X^2}{N} = \frac{200^2}{25} = 1600.00$

Notes

2. SS_t = Sum of squares of all scores in the cells minus the correction term.

= 1948.0 – 1600.0 = 348.0

3. SS_{age} = Sum of squares of all the row totals divided by N minus the correction term.

$$= \frac{38^2 + 43^2 + 42^2 + 38^2 + 39^2}{5} - 1600$$

$$= \frac{8022}{5}$$

= 1600=4.4

4. SS_{Days} = Sum of squares of column totals divided by column n

$$= \frac{36^2 + 33^2 + 38^2 + 46^2 + 47^2}{5} - 1600$$

= 1630.8– 1600 = 30.8

We, now, have to find out the sum of all scores for A, B, C, D, and E with the helAp of table -1 – How to do it is demonstrated below:

A = 8 + 10 + 12 + 14 + 10 = 54, i.e., In A sum of all five columns datas

B = 5 + 7 + 4 + 6 + 9 = 31, i.e., In B sum of all datas of five columns

C = 3 + 2 + 4 + 5 + 3 = 17 i.e., In C sum of all datas of five columns

D = 8 + 4 + 6 + 7 + 10 = 35, i.e, In D sum of all datas of five columns

E = 12 + 10 + 12 + 14 + 15 = 63, i.e., In E sum of all dates of five columns

5. $$SS_{Tr} = \frac{54^2 + 31^2 + 17^2 + 35^2 + 63^2}{5} - 1600$$

$$= \frac{9360}{5} - 1600 = 1872 - 1600 = 272.0$$

6. $SS_W = SS_{TC} - (SS_{age} + SS_{days} + SS_{Tr})$

=348-(4.4+30.8+272.0)=348-307.2=40.8

Table 12.9 (C): Summary of Results

Sources of Variance	*SS*	*df*	*Mean SS*	*F*
Age	4.4	4	1.1	
Day	30.8	4	7.7	2.26
Drugs	272	4	68	20
Mistakes	40.8	12	3.4	

Thus, it is clear from obtained results that drugs' different doses effect the significance on the level of mice activities separately, because treatments' variances' difference at level .01 is significant. In the above example, treatments' F-value is (20.0) which at 4, 12 *df* at level 01 in relevant. Hence the conclusion is drawn that drugs affect activities of mice.

Uses of Latin Square Design

Notes

This design is much useful in that condition where researcher wants to control two true study variables effects. These variables are those which are related to dependent variable and if it does not control then result's accuracy and validity becomes doubtful. Hence by medium of blocking they are done inactive. In Biology and medicine science L.S.D has great importance because in these effects of differences lies between units are ended, drugs or independents variables' effects studies creatures' behaviour effects. Through medium of blocking these true studies' effects are become Zero in this design on these variables matched the unit keeps them in one block. Thus to make several blocks this design is used. For example if block is to be prepared on weight and size basis then prepare in this type.

Table 12.10

Weight	Small	Medium	High
40 – 50 kg.			
60 – 70 kg.			
80 – 90 kg.			

Thus weight and size different levels can be distributed in columns and rows and their effects can be eliminated from experiment

Its second utility is in treatment counter balancing as is shown in following table.

Table 12.11

	S-1	S-2	S-3
Group-1	a_3	a_1	a_2
Group-2	a_1	a_2	a_3
Group-3	a_2	a_3	a_1

It is clear from above table each treatment under counter is equally happened.

Comparison of error of perfect randomized groups and block design and L.S.D

Example

From the point of view of accuracy and validity of research. L.S.D is superior to both–perfect randomized groups design and randomized block design because size of error is less than both of them.

It can be cleared by as follows:

Perfect randomized group design error-perfect variance (SS_T) – treatment variance (SS_{TT}). Error in randomized block design error = $SS_{TT} - SS_{blocks}$ i.e., $SS_{Tr} - SS_{rows}$

But error in L.S.D = $SS_{Tr} - (SS_{rows} + SS_{columns})$

It is clear that in L.S.D., due to occurred error by SS_{column}, less on extra part it becomes less with respect to both of above.

Notes

Randomized Groups Independent Replication

In this design degrees of freedom becomes less because both columns and rows *df* from *df* of SS_w is to be subtracted. Due to this reason it become essential with more *df* SS_w calculated. To fulfill this purpose beside whole experiment taken Latin Square indecently replication should be done. These Latin Squares should be free from Pervious Squares and they should be selected by randomization. Including all those Latin squares doing analysis number of *df* increased and obtained the result more accurate. In this effect of treatments seems clearly.

12.9 Summary

- Like T-Test this test which is called F-test too, is a parametrical test. This is a most used test which has a a number of changed form and which can be used in many experimental conditions.
- The meaning of variance in literally is difference, but in statistical language its meaning is "average of squares of differences". These differences are taken frm scores and mean.
- Between group variances is that variances which each group mean and collective mean of three groups' differences squares gets by divided N-1.
- SS_A/SS_w is called F-ratio. **Sir Ronald Fisher's** one pupil **Senedecor** as a respect to him named this ratio.

12.10 Keywords

1. **Variance –** difference, under statistics obtained data-average of square of differences.
2. **Design –** Construction, Types form.

12.11 Review Questions

1. What do you mean by single – way variances analysis?
2. What is randomized groups design? Mention with examples?
3. Throw light on repeated measures design

   Or

   Illustrate single way ANOVA with example.
4. What do you mean by two way analysis of variance? Explained in detail.
5. What is called randomized block design? Illustrate with example
6. Explain Latin square design (L.S.D).

Answers: Self Assessment

1. Parametrical	2. Average of square of difference		3. One ended
4. (c) group	5. (b) data	6. (a) control	7. (b) Latin square

12.12 Further Readings

Books

1. Academic Technique and Assessment – *Dr. Rampal Singh, Bhatt Brothers*
2. Education Technique – *S. S. Mathur, Bhatt Brothers.*
3. Academic Technique Management and Assessment – *J C Agarwal, Bhatt Brothers.*
4. Education Technique- R. A. Sharma, Bhatt Brothers.