Data Analysis and Interpretation of Results

Unit 5 Contents

UNIT 5

New HRD Module on Action Research

Unit 5: Data Analysis and Interpretation of Results

Unit introduction

In unit 3 and 4 we looked in detail the various methods of data collection, sampling techniques and research tools. One very important aspect after collection of data is data analysis and interpretation of results. In unit 5, you will learn about various schemes of coding, organization of data, data entry and data cleaning. You will also learn about type of measurement and types of variables. The main emphasis in this unit is the analysis of data. Measures of central tendency and measures of variability are discussed in detail along with their interpretation. In order to study the relationship between any two variables a concept called "correlation" is introduced in this unit. Unit 5 ends with inferential statistics which basically include methods of comparison.

UNIT 5

Unit learning outcomes

When you have worked through this unit you should have advanced your ability to:

- be aware of the coding, organization, data entry and data cleaning methods
- distinguish between quantitative and qualitative measurements
- distinguish between discrete and continuous variables
- analyse data through descriptive statistics, correlation coefficient and inferential statistics.

Coding

Once the data has been collected from the field of enquiry, it must be subjected to its appropriate analysis by computer. When the data is being analysed, one can draw valid inferences based on results. The first step in data analysis is coding.

Coding is the process whereby data is transferred into a format suitable to be processed by a computer or even manually. It involves three different phases:

1. Organization of data
2. Coding and data entry
3. Data cleaning.

1. Organization of data

First of all, one should be familiar with the analytical software packages by which the data is to be analysed. Each software package has its own limitations and formats that can be handled by an analyzer.

At this stage, it is important that each respondent is given a serial number for identification. This helps in checking completeness of data files. Further, the data should be coded in the order it appears on the questionnaire so that symmetry of the data is maintained.

2. Coding and data entry

After the questionnaire has been designed, it is necessary to translate answers into numerical numbers in terms of codes. For example: if questions related to socio-economic status are present in the information then it has to be coded as follows:

Serial No.	☐☐	(Starting from 01to 50 if 50 individuals are present or three columns may be used, if more than 100 individuals are present i.e. 001 to 100 etc.
Age	☐☐	(in years)
Sex	☐	1. Male 2. Female
Marital status	☐	1. Married 2. Unmarried 3. Any other
Educational qualification	☐	1. Illiterate 2. 1 to 8th class 3. High school 4. Graduate 5. Post graduate & above
Locality	☐	1. Urban 2. Rural
Religion	☐	1. Hindu 2. Muslim 3. Sikh 4. Christian

UNIT 5

It should be noted that each individual of the study must be coded in the same sequence so that while entering the data age column is below age, sex column below sex etc. for all individuals. This will avoid ambiguity present in the data. Most of the time it is recommended that the number of response categories must be less than 10, so that these can be accommodated in a single column by a computer (note that in one column you can enter numbers from 0 to 9 i.e. one digit in one column). The coding format may look like:

Column No.	Variable
1 2	Serial No.
3 4	Age (in years)
5	Sex
6	Marital status
.	.
.	.
.	.

Here for serial number and age, we are using two columns and for other variables we are using only one column. Depending upon the type of the answer one has to decide whether to use one column, two columns or three columns. Let us take one more example to illustrate this. Suppose you want to enter income per month of the respondent that varies from 50 to 9999 dollars and rupees, then use four columns and enter exact income in the column 'code' as:

Column No.	Variable
11 12 13 14	1 0 0 0

Number of computer packages is available to enter the data. The most famous of these is Excel Worksheet in which coded data can be easily entered. In the Excel Worksheet the data entered looks like:

	Sr. No.	Age	Sex	Marital Status	Educational Qualification	Locality	Religion
1.	001	24	1	1	3	1	1
2.	002	29	2	3	4	1	3
3.	003	45	2	2	5	2	4
-	-	-	-	-	-	-	-
-	-	-	-	-	-	-	-
-	-	-	-	-	-	-	-

One may enter all the individuals in this Worksheet which specify age below age, sex below sex, locality below locality. This Worksheet can be used for analysis of variables present in the data.

3. Data cleaning

Data cleaning is the process of doing final check on the data file for accuracy, completeness and consistency before the analysis is undertaken. Once responses have been coded and the data is entered, it needs to be verified. This process is also called debugging. Checks can be made to verify coding range as well as relationship. For example: if code for sex is 1 or 2 and by mistake it has 3 in that column then it needs to be corrected. Relationship also needs verification e.g. if response is student, one would normally expect no income.

UNIT 5

Once the errors are found, the original source must be consulted and errors should be corrected. Checks should be run until the data is 100% correct.

After the data is cleaned it is ready for analysis. Some statistical methods can be applied to find out average, sum, relationship and other connections among various variables.

Before we subject data to statistical analysis, we should be familiar with type of measurement and types of variables.

Types of measurement

There are two types of measurement namely, quantitative and qualitative.

Quantitative measurement

The variables which can be measured by some suitable scale are called quantitative variables. For example: height, weight, blood pressure, head circumference or chest measurement etc.

Qualitative measurement

This measurement distinguishes one class of objects from another. Moreover, there is no measurement scale for these variables. For example: variables like honesty, intelligence or beauty etc. Similarly, one may distinguish Buddhist from Christian, Chinese from Indians and rural from urban. There is no underlying scale to measure these variables.

UNIT 2

Types of variables

Discrete variables

If a variable takes at most a countable number of isolated values, it is called a discrete variable. For example: number of members in a family or number of persons suffering from a particular disease.

Continuous variables

A variable is said to be continuous if it takes all possible values (depending upon the degree of accuracy required) between certain limits e.g. height of an individual or weight of an individual etc.

Self-help question 5.1

Specify which one is quantitative or qualitative measurement. Tick mark in the appropriate box.

		Quantitative	Qualitative
1.	Height	☐	☐
2.	Honesty	☐	☐
3.	Intelligence	☐	☐
4.	Blood pressure	☐	☐
5.	Sex	☐	☐
6.	Religion	☐	☐
7.	Locality	☐	☐
8.	Age	☐	☐
9.	Head circumference	☐	☐
10.	Chest measurement	☐	☐

Compare your answers with those provided at the end of the unit.

UNIT 5

? Self-help question 5.2

Specify which one is discrete or continuous variable. Tick mark in the appropriate box.

	Continuous	Discrete
1. No. of balls in a box	☐	☐
2. Atmospheric pressure	☐	☐
3. Members in a family	☐	☐
4. Rainfall in a particular region	☐	☐
5. Cancer cases in a given community	☐	☐

Compare your answers with those provided at the end of the unit.

Analysis

1. Descriptive statistics

Descriptive statistics is divided into two parts namely, measures of central tendency and measures of variability.

A. Measures of central tendency

It gives us an idea about the concentration of the values in the central part of the distribution. There are three main ways of describing the central value of a data viz. arithmetic mean, median and mode.

(i) Arithmetic mean – Arithmetic mean of a set of observations is their sum divided by the number of observations e.g. the arithmetic mean (A.M) of n observations $x_1, x_2, \ldots, x_n$ denoted by $\bar{x}$, is given by

$$\bar{x} = \frac{x_1 + x_2 + ... + x_n}{n} = \frac{1}{n}\sum_{i=1}^{n} x_i$$

where, $\sum$ is called summation sign.

In case of group frequency data or continuous frequency distribution, x can be taken as the mid-value of the corresponding class and if 'f' be the corresponding frequency, then

$$\bar{x} = \frac{f_1 x_1 + f_2 x_2 + ... + f_n x_n}{f_1 + f_2 + ...f_n} = \frac{\sum_{i=1}^{n} f_i x_i}{\sum_{i=1}^{n} f_i} = \frac{1}{N}\sum_{i=1}^{n} f_i x_i,\ N = \sum_{i=1}^{n} f_i$$

Example: (a) Find the arithmetic mean of the following frequency data.

X:	1	2	3	4	5	6	7
f:	5	9	12	17	14	10	6

(b) Calculate arithmetic mean of the marks from the following data:

Marks:	0-10	10-20	20-30	30-40	40-50	50-60
No. of Students (frequency)	12	18	27	20	17	6

Solution (a)

x	f	fx
1	5	5
2	9	18
3	12	36
4	17	68
5	14	70
6	10	60
7	6	42
Total	73=N	299 = $\sum$fx

UNIT 5

$$\bar{x} = \frac{1}{N}\sum_{i=1}^{7} f_i x_i = \frac{299}{73} = 4.09$$

Solution (b)

Marks	No. of Students (f)	Mid-point (x)	fx
0-10	12	5	60
10-20	18	15	270
20-30	27	25	675
30-40	20	35	700
40-50	17	45	765
50-60	6	55	330
Total	$\sum_{i=1}^{n} f_i = N = 100$		$\sum_{i=1}^{n} f_i x_i = 2800$

$$\bar{x} = \frac{1}{N}\sum_{i=1}^{7} f_i x_i = \frac{2800}{100} = 28$$

(ii) Median – Median of a distribution in the value of the variable which divides it into two equal parts. It is the value which exceeds and is exceeded by the same number of observations.

In case of simple data, if the no. of observation is odd then median is the middle value after the observations have been arranged in ascending or descending order of magnitude. In case of even number of observations, there are two middle terms and median is obtained by taking the arithmetic mean of the middle terms. For example: the median of the values 25, 20, 15, 35, 18 i.e. 15, 18, 20, 25, 35 is 20 and the median of 8, 20, 50, 25, 15, 30 i.e. 8, 15, 20, 25, 30, 50 is $\frac{1}{2}$ (20+ 25) = 22.5

In case of frequency data, as mentioned in above example part (a) the median can be determined as follows:

Step I: Compute $\frac{N}{2}$

Step II: See the cumulative frequency greater than $\frac{N}{2}$

Step III: The corresponding value of x is median

Consider part (a) data

X	f	Cumulative frequency*
1	5	5
2	9	14
3	12	26
4	← 17	—— 43
5	14	57
6	10	67
7	6	73
	N= 73	

*(Cumulative frequency is obtained by adding successive frequencies till the last frequency)

In this example, $\frac{N}{2} = \frac{73}{2} = 36.5$

Cumulative frequency more than 36.5 is 43.

The corresponding value of x i.e. 4 is the median.

In case of continuous frequency distribution, the class corresponding to the cumulative frequency just greater than $\frac{N}{2}$ is called the median class and the

UNIT 5

value of the median is obtained by the following formula.

$$\text{Median} = l + \frac{h}{f}\left(\frac{N}{2} - C\right)$$

Where,
l = lower limit of median class
f = frequency of the median class
h = the magnitude of the median class
C = cumulative frequency (c.f.) preceding the median class.

Consider example, part (b)

Marks	Mid-point	Frequency frequency	Cumulative
	(x)	(f)	(c.f)
0-10	05	12	12
10-20	15	18	30=C
20-30	25	27	57 Median class
30-40	35	20	77
40-50	45	17	94
50-60	55	06	100
		$\sum f_i = N = 100$	

Here $\frac{N}{2}$ = 50, cumulative frequency greater than 50 is 57 and the corresponding class is median class is 20-30.

Hence, using the formula

l = 20

f = 25

h = 10

C = 30

$$\text{Median} = l + \frac{h}{f}\left(\frac{N}{2} - C\right)$$

$$= 20 + \frac{10}{25}\left(\frac{100}{2} - 30\right)$$

$$= 20 + \frac{2}{5}(50 - 30)$$

$$= 20 + 8$$

$$= 28$$

(iii) Mode – Consider the following statements:

(i) The average height of an Indian (male) is 5 feet 7 inches

(ii) The average size of the shoes sold in a shop is 7

(iii) The average amount spend by a student in a hostel is Rs. 1500.

In all the above cases, the average referred to mode. Mode is the value which occurs most frequently in a set of observations and around which other items of the set cluster densely. In other words, mode is the value which is predominant in the series.

Thus, in case of frequency distribution, mode is the value of x corresponding to maximum frequency. For example: consider part (a) of above example:

x:	1	2	3	4	5	6	7
f:	5	9	12	17	14	10	6

The value corresponding to maximum frequency viz. 17 is 4. Hence mode is 4.

UNIT 5

Activity 5.1

Calculate arithmetic mean and median from the following frequency distribution of marks obtained by students in Mathematics test.

Marks	No. of students
05	20
10	43
15	75
20	67
25	72
30	45
35	39
40	09
45	08
50	06

Activity 5.2

The following table gives the monthly income of twelve families in a town:

Sr. No.	Monthly income (Rs.)	Sr. No.	Monthly income (Rs.)
1	280	7	80
2	180	8	84
3	96	9	100
4	98	10	75
5	104	11	600
6	75	12	200

Calculate the arithmetic average, media and mode of the above incomes. What average would represent the above series best?

B. Measures of variability

Although averages or measures of central tendency gives us an idea of the concentration of observation about the central part of the distribution, however, if we know the average alone we cannot form a complete idea about the distribution. For example: Consider the series (i) 7, 8, 9, 10, 11

(ii) 3, 6, 9, 12, 15

(iii) 1, 5, 9, 13, 17

In all, these cases we see that number of observation is 5 and mean $\bar{x} = 9$

—10 (A)

Thus, if we are given that the mean of 5 observations is 9, we cannot form an idea as to whether it is the average of first series, second series or the third series. Therefore, we see that measures of central tendency are inadequate to give us a complete idea of the distribution. They may be supplemented by some other measures. One such measure is variability or dispersion. Variability means scatteredness. We study dispersion to have an idea about the homogeneity or heterogeneity of the distribution. In the above cases we see that series (i) is more homogeneous (less dispersed) than series (ii) or (iii) or we say that series (iii) is more heterogeneous (i.e. more scattered) than the series (i) or (ii).

There are three measures of dispersion (i) range (ii) variance (iii) standard deviation.

(i) Range – The range is the difference between two extreme observations of the distribution. If A & B are the greatest and smallest observations respectively, then its range is A – B. Since it is based on two extreme observations which themselves are subject to chance fluctuations, it is not at all a reliable measure of dispersion. Thus, range is a very crude measure of dispersion.

(ii) The Variance – The arithmetic mean of the squares of the deviations of the given values from their mean, is called variable. For example: if the mean of observations is $\bar{x}$, then variance, denoted by s^2 , is defined σ^2as:

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

—10(B)

where $\bar{x} = \frac{\sum_{i=1}^{n} x_i}{n}$

If we are dealing with frequency data , then

x	f	fx	$fx^2.x_1$
x_1	f_1	$f_1 x_1$	$f_1 x_1^2$
x_2	f_2	$f_2 x_2$	$f_2 x_2^2$
.	.	.	.
.	.	.	.
x_n	f_n	$f_n x_n$	$f_n x_n^2$
Total	$n = \sum_{i=1}^{n} f_i$	$\sum f_i x_i$	$\sum_{i=1}^{n} f_i x_i^2$

$$\bar{x} = \frac{1}{n} \sum f_i x_i$$

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

which can be simplified to

$$\sigma^2 = \frac{1}{n} \sum_{i=1}^{n} f_i x_i^2 - \left(\frac{1}{n} \sum_{i=1}^{n} f_i x_i\right)^2$$

UNIT 5

The estimate of σ^2 from sample values is denoted by s^2.

Since variance is based on all the observations and gives equal weightage to every observation and hence from statistical point of view, it is the most appropriate measure of variability.

(iii) Standard deviation – The square root of variance is called standard deviation and is denoted by σ and

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

Also for frequency data

$$\sigma = \sqrt{\frac{1}{n}\sum_{i=1}^{n} f_i\left(x_i - \bar{x}\right)^2} \quad , \quad \bar{x} = \frac{\sum f_i x_i}{n}$$

Here $n = \sum_{i=1}^{n} f_i$

The estimate of σ from sample values is denoted by 's'.

Activity 5.3

Following data show the distribution of the weights of .22 persons:

Weight (in lb)	No. of persons
100 - 110	4
110 - 120	6
120 - 130	20
130 - 140	32
140 - 150	33
150 - 160	17
160 - 170	8
170 - 180	2

From the above data, compute

i. Mean and Median
ii. Variance
iii. Standard Derivation

Why is standard derivation the most appropriate measure of variability?

2. Inferential statistics

Most social sciences, medical sciences, agriculturists, psychologists and industrialists are interested to compare differences among two or more groups. For example: one may be interested in testing the differences in the gain in weight of two random samples of rats fed on two different diets A and B. Similarly, one may be interested in testing significant differences between levels of education of private school and government school. In industry, the life lengths of electronic components manufactured by two different units can be compared. Statistically tests help us to decide this.

In order to test such differences we generally compute student's t-statistic and the computed value of this statistics can be compared with tabulated values (available in statistical tables) to know whether the differences are significant or non-significant. The concept of p-value, obtained through computer packages also decides whether differences are significant or not. The smaller the p-value greater is the evidence of significance.

Suppose, we have two independent random samples *i.e.* $x_1, x_2, \ldots x_{n1}$; *and* $y_1, y_2, \ldots y_{n2}$ of sizes n_1 and n_2 drawn from two different populations. Let the observations from group 1 has mean $\bar{x}$ and variance s_1^2 and from group 2 has mean $\bar{y}$ and variance s_2^2.

	Group-1	Group-2
Sample size	n_1	n_2
Mean	$\bar{x}$	$\bar{y}$
Sample variance	s_1^2	s_2^2

Here

$$\bar{x} = \frac{1}{n_1}\sum_{i=1}^{n_1} x_i ; \bar{y} = \frac{1}{n_2}\sum_{i=1}^{n_2} y_i$$

$$s_1^2=\frac{1}{n_1-1}\sum_{i=1}^{n_1}(x_i-\bar{x})^2,\quad s_2^2=\frac{1}{n_2-1}\sum_{i=1}^{n_2}(y_i-\bar{y})^2$$

Then the differences in the means of the two groups can be compared by using t statistics given below:

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

, which has t-distribution with (n_1+n_2-2) degrees of freedom, where

$$S^2=\frac{(n_1-1)s_1^2+(n_2-1)s_2^2}{n_1+n_2-2}$$

Calculated value of t- statistics can be compared with tabulated value (given in statistical tables) for various degrees of freedom and at a particular level of significance. If calculated value is less than tabulated, differences are non-significant, otherwise they are significant. Statistical tables for t– statistics are available in standard text books of statistics for various degrees of freedom and at a particular level of significance. The statistical packages always give results in terms of p-value.

Example: Samples of two types of electric light bulbs were tested for length of life and following data were obtained:

	Type-1	Type-2
Sample size	$n_1=8$	$n_2=7$
Sample mean	$\bar{x}_1=1234$ hrs	$\bar{x}_2=1036$ hrs
Sample variance	$s_1^2=1296$	$s_2^2=1600$

UNIT 5

Do you think that mean differences in life lengths sufficiently warranty that type 1 is superior to type 2?

Computation of test statistics:

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

Now, by using above data

$$S^2 = \frac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2}{n_1 + n_2 - 2} = 1659.08$$

and t = 9.39

Tabulated value of 't' for $(n_1 + n_2 - 2) = 13$ degree of freedom and at 1% level of significance is 1.77 (from t-value of table).

Since calculated value of 't' is much greater than tabulated value and hence difference is significant. Thus, we conclude that Type I is definitely superior to Type II.

Activity 5.4

Determination of pH levels was made on two independent random samples of seventh grade school children. Sample A children were caries-free while sample B children had a high incidence of caries. The result follows:

A: 7.14 7.11 7.61 7.98 7.21 7.16 7.89 7.24 7.86 7.47 7.62

B: 7.36 7.04 7.19 7.10 7.15 7.36 7.57 7.64 7.00 7.25 7.19

Test whether pH levels of caries-free children on an average were higher than those for having high incidence of caries? Let level of significance =.05

There are other statistical methods where one may compare more than two groups, like analysis of variance (ANOVA), if one is dealing with quantitative data, and chi-square method if one is dealing with frequency data. However, these advanced topics are beyond the scope of this unit.

Correlation coefficient

Sometimes, we may be interested to find out if there is any correlation or relationship between the two variables under study. For example: if we measure the height and weight of a certain group of persons, we shall be interested to find out how heights and weights are related.

If the change in one variable affects a change in other variable, the variables are said to be correlated. If the two variables deviate in same direction i.e. if the increase(or decrease) in one results in a corresponding increase (or decrease) in the other, correlation is said to be positive. But if they constantly deviate in the opposite directions i.e. if increase (or decrease) in one results in corresponding decrease (or increase) in the other, correlation is said to be negative.

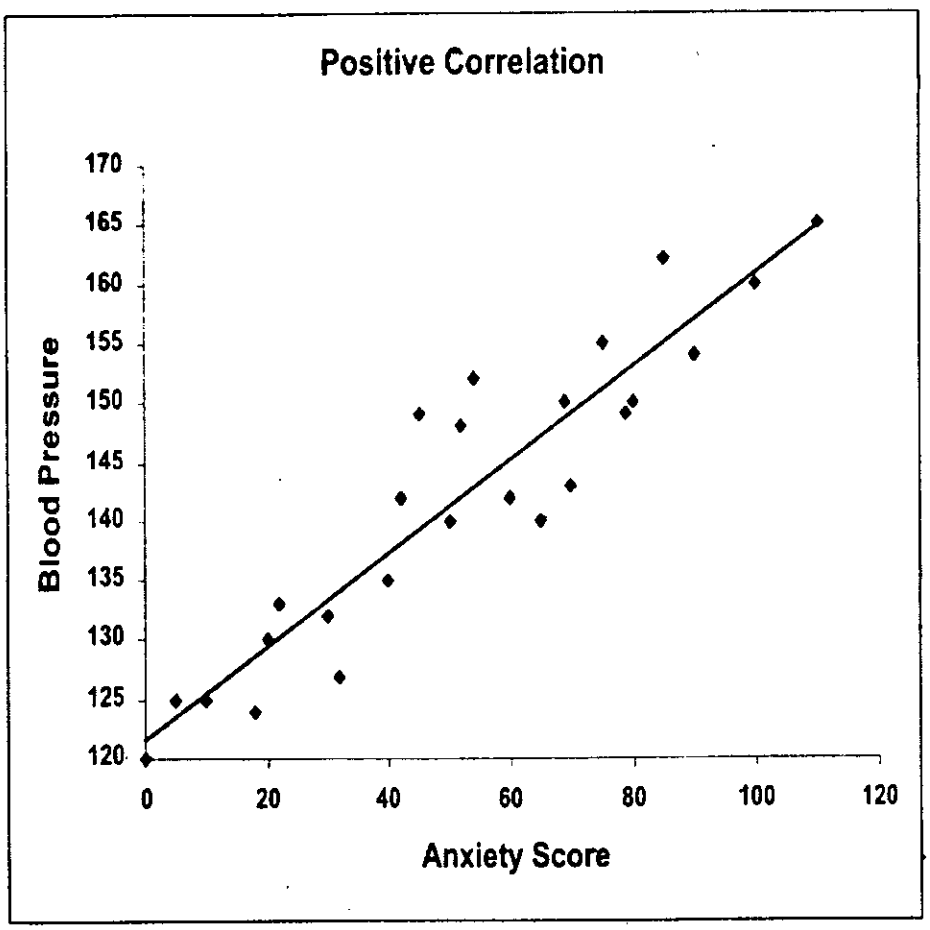
The examples of positive correlation are (i) the height and weights of a group of persons (ii) the income and expenditure. Similarly, the example of negative correlation are (i) price and demand of a commodity (ii) the volume and pressure of a gas.

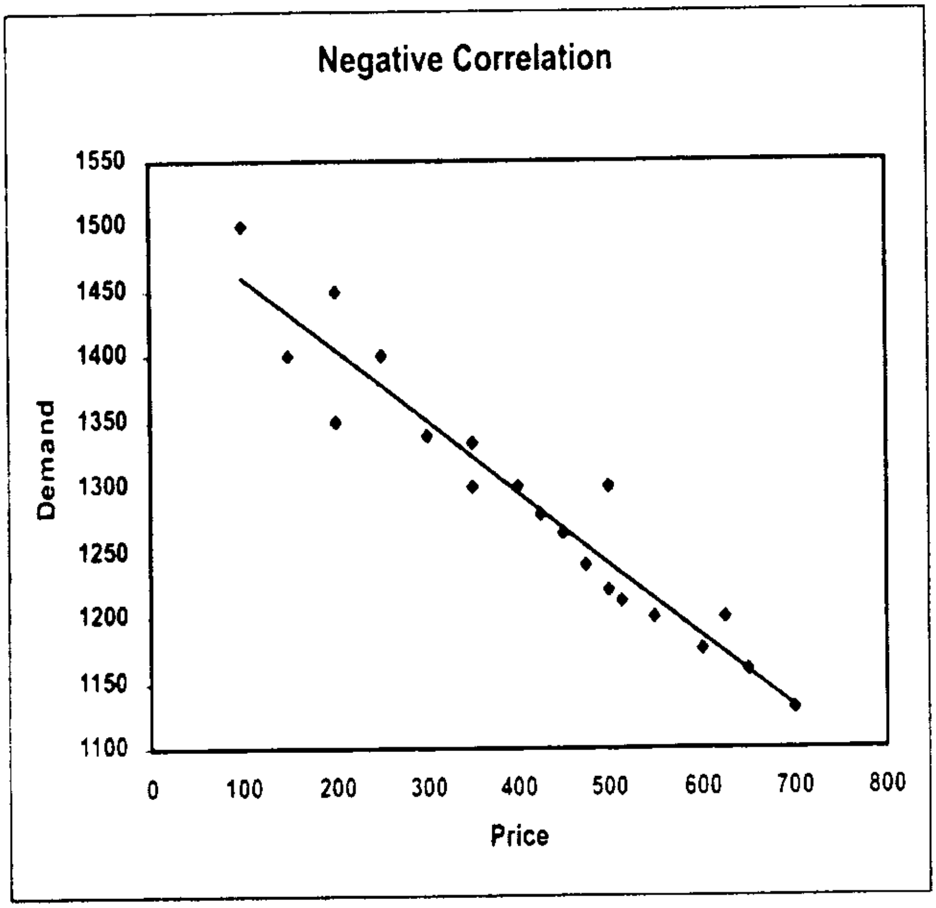
The correlation is said to be perfect (positive or negative) if the deviation in one variable is followed by the corresponding and proportional deviation in the other.

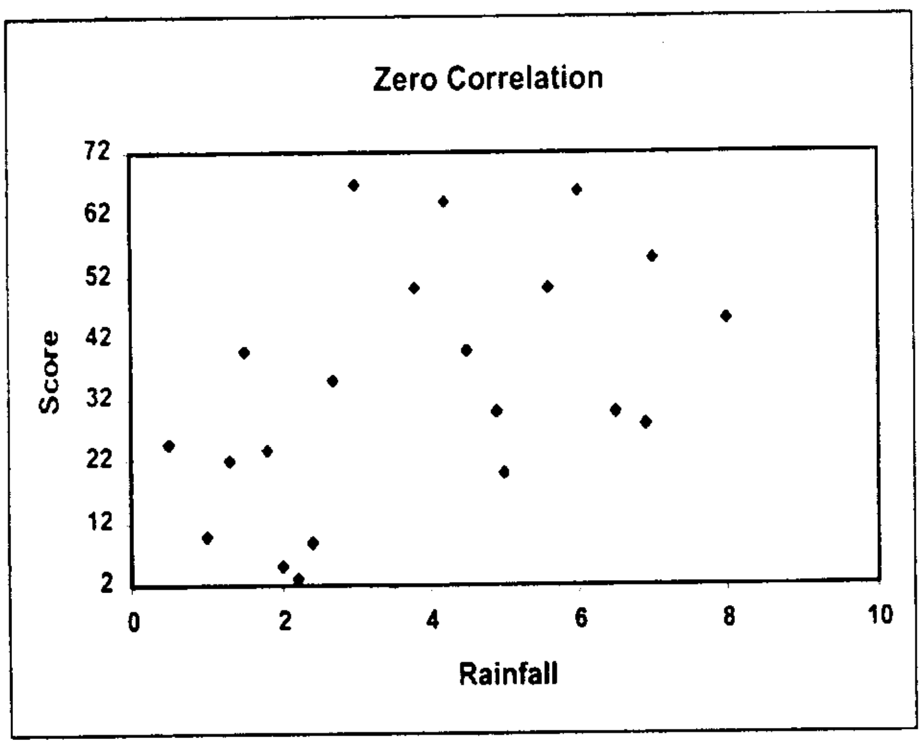
UNIT 5

If the changes in one variable do not affect the value of other, there will be "zero" correlation between two variables.

Scatter diagram is the simplest way of judging the correlation between two variables. If the values of variables X and Y be plotted along the X-axis and Y-axis respectively, the diagram of dots so obtained is known as scatter diagram. If the points are very dense i.e. very close to each other, we should expect a fairly good amount of correlation between the variables and if the points are widely scattered, a poor correlation is expected. We present below some pictures of positive, negative or no correlation between two variables.




Negative Correlation
Demand
1550
1500
1450
1400
1350
1300
1250
1200
1150
1100
0
100
200
300
400
500
600
700
800
Price


Zero Correlation
Score
72
62
52
42
32
22
12
2
0
2
4
6
8
10
Rainfall

UNIT 5

When the number of observations is fairly large, it becomes cumbersome to use method of scatter diagram.

Mathematical measure of correlation

Suppose we have two variables with n values say, $x_1, x_2, \ldots\ldots x_n$ and $y_1, y_2, \ldots\ldots y_n$.

From the following tables, the correlation coefficient, denoted by "r" can be easily computed

	x_i	y_i	$x_i y_i$	x_i^2	y_i^2
	x_1	y_1	$x_1 y_1$	x_1^2	y_1^2
	x_2	y_2	$x_2 y_2$	x_2^2	y_2^2
	.	.	.	.	.
	.	.	.	.	.
	.	.	.	.	.
	x_n	y_n	$x_n y_n$	x_n^2	y_n^2
Total	$\sum x_i$	$\sum y_i$	$\sum x_i y_i$	$\sum x_i^2$	$\sum y_i^2$

Mathematically, the values of correlation coefficient "r" is given by

$$r = \frac{n\sum x_i y_i - (\sum x_i)(\sum y_i)}{\sqrt{n\sum x_i^2 - (\sum x_i)^2}\sqrt{n\sum y_i^2 - (\sum y_i)^2}}$$

The value of "r"always lies between –1 to +1. If r= +1, the correlation is said to be perfect and positive and if r =-1, then correlation is said to be perfect and negative.

If the value of "r" is close to "zero", then the poor correlation is expected between two variables.

It may be noted that "r" is a measure of only linear relationship (i.e. points lies along a straight line) between two variables, r = 0 simply implies the absence of any linear relationship between the variables X and Y. However, there may exist some other form of relationship between them e.g. quadratic, cubic etc.

Activity 5.5

The following data relate to stature (x) and sitting height (y) both in cm, for each of 10 people of a particular Indian Caste.

(X)	(Y)
172.8	83.9
166.0	83.6
164.1	81.3
164.4	85.4
168.8	83.9
165.2	81.1
[illegible]	84.6
163.5	81.1
169.4	84.9
159.1	79.6

Represent the data by means of scatter diagram and compute the correlation coefficient between (X) and (Y). How will you interpret the amount of correlation?

UNIT 5

Unit summary

In this unit, you have covered the following main points:

- Coding and organization of data
- Data entry by using computer packages
- Debugging procedure
- Types of measurement
  - Quantitative
  - Qualitative
- Types of variables
  - Discrete variable
  - Continuous variable
- Data analysis
  - Using descriptive statistics
  - Using variability
  - Using measure of association
  - Using inferential statistics

Answers to self-help questions

Self-help question 5.1

1. Quantitative
2. Qualitative
3. Qualitative
4. Quantitative
5. Qualitative
6. Qualitative
7. Qualitative
8. Quantitative
9. Quantitative
10. Quantitative

Self-help question 5.2

1. Discrete
2. Continuous
3. Discrete
4. Continuous
5. Discrete

UNIT 5

Module Summary

The aim of this module is to help you to gain confidence, skills and determination required for preparing an action research project. Note that you have completed this module "New HRD Module on Action Research", you should have a good idea in detail and techniques used in action research and an understanding of its underlying philosophy and principles. You should also be knowledgeable and enrich yourself about various sources of information for data collection, various available sampling techniques and research tools, monitoring and evaluation of projects through data analysis and interpretation. We sincerely hope that the important concept of project impact should be clear to you after working through this module. We also hope that you should have learned and reflected on your own accountability and teachability as a project manager or participant.

Now, after going through the entire module you should :

1. be able to develop and understand the meaning, need and importance of research
2. be familiar with principles of action research
3. have a detailed knowledge of action research
4. be able to formulate action research design
5. be aware of the different sources of gathering information

6. be aware of sampling techniques and research tools
7. plan related projects on the basis of need analysis
8. create systems for monitoring the progress of projects through data analysis and interpretation of results
9. develop appropriate methods for evaluating projects.

We hope you have found your study of this module enjoyable.