ANALYSIS OF DATA

OBJECTIVES

After reading this material and performing the activities, you will be able to

- ③ Group the data you have collected to facilitate statistical analysis
- ③ Use some simple statistical tools for analyzing the data.
- ③ Interpret the data/out come of the analysis
- ③ Verify the action hypothesis

OVERVIEW

Whenever we conduct an activity we make some observations and see if our effort has been fruitful. If not, we undertake further work until we get the desired result. This is facilitated by the analysis of data. We arrive at some information as a result of the observations made, tests conducted in the course of the action research project. These data are the indicators of the success of a treatment given or its inadequacy. To interpret the data we should know systematic ways of analyzing it. Statistical tools allow an investigator to test a hypothesis objectively. The hypothesis is either accepted or rejected based on the analysis of the scores and not based on opinions.

In this chapter you will be introduced to some simple statistical tools to help you analyse the data and interpret it. Please note that what we present here is not a complete treatise on statistical analysis. You are just being initiated into the world of systematic analysis of the information possible. This is done so because the analysis you gather in the course of your project. You could read a detailed treatment of these topics in any book on educational statistics. Here we have made an attempt to explain what statistical measures express. We have tried to present the ideas in simple prose and avoided technicalities as far as possible. This is done so because the analysis of data can be done through computers in modem times. What is needed is an ability to interpret the results and make decisions about the outcome of the experiment conducted.

WHAT IS STATISTICS?

When we undertake action research we also decide on the nature of data we are going to collect. Sometimes we may arrive at qualitative data in the form

of anecdotal records or notes of an interview. We will have to study such information and describe our conclusions at length while reporting in such cases.

We may also design our tools in such a way that we get a set of scores. Data is the word used for the observations, records, test scores or field notes and the like made during the implementation of the action plan. All this information can be organized, tabulated or arranged in a meaningful way to help the investigator answer the questions raised. The organized scores are ready for statistical analysis.

Statistics is a body of mathematical techniques or processes for gathering, organizing, analyzing and interpreting numerical data.

UNGROUPED AND GROUPED DATA

When a researcher gathers information on the outcome of his/her project in the form of numerical data or scores, they are not grouped. The scores are just collected. Look at the following.

Given below is a table showing scores of 25 class III students on a mathematics test. The scores here are arranged in the descending order.

98	88	85	80	78
97	87	84	80	78
95	87	82	80	78
93	85	82	80	77
90	85	82	80	76

The same data can be arranged in an organised form at the scores are more clearly presented when they are grouped in frequency tables with class intervals. Now look at the following table showing the scores of 25 Class III students on a Mathematics test. The scores are grouped in intervals of three.

Class Interval	Tallies	Frequency (f)
97-99	II	2
94-96	I	1
91-93	I	1
88-90	II	2
85-87	IIII	5

82-84	IIII	4
79-81	IIIII	5
76-78	IIIII	5
Total		**25**

The scores arranged in class intervals immediately gives you an idea about the concentration or spread of the scores. One can see that most of the scores fall below 87. Similarly these scores could be arranged in class intervals of five. If you have a very large set of scores, you could even go for higher class intervals. We need to group the scores that we obtain in this fashion to facilitate the calculation of different statistical measures. You will know more about the usefulness of such an organization when you learn about the methods of calculating different measures statistics.

Activity

Arrange the above data in intervals of five. Discuss with your colleagues how these data appear different from the earlier one.

STATISTICAL MEASURES

You saw some scores in the above distributions. What do these scores mean to you? Can you tell anything about the performance of the students by looking at these scores? Would you say the performance has been average or below average or above average? What are the criteria for making such judgments? What is the average performance of the group?

1. Measures of central tendency

2. Measures of variability

3. Significance of the difference between means

MEASURES OF CENTRAL TENDENCY

Let us suppose that you conducted a test in Mathematics after giving some special inputs to below averages in your class. The following are the scores obtained by 9 students.

3 5 5 6 8 9 11 12 15

What do these scores indicate? What do you think is the general performance of the group? Would you say that the performance has been average or below

average? What are the criteria for making such judgments? What is the average or the **mean** performance of the group?

The total of the 9 scores is 74 when divided by 9 given an average 8.22. such averages tell us about the general performance of the group.

Now assume that one score among them is 60 instead of 15.

3 5 5 6 8 9 11 12 60

What will happen to the average?

The average will be 13. One extreme score has influenced the average to a great extent. We can not say that it is the most representative score or the general performance of the group. In such situation we can think of considering the *exact mid-point of the score or the* **median.**

The median for the ungrouped data is = N+1 divided by 2. (9+1)/2=5. This means that the 5th score of the group happens to be the exact mid point. This is called the median. The median for the above score is 8.8 as a measure of the central tendency of the group is more representative than 13. It is nearer to the general performance of the group.

Sometimes we would like to know the typical score of the group. In other words which score repeats again and again? It is 5 that repeat twice. This is the point at which more measures fall than at any other point. This is called the **mode**

- ③ It is the most useful among the measures of central tendency, .
- ③ It is the most representative score,
- ③ It is that score which speaks of the performance of the whole group,
- ③ It is used in the calculation of other measures like the standard deviation.

CALCULATING THE MEAN BY THE SHORT METHOD FOR GROUPED DATA

In this method of calculating mean, the investigator 'guesses' or 'assumes' a mean at the outset and later applies a correction to this assumed mean (AM) in order to obtain the actual mean(M).

There is no set rule for assuming a mean. The best way is to take the mid point of the Class Interval with the largest frequency.

Table showing the scores of 25 students of class III on a mathematics test

Class Interval	Mid Point X	F	x′	Fx′
76-80	78	10	0	0
81-85	83	7	+1	7
86-90	88	4	+2	8
91-95	93	2	+3	6
96-100	98	2	+4	8
		N = 25		Σfx′=29

The steps are as follows

③ Write the mid point for each class interval in the frequency distribution

③ Write the frequency of 'f' for each interval

③ Assume the mean in an interval with the largest score and write the distance from the AM for each CI in units of class interval in the fourth column-x'

③ Now fill up the 5th column for each CI with the x' multiplied by f

③ Write N and Σfx'

③ Calculate the mean by using the following formula.

MEASURES OF VARIABILITY:

A given set of scores ranges from the lowest to the highest score. This range shows how the scores are spread. You would like to know how far the scores differ on an average, from the general performance of the group. In the above set of scores see how the score 15 is so far off the average of 8.22. similarly 3 is also quite far off the average performance of the group. A teacher who sees such a variability would immediately call the former a high achiever group and the latter a low achiever group. What is the tendency of the group to deviate from the average score? Smaller the deviation greater the homogeneity of the group and vice versa. Two sets of scores may have the same mean. This does not mean that the groups are equally comparable. Look at the following scores.

1 3 5 7 9

3 4 5 6 7

For both the sets the mean is 5. But observe how much the scores deviate from the mean. In the first set the extreme scores differ as much as four measures whereas in the second set the extreme scores differ only two measures from the mean. This definitely speaks of more homogeneity in the second group. Thus scores that seemingly look comparable might have a lot of variation in them. Unless a researcher takes into consideration such variations, the interpretations would not reflect the true nature of the performance of the group.

These interpretations have their own implications for classroom work. A heterogeneous group needs a different treatment as compared to a homogeneous group. The teacher may have to work with smaller groups employing techniques that suit the levels of the group.

There are different measures of variability like the quartile deviation, the average deviation and the standard deviation. For the present purpose we limit ourselves to the standard deviation which is again, like the mean, a very useful measure because

- ③ This measure can withstand the extreme deviations
- ③ It is used in calculating other measures like the coefficients of correlation
- ③ It helps us get a better picture of the mean performance of a group.

CALCULATING THE SD BY THE SHORT METHOD FOR GROUPED DATA

Table showing the scores of 25 students of class III on a mathematics test

Class Interval	Mid Point X	f	$x\|$	$fx\|$	$fx\|^2$
76-80	78	10	0	0	0
81-85	83	7	+1	7	49
86-90	88	4	+2	8	64
91-95	93	2	+3	6	36
96-100	98	2	+4	8	64

N = 25 $\Sigma fx'$ = 29 $\Sigma fx'^2$ = 213

$$SD = i\sqrt{\frac{fx^2}{N} - c^2}$$

$$SD = 5\sqrt{\frac{213}{25} - 1.3456} = 13.39$$

f = frequency

x' = distance from AM in units of class interval

fx' = x' multiplied for 'weighed' by f

i = the length of the class interval

c = the algebraic sum of column fx' divided by N. = 1.16

Note: 'c' is the correction obtained by summing up fx' column divided by N which shows the distance of the scores in different class intervals from the assumed mean.

OTHER STATISTICAL MEASURES

The difference between the means obtained from the scores of pre-test and post-test may or may not be significant. Let us say that the mean of 83.8 is the mean score of the group in the pre-test.In the post-test, the same group scores give you a mean of 85. Can you say that there has been a definite improvement in the performance of the group? Is it not possible that the improvement shown is due to chance factors? May be, just one or two scores have been high and have influenced the mean. What is the chance or probability that the same difference will be shown if the two tests are conducted many times? What is the chance factor in the difference. We can find out whether the difference between two means is really sifnificant through statistical procedures.

Similarly you would also like to see if the performance of the group in one test corresponds with their performance in another. For this we can find out the coefficient of correlation. It is possible that the student who has scored the highest marks in one test has also scored the highest marks in the other and the like. It is also possible that the one who has scored the highest marks in one test has scored the highest marks in the other and the like. It is also

possible that the one who has scored the highest marks in one test has scored the lowest in another. The former is a case of positive correlation. The latter is a case of negative correlation. We can thus find out if a good performance in mathematics is correlated with their performance in logical reasoning. For the question whether high achievement in mother tongue indicates a high achievement in the 2nd language too we just need to find out the coefficient of correlation of the scores of the students in their mother tongue and the 2nd language.

Once the analysis of the data is done, we know how far the goals of the project have been realized. In other words, we can compare the results with the hypothesis made. The hypothesis may turn out to be true which would mean that the guessed problem solving strategy was effective. If the hypothesis is found not true, then it means that the guessed strategy was not effective. Strictly speaking there are statistical tools to decide whether a hypothesis could be accepted. In action research it may not be necessary to apply statistical tools so rigorously. If the teacher-researcher can make out a clear difference in the scores of the pre-test and the post-test, he/she can as well accept or reject the hypothesis as the case may be. One who has observed the problem should also be able to observe the absence of the problem. If you have gone for statistical tools, you can as well find out if the difference between the means is significant and accordingly accept or reject the hypothesis.

SUMMARY

- ③ Analysis of data gives the researcher feedback on the adequacy of the treatment given.
- ③ The data collected needs to be grouped for under taking statistical analysis.
- ③ The arithmetic mean tells the researcher about the general performance of the group.
- ③ The mean is the most representative single score of a group of scores.
- ③ The Standard Deviation tells the researcher how much the scores vary from each other in a group of scores on an average.

ACTIVITY

Here are two sets of marks of class V students in dictation. One statistic of marks shows the pre-test performance and the other test shows the post-test performance of the students. Find out if there is any improvement in the group because of action research intervention.

Name of the Students	Pre-test	Post-test
Amold	12	15
Balu	08	14
Chamundeswari	07	09
Mary	11	13
Elumalai	06	10
Farook	04	09
Govind	11	14
Hassina	11	14
Izzech	10	15
Joseph	12	14
Krishna	09	12
Laxmi	05	11

TRY IT OUT

Whenever you get some test scores try to apply the statistical measures and interpret the scores. Discuss with your colleagues your interpretations.

READING

Read the following books for more information on statistical procedures.

1. Best, J.W. – Research in Education
2. Garret – Statistics in Psychology and Education.
3. Kerlinger – Foundations of Behavioral Research

Activity

Students	Pre-test Scores Max. marks 10	Post-test Scores Max. marks 10
1	4	6

2	3	5
3	5	5
4	4	7
5	2	6
6	5	4
7	2	7
8	4	6

What do you think is the effect of 'treatment' given?

Can you say that the gain in the post-test is significant or substantial?