SOW 305

Introduction to Statistics for the Helping Profession

Ibadan Distance Learning Centre Series

SOW 303
Psychology of Social Work

By
E. M. Ajala Ph.D.
Department of Social Work
University of Ibadan

Published by
Distance Learning Centre
University of Ibadan

First Published 2011

ISBN

General Editor: *Prof. Bayo Okunade*
Series Editor: *Mr. Akisanya O.O.*

Typesetted @ Distance Learning Centre, University of Ibadan

Table of Contents

Vice-Chancellor's Message

I congratulate you on being part of the historic evolution of our Centre for External Studies into a Distance Learning Centre. The reinvigorated Centre, is building on a solid tradition of nearly twenty years of service to the Nigerian community in providing higher education to those who had hitherto been unable to benefit from it.

Distance Learning requires an environment in which learners themselves actively participate in constructing their own knowledge. They need to be able to access and interpret existing knowledge and in the process, become autonomous learners.

Consequently, our major goal is to provide full multi media mode of teaching/learning in which you will use not only print but also video, audio and electronic learning materials.

To this end, we have run two intensive workshops to produce a fresh batch of course materials in order to increase substantially the number of texts available to you. The authors made great efforts to include the latest information, knowledge and skills in the different disciplines and ensure that the materials are user-friendly. It is our hope that you will put them to the best use.

Professor Isaac F. Adewole
Vice-Chancellor

Foreword

The University of Ibadan Distance Learning Programme has a vision of providing lifelong education for Nigerian citizens who for a variety of reasons have opted for the Distance Learning mode. In this way, it aims at democratizing education by ensuring access and equity.

The U.I. experience in Distance Learning dates back to 1988 when the Centre for External Studies was established to cater mainly for upgrading the knowledge and skills of NCE teachers to a Bachelors degree in Education. Since then, it has gathered considerable experience in preparing and producing course materials for its Programmes. The recent expansions of the programme to cover Agriculture and the need to review the existing materials have necessitated an accelerated process of course materials production. To this end, one major workshop was held in December 2006 which have resulted in a substantial increase in the number of course materials. The writing of the courses by a team of experts and rigorous peer review has ensured the maintenance of the University's high standards. The approach is not only to emphasize cognitive knowledge but also skills and humane values which are at the core of education, even in an ICT age.

The materials have had the input of experienced editors and illustrators who have ensured that they are accurate, current and learner friendly. They are specially written with distance learners in mind, since such people can often feel isolated from the community of learners. Adequate supplementary reading materials as well as other information sources are suggested in the course materials.

The Distance Learning Centre also envisages that regular students of tertiary institutions in Nigeria who are faced with a dearth of high quality textbooks will find these books very useful. We are therefore delighted to present these new titles to both our Distance Learning students and the University's regular students. We are confident that the books will be an invaluable resource to them.

We would like to thank all our authors, reviewers and production staff for the high quality of work.

Best wishes.

Professor Bayo Okunade
Director

General Introduction and Course Objectives

This course is designed to introduce and acquaint you with basic principles in statistics and its application in social work. The course will equip you with requisite knowledge from various topics so that you can:

- discuss the meaning and uses of statistics in social work;
- differentiate types of measurement scales in statistics;
- present statistical data in tabular forms;
- statistically present data in frequency tables;
- graphically present frequency distribution in forms of histogram, polygon and ogive, etc;
- diagrammatically present data in form of bar and pie charts;
- comprehend the various measures of central tendency, such as mean, median and mode; and
- discuss the various types of measures of dispersion and variability, such as range and deviations.

LECTURE ONE

Meaning and Uses of Statistics

Introduction

Statistics deals with the collection, classification, analysis, and interpretation of data, leading to usable information. It is the collection of information in the form of data, evaluating it, and drawing conclusions from it. It is against this background that this lecture looks at the meaning and uses of statistics in everyday life experience.

Objectives

At the end of this lecture, you should be able to:

1. explain the meaning of statistics;
2. state the uses of statistics in everyday life; and
3. discuss what $\overline{X}$ and Σ stand for?

Pre – Test

1. What do you mean by statistics?
2. Why do you need statistics for your daily activities?

CONTENT

1. Meaning of Statistics

Statistics dates back to the beginning of man. History has it that rulers (Chiefs, Obas, Obis, etc.) obtained numerical facts about number of people in their domains so as to aid the planning and administration of their societies. The Biblical instance of the use of statistics was when Moses and David were commissioned to number their people.

The publication of Grant on natural and political observation in 1662 on mortality and political observation was a step towards a scientific method of statistics in the social sciences.

Decisions are made on the basis of available information otherwise called data. According to Speigel (1961), statistics is the scientific procedure for collecting, compiling, analyzing, summarizing, interpreting and presenting information in numerical form with a view to making valid inferences and decisions.

2. Importance of Statistics in Everyday Life

a. The development of statistics shows that many aspects of human progress depend on the correct analysis of numerical information, particularly, in economics and business.

b. Rational decisions are better taken if numerical figures are considered rather than taking decisions based on hunches.

c. It is helpful in making calculated predictions.

d. It is used to pass to the public or users some useful information.

3. Some Selected Statistical Terminologies/Symbols

a. Population: This is the totality of the objects (living or non-living, which is countable) in a defined area of interest.

b. Sample: A sample is a subject of the population, that is, it is a subset (part) of the population selected. It is representative of the whole population.

c. Sampling: This is the process of selecting a representative from the larger population.

d. $\overline{X}$ (called "bar X''): This is the arithmetic mean of the values of that variable.

e. f = This stands for frequency, that is, the number of times a given value occurs in a collection of figures.

f. n = This stands for the number of items (or pairs of items) in a collection of figures.

g. Σ also called sigma: This means "the sum of" and simply indicates that the numbers following it should be added.

h. r = This stands for "coefficient of correlation".

i. **Data:** This is measurement that is made on the subjects of an experiment. Usually, data consist of the measurements of the dependent variable. Data as originally measured are often referred to as raw or original scores.

j. **Variable:** A variable is any property or characteristics of some events, objects, or person that may have different values at different times depending on the conditions.

k. **Independent Variable(s):** These are variables that are systematically manipulated by the investigator.

l. **Dependent Variable(s):** These are variables that the investigator measure to determine the effect of the independent variable(s).

Summary

In this lecture, you have been acquainted with different definitions of statistics, uses of statistics in everyday life and the terminologies/symbols to be employed throughout this course.

Post – Test

1. Give the meaning of statistics.
2. State the uses of statistics in everyday life.
3. What do $\overline{X}$ and Σ stand for?

References

Hammed, Ayo (2002).*Statistics in Education: A Basic Text*. Ibadan: Stirling-Horden Publishers (Nig).

Spiegel, M. R. (1961). *Theory and Problems of Statistics:* Schaum's *Outline Series*. London: McGraw Hill Book Company.

Pagano, R.R. (2001). *Understanding Statistics in the Behavioural Sciences.* USA: Wadsworth.

LECTURE TWO

Scales of Measurement in Statistics

Introduction

Measurement is concerned with the real world in terms of purpose and operations. As you know, all scientific observation is usually a collection of measurements and these measurements are called data. As discussed in the earlier chapter, statistics is a tool for accurate collection of these measurements or data. The lecture focuses on the discussion of various scales of measurement in statistics.

Objectives

At the end of this lecture, you should be able to:

1. discuss the concept of measurement; and
2. explain the various scales of measurement.

Pre-Test

1. Discuss the three levels of measurement.
2. List types of measurement scale you know.

CONTENT

A measurement scale is an established set of rules for assigning scores to variables so as to indicate the quality or quantity of the variables involved. A good measurement scale provides assignment of numbers to represent all potentially observable values of variables and also provides for exclusiveness in the assignment of numbers to the observed quality or quantity of the variables. The type of measuring scales employed in collecting the data helps determine which statistical inference test is used to analyze the data.

Levels of Measurement

For logical manipulations to obtain an acceptable result, a measurement scale should possess the following characteristics:

1. Magnitude

A scale has magnitude if a point or unit of the attribute being measured can be judged as greater than, less than, or equal to, another point or unit. For example, if student A obtains a score of 35 on an "Aggressiveness scale" and student B scores 55, then we can confirm that student B is more aggressive than student A.

2. Equal Interval

A scale has equal interval if the magnitude represented by one unit of measurement on the scale is equal, regardless of where the unit falls on the scale. For example, car A has 20 litres of petrol left after a journey, which started with 40 litres, while car B started with 50 litres and ended with 30 litres. Their difference with the left-over of petrol in the cars is equal.

3. Absolute Zero Point

This is the point where the measurement scale shows that nothing at all of the attribute being measured exists. It represents the absolute starting point of the thing being measured. Thus, "0 litres" of volume is a scale value that implies no volume whatsoever-absolute zero.

It must be noted that the discussed characteristics of measurement help in dividing the measurement scale into four different types;

a. Nominal scale
b. Ordinal scale
c. Interval scale
d. Ratio scale

a. ***Nominal Scale:*** This is classification of objects and giving them the name of the category to which they belong. The variables have only the characteristics of exhaustiveness and mutual exclusiveness. Examples include gender, ethnicity, religious affiliation, birthplace etc. Therefore, nominal scale is most often used with variables that are qualitative in nature rather than quantitative. A fundamental property of nominal scale is that of equivalence, that is, all members of a given class are the same from the standpoint of the classification variable. It merely allows categorisation of objects into mutually exclusive categories.

b. ***Ordinal Scale:*** This is also known as ranking scale. With an ordinal scale, we rank-order the objects being measured according to whether they possess more, less, or the same amount of the variable being measured. Thus, an ordinal scale allows determination of whether A>B, A=B, or B<A. For example, ranking of five speakers as 1^{st}, 2^{nd}, 3^{rd}, 4^{th}, and 5th has an intrinsic meaning; however, it says nothing about the magnitude of the difference between adjacent units on the scale. Therefore, ordinal scale does not have the property of equal intervals between adjacent units, nor tell the absolute level of the variable.

c. ***Interval Scale:*** On an interval scale, intervals are organised sequentially, as all categories are the same size. Thus, it is a series of equal intervals like centimeters on a ruler, seconds in time measurement, kilograms in weight and Celsius scale of temperature measurement. Interval scale possesses the properties of the ordinal scale and has equal interval between adjacent units. The equality of intervals makes it possible to determine the magnitude of differences. Another characteristic of an interval scale is that it has an arbitrary zero point. That is, the value zero (0) is assigned to a particular location on the scale simply as a matter of convenience or reference.

d. ***Ratio Scale:*** It has all the properties of an interval scale and, in addition, has an absolute zero point. Without an absolute zero point, it is not legitimate to compute ratios with the scale reading. With ratio scale, one can construct ratios and perform all the other mathematical operations usually associated with numbers (i.e. “+”, “−”, “x”, “÷”). Ratio scale not only allows measurement according to direction as magnitude of the differences between measurements but also describe differences in terms of ratios.

Summary

In this lecture, we have discussed the attributes of any scale of measurement in statistics. The attributes discussed are magnitude, equal interval and absolute zero point. Furthermore, a brief listing of types of measurement scale was done.

Post-Test

1. Discuss in detail the following attributes of measurement:
    a. Absolute zero point
    b. Equal interval
    c. magnitude
2. List the major types of measurement scale.

References

Adamu, S. O. and Johnson, T. L. (2004). *Statistics for Beginners*. Ibadan: Evans Brothers (Nig Publishers).

Hammed, Ayo (2002). *Statistics in Education: A Basic Text*. Ibadan: Stirling-Horden Publishers (Nig).

Pagano, R.R. (2001). *Understanding Statistics in the Behavioural Sciences.* USA: Wadsworth.

LECTURE THREE

Tabulation: Presentation of Statistical Data

Introduction

In any statistical investigation, masses of data are obtained. In its original form, such a mass of data fails to give us any detailed information. It is necessary to summarize them. Therefore, the next step after figures have been collected is to lay them out in an orderly way so that they are more easily comprehended.

The first task of any statistician is to reduce and simplify the details of the data so collected into such a form that the salient features may be brought out, while still facilitating the interpretation of the assembled data. Therefore, the purpose of tabulation is to condense and thereby facilitate comparison of data. Tabulation is the first kind of elementary summary work in statistical investigation.

The lecture discusses the presentation of data collected in form of tabulation.

Objectives

At the end of this lecture, you should be able to:

1. identify types of statistical tables;
2. describe the principles of table construction; and
3. list the characteristics of a statistical table.

Pre – Test

1. Mention types of statistical tables you know.
2. List the main characteristics of a good statistical table.
3. State the main advantages of any tabular layout in statistics.

CONTENT

Presentation of statistical data in a tabular form requires that you have the knowledge on:

1. types of tables;
2. principles of table construction;
3. characteristics of a table; and
4. advantages of the table so constructed.

1. Types of Tables

Three types of tables are usually used in presenting statistical work. They are:

a. **Source or reference table:** This is a table on which further analysis is based.
b. **Working table:** This is a sheet on which initial calculations are done before the final tables are arrived at.
c. **Summary or text table:** This is a table found in books either to support the arguments presented or to have easy reference to them.

2. Principles of table construction

In constructing a table, student must know from the very beginning the purpose of the table you are to construct. A table should be constructed such that it is able to achieve its objectives as much as possible. To this extent the following principles are to be observed

a. The original figures should be presented in an orderly manner.
b. The table should show a distinct pattern in the figures.
c. The table should summarize the figures effectively.
d. The table should show salient figures, which other people may use in future statistical studies.

3. Characteristics of a Table

A statistical table has some general features, which are given below:

- **a. Simplicity:** A table with too many details or which is too complex would be hard to understand and would, therefore, be of no use. It is better to construct a table, showing only a little but which is easily understandable.
- **b. A comprehensive, short and explanatory title:** A general title that is self explanatory of the purpose of the table is preferable to a wordy title.
- **c. Source:** All the figures on a table come from somewhere; thus, the source or sources of the figures must be stated, usually, at the bottom of the table. Sometimes the information about the sources of the figure may be accompanied by a footnote. Footnotes explain any irregular figure or other aspects of the table.
- **d. State the units used.** There should be a clear indication of the units in which the data in the table are given, as in N = Nigerian Naira, £ = pound sterling, 000 = per thousand.
- **e. Headings of columns and rows should be clearly stated.** No form of ambiguity in meaning should exist in the headings of columns and rows.
- **f. Totals should be shown where appropriate.** Totals are used in a table for the following purposes:
  - i. give the overall total of a main class;
  - ii. indicate that the preceding figures are sub-divisions of the total; and
  - iii. indicate that all the items on the table have been accounted for.

4. Advantages of a Tabular Layout

The presentation of information in a tabular form, rather than in a narrative form, has the following advantages:

- a. It enables the required figure or information to be located more quickly.
- b. Comparison between different classes can easily be made.
- c. A given pattern within the figures, which cannot be seen in the narrative form, is easily seen.

d. It is concise and less voluminous. It occupies less space than the narrative form.

Summary

In this lecture, we have discussed the issue of tabulation of statistical data. In this process, we mentioned the importance of tabulation, types of tables in presenting statistical work and the guiding principles in table construction. To make statistical tables meaningful, the major characteristics were discussed, and we also briefly examined the advantages of presenting statistical data in a tabular form.

Post – Test

1. List the three major types of statistical tables you know.
2. What are the major characteristics of a statistical table?
3. Of what importance is a statistical table to its user?

References

Adamu, S. O. and Johnson, T. L. (2004). *Statistics for Beginners*. Ibadan: Evans Brothers (Nig Publishers).

Hammed, Ayo (2002). *Statistics in Education: A Basic Text*. Ibadan: Stirling-Horden Publishers (Nig).

Harper, W. M. (1975). *Statistics*. London: Macdonald and Evans Ltd.

Pagano, R.R. (2001). *Understanding Statistics in the Behavioural Sciences.* USA: Wadsworth.

LECTURE FOUR

Data Organisation and Frequency Tables

Introduction

Collection of data for solving problems leads to a large mass of data. In its original, unprocessed state, data so collected appear meaningless; they need to be condensed or understandably before they can be of much use. In this chapter, we shall look at how frequency tables via tally system, both for ungrouped or grouped data, are constructed for meaningful data presentation.

Objectives

At the end of the lecture, you should be able to:

1. define frequency distribution; and
2. prepare a frequency table from a non-grouped and grouped data.

Pre – Test

1. What is a tally in statistics?
2. What is a frequency distribution?

CONTENT

Data are a set of usable information collected in numerical form. The set of data could be a set of numbers, figures or measurements derived from observation. However, the data so collected could be arranged in such a manner that a meaningful interpretation could be easily made. The first step to make a meaning out of data so collected is to arrange them into what is called frequency distribution.

Listing of Data

Listing of the data is writing the data out. Listing is a prerequisite of grouping for analysis, a pattern that assists in cross-checking. When a large number of observations are involved in the listing, tabulation comes in to prevent flaws that might crop up.

For example: Table 1

In a statistics test for 40 students, the following marks were obtained.

2,	3,	5,	3,	4,	6,	8,	8,	4,	3.
7,	4,	5,	6,	9,	7,	3,	5,	7,	2.
6,	7,	4,	6,	6,	5,	4,	2,	8,	7.
9,	8,	5,	5,	6,	4,	2,	3,	3,	4.

These scores are termed raw scores because they have not been meaningfully arranged.

Frequency Table

A frequency table is a tally of the number of times each score value occurs in a group of scores. Therefore, the first step to be taken in making the raw data more meaningful is to re-list the figures in order of size, that is, re-arrange them so that they run from the lowest to the highest. Such a list of figures is called an array.

Since some of the values occurred more than once, it would be clearer if each figure were listed once and the number of times it occurred written alongside as shown below (Table 2)

a. Ungrouped Data

In statistics, the number of occurrence is called the frequency (and symbolised as f) and what you have in Table 2 is called an ungrouped frequency distribution table. Therefore, an ungrouped frequency distribution is a list of figures occurring in the raw data, together with the frequency of each figure.

The frequency for each figure is easily obtained through the use of tally (strokes). In order to prepare a frequency table, therefore, we count the number of strokes (tally) for each mark to obtain the frequency for the mark. For example, **p**repare the data in Table 1 into a tally and frequency table.

Table 2: Frequency Distribution Table

Marks	**Tally**	**Frequency**
2	IIII	4
3	~~IIII~~ I	6
4	~~IIII~~ II	7
5.	~~IIII~~ I	6
6	~~IIII~~ I	6
7	~~IIII~~	5
8	IIII	4
9	II	2
	Total	**40**

The scores in Table 1 have been arranged into a frequency distribution as shown in Table 2. The data now are more meaningful. It is easy to see and attach value (occurrence) to the marks scored. Frequency distribution thus present the scores in such a way as to facilitate ease of understanding and interpretation.

b. Grouped Data

When there are too many figures for the mind to easily comprehend, the information is effectively presented by tallying. In other words, you simplify the understanding of information being given by the figures by grouping them. For example, assume you use the scores for another set of 40 students in statistics examination thus:

Table 3

42,	24,	35,	52,	49,	42,	47,	31,	47,	44.
47,	34,	46,	54,	50,	61,	32,	46,	27,	46.
43,	50,	32,	29,	61,	44,	30,	20,	44,	40.
39,	33,	57,	59,	48,	31,	51,	56,	32,	55

To construct a frequency distribution for the above table requires three major steps:

1. choose the classes into which the data are to be grouped. When grouping data, do not let the width be too large, otherwise there will be the lose of information; nor too narrow (you will be back to the problem of too large numbers to manage);
2. tallying; and
3. counting the number in each class.

In choosing the classes, the following should be done:

1. Find the range of the data (that is, the value of the largest mark minus the value of the smallest mark) e.g. 61 – 20 = 41
2. Estimate the size of the class intervals. It may be less than 10 if the total number of cases is small, and it may be more if the total is very large. e.g. 5 intervals.
3. Decide the number of the class intervals. This is done by dividing the range by the number of interval selected in (2.) above. Round off the result to the nearest whole number so as to accommodate all variables. A trade-off between losing information (wide class interval) and presenting a meaningful visual display, choose an interval width neither too wide nor too narrow.

   E.g. $\frac{61 - 20}{5} = \frac{41}{5} = 8.2 = 9$
4. Select the lowest class interval so that its lowest stated limit is evenly divisible by the size of the interval. E.g. 20 – 24, 25 – 29, etc;
5. Make sure that the observations belong to only one class; and
6. Make the classes of equal width, so that each will cover the same range of values.

   Now, construction of the tally and frequency distribution for Table 3.

Table 4: Frequency Distribution Table

Class Interval	Class Mark	Tally	Frequency
20 – 24	22	II	2
25 – 29	27	II	2
30 – 34	32	~~IIII~~ III	8
35 – 39	37	II	2
40 – 44	42	~~IIII~~ II	7
45 – 49	47	~~IIII~~ III	8
50 – 54	52	~~IIII~~	5
55 – 59	57	IIII	4
60 – 64	62	II	2
		Total	**40**

Summary

In this lecture, we have discussed how data collected are organized into frequency distribution. The frequency distribution for both grouped and non-grouped data were examined, and the steps involved in the construction of frequency were also taught. Frequency distribution was seen as a system for classifying data, usually in a tabular form, so that the frequency of observations in each class can be ascertained.

Post – Test

Tabulate the following data into frequency table.

1. In a statistical class of 20 students, the following marks were obtained after a test.

   4, 7, 6, 8, 9, 6, 5, 3, 2, 8.

   3, 4, 5, 3, 2, 4, 5, 7, 3, 2.

2. From the attendance of students in classes within thirty days, the following data was collated:

   42, 53, 27, 33, 40, 54, 37, 45, 34, 37.

   51, 36, 47, 46, 45, 41, 40, 37, 45, 26.

   39, 56, 57, 55, 30, 37, 50, 40, 31, 41.

References

Adamu, S. O. and Johnson, T. L. (2004). *Statistics for Beginners*. Ibadan: Evans Brothers (Nig Publishers).

Hammed, Ayo (2002). *Statistics in Education: A Basic Text*. Ibadan: Stirling-Horden Publishers (Nig).

Harper, W. M. (1975). *Statistics*. London: Macdonald and Evans Ltd.

Pagano, R.R. (2001). *Understanding Statistics in the Behavioural Sciences.* USA: Wadsworth.

LECTURE FIVE

Graphical Presentation of Frequency Distribution

Introduction

As earlier discussed, a frequency table is a numerical presentation of data in an organised summary form. To make this data information more attractive and easier for visualisation of important features of the data. We use diagrams, graphs and pictures. For this reason, this chapter will discuss the graphical presentation of data under the following:

1. Histogram;
2. Frequency polygon; and
3. Ogive

Objectives

At the end of this lecture, you should be able to:

1. plot a histogram graph from frequency table;
2. plot frequency polygon; and
3. plot an ogive.

Pre – Test

1. What is a histogram in statistics?
2. Define a frequency polygon in statistics.
3. Calculate the cumulative frequency for tables in the post-test of Chapter 4.

CONTENT

Once data collected from observations have been grouped, tallied, the next step is to present the result pictorially so that their salient features may be made clear at a glance. In making a graphical representation of observations, subjectivity should be eliminated through the following principles:

1. Do not eliminate the zero point from the vertical axis.
2. Make the height of the maximum frequency on the vertical axis about three-quarters the length of the horizontal axis. This is just an arbitrary rule to prevent subjectivity from creeping in.
3. Label graph clearly, so that they can be self explanatory as much as possible.

Graphical Presentation of a Frequency Distribution

The following points should be noted while constructing graphs

1. A graph has two axes: vertical (ordinate, or Y-axis) and horizontal (abscissa or X-axis).
2. The scores are plotted along the horizontal axis, and the same characteristic of the scores is plotted on the vertical axis.
3. Suitable units for plotting scores should be chosen along the axes.
4. Set the intersection of the two axes at zero and then choose scales for the axes such that the height of the graphed data is about three-fourths of the width.
5. When the intersection of the two axes is not at zero, break the relevant axis near the intersection.
6. Each axis should be labeled, and the title of the graph should be short and explicit.

Three most common ways of presenting frequency distribution graphically are through histogram, frequency polygon and the cumulative frequencies (ogive).

1. Histogram

This is a graphical representation of a frequency distribution in which class frequencies are plotted against class marks. Then, class frequencies are represented by vertical bars, centered at the class marks, whose area is proportional to the frequency. Since the intervals are continuous, the vertical bars must touch each other.

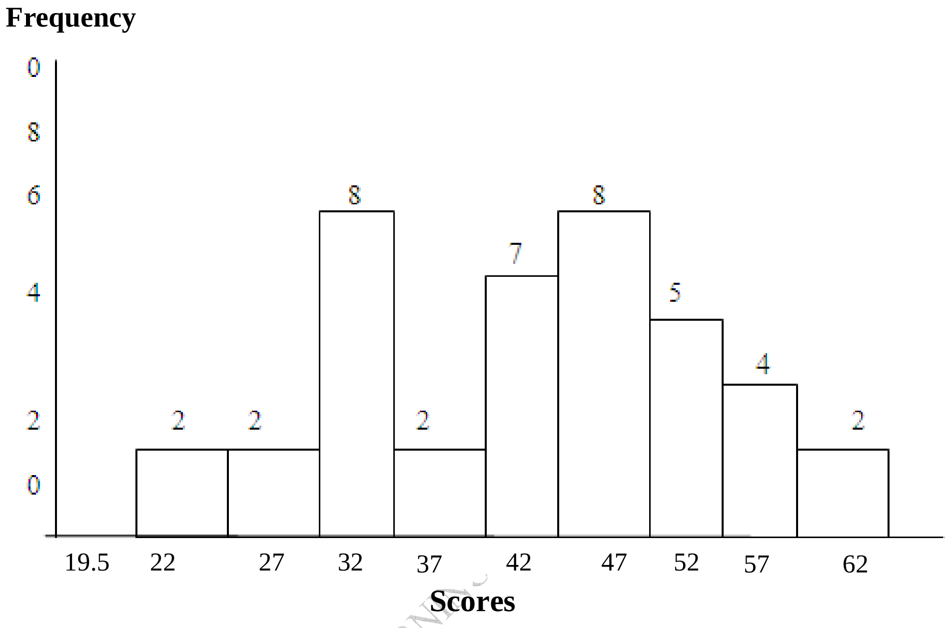


Figure1 – Histogram of Scores in Statistics

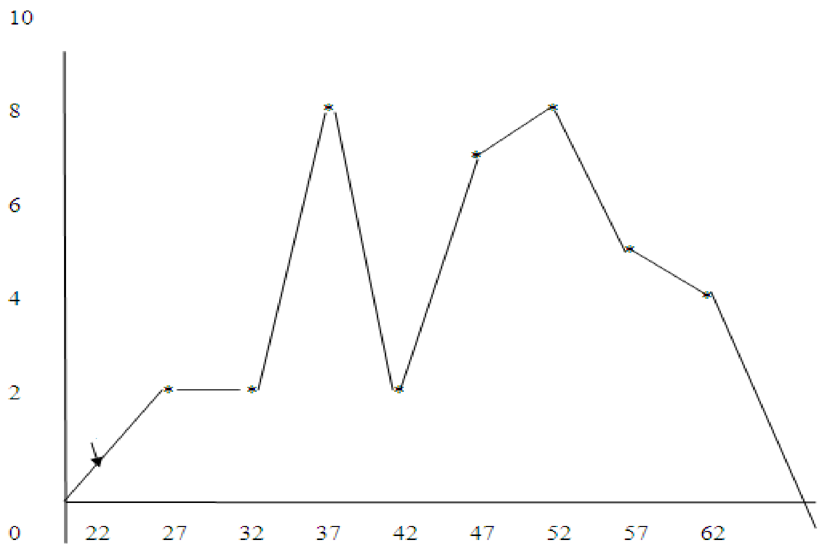
2. Frequency Polygon

A frequency polygon is a graphical representation of a frequency distribution in which class frequencies are plotted against class marks. These are then joined by straight line segments.

Steps in drawing a frequency polygon

a. Mark the mid-point on the x – axis.
b. Make a dot corresponding to the frequency of the mid-point on the y – axis.
c. Join the mid-point values with straight lines.
d. Extend the curve downwards at both ends so that it cuts the axis at points which are half a class interval beyond the outside limits of the end classes. This should be done even if it means that the curve crosses the vertical axis and ends in the minus part of the graph.

Frequency



Scores

Fig. 2 – Frequency polygon of scores in statistics

The major difference between a histogram and a frequency polygon is that while histogram displays the scores as though they were equally distributed over the interval, the frequency polygon displays the scores as though they were all concentrated at the mid-point of the interval and continuously distributed.

3. Ogive

Ogive is the name given to the curve obtained when the cumulative frequencies of a distribution are graphed.

Construction Procedure

a. Compute the cumulative frequencies of the distribution.
b. Let the graph have the cumulative frequency on the vertical axis and the class mark notation on the horizontal axis.
c. Plot a starting point at zero on the vertical scale and the lower class limits on the first class.

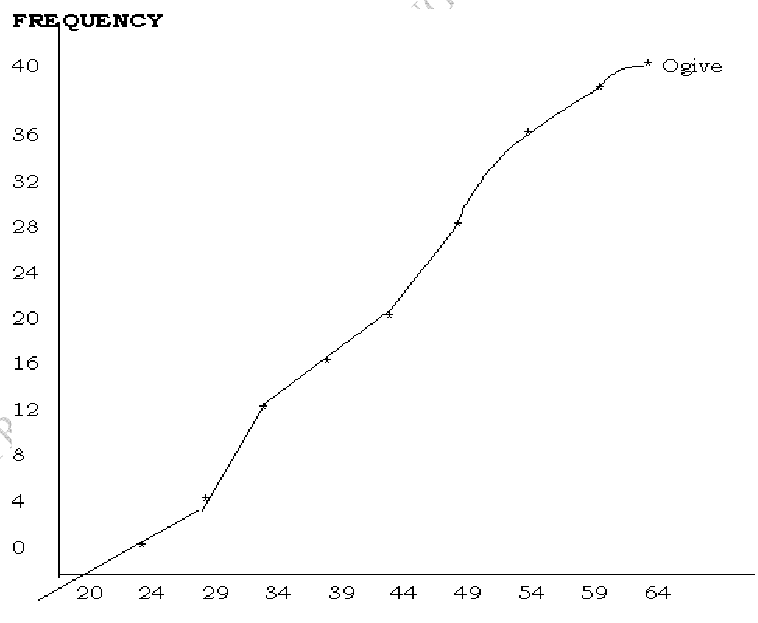
d. Plot the cumulative frequencies on the graph at the upper class limits of the classes to which they refer.

e. Join all the points.

Table 5

Class Interval	Frequency	Cumulative Frequency
20 – 24	2	2
25 – 29	2	4
30 – 34	8	12
35 – 39	2	14
40 – 44	7	21
45 – 49	8	29
50 – 54	5	34
55 – 59	4	38
60 – 64	2	40

Cumulative



Scores

Figure 3: Cumulative Frequency of Scores in Statistics

Summary

In this lecture, we have has discussed the graphical presentation of frequency distribution. The topics taught are the procedure for plotting histogram, frequency polygon and ogive. We stated that cumulative frequency is an off-shot of the addition of successive frequency for his class. The idea of class marks and class interval were equally emphasized.

Post – Test

1. What is a histogram in statistics?
2. What is a frequency polygon in statistics?
3. Calculate the cumulative frequency and plot the ogive for question 2 of the Post Test of Lecture Four.

References

Adamu, S. O. and Johnson, T. L. (2004). *Statistics for Beginners*. Ibadan: Evans Brothers (Nig Publishers).

Hammed, Ayo (2002). *Statistics in Education: A Basic Text*. Ibadan: Stirling-Horden Publishers (Nig).

Harper, W. M. (1975). *Statistics*. London: Macdonald and Evans Ltd.

Pagano, R.R. (2001). *Understanding Statistics in the Behavioural Sciences.* USA: Wadsworth.

LECTURE SIX

Diagrammatic Presentation of Data

Introduction

Graphs are not the only way visual presentation of data. Other alternative methods are:

1. pictorial representatives;
2. bar charts; and
3. pie charts.

For the purpose of this lecture, we shall consider bar charts and pie charts.

Objectives

At the end of this lecture, you should be able to:

1. define and construct a bar chart; and
2. define and construct a pie chart

Pre – Test

1. How many degrees have a pie chart?
2. Define a bar chart.
3. List the types of bar charts you know.

CONTENT

Bar Charts

Bar charts can be classified into:

1. simple bar charts;
2. component bar charts; and
3. multiple bar charts.

Bar charts are useful for depicting in a simple manner a series of changes in major figures. They depict data more accurately than pictograms. Bar charts are used to indicate the sizes of component figures.

1. Simple Bar Chart

Data are represented in a simple bar chart by a series of bars. The height (or length) of each bar indicates the size of the figure represented. Such bars may be drawn vertically or horizontally.

A bar chart is very easy to construct, as it always starts with zero for the height measurement with equal size of the width for the bars and spaces between the bars. The bars for each category in a bar graph do not touch each other. This further emphasizes the lack of a quantitative relationship between the categories.

A simple bar chart is used where changes in totals only are required. For example, construct a bar chart for the following data:

Table 6

The data below show the attendance of patients at a clinic per week.

Days	**Mon**	**Tues**	**Wed**	**Thurs**	**Fri**	**Sat**	**Sun**
Attendance rate	120	150	300	220	100	250	60

Scale: 1cm rep. 50 patients on the Y – axis
½cm rep. each day on the X – axis

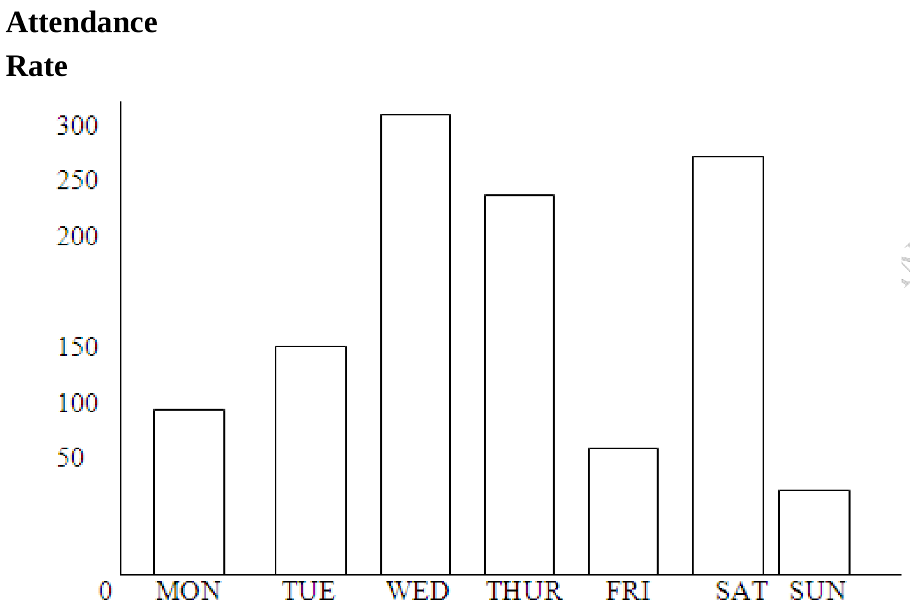


Figure 4 Simple Bar Chart: Attendance at Medical Clinic for the Week

2. Component Bar Chart

This is constructed when the overall height of the bar and the individual component lengths represent actual figures, that is, the bar is constructed when each total figure is built up from two or more component figures. The bars are drawn with equal thickness but the length of each varies in proportion to the data or figure it represents. This is used where changes in totals and an indication of the size of each component figure are required. For example, draw a component bar chart to show the output (per tones) of three farmers from three crops in a year.

Table 7

Farmer/Crops	**A**	**B**	**C**
MAIZE	300	400	400
YAM	400	350	450
CASSAVA	150	100	150
TOTAL	**850**	**850**	**1000**

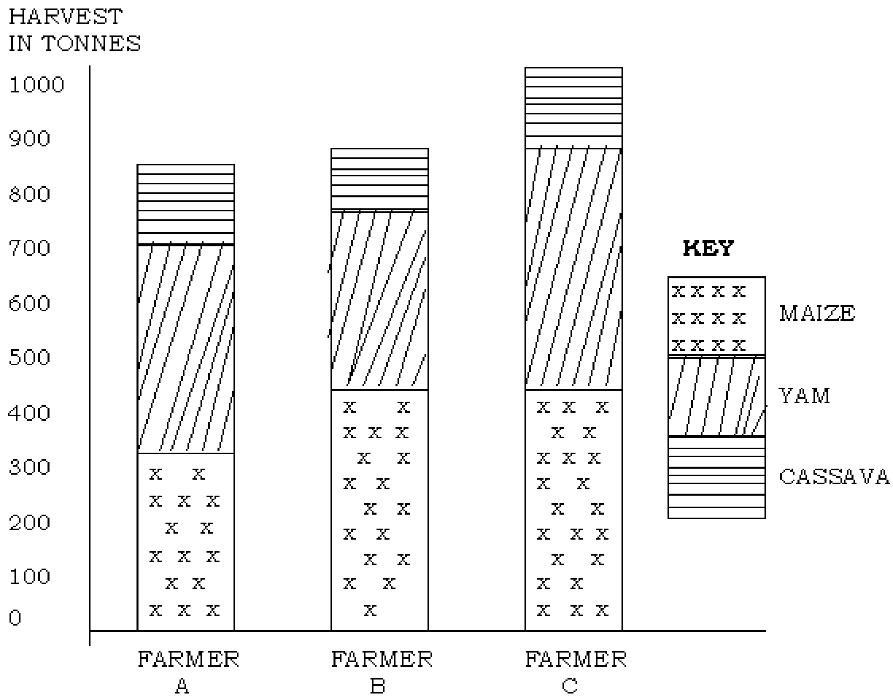


Figure 5: Component Bar Chart

3. Multiple Bar Chart

In this chart, the component figures are shown as separate bars adjoining each other. The height of each bar represents the actual value of the component figure. Multiple bar charts are used, where changes in the actual values of component figures only are required, and the overall total is of no importance.

For example, compute Table 7 into a multiple bar charts

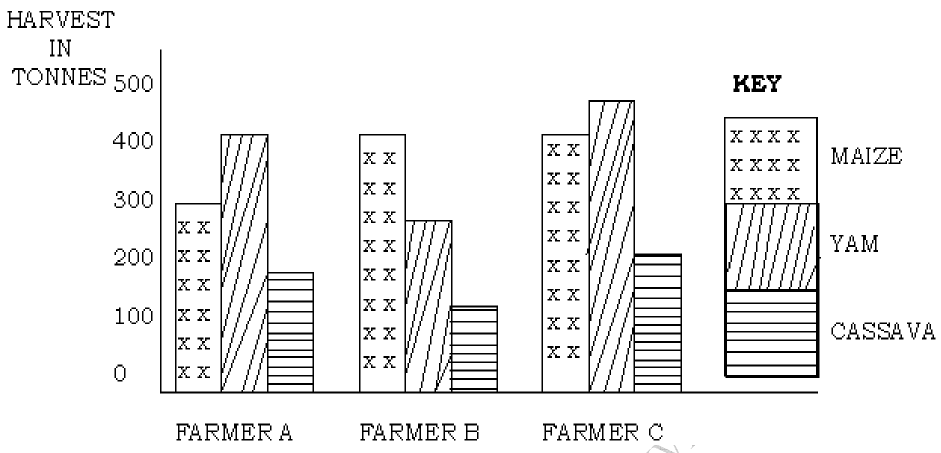


Figure 6: Multiple Bar Charts

Pie Charts

Component and multiple bar charts can only be used when there are not more than three or four components. When a large number of components have to be shown, a pie chart is more suitable.

A pie chart is a circle which is divided into sections by radial lines so that the area of each section is proportional to the size of the figure represented.

Steps in the construction of pie charts

1. Calculate the percentage of each component.
2. Convert step 1 (percentages) to degrees.
3. Mark out the degrees in a drawn circle.

For example, draw a pie chart for Table 6

Table 8

Days	Attendance	Percentage	Degree
Mon	120	$\frac{120}{1200} \times \frac{100}{1} = 10.0$	36^0 $(\frac{120}{1200} \times 360^0)$
Tue	150	$\frac{150}{1200} \times \frac{100}{1} = 12.5$	45^0 $(\frac{150}{1200} \times 360^0)$
Wed	300	$\frac{300}{1200} \times \frac{100}{1} = 25$	90^0 $(\frac{300}{1200} \times 360^0)$
Thurs	220	$\frac{220}{1200} \times \frac{100}{1} = 18.33$	66^0 $(\frac{220}{1200} \times 360^0)$
Fri	100	$\frac{100}{1200} \times \frac{100}{1} = 8.33$	30^0 $(\frac{100}{1200} \times 360^0)$
Sat	250	$\frac{250}{1200} \times \frac{100}{1} = 2.08$	75^0 $(\frac{250}{1200} \times 360^0)$
Sun	60	$\frac{60}{1200} \times \frac{100}{1} = 5.0$	18^0 $(\frac{60}{1200} \times 360^0)$
Total	**1200**		**360^0**

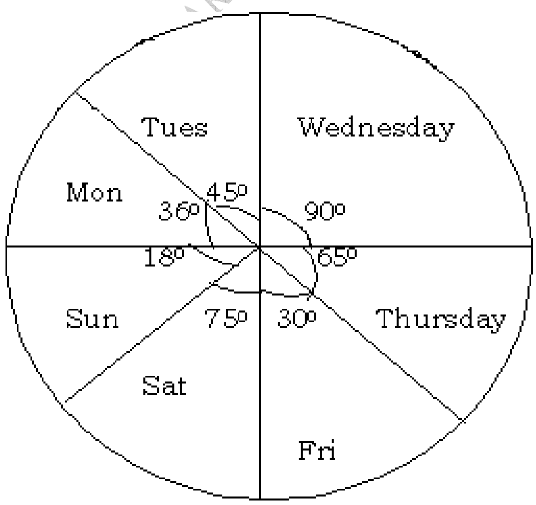


Figure 7: Pie Chart: Attendance at Medical Clinic for the Week

Summary

In this lecture, we have established that other ways of presenting data visually are through bar and pie charts. We noted that there are three types of bar charts; the simple, the component and the multiple bar charts. We also discussed their construction and features. Furthermore, we examined the construction of pie charts. In highlighting the differences between bar charts and pie charts we stated that a bar chart is a vertical graph in which the height denotes the frequency; however a pie chart is a circular graph, which is divided into sections using the sum of angles at point – 360^0.

Post – Test

1. How many degrees make a pie chart?
2. Define a bar chart.
3. List the types of bar charts you know.
4. Construct a bar chart for the following data obtained from the throwing of a dice for twenty times.

   3, 2, 3, 1, 2, 6, 4, 3, 4, 6.

   5, 1, 6, 2, 3, 5, 4, 4, 5, 3.
5. Plot Question 4 into a pie chart.

References

Adamu, S. O. and Johnson, T. L. (2004). *Statistics for Beginners*. Ibadan: Evans Brothers (Nig Publishers).

Hammed, Ayo (2002) *Statistics in Education: A Basic Text*. Ibadan: Stirling-Horden Publishers (Nig).

Harper, W. M. (1975) *Statistics*. London: Macdonald and Evans Ltd.

Pagano, R.R. (2001) *Understanding Statistics in the Behavioural Sciences.* USA: Wadsworth.

LECTURE SEVEN

Measures of Central Tendency: Mean

Introduction

When a series of observations has been tabulated, that is, put in the form of a frequency distribution, the next step is the calculation of certain values, which may be used as descriptive of the characteristics of that distribution. These values will enable comparison to be made between one series of observations and another. The two most important types of numerical summaries are (1) those that locate the "centre" of a distribution of data, called "measure of central tendency", and (2) those that describe the "speed" of data values about the center location called "measures of dispersion" or variability.

In this lecture, the first measure of central tendency will be considered. This is the mean.

Objectives

At the end of this lecture, you should be able to:

1. calculate the mean, for grouped and ungrouped, data; and
2. calculate the mean via assumed mean, coding method.

Pre – Test

1. What do you understand by "arithmetic mean"?
2. What are the properties of the mean?

CONTENT

The Mean

This is the most commonly used. However, there are situations where the mean does not provide a good measure of central tendency or where it is impossible to compute a mean. These are:

a. when a sample contains a few extreme scores unusually high or unusually low values- the mean tends to be distorted by the extreme values so that it is not a good central, representative value.

b. when sample data consist of measurements from a nominal scale and are not numerical values. For example gender, occupation, eye-colour etc. Because no numerical values are involved, it is impossible to compute a mean value for such data.

The mean is computed by summing scores and dividing the scores by the number of individual. The mean for a population is represented by μ, the Greek letter mμ, while sample population is represented by $\overline{X}$ (X-bar). The mean has various types. In this lecture, we shall discuss only the Arithmetic mean.

Properties of the Mean

The mean has many important properties namely:

1. The mean is sensitive to the exact value of all the scores in the distribution. Any change in any of the scores to be calculated will cause a change in the mean because you have to add all the scores to obtain the calculated mean. This is not true of the median or the mode.

2. The sum of the deviations about the mean equals zero.

   That is, $\sum(Xi - \overline{X}) = 0$.

3. The mean is very sensitive to extreme scores. The mean would have to shift a considerable distance to re-establish balance when extreme values are involved.

4. The sum of the squares deviations of all the scores about their mean is a minimum. That is, $\sum(Xi - \overline{X})^2$ is a minimum.

5. Under most circumstances, of the measures used for central tendency, the mean is least subject to sampling variations.

Arithmetic Mean: This is the total of the values of a variance divided by the number of values.

I. For Ungrouped data without frequency, the mean, represented by $\overline{X}$ (read "x – bar") is written as:

$$\overline{X} = \frac{x_1 + x_2 + \cdots x_N}{N}$$

$$= \frac{\sum_{i=1}^{N} Xi}{N}$$

Where $\sum$ = sigma; addition of all variables

X = observed variable individually

N = number of values of the variables

i = takes value from 1 up to n.

Example: A variable has the following values: 2, 4, 6, 8 and 10. Calculate the mean.

Table 9

Solution: $\frac{2 + 4 + 6 + 8 + 10}{5}$

$= \frac{30}{5}$

$= \mathbf{\underline{6}}$

II. For ungrouped data with frequency, the mean is calculated using

$$\overline{X} = \frac{\sum fx}{\sum f}$$

Where $\overline{X}$ = mean

$\sum fx$ = total product of value of x and corresponding frequency

$\sum f$ = total of all frequency

For example, find the mean for the score in statistics test in Table 2

Table 10

Mark (x)	Frequency (f)	(fx)
2	4	8
3	6	18
4	7	28
5	6	30
6	6	36
7	5	35
8	4	32
9	2	18
	40	**205**

$\sum f = 40$

$\sum fx = 205$

$$\overline{X} = \frac{\sum fx}{\sum f} = \frac{205}{40}$$

$$= \underline{\mathbf{5.125}}$$

III. To calculate the mean from data with frequency and class interval. Since values of variables are large and are to be grouped to obtain the frequency for each group, the mid-point of the class interval will now serve as the value X. Let us illustrate this with a worked example.

Find the mean to the data in Table 4.

Table 11

1 Class Interval	2 Frequency (f)	3 Class mark (X)	4 fx
20 – 24	2	22	44
25 – 29	2	27	54
30 – 34	8	32	256
35 – 39	2	37	74
40 – 44	7	42	294
45 – 49	8	47	376
50 – 54	5	52	260
55 – 59	4	57	228
60 – 64	2	62	124
	40		**1710**

Solution

1. Create column (3) to show the mid-point (x) or the class mark of the class interval. For example, for class interval (20-24), the class mark is obtained by adding the lower class interval plus the upper class interval divided by 2

   i.e. $\frac{20 + 24}{2} = \frac{44}{2} = 22$

2. Create Column (4) = fx (Column 2 multiply by column 3)
3. Using the formula

   $$\overline{X} = \frac{\sum fx}{\sum f}$$

   $= \frac{1710 \text{ (summation of column 4)}}{40 \text{ (summation of column 2)}}$

   $= \underline{\mathbf{42.75}}$

IV. To calculate the mean from data with frequency and class interval with assumed mean:

In using the assumed mean, we use the formula

$$\overline{X} = A + \frac{\sum fd}{\sum f}$$

Where $\overline{X}$ = mean

A= an arbitrary mean, just assumed

d= deviation (X−A): difference between the assumed mean and the variables.

$\sum fd$= summation of the frequency (f) multiplied by the deviation (d).

$\sum f$= summation of all the frequencies.

Example: using an assumed mean of 40, calculate the mean for Table 11.

Solution:

1. with column 1, 2, 3 from Table 11, create column (4) = d
2. create column 5 = fd

Table 12

1 **Class Interval**	**2** **Freq (f)**	**3** **Class mark (X)**	**4** **d** **(X**−A)	**5** **fd**
20 – 24	2	22	22−40 = −18	−36
25 – 29	2	27	27−40 = −13	−26
30 – 34	8	32	32−40 = −8	−64
35 – 39	2	37	37−40 = −3	−6
40 – 44	7	42	42−40 = 2	14
45 – 49	8	47	47−40 = 7	56
50 – 54	5	52	52−40 = 12	60
55 – 59	4	57	57−40 = 17	68
60 – 64	2	62	62−40 = 22	44
	40		**18**	**110**

3. calculate using $\overline{X} = A + \frac{\sum fd}{\sum f}$

$$\overline{X} = 40 + \frac{110}{40}$$

$$\overline{X} = 40 + 2.75$$

$$\overline{X} = \underline{\mathbf{42.75}}$$

V. To calculate the mean from data with frequency and class interval with coding method:

In using the coding method, use the formula:

$$\overline{X} = A + \frac{(\sum fu)c}{\sum f}$$

Where $\overline{X}$ = mean

A= assumed mean

c= class interval

u= d/c: deviation (X−A) divided by class size

Example: using Table 12, calculate the mean using the coding method

Solution:

1. using column 1, 2, 3, 4 from Table 12, create column 5 = u
2. create column 6 = fu

Table 13

1 Class Interval	2 Freq (f)	3 Class mark (X)	4 D= (X−A)	5 U= d/c	6 Fu
20 – 24	2	22	22−40 = −18	−18÷5= −3.6	−7.2
25 – 29	2	27	27−40 = −13	−13÷5= −2.6	−5.2
30 – 34	8	32	32−40 = −8	−8÷5= −1.6	−12.8
35 – 39	2	37	37−40 = −3	−3÷5= −0.6	−1.2
40 – 44	7	42	42−40 = 2	2÷5= 0.4	2.8
45 – 49	8	47	47−40 = 7	7÷5= 1.4	11.2
50 – 54	5	52	52−40 = 12	12÷5= 2.4	12
55 – 59	4	57	57−40 = 17	17÷5= 3.4	13.6
60 – 64	2	62	62−40 = 22	22÷5= 4.4	8.8
	40		**18**	**3.6**	**22**

(Note: Class size is 5)

Using $\overline{X} = A + \frac{(\sum fu)c}{\sum f}$

$\overline{X} = 40 + \frac{(22)5}{40}$

$\overline{X} = 40 + 2.75$

$\overline{X} = \underline{\mathbf{42.75}}$

VI. To calculate the grand mean sometimes, the situation may arise where we know the mean of several groups of scores and we want to calculate the mean of all the scores combined. We calculate using the formula:

$$\overline{X}\text{ overall} = \frac{n_1\overline{X}_1 + n_2\overline{X}_2 + \ldots\ldots + n_k\overline{X}_k}{n_1 + n_2 + \ldots\ldots + n_k}$$

Where:

$n_1, n_2 \ldots n_k$ = number of scores in the first, second... last group

$\overline{X}_1, \overline{X}_2 \ldots \overline{X}_k$ = mean of groups 1, 2 ... last group respectively

Example: The mean of marks scored by 12 students of MSW I in statistics is 55, while the mean of 15 MSW II students is 70. 10 PGDSW students scored 65 as mean in statistics too. Find the mean scored mark for all students in the department of social work.

Solution:

Using the formula: $\overline{X}\text{ overall} = \dfrac{n_1\overline{X}_1 + n_2\overline{X}_2 + n_3\overline{X}_3}{n_1 + n_2 + n_3}$

where $\overline{X}_1 = 55$, $n_1 = 12$; $\overline{X}_2 = 70$, $n_2 = 15$; $\overline{X}_3 = 65$, $n_3 = 10$

$$\overline{X}\text{ overall} = \frac{12(55) + 15(70) + 10(65)}{12 + 15 + 10}$$

$$= \frac{660 + 1050 + 650}{37}$$

$$= \frac{2360}{37}$$

$\overline{X}$ overall = **63.784**

Summary

In this lecture, we have discussed one of the three major measures of central tendency, the mean. Arithmetic mean is the total of the values of a variable divided by the number of values. For the ungrouped data it is $\overline{X} = \dfrac{\sum x}{n}$ while for grouped data it is $\overline{X} = \dfrac{\sum fx}{\sum f}$. We also calculated mean for data with frequencies and class interval, with assumed means and with coding method.

Post – Test

1. What is Arithmetic mean?
2. Using Question 4 of the Post-test in Lecture Six, calculate the mean.

References

Adamu, S. O. and Johnson, T. L. (2004). *Statistics for Beginners*. Ibadan: Evans Brothers (Nig Publishers).

Hammed, Ayo (2002). *Statistics in Education: A Basic Text*. Ibadan: Stirling-Horden Publishers (Nig).

Harper, W. M. (1975). *Statistics*. London: Macdonald and Evans Ltd.

Gravether, F.J. and Forzano, L.B (2003). *Research Methods for the Behavioural Sciences*. U.S.A.: Wadsworth/Thomson Learning.

LECTURE EIGHT

Measures of Central Tendency: Median

Introduction

The second most frequently encountered measure of central tendency is the median. Usually, the median is used for data sets where the mean does not provide a good representative value. In a situation of extreme values in a set of data, the mean is not properly placed at the center; hence, the median is used to provide a better measure of central tendency.

Objectives

At the end of this lecture, you should be able to:

1. define what the median is; and
2. calculate the median, for grouped and ungrouped data

Pre – Test

1. What do you understand by the median in statistics?
2. What are the properties of the median?

CONTENT

The Median

Another measure of central tendency which is used in statistics is the median. The median, as the name implies, is that value in a set of observations which, after ordering according to magnitude, has an equal number of observations above and below it. This implies that 50 percent of the observations are above the median and the remaining 50 percent are below. The median also is called the 50^{th} percentile. It is a back-up measure of central tendency that is used in situations where the mean does not work well.

Properties of the Median

There are two properties of the median worth mentioning:

1. The median is less sensitive than the mean to extreme scores. Since the median is not responsive to each individual score, but rather divides the distribution in half, it is not as sensitive to extreme scores as is the mean.

2. Under usual circumstances, the median is more subjective to sampling variability than the mean but less subjective to sampling variability than the mode. Because the median is usually less stable than the mean from sample to sample, it is not as useful in inferential statistics.

I. Median of Ungrouped data

Median computations under ungrouped data can be done under two conditions:

i. Distribution having odd numbers of observations

The calculation of the median is very simple. For example, to calculate the median for a set of scores; 27, 40, 25, 21, 14 which are odd in number is thus: first, rearrange in descending or ascending order, 14, 21, 25, 27, 40. The middle value is the median that is 25.

ii. Distribution having even numbers of observations

If there had been even observation (for instance, 14, 21, 25, 27, 27, 40) then the median would be the summation of the two middle observations, divided by 2.

i.e. 14, 21, 25, 27, 27, 40. (after re-arrangement)

$$\text{Median} = \frac{\text{summation of the two middle observations}}{2}$$

$$= \frac{25 + 27}{2} = \frac{52}{2} = 26$$

II. The Median of Grouped Data

The median of a frequency distribution with class interval is illustrated with a worked example.

Find the median for table 11.

Table 14

1	**2**	**3**	**4**	**5**
Class Interval	**Frequency (f)**	**Class Mark**	**Cumulative frequency**	**Class Boundary**
20 – 24	2	22	2	19.5 – 24.5
25 – 29	2	27	4	24.5 – 29.5
30 – 34	8	32	12	29.5 – 34.5
35 – 39	2	37	14	34.5 – 39.5
40 – 44	7	42	21	39.5 – 44.5
45 – 49	8	47	29	44.5 – 49.5
50 – 54	5	52	34	49.5 – 54.5
55 – 59	4	57	38	54.5 – 59.5
60 – 64	2	62	40	59.5 – 64.5

Steps to solution

1. Retain columns (1), (2) and (3).
2. Create column (4), which is cumulative frequency column.
3. Create column (5) as class Boundary.
4. Calculate half of the total number of observations = $\frac{N}{2}$

Solution: Using the next formula to solve for table 14.

$$L + \left(\frac{\frac{N}{2} - fc}{fm}\right) C$$

Where

L = Lower boundary limit of the median class (39.5).

N = The total number of observations (40).

fc = Cumulative frequency up to the median class (14).

fm = Frequency of the median class (7).

C = Width of class interval (5).

i. First determine the median, i.e. $\frac{N}{2} = \frac{40}{2} =$ **20th** observation.

ii. 20th observation is used to determine the class interval after due consultation of the cumulative frequency.

iii. From column (4) above, 20th observation falls into the class interval 40- 44 and class boundary 39.5 - 44.5.

iv. The values are thus derived.

L = 39.5

N = 40

fc = 14

fm = 7

c = 5

v. Substituting:

$$= \quad 39.5 + \left(\frac{\frac{40}{2} - 14}{7}\right) 5$$

$$= \quad 39.5 + \left(\frac{20 - 14}{7}\right) 5$$

$$= \quad 39.5 + \left(\frac{6}{7}\right) 5$$

$$= \quad 39.5 + 4.286$$

Median = <u>43.786</u>

Summary

In this lecture, we have discussed the median as a measure of central tendency. Median is the scale value below which 50 percent of the scores fall. For the ungrouped data, it is the central value of observation after re-arrangement either in ascending or descending order. For the grouped data, it is

$$L + \left(\frac{\frac{N}{2} - fc}{Fm}\right)C$$

Post – Test

1. What is Median?
2. Using Question 4 of the Post-test in Lecture Six, calculate the median.

References

Adamu, S. O. and Johnson, T. L. (2004). *Statistics for Beginners*. Ibadan: Evans Brothers (Nig Publishers).

Hammed, Ayo (2002). *Statistics in Education: A Basic Text*. Ibadan: Stirling-Horden Publishers (Nig).

Harper, W. M. (1975). *Statistics*. London: Macdonald and Evans Ltd.

Gravether, F.J and Forzano, L.B (2003). *Research Methods for the Behavioural Sciences*. U.S.A.: Wadsworth/Thomson Learning.

LECTURE NINE

Measures of Central Tendency: Mode

Introduction

The mode is the third measure of central tendency, and it is the most frequently occurring score in a distribution. The mode is found by inspection of the scores.

Objectives

At the end of this lecture, you should be able to:

1. define what the mode is;
2. calculate the mode for both grouped and ungrouped data; and
3. differentiate between unimodal and bimodal data.

Pre – Test

1. What do you understand by the mode?
2. What are the properties of the mode?

CONTENT

The Mode

The mode is simply the most frequent score in the distribution simply put, the value that occurs with the highest frequency on a set of data is the mode. A data that contains only one mode is referred to as unimodal, while a distribution that has two modes is called bimodal. A distribution with more than two modes is called multimodal.

The mode does not lend itself easily for further algebraic or statistical manipulation; however, it can be used for both qualitative and quantitative data.

I. Mode of Ungrouped Data

The mode for ungrouped data is easy to determine and can be found almost by inspection.

Example 1: In the following observations: 1, 2, 1, 2, 3, 5, 2, 3, 7, 8, 8. 2 is the mode since 2 occurs most often in the distribution.

Example 2: If a distribution has two non-adjacent items that occur most frequently, then the distribution has two modes. It is called a bi-modal distribution. For example, the distribution whose items are 3. 5. 7, 5, 5, 9, 9, 9, 11, is a bi-modal distribution. Its modes are 5 and 9 respectively.

II. Mode of Grouped Data

In a group data we can obtain the mode by using the following formula:

$$\text{Mode} = L_1 + \left(\frac{d_1}{d_1 + d_2}\right)C$$

Where L = Lower boundary limit of modal class

C = width of the modal class

$(f_m - f_b) = d_1$ = difference between frequency of the modal class and the frequency of the next lower class

$(f_m - f_a) = d_2$ = difference between frequency of the modal class and the frequency of the next higher class.

f_m = frequency of the modal class

f_a = frequency above the modal class

f_b = frequency below the modal class.

For example, find the mode of Table 15.

Table 15

Class Interval	**Frequency**
20 – 24	2
25 – 29	2
30 – 34	8
35 – 39	5
40 – 44	4
45 – 49	12
50 – 54	5
55 – 59	2
60 – 64	1

Solution

From the above, we have the following:

$L_1 = 45$ (By observation, it is assumed that our mode will fall within the class with the highest frequency).

$C = 5$

$d_1 = f_m - f_b = 12 - 4 = 8$

$d_2 = f_m - f_a = 12 - 5 = 7$

Substituting: $= 45 + \left(\frac{8}{8+7}\right)5$

$= 45 + \left(\frac{8}{15}\right)5$

$= 45 + 2.667$

$= \underline{\mathbf{47.667}}$

Summary

In this lecture, we have discussed the mode as a measure of central tendency. The mode is the value that occurs most in a set of data. For the ungrouped data, it is the most occurred value in a set of data which could be unimodal or bimodal. For the ungrouped data it is

$$L_1 + \left(\frac{d_1}{d_1 + d_2}\right)C$$

Post–Test

1. What is the mode?
2. Using Question 4 of the Post-test in Lecture Six, calculate the mode.

References

Adamu, S. O. and Johnson, T. L. (2004). *Statistics for Beginners*. Ibadan: Evans Brothers (Nig Publishers).

Hammed, Ayo (2002) *Statistics in Education: A Basic Text*. Ibadan: Stirling-Horden Publishers (Nig).

Harper, W. M. (1975) *Statistics*. London: Macdonald and Evans Ltd.

Gravether, F.J and Forzano, L.B (2003). *Research Methods for the Behavioural Sciences*. U.S.A.: Wadsworth/Thomson Learning.

LECTURE TEN

Measure of Dispersions and Variability I

Introduction

In order to adequately describe data, we need, in addition to a measure of the centre of the data, like the mean, a measure of the variation of the data about this centre. That is, statisticians and researchers are often interested in knowing how data are spread in either direction, from the centre of a distribution. The most commonly used measures of variability are:

1. the range;
2. the quartile deviation/semi-inter quartile range;
3. the mean deviation;
4. the variance; and
5. the standard deviation

In this lecture, we shall deal with the range, the semi-inter quartile and the mean deviation.

Objectives

At the end of this lecture, you should be able to calculate:

1. the range;
2. the quartile deviation; and
3. the mean deviation.

Pre – Test

1. What is the range of these figures: 2, 8, 15, 18, and 20?
2. What does Q_1, Q_3 stand for in quartile deviation?
3. What is the purpose of mean deviation?

CONTENT

1. The Range

The range is the difference between the largest and the smallest observations in a set. In equation form: Range = $X_{max} - X_{min.}$

= Highest score –Lowest score

The range is a very simple measure of dispersion of data; it is easy for the layman to interpret. However, it is limited in value because it takes care of the highest and the lowest values but tells nothing about the variability of the data lying between those two values.

For example, in the set of the following scores, 3, 2, 9, 8, 15, 20, the range is obtained thus:

a. look for the highest score, which is 20; look for the lowest score, which is 2; and
b. deduct the lowest from the highest.

20 – 2

range = **<u>18</u>**

2. The Quartile Deviation/Semi-Inter quartile range

This takes care of the problem of the range by looking at the middle (50%) of the distribution without making use of the extreme scores.

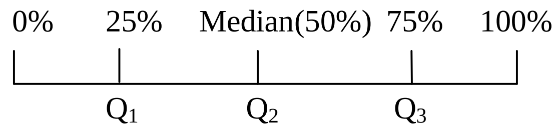
We define the three quartiles Q_1, Q_2 and Q_3 of a set of data as follows:

Q_1 = 25% of the data are less than or equal to the first quartile

Q_2 = 50% of the data are less or equal to the second quartile.

Q_3 = 75% of the data are less or equal to the third quartile.

The above can be represented diagrammatically thus:



Semi-Inter Quartile Range

i. Calculating Q_1 and Q_3

Recall the formula for finding Q_2 (median) and substituting Q_1, Q_3, we have the following formula:

$$Q_1 = L + \left(\frac{\frac{N}{4} - fc}{fM} \right) C$$ calculating 25% (1/4) ; first quartile

Locating Q_1 class interval is through $\frac{N}{4}$ th value

$$Q_3 = L + \left(\frac{\frac{3N}{4} - fc}{fM} \right) C$$ calculating 75% (3/4); third quartile

Locating Q_3 class interval is through $\frac{3N}{4}$ th value

(From Table 14)

$Q_1 = L = 29.5$	$Q_3 = L = 49.5$
$\frac{N}{4} = 10^{th}$ value	$\frac{3N}{4} = 30^{th}$ value
$f_c = 4$	$f_c = 29$
$f_m = 8$	$f_m = 5$
$C = 5$	$C = 5$

Substituting: $Q_1 = 29.5 + \left(\frac{10 - 4}{8} \right) 5$

$= 29.5 + \left(\frac{6}{8} \right) 5$

$= 29.5 + 3.75$

$= 33.25$

$Q_3 = 49.5 + \left(\frac{30 - 29}{5} \right) 5$

$= 49.5 + \left(\frac{1}{5} \right) 5$

$= 49.5 + 1$

$= 50.5$

ii. Calculating the Quartile deviation

The quartile deviation is one-half of the difference between the upper and the lower quartile (this is also referred to as the inter-quartile range). This is measured thus:

$$\frac{Q_3 - Q_1}{2}$$

Note that $\frac{Q_3 - Q_1}{2}$ is called the semi-inter quartile range

A better understanding will be obtained by working an example. Calculate the quartile deviation for Table 14.

Table 16

1 **Class Interval**	2 **Frequency (f)**	3 **Class Mark (X)**	4 **Cumulative frequency**	5 **Class Boundary**
20 – 24	2	22	2	19.5 – 24.5
25 – 29	2	27	4	24.5 – 29.5
30 – 34	8	32	12	29.5 – 34.5
35 – 39	2	37	14	34.5 – 39.5
40 – 44	7	42	21	39.5 – 44.5
45 – 49	8	47	29	44.5 – 49.5
50 – 54	5	52	34	49.5 – 54.5
55 – 59	4	57	38	54.5 – 64.5
60 – 64	2	62	40	59.5 – 64.5

After creating columns (3), (4), and (5) determine Q_1 and Q_3.

$Q_1 = \frac{N}{4}$

$Q_3 = \frac{3N}{4}$

Where N is total number of observation

Solution

Step 1: $Q_1 = \frac{N}{4} = \frac{40}{4} = 10^{th}$ value

Step 2: $Q_3 = \frac{3N}{4} = \frac{3(40)}{4} = 30^{th}$ value

$$Q.D = \frac{Q_3 - Q_1}{2}$$

$$= \frac{50.5 - 33.25}{2}$$

$$= \frac{17.25}{2}$$

$$= \mathbf{8.625}$$

3. The Mean Deviation

The mean deviation measures the dispersion of the data taking into consideration all the values in a set of data. It indicates how far the values are from the arithmetic mean or the median.

The mean deviation is measured thus:

$$MD = \frac{\Sigma |X - \bar{X}|}{n}$$

Where X = value of each observation

$\bar{X}$ = arithmetic mean of the value

n = number of observations in the sample

| | = absolute value.

Σ = Summation

a. The Mean Deviation for Ungrouped data

For Example: Find the mean deviation for the following data: 27, 40, 25, 20 and 13.

Solution: Step 1. Create column (1): x value

Step 2. Calculate the arithmetic mean.

Step 3. Create column (2): Mean Deviation.

Step 4. Create column (3): Modulus Value.

Step 5. Calculate the Mean deviation

Table 17

(1) X	(2) Deviation $X - \bar{X}$	(3) Deviation $\lvert x - \bar{x} \rvert$
27	27 – 25 = 2	2
40	40 – 25 = 15	15
25	25 – 25 = 0	0
20	20 – 25 = -5	5
13	13 – 25 = -12	12

Calculation: $M.D = \frac{2 + 15 + 0 + 5 + 12}{5} = \frac{34}{5} =$ **6.8**

b. The Mean Deviation for a grouped data

When frequencies are attached to grouped data, the formula is modified to read:

$$MD = \frac{\Sigma f \lvert X - \bar{X} \rvert}{\Sigma f}$$

Where f = frequency of class interval

X = midpoint of each class interval

$\bar{X}$ = mean of the group frequency distribution

Σf = Summation of the frequency

For Example: Find the mean deviation of the following data in Table 17.

Table 18

(1) Class Interval	(2) Frequency (f)	(3) Mid point X	(4) fx	(5) Deviation $(x - \bar{x})$	(6) $f /x - \bar{x}/$
5 – 9	25	7	175	7 – 16.05 = -9.05	226.25
10 – 14	18	12	216	12 – 16.05 = -4.05	72.9
15 – 19	20	17	340	17 – 16.05 = 0.95	19.0
20 – 24	14	22	308	22 – 16.05 = 5.95	83.3
25 – 29	18	27	486	27 – 16.05 = 10.95	197.1
	95		1525		598.55

Solution:

Step 1 – Create column 3: class midpoint X

Step 2 – create column 4: fx = column (2) x column (3)

Step 3 – calculate arithmetic mean = $\overline{X} = \frac{1525}{95} = 16.05$

Step 4 – create column 5: $(x - \overline{x})$

Step 5 – create column 6: f /x – x/

Step 6 – calculate MD = $\frac{\Sigma f \, |X - X|}{\Sigma f}$

$= \frac{598.55}{95}$

$= \underline{\mathbf{6.3}}$

Summary

In this lecture, we have discussed three measures of variability; the range, the quartile deviation and the mean deviation. The range is the difference between the values of the largest and the smallest observation scores. The mean deviation is the arithmetic mean of the absolute deviations from the arithmetic means. The quartile deviation is one-half of the difference between the upper (Q_3) and the lower quartile (Q_1).

Post – Test

1. What is the range of 2, 8, 15, 18, 20.
2. What do Q_1 and Q_3 stand for?
3. What is the purpose of mean deviation?
4. Find the mean deviation of the following set of numbers: 32, 34, 36, 38, 40
5. Calculate the quartile deviation for question 2 of the post-test in Lecture 4.

References

Adamu, S. O. and Johnson, T. L. (2004). *Statistics for Beginners*. Ibadan: Evans Brothers (Nig Publishers).

Hammed, Ayo (2002). *Statistics in Education: A Basic Text*. Ibadan: Stirling-Horden Publishers (Nig).

Harper, W. M. (1975). *Statistics*. London: Macdonald and Evans Ltd.

LECTURE ELEVEN

Measure of Dispersions and Variability II

Introduction

In the discussion of the mean deviation, we considered the differences of X_i – X. We had to take their absolute values; otherwise, their sum would have been zero. Another way of converting the differences X_i – X into positive quantities is to square them and add the squared differences. This sum, when divided by N (the number of observations), is called the variance of the observations.

In view of the above, this lecture focuses on the calculation of the variance and the standard deviation for both grouped and ungrouped data.

Objectives

At the end of this lecture, you should be able to calculate:

1. the variance for both grouped and ungrouped data; and
2. the standard deviation for both grouped and ungrouped data.

Pre – Test

1. What is the variance for 3, 4, 5, 6, 5, 7.
   a. 1.67 b. 2 c. 2.5 d. 1.41
2. What is the standard deviation if the variance is 22.7
   a. 10.23 b. 4.76 c. 11.35 d. 5.25

CONTENT

The Variance and the Standard Deviation

By definition, variance is the average of the sum of squared deviations from the arithmetic mean. The essence of squaring the deviation, as earlier mentioned, is to eliminate the problem that may be posed by the negative signs that will eventually give zero summation. Here the formula for calculating variance is,

a. Using Deviation Method b. using raw score method

$$S^2 = \frac{\sum_{i=1}^{N} (X - \overline{X})^2}{N} \quad \text{(for ungrouped data)} \quad S^2 = \frac{\Sigma X^2 - \frac{(\Sigma X)^2}{N}}{N}$$

$$S^2 = \frac{\sum_{i=1}^{N} f (X - \overline{X})^2}{N} \quad \text{(for grouped data)} \quad S^2 = \frac{\Sigma x^2 - \frac{(\Sigma X)^2}{N}}{N-1}$$

For the purpose of estimation, it is better and in common use to calculate variance by introducing Bessel's correction factor by dividing the squared mean deviation by N – 1 rather than N.

In this lecture, we shall calculate the variance both for grouped and ungrouped data using:

$$S^2 = \frac{\sum_{i=1}^{N} (X - \overline{X})^2}{N - 1} \quad \text{(for ungrouped data)}$$

$$S^2 = \frac{\sum_{i=1}^{N} f (X - \overline{X})^2}{N - 1} \quad \text{(for grouped data)}$$

The standard deviation is the positive root of the variance. Therefore, the same logic applies to this measure of variability. The formula for standard deviation is,

$$S = \sqrt{\frac{\sum_{i=1}^{N} (X_i - \overline{X})^2}{N - 1}} \quad \text{(Ungrouped data)}$$

$$S = \sqrt{\frac{\sum_{i=1}^{N} f(X_i - \overline{X})^2}{n - 1}} \quad \text{(Grouped data)}$$

The following steps are to be taken when finding either/both the variance and the standard deviation.

1. Find the arithmetic mean of the distribution (X)
2. Find the deviation from the mean (each value of X in raw scores, and mid-point value X for grouped data) $(X - \overline{X})$.
3. Square each deviation $(X - \overline{X})^2$.
4. Multiply each step (3) by the class frequency to obtain $f(x - \overline{X})^2$ (applicable only to group data).
5. Sum these products, by either step 3 for ungrouped or step (4) for grouped data.
6. Divide the sum of these squared deviations (i.e. step 5) by number of cases in the distribution less one (n – 1) to get the variance.
7. Find the square root of the variance (i.e. step 6) to obtain the standard deviation.

Properties of Standard deviation

The following are the main characteristics of standard deviation

i. It gives a measure of dispersion relative to the mean. This differs from the range, which gives an absolute measure of spread between the two most extreme scores.

ii. It is sensitive to each score in the distribution. If a score is moved closer to the mean, then the standard deviation will become smaller. Conversely, if a score shifts away from the mean, the standard deviation will increase.

iii. It is stable with regard to sampling fluctuations. This makes it useful for reporting variability.

iv. It can be manipulated algebraically, hence, the use in inferential statistics.

1. For Ungrouped Data (Without Frequency)

Find the variance and the standard deviation for the following scores in statistics. 5, 7, 8, 12, 18

Solution

Step 1. Find the mean = $\frac{5 + 7 + 8 + 12 + 18}{5} = \frac{50}{5} = 10$

Step 2. Find the deviation (column 2)

Step 3. Square each deviation (column 3).

Step 4. Sum the product of column 3 = 106

Step 5. Divide step 4 by n – 1 = $\frac{106}{5-1} = \frac{106}{4} = 26.5$ (variance)

Step 6. Find the square root of step 5 = 26.5 = 5.148 (Standard Deviation)

Table 19

(1) **X**	**(2)** $(X - \bar{X})$	**(3)** $(X - \bar{X})^2$
5	5 – 10 = -5	25
7	7 – 10 = -3	9
8	8 – 10 = -2	4
12	12 – 10 = +2	4
18	18 – 10 = +8	64
		106

$$\text{Variance} = \frac{\Sigma(x - \bar{X})^2}{n-1} = \frac{106}{4} = 26.5$$

Standard deviation = $\sqrt{\frac{\Sigma(x - \bar{X})^2}{n - 1}} = \sqrt{26.5} = 5.148$

Another way of working the variance and the standard deviation is

$S^2 = \sqrt{\frac{\Sigma X^2 - N\bar{X}^2}{N - 1}}$ = for ungrouped data

$S^2 = \sqrt{\frac{\Sigma f X^2 - N\bar{X}^2}{N - 1}}$ = for grouped data

Using Table 19 and the new formula, calculate the variance and the standard deviation.

Table 20

(1) **X**	**(2)** **X^2**
5	25
7	49
8	64
12	144
18	326
	606

Formula $S^2 = \sqrt{\frac{\Sigma X^2 - N\bar{X}^2}{N - 1}}$

Step 1. Calculate the arithmetic mean

$= \frac{5 + 7 + 8 + 12 + 18}{5} = \frac{50}{5} = 10$

Step 2. Find the value of X^2 (column 2)

Step 3. Sum-up the value in column 2 = 606

Step 4. Find N – 1 = 5 – 1 = 4; and N = 5

Substituting: $S^2 = \frac{606 - 5(10)^2}{5 - 1} = \frac{606 - 5(100)}{4}$

$$= \frac{606 - 500}{4} = \frac{106}{4} = 26.5 \text{ (Variance)}$$

$$= \sqrt{26.5} = \underline{\mathbf{5.148}} \text{ (Standard Deviation)}$$

2. Variance and Standard Deviation for Data with Frequency but not Grouped.

Find the standard deviation and the variance for the following age distribution of students in a class.

Age	No in Class
5	2
6	3
7	5
8	6
9	4

Table 21

Solution

(1) Age (X)	(2) No in Class (f)	(3) fx	(4) $(x - \overline{x})$	(5) $(x - \overline{x})^2$	(6) $f(x - \overline{x})^2$
5	2	10	5 – 7.35 = -2.35	5.5225	11.045
6	3	18	6 – 7.35 = -1.35	1.8225	5.4675
7	5	35	7 – 7.35 = -0.35	0.1225	0.6125
8	6	48	8 – 7.35 = 0.65	0.4225	2.535
9	4	36	9 – 7.35 = 1.65	2.7225	10.89
	20	**147**			**30.55**

Step 1. Calculate arithmetic mean $= \frac{\Sigma fx}{\Sigma f} = \frac{147}{20} = 7.35$

Step 2. Calculate the deviation: column 4

Step 3. Square column 4 to give column 5

Step 4. Multiply each manipulation in column 5 by column 2 to give column 6.

Step 5. Sum-up column 6 for $\Sigma f(x - \overline{x})^2 = 30.55$.

Substituting in Equation

$$S = \sqrt{\frac{\sum_{i=1}^{N} f (X - \overline{X})^2}{n - 1}}$$

$$= \sqrt{\frac{30.55}{20 - 1}} = \frac{30.55}{19} = 1.608 \text{ (Variance)}$$

$$= \sqrt{1.608} = 1.268 \text{ (Standard Deviation)}$$

Using an alternative method, calculate the standard deviation and the variance of table 21.

Solution: Using the formula:

$$S = \sqrt{\frac{\Sigma fx^2 - N(\overline{x})^2}{N - 1}}$$

Table 20

1	2	3	4	5
Age (X)	**No in Class (f)**	**Fx**	**x^2**	**fx^2**
5	2	10	25	50
6	3	18	36	108
7	5	35	49	245
8	6	48	64	384
9	4	36	81	324
	20	**147**		**1111**

Step 1. Calculate the mean = $\frac{\Sigma fx}{\Sigma f} = \frac{147}{20} = 7.35$

Step 2. Calculate X^2 (column 3)

Step 3. Calculate fx^2 (column 4) and sum-up = 1111

Substituting in the formula

$$S = \sqrt{\frac{\Sigma fx^2 - N(\overline{x})^2}{N - 1}}$$

$$= \sqrt{\frac{1111 - 20\,(7.35)^2}{20 - 1}} \qquad = \sqrt{\frac{1111 - 20\,(54.0225)}{19}}$$

$$= \sqrt{\frac{1111 - 1080.45}{19}} \qquad = \sqrt{\frac{30.55}{19}}$$

$$= \sqrt{1.608 \text{ (Variance)}}$$

$$= 1.268 \text{ (Standard Deviation)}$$

3. Co-efficient of Variation

It sometimes happens such that we need to compare the variability of two or more sets of figures, which are expected in different units. The co-efficient of variation then reduces the various measures into ordinary numbers so that comparisons can therefore be made. This reduces the various measures into ordinary numbers so that comparison can then be made. The popular ones are:

1. Co-efficient of quartile deviations

$$= \frac{\text{Quartile Deviation}}{\text{Mean}}$$

2. Co-efficient of mean deviation

$$= \frac{\text{Mean Deviation}}{\text{Mean}}$$

3. Co-efficient of standard deviation

$$= \frac{\text{Standard Deviation}}{\text{Mean}}$$

4. Co-efficient of Variation (CV)

$$= \frac{\text{Standard Deviation}}{\text{Mean}} \times 100\%$$

e.g. CV for Table 19 is

$$CV = \frac{SD}{\text{Mean}} \times \frac{100}{1}$$

$$= \frac{5.148}{10} \times \frac{100}{1}\% = \mathbf{\underline{51.48\%}}$$

Summary

In this lecture, we have examined the calculation of the variance and the standard deviation of both the grouped and ungrouped data. Variance is the average of the sum of squared deviations from the arithmetic mean. The standard deviation is the square root of the mean of the squared deviations.

Post – Test

1. What is the variance for 3, 4, 5, 6, 5, 7

   a. 1.67 b. 2 c. 2.5 d. 1.41
2. What is the standard deviation if the variance is 227

   a. 10.23 b. 4.76 c. 11.35 d. 5.25
3. Calculate the variance and the standard deviation for Q4 of the Post-Test of Lecture 10.

References

Adamu, S. O. and Johnson, T. L. (2004). *Statistics for Beginners*. Ibadan: Evans Brothers (Nig Publishers).

Hammed, Ayo (2002). *Statistics in Education: A Basic Text*. Ibadan: Stirling-Horden Publishers (Nig).

Harper, W. M. (1975). *Statistics*. London: Macdonald and Evans Ltd.

Solution to Post – Test

LECTURE ONE

Q3: X = mean, Σ = Summation

LECTURE FOUR

Q1:

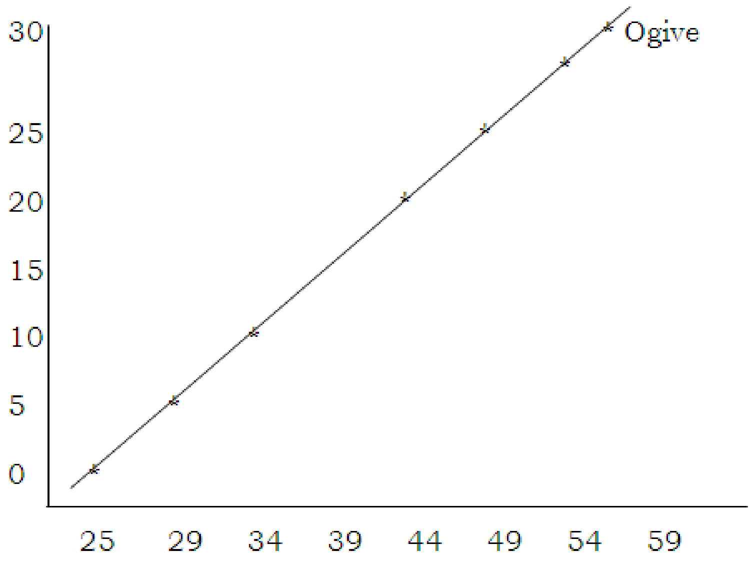
		Tally	Frequency
2	–	III	3
3	–	IIII	4
4	–	III	3
5	–	III	3
6	–	II	2
7	–	II	2
8	–	II	2
9	–	I	1

Q2:

Class Interval	**Tally**	**Frequency**	**Cum. Freq**
25 – 29	II	2	2
30 – 34	III	4	6
35 – 39	~~IIII~~ I	6	12
40 – 44	~~IIII~~ I	6	18
45 – 49	~~IIII~~	5	23
50 – 54	IIII	4	27
55 – 59	III	3	30

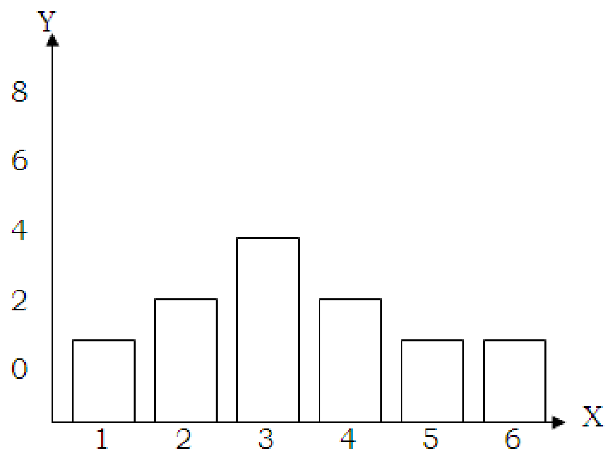
LECTURE FIVE: Solution to Post Test

Q3:



LECTURE SIX: Solution to Post – Test

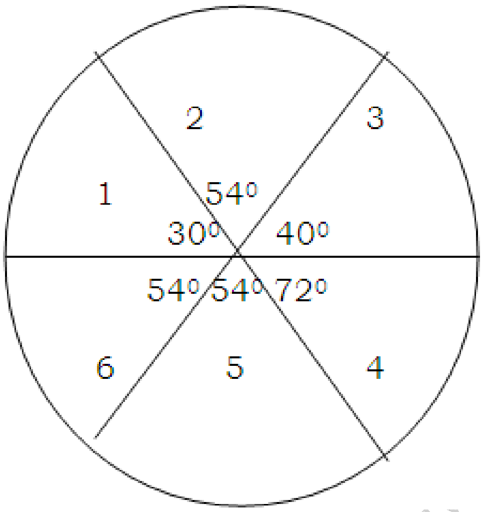
Q4:



Scale: 2cm rep 2 unit on Y – axis

2cm rep 1 unit on X – axis

Q5:



PIE CHART

LECTURE SEVEN - NINE: Solution to Post – Test

$$\text{Mean} = \frac{72}{20} = 3.6$$

$$\text{Median} = \frac{3 + 4}{2} = 3.5$$

Mode = 3

LECTURE TEN: Solution to Post – Test

Q1: 18

Q2: Lower quartile, upper quartile

Q4: 2.4

Q5: 6.625

LECTURE NINE: Solution to Post –Test

Q1: d = 1.41

Q2: b = 4.76

Q3: variance = 10; SD = 3.162