SAWA UNIVERSITY

# الاحصاء - SPSS

## المحاضرة الثالثة


جامعـــــــة ســـــــــاوة

الاهلية

كليـــــة التقنيـــات الصحيـة

والطبيـة

قسم تقنيات المختبرات

الطبية

المرحلـــــــــــــة الثانية.

تدريسي المادة : م.م سلام النجيب

**Sawa University**

**College of health and medical techniques**

**Department of Medical Laboratories**

**. 2nd Stage**

محاضرة رقم ٣

**Lecture No. 3**

الجانب النظري

**Theoretical**

# Introduction

- The data collected in a survey is called **raw data**.
- In most cases, useful information is not immediately evident from the mass of these raw data.
- Collected data need to be **organized** in such a way as to condense the information they contain in a way that will show patterns of variation clearly.
- Therefore, for the raw data to be more easily appreciated and to draw quick comparisons, it is often useful to present the data in the form of:

# Introduction

1. Ordered arrays – If the number of observations is not too large (usually less than 20), a first step in organizing these data is the preparation of an ordered array. It is an arrangement of the figures in increasing or decreasing order.
2. Tables – It An orderly and systematic presentation of data in rows and columns.
3. Graph – Graphs usually represent quantitative data on an 2D image with an x–y axis, using lines or bars

## A- Tabular presentation

As indicated above, tabular presentation refers to the systematic arrangement of data in rows and columns. A table should not be a misleading one. It should present a truthful impression of the data.

Parts of a table

a) Title : it explains

- what the data are about
- from where the data are collected
- time period when the data are collected
- how the data are classified

b) **Captions**: The headings of the columns are given in captions. In case there is a sub-division of any column, there would be sub-caption headings too.

c) **Stubs** : the headings of rows are called stubs.

d) **Body** : contains the numerical data.

e) **Head note** : a statement below the title which clarifies the content of the title.

f) **Foot note**: a statement below the table which clarifies some specific items given in the table. For example, it explains omissions, etc. in the table.

g) **Source** : the source of the data should be stated.

The title, body and captions/stubs are present in all tables while the presence of the other parts depends upon the type of data and specific purpose.

A study which involved 450 persons (aged 18 to 60 years) was undertaken to estimate the amount of caffeinated coffee taken everyday (on the average) in District Y, 1999. The following table gives the distribution of the data collected.

| The average number of cups of caffeinated coffee taken per day by each person | Frequency (number of persons) | Relative frequency (%) |
|---|---|---|
| 0 | 20 | 5.0 |
| 1 | 40 | 10.0 |
| 2 | 100 | 15.0 |
| 3 | 150 | 37.5 |
| 4 | 90 | 22.5 |
| 5 | 35 | 7.5 |
| 6 | 15 | 2.5 |
| Total | 450 | 100.0 |

# Grouped frequency distribution

Consider the problem of an investigator who wants to study the ages of persons who had car accidents during one year in a country. In connection with large sets of data, a good overall picture and sufficient information can often be conveyed by grouping the data into a number of class intervals as shown below..

| Age (years) | Number of persons |
|---|---|
| 18 – 24 | 4860 |
| 25 – 34 | 3240 |
| 35 – 44 | 1620 |
| 45 – 54 | 756 |
| 55 and over | 324 |
| Total | 10800 |

This kind of frequency distribution is called grouped frequency distribution.

# Grouped frequency distribution

Choosing suitable classification involves choosing the number of classes and the range of values each class should cover, following are some rules.

## 1) Determine the number of classes

The **number of classes (k)** can be the Sturge's Formula, given by:

$$K = 1 + 3.322 \times \log(n)$$

where n is the number of observations.

The length or **width of the class interval (w)** can be calculated by:

$$W = (\text{Maximum value} - \text{Minimum value}) / K = \text{Range} / K$$

$$\text{Range} = \text{maximum value} - \text{minimum value}$$

## B- Diagrammatic Representation of Data

- Appropriately drawn graph allows readers to obtain rapidly an overall grasp of the data presented.
- The relationship between numbers of various magnitudes can usually be seen more quickly and easily from a graph than from a table.
- Figures are not always interesting, and as their size and number increase they become confusing and uninteresting.

## Diagrammatic Representation of Data

- The choice of the particular graph among the different possibilities will depend largely on the type of the data.

- Bar chart
- Line graph
- Pie chart
- Histograms

There are, however, general rules that are commonly accepted about construction of graphs.

## Diagrammatic Representation of Data

- A proper scale should be selected and the units in to which The scale is divided should be clearly indicated.
- The vertical and horizontal scales should be clearly shown on The diagram itself – the former on the left hand side and the latter at the bottom of the diagram.
- The numerical scale representing frequency must start at zero or a break in the line should be shown.
- Titles of diagrams should be self explanatory. That is, the type of data (what), place that the data were collected (where), time period (when) and How the data were classified should be shown.
- Diagrams should be as simple as possible.
- Legends or keys should be used to differentiate variables if more than one is shown.
- Neatness should be strictly observed.
- Source should be given (if data are not original).

**Among the kinds of diagrams in common use are:**

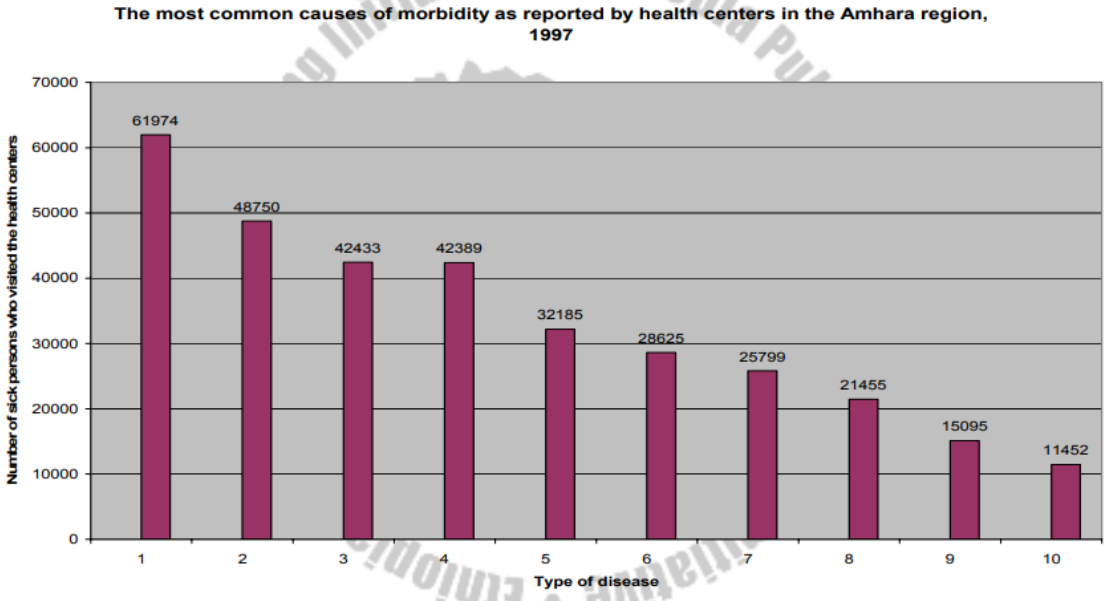
**1. Bar graph**

- Bar diagrams are used to represent and compare the frequency distribution of discrete variables and attributes of categorical series.
- There are different types of bar diagrams, the most important ones are:

A. A. Simple bar chart: It is a one-dimensional diagram in which the bar represents the whole of the magnitude. The height or length of each bar indicates the size (frequency) of the figure represented.

# Among the kinds of diagrams in common use are:

## 1. Bar graph

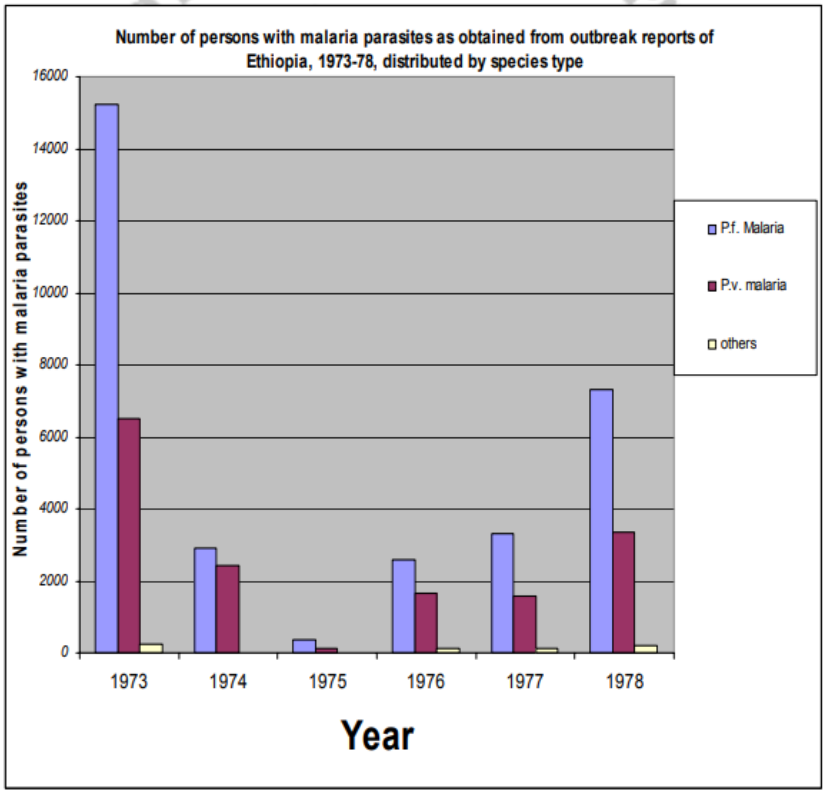



**Key**

| Type of disease | Code numbers given |
|---|---|
| Intestinal parasites | 1 |
| Malaria | 2 |
| Skin diseases | 3 |
| Upper respiratory tract infections | 4 |
| Pneumonia | 5 |
| Gastritis | 6 |
| Diarrhoea | 7 |
| STI (sexually transmitted infections) | 8 |
| Eye diseases | 9 |
| Tuberculosis | 10 |

## B. Multiple bar chart:

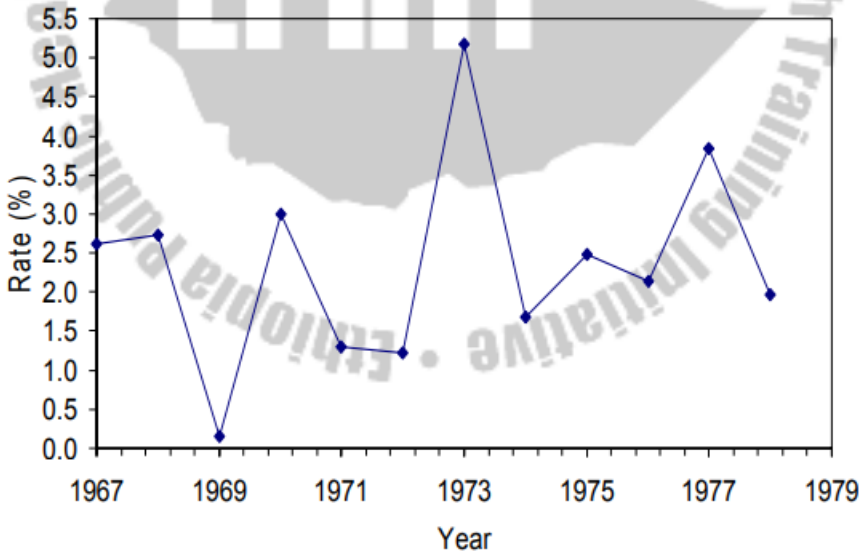
- In this type of chart the component figures are shown as separate bars adjoining each other.
- The height of each bar represents the actual value of the component figure.
- It depicts distributional pattern of more than one variable



## 2. The line graph

- The line graph is especially useful for the study of some variables according to the passage of time.
- The time, in weeks, months or years is marked along the horizontal axis; and
- the value of the quantity that is being studied is marked on the vertical axis.

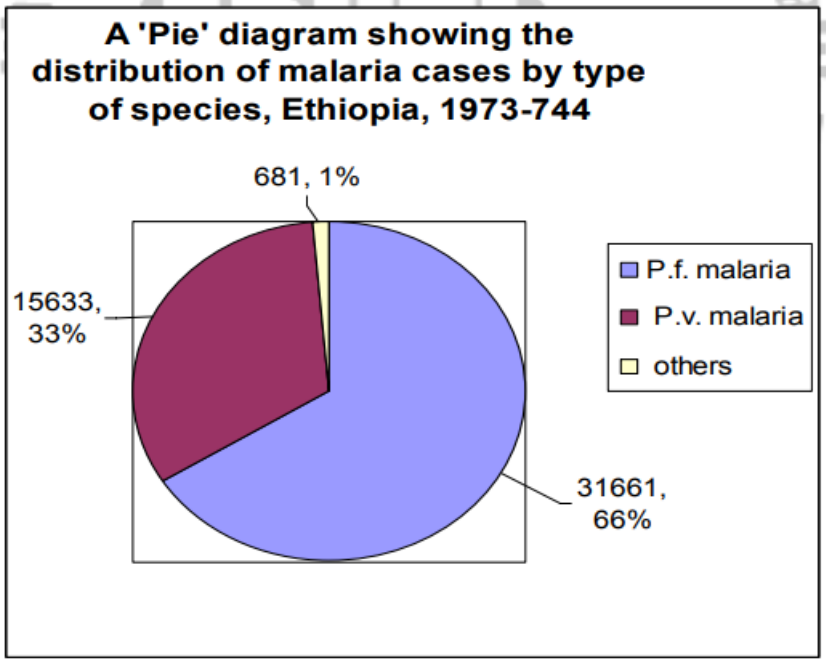
**Example**: Malaria situation of Ethiopia as obtained from malaria seasonal blood survey results,1967-79 E.C., classified by slide positivity rate and year.



# 3. 'Pie' chart (qualitative or quantitative discrete data):

A circle is divided into sectors so that the areas of the sectors are proportional to the frequencies.

Example: Distribution of malaria cases by species type, Ethiopia, 1973-78 (eth.c.).



## 4. Histograms (quantitative continuous data)

A histogram is the graph of the frequency distribution of continuous measurement variables. It is constructed on the basis of the following principles:

- The horizontal axis is a continuous scale running from one extreme end of the distribution to the other. It should be labelled with the name of the variable and the units of measurement.
- For each class in the distribution a vertical rectangle is drawn with (i) its base on the horizontal axis extending from one class boundary of the class to the other class boundary, there will never be any gap between the histogram rectangles. (ii) the bases of all rectangles will be determined by the width of the class intervals.

# 4. Histograms (quantitative continuous data)