



ITUC–Africa Capacity Building Workshop for Researchers

Session 6: Elementary Statistics for Trade Unions

Grayson Koyi
Institute of Economic and Social Research,
University of Zambia

12–16 April, 2010
Panafic Hotel, Nairobi



Module 1: Concept and role of statistics

- This module introduces the subject and focuses on:
 - Definition of statistics
 - Role of statistics
 - Two major divisions of statistics
 - Types of data
 - Strengths and weaknesses of subject of statistics
 - The language of statistics
 - Revision exercise



1.1 Statistics defined

- Statistics, as a word, connotes at least two meanings:
 - Numerical data expressed in quantitative terms
 - Techniques used to analyse data
- Formally, statistics is **the science concerned with gathering, analyzing and interpreting data for decision-making.**
 - Data can simply be defined as the **numerical facts or any materials that serve as a basis for drawing conclusions.**



1.2. Role of Statistics

- The role of statistics is to facilitate decision-making and informed action.
 - It does this through various ways. For instance:
 - By providing methods of bringing out the latent characteristics present in a set of numerical data
 - By providing methods for drawing inferences, or making broad generalisations.
- Statistics plays a significant role in such areas as planning, business forecasting, CB negotiations, time series analysis, regression & correlation analysis etc.



1.3 Divisions of Statistics

- Basically statistical methods involved in data analysis are divided into two categories,
 - descriptive statistics, and
 - inferential statistics.
- Descriptive statistics deals with collecting, summarizing and simplifying data;
 - presenting it in a convenient, usable and understandable form, often displaying it in tables or diagrams to illustrate the main features.



1.3 Divisions of Statistics

- Inferential statistics, goes beyond mere descriptions to methods used to draw conclusions & make broader generalizations about the population.
 - Obtaining a particular value from the sample information and using it for drawing an inference about the entire population underlies the subject matter of inferential statistics.
 - One consideration when employing inferential methods is that it is contingent upon the correct methodological framework (i.e. sampling plan).



1.4 Types of data

- In statistics, data are classified into two broad categories: Quantitative data and Qualitative data
 - Quantitative data are those that can be quantified in definite units of measurement(data in numbers).
 - Quantitative data can further be categorized as continuous or discrete data.
 - Continuous data represent the numerical values of a continuous variable e.g those of weight, length, height, thickness which represent continuous variables.
 - Discrete data are values assumed by a discrete variable. Discrete variables are measured in fixed numbers. Essentially emanate from counting.e.g no. of customers visiting a shop.



1.4 Types of data

- Qualitative data refer to qualitative characteristics of a subject or object (data in words).
- A characteristic is qualitative in nature when its observations are defined and noted in terms of the presence of a certain attribute in discrete numbers.
- These data are further classified as nominal and rank data



1.4 Types of data

- Nominal data are the outcome of classification into two or more categories of items or units according to some quality characteristic
 - E.g classification of student according to sex (as males or females), of workers according to skill (as skilled, semi-skilled, unskilled).
- Rank data, on the other hand, are the result assigning ranks to specify order e.g 1st 2nd 3rd.



1.4 Types of data

- The above data types / clarifications enable us to notice the following:
 - That quantitative data (both continuous and discrete) possesses the characteristic of being numerically expressed to a marked degree of precision:
 - They represent the values associated with the measurement of variables.
 - That qualitative data (nominal and rank) refer to variables which do not give rise to numerical values in the same way as quantitative variables.
 - These are the result of observing a variable with respect to a certain quality characteristic or an attribute.



1.5 Evaluating Statistics

- Being a subject of much practical utility and having wide-ranging applications, statistics displays a unique strength. It suffers from an important weakness as well
 - The strength:
 - The greatest strength lies in developing a statistical mode of thinking, imparting an orientation to the mind to think statistically. Imagine what would happen if government didn't collect data on such characteristics of the state as education, occupation, sex, age, marital status as are necessary for effective planning!



1.5 Evaluating Statistics

- The weakness: the general feeling of distrust in it is an important weakness.
 - It emanates from the general held view that the data, to which statistical methods are often applied, lack the desired element of accuracy.
 - To the extent that this apprehension may be taken as based on factual situations, the real culprits are those who compile, collect, and project data in a given light.



1.6 The language of statistics

- Notations for variables
 - Variable usually represented by Roman capital letter, X , the possible values that it may take X_1, X_2, X_3, \dots
- Notations for observed data
 - Observed values of X are represented by the Roman small letter x as x_1, x_2, x_3, \dots
- Notations in sampling
 - Population size denoted by N
 - Sample size denoted by n
- Notation for average
 - Population mean, μ
 - Sample mean



1.7 Revision exercise

1. Define statistics as a discipline. Also bring out its scope.
2. Differentiate between descriptive and inferential statistics.
3. State the various factors determining the classification of data.
4. Define a continuous and a discrete variable. Give suitable examples to show the difference between the two.



1.7 Revision exercise

5. Differentiate between nominal and ordinal data by citing suitable examples.
6. Justify that all qualitative data are essentially discrete data.
7. Evaluate statistics as a subject in terms of its basic strength and weakness.
8. What are the causes of mistrust in statistics? Is it really well-based?



Module 2

Sources of data



2.1 Sources of data

- Data sources could be seen as of two types, viz.,
 - Secondary
 - Primary
- A data source is secondary if it already contains the needed data in one form or other.
- A primary data source is invariably a sample or population survey undertaken solely with a view to collecting the needed data when the same are not available from an existing source.



2.1 Sources of data

- Thus, depending on this broad classification of possible data sources, statistical data may be distinguished as secondary data and primary data.
 - Secondary data are secondary in the sense that they already exist in some form– published or unpublished–in an identifiable secondary source.
 - Generally, they are available from published sources, though not necessarily in the form actually required.



2.1 Sources of data

- Primary data, on the contrary, are those that do not already exist in any form, and thus have to be collected for the first time from the primary source(s)
 - By their very nature, these data require fresh and first-time collection covering the whole population or a sample drawn from it.



2.2 Issues in secondary data

- The issues confronted in the collection of secondary data are not many. The task is relatively simple. It involves the following:
 - Identifying the sources of data, various publications where the required data may be found available.
 - Examining the available data, if these satisfy the needs of the proposed research investigation.
 - Compiling and re-organising the available data in the manner necessary for the investigation at hand.



2.2 Issues in secondary data

- To properly answer each of these concerns, it is important to be adequately clear about the objectives of the investigation.
 - Because it is the objective of your research that is going to help identify actual data requirements:
 - whether the existing data will suffice or you will require some kind of compilation and/or reorganisation in any form.
- Once the data requirements of a given research are clearly identified, it then becomes important to locate and reach relevant data sources.



2.2 Issues in secondary data

- Secondary data should be used with abundant caution and care.
 - They may not be available in the form required;
 - Their classification and tabulation may be limited, especially those of a specific-purpose secondary data.
 - However, they offer unique advantages;
 - Generally easy availability
 - Convenient to reach and access
 - Less costly in terms of collection
 - Time required in collection/compilation not long



2.3 Issues in primary data

- Two major issues which call for thoughtful decision in the collection of primary data, relate to
 - Choosing the right type of sampling method to identify the respondents.
 - Selecting an appropriate technique to use for the data collection.



2.3 Issues in primary data

- Issues relating to sampling have been discussed separately in earlier modules.
- Suffice to say no survey whether population or sample, can be undertaken without due consideration of cost, time and effort involved in data collection, and the reliability of data generated.



2.4 Practical exercise

- Explain by citing examples primary and secondary sources of data.
- Spell out the issues involved in the collection of secondary data.
- Write short notes on (a) cautions in using secondary data, and (b) merits and demerits of using secondary data.
- What are the advantages and disadvantages of using primary sources of data?



Modules 3

Data collection and analysis



3.1 Introduction

- Collection of data is the first important step with which statistics start.
- Once the data has been collected it has to be organised in a systematic manner for it to make sense.
 - This process is known as data analysis.
- The task of data collection covers both secondary and primary data



3.2 Data collection defined

- Data collection is the act of assembling & gathering the needed numerical information in the context of a specified research investigation.
- Given the objectives and research design of the investigation, the data required for investigation may be secondary or primary
 - Collection is involved in one form or the other in the case of both.
 - However, as discussed in the previous module, the issues involved in collecting secondary data are different from those concerning primary data.



3.3 Techniques of data collection

- A technique of data collection refers to the method by which we actually go about collecting the desired information in a survey
- As information is always elicited from the respondents, three alternative techniques of data collection have come to be adopted:
 - Personal interviews
 - Telephone interviews
 - questionnaire

3.3 Techniques of data collection



- An important note on primary data collection:
 - Remember that no survey based primary data offer adequately reliable results unless the research investigation at hand is properly conceived, effectively designed, and meticulously implemented.
 - Preparing survey instruments and getting them answered/filled require serious thought and imagination
 - Otherwise, the data collected may lack the needed reliability to an extent that seriously impairs the worth of the inferences that may be drawn.



3.4 Practical exercise

- What considerations must be borne in mind while preparing a questionnaire. Explain
- Briefly state the advantages and disadvantages of:
 - Personal interviewing
 - Telephone interview
 - Mailed questionnaire



Module 4

Data presentation and Interpretation



4.1 Introduction

- The foremost business after data collection is to present them in a meaningful way
- It means systematically arranging the raw data and abridging them in a convenient and easy-to understand form
- Data presentation is basically an exercise in ordering and smarting data with a view to revealing and highlighting the latent characteristics.



4.1 Introduction

- Thus, when presenting your data ensure that the data:
 - becomes manageable
 - appear more meaningful
 - offer more useful interpretations
 - reveal more clearly the broad trends in data variations



4.2 Major forms of data presentation

- Broadly, there are two forms of data presentation:
 - Tabular and graphic
 - This module will deal with tabular presentation while the next module looks at graphic



4.3 Tabular Presentation

- Tabular presentation means tabulating the data in the form of appropriate tables.
 - The basic function of the table is to simplify data and to present it in a manner that facilitates comparison
 - Simplifying data means that the information desired become easy to locate
 - Comparison involves bringing all related data together at one place such that a relational picture can be conveniently and efficiently drawn.



4.4 Types of tables

- The main types of tables used in data presentation are of the following type:
 - Single column or single row tables
 - Multiple column and multiple-row tables
 - Reference tables
 - Present extensive information on any subject.
 - For example, all data published in the annual survey of industries are in the form of reference tables which offer all relevant data on industries in detail.



4.5 Practical exercise

- Components of a table are functional parts that constitute the structure of the table.
 - Using table 4.0 provided, identify the eight components of a statistical table.



Module 5

Curves and diagrams



5.1 Introduction

- Graphic presentation means displaying tabular data visually, often in the form of various graphs



Types of graphs

- More popularly used graphs:
 - Simple bar chart
 - Figures one wishes to compare are represented by bars.
 - The height or length of a bar is proportional to the size of the figure being illustrated
 - Bar charts can be vertical or horizontal
 - Component bar chart
 - These are a special form of bar charts.
 - These charts are meant to exhibit the changes in the component or parts of a given total in relative terms
 - Pie chart
 - Pie charts are basically circle charts, which are usually drawn for component wise percent data



Illustrated Example: bar chart

- Suppose we have some data relating to numbers of cars sold each year, over a period of four years by Kenya motors.

Year	1990	1992	1994	1996
Nr. Of cars	10	20	30	15

Illustrated

example: component bar chart

- Draw a component bar chart based on the percentage shares of three traditional broad sector of GDP in country Z for period 2008 to 2010

	2007	2008	2009	2010
Tertiary	42.8%	42.8%	44.1%	43.9%
Secondary	23.7%	24.2%	25.3%	25.1%
Primary	33.5%	33.0%	30.6%	30.6%



Illustrated example: Pie chart

- Draw a pie chart based on sector wise data on GDP for country Z for the year 2007

	2007	2008	2009	2010
Tertiary	42.8%	42.8%	44.1%	43.9%
Secondary	23.7%	24.2%	25.3%	25.1%
Primary	33.5%	33.0%	30.6%	30.6%



Steps in drawing pie chart

- As the sum of angles in circle is 360 degrees, all percent components be converted into angles first
- Since all percent components add to 100, the corresponding component angles are computed by multiplying each component percent by the constant factor $360/100 = 3.6$
- Component angles are shown in the circle from 12 o'clock position, assuming the circle is a time clock.
- Moving clock wise from the starting point, the largest angle is displayed first, followed by the second and so on, in descending order.
- Each angle is depicted by using different shading or colour. What each colour mean is explained is explained in the margin of the pie chart.