

MATH 353: STATISTICS

Dr. Kofi Agyarko Ababio

kaababs@gmail.com

0244980077

Concept of Statistics

What is statistics

Statistics is the science
of collecting, organizing,
presenting, analyzing,
and interpreting
numerical data to assist
in making more
effective decisions.

μ

Σ

β

λ

σ

Uses of Statistics

Who Uses Statistics?

Statistical techniques are used extensively by

- ✓ Engineers
- ✓ Quality
- ✓ Hospital Administrators
- ✓ Accountants
- ✓ Educators
- ✓ Politicians
- ✓ Others.



Duties of Statistician

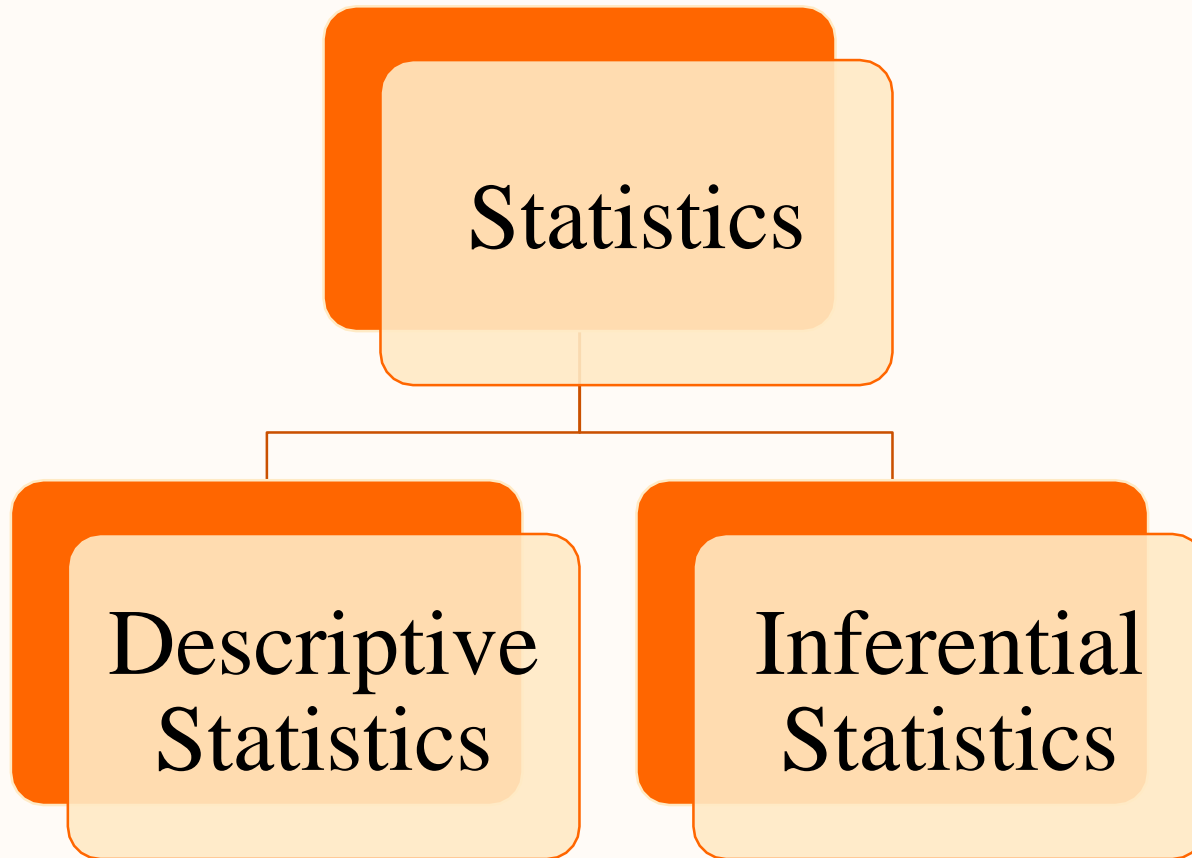
Duties of people with knowledge in statistics

- ✓ Design surveys and experiments to obtain data (un-processed information) at a minimum cost
- ✓ Seek for appropriate statistical method for analyzing the data
- ✓ Assist in making inference from the analysis
- ✓ Support decision making

Notable packages for statistical analysis

- **Statistical Package for Social Sciences (SPSS)**
- **SAS (Statistical Analysis System)**
- **Stata**
- **MINITAB**
- **MATLAB**
- **R**
- **GENSTAT**
- **Microsoft Excell**

Branches of Statistics



Branches of Statistics

Descriptive Statistics: consist of organization, summarisation, and presentation of data in an informative way. This includes tabular presentation, graphical and numerical summaries.

Inferential Statistics: consist of generalization of the results from sample to population. This includes estimation, hypothesis testing, prediction and determination of relationships among variables.

Definition of terms

A **Population** consist of all possible individuals, objects, or measurements of interest.

A **Sample** is a portion, or part, of the population of interest

A **variable** is a characteristic or attribute that can assume different values

Data are the values (measurements or observations) that the variable can assume.

Data set is a collection of data values. Each value in a data set is called a **data value** or **datum**.

Variable

A **variable** is a characteristic or attribute measured on a sample or population elements and can assume different values.

There are two types of variables

- ✓ Quantitative variable
- ✓ Qualitative Variable

Quantitative variables: are variables that assume numeric values.

- ✓ Example: weight of soil sample, number of sensors

Qualitative variables: are variables that assume non-numeric values.

- ✓ Example: Gender of a person, colour of soil sample,

Variable

The quantitative variable can further be classified into two groups:

- ✓ Discrete variable
- ✓ Continuous variable

Discrete variables: are variables that assume values that can be counted. They are obtained by counting.

- ✓ Example: number of sensors, number of vehicle crashes.

Continuous variables: are variables that assume all values between any two specific values. They are obtained by measuring

- ✓ Example: distance between two points, weight of soil sample, pressure in tire

Scale of Measurement

Data can also be classified according to how they are categorised, counted or measured

There are four scales of measuring data:

- ✓ Nominal scale
- ✓ Ordinal Scale
- ✓ Interval Scale
- ✓ Ratio scale

Scale of Measurement

Nominal Scale:

In this scale data are classified into mutually exclusive categories and can not be arranged in a certain order. This type of data are non-numeric.

Example:

- ✓ Colour of soil sample: brown, black, yellow
- ✓ Gender: male or female
- ✓ Marital status: single, married, divorced
- ✓ Employment status: employed, unemployed
- ✓ Religious affiliation

Scale of Measurement

Ordinal Scale:

In this scale data are classified into categories that can be ranked or ordered; however the precise differences between the ranks does not exist.

Example:

- ✓ Severity of injury: fatal, serious, minor, no injury
- ✓ Academic performance: Excellent, very good, poor
- ✓ Socio-economic status: High, middle, low
- ✓ Taste of food: Good, moderate, bad

Scale of Measurement

Interval Scale:

In this scale, data can be ranked and the precise differences between the ranks does exist; however, there is no meaningful zero

Example:

- ✓ Temperature: the difference between temperature of 10° and 50° indicates that one is warmer than the other. However, temperature of zero does not mean that there is no temperature
- ✓ IQ of a person: the difference between IQ of 50 and 110 indicates that one is more intelligent than the other. However, IQ of zero does not mean that the person has no intelligence.

Scale of Measurement

Ratio Scale:

The ratio scale of measurement possesses all the characteristics of interval scale and there is meaningful zero.

Example:

- ✓ Distance between two locations
- ✓ Weight of soil sample
- ✓ Area of a landfill
- ✓ Volume of a landfill

Data Collection

Data collection

- Data can be collected in a variety of ways. It is the most important part of statistical procedure because valid conclusions can only result from data which has been collected properly.
- If proper procedure is used to collect data then the issue of representativeness can be guaranteed. The use of wrong or faulty data collection methods would result in wrong conclusions because no good statistical tool can produce good results from wrongly collected data.
- Where do data come from?
- There are two main types of data:
 - ✓ Primary data
 - ✓ Secondary data

Secondary Source of Data

- **Secondary Data:** This refers to data that was collected by someone other than the user. That is, data collected for other research purposes.

- **Sources of secondary data include:**
 - ✓ Government Records
 - ☐ Census data,
 - ☐ population statistics,
 - ☐ health records,
 - ☐ educational institutes records
 - ✓ Private Organisations/Companies
 - ✓ Published materials (i.e. Books, Journals, websites)

Secondary Source of Data

➤ Advantages of using secondary data.

- ✓ It saves time
- ✓ It saves money
- ✓ It may be very accurate
- ✓ It has great exploratory value

➤ Limitations of using secondary data.

- ✓ It may be incomplete
- ✓ It may not be exactly what you need
- ✓ It may not be consistent/reliable
- ✓ It may be outdated

Primary Source of Data

- **Primary data:** This refers to data that was collected by the user. That is, data collected for the first time by the researcher for a defined purposes.

- **Sources of primary data:**

There are several methods of collecting primary data, and sometimes you have to think about creating a particular kind to suit your research.

- Three of the most popular methods are:
 - ✓ Experiment
 - ✓ Observation
 - ✓ Surveys

Primary Source of Data

➤ Advantages of using primary data.

- ✓ It gives original research quality and does not carry bias or opinion of third parties
- ✓ Ability to change the content or the course of study when ever needed
- ✓ What needed is what is obtained (if it is well designed)

➤ Limitations of using primary data.

- ✓ Difficulty in designing suitable approach
- ✓ Cost involve
- ✓ Time consuming

Choice of Data Sources

- Should I use secondary data or primary data?
- Ask your self certain questions?
 - ✓ Will the data answer my research questions?
 - ❑ You must first decide what your research questions are
 - ❑ Then you need to decide what variables are needed to answer the questions scientifically
- If that data exist in secondary form, then use them to the extent you can, keeping in mind limitations
- But if it does not, and you are able to fund primary data collection, then it is the method of choice.