



# Introduction to Biostatistics

Department of Epidemiology and Biostaistics

Alemakef Wagnew M.(Bsc. in statistics and MPH)

September 25, 2019

# introduction to Biostatistics

## **Course objective:**

- ▶ Definition and classification of Statistics
- ▶ Stages in statistical investigation
- ▶ Definition of Some Basic terms
- ▶ Applications, uses and limitations of Statistics
- ▶ Types of variables and measurement scales

# Chapter one: Introduction

## Definition and classification of Biostatistics

**Biostatistics** is a growing field with applications in many areas of biology including epidemiology, medical sciences, health sciences, educational research and environmental sciences.

# Introduction

## concern of Biostatistics

**Applied Statistics:** the application of statistical methods to solve real problems involving randomly generated data and the development of new statistical methodology motivated by real problems.

**Biostatistics** is the branch of applied statistics directed toward applications in the health sciences and biology.

**Biostatistics:** The tools of statistics are employed in many fields - business, education, psychology, agriculture, and economics, to mention only few. When the data being analyzed are derived from the public health data, biological sciences and medicine, we use the term biostatistics to distinguish this particular application of statistical tools and concepts.

# Introduction

## classification of Biostatistics

**Descriptive statistics:** A statistical method that is concerned with the collection, organization, summarization, and analysis of data from a sample of population.

**Inferential statistics** A statistical method that is concerned with the drawing conclusions/infering about a particular population by selecting and measuring a random sample from the population.

# introduction

## Descriptive Statistics

### Descriptive Statistics

Statistical procedures used to summarise, organise, and simplify data. This process should be carried out in such a way that reflects overall findings.

- ▶ Raw data is made more manageable
- ▶ Raw data is presented in a logical form
- ▶ Patterns can be seen from organized data

## descriptive statistics

- Some statistical summaries which are especially common in descriptive analyses are:
  - ▶ Measures of central tendency
  - ▶ Measures of dispersion
  - ▶ Measures of association
  - ▶ Cross-tabulation, contingency table
  - ▶ Histogram
  - ▶ Quantile, Q-Q plot
  - ▶ Scatter plot
  - ▶ Box plot

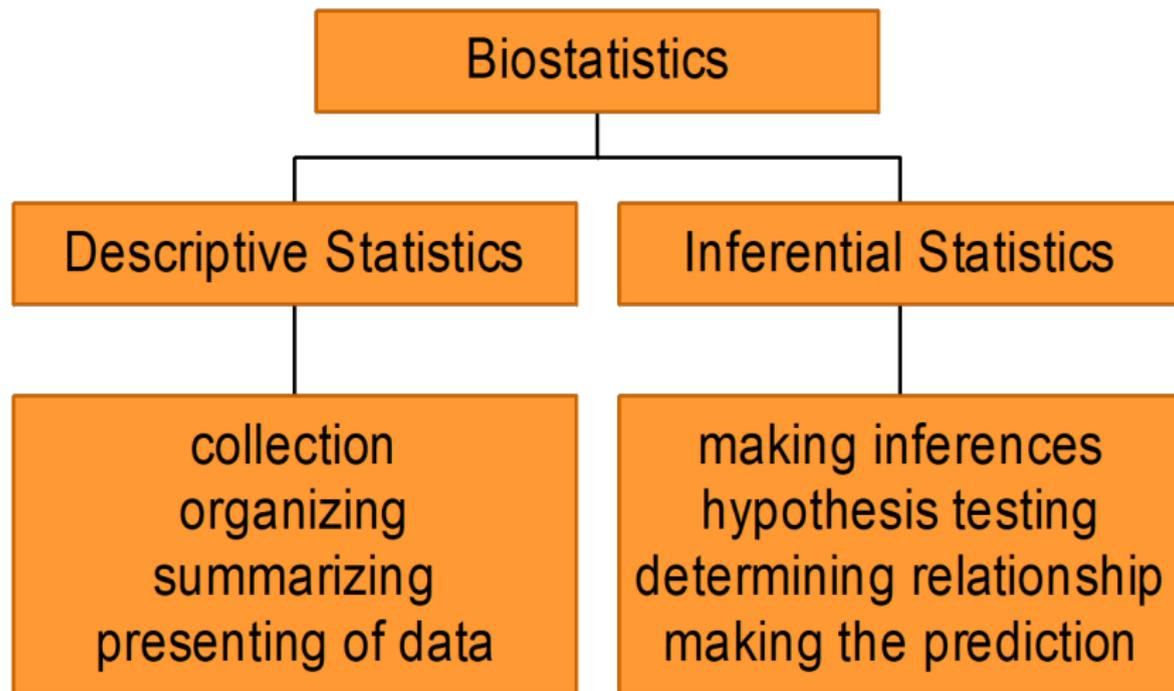
# Inferential Statistics

- This branch of statistics deals with techniques of making conclusions about the population
- Inferential statistics builds upon descriptive statistics
- The inferences are drawn from particular properties of sample to particular properties of population
- Inferential statistics are used to make generalizations from a sample to a population.
- They encompasses a variety of procedures to ensure that the inferences are sound and rational, even though they may not always be correct

# Inferential statistics

- In short, inferential statistics enables us to make confident decisions in the face of uncertainty
  - ▶ E.g. Antibiotics reduce the duration of viral throat infections by 1-2 days
  - ▶ Five percent of women aged 30-49 consult their GP each year with heavy menstrual bleeding

# Statistical Methods



## 1.2 Stages in statistical investigation

### **There are five stages or steps in any statistical investigation**

1. Collection of data
  - ▶ The process of obtaining measurements or counts.
2. Organization of data
  - ▶ Includes editing, classifying, and tabulating the data collected
3. Presentation of data
  - ▶ overall view of what the data actually looks like
  - ▶ facilitate further statistical analysis
  - ▶ Can be done in the form of tables and graphs or diagrams
4. Analysis of data
  - ▶ To dig out useful information for decision making
  - ▶ It involves extracting relevant information from the data (like mean, median, mode, range, variance. . . )
5. Interpretation of data
  - ▶ Concerned with drawing conclusions from the data collected and analyzed; and giving meaning to analysis results
  - ▶ A difficult task and requires a high degree of skill and experience

## definitions of Some basic terms

- Population
- Census
- Sample
- Parameter
- Statistic, Statistics
- Sampling
- sample size
- Variable
- Data

## 1.3 Definition of Some basic terms

**Population:** is the complete set of possible measurements for which inferences are to be made.

**Census:** a complete enumeration of the population. But in most real problems it cannot be realized, hence we take sample.

**Sample:** A sample from a population is the set of measurements that are actually collected in the course of an investigation.

**Parameter:** Characteristic or measure obtained from a population.

**Statistic:** A statistic refers to a numerical quantity computed from sample data (e.g. the mean, the median, the maximum...).

**Data:** Refers to a collection of facts, values, observations, or measurements that the variables can assume.

## 1.3 Definitions of Some basic terms

**Statistics:** is a branch of mathematics dealing with data collection, organization, analysis, interpretation and presentation.

**Sampling:**The process or method of sample selection from the population.

**Sample Size:**The number of elements or observation to be included in the sample.

**Variable:**It is an item of interest that can take on many different numerical values.

- Some examples of variables include:
  - ▶ Diastolic blood pressure,
  - ▶ heart rate, heights,
  - ▶ The weights,
  - ▶ Stage of bladder cancer patients,

## 1.4 Applications, Uses and Limitations of statistics.

### **Applications of Statistics**

- ▶ In almost all fields of human endeavor
- ▶ Almost all human beings in their daily life are subjected to obtaining numerical facts e.g. about price.
- ▶ Applicable in some process e.g. invention of certain drugs, extent of environmental pollution.
- ▶ In industries especially in quality control area

# Uses of Statistics

**The main function of statistics is to enlarge our knowledge of complex phenomena. The following are some uses of statistics:**

- i It presents facts in a definite and precise form.
- ii Data reduction.
- iii Measuring the magnitude of variations in data.
- iv Furnishes a technique of comparison.
- v Estimating unknown population characteristics.
- vi Testing and formulating of hypothesis.
- vii Studying the relationship between two or more variable.
- viii Forecasting future events

# Limitations of statistics

**As a science statistics has its own limitations. The following are some of the limitations:**

- I Deals with only quantitative information.
- II Deals with only aggregate of facts and not with individual data items.
- III Statistical data are only approximately and not mathematical correct.
- IV Statistics can be easily misused and therefore should be used by experts

## 1.5 Types of Variables and Measurement Scales

### **variable**

A variable is a characteristic or attribute that can assume different values in different persons, places, or things.

### **Example:**

- ▶ Age,
- ▶ Diastolic blood pressure,
- ▶ Heart rate,
- ▶ The height of adult males,
- ▶ The weights of preschool children,
- ▶ Gender of Biostatistics students,
- ▶ Marital status of instructors at University of Gondar,
- ▶ Ethnic group of patients

## Types of Variables 2

A. Depending on the characteristic of the measurement, variable can be:

### 1. **Qualitative(Categorical) variable**

- ▶ A variable or characteristic which cannot be measured in quantitative form but can only be identified by name or categories,
- ▶ for instance place of birth, ethnic group, type of drug, stages of breast cancer (I, II, III, or IV), degree of pain (minimal, moderate, sever or unbearable).
- ▶ The categories should be clear cut, not overlapping, and cover all the possibilities. For example, sex (male or female), vital status (alive or dead), disease stage (depends on disease), ever smoked (yes or no).

# Types of Variables 3

A. Depending on the characteristic of the measurement, variable can be:

## 2. Quantitative(Numerical) variable:

- ▶ is one that can be measured and expressed numerically.

### Example:

- ▶ survival time
- ▶ systolic blood pressure
- ▶ number of children in a family
- ▶ height, age, body mass index.

# Types of Variables 4

A. Depending on the characteristic of the measurement, variable can be:

**They can be of two types**

## 1. Discrete Variables

- ▶ Have a set of possible values that is either finite or countably infinite.
- ▶ The values of a discrete variable are usually whole numbers.
- ▶ Numerical discrete data occur when the observations are integers that correspond with a count of some sort

## 2. continuous variables

- ▶ A continuous variable has a set of possible values including all values in an interval of the real line.
- ▶ No gaps between possible values.
- ▶ Each observation theoretically falls somewhere along a continuum

# Types of variables 5

## Examples of discrete variables

- ▶ Number of pregnancies,
- ▶ The number of bacteria colonies on a plate,
- ▶ The number of cells within a prescribed area upon microscopic examination,
- ▶ The number of heart beats within a specified time interval,
- ▶ A mother's history of numbers of births ( parity) and pregnancies (gravidity),
- ▶ The number of episodes of illness a patient experiences during some time period, etc.

# Types of Variables 6

## Examples of Continuous variables

- ▶ Body mass index
  - ▶ Height
  - ▶ Blood pressure
  - ▶ Serum cholesterol level
  - ▶ Weigh,
  - ▶ Age etc...
- 
- Observations are not restricted to take on certain numerical values: Often measurements (e.g., height, weight, age)
  - Continuous data are used to report a measurement of the individual that can take on any value within an acceptable range

# Types of Variables 6

B. On the basis of Scales of measurement:

**There are four types of measurement scales:**

## 1. Nominal scales of measurement

- ▶ Only "naming" and classifying observations is possible. When numbers are assigned to categories, it is only for coding purposes and it does not provide a sense of size **Example:**
  - ▶ Sex of a person (M, F)
  - ▶ eye color (e.g. brown, blue)
  - ▶ religion (Muslim, Christian)
  - ▶ place of residence (urban, rural) etc

# Types of Variables 7

## B. On the basis of scale of Measurement

### 2.Ordinal Scales of Measurement

- Categorization and ranking (ordering) observations is possible
  - ▶ We can talk of greater than or less than and it conveys meaning to the value but;
  - ▶ Impossible to express the real difference between measurements in numerical terms

#### **Example:**

- ▶ Socio-economic status (very low, low, medium, high, very high)
- ▶ severity(mild, moderate, sever)
- ▶ blood pressure (very low, low, high, very high) etc.

# Types of variables 8

## B. On the basis of scale of Measurement

### 3.Interval Scales of Measurement

- ▶ Possible to categorize, rank and tell the real distance between any two measurements
- ▶ Zero is not absolute

#### **Example:**

- ▶ Body temperature in degree F. and Celsius (measured in degrees).
- ▶ It is a meaningful difference

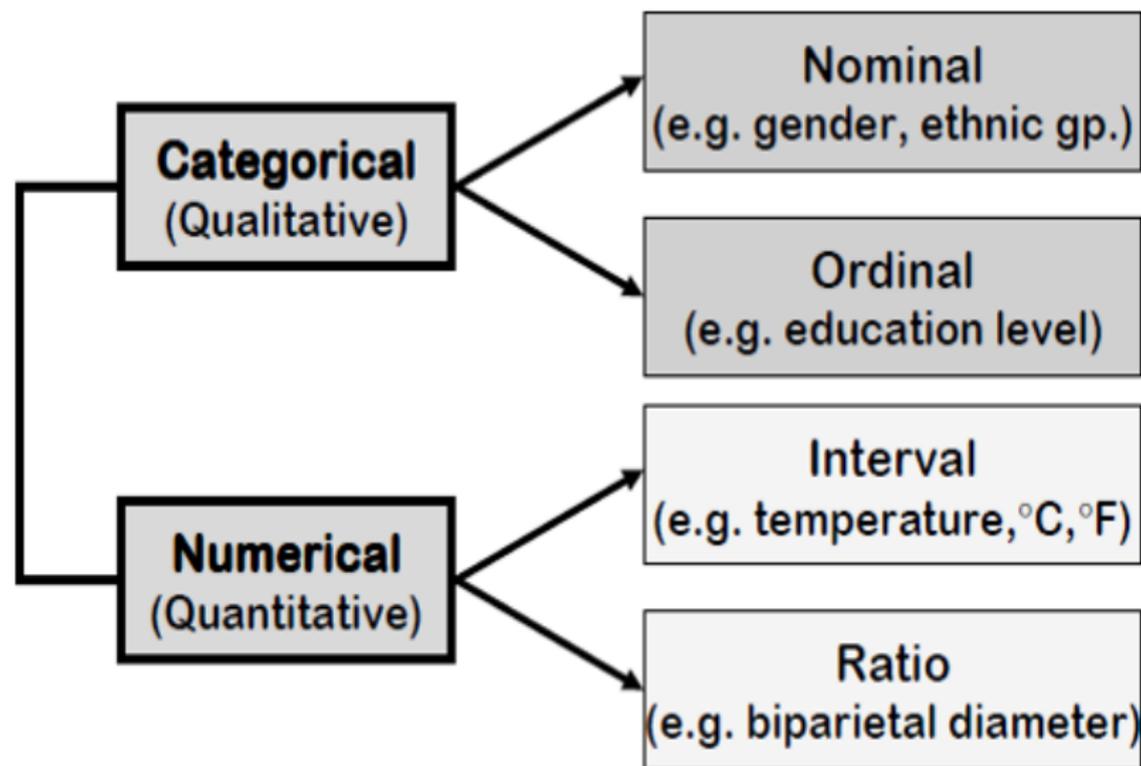
# Types of variables 9

## B. On the basis of scale of Measurement

### 4. Ratio scales of Measurement

- ▶ the highest level of measurement scale, characterized by the fact that equality of ratios as well as equality of intervals can be determined
- ▶ There is a true zero point. i.e. zero is absolute **Example:**
  - ▶ volume
  - ▶ height
  - ▶ weight
  - ▶ length
  - ▶ time until death, etc...

## Types of Variables 10



# Types of Variables 11

## C. On the basis of source of data

### 1. **Primary Data:**

- ▶ Data generated for the first time primarily/originally for the study in question
- ▶ It needs the involvement of the researcher himself. Census and sample survey are sources of primary types of data

### 2. **secondary Data:**

- ▶ Obtained from other pre-existing/ priorly collected sources
- ▶ In this case data were obtained from already collected sources like newspaper, magazines, DHS, hospital records and existing data like:
  - ▶ Mortality reports
  - ▶ Morbidity reports
  - ▶ Epidemic reports
  - ▶ Reports of laboratory utilization (including laboratory test results)

# Primary scale of Measurements

|                 |                                              |                                                                                   |                                                                                    |                                                                                     |
|-----------------|----------------------------------------------|-----------------------------------------------------------------------------------|------------------------------------------------------------------------------------|-------------------------------------------------------------------------------------|
| <b>Nominal</b>  | <b>Numbers assigned to runners</b>           |  |  |  |
| <b>Ordinal</b>  | <b>Rank order of winners</b>                 |  |  |  |
|                 |                                              | <b>Third Place</b>                                                                | <b>Second Place</b>                                                                | <b>First Place</b>                                                                  |
| <b>Interval</b> | <b>Performance rating on a 0 to 10 Scale</b> | <b>8.2</b>                                                                        | <b>9.1</b>                                                                         | <b>9.6</b>                                                                          |
| <b>Ratio</b>    | <b>Time to finish in 20 seconds</b>          | <b>15.2</b>                                                                       | <b>14.1</b>                                                                        | <b>13.4</b>                                                                         |

# Statistics

| STATISTICS |                                  |                                                             |
|------------|----------------------------------|-------------------------------------------------------------|
| SCALE      | DESCRIPTIVE                      | INFERENCEAL                                                 |
| Nominal    | Percentages, Mode                | Chi-square, Binomial test                                   |
| Ordinal    | Percentile, Median               | Rank-order, Correlation, ANOVA                              |
| Interval   | Range, Mean, SD                  | Correlations, t-tests, ANOVA<br>Regression, Factor Analysis |
| Ratio      | Geometric Mean,<br>Harmonic Mean | Coefficient of Variation (CV)                               |

## Exercise 1

**The following are list of different attributes/ variables or data. Classify the variables/data in to different measurement scales**

1. Your checking account number as a name for your account.
2. Your score on Bio-statistics test as a measure of your knowledge of Bios-tatistics.
3. A response to the statement "Abortion is a woman's right" where "Strongly Disagree" = 1, "Disagree" = 2, "No Opinion" = 3, "Agree" = 4, and "Strongly Agree" = 5, as a measure of attitude toward abortion.
4. Times for swimmers to complete a 50-meter race
5. Months of the year as September, October. . .
6. Economic status of a family when classified as low, middle and upper classes.
7. Blood type of individuals as A, B, AB and O.
8. Regions of Ethiopia as region 1, region 2, region 3. . .

## Assignment One

**Categorize the following variables into nominal, ordinal, interval or ratio**

- ▶ Gender
- ▶ Grade(A, B, C, D and F )
- ▶ Rating scale(poor, good, excellent)
- ▶ Eye color
- ▶ Political affiliation
- ▶ Religious affiliation
- ▶ Ranking of tennis players
- ▶ Major field
- ▶ Nationality
- ▶ Height
- ▶ Weight
- ▶ Time
- ▶ Age
- ▶ IQ
- ▶ Temperature
- ▶ Salary