جامعة المستقبل

كلية التقنيات الطبية والصحية

قسم تقنيات البصريات

# محاضرات التطبيقات الإحصائية 1

الكورس الأول

المحاضرة الأولى

المرحلة الثانية

2024-2023

م.د. ايناس رضا علي

# Introduction to biostatistics

## Introduction

We are frequently reminded of the fact that we are living in the information age. Appropriately, then, this subject is about information how it is obtained, how it is analyzed, and how it is interpreted. The information about which we are concerned we call data, and the data are available to us in the form of numbers.

The objectives are twofold:

(1) to teach the student to organize and summarize data, and

(2) to teach the student how to reach decisions about a large body of data by examining only a small part of it. The concepts and methods necessary for achieving the first objective are presented under the heading of descriptive statistics, and the second objective is reached through the study of what is called inferential statistics. These lectures discuss descriptive statistics. Because this lecture is designed for persons preparing for or already pursuing a career in the health field, the illustrative material and exercises reflect the problems and activities that these persons are likely to encounter in the performance of their duties.

## Some Basic Concepts

Like all fields of learning, statistics has its own vocabulary. Some words and phrases encountered in the study of statistics will be new to those not previously exposed to the subject. Other terms, though appearing to be familiar, may have specialized meanings that are different from the meanings that we are accustomed to associating with these terms. The following are some terms.

**1.Data:** The raw material of statistics is data. For our purposes, we may define data as numbers. The two kinds of numbers that we use in statistics are numbers that result from the taking in the usual sense of the term of measurement, and those that result from the process of counting. For example, when a nurse weighs a patient or takes a patient's temperature, a measurement, consisting

of a number such as 150 pounds or 100 degrees Fahrenheit, is obtained. Quite a different type of number is obtained when a hospital administrator counts the number of patients—perhaps 20—discharged from the hospital on a given day. Each of the three numbers is a datum, and the three taken together are data.

**2. Statistics:** The meaning of statistics is implicit in the previous section. More concretely, however, we may say that statistics is a field of study concerned with

(1) the collection, organization, summarization, and analysis of data; and
(2) the drawing of inferences about a body of data when only a part of the data is observed.

The person who performs these statistical activities must be prepared to interpret and to communicate the results to someone else as the situation demands. Simply put, we may say that data numbers, numbers contain information, and the purpose of statistics is to investigate and evaluate the nature and meaning of this information.

**3. Variable:** If, as we observe a characteristic, we find that it takes on different values in different persons, places, or things, we label the characteristic a variable. We do this for the simple reason that the characteristic is not the same when observed in different possessors of it. Some examples of variables include diastolic blood pressure, heart rate, the weights of preschool children, and the ages of patients seen in a dental clinic.

**4. Population:** The average person thinks of a population as a collection of entities, usually people. A population or collection of entities may, however, consist of animals, machines, places, or cells. For our purposes, we define a population of entities as the largest collection of entities in which we have an interest at a particular time. If we take a measurement of some variable on each of the entities in a population, we generate a population of values of that variable. We may, therefore, define a population of values as the largest collection of values of a random variable for which we have an interest at a particular time. If, for example, we are interested in the weights of all the children enrolled in a certain county elementary school system, our population consists of all these weights. If our interest lies only in the weights of first-grade students in the system, we have a different population—weights of first-grade students enrolled in the school system. Hence, populations are determined or defined by our sphere of interest. Populations may be finite or infinite. If a population of values consists of a fixed number of these values, the population is said

to be finite. If, on the other hand, a population consists of an endless succession of values, the population is an infinite one.

**5. Sample:** A sample may be defined simply as a part of a population. Suppose our population consists of the weights of all the elementary school children enrolled in a certain county school system. If we collect for analysis the weights of only a fraction of these children, we have only a part of our population of weights, that is, we have a sample.

## Types of data (variables)

The purpose of most studies is to collect data to obtain information about a particular area of research. Our data comprise observations on one or more variables; any quantity that varies is termed a variable. For example, we may collect basic clinical and demographic information on patients with a particular illness.

The variables of interest may include the sex, age, and height of the patients. Our data are usually obtained from a sample of individuals which represents the population of interest. Our aim is to condense these data in a meaningful way and extract useful information from them. Statistics encompasses the methods of collecting, summarizing, analyzing, and drawing conclusions from the data: we use statistical techniques to achieve our aim.

Data may take many forms. We need to know what form every variable takes before we can make a decision regarding the most appropriate statistical methods to use. Each variable and the resulting data will be one of two types: categorical or numerical (Fig. 1.1).

- **Categorical (qualitative) data:** These occur when each individual can only belong to one of a number of distinct categories of the variable.

1. Nominal data: the categories are not ordered but simply have names. Examples include blood group (A, B, AB, and O) and marital status (married/widowed/single, etc.).
2. Ordinal data: the categories are ordered in some way. Examples include disease staging systems (advanced, moderate, mild, none) and degree of pain (severe, moderate, mild, none).

- **Numerical (quantitative) data:** These occur when the variable takes some numerical value. We can subdivide numerical data into two types.

1. Discrete data: occurs when the variable can only take certain whole numerical values. These are often counted of numbers of events, such as the number of visits to a GP in a year or the number of episodes of illness in an individual over the last five years.
2. Continuous data: occurs when there is no limitation on the values that the variable can take, e.g. weight or height, other than that which restricts us when we make the measurement.

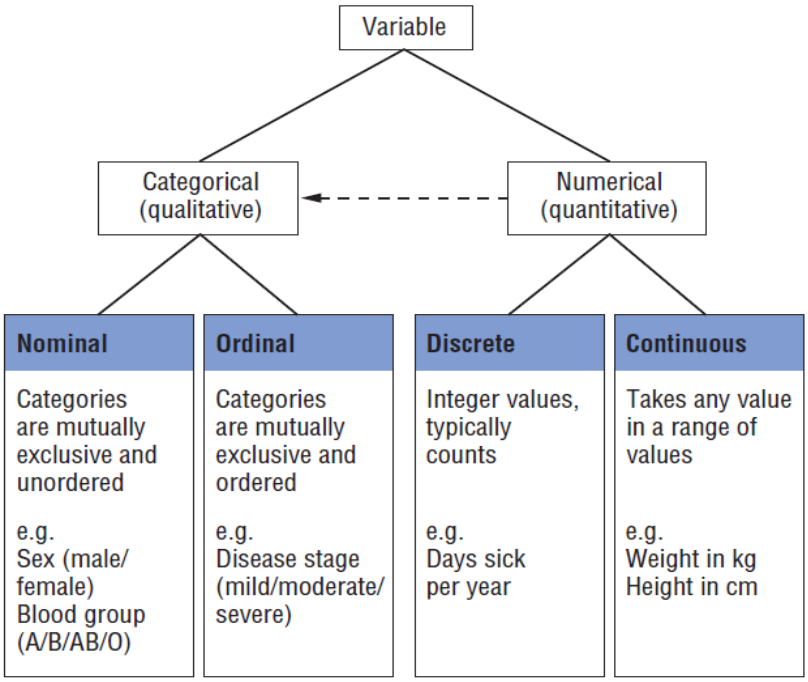


Figure 1.1 Diagram showing the different types of variables.

Categorical variable is binary or dichotomous when there are only two possible categories. Examples include 'Yes/No', 'Dead/Alive' or 'Patient has disease/Patient does not have disease'.

- **Derived data:** We may encounter a number of other types of data in the medical field. These include:

1. Percentages: These may arise when considering improvements in patients following treatment, e.g. a patient's lung function (forced expiratory volume in 1 second, FEV1) may increase by 24% following treatment with a new drug. In this case, it is the level of improvement, rather than the absolute value, which is of interest.

2. Ratios or quotients: Occasionally you may encounter the ratio or quotient of two variables. For example, body mass index (BMI), calculated as an individual's weight (kg) divided by her/his height squared (m2), is often used to assess whether s/he is over-or under-weight.
3. Rates: Disease rates, in which the number of disease events occurring among individuals in a study is divided by the total number of years of follow-up of all individuals in that study (Chapter 31), are common in epidemiological studies.
4. Scores: We sometimes use an arbitrary value, i.e. a score, when we cannot measure a quantity. For example, a series of responses to questions on the quality of life may be summed to give some overall quality of life score on each individual. All these variables can be treated as numerical variables for most analyses. Where the variable is derived using more than one value (e.g. the numerator and denominator of a percentage), it is important to record all of the values used. For example, a 10% improvement in a marker following treatment may have different clinical relevance depending on the level of the marker before treatment.

➢ **Censored data:** We may come across censored data in situations illustrated by the following examples.

1. If we measure laboratory values using a tool that can only detect levels above a certain cut-off value, then any values below this cut-off will not be detected. For example, when measuring virus levels, those below the limit of detectability will often be reported as 'undetectable' even though there may be some virus in the sample.
2. We may encounter censored data when following patients in a trial in which, for example, some patients withdraw from the trial before the trial has ended.

## Sources of Data

a. **Routinely kept records.**
b. **Surveys**.
c. **Experiments.**
d. **External sources.**
e. **Biostatistics.**

# Exercises

Q1: For each of the following variables indicate whether it is quantitative or qualitative variable:

(a)The blood type of some patients in the hospital. Qualitative Nominal

(b) Blood pressure level of a patient. Qualitative ordinal

(c) Weights of babies born in a hospital during a year. Quantitative continues

(d) Gender of babies born in a hospital during a year. Qualitative nominal

(e) The distance between the hospital to the house. Quantitative continues

(f) Under-arm temperature of day-old infants born in a hospital. Quantitative continues

---

Q2: For each of the following situations, answer questions a through d:

(a) What is the population?

(b) What is the sample in the study?

(c) What is the variable of interest?

(d) What is the type of the variable?

Situation A: A study of 300 households in a small southern town revealed that if she has a school age child present.

(a) Population: All households in a small southern town.

(b) Sample: 300 households in a small southern town.

(c) Variable: Do households have school-age children present.

(d) Variable is qualitative nominal.

Situation B: A study of 250 patients admitted to a hospital during the past year revealed that, on average, the patients lived 15 miles from the hospital.

(a) Population: All patients admitted to a hospital during the past year.

(b) Sample: 250 patients admitted to a hospital during the past year.

(c) Variable: Distance the hospital live away from the hospital

(d) Variable is Quantitative continuous.

---

Q3: Choose the right answer:

1. The variable is an

a. subset of the population.

b. parameter of the population.

c. relative frequency.

d. characteristic of the population to be measured.

e. class interval.

2. Which of the following is an example of discrete variable

a. the number of students taking statistics in this term at ksu.

b. the time to exercise daily.

c. whether or not someone has a disease

d. height of certain buildings

e. Level of education

3. Which of the following is not an example of discrete variable

a. the number of students at the class of statistics.

b. the number of times a child cry in a certain street.

c. the time to run a certain distance.

d. the number of buildings in a certain street.

e. number of educated persons in a family.

4. Which of the following is an example of qualitative variable

a. the blood pressure level.

b. the number of times a child brush his/her teeth.

c. whether or not someone fails in an exam.

d. weight of babies at birth.

e. the time to run a certain distance.

5.The continuous variable is a

a. variable with a specific number of values.

b. variable which cant be measured.

c. variable takes on values within intervals.

d. variable with no mode.

e. qualitative variable.

6. which of the following is an example of continuous variable

a. The number of visitors of the clinic yesterday.

b. The time to finish the exam.

c. The number of patients suffering from certain disease.

d. Whether or not the answer is true.

7. The discrete variable is

a-qualitative variable.

b-variable takes on values within interval.

c-variable with a specific number of values.

d-variable with no mode.

8-Which of the following is an example of nominal variable:

a-age of visitors of a clinic.

b-The time to finish the exam.

c-Whether or not a person is infected by influenza.

d-Weight for a sample of girls.

9-The nominal variable is a

a-A variable with a specific number of values

b-Qualitative variable that can't be ordered.

c-variable takes on values within interval.

d-Quantitative variable.

10-Which of the following is an example of nominal variable:

a-The number of persons who are injured in accident.

b-The time to finish the exam.

c-Whether or not the medicine is effective.

d-Socio-economic level.

11-The ordinal variable is:

a-variable with a specific number of values.

b-variable takes on values within interval.

c-Qualitative variable that can be ordered.

d-Variable that has more than mode.