Biostatistics

Variables

Dr Abubakir Majeed Saleh
Assist. Prof of Community Medicine

Outline

Types of variables

- First classification, by type
  - Qualitative (categorical) variables
    - Binary
    - Categorical
    - Ordinal
  - Quantitative variables
    - Continuous
    - Discrete
- Second classification, by relations
  - Dependent (outcome) variable
  - Independent (explanatory) variable

Data

- Data are the raw material of statistics.
- Simply defined as numbers.
- Statistic is anything calculated from the data.
- Two main kinds of data:
  - Result from <u>measurement</u> (e.g. body weight).
  - Result from <u>counting</u> (e.g. No. of patients).

Sources of data:

- Routinely kept records. E.g.: hospital medical records.
- Surveys
- Experiments
- External sources. E.g.: published reports, data banks, research literatures

Variables

- The term **variable** is used to mean a **quality** or **quantity** which varies from one member of a sample or population to another.

- **Quantity**: Blood pressure is a variable, which varies both from person to person and from measurement to measurement within the same person.

- **Quality**: Sex is a variable, people are either male or female.

Types of variables

- It is useful to think of data as being of several different types
- The type of data is important in deciding which methods of presentation and analysis we should adopt.
- First classification: Qualitative & Quantitative variables
- Second classification: Independent & Dependent variables

Types of variables
1st classification

Qualitative & Quantitative variables

Qualitative variables

- Qualitative data arise when individuals may fall into separate classes. E.g.:
  - Sex: male/female,
  - Severity of pain: mild/moderate/severe
  - Tobacco smoking: yes/no

- **3 main types**
  1. **Binary variables**
  2. **Categorical variables**
  3. **Ordered categorical variables**

Qualitative variables

1. Binary variables:

The values are of two different categories; e.g:

Sex: male/female

Disease status: having disease/not having disease

Smoking status: smoker/ not smoker

2. Categorical variables:

The values take several different categories that are distinct from each other; e.g.:

- Ethnic group: Kurd/Arab/Turkman/etc
- Marital status: single/married/widow/divorced

3. Ordered categorical variables:

The different categories are ordered on some scale; e.g.:

- Age groups: Child/Adolescent/Adult/Old
- Severity of disease: mild/moderate/severe

Quantitative variables

- Quantitative variables are **numerical**, arising from **measurements** or **counts** (measured on a well defined scale with units)

- **Measurements**

  If the values of the measurements can take any number in a range, such as height or weight, the data are said to be **continuous**. E.g.:
  - Weight - kg: 50.5, 51.6,52.2,53.8,etc
  - Blood pressure – 100,101,102,103,etc

- **Counts**

  If the values of the measurements can only take a few separate values, often integers (whole numbers) those data are said to be **discrete**. E.g.:
  - Family size – 2,3,4,5,6
  - Number of episodes of diarrhoea over 1 year – 0,1,2,3,4

Changing quantitative to qualitative variables

- Sometimes we change **continuous** or **discrete** variable to **categorical** (usually **ordered categorical**) variable for the sake of easy presentation or analysis

E.g.

- Age to categories of 10 years (0-10, 11-20, 21-30, etc)
- BMI (<18.5 underweight, 18.5-25 normal, 26-30 overweight, >30 obese)
- Number of pregnancies (0, 1-3, 4 and more)
- Hemoglobin level (low, normal, high)

What type of variable is each of the following?

• **BCG scar or not**	**Binary**
• **Height**	**Continuous numerical**
• **Child or adult**	**Binary**
• **Age (years)**	**Continuous numerical**
• **Social class** (poor, fair, wealthy)	**Categorical**
• **BMI** (<18, 18-25, 26-30, >30)	**Ordered categorical**
• **Number of pregnancies**	**Discrete numerical**
• **Job** (Governmental employee, private work, jobless)	**Categorical**

Types of data
2nd classification

- A variable can usually be one of two types:-

1. An outcome of interest.
   These are **outcome, response** or **dependent** variables

2. A factor that influences (or might influence) the outcome.
   These are often called **explanatory** or **independent** variables

Independent		**Dependent**
Smoking	→	Lung cancer
No physical exercise	→	Obesity (BMI)
High blood pressure	→	Heart attack
Reckless driving	→	Car accident
Alcohol drinking	→	Liver cirrhosis
New drug	→	Change in blood pressure

Examples of independent and dependent variables

- A study assessed the relation between high blood cholesterol and heart disease

  High blood cholesterol ------ Heart disease

- A research paper studied the effect of increasing age on blood cholesterol level

  Age --------- Blood cholesterol

- We studied the level of stress among medical students

  Study medicine ------ Stress

Questions on Variables?