EDP 613 Fall 2020

Chapter 1 Slides

Abhik Roy

Abhik.Roy@mail.wvu.edu

West Virginia University

Idea: Types of Data

Basic Definitions

- Information is collected on *elements* or *individuals*.
- The characteristics of the individuals about which we collect information are called *variables*.
- The values of the variables that we obtain are called *data*.

Qualitative and Quantitative Variables

Basic Definitions

- *Qualitative variables* or *Categorical variables* classify elements into categories.
- *Quantitative variables* troll how much or how many of something there is.

Example

Which of the following variables are qualitative and which are quantitative?

1. The name of the schools in your district.
2. The number of schools in your district.
3. The amount of each ingredient in a cake.
4. The ingredients in a cake.

Example Solutions

Which of the following variables are ordinal and which are nominal?

1. The name of the schools in your district. Qualitative
2. The number of schools in your district. Quantitative
3. The amount of each ingredient in a cake. Quantitative
4. The ingredients in a cake. Qualitative

Ordinal and Nominal Variables

Basic Definitions

- *Ordinal variables* are qualitative variables whose categories have natural ordering.
- *Nominal variables* are qualitative variables whose categories have no natural ordering.

Example

Which of the following variables are ordinal and which are nominal?

1. The (typical) letter grade distribution in a school.
2. Toppings on a cheeseburger.
3. Social economic status.
4. A telephone number.

Example Solutions

Which of the following variables are ordinal and which are nominal?

1. The (typical) letter grade distribution in a school. Ordinal
2. Toppings on a cheeseburger. Nominal
3. Social economic status. Ordinal
4. A telephone number. Ordinal

Discrete and Continuous Variables

Basic Definitions

- *Discrete variables* are quantitative variables whose possible values can be listed (possible infinite). (obtained by counting)
- *Continuous variables* are quantitative variables that can take on any value in some interval. (obtained by measuring)

Example

Which of the following variables are discrete and which are continuous?

1. Time it takes to get to school.
2. Water temperature.
3. Ratings on a 5-point rating scale.
4. Number of cars currently in a parking lot.

Example Solutions

Which of the following variables are discrete and which are continuous?

1. Time it takes to get to school. Continuous
2. Water temperature. Continuous
3. Ratings on a 5-point rating scale. Discrete
4. Number of cars currently in a parking lot. Discrete

Idea: Designs of Experiments

The textbook do not discuss these items directly but we will be using these ideas throughout the term.

Basic Definitions

- The *experimental units* are the individuals that are studies usually called subjects.
- The *outcome* or *response* is what is measured on each experimental unit.
- The *treatments* are the procedures applied to each experimental unit. The textbook definition is incorrect when it states that there are always two or more treatments. The purpose of a treatment is to measure an outcome.

Basic Definitions (cont.)

- A *randomized experiment* is a study in which the investigator assigns the treatments to experimental units at random.
- An *observational study* is one in which the assignment to treatment groups is not made by the investigator.

Randomized Experiments

Completely randomized experiment

There is no restriction on which subjects may be assigned which treatment (gold standard in experimental design).

Example

- Purpose: To determine how a new type of short wave UVA-blocking sunscreen affects the general health of skin in comparison to a regular long wave UVA-blocking sunscreen.
- Sample: 40 trial participants were randomly separated into equal groups of 20: an experimental group and a control group.

Example (cont.)

- Method: All participants' skin health was then initially evaluated.
  - Control group: wore the long wave UVA-blocking sunscreen daily.
  - Experimental group: wore the short wave UVA-blocking sunscreen daily.
- Analysis: After one year, the general health of the skin was measured in both groups are statistically analyzed.

Example (cont.)

- Result:
  - Control group: wearing long wave UVA-blocking sunscreen daily led to improvements in general skin health for 60% of the participants.
  - Experimental group: wearing short wave UVA-blocking sunscreen daily led to improvements in general skin health for 75% of the participants.

Double-blind Experiment

Neither the investigator nor the subjects know who had been assigned to which treatment

Example

- Purpose: To determine how a new treatment for colds affects users.
- Sample: 60 trial participants were randomly separated into equal groups of 30: an experimental group and a control group.

Example (cont.)

- Method:
  - Doctors in charge label the treatments with a code number (e.g. 101).
  - Patients receive a treatment labeled with a code number.
  - The nurses who give the treatments and record the responses know the treatment by its code number.
  - Neither the nurse nor the patients know if the treatment being imposed is an experimental drug or a placebo.

Example (cont.)

- Analysis: The patients in both groups are statistically analyzed. Only then they are placed into the pre-defined control and experimental groups.
- Result:
  - Control group: 1 patient deceased. 29 with no change.
  - Experimental group: 5 patients deceased. 8 patents admitted to the ICU. 12 patients with no change.

Randomized Block Experiment

Subjects are first divided into homogeneous blocks before they are randomly assigned to a treatment group.

Example

- Purpose: To determine if age might be a significant factor in the effect of a given medication.
- Sample: 100 trial participants.

Example (cont.)

- Method:
  - Researchers may choose to first divide the experimental subjects into age groups, such as under 30 years old, 30-60 years old, and over 60 years old.
  - Then, within each age level, individuals would be assigned to treatment groups using a completely randomized design.

Example (cont.)

- Analysis: The patients in both groups are statistically analyzed and then placed into the pre-defined control and experimental groups.
- Result:
  - Control group: $<$ 30 years: 12 normal, (30 years, 60 years): 1 deceased, 13 normal, $>$ 60 years: 4 deceased, 5 normal.
  - Experimental group: $<$ 30 years: 22 normal, (30 years, 60 years): 4 deceased, 32 normal, $>$ 60 years: 6 deceased, 2 normal.

Observational Studies

Cohort Study

- One or more samples (aka cohorts) are followed prospectively and subsequent status evaluations with respect to a disease or outcome are conducted to determine which initial participants exposure characteristics (risk factors) are associated with it.
- As the study is conducted, outcome from participants in each cohort is measured and relationships with specific characteristics determined.

Types of Cohort Studies

- In a *prospective cohort study*, the subjects are followed over time.
- In a *cross-sectional cohort study*, measurements are taken at one point in time.
- In a *retrospective cohort study*, the subjects are sampled after the outcome has occurred.

Example

- Purpose: To assess the impact of sun exposure on skin damage in beach volleyball players.
- Sample: Players from a weekend beach tournament.
- Method: Players from one team wore waterproof, SPF 35 sunscreen, while players from the other team did not wear any sunscreen.

Example

- Analysis: At the end of the volleyball tournament players' skin from both teams was analyzed for texture, sun damage, and burns and comparisons were made of skin damage were then made based on the use of sunscreen.
- Result: Significant difference between the cohorts in terms of the skin damage based on color, texture, and poisoning factors.

Case-control Study

- Patients who have a disease or outcome of interest (aka cases) with patients who do not have the disease or outcome (aka controls), and looks back retrospectively to compare how frequently the exposure to a risk factor is present in each group to determine the relationship between the risk factor and the disease.

Example

- Purpose: To investigate if exposure to zinc oxide is a more effective skin cancer prevention measure.
- Sample: 12 Lifeguards (of the same average age, lifeguard history for a similar number of seasons and amount of time per season) with (a) cheek and nose cancer and (b) ones without.
- Method: Retrospective interviews.

Example

- Analysis: Former lifeguards asked to recall which type of sunscreen they used on their face and approximately how often.
- Result: A survey of the differing sunscreens found that the 7 lifeguards with cheek and nose cancer did not use sunscreen with zinc oxide.

Confounding

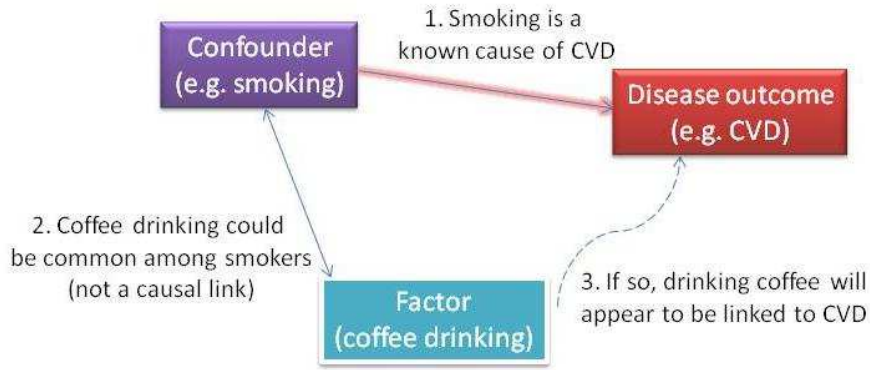
Observational studies have one major problem!: It is difficult to ascertain whether a difference in the outcome is due to the treatment or some other difference between the treatment and control groups.

Example

If drinking coffee is associated with smoking (perhaps people who drink coffee also tend to smoke) the smoking forms a confounding variable that can make it falsely appear as though coffee drinking leads to heart disease.

Example

If drinking coffee is associated with smoking (perhaps people who drink coffee also tend to smoke) the smoking forms a confounding variable that can make it falsely appear as though coffee drinking leads to heart disease.



Homework

Chapter 1 Exercises: 3, 4, 6, 7, and 8. Please turn it in to the Submission Portal on ecasmpus by 11:59 PM next Wednesday.