Chapter 4.2 Lecture Notes and Examples Part 1

Section 4.2 (Part 1) pp. 231-236
Experiments

A sample survey aims to gather information about a population without disturbing the population in the process. Sample surveys are a type of **observational study**.

1. Observational Study versus Experiment

Experiments do not just observe individuals or ask them questions. They actively impose some sort of *treatment* in order to measure the response. For example, "Does aspirin reduce the chance of a heart attack?"

> **Definition**:
> An **observational study** observes individuals and measures variables of interest but does not attempt to influence the responses.
>
> An **experiment** deliberately imposes some treatment on individuals to measure their responses.

- The goal of an *observational study* can be to describe some group or situation, to compare groups, or to examine relationships between variables.
- The purpose of an *experiment* is to determine whether a treatment causes a change in the response.

****

When our goal is to understand cause and effect, experiments are the only source of fully convincing data.

****

Example - For a long time, scientists have believed that soy foods in Asian diets explain lower rates of breast cancer, prostate cancer, osteoporosis, and heart disease in places like China and Japan. However, when experiments were conducted, soy either had no effect or a small effect on the health of the participants. For example, several studies randomly assigned elderly women to either soy or placebo, and none of the studies showed that soy was more beneficial for preventing osteoporosis.

So what explains the lower rates of osteoporosis in Asian cultures? **Genetics**? **Other dietary factors**? **Other differences between cultures**?

> **Definition**:
> A **response variable** measures an outcome of a study.
>
> An **explanatory variable** may help explain or influences changes in a response variable.

In this example, the *explanatory variable* was whether women got the soy or not, and the *response variable* was whether or not they got osteoporosis. In the **vocabulary of experimental design**, another name for explanatory variable is ***treatment***.

In this example, the effect of eating soy was mixed up with the characteristics of women who ate the soy. These characteristics are *confounding variables*.

> **Definition**: **Confounding** occurs when two variables are associated in such a way that their effects on a response variable cannot be distinguished from each other.

AP TIP! Note: If you are asked to identify a possible confounding variable in a given setting, you are expected to **explain how the variable you choose** (1) is associated with the explanatory variable, and (2) affects the response variable.

Example - A common definition of "binge drinking" is 5 or more drinks at one sitting for men and 4 or more for women. An observational study finds that students who binge drink have lower average GPA than those who do not. Identify a variable that may be confounded with the effects of binge drinking. Explain how confounding might occur.

Check Your Understanding

1) Does reducing screen brightness increase battery life in laptop computers? To find out, researchers obtained 30 new laptops of the same brand. They chose 15 of the computers at random and adjusted their screens to the brightest setting. The other 15 laptop screens were left at the default setting—moderate brightness. Researchers then measured how long each machine's battery lasted. Was this an observational study or an experiment? Justify your answer.

2) Does eating dinner with their families improve students' academic performance? According to an ABC News article, "Teenagers who eat with their families at least five times a week are more likely to get better grades in school. This finding was based on a sample survey conducted by researchers at Columbia University.

a) Was this an observational study or experiment? Justify your answer.

b) What are the explanatory and response variables?

c) Explain clearly why such a study cannot establish a cause-and-effect relationship. Suggest a variable that may be confounded with whether families eat dinner together.

When discussing confounding, it is best to explain what is happening in the context of the problem. Do not overly rely on statistical vocabulary without additional explanation. Just explain what you see happening.

2. The Language of Experiments

Definitions: A **treatment** is a specific condition applied to the individuals in an experiment. If an experiment has several explanatory variables, a treatment is a combination of specific values of these variables. The **experimental units** are the smallest collection of individuals to which treatments are applied. When the units are human beings, they are often called **subjects**.

Example - A study published in the *New England Journal of Medicine* compared two medicines to treat head lice: an oral medication called ivermectin and a topical lotion containing Malathion. Researchers studied 812 people in 376 households in seven areas around the world. Of the 185 households randomly assigned to ivermectin, 171 were free from head lice after two weeks compared to only 151 of the 191 households randomly assigned to Malathion.

Identify the experimental units, explanatory and response variables, and the treatments in this experiment.

The advantage of experiments over observational studies is that experiments can give good evidence for *causation*.

Sometimes, the explanatory variables in an experiment are called **factors**. Many experiments study the joint effect of several factors. In such an experiment, each treatment is formed by combining a specific value (called a **level**) of each of the factors.

Example - Does adding fertilizer affect the productivity of tomato plants? How about the amount of water given to the plants? A gardener plants 24 similar tomato plants in identical pots in his greenhouse. He will add fertilizer to the soil in half of the pots. Also, he will water 8 of the plants with 0.5 gallon of water per day, 8 of the plants with 1 gallon, and the remaining 8 with 1.5 gallons. At the end of three months he will record the weight of tomatoes produced on each plant.

Identify: (a) explanatory and response variables, (b) experimental units, (c) all treatments.

3. How to Experiment Badly

A bad experiment: A high school regularly offers a review course to prepare students for the SAT. This year, budget cuts will allow the school to offer only an online version of the course. Over the past 10 year, the average SAT score of the students in the classroom course was 1620. The online group gets an average score of 1780. That's roughly 10% higher than the long-time average for those who took the classroom review course. Is the online course more effective?

This experiment has a very simple design. A group of subjects, (the students) were exposed to a treatment (the online course), and the outcome (SAT scores) was observed. Here is the design:

Students -> Online Course -> SAT Scores

A closer look showed that the students in the online review course were quite different from the students who took the classroom course in the past years. They had higher GPAs and were taking more AP classes.

The effect of online versus in-class instruction is mixed up with the effect of these confounding variables. Maybe the online students earned higher SAT scores because they were smarter to begin with, not because the online course prepared them better. This confounding prevents us from concluding that the online course is more effective than classroom instruction.

HW 4.2 Part A page 253 – 254 45 – 53 odd, and page 230 37 - 42