

Types of Statistical Studies

Alignments to Content Standards: S-IC.B.3

Task

The following are some common methods of data collection in statistical studies that involve people.

Broadly speaking, a **survey** is a way of learning about a group of people by having some of the people in the group answer questions. A survey might involve completing a form, completing a questionnaire online, participating in a personal interview, or answering questions over the phone.

An **observational study** is a type of study where a researcher uses observed information to learn about a group of individuals. The individuals being observed are specifically **not** interfered with aside from the measuring of their responses. For example, if we wanted to learn if a person's hair color is related to his/her favorite soda, we would record values of the variables "hair color" and "favorite soda" from observation or from reported observations, such as through a survey. (Surveys are one type of observational study.)

A **sample survey** is a survey that is carried out using a sample of people who are intended to represent a larger population. For example, if we wanted to know more about the opinions of the one million adults who live in a particular city, and if we selected a *sample* of 900 individuals from those one million people, a survey conducted using those 900 people would be considered a *sample* survey. Well designed sample surveys can provide data that can be used in a variety of ways. For example, a survey might be used to assess the views of the general public, to examine trends in customer behavior, or to estimate the values of population characteristics when a census of an entire population is not practical.

An **experiment** is different from an observational study in that a researcher deliberately imposes different **treatments** on different groups and then compares the groups to determine if there is any difference in a specific variable of interest (called a **response**). A well designed experiment allows researchers to decide if the different treatments result in different responses.

1) Using the descriptions above, determine if each study described below is a sample survey, an observational study, or an experiment.

a) A pharmaceutical company is interested in comparing three brands of pain relievers. The company wants to know if one brand is much better than the other brands in terms of mobility improvement. Three different groups of 30 people participate in the study. Each group gets one of the three brands of pain relievers. The improvement in mobility of these 90 people is recorded after 5 weeks of taking the medicine.

b) Medical records of a group of long-time residents of a town are compared to the medical records of a group of long-time residents of another town. Researchers are curious to see if one town appears to be “healthier” in general.

c) A group of 600 registered voters in a given county are asked how they intend to vote in an upcoming election. A summary of their responses is posted on a news web-site, and it is implied that this group is representative of all registered voters in that county. The responses are used to predict the outcome of the election.

d) In a study to see if administering a medication orally or by injection makes a difference in how quickly people with back pain feel relief, a group of 400 adults suffering from back pain is divided into two groups. One group of 200 people is given the medication in pill form, while the other group of 200 people is given the medication in the form of a shot. Results for each group are compared to see if there is a difference in pain relief between the two methods.

Why Randomization Matters

Randomization plays an important part in data collection in two distinct ways:

Random Selection – In the case of a sample survey or observational study, we want the sample of people involved to be representative of their greater population. To help achieve this, we want the people to be *impartially* selected so that we do not introduce

any favoritism in the selection which could potentially distort conclusions drawn from the data.

Random Assignment – In the case of an experiment, we want to *impartially* assign the people (called **subjects**) to the various treatment groups. If we do not do this, we might inadvertently create a situation where any observed differences between our groups may have been caused by some factor other than the treatments we are interested in.

2) The four studies mentioned in Question 1 are presented here again. However, additional information is now provided regarding how the people involved were chosen or assigned. Each of these studies contains a flaw in terms of how the people involved were chosen or assigned. For each study, explain why the data collection method might cause a problem, and how randomization should be properly employed.

a) A pharmaceutical company is interested in comparing three brands of pain relievers. The company wants to know if one brand is much better than the other brands in terms of mobility improvement. Three different groups of 30 people participate in the study. Each group gets one of the three brands of pain relievers. The improvement in mobility of these 90 people is recorded after 5 weeks of taking the medicine. **The first group of 30 patients were patients of Dr. Smith, the second group of 30 patients came from Dr. Jones, and the last group of 30 came from Dr. McGillicuddy.**

b) Medical records of a group of long-time residents of a town are compared to the medical records of a group of long-time residents of another town. Researchers are curious to see if one town appears to be “healthier” in general. **In order to save time and money, all of the records were selected from the largest of the four hospitals in Town X and from the largest of the three hospitals in Town Y.**

c) A group of 600 registered voters in a given county are asked how they intend to vote in an upcoming election. A summary of their responses is posted on a news web-site, and it is implied that this group is representative of all registered voters in that county. The responses are used to predict the outcome of the election. **The 600 voters submitted their opinions via a webpage and were able to choose to participate after seeing an advertisement on television.**

d) In a study to see if administering a medication orally or by injection makes a difference in how quickly people with back pain feel relief, a group of 400 adults suffering from back pain is divided into two groups. One group of 200 people is given

the medication in pill form, while the other group of 200 people is given the medication in the form of a shot. Results for each group are compared to see if there is a difference in pain relief between the two methods.. **People were allowed to select whether they would take the medicine orally or by shot. Once 200 people had filled up one of the groups, all remaining people were immediately placed in the other group.**

IM Commentary

The purpose of this task is to provide students with experience distinguishing between the various types of statistical studies and to understand the purpose of random selection in surveys and observational studies vs. random assignment to treatments in experiments. Students should recognize that data collection methods differ and that all surveys do not necessarily properly represent a larger group. Students should also note that randomization brings a needed impartiality to data collection and that the word “random” does not mean “out of the ordinary” or “haphazard” (e.g. “that was random”). As a segue to more robust discussions regarding data collection, students can also begin to appreciate that since there is no random *assignment* in observational studies, uncontrolled factors are far more likely to be potentially responsible for any observed relationships between two variables.

[Edit this solution](#)

Solution

1.

a) Experiment – treatments are imposed on separate groups of people.

b) Observational Study – the people involved are not interfered with. (Note: an answer of “sample survey” may also be considered as correct if the student thinks of this as a survey of medical records.

c) Sample Survey – the 600 individuals are intended to represent a larger population.

d) Experiment – treatments are imposed on separate groups of people.

2) Answers may vary. Sample answers are shown here.

a) It could be that the one doctor has already treated his/her individual patients more effectively than the other doctors, or it could be that one doctor tends to get the patients who are in the most pain for some reason (e.g., referrals). To avoid these potentially confounding factors, we should randomly assign the 90 patients to the three groups of pain relievers.

b) It could be that one hospital in any multi-hospital town gets patients who are of poorer health, more urgent care needs, etc. To properly represent the town, a mix of files should be randomly selected from each hospital in the town, not just the largest one.

c) It could be that the 600 people who responded were not representative of the population. This method of collection impedes participation of those who are not as comfortable or familiar with web polls, and it completely excludes those who were unaware of the poll to begin with. A random selection of 600 registered voters will reduce the potential for these issues and will reach more subgroups of the voting population.

d) By allowing people to choose groups, several confounding factors might be introduced. For example, those who resist shots may have skin sensitivity, muscle pain, fear, etc. and it may be that those characteristics are unknowingly related to the variable of interest. Likewise, those who resist pills due to choking discomfort or gastrointestinal distress may have tendencies or conditions that are related to the variable of interest, and this could affect the results. It is better to impartially and randomly allocate people to the two groups so that the potential problems caused by self-allocation are minimized.



Types of Statistical Studies
Typeset May 4, 2016 at 21:37:55. Licensed by Illustrative Mathematics under a
Creative Commons Attribution-NonCommercial-ShareAlike 4.0 International License .