

1

Experimental Design

Learning Outcomes

By the end of this chapter, you should be able to:

1. Define key terms related to sampling and variables.
2. Describe the relationship between a population and a sample in making a statistical estimate.
3. Determine the independent and dependent variables within a given scenario.
4. Formulate a study with an appropriate sampling design that limits bias and error.

1.1 Experimental Design Background

As scientists, our knowledge of the natural world comes from direct observations and experiments. A good experimental design is essential for making inferences and drawing appropriate conclusions from our observations. Experimental design starts by formulating an appropriate question and then knowing how data can be collected and analyzed to help answer your question. Let us take the following example.

Case Study

Observation: A healthy body weight is correlated with good diet and regular physical activity. One component of a good diet is consuming enough fiber; therefore, one **question** we might ask is: do Americans who eat more fiber on a daily basis have a healthier body weight or body mass index (BMI) score?

How would we go about answering this question?

In order to get the most accurate data possible, we would need to design an experiment that would allow us to survey the entire **population** (all possible test subjects – all people living in the United States) regarding their eating habits and then match those to their BMI scores. However, it would take a lot of time and money to survey every person in the country. In addition, if too much time elapses from the beginning to the end of collection, then the accuracy of the data would be compromised.

More practically, we would choose a representative sample with which to make our inferences. For example, we might survey 5000 men and 5000 women to serve as a representative **sample**. We could then use that smaller sample as an **estimate** of our population to evaluate our question. In order to get a proper (and unbiased) sample and estimate of the population, the researcher must decide on the best (and most effective) sampling design for a given question.

1.2 Sampling Design

Below are some examples of sampling strategies that a researcher could use in setting up a research study. The strategy you choose will be dependent on your research question. Also keep in mind that the sample size (N) needed for a given study varies by discipline. Check with your mentor and look at the literature to verify appropriate sampling in your field.

Some of the sampling strategies introduce bias. **Bias** occurs when certain individuals are more likely to be selected than others in a sample. A biased sample can change the predictive accuracy of your sample; however, sometimes bias is acceptable and expected as long as

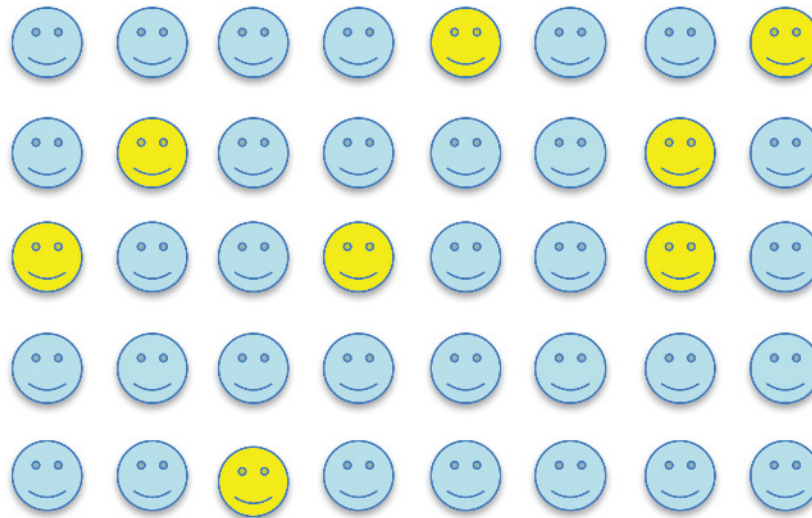


Figure 1.1 A representation of a random sample of individuals within a population.

it is identified and justifiable. Make sure that your question matches and acknowledges the inherent bias of your design.

Random Sample

In a random sample all individuals within a population have an equal chance of being selected, and the choice of one individual does not influence the choice of any other individual (as illustrated in Figure 1.1). A random sample is assumed to be the best technique for obtaining an accurate representation of a population. This technique is often associated with a random number generator, where each individual is assigned a number and then selected randomly until a preselected sample size is reached. A random sample is preferred in most situations, unless there are limitations to data collection or there is a preference by the researcher to look specifically at subpopulations within the larger population.

In our BMI example, a person in Chicago and a person in Seattle would have an equal chance of being selected for the study.

4 | *An Introduction to Statistical Analysis in Research*

Likewise, selecting someone in Seattle does not eliminate the possibility of selecting other participants from Seattle. As easy as this seems in theory, it can be challenging to put into practice.

Systematic Sample

A systematic sample is similar to a random sample. In this case, potential participants are ordered (e.g., alphabetically), a random first individual is selected, and every k th individual afterward is picked for inclusion in the sample. It is best practice to randomly choose the first participant and not to simply choose the first person on the list. A random number generator is an effective tool for this. To determine k , divide the number of individuals within a population by the desired sample size.

This technique is often used within institutions or companies where there are a larger number of potential participants and a subset is desired. In Figure 1.2, the third person (going down the first column) is the first individual selected and every sixth person afterward is selected for a total of 7 out of 40 possible.



Figure 1.2 A systematic sample of individuals within a population, starting at the third individual and then selecting every sixth subsequent individual in the group.



Figure 1.3 A stratified sample of individuals within a population. A minimum of 20% of the individuals within each subpopulation were selected.

Stratified Sample

A stratified sample is necessary if your population includes a number of different categories and you want to make sure your sample includes all categories (e.g., gender, ethnicity, other categorical variables). In Figure 1.3, the population is organized first by category (i.e., strata) and then random individuals are selected from each category.

In our BMI example, we might want to make sure all regions of the country are represented in the sample. For example, you might want to randomly choose at least one person from each city represented in your population (e.g., Seattle, Chicago, New York, etc.).

Volunteer Sample

A volunteer sample is used when participants volunteer for a particular study. Bias would be assumed for a volunteer sample because people who are likely to volunteer typically have certain characteristics in common. Like all other sample types, collecting demographic data would be important for a volunteer study, so that you can determine most of the potential biases in your data.

Sample of Convenience

A sample of convenience is not representative of a target population because it gives preference to individuals within close proximity. The reality is that samples are often chosen based on the availability of a sample to the researcher.

Here are some examples:

- A university researcher interested in studying BMI versus fiber intake might choose to sample from the students or faculty she has direct access to on her campus.
- A skeletal biologist might observe skeletons buried in a particular cemetery, although there are other cemeteries in the same ancient city.
- A malacologist with a limited time frame may only choose to collect snails from populations in close proximity to roads and highways.

In any of these cases, the researcher assumes that the sample is biased and may not be representative of the population as a whole.

Replication is important in all experiments. Replication involves repeating the same experiment in order to improve the chances of obtaining an accurate result. Living systems are highly variable. In any scientific investigation, there is a chance of having a sample that does not represent the norm. An experiment performed on a small sample may not be representative of the whole population. The experiment should be replicated many times, with the experimental results averaged and/or the median values calculated (see Chapter 2).

For all studies involving living human participants, you need to ensure that you have submitted your research proposal to your campus' Institutional Review Board (IRB) or Ethics Committee prior to initiating the research protocol. For studies involving animals, submit your research proposal to the Institutional Animal Care and Use Committee (IACUC).

Counterbalancing

When designing an experiment with paired data (e.g., testing multiple treatments on the same individuals), you may need to consider

counterbalancing to control for bias. Bias in these cases may take the form of the subjects learning and changing their behavior between trials, slight differences in the environment during different trials, or some other variable whose effects are difficult to control between trials. By counterbalancing we try to offset the slight differences that may be present in our data due to these circumstances. For example, if you were investigating the effects of caffeine consumption on strength, compared to a placebo, you would want to counterbalance the strength session with placebo and caffeine. By dividing the entire test population into two groups (A and B), and testing them on two separate days, under alternating conditions, you would counterbalance the laboratory sessions. One group (A) would present to the laboratory and undergo testing following caffeine consumption and then the other group (B) would present to the laboratory and consume the placebo on the same day. To ensure washout of the caffeine, each group would come back one week later on the same day at the same time and undergo the strength tests under the opposite conditions from day 1. Thus, group B would consume the caffeine and group A would consume the placebo on testing day 2. By counterbalancing the sessions you reduce the risk of one group having an advantage or a different experience over the other, which can ultimately impact your data.

1.3 Sample Analysis

Once we take a sample of the population, we can use **descriptive statistics** to characterize the population. Our estimate may include the **mean** and **variance** of the sample group. For example, we may want to compare the mean BMI score of men who intake greater than 38 g of dietary fiber per day with those who intake less than 38 g of dietary fiber per day (as indicated in Figure 1.4). We cannot sample all men; therefore, we might randomly sample 100 men from the larger population for each category (<38 g and >38 g). In this study, our sample group, or subset, of 200 men ($N = 200$) is assumed to be representative of the whole.

Although this estimate would not yield the exact same results as a larger study with more participants, we are likely to get a good estimate that approximates the population mean. We can then use

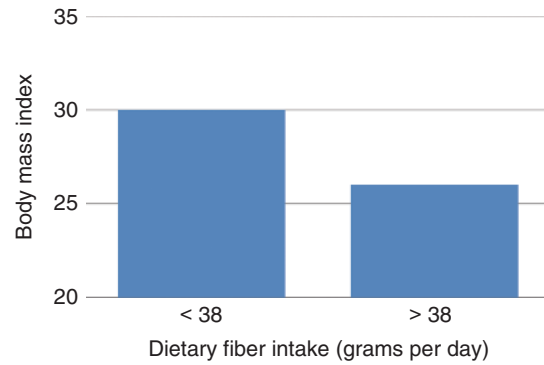
8 | *An Introduction to Statistical Analysis in Research*

Figure 1.4 Bar graph comparing the body mass index (BMI) of men who eat less than 38 g of fiber per day to men who eat more than 38 g of fiber per day.

inferential statistics to determine the quality of our estimate in describing the sample and determine our ability to make predictions about the larger population.

If we wanted to compare dietary fiber intake between men and women, we could go beyond descriptive statistics to evaluate whether the two groups (populations) are different, as in Figure 1.5. Inferential statistics allows us to place a **confidence interval** on whether the two samples are from the same population, or whether they are really two different populations. To compare men and women, we could use an independent *t*-test for statistical analysis. In this case, we would receive both the means for the groups, as well as a ***p*-value**,

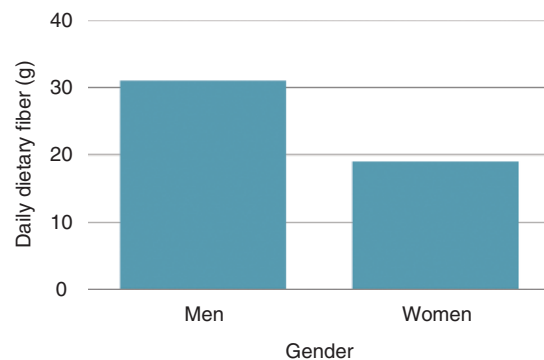


Figure 1.5 Bar graph comparing the daily dietary fiber (g) intake of men and women.

which would give us an estimated degree of confidence in whether the groups are different from each other.

1.4 Hypotheses

In essence, statistics is hypothesis testing. A **hypothesis** is a testable statement that provides a possible explanation to an observable event or phenomenon. A good, testable hypothesis implies that the independent variable (established by the researcher) and dependent variable (also called a response variable) can be measured. Often, hypotheses in science laboratories (general biology, cell biology, chemistry, etc.) are written as “If...then...” statements; however, in scientific publications, hypotheses are rarely spelled out in this way. Instead, you will see them formulated in terms of possible explanations to a problem. In this book, we will introduce formalized hypotheses used specifically for statistical analysis. Hypotheses are formulated as either the null hypothesis or alternative hypotheses. Within certain chapters of this book, we indicate the opportunity to formulate hypotheses using this symbol \textcircled{H} .

In the simplest scenario, the **null hypothesis (H_0)** assumes that there is no difference between groups. Therefore, the null hypothesis assumes that any observed difference between groups is based merely on variation in the population. In the dietary fiber example, our null hypothesis would be that there is no difference in fiber consumption between the sexes.

The **alternative hypotheses (H_1 , H_2 , etc.)** are possible explanations for the significant differences observed between study populations. In the example above, we could have several alternative hypotheses. An example for the first alternative hypothesis, H_1 , is that there will be a difference in the dietary fiber intake between men and women.

Good hypothesis statements will include a rationale or reason for the difference. This rationale will correspond with the background research you have gathered on the system.

It is important to keep in mind that difference between groups could be due to other variables that were not accounted for in our experimental design. For example, if when you were surveying men and women over the telephone, you did not ask about other dietary

choices (e.g., Atkins, South Beach, vegan diets), you may have introduced bias unexpectedly. If by chance, all the men were on a high protein diet and the women were vegan, this could bring bias into your sample. It is important to plan out your experiments and consider all variables that may influence the outcome.

1.5 Variables

An important component of experimental design is to define and identify the variables inherent in your sample. To explain these variables, let us look at another example.

Case Study

In 1995, wolves were reintroduced to Yellowstone National Park after an almost 70-year absence. Without the wolf, many predator–prey dynamics had changed, with one prominent consequence being an explosion of the elk population. As a result, much of the vegetation in the park was consumed, resulting in obvious changes, such as to nesting bird habitat, but also more obscure effects like stream health. With the reintroduction of the wolf, park rangers and scientists began noticing dramatic and far reaching changes to food webs and ecosystems within the park. One question we could ask is how trout populations were impacted by the reintroduction of the wolf. To design this experiment, we will need to define our variables.

The **independent variable**, also known as the treatment, is the part of the experiment established by or directly manipulated by the research that causes a potential change in another variable (the dependent variable). In the wolf example, the independent variable is the presence/absence of wolves in the park.

The **dependent variable**, also known as the response variable, changes because it “depends” on the influence of the independent variable. There is often only one independent variable (depending on the level of research); however, there can potentially be several dependent variables. In the question above, there is only one dependent variable – trout abundance. However, in a separate question, we could examine how wolf introduction impacted populations of beavers, coyotes, bears, or a variety of plant species.

Controlled variables are other variables or factors that cause direct changes to the dependent variable(s) unrelated to the changes caused by the independent variable. Controlled variables must be carefully monitored to avoid error or bias in an experiment. Examples of controlled variables in our example would be abiotic factors (such as sunlight) and biotic factors (such as bear abundance). In the Yellowstone wolf/trout example, researchers would need to survey the same streams at the same time of year over multiple seasons to minimize error.

Here is another example: In a general biology laboratory, the students in the class are asked to determine which fertilizer is best for promoting plant growth. Each student in the class is given three plants; the plants are of the same species and size. For the experiment, each plant is given a different fertilizer (A, B, and C). What are the other variables that might influence a plant's growth?

Let us say that the three plants are not receiving equal sunlight, the one on the right (C) is receiving the most sunlight and the one on the left (A) is receiving the least sunlight. In this experiment, the results would likely show that the plant on the right became more mature with larger and fuller flowers. This might lead the experimenter to determine that company C produces the best fertilizer for flowering plants. However, the results are biased because the variables were not controlled. We cannot determine if the larger flowers were the result of a better fertilizer or just more sunlight.

Types of Variables

Categorical variables are those that fall into two or more categories. Examples of categorical variables are nominal variables and ordinal variables.

Nominal variables are counted not measured, and they have no numerical value or rank. Instead, nominal variables classify information into two or more categories. Here are some examples:

- Sex (male, female)
- College major (Biology, Kinesiology, English, History, etc.)
- Mode of transportation (walk, cycle, drive alone, carpool)
- Blood type (A, B, AB, O)

Ordinal variables, like nominal variables, have two or more categories; however, the order of the categories is significant. Here are some examples:

- Satisfaction survey (1 = “poor,” 2 = “acceptable,” 3 = “good,” 4 = “excellent”)
- Levels of pain (mild, moderate, severe)
- Stage of cancer (I, II, III, IV)
- Level of education (high school, undergraduate, graduate)

Ordinal variables are ranked; however, no arithmetic-like operations are possible (i.e., rankings of poor (1) and acceptable (2) cannot be added together to get a good (3) rating).

Quantitative variables are variables that are counted or measured on a numerical scale. Examples of quantitative variables include height, body weight, time, and temperature. Quantitative variables fall into two categories: discrete and continuous.

Discrete variables are variables that are counted:

- Number of wing veins
- Number of people surveyed
- Number of colonies counted

Continuous variables are numerical variables that are measured on a continuous scale and can be either ratio or interval.

Ratio variables have a true zero point and comparisons of magnitude can be made. For instance, a snake that measures 4 feet in length can be said to be twice the length of a 2 foot snake. Examples of ratio variables include: height, body weight, and income.

Interval variables have an arbitrarily assigned zero point. Unlike ratio data, comparisons of magnitude among different values on an interval scale are not possible. An example of an interval variable is temperature (Celsius or Fahrenheit scale).