

Introduction

In this chapter we will look at a series of examples of areas in the life sciences in which statistics is used, with the goal of understanding the scope of the field of statistics. We will also

- explain how experiments differ from observational studies.
- discuss the concepts of blinding.
- discuss the role of random sampling in statistics.

Statistics and the Life Sciences

- Data from life sciences unavoidably exhibit **variability**.
- Statistics is the science of understanding data and of making decisions in the face of variability and uncertainty.

Example: bacteria and cancer

To study the effect of bacteria on tumor development, researchers used a strain of mice with a naturally high incidence of liver tumors. One group of mice were maintained entirely germ free, while another group were exposed to the intestinal bacteria *Escherichia coli*.

| Response | Treatment | |
|---------------------------|-----------|-----------|
| | E. coli | Germ free |
| Liver tumors | 8 | 19 |
| No liver tumors | 5 | 30 |
| Total | 13 | 49 |
| Percent with liver tumors | 62% | 39% |

Observational Study v.s. Experimental Study

- Observational Study:
 - Observing and collecting data without intervening or changing variables.
 - Researchers gather information by watching and recording natural events or existing data.
 - Used to find patterns or relationships but doesn't involve actively manipulating variables.
- Experimental Study:
 - Actively manipulating variables to see how they affect the outcome.
 - Researchers set up controlled experiments with experimental and control groups.

- Used to determine cause-and-effect relationships and assess the impact of specific interventions.

Blinding

Blinding:

- Withholding treatment information from participants or researchers.
- Prevents biases and expectations that could influence results.
- Participants may not know if they are receiving treatment or placebo.
- Researchers administering treatment may be unaware of group assignments.

Double-blinding:

- Goes a step further than blinding.
- Neither participants nor researchers know treatment assignments.
- Independent third party assigns treatments and provides coded labels.
- Treatment assignments revealed after study completion and data analysis.

Blinding and double-blinding help ensure rigorous and unbiased experimental research by concealing treatment assignments from participants and researchers. This reduces biases and improves the validity and reliability of study results.

Examples

For each of the following cases, state whether the study should be observational or experimental and whether the study should be run blind, double-blind, or neither. If the study should be run blind or double-blind, who should be blinded?

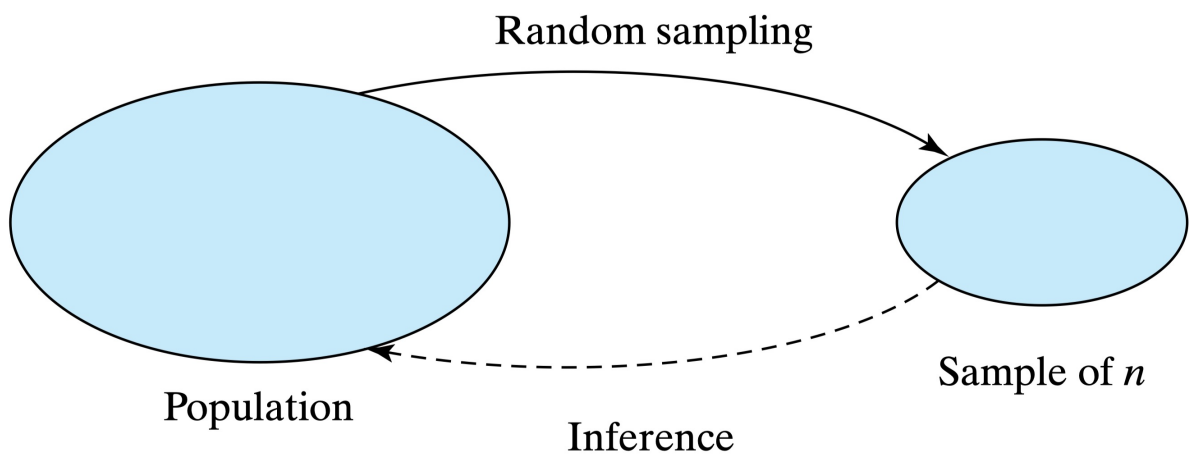
- An investigation of whether the size of the midsagittal plane of the anterior commissure (a part of the brain) of a man is related to the sexual orientation of the man.
- An investigation of whether taking aspirin reduces one's chance of having a heart attack.
- An investigation of whether drinking more than 1 liter of water per day helps with weight loss for people who are trying to lose weight.
- An investigation of whether babies born into poor families (family income below \$25,000) are more likely to weigh less than 5.5 pounds at birth than babies born into wealthy families (family income above \$65,000).
- This should be an observational study and should be single blind; the people who measure the brains should be blinded to the sexual orientation of the men.
- This should be an experiment and should be double-blind (i.e., neither the subjects nor the evaluating physicians should know who is in which group).
- This should be an experiment, with people randomly assigned to one of the two groups: high (more than 1 liter per day) vs low water intake. It should be single blind; the people

who measure the weights of the subjects should be blinded to the group assignments.

- This should be an observational study. There is no need for blinding here.

Random Sampling

- **Population:** all subjects/animals/specimens/plants, and so on, of interest.
 - All squirrels in the main campus of UC Davis.
 - All birch tree seedlings in Florida.
 - All black bears in Yosemite National Park.
- **Sample:** A subset of the population.
 - A *simple random sample* of n items is a sample in which (a) every member of the population has the *same chance* of being included in the sample, and (b) the members of the sample are chosen *independently* of each other.



Random cluster sampling:

- In random cluster sampling, the population is divided into clusters (e.g., geographical regions, schools, or households).
- A subset of clusters is randomly selected, often using simple random sampling.
- All individuals within the selected clusters are included in the sample.
- It is a cost-effective approach when clusters are easier to sample than individuals.
- Analysis requires accounting for the clustering effect in the data.

Examples:

- Conducting a health survey in a country by randomly selecting a few cities or districts as clusters, then sampling individuals from within those selected clusters.
- Studying the impact of a new educational program by randomly selecting a few schools as clusters, and then collecting data from all students within those selected schools.
- Assessing the quality of a product by randomly selecting retail stores as clusters and evaluating the product in all sampled stores.

Stratified random sampling:

- Stratified random sampling involves dividing the population into distinct subgroups or strata based on certain characteristics (e.g., age, gender, or location).
- Random samples are then independently selected from each stratum.
- Ensures representation from each subgroup, increasing precision and reducing sampling variability.
- Useful when strata differ significantly and you want to ensure proportional representation.
- Analysis may involve combining results from each stratum to make population-level inferences.

Examples

- Conducting a political opinion poll by dividing the population into strata based on age groups (e.g., 18-25, 26-40, 41-60, 61+), and then randomly selecting a sample from each age group.
- Conducting a market research survey on smartphone preferences by dividing the population into strata based on income levels (e.g., low, medium, high), and then randomly selecting participants from each income group.
- Studying the effects of a medication by stratifying the population into different severity levels of the condition (e.g., mild, moderate, severe), and randomly selecting participants from each severity group.