Data Types and Sampling

Summary

1. Statistics is a tool used to gather info about populations through samples.

2. Data can be qualitative or quantitative; of which can be discrete or continuous.

3. Experiments try to isolate the effect(s) of a treatment(s).

4. Various sampling methods exist to gather data.

Statistics

Statistics is the process of obtaining, organizing, summarizing, interpreting, and drawing conclusions based on observable values called *data*.
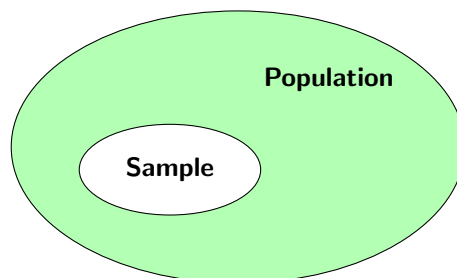
Population

The **population** is composed of all entities (*data values*) to be observed. The information we obtain from a population is referred to as a **parameter**.

Sample

The **sample** is a subgroup (a.k.a. *subset*) of the population. The information we obtain from a sample is referred to as a **statistic**.

A sample drawn from a population

- should be a good representation of that population

- should be big enough to include a variety of observations.

Example 1. Identify the population and the sample of each of the following.

(a) 100 people at the mall were surveyed as to whether or not they like the mall.

(b) Doctors analyzed the MRIs of 38 professional boxers for possible brain injury.

Statistical vs. Practical significance

As we progress, keep in mind that chance and randomness are always a factor in the study of statistics, and things are not as foolproof as they are in other math courses.

> **Statistically Significant**
>
> **Statistically significant** results are those that are not likely obtained by chance.

Example 2. Determine if each outcome would be considered statistically significant.

(a) Flipping a coin 100 times and getting tails 94 times.

(b) Flipping a coin 100 times and getting tails 53 times.

> **Practically Significant**
>
> **Practically significant** results are those that are useful in the real world.

Qualitative and Quantitative Data

> **Qualitative Data**
>
> **Qualitative data** (a.k.a. *categorical data*) is data that is based on some quality or characteristic.

For instance:

- Your name
- Blood type
- Zip code

> **Quantitative Data**
>
> **Quantitative data** is data that is based on some _measurable_ numeric value.

Not all numeric data is quantitative.

If two data values can be added together (or subtracted) to produce **meaningful** results, then the data is quantitative. Else, it is qualitative.

Example 3. Determine if each of the following represents qualitative or quantitative data.

(a) The amount of water a household uses in a month.

(b) Each student's favorite color in a statistics class.

(c) Social security numbers.

(d) How much money you have on you right now.

Discrete vs. Continuous Data

Within the realm of quantitative data, there are two types: discrete and continuous.

Quantitative Data

Discrete   **Continuous**

Countable   Measurable
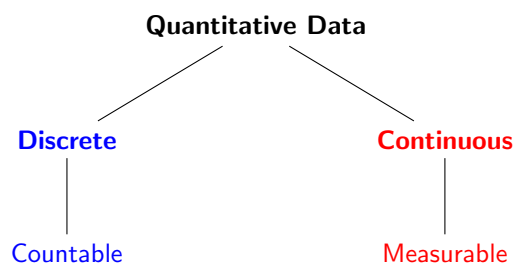
Example 4. Determine whether each quantitative variable is discrete or continuous.

(a) Number of free throws made.

(b) Time it takes to finish a book.

(c) Water pressure from a fire hose.

(d) The amount of money in a retirement account.

Observational study vs. Experiment

> ### Observational Study
>
> An **observational study** is a method to obtain data in which the collector (researcher) does not get involved.

Observational studies try to take as much of a *hands-off approach* as possible.

Researcher does not interject themselves into the study.

> ### Experiment
>
> In an **experiment**, the researcher
>
> - divides subjects into groups
>
> - applies a treatment to 1 group
>
> - notes the effects (if any) between the groups

Typically divided into 2 groups:

- **Experimental**: group that receives the treatment.

- **Control**: group that either does not receive treatment or receives a "fake" treatment (such as a *placebo*).

Example 5. Classify each as either an observational study or an experiment.

(a) A survey of 1,000 people is conducted to determine the best dog breed.

(b) 83 patients are given a new anxiety medication and 75 patients are given a sugar pill.

Sampling Methods

There are various ways in which to take samples, and depending on the research, one might be more appropriate than another.

However, keep in mind that **good sampling incorporates randomness into the process.**

- **Random sampling**

  - Each member of the population has an equal chance of being selected.

- **Stratified sampling**

  - Divide the population into non-overlapping groups (*strata*).
  - Randomly sample from each group.
  - (Some from all)

- **Cluster sampling**
  - Divide the population into non-overlapping groups.
  - Randomly pick groups and obtain information from everyone in those groups.
  - (All from some)

- **Systematic sampling**
  - Subjects are placed in some order.
  - Pick a random starting value (n).
  - Pick a random value to count by (k).
  - Starting at n, take every k^{th} subject thereafter.

- **Convenience sampling**
  - Subjects select themselves.
  - Results are easily obtained.
  - Least effective and desirable method.
  - Also known as a *voluntary response sample*.

Example 6. Classify each by the sampling method used.

(a) An achievement test is given to all 9th and 12th grade students at a local high school.

(b) A radio station asks listeners to call in with their opinion on a political issue.

(c) A quality control manager selects the 5th circuit board on an assembly line and then selects every 14th circuit board after that.

(d) 10 seniors, 12 juniors, 13 sophomores, and 10 freshmen are asked to name their favorite food.

(e) 10 names are drawn out of a hat containing 50 names.