TOPIC 1.1

An Overview of Statistics

OBJECTIVES Define statistics. Tell a population from a sample and a parameter from a statistic. Tell descriptive statistics from inferential statistics.

GETTING STARTED A school newspaper wants to know how many hours per week the 1,200 students at its high school spend on homework. A reporter asks 80 students in the cafeteria at lunch.

a.) Which group does the newspaper want to learn about? _____

b.) Which group did the reporter actually ask? _____

c.) Will the average for the 80 students be exactly the same as the average for all 1,200 students? Explain.

What Is Statistics?

DEFINITION Data and Statistics

Data are information that comes from observations, counts, measurements, or responses.

Statistics is the science of _____, _____, _____, and _____ data in order to make decisions.

Populations and Samples

DEFINITION Population and Sample

A **population** is the collection of _____ outcomes, responses, measurements, or counts that are of interest.

A **sample** is a _____, or part, of the population. (Strictly, both are sets of responses or measurements, such as “the heights of all students.”)

A **census** collects data from _____ member of the population. A **sampling** collects data from part of the population.

Example 1 An inspector measures the amount of juice in 50 bottles chosen from one day's production of 20,000 bottles.

a.) Population _____

b.) Sample _____

Example 2 In a survey of 1,500 U.S. adults, 41% said they get most of their news from social media.

a.) Population _____

b.) Sample _____

Parameters and Statistics

DEFINITION Parameter and Statistic

A **parameter** is a number that describes a _____ characteristic.

A **statistic** is a number that describes a _____ characteristic.

Parameters use _____ letters, such as μ (mean) and σ (standard deviation). Statistics use _____ letters, such as $\bar{x}$ and s.

TIP Match the Letters

Population goes with **p**arameter. **S**ample goes with **s**tatistic. A parameter is a fixed number (often unknown). A statistic changes from sample to sample. **Clue words:** all, every, entire, census → population. Some, part, a poll of, randomly selected → sample.

Compare	Population	Sample
Is	all members of interest	part of the population
Described by a	_____	_____
Symbols	Greek: μ, σ	Roman: $\bar{x}$, s

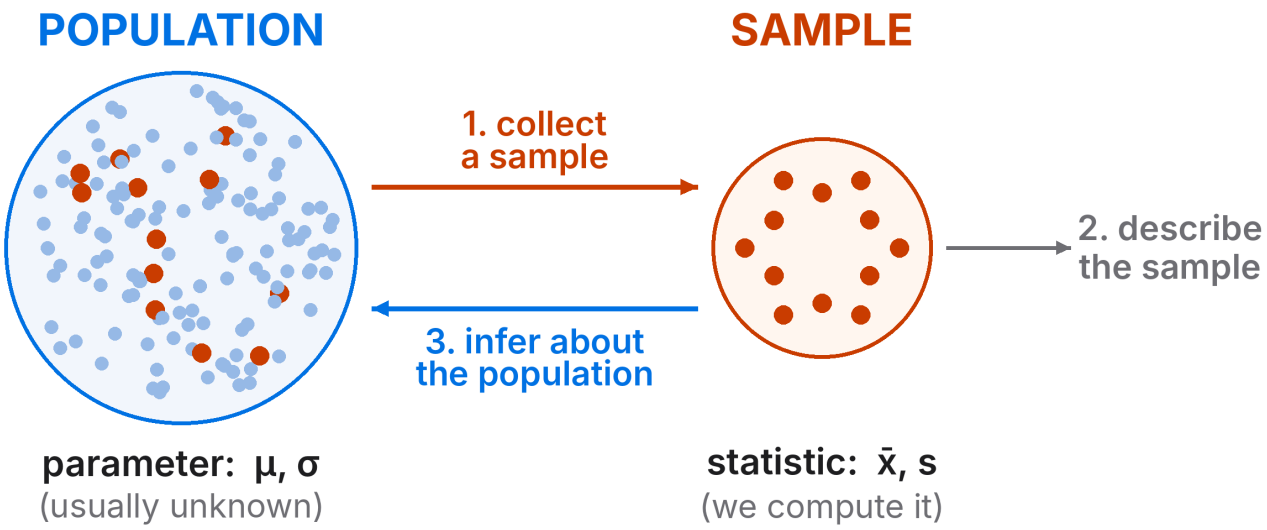
Example 3 Decide whether each number is a parameter or a statistic. Explain.

a.) In a survey of 400 voters in a city, 54% support building a new park. _____

b.) All 212 graduates of a nursing program had a mean starting salary of $68,400. _____

c.) A check of 60 randomly selected homes in a town found a mean water use of 290 gallons per day. _____

d.) Of the 32 teams in the NFL, 2 are based in Ohio (the Browns and the Bengals). _____



Branches of Statistics

DEFINITION **Descriptive and Inferential Statistics**

Descriptive statistics is the branch of statistics that involves organizing, summarizing, and _____ data.

Inferential statistics is the branch of statistics that uses a _____ to draw conclusions about a _____. Probability is a basic tool of inferential statistics.

Example 4 A study of 2,000 adults ages 20 to 40 found that adults who slept less than 6 hours a night drank an average of 2.4 cups of coffee per day. Adults who slept at least 8 hours drank an average of 1.1 cups per day.

a.) Identify the population. _____

b.) Identify the sample. _____

c.) Which part of the study is descriptive statistics?

d.) What inference might be drawn from the study?

WATCH OUT Inferences Can Be Wrong

A sample result is an **estimate**. A different sample would probably give a slightly different result. Also, a link between two variables does not prove that one **causes** the other.

Example 5 Decide whether each statement uses descriptive or inferential statistics. Explain.

a.) In last year's graduating class, 72% of students took at least one AP or CCP course. _____

b.) Based on a survey of 600 drivers, about 1 in 4 drivers in the state text while driving. _____

CHECK YOUR UNDERSTANDING

Which statement describes a **parameter**?

(A) The mean height of 25 students chosen from a school of 900 students is 66 inches.
(B) In a poll of 1,000 adults, 38% own an electric car.
(C) The mean age of all 14 employees of a bakery is 31 years.
(D) In a sample of 50 phones, 3 had a defective battery.

TOPIC 1.2

Data Classification

OBJECTIVES Tell qualitative data from quantitative data. Classify data by level of measurement: nominal, ordinal, interval, or ratio.

GETTING STARTED A customer survey card asks for three things: (1) your ZIP code, (2) your age in years, and (3) a rating of the service: poor, fair, good, or excellent.

a.) For which item would it make sense to find an average? _____

b.) Which item has answers that can be put in a meaningful order but are not numbers? _____

Types of Data

DEFINITION Qualitative and Quantitative Data

Qualitative data consist of attributes, labels, or _____ entries.

Quantitative data consist of numerical _____ or _____.

WATCH OUT Numbers Can Be Labels

A number is quantitative only if it measures or counts something. ZIP codes, jersey numbers, and phone numbers are **qualitative**: they are labels.

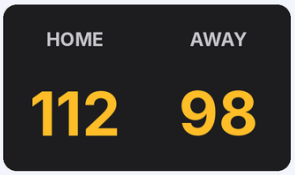
QUALITATIVE: labels



The number 23 is a name tag, not an amount.

Averaging jersey numbers means nothing.

QUANTITATIVE: counts or measures



Points are counts.

112 − 98 = 14 makes sense: a 14-point win.

Example 1 Decide whether each variable is qualitative or quantitative.

a.) Eye color _____ b.) Number of siblings _____

c.) Jersey number of an NBA player (LeBron James has worn 23 and 6) _____

d.) Time to run a mile, in minutes _____

e.) Type of phone (brand) _____

Levels of Measurement

DEFINITION The Four Levels of Measurement

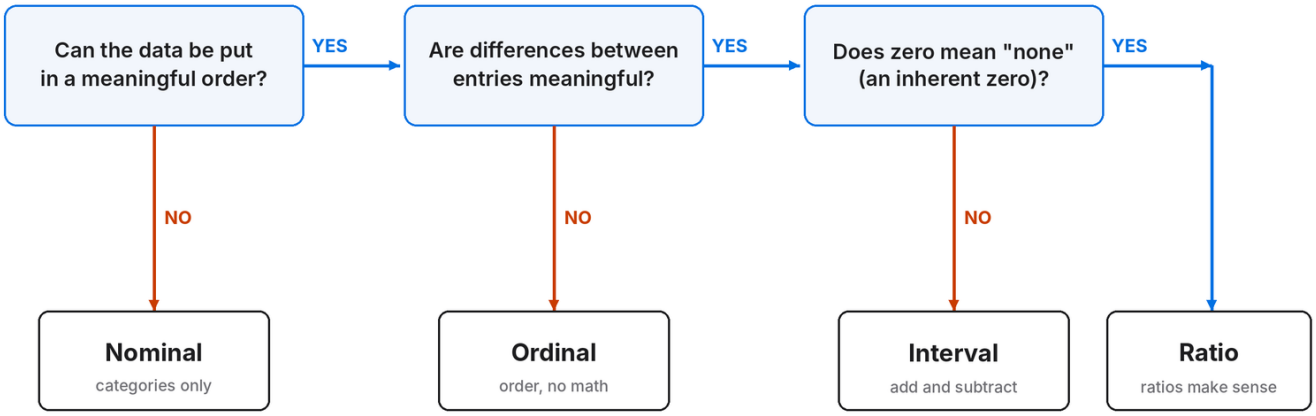
Nominal: qualitative data that can only be sorted into categories by name, label, or quality. The data _____ be put in a meaningful order.

Ordinal: data that can be put in a meaningful _____, but differences between entries are not meaningful.

Interval: quantitative data that can be ordered, and _____ between entries are meaningful. Zero is only a position on a scale; it does not mean "none."

Ratio: like interval, but zero is an **inherent zero** (zero means "none"). Ratios are meaningful, so "twice as much" makes sense.

Deciding the level: start at the left and stop at the first NO. Data are classified at the highest level they fit.



Example 2 Complete the table. Write Yes or No.

Level	Put in categories	Put in order	Subtract (differences)	Ratios ("twice as")
Nominal	_____	_____	_____	_____
Ordinal	_____	_____	_____	_____
Interval	_____	_____	_____	_____
Ratio	_____	_____	_____	_____

TIP Inherent Zero Test

Ask: “Does zero mean **none** of the quantity?” 0 pounds means no weight, so weight is ratio. 0°F does not mean no temperature, so °F is interval.

Example 3 Identify the level of measurement of each data set. Explain.

a.) The colors of the cars in a parking lot _____

b.) The final AP Top 25 college football rankings (1st, 2nd, 3rd, …) _____

c.) Years the Cincinnati Reds won the World Series: 1919, 1940, 1975, 1976, 1990 _____

d.) The heights of NBA players, in inches _____

e.) Daily high temperatures, in degrees Fahrenheit _____

f.) The number of home runs each MLB player hit last season _____

Example 4 On Monday the high temperature was 80°F. On Tuesday it was 40°F. Is it correct to say Monday was twice as hot as Tuesday? Explain.

Example 5 Movie ratings G, PG, PG-13, and R are used to describe how suitable a film is for young viewers.

a.) Qualitative or quantitative? _____

b.) Level of measurement? _____

CHECK YOUR UNDERSTANDING

Which data set is at the **ratio** level of measurement?

(A) The jersey numbers of a basketball team
(B) Customer ratings of 1 to 5 stars
(C) The distances, in miles, that employees drive to work
(D) Body temperatures, in degrees Celsius

TOPIC 1.3

Data Collection and Experimental Design

OBJECTIVES Design a statistical study. Tell an observational study from an experiment. Design an experiment. Choose and identify sampling techniques, and spot a biased sample.

GETTING STARTED A company wants to know whether its new energy drink improves reaction time.

Plan A: Find students who already drink it and students who do not. Compare their reaction times.

Plan B: Take 40 volunteers. Randomly assign 20 to drink the energy drink and 20 to drink a look-alike drink with no caffeine. Compare their reaction times.

Which plan gives better evidence that the drink **causes** faster reactions? Explain. _____

1. **PROCEDURE** **Designing a Statistical Study**

1. Identify the **variable(s)** of interest and the **population** of the study.
2. Plan how to collect the data. If you use a sample, make sure it **represents** the population.
3. Collect the data.
4. Describe the data with **descriptive** statistics.
5. Interpret the data and make decisions about the population with **inferential** statistics.
6. Identify any possible **errors**.

Ways to Collect Data

DEFINITION Four Ways to Collect Data

Observational study: researchers observe and measure characteristics of interest but do _____ change existing conditions.

Experiment: a _____ is applied to part of a population, called the **treatment group**, and responses are observed. Another part, the **control group**, may get no treatment or a fake treatment.

Simulation: uses a mathematical or physical _____ to reproduce the conditions of a situation or process. Often done with computers.

Survey: an investigation of one or more characteristics of a population, usually by asking people _____.

Example 1 Decide which method of data collection fits best: observational study, experiment, simulation, or survey.

a.) Testing whether a new fertilizer increases the number of tomatoes per plant _____

b.) Since 1948, the Framingham Heart Study has followed thousands of residents of Framingham, Massachusetts, recording their health and habits _____

c.) Using a computer model to predict damage to a car in a crash _____

d.) Finding the percent of students who want a later school start time _____

Experimental Design

DEFINITION Key Elements of an Experiment

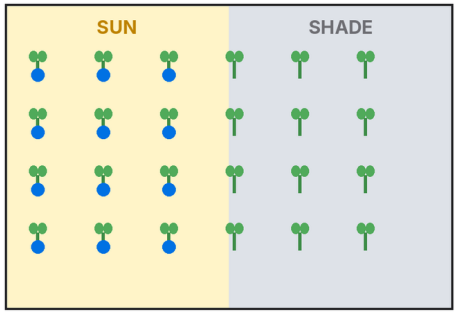
Control: limit the influence of other factors. A **confounding variable** occurs when you cannot tell the difference between the effects of different factors on a variable.

Placebo: a harmless, fake treatment that looks like the real one. The **placebo effect** occurs when a subject reacts favorably to a placebo.

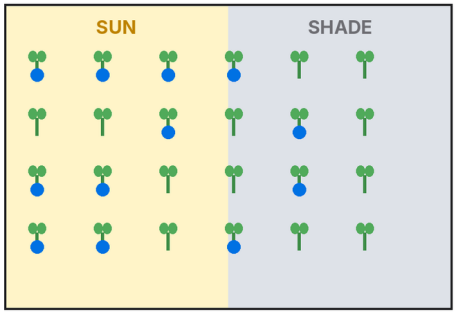
Blinding: subjects do not know whether they get the treatment or the placebo. **Double-blind:** _____ the subjects nor the experimenter knows.

Randomization: subjects are assigned to groups _____.

Replication: the experiment is repeated on a large group of subjects. The **sample size** is the number of subjects.



Confounded: fertilizer only in the sun



Better: fertilizer assigned at random

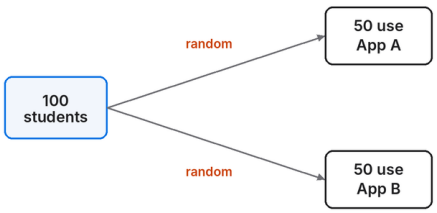
● = plant given the new fertilizer

DEFINITION Two Common Designs

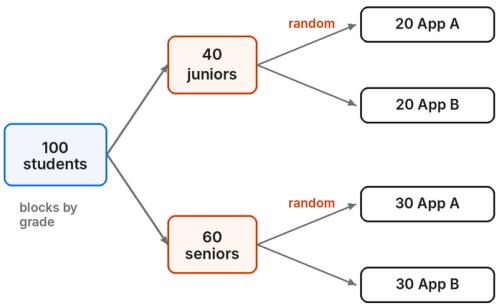
Randomized block design: first divide subjects with similar characteristics into _____. Then randomly assign treatments within each block.

Matched-pairs design: subjects are paired up by similar characteristics. One subject in each _____ gets one treatment and the other gets a different treatment.

Completely Randomized Design



Randomized Block Design



Example 2 A real experiment. In 1954, about 400,000 children took part in a test of the new Salk polio vaccine. About half were randomly assigned to get the vaccine. The other half got a placebo shot that looked the same. Neither the children, their parents, nor the doctors knew who got which. (Source: Francis Report, 1955)

a.) Describe the treatment group. _____

b.) Describe the control group. _____

c.) Why give the control group a placebo shot instead of nothing?

d.) Was the experiment blind or double-blind? Explain. _____

e.) Polio struck young children more often than older children. Describe a randomized block design.

Example 3 A gym tests a new protein shake. Members who lift weights 5 days a week drink the shake. Members who lift 2 days a week do not. After 8 weeks, the shake group gained more strength.

a.) Identify a problem with this experiment.

b.) Suggest a way to fix it.

Example 4 A company tests a new energy bar. Four runners eat the bar before a race and four do not. The bar group ran faster.

a.) Identify a problem with this experiment.

b.) Suggest a way to fix it.

Sampling Techniques

DEFINITION Sampling Techniques

Random sample: every member of the population has an _____ chance of being selected.

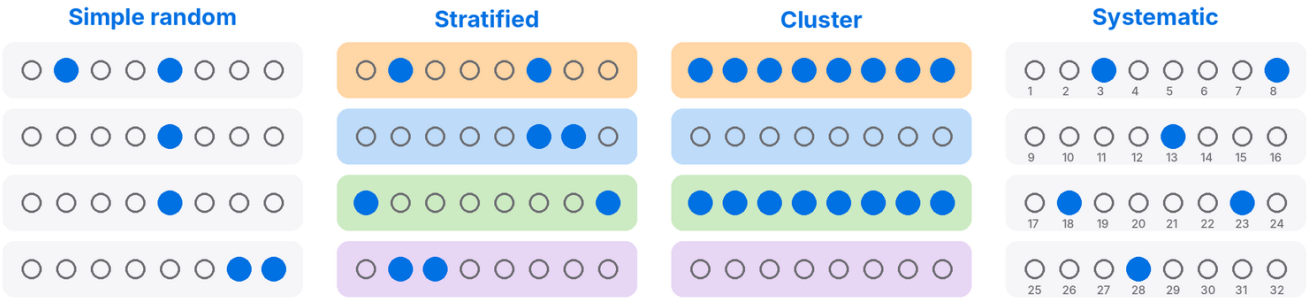
Simple random sample: every possible sample of the same size has the same chance of being selected.

Stratified sample: divide the population into groups (**strata**) that share a characteristic, then randomly select members from _____ stratum.

Cluster sample: divide the population into groups (**clusters**), randomly select some clusters, and use _____ members of the chosen clusters.

Systematic sample: number the members, pick a random starting number, then select every *k*th member.

Convenience sample: uses only members who are easy to reach. It is usually _____.



TIP Why Stratify?

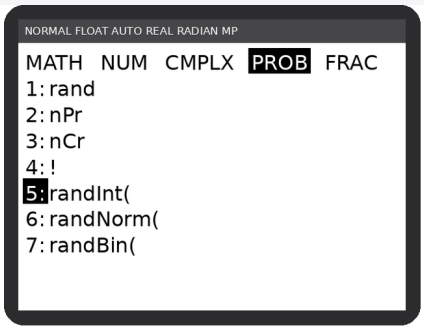
A simple random sample of 20 students could, just by chance, be mostly freshmen. Randomly choosing 5 students from **each** grade guarantees that every grade is represented.

WATCH OUT Biased Samples and Sampling Error

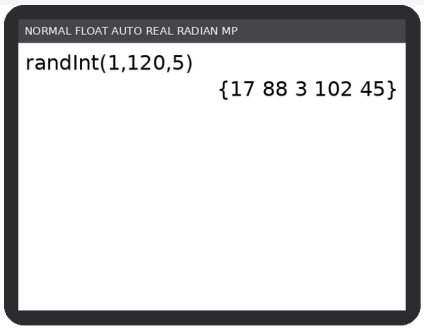
A sample is **biased** if some part of the population is favored or left out. Even a good random sample will not match the population exactly. The difference between a sample result and the population result is **sampling error**.

TIP TI-84: Random Integers

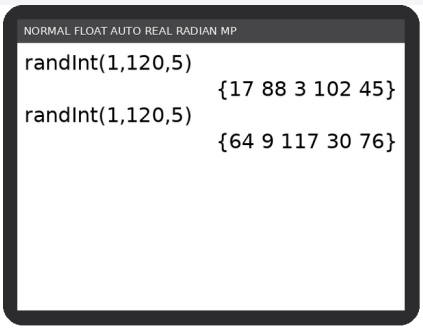
To choose 5 random numbers from 1 to 120: press **MATH**, arrow right to **PROB**, choose **5: randInt(**, and enter **randInt(1, 120, 5)**. If a number repeats, skip it and press **ENTER** again for more.



1 MATH, arrow right to PROB, choose 5:randInt(



2 Type 1, 120, 5 and press ENTER



3 ENTER again gives a new set

Example 5 Identify the sampling technique. Then tell whether the sample could be biased.

a.) A school picks a random name among the first 15 names on its alphabetical roster, then selects every 15th name after it. _____

b.) College athletes are divided by NCAA division (I, II, III), and 100 athletes are randomly selected from each division. _____

c.) A researcher randomly selects 5 of the 30 MLB teams and surveys every player on those 5 teams. _____

d.) A reporter interviews the first 30 people who walk out of a coffee shop. _____

e.) A company puts every employee ID number into a random number generator and chooses 50 numbers. _____

Example 6 **A famous biased sample.** In 1936, *Literary Digest* magazine mailed poll cards to about 10 million people, chosen mostly from telephone books and car registration lists. About 2.4 million replied, and the poll predicted that Alf Landon would easily beat President Franklin Roosevelt. Roosevelt won 46 of the 48 states.

a.) Why was the sample biased?

b.) The sample had 2.4 million people. Why didn't the large size fix the problem?

Example 7 A teacher numbers 28 students from 1 to 28 and wants to choose 4 of them at random. What should the teacher enter on the TI-84? _____

Example 8 Decide whether you would take a census or use a sample. Explain.

a.) The mean height of the 15 players on an NBA team's roster _____

b.) The mean number of hours a brand of light bulb lasts _____

CHECK YOUR UNDERSTANDING

A researcher divides a town into 200 city blocks, randomly selects 10 blocks, and surveys every household on those blocks. Which sampling technique is this?

(A) stratified
(B) cluster
(C) systematic
(D) convenience