Five-minute review

1. What information could you collect to compare two students' travel to school?

2. Five students report 12, 15, 10, 18 and 20 minutes. Find the total and average.

Write your answers before we discuss them.

Population, sample and statistical questions

What we will learn

Distinguish populations, samples, parameters and statistics; formulate a statistical question.

Why does it matter?

A library wants to decide how many books to order. It surveys a random group rather than asking every student.

Definitions: Population sample and statistical questions

Statistics: the study of how data are collected, organized, analyzed and interpreted.

Population: the complete group we want to understand; sample: a subset of that group.

Parameter: a numerical summary of a population; statistic: a numerical summary of a sample.

Illustration and interpretation



Notice the pattern

A library wants to decide how many books to order. It surveys a random group rather than asking every student.

Illustrative data / model

Important ideas and notation

Statistics turns data into evidence: ask a question, collect data, summarize, analyze and interpret. A statistical question anticipates variability.

Population = all individuals of interest (size N); sample = a subset (size n). A census measures the entire population.

A parameter summarizes a population; a statistic summarizes a sample. Descriptive statistics summarize observed data; inference draws conclusions beyond the sample.

Example 1.1.1

A school has 480 students. A random sample of 40 reports a mean commute of 18 minutes. Identify the population, sample and statistic.

Solution:

Population: all 480 students

sample: the 40 surveyed

statistic: 18 minutes. The population mean is unknown.

Example 1.1.2

All 25 students in a class report their number of books read. The mean is 3.2. Is 3.2 a parameter?

Solution:

Yes, if the population of interest is that class. The same mean is a statistic if the target is the entire school.

Example 1.1.3 • Reading habits

Which number is a statistic? Why is it useful?

Group	Number	Mean books
All students	240	Unknown
Surveyed students	30	4.2

Illustrative classroom dataset

Example 1.1.3 • Solution

Which number is a statistic? Why is it useful?

Solution:

4.2 books

it estimates the population mean without surveying all 240.

Individual practice

1. A survey of 60 of 900 residents finds 36 prefer a park. Identify N, n and the sample proportion.

2. Which is a statistical question: How tall is one specified student? or How tall are students in Grade 10?

3. A mean of 72 is calculated for all 18 team members. Parameter or statistic for that team?

Group activity

Design a question about student sleep; identify the population, a feasible sample and a numerical statistic.

Show your calculation or graph. Write a sentence explaining the result in context.

Check your answers

1. $N = 900$; $n = 60$; $\frac{36}{60} = 0.60$.

2. The second, because heights vary.

3. Parameter.

Group solution

Solution:

Example: hours of sleep on school nights

all Grade 10 students

random sample of 30

mean hours.

Compare your conclusion with the question. Check your units and explain what the numbers represent.

Exit ticket • 5 minutes

1. Why can a sample statistic differ from a population parameter?

2. A school has 600 students. A random sample of 50 has an average of 7 hours of sleep. Identify the parameter of interest and its estimate.

Show your reasoning. Submit before leaving.

Five-minute review

1. Why can a sample statistic differ from a population parameter?

2. A survey of 60 of 900 residents finds 36 prefer a park. Identify N, n and the sample proportion.

Write your answers before we discuss them.

Types of data and measurement scales

What we will learn

Classify categorical and quantitative variables and identify measurement levels.

Why does it matter?

A number in a spreadsheet is not always a measurement. Look at what the column actually records.

Definitions: Types of data and measurement scales

Categorical data: labels or groups, even when represented by numbers.

Quantitative data: meaningful numerical counts or measurements.

Discrete data: countable possible values; continuous data: measurements that can vary along an interval.

Important ideas and notation

Categorical variables identify groups. Quantitative variables represent meaningful counts or measurements; discrete values are countable, continuous values lie on a continuum.

Nominal: unordered categories. Ordinal: ordered categories without meaningful equal differences. Interval: meaningful differences but arbitrary zero (°C). Ratio: meaningful differences and true zero (length).

Using the method correctly

Numeric labels are not necessarily quantitative. A measured variable can be recorded after rounding without becoming conceptually discrete.

Example 1.2.1

Classify student ID, quiz points out of 20, and time to finish.

Solution:

ID: nominal categorical
quiz points: discrete quantitative
completion time: continuous quantitative.

Example 1.2.2

Compare 10°C and 20°C. Is 20°C twice as hot?

Solution:

No

Celsius is interval. A 10°C difference is meaningful but zero is arbitrary.

Example 1.2.3 • A school travel survey

Classify travel mode, minutes and satisfaction.

Student ID	Travel mode	Minutes	Satisfaction
101	Bus	18	High
102	Walk	12	Medium
103	Car	25	Low

Illustrative classroom dataset

Example 1.2.3 • Solution

Classify travel mode, minutes and satisfaction.

Solution:

Nominal categorical
continuous ratio measurement (recorded to minutes)
ordinal categorical.

Individual practice

1. Classify favorite genre, rating poor/fair/good, and distance walked.

2. Classify number of absences.

3. Is a ZIP code nominal or ratio?

Group activity

Make a survey with one variable at each measurement level.

Show your calculation or graph. Write a sentence explaining the result in context.

Check your answers

1. Nominal; ordinal; ratio (continuous).

2. Discrete quantitative, ratio.

3. Nominal; arithmetic on ZIP codes is not meaningful.

Group solution

Solution:

Example: transport type, satisfaction rating, temperature in °F, minutes of exercise.

Compare your conclusion with the question. Check your units and explain what the numbers represent.

Exit ticket • 5 minutes

1. Explain why a rank of 2 is not twice a rank of 1.

2. Classify jersey number, number of siblings and mass in kilograms.

Show your reasoning. Submit before leaving.

Quick review — 5 minutes

1. What is the difference between a population and a sample?

2. Classify the following as qualitative or quantitative: favorite subject, number of siblings, height.

Lesson 1.3: Sampling methods and bias

What we will learn

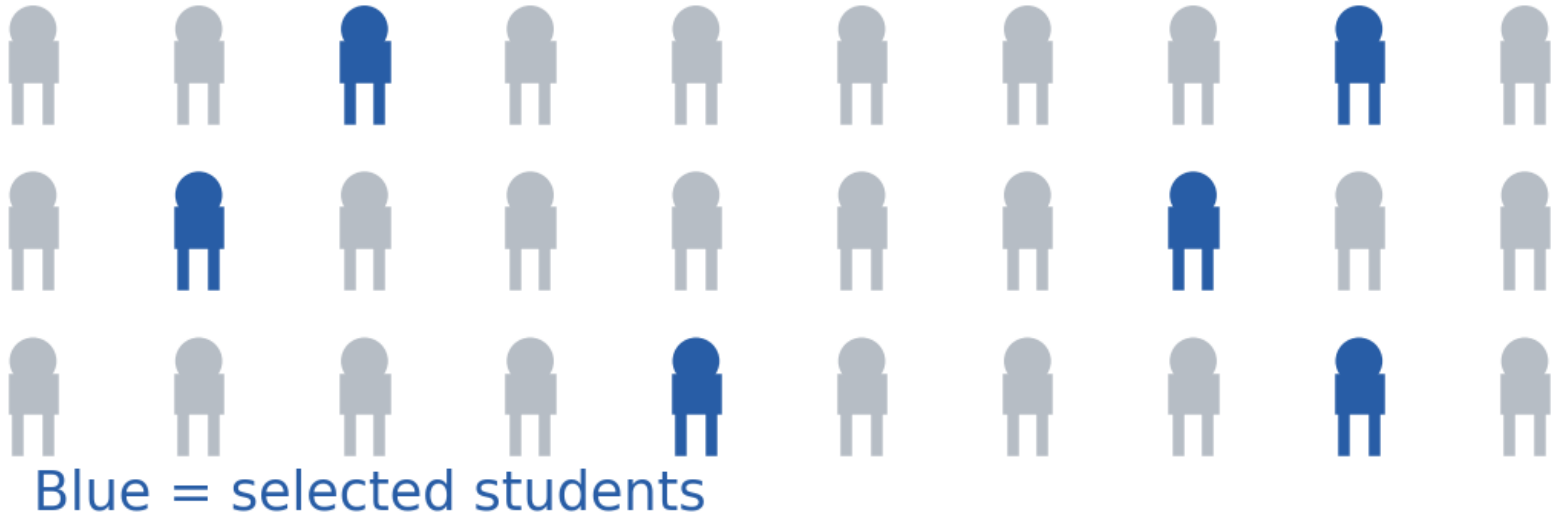
- Identify five common methods of sampling.
- Distinguish stratified sampling from cluster sampling.
- Explain how sampling bias can affect a conclusion.

Why does it matter?

A school wants to improve its lunch menu. Asking every student takes time, so a well-chosen sample can help.

Why do we need a sample?

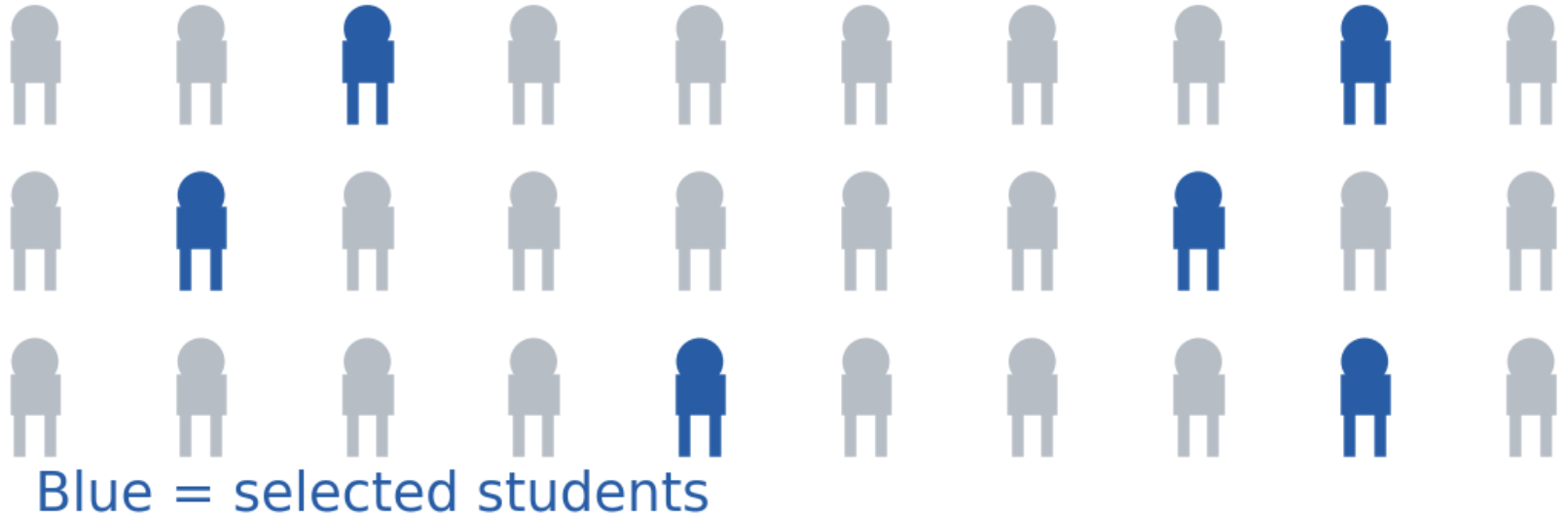
A population is the complete group of individuals under study. A sample is a subset of that population.



Here, N = 30 students and n = 6 students are selected.

Simple random sampling

Every possible sample of a fixed size has an equal chance of being selected.



Example 1.3.1: Randomly draw six names from a list of 30 students.

Systematic sampling

Choose a random starting position and then select every kth member of the population.



Example 1.3.2: Start with student 3 and select every fifth name.

Convenience sampling

Select members of the population who are easiest to reach.



Example 1.3.3: Ask the six students sitting closest to the teacher.

Stratified sampling

Divide the population into groups (strata), then randomly sample from every group.



Example 1.3.4: Randomly choose two students from each of three grades.

Cluster sampling

Divide the population into groups (clusters), randomly choose entire groups, and survey all members of the selected groups.



Example 1.3.5: Randomly select one homeroom and survey everyone in it.

Sampling bias

Sampling bias occurs when the selection procedure systematically favors some members of a population over others.

Example: A survey of only students in the school cafeteria may miss students who bring lunch from home.

A large sample does not fix a biased selection method.

Individual practice

1. Choose every 10th name after a random start. Identify the method.

2. Survey only your friends about lunch. What is the method and possible bias?

3. Randomly select two entire classrooms. Identify the method.

Group activity

A school has three grades with 60, 90 and 90 students.

Design a proportional stratified sample of 24 students. Explain how you will select students within each grade.

Solutions

1. Systematic sampling.

2. Convenience sampling; friends may not represent all students.

3. Cluster sampling.

Group activity: select 6, 9 and 9 students randomly from the three grades.

Exit ticket — 5 minutes

1. Explain the difference between stratified and cluster sampling.

2. Why might a convenience sample give a misleading result?

Five-minute review

1. What distinguishes stratified from cluster sampling?

2. Choose every 10th name after a random start.

Write your answers before we discuss them.

Observational studies, experiments and ethics

What we will learn

Design a controlled study and recognize confounding and biased questions.

Why does it matter?

Compare two practice methods by randomly assigning volunteers, then giving everyone the same quiz.

Definitions: Observational studies experiments and ethics

Observational study: researchers record characteristics without assigning a treatment.

Experiment: researchers deliberately assign a treatment and measure the response.

Confounder: a variable that makes it difficult to separate the effects of the explanatory variables.

Important ideas and notation

Observational studies measure without assigning treatment. Experiments assign treatments; random assignment helps balance confounders.

Use a comparison group, consistent measurement, random assignment and replication. Random sampling supports generalization; random assignment supports causal inference.

Avoid leading questions; report sources, sample design, limitations and protect participants' privacy. Association alone is not causation.

Example 1.4.1

Students who choose tutoring have higher scores. What can we conclude?

Solution:

Association only

motivation and prior achievement may confound results.

Example 1.4.2

Randomly assign volunteers to two review methods, then give the same test.

Solution:

Randomized experiment

review method is treatment, test score response. Causal evidence applies to study participants under appropriate conditions.

Example 1.4.3 • Testing a study method

What should stay the same besides the practice method?

Group	Students	Practice
A	20	Flashcards
B	20	Practice questions

Illustrative classroom dataset

Example 1.4.3 • Solution

What should stay the same besides the practice method?

Solution:

Time, quiz, instructions and testing conditions

random assignment helps reduce confounding.

Individual practice

1. Rewrite: Don't you agree our excellent library needs more funding?

2. Is recording daily rainfall an experiment?

3. Why use random assignment?

Group activity

Design a study comparing two practice methods with 40 volunteers.

Show your calculation or graph. Write a sentence explaining the result in context.

Check your answers

1. How do you feel about increasing library funding? Offer balanced choices.

2. No, observational.

3. To reduce systematic differences between treatment groups.

Group solution

Solution:

Randomly assign 20 per group, equal time and content, same assessment

compare scores and acknowledge volunteer sample.

Compare your conclusion with the question. Check your units and explain what the numbers represent.

Exit ticket • 5 minutes

1. Can a random sample alone establish causation?

2. A survey finds that students who walk to school sleep more. Can we conclude walking causes more sleep?

Show your reasoning. Submit before leaving.

Five-minute review

1. Can a random sample alone establish causation?

2. Rewrite: Don't you agree our excellent library needs more funding?

Write your answers before we discuss them.

Evaluating statistical claims

What we will learn

Evaluate sources, bias, wording and causal claims before accepting a conclusion.

Why does it matter?

A poll shared on a homework app reaches students who already use that app. The headline may overstate the evidence.

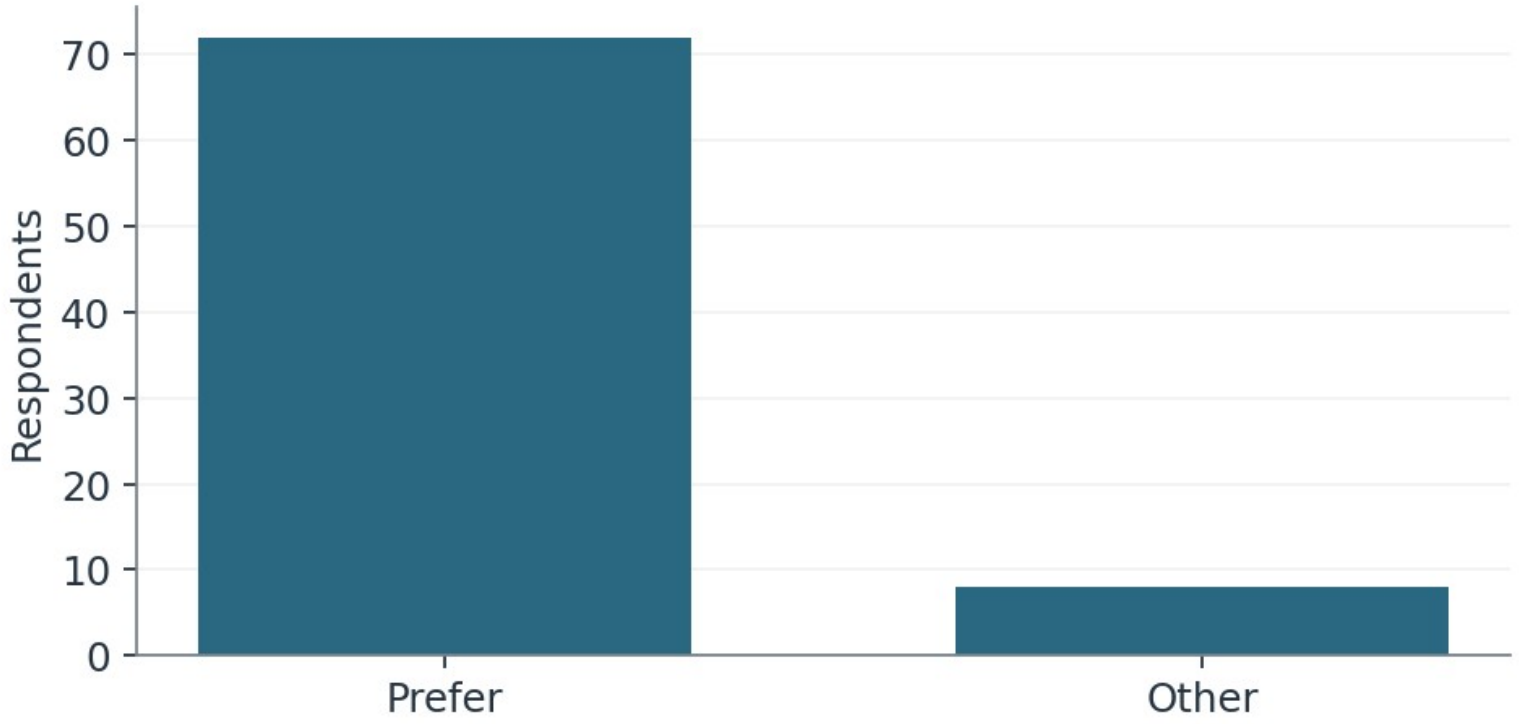
Definitions: Evaluating statistical claims

Statistical claim: a statement about a population or pattern supported by data.

Nonresponse bias: respondents differ systematically from people who do not respond.

Leading question: wording that encourages a particular answer.

Illustration and interpretation



Notice the pattern

A poll shared on a homework app reaches students who already use that app. The headline may overstate the evidence.

Illustrative data / model

Important ideas and notation

Check who collected the data, the target population, sample selection, sample size and nonresponse. A large biased sample is still biased.

Compare the headline with the actual question and data. Leading questions can change responses; an association does not establish causation.

A stronger report states the method, limitations and exact denominator. Ask whether the sample supports the scope of the claim.

Example 1.5.1

An online poll of 80 volunteers finds 72 prefer digital homework. A headline claims 90% of all students prefer it. Evaluate.

Solution:

$\frac{72}{80}$ = 90% of respondents. Voluntary response may bias the result

the population claim is unsupported.

Example 1.5.2

A survey asks: “Do you agree our excellent library should stay open longer?” Revise the question.

Solution:

Use neutral wording: “Should the library close earlier, at the same time, or later?”

Example 1.5.3 • Read beyond the headline

Can we claim that 90% of every student in the school prefers the app?

Response	Count
Prefer app	72
Do not prefer app	8
Total respondents	80

Illustrative classroom dataset

Example 1.5.3 • Solution

Can we claim that 90% of every student in the school prefers the app?

Solution:

No. $\frac{72}{80}$ is the respondent proportion

volunteers may differ from the population.

Individual practice

1. Of 50 students surveyed in an advanced class, 40 like math. Can this represent every student?

2. A report links more exercise with better grades. Does it prove exercise causes higher grades?

3. A graph shows 51% versus 53% with a vertical axis starting at 50%. Explain the concern.

Group activity

Write a short fictional headline about student transportation. Design a fair survey and rewrite the headline so it accurately reflects a sample of 60 with 36 bus riders.

Show your calculation or graph. Write a sentence explaining the result in context.

Check your answers

1. 80% of that sample; not necessarily representative of all students.

2. No; other variables may explain the association.

3. The truncated axis can exaggerate a 2-percentage-point difference.

Group solution

Solution:

Example: randomly select 60 students from the roster
ask their usual transport mode neutrally. Headline: 60% of surveyed students usually ride the bus. Mention sampling uncertainty.

Compare your conclusion with the question. Check your units and explain what the numbers represent.

Exit ticket • 5 minutes

1. Why does a large sample not automatically eliminate bias?

2. A poll of 200 gym members finds 180 exercise daily. Is that evidence that 90% of all residents exercise daily?

Show your reasoning. Submit before leaving.