

AP Statistics · Collecting Data: Sampling and Experiments · Lesson

Common Core AP Statistics Unit 1 (2026-27) · mathovore.fr

This ap statistics collecting data lesson explains collecting data: sampling and experiments in one printable page, with definitions, a method and a worked example.

Statistics begins with how data are collected. Random sampling methods, such as simple random, stratified, cluster and systematic samples, avoid bias when generalizing to a population. Only an experiment with random assignment can support a cause-and-effect conclusion, while an observational study shows association only.

AP StatisticsCommon Core: AP Statistics Unit 1 (2026-27)

AP Statistics Collecting Data Lesson: key facts

- Simple random, stratified, cluster and systematic samples use chance; convenience and voluntary response samples are biased.
- Sources of bias include undercoverage, nonresponse and response bias such as a leading question; a bigger sample does not fix bias.
- Observational studies show association; experiments with random assignment can show cause and effect.
- Good experiments use comparison, random assignment, replication and blocking, and may be blind.
- Random selection allows generalizing to a population; random assignment allows causal claims.

What is a statistical investigative question?

A **statistical investigative question** expects answers that vary and can be settled only by collecting data. “How many minutes did Maya study last night?” has one answer. “How many minutes do students at Roosevelt High typically study on a school night?” is statistical because the answers differ from student to student.

Key vocabulary

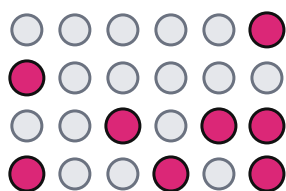
- The **population** is the whole group we want to learn about; the **sample** is the part we measure.

- A **parameter** describes a population. A **statistic** describes a sample, such as $\hat{p}$ or $\bar{x}$.

How do the random sampling methods differ?

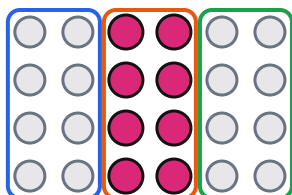
A **simple random sample (SRS)** gives every group of the chosen size the same chance of selection. A **stratified** sample splits the population into similar groups (strata) and takes an SRS from each. A **cluster** sample splits it into mixed groups (clusters), randomly picks some clusters, and surveys everyone in them. A **systematic** sample picks a random start and then every k th individual.

Simple random sample



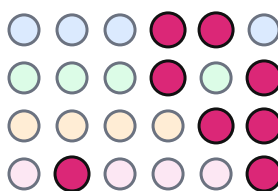
8 of 24 chosen at random

Cluster (column pairs)



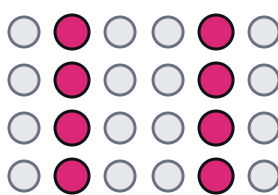
1 of 3 clusters, all 8 surveyed

Stratified (rows)



2 chosen at random per row

Systematic (every 3rd)



start at 2, then every 3rd

Four random sampling methods applied to the same 24 individuals. Pink dots are the 8 chosen.

What causes bias in a sample?

Bias means the method systematically favors some outcomes. A **convenience** sample uses whoever is easy to reach, and a **voluntary response** sample lets people select themselves. **Undercoverage** leaves part of the population with no chance of selection, **nonresponse** happens when chosen people do not answer, and **response bias** comes from leading questions or untruthful answers. A larger sample never repairs a biased method.

How is an experiment different from an observational study?

An **observational study** records data without imposing anything, so it can show association but not cause and effect: a **confounding variable** may be mixed up with the explanatory variable. An **experiment** assigns **treatments** to **experimental units** and measures a **response variable**.

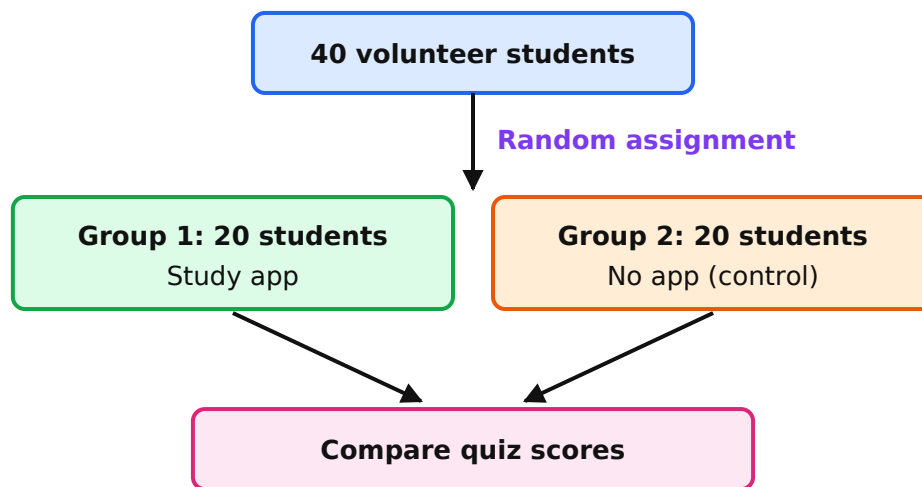
What makes an experiment well designed?

Four principles: **comparison** (a control group or placebo), **random assignment** (balances other variables across groups), **replication** (enough units), and **blocking** (grouping similar units first). In a **blind** experiment participants, evaluators, or both do not know who got which treatment.

How to run a completely randomized design

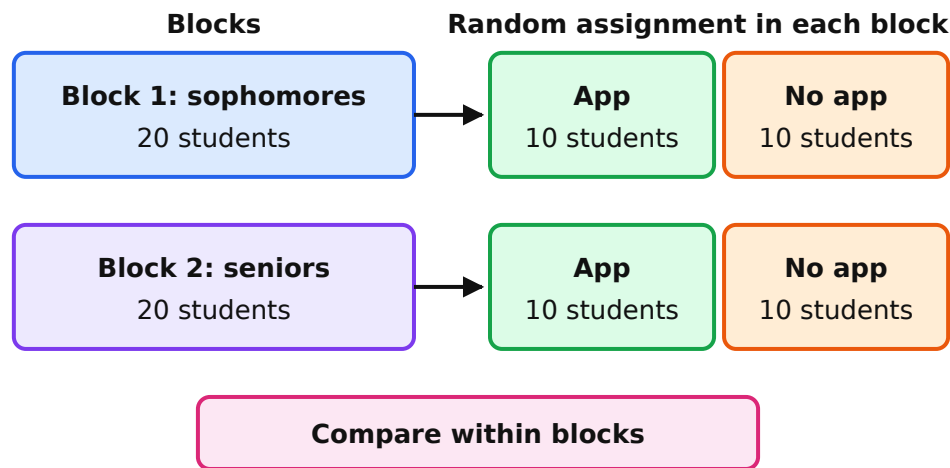
1. Name the units, the treatments, and the response variable.
2. Label every unit, then use random digits to pick the units for the first treatment.
3. Give the remaining units the other treatment, and treat both groups alike in every other way.
4. Measure the response and compare the groups.

Example. Forty volunteers label themselves 01 to 40. Random digits select 20 of them for the study app; the other 20 get no app. Then compare quiz scores.



A completely randomized design with two treatments.

If sophomores and seniors might react differently, **block** by grade first and randomize inside each block. A **matched-pairs** design is a block design with blocks of size two, such as one person trying both treatments in random order.



Blocking by grade level, then randomizing within each block.

What conclusions are justified?

Random selection lets us generalize to the population sampled. Random assignment lets us claim that the treatment caused the difference.

	Random assignment	No random assignment
Random sample	Cause and effect for the whole population	Generalize to the population only
Not random	Cause and effect, volunteers only	Neither conclusion is justified

Random selection controls what we can generalize to. Random assignment controls whether we can claim causation.

Collecting Data: Sampling and Experiments: frequently asked questions

What is the difference between a stratified sample and a cluster sample?

A stratified sample divides the population into similar groups and samples some individuals from every group. A cluster sample divides it into mixed groups, picks a few whole groups at random, and surveys everyone in those groups.

Does a larger sample remove bias?

No. Bias comes from the sampling method, so a biased method stays biased at any sample size. A larger random sample only reduces chance variation.

Why is random assignment so important in an experiment?

Random assignment tends to balance all other variables, known and unknown, across the treatment groups. Differences in the response can then be credited to the treatments.

What is a confounding variable?

A confounding variable is related to both the explanatory and the response variable, so their effects cannot be separated. Hot weather confounds ice cream sales and sunburns, for example.

When can we generalize results to a population?

When the subjects were chosen by random selection from that population. Generalizing to a larger group from volunteers or a convenience sample is not justified.

How to use this ap statistics collecting data lesson

Use this ap statistics collecting data lesson before the worksheet, then check what you learned with the quiz.

