

Sampling Bias

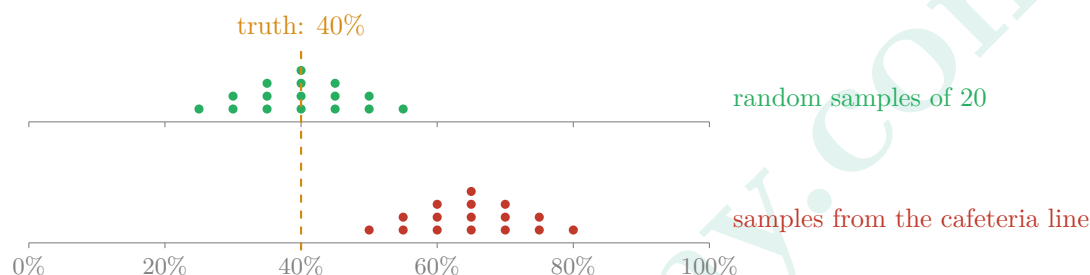
AP Statistics | Collecting Data | what goes wrong when a sample is not representative

1. What bias means

Bias

A sampling method is **biased** if it *systematically* overestimates or underestimates the true value: it misses in the same direction, every time.

Random error shrinks as the sample grows. **Bias does not.** A bigger biased sample gives the wrong answer with more confidence.



In a school where 40% of students like the cafeteria food, random samples scatter *around* 40%; samples taken only from the cafeteria line scatter around the wrong value. (Illustration; hypothetical data.)

2. Biased ways to choose a sample

- **Convenience sample:** choose whoever is easiest to reach. The first 20 students in the cafeteria line: 13 of 20 like the food (65%, truth 40%). People who are easy to reach differ from everyone else.
- **Voluntary response sample:** people decide for themselves whether to take part (online polls, call-ins, reviews). Those who respond tend to have strong opinions.

3. Problems that can hit even a random sample

Problem	What happens	Example
Undercoverage	part of the population has a smaller chance (or no chance) of being chosen	sampling from a school-email list that 30 students are not on
Nonresponse	people who <i>were</i> chosen cannot be reached or refuse	12 of the 20 chosen students never answer
Response bias	people answer, but inaccurately: loaded wording, sensitive questions, the interviewer	“Who has cheated on a test?” asked by the teacher, by show of hands

Voluntary response vs nonresponse: in voluntary response nobody was chosen; in nonresponse people were chosen and then did not answer.

Nonresponse vs response bias: no answer vs a wrong answer.

4. Writing the answer

A bias answer has three parts

1. **Name** the problem, in context: *undercoverage: students without a school email cannot be chosen.*
2. **Explain** how those people are likely to differ: *they may eat at home more and like the cafeteria less.*
3. Give the **direction**: *so the sample proportion likely overestimates the true proportion of students who like the food.*

1936, Literary Digest. About 2.4 million responses predicted Landon would win with 57%; Roosevelt won 46 of 48 states. *Undercoverage*: names came from phone books, car registrations and subscribers (wealthier voters). *Nonresponse*: 10 million ballots mailed, about three in four never returned. A huge sample, chosen badly.

Where students lose points

1. Naming the bias without the *direction* (too high or too low?).
2. “The sample is biased” with no reason in context.
3. Mixing up nonresponse and voluntary response.
4. Mixing up nonresponse and response bias.
5. Saying a larger sample would reduce the bias.

Now practise it

Theory sticks once you use it. On **sad-academy.com** you'll find practice questions for this exact subtopic in the AP format, and an AI that grades your free-response answers against the scoring guidelines, the way an AP reader would.

sad-academy.com

Unit 3 in the 2019 AP Statistics course description; collecting data is in Unit 1 from fall 2026. School data are hypothetical; poll figures are approximate. AP® is a registered trademark of College Board, which is not affiliated with and does not endorse this material.