

# AP Statistics - Sampling Methods and Bias in Surveys - Lesson

Common Core AP STAT 3.1, AP STAT 3.2, AP STAT 3.3, HSS-IC.B.3 · [maths-pdf.fr/en/](https://maths-pdf.fr/en/)

**A sample is the part of a population we actually measure, and a statistic from a random sample estimates a population parameter. Random sampling methods include simple random, stratified, cluster, and systematic samples. Convenience and voluntary response samples are biased, and a larger sample does not remove bias from a poor method.**

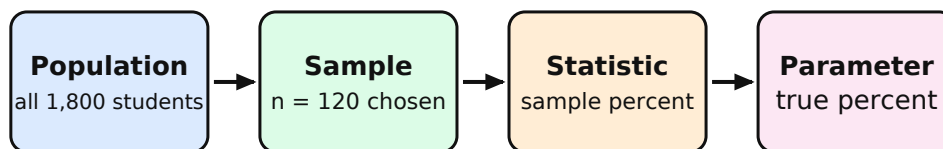
**AP Statistics** · Common Core State Standards: **AP STAT 3.1, AP STAT 3.2, AP STAT 3.3, HSS-IC.B.3**

## Key facts

- Population vs sample; parameter vs statistic.
- SRS gives every group of  $n$  individuals an equal chance; stratified takes an SRS from every stratum; cluster surveys all members of randomly chosen clusters.
- Convenience and voluntary response samples are biased.
- Sources of bias: undercoverage, nonresponse, response bias, voluntary response.
- A larger random sample reduces sampling variability but never fixes bias.

## What are populations, samples, parameters, and statistics?

A **population** is the entire group we want information about. A **sample** is the part of the population that we actually examine. A number that describes the population is a **parameter**; a number that describes the sample is a **statistic**.

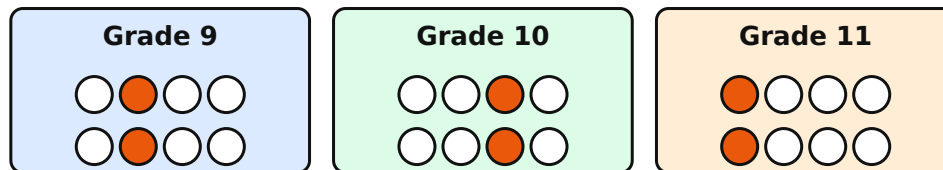


A statistic describes a sample; a parameter describes a population.

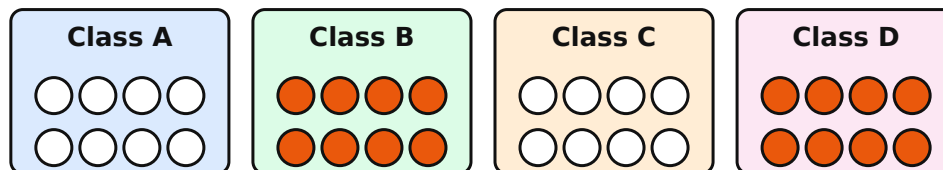
For example, if 54 of 120 students in a random sample of a school's 1,800 students ride the bus, the statistic is  $\hat{p} = \frac{54}{120} = 0.45$ . The parameter, the percent of all 1,800 students who ride the bus, is unknown.

## What are the main random sampling methods?

- **Simple random sample (SRS):** every group of  $n$  individuals has the same chance of being the sample. Label the individuals, then choose labels with random digits or a random number generator.
- **Stratified random sample:** split the population into similar groups (strata), then take an SRS from each stratum.
- **Cluster sample:** split the population into naturally occurring groups (clusters), randomly choose some clusters, and survey everyone in them.
- **Systematic random sample:** choose a random starting point, then take every  $k$ th individual.



Stratified random sample: split the population into strata, then take an SRS from every stratum.



Cluster sample: randomly select whole clusters and survey every member of those clusters.

Strata should be alike inside and different from each other; clusters should each look like a small version of the population.

## What causes bias?

A sampling method is **biased** if it systematically favors some outcomes. Convenience samples (asking whoever is nearby) and voluntary response samples (call-in polls) are almost always biased. Increasing the sample size does not fix bias; only a better selection method does.

| Source of bias            | What goes wrong                                    |
|---------------------------|----------------------------------------------------|
| <b>Undercoverage</b>      | Some group has no chance of selection              |
| <b>Voluntary response</b> | People choose to respond; strong opinions dominate |
| <b>Nonresponse</b>        | Chosen people do not answer                        |
| <b>Response bias</b>      | Wording or setting pushes answers                  |

Four common sources of bias in surveys.

## How do you select a simple random sample with random digits?

1. Label every individual with a number of the same length, such as 01 to 60.
2. Read random digits in groups of that length.
3. Ignore labels that are out of range or repeated.
4. Continue until you have  $n$  individuals, then name them.

Example: with labels 01 to 60 and digits 8 4 | 1 7 | 6 0 | 1 7 | 0 3, the first label 84 is skipped, 17 and 60 are chosen, the repeat 17 is skipped, and 03 is chosen.

## Questions about sampling methods and bias in surveys

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

A stratified sample takes some individuals from every group (strata). A cluster sample picks a few whole groups at random and includes every member of them. Strata are chosen to be alike inside; clusters are meant to mirror the population.

### Does a bigger sample fix bias?

No. A large biased sample is still biased. A voluntary response poll with 50,000 responses is less trustworthy than a random sample of 1,000.

### What is undercoverage?

Undercoverage happens when some members of the population have no chance of being selected, such as a phone survey that excludes people without a landline.

### What does it mean for a statistic to estimate a parameter?

A statistic from a random sample is used as the best guess for the unknown population parameter. It will vary from sample to sample, but a random method makes the error predictable.

### Why use random digits instead of picking names myself?

Humans unconsciously favor some individuals. Random digits give each individual a known, fair chance of being chosen, which prevents selection bias.