

AP Statistics - Variables and Categorical Data Displays - Lesson

Common Core AP STAT 1.1, AP STAT 1.2, AP STAT 1.3, HSS-ID.A.1 · maths-pdf.fr/en/

A **categorical variable** sorts individuals into groups, such as favorite lunch. Its distribution tells which categories occur and how often, using counts or relative frequencies (proportions or percents). Display it with a bar graph or, when the categories add to a whole, a pie chart. Compare groups of different sizes with relative frequencies.

AP Statistics · Common Core State Standards: **AP STAT 1.1, AP STAT 1.2, AP STAT 1.3, HSS-ID.A.1**

Key facts

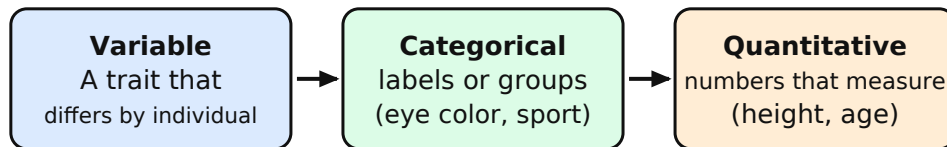
- Individuals are the objects described by data; a variable is a trait that varies among them.
- Categorical variables name groups; quantitative variables are measured with numbers.
- Relative frequency = category count divided by total, and relative frequencies add to 1.
- Bar graphs have a zero baseline and gaps between bars; pie charts need parts of one whole.
- Compare different-sized groups with percents, not raw counts.

What are individuals and variables?

Statistics begins with data. The **individuals** are the people, animals, or things we collect information about. A **variable** is any characteristic that can take different values from one individual to another.

Two kinds of variables

A **categorical variable** puts each individual into a group or category, such as favorite lunch or the sport a student plays. A **quantitative variable** takes numerical values for which arithmetic makes sense, such as height in inches or minutes spent on homework.

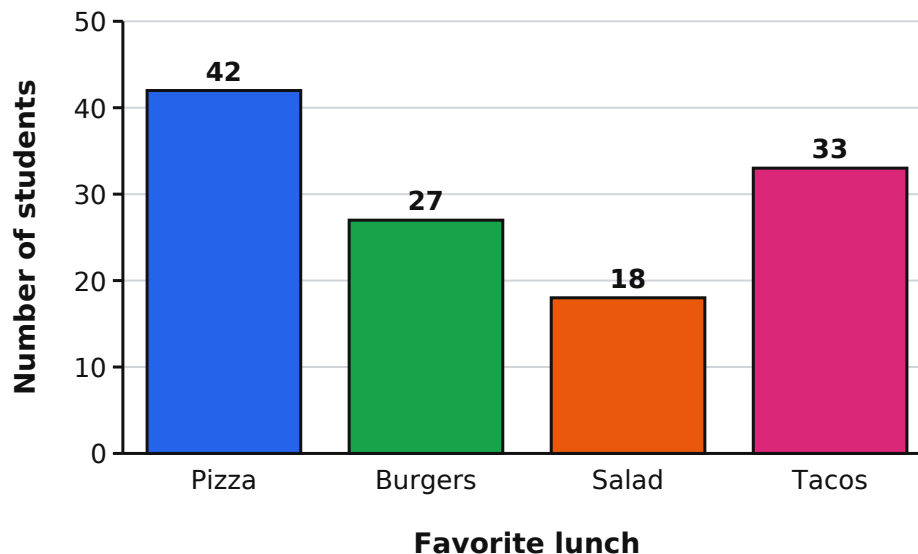


A variable is categorical when it places individuals into groups and quantitative when it measures them with numbers.

Be careful with numbers that act like labels. Zip codes and jersey numbers are categorical because averaging them has no meaning.

How do you display a categorical distribution?

The **distribution** of a categorical variable lists the categories and how often each occurs. A **frequency table** gives counts. A **relative frequency table** gives each count as a proportion or percent of the total.



Bar graph of the lunch survey. Each bar height is a count; the bars do not touch.

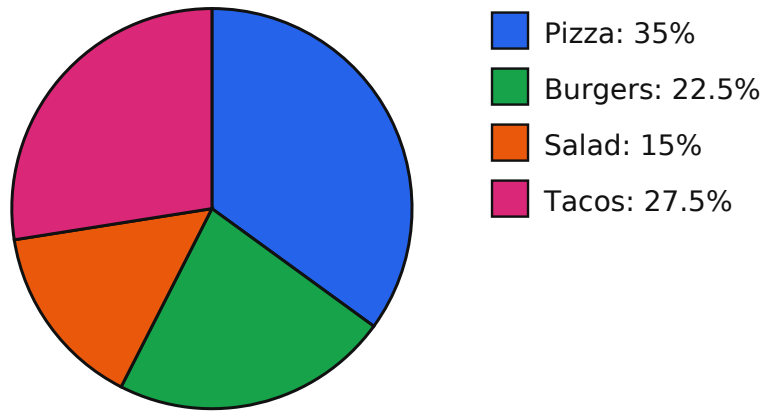
Lunch	Count	Relative frequency
Pizza	42	0.350
Burgers	27	0.225
Salad	18	0.150
Tacos	33	0.275
Total	120	1.000

Relative frequency = count divided by total. The relative frequencies add to 1.

In the survey of 120 students, $\frac{42}{120} = 0.35$ chose pizza, so 35% of the sample chose pizza.

When should you use a bar graph or a pie chart?

A **bar graph** works for any categorical data. Bars have equal widths, a gap between them, and a vertical axis that starts at 0. A **pie chart** is appropriate only when the categories cover the whole group exactly once.

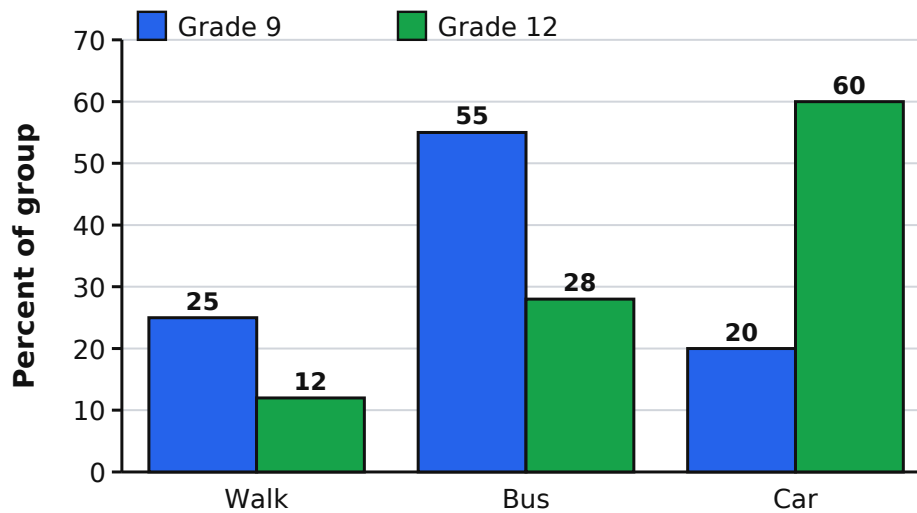


A pie chart shows the same data as parts of one whole. Use it only when the slices add to 100%.

1. Name the variable and the individuals.
2. Count the individuals in each category and find the total n .
3. Divide each count by n to get relative frequencies.
4. Check that the relative frequencies add to 1 (or 100%).
5. Graph the result, then describe the distribution in context.

How do you compare two categorical distributions?

When groups have different sizes, compare relative frequencies, not raw counts. In the graph below, only 16 ninth graders but 75 twelfth graders drive, yet the groups have 80 and 125 students, so the fair comparison is 20% against 60%.



How students get to school

Comparing groups of different sizes works best with percentages (relative frequencies).

A good description names the most and least common categories, uses percents, and mentions the context.

Questions about variables and categorical data displays

Is a zip code a categorical or quantitative variable?

A zip code is categorical. Although it is written with digits, adding or averaging zip codes tells you nothing, so the digits only act as labels for locations.

What is the difference between frequency and relative frequency?

Frequency is the count of individuals in a category. Relative frequency is that count divided by the total number of individuals, written as a decimal or a percent.

Why must relative frequencies add to 1?

Every individual falls in exactly one category, so the pieces together account for the whole group. If they add to 0.94, a category is missing or a count was entered wrongly.

When is a pie chart a poor choice?

Avoid a pie chart when respondents could pick more than one category or when the categories leave out part of the whole. A bar graph handles these cases without distortion.

How do you describe a categorical distribution on the AP exam?

Name the variable in context, state the most and least common categories with percents or counts, and compare groups using relative frequencies. Never answer with only a list of

numbers.