Categorical and numerical data

Some data are names of groups, like a fruit or a color. Some are amounts, like a mass or a count. How do you tell them apart, and why does it matter?

WORDS TO KNOW

Categorical variable: each value names a group, or **category**, such as plum, cash or red.

Label: a number used as a name, such as a jersey number or a ZIP code.

Continuous: the values come from measuring a length, mass or time. Any value in a range is possible.

Numerical variable: each value is a number that counts or measures an amount, in a unit.

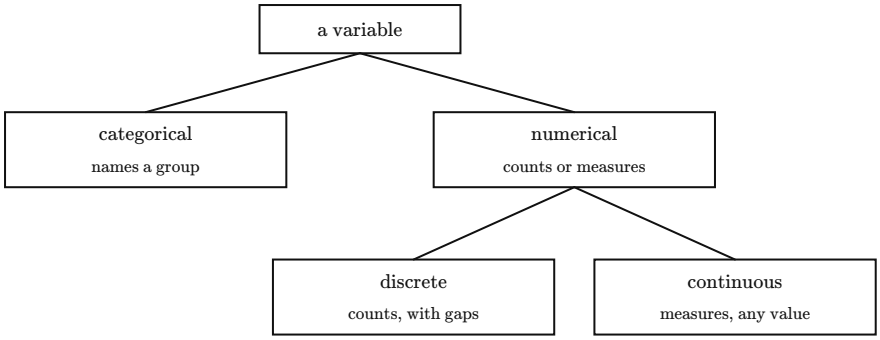
Discrete: the values come from counting. A count can be 3 or 4, never 3.5.

KEY IDEAS

1. The test: does an average make sense?

Add the values and share the total out equally. If the result is an amount of the same thing, the variable is numerical. If it means nothing, the variable is categorical.

> Numbers that are labels are categorical. Jersey numbers, ZIP codes, phone numbers and room numbers are names written with digits. Their average is not an amount of anything.



2. Counts are discrete, measurements are continuous

A count answers "how many?", so it is a whole number with gaps between. A measurement can fall anywhere: between any two possible masses there is always another one.

> A measured value is always rounded to what the tool can show. It stays continuous even though the recorded values jump in steps.

3. The type decides the summary

Categorical	Numerical
Count each category	Put the values in order
Name the most common one	Add, subtract and average
Give each count as a fraction of the whole	Describe the spread

Free sample from math.wiki for classroom use. Do not post. math.wiki

WATCH OUT

- **An order does not make a variable numerical.** Small, medium and large, or 1st, 2nd and 3rd, can be put in order, but the gap between them is not an amount. They are categorical.
- **An average of a count need not be a possible count.** 71 pieces of fruit shared by 5 customers is 14.2 pieces each. Nobody buys 14.2 pieces, but the average still means something.
- **Steps on a scale do not make mass discrete.** The steps come from the scale, not from the thing weighed.

WORKED EXAMPLES

Example 1. A pet shop records, for each fish tank: (a) the kind of fish in it, (b) the number of fish, (c) the water temperature, (d) the tank number, (e) the length of the tank in centimeters. Sort each variable.

1. (a) The kind of fish names a group: categorical.
2. (b) The number of fish is counted: numerical and discrete.
3. (c) Temperature is measured: numerical and continuous.
4. (d) The tank number only names the tank. Its average means nothing: categorical.
5. (e) Length is measured: numerical and continuous.

Answer: (a) and (d) categorical; (b) numerical, discrete; (c) and (e) numerical, continuous.

Example 2. Four city buses run on routes 4, 9, 15 and 22. On one trip they carry 30, 12, 25 and 17 riders, in that order. Find the average of each list. Which average means something?

1. Routes: $4 + 9 + 15 + 22 = 50$, and $50 \div 4 = 12.5$.
2. Riders: $30 + 12 + 25 + 17 = 84$, and $84 \div 4 = 21$.
3. 84 riders shared equally over 4 buses is 21 riders each, an amount.
4. A route number only names a route. If route 22 were renamed route 40, the buses would carry the same people, but the average route number would change.

Answer: only the riders average, 21 riders per bus, means something. Riders are numerical and discrete; route numbers are categorical.

Example 3. A stopwatch shows time to the nearest tenth of a second. A runner's time reads 12.3 seconds. What could the true time be? Is time discrete?

1. A time rounds to 12.3 when it is at least $12.3 - 0.05 = 12.25$ seconds.
2. It must also be less than $12.3 + 0.05 = 12.35$ seconds, or it would round to 12.4.
3. Time is measured, so it is continuous. A finer stopwatch would show values between the tenths.

Answer: at least 12.25 s and less than 12.35 s. Time is continuous.

Free sample from math.wiki for classroom use. Do not post. math.wiki