

# Language of Variables

1. In an AP Statistics study, the variable "hours of sleep per night" is recorded for each student. What type of variable is this?

- A. Categorical
- B. Quantitative
- C. Explanatory label
- D. Identifier

---

2. A survey records each student's "favourite school subject" as Math, English, Science, or History. Which statement is correct?

- A. The variable is quantitative because it can be counted.
- B. The variable is categorical because it places individuals into groups.
- C. The variable is continuous because there are many subjects.
- D. The variable is discrete because there are four choices.

1. In an AP Statistics study, the variable "hours of sleep per night" is recorded for each student. What type of variable is this?

- A. Categorical
- B. Quantitative
- C. Explanatory label
- D. Identifier

**Answer:** B. Quantitative

---

2. A survey records each student's "favourite school subject" as Math, English, Science, or History. Which statement is correct?

- A. The variable is quantitative because it can be counted.
- B. The variable is categorical because it places individuals into groups.
- C. The variable is continuous because there are many subjects.
- D. The variable is discrete because there are four choices.

**Answer:** B. The variable is categorical because it places individuals into groups.

- **Variable (变量)** – a characteristic or attribute that can take on different values.
- **Categorical variable (分类变量 / 类别变量)** – a variable that places individuals into categories.
- **Quantitative variable (数值变量 / 定量变量)** – a variable measured with numbers that make sense to average or compute differences.
- **Explanatory variable (解释变量 / 自变量)** – a variable that may explain or predict changes in another variable.
- **Response variable (响应变量 / 因变量)** – a variable that measures an outcome or result of the study.

Here are the **key points for the “language of variation” and variables** (AP Statistics–friendly and concise):

- **Variables** describe characteristics that can take different values across individuals.
  - **Variation** means data values differ; no two data sets are exactly the same.
    - Statistics focus on describing, explaining, and measuring this variation.
      - Both **center** and **spread** are needed to fully describe a distribution.
  - Conclusions must be stated **in context**, acknowledging natural variability in data.

- **Individuals** → the objects described by the data (students, hospitals, cars, states, plants, etc.)
- **Variable** → a characteristic recorded for each individual (height, blood type, major, income, etc.)
- **Data** → the actual values of the variable (65 inches, A+, biology, 42000 dollars, etc.)

Here's how they fit together in one chain:



## Categorical Variables

A [categorical variable](#) takes values that are group labels or category names.

Examples:

- Dominant hand (left/right)
- Age group (child, teen, adult)
- Highest degree earned
- Political party
- Eye color

Key features:

- Values are labels, not numbers with meaning.
- You cannot average them.
- Some have a natural order (small, medium, large), but they are still categorical.

## Quantitative Variables

A [quantitative variable](#) takes numerical values for a measured or counted quantity.

Examples:

- Age of a structure (years)
- Height of a child (inches or cm)
- Concentration of a sample
- Number of siblings

Arithmetic makes sense here. You can compute a mean or standard deviation.

Quantitative variables come in two forms.

### Discrete

- Counts
- Usually whole numbers
- Example: number of injuries, number of pets

### Continuous

- Measurements
- Can take any value in an interval
- Example: height, time, concentration

# Discrete or Continuous

**Instructions:** For each of the following variables, determine whether it is **continuous** or **discrete**.

1. The number of books a student reads in a month.
2. The height of a sunflower in centimeters.
3. The number of cars that pass through a toll booth in an hour.
4. The amount of milk (in liters) in a bottle.
5. The number of siblings a person has.
6. The time (in seconds) it takes to run 100 meters.
7. The number of goals scored in a soccer match.
8. The weight of a backpack in kilograms.
9. The number of students absent from class on a day.
10. The temperature in a city at noon.

# Discrete or Continuous

1. Discrete
2. Continuous
3. Discrete
4. Continuous
5. Discrete
6. Continuous
7. Discrete
8. Continuous
9. Discrete
10. Continuous

A researcher collects data on 50 students, recording their **study hours per week** and their **final exam scores**.

- (a) Identify the **explanatory** and **response** variables.
- (b) State whether each variable is **categorical** or **quantitative**.
- (c) Explain how variation in the explanatory variable might affect the response variable.

**(a) Explanatory and Response Variables:**

- **Explanatory variable:** Study hours per week
- **Response variable:** Final exam scores

**(b) Type of Variable:**

- Study hours → **Quantitative**
- Final exam scores → **Quantitative**

**(c) Explanation of Variation:**

- Differences in study hours may lead to differences in exam scores; typically, more study hours correspond to higher scores, showing how variation in the explanatory variable affects the response variable.

1. A school survey collects the following information from students:

- Year group
- Height in centimetres
- Favourite music genre
- Number of siblings

Classify each variable as either **categorical** or **quantitative**. Explain your reasoning for each.

---

2. A researcher wants to study whether time spent exercising affects resting heart rate.

- Identify the explanatory variable.
  - Identify the response variable.
  - State whether each variable is categorical or quantitative.
  - Explain why distinguishing between explanatory and response variables is important in statistics.
- 

3. A sports analyst records the following variables for basketball players:

- Position played
- Points scored per game
- Shoe size
- Team name

- Identify all categorical variables.
  - Identify all quantitative variables.
  - Which quantitative variable is most likely continuous? Explain.
  - Which quantitative variable is most likely discrete? Explain.
- 

4. A company collects customer data including:

- Customer ID number
- Satisfaction rating from 1–10
- Preferred payment method
- Monthly spending in dollars

Explain why some numbers in a data set do not necessarily represent quantitative variables. Use examples from the data above in your answer.

**1.**

- Year group — categorical
  - Height — quantitative
  - Favourite music genre — categorical
  - Number of siblings — quantitative
- 

**2.**

- a) Explanatory: time spent exercising
  - b) Response: resting heart rate
  - c) Both are quantitative
  - d) It helps identify which variable may influence the other.
- 

**3.**

- a) Categorical: position played, team name
  - b) Quantitative: points scored per game, shoe size
  - c) Continuous: shoe size
  - d) Discrete: points scored per game
- 

**4.**

Customer ID numbers are labels, not quantitative variables, because arithmetic with them is meaningless.  
Satisfaction rating and monthly spending are quantitative.  
Preferred payment method is categorical.

# One variable data- descriptive statistics

- **Variable** — 变量
- **Categorical variable** — 分类变量
- **Quantitative variable** — 定量变量
- **Distribution** — 分布
- **Center** — 中心
- **Spread** — 离散程度 / 变异性
- **Outlier** — 离群值
- **Mean** — 平均数
- **Median** — 中位数
- **Standard deviation** — 标准差

1. A variable can be **categorical or quantitative**, and this determines how it is analysed.
2. The **distribution** of data describes its overall pattern.
3. Key features of a distribution are **shape, center, and spread**.
4. The **mean and median** are measures of center.
5. The **standard deviation and range** measure spread.
6. The **mean is sensitive to outliers**, while the median is resistant.
7. An **outlier** is a value that lies unusually far from the rest of the data.
8. A distribution can be **symmetric or skewed (left/right)**.
9. Graphs such as **histograms, boxplots, and dotplots** are used to display data.
10. Describing data involves **both numerical summaries and visual interpretation**.

1. A variable can be **categorical or quantitative**, and this determines how it is analysed.
2. The **distribution** of data describes its overall pattern.
3. Key features of a distribution are **shape, center, and spread**.
4. The **mean and median** are measures of center.
5. The **standard deviation and range** measure spread.
6. The **mean is sensitive to outliers**, while the median is resistant.
7. An **outlier** is a value that lies unusually far from the rest of the data.
8. A distribution can be **symmetric or skewed (left/right)**.
9. Graphs such as **histograms, boxplots, and dotplots** are used to display data.
10. Describing data involves **both numerical summaries and visual interpretation**.

A teacher records the test scores (out of 100) of 12 students:

42, 55, 61, 63, 65, 67, 70, 72, 74, 78, 85, 96

---

(a)

Calculate the **mean** and **median** of the data.

---

(b)

Describe the **shape of the distribution**. Justify your answer.

---

(c)

Identify any **outliers** using the  $1.5 \times \text{IQR}$  rule.

---

(d)

Which measure of center (mean or median) is more appropriate to represent this data? Explain.

---

(e)

Describe the **spread** of the data using an appropriate numerical measure.

✓ Concise Answers

(a)

- Mean = 69.0
  - Median = 68.5
- 

(b)

- Distribution is **slightly right-skewed** (due to high value 96).
- 

(c)

- Q1 = 62, Q3 = 76
  - IQR = 14
  - Lower fence = 41, Upper fence = 97
  - **No outliers**
- 

(d)

- **Mean**, because there are **no outliers** and data is fairly balanced (though slight skew).
- 

(e)

- Standard deviation  $\approx$  13.4 (or) Range = 54

# TRUE or FALSE

- The range is a resistant measure of spread.

# TRUE or FALSE

- A skewed distribution is not symmetric.

# TRUE or FALSE

- Boxplots show median and quartiles.