

Please . . .

L

F

A

R

N

Introducing Statistics – do the data tell the truth

Multiple-Choice Questions

1. Which sampling method gives every individual in the population an equal chance of being selected?

- A. Convenience Sample
- B. Systematic Sample
- C. Simple Random Sample
- D. Voluntary Response Sample

2. A principal divides students into Grade 9, 10, 11, and 12 groups and randomly selects 20 students from each grade. What type of sampling is this?

- A. Cluster Sampling
- B. Stratified Sampling
- C. Convenience Sampling
- D. Systematic Sampling



Changes in AP statistics

Major Changes

1. Reorganized Course Structure

The course has been streamlined and reorganized into fewer, larger units. Some topics have been reduced or removed, while greater emphasis is placed on statistical reasoning, data analysis, and interpretation.

2. Changes to Exam Format

The AP Statistics exam will include:

42 multiple-choice questions (up from 40)

4 answer choices per question instead of 5

New multiple-question sets that assess related statistical concepts together.

3. Fully Digital Exam (Starting May 2027)

The exam will move from the current hybrid format to a **fully digital exam in Bluebook**:

Multiple-choice and free-response questions will both be completed digitally.

Students may still use an approved graphing calculator.

A built-in Desmos calculator will also be available.

Changes in AP statistics

4. Content Changes

Topics reported as removed or reduced include:

Cumulative frequency graphs (ogives)

Outliers, high-leverage points, and influential points in regression analysis

Some inference procedures and probability distributions have been streamlined.

5. Greater Emphasis on Statistical Investigation

Students are expected to:

Ask statistical questions

Analyze data in context

Interpret results

Communicate conclusions clearly

This reflects the growing focus on data literacy and real-world applications of statistics.

The NEW AP Statistics Units — 2026–27

New Unit	Title	MC Exam Weight 
Unit 1	Exploring One-Variable Data and Collecting Data	20–30%
Unit 2	Probability, Random Variables, and Probability Distributions	15–25%
Unit 3	Inference for Categorical Data: Proportions	15–25%
Unit 4	Inference for Quantitative Data: Means	10–20%
Unit 5	Regression Analysis	10–20%

- **Bias (偏差 / 偏倚)** – systematic error that favors certain outcomes.
- **Sampling bias (抽样偏差)** – when some members of the population are more likely to be selected than others.
- **Response bias (回应偏差)** – when participants give inaccurate or dishonest answers.
- **Confounding variable (混杂变量)** – a variable that affects the response and is related to the explanatory variable.
- **Causation (因果关系)** – when changes in one variable directly cause changes in another.

- Data do not automatically tell the truth without proper analysis.
- Bias in data collection can lead to misleading conclusions.
- The sample must be representative of the population. Graphs and statistics can be used to mislead if not chosen carefully.
- Sound conclusions require attention to context, variability, and how the data were collected.

A school wants to determine whether a new tutoring program improves math test scores. Participation in the program is voluntary, and students who choose to participate tend to be more motivated than those who do not. After one semester, students in the tutoring program have higher average test scores than students who did not participate.

- (a) Identify a potential **source of bias** in this study.
- (b) Name a possible **confounding variable**.
- (c) Explain why the results of this study **do not necessarily show causation**.

(a) Source of bias:

- **Sampling bias (抽样偏差):** Since participation in the tutoring program is voluntary, students who choose to participate may be more motivated or have higher initial abilities, which could skew the results.

(b) Possible confounding variable:

- **Motivation (动机):** More motivated students may be more likely to participate in the tutoring program and also may perform better on the test, making it difficult to isolate the effect of the tutoring itself.

(c) Why the results do not necessarily show causation:

- **Confounding:** The results may be due to the fact that more motivated students participated in the tutoring program, not because the program itself caused the improvement in test scores. Without controlling for motivation or other factors, we cannot conclude that the tutoring program caused the higher scores.

Stretch and Challenge

1. Chocolate and Exam Scores

A study finds that students who eat chocolate before an exam score 10% higher than students who do not.

Questions:

- Does this prove chocolate improves exam performance?
 - What lurking variables might explain the relationship?
 - How could an experiment be designed to test causation?
-

2. Social Media Poll

An online poll asks:

"Should schools stop wasting money on unnecessary homework?"

Results show that 85% of respondents agree.

Questions:

- What sources of bias are present?
- Is the wording neutral?
- Can the results be generalized to all students?

3. The Miracle Diet

An advertisement claims:

"95% of our customers lost weight using our diet plan."

Questions:

- What additional information would you want before believing this claim?
 - How might the data have been collected?
 - What information might be missing?
-

4. Ice Cream and Drowning

A city finds that on days when more ice cream is sold, there are more drowning incidents.

Questions:

- Does eating ice cream cause drowning?
- What is a likely confounding variable?
- Explain the difference between correlation and causation.

1. **Chocolate & exams:** No causation. Possible confounders: sleep, diet, study habits. Use a randomized experiment.
2. **Social media poll:** Voluntary response bias; leading wording. Cannot generalize to all students.
3. **Miracle diet:** Need sample size, selection method, time period, control group, and dropout information.
4. **Ice cream & drowning:** No. **Temperature/summer** is a confounding variable.

Statistical Thinking – 5 Key Words

English	Chinese	Definition
Data	数据	Information collected for analysis.
Variable	变量	A characteristic that can have different values.
Population	总体	The entire group being studied.
Sample	样本	A smaller group selected from the population.
Bias	偏差	A systematic problem that can cause results to be misleading.

Statistical Thinking – 5 Key Concepts

1. **Data (数据)** – Information collected to answer a question.
2. **Variability (变异性)** – Data values naturally differ from one another.
3. **Association (关联)** – A relationship between two variables.
4. **Causation (因果关系)** – One variable directly causes a change in another.
5. **Inference (统计推断)** – Using sample data to draw conclusions about a population.

FRQ: Do the Data Tell the Truth?



A school newspaper reports:

"Students who eat breakfast every morning have higher mathematics grades."

The newspaper bases its claim on a survey of **500 students**. Students were asked whether they ate breakfast every morning and were asked to report their average mathematics grade. The study found that students who ate breakfast every morning had an average grade of **82**, while students who did not had an average grade of **74**.

- (a) Identify the **population** of interest in this study.
- (b) Identify the **sample**.
- (c) Is this an **observational study or an experiment**? Explain.
- (d) The newspaper concludes that eating breakfast **causes** higher mathematics grades. Explain why this conclusion is not justified.
- (e) Identify **one possible confounding variable** that could explain the association.
- (f) Explain how **bias** in the survey could affect the results.
- (g) Describe how researchers could design an experiment to investigate whether eating breakfast causes higher mathematics grades.

- (a) All students at the school.
- (b) The 500 students surveyed.
- (c) **Observational study** — researchers did not assign students to eat or skip breakfast.
- (d) Association does not imply causation; other variables may explain the difference.
- (e) Possible answers: sleep, socioeconomic status, study habits, attendance, motivation, overall health.
- (f) Students may give inaccurate answers or the students who respond may not represent the whole school.
- (g) Randomly assign students to a breakfast group or control group, control other conditions, measure mathematics performance, and compare the groups.

True or False

- **Data can be used to answer questions and make informed decisions.**

True or False

- **A sample is a smaller group selected from a population.**

True or False

A larger sample automatically eliminates all bias.