

1. Which graph is most appropriate for displaying the relationship between two quantitative variables?

- A. Histogram
 - B. Scatterplot
 - C. Boxplot
 - D. Bar chart
-

2. In a scatterplot, what does each point represent?

- A. A summary of a group
- B. A single observation
- C. A category total
- D. A conditional mean

Scatterplot

散点图

Quantitative variable

定量变量

Explanatory variable

自变量

Response variable

因变量

Correlation

相关

- Use a **scatterplot** to display the relationship between the two variables.
- Examine the **direction** (positive, negative, or none) of the association.
 - Describe the **form** (linear or non-linear).
 - Assess the **strength** of the relationship.
- Identify any **outliers or influential points** and interpret results **in context**.

A researcher collects data on **hours studied** and **exam scores** for 30 students.

(a) Identify the **explanatory** and **response** variables.

(b) Suggest the most appropriate graph to display the relationship between the two variables and explain why.

(c) Explain how you could describe the **strength and direction** of the relationship.

(a)

- **Explanatory variable:** Hours studied (independent variable)
- **Response variable:** Exam score (dependent variable)

(b) A **scatterplot** is most appropriate because it shows the relationship between two **quantitative variables** and allows us to see trends, clusters, and outliers.

(c) The **strength** and **direction** of the relationship can be described by:

- **Direction:** positive (as hours increase, scores tend to increase) or negative (as hours increase, scores decrease)
- **Strength:** how closely the points follow a straight line (strong, moderate, weak)
- Outliers and unusual points can also be noted.

Stretch & Challenge FRQ — Two Quantitative Variables (AP Statistics)**Question**

A researcher investigates the relationship between **hours studied per week** (x) and **AP Statistics exam score** (y) for 12 students. The data are shown below:

Student	Hours Studied (x)	Exam Score (y)
1	1	52
2	2	55
3	3	57
4	4	60
5	5	63
6	6	65
7	7	70
8	8	72
9	9	75
10	10	78
11	11	80
12	12	95

- (a) Describe the relationship between hours studied and exam score.
- (b) Identify any unusual feature in the data and explain its effect.
- (c) Would a linear model be appropriate? Justify your answer.
- (d) Explain whether it is reasonable to use this model to predict a student who studies 20 hours per week.
- (e) Does this data prove that studying more causes higher exam scores? Explain.

(a)

Strong positive linear association: as hours studied increase, exam score increases.

(b)

Student 12 is an outlier (much higher score than expected), which may inflate the correlation.

(c)

Yes, a linear model is reasonable, though slightly influenced by the outlier.

(d)

No. This is extrapolation (20 hours is outside the data range), so predictions are unreliable.

(e)

No. This is observational data, so we cannot conclude causation.

1. Correlation — 相关性
2. Positive Correlation — 正相关
3. Negative Correlation — 负相关
4. Scatterplot — 散点图
5. Linear Relationship — 线性关系
6. Strength — 强度
7. Direction — 方向
8. Outlier — 离群值
9. Influential Point — 影响点
10. Correlation Coefficient (r) — 相关系数 (r)

1. Correlation measures the **strength and direction** of a linear relationship between two quantitative variables.
2. Correlation is represented by the symbol r .
3. The value of r is always between -1 and $+1$.
4. A **positive** correlation means both variables tend to increase together.
5. A **negative** correlation means one variable increases as the other decreases.
6. The closer r is to ± 1 , the **stronger** the linear relationship.
7. The closer r is to 0 , the **weaker** the linear relationship.
8. Correlation only measures **linear** relationships.
9. Correlation is **not resistant** to outliers; unusual points can greatly affect r .
10. **Correlation does not imply causation**; a strong association does not prove that one variable causes changes in the other.

A researcher wants to investigate whether the number of hours students spend studying each week is related to their AP Statistics test score.

The researcher collects data from 12 students and creates a scatterplot. The scatterplot shows a positive, roughly linear pattern with one possible outlier. The correlation between study hours and test score is $r = 0.82$.

Questions

- (a) Interpret the value of the correlation coefficient $r = 0.82$ in the context of this study.
- (b) Explain whether the correlation coefficient alone proves that studying more causes higher AP Statistics test scores.
- (c) Describe how the possible outlier might affect the value of the correlation coefficient.
- (d) The researcher calculates a least-squares regression line for the data. Explain how the correlation coefficient supports the use of a linear regression model.
- (e) Identify one additional variable that could influence both study hours and AP Statistics test scores and explain why it may be a confounding variable.

- (a)** There is a strong positive linear association between study hours and AP Statistics test scores.
- (b)** No. Correlation does not imply causation. Other variables may affect the relationship.
- (c)** An outlier can strengthen or weaken the correlation, depending on its location.
- (d)** Since $r = 0.82$ is close to 1 and the pattern is roughly linear, a linear regression model is reasonable.
- (e)** Prior mathematics ability, tutoring, motivation, or attendance could affect both study hours and test scores, making it a confounding variable.

TRUE or FALSE

- A correlation of $r=0$ always means there is no relationship between two variables.

TRUE or FALSE

- Correlation measures the strength and direction of a linear relationship between two quantitative variables.

TRUE or FALSE

- Correlation is not affected by the units of measurement of the variables.