Introducing statistics – do these points align

Question 1

A scatterplot shows points that lie almost exactly on a straight line. What does this suggest?

A. There is no relationship between the variables.
B. There is a strong linear relationship between the variables.
C. The variables are categorical.
D. The graph must be incorrect.



Question 2

A scatterplot shows points scattered randomly with no obvious pattern. What is the best conclusion?

A. There is likely little or no association between the variables.
B. The variables have a perfect positive correlation.
C. One variable causes the other.
D. The data were collected incorrectly.



- **Scatterplot** — 散点图
- **Association / relationship** — 关联 / 关系
- **Positive association** — 正相关
- **Negative association** — 负相关
- **Outlier** — 离群点

1. Results vary naturally, so observed data may differ from expectations.
2. Statistics measures this variability and helps determine what is typical or unusual.
3. For categorical data, expected counts are calculated to model “no difference” or “no association.”
4. Chi-Square tests (Goodness-of-Fit, Homogeneity, Independence) compare observed and expected counts.
5. Use the test statistic and p-value to decide if results are consistent with chance or show a real pattern.

Question 1:

A researcher plots the number of hours students study per week (x) against their exam scores (y) and observes the points on a scatterplot.

a. Describe the **direction** of the association.

b. Describe the **strength** of the association.

c. Identify any **outliers** that might affect the relationship.

Given: Scatterplot of hours studied vs. exam scores

a. Direction of association:

- Most points **increase together → positive association (正相关)**

b. Strength of association:

- Points are **closely clustered around a line → strong association (强关联)**

c. Outliers:

- If any points are **far from the trend** (e.g., someone studies a lot but scores low) → **outlier (离群点)** that could **affect the relationship**

Stretch and Challenge

1. The Mysterious Outlier

A scatterplot shows a strong positive linear relationship between height and arm span. One point is far from the overall pattern. Should the point always be removed? Explain.

---

2. Firefighters and Damage

A study finds a correlation of $r = 0.92$ between the number of firefighters at a fire and the amount of property damage. Does this mean firefighters cause more damage? Explain.

3. Line or Curve?

Two students model the same dataset. One uses a straight line, and the other uses a curve. The data rise quickly and then level off. Which model is likely more appropriate? Why?

---

4. Beyond the Data

A regression line is created using data from children aged 5–12. A researcher uses the model to predict the reading ability of a 40-year-old. What concerns might you have about this prediction?

1. The Mysterious Outlier

Answer: No. Investigate whether it is an error or a valid observation first.

---

2. Firefighters and Damage

Answer: No. A lurking variable (fire size) may explain the relationship.

---

3. Line or Curve?

Answer: The curve, because the pattern is nonlinear.

---

4. Beyond the Data

Answer: It is an extrapolation beyond the observed data range.

Analyzing Relationships Between Quantitative Variables

Scatterplot	散点图
Quantitative Variable	定量变量
Association	关联 / 相关关系
Positive Association	正相关
Negative Association	负相关
Correlation	相关系数 / 相关性
Linear Relationship	线性关系
Outlier	离群值 / 异常值
Influential Point	影响点
Lurking Variable	潜在变量 / 隐藏变量

1. **Identify the Variables**
   Determine the two quantitative variables being compared.
2. **Describe the Scatterplot**
   Describe the **direction, form, strength, and unusual features**.
3. **Direction**
   A relationship can be **positive, negative**, or show little/no association.
4. **Form**
   Determine whether the relationship is **linear or nonlinear**.
5. **Strength**
   Strong relationships have points clustered closely around a clear pattern.
6. **Correlation r**
   Measures the **direction and strength of a linear relationship** between two quantitative variables.
7. **Correlation Does Not Imply Causation**
   An association between two variables does not prove that one causes the other.
8. **Outliers**
   An unusual observation can strongly affect correlation and regression results.
9. **Lurking Variables**
   A third variable may help explain an observed association between two variables.
10. **Extrapolation**
    Predictions outside the range of observed data may be unreliable.

Question 1: Describing a Relationship

A teacher records the **hours students spend studying** and their **scores on a statistics test**. A scatterplot shows a moderately strong, positive, approximately linear relationship.

a. Describe the direction of the association.
b. Describe the form of the association.
c. Describe the strength of the association.
d. Interpret the relationship in context.

Question 2: Correlation and Causation

A researcher finds a strong positive correlation between **ice cream sales** and **number of swimming accidents**.

a. Explain what the positive correlation means.
b. Does this provide evidence that eating ice cream causes swimming accidents? Explain.
c. Identify a possible lurking variable.

1. Describing a Relationship

a. Positive

b. Approximately linear

c. Moderately strong

d. More study time tends to be associated with higher test scores.

2. Correlation and Causation

a. Ice cream sales and swimming accidents increase together.

b. No; correlation does not establish causation.

c. Temperature / hot weather.

TRUE or FALSE

- **A scatterplot is used to display the relationship between two quantitative variables.**

TRUE or FALSE

- **A positive association means that as one variable increases, the other tends to decrease.**

TRUE or FALSE

- **Correlation measures the direction and strength of a linear relationship.**