

Algebra 1 - Scatter Plots, Lines of Fit and Two-Way Tables - Lesson

Common Core S-ID.B.5, S-ID.B.6, S-ID.C.7, S-ID.C.8, S-ID.C.9 · maths-pdf.fr/en/

A scatter plot shows the association between two numerical variables. A line of best fit models the trend, and its slope and intercept are interpreted in context. A residual is actual minus predicted. The correlation coefficient r measures the strength of a linear trend, but correlation does not prove causation. Two-way tables show relative frequencies.

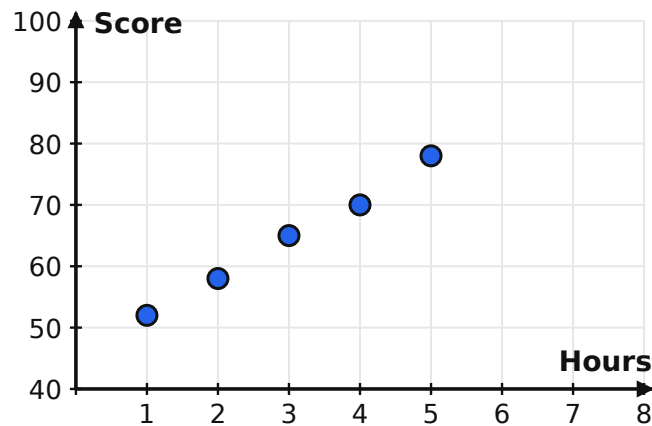
Algebra 1 · Common Core State Standards: **S-ID.B.5, S-ID.B.6, S-ID.C.7, S-ID.C.8, S-ID.C.9**

Key facts

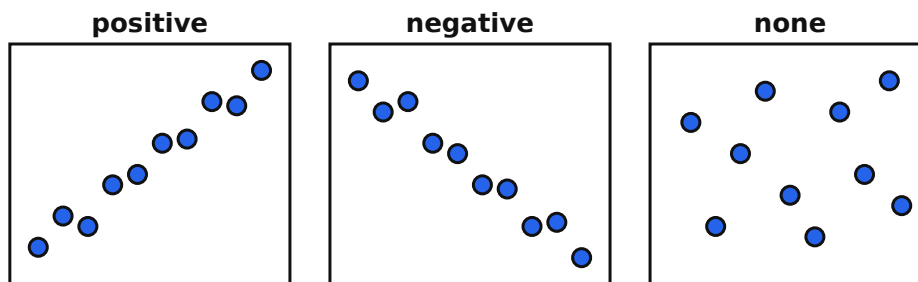
- A scatter plot can show positive, negative, or no association.
- In a line of best fit, the slope is the rate of change and the intercept is the predicted value at $x = 0$.
- Residual = actual value - predicted value.
- The correlation coefficient r is between -1 and 1; correlation does not imply causation.
- Two-way tables give joint, marginal, and conditional relative frequencies.

What does a scatter plot show?

A **scatter plot** graphs pairs of numbers as points. The pattern shows the **association** between the variables: positive (both increase), negative (one increases as the other decreases), or none.



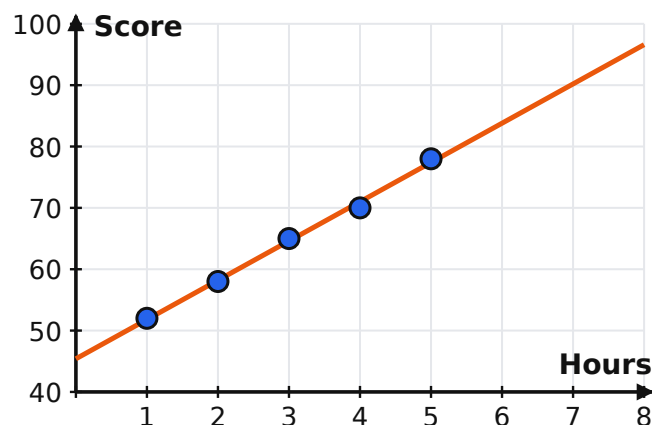
A scatter plot shows the relationship between two numerical variables. Here the points rise: a positive association.



The correlation coefficient r is near 1 for a strong positive trend, near -1 for a strong negative trend, and near 0 when there is no linear trend.

How do you use a line of fit?

A **line of best fit** is a line that follows the trend. Its equation $y = mx + b$ can predict values. For the study data, $y = 6.4x + 45.4$; each additional hour of studying is associated with about 6.4 more points, and 45.4 is the predicted score for 0 hours.



The line of best fit is $y = 6.4x + 45.4$. Each dashed segment is a residual: actual minus predicted.

1. Decide whether a linear model fits the pattern.
2. Write the equation of the line, using technology or two points on the line.
3. Interpret the slope as the rate of change and the intercept as the starting value, in context.
4. Predict only within the range of the data. Predicting far outside it is **extrapolation** and can be unreliable.

What is a residual?

A **residual** is the actual value minus the predicted value. If a student studied 4 hours and scored 70, the predicted score is $6.4(4) + 45.4 = 71$ and the residual is $70 - 71 = -1$. A residual plot with no pattern supports a linear model.

What does the correlation coefficient measure?

The **correlation coefficient** r is between -1 and 1 . Its sign gives the direction, and its size gives the strength of the linear relationship. Correlation does not prove **causation**: ice cream sales and sunburns both rise in summer because of hot weather, a lurking variable.

How do two-way tables work?

A two-way table shows counts for two categorical variables. The **joint relative frequency** divides a cell by the grand total, the **marginal** relative frequency divides a row or column total by the grand total, and the **conditional** relative frequency divides a cell by its row or column total.

	Walk	Bike	Total
Grade 7	30	20	50
Grade 8	10	40	50
Total	40	60	100

A two-way table counts how many people fall in each pair of categories.

In the table, the joint frequency for grade 8 and bike is $\frac{40}{100} = 40\%$. Among grade 8 students, the conditional frequency of biking is $\frac{40}{50} = 80\%$.

Questions about scatter plots, lines of fit and two-way tables

What is a line of best fit?

It is a straight line that follows the trend of a scatter plot closely, with points above and below it. It is used to describe and predict the relationship.

How do you interpret the slope of a line of best fit?

It is the predicted change in y for each 1-unit increase in x . For $y = 6.4x + 45.4$ with hours studied, each extra hour predicts 6.4 more points.

What does a residual tell you?

A residual is actual minus predicted. A positive residual means the point is above the line, and a negative residual means it is below.

Does a strong correlation prove causation?

No. Two variables can move together because of a lurking variable or by chance. Only a well-designed experiment can show cause and effect.

What is a conditional relative frequency?

It is a count divided by a row or column total, not the grand total. For example, 40 of the 50 grade 8 students bike, so the conditional frequency is 80%.