

Reading Two-Way Tables and Scatter Plots

Grade 9 – Algebra 1 math · Lesson · Standards: S-ID.B.5, S-ID.B.6, S-ID.C.7, S-ID.C.8, S-ID.C.9

Quick answer: A two-way table shows how two categories relate, using joint, marginal and conditional relative frequencies. For two numerical variables, a scatter plot and a line of best fit describe the trend. The slope gives the predicted change per unit, residuals measure errors, and the correlation r shows strength and direction, but correlation alone never proves causation.

Standards: S-ID.B.5, S-ID.B.6, S-ID.C.7, S-ID.C.8, S-ID.C.9 · **Level:** Grade 9 Algebra 1

How do two-way tables show relationships between categories?

A two-way table counts people in two categories at once. In a survey of 120 students, 38 ninth graders and 30 tenth graders buy lunch in the cafeteria, while 22 ninth graders and 30 tenth graders pack lunch.

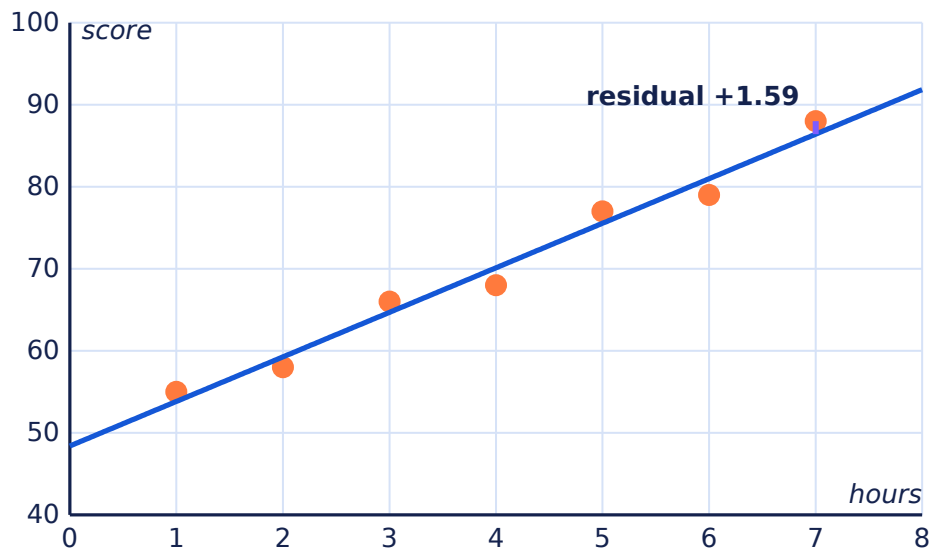
	Cafeteria	Packed	Total
9th grade	38	22	60
10th grade	30	30	60
Total	68	52	120

Three kinds of relative frequency

- **Joint:** a cell divided by the grand total. $38 \div 120 \approx 31.7\%$ are ninth graders who buy lunch.
- **Marginal:** a row or column total divided by the grand total. $68 \div 120 \approx 56.7\%$ buy lunch.
- **Conditional:** a cell divided by its row or column total. Of the ninth graders, $38 \div 60 \approx 63.3\%$ buy lunch.

How do you fit a line to a scatter plot?

A scatter plot pairs two numerical variables. If the points follow a straight trend, a line of best fit describes it. The line is written $\hat{y} = mx + b$, where $\hat{y}$ is the predicted value. For the study-hours data below, a calculator gives $\hat{y} = 5.43x + 48.4$.



Points near the line $\hat{y} = 5.43x + 48.4$. The vertical gap is a residual.

Method

1. Enter the data pairs in a calculator or spreadsheet.
2. Run linear regression to get the slope m , the intercept b and the correlation r .
3. Interpret m as the predicted change in y for each 1-unit increase in x .
4. Interpret b as the predicted y when $x = 0$, if that makes sense in context.

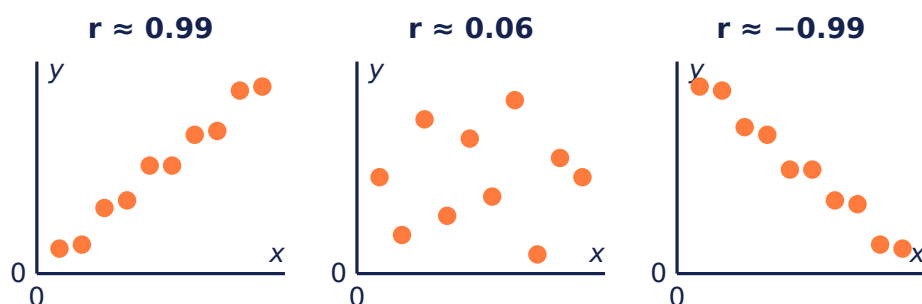
Worked example

The slope 5.43 means each extra hour of study is associated with about 5.4 more points. At 7 hours the line predicts $5.43(7) + 48.4 = 86.41$, but the actual score is 88, so the residual is $88 - 86.41 = 1.59$.

A residual is observed minus predicted. If the residuals look random around zero, a line is a good model. A curved pattern means a line is not appropriate.

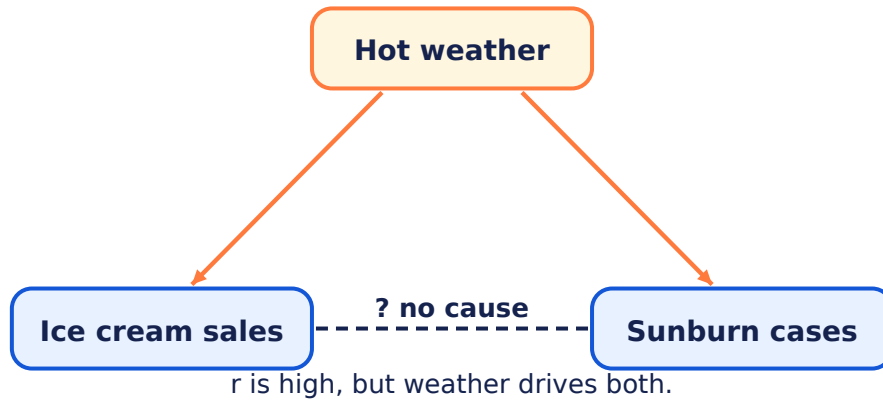
What does the correlation coefficient tell you?

The correlation coefficient r is a number from -1 to 1 . Its sign gives the direction, and its distance from 0 gives the strength of the straight-line relationship.



Does correlation mean causation?

No. Two variables can move together because a third one drives both. Ice cream sales and sunburns both rise in hot weather, yet neither causes the other. Only a well-designed experiment can show cause and effect.



A lurking variable can create correlation without causation.

Common mistake

Using a line far outside the data. Predicting beyond the data is extrapolation, and the trend may not continue.

Mathbot's tip

Say "associated with" instead of "causes" unless the data come from an experiment.

Key points to remember

- Joint, marginal and conditional relative frequencies come from a two-way table.
- The line of best fit has the form $\hat{y} = mx + b$; the slope is the predicted change in y per 1 unit of x .
- Residual = observed value – predicted value.
- The correlation r lies between -1 and 1 ; values near ± 1 show a strong linear relationship.
- Correlation does not prove causation, and extrapolating is risky.

Reading Two-Way Tables and Scatter Plots: frequently asked questions

What is a conditional relative frequency?

It is a cell count divided by a row total or a column total instead of the grand total. It answers questions such as "Of the ninth graders, what fraction buy lunch?"

What does the slope of a regression line mean?

The slope is the predicted change in y for each increase of 1 in x . In the study data, a slope of 5.43 means about 5.4 extra points for each extra hour studied.

What is a residual?

A residual is the actual y -value minus the value the line predicts. A positive residual means the point lies above the line, and a negative residual means it lies below.

What do values of r near 1, 0 and -1 mean?

A value near 1 means a strong positive linear relationship, near -1 a strong negative one, and near 0 little or no linear relationship.

Why does correlation not imply causation?

A lurking variable, such as hot weather, can affect both variables. The variables then change together even though neither one causes the other.