The correlation coefficient r

Two clouds of dots both climb. Which one keeps closer to a straight line, and can a single number say so?

WORDS TO KNOW

Correlation coefficient r: a number from −1 to 1 that measures how closely the points of a scatter plot cluster about a straight line.

Sign of r: the direction. Positive when the cloud climbs, negative when it falls.

Size of r: its distance from 0. It gives the strength.

Center of the cloud: the point (mean of the x-values, mean of the y-values).

KEY IDEAS

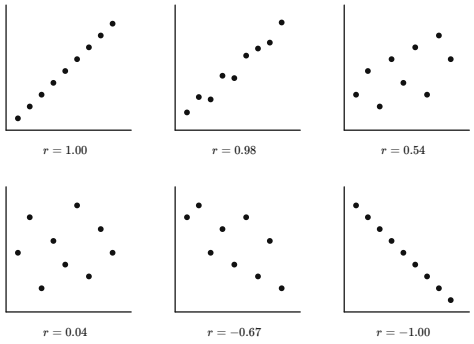
1. Sign for direction, size for strength

You find r with technology: a calculator, spreadsheet or app takes the pairs and reports it. Round it to two decimal places.

$$-1 \le r \le 1$$

$r > 0$: the cloud climbs. $r < 0$: it falls. The sign of r matches the sign of the slope of the line of best fit.

$r = 1$ or $r = -1$ only when every point lies on one straight line. r near 0 means no straight-line pattern.



2. Putting r into words

Words for r are a judgment, not a rule. One rough guide for the size of r:

Size of r	about 0.8 or more	about 0.5 to 0.8	about 0.3 to 0.5	below about 0.3
Strength	strong	moderate	weak	little or none

The order is not a judgment: a larger size always means a tighter straight-line pattern, whatever the sign.

3. Why the sign works

Draw a vertical and a horizontal line through the center of the cloud. Points in the top-right and bottom-left boxes push r up; points in the top-left and bottom-right boxes push it down. A climbing cloud sits mostly top-right and bottom-left, so r comes out positive.

WATCH OUT

- r **only sees straight lines.** A strong curve can have r near 0 (an arch) or near 1 (a gentle bend). Always look at the scatter plot too.
- r **says nothing about steepness.** A steep cloud and a nearly flat one can have the same r.
- r **does not show cause,** and it needs two numerical variables measured in pairs.

Free sample from math.wiki for classroom use. Do not post. math.wiki

WORKED EXAMPLES

Example 1. For twelve used cars, technology gives $r = -0.94$ for the pairs (age in years, price in dollars). Describe the relationship.

1. The sign is negative, so the cloud falls from left to right: older cars tend to cost less.
2. The size, 0.94, is close to 1, so the points cluster closely about a straight line.
3. It is not −1, so the points do not all lie exactly on one line: they scatter a little above and below it.

Answer: a strong, negative, linear relationship: older cars tend to be cheaper.

Example 2. Three data sets have $r = 0.62$, $r = -0.97$ and $r = 0.04$. Order them from the strongest straight-line relationship to the weakest.

1. Strength is the size of r, ignoring the sign: 0.62, 0.97 and 0.04.
2. Largest first: 0.97, then 0.62, then 0.04.

Answer: $r = -0.97$ (strong, falling), then $r = 0.62$ (moderate, climbing), then $r = 0.04$ (little or no straight-line pattern). The minus sign does not make −0.97 weaker.

Example 3. A culture of cells doubles every hour. The pairs (hours, thousands of cells) are $(0, 1)$, $(1, 2)$, $(2, 4)$, $(3, 8)$, $(4, 16)$, $(5, 32)$, $(6, 64)$. Technology gives $r = 0.88$. Does that show a straight-line relationship?

1. Look at the steps. Each hour the count grows by 1, 2, 4, 8, 16, 32: a bigger step every time.
2. Points on a straight line grow by the same step each time, so these points lie on a curve that gets steeper.
3. r is large only because a straight line passes fairly close to the points; it does not say a line is the right shape.

Answer: no. The hours fix the count exactly, but the pattern is a curve. A large r is not proof of a line: check the plot.

Free sample from math.wiki for classroom use. Do not post. math.wiki