Statistics Unit 2 worksheet: The correlation coefficient (answer key)

ANSWERS AT A GLANCE

1. (a) −0.96, 0.91, −0.83, −0.50, 0.45, 0.07 (b) the sets with $r = -0.83$, $r = -0.50$ and $r = -0.96$ (c) the set with $r = 0.07$
2. (a) negative (b) positive (c) negative (d) near 0 (e) no r: favorite color is a category, not a number (f) positive
3. (a) $r = -1$ (b) cannot be found (c) $r = 1$ (d) none: positive, $r = 0.96$ (e) cannot be found (f) none: negative, $r = -0.98$
4. (a) (9, 202) (b) (6, 140), (10, 230) and (12, 260) push up; (8, 210) pushes down; (9, 170) pushes neither way. (c) positive; $r = 0.89$, a strong positive linear relationship
5. (a) Station A: $r = 0.93$; Station B: $r = 0.09$ (b) Station A: strong, positive and linear; Station B: little or no straight-line pattern (c) No: $r = 0.09$ is so close to 0 that its sign gives no direction worth reporting.
6. (a) r is not a rate of change; it measures how closely the points cluster about a line, not how steeply they climb. (b) The count tends to rise year by year, and the points cluster tightly about a rising line: a strong positive linear relationship. (c) The slope of the line of best fit, found with technology from the same pairs.

WORKED SOLUTIONS

1. Six sets of paired data have these correlation coefficients:

$$r = 0.45, \quad r = -0.83, \quad r = 0.07, \quad r = -0.50, \quad r = 0.91, \quad r = -0.96$$

(a) List the six values in order, from the strongest straight-line relationship to the weakest.

(b) Which of the six scatter plots fall from left to right?

(c) Which set shows little or no straight-line pattern?

ANSWER

(a) −0.96, 0.91, −0.83, −0.50, 0.45, 0.07 (b) the sets with $r = -0.83$, $r = -0.50$ and $r = -0.96$ (c) the set with $r = 0.07$

SOLUTION

(a) Strength is the size of r, its distance from 0, whatever the sign. The sizes are 0.45, 0.83, 0.07, 0.50, 0.91 and 0.96. From largest to smallest: −0.96, 0.91, −0.83, −0.50, 0.45, 0.07.

(b) A cloud falls from left to right exactly when $r < 0$: the sets with $r = -0.83$, $r = -0.50$ and $r = -0.96$.

(c) The set with $r = 0.07$: its size is far below about 0.3, so a straight line describes its cloud hardly at all.

2. For each pair of variables, measured on many individuals, say whether you would expect r to be positive, negative or near 0, and why. If r cannot be found for the pair at all, say so and explain.

(a) The outdoor temperature on a winter day and the number of cups of hot chocolate a ski lodge cafe sells that day.

(b) The number of pages in a hardcover book and the mass of the book.

(c) The age of a used car and its selling price.

(d) A student's height and the last digit of the student's house number.

(e) A student's favorite color and the student's height.

(f) The length of a running race and the winning time.

ANSWER

(a) negative (b) positive (c) negative (d) near 0 (e) no r: favorite color is a category, not a number (f) positive

SOLUTION

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

(a) Negative: colder days tend to bring more hot chocolate sales, so as temperature goes up, sales tend to go down.

(b) Positive: books with more pages tend to be heavier.

(c) Negative: older cars tend to sell for less.

(d) Near 0: the last digit of a house number has nothing to do with height, so the cloud would show no direction.

(e) r cannot be found. It needs two numerical variables, and a favorite color is a category. A two-way table is the tool for such data.

(f) Positive: longer races take longer to win.

3. Without technology, decide for each data set whether $r = 1$, $r = -1$, r cannot be found, or none of these. When it is none of these, say whether r is positive or negative and then find r with technology.

(a) $(1, 7), (2, 5), (3, 3), (4, 1)$

(b) $(2, 4), (2, 6), (2, 9)$

(c) $(0, 3), (5, 13), (10, 23)$

(d) $(1, 2), (2, 4), (3, 8), (4, 16)$

(e) $(-2, 5), (0, 5), (3, 5)$

(f) $(1, 10), (2, 8), (3, 7), (4, 4)$

ANSWER

(a) $r = -1$ (b) cannot be found (c) $r = 1$ (d) none: positive, $r = 0.96$ (e) cannot be found (f) none: negative, $r = -0.98$

SOLUTION

(a) Each step of 1 in x lowers y by 2, so all four points lie on one falling line: $r = -1$.

(b) Every x-value is 2. The points sit on a vertical line, there is no spread in x, and r cannot be found.

(c) Each step of 5 in x raises y by 10, so the three points lie on one rising line: $r = 1$.

(d) Each step of 1 in x raises y by 2, then 4, then 8. The steps are not equal, so the points are not on one line and r is neither 1 nor −1. The cloud climbs, so r is positive. Technology gives $r = 0.96$: a curve that a line still passes close to.

(e) Every y-value is 5. The points sit on a horizontal line, there is no spread in y, and r cannot be found.

(f) Each step of 1 in x lowers y by 2, then 1, then 3. The steps are not equal, so r is not −1. The cloud falls, so r is negative. Technology gives $r = -0.98$.

4. A library recorded, on five weekdays, the hours it was open and the number of books borrowed (made-up data), as pairs (hours, books): (6, 140), (8, 210), (9, 170), (10, 230), (12, 260).

(a) Find the center of the cloud: the mean number of hours and the mean number of books.

(b) For each point, say whether it pushes r up, pushes it down, or pushes neither way, and why.

(c) From (b), predict the sign of r. Then find r with technology and describe the relationship.

ANSWER

(a) (9, 202) (b) (6, 140), (10, 230) and (12, 260) push up; (8, 210) pushes down; (9, 170) pushes neither way. (c) positive; $r = 0.89$, a strong positive linear relationship

SOLUTION

(a) Mean hours: $\frac{6 + 8 + 9 + 10 + 12}{5} = \frac{45}{5} = 9$. Mean books: $\frac{140 + 210 + 170 + 230 + 260}{5} = \frac{1010}{5} = 202$. The center is (9, 202).

(b) Draw a vertical line at 9 hours and a horizontal line at 202 books.

- (6, 140): both values below their means, bottom-left box, pushes r up.

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

- (8, 210): hours below the mean but books above, top-left box, pushes r down.
- (9, 170): on the vertical center line, pushes neither way.
- (10, 230) and (12, 260): both values above their means, top-right box, push r up.

(c) Three points push up and only one pushes down, and (8, 210) is close to both center lines, so its push is small. Predict $r > 0$. Technology gives $r = 0.89$: on days the library is open longer, more books tend to be borrowed, and the points cluster closely about a rising line. The relationship is strong, positive and linear.

5. Two weather stations each recorded, on six summer days, the hours of sunshine and the highest temperature in degrees Celsius (made-up data), as pairs (hours, temperature).

Station A: (2, 14), (3, 16), (5, 15), (7, 19), (9, 20), (10, 24)

Station B: (1, 19), (4, 23), (5, 17), (6, 22), (8, 18), (12, 21)

(a) Find r for each station with technology.

(b) At which station does sunshine have the stronger straight-line relationship with temperature? Describe the relationship at each station.

(c) The forecaster at Station B says, "Our r is positive too, so here as well sunnier days tend to be warmer." Is that a fair reading of Station B's r? Explain.

ANSWER

(a) Station A: $r = 0.93$; Station B: $r = 0.09$ (b) Station A: strong, positive and linear; Station B: little or no straight-line pattern (c) No: $r = 0.09$ is so close to 0 that its sign gives no direction worth reporting.

SOLUTION

(a) Entering each list into technology gives $r = 0.93$ for Station A and $r = 0.09$ for Station B.

(b) Station A, since 0.93 is much farther from 0 than 0.09. At Station A sunnier days tend to be warmer, and the points cluster closely about a rising line: a strong positive linear relationship. At Station B the size 0.09 is far below about 0.3, so there is little or no straight-line pattern between sunshine and temperature.

(c) No. A size as small as 0.09 means the points show no straight-line trend at all: at Station B the 1-hour day was warmer than the 5-hour day and the 8-hour day. The plus sign carries no direction worth reporting when r is this close to 0.

6. A town counted its public charging points for electric cars at the end of each year from 2016 to 2025. For the pairs (year, number of charging points), technology gives $r = 0.95$. A news report says: "The number of charging points rose by 0.95 each year."

(a) Explain what is wrong with the report.

(b) What does $r = 0.95$ tell you about these data?

(c) What number should the report have used to say how fast the count rose each year, and how would you find it?

ANSWER

(a) r is not a rate of change; it measures how closely the points cluster about a line, not how steeply they climb. (b) The count tends to rise year by year, and the points cluster tightly about a rising line: a strong positive linear relationship. (c) The slope of the line of best fit, found with technology from the same pairs.

SOLUTION

(a) The report reads r as an amount added each year. But r is a position on the scale from -1 to 1 that measures how closely the points cluster about a straight line. It says nothing about how steep the cloud is,

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

so it cannot tell how many charging points were added a year. (A count of charging points cannot even rise by 0.95.)

(b) The sign is positive, so the count tends to grow from year to year. The size 0.95 is close to 1, so the ten points lie close to one rising line: a strong positive linear relationship.

(c) The slope of the line of best fit, in charging points per year. Technology fits the line to the ten (year, count) pairs and reports its slope, often next to r.

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