What is Correlation?

Worksheet

Correlation is a coefficient, r = Cov(x,y) / (xy), that measures the direction (positive or negative) and strength (from -1 to 1) of the linear relationship between two variables.

$$r = \frac{\sum(x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum(x_i - \bar{x}}$$

Questions

1. What does a correlation coefficient of r = -1 mean?

A) No relationship
B) Perfect positive relationship
C) Perfect negative relationship
D) Weak relationship

2. The range of the correlation coefficient r is:

A) 0 to 1
B) -1 to 1
C) -100 to 100
D) 0 to 100

3. Ice cream sales and drowning deaths show a high correlation. This means:

A) Ice cream causes drowning
B) A hidden variable (summer heat) likely drives both
C) Drowning causes ice cream sales
D) The data must be wrong

4. If r = 0.9, the relationship is:

A) Weak and negative
B) Strong and positive
C) No relationship
D) Perfectly negative

5. Find the correlation coefficient for x = (1, 2, 3) and y = (2, 4, 6).

6. Find the correlation coefficient for x = (1, 2, 3) and y = (6, 4, 2).

7. Find the correlation coefficient for x = (1, 2, 3, 4) and y = (2, 3, 5, 4).

8. Define: What is correlation?

9. Define: What does r = 0 mean?

10. Define: What's the difference between correlation and causation?

Answer Key

1. C) Perfect negative relationship - r = -1 means a perfect negative relationship - as x rises, y falls in perfect lockstep.

2. B) -1 to 1 - r is always between -1 and +1.

3. B) A hidden variable (summer heat) likely drives both - This is a classic confounding-variable example - there's no causal link.

4. B) Strong and positive - 0.9 is close to 1, indicating a strong positive relationship.

5. x = 2, = 4 Deviations x: -1, 0, 1 - y: -2, 0, 2 (x-x)(y-) = 2 + 0 + 2 = 4 (x-x) = 2, (y-) = 8 r = 4 / (28) = 4/4 = 1

6. x = 2, = 4 Deviations x: -1, 0, 1 - y: 2, 0, -2 (x-x)(y-) = -2 + 0 - 2 = -4 (x-x) = 2, (y-) = 8 r = -4 / (28) = -4/4 = -1

7. x = 2.5, = 3.5 Deviations x: -1.5, -0.5, 0.5, 1.5 - y: -1.5, -0.5, 1.5, 0.5 (x-x)(y-) = 2.25+0.25+0.75+0.75 = 4 (x-x) = 5, (y-) = 5 r = 4 / (55) = 4/5 = 0.8

8. A statistic (r) measuring the strength and direction of a linear relationship between two variables, from -1 to 1.

9. No linear relationship between the variables (though a nonlinear one may still exist).

10. Correlation shows two variables move together; it does not prove one causes the other.