Patterns in the Past

Reading correlation coefficients in historical research

P.1.6 "Define correlation coefficients and explain their appropriate interpretation."

NAME _____ DATE _____

1 READ AND MARK

Read it twice

YOUR JOB Underline the sentence that defines r, circle the signs on the four coefficients in the table, and number the coefficients from 1 (strongest linear association) to 4 (weakest) to track their meaning and strength.

What County Records Can Show

1. Imagine a historian comparing records from twenty fictional counties in 1932. Each county has two measurements, such as the number of doctors and the number of residents. A correlation coefficient, written r, summarizes how closely two numerical variables follow a straight-line pattern across counties. Its value runs from -1 to +1; a value near 0 shows little linear association.

2. The sign shows direction: positive means the variables tend to rise together; negative means one tends to fall as the other rises. The farther r is from zero, the stronger the linear association. Compare absolute values to judge strength. Zero does not rule out a curved relationship. The invented county records below are for practice, not historical evidence.

HYPOTHETICAL CORRELATIONS ACROSS TWENTY COUNTIES IN 1932

Pair of county measurements	r
Library branches and population	+0.82
Average distance to a rail station and weekly market trips	-0.73
Newspaper circulation and voter turnout	+0.06
Unemployed workers and relief spending	+0.57

3. Correlations describe patterns, not causes. A county with more doctors may also have more residents because larger populations need more services. The coefficient alone cannot show which variable changes the other. Historians must check how records were collected and consider other possible explanations. A high coefficient may suggest a question, but it cannot prove a cause.

2 ANSWER WITH EVIDENCE

Use the passage

1. What does +0.82 suggest about library branches and population across these counties?

2. What does -0.73 suggest about distance to a rail station and weekly market trips?

3. What does +0.06 tell you about newspaper circulation and voter turnout, and what does it not rule out?

4. Which pair has a stronger linear association: rail distance and market trips, or unemployed workers and relief spending? Explain your choice and whether either coefficient proves a cause.

3 YOUR TURN

Your own words

Invent a different pair of numerical county measurements a historian might compare in 1932 and give it a hypothetical r. In three sentences, explain its direction and strength, then name another factor that might affect both measurements.

Patterns in the Past

PART 1 · READ AND MARK

Underline the sentence that defines r, circle the signs on the four coefficients in the table, and number the coefficients from 1 (strongest linear association) to 4 (weakest) to track their meaning and strength. Accept any marking that follows this job and points at the right text.

PART 2 · ANSWERS

#	QUESTION	ANSWER
1	What does +0.82 suggest about library branches and population across these counties?	**Counties with larger populations tend to have more library branches. The positive linear association is strong.**
2	What does -0.73 suggest about distance to a rail station and weekly market trips?	**Counties farther from a rail station tend to have fewer weekly market trips. The negative linear association is fairly strong.**
3	What does +0.06 tell you about newspaper circulation and voter turnout, and what does it not rule out?	**It shows little linear association. It does not rule out a curved relationship.**
4	Which pair has a stronger linear association: rail distance and market trips, or unemployed workers and relief spending? Explain your choice and whether either coefficient proves a cause.	**Rail distance and market trips have the stronger linear association because 0.73 is farther from zero than 0.57. Neither coefficient proves causation.**

PART 3 · YOUR TURN (SAMPLE ANSWER)

Across fictional 1932 counties, the number of factories and the number of grocery stores have a hypothetical r of +0.76. This suggests a fairly strong positive linear association: counties with more factories tend to have more grocery stores. Population size could affect both, so the coefficient does not prove that factories cause stores to open.

Accept reasonable descriptions of strength that match how far r is from zero. For Part 3, the invented coefficient must be between -1 and +1, and its interpretation must match its sign and magnitude.