

Correlation & Causation

Computing

What a correlation coefficient tells you, and the four things it might mean.

r runs from -1 to $+1$. Zero means no LINEAR relationship — not no relationship.

Reading r

± 0.9 to 1.0 – very strong

± 0.7 to 0.9 – strong

± 0.4 to 0.7 – moderate

± 0.1 to 0.4 – weak

0 – no linear relationship

If A correlates with B

A causes B

B causes A

C causes both — confounding

Coincidence, especially in small samples

Watch for

Selection bias in who was measured

Simpson's paradox — reversal within subgroups

A curved relationship that *r* reports as zero

Multiple comparisons finding something by chance

Anscombe's quartet: four datasets with identical means, variances and *r*, and completely different shapes. Plot the data.