🧀 01 The Units Trap: covariance changes, correlation doesn't

A researcher measures temperature T (°C) and ice-cream sales S. They convert T to Fahrenheit:

$$F = 1.8T + 32.$$

- a. Using covariance properties, express $\mathrm{Cov}(F, S)$ in terms of $\mathrm{Cov}(T, S)$.
- b. Prove that the correlation is unchanged: $\rho_{F,S} = \rho_{T,S}$, using the definition of correlation.
- c. Explain (2–4 sentences) why correlation is "unit-free," while covariance is not.

$\mathrm{Cov}(X,Y)$

$\mathbb{E}[(X-\mathbb{E}[X])(Y-\mathbb{E}[Y])]$

$1.8T+3.2$

$\sigma_F = \sqrt{\mathrm{Var}(F)} = \sqrt{\mathrm{Var}(1.8T+32)} = \sqrt{(1.8)^2 \mathrm{Var}\, T}$

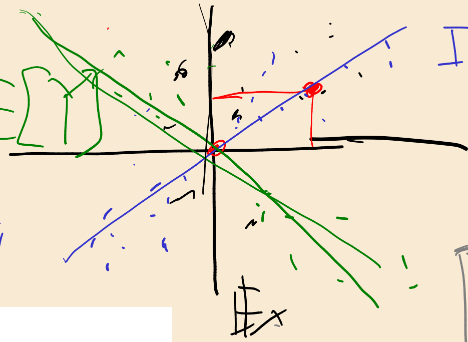
$\mathrm{Cov}(F,S) = \mathrm{Cov}(1.8T+32, S) = \mathbb{E}[XY] - \mathbb{E}[X]\mathbb{E}[Y]$

$= 1.8\,\mathrm{Cov}(T,S) + \mathrm{Cov}(32,S)$

$\rho(F,S) = \dfrac{1.8\,\mathrm{Cov}(T,S)}{1.8\,\sigma_T\sigma_S} = \rho(T,S)$

$X - \mathbb{E}[X]$

$32 - 32 \to 0$



$E[Y]$

$1 \to$ ... $-1 \to$ ... 0

🧀 02 Correlation $= \pm 1$ as a detective test (constructive, not computational) 🔗

$(f(X)-y)^2$ MSE 0.98

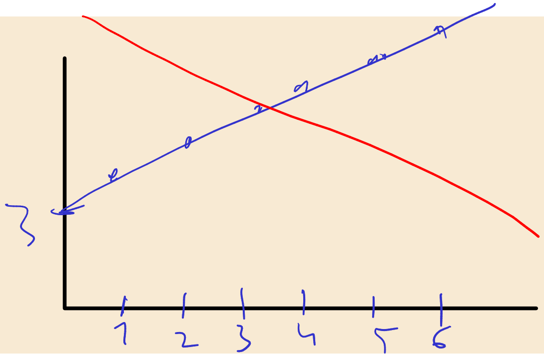
Cor ≠ causal

$\rho_{X,Y} = \dfrac{\mathrm{Cov}(X,Y)}{\sigma_X \sigma_Y} \to [-1,1]$

$\sigma_x = \sqrt{\mathrm{Var}(x)}$

You are given $x = [1, 2, 3, 4, 5, 6]$.

- a. Construct integer-valued y such that the correlation is exactly $+1$.
- b. Construct integer-valued y such that the correlation is exactly -1.
- c. Justify both by referencing the condition for extremal correlation (the $Y = aX + b$ characterization).



$y = 2x + 3$

$\rho(x,x) = 1$

$\iff$

$y = -|k|x + b$

E

03 Dependent but uncorrelated (the “nonlinear dependence” trap)

Let X be uniform on $\{-2,-1,1,2\}$, and define $Y = X^2$.

- a. Compute $\mathbb{E}[X]$, $\mathbb{E}[Y]$, and $\mathbb{E}[XY]$.
- b. Use $\mathrm{Cov}(X,Y) = \mathbb{E}[XY] - \mathbb{E}[X]\,\mathbb{E}[Y]$ to show $\mathrm{Cov}(X,Y) = 0$.
- c. Explain why this does not mean independence (connect explicitly to the “correlation only measures linear dependence” warning).

$c_1X_1 + c_2X_2$

$(0,0)$

$\{-2,-1,1,2\}$ $\quad Y = x^2$ $\quad \mathrm{Cov}(X,Y) = \mathbb{E}[XY] - \mathbb{E}[X]\,\mathbb{E}[Y] = 0$

$\mathbb{E}[X] = \frac{1}{4}(-2-1+1+2) = 0$

kendall

$X \quad Y$

$4 \quad 1 \quad 1 \quad 4$

X^2, Y

$\mathbb{E}[Y] = \frac{4+1+1+4}{4} = 2.5$

$\mathbb{E}[XY] = \frac{1}{4}(-8-1+1+8) = 0$

04 Pearson vs Spearman: monotonic but not linear

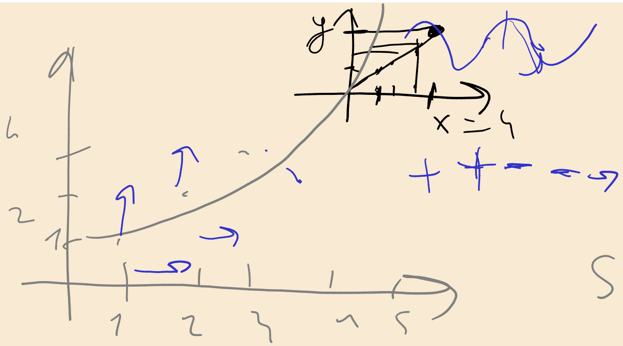
Consider $x = [1,2,3,4,5,6]$ and $y = [1,2,4,8,16,\cancel{32}]$. 50

- a. Argue (without calculating Pearson exactly) why Pearson correlation is not 1 (use the “linear relationship” criterion).
- b. Compute Spearman rank correlation exactly (no ties here).
- c. One sentence: why Spearman is the right tool here.

$(X - \mathbb{E}[X])(Y - \mathbb{E}[Y])$

$P(x,y) = P(x)P(y)$

$P(x) = P(x \mid Y)$



$Y = 2^x \qquad Y = kx + b$

$S = \rho(\mathrm{rank}(X), \mathrm{rank}(Y)) = 1$

32

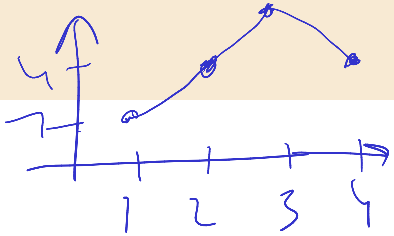
🧀 🧀 05 One outlier can flip the story

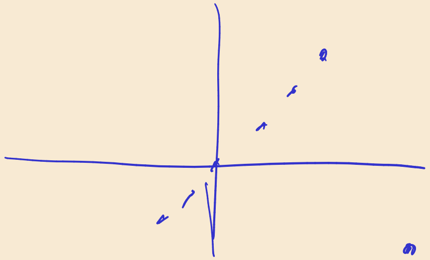
Common points:

$$(1,1), (2,2), (3,3), (4,4), (5,5).$$

Dataset A adds $(6,6)$. Dataset B adds $(6,-20)$.

- a. For each dataset, decide whether the sample correlation is positive or negative (justify using the sign of "products of deviations," not full computation).
- b. Explain how one point can dominate this "one-number summary."





$\mathbb{E}[XY] = \frac{1}{4}(1+4+12+12) =$

$\mathbb{E}[X] = \frac{1}{4}\left(\frac{4\cdot 5}{2}\right) = \frac{10}{4}$

$\frac{1}{4}\left(1+4+12+12 - \frac{10\cdot 10}{4}\right) =$

$1 - \frac{\cancel{6}\cdot 2}{\cancel{60}_{10}} = 0.8$

🧀 🧀 06 Diversification as a decision: when does combining reduce risk? 🔗

Two daily returns R_1, R_2 satisfy:

$$\mathrm{Var}(R_1) = 4, \quad \mathrm{Var}(R_2) = 9, \quad \mathrm{Cov}(R_1, R_2) = c.$$

Let $P = R_1 + R_2$.

$\mathrm{Var}(R_1+R_2) = \mathrm{Var}(R_1) + \mathrm{Var}(R_2) + 2\,\mathrm{cov}$

- a. Express $\mathrm{Var}(P)$ as a function of c.
- b. Compare $c = +5,\ 0,\ -5$. Which case yields the smallest portfolio variance, and why?
- c. Translate the result into plain English (what does negative covariance "do" for you?).

$\mathrm{Var}(R_1) = 4 \qquad \mathrm{Var}(R_2) = 9$

$\mathrm{Cov}(R_1, R_2) = c$

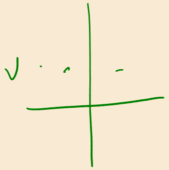
$4 + 9 + 2c$

$2 \quad 3$

$\boxed{13}$

3

cov

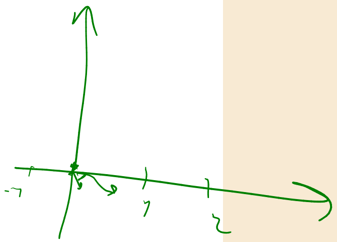


🧀 🧀 07 Correlation as geometry: angle between centered vectors

You are given mean-centered vectors:

$$u = (1, 0, -1), \qquad v = (2, -1, -1).$$

- a. Compute $\cos(\theta) = \dfrac{u \cdot v}{\|u\|\,\|v\|}$.
- b. Interpret the sign and magnitude of $\cos(\theta)$ as a correlation analogue.
- c. Decide whether the variables are "roughly aligned," "roughly orthogonal," or "roughly opposite."



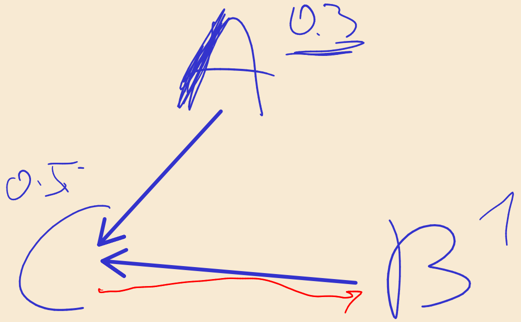
E

Var

$u \cdot v = 2 + 0 + 1 = 3$

$\|u\| = \sqrt{2} \qquad \|v\| = \sqrt{6}$

$\cos(\theta) = \dfrac{3}{\sqrt{2}\sqrt{6}} = \dfrac{3}{2\sqrt{3}} = \dfrac{\sqrt{3}}{2} = 0.85$



A, B, C, A