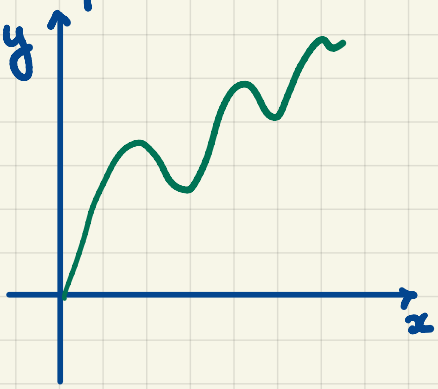
M358K: December 1st, 2023.

Correlation. Linear Transform.

x ... independent variable (say, time) : on the horizontal axis
y ... dependent variable (say, position): on the vertical axis



$$x \xrightarrow{f} y$$
$$y = f(x)$$

Recall: Covariance.

Let X and Y be a pair of numerical random variables.

Set:
- μ_X, μ_Y ... the means (both finite)
- $\mathrm{Var}[X]$, $\mathrm{Var}[Y]$ .... the variances (both finite)
- σ_X, σ_Y ... the standard deviations

Def'n. The covariance between X and Y is:

$$\mathrm{Cov}[X,Y] := \mathbb{E}[(X-\mu_X)(Y-\mu_Y)]$$
$$= \mathbb{E}[X \cdot Y] - \mu_X \cdot \mu_Y$$

Q: $\mathrm{Cov}[X,X] = \mathbb{E}[(X-\mu_X)(X-\mu_X)] = \mathbb{E}[(X-\mu_X)^2] = \mathrm{Var}[X]$

Q: If above-average values of X are associated w/ above-average values of Y, then $\mathrm{Cov}[X,Y] > 0$

Q: If above-average values of X are associated w/ below-average values of Y, then $\mathrm{Cov}[X,Y] < 0$

Q: If X and Y are independent, then $\mathrm{Cov}[X,Y] = 0$

Note: Let α and β be two constants.

$$\mathrm{Var}[\alpha X + \beta Y] = \alpha^2 \cdot \mathrm{Var}[X] + 2\alpha\beta\,\mathrm{Cov}[X,Y] + \beta^2 \cdot \mathrm{Var}[Y]$$

Correlation (coefficient).

Defn.

$$\rho_{X,Y} = \text{corr}[X,Y] = \frac{\text{Cov}[X,Y]}{\sigma_X \cdot \sigma_Y}$$

Q: In what units is the correlation?

Unitless.

Q: What values can the correlation take?

$$-1 \leq \rho_{X,Y} \leq 1$$

Q: What if $\rho_{X,Y} = 1$?

$$\text{Var}\left[Y - \frac{\sigma_Y}{\sigma_X} X\right] =$$

$$= \text{Var}[Y] - 2 \cdot \frac{\sigma_Y}{\sigma_X} \cdot \text{Cov}[Y,X] + \frac{\sigma_Y^2}{\sigma_X^2} \cdot \text{Var}[X]$$

$$= \sigma_Y^2 - 2 \cdot \frac{\sigma_Y}{\sigma_X} \cdot \rho_{X,Y} \cdot \sigma_X \sigma_Y + \frac{\sigma_Y^2}{\sigma_X^2} \cdot \sigma_X^2 = 0$$

($\rho_{X,Y} = 1$)

$$\Rightarrow \text{Var}\left[Y - \frac{\sigma_Y}{\sigma_X} \cdot X\right] = 0$$

$$\Rightarrow Y - \frac{\sigma_Y}{\sigma_X} \cdot X \text{ is constant} \quad \text{(say } b\text{)}$$

$$\Rightarrow Y = b + \underbrace{\frac{\sigma_Y}{\sigma_X}}_{=a} X = b + a \cdot X$$

$\Rightarrow$ Y is a linear transform of X.

Q: What if $\rho_{X,Y} = -1$? Think @ home.

Def'n. Let $(x_1, y_1), (x_2, y_2), \ldots, (x_n, y_n)$ be the observations. We define the sample correlation as

$$r = \frac{1}{n-1} \sum_{i=1}^{n} \left(\frac{x_i - \bar{x}}{s_x} \right) \left(\frac{y_i - \bar{y}}{s_y} \right)$$

UNIVERSITY OF TEXAS AT AUSTIN

Problem set # 21

Correlation. Linear transforms.

21.1. **Correlation.** Provide your final answer only to the following questions.

Problem 21.1. The following statement makes sense:

"The correlation between nationality and personal wealth is −0.89."

True or false?

Problem 21.2. The correlation coefficient equals the proportion of times that two variables lie on a straight line. *True or false?*

Problem 21.3. If r is the sample correlation between x and y, then $2r$ is the correlation between $2x$ and y. *True or false?*

Problem 21.4. A college newspaper interviews a psychologist about a proposed system for rating the teaching ability of faculty members. The psychologist says, "The evidence indicates that the correlation between a faculty member's research productivity and teaching rating is close to zero." The correct interpretation of this statement is:

Good researchers are just as likely to be good teachers as they are bad teachers. Likewise for poor researchers.

True or false?

Provide your complete solutions for the following problems.

Problem 21.5. Which of the following is a valid statement?

a. The correlation between students' majors and their GPA is 0.67.
b. The correlation between students' IQ scores and their foot size is 0.67.
c. The correlation between students' incomes and their GPA is 1.2.

Explain!

Problem 21.6. *What is wrong about the following report?*

An experiment is conducted to study the bonding strength of adhesives that contain varying amounts of a particular chemical additive. Wafers of a specified material are glued together using the adhesive with each amount of additive, allowed to set for 24 hours, and then the strength needed to separate the wafers is determined. It is reported that the correlation between strength required and amount of additive was 0.86 pounds-force per square inch.

21.2. **Linear transforms.** Provide your complete solutions for the following problems.

Problem 21.7. Raw scores on a certain exam range from 400 to 1600. Find a **linear transformation** that will transform the raw-score interval to the modified range from 100 to 1000 (i.e., the score of 400 will get transformed to 100, the score of 1600 to 1000, and all the scores in between will have a linear correspondence).

Problem 21.8. Suppose we apply a linear transformation $y_{new} = 12 - 34y_{old}$ to a dataset. After a while we choose to apply another linear transformation $y_{newest} = -3 + 4y_{new}$. Is y_{newest} a linear transformation of y_{old}? If so, what are its coefficients a and b. If not, why not?
