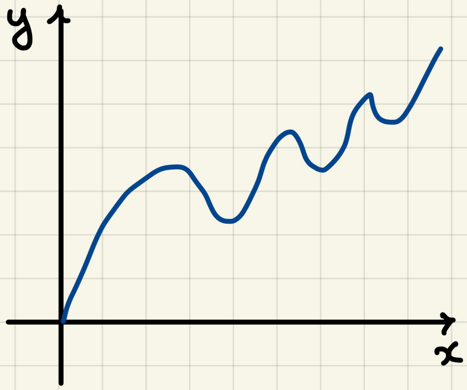
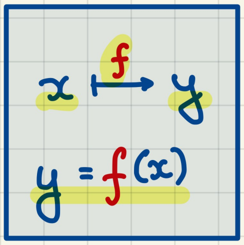
M358K: December 3rd, 2021.

Correlation. Linear Transform.

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





Recall: Covariance.

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

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

Def'n. The covariance between X and Y is

$$\boxed{\begin{aligned} Cov[X,Y] &:= \mathbb{E}\left[(X-\mu_X)(Y-\mu_Y)\right] \\ &= \mathbb{E}[X\cdot Y] - \mu_X \cdot \mu_Y \end{aligned}}$$

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

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

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

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

Note: Let α and β be two constants.

$$\boxed{\begin{aligned}\text{Var}[\alpha\cdot X+\beta\cdot Y] &= \text{(use the def'n of the variance; use linearity of expectation; tidy up)}\\ &= \alpha^2\cdot\text{Var}[X]+2\alpha\beta\,\text{Cov}[X,Y]+\beta^2\cdot\text{Var}[Y]\end{aligned}}$$

✓

Correlation (coefficient).

Def'n.

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

Q: In what units is the correlation?

$\boxed{\text{Unitless.}}$

Q: What values can the correlation take?

$$\boxed{-1\le\rho_{X,Y}\le 1}$$

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

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

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

$$=\sigma_Y^2-2\cdot\frac{\sigma_Y}{\cancel{\sigma_X}}\underbrace{\rho_{X,Y}}_{=1}\cancel{\sigma_X}\cdot\sigma_Y+\frac{\sigma_Y^2}{\cancel{\sigma_X^2}}\cdot\cancel{\sigma_X^2}=0$$

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

$\Rightarrow\ \boxed{Y-\frac{\sigma_Y}{\sigma_X}\cdot X\ \text{is constant}}$

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

$\Rightarrow$ 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

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