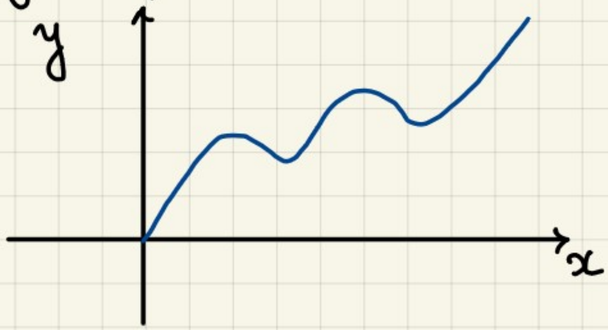
M358K: December 2nd, 2020.

Correlation. Linear Transform.

Recall: Variables in Calculus.

x... independent variable (say, time): on the horizontal axis

y... dependent variable (say, position): on the vertical axis



$x \overset{f}{\longmapsto} y$

$y = f(x)$

Recall: Covariance

Say that X and Y are two numerical random variables.

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 defined as

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

Q: If X and Y are independent, $Cov[X,Y] = ?$

$\rightarrow$: $Cov[X,Y] = \mathbb{E}[XY] - \mu_X \cdot \mu_Y$

$$\underset{\uparrow \text{ independence}}{=} \mathbb{E}[X] \cdot \mathbb{E}[Y] - \mu_X \cdot \mu_Y = 0 \checkmark$$

Q: $\mathrm{Cov}[X, X] = \cancel{?}\ \mathrm{Var}[X]$.

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

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

α and β are two real constants

$\mathrm{Var}[\alpha \cdot X + \beta \cdot Y]$ = (write the variance down by def'n; use the linearity of expectation; tidy up)

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

<u>Correlation (coefficient).</u>

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

Q: In which units is the correlation?

→: Unitless!

Q: What values can the correlation take?

→: $\boxed{-1 \leq \rho_{X,Y} \leq 1}$

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

→: $\mathrm{Var}\left[\boxed{Y - \frac{\sigma_Y}{\sigma_X} \cdot X}\right] =$

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

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

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

$\Rightarrow Y - \frac{\sigma_Y}{\sigma_X} \cdot X = \text{constant} = b$

$\Rightarrow Y = \underbrace{\left(\frac{\sigma_Y}{\sigma_X}\right)}_{=a} X + b$

$\Rightarrow$ <u>Y is a linear transform of X.</u>

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

Def'n. For observations $(x_1, y_1), (x_2, y_2), \dots, (x_n, y_n)$, we define the sample correlation as

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