What's the most important thing about a class?

	Travis	strictness	challenging work	
students	150	50	50	→ 250
Teacher	50	25	125	→ 200
Admin	10	25	15	→ 50

Q1 what % of survey were students → $\frac{250}{500}$ = 50%

Q2 what group is most likely to pick strictness

strict students: $\frac{50}{250}$

teachers: $\frac{25}{200}$

admins: $\frac{25}{50}$

Q3 what % of students picked "Travis?"

$\frac{150}{250}$

X, Y are data sets

are X & Y related to each other?

① correlation: correlation ≠ causation

← association "symmetric relationship"

② regression ← make predictions

← use to prove causality

① response variable ⟷ Y

② explanatory variable ⟺ X

← (one or more)

---

covariance:

$X = \{x_1, \ldots, x_n\}$

$Y = \{y_1, \ldots, y_n\}$

equal # of data pts

n

$$\operatorname{cov}(X, Y) = \frac{1}{n} \cdot \sum_{i=1} (x_i - \mu_X) \cdot (y_i - \mu_Y)$$

* changing order
order of X or Y gives different results!

$X = \{1, 2, 3\} \rightarrow \mu_X = 2$

$Y = \{7, 6, 2\} \quad \mu_Y = 5$

$$\operatorname{cov}(X, Y) = \frac{1}{3} \left(\begin{array}{c} (1-2) \cdot (7-5) \\ + \\ (2-2)(6-5) \\ + \\ (3-2)(2-5) \end{array} \right)$$

$$= \frac{1}{3} \{ 2 + 0 + -3 \}$$

$= 3$

Rmk: $\text{cov}(X, X) = \text{Var}(X) = \sigma_X^2$.

$\text{cov}(X, Y) > 0 \rightsquigarrow (x_i - \mu_x)(y_i - \mu_y)$ is more $+$ than $-$

$$\text{cov}(X,Y) = \frac{\sum \boxed{(x_i - \mu_x)(y_i - \mu_y)}}{n}$$

$\rightsquigarrow$ x_i and y_i will be both above means or both below means (more often than not) gives us a measure how change in X is associated to changes in Y.

linear correlation "r"

when 2 variables are linearly associated.

$$\boxed{r = \frac{cov(X,Y)}{\sigma_X \sigma_Y}}$$

always in b/t $[-1,1]$

$\{ r \to -1$ negative

linear correlation

$r \to +1$ positive linear correlation

$r \to 0$ non linear correlation.

X, Y are ~~independent~~

if $cov(X,Y) = 0$

independence $\Longrightarrow r = 0$

but $r = 0 \not\Longrightarrow$ indep!