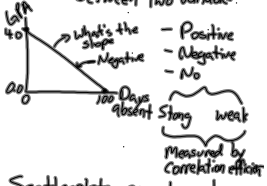
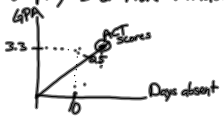
Correlation

- Measures the relationship between two or more variables
  - Scatterplot visualizes the relationship/correlation between two variables.

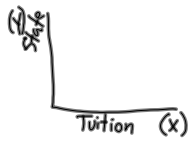


Scatterplots cannot easily display 3 or more variables.



- Correlation (aka. r, Pearson's correlation coefficient)
  - Measures the direction of relationship
  - Measures the strength
- Slope
  - Measures direction of relationship.
  - Measures the trade-off.

Ex: Tuition & State Rev.



Calculate the slope of the trade-off between tuition & state revenues between 1980 & 2010.

$$\text{Slope} = \frac{\text{Change in Y values}}{\text{Change in X values}}$$

→ Chg in State / Chg in Tuition

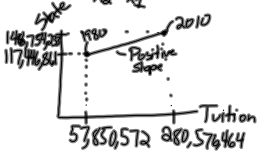
$$\text{Change in } Y = \Delta Y = Y_2 - Y_1$$

"to" 2010 — "from" 1980

$$\text{Change in } X = \Delta X = X_2 - X_1$$

tuition in 2010 — tuition in 1980

$$\text{Slope} = \frac{Y_2 - Y_1}{X_2 - X_1}$$



$$\text{Slope} = \frac{148,754,233 - 117,446,861}{280,576,464 - 57,850,572}$$

$$= \frac{31,307,372}{222,725,892} = \boxed{0.141}$$

→ positive

For every $1 change in tuition revenues, state revenues also change by $0.141.

→ Tuition increases $1, then state revenues increase 14¢

→ Tuition increases $10, then state revenues increase $1.41 (10 × 0.141)

→ State revenues increase $1, then tuition revenues increase $7.09

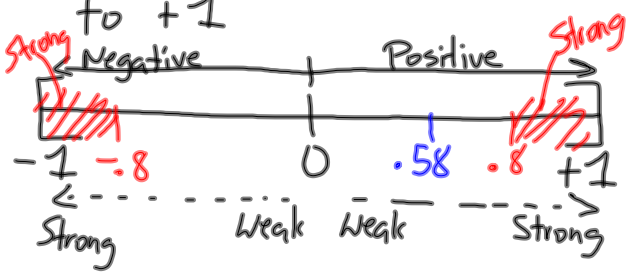
(X) → (Y) Tuition to state = slope

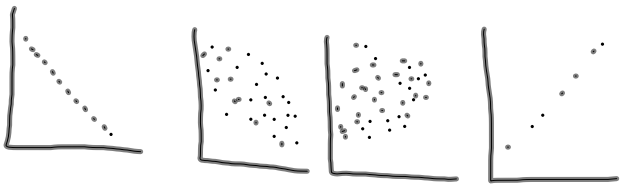
State to tuition = 1/slope → 1/0.141 = 7.09

(Y) → (X)

Pearson Correlation Coefficient (a.k.a. r, R)

R can be between and include anywhere from −1 to +1





Ex: Excel use =PEARSON(·)

= .58

↳ Positive relationship

↳ Intepretation depends on your field.

↳ An R greater than 0.8 or less than −0.8 is "strong."
↳ unless you anticipate stronger.

↳ A correlation that is greater than 0.1 or less than −0.1 is "sufficient."