

2/20/2019

## Univariate analysis

one-variable

measures of center

mean

median

measures of spread

st deviation

range (max - min)

quartile

IQR ( $Q3 - Q1$ )

## Bivariate data

two variable ~~characteristics~~

How the two variables are related.

ex: Highest score of completion vs

crime rate.

Are 2 variable related.

is ~~that~~ <sup>that</sup> relationship linear?

↳ rate of change is constant.

Explanatory variable:

Explains "Input that makes the response variable change",  $x$ .

Response variable.

"Respond" output that changes by varying the value of the explanatory variable.

Correlation is not causation!

i.e. just because two variables are related does not mean one causes the other.

exp = HS grad rate =  $x$

resp = crime rate =  $y$

$$y = \frac{-49.57x}{m} + \frac{6990.2}{b}$$

If grad  $x=0$ , then crime rate is 6990.2 in 100,000.

as grad rate increase, crime rate decrease

$R^2 = 0.1817$ , 18% of the data is explained by the data.

Extrapolation is using the model to estimate values outside the range of the model.

Actual data point is (78, 3199.7)

model estimate is

$$\hat{y} = -49.57 \times 78 + 6990.2$$
$$= 3123.74$$

pearson coefficient.

$r = 0$ , no relationship

$r = -1$ , perfect negative relationship

$r = 1$ , perfect positive relationship

