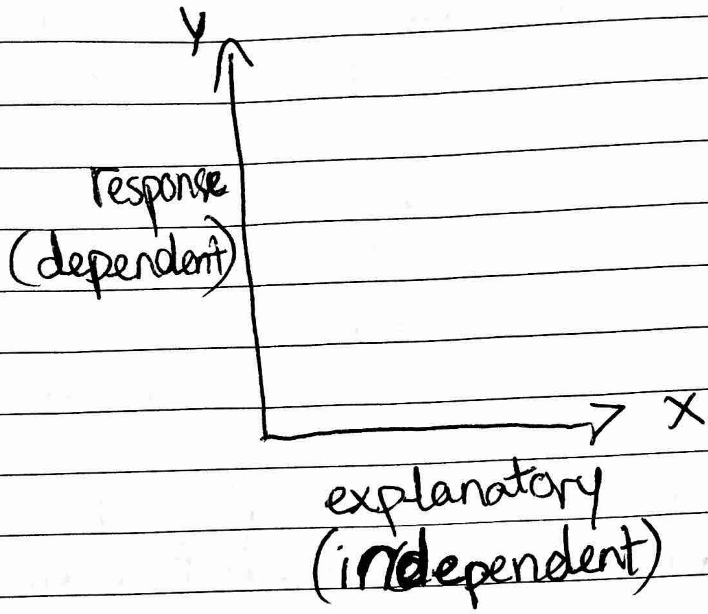
Math 13 Notes 02/11/2019

- standard deviation: average distance from the mean (between data and mean) ; says how spread out data is from the mean (center)
  - Use SD to create a typical range of values: $\bar{x} - SD$ to $\bar{x} - SD$
- If data is unimodal and symmetric, generally ~68% of data falls between $\bar{x} - \sigma$ to $x + \sigma$
- Outliers for symmetric graphs/data: below $(\bar{x} - 2 \cdot SD)$ and above $(\bar{x} + 2 \cdot SD)$, where 95% of data is.

*- Modules 4 - 13 will be the content of the test

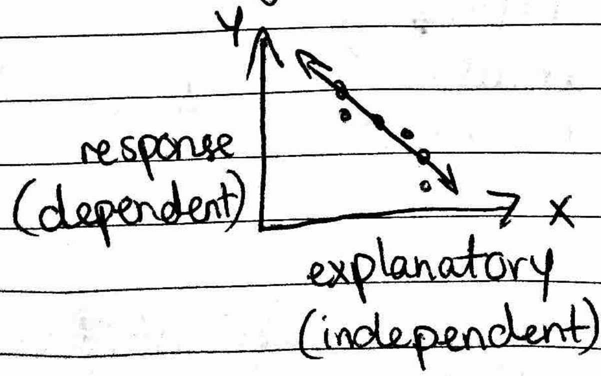
- Before:
  - We looked at one numerical variable (values that can be averaged like money, temperatures, weight, etc.)
    And, we looked at one numerical/quantitative variable and one categorical variable (categories like gender, zip code, etc.)
- Now (Unit 4):
  - We're examining 2 quantitative variables (numerical) and their association/relationship

Ex. of graph w/ explanatory and response variables:



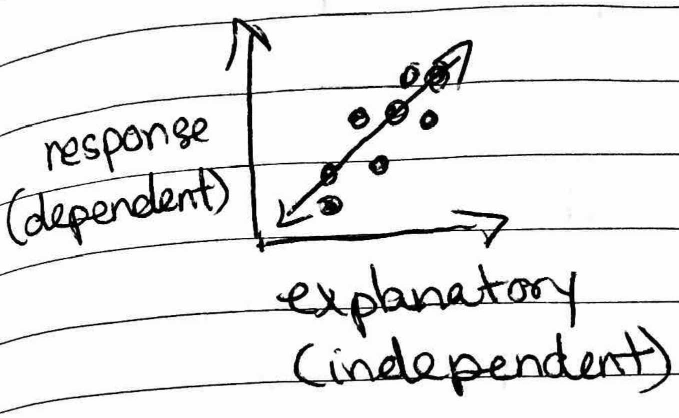
- When looking at scatterplots, we will use direction, form, strength and unusual factors to describe them

- Direction: can be positive (+) or negative (−); think of slope *



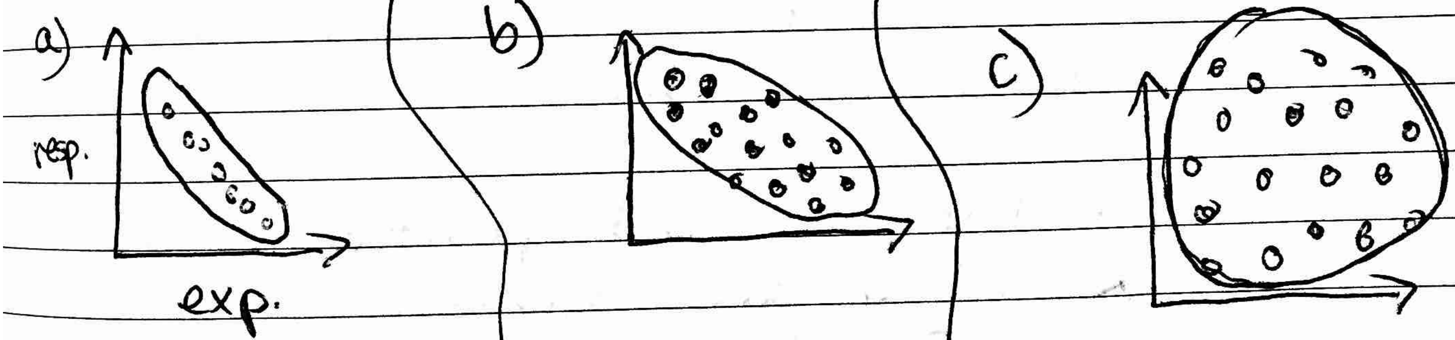
* Negative Direction: as y increases, x decreases

- Ex. of positive direction:



*Positive Direction: as y ~~decreases~~ increases, x increases

- Ex. of associations:



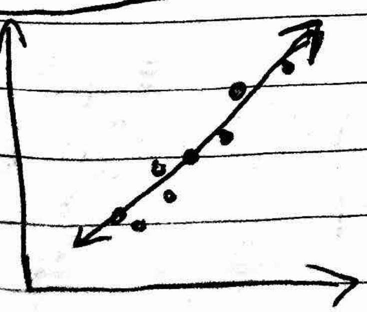
ⓐ is a strong association

ⓑ is a moderate association

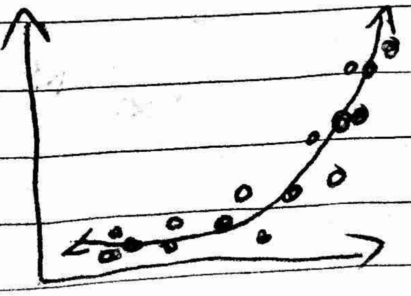
ⓒ has no linear association

- Form: can be linear, curved, blob of points (no association)

Ex. of linear form

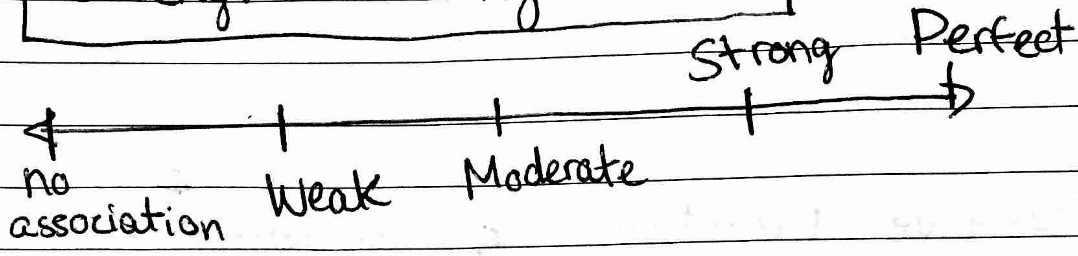


Ex. of curved form



- Strength: sliding scale associated with shape of points and how close together (condensed) they are.

Strength sliding scale:



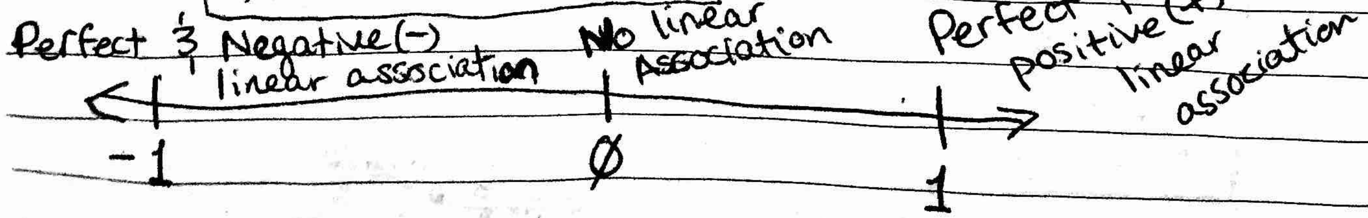
- Correlation does NOT establish cause and effect (i.e. correlation is not causation).
- You need a well designed experiment to establish cause

* Positive Direction and positive association = x and y are increasing ; Negative Direction and negative association = x is increasing and y is decreasing

- Correlation measures the strength of a LINEAR association. If its not linear, we don't calculate it.

- Correlation is a number between -1 and 1.

Ex. of correlation



- Equation for correlation coefficient (r):

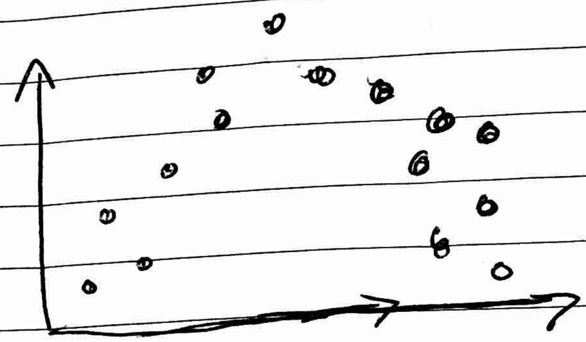
$$r = \frac{\Sigma (y-\bar{y})(x-\bar{x})}{SD_y \, SD_x \, (n-1)}$$

- SD_y: Standard deviation of y
- SD_x: Standard deviation of x
- $(y-\bar{y})$: Each y value minus the mean for the y values
- $(x-\bar{x})$: Each x value minus the mean for the x values
- $(n-1)$: # of numbers (# of data points) minus 1

- Strength can be dependent on the field. For example, in psychology +0.6 for a correlation is strong. However, in other fields of science +Ø.6 is moderate. In both cases, +Ø.6 is a positive correlation but how strong it is said to be is dependent on the field. For our purposes, we can see +Ø.6 as a moderate correlation because of it's relative distance from +1.
- Correlation: requires 2 numerical variables for x and y, has no units, positive correlation value will indicate positive direction of the data, neg. correlation value indicates neg. direction,

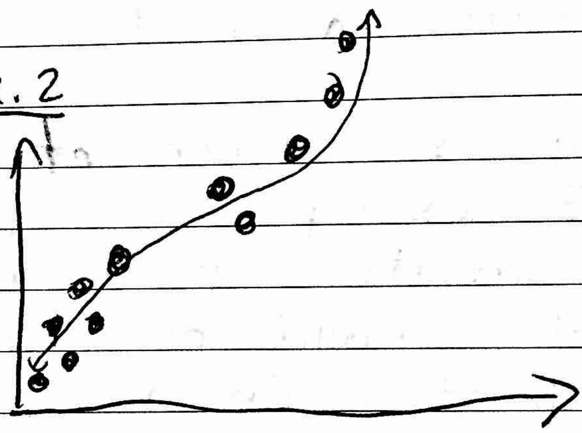
• Don't confuse association with correlation

Ex. 1



* There is an association but there is no linear association.

Ex. 2



* The correlation here could be calculated to be about +ø.9 (a strong correlation), but it's curved. Therefore, this value isn't completely accurate.

Unit 4 Class Exercise:

1a) As distance increases, time increases. Time is dependent on distance, so time is y and distance is x.
— Positive correlation + strong relationship

1b) Brightness is more when you are closer (distance lower). Brightness

1b) ... is dependent on distance. Brightness is y and distance is x. This is a negative correlation.