Section 6.1: Measures of Centre:

Mean and Median

numerical measures - another way to summarize
a data set

(graphs → visual summary)

particularly useful when comparing data
sets whole values are
very similar

numerical measures are quantities (numbers) that describe
a data set

→ called <u>parameters</u> for populations

→ called <u>statistics</u> for samples

we'll look at three types of numerical measures:

- measures of centre
   - where most of the data points
      are located

- measures of variability
   - is there a lot of spread in the data?

- measures of relative standing
   - where is <u>this</u> data point located
      in comparison to all the others?

measures of centre :

mean (average) - add all of the values of the data points together and divide by the number of points

population mean:

$$\mu \quad = \quad \frac{\sum_{i=1}^{n} x_i}{n}$$

Greek letter "mu"

where x_i is the ith measurement and n is the size of the population

sample mean:

$$\overline{x} \quad = \quad \frac{\sum x_i}{n}$$

"x-bar"

one problem - mean is greatly affected by outliers

median - if you write your data in an ordered list, it's the middle value

- if you have an even number of points, it's the average of the two middle points

you can, if you wish, write the median as $\tilde{x}$

example: starting salary of Camosun graduates*

*totally made-up data

$35,000

$45,000

$37,000

$60,000

$2,400,000

find the mean and median of this data set

answer: mean = $515,400

median = $45,000

the mean here is misleading because the outlier
has dragged the mean in that direction

for the median, if there are lots of data points
how do you determine the position of the middle value?

$$position = \frac{1}{2}(n+1)$$

for 75 points, the median is the 38th point

for 76 points, the median is the 38.5th point

take average of the

take average of the 38th and 39th point

mean and median for skewed distributions

mean and median

symmetrical

median
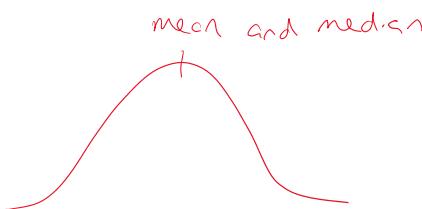
mean - pulled in the direction of the tail
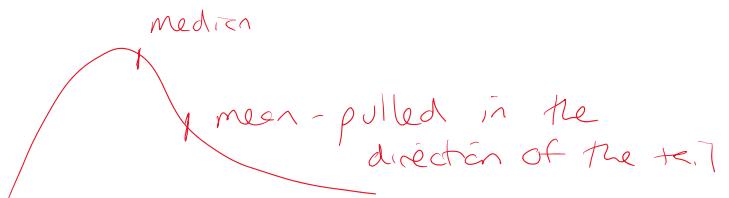
skewed to the right

note: where's the median located on an asymmetrical graph?

median
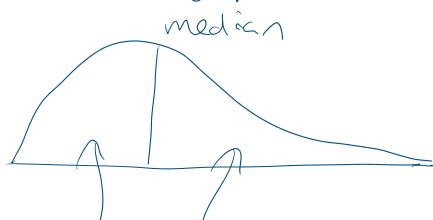
these two areas are equal

where is the mean located?

mean

pivot point
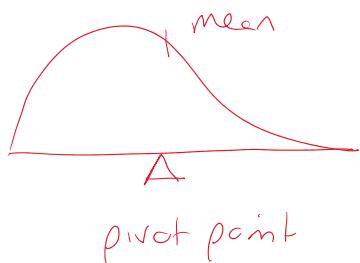
weighted average:

suppose you have the following data set:

1, 1, 1, 1, 1, 2, 2, 3, 3, 3, 3

you could summarize this data in a frequency tables

x	f
1	5
2	2
3	4

↑ values of x

↑ number of times this value occurs in the data set

calculating the mean:

for a sample, $\bar{x} = \dfrac{\sum x_i \cdot f_i}{n} = \dfrac{\sum x_i f_i}{\sum f_i}$

a note on rounding means:

- you can generally report the mean of a set of measurement to at least one more decimal place than the measurements themselves

25, 37, 42, ...

→ can report the mean as 32.5
at least one more decimal place

(to do it properly, need to calculate the standard deviation first)