Section 6.1: Measures of Centre

Wednesday, March 13, 2024 11:10 AM

numerical measure - another way to summarize a data set

(graphs are a visual summary)

will look at three types of numerical measures

- measures of centre
  - where is the "middle" of the data
- measures of variability
  - is there a lot of "spread" in the data
- measures of relative standing
  - where is this data point located in comparison to all the others

---

measures of centre:

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

population mean:

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

n: size of population

x_i: the i^{th} measurement

μ: Greek letter "mu"

Sample mean:

$$\bar{x} = \frac{\sum x_i}{n}$$

n ← size of sample

one problem: mean is greatly affected by outliers

---

median - if you write your data as 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

example: starting salary of Camosun graduates *

* totally made-up data

$35 000
$45 000
$37 000
$60 000
$2 400 000

35, 37, 45, 60, 2400

find the mean and median of this data set

answer: mean = $515 400
median = $45 000

the mean value can be misleading because the outlier has dragged the mean in

the mean will be misleading because the outlier has dragged the mean in that direction

note: there is a notation for median which we're not going to use:

mean $\bar{x}$

median $\tilde{x}$

---

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

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

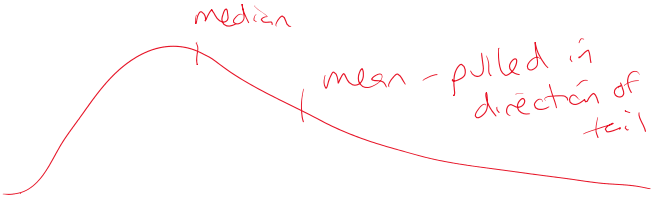
for 75 points, the median is the 38th point

for 76 points, the median is the 38.5th point

average of 38th and 39th points

mean and median for skewed distributions

symmetrical



skewed to the right

---

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 table

x	f
1	5
2	2
3	4

↑ values of x

↑ number of times the value of x appears in the data set

calculating the mean:

for a sample, $\bar{x} = \frac{5(1) + 2(2) + 4(3)}{5 + 2 + 4}$

$= \frac{\sum x_i \cdot f_i}{\sum f_i}$

a note on rounding means:

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

25, 37, 42, ... → can report the mean as 32.5

=

one more decimal place

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