# Section 1.4: Measures of Centre and Variability

numerical measures - another way to summarize a data set

(graphs → visual summary)

particularly of use when comparing two data sets whose values look very similar

numerical measures are quantities that describe a data set

→ called parameters for populations

→ called statistics for samples

we'll look at 2 types:

- measures of centre
- measures of variability

---

measures of centre:

<u>mean</u> (average): add up the data points and divide by the number of data points

population mean:

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

Greek letter "mu"

where $x_i$ is the $i^{th}$ data point

sample mean :

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

one problem :  affected greatly 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 values

example:   starting salary for Civil Tech graduates

$60 000
$45, 000
$30, 000
$30, 000
$5, 000, 000

30, 30, ⑤45, 60, 5000

calculate the mean and median

mean   =   $1, 033, 000

median   =   $45, 000

note:   if there are lots of data points, how do you decide which is the "middle" one?

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

for 75 data points, median is the 38th point

for 76 data points, median is the (38.5th) point

average of
38th and 39th

---

weighted averages:

| x | f |
|---|---|
| 1 | 5 |
| 2 | 2 |
| 3 | 4 |

← frequency table — the data set is

$1,1,1,1,1, 2,2, 3,3,3,3$
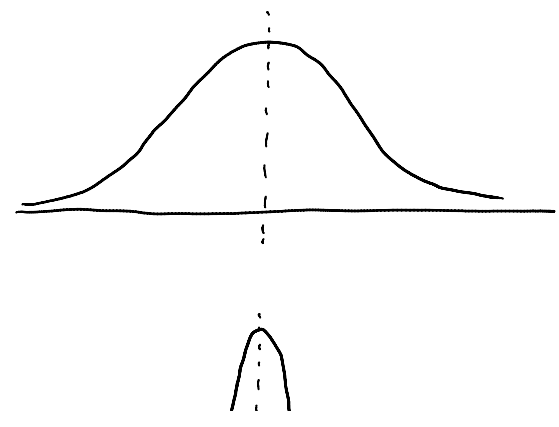
↑ values of x

↑ number of times that value occurs in the data set
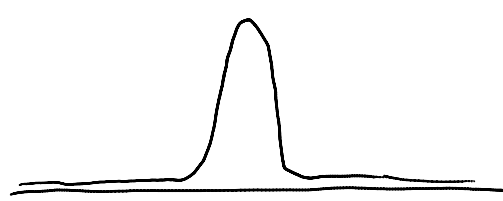
if it's a sample:

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

$$= \frac{1 \cdot 5 + 2 \cdot 2 + 3 \cdot 4}{11}$$

---

measures of variability:



two distributions with same mean value

but different "widths"

but different "widths"

– we want to characterize the "spread" of the distribution

note: when do you want a small spread?
   when trying to make identical widgets

when do you want a large spread?
   when you are try to make distinctions
      → you are ranking objects

———————————

easiest measure of variability to calculate:

range – difference between the max and
      min values

good part: easy to calculate

bad part: almost completely useless

      → heavily influenced by
                     outliers
      → only depends on two
            values from entire
            data set