Statistics Suggested Homework
   is on website.

Quiz Tues March 12   Section 3.3

Test Wed March 13
   3.1-3.3, 4.1-4.3, 5.1
      (8 Questions)
   Practice Questions on website
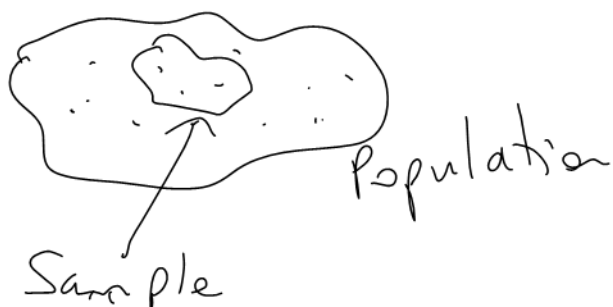   Formula sheet will be provided

Ch 6   Describing Data with Numbers

   6.1   Mean and Median

Recall
Population : The set of all measurements of interest.
 Sample : A subset of the population.

Sample
Population

The mean of a data set is commonly
   known as the average.

A population mean is written μ.
A sample mean is written $\bar{x}$.

Ex: Find the mean of:
  a) the population 1, 1, 2, 8, 12, 15

$$\mu = \frac{1+1+2+8+12+15}{6}$$

$$= 6.5$$

  b) the sample   1, 2, 15

$$\bar{x} = \frac{1+2+15}{3}$$

$$= 6$$

The median is the middle value when the data is ordered.
If the number of measurements is even, average the middle two measurements.

Ex: Find the median:
  a)   12, 2, 9
ordered: 2, 9, 12   median = 9

b)   12, 8, 2, 9

ordered: 2, 8, 9, 12

$$\text{median} = \frac{8+9}{2} = 8.5$$

Ex: Salary of 5 people (in thousands of $):

35, 45, 47, 60, 2400

a) Find the mean

$$\text{mean} = \frac{35 + 45 + 47 + 60 + 2400}{5} = 517.4$$

b) Find the median

ordered ✓
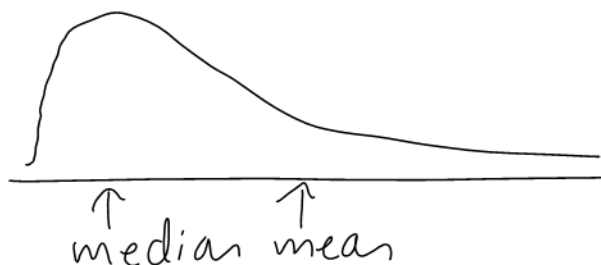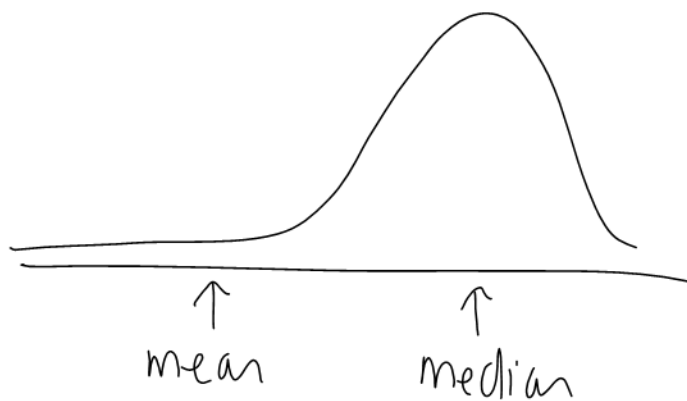
median = 47

c) which is more representative?

median

The mean is pulled in the direction of the outlier(s).

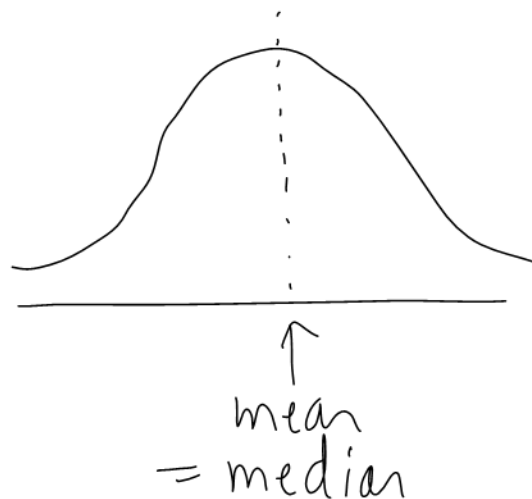If the data is skewed right then mean > median.

If the data is skewed left then
mean < median

mean median

If the data is symmetric

mean = median

mean
= median

Ex: The mean of 4 measurements is 14. The mean of 3 of the measurements is 13. Find the 4th measurement.

Call the measurements a, b, c, d.

$$\frac{a+b+c+d}{4} = 14$$

Mult. by 4 : $a + b + c + d = 56$ ①

$$\frac{a + b + c}{3} = 13$$

Mult. by 3: $a + b + c = 39$ ②

① − ② : $d = 17$

Ex: Consider 130, 170, 180.
    The mean is 160.
    The median is 170.
    Find the new mean and median
    if every measurement is :

a) increased by 40.
   New data set : 170, 210, 220
   new mean = $\frac{170 + 210 + 220}{3}$ = 200
   new median = 210
        Notice : Both increased by 40.

b) increased by 10%
        (multiplied by 1.1)

Multiplied by 10% $\times 0.1$
Increased by 10% $\times (1 + 0.1)$

New data set: 143, 187, 198

new mean = $\dfrac{143 + 187 + 198}{3} = 176$

new median = 187

Notice: Both mean and median
       increased by 10%

FACTS

If every measurement is multiplied by k
then mean and median get multiplied
    by k.

If every measurement is increased by k
then mean and median get
    increased by k.
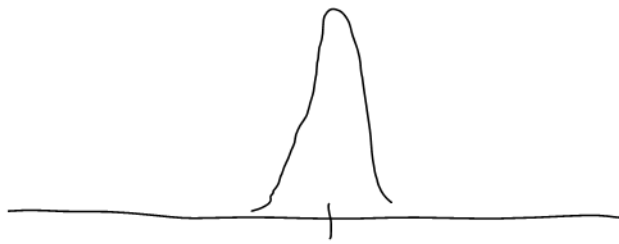
6.2 Measures of Spread

Two data sets with the same mean:

larger spread

smaller spread

Range = maximum value − minimum value

Quick Ex: 7, 54, 42, 3, 15

ordered: 3, 7, 15, 42, 54

range = 54 − 3 = 51

Range is too simple to be useful.

Population Variance: The average of the squared distances to the mean. It's written σ^2.

σ is the <u>population standard deviation.</u>

Ex: Consider the population 20, 50, 77.
Calculate σ^2 and σ.

mean $\mu = \dfrac{20+50+77}{3} = 49$

x	$x-49$	$(x-49)^2$
20	-29	841
50	1	1
77	28	784

population variance $\sigma^2 = \dfrac{841+1+784}{3}$

$\qquad\qquad\qquad\qquad = 542$

population standard deviation $\sigma = \sqrt{542}$
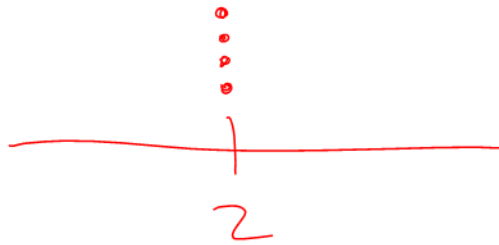
$\qquad\qquad\qquad\qquad\qquad\qquad \approx 23$

Won't be tested

FACT
The larger σ is, the more spread out
the data is.

Ex: Without calculating,
find the standard deviation
of   2, 2, 2, 2.

$$\sigma = 0$$

Ex: Data Set A has $\sigma = 3$
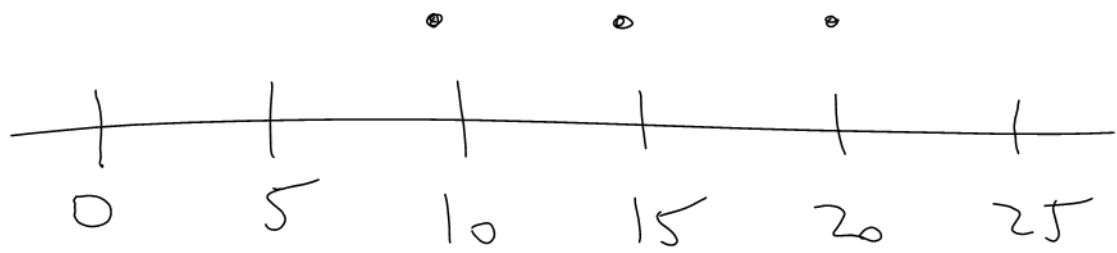"   B   " $\sigma = 5$
Which is more spread out?
Data Set B

Ex: Data Set A: 10, 15, 20
"   B: 5, 15, 25
which is more spread out?

Ⓐ

Ⓑ

Data Set B

FACT

If all measurements are increased by k, then σ is unchanged.

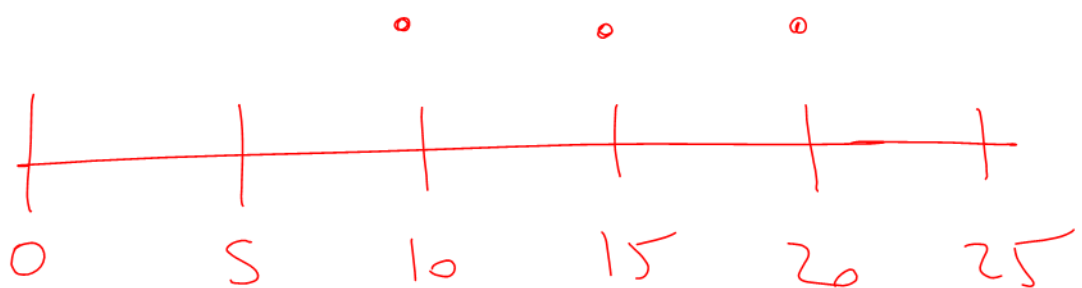If all measurements are multiplied by the positive number k, then σ is multiplied by k.

Ex : Data Set A : 10, 15, 20
Data Set B : 15, 20, 25
Which has the larger standard deviation?

Notice: We added 5 to each measurement.

A

```
    +    +    +    +    +    +    +
    0    5    10   15   20   25
```

B

The standard deviations
are the same.

Ex: Data Set A: 10, 15, 20
      "   B: 20, 30, 40

Which is more spread out?

Notice: We multiplied by 2.

(A)

```
    +    +    +    +    +    +
    0    10   20   30   40
```

(B)

Data Set B