

4.1 Measures of Central Tendency

A *measure of central tendency* is a number or piece of data used to represent a typical value for a data set.

Three common measures of central tendency are the *mean*, *median* and *mode*.

mean - the mathematical average of n numbers

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

Σ sigma
- the sum of
 n - how many items
 x_i arbitrary items

{ 8, 1, 4, 20, 69 }
 $x_1 = 8$ $x_2 = 1$ $x_3 = 4$... $x_5 = 69$

median - for n numbers, the median is the middle number in a ordered set of numbers, or the mean of the middle two.

mode - the most frequently occurring number in a set

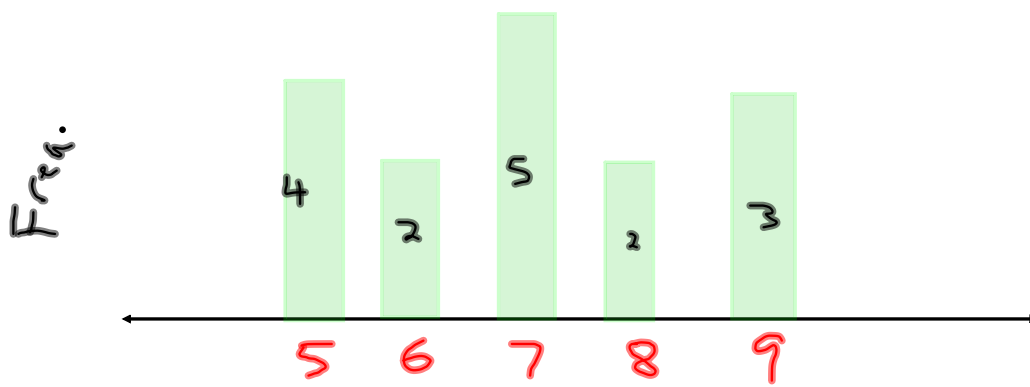
- there can be no mode, one mode, or many modes
- NOTE: MS Excel will only find the first mode

3, 4, 5

mode = no mode
mode(s) = 3, 4, 5

3, 3, 4, 4, 5, 5

mode(s) = 3, 4, 5



MODE = 7

MEDIAN: 7

5, 5, 5, 5, 6, 6, 7, 7, 7, 7, 7, 8, 8, 9, 9, 9

MEAN:

...

Ex. 1 Find the mean, median and mode for the following set:
90, 102, 120, 180, 90, 85, 90, 137, 120, 145, 97, 93, 120

mean: 113

MED. : 102

MODE(S) = 90 & 120

Can you explain the reasoning behind the mean being so much greater than the median?

Outliers

Practice: find the mean, median and mode for each set

a) 62, 23, 27, 56, 52, 34, 42, 40, 70, 45, 77

b) 38, 40, 29, 34, 8, 40, 47, 26, 43, 49, 14, 40

Appropriate Measures of Central Tendency

For a set of data, the three measures of central tendency may be nearly identical or may vary widely. Determining which of the three measures is most useful depends on the context in which the data are being used.

ex. school periods (discrete)

Numeric vs. Non-Numeric Data

Which measure(s) can be used with numeric data?

if conts.

mean, med., mode

Which measure(s) can be used with non-numeric data?

mode only

Discrete vs. Continuous Data

A *discrete* data set contains a ^{finite} countable number of items in the set
ex. sizes of milk containers (1 gal, 1/2 gal, quart, pint)

A *continuous* data set contains items that can take on an infinite number of values
ex. the temperature outside (76, 76.000000001, 76.000000002, etc.)

Which measure(s) can be used with discrete data? ^{Median} ~~mean~~ ^{sometimes} ^{MODE}

Which measure(s) can be used with continuous data? all 3

shoes : 6, 7, 8½
 $\bar{x} = 7\frac{1}{6}$ WTF?
5, 7½, 8, 9
med: 7¾ WTF?

Outliers

Outliers are items in a set of data that vary greatly from the other items in the set. In a few weeks, we will be mathematically testing to see if an item in a set is indeed an outlier.

Which measure(s) should be used when a set contains outliers?

ages: 16, 17, 17, 17, 17, 17, 18, 18, 33
DON'T USE THE \bar{x}

Ex. 2 Choosing Mean, Median or Mode

Tell whether mean, median or mode best represents the data. Explain your reasoning.

NOT one is best b/c...

The other 2 SUCK b/c ...

a) waiting times to see the doctor (in minutes)

24, 19, 39, 22, 26, 58, 16, 64, 27

median

there is no mode

outliers \Rightarrow ~~mean~~

b) money (in dollars) earned baby-sitting

30, 10, 15, 25, 10, 35, 20, 25, 30, 10

mode b/c popular
~~mean or median~~

$$\bar{x} = 21 \quad \text{med} = 22.50$$

c) Shoe sizes of females in a bowling alley

5.5, 9, 6, 7.5, 6, 8, 6, 8.5, 6, 9.5, 6, 7, 6.5, 9

Discrete \Rightarrow mode

Practice

Mean, median or mode. Why?

a) 21, 25, 10, 18, 7, 9, 3, 8, 14, 30, 12, 29, 11, 6, 35, 30

(Med)
b/c range

b) 3.65, 4, 3.25, 4.15, 3, 3.05, 3.25, 3.85

n, c, no out. (\bar{x})

c) 6, 6, 5, 6, 6, 20, 6, 6, 6, 6

MODE b/c

↓
out. ~~✗~~

For a set containing all whole numbers, are the mean, median and mode(s) necessarily whole numbers? Explain.

\bar{x} not always

MED not always (it will when n is odd)

MODE - always (assuming it exists)

