

Math 0212-4b

More Measures for Central Tendency

- Mean
- Median
- Midrange
- Mode

The **median** is the midpoint of the data array. The symbol for the median is MD.

Steps in computing the median of a data array

Step 1 Arrange the data in order.

Step 2 Select the middle point.

3, 3.3, 1, 0, 2, 22, 10, 5, 6.1

Total 9 numbers!

Order from the smallest to the largest:


0, 1, 2, 3, 3.3, 5, 6.1, 10, 22

The middle position is at the 5th position, where the value is 3.3, therefore

MD = 3.3

The median means:

0, 1, 2, 3, **3.3**, 5, 6.1, 10, 22



- There are at least 50% of the data **at or below** this median value;

$5/9 = 56\%$

The median means:

0, 1, 2, 3, **3.3**, 5, 6.1, 10, 22

• There are at least 50% of the data **at or above** this median value.

$5/9 = 56\%$

What if there are even number of data in the data set?

For example:

0, 1, 2, 3, 3.3, **5**, 6.1, 10, 22, 30

There are 10 numbers, and 2 are at the middle.

We define:

$$MD = \frac{3.3 + 5}{2} = \frac{8.3}{2} = 4.15$$

General rules to find the median:

- If we have odd number of values in the data set, order the values from the smallest to the largest, the value in the middle is the value of the median;
- If we have even number of values, then there are two values in the middle, the value of the median is the average value of these two.

Properties and use of median:

1. Median can be used for ordinal level of variables.
2. Median is not sensitive to the outliers.
3. We use median if we want to find the mid value of the data set that separates the top half from the bottom half.

Estimate MD in a grouped frequency table:

Selling Price (\$000)	Count (of Houses Sold)	Percent (%)
0 – 49	18	9%
50 – 99	28	14%
100 – 149	70	35%
150 – 199	34	17%
200 – 249	26	13%
250 – 299	20	10%
300 – 349	4	2%

Selling Price (\$000)	Count (of Houses Sold)	Percent (%)	Cumulative
0 – 49	18	9%	18
50 – 99	28	14%	46
100 – 149	70	35%	116
150 – 199	34	17%	150
200 – 249	26	13%	176
250 – 299	20	10%	196
300 – 349	4	2%	200
Total	200		

- Step 1: Total number of values = 200;
 Step 2: $200 / 2 = 100$. Two middle values are at 100th and 101st position.
 Step 3: Use cumulative frequencies to identify the location of these 2 middle numbers. They are both in the class of "100-149".
 Step 4: Take the middle value of the class MD ≈ 124.5

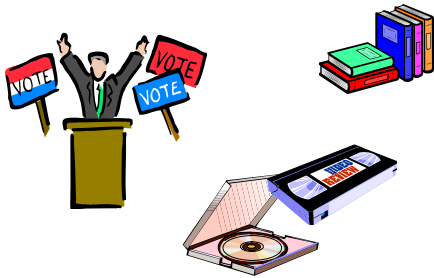
The Mode

- The **mode** is defined to be the value that occurs most often in a data set.
- A data set can have more than one mode (Unimodal, Bimodal, and Multimodal).
- A data set is said to have **no mode** if all values occur with equal frequency.

Examples:

- { 1, 2, 2, 1, 3, 3, 5, 5, 5}
has the mode 5;
- {2, 1, 1, 3, 3, 5, 6, 10}
has two modes 1 and 3 (bimodal).
- {0, 3, 4, 1, 9}
has no mode.

The most popular applications of mode:



Modal class in a grouped frequency table:

Selling Price (\$000)	Count (of Houses Sold)	Percent (%)
0 – 49	18	9%
50 – 99	28	14%
100 – 149	70	35%
150 – 199	34	17%
200 – 249	26	13%
250 – 299	20	10%
300 – 349	4	2%

Properties and uses of mode:

- Mode is generally used for categorical data.
- Mode is used to find the most popular value(s) in the data set.
- Mode should not be used for continuous numerical data.

The **midrange** is defined as the sum of the lowest and highest values in the data set, divided by 2. The symbol MR is used for the midrange.

$$MR = \frac{\text{lowest value} + \text{highest value}}{2}$$

- Simple and fast;
- Extremely sensitive to outliers.

Example:

2, 4, 6, 1, 9, 10, 0, -4, 229

$$MR = \frac{(-4) + 229}{2} = 112.5$$

Comparison of Central Tendency Measures

- In a perfect world, the mean, median & mode would be the same.
- However, the world is not perfect & very often, the mean, median and mode are not the same.

For a data set, the mean, median, and mode can be quite different.

Staff	Salary
Owner	\$50,000
Manager	\$20,000
Salesperson	\$12,000
Technician A	\$9,000
Technician B	\$9,000

The average salary in this company is
Mean = \$20,000


Staff	Salary
Owner	\$50,000
Manager	\$20,000
Salesperson	\$12,000
Technician A	\$9,000
Technician B	\$9,000

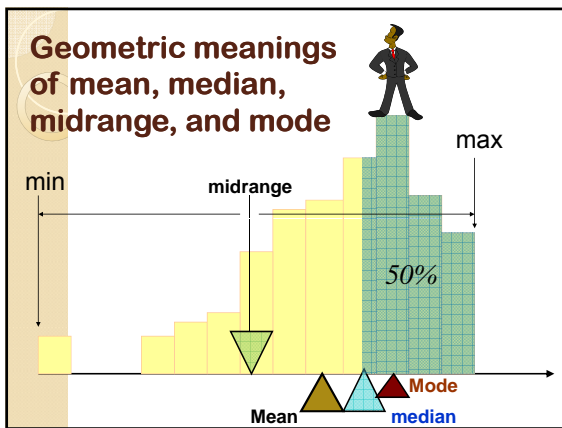
Mean = \$20,000

The median is
MD = \$12,000
i.e. there are at least 50% staff whose salaries are at or below \$12,000,
and at least 50% staff at or above \$12,000

Staff	Salary
Owner	\$50,000
Manager	\$20,000
Salesperson	\$12,000
Technician A	\$9,000
Technician B	\$9,000

Mean = \$20,000
MD = \$12,000
Mode = \$9,000
 i.e. the most frequently seen salary level in this company is \$9,000





Comparison of Central Tendency Measures

- **Use Mean** when distribution is reasonably symmetrical, with few extreme scores and has one mode.
- **Use Median** with nonsymmetrical distributions because it is not sensitive to skewness.
- **Use Mode** when dealing with frequency distribution for nominal data
