

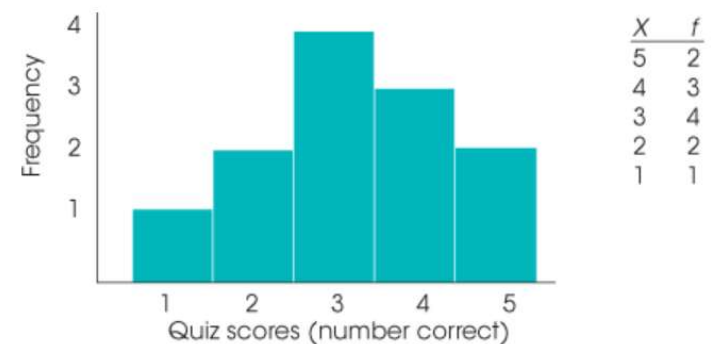
Central Tendency

2.1

Frequency Distributions

- In the previous module we saw that we can efficiently summarize data using frequency distributions.

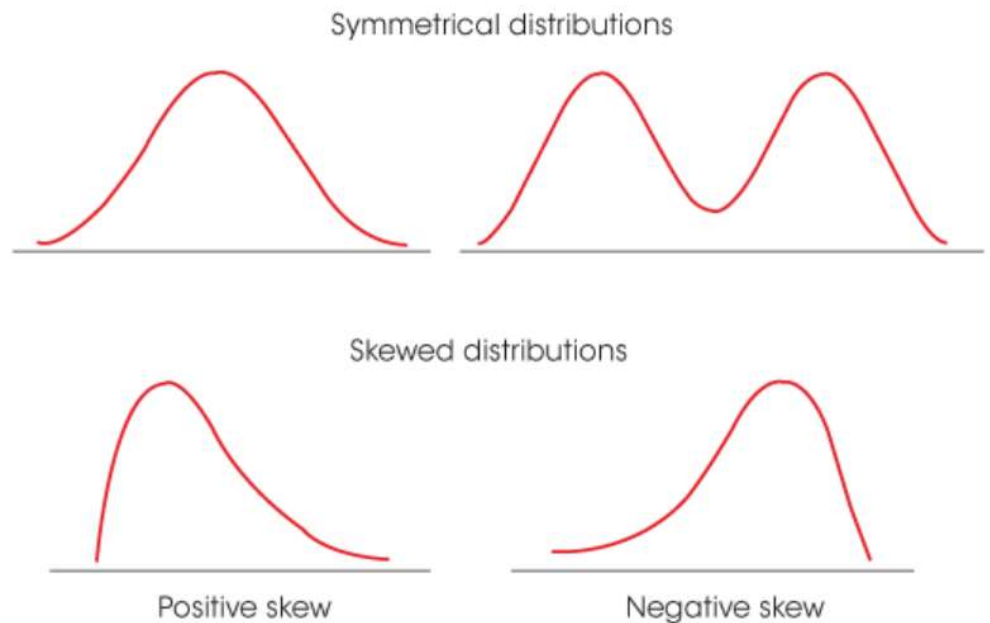
Score	f	p	percentage
2	1	1/7	14.29%
3	2	2/7	28.57%
4	3	3/7	42.86%
5	1	1/7	14.29%





Frequency Distributions

- But! We can summarize even more!
- Especially symmetrical and unimodal (top left) distribution can be summarized efficiently
 - Just need a measure of their central tendency, and
 - a measure of their spread.





Central Tendency

- Central Tendency:
 - a statistic that captures the center of a distribution in a single value.
 - Goal: find the score that is most typical/representative of a group.
- You all know the mean, but there are other options.
 - Which one is best depends on...measurement level and shape of a distribution.



Mean

- The mean is calculated by adding all scores together and dividing by the number of scores.

Population:

$$\mu = \frac{\sum X}{N}$$

Sample:

$$M = \frac{\sum X}{n}$$

Participant	# of shoes
Lennie	4
Joran	5
Leonie	4
Zaïra	4
Vince	3
Natascha	3
Sacha	2



Mean

- The mean is calculated by adding all scores together and dividing by the number of scores.

Population:

$$\mu = \frac{\sum X}{N}$$

Sample:

$$M = \frac{\sum X}{n}$$

Participant	# of shoes
Lennie	4
Joran	5
Leonie	4
Zaïra	4
Vince	3
Natascha	3
Sacha	2

$$4+5+4+4+3+3+2 = 25$$

$$n = 7$$

$$M = \frac{25}{7} = 3.57$$



Mean

- Can also calculate the mean from a frequency table

Participant	# of shoes
Lennie	4
Joran	5
Leonie	4
Zaïra	4
Vince	3
Natascha	3
Sacha	2

$$4+5+4+4+3+3+2 = 25$$

$$n = 7$$

$$M = \frac{25}{7} = 3.57$$

Score	f	f*Score
2	1	2
3	2	6
4	3	12
5	1	5

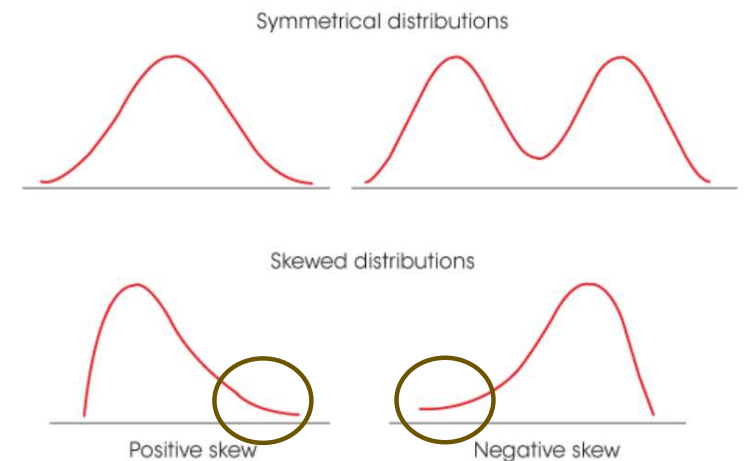
$$\sum f = n = 7 \quad \sum f * Score = 25$$

$$M = \frac{25}{7} = 3.57$$



Mean

- ◎ The mean is a good measure of central tendency for:
 - Continuous data
 - That has a symmetrical distribution
 - Is unimodal
- ◎ The mean is “sensitive” to extreme values:
 - If you add one very large or very small number to your data, the mean changes a lot.
 - This also mean that in a skewed distribution the high (or low) scores in the tail of the distribution have a large influence on the mean of the distribution.



Median

- ⦿ The median is the score that has 50% of scores above it and 50% below.
- ⦿ In other words, it's the 50th percentile.

Participant	# of shoes
Lennie	4
Joran	5
Leonie	4
Zaïra	4
Vince	3
Natascha	3
Sacha	2



Median

- For discrete variables:
 - You can find it by ordering the scores from smallest to highest and finding the middle one

Participant	# of shoes
Lennie	4
Joran	5
Leonie	4
Zaïra	4
Vince	3
Natascha	3
Sacha	2

2, 3, 3, 4, 4, 4, 5





Median

- For discrete variables:
 - You can find it by ordering the scores from smallest to highest and finding the middle one.
 - But what if we have an odd number of observations?! 😱.....
 -take the mean of the middle two numbers.

Participant	# of shoes
Lennie	4
Joran	5
Leonie	4
Zaïra	4
Vince	3
Natascha	3
Sacha	2

3, 3, 4, 4, 4, 5



$$\frac{4 + 4}{2} = 4$$



Median

- For continuous variables:
 - Find the 50th percentile while taking the real limits into account (see video 1.5).

Participant	# of shoes
Lennie	4
Joran	5
Leonie	4
Zaïra	4
Vince	3
Natascha	3
Sacha	2



Mean vs Median

- The mean and the median define “middle” in different ways.
 - The mean defines the middle as the place where the total distance between the scores below the mid-point and the mid-point, AND the scores above the mid-point and the mid-point are the same.

Participant	# of shoes
Lennie	4
Joran	5
Leonie	4
Zaïra	4
Vince	3
Natascha	3
Sacha	2

$$(2-3.75) + (3-3.75) + (3-3.75) = -2,71$$

$$(4-3.75) + (4-3.75) + (4-3.75) + (5-3.75) = 2,71$$



Mean vs Median

- The mean and the median define “middle” in different ways.
 - The mean defines the middle as the place where the total distance between the scores below the mid-point and the mid-point, AND the scores above the mid-point and the mid-point are the same.
 - The median is the middle score.

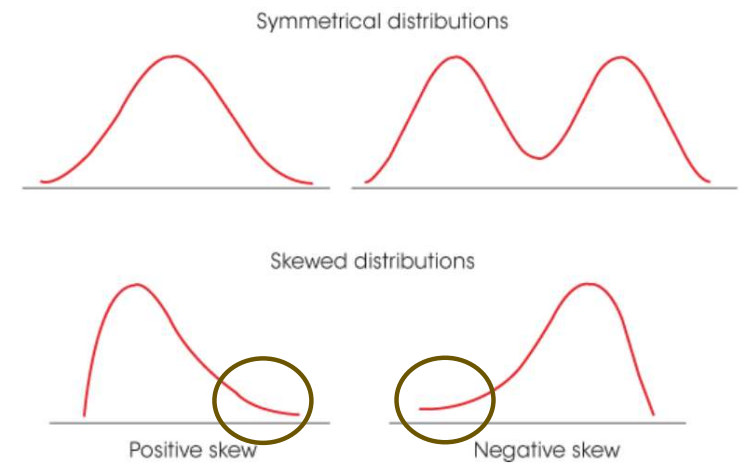
Participant	# of shoes
Lennie	4
Joran	5
Leonie	4
Zaïra	4
Vince	3
Natascha	3
Sacha	2

$$(2-3.75) + (3-3.75) + (3-3.75) = -2,71$$

$$(4-3.75) + (4-3.75) + (4-3.75) + (5-3.75) = 2,71$$

Median

- ◎ The median is a good measure of central tendency for:
 - Continuous and ordinal data
 - That has a skewed distribution
 - Is unimodal



Mode

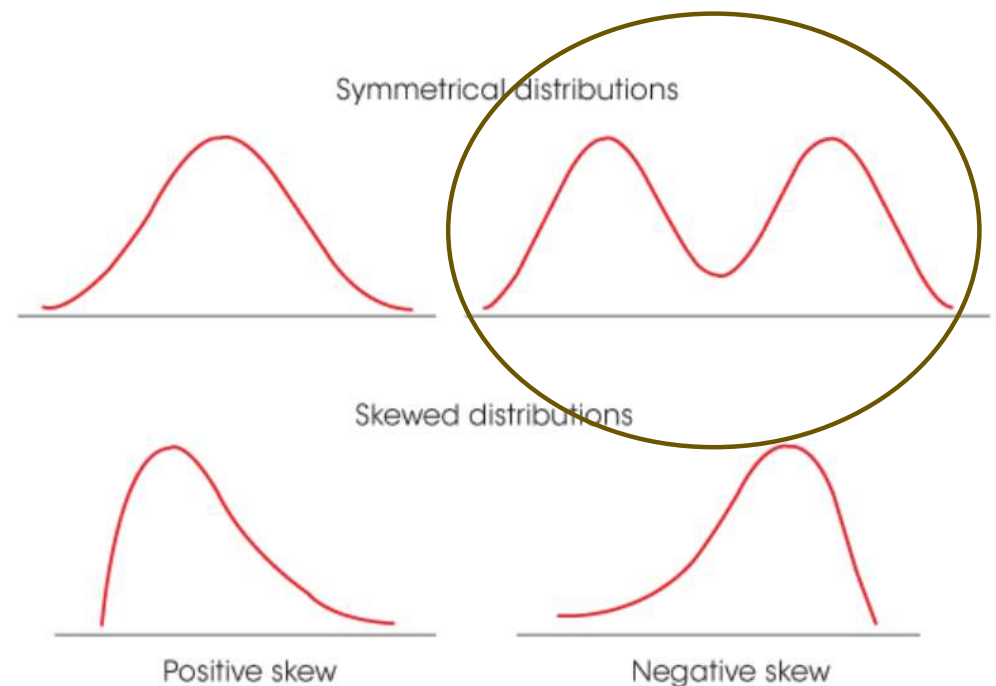
- The mode is simply the score that occurs most often.

Participant	# of shoes
Lennie	4
Joran	5
Leonie	4
Zaïra	4
Vince	3
Natascha	3
Sacha	2

Score	f
2	1
3	2
4	3
5	1

Mode

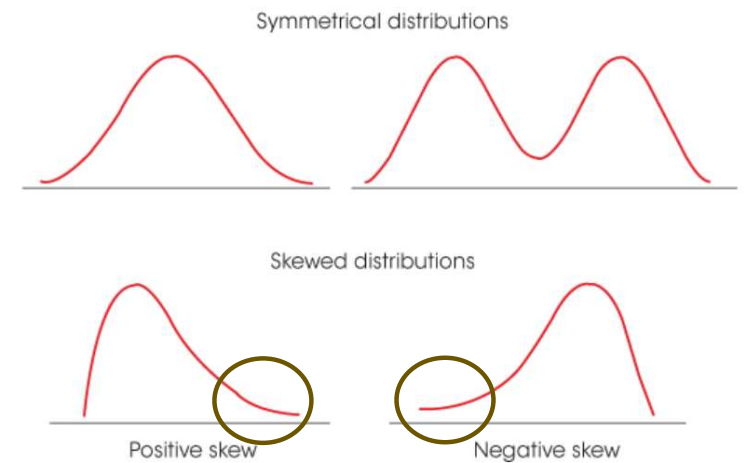
- While a distribution always has one mean and median, it can have multiple modes.
 - If the peaks are not equally large, we still call them modes.
 - The “higher one” is the major mode.
 - The “lower one” is the minor mode.





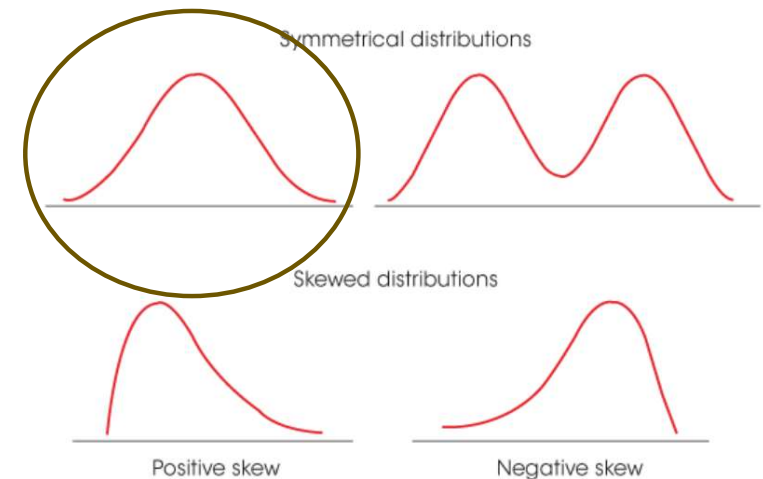
Mode

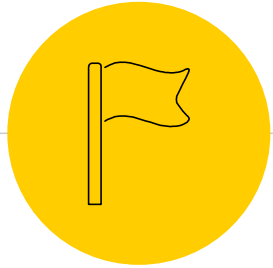
- ◎ The mode is a good measure of central tendency for:
 - Nominal and ordinal (discrete) data, but can be used with continuous data as well
 - That has a symmetric or skewed distribution
 - Is either uni- or multimodal



Mode

- ☉ The mode is a good measure of central tendency for:
 - Nominal and ordinal (discrete) data, but can be used with continuous data as well
 - That has a symmetric or skewed distribution
 - Is either uni- or multimodal
- ☉ Note that for a continuous, unimodal, symmetrical distribution the mean, median, and mode all have the same value.





The End

Any **questions** ?

Go to Canvas Discussions and ...

- ask your question
- like relevant questions of others