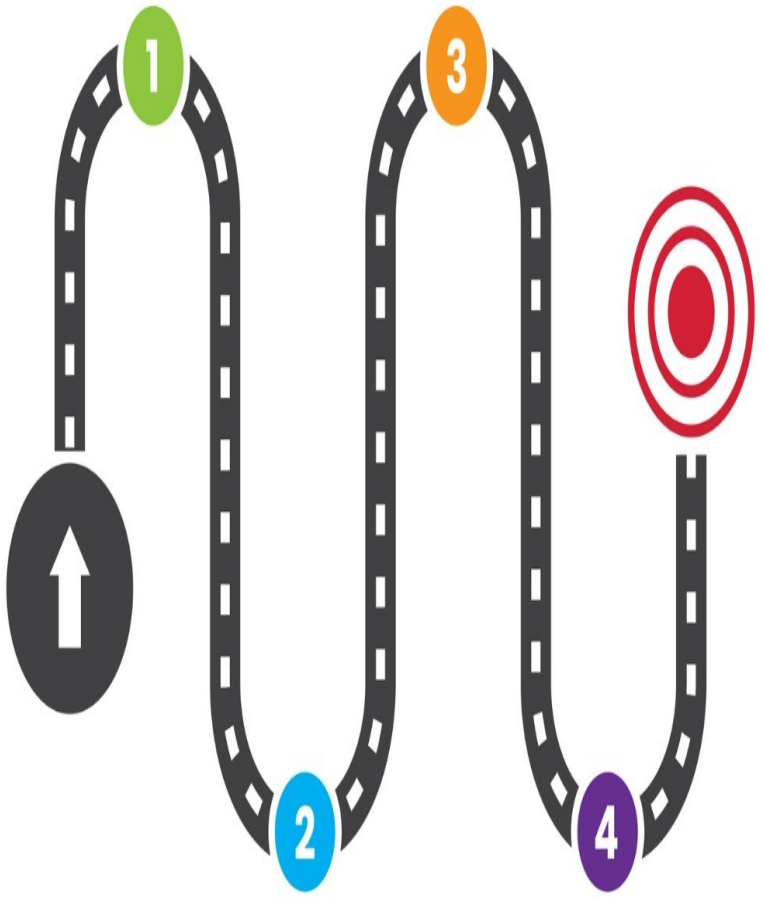
Chapter 3 : Measures of Central Tendency

STATISTICAL METHODS IN SOCIOLOGY

Roadmap

1. Reviewing Ch3 RQ
2. Quick recap
3. Working on Ch3 LA together



Reviewing Ch3 RQ

Q7. Take a look at Table 3.1 from your book. What is the mode of this distribution?

Table 3.1 Ten Most Common Foreign Languages Spoken in the United States, 2009–2013

Language	Number of Speakers
Spanish	37,458,624
Chinese	2,896,766
Tagalog	1,613,346
Vietnamese	1,399,936
French	1,307,742
Korean	1,117,343
German	1,063,773
Arabic	924,374
Russian	879,434
French Creole	739,725

Source: U.S. Census Bureau, *Census Bureau Reports at Least 350 Languages Spoken in U.S. Homes*, 2015.

Source: U.S. Census Bureau, *Census Bureau Reports at Least 350 Languages Spoken in U.S. Homes*, 2015.

Figure 3.1 Respondent's Prospects of Improving Standard of Living (GOODLIFE), GSS 2014

a) French Creole
b) 739,725
c) Spanish
d) 37,458,624

- The mode is always **a category** or score, not its frequency (p.76).
- The mode is Spanish, not 37,458,624

Reviewing Ch3 RQ

Q8. Consider the following information about hate crimes, by state, in the United States. Number of hate crimes is an interval-ration level variable. What is the median number of hate crimes in this distribution? (even number)

1. Order the number of hate crimes
2. Find the middle position (N=8) = (8+1) ÷2 = 4.5
3. Get the median = (140 + 145) ÷ 2 = 142.5

Number of Hate Crimes	States
50	Pennsylvania
65	Florida
109	Illinois
140	North Carolina
145	Texas
311	Michigan
759	California
545	New York
Total (*N*) = 8	

Median (ex: raw data, interval, odd)

Find the median of the weight listed in the first example.

274 235 223 268 290 285 235 **(total number : 7)**

1. Order the data.

(min) 223 235 235 268 274 285 290 (max)

2. Count the cases and find the middle ones

= Location of median: (n + 1) / 2 = (7+1)/2 = 4

The median weight of the adults is 268 pounds.

Reviewing Ch3 RQ

Q9. In what category of this variable does the 50th percentile (also known as the median) fall?

Table 3.3 Days per Week Having Dinner With Parent(s), Monitoring the Future 2014

Days per Week Having Dinner With Parent	Frequency (f)	Percentage (%)	Cumulative Percentage (C%)
Less than 1 day	1,445	16	16
1 day per week	492	6	22
2 days per week	711	8	30
3 days per week	1,024	11	41
4–5 days per week	1,909	21	62
6–7 days per week	3,326	37	99
Total (*N*)	8,907	99[a]	

Note: a. Due to rounding, the percentage total does not add up to 100.

Note: a. Due to rounding, the percentage total does not add up to 100.

- **A percentile:** a score at or below which a specific percentage of the distribution falls (Cumulative percentage).
- Tips: find the category that goes over **cumulative value 50% for the first time**!

Interpretation: the median of days per week having dinner with parents is “4-5 days per week.”

Getting Ready for Ch3 LA: Mode

- **Mode:** The category or score with the *greatest* frequency of percentage/ *the most typical* value in the distribution.

- **nominal**

Political Party	Frequency
Democrat	56
Republican	44
Independent	39
Other	12

- **ordinal**

Light exercise	Frequency
Per day	2,284
Per week	3,119
Per month	135
Per year	99
Unable to do sports	1576
Total	7,213

- **interval-ratio**

Number of Children	Frequency
0	105
1	207
3	398
4	120
5	57
6	6

Getting Ready for Ch3 LA: Median

- **Median:** the score that divides the distribution into *two equal parts* so that half the cases are above it and half are below it.

- **Steps to determine the median:**
  1. Arrange the responses in order.
  2. Count the number of cases or observations.
  3. Find the response associated with the middle case in the distribution:
     Location of median = (n + 1) ÷ 2
  4. Match the location of the median to the value in the distribution.

Median (ex: frequency distribution)

Table 4. Frequency distribution of self-reported age at first sex

Age at First Sex	Frequency	Percent	Cumulative Frequency
0-2	1	0.01	1
3-5	68	0.89	69
6-8	66	0.86	135
9-11	146	1.91	281
12-14	1832	23.97	2113
15-17	3865	50.58	5978
18-20	1485	19.43	7463
21-23	171	2.24	7634
24-26	8	0.10	7642
Total (N)	7642	100.00	

Source: National Longitudinal Survey of Youth 1997 (Waves 1-8)

Number Distribution:

0-2 (1 time), 3-5 (68 times), … 24-26 (8 times)

Total Cases: 7642

Median Position: (7642 + 1) ÷ 2 = 3831.5

Median: Both 3831st and 3832nd cases fall within the range "15-17".

Median (ex: frequency distribution)

Table 4. Frequency distribution of self-reported age at first sex

Age at First Sex	*Frequency*	*Percent*	*Cumulative Frequency*	*Cumulative Percent*
0-2	1	0.01	1	0.01
3-5	68	0.89	69	0.90
6-8	66	0.86	135	1.77
9-11	146	1.91	281	3.68
12-14	1832	23.97	2113	27.65
15-17	3865	50.58	5978	78.23
18-20	1485	19.43	7463	97.66
21-23	171	2.24	7634	99.90
24-26	8	0.10	7642	100.00
Total (*N*)	7642	100.00		

Source: National Longitudinal Survey of Youth 1997 (Waves 1-8)

Identify the response category that **just surpasses 50%.**

We know that the category **"15-17"** is the median, because it contains between 27.65% and 78.23% of cases—*including* the case at 50%.

Getting Ready for Ch3 LA: Mean

- **Mean:** the *arithmetic average* obtained by adding up all the scores and then dividing by the total number of scores

$$\overline{Y} = \frac{\sum Y}{N}$$

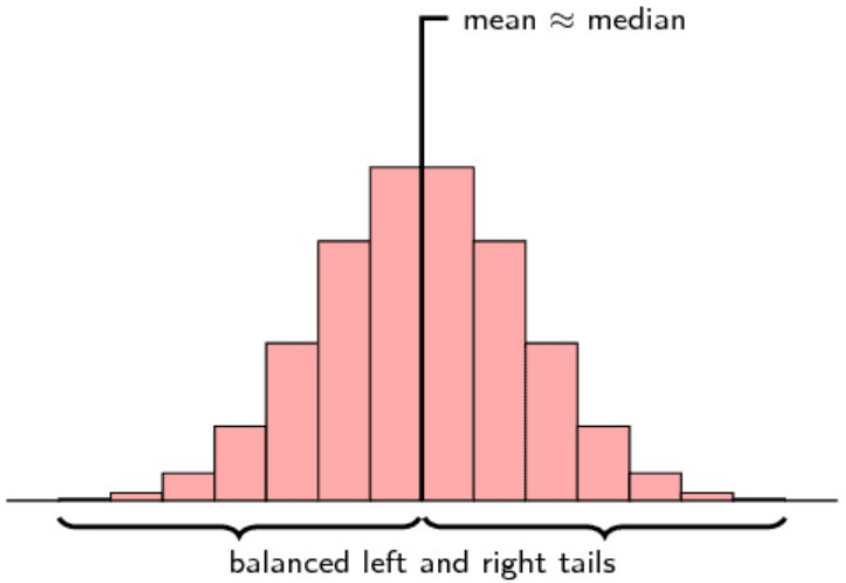
$\overline{Y}$ = mean

Y = case's score on the variable Y

N = number of cases

Getting Ready for Ch3 LA: Skewness

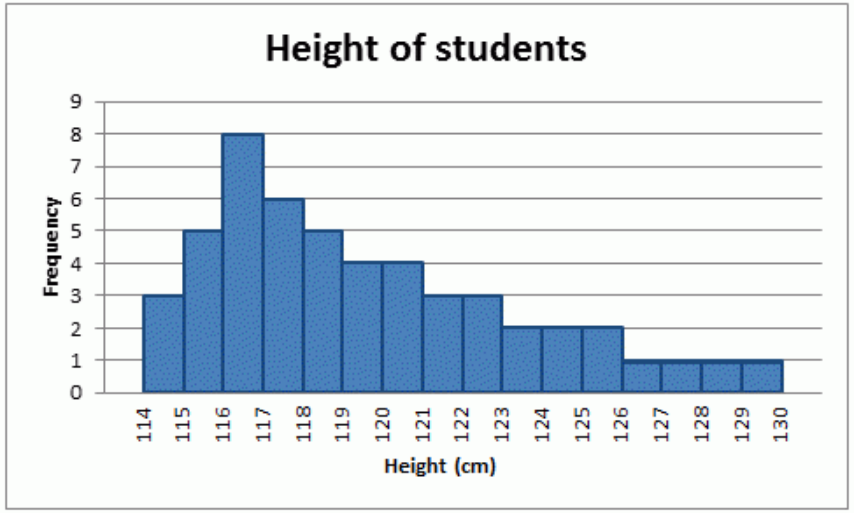
- **Symmetrical distribution:** the frequencies at the right and the left tails of the distribution are identical such that each half of the distribution is *the mirror image of the other*



Mode = Median = Mean

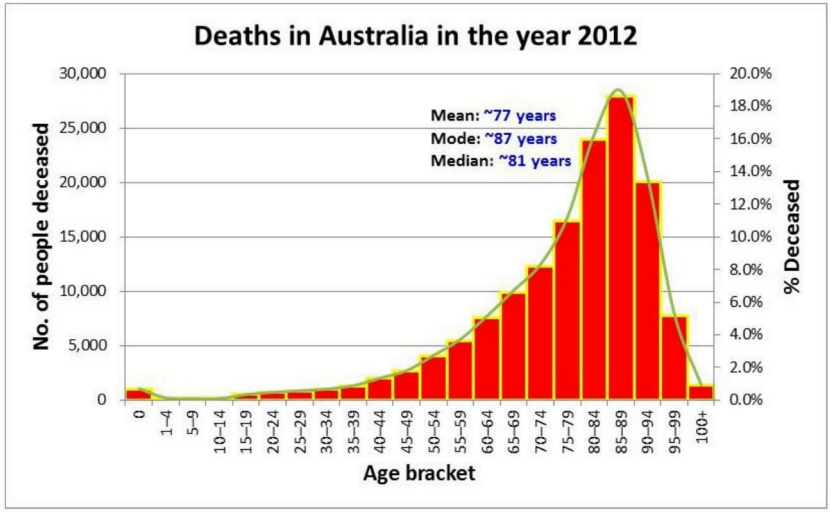
The shape of distribution

Positively Skewed



A few **extremely large** values in the dataset:
Mode < Median < Mean

Negatively Skewed

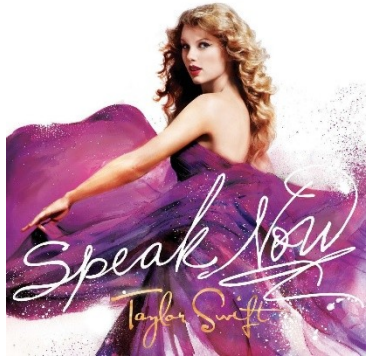


A few **extremely small** values in this dataset:
Mean < Median < Mode

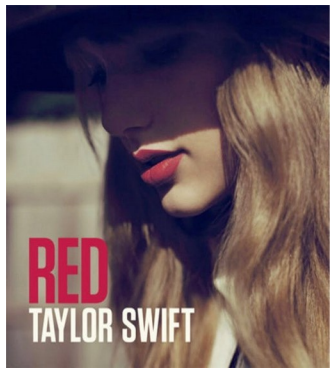
Central tendency - mean, median and mode

Let's recap with the example: TS's age and her album release since age 21

Age: 21 | Age: 23 | Age: 25 | Age: 28 | Age: 30 | Age: 31 | Age: 32

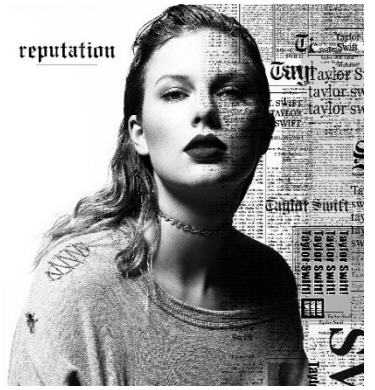


Speak Now (2010)



Red (2012)

1989 (2014)



Reputation (2017)

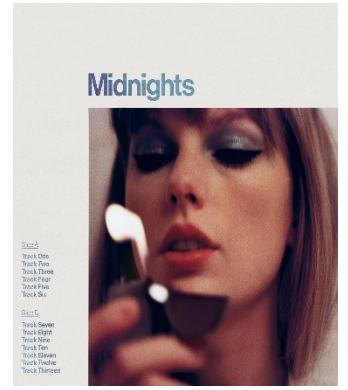
Lover (2019)

Folklore (2020)

Evermore (2020)

Fearless (Taylor's version, 2021)

Red (Taylor's version, 2021)



Midnight (2022)

Source: https://www.lofficielusa.com/pop-culture/every-taylor-swift-album-era-style

TS’s age and album releases (N = 10, even)

- Mode (the most frequent score)

21 23 25 28 30 **31 31 31 31** 32

The mode of TS’s album release age is 31.

- Median (the middle point)

21 23 25 28 **30 31** 31 31 31 32

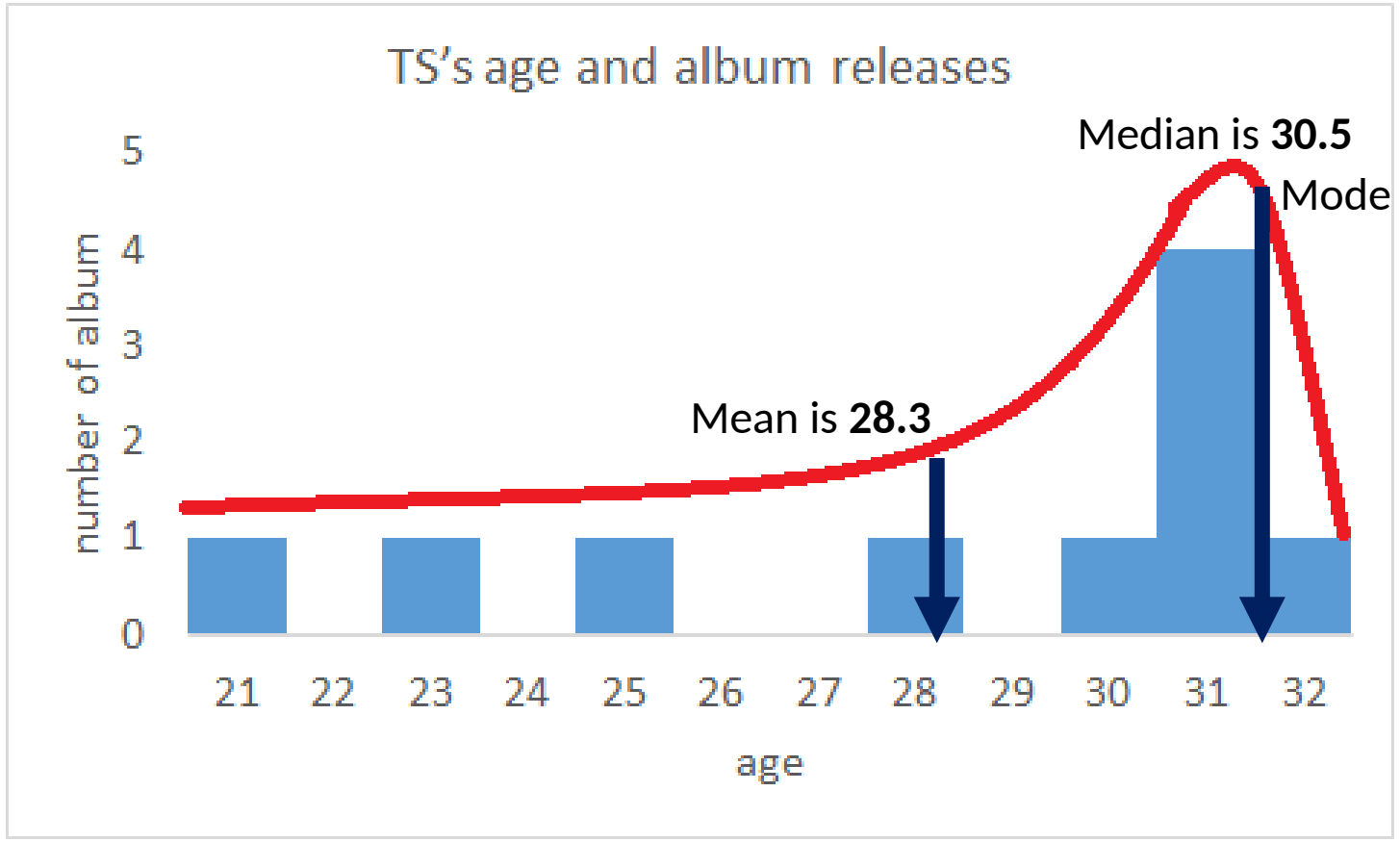
The location of median: (10+1)/2 = 5.5
The median: (30+31)/2 = 30.5

- Mean (Average)

21 23 25 28 30 31 31 31 31 32

(21+23+25+28+30+31+31+31+31+32)/10 = 28.3

The Shape of Distribution

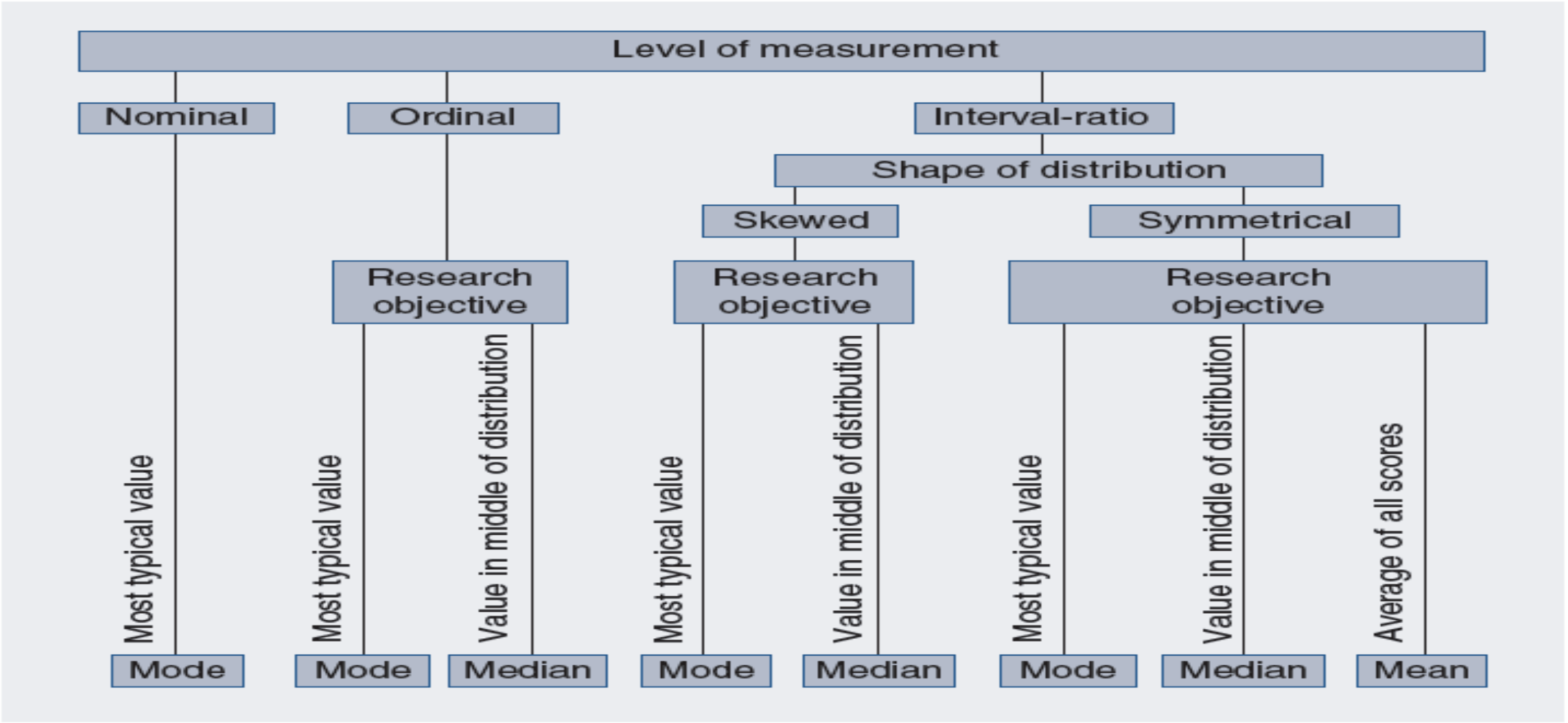


mean < median < mode:

Negatively skewed distribution

Major Takeaway

Figure 3.11 How to Choose a Measure of Central Tendency



Ch3 Learning assessment

- The quiz comprises **14 questions** covering measures of central tendency and the shape of distribution.
- If you have any questions don't hesitate to ask me.
- Next week: Ch4. Measures of Variability.
- Ch4. RQ is due **Next Tuesday 9:30 AM**.
- Office hours: **Tues/ Thursday 12:00 – 1:30 PM, S408**