# Central Tendency, Modality, Skewness

Lecture 3

Emma Ning, M.A.

# Last Week...

**Samples** vs. **Populations**

**Levels of Measurement**
(nominal, ordinal, interval, ratio)

**Frequency Tables**
(absolute vs. relative frequency)

**Data Visualizations**
(bar charts, pie charts, histograms, line graphs, box plots)

# With last week's knowledge...

Imagine you are considering a job at a company.

 A friend who works there says:
"The annual salaries here range from $40K to $2 million."

How helpful is that? Do you feel like you understand what you might get paid? What about seeing a histogram?

 Your other friend says:
"Most people here make around $70K."

That feels different, right? Why is that more useful?

# TODAY'S PLAN

01

## Shape & Modality

What words are typically used to describe these?

02

## Central Tendency

Hey! We are glorifying the meaning of "average"!

03

## Picking the Best Summary for the Data

Which one(s) from Section 01 and 02 should I use?

04

## Wrap Up

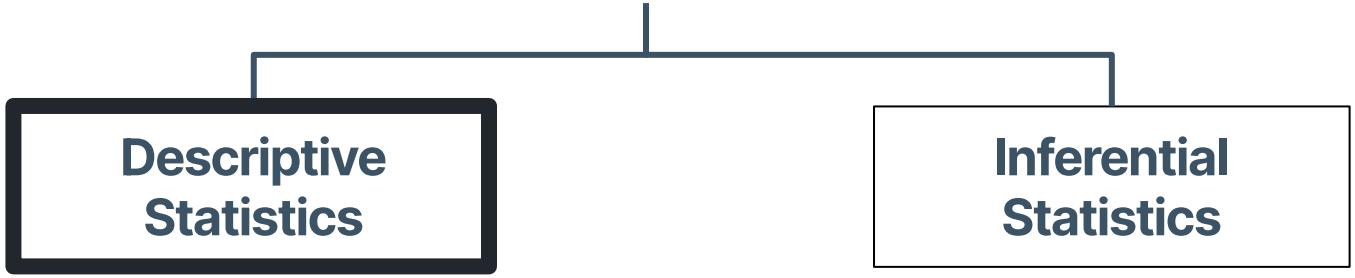
Review + Reminders

# Learning objectives

- Calculate and interpret the three measures of **central tendency** (mean, median, mode).
- Interpret the **modality** of a given distribution or graph (unimodal, bimodal, multimodal).
- Identify the **shape** (symmetrical, positively skewed, or negatively skewed) of a distribution.
- Explain how the three measures of central tendency — mean, median, and mode — are **related** to each other for symmetrical and skewed distributions.


# Shape & Modality

# Statistics

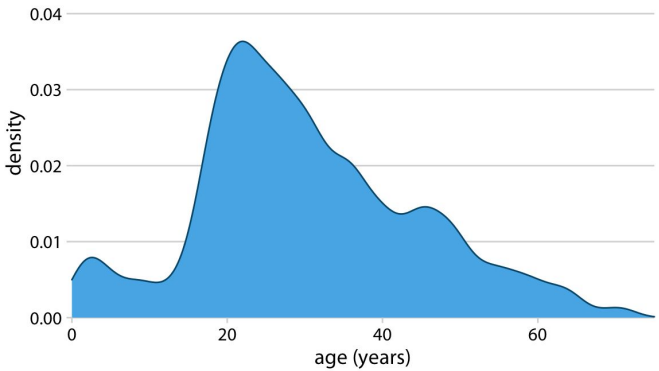



used to **summarize** and **describe** data

techniques used to **make generalizations** about samples and apply them to populations

# Descriptive Statistics

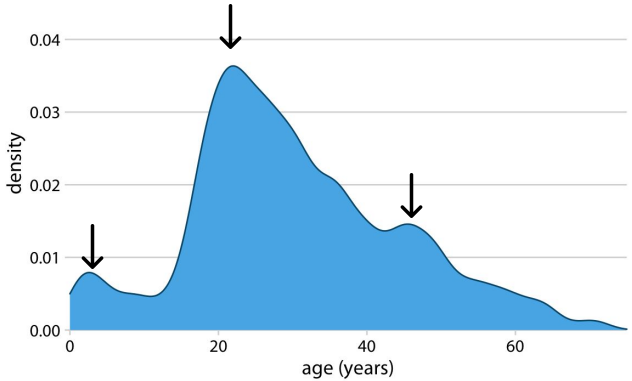
Remember this density plot from the last week?




Think of **descriptive statistics** this way:

How would you describe the overall shape of this data to a friend—without showing them a graph—and you're in a hurry?

# Shape & Modality, in a nutshell




A **distribution** shows us how data are spread out—what values are common, and which ones are rare.
Imagine tossing a handful of sand onto the table—each grain is a data point. Where the sand piles up, you have more data; where it's sparse, fewer. That pattern is your distribution.

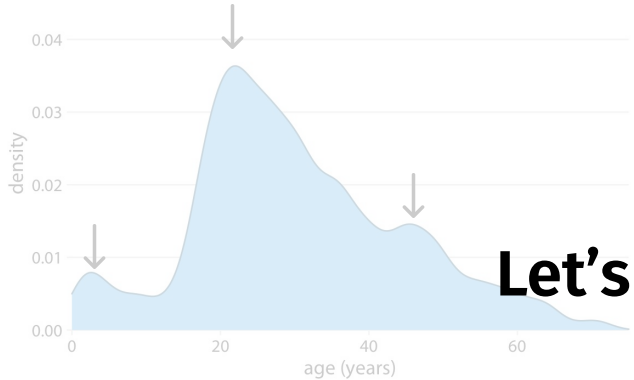
Shape & Modality refer to two aspects:

1. Is it symmetrical? → **symmetry/shape**
2. How many "peaks" does it have? → **modality**

Here, the distribution is not symmetrical, therefore it is **skewed**.

The distribution also seems to be **multimodal**.

# Shape & Modality, in a nutshell




A **distribution** shows us how data are spread out—what values are common, and which ones are rare.
Imagine tossing a handful of sand onto the table—each grain is a data point. Where the sand piles up, you have more data; where it's sparse, fewer. That pattern is your distribution.

Shape & Modality refer to two aspects:

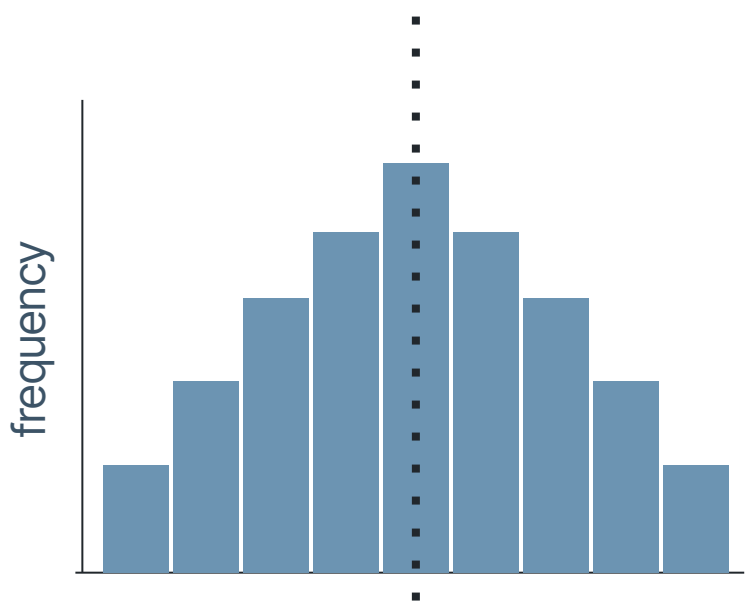
1. Is it symmetrical? → **symmetry**
2. How many "peaks" does it have? → **modality**

**Let's see more examples!**

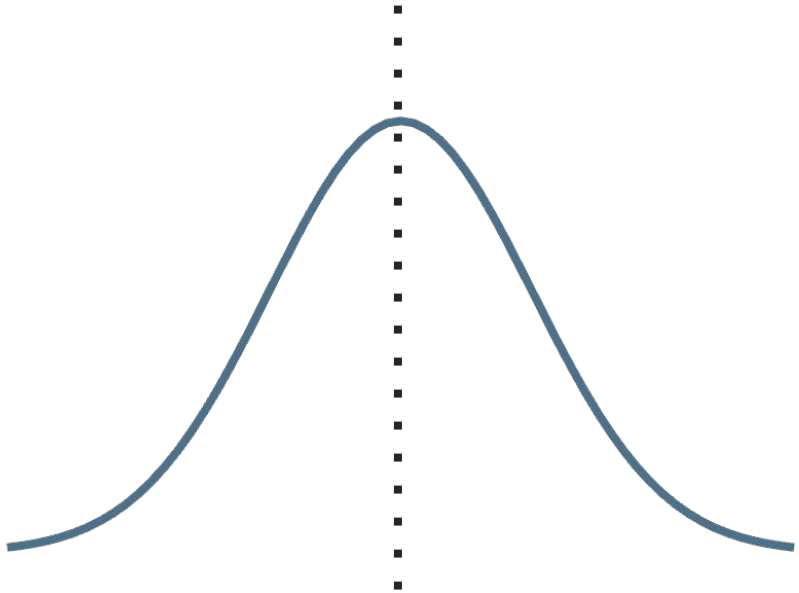
Here, the distribution is not symmetrical, therefore it is **skewed**.

The distribution also seems to be **multimodal**.

# **First, Skewness** (symmetry)



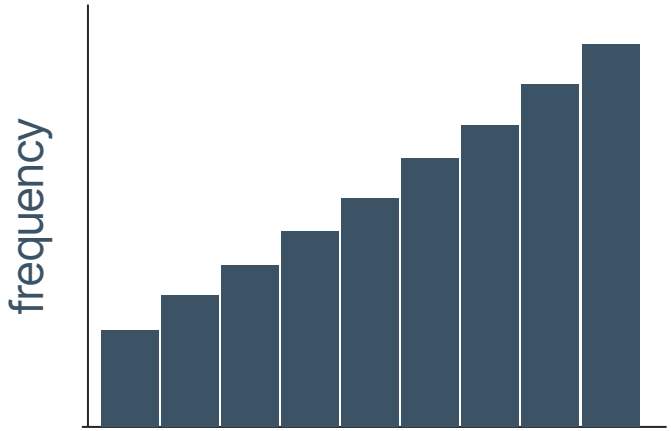

Symmetrical



Normal Curve

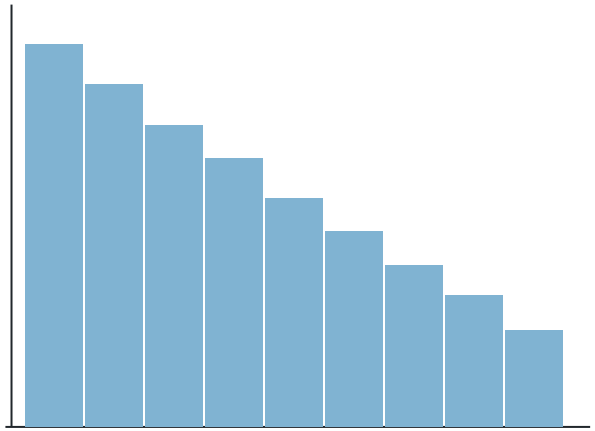
"they are **symmetrical/normal**"
(non-skewed is not wrong, but not what scientists say)

# Skewness continued




**Negatively** Skewed

"skewed to the **left**"



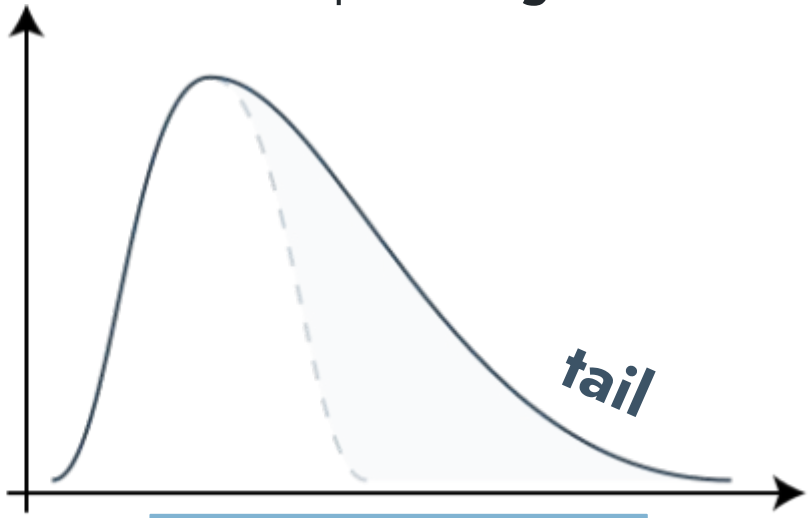
**Positively** Skewed

"skewed to the **right**"

# TIP: FOLLOW THE TAIL!

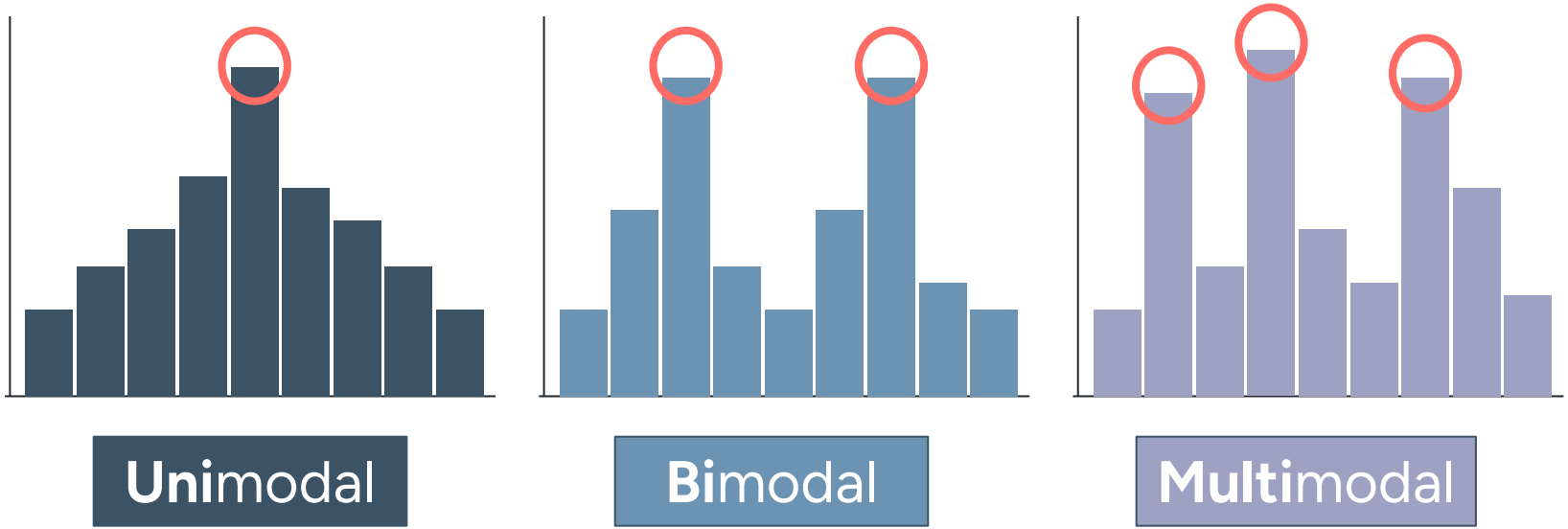
A good way to remember which skewed distribution is which is to follow the tail. The tail of a **negatively skewed** distribution points **left**; the tail of the **positively skewed** distribution points **right**.


**Negatively** Skewed/
**Left** Skewed




**Positively** Skewed/
**Right** Skewed

# Next, Modality

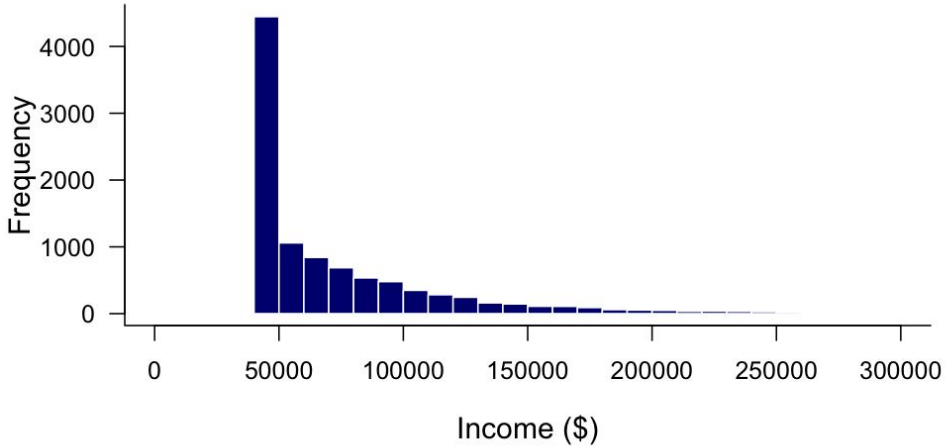



Note: Modality can be quite subjective in real data

# Let's go back to our first example

"The salaries here range from $40K to $2 million."

"Most people here make around $70K."




The income distribution at this company is **unimodal**, **right-skewed** (aka **positively skewed**).

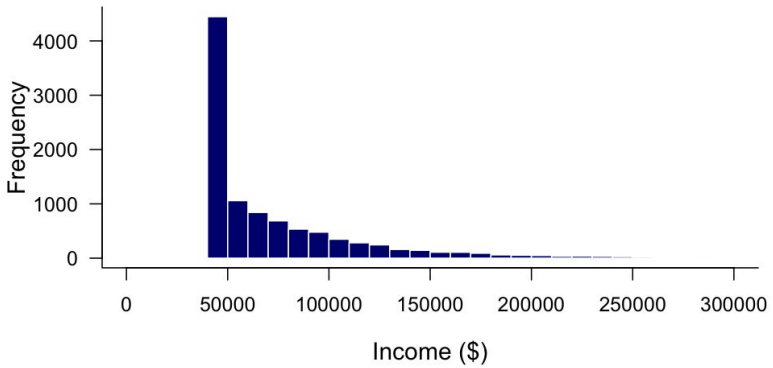

# Central Tendency

# 3 Types of Central Tendency

**1** **Mean**

**2** **Median**

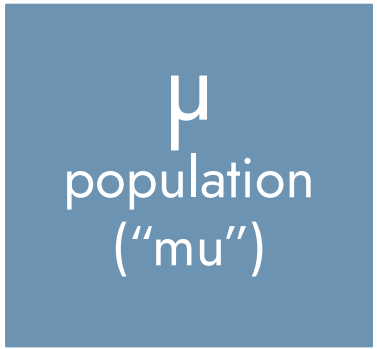
**3** **Mode**

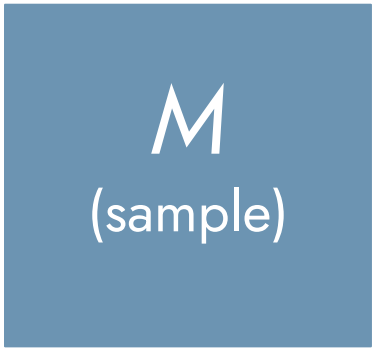



Think of central tendency this way: If you know the overall shape of a distribution—its bumps, peaks, or symmetry—how do you figure out where it balances on the x-axis? That balance point is the intuition behind central tendency.

# The **Mean**

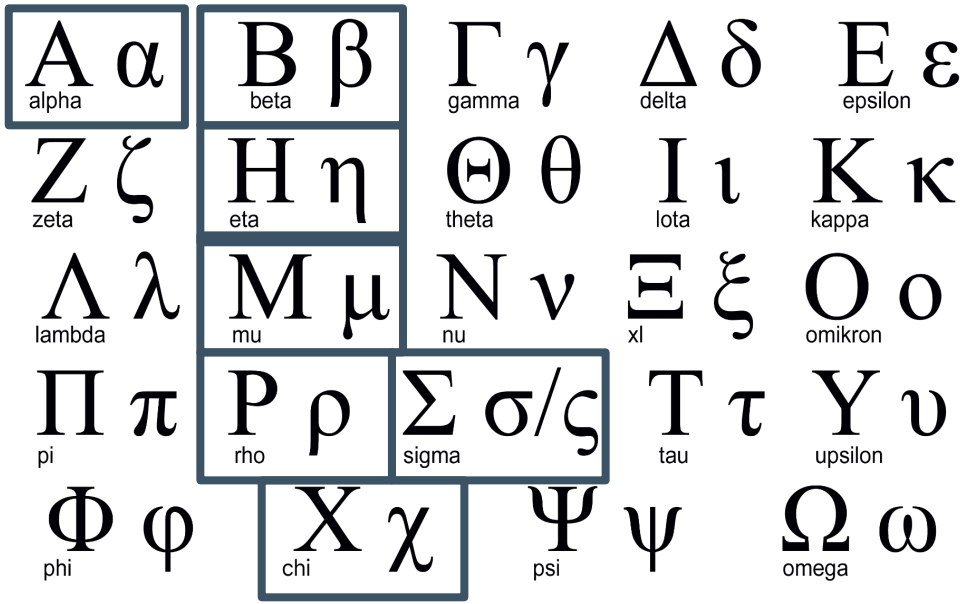
- The "**average**" of a set of scores
- A very commonly used measure of central tendency
- The mean requires at least **interval/ratio data** (you cannot compute it for either nominal or ordinal data).



# QUICK TIP

In statistics, **Greek letters** represent **population parameters**; **roman letters** represent **sample statistics**.



# QUICK TIP

In statistics, **Greek letters** represent **population parameters**; **roman letters** represent **sample statistics**.

| Meaning | Population | Sample |
|---|---|---|
| Mean | μ | *M* |
| Standard Deviation | σ | *s* or *sd* |

# Mean Formula *Conceptually*

$$M = \frac{\text{add up all scores}}{\text{number of scores}}$$

# Unpacking the Mean Formula

(sample)

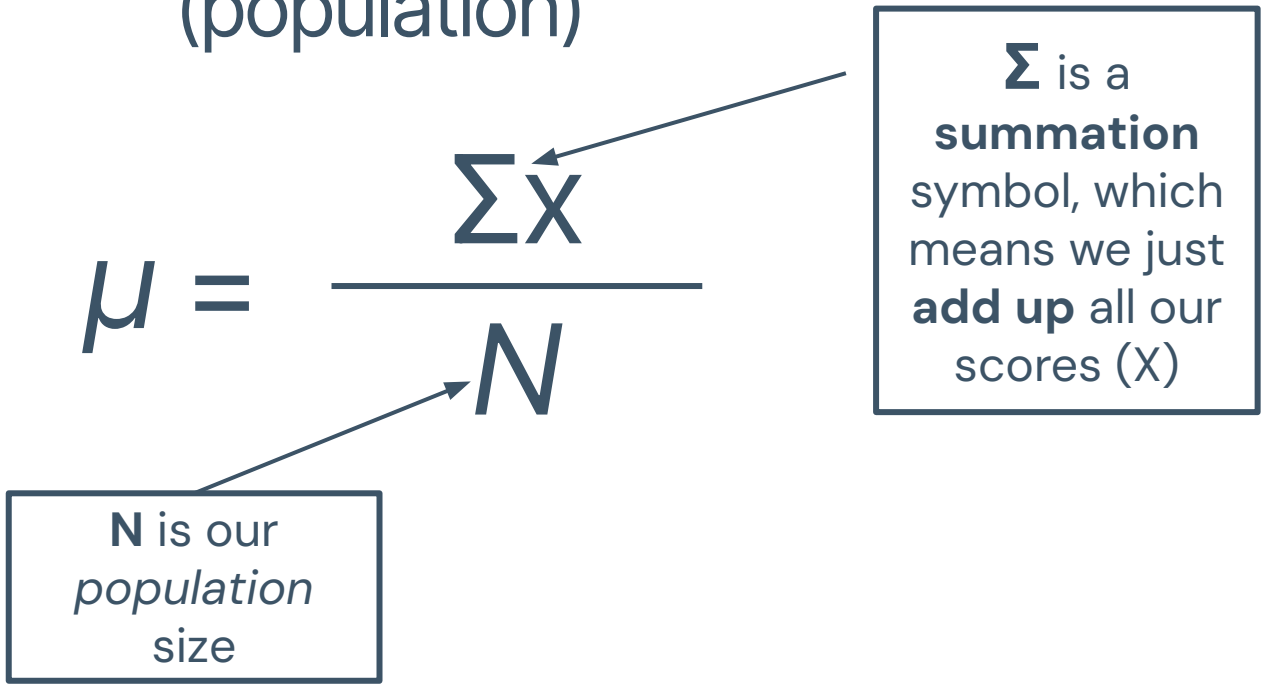
$$M = \frac{\Sigma x}{n}$$

**Σ** is a **summation** symbol, which means we just **add up** all our scores (X)

**n** is our *sample* size

# Unpacking the Mean Formula

(population)

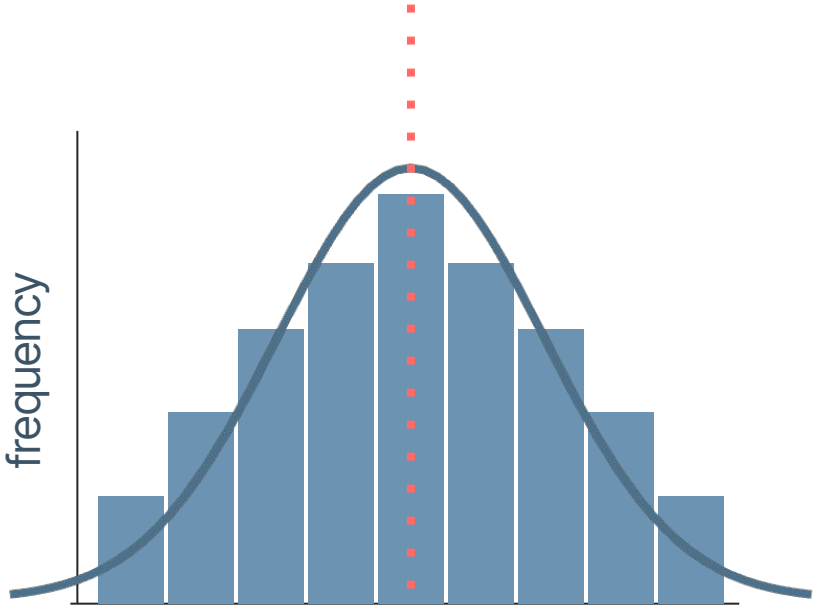


# Quick Example

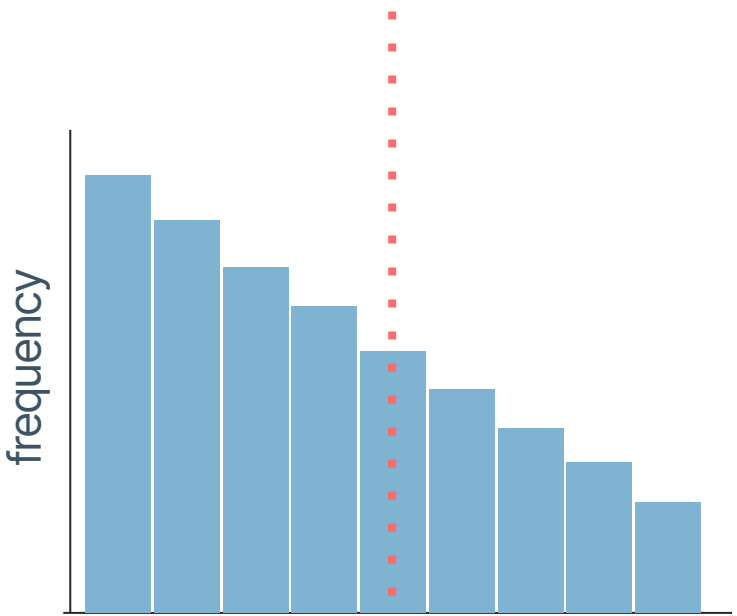
4 1 6 5

$$M = \frac{\Sigma x}{n} = \frac{4 + 1 + 6 + 5}{4} = \mathbf{4}$$

# Mean - Imagining the “balancing point”




Normal Curve

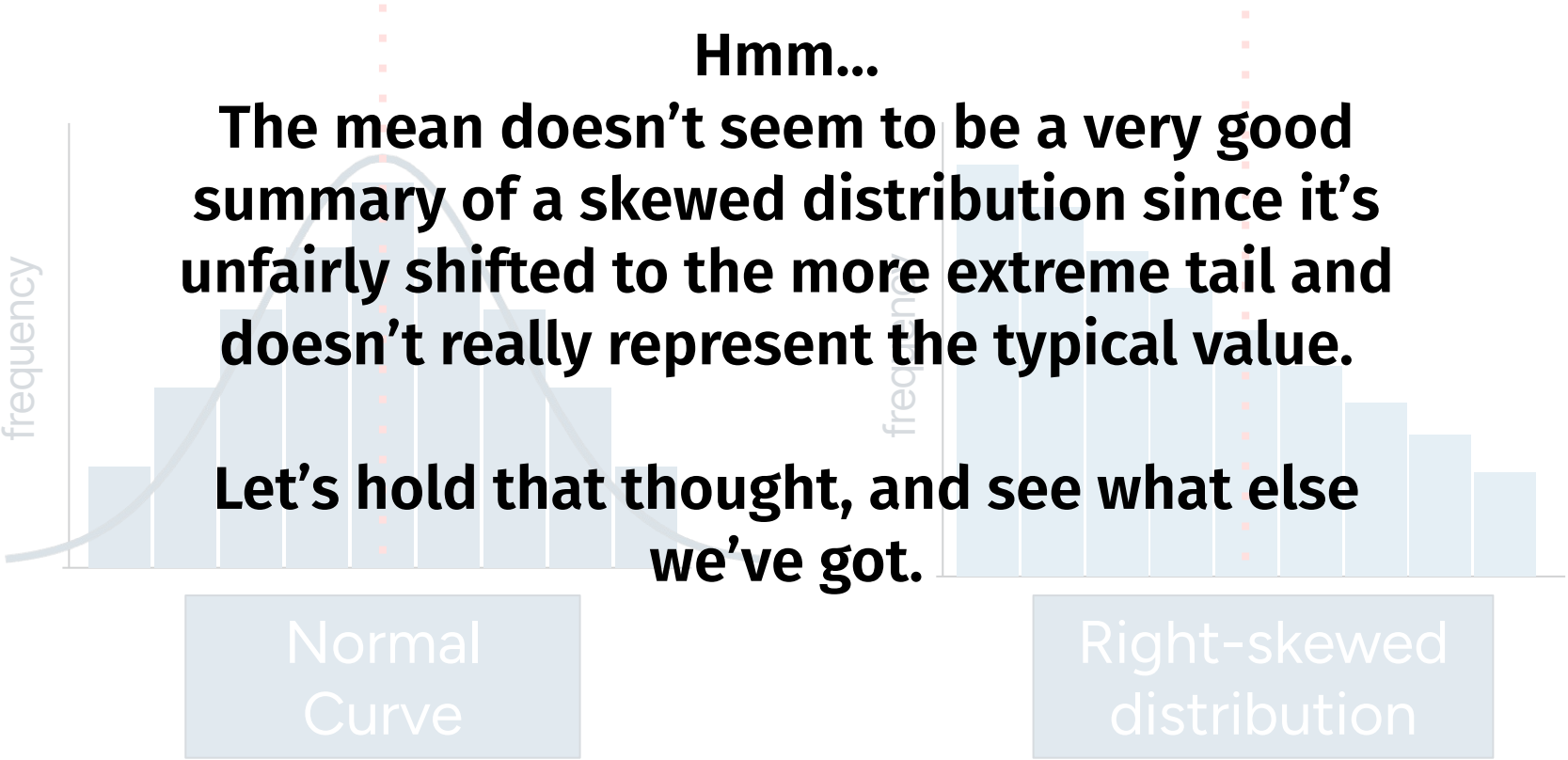



Right-skewed distribution

# Mean - Imagining the “balancing point”

**Hmm...**
**The mean doesn’t seem to be a very good summary of a skewed distribution since it’s unfairly shifted to the more extreme tail and doesn’t really represent the typical value.**

**Let’s hold that thought, and see what else we’ve got.**



# The **Median**

- The score that **divides the amount of data exactly in half** – that is, half of the scores are below the median and half are above the median.
- It is the precise midpoint (also referred to as the **50th percentile**).
- The median requires **at least ordinal data** (you cannot compute it for nominal data).

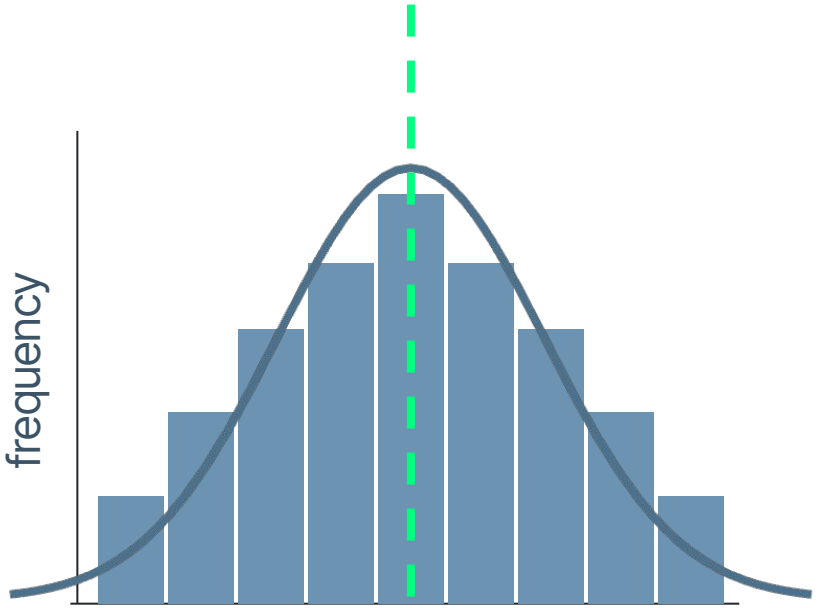
Mdn

# **Median** Calculation

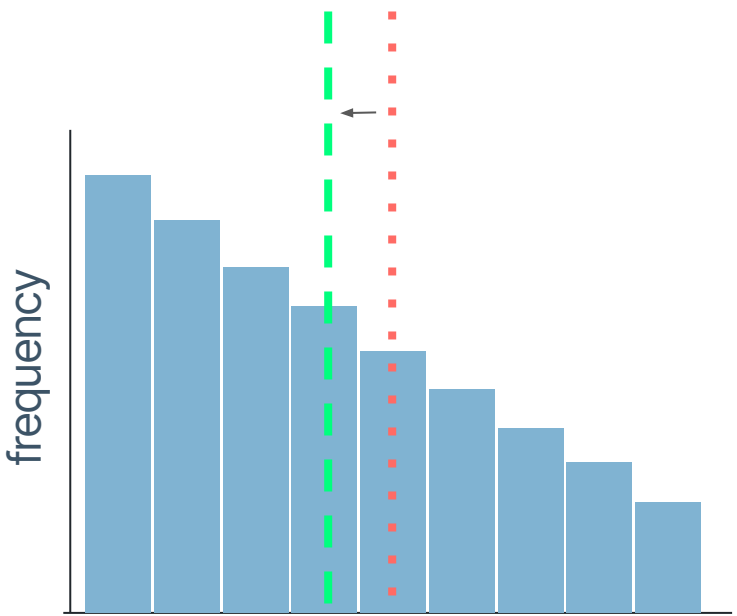
The computation of the median depends on whether there are an **odd** number or **even** number of observations (N).

| | **ODD** | **EVEN** |
|---|---|---|
| | 6, 7, 1, 2, 5, 8, 8 | 1, 3, 7, 4, 6, 1 |
| |  |  |
| **Rank it** | 1, 2, 5, 6, 7, 8, 8 | 1, 1, 3, 4, 6, 7 |
| **Find the middle number** | ~~1, 2, 5,~~ 6, ~~7, 8, 8~~ | ~~1, 1,~~ 3, 4, ~~6, 7~~ |
| | | 3.5 |
| | | (<u>average</u> of two middle numbers) |

# Median - A different “balancing point”




Normal Curve




Right-skewed distribution

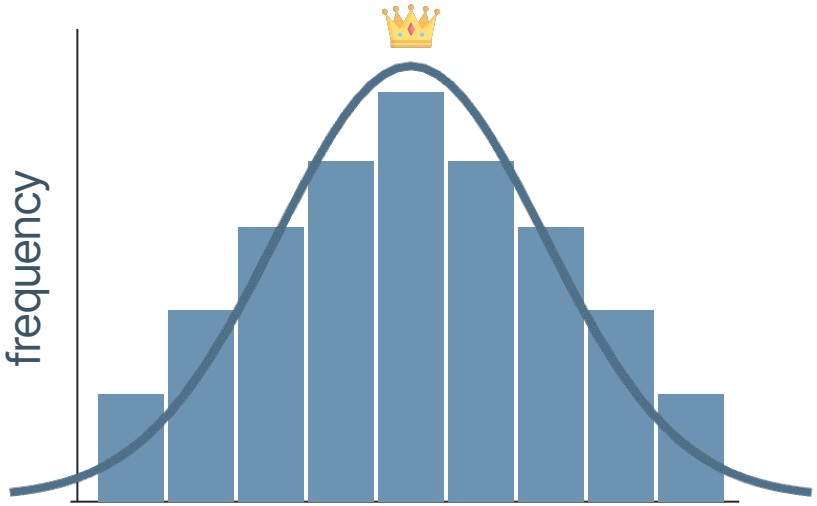
# The Mode

- The mode is the **most frequent score** in the data/distribution.
- It is the "typical" value.
- The mode only requires nominal data, but you can also compute it for ordinal, interval, and ratio (**any** type of data)

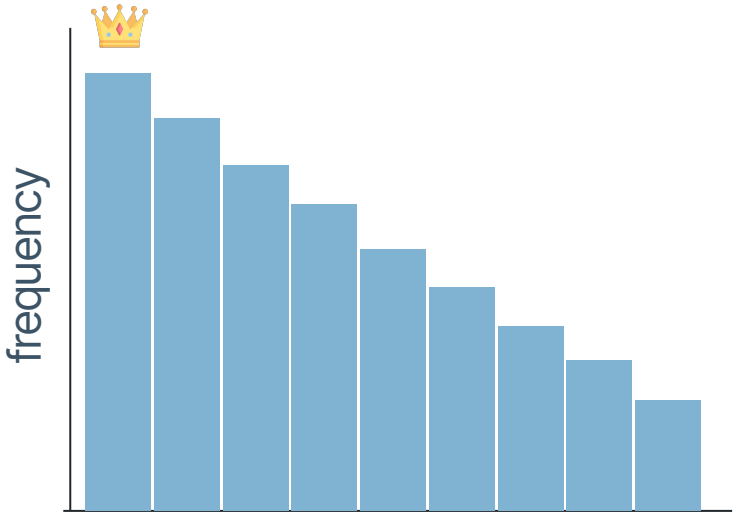
1, 3, 4, 5, 7, 7, 7, 9, 10, 10

1, 3, 4, 5, **7, 7, 7**, 9, 10, 10

# Mode - The populist



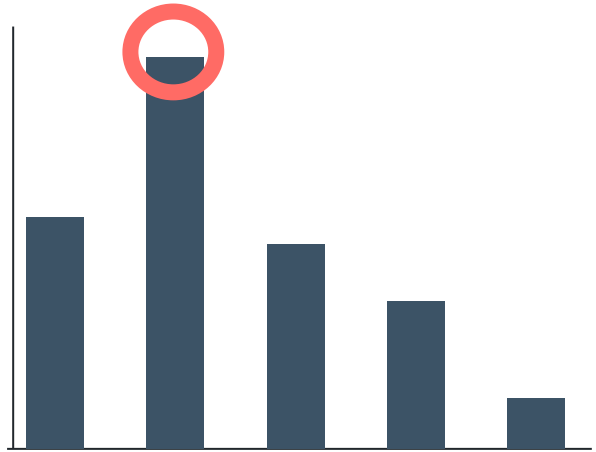

Normal Curve

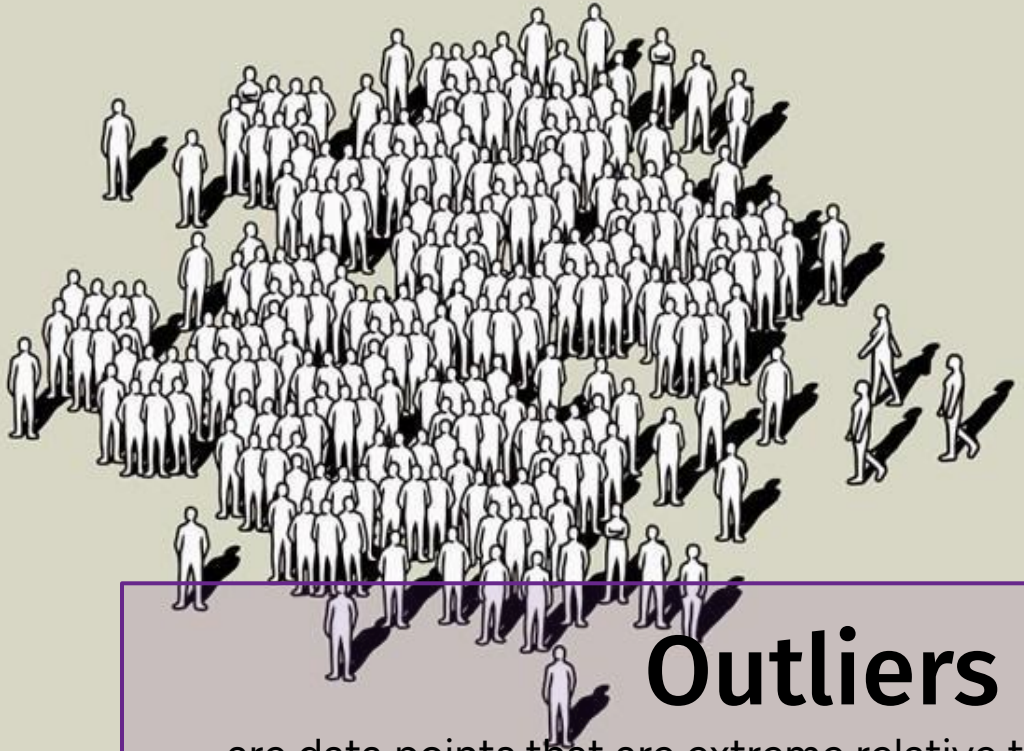



Right-skewed distribution

# The mode can be easily spotted with a **frequency table** or **bar graph/histogram.**

| Diagnosis | F | P |
|---|---|---|
| ADHD | 9 | 22.5% |
| Major Depression | 15 | 37.5% |
| PTSD | 8 | 20% |
| Generalized Anxiety | 6 | 15% |
| Schizophrenia | 2 | 5% |
| **N** | 40 | 100% |





# Outliers

are data points that are extreme relative to others in a dataset

# **TRUE** or **FALSE?**

| | | |
|---|---|---|
| 1 | The mean is the most commonly used measure of central tendency in psychological statistics. | |
| 2 | The median can be determined for ordinal data as well as interval and ratio data. | |
| 3 | For nominal/categorical data, the only measure of central tendency we can use is the mode. | |
| 4 | Greek letters are often used for sample data, whereas Roman letters are often used for populations. | |

# TRUE or FALSE?

| | | |
|---|---|---|
| 1 | The mean is the most commonly used measure of central tendency in psychological statistics. | ✗ |
| 2 | The median can be determined for ordinal data as well as interval and ratio data. | ✓ |
| 3 | For nominal/categorical data, the only measure of central tendency we can use is the mode. | ✓ |
| 4 | Greek letters are often used for sample data, whereas Roman letters are often used for populations. | ✗ |

# Example Descriptive Statistics Table

**Table 1**
*Participant Characteristics*

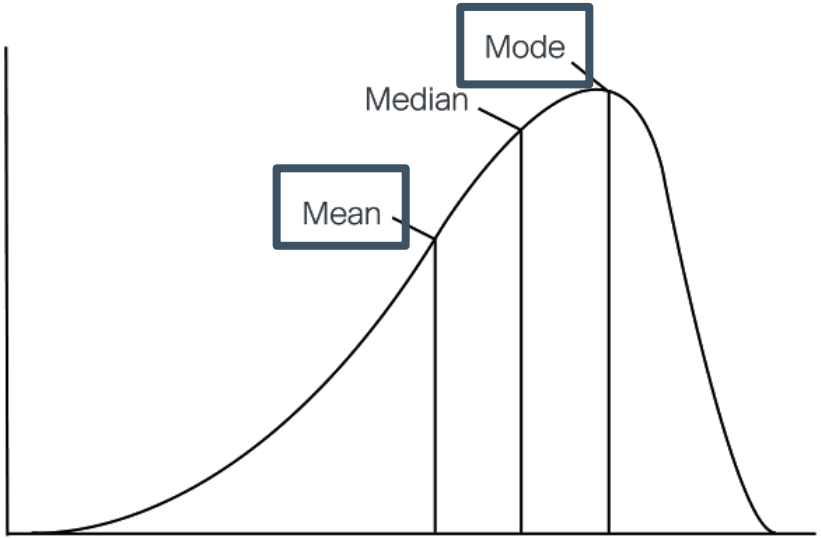
| Variable | Total sample | ACE group | |
|---|---|---|---|
| | | Low | High |
| *N* | 211 | 143 | 68 |
| Age: *M* (*SD*) | 44.09 (17.14) | 45.03 (17.62) | 42.12 (16.03) |
| Education: *M* (*SD*) | 13.82 (2.96) | 13.84 (3.09) | 13.79 (2.69) |
| Sex: *n* (%) | | | |
| Male | 112 (53.1%) | 79 (55.2%) | 33 (48.5%) |
| Female | 99 (46.9%) | 64 (44.8%) | 35 (51.5%) |

"The sample had a **mean age** of **44.09 years** (*SD* = 17.14) and education of **13.82** years (*SD* = 2.96; range: 6–20). Sex distribution was **53.1% male** (*n* = 112) and **46.9%** female (*n* = 99)."

Bing-Canar et al. (2024)


# Picking the Best Summary for the Data

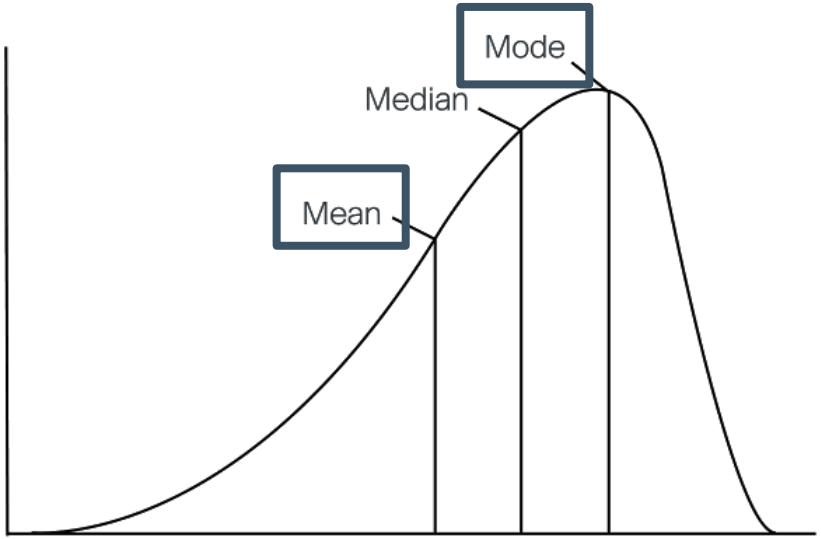
**Skewed distributions** can affect our measures of central tendency.

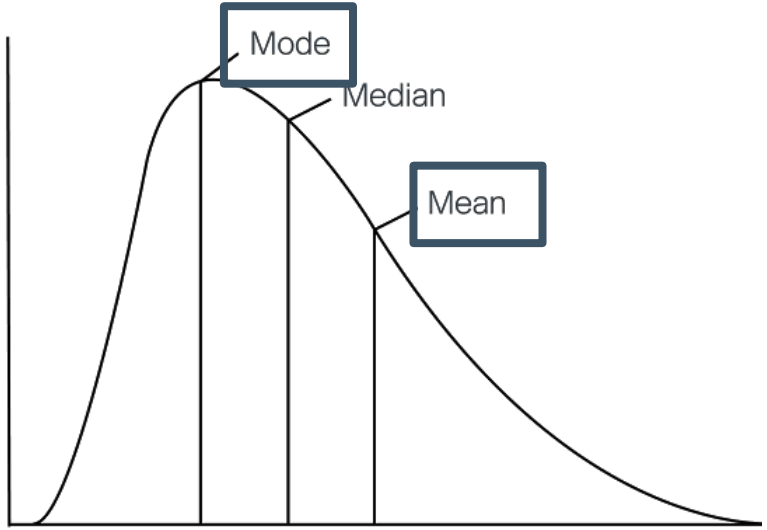



The **mode** will always be at the very top (most common), as it is **unaffected by outliers**.

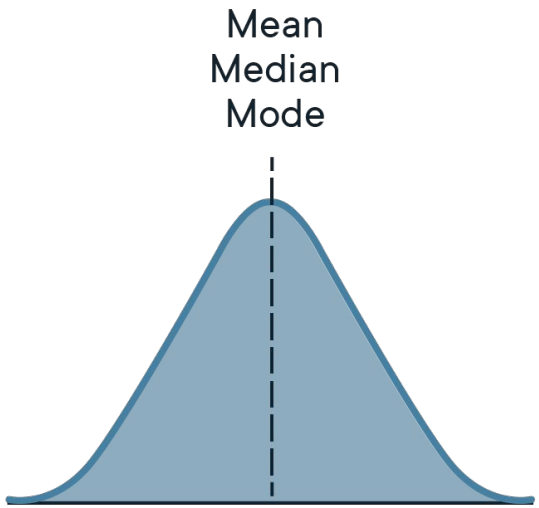
The **mean** will always "follow" the tail, as the **mean is the most affected by outliers**.

**Skewed distributions** can affect our measures of central tendency.



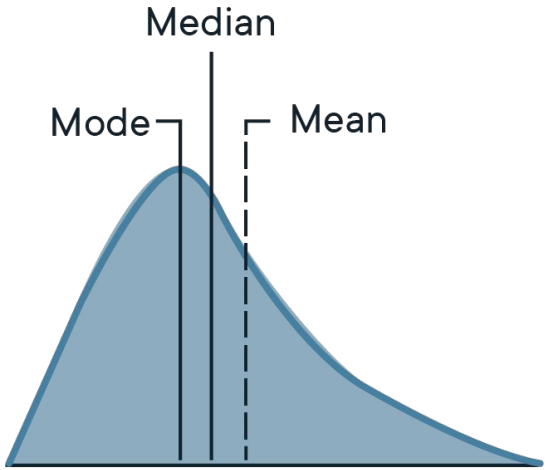



# Summary

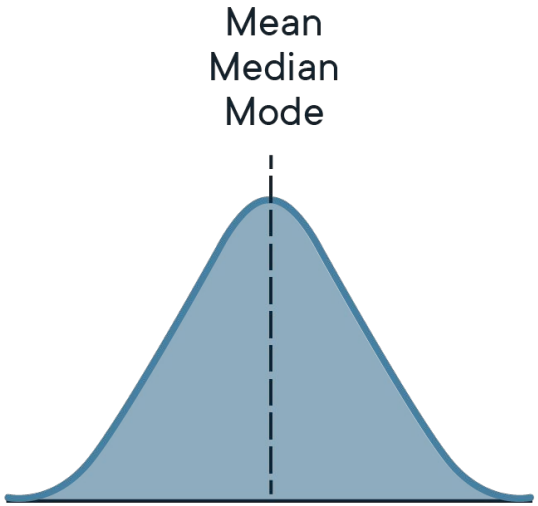



Symmetrical

# Summary

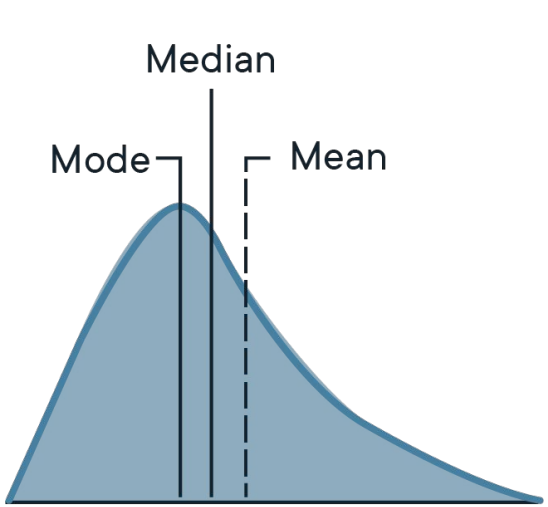



Positive Skew

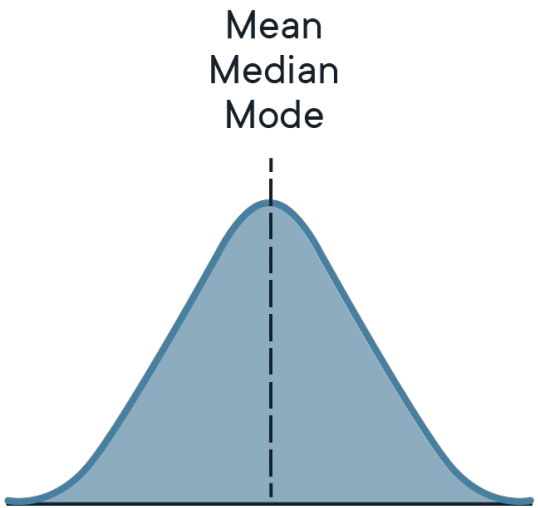



Symmetrical

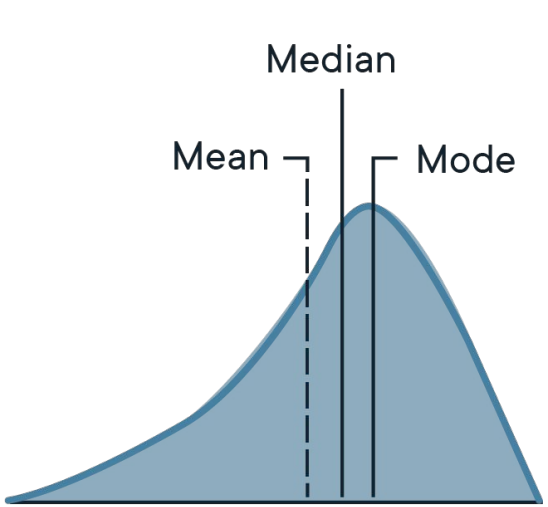
# Summary




Positive Skew




Symmetrical

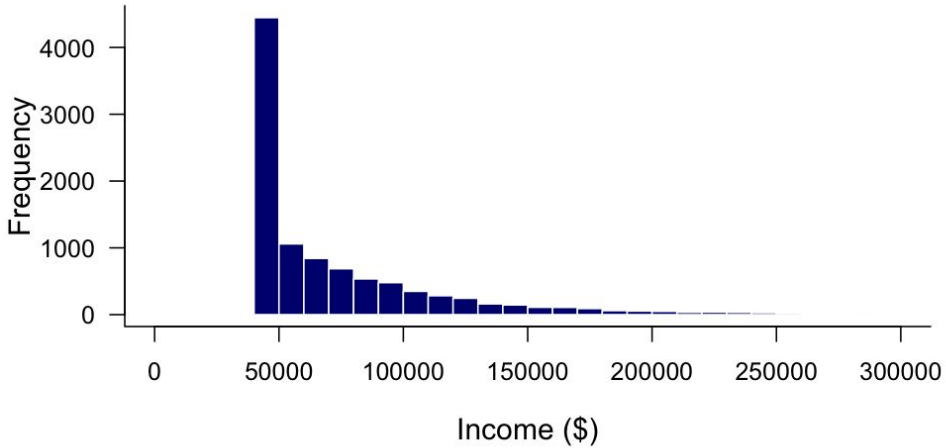



Negative Skew

# First example, again

"The salaries here range from $40K to $2 million."

"Most people here make around $70K."




What is the best measure of central tendency for this distribution if you are the job applicant?

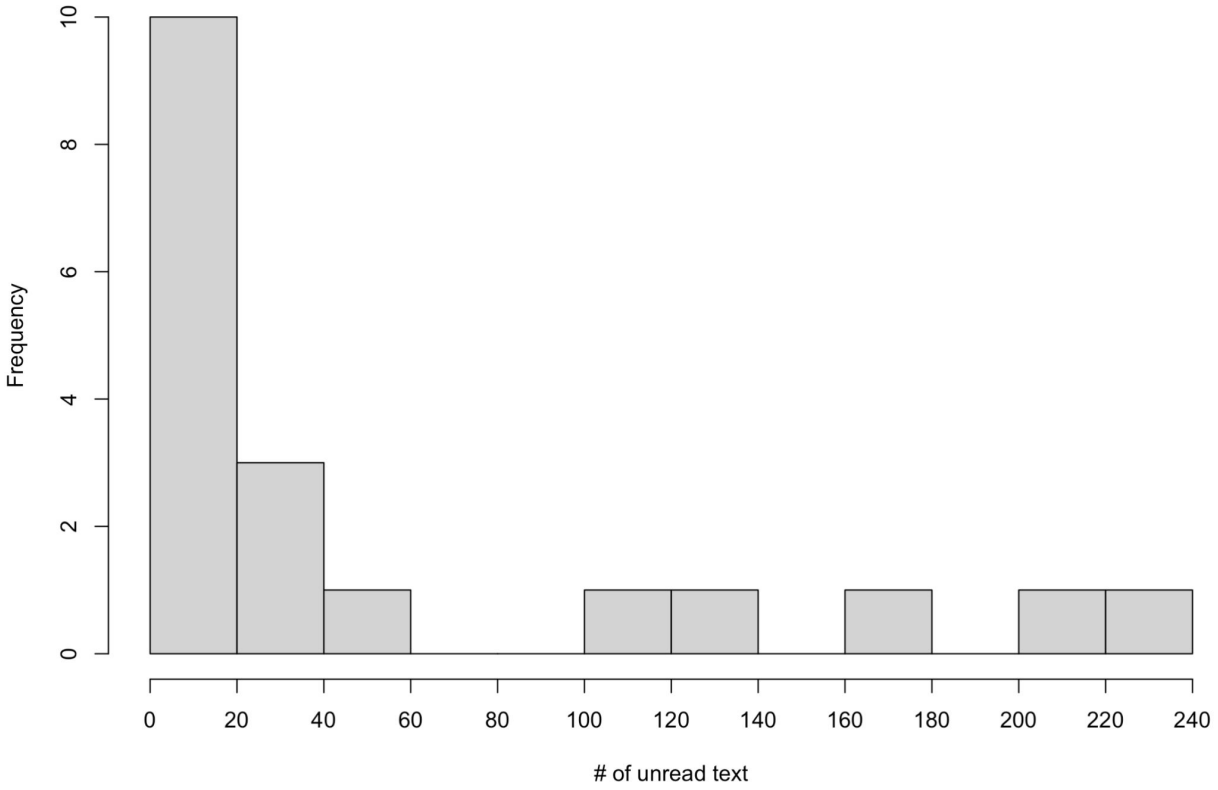
What is the best measure of central tendency for the employers to put in the job ad?

# ICA 3

You will be given a small dataset about your classmates. Work with your group to do the following:

1. Calculate the **mean**, **median**, and **mode**.
2. Create a **histogram**. **Label** the x and y axes.
3. Describe the **shape** and/or **modality** of your distribution.

# # of unread texts, from our class




***M*** = 54.42
***Mdn*** = 17
**Mode** = 1

Unimodal, very right-skewed

I am comfortable working with numbers and data.

1 ○ Strongly agree

2 ○ Agree

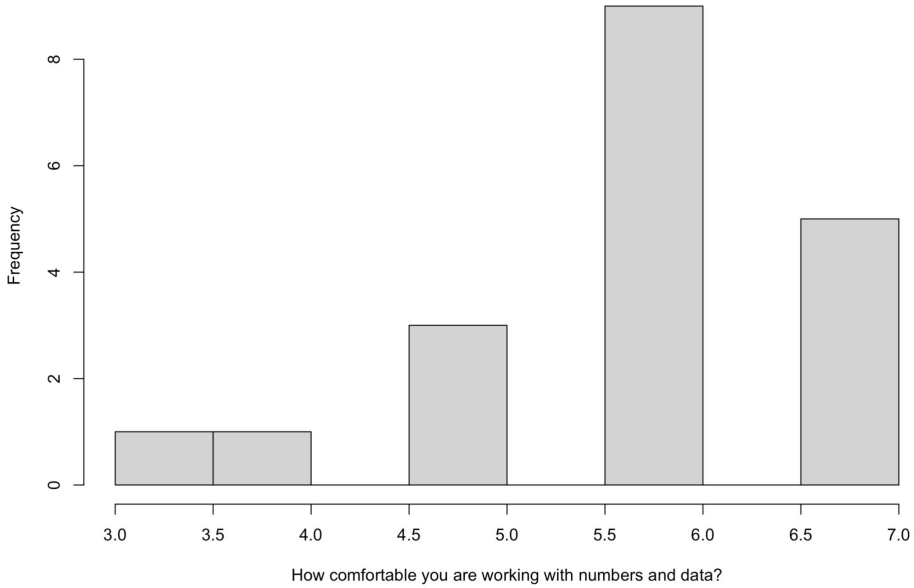
3 ○ Slightly agree

4 ○ Neither agree nor disagree

5 ○ Slightly disagree

6 ○ Disagree

7 ○ Strongly disagree




**Seeing the modality and shape of this distribution, what does this tell me about how I should design my class? Should I worry about math anxiety too much?**


# Wrap Up

# How does this relates to our last class?

| What You Can/Cannot Do | Nominal | Ordinal | Interval | Ratio |
|---|---|---|---|---|
| Count | ✓ | ✓ | ✓ | ✓ |
| Order | ✗ | ✓ | ✓ | ✓ |
| Add; Subtract | ✗ | ✗ | ✓ | ✓ |
| Multiply; Divide; Average | ✗ | ✗ | ✗ | ✓ |

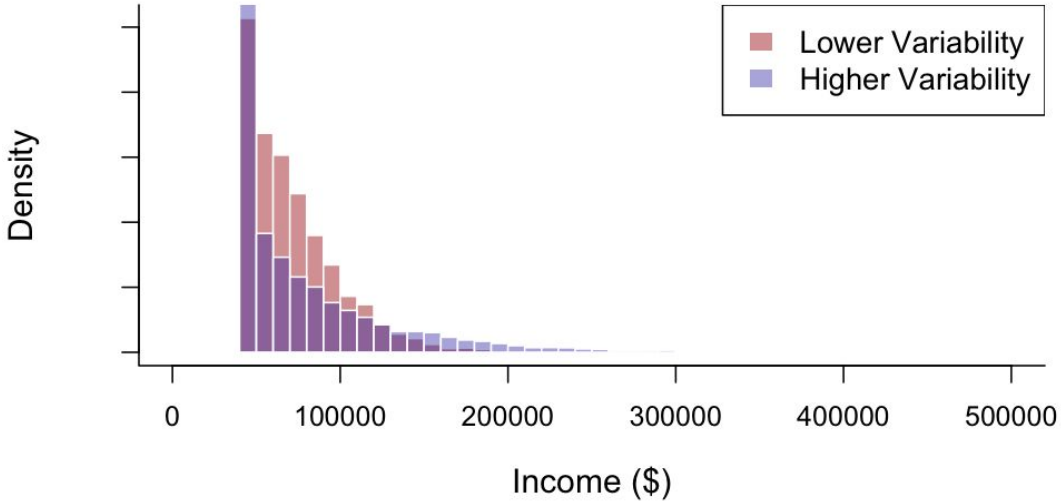
# Which of the things from today can you calculate?

**What You Can/Cannot Do**

| | Nominal | Ordinal | Interval | Ratio |
|---|---|---|---|---|
| Mode | ✓ | ✓ | ✓ | ✓ |
| Median; Percentile | ✗ | ✓ | ✓ | ✓ |
| Mean | ✗ | ✗ | ✓ | ✓ |
| Everything above + ratios | ✗ | ✗ | ✗ | ✓ |
| Modality of distribution | ✓ | ✓ | ✓ | ✓ |

# Being a job applicant again

You are choosing between 2 companies: A  or B 




You don’t know your contract yet, which offer are you leaning toward?