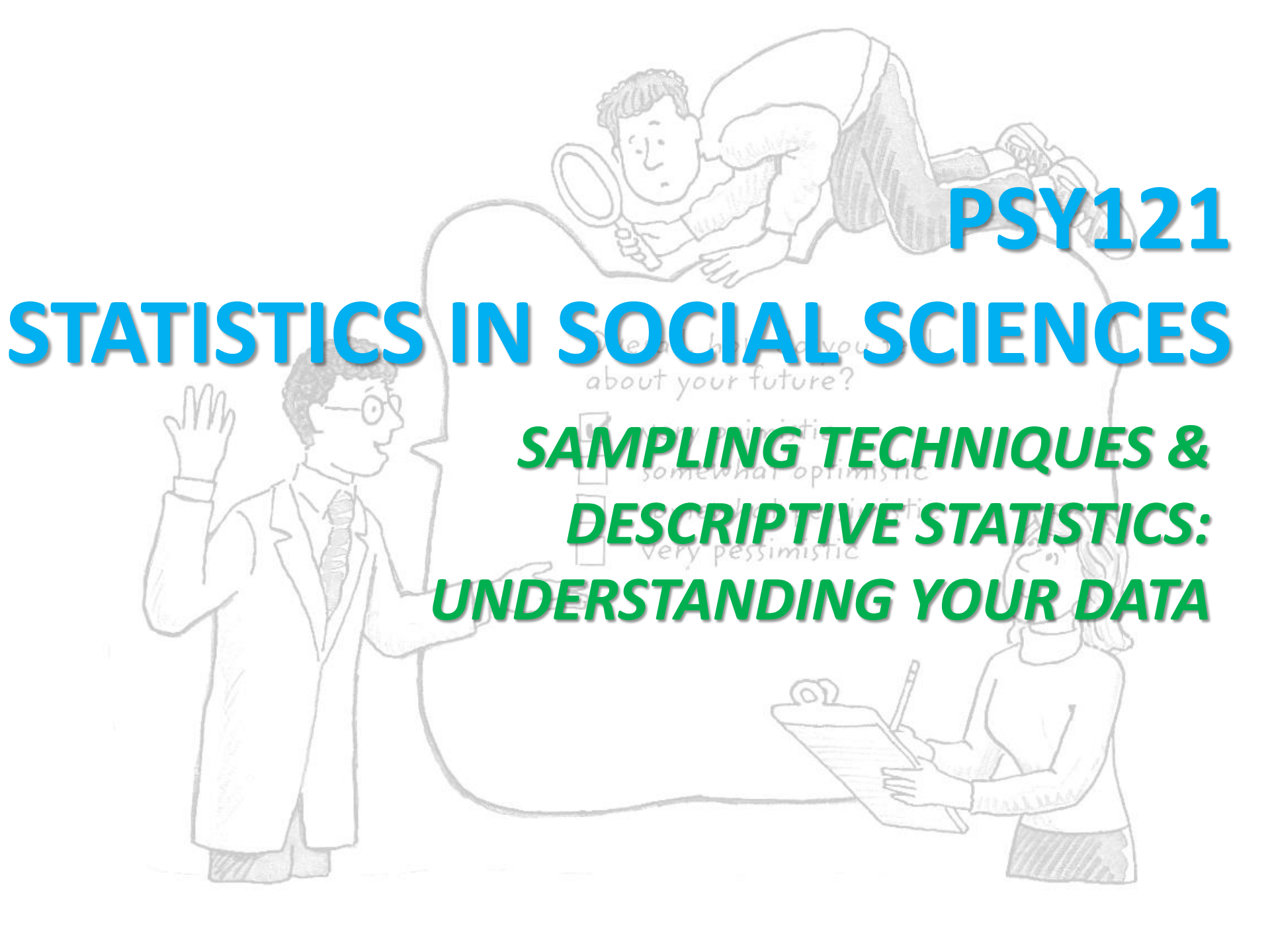




PSY121

STATISTICS IN SOCIAL SCIENCES

How do you feel about your future?

☒ very optimistic
☐ somewhat optimistic
☐ somewhat pessimistic
☐ very pessimistic

***SAMPLING TECHNIQUES &
DESCRIPTIVE STATISTICS:
UNDERSTANDING YOUR DATA***

let's remember...



- Why do researchers (us😊) select a sample instead of studying the entire population?

so many reasons:

1. Studying the entire population is usually **impractical**.
2. Populations can be **very large, spread out, or constantly changing**.
3. Collecting data from everyone would take **too much time, effort, and money**.
4. A **sample** is a smaller, manageable group selected from the population.
5. If the sample is **representative**, it can **accurately reflect** the population.
6. A sample is like a “**small window**” to understand the bigger picture efficiently.



Sampling

- The process of selecting members of the population to be included in the sample.
- Research uses data obtained from a sample to make inferences about the population.

Population

- A defined group of individuals from which a sample is taken
(e.g., married individuals, psychology students etc.)
- The sample must reflect the population!!!
Otherwise, we face the risk of **sampling bias... (örnekleme yanlılığı)** – not reflecting that sample or population!

Sampling

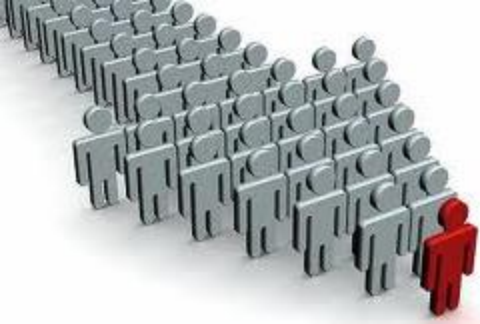


Ideal: Including the entire population... (nüfus sayımı)

- Determining a sample from the population of interest...
Adolescents, adults, prisoners, teachers...
- Good sampling allows us to make inferences about the **entire population**.
- **Statistical Theory**
Sample → population
Statistical significance $p < .05$ - (later on)

Population/Sample (Denek)

- Each unit or individual in the population and the sample
 - **N**: Number of subjects in the population
 - **n**: Number of subjects in the sample
- Although the study is conducted on a sample, find an article that **reports the number of subjects as N...**



Sampling Techniques

- **Probability Sampling**
 - Simple Random Sampling – equal chance of selection
 - Stratified Random Sampling
 - Cluster Sampling
- **Nonprobability Sampling**
 - Haphazard sampling - *Ulaşılabilirlik*
 - Purposive Sampling
 - Quota Sampling
 - Snowball Sampling



Probability Sampling

- You can determine the probability of any member of the population being included in the sample.
- The best way to make inferences about a population...

Probability Sampling

Simple Random Sampling

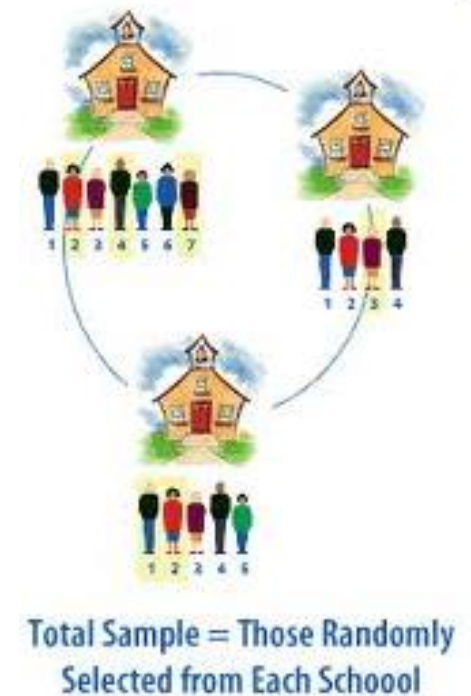
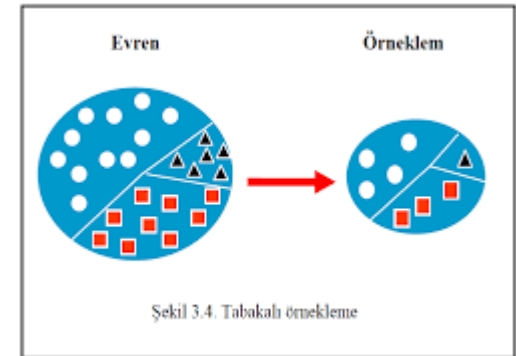
- The probability of sampling all individuals in the population is equal.
- N's are randomly selected using various methods;
e.g., psychology students.
- If you are interested in surveying ÇAĞ UNIVERSITY students, you will need a complete list of all students...

Probability Sampling

Stratified Sampling

Tabakalı Rastgele Örnekleme

- **Dividing the population into subgroups**
e.g., age, gender, education level, political views.
- **Random selection within strata/layers**
e.g., we want to select 9 cubes from a total of 30...
- When categorized by color (blue, green, pink) and taking 3 from each group, we get 9 cubes...

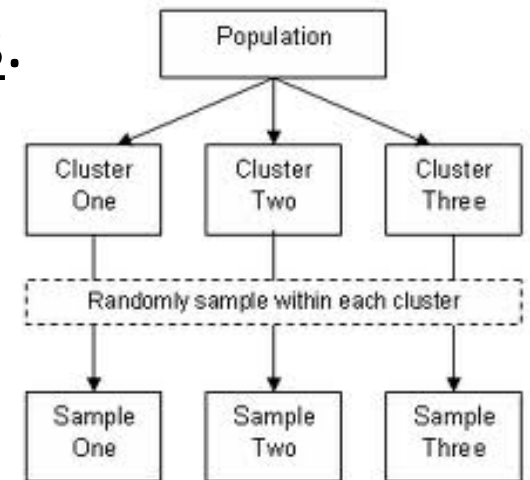


Probability Sampling

Cluster Sampling

Küme Örnekleme

- **Population lists are usually unavailable!!**
- Used when you can't identify individuals.
- Randomly sample clusters of people in identifiable groups.
 - e.g., sample from each ENG102 section.
- Collect data from all people within the sampled clusters.



FARKLARI??

- **Tabakalı Örnekleme**: Alt gruplar (tabakalar) homojendir.
- Aynı tabakada yer alan bireyler benzer özelliklere sahip!!
 - Örneğin, yaş, cinsiyet ya da gelir seviyesine göre tabakalar oluşturulabilir.
- **Küme Örnekleme**: Kümeler heterojendir.
- Her küme popülasyonun küçük bir temsili olarak kabul edilir.
 - Küme içindeki bireyler çeşitli özellikler taşıyabilir.

Nonprobability Sampling

- A type of sampling procedure in which one cannot specify the probability that any member of the population will be included in the sample
- Accidental or convenience sample
- Cheap and convenient
- Introduces biases – big problem when people select themselves to be part of the survey (return a magazine survey, for example)

Nonprobability Sampling

Haphazard Sampling

- Kolayda Örnekleme (**Convenience Sampling**)
- Select Ss “conveniently” – En kolay yol ile
- Population: ÇAĞ UNIVERSITY. Stand in front of the cafétéria between 12:00 and 13:30 and ask people to participate your research.
- Disadvantages
 - Biases → e.g., **you’d exclude students who eat at dorms...**
 - Difficulty in sample → population (generalisation)

Nonprobability Sampling

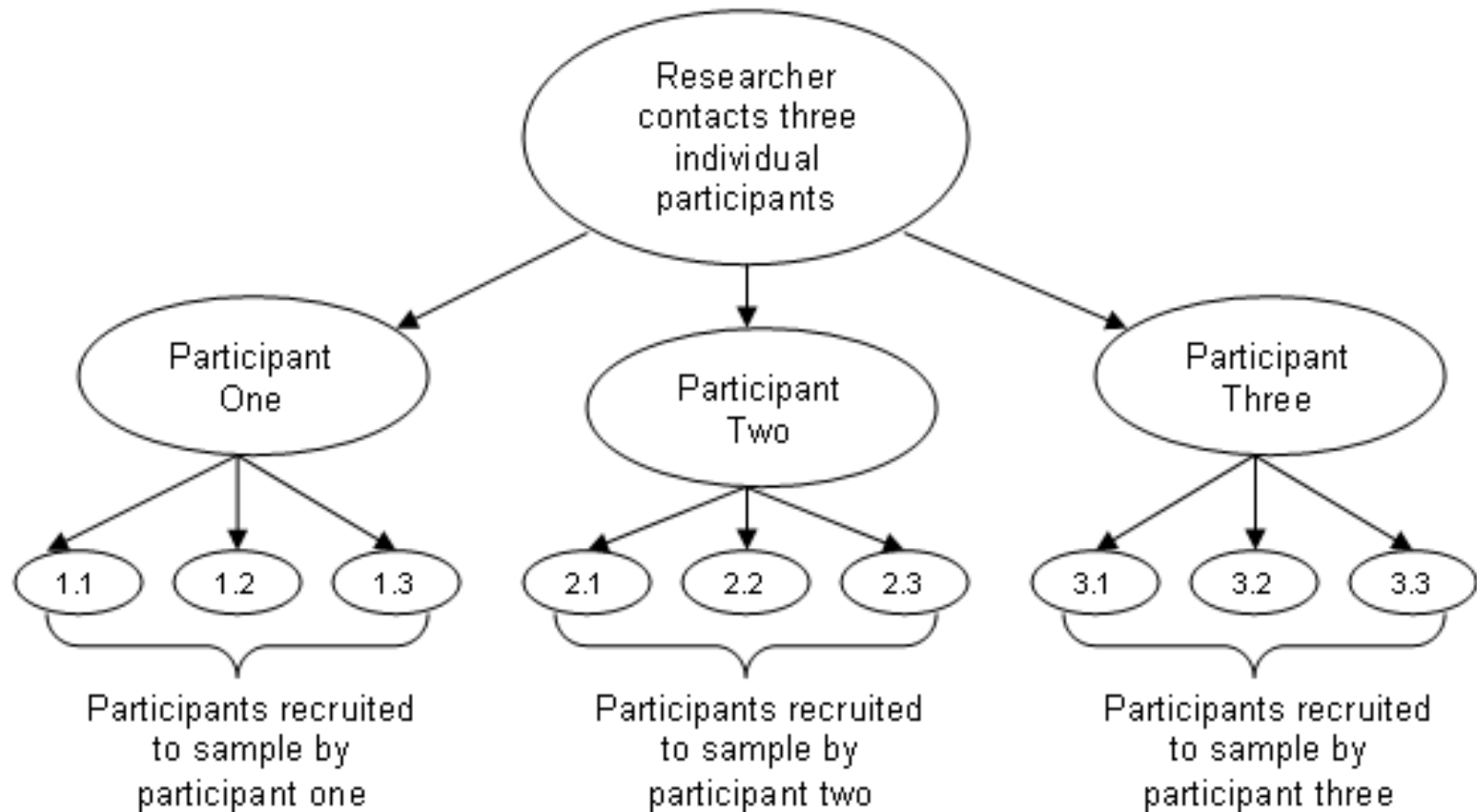
Purposive Sampling

Amaçlı Örnekleme

- Aim → selecting Ss who meet predetermined criteria.
- e.g., population: young theatre goers
 - Not just anyone walking towards the theatre
 - Under, say, 30 years of age.
 - Students (or who look like students)
- A smart practical way of approaching to having a good sample
- But, recall that not a probabilistic sampling process


Nonprobability Sampling

Snowball Sampling



Sampling Frame

- Sampling frame → the actual population of interest.
- Population: residents of Tarsus 😊
- Possible sampling frames can be:
 - *People who go to Kebapçı Eyüp* 😊

**DESCRIPTIVE STATISTICS: UNDERSTANDING
YOUR DATA**

Overall, how do you feel
about your future?

- ☒ very optimistic
- ☐ somewhat optimistic
- ☐ somewhat pessimistic
- ☐ very pessimistic

Frequency Distributions

- The collection of **unprocessed numbers obtained through observation or recording**, which have not yet been organized or made meaningful, is considered ***raw data***.
- One of the most common methods used to organize, summarize, and make raw data meaningful and understandable is to present their *frequency distributions*.
- By presenting frequency distributions, scattered scores are compiled, arranged in descending or ascending order, and meaningful interpretations about the scores can be made.
- **Frequency (f)**

Frequency Distributions

- Raw Data

88	74	76	60	38	65	82	28	54	68
82	67	58	42	94	54	90	63	82	35

- Ranked Scores

94	82	65	54
90	76	63	42
88	74	60	38
82	68	58	35
82	67	54	28

Simple Frequency Distribution

- A *simple frequency distribution* is a way of organizing data that shows how often each individual value (or score) occurs in a dataset.

In this type of distribution:

- Each unique value is listed separately.
- The number of times (frequency) that each value appears is recorded.

Exmpl.: The exam scores of 15 students: 40, 50, 50, 55, 60, 60, 60, 65, 70, 70, 75, 75, 80, 85, 90.

- This is called a *frequency distribution* — it shows *how many times each value is repeated*.

Let's write down SFD table!

Grouped Frequency Distribution

- When the data set is **large** or the values vary widely, listing every single value separately becomes **impractical**.
- In such cases, the data are **grouped into intervals (classes)**, and the **frequency** of each interval is shown.
- Helps us **see the overall pattern of the data more easily** and **summarize large data sets** in a simple and understandable way.

Score Range	Frequency (f)
40–49	1
50–59	3
60–69	4
70–79	4
80–89	2
90–99	1

The Range (ranj)

- The *range* is a simple measure that shows how spread out the data are.

It is found by subtracting the **smallest value** from the **largest value** in the dataset.

Formula:

- $\text{Range} = \text{Highest Score} - \text{Lowest Score}$

Example:

The scores: 40, 50, 50, 55, 60, 60, 60, 65, 70, 70, 75, 75, 80, 85, 90.

- $\text{Range} = 90 - 40 = 50$
- So, the range is **50**.
- **It shows that the data values spread across 50 points.**

The Range (ranj)

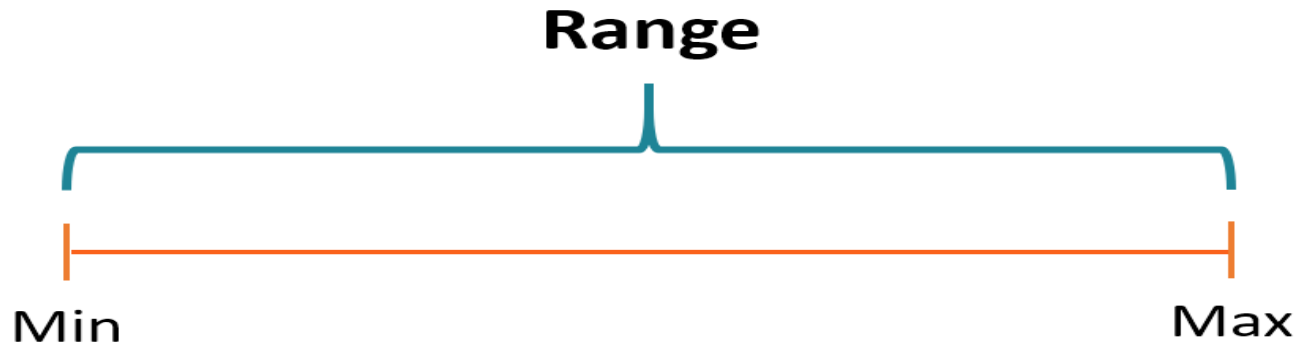
BUT YOU SHOULD BE CAREFUL!!

- The range is always POSITIVE 😊
- Affected by **extreme values!**

Formula: $\text{Range} = X_{max} - X_{min}$

X_{max} = largest value

X_{min} = smallest value



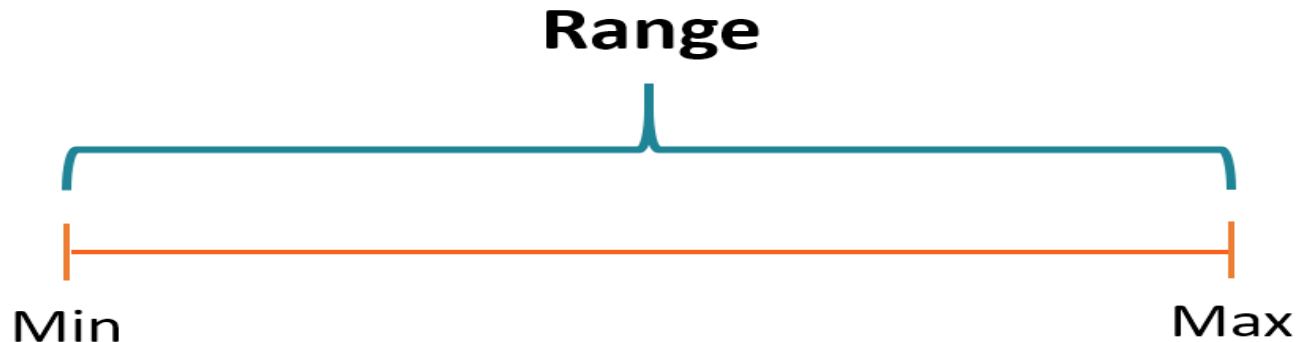
The Range (ranj)

Exmpl.:

1) Data: 59, 59, 59, 60, 61, 61, 61 ise Range=?

2) Data 2: 30, 40, 50, 60, 70, 80, 90 ise Range=?

- Bu iki dağılımda aritmetik ortalama ve medyanlar eşit olmasına karşın ranjları farklıdır.
- Dağılımın ranji azaldıkça dağılımdaki puanlar birbirine yaklaşır, ranj arttıkça puanlar birbirinden uzaklaşır ya da puanlar arası fark artar.



Mean (Aritmetik Ortalama)

- The **mean** is the sum of all values divided by the total number of values. It represents the *average* of the data.

$$\bar{X} = \frac{X_1 + X_2 + X_3}{N}$$

Formula:

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

$\bar{x}$: Mean

$\sum X$: Sum of all data values

N : Number of observations

Standard Deviation (SD)

- The **standard deviation** is a measure that tells us **how spread out** or **how much variation** there is in a set of data values.
- In other words, it shows **how far each value is from the MEAN (average)**.
- It is equal to the square root of the average of the squared differences of the scores from the mean (variance).
- Variance is equal to the square of the standard deviation.
- It does not accept negative values.
- If all values in the data group are the same, the SS is zero.

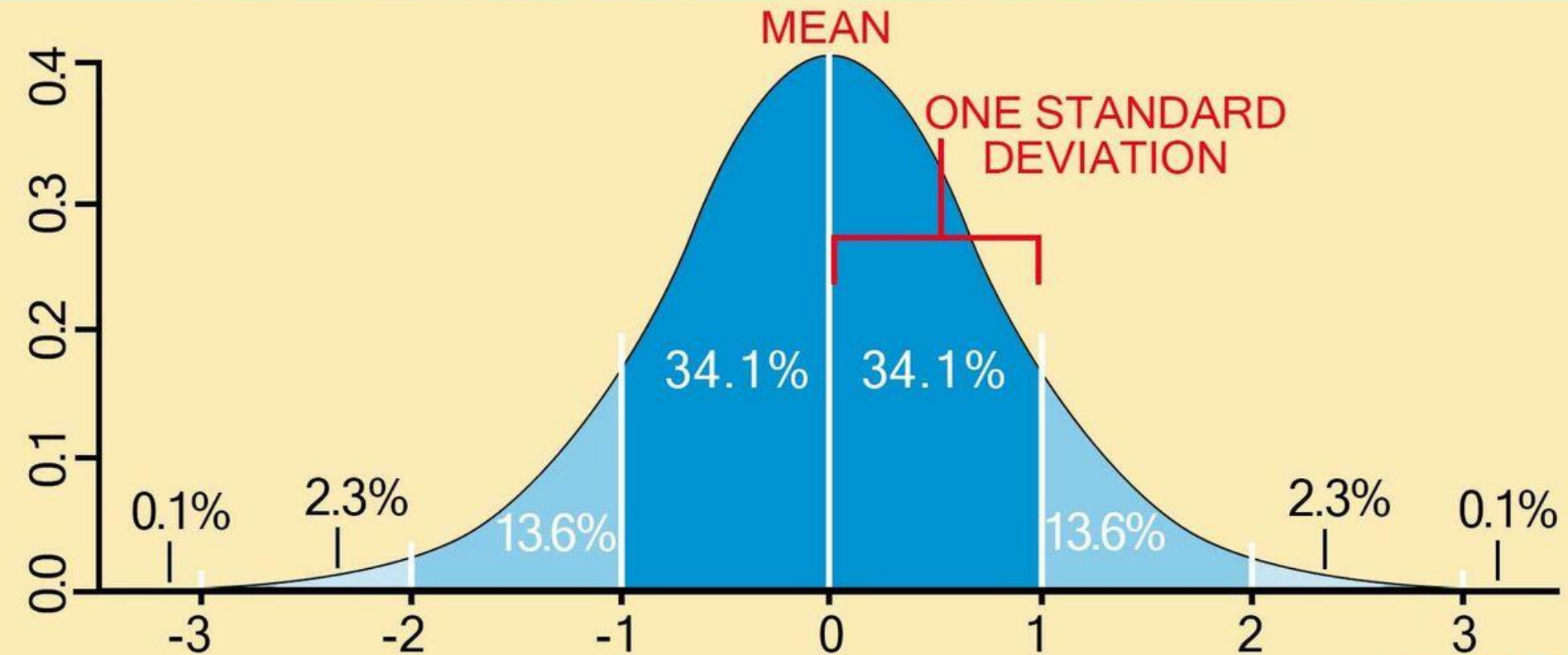
X = each data value

$\bar{X}$ = sample mean

n = sample size

$$S = \sqrt{\frac{\sum (X - \bar{X})^2}{n-1}}$$

STANDARD DEVIATION



Median

- The **median** is the *middle value* when data are arranged in order (from smallest to largest).
- If there's an even number of values, the median is the *average of the two middle values*.

Example:

Scores = 4, 5, 7, 8, 10

→ **Median** = 7

Scores = 4, 5, 7, 8

→ **Median** = $(5 + 7) / 2 = 6$

Mode

- The **mode** is the value that appears *most frequently* in the data.

There can be:

- **One mode** → *Unimodal*
- **Two modes** → *Bimodal*
- **More than two** → *Multimodal*
- **Example:**

Scores = 4, 5, 7, 7, 8, 9

→ Mode = 7

WHAT IF?

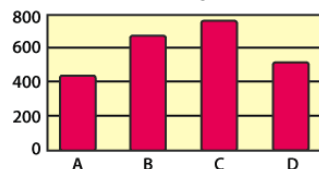
- Scores = 7, 8, 9, 9, 9, 6, 6, 10, 10, 11, 11, 14, 14, 18, 18, 19, 19, 19
- → Mode = ???

Graphical Representation of Data

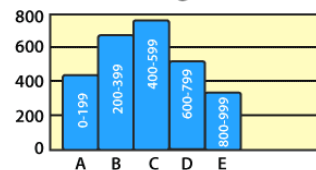
TYPES OF GRAPHICAL REPRESENTATION



Bar Graphs



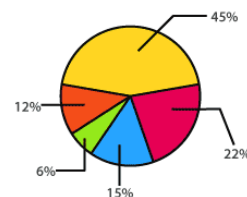
Histograms



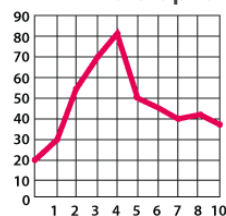
Frequency Table

Rulers of France		
Reign (Years)	Tally	Frequency
1-15		18
16-30		11
31-45		6
46-60		4
61-75		1

Circle Graph



Line Graphs



Stem and Leaf Plot

Stem	Leaf
0	1, 1, 2, 2, 3, 4, 4, 4, 4, 5, 8
1	0, 0, 0, 1, 1, 3, 7, 9
2	5, 5, 7, 7, 8, 8, 9, 9
3	0, 1, 1, 1, 2, 2, 2, 4, 5
4	0, 4, 8, 9
5	2, 6, 7, 7, 8
6	3, 6

Key: 6 | 3 = 63 Year

Line Plot



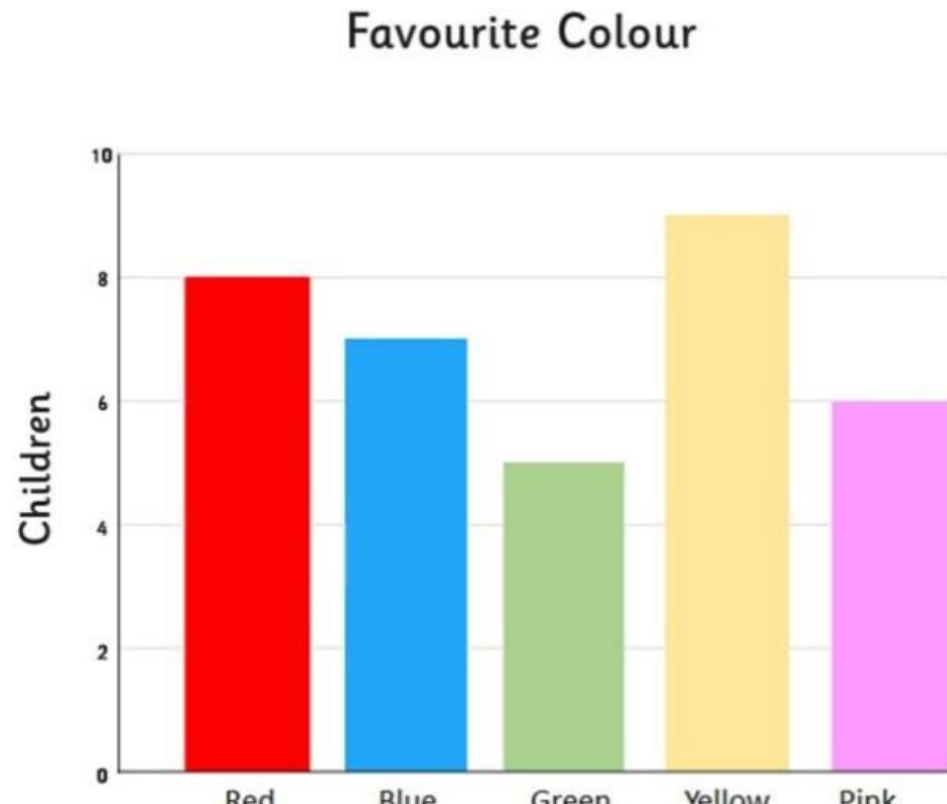
Box and Whisker Plot



1. Bar Graph

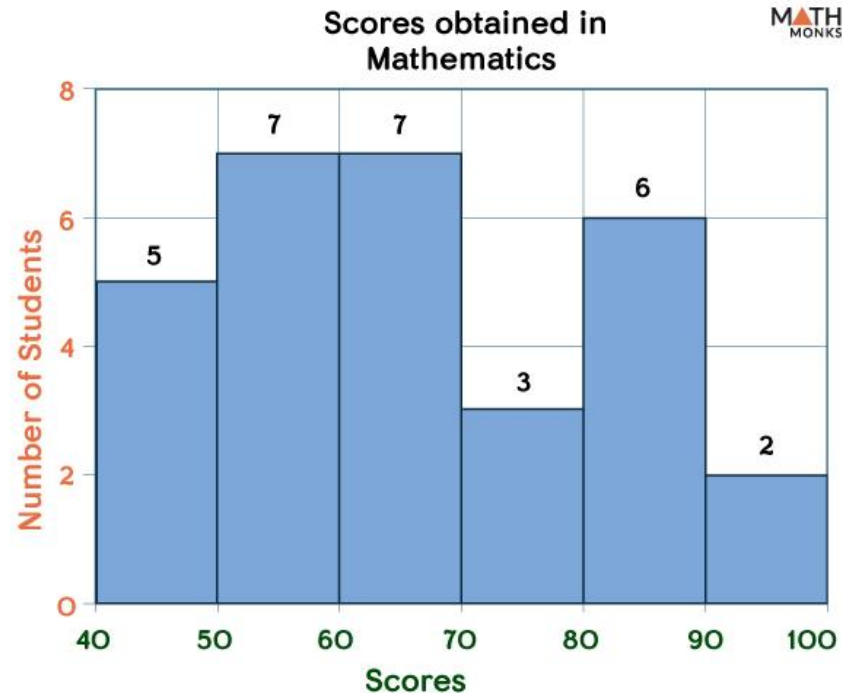
- Used to display **categorical data**.
- Each category is represented by a bar.
- The height of the bar shows frequency.
- There are spaces between bars.

Example: Favorite colors of children



2. Histogram

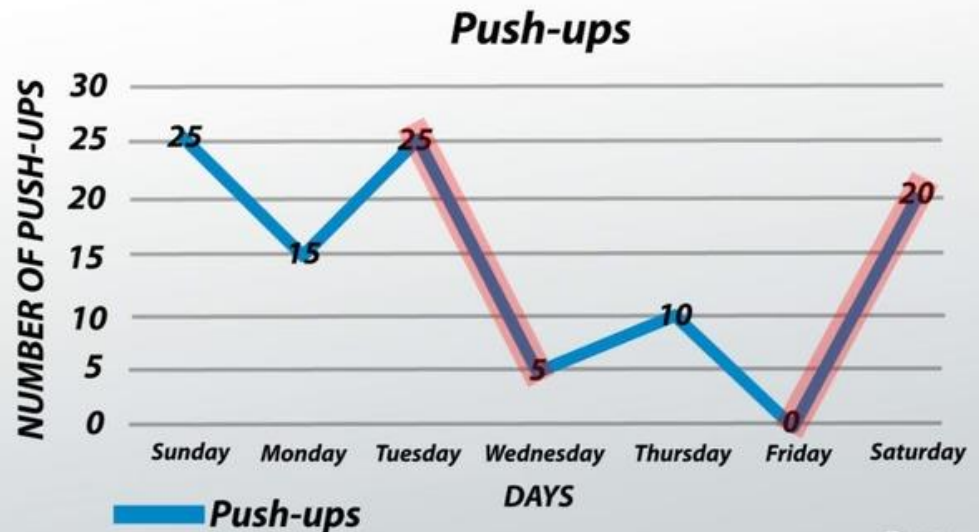
- Used for continuous numerical data.
- Bars are adjacent (no gaps).
- Shows frequency distribution of intervals.
- Example: Distribution of exam scores (40–49, 50–59, 60–69, 70–79).



3. Line Graph

- Used to display changes over time.
- Points are plotted and connected by lines.
- Useful for showing trends (increase/decrease).
- Example: Number of push-ups in a week 😊

EXAMPLE OF A LINE GRAPH



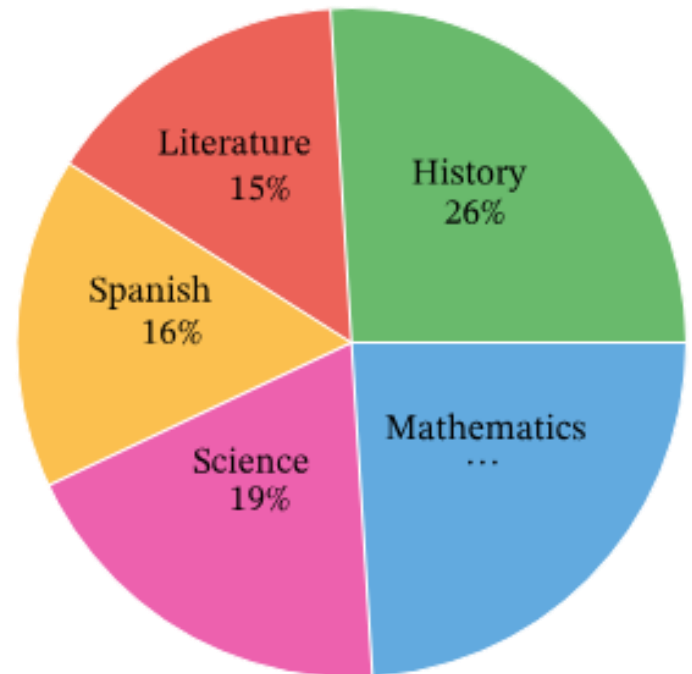
4. Circle Graph (Pie Chart)

- Used to show proportions or percentages of a whole.
- Circle represents 100% (360°).
- Each slice shows one category.
- Example: Literature– 15%, Spanish 40%

Circle Graph



How many percent of students are taking Mathematics?

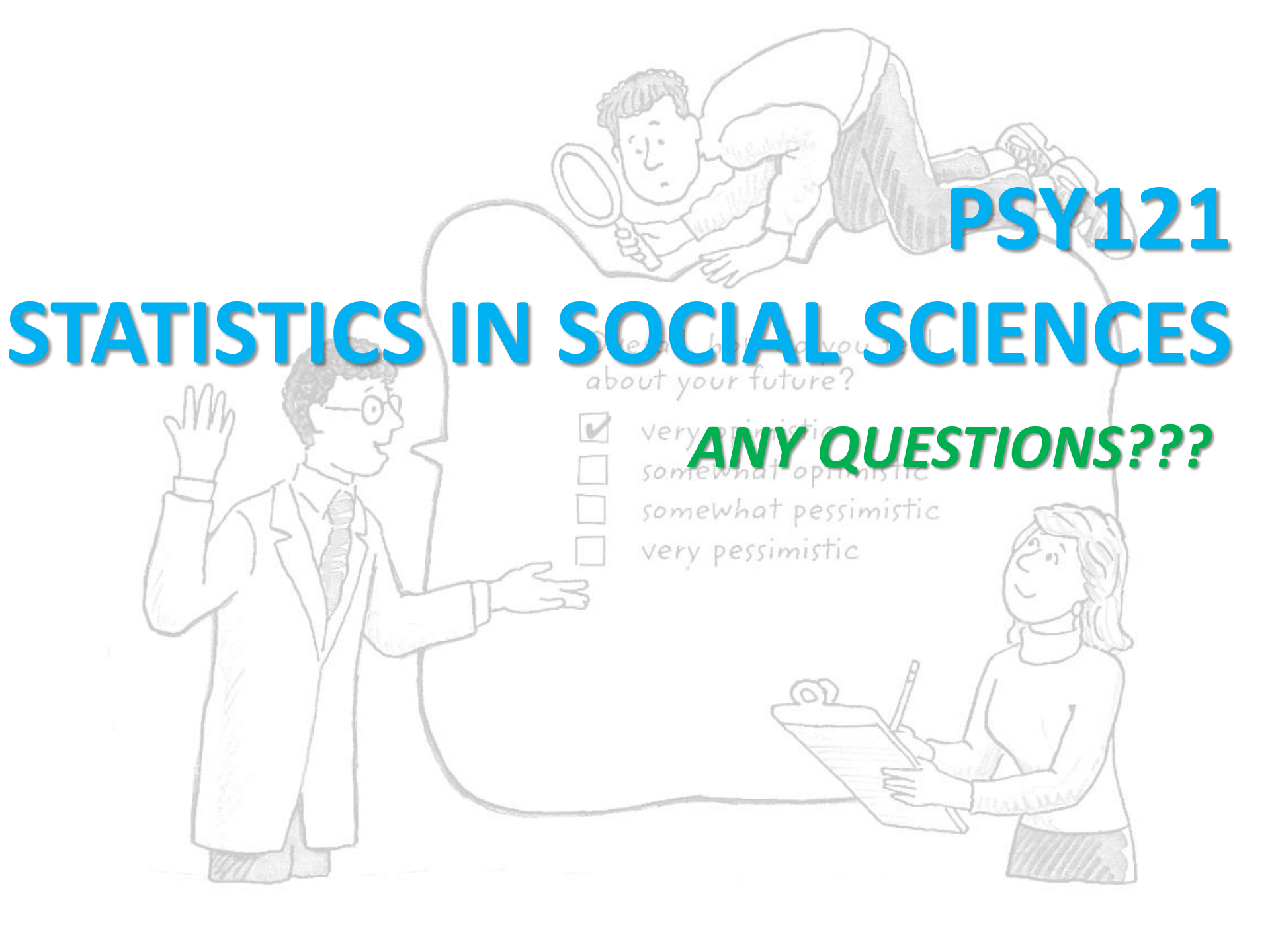


5. Steam and Leaf Plot

- A **stem-and-leaf plot** is a method of organizing numerical data so you can see both:
 - the **distribution (shape)** of the data, and
 - the **exact values** themselves.
- It's especially useful for **small to medium-sized data sets**.
- Each number is **split into two parts**:
 - **Stem**: the leading digit(s)
 - **Leaf**: the last digit

stem	leaf
0	1, 1, 2, 2, 3, 4, 4, 4, 4, 5, 8
1	0, 0, 0, 1, 1, 3, 7, 9
2	5, 5, 7, 7, 8, 8, 9, 9
3	0, 1, 1, 1, 2, 2, 2, 4, 5
4	0, 4, 8, 9
5	2, 6, 7, 7, 8
6	3, 6

Key: 6|3 = 63 years old



PSY121

STATISTICS IN SOCIAL SCIENCES

How do you feel about your future?

- ☒ very optimistic
- ☐ somewhat optimistic
- ☐ somewhat pessimistic
- ☐ very pessimistic

ANY QUESTIONS???