Department of CSE
(School of Technology)
MATH2361: Probability and Statistics

(No. of hrs/week: 3 Credits: 3)

@Semester –IV

UNIT-I
Introduction to Statistics –Scales of Measurement

Dr Malikarjuna Reddy Doodipala
MSc, M.Phil, PGDCA, Ph D,
Associate Professor
Department of Mathematics
GITAM Hyderabad

Learning Objectives

By the end of this topic, students should be able to:

- Know Scales of Measurement nominal, ordinal, interval and ratio scale
- Understand Graphical representation of the data
- Analyze the given data using MS-Excel

Learning Outcomes

Upon successful completion of this topic, students will be able to:

- Use Know Scales of Measurement nominal, ordinal, interval and ratio scale
- Show that Graphical representation of the data
- Analyze the given data using MS-Excel

Contents

- Scales of Measurement-nominal, ordinal, interval and ratio scale
- Graphical representation of the data
- Analysis of the data using MS-Excel

Prerequisite: Data and Data Sets

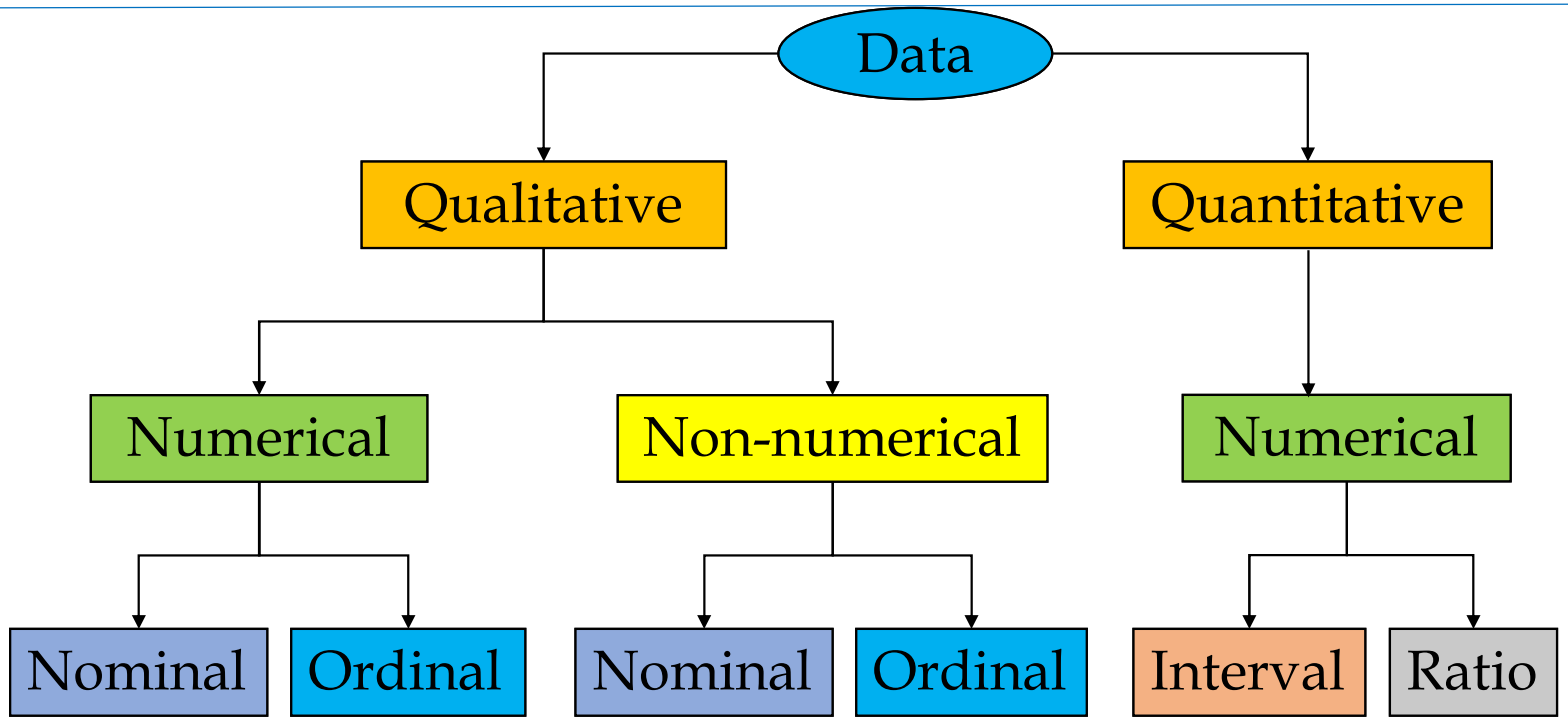
- Data are the facts and figures collected, summarized, analyzed, and interpreted.

  - The data collected in a particular study are referred to as the data set.

Prerequisite: Elements, Variables, and Observations

- The elements are the entities on which data are collected.
- A variable is a characteristic of interest for the elements.
- The set of measurements collected for a particular element is called an observation.
- The total number of data values in a data set is the number of elements multiplied by the number of variables.

Scales of Measurement



Scales of Measurement

Scales of measurement include:

- Nominal
- Interval
- Ordinal
- Ratio

The scale determines the amount of information contained in the data.

The scale indicates the data summarization and statistical analyses that are most appropriate.

Scales of Measurement- Nominal

Data are labels or names used to identify an attribute of the element.

A nonnumeric label or numeric code may be used.

Scales of Measurement- Nominal

Example:

Students of a university are classified by the residence that they live in using a nonnumeric label such as Farley, Keena, Azhar, Phillips, and so on.

A numeric code can be used for the school/college variable (e.g. 1: Farley, 2: Keena, 3: Azhar, and so on).

Scales of Measurement- Ordinal

- The data have the properties of nominal data and the order or rank of the data is meaningful.

A nonnumeric label or numeric code may be used.

- An Ordinal Number is a number that tells the position of something in a list, such as 1st, 2nd, 3rd, 4th, 5th etc.

Scales of Measurement- Ordinal

Example:

Students of a university are classified by their class standing using a nonnumeric label such as Fresher, Junior, or Senior.

A numeric code can be used for the class standing variable (e.g. 1 denotes Fresher, 2 denotes junior, and so on).

Scales of Measurement- Interval

The data have the properties of ordinal data, and the interval between observations is expressed in terms of a fixed unit of measure.

Interval data are always numeric.

Scales of Measurement- Interval

Example: Average Salary Offer 2020 per month
Economics/Finance: RS 30,084
History: Rs 52,108
English: Rs 37,454

Scales of Measurement- Ratio

The data have all the properties of interval data and the ratio of two values is meaningful.

Variables such as distance, height, weight, and time use the ratio scale.

This scale must contain a zero value that indicates that nothing exists for the variable at the zero point.

Qualitative and Quantitative Data

Data can be qualitative or quantitative.

The appropriate statistical analysis depends on whether the data for the variable are qualitative or quantitative.

There are more options for statistical analysis when the data are quantitative.

Qualitative Data

Labels or names used to identify an attribute of each element. E.g., Black or white, male or female.

Referred to as categorical data

Use either the nominal or ordinal scale of measurement

Can be either numeric or nonnumeric

Appropriate statistical analyses are rather limited

Quantitative Data

Quantitative data indicate how many or how much:

- Discrete, if measuring how many. E.g., number of students attending my class
- Continuous, if measuring how much. E.g., Weights *(kg)

Quantitative data are always numeric.

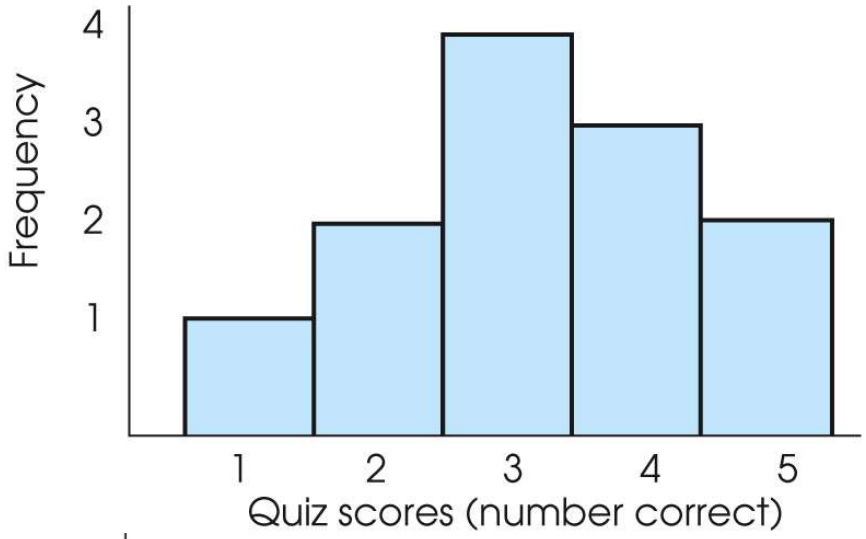
Ordinary arithmetic operations are meaningful for quantitative data.

Graphical representation: Frequency Distribution

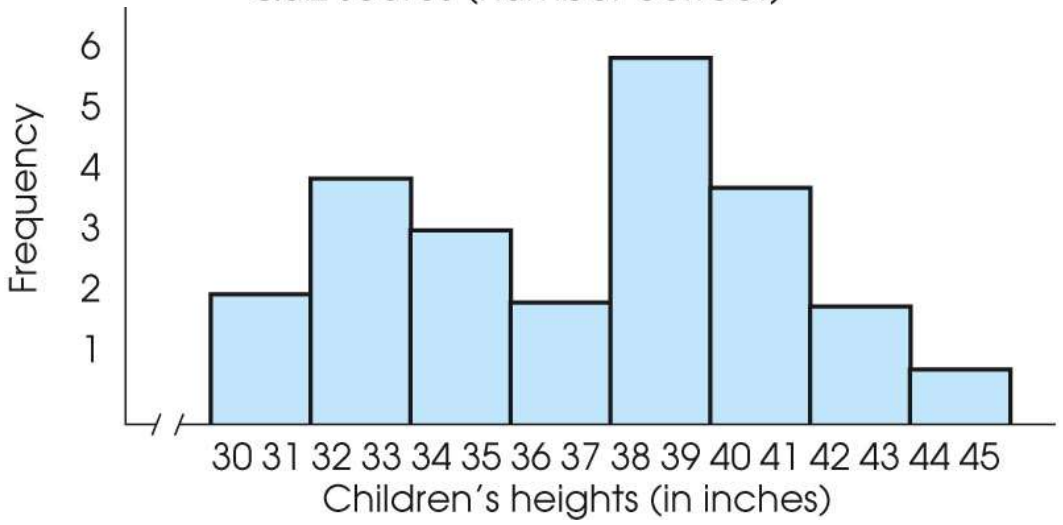
- In a **frequency distribution graph**, the score categories (X values) are listed on the X axis and the frequencies are listed on the Y axis.
- When the score categories consist of numerical scores from an interval or ratio scale, the graph should be either a histogram or a polygon.

Histograms

- In a **histogram**, a bar is centered above each score (or class interval) so that the height of the bar corresponds to the frequency and the width extends to the real limits, so that adjacent bars touch.



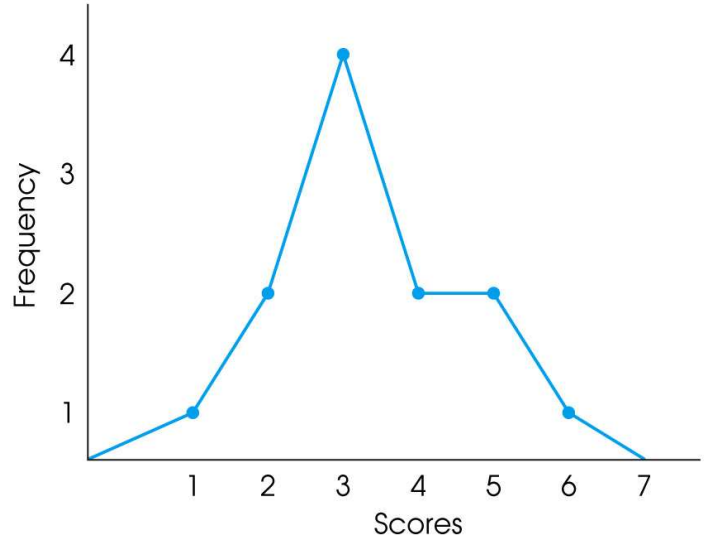
X	f
5	2
4	3
3	4
2	2
1	1



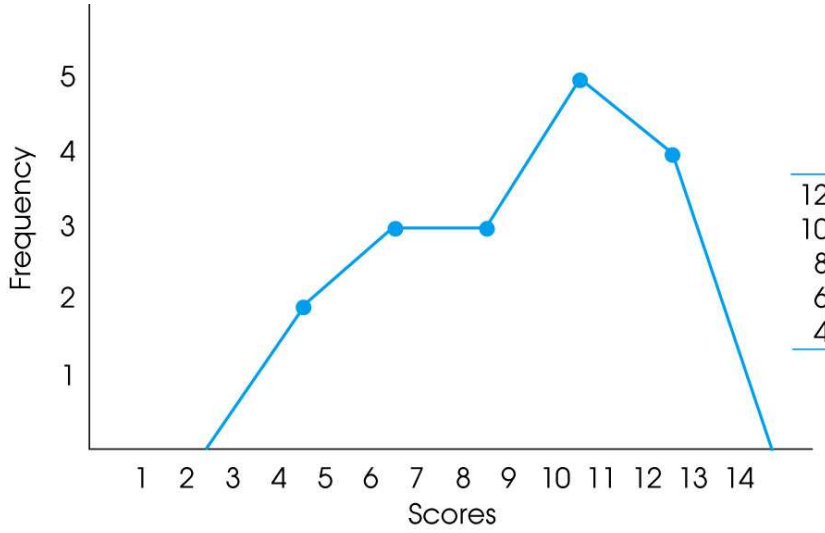
X	f
44–45	1
42–43	2
40–41	4
38–39	6
36–37	2
34–35	3
32–33	4
30–31	2

Polygons

- In a **polygon**, a dot is centered above each score so that the height of the dot corresponds to the frequency.
- The dots are then connected by straight lines.
- An additional line is drawn at each end to bring the graph back to a zero frequency.



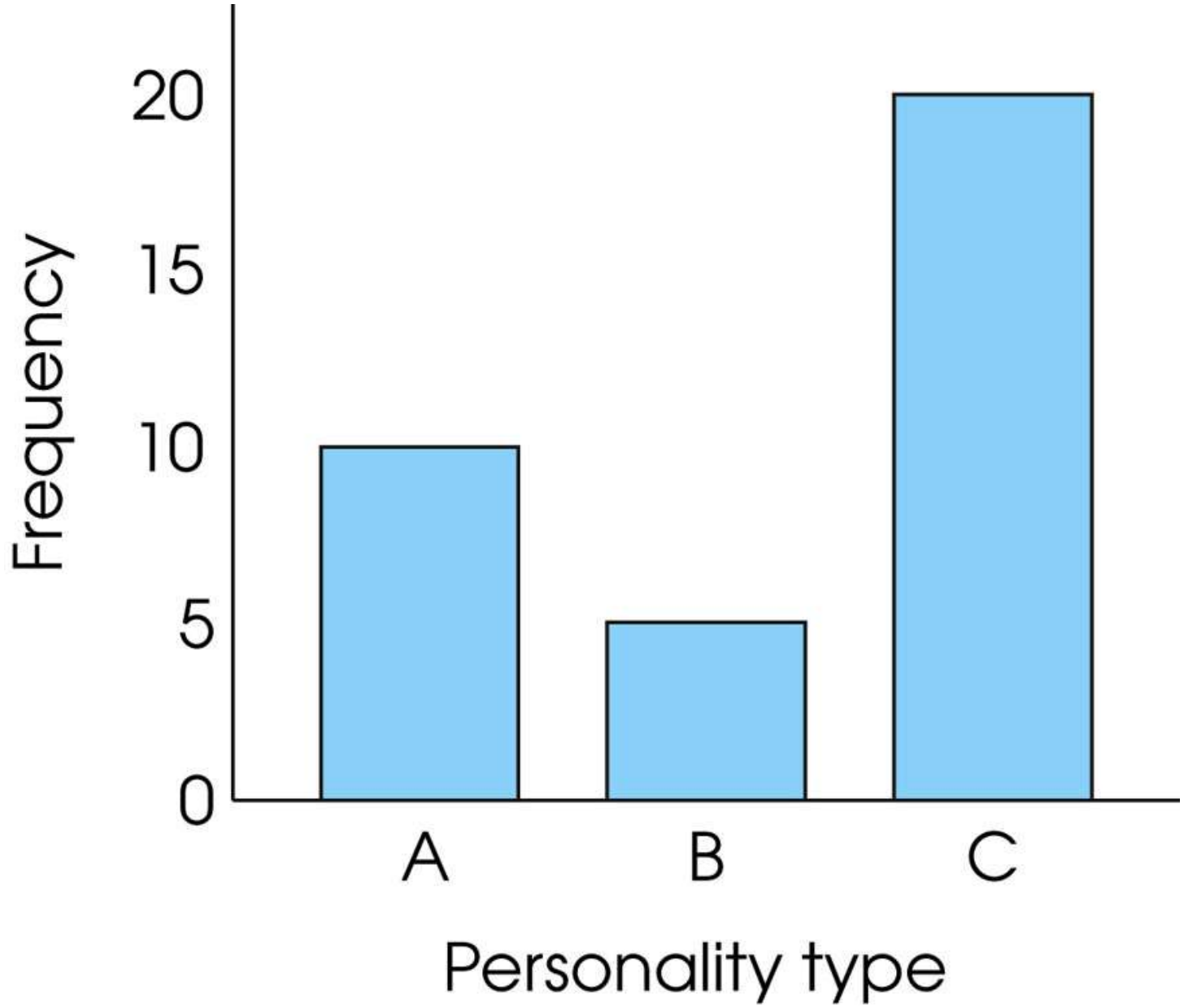
X	f
6	1
5	2
4	2
3	4
2	2
1	1



X	f
12–13	4
10–11	5
8–9	3
6–7	3
4–5	2

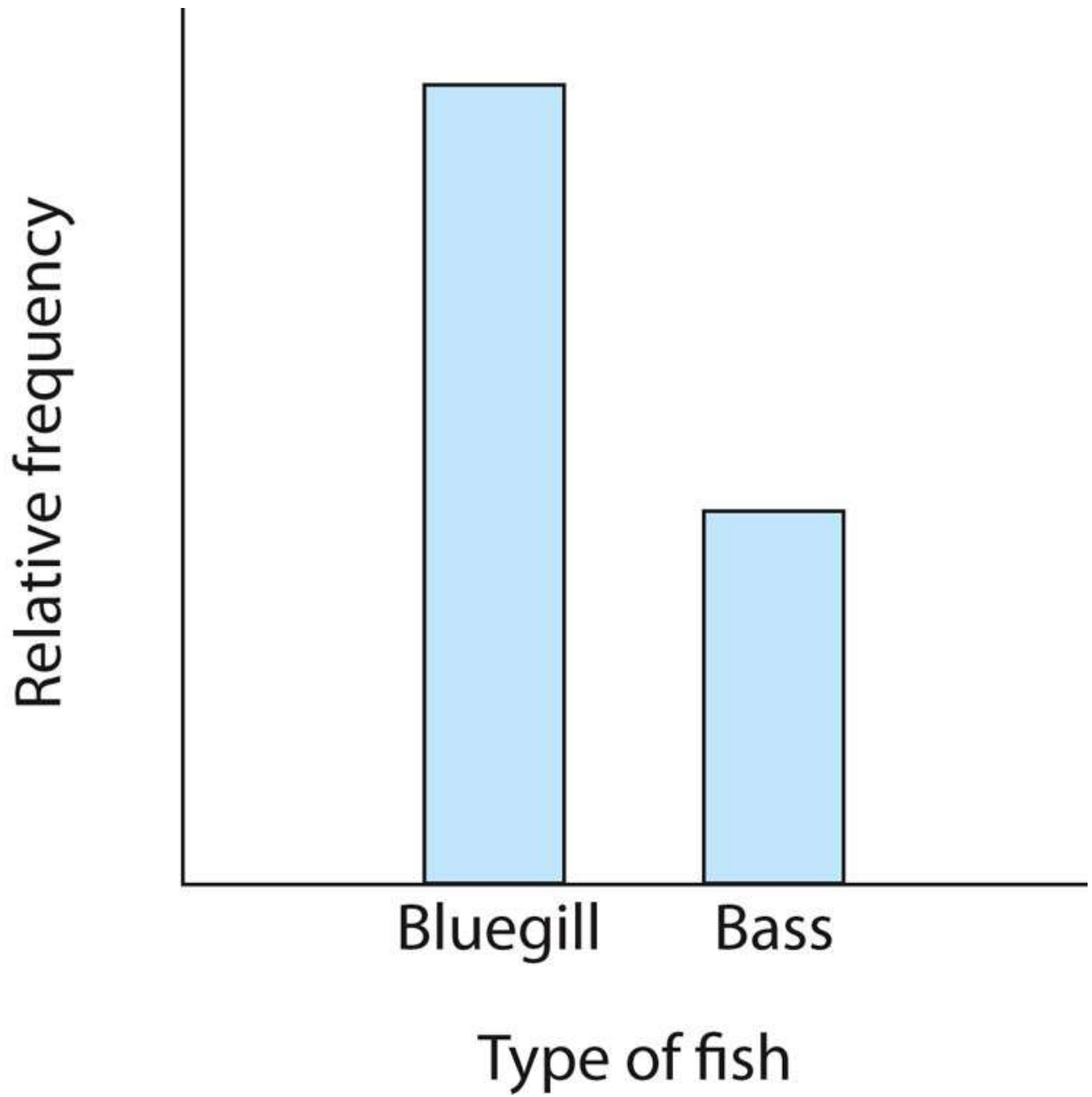
Bar graphs

- When the score categories (X values) are measurements from a nominal or an ordinal scale, the graph should be a bar graph.
- A **bar graph** is just like a histogram except that gaps or spaces are left between adjacent bars.


Frequency
20
15
10
5
0
A
B
C
Personality type

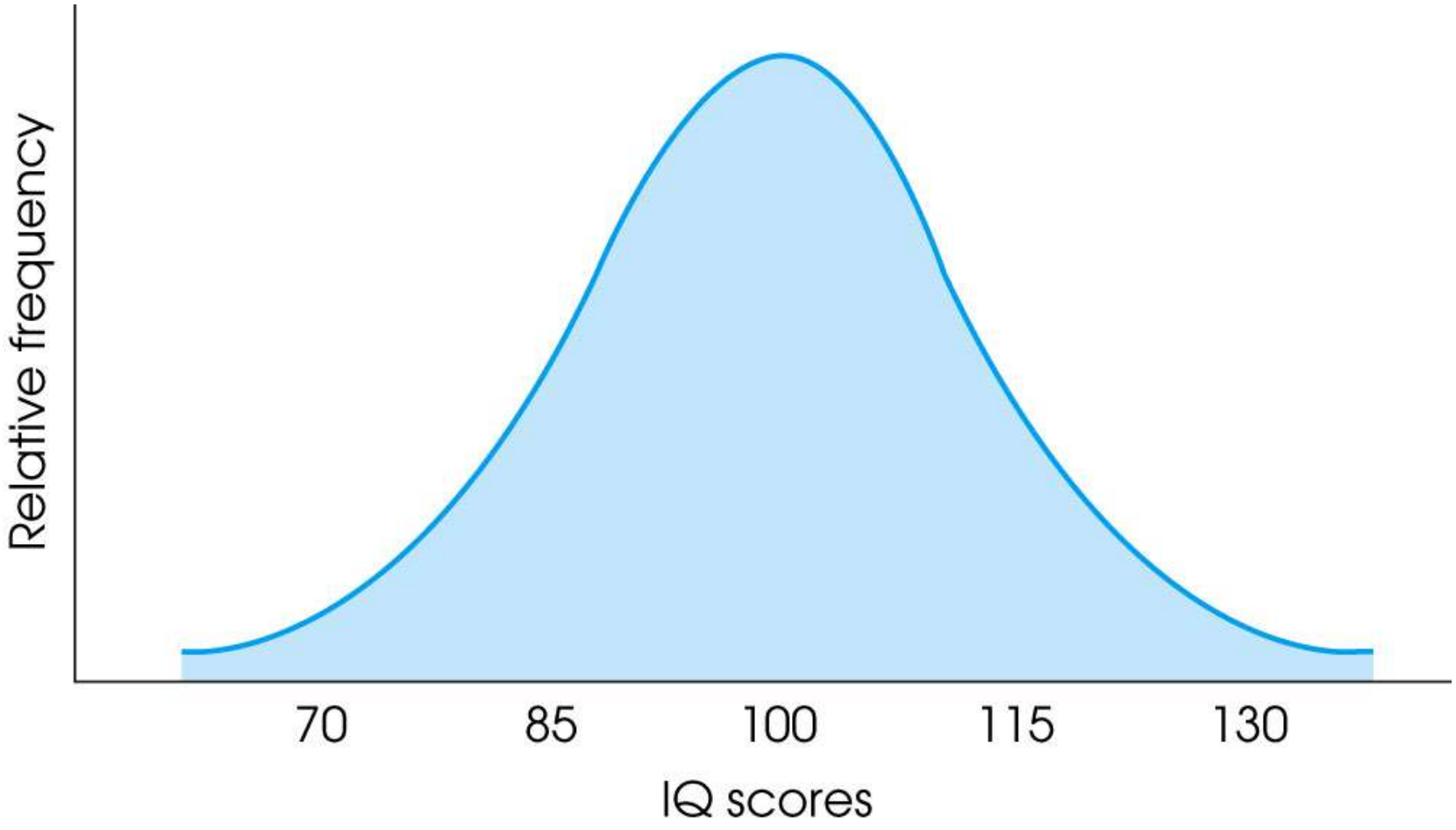
Relative frequency

- Many populations are so large that it is impossible to know the exact number of individuals (frequency) for any specific category.
- In these situations, population distributions can be shown using **relative frequency** instead of the absolute number of individuals for each category.


Relative frequency
Bluegill
Bass
Type of fish

Smooth curve

- If the scores in the population are measured on an interval or ratio scale, it is customary to present the distribution as a **smooth curve** rather than a jagged histogram or polygon.
- The smooth curve emphasizes the fact that the distribution is not showing the exact frequency for each category.


Relative frequency
70
85
100
115
130
IQ scores

Frequency distribution graphs

- Frequency distribution graphs are useful because they show the entire set of scores.
- At a glance, you can determine the highest score, the lowest score, and where the scores are centered.
- The graph also shows whether the scores are clustered together or scattered over a wide range.

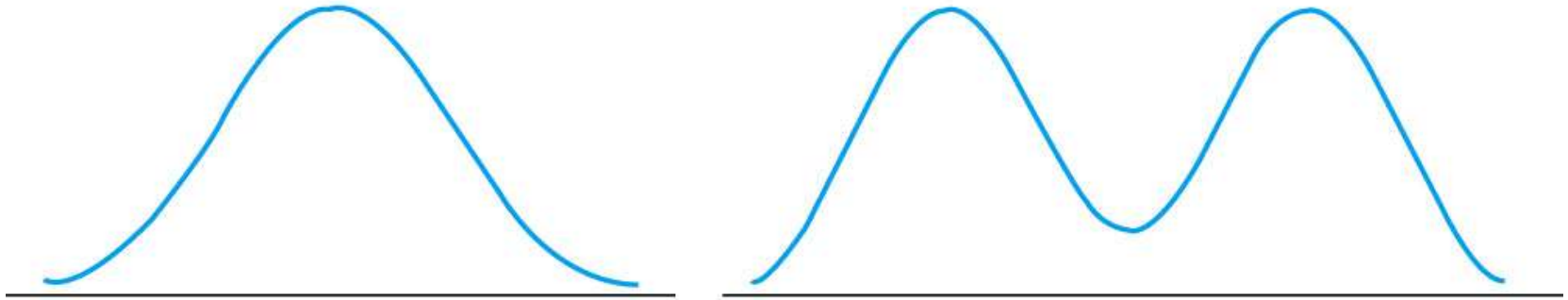
Shape

- A graph shows the **shape** of the distribution.
- A distribution is **symmetrical** if the left side of the graph is (roughly) a mirror image of the right side.
- One example of a symmetrical distribution is the bell-shaped normal distribution.
- On the other hand, distributions are **skewed** when scores pile up on one side of the distribution, leaving a "tail" of a few extreme values on the other side.

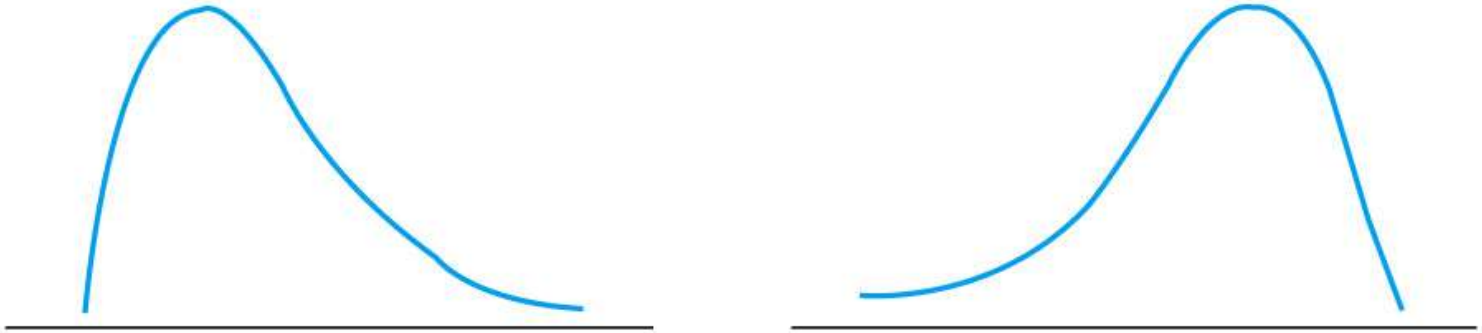
Positively and Negatively Skewed Distributions

- In a **positively skewed** distribution, the scores tend to pile up on the left side of the distribution with the tail tapering off to the right.
- In a **negatively skewed** distribution, the scores tend to pile up on the right side and the tail points to the left.

Symmetrical distributions



Skewed distributions



Positive skew

Negative skew

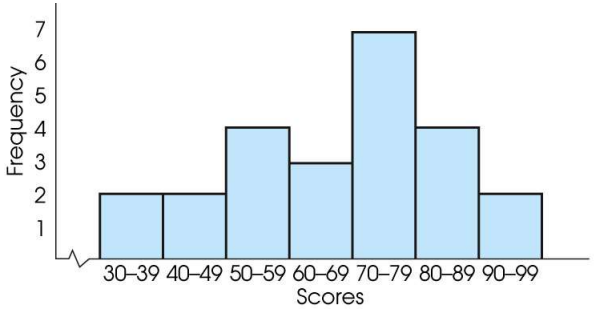
Stem-and-Leaf Displays

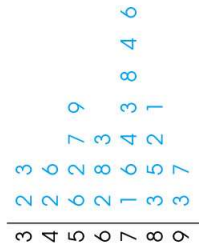
- A **stem-and-leaf display** provides a very efficient method for obtaining and displaying a frequency distribution.
- Each score is divided into a **stem** consisting of the first digit or digits, and a **leaf** consisting of the final digit.
- Finally, you go through the list of scores, one at a time, and write the leaf for each score beside its stem.
- The resulting display provides an organized picture of the entire distribution. The number of leafs beside each stem corresponds to the frequency, and the individual leafs identify the individual scores.

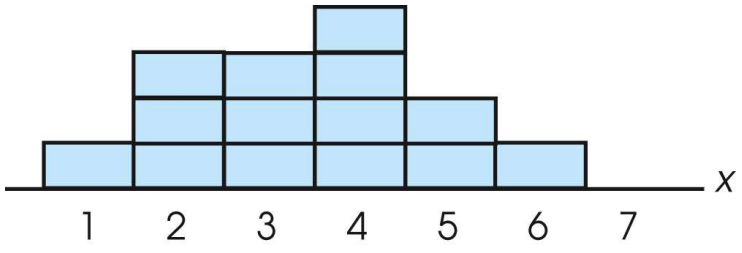
TABLE 2.3

A set of $N = 24$ scores presented as raw data and organized in a stem and leaf display.

Data			Stem and Leaf Display	
83	82	63	3	23
62	93	78	4	26
71	68	33	5	6279
76	52	97	6	283
85	42	46	7	1643846
32	57	59	8	3521
56	73	74	9	37
74	81	76		







Summery

- Data types
- Frequency distributions
- Ungrouped and grouped
- Graphical representations of the data

Quiz
https://docs.google.com/forms/d/e/1FAIpQLSefDZxdkOOEdECeYfs3JOZeXcCvDlaj2Z7W8HsJL8KU6IU8Vg/viewform

Thank You !

Feedback to mallik.reddyd@gmail.com