

Introduction to Probability and Statistics

Topic (1): “Describing Data with Graphs”



Dr. Heba Ayyoub

Philadelphia University

Topic (1): “Describing Data with Graphs”

1 Variables and Data

Main concepts:

1) Variable: is a characteristic that changes over time and/or varies from person to other.

Example (1): Body temperature, height, weight, sex, age, color, major, GPA.

2) Experimental unit: is the person or object on which a variable is measured.

Example (2): We need to study a GPA of 50 students.

➤ **Variable:**

➤ **Experimental Unit:**

Note: When the variable measured on a set of experimental unit a result is a set of measurement / data.

Example (3): GPA of students are 85, 67, 93, 87, ... are called data / set of measurement.

3) data: is a measurements of the sample based on variable(s).

- **Types of data:**

- **Univariate data:** **One variable** is measured on a single experimental unit.
- **Bivariate data:** **Two variables** are measured on a single experimental unit.
- **Multivariate data:** **More than two variables** are measured on a single experimental unit.

4) Population: is the set of all measurements of interest to the investigator.

5) Sample: is a subset of measurements selected from the population of interest.



Example (4): We need to study GPA of 18 students from mathematics department at Philadelphia University.

Population:

Sample:

Variable:

Experimental Unit:

6) Parameter: Numerical description of population characteristic.

Example (5): Average age of all people in Jordan.

7) Statistic: Numerical description of sample characteristic.

Example (6): Average age of people from three cities in Jordan.

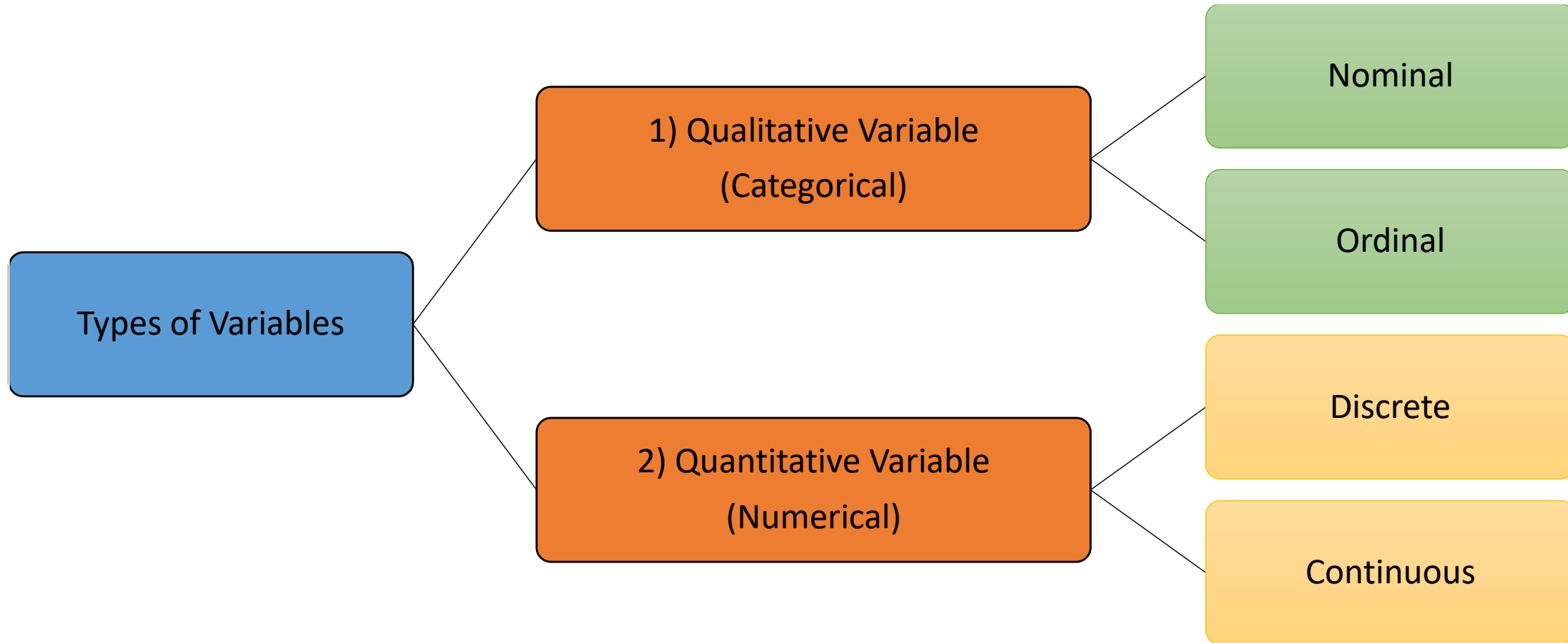
Example (7): A set of 4 students are selected from Introduction to Probability and Statistics class and the following measurement are recorded:

Student	GPA	Gender	Major
1	75	F	CS
2	80	M	Mathematics
3	85	M	Bio Technology
4	72	F	Mathematics

- **Population:**
- **Sample:**
- **Experimental Unit:**
- **How many variables we have:**
- **Data:**
- **Types of data:**

2 Types of Variables

Variables can be classified into one of two categories:



1) Qualitative variables:

Words or attributes.

Types of Qualitative:

➤ **Nominal Variables:**

Example (8):

- Blood type (A, B, AB, O).
- Nationality (Jordanian, Palestinian, Syrian, Iraqi, Egyptian, Omani, Saudi, ...).
- Sex (Male, Female)
- State of birth (Irbid, Jerash, Amman, ...).

➤ **Ordinal Variables:**

Example (9):

- Educational level (Elementary, Intermediate, ...).
- Students grade (A, B, C, D, F).
- Military rank.

2) Quantitative variables:

Numbers.

Types of Quantitative:

➤ **Discrete variable:**

Can assume only a finite or countable number of values.

Example (10):

- Number of family members.
- Number of new car sales.

➤ **Continuous variable:**

Can assume the infinitely many values corresponding to the points on a line interval.

Example (11): Height, weight, time, distance, and volume.

Example (13): Choose the correct answer:

1. The age is an example of:

- a) Qualitative variable
- b) Discrete Quantitative
- c) Continuous Quantitative
- d) Continuous Qualitative

2. The number of family members is an example of:

- a) Qualitative variable
- b) Quantitative variable
- c) a + b
- d) None

Exercises (1): Experimental Units Identify the experimental units on which the following variables are measured:

- a. Gender of a student.
- b. Number of errors on a midterm exam.
- c. Age of a cancer patient.
- d. Number of flowers on an azalea plant.
- e. Color of a car entering a parking lot.

Experimental Units: The student.

Exercises (2): Qualitative or Quantitative? Identify each variable as quantitative or qualitative:

- a. Amount of time it takes to assemble a simple puzzle.
- b. Number of students in a first-grade classroom.
- c. Rating of a newly elected politician (excellent, good, fair, poor).
- d. State in which a person lives.

Quantitative

Exercises (3): Discrete or Continuous? Identify the following quantitative variables as discrete or continuous:

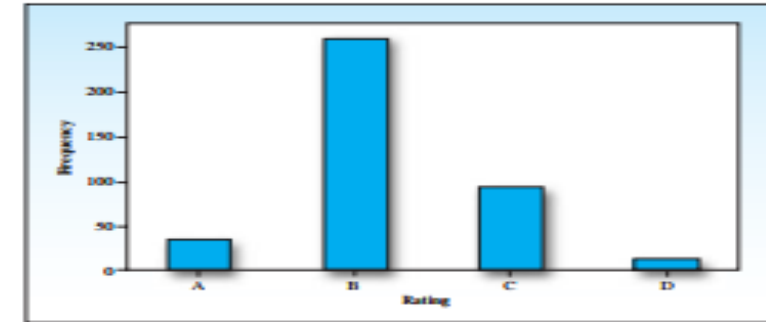
- a. Population in a particular area in Jordan.
- b. Weight of newspapers recovered for recycling on a single day.
- c. Time to complete a mathematics exam.

Discrete

3 Graphs for Categorical Data

1) Bar chart

- Bar charts display data using rectangular bars.
- Each bar represents a category with its height showing frequency.
- Useful for comparing different categories.



- Frequency (f): Number of measurement in each category.
- Relative Frequency ($r.f$): Proportion of measurement in each category.

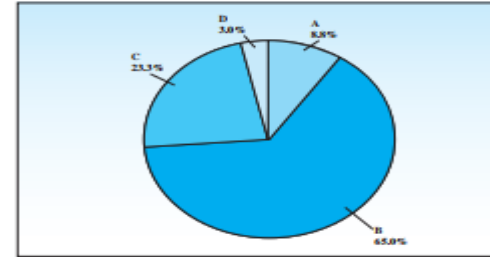
$$r.f = \frac{f}{\sum f}$$

- Percentage (p) = $r.f \times 100\%$
- Sample size (n): $n = \sum f$

- $\sum r.f = 1$
- $\sum p = 100\%$

2) Pie Chart

- A pie chart is a circular graph divided into slices.
- Each slice represents a proportion of the whole.
- Best for showing percentages and parts of a whole.



1. Draw a circle
2. $\theta = r.f \times 360^\circ$

$$\sum \theta = 360^\circ$$

Example (14): Number of students in Genetics class 13, Mathematics class 12, and IT class 10, then the relative frequency of Genetics students is

Example (15):

Department	Mathematics	Genetics	IT
$r.f$	0.1	0.3	A

1) Find the $r.f$ of the IT department (Find the value of A)?

2) Draw a bar chart?

3) Draw a pie chart?

Example (16): Given the following grades of 20 students in Philadelphia University:

A, A, B, B, B, C, C, C, A, B, C, A, B, B, C, C, A, C, B, C.

1) Construct a Frequency table.

Category	f	$r.f$	Cumulative f	Percentage	θ
A					
B					
C					

2) Draw a bar chart.

3) Draw a pie chart.

Exercises (4): Fill-in the blanks for the given frequency distribution.

Category	f	$r.f$	Cumulative f
A			32
B		0.27	
C			
D	21		100

4 Graphs of Quantitative Variables

- 1) Bar chart
- 2) Pie chart
- 3) Line chart
- 4) Dot plot
- 5) Stem and leaf
- 6) Histogram
- 7) Box plot

3) Line chart connect data points with a line and usually used when a Quantitative variable is recorded over time.

Example (17): Draw a line chart for the following data:

time	temperature
6	36
7	38
8	37
9	36
10	38
11	35

4) Dot plot: is a simple type of graph where each data point is represented by a dot. It's used to show how often something happens in a set of data.

Example (18): Given the following grades:

9, 8, 7, 7, 8, 9, 10, 10, 10, 10, 1, 3, 5. Draw dot-plot.

5) Stem and leaf: Each number is separated into a leaf (the right most digit) and a stem (the remaining left most digits). Used to display distribution of Quantitative variable.

Example (19): Given the following data:

3.2, 2.1, 2.2, 2.9, 5, 6.9, 4, 4.1, 4.2, 4.3, 4.2, 3.4, 3.5, 5.8, 5.7, 3.6, 4.4, 5.9, 4.9, 4.8, 4.8, 4.5, 4.6. Draw stem and leaf?

Note: Number of leaves = Number of data.

Example (20): 32 → Leaf unit =

3.2 → Leaf unit =

35.2 → Leaf unit =

3.52 → Leaf unit =

3.255 → Leaf unit =

Example (21):

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

- 1) If the leaf unit is 0.1, what is the min and max values.
- 2) Find the $r.f$ of the observations that are less than 6.3?
- 3) Find the $r.f$ of the observations that are less than or equal 6.4?
- 4) Find the $r.f$ of the observations that are greater than or equal 8.4?

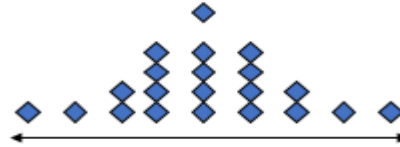
Example (22): Given the following data:

12, 11, 2, 25, 21, 24, 23, 39, 39, 36, 38, 35, 38, 34, 29. Construct stem and leaf plot?

Shapes of Distributions

1) Symmetric

If the right and left sides of the distribution when divided at the middle value form a mirror image.



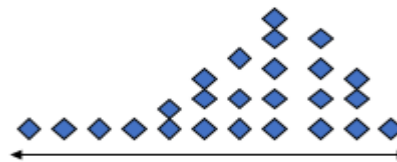
2) Skewed to the right

If a greater proportion of the measurement lie to the right of the peak value.



3) Skewed to the left

If a greater proportion of the measurement lie to the left of the peak value.

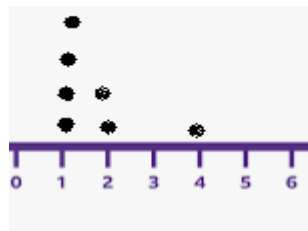
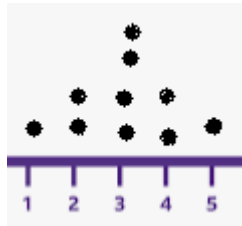


Example (23): Given the following stem and leaf plot:

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

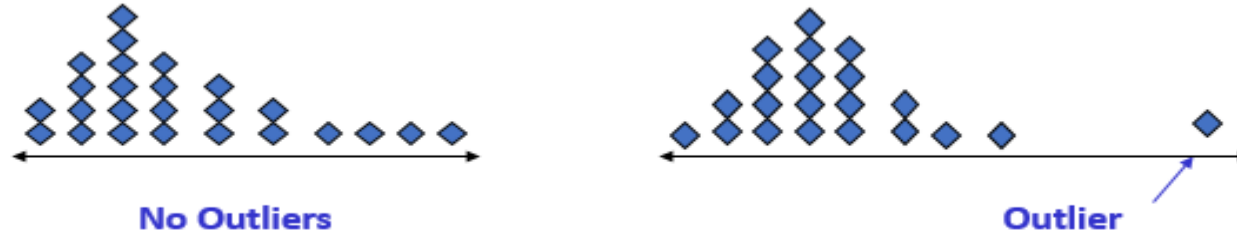
Describe the shape of distribution?

Example (24): Given the following dot plot, describe the shape of distribution?



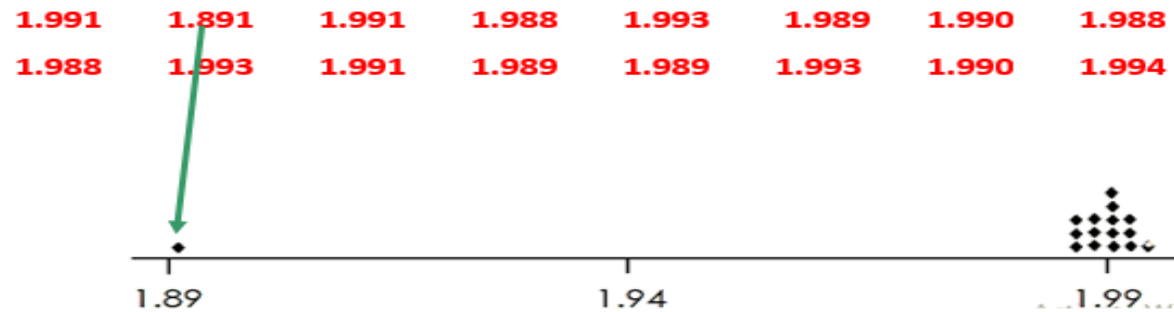
Outliers

One observation or more that are not compatible with rest of observations.



Example (25):

Are there any strange or unusual measurements that stand out in the data set?

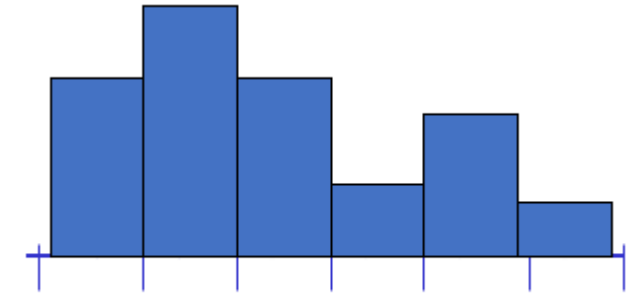


5 Relative Frequency Histograms

A relative frequency histogram for a quantitative data set is a bar chart in which the height of the bar shows measurements fall in a particular class or subinterval.

Note:

- The classes do not overlap.
- Each class has a lower-class limit, which is the least number that can belong to the class, an upper-class limit, which is the greatest number that can belong to the class.
- Class Length = Upper Limit of a Class - Lower Limit of that Class.
- Class Width = distance between lower (or upper) limits of consecutive classes.
For instance, the class width in the frequency distribution shown is $6 - 1 = 5$.
- Class Length = Class Width — 1.
- **Mid-Points of Class Intervals:**



Stack and draw bars

$$\text{Mid-point} = \frac{\text{lower class limit} + \text{upper class limit}}{2}$$

Grouping data (grouping frequency table)

How to construct a group frequency table?

Step 1: Choose the number of classes/intervals. [Usually between 5 and 10 and usually given]

Step 2: Find the range. $R = \text{max} - \text{min}$

Step 3: Calculate the approximate class width.

$$\text{class width} = \left\lceil \frac{R}{\# \text{ of calsses}} \right\rceil$$

where $\lceil . \rceil$ is the ceiling function. For example, $\lceil 2.8 \rceil = 3$, $\lceil 1.5 \rceil = 2$, and $\lceil 6.1 \rceil = 7$.

Example (26): Given the following grades for 40 students: 94, 82, 77, 75, 70, 62, 89, 57, 35, 92, 80, 85, 55, 45, 69, 75, 72, 90, 82, 62, 50, 44, 67, 53, 82, 90, 91, 57, 47, 82, 70, 82,78, 62, 49, 81, 61, 55, 70, 80.

Construct grouped Frequency tables with 5 classes.

a) Draw a histogram.

Interval	f	r.f	Mid-point

b) Describe the shape of the distribution of the grades?

c) Find the proportion of the student that have grades 71 or higher?

d) Find the proportion of the student that have grades less than 59?

Example (27): $r.f = 0.75$ and $\sum f = 40$ what is the value of f ?

Exercises (5): Given the following data:

21.8, 20.3, 20.9, 21.9, 20.8, 19.9, 18.8, 18.9, 19.8, 19.9, 18.9, 18, 18.7, 18.1, 18.2, 19.4, 19.5, 19.5, 18.6.

Draw stem and leaf plot?

Exercises (6): Given the following data:

8, 8, 9, 9, 10, 10, 7, 7, 7, 8, 8, 9, 10, 10, 10, 10, 10, 8, 6, 6.

- 1) Construct a Frequency table.
- 2) Draw a bar chart.
- 3) Describe the shape of the data?

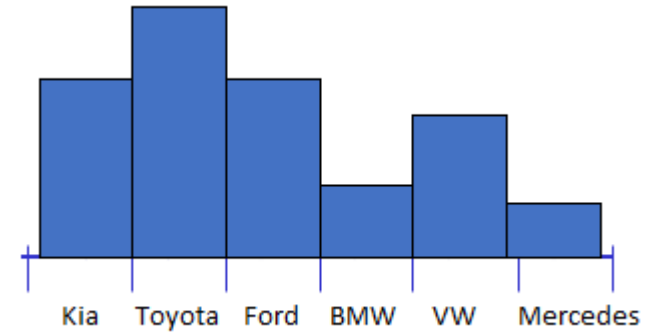
Exercises (7): Given the following data:

1, 2, 1, 0, 2, 2, 1, 1, 0, 0, 2, 2, 1, 1, 0, 0, 1, 2, 1, 1. Where the leaf unit = 0.1.

- a) Construct a stem and leaf plot.
- b) Proportion of measurement less than 1?
- c) Proportion of measurement less than 2?
- d) Describe the shape of the data?

Example (28): Describe the shape of the following Car Type Histogram?

- a) Skewed to the left.
- b) Skewed to the right.
- c) Symmetric.
- d) Unknown.



Exercises (8): True or False:

- 1) A histogram is a graphical representation of the frequency distribution of a dataset.
- 2) In a bar chart, the height of each bar represents the frequency or relative frequency of a category.
- 3) A line chart is best suited for displaying categorical data.
- 4) The shape of a distribution can be described as symmetric, skewed left, or skewed right.
- 5) The area of each section in a pie chart corresponds to its proportion in the dataset.