Topic 4

DESCRIPTIVE STATISTICS

Contents:

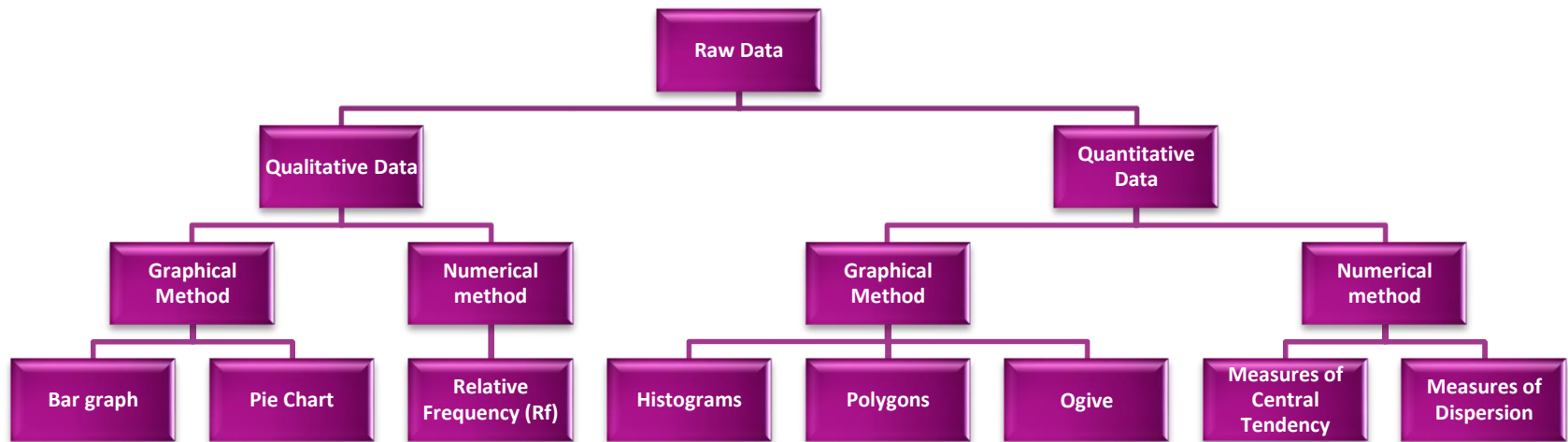**4.1 Introduction**

4.2 Organizing data

4.3 Measurement of central tendency and dispersion

1

Subtopics:

- 4.2.1 RAW DATA
- 4.2.2 ORGANIZING AND DISPLAYING QUALITATIVE DATA
- 4.2.3 ORGANIZING AND DISPLAYING QUANTITATIVE DATA

OVERVIEW

4.2.1 RAW DATA

Discrete variables

Values that can be counted

- The number of cars sold per month by a car sales executive.
- The number of students who attend Statistics class.
- The number of burgers sold per day in McDonalds.
- The number of books read by Chong per year.
- The number of sixes occur when a dice is tossed three times.

Continuous variables

Values that are measured in certain intervals

- The amount of milk that cows produce.
- The heights of children in Kindergarten Sunshine.
- The weight of engineering students.
- The temperature in a frozen room in a restaurant.
- The prices of books in a book store.

4.2.1 RAW DATA

Definition of raw data:

- ❖ Data that have been collected or recorded but have not been arranged or processed yet
- ❖ Ungrouped data

4.2.1   RAW DATA

Example 1

Qualitative Data:

• Type of blood for 10 MMU students:

A	O	A	AB	A
B	B	O	O	AB

Quantitative Data:

• The weight (in kg) of 12 engineers in ABC semiconductor:

56	78	49	55	65	49
101	89	72	67	77	50

4.2.2 ORGANIZING AND DISPLAYING QUALITATIVE DATA

4.2.2.1 Frequency Distributions

Definition of frequency distribution:

Distribution that lists **all** the **categories** and the **number of elements** that belong to each of the categories.

Example 2

Below data set gives the preferred leisure-time physical activity for each of 20 Malaysian women. The following coding is used:
W = walking, T = weight training, C = cycling, G = gardening, A = aerobics

W	T	A	W	G	T	W	W	C	W
T	W	A	T	T	W	G	W	T	A

Construct a frequency distribution table for these data.

Solution:

1) Classified variable into categories (1st column)
2) Read each of the response
3) Mark a tally (2nd column)
4) Record the total of the tallies. (3rd column)

Category	Tally	Frequency
W	‖‖‖ ‖‖	8
T	‖‖‖	6
C			1
G	||	2
A	|||	3

8

Exercise:

A sample was taken of 15 Foundation Engineering students who are going to Beta level. Each of the students was asked which of the majors he or she intended to choose. The responses of these students are as follows

telecommunications	electronics	telecommunications
computer	telecommunications	electronics
electronics	electronics	computer
robotics	computer	robotics
telecommunications	telecommunications	robotics

Construct a frequency distribution table for these data.

Solution:

Category	Tally	Frequency
Telecommunications	⊬⊬	5
Computer	|||	3
Electronics	||||	4
Robotics	|||	3

9

4.2.2 ORGANIZING AND DISPLAYING QUALITATIVE DATA

4.2.2.2 Relative Frequency and Percentage Distributions

Relative Frequency:

• Show the fraction of total frequency belongs to a corresponding category:

$$\text{Relative frequency of a class} = \frac{\text{Frequency of a class}}{\text{Total number of observations}}$$

Percentage distribution:

• Shows the percentages for all categories:

$$\text{Percentage} = \text{Relative frequency of a class} \times 100$$

Example 3

Determine the relative frequency and percentage distributions for the following data.

Solution:

Category	Frequency (f)	Relative frequency (Rf)	Percentage (%)
W	8	$\frac{8}{20} = 0.4$	0.40(100) = 40
T	6	$\frac{6}{20} = 0.3$	0.30(100) = 30
C	1	$\frac{1}{20} = 0.05$	0.05(100) = 5
G	2	$\frac{2}{20} = 0.1$	0.10(100) = 10
A	3	$\frac{3}{20} = 0.15$	0.15(100) = 15
TOTAL	20	1.00	100

Exercise:

Determine the relative frequency and percentage distributions for the following data.

Category	Tally	Frequency
Telecommunications	⌶⌶⌶⌶	5
Computer	|||	3
Electronics	||||	4
Robotics	|||	3

Solution:

Category	Frequency (f)	Relative frequency (Rf)	Percentage (%)
Telecommunications	5	$\frac{5}{15} = 0.33$	0.33(100) = 33
Computer	3	$\frac{3}{15} = 0.2$	0.2(100) = 20
Electronics	4	$\frac{4}{15} = 0.27$	0.27(100) = 27
Robotics	3	$\frac{3}{15} = 0.2$	0.2(100) = 20
TOTAL	15	1.00	100

4.2.2 ORGANIZING AND DISPLAYING QUALITATIVE DATA

4.2.2.3 Graphical Representations for Qualitative Data
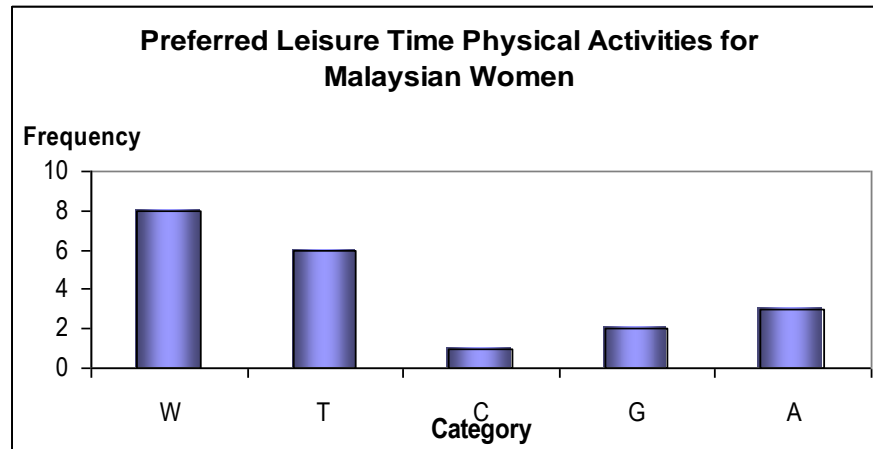
4.2.2.3.1 Bar graph

- Represents a frequency distribution and use for categorical data.
- The height of the bars represents the frequency and can be drawn either vertically or horizontally.
- The small gap is left in between two bars and all the bars are same width.

Example 4

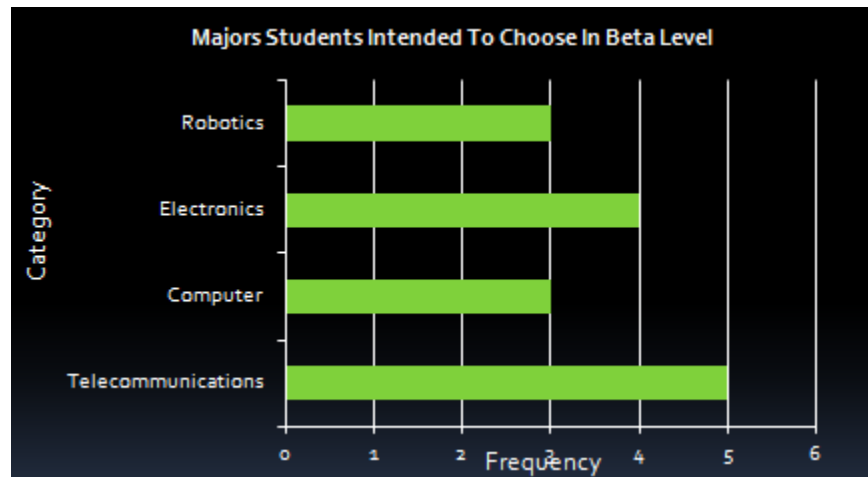Construct the bar graph for the following data.

Solution:

Category	Frequency
W	8
T	6
C	1
G	2
A	3

Exercise:

Construct the bar graph for the following data.

Category	Frequency
Telecommunications	5
Computer	3
Electronics	4
Robotics	3

Solution:

Majors Students Intended To Choose In Beta Level

4.2.2 ORGANIZING AND DISPLAYING QUALITATIVE DATA

4.2.2.3 Graphical Representations for Qualitative Data

4.2.2.3.2 Pie Chart

- Illustrate proportion of the whole data set for various categories
- It is round in shape and contains more than one wedge (or V-shaped piece)
- Calculate the angle sizes for various categories
- Angle = Relative Frequency x 360

Example 5

Construct the pie chart for the following data.

Solution:

Category	Frequency (f)	Relative frequency (Rf)	Angle size (°)
W	8	$\frac{8}{20} = 0.4$	0.40(360°) = 144°
T	6	$\frac{6}{20} = 0.3$	0.30(360°) = 108°
C	1	$\frac{1}{20} = 0.05$	0.05(360°) = 18°
G	2	$\frac{2}{20} = 0.1$	0.10(360°) = 36°
A	3	$\frac{3}{20} = 0.15$	0.15(360°) = 54°

A pie chart showing: W 40%, T 30%, A 15%, G 10%, C 5%
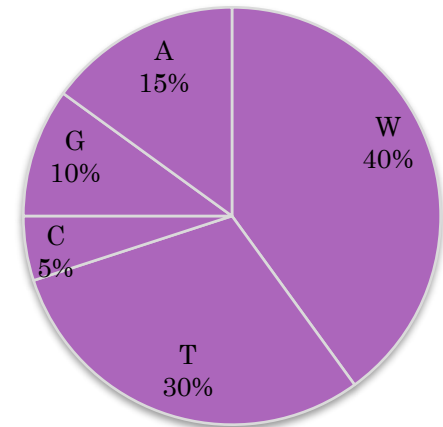
4.2.3 ORGANIZING AND DISPLAYING QUANTITATIVE DATA

4.2.3.1 Frequency Distributions

Frequency distribution table

- Lists all the classes and their frequencies of the values that belong to each class
- Data provided in this form is called grouped data

Class

- An interval that includes all the values that fall within two numbers, the lower and upper limits

Class boundary

- Given by the midpoint of the upper limit of one class and the lower limit of the next class

Class width = Upper boundary – Lower boundary

$$\text{Class midpoint} = \frac{\text{Lower limit} + \text{Upper limit}}{2} \text{ or } \frac{\text{Lower boundary} + \text{Upper boundary}}{2}$$

Example 6

Weekly earnings of 100 employees in *ABC* Semiconductors:

Weekly earnings (RM)	Tally	Number of Employees (f)
301- 400	‖‖‖ ||||	9
401- 500	‖‖‖ ‖‖‖ ‖‖‖ |	16
501- 600	‖‖‖ ‖‖‖ ‖‖‖ ‖‖‖ ‖‖‖ ‖‖‖ |||	33
601- 700	‖‖‖ ‖‖‖ ‖‖‖ ‖‖‖	20
701- 800	‖‖‖ ‖‖‖ ||||	14
801- 900	‖‖‖ |||	8

Find the class midpoint, class boundaries and class width.

Solution:

Class limits	Class boundaries	Class Width	Class Midpoint
301- 400	300.5 – 400.5	100	350.5
401- 500	400.5 – 500.5	100	450.5
501- 600	500.5 – 600.5	100	550.5
601- 700	600.5 – 700.5	100	650.5
701- 800	700.5 – 800.5	100	750.5
801- 900	800.5 – 900.5	100	850.5

STEPS TO CONSTRUCT A FREQUENCY DISTRIBUTION TABLE:

1
- Determine number of classes, *c*
- Sturge's Formula: *c* = 1 + 3.3 *log n*
  where n = number of observations

2
- Determine the class width, $\text{Class width} = \dfrac{\text{Largest value} - \text{smallest value}}{\text{number of classes}}$

3
- Determine the lower limit of the first class

4
- Determine class boundary

5
- Determine number of observations that fall in each class

Example 7:

Below is the payrolls (in million dollars) for 30 major league soccer teams. Construct a frequency distribution table.

51	70	79	75	72	55	25	38	74	54
37	15	56	17	77	43	16	15	72	92
25	30	24	46	47	46	45	38	81	49

Solution:

1) $C = 1 + 3.3 \log 30 = 5.87$ (Choose either 5 or 6)

2) Class width $= \dfrac{92 - 15}{6} = 12.83 (\text{round up}) = 13$

3) Let's choose 15 (the smallest data value) as the starting point.

Class limits	Class boundaries	Tally	Frequency
15 – 27	14.5 – 27.5	++++ | |	7
28 – 40	27.5 – 40.5	| | | |	4
41 – 53	40.5 – 53.5	++++ | |	7
54 – 66	53.5 – 66.5	| | |	3
67 – 79	66.5 – 79.5	++++ | |	7
80 - 92	79.5 – 92.5	| |	2

4.2.3 ORGANIZING AND DISPLAYING QUANTITATIVE DATA

4.2.3.2 Relative frequency and Percentage Distributions

$$\text{Relative frequency of a class} = \frac{\text{Frequency of a class}}{\text{Total number of observations}}$$

$$\text{Percentage} = (\text{Relative frequency of a class}) \times 100$$

4.2.3 ORGANIZING AND DISPLAYING QUANTITATIVE DATA

4.2.3.3 Cumulative Frequency Distribution

- ❖ Gives the total number of values that fall below the upper boundary of each class.

- ❖ Each class has the same lower limit but a different upper limit.

23

Example 9

Prepare a cumulative frequency distribution for the data below.

Weekly earnings (RM)	Number of Employees (f)
301 - 400	9
401 - 500	16
501 - 600	33
601 - 700	20
701 - 800	14
801 - 900	8

Solution:

Weekly earnings (RM)	Number of Employees (f)	Cumulative Frequency
301 - 400	9	**9**
401 - 500	16	9+16 = **25**
501 - 600	33	9 +16+ 33= **58**
601 - 700	20	9+16+33+20= **78**
701 - 800	14	9+16+33+20+14= **92**
801 - 900	8	9+16+33+20+14+8= **100**

4.2.3 ORGANIZING AND DISPLAYING QUANTITATIVE DATA

4.2.3.4 Graphical Representations of Quantitative Data

❖ Illustrate frequency distribution /relative frequency distribution /percentage distribution: **Histogram and Polygon**

❖ Illustrate cumulative distribution: **Ogive**

4.2.3 ORGANIZING AND DISPLAYING QUANTITATIVE DATA

4.2.3.4 Graphical Representations of Quantitative Data
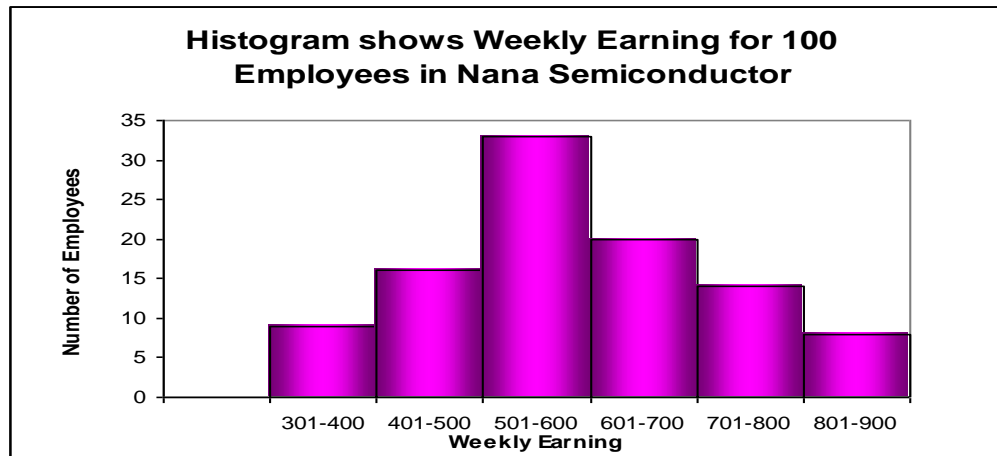
4.2.3.4.1 Histogram

- Similar with bar graph except that histogram has no gap among the bars.
- Height of the bars represents the frequency and can be drawn either vertically or horizontally.

Example 10

Draw a histogram for the data below.

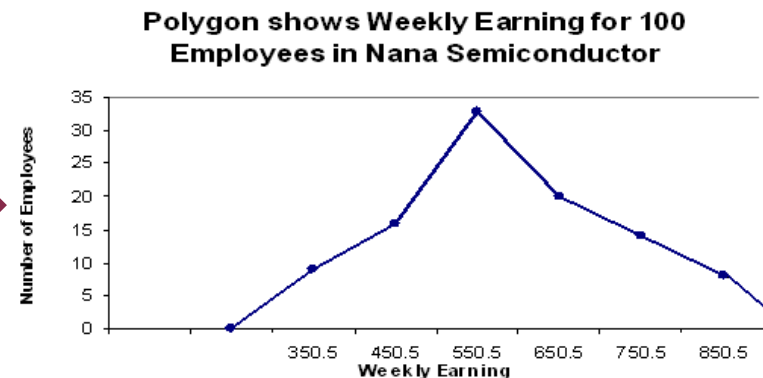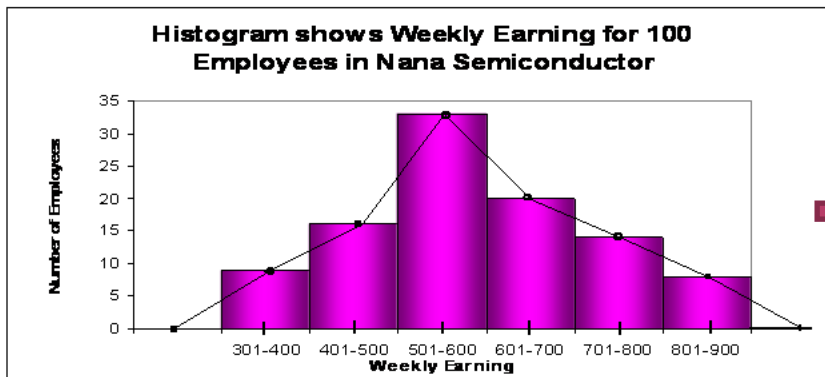| Weekly earnings (RM)	Number of Employees (f)
301 – 400	9
401 – 500	16
501 – 600	33
601 – 700	20
701 – 800	14
801 – 900	8

Solution:

"||" is called truncation. It represents a break. It is used when the scale on the axes does not start at "0", that is, the entire scale on the axes is not shown on the graph.

4.2.3 ORGANIZING AND DISPLAYING QUANTITATIVE DATA

4.2.3.4 Graphical Representations of Quantitative Data

4.2.3.4.2 Polygon

- A graph designed by drawing a straight line joining all the midpoints which are located on the tops of sequential bars in a histogram.

Histogram shows Weekly Earning for 100 Employees in Nana Semiconductor

Polygon shows Weekly Earning for 100 Employees in Nana Semiconductor

4.2.3 ORGANIZING AND DISPLAYING QUANTITATIVE DATA

4.2.3.4 Graphical Representations of Quantitative Data

4.2.3.4.3 Ogive

- A curve obtained only to display cumulative frequency or cumulative relative frequency, or cumulative percentage distribution.

It is obtained as follows:

1
- On a vertical axis we mark cumulative frequency.

2
- On a horizontal axis we mark the upper boundaries of all classes. However, the lower boundary of the first class will be the starting point.

3
- Then, the values for all upper limits for each class are marked on the graph.

4
- Finally, a smooth curve is drawn joining all these points.

Example 10

Draw a frequency ogive for the data below.

Weekly earnings (RM)	Number of Employees (f)
301 – 400	9
401 – 500	16
501 – 600	33
601 – 700	20
701 – 800	14
801 – 900	8

Solution:

Class limits	Class boundaries	f	CF
301 – 400	300.5 – 400.5	9	9
401 – 500	400.5 – 500.5	16	25
501 – 600	500.5 – 600.5	33	58
601 – 700	600.5 – 700.5	20	78
701 – 800	700.5 – 800.5	14	92
801 – 900	800.5 – 900.5	8	100

Ogive shows Weekly Earning for 100 Employees in Nana Semiconductor