



UTM
UNIVERSITI TEKNOLOGI MALAYSIA

FACULTY OF COMPUTING
UTM Johor Bahru

Semester II 2024/2025

Subject : SECI2143 PROBABILITY & STATISTICAL DATA ANALYSIS
Task : Chapter 1 & Chapter 2

INSTRUCTION:

1. This is a **GROUP** assignment. Please clearly write the group members' names and matric numbers on the front page of the submission.
 2. This assignment contributes to **5%** of overall course marks.
 3. Only **HANDWRITTEN** submission is accepted:
 - a. Submissions using any reporting or statistical tools (e.g.: MS Word, MS Excel, etc.,) will be **REJECTED**.
 - b. Make sure the submission is neatly written. Any submission with handwriting that is unreadable, will be **REJECTED**.
 - c. For answers that need to draw graphs, using graph paper(s) is optional. You can use plain paper.
 - d. Round your answers to **TWO** decimal places.
 - e. Please scan/snapshot your work and save it as a **PDF** file.
 4. Submission via eLearning – only **ONE** group member needs to submit on behalf of the group.
-

GROUP 5

	Team Members	Matrix Number
1	ASWINI A/P CHANDRASAGARAN	SX242452ECRHF01
2	IZWAN AZIZ BIN ISMAIL @ ABD MALEK	SX241894ECJHF01
3	FATIN SYAHIRAH BINTI NOR RASHID	SX241920ECRHF01
4	SITI NURNAJIHAH BINTI MOHAMAD ANUAR	SX232351ECRHF04
5	YUARAJ A/L PARTHIPAN	SX241919ECRHF01

QUESTION 1[17 MARKS]

A pizza shop wants to improve its business by collecting customer data:

- Customer name (e.g., "Ford Cash")
- Age group (Child, Teen, Adult, Senior)
- Favorite pizza topping (Pepperoni, Veggie, Cheese)
- Rating of service (1-5 stars)
- Number of slices ordered
- Total bill amount (RM)
- Time spent eating (in minutes)

Answer the following questions:

- a. Which collected customer data are qualitative, and which are quantitative? [7 marks]

a) Which collected customer data are qualitative, and which are quantitative?

Qualitative: i) Customer name
ii) Age group
iii) Favorite pizza topping
iv) Rating of service (1 to 2 stars ≠ 3 to 4 stars (Unequal Intervals).)

Quantitative: i) Number of slices ordered
ii) Total bill amount
iii) Time spent eating (in minutes)

- b. Based on the quantitative data from the answer (a), which is/are discrete, and which is/are continuous? Justify your answers with examples. [3 marks]

15	b) Based on the quantitative data from the answer (a), which is/are discrete, and which is/are continuous? Justify your answers with examples.
	Discrete : i) Number of slices ordered
20	Reason : Discrete data can only take on certain individual values. This means that the data must be in whole numbers, like a slice (1) or three slices (3). Although the number of slices ordered can be in decimals (eg: 1.5, one and a half slices) but this is rare and it's safe to ignore it.
	Continuous : ii) Total bill amount
	iii) Time spent eating (in minutes)
25	Reason : Continuous data can take on any value in a certain range. This means that the data can be in float numbers (decimals). Examples; Total bill amount (RM) totaled RM13.40, and time spent eating (in minutes) was 5.6 minutes each.

- c. Based on the collected customer data, identify the level of measurement for the data in terms of nominal, ordinal, interval, and ratio. Justify your answers with examples. [7 marks]

	<p>c) Based on the collected customer data, identify the level of measurement for the data in terms of nominal, ordinal, interval, and ratio. Justify your answers with examples.</p>
5	<p>Nominal: i) Customer name ii) Favorite pizza topping</p> <p>Reason: Measurement has just one property: classification. The data doesn't necessarily be put in orders.</p>
10	<p>Example: i) Ford Cash, Teal Mosby and Walter White is true, Walter White, Ford Cash and Teal Mosby is also true.</p> <p>ii) Pepperoni, Mushrooms and onions is true, Mushrooms, Pepperoni and Onions is also true.</p>
15	<p>Ordinal: i) Age group ii) Rating of service</p> <p>Reason: Measurement has two properties: classification and order. The data must be in orders.</p> <p>Example: i) Child, Teen, Senior, Adult clearly shows classification, but the order doesn't reflect real life situation. "Senior" must come last.</p> <p>ii) "1 star" means the service is poor, and having "3 stars" at the start of the classification will cause the rating unreliable.</p>
20	<p>Ratio: i) Number of slices ordered ii) Total bill amount (RM) iii) Time spent eating (in minutes)</p> <p>Reason: Data has four properties: classification, order, equal intervals and true zero.</p>
25	<p>Example: - As the interval and ratio were both under Quantitative data, the most important point is the "true zero", which means there's no value before 0. We can't have -3 slices, -12.60 ringgit and -4.2 minutes.</p> <p>- This also explains why there's no interval LOM in the all 7 customers data.</p>

QUESTION 2[13 MARKS]

A pizza shop surveyed 100 customers about their favorite pizza topping:

- Pepperoni (45 customers)
- Veggie (30 customers)
- Cheese (25 customers)

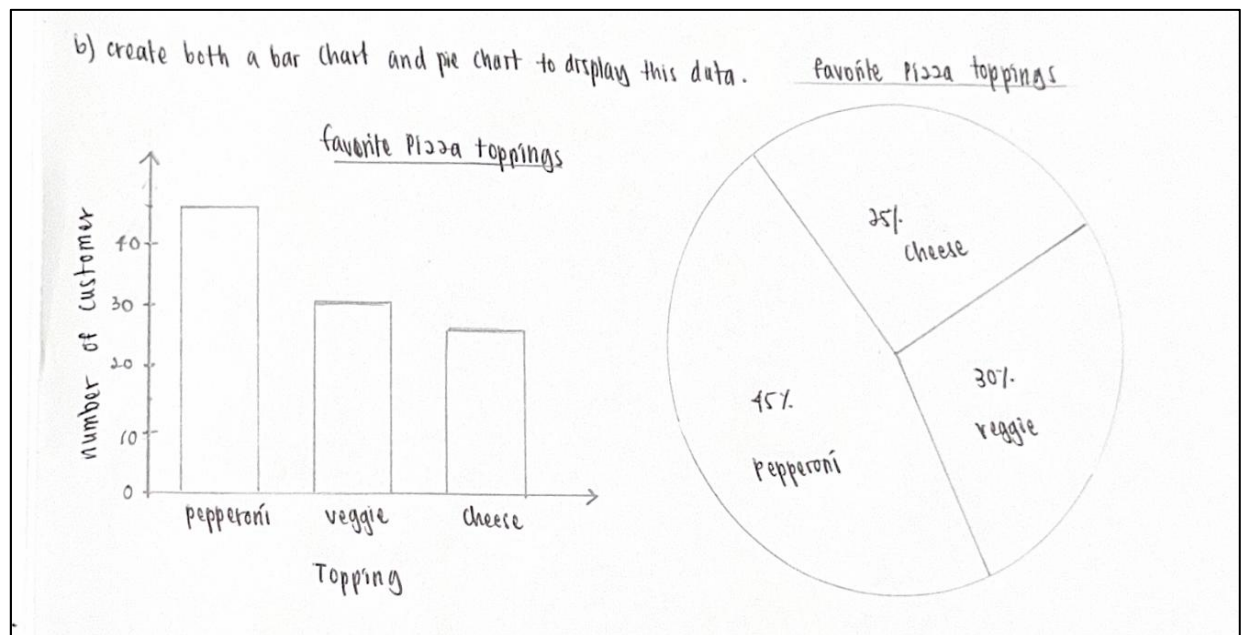
Answer the following questions:

- a. Construct a frequency table including relative frequencies (percentages). [4 marks]

a) construct a frequency table including relative frequencies (percentages).

Topping	frequency	relative frequency
pepperoni	45	45%
veggie	30	30%
cheese	25	25%
total	100	100%

- b. Create both a bar chart and a pie chart to display this data. (Note: Label all axes/sections clearly.) [8 marks]



- c. Explain why a pie chart might be less effective if the shop considers adding five more topping options (making eight total). [1 mark]

c) explain why pie chart might be less effective if the shop considers adding five more topping (making eight total).

= a pie chart becomes less effective when there are too many categories (like eight toppings) because it becomes hard to visually distinguish between similar-sized slices.

QUESTION 3 [15 MARKS]

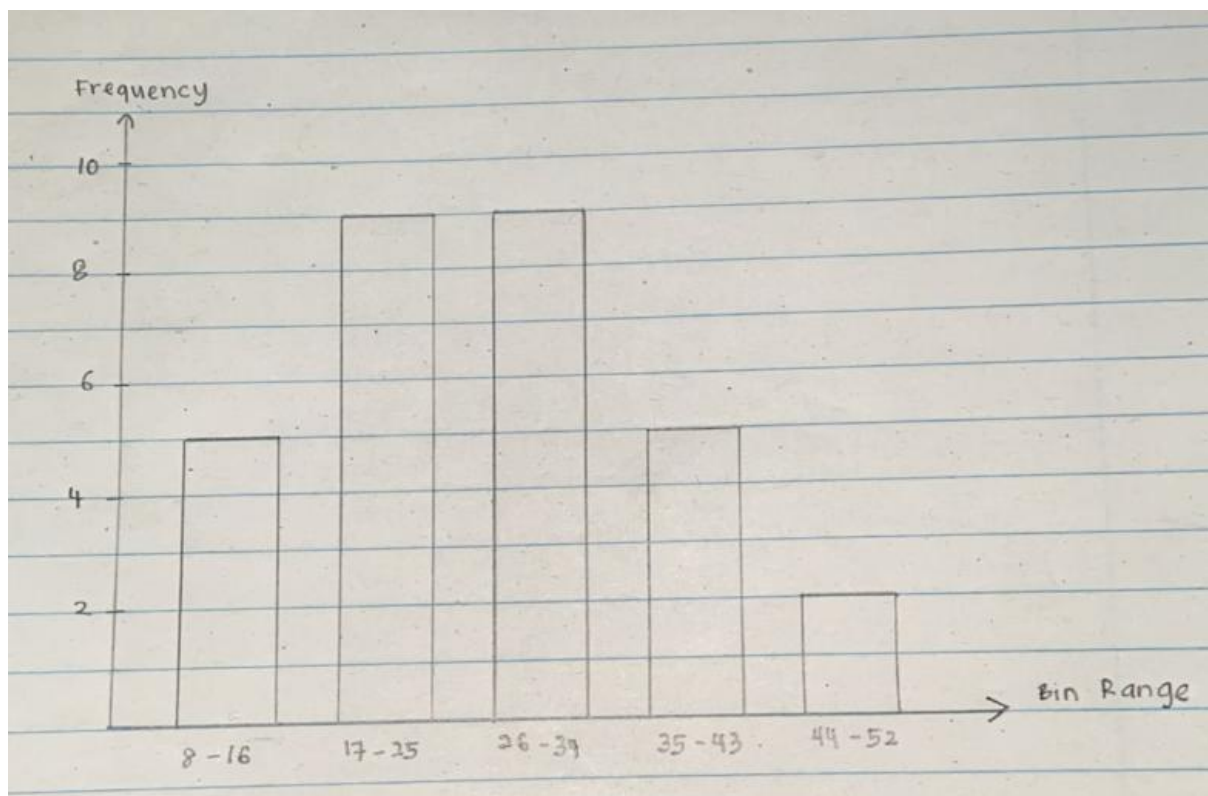
The pizza shop recorded the time (in minutes) taken to prepare 30 orders:

8	10	12	15	16	18	19	20	21	21
22	23	24	25	26	27	28	29	30	31
32	33	34	35	36	38	40	42	45	50

Answer the following questions:

- a. Construct a histogram with five bins (show bin ranges and frequencies). (Note: Label axes and title clearly.) [4 marks]

Bin Range	Frequency	
8 - 16	5	Min = 8
17 - 25	9	Max = 50
26 - 34	9	Range = $50 - 8 = 42$
35 - 43	5	Bin width = $42 \div 5$
44 - 52	2	$= 8.4 \rightarrow \text{round } 9$



- b. Calculate the minimum, first quartile (Q1), median, third quartile (Q3), and maximum based on the recorded time. [5 marks]

b) calculate the minimum, first quartile (Q1), median, third quartile (Q3) and maximum.

min = 8
max = 50

median :
(middle value) 30 value → average median 15 and 16
- 15 value = 26
- 16 value = 27
= Median = $(26 + 27) \div 2 = 26.5$

Q1 = lower half (first 15 value)
middle (8 value)
Q1 = 20

Q3 = upper half (last 15 value)
middle (8 value)
Q3 = 34

Answer	
Min	= 8
Q1	= 20
median	= 26.5
Q3	= 34
Max	= 50

- c. Compute the interquartile range (IQR). [1 mark]

c) compute the IQR

$$\begin{aligned} \text{IQR} &= Q_3 - Q_1 \\ &= 34 - 20 \\ &= 14 \end{aligned}$$

- d. Identify any outliers using the 1.5×IQR Rule. [2 marks]

d) Identify any outliers using the 1.5 × IQR rule.

lower limit = $Q_1 - 1.5 \times \text{IQR}$
 $= 20 - 1.5 \times 14$
 $= -1$

upper limit = $Q_3 + 1.5 \times \text{IQR}$
 $= 34 + 1.5 \times 14$
 $= 55$

- e. Draw a modified box plot showing outliers (if any). [3 marks]

