



UTM
UNIVERSITI TEKNOLOGI MALAYSIA

FACULTY OF COMPUTING
UTM Johor Bahru

Semester II 2024/2025

Subject : SECI1143 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.

Angela Ngu Xin Yi A24CS0226

Tan Xin Tian A24CS0198

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]
- 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]
- 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]

Question 1

Qualitative	Quantitative
Customer name	Rate of service
Age group	Number of slices ordered
Favourite pizza topping	Total bill amount (RM)
	Time spent eating (in minutes)

b. Discrete:

Rating of services is discrete because the data can only take on certain individual values which are specific whole number (1-5). For example, a customer gives a rating of 3 stars but no 3.24 stars.

Number of slices ordered is also discrete because it can only take on certain individual values which are whole numbers. For example, the customer order 4 slices of pizza but not 4.54 slices of pizza.

Continuous:

Total bill amount in RM is continuous because the data can take on any value in certain range including decimals. For example, the total bill is RM 15.65.

Time spent eating in minutes is continuous because it can take on any value in certain range even in fractions or decimals. For example, the customer spent 10.5 minutes eating.

c. Nominal:

Customer name is nominal because it does not have any quantitative value and any ordering of categories is arbitrary. It only indicates difference between customers without direction or exact amount. For example, Adam and Bruce.

Favourite pizza topping is nominal because it does not have quantitative value and there is no ranking between the toppings. For example, customer A likes cheese while customer B likes veggie.

Ordinal:

Age group is ordinal because we know the difference in degree but we do not know actually how much more. For example, we know that senior has older age than teen but we have no way of knowing how much older senior is than teen, maybe 40 or 50 or even 60.

Rating of service is ordinal because it indicates differences and directions but there is no exact amount of differences stated. For example, the customer gives a rating of 4 out of 5 but we could not assume that the pizza is 4 times better than other pizzas.

c. Ratio:

Number of slices ordered is ratio because it has a true zero. For example, the customer order 0 piece of pizza means that no pizza is ordered.

Total bill amount is ratio because it has natural zero despite it is measurable and compatible. For example, RM0 means no amount of money and bill of RM10 is twice expensive than a RM5-bill.

Time spent eating is ratio because it has true zero and there is amount of difference. For example, 0 minute means no time is spent eating and 30 minutes is triple 10 minutes.

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]
- 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]

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

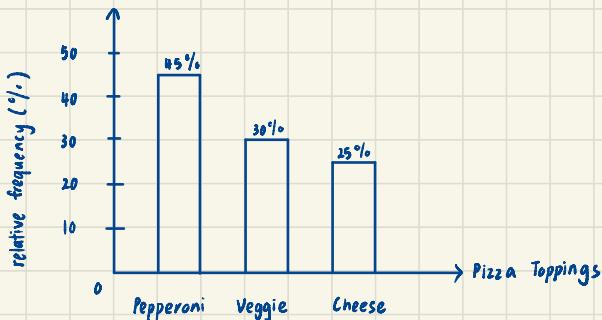
Answer the following questions:

- a. Construct a histogram with five bins (show bin ranges and frequencies). (Note: Label axes and title clearly.) [4 marks]
- b. Calculate the minimum, first quartile (Q1), median, third quartile (Q3), and maximum based on the recorded time. [5 marks]
- c. Compute the interquartile range (IQR). [1 mark]
- d. Identify any outliers using the $1.5 \times \text{IQR}$ Rule. [2 marks]
- e. Draw a modified box plot showing outliers (if any). [3 marks]

Topping	Frequency	Relative Frequency
1. Pepperoni	45	$45 \div 100 \times 100\% = 45\%$
2. Veggie	30	$30 \div 100 \times 100\% = 30\%$
3. Cheese	25	$25 \div 100 \times 100\% = 25\%$
Total	100	100%

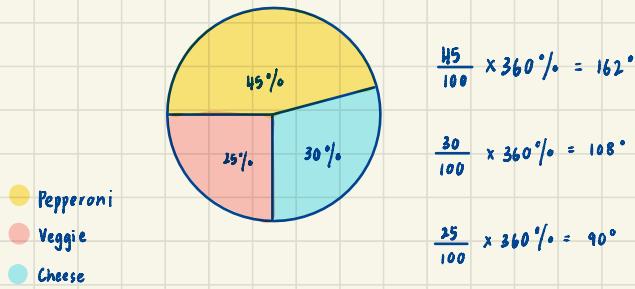
(b) Bar chart

Favourite Pizza Toppings Among 100 Customers



Pie Chart

Percentage of Favourite Pizza Toppings Among 100 Customers

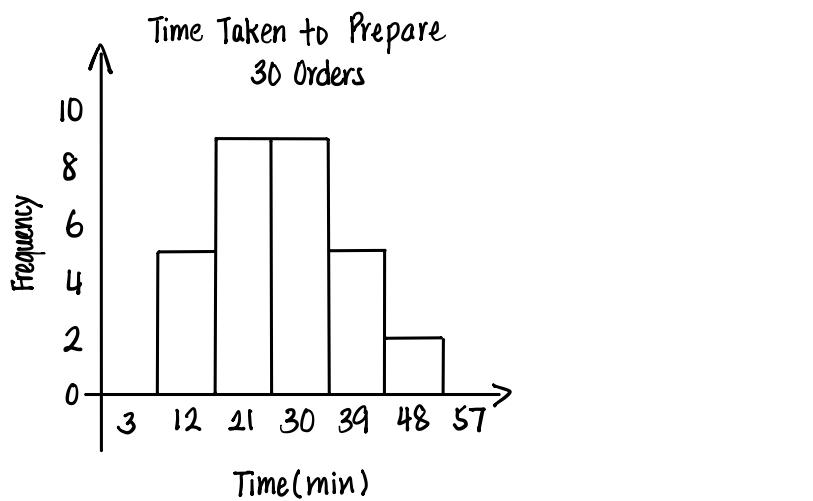


(c) A pie chart becomes less effective with many categories because too many small slices make it hard to compare sizes accurately.

Question 3

Time(min)	Cell Boundaries	Cell Midpoint	Tabulation	Frequency
8-16	7.5-16.5	12		5
17-25	16.5-25.5	21		9
26-34	25.5-34.5	30		9
35-43	34.5-43.5	39		5
44-52	43.5-52.5	48		2

$$\begin{aligned}\text{Class width} &= \frac{50 - 8}{5} \\ &= 8.4 \\ &\approx 9\end{aligned}$$



b. minimum = 8

First Quartile, Q1

$$i = \frac{1}{4} \times 30 = 7.5$$

$$k = 8$$

$$Y[8] = 20$$

$$Q1 = 20$$

Median

$$i = \frac{1}{2} \times 30 = 15$$

$$k = 15$$

$$\frac{Y[15] + Y[16]}{2} = \frac{26 + 27}{2} = 26.5$$

Third Quartile, Q3

$$i = \frac{3}{4} \times 30 = 22.5$$

$$k = 23$$

$$Y[23] = 34$$

$$Q3 = 34$$

Maximum = 50

c. IQR = Q3 - Q1

$$= 34 - 20$$

$$= 14$$

d. Lower Limit : $Q1 - 1.5 \times IQR$

$$= 20 - 1.5 \times 14$$

$$= -1$$

Upper Limit : $Q3 + 1.5 \times IQR$

$$= 34 + 1.5 \times 14$$

$$= 55$$

\therefore no outliers

e.

