UTM

UNIVERSITI TEKNOLOGI MALAYSIA

FACULTY OF COMPUTING

SEMESTER 2

2023/2024

SECI1143 - PROBABILITY & STATISTICAL DATA ANALYSIS

SECTION 02

ASSIGNMENT 1 - CHAPTER 1 & CHAPTER 2

LECTURER: DR. NOORFA HASZLINNA MUSTAFFA

NAME	MATRIC NUMBER
NABIL AFLAH BOO BINTI MOHD YOSUF BOO YONG CHONG	A23CS0252
LUBNA AL HAANI BINTI RADZUAN	A23CS0107
NUR FIRZANA BINTI BADRUS HISHAM	A23CS0156
NAWWARAH AUNI BINTI NAZRUDIN	A23CS0143

Question 1

(a) Population: The entire collection of students of Faculty of Computing

Sample: A subset of the faculty students', for example the undergraduate students from the faculty.

b) i. student's gender (Discrete, categorical, nominal)

ii. Number of siblings student's have (Discrete, numerical, ratio)

iii. Student's household income (Continuous, numerical, ratio)

iv. Student's academic GPA (continuous, numerical, ordinal)

v. Time in hours student's spend to study (continuous, numerical, interval)

vi. Student's ethnicity (Discrete, categorical, nominal)

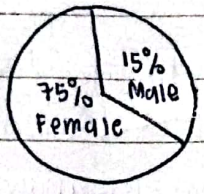
Example of sample data:

Student's gender = F (female)
Number of siblings student's have = 5
Student's household income = RM 10,000
Student's academic GPA = 3.64
Time in hours student's spend to study = 5
Student's ethnicity = Malay

c) Questionnaires distributed to undergraduate students' of Faculty of Computing

d) Academic records from the university's database

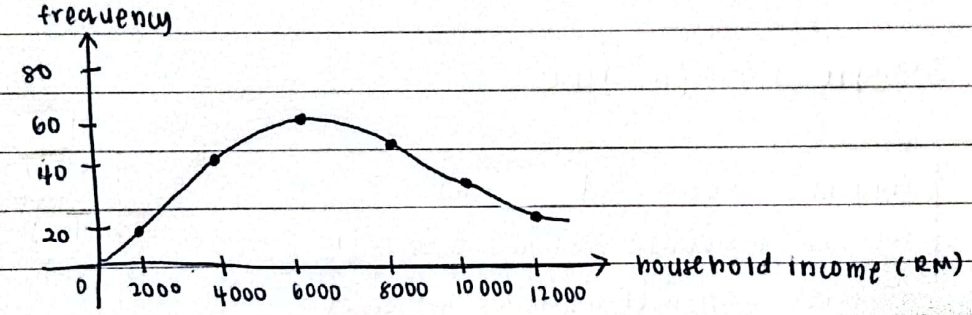
e) i. Student's gender



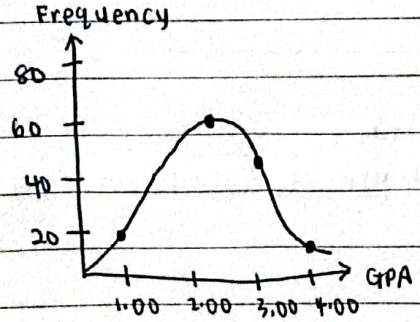
ii.

Number of siblings student's have	Frequency
3	23
4	12
6	7
8	3

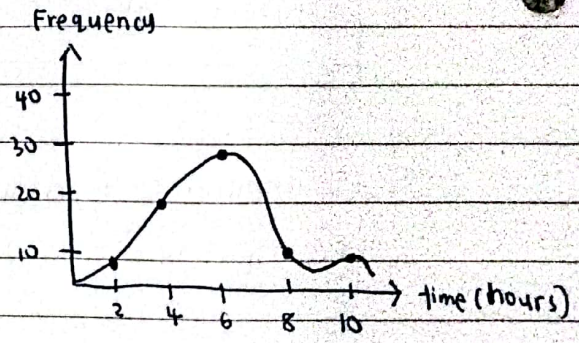
iii. Student's household income



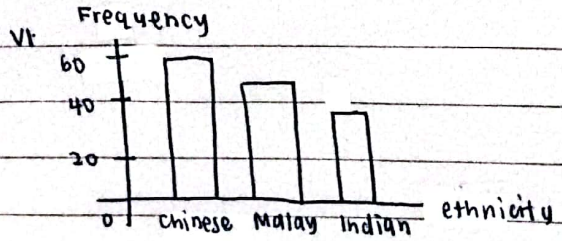
iv. Student's academic GPA



v. Time in hours student's study



vi.



QUESTION 2 [10 marks]

A student trainer worked in training room in the physical education building. The number of students coming in the taped or for treatment of other injuries was recorded each half hour, yielding the following data. Calculate the frequency and relative frequency of the data.

3	3	2	0	4	5	6	4	4	3
2	1	2	3	0	5	5	3	2	3
5	4	1	2	0	3	2	4	2	6

Number of students coming in the taped or for treatment	frequency	relative frequency
0	3	0.10
1	2	0.07
2	7	0.23
3	7	0.23
4	5	0.17
5	4	0.13
6	2	0.07

$$\text{relative frequency} = \frac{\text{frequency}}{\text{total no.}}$$

NO.

Date

Question 3

Weights	Tabulation	Frequency
450	𝍸 𝍸	10
451	II	2
452	II	2
453	𝍸 II	7
454	III	3
455	𝍸 I	6
456	IIII	4
457	𝍸 I	6
458	IIII	4
459	III	3

Weights	Tabulation	Frequency
460	III	3
461	II	2
462	II	2
463	III	3
464	IIII	4
465	III	3
466	IIII	4
467	𝍸	5
468		0
469	IIII	4
470	𝍸	5
471	𝍸 III	8
472	𝍸 IIII	9
473	I	1

Total frequency : 100

Range $= X_n - X_1$

$= 473 - 450$

$= 23$

$i = \frac{R}{1 + 3.322 \log n}$

$= \frac{23}{1 + 3.322 \log(100)}$

$= 3.01 \approx 3$

Trial error, $h = \frac{R}{i}$

$= \frac{23}{3}$

$= 7.66 \approx 8$

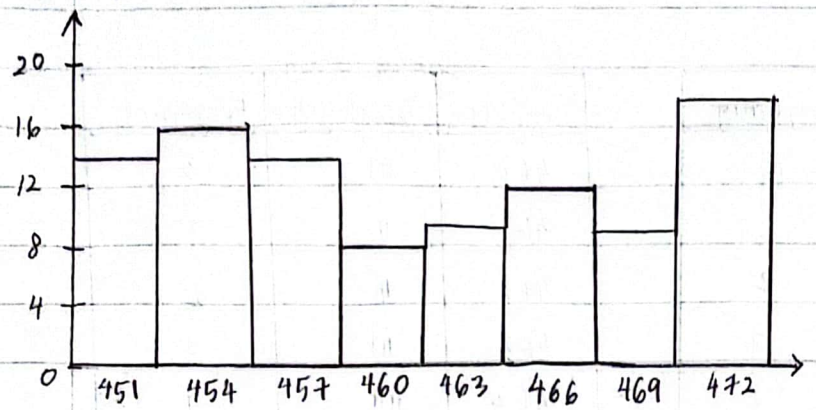
weights	cell boundaries	cell midpoint	Frequency
450 – 452	449.5 – 452.5	451	14
453 – 455	452.5 – 455.5	454	16
456 – 458	455.5 – 458.5	457	14
459 – 461	458.5 – 461.5	460	8
462 – 464	461.5 – 464.5	463	9
465 – 467	464.5 – 467.5	466	12
468 – 470	467.5 – 470.5	469	9
471 – 473	470.5 – 473.5	472	18
Total			100

$MP_L = X_L + \frac{i}{2}$

$= 450 + \frac{3}{2}$

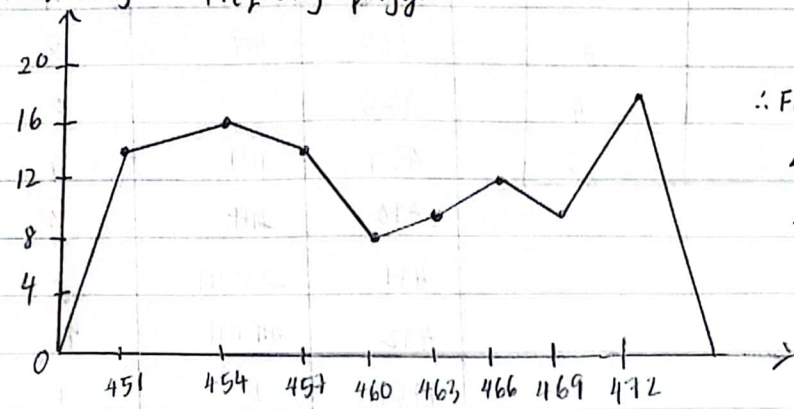
$= 451.5 \approx 451$

a) Histogram



∴ The graph shows that 56 packages don't fulfill the company goal as they already exceed the range given which 450 – 458 grams.

b) Frequency polygon



∴ From the graph, total of 44 packages achieve the company goal to fill with at least 450 to 458 grams.

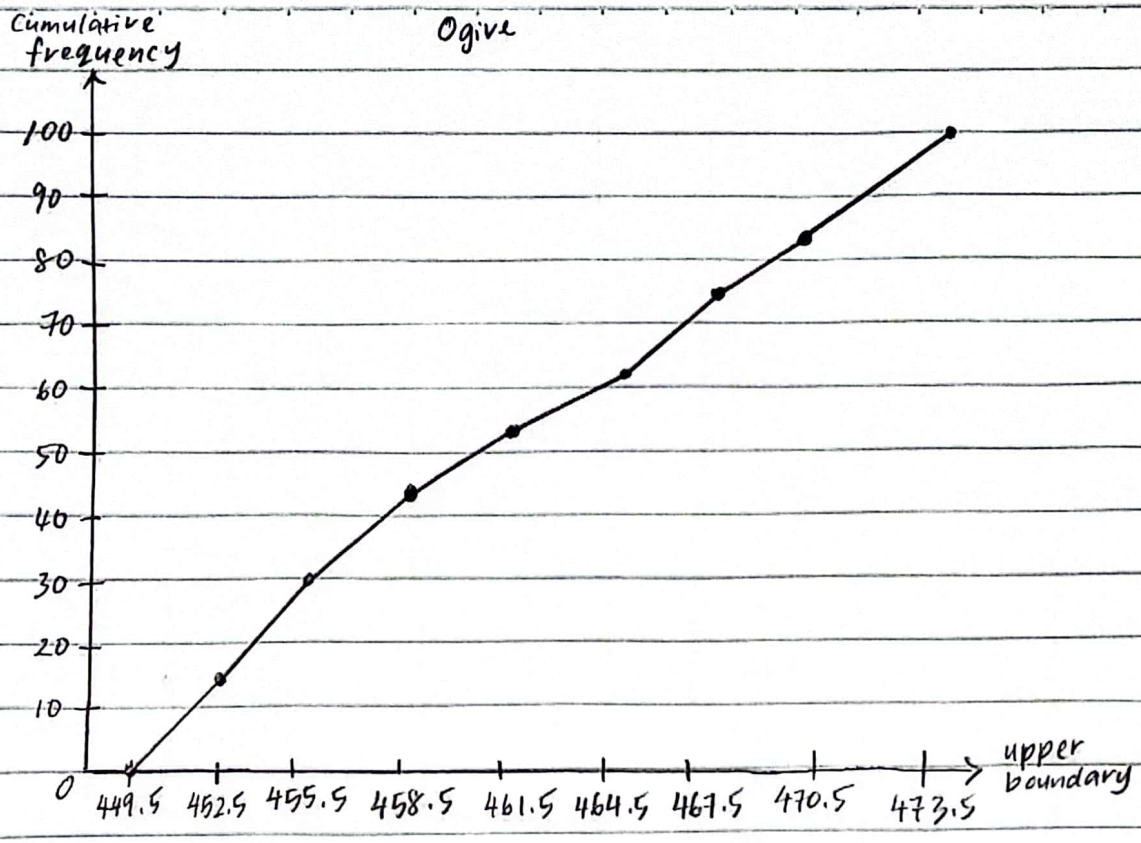
c)

weights	cell boundaries	Frequency	cumulative frequency	relative frequency
450 – 452	449.5 – 452.5	14	14	0.14
453 – 455	452.5 – 455.5	16	30	0.16
456 – 458	455.5 – 458.5	14	44	0.14
459 – 461	458.5 – 461.5	8	52	0.08
462 – 464	461.5 – 464.5	9	61	0.09
465 – 467	464.5 – 467.5	12	73	0.12
468 – 470	467.5 – 470.5	9	82	0.09
471 – 473	470.5 – 473.5	18	100	0.18
Total,		100		1.00

NO.

Date

c) Ogive



∴ There are about 44 package that fulfill the company goal with at least 450 grams and at most 458 grams.

Ref: Date:

ASSIGNMENT 1 : Chapter 1 and Chapter 2 (PSDA)

Question 4

Data value in ~~order~~ Order (left to right):

78	702	765	811	832	855	896	902	905	918	~~919~~ ~~920~~
919	920	923	929	936	938	948	950	956	958	
956	970	972	978	1009	1009	1022	TO 35	1037	1045	
1067	1085	1092	1102	1122	1126	1151	1156	1157	1157	
1162	1170	1195	1195	1196	1217	1237	1311	1333	1340	

Quartile 1 (p=25)

$i = \frac{25}{100}(50)$

$= 12.5 \approx 13$

$\therefore Q1 = Y[13] = 923$

Quartile 2 (p=50)

$i = \frac{50}{100}(50)$

$= 25$

$Q2 = (Y[25] + Y[26]) \div 2$

$= (1009 + 1009) \div 2$

$\therefore = 1009$

Quartile 3 (p=75)

$i = \frac{75}{100}(50)$

$= 37.5 \approx 38$

$\therefore Q3 = Y[38] = 1156$

$iqr = 1156 - 923 = 233$

lower limit $= Q1 - 1.5 \times iqr$

$= 923 - 1.5 \times 233$

$= 573.5$

Upper limit $= Q3 + 1.5 \times iqr$

$= 1156 + 1.5 \times 233$

$= 1505.5$

lowest limit $= 923 - 3 \times 233$

$= 224$

$\therefore$ Since $78 < 224$, thus 78 is an extreme outlier

mean : $\frac{\Sigma x}{50}$

$= \frac{50647}{50} = 1012.94$

Modified Box-Plots :