



UTM
UNIVERSITI TEKNOLOGI MALAYSIA

FACULTY OF COMPUTING

SEMESTER 2 2023/2024

SECI1143 – PROBABILITY & STATISTICAL DATA ANALYSIS

SECTION 3

ASSIGNMENT 1

LECTURER: DR. NOR HAIZAN BT MOHAMED RADZI

STUDENT NAME	MATRIC NO
EVELYN GOH YUAN QI	A23CS0222
JOANNE CHING YING XUAN	A23CS0227
CHUA JIA LIN	A23CS0069

i.a.- A population is the entire group on which data is being gathered and analyzed.
Example of population is UTM students.

- A sample is a subset of the population, selected for study in some prescribed manner.
Example of sample is UTM Faculty of Computing students.

bi.- Discrete data example: Number of students in a class.

Characteristic: It contains finite value that have nothing in-between. For example, we cannot have 30.5 students in a class. We can only have 30 or 31 students in a class.

- Continuous data example: Family income

Characteristic: It contains data that include fractions and decimals. For example, family income can be RM 6550.50.

ii.- Categorical data example: Colour of a car.

Characteristic: It can be put into categories or labels. For example, the colour of a car can be black or white.

- Numerical data example: Age of students.

Characteristic: It can be expressed as number. For example, age of students can be 20 or 21.

iii.- Nominal measurement example: Gender of students.

Characteristic: It is different in kind. For example, gender of students can be male or female.

- Ordinal measurement example: Restaurant ratings.

Characteristic: It is different in degree. For example, a 5 star restaurant is greater than a 4 star restaurant in aspects such as cleanliness and taste of food.

- Ratio measurement example: Income.

Characteristic: It is measurable differences in total amount. For example, income of a student is RM0 whereas income of a lecturer is RM5000.

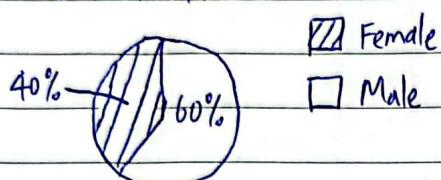
- Interval measurement example: Degrees of temperature.

Characteristic: It is measurable differences in amount. For example, degree of temperature is 30°C.

c. Primary data source are facts and figures newly collected for the project.
Example: Interviews conducted by students for their assignments and projects.

d. Secondary data source are facts and figures already recorded prior to the project.
Example: Movie reviews.

e. Gender of students



2. No. of students come to training room in half hour	Frequency	Relative frequency
0	3	$3/30 = 0.10$
1	2	$2/30 = 0.07$
2	7	$7/30 = 0.23$
3	7	$7/30 = 0.23$
4	5	$5/30 = 0.17$
5	4	$4/30 = 0.13$
6	2	$2/30 = 0.07$
Total	30	1.00

3.	Weight(g)	Tabulation	Frequency	Weight(g)	Tabulation	Frequency	Weight(g)	Tabulation	Frequency
	450		10	460		3	470		5
	451		2	461		2	471		8
	452		2	462		2	472		9
	453		7	463		3	473		1
	454		3	464		4			
	455		6	465		3			
	456		4	466		4			
	457		6	467		5			
	458		4	468		0			
	459		3	469		4			

$$\text{Range} = 473 - 450 \\ = 23$$

$$\text{Cell mid point, } MP_i = X_0 + \frac{i}{2} \\ = 450 + \frac{3.01}{2}$$

$$= 451.505$$

$$\text{Cell interval, } i = \frac{R}{1+3.322 \log n} \\ = \frac{23}{1+3.322 \log(100)} \\ = 3.0088$$

$$\approx 3.01$$

$$i = 1.01, h = \frac{R}{i} = \frac{23}{1.01} = 22.77 \approx 23$$

Cell Midpoint

$$451.505$$

$$i = 3.01, h = \frac{R}{i} = \frac{23}{3.01} = 7.64 \approx 8$$

454.515

$$i = 5.01, h = \frac{R}{i} = \frac{23}{5.01} = 4.59 \approx 5$$

457.525

$\therefore 3.01$ will give the best presentation of the data.

460.535

463.545

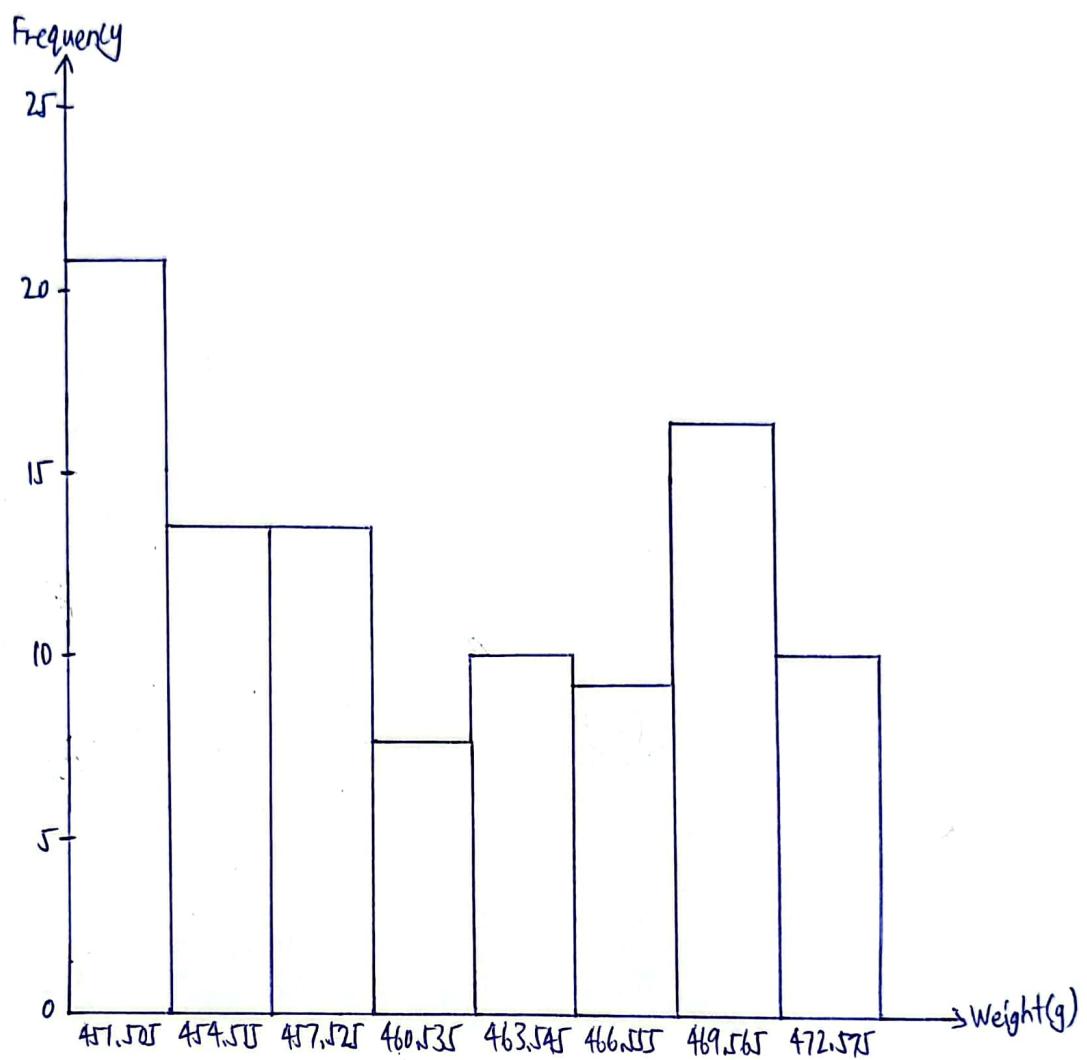
466.555

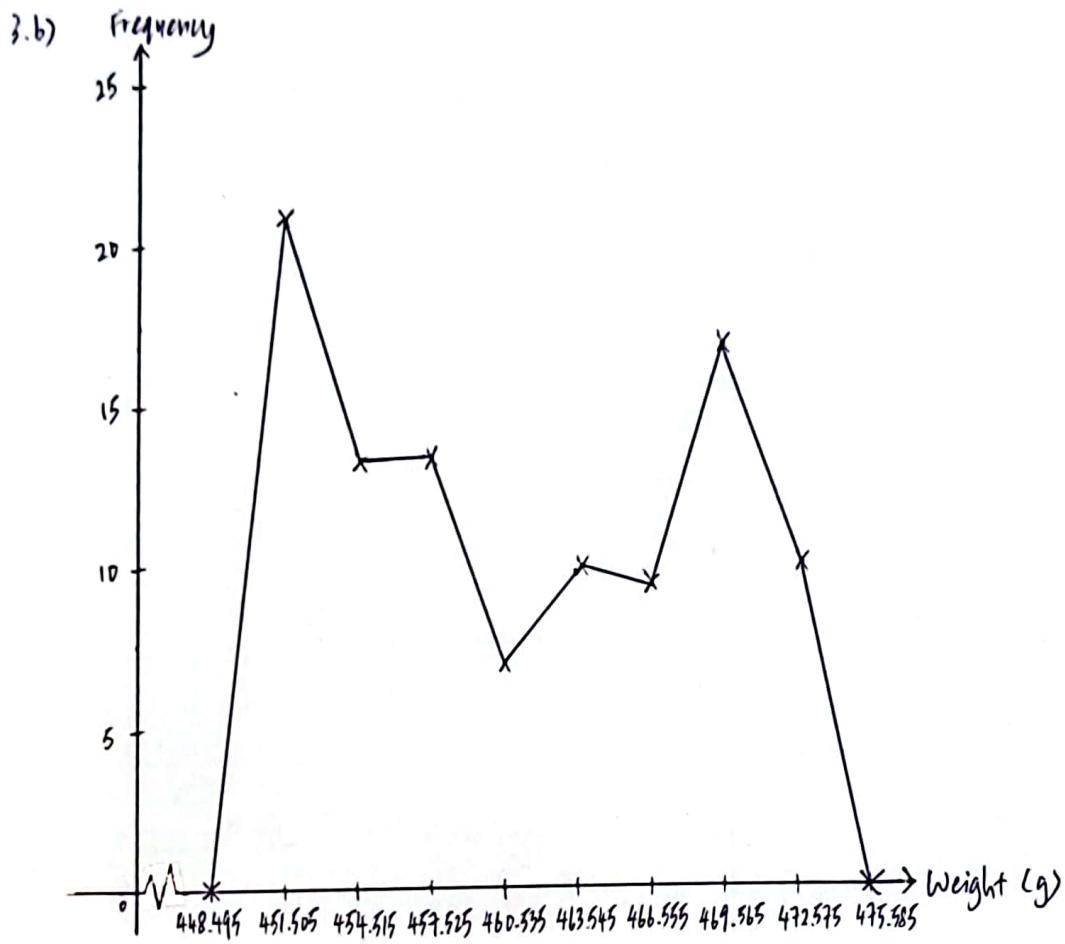
469.565

Cell boundaries	Cell Midpoint	Frequency	Cumulative Frequency
450.00 - 453.01	451.505	21	21
453.01 - 456.02	454.515	13	34
456.02 - 459.03	457.525	13	47
459.03 - 462.04	460.535	7	54
462.04 - 465.05	463.545	10	64
465.05 - 468.06	466.555	9	73
468.06 - 471.07	469.565	17	90
471.08 - 474.08	472.575	10	100



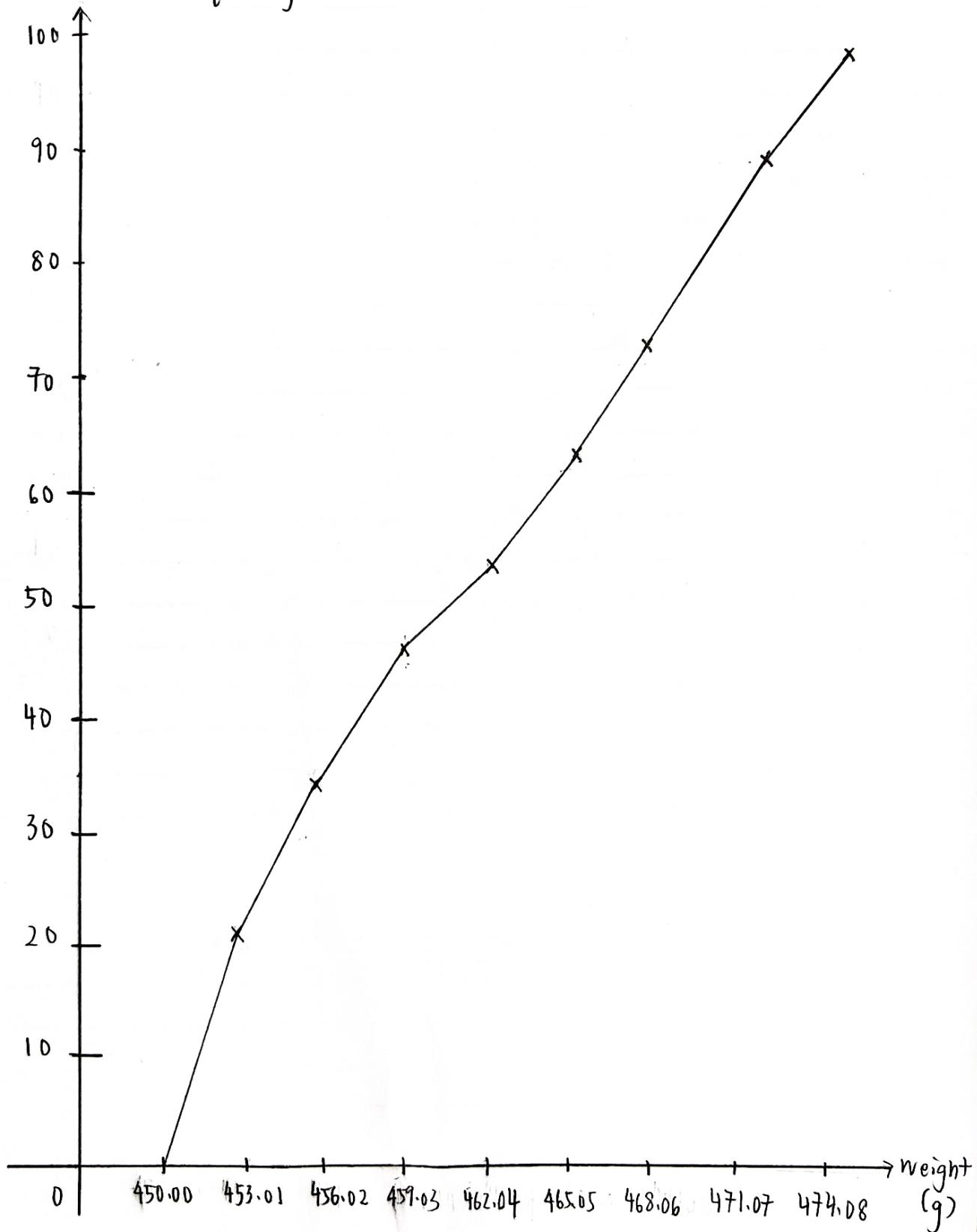
3a. Histogram





c) Ogive

Cummulative frequency



4. Arrange in order:

78, 702, 765, 811, 832, 855, 896, 902, 905, 918
 919, 920, 923, 929, 936, 938, 948, 950, 956, 958
 958, 970, 972, 978, 1009, 1009, 1022, 1035, 1037, 1045
 1067, 1085, 1092, 1102, 1122, 1126, 1151, 1156, 1157, 1157
 1162, 1170, 1195, 1195, 1196, 1217, 1237, 1311, 1333, 1340

$$\frac{25}{100} (50) = 12.5 \approx 13$$

$$Q_1 = 923$$

$$Q_3 - Q_1 = 233$$

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

lower limit:

$$\frac{Y[25] + Y[26]}{2} = \frac{1009 + 1009}{2}$$

$$923 - (1.5 \times 233) = 573.50$$

$$Q_2 = 1009$$

upper limit:

$$1156 + (1.5 \times 233) = 1505.50$$

$$\frac{75}{100} (50) = 37.5 \approx 38$$

$\therefore 78$ is outlier because it is lower than the lower limit

$$Q_3 = 1156$$

