

PSDA (Assignment 3)

Question 1

(a)	Variable	Data Type	Measuring Scales
i)	Case No.	discrete, quantitative	nominal
ii)	Gender	discrete, qualitative	nominal
iii)	Age* (only displayed in years)	continuous, quantitative	ratio
iv)	Nationality	discrete, qualitative	nominal
v)	Date confirmed	discrete, quantitative	interval
vi)	Date discharged	discrete, quantitative	interval
vii)	Hospital	discrete, qualitative	nominal

(b) Secondary data source, because the source is from a website where the data is collected and recorded from researches by the Kementerian Kesihatan Malaysia. Data collected by KKM is primary data source, while those data used back to be published by the website is secondary data source.

(c) i) Gender : Pie chart / Bar graph ii) Age : Histogram / Dot plot iii) Date discharged : Scatter plot

Question 2

(a) i)	Stem	Leaf	ii) 1 st Quartile:	2 nd Quartile:	3 rd Quartile:
	35	5 7 8 9	$P=25, N=30$	$P=50, N=30$	$P=75, N=30$
	36	1 1 3 4 5 6 7 7 7 9	$i = \frac{25}{100}(30) = 7.5$	$i = \frac{50}{100}(30) = 15$	$i = \frac{75}{100}(30) = 22.5$
	37	0 0 0 1 2 2 4 5 7 7 8	$k=8$	$k=15$	$k=23$
	38	0 1 1 3 7	$Y[8] = 36.4$	$\frac{Y[15] + Y[16]}{2} = \frac{37 + 37}{2} = 37.0$	$Y[23] = 37.7$
	Key: 36 1 means 36.1°C		$\therefore Q_1 = 36.4$	$\therefore Q_2 = 37.0$	$\therefore Q_3 = 37.7$

iii) $Q_1 = 36.4, Q_2 = 37.0, Q_3 = 37.7$

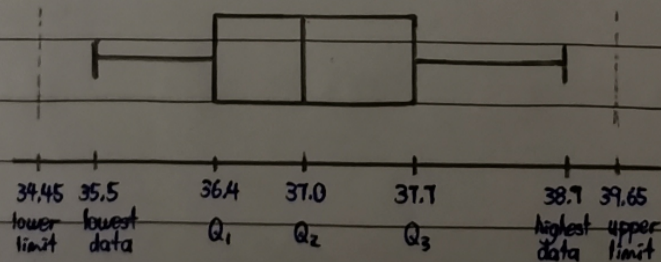
$$IQR = Q_3 - Q_1 = 37.7 - 36.4 = 1.3$$

Max observation = 38.7, Min observation = 35.5

$$\text{Upper limit} = Q_3 + 1.5 IQR = 37.7 + 1.5(1.3) = 39.65$$

$$\text{Lower limit} = Q_1 - 1.5 IQR = 36.4 - 1.5(1.3) = 34.45$$

iv) Since the maximum observation (38.7) falls within upper limit (39.65) and the minimum observation (35.5) falls in front lower limit (34.45), the dataset has no outliers. Hence, all the data are reliable, with the distribution slightly positively skewed (Q_3 farther from Q_2 than Q_1).



No: _____

(b) i) Total frequency

$$= 10 + 27 + 38 + 47 + 86 + 102 + 78 + 56$$

$$= 444$$

Age (years)

Frequency

Relative Frequency

Cumulative Frequency

Relative Cumulative Frequency

30 - under 35

10

$$\frac{10}{444} = 0.0225$$

10

$$0.0225$$

35 - under 40

27

$$\frac{27}{444} = 0.0608$$

$$10 + 27 = 37$$

$$0.0225 + 0.0608 = 0.0833$$

40 - under 45

38

$$\frac{38}{444} = 0.0856$$

$$37 + 38 = 75$$

$$0.0833 + 0.0856 = 0.1689$$

45 - under 50

47

$$\frac{47}{444} = 0.1059$$

$$75 + 47 = 122$$

$$0.1689 + 0.1059 = 0.2748$$

50 - under 55

86

$$\frac{86}{444} = 0.1937$$

$$122 + 86 = 208$$

$$0.2748 + 0.1937 = 0.4685$$

55 - under 60

102

$$\frac{102}{444} = 0.2297$$

$$208 + 102 = 310$$

$$0.4685 + 0.2297 = 0.6982$$

60 - under 65

78

$$\frac{78}{444} = 0.1757$$

$$310 + 78 = 388$$

$$0.6982 + 0.1757 = 0.8739$$

65 - under 70

56

$$\frac{56}{444} = 0.1261$$

$$388 + 56 = 444$$

$$0.8739 + 0.1261 = 1.0000$$

ii) Relative Ogive :

Relative Ogive

Cumulative
Relative Frequency

1.00

0.90

0.80

0.70

0.60

0.50

0.40

0.30

0.20

0.10

0 24.5

29.5

34.5

39.5

44.5

49.5

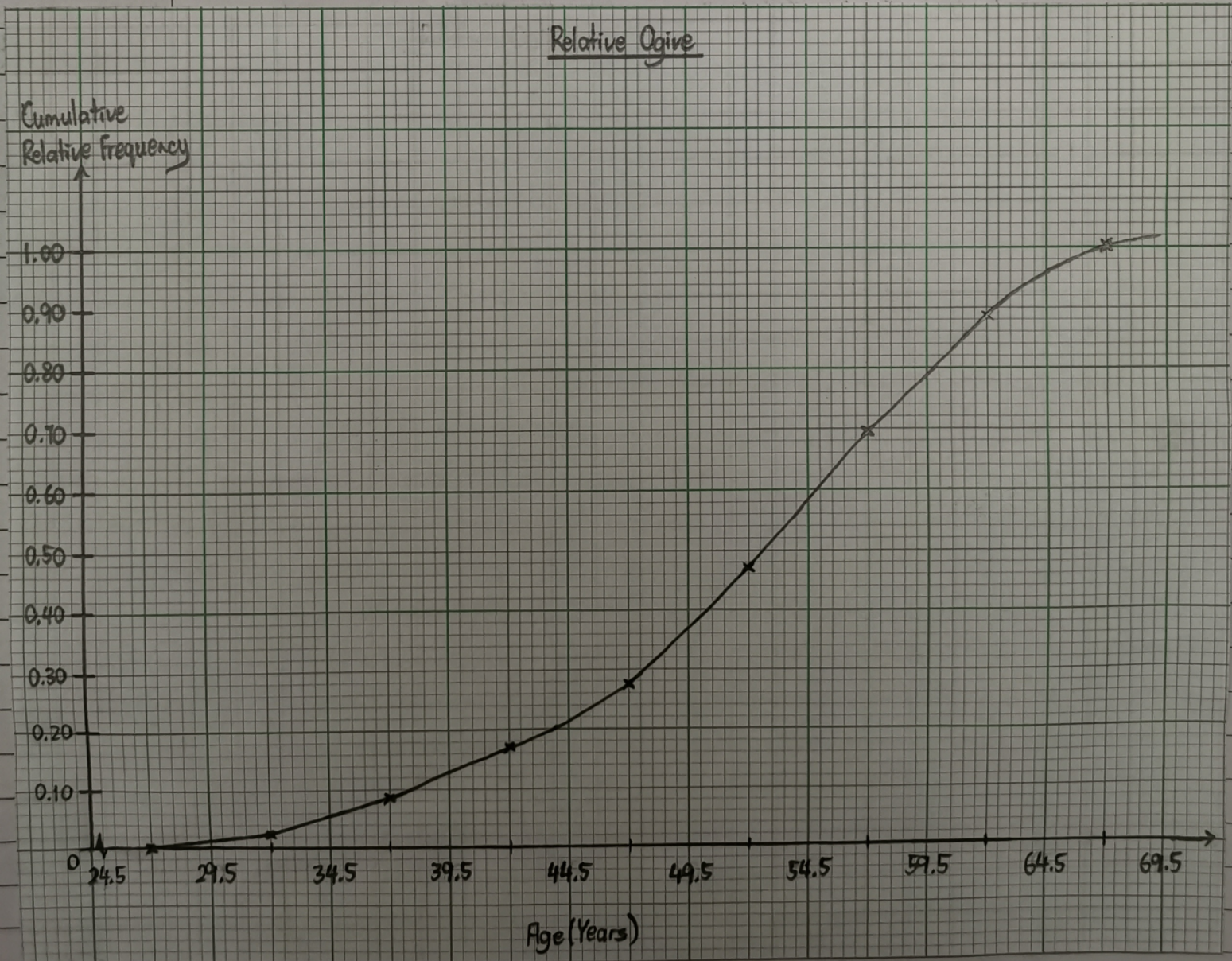
54.5

59.5

64.5

69.5

Age (Years)



Question 3

(a)	Age Group	Midpoint (x)	No. of cases (f)	Cumulative frequency (cf)	Midpoint x no. of cases (fx)	Class Boundaries	
	1-5	3	27	27	81	0.5-5.5	Total no. of cases, $\Sigma f = 1637$
	6-10	8	23	50	184	5.5-10.5	
	11-15	13	43	93	559	10.5-15.5	$\Sigma fx = 81 + 184 + 559 + 1404 + 2829$
	16-20	18	78	171	1404	15.5-20.5	$+ 5040 + 4752 + 5168 + 5074 + 6384$
	21-25	23	123	294	2829	20.5-25.5	$+ 7579 + 10556 + 8631 + 6120 + 3358$
	26-30	28	180	474	5040	25.5-30.5	$+ 1560 + 1162$
	31-35	33	144	618	4752	30.5-35.5	$\Sigma fx = 70441$
	36-40	38	136	754	5168	35.5-40.5	
	41-45	43	118	872	5074	40.5-45.5	
	46-50	48	133	1005	6384	45.5-50.5	
	51-55	53	143	1148	7579	50.5-55.5	
	56-60	58	182	1330	10556	55.5-60.5	
	61-65	63	137	1467	8631	60.5-65.5	
	66-70	68	90	1557	6120	65.5-70.5	
	71-75	73	46	1603	3358	70.5-75.5	
	76-80	78	20	1623	1560	75.5-80.5	
	81-85	83	14	1637	1162	80.5-85.5	

i) Modal class : 56-60

$$l = 55.5, h = 5,$$

$$f_1 = 182, f_0 = 143, f_2 = 137$$

 $\therefore$ Mode

$$= l + h \left[\frac{f_1 - f_0}{2f_1 - f_0 - f_2} \right]$$

$$= 55.5 + 5 \left[\frac{182 - 143}{2(182) - 143 - 137} \right]$$

$$= 57.82$$

ii) $N = 1637, \frac{N}{2} = \frac{1637}{2} = 818.5$

Median class = 41-45

$$L = 40.5, cf_p = 754,$$

$$f_{med} = 118, w = 5$$

 $\therefore$ Median

$$= L + \frac{\frac{N}{2} - cf_p}{f_{med}} (w)$$

$$= 40.5 + \frac{818.5 - 754}{118} (5)$$

$$= 43.23$$

iii) $\Sigma f, N = 1637, \Sigma fx = 70441$

$$\therefore \text{Mean} = \frac{\Sigma fx}{N}$$

$$= \frac{70441}{1637}$$

$$= 43.03$$

(b) i) $N = 20$

$$\text{Mean} = \frac{\sum f}{N}$$

$$\begin{aligned} & 36.5 + 36.5 + 36.5 + 36.6 + 36.6 + 36.6 \\ & + 36.7 + 36.7 + 36.7 + 36.7 + 36.7 + 36.7 \\ & + 36.7 + 36.7 + 36.8 + 36.8 + 36.8 \\ & + 36.9 + 36.9 + 36.9 \end{aligned}$$

20

$$= \frac{734}{20}$$

 $\therefore \text{Mean temperature} = 36.7^\circ\text{C}$

 ii) Mode temperature = 36.7°C
(occur 8 times)

iii) Arranging data:

36.5, 36.5, 36.5, 36.6, 36.6, 36.6, 36.7,

36.7, 36.7, 36.7, 36.7, 36.7, 36.7, 36.7,

36.8, 36.8, 36.8, 36.9, 36.9, 36.9

$$\frac{N}{2} = \frac{20}{2} = 10 \text{ (take 10th and 11th value)}$$

$$\text{Median} = \frac{Y_{[10]} + Y_{[11]}}{2} = \frac{36.7 + 36.7}{2}$$

 $\therefore \text{Median temperature} = 36.7^\circ\text{C}$

iv) Arranging data:

36.5, 36.5, 36.5, 36.6, 36.6, 36.6, 36.7, 36.7,

36.7, 36.7, 36.7, 36.7, 36.7, 36.7, 36.8, 36.8,

36.9, 37.0, 37.0, 37.0

Percentile of value 36.9

$$= \frac{\text{no. of values less than 36.9}}{\text{total no. of values}} (100)$$

$$= \frac{16}{20} (100)$$

$$= 80$$

 $\therefore 36.9$ is the 80th percentile.

v)

Temperature ($^\circ\text{C}$)

Frequency

$$\text{Mean} = \frac{\sum fx}{N}$$

36.5

3

$$\frac{3(36.5) + 3(36.6) + 8(36.7)}{20}$$

20

36.6

3

36.7

8

$$\text{Mean, } \bar{x} = \frac{734.4}{20} = 36.72$$

36.8

2

Standard deviation, s

36.9

1

$$= \sqrt{\frac{\sum f(x - \bar{x})^2}{n - 1}}$$

37.0

3

$$= \sqrt{\frac{3(36.5 - 36.72)^2 + 3(36.6 - 36.72)^2 + 8(36.7 - 36.72)^2 + 2(36.8 - 36.72)^2 + (36.9 - 36.72)^2 + 3(37.0 - 36.72)^2}{20 - 1}}$$

$$\therefore s = \sqrt{\frac{0.472}{19}} = 0.1576$$

vi) Mean, $\bar{x} = 36.72$

$$\text{Skewness} = \frac{\sum f(x - \bar{x})^3}{(n - 1) s^3}$$

$$\begin{aligned} & 3(36.5 - 36.72)^3 + 3(36.6 - 36.72)^3 + 8(36.7 - 36.72)^3 \\ & + 2(36.8 - 36.72)^3 + (36.9 - 36.72)^3 + 3(37.0 - 36.72)^3 \end{aligned}$$

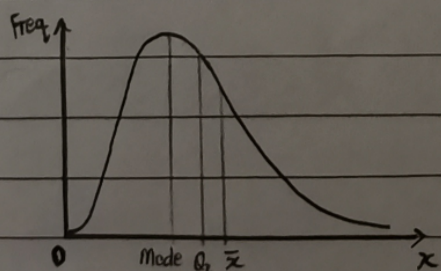
$$(20 - 1) (0.1576)^3$$

$$= \frac{0.03552}{19(0.1576)^3}$$

$$\therefore \text{Skewness} = 0.4776$$

- Since skewness > 0 , the distribution is asymmetrical, pointing in positive direction and positively skewed.

- Also, mode $<$ median $<$ mean

vii) $\bar{x} = 36.72$, $n = 20$

$$\text{Kurtosis} = \frac{\sum f(x - \bar{x})^4}{(n - 1) s^4}$$

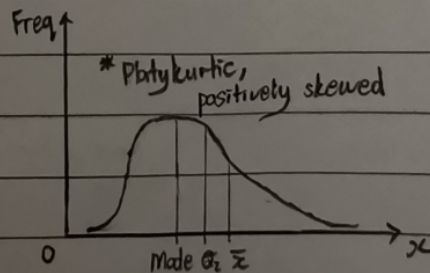
$$\begin{aligned} & 3(36.5 - 36.72)^4 + 3(36.6 - 36.72)^4 + 8(36.7 - 36.72)^4 \\ & + 2(36.8 - 36.72)^4 + (36.9 - 36.72)^4 + 3(37.0 - 36.72)^4 \end{aligned}$$

$$(20 - 1) (0.1576)^4$$

$$= \frac{0.0272224}{19(0.1576)^4}$$

$$\therefore \text{Kurtosis} = 2.3225, \text{ Excess Kurtosis} = 2.3225 - 3 = -0.6775 < 0$$

- Since kurtosis < 3 and excess kurtosis < 0 , the distribution is platykurtic, where tails of distribution are shorter and thinner, and the central peak lower and broader.



Question 4

- (a) Arranging data: 2, 4, 7, 8, 9, 9, 11, 12, 16, 16, 20, 21, 24, 26, 27, 27, 28, 29, 30, 32, 32, 32, 32, 33, 34, 35, 35, 35, 36, 36, 36, 36, 37, 37, 37, 38, 38, 38, 39, 40, 40, 40, 40, 40, 41, 41, 41, 42, 43, 44, 45, 45, 45, 46, 46, 46, 47, 47, 48, 48, 48, 49, 49, 49, 49, 50, 50, 50, 50, 50, 51, 51, 51, 51, 51, 52, 52, 52, 52, 53, 53, 53, 53, 53, 54, 55, 55, 56, 56, 56, 56, 57, 57, 57, 58, 58, 58, 58, 59, 59, 60, 60, 60, 60, 60, 60, 61, 61, 61, 61, 61, 61, 62, 62, 62, 62, 62, 63, 63, 63, 63, 63, 64, 64, 64, 65, 66, 66, 66, 67, 68, 68, 68, 69, 69, 69, 69, 70, 70, 71, 72, 73, 73, 73, 74, 75, 76, 76, 77, 78, 79, 80, 81, 83, 83, 84, 85, 85, 91

$$\text{Range} = x_{\text{highest}} - x_{\text{lowest}}$$

$$= 91 - 2$$

$$= 89$$

$$n = 160$$

$$\text{Mean}, \bar{x} = \frac{\sum fx}{n}$$

$$\begin{aligned} & 2+4+7+8+2(9)+11+12+2(16)+20+21+24+26+2(27)+28+29+30+4(32)+33+34+3(35)+4(36) \\ & +3(37)+3(38)+39+5(40)+3(41)+42+43+44+3(45)+3(46)+2(47)+3(48)+4(49)+5(50)+5(51) \\ & +4(52)+6(53)+54+2(55)+4(56)+3(57)+4(58)+2(59)+6(60)+6(61)+5(62)+5(63)+3(64)+65 \\ & +3(66)+67+3(68)+4(69)+2(70)+71+72+3(73)+74+75+2(76)+77+78+79+80+81+2(83) \\ & +84+2(85)+91 \end{aligned}$$

$$160$$

$$\bar{x} = \frac{8195}{160} = 51.21875 \approx 51.22$$

$$\text{Variance}, s^2 = \frac{\sum f(x - \bar{x})^2}{n - 1}$$

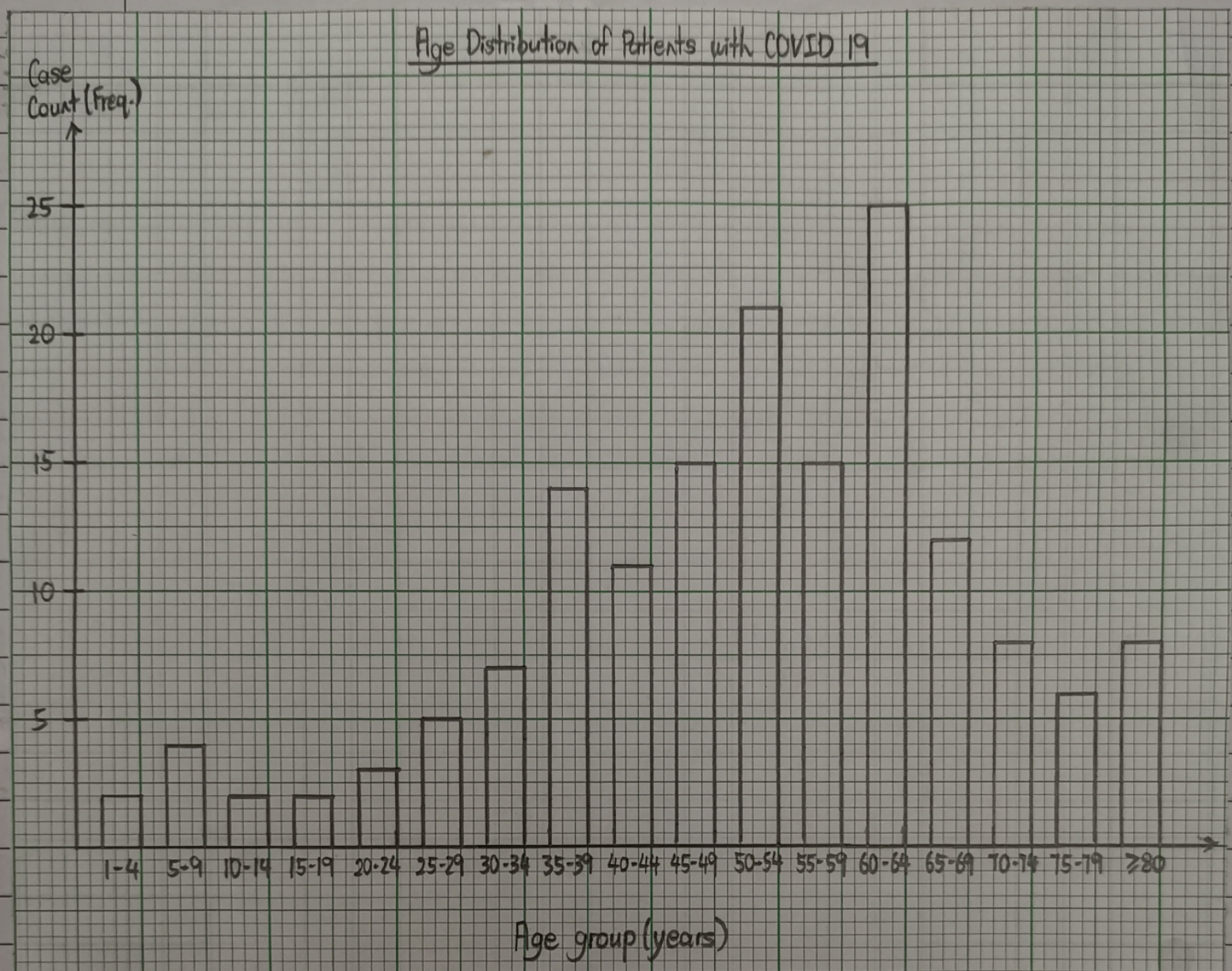
$$\begin{aligned} & (2-51.22)^2 + (4-51.22)^2 + (7-51.22)^2 + (8-51.22)^2 + 2(9-51.22)^2 + (11-51.22)^2 + (12-51.22)^2 + 2(16-51.22)^2 + (20-51.22)^2 + (21-51.22)^2 + (24-51.22)^2 \\ & + (26-51.22)^2 + 2(27-51.22)^2 + (28-51.22)^2 + (29-51.22)^2 + (30-51.22)^2 + 4(32-51.22)^2 + (33-51.22)^2 + (34-51.22)^2 + 3(35-51.22)^2 + 4(36-51.22)^2 \\ & + 3(37-51.22)^2 + 3(38-51.22)^2 + (39-51.22)^2 + 5(40-51.22)^2 + 3(41-51.22)^2 + (42-51.22)^2 + (43-51.22)^2 + (44-51.22)^2 + 3(45-51.22)^2 + 3(46-51.22)^2 \\ & + 2(47-51.22)^2 + 3(48-51.22)^2 + 4(49-51.22)^2 + 5(50-51.22)^2 + 5(51-51.22)^2 + 4(52-51.22)^2 + 6(53-51.22)^2 + (54-51.22)^2 + 2(55-51.22)^2 + 4(56-51.22)^2 \\ & + 3(57-51.22)^2 + 4(58-51.22)^2 + 2(59-51.22)^2 + 6(60-51.22)^2 + 6(61-51.22)^2 + 5(62-51.22)^2 + 5(63-51.22)^2 + 3(64-51.22)^2 + (65-51.22)^2 + 3(66-51.22)^2 \\ & + (67-51.22)^2 + 3(68-51.22)^2 + 4(69-51.22)^2 + 2(70-51.22)^2 + (71-51.22)^2 + (72-51.22)^2 + 3(73-51.22)^2 + (74-51.22)^2 + (75-51.22)^2 + 2(76-51.22)^2 \\ & + (77-51.22)^2 + (78-51.22)^2 + (79-51.22)^2 + (80-51.22)^2 + (81-51.22)^2 + 2(83-51.22)^2 + (84-51.22)^2 + 2(85-51.22)^2 + (91-51.22)^2 \end{aligned}$$

$$160 - 1$$

$$\therefore s^2 = \frac{52791.344}{159} = 332.02$$

$$\text{Standard deviation}, s = \sqrt{s^2} = \sqrt{332.02} = 18.22$$

(b) i) Age Distribution of Patients with COVID 19



ii)	Age Group	Class Boundary	Midpoint(x)	Case Count(f)	Age Group	Class Boundary	Midpoint(x)	Case Count(f)	$\Sigma f = 160$
	1-4	0.5-4.5	2.5	2	45-49	44.5-49.5	47	15	
	5-9	4.5-9.5	7	4	50-54	49.5-54.5	52	21	
	10-14	9.5-14.5	12	2	55-59	54.5-59.5	57	15	
	15-19	14.5-19.5	17	2	60-64	59.5-64.5	62	25	
	20-24	19.5-24.5	22	3	65-69	64.5-69.5	67	12	
	25-29	24.5-29.5	27	5	70-74	69.5-74.5	72	8	
	30-34	29.5-34.5	32	7	75-79	74.5-79.5	77	6	
	35-39	34.5-39.5	37	14	≥80	79.5-84.5	82	8	
	40-44	39.5-44.5	42	11					

(b) ii) Mean, $\bar{x} = \frac{\sum fx}{N}$

$$= \frac{2(2.5) + 4(7) + 2(12) + 2(17) + 3(22) + 5(27) + 7(32) + 14(37) + 11(42) + 15(47) + 21(52) + 15(57) + 25(62) + 12(67) + 8(72) + 6(77) + 8(82)}{160}$$

$$\therefore \bar{x} = \frac{8196}{160} = 51.225$$

Standard deviation, $s = \sqrt{\frac{\sum f(x-\bar{x})^2}{N-1}}$

$$= \sqrt{\frac{2(2.5-51.225)^2 + 4(7-51.225)^2 + 2(12-51.225)^2 + 2(17-51.225)^2 + 3(22-51.225)^2 + 5(27-51.225)^2 + 7(32-51.225)^2 + 14(37-51.225)^2 + 11(42-51.225)^2 + 15(47-51.225)^2 + 21(52-51.225)^2 + 15(57-51.225)^2 + 25(62-51.225)^2 + 12(67-51.225)^2 + 8(72-51.225)^2 + 6(77-51.225)^2 + 8(82-51.225)^2}{160-1}}$$

$$s = \sqrt{\frac{31529.4}{159}} = 18.0023$$

Skewness

$$= \frac{\sum f(x-\bar{x})^3}{(N-1)s^3}$$

$$\begin{aligned} &= \frac{2(2.5-51.225)^3 + 4(7-51.225)^3 + 2(12-51.225)^3 + 2(17-51.225)^3 + 3(22-51.225)^3 + 5(27-51.225)^3 + 7(32-51.225)^3 + 14(37-51.225)^3 + 11(42-51.225)^3 + 15(47-51.225)^3 + 21(52-51.225)^3 + 15(57-51.225)^3 + 25(62-51.225)^3 + 12(67-51.225)^3 + 8(72-51.225)^3 + 6(77-51.225)^3 + 8(82-51.225)^3}{(160-1)(18.0023)^3} \end{aligned}$$

$$\therefore \text{Skewness} = \frac{-535069.4175}{159(18.0023)^3} = -0.5768$$

Kurtosis

$$= \frac{\sum f(x-\bar{x})^4}{(N-1)s^4}$$

$$\begin{aligned} &= \frac{2(2.5-51.225)^4 + 4(7-51.225)^4 + 2(12-51.225)^4 + 2(17-51.225)^4 + 3(22-51.225)^4 + 5(27-51.225)^4 + 7(32-51.225)^4 + 14(37-51.225)^4 + 11(42-51.225)^4 + 15(47-51.225)^4 + 21(52-51.225)^4 + 15(57-51.225)^4 + 25(62-51.225)^4 + 12(67-51.225)^4 + 8(72-51.225)^4 + 6(77-51.225)^4 + 8(82-51.225)^4}{(160-1)(18.0023)^4} \end{aligned}$$

$$\therefore \text{Kurtosis} = \frac{51988603.14}{(160-1)(18.0023)^4} = 3.1131$$

iii) • For graph in 4(b) i), if we get data from the skewness and kurtosis, the skewness (-0.5768) < 0 shows that the distribution is negatively skewed and skewed to the left. If we look at the graph distribution we see most data (bars) located on right side and small values on the left side, this shows the distribution skewed to the left. For graph from China database, looking at the graph shows most data are on the middle and about the same small values for left and right side so we can say that the data are almost symmetric.

• In terms of kurtosis, from 4(b) ii) we get value 3.1131, excess kurtosis = 3.1131 - 3 = 0.1131 > 0, this shows that the distribution is leptokurtic. If we look at the graph from 4(b) i) the data shows that the central peak is higher and sharper while the tails on both sides longer and fatter, hence distribution is leptokurtic. For distribution in China, by looking at the graph we can see a normal distribution curve where the peak is not too high or low, and the tails not too thin or fat, the kurtosis is about 3 in value, excess kurtosis ≈ 0 and the distribution is mesokurtic.

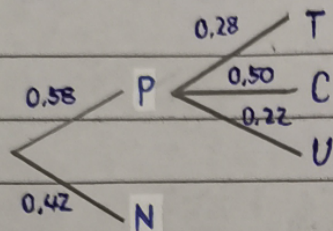
• For the data, a negatively skewed distribution in 4(b) i) will show mode > median > mean which is true because mode is 62 years (25 case count) > median is around 52 years ($\frac{V[80] + V[81]}{2}$) > mean is 51.225 years (from 4(b) ii). While the distribution is China which is almost symmetrical will show mean ≈ median ≈ mode.

Question 5

- (a) Let P = test positive,
 N = test negative,
 T = travelling,
 C = close contact,
 U = unknown cluster

(b) $P(P) = 58\% = 0.58$
 $P(N) = 1 - P(P)$
 $= 1 - 0.58$
 $\therefore P(N) = 0.42$

(c) $P(N) = 0.42$, $n = 72$
 People with negative results
 $= 72(0.42)$
 $= 30.24$
 ≈ 31 people



(d) $(P \cap T) \cup (P \cap C)$

$$= (0.58)(0.28) + (0.58)(0.50)$$

$$= 0.1624 + 0.2900$$

$$= 0.4524$$

$P(P) = 0.58$
 $P(T) = 0.28$
 $P(C) = 0.50$

(e) i) Let D = death,

R = recover,

T = treatment

$P(D) = 1.6\% = 0.016$

$P(R) = 34\% = 0.34$

$\therefore P(T) = 1 - P(D) - P(R)$

$= 1 - 0.016 - 0.34$

$P(T) = 0.644$

ii) $P(T) = 0.644$

$n = 2766$

People under treatment

$= 2766(0.644)$

$= 1781.304$

≈ 1782 people

Question 6

(a) i) $P(X \leq 2)$

$$= P(X=0) + P(X=1) + P(X=2)$$

$$= 0.091 + 0.182 + 0.182$$

$$= 0.455$$

ii) $P(X \geq 5)$

$$= P(X=5) + P(X=6) + P(X=7) + P(X=8)$$

$$= 0.091 + 0.045 + 0.000 + 0.045$$

$$= 0.181$$

iii) $P(3 \leq X \leq 5)$

$$= P(X=3) + P(X=4) + P(X=5)$$

$$= 0.227 + 0.137 + 0.091$$

$$= 0.455$$

iv) Mean, $\mu_x = \sum x p(x)$

$$= 0(0.091) + 1(0.182) + 2(0.182) + 3(0.227) + 4(0.137)$$

$$+ 5(0.091) + 6(0.045) + 7(0.000) + 8(0.045)$$

$\therefore \mu_x = 2.86 \approx 3$ deaths

Variance, $\sigma_x^2 = \sum (x - \mu_x)^2 p(x)$

$$= (0-2.86)^2(0.091) + (1-2.86)^2(0.182) + (2-2.86)^2(0.182)$$

$$+ (3-2.86)^2(0.227) + (4-2.86)^2(0.137) + (5-2.86)^2(0.091)$$

$$+ (6-2.86)^2(0.045) + (7-2.86)^2(0.000) + (8-2.86)^2(0.045)$$

$\therefore \sigma_x^2 = 3.7404 \approx 4$ deaths

(b) i) $n=8, p=0.105, 1-p=1-0.105=0.895$ ii) $n=4, p=0.073, 1-p=1-0.073=0.927$ iii) $n=10, p(\text{die})=0.063, p(\text{recover})=0.937$

$P(X \leq 2)$ $P(X \geq 2)$ $P(X \leq 1)$

$= P(X=0) + P(X=1) + P(X=2)$ $= P(X=2) + P(X=3) + P(X=4)$ $= P(X=0) + P(X=1)$

$= {}^8C_0(0.105)^0(0.895)^8 + {}^8C_1(0.105)^1(0.895)^7$ $= {}^4C_2(0.073)^2(0.927)^2 + {}^4C_3(0.073)^3(0.927)^1$ $= {}^{10}C_0(0.937)^0(0.063)^{10} + {}^{10}C_1(0.937)^1(0.063)^9$

$+ {}^8C_2(0.105)^2(0.895)^6$ $+ {}^4C_4(0.073)^4(0.927)^0$ $\therefore P(X \leq 1) = 1.4747 \times 10^{-10}$

$\therefore P(X \leq 2) = 0.9568$ $\therefore P(X \geq 2) = 0.0289$

iv) $n=5, p(\text{die})=0.056, p(\text{recover})=1-0.056=0.944$ v) $n=56-8-4-10-8-5=21$

$P(2 \leq X \leq 4)$ $p(\text{die})=0.009, p(\text{recover})=1-0.009=0.991$

$= P(X=2) + P(X=3) + P(X=4)$ $P(X=21) = {}^{21}C_{21}(0.991)^{21}(0.009)^0$

$= {}^5C_2(0.944)^2(0.056)^3 + {}^5C_3(0.944)^3(0.056)^2 + {}^5C_4(0.944)^4(0.056)^1$ $P(X=21) = 0.8271$

$\therefore P(2 \leq X \leq 4) = 0.2503$

(c) i) Variable X is a geometric distribution variable where X is the number of trials until the first success trial (dying) is observed. Each trial can result in one of two possible outcomes, either success (dying) or failure (recover). The probability of success (dying) for all trials is the same.

ii) $p=0.028$ iii) $p(\text{dying})=0.047, p(\text{recover})=1-0.047=0.953$

$p(x=3) = (1-0.028)^{3-1}(0.028)$ $p(x=5) = (1-0.953)^{5-1}(0.953)$

$= (0.972)^2(0.028)$ $= (0.047)^4(0.953)$

$\therefore p(x=3) = 0.0265$ $\therefore p(x=5) = 4.6503 \times 10^{-6}$

(d) i) $\mu=4, \sigma=5$ ii) $\mu=4, \sigma=5$ iii) $\mu=4, \sigma=5$

$P(X > 6) = P(Z > \frac{X-\mu}{\sigma})$ $P(2 \leq X \leq 4) = P(\frac{2-4}{5} \leq Z \leq \frac{4-4}{5})$ $P(X < 0) = P(Z < \frac{0-4}{5})$

$= P(Z > \frac{6-4}{5})$ $= P(-0.4 \leq Z \leq 0)$ $= P(Z < -0.8)$

$= P(Z > 0.4)$ $= P(0 \leq Z \leq 0.4)$ $= P(Z > 0.8)$

$= 0.5 - 0.1554$ $\therefore P(2 \leq X \leq 4) = 0.1554$ $= 0.5 - 0.2881$

$\therefore P(X > 6) = 0.3446$ $\therefore P(X < 0) = 0.2119$

