Subject : Date:

Nama : LOH YEW CHONG
Matric number : A19EC0076
Section : 2

1.(a)

Variable	Data type	Measuring
Age	Discrete and quantitative	Ratio Scale
Gender	Discrete and qualitative	Norminal Scale
Nationality	Discrete and qualitative	Norminal Scale
Date confirmed	Discrete and quantitative	Interval Scale
Date discharge	Discrete and quantitativ	Interval Scale
Hospital	Discrete and qualitative	Nominal Scale
Case Number	Discrete and quantitative	Ratio Scale

1.(b) Table 1 is considered as secondary data. This is because the data to make the table is not collected by us. The table already catagories or tabulated for us already.

1.(c) (i) Bar plot
(ii) Histogram
(iii) Scatter plot

2.(a)(i)

stem	leaf
35	5 7 8 9
36	1 1 3 4 5 6 7 7 7 9
37	0 0 0 1 2 2 4 5 7 7 8
38	0 1 1 3 7

key : 35 | 5 = 35.5

2(a)(ii) 1st Quartile $= \frac{25}{100} \times$ total number of patients

$= \frac{1}{4} \times 30 = 7.5 = 8\text{th} = 36.4$

~~Total number = 35.5 + 35.7 + 35.8 + 35.9 + 36.1 + 36.1 + 36.3 + 36.4~~
~~+ 37.0 + 36.5 + 36.6 + 36.7 + 36.7 + 36.7 + 36.9~~

Subject : Date:

2(a)(ii) 2^{nd} quartile = median = $\frac{1}{2} \times$ (number of patients)

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

$= 15$

$= \frac{15th + 16th}{2}$

$= \frac{37.0 + 37.0}{2}$

$= 37.0$

3^{rd} quartile $= \frac{75}{100} \times$ (number of patients)

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

$= 22.5$

$= 23 = 23th = 37.7$

2(a)(iii) Interquartile range $= Q_3 - Q_1$

$= 37.7 - 36.4$

$= 1.3$

lower limit $= Q1 - 1.5 \times IQR$

$= 36.4 - (1.5 \times 1.3)$

$= 34.45$

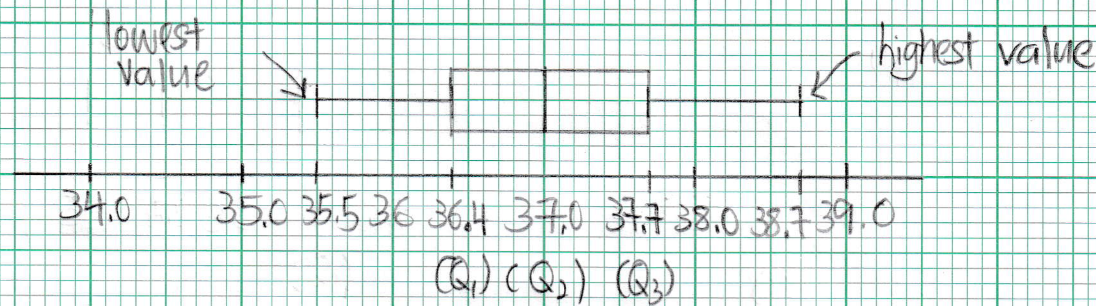
upper limit $= Q_3 + (1.5 \times IQR)$

$= 37.7 + (1.5 \times 1.3)$

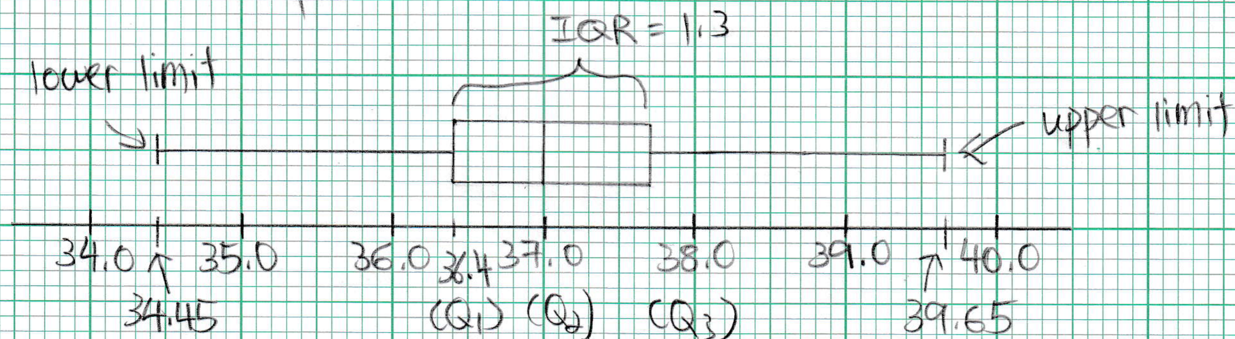
$= 39.65$

2.(a)(iii)

Skeletal box plot



Modified Box plot



Subject :

Date:

2.(a)(iv) The temperature that is greater than 39.65 °C and lower than 34.45 °C will affect the mean of the dataset. It is skewed to the right because the $(Q_3 - Q_2)$ is greater than $(Q_2 - Q_1)$.

Subject :

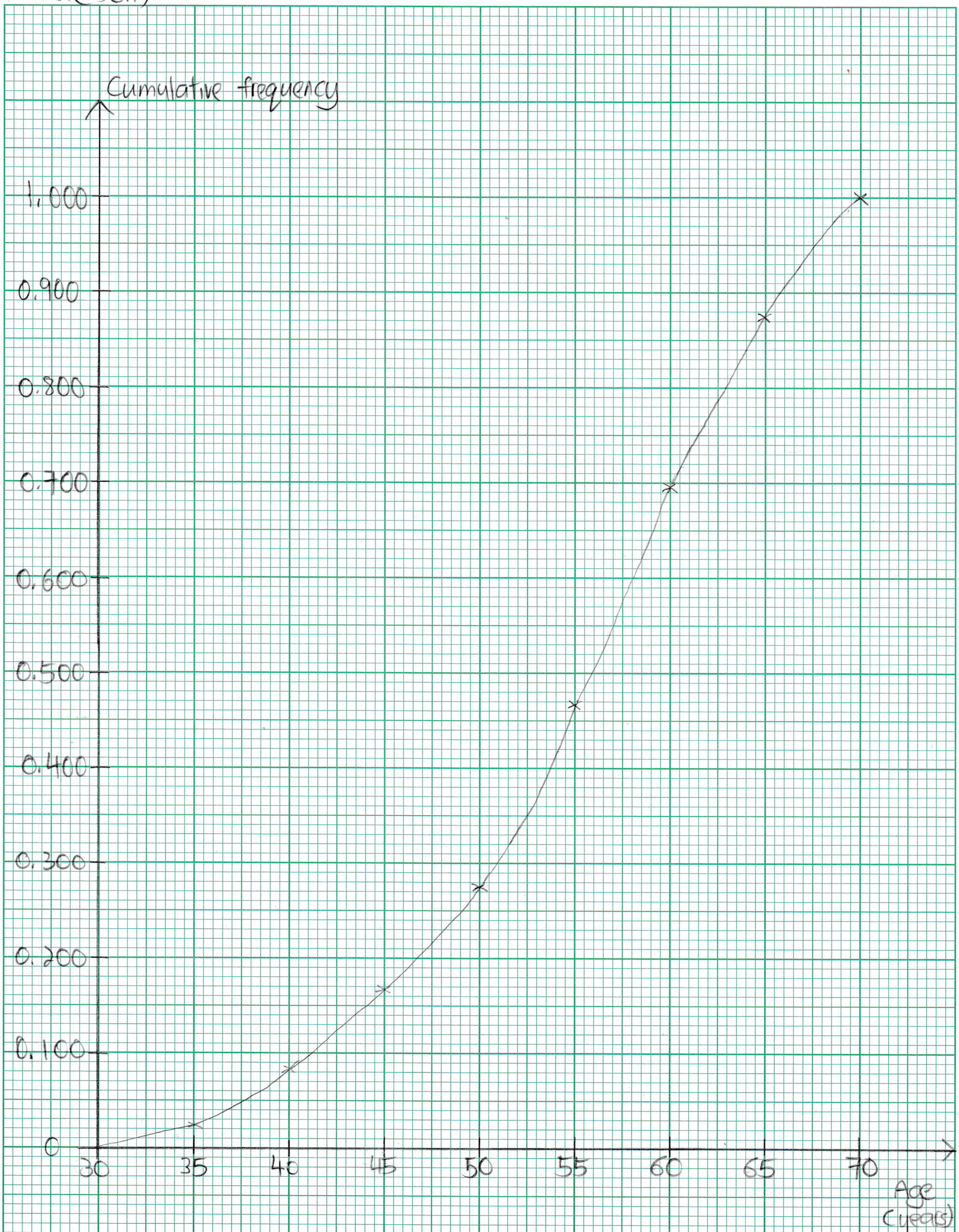
Date:

2(b)(i)

Age (years)	Frequency	Relative frequency	Cumulative relative frequency
30 - under 35	10	$\frac{10}{444} = 0.0225$	0.0225
35 - under 40	27	$\frac{27}{444} = 0.0608$	0.0833
40 - under 45	38	$\frac{38}{444} = 0.0856$	0.1689
45 - under 50	47	$\frac{47}{444} = 0.1059$	0.2748
50 - under 55	86	$\frac{86}{444} = 0.1937$	0.4685
55 - under 60	102	$\frac{102}{444} = 0.2297$	0.6982
60 - under 65	78	$\frac{78}{444} = 0.1757$	0.8739
65 - under 70	56	$\frac{56}{444} = 0.1261$	1.0000

2(b)(ii)

2(b)(ii)



Cumulative frequency
1.000
0.900
0.800
0.700
0.600
0.500
0.400
0.300
0.200
0.100
0
30
35
40
45
50
55
60
65
70
Age (years)

Subject : Date:

3(a)

Age Group	Class mark	No. of case	Cumulative frequency
1-5	3	27	27
6-10	8	23	50
11-15	13	43	93
16-20	18	78	171
21-26	23	123	294
27-30	28	180	474
31-35	33	144	618
36-40	38	136	754
41-45	43	118	872
46-50	48	133	1005
51-55	53	143	1148
56-60	58	182	1330
61-65	63	137	1467
66-70	68	90	1557
71-75	73	46	1603
76-80	78	20	1623
81-85	83	14	1637

3(a)(i) mode $= l + h \times \left(\frac{f_1 - f_0}{2f_1 - f_0 - f_2}\right)$

$$= 55.5 + \left(\frac{182 - 143}{2(182) - 143 - 137}\right)(5) = 57.8214$$

3(a)(ii) median $=$ lower boundary $+ \frac{\frac{N}{2} - cf_p}{f_{med}}(W)$

$$= 40.5 + \left(\frac{\frac{1637}{2} - 754}{118}\right)(5) = 43.23305$$

$$= 43.2331$$

Subject : Date:

3.(a)(iii) mean $= \frac{fx}{\Sigma x}$

$$= \frac{(3\times 27)+(8\times 23)+(13\times 43)+(18\times 78)+(23\times 123)+(28\times 180)+(33\times 144)+(38\times 136)+(43\times 118)+(48\times 133)+(53\times 143)+(58\times 182)+(63\times 137)+(68\times 90)+(73\times 46)+(78\times 20)+(83\times 14)}{1637}$$

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

$$= 43.0305$$

3(b) 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

3.(b)(i) mean temperature $= \frac{\text{total of temperature}}{\text{total number of temperature}}$

$$= \frac{(36.5\times 3)+(36.6\times 3)+(36.7\times 8)+(36.8\times 3)+(36.9\times 3)}{20}$$

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

$$= 36.7$$

3(b)(ii) mode temperature = 36.7

3(b)(iii) median temperature $= \frac{50}{100} \times$ total number of temperature

$$= \frac{1}{2}\times 20$$

$$= 10$$

$$= \frac{10th + 11th}{2}$$

$$= \frac{36.7 + 36.7}{2}$$

Subject : Date:

3.(b) (iv) 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

$\uparrow$ (16th)

$$P_{36.9} = \frac{\text{number of temperature less than } 36.9}{\text{total number of temperature}} \times 100$$

$$= \frac{16}{20} \times 100$$

$= 80\%$

$= 80$ th percentile

3(b) (v) $$\text{mean} = \frac{(36.5 \times 3) + (36.6 \times 3) + (36.7 \times 8) + (36.8 \times 2) + 36.9 + (37.0 \times 3)}{20}$$

$$= \frac{734.4}{20}$$

$= 36.72$

$$\text{standard deviation} = \sqrt{\frac{\Sigma (x - \bar{x})^2}{n-1}}$$

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

$$= \sqrt{\frac{0.472}{19}}$$

$$= \sqrt{\frac{59}{2375}}$$

$= 0.1576$

Subject : Date:

3(b) (vi) skewness $= \dfrac{\sum (x-\bar{x})^3}{(N-1)s^3}$

$$= \frac{[(36.5-36.72)^3\times3]+[(36.6-36.72)^3\times3]+[(36.7-36.72)^3\times8]+[(36.8-36.72)^3\times2]+(36.9-36.72)^3+[(37.0-36.72)^3\times3]}{(20-1)(0.1576)^3}$$

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

$$= 0.4776$$

The distribution is skewed to the right or skewed positively. This is because the value of skewness is greater than zero. The median is greater than mode and the mean is greater than median.

3(b)(vii) kurtosis $= \dfrac{\sum (x-\bar{x})^4}{(N-1)(s)^4}$

$$= \frac{[(36.5-36.72)^4\times3]+[(36.6-36.72)^4\times3]+[(36.7-36.72)^4\times8]+[(36.8-36.72)^4\times2]+(36.9-36.72)^4+[(37.0-36.72)^4\times3]}{(20-1)(0.1576)^4}$$

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

$$= 2.3225$$

The value of kurtosis is 2.3225. This is smaller or lower than the value 3. Hence this distribution is a platykurtic. The tails of this platykurtic distribution are shorter and thinner when compare its with the tails of normal distribution. The central peak is lower and also broader compared to mesokurtic that has the kurtosis value of 3. This indicate the data is lack of outliers. This is because the extreme value are less than that of normal distribution.

Subject : Date:

4. (a) Using Rstudio,

range = 91 - 2 = 89

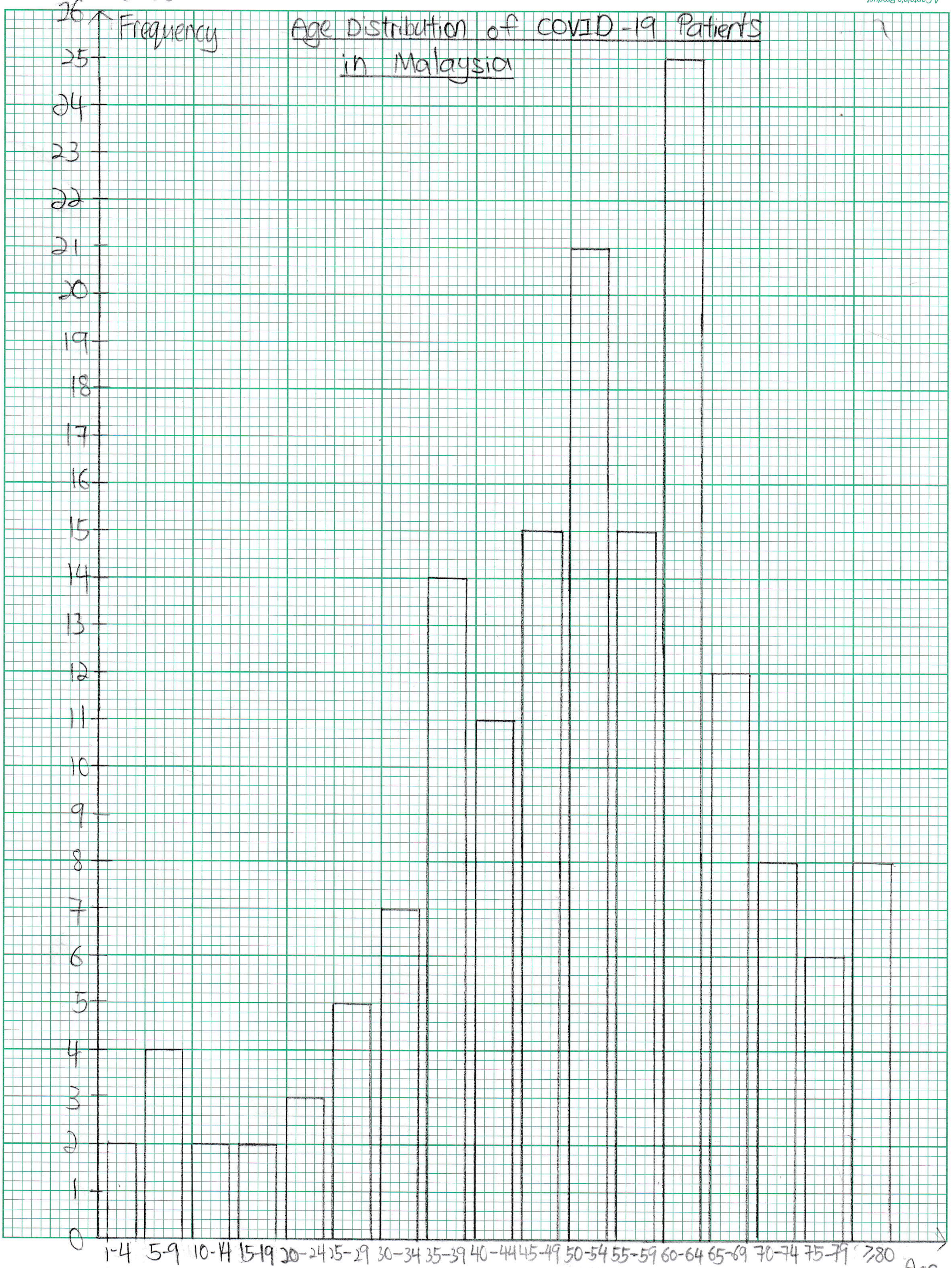
varience, $\sigma^2 = 333.021$

standard deviation = $\sigma = 18.22144$

4.b)(i) Using Rstudio,

Age group (year)	Frequency
1 - 4	2
5-9	4
10-14	2
15-19	2
20-24	3
25-29	5
30 - 34	7
35 - 39	14
40 - 44	11
45 - 49	15
50 - 54	21
55 - 59	15
60 - 64	25
65 - 69	12
70 - 74	8
75 - 79	6
≥ 80	8

4.(b)(i)



Age Distribution of COVID-19 Patients in Malaysia
Frequency
0 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26
1-4 5-9 10-14 15-19 20-24 25-29 30-34 35-39 40-44 45-49 50-54 55-59 60-64 65-69 70-74 75-79 ≥80

Subject :

Date:

4(b)(ii) Using Rstudio,

skewness = -0.4765937

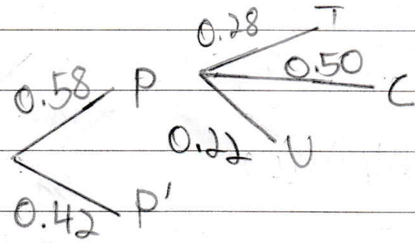
kurtosis = 3.067769

(iii) The distribution generated in (i) has a skewness that is lower than zero and has excess kurtosis of 0.067769. The distribution is skewed to the left or skewed negatively. Skewness is a measure of dataset symmetry. The skewness is between -0.5 and 0.5, here the data are fairly symmetrical. The distribution is a leptokurtic. The distribution has tails that are longer and fatter and the peak is higher and sharper if we compared to normal distribution.

The distribution of COVID-19 Patients in China is skewed to the left and skewed negatively and it skewed more than the distribution generated in (i). The distribution of COVID-19 Patients in China is leptokurtic. It has peak that is higher and sharper than that in distribution generated in (i). The tails of this

Subject : Date:

5. (a)



let

P' - screening test get positive

P' - screening test get negative

T - Has history travelling to country with COVID-19 outbreak.

C - Has contact with confirmed positive person.

U - Unknown cluster

5(b) Probality = 100 - 58

= 42

$= \frac{42}{100} = 0.42$

5.(c) Number of people = 72 × 0.42

= 30.24 = 31 person

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

$= (0.58 \times 0.28) + (0.58 \times 0.5)$

$= 0.4524$

5(e)(i) Probability = 1 - 0.016 - 0.34

= 0.644

5.(e)(ii) Number of people

= 0.644 × total number of people

= 0.644 × 2766

= 1781.304 = 1782

6.(a) let x = number of death in a day.

Subject :

Date:

6(a)(i) $P(x \leq 2) = P(x=0) + P(x=1) + P(x=2)$

$= 0.091 + 0.182 + 0.182$

$= 0.455$

6.(a)(ii) $P(x \geq 5) = P(x=5) + P(x=6)$

$+ P(x=7) + P(x=8)$

$= 0.091 + 0.045 + 0.$

$+ 0.045$

$= 0.181$

6.(a)(iii) $P(3 \leq x \leq 5) = P(x=3) + P(x=5) + P(x=4)$

$= 0.277 + 0.091 + 0.137$

$= 0.455$

6.(a)(iv) mean $\bar{x} = \sum x\, p(x)$

$= (0 \times 0.091) + (1 \times 0.182) + (2 \times 0.182)$

$+ (3 \times 0.227) + (4 \times 0.137) + (5 \times 0.091)$

$+ (6 \times 0.045) + (7 \times 0) + (8 \times 0.045)$

$= 2.86$

variance $= \sigma^2 = \sum (x - \mu_x)^2\, p(x)$

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

$+ (2-2.86)^2(0.182) + (3-2.86)^2(0.227) + (4-2.86)^2(0.137)$

$+ (5-2.86)^2(0.091) + (6-2.86)^2(0.045)$

$+ (7-2.86)^2(0) + (8-2.86)^2(0.045)$

$= 3.7404$

6(b) Subject: x - number of people die.
let pre-existing condition = PD

Date:

6(b)(i) $n=8$, $p=0.105$, PD = cardiovascular disease.

$P(x \leq 2) = P(x=0) + P(x=1) + P(x=2)$

$= ({}^8C_0)(0.105)^0(1-0.105)^8 + ({}^8C_1)(0.105)^1(1-0.105)^8 + ({}^8C_2)(0.105)^2(1-0.105)^6$

$= 0.95677 = 0.9568$

6(b)(ii) $n=4$, $p=0.073$, PD = Diabetes

$P(x \geq 2) = 1 - P(x \leq 1)$

$= 1 - (P(x=0) + P(x=1))$

$= 1 - ({}^4C_0(0.073)^0(1-0.073)^4 + {}^4C_1(0.073)^1(1-0.073)^4)$

$= 1 - 0.97105 = 0.02895$

6(b)(iii) let y - number of people that recover

$n = 10$

$q = 0.063$ PD: Chronic respiratory cancer.

$p = 1 - 0.063 = 0.937$

$P(y \leq 1) = P(y=0) + P(y=1)$

$= ({}^{10}C_0)(0.937)^0(0.063)^{10} + ({}^{10}C_1)(0.937)^1(0.063)^9$

$= 1.474738 \times 10^{-10}$

$= 1.4747 \times 10^{-10}$

6(b)(iv) let y - number of people that recover

$n = 5$

$q = 0.056$ PD: Cancer

$p = 1 - 0.056 = 0.944$

$P(2 \leq y \leq 4) = P(y=2) + P(y=3) + P(y=4)$

$= ({}^5C_2)(0.944)^2(0.056)^3 + ({}^5C_3)(0.944)^3(0.056)^2$

$+ {}^5C_4(0.944)^4(0.056)^1$

$= 0.2503$

Subject :

Date:

6(b)(v) let a - number of people that do not have pre-existing disease cure.

$n = 56 - (8 + 4 + 10 + 5)$
$= 21$

$q = 0.009$
$p = 1 - 0.009 = 0.991$

$P(a = 21) = ({}^{21}C_{21})(0.991)^{21}(0.009)^{0}$
$= 0.82708$
$= 0.8271$

6(c)(i) Variable x is the number of people that infected with COVID-19 die. It is a discrete variable. It is binomially distributed as there are fixed number of case for it.

6(a)(ii) Gender : Female

$p = 0.028$
$q = 1 - 0.028 = 0.972$

x is the random variable with probability of death equal to p for each trial.

$p(x) = (1-p)^{x-1}p$

$P(x=3) = (1 - 0.028)^{3-1}(0.028)$
$= 0.0265$

6.(c)(iii) Gender : Male

$q = 0.045$
$p = 1 - 0.045 = 0.953$

x is the random variable with probability of recovery equal to p for each trial.

$p(x) = (1-p)^{x-1}p$

$P(x=5) = (1 - 0.953)^{5-1}(0.953)$
$= 4.650 \times 10^{-6}$

Subject : Date:

6.(d) mean, $\mu = 4$ days

standard deviation, $\sigma = 5$ days

let x – the time duration between infector and infectee (day)

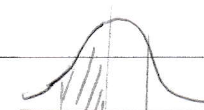
6.(d)(i) $P(x > 6) = P\left(Z > \frac{6-\mu}{\sigma}\right)$

$= P\left(Z > \frac{6-4}{5}\right)$

$= P(Z > 0.4)$

$= 0.3446$

standard normal distribution:



6.(d)(ii) $P(2 < x < 4) = P\left(\frac{2-4}{5} < Z < \frac{4-4}{5}\right)$

$= P(-0.4 < Z < 0)$

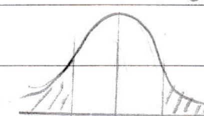
$= 0.5 - P(Z > -0.4)$

$= 0.5 - P(Z > 0.4)$

$= 0.5 - 0.3446$

$= 0.1554$

standar normal distribution:



$P(Z > 0.4) = P(Z < 0.4)$

as probability greater than zero and lower than zero for Z are the same which is 0.5.

6.(d)(iii) $P(x < 0) = P\left(Z < \frac{0-4}{5}\right)$

$= P(Z < -0.8)$

$= P(Z > 0.8)$

$= 0.2119$