Lee ~~Sze~~ Yuan Assignment 3

NO:

DATE: 10-4-2020

Question 1)

Variable	Data Type	Measuring Scale
Gender	discrete, qualitative	Nominal
Age	discrete, quantitative	Ratio
Nationality	discrete, qualitative	Nominal
Date confirmed	discrete, quantitative	Interval
Date discharged	discrete, quantitative	Interval
Hospital	discrete, qualitative	Nominal

b) Table 1 is secondary data because the data we have is not collected first-hand

NO:

DATE:

c) (i) bar chart

(ii) Histogram

(ii) Scatter plot

POP bazic™

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

$N = 30$

Key: $35|5 = 35.5\,^\circ C$

ii) 1st Quartile

$i = \frac{25}{100} N$

$i = \frac{1}{4}(30)$

$= 7.5$ round up

$= 8$

$Q_1 = x_8$

$= 36.4\,^\circ C$

2nd Quartile

$i = \frac{50}{100} N$

$= \frac{1}{2}(30)$

$= 15$ integer

$Q_2 = \frac{x_{15} + x_{15+1}}{2}$

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

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

$= 37.0\,^\circ C$

3rd quartile

$i = \frac{75}{100} N$

$= \frac{3}{4}(30)$

$= 22.5$ round up

$= 23$

$Q_3 = x_{23}$

$= 37.7\,^\circ C$

iii) Lower Limit $= Q1 - 1.5(\text{Interquartile range})$

$= Q1 - 1.5(Q3 - Q1)$

$= 36.4 - 1.5(37.7 - 36.4)$

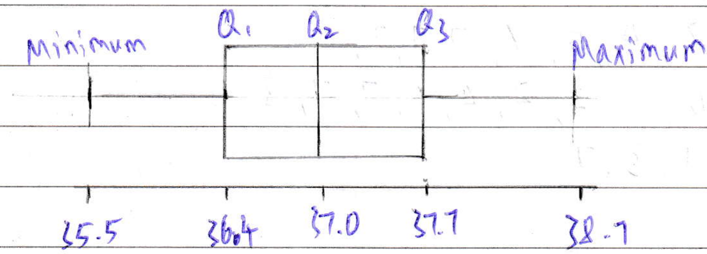
$= 34.45$

Upper Limit $= Q3 + 1.5(\text{Interquartile range})$

$= 37.7 + 1.5(37.7 - 36.4)$

$= 39.65$

$\therefore$ No outliers



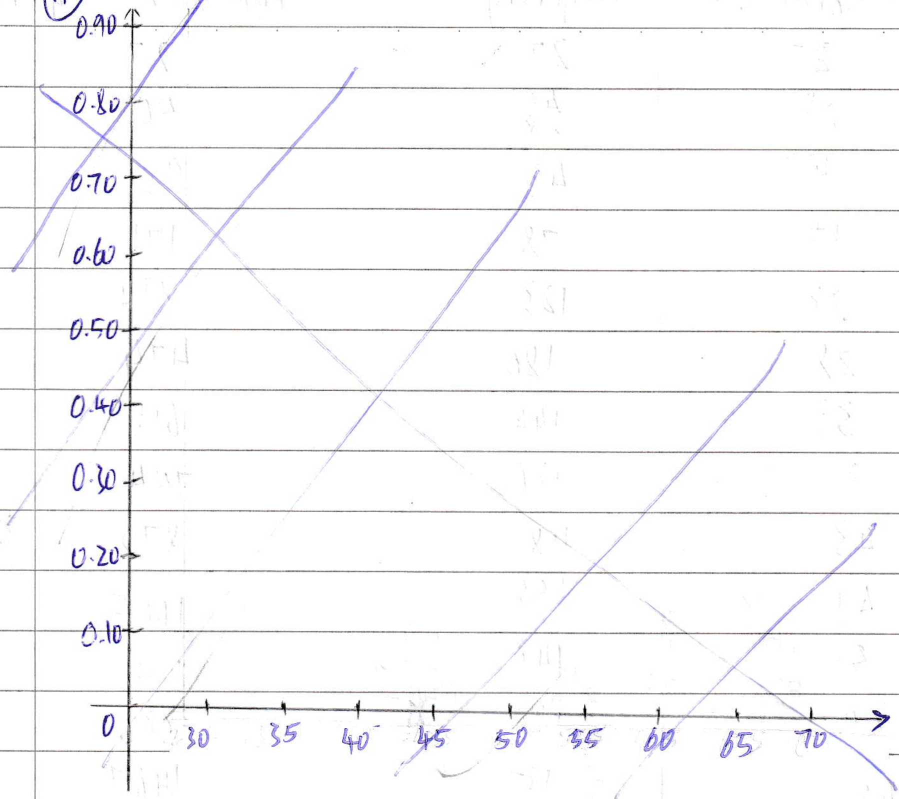
iv) Based on the boxplot, we can conclude that it skewed to the right because $(Q_3 - Q_2) > (Q_2 - Q_1)$

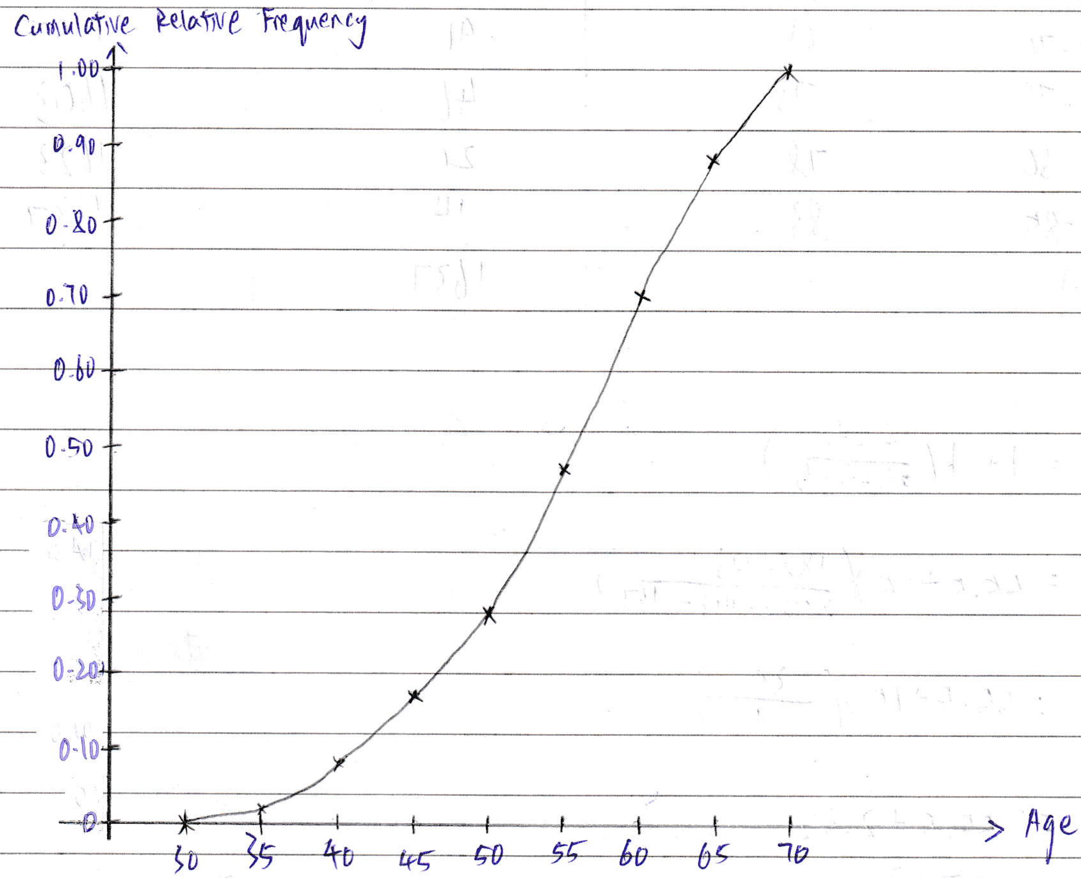
b) i)

Age	Frequency	Relative frequency	Cumulative relative frequency
30 - under 35	10	0.02	0.02
35 - under 40	27	0.06	0.08
40 - under 45	38	0.09	0.17
45 - under 50	47	0.11	0.28
50 - under 55	86	0.19	0.47
55 - under 60	102	0.23	0.70
60 - under 65	78	0.18	0.88
65 - under 70	56	0.12	1.00
Total	444		

NO:.............................. DATE:..............................

(ii)


0.90
0.80
0.70
0.60
0.50
0.40
0.30
0.20
0.10
0
30
35
40
45
50
55
60
65
70


Cumulative Relative Frequency
1.00
0.90
0.80
0.70
0.60
0.50
0.40
0.30
0.20
0.10
0
30
35
40
45
50
55
60
65
70
Age

NO: Question 3 DATE:

(a)

Age Group	Median	Frequency	Cumulative Frequency
1-5	3	27	27
6-10	8	23	50
11-15	13	43	93
16-20	18	78	171
21-25	23	123	294
26-30	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
Total		1637	

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

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

$= 55.5 + (5)\left(\frac{39}{84}\right)$

$= 55.5 + 2.32$

$= 57.82$

NO:.............................. DATE:..............................

ii) $\text{Median} = L + \left(\frac{\frac{N}{2} - cf_p}{f_{med}}\right)(w)$ $\frac{N}{2} = \frac{1637}{2} = 818.5$

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

$= 40.5 + \left(\frac{64.5}{118}\right)(5)$

$= 40.5 + 2.73$

$= 43.23$

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

$= \frac{70441}{1637}$

$= 43.03$

(b) i) $\text{mean temperature} = \frac{\Sigma x}{n} = \frac{36.5 + 36.5 + 36.5 + 36.6 + 36.6 + 36.6 + 36.7 + 36.7 + 36.7 + 36.7 + 36.7 + 36.7 + 36.7 + 36.7 + 36.8 + 36.8 + 36.8 + 36.9 + 36.9 + 36.9}{20}$

$= \frac{734}{20}$

$= 36.7$

ii) Mode temperature = 36.7 (It appears 8 times, it has the highest number of occurence)

iii)

$i = \frac{1}{2}(N)$

$= \frac{1}{2}(20)$

$= 10$ integer

$\text{Median temperature} = \frac{x_{10} + x_{11}}{2}$

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

$= 36.7$

NO:..............................   DATE:..............................

iv) 36.9 is the 17th value

$$\text{Percentile of } 36.9 = \frac{\text{number of values} < 36.9}{\text{Total number of values}} \times 100$$

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

$$= 80^{th}$$

36.9 is the 80th percentile

v) $\bar{x} = \frac{\Sigma x}{n} = \frac{36.5+36.5+36.5+36.6+36.6+36.6+36.7+36.7+36.7+36.7+36.7+36.7+36.7+36.7+36.8+36.8+36.9+37.0+37.0+37.0}{20}$

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

$$= 36.72$$

$$s^2 = \frac{\Sigma(x-\bar{x})^2}{n-1} = \frac{(36.5-36.72)^2(3) + (36.6-36.72)^2(3) + (36.7-36.72)^2(8) + (36.8-36.72)^2(2) + (36.9-36.72)^2 + (37.0-36.72)^2(3)}{20-1}$$

$$s^2 = \frac{0.472}{19}$$

$$s = 0.1576$$

NO:.............................. DATE:..............................

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

$$= \frac{(36.5-36.72)^3(3)+(36.6-36.72)^3(3)+(36.7-36.72)^3(8)+(36.8-36.72)^3(2) + (36.9-36.72)^3+(37.0-36.72)^3(3)}{(20-1)(0.1576)^3}$$

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

$$= 0.4775$$

But from previous calculation, we know that mean = median = mode
The distribution is considered symmetrical and did not skewed to the left or right

vii) Kurtosis $= \dfrac{\Sigma(x-\bar{x})^4}{(N-1)(s)^4}$

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

$$= 2.3225$$

its kurtosis is less than 3. It is platykurtic. Its tails are shorter and thinner, and often its central peak is lower and broader

NO:.................................. DATE:..................................

Question 4 I used R studio to calculate

a) range = maximum value - minimum value

$= 91 - 2 = 89$

b) variance $= \frac{\Sigma(x_i - \bar{x})^2}{N-1} = \frac{52791.339}{160-1} = 332.021$

standard deviation $= \sqrt{\text{variance}}$

$= 18.22144$

b) i) refer to R.script

I used library e1071

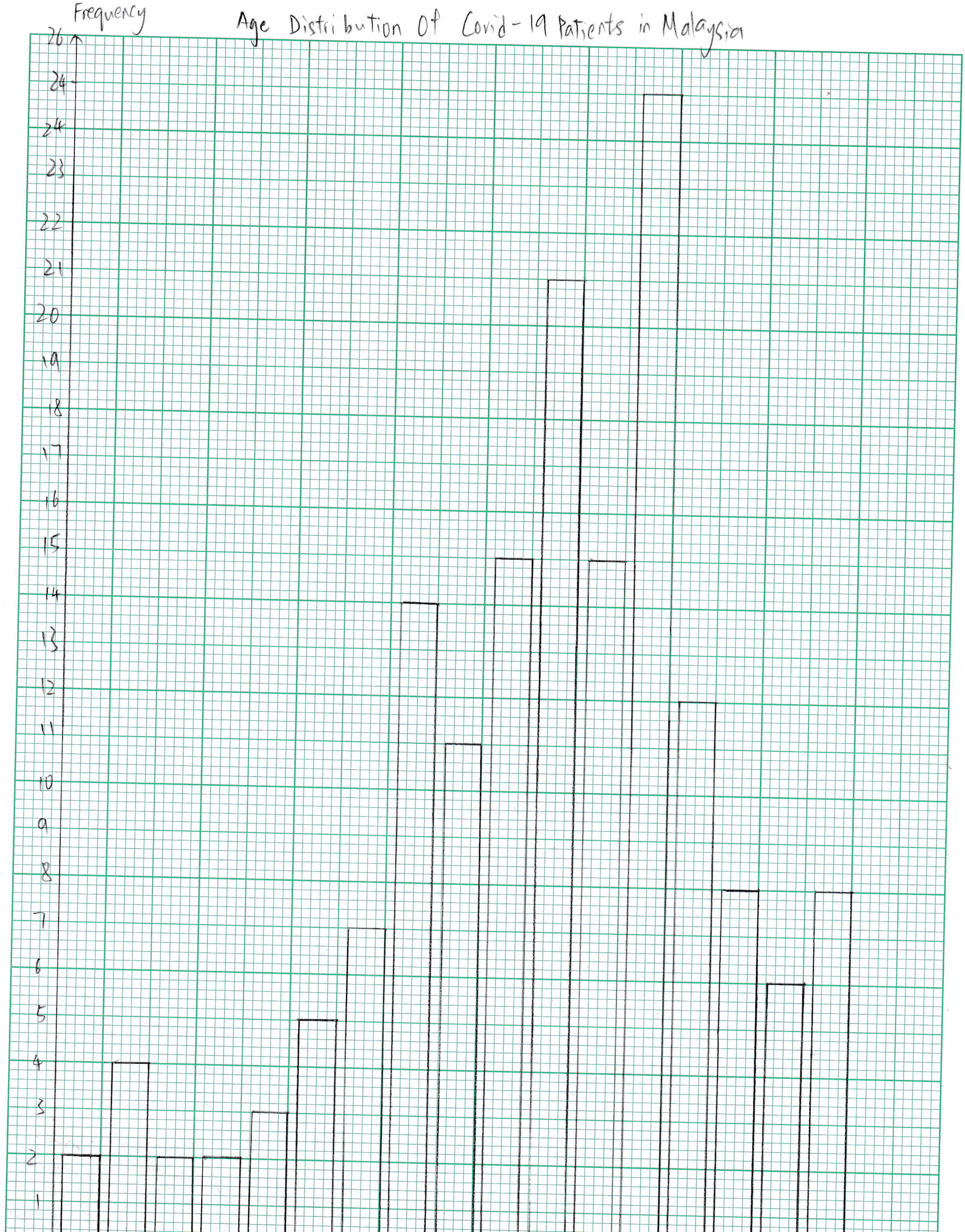
ii) Skewness $= \frac{\Sigma(x-\bar{x})^3}{(N-1)(S)^3} = \frac{-454128.9734}{(160-1)(18.22144)^3} = -0.4721$

Kurtosis $= \frac{\Sigma(x-\bar{x})^4}{(N-1)(S)^4} = \frac{53101266.98}{(160-1)(18.22144)^4} = 3.029542$

iii) For Malaysia, the distribution is skewed to the left and more negative compared to that of China. Because most of the patient in Malaysia is at age of around 50 to 70 years old, but most of the patient in China is at a age of around 30 to 50 years old

The age distribution in Malaysia is leptokurtic and have a bigger kurtosis than that of china. Because the number of patients in Malaysia who are at a age ranged from 35 to 70 is more than patients from other age group.

2(b)(i)



Frequency
Age Distribution Of Covid-19 Patients in Malaysia

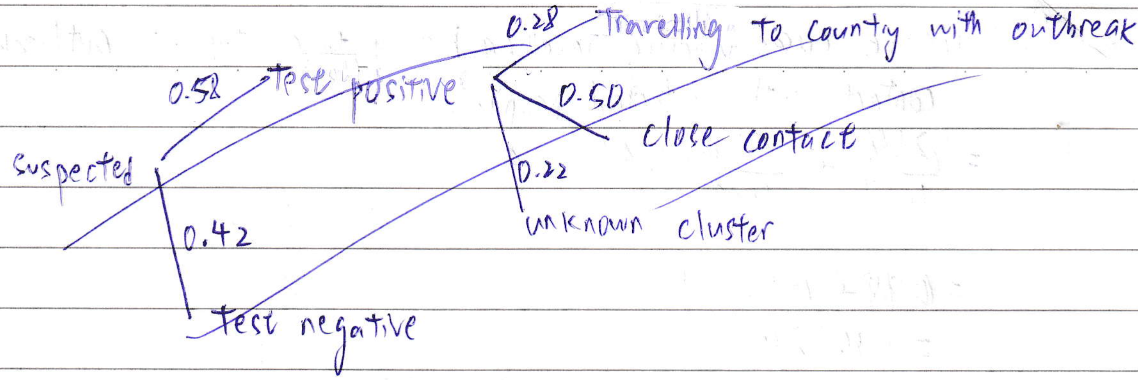
NO:................................ DATE:................................

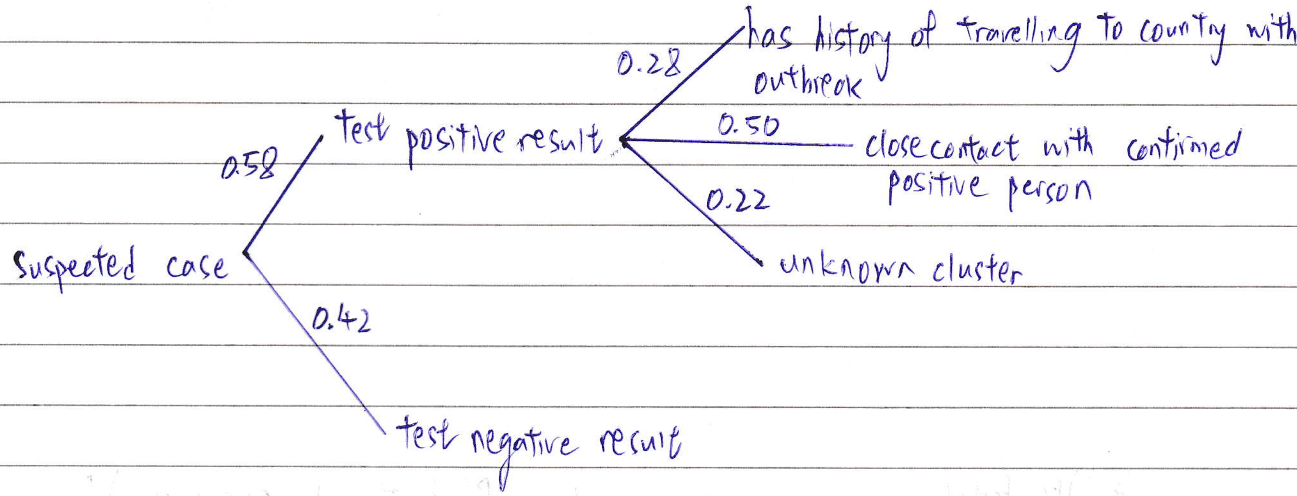
iii) For Malaysia, the distribution is skewed to the left but less negative than that of China which also skewed to the left. In my opinion, this because most of the patient in Malaysia is at age of around 50 to 70, but most of the patient in China is at a age of around 30 to 50 years old.

The age distibution in Malaysia is leptokurtic and have a bigger kurtosis than that of china. Because the number of patients in Maaycia is more centered at the range of age from 35 to 70.

NO:.................................. DATE:..................................

Question 5 (a)





b) P(test negative result) = P(test positive result)'

= 1 − P(test positive result)

= 1 − 0.58

= 0.42

c) $n = 72$ person-under-investigation

p − negative screening test result $p = 0.42$

Based on Binomial distribution

No. of people who get negative result $= np$

$= (72)(0.42)$

$= 30.24$

≈ 31

NO:..............................

DATE:..............................

d) P(positive cases having travelling history to country with outbreak or contact with infected people)

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

$= 0.1624 + 0.29$

$= 0.4524$

e) i) P(infected person under treatment) = P(death ∪ recovery)'

$= 1 - P(\text{death} \cup \text{recovery})$

$= 1 - \left(\frac{1.6\%}{100\%} + \frac{34\%}{100\%}\right)$

$= 1 - (0.016 + 0.34)$

$= 1 - 0.356$

$= 0.644$

ii) Based on binomial distribution

$n = 2766$ infected people

$p =$ probability of still under treatment $= 0.644$

no. of people under treatment $= np$

$= (2766)(0.644)$

$= 1781.304$

≈ 1782

NO:.................................   DATE:.................................

...stion 6 x - no. of deaths in a day

(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) ~~$P(x=0)$~~ $\mu_x = \sum x p(x)$

$= 0(0.091) + 1(0.182) + 2(0.182) + 4(0.137) + 5(0.091) + 6(0.045)$
$+ 7(0.000) + 8(0.045) + 3(0.227)$

$= 2.86$

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

$= (0-2.86)^2(0.091) + (1-2.86)^2(0.182) + (2-2.86)^2(0.182) + (4-2.86)^2(0.182)$
$+ (3-2.86)^2(0.227) + (5-2.86)^2(0.091) + (6-2.86)^2(0.045)$
$+ (7-2.86)^2(0.000) + (8-2.86)^2(0.045)$

$= 3.7404$

NO:................................ DATE:................................

(b) i) P – probability of dying among patient with cardiovascular disease $P = 0.105$

$n = 8$ $q = 1 - p = 0.895$

$$P(x \leq 2) = P(x=0) + P(x=1) + P(x=2)$$
$$= {}^8C_0(0.105)^0(0.895)^8 + {}^8C_1(0.105)^1(0.895)^7 + {}^8C_2(0.105)^2(0.895)^6$$
$$= 0.9568$$

ii) P – probability of dying among patient with diabetes

$P = 0.073$ $q = 1 - p = 1 - 0.073 = 0.927$

$n = 4$

$$P(x \geq 2) = 1 - P(x=0) - P(x=1)$$
$$= 1 - {}^4C_0(0.073)^0(0.927)^4 - {}^4C_1(0.073)^1(0.927)^3$$
$$= 0.0289$$

iii) P – probability of recovering among patient with chronic respiratory disease

$p = 1 - 0.063 = 0.937$ $q = 0.063$

$n = 10$

$$P(x \leq 1) = P(x=0) + P(x=1)$$
$$= {}^{10}C_0(0.937)^0(0.063)^{10} + {}^{10}C_1(0.937)^1(0.063)^9$$
$$= 1.474 \times 10^{-10}$$

iv) P – probability of recovering among patient with cancer

$p = 1 - 0.056 = 0.944$ $q = 0.056$

$n = 5$

$$P(2 \leq x \leq 4) = P(x=2) + P(x=3) + P(x=4)$$
$$= {}^5C_2(0.944)^2(0.056)^3 + {}^5C_3(0.944)^3(0.056)^2 + {}^5C_4(0.944)^4(0.056)^1$$
$$= 0.2503$$

NO:...............................   DATE:...............................

V) P - probability of recovering among patient with no pre-existings conditions

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

$n = 21$

$P(x=21) = {}^{21}C_{21} (0.991)^{21} (0.009)^{0}$

$= 0.8271$

NO:..............................

DATE:..............................

c) i) X is number of death among patient infected with COVID-19

X is binomially distributed

ii)

For female, $p = 0.028$

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

$= (0.972)^2(0.028)$

$= 0.0265$

iii) $p = 1-0.047 = 0.953$

$P(x=5) = (1-0.953)^{5-1}(0.953)$

$= 4.650 \times 10^{-6}$

NO:.................................. DATE:..................................

d) X - time duration between infector and infectee

$\mu = 4 \quad \sigma = 5$

i) $P(X > 6) = P(Z > \frac{6-\mu}{\sigma})$

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

$= P(Z > \frac{2}{5})$

$= P(Z > 0.4)$

$= 0.5 - 0.1554$

$= 0.3446$

ii) $P(2 < X < 4) = P(\frac{2-\mu}{\sigma} < Z < \frac{4-\mu}{\sigma})$

$= P(\frac{2-4}{5} < Z < \frac{4-4}{5})$

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

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

$= 0.5000 - [1 - P(Z < 0.4)]$

$= 0.5000 - [1 - 0.6554)$

$= 0.5000 - 0.3446$

$= 0.1554$

iii) $P(X < 0) = P(Z < \frac{0-\mu}{\sigma})$

$= P(Z < \frac{0-4}{5})$

$= P(Z < -0.8)$

$= P(Z > 0.8)$

$= 1 - P(Z < 0.8)$

$= 1 - 0.7881$

$= 0.2119$