MOHAMAD AMIN HAZEEQ BIN HISHAM A19EC0087

ASSIGNMENT 3

Question 1

Variable	Data Type	Measuring scale

a) Case No - discrete, quantitative, Ratio scale

Gender - discrete, qualitative, Nominal scale

Age - discrete, quantitative, Ratio scale

Nationality - discrete, qualitative Nominal scale

Date confirmed - discrete, quantitative, Interval Scale

Date discharged - discrete, quantitative, Interval scale

Hospital - discrete, qualitative, Nominal scale

b) Data in table 1 is secondary storage source because the data has been collected by individuals or agencies for purpose in research study. Data taken is already recorded and presented in table 1 and not collected by us.

c) i) Gender - Barplot / Pie Chart

ii) Age - Histogram

iii) Data discharged - Barplot / Pie Chart

Question 2

a) i) The temperature (in °C) measured

key: 35/5 means ↓ 35.5°C

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

a) ii) $i = \frac{P}{100}(N)$ $N = 30$

1st quartile => $Q_1 = \frac{25}{100}(30)$

$= 7.5$ (not integer)

≈ 8

$Y[8] = 36.4$

2nd quartile => $Q_2 = \frac{50}{100}(30)$

$= 15$ (integer)

$= (Y[15] + Y[16]) \div 2$

$= (37.0 + 37.0) \div 2$

$= 37.0$

3rd quartile => $Q_3 = \frac{75}{100}(30)$

$= 22.5$ (not integer)

≈ 23

$Y[23] = 37.7$

Question 2

a) iii) minimum = 35.5 $Q_1 = 36.4$ $IQR = Q_3 - Q_1$

maximum = 38.7 $Q_2 = 37.0$ $= 1.3$

$Q_3 = 37.7$

lower limit $= Q_1 - 1.5 \times IQR$

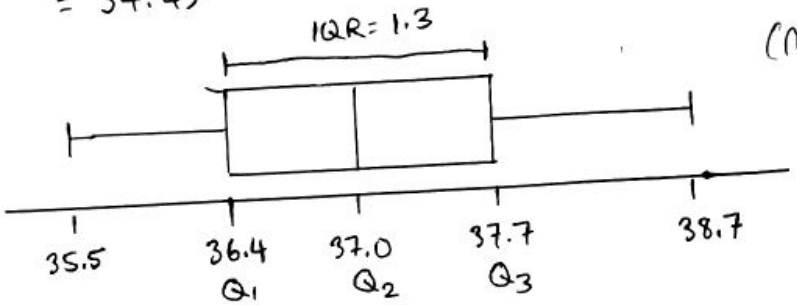
$= 36.4 - 1.5(1.3)$

$= 34.45$

Upper limit $= Q_3 + 1.5 \times IQR$

$= 37.7 + 1.5(1.3)$

$= 39.65$



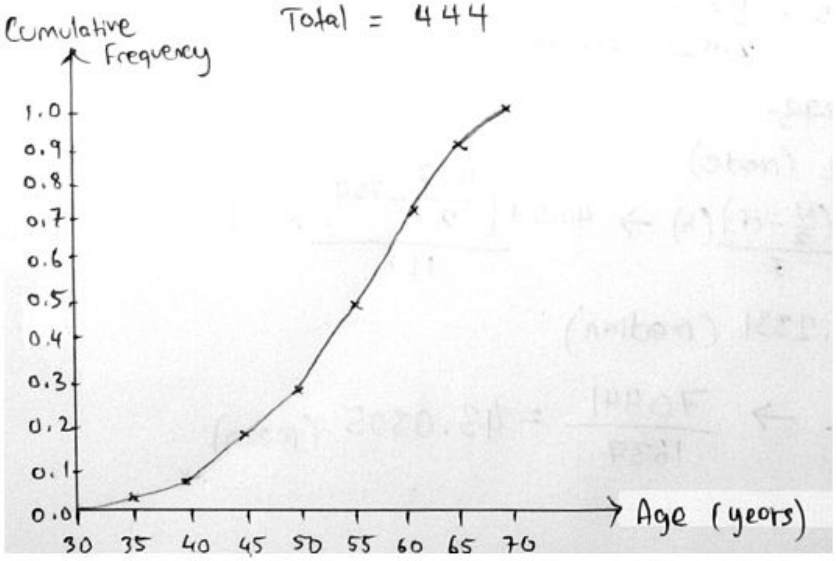
(Modified Box Plot)

a) iv) It is skewed to the right because 3rd quartile − 2nd quartile is = 0.7 greater than 2nd quartile − 1st quartile is = 0.6. This boxplot does not has outlier, so not affect the mean of dataset.

b) i)

×	Frequency	Relative-Frequency	Cumulative Relative Frequency
30 – 34	10	10 ÷ 444 = 0.0225	0.0225
35 – 39	27	27 ÷ 444 = 0.0608	0.0833
40 – 44	38	38 ÷ 444 = 0.0856	0.1689
45 – 49	47	47 ÷ 444 = 0.1059	0.2748
50 – 54	86	86 ÷ 444 = 0.1937	0.4685
55 – 59	102	102 ÷ 444 = 0.2297	0.6982
60 – 64	78	78 ÷ 444 = 0.1757	0.8739
65 – 69	56	56 ÷ 444 = 0.1261	1.0000
Total =	444		

b) ii)



Question 3

Boundaries	midpoint, x	frequency, f	Cumulative frequency	fx
0.5 - 5.5	3	27	27	81
5.5 - 10.5	8	23	50	184
10.5 - 15.5	13	43	93	559
15.5 - 20.5	18	78	171	1404
20.5 - 25.5	23	123	294	2829
25.5 - 30.5	28	180	474	5040
30.5 - 35.5	33	144	618	4752
35.5 - 40.5	38	136	754	5168
40.5 - 45.5	43	118	872	5074
45.5 - 50.5	48	133	1005	6384
50.5 - 55.5	53	143	1148	7579
55.5 - 60.5	58	182	1330	10 556
60.5 - 65.5	63	137	1467	8631
65.5 - 70.5	68	90	1557	6120
70.5 - 75.5	73	46	1603	3358
75.5 - 80.5	78	20	1623	1560
80.5 - 85.5	83	14	1637	1162
				Total = 70 441

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

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

$= 28.0893$

$= 57.82$ (mode)

a) ii) $\text{median} = L + \frac{\left(\frac{N}{2} - cf\right)}{f}(h) \rightarrow 40.5 + \frac{\left(\frac{1637}{2} - 754\right)}{118} \times (5)$

$= 43.2331$ (median)

a) iii) $\text{mean} = \frac{\Sigma fx}{n} \rightarrow \frac{70441}{1637} = 43.0305$ (mean)

Question 3

(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 = 734 (Total)

i) $\frac{734}{20} = 36.7$ (mean)

ii) mode = 36.7

iii) median = $\frac{36.7 + 36.7}{2}$

= 36.7 (median)

(3) b) iv) $i = \frac{\text{n less than } 36.9}{\text{total number}} \times 100$ $n = 20$

∴ there are 16 values before 36.9

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

$= 80$ ∴ 36.9 is in the 80^{th} percentile. ※

(3) b) v) 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 = 734.4

$n = 20$ $\bar{x} = \frac{734.4}{20} \Rightarrow 36.72$ (mean) $s = \sqrt{\frac{(\Sigma(x-\bar{x})^2}{n-1}}$ → $(x-\bar{x})^2$

$s = \sqrt{\frac{0.472}{20-1}}$

$= \sqrt{0.0248}$

$= 0.1576$ (standard deviation)

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

$= 36.70$ (median)

mode = 36.70

0.0484
0.0484
0.0484
0.0144
0.0144
0.0144
4×10^{-4}
4×10^{-4}
4×10^{-4}
4×10^{-4}
4×10^{-4}
4×10^{-4}
4×10^{-4}
4×10^{-4}
6.4×10^{-3}
6.4×10^{-3}
0.0324
0.0784
0.0784
0.0784

$\Sigma(x-\bar{x})^2 = 0.472$

Question 3

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

$= \dfrac{0.035664}{(20-1)(0.1576)^3}$

$= 0.47952$

$= 0.4795$

$(x-\bar{x})^3$

-0.0106
-0.0106
-0.0106
-1.728×10^{-3}
-1.728×10^{-3}
-1.728×10^{-3}
-8×10^{-6}
-8×10^{-6}
-8×10^{-6}
-8×10^{-6}

→

-8×10^{-6}
-8×10^{-6}
-8×10^{-6}
-8×10^{-6}
5.12×10^{-4}
5.12×10^{-4}
5.832×10^{-3}
0.021952
0.021952
0.021952

↓ Total sum:

$\sum (x-\bar{x})^3 = 0.035664$

mean = 36.72

median = 36.70

mode = 36.70

∴ The distribution is skewed to the right or skewed positively. because the skewness value is greater than zero ($0.4795 > 0$) The mean is greater than mode and also median

Question 3

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

$= \dfrac{0.027224}{(20-1)(0.1576)^4}$

$= 2.32259$

$= 2.3226$

$(x-\bar{x})^4$

2.34256×10^{-3}
2.34256×10^{-3}
2.34256×10^{-3}
2.0736×10^{-4}
2.0736×10^{-4}
2.0736×10^{-4}
1.6×10^{-7}
1.6×10^{-7}
1.6×10^{-7}
1.6×10^{-7}

→

1.6×10^{-7}
1.6×10^{-7}
1.6×10^{-7}
1.6×10^{-7}
4.096×10^{-5}
4.096×10^{-5}
1.04976×10^{-3}
6.14656×10^{-3}
6.14656×10^{-3}
6.14656×10^{-3}

↓ Total sum:

$\sum (x-\bar{x})^4 = 0.027224$

∴ This distribution is platykurtic because kurtosis is less than 3 ($2.3226 < 3$)
so the tail of platykurtic distribution is shorter, central peak is lowered compare to other distribution.

Question 4

a)

$\sum (x-\bar{x})^2$

1617.5478
2422.4853
189.9228
125.8603
231.6103
2229.6103
0.6103
4.9228
104.4228
138.7978
0.0478
3.1728
3.1728
1538.1103
77.1103
281.6103
635.9853
218.4853
45.9853
296.4853
38.6728
1782.4228
125.8603
202.1728
60.5478
249.0478
163.3603
369.3603
149.2978
1955.2978
828.3603
27.2353
474.4228
77.1103
77.1103
281.6103
1240.3603
45.9853
3.1728
116.2353

369.3603
1010.0478
3.1728
104.4228
7.7353
0.61030
974.61030
38.6728
263.0475
174.7353
138.7978
174.7353
77.1103
95.6728
431.8603
518.9853
33.4228
77.1103
0.0478
614.1103
1.4853
67.5478
45.9853
125.8603
0.0478
4.9228
138.7978
77.1103
38.6728
4.9228
316.1728
163.3603
664.6728
218.4853
10.3603
4.9228
771.7978
125.8603
116.2353
263.0478

586.5478
174.7353
60.5478
116.2353
450.2353
104.4228
116.2353
14.2978
0.0478
22.8603
352.7353
352.7353
1.4853
281.6103
202.1728
218.4853
95.6728
125.8603
3.1728
565.5478
1782.4228
84.9853
1.4853
27.2353
231.6103
1.4853
1240.3603
493.6728
95.6728
138.7978
95.6728
138.7978
22.8603
316.1728
10.3603
116.2353
17.7978
33.4228
1010.0478
474.4228

45.9853
1.4853
1867.8603
539.1103
0.6103
95.6728
331.9728
231.6103
369.3603
0.0478
0.6103
163.3603
1141.1728
95.6728
10.3603
1582.5478
614.1103
586.5478
27.2353
1074.6103
22.8603
52.1103
3.1728
263.0478
913.1728
740.8603
33.4228
17.7978
369.3603
231.6103
1141.1728
316.1728
316.1728
886.9228
474.4228
22.8603
202.1728
717.2353
391.2978
14.2978

$\bar{X} = \frac{8195}{160}$

$= 51.21875$ (mean)

Range $= 91 - 2$

$= 89$ ✱

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

$= \frac{52791.3355}{160-1}$

$\sigma^2 = 332.0209$ ✱✱

Standard deviation $= \sqrt{332.0209}$

$\sigma = 18.2214$ ✱

$\sum (x-\bar{x})^2 = 52791.3355$

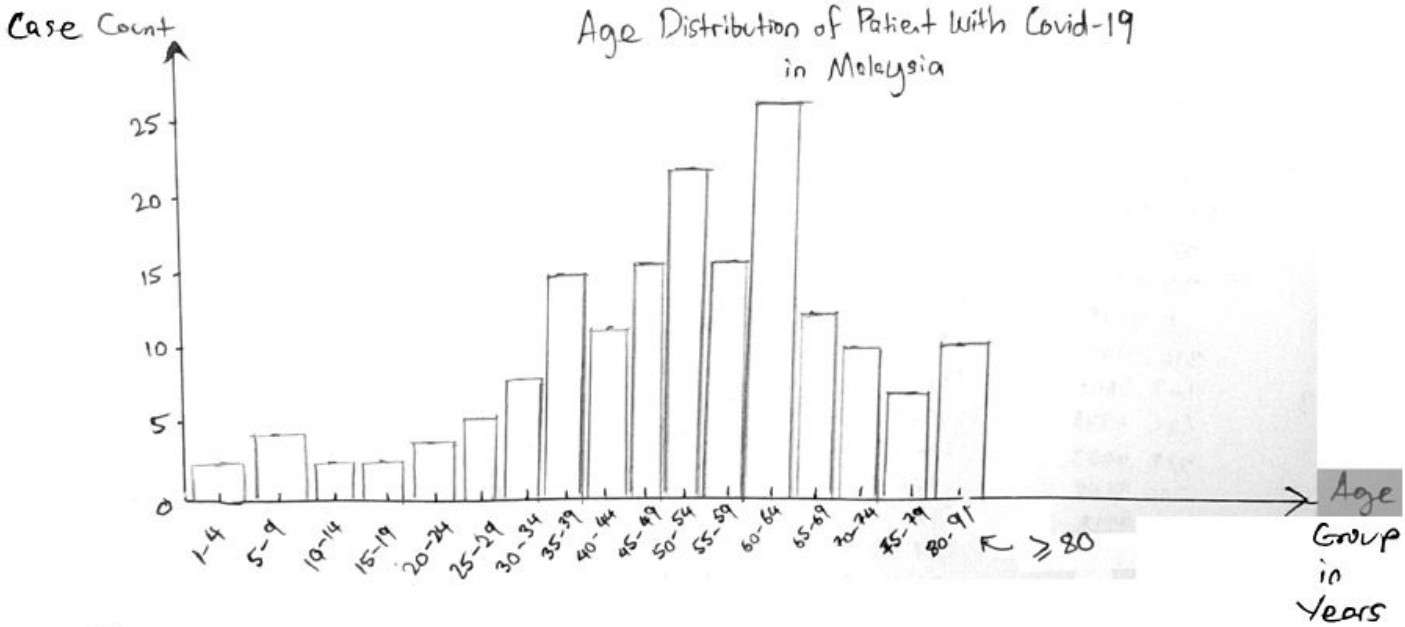
Question 4

b)

i)

Age Group	Case Count	Midpoint
1-4	2	2.5
5-9	4	7
10-14	2	12
15-19	2	17
20-24	3	22
25-29	5	27
30-34	7	32
35-39	14	37
40-44	11	42
45-49	15	47
50-54	21	52
55-59	15	57
60-64	25	62
65-69	12	67
70-74	8	72
75-79	6	77
(≥80) 80-91	8	85.5

Total = 160



ii) skewness = -0.4765937 (refer to my Rscript)

kurtosis = 3.067769 (refer to my Rscript)

Question 4

b) iii)

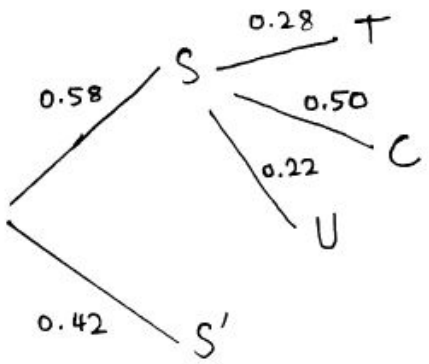
In Malaysia, the distribution is more negatively skewed compare to negatively skewed distribution in China. It is because patient in Malaysia have many number of patients that age are around 50 to 60 but in China, many case at age around 35 to 45.

In Malaysia, the distribution is leptokurtic and bigger kurtosis than in China. It is leptokurtic because ~~ther~~ the kurtosis is greater than 3 ($3.067769 > 3$). In Malaysia, central peak is higher compare to China because in Malaysia age in range 50 - 60 is the only has higher case count and the remaining grouped of age only has less case count which is ~~less than 15 cases~~. not more than 15 cases. But in China, the distribution shows that the number of case count has not much different in range of 30 - 55 ages thus will make the ~~dis~~ central peak broader.

Question 5

a) S - screen test positive T - travel to other country with covid-19 outbreak ~~C - close contact~~ U - unknown cluster

C - contact with confirmed positive person



b) $= 100 - 58$

$= 42$

$= \frac{42}{100}$

$= 0.42$

c) Number of people $\Rightarrow \frac{42}{100} \times 72 = 30.24$

$= 30.24$

$= 30$ people

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

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

$= 0.4524$

e) i) recovery = 34% (0.34)
death = 1.6% (0.016)

undertreatment = $1 - 0.34 - 0.016$
$= 0.644$

ii) Number of people

$\Rightarrow = 0.644 \times$ total number of people

$= 0.644 \times 2766$

$= 1781.304$

$= 1781$ people under treatment

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

$= 0.091 + 0.045 + 0.00 + 0.045$

$= 0.181$

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, $\bar{X} = \Sigma x P(x)$

$= (0 \times 0.091) + (1 \times 0.182) + (2 \times 0.182) + (3 \times 0.227)$
$+ (4 \times 0.137) + (5 \times 0.091) + (6 \times 0.045) + (7 \times 0.00) + (8 \times 0.045)$

$= 2.86$

Variance, $\Sigma (x - \bar{X})^2 \times P(x)$

$= \left((0-2.86)^2(0.091)\right) + \left((1-2.86)^2(0.182)\right) + \left((2-2.86)^2(0.182)\right) + \left((3-2.86)^2(0.227)\right)$
$+ \left((4-2.86)^2(0.137)\right) + \left((5-2.86)^2(0.091)\right) + \left((6-2.86)^2(0.045)\right) +$
$\left((7-2.86)^2(0.0)\right) + \left((8-2.86)^2(0.045)\right)$

$= 0.74434 + 0.62964 + 0.13460 + 4.4492 \times 10^{-3}$
$+ 0.17804 + 0.41674 + 0.44368 + 0.0 + 1.18882$

$= 3.7403$

Question 6

b) i) X – number of people die Probability death = 0.105

condition – cardiovascular disease

$n = 8$

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

$= \left({}^{8}C_0 (0.105)^0 (1-0.105)^8\right) + \left({}^{8}C_1 (0.105)^1 (1-0.105)^7\right) + \left({}^{8}C_2 (0.105)^2 (1-0.105)^6\right)$

$= 0.956768977$

$= 0.9568$

b) ii) condition = Diabetes Probability death = 0.073 X – number people die

$n = 4$

$P(x \geq 2) = P(x=2) + P(x=3) + P(x=4)$

$= \left({}^{4}C_2 (0.073)^2 (1-0.073)^2\right) + \left({}^{4}C_3 (0.073)^3 (1-0.073)^1\right) + \left({}^{4}C_4 (0.073)^4 (1-0.073)^0\right)$

$= 0.02894705$

$= 0.0289$

b) iii) Chronic respiratory Probability death = 0.063

recover = 1 – 0.063 = 0.937 X – number recovery

$n = 10$

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

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

$= 1.474737689 \times 10^{-10}$

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

b) iv) Cancer condition. Probability death = 0.056

recover = 1 – 0.056 = 0.944 X – number recovery

$n = 5$

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

$= \left({}^{5}C_2 (0.944)^2 (0.056)^3\right) + \left({}^{5}C_3 (0.944)^3 (0.056)^2\right) + \left({}^{5}C_4 (0.944)^4 (0.056)^1\right)$

$= 0.2503005$

$= 0.2503$

b) v) $56 - (8+4+10+8+5) = 21$ Probability death = 0.009

recover = 1 – 0.009 = 0.991 X – number recover

$n = 21$

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

$= 0.8271$

Question 6

c) male = 0.047 (death rate)
female = 0.028 (death rate)

i) Variable X is number of male patient that are infected with covid-19. X is binomial distribution.

ii) $P(Y=3) = (1-0.028)^{3-1}(0.028)$
$= 0.02645$

Y - (female) number of patient infected.

iii) $P(X=5) \Rightarrow$ recovery $= 1-0.047$
$= 0.953$

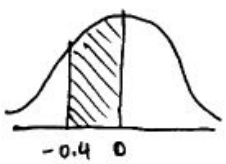
$= (1-0.953)^{5-1}(0.953)$
$= 4.6503 \times 10^{-6}$

d)

i) $n = 468 \quad X \sim N(4,5) \quad z = \frac{x-\mu}{\sigma}$

$P(x>6) = P\left(z > \left(\frac{6-4}{5}\right)\right)$
$= P(z > 0.4) \Rightarrow 1 - P(z<0.4)$
$= 1 - 0.65542$
$= 0.34458$

ii) $P(2<x<4) = P\left(\frac{2-4}{5} < z < \frac{4-4}{5}\right)$



$= P(-0.4<z) \quad P(z<0)$
$= P(z>-0.4) \quad P(z<0) = 0.5$
$= P(z<0) - P(z<-0.4)$
$= 0.5 - 0.34458$
$= 0.15542$

Question 6

d) iii) $P(0 > x) = P\left(\frac{0-4}{5} > z\right)$

$= P(-0.8 > z)$

$= P(z < -0.8)$

$= 0.21186$