Assigment 3 (PSDA)

Ref: Date:

Question 1

(a) case no => discrete, quantitive, ratio

gender => discrete, qualitative, nominal

age => continuous, quantitative, ordinal

nationality => discrete, qualitative, nominal

date confirmed => discrete, quantitative, ordinal

date discharged => discrete, quantitative, ordinal

hospital => discrete, qualitative, nominal.

(b) Primary data. Because the data is collected directly from the primary sources.

(c) Gender => Bar graph

Age => Pie chart

Date discharged => Stem and leaf

Ref: Date:

Question 2

(a) i) Temperature of COVID-19 Patients

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 => 36|1 = 36.1

ii) First quartile, Q_1 => 1/4 (30)
= 7.5
= 8th
= 36.4

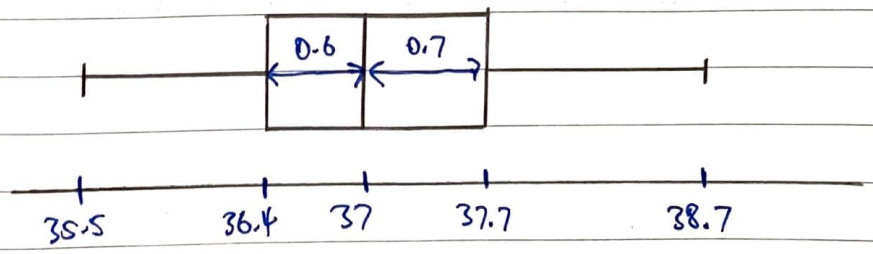
Second quartile, Q_2 => 1/2 (30)
= 15
= (37.0 + 37.0)/2
= 37.0

Third quartile, Q_3 => 3/4 (30)
= 22.5
= 23rd
= 37.7

iii) min value = 35.5
max value = 38.7
IQR = 37.7 - 36.4
= 1.3

Lower limit = $Q_1 - 1.5(IQR)$
= 36.4 - 1.5(1.3)
= 34.45

Upper limit = $Q_3 + 1.5(IQR)$
= 37.7 + 1.5(1.3)
= 39.65



iv) Data skewed to the left, negatively skewed.

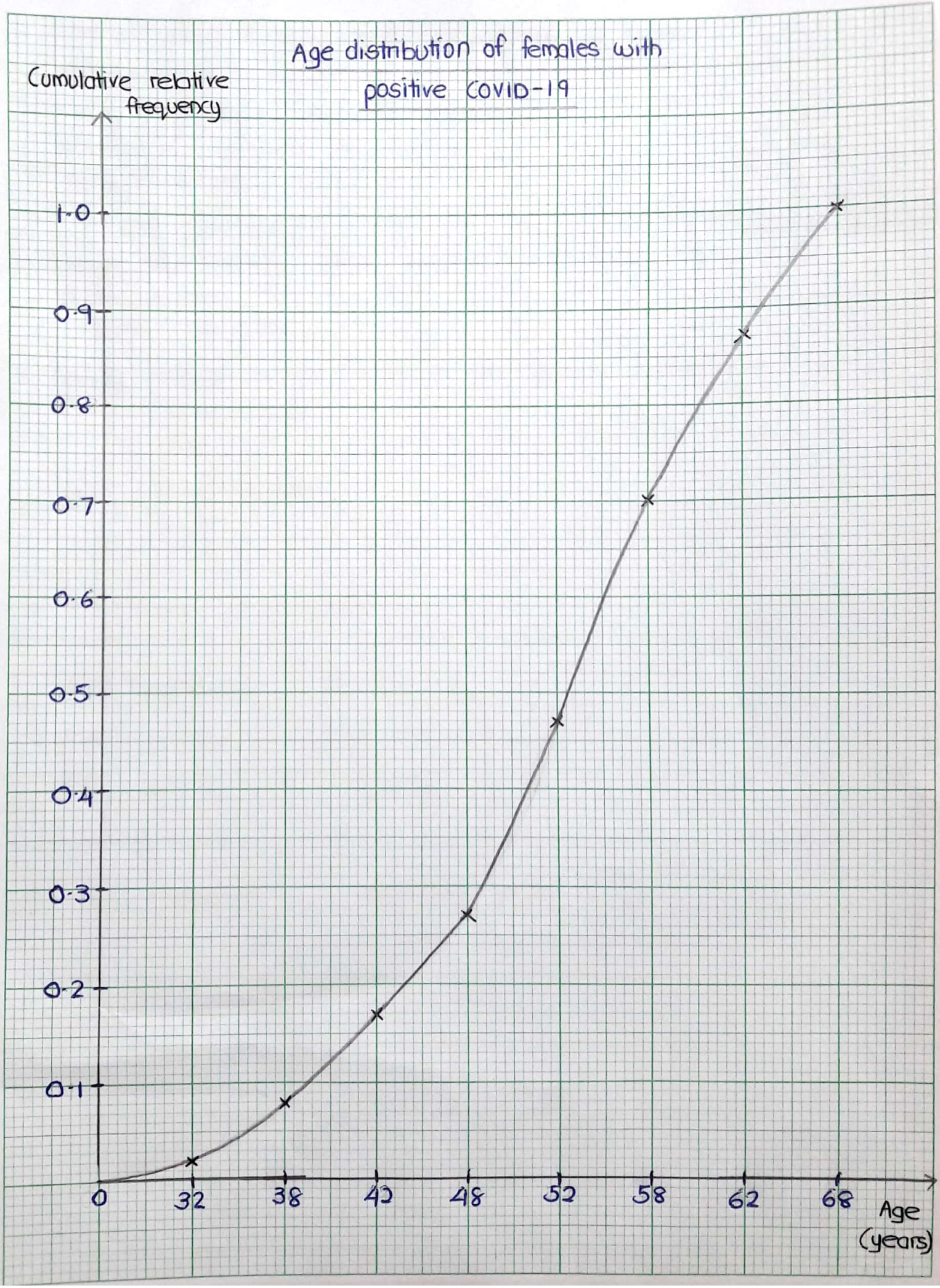
Ref: Date:

(b) i)

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

ii)

* Ogive on graph


Age distribution of females with positive COVID-19
Cumulative relative frequency
1·0
0·9
0·8
0·7
0·6
0·5
0·4
0·3
0·2
0·1
0
32
38
43
48
52
58
62
68
Age (years)

Ref: Date:

Question 3

(a)

Age group	Midpoint, x	Frequency, f	Cumulative frequency, cf
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

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

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

$= 57.82$

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

$= 40.5 + \left(\frac{1637/2 - 154}{118}\right)(5)$

$= 43.23$

iii) $\text{Mean} = \frac{[27(3) + 23(8) + 13(43) + 78(18) + 123(23) + 180(28) + 144(33) + 136(38) + 118(43) + 133(48) + 143(53) + 182(58) + 137(63) + 90(68) + 46(73) + 20(78) + 14(83)]}{1687}$

$= 43.03$

Ref: Date:

(b) i) Mean temp = $\frac{36.5(3) + 36.6(3) + 36.7(8) + 36.8(2) + 36.9 + 37.0(3)}{20}$

$= 36.72$

Temperature	$(x-\bar{x})^2$	$(x-\bar{x})^4$
36.5	0.0484	0.0023
36.5	0.0484	0.0023
36.5	0.0484	0.0023
36.6	0.0144	0.0002
36.6	0.0144	0.0002
36.6	0.0144	0.0002
36.7	0.0004	0.0000
36.7	0.0004	0.0000
36.7	0.0004	0.0000
36.7	0.0004	0.0000
36.7	0.0004	0.0000
36.7	0.0004	0.0000
36.7	0.0004	0.0000
36.7	0.0004	0.0000
36.8	0.0064	0.0000
36.8	0.0064	0.0000
36.9	0.0324	0.0010
37.0	0.0784	0.0061
37.0	0.0784	0.0061
37.0	0.0784	0.0061
Total =>	0.472	0.0268

ii) Mode = 36.7

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

$= 10^{th}$

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

$= 36.7$

iv) $x = 36.9$

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

$= 80^{th}$ percentile

v) $\sigma = \sqrt{\frac{\Sigma(x-\bar{x})^2}{N-1}}$

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

$= 0.1576$

vi) Mean = 36.7, Median = 36.7, Mode = 36.7

∴ Mean = median = mode, all distribution fall at the same point, known as symmetrical distribution.

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

$= \frac{0.0268}{(19)(0.16)^4} = 2.20$

∴ Kurtosis < 3, platykurtic.

Ref: Date:

Question 4

(a) Variance, $\sigma^2 = 328.287$

Standard deviation, $\sigma = \sqrt{328.287}$

$= 18.12$

Range $= 91 - 2$

$= 89$

(b)(i)

Age group	f	x	f(x)	cf	$(x-\bar{x})^3$	$(x-\bar{x})^4$
1-4	2	2.5	5	2	-116142.84	5666609.33
5-9	4	7	28	6	-86879.45	3847890.65
10-14	2	12	24	8	-60652.13	2383022.35
15-19	2	17	34	10	-40318.32	1382515.28
20-24	3	22	66	13	-25128.01	735999.44
25-29	5	27	135	18	-14331.20	348104.84
30-34	7	32	224	25	-7177.89	138461.46
35-39	14	37	518	39	-2918.08	41699.31
40-44	11	42	462	50	-801.77	7448.40
45-49	15	47	705	66	-78.95	338.71
50-54	21	52	1092	86	0.36	0.25
55-59	15	57	855	101	186.17	1063.03
60-64	25	62	1550	127	1228.48	13157.03
65-69	12	67	804	138	3877.29	60912.26
70-74	8	72	576	146	8882.60	183958.73
75-79	6	77	462	152	16994.42	436926.42
80-84	5	82	410	157	28962.73	889445.34
85-89	2	87	174	159	45537.54	1626145.50
90-94	1	92	92	160	67468.85	2746656.88
Total =>	160		8206		-402135.97	20744015.4

Ref: Date:

Mean, $\bar{x} = \frac{\Sigma f(x)}{f}$

$= \frac{8206}{160} \Rightarrow \bar{x} = 51.29$

ii) Skewness $= \frac{\Sigma (x_i - \bar{x})^3}{(N-1)s^3}$ Kurtosis $= \frac{\Sigma (x_i - \bar{x})^4}{(N-1)s^4}$

$= \frac{-402135.97}{(160-1)(18.19)^3}$ $= \frac{2074401.54}{(160-1)(18.19)^4}$

$= -0.42$ $= 0.119$

iii) - The skewness from the data, $s < 0$ shows the distribution points in the negative direction. The skewness is negatively skewed, and skewed to the left.

- The kurtosis, $k < 3$ shows that it is a platykurtic distribution.

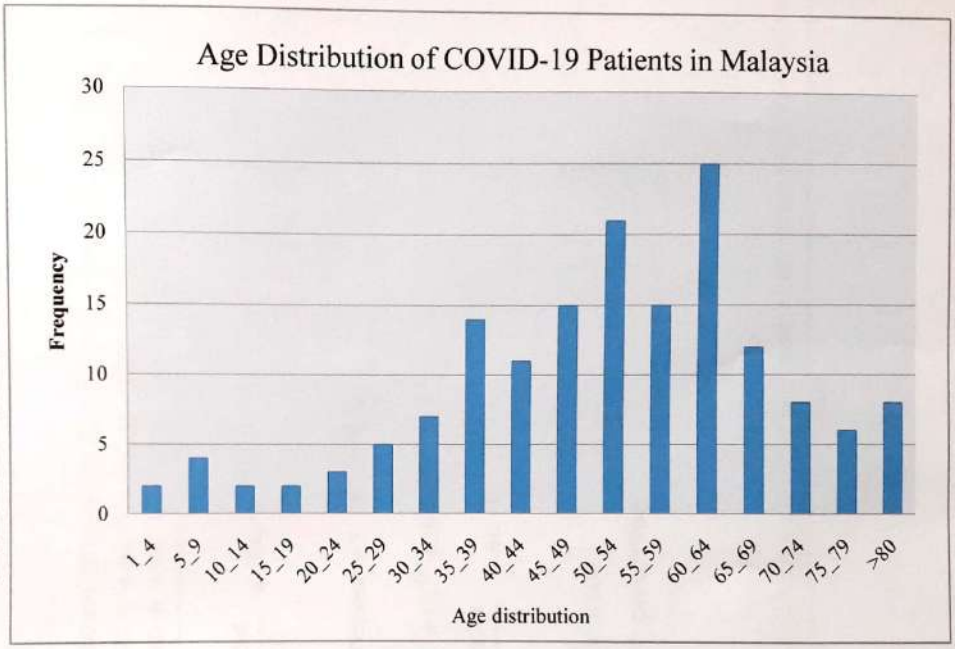
QUESTION 4 a)

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

Variance	332.021
Standard deviation	18.22144
Max value	91
Min value	2
Range	89

QUESTION 4 b ii)

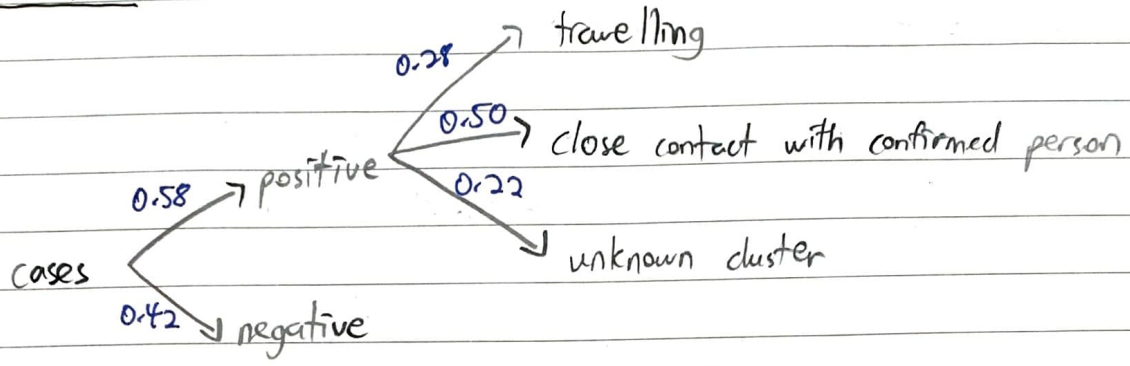
Age group	Case count
1_4	2
5_9	4
10_14	2
15_19	2
20_24	3
25_29	5
30_34	7
35_39	14
40_44	11
45_49	15
50_54	21
55_59	15
60_64	25
65_69	12
70_74	8
75_79	6
>80	8



Ref: Date:

Question 5

(a)



(b) Cases with negative screening, $P(N) = 1 - P(P)$
$= 1 - 0.58$
$= 0.42$

(c) Negative results => 0.42×72 people
$= 30.24$
$= 30$ people

(d) Positive cases with travelling history & close contact => $(0.58 \times 0.28) + (0.58 \times 0.50)$
$= 0.45$

(e) Death rate = 1.6%
$= 0.016$
, Recovery rate = 34%
$= 0.34$

i) Infected person under treatment => $1 - 0.016 - 0.34$
$= 0.644$

ii) Number of people under treatment => 0.644×2766
$= 1781.304$
$= 1781$ people

Ref: Date:

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, $\bar{x} = \dfrac{[0(0.091)+1(0.182)+2(0.182)+3(0.227)+4(0.137)+5(0.091)+6(0.045)+7(0.000)+8(0.045)]}{0.091+0.182+0.182+0.227+0.137+0.091+0.045+0.000+0.045}$

$\bar{x} = 2.86$

Variance, $\sigma^2 = [0^2(0.091)+1^2(0.182)+2^2(0.182)+3^2(0.227)+4^2(0.137)+5^2(0.091)+6^2(0.045)+7^2(0.000)+8^2(0.045)] - \bar{x}^2$

$\sigma^2 = 11.92 - (2.86)^2$

$\sigma^2 = 3.74$

Ref: Date:

(b)

Pre-existing condition	Probability, P(X)	Num. of Patients
Cardiovasaular disease	0.105	8
Diabetes	0.073	4
Chronic respiratory disease	0.063	10
Hypertension	0.060	8
Cancer	0.056	5
No pre-existing conditions.	0.009	21

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

$= [{}^{8}C_0 (0.105)^0 (0.895)^8] + [{}^{8}C_1 (0.105)^1 (0.895)^7] + [{}^{8}C_2 (0.105)^2 (0.895)^6]$

$= 0.9568$

ii) $P(X \geq 2) = P(X=2) + P(X=3) + P(X=4)$

$= [{}^{4}C_2 (0.073)^2 (0.927)^2] + [{}^{4}C_3 (0.073)^3 (0.927)^1] + [{}^{4}C_4 (0.073)^4 (0.927)^0]$

$= 0.029$

iii) $P(X \leq 1) = P(X=0) + P(X=1)$

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

$= 1.475 \times 10^{-10}$

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

$= [{}^{5}C_2 (0.944)^2 (0.006)^3] + [{}^{5}C_3 (0.944)^3 (0.006)^2] + [{}^{5}C_4 (0.944)^4 (0.006)^1]$

$= 0.2503$

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

$= 0.8271$

Ref: Date:

(c) male dying = 0.047 (n=7), female dying = 0.028 (n=5)

i) x = probability of infected patients dying.

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

iii) $P(5) = (1-0.953)^{5-1}(0.953)$ $\quad 1-0.047 = 0.953$ (recovery)
$= 4.65 \times 10^{-6}$

(d) $N = 4$, $\sigma = 5$

i) $P(X>6) = P\left(z > \frac{6-4}{5}\right)$
$= P(z > 0.4)$
$= 0.6554$

ii) $P(2<x<4) = P\left(\frac{2-4}{5} < z < \frac{4-4}{5}\right)$
$= P(-0.4 < z < 0)$
$= P(z<0) - [1 - P(z<0.4)]$
$= 0.50 - [1 - 0.65542]$
$= 0.1554$

iii) $P(x<0) = P\left(z < \frac{0-4}{5}\right)$
$= P(z < -0.8)$
$= 1 - P(z > 0.8)$
$= 1 - 0.7881$
$= 0.2119$