No. :

Date :

KHOR YONG XIN

PSDA - Assignment 3

① a.

Variable	Discrete / Continuous	Qualitative / Quantitative	Measuring scale
Case number	Discrete	Quantitative	Ratio
Gender	Discrete	Qualitative	Nominal
Age	Continuous	Quantitative	Ratio
Nationality	Discrete	Qualitative	Nominal
Date confirmed	Discrete	Quantitative	Interval
Date discharged	Discrete	Quantitative	Interval
Hospital	Discrete	Qualitative	Nominal
Total of cured patient	Discrete	Quantitative	Ratio

b. The data in Table 1 is considered as primary data source. This is because the data is collected by the government and not from the existing data bases or record reviews. All the data collected are the investigator that conducting the survey. The data is not exist before.

c. i) Pie chart

ii) Histogram

iii) Bar chart

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: 36|7 = 36.7°C

ii). $\frac{30}{4} = 7.5$, $Q_1 = X[8]$

$= 36.4°C$

$\frac{30}{2} = 15$, $Q_2 = \frac{X[15] + X[16]}{2}$

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

$= 37.0°C$

$\frac{30 \times 3}{4} = 22.5$, $Q_3 = X[23]$

$= 37.7°C$

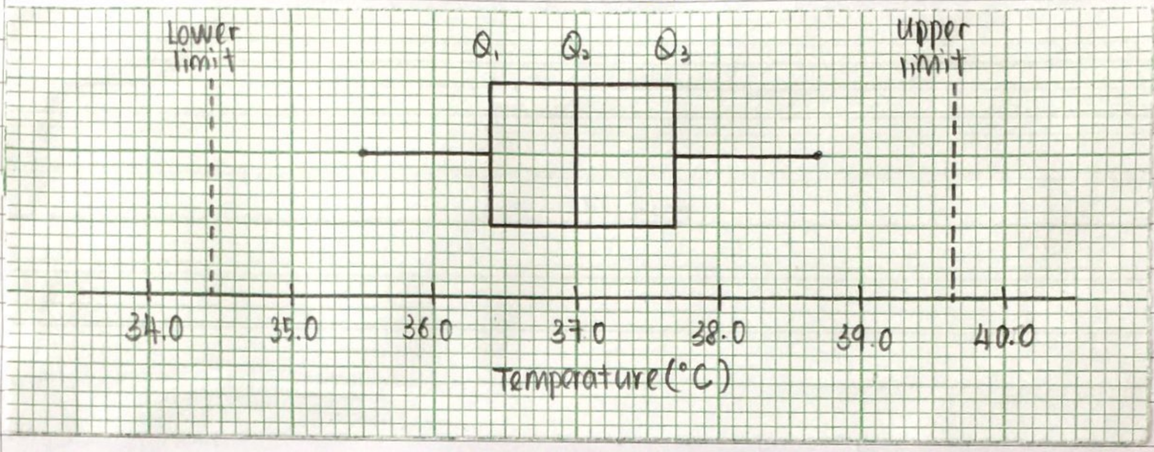
No. :

Date :

iii). $IQR = Q_3 - Q_1$
$= 37.7 - 36.4$
$= 1.3$

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

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

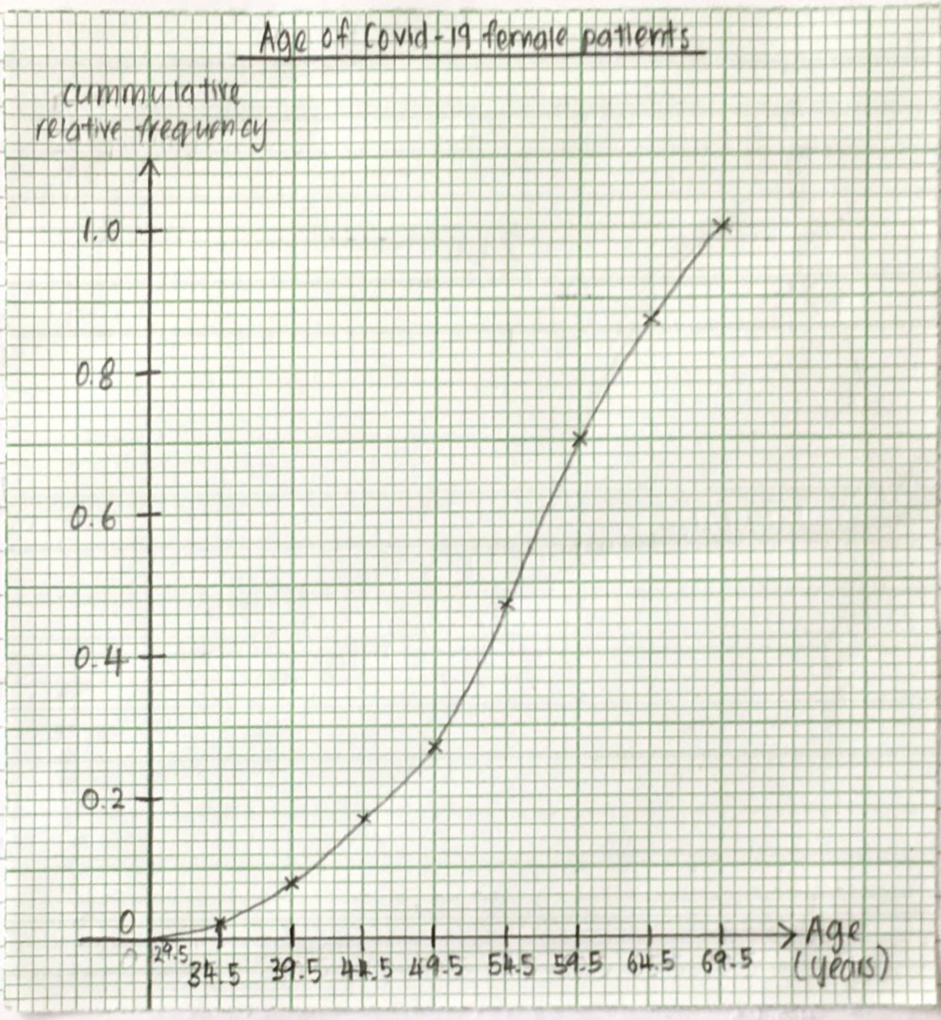


iv). There is neither mild nor extreme outlier. The lowest data value or called lowest temperature measured is 35.5°C while the highest temperature measured / highest data value is 38.7°C. The data ~~is~~ can be said to be normal distribution as the mode, median and mean lies on 37.0°C. But, there are two modes (bimodal), 36.7 and 37.0°C thus, it is positive distribution because mode < median and mean.

b. i).

Age (years)	Frequency	Relative Frequency	Cummulative 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
Total	444	1	

ii).



③ a.i). modal class : 56-60

$L = \frac{55+56}{2} = 55.5$

$W = 60.5 - 55.5 = 5$

$$\text{Mode} = L + \left(\frac{f_1 - f_0}{2f_1 - f_0 - f_2}\right)W$$

$$= 55.5 + \left(\frac{182 - 143}{2\times182 - 143 - 137}\right)(5)$$

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

$$= 57.82$$

ii). $N = 27+23+43+78+123+180+144+136+118+133+143+182+137+90+46+20+14$

$= 1637$

$$Q_2 = \frac{x}{2}$$

$$= \frac{1637}{2}$$

$$= 818.5$$

Median class = 41-45

No. :

Date :

$$\text{Median} = L + \left(\frac{\frac{N}{2} - CF_p}{f}\right) W$$

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

$$= 43.23$$

$$L = \frac{40+41}{2} = 40.5$$

$$W = 45.5 - 40.5 = 5$$

(iii).

Age Group	Midpoint, x	No. of cases, f	fx	Age Group	Midpoint, x	No. of cases, f	fx
1–5	$\frac{0.5+5.5}{2} = 3$	27	$27 \times 3 = 81$	46-50	$\frac{45.5+50.5}{2} = 48$	133	$133 \times 48 = 6384$
6–10	$\frac{5.5+10.5}{2} = 8$	23	$23 \times 8 = 184$	51-55	$\frac{50.5+55.5}{2} = 53$	143	$143 \times 53 = 7579$
11–15	$\frac{10.5+15.5}{2} = 13$	43	$43 \times 13 = 559$	56-60	$\frac{55.5+60.5}{2} = 58$	182	$182 \times 58 = 10556$
16-20	$\frac{15.5+20.5}{2} = 18$	78	$78 \times 18 = 1404$	61-65	$\frac{60.5+65.5}{2} = 63$	137	$137 \times 63 = 8631$
21-25	$\frac{20.5+25.5}{2} = 23$	123	$123 \times 23 = 2829$	66-70	$\frac{65.5+70.5}{2} = 68$	90	$90 \times 68 = 6120$
26-30	$\frac{25.5+30.5}{2} = 28$	180	$180 \times 28 = 5040$	71-75	$\frac{70.5+75.5}{2} = 73$	46	$46 \times 73 = 3358$
31-35	$\frac{30.5+35.5}{2} = 33$	144	$144 \times 33 = 4752$	76-80	$\frac{75.5+80.5}{2} = 78$	20	$20 \times 78 = 1560$
36-40	$\frac{35.5+40.5}{2} = 38$	136	$136 \times 38 = 5168$	81-85	$\frac{81.5+85.5}{2} = 83$	14	$14 \times 83 = 1162$
41-45	$\frac{40.5+45.5}{2} = 43$	118	$118 \times 43 = 5074$				

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

$$= \frac{81+184+559+1404+2829+5040+4752+5168+5074+6384+7579+10556+8631+6120+3358+1560+1162}{1637}$$

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

$$= 43.03$$

b. i). $$\text{Mean} = \frac{\sum x}{n}$$

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

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

$$= 36.7^\circ C$$

ii). Mode $= 36.7^\circ C$

iii). $\frac{20}{2} = 10$

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

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

$$= 36.7^\circ C$$

Scanned with CamScanner

No. :

Date :

iv). Percentile for 36.9 $= \frac{\text{Number of values less than } 36.9}{\text{Total number of values}} \times 100$

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

$=$ 80th percentile

v). $s = \sqrt{\frac{\Sigma(x-\bar{x})^2}{n-1}}$

$\bar{x} = \frac{36.5\times 3 + 36.6 \times 3 + 36.7 \times 8 + 36.8 \times 2 + 36.9 + 37.0 \times 3}{20}$

$= \frac{734.4}{20}$

$= 36.72$

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

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

$= 0.1576$

vi). skewness $= \frac{\Sigma(x-\bar{x})^3}{(n-1)s^3}$

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

$= \frac{0.03552}{0.07437}$

$= 0.4776 > 0$

~~∴ The data is normally skewed because the value of skewness is close to zero. The central of measurement : mean, mode, median are all lie on 36.7°C closely~~

∴ Mode is 36.7, median is also 36.7 while the mean for the data is 36.72. From this, it can be seen that mean > mode, but the different is not large. Hence, the distribution is slightly positive-skewed. According to the value of skewness calculated, it is 0.4776 which is slightly greater than zero. Thus, it can be concluded than the distribution is slightly positively skewed.

Uni

Scanned with CamScanner

No. :

Date :

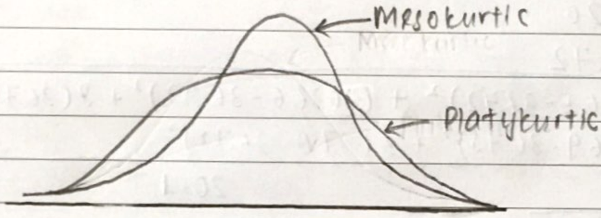
vii). Kurtosis $= \dfrac{\sum (x-\bar{x})^4}{(n-1)s^4}$

$$= \frac{3(36.5-36.72)^4 + 3(36.6-36.72)^4 + 8(36.7-36.72)^4 + 2(36.8-36.72)^4 + (36.9-36.72)^4 + 3(37.0-36.72)^4}{(20-1)(0.1576)^4}$$

$$= \frac{0.0272}{0.0117}$$

$$= 2.32 < 3$$

∴ The distribution is platykurtic because the value of kurtosis is less than 3. When compare to normal distribution or called mesokurtic, its height is shorter or central peak is lower, tails are shorter and thinner.



4 a. range = 91 - 2
= 89

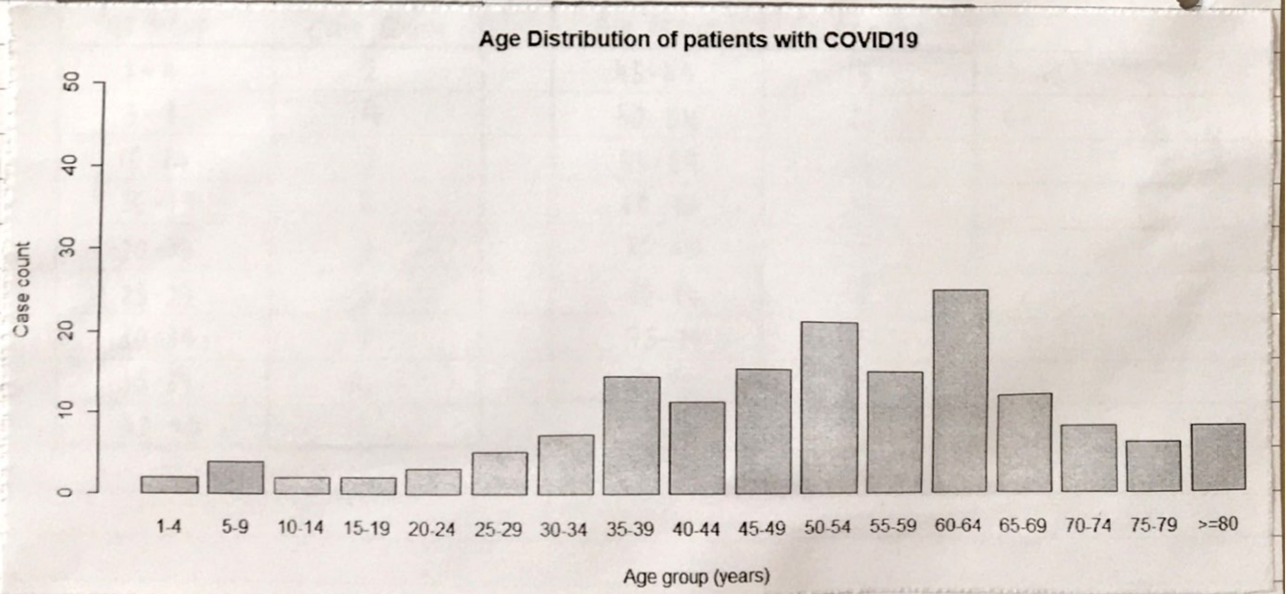
by using R studio:

mean = 51.22 , mean $= \frac{\sum x}{n} = \frac{8195}{160} = 51.22$

variance = 332.021 , variance $= \frac{\sum (x-\bar{x})^2}{n-1} = \frac{(11-51.22)^2 + (2-51.22)^2 + \ldots + (55-51.22)^2}{160-1} = 332.02$

standard deviation $= \sqrt{332.021}$, $s = \sqrt{\text{variance}}$
$= 18.221$

b.i).



No. :

Date :

ii) $\text{skewness} = \frac{\sum f(m-\bar{x})^3}{(n-1)s^3}$, $\bar{x} = 51.22$, $s = 18.221$, $n = 160$

$$= \frac{\begin{aligned}&2(2.5-51.22)^3 + 4(7-51.22)^3 + 2(12-51.22)^3 + 2(17-51.22)^3 + 3(22-51.22)^3 \\ &+ 5(27-51.22)^3 + 7(32-51.22)^3 + 14(37-51.22)^3 + 11(42-51.22)^3 + 15(47-51.22)^3 \\ &+ 21(52-51.22)^3 + 15(57-51.22)^3 + 25(62-51.22)^3 + 12(67-51.22)^3 + 8(72-51.22)^3 \\ &+ 6(77-51.22)^3 + 8(85.5-51.22)^3\end{aligned}}{(160-1)(18.221)^3}$$

$$= \frac{-852805.6896 + (-170742.9553) + 578226.794}{159(18.221)^3}$$

$$= -0.4630 < 0$$

According to generated distribution in (i), modal class is 60-64, mean = 51.22, mean < mode. The value of skewness calculated is less than zero. Hence the distribution is negatively-skewed.

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

$$= \frac{\begin{aligned}&2(2.5-51.22)^4 + 4(7-51.22)^4 + 2(12-51.22)^4 + 2(17-51.22)^4 + 3(22-51.22)^4 + \\ &5(27-51.22)^4 + 7(32-51.22)^4 + 14(37-51.22)^4 + 11(42-51.22)^4 + 15(47-51.22)^4 + \\ &21(52-51.22)^4 + 15(57-51.22)^4 + 25(62-51.22)^4 + 12(67-51.22)^4 + 8(72-51.22)^4 + \\ &6(77-51.22)^4 + 8(85.5-51.22)^4\end{aligned}}{159(18.221)^4}$$

$$= \frac{36224464.41 + 3332468.151 + 1595271.522}{159(18.221)^4}$$

$$= 2.3480 < 3$$

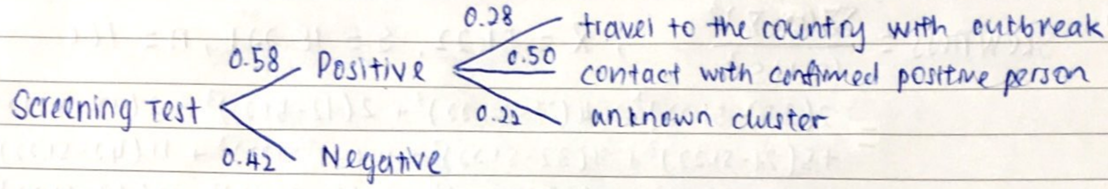
The value of kurtosis is smaller than 3. Hence the distribution is platykurtic.

(iii). By comparing the skewness of both graphs, the graph generated in (i) skew more to the left than the distribution in Figure 1. Hence it can be said that the graph generated in (i) skewed more negative than the graph in Figure 1.
While by comparing the kurtosis of both graphs, the peak of in the graph generated in (i) is lower than in ~~Figu~~ the graph in Figure 1. The tails are shorter and thinner in the graph generated in (i) compared to the graph in Figure 1. If the graph generated in (i) is platykurtic, than the graph in Figure 1 may be leptokurtic.

No. :

Date :

⑤ a.



b. P(negative) = 1 - 0.58
= 0.42

c. n(negative) = 0.42 × 72
= 30.24
≈ 31

d. P((positive ∩ travel to the country with outbreak) ∪ (positive ∩ close contact with person infected
= 0.58 × 0.28 + 0.58 × 0.50
= 0.4524

e. P(death) = 0.016 , P(Recovery) = 0.34 , n = 2766

i). P(infected person under treatment)
= 1 - 0.016 - 0.34
= 0.644

ii) n = 0.644 × 2766
= 1781.304
≈ 1782

No. :

Date :

⑥ a. i). $P(X \leq 2) = P(X=0) + P(X=1) + P(X=2)$
$= 0.091 + 0.182 + 0.182$
$= 0.455$

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 $= \sum x p(x)$
$= 0(0.091) + 1(0.182) + 2(0.182) + 3(0.227) + 4(0.137) + 5(0.091) + 6(0.045) + 7(0.000) + 8(0.045)$
$= 0 + 0.182 + 0.364 + 0.681 + 0.548 + 0.455 + 0.27 + 0.36$
$= 2.86$

Variance $= \sum (x-\mu)^2 p(x)$
$= (0-2.86)^2(0.091) + (1-2.86)^2(0.182) + (2-2.86)^2(0.182) + (3-2.86)^2(0.227) + (4-2.86)^2(0.137) + (5-2.86)^2(0.091) + (6-2.86)^2(0.045) + (7-2.86)^2(0.000) + (8-2.86)^2(0.045)$
$= 3.7404$

b. $\overset{8}{P(CD)} = 0.105$, $\overset{4}{P(D)} = 0.073$, $\overset{10}{P(CR)} = 0.063$, $\overset{8}{P(H)} = 0.060$, $\overset{5}{P(C)} = 0.056$,
$P(N) = 0.009$, $n = 56$

i). $P(X \leq 2) = P(X=0) + P(X=1) + P(X=2)$
$= {}^8C_0(0.105)^0(1-0.105)^8 + {}^8C_1(0.105)^1(1-0.105)^7 + {}^8C_2(0.105)^2(1-0.105)^6$
$= 0.4117 + 0.3864 + 0.1587$
$= 0.9568$

ii). $P(X \geq 2) = P(X=2) + P(X=3) + P(X=4)$
$= {}^4C_2(0.073)^2(1-0.073)^2 + {}^4C_3(0.073)^3(1-0.073)^1 + {}^4C_4(0.073)^4(1-0.073)^0$
$= 0.0275 + 1.4425 \times 10^{-3} + 2.8398 \times 10^{-5}$
$= 0.02897$

iii). $P(X \leq 1) = P(X=0) + P(X=1)$
$= {}^{10}C_0(1-0.063)^0(0.063)^{10} + {}^{10}C_1(1-0.063)^1(0.063)^9$
$= 9.8493 \times 10^{-13} + 1.46249 \times 10^{-10}$
$= 1.4747 \times 10^{-10}$

Scanned with CamScanner

No. :

Date :

iv). $P(2 \leq X \leq 4) = P(X=2) + P(X=3) + P(X=4)$
$= {}^5C_2 (1-0.056)^2 (0.056)^3 + {}^5C_3 (1-0.056)^3 (0.056)^2 + {}^5C_4 (1-0.056)^4 (0.056)$
$= 1.5650 \times 10^{-3} + 0.0264 + 0.2224$
$= 0.2504$

v). $X = 56 - 8 - 4 - 10 - 8 - 5$
$= 21$
$P(X=21) = {}^{21}C_{21} (1-0.009)^{21} (0.009)^0$
$= 0.8271$

c. $P(M) = 0.047$, $P(F) = 0.028$, $n(M) = 7$, $n(F) = 5$

i). X is the number of death among female or male cases until the first death is observed. X is a geometric distribution.

ii). $P(X=3) = (1-0.028)^3 (0.028)$
$= 0.0265$

iii). $P(X=5) = (0.047)^4 (1-0.047)$
$= 4.6503 \times 10^{-6}$

d. $n = 468$, $X \sim N(4, 5^2)$, $\mu = 4$, $\sigma = 5$, $Z = \frac{X-\mu}{\sigma}$

i). $P(X > 6) = P(Z > \frac{6-4}{5})$
$= P(Z > 0.4)$
$= 1 - P(Z \leq 0.4)$
$= 1 - 0.6554$
$= 0.3446$

ii). $P(2 \leq X \leq 4) = P(\frac{2-4}{5} \leq Z \leq \frac{4-4}{5})$
$= P(-0.4 \leq Z \leq 0)$
$= P(Z < 0) - P(Z < -0.4)$
$= P(Z < 0) - [1 - P(Z < 0.4)]$
$= 0.5 - (1 - 0.6554)$
$= 0.1554$

iii). $P(X < 0) = P(Z < \frac{0-4}{5})$
$= P(Z < -0.8)$
$= 1 - P(Z < 0.8)$
$= 1 - 0.7881$
$= 0.2119$

Scanned with CamScanner