MA4413 2015 Solutions.

Question 1

(a)(i) $P_r(A \cup B) = P_r(A) + P_r(B) - P_r(A \cap B)$

$= 0.7 + 0.6 - 0.5$

$= 0.8$

(ii) $P_r(A^c \cup B^c) = 1 - P_r((A^c \cup B^c)^c)$

$= 1 - P_r(A \cap B)$.

$= 1 - 0.5$

$= 0.5$

(iii) $P_r(A|B) = \dfrac{P_r(A \cap B)}{P_r(B)} = \dfrac{0.5}{0.6}$

$= 0.833$

(b) (i)

Position	1	2	3	4	5	6	7	8	9	10	11	12
Value	17	19	24	24	24	26	29	32	32	33	34	34

Position of $Q1 = \frac{n+1}{4} = \frac{12+1}{4} = \frac{13}{4} = 3.25 \Rightarrow$ between 3 & 4

$Q2 = 2\left(\frac{n+1}{4}\right) = 2(3.25) = 6.5 \Rightarrow$ between 6 & 7

$Q.3 = 3\left(\frac{n+1}{4}\right) = 3(3.25) = 9.75 \Rightarrow$ between 9 & 10.

$\Rightarrow Q1 = \frac{24+24}{2} = 24$

$Q2 = \frac{26+29}{2} = 27.5$

$Q3 = \frac{32+33}{2} = 32.5$

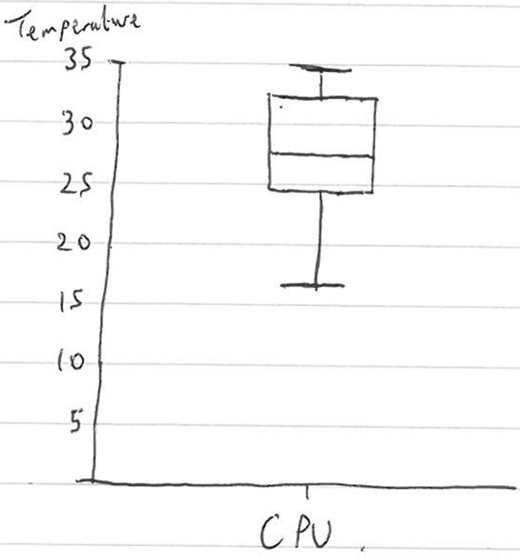
(ii) $IQR = Q3 - Q1 = 32.5 - 24 = 8.5$

$UF = Q3 + 1.5\,IQR = 32.5 + 1.5(8.5) = 45.25$

$LF = Q1 - 1.5\,IQR = 24 - 1.5(8.5) = 11.25$.

No values above UF or below LF $\Rightarrow$ no outliers

(iii)



(iv) No - most values are above 22°.

(c) (i) μ = unknown

(ii) $H_0: \mu = 2$
$H_a: \mu \neq 2$

(iii) P-value = 0.001195 is very small (less than 0.01)
$\Rightarrow$ there is strong evidence to reject H_0.
i.e., it appears that $\mu \neq 2$.

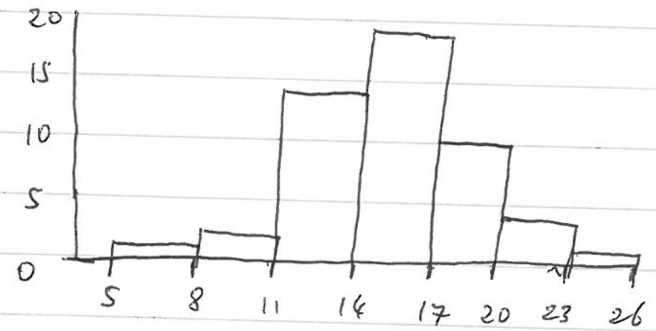
(iv) We are 95% confident that μ is in the range [2.36, 3.11].
This does not support the possibility that $\mu = 2$ since $\mu = 2$ is not in the C.I., i.e., we reject H_0.

Question 2

(a) (i) Class width $= \frac{\text{range}}{\text{\# of classes}} = \frac{16}{7} = 2.28 \Rightarrow 3.$

Class	Frequency
5 – 7.9	1
8 – 10.9	2
11 – 13.9	14
14 – 16.9	18
17 – 19.9	10
20 – 22.9	4
23 – 25.9	1
	50

(ii)



(iii) Symmetric ⇒ the mean is an appropriate measure of centrality.

(b) (i) Standard deviation: a measure of spread around the mean
IQR: the range of the middle 50% of data.

(ii) When the data is skewed.

(iii) It is usually not feasible (or possible) to access the whole population. Thus, we collect a sample of data and calculate a statistic (e.g. $\bar{x}$ or $\hat{p}$) to estimate a parameter (μ or p).

(c) (i) $Pr(R_1) = 0.3$ $\quad$ $Pr(L \mid R_1) = 0.15$

$Pr(R_2) = 0.7$ $\quad$ $Pr(L \mid R_2) = 0.04$

$$\Rightarrow Pr(L \cap R_1) = Pr(R_1)\,Pr(L \mid R_1) = 0.3(0.15) = 0.045$$

$$Pr(L \cap R_2) = Pr(R_2)\,Pr(L \mid R_2) = 0.7(0.04) = 0.028$$

(ii) $$Pr(L) = Pr(L \cap R_1) + Pr(L \cap R_2) = 0.045 + 0.028 = 0.073$$

(iii) $$Pr(R_1 \mid L^c) = \frac{Pr(R_1 \cap L^c)}{Pr(L^c)}$$

$$= \frac{Pr(R_1)\,Pr(L^c \mid R_1)}{Pr(L^c)}$$

$$= \frac{Pr(R_1)\left[1 - Pr(L \mid R_1)\right]}{1 - Pr(L)}$$

$$= \frac{0.3(1 - 0.15)}{1 - 0.073}$$

$$= 0.2751$$

Question 3

(a) (i) $\bar{x} = \frac{\sum x}{n} = \frac{5+2+2+3+1+3}{6} = \frac{16}{6} = 2.66667$.

(ii) $\sum x^2 = 5^2 + 2^2 + 2^2 + 3^2 + 1^2 + 3^2$
$= 25 + 4 + 4 + 9 + 1 + 9 = 52$

$\Rightarrow s^2 = \frac{\sum x^2 - n\bar{x}^2}{n-1} = \frac{52 - 6(2.66667^2)}{5}$

$= \frac{9.333}{5} = 1.86667$

$\Rightarrow s = \sqrt{s^2} = \sqrt{1.86667} = 1.36626$

(b) (i) Categorical ("Yes" / "No")

(ii) p = unknown

(iii) $\hat{p} = \frac{50}{168} = 0.2976$

(iv) $\hat{p} \pm Z_{\alpha/2}\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$ where $n = 168$
$\alpha = 0.01 \Rightarrow \alpha/2 = 0.005$
$\Rightarrow Z_{0.005} = 2.58$

$0.2976 \pm 2.58\sqrt{\frac{0.2976(0.7024)}{168}}$

0.2976 ± 0.091

$[0.2066, 0.3886]$ → This does not support the researcher's belief that $p = 0.4$ since $p = 0.4$ is not contained in the interval.

(v) We require $\hat{p} \pm 0.03$

i.e., $2.58\sqrt{\frac{p(1-p)}{n}} = 0.03$

Using $\hat{p}$ in place of p we have

$$2.58\sqrt{\frac{0.2976(0.7024)}{n}} = 0.03$$

$$\sqrt{\frac{0.2976(0.7024)}{n}} = \frac{0.03}{2.58}$$

$$\frac{1}{n} = \left(\frac{0.03}{2.58}\right)^2 \frac{1}{0.2976(0.7024)}$$

$$\Rightarrow n = \left(\frac{2.58}{0.03}\right)^2 0.2976(0.7024)$$

$$= 1546.$$

(c)(i) $H_0: \mu_1 - \mu_2 = 0$

$H_a: \mu_1 - \mu_2 \neq 0.$

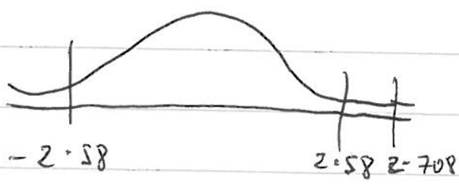
(ii) $$Z = \frac{(\bar{x}_1 - \bar{x}_2) - 0}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}} = \frac{83.1 - 80.1}{\sqrt{\frac{30.6}{40} + \frac{18.5}{40}}}$$

$$= \frac{3}{1.108} = 2.708$$

Two-tailed test $\Rightarrow$ critical values are $\pm Z_{\alpha/2} = \pm Z_{0.005} = \pm 2.58$

(since $\alpha = 0.01$, $\alpha/2 = 0.005$).

(iii)



- 2.708 is outside of $\pm 2.58 \Rightarrow$ we reject H_0, i.e., it appears $\mu_1 \neq \mu_2$.
- The hours of gameplay are not equal, "Game 1" has more gameplay hours.

Question 4

(a) (i) $k = 1 - (0.1 + 0.4 + 0.2) = 1 - 0.7 = 0.3$

(ii) $$\begin{aligned} EX &= 0(0.1) + 3(0.4) + 6(0.3) + 9(0.2) \\ &= 0 + 1.2 + 1.8 + 1.8 \\ &= 4.8 \end{aligned}$$

(iii) $$\begin{aligned} EX^2 &= 0^2(0.1) + 3^2(0.4) + 6^2(0.3) + 9^2(0.2) \\ &= 30.6 \end{aligned}$$

$$\begin{aligned} Var\,X &= EX^2 - (EX)^2 \\ &= 30.6 - (4.8)^2 = 7.56 \end{aligned}$$

$$Sd\,X = \sqrt{Var\,X} = \sqrt{7.56} = 2.749$$

(b)(i) $EX = np = 80(0.04) = 3.2$

(ii) $n = 15 \Rightarrow p(x) = \binom{n}{x} p^x (1-p)^{n-x} = \binom{15}{x} 0.04^x \cdot 0.96^{15-x}$

$$\begin{aligned} Pr(2 \leq X \leq 5) &= p(2) + p(3) + p(4) + p(5) \\ &= 0.0988 + 0.0178 + 0.0022 + 0.0002 \\ &= 0.119 \end{aligned}$$

(iii) $Pr(X > 8) = Pr(X \geq 9) = 0.0190$ (from tables).
($n=100$, $p=0.04$, $r=9$).

(iv) The binomial distribution arises as a sequence of independent Bernoulli trials. If the disease is contagious then the occurrence of the disease is not independent as one person can pass it on to another. Similarly, individuals in the same family will be more alike and, hence, not independent.

(c) (i) $\lambda = 7/hr.$

$\Rightarrow \lambda = 7(\frac{1}{2}) = 3.5 / \frac{1}{2}hr. \quad \Rightarrow p(x) = \frac{\lambda^x e^{-\lambda}}{x!} = \frac{3.5^x e^{-\lambda}}{x!}$

$$P_r(X \geq 3) = 1 - P_r(X < 3)$$
$$= 1 - [p(0) + p(1) + p(2)]$$
$$= 1 - [0.0302 + 0.1057 + 0.1850]$$
$$= 1 - 0.3209 = 0.6791$$

Or, using, tables $P_r(X \geq 3) = 0.6792 \quad (m = 3.5,\ r = 3)$

(ii) $\lambda = 7(3) = 21 / 3\text{-hrs.}$

$$P_r(15 \leq X \leq 25) = P_r(X \geq 15) - P_r(X \geq 26)$$
$$= \underset{(m=21,\ r=15)}{0.9284} - \underset{(m=21,\ r=26)}{0.1623}$$

$$= 0.7661$$

(iii) $T \sim \text{Exponential}(\lambda = 7)$

$$P_r(T < \tfrac{5}{60}) = 1 - P_r(T > \tfrac{5}{60}) = 1 - e^{-7(\frac{5}{60})}$$
$$= 1 - 0.558$$
$$= 0.442$$

Since we are working in hours

Question 5

(a) (i)

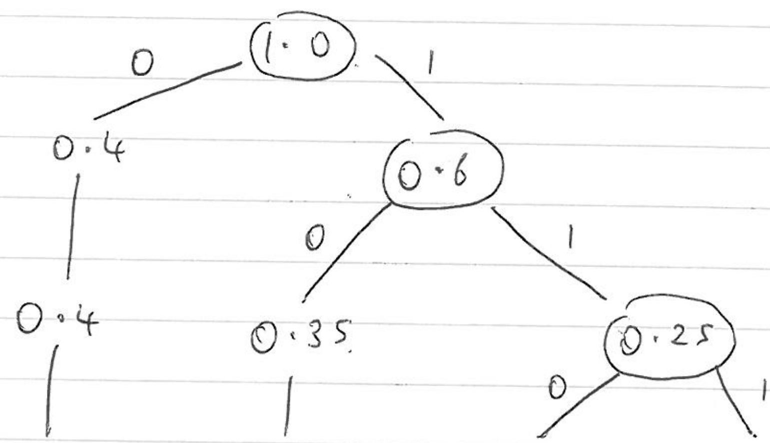
x	a	b	c	d
$p(x)$	0.4	0.1	0.35	0.15
$h(x)$	1.323	3.323	1.515	2.737

$$H(X) = E(h(X))$$

$$= 1.323(0.4) + 3.323(0.1) + 1.515(0.35) + 2.737(0.15)$$

$$= 1.8018 \text{ bits.}$$

(ii)



$p(x)$	0.4	0.35	0.15	0.1
x	a	c	d	b
$C(x)$	0	10	110	111
(iii) $l(x)$	1	2	3	3

$$E(L) = 1(0.4) + 2(0.35) + 3(0.15) + 3(0.1)$$

$$= 0.4 + 0.7 + 0.45 + 0.3$$

$$= 1.85$$

$$\Rightarrow \text{efficiency} = \frac{H(X)}{E(L)} = \frac{1.8018}{1.85} = 0.97$$

i.e., 97%.

(b) (i) $P_r(X < 25) = P_r\left(Z < \frac{25-20}{3}\right)$

$= P_r(Z < 1.67)$

$= 1 - P_r(Z > 1.67)$

$= 1 - 0.0475$

$= 0.9525.$

(ii) $P_r(23.5 < X < 28.4) = P_r(X > 23.5) - P_r(X > 28.4)$

$= P_r\left(Z > \frac{23.5-20}{3}\right) - P_r\left(Z > \frac{28.4-20}{3}\right)$

$= P_r(Z > 1.17) - P_r(Z > 2.8)$

$= 0.1210 - 0.00256$

$= 0.11844$

(iii) $P_r(X > x) = 0.35$

$\Rightarrow P_r\left(Z > \frac{x-20}{3}\right) = 0.35$

But $P_r(Z > 0.39) = 0.3483 \simeq 0.35$

$\Rightarrow \frac{x-20}{3} = 0.39 \Rightarrow x = 20 + 0.39(3)$

$= 21.17$

(iv) $\bar{X} \sim N\left(\mu, \frac{\sigma}{\sqrt{n}}\right) = N\left(20, \frac{3}{\sqrt{45}} = 0.4472\right)$

$\Rightarrow P_r(\bar{X} > 20.8) = P_r\left(Z > \frac{20.8-20}{0.4472}\right)$

$= P_r(Z > 1.79) = 0.0367$

(v) $X_1 + X_2 \sim N\left(\mu_1 + \mu_2, \sqrt{\sigma_1^2 + \sigma_2^2}\right) = N(40, \sqrt{18} = 4.2426).$

$P_r(X_1 + X_2 > 45.7) = P_r\left(Z > \frac{45.7-40}{4.2426}\right)$

$= P_r(Z > 1.34) = 0.0901$

Question 6

(a) (i) $H_0: \mu_1 - \mu_2 = 0$

$H_a: \mu_1 - \mu_2 \neq 0$

(ii) $n_1 = 8, \quad \bar{X}_1 = 7.96, \quad s_1 = 0.73$

$n_2 = 7, \quad \bar{X}_2 = 6.83, \quad s_2 = 2.36$

95% C.I. $\Rightarrow \alpha = 0.05 \Rightarrow \alpha/2 = 0.025$

Two small samples, unequal variance $\Rightarrow t_{v, \alpha/2}$

where $v = \dfrac{(a+b)^2}{\frac{a^2}{n_1-1} + \frac{b^2}{n_2-1}}$

$$\left[\begin{array}{l} a = \dfrac{s_1^2}{n_1} = \dfrac{0.73^2}{8} = 0.0666 \\ b = \dfrac{s_2^2}{n_2} = \dfrac{2.36^2}{7} = 0.7957 \end{array}\right]$$

$$= \frac{(0.0666 + 0.7957)^2}{\frac{0.0666^2}{7} + \frac{0.7957^2}{6}} = 7.004 \simeq 7.$$

t-tables only have whole numbers.

$$(\bar{X}_1 - \bar{X}_2) \pm t_{7, 0.025} \sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}$$

$$(7.96 + 6.83) \pm 2.365 \sqrt{\frac{0.73^2}{8} + \frac{2.36^2}{7}}$$

$1.13 \pm 2.365(0.4286)$

1.13 ± 2.196

$[-1.066, 3.326]$

(iii) The interval includes $\mu_1 - \mu_2 = 0$ which supports H_0, i.e., no difference between means. It appears that customers spend equal amounts using both website designs.

(b)(i) $\lambda_s = \frac{1}{E(T_s)} = \frac{1}{0.05} = 20$ cust/hr

(ii) $\rho = \frac{\lambda_a}{\lambda_s} = \frac{15}{20} = 0.75$

$\Rightarrow$ The service node is in use 75% of the time and idle 25% of the time.

(iii) $T \sim Exp(\lambda)$ where $\lambda = \lambda_s - \lambda_a = 20 - 15 = 5$.

$\Rightarrow E(T) = \frac{1}{\lambda} = \frac{1}{5}$ hours $= \frac{1}{5}(60) = 12$ minutes.

$$E(T_q) = E(T) - E(T_s) = \frac{1}{5} - 0.05$$
$$= 0.2 - 0.05$$
$$= 0.15 \text{ hours}$$
$$= 0.15(60) = 9 \text{ minutes}$$

(iv) $E(N) = \lambda_a E(T) = 15(0.2) = 3$ customers

$E(N_q) = \lambda_a E(T_q) = 15(0.15) = 2.25$ customers

(v) $Pr(T > \frac{45}{60}) = Pr(T > 0.75) = e^{-5(0.75)}$
$= 0.0235$

(vi) $ET = \frac{1}{\lambda} = \frac{1}{\lambda_s - \lambda_a} = \frac{1}{\lambda_s - 15} = \frac{5}{60}$ (5 mins in hours).

$\Rightarrow \lambda_s - 15 = \frac{60}{5} = 12$
$\lambda_s = 15 + 12 = 27$ cust/hr

$EN = \lambda_a ET = 15(\frac{5}{60}) = 1.25$ customers.