Statistics for Computing 2014

Q.1 (a)

Type A:	2.7	2.9	4.1	4.3	5.7	6.6	9.0	15.0
Type B:	1.0	1.0	1.8	2.3	2.3	3.3	3.5	5.5

$n = 8 \Rightarrow$ Position of $Q_1 = \frac{n+1}{4} = \frac{9}{4} = 2.25 \Rightarrow 2 \ \& \ 3$

$Q_2 = 2\left(\frac{n+1}{4}\right) = 2(2.25) = 4.5 \Rightarrow 4 \ \& \ 5$

$Q_3 = 3\left(\frac{n+1}{4}\right) = 3(2.25) = 6.75 \Rightarrow 6 \ \& \ 7$

Type A

(i) $Q_1 = \frac{2.9+4.1}{2} = 3.5$

$Q_2 = \frac{4.3+5.7}{2} = 5$

$Q_3 = \frac{6.6+9.0}{2} = 7.8$

(ii) $IQR = Q_3 - Q_1 = 7.8 - 3.5$
$= 4.3$

$UF = Q_3 + 1.5\, IQR$
$= 7.8 + 1.5(4.3)$
$= 14.25$

$LF = Q_1 - 1.5\, IQR$
$= 3.5 - 1.5(4.3)$
$= -2.95$

Type B

$Q_1 = \frac{1.0+1.8}{2} = 1.4$

$Q_2 = \frac{2.3+2.3}{2} = 2.3$

$Q_3 = \frac{3.3+3.5}{2} = 3.4$

$IQR = Q_3 - Q_1 = 3.4 - 1.4$
$= 2.0$

$UF = Q_3 + 1.5\, IQR$
$= 3.4 + 1.5(2.0)$
$= 6.4$

$LF = Q_1 - 1.5\, IQR$
$= 1.4 - 1.5(2.0)$
$= -1.6$

Type A

⇒ 15 is an outlier

From non-outliers: min = 2.7

max = 9.0

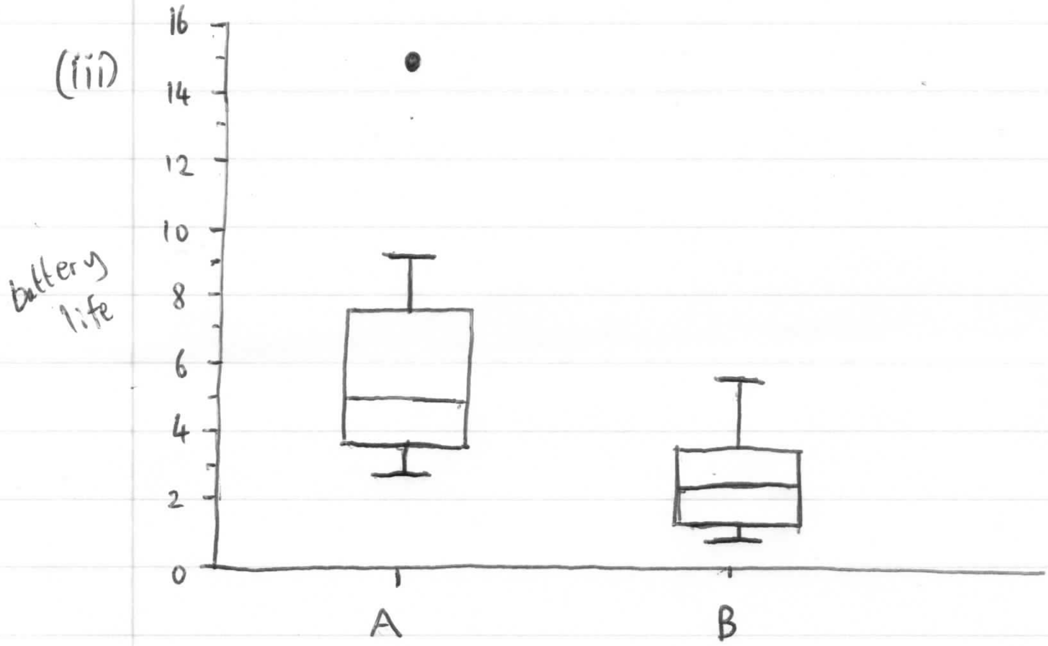
Type B.

⇒ no outliers.

min = 1.0

max = 5.5

(iii)



(iv) Type A appears to have a greater battery life.

(b) (i) Numeric continuous

(ii) Categorical

(iii) Numeric discrete

(iv) Categorical

(v) Numeric continuous.

(c) (i) The amount that the machine truly puts into each bottle on average.

μ = unknown.

(ii) The average for the sample.

$\bar{x} = 501.5$

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

$n = 40$ (large sample) $\Rightarrow Z_{0.025} = 1.96$.

$\bar{x} \pm Z_{0.025} \frac{s}{\sqrt{n}}$

$501.5 \pm 1.96 \left(\frac{3.05}{\sqrt{40}}\right)$

$501.5 \pm 1.96\,(0.482)$

501.5 ± 0.9452

$[500.55,\ 502.45]$

95% confident that μ is in this interval.
It does not include $\mu = 500$.
$\Rightarrow$ It appears that the machine is not working as programmed.
(average content is higher than it should be).

(iv) We require

$$Z \frac{s}{\sqrt{n}} = 0.5$$

$$\frac{1}{\sqrt{n}} = \frac{0.5}{Zs}$$

$$\sqrt{n} = \frac{Zs}{0.5}$$

$$n = \left(\frac{Zs}{0.5}\right)^2 = \left(\frac{1.96\,(3.05)}{0.5}\right)^2 = 142.95$$

i.e., we need 143 bottles to achieve this.

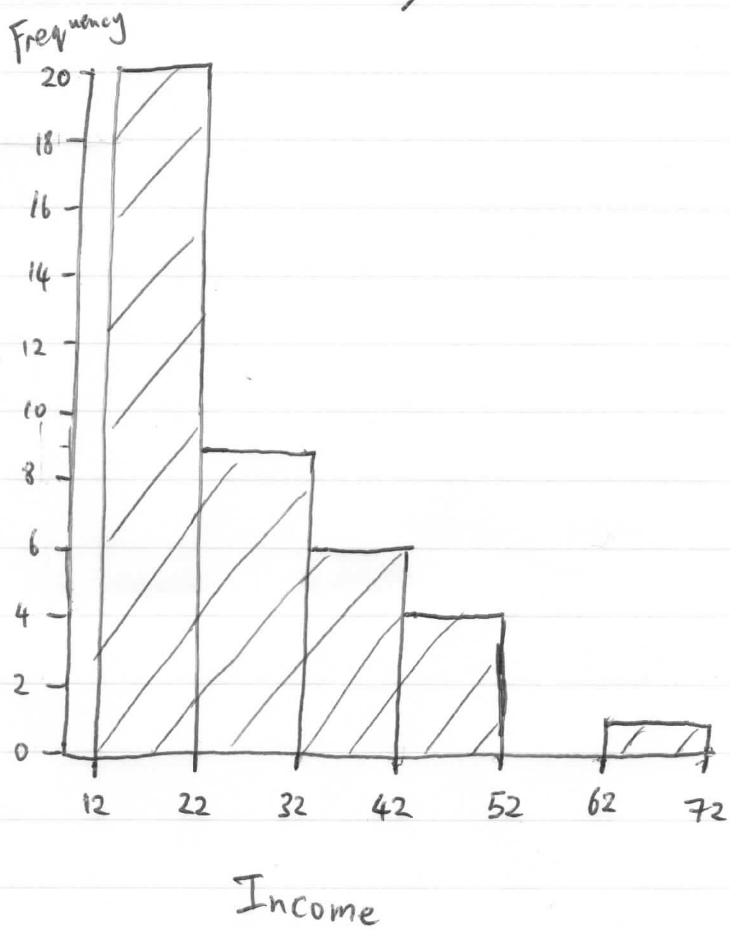
Q.2 (a)

Classes = 6 $\Rightarrow$ width = $\frac{max - min}{classes} = \frac{71-14}{6} = 9.5$

$\Rightarrow$ 10

(i)
(ii)

Class	Frequency
12 - 21.9	20
22 - 31.9	9
32 - 41.9	6
42 - 51.9	4
52 - 61.9	0
62 - 71.9	1
	40



(iii) Data skewed $\Rightarrow$ median.

Position of $Q_2 = 2\left(\frac{n+1}{4}\right) = \frac{n+1}{2} = \frac{41}{2} = 20.5 \Rightarrow 20 \ \& \ 21$

$\Rightarrow Q_2 = \frac{21 + 22}{2} = 21.5$

b) (i) $H_0: \sigma_1^2 = \sigma_2^2$
$H_a: \sigma_1^2 \neq \sigma_2^2$.

p-value = 0.7168 ⇒ no evidence to reject H_0,
ie., we may assume equal variances.

(ii) • $H_0: \mu_1 - \mu_2 = 0$
$H_a: \mu_1 - \mu_2 \neq 0$.

As we are assuming equal variances:

$$s_p^2 = \frac{(n_1 - 1)\, s_1^2 + (n_2 - 1)\, s_2^2}{n_1 + n_2 - 2}$$

$$= \frac{(8-1)\, 0.05 + (12-1)\, 0.04}{8 + 12 - 2}$$

$$= \frac{0.79}{18} = 0.04389$$

• Test statistic: $t = \dfrac{(\bar{x}_1 - \bar{x}_2) - (\mu_{01} - \mu_{02})}{\sqrt{\dfrac{s_p^2}{n_1} + \dfrac{s_p^2}{n_2}}}$

$$= \frac{(1.211 - 0.87) - 0}{\sqrt{\dfrac{0.04389}{8} + \dfrac{0.04389}{12}}}$$

$$= \frac{0.341}{0.096} = 3.566$$

- Two-tailed test $\Rightarrow \frac{\alpha}{2} = \frac{0.01}{2} = 0.005$

  Equal variances $\Rightarrow V = n_1 + n_2 - 2 = 8 + 12 - 2 = 18$.

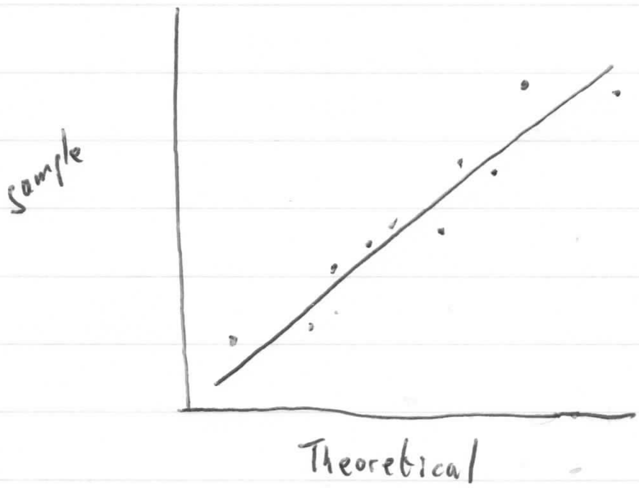
  $\Rightarrow$ Critical values: $\pm t_{18, 0.005} = \pm 2.878$

  Test statistic $t = 3.566$ is outside of ± 2.878.

- Conclusion: Reject the null hypothesis that the means are equal. (at the 1% level).

  $\Rightarrow$ Design 1 is a faster CPU.

(iii) The Q-Q plot which compares theoretical normal quantiles to the quantiles of the data sample.



If the points lie on the line (approximately) then we conclude that the data is normally distributed.

Q.3 (a)

						Σ
x	13	9	8	10	5	45
x^2	169	81	64	100	25	439

(i) $\bar{x} = \dfrac{\Sigma x}{n} = \dfrac{45}{5} = 9$ lines

(ii) $s^2 = \dfrac{\Sigma x^2 - n\bar{x}^2}{n-1} = \dfrac{439 - 5(9^2)}{5-1} = \dfrac{439-405}{4} = 8.5$ lines^2

(iii) $\Rightarrow s = \sqrt{8.5} = 2.915$ lines.

(iv) 90% C.I. $\Rightarrow \alpha/2 = \dfrac{0.1}{2} = 0.05$

Small sample $\Rightarrow$ t-value with $\nu = n-1 = 5-1 = 4$.

$\Rightarrow t_{4,0.05} = 2.132$.

C.I: $\bar{x} \pm t_{4,0.05} \dfrac{s}{\sqrt{n}}$

$9 \pm 2.132 \left(\dfrac{2.915}{\sqrt{5}}\right)$

$9 \pm 2.132(1.304)$

9 ± 2.78

$[6.22,\ 11.78]$

$\Rightarrow$ 90% confident that (on average) the task requires 6.22 – 11.78 lines of code.

(b) $P_r(A) = 0.4, \quad P_r(B) = 0.8, \quad P_r(A \cap B) = 0.3$

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

$= 0.4 + 0.8 - 0.3$

$= 0.9$

(ii) $P_r(B \mid A) = \dfrac{P_r(A \cap B)}{P_r(A)} = \dfrac{0.3}{0.4} = 0.75$

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

$= 1 - 0.3$

$= 0.7$

(c) (i) $P_r(X > 31 \mid C_1) = P_r\left(Z > \frac{31-30}{1}\right) = P_r(Z > 1) = 0.1587$

$= 0.16$

$P_r(X > 31 \mid C_2) = P_r\left(Z > \frac{31-29}{5}\right) = P_r(Z > 0.4) = 0.3446$

$= 0.34$

(ii) $P_r(C_1) = 0.8 \Rightarrow P_r(X > 31 \cap C_1) = P_r(C_1)\, P_r(X > 31 \mid C_1)$

$= 0.8(0.16)$

$= 0.128$

$P_r(C_2) = 0.2 \Rightarrow P_r(X > 31 \cap C_2) = P_r(C_2)\, P_r(X > 31 \mid C_2)$

$= 0.2(0.34)$

$= 0.068$

(iii) $Pr(X > 31) = Pr(X > 31 \cap C_1) + Pr(X > 31 \cap C_2)$

$= 0.128 + 0.068$

$= 0.196.$

(iv) $$Pr(C_1 \mid X < 31) = \frac{Pr(C_1 \cap X < 31)}{Pr(X < 31)}$$

$$= \frac{Pr(C_1)\, Pr(X < 31 \mid C_1)}{Pr(X < 31)}$$

$$= \frac{Pr(C_1)\,[1 - Pr(X > 31 \mid C_1)]}{1 - Pr(X > 31)}$$

$$= \frac{0.8(1 - 0.16)}{1 - 0.196}$$

$$= \frac{0.8(0.84)}{0.804} = 0.836.$$

Q. 4(a)

(i) $Pr(\text{RAID 1 fails}) = Pr(H_1^c \cap H_2^c)$

$= Pr(H_1^c)\, Pr(H_2^c)$

$= 0.8(0.8) = 0.64.$

(ii) We require $0.8^n = 0.01$

$n \log 0.8 = \log 0.01 \Rightarrow n = \frac{\log 0.01}{\log 0.8} = 20.63$

$\Rightarrow$ 21 hard disks required.

(b) $p = 0.1$

(i) $n = 10$

$Pr(X \geqslant 2) = 0.2639$ (using tables).

[Could also do $Pr(x \geqslant 2) = 1 - P(x < 2)$
$= 1 - [p(0) + p(1)]$
$= 1 - (0.3487 + 0.3874)$]

(ii) $n = 100$

$Pr(X < 15) = 1 - P(X \geqslant 15)$
$= 1 - 0.0726$ (using tables)
$= 0.9274$

(iii) $n = 20 \Rightarrow E[X] = np = 20(0.1) = 2$ cables
$sd(X) = \sqrt{np(1-p)} = \sqrt{20(0.1)(0.9)} = 1.34$ cables

(c) $\lambda = 4 \setminus hr.$

(i) $Pr(X = 4) = \frac{\lambda^x}{x!} e^{-\lambda} = \frac{4^4}{4!} e^{-4} = 0.1954$

[or $Pr(x = 4) = P(x \geqslant 4) - Pr(x \geqslant 5)$ using tables]
$= 0.5665 - 0.3712$

(ii) $\lambda = 4 \setminus hr \Rightarrow \lambda = 2 \setminus$ half-hr

$P(2 \leq X \leq 4) = \frac{2^2}{2!} e^{-2} + \frac{2^3}{3!} e^{-2} + \frac{2^4}{4!} e^{-2} = (1 + 1.333 + 0.666) e^{-2}$
$= 0.5413$

$$\left[\text{Or } P_r(2 \le X \le 4) = P_r(X \ge 2) - P_r(X \ge 5) \text{ using tables}\right]$$
$$= 0.5940 - 0.0527$$

(iii) $\lambda = 4 \setminus hr \Rightarrow \lambda = 20 \setminus 5\text{-hrs}.$

$$P_r(10 \le X \le 20) = P_r(X \ge 10) - P_r(X \ge 21)$$
$$= 0.9950 - 0.4409 \quad \text{(using tables)}$$
$$= 0.5541$$

(iv) 30 mins = 0.5 hrs.

Time $\Rightarrow T \sim \text{Exponential}(\lambda = 4)$

original $\lambda = 4 \setminus hr.$

$$P_r(T > 0.5) = e^{-\lambda t} = e^{-4(0.5)}$$
$$= e^{-2}$$
$$= 0.1353$$

Q.5 (a)

(i) $P_r(x=1) = 1 - [p(3) + p(5) + p(10)]$
$$= 1 - [0.2 + 0.2 + 0.1]$$
$$= 0.5$$

(ii) $E[X] = \sum x_i \, p(x_i) = 1(0.5) + 3(0.2) + 5(0.2) + 10(0.1)$
$$= 0.5 + 0.6 + 1.0 + 1.0$$
$$= 3.1$$

On average we expect to get a value of 3.1 from this distribution.

(iii) $E[X^2] = 1^2(0.5) + 3^2(0.2) + 5^2(0.2) + 10^2(0.1)$

$= 0.5 + 1.8 + 5 + 10$

$= 17.3$

$\Rightarrow Var[X] = E[X^2] - (E[X])^2$

$= 17.3 - (3.1)^2$

$= 7.69$

$\Rightarrow Sd(X) = \sqrt{Var(X)} = \sqrt{7.69} = 2.77.$

(b)

(i) Categorical

(ii) $H_0: p \leq 0.5$

$H_a: p > 0.5$

(iii) $\hat{p} = \frac{43}{65} = 0.662$

$$Z = \frac{\hat{p} - p_0}{\sqrt{\frac{p_0(1-p_0)}{n}}} = \frac{0.662 - 0.5}{\sqrt{\frac{0.5(0.5)}{65}}} = \frac{0.162}{0.062}$$

$= 2.61$

$\Rightarrow$ p-value $= P_r(Z > 2.61) = 0.00453$

(iv) P-value is small $\Rightarrow$ strong evidence <u>against</u> H_0

$\Rightarrow$ Reject H_0.

It appears that more people prefer the <u>new</u> recipe.

(c) $X \sim N(20, 3)$

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

$= P_r(Z < 1.67)$

$= 1 - P_r(Z > 1.67)$ (complement)

$= 1 - 0.0475$ (using tables)

$= 0.9525$

(ii) $P_r(X > x) = 0.1$

$P_r\left(Z > \frac{x-20}{3}\right) = 0.1$

From tables $P_r(Z > 1.28) = 0.1003 \approx 0.1$

$\Rightarrow \frac{x-20}{3} = 1.28$

$x - 20 = 1.28(3)$

$x = 20 + 1.28(3)$

$= 23.84$.

(iii) $\bar{X} \sim N(\mu, \sigma/\sqrt{n}) = N(20, 3/\sqrt{40})$

$= 0.474$

$\Rightarrow P_r(\bar{x} > 19.5) = P_r\left(Z > \frac{19.5 - 20}{0.474}\right)$

$= P_r(Z > -1.05)$

$= P_r(Z < 1.05)$ (symmetry)

$= 1 - P_r(Z > 1.05)$ (complement)

$= 1 - 0.1469$

$= 0.8531$

Q.6 (a)

(i) $E[T_s] = 0.025$ hrs $\Rightarrow \lambda_s = \frac{1}{0.025} = 40$ /hr.

(ii) $T \sim Exp(\lambda = \lambda_s - \lambda_a)$
$= 40 - 30$
$= 10.$

$\Rightarrow E[T] = \frac{1}{10}$ hrs $= 6$ mins.

$E[N] = \lambda_a E[T] = 30(\frac{1}{10}) = 3$ customers.

$E[T_q] = E[T] - E[T_s]$
$= 0.1 - 0.025$
$= 0.075$ hrs $= 4.5$ mins.

$E[N_q] = \lambda_a E[T_q] = 30(0.075) = 2.25$ customers

(iii) $\rho = \frac{\lambda_a}{\lambda_s} = \frac{30}{40} = 0.75 \Rightarrow 75\%.$

The service counter is busy 75% of the time.
(and idle 25% of the time).

(iv) 12 minutes $= \frac{12}{60} = 0.2$ hrs.

$\Rightarrow P_r(T < 0.2) = 1 - 0.1353$
$= 0.8647$

$P_r(T > 0.2) = e^{-\lambda t} = e^{-10(0.2)} = e^{-2} = 0.1353.$

(v) Exit rate is $\lambda_a = 30$ /hr. (same as arrival).
$\Rightarrow \lambda_a = 30(\frac{1}{6}) = 5$ / 10 mins i.e. $\frac{1}{6}$ of an hr.

$P_r(X > 7) = P_r(X \geq 8) = 0.1334.$

(b)

(i)

x	a	b	c	d
$p(x)$	0·5	0·25	0·2	0·05
$h(x)$	1·00	2·00	2·32	4·32

$h(x) = -\log_2 p(x)$

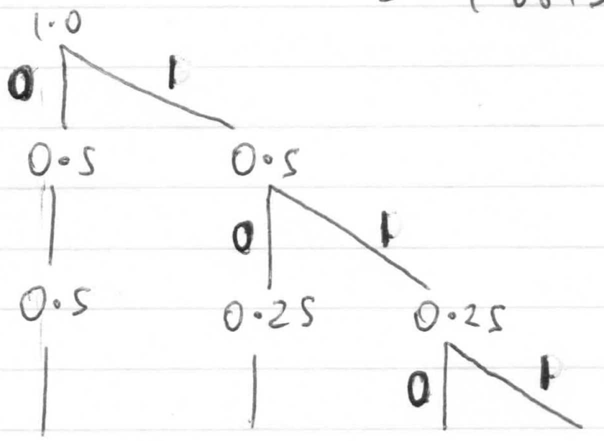
$$\Rightarrow H(X) = E(h(X)) = \sum h(x_i)\, p(x_i)$$

$$= 1{\cdot}00(0{\cdot}5) + 2{\cdot}00(0{\cdot}25) + 2{\cdot}32(0{\cdot}2) + 4{\cdot}32(0{\cdot}05)$$

$$= 0{\cdot}5 + 0{\cdot}5 + 0{\cdot}464 + 0{\cdot}2175$$

$$= 1{\cdot}6815 \text{ bits}$$

(ii)



$p(x)$	0·5	0·25	0·2	0·05
x	a	b	c	d
$c(x)$	0	10	110	111
$\ell(x)$	1	2	3	3

(iii) $E[L] = \sum \ell(x)\, p(x_i) = 1(0{\cdot}5) + 2(0{\cdot}25) + 3(0{\cdot}2) + 3(0{\cdot}05)$

$$= 0{\cdot}5 + 0{\cdot}5 + 0{\cdot}6 + 0{\cdot}15$$

$$= 1{\cdot}75 \text{ bits}$$

$$\Rightarrow e = \frac{H(X)}{E(L)} = \frac{1{\cdot}6815}{1{\cdot}75} = 0{\cdot}96 \Rightarrow 96\% \text{ efficient.}$$