Possible Solutions to 2020 FM Paper 2

1 Let S be the solution space of the following system of equations.

$$\begin{aligned} x_1 - x_2 + 2x_3 - x_4 + 5x_5 &= 0 \\ x_1 + 3x_2 - 2x_3 + x_4 - x_5 &= 0 \\ x_1 + x_3 + 3x_4 &= 0 \\ x_2 - x_3 + x_4 - 2x_5 &= 0 \end{aligned}$$

(i) Find a basis for S. [5]

(ii) State the dimension of S. [1]

[Solution]

1 The system of equation can be represented by

$$\begin{pmatrix} 1 & -1 & 2 & -1 & 5 \\ 1 & 3 & -2 & 1 & -1 \\ 1 & 0 & 1 & 1 & 0 \\ 0 & 1 & -1 & 1 & -2 \end{pmatrix}\begin{pmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \\ x_5 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ 0 \\ 0 \end{pmatrix}$$

It can be reduced to $\begin{pmatrix} 1 & 0 & 1 & 0 & 3 \\ 0 & 1 & -1 & 0 & -1 \\ 0 & 0 & 0 & 1 & -1 \\ 0 & 0 & 0 & 0 & 0 \end{pmatrix}\begin{pmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \\ x_5 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ 0 \\ 0 \end{pmatrix} \Rightarrow \begin{aligned} x_1 + x_3 + 3x_5 &= 0 \\ x_2 - x_3 - x_5 &= 0 \\ x_4 - x_5 &= 0 \end{aligned}$

The solution is $\begin{pmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \\ x_5 \end{pmatrix} = \begin{pmatrix} -x_3 - 3x_5 \\ x_3 + x_5 \\ x_3 \\ x_5 \\ x_5 \end{pmatrix} = x_3\begin{pmatrix} -1 \\ 1 \\ 1 \\ 0 \\ 0 \end{pmatrix} + x_5\begin{pmatrix} -3 \\ 1 \\ 0 \\ 1 \\ 1 \end{pmatrix}$

The two vectors $\begin{pmatrix} -1 \\ 1 \\ 1 \\ 0 \\ 0 \end{pmatrix}$ and $\begin{pmatrix} -3 \\ 1 \\ 0 \\ 1 \\ 1 \end{pmatrix}$ are linearly independent.

A basis is $\left\{ \begin{pmatrix} -1 \\ 1 \\ 1 \\ 0 \\ 0 \end{pmatrix}, \begin{pmatrix} -3 \\ 1 \\ 0 \\ 1 \\ 1 \end{pmatrix} \right\}$

(ii) dimension = 2

2 (i) By considering the 7th roots of unity, express $z^6 + z^5 + z^4 + z^3 + z^2 + z + 1$ as the product of three quadratic factors with real coefficients. Give each coefficient, where appropriate, in an exact trigonometric form. [3]

(ii) Deduce that

$$\cos\left(\tfrac{1}{7}\pi\right) - \cos\left(\tfrac{2}{7}\pi\right) + \cos\left(\tfrac{3}{7}\pi\right) = \tfrac{1}{2}$$

and determine the exact value of

$$\cos\left(\tfrac{1}{7}\pi\right)\cos\left(\tfrac{2}{7}\pi\right)\cos\left(\tfrac{3}{7}\pi\right).$$

Full working must be shown. [5]

[Solution]

(i) Considering $z^7 = 1$

$$z^7 = 1 = e^{i2k\pi} \text{ where } k = 0, \pm1, \pm2 \text{ and } \pm3$$

$$\Rightarrow z = e^{i\frac{2k\pi}{7}}$$

$$z^6 + z^5 + z^4 + \ldots z + 1 = \frac{z^7 - 1}{z - 1},$$

$$z^7 - 1 = (z-1)\left(z - e^{i\frac{2\pi}{7}}\right)\left(z - e^{-i\frac{2\pi}{7}}\right)\left(z - e^{i\frac{4\pi}{7}}\right)\left(z - e^{-i\frac{4\pi}{7}}\right)\left(z - e^{i\frac{6\pi}{7}}\right)\left(z - e^{-i\frac{6\pi}{7}}\right)$$

$$= (z-1)\left(z^2 - \left(e^{i\frac{2\pi}{7}} + e^{-i\frac{2\pi}{7}}\right)z + 1\right)\left(z^2 - \left(e^{i\frac{4\pi}{7}} + e^{-i\frac{4\pi}{7}}\right)z + 1\right)\left(z^2 - \left(e^{i\frac{6\pi}{7}} + e^{-i\frac{6\pi}{7}}\right)z + 1\right)$$

$$= (z-1)\left(z^2 - 2z\cos\frac{2\pi}{7} + 1\right)\left(z^2 - 2z\cos\frac{4\pi}{7} + 1\right)\left(z^2 - 2z\cos\frac{6\pi}{7} + 1\right)$$

$$z^6 + z^5 + z^4 + \ldots z + 1 = \frac{z^7 - 1}{z - 1}$$

$$= (z^2 - 2z\cos\frac{2\pi}{7} + 1)(z^2 - 2z\cos\frac{4\pi}{7} + 1)(z^2 - 2z\cos\frac{6\pi}{7} + 1)$$

(i) [Method 1]

Comparing the coefficient of the z term:

$$1 = -2\cos\frac{2\pi}{7} - 2\cos\frac{4\pi}{7} - 2\cos\frac{6\pi}{7}$$

$$\cos\frac{2\pi}{7} + \cos\frac{4\pi}{7} + \cos\frac{6\pi}{7} = -\frac{1}{2}$$

$$\Rightarrow \cos\frac{2\pi}{7} - \cos\frac{3\pi}{7} - \cos\frac{\pi}{7} = -\frac{1}{2} \quad \text{using } \cos(\pi - \theta) = -\cos\theta.$$

$$\Rightarrow \cos\frac{\pi}{7} - \cos\frac{2\pi}{7} + \cos\frac{3\pi}{7} = \frac{1}{2}$$

[Method 2]

The sum of roots of the equation $z^6 + z^5 + z^4 + \ldots z + 1 = 0$ is

$$-1 = e^{i\frac{2\pi}{7}} + e^{-i\frac{2\pi}{7}} + e^{i\frac{4\pi}{7}} + e^{-i\frac{4\pi}{7}} + e^{i\frac{6\pi}{7}} + e^{-i\frac{6\pi}{7}}$$

$$\Rightarrow \ -1 = 2\cos\frac{2\pi}{7} + 2\cos\frac{4\pi}{7} + 2\cos\frac{6\pi}{7}$$

$$\Rightarrow \cos\frac{\pi}{7} - \cos\frac{2\pi}{7} + \cos\frac{3\pi}{7} = \frac{1}{2}$$

(ii) [Method 1]: Let z = i in

$$z^6 + z^5 + z^4 + ...z + 1 = \frac{z^7 - 1}{z - 1} = (z^2 - 2z\cos\frac{2\pi}{7} + 1)(z^2 - 2z\cos\frac{4\pi}{7} + 1)(z^2 - 2z\cos\frac{6\pi}{7} + 1)$$

$$i^7 - 1 = (i - 1)(i^2 - 2i\cos\frac{2\pi}{7} + 1)(i^2 - 2i\cos\frac{4\pi}{7} + 1)(i^2 - 2i\cos\frac{6\pi}{7} + 1)$$

$$-i - 1 = (i - 1)\ (-2i\cos\frac{2\pi}{7})(-2i\cos\frac{4\pi}{7})(-2i\cos\frac{6\pi}{7})$$

$$-i - 1 = 8i\ (i - 1)\ \cos\frac{2\pi}{7}\cos\frac{4\pi}{7}\cos\frac{6\pi}{7}$$

$$\Rightarrow \cos\frac{2\pi}{7}\cos\frac{4\pi}{7}\cos\frac{6\pi}{7} = \frac{1}{8}$$

$\Rightarrow \cos\frac{2\pi}{7}\cos\frac{3\pi}{7}\cos\frac{\pi}{7} = \frac{1}{8}$ using $\cos(\pi - \theta) = -\cos\theta$,

[Method 2]: Comparing the coefficient of z^3 term in

$$z^6 + z^5 + z^4 + ...z + 1 = (z^2 - 2z\cos\frac{2\pi}{7} + 1)(z^2 - 2z\cos\frac{4\pi}{7} + 1)(z^2 - 2z\cos\frac{6\pi}{7} + 1)$$

$$1 = 2(-2\cos\frac{2\pi}{7} - 2\cos\frac{4\pi}{7} - 2\cos\frac{6\pi}{7}) - 8\cos\frac{2\pi}{7}\cos\frac{4\pi}{7}\cos\frac{6\pi}{7}$$

$$\Rightarrow 8\cos\frac{2\pi}{7}\cos\frac{4\pi}{7}\cos\frac{6\pi}{7} = -4(\cos\frac{2\pi}{7} + \cos\frac{4\pi}{7} + \cos\frac{6\pi}{7}) - 1$$

$$\Rightarrow 8\cos\frac{\pi}{7}\cos\frac{2\pi}{7}\cos\frac{3\pi}{7} = -4(-\frac{1}{2}) - 1 = 1$$

Thus $\cos\frac{\pi}{7}\cos\frac{2\pi}{7}\cos\frac{3\pi}{7} = \frac{1}{8}$

3 The plane transformation T, from $\mathbb{R}^2 \to \mathbb{R}^2$, is defined by

$$\mathrm{T}: \begin{pmatrix} x \\ y \end{pmatrix} \mapsto \mathbf{A}\begin{pmatrix} x \\ y \end{pmatrix}, \text{ where } \mathbf{A} = \begin{pmatrix} -1 & 7 \\ 2 & -6 \end{pmatrix}.$$

(i) Find

(a) the image of the point with coordinates (6, 1), [1]

(b) the coordinates of the point whose image point has coordinates (5, 14), [2]

(c) the image of the line $y = 1$. [3]

(ii) The unit square $OABC$, where $O = (0, 0)$, $A = (1, 0)$, $B = (1, 1)$ and $C = (0, 1)$, is transformed under T to the parallelogram $OA'B'C'$. Determine the area of $OA'B'C'$. [3]

[Solution]

(i)(a) $\begin{pmatrix} -1 & 7 \\ 2 & -6 \end{pmatrix}\begin{pmatrix} 6 \\ 1 \end{pmatrix} = \begin{pmatrix} -6+7 \\ 12-6 \end{pmatrix} = \begin{pmatrix} 1 \\ 6 \end{pmatrix}$

(b) $\begin{pmatrix} -1 & 7 \\ 2 & -6 \end{pmatrix}\begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} 5 \\ 14 \end{pmatrix}$

$\begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} -1 & 7 \\ 2 & -6 \end{pmatrix}^{-1}\begin{pmatrix} 5 \\ 14 \end{pmatrix} = \begin{pmatrix} 16 \\ 3 \end{pmatrix}$ using a GC

Coordinates of the point is (16, 3)

(c) $\begin{pmatrix} -1 & 7 \\ 2 & -6 \end{pmatrix}\begin{pmatrix} x \\ 1 \end{pmatrix} = \begin{pmatrix} 7-x \\ 2x-6 \end{pmatrix}$

[A 3-mark question. So examiner expects you to say more about this image]

$\begin{pmatrix} 7-x \\ 2x-6 \end{pmatrix} = \begin{pmatrix} 7 \\ -6 \end{pmatrix} + x\begin{pmatrix} -1 \\ 2 \end{pmatrix}, x \in \mathbb{R}$

The image are position vectors of points that lie on a line that passes through the point (7, –6) and is parallel to the vector $\begin{pmatrix} -1 \\ 2 \end{pmatrix}$ or points on the line $y = -2x + 8$.

(ii) For A' : $\begin{pmatrix} -1 & 7 \\ 2 & -6 \end{pmatrix}\begin{pmatrix} 1 \\ 0 \end{pmatrix} = \begin{pmatrix} -1 \\ 2 \end{pmatrix}$

B': $\begin{pmatrix} -1 & 7 \\ 2 & -6 \end{pmatrix}\begin{pmatrix} 1 \\ 1 \end{pmatrix} = \begin{pmatrix} 6 \\ -4 \end{pmatrix}$ and C' : $\begin{pmatrix} -1 & 7 \\ 2 & -6 \end{pmatrix}\begin{pmatrix} 0 \\ 1 \end{pmatrix} = \begin{pmatrix} 7 \\ -6 \end{pmatrix}$

Area of the parallelogram = $\left|\overrightarrow{OA'} \times \overrightarrow{OC'}\right|$

$= \left|\begin{pmatrix} -1 \\ 2 \\ 0 \end{pmatrix} \times \begin{pmatrix} 7 \\ -6 \\ 0 \end{pmatrix}\right| = \left|\begin{pmatrix} 0 \\ 0 \\ 8 \end{pmatrix}\right| = 8 \text{ units}^2$

4 The normal to the rectangular hyperbola $xy = c^2$ at the point $P\left(cp, \frac{c}{p}\right)$, $p > 0$, meets the curve again at the point Q.

(i) Determine the coordinates of Q. [7]

(ii) Prove that $(PQ)^2 = 3(OP)^2 + (OQ)^2$. [5]

[Solution]

(i) $xy = c^2 \Rightarrow x\frac{dy}{dx} + y = 0$

At $P(cp, \frac{c}{p})$, $\frac{dy}{dx} = \frac{-\frac{c}{p}}{cp} = -\frac{1}{p^2}$

Equation of the normal at P is $\frac{y - \frac{c}{p}}{x - cp} = p^2 \Rightarrow y - \frac{c}{p} = p^2(x - cp)$

When the normal meets the curve again: $\frac{c^2}{x} - \frac{c}{p} = p^2(x - cp) \Rightarrow pc^2 - cx = p^3(x^2 - xcp)$

$p^3x^2 + x(c - cp^4) - pc^2 = 0$

$(x - cp)(xp^3 + c) = 0 \Rightarrow x = cp$ (point P) or $x = -\frac{c}{p^3}$ (point Q)

Coordinates of Q are $(-\frac{c}{p^3}, -cp^3)$

(ii) $P(cp, \frac{c}{p})$,

$PQ^2 = (cp + \frac{c}{p^3})^2 + (\frac{c}{p} + cp^3)^2$

$= (cp)^2 + (\frac{c}{p^3})^2 + \frac{2c^2}{p^2} + (\frac{c}{p})^2 + (cp^3)^2 + 2c^2p^2$

$= (cp)^2 + (\frac{c}{p})^2 + (\frac{c}{p^3})^2 + (cp^3)^2 + 2c^2p^2 + \frac{2c^2}{p^2}$

$= 3[(cp)^2 + (\frac{c}{p})^2] + (\frac{c}{p^3})^2 + (cp^3)^2$

$= 3OP^2 + OQ^2$

5 The curve C is defined parametrically by $x = 4t^3 + 9$, $y = 24t\sqrt{t}$ for $0 \le t \le a$, where a is a positive constant.

When C is rotated about the x-axis, the surface area generated is $X(a)$.

When C is rotated about the y-axis, the surface area generated is $Y(a)$.

(i) Show that $X(a) = k\pi\left(\left(a^3+9\right)^{1.5} - 27\right)$ for some integer k to be determined. [7]

(ii) Find an expression, in terms of a, for $Y(a)$. [5]

(iii) Find, correct to 4 decimal places, for all values of a for which $X(a) = Y(a)$ and justify that there are no others. [3]

(i)	$\frac{dx}{dt} = 12t^2$, $\frac{dy}{dt} = 36t^{\frac{1}{2}}$ $\left(\frac{dx}{dt}\right)^2 + \left(\frac{dy}{dt}\right)^2 = 12^2t^4 + 36^2t$ $= 12^2t\left(t^3+9\right)$ $X(a) = 2\pi\int_0^a y\sqrt{\left(\frac{dx}{dt}\right)^2 + \left(\frac{dy}{dt}\right)^2}\,dt$ $= 2\pi\int_0^a 24t^{\frac{3}{2}}\sqrt{12^2t\left(t^3+9\right)}\,dt$ $= 576\pi\int_0^a t^2\sqrt{\left(t^3+9\right)}\,dt$ $= 192\pi\int_0^a 3t^2\sqrt{\left(t^3+9\right)}\,dt$ $= 192\pi\left[\frac{\left(t^3+9\right)^{\frac{3}{2}}}{\frac{3}{2}}\right]_0^a$ $= 128\pi\left[\left(a^3+9\right)^{1.5} - 27\right]$, where $k = 128$
(ii)	$Y(a) = 2\pi\int_0^a x\sqrt{\left(\frac{dx}{dt}\right)^2 + \left(\frac{dy}{dt}\right)^2}\,dt$ $= 2\pi\int_0^a \left(4t^3+9\right)\sqrt{12^2t\left(t^3+9\right)}\,dt$ $= 24\pi\int_0^a \left(4t^3+9\right)\left(t^4+9t\right)^{\frac{1}{2}}\,dt$ $= 24\pi\left[\frac{\left(t^4+9t\right)^{\frac{3}{2}}}{\frac{3}{2}}\right]_0^a$ $= 16\pi\left(a^4+9a\right)^{\frac{3}{2}}$
(iii)	When $X(a) = Y(a)$,

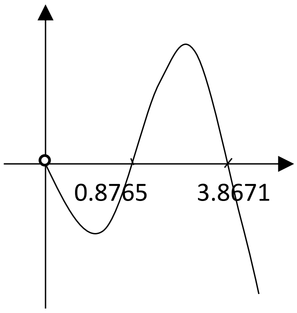
$$128\pi\left[\left(a^3+9\right)^{1.5}-27\right]=16\pi\left(a^4+9a\right)^{1.5}$$

$$8\left(a^3+9\right)^{1.5}-a^{1.5}\left(a^3+9\right)^{1.5}-216=0$$

$$\left(8-a^{1.5}\right)\left(a^3+9\right)^{1.5}-216=0$$

Solving (refer to graph of $y=\left(8-a^{1.5}\right)\left(a^3+9\right)^{1.5}-216$):

$a=0\left(\text{rej. } \because a>0\right)$ or a = 0.8765 or a = 3.8671.



If $a>3.8671$, then $8-a^{1.5}<0$ and $\left(a^3+9\right)^{1.5}>0$

$$\left(8-a^{1.5}\right)\left(a^3+9\right)^{1.5}-216<0.$$

$\therefore$ for all $a > 0$, a = 0.8765 and a = 3.8671 are the only two solutions of $X\left(a\right)=Y\left(a\right)$.

Section B: Probability and Statistics [50 marks]

6 Cheng sells insurance policies by phone. He knows from past experience that 17% of the calls he makes will be successful in selling a policy.

Cheng denotes by X the number of calls he makes each day up to and including his first successful call.

(i) State two conditions under which X can be well modelled using a geometric distribution. [2]

You are now given that X follows a geometric distribution.

(ii) Find the probability that the fifth call Cheng makes one day is the first to be successful. [1]

(iii) Find how many calls Cheng must make in order to be 95% certain that at least one will be successful. [3]

(iv) Write down the mean, a, and standard deviation, b, of X.

Hence find $\mathrm{P}\left(X>a+b\right)$. [3]

(i) Whether a call is successful or not is independent of any other call.
The probability of a call being successful is a fixed constant 0.17 for every call.

(ii) $X\sim\mathrm{Geo}\left(0.17\right)$

$\mathrm{P}\left(X=5\right)=0.080679\approx0.0807$

(iii) Let Y = Number of successful calls out of n calls.

$Y\sim\mathrm{B}\left(n,0.17\right)$

$\mathrm{P}\left(Y\geq1\right)\geq0.95$

$\mathrm{P}\left(Y=0\right)\leq0.05$

n	$\mathrm{P}\left(Y=0\right)$

16	0.05073
17	0.0421

There must be 17 calls to be 95% certain that at least one will be successful.

Or $\mathrm{P}(Y=0) \le 0.05$

$(0.83)^n \le 0.05$

$n \ge \dfrac{\lg 0.05}{\lg(0.83)}$

$n \ge 16.08$

$\therefore$ Least n = 17

(iv) $a = \mathrm{E}(X) = \dfrac{1}{0.17} = \dfrac{100}{17}$

$\mathrm{Var}(X) = \dfrac{1-0.17}{0.17^2} = \dfrac{8300}{289}$

Standard deviation of X, $b = \sqrt{\dfrac{8300}{289}} \approx 5.36$

$$\begin{aligned} \mathrm{P}(X > a+b) &= \mathrm{P}(X > 11.2414) \\ &= \mathrm{P}(X > 11) \\ &= (1-0.17)^{11} \\ &\approx 0.129 \end{aligned}$$

Or

$$\begin{aligned} \mathrm{P}(X > a+b) &= \mathrm{P}(X > 11.2414) \\ &= 1 - \mathrm{P}(X \le 11) \quad \text{(using GC)} \\ &\approx 0.129 \end{aligned}$$

7 An internet service provider (ISP) monitors the download speeds available to its domestic customers. Any occasion on which the speed drops below 5 megabits per second (Mbps) for a minute or more is referred to as an 'outage'. The ISP models the times between outages by the random variable T with cumulative distribution function

$$\mathrm{F}(t) = 1 - \mathrm{e}^{-kt} \text{ , for suitable } k \text{ and } t \geq 0 .$$

In this distribution, T is measured in days.

(i) Obtain the probability density function for T, name the distribution and state its mean and variance. [3]

(ii) Show that $P(2 \leq T < 3) = \mathrm{e}^{-2k} - \mathrm{e}^{-3k}$. [1]

The ISP claims in its technical literature that the times between outages follow the given distribution with $k = 0.5$.

(iii) What does this claim imply about the mean time between outages? [1]

A group of customers wish to test the accuracy of the claim. They obtain a random sample of 100 time intervals between outages. The data are summarised as follows.

Time between outages (days)	$0 \leq t < 1$	$1 \leq t < 2$	$2 \leq t < 3$	$3 \leq t < 4$	$4 \leq t < 5$	$5 \leq t$
Frequency	56	19	12	6	3	4

(iv) Carry out a chi-squared goodness-of-fit test for the ISP's model.
Discuss what the test indicates about the ISP's claim. You should refer to the p-value for your test and to the two largest contributions to the test statistic. [9]

7 $\mathrm{F}(t) = 1 - \mathrm{e}^{-kt}$

(i) $\mathrm{f}(t) = \mathrm{F}'(t) = \begin{cases} k\mathrm{e}^{-kt}, & t \geq 0 \\ 0 & \text{otherwise} \end{cases}$

This is an exponential distribution with

$\mathrm{E}(T) = \dfrac{1}{k}$ and $\mathrm{Var}(T) = \dfrac{1}{k^2}$

(ii) $\mathrm{P}(2 \leq T < 3) = \mathrm{F}(3) - \mathrm{F}(2)$

$= 1 - \mathrm{e}^{-3k} - (1 - \mathrm{e}^{-2k})$

$= \mathrm{e}^{-2k} - \mathrm{e}^{-3k}$

(iii) $\mathrm{E}(T) = \dfrac{1}{0.5} = 2$

It takes on average, 2 days after an outage for the next outage to occur.

(iv)

$H_o : T \sim \text{Exp}(0.5)$

$H_1 : T$ does not follow $\text{Exp}(0.5)$

Under H_o, expected frequency $E_i = 100 \times P(t_i \le t < t_{i+1})$

Time between outages	$0 \le t < 1$	$1 \le t < 2$	$2 \le t < 3$	$3 \le t < 4$	$4 \le t < 5$	$5 \le t$
Observed Freq (O_i)	56	19	12	6	3	4
Expected Freq (E_i)	39.35	23.87	14.47	8.78	5.33	8.20
Contribution	7.048	0.9918	0.423	0.8799	1.01516	2.15769

Constraint : $\sum O_i = \sum E_i$ (Check!)

Degree of freedom = 6 – 1 = 5

Under H_o, $\chi^2_{\text{cal}} = \sum_{i=1}^{6} \frac{(O_i - E_i)^2}{E_i} = 12.516$ (using GC)

$p\text{-value} = 0.02836 < 0.05$

Reject H_0 at 5% level of significance

There is sufficient evidence at 5% level of significance to conclude that T does not follow the exponential distribution with $k = 0.5$.

The null hypothesis is rejected at the 5% significance level, but not at the 1% level. There is therefore some, but not very strong evidence that the distribution is not an appropriate model for the data.

The largest contributions to the test statistic are from the cell $0 \le t < 1$, where the observed frequencies are higher than expected, and from the last cell $5 \le t$, where they are lower than expected, suggesting that the fit of the model is not as good for extreme values of t, and that perhaps k is greater than 0.5, i.e the mean of the distribution is lower than 2.

For comment on contributions of cells: comment whether the observed frequencies are higher or lower than expected.

8 AutoLot is a company which sells used cars. The company has recently employed a new salesperson, Grace, and part of her job is to set the prices at which cars should be advertised. In order to investigate whether price that Grace sets is correct, AutoLot chooses a random sample of the cars that are currently for sale and have already been priced. Grace is asked to set a price for each of them. She is not told what the current price is.

The table belows shows the prices (in thousands of dollars) at which the cars are on sale and the prices that Grace sets.

Car	*A*	*B*	*C*	*D*	*E*	*F*	*G*	*H*	*I*
Current price	21.0	12.2	17.5	15.3	16.5	23.5	34.4	47.5	85.0
Grace's price	19.9	12.4	17.0	15.2	16.8	22.8	32.1	42.0	79.0
Current – Grace's price	1.1	-0.2	0.5	0.1	-0.3	0.7	2.3	5.5	6.0

(i) Explain why it might not be appropriate to carry out a test based on the *t*-distribution using these data. [1]

(ii) Carry out a Wilcoxon test, at the 5% level of significance using these data, and state the conclusions AutoLot should reach. [6]

One of AutoLot's managers comments that the same conclusion would be reached by carrying out a sign test.

(iii) Check whether the manager is correct. [3]

(iv) Discuss briefly the relative merits of the sign test and the Wilcoxon test. [2]

(i) To carry out a test based on the *t*-distribution, we need the **distribution of the difference** in the current price and Grace's price **to follow a normal distribution**. However, it is not known if the difference follows a normal distribution, it is not appropriate to carry out a test based on the *t*-distribution.

(ii) Let D = current price – Grace's price.

Let m_D be the population median of D.

$H_0: m_D = 0$

$H_1: m_D \neq 0$

Level of significance: 5%

Car	*A*	*B*	*C*	*D*	*E*	*F*	*G*	*H*	*I*
Current price	21.0	12.2	17.5	15.3	16.5	23.5	34.4	47.5	85.0
Grace's price	19.9	12.4	17.0	15.2	16.8	22.8	32.1	42.0	79.0
Current – Grace's price	1.1	-0.2	0.5	0.1	-0.3	0.7	2.3	5.5	6.0
Rank	6	2	4	1	3	5	7	8	9

$P =$ sum of positive ranks = 40

$Q =$ sum of negative ranks = 5

$T = \min(P, Q) = 5$

From MF26, for $n = 9$, 2-tailed test at 5% significance level, critical region is $T \leq 5$.

Since $T = 5$ lies inside the critical region, we reject H_0 and conclude that there is sufficient evidence at 5% significance level that the price that Grace sets is not correct.

(iii) Let $D =$ current price – Grace's price.

Let m_D be the population median of D.

$H_0 : m_D = 0$

$H_1 : m_D \neq 0$

Level of significance: 5%

Car	*A*	*B*	*C*	*D*	*E*	*F*	*G*	*H*	*I*
Current price	21.0	12.2	17.5	15.3	16.5	23.5	34.4	47.5	85.0
Grace's price	19.9	12.4	17.0	15.2	16.8	22.8	32.1	42.0	79.0
Current – Grace's price	1.1	-0.2	0.5	0.1	-0.3	0.7	2.3	5.5	6.0

Test Statistic: $S = \text{Min}\{ S_+ , S_- \} \sim B(9, 0.5)$

$\boldsymbol{S_+}$ = no. of "+" signs = **7**, $\boldsymbol{S_-}$ = no. of "–" signs = 2

If H_0 is true, $S = \boldsymbol{S_-} \sim B(9, 0.5)$

Observed value of $S = \boldsymbol{S_-} = 2$, p-value = $2P(S \leq 2) = 0.180 > 0.05$

Hence we do not reject H_0 and conclude that there is insufficient evidence at 5% significance level that the price that Grace sets is not correct. Since the conclusion is different from the Wilcoxon test, the manager is not correct.

(iv) The sign test is fast and easy to carry out while the Wilcoxon test provides more information by considering the magnitudes of the differences.

9 An American newspaper commissioned a survey of doctors' salaries in order to investigate whether the mean salary is higher for males than for females. A random sample of 200 doctors were asked to state their annual salary and their gender. The results were as follows.

	Sample size	Mean ($)	Standard deviation ($)
Males	107	265 500	28 400
Females	93	254 000	26 700

(i) Carry out a suitable hypothesis test. State the *p*-value for the test and explain what it indicates. [5]

(ii) On examining the individual salaries for the two groups of doctors, it is clear that they are not normally distributed. Explain what implications, if any, this has for the test carried out in part **(i).** [2]

(i) Let X and Y be the amount of salary of male and female doctors respectively. Let μ_x and μ_y be the mean amount of salary of male and female doctors respectively.

$H_\circ : \mu_x = \mu_y$

$H_1 : \mu_x > \mu_y$

Level of significance: unknown

$n_x = 107, \bar{x} = 265500,$ sample s.d. $= 28400, s_x^2 = \frac{107}{106}\left(28400^2\right) = 814169056.6$

$n_y = 93, \bar{y} = 254000,$ sample s.d. $= 26700, s_y^2 = \frac{93}{92}\left(26700^2\right) = 720638804.3$

Under H_0, since both samples sizes are large, by Central Limit Theorem, $\bar{X}$ and $\bar{Y}$ are approximately normally distributed, hence

$\bar{X} - \bar{Y} \sim N\left(0, \frac{s_x^2}{107} + \frac{s_y^2}{93}\right)$ approx..

Test statistic of a 2-sample Z test :

$$\frac{\bar{X} - \bar{Y}}{\sqrt{\frac{28533.6478^2}{107} + \frac{26844.71651^2}{93}}} \sim \mathrm{N}(0,1) \text{ approx.}$$

Using GC, $z_{cal} = 2.934$

p-value = 0.00167 (i.e $\mathrm{P}\left(\bar{X} - \bar{Y} \geq 265500 - 254000\right)$)

p-value of 0.00167 indicates that H_0 will only be rejected if level of significance is at least 0.167% which is small. This means that there is very strong evidence that male doctors' salaries are higher than those of female doctors.

Note that question is not asking for explanation of p-value in the context of the question but rather the interpretation of the value obtained.

(ii) There is no implications to the test in **(i)** because both sample sizes are large, so by Central Limit Theorem, the sample mean salary of the two independent groups ($\overline{X}$ and $\overline{Y}$) each follow a normal distribution and hence $\overline{X}-\overline{Y}$ follows a normal distribution.

10 A survey carried out in 2006 in the UK asked 2112 adults about their views on evolution. 1014 of the adults said that they accepted the theory of evolution.

(i) Find a 95% confidence interval for p_{2006}, the proportion of UK adults in 2006 who accepted the theory of evolution. [3]

A scientific paper published in 2012 stated that a survey had shown that 69% of UK adults accepted the theory of evolution. The paper also stated that the 99% confidence interval was $69\% \pm 4.7\%$.

(ii) Determine the sample size in the 2012 paper. [2]

A researcher in 2018 wanted to investigate the proportion, p_{2018}, of UK adults in 2018 who accepted the theory of evolution. She wanted to produce a 95% confidence interval of width at most 6%.

(iii) Determine the smallest sample size that will give a confidence interval of the required width whatever the value of p_{2018}. [3]

(i) Sample proportion $p_s = \dfrac{1014}{2112}$

Since n = 2112 is large, $\dfrac{P_s - p_{2006}}{\sqrt{\dfrac{p_{2006}(1-p_{2006})}{n}}} \sim N(0,1)$ approximately. (optional)

95% confidence limits for $p_{2006} = p_s \pm 1.96\sqrt{\dfrac{p_s(1-p_s)}{n}}$

The 95% confidence interval is $(0.4588, 0.5014)$.

(ii) Let n be the required sample size.

In 2012, sample proportion = 0.69

and since it is a 99% confidence interval, z-value = 2.57583.

$$\therefore 2.57583\sqrt{\frac{0.69(1-0.69)}{n}} = 0.047$$

$$n \approx 642.46$$

Sample size is 642.

(iii) $2\times 1.96\sqrt{\dfrac{p_{2018}(1-p_{2018})}{n}} \le 0.06$

$n \ge 4268.44 p_{2018}(1-p_{2018})$

$n \ge 1067.11 - 4268.44(p_{2018}-0.5)^2$

In order for $n \ge 1067.11 - 4268.44(p_{2018}-0.5)^2$ for all values of p_{2018},

n has to be $\ge$ maximum value of $1067.11 - 4268.44(p_{2018}-0.5)^2$

maximum value of $1067.11 - 4268.44(p_{2018}-0.5)^2 = 1067.11$ when $p_{2018} = 0.5$

$\therefore n \ge 1067.11$

Least n = 1068