Please use this proforma at the beginning of your TMA to indicate how/if you have used generative AI.

I have used Generative AI in this TMA (such as Copilot, Gemini or ChatGPT) to help with the following:
[please tick all that apply]

☐ As a starting point or inspiration with a part of the TMA.

☐ To improve my own work, like the interpretation/summary of results.

☐ To summarise materials I found on the web for this TMA

☒ I did not use generative AI to help me with this TMA

Q 1.

(a)

(i) Let the random variable X take the value 1 if an O-ring seal fails and 0 otherwise. The appropriate distribution to describe the failure or non-failure of a particular O-ring seal on a particular flight is $X \sim \text{Bernoulli}(p)$.

2/2

(ii) Data were recorded for 23 Space Shuttle flights, each with 6 O-rings, giving a total of 138 O-rings. Table 1 shows that 5 flights experienced a single O-ring failure, 2 flights experienced 2 O-ring failures, giving a total of 9 failures. If we consider the proportion of failed O-rings to be an estimate for the probability of failure, this is given by

$$p(x) = \frac{\text{Number of failed O-rings}}{\text{Number of O-rings}}$$

2/2

This gives an estimate of $\frac{9}{138} = \frac{3}{46} \simeq 0.065$, as required.

(iii) The Binomial model assumes that the probability of a failure is the same for all O-rings on a flight and that whether or not one O-ring fails has no impact on the probability of any other O-ring failure (i.e. failure events are independent).

3/3

The binomial model does seem appropriate as the number of O-ring failures on a flight can be modeled as a series of 6 Bernoulli trials, assuming equal probability of failure and independence among the O-rings.

(iv)

x	pmf	cdf
0	0.668143	0.66814
1	0.278691	0.94683
2	0.048436	0.99527
3	0.004490	0.99976
4	0.000234	0.99999
5	0.000007	1.00000
6	0.000000	1.00000

3/3

Figure 1: Screenshot of p.m.f. and c.d.f. as calculated by Minitab.

(v)

3/3

Number of failed O-rings	Observed proportion	Probability
0	0.696	0.668
1	0.217	0.279
2	0.087	0.048
3	0.000	0.004
4	0.000	0.000
5	0.000	0.000
6	0.000	0.000

The observed proportions are close to the probabilities estimated by the model, suggesting $B(6, 0.065)$ is an appropriate model for the number of O-ring failures.

(b)

(i)

Let the number of readings classified as low in 10 selected at random be the random variable X, which can be modeled by $X \sim \text{B}\left(10, \frac{1}{3}\right)$. The Binomial p.m.f. is

1

$$p(x) = \binom{n}{x} p^x (1-p)^{n-x}, \qquad x = 0, 1, 2, ..., n$$

so the probability that at least 3 readings will be low is given by

$$P(X \geq 3) = \sum_{x=3}^{10} \binom{10}{x} p^x \left(1 - \tfrac{1}{3}\right)^{10-x}$$

or equivalently

Error Here

$$P(X \geq 3) = 1 - P(X \cancel{\geq} 8) = 1 - \sum_{x=8}^{10} \binom{10}{x} \left(\tfrac{2}{3}\right)^x \left(1 - \tfrac{2}{3}\right)^{10-x}$$

Your method is incorrect because the binomial distribution is only symmetric when p = 0.5

$$\boxed{P(X \geq 3) = 1 - P(X \leq 2) \quad \cdots = 0.7009 \text{ (4 d.p.)}}$$

$$= 1 - \left[\binom{10}{8}\left(\tfrac{2}{3}\right)^8\left(\tfrac{1}{3}\right)^2\right] + \left[\binom{10}{9}\left(\tfrac{2}{3}\right)^9\left(\tfrac{1}{3}\right)^1\right] + \left[\binom{10}{10}\left(\tfrac{2}{3}\right)^{10}\left(\tfrac{1}{3}\right)^0\right]$$

$$= 1 - 0.1950... + 0.0867... + 0.0173...$$

Follows correctly

$$= 0.7008...$$

1

2/4

So the probability that out of 10 readings selected at random, at least 3 will be classified as low, is 0.7008 (4 d.p.).

(ii)

Let the number of trials up to and including the first low reading be the random variable X, which can be modeled by $X \sim \mathrm{G}\left(\frac{1}{3}\right)$. The geometric p.m.f. is

$$p(x) = (1-p)^{x-1}p, \qquad x = 0, 1, 2, ..., n$$

Substituting $p = \frac{1}{3}$ gives

$$\begin{aligned} P(X=9) &= \left(1 - \frac{1}{3}\right)^8 \frac{1}{3} \\ &= \frac{2}{3}^8 \frac{1}{3} \\ &= 0.0130... \end{aligned}$$

2/2

So the probability that the first reading classified as low will be the ninth reading selected at random, is 0.0130 (4 d.p.).

(c) The random samples drawn from the two groups by Minitab are shown below.

GroupID = 2: 3.3, 2.4, 6.9, 7.1, 0.7, 3.3, 2.5, 4.1, 9.9, 1.8

GroupID = 40: 2.6, 8.9, 3.2, 7.1, 2.6, 1.7, 2.2, 6.1, 3.1, 0.7

0.7 and 1.8 are 'low', so X_2 = 2

1.7 and 0.7 are 'low', so X_3 = 2

(−1)

The model used in part (b)(i) suggests the probability of observing at least 3 readings in 10 selected at random, is $\simeq$0.7008, so these samples both contain fewer low values than expected. (−1)

Error Here See below

Running commentary TMA01 Q1b: Description, Similarities and Differences for Group 2 and Group 40.

As shown in Figure 1 (TMA01), the PM2.5 variable has a similar central location in Groups 2 and 40, with the mean and median slightly higher in Group 40. Group 40 has a slightly higher standard deviation and, as shown in Figure 2 (TMA01), interquartile range. Both groups have the same minimum value of 0.3, but Group 40's maximum value is more than twice the largest value in Group 2. Figure 2 (TMA01) suggests both data distributions exhibit positive skew, with Group 40's skew being slightly more pronounced.

2/5

The number of low PM2.5 values in a sample of 10 readings from Group 2 is fewer than expected and in a sample of 10 readings from Group 40 is fewer than expected.

From (b)(i) we know one third of PM2_5 records are 'low', so in a random sample of ten expect 10 x 1/3 = 3.33 'low' records. (−1)

TMA02 Q1b and c:
The number of low PM2.5 values in a sample of 10 readings from Group 2 is fewer than expected and the number of low PM2.5 values in a sample of 10 readingsfrom Group 40 is fewer than expected.

(d)

Let the number of flaws in a fibre optic cable be the random variable X, then probability that there are 2 or fewer flaws in a cable chosen at random is given by $P(X \leq 2) = p(0)+p(1)+p(2)$, where $p(x)$ is the value of the Poisson p.m.f. at x:

$$p(x) = \frac{e^{-\lambda}\lambda^x}{x!}, \qquad x = 0, 1, 2, ...$$

Substituting $\lambda = 1.25$ gives

$$\begin{aligned} P(X \leq 2) &= \left(\frac{e^{-1.25}1.25^0}{0!}\right) + \left(\frac{e^{-1.25}1.25^1}{1!}\right) + \left(\frac{e^{-1.25}1.25^2}{2!}\right) \\ &= 0.2865... + 0.3581... + 0.2238... \\ &= 0.8684... \end{aligned}$$

$\frac{3}{3}$

So the probability that there are 2 or fewer flaws is 0.868 (3 d.p.).

$\frac{22}{27}$

Q 2.

(a)

(i)

If X is a discrete random variable with p.m.f. $p(x)$, the population mean of X is denoted by μ and is given by

$$\mu = \sum_x xp(x)$$

Substituting the values of the p.m.f. in the range of X gives

$$\mu = (0 \times 0.3) + (1 \times 0.2) + (2 \times 0.2) + (3 \times 0.1) + (4 \times 0.1) + (5 \times 0.05) + (6 \times 0.05)$$
$$= 1.85$$

So the mean number of bicycles available at the docking station each morning is 1.85.

$\frac{2}{2}$

(ii)

If X is a discrete random variable with p.m.f. $p(x)$, the population variance of X is denoted by σ^2 and is given by

$$\sigma^2 = \sum_x (x-\mu)^2 p(x)$$

Substituting the values of the p.m.f. in the range of X gives

$$\sigma^2 = 0.3(-1.85)^2 + 0.2(-0.85)^2 + 0.2(0.15)^2 + 0.1(1.15)^2 + 0.1(2.15)^2 + 0.05(3.15)^2 + 0.05(4.15)^2$$
$$= 3.1275$$

So the variance of the number of bicycles available at the docking station each morning is 3.13 (2 d.p.).

$\frac{3}{3}$

(b)

(i)

2/2

The random variable X can me modeled as $X \sim \text{Bernoulli}(0.2)$. The population mean of a Bernoulli distribution is its parameter p. Therefore, the mean of X is 0.2.

(ii)

2/2

Let the number of years with rainfall be the random variable Y, which can be modeled as $Y \sim \text{B}(100, 0.2)$. The expected value (population mean) of a binomial distribution is the product of its parameters $= np$. Therefore, the expected number of years with some rainfall in a period of 100 years is $100 \times 0.2 = 20$.

(iii)

2/2

Let the number of years up to and including the first year in which there is some rainfall, be the random variable Z, which can be modeled as $Z \sim \text{G}(0.2)$. The expected value (population mean) of a geometric distribution is the reciprocal of its parameter $= 1/p$. Therefore, the expected number of years up to and including the first year in which there is some rainfall is $1/0.2 = 5$.

(c)

(i)

The population mean of a continuous random variable is given by

$$\mu = \int x f(x)\, dx$$

where the integral is taken over the range of x. Substituting the p.d.f. and range gives

$$\begin{aligned}
\mu &= \int_0^1 x\left(\frac{3}{2}\sqrt{x}\right)\, dx \\
&= \frac{3}{2}\int_0^1 x^{3/2}\, dx \\
&= \frac{3}{2}\left[\frac{2x^{5/2}}{5}\right]_0^1 \\
&= \frac{3}{5}
\end{aligned}$$

3/3

So the model predicts the mean length of sticks that the scout brings back is 0.6m.

(ii)

The population variance of a continuous random variable is given by

$$\sigma^2 = \int (x-\mu)^2 f(x)\ dx$$

where the integral is taken over the range of x. Substituting the p.d.f., range, and μ from the previous question gives

$$\begin{aligned}\sigma^2 &= \int_0^1 (x-0.6)^2 \left(\frac{3}{2}\sqrt{x}\right)\ dx \\ &= \int_0^1 \frac{3x^{5/2}}{2} - \frac{9x^{3/2}}{5} + \frac{27\sqrt{x}}{50}\ dx \\ &= \frac{3}{7} - \frac{18}{25} + \frac{9}{25} \\ &= \frac{12}{175}\end{aligned}$$

5/5

The standard deviation is the square root of the variance, and is therefore $\sigma = \frac{\sqrt{12}}{\sqrt{175}} = 0.262$m (3 s.f.).

The integration is easier if you use Equation (29), which leads to:

$$V(X) = E(X^2) - \mu^2 = \frac{3}{7} - \left(\frac{3}{5}\right)^2 = \frac{3}{7} - \frac{9}{25} = \frac{12}{175}$$

Do you agree?

(iii)

1/1

The mean and standard deviation are both in units of metres.

(d)

Y is a linear function of the random variable X. The mean of a linear function of a random variable $Y = aX + b$, is $E(Y) = aE(X) + b$, where $E(X)$ is the expected value (population mean) of X. Substituting $a = -3.5$, $b = 105$, and $E(X) = 24$ gives

$$\begin{aligned}\mu = E(Y) &= 105 - 3.5 \times 24 \\ &= 21\end{aligned}$$

3/3

The variance of a linear function of a random variable is $V(Y) = a^2V(X)$, where $V(X)$ is the population variance of X. Substituting $a = -3.5$ and $V(X) = 36$ gives

$$\begin{aligned}\sigma^2 = V(Y) &= (-3.5)^2 \times 36 \\ &= 441 \\ \sigma &= 21\end{aligned}$$

So the mean and standard deviation of the loss are both £21.

23/23 Very Good