Dr Oliver Mathematics
Advance Level Mathematics
Statistics: Calculator
2 hours

The total number of marks available is 50.
You must write down all the stages in your working. Note: It goes with the Mechanics paper.

1. Three bags, A, B, and C, each contain 1 red marble and some green marbles.

   Bag A contains 1 red marble and 9 green marbles only.
   Bag B contains 1 red marble and 4 green marbles only.
   Bag C contains 1 red marble and 2 green marbles only.

   Sasha selects at random one marble from A.
   If he selects a red marble, he stop selecting.
   If the marble is green, he continues by selecting at random one marble from B.
   If he selects a red marble, he stop selecting.
   If the marble is green, he continues by selecting at random one marble from C.

   (a) Draw a tree diagram to represent this information. (2)

   (b) Find the probability that Sasha selects 3 green marbles. (2)

   (c) Find the probability that Sasha selects at least 1 marble of each colour. (2)

   (d) Given that Sasha selects a red marble, find the probability that he selects from bag B. (2)

2. The partially completed box plot in Figure 1 shows the distribution of daily air temperatures using the data from the large data set for Beijing in 2015.

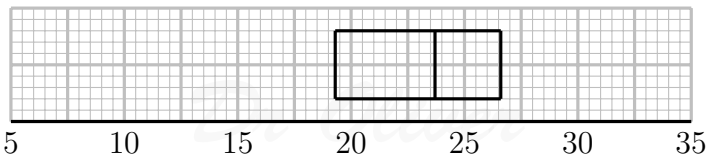

Figure 1: Temperature (°C)

   An outlier is defined as a value

   - more than 1.5 × IQR below Q_1, or
   - more than 1.5 × IQR above Q_3.

   The three lowest air temperature in the data set are 7.6°, 8.1°, and 9.1°.
   The highest air temperature in the data set is 32.5°

(a) Complete the box plot Figure 1, showing clearly any outliers. (4)

(b) Using your knowledge of the large data set, suggest from which two months outliers are likely to have come. (1)

Using the data the large data set, Simon produced the following summary statistics for the daily mean air temperatures, x°, for Beijing in 2015.

$$n = 184, \sum x = 4\,153.6, \; S_{xx} = 4\,952.906.$$

(c) Show that, to 3 significant figures, the standard deviation is 5.19°C. (1)

Simon decides to model the daily air temperatures with the random variable

$$T \sim \mathrm{N}(22.6, 5.19^2).$$

(d) Using Simon's model, calculate the 10th to 90th interpercentile range. (3)

Simon wants to model another variable from the large data set for Beijing using a normal distribution.

(e) Stats two variables from the large data set for Beijing that are **not** suitable to be modelled by a normal distribution. Give a reason for your answer. (2)

3. Barbara is investigating the relationship between average income (GDP per capita), x US dollars, and annual carbon dioxide (CO_2) emissions y tonnes, for different countries.

She takes a random sample of 24 countries and finds the product moment correlation coefficient between average annual CO_2 emissions and average income to be 0.446.

(a) Stating your hypotheses clearly, test, at the 5% level of significance, whether or not the product moment correlation coefficient for all countries is greater than zero. (3)

Barbara believes that a non-linear model would be a better fit to the data. She codes the data using the coding

$$m = \log_{10} x \text{ and } c = \log_{10} y$$

and obtains the model

$$c = -1.82 + 0.89m.$$

The product moment correlation coefficient between c and m is found to be 0.882.

(b) Explain how this value supports Barbara's belief. (1)

(c) Show that this relationship between y and x can be written in the form (5)

$$y = ax^n,$$

where a and n are constants to be found.

4. Magali is studying the mean total cloud cover, in oktas, for Leuchars in 1987 using data from the large data set. The daily mean total cloud cover for all 184 days from the large data set is summarised in the table below.

Daily mean total cloud cover (oktas)	0	1	2	3	4	5	6	7	8
Frequency (number of days)	0	1	4	7	10	30	52	52	28

One of the 184 days is selected at random.

(a) Find the probability that it has a daily mean total cloud cover of 6 or greater. (1)

Magali is investigating whether the daily mean total cloud cover can be modelled using a binomial distribution.

She uses the random variable X to denote the daily mean total cloud cover and believes that

$$X \sim \mathrm{B}(8, 0.76).$$

Using Magali's method,

(b) (i) Find $\mathrm{P}(X \geqslant 6)$, (2)

(ii) find, to 1 decimal place, the expected number of days in a sample of 184 days with a daily mean total cloud cover of 7. (2)

(c) Explain whether or not your answers to part (b) support the use of Magali's model. (1)

There were 28 days that had a daily mean total cloud cover of 8.
For these 28 days, the daily mean total cloud cover for the **following** days is shown in the table below.

Daily mean total cloud cover (oktas)	0	1	2	3	4	5	6	7	8
Frequency (number of days)	0	0	1	1	2	1	5	9	9

(d) Find the proportion of these days when the daily mean total cloud cover was 6 or greater. (1)

(e) Comment on Magali's model in the light of your answer to part (d). (2)

5. A machine puts liquid into bottles of perfume. The amount of liquid put into each bottle, D ml, follows a normal distribution with mean 25 ml.

Given that 15% of bottles contain less less than 24.63 ml,

(a) find, to 2 decimal places, the value of k such that (5)

$$\mathrm{P}(24.63 < D < k) = 0.45.$$

A random sample of 200 bottles is taken.

(b) Using a normal approximation, find the probability that fewer than half of these bottles contain between 24.63 ml and k ml. (3)

The machine is adjusted so that the standard deviation of the liquid put in the bottles is now 0.16 ml.

Following the adjustments, Hannah believes that the mean amount of liquid put in the bottle is less than 25 ml.

She takes a random sample of 20 bottles and finds the mean amount of liquid to be 24.94 ml.

(c) Test Hannah's belief at the 5% level of significance. (5)
You should state your hypotheses clearly.