Please check the examination details below before entering your candidate information

Candidate surname	Other names

Pearson Edexcel Level 3 GCE

Centre Number

Candidate Number

Friday 14 June 2019

Afternoon | Paper Reference **9MA0-31**

Mathematics

Advanced
Paper 31: Statistics

You must have:
Mathematical Formulae and Statistical Tables, calculator

Total Marks

Candidates may use any calculator allowed by Pearson regulations. Calculators must not have the facility for symbolic algebra manipulation, differentiation and integration, or have retrievable mathematical formulae stored in them.

Instructions

- Use **black** ink or ball-point pen.
- If pencil is used for diagrams/sketches/graphs it must be dark (HB or B).
- **Fill in the boxes** at the top of this page with your name, centre number and candidate number.
- Answer **all** questions and ensure that your answers to parts of questions are clearly labelled.
- Answer the questions in the spaces provided – *there may be more space than you need.*
- You should show sufficient working to make your methods clear. Answers without working may not gain full credit.
- Answers should be given to three significant figures unless otherwise stated.

Information

- A booklet 'Mathematical Formulae and Statistical Tables' is provided.
- The total mark for this part of the examination is 50. There are 5 questions.
- The marks for **each** question are shown in brackets – *use this as a guide as to how much time to spend on each question.*

Advice

- Read each question carefully before you start to answer it.
- Try to answer every question.
- Check your answers if you have time at the end.

Turn over ▶

Answer ALL questions. Write your answers in the spaces provided.

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

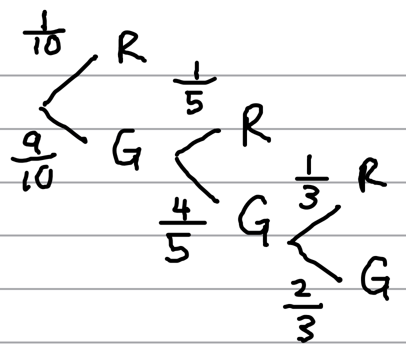
(a) Draw a tree diagram to represent this information.
(2)

(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 it from bag B.
(2)

a)



b) $\frac{9}{10} \times \frac{4}{5} \times \frac{2}{3} = \frac{12}{25}$

c) $\left(\frac{9}{10} \times \frac{1}{5}\right) + \left(\frac{9}{10} \times \frac{4}{5} \times \frac{1}{3}\right) = \frac{9}{50} + \frac{6}{25}$

$= \frac{21}{50}$

d) P (red from B | red selected) $= \dfrac{\frac{9}{10} \times \frac{1}{5}}{\frac{1}{10} + \frac{21}{50}}$

$= \frac{9}{26}$

2.

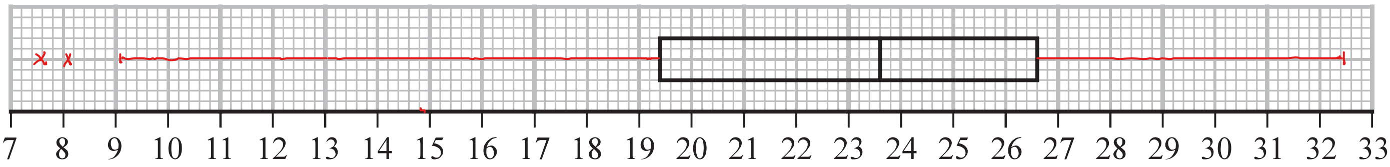


Figure 1

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

An outlier is defined as a value
more than $1.5 \times$ IQR below Q_1 or
more than $1.5 \times$ IQR above Q_3

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

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

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

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

$$n = 184 \qquad \sum x = 4153.6 \qquad S_{xx} = 4952.906$$

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

Simon decides to model the 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) State two variables from the large data set for Beijing that are **not** suitable to be modelled by a normal distribution. Give a reason for each answer.
(2)

a) IQR = 26.6 − 19.4
= 7.2
7.2(1.5) = 10.8

Question 2 continued

$19.4 - 10.8 = 8.6$

$26.6 + 10.8 = 37.4$

b) October, since it is the month with the coldest temperatures between May and October in Beijing.

c) $\sigma = \sqrt{\frac{S_{xx}}{n}}$

$= \sqrt{\frac{4952.906}{184}}$

$= 5.18825$

$\simeq 5.19$

d) $z \approx \pm 1.2816$

$2 \times 1.2816 \times 5.19 = 13.303$

$\simeq 13.3$

e) Rainfall since it is not symmetrical as there are a lot of days without rainfall. Daily mean windspeed given using the Beauford scale is qualitative so it is also not suitable.

Turn over for a spare grid if you need to redraw your box plot.

3. Barbara is investigating the relationship between average income (GDP per capita), x US dollars, and average 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 $m = \log_{10}x$ 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 the relationship between y and x can be written in the form $y = ax^n$ where a and n are constants to be found.

(5)

a) $H_0 : \rho = 0$

$H_1 : \rho > 0$ 5% S.L. , $n=24$

CV = 0.3438

$0.446 > 0.3438$

∴ H_0 is rejected. Evidence shows that pmcc for all countries is greater than 0.

b) 0.882 is closer to 1 so there is a stronger positive correlation.

c) $\log_{10} y = -1.82 + 0.89(\log_{10} x)$

$y = 10^{-1.82 + 0.89(\log_{10} x)}$

$= 10^{-1.82} \times 10^{\log_{10} x^{0.89}}$

$= 10^{-1.82} x^{0.89}$

$= 0.015 x^{0.89}$

$a = 0.015$, $n = 0.89$

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 model,

(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** day 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 light of your answer to part (d).

(2)

a) $P(X \geq 6) = \dfrac{52+52+28}{184}$

$= \dfrac{132}{184}$

$= \dfrac{33}{46}$

$= 0.717$

DO NOT WRITE IN THIS AREA

DO NOT WRITE IN THIS AREA

DO NOT WRITE IN THIS AREA

Question 4 continued

bi) $P(X \geq 6) = P(X=6) + P(X=7) + P(X=8)$

$= {}^8C_6 (0.76)^6 (1-0.76)^2 + {}^8C_7 (0.76)^7 (1-0.76) + (0.76)^8$

$= 0.703$

ii) $P(X=7) \times 184 = 0.28119 (184)$

$= 51.7$ days

c) The probabilities calculated for a) and bi) are similar, and the answer for bii) is similar to the frequency for daily mean total cloud of 7 from the table, so Magali's model is supported

d) $P(X \geq 6) = \frac{9+9+5}{28}$

$= \frac{23}{28}$

≈ 0.821

e) The probability calculated for part d) is different from a) and bi) so the model is not supported.

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 than 24.63 ml

(a) find, to 2 decimal places, the value of k such that $P(24.63 < D < k) = 0.45$
(5)

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 each 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.
You should state your hypotheses clearly.
(5)

a) $\frac{24.63-25}{\sigma} = -1.0364$

$\sigma = 0.357$

$P(D < k) = 0.15 + 0.45$

$= 0.6$

$P(D > k) = 0.4$

$\frac{k-25}{0.357} = 0.2533$

$k = 25.0904$

$\simeq 25.09$

b) $W \sim N(200 \times 0.45,\ 200 \times 0.45(1-0.45))$

$W \sim N(90, 49.5)$

$P(Y < 100) \approx P(W < 99.5)$

$= P\left(Z < \frac{99.5-90}{\sqrt{49.5}}\right)$

$= P(Z < 1.35)$

$= 0.9115$

$\simeq 0.912$

Question 5 continued

c) $H_0: \mu = 25$

$H_1: \mu < 25$ 5% s.l.

$\bar{D} \sim N\left(25, \frac{0.16^2}{20}\right)$

$\bar{D} \sim N(25, 0.00128)$

$P(\bar{D} < 24.94) = P\left(Z < \frac{24.94 - 25}{\sqrt{0.00128}}\right)$

$= P(Z < -1.677)$

$\approx 0.046766 < 0.05$

$\therefore$ Reject H_0. There is sufficient evidence to support Hannah's belief.