Please check the examination details below before entering your candidate information

Candidate surname	Other names

Pearson Edexcel Level 3 GCE

Centre Number	Candidate Number

Wednesday 22 May 2019

Morning	Paper Reference **8MA0-21**

Mathematics

Advanced Subsidiary
Paper 21: 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 30. 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 ▶

P63360A

1/1/1/1/1/

P 6 3 3 6 0 A 0 1 1 6

Pearson

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

1. A sixth form college has 84 students in Year 12 and 56 students in Year 13

The head teacher selects a stratified sample of 40 students, stratified by year group.

(a) Describe how this sample could be taken.

(3)

The head teacher is investigating the relationship between the amount of sleep, s hours, that each student had the night before they took an aptitude test and their performance in the test, p marks.
For the sample of 40 students, he finds the equation of the regression line of p on s to be

$$p = 26.1 + 5.60s$$

(b) With reference to this equation, describe the effect that an extra 0.5 hours of sleep may have, on average, on a student's performance in the aptitude test.

(1)

(c) Describe one limitation of this regression model.

(1)

a Number year 12 students 1-84 and year 13 students 1-56, then use a random number source e.g. calculator to choose a sciple random sample:

Yr 12 : Yr 13 = 84 : 56 = 3 : 2

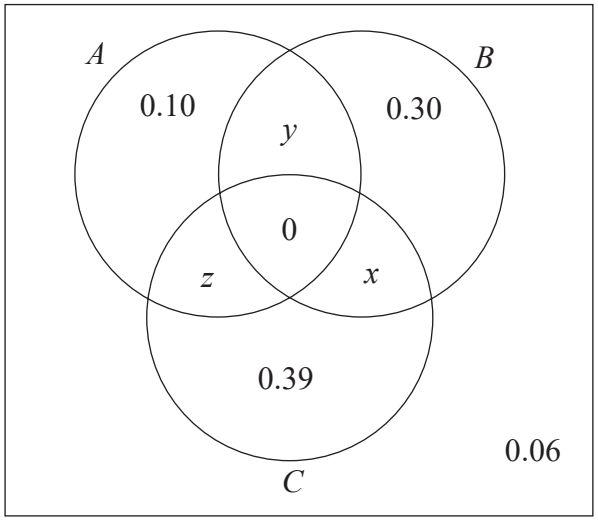
40 split into 3:2 ratio: $\frac{40}{5} = 8$, 24 : 16

Choose 24 year 12 students and 16 year 13.

b. $\Delta p = 5.6 \times 0.5 = 2.8 \Rightarrow$ 2.8 mark increase

c. The model has no defined range $\Rightarrow$ students who have infinite sleep (don't wake up) will perform best.

2. The Venn diagram shows three events, A, B and C, and their associated probabilities.



Events B and C are mutually exclusive.
Events A and C are independent.

Showing your working, find the value of x, the value of y and the value of z.

(5)

B and C mutually exclusive $\Rightarrow x = 0$

$P(A) = 0.1 + y + z$

$P(C) = 0.39 + z + (x) \rightarrow = 0$

$P(A \text{ and } C) = z$

A and C independant $\Rightarrow P(A \text{ and } C) = P(A) \times P(C)$

$\Rightarrow z = (0.1 + y + z)(0.39 + z)$

Σ probability $= 1$. $1 = 0.06 + 0.1 + 0.39 + 0.3 + y + z$

$y + z = 0.15$

$\Rightarrow z = (0.1 + 0.15)(0.39 + z)$

$0.75z = 0.0975$

$\downarrow$

Question 2 continued

$z = \frac{0.0975}{0.75} = 0.13$

$y = 0.15 - z = 0.15 - 0.13 = 0.02$

$x = 0$
$y = 0.02$
$z = 0.13$

(Total for Question 2 is 5 marks)

3. A fair 5-sided spinner has sides numbered 1, 2, 3, 4 and 5

The spinner is spun once and the score of the side it lands on is recorded.

(a) Write down the name of the distribution that can be used to model the score of the side it lands on.

(1)

The spinner is spun 28 times.

The random variable X represents the number of times the spinner lands on 2

(b) (i) Find the probability that the spinner lands on 2 at least 7 times.

(ii) Find $P(4 \leqslant X < 8)$

(5)

a. Discrete uniform distribution.

b. i. At least 7 times: $P(x \geq 7) = 1 - P(x \leq 6)$

$= 1 - 0.6784$

$= 0.322$ (3dp)

ii. $P(4 \leq x < 8) = P(x \leq 7) - P(x \leq 3)$

$= 0.818 - 0.160$

$= 0.658$ (3dp)

4. Joshua is investigating the daily total rainfall in Hurn for May to October 2015

Using the information from the large data set, Joshua wishes to calculate the mean of the daily total rainfall in Hurn for May to October 2015

(a) Using your knowledge of the large data set, explain why Joshua needs to clean the data before calculating the mean.

(1)

Using the information from the large data set, he produces the grouped frequency table below.

Daily total rainfall (r mm)	Frequency	Midpoint (x mm)
$0 \leqslant r < 0.5$	121	0.25
$0.5 \leqslant r < 1.0$	10	0.75
$1.0 \leqslant r < 5.0$	24	3.0
$5.0 \leqslant r < 10.0$	12	7.5
$10.0 \leqslant r < 30.0$	17	20.0

You may use $\sum fx = 539.75$ and $\sum fx^2 = 7704.1875$

(b) Use linear interpolation to calculate an estimate for the upper quartile of the daily total rainfall.

(2)

(c) Calculate an estimate for the standard deviation of the daily total rainfall in Hurn for May to October 2015

(2)

(d) (i) State the assumption involved with using class midpoints to calculate an estimate of a mean from a grouped frequency table.

(ii) Using your knowledge of the large data set, explain why this assumption does not hold in this case.

(iii) State, giving a reason, whether you would expect the actual mean daily total rainfall in Hurn for May to October 2015 to be larger than, smaller than or the same as an estimate based on the grouped frequency table.

(3)

a. Trace data must be converted to numbers to allow calculations to be carried out.

b. $1 + \frac{138 - 131}{24} \times 6 = 2.17$

Question 4 continued

c. $\sigma = \sqrt{\frac{\Sigma fx^2}{\Sigma f} - \left(\frac{\Sigma fx}{\Sigma f}\right)^2} = \sqrt{\frac{77061875}{184} - \left(\frac{539.75}{184}\right)^2}$

$= 5.77$

d. i. Using midpoints assumes the data is distributed uniformly throughout each class.

ii. Most of the data in the first class are 0.

iii. A true mean is likely to be smaller as the first group has more values at and close to 0 than the calculated mean takes into account.

5. Past records show that 15% of customers at a shop buy chocolate. The shopkeeper believes that moving the chocolate closer to the till will increase the proportion of customers buying chocolate.

After moving the chocolate closer to the till, a random sample of 30 customers is taken and 8 of them are found to have bought chocolate.

Julie carries out a hypothesis test, at the 5% level of significance, to test the shopkeeper's belief.

Julie's hypothesis test is shown below.

$H_0 : p = 0.15$

$H_1 : p \geqslant 0.15$

Let X = the number of customers who buy chocolate.

$X \sim B(30, 0.15)$

$P(X = 8) = 0.0420$

$0.0420 < 0.05$ so reject H_0

There is sufficient evidence to suggest that the proportion of customers buying chocolate has increased.

(a) Identify the first two errors that Julie has made in her hypothesis test.

(2)

(b) Explain whether or not these errors will affect the conclusion of her hypothesis test. Give a reason for your answer.

(1)

(c) Find, using a 5% level of significance, the critical region for a one-tailed test of the shopkeeper's belief. The probability in the tail should be less than 0.05

(2)

(d) Find the actual level of significance of this test.

(1)

a. H_1 should be: $p > 0.15$ rather than $p \geq 0.15$

Test statistic should be calculated by $P(x \geq 8)$, not $P(x = 8)$.

b. Yes, $P(x \geq 8) = 0.0698$, > 0.05 so this finding is in fact significant.

$1 - P(X \leq 7) = 1 - 0.9302$

Question 5 continued

c. $P(x \leq 8) = 0.9722 > 0.95$

$P(x \geq 9) = 0.0278 < 0.05$

$\Rightarrow$ critical region : $x \geq 9$

d. 0.0278, 2.78%