Please check the examination details below before entering your candidate information

Candidate surname	Other names

Centre Number | Candidate Number

Pearson Edexcel Level 3 GCE

Paper reference **8MA0/21**

Mathematics

Advanced Subsidiary
PAPER 21: Statistics

You must have:
Mathematical Formulae and Statistical Tables (Green), calculator

Total Marks

Candidates may use any calculator permitted 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.
- Values from statistical tables should be quoted in full. If a calculator is used instead of tables the value should be given to an equivalent degree of accuracy.
- Inexact 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 ▶

P69599A

Q:1/1/1/

P 6 9 5 9 9 A 0 1 1 2

Pearson

1. The relationship between two variables p and t is modelled by the regression line with equation

$$p = 22 - 1.1\,t$$

The model is based on observations of the independent variable, t, between 1 and 10

(a) Describe the correlation between p and t implied by this model.

(1)

Given that p is measured in centimetres and t is measured in days,

(b) state the units of the gradient of the regression line.

(1)

Using the model,

(c) calculate the change in p over a 3-day period.

(2)

Tisam uses this model to estimate the value of p when $t = 19$

(d) Comment, giving a reason, on the reliability of this estimate.

(1)

a) $p = 22 - 1.1t$ — gradient

∴ Negative since gradient of regression line is negative. (1)

b) $p\,(cm) = 22 - 1.1\left(\frac{cm}{day}\right) \times t\,(day)$

unit LHS = RHS

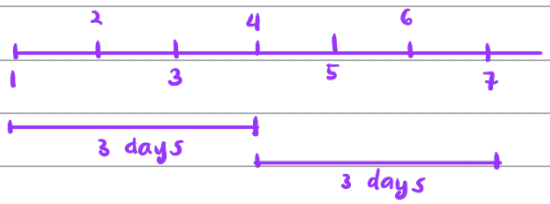
units are $\frac{cm}{day}$ (1)

c) Change in p over a 3-day period:

$-1.1 \times 3 = -3.3$ (1)

∴ decrease of 3.3 cm (1)

EXAMPLE OF 3 DAY PERIOD



d) 19 is outside the range [1, 10], so the result will be unreliable. (1)

Question 1 continued

(Total for Question 1 is 5 marks)

2. A manufacturer of sweets knows that 8% of the bags of sugar delivered from supplier *A* will be damp.
A random sample of 35 bags of sugar is taken from supplier *A*.

(a) Using a suitable model, find the probability that the number of bags of sugar that are damp is

(i) exactly 2

(ii) more than 3

(3)

Supplier *B* claims that when it supplies bags of sugar, the proportion of bags that are damp is less than 8%

The manufacturer takes a random sample of 70 bags of sugar from supplier *B* and finds that only 2 of the bags are damp.

(b) Carry out a suitable test to assess supplier *B*'s claim.
You should state your hypotheses clearly and use a 10% level of significance.

(4)

a) Let D = number of bags that are damp from A

$D \sim B(35, 0.08)$ (1)

(i) $P(D = 2) = 0.243$ (3 s.f.) (1)

$P(D > 3) = 1 - P(D \leq 3)$

$= 1 - 0.6939\ldots = 0.306$ (3 s.f.) (1)

b) Let X = number of bags that are damp from B

$X \sim B(70, 0.08)$ (1)

$H_0 : p = 0.08$, $H_1 : p < 0.08$ (1)

$\alpha = 0.1$

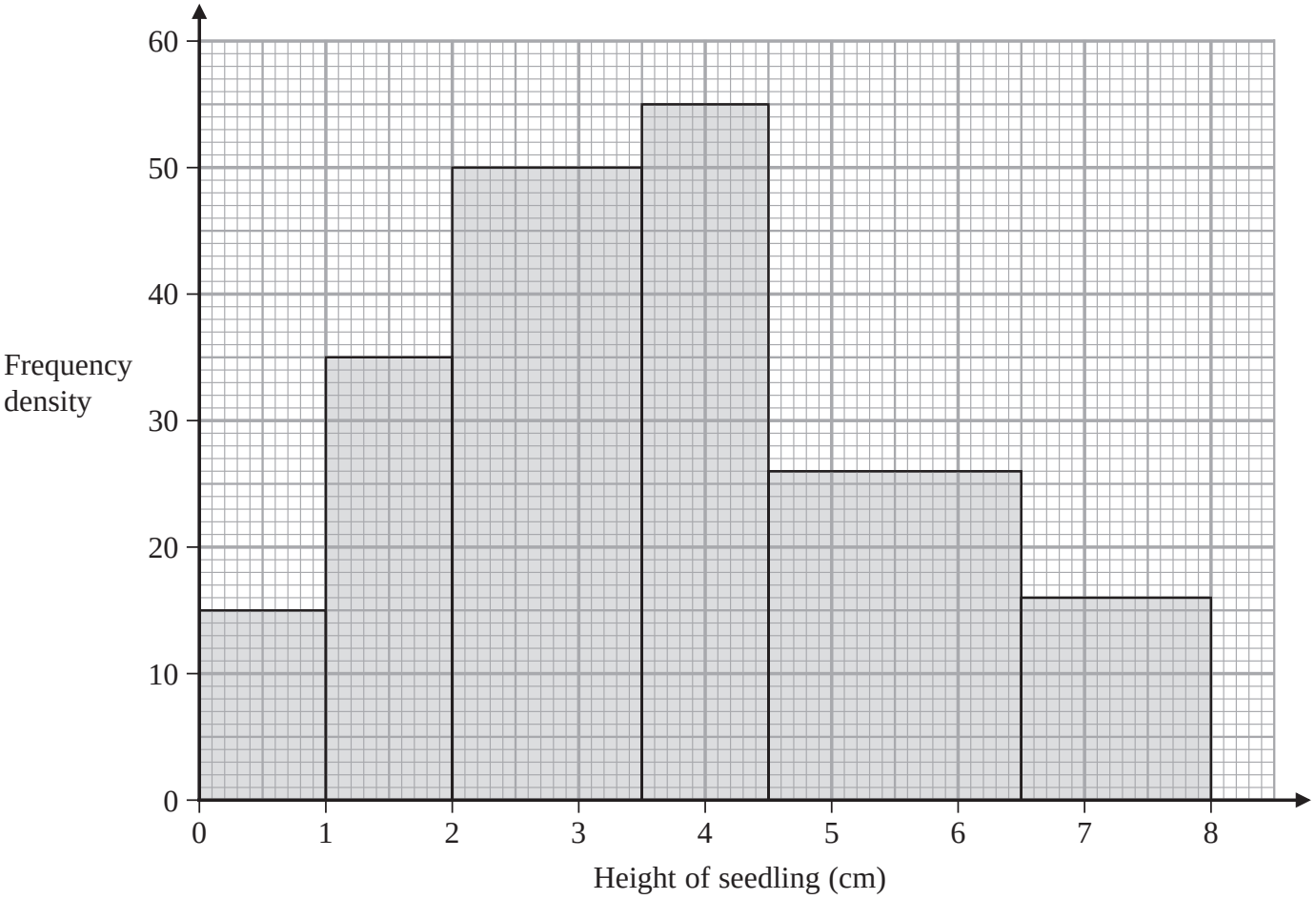
$P(X \leq 2) = 0.0740$ (3 s.f.) (1)

0.0740 is < 0.1, so reject H_0 since sufficient evidence to support B's claim (1)

Question 2 continued

(Total for Question 2 is 7 marks)

3. The histogram summarises the heights of 256 seedlings two weeks after they were planted.



(a) Use linear interpolation to estimate the median height of the seedlings.

(4)

Chris decides to model the **frequency density** for these 256 seedlings by a curve with equation

$$y = kx(8 - x) \qquad 0 \leqslant x \leqslant 8$$

where k is a constant.

(b) Find the value of k

(3)

Using this model,

(c) write down the median height of the seedlings.

(1)

Question 3 continued

a)

class	frequency	cumulative frequency
0 - 1	15	15 (1)
1 - 2	35	50
2 - 3.5	75	125
3.5 - 4.5	55	180
4.5 - 6.5	52	232
6.5 - 8	24	256 (1)

128th term is within this class → 3.5 - 4.5

$\frac{256}{2} = 128^{th}$ → median is the 128^{th} term

$$\text{Median } (Q_2) = 3.5 + \frac{\frac{256}{2} - 125}{55} \times (4.5 - 3.5) \quad (1)$$

$$= 3.55 \ (3 \text{ s.f.}) \quad (1)$$

b) $\int_0^8 kx(8-x)\,dx = 256$ (1) ← which is the total frequency (area under the curve)

$$k\int_0^8 x(8-x)\,dx = 256$$

$$k\int_0^8 8x - x^2\,dx = 256$$

$$k\left[4x^2 - \frac{1}{3}x^3\right]_0^8 = 256 \quad (1)$$

$$k\left[4(8)^2 - \frac{1}{3}(8)^3\right] = 256 \quad (1)$$

$$k\left(\frac{256}{3}\right) = 256$$

$k = 3$ (1)

since the quadratic $y = 3x(8-x)$ has roots 0 and 8, the median value must be between them since parabolas are symmetrical

c) By symmetry, median = 4 (1)

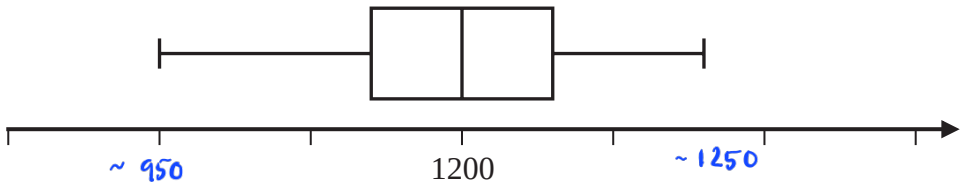
Question 3 continued

(Total for Question 3 is 8 marks)

4. Jiang is studying the variable Daily Mean Pressure from the large data set.

He drew the following box and whisker plot for these data for one of the months for one location using a linear scale but

- he failed to label all the values on the scale
- he gave an incorrect value for the median



Daily Mean Pressure (hPa)

Using your knowledge of the large data set, suggest a suitable value for

(a) the median, **(1)**

(b) the range. **(1)**

(*You are not expected to have memorised values from the large data set. The question is simply looking for sensible answers.*)

For this sort of questions, the only way is to familiarise yourself with range of values for various data sets.

a) 1000 hPa (1)

b) 30 hPa (1)

(Total for Question 4 is 2 marks)

5. Manon has two biased spinners, one red and one green.

The random variable R represents the score when the red spinner is spun.
The random variable G represents the score when the green spinner is spun.

The probability distributions for R and G are given below.

r	2	3
$P(R = r)$	$\frac{1}{4}$	$\frac{3}{4}$

g	1	4
$P(G = g)$	$\frac{2}{3}$	$\frac{1}{3}$

Manon spins each spinner once and adds the two scores.

(a) Find the probability that

(i) the sum of the two scores is 7

(ii) the sum of the two scores is less than 4

(3)

The random variable $X = mR + nG$ where m and n are integers.

$$P(X = 20) = \frac{1}{6} \quad \text{and} \quad P(X = 50) = \frac{1}{4}$$

(b) Find the value of m and the value of n

(5)

a) (i) The sum of the two scores is 7 when:

R = 3 , G = 4

∴ P(two scores is 7) = $\frac{3}{4} \times \frac{1}{3}$ (1)

= $\frac{1}{4}$ (1)

(ii) The sum of the two scores is less than 4 when:

R = 2 , G = 1

∴ P(two scores is less than 4) = $\frac{1}{4} \times \frac{2}{3}$

= $\frac{1}{6}$ (1)

Question 5 continued

b) $P(x = 50) = \frac{1}{4}$ ∴ means R = 3 and G = 4 (1)

(from (a))

$P(x = 20) = \frac{1}{6}$ ∴ means R = 2 and G = 1

(from (a))

So for $P(x = 50)$: $50 = 3m + 4n$ — (1) (1)

for $P(x = 20)$: $20 = 2m + n$ — (2) (1)

substitute (2) into (1)

$50 = 3m + 4(20 - 2m)$ (1)

$= 3m + 80 - 8m$

$5m = 30$

$m = 6$ (1)

Substitute $m = 6$ into (2)

$20 = 2(6) + n$

$n = 8$

Question 5 continued

(Total for Question 5 is 8 marks)

TOTAL FOR STATISTICS IS 30 MARKS