



Edexcel AS Mathematics

8MA0/21 - Statistics

Thursday 9 June 2022

Questions and Fully Worked Model Solutions

This booklet places the **question and the solution together**. The mathematics is set out using proper fractions, powers, roots and exact notation, with extra written explanation so that a student can see *why* each step is being taken.

Solutions have been checked against the Pearson Edexcel Summer 2022 mark scheme. The aim is a clear **full-mark model response**, not merely a list of final answers.

Contents

Question 1	2
Question 2	3
Question 3	4
Question 4	6
Question 5	7

www.mathswithmatt.co.uk

Question 1

Question 1

The relationship between two variables p and t is modelled by the regression line

$$p = 22 - 1.1t.$$

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

- Describe the correlation between p and t implied by this model.
- Given that p is measured in centimetres and t is measured in days, state the units of the gradient of the regression line.
- Using the model, calculate the change in p over a 3-day period.
- Tisam uses this model to estimate the value of p when $t = 19$. Comment, giving a reason, on the reliability of this estimate.

Model solution

(a) The gradient is negative, so as t increases, p tends to decrease. Thus the implied correlation is negative.

Negative correlation.

(1)

(b) Gradient = $\frac{\text{change in } p}{\text{change in } t}$, so its units are centimetres per day.

cm day⁻¹ (or cm/day).

(1)

(c) For an increase of 3 days, the change predicted by the regression line is

$$(-1.1) \times 3 = -3.3 \text{ cm.}$$

p decreases by 3.3 cm.

(2)

(d) The observed values of t lie only between 1 and 10. Since 19 is outside this range, using the model at $t = 19$ is extrapolation.

The estimate is unreliable, because it is an extrapolation beyond the data used to form the model.

(1)

Question 2

Question 2

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

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

- (i) exactly 2;
- (ii) more than 3.

Supplier B claims that the proportion of bags it supplies that are damp is less than 8%. A random sample of 70 bags from supplier B contains only 2 damp bags.

(b) Carry out a suitable test to assess supplier B's claim. State the hypotheses clearly and use a 10% level of significance.

Model solution

(a) Let D be the number of damp bags in the sample from A. Each bag is classified as damp or not damp, with probability 0.08 of being damp, so

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

(i)

$$P(D = 2) = \binom{35}{2} (0.08)^2 (0.92)^{33} = 0.243049 \dots$$

$$P(D = 2) = 0.243 \text{ (3 s.f.)}$$

(1) (ii)

$$P(D > 3) = 1 - P(D \leq 3) = 1 - 0.69397 \dots = 0.30602 \dots$$

$$P(D > 3) = 0.306 \text{ (3 s.f.)}$$

(2)

(b) Let p be the true proportion of supplier B's bags that are damp. Supplier B's claim gives a one-tailed lower-tail test:

$$H_0 : p = 0.08, \quad H_1 : p < 0.08.$$

Under H_0 , if X is the number of damp bags in a sample of 70, then $X \sim B(70, 0.08)$. The observed value is $x = 2$, so the one-tailed p-value is

$$P(X \leq 2) = 0.0739756 \dots \approx 0.074.$$

Since $0.074 < 0.10$, the result is significant at the 10% level. Reject H_0 .

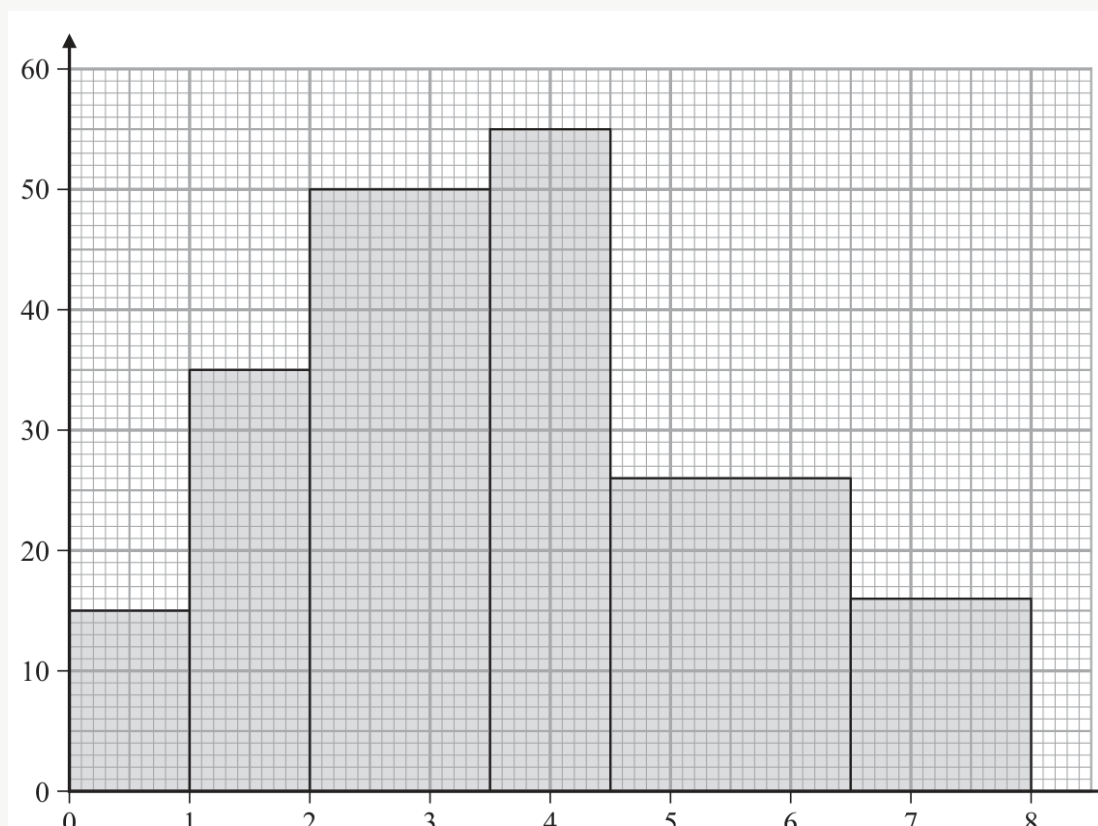
There is sufficient evidence, at the 10% level, to support supplier B's claim that the proportion of damp bags is less than 8%.

(4)

Question 3

Question 3

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



- (a) Use linear interpolation to estimate the median height of the seedlings.
 (b) Chris models the frequency density by the curve

$$y = kx(8 - x), \quad 0 \leq x \leq 8,$$

where k is a constant. Find k .

- (c) Using this model, write down the median height of the seedlings.

Model solution

(a) Frequency is the area of a histogram bar. The relevant grouped frequency table is

Height (cm)	frequency	cumulative frequency
$0 \leq x < 1$	$15 \times 1 = 15$	15
$1 \leq x < 2$	$35 \times 1 = 35$	50
$2 \leq x < 3.5$	$50 \times 1.5 = 75$	125
$3.5 \leq x < 4.5$	$55 \times 1 = 55$	180

The median is the $\frac{256}{2} = 128$ th value. It lies in the class $3.5 \leq x < 4.5$, since the cumulative frequency rises from 125 to 180 in that class. Linear interpolation gives

$$\text{median} = 3.5 + \frac{128 - 125}{55}(4.5 - 3.5) = 3.5545 \dots$$

Estimated median height = 3.55 cm (3 s.f.).

(4)

(b) The total area under a frequency-density curve is the total frequency, 256. Hence

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

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

$$k \left(4 \times 8^2 - \frac{8^3}{3} \right) = 256 \Rightarrow k \left(256 - \frac{512}{3} \right) = 256.$$

$$\frac{256k}{3} = 256 \Rightarrow k = 3.$$

$k = 3.$

(3)

(c) The curve $y = kx(8-x)$ is symmetric about $x = 4$, so equal areas, and hence equal frequencies, lie on each side of $x = 4$.

Median height = 4 cm.

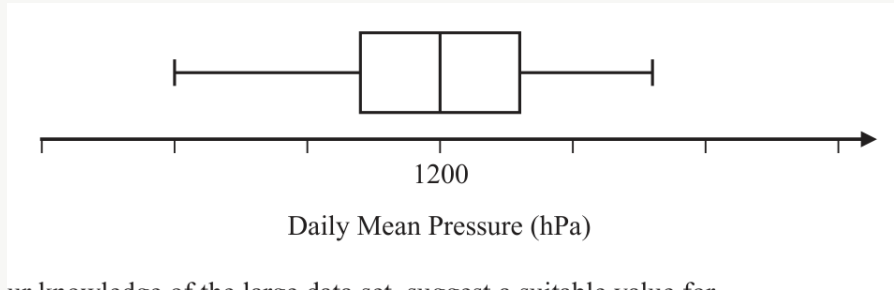
(1)

www.mathswithmatt.co.uk

Question 4

Question 4

Jiang is studying Daily Mean Pressure from the large data set. He drew the following box-and-whisker plot for one month at one location, on a linear scale, but failed to label all scale values and gave an incorrect value for the median.



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

- (a) the median;
- (b) the range.

You are not expected to have memorised values from the large data set; sensible answers are required.

Model solution

(a) Daily mean atmospheric pressure is normally close to 1000 hPa. A suitable corrected median is therefore near this value.

1010 hPa (a suitable value).

(1)

(b) For daily mean pressures in one month, a spread of a few tens of hPa is sensible.

20 hPa (a suitable value).

(1)

Question 5

Question 5

Manon has two biased spinners, one red and one green. The random variables R and G are their respective scores:

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 is 7;
- (ii) the sum is less than 4.

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

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

(b) find m and n .

Model solution

(a)(i) The sum is 7 only when $R = 3$ and $G = 4$. As the spinners are spun independently,

$$P(R = 3 \text{ and } G = 4) = \frac{3}{4} \times \frac{1}{3} = \frac{1}{4}.$$

$$\frac{1}{4}.$$

(2) (a)(ii) The only possible sum less than 4 is $2 + 1 = 3$, so $R = 2$ and $G = 1$. Thus

$$P(R = 2 \text{ and } G = 1) = \frac{1}{4} \times \frac{2}{3} = \frac{1}{6}.$$

$$\frac{1}{6}.$$

(1)

(b) The probability $P(X = 50) = \frac{1}{4}$ corresponds to $R = 3$, $G = 4$, because

$$P(R = 3, G = 4) = \frac{3}{4} \times \frac{1}{3} = \frac{1}{4}.$$

Therefore $3m + 4n = 50$. Similarly, $P(X = 20) = \frac{1}{6}$ corresponds to $R = 2$, $G = 1$, because

$$P(R = 2, G = 1) = \frac{1}{4} \times \frac{2}{3} = \frac{1}{6},$$

so $2m + n = 20$. Solving, $n = 20 - 2m$, and hence

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

$$-5m = -30 \Rightarrow m = 6.$$

$$n = 20 - 2(6) = 8.$$

$$m = 6 \text{ and } n = 8.$$

(5)

www.mathswithmatt.co.uk