



Edexcel AS Mathematics

8MA0/21 - Statistics

Thursday 25 May 2023

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 2023 mark scheme. The aim is a clear **full-mark model response**, not merely a list of final answers.

Contents

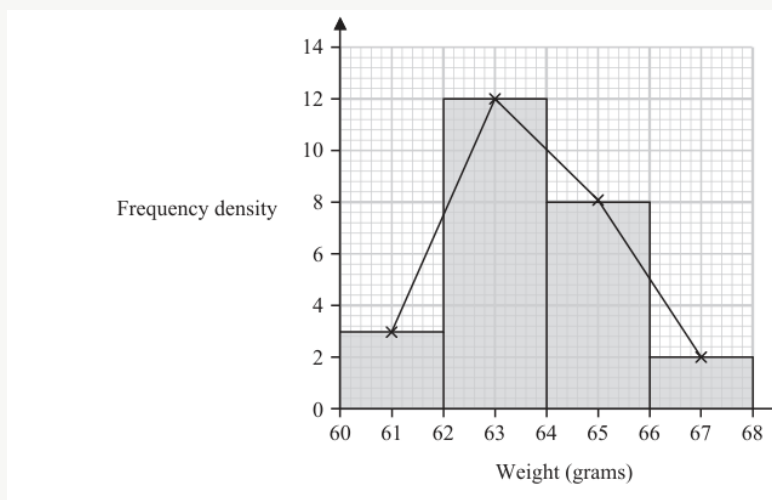
Question 1	2
Question 2	4
Question 3	6
Question 4	7
Question 5	8

www.mathswithmatt.co.uk

Question 1

Question 1

The histogram and frequency polygon give information about the weights, in grams, of 50 plums.



(a) Show that an estimate for the mean weight of the 50 plums is 63.72 grams.

(b) Calculate an estimate for the standard deviation of the 50 plums.

Later it was discovered that the scales used to weigh the plums were broken. Each plum actually weighs 5 grams less than originally thought.

(c) State the effect this will have on the estimate of the standard deviation in part (b). Give a reason for your answer.

Model solution

The class intervals have width 2, so frequency is frequency density $\times 2$.

weight (g)	$60 \leq w < 62$	$62 \leq w < 64$	$64 \leq w < 66$	$66 \leq w < 68$
midpoint x	61	63	65	67
frequency f	6	24	16	4

(a)

$$\bar{x} = \frac{61(6) + 63(24) + 65(16) + 67(4)}{50} = \frac{3186}{50} = 63.72.$$

Estimated mean = 63.72 g.

(2)

(b)

$$\sum fx^2 = 61^2(6) + 63^2(24) + 65^2(16) + 67^2(4) = 203,138.$$

$$s = \sqrt{\frac{\sum fx^2}{\sum f} - \bar{x}^2} = \sqrt{\frac{203138}{50} - 63.72^2} = 1.58795 \dots$$

Estimated standard deviation = 1.59 g.

(2)

(c) Subtracting the same constant from every value does not change the distances of the values from their mean, so it does not change their spread.

There is no effect: the estimated standard deviation remains 1.59 g.

(1)

www.mathswithmatt.co.uk

Question 2

Question 2

Fred and Nadine are investigating whether there is a linear relationship between Daily Mean Pressure, p hPa, and Daily Mean Air Temperature, t °C, in Beijing using data from the large data set. Fred randomly selects one month and draws Figure 1. The scale has been left off the horizontal axis.

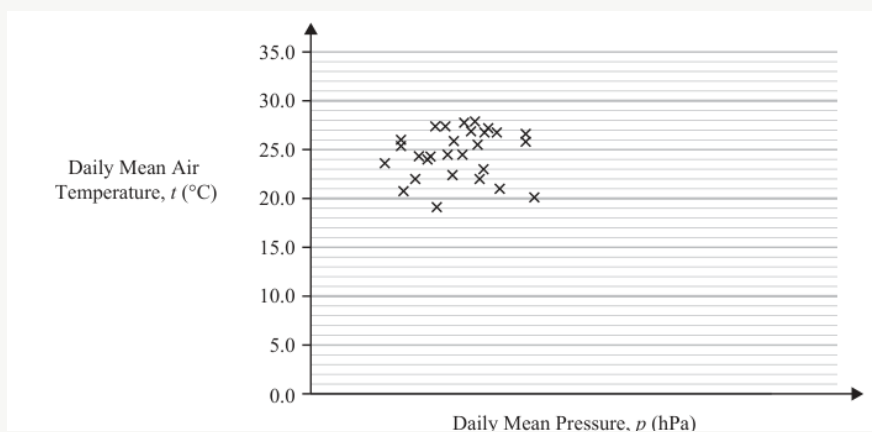


Figure 1

(a) Describe the correlation shown in Figure 1.

Nadine uses all the data for Beijing from 2015 and draws Figure 2, using the same scales as Fred.

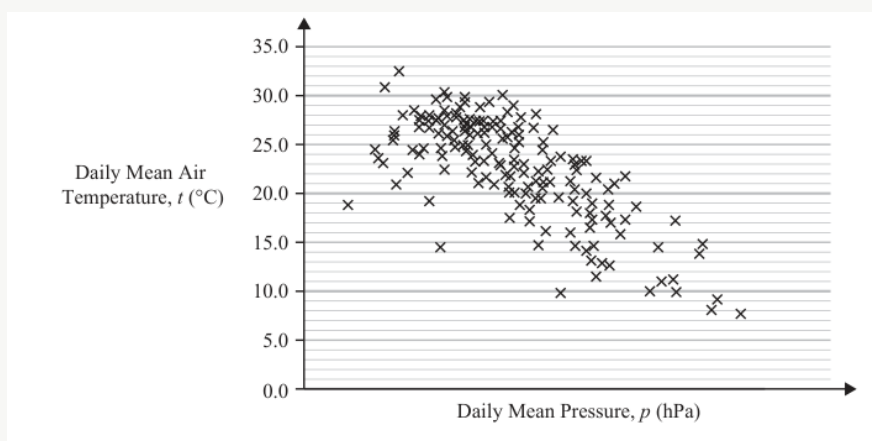


Figure 2

(b) Explain, in context, what Nadine can infer about the relationship between p and t using Figure 2.

(c) Using your knowledge of the large data set, state a value of p for which interpolation can be used with Figure 2 to predict a value of t .

(d) Using your knowledge of the large data set, explain why it is not meaningful to look for a linear relationship between Daily Mean Wind Speed (Beaufort Conversion) and Daily Mean Air Temperature in Beijing in 2015.

Model solution

(a) *The points have no clear linear pattern.*

There is no correlation (or only weak correlation).

(1)

(b) Figure 2 shows a negative association: as daily mean pressure p increases, daily mean air temperature t tends to decrease.

As pressure increases, temperature decreases.

(1)

(c) The pressure must be within the range of pressure values represented in the data; for example, $p = 1000$ hPa lies within that range.

$p = 1000$ hPa.

(1)

(d) Daily Mean Wind Speed (Beaufort Conversion) is a qualitative (categorical) variable, rather than a quantitative variable. A linear relationship is therefore not meaningful.

Wind speed in Beaufort categories is qualitative, so a linear relationship cannot meaningfully be investigated.

(1)

Question 3

Question 3

In an after-school club, students can choose to take part in Art, Music, both or neither. There are 45 students. Of these, 25 take part in Art, 12 take part in both Art and Music, and the number who take part in Music is x .

(a) Find the range of possible values of x .

One student is selected at random. Event A is that the student takes part in Art and event M is that the student takes part in Music.

(b) Determine whether or not it is possible for events A and M to be independent.

Model solution

(a) There are $25 - 12 = 13$ students in Art only. Since the 12 students who take both must be included in Music,

$$x \geq 12.$$

At most all 20 students not in Art can also take Music, giving $12 + 20 = 32$ students in Music. Hence

$$12 \leq x \leq 32.$$

(2)

(b) If A and M were independent, then

$$P(A)P(M) = P(A \cap M).$$

Here

$$P(A) = \frac{25}{45}, \quad P(A \cap M) = \frac{12}{45}, \quad P(M) = \frac{x}{45}.$$

Therefore independence would require

$$\frac{25}{45} \cdot \frac{x}{45} = \frac{12}{45},$$

so

$$x = \frac{12}{25} \times 45 = 21.6.$$

The number of students taking Music must be a whole number, whereas 21.6 is not.

It is not possible for A and M to be independent.

(4)

Question 4

Question 4

Past information shows that 25% of adults in a large population have a particular allergy. Rylan believes that the proportion with the allergy differs from 25%. He takes a random sample of 50 adults and tests $H_0 : p = 0.25$ at the 5% level of significance.

- Write down the alternative hypothesis.
 - Find the critical region. State the probability associated with each tail, each as close to 2.5% as possible.
 - State the actual probability of incorrectly rejecting H_0 for this test.
- Rylan finds that 10 adults in his sample have the allergy.
- State the conclusion of Rylan's hypothesis test.

Model solution

(a)

$$H_1 : p \neq 0.25.$$

(1)

(b) Under H_0 , let X be the number in the sample with the allergy. Then

$$X \sim B(50, 0.25).$$

For the lower tail,

$$P(X \leq 6) = 0.0194,$$

and for the upper tail,

$$P(X \geq 19) = 0.0287.$$

Both are as close to 0.025 as possible while keeping the total significance probability below 0.05.

Critical region: $X \leq 6$ or $X \geq 19$; $P(X \leq 6) = 0.0194$ and $P(X \geq 19) = 0.0287$.

(4)

(c)

$$P(\text{reject } H_0 \mid H_0 \text{ true}) = 0.0194 + 0.0287 = 0.0481.$$

The actual significance level is 0.0481 (approximately 0.048).

(1)

(d) The observed value is $X = 10$, which is not in the critical region. Do not reject H_0 .

There is insufficient evidence at the 5% level to suggest that the proportion of adults with the allergy differs from 25%.

(1)

Question 5

Question 5

Julia selects 3 letters at random, one at a time without replacement, from the word

VARIANCE.

The discrete random variable X represents the number of times she selects the letter A.

(a) Find the complete probability distribution of X .

Yuki selects 10 letters at random, one at a time with replacement, from the word

DEVIATION.

(b) Find the probability that she selects the letter E at least 4 times.

Model solution

(a) There are two As and six non-As among the eight letters. Thus X can take the values 0, 1 and 2.

$$P(X = 0) = \frac{6}{8} \cdot \frac{5}{7} \cdot \frac{4}{6} = \frac{5}{14}.$$

For exactly one A, choose the one of the three selection positions occupied by A:

$$P(X = 1) = 3 \left(\frac{2}{8} \cdot \frac{6}{7} \cdot \frac{5}{6} \right) = \frac{15}{28}.$$

For two As, choose the two positions occupied by A:

$$P(X = 2) = 3 \left(\frac{2}{8} \cdot \frac{1}{7} \cdot \frac{6}{6} \right) = \frac{3}{28}.$$

Hence the complete distribution is

x	0	1	2
$P(X = x)$	$\frac{5}{14}$	$\frac{15}{28}$	$\frac{3}{28}$

$$P(X = 0) = \frac{5}{14}, P(X = 1) = \frac{15}{28}, P(X = 2) = \frac{3}{28}.$$

(5)

(b) There is one E in the nine letters of DEVIATION, and selections are with replacement. Let J be the number of Es selected. Then

$$J \sim B\left(10, \frac{1}{9}\right).$$

$$P(J \geq 4) = 1 - P(J \leq 3) = 1 - \sum_{j=0}^3 \binom{10}{j} \left(\frac{1}{9}\right)^j \left(\frac{8}{9}\right)^{10-j} = 0.0184\dots$$

$$P(J \geq 4) = 0.0184 \text{ (approximately).}$$

(3)