Maths with Matt

Physics and maths resources

Edexcel A Level Mathematics

9MA0/31 - Statistics

Thursday 20 June 2024

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 2024 mark scheme. The aim is a clear **full-mark model response**, not merely a list of final answers.

Question wording is reproduced from the user-provided 2024 paper for study/reference. Pearson retains copyright in the original examination material.

Contents

Question 1 2

Question 2 4

Question 3 7

Question 4 9

Question 5 11

Question 6 13

www.mathswithmatt.co.uk

Question 1

Question 1

Xian rolls a fair die 10 times.
The random variable X represents the number of times the die lands on a six.
(**a**) Using a suitable distribution for X, find
(**i**) $P(X = 3)$
(**ii**) $P(X < 3)$ **(3)**
Xian repeats this experiment each day for 60 days and records the number of days when $X = 3$.
(**b**) Find the probability that there were at least 12 days when $X = 3$. **(3)**
(**c**) Find an estimate for the total number of sixes that Xian will roll during these 60 days. **(1)**
(**d**) Use a normal approximation to estimate the probability that Xian rolls a total of more than 95 sixes during these 60 days. **(4)**

Model solution

(a)
Each roll is an independent trial with a constant probability $\frac{1}{6}$ of a six, so

$$X \sim B\left(10, \frac{1}{6}\right).$$

(i)

$$P(X = 3) = \binom{10}{3}\left(\frac{1}{6}\right)^3\left(\frac{5}{6}\right)^7 = 0.155045\ldots$$

$$P(X = 3) = 0.155 \text{ to 3 s.f.}$$

(ii)
The event $X < 3$ includes the three values 0, 1 and 2.

$$\begin{aligned} P(X < 3) &= P(X \leqslant 2) \\ &= \sum_{x=0}^{2} \binom{10}{x}\left(\frac{1}{6}\right)^x\left(\frac{5}{6}\right)^{10-x} \\ &= 0.775226\ldots \end{aligned}$$

$$P(X < 3) = 0.775 \text{ to 3 s.f.}$$

(b)
Let D be the number of days, out of 60, on which $X = 3$. The daily experiments are independent and the probability from part (a)(i) is constant, so

$$D \sim B(60, 0.155045\ldots).$$

"At least 12" means $D \geqslant 12$, so it is efficient to use the complement $D \leqslant 11$.

$$\begin{aligned} P(D \geqslant 12) &= 1 - P(D \leqslant 11) \\ &= 1 - 0.788190\ldots \\ &= 0.211809\ldots \end{aligned}$$

$$P(D \geqslant 12) = 0.212 \textit{ to 3 s.f.}$$

(c)

There are $60 \times 10 = 600$ *rolls in total. Hence the expected number of sixes is*

$$600\left(\frac{1}{6}\right) = 100.$$

$$\textit{Estimated total number of sixes} = 100.$$

(d)

Let S *be the total number of sixes in the 600 rolls. Then*

$$S \sim B\left(600, \frac{1}{6}\right),$$

with

$$\mu = np = 100, \qquad \sigma^2 = np(1-p) = 600\left(\frac{1}{6}\right)\left(\frac{5}{6}\right) = \frac{250}{3}.$$

Since np *and* $n(1-p)$ *are both large, use*

$$Y \sim N\left(100, \frac{250}{3}\right)$$

as an approximation to S.

Because S *is discrete,* $S > 95$ *means* $S \geqslant 96$. *The continuity-corrected boundary is therefore* 95.5.

$$\begin{aligned} P(S > 95) &\approx P(Y > 95.5) \\ &= P\left(Z > \frac{95.5 - 100}{\sqrt{250/3}}\right) \\ &= P(Z > -0.492950\ldots) \\ &= 0.688976\ldots \end{aligned}$$

$$P(S > 95) \approx 0.689 \textit{ to 3 s.f.}$$

Question 2

www.mathswithmatt.co.uk

Question 2

Amar is studying the flight of a bird from its nest.
He measures the bird's height above the ground, h metres, at time t seconds for 10 values of t.
Amar finds the equation of the regression line for the data to be

$$h = 38.6 - 1.28t.$$

(**a**) Interpret the gradient of this line. **(1)**

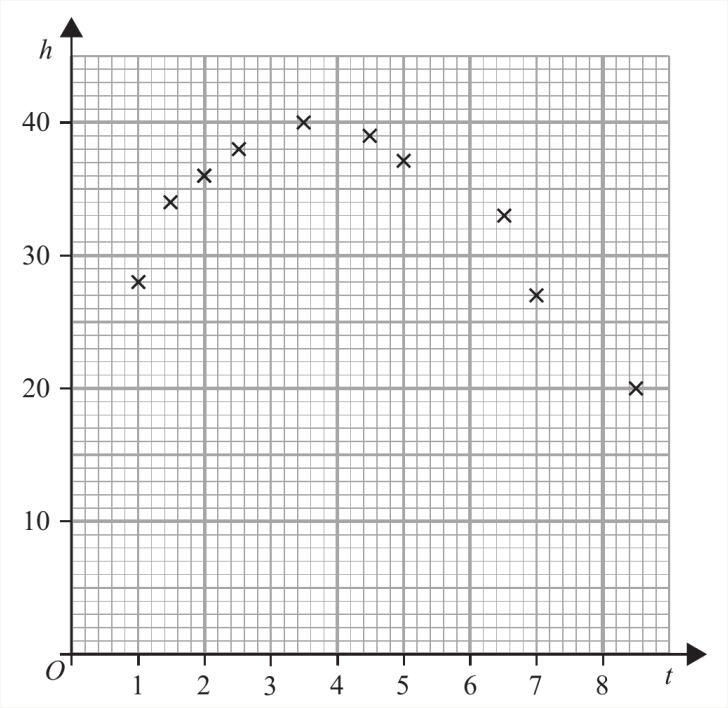
The product moment correlation coefficient between h and t is -0.510.

(**b**) Test whether or not there is evidence of a negative correlation between the height above the ground and the time during the flight.
You should

- state your hypotheses clearly,
- use a 5% level of significance,
- state the critical value used.

(3)

Jane draws the following scatter diagram for Amar's data.



(**c**) With reference to the scatter diagram, state, giving a reason, whether or not the regression line $h = 38.6 - 1.28t$ is an appropriate model for these data. **(1)**

Jane suggests an improved model using the variable

$$u = (t - k)^2,$$

where k is a constant. She obtains the equation

$$h = 38.1 - 0.78u.$$

(**d**) Choose a suitable value for k to write Jane's improved model for h in terms of t only. **(1)**

Model solution

(a)

The gradient is -1.28, so the model predicts that the bird's height above the ground decreases by approximately 1.28 metres for each additional second of the flight.

Height decreases by approximately $1.28\,\mathrm{m}$ per second.

(b)

Let ρ be the population product moment correlation coefficient between height and time.

$$H_0 : \rho = 0, \qquad H_1 : \rho < 0.$$

This is a one-tailed test because the alternative hypothesis specifies negative correlation. For $n = 10$ at the 5% significance level, the critical value has magnitude 0.5494; in the lower tail the critical region is

$$r < -0.5494.$$

The observed value is $r = -0.510$, and

$$-0.510 > -0.5494,$$

so it is not in the critical region. We do not reject H_0.

There is insufficient evidence at the 5% level of a negative correlation between the bird's height above the ground and the time during its flight.

(c)

The regression line is not appropriate because the scatter diagram shows a curved pattern: height initially rises and then falls, whereas the proposed model is a straight line with a constant negative gradient.

No; the data follow a curved rather than linear pattern.

(d)

The scatter diagram reaches its maximum height at about $t = 3.5$, so choose $k = 3.5$. Substituting this value into $u = (t - k)^2$ gives

$$h = 38.1 - 0.78(t - 3.5)^2.$$

$$h = 38.1 - 0.78(t - 3.5)^2$$

Question 3

Question 3

Ming is studying the large data set for Perth in 2015.
He intended to use all the data available to find summary statistics for the Daily Mean Air Temperature, x °C.
Unfortunately, Ming selected an incorrect variable on the spreadsheet. This incorrect variable gave a mean of 5.3 and a standard deviation of 12.4.

(**a**) Using your knowledge of the large data set, suggest which variable Ming selected. **(1)**

The correct values for the Daily Mean Air Temperature are summarised as

$$n = 184, \qquad \sum x = 2801.2, \qquad \sum x^2 = 44695.4.$$

(**b**) Calculate the mean and standard deviation for these data. **(3)**

One of the months from the large data set for Perth in 2015 has

$$\bar{x} = 19.4, \qquad \sigma_x = 2.83$$

for Daily Mean Air Temperature.

(**c**) Suggest, giving a reason, a month these data may have come from. **(2)**

Model solution

(a)

A mean of 5.3 *with a much larger standard deviation of* 12.4 *is consistent with daily rainfall, since many days have little or no rain but a few days can have much larger totals.*

Daily Total Rainfall.

(b)

The mean is

$$\bar{x} = \frac{\sum x}{n} = \frac{2801.2}{184} = 15.2239\ldots$$

so

$$\bar{x} = 15.2\,^\circ\text{C} \quad \textit{to 3 s.f.}$$

Using the population standard-deviation formula,

$$\begin{aligned}\sigma_x &= \sqrt{\frac{\sum x^2}{n} - \bar{x}^2} \\ &= \sqrt{\frac{44695.4}{184} - \left(\frac{2801.2}{184}\right)^2} \\ &= 3.33800\ldots\end{aligned}$$

$\bar{x} = 15.2\,^\circ\text{C}$ *and* $\sigma_x = 3.34\,^\circ\text{C}$, *to 3 s.f.*

(c)

The monthly mean 19.4 °C is higher than the overall mean 15.2 °C, so the data should come from one of Perth's warmer months. Perth is in the southern hemisphere and, among the months included in this large data set, October is the appropriate warmer month.

October, because its mean temperature is higher than the overall mean and it is a warmer spring month in Perth.

www.mathswithmatt.co.uk

Question 4

Question 4

The proportion of left-handed adults in a country is 10%.
Freya believes that the proportion of left-handed adults under the age of 25 in this country is different from 10%.
She takes a random sample of 40 adults under the age of 25 from this country to investigate her belief.
(**a**) Find the critical region for a suitable test to assess Freya's belief.
You should

- state your hypotheses clearly,
- use a 5% level of significance,
- state the probability of rejection in each tail.

(4)

(**b**) Write down the actual significance level of your test in part (a). **(1)**

In Freya's sample 7 adults were left-handed.

(**c**) With reference to your answer in part (a), comment on Freya's belief. **(1)**

Model solution

Let p be the proportion of adults under 25 in the country who are left-handed, and let X be the number of left-handed adults in Freya's sample.

(a)

Freya believes the proportion is different from 10%, so a two-tailed test is required:

$$H_0 : p = 0.10, \qquad H_1 : p \neq 0.10.$$

Under H_0,

$$X \sim B(40, 0.10).$$

For the lower tail,

$$P(X = 0) = 0.9^{40} = 0.0147809\ldots < 0.025.$$

The only possible lower-tail critical value is therefore $X = 0$.
For the upper tail,

$$P(X \geqslant 9) = 1 - P(X \leqslant 8) = 0.0154953\ldots < 0.025,$$

whereas

$$P(X \geqslant 8) = 0.04190\ldots > 0.025.$$

Hence the upper tail begins at $X = 9$.

Critical region: $X = 0$ or $X \geqslant 9$. The rejection probabilities are 0.0148 in the lower tail and 0.0155 in the upper tail.

(b)

The actual significance level is the total probability of the two tails:

$$0.0147809\ldots + 0.0154953\ldots = 0.0302762\ldots$$

Actual significance level $= 0.0303$, *or* 3.03%, *to 3 s.f.*

(c)

The observed value $X = 7$ *is not in the critical region, so we do not reject* H_0.

There is insufficient evidence to support Freya's belief that the proportion of left-handed adults under 25 is different from 10%.

www.mathswithmatt.co.uk

Question 5

Question 5

The records for a school athletics club show that the height, H metres, achieved by students in the high jump is normally distributed with mean 1.4 metres and standard deviation 0.15 metres.

(**a**) Find the proportion of these students achieving a height of more than 1.6 metres. **(1)**

The records also show that the time, T seconds, to run 1500 metres is normally distributed with mean 330 seconds and standard deviation 26 seconds.

The school's Head would like to use these distributions to estimate the proportion of students from the school athletics club who can jump higher than 1.6 metres and can run 1500 metres in less than 5 minutes.

(**b**) State a necessary assumption about H and T for the Head to calculate an estimate of this proportion. **(1)**

(**c**) Find the Head's estimate of this proportion. **(3)**

Students in the school athletics club also throw the discus.

The random variable $D \sim N(\mu, \sigma^2)$ represents the distance, in metres, that a student can throw the discus.

Given that

$$P(D < 16.3) = 0.30 \qquad \text{and} \qquad P(D > 29.0) = 0.10,$$

(**d**) calculate the value of μ and the value of σ. **(5)**

Model solution

(a)

Since $H \sim N(1.4, 0.15^2)$,

$$\begin{aligned} P(H > 1.6) &= P\left(Z > \frac{1.6 - 1.4}{0.15}\right) \\ &= P(Z > 1.3333\ldots) \\ &= 0.091211\ldots \end{aligned}$$

Proportion $= 0.0912$ *to 3 s.f.*

(b)

To multiply the two probabilities, the Head must assume that high-jump height and 1500-metre time are independent.

H *and* T *are independent.*

(c)

Five minutes is 300 *seconds. Since* $T \sim N(330, 26^2)$,

$$\begin{aligned} P(T < 300) &= P\left(Z < \frac{300 - 330}{26}\right) \\ &= P(Z < -1.15385\ldots) \\ &= 0.124282\ldots \end{aligned}$$

Using the independence assumption,

$$\begin{aligned} \mathrm{P}(\mathrm{H} > 1.6 \textit{ and } \mathrm{T} < 300) &= \mathrm{P}(\mathrm{H} > 1.6)\,\mathrm{P}(\mathrm{T} < 300) \\ &= (0.0912112\ldots)(0.124282\ldots) \\ &= 0.0113359\ldots \end{aligned}$$

Estimated proportion $= 0.0113$ to 3 s.f.

(d)

Convert each stated probability to its corresponding standard-normal quantile. The upper-tail probability 0.10 above 29 means that the cumulative probability below 29 is 0.90.

$$\mathrm{P}(Z < z_1) = 0.30 \quad \Longrightarrow \quad z_1 = -0.524400\ldots,$$

and

$$\mathrm{P}(Z < z_2) = 0.90 \quad \Longrightarrow \quad z_2 = 1.281552\ldots.$$

Standardising the two distances gives

$$\frac{16.3 - \mu}{\sigma} = -0.524400\ldots$$

and

$$\frac{29.0 - \mu}{\sigma} = 1.281552\ldots.$$

Equivalently,

$$16.3 = \mu - 0.524400\ldots\sigma$$

and

$$29.0 = \mu + 1.281552\ldots\sigma.$$

Subtracting the first equation from the second eliminates μ:

$$12.7 = (1.281552\ldots + 0.524400\ldots)\sigma.$$

Therefore

$$\sigma = 7.03230\ldots$$

and then

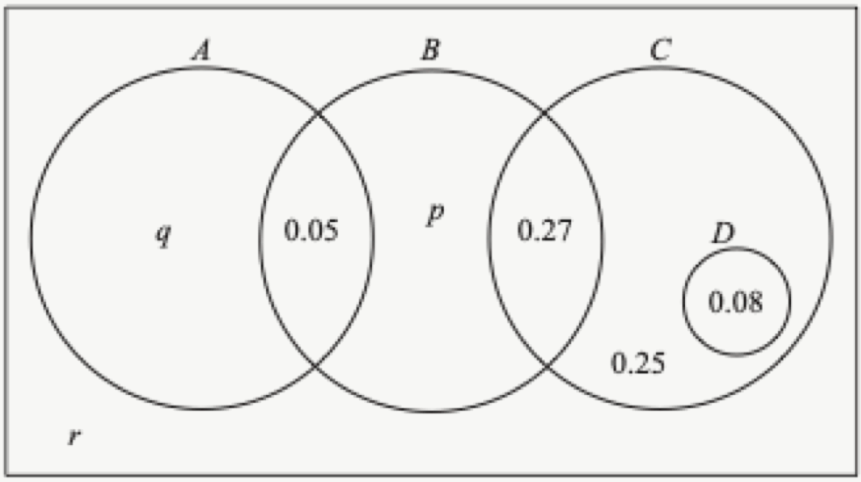
$$\mu = 16.3 + 0.524400\ldots(7.03230\ldots) = 19.9877\ldots$$

$\mu = 20.0\,\mathrm{m}$ and $\sigma = 7.03\,\mathrm{m}$, to 3 s.f.

Question 6

Question 6

The Venn diagram, where p, q and r are probabilities, shows the events A, B, C and D and associated probabilities.



(**a**) State any pair of mutually exclusive events from A, B, C and D. **(1)**

The events B and C are independent.

(**b**) Find the value of p. **(2)**

(**c**) Find the greatest possible value of $P(A \mid B')$. **(3)**

Given that

$$P(B \mid A') = 0.5,$$

(**d**) find the value of q and the value of r. **(3)**

(**e**) Find

$$P\left((A \cup B)' \cap C\right).$$

(1)

(**f**) Use set notation to write an expression for the event with probability p. **(1)**

Model solution

(a)

The circles for A and C do not overlap, so these events cannot occur together.

A and C are mutually exclusive.

(b)

From the diagram,

$$P(C) = 0.27 + 0.25 + 0.08 = 0.60,$$

$$P(B) = 0.05 + p + 0.27 = p + 0.32,$$

and

$$P(B \cap C) = 0.27.$$

Independence means $P(B \cap C) = P(B)P(C)$. *Hence*

$$0.27 = (p + 0.32)(0.60).$$

Therefore

$$p + 0.32 = 0.45 \qquad \Longrightarrow \qquad p = 0.13.$$

$$\boxed{p = 0.13}$$

(c)

Adding every region in the sample space gives

$$q + 0.05 + 0.13 + 0.27 + 0.25 + 0.08 + r = 1,$$

so

$$q + r = 0.22.$$

Also,

$$P(B') = 1 - P(B) = 1 - 0.45 = 0.55.$$

The part of A *that also lies in* B' *is the* A*-only region, with probability* q*. Thus*

$$P(A \mid B') = \frac{P(A \cap B')}{P(B')} = \frac{q}{0.55}.$$

Since $q + r = 0.22$ *and probabilities cannot be negative, the greatest possible value of* q *occurs when* $r = 0$*, giving* $q = 0.22$*. Therefore*

$$\max P(A \mid B') = \frac{0.22}{0.55} = 0.4.$$

$$\boxed{\textit{Greatest possible value} = 0.4.}$$

(d)

The part of B *outside* A *has probability*

$$P(B \cap A') = p + 0.27 = 0.13 + 0.27 = 0.40.$$

The regions outside A *give*

$$P(A') = 0.13 + 0.27 + 0.25 + 0.08 + r = 0.73 + r.$$

Using the given conditional probability,

$$\frac{0.40}{0.73 + r} = 0.5.$$

Hence

$$0.40 = 0.365 + 0.5r \qquad \Longrightarrow \qquad r = 0.07.$$

Since $q + r = 0.22$*,*

$$q = 0.22 - 0.07 = 0.15.$$

$$q = 0.15 \text{ and } r = 0.07.$$

(e)

The event $(A \cup B)' \cap C$ *consists of the regions inside* C *but outside both* A *and* B*. Their probabilities are* 0.25 *and* 0.08*, so*

$$P((A \cup B)' \cap C) = 0.25 + 0.08 = 0.33.$$

$$0.33$$

(f)

The probability p *labels the region inside* B *but outside both* A *and* C*.*

$$B \cap A' \cap C'$$

www.mathswithmatt.co.uk