Maths with Matt

Physics and maths resources

OCR AS Level Further Mathematics A

H235/Y532 - Statistics

Friday 16 May 2025

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 OCR Friday 16 May 2025 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 2025 paper for study/reference. OCR retains copyright in the original examination material.

Contents

Question 1 2

Question 2 3

Question 3 5

Question 4 7

Question 5 9

Question 6 10

www.mathswithmatt.co.uk

Question 1

Question 1

A six-sided dice may or may not be equally likely to land on any one of its faces. It has two faces numbered 5 and the other four faces are numbered 1, 2, 3 and 4. The dice is thrown 60 times, and the outcome on each throw is recorded. The results are shown in the table.

Outcome	1	2	3	4	5
Frequency	7	13	11	15	14

It is required to carry out a test at the 5% significance level of whether the probabilities of the outcomes 1, 2, 3, 4 and 5 are in the ratio $1 : 1 : 1 : 1 : 2$.

(a) State suitable hypotheses for the test. [1]

(b) Carry out the test. [6]

Model solution

(a) The null hypothesis states the claimed probability ratio; the alternative is that this ratio does not hold.

H_0: *the probabilities of outcomes* $1, 2, 3, 4, 5$ *are in the ratio* $1 : 1 : 1 : 1 : 2$.
H_1: *the probabilities are not in this ratio.* **(1)**

(b) A chi-squared goodness-of-fit test is appropriate because the observed frequencies are being compared with frequencies predicted by a stated distribution. The ratio contains six equal parts, so the expected frequencies are $10, 10, 10, 10, 20$.

Outcome	1	2	3	4	5
O	7	13	11	15	14
E	10	10	10	10	20
$\frac{(O-E)^2}{E}$	0.9	0.9	0.1	2.5	1.8

$$\chi^2 = 0.9 + 0.9 + 0.1 + 2.5 + 1.8 = 6.2.$$

There are $5-1 = 4$ *degrees of freedom. At the* 5% *level, the critical value is* 9.488. *Since* $6.2 < 9.488$, *the statistic is not in the critical region.*

Do not reject H_0. *There is insufficient evidence that the probabilities are not in the ratio* $1 : 1 : 1 : 1 : 2$. **(6)**

Question 2

Question 2

On any day, flights of helicopters and aeroplanes pass within sight of a certain building only during a daytime period of 12 hours, from 7.00 am to 7.00 pm.
The random variable H is the number of helicopters that pass within sight of the building during a randomly chosen daytime period of 12 hours.
(a) State two assumptions needed for H to be well modelled by a Poisson distribution. [2]
Assume now that H can be well modelled by the distribution Po(1.8).
(b) Find the probability that, during a randomly chosen daytime period of 12 hours, the number of helicopters that pass within sight of the building is 2, 3 or 4. [2]
(c) During a period of t hours (within the daytime period of 12 hours), the probability that no helicopters pass within sight of the building is 0.95.
Use an algebraic method to determine the value of t. [3]
(d) During a randomly chosen daytime period of t hours (where t is the value found in part (c)) two helicopters pass within sight of the building.
Explain whether this casts doubt on the validity of the model. You do not need to carry out any calculations. [1]
Throughout the daytime period of 12 hours on any day, aeroplanes pass within sight of the building at a fixed constant rate of 1 aeroplane every 4 minutes.
(e) Explain whether the number of aeroplanes that pass within sight of the building on any day is likely to be well modelled by a Poisson distribution. [1]
(f) Find the expected value of the total number of helicopters and aeroplanes that pass within sight of the building on a randomly chosen day. [1]

Model solution

(a) For a Poisson model, events should occur independently and at a constant average rate over the interval.

Helicopters pass within sight independently of one another, and they pass at a constant average rate. **(2)**

(b) The required event is $2 \leq H \leq 4$, so subtract the probabilities of 0 or 1 from the cumulative probability up to 4.

$$P(2 \leq H \leq 4) = P(H \leq 4) - P(H \leq 1) = 0.9639 - 0.4628 = 0.5008.$$

$P(2 \leq H \leq 4) = 0.501$ *(to 3 significant figures).* **(2)**

(c) The mean number in t hours is proportional to time. The rate is $1.8/12 = 0.15$ per hour, so the mean is $0.15t$. For a Poisson variable, $P(H = 0) = e^{-\lambda}$.

$$e^{-0.15t} = 0.95$$

$$-0.15t = \ln(0.95), \qquad t = \frac{\ln(0.95)}{-0.15} = 0.34196\ldots$$

$t = 0.342$ *hours, or approximately* 20.5 *minutes.* **(3)**

(d) The interval is short, so an unusual outcome can occur by chance. One observation of two helicopters is not, by itself, enough to invalidate the model.

No, not necessarily: two helicopters could pass within sight by chance in this short time interval. *(1)*

(e) A Poisson model requires random independent occurrences. Here the aeroplanes pass at a fixed, regular rate rather than randomly.

No. The aeroplanes do not pass within sight randomly or independently, since their rate is fixed and constant. *(1)*

(f) In 12 hours there are $12 \times 60/4 = 180$ *expected aeroplanes. Expected numbers add, so add the helicopter mean* 1.8*.*

$$E(\textit{total}) = 180 + 1.8 = 181.8.$$

181.8 *(1)*

www.mathswithmatt.co.uk

Question 3

Question 3

(a) It is required to test whether there is a relationship between two numerical variables.
State why in certain circumstances it would be better to use a test based on Spearman's rank correlation coefficient rather than a test based on the product-moment correlation coefficient. [1]

(b) A publisher wants to test whether there is any association between the ratings given by two readers, A and B, to books. The publisher selects 7 books at random and gives a copy of each book to each reader. Each reader gives a rating out of 20, where 20 is high and 0 is low, to each of the 7 books. The ratings are shown in the table.

Book	P	Q	R	S	T	U	V
Reader A	3	7	8	12	13	18	19
Reader B	17	15	16	14	4	5	13

(i) In the Printed Answer Booklet, complete the table giving the ranks of the ratings given by Reader A and Reader B.

Book	P	Q	R	S	T	U	V
Reader A	1	2					
Reader B							

[2]

(ii) Use Spearman's rank correlation coefficient to test at the 5% significance level whether there is any association between the readers' ratings of books. [7]

Model solution

(a) Spearman's rank test detects monotonic association, whereas the product-moment correlation test is specifically for linear correlation and has distributional assumptions.

Spearman's test is preferable when the relationship may be monotonic but not linear (or when the conditions for a product-moment correlation test are not met). (1)

(b)(i) Rank each reader's ratings from lowest, rank 1, to highest, rank 7.

Book	P	Q	R	S	T	U	V
Reader A rank	1	2	3	4	5	6	7
Reader B rank	7	5	6	4	1	2	3

The completed ranks are as shown. (2)

(b)(ii) The question asks about any association, so use a two-tailed Spearman test. Let ρ_s denote the population Spearman rank correlation coefficient for the readers' rankings of books.

$$H_0 : \rho_s = 0, \qquad H_1 : \rho_s \neq 0.$$

The rank differences are $-6, -3, -3, 0, 4, 4, 4$, giving the following calculation.

$$\sum d^2 = 6^2 + 3^2 + 3^2 + 0^2 + 4^2 + 4^2 + 4^2 = 102,$$

$$r_s = 1 - \frac{6\sum d^2}{n(n^2-1)} = 1 - \frac{6(102)}{7(7^2-1)} = -0.8214\ldots = -\frac{23}{28}.$$

For $n = 7$ at the 5% two-tailed level, the critical value is 0.7857. Since $|r_s| = 0.8214 > 0.7857$, the result is significant.

Reject H_0. There is significant evidence of an association between the readers' ratings of books. *(7)*

www.mathswithmatt.co.uk

Question 4

Question 4

In a science experiment, the values of a variable W are selected by the scientist. The experiment involves measuring the corresponding values of the variable V. The results are summarised as follows.

$$n = 10, \quad \sum v = 176.2, \quad \sum w = 87.5, \quad \sum v^2 = 4077.64, \quad \sum w^2 = 951.25, \quad \sum vw = 1963.4.$$

(a) Calculate the value of the product-moment correlation coefficient. [2]
(b) By identifying the controlled variable in the experiment, calculate the equation of the appropriate least squares regression line. [3]
It is required to obtain an estimate of the value of the response variable when the value of the controlled variable is 10.0.
(c) (i) Find the value of the required estimate. [1]
(ii) Use the appropriate mean and standard deviation to determine whether 10.0 is likely to be within the range of the data for the controlled variable. [2]
(iii) Explain whether the estimate found in part (c)(i) is likely to be reliable. [2]
(d) Explain what is meant by the words "least squares" in the phrase "least squares regression line". [1]

Model solution

(a) First calculate the corrected sums of squares and products, then use the product-moment correlation coefficient formula.

$$S_{vv} = 4077.64 - \frac{176.2^2}{10} = 972.996,$$

$$S_{ww} = 951.25 - \frac{87.5^2}{10} = 185.625, \qquad S_{vw} = 1963.4 - \frac{176.2(87.5)}{10} = 421.65.$$

$$r = \frac{S_{vw}}{\sqrt{S_{vv}S_{ww}}} = \frac{421.65}{\sqrt{972.996(185.625)}} = 0.992152\ldots$$

$r = 0.992$ ***(to 3 significant figures).*** ***(2)***

(b) The scientist selects W, so W is the controlled (independent) variable and V is the response. Therefore regress v on w.

$$b = \frac{S_{vw}}{S_{ww}} = \frac{421.65}{185.625} = 2.2715\ldots,$$

$$\bar{v} = 17.62, \qquad \bar{w} = 8.75,$$

$$a = \bar{v} - b\bar{w} = 17.62 - (2.2715\ldots)(8.75) = -2.255\ldots$$

$v = 2.2715w - 2.255$ ***(3)***

(c)(i) Substitute the controlled value $w = 10.0$ into the regression line.

$$v = 2.2715\ldots(10.0) - 2.255\ldots = 20.460\ldots$$

The estimate of v is 20.5. ***(1)***

(c)(ii) Check the selected variable W, not the response variable. Its mean and standard deviation

are

$$\bar{w} = \frac{87.5}{10} = 8.75, \qquad s_w = \sqrt{\frac{185.625}{10}} = 4.31 \text{ (approximately)}.$$

$$8.75 - 2(4.31) < 10.0 < 8.75 + 2(4.31).$$

Yes. 10.0 is within two standard deviations of the mean of W, so it is likely to be within the range of the W-data. **(2)**

(c)(iii) An estimate is more reliable when it is interpolation, rather than extrapolation, and when the correlation is strong. Both conditions hold here.

Yes, it is likely to be reliable: 10.0 is likely to be within the range of the controlled-variable data and $r = 0.992$ is close to 1. **(2)**

(d) The line is chosen by considering the residuals, the vertical differences between observed values and the line's predicted values.

It is the line whose coefficients minimise the sum of the squares of the vertical deviations (residuals) of the data points from the line. **(1)**

www.mathswithmatt.co.uk

Question 5

Question 5

The probability distribution of a discrete random variable U is given in the table.

u	0	1	2	3
$P(U=u)$	r	0.3	p	$0.5-p$

(a) Find the value of the constant r. [1]
(b) Determine an expression for $\text{Var}(U)$ in terms of p. Give your answer in the form $a-(p+b)^2$, where a and b are constants. [4]
(c) Use your expression in part (b) to determine the smallest possible value of $\text{Var}(U)$. [3]
(d) Find the value of $E(U)$ that corresponds to the smallest possible value of $\text{Var}(U)$. [1]

Model solution

(a) The probabilities in a distribution sum to 1.

$$r+0.3+p+(0.5-p)=1, \qquad r=0.2.$$

$r=0.2$ *(1)*

(b) Calculate $E(U)$ and $E(U^2)$, then use $\text{Var}(U)=E(U^2)-[E(U)]^2$.

$$E(U)=0(0.2)+1(0.3)+2p+3(0.5-p)=1.8-p,$$

$$E(U^2)=0^2(0.2)+1^2(0.3)+2^2p+3^2(0.5-p)=4.8-5p.$$

$$\begin{aligned}\text{Var}(U)&=4.8-5p-(1.8-p)^2\\&=1.56-1.4p-p^2\\&=2.05-(p+0.7)^2.\end{aligned}$$

$\text{Var}(U)=2.05-(p+0.7)^2$ *(4)*

(c) Since $p\geq 0$ and $0.5-p\geq 0$, $0\leq p\leq 0.5$. To make $2.05-(p+0.7)^2$ as small as possible, make the squared term as large as possible. This occurs at $p=0.5$.

$$\text{Var}(U)_{\text{min}}=2.05-(0.5+0.7)^2=2.05-1.44=0.61.$$

The smallest possible variance is 0.61. *(3)*

(d) At the minimum variance, $p=0.5$. Substitute this into the expectation found in part (b).

$$E(U)=1.8-0.5=1.3.$$

$E(U)=1.3$ *(1)*

Question 6

Question 6

A spinner has probability 0.2 of landing on an edge coloured blue. The spinner is spun repeatedly. The number of spins required, up to and including the first spin on which the spinner lands on an edge coloured blue, is denoted by X.

(a) Find $P(X \leq 9)$. [3]

(b) Determine the value of $E(X^2)$. [4]

(c) In this question you must show detailed reasoning.

The spinner is spun n times. The probability that X takes the value n is $0.041\,943$ correct to 6 decimal places. In these n spins, the number of times the spinner lands on an edge coloured blue is denoted by Y.

Find the variance of Y. [6]

Model solution

(a) Each spin is independent and the first blue edge is a success with probability 0.2, so X has a geometric distribution. It is simpler to use the complement: $X > 9$ means that the first nine spins are all non-blue.

$$P(X \leq 9) = 1 - P(X > 9) = 1 - 0.8^9 = 0.865782272.$$

$P(X \leq 9) = 0.866$ ***(to 3 significant figures).*** ***(3)***

(b) For a geometric random variable with success probability $p = 0.2$, find the mean and variance first, then use $E(X^2) = \mathrm{Var}(X) + [E(X)]^2$.

$$E(X) = \frac{1}{0.2} = 5, \qquad \mathrm{Var}(X) = \frac{1 - 0.2}{0.2^2} = 20.$$

$$E(X^2) = 20 + 5^2 = 45.$$

$E(X^2) = 45$ ***(4)***

(c) For the first blue result to occur on spin n, the first $n - 1$ spins must be non-blue and spin n must be blue. Thus

$$0.2(0.8)^{n-1} = 0.041943.$$

$$(0.8)^{n-1} = \frac{0.041943}{0.2} = 0.209715,$$

$$n - 1 = \frac{\ln(0.209715)}{\ln(0.8)} = 7.00\ldots, \qquad n = 8.$$

In eight independent spins, Y counts the number of blue results. Hence $Y \sim \mathrm{B}(8, 0.2)$, and the binomial variance is $np(1-p)$.

$$\mathrm{Var}(Y) = 8(0.2)(0.8) = 1.28.$$

$\mathrm{Var}(Y) = 1.28$ ***(6)***