



Edexcel AS Further Mathematics

8FM0/23 - Further Statistics 1

Friday 17 May 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.

Contents

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

www.mathswithmatt.co.uk

Question 1

Question 1

Sharma samples 100 people who play computer games. Their preferred game is recorded by age range.

Age range	A	B	C	Total
< 20	8	15	6	29
20–30	21	12	9	42
> 30	6	10	13	29
Total	35	37	28	100

- (a) Write hypotheses for a suitable test of Sharma's belief. (1)
- (b) Calculate the expected frequencies for (i) under-20 players preferring game C and (ii) 20–30-year-olds preferring game A. (2)
- (c) State the degrees of freedom. (1)
- The test statistic is 11.542. (d) At the 5% significance level, state the critical value and comment on Sharma's belief. (2)

Model solution

(a) The two variables are age range and preferred game, so a chi-squared test of association is appropriate. The null hypothesis represents Sharma's belief.

H_0 : there is no association between age range and preferred game, H_1 : there is an association between age range and preferred game.

H_0 : no association; H_1 : an association between age range and preferred game.

(1)

(b) Under H_0 , an expected frequency is found from $\frac{(\text{row total})(\text{column total})}{\text{grand total}}$. This is the count expected if the two classifications are independent.

$$E_{<20,C} = \frac{29 \times 28}{100} = 8.12, \quad E_{20-30,A} = \frac{42 \times 35}{100} = 14.7.$$

(i) 8.12, (ii) 14.7.

(2)

(c) There are three age rows and three game columns. No pooling is needed because the expected frequencies are sufficiently large, so $df = (r - 1)(c - 1)$.

$$df = (3 - 1)(3 - 1) = 4.$$

4.

(1)

(d) This is a right-tailed chi-squared test because large discrepancies between observed and expected frequencies count against H_0 .

$$\chi^2_{4,0.05} = 9.488.$$

Since $11.542 > 9.488$, reject H_0 . There is significant evidence of an association between age range and preferred game, so Sharma's belief is not supported.

Reject H_0 ; Sharma's belief is not supported at the 5% level.

(2)

www.mathswithmatt.co.uk

Question 2

Question 2

Accidents in a canteen occur randomly at an average rate of 2.7 per month.

- (a) Give a reason why a Poisson distribution could be suitable. (1)
- (b) Assuming a Poisson model, find the probability of (i) at least 3 accidents next month, (ii) no more than 10 accidents in 3 months, and (iii) at least 2 months with no accidents in 8 months. (7)
- (c) Two staff members report the same collision. State a Poisson property relevant to recording it, and whether it should be one or two accidents. (1)
- After new procedures, 22 accidents occur in 12 months. (d) At the 5% level, test the manager's claim that the monthly accident rate has reduced, stating hypotheses and the p-value. (4)

Model solution

(a) A Poisson model counts random events in a fixed time interval. It is suitable here because accidents occur randomly and independently at a constant average rate.

Accidents occur randomly and independently at a constant average rate.

(1)

(b)(i) Let A be the number of accidents in the next month. The given monthly mean gives $A \sim \text{Po}(2.7)$. Use the complement because a Poisson table/calculator gives the lower tail directly.

$$P(A \geq 3) = 1 - P(A \leq 2) = 1 - 0.49362 \dots = 0.50637 \dots$$

$$P(A \geq 3) = 0.506(3\text{s.f.}).$$

(1)

(b)(ii) Means add over independent time intervals, so in three months the Poisson mean is $3(2.7) = 8.1$.

$$T \sim \text{Po}(8.1), \quad P(T \leq 10) = 0.805837 \dots$$

$$0.806(3\text{s.f.}).$$

(2)

(b)(iii) First find the probability that one month has no accidents: $P(A = 0) = e^{-2.7}$. Across eight independent months, the number M of zero-accident months is therefore binomial.

$$M \sim B(8, e^{-2.7}) = B(8, 0.067205 \dots).$$

Use the complement of zero or one such month to obtain "at least two".

$$P(M \geq 2) = 1 - P(M \leq 1) = 1 - 0.903542 \dots = 0.096457 \dots$$

$$0.0965(3\text{s.f.}).$$

(4)

(c) For a Poisson model, events must occur singly (and independently). The two reports describe one collision rather than two separate events.

Events must occur singly; record this as one accident.

(1)

(d) The claim is a reduction, so this is a lower-tailed Poisson test. Under the original rate, the expected annual total is $12 \times 2.7 = 32.4$.

$$H_0 : \lambda = 2.7 \text{ accidents per month}, \quad H_1 : \lambda < 2.7 \text{ accidents per month.}$$

$$Y = \text{number of accidents in 12 months} \sim \text{Po}(32.4) \quad \text{under } H_0.$$

The observed total is 22, so the appropriate one-tailed p -value is the probability of 22 or fewer accidents.

$$p = P(Y \leq 22) = 0.03512 \dots$$

Since $0.03512 < 0.05$, reject H_0 . There is sufficient evidence that the average number of accidents per month has decreased.

$p = 0.0351$; reject H_0 . The manager's claim is supported.

(4)

Question 3

Question 3

The discrete random variable X has distribution

x	-1	0	1	3	7
$P(X=x)$	p	r	p	0.3	r

Given that $E(X) = 1.95$, find the exact value of $E(\sqrt{X+1})$, in the form $a + b\sqrt{2}$, where a and b are rational. (6)

Model solution

Use the given expectation first: it supplies one equation involving r . Then use the fact that all probabilities sum to one to find p .

$$E(X) = (-1)p + 0r + 1p + 3(0.3) + 7r = 0.9 + 7r.$$

$$0.9 + 7r = 1.95 \Rightarrow 7r = 1.05 \Rightarrow r = 0.15.$$

$$2p + 0.3 + 2r = 1 \Rightarrow 2p + 0.3 + 0.3 = 1 \Rightarrow p = 0.2.$$

Now apply the transformation $Y = \sqrt{X+1}$. Each possible X -value is transformed before multiplying by its probability.

X	-1	0	1	3	7
$Y = \sqrt{X+1}$	0	1	$\sqrt{2}$	2	$2\sqrt{2}$
P	0.2	0.15	0.2	0.3	0.15

$$\begin{aligned} E(Y) &= 0(0.2) + 1(0.15) + \sqrt{2}(0.2) + 2(0.3) + 2\sqrt{2}(0.15) \\ &= 0.75 + 0.5\sqrt{2}. \end{aligned}$$

$$E(\sqrt{X+1}) = \frac{3}{4} + \frac{1}{2}\sqrt{2}.$$

(6)

Question 4

Question 4

Robin shoots 8 arrows each day for 100 days. The daily number of hits is

Number of hits	0	1	2	3	4	5	6	7	8
Frequency	1	10	30	34	17	4	2	0	2

Misha proposes a binomial model.

- State, in context, two assumptions implied by this model. (2)
- Estimate the proportion of arrows that hit the target. (2)
- The expected frequencies for 0 to 8 hits are 2.81, 12.67, r , 28.05, 19.73, s , 2.50, 0.40, 0.03. Find r and s . (3)
- (i) Explain why the goodness-of-fit test has 3 degrees of freedom. (ii) Complete a 5% test, stating hypotheses, test statistic, critical value and conclusion. (8)

Model solution

(a) A binomial distribution models a fixed number of independent trials with the same success probability. Here each arrow is a trial and a hit is a success.

Each arrow has the same probability of hitting the target, and the arrows are shot independently.

(2)

(b) Estimate the hit probability by $\hat{p} = \frac{\text{total hits}}{\text{total arrows}}$. There are 100×8 arrows in total.

$$\begin{aligned} \text{total hits} &= 0(1) + 1(10) + 2(30) + 3(34) + 4(17) + 5(4) + 6(2) + 7(0) + 8(2) \\ &= 288, \end{aligned}$$

$$\hat{p} = \frac{288}{100 \times 8} = 0.36.$$

$$\hat{p} = 0.36.$$

(2)

(c) With eight arrows per day and estimated hit probability 0.36, the model is $X \sim B(8, 0.36)$. An expected frequency is $100P(X = x)$.

$$r = 100 \binom{8}{2} (0.36)^2 (0.64)^6 = 24.936 \dots = 24.94.$$

$$s = 100 \binom{8}{5} (0.36)^5 (0.64)^3 = 8.876 \dots = 8.88.$$

$$r = 24.94, s = 8.88(\text{to 2 d.p.}).$$

(3)

(d)(i) A chi-squared goodness-of-fit test requires every expected frequency after pooling to exceed 5. Pooling gives five groups: 0 or 1, 2, 3, 4, and 5--8. There are two constraints: the expected frequencies total 100 and p was estimated from the data.

$$df = 5 - 1 - 1 = 3.$$

Pool the tail to make expected frequencies exceed 5; $df = 5 - 2 = 3$.

(2)

(d)(ii) Use a chi-squared goodness-of-fit test because observed daily-hit frequencies are being compared with those predicted by $B(8, 0.36)$.

$H_0 : B(8, 0.36)$ is a suitable model for the data, $H_1 : B(8, 0.36)$ is not a suitable model for the data.

Pool the low expected-frequency tail categories before calculating the statistic. The final expected values below use unrounded binomial probabilities.

Hits	0 or 1	2	3	4	5--8
O	11	30	34	17	8
E	15.48	24.94	28.05	19.73	11.80

$$\begin{aligned}
 \chi^2 &= \sum \frac{(O - E)^2}{E} \\
 &= \frac{(11 - 15.48)^2}{15.48} + \frac{(30 - 24.94)^2}{24.94} + \frac{(34 - 28.05)^2}{28.05} \\
 &\quad + \frac{(17 - 19.73)^2}{19.73} + \frac{(8 - 11.80)^2}{11.80} \\
 &\approx 5.19.
 \end{aligned}$$

$$\chi_{3,0.05}^2 = 7.815.$$

Since $5.19 < 7.815$, do not reject H_0 . The data are compatible with Misha's belief, so a binomial model is suitable at the 5% level.

$\chi^2 \approx 5.19 < 7.815$: do not reject H_0 ; $B(8, 0.36)$ is suitable.

(6)