



## Edexcel AS Further Mathematics

### 8FM0/23 - Further Statistics 1

Friday 19 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 | 3 |
| Question 3 | 5 |
| Question 4 | 7 |

[www.mathswithmatt.co.uk](http://www.mathswithmatt.co.uk)

## Question 1

### Question 1

The discrete random variable  $X$  has distribution

| $x$        | 0   | 1   | 2             | 3             | 4             |
|------------|-----|-----|---------------|---------------|---------------|
| $P(X = x)$ | $r$ | $k$ | $\frac{k}{2}$ | $\frac{k}{3}$ | $\frac{k}{4}$ |

where  $r$  and  $k$  are positive constants. The standard deviation of  $X$  equals the mean of  $X$ . Find the exact value of  $r$ .

### Model solution

*We need both the mean and variance: the condition involving the standard deviation can then be converted into an equation for  $k$ .*

$$E(X) = 0r + 1(k) + 2\left(\frac{k}{2}\right) + 3\left(\frac{k}{3}\right) + 4\left(\frac{k}{4}\right) = 4k.$$

$$E(X^2) = 0^2r + 1^2k + 2^2\left(\frac{k}{2}\right) + 3^2\left(\frac{k}{3}\right) + 4^2\left(\frac{k}{4}\right) = 10k.$$

*Thus the variance is found from  $\text{Var}(X) = E(X^2) - [E(X)]^2$ .*

$$\text{Var}(X) = 10k - (4k)^2 = 10k - 16k^2.$$

*Since the standard deviation equals the mean, their squares are equal. Hence  $\text{Var}(X) = [E(X)]^2$ .*

$$10k - 16k^2 = 16k^2 \Rightarrow 10k = 32k^2 \Rightarrow k = \frac{5}{16},$$

*where  $k = 0$  is rejected because  $k$  is positive. The probabilities must now add to 1.*

$$r + k + \frac{k}{2} + \frac{k}{3} + \frac{k}{4} = 1$$

$$r + \frac{25k}{12} = 1 \Rightarrow r = 1 - \frac{25}{12}\left(\frac{5}{16}\right) = \frac{67}{192}.$$

$$r = \frac{67}{192}.$$

(6)

## Question 2

### Question 2

A bag contains a large number of equal-sized, equal-weight balls coloured Red, Blue or Yellow. Each of 150 children, with eyes closed, selects a ball, shows it to Jasmine, replaces it, and then selects a second ball. A prize is given when the two balls have the same colour. The results are

| 2nd colour \ 1st colour | Red | Blue | Yellow | Total |
|-------------------------|-----|------|--------|-------|
| Red                     | 31  | 11   | 18     | 60    |
| Blue                    | 8   | 10   | 9      | 27    |
| Yellow                  | 21  | 9    | 33     | 63    |
| Total                   | 60  | 30   | 60     | 150   |

Jasmine tests, at the 5% level, whether the colour of the second ball is independent of the colour of the first ball.

(a) Calculate the expected frequencies where both balls are the same colour.

The test statistic is 12.712.

(b) Use this value to complete the test, stating the critical value and conclusion clearly.

(c) With reference to part (a) and to the experiment, give a plausible reason for the conclusion in part (b).

### Model solution

(a) Under independence, an expected cell frequency is  $\frac{(\text{row total})(\text{column total})}{\text{grand total}}$ . We apply this to the three diagonal cells, which represent matching colours.

$$E(RR) = \frac{60 \times 60}{150} = 24, \quad E(BB) = \frac{27 \times 30}{150} = 5.4, \quad E(YY) = \frac{63 \times 60}{150} = 25.2.$$

$$E(RR) = 24, E(BB) = 5.4, E(YY) = 25.2.$$

(2)

(b) This is a  $3 \times 3$  independence table, so the degrees of freedom are  $(3 - 1)(3 - 1) = 4$ . All expected frequencies are at least 5, so no categories need pooling.

$H_0$  : the colours of the first and second balls are independent,

$H_1$  : the colours of the first and second balls are not independent.

The test is upper-tailed because a large chi-squared statistic indicates a poor fit to independence. For 4 degrees of freedom at 5%,

$$\chi_{0.95,4}^2 = 9.488.$$

$$12.712 > 9.488,$$

so the statistic lies in the critical region and  $H_0$  is rejected.

*There is significant evidence that the colour of the second ball is not independent of the colour of the first ball.*

(3)

(c) *The observed matching-colour frequencies are all larger than their expected values:  $31 > 24$ ,  $10 > 5.4$  and  $33 > 25.2$ . This suggests that the second selection was more likely to match the first than independence predicts.*

*For example, some children may have looked when choosing the second ball in order to win a prize, so their second choice was not independent of the first.*

(1)

www.mathswithmatt.co.uk

## Question 3

### Question 3

A machine produces cloth in which faults occur randomly at a rate of 0.4 per square metre. Tablecloths have area  $A \text{ m}^2$ , and the probability that a randomly selected tablecloth has no faults is 0.0907.

(a) Find  $A$ .

Tablecloths are sold in packets of 20.

(b) Find the probability that more than one tablecloth in a randomly selected packet has no faults.

A hotel orders 100 tablecloths of area  $A \text{ m}^2$ . Let  $X$  be the number with no faults.

(c) Find (i)  $E(X)$  (ii)  $\text{Var}(X)$ .

(d) Use a Poisson approximation to estimate  $P(X = 10)$ .

A new machine is claimed to produce faults at a rate less than 0.4 per square metre. A random  $30 \text{ m}^2$  piece of its cloth has 6 faults.

(e) Stating hypotheses clearly, test the claim at the 5% level.

(f) Write down the p-value for the test in part (e).

### Model solution

(a) Random faults at a constant rate over an area are modelled by a Poisson distribution. If  $F$  is the number of faults in one tablecloth, then  $F \sim \text{Po}(0.4A)$ .

$$P(F = 0) = e^{-0.4A} = 0.0907.$$

$$-0.4A = \ln(0.0907) \Rightarrow A = \frac{-\ln(0.0907)}{0.4} = 6.$$

$$A = 6 \text{ m}^2.$$

(3)

(b) Each tablecloth independently has probability 0.0907 of having no faults. Thus, for  $T$  = number with no faults in the packet,  $T \sim B(20, 0.0907)$ . "More than 1" means  $T \geq 2$ , so it is shortest to use the complement.

$$P(T > 1) = 1 - P(T \leq 1) = 1 - \left[ \binom{20}{0} (0.0907)^0 (0.9093)^{20} + \binom{20}{1} (0.0907) (0.9093)^{19} \right] = 0.55276 \dots$$

$$P(T > 1) = 0.553(3\text{s.f.}).$$

(3)

(c) Each of the 100 tablecloths has the same probability 0.0907 of being fault-free, so  $X \sim B(100, 0.0907)$ . For a binomial variable,  $E(X) = np$  and  $\text{Var}(X) = np(1 - p)$ .

$$(i) \quad E(X) = 100(0.0907) = 9.07,$$

$$(ii) \quad \text{Var}(X) = 100(0.0907)(1 - 0.0907) = 8.247351.$$

$$E(X) = 9.07; \text{Var}(X) = 8.25(3\text{s.f.}).$$

(3)

(d) Here  $n = 100$  is large and  $p = 0.0907$  is small, so the binomial distribution is approximated by  $\text{Po}(np) = \text{Po}(9.07)$ .

$$P(X = 10) \approx e^{-9.07} \frac{9.07^{10}}{10!} = 0.11947 \dots$$

$$P(X = 10) \approx 0.1195.$$

(2)

(e) The claim is a lower rate, so the alternative hypothesis is lower-tailed. Under  $H_0$ , a  $30 \text{ m}^2$  piece has mean number of faults  $30(0.4) = 12$ .

$$H_0 : \lambda = 0.4 \text{ faults per m}^2, \quad H_1 : \lambda < 0.4 \text{ faults per m}^2.$$

$$Y = \text{number of faults in } 30 \text{ m}^2 \sim \text{Po}(12) \text{ under } H_0.$$

Six faults is evidence for the claim only if it is unusually small, so calculate the lower-tail probability.

$$P(Y \leq 6) = \sum_{y=0}^6 e^{-12} \frac{12^y}{y!} = 0.04582 \dots$$

This is below 0.05, so reject  $H_0$ .

There is significant evidence at the 5% level to support the claim that the new machine

(4)

(f) The  $p$ -value is the probability, assuming  $H_0$ , of observing 6 or fewer faults: the result at least as supportive of the lower-rate claim as the observed result.

$$p\text{-value} = 0.04582 \dots \approx 0.0458.$$

(1)

## Question 4

### Question 4

In 60 months, the numbers of car breakdowns in the Snoreap district were

| Number of breakdowns | 0  | 1  | 2  | 3  | 4 | 5 |
|----------------------|----|----|----|----|---|---|
| Frequency            | 12 | 11 | 19 | 14 | 3 | 1 |

Anja believes the number of breakdowns per month can be modelled by a Poisson distribution.

- State suitable hypotheses to investigate this belief.
- Explain why the final class is labelled  $\geq 5$  when calculating expected frequencies.
- Complete the expected-frequency row:

| Number of breakdowns         | 0    | 1  | 2  | 3    | 4    | $\geq 5$ |
|------------------------------|------|----|----|------|------|----------|
| Observed frequency ( $O_i$ ) | 12   | 11 | 19 | 14   | 3    | 1        |
| Expected frequency ( $E_i$ ) | 9.92 |    |    | 9.64 | 4.34 |          |

- Find  $\frac{(O_i - E_i)^2}{E_i}$  when the number of breakdowns is (i) 1 (ii) 3.

- Explain why 3 degrees of freedom are used.

The test statistic is 6.54.

- At the 5% level, state the critical value and complete the test.

### Model solution

(a) A goodness-of-fit test compares the observed distribution with the proposed Poisson model.

$H_0$  : the number of car breakdowns per month follows a Poisson distribution,

$H_1$  : the number of car breakdowns per month does not follow a Poisson distribution.

$H_0$  : monthly car breakdowns follow a Poisson distribution;  $H_1$  : they do not.

(1)

(b) A Poisson random variable can take any non-negative integer value. Although no month in the sample had more than 5 breakdowns, the model assigns probability to values above 5, so they must be included in the final expected-frequency class.

The class must include all values 5, 6, 7, ..., because a Poisson distribution assigns

(1)

(c) First estimate the Poisson parameter by the sample mean. This is the appropriate parameter estimate for a Poisson goodness-of-fit test.

$$\hat{\lambda} = \frac{0(12) + 1(11) + 2(19) + 3(14) + 4(3) + 5(1)}{60} = \frac{108}{60} = 1.8.$$

Under  $H_0$ ,  $X \sim \text{Po}(1.8)$ , so  $E_i = 60P(X = i)$ .

| Number of breakdowns         | 0            | 1                 | 2                            | 3                            | 4                            | $\geq 5$        |
|------------------------------|--------------|-------------------|------------------------------|------------------------------|------------------------------|-----------------|
| Expected frequency ( $E_i$ ) | $60e^{-1.8}$ | $60e^{-1.8}(1.8)$ | $60e^{-1.8}\frac{1.8^2}{2!}$ | $60e^{-1.8}\frac{1.8^3}{3!}$ | $60e^{-1.8}\frac{1.8^4}{4!}$ | $60P(X \geq 5)$ |
|                              | 9.92         | 17.85             | 16.07                        | 9.64                         | 4.34                         | 2.18            |

For the final entry,  $60P(X \geq 5)$  can be found as  $60 - (9.92 + 17.85 + 16.07 + 9.64 + 4.34)$  before rounding.

The missing expected frequencies are 17.85, 16.07 and 2.18; the complete row is 9.92,

(4)

(d) Substitute the observed and expected frequencies for each specified class into the corresponding contribution to the chi-squared statistic.

$$(i) \frac{(11 - 17.8523 \dots)^2}{17.8523 \dots} = 2.630 \dots$$

$$(ii) \frac{(14 - 9.6402 \dots)^2}{9.6402 \dots} = 1.971 \dots$$

(i) 2.63; (ii) 1.97 (both to 3 s.f.).

(2)

(e) The expected frequencies for 4 and  $\geq 5$  are below 5, so those two classes are pooled. This leaves 5 classes: 0, 1, 2, 3,  $\geq 4$ . One degree of freedom is lost because the frequencies total 60, and one more is lost because  $\lambda$  was estimated from the data.

$$\text{degrees of freedom} = 5 - 1 - 1 = 3.$$

There are 3 degrees of freedom.

(2)

(f) A goodness-of-fit test has a large-value critical region. With 3 degrees of freedom, the 5% critical value is

$$\chi_{0.95,3}^2 = 7.815.$$

$$6.54 < 7.815,$$

so the statistic is not in the critical region and we do not reject  $H_0$ .

There is insufficient evidence at the 5% level to reject Anja's belief that monthly breakdowns are Poisson distributed.

(2)