



Edexcel AS Further Mathematics

8FM0/23 - Further Statistics 1

November 2021

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 2021 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	8

www.mathswithmatt.co.uk

Question 1

Question 1

Flobee sells tomato seeds in packets of 40 and advertises that 4% do not germinate. Amodita plants 125 packets and obtains

Number not germinating	0	1	2	3	4	5	6 or more
Frequency	15	35	38	22	10	5	0

For the model $X \sim B(40, 0.04)$, the expected frequencies are

Number not germinating	0	1	2	3	4	5 or more
Expected frequency	24.42	40.70	r	17.45	6.73	s

- (a) Find r and s. (2)
 (b) At the 5% level, test whether $B(40, 0.04)$ is a suitable model. State hypotheses, degrees of freedom, critical value and conclusion. (6)
 (c) Using a new model $B(40, p)$, suggest a more realistic value of p, showing your working. (2)

Model solution

(a) The expected frequency for a particular outcome is the model probability multiplied by the 125 packets. Here r corresponds to exactly two seeds failing to germinate.

$$\begin{aligned} r &= 125 P(X = 2) \\ &= 125 \binom{40}{2} (0.04)^2 (0.96)^{38} \\ &= 33.0698 \dots \end{aligned}$$

For s, all outcomes from five upwards belong in the final class, so it is most efficient to use a complement.

$$\begin{aligned} s &= 125 P(X \geq 5) \\ &= 125 [1 - P(X \leq 4)] \\ &= 2.6278 \dots \end{aligned}$$

$$r = 33.07 \text{ and } s = 2.63, \text{ to 2 decimal places.}$$

(2)

(b) This is a chi-squared goodness-of-fit test, because observed frequencies are being compared with the frequencies predicted by one stated distribution.

H_0 : $X \sim B(40, 0.04)$ is a suitable model for the number of seeds that do not germinate,
 H_1 : $X \sim B(40, 0.04)$ is not a suitable model.

The expected frequencies in the last two classes are below 5, so combine them before calculating the test statistic. Their observed frequency is $10 + 5 = 15$, and their expected

frequency is $6.73 + 2.63 = 9.36$.

$$\begin{aligned}\chi^2 &= \frac{(15 - 24.42)^2}{24.42} + \frac{(35 - 40.70)^2}{40.70} + \frac{(38 - 33.07)^2}{33.07} \\ &\quad + \frac{(22 - 17.45)^2}{17.45} + \frac{(15 - 9.36)^2}{9.36} \\ &= 9.752 \dots\end{aligned}$$

There are now five classes. No model parameters have been estimated from the data for this test, so the degrees of freedom are $5 - 1 = 4$.

$$\chi^2 \sim \chi^2_4, \quad \text{critical value at the 5\% level} = 9.488.$$

Since $9.752 > 9.488$, the statistic is in the critical region.

Reject H_0 . At the 5% level there is significant evidence that $B(40, 0.04)$ is not a suitable model.

(6)

(c) For a binomial model, p is estimated by the total number of failures divided by the total number of seeds observed.

$$\begin{aligned}\hat{p} &= \frac{0(15) + 1(35) + 2(38) + 3(22) + 4(10) + 5(5)}{125 \times 40} \\ &= \frac{242}{5000} = 0.0484.\end{aligned}$$

A more realistic value is $p = 0.0484$ (about 4.84%).

(2)

Question 2

Question 2

In a 30-minute period, $R \sim \text{Po}(28)$ is the number checked in by Rowan.

(a) Calculate $P(R \geq 23)$. (1)

Alex's number is $A \sim \text{Po}(16)$, independently of R . (b) Find the probability that exactly 42 passengers in total are checked in in a randomly selected 30-minute period. (2)

(c) For 150 non-overlapping 60-minute periods, use a Poisson approximation to find the probability that in at least 25 periods their combined total is fewer than 80. (4)

(d) In a random 90-minute period Rowan checks in 67 people. At the 5% level, test whether there is evidence that the system is working slower than normal. (4)

Model solution

(a) The required event is a lower-tail complement, since standard Poisson cumulative probabilities give $P(R \leq 22)$.

$$P(R \geq 23) = 1 - P(R \leq 22) = 0.8517 \dots$$

$$P(R \geq 23) = 0.852$$

(1)

(b) Independent Poisson counts over the same time interval add: their means add as well.

$$R + A \sim \text{Po}(28 + 16) = \text{Po}(44).$$

$$P(R + A = 42) = e^{-44} \frac{44^{42}}{42!} = 0.05866 \dots$$

$$0.0587$$

(2)

(c) First find the chance of a "fewer than 80" period. In 60 minutes the combined mean is twice the 30-minute mean, so the combined count has distribution $\text{Po}(88)$.

$$p = P(\text{Po}(88) < 80) = P(\text{Po}(88) \leq 79) = 0.183 \dots$$

If N is the number of such periods among 150, then $N \sim B(150, 0.183 \dots)$. The question requests a Poisson approximation, which is appropriate for counting occurrences across many independent periods; its mean is np .

$$N \approx \text{Po}(150 \times 0.183 \dots) = \text{Po}(27.48 \dots).$$

$$P(N \geq 25) = 1 - P(N \leq 24) = 1 - 0.2922 \dots = 0.7078 \dots$$

$$0.708$$

(4)

(d) A slower system means fewer check-ins, so this is a one-tailed lower-tail test. Under normal operation Rowan's mean for 90 minutes is $3 \times 28 = 84$.

$$H_0: \lambda = 84 \quad (\text{normal rate}),$$

$$H_1: \lambda < 84 \quad (\text{slower than normal}).$$

$$J \sim \text{Po}(84) \quad \text{under } H_0, \quad P(J \leq 67) = 0.03246 \dots$$

The lower-tail probability is below 0.05 (equivalently, the 5% critical region is $J \leq 68$, which contains 67).

Reject H_0 . There is evidence at the 5% level that the system is working slower than normal.

(4)

www.mathswithmatt.co.uk

Question 3

Question 3

x	-3	-2	-1	0	2	5
$P(X = x)$	0.3	0.15	0.1	0.15	0.1	0.2

(a) Find $E(X)$. (1)

Given $\text{Var}(X) = 8.79$, (b) find $E(X^2)$. (2)

y	-2	-1	0	1	2
$P(Y = y)$	$3a$	a	b	a	c

where $P(Y \leq 0) = 0.75$ and $E(Y^2 + 3) = 5$. (c) Find a , b and c . (5)

The variables X and Y are independent, $W = Y - X$, and $T = 3W - 8$. (d) Calculate $P(W > T)$. (4)

Model solution

(a) The expectation is the probability-weighted mean of all possible values.

$$E(X) = (-3)(0.3) + (-2)(0.15) + (-1)(0.1) + 0(0.15) + 2(0.1) + 5(0.2) = -0.1.$$

$$E(X) = -0.1$$

(1)

(b) Use $\text{Var}(X) = E(X^2) - [E(X)]^2$, because the variance and mean have both been supplied.

$$8.79 = E(X^2) - (-0.1)^2, \quad E(X^2) = 8.79 + 0.01 = 8.8.$$

$$E(X^2) = 8.8$$

(2)

(c) The information gives three equations. First use the expectation involving Y^2 ; the added 3 is a constant.

$$E(Y^2) = 5 - 3 = 2.$$

$$4(3a) + 1(a) + 0(b) + 1(a) + 4c = 2 \Rightarrow 7a + 2c = 1.$$

Next use the stated cumulative probability, then the fact that all probabilities sum to 1.

$$3a + a + b = 0.75 \Rightarrow 4a + b = 0.75,$$

$$3a + a + b + a + c = 1 \Rightarrow 5a + b + c = 1.$$

$$7a + 2c = 1,$$

$$4a + b = 0.75, \quad \Rightarrow \quad a = 0.1, \quad b = 0.35, \quad c = 0.15.$$

$$5a + b + c = 1$$

$$a = 0.1, b = 0.35, c = 0.15$$

(5)

(d) Begin by turning the comparison between W and T into one condition on W .

$$W > T \iff W > 3W - 8 \iff 2W < 8 \iff W < 4.$$

It is shorter to find the complement. Since $W = Y - X$ is discrete, $W \geq 4$ occurs only for $(X, Y) = (-3, 1), (-3, 2), (-2, 2)$. Independence allows the probabilities of each pair to be multiplied.

$$\begin{aligned} P(W \geq 4) &= P(X = -3)\{P(Y = 1) + P(Y = 2)\} + P(X = -2)P(Y = 2) \\ &= 0.3(0.1 + 0.15) + 0.15(0.15) \\ &= 0.0975. \end{aligned}$$

$$P(W > T) = P(W < 4) = 1 - 0.0975 = 0.9025.$$

$$0.9025$$

(4)

Question 4

Question 4

Age / main investment	Bonds	Cash	Stocks
25–44	a	b – e	e
45–75	c	d – 59	59

(a) Find, in terms of a, b, c, d, the difference $O - E$ for the group aged 45–75 whose main investment is Bonds. Give a single simplified fraction. (4)

Given $\sum \frac{(O-E)^2}{E} = 9.62$, (b) test at the 5% level whether age and main type of investment are associated. State hypotheses, critical value and conclusion. No cells need be combined. (3)

Model solution

(a) For a contingency table, an expected frequency is $\frac{\text{row total} \times \text{column total}}{\text{grand total}}$. The required cell has observed value c, row total $c + d$, column total $a + c$, and grand total $a + b + c + d$.

$$E = \frac{(c + d)(a + c)}{a + b + c + d}.$$

Subtract this expected value from the observed value, then put the result over the common denominator:

$$\begin{aligned} O - E &= c - \frac{(c + d)(a + c)}{a + b + c + d} \\ &= \frac{c(a + b + c + d) - (c + d)(a + c)}{a + b + c + d} \\ &= \frac{ac + bc + c^2 + cd - ac - c^2 - ad - cd}{a + b + c + d} \\ &= \frac{bc - ad}{a + b + c + d}. \end{aligned}$$

$$O - E = \frac{bc - ad}{a + b + c + d}.$$

(4)

(b) This is a chi-squared test of association, because the data are classified by two categorical variables.

H_0 : There is no association between age and main type of investment,

H_1 : There is an association between age and main type of investment.

The table has 2 rows and 3 columns, so the degrees of freedom are $(2 - 1)(3 - 1) = 2$.

$$\chi^2 \sim \chi^2_2, \quad 5\% \text{ critical value} = 5.991.$$

The test statistic is 9.62, which is greater than 5.991, so it lies in the critical region.

Reject H_0 . At the 5% level, there is evidence of an association between a person's age

(3)