



Edexcel AS Further Mathematics

8FM0/23 - Further Statistics 1

Thursday 17 May 2018

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 2018 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

A researcher counts the numbers of orchids in 150 randomly selected, non-overlapping 1 m² areas of a field. The observed frequencies are

number of orchids	0	1	2	3	4	5	6
frequency	30	42	35	26	11	6	0

He proposes the model $X \sim \text{Po}(1.75)$. The expected frequencies for 0, 1, 2, 3, 4, 5, and more than 5 orchids are respectively

26.07, 45.62, 39.91, 23.28, 10.19, 3.57, r .

- Find r , to 2 decimal places. (1)
- State hypotheses for a 5% test of the proposed model. (1)
- The test statistic is 2.0 and 4 degrees of freedom are to be used. Explain fully why. (2)
- State the critical value and determine whether the data support the researcher's belief. (2)
- In a second field, orchid counts per square metre have distribution $\text{Po}(1.5)$. For 200 randomly selected squares, use a Poisson approximation to show that the probability of finding at least one square containing exactly 6 orchids is 0.506 to 3 decimal places. (4)

Model solution

(a) The expected frequency is 150 times the relevant Poisson probability.

$$r = 150P(X \geq 6) = 150 \left(1 - \sum_{x=0}^5 e^{-1.75} \frac{1.75^x}{x!} \right) = 1.36 \dots$$

$$r = 1.36$$

(b) The null hypothesis says that the specified distribution models the counts.

$H_0 : X \sim \text{Po}(1.75)$ is a suitable model, $H_1 : X \sim \text{Po}(1.75)$ is not a suitable model.

(c) The classes for 4, 5 and more than 5 orchids must be combined, since the expected frequencies for 5 and more than 5 are below 5. Thus there are five classes: 0, 1, 2, 3, and ≥ 4 . One degree of freedom is lost because the frequencies total 150 (equivalently, only four class frequencies can be chosen independently); no parameter has been estimated from these data because $\lambda = 1.75$ is given. Hence $5 - 1 = 4$ degrees of freedom.

(d) For a 5% upper-tail chi-squared goodness-of-fit test with 4 degrees of freedom,

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

Since $2.0 < 9.488$, the statistic is not in the critical region, so do not reject H_0 .

The data support the researcher's belief that $\text{Po}(1.75)$ is a good model.

(e) For one square in the second field,

$$p = P(X = 6) = e^{-1.5} \frac{1.5^6}{6!} = 0.003529 \dots$$

If Y is the number, out of 200 squares, with exactly 6 orchids, then $Y \sim \text{Bin}(200, p)$. Using the Poisson approximation,

$$Y \sim \text{Po}(200p) = \text{Po}(0.7059 \dots).$$

Therefore

$$P(Y \geq 1) = 1 - P(Y = 0) = 1 - e^{-0.7059 \dots} = 0.506 \dots$$

$$P(Y \geq 1) = 0.506 \text{ (3 d.p.)}$$

www.mathswithmatt.co.uk

Question 2

Question 2

The daily numbers of heaters bought from Warmup and Pumraw supermarkets are independent random variables $H \sim \text{Po}(0.7)$ and $G \sim \text{Po}(3)$ respectively.

- (a) Calculate $P(H \geq 2)$. (1)
- (b) Show that the probability that fewer than 4 heaters in total are bought in one day is 0.494 to 3 decimal places. (2)
- (c) Find the probability that this occurs on at least 5 of 6 randomly chosen days. (3)
- (d) In December, two randomly selected days had a combined total of 14 heaters. Test, at the 5% level, whether the mean daily total has increased. State hypotheses clearly. (5)

Model solution

(a)

$$P(H \geq 2) = 1 - P(H = 0) - P(H = 1) = 1 - e^{-0.7}(1 + 0.7) = 0.1558 \dots$$

$$0.156$$

(b) The sum of independent Poisson random variables is Poisson. Let $T = H + G$.

$$T \sim \text{Po}(0.7 + 3) = \text{Po}(3.7), \quad P(T < 4) = P(T \leq 3) = \sum_{t=0}^3 e^{-3.7} \frac{3.7^t}{t!} = 0.4939 \dots$$

$$P(T < 4) = 0.494 \text{ (3 d.p.)}$$

(c) Let K be the number of the 6 days on which fewer than 4 heaters are bought. The days are independent, so $K \sim \text{Bin}(6, 0.4939 \dots)$.

$$P(K \geq 5) = 1 - P(K \leq 4) = \binom{6}{5} p^5 (1 - p) + p^6 = 0.1039 \dots, \quad p = 0.4939 \dots$$

$$0.104$$

(d) Under the original model the mean daily total is 3.7. The alternative is one-tailed because only an increase is being tested.

$$H_0 : \lambda = 3.7, \quad H_1 : \lambda > 3.7.$$

For the combined total J over two days, under H_0 ,

$$J \sim \text{Po}(2 \times 3.7) = \text{Po}(7.4).$$

The upper-tail probability for the observed total is

$$P(J \geq 14) = 1 - P(J \leq 13) = 0.0195 \dots$$

Equivalently, $P(J \geq 12) = 0.0735\dots$ whereas $P(J \geq 13) = 0.0391\dots$, so the 5% critical region is $J \geq 13$. Since $14 \geq 13$ (and $0.0195 < 0.05$), reject H_0 .

There is evidence at the 5% level that the mean total number of heaters bought from the two supermarkets has increased.

www.mathswithmatt.co.uk

Question 3

Question 3

A fair black die has faces 1, 2, 2, 3, 3, 4; its score is B . A white die has faces 1, 1, 2, 4, 5, c , where $c > 5$; its score is W , with distribution

w	1	2	4	5	c
$P(W = w)$	$a + b$	a	0.3	a	b

Greg rolls the black die and Nilaya the white die. Greg wins if his score is at least two more than Nilaya's, and $P(\text{Greg wins}) = \frac{1}{6}$. Also $X = 2W - 5$ and $E(X) = 2.6$.

- (a) Find $E(B)$. (1)
- (b) Find a and b , showing working. (5)
- (c) Find the exact value of $\text{Var}(X)$. (6)

Model solution

(a)

$$E(B) = \frac{1 + 2 + 2 + 3 + 3 + 4}{6} = \frac{5}{2}.$$

$$E(B) = 2.5$$

(b) The possible winning score pairs (B, W) are $(4, 1)$, $(4, 2)$ and $(3, 1)$. Hence

$$\frac{1}{6}(a + b) + \frac{1}{6}a + \frac{2}{6}(a + b) = \frac{1}{6},$$

so

$$4a + 3b = 1.$$

Also the probabilities in the distribution sum to 1, giving

$$(a + b) + a + 0.3 + a + b = 1, \quad 3a + 2b = 0.7.$$

Solving $4a + 3b = 1$ and $3a + 2b = 0.7$ gives

$$a = 0.1, \quad b = 0.2.$$

$$a = 0.1, \quad b = 0.2$$

(c) First use the given mean to determine c .

$$E(X) = 2E(W) - 5 = 2.6 \Rightarrow E(W) = 3.8.$$

Using $a = 0.1$ and $b = 0.2$,

$$E(W) = 1(0.3) + 2(0.1) + 4(0.3) + 5(0.1) + c(0.2) = 2.2 + 0.2c.$$

Thus $2.2 + 0.2c = 3.8$, so $c = 8$, which satisfies $c > 5$. Next,

$$E(W^2) = 1^2(0.3) + 2^2(0.1) + 4^2(0.3) + 5^2(0.1) + 8^2(0.2) = 20.8.$$

Therefore

$$\text{Var}(W) = E(W^2) - [E(W)]^2 = 20.8 - 3.8^2 = 6.36,$$

and, since $X = 2W - 5$,

$$\text{Var}(X) = 2^2 \text{Var}(W) = 4(6.36) = 25.44.$$

$$\text{Var}(X) = 25.44 = \frac{636}{25}.$$

www.mathswithmatt.co.uk

Question 4

Question 4

In a random sample of 125 plants with fungus, the outcomes for three treatments were:

outcome\treatment	No action	Sprayed once	Sprayed daily	total
Died within a month	15	16	25	56
Survived 1–6 months	8	25	10	43
Survived beyond 6 months	7	14	5	26
total	30	55	40	125

For a suitable test, seven expected frequencies are 17.92, 10.32, 18.92, 13.76, 6.24, 11.44, 8.32, and the sum of the chi-squared contributions for these seven cells is 8.29. Test at the 2.5% level whether there is an association between treatment and survival. State hypotheses and conclusion clearly. (7)

Model solution

Use a chi-squared test of association.

H_0 : there is no association between treatment and plant survival/outcome,

H_1 : there is an association between treatment and plant survival/outcome.

Under H_0 , each expected frequency is $\frac{(\text{row total})(\text{column total})}{125}$. The two missing entries in the first row are

$$E(\text{died, no action}) = \frac{56 \times 30}{125} = 13.44, \quad E(\text{died, sprayed once}) = \frac{56 \times 55}{125} = 24.64.$$

Thus the full expected-frequency table is

	No action	Sprayed once	Sprayed daily
Died within a month	13.44	24.64	17.92
Survived 1–6 months	10.32	18.92	13.76
Survived beyond 6 months	6.24	11.44	8.32

The two additional contributions are

$$\frac{(15 - 13.44)^2}{13.44} = 0.1814\dots, \quad \frac{(16 - 24.64)^2}{24.64} = 3.0296\dots$$

so

$$\chi^2 = 8.29 + 0.1814\dots + 3.0296\dots = 11.50\dots$$

There are $(3 - 1)(3 - 1) = 4$ degrees of freedom. At the 2.5% significance level,

$$\chi_{4,0.025}^2 = 11.143.$$

Since $11.50\dots > 11.143$, reject H_0 .

There is evidence at the 2.5% level of an association between the treatment of the plants and their survival/outcome.