



Edexcel AS Further Mathematics

8FM0/23 - Further Statistics 1

Tuesday 24 May 2022

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 2022 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

Stuart applies a treatment to a random sample of 400 fruit trees. The numbers of years, y , for which the trees remain free from disease are recorded.

Number of years free from disease	Treatment applied	Treatment not applied
$y < 1$	15	25
$1 \leq y < 2$	35	61
$2 \leq y$	124	140

- (a) Calculate the expected frequencies for (i) treatment applied and $y < 1$, and (ii) treatment not applied and $1 \leq y < 2$. (2)

For the other four classes, $\sum (O - E)^2/E = 2.642$, to 3 decimal places.

- (b) At the 5% level, test whether there is an association between treatment application and the number of years a tree remains disease-free. State hypotheses, test statistic, critical value and conclusion. (5)

Model solution

(a) This is a contingency table. Under the assumption of no association, an expected frequency is $\frac{(\text{row total})(\text{column total})}{\text{grand total}}$. The treatment-applied column total is 174, the untreated column total is 226, and the first two row totals are 40 and 96.

$$E(\text{applied and } y < 1) = \frac{174 \times 40}{400} = 17.4.$$

(a)(i) 17.4

$$E(\text{not applied and } 1 \leq y < 2) = \frac{226 \times 96}{400} = 54.24.$$

(a)(ii) 54.24

(2)

(b) A chi-squared test of association is appropriate because both variables are categorical. We compare the observed table with the frequencies expected if treatment and disease-free time were independent.

H_0 : there is no association between treatment application and years disease-free,

H_1 : there is an association between treatment application and years disease-free.

$$\begin{aligned}\chi^2 &= \frac{(15 - 17.4)^2}{17.4} + \frac{(61 - 54.24)^2}{54.24} + 2.642 \\ &= 3.815 \dots \approx 3.82.\end{aligned}$$

There are $(3 - 1)(2 - 1) = 2$ degrees of freedom, since the table has three rows and two columns. At the 5% level, the upper-tail critical value is $\chi^2_{2,0.05} = 5.991$.

$$3.82 < 5.991.$$

The statistic is not in the critical region, so we do not reject H_0 .

There is no evidence at the 5% level of an association between treatment application free.

(5)

www.mathswithmatt.co.uk

Question 2

Question 2

Xena catches fish at random at a constant rate of 0.6 per hour.

(a) Find the probability that she catches exactly 4 fish in 5 hours. (2)

The probability that she catches no fish in t hours is less than 0.16.

(b) Find the minimum value of t , to 1 decimal place. (3)

Independently, Zion catches fish at a mean rate of 0.8 per hour. After using new bait, the total caught by Xena and Zion in a randomly selected 4-hour period is 12.

(c) At the 5% level, use a suitable test to determine whether there is evidence that the catching rate has increased. State hypotheses and the p-value used. (5)

Model solution

(a) Counts of independent random events occurring at a constant rate are modelled by a Poisson distribution. In 5 hours the mean count is $\lambda = 0.6 \times 5 = 3$, so $X \sim \text{Po}(3)$.

$$P(X = 4) = \frac{e^{-3} 3^4}{4!} = 0.1680 \dots$$

$$0.168$$

(2)

(b) Over t hours the mean is $0.6t$. For a Poisson variable, $P(X = 0) = e^{-\lambda}$, so translate the stated probability into an inequality.

$$e^{-0.6t} < 0.16.$$

$$-0.6t < \ln(0.16) \implies t > \frac{-\ln(0.16)}{0.6} = 3.054 \dots$$

The time must be greater than this value. To one decimal place, the minimum stated time is therefore 3.1 hours.

$$t = 3.1 \text{ hours}$$

(3)

(c) Under no increase, the combined rate is $0.6 + 0.8 = 1.4$ fish per hour. We use an upper-tailed Poisson test because the alternative is an increased rate.

$$H_0 : \lambda = 1.4 \text{ fish per hour}, \quad H_1 : \lambda > 1.4 \text{ fish per hour.}$$

For 4 hours, the total J would have distribution $\text{Po}(4 \times 1.4) = \text{Po}(5.6)$ if H_0 were true. The p-value is the probability of an outcome at least as large as the observed 12.

$$\begin{aligned} p &= P(J \geq 12) \\ &= 1 - P(J \leq 11) \\ &= 1 - 0.9875 \dots = 0.01248 \dots \end{aligned}$$

Since $0.01248 < 0.05$, reject H_0 .

There is evidence at the 5% level that the rate at which fish are caught has increased (value ≈ 0.0125).

(5)

www.mathswithmatt.co.uk

Question 3

Question 3

A coin is spun 5 times and the number of heads is recorded.

(a) Tao suggests playing the game only 20 times and using a chi-squared test to investigate whether the coin might be biased. Explain why this may cause problems. (1)

Chris plays the game 500 times. Her observed and expected frequencies, where the proposed model is $B(5, 0.6)$, are

Number of heads	0	1	2	3	4	5
O	2	27	93	181	146	51
E	5.12	38.40	115.20	172.80	129.60	38.88

(b) State the degrees of freedom for Chris' test, giving a reason. (1)

(c) Carry out the test at the 5% level, stating hypotheses, test statistic, critical value and conclusion. (5)

(d) Show working to find an alternative model that better fits Chris' data. (2)

Model solution

(a) A chi-squared approximation requires expected frequencies to be sufficiently large, conventionally at least 5. With only 20 games, some expected frequencies are likely to be below 5.

The sample is too small for all expected frequencies to be likely to exceed 5, so the chi-squared test conditions may fail.

(1)

(b) There are six outcome classes. No parameter has been estimated from these data, because the proposed value $p = 0.6$ is specified. Hence the degrees of freedom are $6 - 1 = 5$.

5degreesoffreedom.

(1)

(c) This is a goodness-of-fit test: it checks whether the observed distribution of numbers of heads is consistent with the specified binomial model.

H_0 : the number of heads has distribution $B(5, 0.6)$, H_1 : this is not a suitable model.

$$\begin{aligned}
 \chi^2 &= \sum \frac{(O - E)^2}{E} \\
 &= \frac{(2 - 5.12)^2}{5.12} + \frac{(27 - 38.40)^2}{38.40} + \frac{(93 - 115.20)^2}{115.20} \\
 &\quad + \frac{(181 - 172.80)^2}{172.80} + \frac{(146 - 129.60)^2}{129.60} + \frac{(51 - 38.88)^2}{38.88} \\
 &= 15.8063 \dots \approx 15.8.
 \end{aligned}$$

With 5 degrees of freedom, the 5% upper-tail critical value is $\chi^2_{5,0.05} = 11.070$.

$$15.8 > 11.070.$$

The statistic is in the critical region, so we reject H_0 .

At the 5% level, $B(5, 0.6)$ is not a suitable model for Chris' numbers of heads.

(5)

(d) Keep the binomial number of trials $n = 5$, but estimate its head probability from the sample mean number of heads. The total number of heads is obtained by weighting each frequency by its number of heads.

$$\bar{x} = \frac{0(2) + 1(27) + 2(93) + 3(181) + 4(146) + 5(51)}{500} = 3.19.$$

For $B(5, p)$, $E(X) = 5p$. Equating this to the observed mean gives $p = 3.19/5$.

$$p = 0.638.$$

An improved model is $X \sim B(5, 0.638)$.

(2)

www.mathswithmatt.co.uk

Question 4

Question 4

The discrete random variable X has distribution

x	0	2	3	6
$P(X = x)$	p	0.25	q	0.4

(a) Find, in terms of q , (i) $E(X)$, and (ii) $E(X^2)$. (2)

Given that $\text{Var}(X) = 3.66$, (b) show that $q = 0.3$. (3)

The score in each independent game is X . (c) Find the probability that the total score after 4 games is exactly 20. (3)

A round consists of 4 games and, only when their total is exactly 20, 2 bonus games. A £10 prize is awarded if 6 games are played and the total is at least 27. Bobby plays 3 rounds. (d) Find the probability that Bobby wins at least £10. (6)

Model solution

(a) The expectation of a discrete random variable is found by multiplying each value by its probability and adding.

$$\begin{aligned} E(X) &= 0p + 2(0.25) + 3q + 6(0.4) \\ &= 2.9 + 3q. \end{aligned}$$

$$(a)(i) E(X) = 2.9 + 3q$$

For $E(X^2)$, square each possible score before weighting it by its probability.

$$\begin{aligned} E(X^2) &= 0^2p + 2^2(0.25) + 3^2q + 6^2(0.4) \\ &= 15.4 + 9q. \end{aligned}$$

$$(a)(ii) E(X^2) = 15.4 + 9q$$

(2)

(b) Use $\text{Var}(X) = E(X^2) - [E(X)]^2$, substituting the two expressions found in part (a).

$$(15.4 + 9q) - (2.9 + 3q)^2 = 3.66.$$

$$15.4 + 9q - (8.41 + 17.4q + 9q^2) = 3.66,$$

$$9q^2 + 8.4q - 3.33 = 0.$$

$$(10q - 3)(30q + 37) = 0.$$

$$q = 0.3 \quad \text{or} \quad q = -\frac{37}{30}.$$

A probability cannot be negative, so the second root is impossible.

$$q = 0.3$$

(3)

(c) The probabilities are now $P(X = 2) = 0.25$ and $P(X = 6) = 0.4$. A total of 20 from four possible scores can only be made by $6 + 6 + 6 + 2$. There are four positions for the score 2.

$$P(\text{total of 4 games} = 20) = 4(0.4)^3(0.25) = 0.064.$$

$$0.064$$

(3)

(d) To win in one round, the first four games must total 20, which has probability 0.064. Given this, the two bonus games must add at least 7, because $20 + 7 = 27$. The possible bonus-score pairs are (6, 6), (6, 3), (3, 6), (6, 2), (2, 6).

$$\begin{aligned} P(X_5 + X_6 \geq 7) &= (0.4)^2 + 2(0.4)(0.3) + 2(0.4)(0.25) \\ &= 0.6. \end{aligned}$$

Thus the probability of a prize in one round is the product of the probability of reaching the bonus games and the conditional probability of reaching 27.

$$P(\text{prize in one round}) = 0.064(0.6) = 0.0384.$$

The rounds are independent, so if Y is the number of prizes in 3 rounds, then $Y \sim B(3, 0.0384)$. "At least £10" means at least one prize. Use the complement of no prizes.

$$\begin{aligned} P(Y \geq 1) &= 1 - P(Y = 0) \\ &= 1 - (1 - 0.0384)^3 \\ &= 0.1108\dots \end{aligned}$$

$$P(\text{Bobby wins at least £10}) \approx 0.111$$

(6)