Maths with Matt

Physics and maths resources

Edexcel A Level Further Mathematics

9FM0/3B - Further Statistics 1

Tuesday 18 June 2019

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 2019 mark scheme. The aim is a clear **full-mark model response**, not merely a list of final answers.

Question wording is reproduced from the user-provided 2019 paper for study/reference. Pearson retains copyright in the original examination material.

Contents

Question 1 2

Question 2 3

Question 3 4

Question 4 5

Question 5 7

Question 6 9

Question 7 11

www.mathswithmatt.co.uk

Question 1

Question 1

A token is placed independently in each bar with probability 0.02. Andreia buys 40 bars. Find the probability that she obtains at least 3 tokens. Barney buys bars until obtaining his third token. Find the probability that his third token is in bar 40, and the expected number of bars he buys.

Model solution

(a) There is a fixed number of independent bars, each with the same token probability, so a binomial model is appropriate. A prize needs at least three tokens.

$$X \sim B(40, 0.02).$$

It is shorter to use the complement, since the calculator can add the probabilities for 0, 1 *and* 2 *tokens.*

$$P(X \geqslant 3) = 1 - P(X \leqslant 2) = 1 - \sum_{x=0}^{2} \binom{40}{x} (0.02)^x (0.98)^{40-x} = 0.04567\ldots$$

$$P(\textit{Andreia can claim a prize}) = 0.0457.$$

(2)

(b) Here the number of trials continues until the third success, so use a negative binomial distribution. For the third success to occur on bar 40*, there must be exactly two tokens in the first* 39 *bars and a token in bar* 40*.*

$$Y \sim \text{NegBin}(3, 0.02), \qquad P(Y = 40) = \binom{39}{2} (0.02)^3 (0.98)^{37} = 0.0028071\ldots$$

$$P(Y = 40) = 0.00281.$$

(3)

(c) For a negative binomial variable counting trials to the r*th success,* $E(Y) = r/p$*.*

$$E(Y) = \frac{3}{0.02} = 150.$$

$$\text{Theexpectednumberofbarsis150.}$$

(1)

Question 2

Question 2

Calls arrive randomly at an average rate of 80 in 4 hours. Find: (a) the probability of more than 4 calls in 20 minutes; (b) the probability that exactly 3 of four separate 5-minute breaks have no calls; (c) the probability of exactly one missed call in each of a 5-minute and a 15-minute break.

Model solution

(a) Random arrivals at a constant average rate are modelled by a Poisson distribution. Twenty minutes is $\frac{1}{12}$ of four hours, so the mean for this period is $80/12 = \frac{20}{3}$.

$$C \sim \text{Po}\left(\frac{20}{3}\right).$$

Use a complement to find a probability of more than 4.

$$P(C > 4) = 1 - P(C \leqslant 4) = 0.79437\ldots$$

0.794.

(3)

(b) In five minutes the Poisson mean is $\frac{80}{48} = \frac{5}{3}$, so the probability of no call in one break is $e^{-5/3}$. The four non-overlapping breaks are independent, hence a binomial count of empty breaks can be used.

$$N \sim B\left(4, e^{-5/3}\right), \qquad P(N = 3) = \binom{4}{3}(e^{-5/3})^3(1 - e^{-5/3}) = 0.021861\ldots$$

0.0219.

(2)

(c) The two breaks have different lengths, so find the Poisson probability of one call for each period and multiply: the periods are disjoint and therefore independent.

$$P(\text{Po}(5/3) = 1) = \frac{5}{3}e^{-5/3}, \qquad P(\text{Po}(5) = 1) = 5e^{-5}.$$

$$P(\textit{one in each}) = \left(\frac{5}{3}e^{-5/3}\right)(5e^{-5}) = 0.010605\ldots$$

0.0106.

(3)

Question 3

Question 3

A spinner has probabilities $0.3, 0.1, 0.2, 0.1, 0.3$ of landing on $1, 2, 3, 4, 5$ respectively. It is spun 80 times. Estimate the probability that the sample mean exceeds 3.25.

Model solution

Because the question concerns the mean of 80 independent spins, use the central limit theorem after finding the mean and variance of one spin.

$$\mu = E(X) = 1(0.3) + 2(0.1) + 3(0.2) + 4(0.1) + 5(0.3) = 3,$$

$$E(X^2) = 1^2(0.3) + 2^2(0.1) + 3^2(0.2) + 4^2(0.1) + 5^2(0.3) = 11.6,$$

$$\text{Var}(X) = E(X^2) - [E(X)]^2 = 11.6 - 3^2 = 2.6.$$

Thus the sample mean has approximately a normal distribution with the same mean and variance divided by 80.

$$\overline{X} \sim N\left(3, \frac{2.6}{80}\right).$$

$$P(\overline{X} > 3.25) = P\left(Z > \frac{3.25 - 3}{\sqrt{2.6/80}}\right) = P(Z > 1.3867\ldots) = 0.082758\ldots$$

The estimated probability is 0.0828.

(6)

Question 4

Question 4

In 80 equal squares, the frequencies for $0, 1, 2, 3, 4, 5, 6, 7+$ oak trees are $1, 4, 21, 23, 13, 11, 7, 0$. Liam proposes six trees originally per square, each surviving independently with probability p. Complete goodness-of-fit tests for Liam's binomial model and Simone's Poisson model; for the latter the mean is 3.3 and the expected frequencies for 3 and 5 trees are s and t.

Model solution

(a) Each square starts with six trees and each either survives or does not survive with constant probability p, which is precisely the binomial setting.

$$T \sim B(6, p).$$

$$T \sim B(6, p).$$

(2)

(b) Use a chi-squared goodness-of-fit test. Classes with expected frequency below 5 must be pooled, so combine the observed classes 5 and 6 (including 7+) to give observed frequency 18 and expected frequency $10.87 + 2.21 = 13.08$.

class	0 *or* 1	2	3	4	5 *or more*
O_i	5	21	23	13	18
E_i	5.53	14.89	24.26	22.24	13.08
$\frac{(O_i - E_i)^2}{E_i}$	0.051	2.507	0.065	3.839	1.851

$$\chi^2 = \sum \frac{(O_i - E_i)^2}{E_i} = 8.313.$$

There are five pooled classes and one parameter, p, has been estimated (here $\hat{p} = 0.55$). Therefore $\nu = 5 - 1 - 1 = 3$. At the 5% level the critical value is $\chi^2_{3,0.95} = 7.815$.

$$8.313 > 7.815.$$

$$\text{Reject} H_0 : \text{at the 5\% level, Liam's binomial model is not suitable.}$$

(7)

(c) For Simone's model, $R \sim Po(3.3)$, so an expected frequency is $80P(R = r)$.

$$s = 80e^{-3.3}\frac{3.3^3}{3!} = 17.6729\ldots = 17.67, \qquad t = 80e^{-3.3}\frac{3.3^5}{5!} = 9.6229\ldots = 9.62.$$

$$s = 17.67 \text{ and } t = 9.62.$$

(4)

(d) A goodness-of-fit test compares whether the stated distribution provides an adequate model.

H_0 : *the number of oak trees per square follows a Poisson distribution;* H_1 : *it does not.*

$$H_0 : \text{Poisson is a good fit; } H_1 : \text{Poisson is not a good fit.}$$

(1)

(e) All six Poisson classes have adequate expected frequencies, and the mean 3.3 *has been estimated. Hence* $\nu = 6 - 1 - 1 = 4$.

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

The supplied statistic is 8.749, *which is below the critical value.*

$$8.749 < 9.488.$$

$$\text{Do not reject } H_0 : \text{at the 5\% level, Simone's Poisson model is suitable.}$$

(3)

(f) A Poisson model describes random occurrences, whereas the rejected binomial model represented deliberately planted trees with independent survival.

The woodland is more likely to be wild than cultivated.

(2)

Question 5

Question 5

Monthly bypass accidents previously followed a Poisson distribution with mean 2.5. Observations are made in 3-month periods after roadworks. Construct a two-sided 5% test for a changed mean, then find its Type I error probability, the chance at least two of eight periods are significant, and its Type II error probability if the new monthly mean is 2.1.

Model solution

(a) The wording "a change" requires a two-sided test. In a three-month period the null mean is $3(2.5) = 7.5$.

$$H_0 : \lambda = 2.5, \qquad H_1 : \lambda \neq 2.5, \qquad X \sim \text{Po}(7.5) \text{ under } H_0.$$

Put no more than 0.025 *in each tail, selecting the most extreme integer values that keep each tail within this limit.*

$$P(X \leqslant 2) = 0.020256\ldots \approx 0.0203,$$

$$P(X \geqslant 14) = 1 - P(X \leqslant 13) = 0.021565\ldots \approx 0.0216.$$

$$\text{Critical region : } X \leqslant 2 \text{ or } X \geqslant 14; \text{ the tail probabilities are } 0.0203 \text{ and } 0.0216.$$

(5)

(b) A Type I error is rejecting H_0 *when it is true, so its probability is the total size of the actual discrete critical region.*

$$P(\textit{Type I error}) = 0.020256\ldots + 0.021565\ldots = 0.041821\ldots$$

$$0.0418 \text{ (approximately } 0.0419).$$

(1)

(c) Two years give eight separate three-month periods. Under H_0*, each period is significant with probability* $0.041821\ldots$*, so the number significant has a binomial distribution.*

$$M \sim B(8, 0.041821\ldots).$$

$$P(M \geqslant 2) = 1 - P(M \leqslant 1) = 1 - \left[(1-p)^8 + 8p(1-p)^7\right] = 0.04139\ldots$$

$$0.0414 \text{ (or } 0.0415 \text{ using rounded probabilities}).$$

(3)

(d) *If the true monthly mean is* 2.1*, the mean for a three-month period is* 6.3*.* ***A Type II error*** *means failing to reject, so retain the values outside the critical region.*

$$Y \sim \text{Po}(6.3), \qquad P(\textbf{\textit{Type II error}}) = P(3 \leqslant Y \leqslant 13).$$

$$P(3 \leqslant Y \leqslant 13) = P(Y \leqslant 13) - P(Y \leqslant 2) = 0.9945147\ldots - 0.049846\ldots = 0.94467\ldots$$

$$P(\textbf{\textit{Type II error}}) = 0.945.$$

(3)

www.mathswithmatt.co.uk

Question 6

Question 6

The probability generating function of a discrete random variable X is

$$G_X(t) = k \ln\left(\frac{2}{2-t}\right).$$

Find (a) k, (b) Var(X) exactly, and (c) $P(X = 3)$.

Model solution

(a) Every probability generating function satisfies $G_X(1) = 1$. This condition determines the constant.

$$1 = G_X(1) = k \ln\left(\frac{2}{1}\right) = k \ln 2, \qquad k = \frac{1}{\ln 2}.$$

$$k = \frac{1}{\ln 2}.$$

(1)

(b) For a pgf, $G'_X(1) = E(X)$ and $G''_X(1) = E[X(X-1)]$. Differentiating first gives the mean, then differentiating again provides the factorial second moment needed for the variance.

$$G_X(t) = \frac{1}{\ln 2}\left[\ln 2 - \ln(2-t)\right], \qquad G'_X(t) = \frac{1}{(2-t)\ln 2},$$

$$E(X) = G'_X(1) = \frac{1}{\ln 2}.$$

$$G''_X(t) = \frac{1}{(2-t)^2 \ln 2}, \qquad G''_X(1) = \frac{1}{\ln 2}.$$

Since $E(X^2) = G''_X(1) + G'_X(1)$, subtract the square of the mean.

$$\mathrm{Var}(X) = G''_X(1) + G'_X(1) - [G'_X(1)]^2 = \frac{1}{\ln 2} + \frac{1}{\ln 2} - \left(\frac{1}{\ln 2}\right)^2 = \frac{1}{\ln 2}\left(2 - \frac{1}{\ln 2}\right).$$

$$\mathrm{Var}(X) = \frac{1}{\ln 2}\left(2 - \frac{1}{\ln 2}\right).$$

(7)

(c) In a pgf, the coefficient of t^3 is $P(X = 3)$. Equivalently, the third derivative at zero is $3!$ times this coefficient.

$$G'''_X(t) = \frac{2}{(2-t)^3 \ln 2}, \qquad P(X = 3) = \frac{G'''_X(0)}{3!} = \frac{1}{6} \cdot \frac{2}{2^3 \ln 2} = \frac{1}{24 \ln 2}.$$

$$P(X = 3) = \frac{1}{24 \ln 2} = 0.0601 \text{ (approximately)}.$$

(4)

www.mathswithmatt.co.uk

Question 7

Question 7

A spinner lands blue with probability $\frac{1}{3}$. Let B be the number of the spin on which blue first occurs. Find (a)(i) $P(B = 4)$, (ii) $P(B \leqslant 5)$, and (b) $E(B^2)$. In a game, choosing red gives score e^X when red first occurs, while choosing blue gives score X^2 when blue first occurs. Recommend a colour.

Model solution

(a)(i) The number of the trial of the first blue result has a geometric distribution with success probability $\frac{1}{3}$. For blue first on spin 4, the first three spins must be red.

$$B \sim \text{Geo}\left(\frac{1}{3}\right), \qquad P(B = 4) = \left(\frac{2}{3}\right)^3 \left(\frac{1}{3}\right) = \frac{8}{81}.$$

$$\frac{8}{81}.$$

(2)

(a)(ii) It is simplest to subtract the probability that the first blue is later than spin 5, namely five consecutive red results.

$$P(B \leqslant 5) = 1 - P(B > 5) = 1 - \left(\frac{2}{3}\right)^5 = \frac{211}{243}.$$

$$\frac{211}{243}.$$

(2)

(b) Use $E(B^2) = \text{Var}(B) + [E(B)]^2$. For a geometric variable, $E(B) = 1/p$ and $\text{Var}(B) = (1-p)/p^2$.

$$E(B) = 3, \qquad \text{Var}(B) = \frac{2/3}{(1/3)^2} = 6,$$

$$E(B^2) = 6 + 3^2 = 15.$$

$$E(B^2) = 15.$$

(3)

(c) The blue score has expected value $E(B^2) = 15$. For red, the success probability is $\frac{2}{3}$, so $X \sim \text{Geo}(\frac{2}{3})$. Because the score is e^X, calculate its expected value directly rather than substituting $E(X)$ into the exponential.

$$E(e^X) = \sum_{x=1}^{\infty} e^x \left(\frac{1}{3}\right)^{x-1} \left(\frac{2}{3}\right) = \frac{2e}{3} \sum_{x=1}^{\infty} \left(\frac{e}{3}\right)^{x-1}.$$

This is a convergent geometric series because $e/3 < 1$.

$$E(e^X) = \frac{2e}{3} \cdot \frac{1}{1 - e/3} = \frac{2e}{3 - e} = 19.297\ldots$$

Recommendred, sinceitsexpectedscore19.3 . . . isgreaterthantheblueexpectedsc

(5)

www.mathswithmatt.co.uk