Maths with Matt

Physics and maths resources

Edexcel A Level Further Mathematics

9FM0/3B - Further Statistics 1

Friday 16 June 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.

Question wording is reproduced from the user-provided 2023 paper for study/reference. Pearson retains copyright in the original examination material.

Contents

Question 1 2

Question 2 3

Question 3 4

Question 4 6

Question 5 7

Question 6 8

Question 7 10

www.mathswithmatt.co.uk

Question 1

Question 1

The discrete random variable X has distribution

x	-2	-1	0	1	3
$P(X = x)$	0.25	a	b	a	0.30

where a and b are probabilities. (a) Find $E(X)$. Given that $\text{Var}(X) = 3.9$, (b) find a and b. The independent variables X_1, X_2 each have this distribution. (c) Find $P(X_1 + X_2 > 3)$.

Model solution

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

$$E(X) = (-2)(0.25) + (-1)a + 0b + 1a + 3(0.30) = -0.5 - a + a + 0.9 = 0.4.$$

$$\boxed{E(X) = 0.4.}$$

(2)

(b) The given variance lets us use $\text{Var}(X) = E(X^2) - [E(X)]^2$. *First calculate* $E(X^2)$, *keeping* a *in the expression.*

$$E(X^2) = (-2)^2(0.25) + (-1)^2a + 0^2b + 1^2a + 3^2(0.30) = 1 + 2a + 2.7 = 3.7 + 2a.$$

$$3.9 = (3.7 + 2a) - (0.4)^2, \qquad 3.9 = 3.54 + 2a, \qquad a = 0.18.$$

The probabilities must sum to 1, *so this now determines* b.

$$0.25 + a + b + a + 0.30 = 1 \quad \Rightarrow \quad 2a + b = 0.45 \quad \Rightarrow \quad b = 0.45 - 2(0.18) = 0.09.$$

$$\boxed{a = 0.18 \text{and} b = 0.09.}$$

(4)

(c) List the pairs whose sum exceeds 3. *Since the largest value is* 3, *the only possibilities are* $(1,3)$, $(3,1)$ *and* $(3,3)$. *Independence means that each pair probability is the product of its two probabilities.*

$$P(X_1 + X_2 > 3) = 0.18(0.30) + 0.30(0.18) + 0.30(0.30) = 0.198.$$

$$\boxed{P(X_1 + X_2 > 3) = 0.198.}$$

(3)

Question 2

Question 2

Calls arrive randomly at a call centre at an average rate of 1.7 per minute. After a one-week closure, 25 calls are received in a random 10-minute sample. (a) At the 5% level, test whether the rate has changed. Only 1.2% of calls last longer than 8 minutes. (b) Find the probability that more than 2 of 70 calls last longer than 8 minutes. Records show that 95% of days have at least one call lasting longer than 30 minutes. On a day with 900 calls, none lasts longer than 30 minutes. (c) Using a Poisson approximation, estimate the proportion of calls lasting longer than 30 minutes.

Model solution

(a) Random arrivals at a constant average rate are modelled by a Poisson distribution. The phrase "has changed" requires a two-tailed test.

$$H_0 : \lambda = 1.7 \text{ calls per minute}, \qquad H_1 : \lambda \neq 1.7 \text{ calls per minute}.$$

For the 10-minute period, let X *be the number of calls. Under* H_0, $X \sim \text{Po}(10\times1.7) = \text{Po}(17)$. *The observed result is 25, so its upper-tail probability is*

$$P(X \geqslant 25) = 1 - P(X \leqslant 24) = 0.0406463\ldots$$

For a two-tailed 5% test, compare this with 0.025. *It is not sufficiently extreme (equivalently, the upper critical region is* $X \geqslant 27$*).*

$$\text{Do not reject } H_0. \text{ There is insufficient evidence at the 5\% level that the rate of calls sh}$$

(4)

(b) Each call either does or does not exceed 8 minutes, with a fixed probability 0.012, *so use a binomial model.*

$$T \sim \text{Bin}(70, 0.012).$$

"More than 2" means $T \geqslant 3$, *so use the complement of* $T \leqslant 2$.

$$P(T > 2) = 1 - P(T \leqslant 2) = 1 - 0.947725\ldots = 0.052275\ldots$$

$$P(T > 2) = 0.0523 \text{ (to 3 significant figures)}.$$

(3)

(c) Let C *be the number, among 900 calls, lasting more than 30 minutes, and let* p *be the required proportion. Since* n *is large and the event is rare, approximate* $\text{Bin}(900, p)$ *by* $\text{Po}(900p)$.

$$C \sim \text{Po}(900p), \qquad P(C = 0) \approx e^{-900p}.$$

The stated 95% probability of at least one means $P(C = 0) = 0.05$. *Hence*

$$e^{-900p} = 0.05 \quad \Rightarrow \quad 900p = -\ln(0.05) = 2.9957\ldots \quad \Rightarrow \quad p = 0.003328\ldots$$

$$\text{Estimated proportion} = 0.00333 \text{ (about } 0.333\%).$$

(4)

Question 3

Question 3

Over 170 days, a child spins a cardboard coin 5 times; the frequencies of $0,1,2,3,4,5$ heads are respectively $3,10,45,62,38,12$. Marcus proposes $B(5,0.5)$, whose expected frequencies are $r, 26.56, s, s, 26.56, r$. (a) Find r, s. (b) Test this model at 5%. Nima proposes $B(5,p)$. (c) Estimate p. Her expected frequencies are $2.07, 14.65, 41.44, 58.63, 41.47, 11.74$, and her test statistic is 1.62. (d) State, with reason, the degrees of freedom and the 5% critical value. (e) Comment on the probability of heads and independence of spins.

Model solution

(a) Under Marcus's model, the number of heads is $X \sim \text{Bin}(5, 0.5)$. *Expected frequency equals* 170 *times the relevant binomial probability.*

$$r = 170P(X = 0) = 170(0.5)^5 = 5.3125,$$

$$s = 170P(X = 2) = 170\binom{5}{2}(0.5)^2(0.5)^3 = 53.125.$$

$$r = 5.31 \text{ and } s = 53.1 \text{ (to 2 decimal places).}$$

(3)

(b) This is a goodness-of-fit test, comparing the observed frequencies O *with the expected frequencies* E.

$H_0 : B(5, 0.5)$ *is a suitable model;* $H_1 : B(5, 0.5)$ *is not a suitable model.*

$$\chi^2 = \sum \frac{(O-E)^2}{E} = \frac{(3-5.3125)^2}{5.3125} + \frac{(10-26.5625)^2}{26.5625} + \frac{(45-53.125)^2}{53.125}$$
$$+ \frac{(62-53.125)^2}{53.125} + \frac{(38-26.5625)^2}{26.5625} + \frac{(12-5.3125)^2}{5.3125} = 27.45\ldots$$

There are six categories and no parameter has been estimated, so $\text{df} = 6 - 1 = 5$. *At 5%, the critical value is* $\chi^2_{5,0.95} = 11.070$.

$$27.45 > 11.070.$$

$$\text{Reject } H_0. \text{ At the 5\% level, } B(5, 0.5) \text{ is not a good fit to the data.}$$

(6)

(c) For a binomial model, estimate p *by the total number of heads divided by the total number of spins.*

$$\hat{p} = \frac{0(3) + 1(10) + 2(45) + 3(62) + 4(38) + 5(12)}{170 \times 5} = \frac{498}{850} = 0.58588\ldots$$

$$\hat{p} = 0.586.$$

(1)

(d)(i) The expected frequency for zero heads is below 5, so the first two cells must be pooled. This leaves five groups. One degree of freedom is lost because frequencies total 170, and one because p was estimated from these data.

$$\mathrm{df} = 5 - 2 = 3.$$

$$\mathrm{Degrees of freedom} = 3.$$

(2)

(d)(ii) With 3 degrees of freedom, the 5% upper-tail chi-squared critical value is

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

$$\mathrm{Critical value} = 7.815.$$

(1)

(e)(i) Nima's statistic satisfies $1.62 < 7.815$, whereas Marcus's model was rejected. Thus a binomial model with $p \approx 0.586$ fits, rather than one with $p = 0.5$.

$$\mathrm{The coin appears biased, with probability of heads approximately 0.586 (about 0.6}$$

(1)

(e)(ii) Nima's binomial model is not rejected. A binomial model requires independent spins with the same probability of a head, so the data support the independence assumption.

It is reasonable to assume that the spins are independent.

(1)

Question 4

Question 4

There are 32 students. Each rolls a fair die repeatedly until their fourth six, and records the total number of rolls. Estimate the probability that the class mean number of rolls is less than 27.2.

Model solution

Let X be one student's number of rolls. This counts trials up to the fourth occurrence of a success (a six), so $X \sim \text{NegBin}(4, 1/6)$. For this convention,

$$E(X) = \frac{r}{p} = \frac{4}{1/6} = 24, \qquad \text{Var}(X) = \frac{r(1-p)}{p^2} = \frac{4(5/6)}{(1/6)^2} = 120.$$

The class mean $\overline{X}$ averages 32 independent student results. The central limit theorem gives

$$\overline{X} \sim N\left(24, \frac{120}{32}\right), \qquad \text{sd}(\overline{X}) = \sqrt{\frac{120}{32}}.$$

$$P(\overline{X} < 27.2) = P\left(Z < \frac{27.2 - 24}{\sqrt{120/32}}\right) = P(Z < 1.652\ldots) = 0.95078\ldots$$

$$P(\overline{X} < 27.2) \approx 0.951.$$

(6)

Question 5

Question 5

Carton contents are normally distributed with mean μ ml and standard deviation 8 ml. A random sample of 25 has mean $\overline{X}$. (a) At 5%, test whether $\mu < 330$, giving the critical region. A Director takes a sample of 55 with mean $\overline{Y}$ and uses critical region $\overline{Y} < 328$ or $\overline{Y} > 332$. (b) Find the Type I error probability. When $\mu = 325$, (c) find the Type II error probability for part (a).

Model solution

(a) "Less than" gives a lower-tailed test. Under the null hypothesis, the sample mean has a normal distribution with standard error $8/\sqrt{25}$.

$$H_0 : \mu = 330, \qquad H_1 : \mu < 330, \qquad \overline{X} \sim N\left(330, \left(\frac{8}{\sqrt{25}}\right)^2\right) \text{ under } H_0.$$

Choose c *so that* $P(\overline{X} < c \mid H_0) = 0.05$. *Since* $z_{0.05} = -1.6449$,

$$\frac{c - 330}{8/5} = -1.6449 \quad \Rightarrow \quad c = 327.368\ldots$$

$$\text{Reject} H_0 \text{if} \overline{X} < 327.37\text{ml} (\text{or} \overline{X} < 327\text{ml}).$$

(4)

(b) A Type I error is rejecting the true hypothesis $\mu = 330$. *The Director's region has equal lower and upper tail probabilities by symmetry.*

$$\overline{Y} \sim N\left(330, \left(\frac{8}{\sqrt{55}}\right)^2\right),$$

$$P(\textit{Type I error}) = 2P(\overline{Y} < 328) = 2P\left(Z < \frac{328 - 330}{8/\sqrt{55}}\right) = 0.063732\ldots$$

$$P(\textit{Type I error}) = 0.0637.$$

(2)

(c) A Type II error here means failing to reject H_0 *when the true mean is 325 ml. Therefore use the non-rejection region from part (a),* $\overline{X} \geqslant 327.368\ldots$, *with the distribution centred at 325.*

$$\overline{X} \sim N\left(325, \left(\frac{8}{5}\right)^2\right),$$

$$P(\textit{Type II error}) = P(\overline{X} > 327.368\ldots \mid \mu = 325) = P\left(Z > \frac{327.368\ldots - 325}{1.6}\right) = 0.069423\ldots$$

$$P(\textit{Type II error}) = 0.0694.$$

(2)

Question 6

Question 6

The discrete random variable X has probability generating function

$$G_X(t) = \left(\frac{t}{3-2t}\right)^2.$$

(a) Specify the distribution of X. (b) A fair die is rolled repeatedly: describe an outcome modelled by X. (c) Use calculus and $G_X(t)$ to find (i) $E(X)$ and (ii) $\mathrm{Var}(X)$. The random variable Y has pgf

$$G_Y(t) = \left(\frac{t^{10}}{(3-2t^3)^2}\right).$$

(d) Find the exact value of $P(Y = 19)$.

Model solution

(a) Compare the given pgf with the negative-binomial pgf

$$G(t) = \left(\frac{pt}{1-(1-p)t}\right)^r.$$

Here $r = 2$ *and* $p = 1/3$*, since* $pt/[1-(1-p)t] = (t/3)/(1-2t/3) = t/(3-2t)$.

$$X \sim \mathrm{NegBin}(2, 1/3).$$

(2)

(b) On a fair die, a 5 or 6 has probability $2/6 = 1/3$*. The negative-binomial variable counts the number of rolls up to the second success.*

Xcanbethenumberofrollsneededtoobtaintworollsthatare5or6.

(1)

(c)(i) For a pgf, $E(X) = G'_X(1)$*. Differentiate before substituting* $t = 1$*:*

$$G'_X(t) = 2\left(\frac{t}{3-2t}\right)\left(\frac{(3-2t)-t(-2)}{(3-2t)^2}\right) = \frac{6t}{(3-2t)^3}.$$

$$E(X) = G'_X(1) = \frac{6}{(3-2)^3} = 6.$$

$$E(X) = 6.$$

(3)

(c)(ii) The second pgf derivative gives the factorial moment $E[X(X-1)]$.

$$G''_X(t) = \frac{d}{dt}\left(6t(3-2t)^{-3}\right) = \frac{6}{(3-2t)^3} + \frac{36t}{(3-2t)^4} = \frac{18+24t}{(3-2t)^4}.$$

$$G''_X(1) = 42.$$

Now $E(X^2) = G''_X(1) + G'_X(1)$, *so*

$$\text{Var}(X) = G''_X(1) + G'_X(1) - [G'_X(1)]^2 = 42 + 6 - 6^2 = 12.$$

$$\text{Var}(X) = 12.$$

(4)

(d) The pgfs reveal a useful transformation:

$$G_Y(t) = t^4 G_X(t^3),$$

so $Y = 3X + 4$. *The event* $Y = 19$ *is therefore exactly* $X = 5$. *For a negative-binomial variable, the second success occurs on roll 5 when there is one success in the first four rolls and a success on roll 5.*

$$P(Y = 19) = P(X = 5) = \binom{4}{1}\left(\frac{1}{3}\right)^2\left(\frac{2}{3}\right)^3 = \frac{32}{243}.$$

$$P(Y = 19) = \frac{32}{243}.$$

(3)

Question 7

Question 7

Each spin has probability 0.2 of red. (a) Find the probability of red (i) for the first time on spin 4, (ii) for the third time on spin 8, (iii) exactly four times in ten spins. Yellow has probability 0.4 each spin. In a game, R is red for the first time in at most 4 spins and Y is yellow for the third time in at most 7 spins. (b) Determine which event has greater probability, showing calculations.

Model solution

(a)(i) A first success on trial 4 has three failures followed by a success, so it is geometric.

$$\mathrm{P}(\textit{first red on spin 4}) = (0.8)^3(0.2) = 0.1024.$$

$$0.1024.$$

(2)

(a)(ii) For the third success to occur on spin 8, there must be exactly two reds in the first seven spins and a red on spin 8. This is a negative-binomial situation.

$$\mathrm{P} = \binom{7}{2}(0.2)^2(0.8)^5(0.2) = 0.05505024.$$

$$0.0551 (\mathrm{to\,3\,significant\,figures}).$$

(2)

(a)(iii) A fixed number of reds in a fixed ten spins is binomial.

$$\mathrm{P}(\textit{exactly 4 reds}) = \binom{10}{4}(0.2)^4(0.8)^6 = 0.088080384.$$

$$0.0881 (\mathrm{to\,3\,significant\,figures}).$$

(2)

(b) Event R *is at least one red in four spins. It is quickest to use its complement, no red in four spins.*

$$\mathrm{P}(\mathrm{R}) = 1 - (0.8)^4 = 0.5904.$$

For Y*, let* W *be the number of yellow outcomes in seven spins. The event "third yellow in at most 7 spins" is* $W \geqslant 3$*, so use a binomial model and its complement.*

$$W \sim \mathrm{Bin}(7, 0.4),$$

$$\mathrm{P}(\mathrm{Y}) = \mathrm{P}(W \geqslant 3) = 1 - \mathrm{P}(W \leqslant 2)$$

$$= 1 - \left[\binom{7}{0}(0.4)^0(0.6)^7 + \binom{7}{1}(0.4)(0.6)^6 + \binom{7}{2}(0.4)^2(0.6)^5\right] = 1 - 0.419904 = 0.580096.$$

Comparing the probabilities gives $0.5904 > 0.580096$.

Rhasthegreaterprobability.

(7)

www.mathswithmatt.co.uk