Maths with Matt

Physics and maths resources

Edexcel A Level Further Mathematics

9FM0/3B - Further Statistics 1

Saturday 18 June 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.

Question wording is reproduced from the user-provided 2022 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 8

Question 6 9

Question 7 11

www.mathswithmatt.co.uk

Question 1

Question 1

A researcher models the number of female cubs in a litter of size 4 by $B(4, 0.5)$. In 100 litters, the observed frequencies for $0, 1, 2, 3, 4$ female cubs are respectively $10, 33, 33, 15, 9$. The corresponding expected frequencies are $6.25, r, s, r, 6.25$.

(a) Find r and s. **(3)**

(b) At the 5% level, test whether $B(4, 0.5)$ is a suitable model. State hypotheses and the critical value. **(6)**

Model solution

(a) The expected frequency is the model probability multiplied by the 100 litters. The binomial probabilities are symmetric because $p = 0.5$.

$$r = 100\binom{4}{1}(0.5)^1(0.5)^3 = 25, \qquad s = 100\binom{4}{2}(0.5)^2(0.5)^2 = 37.5.$$

$$r = 25 \text{ and } s = 37.5.$$

(3)

(b) The data are frequencies in mutually exclusive categories with fully specified expected probabilities, so a chi-squared goodness-of-fit test is appropriate.

$H_0: X \sim B(4, 0.5)$ *is a suitable model,* $H_1: X \sim B(4, 0.5)$ *is not a suitable model.*

$$\chi^2 = \sum \frac{(O-E)^2}{E} = \frac{(10-6.25)^2}{6.25} + \frac{(33-25)^2}{25} + \frac{(33-37.5)^2}{37.5} + \frac{(15-25)^2}{25} + \frac{(9-6.25)^2}{6.25} = 10.56.$$

There are five categories and no parameters have been estimated from the data, so $\nu = 5 - 1 = 4$. *At 5%, the upper-tail critical value is* $\chi^2_{4,0.95} = 9.488$.

$$10.56 > 9.488.$$

The statistic is in the critical region, so reject H_0. *The discrepancies are too large to be attributed to random variation at the 5% level.*

There is sufficient evidence that $B(4, 0.5)$ *is not a suitable model.*

(6)

Question 2

Question 2

X has values $-5, -1, 0, 5, b$ with probabilities $0.3, 0.25, 0.1, 0.15, 0.2$, where $b > 5$.
(a) Find $E(X)$ in terms of b. **(1)**
Given $\text{Var}(X) = 34.26$, (b) find b. **(4)**
(c) Find $P(X^2 < 2 - 3X)$. **(4)**

Model solution

(a) Use $E(X) = \sum xP(X = x)$.

$$E(X) = -5(0.3) - 1(0.25) + 0(0.1) + 5(0.15) + 0.2b = 0.2b - 1.$$

$$E(X) = 0.2b - 1.$$

(1)

(b) The given variance links the unknown value to the second moment through $\text{Var}(X) = E(X^2) - [E(X)]^2$.

$$E(X^2) = 25(0.3) + 1(0.25) + 0 + 25(0.15) + 0.2b^2 = 11.5 + 0.2b^2.$$

$$34.26 = 11.5 + 0.2b^2 - (0.2b - 1)^2.$$

$$0.16b^2 + 0.4b - 23.76 = 0, \qquad b = 11 \text{ or } -13.5.$$

The condition $b > 5$ *selects the positive admissible root.*

$$b = 11.$$

(4)

(c) Test the inequality only at the possible values of the discrete random variable. Equivalently, solve $X^2 + 3X - 2 < 0$.

x	-5	-1	0	5	11
$x^2 + 3x - 2$	8	-4	-2	38	152

Only $x = -1$ *and* $x = 0$ *make the expression negative.*

$$P(X^2 < 2 - 3X) = P(X = -1) + P(X = 0) = 0.25 + 0.10 = 0.35.$$

$$0.35$$

(4)

Question 3

Question 3

In summer, team A receives calls randomly at rate 0.4 per day.
(a) Find the probability of at least 19 calls in 28 days. **(2)**
(b) In 250 summer days, estimate, using a Poisson approximation, the probability that at least 20 days have more than one call. **(4)**
(c) There are 8 calls in 42 winter days. Test at 5% whether the winter daily rate is lower. **(4)**
Team B receives independent summer calls at rate 0.2 per day. If C is the total in n days and $P(C = 0) < 0.001$, (d) find the minimum n. **(3)**
(e) State an assumption needed for the model. **(1)**

Model solution

(a) A count in a fixed interval from a constant-rate random process is Poisson. The mean over 28 days is $28(0.4) = 11.2$.

$$W \sim \text{Po}(11.2), \qquad P(W \geqslant 19) = 1 - P(W \leqslant 18) = 0.020776\ldots$$

0.0208 *(approximately)*

(2)

(b) First find the chance that one day is a ``more than 1 call'' day. Then the number of such days is binomial, which may be approximated by Poisson because there are many days and the event is relatively rare.

$$S \sim \text{Po}(0.4), \qquad P(S > 1) = 1 - P(S = 0) - P(S = 1) = 0.061551\ldots$$

$$X \sim \text{B}(250, 0.061551\ldots) \approx Y \sim \text{Po}(250 \times 0.061551\ldots) = \text{Po}(15.3879\ldots).$$

$$P(X \geqslant 20) \approx P(Y \geqslant 20) = 1 - P(Y \leqslant 19) = 0.14751\ldots$$

0.148 *(approximately)*

(4)

(c) Under the summer rate, the total over 42 days has mean $42(0.4) = 16.8$. *A lower-tailed Poisson test matches the claim that winter has fewer calls.*

$$H_0 : \lambda = 16.8 \quad (\textit{equivalently rate } 0.4 \textit{ per day}), \qquad H_1 : \lambda < 16.8.$$

$$U \sim \text{Po}(16.8), \qquad P(U \leqslant 8) = 0.0141\ldots < 0.05.$$

Reject H_0*; the observed total is sufficiently low for this one-tailed test.*

Thereissufficientevidenceatthe5%levelthatthedailycallrateislowerinwinter.

(4)

(d) Independent Poisson counts add, so the combined rate is $0.4 + 0.2 = 0.6$ *calls per day.*

$$C \sim \text{Po}(0.6n), \qquad P(C = 0) = e^{-0.6n} < 0.001.$$

$$-0.6n < \ln(0.001) \quad \Rightarrow \quad n > 11.513\ldots$$

The minimum integer is $n = 12$ *days.*

(3)

(e) A Poisson-process model needs a constant rate (and independent counts in non-overlapping intervals).

Assume the rate of calls per day is constant.

(1)

www.mathswithmatt.co.uk

Question 4

Question 4

A spinner lands blue with probability 0.11 on each spin.
(a) Find the probability of blue (i) first on spin 6, (ii) first before spin 6, (iii) exactly 4 times in the first 6 spins, (iv) for the fourth time on or before spin 6. **(2,2,2,4)**
Izana spins first and Zac second; the first blue wins. (b) Show that Zac's probability of winning is 0.471 to 3 significant figures. **(3)**

Model solution

(a)(i) This is a geometric setting: the first five spins must be non-blue, followed by blue.

$$(0.89)^5(0.11) = 0.06142\ldots$$

0.0614

(2)

(a)(ii) Use the complement: "first before spin 6" means at least one blue in the first five spins.

$$1 - (0.89)^5 = 0.44159\ldots$$

0.442

(2)

(a)(iii) The number of blues in six independent spins is binomial.

$$X \sim B(6, 0.11), \qquad P(X = 4) = \binom{6}{4}(0.11)^4(0.89)^2 = 0.001739\ldots$$

0.00174

(2)

(a)(iv) The fourth blue occurs on or before spin 6 exactly when there are at least four blues in the first six spins.

$$P(X \geqslant 4) = \sum_{r=4}^{6} \binom{6}{r}(0.11)^r(0.89)^{6-r} = 0.001827\ldots$$

0.00183

(4)

(b) Zac can win on spins $2, 4, 6, \ldots$. Before each such opportunity, a complete Izana--Zac pair must have contained no blue. This gives a geometric series.

$$P(\textit{Zac wins}) = (0.89)(0.11) + (0.89)^3(0.11) + (0.89)^5(0.11) + \cdots$$

$$= \frac{0.89(0.11)}{1-(0.89)^2} = 0.470899\ldots$$

$\mathrm{P}(\textit{Zac wins}) = 0.471$ *to 3 significant figures.*

(3)

www.mathswithmatt.co.uk

Question 5

Question 5

A random sample of 150 observations is taken from a geometric distribution with parameter 0.3.
Estimate $P(\overline{X} < 3.45)$. **(5)**

Model solution

For the geometric distribution on $1, 2, \ldots$, first obtain its mean and variance. Since the sample is large, the central limit theorem gives an approximate normal distribution for the sample mean.

$$\mu = \frac{1}{0.3} = \frac{10}{3}, \qquad \sigma^2 = \frac{1 - 0.3}{0.3^2} = \frac{70}{9}.$$

$$\overline{X} \overset{\text{approximately}}{\sim} N\left(\frac{10}{3}, \frac{70/9}{150}\right) = N\left(\frac{10}{3}, \frac{7}{135}\right).$$

$$P(\overline{X} < 3.45) = P\left(Z < \frac{3.45 - 10/3}{\sqrt{7/135}}\right) = P(Z < 0.5123\ldots) = 0.69579\ldots$$

0.696

(5)

www.mathswithmatt.co.uk

Question 6

Question 6

V has values 2, 3, 4 with probabilities 9/25, 12/25, 4/25.
(a) Show that $G_V(t) = t^2\left(\frac{2}{5}t + \frac{3}{5}\right)^2$. **(2)**
$G_W(t) = t\left(\frac{2}{5}t + \frac{3}{5}\right)^5$. (b) Use calculus to find (i) $E(W)$ and (ii) $\text{Var}(W)$. **(2,4)**
Given independence, (c) find $G_X(t)$ for $X = V + W$. **(2)**
For $Y = 2X + 3$, (d) find $G_Y(t)$ and (e) find $P(Y = 15)$. **(2,2)**

Model solution

(a) A probability generating function is $\sum P(V = v)t^v$. Factoring the resulting polynomial gives the requested form.

$$G_V(t) = \frac{9}{25}t^2 + \frac{12}{25}t^3 + \frac{4}{25}t^4 = t^2\left(\frac{3}{5} + \frac{2}{5}t\right)^2 = t^2\left(\frac{2}{5}t + \frac{3}{5}\right)^2.$$

$$\boxed{G_V(t) = t^2\left(\tfrac{2}{5}t + \tfrac{3}{5}\right)^2.}$$

(2)

(b)(i) For a PGF, $E(W) = G'_W(1)$. Differentiate using the product rule because G_W is a product.

$$G'_W(t) = \left(\frac{2}{5}t + \frac{3}{5}\right)^5 + t \cdot 5\left(\frac{2}{5}t + \frac{3}{5}\right)^4\left(\frac{2}{5}\right).$$

$$E(W) = G'_W(1) = 1 + 2 = 3.$$

$$\boxed{E(W) = 3.}$$

(2)

(b)(ii) The second derivative gives $E[W(W-1)]$, so $\text{Var}(W) = G''_W(1) + G'_W(1) - [G'_W(1)]^2$.

$$G''_W(t) = 4\left(\frac{2}{5}t + \frac{3}{5}\right)^4 + \frac{16}{5}t\left(\frac{2}{5}t + \frac{3}{5}\right)^3, \qquad G''_W(1) = \frac{36}{5}.$$

$$\text{Var}(W) = \frac{36}{5} + 3 - 3^2 = \frac{6}{5}.$$

$$\boxed{\text{Var}(W) = \tfrac{6}{5}.}$$

(4)

(c) For independent variables, PGFs multiply. Both PGFs have the same binomial factor, so the powers combine.

$$G_X(t) = G_V(t)G_W(t) = t^3\left(\frac{2}{5}t + \frac{3}{5}\right)^7.$$

$$G_X(t) = t^3\left(\tfrac{2}{5}t + \tfrac{3}{5}\right)^7.$$

(2)

(d) Since $Y = 2X + 3$, substitute t^2 into G_X and multiply by t^3.

$$G_Y(t) = t^3 G_X(t^2) = t^9\left(\frac{2}{5}t^2 + \frac{3}{5}\right)^7.$$

$$G_Y(t) = t^9\left(\tfrac{2}{5}t^2 + \tfrac{3}{5}\right)^7.$$

(2)

(e) The coefficient of t^{15} is required. In the binomial factor this means choosing the t^2 term three times.

$$P(Y = 15) = \binom{7}{3}\left(\frac{2}{5}\right)^3\left(\frac{3}{5}\right)^4 = \frac{22680}{78125} = 0.290304.$$

$$P(Y = 15) = 0.290304.$$

(2)

Question 7

Question 7

The weight X grams is normal with mean μ and standard deviation 30. A sample of 10 has $\bar{x} = 1020$.
(a) At 5%, test $H_0 : \mu = 1000$ against $H_1 : \mu \neq 1000$. **(4)**
For a second sample of size n, the one-sided critical region is $\bar{X} > c$ for $H_0 : \mu = 1000$ against $H_1 : \mu > 1000$. (b) Find an equation for c in terms of n. **(2)**
When $\mu = 1020$, the Type II error probability is 0.0050 to 2 significant figures. (c) Find n and c. **(5)**

Model solution

(a) The population standard deviation is known, so the sample mean under H_0 is normal with standard deviation $30/\sqrt{10}$. This is a two-tailed test because the alternative is "not equal".

$$\bar{X} \sim N\left(1000, \frac{30^2}{10}\right), \qquad z = \frac{1020 - 1000}{30/\sqrt{10}} = 2.108\ldots$$

$$P(\bar{X} > 1020) = 0.0175\ldots < 0.025 \quad \textit{(equivalently, } 2.108 > 1.96\textit{).}$$

Reject H_0. The sample mean lies in the two-sided 5% critical region.

There is evidence that the mean flour weight is not 1000 g.

(4)

(b) For an upper-tailed 5% test, c is the 95th percentile under H_0, so use $z = 1.6449$.

$$\frac{c - 1000}{30/\sqrt{n}} = 1.6449 \qquad \Rightarrow \qquad c = 1000 + \frac{49.347}{\sqrt{n}}.$$

$$c = 1000 + \frac{49.347}{\sqrt{n}}.$$

(2)

(c) A Type II error at true mean 1020 is failing to reject, namely $\bar{X} \leqslant c$. Standardise using the distribution under $\mu = 1020$ and use the lower-tail probability 0.0050, whose z-value is -2.5758.

$$\frac{c - 1020}{30/\sqrt{n}} = -2.5758.$$

$$\frac{1000 + 49.347/\sqrt{n} - 1020}{30/\sqrt{n}} = -2.5758 \quad \Rightarrow \quad \sqrt{n} = \frac{49.347 + 77.274}{20} = 6.331\ldots$$

$$n = 40.08\ldots,$$

The calculation gives $n = 40.08\ldots$. Following the mark-scheme rounding, take $n = 40$. If 0.0050 is instead treated as a maximum Type II error probability (rather than a value rounded to 2 significant figures), round up and take $n = 41$. Substituting into the critical-value equation gives

$$c = 1000 + \frac{49.347}{\sqrt{40}} = 1007.80\ldots$$

Using the mark-scheme rounding, $n = 40$ *and* $c \approx 1007.8$ *g (about* 1010 *g). The conservative alternative is* $n = 41$ *and* $c \approx 1007.7$ *g.*

(5)

www.mathswithmatt.co.uk