Maths with Matt

Physics and maths resources

Edexcel A Level Further Mathematics

9FM0/3B - Further Statistics 1

Friday 14 June 2024

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 2024 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 2024 paper for study/reference. Pearson retains copyright in the original examination material.

Contents

Question 1 2

Question 5 3

Question 6 5

Question 7 7

Question 2 9

Question 3 10

Question 4 11

www.mathswithmatt.co.uk

Question 1

Question 1

The discrete random variable X has values $-1, 0, 1, 3, 5$ with respective probabilities $0.2, 0.1, 0.2, 0.25, 0.25$. (a) Find $\mathrm{Var}(X)$. (b) Find $\mathrm{Var}(X^2)$.

Model solution

(a) Use $\mathrm{Var}(X) = \mathrm{E}(X^2) - [\mathrm{E}(X)]^2$. ***The probability table gives both moments directly.***

$$\mathrm{E}(X) = (-1)(0.2) + 0(0.1) + 1(0.2) + 3(0.25) + 5(0.25) = 2,$$

$$\mathrm{E}(X^2) = 1(0.2) + 0 + 1(0.2) + 9(0.25) + 25(0.25) = 8.9.$$

$$\mathrm{Var}(X) = 8.9 - 2^2 = 4.9.$$

$$\mathrm{Var}(X) = 4.9.$$

(3)

(b) Apply the same variance identity to the random variable X^2. ***Its second moment is*** $\mathrm{E}((X^2)^2) = \mathrm{E}(X^4)$.

$$\mathrm{E}(X^4) = 1(0.2) + 0 + 1(0.2) + 81(0.25) + 625(0.25) = 176.9.$$

$$\mathrm{Var}(X^2) = \mathrm{E}(X^4) - [\mathrm{E}(X^2)]^2 = 176.9 - (8.9)^2 = 97.69.$$

$$\mathrm{Var}(X^2) = 97.69 \text{ ***(or 97.7)***.}$$

(3)

Question 5

Question 5

The required defective proportion is no more than 3%. A random sample of n components contains D_n defectives. (a) State hypotheses for testing whether the proportion is greater than 0.03. Test A uses $n = 100$ and rejects for $D_{100} \geqslant 5$. Test B uses $n = 80$, rejects for $D_{80} \geqslant 5$, accepts for $D_{80} \leqslant 3$, and when $D_{80} = 4$ takes a second sample of 80 and rejects if it has at least one defective. (b) Find the size of A. (c) Find the size of B. Given $p = 0.06$, (d) find the power of A and the expected number sampled in B. Given the power of B is 0.713, (e) recommend a test.

Model solution

(a) The management requirement gives the null value, and the concern is specifically an increase in the defective proportion, so this is a one-tailed test.

$$H_0 : p = 0.03, \qquad H_1 : p > 0.03.$$

(1)

(b) The size is the probability of rejection when H_0 is true. Under H_0, $D_{100} \sim B(100, 0.03)$.

$$P(D_{100} \geqslant 5) = 1 - P(D_{100} \leqslant 4) = 0.18214\ldots$$

Size of test A $= 0.182$ (approximately).

(2)

(c) Under H_0, let $D, D' \sim B(80, 0.03)$ independently represent the first and second samples. Rejection happens immediately for $D \geqslant 5$, or after the follow-up route $D = 4$ and $D' \geqslant 1$.

$$\begin{aligned} P(\textit{reject}) &= P(D \geqslant 5) + P(D = 4)P(D' \geqslant 1) \\ &= 0.09279\ldots + (0.12654\ldots)(0.91255\ldots) \\ &= 0.20826\ldots \end{aligned}$$

Size of test B $= 0.208$ (approximately).

(3)

(d)(i) Power is the probability of rejecting when the actual defective proportion is 0.06. For test A, $D_{100} \sim B(100, 0.06)$.

$$P(D_{100} \geqslant 5) = 0.72322\ldots$$

(i) Power of test A $= 0.723$ (approximately).

(d)(ii) Test B always takes the first 80 components. It takes another 80 only when the first sample has exactly four defectives. With $Y \sim B(80, 0.06)$,

$$E(\textit{number sampled}) = 80 + 80P(Y = 4) = 80 + 80(0.1859\ldots) = 94.87\ldots$$

(ii) Expected number of components sampled $= 94.9$ (approximately), or 95.

(3)

(e) Compare both effectiveness and sampling cost. Test B has similar size and power to test A, but its expected sample size when $p = 0.06$ is below 100.

Use test B, because it has similar size and power but requires fewer components on average.

(1)

www.mathswithmatt.co.uk

Question 6

Question 6

The random variable X has probability generating function $G_X(t) = 1/\sqrt{4-3t}$. (a) Use calculus to find $\mathrm{Var}(X)$. (b) Find the exact value of $P(X \leqslant 2)$. Independent X_1, X_2 have the distribution of X, and $Y = X_1 + X_2 + 1$. (c) Find $G_Y(t)$ and state the distribution of Y. (d) Find $P(X_1 + X_2 > 5)$.

Model solution

(a) For a probability generating function, $G'_X(1) = E(X)$ *and* $G''_X(1) = E[X(X-1)]$. *Differentiate twice so that the variance can be found from* $E(X^2) = E[X(X-1)] + E(X)$.

$$G'_X(t) = \frac{3}{2}(4-3t)^{-3/2}, \qquad E(X) = G'_X(1) = \frac{3}{2}.$$

$$G''_X(t) = \frac{27}{4}(4-3t)^{-5/2}, \qquad E[X(X-1)] = G''_X(1) = \frac{27}{4}.$$

$$E(X^2) = \frac{27}{4} + \frac{3}{2} = \frac{33}{4},$$

$$\mathrm{Var}(X) = E(X^2) - [E(X)]^2 = \frac{33}{4} - \left(\frac{3}{2}\right)^2 = 6.$$

$$\mathrm{Var}(X) = 6.$$

(6)

(b) The coefficient of t^r *in the pgf is* $P(X = r)$. *Expand about* $t = 0$ *using* $(1-u)^{-1/2} = 1 + \frac{1}{2}u + \frac{3}{8}u^2 + \cdots$, *with* $u = 3t/4$.

$$G_X(t) = \frac{1}{2}\left(1 - \frac{3t}{4}\right)^{-1/2} = \frac{1}{2} + \frac{3}{16}t + \frac{27}{256}t^2 + \cdots$$

$$P(X \leqslant 2) = \frac{1}{2} + \frac{3}{16} + \frac{27}{256} = \frac{203}{256}.$$

$$P(X \leqslant 2) = 203/256.$$

(4)

(c) Independence means the pgf of a sum is the product of the pgfs. Adding 1 multiplies the pgf by t.

$$G_Y(t) = tG_{X_1}(t)G_{X_2}(t) = \frac{t}{4-3t} = \frac{\frac{1}{4}t}{1-\frac{3}{4}t}.$$

This is the pgf of a geometric variable on $1, 2, \ldots$, *with success probability* $1/4$.

$$Y \sim \mathrm{Geo}(1/4).$$

(4)

(d) Since $Y = X_1 + X_2 + 1$*, the required event is* $Y > 6$*. For* $Y \sim \text{Geo}(1/4)$*,* $P(Y > 6)$ *is the probability of six failures before the first success.*

$$P(X_1 + X_2 > 5) = P(Y > 6) = \left(\frac{3}{4}\right)^6 = 0.177978\ldots$$

0.178 *(approximately).*

(2)

www.mathswithmatt.co.uk

Question 7

Question 7

Each independent game of Pento wins a prize with probability $1/5$. (a) Sam plays 20 games: find the probability of at least 4 prizes. (b) Find the probability that Tessa wins her fourth prize on game 20. Rama charges £1 per game and pays £2 for a first prize, £4 for a second prize and £$(2w)$ for a wth prize. (c) Show Sam's expected profit from n games is £$(n^2 - 16n)/25$. Given $n = 15$, (d) find the probability Sam does not make a loss. (e) If Tessa plays until her rth prize, find her expected profit.

Model solution

(a) Let X be Sam's number of prizes. There is a fixed number of independent games with constant win probability, so $X \sim B(20, 0.2)$.

$$P(X \geqslant 4) = 1 - P(X \leqslant 3) = 0.58855\ldots$$

0.589 *(approximately).*

(2)

(b) For the fourth win to occur on game 20, there must be exactly three wins in the first 19 games and then a win on game 20. This is a negative-binomial situation.

$$P(\textit{fourth win on game 20}) = \binom{19}{3}(0.2)^4(0.8)^{16} = 0.043639\ldots$$

0.0436 *(approximately).*

(2)

(c) Let S be the number of prizes Sam wins in n games. Then $S \sim B(n, 0.2)$. If he wins S prizes, his total winnings are the arithmetic series $2 + 4 + \cdots + 2S = S(S+1)$, while his cost is n pounds.

$$\textit{profit} = S(S+1) - n = S^2 + S - n.$$

Use $E(S^2) = Var(S) + [E(S)]^2$:

$$E(S) = 0.2n, \qquad Var(S) = 0.16n,$$

$$E(S^2) = 0.16n + 0.04n^2.$$

$$E(\textit{profit}) = 0.16n + 0.04n^2 + 0.2n - n = \frac{n^2 - 16n}{25}.$$

Sam's expected profit is £$\dfrac{n^2 - 16n}{25}$.

(6)

(d) With $n = 15$*, a non-loss requires* $S^2 + S - 15 \geqslant 0$*. Solve this inequality before finding the binomial probability.*

$$S \geqslant \frac{-1 + \sqrt{61}}{2} = 3.405\ldots$$

As S *is an integer, the first non-loss is at* $S = 4$*. Thus, with* $S \sim B(15, 0.2)$*,*

$$P(\textit{no loss}) = P(S \geqslant 4) = 0.35183\ldots$$

0.352 *(approximately).*

(4)

(e) Let T *be the number of games needed for Tessa's* r*th win. Then* $T \sim \text{negB}(r, 0.2)$*, so* $E(T) = r/0.2 = 5r$*. Her prize money is* $2 + 4 + \cdots + 2r = r(r+1)$*.*

$$E(\textit{profit}) = r(r+1) - E(T) = r(r+1) - 5r = r^2 - 4r.$$

Tessa's expected profit is £$(r^2 - 4r)$*.*

(4)

Question 2

Question 2

The number of errors made by a secretary is Poisson with mean 2.4 per 100 words. (a) For 100 words, find the probabilities of (i) exactly 3 errors and (ii) fewer than 2 errors. After a holiday, a 250-word piece is examined. (b) State hypotheses and find a 5% critical region for a test of whether the rate has changed. (c) Find the probability of a Type I error.

Model solution

(a) For a 100-word piece, let $E \sim \text{Po}(2.4)$. *A Poisson model is appropriate because the question gives a mean count over a fixed length.*

$$P(E = 3) = \frac{e^{-2.4}2.4^3}{3!} = 0.20901\ldots$$

(i) 0.209

$$P(E < 2) = P(E = 0) + P(E = 1) = e^{-2.4}(1 + 2.4) = 0.30844\ldots$$

(ii) 0.308

(2)

(b) "Has changed" means departures in either direction matter, so use a two-tailed test. In 250 words the null mean is $2.5 \times 2.4 = 6$.

$$H_0 : \lambda = 6 \quad \textit{and} \quad H_1 : \lambda \neq 6, \qquad E \sim \text{Po}(6) \textit{ under } H_0.$$

Choose tails whose total probability is at most 0.05*:*

$$P(E \leqslant 1) = 0.01735\ldots, \qquad P(E \geqslant 12) = 0.02009\ldots.$$

Including $E = 2$ *or* $E = 11$ *would make the relevant tail too large, so these are the most extreme allowable cut-offs.*

Critical region: $E \leqslant 1$ *or* $E \geqslant 12$.

(4)

(c) A Type I error is rejecting H_0 *when its model is true, so its probability is the probability of the critical region under* $\text{Po}(6)$.

$$0.017351\ldots + 0.020092\ldots = 0.037443\ldots$$

$P(\textit{Type I error}) = 0.0375$ *(approximately).*

(1)

Question 3

Question 3

Tisam's favourite-colour survey is

Year group/Colour	Red	Blue	Green	Yellow	Black	Total
1–5	34	15	14	22	3	88
6–9	23	32	12	9	8	84
10–12	5	28	19	8	8	68
Total	62	75	45	39	19	240

She tests for association. (a) State hypotheses. (b) Identify the one cell that need be checked for an expected frequency below 5, and calculate it. (c) Given test statistic 38.449, complete the test at 1%.

Model solution

(a) This is a χ^2 test of association because both variables are categorical.

H_0: *favourite colour and year group are independent (there is no association).*
H_1: *favourite colour and year group are associated.*

(1)

(b) An expected frequency is $\dfrac{(\textit{row total})(\textit{column total})}{\textit{grand total}}$. *Therefore the smallest expectation must be at the intersection of the smallest row and column totals.*

(i) The Year 10--12, Black cell, since 68 *and* 19 *are the smallest row and column totals.*

$$\mathrm{E}(\textit{Year 10--12, Black}) = \frac{68 \times 19}{240} = 5.3833\ldots$$

(ii) 5.38 *(approximately), so all expected frequencies are at least* 5.

(2)

(c) There are $(3-1)(5-1) = 8$ *degrees of freedom. Compare the supplied statistic with the upper-tail 1% critical value.*

$$\chi^2_{0.99,8} = 20.090, \qquad 38.449 > 20.090.$$

The statistic is in the critical region, so reject the independence model.

Thereissignificantevidenceatthe1%levelofanassociationbetweenfavouriteco

(3)

Question 4

Question 4

Geethaka rolls a fair die repeatedly until a 3 occurs; X is the number of rolls. (a) Model X. (b) Show $P(X \leqslant 3) = 91/216$. (c) After 64 mornings estimate $P(5.6 < \overline{X} < 7.2)$. (d) Nira obtains $x = 16$ when testing whether $P(3) < 1/6$, at 5%.

Model solution

(a) Each roll is independent and the experiment stops at the first success (a 3), which is exactly the setting for a geometric distribution.

$$X \sim \text{Geo}(1/6).$$

(1)

(b) The complement of $X \leqslant 3$ is no 3 in the first three rolls.

$$P(X \leqslant 3) = 1 - P(X > 3) = 1 - \left(\frac{5}{6}\right)^3 = \frac{91}{216}.$$

$$P(X \leqslant 3) = 91/216.$$

(2)

(c) The sample size is large, so use the central limit theorem for the mean of geometric variables. For $X \sim \text{Geo}(p)$, $E(X) = 1/p$ and $\text{Var}(X) = (1-p)/p^2$.

$$E(X) = 6, \qquad \text{Var}(X) = 30,$$

$$\overline{X} \sim N\left(6, \frac{30}{64}\right).$$

Standardising the two bounds gives

$$P(5.6 < \overline{X} < 7.2) = \Phi\left(\frac{7.2-6}{\sqrt{30/64}}\right) - \Phi\left(\frac{5.6-6}{\sqrt{30/64}}\right) = 0.68064\ldots$$

$$0.681 \textit{ (approximately).}$$

(5)

(d) A probability lower than $1/6$ makes a long wait for a 3 more likely, so this is an upper-tailed test in X.

$$H_0 : p = \frac{1}{6}, \qquad H_1 : p < \frac{1}{6},$$

$$P(X \geqslant 16 \mid H_0) = \left(\frac{5}{6}\right)^{15} = 0.0649\ldots$$

This is the one-tailed p-value. Since it exceeds 0.05, the observation is not sufficiently unusual under a fair die.

Do not reject H_0: there is insufficient evidence at the 5% level that the probability of rolling a 3 is less than $1/6$ (or that the die is biased in this way).

(4)

www.mathswithmatt.co.uk