Maths with Matt

Physics and maths resources

Edexcel A Level Further Mathematics

9FM0/4B - Further Statistics 2

October 2020

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 2020 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 2020 paper for study/reference. Pearson retains copyright in the original examination material.

Contents

Question 1 2

Question 5 3

Question 6 4

Question 2 5

Question 3 6

Question 4 7

Question 7 8

Question 8 10

www.mathswithmatt.co.uk

Question 1

Question 1

Gina samples 7 packages from company A and 10 independent packages from company B. The summaries are $a = 300$, $S_{aa} = 145\,496$, $b = 233.4$ and $S_{bb} = 56\,364.4$. Assuming both populations are normal, test at the 5% level Gina's belief that company A has the greater variance.

Model solution

This is a comparison of two variances from independent normal populations, so an F-test is appropriate. Gina's belief makes this a one-tailed test.

$$H_0 : \sigma_A^2 = \sigma_B^2, \qquad H_1 : \sigma_A^2 > \sigma_B^2.$$

First calculate the two unbiased sample variances, using $s^2 = S_{xx}/(n-1)$.

$$s_A^2 = \frac{145\,496}{7-1} = 24\,249.333\ldots, \qquad s_B^2 = \frac{56\,364.4}{10-1} = 6262.711\ldots$$

$$F = \frac{s_A^2}{s_B^2} = \frac{24\,249.333\ldots}{6262.711\ldots} = 3.872\ldots$$

Under H_0, this has an $F_{6,9}$ distribution. For a 5% upper-tailed test the critical value is 3.37. The statistic lies in the critical region because $3.872 > 3.37$.

Reject H_0. At the 5% level there is sufficient evidence that the variance of package weights from company A is greater than that from company B, supporting Gina's belief.

(6)

Question 5

Question 5

$$f(x) = \begin{cases} k(1-\cos x), & 0 \leqslant x \leqslant 2\pi, \\ 0, & \text{otherwise.} \end{cases}$$

(a) Show that $k = \dfrac{1}{2\pi}$.

$Y \sim N(\mu, \sigma^2)$, $E(Y) = E(X)$, and the density g of Y satisfies $g(\mu) = f(\mu)$.

(b) Find the exact value of σ.

(c) Calculate the error in using $P(\pi/2 < Y < 3\pi/2)$ as an approximation to $P(\pi/2 < X < 3\pi/2)$.

Model solution

(a) A probability density must have total area 1, so integrate f over its support.

$$1 = \int_0^{2\pi} k(1-\cos x)\,dx = k\,[x - \sin x]_0^{2\pi} = k(2\pi).$$

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

(3)

(b) The density is symmetric about $x = \pi$, so $E(X) = \pi$, hence $\mu = \pi$. Evaluate the two densities at this common mean.

$$f(\mu) = f(\pi) = \frac{1}{2\pi}(1-\cos\pi) = \frac{1}{\pi}, \qquad g(\mu) = \frac{1}{\sigma\sqrt{2\pi}}.$$

$$\frac{1}{\sigma\sqrt{2\pi}} = \frac{1}{\pi} \implies \sigma^2 = \frac{\pi}{2}.$$

$$\sigma = \sqrt{\frac{\pi}{2}}.$$

(3)

(c) Calculate the exact probability for X by integrating its density, then calculate the normal probability for Y.

$$P\left(\frac{\pi}{2} < X < \frac{3\pi}{2}\right) = \frac{1}{2\pi}\,[x - \sin x]_{\pi/2}^{3\pi/2} = \frac{\pi+2}{2\pi} = 0.81830\ldots$$

$$P\left(\frac{\pi}{2} < Y < \frac{3\pi}{2}\right) = P\left(-\sqrt{\frac{\pi}{2}} < Z < \sqrt{\frac{\pi}{2}}\right) = 0.7899\ldots$$

The error is $0.81830\ldots - 0.7899\ldots = 0.0284$ (approximately).

(4)

Question 6

Question 6

Kim's sample of 25 normally distributed steel-rod lengths gives the 95% t-interval $(1.193, 1.367)$ metres. The usual standard deviation is 0.175 m.
(a) At the 10% level, test whether Kim's sample suggests that the standard deviation differs from 0.175 m.
(b) Using Kim's sample and the normal distribution with standard deviation 0.175 m, find a 95% confidence interval for the mean.

Model solution

(a) The centre of Kim's interval is the sample mean, and its half-width lets us recover the sample standard deviation.

$$\bar{x} = \frac{1.193 + 1.367}{2} = 1.280.$$

$$1.367 = 1.280 + 2.064\frac{s}{\sqrt{25}} \quad \Longrightarrow \quad s = 0.210755\ldots$$

To test a population variance from a normal population, use a chi-squared test. The wording requires a two-tailed test.

$$H_0 : \sigma = 0.175, \qquad H_1 : \sigma \neq 0.175.$$

$$\frac{(n-1)s^2}{\sigma_0^2} \sim \chi^2_{24} \quad \textit{under } H_0, \qquad \chi^2 = \frac{24(0.210755\ldots)^2}{0.175^2} = 34.809\ldots$$

At the 10% two-tailed level, the critical region is $\chi^2 < 13.848$ or $\chi^2 > 36.415$. The test statistic is not in this region.

Do not reject H_0. There is insufficient evidence that the standard deviation differs from 0.175 m; retaining the usual standard deviation is reasonable.

(9)

(b) Here the standard deviation is assumed known, so use a normal critical value rather than t.

$$\bar{x} \pm 1.96\frac{0.175}{\sqrt{25}} = 1.280 \pm 0.0686 = (1.2114,\ 1.3486).$$

A 95% confidence interval is $(1.211$ m, 1.349 m$)$, approximately $(1.21$ m, 1.35 m$)$.

(3)

Question 2

Question 2

For a normal population of jar weights, a sample of 8 gives $\sum x = 3552$ and $\sum x^2 = 1\,577\,314$.
(a) Calculate a 95% confidence interval for the population mean.
(b) The labels state that the average contents weigh 440 g. State whether Jemima should be concerned.

Model solution

(a) The population standard deviation is unknown and the sample is small, so use a t-confidence interval with 7 degrees of freedom.

$$\bar{x} = \frac{3552}{8} = 444, \qquad s^2 = \frac{\sum x^2 - n\bar{x}^2}{n-1} = \frac{1\,577\,314 - 8(444)^2}{7} = 32.2857\ldots$$

For 95% confidence, $t_{7,0.025} = 2.365$. *Therefore*

$$\mu \in \bar{x} \pm t_{7,0.025}\frac{s}{\sqrt{n}} = 444 \pm 2.365\sqrt{\frac{32.2857\ldots}{8}} = (439.248\ldots, 448.752\ldots).$$

A 95% confidence interval is $(439.25\ g, 448.75\ g)$, *or* $(439\ g, 449\ g)$ *to the nearest gram.*

(5)

(b) Compare the labelled value with the interval: a value inside a 95% confidence interval is consistent with the sample evidence.

Jemima need not be concerned: $440\ g$ *lies in the 95% confidence interval.*

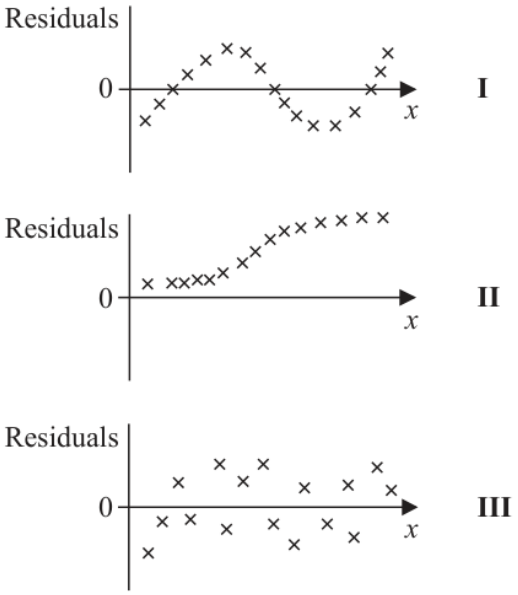
(1)

Question 3

Question 3

For each of the three residual plots below, state whether it is feasible and, if feasible, whether it suggests a linear or non-linear relationship between y and x, giving a reason.

Below are 3 sketches from some students of the residuals from their linear regressions of y on x.



Model solution

For a least-squares regression with an intercept, the residuals sum to zero. A suitable linear model also has residuals randomly scattered about zero; a systematic pattern suggests that a curve would model the relationship better.

Plot I is feasible, but it suggests a non-linear relationship. Its residuals follow a systematic curved pattern rather than being randomly scattered about zero.

I is feasible and suggests a non-linear relationship.

(2)

Plot II is not feasible as a residual plot from a linear regression. All its residuals are above zero, so they cannot sum to zero.

II is not feasible.

(2)

Plot III is feasible and suggests a linear relationship. The points are randomly scattered on both sides of zero, with no discernible pattern.

III is feasible and suggests a linear relationship.

(2)

Question 4

Question 4

A biased coin has probability p of heads. Simon observes X heads in n spins and Taruni observes Y heads in m spins, where $m \neq n$. They propose

$$S = \frac{X+Y}{n+m}, \qquad T = \frac{1}{2}\left(\frac{X}{n} + \frac{Y}{m}\right).$$

(a) Show that both estimators are unbiased for p.
(b) Prove that S is the better estimator for all permitted m, n.

Model solution

(a) Head counts from a fixed number of independent coin spins are binomial, so their expectations can be used to test unbiasedness.

$$X \sim B(n,p), \quad Y \sim B(m,p), \qquad E(X) = np, \quad E(Y) = mp.$$

$$E(S) = \frac{E(X)+E(Y)}{n+m} = \frac{np+mp}{n+m} = p,$$

$$E(T) = \frac{1}{2}\left(\frac{E(X)}{n} + \frac{E(Y)}{m}\right) = \frac{1}{2}(p+p) = p.$$

Both S and T are unbiased estimators of p.

(3)

(b) Since both estimators are unbiased, the one with the smaller variance is better. The two sets of spins are independent, so variances add.

$$\mathrm{Var}(S) = \frac{\mathrm{Var}(X)+\mathrm{Var}(Y)}{(n+m)^2} = \frac{p(1-p)}{n+m},$$

$$\mathrm{Var}(T) = \frac{1}{4}\left(\frac{\mathrm{Var}(X)}{n^2} + \frac{\mathrm{Var}(Y)}{m^2}\right) = \frac{p(1-p)(m+n)}{4mn}.$$

As $0 < p < 1$, $p(1-p) > 0$, and m, n are positive, denominators may be cleared without reversing inequalities.

$$\mathrm{Var}(S) < \mathrm{Var}(T) \iff \frac{1}{m+n} < \frac{m+n}{4mn} \iff 4mn < (m+n)^2$$

$$\iff 0 < m^2 - 2mn + n^2 = (m-n)^2.$$

The final inequality is true because $m \neq n$.

$\mathrm{Var}(S) < \mathrm{Var}(T)$ for every permitted m, n, so S is always the better estimator.

(4)

Question 7

Question 7

Large-panel lengths are $N(198, 5^2)$ cm and small-panel lengths are $N(74, 3^2)$ cm.
(a) Find the probability that the total length of 3 large panels exceeds that of 8 small panels.
(b) Find the probability that one large panel is more than 8/3 times the length of one small panel.
(c) Rosa buys 5 large panels at £80 each. If their total is below 1000 cm, she also buys one small panel at £30. Decide whether her expected cost is cheaper than buying 14 small panels.

Model solution

(a) Sums of independent normal variables are normal. Form the difference of the two totals so that the required comparison becomes a probability against zero.

$$L_1 + L_2 + L_3 \sim N(3 \times 198, 3 \times 5^2) = N(594, 75),$$

$$S_1 + \cdots + S_8 \sim N(8 \times 74, 8 \times 3^2) = N(592, 72).$$

$$D = (L_1 + L_2 + L_3) - (S_1 + \cdots + S_8) \sim N(2, 147).$$

$$P(D > 0) = P\left(Z > \frac{0-2}{\sqrt{147}}\right) = 0.56551\ldots$$

The probability is 0.566 *(3 s.f.).*

(6)

(b) Rearrange the stated comparison as $L - \frac{8}{3}S > 0$*, which again can be modelled by one normal random variable.*

$$W = L - \frac{8}{3}S \sim N\left(198 - \frac{8}{3}(74),\ 5^2 + \left(\frac{8}{3}\right)^2 3^2\right) = N\left(\frac{2}{3}, 89\right).$$

$$P\left(L > \frac{8}{3}S\right) = P(W > 0) = P\left(Z > \frac{-2/3}{\sqrt{89}}\right) = 0.528168\ldots$$

The probability is 0.528 *(3 s.f.).*

(5)

(c) The extra small panel is bought precisely when the total of the five large panels is below 1000 cm. Find that probability before taking the expected cost.

$$F = L_1 + \cdots + L_5 \sim N(5 \times 198, 5 \times 5^2) = N(990, 125).$$

$$P(F < 1000) = P\left(Z < \frac{1000 - 990}{\sqrt{125}}\right) = 0.814455\ldots$$

The cost is £430 when the small panel is needed and £400 otherwise.

$$\mathrm{E}(\textit{cost}) = 430(0.814455\ldots) + 400(1 - 0.814455\ldots) = \$424.43\ldots$$

$$\textit{Cost of 14 small panels} = 14(30) = \$420.$$

Rosa's expected cost is about \$424.43, so it is not cheaper than buying 14 small panel

(6)

www.mathswithmatt.co.uk

Question 8

Question 8

The radius X cm has a rectangular distribution on $[0, \pi]$.
(a) Find P(area of the circle > 10 cm^2).
(b) State whether the median area is greater or less than 10 cm^2, giving a reason.
For triangle OAB, OA = OB = X and $\angle AOB = X$ radians.
(c) Use algebraic integration to find the exact expected area of triangle OAB.

Model solution

(a) The circle area is πX^2. Since $X \geqslant 0$, taking square roots preserves the inequality. A rectangular distribution on $[0, \pi]$ has constant density $1/\pi$.

$$P(\pi X^2 > 10) = P\left(X > \sqrt{\frac{10}{\pi}}\right) = \frac{\pi - \sqrt{10/\pi}}{\pi} = 0.43209\ldots$$

$$P(\textit{area} > 10\ \textit{cm}^2) = 0.432\ \textit{(3 s.f.).}$$

(3)

(b) By definition, the probability of exceeding the median is 0.5. The probability of exceeding 10 cm^2 found in part (a) is less than 0.5.

The median area is less than 10 cm^2.

(1)

(c) The area of a triangle with two sides x, x and included angle x is $\frac{1}{2}x^2 \sin x$. Use $E[h(X)] = \int h(x) f_X(x)\, dx$, then integrate by parts twice.

$$E(\textit{area}) = \frac{1}{2\pi}\int_0^{\pi} x^2 \sin x\, dx.$$

$$\int x^2 \sin x\, dx = -x^2\cos x + \int 2x\cos x\, dx = -x^2\cos x + 2x\sin x - \int 2\sin x\, dx = -x^2\cos x + 2x\sin x + 2\cos x.$$

$$E(\textit{area}) = \frac{1}{2\pi}\left[-x^2\cos x + 2x\sin x + 2\cos x\right]_0^{\pi} = \frac{1}{2\pi}(\pi^2 - 4).$$

$$E(\textit{area of OAB}) = \frac{\pi}{2} - \frac{2}{\pi}\ \textit{cm}^2.$$

(7)