Maths with Matt

Physics and maths resources

Edexcel A Level Further Mathematics

9FM0/4B - Further Statistics 2

Monday 24 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 6 4

Question 7 6

Question 8 8

Question 2 9

Question 3 10

Question 4 11

Question 5 12

www.mathswithmatt.co.uk

Question 1

Question 1

Two students fit the models $t = a + b\sqrt{h}$ and $t = c + dh$ to the water-bottle data. For $v = t - 40$ and $w = \sqrt{h}$, $\sum v = 626$, $\sum v^2 = 64678$, $\sum w = 22.47$, $S_{ww} = 4.52$, and $S_{vw} = -338.83$. Answer the four parts about the regression model t on $\sqrt{h}$.

Model solution

(a) Because the proposed model is linear in $w = \sqrt{h}$, first find the regression line of v on w, then undo the coding.

$$b = \frac{S_{vw}}{S_{ww}} = \frac{-338.83}{4.52} = -74.963\ldots$$

$$\overline{v} = \frac{626}{10} = 62.6, \qquad \overline{w} = \frac{22.47}{10} = 2.247,$$

$$a = \overline{v} - b\overline{w} = 62.6 - (-74.963\ldots)(2.247) = 231.040\ldots$$

Thus $v = 231.040\ldots - 74.963\ldots w$. *Since* $v = t - 40$ *and* $w = \sqrt{h}$,

$$t - 40 = 231.040\ldots - 74.963\ldots\sqrt{h}.$$

$$t = 271.04\ldots - 74.96\ldots\sqrt{h}.$$

(4)

(b) A residual is observed value minus the value predicted by the fitted model. Here $h = 9$, *so* $\sqrt{h} = 3$.

$$\widehat{t} = 271.04\ldots - 74.96\ldots(3) = 46.153\ldots$$

$$e = 47 - 46.153\ldots = 0.8466\ldots$$

Residual $= 0.85$ *seconds.*

(2)

(c) For a linear regression, $\text{RSS} = \sum v^2 - (\sum v)^2/n - S_{vw}^2/S_{ww}$. *Coding by a shift does not change residuals, so this is also the RSS for the original model.*

$$\text{RSS} = 64678 - \frac{626^2}{10} - \frac{(-338.83)^2}{4.52} = 90.89\ldots$$

$\text{RSS} = 90.9$ *(approximately).*

(2)

(d) The smaller RSS represents the closer overall fit.

Student A's model is more suitable, because $90.9 < 980$.

(1)

www.mathswithmatt.co.uk

Question 6

Question 6

Eight pairs of twin lambs were used; one twin in each pair received a food supplement. Their weight gains (with minus without supplement) are to be analysed using a paired t-procedure. Answer all five parts.

Model solution

(a) A two-sample t-test requires independent samples, whereas the two observations in each twin pair are linked.

A two-sample t-test is unsuitable because the samples are not independent.

(1)

(b) Relevant characteristics should be controlled so that differences in gain are not caused by a confounding factor.

For example, control birth weight and gender.

(2)

(c) The paired test is applied to the within-pair differences, not to the two original samples.

Assume that the distribution of the differences in weight gain between twin lambs is normal.

(1)

(d) Use the paired differences $W =$ gain with supplement $-$ gain without supplement, then form a t-interval because the population standard deviation is unknown and $n = 8$.

pair	A	B	C	D	E	F	G	H
w	-0.9	0.5	0.8	0.2	0.8	0.3	0.1	-0.1

$$\overline{w} = \frac{1.7}{8} = 0.2125, \qquad s^2 = 0.3041\ldots$$

For 7 degrees of freedom, the central 98% critical value is $t_{7,0.01} = 2.998$. Hence

$$\overline{w} \pm t\frac{s}{\sqrt{8}} = 0.2125 \pm 2.998\sqrt{\frac{0.3041\ldots}{8}} = 0.2125 \pm 0.5845\ldots$$

98% CI for μ_W: $(-0.372, 0.797)$ kg (approximately).

(5)

(e) The claim is that the mean difference exceeds 200 g, which is 0.2 kg; this gives a one-tailed alternative. If 0.2 is in the corresponding 98% interval, the data do not support a 1% one-tailed claim.

$$H_0 : \mu_W = 0.2 \text{ kg}, \qquad H_1 : \mu_W > 0.2 \text{ kg}.$$

Since

$$-0.372 < 0.2 < 0.797,$$

Do not reject H_0: there is insufficient evidence that the mean gain with the supplement is more than 200 g greater.

(3)

www.mathswithmatt.co.uk

Question 7

Question 7

Independent surveys give $X \sim \text{Bin}(m, p)$ and $Y \sim \text{Bin}(n, p)$. The proposed unbiased estimators are $Q = k(X/m + Y/n)$ and $R = aX/m + bY/n$. Show the stated conditions and, for $m = 100$, $n = 200$, find the possible values of a for which R is better than Q.

Model solution

(a) An unbiased estimator has expectation p. Use $E(X) = mp$ and $E(Y) = np$.

$$E(Q) = k\left(\frac{E(X)}{m} + \frac{E(Y)}{n}\right) = k\left(\frac{mp}{m} + \frac{np}{n}\right) = 2kp.$$

For unbiasedness, $2kp = p$, hence $2k = 1$.

$$k = \tfrac{1}{2}.$$

(2)

(b) Apply the same unbiasedness condition to R.

$$E(R) = \frac{aE(X)}{m} + \frac{bE(Y)}{n} = \frac{amp}{m} + \frac{bnp}{n} = (a+b)p.$$

Equating this to p gives

$$a + b = 1.$$

(2)

(c) For two unbiased estimators, "better" means having the smaller variance. The survey results are independent, so variances of their weighted sum add.

$$\text{Var}(Q) = \frac{1}{4}\left(\frac{p(1-p)}{m} + \frac{p(1-p)}{n}\right).$$

$$\text{Var}(R) = p(1-p)\left(\frac{a^2}{m} + \frac{b^2}{n}\right).$$

With $m = 100$, $n = 200$, and $b = 1 - a$, the condition $\text{Var}(R) < \text{Var}(Q)$ becomes, after cancelling the positive factor $p(1-p)$,

$$\frac{a^2}{100} + \frac{(1-a)^2}{200} < \frac{1}{4}\left(\frac{1}{100} + \frac{1}{200}\right).$$

Multiplying by 800 and collecting terms,

$$8a^2 + 4(1 - 2a + a^2) < 3,$$

$$12a^2 - 8a + 1 < 0.$$

The roots are

$$a = \frac{8 \pm \sqrt{64 - 48}}{24} = \frac{8 \pm 4}{24},$$

so they are $1/6$ and $1/2$. Since the quadratic opens upwards, it is negative between its roots.

$$\frac{1}{6} < a < \frac{1}{2}.$$

(7)

www.mathswithmatt.co.uk

Question 8

Question 8

Small-bag weight is $N(500, 5^2)$ g and large-bag weight is $N(1020, 20^2)$ g. Find the requested probability for three small bags, then for one independently selected large bag and small bag.

Model solution

(a) The sum of independent normal variables is normal. Add the means and variances of the three small-bag weights.

$$T = S_1 + S_2 + S_3 \sim N\left(3(500),\, 3(5^2)\right) = N(1500, 75).$$

Therefore

$$P(1490 < T < 1530) = P\left(\frac{-10}{\sqrt{75}} < Z < \frac{30}{\sqrt{75}}\right) = 0.8756\ldots$$

$$P(1490 < T < 1530) = 0.876 \text{ (approximately).}$$

(3)

(b) The required event is $L \geqslant 2S + 30$. Form a normal variable whose being positive is exactly this event, then use its mean and variance.

$$W = L - 2S - 30.$$

Because the selected bags are independent,

$$E(W) = 1020 - 2(500) - 30 = -10,$$

$$\mathrm{Var}(W) = 20^2 + (-2)^2(5^2) = 400 + 100 = 500.$$

Thus $W \sim N(-10, 500)$, and

$$P(L \geqslant 2S + 30) = P(W \geqslant 0) = P\left(Z \geqslant \frac{0 - (-10)}{\sqrt{500}}\right) = P(Z \geqslant 0.4472\ldots) = 0.3273\ldots$$

The probability is 0.327 (approximately).

(6)

Question 2

Question 2

Two customers rank house features A to G. Customer 1 gives ranks $1, 6, 3, 7, 2, 4, 5$, and Customer 2 gives ranks $5, 7, 3, 6, 1, 2, 4$, respectively. Find Spearman's rank correlation coefficient and test whether they are generally in agreement at the 5% level.

Model solution

(a) The data are paired rankings with no ties, so Spearman's coefficient is appropriate. Match each feature's two ranks before finding the rank differences.

feature	A	B	C	D	E	F	G
d	-4	-1	0	1	1	2	1
d^2	16	1	0	1	1	4	1

$$\sum d^2 = 24.$$

$$r_s = 1 - \frac{6\sum d^2}{n(n^2-1)} = 1 - \frac{6(24)}{7(7^2-1)} = \frac{4}{7} = 0.5714\ldots$$

$$r_s = 0.571 \text{ (approximately).}$$

(4)

(b) "In agreement" means a positive population rank correlation, so use a one-tailed test.

$$H_0 : \rho_s = 0, \qquad H_1 : \rho_s > 0.$$

For $n = 7$, the 5% critical value is 0.7143. Since

$$0.5714 < 0.7143,$$

the statistic is not in the critical region.

Do not reject H_0: there is insufficient evidence at the 5% level that the customers are generally in agreement.

(3)

Question 3

Question 3

Bolt lengths are normal with mean μ mm and standard deviation 0.868 mm. A sample of 15 has mean 30.03 mm. Find a 90% confidence interval, a 10% upper-tail critical region for S^2, and assess a later sample with mean 30.06 mm and standard deviation 1.02 mm.

Model solution

(a) The population standard deviation is known and the population is normal, so a z-interval for μ is used. A central 90% interval uses $z = 1.6449$.

$$30.03 \pm 1.6449\frac{0.868}{\sqrt{15}} = 30.03 \pm 0.3687\ldots$$

$$29.661\ldots < \mu < 30.399\ldots, \text{ i.e. } (29.7, 30.4) \text{ mm.}$$

(3)

(b) To test whether a normal population variance has increased, use $(n-1)S^2/\sigma_0^2 \sim \chi^2_{n-1}$, with an upper-tail critical region.

$$\frac{14S^2}{0.868^2} > \chi^2_{14}(0.10) = 21.064.$$

Hence

$$S^2 > \frac{21.064(0.868)^2}{14} = 1.133\ldots$$

Critical region: $S^2 > 1.13$ mm^2 (approximately).

(3)

(c) Compare the new mean with the confidence interval and the new variance with the critical region.

$$29.661\ldots < 30.06 < 30.399\ldots, \qquad 1.02^2 = 1.0404 < 1.133\ldots$$

There is insufficient evidence that the machine is not working properly: the mean is within the interval and the sample variance is not in the critical region.

(2)

Question 4

Question 4

$G \sim U[-3, 15]$, $H \sim U[2, w]$, G and H are independent, and $E(H) = 10$. Also, $k \sim U[5, 10]$ defines the area bounded by $y = -3$ and $y = -4|x| + k$. Answer all parts.

Model solution

(a) A uniform probability is a favourable interval length divided by the full interval length.

$$P(G > 12) = \frac{15 - 12}{15 - (-3)} = \frac{3}{18}.$$

$$P(G > 12) = \tfrac{1}{6}.$$

(1)

(b) The mean of $U[2, w]$ is the midpoint, which determines w. Independence then allows multiplication of the two probabilities.

$$\frac{2 + w}{2} = 10 \quad \Rightarrow \quad w = 18.$$

$$P(G > 12 \text{ and } H > 12) = P(G > 12)P(H > 12) = \frac{1}{6} \times \frac{18 - 12}{18 - 2} = \frac{1}{6} \times \frac{3}{8} = \frac{1}{16}.$$

$$P(G > 12 \text{ and } H > 12) = \tfrac{1}{16}.$$

(3)

(c) For a fixed k, the two lines enclose an isosceles triangle. Find its base from the intersections, then average its area over the uniform distribution of k.

$$-3 = -4|x| + k \quad \Rightarrow \quad |x| = \frac{k + 3}{4}.$$

Thus the base is $(k + 3)/2$ and the height is $k + 3$, so

$$A = \frac{1}{2}\left(\frac{k + 3}{2}\right)(k + 3) = \frac{(k + 3)^2}{4}.$$

Since the density of $K \sim U[5, 10]$ is $1/5$,

$$E(A) = \frac{1}{5}\int_5^{10} \frac{(k + 3)^2}{4}\, dk = \frac{1}{60}\left[(k + 3)^3\right]_5^{10} = \frac{2197 - 512}{60} = \frac{337}{12}.$$

$$E(A) = \tfrac{337}{12} \text{ square units.}$$

(5)

Question 5

Question 5

X has density $f(x) = ax^{-2} - bx^{-3}$ for $x \geqslant 2$, and zero otherwise. Given $P(X \leqslant 4) = 3/8$, show $a = 3$ and find the exact median.

Model solution

(a) A density must have total area 1. Use that condition and the given probability to form two equations in a and b.

$$\int \left(ax^{-2} - bx^{-3}\right) dx = -\frac{a}{x} + \frac{b}{2x^2}.$$

First,

$$1 = \int_2^\infty f(x)\, dx = \left[-\frac{a}{x} + \frac{b}{2x^2}\right]_2^\infty = \frac{a}{2} - \frac{b}{8}.$$

Also,

$$\frac{3}{8} = P(X \leqslant 4) = \int_2^4 f(x)\, dx = \left[-\frac{a}{x} + \frac{b}{2x^2}\right]_2^4 = \frac{a}{4} - \frac{3b}{32}.$$

Equivalently,

$$4a - b = 8, \qquad 8a - 3b = 12.$$

Subtracting the first equation multiplied by 3 from the second gives

$$-4a = -12, \qquad a = 3.$$

$$a = 3.$$

(6)

(b) The total-area equation now gives $b = 4$. The median m satisfies $P(X \leqslant m) = 1/2$, with $m \geqslant 2$.

$$4(3) - b = 8 \quad \Rightarrow \quad b = 4,$$

$$\frac{1}{2} = \int_2^m \left(\frac{3}{x^2} - \frac{4}{x^3}\right) dx = \left[-\frac{3}{x} + \frac{2}{x^2}\right]_2^m = 1 - \frac{3}{m} + \frac{2}{m^2}.$$

Multiplying by $2m^2$ gives

$$m^2 - 6m + 4 = 0,$$

so

$$m = 3 \pm \sqrt{5}.$$

Only $3 + \sqrt{5}$ lies in the support $x \geqslant 2$.

$$\textit{Median} = 3 + \sqrt{5}.$$

(4)