

A LEVEL FURTHER MATHS • EXTREME PAPERS

MATHS



MODEL ANSWERS

EXTREME PAPER: FURTHER STATISTICS

SCORE

/90

PERCENTAGE

%

STEMUP

Paper: Further Statistics (90 marks)

1 (11 marks) Edexcel | CIE

$$(a) G_X(t) = \sum_{r=1}^{\infty} q(1-q)^{r-1}t^r = qt \sum_{r=1}^{\infty} [(1-q)t]^{r-1} \quad \text{M1}$$

This is a geometric series with first term 1 and common ratio $(1-q)t$, so the sum is $\frac{1}{1-(1-q)t}$.

$$\text{Hence } G_X(t) = \frac{qt}{1-(1-q)t}. \quad \text{A1}$$

Answer: Shown

2 marks

(b) Key idea: condition on the value of N.

$$G_S(t) = E(t^S) = \sum_{n=1}^{\infty} P(N=n) E(t^{X_1+\dots+X_n}) \quad \text{M1}$$

$$\text{By independence, } E(t^{X_1+\dots+X_n}) = [G_X(t)]^n. \quad \text{M1}$$

$$\text{So } G_S(t) = \sum_{n=1}^{\infty} P(N=n)[G_X(t)]^n = G_N(G_X(t)). \quad \text{A1}$$

Answer: Shown

3 marks

$$(c) G_N(s) = \frac{ps}{1-(1-p)s}, \text{ so } G_S(t) = \frac{p G_X}{1-(1-p)G_X} \text{ with } G_X = \frac{qt}{1-(1-q)t}.$$

$$\text{Multiply numerator and denominator by } 1-(1-q)t: G_S(t) = \frac{pqt}{[1-(1-q)t] - (1-p)qt} \quad \text{M1}$$

$$\text{Denominator} = 1-(1-q)t - (1-p)qt = 1-(1-pq)t. \text{ So } G_S(t) = \frac{pqt}{1-(1-pq)t}, \text{ the PGF of} \quad \text{A1}$$

a geometric distribution with parameter pq.

Answer: Geometric with parameter pq

2 marks

(d) From (c), $P(S=3) = pq(1-pq)^2$.

B1

$P(N=2) = p(1-p)$. Given $N=2$, $S=3$ requires $(X_1, X_2) = (1, 2)$ or $(2, 1)$.

M1

$$P(S=3 | N=2) = 2 \times q \times q(1-q) = 2q^2(1-q).$$

A1

$$P(N=2 | S=3) = \frac{p(1-p) \times 2q^2(1-q)}{pq(1-pq)^2} = \frac{2(1-p)q(1-q)}{(1-pq)^2}.$$

M1

Answer: Shown

4 marks

2 (11 marks) AQA | OCR A | OCR MEI | CIE

$$(a) \text{ Each of } r \text{ and } s \text{ is a permutation of } 1, 2, \dots, n, \text{ so } \sum r = \sum s = \frac{n(n+1)}{2} \text{ and } \sum r^2 = \sum s^2 = \frac{n(n+1)(2n+1)}{6}. \quad \text{B1}$$

$$S_{rr} = S_{ss} = \frac{n(n+1)(2n+1)}{6} - \frac{n(n+1)^2}{4} = \frac{n(n^2-1)}{12}. \quad \text{M1}$$

$$\sum d^2 = \sum r^2 + \sum s^2 - 2 \sum rs, \text{ so } \sum rs = \frac{n(n+1)(2n+1)}{6} - \frac{1}{2} \sum d^2.$$

$$S_{rs} = \sum rs - \frac{(\sum r)(\sum s)}{n} = S_{rr} - \frac{1}{2} \sum d^2. \quad \text{M1}$$

$$\rho = \frac{S_{rs}}{\sqrt{S_{rr}S_{ss}}} = \frac{S_{rs}}{S_{rr}} = 1 - \frac{6 \sum d^2}{n(n^2-1)}. \quad \text{A1}$$

Answer: Shown

4 marks

(b) (i) d_i^2 and d_i have the same parity, so $\sum d^2 \equiv \sum d = \sum r - \sum s = 0 \pmod{2}$.

M1 A1

(ii) For $n=6$, $n(n^2-1) = 210$, so $\rho = 1 - \frac{\sum d^2}{35}$. $\rho = 0$ would need $\sum d^2 = 35$, which is odd.

M1

As $\sum d^2$ is even, the closest values to 35 are 34 and 36, giving $|\rho| = \frac{1}{35}$.

Attained by judge A ranking 1, 2, 3, 4, 5, 6 and judge B ranking the same performers 1, 4, 6, 5, 3, 2: $d = 0, -2, -3, -1, 2, 4$ and $\sum d^2 = 34$.

A1

Answer: Smallest $|\rho| = 1/35$, e.g. B ranks 1, 4, 6, 5, 3, 2

4 marks

(c) The performer ranked i by A is ranked $m + 1 - i$ by B, so $d_i = 2i - m - 1$ for $i \leq m$ and $\sum$ M1

$$d^2 = \sum_{i=1}^m (m + 1 - 2i)^2 = \frac{m(m^2 - 1)}{3}.$$

$$n(n^2 - 1) = 504, \text{ so } \rho = 1 - \frac{6m(m^2 - 1)}{3 \times 504} = 1 - \frac{m(m^2 - 1)}{252}. \quad \text{A1}$$

$m = 4: \rho = 1 - \frac{60}{252} = 0.762 > 0.738$. $m = 5: \rho = 1 - \frac{120}{252} = 0.524 < 0.738$. As ρ decreases with m , the largest m is 4. A1

Answer: $m = 4$ 3 marks

3 (10 marks) Edexcel | OCR A | OCR MEI | CIE

(a) Key idea: every cell has the same size of deviation $O - E$.

$$\text{Cell a: } a - \frac{(a+b)(a+c)}{n} = \frac{na - (a+b)(a+c)}{n} = \frac{ad - bc}{n}. \quad \text{M1}$$

Since row and column totals of the $O - E$ table are zero, the deviations in the other cells are A1

$$-\frac{ad - bc}{n}, -\frac{ad - bc}{n} \text{ and } +\frac{ad - bc}{n}; \text{ all have square } \frac{(ad - bc)^2}{n^2}.$$

$$\text{So } \chi^2 = \frac{(ad - bc)^2}{n^2} \sum \frac{1}{E}, \text{ with } E = \frac{r_i c_j}{n} \text{ for row totals } r_i \text{ and column totals } c_j.$$

$$\sum \frac{1}{E} = n(1/r_1 + 1/r_2)(1/c_1 + 1/c_2) = n \times \frac{n}{r_1 r_2} \times \frac{n}{c_1 c_2} = \frac{n^3}{r_1 r_2 c_1 c_2}. \quad \text{M1}$$

$$\text{Hence } \chi^2 = \frac{n(ad - bc)^2}{r_1 r_2 c_1 c_2}, \text{ as required.} \quad \text{A1}$$

Answer: Shown 4 marks

(b) Clinic 1: $(a, b, c, d) = (53, 97, 1, 19)$, $n = 170$: $\chi^2 = 7.49 > 3.841$, so reject independence; cure rates $53/150 = 35.3\%$ for A against $1/20 = 5\%$ for B. M1 A1

Clinic 2: $(36, 4, 85, 35)$, $n = 160$: $\chi^2 = 5.98 > 3.841$, so reject independence; cure rates 90% for A against 70.8% for B. A1

Combined: $(89, 101, 86, 54)$, $n = 330$: $\chi^2 = 6.89 > 3.841$, so reject independence, but now A has cure rate $89/190 = 46.8\%$ against $86/140 = 61.4\%$ for B. A1

All expected frequencies are at least 5 (the smallest are 6.35, 9.75 and 65.76 respectively), so the χ^2 approximation is valid.

Answer: $\chi^2 = 7.49, 5.98, 6.89$; A better in each clinic, B better when pooled 4 marks

(c) Treatment A: 150 of 190 patients (78.9%) were at Clinic 1; treatment B: 20 of 140 (14.3%). M1

Clinic 1 has much lower cure rates for both treatments, and treatment A was used mostly there, so pooling mixes the clinic effect (a confounding variable) with the treatment effect and reverses the apparent comparison (Simpson's paradox). A1

Answer: 78.9% against 14.3%; Clinic 1 confounds the pooled comparison 2 marks

4 (12 marks) Edexcel | OCR A | OCR MEI | CIE

(a) Let $N = B + K \sim \text{Po}(2t)$. $P(B = b, K = k) = P(N = b + k) \times P(B = b | N = b + k)$. M1

$$= e^{-2t} \times (2t)^{b+k} / (b+k)! \times \frac{(b+k)!}{b!k!} \times \left(\frac{1}{4}\right)^b \left(\frac{3}{4}\right)^k. \quad \text{M1}$$

$$(b+k)! \text{ cancels, } e^{-2t} = e^{-t/2} e^{-3t/2}, \text{ and } (2t)^{b+k} \left(\frac{1}{4}\right)^b \left(\frac{3}{4}\right)^k = \left(\frac{t}{2}\right)^b \left(\frac{3t}{2}\right)^k. \quad \text{M1}$$

So $P(B = b, K = k) = [e^{-t/2} (t/2)^b / b!] \times [e^{-3t/2} (3t/2)^k / k!]$, a product of a function of b and a function of k , so $B \sim \text{Po}(\frac{t}{2})$ and $K \sim \text{Po}(\frac{3t}{2})$ are independent. A1

Answer: Independent; $B \sim \text{Po}(t/2)$, $K \sim \text{Po}(3t/2)$ 4 marks

(b) Customer types are independent of the arrival times. At least two browsers before the first buyer means the first two customers are both browsers. M1

$$\text{Probability} = \left(\frac{3}{4}\right)^2 = \frac{9}{16}. \quad \text{A1}$$

Answer: 9/16 2 marks

(c) Each customer is independently a buyer with probability $\frac{1}{4}$, so N is the sum of three B1

independent geometric random variables, each with mean 4.

$$E(N) = 3 \times 4 = 12.$$

M1

The 12 expected customers include the 3 buyers, so the expected number of browsers is $12 - 3 = 9$.

A1

Answer: $E(N) = 12$; expected number of browsers = 9

3 marks

(d) The event needs: 0 buyers in (0, 2], 2 buyers in (2, 4], and 3 browsers in (0, 4]. These are independent, with means 1, 1 and 6.

M1

$$\text{Probability of the event} = e^{-1} \times \frac{e^{-1}}{2} \times \frac{e^{-6} 6^3}{3!} = 18e^{-8}.$$

M1

$$\text{Divide by } P(N(4) = 5) = e^{-8} \frac{8^5}{120}; \text{ probability} = \frac{18 \times 120}{32768} = \frac{135}{2048}.$$

A1

Answer: 135/2048

3 marks

5 (11 marks) Edexcel | OCR A | OCR MEI

(a) $M \leq x$ iff all n observations are $\leq x$, so $P(M \leq x) = \left(\frac{x}{\theta}\right)^n$ for $0 \leq x \leq \theta$, and $f(x) = \frac{n x^{n-1}}{\theta^n}$.

B1

$$E(M) = \int_0^\theta \frac{n x^n}{\theta^n} dx = \frac{n\theta}{(n+1)}.$$

M1

$$E(M^2) = \int_0^\theta \frac{n x^{n+1}}{\theta^n} dx = \frac{n\theta^2}{(n+2)}.$$

M1

$$\text{Var}(M) = \frac{n\theta^2}{n+2} - \frac{n^2\theta^2}{(n+1)^2} = \frac{n\theta^2}{(n+1)^2(n+2)}.$$

A1

Answer: Shown

4 marks

(b) $E(T) = \frac{n+1}{n} \times \frac{n\theta}{(n+1)} = \theta$, so T is unbiased.

B1

$$\text{MSE}(T) = \text{Var}(T) = \left(\frac{n+1}{n}\right)^2 \times \frac{n\theta^2}{(n+1)^2(n+2)} = \frac{\theta^2}{n(n+2)}.$$

M1

$$\text{MSE}(T_c) = c^2 \text{Var}(M) + [cE(M) - \theta]^2 = \theta^2 \left[\frac{nc^2}{(n+1)^2(n+2)} + \left(\frac{nc}{n+1} - 1\right)^2 \right].$$

M1

The coefficient of c^2 is $\frac{n}{(n+1)^2(n+2)} + \frac{n^2}{(n+1)^2} = \frac{n[1+n(n+2)]}{(n+1)^2(n+2)} = \frac{n}{n+2}$, since $1+n(n+2) = (n+1)^2$; the other terms give $-\frac{2nc}{n+1} + 1$.

A1

Differentiate: $\frac{2nc}{n+2} - \frac{2n}{n+1} = 0$, so $c = \frac{n+2}{n+1}$ (the coefficient of c^2 is positive, so this is a minimum).

A1

Answer: $c = (n+2)/(n+1)$

5 marks

(c) Substituting, $\text{MSE} = \theta^2 \left[\frac{n(n+2)}{(n+1)^2} - \frac{2n(n+2)}{(n+1)^2} + 1 \right] = \theta^2 \frac{1-n(n+2)}{(n+1)^2} = \frac{\theta^2}{(n+1)^2}.$

M1

$(n+1)^2 = n(n+2) + 1 > n(n+2)$, so $\frac{\theta^2}{(n+1)^2} < \frac{\theta^2}{n(n+2)}$. For $n = 10$ the ratio is $\frac{121}{120}$.

A1

Answer: Min MSE = $\theta^2/(n+1)^2$; ratio 121/120

2 marks

6 (11 marks) Edexcel | OCR A

(a) Reject H_0 if $X \geq k$. Under H_0 , $P(X \geq k) = 0.8^{k-1}$.

M1

Need $0.8^{k-1} \leq 0.05$, so $k-1 \geq \frac{\ln 0.05}{\ln 0.8} = 13.4$, hence $k = 15$. ($k = 14$ gives 0.0550, too large.)

M1 A1

Critical region $X \geq 15$; size = $0.8^{14} = 0.0440$.

A1

Power when $p = 0.1$: $P(X \geq 15) = 0.9^{14} = 0.2288$.

A1

Answer: $X \geq 15$; size 0.0440; power 0.2288

5 marks

(b) Key idea: S is the number of trials needed to obtain two successes in one long sequence of independent trials.

$S > m$ iff there are fewer than 2 successes in the first m trials.

M1

$P(0 \text{ successes}) = q^m$, $P(1 \text{ success}) = m p q^{m-1}$ (binomial).

M1

So $P(S > m) = q^m + m p q^{m-1}$; taking $m = s-1$ gives $P(S \geq s) = q^{s-1} + (s-1)p q^{s-2}$.

A1

Answer: Shown

3 marks

(c) Under H_0 , $P(S \geq s) = 0.8^{s-1} + 0.2(s-1)0.8^{s-2}$. $s = 22$: $0.0576 > 0.05$; $s = 23$: $0.0480 \leq 0.05$.

M1

Critical region $S \geq 23$, size 0.0480.

A1

Power at $p = 0.1$: $0.9^{22} + 22(0.1)0.9^{21} = 0.3392$, compared with 0.2288 for the single observation: the S test is more powerful (about 1.48 times as powerful), for a size that is only slightly larger.

A1

Answer: $S \geq 23$; size 0.0480; power 0.3392 (S test more powerful)

3 marks

7 (12 marks) Edexcel | OCR A | OCR MEI | CIE

(a) $P(M \leq x) = x^3$, so $f_M(x) = 3x^2$ for $0 \leq x \leq 1$.

M1

$$E(M) = \int_0^1 3x^3 dx = \frac{3}{4}.$$

A1

$$E(M^2) = \int_0^1 3x^4 dx = \frac{3}{5}, \text{ so } \text{Var}(M) = \frac{3}{5} - \frac{9}{16} = \frac{3}{80}.$$

A1

Answer: Shown

3 marks

(b) Key idea: split according to which variable is the smallest.

Ties have probability zero and, by symmetry, each X_i is the smallest with equal contribution, so

M1

$P(R \leq r) = 3 \times P(X_1 \text{ smallest and } R \leq r)$.

If $X_1 = a$, then X_1 is the smallest with $R \leq r$ iff both X_2 and X_3 lie in $[a, a+r] \cap [a, 1]$, which

M1

has length $\min(1-a, r)$. Hence $P(R \leq r) = 3 \int_0^1 [\min(1-a, r)]^2 da$.

Split at $a = 1-r$: for $0 \leq a \leq 1-r$ the length is r , and for $1-r \leq a \leq 1$ it is $1-a$.

M1

$$P(R \leq r) = 3[r^2(1-r) + \int_{1-r}^1 (1-a)^2 da] = 3[r^2(1-r) + \frac{r^3}{3}].$$

A1

$$= 3r^2 - 3r^3 + r^3 = 3r^2 - 2r^3.$$

A1

Answer: Shown

5 marks

$$(c) f_R(r) = 6r - 6r^2 = 6r(1-r). E(R) = \int_0^1 6r^2(1-r) dr = 6(\frac{1}{3} - \frac{1}{4}) = \frac{1}{2}.$$

M1

$$E(R^2) = 6(\frac{1}{4} - \frac{1}{5}) = \frac{3}{10}, \text{ so } \text{Var}(R) = \frac{3}{10} - \frac{1}{4} = \frac{1}{20}.$$

A1

Answer: $E(R) = 1/2$, $\text{Var}(R) = 1/20$

2 marks

(d) Key idea: m has the same distribution as $1-M$, so $\text{Var}(m) = \text{Var}(M) = \frac{3}{80}$.

$$\text{Var}(R) = \text{Var}(M) + \text{Var}(m) - 2\text{Cov}(m, M), \text{ so } \frac{1}{20} = \frac{6}{80} - 2\text{Cov}(m, M), \text{ giving } \text{Cov}(m, M) = \frac{1}{80}.$$

M1

$$\text{Correlation} = \frac{\text{Cov}(m, M)}{\sqrt{\text{Var}(m)\text{Var}(M)}} = \frac{1}{80} \div \frac{3}{80} = \frac{1}{3}.$$

A1

Answer: $1/3$

2 marks

8 (12 marks) OCR MEI | CIE

$$(a) M_X(t) = \int_{-\infty}^{\infty} \frac{1}{2} e^{tx - |x|} dx. \text{ Split at } x = 0.$$

M1

$$\int_{-\infty}^0 \frac{1}{2} e^{(t+1)x} dx = \frac{1}{2(1+t)} \text{ (valid for } t > -1); \int_0^{\infty} \frac{1}{2} e^{(t-1)x} dx = \frac{1}{2(1-t)} \text{ (valid for } t < 1).$$

M1

$$\text{Sum} = \frac{1}{2} \left[\frac{1}{1+t} + \frac{1}{1-t} \right] = \frac{1}{1-t^2}.$$

A1

Answer: Shown

3 marks

$$(b) M_{-E_2}(t) = E(e^{-tE_2}) = \frac{1}{1+t} \text{ for } t > -1.$$

M1

By independence $M_{E_1-E_2}(t) = \frac{1}{(1-t)(1+t)} = \frac{1}{1-t^2} = M_X(t)$; equal moment generating functions mean equal distributions.

A1

Answer: Shown

2 marks

$$(c) M_X(t) = 1 + t^2 + t^4 + \dots, \text{ and the coefficient of } t^k \text{ is } \frac{E(X^k)}{k!}.$$

M1

$$E(X) = 0 \text{ (no } t \text{ term)}, E(X^2) = 2! \times 1 = 2, \text{ so } \text{Var}(X) = 2; E(X^4) = 4! \times 1 = 24.$$

A1

Answer: $\text{Var}(X) = 2$, $E(X^4) = 24$

2 marks

(d) Key idea: given T , W is normal with variance $2T$.

$$E(e^{tW} | T) = E(e^{t\sqrt{2T}Z} | T) = e^{(t\sqrt{2T})^2/2} = e^{t^2T}.$$

M1

$$E(e^{tW}) = E(e^{t^2T}) = \int_0^\infty e^{t^2x} e^{-x} dx$$

M1

$$= \frac{1}{1-t^2} \text{ for } |t| < 1, \text{ which is } M_X(t), \text{ so } W \text{ and } X \text{ have the same distribution.}$$

A1

Answer: Shown

3 marks

$$(e) M_{Y_n}(t) = [M_X(t/\sqrt{n})]^n = (1 - \frac{t^2}{n})^{-n}.$$

M1

As $n \rightarrow \infty$, $(1 - \frac{t^2}{n})^n \rightarrow e^{-t^2}$, so $M_{Y_n}(t) \rightarrow e^{t^2}$, the moment generating function of $N(0, 2)$, consistent with $\text{Var}(X) = 2$.

A1

Answer: $(1 - t^2/n)^{-n} \rightarrow e^{t^2}; N(0, 2)$

2 marks