

A LEVEL FURTHER MATHS • EXTREME PAPERS

MATHS



EXAM PAPER

EXTREME PAPER: FURTHER STATISTICS

SCORE

/90

PERCENTAGE

%

STEMUP

Paper: Further Statistics (90 marks)

1 Edexcel | CIE

A discrete random variable X has a geometric distribution with parameter q , so that $P(X = r) = q(1 - q)^{r-1}$ for $r = 1, 2, 3, \dots$

The random variable N has a geometric distribution with parameter p . Independent random variables $X_1, X_2, X_3, \dots$ each have the same distribution as X , and are also independent of N .

Define $S = X_1 + X_2 + \dots + X_N$, so that S is a random sum of N terms.

- Show that the probability generating function of X is $G_X(t) = \frac{qt}{1 - (1 - q)t}$. [2]
- Show that $G_S(t) = G_N(G_X(t))$, where G_N is the probability generating function of N . [3]
- Hence show that S has a geometric distribution, and state its parameter. [2]
- Given that $S = 3$, show that the probability that $N = 2$ is $\frac{2(1 - p)q(1 - q)}{(1 - pq)^2}$. [4]

2 AQA | OCR A | OCR MEI | CIE

Two judges rank the same n performers with no tied ranks. The ranks given to the i th performer are r_i and s_i , and $d_i = r_i - s_i$. Spearman's rank correlation coefficient is the product moment correlation coefficient calculated from the ranks (r_i, s_i) .

- Starting from the product moment correlation coefficient of the ranks, show that

$$\text{Spearman's coefficient is } \rho = 1 - \frac{6 \sum d_i^2}{n(n^2 - 1)}.$$

You may use $\sum i = \frac{n(n+1)}{2}$ and $\sum i^2 = \frac{n(n+1)(2n+1)}{6}$. [4]

- Prove that $\sum d_i^2$ is always even.
 - Hence show that ρ cannot be 0 when $n = 6$. Find the smallest possible value of $|\rho|$ when $n = 6$, and give a pair of rankings of six performers for which it is achieved. [4]
- For $n = 8$, judge B's ranking is identical to judge A's except that the m performers ranked 1, 2, ..., m by judge A ($2 \leq m \leq 8$) are ranked in exactly the reverse order by judge B. The two-tailed 5% critical value of ρ for $n = 8$ is 0.738. Find the largest m for which the two judges' rankings show significant positive correlation at the 5% level. [3]

3 Edexcel | OCR A | OCR MEI | CIE

A 2×2 contingency table has observed frequencies a, b in the first row and c, d in the second row, with total $n = a + b + c + d$. For a test of independence, the expected frequency in each cell is (row total) $\times$ (column total) $\div n$. Do not use a continuity correction anywhere in this question.

- Show that the χ^2 test statistic for the table is $\frac{n(ad - bc)^2}{(a + b)(c + d)(a + c)(b + d)}$. [4]
- Two clinics trial treatments A and B. The numbers of patients cured and not cured are as follows.
Clinic 1: Treatment A 53 cured, 97 not cured; Treatment B 1 cured, 19 not cured.
Clinic 2: Treatment A 36 cured, 4 not cured; Treatment B 85 cured, 35 not cured.
 Test at the 5% level (critical value 3.841 for 1 degree of freedom) whether cure and treatment are independent (i) in Clinic 1, (ii) in Clinic 2, and (iii) for the two clinics combined. In each case state which treatment has the higher observed cure rate, and verify that the model is valid (every expected frequency at least 5). [4]
- The pooled result contradicts both clinics. Calculate the proportion of each treatment's patients who were treated at Clinic 1, and use these to explain the contradiction. [2]

4 Edexcel | OCR A | OCR MEI | CIE

Customers arrive at a shop according to a Poisson process of rate 2 per minute. Independently of everything else, each customer is a buyer with probability $\frac{1}{4}$ and a browser with probability $\frac{3}{4}$. Let $B(t)$ and $K(t)$ be the numbers of buyers and browsers who arrive in a time interval of length t minutes.

- By conditioning on the total number of arrivals, show that $B(t)$ and $K(t)$ are independent Poisson random variables, and state their means. [4]
- The shop opens at time 0. Find the probability that at least two browsers arrive before the first buyer. [2]
- Let N be the number of customers who arrive up to and including the third buyer. Find $E(N)$, and find the expected number of browsers who arrive before the third buyer. [3]
- Given that exactly 5 customers arrive in the first 4 minutes, find the probability that exactly two of them are buyers and both of these arrive in the last 2 of those 4 minutes. [3]

5 Edexcel | OCR A | OCR MEI

A random sample $X_1, X_2, \dots, X_n$ ($n \geq 2$) is taken from the continuous uniform distribution on $[0, \theta]$, where $\theta > 0$ is unknown. Let M be the largest of the n observations.

For an estimator T of θ , its mean square error is $\text{MSE}(T) = E[(T - \theta)^2]$, which equals $\text{Var}(T) + [E(T) - \theta]^2$.

- Find the cumulative distribution function and probability density function of M . Hence show that $E(M) = \frac{n\theta}{n+1}$ and $\text{Var}(M) = \frac{n\theta^2}{(n+1)^2(n+2)}$. [4]
- Show that $T = \frac{n+1}{n}M$ is an unbiased estimator of θ with $\text{MSE}(T) = \frac{\theta^2}{n(n+2)}$.
For a constant c , let $T_c = cM$. Show that $\text{MSE}(T_c) = \theta^2 \left[\frac{nc^2}{n+2} - \frac{2nc}{n+1} + 1 \right]$, and find the value of c that minimises it. [5]
- Find the minimum value of $\text{MSE}(T_c)$ and hence show that, for every n , the biased estimator with this value of c has a smaller mean square error than the unbiased estimator T . Find the ratio $\text{MSE}(T) / \text{MSE}(T_c)$ when $n = 10$. [2]

6 Edexcel | OCR A

A scientist counts the number of trials X up to and including the first success in a sequence of independent trials, each with probability of success p . She tests $H_0: p = 0.2$ against $H_1: p < 0.2$ at a significance level of at most 5%, rejecting H_0 when the number of trials needed is large.

- She first uses a single observation of X . Find the critical region, the actual size of the test, and the power of the test when $p = 0.1$. [5]
- She now uses $S = X_1 + X_2$, the sum of two independent observations of X , each with success probability p , where $q = 1 - p$. By considering a single sequence of Bernoulli trials, show that $P(S > m) = q^m + m p q^{m-1}$ for $m \geq 2$ and hence that $P(S \geq s) = q^{s-1} + (s-1)p q^{s-2}$. [3]
- Find the critical region of the test based on S at a significance level of at most 5%, its actual size, and its power when $p = 0.1$. Hence state which of the two tests is more powerful and by how much. [3]

7 Edexcel | OCR A | OCR MEI | CIE

X_1, X_2 and X_3 are independent random variables, each uniformly distributed on $[0, 1]$. Let m and M be the smallest and largest of the three, and let $R = M - m$.

- Find the probability density function of M , and show that $E(M) = \frac{3}{4}$ and $\text{Var}(M) = \frac{3}{80}$. [3]
- Show that $P(R \leq r) = 3r^2 - 2r^3$ for $0 \leq r \leq 1$. [5]
- Hence find $E(R)$ and $\text{Var}(R)$. [2]
- Hence find the correlation coefficient between m and M , using $\text{Var}(m - M) = \text{Var}(m) + \text{Var}(M) - 2\text{Cov}(m, M)$ and correlation = $\frac{\text{Cov}(m, M)}{\sqrt{\text{Var}(m)\text{Var}(M)}}$. [2]

8 OCR MEI | CIE

The random variable X has probability density function $f(x) = \frac{1}{2}e^{-|x|}$ for $-\infty < x < \infty$. The moment generating function of a random variable Y is $M_Y(t) = E(e^{tY})$; you may assume that it determines the distribution of Y uniquely, and that $M_Y(t) = \sum E(Y^k) \frac{t^k}{k!}$.

(a) Show that the moment generating function of X is $M_X(t) = \frac{1}{1-t^2}$ for $|t| < 1$. [3]

(b) E_1 and E_2 are independent random variables, each with probability density function e^{-x} for $x \geq 0$. Given that E_1 has moment generating function $\frac{1}{1-t}$ for $t < 1$, show that X has the same distribution as $E_1 - E_2$. [2]

(c) Use the series expansion of $M_X(t)$ to find $\text{Var}(X)$ and $E(X^4)$. [2]

(d) Z is a standard normal random variable with moment generating function $E(e^{sZ}) = e^{s^2/2}$, and T is an exponential random variable with probability density function e^{-x} ($x \geq 0$), independent of Z . Define $W = \sqrt{2T} Z$.
By first finding $E(e^{tW} | T)$, show that W has the same distribution as X . [3]

(e) $X_1, X_2, \dots, X_n$ are independent random variables, each with the same distribution as X .
Find the moment generating function of $Y_n = \frac{X_1 + X_2 + \dots + X_n}{\sqrt{n}}$ and show that it tends to e^{t^2} as $n \rightarrow \infty$. Identify the limiting distribution, given that $N(0, \sigma^2)$ has moment generating function $e^{\sigma^2 t^2/2}$. [2]