

BSc, MSci and MSc EXAMINATIONS (MATHEMATICS)
Summer 2026

This paper is also taken for the relevant examination for the
Associateship of the Royal College of Science

Probability and Statistics

Date: Friday, 01 May, 2026

Time: Start time 14:00 – End time 16:00 (BST)

Time Allowed: 2 hours

This paper has 4 Questions.

Please Answer Each Question in a Separate Answer Booklet

This is a closed book examination.

Candidates should start their solutions to each question on a new sheet of paper.

Supplementary books may only be used after the relevant main book(s) are full.

Any required additional material(s) will be provided.

Credit will be given for all questions attempted.

Allow margins for marking.

DO NOT OPEN THIS PAPER UNTIL THE INVIGILATOR TELLS YOU TO DO SO

Please remember to justify all your answers and state carefully which results from the lectures you apply in your proofs.

Statistical tables are provided on the last pages of the exam paper.

1. (a) Briefly describe the probability space $(\Omega, \mathcal{F}, P)$ for rolling a fair six-sided die twice. (3 marks)

- (b) A player decides to play *EuroMillions*, a popular lottery played in the United Kingdom and across Europe. A ticket consists of a selection of:

- * 5 distinct main numbers from the set $\{1, \dots, 50\}$, and
- * 2 distinct Lucky Stars from the set $\{1, \dots, 12\}$,

where the order of the numbers does not matter.

For the purposes of this question, a secondary prize is awarded for matching 3 main numbers and both Lucky Stars.

- (i) How many EuroMillions outcomes are possible? (3 marks)
- (ii) A ticket matches exactly 3 main numbers. What is the probability it also matches both Lucky Stars, i.e. wins the secondary prize? (3 marks)
- (iii) What is the probability that a randomly chosen ticket wins the secondary prize over all possible tickets? (4 marks)
- (iv) Suppose now that the two Lucky Stars are drawn *with replacement*, so that the same Lucky Star may appear twice. How many different Lucky Star outcomes are possible? (2 marks)

[For parts (i)-(iii), you may leave your answers in binomial forms.]

- (c) Let X follow a Shifted Geometric distribution with parameters n (shift parameter) and p , and its probability mass function is given by:

$$p_X(x) = \begin{cases} p(1-p)^{x-n}, & x = n, n+1, n+2, \dots \\ 0, & \text{otherwise.} \end{cases}$$

Show that the probability mass function is valid, and find the cumulative distribution function. (5 marks)

(Total: 20 marks)

2. (a) Construct three random variables that are *pairwise* independent but not independent. Justify your construction. (3 marks)

- (b) Consider jointly continuous random variables X and Y with joint distribution function given by:

$$F_{X,Y}(x, y) = \begin{cases} 1 - e^{-x} - e^{-y} + e^{-x-y}, & x \geq 0, y \geq 0, \\ 0, & \text{otherwise.} \end{cases}$$

- (i) Compute the marginal distribution functions of X , and Y , $F_X(x)$ and $F_Y(y)$, and their densities. Identify the distributions and their parameter(s). (5 marks)
 - (ii) Are X and Y independent? (2 marks)
 - (iii) Find the moment generating function of X . (3 marks)
 - (iii) Find the cumulative distribution function of $Z = X^2$. (3 marks)
- (c) Suppose there are n students in the Clore and d possible birthdays in a year (assume each day is equally likely and ignore leap years), and each student's birthday is independent of the others.
- Let Y_i be an indicator random variable which equals 1 if student i has a birthday on 1 January, and 0 otherwise, and let B denote the total number of students with a birthday on 1 January.
- (i) Determine a suitable distribution for the random variable Y_i and the random variable B , and state clearly their parameters. (2 marks)
 - (ii) Find the probability that at least one student has a birthday on 1 January. (2 marks)

(Total: 20 marks)

3. (a) Let X be a random variable with finite first and second moments that follows the distribution F_X . Suppose a , b and c are values such that

$$a = E[(X - b)^2] = \min_{d \in \mathbb{R}} E[(X - d)^2], \quad \text{and} \quad E[|X - c|] = \min_{e \in \mathbb{R}} E[|X - e|],$$

where d , e are arbitrary real values. What are a , b and c in relation to X or its distribution F_X ? (3 marks)

- (b) Prove that the mean squared error of an estimator $\hat{\Theta}$ of a parameter θ can be expressed in terms of its bias $b_\theta(\hat{\Theta})$ and its variance $\text{Var}(\hat{\Theta})$ as

$$E[(\hat{\Theta} - \theta)^2] = [b_\theta(\hat{\Theta})]^2 + \text{Var}(\hat{\Theta}).$$

(3 marks)

- (c) Suppose we are given a dataset that lists the prices of 78 houses and also gives the number of bedrooms each house has and the area in which it is located. In this dataset, all of the houses have between 2 and 7 bedrooms, and the areas are either South Kensington, Chelsea, Fulham or Clapham. The first three rows of the dataset are:

House ID	Number of bedrooms	Area	Price in £
856018	4	South Kensington	4,750,000
134338	3	Clapham	1,400,000
483579	5	Chelsea	7,200,000
$\vdots$	$\vdots$	$\vdots$	$\vdots$

- (i) Which type of plot would be best to visualise the data in a way that shows the relationship between the number of bedrooms and the price of the houses? (1 mark)
- (ii) Which type of plot would be best for displaying the median and spread of the house prices in South Kensington and for indicating if any prices are outliers? (1 mark)
- (d) The heights of a group of students are assumed to be independent and follow a normal distribution with mean μ and variance σ^2 , where μ and σ^2 are both unknown. Twelve students are randomly selected from the group and their heights are measured as $x_1, \dots, x_{12}$, and the sample mean and variance of these measurements are computed as $\bar{x} = 178.5$ cm and $s^2 = 3$ cm², respectively. For the questions below, Tables 1-4 on pages 8-11 may be helpful.
- (i) Compute a 99% confidence interval for the unknown variance, σ^2 , with full justification. (4 marks)
- (ii) Test the hypothesis that $\mu = 180$, stating the null and alternative hypotheses and computing a p -value for the appropriate test statistic to two decimal places. (5 marks)
- (e) Define what a p -value is and how to interpret it, referencing any particular distribution it may follow and under what conditions this may occur. (3 marks)

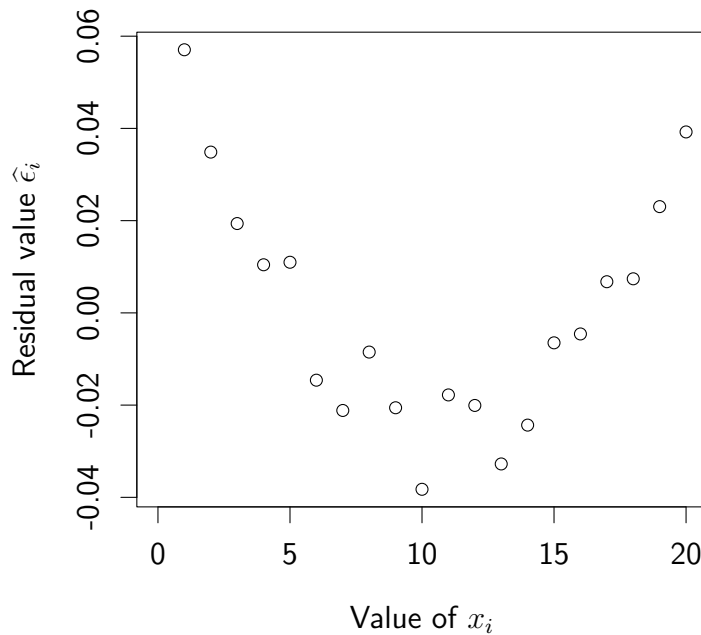
(Total: 20 marks)

4. (a) Suppose that X and Y are two normally distributed random variables that are neither independent nor identically distributed. In fact, suppose it is known that $X \sim N(8, 4)$ and $Y \sim N(7, 9)$, and their correlation is $\text{Cor}(X, Y) = \frac{1}{3}$. Defining the new random variable $Z = 2X + 3Y$, compute the correlation $\text{Cor}(Y, Z)$. (4 marks)

- (b) A statistician has attempted to fit the simple linear regression model

$$Y_i = \beta_0 + \beta_1 x_i + \epsilon_i, \quad \epsilon_i \stackrel{\text{i.i.d.}}{\sim} N(0, \sigma^2),$$

to the data $(x_1, y_1), (x_2, y_2), \dots, (x_n, y_n)$. Given the plot of the fitted residuals below, and the fact that $R^2 = 0.96$, briefly state and justify whether or not the model fits the data well. If you believe the model does not fit the data well, suggest what could be done to possibly improve the fit. (3 marks)

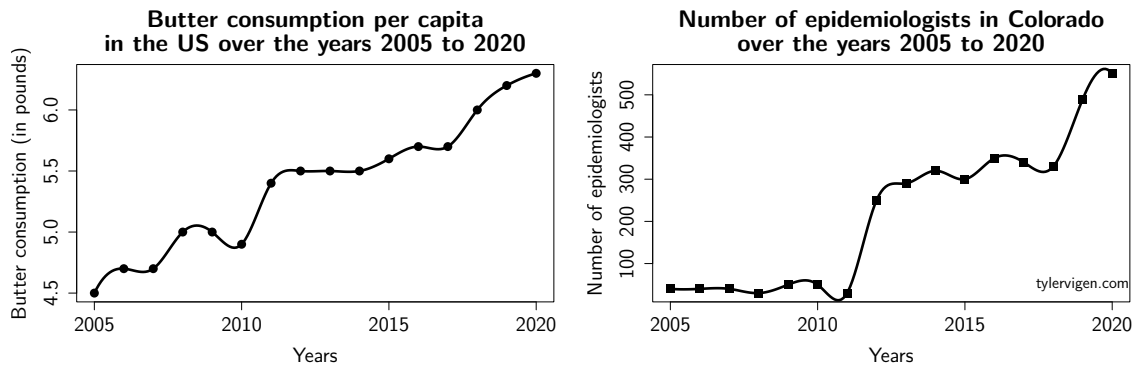


- (c) Briefly describe a real-world example in which linear regression can be used to investigate a relationship between two quantities X and Y . (2 marks)

[Question 4 continues on the next page]

[Question 4 continues on this page]

- (d) The figures below show X , the amount of butter consumed per capita (in pounds) in the US for the years 2005 to 2020, and Y , the number of epidemiologists in the state of Colorado over the same time period. The graphs appear to be very similar, and if one computes the sample correlation between the two data sets for X and Y , one obtains a value of $r = 0.92$. Taking this sample correlation value into account, comment on whether or not we can conclude that the two quantities X and Y are related and have any influence on each other. (2 marks)



- (e) Suppose that a medical researcher records survival times for mice which have been given a new treatment. The data, which show the survival times in days, are:

$$\{25, 34, 39, 46, 52\}.$$

For each of the following samples, state whether or not they are bootstrap samples of this data, and provide justification.

- (i) $\{34, 25, 46, 52, 52\}$ (2 marks)
- (ii) $\{25, 34, 46, 39\}$ (2 marks)

- (f) Suppose that $x_1, x_2, \dots, x_n$ are observations of the random variables $X_1, X_2, \dots, X_n$, which are all assumed to be independent and distributed according to a uniform distribution on the interval $[\theta, \theta + 1]$, where θ is unknown. Compute the maximum likelihood estimate of θ . (5 marks)

(Total: 20 marks)

Name of distribution of X	$\text{Im}X$	Parameters	p.m.f. $P(X = x)/\text{p.d.f } f_X(x)$
Discrete distributions			
Bernoulli: $X \sim \text{Bern}(p)$	$\{0, 1\}$	$p \in (0, 1)$	$p^x(1-p)^{1-x}$
Binomial: $X \sim \text{Bin}(n, p)$	$\{0, 1, \dots, n\}$	$n \in \mathbb{N}, p \in (0, 1)$	$\binom{n}{x} p^x (1-p)^{n-x}$
Hypergeometric: $X \sim \text{HGeom}(N, K, n)$	$\{0, 1, \dots, \min(n, K)\}$	$N \in \mathbb{N} \cup \{0\},$ $K, n \in \{0, 1, \dots, N\}$	$\frac{\binom{K}{x} \binom{N-K}{n-x}}{\binom{N}{n}},$ for $x \in \{0, 1, \dots, K\},$ $n-x \in \{0, 1, \dots, N-K\}$
Poisson: $X \sim \text{Poi}(\lambda)$	$\mathbb{N}_0 = \{0, 1, 2, \dots\}$	$\lambda > 0$	$\frac{\lambda^x}{x!} e^{-\lambda}$
Geometric (1): $X \sim \text{Geom}(p)$	$\mathbb{N} = \{1, 2, \dots\}$	$p \in (0, 1)$	$(1-p)^{x-1} p$
Geometric (2): $X \sim \text{Geom}(p)$	$\mathbb{N}_0 = \{0, 1, 2, \dots\}$	$p \in (0, 1)$	$(1-p)^x p$
Negative binomial: $X \sim \text{NBin}(r, p)$	$\mathbb{N}_0 = \{0, 1, 2, \dots\}$	$r \in \mathbb{N}, p \in (0, 1)$	$\binom{x+r-1}{r-1} p^r (1-p)^x$
Continuous distributions			
Uniform: $X \sim \text{U}(a, b)$	(a, b)	$a, b \in \mathbb{R}, a < b$	$\frac{1}{b-a}$
Exponential: $X \sim \text{Exp}(\lambda)$	$(0, \infty)$	$\lambda > 0$	$\lambda e^{-\lambda x}$
Gamma: $X \sim \text{Gamma}(\alpha, \beta)$	$(0, \infty)$	$\alpha, \beta > 0$	$\frac{\beta^\alpha}{\Gamma(\alpha)} x^{\alpha-1} e^{-\beta x},$ where $\Gamma(\alpha) = \int_0^\infty x^{\alpha-1} e^{-x} dx$
Chi-squared: $X \sim \chi^2(n)$ $(X \sim \chi_n^2)$	$(0, \infty)$	$n \in \mathbb{N}$	$\frac{1}{2\Gamma(n/2)} \left(\frac{x}{2}\right)^{n/2-1} e^{-x/2}$
Beta: $X \sim \text{Beta}(\alpha, \beta)$	$(0, 1)$	$\alpha, \beta > 0$	$\frac{1}{B(\alpha, \beta)} x^{\alpha-1} (1-x)^{\beta-1},$ where $B(\alpha, \beta) = \int_0^1 x^{\alpha-1} (1-x)^{\beta-1} dx$ $= \frac{\Gamma(\alpha)\Gamma(\beta)}{\Gamma(\alpha+\beta)}$
Normal: $X \sim \text{N}(\mu, \sigma^2)$	$\mathbb{R}$	$\mu \in \mathbb{R}, \sigma > 0$	$\frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$
Student-t: $X \sim \text{Student}(\nu)$	$\mathbb{R}$	$\nu > 0$	$\frac{\Gamma(\frac{\nu+1}{2})}{\sqrt{\nu\pi}\Gamma(\frac{\nu}{2})} \left(1 + \frac{x^2}{\nu}\right)^{-\frac{\nu+1}{2}}$

Table 1: $P(Z < z)$ where $Z \sim N(0, 1)$ for values of z between 1.00 and 3.99. One reads the table as $P(Z < z) = p_{r,c}$, where $z = z_r + z_c$ and z_r are the first two digits of z shown in row r , z_c is the third digit of z shown in column c , and $p_{r,c}$ is the (r, c) th entry. For example, $P(Z < 1.02) = 0.8461$, looking in row labelled 1.0 and column labelled 0.02.

z	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
1.0	0.8413	0.8438	0.8461	0.8485	0.8508	0.8531	0.8554	0.8577	0.8599	0.8621
1.1	0.8643	0.8665	0.8686	0.8708	0.8729	0.8749	0.8770	0.8790	0.8810	0.8830
1.2	0.8849	0.8869	0.8888	0.8907	0.8925	0.8944	0.8962	0.8980	0.8997	0.9015
1.3	0.9032	0.9049	0.9066	0.9082	0.9099	0.9115	0.9131	0.9147	0.9162	0.9177
1.4	0.9192	0.9207	0.9222	0.9236	0.9251	0.9265	0.9279	0.9292	0.9306	0.9319
1.5	0.9332	0.9345	0.9357	0.9370	0.9382	0.9394	0.9406	0.9418	0.9429	0.9441
1.6	0.9452	0.9463	0.9474	0.9484	0.9495	0.9505	0.9515	0.9525	0.9535	0.9545
1.7	0.9554	0.9564	0.9573	0.9582	0.9591	0.9599	0.9608	0.9616	0.9625	0.9633
1.8	0.9641	0.9649	0.9656	0.9664	0.9671	0.9678	0.9686	0.9693	0.9699	0.9706
1.9	0.9713	0.9719	0.9726	0.9732	0.9738	0.9744	0.9750	0.9756	0.9761	0.9767
2.0	0.9772	0.9778	0.9783	0.9788	0.9793	0.9798	0.9803	0.9808	0.9812	0.9817
2.1	0.9821	0.9826	0.9830	0.9834	0.9838	0.9842	0.9846	0.9850	0.9854	0.9857
2.2	0.9861	0.9864	0.9868	0.9871	0.9875	0.9878	0.9881	0.9884	0.9887	0.9890
2.3	0.9893	0.9896	0.9898	0.9901	0.9904	0.9906	0.9909	0.9911	0.9913	0.9916
2.4	0.9918	0.9920	0.9922	0.9925	0.9927	0.9929	0.9931	0.9932	0.9934	0.9936
2.5	0.9938	0.9940	0.9941	0.9943	0.9945	0.9946	0.9948	0.9949	0.9951	0.9952
2.6	0.9953	0.9955	0.9956	0.9957	0.9959	0.9960	0.9961	0.9962	0.9963	0.9964
2.7	0.9965	0.9966	0.9967	0.9968	0.9969	0.9970	0.9971	0.9972	0.9973	0.9974
2.8	0.9974	0.9975	0.9976	0.9977	0.9977	0.9978	0.9979	0.9979	0.9980	0.9981
2.9	0.9981	0.9982	0.9982	0.9983	0.9984	0.9984	0.9985	0.9985	0.9986	0.9986
3.0	0.9987	0.9987	0.9987	0.9988	0.9988	0.9989	0.9989	0.9989	0.9990	0.9990
3.1	0.9990	0.9991	0.9991	0.9991	0.9992	0.9992	0.9992	0.9992	0.9993	0.9993
3.2	0.9993	0.9993	0.9994	0.9994	0.9994	0.9994	0.9994	0.9995	0.9995	0.9995
3.3	0.9995	0.9995	0.9995	0.9996	0.9996	0.9996	0.9996	0.9996	0.9996	0.9997
3.4	0.9997	0.9997	0.9997	0.9997	0.9997	0.9997	0.9997	0.9997	0.9997	0.9998
3.5	0.9998	0.9998	0.9998	0.9998	0.9998	0.9998	0.9998	0.9998	0.9998	0.9998
3.6	0.9998	0.9998	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999
3.7	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999
3.8	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999
3.9	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000

Table 2: Values of t for $P(T < t)$, where T has Student's t -distribution with ν degrees of freedom, for $\nu \in \{1, 2, \dots, 30\}$. In this table, $P(T < t_{\nu,p}) = p$, where ν is the ν th row, p is the column with heading p , and $t_{\nu,p}$ is the corresponding entry in the table. For example, $t_{3,0.60} = 0.277$, looking in row labelled 3 and column labelled 0.60.

ν	0.60	0.667	0.75	0.80	0.875	0.90	0.95	0.975	0.99	0.995	0.999
1	0.325	0.577	1.000	1.376	2.414	3.078	6.314	12.706	31.821	63.657	318.31
2	0.289	0.500	0.816	1.061	1.604	1.886	2.920	4.303	6.965	9.925	22.327
3	0.277	0.476	0.765	0.978	1.423	1.638	2.353	3.182	4.541	5.841	10.215
4	0.271	0.464	0.741	0.941	1.344	1.533	2.132	2.776	3.747	4.604	7.173
5	0.267	0.457	0.727	0.920	1.301	1.476	2.015	2.571	3.365	4.032	5.893
6	0.265	0.453	0.718	0.906	1.273	1.440	1.943	2.447	3.143	3.707	5.208
7	0.263	0.449	0.711	0.896	1.254	1.415	1.895	2.365	2.998	3.499	4.785
8	0.262	0.447	0.706	0.889	1.240	1.397	1.860	2.306	2.896	3.355	4.501
9	0.261	0.445	0.703	0.883	1.230	1.383	1.833	2.262	2.821	3.250	4.297
10	0.260	0.444	0.700	0.879	1.221	1.372	1.812	2.228	2.764	3.169	4.144
11	0.260	0.443	0.697	0.876	1.214	1.363	1.796	2.201	2.718	3.106	4.025
12	0.259	0.442	0.695	0.873	1.209	1.356	1.782	2.179	2.681	3.055	3.930
13	0.259	0.441	0.694	0.870	1.204	1.350	1.771	2.160	2.650	3.012	3.852
14	0.258	0.440	0.692	0.868	1.200	1.345	1.761	2.145	2.624	2.977	3.787
15	0.258	0.439	0.691	0.866	1.197	1.341	1.753	2.131	2.602	2.947	3.733
16	0.258	0.439	0.690	0.865	1.194	1.337	1.746	2.120	2.583	2.921	3.686
17	0.257	0.438	0.689	0.863	1.191	1.333	1.740	2.110	2.567	2.898	3.646
18	0.257	0.438	0.688	0.862	1.189	1.330	1.734	2.101	2.552	2.878	3.610
19	0.257	0.438	0.688	0.861	1.187	1.328	1.729	2.093	2.539	2.861	3.579
20	0.257	0.437	0.687	0.860	1.185	1.325	1.725	2.086	2.528	2.845	3.552
21	0.257	0.437	0.686	0.859	1.183	1.323	1.721	2.080	2.518	2.831	3.527
22	0.256	0.437	0.686	0.858	1.182	1.321	1.717	2.074	2.508	2.819	3.505
23	0.256	0.436	0.685	0.858	1.180	1.319	1.714	2.069	2.500	2.807	3.485
24	0.256	0.436	0.685	0.857	1.179	1.318	1.711	2.064	2.492	2.797	3.467
25	0.256	0.436	0.684	0.856	1.178	1.316	1.708	2.060	2.485	2.787	3.450
26	0.256	0.436	0.684	0.856	1.177	1.315	1.706	2.056	2.479	2.779	3.435
27	0.256	0.435	0.684	0.855	1.176	1.314	1.703	2.052	2.473	2.771	3.421
28	0.256	0.435	0.683	0.855	1.175	1.313	1.701	2.048	2.467	2.763	3.408
29	0.256	0.435	0.683	0.854	1.174	1.311	1.699	2.045	2.462	2.756	3.396
30	0.256	0.435	0.683	0.854	1.173	1.310	1.697	2.042	2.457	2.750	3.385

Table 3: Values of x for $P(X < x)$, where $X \sim \chi^2(\nu)$, a chi-squared distribution with ν degrees of freedom. In this table, $P(T < \chi^2_{\nu,p}) = p$, where ν is the ν th row, p is the column with heading p , and $\chi^2_{\nu,p}$ is the corresponding entry in the table. For example, $\chi^2_{3,0.005} = 0.072$, looking in row labelled 3 and column labelled 0.005.

ν	0.001	0.005	0.01	0.025	0.05	0.1	0.125	0.2	0.25	0.333	0.5
1	0.000	0.000	0.000	0.001	0.004	0.016	0.025	0.064	0.102	0.186	0.455
2	0.002	0.010	0.020	0.051	0.103	0.211	0.267	0.446	0.575	0.811	1.386
3	0.024	0.072	0.115	0.216	0.352	0.584	0.692	1.005	1.213	1.568	2.366
4	0.091	0.207	0.297	0.484	0.711	1.064	1.219	1.649	1.923	2.378	3.357
5	0.210	0.412	0.554	0.831	1.145	1.610	1.808	2.343	2.675	3.216	4.351
6	0.381	0.676	0.872	1.237	1.635	2.204	2.441	3.070	3.455	4.074	5.348
7	0.598	0.989	1.239	1.690	2.167	2.833	3.106	3.822	4.255	4.945	6.346
8	0.857	1.344	1.646	2.180	2.733	3.490	3.797	4.594	5.071	5.826	7.344
9	1.152	1.735	2.088	2.700	3.325	4.168	4.507	5.380	5.899	6.716	8.343
10	1.479	2.156	2.558	3.247	3.940	4.865	5.234	6.179	6.737	7.612	9.342
11	1.834	2.603	3.053	3.816	4.575	5.578	5.975	6.989	7.584	8.514	10.341
12	2.214	3.074	3.571	4.404	5.226	6.304	6.729	7.807	8.438	9.420	11.340
13	2.617	3.565	4.107	5.009	5.892	7.042	7.493	8.634	9.299	10.331	12.340
14	3.041	4.075	4.660	5.629	6.571	7.790	8.266	9.467	10.165	11.245	13.339
15	3.483	4.601	5.229	6.262	7.261	8.547	9.048	10.307	11.037	12.163	14.339
16	3.942	5.142	5.812	6.908	7.962	9.312	9.837	11.152	11.912	13.083	15.338
17	4.416	5.697	6.408	7.564	8.672	10.085	10.633	12.002	12.792	14.006	16.338
18	4.905	6.265	7.015	8.231	9.390	10.865	11.435	12.857	13.675	14.931	17.338
19	5.407	6.844	7.633	8.907	10.117	11.651	12.242	13.716	14.562	15.859	18.338
20	5.921	7.434	8.260	9.591	10.851	12.443	13.055	14.578	15.452	16.788	19.337
21	6.447	8.034	8.897	10.283	11.591	13.240	13.873	15.445	16.344	17.720	20.337
22	6.983	8.643	9.542	10.982	12.338	14.041	14.695	16.314	17.240	18.653	21.337
23	7.529	9.260	10.196	11.689	13.091	14.848	15.521	17.187	18.137	19.587	22.337
24	8.085	9.886	10.856	12.401	13.848	15.659	16.351	18.062	19.037	20.523	23.337
25	8.649	10.520	11.524	13.120	14.611	16.473	17.184	18.940	19.939	21.461	24.337
26	9.222	11.160	12.198	13.844	15.379	17.292	18.021	19.820	20.843	22.399	25.336
27	9.803	11.808	12.879	14.573	16.151	18.114	18.861	20.703	21.749	23.339	26.336
28	10.391	12.461	13.565	15.308	16.928	18.939	19.704	21.588	22.657	24.280	27.336
29	10.986	13.121	14.256	16.047	17.708	19.768	20.550	22.475	23.567	25.222	28.336
30	11.588	13.787	14.953	16.791	18.493	20.599	21.399	23.364	24.478	26.165	29.336
35	14.688	17.192	18.509	20.569	22.465	24.797	25.678	27.836	29.054	30.894	34.336
40	17.916	20.707	22.164	24.433	26.509	29.051	30.008	32.345	33.660	35.643	39.335
45	21.251	24.311	25.901	28.366	30.612	33.350	34.379	36.884	38.291	40.407	44.335
50	24.674	27.991	29.707	32.357	34.764	37.689	38.785	41.449	42.942	45.184	49.335

Table 4: Values of x for $P(X < x)$, where $X \sim \chi^2(\nu)$, a chi-squared distribution with ν degrees of freedom. In this table, $P(T < \chi^2_{\nu,p}) = p$, where ν is the ν th row, p is the column with heading p , and $\chi^2_{\nu,p}$ is the corresponding entry in the table. For example, $\chi^2_{3,0.6} = 2.946$, looking in row labelled 3 and column labelled 0.6.

ν	0.6	0.66	0.75	0.80	0.87	0.9	0.95	0.975	0.99	0.995	0.999
1	0.708	0.936	1.323	1.642	2.354	2.706	3.841	5.024	6.635	7.879	10.828
2	1.833	2.197	2.773	3.219	4.159	4.605	5.991	7.378	9.210	10.597	13.816
3	2.946	3.405	4.108	4.642	5.739	6.251	7.815	9.348	11.345	12.838	16.266
4	4.045	4.579	5.385	5.989	7.214	7.779	9.488	11.143	13.277	14.860	18.467
5	5.132	5.730	6.626	7.289	8.625	9.236	11.070	12.833	15.086	16.750	20.515
6	6.211	6.867	7.841	8.558	9.992	10.645	12.592	14.449	16.812	18.548	22.458
7	7.283	7.992	9.037	9.803	11.326	12.017	14.067	16.013	18.475	20.278	24.322
8	8.351	9.107	10.219	11.030	12.636	13.362	15.507	17.535	20.090	21.955	26.125
9	9.414	10.215	11.389	12.242	13.926	14.684	16.919	19.023	21.666	23.589	27.877
10	10.473	11.317	12.549	13.442	15.198	15.987	18.307	20.483	23.209	25.188	29.588
11	11.530	12.414	13.701	14.631	16.457	17.275	19.675	21.920	24.725	26.757	31.264
12	12.584	13.506	14.845	15.812	17.703	18.549	21.026	23.337	26.217	28.300	32.910
13	13.636	14.595	15.984	16.985	18.939	19.812	22.362	24.736	27.688	29.819	34.528
14	14.685	15.680	17.117	18.151	20.166	21.064	23.685	26.119	29.141	31.319	36.123
15	15.733	16.761	18.245	19.311	21.384	22.307	24.996	27.488	30.578	32.801	37.697
16	16.780	17.840	19.369	20.465	22.595	23.542	26.296	28.845	32.000	34.267	39.252
17	17.824	18.917	20.489	21.615	23.799	24.769	27.587	30.191	33.409	35.718	40.790
18	18.868	19.991	21.605	22.760	24.997	25.989	28.869	31.526	34.805	37.156	42.312
19	19.910	21.063	22.718	23.900	26.189	27.204	30.144	32.852	36.191	38.582	43.820
20	20.951	22.133	23.828	25.038	27.376	28.412	31.410	34.170	37.566	39.997	45.315
21	21.991	23.201	24.935	26.171	28.559	29.615	32.671	35.479	38.932	41.401	46.797
22	23.031	24.268	26.039	27.301	29.737	30.813	33.924	36.781	40.289	42.796	48.268
23	24.069	25.333	27.141	28.429	30.911	32.007	35.172	38.076	41.638	44.181	49.728
24	25.106	26.397	28.241	29.553	32.081	33.196	36.415	39.364	42.980	45.559	51.179
25	26.143	27.459	29.339	30.675	33.247	34.382	37.652	40.646	44.314	46.928	52.620
26	27.179	28.520	30.435	31.795	34.410	35.563	38.885	41.923	45.642	48.290	54.052
27	28.214	29.580	31.528	32.912	35.570	36.741	40.113	43.195	46.963	49.645	55.476
28	29.249	30.639	32.620	34.027	36.727	37.916	41.337	44.461	48.278	50.993	56.892
29	30.283	31.697	33.711	35.139	37.881	39.087	42.557	45.722	49.588	52.336	58.301
30	31.316	32.754	34.800	36.250	39.033	40.256	43.773	46.979	50.892	53.672	59.703
35	36.475	38.024	40.223	41.778	44.753	46.059	49.802	53.203	57.342	60.275	66.619
40	41.622	43.275	45.616	47.269	50.424	51.805	55.758	59.342	63.691	66.766	73.402
45	46.761	48.510	50.985	52.729	56.052	57.505	61.656	65.410	69.957	73.166	80.077
50	51.892	53.733	56.334	58.164	61.647	63.167	67.505	71.420	76.154	79.490	86.661

BSc and MSci EXAMINATIONS (MATHEMATICS)

May 2026

This paper is also taken for the relevant examination for the Associateship.

MATH40005

Probability and Statistics (Solutions)

Setter's signature

.....

Checker's signature

.....

Editor's signature

.....

1. (a) The probability space $(\Omega, \mathcal{F}, \mathbb{P})$ for rolling a fair six-sided die twice can be described as follows:

seen/sim.seen ↓

Sample Space (Ω): The sample space consists of all possible ordered outcomes of rolling the die twice. Since each die has six faces, we have:

$$\Omega = \{(1, 1), (1, 2), \dots, (1, 6), (2, 1), \dots, (6, 6)\}.$$

1, A

Event Space ($\mathcal{F}$): The event space is the collection of all subsets of Ω , representing all possible events. This includes the empty set, individual outcomes, combinations of outcomes, and the entire sample space:

$$\mathcal{F} = \{\emptyset, \{(1, 1)\}, \{(1, 2)\}, \dots, \{(6, 6)\}, \{(1, 1), (1, 2)\}, \dots, \Omega\}.$$

1, A

Probability Measure ($\mathbb{P}$): The probability measure assigns a real number between 0 and 1 to each event in $\mathcal{F}$. For a fair die, each outcome is equally likely, so

$$\mathbb{P}(\{(i, j)\}) = \frac{1}{36}, \quad \text{for each } (i, j) \in \Omega.$$

For any event $A \in \mathcal{F}$, the probability is:

$$\mathbb{P}(A) = \sum_{(i,j) \in A} \mathbb{P}(\{(i, j)\}) = \frac{|A|}{36},$$

where $|A|$ is the number of outcomes in A .

1, A

- (b) (i) An outcome of the EuroMillions draw consists of (i) choosing 5 distinct main numbers from 50, and (ii) choosing 2 distinct Lucky Stars from 12.

meth seen ↓

Since the order of selection does not matter in either case, the number of ways to choose the main numbers is $\binom{50}{5}$, and the number of ways to choose the Lucky Stars is $\binom{12}{2}$.

As these choices are independent, the total number of possible outcomes is:

$$\binom{50}{5} \cdot \binom{12}{2}.$$

[1 mark for counting the main numbers, 1 mark for counting the Lucky Stars, 1 mark for combining the two counts]

3, A

- (ii) Let S denote the event that a ticket matches both Lucky Stars, and let M_3 denote the event that a ticket matches exactly 3 main numbers.

We are asked to compute the conditional probability $P(S \mid M_3)$.

Given that the ticket already matches exactly 3 main numbers, the only remaining uncertainty is whether the two Lucky Stars match.

There are $\binom{12}{2}$ possible Lucky Star pairs in total, and exactly one of these pairs matches the drawn Lucky Stars. Hence,

$$P(S \mid M_3) = \frac{1}{\binom{12}{2}} = \frac{1}{66}.$$

[1 mark for introducing notation, 1 mark for the justification, 1 mark for the correct probability]

3, B

- (iii) To win the secondary prize, a ticket must match exactly 3 of the 5 drawn main numbers, and match both Lucky Stars.

unseen ↓

The number of ways to choose exactly 3 correct main numbers from the 5 drawn is $\binom{5}{3}$, and the number of ways to choose the remaining 2 incorrect numbers from the remaining 45 numbers is $\binom{45}{2}$.

Hence, the probability that a ticket matches exactly 3 main numbers is:

$$P(M_3) = \frac{\binom{5}{3} \binom{45}{2}}{\binom{50}{5}}.$$

Using the conditional probability from part (ii),

$$P(\text{secondary prize}) = P(M_3) P(S \mid M_3) = \frac{\binom{5}{3} \binom{45}{2}}{\binom{50}{5}} \cdot \frac{1}{\binom{12}{2}}.$$

[1 mark for counting exactly 3 matches, 2 marks for the justification, 1 mark for the correct final expression]

4, D

- (iv) We now assume that the two Lucky Stars are drawn with replacement, so the same star may appear twice. The order of the stars does not matter.

meth seen ↓

This is equivalent to counting the number of multisets of size 2 chosen from 12 stars.

Using the stars and bars argument, the number of such multisets is:

$$\binom{12 + 2 - 1}{2} = \binom{13}{2}.$$

Therefore, there are $\binom{13}{2} = 78$ different Lucky Star outcomes.

[1 mark for recognising a stars and bars argument, 1 mark for the correct answer]

Alternative solutions:

[Alternative solution 1] We have 12^2 ordered outcomes obtained when two Lucky Stars are drawn with replacement. This counts all ordered pairs (a, b) where $a, b \in \{1, \dots, 12\}$. However, in the lottery the order of the two Lucky Stars does not matter.

Among these 12^2 outcomes, there are 12×11 ordered outcomes where the two stars are different. Each unordered pair of distinct stars appears twice in this list (for example (a, b) and (b, a) correspond to the same outcome when order is ignored). Hence the ordered count includes $\frac{12 \times 11}{2}$ duplicate outcomes, which must be removed. Therefore the number of different Lucky Star outcomes is:

$$12^2 - \frac{12 \times 11}{2} = 144 - 66 = 78.$$

[Alternative solution 2] We can count directly by separating the cases where the two stars are different and where they are the same.

If the two Lucky stars are different, we must choose two distinct stars from the 12 available, and the order does not matter. Hence the number of such outcomes is:

$$\binom{12}{2} = \frac{12 \times 11}{2}.$$

In addition, there are 12 outcomes in which both Lucky Stars are the same (namely $(1, 1), (2, 2), \dots, (12, 12)$). Therefore the total number of possible Lucky Star outcomes is:

$$\frac{12 \times 11}{2} + 12 = 66 + 12 = 78.$$

[Provided the reasoning is clearly explained, such arguments would also receive full marks.]

2, B

(c) Clearly, $p_X(x) \geq 0$ for all x . To check that it sums to 1:

meth seen ↓

$$\sum_{x=n}^{\infty} p_X(x) = \sum_{x=n}^{\infty} p(1-p)^{x-n} \stackrel{k=x-n}{=} p \sum_{k=0}^{\infty} (1-p)^k = p \cdot \frac{1}{1-(1-p)} = 1.$$

Hence, p_X is a valid probability mass function.

2, A

For $x \geq n$, the cumulative distribution function is:

$$F_X(x) = P(X \leq x) = \sum_{k=n}^{\lfloor x \rfloor} p_X(k) = \sum_{k=n}^{\lfloor x \rfloor} p(1-p)^{k-n} = p \sum_{k=0}^{\lfloor x \rfloor - n} (1-p)^k.$$

Using the finite geometric sum formula:

$$F_X(x) = p \cdot \frac{1 - (1-p)^{\lfloor x \rfloor - n + 1}}{1 - (1-p)} = 1 - (1-p)^{\lfloor x \rfloor - n + 1}.$$

For $x < n$, clearly $F_X(x) = 0$. Hence, for general $x \in \mathbb{R}$, we have:

$$F_X(x) = \begin{cases} 0, & x < n, \\ 1 - (1-p)^{\lfloor x \rfloor - n + 1}, & x \geq n. \end{cases}$$

[1 mark non-negativity condition, 1 mark for sum equal to 1, 2 marks for obtaining the cdf, 1 mark for the justification]

3, C

2. (a) We roll two fair dice and write the sample space as $\Omega = \{(i, j) : i, j = 1, \dots, 6\}$.
All outcomes are equally likely, so $\text{card}(\Omega) = 6^2 = 36$.

seen $\Downarrow$

We define the following events:

$$A_1 = \{\text{first roll is odd}\}, A_2 = \{\text{second roll is odd}\}, A_3 = \{\text{sum of the two rolls is odd}\}.$$

Each event has probability:

$$P(A_1) = P(A_2) = P(A_3) = \frac{1}{2}.$$

We compute:

$$P(A_1 \cap A_2) = \frac{1}{4} = P(A_1)P(A_2),$$

$$P(A_1 \cap A_3) = \frac{1}{4} = P(A_1)P(A_3),$$

$$P(A_2 \cap A_3) = \frac{1}{4} = P(A_2)P(A_3).$$

Hence, A_1 , A_2 , and A_3 are pairwise independent.

However,

$$P(A_1 \cap A_2 \cap A_3) = 0,$$

since the sum of two odd numbers is even.

On the other hand,

$$P(A_1)P(A_2)P(A_3) = \frac{1}{2} \cdot \frac{1}{2} \cdot \frac{1}{2} = \frac{1}{8}.$$

Since

$$P(A_1 \cap A_2 \cap A_3) \neq P(A_1)P(A_2)P(A_3),$$

the events A_1 , A_2 , and A_3 are not independent.

[Multiple examples possible, 1 mark for providing an example and describing the experiment setup, 2 marks for the justification]

3, A

- (b) (i) For any $x \in \mathbb{R}$,

meth seen $\Downarrow$

$$F_X(x) = \lim_{y \rightarrow \infty} F_{X,Y}(x, y).$$

If $x < 0$, then $F_X(x) = 0$. For $x \geq 0$,

$$F_X(x) = \lim_{y \rightarrow \infty} (1 - e^{-x} - e^{-y} + e^{-x-y}) = 1 - e^{-x}.$$

Thus,

$$F_X(x) = \begin{cases} 1 - e^{-x}, & x \geq 0, \\ 0, & x < 0. \end{cases}$$

Differentiating, the density of X is given by:

$$f_X(x) = \begin{cases} e^{-x}, & x \geq 0, \\ 0, & x < 0. \end{cases}$$

Similarly,

$$F_Y(y) = \lim_{x \rightarrow \infty} F_{X,Y}(x, y).$$

If $y < 0$, then $F_Y(y) = 0$. For $y \geq 0$,

$$F_Y(y) = \lim_{x \rightarrow \infty} (1 - e^{-x} - e^{-y} + e^{-x-y}) = 1 - e^{-y}.$$

Hence,

$$F_Y(y) = \begin{cases} 1 - e^{-y}, & y \geq 0, \\ 0, & y < 0. \end{cases}$$

Differentiating, the density of Y is given by:

$$f_Y(y) = \begin{cases} e^{-y}, & y \geq 0, \\ 0, & y < 0. \end{cases}$$

Both X and Y have density e^{-t} for $t \geq 0$, so

$$X \sim \text{Exponential}(1), \quad Y \sim \text{Exponential}(1).$$

[1 mark for each marginal, 1 mark for each density, 1 mark for identifying the distributions and the parameters]

5, A

(ii) For $x \geq 0$ and $y \geq 0$,

$$F_X(x)F_Y(y) = (1 - e^{-x})(1 - e^{-y}) = 1 - e^{-x} - e^{-y} + e^{-x-y} = F_{X,Y}(x, y).$$

Thus,

$$F_{X,Y}(x, y) = F_X(x)F_Y(y),$$

which shows that X and Y are independent.

[1 mark for showing independence, 1 mark for the justification]

2, B

(iii) The moment generating function of the random variable X is given by:

$$M_X(t) = E[e^{tX}] = \int_0^\infty e^{tx} f_X(x) dx = \int_0^\infty e^{tx} e^{-x} dx = \int_0^\infty e^{-(1-t)x} dx = \frac{1}{1-t}, t < 1.$$

[1 mark for applying the definition, 1 mark for the justification, 1 mark for the correct answer]

3, C

(iv) Let $Z = X^2$. Then the cumulative distribution function is given by:

$$F_Z(z) = P(Z \leq z) = P(X^2 \leq z) = P(X \leq \sqrt{z}) = F_X(\sqrt{z}) = 1 - e^{-\sqrt{z}}, \quad z \geq 0.$$

[1 mark for applying the definition, 1 mark for the justification, 1 mark for the correct answer]

3, B

(c) (i) Each Y_i is an indicator of whether student i 's birthday is on January 1. Since each birthday is equally likely:

unseen ↓

$$P(Y_i = 1) = \frac{1}{d}, \quad P(Y_i = 0) = 1 - \frac{1}{d},$$

hence, $Y_i \sim \text{Bernoulli}(\frac{1}{d})$.

The total number of matches is $B = \sum_{i=1}^n Y_i$. Since the Y_i are independent, B is a sum of independent Bernoulli random variables, hence, $B \sim \text{Binomial}(n, \frac{1}{d})$.

[1 mark for each distribution and its parameter(s)]

2, D

(ii) The probability that at least one student has a birthday on January 1 is the complement of no one having it:

$$P(B \geq 1) = 1 - P(B = 0) = 1 - \left(1 - \frac{1}{d}\right)^n.$$

[1 mark for the justification, 1 mark for the correct answer]

2, D

3. (a) $a = \text{Var}[X]$.

seen ↓

$b = E[X]$.

1, A

1, A

c is the median of F_X (or c is the median of X).

1, B

(b) For any random variable X , an exercise in the notes gives

$$\begin{aligned}\text{Var}[X] &= E[X^2] - (E[X])^2 \\ \Rightarrow E[X^2] &= (E[X])^2 + \text{Var}[X].\end{aligned}$$

1, A

Applying this identity to the estimation error $\hat{\Theta} - \theta$, one obtains the mean squared error as

$$\begin{aligned}E[(\hat{\Theta} - \theta)^2] &= (E[\hat{\Theta} - \theta])^2 + \text{Var}[\hat{\Theta} - \theta] \\ \Rightarrow E[(\hat{\Theta} - \theta)^2] &= (b_\theta(\hat{\Theta}))^2 + \text{Var}[\hat{\Theta}],\end{aligned}$$

since $\text{Var}[\hat{\Theta} - \theta] = \text{Var}[\hat{\Theta}]$, by a property of the variance, and since the bias is defined as $b_\theta(\hat{\Theta}) = E[\hat{\Theta}] - \theta = E[\hat{\Theta} - \theta]$, by the linearity of expectation.

2, A

[1 mark for recalling identity and 2 marks mentioning and using linearity of expectation and property of the variance.]

(c) (i) A scatterplot.

sim. seen ↓

1, A

(ii) A box plot.

1, A

- (d) (i) There is a theorem in the notes which proves, for $X_1, \dots, X_n \stackrel{\text{i.i.d.}}{\sim} N(\mu, \sigma^2)$, that

$$\frac{(n-1)S^2}{\sigma^2} \sim \chi_{n-1}^2.$$

1, A

We can construct a $1 - \alpha$ confidence interval as follows:

$$\begin{aligned} P\left(\chi_{n-1, \alpha/2}^2 < \frac{(n-1)S^2}{\sigma^2} < \chi_{n-1, 1-\alpha/2}^2\right) &= 1 - \alpha \\ P\left(\left\{\chi_{n-1, \alpha/2}^2 < \frac{(n-1)S^2}{\sigma^2}\right\} \text{ and } \left\{\frac{(n-1)S^2}{\sigma^2} < \chi_{n-1, 1-\alpha/2}^2\right\}\right) &= 1 - \alpha \\ P\left(\left\{\frac{\sigma^2}{(n-1)S^2} < \frac{1}{\chi_{n-1, \alpha/2}^2}\right\} \text{ and } \left\{\frac{1}{\chi_{n-1, 1-\alpha/2}^2} < \frac{\sigma^2}{(n-1)S^2}\right\}\right) &= 1 - \alpha \\ P\left(\frac{1}{\chi_{n-1, 1-\alpha/2}^2} < \frac{\sigma^2}{(n-1)S^2} < \frac{1}{\chi_{n-1, \alpha/2}^2}\right) &= 1 - \alpha \\ P\left(\frac{(n-1)S^2}{\chi_{n-1, 1-\alpha/2}^2} < \sigma^2 < \frac{(n-1)S^2}{\chi_{n-1, \alpha/2}^2}\right) &= 1 - \alpha. \end{aligned}$$

1, A

Since $1 - \alpha = 0.99 \Rightarrow \alpha = 0.01 \Rightarrow \alpha/2 = 0.005$, and $n - 1 = 11$, using Table 3 we compute $\chi_{n-1, \alpha/2}^2 = \chi_{11, 0.005}^2 = 2.603$. Then since $1 - \alpha/2 = 0.995$, from Table 4 one obtains $\chi_{n-1, 1-\alpha/2}^2 = \chi_{11, 0.995}^2 = 26.757$.

Then the realised confidence interval is

$$\begin{aligned} \left(\frac{(n-1)s^2}{\chi_{n-1, 1-\alpha/2}^2}, \frac{(n-1)s^2}{\chi_{n-1, \alpha/2}^2}\right) &= \left(\frac{(11)(3)}{(26.757)}, \frac{(11)(3)}{(2.603)}\right) \\ &\approx (1.23, 12.68), \end{aligned}$$

where the final expression can be left as a pair of fractions.

2, A

[1 mark for stating that scaled S^2 follows χ_{n-1}^2 , 1 mark for correctly deriving/writing down the theoretical $1 - \alpha$ confidence interval, 2 marks for obtaining the correct critical values from the tables and substituting in correctly.]

- (ii) The null hypothesis is $\mu = 180$, and the alternative hypothesis is $\mu \neq 180$.

1, A

Since the variance σ^2 is unknown, we can define the statistic

$$T = \frac{\mu - \bar{X}}{S/\sqrt{n}}, \quad \text{where } S^2 = \frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X})^2 \quad \text{and} \quad \bar{X} = \frac{1}{n} \sum_{i=1}^n X_i,$$

and then under the null hypothesis $T \sim t_{n-1}$, a t -distribution with $n - 1$ degrees of freedom.

1, A

Since under the null hypothesis $\mu = 180$, and for the observed data $\bar{x} = 178.5$, $s = \sqrt{3}$ and $n = 12$, the realised test statistic is

$$t = \frac{\mu - \bar{x}}{s/\sqrt{n}} = \frac{180 - 178.5}{\sqrt{3}/\sqrt{12}} = \frac{1.5}{1/\sqrt{4}} = \frac{1.5}{0.5} = 3.$$

1, A

Looking at the t -distribution table, if F_T is the t -distribution with $n - 1 = 11$ degrees of freedom, Table 3 gives $F_T(2.718) = 0.99$ and $F_T(3.106) = 0.995$.

1, C

Since this is a two-sided test, the p -value for this test statistic is

$$p = 2 \cdot \min\{F_T(3), 1 - F_T(3)\}.$$

Since

$$2 \cdot \min\{F_T(3.106), 1 - F_T(3.106)\} = 2 \cdot \min\{0.995, 0.005\} = 0.01,$$

and

$$2 \cdot \min\{F_T(2.718), 1 - F_T(2.718)\} = 2 \cdot \min\{0.99, 0.01\} = 0.02,$$

then

$$p = 2 \cdot \min\{F_T(3), 1 - F_T(3)\} \in (0.01, 0.02).$$

1, D

- (e) Suppose that a test statistic $t(\mathbf{x})$ is computed from data $\mathbf{x}$. The p -value can be interpreted as the probability of observing a test statistic at least as extreme as $t(\mathbf{x})$ under the assumption that the null hypothesis H_0 is true.

2, B

If the test statistic $t(\mathbf{X})$ follows a continuous distribution under H_0 , then the p -value will follow a $U(0, 1)$ distribution.

1, C

4. (a) Using the linearity and symmetry of the covariance function,

sim. seen ↓

$$\begin{aligned}\text{Cov}(Y, Z) &= \text{Cov}(Y, 2X + 3Y) \\ &= \text{Cov}(Y, 2X) + \text{Cov}(Y, 3Y) \\ &= \text{Cov}(2X, Y) + 3\text{Cov}(Y, Y) \\ &= \text{Cov}(2X, Y) + 3\text{Var}(Y) \\ &= 2\text{Cov}(X, Y) + 3\text{Var}(Y).\end{aligned}$$

1, C

Now, since the correlation is defined as $\text{Cor}(X, Y) = \frac{\text{Cov}(X, Y)}{\sqrt{\text{Var}(X)}\sqrt{\text{Var}(Y)}}$, we have

$$\text{Cov}(X, Y) = \sqrt{\text{Var}(X)}\sqrt{\text{Var}(Y)}\text{Cor}(X, Y) = \sqrt{4}\sqrt{9}\left(\frac{1}{3}\right) = 2 \cdot 3\left(\frac{1}{3}\right) = 2.$$

Then,

$$\begin{aligned}\text{Cov}(Y, Z) &= 2\text{Cov}(X, Y) + 3\text{Var}(Y) \\ &= 2(2) + 3(9) \\ &= 4 + 27 \\ &= 31.\end{aligned}$$

1, A

Now, by the properties of the variance,

$$\begin{aligned}\text{Var}(Z) &= \text{Var}(2X + 3Y) \\ &= \text{Var}(2X) + \text{Var}(3Y) + 2\text{Cov}(2X, 3Y) \\ &= 4\text{Var}(X) + 9\text{Var}(Y) + 2(2)(3)\text{Cov}(X, Y) \\ &= 4(4) + 9(9) + 12(2) \\ &= 16 + 81 + 24 \\ &= 121.\end{aligned}$$

1, B

Finally, we can compute the correlation of Y and Z as

$$\begin{aligned}\text{Cor}(Y, Z) &= \frac{\text{Cov}(Y, Z)}{\sqrt{\text{Var}(Y)}\sqrt{\text{Var}(Z)}} = \frac{31}{\sqrt{9}\sqrt{121}} = \frac{31}{3 \cdot 11} \\ &= \frac{31}{33}.\end{aligned}$$

1, A

- (b) Despite the value of $R^2 = 0.96$ being close to 1, the residual plot indicates that the data does not fit the model well.

1, C

This is because the residuals form a “U-shape” which implies that the residuals are correlated and not independent.

1, D

The fit of the model could potentially be improved by first transforming the quantities before fitting the simple linear regression model, i.e. fitting

$$g(Y_i) = \beta_0 + \beta_1 f(x_i) + \epsilon_i, \quad \epsilon_i \stackrel{\text{i.i.d.}}{\sim} N(0, \sigma^2),$$

where f and g are transformations such as $f(x) = x^2$, $g(Y) = \log(Y)$, etc.

1, D

- (c) There are several possible real-world examples for the two quantities X and Y , such as:

- * The boiling point of water vs air pressure (seen in class).
- * The temperature on a particular day vs the amount of electricity (or gas, or water) consumption.
- * The amount of revenue a film generates through ticket sales vs the advertising budget for the film.
- * The selling price of a house vs the property's floor area (size).
- * The cost of a flight ticket vs the duration of the flight in hours and minutes.

2, D

[As long as the suggestion is reasonable, 2 marks can be awarded].

- (d) Although the correlation is 0.92, a high correlation does not imply causation, and so one cannot conclude that the number of epidemiologists in Colorado is related to or influenced by the amount of butter consumed in the US, or vice versa.

2, C

- (e) (i) This is a bootstrap sample of the data, because it has the same number of observations as the original sample and all of the values in the bootstrap sample are values in the original sample.

2, B

- (ii) This is not a bootstrap sample, because it does not have the same number of values as the original sample (this sample has 4 values while the original sample has 5 values).

2, B

(f) The probability density function of each X_i given θ , for $i = 1, 2, \dots, n$ is

sim. seen $\Downarrow$

$$f(x_i|\theta) = \begin{cases} \frac{1}{(\theta+1) - \theta} & \text{if } x_i \in [\theta, \theta+1], \\ 0 & \text{otherwise.} \end{cases}$$

$$= \begin{cases} 1 & \text{if } \theta \leq x_i \leq \theta+1, \\ 0 & \text{otherwise.} \end{cases}$$

1, B

Therefore, the joint likelihood is

$$\begin{aligned} L(\theta|\mathbf{x}) &= f(\mathbf{x}|\theta) \\ &= \prod_{i=1}^n f(x_i|\theta) \quad (\text{by independence}) \\ &= \begin{cases} 1 & \text{if } \theta \leq x_i \leq \theta+1, \text{ for all } i = 1, 2, \dots, n \\ 0 & \text{otherwise.} \end{cases} \end{aligned}$$

1, B

The condition that $\theta \leq x_i$ for all $i = 1, 2, \dots, n$ is equivalent to the condition that $\theta \leq \min\{x_1, \dots, x_n\}$.

Similarly, the condition that $x_i \leq \theta+1$ for all $i = 1, 2, \dots, n$ is equivalent to the condition that $x_i - 1 \leq \theta$ for all $i = 1, 2, \dots, n$, which in turn is equivalent to $\max\{x_1, \dots, x_n\} - 1 \leq \theta$.

Therefore the joint likelihood can be rewritten as

$$L(\theta|\mathbf{x}) = \begin{cases} 1 & \text{if } \max\{x_1, \dots, x_n\} - 1 \leq \theta \leq \min\{x_1, \dots, x_n\}, \\ 0 & \text{otherwise.} \end{cases}$$

2, D

Therefore, any value in the interval

$$[\max\{x_1, \dots, x_n\} - 1, \min\{x_1, \dots, x_n\}],$$

could be chosen as a maximum likelihood estimate of θ , and the maximum likelihood estimate is not unique.

1, D

Review of mark distribution:

Total A marks: 32 of 32 marks

Total B marks: 20 of 20 marks

Total C marks: 12 of 12 marks

Total D marks: 16 of 16 marks

Total marks: 80 of 80 marks