Imperial College London

MATH40005

BSc, MSci and MSc EXAMINATIONS (MATHEMATICS)
May 2023

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

Probability and Statistics

Date: 11 May 2023

Time: 14:00 – 17:00 (BST)

Time Allowed: 3hrs

This paper has 6 Questions.

Please Answer Each Question in a Separate Answer Booklets

Candidates should start their answers 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.

Allow margins for marking.

Credit will be given for all questions attempted.

Each question carries equal weight.

DO NOT OPEN THIS PAPER UNTIL THE INVIGILATOR TELLS YOU TO

Throughout the exam, we assume that $(\Omega, \mathcal{F}, \mathrm{P})$ denotes a probability space.

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

1. (a) Consider a school class consisting of 20 pupils. What is the probability that at least two pupils have their birthdays in the same week? Please state clearly any (additional) assumptions you make when solving this problem. (5 marks)

(b) Suppose that, in Imperial's Mathematics MSc programmes in 2022-2023, there are n MSc students with $n \in \mathbb{N}, 3 \leq n \leq 300$. Suppose that none of the MSc students was born in a leap year. What is the probability, as a function of n, that exactly two MSc students share the same birthday? Please state clearly any (additional) assumptions you make when solving this problem.

(5 marks)

(c) Suppose that a father takes his sick child to the GP. The GP knows that 95% of the children who are sick at the moment have the flu, while the other 5% have measles. Luckily, there are no other illnesses circulating in the neighbourhood at the moment and you may assume that a child does not suffer from the flu and measles simultaneously. A common symptom of measles is a rash. Suppose that the probability of having a rash given that a child has measles is 90%. However, in rare cases, a child suffering from flu might develop a rash as well. It is known that the probability that a child develops a rash given that it has the flu is 1%. The GP examines the child and finds that it has a rash. What is the probability that the child has measles?

(5 marks)

(d) Suppose that X and Y are discrete random variables on $(\Omega, \mathcal{F})$ and that $A \in \mathcal{F}$. Show that $Z : \Omega \to \mathbb{R}$, with

$$Z(\omega) = \begin{cases} X(\omega), & \text{if } \omega \in A, \\ Y(\omega), & \text{if } \omega \in A^c, \end{cases}$$

is a discrete random variable on $(\Omega, \mathcal{F})$.

(5 marks)

(Total: 20 marks)

2. (a) Consider non-empty sets $U, V \subseteq \mathbb{R}$ and a function $f : U \to V$. For any subset $A \subseteq V$, define

$$f^{-1}(A) = \{u \in U : f(u) \in A\}.$$

(i) Show that, for all $u \in U$, $\mathbb{I}_A(f(u)) = \mathbb{I}_{f^{-1}(A)}(u)$. [*Hint:* For an arbitrary non-empty set B, $\mathbb{I}_B(\cdot)$ denotes the indicator function associated with the set B.]

(2 marks)

(ii) Show that, if $\mathcal{V}$ is a sigma-algebra on V, then

$$f^{-1}(\mathcal{V}) = \{f^{-1}(A) : A \in \mathcal{V}\}$$

is a sigma-algebra on U.

(6 marks)

(b) Let $X_1, X_2, \ldots$ denote independent and identically distributed random variables with uniform distribution on the interval $(0, 1)$. Let N denote a random variable with Poisson distribution with rate $\lambda > 0$, which is independent of $X_1, X_2, \ldots$. For $0 < x < 1$, find

$$\mathrm{P}(\min\{X_1, X_2, \ldots, X_{N+1}\} > x).$$

(6 marks)

(c) Let X and Y denote independent and standard normally distributed random variables. Find

$$\mathrm{P}(X > 0, Y > -X).$$

(6 marks)

(Total: 20 marks)

3. (a) Suppose that $X \sim \mathrm{Poi}(\lambda), Y \sim \mathrm{Poi}(\mu)$, for $\lambda > 0, \mu > 0$. Does $X+Y$ follow a Poisson distribution? If so, prove it, otherwise, provide a counterexample. (3 marks)

(b) Consider two jointly continuous random variables X and Y with joint probability density function given by

$$f_{X,Y}(x,y) = \begin{cases} c(x+y), & \text{for } 0 < \sqrt{x} < y < 1, \\ 0, & \text{otherwise,} \end{cases}$$

for a constant $c \in \mathbb{R}$.

(i) Show that $c = \frac{20}{7}$ (1 mark)

(ii) Find the marginal density functions of X and Y. (2 marks)

(iii) Explain whether X and Y are independent. (1 mark)

(iv) For $y \in \mathbb{R}$, find the conditional distribution function $\mathrm{P}(Y \leq y | X = x)$ and specify for which values of x it is well-defined. (4 marks)

(c) Consider the continuous random variable $X \sim \mathrm{U}(1,3)$.

(i) Find the cumulative distribution function and probability density function of X^4.
(3 marks)

(ii) Calculate $\mathrm{E}(X^4)$. (2 marks)

(d) Let $X \sim \mathrm{Bin}(n,p)$ and $Y \sim \mathrm{Bin}(m,p)$ denote independent Binomial random variables, where $n, m \in \mathbb{N}, p \in (0,1)$. Compute the moment generating function of $X+Y$. Name the distribution of $X+Y$. (4 marks)

(Total: 20 marks)

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

$$\mathrm{E}[(\widehat{\Theta}-\theta)^2] = \left[b_\theta(\widehat{\Theta})\right]^2 + \mathrm{Var}(\widehat{\Theta}).$$

(3 marks)

(b) Suppose that a population is taking part in a vote with two options, labelled A and B. An unknown proportion p of the voters supports option B. Suppose it is possible to interview a sample of n randomly selected voters and record $\widehat{p}$, the proportion of that sample that supports option B. Stating all assumptions and results used, what is the minimum value of n that needs to be chosen so that, with confidence at least 99%, $\widehat{p}$ is within 0.01 of p?

(5 marks)

(c) For this question, Tables 1 and 2 on pages 10 and 11 may be helpful.

(i) Suppose $X_1, X_2, \ldots, X_n$ are independent and identically distributed random variables following a normal distribution with mean θ and variance σ^2. The value of θ is unknown, but σ^2 is known to be $\sigma^2 = 5$. Suppose the random variables $\mathbf{X} = (X_1, X_2, \ldots, X_n)$ are observed as $\mathbf{x} = (x_1, x_2, \ldots, x_n)$. Given that $\overline{x} = 10.2$ and $n = 20$, construct a 97% confidence interval for θ. (4 marks)

(ii) Suppose a 99% confidence interval has been computed for θ as $(5.6, 14.8)$. From the frequentist view of statistics, provide a concise interpretation of this confidence interval referencing θ, the data, the interval and the level of confidence. (3 marks)

(d) Suppose that the following ten observations are recorded:

$$\{14.3,\ 8.3,\ 15.9,\ 12.7,\ 9.4,\ 24.0,\ 7.2,\ 6.1,\ 11.6,\ 10.5\}$$

Showing all working:

(i) Compute the lower quartile $q_{0.25}$ and the upper quartile $q_{0.75}$ of this dataset. (2 marks)

(ii) State Tukey's criterion for identifying outliers, and use this rule to identify any outliers in this dataset. (3 marks)

(Total: 20 marks)

5. (a) Suppose that the head of a scientific lab is new to hypothesis testing and has hired you as a consulting statistician to explain this concept to their team.

   (i) Concisely explain what a p-value is. (2 marks)

   (ii) Supposing that a p-value for one of the lab's experiments has been computed as 0.06, how should this value be interpreted in relation to the experiment? (3 marks)

(b) Suppose that Imperial's Year 1 Mathematics cohort decides to form a football team and a basketball team. The teams consist of different students; no student in the cohort is in both teams. Suppose that the heights of the students in the football team follow a $\mathrm{N}(\mu_1, \sigma_1^2)$ distribution, and the heights of the students in the basketball team follow a $\mathrm{N}(\mu_2, \sigma_2^2)$ distribution. Assuming that $\sigma_1^2 = \sigma_2^2$, and carefully stating any other assumptions and results used, test the hypothesis that $\mu_1 + \mu_2 = 360$ at the significance threshold $\alpha = 0.05$, given that the heights of the football team are measured as $x_1, x_2, \ldots, x_n$ and the heights of the basketball team are measured as $y_1, y_2, \ldots, y_m$, such that

$$\overline{x} = \frac{1}{n}\sum_{i=1}^{n} x_i = 176.5, \qquad s_x^2 = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \overline{x}) = 4, \qquad n = 16,$$

$$\overline{y} = \frac{1}{m}\sum_{i=1}^{m} y_i = 181, \qquad s_y^2 = \frac{1}{m-1}\sum_{i=1}^{m}(y_i - \overline{y}) = 4, \qquad m = 9,$$

where the units for height are cm. Note that Tables 1 and 2 on pages 10 and 11 may be helpful. (6 marks)

(c) 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 $X \sim \mathrm{N}(2, 3)$ and $Y \sim \mathrm{N}(5, 12)$, and their correlation is $\mathrm{Cor}(X, Y) = \frac{1}{4}$. Defining the new random variable $Z = X + 3Y$, compute the correlation $\mathrm{Cor}(Y, Z)$. (6 marks)

(d) The table below shows the median wages for a country in the years 2012 and 2022. Median wages decreased in each of the four education subgroups, but overall median wages increased. What is the name of this phenomenon, and how could it have been caused in this example?

Level of education	**Median wages 2012**	**Median wages 2022**	**% change**
Secondary school uncompleted	£ 15194.34	£ 13963.58	8.1% decrease
Secondary school, no university	£ 20215.98	£ 19263.47	4.7% decrease
Some university	£ 24411.72	£ 22553.90	7.6% decrease
University degree	£ 34729.85	£ 34313.64	1.2% decrease
Overall	£ 22314.71	£ 22553.91	1.1% **increase**

(3 marks)

(Total: 20 marks)

Note that this question is split over two pages. Please turn the page to see the rest of Question 6.

6. (a) Suppose a tennis coach decides to use a geometric distribution with unknown parameter $\theta \in (0, 1]$ and probability mass function

$$\mathrm{P}(X = x) = (1-\theta)^{x-1}\theta, \qquad x \in \{1, 2, \dots\},$$

to model how long it takes until one of their players serves an ace during a match; an ace is a good serve where the opponent does not touch the ball during the serve (and every serve is either an ace or not an ace). Over a sequence of n matches, the coach records data $x_1, x_2, \dots, x_n$, where x_i is the number of serves until the player serves their first ace in match i. Given this data, and stating any assumptions used, find the maximum likelihood estimator for the parameter θ. (6 marks)

(b) Given n pairs of observations $(x_1, y_1), (x_2, y_2), \dots, (x_n, y_n)$ for quantities X and Y, consider the model

$$y_i = \beta_0 + \beta_1 x_i + e_i, \qquad i \in \{1, 2, \dots, n\},$$

where the e_i, $i \in \{1, 2, \dots, n\}$, are unobservable errors. Show that the estimates of $\widehat{\beta}_0$ and $\widehat{\beta}_1$ of the parameters β_0 and β_1, respectively, given by

$$\widehat{\beta}_0 = \overline{y} - \left(\frac{S_{xy}}{S_{xx}}\right)\overline{x}, \qquad \widehat{\beta}_1 = \frac{S_{xy}}{S_{xx}},$$

minimise the residual sum of squares, i.e.

$$\sum_{i=1}^{n}(y_i - \widehat{\beta}_0 - \widehat{\beta}_1 x_i)^2 = \min_{b_0, b_1} \sum_{i=1}^{n}(y_i - b_0 - b_1 x_i)^2,$$

where $\overline{x}$, $\overline{y}$, S_{xx}, S_{yy} and S_{xy} are defined by

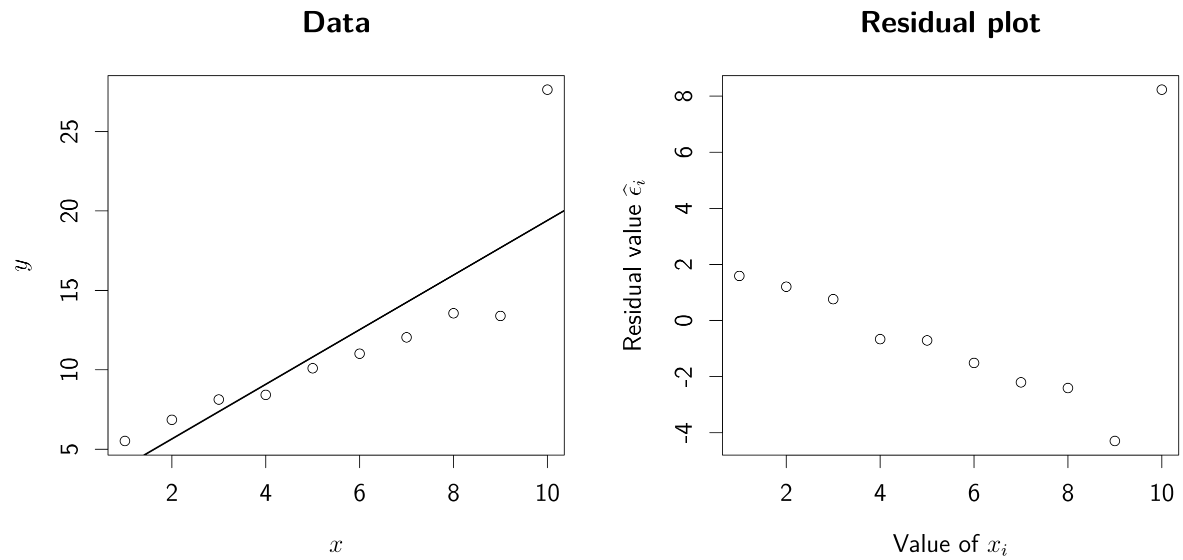
$$\begin{aligned}
\overline{x} &= \frac{1}{n}\sum_{i=1}^{n} x_i, & S_{xx} &= \sum_{i=1}^{n}(x_i - \overline{x})^2, \\
\overline{y} &= \frac{1}{n}\sum_{i=1}^{n} y_i, & S_{yy} &= \sum_{i=1}^{n}(y_i - \overline{y})^2, \\
& & S_{xy} &= \sum_{i=1}^{n}(x_i - \overline{x})(y_i - \overline{y}).
\end{aligned}$$

(6 marks)

[IMPORTANT: Question 6 continues on the next page.]

[**Question 6 continues on this page**]

(c) Consider the plots below which show data $(x_1, y_1), \ldots, (x_n, y_n)$ with a fitted regression line (on the left) and a plot of the fitted residuals $(x_1, \widehat{\epsilon}_1), \ldots, (x_n, \widehat{\epsilon}_n)$ (on the right):



Given these plots, briefly assert and justify whether or not the model fits the data well. If you believe the model does not fit the data well, then suggest what could be done to possibly improve the fit. (3 marks)

(d) In the following, state which type of plot would be the best for the given situation.

(i) Which plot could be used to visualise the distribution of the $n = 573$ observations below

$$\{32.38,\ 15.57,\ 47.01,\ 12.319,\ \ldots,\ 36.233,\ 54.07,\ 20.27\},$$

which are the average daily commute times, in minutes, for employees of a medium-sized company? (1 mark)

(ii) Which plot could be used to check if the data in (i) follow a normal distribution? (1 mark)

(iii) Which plot could be used to visualise the proportions of the categories of the data

$$\{\text{Italy},\ \text{Brazil},\ \text{England},\ \text{Ghana},\ \ldots,\ \text{Brazil},\ \text{Japan},\ \text{England}\},$$

that records the countries from which the footballers of a particular professional club originate? There are $c = 8$ countries and $n = 32$ observations. (1 mark)

(iv) Which plot could be used to visualise the following $n = 657$ observations

$$\{49.57,\ 49.70,\ 49.97,\ 48.30, \ldots,\ 207.63,\ 196.88,\ 202.07\},$$

which are the daily closing prices of a particular stock over three years? (1 mark)

(v) Which plot could be used to visualise the median, lower and upper quartiles of the data in (iv)? (1 mark)

(Total: 20 marks)

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

Table 1: $\mathrm{P}(Z < z)$ where $Z \sim \mathrm{N}(0,1)$ for values of z between 1.00 and 3.99. One reads the table as $\mathrm{P}(Z < z) = p_{r,c}$, 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, $\mathrm{P}(Z < 1.02) = 0.8461$, 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 $\mathrm{P}(T < t)$, where T has Student's t-distribution with ν degrees of freedom, for $\nu \in \{1, 2, \ldots, 30\}$. In this table, $\mathrm{P}(T < t_{\nu,p}) = p$, where ν is the νth 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

1. (a) We shall assume that each pupil's birthday is equally likely to be in any of the 52 weeks of the year. $\boxed{\text{meth seen} \Downarrow}$ $\boxed{1, \text{A}}$

Due to our assumption, the **naive probability** definition is applicable here.

First, we count how many possible ways there are to assign birthday weeks to the 20 pupils in the class. This problem can be viewed as **sampling with replacement**. Hence, the corresponding sample space is given by $\Omega = \{1, 2, \ldots, 52\}^{20}$, and we have $\text{card}(\Omega) = 52^{20}$ possible birthday week combinations (by the **multiplication principle**).

Define the event $A :=$"at least two pupils have their birthdays in the same week". Then $A^c =$"no two people share the same birthday week".

We compute $\text{card}(A^c)$, i.e. the number of scenarios such that no two people share the same birthday week. This number can be computed using **sampling without replacement**, which leads to $\text{card}(A^c) = (52)_{20} = 52 \cdot 51 \cdots 33$ possible outcomes. Then

$$\mathrm{P}(A^c) = \frac{\text{card}(A^c)}{\text{card}(\Omega)} = \frac{(52)_{20}}{52^{20}} \qquad [\approx 0.015],$$

$$\mathrm{P}(A) = 1 - \mathrm{P}(A^c) = 1 - \frac{(52)_{20}}{52^{20}} \qquad [\approx 1 - 0.015 = 0.985].$$

[1 mark for the assumption, 2 marks for the justifications and correct notation, 2 marks for the computations] $\boxed{4, \text{A}}$

(b) We shall assume that each student's birthday is equally likely to be on any of the 365 days of the year. $\boxed{\text{meth seen} \Downarrow}$ $\boxed{1, \text{A}}$

Due to our assumption, the **naive probability** definition is applicable here.

First, we count how many possible ways there are to assign birthdays to the n MSc students. This problem can be viewed as **sampling with replacement**. Hence, the corresponding sample space is given by $\Omega = \{1, 2, \ldots, 365\}^n$, and we have $\text{card}(\Omega) = 365^n$ possible birthday combinations (by the **multiplication principle**).

Define the event $A :=$"exactly two students share the same birthday". Then (recall that $3 \leq n \leq 300$)

$$\begin{aligned}\text{card}(A) &= \binom{n}{2} \cdot 365 \cdot 1 \cdot 364 \cdots (364 - n + 3) = \binom{n}{2} 365 \cdot 364 \cdots (367 - n) \\ &= \binom{n}{2} \frac{365!}{(366 - n)!},\end{aligned}$$

since there are $\binom{n}{2}$ possibilities of picking two out of n students to share the birthday. For the first student, there are 365 possible birthdays, for the one who shares the birthday, there is only 1 choice, and for the next student, there are 364 possible (different) birthdays etc.

Altogether, we get

$$\mathrm{P}(A) = \frac{\text{card}(A)}{\text{card}(\Omega)} = \frac{\binom{n}{2} \frac{365!}{(366-n)!}}{365^n}.$$

[1 mark for the assumption, 2 marks for the justifications and correct notation, 2 marks for the computations] $\boxed{4, \text{B}}$

(c) Define the following events: meth seen ⇓

$$M := \text{"a child has measles"},$$
$$F := \text{"a child has the flu"},$$
$$R := \text{"a child has a rash"}.$$

From the question, we read off the following (conditional) probabilities: $\mathrm{P}(F) = 0.95, \mathrm{P}(M) = 0.05, \mathrm{P}(R|M) = 0.9, \mathrm{P}(R|F) = 0.01$. 1, A

Using the **law of total probability (LTP)**, we get that

$$\mathrm{P}(R) = \mathrm{P}(R|M)\mathrm{P}(M) + \mathrm{P}(R|F)\mathrm{P}(F) = \frac{90}{100}\frac{5}{100} + \frac{1}{100}\frac{95}{100} = \frac{545}{10000} = 0.0545.$$

Using **Bayes' theorem**, we deduce that 2, A

$$\mathrm{P}(M|R) = \frac{\mathrm{P}(R|M)\mathrm{P}(M)}{\mathrm{P}(R)} = \frac{0.9 \cdot 0.05}{0.0545} = \frac{450}{545} = \frac{90}{109} \quad [\approx 0.826].$$

2, A

[1 mark for defining the correct notation and reading off the relevant information from the question, 1 mark for stating the LTP, 1 mark for stating Bayes' theorem, 2 marks for the calculations.]

(d) We note that $Z : \Omega \to \mathbb{R}$ and that the image of Z satisfies

$$\mathrm{Im}Z \subseteq \mathrm{Im}X \cup \mathrm{Im}Y.$$

Since X and Y are discrete random variables, their images are countable and, hence, $\mathrm{Im}X \cup \mathrm{Im}Y$ is countable, too. Hence, $\mathrm{Im}Z$ is countable. 1, B

Next, we consider the pre-images of Z and check whether they are elements of $\mathcal{F}$:
For all $z \in \mathbb{R}$, we have

$$Z^{-1}(\{z\}) = (Z^{-1}(\{z\}) \cap A) \cup (Z^{-1}(\{z\}) \cap A^c).$$

Note that $Z^{-1}(\{z\}) = \{\omega \in \Omega : Z(\omega) = z\}$. Hence

$$\begin{aligned} Z^{-1}(\{z\}) \cap A &= \{\omega \in \Omega : Z(\omega) = z\} \cap A = \{\omega \in A : Z(\omega) = z\} = \{\omega \in A : X(\omega) = z\} \\ &= \{\omega \in \Omega : X(\omega) = z\} \cap A = X^{-1}(\{z\}) \cap A \in \mathcal{F}, \end{aligned}$$

since $X^{-1}(\{z\}) \in \mathcal{F}$ since X is a discrete random variable, $A \in \mathcal{F}$ and $\mathcal{F}$ is a sigma-algebra and, hence, closed under countable intersection. Similarly,

$$\begin{aligned} Z^{-1}(\{z\}) \cap A^c &= \{\omega \in \Omega : Z(\omega) = z\} \cap A^c = \{\omega \in A^c : Z(\omega) = z\} = \{\omega \in A^c : Y(\omega) = z\} \\ &= \{\omega \in \Omega : Y(\omega) = z\} \cap A^c = Y^{-1}(\{z\}) \cap A^c \in \mathcal{F}, \end{aligned}$$

since $Y^{-1}(\{z\}) \in \mathcal{F}$ since Y is a discrete random variable, $A^c \in \mathcal{F}$ (since $A \in \mathcal{F}$ and $\mathcal{F}$ is closed under complements) and $\mathcal{F}$ is a sigma-algebra and, hence, closed under countable intersection.

Since $\mathcal{F}$ is closed under countable intersection, we conclude that, for all $z \in \mathbb{R}$, we have

$$Z^{-1}(\{z\}) = \underbrace{(Z^{-1}(\{z\}) \cap A)}_{\in \mathcal{F}} \cup \underbrace{(Z^{-1}(\{z\}) \cap A^c)}_{\in \mathcal{F}} \in \mathcal{F}.$$

Hence, Z is a discrete random variable on $(\Omega, \mathcal{F})$. 4, C

[1 mark for showing countability of ImZ, 4 marks for proving that $Z^{-1}(\{z\}) \in \mathcal{F}$ for all $z \in \mathbb{R}$.]

2. (a) (i) This follows directly from the definition of the indicator function and of the pre-image: Let $u \in U$, then meth seen ⇓

$$\mathbb{I}_A(f(u)) = \begin{cases} 1, & \text{if } f(u) \in A, \\ 0, & \text{if } f(u) \notin A, \end{cases} = \begin{cases} 1, & \text{if } u \in f^{-1}(A), \\ 0, & \text{if } u \notin f^{-1}(A), \end{cases} = \mathbb{I}_{f^{-1}(A)}(u).$$

2, B

(ii) We note that $f^{-1}(\mathcal{V}) = \{f^{-1}(A) : A \in \mathcal{V}\} = \{\{u \in U : f(u) \in A\} : A \in \mathcal{V}\}$ is a collection of subsets of U by definition. It satisfies the three axioms of a sigma-algebra on U: unseen ⇓

1, B

1. $\emptyset \in f^{-1}(\mathcal{V})$ since $\emptyset \in \mathcal{V}$ (since $\mathcal{V}$ is a sigma-algebra) and, hence, $f^{-1}(\emptyset) = \emptyset \in f^{-1}(\mathcal{V})$. 1, B
2. For any $B \in f^{-1}(\mathcal{V})$ there exists an $A \in \mathcal{V}$ such that $B = f^{-1}(A)$. Since $\mathcal{V}$ is a sigma-algebra, which is closed under complements, $A^c \in \mathcal{V}$. Then, $B^c = (f^{-1}(A))^c = f^{-1}(A^c) \in f^{-1}(\mathcal{V})$. To see that $(f^{-1}(A))^c = f^{-1}(A^c)$, we note that $x \in (f^{-1}(A))^c \Leftrightarrow x \in U \setminus f^{-1}(A) \Leftrightarrow x \in U \setminus \{u \in U : f(u) \in A\} \Leftrightarrow x \in \{u \in U : f(u) \in A^c\} \Leftrightarrow x \in f^{-1}(A^c)$. 2, C
3. For any $B_1, B_2, \ldots \in f^{-1}(\mathcal{V})$, there exist $A_1, A_2, \ldots \in \mathcal{V}$ such that $B_1 = f^{-1}(A_1), B_2 = f^{-1}(A_2), \ldots$. Since $\mathcal{V}$ is closed under countable unions since it is a sigma-algebra, we have that $\cup A_i \in \mathcal{V}$. Further, by a result from lectures, $\cup_i B_i = \cup_i f^{-1}(A_i) = f^{-1}(\cup_i A_i) \in f^{-1}(\mathcal{V})$. Hence, $f^{-1}(\mathcal{V})$ is closed under countable unions. 2, C

(b) Using the law of total probability (LTP) leads to unseen ⇓

$$\begin{aligned} p &:= \mathrm{P}(\min\{X_1, X_2, \ldots, X_{N+1}\} > x) \\ &= \sum_{n=0}^{\infty} \mathrm{P}(\min\{X_1, X_2, \ldots, X_{N+1}\} > x | N = n)\mathrm{P}(N = n) \\ &= \sum_{n=0}^{\infty} \mathrm{P}(\min\{X_1, X_2, \ldots, X_{n+1}\} > x | N = n)\mathrm{P}(N = n) \\ &= \sum_{n=0}^{\infty} \mathrm{P}(\min\{X_1, X_2, \ldots, X_{n+1}\} > x)\mathrm{P}(N = n), \end{aligned}$$

where we used the fact that $X_1, X_2, \ldots$ and N are independent. 2, D

Note that

$$\begin{aligned} \mathrm{P}(\min\{X_1, X_2, \ldots, X_{n+1}\} > x) = \mathrm{P}(X_1 > x, \ldots, X_{n+1} > x) &= (\mathrm{P}(X_1 > x))^{n+1} \\ = (1 - F_{X_1}(x))^{n+1} &= (1 - x)^{n+1}, \end{aligned}$$

where we used that $X_1, X_2, \ldots$ i.i.d.$\sim$U$(0, 1)$. 2, D

Plugging in the p.m.f. of $N \sim \mathrm{Poi}(\lambda)$ leads to

$$\begin{aligned} p &= \sum_{n=0}^{\infty} (1 - x)^{n+1} \frac{\lambda^n}{n!} e^{-\lambda} = (1 - x) \sum_{n=0}^{\infty} \frac{((1 - x)\lambda)^n}{n!} e^{-\lambda} \\ &= (1 - x) \exp((1 - x)\lambda - \lambda) = (1 - x) \exp(-\lambda x). \end{aligned}$$

2, D

(c) Recall that we denote by ϕ and Φ the p.d.f. and the c.d.f., respectively, of the standard normal distribution. $\boxed{\text{unseen} \Downarrow}$

Since X, Y i.i.d. $\sim \mathrm{N}(0,1)$, the joint density of (X,Y) is given by

$$f_{X,Y}(x,y) = f_X(x) f_Y(y) = \phi(x)\phi(y), \qquad x, y \in \mathbb{R}.$$

Define the set $A := \{(x,y) \in \mathbb{R}^2 : x > 0, y > -x\}$. According to lectures,

$$p := \mathrm{P}(X > 0, Y > -X) = \int\int_A f_{X,Y}(x,y)dxdy = \int_0^\infty \int_{-x}^\infty \phi(x)\phi(y)dydx$$
$$= \int_0^\infty \phi(x) \left(\int_{-x}^\infty \phi(y)dy \right) dx.$$

Note that, using properties of the standard normal distribution, $\boxed{3, \mathrm{D}}$

$$\int_{-x}^\infty \phi(y)dy = 1 - \Phi(-x) = \Phi(x).$$

Using the chain rule for integration, we get

$$p = \int_0^\infty \phi(x)\Phi(x)dx = \frac{1}{2}(\Phi(x))^2 \Big|_0^\infty = \frac{1}{2}(1^2 - (\Phi(0))^2) = \frac{1}{2}\left(1 - \frac{1}{4}\right) = \frac{3}{8}.$$

$\boxed{3, \mathrm{D}}$

3. (a) In general, $X+Y$ does not follow a Poisson distribution (unless X and Y are assumed to be independent). unseen ⇓

Consider the case when $X=Y$. Then $X,Y \sim \mathrm{Poi}(\lambda)$, where $\lambda = \mu$. Then $X+Y=2X$, which only takes even values, so it cannot follow a Poisson distribution which is supported on $\mathbb{N} \cup \{0\}$. 1, C

[1 mark for a valid counterexample, 2 marks for the justification; if a student correctly proves that $X+Y$ is Poisson in the independence case, then a partial mark should be awarded (1 out of 3).] 2, A

meth seen ⇓

(b) (i) We note that

$$\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f_{X,Y}(x,y)dxdy = c\int_0^1\int_{\sqrt{x}}^1 (x+y)dydx = c\int_0^1 xy + \frac{1}{2}y^2\Big|_{y=\sqrt{x}}^1 dx$$

$$= c\int_0^1 \left(x + \frac{1}{2} - x^{3/2} - \frac{1}{2}x\right)dx = c\int_0^1\left(\frac{1}{2}x + \frac{1}{2} - x^{3/2}\right)dx$$

$$= c\left(\frac{1}{4}x^2 + \frac{1}{2}x - \frac{2}{5}x^{5/2}\Big|_{y=0}^1\right) = c\left(\frac{1}{4}+\frac{1}{2}-\frac{2}{5}\right) = c\frac{7}{20} = 1 \iff c = \frac{20}{7}.$$

Also, $f_{X,Y}$ is nonnegative for this choice of c. 1, A

(ii) For $x \in (0,1)$,

$$f_X(x) = \int_{-\infty}^{\infty} f_{X,Y}(x,y)dy = c\left(xy + \frac{1}{2}y^2\Big|_{y=\sqrt{x}}^1\right) = \frac{20}{7}\left(\frac{1}{2}x + \frac{1}{2} - x^{3/2}\right),$$

and $f_X(x) = 0$ otherwise. 1, A

For $y \in (0,1)$

$$f_Y(y) = \int_{-\infty}^{\infty} f_{X,Y}(x,y)dx = c\int_0^{y^2}(x+y)dx = c\left(\frac{1}{2}x^2 + xy\Big|_{x=0}^{y^2}\right) = \frac{20}{7}\left(\frac{1}{2}y^4 + y^3\right),$$

and $f_Y(y) = 0$ otherwise. 1, A

(iii) X and Y are not independent since the joint density does not factorise into the product of the marginal densities. E.g. for $x = 1/4, y = 1/8$, we have $f_{X,Y}(1/4,1/8) = 0 \neq f_X(1/4)f_Y(1/8) = c^2\frac{1}{2}((1/8)^4/2 + (1/8)^3)$. 1, A

(iv) We note that the conditional density and conditional distribution function of $Y|X=x$ is well-defined for all x such that $f_X(x) > 0$. I.e. for $x \in (0,1)$, it is given by 1, A

$$f_{Y|X=x}(y|x) = \frac{f_{X,Y}(x,y)}{f_X(x)} = \frac{c(x+y)}{c\left(\frac{1}{2}x + \frac{1}{2} - x^{3/2}\right)} = \frac{x+y}{\frac{1}{2}x + \frac{1}{2} - x^{3/2}},$$

for $\sqrt{x} < y < 1$ and 0 otherwise. 1, B

Let $0 < x < 1$. Then, for $y \leq \sqrt{x}$, $\mathrm{P}(Y \leq y|X=x) = 0$; for $y \geq 1$, $\mathrm{P}(Y \leq y|X=x) = 1$; for $0 < \sqrt{x} < y < 1$, 1, A

$$\mathrm{P}(Y \leq y|X=x) = \int_{-\infty}^{y} f_{Y|X=x}(v|x)dv = \int_{\sqrt{x}}^{y}\frac{x+v}{\frac{1}{2}x + \frac{1}{2} - x^{3/2}}dv$$

$$= \frac{1}{\frac{1}{2}x + \frac{1}{2} - x^{3/2}}\left(xv + \frac{1}{2}v^2\right)\Big|_{v=\sqrt{x}}^{y}$$

$$= \frac{1}{\frac{1}{2}x + \frac{1}{2} - x^{3/2}}\left(xy + \frac{1}{2}y^2 - x^{3/2} - \frac{1}{2}x\right).$$

1, B

(c) (i) Let $Y := X^4$. Since $X \sim \mathrm{U}(1, 3)$, we get that

meth seen $\Downarrow$

$$F_Y(y) = \mathrm{P}(Y \le y) = 0, \qquad y < 1,$$
$$F_Y(y) = \mathrm{P}(Y \le y) = \mathrm{P}(X^4 \le y) = \mathrm{P}(X \le y^{1/4}) = \frac{1}{2}(y^{1/4} - 1), \quad 1 \le y < 3^4 = 81,$$
$$F_Y(y) = \mathrm{P}(Y \le y) = 1, \qquad y \ge 3^4 = 81.$$

Differentiating leads to

$$f_Y(y) = 0, \qquad y < 1,$$
$$f_Y(y) = \frac{1}{8}y^{-3/4}, \qquad 1 \le y < 81,$$
$$F_Y(y) = \mathrm{P}(Y \le y) = 1, \qquad y \ge 81.$$

[1 mark for the cdf, 1 mark for the pdf, 1 mark for stating all the trivial cases correctly ($y < 1, y \ge 81$)]

3, A

(ii) Using the law of the unconscious statistician (LOTUS), we get that

$$\mathrm{E}(Y) = \int_{-\infty}^{\infty} x^4 f_X(x)dx = \int_1^3 x^4 \frac{1}{2}dx = \frac{1}{10}x^5\Big|_1^3 = \frac{1}{10}(3^5 - 1)[= 24.2].$$

Alternative solution using Part (i):

$$\mathrm{E}(Y) = \int_{-\infty}^{\infty} y f_Y(y)dy = \int_1^{3^4} y\frac{1}{8}y^{-3/4}dy = \int_1^{3^4} \frac{1}{8}y^{1/4}dy = \frac{1}{10}y^{5/4}\Big|_1^{3^4} = \frac{1}{10}(3^5 - 1)[= 24.2].$$

2, A

(d) Since X and Y are independent, by a result from lectures, we know that $M_{X+Y}(t) = M_X(t)M_Y(t)$, where M_X denotes the moment generating function associated with random variable X. Since $X \sim \mathrm{Bin}(n, p)$, we have, for $t \in \mathbb{R}$, using LOTUS,

sim. seen $\Downarrow$

$$M_X(t) = \sum_{x=0}^{n} e^{xt}\binom{n}{x}p^x(1-p)^{n-x} = \sum_{x=0}^{n}\binom{n}{x}(e^t p)^x(1-p)^{n-x} = (e^t p + 1 - p)^n,$$

where we used the Binomial theorem in the last step to evaluate the sum. Similarly, for $Y \sim \mathrm{Bin}(m, p)$, we have, for $t \in \mathbb{R}$,

$$M_Y(t) = (e^t p + 1 - p)^m.$$

Hence, for $t \in \mathbb{R}$,

$$M_{X+Y}(t) = M_X(t)M_Y(t) = (e^t p + 1 - p)^{n+m}.$$

We observe that this is the m.g.f. of a Binomial random variable with parameters $(n + m, p)$. Since the m.g.f. characterises the distribution of a random variable uniquely, we deduce that $X + Y \sim \mathrm{Bin}(n + m, p)$.

3, B

1, B

4. (a) For any random variable X, an exercise in the notes gives $\boxed{\text{seen} \Downarrow}$

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

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

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

since $\mathrm{Var}[\widehat{\Theta} - \theta] = \mathrm{Var}[\widehat{\Theta}]$, by a property of the variance, and since the bias is defined as $b_\theta(\widehat{\Theta}) = \mathrm{E}[\widehat{\Theta}] - \theta = \mathrm{E}[\widehat{\Theta} - \theta]$, by the linearity of expectation. $\boxed{3, \mathrm{A}}$

$\boxed{\text{meth seen} \Downarrow}$

(b) Label the sample of n voters such that X_i is the random variable with value $x_i = 1$ if voter i supports option B and $x_i = 0$ othewise, for $i = 1, 2, \ldots, n$.

By this construction, each $X_i \sim \mathrm{Bern}(p)$, where p is the unknown parameter (proportion) to be estimated, and therefore each X_i has mean $\mathrm{E}[X_i] = p$ and variance $\mathrm{Var}[X_i] = p(1-p)$. Moreover, by this construction,

$$\widehat{p} = \frac{1}{n}\sum_{i=1}^{n} x_i = \overline{x}.$$

$\boxed{2, \mathrm{B}}$

Using Proposition 1.2.6 in the notes, $\mathrm{E}[\overline{X}] = p$ and $\mathrm{Var}[X_i] = p(1-p)/n$. For any $\epsilon > 0$, Chebyshev's Inequality gives

$$\mathrm{P}\left(\left|\overline{X} - p\right| \geq \epsilon\right) \leq \frac{\mathrm{Var}(\overline{X})}{\epsilon^2} = \frac{p(1-p)}{n\epsilon^2}.$$

Using a result from the notes regarding the variance of bounded random variables, $p(1-p) \leq 1/4$, and therefore

$$\begin{aligned}&\mathrm{P}\left(\left|\overline{X} - p\right| \geq \epsilon\right) \leq \frac{1}{4n\epsilon^2} \\ \Rightarrow &\mathrm{P}\left(\left|\overline{X} - p\right| < \epsilon\right) \geq 1 - \frac{1}{4n\epsilon^2}.\end{aligned}$$

$\boxed{2, \mathrm{B}}$

Setting $\epsilon = 0.01$ and $1 - 1/(4n\epsilon^2) = 0.99$ and solving for n,

$$\begin{aligned}&1 - \frac{1}{4n(0.01)^2} \geq 0.99 \\ \Rightarrow &\frac{1}{4n(0.01)^2} \leq 0.01 \\ \Rightarrow &4n(0.01)^2 \geq 100 \\ \Rightarrow &n \geq 100(100)^2/4 = 250000 = 2.5 \times 10^5.\end{aligned}$$

$\boxed{1, \mathrm{A}}$

(c) (i) From Proposition 1.2.6, $\overline{X} \sim \mathrm{N}(\theta, \sigma^2/n)$. If we define $\boxed{\text{sim. seen} \Downarrow}$

$$Z = \frac{\theta - \overline{X}}{\sigma/\sqrt{n}},$$

then $Z \sim \mathrm{N}(0,1)$. For any significance level α, if we define z_α to be the value such that $\mathrm{P}(Z < z_\alpha) = \alpha$, then

$$\begin{aligned}
&\mathrm{P}(Z < z_{1-\alpha/2}) = 1 - \alpha/2 \\
&\mathrm{P}(Z < z_{\alpha/2}) = \alpha/2, \\
\Rightarrow &\mathrm{P}(z_{\alpha/2} < Z < z_{1-\alpha/2}) = 1 - \alpha \\
\Rightarrow &\mathrm{P}(z_{\alpha/2} < \frac{\theta - \overline{X}}{\sigma/\sqrt{n}} < z_{1-\alpha/2}) = 1 - \alpha \\
\Rightarrow &\mathrm{P}(\overline{X} + z_{\alpha/2} \cdot \frac{\sigma}{\sqrt{n}} < \theta < \overline{X} + z_{1-\alpha/2} \cdot \frac{\sigma}{\sqrt{n}}) = 1 - \alpha.
\end{aligned}$$

To construct a 97% confidence interval, $\boxed{2, \mathrm{A}}$

$$\begin{aligned}
&1 - \alpha = 0.97 \\
\Rightarrow &\alpha = 0.03 \\
\Rightarrow &\alpha/2 = 0.015 \\
\Rightarrow &1 - \alpha/2 = 0.985
\end{aligned}$$

Using Table 1, we find $z_{0.985} = 2.17$, and therefore by symmetry of the normal distribution, $z_{0.015} = -2.17$. $\boxed{1, \mathrm{C}}$
Since $\mathbf{X}$ is observed as $\mathbf{x}$ and $\overline{x} = 10.2$, a 97% confidence interval is therefore

$$\begin{aligned}
&(\overline{x} + z_{\alpha/2} \cdot \frac{\sigma}{\sqrt{n}}, \overline{x} + z_{1-\alpha/2} \cdot \frac{\sigma}{\sqrt{n}}) \\
=&(10.2 - 2.17 \cdot \frac{\sqrt{5}}{\sqrt{20}}, 10.2 + 2.17 \cdot \frac{\sqrt{5}}{\sqrt{20}}) \\
=&(10.2 - \frac{2.17}{2}, 10.2 + \frac{2.17}{2}).
\end{aligned}$$

$\boxed{1, \mathrm{A}}$

If desired, this can be simplified to

$$(10.2 - 1.085, 10.2 + 1.085) = (9.115, 11.285).$$

$\boxed{\text{seen} \Downarrow}$

(ii) From the frequentist view of statistics, the parameter θ has a true but unknown value. Moreover, the confidence interval $(5.6, 14.8)$ does not necessarily contain θ with any given probability (e.g. 99%). $\boxed{1, \mathrm{C}}$
However, if the assumptions under which the confidence interval was constructed are true, and if the data collection could be repeated an infinite number of times to repeatedly constuct new confidence intervals for θ in the same way, then 99% of the constructed confidence intervals would contain the true value of θ. $\boxed{2, \mathrm{C}}$

(d) (i) The first step is to sort the data into nondecreasing order $\boxed{\text{sim. seen} \Downarrow}$

$$\{6.1,\ 7.2,\ 8.3,\ 9.4,\ 10.5\ 11.6,\ 12.7,\ 14.3,\ 15.9,\ 24.0\}.$$

Since there are 10 values, the median would be any value between the 5th and 6th values. Then the lower quartile $q_{0.25}$ would be the middle of the lower 5 values, i.e. the third value, which is 8.3.
Alternatively, one can compute the index for the lower quartile as follows: the index for the median is $i_m = \frac{n+1}{2} = 5.5$, and the index for $q_{0.25}$ is $i_{0.25} = \frac{1}{2}(\lfloor i_m \rfloor + 1) = \frac{1}{2}(5+1) = 3$, and again, the third value is 8.3.
Similarly, the upper quartile has index $i_{0.75} = 10 - i_{0.25} + 1 = 10 - 3 + 1 = 8$, so $q_{0.75} = 14.3$. $\boxed{2, \text{A}}$

$\boxed{\text{seen} \Downarrow}$

(ii) Tukey's criterion for outliers is that if

$$x \notin (q_{0.25} - 1.5 \cdot \text{IQR},\ q_{0.75} + 1.5 \cdot \text{IQR}),$$

then the observation x is an outlier, where $q_{0.25}$ and $q_{0.75}$ are the lower and upper quartiles, respectively, and $\text{IQR} = q_{0.75} - q_{0.25}$ is the interquartile range. $\boxed{1, \text{B}}$

For the given data, $\text{IQR} = 14.3 - 8.3 = 6$, so

$$q_{0.25} - 1.5 \cdot \text{IQR} = 8.3 - 1.5 \cdot 6 = 8.3 - 9 = -0.7,$$
$$q_{0.75} + 1.5 \cdot \text{IQR} = 14.3 + 1.5 \cdot 6 = 14.3 + 9 = 23.3.$$

There are no values in the dataset less that -0.7, and 24 is the only value larger than 23.3, so this is the only outlier in the dataset. $\boxed{2, \text{A}}$

5. (a) (i) There are a couple of ways to concisely explain what a p-value is; here are two examples: seen ⇓

- A p-value is the probability of observing a test statistic at least as extreme as the computed test statistic, given the data we have observed, under the assumption that the null hypothesis is true.
- A p-value is a transformation of a test statistic into a score in the interval $(0, 1)$. Under the null hypothesis, the p-value will follow a $\mathrm{U}(0, 1)$ distribution, and values less than a small threshold α are considered to be less likely to occur.

2, B

sim. seen ⇓

(ii) A p-value of 0.06 is relatively low, which could indicate the assumptions made by the null hypothesis are not true.
However, deciding to reject the null hypothesis (or not) will depend on the signficance level α chosen, usually before the analysis of the data has started. A common value that is chosen is 0.05, in which case, since $0.06 \not< 0.05$, the null hypothesis would not be rejected. 2, C
However, 0.05 is not the only possible value for α. If a larger significance threshold had been chosen, such as $\alpha = 0.1$, then since $0.06 < 0.1$, and then null hypothesis would be rejected. 1, C

unseen ⇓

(b) The null hypothesis H_0 and the alternative hypothesis H_1 for this experiment are

$$H_0 : \mu_1 + \mu_2 = 360,$$
$$H_1 : \mu_1 + \mu_2 \neq 360.$$

We shall assume that the data $x_1, x_2, \ldots, x_n$ and $y_1, y_2, \ldots, y_m$ are observations of the random variables $X_1, \ldots, X_n \sim \mathrm{N}(\mu_1, \sigma^2)$ and $Y_1, \ldots, Y_m \sim \mathrm{N}(\mu_2, \sigma^2)$, which are all independent of each other (we assume $\sigma^2 = \sigma_1^2 = \sigma_2^2$). 1, B
For random variables $X_1, \ldots, X_n \sim \mathrm{N}(\mu_1, \sigma^2)$ and $Z_1, \ldots, Z_m \sim \mathrm{N}(\mu_3, \sigma^2)$, it has been shown in the notes and problem sheets that the test statistic

$$T = \frac{\overline{X} - \overline{Z} - (\mu_1 - \mu_3)}{S_p\sqrt{\frac{1}{n} + \frac{1}{m}}},$$

follows Student's t-distribution with $n+m-2$ degrees of freedom, where the pooled sample variance S_p^2 is defined as

$$S_p^2 = \frac{1}{n+m-2}[(n-1)S_X^2 + (m-1)S_Z^2],$$

where S_X^2 and S_Z^2 are the sample variances of the two samples $X_1, \ldots, X_n$ and $Z_1, \ldots, Z_m$, respectively.
If we define $Y_j = -Z_j$, for $j = 1, 2, \ldots, m$, then $\overline{Y} = -\overline{Z}$ $\mu_2 = -\mu_3$ and the sample variance $S_Y^2 = S_Z^2$, and so

$$\widetilde{T} = \frac{\overline{X} + \overline{Y} - (\mu_1 + \mu_2)}{\widetilde{S}_p\sqrt{\frac{1}{n} + \frac{1}{m}}},$$

will also follow Student's t-distribution with $n+m-2$ degrees of freedom, since

$$\widetilde{S}_p^2 = \frac{1}{n+m-2}[(n-1)S_X^2 + (m-1)S_Y^2] = S_p^2.$$

and so $T = \widetilde{T}$. $\boxed{3, \text{D}}$

For $\alpha = 0.05$, $n = 16$, $m = 9$, the critical value $t_{n+m-2,1-\alpha/2} = t_{23,0.975} = 2.069$, which we obtain from Table 2. $\boxed{1, \text{A}}$

Then the realised pooled sample variance is

$$\begin{aligned}\widetilde{s}_p^2 &= \frac{1}{23}[(15)4 + (8)4] = \frac{1}{23}[23(4)] = 4\\ \Rightarrow \widetilde{s}_p &= 2,\end{aligned}$$

and

$$\widetilde{s}_p\sqrt{\frac{1}{n} + \frac{1}{m}} = 2\sqrt{\frac{1}{16} + \frac{1}{9}} = 2\sqrt{\frac{25}{16\cdot 9}} = 2\left(\frac{5}{4\cdot 3}\right) = \frac{5}{6}.$$

Then, under the null hypothesis that $\mu_1 + \mu_2 = 360$,

$$\begin{aligned}t &= \frac{(\overline{x} + \overline{y}) - (\mu_1 + \mu_2)}{\widetilde{s}_p\sqrt{\frac{1}{n} + \frac{1}{m}}}\\ &= \frac{(176.5 + 181) - (360)}{\frac{5}{6}}\\ &= \frac{357.5 - 360}{\frac{5}{6}}\\ &= (-2.5)\left(\frac{5}{6}\right)^{-1}\\ &= \frac{-5}{2}\cdot\frac{6}{5}\\ &= -3.\end{aligned}$$

Since $|-3| > 2.069$, we reject the null hypothesis at significance level $\alpha = 0.05$. $\boxed{1, \text{A}}$

(c) We start by computing the covariance of Y and Z. By properties of the covariance function,

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

$\boxed{2, \text{D}}$

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

$$\mathrm{Cov}(X,Y) = \sqrt{\mathrm{Var}(X)}\sqrt{\mathrm{Var}(Y)}\mathrm{Cor}(X,Y) = \sqrt{3}\sqrt{12}(\tfrac{1}{4}) = 6(\tfrac{1}{4}) = \frac{3}{2}.$$

$\boxed{1, \mathrm{D}}$

Then,

$$\begin{aligned}\mathrm{Cov}(Y,Z) &= \mathrm{Cov}(X,Y) + 3\mathrm{Var}(Y)\\ &= \tfrac{3}{2} + 3(12)\\ &= \tfrac{3}{2} + \tfrac{72}{2}\\ &= \tfrac{75}{2}.\end{aligned}$$

$\boxed{1, \mathrm{B}}$

Now, by the properties of the variance (or Exercise 6.1.3 in the notes),

$$\begin{aligned}\mathrm{Var}(Z) &= \mathrm{Var}(X+3Y)\\ &= \mathrm{Var}(X) + \mathrm{Var}(3Y) + 2\mathrm{Cov}(X,3Y)\\ &= \mathrm{Var}(X) + 9\mathrm{Var}(Y) + 6\mathrm{Cov}(X,Y)\\ &= 3 + 9\cdot 12 + 6(\tfrac{3}{2})\\ &= 3 + 108 + 3(3)\\ &= 120.\end{aligned}$$

$\boxed{1, \mathrm{C}}$

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

$$\begin{aligned}\mathrm{Cor}(Y,Z) = \frac{\mathrm{Cov}(Y,Z)}{\sqrt{\mathrm{Var}(Y)}\sqrt{\mathrm{Var}(Z)}} &= \frac{\frac{75}{2}}{\sqrt{12}\sqrt{120}} = \frac{\frac{75}{2}}{\sqrt{12}\sqrt{12}\sqrt{10}}\\ &= \frac{75}{24\sqrt{10}} = \frac{25}{8\sqrt{10}} \approx 0.988.\end{aligned}$$

$\boxed{1, \mathrm{B}}$

$\boxed{\text{sim. seen} \Downarrow}$

(d) This is an example of Simpson's paradox. $\boxed{1, \mathrm{B}}$

It could have been caused in this example by a change in the subgroup sizes. For example, the number of working university graduates (subgroup with highest wages) could have increased, which would increase the overall median, even though in that subgroup, and all subgroups, the median wage decreased. $\boxed{2, \mathrm{D}}$

6. (a) We assume that the data $x_1, x_2, \ldots, x_n$ are observations of independent random variables $X_1, X_2, \ldots, X_n$. The likelihood is computed as

$\boxed{\text{meth seen} \Downarrow}$

$$
\begin{aligned}
L(\theta|\mathbf{x}) = f(\mathbf{x}|\theta) &= \prod_{i=1}^{n} f(x_i|\theta) \\
&= \prod_{i=1}^{n} \left[(1-\theta)^{x_i-1}\theta\right] \\
&= \left[(1-\theta)^{\sum_{i=1}^{n}(x_i-1)}\right]\theta^n \\
&= \theta^n(1-\theta)^{n\overline{x}-n}
\end{aligned}
$$

$\boxed{1, \text{A}}$

Noting that $\overline{x} \geq 1$ (since each $x_i \geq 1$), if all $x_i = 1$, then $\overline{x} = 1$ and then the likelihood is $\log L(\theta|\mathbf{x}) = \theta^n$, which is an increasing function and has a maximum when $\theta = 1 = \overline{x} = (\overline{x})^{-1}$. We exclude the special case $\overline{x} = 1$ and $\theta = 1$ for the moment, and consider $\theta \in (0, 1)$.

In order to find the MLE, we need to maximise this likelihood. While it is possible to maximise this likelihood directly, it is easier to maximise the log-likelihood

$$\log L(\theta|\mathbf{x}) = n\log\theta + (n\overline{x} - n)\log(1-\theta).$$

Note that the log-likelihood is well-defined for all $\theta \in (0, 1)$. We take the derivative

$$\frac{\mathrm{d}}{\mathrm{d}\theta}\log L(\theta|\mathbf{x}) = \frac{n}{\theta} - \frac{(n\overline{x} - n)}{1-\theta},$$

and solve for θ when the derivative of $\log L(\theta|\mathbf{x})$ is zero:

$$
\begin{aligned}
&\frac{\mathrm{d}}{\mathrm{d}\theta}\log L(\theta|\mathbf{x}) = 0 \\
\Rightarrow &\frac{n}{\theta} - \frac{(n\overline{x} - n)}{1-\theta} = 0 \\
\Rightarrow &\frac{n}{\theta} = \frac{(n\overline{x} - n)}{1-\theta} \\
\Rightarrow &n(1-\theta) = (n\overline{x} - n)\theta \\
\Rightarrow &n - n\theta = n\overline{x}\theta - n\theta \\
\Rightarrow &n = n\overline{x}\theta \\
\Rightarrow &\theta = (\overline{x})^{-1}.
\end{aligned}
$$

$\boxed{1, \text{B}}$

However, we need to check that this value indeed maximises the log-likelihood. We compute the second derivative

$$\begin{aligned}
\frac{\mathrm{d}^2}{\mathrm{d}\theta^2} \log L(\theta|\mathbf{x}) &= \frac{-n}{\theta^2} - (-1)\frac{(n\overline{x} - n)}{(1-\theta)^2}(-1) \qquad \text{(last } (-1) \text{ from chain rule)} \\
&= \frac{1}{\theta^2(1-\theta)^2}\left[-n(1-\theta)^2 - (n\overline{x} - n)\theta^2\right] \\
&= \frac{1}{\theta^2(1-\theta)^2}\left[-n(1-\theta)^2 - n(\overline{x} - 1)\theta^2\right] \\
&= \frac{-n}{\theta^2(1-\theta)^2}\left[(1-2\theta+\theta^2) + (\overline{x} - 1)\theta^2\right] \\
&= \frac{-n}{\theta^2(1-\theta)^2}\left[1-2\theta+\theta^2 + \overline{x}\theta^2 - \theta^2\right] \\
&= \frac{-n}{\theta^2(1-\theta)^2}\left[1-2\theta + \overline{x}\theta^2\right].
\end{aligned}$$

$\boxed{1, \mathrm{C}}$

The first term $\dfrac{-n}{\theta^2(1-\theta)^2}$ is negative. We already considered the special case above that all $x_i = 1$, which implies $\overline{x} = 1$ and then $\theta = 1 = (\overline{x})^{-1}$ maximises the likelihood. Otherwise, if at least one $x_i > 1$, then $\overline{x} > 1$ and so

$$\begin{aligned}
&\overline{x} > 1 \\
\Rightarrow &(\overline{x})^{-1} < 1 \\
\Rightarrow &-(\overline{x})^{-1} > -1 \\
\Rightarrow &-(\overline{x})^{-1} + 1 > 0.
\end{aligned}$$

Therefore, the second term evaluated at $\theta = (\overline{x})^{-1}$ is

$$1 - 2((\overline{x})^{-1}) + \overline{x}(\overline{x})^{-2} = 1 - 2((\overline{x})^{-1}) + (\overline{x})^{-1} = 1 - (\overline{x})^{-1} > 0,$$

and so overall, when $\theta = (\overline{x})^{-1}$,

$$\frac{\mathrm{d}^2}{\mathrm{d}\theta^2} \log L(\theta|\mathbf{x})\bigg|_{\theta=(\overline{x})^{-1}} = \frac{-n}{\theta^2(1-\theta)^2}\left[1-2\theta+\overline{x}\theta^2\right]\bigg|_{\theta=(\overline{x})^{-1}} < 0.$$

which shows that $\theta = (\overline{x})^{-1}$ is a local maximum. $\boxed{1, \mathrm{D}}$

To show it is a global maximum, first note that

$$L(\theta|\mathbf{x})\big|_{\theta=(\overline{x})^{-1}} = (\overline{x})^{-n}(1-(\overline{x})^{-1})^{n\overline{x}-n} > 0,$$

and then since $\theta \in (0,1]$,

$$\begin{aligned}
\lim_{\theta\to 0^+} L(\theta|\mathbf{x}) &= \lim_{\theta\to 0^+} \left[\theta^n(1-\theta)^{n\overline{x}-n}\right] = 0,\\
\lim_{\theta\to 1^-} L(\theta|\mathbf{x}) &= \lim_{\theta\to 1^-} \left[\theta^n(1-\theta)^{n\overline{x}-n}\right] = 0.
\end{aligned}$$

$\boxed{1, \text{D}}$

Alternatively, one can show that $\theta = (\overline{x})^{-1}$ is a global maximum by showing the second derivative is negative everywhere. Recall that the second derivative was computed as

$$\frac{\mathrm{d}^2}{\mathrm{d}\theta^2}\log L(\theta|\mathbf{x}) = \frac{-n}{\theta^2(1-\theta)^2}\left[1-2\theta+\overline{x}\theta^2\right].$$

Since $\overline{x} \geq 1$, one can show

$$\begin{aligned}
\overline{x}\theta^2 &\geq \theta^2\\
\Rightarrow -\overline{x}\theta^2 &\leq -\theta^2\\
\Rightarrow (-1+2\theta)-\overline{x}\theta^2 &\leq (-1+2\theta)-\theta^2\\
\Rightarrow -(1-2\theta+\overline{x}\theta^2) &\leq -(1-2\theta+\theta^2)\\
\Rightarrow \frac{-n}{\theta^2(1-\theta^2)}(1-2\theta+\overline{x}\theta^2) &\leq \frac{-n}{\theta^2(1-\theta^2)}(1-2\theta+\theta^2)\\
&= \frac{-n}{\theta^2(1-\theta^2)}(1-\theta)^2\\
&= \frac{-n}{\theta^2} \qquad\qquad (\text{since } \theta\in(0,1))\\
\Rightarrow \frac{\mathrm{d}^2}{\mathrm{d}\theta^2}\log L(\theta|\mathbf{x}) &\leq \frac{-n}{\theta^2} < 0, \qquad (\text{since } n,\theta>0),
\end{aligned}$$

which shows that the function is concave everywhere, and so the local maximum $\theta = (\overline{x})^{-1}$ is a global maximum. Note this also includes the special case $(\overline{x},\theta) = (1,1)$.

Therefore, $\widehat{\theta} = (\overline{x})^{-1}$ is a global maximum and is the maximum likelihood estimate of θ, and $\widehat{\Theta} = (\overline{X})^{-1}$ is the maximum likelihood estimator for θ. $\boxed{1, \text{A}}$

[2 mark for showing second derivative is negative when $\theta = (\overline{x})^{-1}$, and so is a local maximum, and 1 mark for showing it is a global maximum by looking at boundary, or 3 marks for alternative solution showing second derivative is negative everywhere.]

(b) There are multiple ways to minimise $\boxed{\text{meth seen} \Downarrow}$

$$f(b_0, b_1) = \sum_{i=1}^{n} (y_i - b_0 - b_1 x_i)^2,$$

(e.g. using Calculus and partial derivatives) but one approach is to let $z_i = y_i - b_1 x_i$, and then recall from an exercise in the notes that the expression

$$f(b_0, b_1) = \sum_{i=1}^{n} (z_i - b_0)^2$$

is minimised when $b_0 = \overline{z}$, since $\sum_{i=1}^{n}(z_i - b_0)^2 \geq \sum_{i=1}^{n}(z_i - \overline{z})^2$. Therefore, we set $\widehat{\beta}_0 = \overline{z} = \overline{y} - b_1\overline{x}$, and need to solve for b_1 which minimises this expression. $\boxed{2, \text{B}}$

Since $f(b_0, b_1) \leq f(\widehat{\beta}_0, b_1)$,

$$\begin{aligned}
f(\widehat{\beta}_0, b_1) &= \sum_{i=1}^{n} [y_i - \widehat{\beta}_0 - b_1 x_i]^2 \\
&= \sum_{i=1}^{n} [y_i - b_1 x_i - \widehat{\beta}_0]^2 \\
&= \sum_{i=1}^{n} [y_i - b_1 x_i - (\overline{y} - b_1 \overline{x})]^2 \\
&= \sum_{i=1}^{n} [(y_i - \overline{y}) - b_1(x_i - \overline{x})]^2 \\
&= \sum_{i=1}^{n} [(y_i - \overline{y})^2 - 2b_1(x_i - \overline{x})(y_i - \overline{y}) + b_1^2(x_i - \overline{x})^2] \\
&= \sum_{i=1}^{n} (y_i - \overline{y})^2 - 2b_1 \sum_{i=1}^{n}(x_i - \overline{x})(y_i - \overline{y}) + b_1^2 \sum_{i=1}^{n}(x_i - \overline{x})^2 \\
&= S_{yy} - 2b_1 S_{xy} + b_1^2 S_{xx}.
\end{aligned}$$

One can complete the square to obtain:

$$\begin{aligned}
f(\widehat{\beta}_0, b_1) &= S_{yy} - 2b_1 S_{xy} + b_1^2 S_{xx} \\
&= S_{xx}\left(b_1^2 - 2b_1 \frac{S_{xy}}{S_{xx}}\right) + S_{yy} \\
&= S_{xx}\left(b_1^2 - 2b_1 \frac{S_{xy}}{S_{xx}} + \left\{\frac{S_{xy}}{S_{xx}}\right\}^2\right) + S_{yy} - S_{xx}\left\{\frac{S_{xy}}{S_{xx}}\right\}^2 \\
&= S_{xx}\left(b_1 - \frac{S_{xy}}{S_{xx}}\right)^2 + \left[S_{yy} - \frac{(S_{xy})^2}{S_{xx}}\right],
\end{aligned}$$

which shows that $f(\widehat{\beta}_0, b_1)$ is minimised when $b_1 = S_{xy}/S_{xx}$. $\boxed{3, \text{A}}$

Substituting this value into $\widehat{\beta}_0 = \overline{y} - b_1\overline{x}$, we have $\widehat{\beta}_0 = \overline{y} - (S_{xy}/S_{xx})\overline{x}$, and so the values that minimise $f(b_0, b_1)$ are

$$\begin{aligned}
b_0 &= \widehat{\beta}_0 = \overline{y} - \left(\frac{S_{xy}}{S_{xx}}\right)\overline{x} \\
b_1 &= \widehat{\beta}_1 = \frac{S_{xy}}{S_{xx}},
\end{aligned}$$

as required. $\boxed{1, \text{A}}$

(c) No, the residual plot shows that the model does not fit the data well, because there is a clear trend in the residuals (a "U"-shape; the first three residuals are positive, the next six are negative and the last is positive). part seen ⇓ 1, B

It seems that the last observation (with largest x-value) has a very large y-value (and residual) relative to the other observations. It is possibly an outlier, and seems to be skewing the fitted regression line. If this observation were removed, it is possible that the fit of the model would be better. 2, D

seen ⇓

(d) (i) Histogram. 1, A

(ii) Q-Q plot. 1, A

(iii) Pie chart 1, A

(iv) Line plot. 1, A

(v) Box plot. 1, A

Review of mark distribution:

Total A marks: 48 of 48 marks

Total B marks: 30 of 30 marks

Total C marks: 18 of 18 marks

Total D marks: 24 of 24 marks

Total marks: 120 of 120 marks

If your module is taught across multiple year levels, you might have received this form for each level of the module. You are only required to fill this out once for each question.

ExamModuleCode	**QuestionNumber**	**Comments for Students**
MATH40005	1	Q1a) followed very closely an example computed in the lectures, but there were still quite a few students who could not come up with the correct result. Many students forgot to justify how they derived their results.There were a wide range of assumptions quoted to compute the problem including a wide range of possible numbers of weeks per year. As long as the computations were done correctly, full marks were awarded for working with various numbers of weeks in the year. Students needed to state that they work under assumptions which ensure that the naive probability is applicable. Just assuming a certain number of weeks in the year was not sufficient as an assumption.Q1b) Many students who managed to do a) were also able to to b). It turned out that the justifications given in this part were typically better and more detailed.Q1c) was done well by the majority of students. Occasionally, there were some minor calculation errors and a couple of students forgot to introduce their notation.Q1d) The majority of students managed to prove that the Im(Z) is countable. For the measurability condition, there were many good attempts, but very few students got this question completely right.
MATH40005	2	2a Generally well done. In part (ii), a common mistake was to show that $-1($ $\neq$ $)$rather than $-1($ $)$. f $\mathcal{V}$ $\emptyset$ $\emptyset \in f$ $\mathcal{V}$ 2b Many students failed to condition on N = n and evaluate the resulting infintie sum. 2c Many students did not attempt this question. Some students correctly idenitified the correct double integral but only a few of these correctly evaluated it. A handful of students spotted the "easy" solution: the bivariate Gaussian density is invariant to rotation, so the probability is the proportion of 2π covered by the region of interest.

MATH40005	3	General comments: The students didn't do very well overall in this exercise.Part a) was attempted by a lot of students but only very few responded correctly. A correct answer would be a good counterexample to show that the sum of two dependent poisson random variables is not neccessarily poisson distributed. The majority attempted to prove that when the two poisson random variables are independent their sum is poisson (which would give some partial marks) but the statement of independence was missing. Also, in many cases the counterexamples were wrong.Part bi) Almost all of the students attempted this questions and the vast majority responded correctly this part. ii) A good amount of students responded correctly however a lot have wrong pfd's because of using wrong limits in the integral. Also, students gave wrong domain for x and y. Bad notation in many cases where f_x was denoted by F or M_x or other things.iii) The majority of students knew that thew variables are not independent but did not show it with some calculations ie expanding f(x,y) and f(x) x f(y).iv) This question was in general not well answered and even not attempted by a lot of students. A small number of students gave the correct form of of the conditional CDF of Y given X. Many mistakes in the form of the conditional distribution, the limits in the integration and the set in which x and y belong to.c) i) Most students could find or remember or compute the form of the uniform distribution of X but a big number of those did not manage to compute the pdf/cdf of X^4. A source of the problem wa the bad notation as they kept using x for the values of X^4.ii) This part went well overall.d) This question had the extremes with students doing either very well or very bad. In many cases the stuents could identify the resulting distribution but with not enough of a proof or missing the independence statement or neglexting to give the MGF (asked by the examiner). There were some surprisingly odd distributions being suggested as the distribution of the sum of 2 independent Binomials with same rate.

MATH40005	4	Q4a) This question asks for a proof of a standard result, covered in lectures. This question was generally answered well, although in many cases students did not give full justification for certain steps (e.g. "linearity of expectation").Q4b) This question follows a similar example covered in the lecture notes, where the proportion of a vote can be modelled using a Bernoulli distribution, and Chebyshev's inequality needs to be used, along with a result on the variance of bounded random variables. This was generally answered well.Q4ci) This is a standard question for constructing a confidence interval for normal random variables, when the variance is known. This was generally answered well.Q4cii) This question asks for the 99% confidence interval to be interpreted. The correct answer would refer to repeated experiments, rather than the parameter being in the interval with a certain probability; i.e. if the experiment were repeated many times, then the true value of the parameter would be in 0.99 of the constructed confidence intervals (if the assumptions were true). This question was only answered correctly by a small number of students.Q4d) In part (i) the lower and upper quantiles need to be computed, and then in part (ii), any outliers need to be identified after stating Tukey's criterion for outliers. Generally answered well, although a number of students computed the lower/upper quantiles in an unusual manner.

MATH40005	5	Part a)i) Only very few students answered this question correctly. There was a wide variety of wrong answers. My advice is to review this concept as it will be used again in future modules. Part a)ii) Many students described the decision to make if the significant value is greater or less than the p-value, which is fine, but only a few gave a more explanatory answer about the given p-value. Part b) Only a few students got full marks on this question. The most common failure was not establishing the necessary assumptions to carry out the hypothesis test. Other common errors were: incorrect test statistic, incorrect pooled variance, plugging in the pooled variance in the test statistics instead of the pooled standard deviation, incorrect critical value, incorrect decision to make. Some students combined two one-sample hypothesis tests instead of working with a two-sample hypothesis test, which is incorrect. Part c) Several students got full marks on this question. The most common errors were: confusing covariance and correlation coefficient, not considering the term +2Cov(X,Y) when calculating the variance of a sum of dependent random variables, difficulties in calculating COV(Y, X+3Y). Part d) Most of the students coorectly stated that the phenomenon is known as "Simpson's paradox" and related it to the different sizes of the subgroups. Only a few mentioned that it is caused by the change in subgroup sizes, which is what actually explains the paradox, and very few gave an example illustrating how this can happen.
MATH40005	6	There were mixed attempts to this question:6a) Many students could write down the likelihood correctly and derive the maximum likelihood estimator. However, often there were issues in using the correct notation (distinguishing clearly between the random variables and their realisations). Occasionally the proof of a global maximum was incomplete.6b) This question was skipped by various students. Those who attempted the question typically did quite well. Sometimes the justification of why the given estimates minimise the RSS was incomplete.6c) The majority concluded that the model fit was not good and gave a good justification. For the second part, there were many good suggestions of how to improve the fit, but also some incomplete or false answers.6d) Generally answered well; some students struggled to remember the correct terminology though.