Note that there are FIVE questions split across TWO pages.

Provide justification for all solutions unless otherwise stated.

Question 1

Suppose that the random variables $X_1, X_2, \ldots, X_n$, with $n > 3$, are independent and each follows the same distribution which has mean μ and variance σ^2. We decide to define $\widehat{\Theta}$, an estimator of the mean μ, as:

$$\widehat{\Theta} = \frac{1}{n-3} \sum_{i=1}^{n} X_i.$$

Clearly stating any results or properties used:

(i) **(2 points)** Compute $b_\mu(\widehat{\Theta})$, the bias of $\widehat{\Theta}$.

(ii) **(4 points)** Compute the mean squared error of $\widehat{\Theta}$.

Question 2

Suppose the random variable X follows an unknown distribution but is known to only take values in the range $[2, 7]$.

(2 points) Find values a and b such that $a \leq \text{Var}[X] \leq b$ and the interval width, $b - a$, is minimised.

Question 3

Suppose that $X_1, X_2, \ldots, X_n$ are random variables that represent a sample of transaction values for a particular bank account. These transaction values are assumed each independently follow the same distribution with unknown mean μ and variance σ^2.

Suppose further that we observe these random variables as $x_1, x_2, \ldots, x_n$ (positive values represent money received, and negative values represent payments), for $n = 18$, and their sample mean is computed to be $\bar{x} = £4,895$. We decide to assume the standard deviation in transaction values is less than £100.

(i) **(4 points)** Given the data above, construct a 90% confidence interval for the unknown mean μ.

(ii) **(1 point)** Suppose we are now told that the data follow a Cauchy distribution. Would you still be able to follow the same approach as in (a)? Justify your answer.

QUESTION 4 AND QUESTION 5 ARE ON THE NEXT PAGE

Question 4

Consider the sample of the following 11 values:

$$\{5, 12, 7, 4, 10, 8, 11, 15, 17, 16, 18\}.$$

Showing **all working**:

(i) **(1 point)** Compute the sample median.

(ii) **(2 point)** Compute the upper and lower quartiles.

(iii) **(1 point)** Compute the interquartile range.

Note: values provided without working will receive 0 points.

Question 5

For each of the following scenarios, which plot would be the best choice for an exploratory data analysis? In this question you can provide the answer without a full explanation.

(i) **(1 point)** The data is categorical and one wishes to display the proportions of each category recorded.

(ii) **(1 point)** The data is continuous and one wishes to see if it follows a normal distribution.

(iii) **(1 point)** The data is continuous and one wishes to visualise the spread of the data and identify any outliers.

Total: 20 points