Note that there are FOUR questions split across TWO pages.

Question 1

Suppose that the n observations $x_1, x_2, \ldots, x_n$ are recorded, where $n = 25$ and the following summary statistics are computed:

- $x_{(1)} = -1$ (the smallest observation)
- $q_{0.25} = 1$ (the lower quartile)
- $m = q_{0.5} = 2$ (the median)
- $q_{0.75} = 4$ (the upper quartile)
- $x_{(n)} = 7$ (the largest observation)
- $\sum_{i=1}^{n} x_i = 60$
- $\sum_{i=1}^{n} x_i^2 = 264$

Showing **all working** and justifying **any formulae** used:

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

(ii) **(1 point)** Compute the range.

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

(iv) **(2 points)** Compute the sample variance.

Question 2

Suppose that the random variables $X_1, X_2, \ldots, X_n$ are independent and each follows the same distribution which has mean μ and variance σ^2. Recall the definitions of the sample mean and sample variance

$$\overline{X} = \frac{1}{n}\sum_{i=1}^{n} X_i, \qquad S^2 = \frac{1}{n-1}\sum_{i=1}^{n}\left(X_i - \overline{X}\right)^2,$$

where $\overline{X}$ is an estimator of μ and S^2 is an estimator of σ^2. Suppose it is known that for this distribution,

$$\operatorname{Var}\left[\sum_{i=1}^{n}\left(X_i - \overline{X}\right)^2\right] = 2(n-1)\sigma^4.$$

Stating any results used from the notes:

(i) **(1 point)** Show that $b_{\sigma^2}(S^2) = 0$, where $b_{\sigma^2}(S^2)$ is the bias of S^2.

(ii) **(2 point)** Prove that the mean squared error of S^2 is $\dfrac{2\sigma^4}{n-1}$.

(iii) **(1 point)** Suppose that one defines $W = \dfrac{1}{n+1}\sum_{i=1}^{n}\left(X_i - \overline{X}\right)^2$ as an alternative estimator of σ^2. Compute $b_{\sigma^2}(W)$, the bias of W.

(iv) **(1 point)** Compute $\operatorname{Var}(W)$.

(v) **(2 point)** Compute the mean squared error of W, and show that it is less than the mean squared error of S^2.

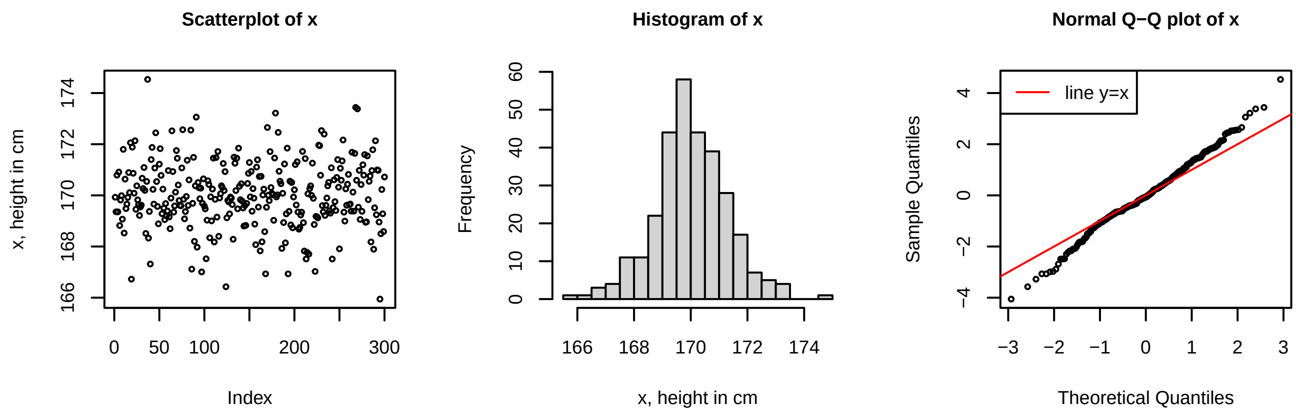
(vi) **(2 points)** Which estimator would you prefer to use to estimate σ^2? Justify your answer, stating the advantages and disadvantages of both estimators.

Question 3 and Question 4 are on the next page.

Question 3

Suppose $X_1, X_2, \ldots, X_n$ are the random variables representing the heights of the $n = 300$ students in a particular module, measured in cm. These random variables are observed as $x_1, x_2, \ldots, x_n$, which are plotted below in (a) a scatterplot of the data, $\mathbf{x} = (x_1, \ldots, x_n)$, (b) a histogram of the data, (c) a Q-Q plot of the data after being standardised by the sample mean and variance.

(2 points) Do these plots suggest that $X_1, X_2, \ldots, X_n$ follow a normal distribution? Provide justification for your answer.



Question 4

Suppose $X_1, X_2, \ldots, X_n$, where $n = 20$, are independent and identically distributed random variables representing the heights of n students measured in cm. Suppose that for $i = 1, 2, \ldots, n$, each X_i is assumed to follow a normal distribution with $\mathrm{E}(X_i) = \theta$ and $\mathrm{Var}(X_i) = \sigma^2$, where θ is unknown but σ^2 is known to be $\sigma^2 = 15$.

Now suppose that the heights of the students are measured as $x_1, x_2, \ldots, x_n$, and from these measurements it is computed that $\overline{x} = 182$cm.

(i) **(3 points)** Given the assumptions and the data above, construct a 99% confidence interval for the unknown mean θ. Table 1 below may be helpful.

(ii) **(1 point)** If the variance σ^2 were unknown, how else could you construct the confidence interval for θ?

Table 1: Selected values of z for $\mathrm{P}(Z < z)$, where Z has a standard normal distribution

z	$\mathrm{P}(Z < z)$
1.281	0.900
1.645	0.950
1.960	0.975
2.326	0.990
2.576	0.995

Total: 20 points