Question 1

Suppose $Z_1, Z_2, \ldots, Z_n$ are independent and identically distributed random variables following an unknown distribution F_Z. The mean μ of the distribution F_Z is unknown, but the variance of F_Z is known to be $\sigma^2 = 7$. Suppose we observe $\mathbf{Z} = (Z_1, Z_2, \ldots, Z_n)$ as $\mathbf{z} = (z_1, z_2, \ldots, z_n)$. Given that the sample mean is $\overline{z} = 6$ and $n = 12$, construct a 95% confidence interval for μ.

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,

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

Stating any results used from the notes:

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

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

(c) 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.

(d) Compute $\mathrm{Var}(W)$.

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

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

Question 3

Suppose you are conducting an experiment and record the following nine measurements:

$$\mathbf{x} = \{5.6,\ 3.2,\ 11.7,\ 3.2,\ 13.8,\ 8.4,\ 8.4,\ 7.5,\ 2.1\}$$

and you want to compute a measure of central tendency and dispersion for this data.

(a) Compute the sample mean and sample variance of $\mathbf{x}$ (if you like, you may use a calculator).

(b) Compute the sample median and the sample interquartile range of $\mathbf{x}$.

(c) Suppose that you did not have a computer or calculator with you, and you were asked to compute either (a) or (b) to two decimal places (or as a fraction). Which option would you choose, and why?

Question 4

Let $\mathbf{x} = \{x_1, x_2, \ldots, x_n\}$, with $n > 1$.

(a) Find a sample $\mathbf{x}$ where its sample median equals its sample mean.

(b) Find a sample $\mathbf{x}$ where its sample median is greater than its sample mean.

(c) Find a sample $\mathbf{x}$ where its sample median is smaller than its sample mean.

(d) Suppose that the sample $\mathbf{x} = \{x_1, x_2, \ldots, x_n\}$ is known to have sample mean $\overline{x} = \mu$, but the precise values of the x_i in the sample $\mathbf{x}$ are unknown. Given **any** other finite value $\mu' \neq \mu$, we will add n' elements to $\mathbf{x}$ to construct $\mathbf{x}'$, i.e. $\mathbf{x}' = \{x_1, x_2, \ldots, x_n, x_{n+1}, \ldots, x_{n+n'}\}$, so that the sample mean of $\mathbf{x}'$ is μ'. What must the smallest value of n' be in order to ensure this, and furthermore what are the values of $x_{n+1}, \ldots, x_{n+n'}$?

(e) Suppose that $\mathbf{x} = \{x_1, x_2, \ldots, x_n\}$ is known to have sample **median** m, but the values of the x_i in the sample $\mathbf{x}$ are unknown . Given any other finite value $m' \neq m$, we will add n' elements to $\mathbf{x}$ to construct $\mathbf{x}'$ so that the sample median of $\mathbf{x}'$ is m'. What must the smallest value of n' be in order to ensure this, and furthermore what are the values of $x_{n+1}, \ldots, x_{n+n'}$? Choose the values of $x_{n+1}, \ldots, x_{n+n'}$ so that the sample median of $\mathbf{x}'$ will be m', no matter the values in the original sample $\mathbf{x}$.