Question 1

Recall from Section 8.3.1 in Prof. Veraart's notes that the p.d.f. of the uniform distribution on the interval (a, b) is

$$f_X(x) = \begin{cases} \frac{1}{b-a}, & \text{if } a < x < b, \\ 0, & \text{otherwise.} \end{cases}$$

We write $X \sim \mathrm{U}(a, b)$ to indicate that the random variable X follows this distribution.

(a) If $X \sim \mathrm{U}(a, b)$, compute $\mathrm{E}(X)$.

(b) If $X \sim \mathrm{U}(a, b)$, compute $\mathrm{Var}(X)$.

Question 2

In each part below, $\mathbf{x} = \{x_1, x_2, \ldots, x_n\}$ and $\mathbf{x}' = \{x_1, x_2, \ldots, x_n, x_{n+1}, \ldots, x_{n+n'}\}$.

(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) Given $\mathbf{x}$ with sample mean μ, for any other finite value $\mu' \neq \mu$, construct $\mathbf{x}'$ so that the sample mean of $\mathbf{x}'$ is μ', and furthermore $\mathbf{x}'$ has the smallest possible number of elements.

(e) Given $\mathbf{x}$ with sample median m, for any other finite value $m' \neq m$, construct $\mathbf{x}'$ so that the sample median of $\mathbf{x}'$ is m', and furthermore $\mathbf{x}'$ has the smallest possible number of elements.

Question 3

Suppose you conduct an experiment and record the following measurements: $\mathbf{x} = \{6, 3, 10, 3, 10, 8, 8, 7, 2\}$.

(a) Compute the sample mean of $\mathbf{x}$.

(b) Compute the sample variance of $\mathbf{x}$.

(c) Compute the sample median of $\mathbf{x}$.

(d) Compute the interquartile range of $\mathbf{x}$.

(e) Which do you think, in general, is more computationally intensive for a large number of observations: computing the sample mean or computing the sample median?

(f) Which do you think, in general, is more computationally intensive for a small number of observations when doing the calculation by hand: computing the variance or computing the interquartile range?

Question 4

Suppose that a population is taking part in a vote and an unknown proportion p of the voters supports a particular option, labelled A. 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 A. What value of n should be chosen so that with high confidence (probability at least 95%) $\widehat{p}$ is within 0.01 of p?