Question 1

The goal of this question is to prove Theorem 1.6.13: for a random variable X with distribution F_X and median m, then for any real value a,

$$\min_a \mathrm{E}\left(|X-a|\right) = \mathrm{E}\left(|X-m|\right).$$

While it is possible to prove certain special cases (e.g. assuming X is continuous and has a p.d.f.) relatively easily, this question breaks down a proof for the general case.

Consider the function $G(X) = |X-a| - |X-m|$. Then, assuming $a \geq m$,

(a) Show that $G(X) = a-m$ when $X \leq m$, and $G(X) \geq -(a-m)$ when $X > m$.

(b) Using the indicator function $\mathbb{I}(X \leq m)$, show that $\mathrm{E}(G(X)\,\mathbb{I}(X \leq m)) = (a-m)\mathrm{P}(X \leq m)$.

(c) Using the indicator function $\mathbb{I}(X > m)$, show that $\mathrm{E}(G(X)\,\mathbb{I}(X > m)) \geq -(a-m)\mathrm{P}(X > m)$

(d) Using Parts (b) and (c), show that $\mathrm{E}(G(X)) \geq (a-m)\left[\mathrm{P}(X \leq m) - \mathrm{P}(X > m)\right]$.

(e) Recalling from the definition of the median that $\mathrm{P}(X \leq m) \geq \frac{1}{2}$, show that $\mathrm{E}(G(X)) \geq 0$.

(f) Conclude that the theorem is true if we can assume $a \geq m$.

(g) Show the theorem is true, no matter the value of a.

Question 2 (R question)

In Question 3 you will create an R Markdown document. This question helps you to ensure that R Markdown is installed.

Open R Studio and then click on the '+' icon in the top-left corner which gives you options to create a new file. Select the option 'R Markdown'. RStudio may then display a message saying you need certain packages to be installed in order to create an R Markdown document, and ask if you want to install these packages; say yes. After a while, the packages will be installed and you can move on to the next question.

Alternative The above is perhaps the easiest way to get the packages installed. There is also a more 'manual' method: open R Studio, and then in the **in the R console** run the command

```
library(rmarkdown)
```

If the command runs without throwing an error, then great - R Markdown is installed! If there is an error, then **in the R console** run

```
install.packages("rmarkdown")
```

which will install the `rmarkdown` package and a few other packages that are necessary. Note that there is a way to ensure that R packages are installed to a specific directory by specifying the `lib` path in the `install.packages` command, but then you will need to make sure that this path is added to `.libPaths()` whenever you start R.

Question 3 (R question)

This question is to ensure you can generate an R Markdown report using R Studio.

(a) Open RStudio and then open the file `example_rmarkdown_to_pdf.Rmd` in RStudio. The file is available to download on Blackboard.

(b) Click the 'Knit' button to generate a PDF version of this file. If you cannot see a 'Knit' icon, look for a 'Preview' icon. Or click on the drop-down arrow next to 'Knit' (or 'Preview') and select 'Knit to PDF'.

(c) A PDF should open showing a scatterplot. There are two possible things that can go wrong in this step, though.

- If the PDF opens, but the scatterplot does not appear, in RStudio click on the drop-down arrow on 'Run' command (top-right corner of editor window) and select 'Run All' and 're-knit' (which is like re-compiling).
- If the PDF does not open at all, and you see an error in the console, then please see the 'Troubleshooting' section below.

(d) Modify the 'subtitle' field at the top of your Rmd file to have your full name and CID number.

(e) Edit the Rmd file to fix the expression for the normal probability density function. This is an example of LaTeX code for creating equations; although you may not have seen it before, try to match the code with the output and change the numbers accordingly. After editing, click 'Knit' again to regenerate the PDF

(f) Edit the Rmd file to plot a histogram instead of a scatterplot.

(g) Increase the number of observations generated from 100 to 10000. Re-knit the document. If the changes do not take effect, re-knit the document a second time.

(h) For the histogram, set the parameter `freq` to `FALSE`, i.e. `hist(z, freq=FALSE)`, and replot the histogram (re-knit the document).

(i) Plot the probability density function for the appropriate normal distribution over the histogram. You will need to use the `lines` command, and set the parameters for the normal distribution's mean and standard deviation inside the `dnorm` command correctly.

(j) Save the PDF of the R Markdown document to your computer.

Troubleshooting

If when you first tried to knit the R markdown file and a PDF did not appear, then it is likely your computer is missing a package allowing LaTeX to be compiled. An easy way to fix this is to use R to install a lightweight version of LaTeX. **In the R console** run the following command:

```
tinytex::install_tinytex()
```

For more information, see: https://bookdown.org/yihui/rmarkdown-cookbook/install-latex.html

Alternative If the above doesn't work for you, don't worry. An alternative is to go to https://rstudio.cloud/ and create a free account which allows you to use RStudio in a browser. This version of RStudio already has everything you need in order to create R Markdown documents, and the created PDF will open in a separate window/tab. The only downside is that the free account only allows users 15 hours of server time per month, but this should be enough for our purposes. If you use up the 15 hours, it will probably ask you to upgrade your account (but this is not necessary!)