Assignment 4

Due Tuesday Feb 11 at 11:59pm on Quercus

As before, the questions without solutions are an assignment: you need to do these questions yourself and hand them in (instructions below). The assignment is due on the date shown above. An assignment handed in after the deadline is late, and may or may not be accepted (see course outline). My solutions to the assignment questions will be available when everyone has handed in their assignment.

You are reminded that work handed in with your name on it must be *entirely your own work*.

Assignments are to be handed in on Quercus. See `https://www.utsc.utoronto.ca/~butler/c32/quercus1.nb.html` for instructions on handing in assignments in Quercus. Markers' comments and grades will be available there as well.

As ever, begin with this:

```
library(tidyverse)
```

1. Work through problems 9.1 and 9.2 in PASIAS. If you like, also work through problem 9.3. (This last problem is not immediately relevant to this assignment, but you have all the background to make sense of it, and doing so will give you some context for what is going on.)

   Hand the next one in.

2. The spreadsheet at `https://github.com/nxskok/datasets/blob/master/goodman%20-%20modern%20statistics/Canadian_Equity_Funds.xls` contains a random sample of equity funds published in 2005. We are interested in the asset value of each fund, shown in column C of the spreadsheet. These are in millions of dollars.

   (a) (3 marks) Read in and display (some of) this spreadsheet. To do this, you will need to follow a number of steps:

   - Go to the URL given above and click Download. This will download the spreadsheet to your computer.
   - If you are using `rstudio.cloud`, make sure you have it open in some project. In the file pane bottom right, click Upload and find the spreadsheet you downloaded. (For me, it goes into the Downloads folder by default.)
   - If you are running R Studio on your own computer, find where the file went (eg. by using `file.choose`).
   - Read the spreadsheet in directly. There is only one worksheet.

   (b) (2 marks) Create a new data frame that is the old one without any rows containing missing values in the `Assets` column. (Hint: `drop_na`, and save the result.)

   (c) (3 marks) Make a suitable plot of the `Assets` values (ignoring the other columns). Why would you prefer a sign test for the median to a t-test for the mean? Explain briefly.

(d) (3 marks) A fund manager thinks that the median asset value (of the funds of which these are a sample) should be 55 (million dollars). Count (using R) the number of funds that have `Assets` above and below this value. Would you expect a sign test to reject a null median of 55 (million dollars), in favour of a lower median? Explain briefly.

(e) (3 marks) Run a sign test, using `smmr`, to see whether there is any evidence that that fund manager is wrong and that the median is actually lower than 55. What do you conclude, in the context of the data?

(f) (3 marks) Obtain a 95% confidence interval for the population median. Is the value 55 inside or outside your interval? Does it make sense to compare this with your hypothesis test? Explain briefly.