STA 304H1F-1003H Winter 2020
Assignment 2-Question 4

Question 4. (10 marks)

The data set **StudentsMarks.csv** contains the term test 1 and term test 2 marks for students enrolled in STA304 Winter 2019. We consider term test 1 marks as the population of study.

Use R to do the following questions below.

You can read the data file using the following R code:

```r
mydata<-read.csv("StudentsMarks.csv")
# use x to record term test 1 marks,
# after removing students who missed the test.
x<-na.omit(mydata$Test.1)
N<-length(x)
N
```

```
## [1] 312
```

For each question, report the R code, the R output, and your answer.

(a) (2 marks) Construct a histogram of population values. Find the mean and the standard deviation

(b) (3 marks) Take a single systematic sample of size 20. Treating this sample as a SRS, estimate the average mark of students. Place a 95% bound on the error of estimation, and give a 95% confidence interval.

(c) (3 marks) Select five repeated systematic samples of size 10 each. Estimate the average mark of students. Place a 95% bound on the error of estimation, and construct a 95% confidence interval.

(d) (2 marks) Which of the two methods in (b) and (c) do you prefer, and why?