STAC53: Assignment-1

Shahriar Shams

Fall 2021

Submission deadline: October 24, 2021, 11.59pm (Local Toronto time)

Late penalty: 10% per day.

Instructions on creating documents for submission

- We will use crowdmark for submission and grading which only accepts PDF, JPG and PNG files.
- I recommend using R-markdown(if you are familiar with it). If you are not familiar with R-markdown, you can write your answers using Microsoft Word and in the end save them as pdfs. Answers that are fully handwritten will not be accepted.
- For questions where you will be doing hand calculations, you may take a picture of your calculations and paste it in to the body of your answer.
- The numerical calculations involved in this assignment are simple and you are already familiar with them (hopefully). Calculations are mostly repetitive in nature!
- For questions where you have been asked to use R, make sure you have provided your codes and outputs (as text or as screenshots).
- Make sure your answers are easy to read and nicely presented.

Academic Integrity

Each student will work alone. You are not allowed to ask anyone for help on any platform. Don't ask for solutions to anyone. Do not share your codes or answers. If you need **clarification** on any of these questions, you are allowed to **ask questions on Ed** or ask questions during office hours (please do not email us). And please do not post your solution on Ed and ask "does it look ok?".

When submitting your assignment on crowdmark, there will be a space for an academic integrity statement. Write this following statement on paper/ipad/surface and upload a screenshot of it.

Statement:

I am attesting to the fact that I, [name] (write your full name here), [stnum] (write your student number here), have abided fully to the Code of Behaviour on Academic Matters. I have not committed academic misconduct, and am aware of the penalties that may be imposed if I have committed an academic offence.

Question 1 (9 points)

Suppose you have a population of size 7 [i.e. N=7]. You measure some quantity (Y) and the corresponding numbers are:

$$11, 12, 13, 14, 15, 16, 17$$

Show hand calculation for each of the following parts of the question

a) Calculate the population mean (μ).

b) Calculate the population variance (σ^2).

c) By using random numbers from the text book (page 14) or from R, draw **two samples without** replacement. Then

   i. calculate the sample mean
   ii. calculate the sample variance
   iii. calculate $V[\bar{y}]$
   iv. calculate estimated variance of $\bar{y}$

d) Create a table of all possible combination of two samples (without replacement) from this population.

   i. calcualte the sample mean for each of these combinations
   ii. calculate the sample variance for each of these combinations
   iii. construct 95% confidence intervals for the population mean using each of the sample mean and sample variance from the previous two steps.
   iv. calculate the proportion of the intervals that truely captures the population mean that you calcualted in part (a).
   v. What value did you expect to see as an answer to part d)iv? Why is your calculated number different from the expected number? Explain briefly.

e) By using the sample variances that you have calculated in part d)ii,

   i. calculate their mean and comapare it to your answer from part(b).
   ii. What comment can you make about sample variance based on this comparison?

Question 2 (7 points)

Use R to answer this question

Copy these following two lines to R and in the first line put your student number inside the parenthesis ()

```
set.seed()
pop=sample(c(1:100),size=10)
```

If you run those two lines you will get 10 numbers saved under the name "pop". Treat these as your population.

a) Construct a table(or matrix) with all possible combination of three samples($n = 3$) **with replacement** from your population.

b) Using your table(or matrix) demonstrate that $\bar{y}$ is an unbiased estimator of the population mean.

c) Calculate $V[\bar{y}]$ by using all possible values of $\bar{y}$. (not by using the formula provided in equation 2.12 on page 19 of the text book)

d) Draw a histogram of $\bar{y}$.

e) Increase the value of n sequentially , and repeat the task of drawing the histogram of $\bar{y}$. Comment on the value of n at which your histogram start to look like a Normal density curve.

Question 3 (4 points)

Use R to answer this question

Copy these following two lines to R and in the first line put your student number inside the parenthesis ()

```
set.seed()
pop=sample(c(1:100),size=10)
```

If you run those two lines you will get 10 numbers saved under the name "pop". Treat these as your population.

a) Construct a table(or matrix) with all possible combination of three samples($n = 3$) **without replacement** from your population.

b) Using your table(or matrix) demonstrate that $\bar{y}$ is an unbiased estimator of the population mean.

c) Calculate $V[\bar{y}]$ by using all possible values of $\bar{y}$. (not by using the formula provided in equation 2.12 on page 19 of the text book)

d) By comparing your answers from question-2(parts b and c) to the answers of parts b and c of this question, comment on the performance of $\bar{y}$ under simple random sampling with vs without replacement.