Central Limit Theorem

Mini-Assignment - MTH 361 A/B - Spring 2023

Instructions:

- Please provide complete solutions for each problem. If it involves mathematical computations, explanations, or analysis, please provide your reasoning or detailed solutions.
- Note that some problems have multiple solutions or ways to solve it. Make sure that your solutions are clear enough to showcase your work and understanding of the material.
- Creativity and collaborations are encouraged. Use all of the resources you have and what you need to complete the mini-assignment. Each student must take personal responsibility and submit their work individually. Please abide by the University of Portland Academic Honor Principle.
- **Please save your work as one pdf file, don't put your name in any part of the document, and submit it to the Teams Assignments for this course. Your document upload will correspond to your name automatically in Teams.**
- If you have questions or concerns, please feel free to ask the instructor.

R Packages:

```
library(tidyverse)
library(openintro)
```

I. Sample Means and Medians

Materials

The Central Limit Theorem (CLT) is a fundamental concept in statistics that states that the sum of a large number of independent and identically distributed random variables will tend towards a normal distribution, regardless of the underlying distribution of the variables.

We will explore this theorem in more detail using a simulated population dataset generated using the code below. This is pretending that this is the population we take samples from.

```
set.seed(42)
n <- 10000 # number of item in the population
rate <- 1/2 # probability
population <- tibble(x = rexp(n,rate)) # population distribution
```

Suppose that the population has an underlying distribution taken from an exponential distribution with a true mean and median computed as follows.

```
population_mean <- mean(population$x) # mean of the population
population_mean
```

```
## [1] 2.013022
```

```
population_median <- median(population$x) # median of the population
population_median
```

```
## [1] 1.377348
```

Exercises

1. Using the `population` variable, plot the distribution of the simulated population data. Use the `ggplot` functions to create the distribution. Describe the distribution.
2. Create an R code that takes a simple random sample of size n from the population data - with replacement - and compute its mean N times. Plot the distribution of means. Repeat the same procedure for median. Describe the resulting distributions and compare it to the population distribution and the population mean. As N get very large (try $N = 10$, $N = 100$, and $N = 1000$), what happens the mean (and median) of the means (and medians).
3. (Outstanding Question) Repeat (2) but - instead of using the mean and median as statistics - use the variance and the interquartile range (IQR). Does the central limit theorem still apply to these statistics?