Jon Woody
Summer 2022: Data Analysis
Homework III
Due: TBA
Submit through Canvas

Instructions: Provided solutions to these questions using this template. Include graphics with your solutions. Put all code an appendix to this homework. Use the *verbatim* command to leave code unchanged.

1. This problem considers important computations for hypothesis testing that use the Gaussian/Normal distribution. Suppose $X \sim N(\mu = 10, \sigma^2 = 16)$.
Solution:

a) Find the constant c such that $P(X > c) = 0.05$.
Solution:
We compute c as follows

$$\begin{aligned} P(X > c) &= P\left(\frac{X = 10}{4} > \frac{c = 10}{4}\right) \\ &= P\left(Z > \frac{c - 10}{4}\right) \\ &= 0.05 \end{aligned}$$

implies that $c =$ since $P(Z > 1.645)$ since $P(Z > 1.645) = 0.05$. Therefore $c = 10 + 4 \times 1.645 = 16.58$
b) Compute $P(8.72 < X < 10.41)$ by first standardizing X. Use the Table in Canvas obtain your numerical results.
Solution:
Using R, we compute

$$\begin{aligned} P(8.72 < 10.41) &= P(X < 10.41) - P(X \leq 8.72) \\ &= 0.1663359 \end{aligned}$$

2. This problem considers the **Law of Large Numbers** (LLN). Let $X_i \sim \text{Bin}\,(n = 20, p = 0.35)$. Use *set.seed(123)* to sample the data $\{X_i\}_{i=1}^{2{,}000}$. Make a plot of $\left(\frac{1}{n}\right)\sum_{i=1}^{n} X_i$ for $n \in \{1, 2, \ldots,\ 2{,}000\}$. What value is $\left(\frac{1}{n}\right)\sum_{i=1}^{n} X_i$ approaching as $n \to \infty$?

Solution:
Below, The graphic below depicts the sample average converges to the mean of $x_i \sim \text{Bin}(n = 20, p = 0.35)$ which is $n \times p = 7.0$

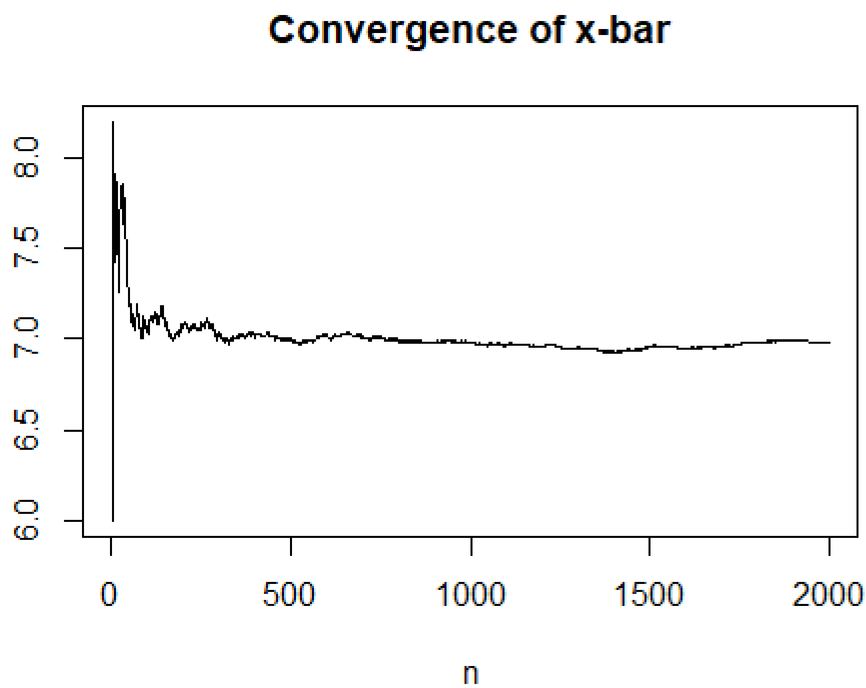

Figure 1: The Law of Large Numbers Works!

3. For this problem we will investigate the **Central Limit Theorem** (CLT).

a) Sample $\{X_i\}_{i=1}^5$ independently where each $X_i \sim \text{Exp}\,(\lambda = 5)$. Compute the sample average $\bar{x}$. Repeat this process 1000 times (that is sample the data and compute $\bar{x}$ for each sample). Make a histogram of the sample averages.
Solution:
Below, Figure (2) depicts the simulation under the parameters in Par (a). Note the graphic does not seem to depict a standard normal random variable.

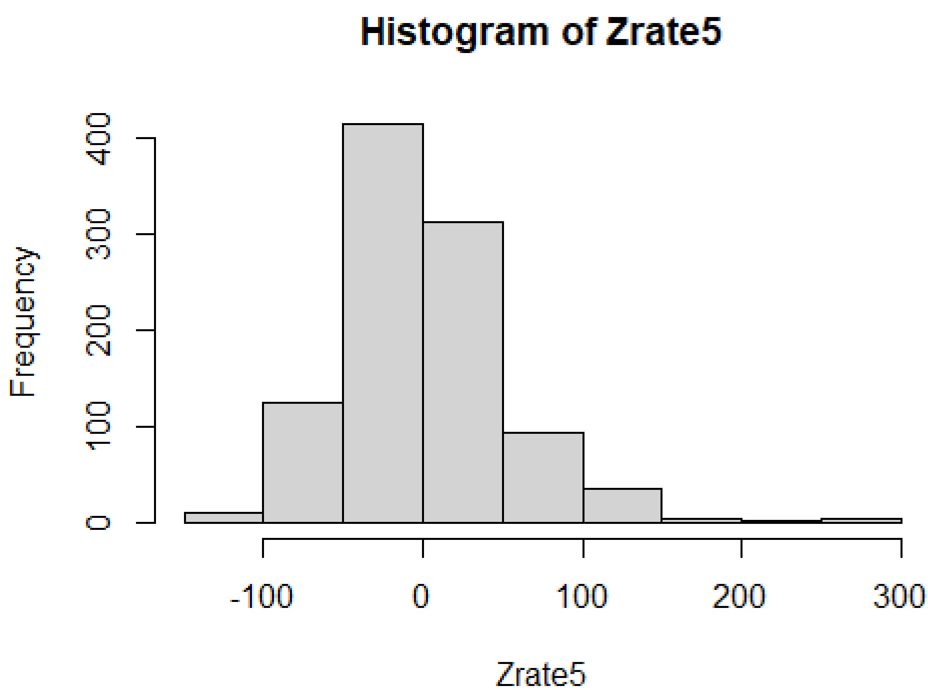

Figure 2: Sample Rate is 5

b) Repeat part a), but set that sample size to $n = 50$. That is you sample $\{X_i\}_{i=1}^{50}$.
Solution:

$$\frac{\sqrt{n(\bar{x} - E[X]}}{\sqrt{Var(X)}} \to Z \qquad\qquad \text{as } n \to \infty,$$

however we programmed

$$\frac{\sqrt{n(\bar{x} - E[X]}}{\sqrt{Var(X)}} \to Z$$

as for $i = 1, 2, \ldots, 1000$

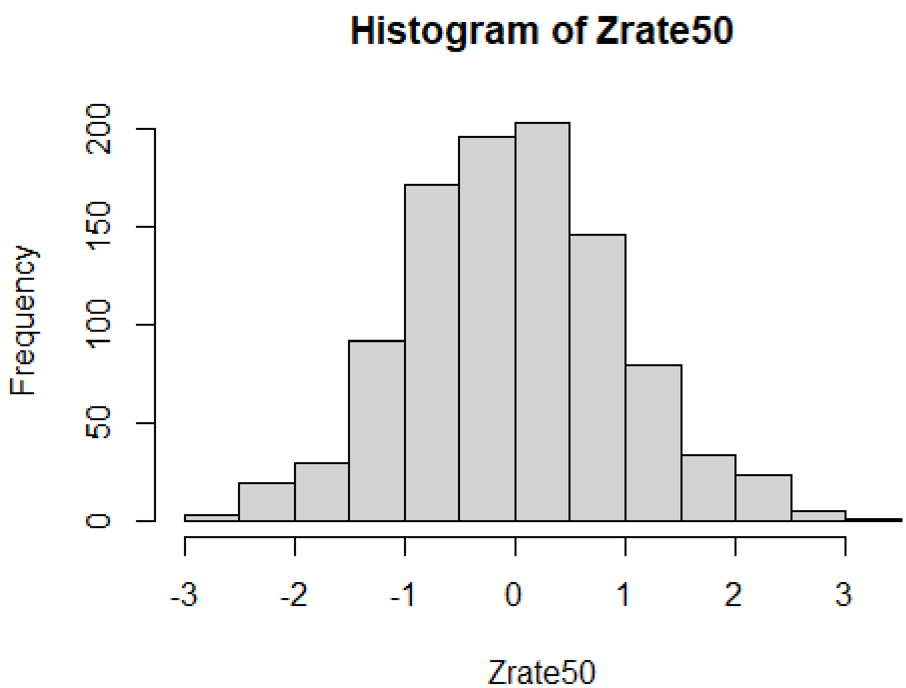

Figure 3: n = 50

c) Repeat parts a) and b), but set $X_i \sim \text{Bin}\,(n = 10, p = 0.30)$.

Solution:

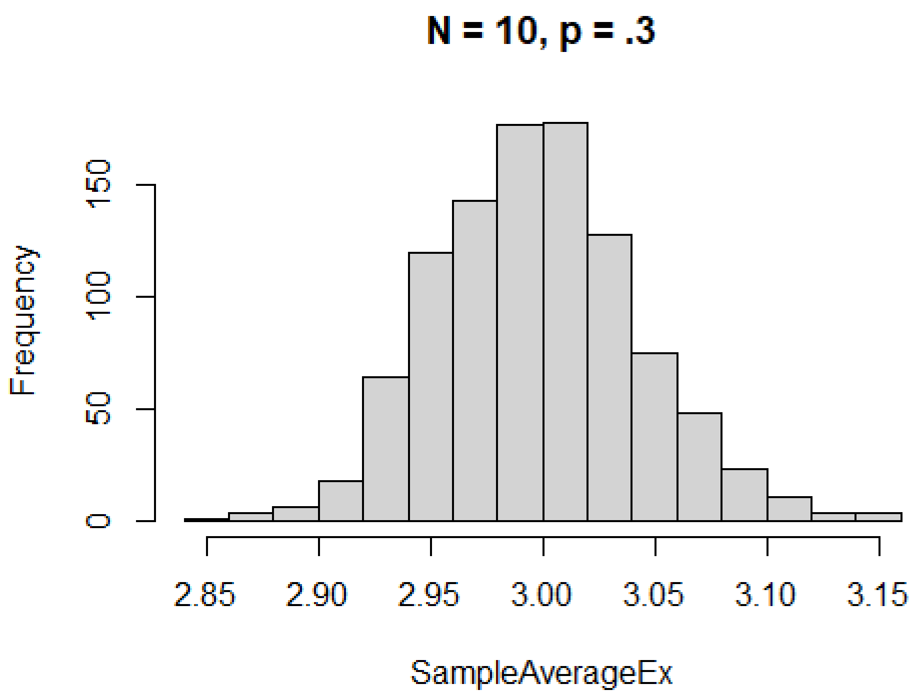

Figure 4: Binomial Distribution of Data

Appendix: Code

Here is the code for question 2):

````
```{r}
# 2

set.seed(123)
A <- rbinom(2000,20,0.35)
hist(A)

```

## Exploratory Data Analysis (EDA)

```{r}

samavg <- rep(0,2000)
n <- c(1:2000)

for(i in n){
  samavg[i] = mean(A[1:i])
}

plot(n, samavg, type = "l", xlab = 'n', ylab = '', main = "Convergence of x-bar")

```
````

Here is the code for Question 3:

```
SampleAverageEx <- rep(0, 1000)
t <- c(1:1000)
Zrate5 <- rep(0,1000)
EX = 1/5
VarX = 1/25

for (i in t){
  E <- rexp(5, rate = 5)
  SampleAverageEx[i] = mean(E)
  Zrate5[i] = sqrt(i)*(SampleAverageEx[i] - EX)/VarX
}

hist(SampleAverageEx)
hist(Zrate5)
```

Question 3 part C:

```
set.seed(123)

SampleAverageEx <- rep(0, 1000)
t <- c(1:1000)
Zrate50 <- rep(0,1000)
EX = 1/5
VarX = 1/25

for (i in t){
  Bin <- rbinom(n = 1000, size = 10, p = .3)
  SampleAverageEx[i] = mean(Bin)
  Zrate50[i] = sqrt(10)*(SampleAverageEx[i] - EX)/sqrt(VarX)
}

hist(SampleAverageEx, main = 'N = 10, p = .3')
hist(Zrate50)
```