Homework 3 - R assignment

(YOUR NAME) - I agree to abide by the Stern Code of Conduct

Question R1

This question will guide you through a *simulation study* in `R` to understand the bias of a certain estimator.

```
# This creates a function that calculates sample variance
# but using n in the denominator instead of n-1
# you don't need to change anything here
varn <- function(x) mean((x - mean(x))^2)
```

Part a.

The code below generates a sample of size n and calculates both the sample variance and the biased version of sample version (the one that has n in the denominator instead of n-1). Modify the code to change n from 10 to some other value between 10 and 50.

```
# Create a sample of data by rolling a 6-sided die n times
n <- 10
data <- sample(1:6, n, replace = TRUE)
data
```

```
##  [1] 2 4 4 6 2 3 6 5 6 4
```

```
# True variance
35/12
```

```
## [1] 2.916667
```

```
# Unbiased estimate of variance
var(data)
```

```
## [1] 2.4
```

```
# Biased estimate of variance
varn(data)
```

```
## [1] 2.16
```

Part b.

Repeat the previous experiment many times and find the expected value of each variance estimator.

```
# Unbiased estimator
mean(replicate(10000, var(sample(1:6, n, replace = TRUE))))
```

```
## [1] 2.906982
```

```
# Biased estimator
# copy the previous line here and change var to varn
```

Part c.

After running the previous code, which estimator's average value in the simulation is closer to the true value (roughly 2.9167)?

Delete this line and type your answer here.

Part d.

Now go back and change n to another, larger value, and rerun all of the code. Do you notice anything about the average of the biased estimator?

Delete this line and type your answer here.

Part e.

Copy the code from part (b) and paste it below here, then change the `mean` function to be `sd` instead.

```
# put code in here
```

Delete this sentence and replace with your answer: which estimator has more variability, the biased or unbiased one?

Question R2

According to a Marketplace/Edison survey in April of 2017, about 23.4% of survey responders agreed with the statement "the economic system in the U.S. is fair to all Americans." In this question we'll use a Bernoulli probability model to analyze this number. Suppose that there were 1,000 survey respondents and 234 agreed with the above quotation. Define a Bernoulli random variable which is 1 if a person agrees and 0 otherwise. Assume the survey was done with independent sampling (with replacement), so these Bernoulli random variables are independent. Then the number of people in the sample of 1,000 who agree is a Binomial random variable.

- We have X_i i.i.d Ber(p) for $i = 1, \dots, 1000$.
- Let $S_n = \sum_{i=1}^{n} X_i$, so S_n is Bin(n, p).

a.

Using the fact that $n\bar{X}_n = S_n$, how could you use the Binomial distribution to calculate $P(a \leq \bar{X}_n \leq b)$? How would you use `pbinom` with the given values of a, b, n, and p?

```
pbinom(something involving a, b, n, p) - pbinom(something involving a, b, n, p)
```

b.

Instead of the Binomial distribution, how would we use the central limit theorem to calculate the same probabilities? Hint: your answer should use `pnorm` and involve $\sqrt{n}$ (and a, b, and p).

```
pnorm(etc) ...
```

c.

Now let $n = 1000$, $p = 0.234$, $b = 0.250, a = 0.239$ and compute the desired probability with both methods.

```
# write code here
```