Problem Set 4 Key

1 Problem 1

The well-known Wechsler IQ test was designed so that the mean score is 100 and variance is 225. a. What is the probability that a sample of 30 individuals will have an average IQ between 130 and 140? [3pts.]

```
n <- 30
se <- sqrt(225/30)

pnorm(140,mean=100,sd=se) - pnorm(130,mean=100,sd=se)
```

```
[1] 0
```

b. What is the standard error of the average IQ for this sample? [1pt.]

```
se <- sqrt(225/30)

se
```

```
[1] 2.738613
```

c. Do we need to impose the finite population correction for 1a? Why or why not? [1pt.]

> **Note**
>
> No need for a finite correction because we can assume that there have been a lot of Wechsler IQ tests implemented and the sample size is less than 5% of the number of all Wechsler IQ tests.

2 Problem 2

Suppose the expected difference in the Alcohol Use Disorder Identification Test (AUDIT) between the control and intervention group is 5 after the intervention has taken place, with the intervention group scoring lower on average. The respective standard deviations of the historical scores for the control and intervention groups are 3.5 and 5.7.

a. If 80 participants are recruited, such that 40 participants are assigned to each group, what is the probability that the average score of the control group exceeds that of the intervention group by a value between 6 and 7? [3 pts.]

```r
diff <- 5
sd1 <- 3.5
sd2 <-5.7
n1 <- 40
n2 <- 40

se <- sqrt(sd1^2/n1 + sd2^2/n2)

pnorm(7,mean=diff,sd=se) - pnorm(6,mean=diff,sd=se)
```

```
[1] 0.1428842
```

b. The AUDIT scores can only take on positive values. What theorem can we impose so we can impose the normal approximation for the sampling distribution? [2 pts.]

> **Tip**
>
> We can impose the central limit theorem because we have a finite mean and variance as well as a relatively sizeable sample size.

3 Problem 3

```
The proportion of residents who are satisfied with the community health center’s service is 
```

a. If we randomly sample 300 residents, what is the probability that at least 80% of them are satisfied with the community health center’s service without the continuity correction? [2pts.]

```r
pi_prop <- 0.67
se <- sqrt((0.67)*(1-0.67)/300)

pnorm(0.8,mean=pi_prop,sd=se,lower.tail=F)
```

```
[1] 8.39696e-07
```

b. If we randomly sample 300 residents, what is the probability that at least 80% of them are satisfied with the community health center's service with the continuity correction? [2pts.]

> **Warning**
>
> The $\hat{p} = 0.80$, which means $\hat{p} > \pi = 0.67$. The continuity correction for the standardized variable z can be written as follows:
>
> $$z = \frac{\hat{p} - (0.5/n) - \pi}{\sqrt{\frac{\pi(1-\pi)}{n}}}$$

```r
numerator <- 0.80-(0.5/300) - 0.67
denominator <- sqrt((0.67)*(1-0.67)/300)

z <- numerator/denominator

## Remember that z follows a standard normal distribution.

pnorm(z,mean=0,sd=1,lower.tail=F)
```

```
[1] 1.138078e-06
```

c. Does the continuity correction affect the mean of the sampling distribution, variance of the sampling distribution, or both? [1pt.]

> **Note**
>
> The continuity correction affects neither one! :) It only affects the calculation of the probabilities.

! Important

On the other hand, the **finite population correction** only affects the variance of the sampling distribution.

4 Problem 4

Suppose an internet merchant has tested two website designs for the local clinic and found that Design A leads to an appointment rate of 10% while Design B leads to an appointment rate of 14%. Suppose 600 prospective patients are shown Design A and 600 prospective patients are shown Design B.

a. Find the probability that the sample appointment rate of Design A exceeds that of Design B. [2.5 pts]

i Note

We are interested in $P(\hat{p}_A - \hat{p}_B > 0)$.

```
n1 <- 600
n2 <- 600
pi1 <- 0.10
pi2 <- 0.14

se <- sqrt((pi1*(1-pi1)/n1) + (pi2*(1-pi2)/n2) )

pnorm(0,mean=pi1-pi2,sd=se,lower.tail=F)
```

```
[1] 0.0163374
```

b. Find the probability that the sample appointment rate of Design B exceeds that of Design A by at least 5%. [2.5 pts]

i Note

We are interested in $P(\hat{p}_B - \hat{p}_A > 0.05)$. or $P(\hat{p}_A - \hat{p}_B < -0.05)$

```r
n1 <- 600
n2 <- 600
pi1 <- 0.10
pi2 <- 0.14

se <- sqrt((pi1*(1-pi1)/n1) + (pi2*(1-pi2)/n2) )

pnorm(-0.05,mean=pi1-pi2,sd=se)
```

```
[1] 0.2966659
```

```r
## OR

pnorm(0.05,mean=pi2-pi1,sd=se, lower.tail=F)
```

```
[1] 0.2966659
```