HW 3 Key

1. (4 points)

From Lohr Exercise 2.10 Which of the following SRS designs will give the most precision (smallest variance) for estimating the population mean? Assume that each population has the same value of the population variance S^2. Please report the variance of the estimator for each scenario as a function of S^2 and then state which is the smallest.

1. An SRS of size 400 from a population of size 4000
2. An SRS of size 30 from a population of size 300
3. An SRS of size 3000 from a population of size 300,000,000

Recall the variance of an estimator of population mean is

$$V(\widehat{\overline{y_U}}) = \frac{S^2}{n}\left(\frac{N-n}{N}\right)$$

1. $V(\widehat{\overline{y_U}}) = \frac{S^2}{400}\left(\frac{4000-400}{4000}\right) = \frac{S^2}{400} \times 0.9 = .00225S^2$
2. $V(\widehat{\overline{y_U}}) = \frac{S^2}{30}\left(\frac{300-30}{300}\right) = \frac{S^2}{30} \times 0.9 = .03S^2$
3. $V(\widehat{\overline{y_U}}) = \frac{S^2}{3000}\left(\frac{300{,}000{,}000-3000}{300{,}000{,}000}\right) = \frac{S^2}{3000} \times 0.999 = .000333S^2$

So option 3 has the smallest variance.

2.

Using the bird data set from Lab 2, a SRS sample of size 15 grid squares has been drawn for you.

```
birds <- read.csv('http://math.montana.edu/ahoegh/teaching/stat446/birdsurvey.csv', header = T)
set.seed(09252019)
N <- 200
n <- 15
sample_values <- sample(birds$bird.counts, size = n)
```

a. (4 points)

Compute an estimate of the *total* number of birds across the entire region composed of 200 grid squares.

```
t_hat <- mean(sample_values) * N
```

The estimate for the total number of birds is 1947

b. (4 points)

Using the FPC compute the variance of the estimator for the population total.

```
var_t <- (sd(sample_values)^2/n) * N * (N - n)
```

The variance for $\hat{t}$ is 2.0560254×10^5.

c. (4 points)

Construct a 95% confidence interval for the population total.

```r
alpha <- .05
t_ci <- t_hat + c(-1,1) * qt(1-alpha/2, df = n-1) * sqrt(var_t)
```

The confidence interval for the population total is 974, 2919.

d. (4 points)

Discuss the similarities and differences for calculating uncertainty in our point estimates between this approach and the repeated sampling procedure from Lab 3.

In the lab we took multiple samples to empirically construct the sampling distribution. In this setting, we only took a single samples. Hence, uncertainty is based on the difference in the sampling units through the sample standard deviation.