

Lab 3: Partial Key

Group Member Names - here

Lab Overview

All students attending class in the group can turn in a single document with each participants name. Students not attending class will need to complete their own lab.

You have been hired to consult on a project by the Montana Department of Natural Resources. The goal is to collect data on the abundance of the Western meadowlark across a study area. Your sampling frame consists of two hundred plots of land and your budget allows sampling of twenty plots. There are four terrain types within the study area:

- desert (40 plots)
- forest (40 plots)
- wetland (20 plots)
- prairie (100 plots)

The dataset can be obtained with the following R code.

contained in the file: 'birdsurvey.csv' which can be accessed on D2L.

```
set.seed(09232019)
birds <- read.csv('http://math.montana.edu/ahoegh/teaching/stat446/birdsurvey.csv', header = T)
population_total <- birds %>% summarize(t = sum(bird.counts)) %>% select(t) %>% pull()
```

1.

a. (4 points)

Take a SRS of size 20 and compute an approximate sampling distribution (by repeated samples) of the population total.

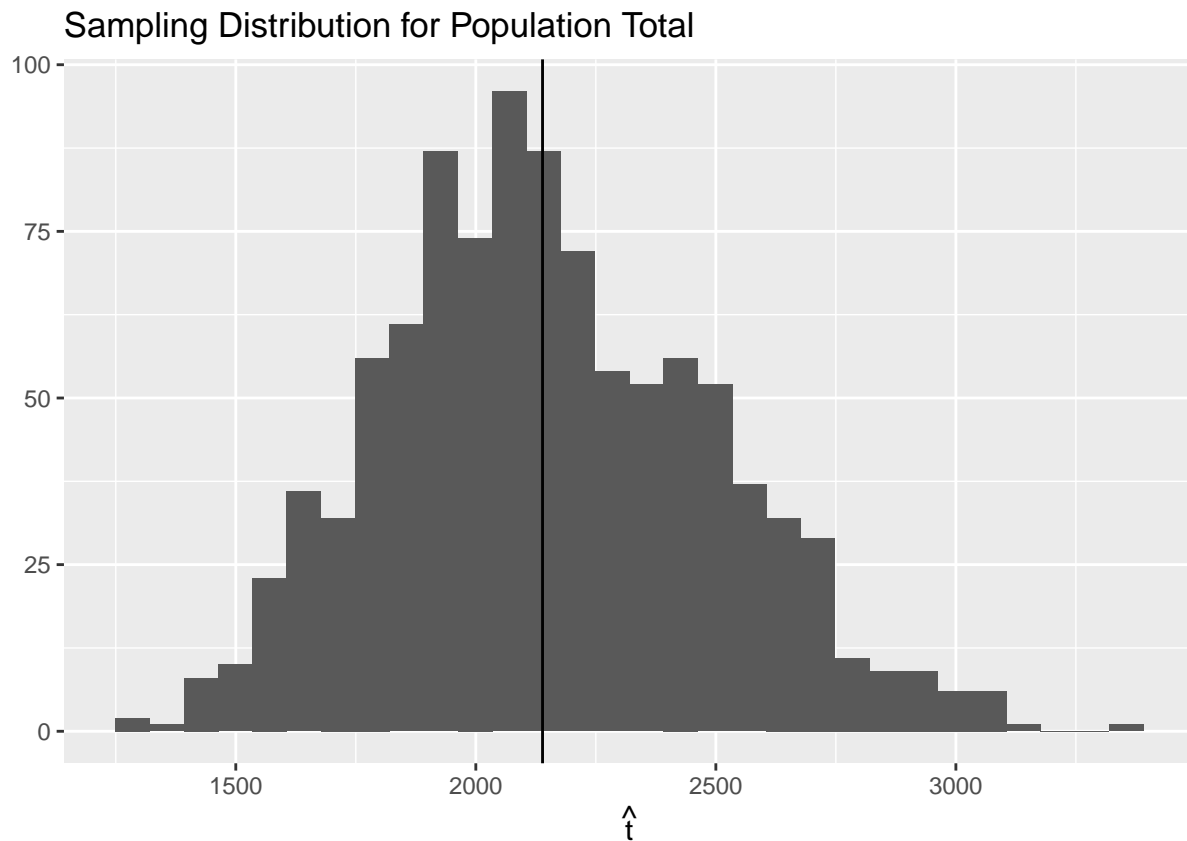
```
samples <- replicate(1000, sample(birds$bird.counts, size = 20)) %>% colMeans() * 200

# or equivalently with a loop
samples_loop <- rep(0, 200)
for (iter in 1:1000){
  samples_loop[iter] <- mean(sample(birds$bird.counts, size = 20)) * 200
}
```

b. (3 points)

Plot the approximate sampling distribution and the true population total.

```
tibble(val = samples, label = 'samples') %>% ggplot(aes(x = samples)) + geom_histogram(bins=30) +
  xlab(expression(hat(t))) + ylab('') + ggtitle('Sampling Distribution for Population Total') +
  geom_vline(xintercept = population_total)
```



c. (3 points)

Compute the MSE of the estimator.

```
mse <- mean((samples - population_total)^2)
#mse2 <- sum((samples - mean(samples))^2)/1000 + (mean(samples) - population_total)^2
```

The MSE of the estimator is 1.15849×10^5 . Often the square root of the MSE (the rMSE) is more intuitive. This value is 340

2.

a. (4 points)

Take a stratified random sample where five samples are drawn from each terrain type. Compute the MSE of this estimator.

```
#Hint this will take one sample
birds %>% group_by(terrain) %>% sample_n(5) %>% ungroup() %>% select(bird.counts) %>%
  summarize(ybar = mean(bird.counts)) %>% pull()
```

```
## [1] 13.9
```

b. (3 points)

Plot the approximate sampling distribution and the true population total.

c. (3 points)

Compute the MSE of the estimator.

3. (5 points)

a. (4 points)

What method was most effective in terms of MSE? What shortcomings did the other methods have?

b. (4 points)

Suppose another option was to use a stratified sample where 10 % of the sampling units in each strata are selected. Do you think this will perform better or worse than the option in part 2, why?