HW 4 Key

1. (6 points)

Modified from Lohr Exercise 3.3 Consider a population with 6 students and suppose we know the test scores of the students.

```
student1 <- 66; student2 <- 59; student3 <- 70
student4 <- 83; student5 <- 82; student6 <- 71
all_students <- c(student1, student2, student3, student4, student5, student6)
```

a. Find the population mean y_U and population variance S^2.

```
yu <- mean(all_students)
S2 <- var(all_students)
```

The population mean $y_U = 71.83$, and the population variance S^2=86.17.

b. Assume you plan to take a simple random sample of size 4. Calculate $V(\bar{y})$ using the following equation

$$V(\bar{y}) = \frac{S^2}{n}\left(1 - \frac{n}{N}\right)$$

```
n <- 4
N <- 6
var_ybar <- S2 / n *(1 - n / N)
```

The variance of $\bar{y} = 7.18$.

c. Now let stratum 1 consist of students 1 - 3 and stratum 2 consist of students 4 - 6. Assume you can take 2 samples from each stratum. Find $V(\bar{y}_{str})$ and compare it to that from SRS.

Recall that the variance of $\bar{y}_{str}$ can be written as

$$V(\bar{y}_{str}) = \sum_{h=1}^{H}\left(1 - \frac{n_h}{N_h}\right)\left(\frac{N_h}{N}\right)^2\left(\frac{S_h^2}{n_h}\right)$$

```
stratum1 <- c(student1, student2, student3)
stratum2<- c(student4, student5, student6)
S2_1 <- var(stratum1)
n_1 <- 2
N_1 <- 3
S2_2 <- var(stratum2)
n_2 <- 2
N_2 <- 3

var_ybar_str <- (1 - n_1 / N_1) * (N_1 / N)^2 *
  S2_1 / n_1 + (1 - n_2 / N_2) * (N_2 / N)^2 * S2_2 / n_2
```

$V(\bar{y}_{str}) = 3.14$

2. (6 points)

Modified from Lohr Exercise 3.6 Suppose that a city has 90,000 dwelling units, of which 35,000 are houses, 45,000 are apartments, and 10,000 are condominium.

a. You believe that the mean electricity usage is about twice as much for houses as for apartments or condominiums, and that the standard deviation is proportional to the mean so that $S_{house} = 2S_{apartment} = 2S_{condo}$. How would you allocate a stratified sample of 900 observations if you wanted to estimate the mean electricity consumption for all dwelling units in the city?

```
N <- 90000
N_house <- 35000
N_apt <- 45000
N_condo <- 10000
n <- 900

# set S_house to 1, the S will cancel out anyway
S_house <- 1
S_apt <- 2 * S_house
S_condo <- 2 * S_house

NhSh_house <- 35000 * S_house
NhSh_apt <- 45000 * S_apt
NhSh_condo <- 10000 * S_condo

Nl_Sl <- NhSh_house + NhSh_apt + NhSh_condo

n_house <- n * ((NhSh_house) / Nl_Sl)
n_apt <- n *((NhSh_apt) / Nl_Sl)
n_condo <- n *((NhSh_condo) / Nl_Sl)
```

Optimum allocation will result in 217 samples from houses, 559 samples from apartments, and 124 samples from condominiums.

b. Now suppose that you take a stratified random sample with proportional allocation and want to estimate the overall proportion of households in which energy conservation is practiced. If 45% of house dwellers, 25% of apartment dwellers, and 3% of condo residents in the sample practice energy conservation, with is p for the population?

```
p_str <- (N_house / N) * 0.45 + (N_apt / N) * 0.25 + (N_condo / N) * 0.03
```

This results in an overall population proportion of 0.303 dwelling units that use energy conservation.