3430 Assignment 4

Ravish Kamath: 213893664

04 April, 2022

Question 1

Return to the data of Table 6.1 and the scenario in Example 6.1. Estimate the percentage change in mean typical values of houses from 1994 to 2002 for the 47 MSAs covered by the American Housing Survey, with an appropriate margin of error. Compare this result to the percentage change in the monthly cost.

Solution

Here we have our data set with the x value being the typical value in 1994 and the y value being the typical value in 2002.

```
x = c(216962, 85378, 86763, 92664, 87615, 70759, 78542,
      97058, 101407, 88269, 127731, 123491, 176277)
y = c(300000, 92402, 125551, 135208, 126492, 99230, 116778,
      136774, 143281, 140490, 179311, 164870, 297458)
data.frame(x,y)
```

```
##         x      y
## 1  216962 300000
## 2   85378  92402
## 3   86763 125551
## 4   92664 135208
## 5   87615 126492
## 6   70759  99230
## 7   78542 116778
## 8   97058 136774
## 9  101407 143281
## 10  88269 140490
## 11 127731 179311
## 12 123491 164870
## 13 176277 297458
```

```
n = 13
N = 47
f = n/N
```

Since its asking for the estimation of the percentage change in mean of the typical values of houses from 1994 to 2002, we will be conducting a ratio estimate with a bound of error. To get the estimated ratio, recall that the formula is $r = \frac{\bar{y}}{\bar{x}}$.

```
r = mean(y)/mean(x)
r
```

```
## [1] 1.436124
```

To calculate the variance we will use the formula: $\hat{V}(r) = \frac{1-f}{n\mu_x^2}(S_y^2 + r^2 S_x^2 - 2r\hat{p}S_x S_y)$.

```
sy2 = 1/(n-1)*sum((y - mean(y))^2)
sx2 = 1/(n-1)*sum((x - mean(x))^2)
sxy = 1/(n-1)*sum((x - mean(x))*(y-mean(y)))
p_hat = sxy/(sqrt(sx2)*sqrt(sy2))

var_hat_r = (1-f)/(n*mean(x)^2)*(sy2 + r^2*sx2 - 2*r*p_hat*sqrt(sx2)*sqrt(sy2))
var_hat_r
```

```
## [1] 0.001301742
```

With the given variance we can now find an appropriate margin of error.

```
B = 2*sqrt(var_hat_r)
r - B
```

```
## [1] 1.363965
```

```
r + B
```

```
## [1] 1.508283
```

Hence we have a ratio **1.4361241** with a bound of error ±0.2966611, which gives us a 95% confidence interval, **(1.139463, 1.732785)**.

Question 2

Exercise 5.4 in Chapter 5 gives data on the typical sales price and typical size of houses for certain MSAs and CMSAs in the United States. Treating these data as coming from a stratified random sample with MSAs and CMSAs as the two strata, estimate the average price per square foot for new one-family homes in the United States. Place a bound on the error of estimation.

A stratified random sample with MSAs and CMSAs as the two strata. Please calculate **the seperate ratio estimator** for the average price per square foot for new family homes in the US, and place a bound on the error of estimation.

Solution

In this question we are trying to use the separate ratio estimator for the average price per square foot with a bound on the error. Notice that this problem deals with a stratified ratio estimator. We have two stratified random sample with MSAs and CMSAs as the 2 strata. We will use the formula $\hat{R}_{sr} = \sum_{h=1}^{H} W_h \frac{\bar{y}}{\bar{x}_h}$ to calculate the separate ratio estimator and for the variance $\sum_{h=1}^{H} W_h^2 \frac{(1-f_h)}{n_h} \frac{1}{n-1} \sum (y_i - rx_i)^2$.

```
Nh = c(250, 18)
N = sum(Nh)
nh = c(18, 10)
fh = nh/Nh
yh = c(2250700, 1522100)
xh = c(36475, 21125)
R_hat_h = yh/xh
Wh = Nh/N
varh = c(725789639, 1415080000)

R_hat_sr = sum(Wh*R_hat_h)
R_hat_sr
```

```
## [1] 62.40021
```

```
var_hat_sr = sum(Wh^2*((1-fh)/nh) *varh)
var_hat_sr
```

```
## [1] 32844629
```

```
B = 2*sqrt(var_hat_sr)

R_hat_sr - B
```

```
## [1] -11399.65
```

```
R_hat_sr + B
```

```
## [1] 11524.45
```

As we can see we get our separate ratio estimate to be **62.40021** with a bound of $\pm$**11462.05**. Hence we will have a confidence interval **(-11399.65, 11524.45)**.

Question 3

The quality control section of an industrial firm uses systematic sampling to estimate the average amount of fill in 12-ounce cans coming off an assembly line. The data in the accompanying table represent a 1-in-50 systematic sample of the production in one day. Estimate μ and place a bound on the error of estimation. Assume $N = 1800$.

Solution

Here is our data set provided to us.

```
y = c(12, 11.91, 11.87, 12.05, 11.75, 11.85,
      11.97, 11.98, 12.01, 11.87, 11.93, 11.98,
      12.01, 12.03, 11.98, 11.91, 11.95, 11.87,
      12.03, 11.98, 11.87, 11.93, 11.97, 12.05,
      12.01, 12, 11.90, 11.94, 11.93, 12.02,
      11.80, 11.83, 11.88, 11.89, 12.05, 12.04)
matrix(y, nrow = 6)
```

```
##       [,1]  [,2]  [,3]  [,4]  [,5]  [,6]
## [1,] 12.00 11.97 12.01 12.03 12.01 11.80
## [2,] 11.91 11.98 12.03 11.98 12.00 11.83
## [3,] 11.87 12.01 11.98 11.87 11.90 11.88
## [4,] 12.05 11.87 11.91 11.93 11.94 11.89
## [5,] 11.75 11.93 11.95 11.97 11.93 12.05
## [6,] 11.85 11.98 11.87 12.05 12.02 12.04
```

In this problem we are trying to estimate μ and place a bound on error. Since it is clear we are using systematic sampling, we will use the formula $\hat{\mu}_{sy} = \frac{\hat{\tau}}{N}$ where $\hat{\tau} = k \sum_{j=2}^{n_i} y_{ij}$.

```
N = 1800
k = 50
n = N/k
f = n/N

mu_hat_sy = k*sum(y)/N

s2 = var(y)
var_hat_sy = (1-f)*s2/n

B = 2*sqrt(var_hat_sy)

mu_hat_sy - B
```

```
## [1] 11.92041
```

```
mu_hat_sy + B
```

```
## [1] 11.9707
```

From the above code, we get our $\hat{\mu}_{sy}$ to **11.94556** with a bound on error $\pm$**0.02514864**. Hence our confidence interval will be **(11.92041, 11.9707)**.

Question 4

A manufacturer of band saws wants to estimate the average repair cost per month for the saws he has sold to certain industries. He cannot obtain a repair cost for each saw, but he can obtain the total amount spent for saw repairs and the number of saws owned by each industry. Thus, he decides to use cluster sampling, with each industry as a cluster. The manufacturer selects a simple random sample of $n = 20$ from the $N = 96$ industries he services. The data on total cost of repairs per industry and number of saws per industry are as given in the accompanying table. Estimate the average repair cost per saw for the past month and place a bound on the error of estimation.

Solution

Here is our data set.

```
repaircost = c(50,110,230,140,60,
               280,240,45,60,230,
               140,130,70,50,10,
               60,280,150,110,120)

saws = c(3,7,11,9,2,
         12,14,3,5,9,
         8,6,3,2,1,
         4,12,6,5,8)
data.frame(saws, repaircost)
```

```
##    saws repaircost
## 1     3         50
## 2     7        110
## 3    11        230
## 4     9        140
## 5     2         60
## 6    12        280
## 7    14        240
## 8     3         45
## 9     5         60
## 10    9        230
## 11    8        140
## 12    6        130
## 13    3         70
## 14    2         50
## 15    1         10
## 16    4         60
## 17   12        280
## 18    6        150
## 19    5        110
## 20    8        120
```

In this problem we are solving for the estimate average repair cost per saw with a bound on error on the estimation. We have the formulas for the estimate average to be $\hat{\mu} = \frac{\sum_{i \in S_c} \tau_i}{\sum_{i \in S_c} M_i}$ and $\hat{V}(\hat{\mu}) = \frac{1-f}{n(\frac{1}{N}\sum M_i)^2} \frac{1}{n-1} \sum_{i \in S_c} (\tau_i - \hat{\mu} M_i)^2$.

```
n = 20
N = 96
f = n/N
ybar = sum(repaircost)/sum(saws)
sr = sd(repaircost - ybar*saws)
var_hat = (1-f)/(n*(sum(saws)/n)^2)*sr^2
B = 2*sqrt(var_hat)

ybar - B
```

```
## [1] 17.95067
```

```
ybar + B
```

```
## [1] 21.51087
```

As we can see from the above code, we have the estimated μ to be **19.73077** with a bound of error $\pm$**1.780103**. Hence we have a 95% confidence interval, **(17.95067, 21.51087)**.