

homework6

Homework 5

Problem one

This is a cluster sampling with one-stage. $M_i = m_i = M = 215$

Step one: SRS 85 claims of 828 claims;

Step two: sample all 215 filed for the claims that have been sampled.

Here: $y_{ij} = 1$ if the i th field in the claim j is error.

(a) mean

```
N <- 828 # number of psus
n <- 85  # number of psus we sample in the first step
M <- 215 # the number of units would be sampled in cluster i (same for every psus here)

yi_bar <- rep(c(4,3,2,1,0), c(1,1,4,22,57))/M
ti_hat <- M * yi_bar
t_hat_unb <- (N/n) * sum(ti_hat)
st_2 <- (1/(n-1)) * sum((ti_hat - t_hat_unb/N)^2)
y_bar_hat <- t_hat_unb / (N*M)
v_y_bar <- (st_2 / (n * M^2)) * (1-n/N)
SE_y_bar <- sqrt(v_y_bar)
CI_y_bar <- y_bar_hat + c(-1,1) * qnorm(0.975) * SE_y_bar
y_bar_hat; SE_y_bar
```

```
## [1] 0.002024624
```

```
## [1] 0.0003570679
```

```
CI_y_bar
```

```
## [1] 0.001324784 0.002724464
```

Thus, the estimated the error rate is 0.002024624, standard error is 0.0003570679 and the confidence interval is [0.001324784, 0.002724464]

(b) total

```
vt_hat <- N^2 * (st_2 /n) * (1-n/N)
SE_t_hat <- sqrt(vt_hat)
CI_t <- t_hat_unb + c(-1,1) * qnorm(0.975) * SE_t_hat
t_hat_unb;SE_t_hat
```

```
## [1] 360.4235
```

```
## [1] 63.56523
```

```
CI_t
```

```
## [1] 235.8380 485.0091
```

Thus, the estimated total number of errors is 360.4235, standard error is 63.56523 and the confidence interval is [235.8380,485.0091]

Problem two

This is a cluster sampling with two stage.

Step one: SRS 12 of 580 cases

Step two: SRS 3 of 24 cans for each cases

Here: y_{ij} is the number of worm fragments of i th can in the j th cases.

In this case the number of psus and ssus is the same which makes the unbaised estimate a better choice.

(a) mean

```

N <- 580
n <- 12
M <- rep(24,n)
m <- rep(3,n)
M0 <- sum(rep(24,N))
y <- matrix(data = c(1,4,0,3,4,0,5,3,7,3,4,0,5,2,1,6,9,7,5,0,3,1,7,0,7,4,2,6,8,3,1,2,
5,4,9,0),nrow = 3,ncol = 12,byrow = T)

yi_hat <- apply(y, 2, mean)
ti_hat <- M * yi_hat
t_unb_hat <- (N/n) * sum(ti_hat)

st_2 <- (1/(n-1)) * sum((ti_hat - t_unb_hat / N)^2)
si_2 <- apply(y, 2, var)
v_t_hat <- N^2 * (st_2 / n) * (1 - n/N) + (N / n) * sum(si_2 * (M^2/m) * (1-m/M))
se_t_hat <- sqrt(v_t_hat)

y_bar_unb <- t_unb_hat / M0
se_y_bar <- se_t_hat / M0
y_bar_unb; se_y_bar

```

```
## [1] 3.638889
```

```
## [1] 0.6101924
```

Thus, the estimated mean number of worm fragments per can is 3.638889, standard error is 0.6101924.

(b) total

```
t_unb_hat;se_t_hat
```

```
## [1] 50653.33
```

```
## [1] 8493.878
```

Thus, the estimated total number of worm fragments is 50653.33, the standard error is 8493.878.

Problem three

In this problem the cluster size varies, thus it could be better to choose the ratio estimation

This is a two stage cluster sampling

step one: SRS 4 of 29 high school

step two: SRS the female student for each school

(a)

```
N <- 29
n <- 4
Mi <- c(792,447,511,800)
mi <- c(25,15,20,40)
yi_bar <- c(10,3,6,27) / mi
ti_hat <- Mi * yi_bar

y_bar_r <- sum(Mi * yi_bar) / sum(Mi)

sr_2 <- (1/(n-1)) * sum(Mi^2 * (yi_bar-y_bar_r)^2)

si_2 <- mi * yi_bar * (1-yi_bar) /(mi-1)

SE_y_bar_r <- sqrt( (sr_2/(n * mean(Mi)^2)) * (1 - n/N) + (1/(n*N*mean(Mi)^2)) * sum(
Mi^2 * (si_2 / mi) * (1- mi/Mi)))
y_bar_r;SE_y_bar_r
```

```
## [1] 0.4311765
```

```
## [1] 0.09910716
```

```
CI_y_bar_r <- y_bar_r + c(-1,1) * qnorm(0.975) * SE_y_bar_r
CI_y_bar_r
```

```
## [1] 0.2369300 0.6254229
```

Thus, the estimated percentage of female high school students in the region who smoke is 0.4311765, the standard error is 0.09910716 and the confidence interval is [0.2369300,0.6254229]

(b)

```
t_unb_hat <- (N/n) * sum(yi_bar * Mi)
st_2 <- (1/(n-1)) * sum((ti_hat-t_unb_hat/N)^2)
SE_t_hat <- sqrt(N^2 * (st_2 / n) * (1 - n/N) + (N / n) * sum(si_2 * (Mi^2/mi) * (1-m
i/Mi)))
t_unb_hat;SE_t_hat
```

```
## [1] 7971.375
```

```
## [1] 2725.67
```

```
CI_t_unb_hat <- t_unb_hat + c(-1,1) * qnorm(0.975) * SE_t_hat  
CI_t_unb_hat
```

```
## [1] 2629.159 13313.591
```

Thus, the estimated total number of female high school students in the region who smoke is 7971.375, the standard error is 2725.67 and the confidence interval is [2629.159,13313.591]