GR5234 - HW4

Mathieu Sauterey - UNI: mjs2364

October 9, 2018

Problem 1

Stratified random sampling was used to estimate the total number of bushels of hard shell clams in Narragansett Bay, Rhode Island. The area of interest was divided into four strata based on preliminary surveys that identified areas in which clams were abundant. Then nh dredge tows were made in stratum h for h = 1, 2, 3, 4. The acreage for each stratum was known, and the area fished during a standard dredge tow was calculated to be 0.04 acres; thus we may use $N_h = 25$ x $Area_h$.

(a) Estimate the total number of bushels of claims in the area, and give the standard error of your estimate.

```r
area = c(222.81, 49.61, 50.25, 197.81)

n.h = c(4, 6, 3, 5)

N.h = round(25*area)

ybar.h = c(0.44, 1.17, 3.92, 1.80)

s.h = sqrt(c(0.068, 0.042, 2.146, 0.794))

t.hat_strat = sum(ybar.h*N.h)

V.hat.t <- sum(N.h^2 * s.h^2 / n.h * (1 - n.h/N.h))

SE.t <- sqrt(V.hat.t)

print(t.hat_strat)
```

```
## [1] 17726.12
```

```r
print(SE.t)
```

```
## [1] 2354.302
```

E[$\hat{t}$] = 17,726.12

SE[$\hat{t}$] = 2,354.30

(b) Another survey was performed at the end of the commercial season. In this survey, strata 1, 2, and 3 were collapsed into a single stratum, called stratum 1 below. Estimate the total number of bushels of clams (with standard error) at the end of the season.

```
area = c(322.67, 197.81)

n.h = c(8, 5)

N.h = round(25*area)

ybar.h = c(0.63, 0.40)

s.h = sqrt(c(0.083, 0.046))

t.hat_strat = sum(ybar.h*N.h)

V.hat.t <- sum(N.h^2 * s.h^2 / n.h * (1 - n.h/N.h))

SE.t <- sqrt(V.hat.t)

print(t.hat_strat)
```

```
## [1] 7060.21
```

```
print(SE.t)
```

```
## [1] 948.2823
```

E$[\hat{t}]$ = 7,060.21

SE$[\hat{t}]$ = 948.28

Problem 2

The data file agsrs contains information on the number of farms and acres devoted to farms, for an SRS of n = 300 counties from the population of N = 3078 in the United States. In 1987, the United States had a total of 2,087,759 farms.

(a) Use ratio estimation to estimate the total number of acres devoted to farming in 1992, using the number of farms in 1987 as the auxiliary variable. Give a 95% confidence interval.

```
# Loads the necessary package
library("SDaA")
```

```
## Warning: package 'SDaA' was built under R version 3.4.4
```

```
# Loads the county data
county_data = agsrs

# Pop size
```

```r
N = 3078
tx_U = 2087759

# Sample data and pop of x
x_87 = county_data$farms87
y_92 = county_data$acres92
xbar.U = tx_U / N

# Ratio estimator function
ratio.estimator.mean <- function(x.samp, y.samp, N, xbar.U)
{
  n <- length(y.samp)
  xbar <- mean(x.samp); ybar <- mean(y.samp);
  B.hat <- ybar / xbar
  ybar.hat.r <- B.hat * xbar.U
  e <- y.samp - B.hat * x.samp
  V.hat <- (xbar.U/xbar)^2 * var(e)/n * (1 - n/N)
  SE <- sqrt(V.hat)
  answer <- c(point.est=ybar.hat.r, std.error=SE)
  return(answer)
}

# Calculates 1992 acres total + SE
result = ratio.estimator.mean(x.samp = x_87, y.samp = y_92, N, xbar.U)
N * result
```

```
## point.est std.error
## 960155061  68446406
```

```r
# 95% CI for the population total of US acres in 1992
N * ( result[1] + c(-1,1) * 1.96 * result[2] )
```

```
## [1]  826000106 1094310016
```

$\text{E}[\hat{t}_{yr}] = 960{,}155{,}061$ acres

$\text{SE}[\hat{t}_{yr}] = 68{,}446{,}406$ acres

[826,000,106 ; 1,094,310,016] is a 95% CI for the total number of acres in the US devoted to farming in 1992.

(b) Repeat part (a), using regression estimation.

```r
# Regression estimator function
regression.estimator.mean <- function(x.samp, y.samp, N, xbar.U)
{
  n <- length(y.samp)
  xbar <- mean(x.samp); ybar <- mean(y.samp);
  fit <- lsfit(x.samp, y.samp)
  B1.hat <- as.numeric(fit$coefficients)[2]
  ybar.hat.reg <- ybar + B1.hat * (xbar.U - xbar)
  e <- fit$residuals
  V.hat <- var(e)/n * (1 - n/N)
  SE <- sqrt(V.hat)
  answer <- c(point.est=ybar.hat.reg, std.error=SE)
  return(answer)
```

```
}

# Calculates 1992 acres total + SE
result <- regression.estimator.mean(x.samp = x_87, y.samp=y_92, N=N, xbar.U=xbar.U)
N * result
```

```
## point.est std.error
## 921406265  58065813
```

```
# 95% CI for the population total of US acres in 1992
N * ( result[1] + c(-1,1) * 1.96 * result[2] )
```

```
## [1]  807597271 1035215259
```

$\text{E}[\hat{t}_{yreg}] = 921{,}406{,}265$ acres

$\text{SE}[\hat{t}_{yreg}] = 58{,}065{,}813$ acres

[807,597,271 ; 1,035,215,259] is a 95% CI for the total number of acres in the US devoted to farming in 1992.

Problem 3

Use the data in agsrs to estimate the total number of acres devoted to farming in 1992 for each of two domains:

(a) counties with fewer than 600 farms.

```
# Sets up domains
domain.samp <- ifelse(county_data$farms92 < 600, 1 , 0)

# Domain estimation function
domain.estimation <- function(y.samp, domain.samp, d, N)
{
 n <- length(y.samp)
 x <- (domain.samp==d); y <- y.samp; u <- x*y;
 u.bar <- mean(u); s2.u <- var(u);
 t.yd.hat <- N * u.bar
 V.hat <- N^2 * s2.u/n * (1 - n/N)
 SE <- sqrt(V.hat)
 answer <- c(point.est=t.yd.hat, std.error=SE)
 return(answer)
}

# Calculates domain total (1992 acres total + SE for domain farms < 600)
domain.estimation(y.samp = y_92, domain.samp, d = 1, N)
```

```
## point.est std.error
## 497939808  55919525
```

$\text{E}[\hat{t}_{yr}] = 497{,}939{,}808$ acres

$\text{SE}[\hat{t}_{yr}] = 55{,}919{,}525$ acres

(b) counties with 600 or more farms

```r
# Calculates domain total (1992 acres total + SE for domain farms >= 600)
domain.estimation(y.samp = y_92, domain.samp, d = 0, N)
```

```
## point.est std.error
## 418987302  38938277
```

$\mathrm{E}[\hat{t}_{yr}] = 418{,}987{,}302$ acres

$\mathrm{SE}[\hat{t}_{yr}] = 38{,}938{,}277$ acres