

Lista2-Muestreo

Hector Mattos

June 9, 2018

Pregunta 1

Pregunta 11

Pregunta 12

Parte a)

Se procede a la carga del archivo y a la realización del muestreo por estratos:

```
library(data.table,quietly = T)
library(sampling,quietly = T)
library(survey,quietly = T)

setwd('D:/Archivos/Maestría/Ciclo 1/Técnicas de Muestreo/Lista2')
load("ce2s16Cz.rdata")
ce2s16Cz=as.data.table(ce2s16Cz)
#Limpiar----
ce2s16Cz=ce2s16Cz[!is.na(ce2s16Cz$M500_M),]
ce2s16Cz=ce2s16Cz[!is.na(ce2s16Cz$Sexo),]
ce2s16Cz=ce2s16Cz[!is.na(ce2s16Cz$Area),]
ce2s16Cz=ce2s16Cz[!is.na(ce2s16Cz$Gestion),]
ce2s16Cz[,Sexo:=droplevels(Sexo)]
#Crear estrato----
ce2s16Cz[,Estrato:=interaction(Area,Gestion)]
#Crear fpc----
ce2s16Cz[,fpc:=.N,by=Estrato]
#Tamaño de estratos
N_h=as.numeric(table(ce2s16Cz$Estrato))
#Muestra piloto
sigmah=numeric(4)
set.seed(351681)
sigmah[1] =
  sd(ce2s16Cz$M500_M[ce2s16Cz$Estrato==levels(ce2s16Cz$Estrato)[1]][sample(N_h[1],10)])
sigmah[2] =
  sd(ce2s16Cz$M500_M[ce2s16Cz$Estrato==levels(ce2s16Cz$Estrato)[2]][sample(N_h[2],10)])
sigmah[3] =
  sd(ce2s16Cz$M500_M[ce2s16Cz$Estrato==levels(ce2s16Cz$Estrato)[3]][sample(N_h[3],10)])
sigmah[4] =
  sd(ce2s16Cz$M500_M[ce2s16Cz$Estrato==levels(ce2s16Cz$Estrato)[4]][sample(N_h[4],10)])
#Hallar proporciones
ah = N_h*sigmah/sum(N_h*sigmah)
#Hallar tamaños de muestras
e=5
d = nrow(ce2s16Cz)*e/qnorm(0.975)
n = sum(((N_h*sigmah)^2)/ah)/(d^2 + sum(N_h*sigmah^2))
n_h = pmax(ceiling(ah*n),2)
```

```

#Ordenar Base según estratos
ce2s16Cz=ce2s16Cz[order(ce2s16Cz$Estrato),]
#muestreo
set.seed(65452)
m=sampling::strata(data = ce2s16Cz,stratanames = c("Estrato"),size=n_h,method="srswor")
samp = getdata(ce2s16Cz,m)
#Diseño
disMAE = svydesign(id=~1,strata=~Estrato,fpc=~fpc,data=samp)

```

Luego, para el cálculo de la media y su respectiva desviación según el dominio “Sexo”

```

#Variables para la aplicación de dominios (Sexo)----
#Estimación del N_dh dado que no se debe conocer
N_dh=round(table(samp$Sexo,samp$Estrato)*
             matrix(c(N_h,N_h),nrow = 2,byrow = T)/matrix(c(n_h,n_h),nrow = 2,byrow = T))
N_d=rowSums(N_dh)
#Estimación del mu_dh dado que no se debe conocer
mu_dh=samp[,mean(M500_M),by=(Sexo,Estrato)]
#Estimación del var_dh dado que no se debe conocer
var_dh=samp[,var(M500_M),by=(Sexo,Estrato)]
#Esperanza de la media para los dominios----
mu_d=numeric(2)
names(mu_d)=c('Hombre','Mujer')
dom='Hombre'
mu_d[dom]=sum((N_h/n_h)*samp[Sexo==dom,sum(M500_M),by=Estrato]$V1)/
           sum((N_h/n_h)*samp[Sexo==dom,length(M500_M),by=Estrato]$V1)
dom='Mujer'
mu_d[dom]=sum((N_h/n_h)*samp[Sexo==dom,sum(M500_M),by=Estrato]$V1)/
           sum((N_h/n_h)*samp[Sexo==dom,length(M500_M),by=Estrato]$V1)
#Varianza para los dominios (Sexo)----
var_dom=numeric(2)
names(var_dom)=c('Hombre','Mujer')
dom='Hombre'
var_dom[dom]=
  (1/N_d[dom]^2)*
  sum((N_h^2/n_h)*(1-n_h/N_h)*
       ((N_dh[dom,]-1)/(N_h-1))*
       var_dh[Sexo==dom,V1]+(N_dh[dom,]/(N_h-1))*
       (1-N_dh[dom,]/N_h)*(mu_dh[Sexo==dom,V1]-mu_d[dom])^2))
dom='Mujer'
var_dom[dom]=
  (1/N_d[dom]^2)*
  sum((N_h^2/n_h)*(1-n_h/N_h)*
       ((N_dh[dom,]-1)/(N_h-1))*
       var_dh[Sexo==dom,V1]+(N_dh[dom,]/(N_h-1))*
       (1-N_dh[dom,]/N_h)*(mu_dh[Sexo==dom,V1]-mu_d[dom])^2))
#Resultados por fórmula----
data.frame(Sexo=c('Hombre','Mujer'),M500_M=mu_d,se=sqrt(var_dom))

```

```

##           Sexo    M500_M      se
## Hombre Hombre 547.4787 3.994217
## Mujer  Mujer 542.1971 3.488693

```

```

#Resultados por svyby----
svyby(formula = ~M500_M,by = ~Sexo,design = disMAE,FUN = svymean)

```

```
##           Sexo   M500_M           se
## Hombre Hombre 547.4787 3.986556
## Mujer  Mujer 542.1971 3.482917
```

Parte b)

Para este caso se colapsarán los estratos para solo quedarnos con “Estatál” y “No estatal”. Se asumirá que cada nuevo estrato se realizó con un MASs. De esta manera, y mediante el IC de confianza de la resta de media se puede hallar lo siguiente:

```
#Colapsar Estratos----
ce2s16Cz[,Estrato2:=
  as.factor(c('Estatál','Estatál','No estatal','No estatal')
    [match(Estrato,c('Urbana.Estatál',
                    'Rural.Estatál',
                    'Urbana.No estatal',
                    'Rural.No estatal'))])]

ce2s16Cz[,fpc2:=.N,by=Estrato2]
samp[,Estrato2:=ce2s16Cz$Estrato2[ID_unit]]
samp[,fpc2:=ce2s16Cz$fpc2[ID_unit]]
#Nuevo diseño----
disMAE2 = svydesign(id=~1,strata=~Estrato2,fpc=~fpc2,data=samp)
#Varianza para M Estatal----
dom='Mujer'
est='Estatál'

n_h=sum(samp$Estrato2==est)
N_h=sum(ce2s16Cz$Estrato2==est)
N_dh=sum(samp$Sexo[samp$Estrato2==est]==dom)*N_h/n_h
N_d=N_dh
mu_dh=samp[Estrato2==est & Sexo==dom,mean(M500_M)]
var_dh=samp[Estrato2==est & Sexo==dom,var(M500_M)]
var_dom1=
  (1/N_d^2)*
  (N_h^2/n_h)*(1-n_h/N_h)*
  (((N_dh-1)/(N_h-1))*var_dh)

est='No estatal'

n_h=sum(samp$Estrato2==est)
N_h=sum(ce2s16Cz$Estrato2==est)
N_dh=sum(samp$Sexo[samp$Estrato2==est]==dom)*N_h/n_h
N_d=N_dh
mu_dh=samp[Estrato2==est & Sexo==dom,mean(M500_M)]
var_dh=samp[Estrato2==est & Sexo==dom,var(M500_M)]
var_dom2=
  (1/N_d^2)*
  (N_h^2/n_h)*(1-n_h/N_h)*
  (((N_dh-1)/(N_h-1))*var_dh)

var_total=var_dom1+var_dom2
#IC por fórmula----
samp[Estrato2=='No estatal' & Sexo==dom,mean(M500_M)]-
```

```
samp[Estrato2=='Estatl' & Sexo==dom,mean(M500_M)]+
qnorm(0.975)*sqrt(var_total)*c(-1,1)

## [1] 18.80882 56.85671

#IC por svyby----
tabla=svyby(formula = ~M500_M,by = ~Sexo+Estrato2,design = disMAE2,FUN = svymean)
tabla$M500_M[tabla$Estrato2=='No estatal' & tabla$Sexo==dom]-
  tabla$M500_M[tabla$Estrato2=='Estatl' & tabla$Sexo==dom]+
  qnorm(0.975)*sqrt(tabla$se[tabla$Estrato2=='No estatal' & tabla$Sexo==dom]^2+
    tabla$se[tabla$Estrato2=='Estatl' & tabla$Sexo==dom]^2)*c(-1,1)

## [1] 18.87435 56.79118
```

De este IC se sabe a un 95% de que las notas de mujeres en Matemáticas es mayor en colegios no estatales vs colegios estatales.

Parte c)

De la misma manera, para este caso se colapsarán los estratos para solo quedarnos con “Estatl” y “No estatal”. Se asumirá que cada nuevo estrato se realizó con un MASs. De esta manera, y mediante el IC de confianza de la resta de media se puede hallar lo siguiente:

```
#Colapsar Estratos----
ce2s16Cz[,Estrato2:=
  as.factor(c('Estatl','Estatl','No estatal','No estatal')
    [match(Estrato,c('Urbana.Estatl',
      'Rural.Estatl',
      'Urbana.No estatal',
      'Rural.No estatal'))])]

ce2s16Cz[,fpc2:=.N,by=Estrato2]
samp[,Estrato2:=ce2s16Cz$Estrato2[ID_unit]]
samp[,fpc2:=ce2s16Cz$fpc2[ID_unit]]

#Nuevo diseño----
disMAE2 = svydesign(id=~1,strata=~Estrato2,fpc=~fpc2,data=samp)

#Varianza para M Estatal----
dom='Hombre'
est='Estatl'

n_h=sum(samp$Estrato2==est)
N_h=sum(ce2s16Cz$Estrato2==est)
N_dh=sum(samp$Sexo[samp$Estrato2==est]==dom)*N_h/n_h
N_d=N_dh
mu_dh=samp[Estrato2==est & Sexo==dom,mean(M500_M)]
var_dh=samp[Estrato2==est & Sexo==dom,var(M500_M)]
var_dom1=
  (1/N_d^2)*
  (N_h^2/n_h)*(1-n_h/N_h)*
  (((N_dh-1)/(N_h-1))*var_dh)

est='No estatal'

n_h=sum(samp$Estrato2==est)
N_h=sum(ce2s16Cz$Estrato2==est)
N_dh=sum(samp$Sexo[samp$Estrato2==est]==dom)*N_h/n_h
```

```

N_d=N_dh
mu_dh=samp[Estrato2==est & Sexo==dom,mean(M500_M)]
var_dh=samp[Estrato2==est & Sexo==dom,var(M500_M)]
var_dom2=
  (1/N_d^2)*
  (N_h^2/n_h)*(1-n_h/N_h)*
  (((N_dh-1)/(N_h-1))*var_dh)

var_total=var_dom1+var_dom2
#IC por fórmula----
samp[Estrato2=='No estatal' & Sexo==dom,mean(M500_M)]-
  samp[Estrato2=='Estatl' & Sexo==dom,mean(M500_M)]+
  qnorm(0.975)*sqrt(var_total)*c(-1,1)

```

```
## [1] 13.44275 57.83736
```

```

#IC por svyby----
tabla=svyby(formula = ~M500_M,by = ~Sexo+Estrato2,design = disMAE2,FUN = svymean)
tabla$M500_M[tabla$Estrato2=='No estatal' & tabla$Sexo==dom]-
  tabla$M500_M[tabla$Estrato2=='Estatl' & tabla$Sexo==dom]+
  qnorm(0.975)*sqrt(tabla$se[tabla$Estrato2=='No estatal' & tabla$Sexo==dom]^2+
    tabla$se[tabla$Estrato2=='Estatl' & tabla$Sexo==dom]^2)*c(-1,1)

```

```
## [1] 13.53439 57.74572
```

De este IC se sabe a un 95% de que las notas de mujeres en Matemáticas es mayor en colegios no estatales vs colegios estatales.

Pregunta 17