

# Muestreo En R

Erika Martinez

November 29, 2015

### 3. TAMAOS DE MUESTRA ESTRATIFICADOS

```
#Cálculo de las proporciones por estrato
df<- ENAHO2013 %>%
  mutate(TamHog=as.numeric(TamHog),
         phu=ifelse(TamHog>1,0,1)) %>%
  select(FACTOR:ZONA,TamHog,phu,-REGION) %>%
  group_by(SEGMENTO,CUESTIONARIO,HOGAR,ZONA) %>%
  summarise(TamHog=mean(TamHog),
            phu=mean(phu)) %>%
  mutate(id=paste0(SEGMENTO,CUESTIONARIO,HOGAR,ZONA))

## Error in eval(expr, envir, enclos): no se pudo encontrar la funcin
"%>%"

r<-0.05

nsizeR<- nsize(x=df$TamHog,r=r,alpha=0.05)

## Error in eval(expr, envir, enclos): no se pudo encontrar la funcin
"nsize"

nsizeR

## Error in eval(expr, envir, enclos): objeto 'nsizeR' no encontrado

Estratos<- df %>%
  select(ZONA,TamHog) %>%
  group_by(ZONA) %>%
  summarise(n=n(),
            s=sd(TamHog)) %>%
  mutate(p=n/sum(n))

## Error in eval(expr, envir, enclos): no se pudo encontrar la funcin
"%>%"

Estratos

## Error in eval(expr, envir, enclos): objeto 'Estratos' no encontrado

Estratos

## Error in eval(expr, envir, enclos): objeto 'Estratos' no encontrado

#Asignacin de la muestra proporcional a los estratos
nsizeProp<-nstrata(n=nsizeR[[2]],wh=Estratos[,4],method="proportional")

## Error in eval(expr, envir, enclos): no se pudo encontrar la funcin
"nstrata"

nsizeProp

## Error in eval(expr, envir, enclos): objeto 'nsizeProp' no encontrado
```

```

#Asignacin de la muestra ptima a los estratos

#Costo de entrevista por estrato
ch<-c(5,10)

nsizeOpt<-nstrata(n=nsizeR[[2]],wh=Estratos[,4],sh=Estratos[,3],ch,method="optimal")

## Error in eval(expr, envir, enclos): no se pudo encontrar la funcin
"nstrata"

nsizeOpt

## Error in eval(expr, envir, enclos): objeto 'nsizeOpt' no encontrado

#Asignacin de la muestra ptima a los estratos(asume costos iguales)

nsizeNeyman<-nstrata(n=nsizeR[[2]],wh=Estratos[,4],sh=Estratos[,3],method="neyman")

## Error in eval(expr, envir, enclos): no se pudo encontrar la funcin
"nstrata"

nsizeNeyman

## Error in eval(expr, envir, enclos): objeto 'nsizeNeyman' no encontrado

```

EJEMPLO: Una pequea ciudad contiene un total de 1800 hogares. La ciudad est dividida en tres distritos que contienen 820, 540 y 440 hogares, respectivamente. Una muestra aleatoria estratificada de 310 hogares contiene 120, 100 y 90 hogares, respectivamente de estos tres distritos. Se pide a los miembros de la muestra que calculen su factura total de electricidad consumida en los meses de invierno. Las respectivas medias muestrales son 290,352 y 427, y las respectivas desviaciones estndar muestrales son 47, 61 y 93.

```

mstr.m=function(N,n,media,s)
{
  Ntot=sum(N)
  f=n/N
  mestr=crossprod(N,media)/Ntot
  varm=(s^2/n)*(1-f)
  vstr=crossprod(N^2,varm)/Ntot^2
  setr=sqrt(vstr)
  a1=mestr-1.96*setr
  b1=mestr+1.96*setr
  a2=Ntot*a1
  b2=Ntot*b1
  cat("media: IC = ",a1, "--",b1,"\n")
  cat("total: IC = ",a2, "--",b2,"\n")
}

N=c(820,540,440)
n=c(120,100,90)
media=c(290,352,427)
s=c(47,61,93)

```

```
mstr.m(N,n,media,s)

## media: IC = 335.7203 -- 348.4574
## total: IC = 604296.6 -- 627223.4
```

EJEMPLO: En una ciudad que tiene tres distritos se quiere conocer la proporcin de hogares con alguna persona profesional. Se toman muestras aleatorias de esos hogares en cada un de los tres distritos y se obtienen los resultados Distrito 1: Ni=1200, ni=180, Profesionales=80, Proporción=0.4444 Distrito 2: Ni=1350, ni=190, Profesionales=50, Proporción=0.2632 Distrito 3: Ni=1050, ni=140, Profesionales=45, Proporción=0.3214

```
N=c(1200,1350,1050)
n=c(180,190,140)
 exitos=c(80,50,45)

{f=n/N
p=exitos/n; q=1-p
pestr=crossprod(N,p)/Ntot
varp=(p*q/(n-1))*(1-f)
vstr=crossprod(N^2,varp)/Ntot^2
setr=sqrt(vstr)
a=pestr-1.96*setr
b=pestr+1.96*setr
cat("proporción: IC = ",a, "--",b,"\n")
}

## Error in eval(expr, envir, enclos): objeto 'Ntot' no encontrado

mstr.p(N,n, exitos)

## Error in eval(expr, envir, enclos): no se pudo encontrar la función
"mstr.p"
```

Muestreo aleatorio por conglomerados:  
Generar el dataframe de UPMS y su respectivo tamaño

```
UPMs<- df %>%
  mutate(TamHog=as.numeric(TamHog)) %>%
  select(SEGMENTO,ZONA,TamHog) %>%
  group_by(ZONA,SEGMENTO) %>%
  summarise(Mta=n()) %>%
  arrange(desc(ZONA),SEGMENTO)

## Error in eval(expr, envir, enclos): no se pudo encontrar la función
"%>%"

head(UPMs)

## Error in head(UPMs): objeto 'UPMs' no encontrado
```

```

#Se asigna proporcionalmente la muestra a cada estrato

#Recordemos la distribucion proporcional de los estratos:
nsizeProp

## Error in eval(expr, envir, enclos): objeto 'nsizeProp' no encontrado

#Clculo del nmero de UPMs por estrato
b<-5
aurbano<- ceiling(nsizeProp[1,1]/b)

## Error in eval(expr, envir, enclos): objeto 'nsizeProp' no encontrado
arural<- ceiling(nsizeProp[2,1]/b)

## Error in eval(expr, envir, enclos): objeto 'nsizeProp' no encontrado

#Unir en un solo objeto para facilitar input de funcin
asizeppt<-rbind(aurbano,arural)

## Error in rbind(aurbano, arural): objeto 'aurbano' no encontrado
asizeppt

## Error in eval(expr, envir, enclos): objeto 'asizeppt' no encontrado

#Seleccin de la muestra con PPT(sistematica)
install.packages("pps")

## Installing package into 'C:/Users/Ery/Documents/R/win-library/3.2'
## (as 'lib' is unspecified)
## Error in contrib.url(repos, "source"): trying to use CRAN without
## setting a mirror

library(pps)
muestrappt<-ppssstrat(sizes=UPMs$Mta, stratum=UPMs$ZONA, n=asizeppt[,1])

## Error in ppssstrat(sizes = UPMs$Mta, stratum = UPMs$ZONA, n = asizeppt[,
## : objeto 'asizeppt' no encontrado

muestraUPMs<-UPMs[muestrappt,]

## Error in eval(expr, envir, enclos): objeto 'UPMs' no encontrado

#Muestra de la seleccin
head(muestraUPMs)

## Error in head(muestraUPMs): objeto 'muestraUPMs' no encontrado

tail(muestraUPMs)

## Error in tail(muestraUPMs): objeto 'muestraUPMs' no encontrado

#Procedimiento para seleccionar submuestras en cada cluster de tamao fijo

#Listado de segmentos seleccionados
segs<- unique(muestraUPMs$SEGMENTO)

```

```

## Error in unique(muestraUPMs$SEGMENTO): objeto 'muestraUPMs' no encontrado

sample<- list()

#Identifica el nmero de columna del id(se requiere para el ciclo)

idcolnum<-which( colnames(df)=="id" )

#Ciclo para seleccionar la muestra en cada segmento
for(i in 1:length(segs)){
  subsample<-data.frame(df[which(df$SEGMENTO==segs[i]),])
  sample[[i]]<-sys.sample(nrow(subsample),b)
  #Previene de los nmeros que se pasan del total del segmento
  #if(sample[[i]][b]>nrow(subsample)){
  #  sample[[i]][b]<-1
  #  }
  #Asigna en cada elemento de las submuestras los elem. seleccionados
  sample[[i]]<-subsample[unlist(sample[[i]]),idcolnum]
}

## Error in eval(expr, envir, enclos): objeto 'segs' no encontrado

#Genera el data frame de ids seleccionados(ya que estaban en una lista)
sampledf<-data.frame(id=unlist(sample))

#Uniendo el data frame de datos con la muestra seleccionada mediante la llave creada
muestrappt<-inner_join(unique(sampledf),unique(df),by="id")

## Error in eval(expr, envir, enclos): no se pudo encontrar la funcin
"inner_join"

#Verificar el procedimiento
head(muestrappt)

## Error in head(muestrappt): objeto 'muestrappt' no encontrado

install.packages("descr")

## Installing package into 'C:/Users/Ery/Documents/R/win-library/3.2'
## (as 'lib' is unspecified)
## Error in contrib.url(repos, "source"): trying to use CRAN without
setting a mirror

library(descr)
freq(muestrappt$ZONA)

## Error in freq(muestrappt$ZONA): objeto 'muestrappt' no encontrado

```