

TÉCNICAS DE MUESTREO

LISTA N° 3

JUSTO ANDRÉS MANRIQUE URBINA – 20091107

Pregunta 1:

- a) Estime la proporción de obras de construcción en Lima Metropolitana que hacen uso de la metodología BIM, reportando su intervalo de confianza al 95%.

Resolución en R:

```
rm(list = ls())

#### Importación de data ####
library(readr)
library(readxl)
DATAEX <- read_excel("D:/DATAEX.xlsx",
                     sheet = "DATAEX")

library(survey)
library(dplyr)
library(sampling)
library(sqldf)

#### Limpieza de data ####
DATAEX_m <- mutate(DATAEX, BIM_fix=
                    case_when(DATAEX$BIM=="SI" ~ 1,
                              DATAEX$BIM=="NO" ~ 0)
                    )

#### Creación del objeto survey ####
disenoM <- svydesign( ids=~CONG+NUM,
                    strata = ~ESTRATO+NULL,
                    data = DATAEX_m,
                    fpc = ~pop_estrato+no_sector,
                    nest=T
                    )

#### Pregunta 1 - Proporción Lima Metropolitana ####
svyciprop( formula = ~BIM_fix,
            design = disenoM,
            method = c("mean"),
            level = 0.95,
            na.rm=T)
```

Respuesta:

	Media	2.5%	97.5%
BIM_fix	0.243	0.181	0.31

Pregunta 2:

- b) Halle la estimación de la proporción del número de obras de construcción en Lima Top que hacen uso de la metodología BIM, junto con su error estándar de estimación.

Resolución en R:

```
#### Pregunta 2 - Proporción Lima Top ####
```

```
svyby( svymean,  
       by = ~ESTRATO,  
       design = disenoM,  
       formula = ~BIM_fix)
```

Respuesta:

	ESTRATO	BIM_fix	se
CALLAO	CALLAO	0.4038462	0.28573903
LIMA CENTRO	LIMA CENTRO	0.2579853	0.08421219
LIMA ESTE	LIMA ESTE	0.2096273	0.08605541
LIMA MODERNA	LIMA MODERNA	0.2603730	0.04546070
LIMA NORTE	LIMA NORTE	0.3355482	0.20119588
LIMA SUR	LIMA SUR	0.1201550	0.05809812
LIMA TOP	LIMA TOP	0.2267393	0.03951196

	Media	Error estándar
Lima Top	0.2267393	0.3951196

Pregunta 3:

- c) Suponga que en lugar de haberse empleado este diseño para Lima Top, usted hubiese empleado un muestreo ppt de 4 sectores, para luego encuestar a todas las obras de los sectores de Lima Top seleccionados. Implemente este diseño, reportando la proporción del número de obras de construcción en Lima Top que hacen uso de la metodología BIM, justo con su error de estimación. Compare finalmente los errores de estimación de este diseño con los del anteriormente tomado.

Resolución en R:

```
#### Pregunta 3 ####
```

```
# Lectura de Marco Muestral
```

```
DATAEX_marco <- read_excel("D:/DATAEX.xlsx",  
                           sheet = "Marco Muestral")
```

```
# Agrupación de obras por sector en por conglomerado en LimaTOP
```

```
td_1 <- aggregate(  
  DATAEX_marco$`Obras por sector`[DATAEX_marco$`Sector Urbano`=="LIMA  
TOP"],  
  by=list(DATAEX_marco$CONG[DATAEX_marco$`Sector Urbano`=="LIMA TOP"]),  
  sum)
```

```
# Determinación de las probabilidades de primer orden
```

```
pik <- inclusionprobabilities(td_1$x,  
                              4)
```

```
# Muestreo ppt mediante UPpoisson
```

```
set.seed(2230)
```

```
pik_select <- UPpoisson(pik = pik)
```

```
pik_select
```

```
# Identificación de sectores en LIMA TOP
```

```
index <- which((DATAEX_m$CONG=="MIRAFLORESB"&DATAEX_m$ESTRATO=="LIMA TOP") |  
              (DATAEX_m$CONG=="SAN BORJAB"&DATAEX_m$ESTRATO=="LIMA TOP") |  
              (DATAEX_m$CONG=="SAN ISIDROA"&DATAEX_m$ESTRATO=="LIMA TOP") |  
              (DATAEX_m$CONG=="SURCOA"&DATAEX_m$ESTRATO=="LIMA TOP")  
              )
```

```
# Extracción de la muestra
```

```
DATAEX_p3 <- DATAEX_m[index,]
```

```
# Determinación del diseño y de la media
```

```
disenoM_limatop <- svydesign(ids=~CONG,data = DATAEX_p3,fpc=~no_sector)
```

```
svymean(~BIM_fix,disenoM_limatop)
```

Respuesta:

```
      mean      SE  
BIM_fix 0.22554 0.0584
```

	Media	Error estándar
Lima Top	0.22554	0.0584

En relación a la pregunta 2, se observa que:

- A nivel de estimación de medias, los valores no difieren mucho entre sí (0.2268 vs 0.225, del muestreo estratificado por conglomerados bietápico y el muestreo ppt respectivamente).
- La variación principal proviene de la estimación del error estándar de la media (0.395 vs 0.058, del muestreo estratificado por conglomerados bietápico y el muestreo ppt respectivamente).