# Técnicas de Muestreo

## Lista N° 3

### Justo Andrés Manrique Urbina – 20091107

**Pregunta 1:**

1. 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:**

1. 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:**

1. 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).