

# Examen 3A

Cruz Martínez Ricardo y Vasquez Arriaga Jorge

## 1 Muestreo bietápico

Un estudiante de posgrado tiene una colección de 400 libros en su librero. Para estimar el número total de palabras en su colección de libros, selecciona una muestra  $s_I$  de dos libros usando un muestreo aleatorio simple sin reemplazo. Para los dos libros seleccionados tiene el número de páginas de cada uno. Posteriormente, en cada libro seleccionado selecciona dos páginas usando un muestreo aleatorio simple sin reemplazo. Una vez seleccionada la página registra de cada una el número de palabras que contiene. La información muestral es como sigue.

Cuadro 1: Muestras seleccionadas en cada etapa y valores observados

Libro seleccionado ( $s_I$ )	Número de páginas ( $N_i$ )	Páginas seleccionadas ( $s_i$ )	Número de Palabras en la página ( $y_k$ )
195	300	61	23
		212	25
288	200	99	20
		111	20

Aquí las páginas conforman la población de interés y los libros son las *upm*.

i) Calcule las probabilidades de inclusión de primer y segundo orden correspondientes a las páginas seleccionadas.

Por el **Resultado 5.4** de las notas, tenemos que las probabilidades de primer orden, considerando el muestreo bietápico están dadas por

$$\pi_k = \pi_{k|i} \pi_{Ii},$$

donde  $\pi_{Ii}$  representa la probabilidad de inclusión de la *upm*  $i$  (en este caso, serían los libros) y  $\pi_{k|i}$  la probabilidad de inclusión del elemento  $k$  cuando se realiza la selección en la *upm*  $i$  (sería la probabilidad de escoger una página del libro  $i$ ) y dado que en la primera y segunda ‘etapa’ se usó un muestreo aleatorio simple sin reemplazo, se tiene que las probabilidades de inclusión correspondientes a las páginas seleccionadas son:

$$\pi_{61} = \frac{2}{400} \cdot \frac{2}{300} = \frac{1}{30000} = \pi_{212}$$

y

$$\pi_{99} = \frac{2}{400} \cdot \frac{2}{200} = \frac{1}{20000} = \pi_{111},$$

pues

$$\pi_{I195} = \frac{2}{400} = \pi_{I288},$$

además

$$\pi_{61|195} = \frac{2}{300} = \pi_{212|195} \quad y \quad \pi_{99|288} = \frac{2}{200} = \pi_{111|288}.$$

Por otra parte, sabemos que las probabilidades de inclusión de segundo orden están dadas por

$$\pi_{kl} = \begin{cases} \pi_{Ii}\pi_{kl|i} & \text{si } i = j \\ \pi_{Iij}\pi_{k|i}\pi_{l|j} & \text{si } i \neq j \end{cases}$$

donde  $\pi_{Iij}$  representa la probabilidad de inclusión de segundo orden para las *upm*  $i$  y  $j$ ,  $i \neq j$  y  $\pi_{kl|i}$  la probabilidad de inclusión de segundo orden de los elementos  $k$  y  $l$  cuando se realiza la selección en la *upm*  $i$ , entonces, se tiene que

$$\pi_{61\&212} = \pi_{I_{195}}\pi_{61\&212|195} = \frac{2}{400} \cdot \frac{2}{300(299)} = \frac{1}{8970000}$$

$$\pi_{99\&111} = \pi_{I_{288}}\pi_{99\&111|288} = \frac{2}{400} \cdot \frac{2}{200(199)} = \frac{1}{3980000}$$

$$\pi_{61\&99} = \pi_{I_{195\&288}}\pi_{61|195}\pi_{99|288} = \frac{2}{400(399)} \cdot \frac{2}{300} \cdot \frac{2}{200} = 8.354218881 \times 10^{-10}$$

$$\pi_{61\&111} = \pi_{I_{195\&288}}\pi_{61|195}\pi_{111|288} = \frac{2}{400(399)} \cdot \frac{2}{300} \cdot \frac{2}{200} = 8.354218881 \times 10^{-10}$$

$$\pi_{212\&99} = \pi_{I_{195\&288}}\pi_{212|195}\pi_{99|288} = \frac{2}{400(399)} \cdot \frac{2}{300} \cdot \frac{2}{200} = 8.354218881 \times 10^{-10}$$

$$\pi_{212\&111} = \pi_{I_{195\&288}}\pi_{212|195}\pi_{111|288} = \frac{2}{400(399)} \cdot \frac{2}{300} \cdot \frac{2}{200} = 8.354218881 \times 10^{-10}$$

ii) Calcule el valor estimado del número total de palabras en la colección de libros.

Por el **Resultado 5.5** el estimador HT para el total se ve como

$$\hat{t}_{y\pi} = \sum_{i \in s_I} \frac{\hat{t}_{i\pi}}{\pi_{Ii}}$$

con  $\hat{t}_{i\pi} = \sum_{k \in s_i} \frac{y_k}{\pi_{k|i}}$ , por ende, tenemos que la estimación quedaría de la siguiente manera:

1. Para  $S_I = 195$ , se tiene que

$$\hat{t}_{195\pi} = \sum_{k \in \{61, 212\}} \frac{300}{2} y_k = 150(23 + 25) = 7200$$

2. Para  $S_I = 288$ , se tiene

$$\hat{t}_{288\pi} = \sum_{k \in \{99, 111\}} \frac{200}{2} y_k = 100(40) = 4000$$

Por último, nuestra estimación del total quedaría como:

$$\hat{t}_{y\pi} = \sum_{i \in \{195, 288\}} 200\hat{t}_{i\pi} = 200(7200 + 4000) = 2240000$$

iii) Dé una estimación insesgada de la varianza del estimador usado en ii).

Sabemos, por el **Resultado 5.6**, que un estimador insesgado de la varianza del estimador del inciso anterior está dado por

$$\hat{V}(\hat{t}_{y\pi}) = \hat{V}_{PSU} + \hat{V}_{SSU} = \sum_{i \in s_I} \sum_{j \in s_I} \hat{\Delta}_{Iij} \frac{\hat{t}_{i\pi}}{\pi_{Ii}} \frac{\hat{t}_{j\pi}}{\pi_{Ij}} + \sum_{i \in s_I} \frac{\hat{V}_i}{\pi_{Ii}},$$

donde  $\hat{\Delta}_{Iij} = \frac{\pi_{Iij} - \pi_{Ii}\pi_{Ij}}{\pi_{Iij}}$  y  $\hat{V}_i = \sum_{k \in s_i} \sum_{l \in s_i} \hat{\Delta}_{kl|i} \frac{y_k}{\pi_{k|i}} \frac{y_l}{\pi_{l|i}}$ , con  $\hat{\Delta}_{kl|i} = \frac{\pi_{kl|i} - \pi_{k|i}\pi_{l|i}}{\pi_{kl|i}}$ , dado que en la primera y segunda etapa se usó un muestreo aleatorio simple sin reposición, tenemos que

$$\sum_{i \in s_I} \sum_{j \in s_I} \hat{\Delta}_{Iij} \frac{\hat{t}_{i\pi}}{\pi_{Ii}} \frac{\hat{t}_{j\pi}}{\pi_{Ij}}$$

es la expresión de la varianza del estimador HT para un muestreo aleatorio simple sin reposición para la primera etapa, por lo que se tiene que

$$\sum_{i \in s_I} \sum_{j \in s_I} \hat{\Delta}_{Iij} \frac{\hat{t}_{i\pi}}{\pi_{Ii}} \frac{\hat{t}_{j\pi}}{\pi_{Ij}} = \frac{N^2}{n} \left(1 - \frac{n}{N}\right) S_{t_{i\pi} s_I}^2$$

$$\text{con } S_{t_{i\pi} s_I}^2 = \frac{\sum_{k \in s_I} (\hat{t}_{i\pi} - \bar{\hat{t}}_{i\pi})^2}{n-1}, \bar{\hat{t}}_{i\pi} = \frac{\sum_{i \in s_I} \hat{t}_{i\pi}}{n}, N = 400 \text{ y } n = 2.$$

```
s_I <-c(195,195, 288, 288)
N_i <-c(300,300, 200, 200)
s_i <-c(61, 212, 99, 111)
y_k <-c(23, 25, 20,20)

Datos <- data.frame(s_I, N_i, s_i, y_k)

t_i = c(7200, 4000)

PSU = var(t_i)*(400^2 / 2) * (1 - 2/400)
PSU
```

```
## [1] 4.07552e+11
```

De manera similar, tenemos que el segundo término del sumando, dado que a cada  $V_i$  la podemos pensar como una suma de varianzas del estimador HT aplicado a las páginas de los libros seleccionados, se tiene que

$$\hat{V}_i = \sum_{k \in s_i} \sum_{l \in s_i} \hat{\Delta}_{kl|i} \frac{y_k}{\pi_{k|i}} \frac{y_l}{\pi_{l|i}} = \frac{N_i^2}{2} \left(1 - \frac{2}{N_i}\right) S_{y_{s_i}}^2$$

$$\text{con } S_{y_{s_i}}^2 = \frac{\sum_{k \in s_i} (y_k - \bar{y}_{s_i})^2}{n-1} \text{ y } \bar{y}_{s_i} = \frac{\sum_{k \in s_i} y_k}{n}$$

```
library(dplyr)
V_i = Datos %>% group_by(s_I)%>%summarise(V_i =unique(N_i)^2/2*(1
  - 2/unique(N_i))*var(y_k))%>%ungroup()

SSU = sum(V_i) * (400 / 2)

SSU
```

```
## [1] 17976600
```

Finalmente, la estimación de la varianza estaría dada por

```
## [1] 4.0757e+11
```

iv) Dé una aproximación de la varianza del estimador usado en ii) que sólo use las probabilidades de inclusión de primer orden o los factores de expansión.

Sabemos, por la ecuación (168) de las notas, que bajo los supuestos de  $\pi_{Ii} = n_I p_i$  y  $m_I = n_I$  usando un muestreo con reemplazo para seleccionar las upm, se tiene que una estimación de la varianza es la siguiente

$$\hat{V}(\hat{t}_{y\pi}) = \frac{n_I}{n_I - 1} \sum_{i=1}^{n_I} \left( \frac{\hat{t}_{i\pi}}{\pi_{Ii}} - \frac{\hat{t}_{y\pi}}{n_I} \right)^2$$

en nuestro caso, tenemos que  $n_I = 2$ ,  $\frac{\hat{t}_{y\pi}}{2} = \frac{\frac{400}{2} \cdot (7200 + 4000)}{2} = \frac{2240000}{2} = 1120000$ , por ende, tenemos lo siguiente

```
t_ypi2 = 1120000
estVar <- 2 * sum((400/2 * t_i - t_ypi2)^2)
estVar
```

```
## [1] 4.096e+11
```

## 2 Comparación de diseños de muestreo

Suponga que se realizará una nueva elección de diputaciones a nivel federal y le han encargado realizar el diseño de muestreo. Para esto cuenta con la información de los resultados a nivel acta (casilla) de los cómputos distritales de 2021 (<https://computos2021.ine.mx/base-de-datos>).

Por simplicidad suponga que se centrará en estudiar los diseños de manera que elegirá el que sirva para realizar de la mejor forma la estimación del porcentaje de votos a favor de una coalición integrada por Morena, PT y P. Verde a nivel nacional.

En el denominador considerará sólo el total de votos válidos, es decir, que no se considerarán los votos nulos para efectos del cálculo del porcentaje.

Consideré los siguientes tres diseños a comparar:

- I. Se seleccionan 1200 casillas de las 163,666 usando un muestreo aleatorio simple sin reemplazo.
- II. Se considera una estratificación a partir de los 300 diferentes distritos electorales del país. El diseño de muestreo considerado en cada estrato corresponde a un muestreo aleatorio simple sin reemplazo de 4 casillas.
- III. Se considera una estratificación considerando las cinco circunscripciones electorales de México. En cada estrato se usa un muestreo aleatorio simple sin reemplazo para seleccionar 8 distritos electorales y en cada distrito electoral seleccionado se usa un muestreo aleatorio simple para seleccionar 30 casillas.

Realice la comparación de los tres diseños a partir de simulaciones. Es decir, repita 1000 veces lo siguiente. Seleccione una muestra con cada diseño y realice la estimación, en este caso deberá usar un estimador de razón.

A partir de las 1000 estimaciones estime el ECM del estimador que se obtendría para cada diseño y con estos resultados indique cuál diseño parece ser el mejor.

Para resolver este problema optamos por resolver manualmente debido a que con `sampling(strata, stage)` y `survey`, se tardaba demasiado tiempo.

```

# libreria para limpiar la base de datos
library(tidyverse)

# leer el csv
setwd("D:/Notas/Muestreo/Exámen/examen02/20210611_1000_CW_diputaciones")
options(digits=10)
set.seed(123)

df <- data.table::fread("diputaciones - copia.csv", sep="|")
# para el INE 34. TOTAL_VOTOS_CALCULADOS - Suma de los votos asentados en las
# actas para: los partidos políticos, combinaciones de estos, candidatos
# independientes, votos para candidaturas no registradas y votos nulos.

# como solo nos interesa el porcentaje sobre la gente que no registro voto nulo
# entonces los restamos
# y solo nos interesa las casillas donde si hubo votaciones
df <- filter(df, TOTAL_VOTOS_CALCULADOS != 0)
df <- df %>%drop_na(TOTAL_VOTOS_CALCULADOS)
df <- df %>%drop_na('VOTOS NULOS')
df$totalvotos = df$TOTAL_VOTOS_CALCULADOS - df$'VOTOS NULOS'
df <- filter(df, totalvotos != 0)

# el número total de casilla sobre el que vamos a trabajar
N <- dim(df)[1]
# tamaño de muestra
n <- 1200

# datos que nos interesan
midf <- df[, c("MORENA", "PVEM", "PT", "PT-MORENA",
               "PVEM-PT", "PVEM-MORENA", "PVEM-PT-MORENA", "totalvotos")]
# calculamos los votos a favor
midf$votos.favor = apply(midf,1, function(x) (sum(x)- x[length(x)]))

# resultado a estimar
prct.votos <- sum(midf$votos.favor)/sum(midf$totalvotos)*100
prct.votos

## [1] 44.28523292

# pesos para el mas
midf$wk.mas = N/n

diseno.mas <- function(){
  #seleccionamos muestra
  seleccion_muestra <- sample(1:N ,size = 1200)

  #estimamos con el estimador de razon de la media
  muestra <- midf[seleccion_muestra, ]
  return(sum(muestra$votos.favor*muestra$wk.mas)/sum(muestra$totalvotos*muestra$wk.mas))
}

```

```
simulacion.mas <- replicate(1000, disenom.mas())
mean(simulacion.mas)*100
```

```
## [1] 44.30034631
```

```
# #estratos
df$distritoelec <- paste(df$ID_ESTADO, df$ID_DISTRITO, df$NOMBRE_DISTRITO, sep="-")
# revisamos el tamaño deben ser 300 unicos
length(unique(df$distritoelec))
```

```
## [1] 300
```

```
midf2 <- df[, c("distritoelec", "totalvotos")]
midf2$votos.favor <- midf2$votos.favor
```

```
# otorgamos id a cada distrito
iddistrito = 1:300
distritoelec = unique(df$distritoelec)
idsdist <- data.frame(iddistrito, distritoelec)
midf2 <- merge(midf2, idsdist)
```

```
# calculo de las probabilidades de inclusión para todos los elementos
prob.inc <- data.frame(midf2 %>% group_by(iddistrito) %>% count())
prob.inc$wk <- 1/(4/prob.inc$n)
```

```
midf2 <- merge(midf2, prob.inc, by="iddistrito")
```

```
# creamos un subid (id dentro del distrito electoral)
subid <- c()
for (i in 1:300){
  s0 = midf2$distritoelec[midf2$iddistrito == i]
  subid = c(subid, 1:length(s0))
}
midf2$subid <- subid
```

```
# creamos un id único en toda la población
midf2$idun <- paste(midf2$iddistrito, midf2$subid, sep = "-")
```

```
# observemos que esta es la función que se repitara 1000 veces para la simulación
# lo que reduce el tiempo de ejecución
```

```
disenost <- function(){
  #seleccionamos muestra con base en el idunico
  idun <- c()
  for (i in 1:300) {
    idun <- c(idun, paste(i, sample(1:prob.inc$n[i], size = 4), sep="-"))
  }
  seleccion <- data.frame(idun)
  #obtenemos los datos de la muestra
  muestra <- left_join(seleccion, midf2, by = "idun")
}
```

```
#estimaciones
```

```

    return(sum(muestra$votos.favor*muestra$wk)/sum(muestra$totalvotos*muestra$wk))
}
simulacion.estr <- replicate(1000, disenost())
mean(simulacion.estr)*100

```

```
## [1] 44.27721997
```

```

# trabajaremos con un dataframe con la siguiente
midf3 <- df[, c("NOMBRE_ESTADO", "NOMBRE_DISTrito", "ID_DISTrito", "ID_ESTADO",
               "distritoelec", "totalvotos")]
# recuperamos las estimaciones hechas anteriormente
midf3$votos.favor <- midf$votos.favor

# asignamos circunscripciones con respecto a los estados
circ <- c(2, 1, 1, 3, 2, 5, 3, 1, 4, 1, 2, 4, 5, 1, 5, 5, 4, 1, 2, 3, 4,
          2, 3, 2, 1, 1, 3, 2, 4, 3, 3, 2)
# recuperamos los nombres de los estados y hacemos un dataframe con los estados
# y circunscripciones
NOMBRE_ESTADO <- unique(df$NOMBRE_ESTADO)
circdf <- data.frame(circ, NOMBRE_ESTADO)
# unimos los dataframes y ahora tenemos la circunscripción en el dataframe
midf3 <- merge(circdf, midf3, by="NOMBRE_ESTADO")

# otorgamos id unico a cada distrito
iddistrito = 1:300
distritoelec = unique(df$distritoelec)
idsdist <- data.frame(iddistrito, distritoelec)
midf3 <- merge(midf3, idsdist)

#agrupa para cada estrato (hacemos un dataframe para cada circunscripción)
circ1 <- midf3 %>% filter(circ == 1)
circ2 <- midf3 %>% filter(circ == 2)
circ3 <- midf3 %>% filter(circ == 3)
circ4 <- midf3 %>% filter(circ == 4)
circ5 <- midf3 %>% filter(circ == 5)

# ahora a cada circunscripción le calcularemos cuantas casillas le
# pertenecen, tambien un sub-id (que sera un id dentro de cada distrito),
# vector que guardara los subids
subid <- c()
# vector que guardara cuantas casillas le pertenecen a cada distrito
npob <- c()
for (i in unique(circ1$iddistrito)){
  # seleccionamos distritos únicos por circunscripción
  s0 = circ1$distritoelec[circ1$iddistrito == i]
  # agregamos el subid
  subid = c(subid, 1:length(s0))
  # total de casillas que le pertenecen al distrito
  npob <- c(npob, rep(length(s0), length(s0)))
}
# agregamos los resultados anteriores al dataframe
circ1$subid <- subid

```

```

circ1$npob <- npob
# calculamos los pesos para cada casilla
circ1$wk <- 1/(8/length(unique(circ1$ididistrito))*30/circ1$npob)
# creamos un idunico para cada casilla con base en ididistrito y subid
circ1$idun <- paste(circ1$ididistrito, circ1$subid, sep="-")
# relaciona el numero de distrito con el numero de casillas
prob.c1 <- data.frame(circ1 %>% group_by(ididistrito) %>% count())
#repetimos este proceso para cada circunscripción

#circunscripción 2
subid <- c()
npob <- c()
for (i in unique(circ2$ididistrito)){
  s0 = circ2$distritoelec[circ2$ididistrito == i]
  subid = c(subid, 1:length(s0))
  npob <- c(npob, rep(length(s0), length(s0)))
}
circ2$subid <- subid
circ2$npob <- npob
circ2$wk <- 1/(8/length(unique(circ2$ididistrito))*30/circ2$npob)
prob.c2 <- data.frame(circ2 %>% group_by(ididistrito) %>% count())
circ2$idun <- paste(circ2$ididistrito, circ2$subid, sep="-")

# circunscripción 3
subid <- c()
npob <- c()
for (i in unique(circ3$ididistrito)){
  s0 = circ3$distritoelec[circ3$ididistrito == i]
  subid = c(subid, 1:length(s0))
  npob <- c(npob, rep(length(s0), length(s0)))
}
circ3$subid <- subid
circ3$npob <- npob
circ3$wk <- 1/(8/length(unique(circ3$ididistrito))*30/circ3$npob)
prob.c3 <- data.frame(circ3 %>% group_by(ididistrito) %>% count())
circ3$idun <- paste(circ3$ididistrito, circ3$subid, sep="-")

# circunscripción 4
subid <- c()
npob <- c()
for (i in unique(circ4$ididistrito)){
  s0 = circ4$distritoelec[circ4$ididistrito == i]
  subid = c(subid, 1:length(s0))
  npob <- c(npob, rep(length(s0), length(s0)))
}
circ4$subid <- subid
circ4$npob <- npob
circ4$wk <- 1/(8/length(unique(circ4$ididistrito))*30/circ4$npob)
prob.c4 <- data.frame(circ4 %>% group_by(ididistrito) %>% count())
circ4$idun <- paste(circ4$ididistrito, circ4$subid, sep="-")

# circunscripción 5
subid <- c()

```



```

npob <- c()
for (i in unique(circ5$iddistrito)){
  s0 = circ5$distrtoelec[circ5$iddistrito == i]
  subid = c(subid, 1:length(s0))
  npob <- c(npob, rep(length(s0), length(s0)))
}
circ5$subid <- subid
circ5$npob <- npob
circ5$wk <- 1/(8/length(unique(circ5$iddistrito))*30/circ5$npob)
prob.c5 <- data.frame(circ5 %>% group_by(iddistrito) %>% count())
circ5$idun <- paste(circ5$iddistrito, circ5$subid, sep="-")

diseno.est.c1 <- function(){
  # seleccionamos 8 distritos de la circunscripción
  seleccion <- sample(unique(circ1$iddistrito), size=8)
  # aquí guardaremos el idunico de las casillas selecciones en la muestra
  idun <- c()
  # para cada distrito de los 8 seleccionados
  for (i in seleccion) {
    # vemos cuantas casillas le pertenecen
    iddist <- prob.c1 %>% filter(iddistrito == i)
    # y seleccionamos 30 elemento con un m.a.s.
    idun <- c(idun, paste(i, sample(iddist$n, size = 30), sep="-"))
  }
  #obtenemos los datos de la muestra
  muestra.c1 <- left_join(data.frame(idun), circ1, by = "idun")
  # recuperamos la muestra de esta circunscripción
  return(muestra.c1)
}
# de manera analoga para el resto de circunscripciones

#circunscripción 2
diseno.est.c2 <- function(){
  seleccion <- sample(unique(circ2$iddistrito), size=8)
  idun <- c()
  for (i in seleccion) {
    iddist <- prob.c2 %>% filter(iddistrito == i)
    idun <- c(idun, paste(i, sample(iddist$n, size = 30), sep="-"))
  }
  muestra.c2 <- left_join(data.frame(idun), circ2, by = "idun")

  return(muestra.c2)
}

#circunscripción 3
diseno.est.c3 <- function(){
  seleccion <- sample(unique(circ3$iddistrito), size=8)
  idun <- c()
  for (i in seleccion) {
    iddist <- prob.c3 %>% filter(iddistrito == i)
    idun <- c(idun, paste(i, sample(iddist$n, size = 30), sep="-"))
  }
}

```

```

}
muestra.c3 <- left_join(data.frame(idun), circ3, by = "idun")

return(muestra.c3)
}

#circunscrición 4
diseno.est.c4 <- function(){
  seleccion <- sample(unique(circ4$iddistrito), size=8)
  idun <- c()
  for (i in seleccion) {
    iddist <- prob.c4 %>% filter(iddistrito == i)
    idun <- c(idun, paste(i, sample(iddist$n, size = 30), sep="-"))
  }
  muestra.c4 <- left_join(data.frame(idun), circ4, by = "idun")

  return(muestra.c4)
}

#circunscrición 5
diseno.est.c5 <- function(){
  seleccion <- sample(unique(circ5$iddistrito), size=8)
  idun <- c()
  for (i in seleccion) {
    iddist <- prob.c5 %>% filter(iddistrito == i)
    idun <- c(idun, paste(i, sample(iddist$n, size = 30), sep="-"))
  }
  #obtenemos los datos de la muestra
  muestra.c5 <- left_join(data.frame(idun), circ5, by = "idun")

  return(muestra.c5)
}

diseno.circs <- function(){
  c1 <- diseno.est.c1()
  c2 <- diseno.est.c2()
  c3 <- diseno.est.c3()
  c4 <- diseno.est.c4()
  c5 <- diseno.est.c5()

  muestra.circs <- union(c1,c2)
  muestra.circs <- union(muestra.circs,c3)
  muestra.circs <- union(muestra.circs,c4)
  muestra.circs <- union(muestra.circs,c5)

  return(sum(muestra.circs$votos.favor*muestra.circs$wk)/sum(muestra.circs$totalvotos*muestra.circs$wk))
}

simulacion.circs <- replicate(1000, diseno.circs())
mean(simulacion.circs)*100

```

```
## [1] 44.26977854
```

```
# ahora calculamos los errores cuadráticos medios de cada diseño
```

```
# del diseño 1
```

```
mean((100*simulacion.mas - prct.votos)^2)
```

```
## [1] 0.2990906677
```

```
# del diseño 2
```

```
mean((100*simulacion.estr - prct.votos)^2)
```

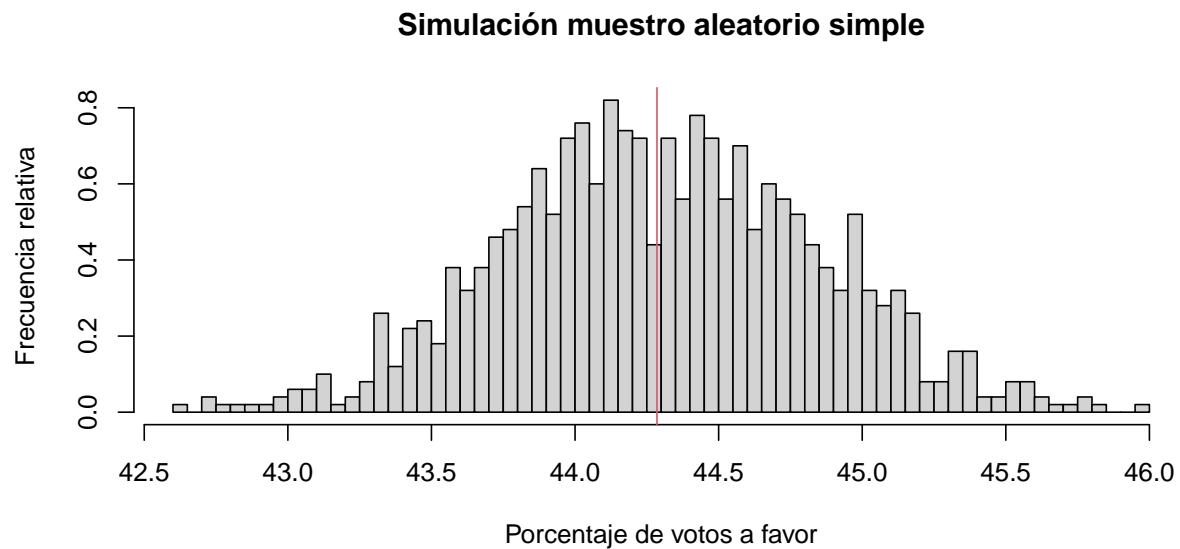
```
## [1] 0.1342909203
```

```
# del diseño 3
```

```
mean((100*simulacion.circs - prct.votos)^2)
```

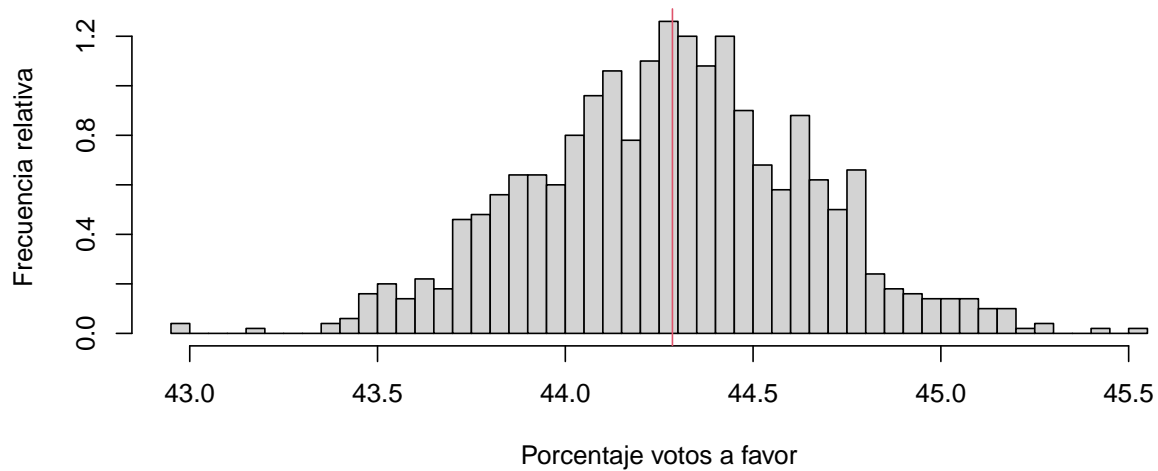
```
## [1] 2.842545481
```

```
hist(100*simulacion.mas, freq = F, breaks = 50,  
     main = "Simulación muestro aleatorio simple",  
     xlab = "Porcentaje de votos a favor", ylab = "Frecuencia relativa")  
abline(v = prct.votos, add = T, col=2)
```



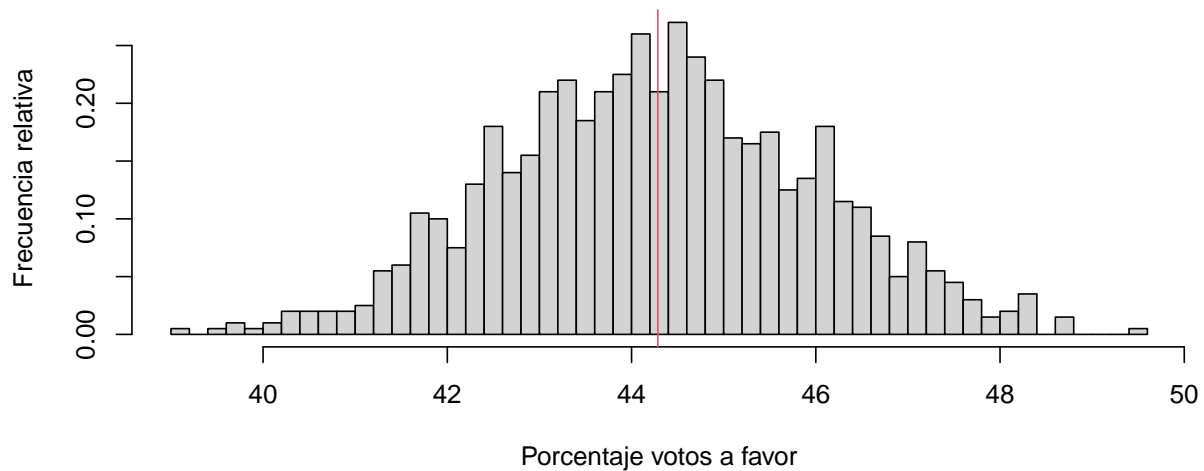
```
hist(100*simulacion.estr, freq = F, breaks = 50,  
     main = "Simulación muestro estratificado",  
     xlab = "Porcentaje votos a favor", ylab="Frecuencia relativa")  
abline(v = prct.votos, add = T, col=2)
```

### Simulación muestro estratificado



```
hist(100*simulacion.circo, freq = F, breaks = 50,  
     main = "Simulacion muestreo estratificado y bietápico",  
     xlab = "Porcentaje votos a favor", ylab = "Frecuencia relativa")  
abline(v = prct.votos, add = T, col=2)
```

### Simulacion muestreo estratificado y bietápico



De lo anterior vemos que el diseño que mejor funciona es el diseño dos, esto es porque a cada distrito le estamos tomando muestra, a diferencia del diseño tres donde solo le tomamos muestra a 40 de 300 distritos, esto hace que sea el peor, sin embargo en cuestiones de costo, puede llegar a ser el más económico, finalmente el diseño que esta en medio es el aleatorio simple.

### 3 Estimación básica de una encuesta con diseño multietápico

Considere la Encuesta Nacional de Vivienda (ENVI) 2020

<https://www.inegi.org.mx/programas/envi/2020/>

Suponga que será el encargado de generar los resultados básicos presentados en el tabulado llamado Cuadro 5.1, ver Figura 1 ([https://www.inegi.org.mx/contenidos/programas/envi/2020/tabulados/envi\\_2020\\_tema\\_05\\_xlsx.zip](https://www.inegi.org.mx/contenidos/programas/envi/2020/tabulados/envi_2020_tema_05_xlsx.zip))

En particular realice lo siguiente

- I. Describa brevemente el diseño de muestreo usado en la encuesta. Es decir, si es muestreo aleatorio simple, tiene estratificación, es por conglomerados, etc.

Se usó un diseño probabilístico, cuyo diseño muestral fue bietápico, donde en la primera etapa se usó un diseño de muestreo estratificado y en la segunda etapa un diseño de muestreo por conglomerados.

- I. Identifique las variables asociadas al diseño de muestreo que están presentes en la base de datos a usar (THOGAR en <https://www.inegi.org.mx/programas/envi/2020/#Microdatos>)
- II. Identifique la pregunta y variable asociada a la identificación de los Hogares con necesidad de rentar, comprar o construir una vivienda independiente de la que habitan.
- III. Con esta información, estime el número total de hogares y el porcentaje de hogares que tienen una necesidad de vivienda a nivel nacional y por entidad federativa.
- IV. Calcule intervalos de confianza para los parámetros estimados en el inciso anterior. Comente sobre los resultados obtenidos.
- V. **Punto extra opcional.** Considerando el porcentaje de hogares que tienen una necesidad de vivienda por entidad federativa, realice un mapa de calor (Geographic Heat Map) y comente los resultados.

Estimaciones puntuales.

Hogares por entidad federativa, según condición de tener necesidad de rentar, comprar o construir una vivienda independiente de la que habitan

Cuadro 5.1

Entidad federativa	Total de hogares	Condición de necesidad de vivienda					
		Sí		No		No especificado	
		Absolutos	Relativos	Absolutos	Relativos	Absolutos	Relativos
Estados Unidos Mexicanos	36 210 467	7 628 562	21.1	28 529 481	78.8	52 424	0.1
Aguascalientes	397 770	61 505	15.5	335 194	84.3	1 071	0.3
Baja California	1 156 528	281 959	24.4	872 108	75.4	2 461	0.2
Baja California Sur	246 920	64 373	26.1	182 547	73.9	0	0.0
Campeche	262 489	67 321	25.6	195 002	74.3	166	0.1
Coahuila de Zaragoza	913 569	163 939	17.9	747 788	81.9	1 842	0.2
Colima	234 272	40 317	17.2	193 611	82.6	344	0.1
Chiapas	1 460 368	400 920	27.5	1 057 487	72.4	1 961	0.1
Chihuahua	1 147 667	180 625	15.7	965 828	84.2	1 214	0.1
Ciudad de México	2 808 652	759 121	27.0	2 046 170	72.9	3 361	0.1

INEGI. Encuesta Nacional de Vivienda. ENVI 2020. Tabulados básicos. 2021

[ÍNDICE](#)

Error estándar.

Hogares por entidad federativa, según condición de tener necesidad de rentar, comprar o construir una vivienda independiente de la que habitan

Cuadro 5.1

Entidad federativa	Total de hogares	Condición de necesidad de vivienda					
		Sí		No		No especificado	
		Absolutos	Relativos	Absolutos	Relativos	Absolutos	Relativos
Estados Unidos Mexicanos	250 260.7	105 958.2	0.3	220 719.5	0.3	7 310.7	0.0
Aguascalientes	9 362.6	3 974.8	0.9	8 564.8	0.9	475.7	0.1
Baja California	30 667.9	19 337.4	1.5	27 919.5	1.5	1 244.3	0.1
Baja California Sur	9 038.1	5 173.9	1.6	6 449.4	1.6	0.0	0.0
Campeche	10 080.2	4 445.6	1.3	8 161.9	1.3	129.1	0.0
Coahuila de Zaragoza	20 581.8	10 105.6	1.0	19 628.4	1.1	934.0	0.1
Colima	8 612.2	3 021.6	1.1	7 590.5	1.1	201.2	0.1
Chiapas	65 259.8	24 523.2	1.2	51 431.1	1.2	1 401.0	0.1
Chihuahua	21 729.7	11 366.6	1.0	22 561.5	1.0	882.9	0.1
Ciudad de México	53 774.2	39 095.9	1.2	48 555.5	1.2	2 378.1	0.1

Figura 1: Parte del cuadro 5.1 de los resultados de la ENVI 2020

## SECCIÓN IIIa. DEMANDA Y NECESIDADES DE VIVIENDA

Pregunta	Nemónico	Tipo	Tamaño	Códigos Válidos	Concepto
(1)	(2)	(3)	(4)	(5)	(6)
3a.1 ¿Alguna de las personas de este hogar tiene necesidad o está planeando rentar, comprar o construir una vivienda, independientemente de la que habitan actualmente?	P3A1_1	Alfanumérico	1	1 2 9	Sí No No sabe

Figura 2: Pregunta de interés, demandas y necesidades de vivienda

```
# leemos la base de datos
setwd("D:/Notas/Muestreo/Exámen/examen02/Bases de datos")

thogar <- data.table::fread("THOGAR.csv")

# P3A1_1 es la variable a la que nos interesa replicar la estimación
# primero observamos que no hay N.A. por lo cual no es necesario hacer correcciones
thogar2 <- thogar %>% drop_na(P3A1_1)
length(thogar2$P3A1_1) == length(thogar$P3A1_1)

## [1] TRUE
```

```
# vemos las variables que usaremos en el diseño muestral
summary(thogar[, c("UPM_DIS", "EST_DIS", "FACTOR")])
```

```
##      UPM_DIS      EST_DIS      FACTOR
## Min.   : 1.000   Min.   : 1.0000   Min.   : 4.0000
## 1st Qu.:2254.000 1st Qu.:135.0000   1st Qu.: 254.0000
## Median :4291.500 Median :268.0000   Median : 473.0000
## Mean   :4306.808 Mean   :267.6786   Mean   : 640.0097
## 3rd Qu.:6329.000 3rd Qu.:397.0000   3rd Qu.: 788.0000
## Max.   :8301.000 Max.   :552.0000   Max.   :8610.0000
```

```
#los que si tienen necesidad
# segun la estructura de archivo
# el número 1 corresponde a si tienen necesidad
a <- sum(thogar$P3A1_1 == 1)
# el número 2 a los que no
b <- sum(thogar$P3A1_1 == 2)
# el número 3 a los que no especificaron
c <- sum(thogar$P3A1_1 == 9)
# efectivamente estos son los únicos resultados
sum(a + b + c) == length(thogar$P3A1_1)
```

```
## [1] TRUE
```

```
# añadimos estos valores al dataframe
# los que tienen necesida
thogar$si <- as.numeric(thogar$P3A1_1 == 1)
# los que no tienen necesidad
thogar$no <- as.numeric(thogar$P3A1_1 == 2)
# los que no especificaran
thogar$ne <- as.numeric(thogar$P3A1_1 == 9)
# ocuparemos un vector de unos para calcular el total
thogar$total <- 1

library(survey)
# en muchos casos sólo hay una upm en cada estrato, lo que
# dificulta la estimación de la varianza por esto ocupamos esta opción
options(survey.lonely.psu="adjust")
# usamos nest=TRUE ya que no hay seguridad de que las claves de las UPM son únicas

# definimos el diseño
dsg.envi <- svydesign(id=~UPM_DIS, strat=~EST_DIS, weight =~FACTOR,
                    data = thogar, nest=T)

# summary(dsg.envi)

# guardamos los resultados
# por nivel nacional
rel.nac <- svymean(~si + no + ne, dsg.envi)*100
abs.nac <- svytotal(~si + no + ne, dsg.envi)
total.nac <- svytotal(~total, dsg.envi)

# por entidad
```

```

rel.ent <- svyby(~si +no + ne,~ENT,design=dsg.envi, svymean)
abs.ent <- svyby(~si+ no + ne,~ENT,design=dsg.envi, svytotal)
total.ent <- svyby(~total, ~ENT,design=dsg.envi, svytotal)

#anexar los nombres de las entidades (para hacer las tablas)
entidades=c("AGU", "BCN", "BCS", "CAM", "COA", "COL", "CHP", "CHH", "CMX",
            "DUR", "GUA", "GRO", "HID", "JAL", "MEX", "MIC", "MOR", "NAY", "NLE",
            "OAX", "PUE", "QUE", "ROO", "SLP", "SIN", "SON", "TAB", "TAM", "TLA",
            "VER", "YUC", "ZAC")

#mostramos los resultados

total.nacdf <- as.data.frame(total.nac)
colnames(total.nacdf) <- c("total", "se")
row.names(total.nacdf) <- c("Nivel nacional")

total.entdf <- total.ent[, c("total", "se")]
row.names(total.entdf) <- entidades

dftotal <- union_all(total.nacdf, total.entdf)
colnames(dftotal) <- c("Total", "Error estándar")

kbl(dftotal, caption = "Total de hogares", booktabs = T) %>%
  kable_styling(latex_options = c("striped", "HOLD_position"))

```



Cuadro 2: Total de hogares

	Total	Error estándar
Nivel nacional	36210467	250260.723090
AGU	397770	9362.560429
BCN	1156528	30667.900587
BCS	246920	9038.064657
CAM	262489	10080.177465
COA	913569	20581.753539
COL	234272	8612.212124
CHP	1460368	65259.754453
CHH	1147667	21729.674891
CMX	2808652	53774.155230
DUR	507158	12451.339215
GUA	1663749	43662.953028
GRO	969487	29227.595625
HID	879538	25147.753568
JAL	2384946	60219.557229
MEX	4801185	154102.191785
MIC	1340554	48980.807753
MOR	593961	20541.243433
NAY	364784	10262.248527
NLE	1702725	47509.691132
OAX	1157915	53248.154463
PUE	1777565	61405.772431
QUE	680255	33271.499208
ROO	563868	22050.121615
SLP	790881	16199.544290
SIN	870827	19436.070996
SON	883105	26162.194677
TAB	694930	25398.594154
TAM	1058100	22381.889965
TLA	347967	9438.489241
VER	2402304	75073.973302
YUC	683612	23033.342983
ZAC	462816	16348.792929

```
abs.nacdf <- as.data.frame(abs.nac)
colnames(abs.nacdf) <- c("Absoluto", "Error estándar de absoluto")
rel.nacdf <- as.data.frame(rel.nac)
colnames(rel.nacdf) <- c("Relativo", "Error estándar de relativo")

n.nac <- cbind(abs.nacdf, rel.nacdf)
rownames(n.nac) <- c("Si", "No", "No sabe")

kbl(n.nac,
    caption = "Condición de hogares con necesidad de vivienda, nivel nacional",
    booktabs = T) %>% kable_styling(latex_options = c("striped", "HOLD_position"))
```

Cuadro 3: Condición de hogares con necesidad de vivienda, nivel nacional

	Absoluto	Error estándar de absoluto	Relativo	Error estándar de relativo
Si	7628562	105958.232900	21.0672842192	0.0025763631
No	28529481	220719.483131	78.7879399622	0.0025834577
No sabe	52424	7310.702217	0.1447758185	0.0002021010

SE significará Error estándar.

```
abs.entdf <- as.data.frame(abs.ent)
abs.entdf <- abs.entdf[c(-1)]
row.names(abs.entdf) <- entidades
colnames(abs.entdf) <- c("Si", "NO", "No sabe", "SE-Si", "SE-No",
                        "SE-No sabe")

kbl(abs.entdf,
     caption = "Condición de hogares con necesidad de vivienda, absoluto nivel entidad",
     booktabs = T) %>% kable_styling(latex_options = c("striped", "HOLD_position"))
```

Cuadro 4: Condición de hogares con necesidad de vivienda, absoluto nivel entidad

	Si	NO	No sabe	SE-Si	SE-No	SE-No sabe
AGU	61505	335194	1071	3974.823389	8564.766518	475.7383443
BCN	281959	872108	2461	19337.419329	27919.457494	1244.3162781
BCS	64373	182547	0	5173.874742	6449.383018	0.0000000
CAM	67321	195002	166	4445.643176	8161.858495	129.0968629
COA	163939	747788	1842	10105.564449	19628.361907	933.9657381
COL	40317	193611	344	3021.593299	7590.486296	201.1989066
CHP	400920	1057487	1961	24523.240800	51431.143341	1401.0014276
CHH	180625	965828	1214	11366.583223	22561.522406	882.9099614
CMX	759121	2046170	3361	39095.924432	48555.506365	2378.1255223
DUR	98252	408397	509	5787.977125	12180.408076	359.9236030
GUA	329132	1330149	4468	18324.931722	40831.519700	2044.6735681
GRO	303266	665626	595	16281.614662	24941.543274	595.0000000
HID	140927	737813	798	9331.489836	23459.519342	566.4379931
JAL	425646	1956569	2731	24713.322114	55730.575500	1931.9578153
MEX	883449	3909988	7748	56765.649370	136796.664492	3909.3186087
MIC	258117	1079968	2469	17019.088202	41651.859837	1348.1427966
MOR	109093	484504	364	6113.353616	20808.299349	364.0000000
NAY	81396	283203	185	5316.975140	8445.249760	185.0000000
NLE	197403	1503188	2134	17956.984197	45934.192957	1237.1960233
OAX	293325	864590	0	21077.694128	42503.529453	0.0000000
PUE	425198	1351558	809	28040.784324	54736.050440	809.0000000
QUE	141001	537897	1357	11145.868274	28477.477284	894.6390334
ROO	138127	424299	1442	9283.650409	20497.536257	663.2073582
SLP	123674	666332	875	8709.998494	14608.143397	623.5006014
SIN	222713	645528	2586	9220.053952	19008.975118	1185.5167650
SON	222551	659580	974	10768.713563	22849.422869	689.8594060
TAB	192553	501518	859	12898.737387	20890.774502	613.4011738
TAM	166002	890294	1804	12338.637054	18993.140649	1053.2302692
TLA	94156	253211	600	5211.782240	8215.154640	346.0028188
VER	536212	1860529	5563	31913.895347	62997.583973	2940.6361557
YUC	162218	520260	1134	9319.293354	21188.653382	641.7367061
ZAC	64071	398745	0	4393.677183	15235.175061	0.0000000

```

rel.entdf <- as.data.frame(rel.ent)
rel.entdf <- rel.entdf[c(-1)]
rel.entdf <- rel.entdf*100
row.names(rel.entdf) <- entidades
colnames(rel.entdf) <- c("Si", "NO", "No sabe", "SE-Si", "SE-No", "SE-No sabe")

kbl(rel.entdf,
  caption = "Condición de hogares con necesidad de vivienda,
relativo nivel entidad",
  booktabs = T) %>% kable_styling(latex_options = c("striped", "HOLD_position"))

```

Cuadro 5: Condición de hogares con necesidad de vivienda, relativo nivel entidad

	Si	NO	No sabe	SE-Si	SE-No	SE-No sabe
AGU	15.46245318	84.26829575	0.2692510747	0.9173243600	0.9131692986	0.1194644400
BCN	24.37978155	75.40742637	0.2127920811	1.4942211922	1.5010807182	0.1077664912
BCS	26.07038717	73.92961283	0.0000000000	1.5627555969	1.5627555969	0.0000000000
CAM	25.64716998	74.28958928	0.0632407453	1.3411186351	1.3407851506	0.0493019537
COA	17.94489524	81.85347795	0.2016268065	1.0390495679	1.0544507667	0.1023115714
COL	17.20948299	82.64367914	0.1468378637	1.1218973803	1.1228392561	0.0860145133
CHP	27.45335422	72.41236455	0.1342812223	1.2105917646	1.2190243948	0.0962142745
CHH	15.73845026	84.15576992	0.1057798124	0.9798720388	0.9821714014	0.0768822884
CMX	27.02794793	72.85238613	0.1196659465	1.2233667823	1.2256745014	0.0847155908
DUR	19.37305534	80.52658146	0.1003632004	1.1069716028	1.1087668608	0.0710157920
GUA	19.78255133	79.94889854	0.2685501238	1.0433200798	1.0429584614	0.1219704690
GRO	31.28107958	68.65754775	0.0613726641	1.4375165230	1.4402629578	0.0610727222
HID	16.02284381	83.88642674	0.0907294511	0.9930538975	0.9921498463	0.0644212040
JAL	17.84719654	82.03829353	0.1145099302	0.9626518326	0.9609447566	0.0808565847
MEX	18.40064484	81.43797833	0.1613768268	1.0401216644	1.0441224162	0.0818720661
MIC	19.25450224	80.56132017	0.1841775863	1.0366917379	1.0617585316	0.1007764780
MOR	18.36703083	81.57168568	0.0612834849	1.1521323133	1.1520266693	0.0612800484
NAY	22.31347866	77.63580640	0.0507149436	1.2116094402	1.2183039255	0.0508143613
NLE	11.59335771	88.28131378	0.1253285175	1.0141366495	1.0150154615	0.0727745498
OAX	25.33217032	74.66782968	0.0000000000	1.3738244479	1.3738244479	0.0000000000
PUE	23.92025046	76.03423785	0.0455116972	1.4095637862	1.4076675669	0.0456509261
QUE	20.72766830	79.07284768	0.1994840170	1.3677842207	1.3558211722	0.1303141533
ROO	24.49633602	75.24793037	0.2557336114	1.5660673312	1.5736470468	0.1179642123
SLP	15.63749793	84.25186596	0.1106361134	1.0016746996	0.9989220344	0.0789018376
SIN	25.57488456	74.12815634	0.2969590975	1.0507983930	1.0690384216	0.1358151994
SON	25.20096704	74.68874030	0.1102926606	1.0695561376	1.0708116277	0.0782073361
TAB	27.70825839	72.16813204	0.1236095722	1.5293479660	1.5338242974	0.0875441274
TAM	15.68868727	84.14081845	0.1704942822	1.0334478251	1.0310601940	0.0991233198
TLA	27.05888777	72.76868209	0.1724301442	1.3022474001	1.2957662277	0.0995935673
VER	22.32073876	77.44769188	0.2315693601	1.1029239966	1.1054092821	0.1223369010
YUC	23.72954249	76.10457394	0.1658835714	1.2799885187	1.2794173714	0.0942916662
ZAC	13.84373055	86.15626945	0.0000000000	0.8878609643	0.8878609643	0.0000000000

```
total.nac.ic <- as.data.frame(confint(total.nac))
rownames(total.nac.ic) <- "Nivel nacional"
total.ent.ic <- as.data.frame(confint(total.ent))
rownames(total.ent.ic) <- entidades

total.ic <- union_all(total.nac.ic, total.ent.ic)

kbl(total.ic,
     caption = "Intervalos confianza, total,
condición de hogares con necesidad de vivienda",
     booktabs = T) %>% kable_styling(latex_options = c("striped", "HOLD_position"))
```

Cuadro 6: Intervalos confianza, total, condición de hogares con necesidad de vivienda

	2.5 %	97.5 %
Nivel nacional	35719964.9960	36700969.0040
AGU	379419.7188	416120.2812
BCN	1096420.0194	1216635.9806
BCS	229205.7188	264634.2812
CAM	242732.2152	282245.7848
COA	873229.5043	953908.4957
COL	217392.3744	251151.6256
CHP	1332461.2316	1588274.7684
CHH	1105077.6198	1190256.3802
CMX	2703256.5924	2914047.4076
DUR	482753.8236	531562.1764
GUA	1578171.1846	1749326.8154
GRO	912201.9652	1026772.0348
HID	830249.3087	928826.6913
JAL	2266917.8367	2502974.1633
MEX	4499150.2542	5103219.7458
MIC	1244553.3809	1436554.6191
MOR	553700.9027	634221.0973
NAY	344670.3625	384897.6375
NLE	1609607.7165	1795842.2835
OAX	1053550.5350	1262279.4650
PUE	1657211.8976	1897918.1024
QUE	615044.0598	745465.9402
ROO	520650.5558	607085.4442
SLP	759130.4766	822631.5234
SIN	832733.0008	908920.9992
SON	831828.0407	934381.9593
TAB	645149.6702	744710.3298
TAM	1014232.3018	1101967.6982
TLA	329467.9010	366466.0990
VER	2255161.7162	2549446.2838
YUC	638467.4773	728756.5227
ZAC	430772.9547	494859.0453

Ahora presentamos los intervalos de conianza, suponiendo que el estimador sigue una distribución Normal.

```
abs.nac.ic <- as.data.frame(confint(abs.nac))
rownames(abs.nac.ic) <- c("Si", "No", "No sabe")

kbl(abs.nac.ic,
     caption = "Intervalos confianza, absoluto nacional,
condición de hogares con necesidad de vivienda",
booktabs = T) %>% kable_styling(latex_options = c("striped", "HOLD_position"))
```

Cuadro 7: Intervalos confianza, absoluto nacional, condición de hogares con necesidad de vivienda

	2.5 %	97.5 %
Si	7420887.67965	7836236.32035
No	28096878.76238	28962083.23762
No sabe	38095.28695	66752.71305

```
rel.nac.ic <- as.data.frame(confint(rel.nac))
rownames(rel.nac.ic) <- c("Si", "No", "No sabe")

kbl(rel.nac.ic,
    caption = "Intervalos confianza, relativo nacional,
condición de hogares con necesidad de vivienda",
    booktabs = T) %>% kable_styling(latex_options = c("striped", "HOLD_position"))
```

Cuadro 8: Intervalos confianza, relativo nacional, condición de hogares con necesidad de vivienda

	2.5 %	97.5 %
Si	21.0622346404	21.0723337980
No	78.7828764781	78.7930034463
No sabe	0.1443797079	0.1451719292

```
abs.ent.ic <- as.data.frame(confint(abs.ent))
abs.ent.ic.si <- abs.ent.ic[1:32, ]
rownames(abs.ent.ic.si) <- entidades
kbl(abs.ent.ic.si,
    caption = "Intervalos confianza, Sí, absoluto entidad,
condición de hogares con necesidad de vivienda",
    booktabs = T) %>% kable_styling(latex_options = c("striped", "HOLD_position"))
```

Cuadro 9: Intervalos confianza, Sí, absoluto entidad, condición de hogares con necesidad de vivienda

	2.5 %	97.5 %
AGU	53714.48931	69295.51069
BCN	244058.35456	319859.64544
BCS	54232.39184	74513.60816
CAM	58607.69949	76034.30051
COA	144132.45764	183745.54236
COL	34394.78596	46239.21404
CHP	352855.33125	448984.66875
CHH	158346.90626	202903.09374
CMX	682494.39617	835747.60383
DUR	86907.77329	109596.22671
GUA	293215.79381	365048.20619
GRO	271354.62165	335177.37835
HID	122637.61600	159216.38400
JAL	377208.77872	474083.22128
MEX	772190.37168	994707.62832
MIC	224760.20007	291473.79993
MOR	97111.04709	121074.95291
NAY	70974.92022	91817.07978
NLE	162207.95770	232598.04230
OAX	252013.47863	334636.52137
PUE	370239.07263	480156.92737
QUE	119155.49961	162846.50039
ROO	119931.37955	156322.62045
SLP	106602.71665	140745.28335
SIN	204642.02632	240783.97368
SON	201444.70926	243657.29074
TAB	167271.93927	217834.06073
TAM	141818.71576	190185.28424
TLA	83941.09452	104370.90548
VER	473661.91451	598762.08549
YUC	143952.52066	180483.47934
ZAC	55459.55096	72682.44904

```
abs.ent.ic.no <- abs.ent.ic[33:64, ]
row.names(abs.ent.ic.no) <- entidades
kbl(abs.ent.ic.no,
    caption = "Intervalos confianza, No, absoluto entidad,
condición de hogares con necesidad de vivienda",
    booktabs = T) %>% kable_styling(latex_options = c("striped", "HOLD_position"))
```

Cuadro 10: Intervalos confianza, No, absoluto entidad, condición de hogares con necesidad de vivienda

	2.5 %	97.5 %
AGU	318407.3661	351980.6339
BCN	817386.8688	926829.1312
BCS	169906.4416	195187.5584
CAM	179005.0513	210998.9487
COA	709317.1176	786258.8824
COL	178733.9202	208488.0798
CHP	956683.8114	1158290.1886
CHH	921608.2286	1010047.7714
CMX	1951002.9563	2141337.0437
DUR	384523.8389	432270.1611
GUA	1250120.6920	1410177.3080
GRO	616741.4735	714510.5265
HID	691833.1870	783792.8130
JAL	1847339.0792	2065798.9208
MEX	3641871.4644	4178104.5356
MIC	998331.8548	1161604.1452
MOR	443720.4827	525287.5173
NAY	266650.6146	299755.3854
NLE	1413158.6361	1593217.3639
OAX	781284.6131	947895.3869
PUE	1244277.3125	1458838.6875
QUE	482082.1702	593711.8298
ROO	384124.5672	464473.4328
SLP	637700.5651	694963.4349
SIN	608271.0934	682784.9066
SON	614795.9541	704364.0459
TAB	460572.8344	542463.1656
TAM	853068.1284	927519.8716
TLA	237109.5928	269312.4072
VER	1737056.0043	1984001.9957
YUC	478731.0025	561788.9975
ZAC	368884.6056	428605.3944

```
abs.ent.ic.ne <- abs.ent.ic[65:96, ]
row.names(abs.ent.ic.ne) <- entidades
# esto es ya que no hay números negativos
abs.ent.ic.ne$`2.5 %` <- pmax(0, abs.ent.ic.ne$`2.5 %`)
kbl(abs.ent.ic.ne,
    caption = "Intervalos confianza, No sabe, absoluto entidad,
condición de hogares con necesidad de vivienda",
    booktabs = T) %>% kable_styling(latex_options = c("striped", "HOLD_position"))
```



Cuadro 11: Intervalos confianza, No sabe, absoluto entidad, condición de hogares con necesidad de vivienda

	2.5 %	97.5 %
AGU	138.56997914	2003.4300209
BCN	22.18490950	4899.8150905
BCS	0.00000000	0.00000000
CAM	0.00000000	419.0252017
COA	11.46079047	3672.5392095
COL	0.00000000	738.3426106
CHP	0.00000000	4706.9123403
CHH	0.00000000	2944.4717260
CMX	0.00000000	8022.0403745
DUR	0.00000000	1214.4372991
GUA	460.51344642	8475.4865536
GRO	0.00000000	1761.1785708
HID	0.00000000	1908.1980659
JAL	0.00000000	6517.5677376
MEX	85.87632281	15410.1236772
MIC	0.00000000	5111.3113273
MOR	0.00000000	1077.4268904
NAY	0.00000000	547.5933371
NLE	0.00000000	4558.8596474
OAX	0.00000000	0.00000000
PUE	0.00000000	2394.6108635
QUE	0.00000000	3110.4602845
ROO	142.13746360	2741.8625364
SLP	0.00000000	2097.0387232
SIN	262.42983761	4909.5701624
SON	0.00000000	2326.0995901
TAB	0.00000000	2061.2442087
TAM	0.00000000	3868.2933951
TLA	0.00000000	1278.1530634
VER	0.00000000	11326.5409567
YUC	0.00000000	2391.7808316
ZAC	0.00000000	0.00000000

```

rel.ent.ic <- as.data.frame(confint(rel.ent)*100)
rel.ent.ic.si <- rel.ent.ic[1:32, ]
row.names(rel.ent.ic.si) <- entidades
kbl(rel.ent.ic.si,
    caption = "Intervalos confianza, Sí, relativo entidad,
condición de hogares con necesidad de vivienda",
    booktabs = T) %>% kable_styling(latex_options = c("striped", "HOLD_position"))

```

Cuadro 12: Intervalos confianza, Sí, relativo entidad, condición de hogares con necesidad de vivienda

	2.5 %	97.5 %
AGU	13.664530469	17.26037588
BCN	21.451161831	27.30840127
BCS	23.007442483	29.13333186
CAM	23.018625753	28.27571420
COA	15.908395509	19.98139497
COL	15.010604534	19.40836145
CHP	25.080637964	29.82607048
CHH	13.817936359	17.65896417
CMX	24.630193093	29.42570276
DUR	17.203430866	21.54267981
GUA	17.737681553	21.82742112
GRO	28.463598969	34.09856019
HID	14.076493937	17.96919369
JAL	15.960433619	19.73395946
MEX	16.362043839	20.43924584
MIC	17.222623774	21.28638071
MOR	16.108892992	20.62516867
NAY	19.938767795	24.68818953
NLE	9.605686398	13.58102901
OAX	22.639523885	28.02481676
PUE	21.157556200	26.68294471
QUE	18.046860490	23.40847611
ROO	21.426900455	27.56577159
SLP	13.674251594	17.60074426
SIN	23.515357558	27.63441157
SON	23.104675533	27.29725855
TAB	24.710791453	30.70572532
TAM	13.663166753	17.71420779
TLA	24.506529765	29.61124577
VER	20.159047446	24.48243007
YUC	21.220811092	26.23827389
ZAC	12.103555040	15.58390607

```
rel.ent.ic.no <- rel.ent.ic[33:64, ]
row.names(rel.ent.ic.no) <- entidades
kbl(rel.ent.ic.no,
    caption = "Intervalos confianza, No, relativo entidad,
condición de hogares con necesidad de vivienda",
    booktabs = T) %>% kable_styling(latex_options = c("striped", "HOLD_position"))
```

Cuadro 13: Intervalos confianza, No, relativo entidad, condición de hogares con necesidad de vivienda

	2.5 %	97.5 %
AGU	82.47851681	86.05807469
BCN	72.46536222	78.34949051
BCS	70.86666814	76.99255752
CAM	71.66169867	76.91747988
COA	79.78679243	83.92016348
COL	80.44295464	84.84440364
CHP	70.02312064	74.80160846
CHH	82.23074935	86.08079050
CMX	70.45010825	75.25466401
DUR	78.35343835	82.69972457
GUA	77.90473752	81.99305956
GRO	65.83468423	71.48041128
HID	81.94184877	85.83100470
JAL	80.15487642	83.92171064
MEX	79.39153600	83.48442066
MIC	78.48031169	82.64232865
MOR	79.31375490	83.82961646
NAY	75.24797458	80.02363821
NLE	86.29192003	90.27070752
OAX	71.97518324	77.36047612
PUE	73.27526011	78.79321558
QUE	76.41548701	81.73020835
ROO	72.16363883	78.33222190
SLP	82.29401475	86.20971717
SIN	72.03287953	76.22343314
SON	72.58998807	76.78749252
TAB	69.16189166	75.17437242
TAM	82.11997760	86.16165929
TLA	70.22902695	75.30833723
VER	75.28112950	79.61425426
YUC	73.59696197	78.61218591
ZAC	84.41609393	87.89644496

```

rel.ent.ic.ne <- rel.ent.ic[65:96, ]
row.names(rel.ent.ic.ne) <- entidades
# esto es ya que no hay números negativos
rel.ent.ic.ne$`2.5 %` <- pmax(0, rel.ent.ic.ne$`2.5 %`)
kbl(rel.ent.ic.ne,
    caption = "Intervalos confianza, No sabe, relativo entidad,
condición de hogares con necesidad de vivienda",
    booktabs = T) %>% kable_styling(latex_options = c("striped", "HOLD_position"))

```

Cuadro 14: Intervalos confianza, No sabe, relativo entidad, condición de hogares con necesidad de vivienda

	2.5 %	97.5 %
AGU	0.0351050750	0.5033970745
BCN	0.0015736396	0.4240105227
BCS	0.0000000000	0.0000000000
CAM	0.0000000000	0.1598707989
COA	0.0010998113	0.4021538017
COL	0.0000000000	0.3154232119
CHP	0.0000000000	0.3228577350
CHH	0.0000000000	0.2564663288
CMX	0.0000000000	0.2857054535
DUR	0.0000000000	0.2395515951
GUA	0.0294923973	0.5076078504
GRO	0.0000000000	0.1810730000
HID	0.0000000000	0.2169926908
JAL	0.0000000000	0.2729859241
MEX	0.0009105259	0.3218431276
MIC	0.0000000000	0.3816958536
MOR	0.0000000000	0.1813901728
NAY	0.0000000000	0.1503092617
NLE	0.0000000000	0.2679640141
OAX	0.0000000000	0.0000000000
PUE	0.0000000000	0.1349858683
QUE	0.0000000000	0.4548950641
ROO	0.0245280039	0.4869392190
SLP	0.0000000000	0.2652808734
SIN	0.0307661982	0.5631519968
SON	0.0000000000	0.2635762226
TAB	0.0000000000	0.2951929090
TAM	0.0000000000	0.3647724190
TLA	0.0000000000	0.3676299492
VER	0.0000000000	0.4713452800
YUC	0.0000000000	0.3506918412
ZAC	0.0000000000	0.0000000000