Tarea 3

Julio Quiñones

1. Repite el análisis hecho para AMLO pero ahora para Meade, Ricardo Anaya y "El Bronco". Es decir, agrega sus respectivas columnas y estima con una muestra de tamaño 2000 sus respectivos números de votos.

Importamos la librería "data.table" y cargamos la base de datos de las elecciones:

```
library(data.table)
data <- data.table::fread('BASE_PRESIDENTE.csv')
colnames(data)
```

```
##  [1] "CIRCUNSCRIPCION"    "ID_ESTADO"          "NOMBRE_ESTADO"
##  [4] "ID_DISTRITO"        "CABECERA_DISTRITAL" "ID_MUNICIPIO"
##  [7] "MUNICIPIO"          "SECCION"            "TIPO_CASILLA"
## [10] "ID_CASILLA"         "EXT_CONTIGUA"       "CASILLA"
## [13] "PAN"                "PRI"                "PRD"
## [16] "PVEM"               "PT"                 "MC"
## [19] "V19"                "MORENA"             "ES"
## [22] "PAN_PRD_MC"         "PAN_PRD"            "PAN_MC"
## [25] "PRD_MC"             "PRI_PVEM_NA"        "PRI_PVEM"
## [28] "PRI_NA"             "PVEM_NA"            "PT_MORENA_ES"
## [31] "PT_MORENA"          "PT_ES"              "MORENA_ES"
## [34] "CAND_IND1"          "CAND_IND2"          "NUM_VOTOS_CAN_NREG"
## [37] "NUM_VOTOS_NULOS"    "TOTAL_VOTOS"        "LISTA_NOMINAL"
## [40] "ESTATUS_ACTA"       "TRIBUNAL"           "OBSERVACIONES"
```

Lo siguiente que queremos es definir las coaliciones que se crearon en las elecciones del 2018 para poder hacer un conteo de los votos totales que tuvieron cada candidato a la presidencia. Para eso, definimos:

- Coalición Todos por México (Meade): PRI, PVEM, Nueva Alianza
- CoaliciónPor México Al Frente (Anaya): PAN, PRD, Movimiento Ciudadano
- Candidato Independiente ("El Bronco")
- Coalición Juntos Haremos Historia (AMLO): Morena, PT, PES

```
# Coalición Meade
data <- data[, MEADE := PRI+PVEM+PRI_PVEM_NA+PRI_PVEM+PRI_NA+PVEM_NA]
# Coalición Anaya
data <- data[, ANAYA := PAN+PRD+MC+PAN_PRD_MC+PAN_PRD+PAN_MC+PRD_MC]
# Coalición AMLO
data <- data[, AMLO := PT+MORENA+ES+PT_MORENA_ES+PT_MORENA+PT_ES+MORENA_ES]

# Limpiamos na's
data <- data[!is.na(data$MEADE),]
data <- data[!is.na(data$ANAYA),]
```

```r
data <- data[!is.na(data$CAND_IND1),]
data <- data[!is.na(data$AMLO),]

# Contamos los votos
votos_meade <- sum(data$MEADE)
votos_anaya <- sum(data$ANAYA)
votos_bronco <- sum(data$CAND_IND1)
votos_amlo <- sum(data$AMLO)

# Contamos los votos finales
data <- data[,TOTAL_FINAL := TOTAL_VOTOS-NUM_VOTOS_NULOS]
votos_finales <- sum(data$TOTAL_FINAL)

# Calculamos porcentajes
porcentual_meade <- votos_meade/votos_finales*100
porcentual_anaya <- votos_anaya/votos_finales*100
porcentual_bronco <- votos_bronco/votos_finales*100
porcentual_amlo <- votos_amlo/votos_finales*100

# Check
porcentual_amlo+porcentual_anaya+porcentual_bronco+porcentual_meade
```

```
## [1] 93.66302
```

Considerando que nuestra población es $\{X_i\}_{i=1}^N$, entonces tomamos una muestra $\{X'_j\}_{j=1}^n$ tal que $n < N$. Con esta muestra entonces queremos crear un estimador de alguna estadística de la población. Tomando inspiración de la clase, proponemos como estimador lo siguiente:

$$estimador = \Sigma_{j=1}^n X'_j \cdot \frac{1}{n} \cdot N \ .$$

La población que vamos a utilizar será el vector de votos totales por candidato por casilla. En nuestro ejercicio, este vector es la respectiva columna que hemos construido para cada candidato. Recordemos que el tamaño de nuestra muestra es de 2000, i.e. $n = 2000$.

```r
# AMLO
# Fijamos <replace = FALSE> para evitar que salga la misma casilla dos veces.
N <- length(data$AMLO)
n <- 2000
casillas <- sample.int(N, size=n, replace=FALSE)
```

Una vez que tenemos el vector de índices aleatorios de casillas obtenemos las respectivas muestras como:

```r
# Sub-muestra AMLO
X_meade <- data$MEADE[casillas]
X_anaya <- data$ANAYA[casillas]
X_bronco <- data$CAND_IND1[casillas]
X_amlo <- data$AMLO[casillas]
```

Finalmente, calculamos las estimaciones de votos totales para cada candidato usando el estimador que hemos propuesto:

```r
estimacion_meade <- N * mean(X_meade)
estimacion_anaya <- N * mean(X_anaya)
estimacion_bronco <- N * mean(X_bronco)
estimacion_amlo <- N * mean(X_amlo)

cat('\n', "Votos Meade estimados - reales: ", estimacion_meade, "-", votos_meade,
    '\n', "Votos Anaya estimados - reales: ", estimacion_anaya, "-", votos_anaya,
    '\n', "Votos Bronco estimados - reales: ", estimacion_bronco, "-", votos_bronco,
    '\n', "Votos AMLO estimados - reales: ", estimacion_amlo, "-", votos_amlo
    )
```

```
##
##  Votos Meade estimados - reales:  8648279 - 8795608
##  Votos Anaya estimados - reales:  12525264 - 12610120
##  Votos Bronco estimados - reales:  29175.03 - 32832
##  Votos AMLO estimados - reales:  30204156 - 30113483
```

Comparando estimaciones con los valores reales, tenemos que:

```r
cat('\n', "Error % para Meade: ", abs(1 - estimacion_meade/votos_meade)*100,
    '\n', "Error % para Anaya: ", abs(1 - estimacion_anaya/votos_anaya)*100,
    '\n', "Error % para Bronco: ", abs(1 - estimacion_bronco/votos_bronco)*100,
    '\n', "Error % para AMLO: ", abs(1 - estimacion_amlo/votos_amlo)*100
    )
```

```
##
##  Error % para Meade:  1.675031
##  Error % para Anaya:  0.6729207
##  Error % para Bronco:  11.13843
##  Error % para AMLO:  0.3011032
```

2. Muestra que el estimador que propusiste es insesgado. Para ello calcula la esperanza de éste.

Calculemos la esperanza de nuestro estimador:

$$\begin{aligned} E(estimador) &= E(\Sigma_{j=1}^{n} X'_j \cdot \frac{1}{n} \cdot N) \\ &= \frac{N}{n} E(\Sigma_{j=1}^{n} X'_j) \ , \end{aligned}$$

y dado que tenemos una muestra aleatoria, por definición sabemos que está compuesta por variables aleatorias independientes e idénticamente distribuidas, lo que nos dice que la esperanza de cada una de las X'_j es la misma, por lo que al usar la linealidad como operador que es la esperanza, tenemos:

$$\begin{aligned} E(estimador) &= \frac{N}{n} \Sigma_{j=1}^{n} E(X'_j) \\ &= \frac{N}{n} n \cdot E(X') \\ &= N \cdot \Sigma_{i=1}^{N} X_i \cdot P(X_i) \\ &= N \Sigma_{i=1}^{N} X_i \frac{1}{N} \\ &= \Sigma_{i=1}^{N} X_i \ . \end{aligned}$$

Así, como la esperanza del estimador es igual a la estadística que estamos buscando obtener de la muestra, decimos que es insesgado.

3. Calcula la varianza del estimador de una muestra de tamaño 1.

Calculemos la varianza del estimador de tamaño 1. Primero tomemos en cuenta que la esperanza de este estimador de tamaño 1, es la misma muestra. Así, desde la definición tenemos que:

$$\begin{aligned} Var(estimador) &= Var(X'_j \cdot N) \\ &= \Sigma_{j=1}^{n=1} \left((X'_j \cdot N) - E(X'_j \cdot N)\right)^2 \ , \end{aligned}$$

como hemos probado anteriormente, la esperanza es el promedio de la muestra, el cual es simplemente la misma muestra, entoncs seguimos que:

$$\begin{aligned} Var(estimador) &= \left(X'_j \cdot N - E(X'_j \cdot N)\right)^2 \\ &= \left(X'_j \cdot N - N \cdot E(X'_j)\right)^2 \\ &= \left(X'_j \cdot N - N \cdot X'_j \cdot P(X'_j)\right)^2 \\ &= \left(X'_j \cdot N - X'_j\right)^2 \\ &= \left(X'_j \cdot (N-1)\right)^2 \\ &= {X'_j}^2 \cdot (N-1)^2 \ . \end{aligned}$$