Examén de EA

DIMAS RAMÍREZ LUIS DANIEL #315334731

8/11/2020

I) Responda las siguientes cuestiones

a) ¿Cuál es la diferencia entre una distribución muestral y una distribución de probabilidad?

La distribución muestral resulta de considerar todas las muestras posibles que pueden ser tomadas de una población.

La distribución de probabilidad surge de la muestra tomada de la población y están identicamente distribuidas.

b) ¿Cuál es la diferencia entre error estándar y el error de muestreo y cómo se relacionan? ^2 El error estandar o también conocido como "desviación estandár" provienen de la varianza $\square = \sqrt{\square^{2}}$, nos da información de la dispersión de los datos.

El error estándar es un error estimado que es probable que cometamos en nuestra estimación.

No se relacionan de manera directa.

c) Si se toma una muestra aleatoria ¿ésta es representativa?

Depende de si es no es representativa. Si es muy pequeña, no sera representativa y depende de cual sea el tamaño de nuestra población.

d) ¿Cuáles son las características de un buen estimador? Describa cada una de ellas.

Insesgado: Cuando se obtiene una estimación puntual de un parametro

Eficiente: que tenga una varianza minima, lo cuál indicara que el la dispersión de los datos es poca

Suficiente: quiere decir que el estimador sí toma en cuenta toda la información de la muestra

e) ¿Por qué muestreamos?

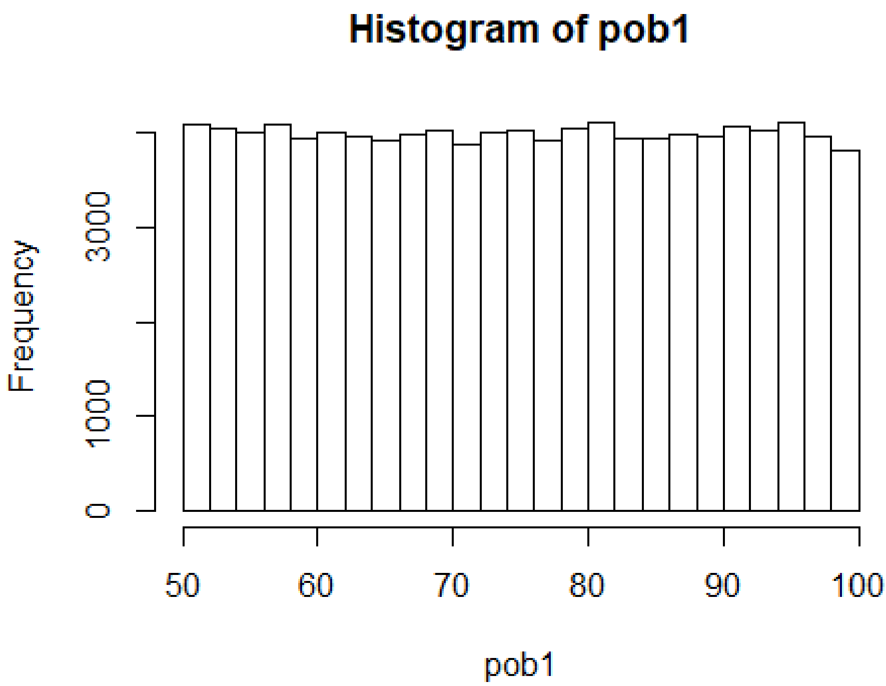
La mayoría de las poblaciones que analizamos suelen ser demasiado grandes como para ser analizadas completamente (censar), o porque no se cuentan con los recursos suficientes.

II) Responda los siguientes ejercicios

Ejercicio 1

Para la siguiente población calcule el tamaño de la muesra con un error de muestreo $\square = 5$ y un coeficiente de confianza $\square\square = 95$ de una población infinita, estime $\square^{2}$ con un muestreo piloto con $\square = 30$ elementos, consideres una población rectangular *runif* con *100,000* elementos $\square = 50$ y $\square = 100$

```
a <- 50
b <- 100
pob1 <- runif(100000, a, b)
hist(pob1)
```



$$\square = \frac{\square^{2}_{\square/2}\square^{2}}{\square^{2}}$$

Tomando n=30 del muestreo piloto

```
mp1 <- 30
n_piloto1 <- sample(pob1, mp1)
n_piloto1

##  [1] 76.11127 91.58432 61.23784 57.00496 80.41559 64.42413 90.55452
74.65262
##  [9] 53.23160 77.34241 60.46060 58.34636 69.00982 80.39129 69.55863
```

```
66.52111
## [17] 85.65045 81.21632 50.61740 60.44547 70.87209 99.37649 59.20212
98.02271
## [25] 84.30224 82.60295 68.72045 52.97742 95.45344 85.39096
```

```
s2piloto1 <- round(var(n_piloto1), 2)
s2piloto1

## [1] 199.52
```

- $S^2 = 199.52$

```
B1 <- 5
alpha1 <-  qnorm(0.975)
n1 <- alpha1^2*s2piloto1/B1^2
n1c <- ceiling(n1)
```

- $n = 31$

Manteniendo el error $B = 5$ calcule el nuevo coeficiente de confianza y los límites inferior LI y límite superior LS.

El nuevo CC estara definido por:

$$Z_{\frac{\alpha}{2}} = \sqrt{\frac{nB^2}{S^2}}$$

```
cc1 <- (n1c*25/s2piloto1)^0.5
cc1

## [1] 1.970868
```

- $CC = 97.61\%$

```
x1 <- mean(n_piloto1)
x1

## [1] 73.52325

li1 <- x1-5
ls1 <- x1+5
```

- $LI = 68.5232536$

- $LS = 78.5232536$

Manteniendo la confianza en CC = 95% calcule el nuevo error y construya los intervalos de confianza.

El nuevo error estara definido como:

$$B = \sqrt{\frac{z_{\frac{\alpha}{2}} s^2}{n}}$$

```
bnuevo1 <- (qnorm(0.975)*s2piloto1/n1c)^0.5
bnuevo1

## [1] 3.551701
```

- $B = 3.5517012$

```
li1nuevo <- x1-bnuevo1
ls1nuevo <- x1+bnuevo1
```

- $LI = 69.9715525$
- $LS = 77.0749548$

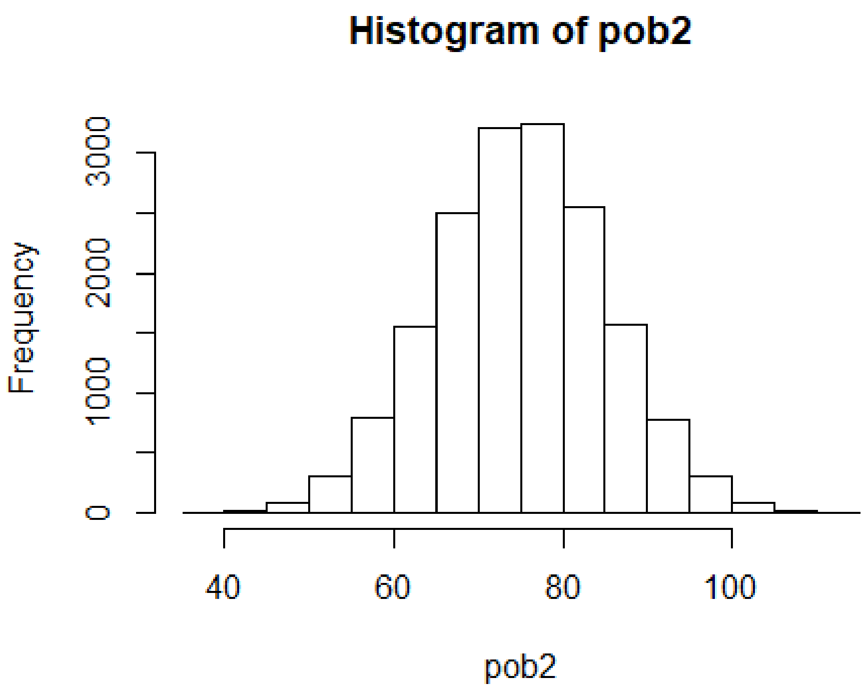
$$IC = \bar{x} + -z_{\alpha/2} s/\sqrt{n}$$

- $IC = $ 73.5232536 $\pm$ 3.5517012

Ejercicio 2

Para una población con *17,000* elementos calcular el tamaño de la muestra, la población tiene una media $\mu = 75$ y una $\sigma = 10$. Utilice un muestreo piloto con n = 30 elementos para calcular s^2 y obtener el tamaño de la muestra para un nivel de error B = 4 y un CC = 97%.

```
mediapob2 <- 75
sd2 <- 10
pob2 <- rnorm(17000, mediapob2, sd2)
hist(pob2)
```




```
s22 <- var(pob2)

mp2 <- sample(pob2, 30)
s2piloto2 <- var(mp2)
```

- $\sigma^2 = 101.6713948$

$$n = \frac{Z_{\alpha/2}^2 \sigma^2}{B^2}$$

```
n2 <- qnorm(0.985)*s2piloto2/16
n2r <- ceiling(n2)
```

- $n = 14$ ***El tamaño de la muestra es muy pequeño***

Manteniendo el error B = 4 calcule el nuevo coeficiente de confianza y los límites inferior LI y límite superior LS.

El nuevo CC estara definido por:

$$Z_{\frac{\alpha}{2}} = \sqrt{\frac{n B^2}{\sigma^2}}$$

```
aplha2 <- (n2r*16/s2piloto2)^0.5
aplha2

## [1] 1.524521
```

- CC 93.45%

```r
x2 <- mean(mp2)
li2 <- x2 - 4
ls2 <- x2+4
```

- $LI = 72.7485347$

- $LS = 80.7485347$

Manteniendo la confianza en CC = 97% calcule el nuevo error B y construya los intervalos de confianza.

El nuevo error estara definido como:

$$B = \sqrt{\frac{Z_{\frac{\alpha}{2}}S^{2}}{n}}$$

```r
b2 <- (qnorm(0.985)*s2piloto2/n2r)^0.5
b2

## [1] 3.865141

li2nuevo <- x2-b2
ls2nuevo <- x2+b2
```

- $B = 3.8651413$

- $LI = 72.8833934$

- $LS = 80.613676$

Ejercicio 3

Calcule el tamaño de la muestra para las proporciones de una población finita de 30,000 elementos con un error del B1 = 1% ,B2 = 2% y B3 = 5% y un CC = 93%

$$n = \frac{N\hat{p}\hat{q}}{(N-1)D^{2} + \hat{p}\hat{q}}$$

$$D = \frac{B}{Z_{\frac{\alpha}{2}}}$$

```r
N <- 30000

p3 <- 0.5 #proporcion que calcula la varinaza más alta
pq3 <- p3*(1-p3)

z3 <- qnorm(.965)
```

```r
b13 <- 0.01
b23 <- 0.02
b33 <- 0.05

d13 <- b13/z3
d23 <- b23/z3
d33 <- b33/z3

n13 <- (pq3*N)/(((N-1)*d13^2)+pq3)
n23 <- pq3*N/((N-1)*d23^2+pq3)
n33 <- pq3*N/((N-1)*d33^2+pq3)
```

- $n_1 = 6444.6153051$
- $n_2 = 1920.5904995$
- $n_3 = 324.7588873$

Ejercicio 4

Para una población con $\sigma_1^2 = 40$, $\sigma_2^2 = 20$ y $\sigma_3^2 = 10$ y
$N_1 = 20{,}000$, $N_2 = 15{,}000$ y $N_3 = 12{,}000$ con un error de B = 3 y un CC = 93%.
Calcular el tamaño de la muestra total y el que corresponde a cada estrato.

$$n = \frac{\sigma^2 N}{(N-1)D^2 + \sigma^2}$$

$$D = \frac{B}{z_{\frac{\alpha}{2}}}$$

```r
D4 <- 3/ qnorm(0.965)

s14 <- 40
s24 <- 20
s34 <- 10

N14 <- 20000
N24 <- 15000
N34 <- 12000

n14 <- s14*N14/((D4^2*(N14-1))+s14)
n24 <- s24*N24/((D4^2*(N24-1))+s24)
n34 <- s34*N34/((D4^2*(N34-1))+s34)

n14r <- ceiling(n14)
n24r <- ceiling(n24)
n34r <- ceiling(n34)

n14r
```

```
## [1] 15
n24r
## [1] 8
n34r
## [1] 4
ntotal4 <- sum(n14r, n24r, n34r)
```

- $n_1 = 15$
- $n_2 = 8$
- $n_3 = 4$
- $n_{total} = 27$

Ejercicio 5

Para una población con $\sigma_1^2 = 45, \sigma_2^2 = 15$ y $\sigma_3^2 = 12$ y $N_1 = 20{,}000, N_2 = 15{,}000$ y $N_3 = 12{,}000$ con un error de B = 2 y un CC = 97%, si los costos $c_1 = 1, c_2 = 2$ y $c_3 = 0.5$. Calcular el tamaño de la muestra total y el que corresponde a cada estrato.

```
b5 <- 2
z5 <- qnorm(0.985)
varianza5 <- c(45, 15, 12)
costosi5 <- c(1,2,0.5)
N5 <- c(20000,15000,12000)

D5 <- b5^2/z5^2

nm5 <-
sum(N5*varianza5^.5*costosi5^.5)*sum(N5*varianza5^.5/costosi5^.5)/(sum(N5
)^2*D5+sum(N5*varianza5))
nmc5<- ceiling(nm5)
est5<-nm5*(N5*varianza5^.5/costosi5^.5)/sum(N5*varianza5^.5/costosi5^.5)
etrc5<-ceiling(est5)
```

Respuesta 1: El n total es de 31

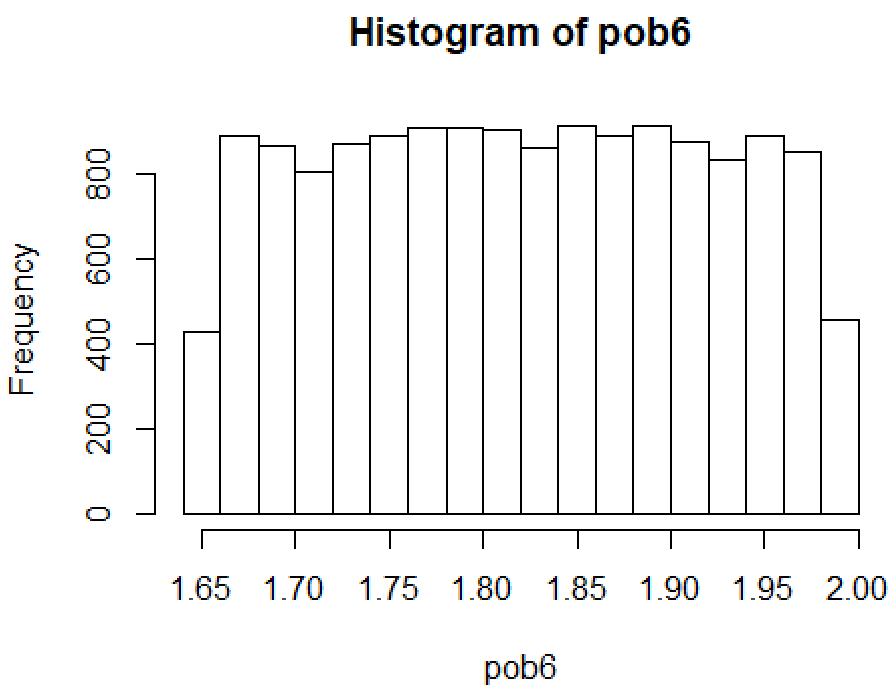
Respuesta 2: El numero de muestras por cada estrato es de 18, 6, 8

Ejercicio 6

Donald Trump sostenía durante campaña que sus seguidores son mejores que los seguidores de Joe Biden en todos los sentidos, incluso en su estatura, afirmando que la altura promedio de los republicanos es de 1.90 metros. Considerando una población de 15,000 republicanos con parámetros a = 1.65 y b = 1.99 y CC = 95% ¿Se puede afirmar que es correcto el comentario de Donald Trump? Mostrar su procedimiento.

$$CC = 95\%$$

```
pob6 <- runif(15000, 1.65, 1.99)
hist(pob6)
```




```
mu06 <-1.90
```

- $H_0: \mu = 1.90$
- $H_a: \mu \neq 1.90$

Tomando una muestras cualquiera

```
ncualquiera6 <- sample(pob6, 1000)
```

Calculando $\bar{x}$ de esa muestra:

```
x6 <- mean(ncualquiera6)
sd6 <- sd(ncualquiera6)
```

De esta muestra podemos notar que la media muestral es 1.8230089

Construyendo los intervalos de confianza para saber si existe evidencia suficiente para rechazar la hipótesis nula es:

$$LIC = \mu_0 - Z_{\alpha/2}\sigma/\sqrt{n}$$

$$LSC = \mu_0 + z_{\alpha/2}\sigma/\sqrt{n}$$

```r
li6 <-  mu06- ((qnorm(0.975)*sd6)/1000^0.5)
ls6 <- mu06 + ((qnorm(0.975)*sd6)/1000^0.5)

li6r <- round(li6, 2)
ls6r <- round(ls6, 2)
```

- $LI = 1.89$
- $LS = 1.91$
- $\bar{x} = 1.8230089$

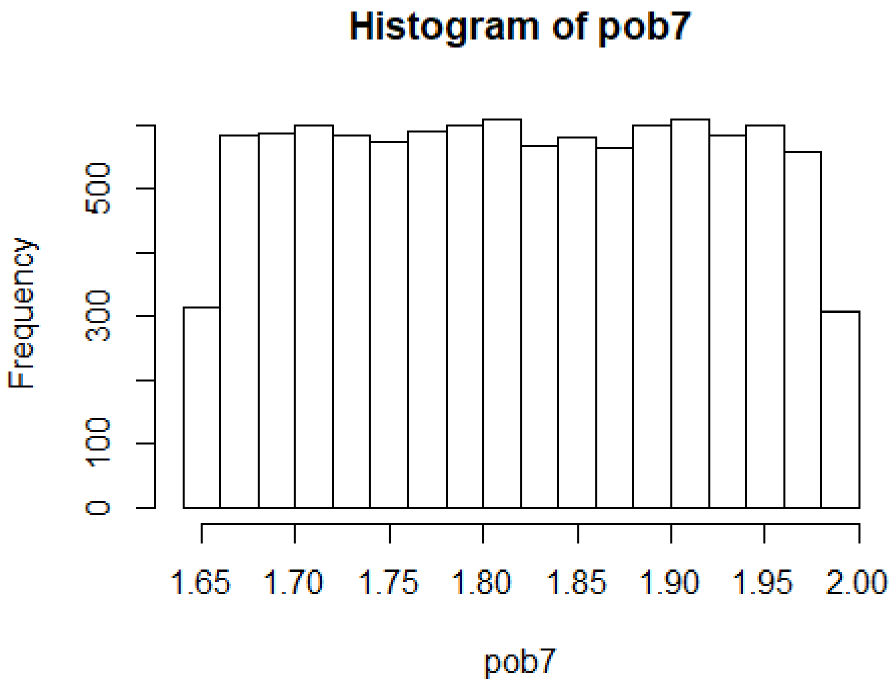
Como el LI es 1.89 y el LS es 1.91 y la media muestral es $\bar{x} = 1.8230089$; y no se encuentra dentro de los límites ***no existe suficiente evidencia para afirmar la*** $\mu_0 = 1.90$ ***del Sr Donald Trump***

Ejercicio 7

Después de sus controversiales comentarios sobre la estatura de sus votantes, el equipo de campaña de Donald Trump decidió destinar recursos para determinar la estatura promedio de los demócratas y demostrar que efectivamente son de menor estatura que los republicanos. Se tiene una población de 10,000 demócratas con parámetros a = 1.65 y b = 1.99, con un coeficiente de confianza CC = 96% y un error de B = 0.075. Determine el tamaño de muestra que se debe extraer de la población y cúal sería el intervalo de confianza en el que se encuentra la estatura promedio de los demócratas.

$$CC = 96\% \; y \; B = 0.075$$

```r
pob7 <- runif(10000, 1.65, 1.99)
hist(pob7)
```



Para determinar n primero sacare una muestra pilto de n=1,000

```
npiloto7 <- sample(pob7, 1000)
s27 <- var(npiloto7)
s27 #varianza muestral

## [1] 0.009984656

x7 <- mean(npiloto7)
x7 #media muestral

## [1] 1.820683

sd7 <- sd(npiloto7)
sd7 #desvacion estandar muestral

## [1] 0.09992325
```

$$n = \frac{z_{\alpha/2}^2 \sigma^2}{b^2}$$

Calculando n…

```
b7 <- 0.075
alpha7 <- qnorm(0.98)

n7 <- alpha7^2*s27/b7^2
n7r <- ceiling(n7)
```

- $\sigma_r = 8$

Construyendo los límites inferiores y superiores…

$$LIC = \bar{x} - z_{\alpha/2}\sigma/\sqrt{n}$$

$$LSC = \bar{x} + z_{\alpha/2}\sigma/\sqrt{n}$$

```
li7 <- x7 - ((qnorm(0.98)* sd7)/1000^0.5)
ls7 <- x7 + ((qnorm(0.98)* sd7)/1000^0.5)

li7r <- round(li7, 2)
ls7r <- round(ls7, 2)
```

- $LI = 1.81$
- $LS = 1.83$

Por lo tanto podemos concluir que la estatura promedio de los democratas y de los republicanos es muy similar

Ejercicio 8

El equipo de campaña de Trump sabe que los comentarios de Donald Trump han mermado su aprobación a unos días de las elecciones y quieren saber en que proporción. Para hacerlo deciden aplicar una encuesta vía internet para conocer su aprobación en una escala del 1 al 100, siendo 100 la mejor calificación posible. Se tiene el correo de 350,000 personas de Texas, 550,000 personas de Wisconsin, 200,000 personas de Arkansas y 400,000 personas de California.

Por ejercicios similares se determinó que la varianzas en las calificaciones para cada estado son de $\sigma^2_{Tex} = 20, \sigma^2_{Wis} = 36, \sigma^2_{Ark} = 13, \sigma^2_{Cal} = 57$, respectivamente. Por la importancia de los votos de cada estado en el colegio electoral se tienen las siguientes fracciones asignadas wTex = 25%, wWis = 25%, wArk = 10% y wCal = 40%. Con un nivel de significancia de α = 5% y un error de B = 1.5 ¿Cuál sería el tamaño de la muestra a la que tienen que aplicar la encuesta y cuántas personas de cada estado?

$$n = \frac{\sum \frac{N_i^2 \sigma_i^2}{w_i}}{N^2 D + \sum N_i \sigma_i^2}$$

$$D = \frac{B}{z^2_{\alpha/2}}$$

```
b8 <- 1.5
alfa8 <- qnorm(0.975)
varianza8 <- c(20, 36, 13, 57)
costos8 <- c(0.25, 0.25, 0.1, 0.4)
N8 <- c(350000,550000,200000,400000)
```

```
d8 <- b8/alfa8

ntotal8 <-
sum((N8^2*varianza8/costos8)/(sum(N8)^2*d8^2+sum(N8*varianza8)))
ntotal18r <- ceiling(ntotal8)

nestra8 <- ceiling(ntotal18r*costos8)
```

- $\square_{\square} = 62$
- $\square_{\square\square\square}, \square_{\square\square\square}, \square_{\square\square\square}, \square_{\square\square\square} = 16, 16, 7, 25$