title: “Muestreo Irrestricto” author: “DIMAS” date: “21/10/2020” output: html_document

#Ejemplo de muestreo irrestricto

La varianza poblacional es

$$\sigma^2 = 20^2$$

Se desea estimar el tamaño de la muestra de una población $N = 1000$ elementos, con un error de $B =$\$ y una confianza del $CC = 95\%$ si se sabe que la varianza pobalcional es $\sigma^2 = 20^2$

```
N <- c(1000)
Varpob <- c(20^2)
alpha <- 0.05
B <- 4
D <- B/qnorm(0.975)

n<- N*Varpob/(D^2*(N-1)+ Varpob)
```

El tamaño de la muestra es 88

#porque el nivel de significancia es 95%, el valor de alpha es 5% pero el valor de alpha medios es 2.5% por eso se le suma al 95% solamente un 2.5% y nos da qnorm(0.975)

Ahora se quiere saber la diferencia si la población fuera finita, por lo tanto

```
n_2 <- qnorm(.975)^2*Varpob/B^2
```

El tamaño de la muestra si la población fuera infinita es 97

#Ejemplo 2: Muestreo irrestricto

Suponga que se les asignan a diez personas calcular el tamaño de la muestra de una población que se distribuye de forma normal, suponga también que conoce la media poblacional la cual es de 70 y su varianza poblacional la cual es de 100. El tamaño de la población es de 10,000 elementos y el límite para el error es de 3. 3/4Cuáles serían algunos de los posibles tamaños de muestra que obtendrían estas diez personas? Suponga que cada uno de ellos toma un muestreo piloto de 30 elementos de forma aleatoria para estimar la varianza poblacional. Tome un nivel de signicancia del 5 %. #Ya que no hay varianza, se toma una muestra poblacional **Nota: en este caso no se genera la población, dado que se supondra que los parametros son los mismos que ya se establecieron para μ y para $\sigma^2 = 100$**

```
#Estudiante uno (rnorm) random numbers with normal distribution
m_1 <- rnorm(30, mean=70, sd=100^0.5)
#Estudiante dos
m_2 <- rnorm(30, mean=70, sd=100^0.5)
#Estudiante tres
m_3 <- rnorm(30, mean=70, sd=100^0.5)
#Estudiante cuatro
m_4 <- rnorm(30, mean=70, sd=100^0.5)
#Estudiante cinco
m_5 <- rnorm(30, mean=70, sd=100^0.5)
#
m_6 <- rnorm(30, mean=70, sd=100^0.5)

m_7 <- rnorm(30, mean=70, sd=100^0.5)
```

```r
m_8 <- rnorm(30, mean=70, sd=100^0.5)

m_9 <- rnorm(30, mean=70, sd=100^0.5)

m_10 <- rnorm(30, mean=70, sd=100^0.5)

matriz_m <- data.frame(m_1, m_2, m_3, m_4,m_5, m_6, m_7, m_8, m_9, m_10) #matriz de muestras
#View(matrizdemuestras)

matriz_m
```

```
##         m_1      m_2      m_3      m_4      m_5      m_6      m_7      m_8
## 1  81.04215 64.62808 52.03280 77.52497 50.59788 69.73135 51.60924 76.21589
## 2  73.57092 71.02720 86.76583 47.35072 71.23248 58.39781 76.72049 64.22085
## 3  86.91661 61.66227 84.11537 62.67563 51.26985 63.83920 70.63981 73.33270
## 4  65.59465 76.19778 69.38324 61.53647 92.24554 87.02503 67.17923 79.99977
## 5  78.53832 69.62602 86.43418 73.59592 51.83810 66.50718 57.58978 74.01741
## 6  70.12188 75.44641 86.97824 67.71249 73.57099 71.43512 47.97188 74.10374
## 7  74.24725 87.58513 83.50003 63.56548 60.14873 82.31926 76.14911 69.56745
## 8  62.28348 91.51246 70.86344 74.74876 68.92293 65.58181 78.64556 59.42914
## 9  71.68882 57.57591 81.37266 79.18433 50.93270 70.77528 53.26952 56.96793
## 10 74.37423 71.35360 66.15637 67.59796 86.89418 70.30750 56.30756 68.28159
## 11 69.57774 72.67006 81.27242 72.97776 49.87670 70.70947 63.96107 86.76785
## 12 57.53995 66.56862 74.83816 85.09495 86.87276 74.76082 68.65603 68.43672
## 13 82.92616 74.65528 76.63229 53.39912 56.99807 61.05439 66.37100 70.65545
## 14 72.93953 81.87281 72.97741 50.35447 74.56047 63.53303 58.34630 67.89768
## 15 57.41803 86.39974 69.13317 64.10878 72.08983 72.46933 70.65156 70.50410
## 16 81.07736 61.83588 55.29717 78.05782 68.97257 77.92182 54.06106 68.19299
## 17 69.27036 72.73176 66.26249 81.20352 77.43711 67.02196 73.33301 70.98040
## 18 71.18285 71.76011 83.50827 77.31534 80.33721 57.93053 60.87592 57.04168
## 19 77.08574 63.53765 54.28993 54.81922 82.70557 82.75997 76.84481 48.23588
## 20 83.37254 79.74481 68.65916 79.85044 70.31409 68.82998 91.27170 55.04243
## 21 79.42429 57.50524 70.29875 70.65414 64.68261 62.68907 71.53883 69.07870
## 22 91.60073 62.76676 58.51641 60.47161 78.71268 85.45890 85.58557 63.88609
## 23 54.85164 54.43394 65.50619 59.28265 70.88144 57.60567 56.73982 89.82601
## 24 69.50500 72.49468 68.63347 67.76174 82.60173 61.57659 68.59692 86.66548
## 25 65.73445 63.82714 58.39026 62.94508 70.95651 68.95208 65.34120 80.61999
## 26 72.96010 72.51607 64.07870 53.46808 68.55277 80.84143 72.52942 82.84771
## 27 68.98947 70.44727 66.04041 58.19373 62.69908 61.45918 89.92116 77.38010
## 28 73.96042 75.53844 77.98951 83.10144 68.39982 68.42600 60.50911 81.45808
## 29 61.14622 60.60551 69.47783 65.08185 71.97769 61.37920 90.13989 75.44525
## 30 72.19989 57.55211 65.86454 69.90636 59.85463 55.46316 78.31594 62.43129
##         m_9     m_10
## 1  79.94395 77.15164
## 2  84.16208 67.27853
## 3  64.99590 64.12485
## 4  83.02655 59.66283
## 5  81.53630 73.67870
## 6  70.77808 75.65323
## 7  74.40741 59.33277
## 8  76.38843 63.10045
## 9  80.60530 68.36408
## 10 47.52721 65.95306
```

```
## 11 59.98897 79.20845
## 12 54.98642 66.96376
## 13 63.38774 63.40628
## 14 74.59578 80.44104
## 15 66.18963 71.78045
## 16 67.36016 77.77233
## 17 81.31876 74.65616
## 18 82.44293 62.72818
## 19 78.52473 87.05085
## 20 70.24520 79.55418
## 21 78.77107 82.23707
## 22 69.20672 88.69741
## 23 55.61414 57.33495
## 24 71.93640 75.14083
## 25 66.64614 73.47953
## 26 64.29463 81.39062
## 27 46.55009 88.35622
## 28 61.16354 62.12075
## 29 72.86427 77.02079
## 30 79.38174 75.74485
```

Calculo de las varianzas muestrales para cada una de las muestras.

```r
var_muestras <- apply(matriz_m,2, var) #el 2 significa que va a trabajar sobre las columnas
```

Calculo del tamaño de las muestras

```r
N <- 10000   #Tamaño de la población finita
B<- 3 #Error dado en el enunciado
D <- B/qnorm(0.975)
n <- var_muestras*N/((N*D^2)+var_muestras)
```

Etapa 2: Muestreo y estimacion de la media poblacional

```r
n_1 <- rnorm(n[1], 70, 10)
n_2 <- rnorm(n[2], 70, 10)
n_3 <- rnorm(n[3], 70, 10)
n_4 <- rnorm(n[4], 70, 10)
n_5 <- rnorm(n[5], 70, 10)
n_6 <- rnorm(n[6], 70, 10)
n_7 <- rnorm(n[7], 70, 10)
n_8 <- rnorm(n[8], 70, 10)
n_9 <- rnorm(n[9], 70, 10)
n_10 <- rnorm(n[10], 70, 10)
```

Estimación puntual de μ a traves de medias_muestrales

```r
mm1 <- mean(n_1)
mm2 <- mean(n_2)
mm3 <- mean(n_3)
mm4 <- mean(n_4)
mm5 <- mean(n_5)
```

```r
mm6 <- mean(n_6)
mm7 <- mean(n_7)
mm8 <- mean(n_8)
mm9 <- mean(n_9)
mm10 <- mean(n_10)

vmm <- c(mm1, mm2, mm3, mm4, mm5, mm6, mm7, mm8, mm9, mm10)
```

Generar los intervalos de confianza

Limite inferior y superior

```r
LI <- vmm-3
LS <- vmm+3

IC<- data.frame(LI, LS )
View(IC)
```

Calculo del coeficiente de cnfianza para cada muestra

$$z_{\alpha/2} = \frac{B}{\sqrt{\frac{s^2}{n}\left(\frac{N-n}{n-1}\right)}}$$

```r
vm1 <- var(n_1)
vm2 <-  var(n_2)
vm3 <-  var(n_3)
vm4 <-  var(n_4)
vm5 <-  var(n_5)
vm6 <-  var(n_6)
vm7 <-  var(n_7)
vm8 <-  var(n_8)
vm9 <-  var(n_9)
vm10 <-  var(n_10)

vvm <- c(vm1, vm2, vm3, vm4, vm5, vm6, vm7, vm8, vm9, vm10) #vector de varianzas muestrales
```

Calculo de $z_{\alpha/2}$

```r
zalpha_2 <- B/(vvm/n*(N-n))^0.5
```

Calculo de los coeficientes de confianza

```r
CC <- (pnorm(zalpha_2)-pnorm(-zalpha_2))*100        #Porque me salen 1? Deberian salir menores a 1

IC<- data.frame(LI, LS, CC, vvm) #Entre mayor varianza, el coeficiente de confianza disminuye

IC
```

```
##             LI       LS       CC       vvm
## m_1  66.07890 72.07890 1.082517 158.25121
## m_2  63.49230 69.49230 1.488181  96.49714
```

```
## m_3  67.75271 73.75271 1.676977  87.95448
## m_4  64.83939 70.83939 1.568341 104.41969
## m_5  68.30702 74.30702 1.691132 115.98539
## m_6  65.87512 71.87512 1.112463 145.89326
## m_7  67.77241 73.77241 1.955790  88.27832
## m_8  69.41220 75.41220 1.882342  68.76046
## m_9  65.33298 71.33298 1.647558  97.12893
## m_10 66.91786 72.91786 1.438635  93.37040
```

```
CC>95
```

```
##   m_1   m_2   m_3   m_4   m_5   m_6   m_7   m_8   m_9  m_10
## FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE
```

##Proporciones La proporción de una muestra

En proporciones la varianza propocional es

$$\sigma^2 = pq$$

donde $q = 1 - p$

y la varizana muestral es

$$s^2 = \hat{p}\hat{q}$$

Suponga que hara una encuesta ver si los estudiantes de la FI de la UNAM ya se han adaptado al nuevo modelo de enseñanza en línea. Se sabe que la población es de $N = 14,000$ estudiantes. Se solicita una confianza del 97% y un error del 2%.

Calcule el tamaño de la muestra n para este estudio.

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

Donde

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

```
N1 <- 14000 #Estudiantes
B1 <- 0.02 #Error
z.97 <- qnorm(0.97+0.03/2)

D1 <- B1/z.97
D1
```

```
## [1] 0.009216206
```

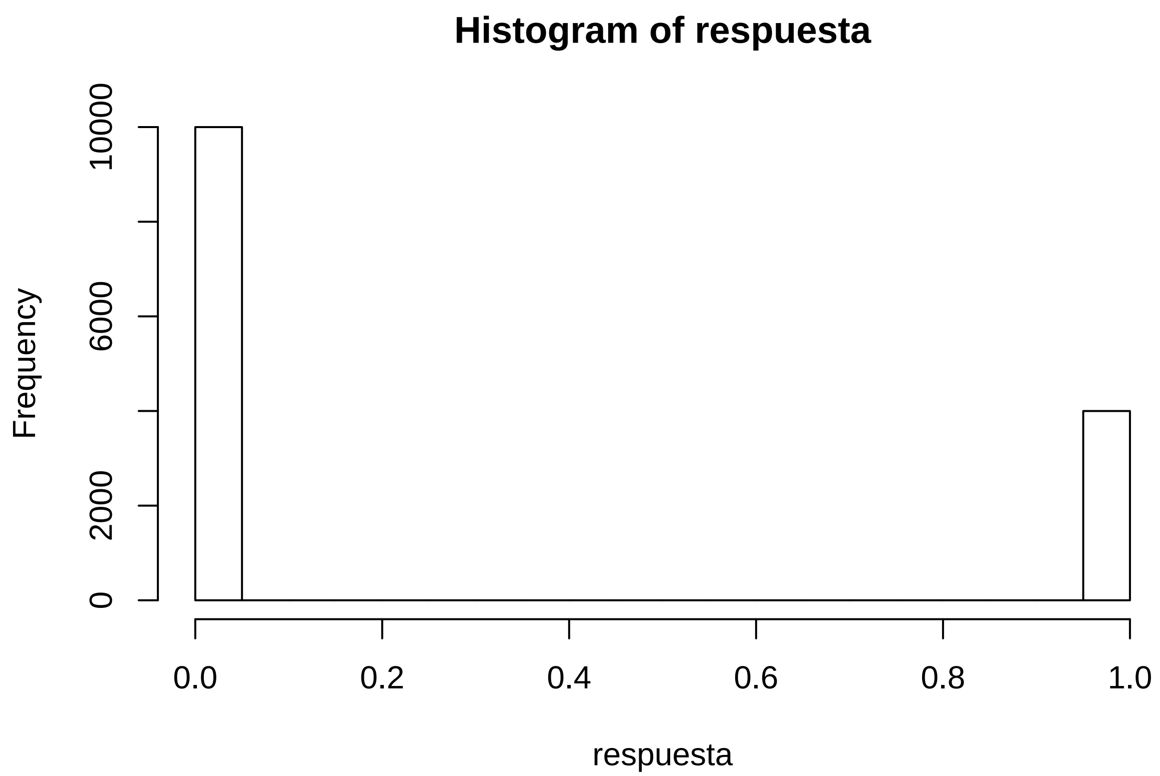
Si tomamos la varianza más alta

Con p $p^* = 0.5$ la varianza más alta $s^{2*} = p^*(1 - p^*) = 0.5(1 - 0.5) = 0.25$

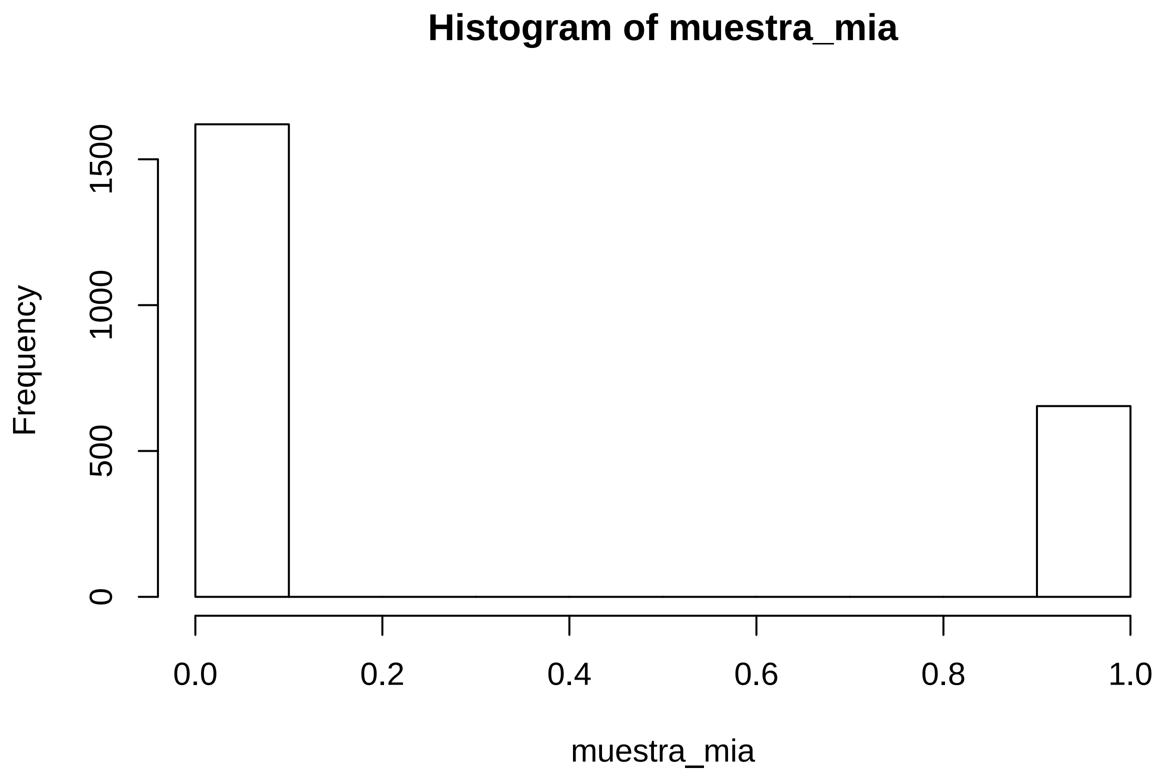
```
pm <- 0.5 #proporcion que calcula la varianza más estimada

n <- pm*(1-pm)*N/(D1^2*(N-1)+pm*(1-pm))
```

Creación de la población

```r
respuesta <- c(rep(1, times=4000), rep(0, times=10000))
hist(respuesta)
```




```r
muestra_mia <- sample(respuesta, n)
hist(muestra_mia)
```




```r
pestimada <-sum(muestra_mia)/n

preal <- 4000/14000

LI <- pestimada-B1
LS <- pestimada+B1

s2estimada <- pestimada*(1-pestimada)
```

De que tamaño hubiera sido la muestra si se toma esta varianza estiamda

```r
pestimada*(1-pestimada)*N/(D1^2*(N-1)+pestimada*(1-pestimada))
```

```
## [1] 1943.474
```