Actividad 6. Distribuciones Muestrales y TCL

Oscar Gutierrez

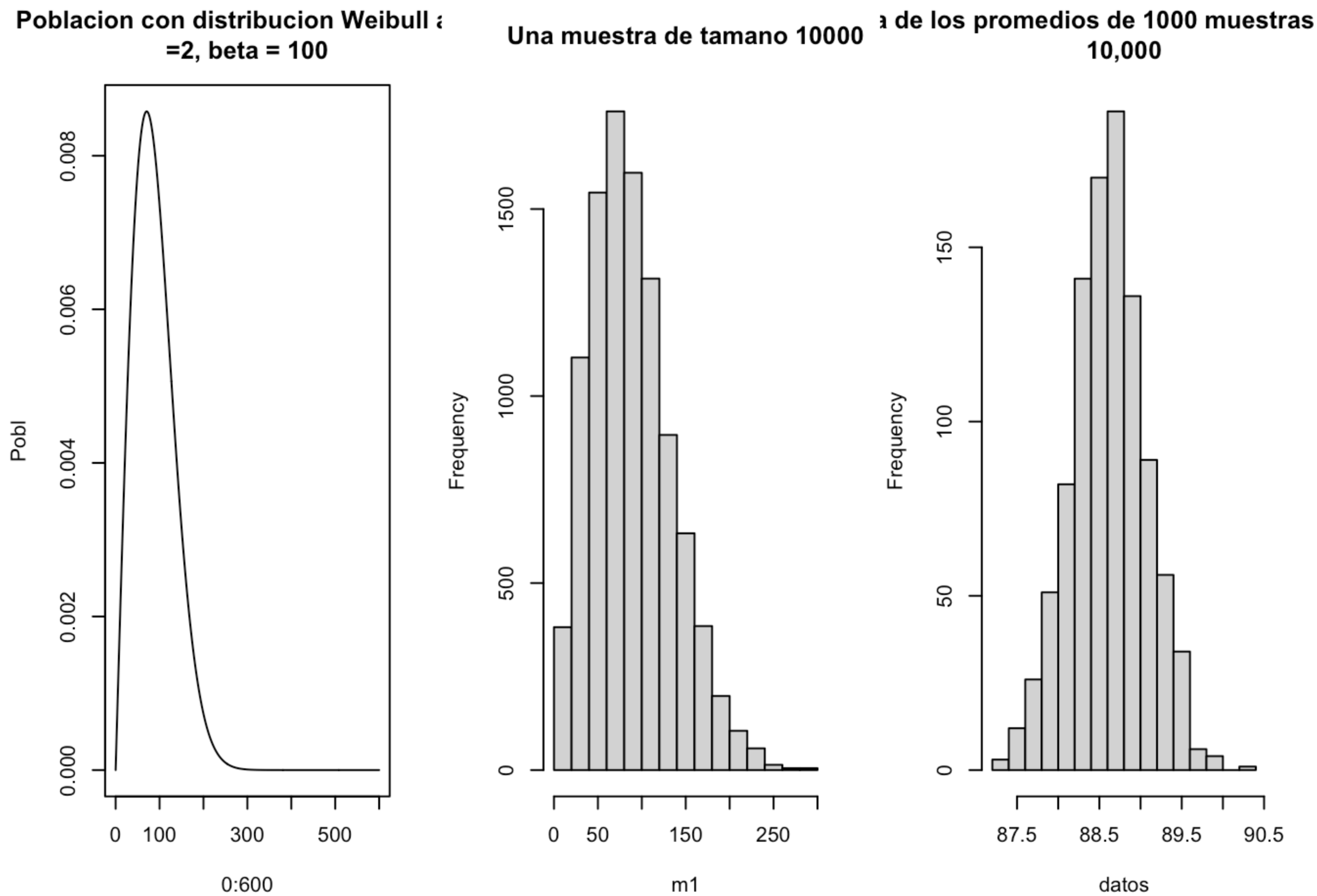
2024-08-16

1. Ensayando distribuciones

Inciso A

Ejercutar el siguiente código de R: DistrsM_enR.txt Download DistrsM_enR.txt. Se esperan tres gráficas, interprete cada una de ellas. Se usa una distribución Weibull, con parámetros alfa = 2 y beta = 100.

```
par(mfrow=c(1,3))
# Graficando una distribucion Weibull de alfa =2, beta = 100
Pobl = dweibull(0:600,2, 100)
plot(0:600,Pobl, type="l", main = "Poblacion con distribucion Weibull alfa
=2, beta = 100")
# Tomando una muestra de 10000 elementos tomados al azar
m1 = rweibull(10000, 2, 100)
hist(m1, main = "Una muestra de tamano 10000")
# Tomando 1000 promedios de las 1000 muestras como la anterior
m =rweibull(10000,2,100)
prom=mean(m)
datos=prom
for(i in 1:999) {
m =rweibull(10000,2,100)
prom=mean(m)
datos=rbind(datos,prom) }
hist(datos, main="Grafica de los promedios de 1000 muestras de tamano
10,000")
```



La gráfica de la izquierda es la población, se observa una distribución asimétrica. La gráfica central es una muestra de la población, la distribución es muy similar. La gráfica de la derecha es la distribución de la media de las muestras, estas se distribuyen normalmente por el TCL.

Inciso B

Cálcula el sesgo y la curtosis de la muestra de tamaño 10000. Aplica una prueba de hipótesis de normalidad. Concluye sobre la normalidad de los datos de la muestra.

```
library(moments)
library(nortest)
cat('sesgo = ',skewness(m1), '\ncurtosis=', kurtosis(m1))
```

```
## sesgo =  0.6487453
## curtosis= 3.248657
```

```
ad.test(m1)
```

```
##
##  Anderson-Darling normality test
##
## data:  m1
## A = 62.208, p-value < 2.2e-16
```

h0: Los datos siguen una distribucion normal h1: Los datos no siguen una distribucion normal

El sesgo es muy diferente de 0, la curtosis sí esta cerca de 3, la prueba de normalidad da un valor de p muy pequeño por lo que se rechaza el hecho de que sea una distribución normal.

Inciso C

Calcula el sesgo y la curtosis de las medias de las 1000 muestras. Aplica la misma prueba de normalidad que aplicaste a la muestra de tamaño 10000. Concluye sobre la normalidad de las medias de las muestras.

```
cat('sesgo = ',skewness(datos), '\ncurtosis=', kurtosis(datos))
```

```
## sesgo =  0.0156967
## curtosis= 3.046426
```

```
ad.test(datos)
```

```
##
##  Anderson-Darling normality test
##
## data:  datos
## A = 0.23701, p-value = 0.7855
```

h0: Los datos siguen una distribucion normal h1: Los datos no siguen una distribucion normal

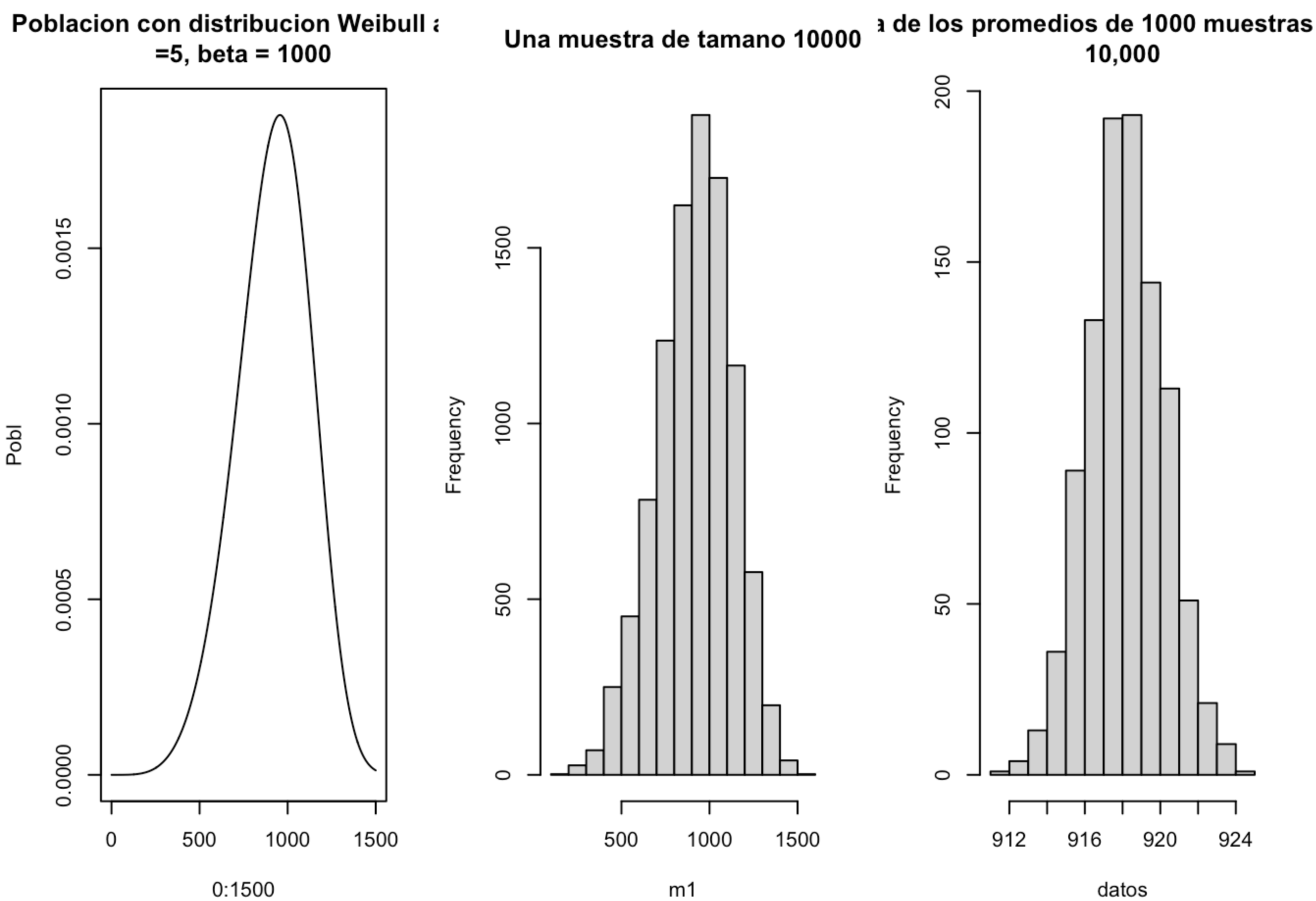
El sesgo esta cerca de 0, la curtosis también esta cerca de 3, la prueba de normalidad da un valor de p suficiente para no rechazar la hipótesis nula.

Inciso D

Repite el procedimiento A, B y C para otras dos distribuciones que no sean simétricas. Puedes cambiar los valores de alfa y beta para lograr sesgo diferente o puedes ensayar con otra distribución, como la uniforme (punif y runif). Interpreta los resultados

Distribucion con alpha = 5 y beta = 1000

```r
par(mfrow=c(1,3))
# Graficando una distribucion Weibull de alfa =5, beta = 1000
Pobl = dweibull(0:1500,5, 1000)
plot(0:1500,Pobl, type="l", main = "Poblacion con distribucion Weibull alfa
=5, beta = 1000")
# Tomando una muestra de 10000 elementos tomados al azar
m1 = rweibull(10000, 5, 1000)
hist(m1, main = "Una muestra de tamano 10000")
# Tomando 1000 promedios de las 1000 muestras como la anterior
m =rweibull(10000,5,1000)
prom=mean(m)
datos=prom
for(i in 1:999) {
m =rweibull(10000,5,1000)
prom=mean(m)
datos=rbind(datos,prom) }
hist(datos, main="Grafica de los promedios de 1000 muestras de tamano
10,000")
```




```r
cat('sesgo muestra = ',skewness(m1), '\ncurtosis muestra=', kurtosis(m1))
```

```
## sesgo muestra =  -0.2721843
## curtosis muestra= 2.850374
```

```
ad.test(m1)
```

```
##
##  Anderson-Darling normality test
##
## data:  m1
## A = 13.387, p-value < 2.2e-16
```

```
cat('sesgo xbar = ',skewness(datos), '\ncurtosis xbar=', kurtosis(datos))
```

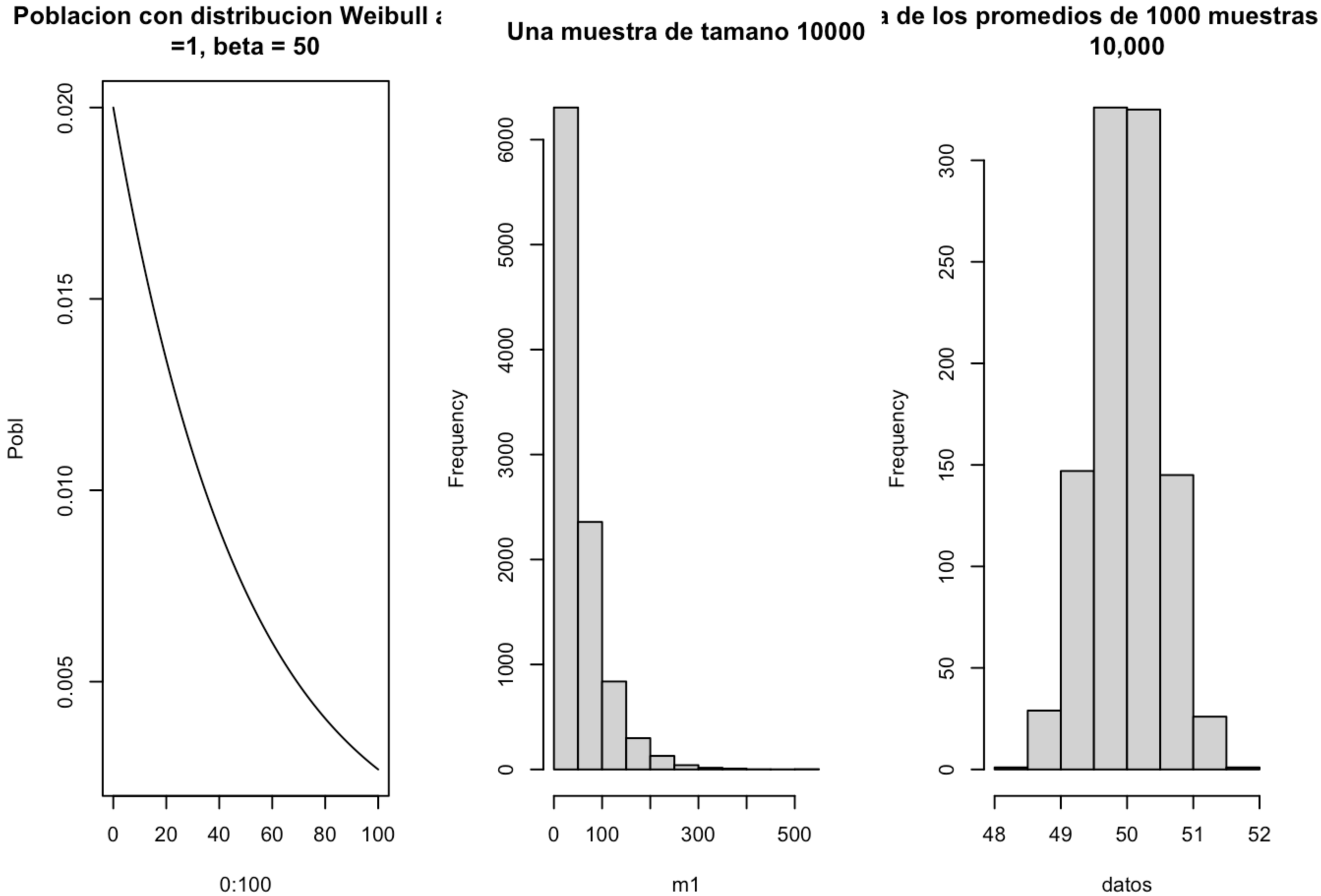
```
## sesgo xbar =  0.03191673
## curtosis xbar= 2.906658
```

```
ad.test(datos)
```

```
##
##  Anderson-Darling normality test
##
## data:  datos
## A = 0.1354, p-value = 0.9785
```

Distribucion con alpha = 1 y beta = 50

```
par(mfrow=c(1,3))
# Graficando una distribucion Weibull de alfa =1, beta = 50
Pobl = dweibull(0:100,1, 50)
plot(0:100,Pobl, type="l", main = "Poblacion con distribucion Weibull alfa
=1, beta = 50")
# Tomando una muestra de 10000 elementos tomados al azar
m1 = rweibull(10000, 1, 50)
hist(m1, main = "Una muestra de tamano 10000")
# Tomando 1000 promedios de las 1000 muestras como la anterior
m =rweibull(10000,1,50)
prom=mean(m)
datos=prom
for(i in 1:999) {
m =rweibull(10000,1,50)
prom=mean(m)
datos=rbind(datos,prom) }
hist(datos, main="Grafica de los promedios de 1000 muestras de tamano
10,000")
```




```
cat('sesgo muestra = ',skewness(m1), '\ncurtosis muestra=', kurtosis(m1))
```

```
## sesgo muestra =  2.168331
## curtosis muestra= 10.67417
```

```
ad.test(m1)
```

```
##
##  Anderson-Darling normality test
##
## data:  m1
## A = 469.69, p-value < 2.2e-16
```

```
cat('sesgo xbar = ',skewness(datos), '\ncurtosis xbar=', kurtosis(datos))
```

```
## sesgo xbar =  0.01523931
## curtosis xbar= 2.795193
```

```
ad.test(datos)
```

```
##
##  Anderson-Darling normality test
##
## data:  datos
## A = 0.24421, p-value = 0.7628
```

Inciso E

Concluye sobre las semejanzas y diferencias entre los tres gráficos generados en cada una de las tres distribuciones teóricas.

En los 3 casos las gráficas de las poblaciones son asimétricas al igual que las gráficas de las muestras de n = 10000, mientras que las gráficas de xbar son aproximadamente normales de acuerdo con la prueba de normalidad realizada.

2. Remaches

La resistencia a la ruptura de un remache tiene un valor medio de 10,000 lb/pulg2 y una desviación estándar de 500 lb/pulg2. Si se sabe que la población se distribuye normalmente,

X: Resistencia a la ruptura de un remache $X \sim N(\mu_x = 10000, \sigma_x = 500)$

Inciso A

¿Cuál es la probabilidad de que la tomar un remache al azar de esa población, éste tenga una resistencia a la ruptura que esté a 100 unidades alrededor de la media? ¿a cuántas desviaciones estándar está de la media?

```
mu = 10000
sigma = 500
a = pnorm(10100, mu, sigma) - pnorm(9900, mu, sigma)
cat("P(9900 < X < 10100) = ", round(a, 4), "\nz = ", (10100-mu)/sigma)
```

```
## P(9900 < X < 10100) =  0.1585
## z =  0.2
```

Inciso B

¿Cuál es la probabilidad de que la resistencia media a la ruptura de la muestra aleatoria de 120 remaches esté 100 unidades alrededor de su media? ¿a cuántas desviaciones estándar está de la media?

$\bar{X} \sim N\left(\mu_{\bar{x}} = 10000, \sigma_{\bar{x}} = \frac{500}{\sqrt{120}}\right)$

$P(9900 < \bar{X} < 10000)$

```
mu = 10000
n = 120
sigma_xbar = 500/sqrt(n)
b = pnorm(10100, mu, sigma_xbar) - pnorm(9900, mu, sigma_xbar)
cat("P(9900 < X < 10100) = ", round(b, 4), "\nz = ", (10100-mu)/sigma_xbar)
```

```
## P(9900 < X < 10100) =  0.9715
## z =  2.19089
```

Inciso C

Si el tamaño muestral hubiera sido 15, en lugar de 120, ¿cuál es la probabilidad de que la resistencia media a la ruptura esté 100 unidades alrededor de la media? ¿a cuántas desviaciones estándar está de la media?

```
mu = 10000
n2 = 15
sigma_xbar_2 = 500/sqrt(n2)
c = pnorm(10100, mu, sigma_xbar_2) - pnorm(9900, mu, sigma_xbar_2)
cat("P(9900 < X < 10100) = ", round(c, 4), "\nz = ", (10100-mu)/sigma_xbar_2)
```

```
## P(9900 < X < 10100) =  0.5614
## z =  0.7745967
```

Inciso D

Un ingeniero recibió un lote muy grande de remaches. Antes de aceptarlo quiso verificar si efectivamente la media de la resistencia de los remaches es de 10 000 lb/pulg2. Para ello tomó una muestra de 120 remaches elegidos al azar tenía media de 9800 lb/pulg2 y rechazó el pedido, ¿hizo lo correcto? ¿por qué?.

```
mu = 10000
n3 = 120
sigma_xbar_3 = 500/sqrt(n3)
d = pnorm(9800, mu, sigma_xbar_3)
cat("P(X < 9800) = ", round(d, 4))
```

```
## P(X < 9800) =  0
```

Si hizo lo correcto, porque la probabilidad de tener una media de 9800 o menos es casi nula.

Inciso E

¿Qué decisión recomiendas al ingeniero si la media obtenida en la media hubiera sido 9925? ¿recomendarías rechazarlo?

```
mu = 10000
n4 = 120
sigma_xbar_4 = 500/sqrt(n4)
d = pnorm(9925, mu, sigma_xbar_4)
cat("P(X < 9925) = ", round(d, 4), "\nz = ", (9925-mu)/sigma_xbar_4)
```

```
## P(X < 9925) =  0.0502
## z =  -1.643168
```

$H_0 : \mu = 10000$ $H_1 : \mu < 10000$ Depende de el criterio de α, si tenemos un valor de alpha de 0.05 no se rechaza la hipótesis nula, si se tiene un alpha de 0.1, se rechaza H_0

3. Embotellando

Una máquina embotelladora puede ser regulada para que se descargue un promedio de μ onzas por botella. Se ha observado que la cantidad de líquido dosificado por una máquina embotelladora está distribuida normalmente con $\sigma = 1$ onza. La máquina embotelladora se calibra cuando la media de una muestra tomada al azar está fuera del 95% central de la distribución muestral. La media de la cantidad de líquido deseada requiere que μ sea de 15 onzas.

Inciso 1

¿A cuántas desviaciones estándar alrededor de la verdadera media μ puede estar la media de una muestra para que esté dentro del estándar establecido del 95% central?

```
cat("z=",qnorm(0.975))
```

```
## z= 1.959964
```

Inciso 2

¿Cuál es la probabilidad de que en una muestra aleatoria de tamaño 10 botellas se obtenga una media mayor a 16 onzas?

```
mu = 15
n = 10
sigma = 1/sqrt(10)
cat("P(Xbar > 16)=", 1- pnorm(16,mu,sigma))
```

```
## P(Xbar > 16)= 0.0007827011
```

Inciso 3

Si en una muestra aleatoria de tamaño 10 botellas se obtuvo una media de 16 onzas, ¿se detendría la producción para calibrar la máquina?

```
cat("z=", qnorm(pnorm(16,mu,sigma)))
```

```
## z= 3.162278
```

Sí porque ese valor de media sale del 95% central.

Inciso 4

¿Cuál es la probabilidad de que en una muestra aleatoria de tamaño 10 botellas se obtenga una media menor a 14.5 onzas?

```
mu = 15
n = 10
sigma = 1/sqrt(10)
cat("P(Xbar < 14.5)=", pnorm(14.5,mu,sigma))
```

```
## P(Xbar < 14.5)= 0.05692315
```

Inciso 5

Si en una muestra aleatoria de tamaño 10 botellas se obtuvo una media de 15.5 onzas, ¿se detendría la producción para calibrar la máquina?

```
mu = 15
n = 10
sigma = 1/sqrt(10)
cat("P(Xbar > 15.5)=", 1- pnorm(15.5,mu,sigma), "z=", qnorm(pnorm(15.5,mu,sigma)))
```

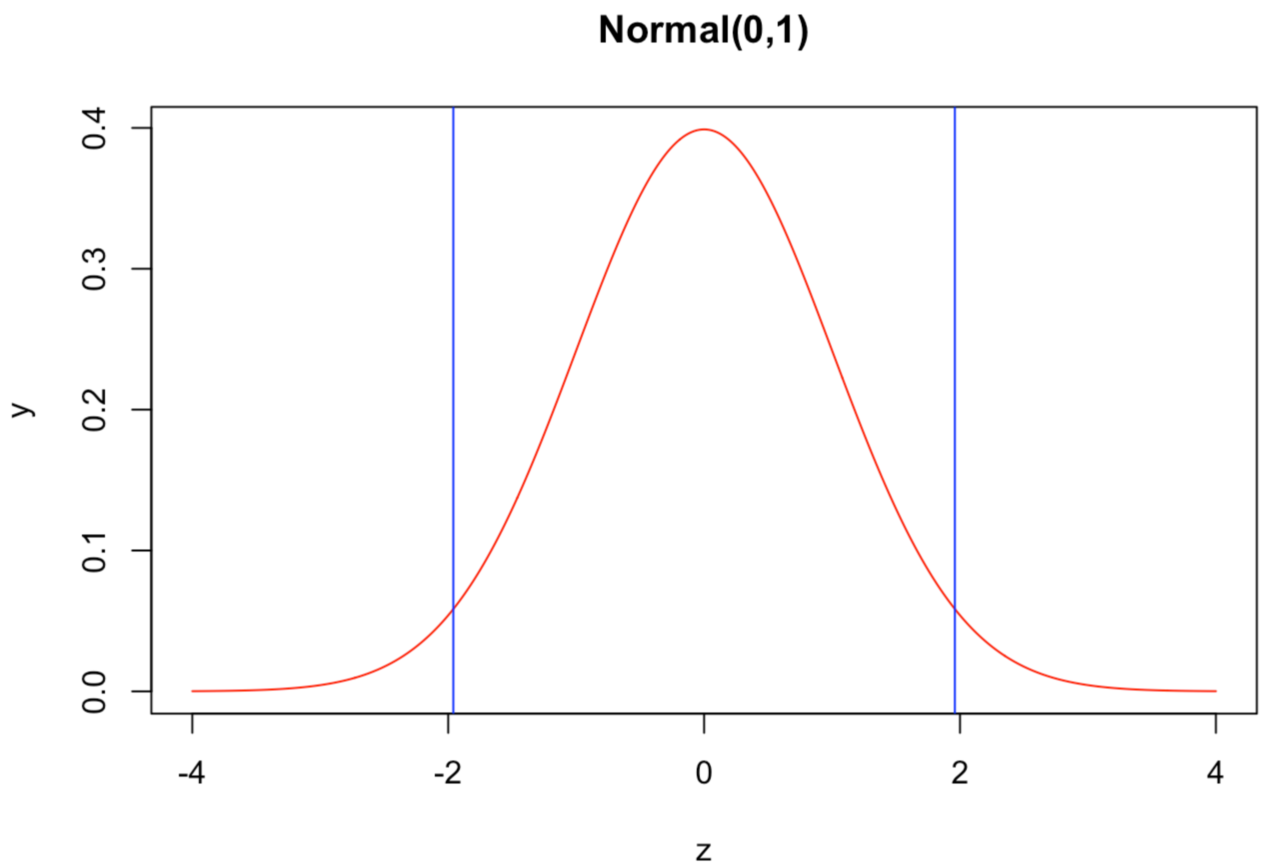
```
## P(Xbar > 15.5)= 0.05692315 z= 1.581139
```

No, ese valor está dentro del 95% central

Inciso 6

Hacer una gráfica del inciso 1.

```
miu = 0
sigma = 1
z = seq(miu - 4*sigma, miu + 4*sigma, 0.01)
y = dnorm(z,miu, sigma)
plot(z,y, type = "l", col = "red", main = "Normal(0,1)")
abline(v= 1.959964, col="blue")
abline(v= -1.959964, col="blue")
```


Normal(0,1)
y
0.0
0.1
0.2
0.3
0.4
-4
-2
0
2
4
z