Act_6_A01742161

Rogelio Lizárraga

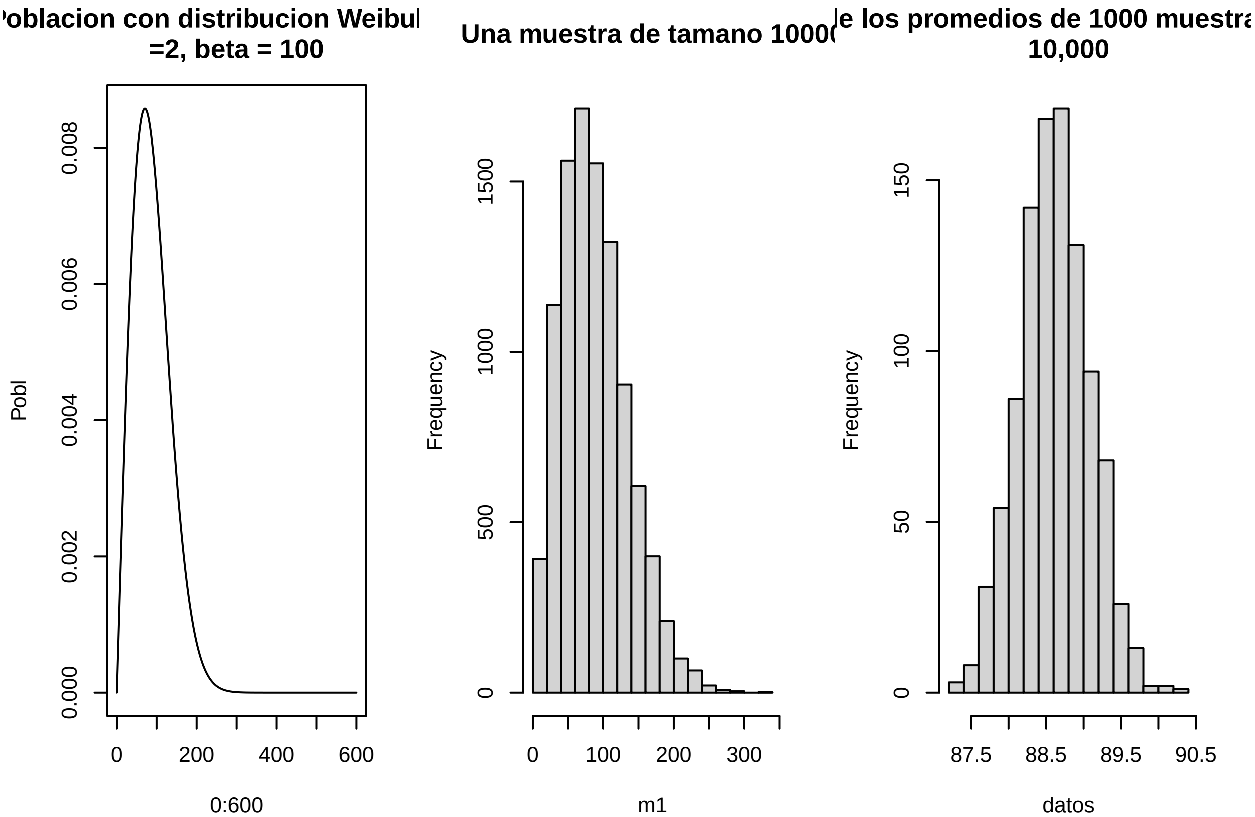
2024-08-16

1. Ensayando Distribuciones

Grafica la Distribución de una variable aleatoria, la de una muestra elegida al azar y la de la Distribución de las medias de 10000 muestras:

A . 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")
```



Observamos que la distribución Weibull tiene una curtosis alta y un sesgo hacia la derecha, así como la gráfica de la muestra de tamaño 10000. Sin embargo, la gráfica de los promedios de 1000 muestras de 10000 sí sigue una distribución normal, pues su distribución es simétrica y no tiene una curtosis o sesgo significativo.

B. Calcula 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.

H_0 = el conjunto de datos se distribuye de manera normal. H_1 = el conjunto de datos NO se distribuye de manera normal.

```
library(e1071)
library(nortest)
kurtosis(m1)
```

```
## [1] 0.333561
```

```
skewness(m1)
```

```
## [1] 0.6807453
```

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

```
##
##  Anderson-Darling normality test
##
## data:  m1
## A = 66.392, p-value < 2.2e-16
```

Observamos que tenemos una curtosis ligeramente elevada de 0.44 y un sesgo elevado de 0.69. Como nuestro valor $p < 0.1$, se rechaza H_0 para el conjunto de datos, por lo que estos NO se distribuyen de manera normal.

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.

$H_0 =$ el conjunto de datos se distribuye de manera normal. $H_1 =$ el conjunto de datos NO se distribuye de manera normal.

```r
library(e1071)
library(nortest)
kurtosis(datos)
```

```
## [1] -0.04137323
```

```r
skewness(datos)
```

```
## [1] 0.06506813
```

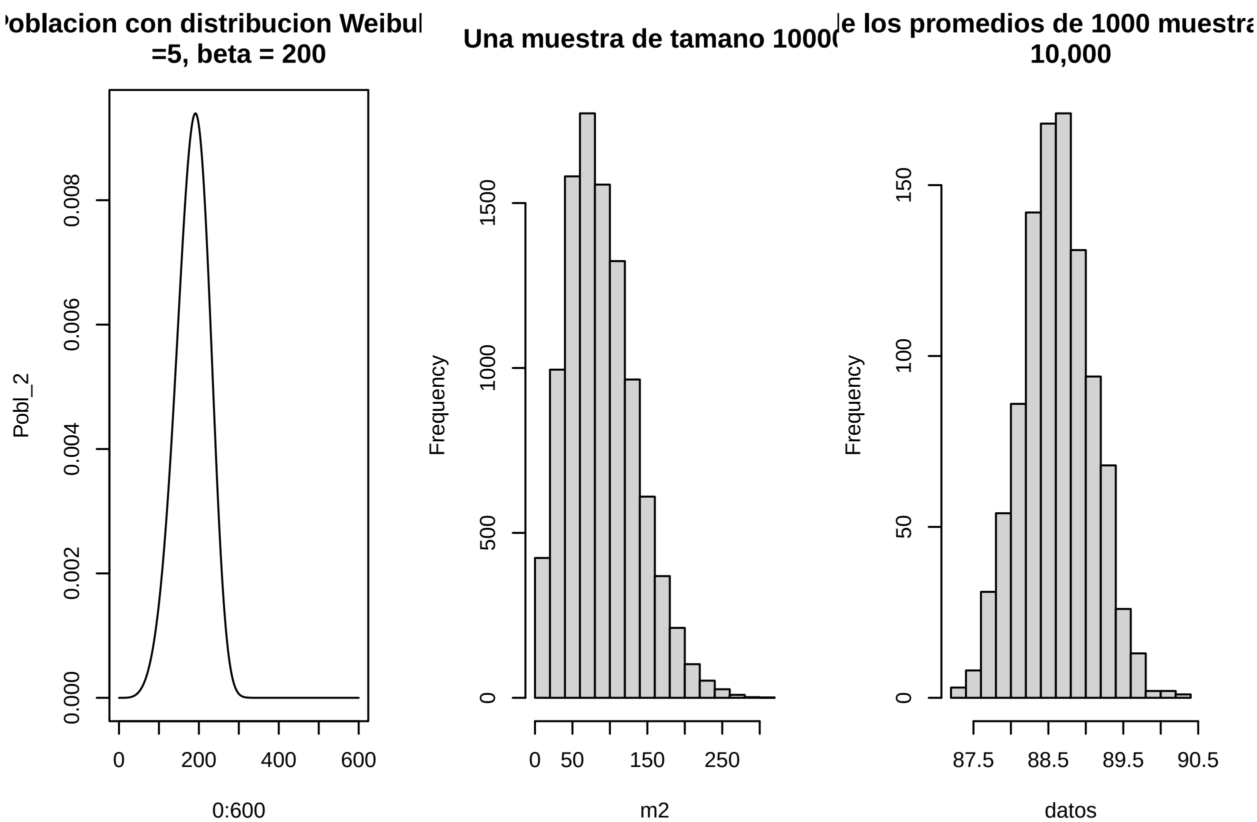
```r
ad.test(datos)
```

```
##
##  Anderson-Darling normality test
##
## data:  datos
## A = 0.26785, p-value = 0.6843
```

Observamos que tenemos una curtosis ligeramente negativa de -0.16 y un sesgo ligero de 0.118. Como nuestro valor $p > 0.1$, NO se rechaza H_0 para el conjunto de datos, por lo que estos se distribuyen de manera normal. S # 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.

```r
par(mfrow=c(1,3))
# Graficando una distribucion Weibull de alfa =5, beta = 200
Pobl_2 = dweibull(0:600, 5, 200)
plot(0:600,Pobl_2, type="l", main = "Poblacion con distribucion Weibull alfa
```

```
=5, beta = 200")
# Tomando una muestra de 10000 elementos tomados al azar
m2 = rweibull(10000, 2, 100)
hist(m2, 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_2=prom
for(i in 1:999) {
m =rweibull(10000,2,100)
prom=mean(m)
datos_2=rbind(datos_2,prom) }
hist(datos, main="Grafica de los promedios de 1000 muestras de tamano
10,000")
```




```
# Cálculo de sesgo, curtosis y pruebas de normalidad

library(e1071)
library(nortest)
kurtosis(m2)
```

```
## [1] 0.3469852
```

```r
skewness(m2)
```

```
## [1] 0.6598039
```

```r
ad.test(m2)
```

```
##
##  Anderson-Darling normality test
##
## data:  m2
## A = 58.528, p-value < 2.2e-16
```

```r
kurtosis(datos_2)
```

```
## [1] -0.2361027
```

```r
skewness(datos_2)
```

```
## [1] 0.1923141
```

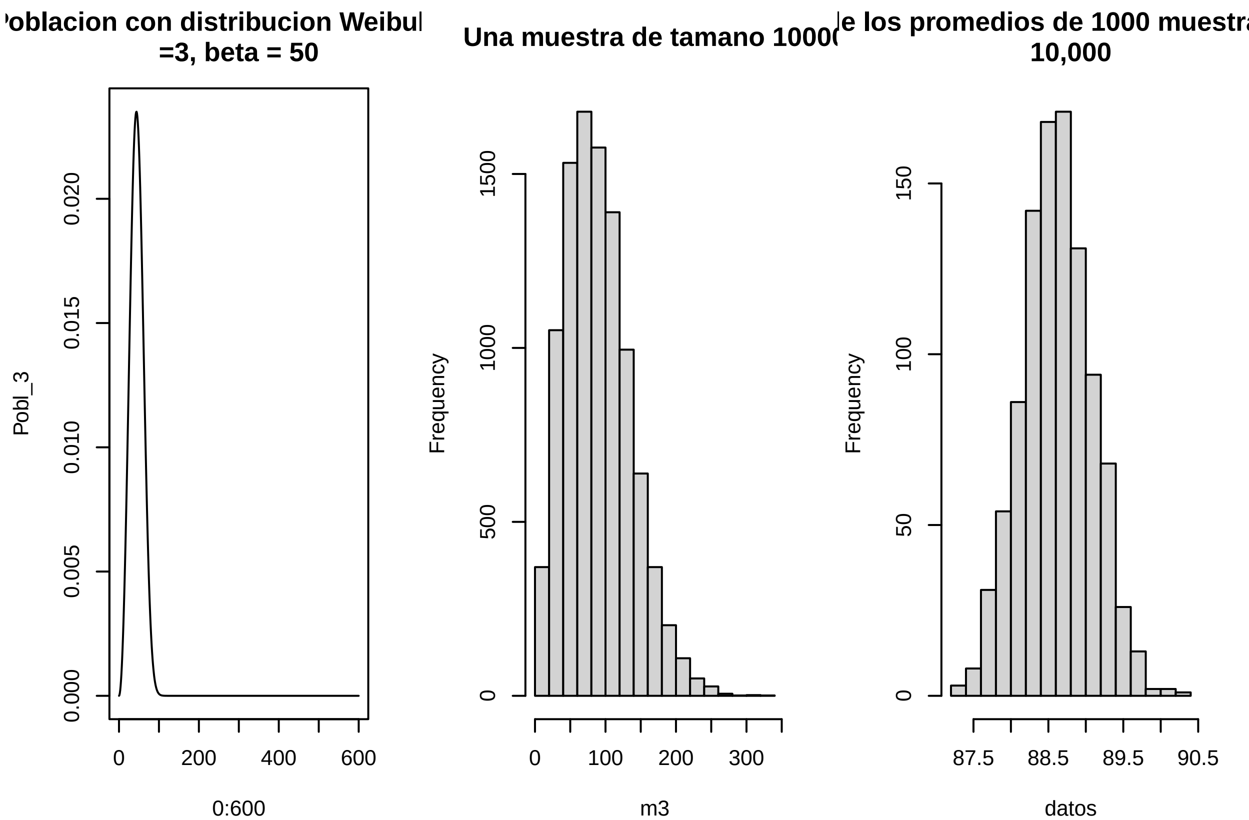
```r
ad.test(datos_2)
```

```
##
##  Anderson-Darling normality test
##
## data:  datos_2
## A = 0.94779, p-value = 0.0165
```

Observamos que la distribución Weibull tiene una curtosis alta y un sesgo hacia la derecha, así como la gráfica de la muestra de tamaño 10000. Sin embargo, la gráfica de los promedios de 1000 muestras de 10000 sí sigue una distribución normal, pues su distribución es simétrica y no tiene una curtosis o sesgo significativo. En valores numéricos, obtenemos una curtosis de 0.33, un sesgo de 0.66 y un valor $p < 0.1$ para la muestra de tamaño, por lo que los datos no se distribuyen de manera normal. Por otro lado, con el promedio de las muestras de 10000, obtenemos curtosis de -0.04, sesgo de -0.07 y un valor $p > 0.1$, por lo que los datos sí se distribuyen de manera normal.

```r
par(mfrow=c(1,3))
# Graficando una distribucion Weibull de alfa =3, beta = 50
Pobl_3 = dweibull(0:600, 3, 50)
plot(0:600,Pobl_3, type="l", main = "Poblacion con distribucion Weibull alfa
=3, beta = 50")
# Tomando una muestra de 10000 elementos tomados al azar
m3 = rweibull(10000, 2, 100)
hist(m3, 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_3=prom
for(i in 1:999) {
m =rweibull(10000,2,100)
```

```
prom=mean(m)
datos_3=rbind(datos_3,prom) }
hist(datos, main="Grafica de los promedios de 1000 muestras de tamano
10,000")
```




```
# Cálculo de sesgo, curtosis y pruebas de normalidad

library(e1071)
library(nortest)
kurtosis(m3)
```

```
## [1] 0.3277245
```

```
skewness(m3)
```

```
## [1] 0.6362989
```

```
ad.test(m3)
```

```
##
##  Anderson-Darling normality test
##
## data:  m3
## A = 53.451, p-value < 2.2e-16
```

```r
kurtosis(datos_3)
```

```
## [1] -0.1393726
```

```r
skewness(datos_3)
```

```
## [1] -0.09387231
```

```r
ad.test(datos_3)
```

```
##
##  Anderson-Darling normality test
##
## data:  datos_3
## A = 0.27308, p-value = 0.6668
```

Observamos que la distribución Weibull tiene una curtosis alta y un sesgo muy hacia la derecha, así como la gráfica de la muestra de tamaño 10000. Sin embargo, la gráfica de los promedios de 1000 muestras de 10000 parece que sigue una distribución normal, pues su distribución es simétrica y no tiene una curtosis o sesgo significativo. En valores numéricos, obtenemos una curtosis de 0.21, un sesgo de 0.64 y un valor $p < 0.1$ para la muestra de tamaño 10000, por lo que los datos no se distribuyen de manera normal. Por otro lado, con el promedio de las muestras de 10000, obtenemos curtosis de -0.08, sesgo de -0.05 y un valor $p > 0.1$, por lo que los datos sí se distribuyen de manera normal.

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

Semejanzas: en todas las distribuciones, a pesar de la forma original, la distribución de los promedios tiende a ser más simétrica y cercana a la normal, debido al teorema central del límite.

Diferencias: las distribuciones originales muestran distintas formas y distintos niveles de asimetría, pues observamos que la última distribución es la más asimétrica de todas, seguida por la primera distribución original y la menos asimétrica sería la segunda distribución 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.

$$X \sim N(\mu_x = 10000, \sigma_x = 500)$$

a. ¿Cuál es la probabilidad de que al 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?

$P(9900 \leq X \leq 11000)$

```
mu = 10000
sigma  = 500
p_a = pnorm(mu + 100, mu, sigma) - pnorm(mu - 100, mu, sigma)
cat('P(9900 < X < 10100) =', p_a)
```

```
## P(9900 < X < 10100) = 0.1585194
```

```
z_a = 100 / sigma
cat('\nz =', z_a)
```

```
##
## z = 0.2
```

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?

$P(9900 \leq \bar{X} \leq 11000)$

```
p_2 = pnorm(mu + 100, mu, sigma/sqrt(120)) - pnorm(mu - 100, mu, sigma/sqrt(120))
cat('P(9900 < X < 10100) =', p_2)
```

```
## P(9900 < X < 10100) = 0.9715403
```

```
z_b = 100/ (sigma / sqrt(120))
cat('\nz =', z_b)
```

```
##
## z = 2.19089
```

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?

```
p_3 = pnorm(mu + 100, mu, sigma/sqrt(15)) - pnorm(mu - 100, mu, sigma/sqrt(15))
cat('P(9900 < X < 10100) =', p_2)
```

```
## P(9900 < X < 10100) = 0.9715403
```

```
z_c = 100/ (sigma / sqrt(15))
cat('\nz =', z_b)
```

```
##
## z = 2.19089
```

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?.

```
z_d = -200/ (sigma / sqrt(120))
z_d
```

```
## [1] -4.38178
```

Hizo lo correcto, pues su media se encontraba a 4.38 desviaciones estándar, lo cual es muy alejado de la media poblacional.

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

$H_0 : \mu = 10000$ $H_1 : \mu < 10000$

```
z_e = -75/ (sigma / sqrt(120))
cat('z =', z_e)
```

```
## z = -1.643168
```

```
cat('\nP(X < 9925)=',pnorm(9925, mu, sigma/sqrt(120)))
```

```
##
## P(X < 9925)= 0.05017412
```

Como el valor \$ p $<$ 0.1\$, se rechaza H_0, por lo que recomendaría al ingeniero rechazar la oferta.

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.

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?

```r
mu =15
sigma = 1
z_1 = abs(qnorm(0.05/2))
z_1
```

```
## [1] 1.959964
```

A 1.96 desviaciones estándar.

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

```r
1 - pnorm(16, mu, sigma/sqrt(10))
```

```
## [1] 0.0007827011
```

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?

```r
z_c = 1/ (sigma / sqrt(10))
z_c
```

```
## [1] 3.162278
```

Se encuentra a 3.16 desviaciones estándar, por lo que está muy alejada de la media poblacional y se detendría la producción.

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

```r
pnorm(14.5, mu, sigma/sqrt(10))
```

```
## [1] 0.05692315
```

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?

$H_0 : \mu = 15$ $H_1 : \mu > 15$

```r
z_5 = 0.5/ (sigma/ sqrt(10))
cat('z =', z_5,'\n')
```

```
## z = 1.581139
```

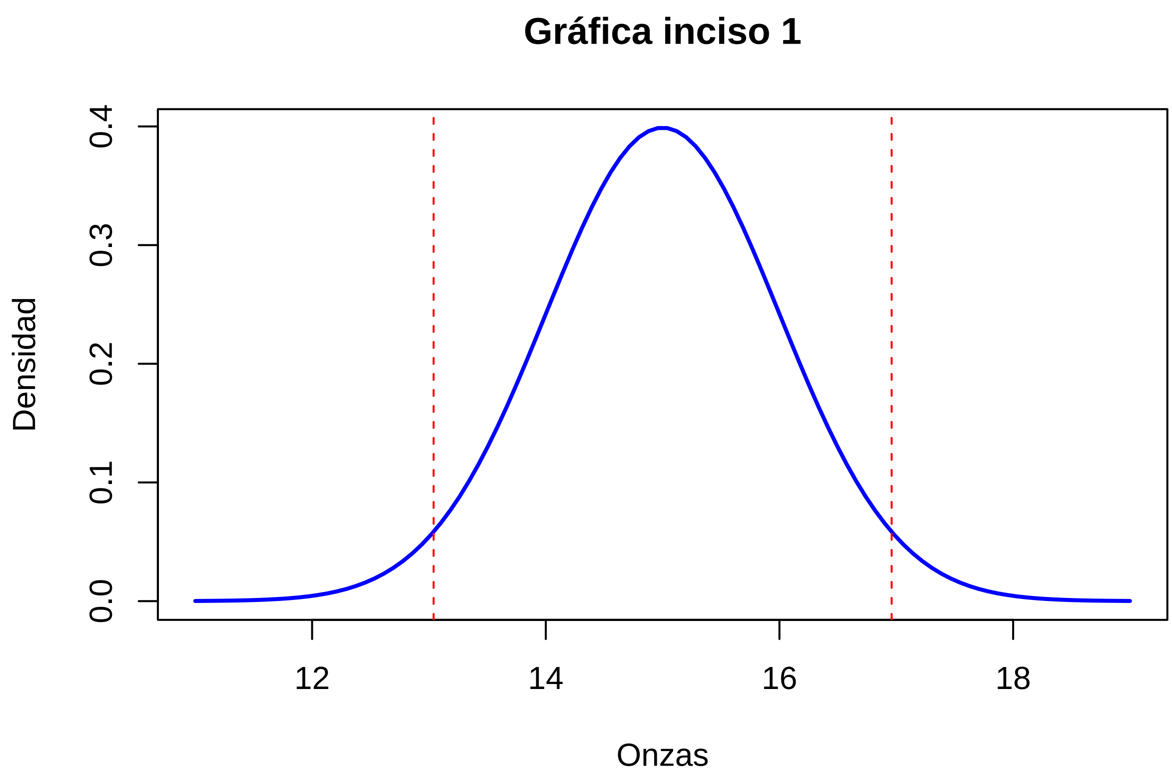
```r
cat('Valor p =', 1 - pnorm(15.5, mu, sigma/sqrt(10)))
```

```
## Valor p = 0.05692315
```

Como el valor \$ p < 0.1\$, se rechaza H_0, por lo que se debería detener la producción para calibrar la máquina.

6. Hacer una gráfica del inciso 1.

```r
x <- seq(mu - 4*sigma, mu + 4*sigma, length=100)
y <- dnorm(x, mean = mu, sd = sigma)
plot(x, y, type="l", lwd=2, col="blue",
     main="Gráfica inciso 1",
     xlab="Onzas", ylab="Densidad")
abline(v = mu - z_1*sigma, col="red", lty=8)
abline(v = mu + z_1*sigma, col="red", lty=8)
```



En la gráfica, las líneas rojas marca el 95% de los datos, los cuales se encuentran alrededor de entre 13 y 17.