Act6 Distribuciones Muestrales y TCL

Luis Maximiliano López Ramírez

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:

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.

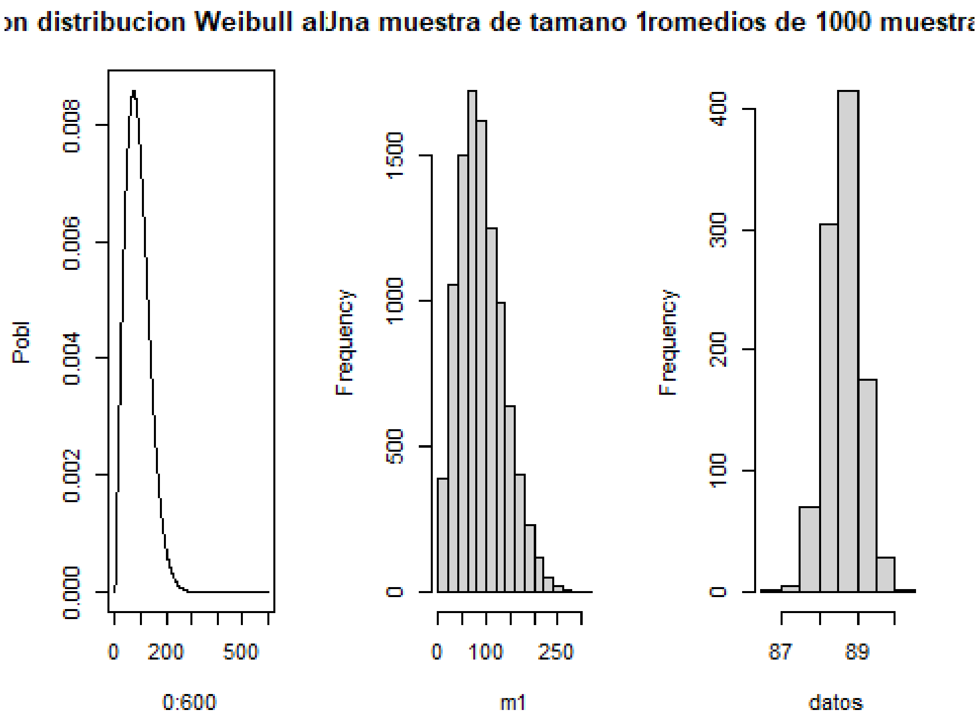
```r
par(mfrow=c(1,3))
# Graficando una distribucion Weibull de alfa =2, beta = 100
Pobl = dweibull(0:600,2, 100)
plot(0:600,Pobl, type="l", main = "Poblacion con distribucion Weibull alfa =2, beta = 100")

# 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")
```



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(e1071)
library(nortest)

sesgo_muestra <- skewness(m1)
curtosis_muestra <- kurtosis(m1)

cat("Sesgo de la muestra de tamaño 10000:", sesgo_muestra, "\n")

## Sesgo de la muestra de tamaño 10000: 0.6182247

cat("Curtosis de la muestra de tamaño 10000:", curtosis_muestra, "\n")

## Curtosis de la muestra de tamaño 10000: 0.1657881

prueba_normalidad_muestra <- ad.test(m1)
cat("Prueba de normalidad de Anderson-Darling para la muestra de tamaño
10000:\n")

## Prueba de normalidad de Anderson-Darling para la muestra de tamaño
10000:

cat("  Estadístico de la prueba:", prueba_normalidad_muestra$statistic,
"\n")
```

```
##   Estadístico de la prueba: 58.31972

cat("  Valor p:", prueba_normalidad_muestra$p.value, "\n")

##   Valor p: 3.7e-24
```

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.

```
sesgo_medias <- skewness(datos)
curtosis_medias <- kurtosis(datos)

cat("Sesgo de las medias de las 1000 muestras:", sesgo_medias, "\n")

## Sesgo de las medias de las 1000 muestras: 0.04787878

cat("Curtosis de las medias de las 1000 muestras:", curtosis_medias,
"\n")

## Curtosis de las medias de las 1000 muestras: 0.1613216

prueba_normalidad_medias <- ad.test(datos)
cat("Prueba de normalidad de Anderson-Darling para las medias de las 1000
muestras:\n")

## Prueba de normalidad de Anderson-Darling para las medias de las 1000
muestras:

cat("  Estadístico de la prueba:", prueba_normalidad_medias$statistic,
"\n")

##   Estadístico de la prueba: 0.4270901

cat("  Valor p:", prueba_normalidad_medias$p.value, "\n")

##   Valor p: 0.3124558
```

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.

Distribución Weibull con parámetros modificados (alfa = 1.5, beta = 50)

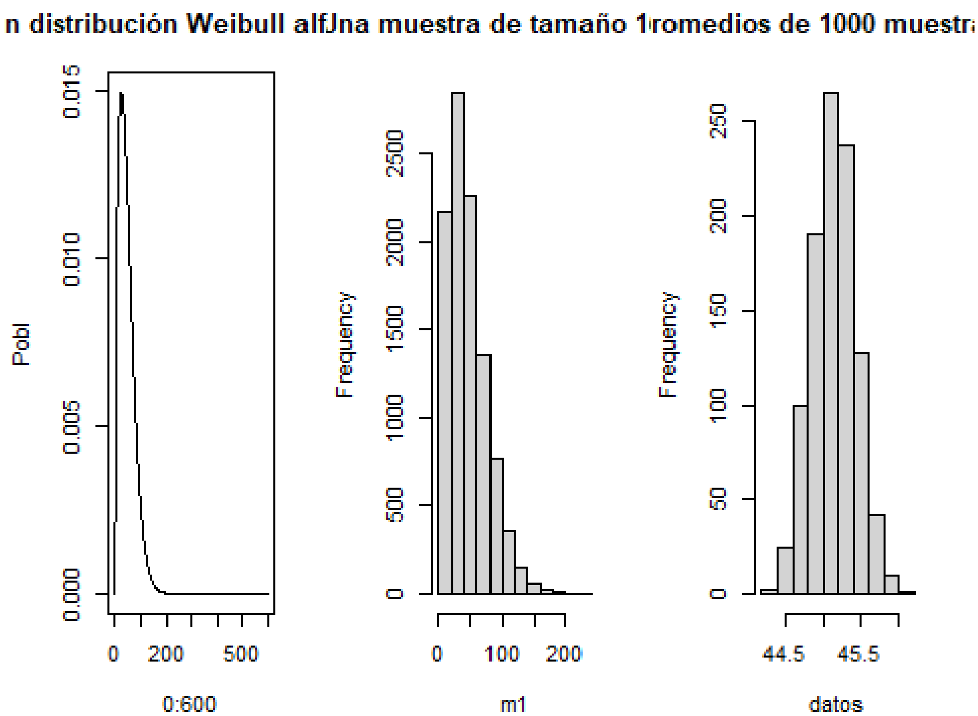
```
library(e1071)
library(nortest)

par(mfrow=c(1,3))
```

```r
# Graficando una distribución Weibull de alfa = 1.5, beta = 50
Pobl = dweibull(0:600, 1.5, 50)
plot(0:600, Pobl, type="l", main = "Población con distribución Weibull
alfa = 1.5, beta = 50")

# Tomando una muestra de 10000 elementos tomados al azar
m1 = rweibull(10000, 1.5, 50)
hist(m1, main = "Una muestra de tamaño 10000")

# Tomando 1000 promedios de las 1000 muestras como la anterior
datos = replicate(1000, mean(rweibull(10000, 1.5, 50)))
hist(datos, main="Gráfica de los promedios de 1000 muestras de tamaño
10000")
```




```r
# Sesgo y curtosis de la muestra de tamaño 10000
sesgo_muestra <- skewness(m1)
curtosis_muestra <- kurtosis(m1)

cat("Sesgo de la muestra de tamaño 10000 (Weibull alfa=1.5, beta=50):",
sesgo_muestra, "\n")

## Sesgo de la muestra de tamaño 10000 (Weibull alfa=1.5, beta=50):
1.074252

cat("Curtosis de la muestra de tamaño 10000 (Weibull alfa=1.5,
beta=50):", curtosis_muestra, "\n")
```

```
## Curtosis de la muestra de tamaño 10000 (Weibull alfa=1.5, beta=50): 
1.439493

# Prueba de normalidad para la muestra de tamaño 10000 (Anderson-Darling)
prueba_normalidad_muestra <- ad.test(m1)
cat("Prueba de normalidad de Anderson-Darling para la muestra de tamaño 
10000 (Weibull alfa=1.5, beta=50):\n")

## Prueba de normalidad de Anderson-Darling para la muestra de tamaño 
10000 (Weibull alfa=1.5, beta=50):

cat("  Estadístico de la prueba:", prueba_normalidad_muestra$statistic, 
"\n")

##   Estadístico de la prueba: 156.549

cat("  Valor p:", prueba_normalidad_muestra$p.value, "\n")

##   Valor p: 3.7e-24

# Sesgo y curtosis de las medias de las 1000 muestras
sesgo_medias <- skewness(datos)
curtosis_medias <- kurtosis(datos)

cat("Sesgo de las medias de las 1000 muestras (Weibull alfa=1.5, 
beta=50):", sesgo_medias, "\n")

## Sesgo de las medias de las 1000 muestras (Weibull alfa=1.5, beta=50): 
0.08190298

cat("Curtosis de las medias de las 1000 muestras (Weibull alfa=1.5, 
beta=50):", curtosis_medias, "\n")

## Curtosis de las medias de las 1000 muestras (Weibull alfa=1.5, 
beta=50): -0.1703769

# Prueba de normalidad para las medias de las 1000 muestras (Anderson-
Darling)
prueba_normalidad_medias <- ad.test(datos)
cat("Prueba de normalidad de Anderson-Darling para las medias de las 1000 
muestras (Weibull alfa=1.5, beta=50):\n")

## Prueba de normalidad de Anderson-Darling para las medias de las 1000 
muestras (Weibull alfa=1.5, beta=50):

cat("  Estadístico de la prueba:", prueba_normalidad_medias$statistic, 
"\n")

##   Estadístico de la prueba: 0.2712722

cat("  Valor p:", prueba_normalidad_medias$p.value, "\n")

##   Valor p: 0.6728363
```

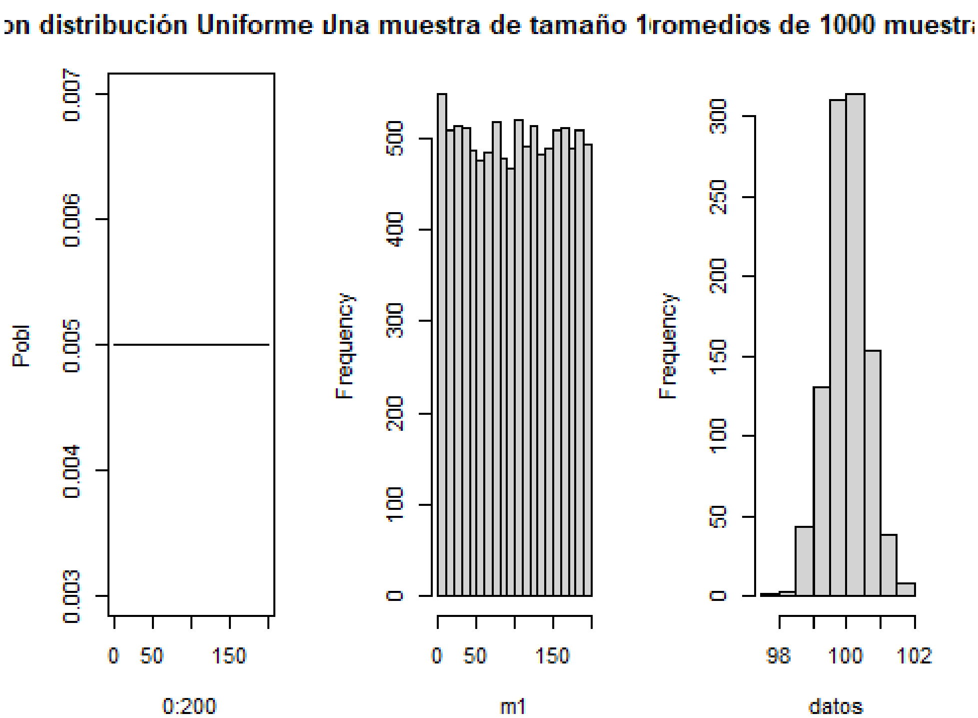
Distribución Uniforme (min = 0, max = 200)

```r
par(mfrow=c(1,3))

# Graficando una distribución uniforme
Pobl = dunif(0:200, min = 0, max = 200)
plot(0:200, Pobl, type="l", main = "Población con distribución Uniforme
min=0, max=200")

# Tomando una muestra de 10000 elementos tomados al azar
m1 = runif(10000, min = 0, max = 200)
hist(m1, main = "Una muestra de tamaño 10000")

# Tomando 1000 promedios de las 1000 muestras como la anterior
datos = replicate(1000, mean(runif(10000, min = 0, max = 200)))
hist(datos, main="Gráfica de los promedios de 1000 muestras de tamaño
10000")
```




```r
# Sesgo y curtosis de la muestra de tamaño 10000
sesgo_muestra <- skewness(m1)
curtosis_muestra <- kurtosis(m1)

cat("Sesgo de la muestra de tamaño 10000 (Uniforme min=0, max=200):",
sesgo_muestra, "\n")

## Sesgo de la muestra de tamaño 10000 (Uniforme min=0, max=200): -
0.00469439
```

```r
cat("Curtosis de la muestra de tamaño 10000 (Uniforme min=0, max=200):",
curtosis_muestra, "\n")

## Curtosis de la muestra de tamaño 10000 (Uniforme min=0, max=200): -
1.214609

# Prueba de normalidad para la muestra de tamaño 10000 (Anderson-Darling)
prueba_normalidad_muestra <- ad.test(m1)
cat("Prueba de normalidad de Anderson-Darling para la muestra de tamaño
10000 (Uniforme min=0, max=200):\n")

## Prueba de normalidad de Anderson-Darling para la muestra de tamaño
10000 (Uniforme min=0, max=200):

cat("  Estadístico de la prueba:", prueba_normalidad_muestra$statistic,
"\n")

##   Estadístico de la prueba: 116.8129

cat("  Valor p:", prueba_normalidad_muestra$p.value, "\n")

##   Valor p: 3.7e-24

# Sesgo y curtosis de las medias de las 1000 muestras
sesgo_medias <- skewness(datos)
curtosis_medias <- kurtosis(datos)

cat("Sesgo de las medias de las 1000 muestras (Uniforme min=0,
max=200):", sesgo_medias, "\n")

## Sesgo de las medias de las 1000 muestras (Uniforme min=0, max=200): -
0.03304625

cat("Curtosis de las medias de las 1000 muestras (Uniforme min=0,
max=200):", curtosis_medias, "\n")

## Curtosis de las medias de las 1000 muestras (Uniforme min=0, max=200):
0.1327517

# Prueba de normalidad para las medias de las 1000 muestras (Anderson-
Darling)
prueba_normalidad_medias <- ad.test(datos)
cat("Prueba de normalidad de Anderson-Darling para las medias de las 1000
muestras (Uniforme min=0, max=200):\n")

## Prueba de normalidad de Anderson-Darling para las medias de las 1000
muestras (Uniforme min=0, max=200):

cat("  Estadístico de la prueba:", prueba_normalidad_medias$statistic,
"\n")

##   Estadístico de la prueba: 0.3142734
```

```
cat("  Valor p:", prueba_normalidad_medias$p.value, "\n")

##   Valor p: 0.5448329
```

Inciso E

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

Semejanzas:

Gráfica de las medias de las 1000 muestras: En las tres distribuciones (Weibull con diferentes parámetros y Uniforme), las gráficas de los promedios de las 1000 muestras se aproximan a una distribución normal, demostrando el efecto del Teorema Central del Límite.

Normalidad en las medias: A pesar de la falta de normalidad en las muestras individuales de 10,000 datos, todas las distribuciones muestran que los promedios de las 1000 muestras son normales, como lo indican las pruebas de normalidad.

Diferencias:

Gráfica de la distribución original: Las distribuciones originales muestran formas distintas: La Weibull con alfa = 2, beta = 100 tiene una forma sesgada a la derecha pero menos pronunciada. La Weibull con alfa = 1.5, beta = 50 es más asimétrica y tiene una cola más larga hacia la derecha. La Uniforme es plana, sin sesgo evidente.

Muestras individuales: Las muestras de 10,000 datos reflejan las características de sus distribuciones originales, con sesgo en las Weibull y una distribución uniforme en la muestra de la distribución uniforme.

 # 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?

P(9900 < X < 10100)

```
miu = 10000
sigma = 500
```

```r
p1 = pnorm(10100, miu, sigma) - pnorm(9900, miu, sigma)

z = (10100 - miu) / sigma

cat("P(9900 < X < 10100) =", p1, "y se encuentra a", z, "desviaciones estándar")

## P(9900 < X < 10100) = 0.1585194 y se encuentra a 0.2 desviaciones estándar
```

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?

$$P(9900 < \bar{x} < 10100)$$

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

```r
muestra = 120
miu_x = miu
sigma_x = sigma / sqrt(muestra)

p2 = pnorm(10100, miu_x, sigma_x) - pnorm(9900, miu_x, sigma_x)

z2 = (10100 - miu_x) / sigma_x

cat("P(9900 < X < 10100) =", p2, "y se encuentra a", z2, "desviaciones estándar")

## P(9900 < X < 10100) = 0.9715403 y se encuentra a 2.19089 desviaciones estándar
```

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?

$$P(9900 < \bar{x} < 10100)$$

$$\bar{X} \sim N\left(\mu_{\bar{x}} = 10000, \sigma_{\bar{x}} = \frac{500}{\sqrt{15}}\right)$$

```r
muestra = 15
miu_x = miu
sigma_x = sigma / sqrt(muestra)

p2 = pnorm(10100, miu_x, sigma_x) - pnorm(9900, miu_x, sigma_x)
```

```r
z2 = (10100 - miu_x) / sigma_x

cat("P(9900 < X < 10100) =", p2, "y se encuentra a", z2, "desviaciones estándar")

## P(9900 < X < 10100) = 0.561422 y se encuentra a 0.7745967 desviaciones estándar
```

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?.

$$H_0: \mu = 10000 \text{ lb/pulg}^2 \quad \text{(La media de la resistencia es de 10,000 lb/pulg}^2\text{)}$$

$$H_A: \mu < 10000 \text{ lb/pulg}^2 \quad \text{(La media de la resistencia es menor de 10,000 lb/pulg}^2\text{)}$$

```r
muestra = 120  # tamaño de la muestra
miu_x = 9800  # media muestral
miu = 10000  # media poblacional bajo H0
sigma = 500  # desviación estándar asumida de la población

# Cálculo del estadístico de prueba
sigma_x = sigma / sqrt(muestra)
z <- (miu_x - miu) / sigma_x

# Valor p para la prueba bilateral
valor_p <- 2 * pnorm(z, lower.tail = TRUE)

# Resultados
cat("Valor de z", z, "y valor p", valor_p)

## Valor de z -4.38178 y valor p 1.177134e-05

print('')

## [1] ""

# Decisión
nivel_significancia = 0.05
if (valor_p < nivel_significancia) {
  decision <- "Rechazar H0: El ingeniero hizo lo correcto al rechazar el lote."
} else {
  decision <- "No rechazar H0: El ingeniero no hizo lo correcto al rechazar el lote."
}
```

```r
cat(decision)

## Rechazar H0: El ingeniero hizo lo correcto al rechazar el lote.
```

Inciso E

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

$$H_0: \mu = 10000 \text{ lb/pulg}^2 \quad (\text{La media de la resistencia es de } 10{,}000 \text{ lb/pulg}^2)$$

$$H_A: \mu < 10000 \text{ lb/pulg}^2 \quad (\text{La media de la resistencia es menor de } 10{,}000 \text{ lb/pulg}^2)$$

```r
muestra = 120  # tamaño de la muestra
miu_x = 9925  # media muestral
miu = 10000  # media poblacional bajo H0
sigma = 500  # desviación estándar asumida de la población

# Cálculo del estadístico de prueba
sigma_x = sigma / sqrt(muestra)
z <- (miu_x - miu) / sigma_x

# Valor p para la prueba bilateral
valor_p <- 2 * pnorm(z, lower.tail = TRUE)

# Resultados
cat("Valor de z", z, "y valor p", valor_p)

## Valor de z -1.643168 y valor p 0.1003482

print('')

## [1] ""

# Decisión
nivel_significancia = 0.05
if (valor_p < nivel_significancia) {
  decision <- "Rechazar H0: Se recomienda rechazar el lote"
} else {
  decision <- "No rechazar H0: Se recomienda NO rechazar el lote"
}

cat(decision)

## No rechazar H0: Se recomienda NO rechazar el lote
```

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 σ = 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 A

¿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?

```
# Valores necesarios
mu <- 15
sigma <- 1
n <- 10
z_alpha_2 <- qnorm(0.975)  # Valor crítico para el 95% de confianza

cat("Valor de z con alpha = 0.05/2:", z_alpha_2, "\n")

## Valor de z con alpha = 0.05/2: 1.959964

# Rango de desviaciones estándar
desviacion_estandar <- z_alpha_2 * sigma / sqrt(n)
cat("Rango de desviaciones estándar alrededor de la media verdadera:",
desviacion_estandar, "\n")

## Rango de desviaciones estándar alrededor de la media verdadera:
0.619795
```

Inciso B

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

```
# Media muestral observada
X_bar <- 16

# Cálculo del Z
Z <- (X_bar - mu) / (sigma / sqrt(n))

# Probabilidad de que la media sea mayor a 16 onzas
probabilidad_mayor_16 <- 1 - pnorm(Z)
cat("Probabilidad de que la media muestral sea mayor a 16 onzas:",
probabilidad_mayor_16, "\n")

## Probabilidad de que la media muestral sea mayor a 16 onzas:
0.0007827011
```

Inciso C

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

```
if (abs(Z) > z_alpha_2) {
  cat("La producción se detendría para calibrar la máquina.\n")
} else {
  cat("La producción no se detendría.\n")
}

## La producción se detendría para calibrar la máquina.

print(Z)

## [1] 3.162278
```

Inciso D

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

```
# Media muestral observada
X_bar_14_5 <- 14.5

# Cálculo del Z
Z_14_5 <- (X_bar_14_5 - mu) / (sigma / sqrt(n))

# Probabilidad de que la media sea menor a 14.5 onzas
probabilidad_menor_14_5 <- pnorm(Z_14_5)
cat("Probabilidad de que la media muestral sea menor a 14.5 onzas:",
probabilidad_menor_14_5, "\n")

## Probabilidad de que la media muestral sea menor a 14.5 onzas:
0.05692315
```

Inciso E

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?

```
# Media muestral observada
X_bar_15_5 <- 15.5

# Cálculo del Z
Z_15_5 <- (X_bar_15_5 - mu) / (sigma / sqrt(n))

if (abs(Z_15_5) > z_alpha_2) {
  cat("La producción se detendría para calibrar la máquina.\n")
} else {
```

```
  cat("La producción no se detendría.\n")
}
```

```
## La producción no se detendría.
```

```
print(Z_15_5)
```

```
## [1] 1.581139
```

Inciso F

Hacer una gráfica del inciso 1.

```
# Crear secuencia de posibles valores de la media muestral
x <- seq(mu - 4*desviacion_estandar, mu + 4*desviacion_estandar,
length=100)

# Función de densidad para la distribución de la media muestral
densidad <- dnorm(x, mean = mu, sd = sigma/sqrt(n))

# Gráfico
plot(x, densidad, type = "l", lwd = 2, main = "Distribución de la media
muestral",
     xlab = "Media muestral", ylab = "Densidad")
abline(v = c(mu - desviacion_estandar, mu + desviacion_estandar), col =
"red", lty = 2)
text(mu, max(densidad)*0.8, labels = "95% central", col = "red")
```