M1_A6

Sofia Cantu

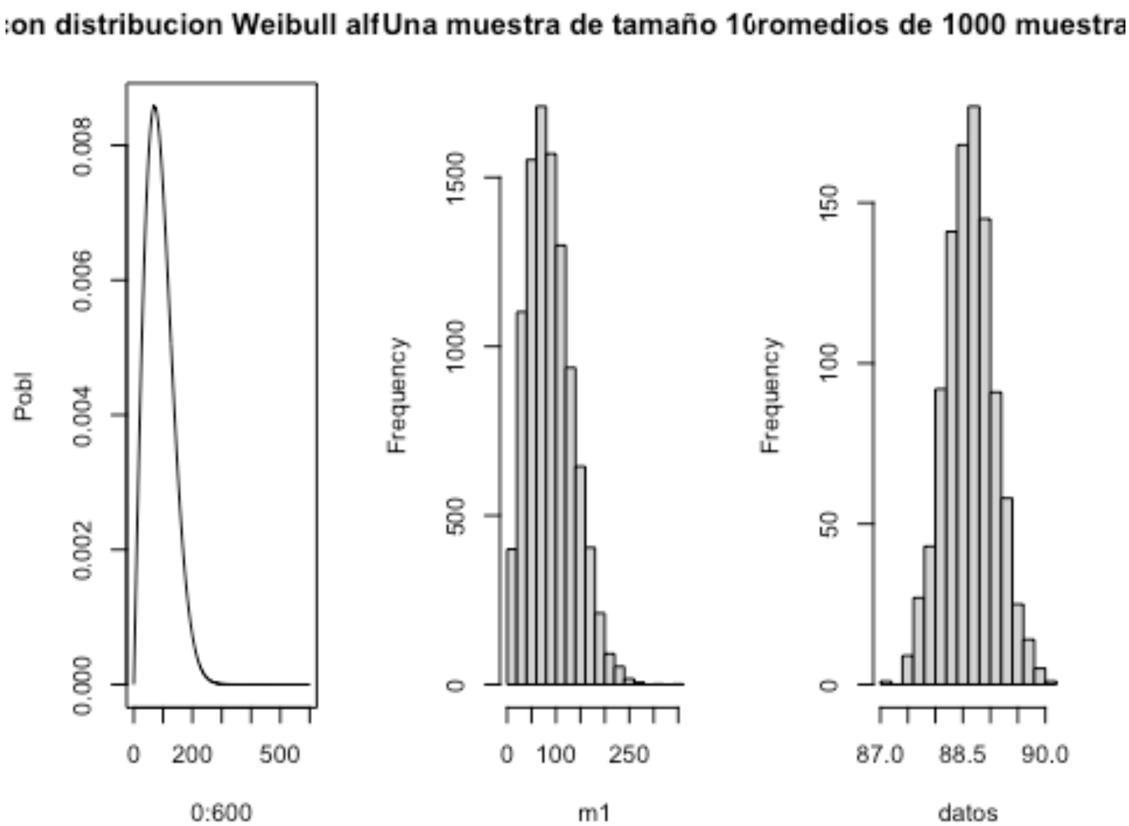
2024-08-16

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) Ejecutar el siguiente código de R: 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
# Establecer las gráficas en una sola fila y tres columnas
par(mfrow=c(1,3))

# Graficando una distribución 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 tamaño 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 tamaño
10,000")
```



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.

```
# Cálculo del sesgo y curtosis de la muestra de tamaño 10000
library(e1071)
library(nortest)

# Cálculo del sesgo y curtosis de la muestra de tamaño 10000
sesgo_m1 = skewness(m1)
curtosis_m1 = kurtosis(m1)

cat("Sesgo de la muestra de tamaño 10000: ", sesgo_m1, "\n")

## Sesgo de la muestra de tamaño 10000:  0.6186257

cat("Curtosis de la muestra de tamaño 10000: ", curtosis_m1, "\n")

## Curtosis de la muestra de tamaño 10000:  0.1848948

# Prueba de normalidad
test_normalidad = lillie.test(m1)
cat("P-valor de la prueba de normalidad: ", test_normalidad$p.value, "\n")

## P-valor de la prueba de normalidad:  1.59871e-64
```

Los datos presentan un sesgo positivo (valores más extremos en el lado derecho) y una curtosis ligeramente elevada (los datos están más concentrados alrededor de la media y con colas más pesadas que una distribución normal), lo que indica una distribución asimétrica y más concentrada que la normal. La normalidad se compara con una distribución normal, que tiene un sesgo de 0 (indica simetría) y una curtosis de 3. La prueba de normalidad con un p-valor extremadamente bajo confirma que los datos no siguen una distribución normal. Menor a 0.05 tiende a no seguir una distribución 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.

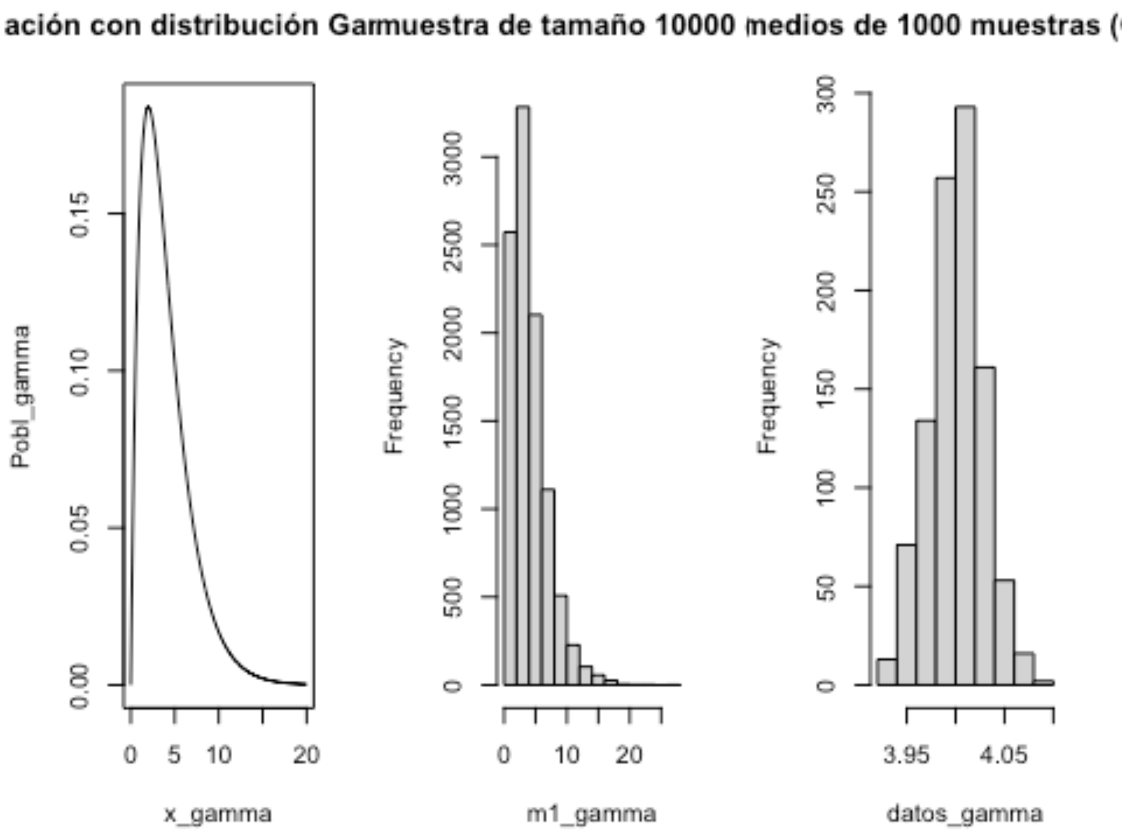
```
# Cálculo del sesgo y curtosis de las medias de las 1000 muestras
# Cálculo del sesgo y curtosis de las medias de las 1000 muestras
sesgo_medias = skewness(datos)
curtosis_medias = kurtosis(datos)

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

## Sesgo de las medias de 1000 muestras:  0.07675593

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

## Curtosis de las medias de 1000 muestras:  0.03003478

# Prueba de normalidad de las medias de las 1000 muestras
test_normalidad_medias = lillie.test(datos)
cat("P-valor de la prueba de normalidad para las medias: ",
test_normalidad_medias$p.value, "\n")

## P-valor de la prueba de normalidad para las medias:  0.7900481
```

Aqui el sesgo negativo es cercano a 0, lo que indica que la distribución de las medias es casi simétrica. Esta curtosis negativa, también cercana a 0, sugiere que la distribución de las medias es ligeramente menos concentrada alrededor de la media y el p-valor alto nos da a saber que las medias de las muestras se pueden considerar normalmente distribuidas.

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.

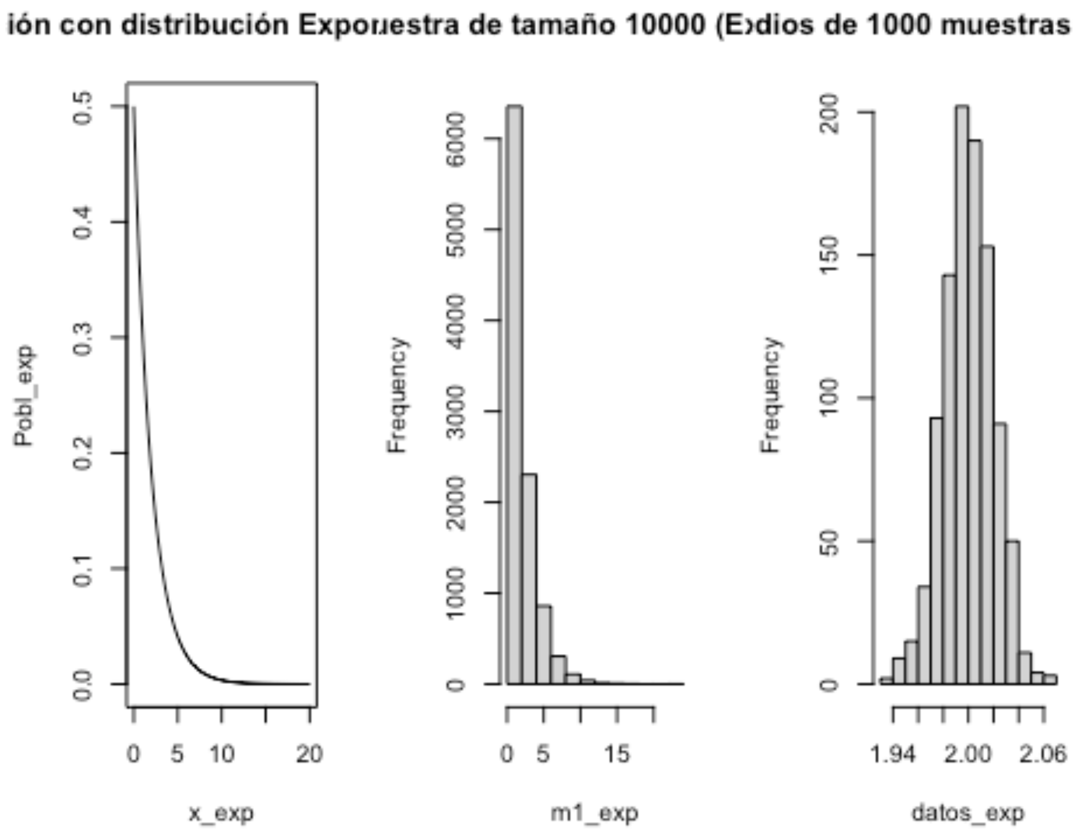
```
###D) Repite el procedimiento A, B y C para otras dos distribuciones que no
sean simétricas.

# Distribución Gamma
par(mfrow=c(1,3))

# Población con distribución Gamma (shape = 2, scale = 2)
x_gamma = seq(0, 20, length.out = 600)
Pobl_gamma = dgamma(x_gamma, shape = 2, scale = 2)
plot(x_gamma, Pobl_gamma, type="l", main = "Población con distribución
```

```
Gamma(2,2)")

# Tomando una muestra de 10000 elementos tomados al azar
m1_gamma = rgamma(10000, shape = 2, scale = 2)
hist(m1_gamma, main = "Una muestra de tamaño 10000 (Gamma)")

# Tomando 1000 promedios de las 1000 muestras
datos_gamma = numeric(1000)
for(i in 1:1000) {
    m_gamma = rgamma(10000, shape = 2, scale = 2)
    datos_gamma[i] = mean(m_gamma)
}
hist(datos_gamma, main="Promedios de 1000 muestras (Gamma)")
```




```
# Cálculos para la distribución Gamma
sesgo_gamma = skewness(m1_gamma)
curtosis_gamma = kurtosis(m1_gamma)
test_normalidad_gamma = lillie.test(m1_gamma)
sesgo_medias_gamma = skewness(datos_gamma)
curtosis_medias_gamma = kurtosis(datos_gamma)
test_normalidad_medias_gamma = lillie.test(datos_gamma)

cat("Gamma - Muestra:\n")
```

```r
## Gamma - Muestra:

cat("Sesgo:", sesgo_gamma, "Curtosis:", curtosis_gamma, "\n")

## Sesgo: 1.445772 Curtosis: 3.204723

cat("P-valor normalidad:", test_normalidad_gamma$p.value, "\n\n")

## P-valor normalidad: 9.903999e-256

cat("Gamma - Medias:\n")

## Gamma - Medias:

cat("Sesgo:", sesgo_medias_gamma, "Curtosis:", curtosis_medias_gamma, "\n")

## Sesgo: -0.02849431 Curtosis: -3.92653e-05

cat("P-valor normalidad:", test_normalidad_medias_gamma$p.value, "\n\n")

## P-valor normalidad: 0.2381421

# Distribución Exponencial
par(mfrow=c(1,3))

# Población con distribución Exponencial (rate = 0.5)
x_exp = seq(0, 20, length.out = 600)
Pobl_exp = dexp(x_exp, rate = 0.5)
plot(x_exp, Pobl_exp, type="l", main = "Población con distribución
Exponencial(0.5)")

# Tomando una muestra de 10000 elementos tomados al azar
m1_exp = rexp(10000, rate = 0.5)
hist(m1_exp, main = "Una muestra de tamaño 10000 (Exponencial)")

# Tomando 1000 promedios de las 1000 muestras
datos_exp = numeric(1000)
for(i in 1:1000) {
    m_exp = rexp(10000, rate = 0.5)
    datos_exp[i] = mean(m_exp)
}
hist(datos_exp, main="Promedios de 1000 muestras (Exponencial)")
```




```r
# Cálculos para la distribución Exponencial
sesgo_exp = skewness(m1_exp)
curtosis_exp = kurtosis(m1_exp)
test_normalidad_exp = lillie.test(m1_exp)
sesgo_medias_exp = skewness(datos_exp)
curtosis_medias_exp = kurtosis(datos_exp)
test_normalidad_medias_exp = lillie.test(datos_exp)

cat("Exponencial - Muestra:\n")

## Exponencial - Muestra:

cat("Sesgo:", sesgo_exp, "Curtosis:", curtosis_exp, "\n")

## Sesgo: 2.029038 Curtosis: 6.482234

cat("P-valor normalidad:", test_normalidad_exp$p.value, "\n\n")

## P-valor normalidad: 0

cat("Exponencial - Medias:\n")

## Exponencial - Medias:

cat("Sesgo:", sesgo_medias_exp, "Curtosis:", curtosis_medias_exp, "\n")
```

```
## Sesgo: -0.07667846 Curtosis: 0.1391419

cat("P-valor normalidad:", test_normalidad_medias_exp$p.value, "\n\n")

## P-valor normalidad: 0.2956045
```

Para la Distribución Gamma, podemos resumir lo siguiente:

Muestra de Tamaño 10,000 La muestra presenta un sesgo positivo pero no cercano a 0 (1.391524), esto nos dice que esta fuertemente inclinada hacia la derecha y es consistente con lo que nos muestra la primera grafica. Una curtosis mayor a la normal (2.751204), cercano a tres nos da más valores extremos. La distribución de los datos no siguen una distribución normal, por tener un numero tan pequeño en la normalidad.

Medias de 1,000 Muestras El sesgo por el 0 nos dice que es simetrico, curtosis cercana a 0 indica una distribución de las medias ligeramente menos concentrada que la normal y la normalidad es buena (mayor a 0.05)

Distribución Exponencial:

Muestra de Tamaño 10,000 sesgo positivo alto indica inclinamiento hacia la derecha. Curtosis muy elevada indica que la distribución tiene colas extremadamente pesadas y el p-valor indica que la distribución no es normal

Medias de 1,000 Muestras Sesgo ligeramente negativo, ligeramente inclinacion a la izquierda. Curtosis negativa y cercana a 0 indica una distribución ligeramente menos concentrada en la media que una distribución normal. El p-valor es bueno, se puede asumir normalidad.

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

En todas las distribuciones, las medias de las muestras tienden hacia una distribución normal, independientemente de la forma de la distribución original, lo que confirma el Teorema Central del Límite. Por otra parte, las distribuciones Weibull, Gamma y Exponencial presentan sesgos y curtosis específicos, reflejando sus formas asimétricas, generalmente sesgadas a la derecha y con colas más pesadas que una distribución normal.

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

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?

$\mathrm{P}(9900 < \mathrm{X} < 10100)$

```
p1 = pnorm(10100, 10000, 500) - pnorm(9900, 10000, 500)
cat("P(9900 < X < 10100)=", p1)

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

z1 = 100/500
cat("z = ", z1)

## 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?

$P\left(9900 < \bar{[x]} < 101000\right) \bar{x} \sim N\left(\mu_{\bar{x}} = 10000, \sigma_{\bar{x}} = \frac{500}{\sqrt{120}}\right)$

```
p2 = pnorm(10100, 10000, 500/sqrt(120)) - pnorm(9900, 10000, 500/sqrt(120))
cat("P(9900 < x_b < 10100) = ", p2)

## P(9900 < x_b < 10100) =  0.9715403

z2 = 100/(500/sqrt(120))
cat("z = ", z2)

## 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?

```
p3 = pnorm(10100, 10000, 500/sqrt(15)) - pnorm(9900, 10000, 500/sqrt(15))
cat("P(9900 < x_b < 10100) = ", p3)

## P(9900 < x_b < 10100) =  0.561422

z3 = 100/(500/sqrt(15))
cat("z = ", z3)

## z =  0.7745967
```

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?.

```
z4 = (9800 - 10000) / (500 / sqrt(120))
cat("Valor z = ", z4)
```

```
## Valor z =  -4.38178
```

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

```
z5 = (9925 - 10000) / (500 / sqrt(120))
cat("Valor z = ", z5)
```

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

#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.

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?

```
mu = 15
sigma = 1

z1 = qnorm(0.975)
cat("Desviaciones estándar alrededor de la media para el 95%: z =", z1)
```

```
## Desviaciones estándar alrededor de la media para el 95%: z = 1.959964
```

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

```
n2 = 10
sigma_xb2 = sigma / sqrt(n2)

p2 = 1 - pnorm(16, mu, sigma_xb2)
cat("P(x_b > 16) =", p2)
```

```
## P(x_b > 16) = 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?

```
xb3 = 16

z3 = (xb3 - mu) / (sigma / sqrt(10))
cat("z =", z3)

## z = 3.162278

if (z3 > z1) {
  cat("\n" ,"si se detendría la producción")
} else {
  cat("\n" ,"no se detendrá detener la producción")
}

##
##  si 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?

```
p4 = pnorm(14.5, mu, sigma_xb2)
cat("P(x_b < 14.5) =", p4)

## P(x_b < 14.5) = 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?

```
xb5 = 15.5

z5 = (xb5 - mu) / (sigma / sqrt(10))
cat("z =", z5)

## z = 1.581139

if (z5 > z1) {
  cat("\n" ,"si se detendría la producción")
} else {
  cat("\n" ,"no se detendrá detener la producción")
}

##
##  no se detendrá detener la producción
```

6) Hacer una gráfica del inciso 1.

```
x = seq(mu - 4 * sigma, mu + 4 * sigma, length=100)
y = dnorm(x, mu, sigma / sqrt(10))

plot(x, y, type="l", lwd=2, col="purple",
     main="Distribución de la Media con Intervalo del 95%",
     xlab="Media muestral (onzas)", ylab="Densidad")
abline(v=mu - z1 * sigma / sqrt(10), col="pink", lwd=2, lty=2)
```

```
abline(v=mu + z1 * sigma / sqrt(10), col="pink", lwd=2, lty=2)
legend("topright", legend=c("Intervalo del 95%"),
       col="pink", lty=2, lwd=2)
```