

Actividad 6. Distribuciones Muestrales y TCL

José Carlos Sánchez Gómez

2024-08-17

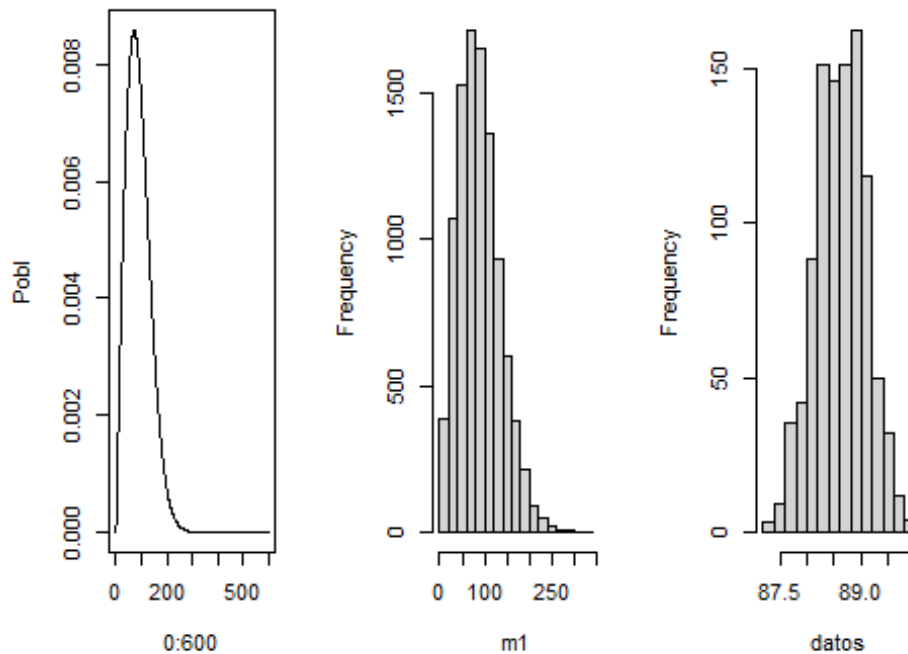
Pregunta 1

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:

- Ejecutar 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 $\alpha = 2$ 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")
```

Gráfica de la distribución Weibull con $\alpha = 2$, $\beta = 100$. Una muestra de tamaño 10,000 y los promedios de 1000 muestras de tamaño 10,000.



En la primera gráfica podemos ver la distribución teórica de Weibull, la cual es asimétrica hacia la derecha, con una cola hacia la derecha. La segunda es un histograma generado con 10,000 valores generados aleatoriamente de la distribución de Weibull, la cual es similar a la gráfica anterior. La última gráfica es el promedio de 1000 muestras, y cada una de ellas es de 10,000 datos, la cuál cumple con el Teorema Central del Limite, la cual dice que independientemente de la distribución original, la distribución de los promedios de un gran número de muestras se asemeja a una distribución normal.

- **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)
cat("Sesgo de la muestra de tamaño 10000: ", skewness(m), "\n")
## Sesgo de la muestra de tamaño 10000: 0.6451094
cat("Curtosis de la muestra de tamaño 10000: ", kurtosis(m), "\n")
## Curtosis de la muestra de tamaño 10000: 0.3297637
ad.test(m)
##
## Anderson-Darling normality test
##
## data: m
## A = 55.195, p-value < 2.2e-16
```

El sesgo dentro de esta muestra nos dice que la distribución tiene una asimetría hacia la derecha, en otras palabras, que tiene una cola muy pronunciada hacia la derecha. El curtosis nos dice que la gran mayoría de los datos se encuentran dentro del pico central. Debido a que el valor de p, dado por la prueba de Anderson-Darling es muy bajo, podemos concluir que los datos no siguen una distribución normal.

- **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.**

```
library(e1071)
library(nortest)
cat("Sesgo de la muestra de tamaño 1000: ", skewness(datos), "\n")

## Sesgo de la muestra de tamaño 1000: -0.05304806

cat("Curtosis de la muestra de tamaño 1000: ", kurtosis(datos), "\n")

## Curtosis de la muestra de tamaño 1000: -0.2270957

ad.test(datos)

##
## Anderson-Darling normality test
##
## data:  datos
## A = 0.65655, p-value = 0.08635
```

El sesgo dentro de esta muestra nos dice que la distribución tiene una asimetría hacia la izquierda, en otras palabras, que tiene una cola pronunciada hacia la izquierda. El valor p dado por la prueba de Anderson-Darling es bueno, y como este es mayor a 0.05 podemos concluir que los datos siguen una distribución normal.

- **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 Gamma

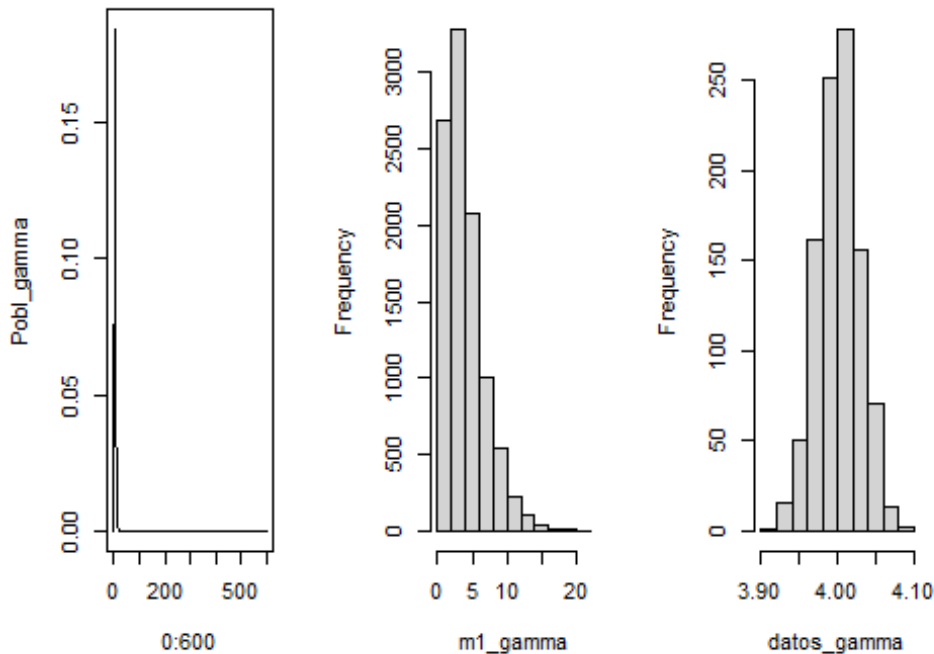
```
par(mfrow=c(1, 3))
Pobl_gamma = dgamma(0:600, shape = 2, scale = 2)
plot(0:600, Pobl_gamma, type="l", main = "Población con distribución
Gamma (shape=2, scale=2)")

m1_gamma = rgamma(10000, shape = 2, scale = 2)
hist(m1_gamma, main = "Muestra de tamaño 10,000 (Gamma)")

prom_gamma = mean(m1_gamma)
datos_gamma = prom_gamma
for(i in 1:999) {
  m = rgamma(10000, shape = 2, scale = 2)
  prom = mean(m)
  datos_gamma = rbind(datos_gamma, prom)
```

```
}
hist(datos_gamma, main="Promedios de 1000 muestras de tamaño 10,000
(Gamma)")
```

En distribución Gamma (sestra de tamaño 10,000 (G 1000 muestras de tamaño



```
sesgo_gamma = skewness(m1_gamma)
curtosis_gamma = kurtosis(m1_gamma)
sesgo_promedios_gamma = skewness(datos_gamma)
curtosis_promedios_gamma = kurtosis(datos_gamma)

cat("Sesgo de la muestra Gamma:", sesgo_gamma, "\n")
## Sesgo de la muestra Gamma: 1.36972

cat("Curtosis de la muestra Gamma:", curtosis_gamma, "\n")
## Curtosis de la muestra Gamma: 2.552493

cat("Sesgo de los promedios Gamma:", sesgo_promedios_gamma, "\n")
## Sesgo de los promedios Gamma: -0.003436166

cat("Curtosis de los promedios Gamma:", curtosis_promedios_gamma, "\n")
## Curtosis de los promedios Gamma: -0.06492726

ad.test(m1_gamma)

##
## Anderson-Darling normality test
```

```
##
## data:  m1_gamma
## A = 234.51, p-value < 2.2e-16

ad.test(datos_gamma)

##
## Anderson-Darling normality test
##
## data:  datos_gamma
## A = 0.12398, p-value = 0.9868
```

– Muestra 10,000

El sesgo dentro de esta muestra nos dice que la distribución tiene una asimetría hacia la derecha, en otras palabras, que tiene una cola muy pronunciada hacia la derecha. El curtosis nos dice que la distribución tiene un pico más bajo que una distribución normal. Debido a que el valor de p, dado por la prueba de Anderson-Darling es muy bajo, podemos concluir que los datos no siguen una distribución normal.

– Medias de 1,000 muestras

El sesgo al ser un valor muy bajo, nos dice que la distribución de los promedios es casi simétrica, lo cual la asemeja a una distribución normal. El valor de la curtosis, a pesar de ser negativo, es muy cercano al cero, por lo que podemos decir que la distribución es un poco más plana al de una distribución normal. El valor p dado por la prueba de Anderson-Darling es bueno, y como este es mayor a 0.05 podemos concluir que los datos siguen una distribución normal.

- Distribucion Log-Normal

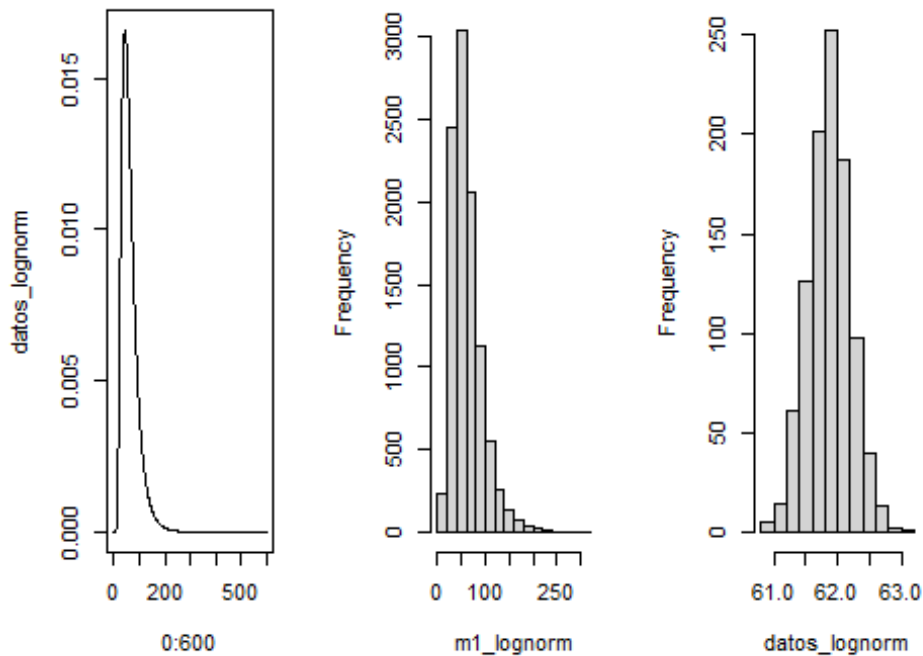
```
library(e1071)
par(mfrow=c(1, 3))

datos_lognorm = dlnorm(0:600, meanlog = 4, sdlog = 0.5)
plot(0:600, datos_lognorm, type="l", main = "Población con distribución
Log-Normal (meanlog=4, sdlog=0.5)")

m1_lognorm = rlnorm(10000, meanlog = 4, sdlog = 0.5)
hist(m1_lognorm, main = "Muestra de tamaño 10,000 (Log-Normal)")

prom_lognorm = mean(m1_lognorm)
datos_lognorm = prom_lognorm
for(i in 1:999) {
  m = rlnorm(10000, meanlog = 4, sdlog = 0.5)
  prom = mean(m)
  datos_lognorm = rbind(datos_lognorm, prom)
}
hist(datos_lognorm, main="Promedios de 1000 muestras de tamaño 10,000
(Log-Normal)")
```

istribución Log-Normal (mtra de tamaño 10,000 (Log000 muestras de tamaño ')



```
sesgo_lognorm = skewness(m1_lognorm)
curtosis_lognorm = kurtosis(m1_lognorm)
sesgo_promedios_lognorm = skewness(datos_lognorm)
curtosis_promedios_lognorm = kurtosis(datos_lognorm)

cat("Sesgo de la muestra Log-Normal:", sesgo_lognorm, "\n")
## Sesgo de la muestra Log-Normal: 1.599149

cat("Curtosis de la muestra Log-Normal:", curtosis_lognorm, "\n")
## Curtosis de la muestra Log-Normal: 4.419608

cat("Sesgo de los promedios Log-Normal:", sesgo_promedios_lognorm, "\n")
## Sesgo de los promedios Log-Normal: 0.04093647

cat("Curtosis de los promedios Log-Normal:", curtosis_promedios_lognorm,
"\n")
## Curtosis de los promedios Log-Normal: 0.0715454

ad.test(m1_lognorm)

##
## Anderson-Darling normality test
##
```

```
## data:  m1_lognorm
## A = 224.74, p-value < 2.2e-16

ad.test(datos_lognorm)

##
## Anderson-Darling normality test
##
## data:  datos_lognorm
## A = 0.21319, p-value = 0.8528
```

– Muestra 10,000

El sesgo dentro de esta muestra nos dice que la distribución tiene una asimetría muy pronunciada hacia la derecha, en otras palabras, que tiene una cola muy pronunciada hacia la derecha. El curtosis nos dice que la distribución tiene un pico mucho más alto que el de una distribución normal. Debido a que el valor de p, dado por la prueba de Anderson-Darling es muy bajo, podemos concluir que los datos no siguen una distribución normal.

– Medias de 1,000 muestras

El sesgo al ser un valor muy bajo, nos dice que la distribución de los promedios es casi simétrica, lo cual la asemeja a una distribución normal. El valor de la curtosis es muy cercano al cero, por lo que podemos decir que la distribución es un poco más plana al de una distribución normal. El valor p dado por la prueba de Anderson-Darling es bueno, y como este es mayor a 0.05 podemos concluir que los datos siguen una distribución normal.

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

En general los tres gráficos eran muy similares, la Gamma teniendo un pico mayor, pero todas seguían la misma tendencia de una asimetría hacia la derecha, de igual manera, todas las distribuciones de las medias de las muestras se comportaron de la misma manera. Todas cumplieron el Teorema del Límite Central, el cual nos dice que estas se iban a comportar como distribuciones normales, y así fue.

Pregunta 2

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

- 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?

$P(9900 \leq X \leq 10100)$ Desviaciones lejos de la media (Z)

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

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

z = (10100 - 10000) / 500
cat("z = ", z)

## 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?

$\bar{X} \sim N\left(\mu_{\bar{x}} = 10000, \sigma_{\bar{x}} = \frac{500}{\sqrt{120}}\right) P(9900 \leq \bar{X} \leq 10100)$

```
p1 = pnorm(10100, 10000, 500 / sqrt(120)) - pnorm(9900, 10000, 500 / sqrt(120))
cat("P(9900 <= x_b <= 10100)", p1)

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

z = (10100 - 10000) / (500 / sqrt(120))
cat("z = ", z)

## 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?

```
p1 = pnorm(10100, 10000, 500 / sqrt(15)) - pnorm(9900, 10000, 500 / sqrt(15))
cat("P(9900 <= x_b <= 10100)", p1)

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

z = (10100 - 10000) / (500 / sqrt(15))
cat("z = ", z)

## 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?.

```
p1 = pnorm(9800, 10000, 500 / sqrt(120))
cat("P(x_b <= 9800) = ", p1)

## P(x_b <= 9800) = 5.88567e-06

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

## z = 4.38178
```

Hizo lo correcto, porque la probabilidad de que los remaches tengan dicha media es muy baja, esto lo podemos observar viendo el valor de z y la probabilidad son muy altas.

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

```
#p1 es el valor p
p1 = pnorm(9925, 10000, 500 / sqrt(120))
cat("P(x_b <= 9925) = ", p1)

## P(x_b <= 9925) = 0.05017412

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

## z = 1.643168
```

En este caso debería de aceptarlo, ya que es más probable que la media sea de 9925, esto lo podemos ver gracias a la desviación estandar y el valor de la z, los cuales son bajos.

Pregunta 3

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.

- 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?

```
d = qnorm(0.025, 0, 1)
d

## [1] -1.959964
```

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

```
p1 = pnorm(15, 16, 1 / sqrt(10))
cat("P(x_b >= 16) = ", p1)

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

- 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?

```
z = (16 - 15) / (1 / sqrt(10))
z

## [1] 3.162278
```

Si se debería de detener la producción, porque la z con 16 de media es mayor a nuestro limite el cual es 1.96.

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

```
p = pnorm(14.5, 15, 1 / sqrt(10))
p

## [1] 0.05692315
```

- 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?

```
z = (15.5 - 15) / (1 / sqrt(10))
z

## [1] 1.581139
```

- d) Hacer una gráfica del inciso 1.

```
mu = 15
sigma = 1
x = seq(mu - 4 * sigma, mu + 4 * sigma, length.out = 1000)
y = dnorm(x, mean = mu, sd = sigma)
plot(x, y, type = "l", col = "blue", lwd = 2, main = "Distribución Normal
de la Cantidad de Líquido por Botella",
      xlab = "Cantidad de Líquido (onzas)", ylab = "Densidad de
Probabilidad")
# Rango del 95% central
z_critical <- 1.96
lower_bound <- mu - z_critical * sigma
upper_bound <- mu + z_critical * sigma

# Sombrear el área del 95% central
x_fill <- seq(lower_bound, upper_bound, length.out = 1000)
y_fill <- dnorm(x_fill, mean = mu, sd = sigma)
polygon(c(lower_bound, x_fill, upper_bound), c(0, y_fill, 0), col =
"lightblue", border = NA)
```

```
# Añadir líneas verticales para Los límites del 95% central  
abline(v = lower_bound, col = "red", lty = 2, lwd = 2)  
abline(v = upper_bound, col = "red", lty = 2, lwd = 2)
```

Distribución Normal de la Cantidad de Líquido por Bc

