

Tarea 6 - Distribuciones Muestrales y TCL

Héctor Hibrán Tapia Fernández - A01661114

2024-08-16

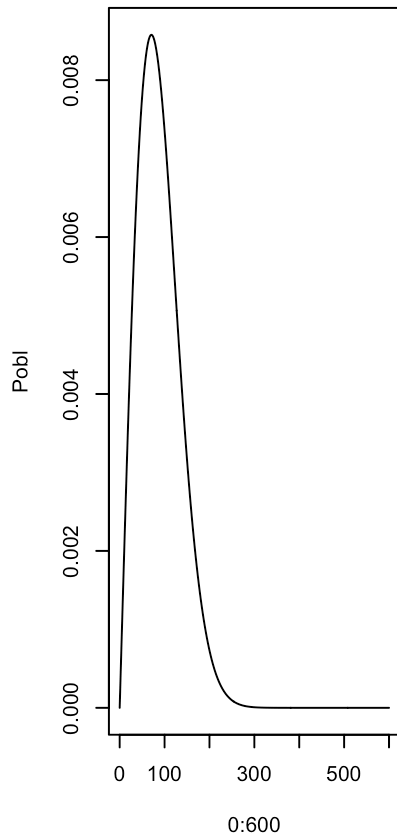
1. Ensayando Distribuciones

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:

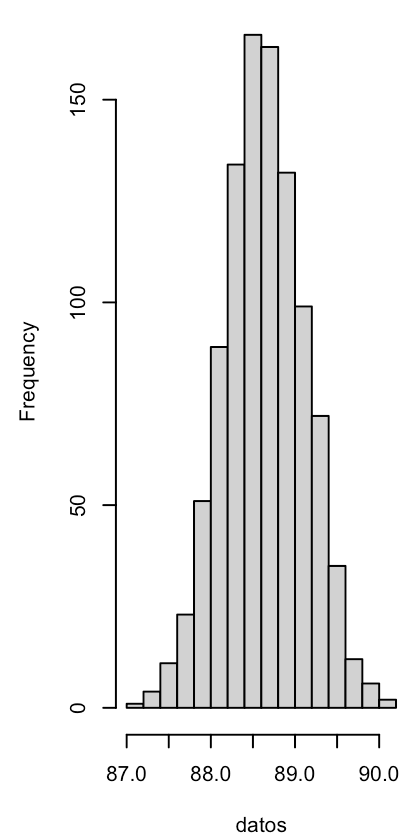
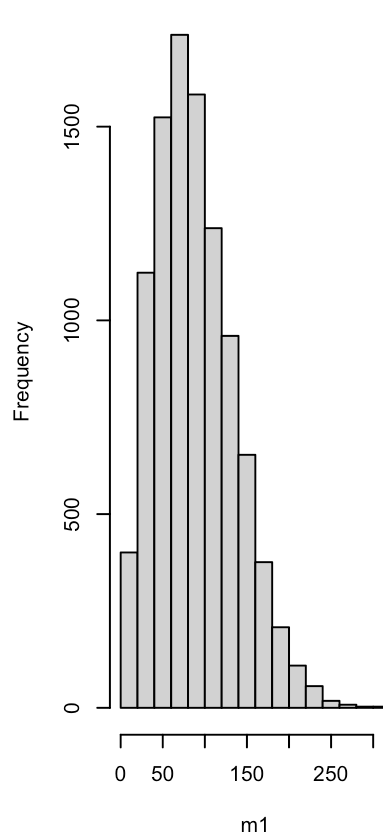
A) Ejercutar el siguiente código de R: DistsM_enR.txt. Se esperan tres gráficas, interprete cada una de ellas. Se usa una distribución Weibull, con parámetros $\alpha = 2$ y $\beta = 100$.

```
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 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")
```

**Poblacion con distribucion Weibull :
=2, beta = 100**



**Una muestra de tamaño 10000 a de los promedios de 1000 muestras
10,000**



Explicación de izquierda a derecha:

1. Primero tenemos el gráfico poblacional que se comporta como la distribución Weibull, se puede notar que es ásimetrica hacia la derecha que es donde se encuentra la “cola”.
2. En este grafo se tomó una muestra de 10,000 observaciones de la primera gráfica de distribución, este histograma sigue mostrando el mismo comportamiento que el grafo anterior, esto es debido al sesgo.
3. Este grafo muestra ahora un histograma de los promedios de 1,000 muestras, donde cada muestra tiene un tamaño 10,000 que fueron extraídas de la misma distribución Weibull. Aquí es donde el TLC entra en acción, apesar de que cada muestra individual sigue la distribución Weibull, los promedios de estas muestras tienden a seguir una distribución normal, es importante destacar que debe la muestra debe ser lo suficientemente grande para que este comportamiento sea notorio.

Está muestra es simétrica.

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) # para sesgo y curtosis
```

```
muestra = rweibull(10000, 2, 100)
sesgo = skewness(muestra)
curtosis = kurtosis(muestra)
```

```
cat("Sesgo: ", sesgo, "\n")
```

```
## Sesgo: 0.6355713
```

```
cat("Curtosis: ", curtosis, "\n")
```

```
## Curtosis: 3.307898
```

```
submuestra = sample(muestra, 5000) # Ocupamos una muestra mas pequeña ya que el shapiro.
test no acepta muestras más grande de 5k de datos
shapiro_test = shapiro.test(submuestra)
print(shapiro_test)
```

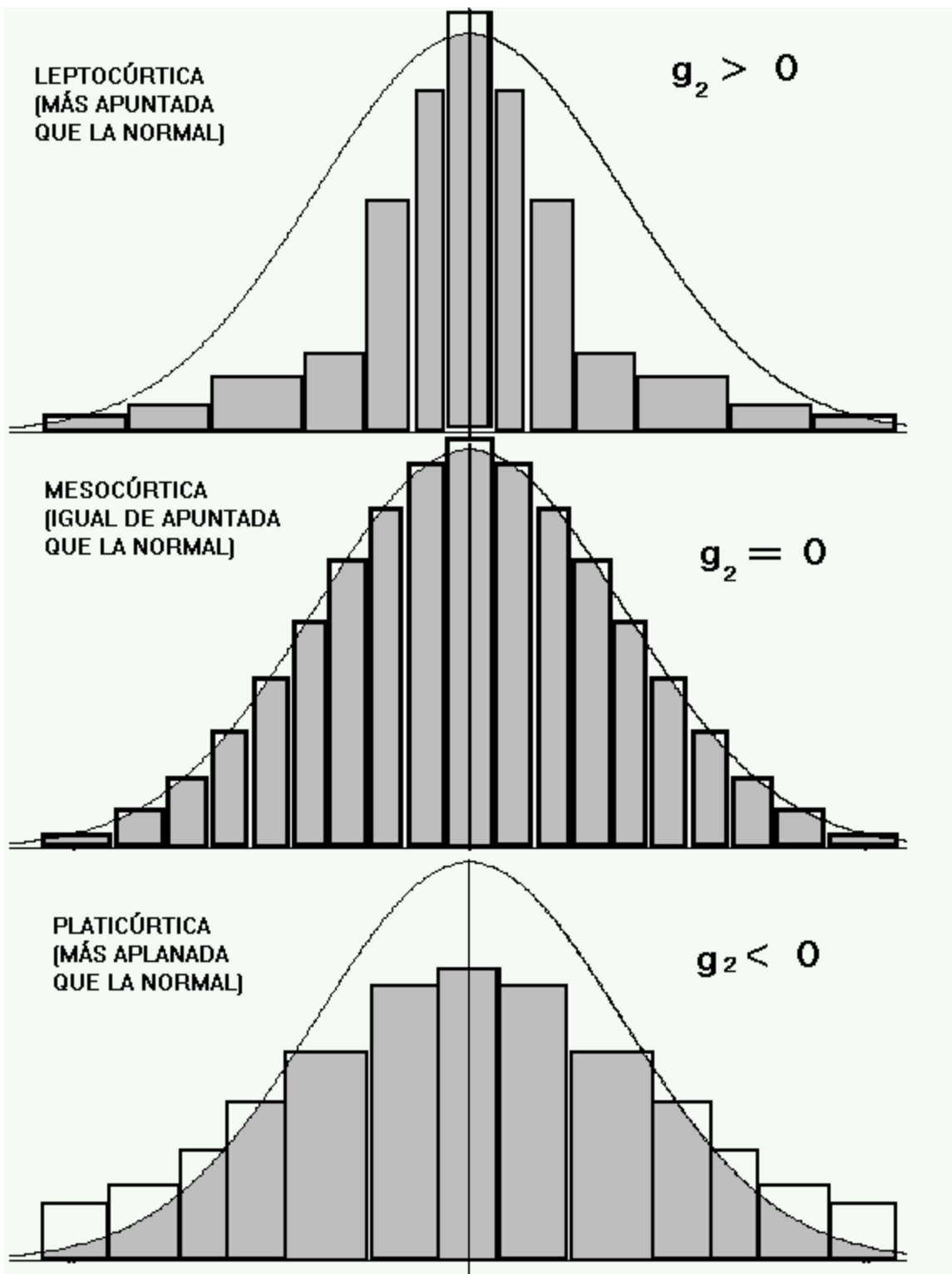
```
##
## Shapiro-Wilk normality test
##
## data: submuestra
## W = 0.97311, p-value < 2.2e-16
```



Sesgo

El sesgo como antes ya mencionado mide hacia donde tiene la cola nuestra distribución o la simetría de esta, si se tiene un sesgo de 0, significa que nuestra distribución es simétrica...

En nuestro caso, el sesgo es de 0.610, lo dice que la distribución tiene una ligera asimetría positiva, la cola derecha es más larga o tiene más valores extremos que la izquierda (como se ve en la imagen).



Curtosis

La *curtosis* mide que tan “puntiagudo” es la distribución, como se muestra en la imagen. En nuestro caso se muestra un valor de 3.163869 lo que nos dice que es ligeramente leptocúrtica.

Por último interpretemos los resultados del *Shapiro Test* que se utiliza para evaluar la normalidad de una muestra:

- $W = 0.97093$, los valores cercanos a 1 nos dicen que la muestra es más consistente con una distribución normal.
- $p\text{-value} < 2.2e-16$, El $p\text{-value}$ demasiado bajo (menor que 0.05), lo nos dice que hay suficiente evidencia para rechazar la hipótesis nula. (la muestra proviene de una distribución normal)

Con toda esta información, (y sin usar una sola gráfica) podemos decir que *la muestra no sigue 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.

```
vector_medias = numeric(1000)

for(i in 1:1000) {
  muestra = rweibull(10000, 2, 100)
  vector_medias[i] = mean(muestra)
}

sesgo_medias = skewness(vector_medias)
curtosis_medias = kurtosis(vector_medias)

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

```
## Sesgo de las medias: -0.03381598
```

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

```
## Curtosis de las medias: 2.850395
```

```
prueba_st_medias = shapiro.test(datos)
print(prueba_st_medias)
```

```
##
## Shapiro-Wilk normality test
##
## data:  datos
## W = 0.99896, p-value = 0.8494
```

En este caso el sesgo es negativo lo que indica que la cola está hacia la izquierda, aunque cabe mencionar que no está demasiado sesgada, ya que el valor está cerca de 0.

La *curtosis* nos da un valor de 3.033333, lo que indica que es ligeramente leptocúrtica.

Por último interpretemos los resultados del *Shapiro Test* que se utiliza para evaluar la normalidad de una muestra:

- $W = 0.99851$, los valores cercanos a 1 nos dicen que la muestra es más consistente con una distribución normal.
- $p\text{-value} = 0.5639$, El $p\text{-value}$ es de 0.5639 que es mayor al nivel de significancia (0.05), lo que nos dice que no hay suficiente evidencia para rechazar la hipótesis nula de normalidad.

Con toda esta información, (y sin usar una sola gráfica) podemos decir que la *muestra sigue una distribución normal*.

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.

Veamos la distribución Gamma:

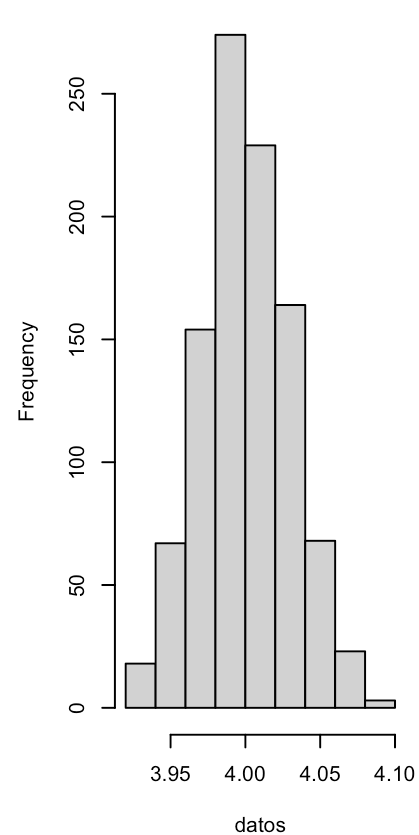
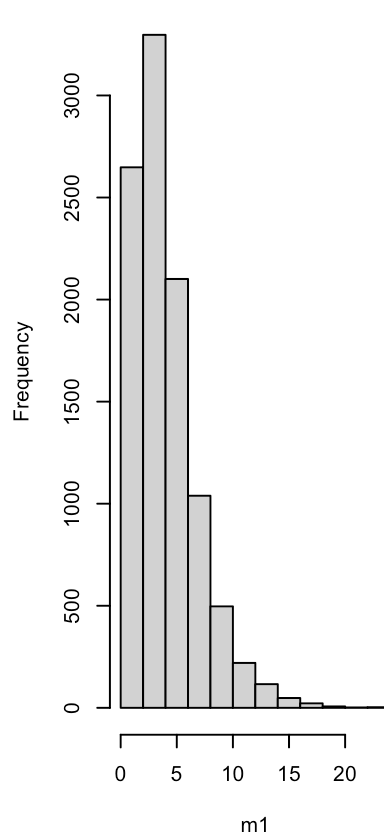
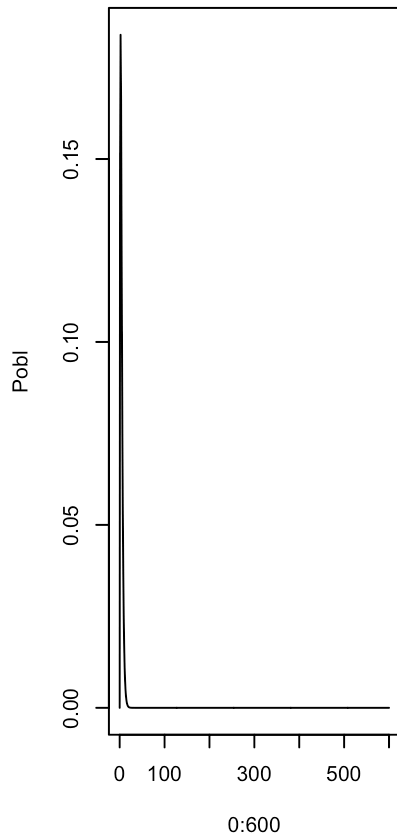
```
par(mfrow=c(1,3))
# Graficando una distribución Gama con alfa = 2, beta = 2
Pobl = dgamma(0:600, shape=2, scale=2)
plot(0:600, Pobl, type="l", main="Población con distribución Gama alfa = 2, beta = 2")

# Tomando una muestra de 10000 elementos al azar
m1 = rgamma(10000, shape=2, scale=2)
hist(m1, main="Una muestra de tamaño 10000")

# Tomando 1000 promedios de las 1000 muestras
m = rgamma(10000, shape=2, scale=2)
prom = mean(m)
datos = prom
for(i in 1:999) {
  m = rgamma(10000, shape=2, scale=2)
  prom = mean(m)
  datos = rbind(datos, prom)
}
hist(datos, main="Gráfica de los promedios de 1000 muestras de tamaño 10,000")
```

ación con distribución Gama alfa = 2,

Una muestra de tamaño 10000 los promedios de 1000 muestras de t



```
library(moments) # para sesgo y curtosis
```

```
muestra = rgamma(10000, shape = 2, scale = 2)
```

```
sesgo = skewness(muestra)
```

```
curtosis = kurtosis(muestra)
```

```
cat("Sesgo: ", sesgo, "\n")
```

```
## Sesgo: 1.48914
```

```
cat("Curtosis: ", curtosis, "\n")
```

```
## Curtosis: 6.267972
```

```
submuestra = sample(muestra, 5000)
```

```
shapiro_test = shapiro.test(submuestra)
```

```
print(shapiro_test)
```

```
##
## Shapiro-Wilk normality test
##
## data:  submuestra
## W = 0.8854, p-value < 2.2e-16
```

Sesgo: Es positivo lo dice que la distribución tiene una ligera asimetría positiva, la cola derecha es más larga o tiene más valores extremos que la izquierda *Curtosis:* Es leptocúrtica, se aleja de la normal. *Shapiro Test:* - $W = 0.8877$: No es lo más cercano de 1 que se podría ser. - $p\text{-value} < 2.2e-16$: El valor es demasiado pequeño, rechazamos hipótesis nula.

Con toda esta información, (y sin usar una sola gráfica) podemos decir que *la muestra no sigue una distribución normal*.

```
vector_medias = numeric(1000)

for(i in 1:1000) {
  muestra = rgamma(10000, shape = 2, scale = 2)
  vector_medias[i] = mean(muestra)
}

sesgo_medias = skewness(vector_medias)
curtosis_medias = kurtosis(vector_medias)

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

```
## Sesgo de las medias: -0.1017918
```

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

```
## Curtosis de las medias: 2.996137
```

```
prueba_st_medias = shapiro.test(datos)
print(prueba_st_medias)
```

```
##
## Shapiro-Wilk normality test
##
## data:  datos
## W = 0.99784, p-value = 0.2206
```

Sesgo: Es positivo lo dice que la distribución tiene una ligera asimetría positiva, la cola derecha es más larga o tiene más valores extremos que la izquierda *Curtosis:* Es leptocúrtica, se aleja ligeramente de la normal. *Shapiro Test:* - $W = 0.9982$: Es muy cercana a 1. - $p\text{-value} = 0.3783$: El valor no es menor al nivel de significancia = 0.05, no rechazamos la hipótesis nula.

Con toda esta información, (y sin usar una sola gráfica) podemos decir que *la muestra sigue una distribución normal*.

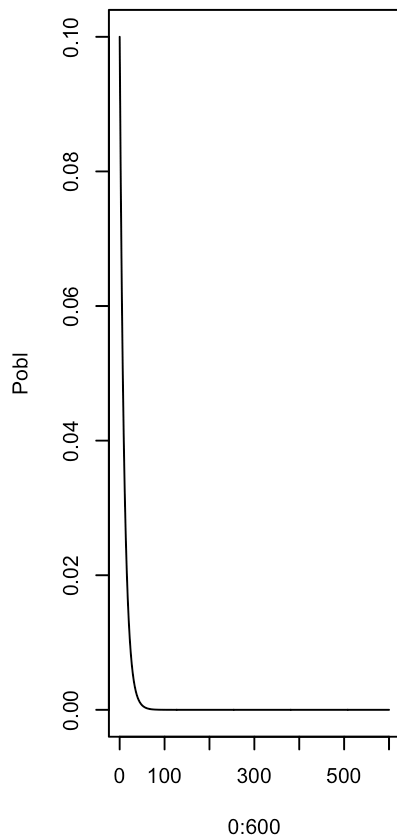
Ahora veamos la distribución Exponencial:

```
par(mfrow=c(1,3))
# Graficando una distribución Exponencial con lambda = 0.1
Pobl = dexp(0:600, rate = 0.1)
plot(0:600, Pobl, type="l", main="Población con distribución Exponencial lambda = 0.1")

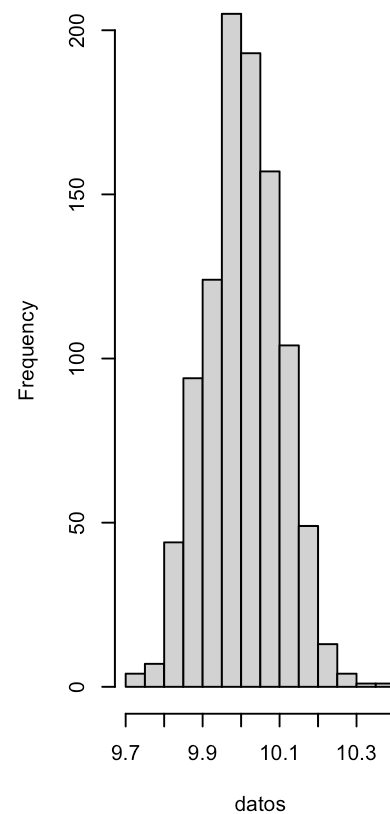
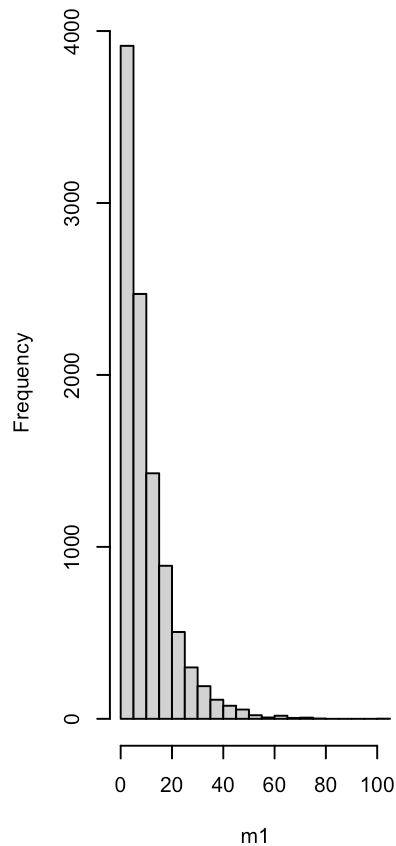
# Tomando una muestra de 10000 elementos al azar
m1 = rexp(10000, rate=0.1)
hist(m1, main="Una muestra de tamaño 10000")

# Tomando 1000 promedios de las 1000 muestras
m = rexp(10000, rate=0.1)
prom = mean(m)
datos = prom
for(i in 1:999) {
  m = rexp(10000, rate=0.1)
  prom = mean(m)
  datos = rbind(datos, prom)
}
hist(datos, main="Gráfica de los promedios de 1000 muestras de tamaño 10,000")
```

ción con distribución Exponencial lar



Una muestra de tamaño 10000 los promedios de 1000 muestras de t



```
library(moments) # para sesgo y curtosis
```

```
muestra = rexp(10000, rate=0.1)
sesgo = skewness(muestra)
curtosis = kurtosis(muestra)

cat("Sesgo: ", sesgo, "\n")
```

```
## Sesgo: 2.090916
```

```
cat("Curtosis: ", curtosis, "\n")
```

```
## Curtosis: 10.37496
```

```
submuestra = sample(muestra, 5000)
shapiro_test = shapiro.test(submuestra)
print(shapiro_test)
```

```
##
## Shapiro-Wilk normality test
##
## data: submuestra
## W = 0.79815, p-value < 2.2e-16
```

Sesgo: Es positivo lo dice que la distribución tiene una ligera asimetría positiva, la cola derecha es más larga o tiene más valores extremos que la izquierda **Curtosis:** Es leptocúrtica, se aleja de la normal. **Shapiro Test:** - $W = 0.82599$: No es muy cercana a 1. - $p\text{-value} < 2.2e-16$: El es menor muy pequeño al nivel de significancia = 0.05, rechazamos la hipótesis nula.

Con toda esta información, (y sin usar una sola gráfica) podemos decir que *la muestra no sigue una distribución normal*.

```
vector_medias = numeric(1000)

for(i in 1:1000) {
  muestra = rexp(10000, rate=0.1)
  vector_medias[i] = mean(muestra)
}

sesgo_medias = skewness(vector_medias)
curtosis_medias = kurtosis(vector_medias)

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

```
## Sesgo de las medias: 0.101113
```

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

```
## Curtosis de las medias: 3.056633
```

```
prueba_st_medias = shapiro.test(datos)
print(prueba_st_medias)
```

```
##
## Shapiro-Wilk normality test
##
## data: datos
## W = 0.99868, p-value = 0.6741
```

Sesgo: Es positivo lo dice que la distribución tiene una ligera asimetría positiva, la cola derecha es más larga o tiene más valores extremos que la izquierda. **Curtosis:** Es leptocúrtica, se aleja ligeramente de la normal. **Shapiro Test:** - $W = 0.99861$: Es muy cercana a 1. - $p\text{-value} = 0.6311$: No es menor al nivel de significancia = 0.05, no rechazamos la hipótesis nula.

Con toda esta información, (y sin usar una sola gráfica) podemos decir que *la muestra sigue una distribución normal*.

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

Distribución Poblacional:

- Semejanzas: Las 3 distribuciones son ásimétricas positivas, la mayor concentración de muestras está a la izquierda y la cola larga hacia la derecha.
- Diferencias: Cada Distribución explica fenómenos diferentes y tiene particularidades muy específicas, por ejemplo la caída de cada distribución depende de esta misma, la exponencial es la que cae más abruptamente.

Distribución de Muestras de Tamaño 10,000:

- Semejanzas: Todas las muestras mantienen la asimetría de la población original y mantienen la misma concentración y cola.
- Diferencias: Cada Distribución explica fenómenos diferentes y tiene particularidades muy específicas, por ejemplo la caída de cada distribución depende de esta misma, la exponencial es la que cae más abruptamente.

Distribución de los Promedios de 1,000 Muestras de Tamaño 10,000:

- Semejanzas: Todas las distribuciones de los promedios son aproximadamente normales, gracias al TLC, que indica que la distribución de los promedios de muestras (lo suficientemente) grandes tiende a una distribución normal, sin importar la distribución original.
- Diferencias: Los promedios varían.

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?

```
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("\nCantidad de desviaciones estándar a la que está de la media =", z1)
```

```
##
## Cantidad de desviaciones estándar a la que está de la media = 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?

$P(9900 < x < 10100)$

```
p2 <- pnorm(10100, 10000, 500/sqrt(120)) - pnorm(9900, 10000, 500/sqrt(120))
cat("\nP(9900 < x < 10100) =", p2)
```

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

```
z2 = 100/(500/sqrt(120))
cat("\nCantidad de desviaciones estándar a la que está de la media =", z2)
```

```
##
## Cantidad de desviaciones estándar a la que está de la media = 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("\nP(9900 < x < 10100) =", p3)
```

```
##
## P(9900 < x < 10100) = 0.561422
```

```
z3 = 100/(500/sqrt(15))
cat("\nCantidad de desviaciones estándar a la que está de la media =", z3)
```

```
##
## Cantidad de desviaciones estándar a la que está de la media = 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?.

```
p4 <- (9800-10000)/(500/sqrt(120))
p4
```

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

```
p5 <- pnorm(9800, 10000, 500/sqrt(120))
p5
```

```
## [1] 5.88567e-06
```

Hipótesis nula = la media de la resistencia de los remaches es de 10 000 lb/pulg².

Rechazo básicamente por el valor del p-value, su valor es extremadamente pequeño (-4.38178), lo que indica que el inge hizo lo correcto al rechazar el lote, ya que el p-value indica que es altamente improbable que la media de la resistencia de los remaches sea la que el desea, o lo que es lo mismo que hay suficiente evidencia para rechazar el lote.

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

```
p6 <- (9925-10000)/(500/sqrt(120))
p6
```

```
## [1] -1.643168
```

```
p7 <- pnorm(9925, 10000, 500/sqrt(120))
p7
```

```
## [1] 0.05017412
```

Acepto porque el p-valor está muy cerca del nivel de significancia típico (0.05), pero diría que la decisión no es muy clara como en el caso anterior. Pero aceptaría básicamente porque no está tan alejado del 0.05, lo que le recomendaría al ingeniero no rechazar el lote, ya que no hay suficiente evidencia para rechazar la hipótesis nula.

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?

```
z1 = qnorm(0.975)
z1
```

```
## [1] 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?

```
sigma = 1
mu = 15

sigma_2 = sigma / sqrt(10)

prob = 1 - pnorm(16, mu, sigma_2)
cat("P(X > 16) =", prob)
```

```
## P(X > 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?

```
media_onzas = 16

z = (media_onzas - mu) / (sigma / sqrt(10))
cat("z =", z, "\n")
```

```
## z = 3.162278
```

```
if (z2 > z1) {
  cat("Se detendrá la producción.")
} else {
  cat("No se detendrá la producción.")
}
```

```
## Se detendrá 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?

```
prob2 = pnorm(14.5, mu, sigma_2)
cat("P(X < 14.5) =", prob2)
```

```
## P(X < 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?

```
media_onzas = 15.5

z = (media_onzas - mu) / (sigma / sqrt(10))
cat("z =", z, "\n")
```

```
## z = 1.581139
```

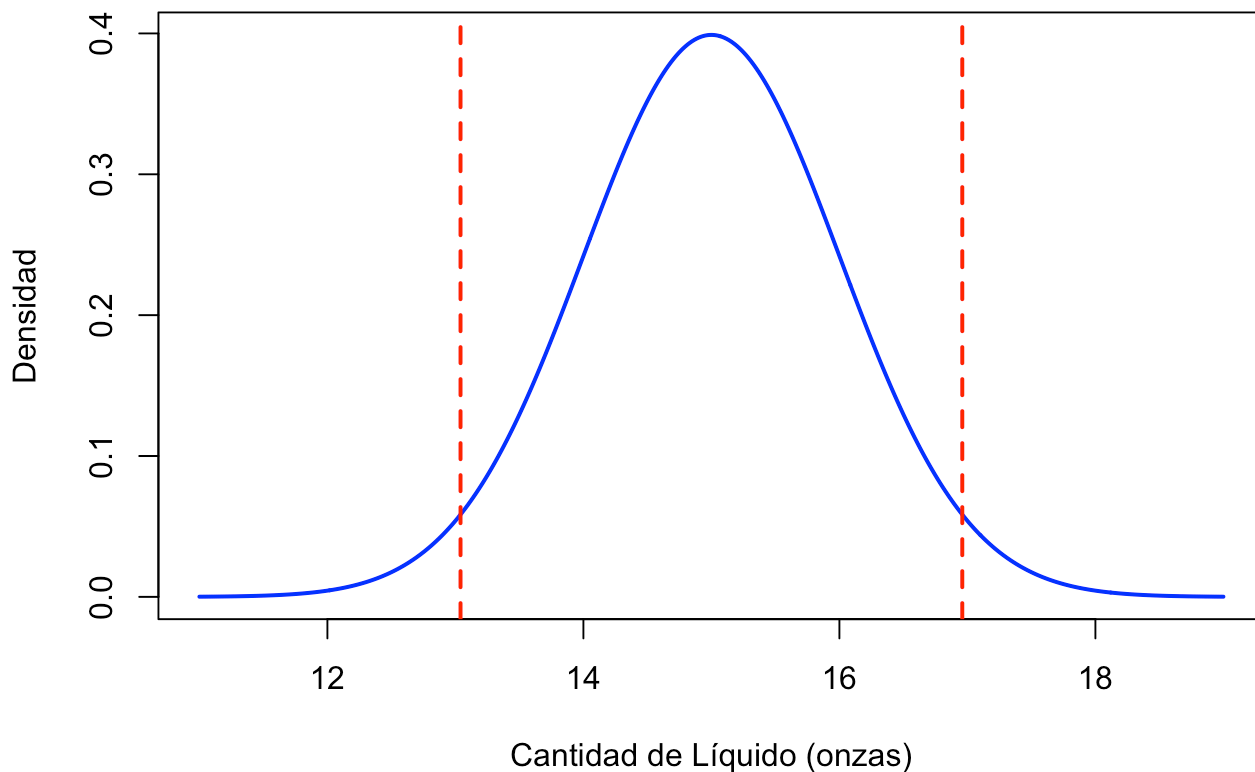
```
if (z2 > z1) {  
  cat("Se detendrá la producción.")  
} else {  
  cat("No se detendrá la producción.")  
}
```

```
## Se detendrá la producción.
```

6. Hacer una gráfica del inciso 1.

```
límite_inferior <- mu - z1 * sigma  
límite_superior <- mu + z1 * sigma  
  
x = seq(mu - 4*sigma, mu + 4*sigma, length = 1000)  
y = dnorm(x, mean = mu, sd = sigma)  
  
plot(x, y, typ = "l", lwd = 2, col = "blue", main = "Distribución Normal de la Máquina E  
mbotelladora",  
      xlab = "Cantidad de Líquido (onzas)", ylab = "Densidad")  
  
abline(v=límite_inferior, col = "red", lwd = 2, lty = 2)  
abline(v=límite_superior, col = "red", lwd = 2, lty = 2)
```

Distribución Normal de la Máquina Embotelladora



Sugerencia: Para contestar los incisos c y e le puede ser de utilidad calcular la Z para los valores de la media obtenidos en cada inciso.