

# Distribuciones Muestrales y TCL

Nallely Serna

2024-08-18

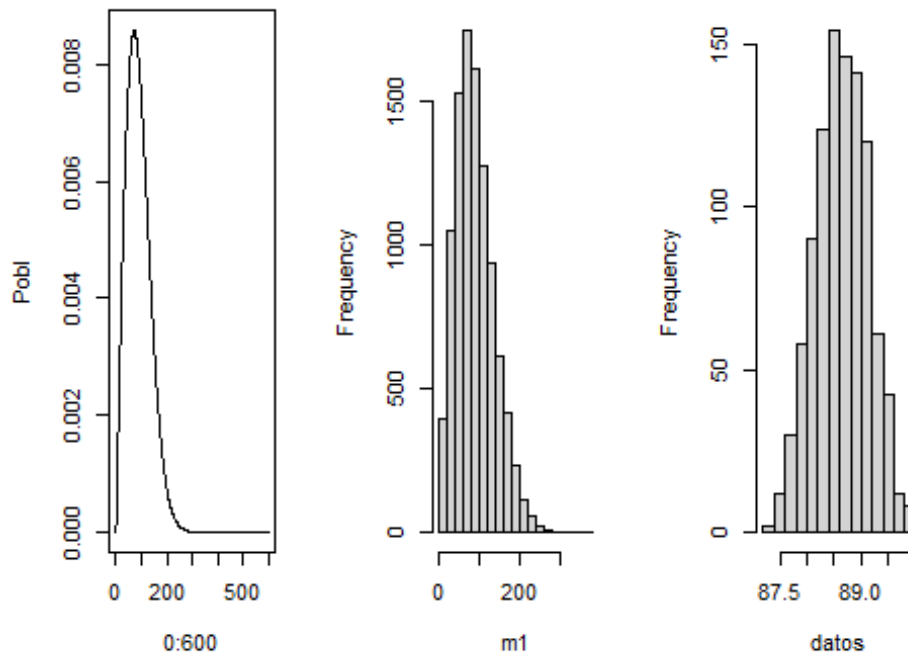
##1. Ensayando Distribuciones Grafica la Distribución de una variable aleatoria, la de una muestra elegida al azar y la de la Distribución de las medias de 10000 muestras:

##A

Ejercutar el siguiente código de R: 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")
```

Generación con distribución Weibull,  $\alpha = 2$ ,  $\beta = 100$ . Una muestra de tamaño 10,000 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.

```
# Cargar Las Librerías necesarias
library(e1071) # Para calcular sesgo y curtosis
library(nortest) # Para La prueba de normalidad

# Sesgo y curtosis de La muestra de tamaño 10,000
sesgo_muestra <- skewness(m1)
curtosis_muestra <- kurtosis(m1)
cat("Sesgo de la muestra:", sesgo_muestra, "\n")

## Sesgo de la muestra: 0.6654008

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

## Curtosis de la muestra: 0.3501355

# Prueba de normalidad para La muestra (Anderson-Darling)
ad_test_muestra <- ad.test(m1)
cat("Prueba de Anderson-Darling para la muestra de tamaño 10,000:\n")

## Prueba de Anderson-Darling para la muestra de tamaño 10,000:
print(ad_test_muestra)
```

```
##  
## Anderson-Darling normality test  
##  
## data: m1  
## A = 64.712, p-value < 2.2e-16
```

##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.

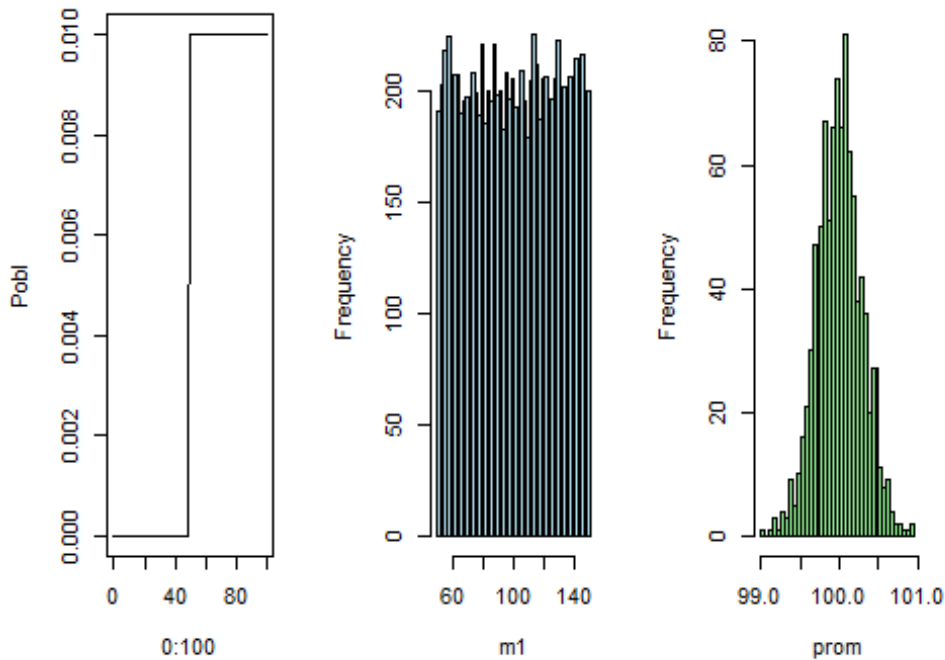
```
# Sesgo y curtosis de Las medias de Las 1000 muestras  
sesgo_medias <- skewness(datos)  
curtosis_medias <- kurtosis(datos)  
cat("Sesgo de las medias:", sesgo_medias, "\n")  
  
## Sesgo de las medias: -0.03250369  
  
cat("Curtosis de las medias:", curtosis_medias, "\n")  
  
## Curtosis de las medias: -0.3317133  
  
# Prueba de normalidad para Las medias  
shapiro_test_medias <- shapiro.test(datos)  
cat("Prueba de Shapiro-Wilk para las medias de 1000 muestras:\n")  
  
## Prueba de Shapiro-Wilk para las medias de 1000 muestras:  
  
print(shapiro_test_medias)  
  
##  
## Shapiro-Wilk normality test  
##  
## data: datos  
## W = 0.99744, p-value = 0.1176
```

##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.

```
# Cambiar a una Distribución Uniforme  
par(mfrow=c(1,3))  
  
# Distribución Uniforme entre 50 y 150  
Pobl = dunif(0:100, min=50, max=150)  
plot(0:100, Pobl, type="l", main="Población con distribución Uniforme  
entre 50 y 150")  
  
# Muestra aleatoria de tamaño 10,000  
m1 = runif(10000, min=50, max=150)  
hist(m1, main="Una muestra de tamaño 10,000", col="lightblue", breaks=50)
```

```
# Distribución de Las medias de 1000 muestras
prom = numeric(1000)
for(i in 1:1000) {
  m = runif(10000, min=50, max=150)
  prom[i] = mean(m)
}
hist(prom, main="Gráfica de los promedios de 1000 muestras de tamaño
10,000", col="lightgreen", breaks=50)
```

con distribución Uniforme una muestra de tamaño 10000 y promedios de 1000 muestras



##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, la distribución de las medias muestrales tiende a ser normal, independientemente de la forma de la distribución original. Esto es una confirmación del teorema central del límite. Las medias muestrales tienden a tener un sesgo cercano a cero y una curtosis cercana a 3, lo que sugiere que se distribuyen normalmente.

Diferencias:

Las distribuciones originales y las muestras grandes mantienen la asimetría y otras características propias de su distribución. Sin embargo, las medias de las muestras son menos sensibles a estas características y tienden hacia la normalidad, lo que muestra cómo el teorema central del límite funciona en la práctica.

## ##2. Remaches

La resistencia a la ruptura de un remache tiene un valor medio de 10,000 lb/pulg<sup>2</sup> y una desviación estándar de 500 lb/pulg<sup>2</sup>. 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 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?

$P(9900 < x < 10100)$

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

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

### Desviaciones lejos de la media (z)

```
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?

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

```
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?

*# Cálculo de la probabilidad para la parte C*

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

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

```
# Cálculo de las desviaciones estándar Lejos de la media para la parte C
```

```
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/pulg<sup>2</sup>. Para ello tomó una muestra de 120 remaches elegidos al azar tenía media de 9800 lb/pulg<sup>2</sup> y rechazó el pedido, ¿hizo lo correcto? ¿por qué?.

```
# Datos
```

```
mu_x <- 10000
```

```
sigma_x <- 500
```

```
n <- 120
```

```
x_bar_d <- 9800
```

```
# Cálculo del valor z
```

```
z_d <- (x_bar_d - mu_x) / (sigma_x / sqrt(n))
```

```
cat("z para X_b = 9800:", z_d, "\n")
```

```
## z para X_b = 9800: -4.38178
```

```
# Comparar con valor crítico
```

```
alpha <- 0.05
```

```
z_critical <- qnorm(1 - alpha / 2)
```

```
cat("Valor crítico de z (al 5% de significancia):", z_critical, "\n")
```

```
## Valor crítico de z (al 5% de significancia): 1.959964
```

```
# Decisión
```

```
if (abs(z_d) > z_critical) {
```

```
  cat("Rechazar H0: La decisión del ingeniero fue correcta.\n")
```

```
} else {
```

```
  cat("No rechazar H0: La decisión del ingeniero no fue correcta.\n")
```

```
}
```

```
## Rechazar H0: La decisión del ingeniero fue correcta.
```

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

```
# Datos
```

```
x_bar_e <- 9925
```

```
# Cálculo del valor z
```

```
z_e <- (x_bar_e - mu_x) / (sigma_x / sqrt(n))
```

```
cat("z para X_b = 9925:", z_e, "\n")
```

```
## z para X_b = 9925: -1.643168
```

```
# Comparar con valor crítico
cat("Valor crítico de z (al 5% de significancia):", z_critical, "\n")

## Valor crítico de z (al 5% de significancia): 1.959964

# Decisión
if (abs(z_e) > z_critical) {
  cat("Rechazar H0: Recomendaría rechazar el lote.\n")
} else {
  cat("No rechazar H0: No recomendaría rechazar el lote.\n")
}

## No rechazar H0: No recomendaría rechazar el lote.
```

### ##3. Embotellando

Una máquina embotelladora puede ser regulada para que se descargue un promedio de  $\mu$  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  $\mu$  sea de 15 onzas.

#### ##1

¿A cuántas desviaciones estándar alrededor de la verdadera media  $\mu$  puede estar la media de una muestra para que esté dentro del estándar establecido del 95% central?

```
# Nivel de significancia
alpha <- 0.05

# Valor crítico para el 95% central (bilateral)
z_critical <- qnorm(1 - alpha / 2)
cat("Valor crítico z para el 95% central:", z_critical, "\n")

## Valor crítico z para el 95% central: 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?

```
# Datos
mu <- 15
sigma <- 1
n <- 10
x_bar <- 16

# Cálculo del valor z
z2 <- (x_bar - mu) / (sigma / sqrt(n))
p2 <- 1 - pnorm(z2)
cat("Probabilidad de obtener una media mayor a 16 onzas:", p2, "\n")
```

```
## Probabilidad de obtener una media mayor a 16 onzas: 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?

```
# Decisión de detener producción
```

```
if (abs(z2) > z_critical) {  
  cat("Detener producción: La media de 16 onzas está fuera del intervalo  
del 95%.\n")  
} else {  
  cat("No detener producción: La media de 16 onzas está dentro del  
intervalo del 95%.\n")  
}
```

```
## Detener producción: La media de 16 onzas está fuera del intervalo del  
95%.
```

```
##4
```

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

```
x_bar <- 14.5
```

```
# Cálculo del valor z
```

```
z4 <- (x_bar - mu) / (sigma / sqrt(n))  
p4 <- pnorm(z4)  
cat("Probabilidad de obtener una media menor a 14.5 onzas:", p4, "\n")
```

```
## Probabilidad de obtener una media menor a 14.5 onzas: 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?

```
x_bar <- 15.5
```

```
# Cálculo del valor z
```

```
z5 <- (x_bar - mu) / (sigma / sqrt(n))
```

```
# Decisión de detener producción
```

```
if (abs(z5) > z_critical) {  
  cat("Detener producción: La media de 15.5 onzas está fuera del  
intervalo del 95%.\n")  
} else {  
  cat("No detener producción: La media de 15.5 onzas está dentro del  
intervalo del 95%.\n")  
}
```



## No detener producción: La media de 15.5 onzas está dentro del intervalo del 95%.

##6

Hacer una gráfica del inciso 1.

*# Gráfica*

```
x <- seq(mu - 4*sigma, mu + 4*sigma, length=100)
y <- dnorm(x, mean=mu, sd=sigma/sqrt(n))
```

*# Graficar la distribución normal*

```
plot(x, y, type="l", lwd=2, col="blue", xlab="Media muestral",
     ylab="Densidad", main="Distribución Normal del 95% Central")
```

*# Agregar el 95% central*

```
abline(v=mu - z_critical * sigma/sqrt(n), col="red", lty=2)
abline(v=mu + z_critical * sigma/sqrt(n), col="red", lty=2)
```

*# Resaltar el área del 95% central*

```
polygon(c(mu - z_critical * sigma/sqrt(n), seq(mu - z_critical *
sigma/sqrt(n), mu + z_critical * sigma/sqrt(n), length=100), mu +
z_critical * sigma/sqrt(n)),
       c(0, dnorm(seq(mu - z_critical * sigma/sqrt(n), mu + z_critical *
sigma/sqrt(n), length=100), mean=mu, sd=sigma/sqrt(n)), 0), col="gray")
```

### Distribución Normal del 95% Central

