

## 6. Distribuciones Muestrales y TCL

Eliezer Cavazos

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:

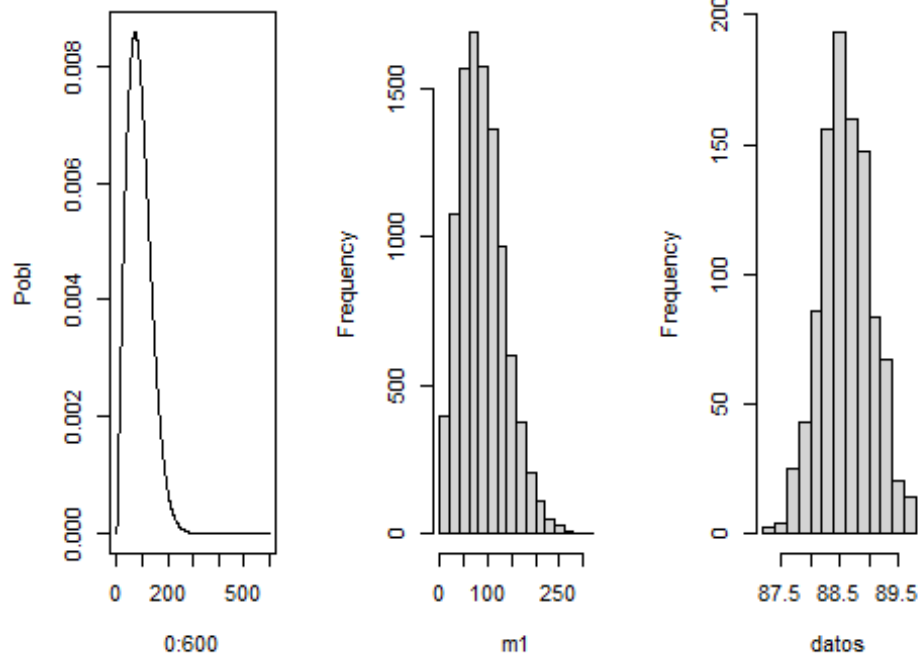
#### Inciso A

A. Ejecutar el siguiente código de R: DistsM\_enR.txt Download 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 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")
```

acion con distribucion We  
 $\alpha = 2$ ,  $\beta = 100$  Una muestra de tamaño 10,000 promedios de 1000 mue



En la primera grafica podemos identificar la distribucion de la poblacion donde la grafica muestra una leptocurtica junto con un sesgo positivo ya que tambien se tiene un sesgo a la derecha, tambien podemos identificar donde se encuentran la mayoria de los valores que en este caso es en el valor de 100.

En la segunda grafica podemos identificar igual una grafica que muestra una leptocurtica con un sesgo positivo y donde la frecuencia de los datos se encuentra entre los 50 a 150, junto tambien que la primera grafica un sesgo a la derecha.

En la tercera grafica podemos identificar un distribución normal ya que la mayor frecuencia de datos se encuentra muy cercano a la media.

## Inciso B

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.

```
sg = skewness(m)
k = kurtosis(m)

cat("Sesgo: ", sg)

## Sesgo: 0.6830507

cat("Curtosis: ", k)

## Curtosis: 3.436946
```

```
jarque.test(m)
```

```
##  
## Jarque-Bera Normality Test  
##  
## data: m  
## JB = 857.15, p-value < 2.2e-16  
## alternative hypothesis: greater
```

Con los datos proporcionados de la prueba de hipotesis podemos ver que el valor de p es muy pequeño, junto con un sesgo positivo, significa que la distribución de los datos no es una Distribución Normal, además de que los valores tienen Sesgo a la Derecha.

### Inciso C

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.

```
sg2 = skewness(datos)  
k2 = kurtosis(datos)  
  
cat("Sesgo_media: ", sg2)  
  
## Sesgo_media: 0.07280247  
  
cat("Curtosis_media: ", k2)  
  
## Curtosis_media: 2.793708  
  
jarque.test(datos[, 1])  
  
##  
## Jarque-Bera Normality Test  
##  
## data: datos[, 1]  
## JB = 2.6566, p-value = 0.2649  
## alternative hypothesis: greater
```

Con la media de los datos podemos identificar que el valor de p es más grande esto significa que no rechaza la hipótesis de normalidad, junto con un sesgo cercano a 0 y una curtosis cercana a 3 podemos identificar que la distribución de las medias de los datos es una Distribución Normal con un ligero Sesgo a la Izquierda

### Inciso D

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.

```

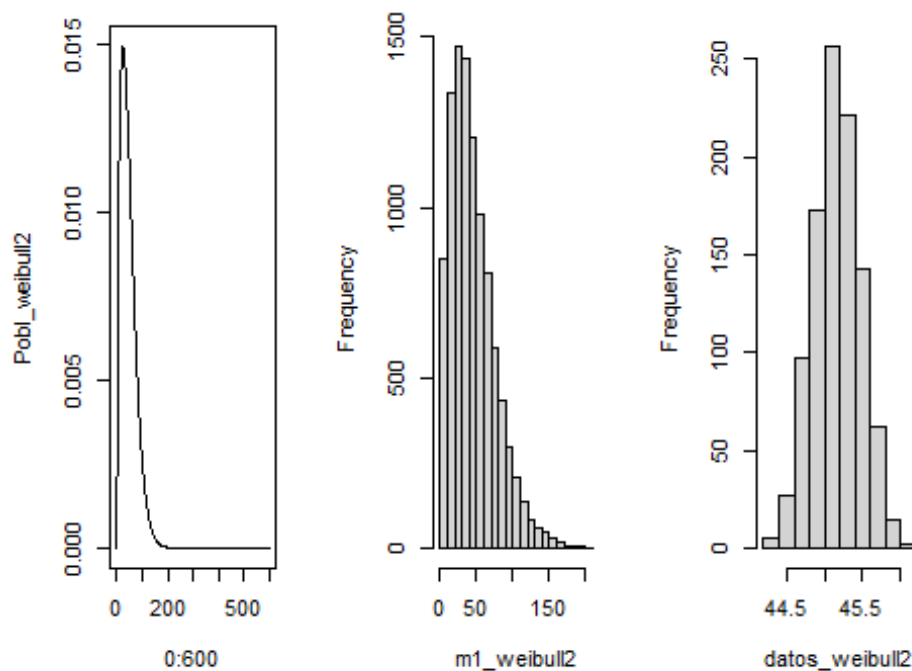
par(mfrow=c(1,3))
# Graficando una distribución Weibull de alfa = 1.5, beta = 50
Pobl_weibull2 = dweibull(0:600, 1.5, 50)
plot(0:600, Pobl_weibull2, type="l", main = "Población con distribución
Weibull alfa = 1.5, beta = 50")

# Tomando una muestra de 10000 elementos tomados al azar
m1_weibull2 = rweibull(10000, 1.5, 50)
hist(m1_weibull2, main = "Muestra de tamaño 10000")

# Tomando 1000 promedios de las 1000 muestras como la anterior
datos_weibull2 = replicate(1000, mean(rweibull(10000, 1.5, 50)))
hist(datos_weibull2, main = "Promedios de 1000 muestras de tamaño 10000")

```

n distribución Weibull alf Muestra de tamaño 10000 los de 1000 muestras de tamaño 10000



En la primera grafica podemos identificar la distribucion de la poblacion donde la grafica muestra una leptocurtica junto con un sesgo positivo y en este caso se puede identificar un gran sesgo a la derecha, tambien podemos identificar donde se encuentran la mayoría de los valores que en este caso es en el valor de 50

En la segunda grafica podemos identificar igual una grafica que muestra una leptocurtica con un sesgo positivo y donde la frecuencia de los datos se encuentra entre los 0 a 50, junto tambien que la primera grafica un sesgo muy pronunciado a la derecha a la derecha.

En la tercera grafica podemos identificar un distribución normal ya que la mayor frecuencia de datos se encuentra muy cercano a la media.

```

# Cálculo para la muestra
sg_weibull12 = skewness(m1_weibull12)
k_weibull12 = kurtosis(m1_weibull12)

cat("Sesgo: ", sg_weibull12)

## Sesgo: 1.096582

cat(" Curtosis: ", k_weibull12)

## Curtosis: 4.438335

jarque.test(m1_weibull12)

##
## Jarque-Bera Normality Test
##
## data: m1_weibull12
## JB = 2866.2, p-value < 2.2e-16
## alternative hypothesis: greater

```

En este caso se puede identificar que el valor de P es muy pequeño así que se rechaza la hipótesis de normalización, junto con un sesgo positivo grande y una curtosis mayor a 3 con esto podemos identificar que la Distribución de los datos no es una Distribución Normal

```

# Cálculo para las medias
sg_medias_weibull12 = skewness(datos_weibull12)
k_medias_weibull12 = kurtosis(datos_weibull12)

cat("Sesgo_media: ", sg_medias_weibull12)

## Sesgo_media: 0.01305556

cat(" Curtosis_media: ", k_medias_weibull12)

## Curtosis_media: 2.745959

jarque.test(datos_weibull12)

##
## Jarque-Bera Normality Test
##
## data: datos_weibull12
## JB = 2.7174, p-value = 0.257
## alternative hypothesis: greater

```

En este caso con las medias de los datos podemos identificar que el valor de P es grande y no rechaza la hipótesis de normalidad, junto con un sesgo negativo muy cercano a 0 y una curtosis muy cercana a 3 con esto podemos identificar que la distribución de las medias de los datos es una Distribución Normal

## Inciso E

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

## 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

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

### Inciso A.

¿Cuál es la probabilidad de que al 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?

Probabilidad de X este entre 100 unidades 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)

```
# z = (x - miu) / sigma
z = (10100 - 10000) / 500
cat("z =", z)
```

```
## z = 0.2
```

### 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?

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

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

```
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
```

### Desviaciones lejos de la media(z)

```
# z = (x - miu) / sigma
z1 = (10100 - 10000) / (500/sqrt(120))
cat("z =", z1)

## z = 2.19089
```

### 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?

$$P(9900 < X^- < 10100)$$

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

```
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
```

### Desviaciones lejos de la media(z)

```
# z = (x - miu) / sigma
z2 = (10100 - 10000) / (500/sqrt(15))
cat("z =", z2)

## z = 0.7745967
```

### 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?.

$$P(X^- \leq 9800)$$

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

```
p4 = pnorm(9800, 10000, 500/sqrt(120))
cat("P( X_b <= 9800 ) =", p4)
```

```
## P( X_b <= 9800 ) = 5.88567e-06
```

### Desviaciones lejos de la media(z)

```
# z = (x - miu) / sigma
z4 = (9800 - 10000) / (500/sqrt(120))
cat("z =", z4)

## z = -4.38178
```

Si hizo bien porque tiene 4 desviaciones estandar de la media y tiene una pequeña probabilidad

### Inciso E.

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

$P(X^- \leq 9995)$

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

```
p5 = pnorm(9925, 10000, 500/sqrt(120))
cat("P( X_b <= 9925 ) =", p5)

## P( X_b <= 9925 ) = 0.05017412
```

### Desviaciones lejos de la media(z)

```
# z = (x - miu) / sigma
z5 = (9925 - 10000) / (500/sqrt(120))
cat("z =", z5)

## z = -1.643168
```

Ahora si se aceptaria

## 3. Embotellado

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.

### Inciso A

¿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?



95% central -> en el centro tenemos el 95% de los datos

```
z6 = qnorm(.025, 0, 1) # USAMOS 0 y 1 porque
cat("z =", z6)

## z = -1.959964
```

### Inciso B

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

$P(16 \geq \bar{X})$

```
p7 = 1 - pnorm(16, 15, 1/sqrt(10)) # 16 > X_b
cat("P( X_b <= 16 ) =", p7)

## P( X_b <= 16 ) = 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?

```
# z = (x - miu) / sigma
z8 = (16 - 15) / (1/sqrt(10))
cat("z =", z8)

## z = 3.162278
```

Si se tendría que detener la producción ya que esta 3 desviaciones estándar lejos de la media

### 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?

$P(14.5 \leq \bar{X})$

```
p9 = pnorm(14.5, 15, 1/sqrt(10)) # 16 > X_b
cat("P( X_b <= 14.5 ) =", p9)

## P( X_b <= 14.5 ) = 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?

```
# z = (x - miu) / sigma
z10 = (15.5 - 15) / (1/sqrt(10))
cat("z =", z10)
```

```
## z = 1.581139
```

Si se podría recalibrar ya que el número de desviaciones estándar es mayor a 1 y si queremos ser más exactos con nuestra media deseada de onzas

Hacer una gráfica del inciso 1.

```
# Parámetros de la distribución normal
mu <- 15      # Media
sigma <- 1    # Desviación estándar

# Valores para el eje x
x <- seq(mu - 4*sigma, mu + 4*sigma, length=100)

# Función de densidad de la distribución normal
y <- dnorm(x, mean=mu, sd=sigma)

# Gráfico de la distribución normal
plot(x, y, type="l", lwd=2, col="blue",
      xlab="Cantidad de líquido (onzas)", ylab="Densidad",
      main="Distribución normal de la cantidad de líquido dosificado")

# Agregar líneas para el intervalo del 95% central
abline(v = mu + qnorm(0.025, 0, 1)*sigma, col="red", lwd=2, lty=2)
abline(v = mu + qnorm(0.975, 0, 1)*sigma, col="red", lwd=2, lty=2)

# Etiquetas para los puntos críticos
text(mu + qnorm(0.025, 0, 1)*sigma, 0.05, round(mu + qnorm(0.025, 0,
1)*sigma, 2), pos=4, col="red")
text(mu + qnorm(0.975, 0, 1)*sigma, 0.05, round(mu + qnorm(0.975, 0,
1)*sigma, 2), pos=4, col="red")
```

### Distribución normal de la cantidad de líquido dosific

