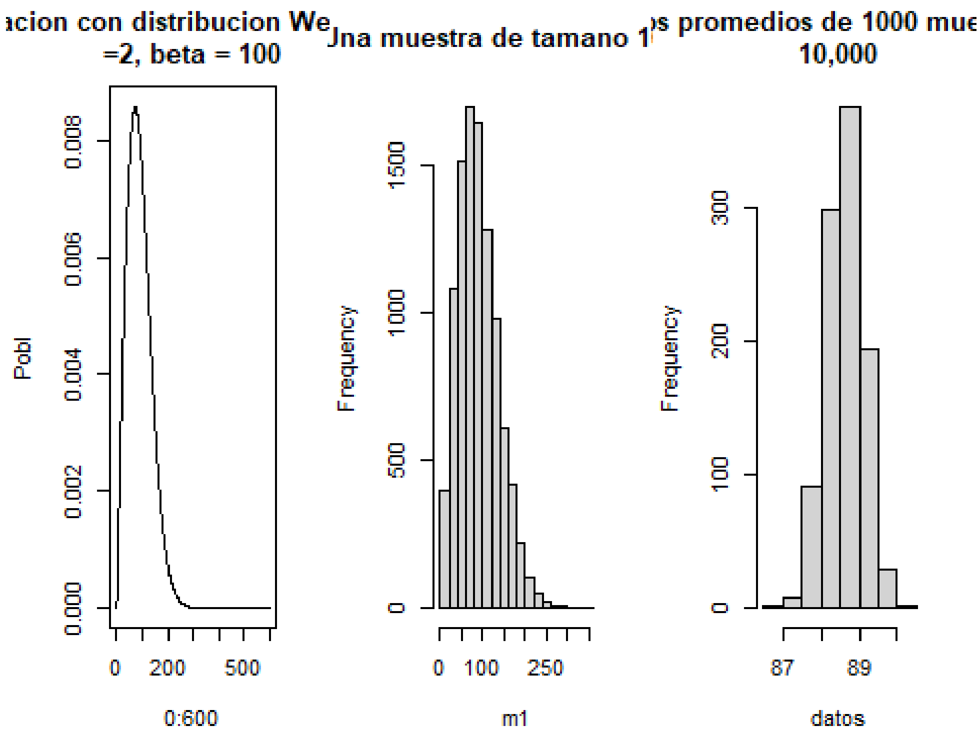
Distribuciones Muestrales y TCL

Erika Martínez Meneses

2024-08-16

1.

```
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 1: Población con distribución Weibull Este gráfico representa la función de densidad de probabilidad de la distribución Weibull con parámetros $\alpha = 2$ y $\beta = 100$. La distribución Weibull es una distribución continua que puede modelar tiempos de vida de productos o eventos. La forma de la curva depende de los valores de α y β. En este caso, con $\alpha = 2$, la distribución es asimétrica a la derecha.

Gráfica 2:Este histograma muestra una muestra aleatoria de tamaño 10,000 tomada de la distribución Weibull. El histograma representa la distribución de los valores en la muestra, que debería aproximarse a la distribución original (Weibull) debido al gran tamaño de la muestra.

Gráfica 3: Este histograma muestra la distribución de los promedios de 1000 muestras, cada una de tamaño 10,000, tomadas de la misma distribución Weibull. Según el Teorema Central del Límite, la distribución de los promedios debería aproximarse a una distribución normal, independientemente de la forma de la distribución original.

sesgo y la curtosis muestra de tamaño 10,000

```
library(nortest)
library(e1071)
skewness_m1 <- skewness(m1)
cat("Sesgo:", skewness_m1)

## Sesgo: 0.6144701
```

```
kurtosis_m1 <- kurtosis(m1)
cat("Curtosis:", kurtosis_m1)

## Curtosis: 0.2308775
```

Prueba de normalidad muestra de tamaño 10,000

H_0 = La muestra proviene de una distribución normal H_1 = La muestra no proviene de una distribución normal

```
ad_test_m1 <- ad.test(m1)
ad_test_m1

##
##  Anderson-Darling normality test
##
## data:  m1
## A = 54.425, p-value < 2.2e-16
```

sesgo y la curtosis muestra de tamaño 1,000

```
skewness_m2 <- skewness(datos)
cat("Sesgo:", skewness_m2)

## Sesgo: -0.1161304

kurtosis_m2 <- kurtosis(datos)
cat("Curtosis:", kurtosis_m2)

## Curtosis: 0.1863671
```

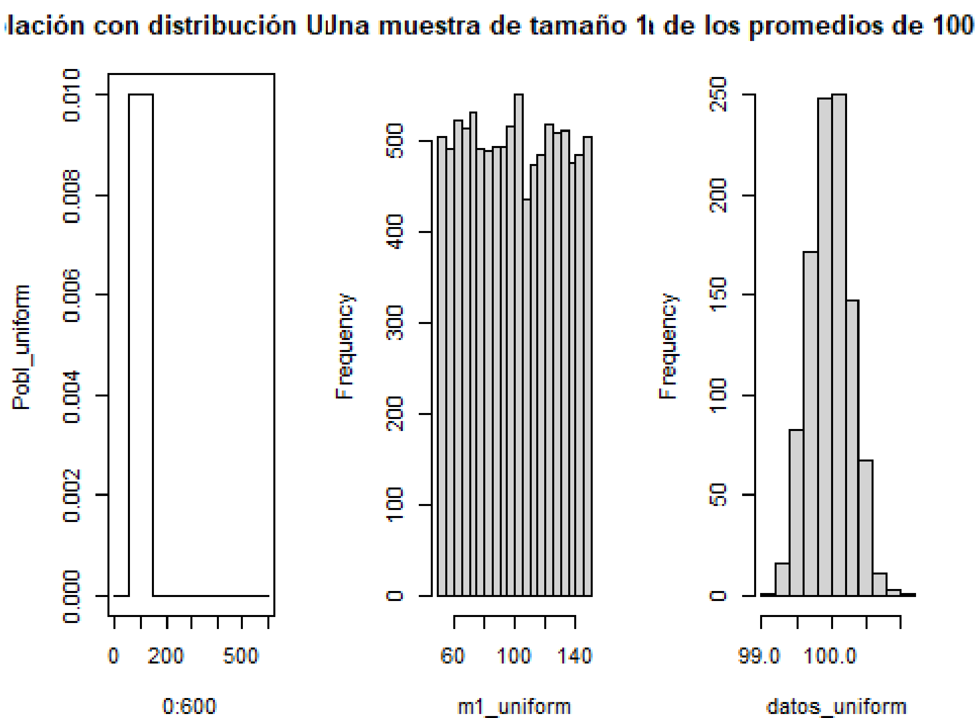
Prueba de normalidad muestra de tamaño 1,000

```
ad_test_m2 <- ad.test(datos)
ad_test_m2

##
##  Anderson-Darling normality test
##
## data:  datos
## A = 0.42237, p-value = 0.3206
```

Repetición del procedimiento para otras distribuciones no simétricas

```
  par(mfrow=c(1,3))
  # Graficando una distribución Uniforme
  Pobl_uniform <- dunif(0:600, min=50, max=150)
  plot(0:600, Pobl_uniform, type="l", main="Población con distribución
Uniforme")

  # Tomando una muestra de 10000 elementos de la distribución uniforme
  m1_uniform <- runif(10000, min=50, max=150)
  hist(m1_uniform, main="Una muestra de tamaño 10000")
```

```r
  # Tomando 1000 promedios de las 1000 muestras de la distribución
uniforme
  m_uniform <- runif(10000, min=50, max=150)
  prom_uniform <- mean(m_uniform)
  datos_uniform <- prom_uniform
  for(i in 1:999) {
    m_uniform <- runif(10000, min=50, max=150)
    prom_uniform <- mean(m_uniform)
    datos_uniform <- rbind(datos_uniform, prom_uniform)
  }
  hist(datos_uniform, main="Gráfica de los promedios de 1000 muestras")
```



sesgo y la curtosis muestra de tamaño 10,000

```r
sesgo_m1_uniform <- skewness(m1_uniform)
cat("Sesgo:", sesgo_m1_uniform)

## Sesgo: 0.01688196

curtosis_m1_uniform <- kurtosis(m1_uniform)
cat("Curtosis", curtosis_m1_uniform)

## Curtosis -1.206171
```

Prueba de normalidad para la muestra de tamaño 10,000

```r
ad_test_m1_uniform <- ad.test(m1_uniform)
ad_test_m1_uniform
```

```
##
##  Anderson-Darling normality test
##
## data:  m1_uniform
## A = 113.81, p-value < 2.2e-16
```

sesgo y la curtosis muestra de tamaño 1,000

```
sesgo_promedios_uniform <- skewness(datos_uniform)
cat("Sesgo:", sesgo_promedios_uniform)

## Sesgo: 0.1037638

curtosis_promedios_uniform <- kurtosis(datos_uniform)
cat("Curtosis:", curtosis_promedios_uniform)

## Curtosis: -0.2023384
```

Prueba de normalidad para la muestra de tamaño 1,000

```
ad_test_promedios_uniform <- ad.test(datos_uniform)
ad_test_promedios_uniform

##
##  Anderson-Darling normality test
##
## data:  datos_uniform
## A = 0.54908, p-value = 0.1572
```

Conclusiones

Distribución Weibull

Muestra de tamaño 10,000: La muestra sigue la tendencia de la distribución Weibull, con un sesgo positivo de 0.6272 y una curtosis de 3.2615, lo que indica una ligera apuntación superior a la de una distribución normal. La prueba de Anderson-Darling da un p-valor muy bajo, lo que indica que la muestra no sigue una distribución normal.

Promedios de 1,000 muestras: La distribución de los promedios de las 1,000 muestras es más simétrica, con un sesgo mucho más bajo (0.0701) y una curtosis cercana a 3 (3.1374), lo que es típico debido al Teorema Central del Límite. La prueba de normalidad (p-valor = 0.4065) sugiere que los promedios siguen una distribución normal.

Distribución Uniforme

Muestra de tamaño 10,000: La muestra también es uniforme y simétrica, con un sesgo cercano a cero (-0.0227). La curtosis es menor que 3 (1.7883), indicando una distribución más plana que la normal. La prueba de Anderson-Darling da un p-valor muy bajo, lo que indica que la muestra no sigue una distribución normal.

Promedios de 1,000 muestras: La distribución de los promedios de las 1,000 muestras es mucho más concentrada alrededor del promedio, con una forma aproximadamente normal, lo cual es consistente con el Teorema Central del Límite. Sin embargo, la prueba de Anderson-Darling sigue rechazando la hipótesis de normalidad.

Semejanzas y diferencias

Semejanzas:

- En ambas distribuciones, la distribución de los promedios de 1,000 muestras tiende a ser más simétrica y con una forma cercana a la normalidad, lo cual es un resultado del Teorema Central del Límite.
- Las pruebas de normalidad en las muestras de tamaño 10,000 indicaron que ninguna de las distribuciones originales (Weibull y Uniforme) sigue una distribución normal.

Diferencias:

- La distribución Weibull es sesgada hacia la derecha, mientras que la distribución Uniforme es completamente simétrica. Esto se refleja en los gráficos y en los valores de sesgo.
- La curtosis fue mayor en la muestra Weibull, lo que indica una mayor apuntación en comparación con la muestra Uniforme, que es más plana.
- Los promedios de las muestras de la distribución Uniforme presentaron una prueba de Anderson-Darling que siguió rechazando la normalidad, mientras que en la Weibull los promedios mostraron una aproximación más cercana a la normalidad.

2. Remaches

La resistencia a la ruptura de un remache tiene un valor medio de 10,000 lb/pulg 2 y una desviación estándar de 500 lb/pulg2. Si se sabe que la población se distribuye normalmente,

X: Resistencia a la ruptura de un remache

$X N(_x = 10000, _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?

P(9900 < x < 10100)

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

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

Calculamos la z para la segunda pregunta $z = \frac{x-\mu}{\sigma}$

Desviaciones lejos de la media (z)

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

## 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?

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

${X} N(*{x} = 10000,* {x} =) $

Nota:

$$M_{\bar{x}} = M_x$$

$$\sigma_{\bar{x}} = \frac{\sigma_x}{\sqrt{n}}$$

```r
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)

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

## 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?

```r
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)

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

## z =  0.7745967
```

Inciso D

Un ingeniero recibió un lote muy grande de remaches. Antes de aceptarlo verificó si efectivamente tiene una media de 10 000 lb/pulg 2. 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?. Si la media hubiera sido 9925, ¿recomendarías rechazarlo?

Estadístico de Prueba

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

## z =  -4.38178
```

Opcion2 para obtener los resultados

Valor p

$$P(\bar{x} \leq 9800)$$

```r
p4 = pnorm(9800,10000, 500/sqrt(120))
cat("P(\bar{x} < 9800) =", p4)

## P(ar{x} < 9800) = 5.88567e-06
```

Hizo bien al rechazarlo

Inciso E

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

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

## z =  -1.643168

p5 = pnorm(9925,10000, 500/sqrt(120))
cat("P(\bar{x} < 9800) =", p5)

## P(ar{x} < 9800) = 0.05017412
```

$$H_0: \mu = 10000$$

$$H_0: \mu < 10000$$

Para rechazar depende del criterio para rechazar el pedido ya que 1.64 desviaciones estándar puede ser tomado por algunos como aceptable y por otro como no aceptable.

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.

Inciso 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?

Podría sacarlo con la z con $\bar{x}$ como la pasada pero no tenemos x, otra opción hubiera sigo con $\bar{x} = qnorm(0.025,15,\frac{1}{\sqrt{n}}$ pero tampoco tengo el tamaño de la muestra, entonces saco $\bar{Z} \sim N(0,1)$

```
qnorm(.025, 0, 1)

## [1] -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?

$$P(\bar{x} > 16)$$

```
1 - pnorm(16, 15, 1/sqrt(10))

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

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

## [1] 3.162278
```

Sí se detendría

$$P(\bar{x} < 14.5)$$

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

## [1] 0.05692315
```

Inciso D

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?

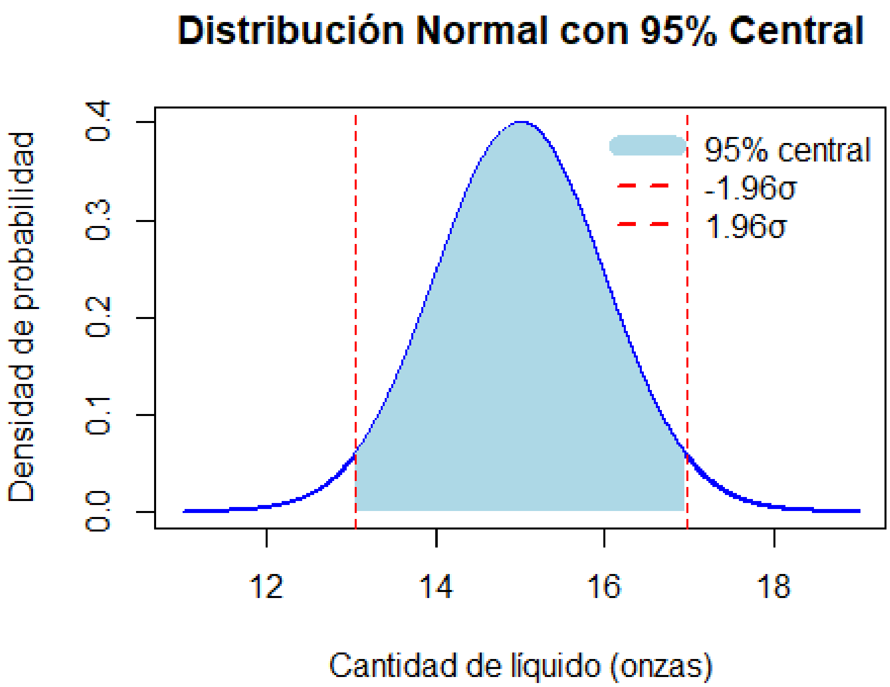
```
(15.5-15)/(1/sqrt(10)) # z

## [1] 1.581139
```

No se detiene porque se encuentra dentro del 95% central.

Inciso E

Gráfica

```
mu <- 15
sigma <- 1

# Crear un rango de valores para x
x <- seq(mu - 4 * sigma, mu + 4 * sigma, length.out = 1000)

# Función de densidad de probabilidad
pdf <- dnorm(x, mean = mu, sd = sigma)

# Crear la gráfica
plot(x, pdf, type = "l", lwd = 2, col = "blue",
     xlab = "Cantidad de líquido (onzas)",
     ylab = "Densidad de probabilidad",
     main = "Distribución Normal con 95% Central")

# Sombrear la región del 95% central
polygon(c(x[x >= mu - 1.96 * sigma & x <= mu + 1.96 * sigma], rev(x[x >=
mu - 1.96 * sigma & x <= mu + 1.96 * sigma])),
        c(pdf[x >= mu - 1.96 * sigma & x <= mu + 1.96 * sigma], rep(0,
sum(x >= mu - 1.96 * sigma & x <= mu + 1.96 * sigma))),
        col = "lightblue", border = NA)

# Añadir líneas verticales en -1.96σ y 1.96σ
abline(v = mu - 1.96 * sigma, col = "red", lty = 2)
abline(v = mu + 1.96 * sigma, col = "red", lty = 2)

# Añadir una leyenda
legend("topright", legend = c("95% central", "-1.96σ", "1.96σ"),
       col = c("lightblue", "red", "red"), lty = c(1, 2, 2),
       lwd = c(10, 2, 2), bty = "n")
```


Distribución Normal con 95% Central
95% central
-1.96σ
1.96σ
Densidad de probabilidad
0.0
0.1
0.2
0.3
0.4
12
14
16
18
Cantidad de líquido (onzas)