6. Distribuciones Muestrales y TCL

Juan Bernal

2024-08-16

Pregunta 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. Ejecutar el siguiente código de R: DistrsM_enR.txt. Se esperan tres gráficas, interprete cad
a una de ellas. Se usa una distribución Weibull, con parámetros alfa = 2 y beta = 100.
```

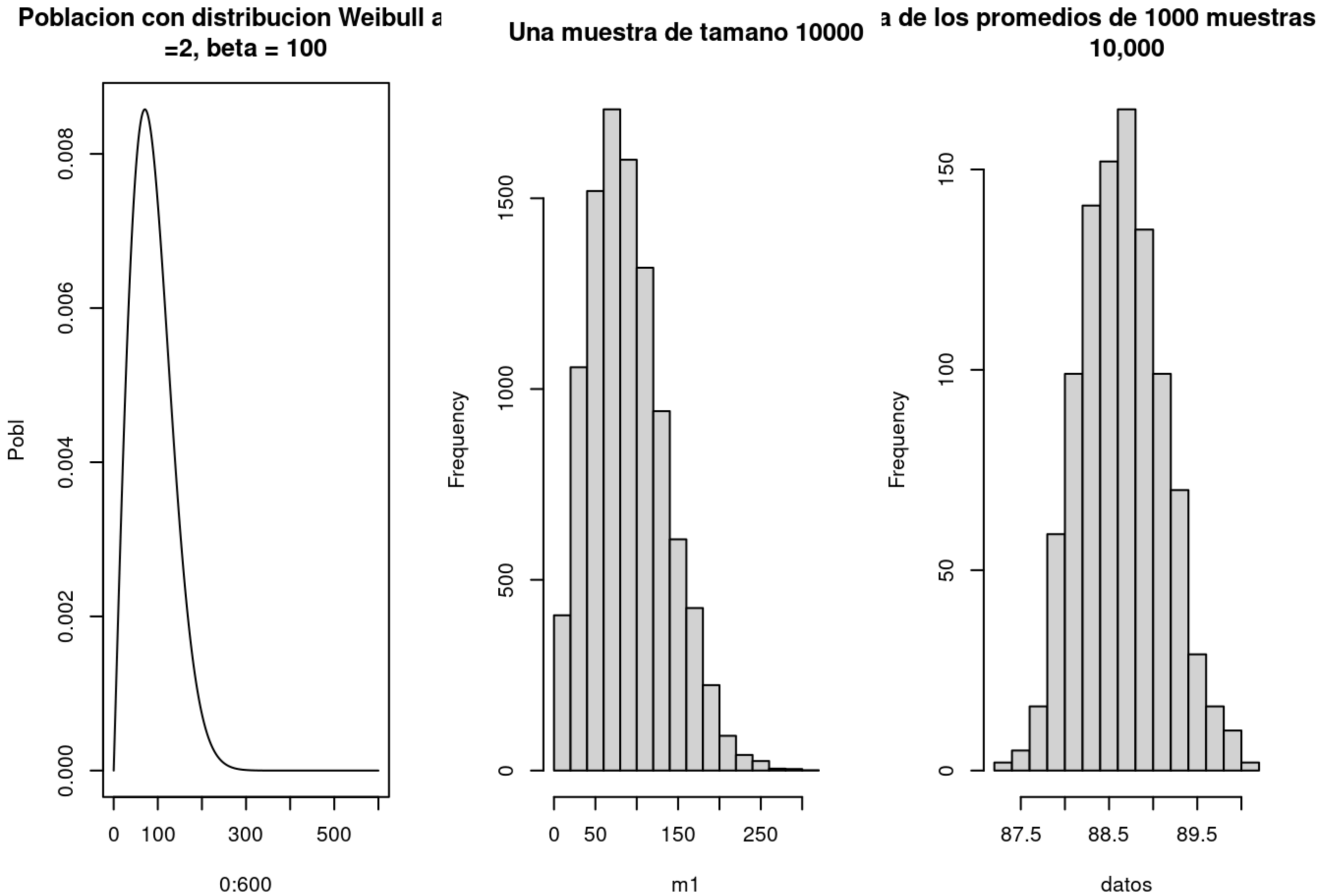
```
par(mfrow=c(1,3))

# 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")
```



La primera gráfica es una distribución Weibull de $X \sim Weibull(\alpha = 2, \beta = 100)$, que parece tener un sesgo a la derecha hacia el infinito.

La segunda gráfica es una muestra de 10,000 datos tomados al azar de la distribución Weibull. Parece presentar el mismo tipo de distribución que la población, con el mismo sesgo a la derecha, pero esta vez no parece irse al infinito.

La tercera y última gráfica presenta 1000 promedios de 1000 muestras de 10,000 datos tomados al azar cada uno. Esta gráfica presenta una acumulación al centro, acercándose a la apariencia de una distribución normal, diferente a la de las primeras 2 gráficas.

```
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.
```

```r
library(moments)
cat("La curtosis de la muestra de tamaño 10000 es ", kurtosis(m1),"y el sesgo es",skewness(m1))
```

```
## La curtosis de la muestra de tamaño 10000 es  3.193843 y el sesgo es 0.6112203
```

```r
library(nortest)
cat("El p-value de la muestra de tamaño 10,000 es ",ad.test(m1)$p)
```

```
## El p-value de la muestra de tamaño 10,000 es  3.7e-24
```

Como podemos observar, la muestra tiene una curtosis muy grande y sesgo pequeño, además de tener un p-value muy pequeño, muhícismo menor al valor de $\alpha = 0.05$ en una prueba de hipótesis. Por lo tanto, la muestra de tamaño 10,000 no es 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 media
s de las muestras.
```

```
cat("La curtosis de las medias de las 1000 muestras es ", kurtosis(datos),"y el sesgo es",skewne
ss(datos))
```

```
## La curtosis de las medias de las 1000 muestras es  2.811889 y el sesgo es 0.1868828
```

```
cat("El p-value de las medias de las 1000 muestras es ",ad.test(datos)$p)
```

```
## El p-value de las medias de las 1000 muestras es  0.06161491
```

Como se puede observar, se muestra una curtosis grande y un sesgo pequeño, por lo que es difícil dar una inferencia acertada. Pero, gracias al p-value sabemos que la distribución de las medias de las 1000 muestras es normal, ya que es mayor a $\alpha = 0.05$ en una prueba de hipótesis, por lo que no rechazamos la hipótesis nula y asumimos que la distribución es 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 distrib
ución, como la uniforme (punif y runif). Interpreta los resultados.
```

Primero probaremos con una distribución Weibull de $\alpha = 5$ y $\beta = 300$.

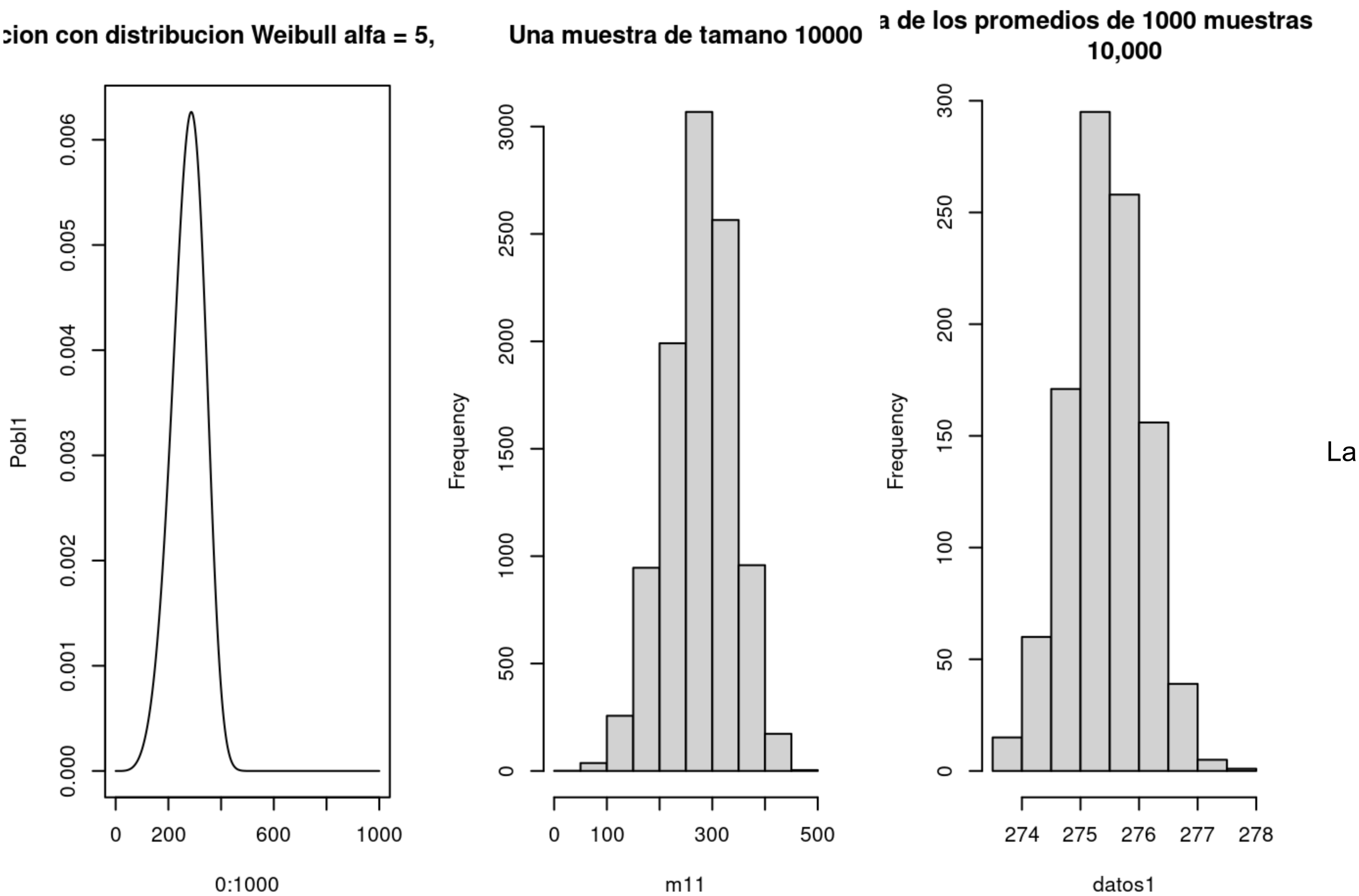
```
par(mfrow=c(1,3))

# Graficando una distribucion Weibull de alfa = 5, beta = 300
Pobl1 = dweibull(0:1000,5, 300)
plot(0:1000,Pobl1, type="l", main = "Poblacion con distribucion Weibull alfa = 5, beta = 300")

# Tomando una muestra de 10000 elementos tomados al azar
m11 = rweibull(10000, 5, 300)
hist(m11, main = "Una muestra de tamano 10000")

# Tomando 1000 promedios de las 1000 muestras como la anterior
  mm =rweibull(10000,5,300)
  prom=mean(mm)
  datos1=prom
for(i in 1:999) {
  mm =rweibull(10000,5,300)
  prom=mean(mm)
  datos1=rbind(datos1,prom) }

hist(datos1, main="Grafica de los promedios de 1000 muestras de tamano
10,000")
```



La primera gráfica es una distribución Weibull de $X \sim Weibull(\alpha = 5, \beta = 300)$, que parece no tener un sesgo, si no que se asemeja a una distribución normal.

La segunda gráfica es una muestra de 10,000 datos tomados al azar de la distribución Weibull. Parece presentar el mismo tipo de distribución que la población, con la misma similitud a una distribución normal.

La tercera y última gráfica presenta 1000 promedios de 1000 muestras de 10,000 datos tomados al azar cada uno. Esta gráfica presenta una acumulación al centro, acercándose a la apariencia de una distribución normal.

```
cat("La curtosis de la muestra de tamaño 10,000 es ", kurtosis(m11),"y el sesgo es",skewness(m1
1))
```

```
## La curtosis de la muestra de tamaño 10,000 es  2.918364 y el sesgo es -0.2436573
```

```
cat("El p-value de la muestra de tamaño 10,000 es ",ad.test(m11)$p)
```

```
## El p-value de la muestra de tamaño 10,000 es  3.7e-24
```

Como podemos observar, la muestra tiene una curtosis muy grande y sesgo pequeño, además de tener un p-value muy pequeño, muhícismo menor al valor de $\alpha = 0.05$ en una prueba de hipótesis. Por lo tanto, la muestra de tamaño 10,000 no es normal.

```
cat("La curtosis de las medias de las 1000 muestras es ", kurtosis(datos1),"y el sesgo es",skewn
ess(datos1))
```

```
## La curtosis de las medias de las 1000 muestras es  2.825018 y el sesgo es -0.0743627
```

```
cat("El p-value de las medias de las 1000 muestras es ",ad.test(datos1)$p)
```

```
## El p-value de las medias de las 1000 muestras es  0.08410977
```

Como se puede observar, se muestra una curtosis grande y un sesgo pequeño, por lo que es difícil dar una inferencia acertada. Pero, gracias al p-value sabemos que la distribución de las medias de las 1000 muestras es normal, ya que es mayor a $\alpha = 0.05$ en una prueba de hipótesis, por lo que no rechazamos la hipótesis nula y asumimos que la distribución es normal.

Ahora probaremos con una distribución uniforme de $a = 1$ y $b = 10$.

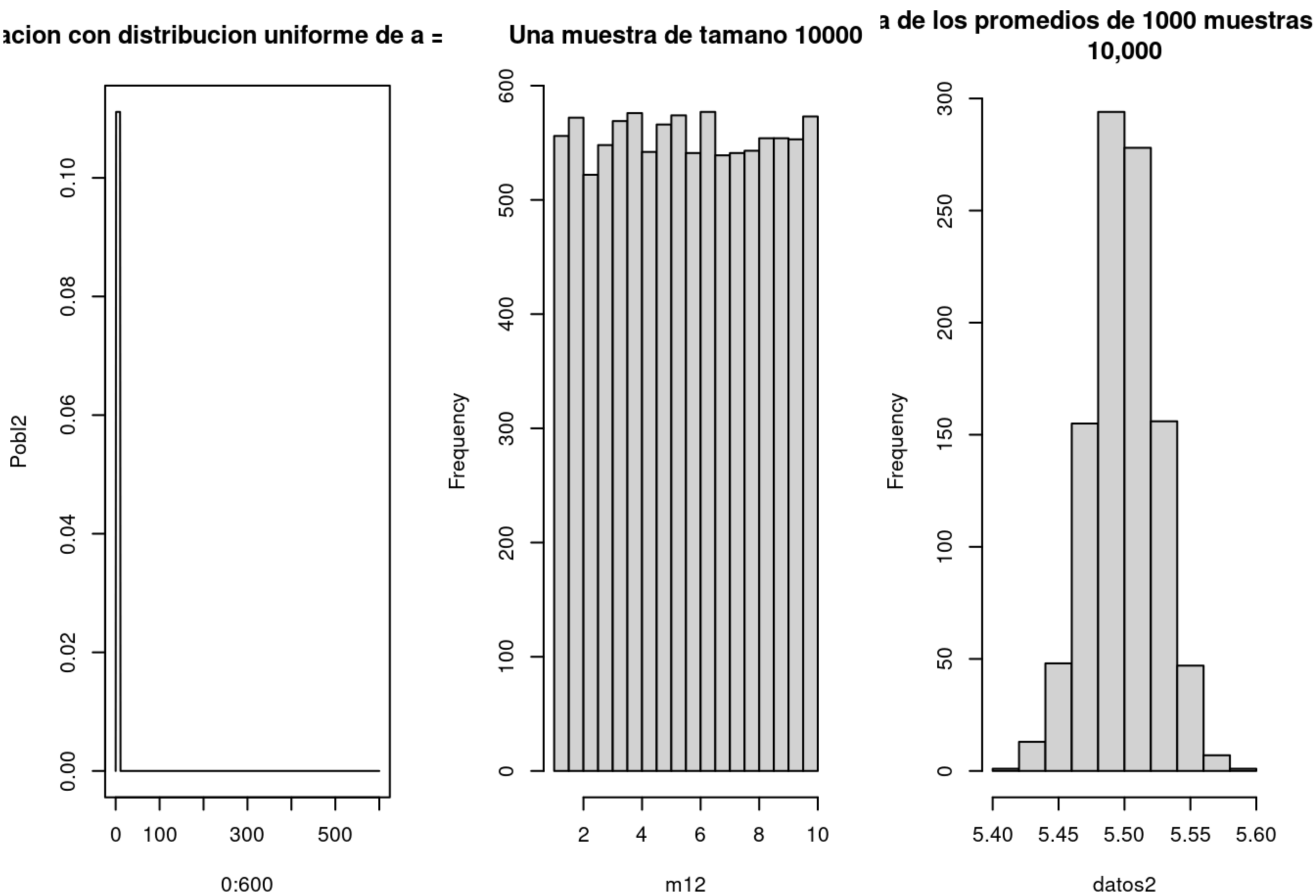
```
par(mfrow=c(1,3))

# Graficando una distribucion uniforme de a = 1, b = 10
Pobl2 = dunif(0:600,1, 10)
plot(0:600,Pobl2, type="l", main = "Poblacion con distribucion uniforme de a = 1, b = 10")

# Tomando una muestra de 10000 elementos tomados al azar
m12 = runif(10000, 1, 10)
hist(m12, main = "Una muestra de tamano 10000")

# Tomando 1000 promedios de las 1000 muestras como la anterior
  m2 =runif(10000,1, 10)
  prom=mean(m2)
  datos2=prom
for(i in 1:999) {
  m2 =runif(10000,1, 10)
  prom=mean(m2)
  datos2=rbind(datos2,prom) }

hist(datos2, main="Grafica de los promedios de 1000 muestras de tamano
10,000")
```



La primera gráfica es una distribución uniforme de $X \sim Unif(a = 1, b = 10)$, que parece solo una línea recta sin pendiente.

La segunda gráfica es una muestra de 10,000 datos tomados al azar de la distribución Uniforme anterior. Aquí no notamos ningún patrón de distribución ya que las frecuencias son todas casi iguales, son "uniformes".

La tercera y última gráfica presenta 1000 promedios de 1000 muestras de 10,000 datos tomados al azar cada uno. Esta gráfica presenta una acumulación al centro, acercándose a la apariencia de una distribución normal.

```
cat("La curtosis de la muestra de tamaño 10,000 es ", kurtosis(m12),"y el sesgo es",skewness(m1
2))
```

```
## La curtosis de la muestra de tamaño 10,000 es  1.805776 y el sesgo es 0.005611495
```

```
cat("El p-value de la muestra de tamaño 10,000 es ",ad.test(m12)$p)
```

```
## El p-value de la muestra de tamaño 10,000 es  3.7e-24
```

Como podemos observar, la muestra tiene una curtosis muy grande y sesgo pequeño, además de tener un p-value muy pequeño, muhícismo menor al valor de $\alpha = 0.05$ en una prueba de hipótesis. Por lo tanto, la muestra de tamaño 10,000 no es normal.

```
cat("La curtosis de las medias de las 1000 muestras es ", kurtosis(datos2),"y el sesgo es",skewn
ess(datos2))
```

```
## La curtosis de las medias de las 1000 muestras es  3.023605 y el sesgo es -0.1013461
```

```
cat("El p-value de las medias de las 1000 muestras es ",ad.test(datos2)$p)
```

```
## El p-value de las medias de las 1000 muestras es  0.2496975
```

Como se puede observar, se muestra una curtosis grande y un sesgo pequeño, por lo que es difícil dar una inferencia acertada. Pero, gracias al p-value sabemos que la distribución de las medias de las 1000 muestras es normal, ya que es mayor a $\alpha = 0.05$ en una prueba de hipótesis, por lo que no rechazamos la hipótesis nula y asumimos que la distribución es normal.

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

A través de los incisos se mencionaron las diferencias y semejanzas entre las gráficas de las distribuciones y qué pasa si solo tomamos muestras y medias, pero la conclusión a la que podemos llegar es que la distribución de las medias de muestras de la población tienden a comportarse de forma normal, es decir, forman una distribución normal, a pesar de la asimetría que la distribución poblacional muestre.

Pregunta 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,

$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?
```

$P(9900 \leq X \leq 10100)$

```
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 l
a media?
```

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

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

```
p2 = pnorm(10100,10000,500/sqrt(120)) - pnorm(9900,10000,500/sqrt(120))
cat("P(9900 < x_bar < 10100) = ",p2)
```

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

Desviaciones lejos de la media (z)

```
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 re
sistencia media a la ruptura esté 100 unidades alrededor de la media? ¿a cuántas desviaciones es
tándar está de la media?
```

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

$\bar{X} \sim N(\mu_{\bar{x}} = 10000, \sigma_{\bar{x}} = \frac{500}{\sqrt{15}})$

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

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

Desviaciones lejos de la media (z)

```
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 ef
ectivamente 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?.
```

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

$\bar{X} \sim N(\mu_{\bar{x}} = 9800, \sigma_{\bar{x}} = \frac{500}{\sqrt{120}})$

Desviaciones lejos de la media (z)

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

```
## Z =  -4.38178
```

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

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

Hizo lo correcto, pues la probabilidad de que la media de la resistencia de los remaches sea menor a 9800 lb/pulg2 es muchísimo menor al 1%. Además, está demasiado desviado de la normalidad de la población general.

```
e) ¿Qué decisión recomiendas al ingeniero si la media obtenida en la media hubiera sido 9925? ¿r
ecomendarías rechazarlo?
```

$P(\bar{x} \leq 9925)$

$\bar{X} \sim N(\mu_{\bar{x}} = 9925, \sigma_{\bar{x}} = \frac{500}{\sqrt{120}})$

Desviaciones lejos de la media (z)

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

```
## Z =  -1.643168
```

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

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

Recomiendo rechazarlo, pues la probabilidad sigue siendo pequeña, de solo un 5% probabilidades de que la media de la resistencia de los remaches sea menor a 9925 lb/pulg2, pero es mejor que el lote con media de 9800 lb/pulg2. Se desvía 1.64 desviaciones estándar de la media poblacional.

Pregunta 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.

$X =$ Descarga en botella por onzas

$X \sim N(\mu_x = 15, \sigma_x = 1)$

```
1. ¿A cuántas desviaciones estándar alrededor de la verdadera media μ puede estar la media de un
a muestra para que esté dentro del estándar establecido del 95% central?
```

Desviaciones lejos de la media (z)

```
z1 = qnorm(0.975,0,1)
cat("La media de una muestra debe estar a ",z1," desviaciones estándar alrededor de la verdadera
media para que esté dentro del estándar establecido del 95% central")
```

```
## La media de una muestra debe estar a  1.959964  desviaciones estándar alrededor de la verdade
ra media para que esté dentro del estándar establecido del 95% central
```

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

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

$\bar{X} \sim N(\mu_{\bar{x}} > 16, \sigma_{\bar{x}} = \frac{1}{\sqrt{10}})$

```
p1 = 1 - pnorm(16,15,1/sqrt(10))
cat("P(16 < x_bar) = ",p1)
```

```
## P(16 < x_bar) =  0.0007827011
```

```
3. Si en una muestra aleatoria de tamaño 10 botellas se obtuvo una media de 16 onzas, ¿se detend
ría la producción para calibrar la máquina?
```

Desviaciones lejos de la media (z)

```
z2 = (16-15)/(1/sqrt(10))
cat("La media de 16 onzas de una muestra de 10 botellas se encuentra ",z2 ,"desviaciones estánda
r alrededor de la verdadera media.")
```

```
## La media de 16 onzas de una muestra de 10 botellas se encuentra  3.162278 desviaciones estánd
ar alrededor de la verdadera media.
```

Se debería detener la máquina para calibrar, ya que más de 1.96 desviaciones estándar alrededor de la media ya se encuentra fuera del estándar establecido del 95% central.

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

$P(\bar{x} \leq 14.5)\ \bar{X} \sim N(\mu_{\bar{x}} < 14.5, \sigma_{\bar{x}} = \frac{1}{\sqrt{10}})$

```
p2 = pnorm(14.5,15,1/sqrt(10))
cat("P(x_bar < 14.5) = ",p2)
```

```
## P(x_bar < 14.5) =  0.05692315
```

```
5. Si en una muestra aleatoria de tamaño 10 botellas se obtuvo una media de 15.5 onzas, ¿se dete
ndría la producción para calibrar la máquina?
```

$\bar{X} \sim N(\mu_{\bar{x}} = 15.5, \sigma_{\bar{x}} = \frac{1}{\sqrt{10}})$

Desviaciones lejos de la media (z)

```
z3 = (15.5-15)/(1/sqrt(10))
cat("La media de 15.5 onzas de una muestra de 10 botellas se encuentra ", z3,"desviaciones están
dar alrededor de la verdadera media.")
```

```
## La media de 15.5 onzas de una muestra de 10 botellas se encuentra  1.581139 desviaciones está
ndar alrededor de la verdadera media.
```

No es necesario que la máquina se detenga, pues se encuentra dentro del intervalo de desviaciones estándar del centro. 1.58<1.96.

```
6. Hacer una gráfica del inciso 1.
```

```
normal_area <- function(mean = 0, sd = 1, lb, ub, acolor = "lightgray", ...) {
    x <- seq(mean - 3 * sd, mean + 3 * sd, length = 100)

    if (missing(lb)) {
       lb <- min(x)
    }
    if (missing(ub)) {
        ub <- max(x)
    }

    x2 <- seq(lb, ub, length = 100)
    plot(x, dnorm(x, mean, sd), type = "n", ylab = "")

    y <- dnorm(x2, mean, sd)
    polygon(c(lb, x2, ub), c(0, y, 0), col = acolor)
    lines(x, dnorm(x, mean, sd), type = "l", ...)
}

normal_area(mean = 0, sd = 1, ub = qnorm(0.975, mean = 0, sd = 1),
            lwd = 2, acolor = rgb(0, 0, 1, alpha = 0.5))
```