

Distribuciones muestrales

Jacobo Hirsch Rodriguez

2024-08-21

##1. Ensayando Distribuciones

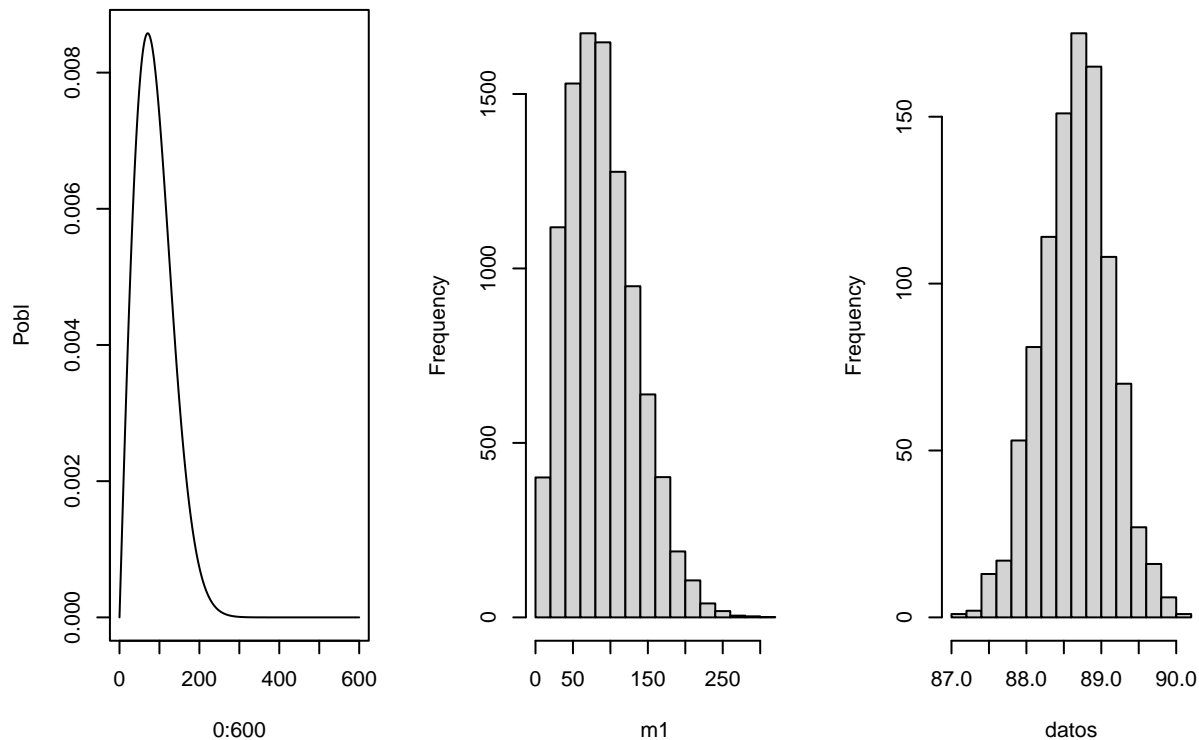
##Grafica la Distribución de una variable aleatoria, la de una muestra elegida al azar y la de la Distribución de las medias de 10000 muestras:

#Ejercutar 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)) #Esta línea de código divide el área de gráficos en 1 fila y 3 columnas. Esto permite  
# 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) } # sirve para concatenar estructuras de datos existentes  
hist(datos, main="Grafica de los promedios de 1000 muestras de tamaño 10,000")
```

on con distribución Weibull alfa=2

Una muestra de tamaño 10000s promedios de 1000 muestras d



en la ultima parte del codigo, especificamenre despues del ciclo for , se grafica un histograma de 1,000 promedios. Esto ilustra la distribución de los promedios de las muestras, que debería aproximarse a una distribución normal debido al teorema del límite central.

la primera grafica muestra una distribucion de weibull teorica con los parametros dados en las que posee un sesgo a la derecha, la segunda nos muestra que si el tamaño de la muestra es lo suficientemente grande será una representación de la distribución teorica original, siguiendo la misma tendencia general. La tercera gráfica muestra el teorema del límite central. Aunque la distribución original es asimétrica, la distribución de los promedios de muchas muestras de tamaño grande tiende a ser normal. Esto sugiere que, independientemente de la forma de la distribución original, si tomamos suficientes muestras grandes, sus promedios seguirán una distribución normal.

##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(moments)
library(nortest)

# Generar la muestra de tamaño 10,000 a partir de la distribución Weibull
set.seed(123) # para reproducibilidad
muestra <- rweibull(10000, 2, 100)

# Calcular el sesgo (skewness)
sesgo <- skewness(muestra)
cat("Sesgo:", sesgo, "\n")
```

Sesgo: 0.6371202

```
# Calcular la curtosis (kurtosis)
curtosis <- kurtosis(muestra)
cat("Curtosis:", curtosis, "\n")
```

```
## Curtosis: 3.250571
```

```
# Aplicar una prueba de normalidad - Prueba de Anderson-Darling
ad_test <- ad.test(muestra)
cat("Prueba de normalidad (Anderson-Darling) - valor p:", ad_test$p.value, "\n")
```

```
## Prueba de normalidad (Anderson-Darling) - valor p: 3.7e-24
```

```
# Conclusión sobre la normalidad de la muestra
if (ad_test$p.value < 0.05) {
  cat("Conclusión: Los datos NO siguen una distribución normal.\n")
} else {
  cat("Conclusión: No se puede rechazar la hipótesis de normalidad.\n")
}
```

```
## Conclusión: Los datos NO siguen una distribución normal.
```

##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. primero convertimos nuestra matriz de datos que se creó anteriormente a un array

```
datos_vector <- as.vector(datos)
```

```
# Calcular el sesgo del vector (skewness)
sesgo_muestras <- skewness(datos_vector)
cat("Sesgo:", sesgo, "\n")
```

```
## Sesgo: 0.6371202
```

```
# Calcular la curtosis del vector (kurtosis)
curtosis_muestras <- kurtosis(datos_vector)
cat("Curtosis:", curtosis, "\n")
```

```
## Curtosis: 3.250571
```

```
# Aplicar una prueba de normalidad - Prueba de Anderson-Darling
ad_test_muestras <- ad.test(datos_vector)
cat("Prueba de normalidad (Anderson-Darling) - valor p:", ad_test_muestras$p.value, "\n")
```

```
## Prueba de normalidad (Anderson-Darling) - valor p: 0.3601598
```

```
# Conclusión sobre la normalidad de la muestra
if (ad_test_muestras$p.value < 0.05) {
  cat("Conclusión: Los datos NO siguen una distribución normal.\n")
} else {
  cat("Conclusión: No se puede rechazar la hipótesis de normalidad.\n")
}
```

Conclusión: No se puede rechazar la hipótesis de normalidad.

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.

Vamos a hacerlo para chi cuadrada

```
# Configurar la semilla para la reproducibilidad
set.seed(123)

# Parámetros
n_muestras = 1000 # número de muestras
tamaño_muestra = 10000 # tamaño de cada muestra
df = 4 # grados de libertad de la distribución chi-cuadrado

# Generar las 1000 muestras y calcular sus medias
datos_chi <- sapply(1:n_muestras, function(x) mean(rchisq(tamaño_muestra, df)))

# Calcular el sesgo
sesgo_chi <- skewness(datos_chi)
cat("Sesgo Chi-cuadrado:", sesgo_chi, "\n")
```

Sesgo Chi-cuadrado: 0.09287594

```
# Calcular la curtosis
curtosis_chi <- kurtosis(datos_chi)
cat("Curtosis Chi-cuadrado:", curtosis_chi, "\n")
```

Curtosis Chi-cuadrado: 3.185495

```
# Prueba de normalidad - Prueba de Anderson-Darling
ad_test_chi <- nortest::ad.test(datos_chi)
cat("Prueba de normalidad (Anderson-Darling) - valor p Chi-cuadrado:", ad_test_chi$p.value, "\n")
```

Prueba de normalidad (Anderson-Darling) - valor p Chi-cuadrado: 0.5069157

```
if (ad_test_chi$p.value < 0.05) {
  cat("Conclusión: Los datos NO siguen una distribución normal.\n")
} else {
  cat("Conclusión: No se puede rechazar la hipótesis de normalidad.\n")
}
```

Conclusión: No se puede rechazar la hipótesis de normalidad.

Vamos a hacerlo para la logaritmica

```
# Configurar la semilla para la reproducibilidad
set.seed(123)

# Parámetros
```

```

lambda = 0.5 # tasa de la distribución exponencial

# Generar las 1000 muestras y calcular sus medias
datos_exp <- sapply(1:n_muestras, function(x) mean(rexp(tamaño_muestra, lambda)))

# Calcular el sesgo
sesgo_exp <- skewness(datos_exp)
cat("Sesgo Exponencial:", sesgo_exp, "\n")

## Sesgo Exponencial: 0.1559132

# Calcular la curtosis
curtosis_exp <- kurtosis(datos_exp)
cat("Curtosis Exponencial:", curtosis_exp, "\n")

## Curtosis Exponencial: 2.986208

# Prueba de normalidad - Prueba de Anderson-Darling
ad_test_exp <- nortest::ad.test(datos_exp)
cat("Prueba de normalidad (Anderson-Darling) - valor p Exponencial:", ad_test_exp$p.value, "\n")

## Prueba de normalidad (Anderson-Darling) - valor p Exponencial: 0.1995271

if (ad_test_exp$p.value < 0.05) {
  cat("Conclusión: Los datos NO siguen una distribución normal.\n")
} else {
  cat("Conclusión: No se puede rechazar la hipótesis de normalidad.\n")
}

```

```
## Conclusión: No se puede rechazar la hipótesis de normalidad.
```

#conclusiones :

En todos los casos, el histograma de los promedios de las 1000 muestras tiende a aproximarse a una distribución normal. Esto es una manifestación del teorema del límite central, que establece que la distribución de los promedios de un número suficientemente grande de muestras independientes se aproxima a una distribución normal, independientemente de la forma de la distribución original.

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

A pesar de las diferencias en la forma de las distribuciones originales, los promedios de las muestras de gran tamaño tienden a una distribución normal en los tres casos, demostrando la robustez del teorema del límite central. Las diferencias más significativas se observan en las distribuciones originales y en las muestras de tamaño 10,000, donde cada distribución teórica muestra su forma y características específicas, como sesgo y curtosis.

##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

$\mu_x = 10000$ //media

$\sigma_x = 500$ //desviacion estandar

Inciso 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?

esto quiere decir, cual es la probabilidad de que x tome un valor entre 9900 y 10100 $P(9900 \leq x \leq 10100)$

```
p1 = pnorm(10100, 10000, 500) - pnorm(9900,10000,500)
```

```
cat("P(9900 < X < 10100) =", p1)
```

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

```
#sabemos que estamos a 100 unidades alrededor de la media por lo que vamos a dividir 100 entre sigma
```

```
p2 <- 100 / 500 # (unidades alrededor de sigma) / sigma
```

```
p2
```

```
## [1] 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?

sabemos que en la media de la muestra es igual a la media de la distribucion original y la desviacion estandas es igual a la sigma de la distribucion sobre la raiz cuadrada de n (el tamaño de la muestra)

```
miu_x_barra = 10000 //media
```

```
sigma_x_barra = 500 / sqrt(120) //desviacion estandar
```

```
sigma_x_barra <- 500 / sqrt(120)
```

```
p3 = pnorm(10100, 10000, sigma_x_barra) - pnorm(9900,10000,sigma_x_barra)
```

```
cat("P(9900 < X < 10100) =", p3)
```

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

```
#sabemos que estamos a 100 unidades alrededor de la media por lo que vamos a dividir 100 entre sigma
```

```
p4 <- 100 / sigma_x_barra # (unidades alrededor de sigma) / sigma
```

```
p4
```

```
## [1] 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?

```

sigma_x_barra_2 <- 500 / sqrt(15)

p5 = pnorm(10100, 10000, sigma_x_barra_2) - pnorm(9900,10000,sigma_x_barra_2)

cat("P(9900 < X < 10100) =", p5)

## P(9900 < X < 10100) = 0.561422

#sabemos que estamos a 100 unidades alrededor de la media por lo que vamos a dividir 100 entre sigma

p6 <- 100 / sigma_x_barra_2 # (unidades alrededor de sigma) / sigma
p6

## [1] 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?.

para poder argumentar eso podemos ver a cuantas desviaciones estandar se encuentra, lo cual nos podría dar la probabilidad de que la media de la muestra apareciera en la distribucion original

$$z = 9800 - 10000 / (500 / \sqrt{120})$$

```

z1 <- (9800-10000) / (500 / sqrt(120))
z1

```

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

```

p7 = pnorm(9800, 10000, sigma_x_barra_2)
p7

```

```
## [1] 0.06066763
```

esto significa que se encuentra muy alejado de la media de la distribucion original por lo que la probabilidad de que se hubiera obtenido es menor

Inciso E

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

hacemos algo similar al inciso anterior por lo que

```

z2 <- (9925-10000) / (500 / sqrt(120))
z2

```

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

se recomienda aceptarlo dependiendo del criterio del comprador

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 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?

```
# Parámetros iniciales
mu <- 15 # media poblacional
sigma <- 1 # desviación estándar poblacional
n <- 10 # tamaño de la muestra

# Calcular los límites del 95% central de la distribución muestral
error_estandar <- sigma / sqrt(n)
z_critico <- qnorm(c(0.025, 0.975)) # percentiles 2.5% y 97.5%

limites <- mu + z_critico * error_estandar

cat("se encuentra a ", z_critico[2], "desviaciones estandar ")
```

```
## se encuentra a 1.959964 desviaciones estandar
```

```
cat("Límites para el 95% central: ", limites, "\n")
```

```
## Límites para el 95% central: 14.3802 15.6198
```

inciso 2

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

```
#Probabilidad de que la media de la muestra sea mayor que 16
prob_mayor_16 <- pnorm(16, mean = mu, sd = error_estandar, lower.tail = FALSE)
# lower.tail hace que pnorm calcule la probabilidad de que sea mayor a la variable en vez de que sea ig

cat("Probabilidad de media > 16 onzas: ", prob_mayor_16, "\n")
```

```
## Probabilidad de media > 16 onzas: 0.0007827011
```

inciso 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?


```
# Calibración con una media de 16 onzas
calibrar_16 <- (16 < limites[1]) | (16 > limites[2])
cat("Necesidad de calibrar con media = 16 onzas: ", calibrar_16, "\n")
```

```
## Necesidad de calibrar con media = 16 onzas: TRUE
```

inciso 4

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

```
#Probabilidad de que la media de la muestra sea menor que 14.5
prob_menor_14_5 <- pnorm(14.5, mean = mu, sd = error_estandar)
cat("Probabilidad de media < 14.5 onzas: ", prob_menor_14_5, "\n")
```

```
## Probabilidad de media < 14.5 onzas: 0.05692315
```

inciso 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?

```
# Calibración con una media de 15.5 onzas
#la siguiente línea compara los límites inferiores y superiores de x cuando se sale del 95% central, si
# entonces aparecerá como true, en caso contrario false
calibrar_15_5 <- (15.5 < limites[1]) | (15.5 > limites[2])
cat("Necesidad de calibrar con media = 15.5 onzas: ", calibrar_15_5, "\n")
```

```
## Necesidad de calibrar con media = 15.5 onzas: FALSE
```

No hay necesidad de detener la producción para calibrar la máquina ya que entra dentro del rango aceptable

inciso 6

Hacer una gráfica del inciso 1.

```
# 6. Gráfica del intervalo de confianza del 95% para la media muestral
curve(dnorm(x, mean = mu, sd = error_estandar), from = mu - 4 * error_estandar, to = mu + 4 * error_estandar)
abline(v = limites, col = "red", lty = 2, lwd = 2)
text(x = limites, y = 0.1, labels = round(limites, 2), col = "red", pos = 3)
```

