

Teorema del límite central a partir de medias muestrales con R

CARLOS GONZALEZ DE LA TORRE*

Universidad de Guadalajara

carlosglezleco@gmail.com

January 28, 2024

Abstract

En este documento se pretende demostrar el teorema central del límite con las herramientas que nos brinda el software estadístico R, mediante la carga de una distribución no normal, obtención de medias muestrales de esta distribución con diferentes tamaños de muestra y las gráficas correspondientes a estas distribuciones muestrales.

I. INTRODUCCIÓN

El teorema central del límite señala que conforme crece el número de elementos de una suma de variables aleatorias independientes, la distribución de la variable resultante de la suma se aproxima a una distribución normal, lo anterior sin importar la distribución que sigan las variables que se sumaron. Esto tam-

bien se aplica a medias muestrales, pues una media es una suma de variables donde cada una está multiplicada por una constante $\frac{1}{n}$ donde n es el tamaño de la muestra.

II. CARGANDO DATOS

Se utilizará el dataset diamonds del paquete ggplot2 cargando la librería tidyverse. La variable que utilizaremos será el precio de los diamantes.

*A thank you or further information

La figura 1 (p. 3) Se produjo con el siguiente código

```
> library(tidyverse)
> ggplot(diamonds, aes(price)) +
+   geom_histogram() +
+   xlab("Precio de los diamantes") +
+   ylab("") +
+   theme(panel.background =
+         element_rect(fill = "white"))
```

También se genera una tabla con estadísticas descriptivas de los datos que nos ayudan a reconocer la forma de la distribución:

	x
Min.	326.00
1st Qu.	950.00
Median	2401.00
Mean	3932.80
3rd Qu.	5324.25
Max.	18823.00

Table 1

Como se puede observar, la distribución tiene un gran sesgo hacia la derecha, en definitiva no muestra la forma de una distribución normal.

Para mayor facilidad en el manejo de los

datos, se creará un dataframe solamente con los precios de los diamantes excluyendo las demás variables del dataset diamonds, se escribe el código a continuación:

```
<<Precio>>=
Precio <- vector("numeric")
Precio <- diamonds$price
Precio <- as.data.frame(Precio)
```

III. REMUESTREO

A continuación se realizará una función cuyo objetivo será obtener un vector con los valores de las medias muestrales.

```
> VMM1 <- function(x, n, m){
+   vmm <- vector("numeric")
+   for (i in 1:m) {
+     muestra <- sample(x,n)
+     mm <- mean(muestra)
+     vmm[i] <- mm
+   }
+   return(vmm)
+ }
```

En la función, x es un vector del cual se obtendrán las muestras, n es el tamaño de la

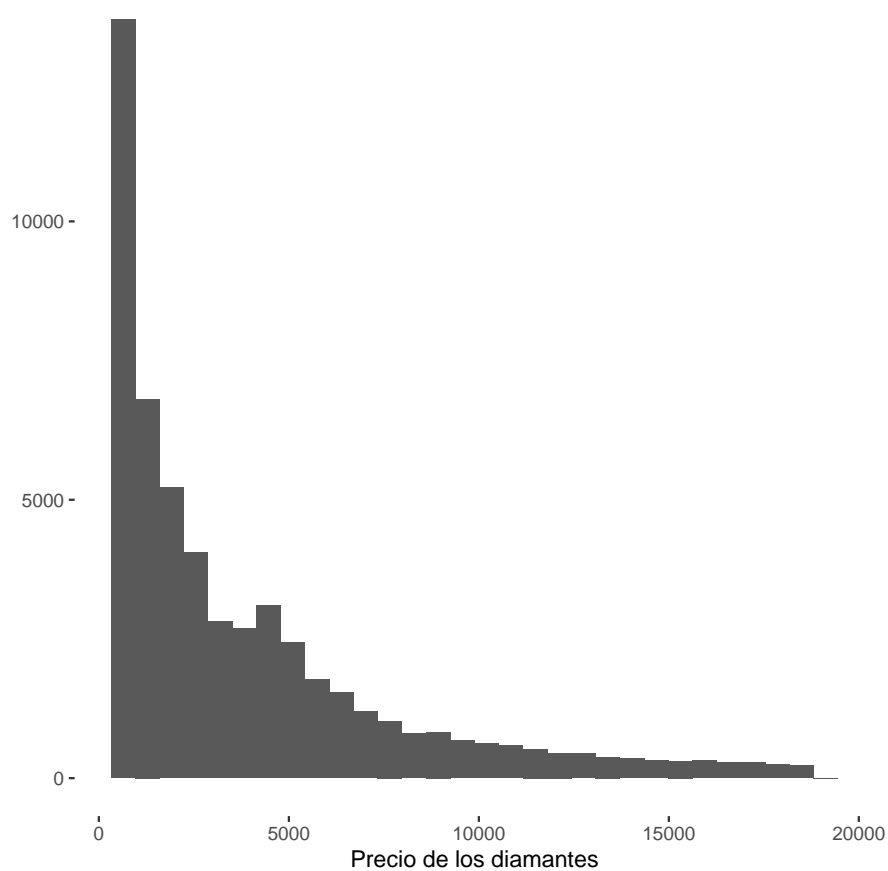


Figure 1: Distribución del precio de los diamantes

muestra y m es el número de remuestreos que se desea. Si utilizamos esta función con diferentes tamaños de muestra, con un número de remuestreo fijo y graficamos el vector de medias muestrales que se obtiene, se podrá ver cómo a medida que aumenta el tamaño de la muestra la distribución gráfica se aproxima a una normal.

```
> VP <- Precio$Precio  
> n5 <- VMM1(VP,5, 500)  
  
> VP <- Precio$Precio  
> n50 <- VMM1(VP,50, 500)  
  
> VP <- Precio$Precio  
> n500 <- VMM1(VP,500, 500)
```

Luego se grafican los resultados:

La figura 2 (p. 5) Se produjo con el siguiente código:

```
> hist(n5, xlab = "", ylab = "Frecuencia")
```

La figura 3 (p. 6) Se produjo con el siguiente código:

```
> hist(n50, xlab = "", ylab = "Frecuencia")
```

La figura 4 (p. 7) Se produjo con el siguiente código:

```
> hist(n500, xlab = "", ylab = "Frecuencia")
```

IV. CONCLUSIONES

Como se puede observar, al menos de forma gráfica, se comprueba el teorema central del límite en la distribución de medias muestrales con diferentes tamaños de muestra, las herramientas que nos ofrece el software R fueron de gran utilidad dentro de este trabajo.

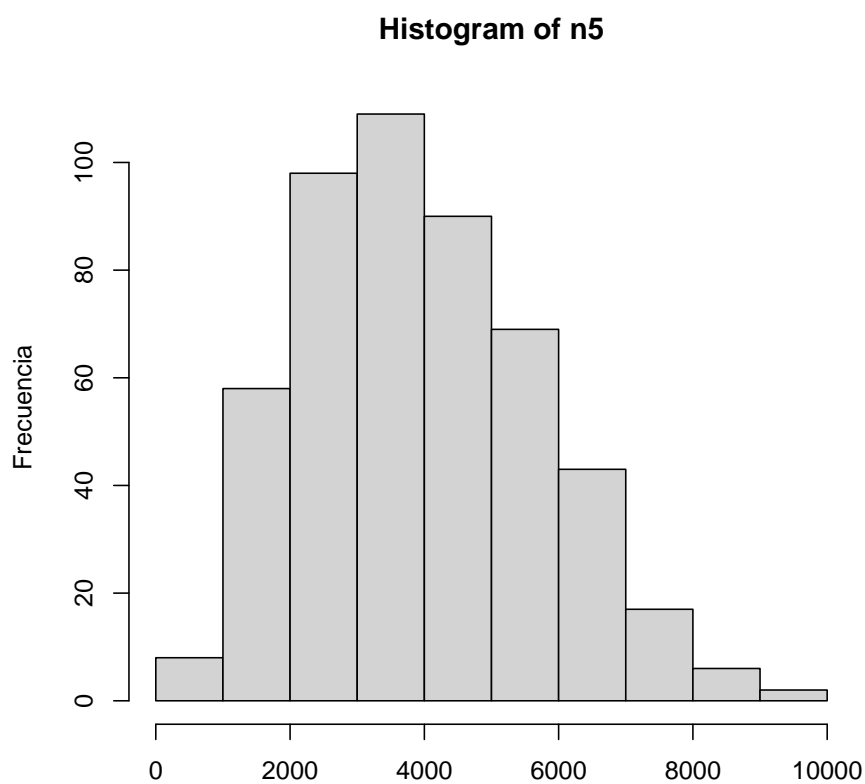


Figure 2: Distribución de la media muestral del precio de los diamantes, $n=5$

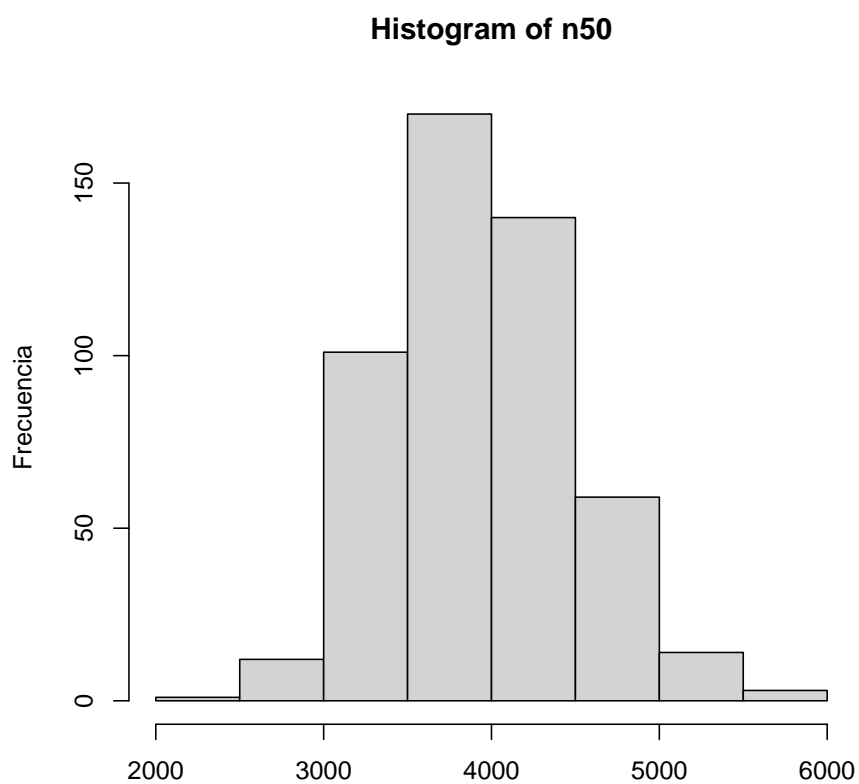


Figure 3: Distribución de la media muestral del precio de los diamantes, $n=50$

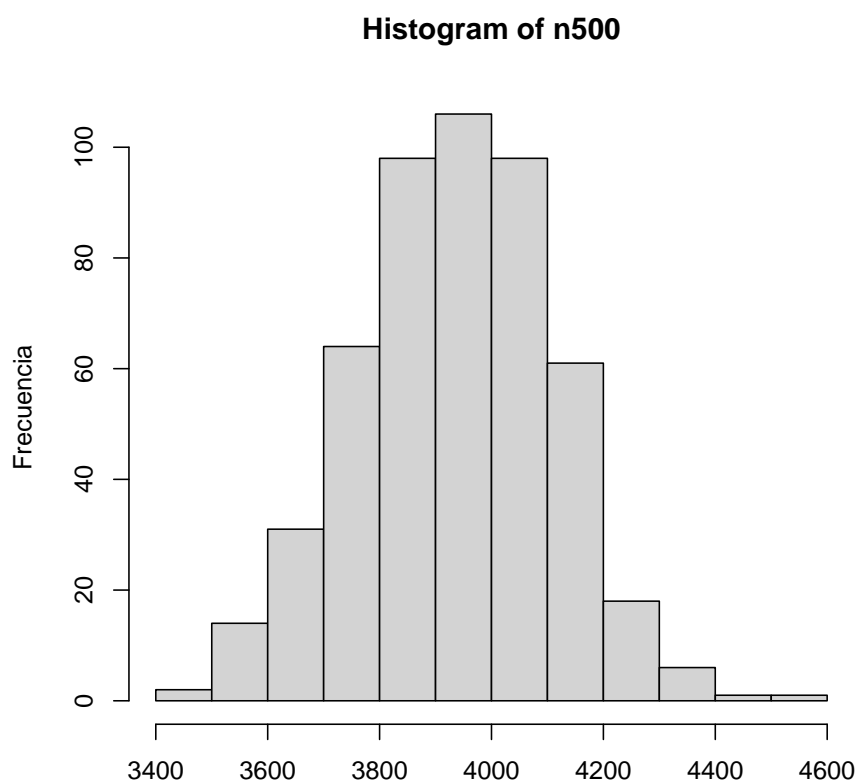


Figure 4: Distribución de la media muestral del precio de los diamantes, $n=500$