Trabajo Práctico 2 en `R`

Teorema Central del Límite

Actividad

Distribución Exponencial

1. **Distribución de una exponencial**

a) $\mathbb{E}(X) = \frac{1}{\lambda}$, $\mathbb{V}(X) = \frac{1}{\lambda^2}$ En este caso, $\mathbb{E}(X) = \frac{1}{12}$, $\mathbb{V}(X) = \frac{1}{144}$

b)
```r
Nrep <- 10000
lambda <- 1/12
exp_N_infty <- rexp(Nrep, rate = lambda)
```

c)
```r
mean(exp_N_infty)
```

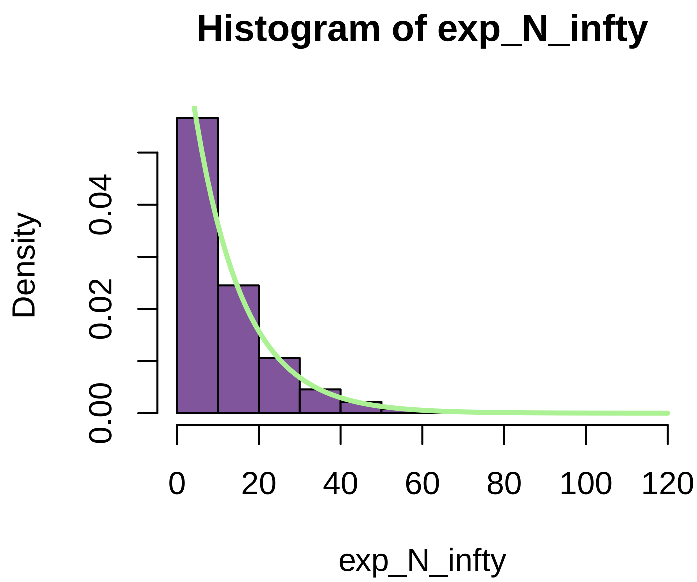
```
## [1] 11.99286
```

```r
var(exp_N_infty)
```

```
## [1] 144.1552
```

Las medidas muestrales se acercan relativamente bastante a los valores correspondientes a su distribución.

d) y e)



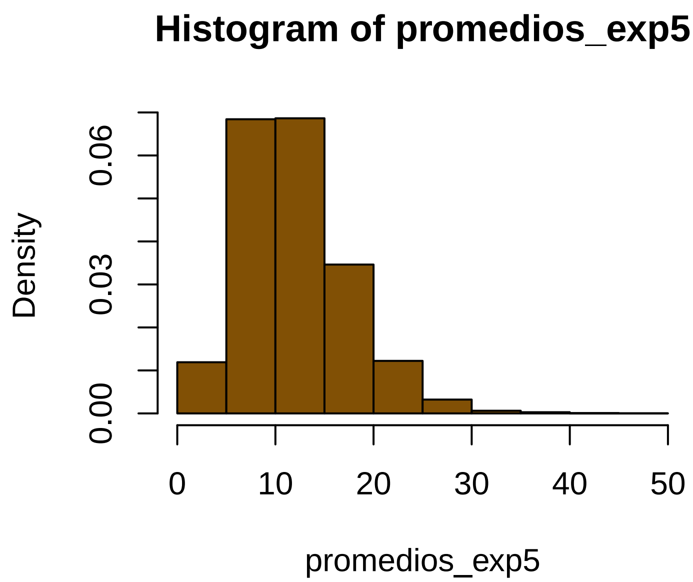
Como puede observarse en el histograma, la distribución se asemeja a la forma de la función de densidad de una exponencial. Esto a su vez se ve corroborado por la similitud con el gráfico de la función de densidad de una exponencial con parámetro $\frac{1}{12}$.

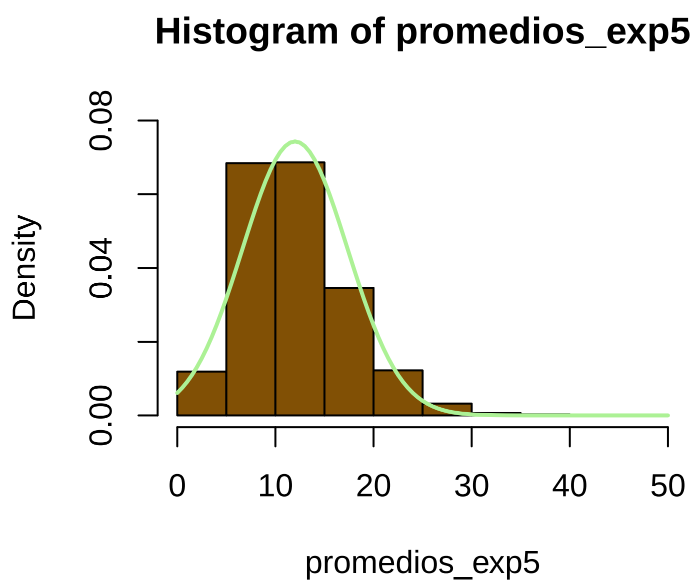
2. **Distribución empírica del promedio de exponenciales**

a)
```
Nrep <- 10000
lambda <- 1/12

promedios_exp5 = c()
for(i in 1:10000){
  promedios_exp5 <- append(promedios_exp5, mean(rexp(5, rate=lambda)))
}
```

b) Histograma de densidad:





c)

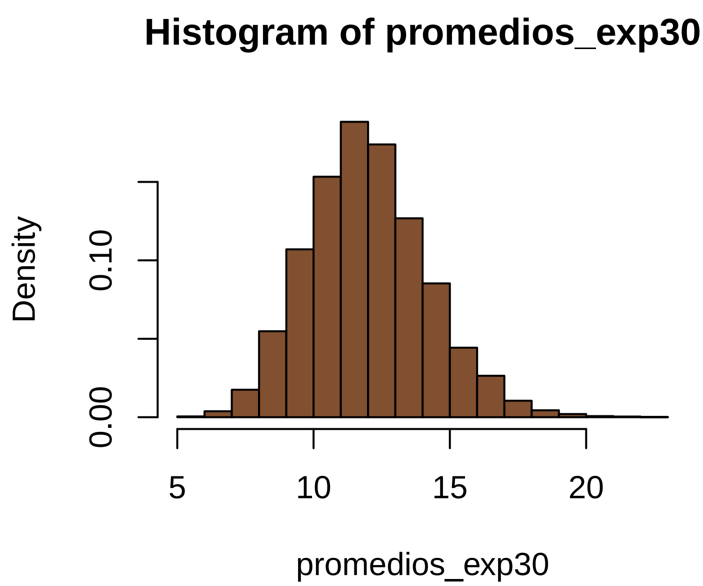
A partir del gráfico, es notorio que el la distribución de los promedios no es normal. Esto se debe probablemente a la asimetría que tiene la distribución exponencial original, por lo que los promedios de exponenciales también se verán "corridos" cuando n es pequeña.

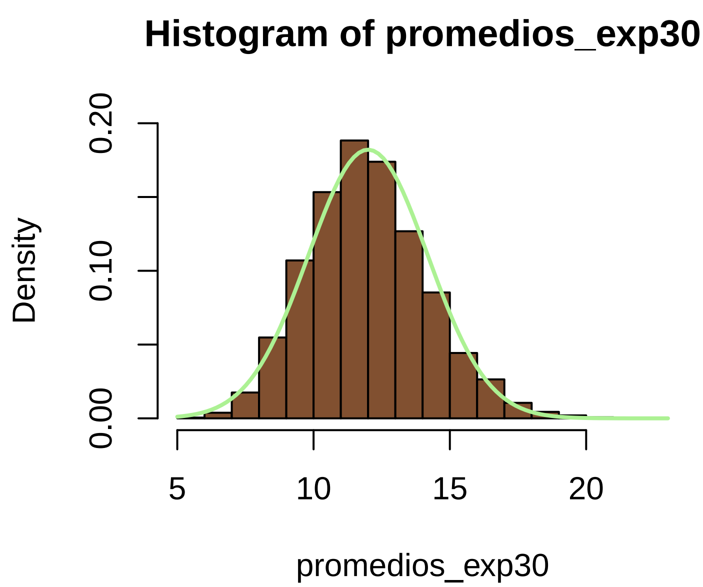
d) **Para $n = 30$:**

```
Nrep <- 10000
lambda <- 1/12

promedios_exp30 = c()
for(i in 1:10000){
  promedios_exp30 <- append(promedios_exp30, mean(rexp(30, rate=lambda)))
}
```

Histograma de densidad:





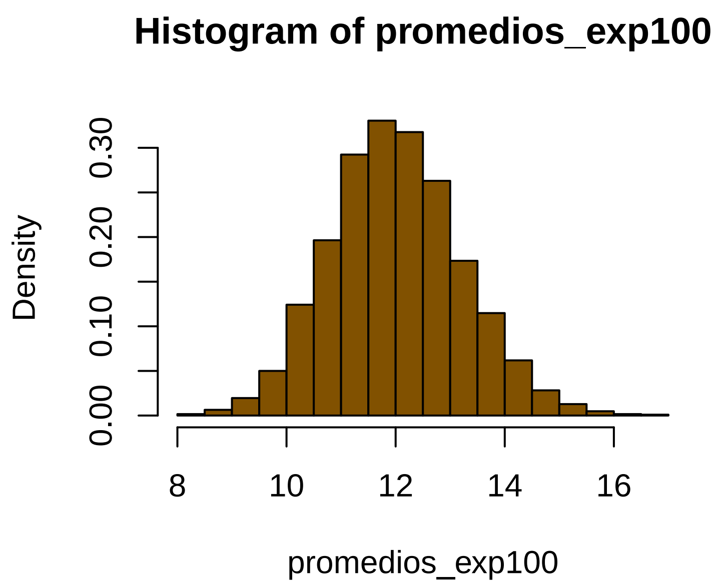
A partir del gráfico con $n = 30$, es más difícil determinar si es un tamaño de muestra razonable. Por un lado, la forma es acampanada y está relativamente centrada, pero al prestar atención se nota algo corrido hacia la izquierda. Aún así, no creemos que esté errado decir que la distribución de los promedios es aproximadamente normal con este tamaño de muestra.

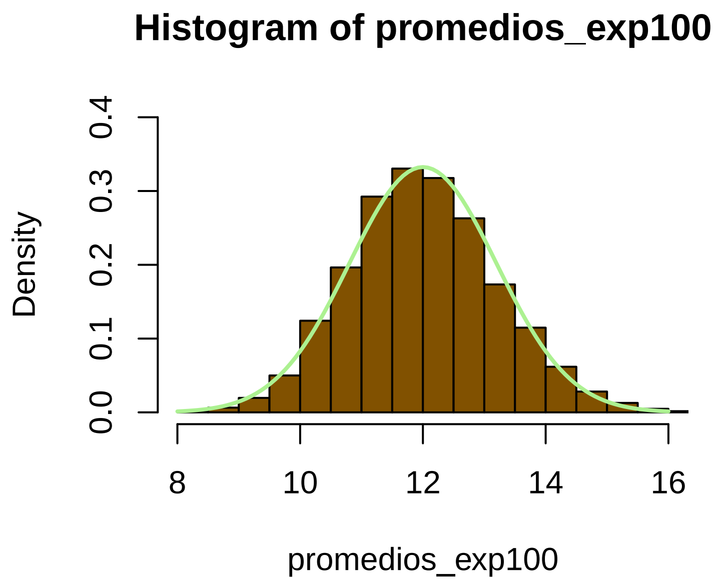
Para $n = 100$**:**

```
Nrep <- 10000
lambda <- 1/12

promedios_exp100 = c()
for(i in 1:10000){
  promedios_exp100 <- append(promedios_exp100, mean(rexp(100, rate=lambda)))
}
```

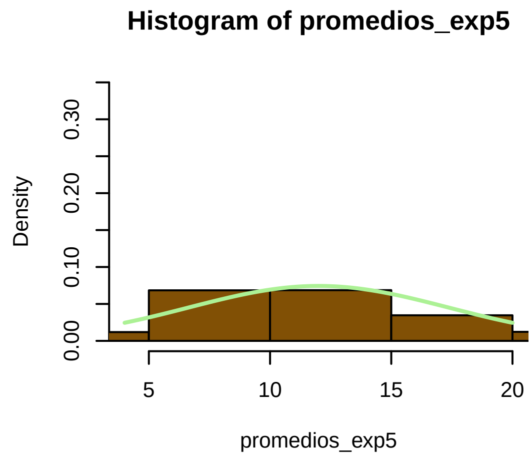
Histograma de densidad:

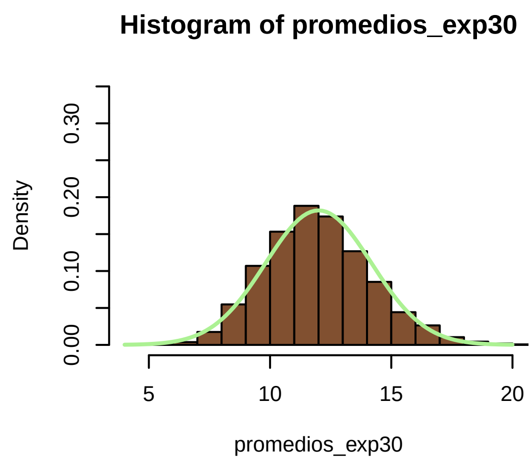


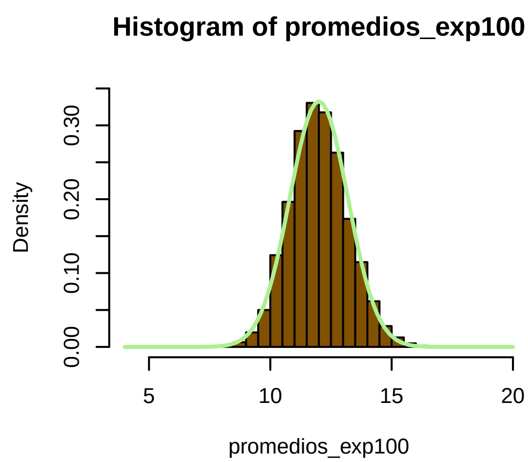


Para el caso de $n = 100$, la forma de la distribución de los promedios se parece mucho más a la curva normal, y está claramente centrada en 12.

e) Histogramas juntos:







La aproximación de los promedios se va ajustando cada vez más a la forma de la normal correspondiente, tal como indica el TCL. La dispersión también se reduce a medida que n aumenta, dado que la varianza de la normal aproximada a partir del promedio de n variables i.i.d. (c/u con distribución F con varianza σ^2) es σ^2/n, de acuerdo con el TCL.

Distribución Uniforme

1)

1. **Distribución de una uniforme**

a) $\mathbb{E}(X) = \frac{a+b}{2}$, $\mathbb{V}(X) = \frac{(b-a)^2}{12}$ En este caso, $\mathbb{E}(X) = 12$, $\mathbb{V}(X) = \frac{4}{3} \approx 1.3$

b)
```
Nrep <- 10000
unif_N_infty <- runif(Nrep, min = 10, max = 14)
```

c)
```
mean(unif_N_infty)
```

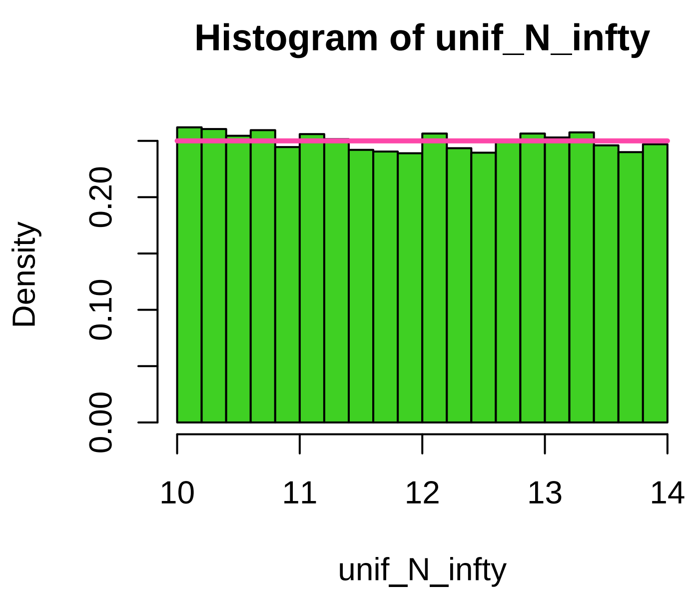
```
## [1] 11.98633
```

```
var(unif_N_infty)
```

```
## [1] 1.346681
```

Nuevamente, la media y varianza muestrales se acercan bastante a los valores correspondientes a su distribución.

d) y e)



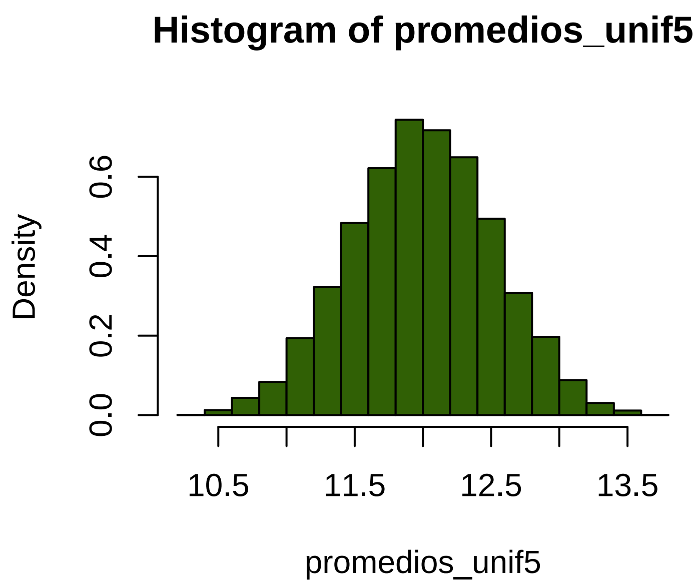
Como puede observarse en el histograma, la distribución se asemeja a la forma de la función de densidad de una uniforme con rango entre 10 y 14, presentando a su vez algunas irregularidades.

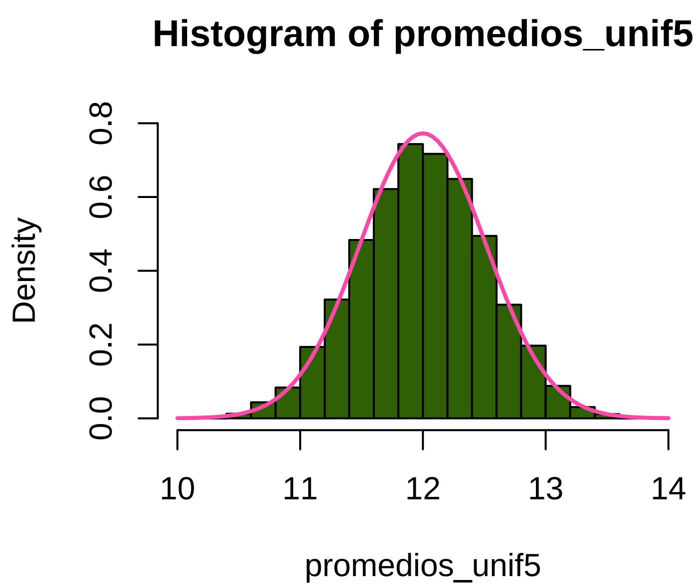
2. **Distribución empírica del promedio de uniformes**

a)
```
Nrep <- 10000

promedios_unif5 = c()
for(i in 1:10000){
  promedios_unif5 <- append(promedios_unif5, mean(runif(5, min=10, max=14)))
}
```

b) Histograma de densidad:





c)

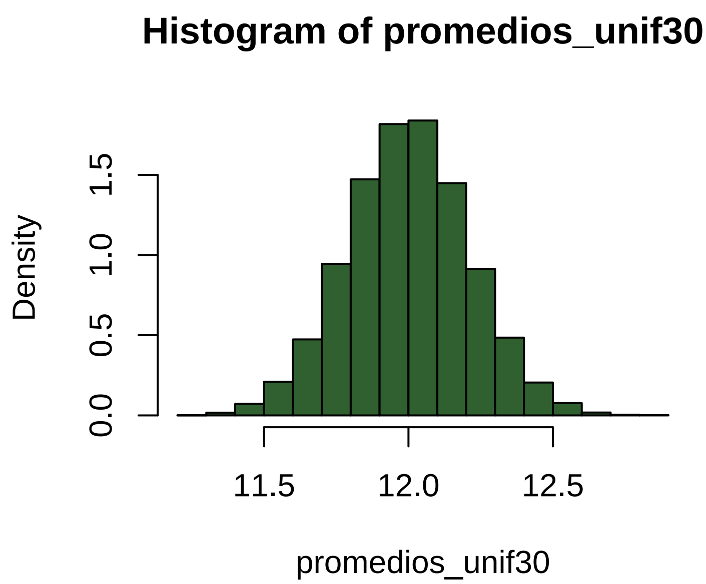
A partir del gráfico, la distribución de los promedios parecería tener forma normal. Esto, a diferencia de la distribución exponencial, se debe a que la distribución de los datos es simétrica. Por ende, en este caso $n = 5$ es un tamaño de muestra razonable siendo los datos de distribución uniforme.

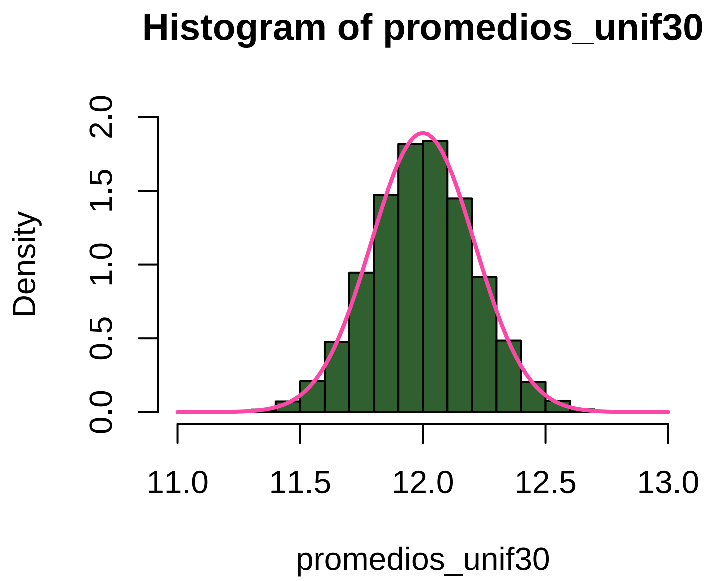
d) **Para** $n = 30$**:**

```
Nrep <- 10000

promedios_unif30 = c()
for(i in 1:10000){
  promedios_unif30 <- append(promedios_unif30, mean(runif(30, min=10, max=14)))
}
```

Histograma de densidad:





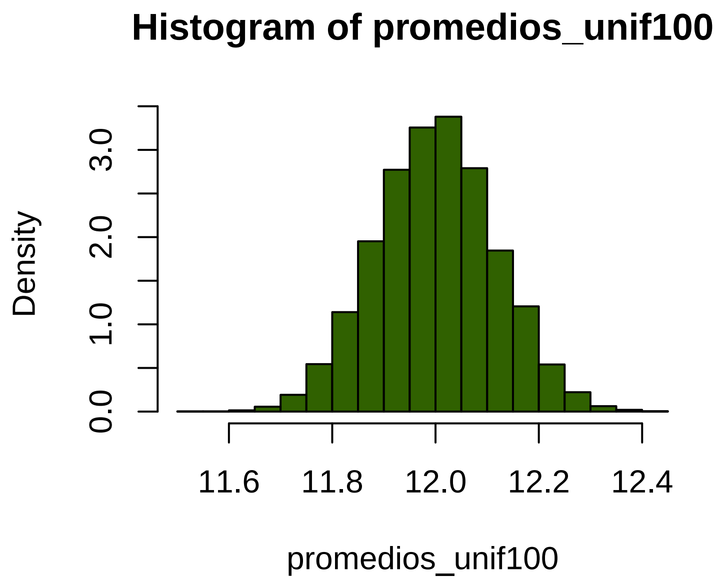
Si con 5 datos ya era suficiente para aproximar los promedios a una normal, con 30 es más similar, y sumado a esto la varianza disminuye y los datos se concentran en la media.

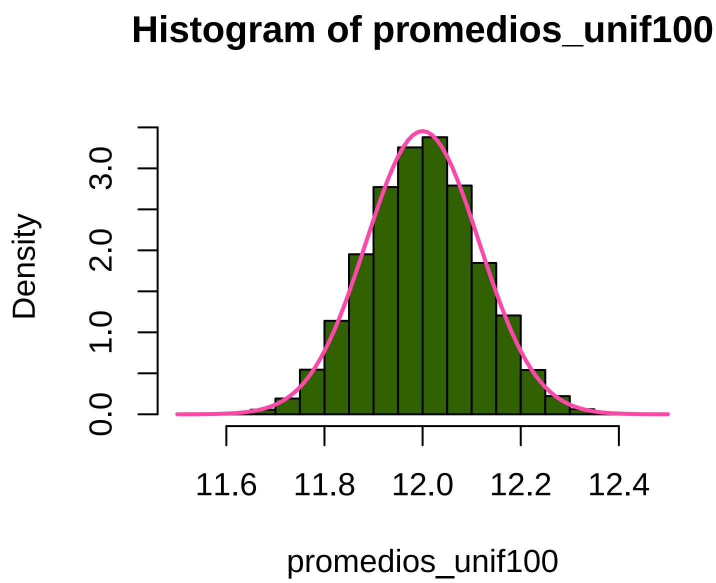
Para $n = 100$**:**

```
Nrep <- 10000

promedios_unif100 = c()
for(i in 1:10000){
  promedios_unif100 <- append(promedios_unif100, mean(runif(100, min=10, max=14)))
}
```

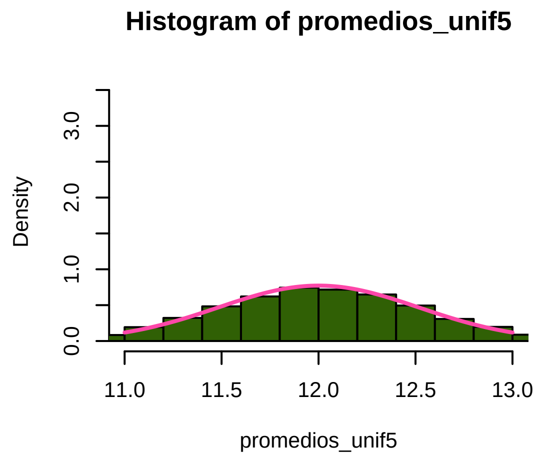
Histograma de densidad:

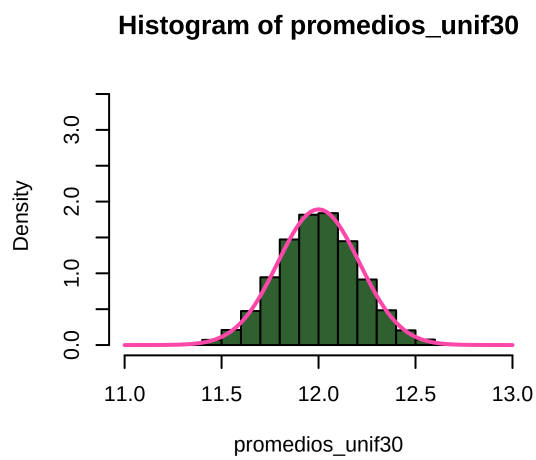


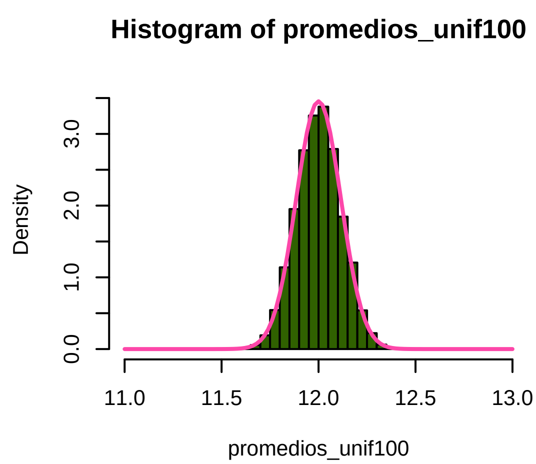


Para el caso de $n = 100$, notar que la normal se concentra más en la media.

Histogramas juntos:







2) La n a partir de la cual la aproximación es buena es menor, siendo en este caso $n = 5$. Esto, nuevamente, se debe probablemente a la simetría de la distribución uniforme, en contraste con la exponencial.

Distribución Normal

1)

1. **Distribución de una normal**

a) Si $X \sim \mathcal{N}(12, 4/3)$,
$\mathbb{E}(X) = 12$, $\mathbb{V}(X) = 4/3$

b)
```
Nrep <- 10000
norm_N_infty <- rnorm(Nrep, mean = 12, sd = sqrt(4/3))
```

c)
```
mean(norm_N_infty)
```

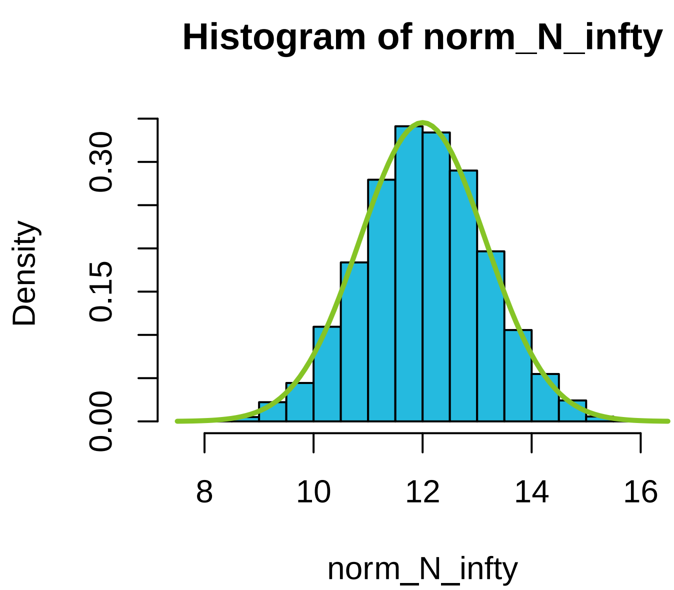
```
## [1] 12.02269
```

```
var(norm_N_infty)
```

```
## [1] 1.295245
```

Nuevamente, las medidas muestrales se acercan bastante a los valores correspondientes a su distribución.

d) y e)



Como puede observarse en el histograma, la distribución se asemeja a la forma de la función de densidad de una normal con media 12 y varianza 4/3.

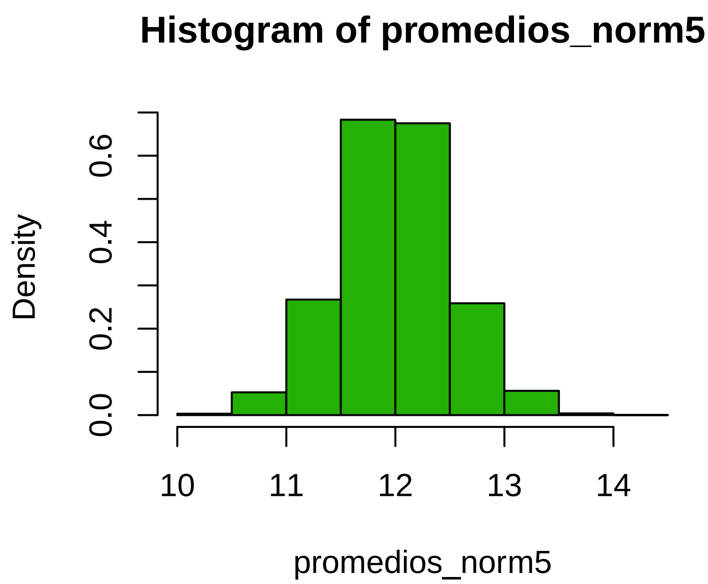
2. **Distribución empírica del promedio de normales**

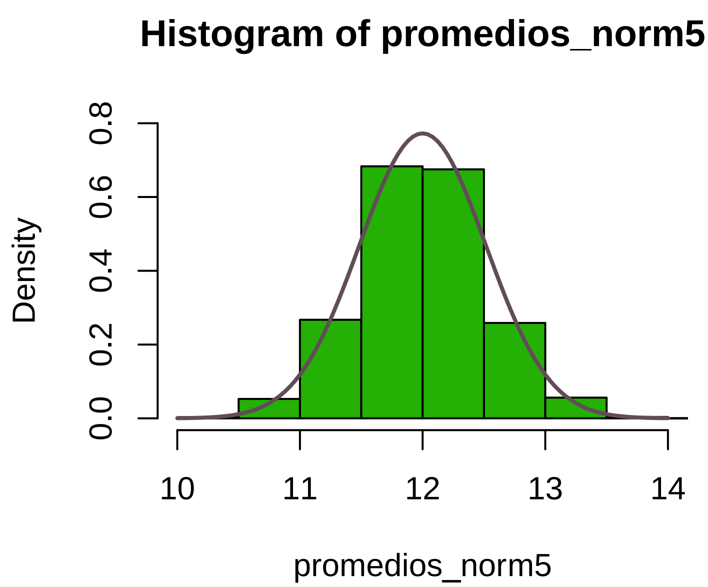
a)

```
Nrep <- 10000

promedios_norm5 = c()
for(i in 1:10000){
  promedios_norm5 <- append(promedios_norm5, mean(rnorm(5, mean=12, sd=sqrt(4/3))))
}
```

b) Histograma de densidad:





c)

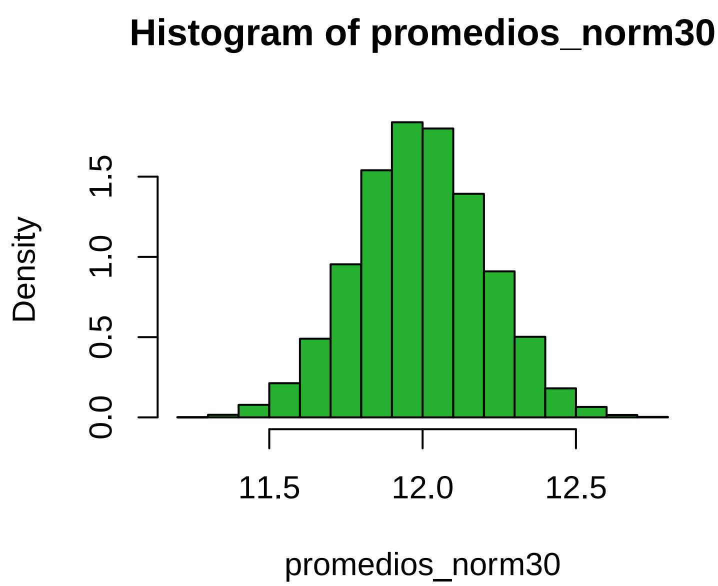
A partir del gráfico, la distribución de los promedios parecería tener forma normal. En este caso, esto se debe a que los datos originales tienen distribución normal.

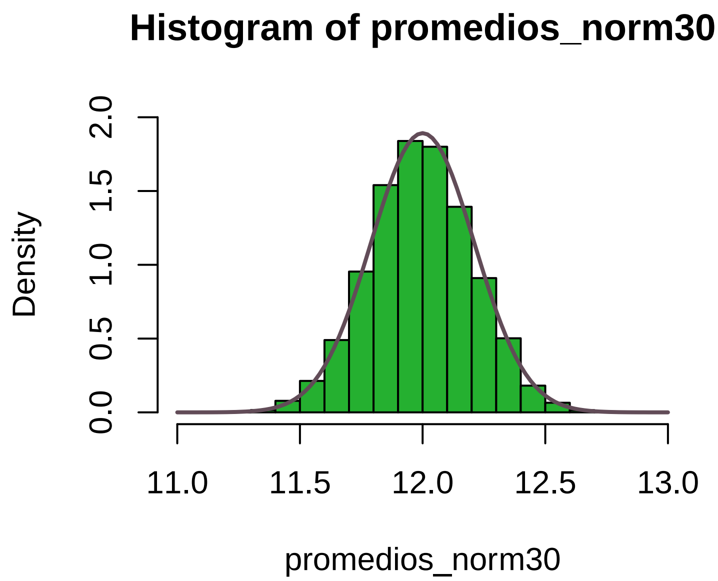
d) **Para** $n = 30$**:**

```
Nrep <- 10000

promedios_norm30 = c()
for(i in 1:10000){
  promedios_norm30 <- append(promedios_norm30, mean(rnorm(30, mean=12, sd=sqrt(4/3))))
}
```

Histograma de densidad:





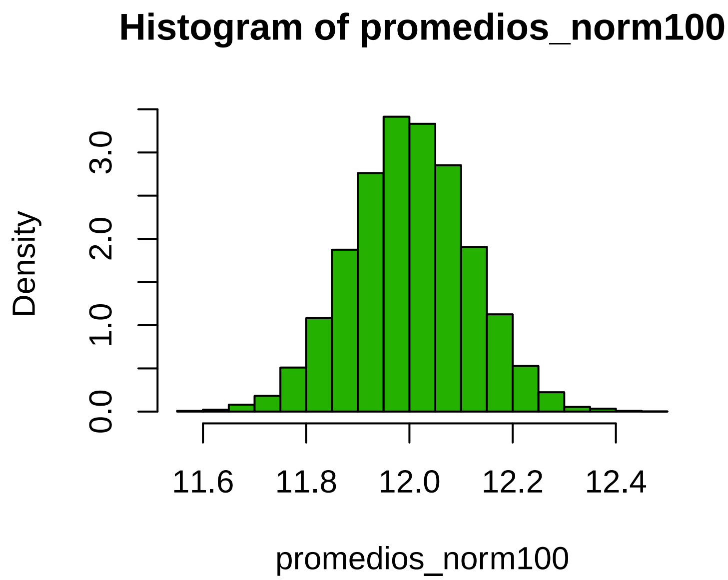
La distribución de los promedios ahora se concentra más alrededor de la media.

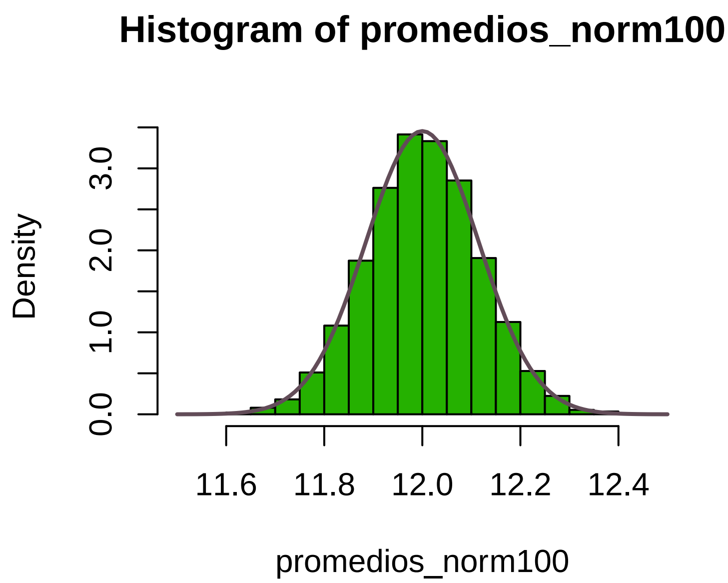
Para $n = 100$**:**

```
Nrep <- 10000

promedios_norm100 = c()
for(i in 1:10000){
  promedios_norm100 <- append(promedios_norm100, mean(rnorm(100, mean=12, sd=sqrt(4/3))))
}
```

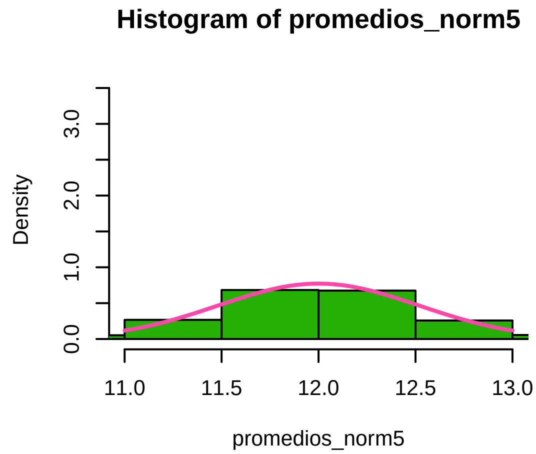
Histograma de densidad:

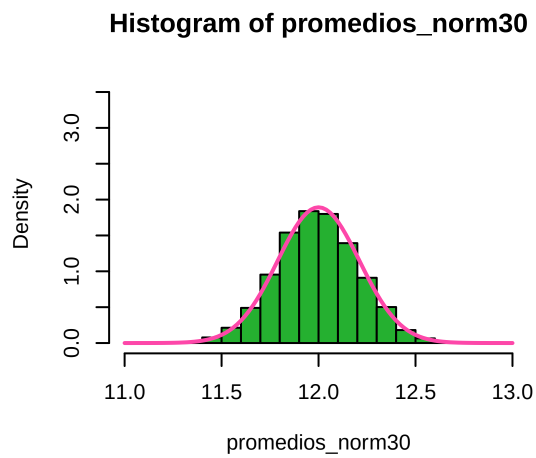


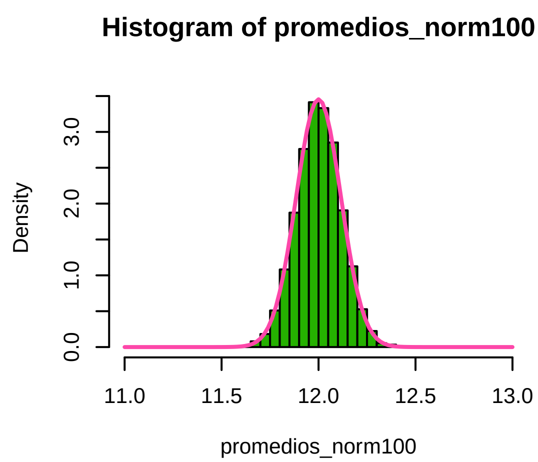


Como con $n = 30$, ahora aún más se concentran los datos alrededor de la media.

Histogramas juntos:







2) En el caso de la distribución normal, como la media de normales es también una normal (no aproximación con el TCL), para cualquier n se puede decir que la distribución del promedio no se aproxima,

sino que *es* normal. En el caso de la distribución uniforme, vimos que gracias a su simetría podíamos aproximar el promedio a una normal con $n = 5$ datos por el TCL. Por último, en el caso de la exponencial recién a partir de $n = 30$ datos pudimos aproximar el promedio a una normal, dado que no era simétrica de base.