Ejercicio 3

Consideremos una población con ley gamma de parámetros $r = 5$ y $\lambda = \dfrac{1}{4}$.

(A)

Lo que se tiene que hacer es simular muestras aleatorias de una distribución gamma, en nuestro caso, Gamma$(5, \frac{1}{4})$ y tomar la media en cada iteración. La gamma se define mediante dos parámetros, forma y escala, en nuestro caso:

- parámetro de forma: 5
- parámetro de escala: $\dfrac{1}{4}$

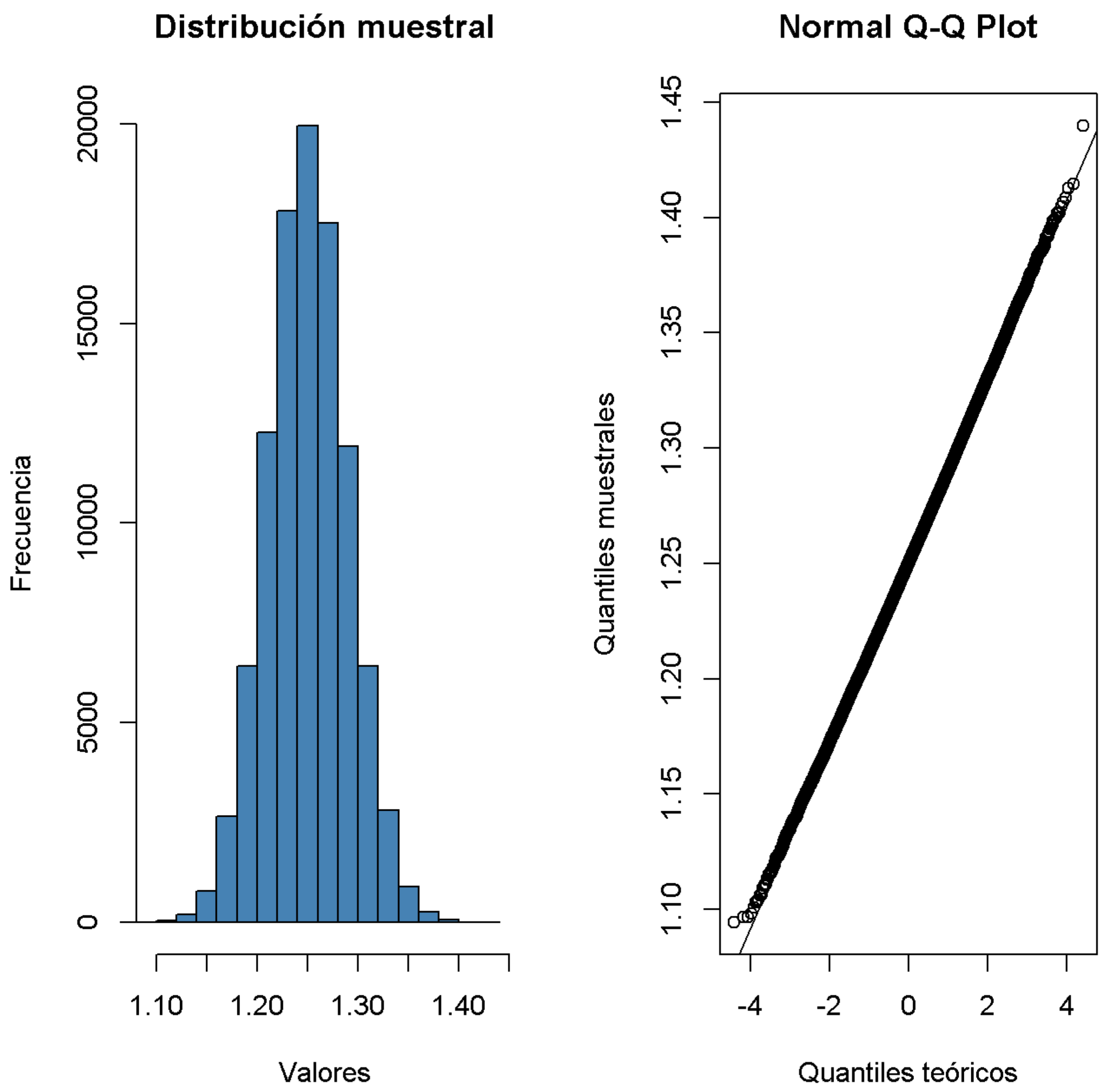
Calcularemos la media de una distribución gamma 10^5 veces con $n_i = 200$, donde n_i es el tamaño de cada muestra, $i = 1, \ldots, 10^5$.

Si $X_1, X_2, \ldots, X_n$ son muestras aleatorias de una ley gamma con forma a y escala b, la suma $T = X_1 + X_2 + \ldots + X_n$ es también una gamma con la misma escala pero con forma an. Por lo que, la distribución de la media muestral se aconsigue mediante la transformación: $(\bar{X}) = T/n$ y tiene densidad $f_{\bar{X}}(x) = nf_T(nx)$.

Esto sigue una gamma con forma an y escala b/n. Esto pasa ya que el valor esperado de una observación es $E(X_i) = ab$. La esperanza de su suma es $E(T) = abn$ y la esperanza de la media muestral es igual al valor esperado de una observación X_i:

$$E(\bar{X}) = (an)(b/n) = ab.$$

Para la varianza no pasa lo mismo ya que esta depende del tamaño de la muestra, digamos n. Para una observación, Var$(X_i) = ab^2$, pero Var$(\bar{X}) = (an)(b/n)^2 = \dfrac{ab^2}{n} \leq ab^2$, para $n \geq 1$.

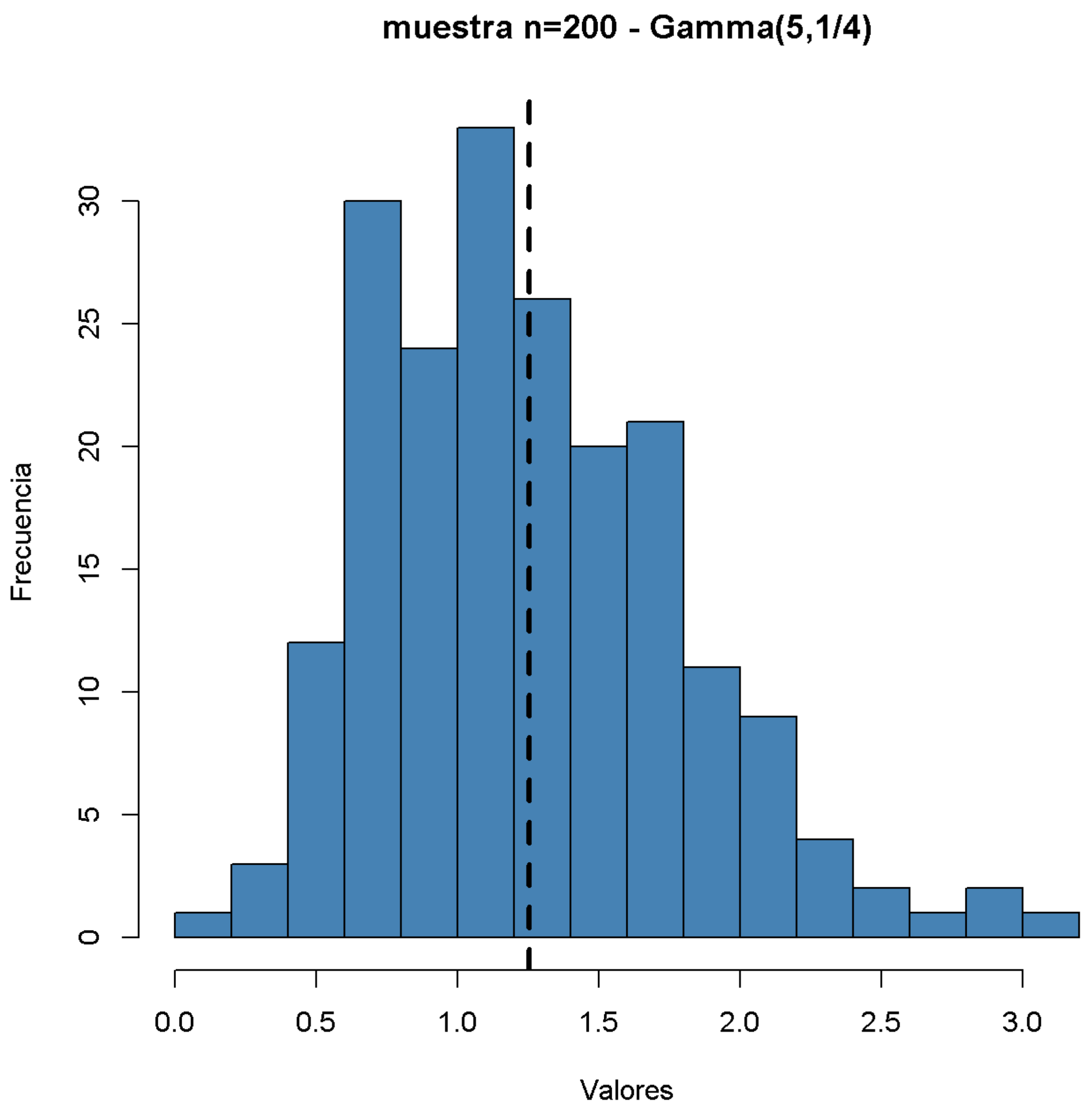


Análisis descriptivo distribución muestral

Min.	1.094
1st Qu.	1.223
Median	1.25
Mean	1.25
3rd Qu.	1.277
Max.	1.44
Var	0.0016

(B)

Generar una muestra aleatoria de tamaño 200 de la ley gamma. La línea vertical representa el valor de la media.

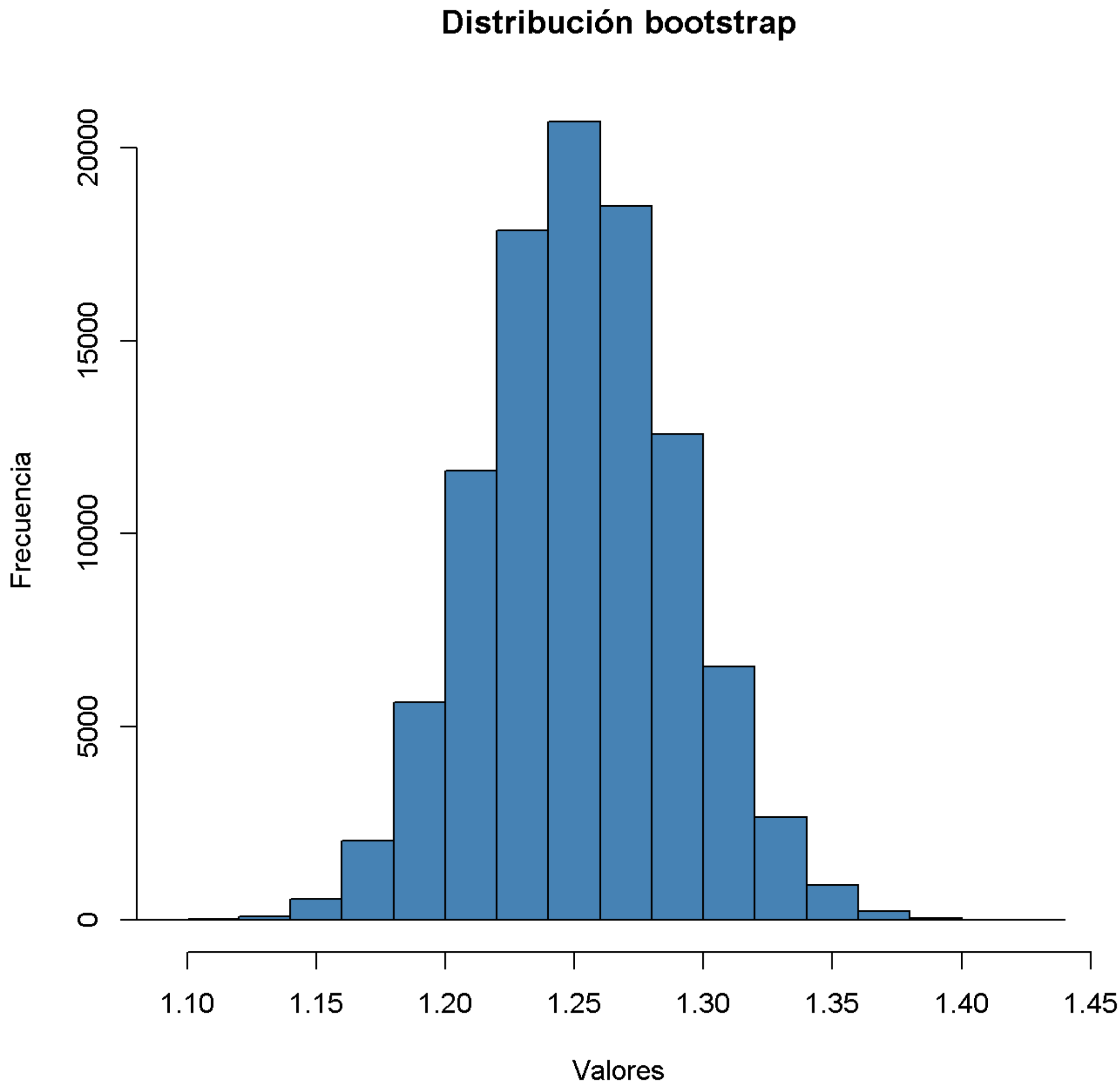


Mean	1.252
SD	0.540

(C)

```
Idea principal del bootstrap:
La muestra original aproxima a la población. Por lo que remuestras de esta muestra se aproximan
a lo que
obtendríamos si tomaramos muchas muestras de la población. La distribución bootstrap de un esta
dístico, basada en muchas remuestras, se aproxima a la distribución de muestreo del estadístic
o, basado en muchas muestras.
```

Lo que me parece más interesante del bootstrap es que teniendo una muestra y suponiendo que es 'representativa' podemos obtener resultados fiables sin necesidad de obtener nuevas muestras, ya que eso en el mundo real, puede ser costoso o imposible de obtener. En este caso, podemos obtener la distribución bootstrap de la media muestral generando m remuestras (muestras con reemplazo) de nuestra muestra de tamaño $n = 200$ y en cada m_i calcular el estadístico muestral, con $i = 1, \dots, 10^5$.



Bootstrap mean1.252
Standard error 0.038

(D)

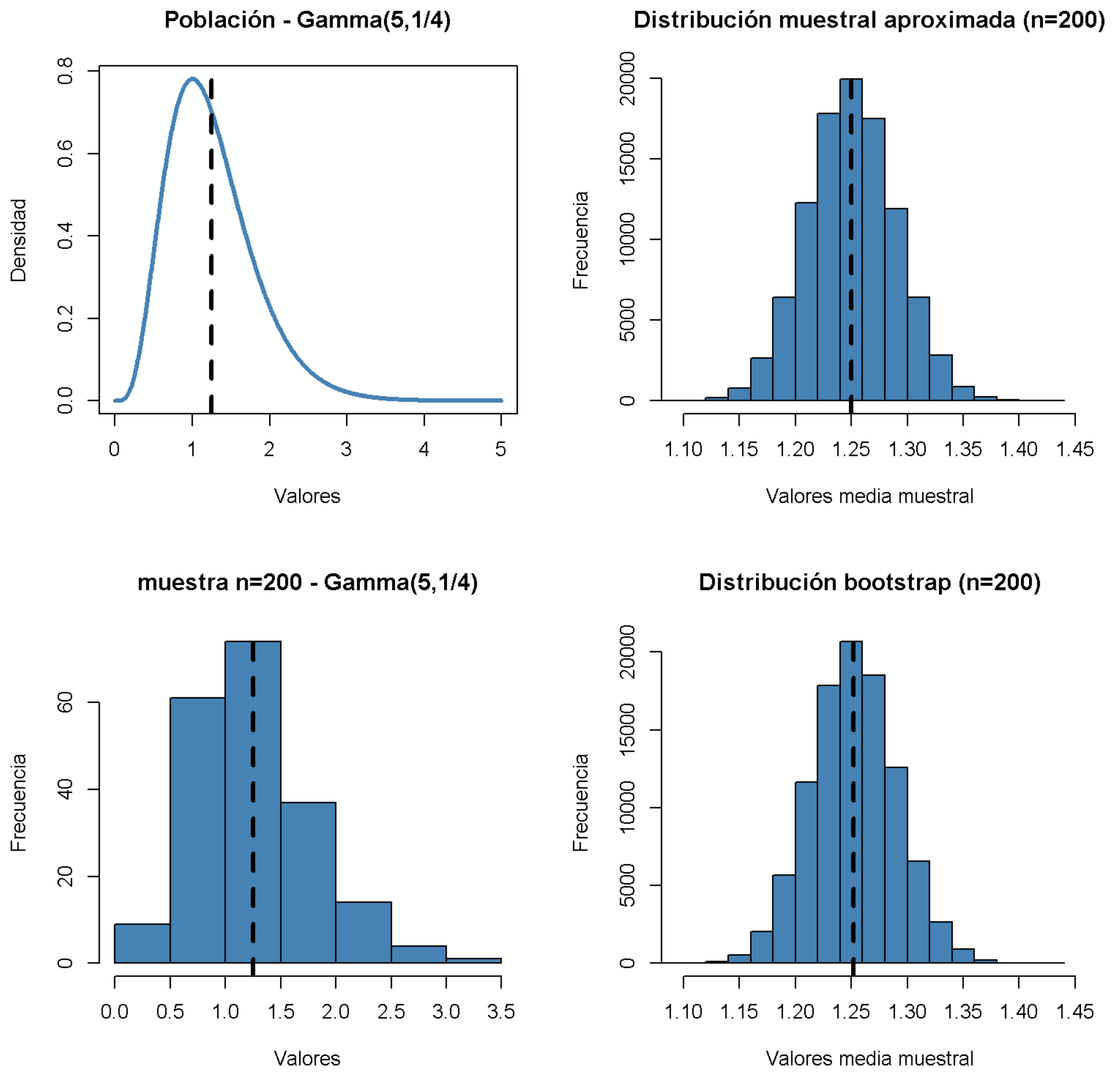
Ahora podemos comparar el resultado obtenido mediante bootstrap con la distribución muestral aproximada teórica (con $n = 200$). Vemos que la distribución bootstrap es muy parecida a la distribución muestral aproximada. Se refleja que la esperanza es igual a la ley gamma poblacional, pero que la desviación decrece en función del tamaño de la muestra. Este resultado también demuestra que con remuestras de una sola muestra (*muestras permutadas con reemplazamiento*) obtenemos resultados muy iguales a los obtenidos con 10^5 muestras.

	Mean	SD
Población	1.2500	0.5590
Distribución muestral aproximada de la media (n=200)	1.2502	0.0395
Muestra n=200	1.2522	0.5397

	Mean	SD
Distribución bootstrap (remuestras con n=200)	1.2521	0.0380

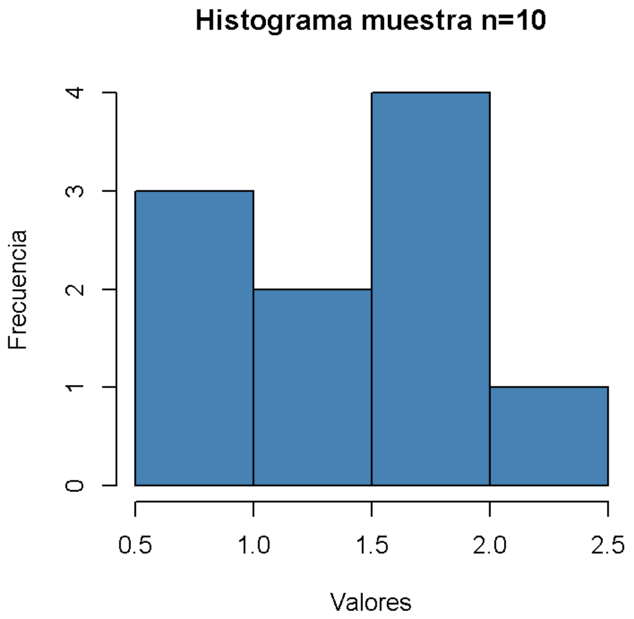
Comparación mediante gráficos

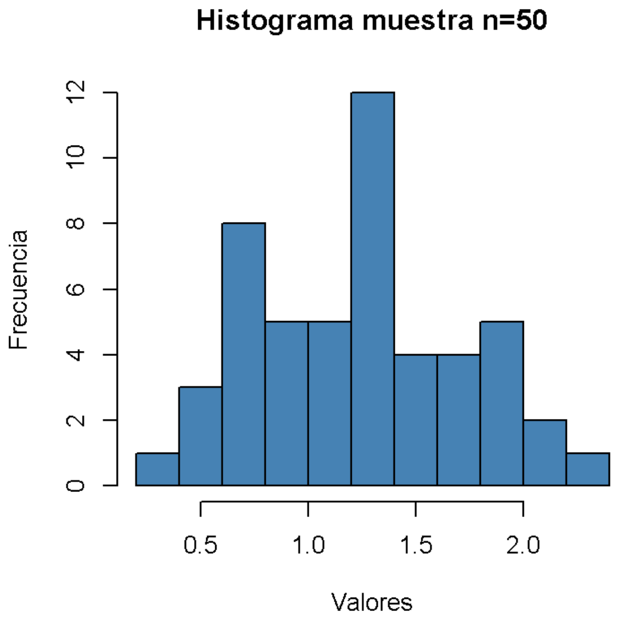
Las líneas verticales representan la media.

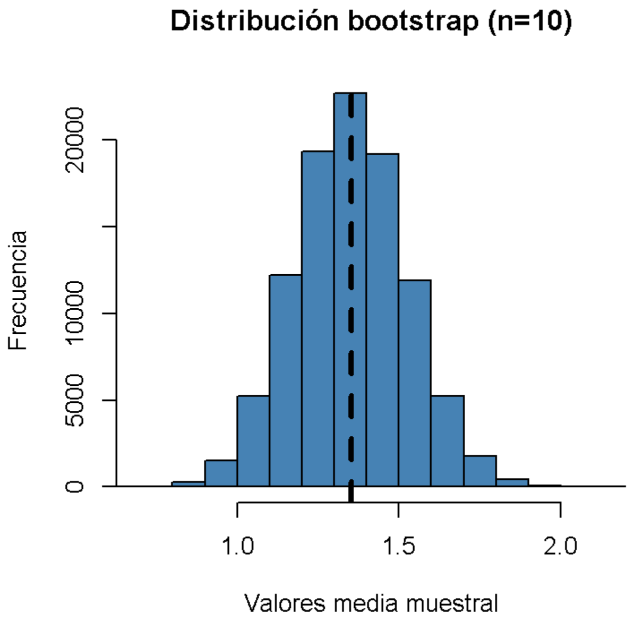


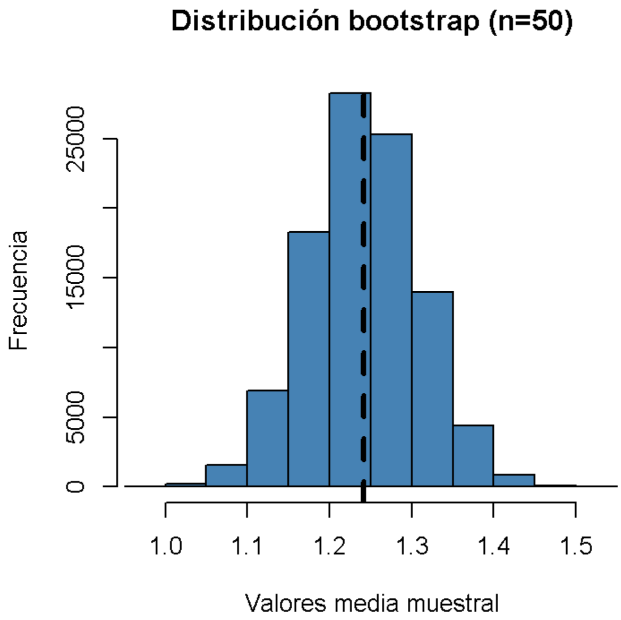
(E)

El problema del bootstrap, es que si cogemos datos no representativos de la población que queremos **dibujar** (por ejemplo, en este caso, una muestra pequeña), los resultados pueden aproximar el estadístico muestral con sesgo. Para $n = 10$, la estimación es sesgada, por lo que los errores son más grandes, error que hará los intervalos bootstrap menos precisos y fiables. La línea vertical representa la media muestral.










```
--------boot10--------
```

```
Mean = 1.351181
```

```
Intervalo para la media muestral
```

```
    2.5%    97.5%
1.020112 1.693709
```

```
--------boot50--------
```

```
Mean = 1.24153
```

```
Intervalo para la media muestral
```

```
    2.5%    97.5%
1.109372 1.374697
```

```
--------boot200--------
```

```
Mean = 1.252087
```

```
Intervalo para la media muestral
```

```
    2.5%    97.5%
1.178679 1.327445
```