

Problema 1. El desarrollo de esta actividad está basado en el problema 1 del tp 1 que trata sobre los talaes bonaerenses: Los talaes bonaerenses constituyen el principal bosque nativo de la provincia de Buenos Aires. Sin embargo, se encuentran en continua reducción. Su superficie actual es mucho menor que la original, que se extendía desde el norte de la provincia de Buenos Aires, en una estrecha franja costera que llegaba hasta Mar del Plata. Esta es la zona más poblada de Argentina y, donde las ciudades crecieron, los talaes desaparecieron casi totalmente. Para poder tomar medidas adecuadas de manejo para evitar su desaparición total, es necesario tener una estimación del estado actual de los remanentes de talar que todavía perduran.

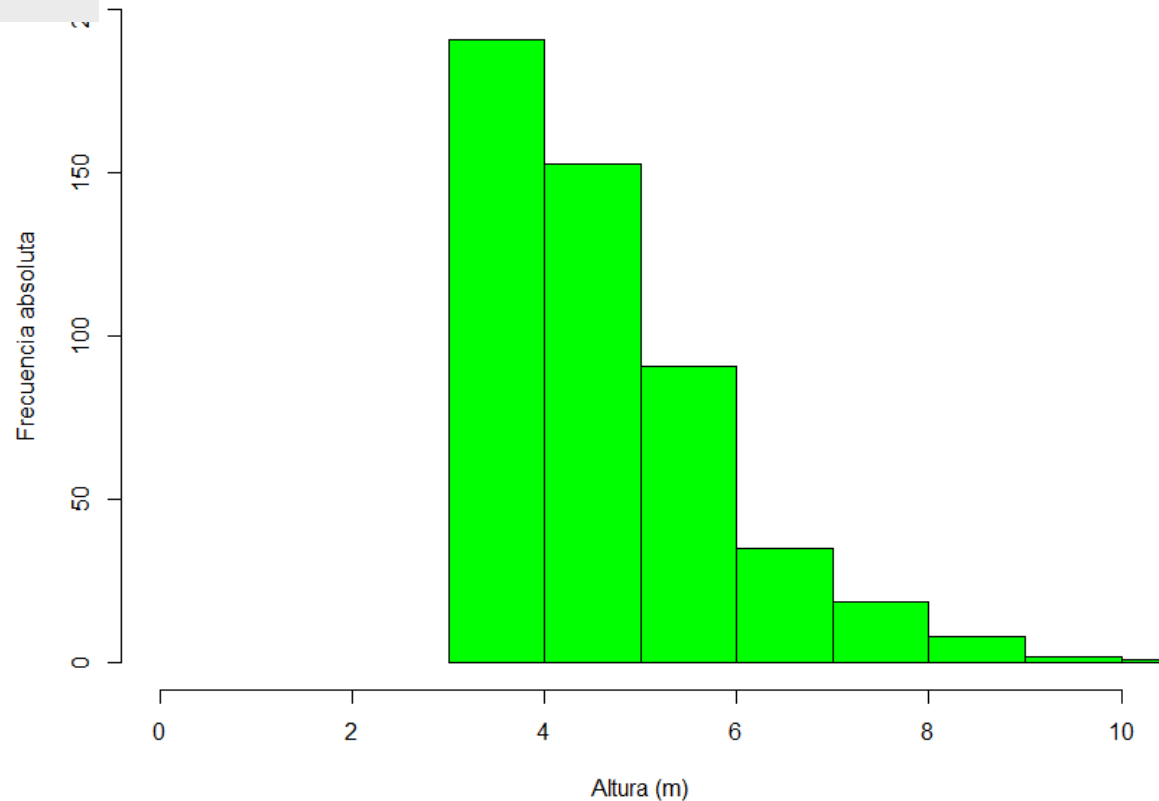
La base de datos **poblacion_talas.txt** muestra las alturas de una población de talas compuesta por 500 árboles de un sector donde han sido extraídos gran parte de los individuos de mayor tamaño.

ACTIVIDAD A #####: .

ACTIVIDAD A #####: .

#¿Cómo es la simetría (forma) de la distribución de esta variable? Justificar gráficamente.

X = altura (m)
N= 500 talas (población)



#¿Cuál es la altura media de los talas de esta población y su dispersión? Esperanza y varianza

$$E(x) = 4.63 \text{ m}$$
$$\text{Var}(x) = 1.57 \text{ m}^2$$

```
varp<-function(x) mean((x-mean(x))^2)
```

#¿Qué nombre reciben estos valores en la población?

Parámetros

ACTIVIDAD B #####: .

1) Vamos a tomar muestras aleatorias de la variable altura que sean de distinto tamaño ($n=5$, $n=20$ y $n=50$), para lo cual usaremos el comando "sample" del paquete "base".

2) Calculen la media y la varianza de cada una de las muestras. Piensen si corresponde usar los mostrados en la actividad A o los que aprendieron en el TP1.

```
# media_muestra_n5 [1] 4.54 var_muestra_n5 [1] 0.98
```

```
# media_muestra_n20 [1] 4.25 var_muestra_n20 [1] 0.95
```

```
# media_muestra_n50 [1] 4.76 var_muestra_n50 [1] 1.54
```

3) Comparen sus resultados con los obtenidos por sus compañeros y discutan la siguiente frase: "Las medias y las varianzas son variables aleatorias". ¿Están de acuerdo? ¿Por qué?

ACTIVIDAD C #####: .

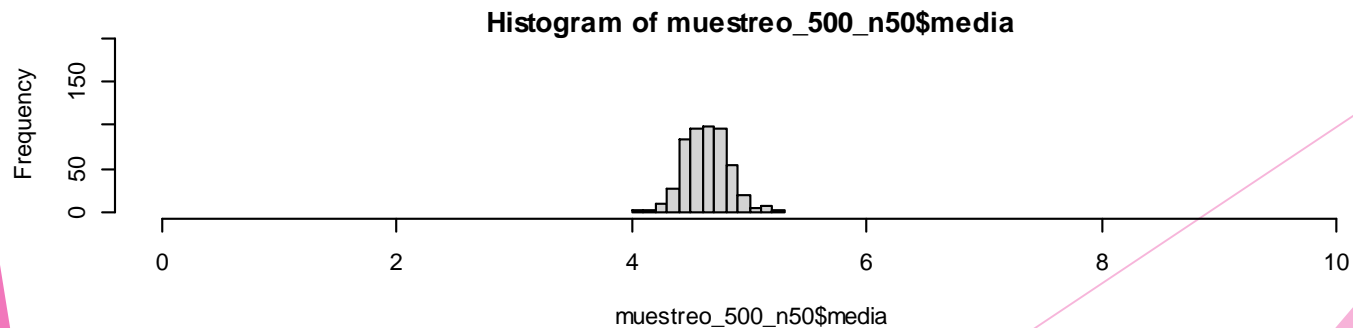
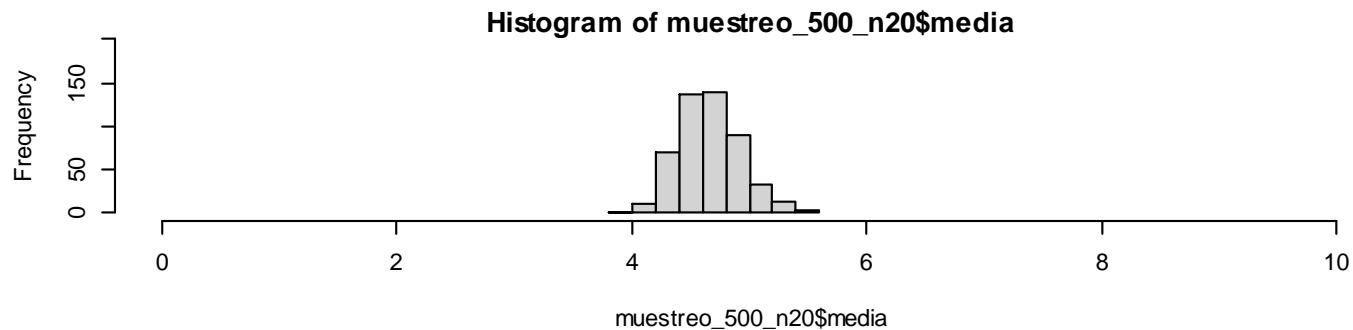
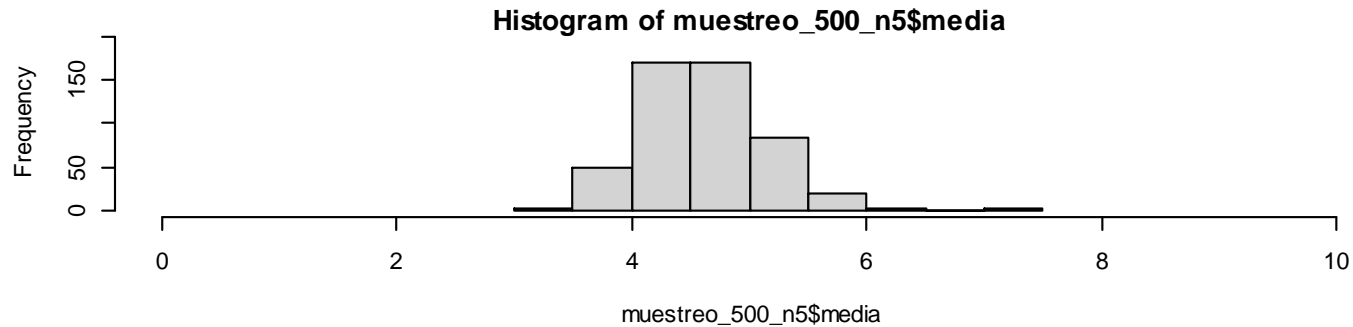
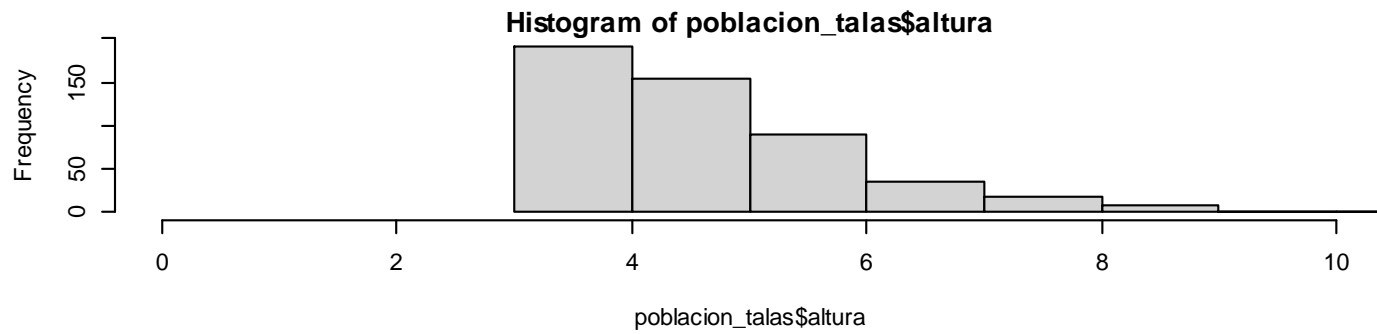
Automatización de la extracción de muestras.

1a) Comparen **la forma** de la distribución de la variable media para los distintos
tamaños de muestra y respecto de la distribución original de la variable altura.

Comparen las cuatro distribuciones

¿Qué conclusiones pueden sacar con respecto a la forma, la media y la varianza?

¿Con qué concepto visto en la teoría de distribuciones muestrales lo pueden relacionar?.



\bar{x}

\bar{x}

\bar{x}

ACTIVIDAD C

El desvío de la variable media recibe el nombre de ERROR ESTANDAR

$$\text{e.s. (media)} = \sigma^2/n = \sigma/\sqrt{n} =$$

1-b) ¿Qué sucede con la varianza de la distribución de medias a medida que aumenta el n?

Comparen la varianza de las medias de muestras de distinto tamaño (n) con la varianza de la variable original.

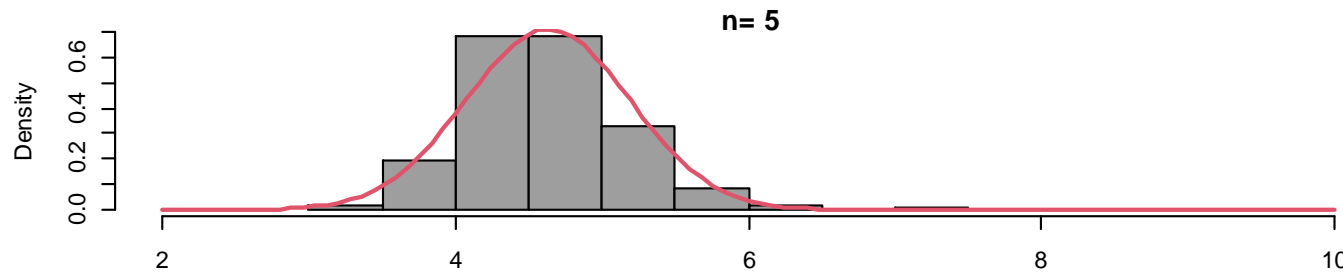
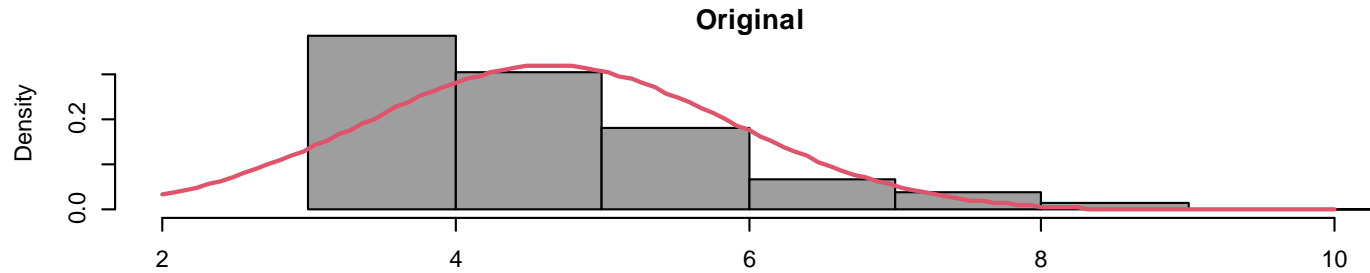
<code>var(muestreo_500_n5\$media)</code>	<code>var(muestreo_500_n5\$media)</code>	0.318
<code>var(muestreo_500_n20\$media)</code>	<code>var(muestreo_500_n20\$media)</code>	0.073
<code>var(muestreo_500_n50\$media)</code>	<code>var(muestreo_500_n50\$media)</code>	0.030
<code>varp(poblacion_talas\$altura)</code>	<code>varp(poblacion_talas\$altura)</code>	1.569

En base a lo que vieron en la teórica, ¿pueden deducir alguna relación por la cual cambia la varianza de la variable media respecto de la varianza original?

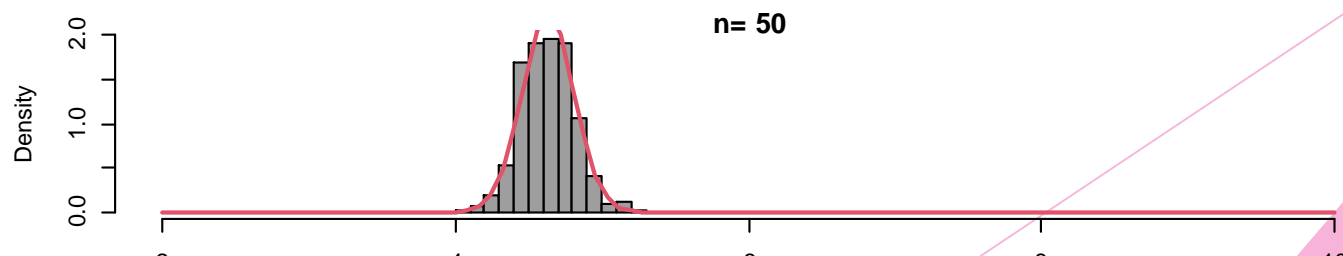
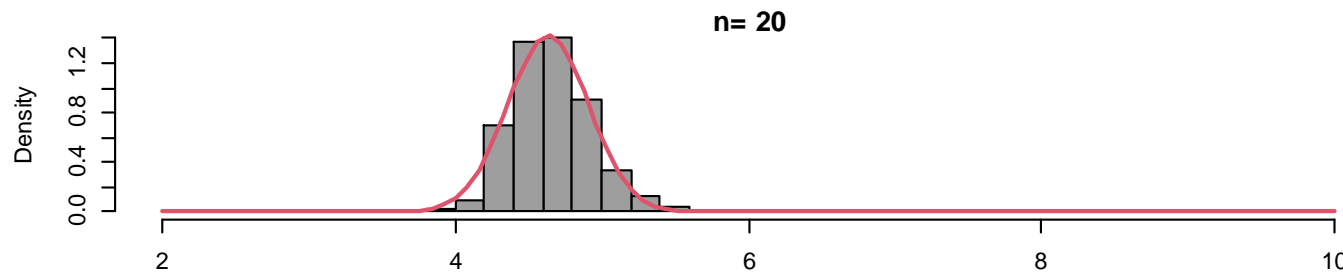
2) Ahora compararemos la forma de la distribución de la variable media de la altura de talas obtenida para cada tamaño de muestra n (5, 20 y 50) con la forma esperada si fuese normal.

Tener en cuenta que el eje y del gráfico ya no es frecuencia sino que es una función de densidad de probabilidad $f(x)$.

Comparen los gráficos obtenidos y concluyan.



$$\bar{x} \sim N(\mu; \sigma/\sqrt{n})$$



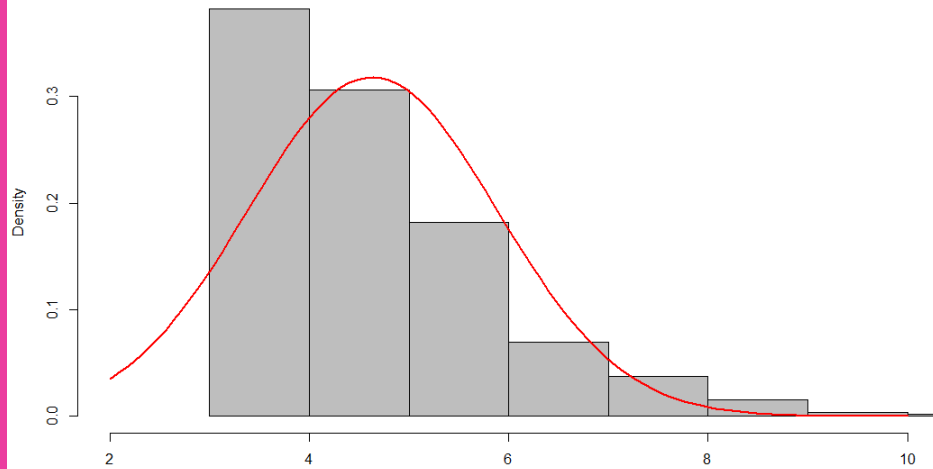
\bar{x}

\bar{x}

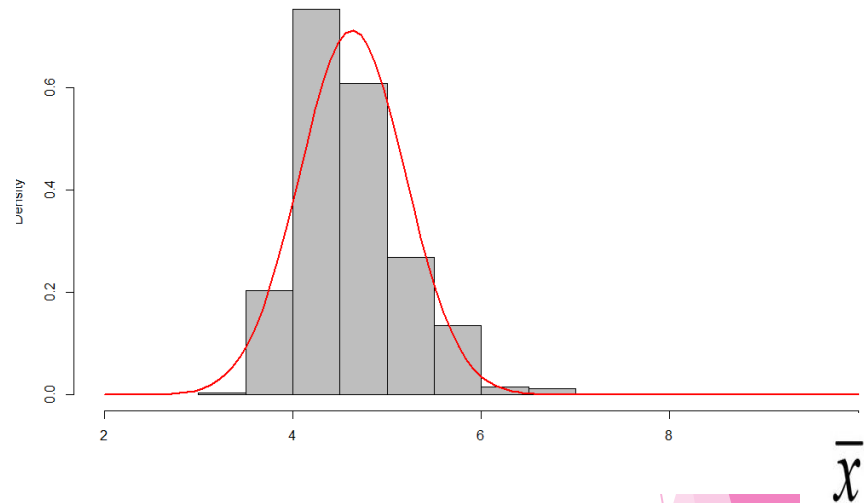
\bar{x}

Comparen los gráficos obtenidos y concluyan.

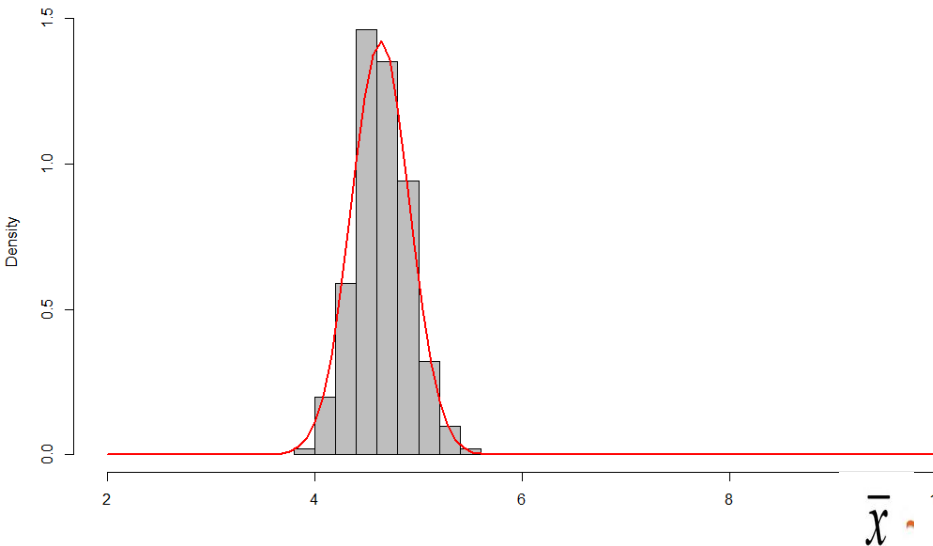
Original



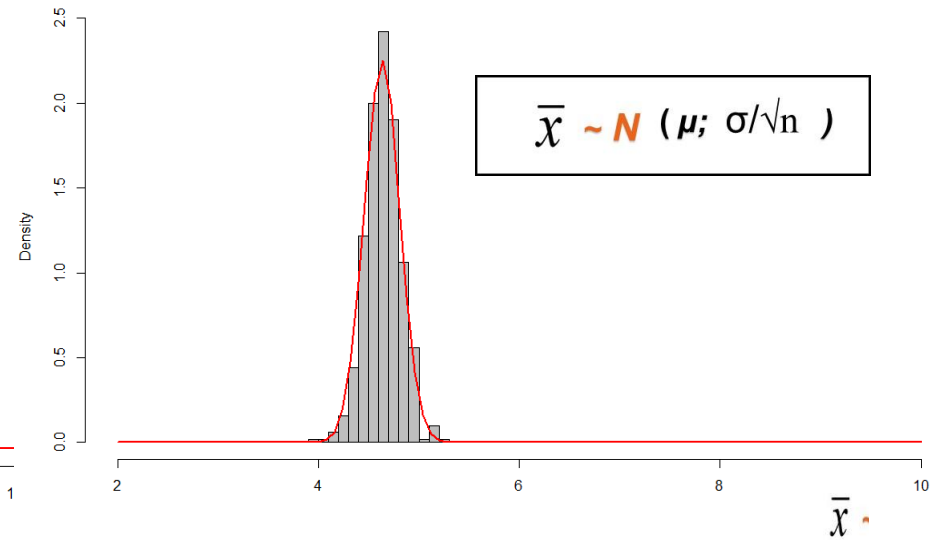
n= 5



n= 20

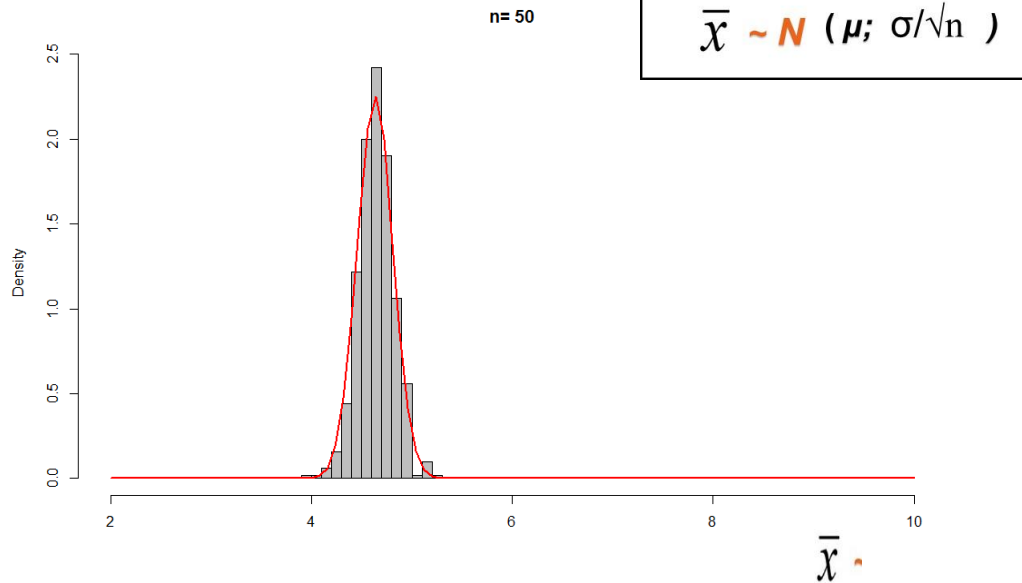


n= 50



$$\bar{x} \sim N(\mu; \sigma/\sqrt{n})$$

Si conozco sigma



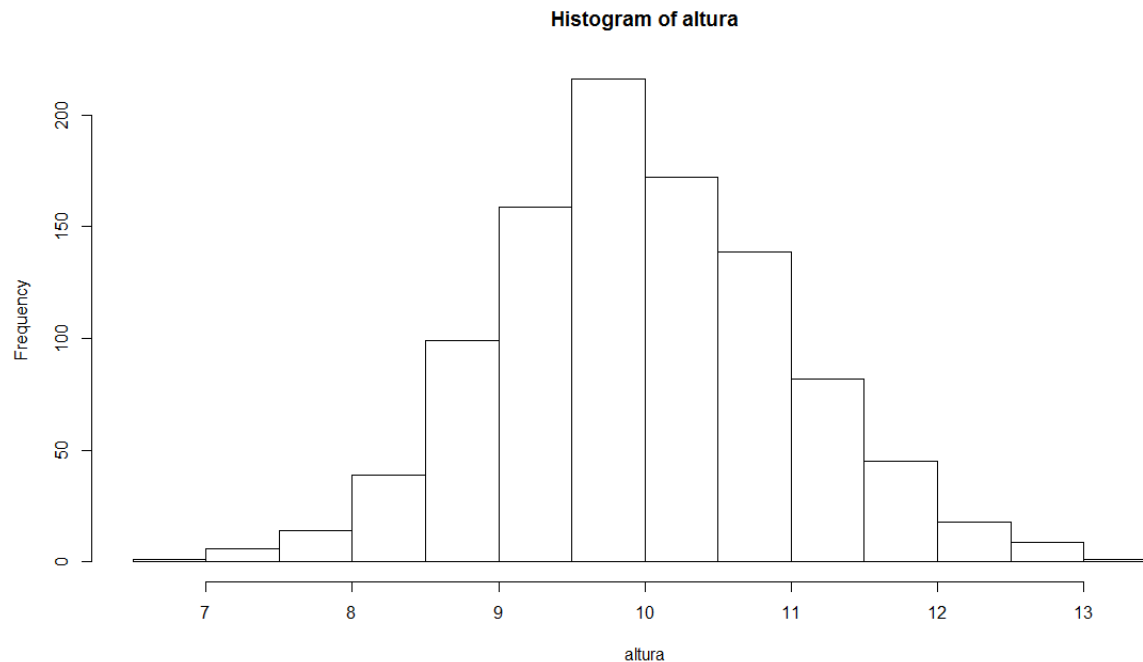
Si no conozco sigma

Estimo el e.s. de la media con el
desvío estándar de la muestra:
$$\text{e.s.} = s\sqrt{n}$$

ACTIVIDAD D ## .

Simularemos una población de ligustros (*Ligustrum lucidum*), un arbusto originario de Asia y muy común en la reserva. Se sabe que la altura de esta población en la reserva presenta una media de 10 m y un desvío de 1 m.

A partir de esta información generen 1000 datos de la variable "altura" con distribución normal con media de 10 m y desvío de 1 m usando la función rnorm:



#Repitan la simulación extrayendo muestras de tamaño n (5, 20 y 50), al igual que en la actividad C, graficando los histogramas correspondientes a la variable altura media.

:

