Práctica 2: Probabilidad, muestras y población

1. Variables aleatorias con distribución conocida

En **R** existen comandos para el cálculo de probabilidades con función de distribución conocida. En el caso de las distribuciones discretas que hemos visto, tenemos la distribución binomial *binom*, Poisson *pois* y Geométrica *geom*. Entre algunas de las funciones de distribuciones contínuas se encuentran la distribución Normal *norm*, t-Student t, χ^2 *chisq*, F-Snedecor f.

En todas ellas hay que añadir un prefijo dependiendo de lo que se desee trabajar:

> d: para funciones de probabilidad,
>
> p: cuando se trata de funciones de distribución,
>
> r: para generar valores de la distribución elegida, y
>
> q: para percentiles.

Por ejemplo, si dada una variable aleatoria $X_1 \sim Po(\lambda = 2)$ queremos calcular $P(X_1 = 4)$,

```
dpois(4,2)
```

```
## [1] 0.09022352
```

Si en cambio, dada la variable aleatoria $X_2 \sim Bi(n = 5, p = 0.25)$, queremos conocer la $P(X_2 \leq 2)$

```
pbinom(2,size=5,prob=0.25)
```

```
## [1] 0.8964844
```

Para calcular una muestra aleatoria de una $N(\mu = 22, \sigma = 7)$ y dibujarla tendríamos que hacer:

```
sample10<-rnorm(10,22,7)
sample10
hist(sample10,freq = FALSE, col = "blue")
```

Con ayuda del **R** podemos calcular las gráficas de las funciones de probabilidad y de distribución. Por ejemplo, para $X \sim Bi(5, 0.2)$:

- Función de probabilidad

```
x <- 0:5
plot(x, dbinom(x, 5,0.2), xlab="x", ylab="probabilidad",
     main="Función de probabilidad, Bi(5,0.2)", type="h")
  points(x, dbinom(x, 5, 0.2), pch=16)
abline(h=0, col="gray")
```

```
remove(x)
```

- Función de distribución:

```
 x <- 0:5
 x <- rep(x, rep(2, length(x)))
    plot(x[-1], pbinom(x, 5,0.2)[-length(x)], xlab="x", ylab="probabilidad",
```

```
     main="Función de distribución, Bi(5,0.2)", type="l")
  abline(h=0, col="gray")
```

```
remove(x)
```

Análogamente en el caso de funciones contínuas como una distrubución $N(50, 10)$:

- Función de densidad

```
x <- seq(10,90 , length=100)
   plot(x, dnorm(x, mean=50, sd=10), xlab="x", ylab="densidad",
 main="Normal(mu=50, sigma=10)", type="l")
abline(h=0, col="gray")
```

```
remove(x)
```

- Función de distribución:

```
x <- seq(10, 90, length=100)
   plot(x, pnorm(x, mean=50, sd=10), xlab="x", ylab="probabilidad",
 main="Normal(mu=50, sigma=10)", type="l")
abline(h=0, col="gray")
```

```
remove(x)
```

Ejercicio 1. En un estudio con ratas de laboratorio se rechazan todas aquellas que pesan menos de 36 gramos. Si el peso sigue una distribución Normal con $\mu = 40$ y $\sigma = 5$,

a. ¿Qué porcentaje de ratas rechazaremos?
b. Supongamos que se escogen 4 ratas al azar:
   (i) ¿cuál es la probabilidad que se rechacen al menos dos de ellas?
   (ii) ¿cuál es la probabilida que el peso medio de ellas sea menor a 36 gramos?
c. Dibuja las funciones de probabilidad/densidad y distribución de las variables aleatorias del problema.

Respuesta:

Apartado a

```
pnorm(36,40,5)*100
```

```
## [1] 21.18554
```

Apartado b.1

```
pbinom(1, 4, pnorm(36,40,5), lower.tail = FALSE)
```

```
## [1] 0.1992705
```

Apartado b.2

Emplearemos que si $X_i \sim N(\mu_i, \sigma_i)$ y $\alpha_i \in \mathbb{R}$, $i = 1, 2, ...n$ entonces

$$Y = \sum_{i=1}^{n} \alpha_i x_i \sim N\left(\sum_{i=1}^{n} \alpha_i \mu_i, \sqrt{\sum_{i=1}^{n} \alpha_i^2 \sigma_i^2}\right)$$

Por lo tanto, la media que estamos estudiando seguirá una distribución

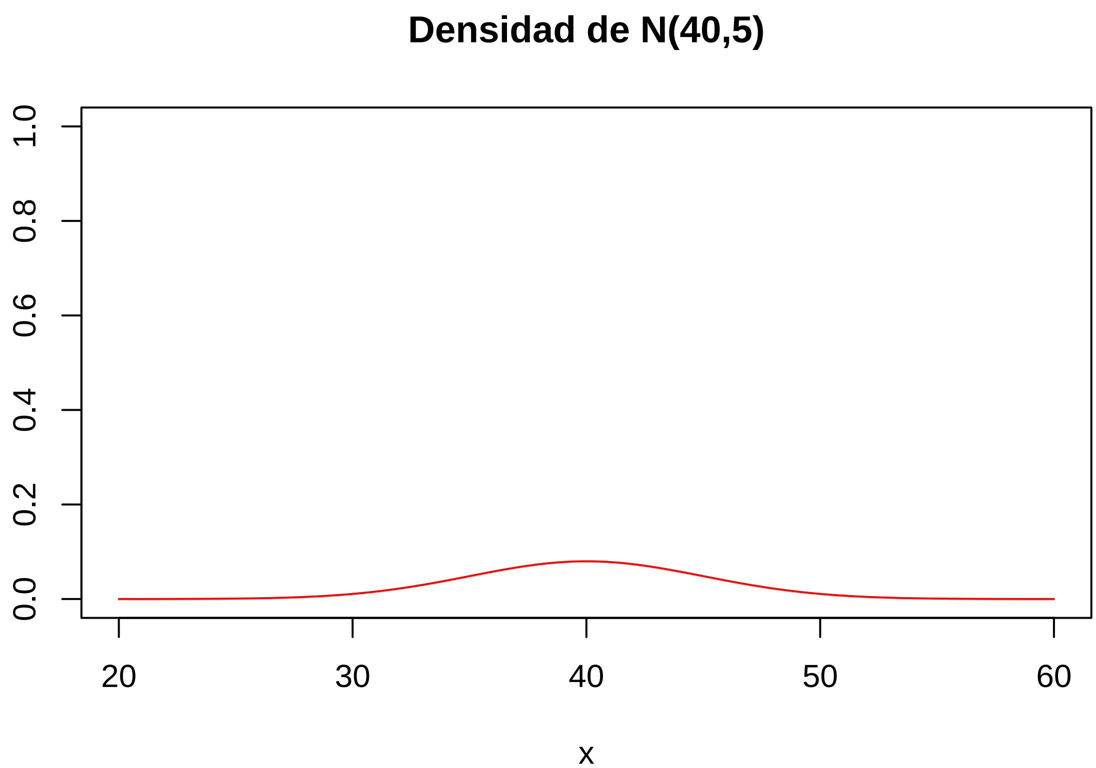
$$\frac{1}{4}(X_1 + X_2 + X_3 + X_4) \sim N\left(\frac{1}{4} \cdot 4 \cdot 40, \sqrt{\sum_{i=1}^{4} \frac{1}{4^2} \cdot 5^2}\right) = N\left(40, \frac{5}{2}\right)$$

```r
pnorm(36, mean=40, sd=5/2)
```

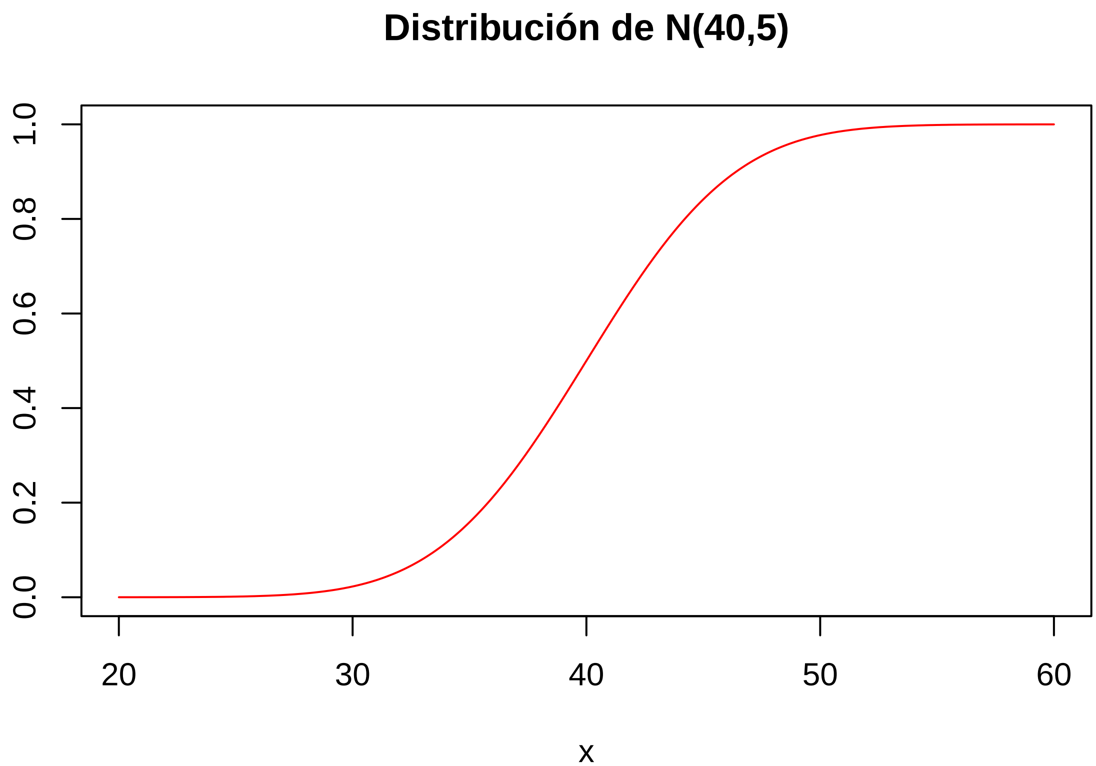
```
## [1] 0.05479929
```

Apartado c

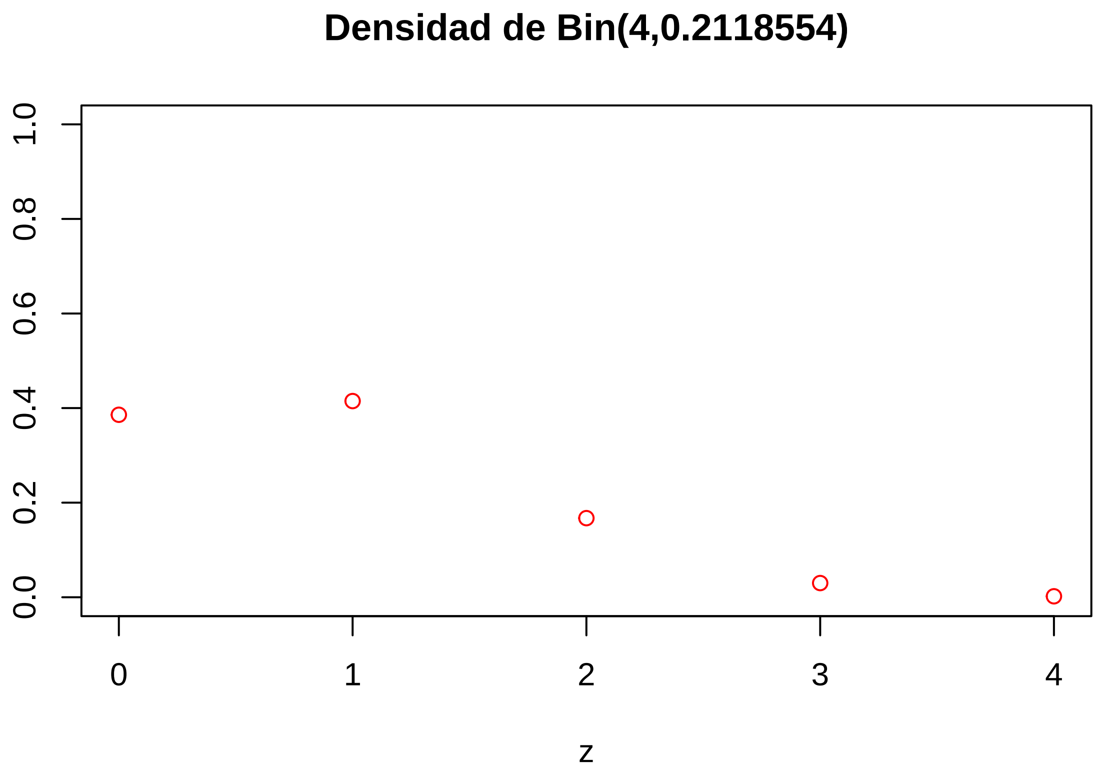
```r
x<-seq(20, 60, 0.1)
z<-seq(0,4,1)
plot(x, dnorm(x, mean=40, sd=5), type="l", ylim=c(0,1), ylab="", col="red", main="Densidad de N(40,5)")
```



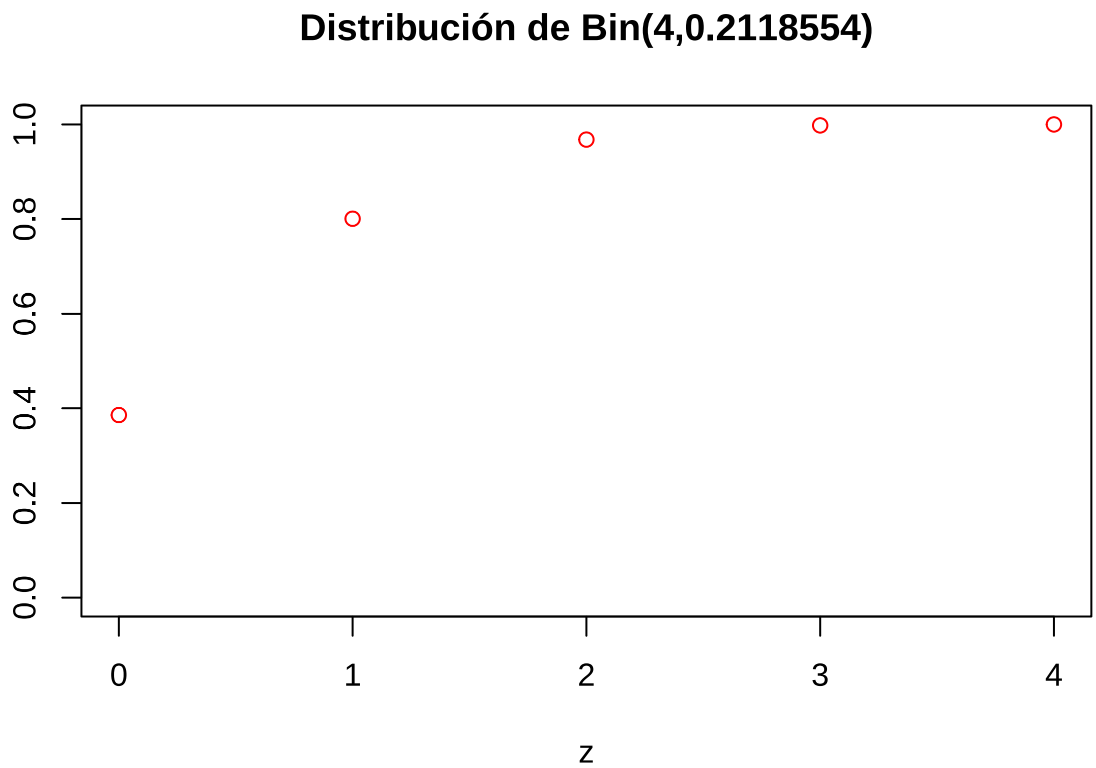

```r
plot(x, pnorm(x, mean=40, sd=5), type="l", ylim=c(0,1), ylab="", col="red", main="Distribución de N(40,5
```



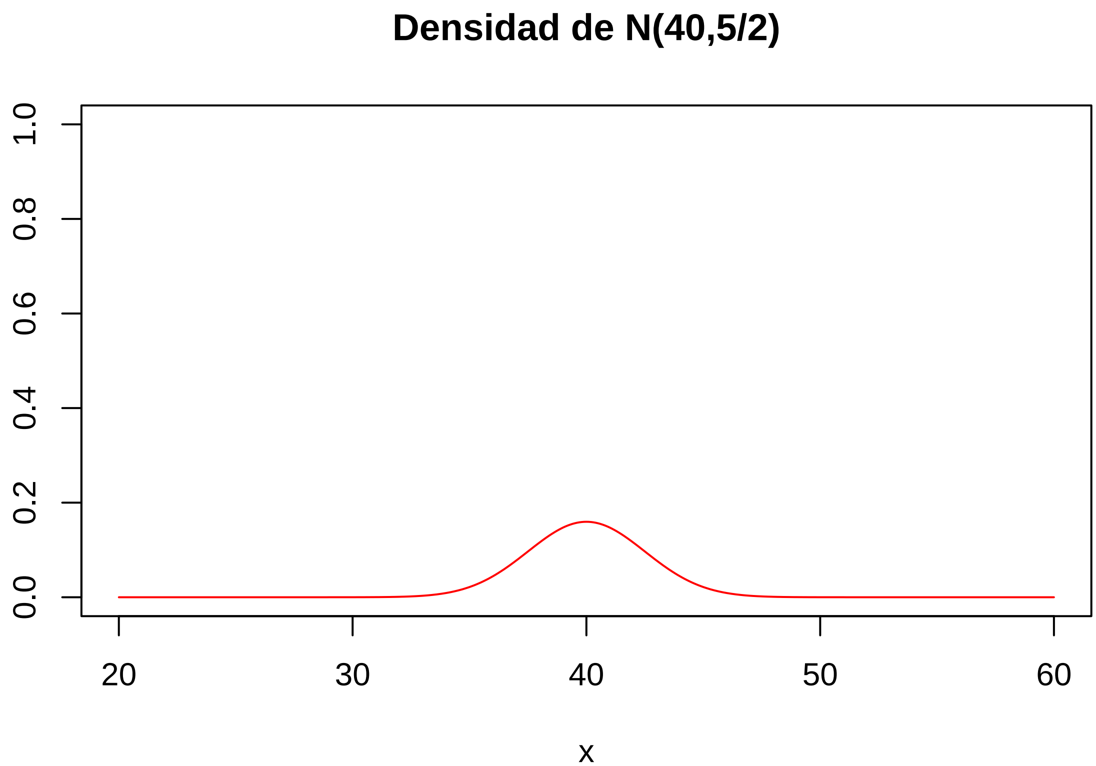

```
plot(z, dbinom(z, size=4, prob=pnorm(36,40,5)), ylim=c(0,1), ylab="", col="red", main="Densidad de Bin(4
```



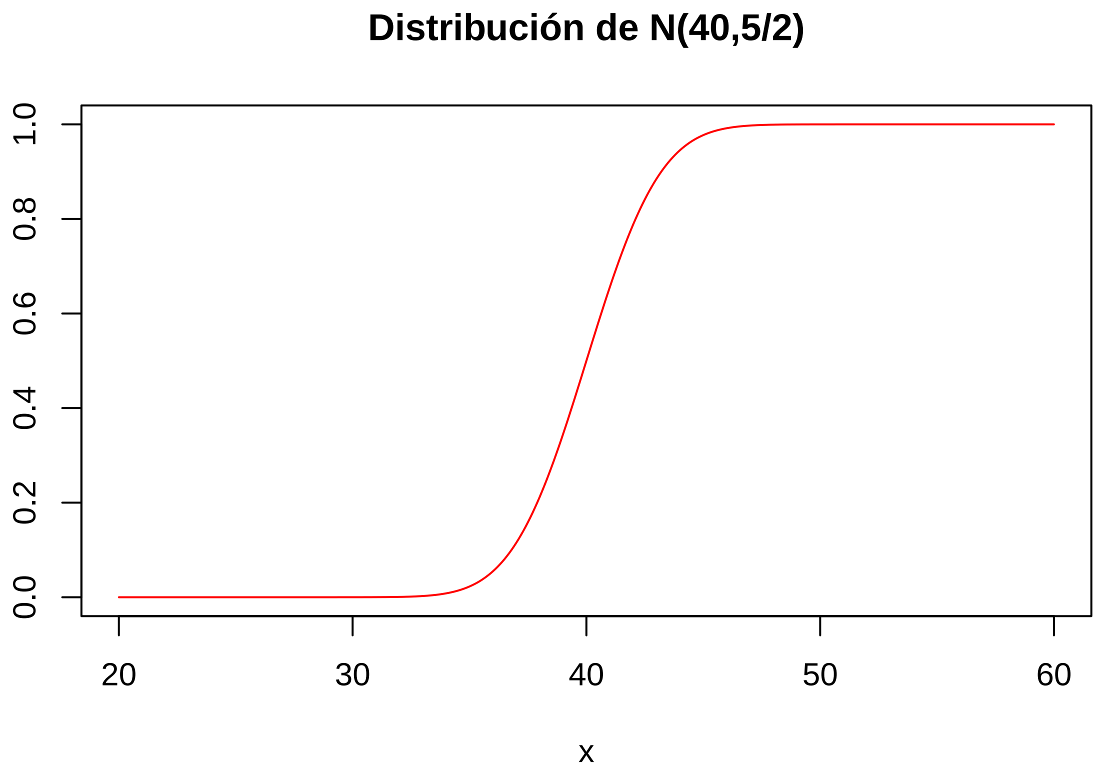

```
plot(z, pbinom(z, size=4, prob=pnorm(36,40,5)), ylim=c(0,1), ylab="", col="red", main="Distribución de B
```




```
plot(x, dnorm(x, mean=40, sd=5/2), type="l", ylim=c(0,1), ylab="", col="red", main="Densidad de N(40,5/2
```



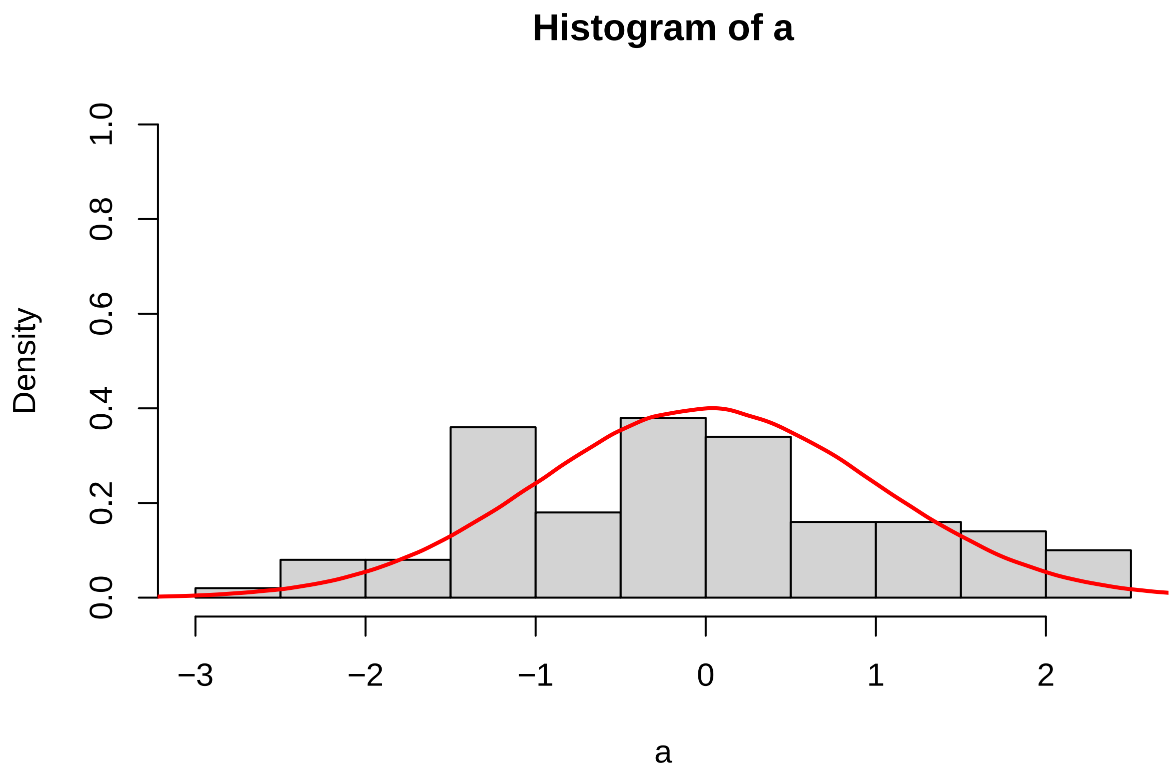

```
plot(x, pnorm(x, mean=40, sd=5/2), type="l", ylim=c(0,1), ylab="", col="red", main="Distribución de N(40
```



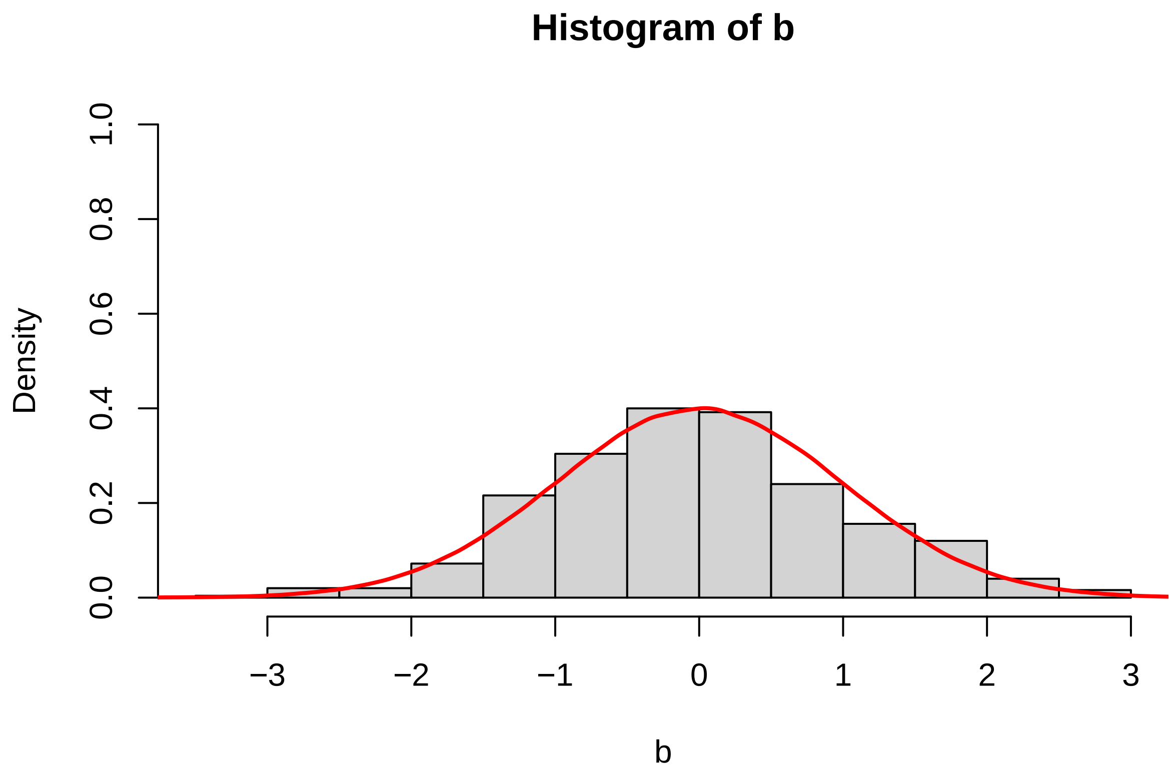
2. Muestras y población

Ejercicio 2. Genera 100, 500, 1000, 5000 observaciones de una población $N(0,1)$. Dibuja el histograma de cada una de ellas. Y añade en cada una de ellas la gráfica de la función de densidad de $N(0,1)$. Calcula la media y desviación típica de cada una de las muestras generadas y compáralas con la media y desviación típica poblacional. Razona qué observas.

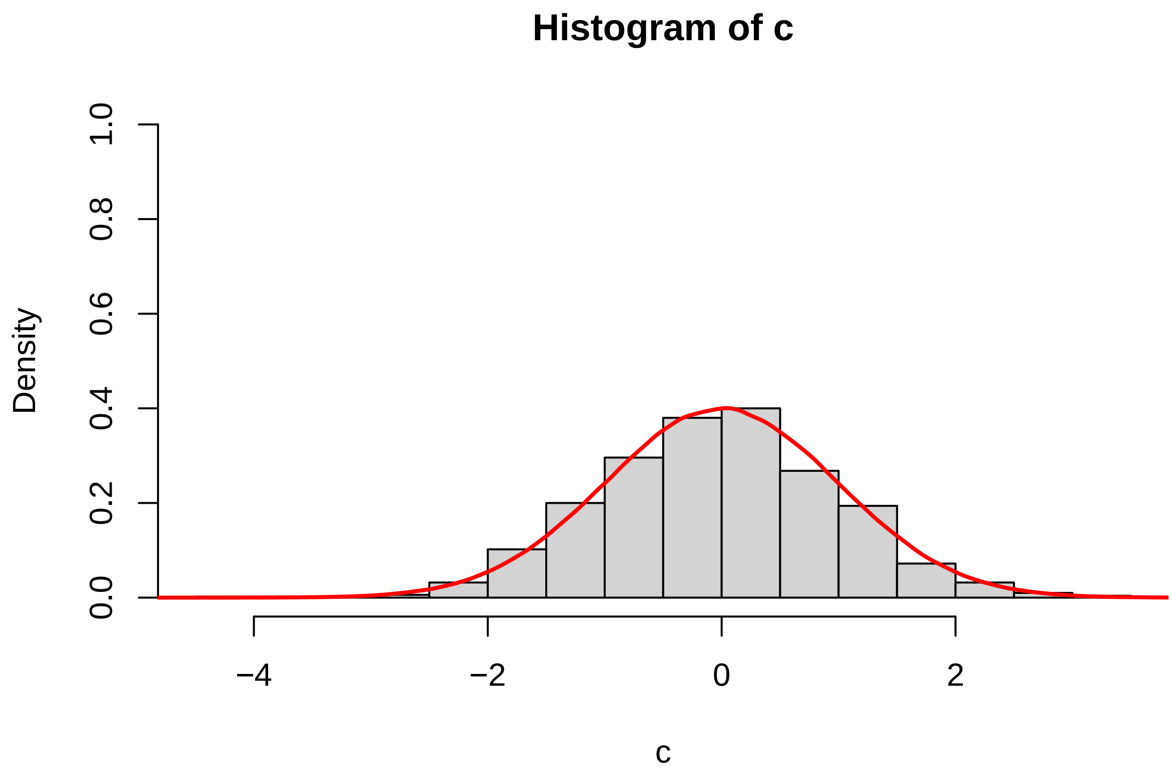
```
a<-rnorm(100)
b<-rnorm(500)
c<-rnorm(1000)
d<-rnorm(5000)
e<-rnorm(1000000)

hist(a, freq=FALSE, ylim=c(0,1))
lines(density(e), lwd=2, col="red")
```



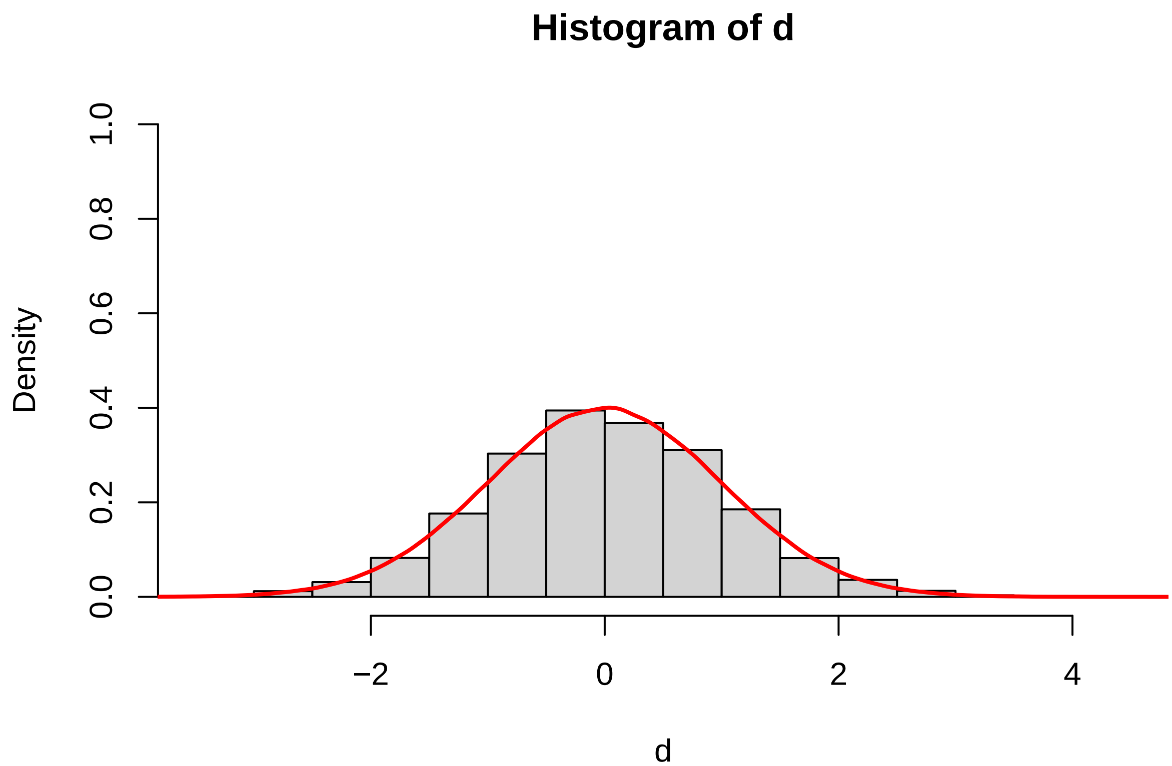

```
hist(b, freq=FALSE, ylim=c(0,1))
lines(density(e), lwd=2, col="red")
```




```
hist(c, freq=FALSE, ylim=c(0,1))
lines(density(e), lwd=2, col="red")
```




```
hist(d, freq=FALSE, ylim=c(0,1))
lines(density(e), lwd=2, col="red")
```




```
Media<-sapply(list(a,b,c,d), mean)
Desviacion<-sapply(list(a,b,c,d), sd)
g<-rbind(Media,Desviacion)
colnames(g)<-c("100", "500", "1000", "5000")
g
```

```
##                   100          500        1000       5000
## Media      -0.1090153 -0.005961009 -0.03611034 0.01775522
## Desviacion  1.1875875  1.026429112  1.00753480 1.00492941
```

Respuesta: observamos cómo va convergiendo la media hacia el valor 0 y la desviación típica hacia 1 a medida que el tamaño de la muestra va aumentando, como era de esperar.