UNIVERSIDAD DE EL SALVADOR

FACULTAD MULTIDISCIPLINARIA DE OCCIDENTE

DEPARTAMENTO DE MATEMÁTICA

Licenciatura en Estadística

Control Estadístico del Paquete R

"UNIDAD TRES"

Alumna:
Erika Beatríz Guillén Pineda

Fecha de elaboración
Santa Ana - 27 de noviembre de 2015

1. DISTRIBUCIONES CONTINUAS.

a) Distribución uniforme.

- **Parámetros.**

$$\text{x} = \text{valor cualquiera}$$

$$\text{p} = \text{probabilidad}$$

$$\text{n} = \text{tama?o de la muestra}$$

$$\text{min} = \text{valor mínimo}$$

$$\text{max} = \text{valor máximo}$$

- **Sintaxis de la función utilizada en R**

1. punif(x, min, max, lower.tail = TRUE, log.p = FALSE)
2. qunif(p, min, max, lower.tail = TRUE, log.p = FALSE)
3. runif(n, min, max)

b) Distribución Normal.

- **Parámetros.**

$$\text{x} = \text{valor cualquiera}$$

$$\text{p} = \text{probabilidad}$$

$$\text{mean} = \text{media}$$

$$\text{sd} = \text{desviación típica}$$

$$\text{n} = \text{tamaño de la muestra}$$

- **Sintaxis de la función utilizada en R**

1. pnorm(x, mean, sd, lower.tail = TRUE, log.p = FALSE)
2. qnorm(p, mean, sd, lower.tail = TRUE, log.p = FALSE)
3. rnorm(n, mean, sd)

c) Distribución T-Student.

- **Parámetros.**

x = valor cualquiera

p = probabilidad

df = grados de libertad

- **Sintaxis de la función utilizada en R**
  1. pt(x, df, lower.tail = TRUE, log.p = FALSE)
  2. pt(x, df, lower.tail = TRUE, log.p = FALSE)
  3. rt(n, df)

d) Distribuión Chi-cuadrado.

- **Parámetros.**

x = valor cualquiera

df = grados de libertad

p = probabilidad

- **Sintaxis de la función utilizada en R**
  1. pchisq(x, df, lower.tail = TRUE, log.p = FALSE)
  2. qchisq(p, df, lower.tail = TRUE, log.p = FALSE)
  3. rchisq(n, df,)

d) Distribución F de Snedecor.

- **Par?metros.**

x,q = vector cuantiles

df1 = grados de libertad en el numerador

df2 = grados de libertad en el denominador

p = vector probabilidad

- **Sintaxis de la función utilizada en R**
  1. pf(q, df1, df2, ncp, lower.tail = TRUE, log.p = FALSE)
  2. qf(p, df1, df2, ncp, lower.tail=TRUE, log.p = FALSE)
  3. rf(n, df1, df2, ncp)

d) **Distribución Exponencial.**

- **Parámetros.**

x,q = vector cuantiles

rate = raz?n = 1/E[X]

p = vector probabilidad

lower.tail = T

- **Sintaxis de la función utilizada en R**

1. pexp(q, rate = 1, lower.tail = TRUE, log.p = FALSE)
2. qexp(p, rate = 1, lower.tail = TRUE, log.p = FALSE)
3. rexp(n, rate = 1)

2. CÁLCULO DE PROBABILIDADES.

- **Ejemplo 1:**

Una persona informal hace esperar a su pareja aleatoriamente entre 0 y 90 minutos. Harto de esta situación, la persona que sufre la espera se plantea un ultimátum; sí al día siguiente su pareja tarda menos de 15 minutos mantiene la relación, sí la espera esté entre 15 y 55 minutos, decide en la siguiente cita con los mismos criterios, mientras que si tarda más de 55 minutos la relación termina en ese momento.

a) Calcule la probabilidad de que la relaci'on continúe hasta la siguiente cita.

```
x <- 55; a=0; b <- 90

# usando la funcion propia de R

punif(x, min=a, max=b, lower.tail=TRUE)

## [1] 0.6111111
```

b) Calcule la probabilidad de que la relación termine en la segunda cita.

Suponiendo que el tiempo de espera en una cita es independiente respecto de otras citas, se calcula la probabilidad P(15<X<55) = P(X<55) - P(X<=15) = 0.6111 - 0.1666 = 0.4445,

```
F55=punif(55, min=a, max=b, lower.tail=TRUE)
F15=punif(15, min=a, max=b, lower.tail=TRUE)
F55-F15

## [1] 0.4444444
```

que es la probabilidad de que aplace la decisión para la segunda cita y, en la segunda cita, la probabilidad de que lo deje definitivamente es P(X>55) = 0.3888,

```
F55=punif(55, min=a, max=b, lower.tail=TRUE);F55

## [1] 0.6111111
```

luego multiplicando ambas probabilidades se obtiene el valor pedido 0.1728.

```
(1-F55)*( F55-F15)

## [1] 0.1728395
```

- **Ejemplo 2:**

Una empresa está buscando personal para su departamento de mercadeo. El perfil solicitado es el de sujetos extrovertidos y creativos. Se han presentado 50 candidatos y la empresa ha establecido como criterio de selección que los candidatos superen el percentil 80 en creatividad y extroversión. Sabiendo que la variable extroversión (X) se distribuye según una Normal de media 5 y desviación típica 1, que la variable creatividad (Y) sigue una t-Student de 10 grados de libertad y que las puntuaciones de creatividad y extroversión son independientes:

a) Cuántos candidatos serán seleccionados?

Al ser X e Y independientes, la probabilidad:

P(X>=P80nY>=P80)=P(X>=P80)P(Y>=P80)=(0.20)(0.20)=0.04.Como se han presentado 50 aspirantes, ser?n seleccionadas (50)(0.04)=2 personas.

b) Qué puntuaciones debe superar un aspirante en creatividad y extroversión para ser admitido?

Según el criterio de selección se debe superar el percentil 80, en ambas variables, para ser admitido. Se calcular? pues el percentil 80 de la variable X e Y,utilizando los cuantiles-normales para la variable X:

```
# y los cuantiles-normales para la variable X:

p <- c(0.80); media=5; d.t=1
qnorm(p, mean=media, sd=d.t, lower.tail=TRUE)

## [1] 5.841621
```

```
#y los cuantiles-t para la variable Y:

p <- c(0.80); g.l <- 10
qt(p, df=g.l, lower.tail=TRUE)

## [1] 0.8790578
```

c) Si se extraen al azar 16 candidatos, cuál es la probabilidad de que su media aritmética en extroversión sea mayor que 4.5? Se sabe que al extraer una muestra de una población normal de tamaño n, lamedia muestral, sigue otra distribución normal de media igual que la poblacional y desviación típica sigma/sqrt(n).

   Como se desea calcular P(media>=4.5):

```
n <- 16; x <- 4.5; mu=5; sigma=1; d.t=sigma/sqrt(n)
pnorm(x, mean=mu, sd=d.t, lower.tail=FALSE)

## [1] 0.9772499
```

- **Ejemplo 3:**

La duración media de un modelo de marcapasos es de 7 años.

a) Cuál es la probabilidad de que dure al menos 5 años? y menos de 3 años?

   Suponiendo que la variable X=”tiempo de funcionamiento del marcapasos”sigue una distribución exponencial.

   La probabilidad P(X>=5) se obtiene así:

```
x <- 5; teta=7
pexp(x, rate=1/teta, lower.tail=FALSE)

## [1] 0.4895417
```

   y de igual forma P(X<3):

```
x <- 3; teta=7
pexp(x, rate=1/teta, lower.tail=TRUE)

## [1] 0.3485609
```

b) Si han transcurrido ya 4 a?os desde su implantación, cuál es la probabilidad de que dure otros 4? Nos piden P(X>=8/X>=4)

   Teniendo en cuenta que la función de distribuci?n es la única distribución continua no tiene memoria resulta que P(X>=8/X>=4)=P(X>=4)=0.5647182

```r
pexp(4, rate=1/teta, lower.tail=FALSE)

## [1] 0.5647181
```

c) Cuánto tiempo debería funcionar un marcapasos para estar entre el 10 % de los que más duran?

Hay que calcular el percentil 90:

```r
p <- 0.9; teta <- 7
qexp(p, rate=1/teta, lower.tail=TRUE)

## [1] 16.1181

#resultando 16.12 a\~nos.
```

d) Calcular el valor que deben tener a y b para que P(X<a)=0.5 PX a 0.5 y P(X>b)=0.32

De forma análoga al apartado anterior, en el primer caso habría que calcular la mediana (percentil 50), a = 4.852,

```r
qexp(0.5, rate=1/teta, lower.tail=TRUE)

## [1] 4.85203
```

```r
# y en el segundo caso, el percentil 68, b = 7.97

qexp(0.68, rate=1/teta, lower.tail=TRUE)

## [1] 7.97604
```

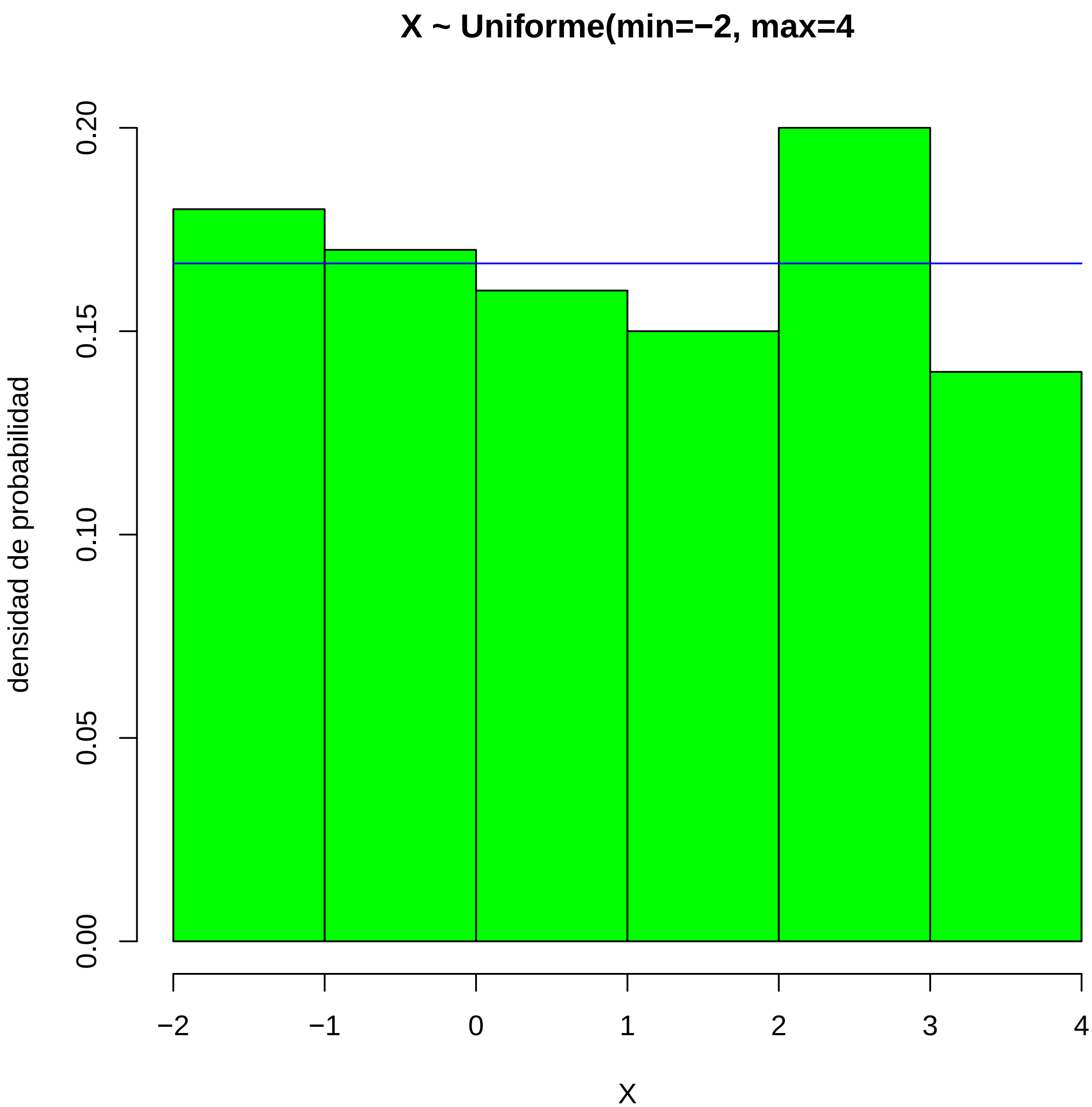
```r
# o de esta otra manera

qexp(0.32, rate=1/teta, lower.tail=FALSE)

## [1] 7.97604
```

3. GENERACIÓN DE MUESTRAS ALEATORIAS DE LAS DISTRIBUCIONES.

- **Ejemplo 1:**

Generar 100 números aleatorios de una distribución Uniforme en [-2, 4]

```
# Definir los par\'ametros apropiados
min <- -2; max <- 4

# generar 100 n\'umeros aleatorios de la distribuci\'on
x = runif(100, min, max); x

##   [1]  3.50599164 -1.41437247 -1.81093095  0.79539394  0.38696249
##   [6] -1.64690599  2.79525553  1.04426434  2.80259385  1.28802595
##  [11]  0.76347609  1.37693846  2.07908200 -0.34463274  0.79260070
##  [16]  1.30526229 -1.36618987 -0.66356981  0.92204716  2.01977357
##  [21]  3.99194739  0.79161218 -0.56457577  1.53208177  2.78693073
##  [26]  3.81870648 -0.76440684 -0.28335532  2.82397081  2.27727619
##  [31] -1.94559681 -0.23901882 -1.59827429 -1.15862641  0.77727450
##  [36]  1.08526759  1.62670488 -1.18364264  1.54727617 -1.12409356
##  [41]  1.64967100  0.49559029  1.99682257  3.66372125  3.52831505
##  [46]  0.71194340  3.29468014 -1.21165677  2.12622471  2.59276867
##  [51]  3.90304602  2.62782402  3.60919158  3.57060329  0.05746331
##  [56]  2.70497557  2.25014903 -0.80994192  0.36072687  3.17023431
##  [61]  1.80305761  2.72509675 -0.49311794 -0.39172836 -1.70066730
##  [66] -0.09144458  2.23457059  3.01191368  2.68741394  2.75500301
##  [71] -0.34996936  0.40522882  1.90744127  2.03871445  0.99306342
##  [76] -1.59047157  1.33492103  1.33921238  0.13325396 -0.78851292
##  [81] -1.99563119 -0.14089075 -1.23043896  0.02821621  2.16237955
##  [86]  2.35488852 -0.13148213  3.64775187 -1.42363359  1.36123238
##  [91] -0.49744195 -1.91749746  0.97899883  3.40141173 -1.29327494
##  [96] -0.40492579  3.01394634  2.52901141 -1.58691624 -0.50944734

# Histograma para la nuestra aleatoria de tama\~no 100
hist(x, main="X ~ Uniforme(min=-2, max=4", xlab="X",
     ylab="densidad de probabilidad", probability=TRUE, col="green")

# Graficar la funci?n de densidad, use la funci\'on curve() para variable continua
curve(dunif(x, min, max), col="blue", add=TRUE)
```

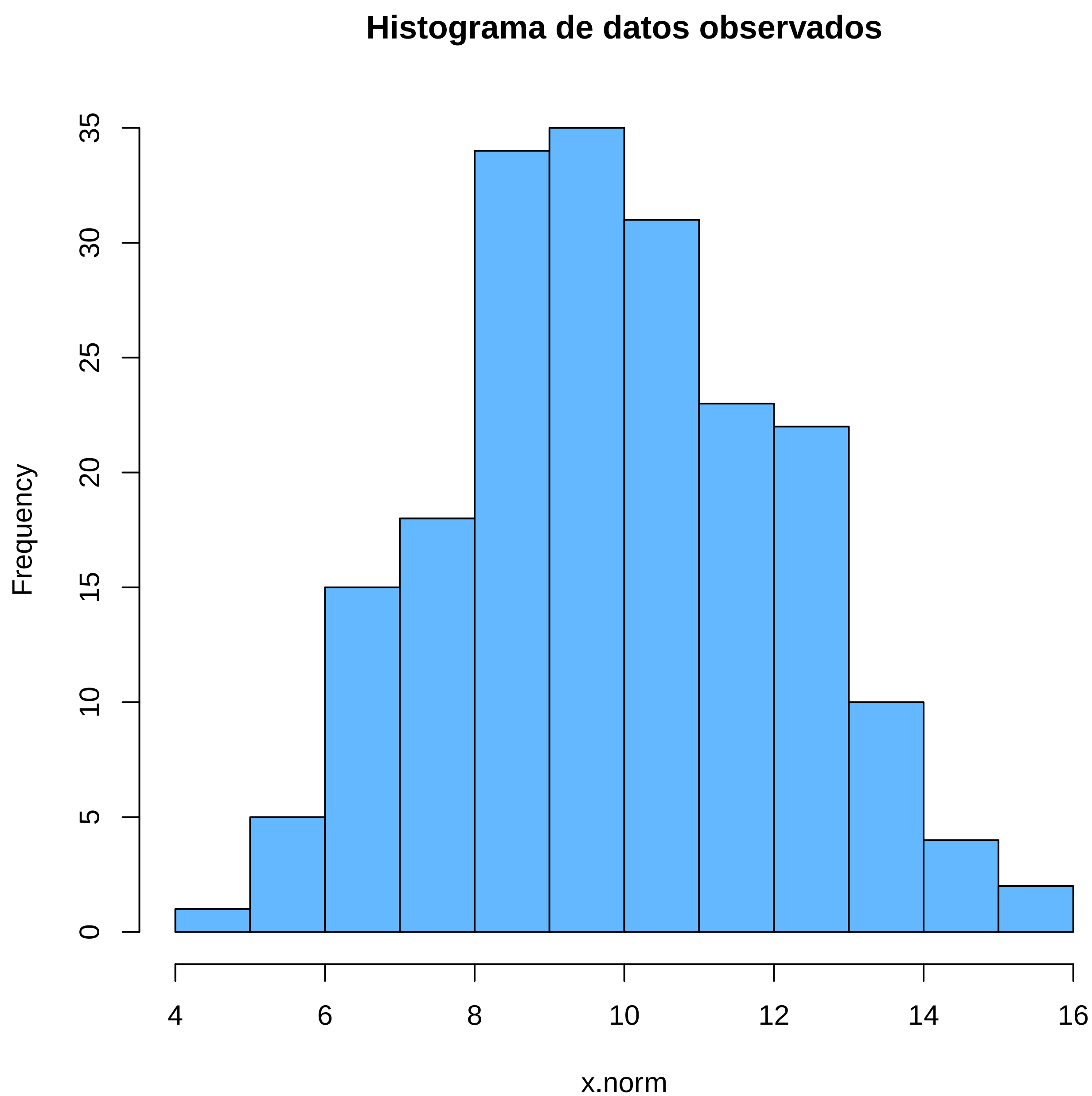


- **Ejemplo 2:**

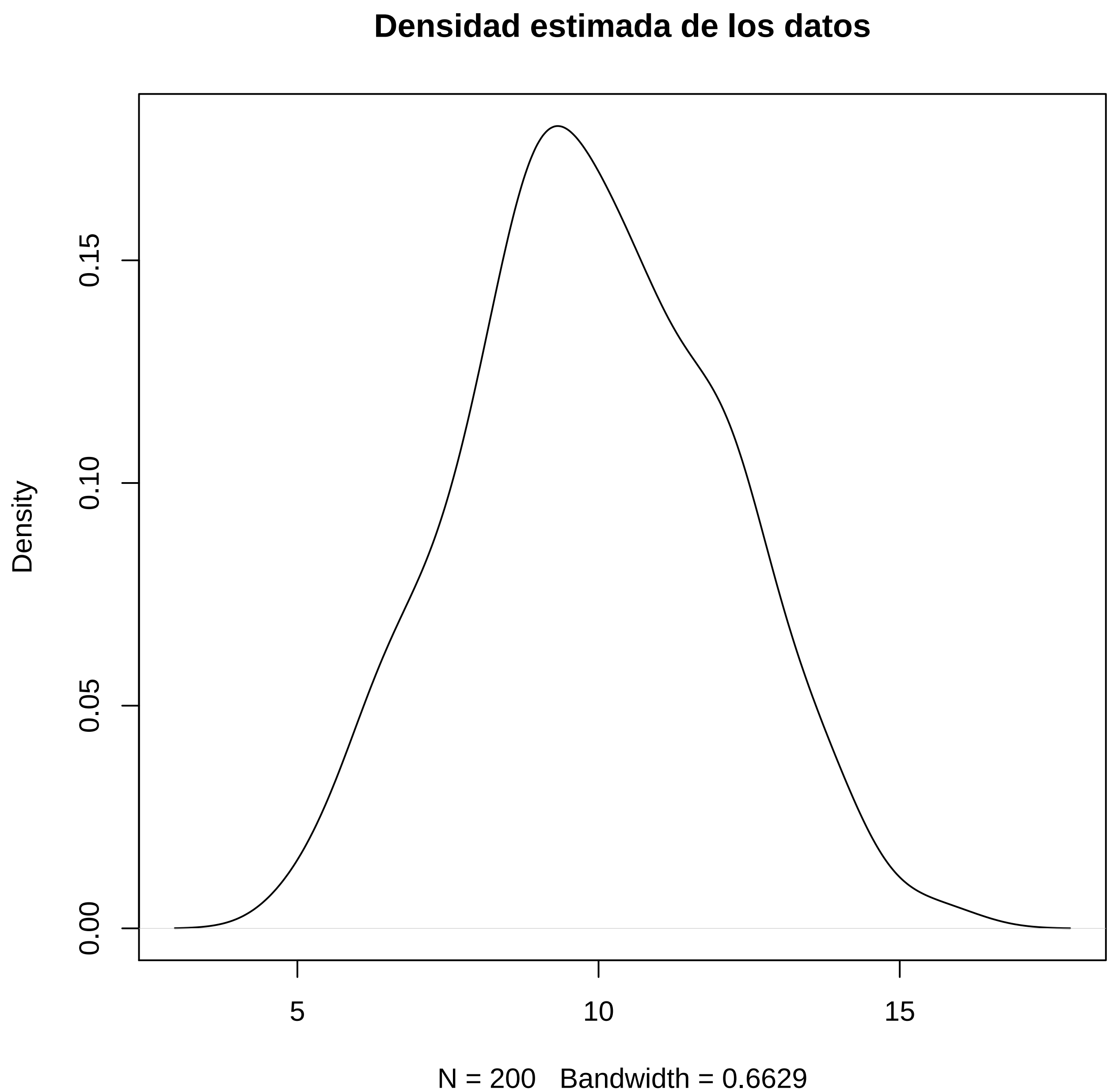
Supongamos que tenemos una muestra de tama?o n=200 perteneciente a una población normal N(10,2) con media igual a 10 y con una desviación estandar igual a 2:

```
# genera los valores aleatorios de la distribuci\'on
x.norm <- rnorm(n=200,mean=10, sd=2)

# Podemos obtener un histograma usando la funci\'on hist()
hist(x.norm, breaks = "Sturges", freq = TRUE, probability = FALSE,
     include.lowest = TRUE, right = TRUE, density = NULL,
```

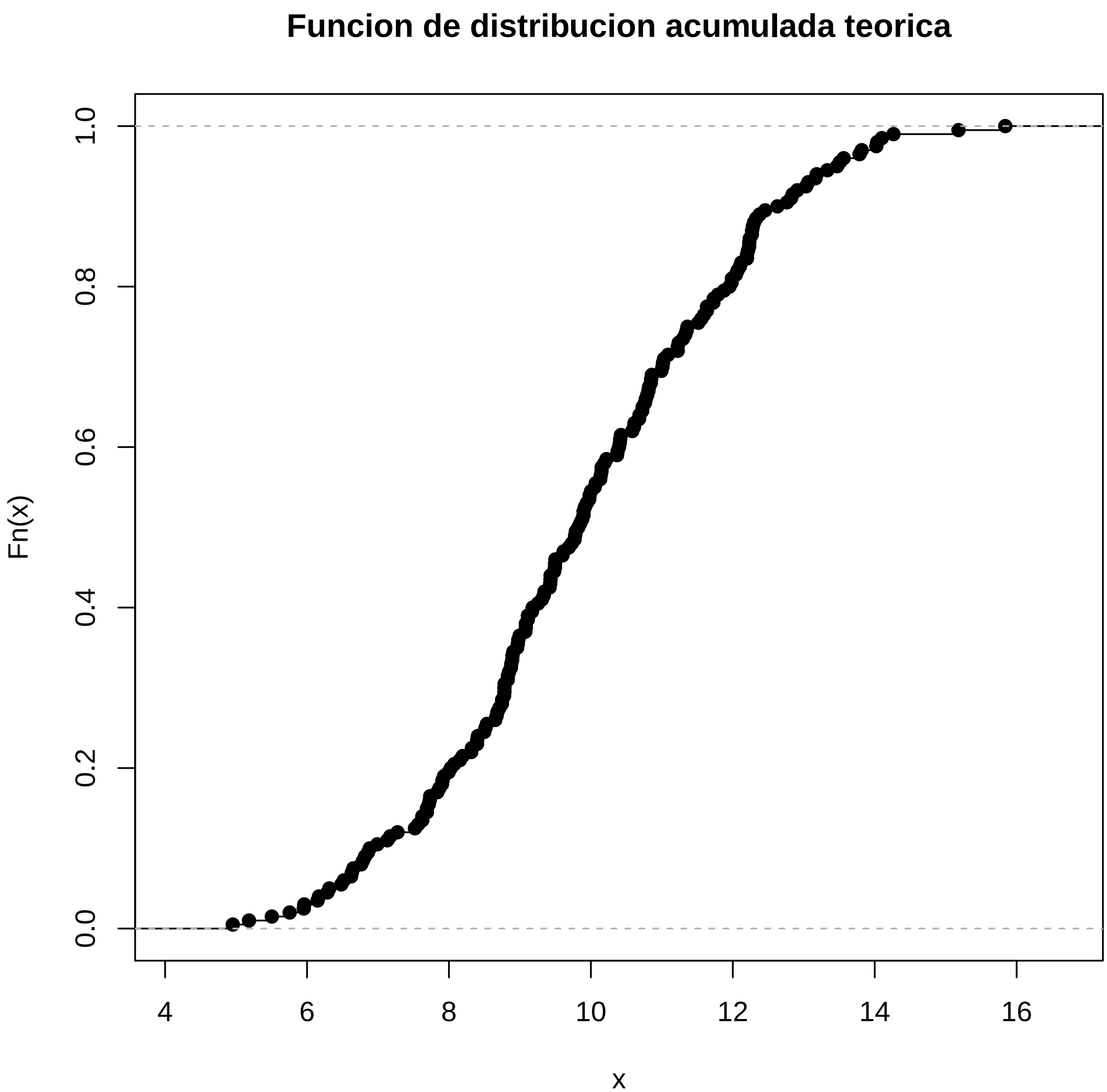
```
angle = 45, col = "steelblue1", border = NULL,
main = "Histograma de datos observados", axes = TRUE,
plot = TRUE, labels = FALSE)
```




```
# Podemos estimar la densidad de frecuencia usando la funci\'on
# density() y plot() para dibujar su "gr\'afica"
plot(density(x.norm), main="Densidad estimada de los datos")
```




```
# R permite calcular la funci\'on de distribuci\'on acumulada te\'erica con ecdf()
plot(ecdf(x.norm),main="Funcion de distribucion acumulada teorica")
```

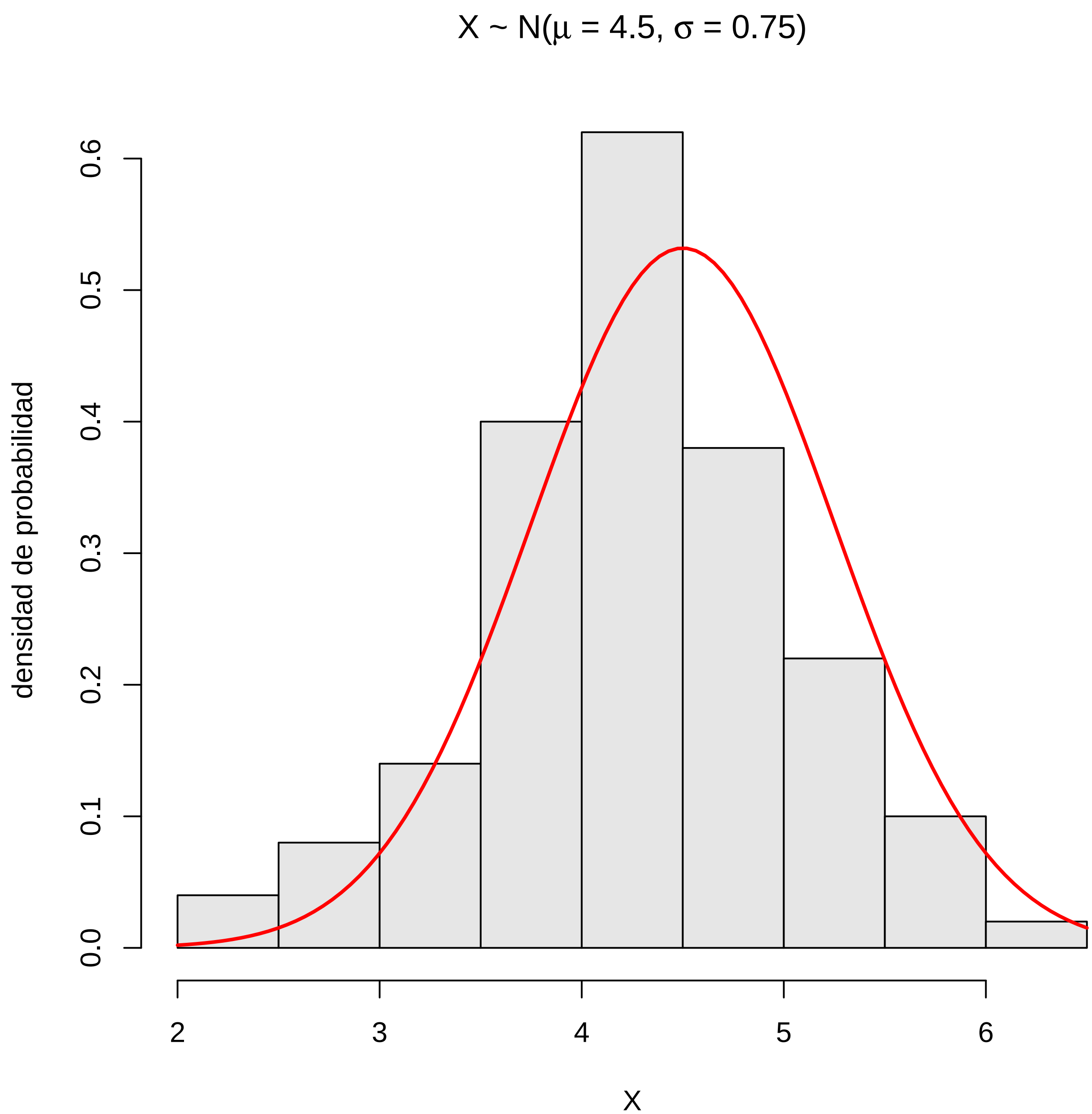


- **Ejemplo 3:**

Generar 100 números aleatorios de una distribución Normal con media 4.5 y desviación estándar 0.75.

```
# Definir los par\'ametros apropiados
media <- 4.5; desviacion <- 0.75

# generar 100 n\'umeros aleatorios de la distribuci\'on
x = rnorm(100, media, desviacion); x

##   [1] 5.114109 2.807644 4.829150 5.171246 3.324950 4.806662 4.467318
```

```
##   [8] 3.822467 4.808454 5.475646 4.288946 6.035396 2.796695 4.633310
##  [15] 4.344524 5.138528 3.687893 4.497320 5.129663 3.855692 5.467961
##  [22] 5.433751 4.242510 3.398178 3.172333 4.440449 4.681502 4.417731
##  [29] 3.910563 5.505933 4.454395 4.282156 3.772408 3.953513 3.592070
##  [36] 4.103023 3.303338 4.960042 4.145200 4.781578 4.423927 5.332887
##  [43] 4.188500 3.830366 5.305644 4.392712 4.124634 3.233090 4.493575
##  [50] 4.508792 4.262792 3.862822 5.890693 3.961443 3.701477 2.362809
##  [57] 4.433050 4.660429 2.928642 5.798794 3.842752 4.021842 4.438447
##  [64] 4.291631 2.820015 4.939433 4.441936 4.702720 4.858302 4.623014
##  [71] 3.903635 3.405556 4.337589 4.901855 4.112485 4.775759 4.720916
##  [78] 3.938021 4.889683 3.802883 2.264697 4.477672 3.510906 3.964042
##  [85] 4.309669 4.369897 3.784581 5.534855 4.089908 5.285568 4.188865
##  [92] 4.044965 3.245994 4.484572 4.532582 5.871118 3.531160 5.417253
##  [99] 3.920462 4.870577

# Histograma para la nuestra aleatoria de tama\~no 100
hist(x,main=expression(paste("X ~ N(", mu, " = 4.5, ", sigma, " = 0.75)")),
xlab="X", ylab="densidad de probabilidad", probability=TRUE, col=gray(0.9))

# Graficar la funci\'on de densidad te\'orica, usando la funci\'on curve()
curve(dnorm(x, media, desviacion), col="red", lwd=2, add=TRUE)
```

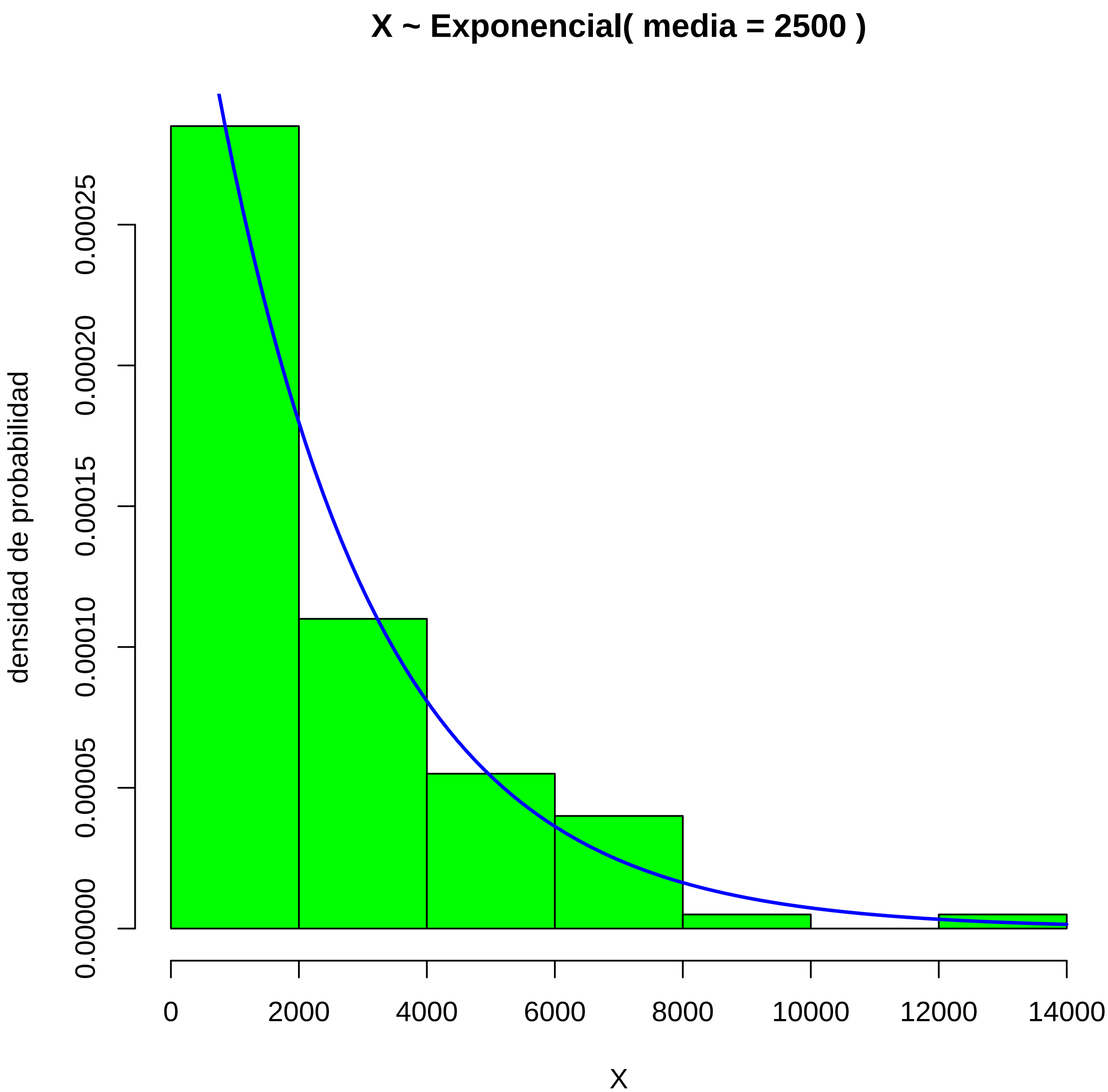


- **Ejemplo 4:**

Generar números aleatorios de una distribución exponencial. Por ejemplo, si la vida media de un bulbo de luz es 2500 horas, uno puede pensar que el tiempo de vida es aleatorio con una distribución exponencial que tiene media 2500. El único parámetro es la razón $= 1/\text{media}$.

```
# Definir el par?metro apropiado
media <- 2500; razon <- 1/media;n=100

# generar 100 n\'umeros aleatorios de la distribuci\'on
x = rexp(n, razon); x
```

```
##   [1]   901.95372   156.02655   314.37034   686.53587  2168.73472
##   [6] 12610.37500  1294.54967  5037.86508   790.48373  1459.89267
##  [11]   544.09025  1792.61143   623.46199  1082.83775  1880.64956
##  [16]  3750.12424   947.67923   171.12897  1378.74409  1827.84653
##  [21]  3514.88976   987.83332  4675.25531   125.15029  2301.93588
##  [26]  4880.12586  4294.12533   453.92363   615.54986  4105.06456
##  [31]  8990.30009  2009.22687  1283.46889  1985.75671   846.18787
##  [36]  7525.72266  1879.99209  1568.82495  4406.31317  3615.10507
##  [41]   710.79347  2806.97688  1207.28220   402.94989   510.31647
##  [46]    13.69918  3920.24207  3300.66863   841.73483  1199.69926
##  [51]   128.54102  1398.71862  2539.67809  1096.25286   595.05019
##  [56]  2913.33995  1608.20961  1941.50742   631.51822  5902.51131
##  [61]  6806.67411  2136.95224  4760.77787   967.67014  3757.59341
##  [66]   378.41600  4074.94374  4715.03219  1862.15400   890.88257
##  [71]    36.01121   732.54459   171.37922  7761.43933  2318.60703
##  [76]  1422.16634  1026.25208   847.59358  2294.66934  1247.00699
##  [81]  2777.53470  2710.36026  5869.33780  1807.00381  2265.43562
##  [86]  6032.54183  2236.81901   247.31747  2863.80081   417.88469
##  [91]  6051.81204  6161.00521   112.80848   835.06331   350.41069
##  [96]  2427.17627   466.93276  7923.93098  3617.30096  7741.74010

# Histograma para la nuestra aleatoria de tama\~no 100
hist(x, main="X ~ Exponencial( media = 2500 )", xlab="X",
     ylab="densidad de probabilidad", probability=TRUE, col="green")

# Graficar la funci?n de densidad, usando la funci?n curve()
curve(dexp(x, razon), col="blue", lwd=2, add=TRUE)
```


X ~ Exponencial(media = 2500)
densidad de probabilidad
0.00000
0.00005
0.00010
0.00015
0.00020
0.00025
0
2000
4000
6000
8000
10000
12000
14000
X

4. FUNCIONES DE DISTRIBUCIÓN Y SU INVERSA (LOS CUANTILES).

En R, las funciones a las que se les antepone una ”p”permiten contestar cu?l es la probabilidad de que una variable aleatoria X sea menor o igual que x, esto es F(x)=P[X<=x]. Las funciones a las que se les antepone una ”q”son lo inverso de esto, ellas permiten conocer qué valor de una variable aleatoria X corresponde a una probabilidad p dada. Esto es el cuantil Xq o punto en el que los datos son partidos, P[X<=x$_q$] $= p$.

- **Ejemplo 1:**

Para una Variable aleatoria X con distribución normal de media 1 y desviación estándar 1, cuál es la probabilidad de que sea menor que 0.7?

```
x <- 0.7
p <- pnorm(x, mean=1, sd=1, lower.tail = TRUE); p

## [1] 0.3820886
```

Observación: lower.tail=TRUE es el valor por defecto, para indicar las probabilidades son P[X<=x], en otro caso será P[X>x].

- **Ejemplo 2:**

Para una variable aleatoria con distribución normal estándar, encontrar P[Z<=0.7] y P[Z>0.7].

```
z <- 0.7
p1 <- pnorm(z, mean=0, sd=1); p1

## [1] 0.7580363

p2 <- pnorm(z, mean=0, sd=1, lower.tail=FALSE); p2

## [1] 0.2419637
```

Observación: ya que P[Z>0.7]=1-[Z<=0.7], obtenemos el mismo resultado con

```
p3 <- 1-pnorm(z, mean=0, sd=1);p3

## [1] 0.2419637
```

- **Ejemplo 3:**

Qué valor de una variable aleatoria con distribución normal estándar, tiene 75 % del ?rea a la izquierda?.

```
p <- 0.75
z <- qnorm(p, mean=0, sd=1, lower.tail = TRUE); z

## [1] 0.6744898
```

Observación: note que el valor de z que resuelve P[Z<=z]=0.75 es el tercer cuartil (Q3), esto es z=0.6744898.

- **Ejemplo 4:**

Cuál es la probabilidad a la derecha de 18.55 para una Variable aleatoria X con distribución Chi-cuadrado de 12 grados de libertad?

```
x <- 18.55; gl <- 12
p <- pchisq(x, gl, lower.tail = FALSE); p

## [1] 0.09998251
```