UNIVERSIDAD DE EL SALVADOR
FACULTAD MULTIDISCIPLINARIA DE OCCIDENTE
DEPARTAMENTO DE MATEMÁTICA

Licenciatura en Estadística

Control Estadistico del Paquete R

"UNIDAD TRES"
Práctica 15 - Distribuciones de probabilidad continuas.

Alumna:
Martha Yoana Medina Sánchez

Fecha de elaboración
Santa Ana - 27 de noviembre de 2015

1. DISTRIBUCIONES CONTINUAS.

a) Distribucíon uniforme.

- **Parámetros.**

x = valor cualquiera
p = probabilidad
n = tamaño de la muestra
min = valor mínimo
max = 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) Distribucíon Normal.

- **Parámetros.**

x = valor cualquiera
p = probabilidad
mean = media
sd = desviaciÃ³*ntípica*
n = 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) Distribucíon 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) Distribucíon Chi-cuadrado.

- **Parámetros.**

$$\begin{gathered}\text{x} = \text{valor cualquiera} \\ \text{df} = \text{grados de libertad} \\ \text{p} = \text{probabilidad}\end{gathered}$$

- **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) Distribucíon F de Snedecor.

- **Parámetros.**

$$\begin{gathered}\text{x,q} = \text{vector cuantiles} \\ \text{df1} = \text{grados de libertad en el numerador} \\ \text{df2} = \text{grados de libertad en el denominador} \\ \text{p} = \text{vector probabilidad}\end{gathered}$$

- **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) Distribucíon Exponencial.

- **Parámetros.**

$$\begin{gathered}\text{x,q} = \text{vector cuantiles} \\ \text{rate} = \text{razón} = 1/\text{E}[\text{X}] \\ \text{p} = \text{vector probabilidad} \\ \text{lower.tail} = \text{T}\end{gathered}$$

- **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 está 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, serán 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 calculará pues el percentil 80 de la variable X e Y,utilizando los cuantiles-normales para la variable X:

```r
# 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
```

```r
#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):

```r
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 años 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 distribución es la única distribución continua no tiene memoria resulta que P(X>=8/X>=4)=P(X>=4)=0.5647182

```
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:

```
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
```

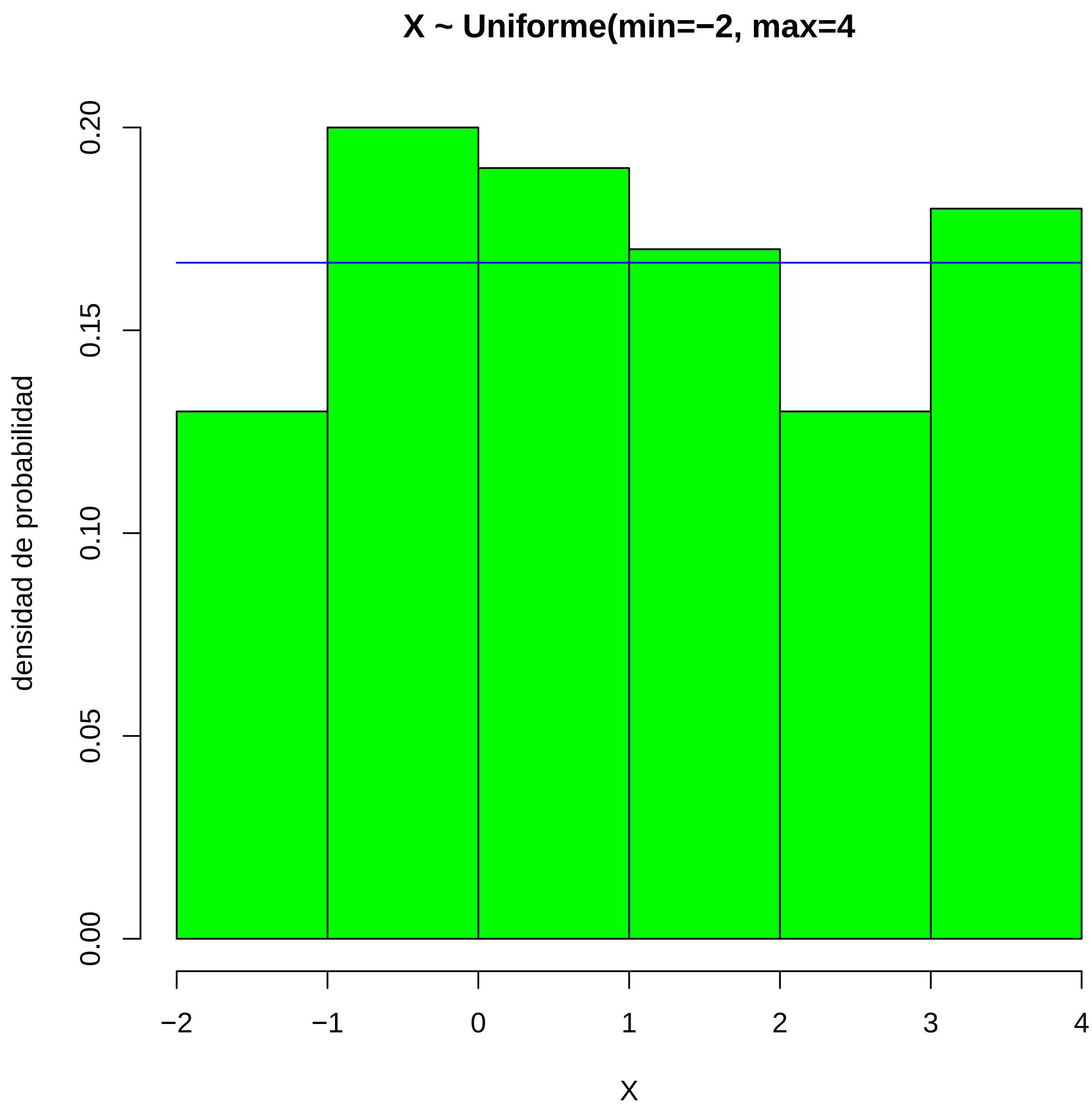
```r
# o de esta otra manera

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]  1.93536984  0.48794002  3.16088643 -1.98006110 -0.12364830
##   [6]  0.36497971 -0.36620827  1.10462600 -0.53854729 -0.31606691
##  [11]  2.44961394  2.54016676 -0.19798646 -1.11801856  2.38775964
##  [16] -0.37759314 -0.87407659  0.98529517 -0.62860891  3.24161901
##  [21] -1.42665437  1.55476034  1.72200572  1.69594416  1.18840346
##  [26] -0.58345596  0.83782765  2.08484564 -0.58484657  1.40401472
##  [31]  0.73504179  3.68171236  0.36219574  1.14706969 -0.45818063
##  [36]  1.02862697 -0.42317659  1.13967025  1.70840789  0.81792576
##  [41]  3.00270698 -1.28102504  3.75752606  3.21890690 -1.23027636
##  [46]  3.63390102  3.88954588  0.62900457  1.01934674  1.53893361
##  [51]  1.82253061 -0.32833368  2.30456683 -0.37287721 -1.35619000
##  [56]  3.66513973  0.28910594 -0.13040638 -1.22198929  2.41731965
##  [61] -1.26216745 -0.31702129  0.71153323  0.47789773  3.54506094
##  [66]  1.40899352  3.49292276  3.06245976  2.84756282  3.55149454
##  [71] -1.68034053  3.10631353  2.64125899 -0.72169913  3.28513325
##  [76] -0.72832528  1.42706489  2.56976939  0.96727307 -0.54222676
##  [81]  0.49741977  0.84228006  3.84567078  3.79354362 -1.88182253
##  [86]  2.71978739 -1.04190263  1.34329914  0.47078387  0.02234774
##  [91] -1.41659411 -0.79560201  2.56052075  0.64063966 -1.59104959
##  [96]  2.12727118  2.12708346  0.64837956  3.33574958  0.59836524

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

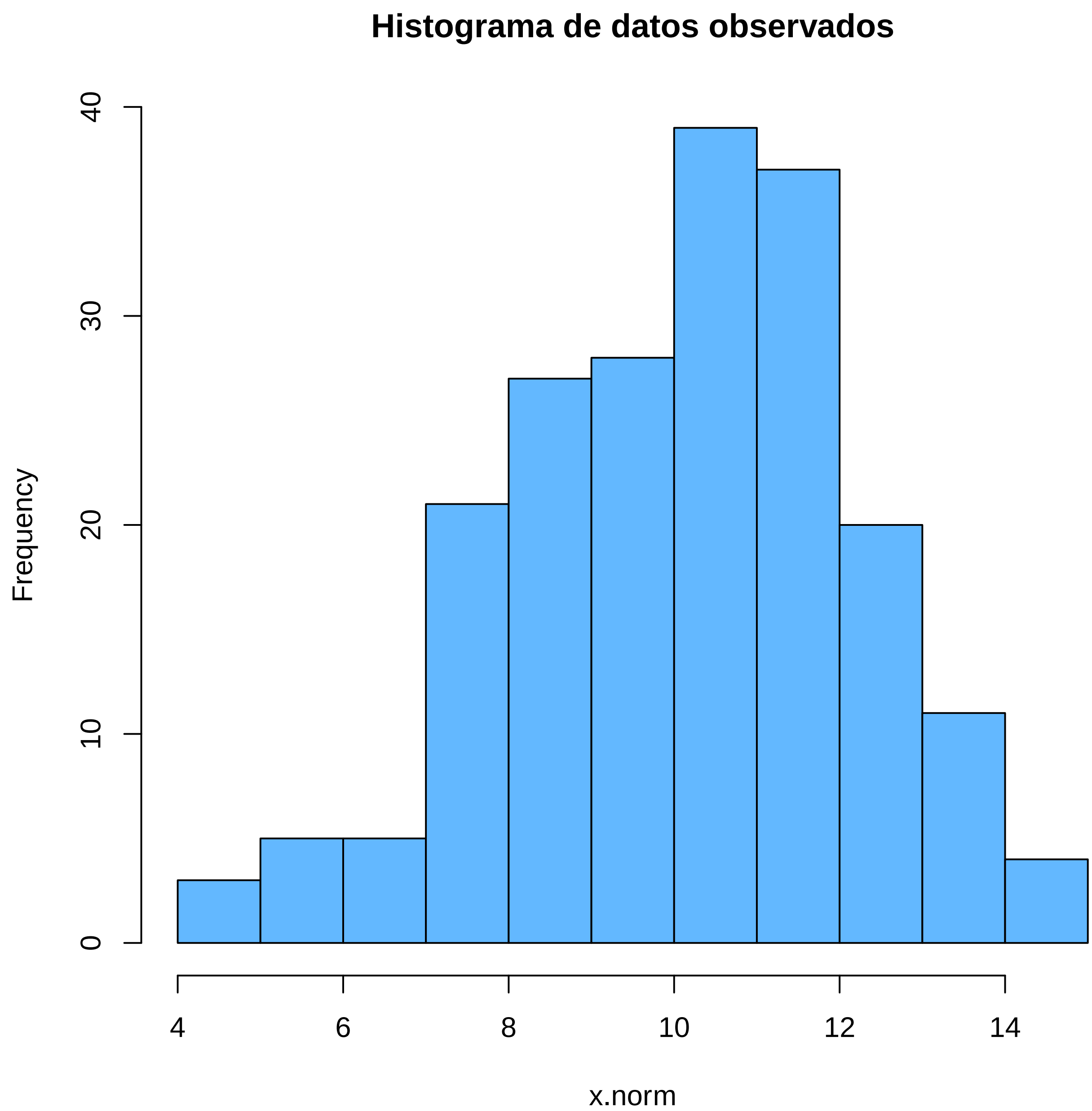


- **Ejemplo 2:**

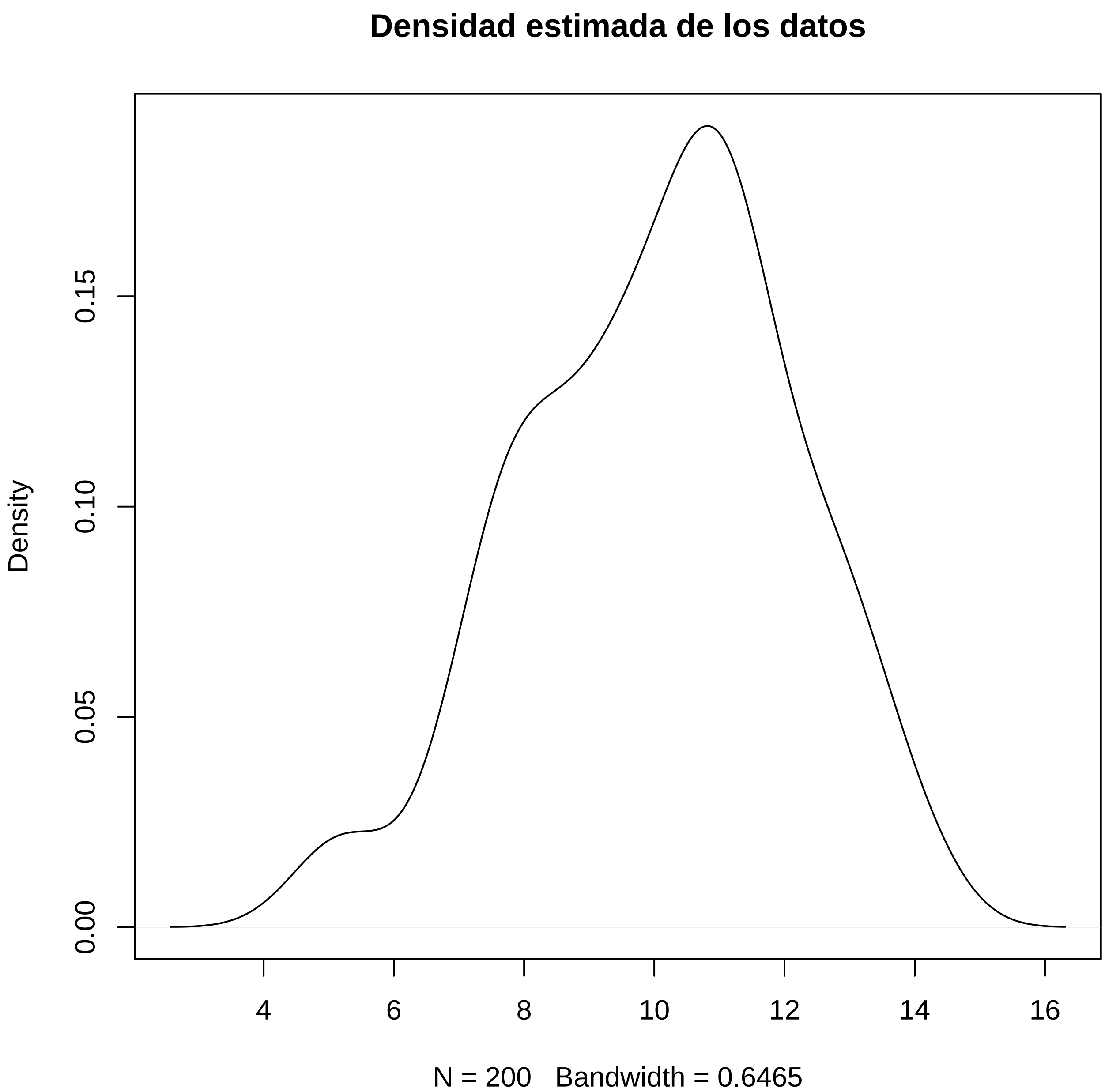
Supongamos que tenemos una muestra de tamaño n=200 perteneciente a una población normal N(10,2) con media igual a 10 y con una desviación estandar igual a 2:

```
# 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,
```

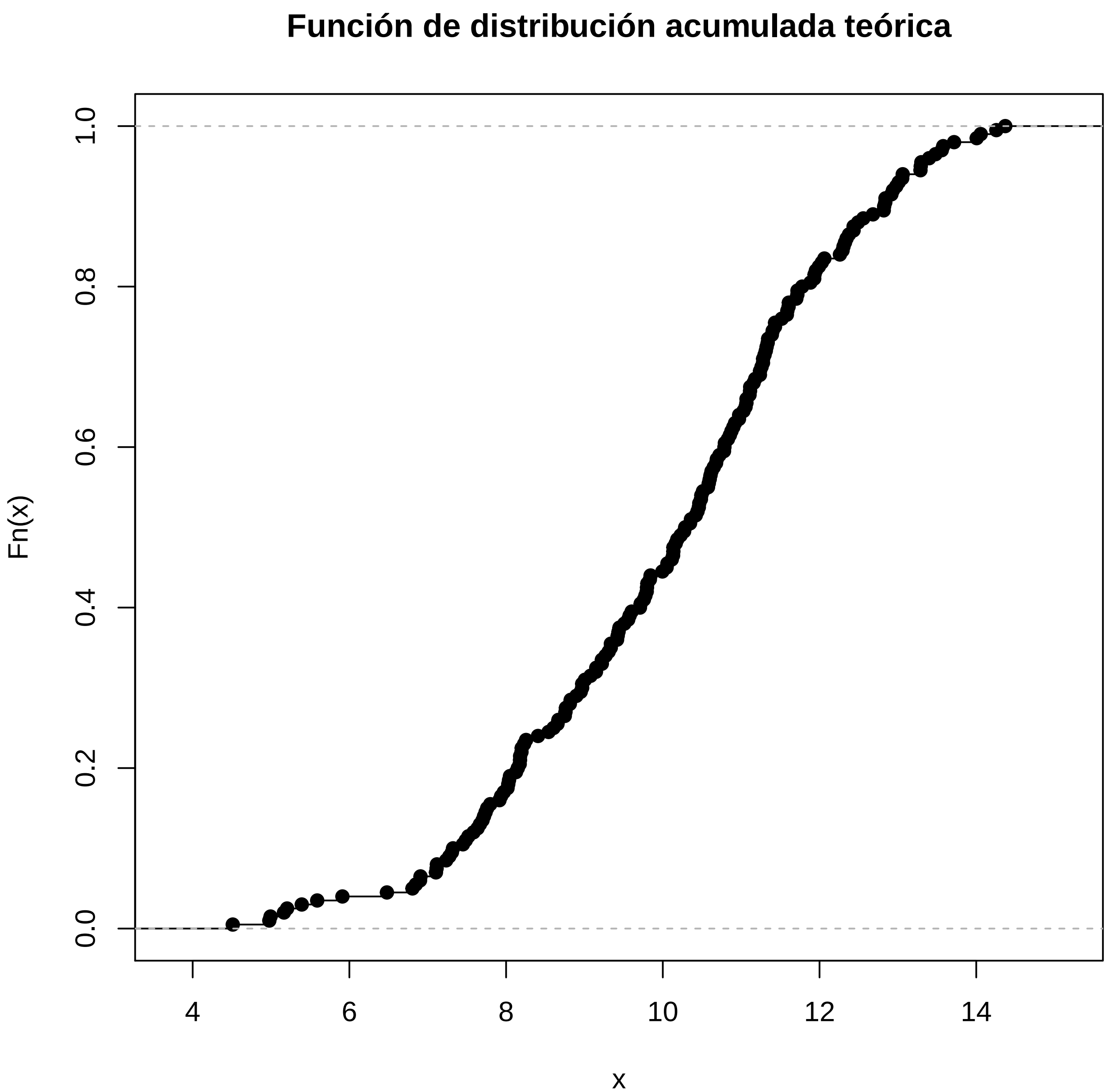
```
angle = 45, col = "steelblue1", border = NULL,
main = "Histograma de datos observados", axes = TRUE,
plot = TRUE, labels = FALSE)
```




```
# Podemos estimar la densidad de frecuencia usando la funci\'on
# density() y plot() para dibujar su
"grÃ¡fica"

## [1] "grÃ¡fica"

plot(density(x.norm), main="Densidad estimada de los datos")
```




```
# R permite calcular la funci\'on de distribuci\'on acumulada te\'orica con ecdf()
plot(ecdf(x.norm),main="FunciÃ³n de distribuciÃ³n acumulada teÃ³rica")
```

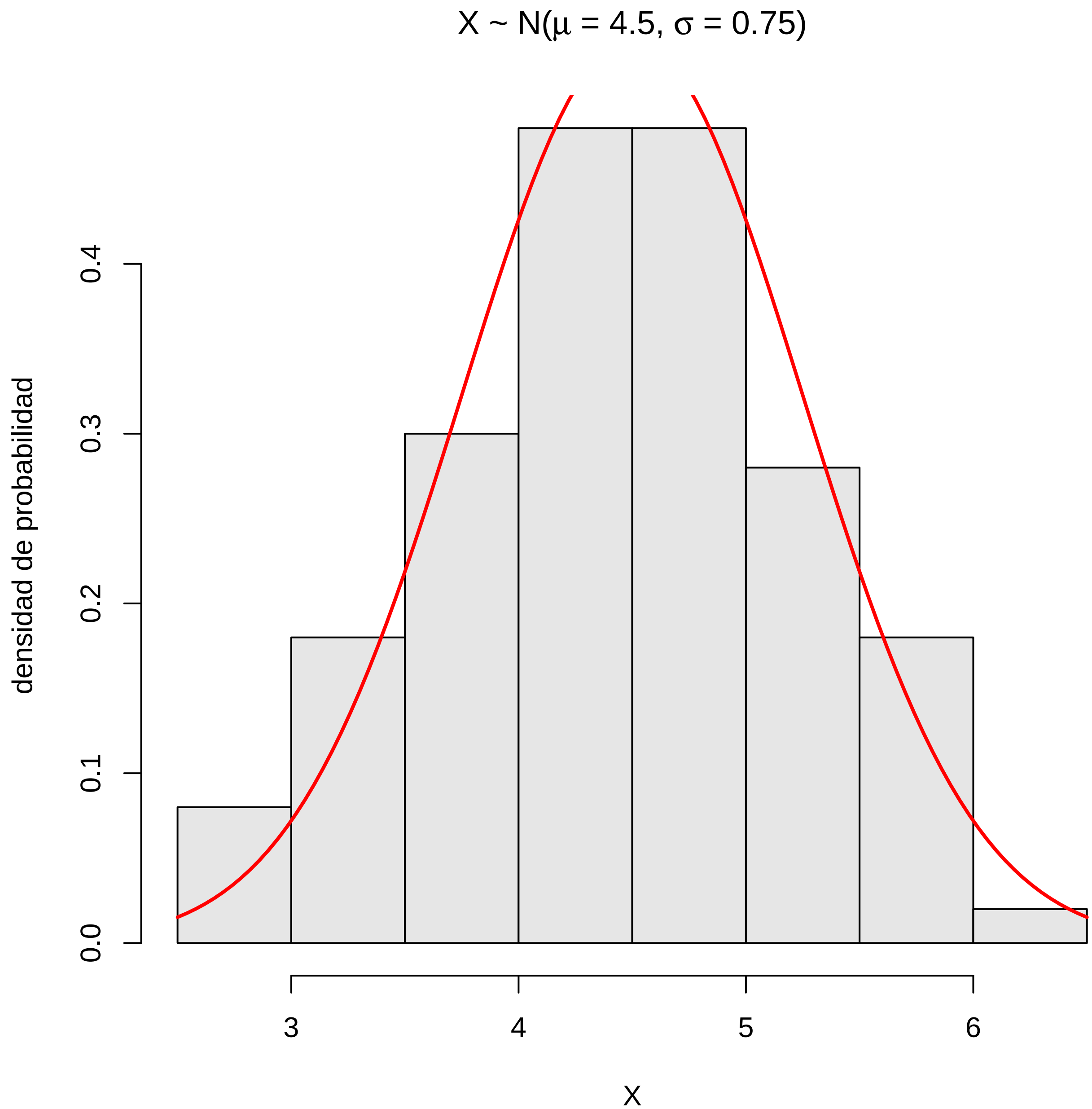


- **Ejemplo 3:**

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] 4.685023 4.085347 4.690376 3.716264 3.538028 4.245896 4.293525
```

```
##   [8] 5.005417 6.363610 4.221063 3.748988 2.983803 4.364624 5.431396
##  [15] 5.469190 4.107875 4.119808 4.043912 3.405542 3.949332 3.828855
##  [22] 5.816811 4.922200 4.774044 4.708622 5.502216 3.998810 5.457184
##  [29] 5.450010 3.789861 4.658435 5.258724 4.371644 4.581935 5.365355
##  [36] 5.777487 2.846609 4.955483 4.981274 4.495799 5.188280 3.793359
##  [43] 3.649243 5.562281 4.093415 4.537764 3.140547 4.435097 3.845908
##  [50] 5.309575 4.703681 4.852680 3.921677 3.118858 4.686166 5.888000
##  [57] 2.917225 4.563399 4.727328 4.074747 4.468296 4.893315 4.260960
##  [64] 5.775534 3.290102 3.280047 4.348795 3.478990 4.848428 5.048595
##  [71] 3.997954 5.127976 4.873776 4.185929 4.733015 4.780149 5.106116
##  [78] 4.343931 4.324382 5.457152 2.985052 3.838946 4.374825 4.796517
##  [85] 4.332632 3.491315 3.624708 3.992557 3.403156 4.329689 4.838943
##  [92] 4.420180 3.375388 5.567997 5.907756 5.577851 5.023092 4.426826
##  [99] 4.743798 4.627503

# 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)
```

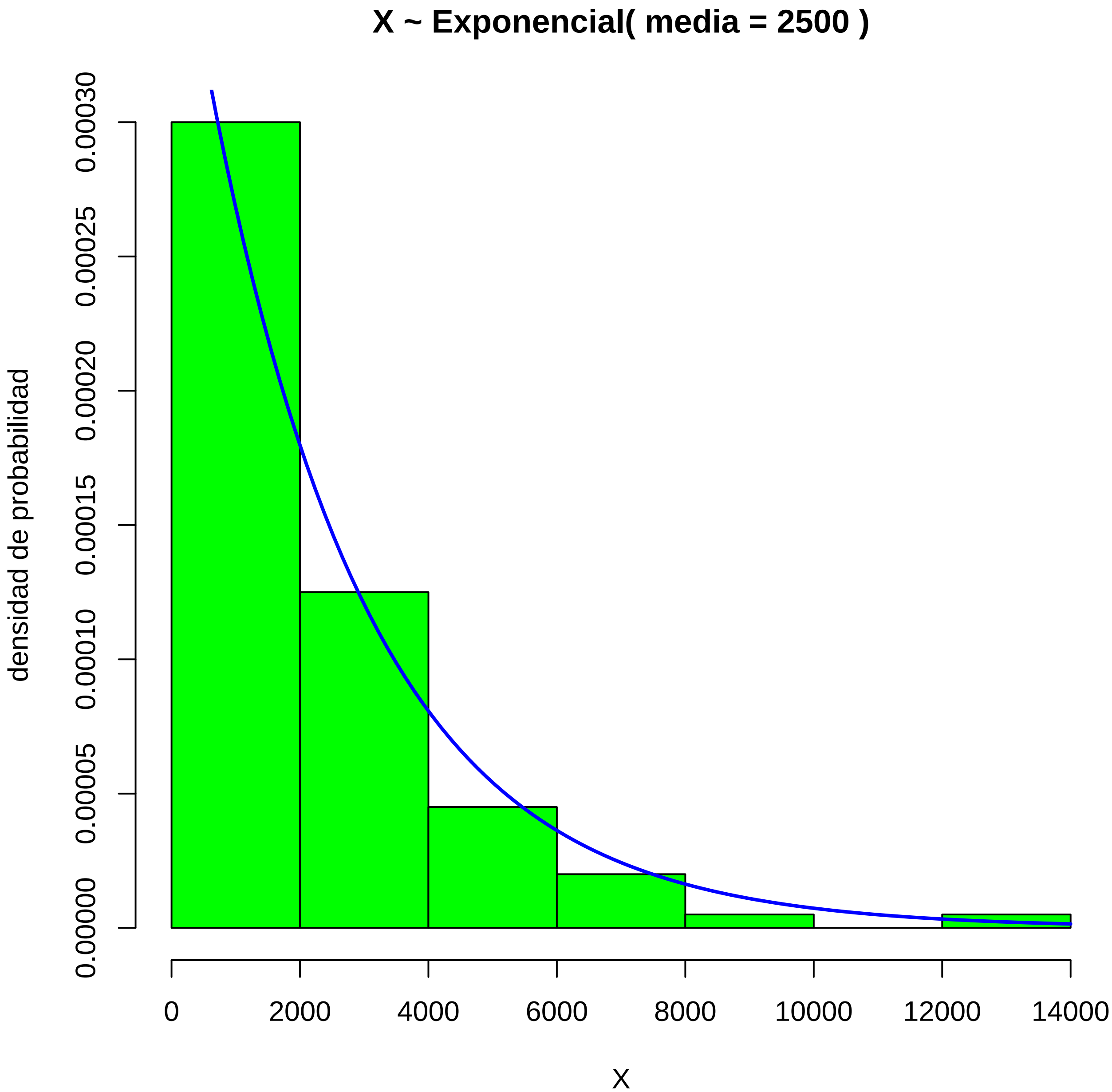


- **Ejemplo 4:**

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/media.

```
# Definir el par\'ametro apropiado
media <- 2500; razon <- 1/media;n=100

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

```
##   [1]  1497.05352  2525.29060  5840.21211  1615.22462  1235.99915
##   [6]  1639.14642  2325.65573  1006.30196  4043.57293   240.55050
##  [11]  2379.63817   902.40007  6424.37764    82.32148  1272.07182
##  [16]  6219.89923  1264.66282  2233.16240  3766.28769  1711.45686
##  [21]   245.22321  4045.29514   299.87246  2378.78900  2387.13219
##  [26]  3349.28066  1365.44929  1915.70584  1144.70006  5624.30359
##  [31]  7427.56721  3022.84775  2382.77466  5968.15114   119.00098
##  [36]   243.42030   558.75517  1920.82684  2966.77171  3632.56684
##  [41]  4580.18906  9928.03359   700.47084  3658.40637   510.15105
##  [46]  1152.04048   579.53950   296.50453  1926.53039  1508.98988
##  [51]  1420.28350  4982.12799  2941.88125   556.30561  3199.28647
##  [56]   825.67994   624.48165  1927.20463  2783.47374  5258.49014
##  [61]   716.79397   942.67333  2638.15169  1649.51976   749.34006
##  [66]  2221.83473 12281.61486   920.61243  1902.75934   976.64014
##  [71]  1381.90697  3352.22399  2490.22503  1354.78209  1554.19017
##  [76]  1012.72515  1755.05837  5745.37882  1138.64851  1708.48026
##  [81]  1065.43021  1308.05210  6920.78363  1333.33177  1435.59109
##  [86]  1334.11109   910.88817  2546.68384   439.87444  2670.88157
##  [91]  2279.90290  1376.39155  3669.47212  1561.55536    74.15215
##  [96]   412.24977    26.54866  1090.88679  3123.18052  1047.65997

# 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\'on de densidad, usando la funci\'on 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
0.00030
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 cuál 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 área a la izquierda?.

```r
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?

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

## [1] 0.09998251
```