UNIVERSIDAD DE EL SALVADOR

FACULTAD MULTIDISCIPLINARIA DE OCCIDENTE

DEPARTAMENTO DE MATEMÁTICAS

PRACTICA 15 DE R

CARRERA:

LICENCIATURA EN ESTADÍSTICA

DOCENTE:

LIC. JAIME ISAAC PEÑA

PRESENTADO POR:

CRISTIAN ALBERTO ZALDAÑA ALVARADO

Índice

1. 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 relación continúe hasta la siguiente cita.

```
> x <- 55
> a=0
> b <- 90
> #usando la función 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 $\mathrm{P}(15 < \mathrm{X} < 55) = \mathrm{P}(\mathrm{X} < 55) - \mathrm{P}(\mathrm{X} \leq 15) = 0{,}6111 - 0{,}1666 = 0{,}4445$,

```
> 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
  $\mathrm{P}(\mathrm{X} \geq \mathrm{P80} \cap \mathrm{Y} \geq \mathrm{P80}) = \mathrm{P}(\mathrm{X} \geq \mathrm{P80})\mathrm{P}(\mathrm{Y} \geq \mathrm{P80}) = (0{,}20)(0{,}20) = 0{,}04$

  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:

```
> #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, la media muestral, sigue otra distribución normal de media igual que la poblacional y desviación típica $\sigma/\sqrt{\mathrm{n}}$.

Como se desea calcular $\mathrm{P}(\overline{\mathrm{x}} \geq 4{,}5)$:

```
> 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) Suponiendo que la variable $X =$ "tiempo de funcionamiento del marcapasos" sigue una distribución exponencial con parámetro $\lambda = 1/\theta = 1/7$ con $\theta = \mathrm{E}[\mathrm{X}]$ tiempo promedio.

La probabilidad $\mathrm{P}(\mathrm{X} \geq 5)$ se obtiene así:

```
> x <- 5
> teta=7
> pexp(x, rate=1/teta, lower.tail=FALSE)
[1] 0.4895417
```

y de igual forma $\mathrm{P}(\mathrm{X} < 3)$:

```
> 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 $\mathrm{P}(\mathrm{X} \geq 8/\mathrm{X} \geq 4)$

Teniendo en cuenta que la función de distribución es la única distribución continua no tiene memoria resulta que $\mathrm{P}(\mathrm{X} \geq 8/\mathrm{X} \geq 4) = \mathrm{P}(\mathrm{X} \geq 4) = 0{,}5647182$

```
> 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 años.
```

d) Calcular el valor que deben tener a y b para que $P(X < a) = 0{,}5$ y $P(X > b) = 0{,}32$

De forma análoga al apartado anterior, en el primer caso habría que calcular la mediana (percentil 50), $a = 4{,}852$,

```
> qexp(0.5, rate=1/teta, lower.tail=TRUE)
[1] 4.85203
> #y en el segundo caso, el percentil 68, b = 7.97
> qexp(0.68, rate=1/teta, lower.tail=TRUE)
[1] 7.97604
> #o de esta otra manera
> qexp(0.32, rate=1/teta, lower.tail=FALSE)
[1] 7.97604
```

2. GENERACIÓN DE MUESTRAS ALEATORIAS DE LAS DISTRIBUCIONES

- **Ejemplo 1:**

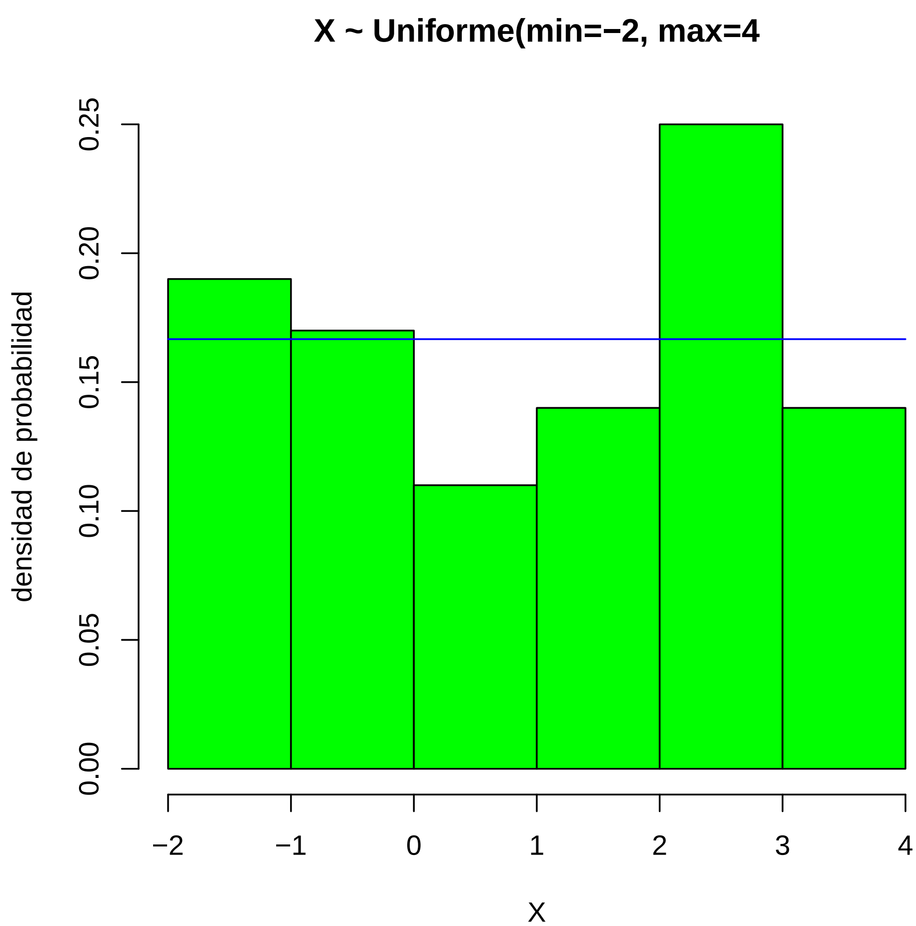
  Generar 100 números aleatorios de una distribución Uniforme en $[-2, 4]$

```
> # Definir los parámetros apropiados
> min <- -2
> max <- 4
> # generar 100 números aleatorios de la distribución
> x = runif(100, min, max); x

  [1]  1.72441885 -1.20258820  2.94122130  3.86366763  2.26986600  2.03473099
  [7]  3.85604805  0.35721927  2.38114040 -0.81553738  2.75105433 -0.33104198
 [13]  2.01224455 -1.08331477  2.13995466  1.56912292  0.30935912  2.14845742
 [19] -0.75401528  2.47878468 -0.63142080 -1.97880560 -0.23055009  2.45344316
 [25] -0.81626618 -0.61981848  3.51962568  3.98986428  0.58031888 -0.05194944
 [31] -1.96741157  0.11994122  1.28991166 -1.61120638  1.92312506  3.75927894
 [37] -1.00149592  2.93107260  1.15787231  1.15737447 -0.39425610 -0.75807252
 [43]  3.32416778  3.62984372  0.03678014  2.82683221  0.35284360  2.59191564
 [49] -1.74496071 -1.21872943  2.40683692  1.13321535 -0.35055926  3.07568893
 [55] -0.53693375  0.11833417  2.26976201  3.39550342  3.90890589 -1.68051258
 [61]  1.07135191  2.66103541 -1.25917957 -1.46222713  1.31241707 -1.69865095
 [67] -1.42791094  2.73738033 -0.71767115 -1.20985951  1.72886385  0.62485700
```

```
[73]  3.92436151 -0.27416710 -0.70545266  3.06295948  2.76390103  3.55843149
[79] -1.83412204  0.78096045  2.16877379  0.93649839  2.12383581  1.42440714
[85]  2.07495609 -1.66591534 -0.59883201  1.34923929  2.79024076 -0.17465591
[91] -1.96548955  2.76086806  1.98413153  2.37679794  1.42089522  2.20179863
[97]  3.01069469 -1.96143564  0.20406957 -1.71320652
```

```
> # Histograma para la nuestra aleatoria de tamaño 100
> hist(x, main="X ~ Uniforme(min=-2, max=4", xlab="X",
+      ylab="densidad de probabilidad",
+ probability=TRUE, col="green")
> # Graficar la función de densidad, use la función curve() para variable continua
> curve(dunif(x, min, max), col="blue", add=TRUE)
```

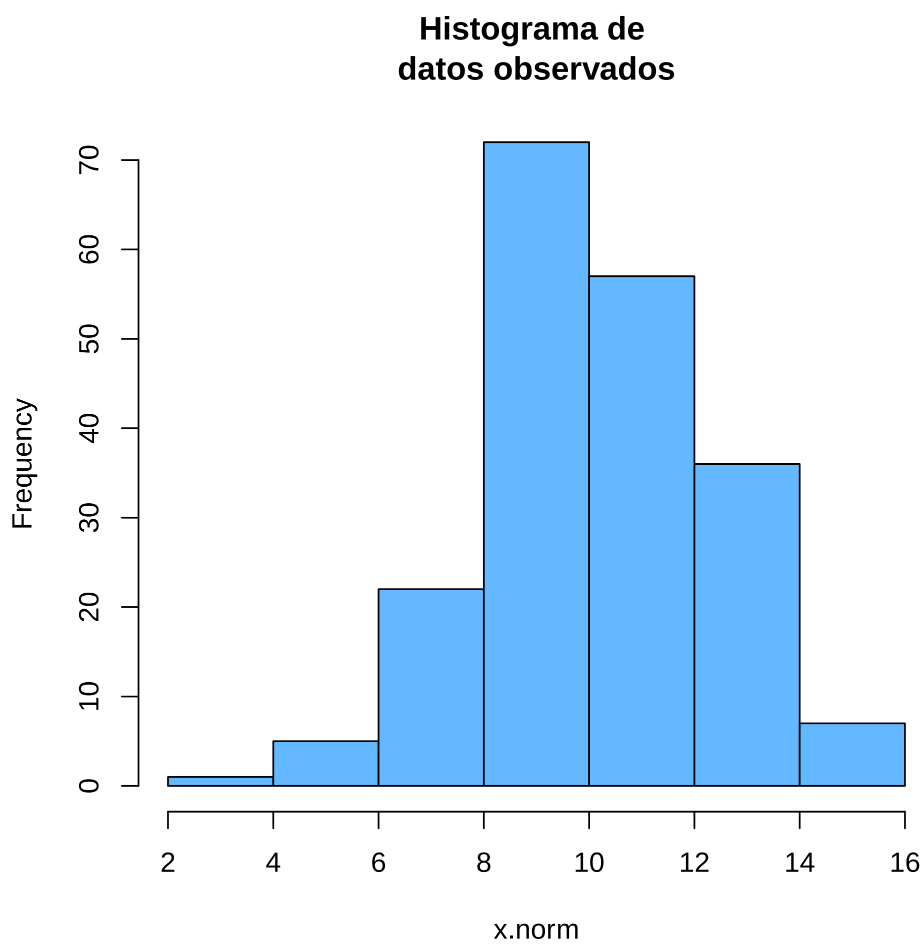


- **Ejemplo 2:**

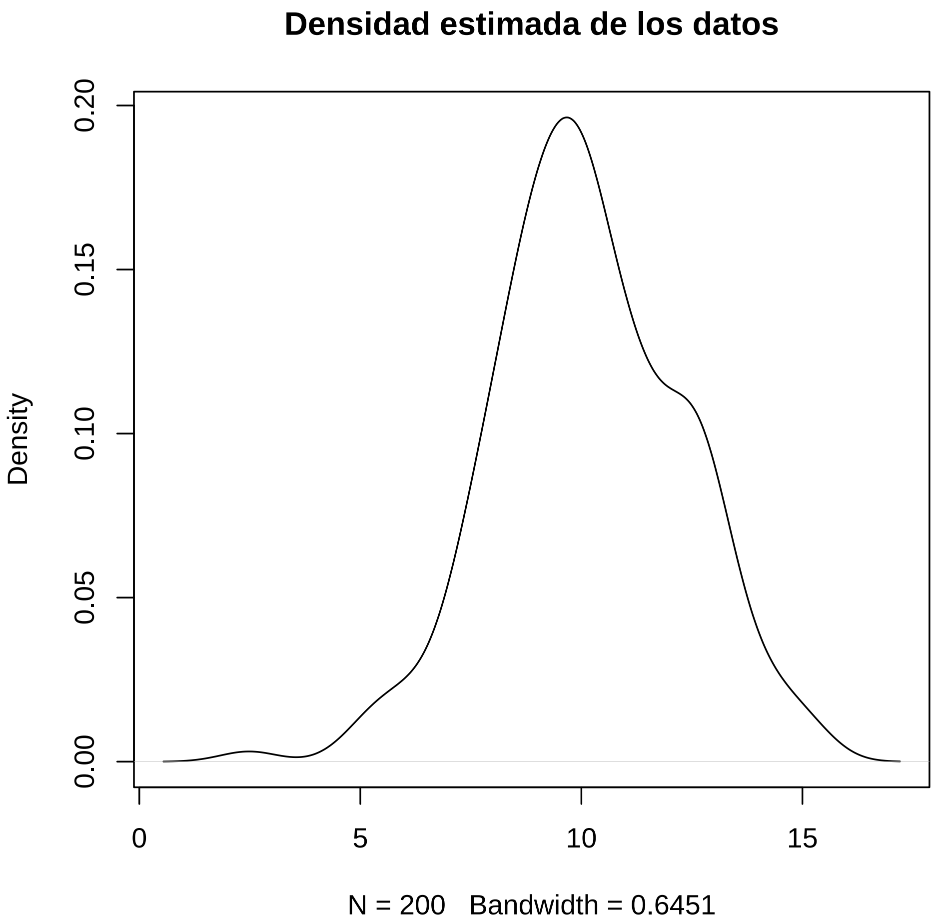
  Supongamos que tenemos una muestra de tamaño n = 200 perteneciente a una población normal $N(10, 2)$ con $\mu = 10$ y $\sigma = 2$:

```
> #genera los valores aleatorios de la distribución
> x.norm <- rnorm(n=200,mean=10, sd=2)
> # Podemos obtener un histograma usando la función hist()
> hist(x.norm, breaks = "Sturges", freq = TRUE, probability = FALSE,
```

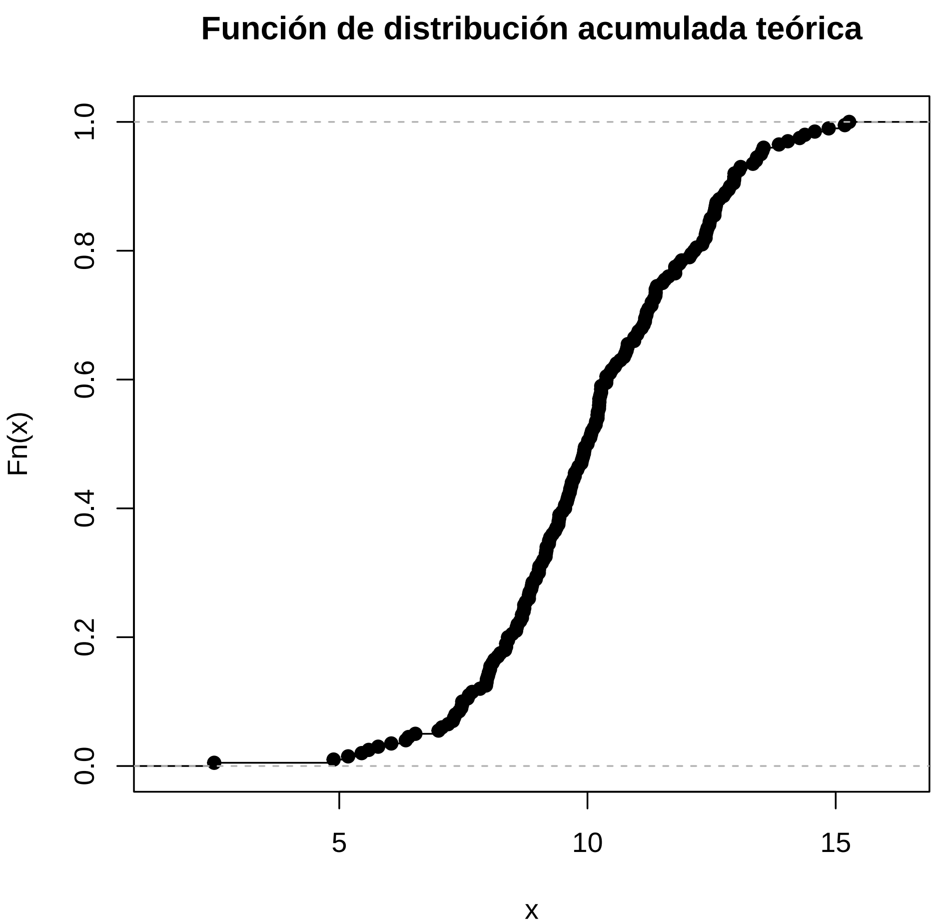
```
+       include.lowest = TRUE, right = TRUE, density = NULL, angle = 45,
+       col = "steelblue1", border = NULL, main = "Histograma de
+ datos observados", axes = TRUE, plot = TRUE, labels = FALSE)
>
>
```




```
> # Podemos estimar la densidad de frecuencia usando la función density()
> # y plot() para dibujar su gráfica
> plot(density(x.norm), main="Densidad estimada de los datos")
>
```




```
> # R permite calcular la función de distribución acumulada teórica con ecdf()
> plot(ecdf(x.norm),main="Función de distribución acumulada teórica")
```



- **Ejemplo 3:**

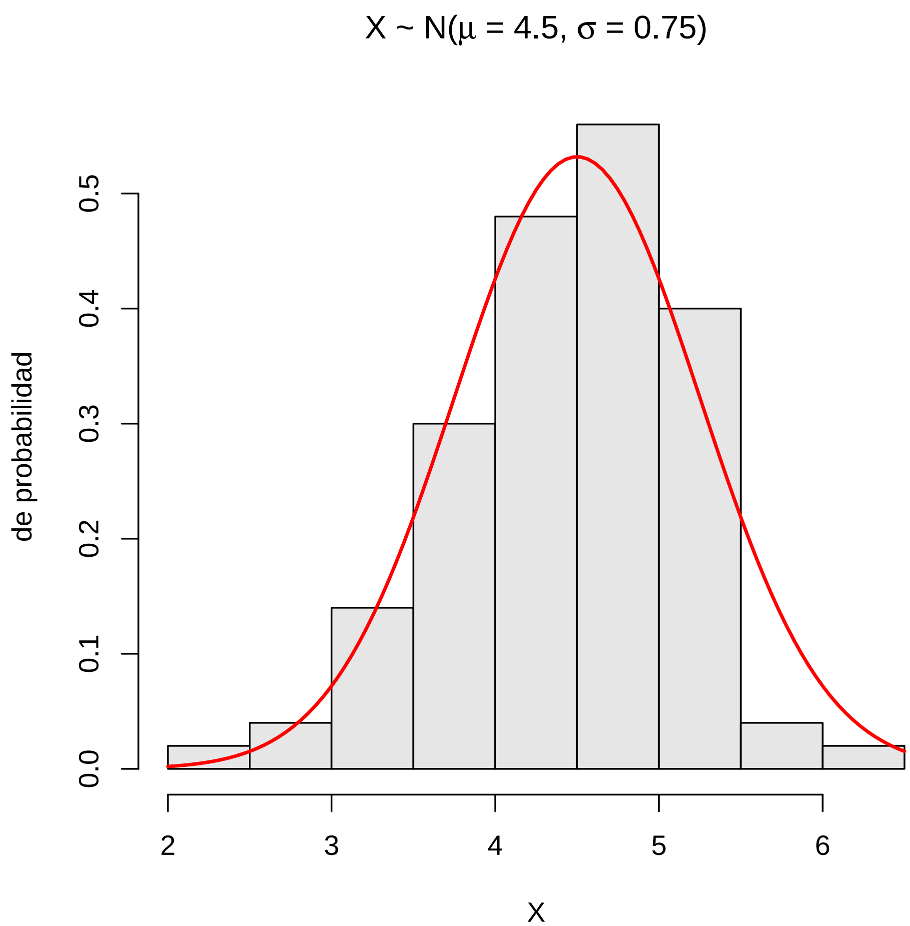
  Generar 100 números aleatorios de una distribución Normal con media 4.5 y desviación estándar 0.75

```
> # Definir los parámetros apropiados
> media <- 4.5
> desviacion <- 0.75
> # generar 100 números aleatorios de la distribución
> x = rnorm(100, media, desviacion); x

  [1] 4.098186 2.415429 5.125352 4.560413 3.968964 4.115018 5.019342 5.212775
  [9] 5.425192 3.176728 4.749502 3.951352 4.846890 5.984273 4.588930 3.286113
 [17] 5.143580 4.034081 4.044894 4.865529 3.924607 3.345973 4.919745 3.913564
 [25] 4.117746 5.029937 4.990285 4.203255 5.236550 3.972723 4.854583 3.924383
 [33] 4.552671 5.022399 3.407094 4.840961 4.575223 5.411974 5.111481 4.047605
 [41] 4.604261 4.654364 5.066934 4.493829 3.280563 5.293406 4.686285 5.566275
 [49] 4.641267 4.789208 4.552209 4.188631 5.045934 4.429961 4.547115 3.918350
 [57] 2.785608 2.668780 4.119112 4.092977 3.695361 4.196763 4.492813 4.958818
 [65] 4.737061 3.513747 4.665086 4.153953 5.468165 3.865902 4.914628 4.009492
 [73] 6.062746 5.151847 3.353902 4.474467 4.388195 4.653558 3.667670 4.661470
 [81] 4.674436 4.197299 5.048429 3.891968 5.253582 3.981630 4.440418 5.257119
```

```
 [89] 4.608504 4.748987 4.266600 4.397642 4.672934 5.354054 4.366479 3.618222
 [97] 3.481835 5.051597 3.712679 4.245569
```

```
> # Histograma para la nuestra aleatoria de tamaño 100
> hist(x,main=expression(paste("X ~ N(", mu, " = 4.5, ", sigma, " = 0.75)")), xlab="X", ylab="densi
+ de probabilidad", probability=TRUE, col=gray(0.9))
> # Graficar la función de densidad teórica, usando la función curve()
> curve(dnorm(x, media, desviacion), col="red", lwd=2, add=TRUE)
```



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.

- **Ejemplo 4:**

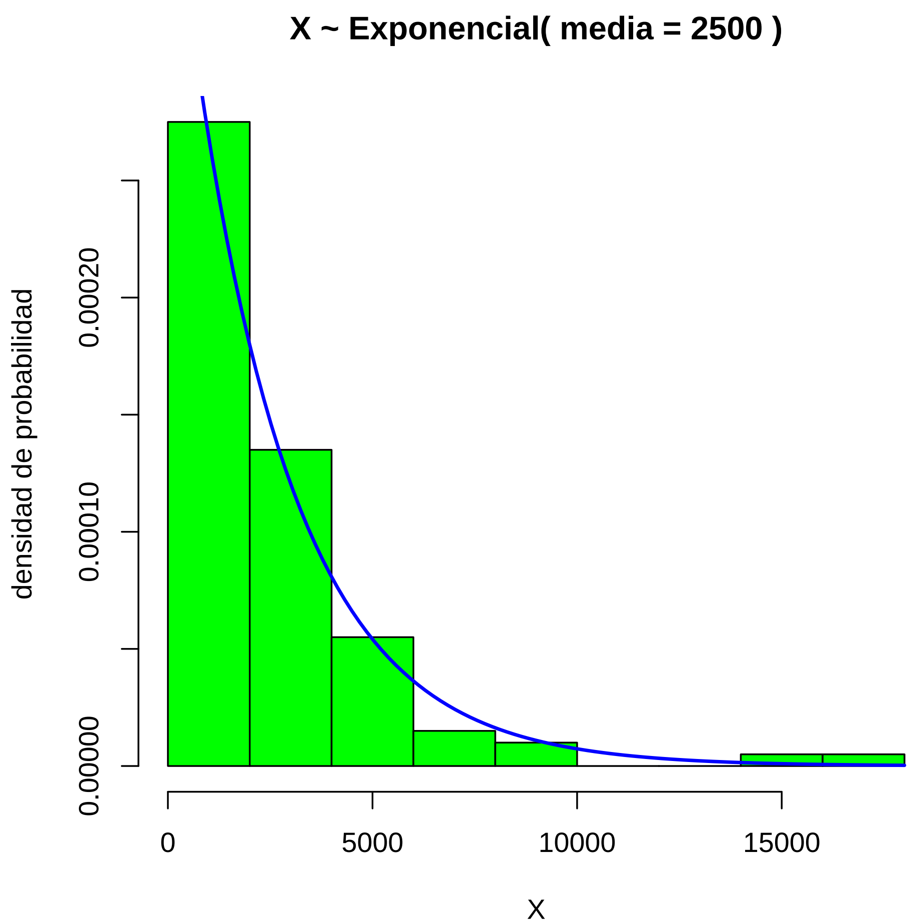
  Generar números aleatorios de una distribución exponencial. Por ejemplo, si la vida media de un bulbo de luz es 2500 horas, uno puede pensar que el tiempo de vida es aleatorio con una distribución exponencial que tiene media 2500. El único parámetro es la razón $= 1/$media.

```
> # Definir el parámetro apropiado
> media <- 2500
> razon <- 1/media
> n=100
```

```
> # generar 100 números aleatorios de la distribución
> x = rexp(n, razon); x

  [1]   442.19691  1257.42291   402.76074  2639.11575   512.59096  3251.74466
  [7] 14408.07396  3248.40159  2821.08092  2178.21193  1367.72952  1384.97707
 [13]  3606.70245   229.77185   723.75281  2679.02411   157.37637   185.53403
 [19]  4095.45979  1214.35516  2690.38725  2387.47888    37.23398  1274.29374
 [25]   719.75545  6265.67775    94.01990   507.84745  4053.12123   423.50616
 [31]  1858.04681  4000.46314  1724.23076   819.23541  5117.93671   649.95663
 [37]  7619.19836  4139.14032  1932.52518  8380.79464  3318.90410   633.70002
 [43]  1660.18852  2131.37916  1588.77678  1159.43023 16260.96231  1432.86499
 [49]   981.50666   469.56988  3215.45098  1038.82091   389.41416  4806.26117
 [55]   426.16676   117.58206  2179.37403  9568.73966   244.33537  1807.55812
 [61]  3353.20186  4279.27543  3322.73148  2273.87650  1934.41521   840.25948
 [67]  2245.71587  2314.37217  3859.66894  1964.95166  5074.47619   894.51077
 [73]  1005.49277  1206.72647  3791.51171  1721.08224   663.60683  7133.50913
 [79]   247.00546   268.80774  5921.67518  1636.96705   165.09792  5245.63163
 [85]   637.24835   148.88069  1358.44872  3066.66927  2665.94710   450.80270
 [91]   287.92623  3346.26737  5132.17208  2982.86455   305.03012   493.73296
 [97]  1233.61203  2182.00770  2639.81463  2272.88139

> # Histograma para la nuestra aleatoria de tamaño 100
> hist(x, main="X ~ Exponencial( media = 2500 )", xlab="X",
+      ylab="densidad de probabilidad",
+ probability=TRUE, col="green")
> # Graficar la función de densidad, usando la función curve()
> curve(dexp(x, razon), col="blue", lwd=2, add=TRUE)
```



3. 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 $\mathrm{F(x) = P[X \leq x]}$. 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 X_q o punto en el que los datos son partidos, $\mathrm{P\,[X \leq X_q] = p}$

- **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 $\mathrm{P[X \leq X]}$, en otro caso será $\mathrm{P[X > x]}$

- **Ejemplo 2:** Para una variable aleatoria con distribución normal estándar, encontrar P[Z $\leq$ 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 $P[Z > 0{,}7] = 1 - P[Z \leq 0{,}7]$, 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?

```
> 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 $P[Z \leq z] = 0{,}75$ 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
```