
Curso OpenCourseWare

**Aprendizaje del Software Estadístico R: un entorno
para simulación y computación estadística**

Alberto Muñoz García

9. Funciones para la exploración de datos



Funciones para estadísticas simples

or	Correlación (admite uno o dos argumentos)
cumsum	Suma acumulativa de un vector
mean	Media aritmética
median	El percentil 0.5: la mediana
min	El mínimo de una serie de números
max	El máximo de una serie de números
prod	El producto de los elementos de un vector
quantile	Los percentiles de una distribución
range	Mínimo y máximo de un vector
sample	Muestreo aleatorio (y permutaciones)
sum	Suma aritmética
var	Varianza y covarianza
summary	Resumen de estadísticas de una serie de datos

Veamos algunos ejemplos:

Estadísticas simples:

```
> x = seq(1:10)
```

```
> x
```

```
[1] 1 2 3 4 5 6 7 8 9 10
```

```
> cumsum(x)
```

```
[1] 1 3 6 10 15 21 28 36 45 55
```

```
> median(x)
```

```
[1] 5.5
```

Generación de 100 datos de una normal estandar, y un sumario de los mismos:

```
> x = rnorm(100)
> summary(x)
  Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
-2.18000 -0.71000 -0.01041  0.05584  0.73730  2.76500
```

Generación de dos muestras correladas, y cálculo de la correlación:

```
> x = seq(1:10)
> y = 2*x+rnorm(10)
> cor(x,y)
[1] 0.9817436
```

Creación de permutaciones:

```
> sample(10)
[1] 5 6 1 4 9 2 8 10 7 3
> sample(10)
[1] 9 1 3 10 7 2 5 4 8 6
```

Muestreo sin repetición:

```
> sample(1:10,5)
[1] 7 8 2 4 6
```

Muestreo con repetición:

```
> sample(1:10,5,rep=T)
[1] 6 10 6 1 6
```

Simulación de 10 tiradas de un dado equilibrado:

```
> sample(1:6,10,rep=T)
```

Cálculo de algunos percentiles de un conjunto de datos:

```
> x = rnorm(200)
> quantile(x,probs=c(0.1,0.4,0.9))
      10%    40%    90%
-1.3194786 -0.3663511  1.0824184
```

Funciones para distribuciones de probabilidad

Densidad de probabilidad (d)	Densidad de probabilidad de la distribución seleccionada
Probabilidades (p)	Probabilidades
Quantiles (q)	Percentiles
Muestras aleatorias (r)	Generación de muestras aleatorias

Distribuciones disponibles:

binom	Binomial
cauchy	Cauchy
chisq	Chi Cuadrado
beta	Beta
exp	Exponencial
gamma	Gamma
geom	Geométrica
hyper	Hipergeométrica
lnorm	Log-normal
logis	Logística
nbinom	Binomial negativa
nchisq	Chi cuadrado no central
norm	Normal
pois	Poisson
signrank	Distribucion del test de Wilcoxon de rangos con signo
t	Student
unif	Uniforme
weibull	Weibull
wilcox	Distribución de la suma de rangos de Wilcoxon

Ejemplo con la distribución normal:

```
> rnorm(1) # Generación de un dato de la normal estandar
[1] -0.4120618
> rnorm(5)
```

```

[1] -0.3220499 -0.5556478 -0.1899898 -0.3450181 -2.5807986
> rnorm(5,mean=1,sd=3) # Generación de un dato de una normal no estandar
[1] -0.4035896 -0.8089832 3.2513373 4.9641722 -1.8603231

> dnorm(0) # Evaluación de la función de densidad normal en el punto 0
[1] 0.3989423
> dnorm(1)
[1] 0.2419707
> dnorm(3)
[1] 0.004431848

> pnorm(0) # Probabilidad acumulada bajo la normal en el punto 0
[1] 0.5
> pnorm(3)
[1] 0.9986501

> qnorm(0.5) # El cuantil 50% de la normal es el 0
[1] 0
> qnorm(0.9986501)
[1] 3.000000

> x<-seq(-4,4,length=200) # Dibujo del gráfico de la distribución normal
> plot(x,dnorm(x),type="l")

```

Otras funciones de utilidad

Una tarea relativamente frecuente es la elaboración de tablas a partir de variables dadas. En un ejemplo simple, podríamos tener el color de ojos de 8 personas y el color de pelo de las mismas guardados en dos variables, y podríamos querer crear una tabla de contingencia a partir de estas dos características:

```

> color.pelo<-c("negro","rubio","negro","rojo","oscuro","oscuro","rubio","negro")
> color.ojos<-c("negro","azul","marron","azul","negro","negro","azul","marron")
> table(color.pelo,color.ojos)

      azul marron negro
negro  0    2    1
oscuro  0    0    2
rojo   1    0    0
rubio  2    0    0

```

Otras funciones de utilidad son las que trabajan sobre matrices:

Funciones sobre matrices

chol	Descomposición de Cholesky
crossprod	Producto cruzado: crossprod(x,y) es lo mismo que t(x) %*% y
diag	Crea una matriz diagonal o extrae la diagonal de una matriz
eigen	Valores propios
outer	Producto exterior de dos vectores
scale	Escala las columnas de una matriz
solve	Resuelve sistemas de ecuaciones lineales y calcula la inversa
svd	Descomposición en valores singulares
qr	Descomposición QR
t	Traspuesta

Por ejemplo, supongamos que queremos resolver el sistema:

$$2x + 3y = 8$$

$$5x + 2y = 9$$

Plantaremos el sistema en la forma $Ax = b$ y resolveremos:

```
> A = matrix(c(2,3,5,2),ncol=2,byrow=T)
```

```
> b = c(8,9)
```

```
> solve(A,b)
```

```
[1] 1 2
```

La solución viene dada por $x = 1$, $y = 2$

Si queremos invertir una matriz:

```
> x = matrix(c(1,2,3,4),nrow=2) # Creamos una matriz
```

```
> x
```

```
 [,1] [,2]
```

```
[1,] 1 3
```

```
[2,] 2 4
```

```
> y = solve(x) # Calculamos la inversa
```

```
> y%%x # Multiplicamos la matriz por su inversa
```

```
 [,1] [,2]
```

```
[1,] 1 4.440892e-16
```

```
[2,] 0 10.000000e-01
```

```
> round(y%%x,2) # Redondeamos a dos decimales
```

significativos

```
 [,1] [,2]
```

```
[1,] 1 0
```

```
[2,] 0 1
```