

Práctica de Laboratorio de R

Curso: Introducción a la Estadística y Probabilidades CM-274

Lecturas Importantes

1. Norman Matloff presenta una guía acerca de lo que es la depuración. Revisar el siguiente enlace <http://heather.cs.ucdavis.edu/~matloff/UnixAndC/CLanguage/Debug.html>.
2. DSL(Domain Specific Languages) en R. <http://adv-r.had.co.nz/dsl.html>.

Preguntas

1. (a) Sea X el número de "unos" obtenidos en 12 lanzamientos de un dado. Entonces X tiene una distribución Binomial ($n = 12, p = 1/3$). Calcule una tabla de probabilidades binomiales para $x = 0, 1, \dots, 12$ por dos métodos:
 - i. Usando la fórmula para la densidad:

$$P(X = K) = \binom{n}{k} p^k (1 - p)^{n-k}$$

y aritmética en R. Usa $0:12$ para la secuencia de x valores y la función `choose` para calcular los coeficientes binomiales $\binom{n}{k}$.

- ii. Usando la función `dbinom` de R y comparar tus resultados con ambos métodos.
- (b) Sea X el número de "unos" obtenidos en 12 lanzamientos de un dado. Entonces X tiene una distribución Binomial ($n = 12, p = 1/3$). Calcule el CDF para $x = 0, 1, \dots, 12$ por dos métodos:
 - i. Usando la función `cumsum` y el resultado del ejercicio anterior.
 - ii. Con el uso de la función `pbinom`. ¿Qué es $P(X > 7)$?
2. (a) Escribe una función llamada `norma` que calcula la norma Euclídea de un vector numérico. La norma Euclídea de un vector $x = (x_1, \dots, x_n)$ es

$$\|x\| = \sqrt{\sum_{i=1}^n x_i^2}.$$

Usa operaciones vectorizadas para calcular la suma. Prueba esta función sobre los vectores $(0, 0, 0, 1)$ y $(2, 5, 2, 4)$ para verificar que el resultado de la función es correcto.

- (b) Construye una matriz con 10 filas y 2 columnas conteniendo data, de la siguiente manera

```
> x = matrix(rnorm(20), 10, 2)
```

Esta es una muestra aleatoria de 10 observaciones de una distribución normal bivariada estándar. Utiliza la función `apply` y la función `norma` del ejercicio anterior para cada una de esas 10 observaciones.

3. El modelo de Regresión Lineal Simple se ajusta a una respuesta y_i mediante una función lineal de una variable predictor x_i .

$$\hat{y}_i = a + bx_i \text{ para } (i = 1, \dots, n).$$

Por lo general, los mínimos cuadrados son utilizados para estimar los parámetros desconocidos a y b , pero a veces se utiliza la menor desviación absoluta. Esto requiere la elección de a y b a fin de minimizar

$$Q(a, b) = \sum_{i=1}^n |y_i - \hat{y}_i|.$$

- (a) Implementa una función que calcule $Q(a, b)$. Debes definir una función de un solo argumento el cual es un vector cuyos primer elemento es a y el segundo elemento b .
 - (b) Explica como usa R la función `optim` para obtener el mejor ajuste de valores de a y b .
4. (a) Escribe dos funciones en R: una que toma una función como argumento de entrada y otra que devuelve una función como salida.
- (b) Escribe una función en R, que toma un entero n como argumento y entonces usa un bucle `for` para generar n números aleatorios usando `runif`. Evalua la performancia de tu función usando la función `system.time`. Por ejemplo calcula la diferencia en el tiempo que toma tu función en retornar 10 millones de números y una llamada directa a `runif`.
- (c) Usa la función `choose` que calcula los coeficientes binomiales para determinar las probabilidades para todas las posibles salidas del siguiente código.

```
> adivina5 <- function(n, fSalida, Fecha_sorteo = "1952-1-17"){
+   if(dia_semana(as.Date(Fecha_sorteo)) == "Lunes")
+     Fecha_sorteo <- as.Date(Fecha_sorteo) +
+       c(0, cumsum(rep(c(3, 4), ceiling((n - 1)/2),
+         length.out = n - 1)))
+   else
+     Fecha_sorteo <- as.Date(Fecha_sorteo) +
+       c(0, cumsum(rep(c(4, 3), ceiling((n - 1)/2),
+         length.out = n - 1)))
+
+   Maquina <- sample(c("A", "B"), n, replace = TRUE)
+   Conjunto <- sample(1:10, n, replace = TRUE)
+   Dia <- substr(dia_semana(Fecha_sorteo), start = 1, stop = 1)
+   info <- cbind(as.character(Fecha_sorteo), Maquina, Conjunto, Dia)
+   A <- replicate(n*10, sample.int(40, 5))
+   for(i in 1:n) {
+     cat(info[i, ], "\n", file = fSalida, append = TRUE)
+     write(A[, (10*i - 9):(10*i)], file = fSalida, ncolumns = 5, append = TRUE)
+   }
+ }
```

5. (a) Fijar cada uno de los siguientes error y corregirlos

```
> mtcars[-1:4, ]
> mtcars[mtcars$cyl<= 5]
> mtcars[mtcars$cyl== 4 | 6, ]
```

- (b) ¿ Por qué `x <- 1:5; x[NA]` produce 5 valores NA?.
- (c) ¿ Por qué `mtcars[1:20]` retorna un error. ¿Cómo difiere del similar `mtcars[1:20,]`?
- (d) ¿Qué hace `df [is.na (df)] <- 0?` ¿Como funciona?.