

Práctica 1 de laboratorio de R

Curso: Introducción a la Estadística y Probabilidades CM-274

Lectura Importante

1. Libro de Aní Adhikari y John DeNero, llamado *Computational and Inferential Thinking* acerca de temas importantes de ciencia de datos. <http://www.inferentialthinking.com/>.

Preguntas

1. Escribe expresiones en R que generen vectores conteniendo las siguientes secuencias de vectores.
 - Las primeras 50 potencias de 3, empezando por 3^1 .
 - $1, 2^2, 3, 4^2, 5, 6^2, \dots, 99, 100^2$.
 - $1, -1, 2, -1, 3, -1, \dots, 20, -1$.
2. Reproduce el siguiente código fuente en R, para mostrar la siguiente tabla de probabilidad de la distribución estándar normal. Explica el uso de la función `outer()`.

```
> id <- 0:4
> dn <- seq(0, .8, by = .2)
> p = outer(id, dn, function(x,y) pnorm(x + y))
> dimnames(p) = list(z = id, "Primer lugar decimal de z " = dn)
> p = round(p, 5)
```

3. (a) Considere la siguiente curiosidad

$$\begin{aligned}8 \times 8 + 13 &= 77 \\8 \times 88 + 13 &= 717 \\8 \times 888 + 13 &= 7117 \\8 \times 8888 + 13 &= 71117 \\8 \times 8888 + 13 &= 711117 \\&\text{etc.}\end{aligned}$$

Escribe código R, que verifica las 10 primeras ecuaciones, imprimiendo los resultados, usando la fórmula dada usando el bucle `for`.

4. Considera el siguiente problema: Dada una matriz numérica X , determinar el índice de la primera fila cuyos elementos son todos números positivos (y que no contiene NA valores). Resuelve este problema usando un bucle `for`.

Sugerencia: usa las función `all` y `is.na()`

5. El conjunto de datos *iris* contiene las medidas de la longitud y el ancho (en cm) de pétalos y sépalos de tres especies: 1: Setosa, 2: versicolor y 3: Virginica.

- (a) Considera el objeto *iris*. ¿ Como está estructurado?. ¿ Cuantas observaciones(lineas) contiene?. ¿ Cuantas variables (columnas) contiene?.

- (b) Para tener una visión general del conjunto de valores, utiliza la función `summary()` del conjunto de dato. ¿Qué información sobre el conjunto de datos proporciona?
- (c) Para la variable `Sepal.Length` verifica los resultados dados, usando las funciones `min()`, `max()`, `mean()`, `median()`, `quantile()`. Si es necesario usa la ayuda `?quantile`.
6. Supongamos que x es un vector numérico. **Explica en detalle**, como las siguientes expresiones son evaluadas y que valores toman

```
> sum(!is.na(x))
> c(x, x[-(1:length(x))])
> x[length(x) + 1]/length(x)
> sum(x > mean(x))
```

7. Ejecuta el siguiente código

```
> genero <- factor(c(rep("f", 91), rep("m", 92)))
> table(genero)
> genero <- factor(genero, levels=c("m", "f"))
> table(genero)
> genero <- factor(genero, levels=c("M", "f"))
> table(genero)
> rm(genero)
```

Explica la salida y explica los números que aparecen.

8. La función `dim()` devuelve las dimensiones (un vector que tiene el número de filas entonces el número de columnas) de matrices y data frames. Utilice esta función para encontrar el número de filas de los data frames de `tinting`, `possum` y `possumsites` del paquete `DAAG`.
9. Datos del siguiente formato son almacenados en un archivo llamado `cdata.txt` (2 puntos)

```
2011-03-25 44 wet
2011-07-03 14 wet
2011-03-27 20 dry
2011-12-31 59 wet
2011-10-31 87 wet
2011-04-02 49 dry
2011-12-12 76 dry
2011-07-05 40 dry
2011-03-29 8 wet
2011-08-29 9 dry
2011-12-04 21 wet
2011-05-20 10 dry
2011-08-02 71 wet
2011-08-18 34 wet
2011-01-22 58 dry
2011-11-27 48 wet
2011-07-08 32 dry
2011-12-10 11 dry
2011-08-27 90 dry
```

- (a) Escribe código R para producir un data frame llamado `cdata` conteniendo las variables `data`, `conteo` y `clima`. Las variables deben tener sus respectivos tipos: `'character'`, `'numeric'`, `'factor'`.
- (b) Escribe código R que reordene las filas del data frame tal que la variable `conteo` esté en orden ascendente.