

Introduction al modelamiento ecológico

Guillermo Torres

Sobre este curso y sus datos

Los datos que se van a utilizar son de la tesis doctoral de Verneaux (Verneaux et al. 2003). Él propuso que se podían usar algunas especies de peces para caracterizar zonas ecológicas a lo largo de ríos y arroyos en Europa. Verneaux demostró que las comunidades de pescados eran buenos indicadores biológicos para estos cuerpos de agua.

Los datos que se van a utilizar consisten en 3 matrices que contienen parte de los datos usados por Verneaux. Estos datos han sido colectados de 30 sitios a lo largo del río *Doubs*, el cual corre entre Francia y Suiza en las montañas Jura. La primera matriz de datos contiene las abundancias de 27 especies de pescados, la segunda matriz contiene 11 variables ambientales relacionadas con la hidrología, geomorfología y química del río. La tercera matriz contiene las coordenadas geográficas (Cartesianas, X y Y) de los sitios.

Análisis exploratorio de los datos

Los análisis exploratorios son métodos que usan herramientas de visualización y calcula descriptores sintéticos que son requeridos para ganar información referente a:

- Tener una visión global de los datos
- Transformar o recodificar algunas variables
- Orientar análisis futuros

Exploración de los datos *Doubs*

Al comienzo de cada sesión de trabajo se recomienda definir el directorio o carpeta de trabajo. Si se usa Rstudio, el directorio de trabajo se puede definir yendo a:

- Sesión
 - Definir directorio de trabajo
 - Escojer directorio

En la consola se puede definir de la siguiente manera:

```
setwd('/direccion/de/la/carpeta/de/trabajo/')
```

Una vez definido el directorio de trabajo vamos a cargar los datos.

```
# Importar los datos desde archivos de texto, ej. formato CSV
# *****
# Matriz de las especies de pescados (data frame de las abundancias de los pescados):
spe <- read.csv('../data/DoubsSpe.csv', row.names=1)
# Matriz de los datos ambientales:
env <- read.csv('../data/DoubsEnv.csv', row.names=1)
# Matriz de la información espacial (Coordenadas geográficas X,Y):
spa <- read.csv('../data/DoubsSpa.csv', row.names=1)
```

Echemos un vistazo a los datos de la comunidad (objeto **spe**, abundancia de los pescados)

```

# funciones básicas
# *****
# Muestra el data frame completo (No recomendado para set de datos grandes):
spe
# Muestra solamente las primeras 5 líneas:
head(spe)

# Muestra la dimensión del datafrmae (No. filas ,No. columnas):
dim(spe)

## [1] 30 27

# Muestra el No. de filas:
nrow(spe)

## [1] 30

# Muestra el No. columnas:
ncol(spe)

## [1] 27

# Muestra solamente 5 filas y 10 columnas:
spe[1:5,1:10]

##   CHA TRU VAI LOC OMB BLA HOT TOX VAN CHE
## 1   0   3   0   0   0   0   0   0   0   0
## 2   0   5   4   3   0   0   0   0   0   0
## 3   0   5   5   5   0   0   0   0   0   0
## 4   0   4   5   5   0   0   0   0   0   1
## 5   0   2   3   2   0   0   0   0   5   2

# Nombre de las columnas (descriptors = especies):
colnames(spe)

## [1] "CHA" "TRU" "VAI" "LOC" "OMB" "BLA" "HOT" "TOX" "VAN" "CHE" "BAR"
## [12] "SPI" "GOU" "BRO" "PER" "BOU" "PSO" "ROT" "CAR" "TAN" "BCO" "PCH"
## [23] "GRE" "GAR" "BBO" "ABL" "ANG"

# Nombre de las filas
rownames(spe)

## [1] "1" "2" "3" "4" "5" "6" "7" "8" "9" "10" "11" "12" "13" "14"
## [15] "15" "16" "17" "18" "19" "20" "21" "22" "23" "24" "25" "26" "27" "28"
## [29] "29" "30"

# Estadísticos descriptivos para las columnas
summary(spe)

```

```

##           CHA           TRU           VAI           LOC
## Min.      :0.00   Min.      :0.00   Min.      :0.000   Min.      :0.000
## 1st Qu.:0.00   1st Qu.:0.00   1st Qu.:0.000   1st Qu.:1.000
## Median :0.00   Median :1.00   Median :3.000   Median :2.000
## Mean     :0.50   Mean     :1.90   Mean     :2.267   Mean     :2.433
## 3rd Qu.:0.75   3rd Qu.:3.75   3rd Qu.:4.000   3rd Qu.:4.000
## Max.     :3.00   Max.     :5.00   Max.     :5.000   Max.     :5.000

```

Usando la mediana y la media de las abundancias. Las distribuciones son simétricas?

```
# Distribución general de las abundancias
# *****
# Mínimo y máximo valor de las abundancias de todo el set de datos:
range(spe)

## [1] 0 5

# Número de especies ausentes
sum(spe==0)

## [1] 435

# Proporción de ceros en el set de datos
sum(spe==0)/(nrow(spe)*ncol(spe))

## [1] 0.537037

# Conteo de especies con 0,1,2,3 o n de abundancia (clases)
ab <- table(unlist(spe))
# Diagrama de barras de la distribución de abundancia de las especies
barplot(ab,las=1,xlab="Clases de abundancia",ylab="Frecuencia",co=gray(5:0/5))
```

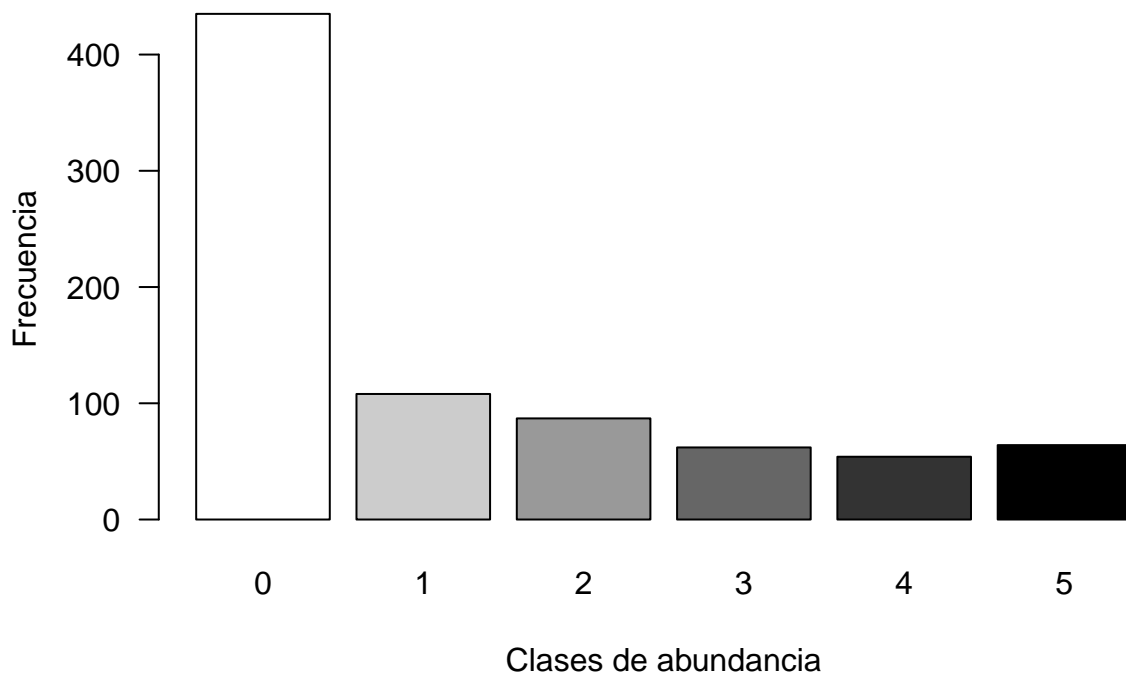
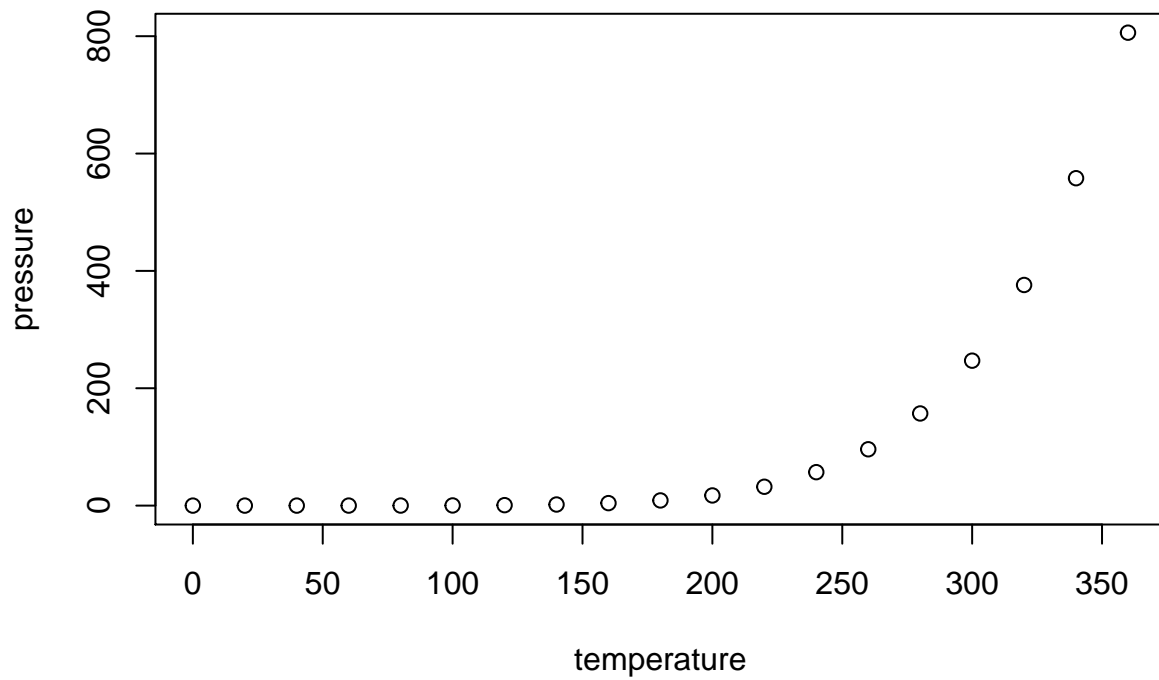


Fig. 1 Diagrama de barras de las clases de abundancia

Mirando el diagrama de barras sobre las clases de abundancia, cómo se interpretaría la alta frecuencia de ceros (ausencias) en el set de datos?

Including Plots

You can also embed plots, for example:



Note that the `echo = FALSE` parameter was added to the code chunk to prevent printing of the R code that generated the plot.