

PEC3

Fernando Moral Algaba

29 de diciembre de 2018

Sección 1 (8 puntos)

1. (1 punto) Buscad un conjunto de datos relacionados con la Bioestadística o Bioinformática. Para ello, podéis utilizar recursos conocidos de la PEC1, por ejemplo, como es el caso de <http://www.bioinformatics.org/sms2/index.html> o de <http://biostat.mc.vanderbilt.edu/wiki/Main/DataSets>. También podéis utilizar otros recursos propios que conozcáis o que sean de vuestro interés, y siempre teniendo en cuenta que sean datos públicos que podéis utilizar. Tenéis que explicar la procedencia de los datos así como incluir las referencias que correspondan y justificar porqué habéis elegido estos datos.

```
# Descomentar para usar uno u otro dataset

# Diabetes data

#mydata <- read.csv("http://biostat.mc.vanderbilt.edu/wiki/pub/Main/DataSets/diabetes.csv",header=T, sep=

# Duchenne muscular dystrophy dataset
mydata <- read.csv("http://biostat.mc.vanderbilt.edu/wiki/pub/Main/DataSets/dmd.csv", header= T, sep=",
t(head(mydata, n=4))
```

```
##           1           2           3           4
## X          1.00000    2.00000    3.0     4.00000
## hospid    657.00000   667.00000   669.0    671.00000
## age       27.00000    31.00000    22.0     25.00000
## sdate     6497.00000  6528.00000  6558.0   6497.00000
## ck        22.00000    29.00000    22.0     41.00000
## h         99.00000    94.00000    85.5     87.29688
## pk        10.79883    11.79883    15.0     15.00000
## ld         NA         NA         NA         NA
## carrier    0.00000    0.00000    0.0      0.00000
## obsno      1.00000    1.00000    1.0      1.00000
```

```
names(mydata)
```

```
## [1] "X"      "hospid" "age"     "sdate"   "ck"      "h"       "pk"
## [8] "ld"     "carrier" "obsno"
```

2. (1 punto) Utilizando R, mostrad y explicad qué tipo de fichero habéis importado y las variables que forman parte de él (tipo, clasificación,...), así como todo aquello que creáis relevante. Incluir capturas de pantalla y las instrucciones en R que habéis utilizado para importar y mostrar los datos.
3. (2 puntos) Realizad un mínimo de seis preguntas objetivo que den una idea de la información contenida en el conjunto de datos escogido. Para ello, podéis basaros en el tipo de consultas realizadas a la Sección 2 de la PEC1 y también utilizando, en alguno de los casos, la definición de funciones tal como se trabaja en el LAB3.
4. (1 punto) Realizad un análisis descriptivo de los datos. Este estudio debe incluir, como se vio en la Sección 3 de la PEC1, un resumen paramétrico de los datos y diversas representaciones gráficas de los mismos basadas en determinados criterios. Dejamos a vuestra elección el tipo de gráficos y los criterios utilizados.

5. (1 punto) Realizad, basándoos en los conceptos trabajados en el LAB4 y PEC2, un mínimo de tres cuestiones que respondan a una cuestión de probabilidad y un mínimo de una cuestión que corresponda a un breve modelo de simulación.
6. (1 punto) Realizad un breve análisis de regresión a partir de las variables que disponéis y utilizando el criterio que responda a alguna pregunta de interés que os hayáis planteado.
7. (1 punto) A partir de los datos de origen y el estudio realizado, haced una valoración final. Para ello, podéis basaros en las siguientes preguntas: “disponemos de conclusiones finales?”, “sería necesario hacer un análisis más avanzado?”, “faltan datos para obtener otro tipo de información como...?”,....

Sección 2 (2 puntos)

A lo largo del curso se ha trabajado con datos cuyo origen era diverso pero, básicamente, correspondían a archivos de tipo texto o hojas de cálculo. En este ejercicio se os pide que realicéis un breve estudio acerca de cómo gestionar la información a partir de una base de datos. En particular, se pide:

- Seleccionar una base de datos de libre acceso y importad, desde Rstudio, estos datos. Mostrad el código utilizado y el resultado obtenido por pantalla.
- Realizad un par de consultas, desde Rstudio, a partir de estos datos y mostrad el código utilizado y resultado obtenido por pantalla.