

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 correspon dan 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
## 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"
```