

Anàlisi de dades òmiques (M0-157). PAC1.

Marc Anton

2025-03-25

Introducció

Aquest arxiu conté les sol·lucions als problemes plantejats en la PAC1 *Anàlisi de dades òmiques (M0-157)*. El projecte total del treball es pot consultar al: meu repositori github, cliqueu **aquí**

A continuació, procedeix a resoldre la PAC. Abans que res, carreguem les dades al nostre entorn de treball. Treballarem amb l'exemple de mostra, veure **aquí**. Carreguem els dos fitxers:

```
DataInfo_S013 <- read.csv("DataInfo_S013.csv", sep = ",")
DataValues_S013 <- read.csv("DataValues_S013.csv", sep = ",")
```

Com que ja sabem que aquests fitxers solen portar molts camps, mirem primer l'estructura dels dos.

```
dim(DataInfo_S013)
```

```
## [1] 695 4
```

```
dim(DataValues_S013)
```

```
## [1] 39 696
```

Aquí ja veiem que en l'arxiu Values tenim 696 columnes, el que correspon a un index únic de cada cas d'estudi més les 695 variables que es tenen de cada un d'ells. Per veure que hi ha, mirem les 20 primeres files d'Info i podem fer-nos una idea

```
head(DataInfo_S013,20)
```

##		X	VarName	varTpe	Description
## 1		SUBJECTS	SUBJECTS	integer	dataDesc
## 2		SURGERY	SURGERY	character	dataDesc
## 3		AGE	AGE	integer	dataDesc
## 4		GENDER	GENDER	character	dataDesc
## 5		Group	Group	integer	dataDesc
## 6		MEDDM_TO	MEDDM_TO	integer	dataDesc
## 7		MEDCOL_TO	MEDCOL_TO	integer	dataDesc
## 8		MEDINF_TO	MEDINF_TO	integer	dataDesc
## 9		MEDHTA_TO	MEDHTA_TO	integer	dataDesc
## 10		GLU_TO	GLU_TO	integer	dataDesc
## 11		INS_TO	INS_TO	numeric	dataDesc
## 12		HOMA_TO	HOMA_TO	numeric	dataDesc
## 13		HBA1C_TO	HBA1C_TO	numeric	dataDesc
## 14	HBA1C.mmol.mol_TO	HBA1C.mmol.mol_TO	HBA1C.mmol.mol_TO	numeric	dataDesc
## 15		PESO_TO	PESO_TO	integer	dataDesc
## 16		bmi_TO	bmi_TO	numeric	dataDesc
## 17		CC_TO	CC_TO	numeric	dataDesc
## 18		CINT_TO	CINT_TO	integer	dataDesc

```
## 19          CAD_TO          CAD_TO    integer    dataDesc
## 20          TAD_TO          TAD_TO    integer    dataDesc
```

Aquí ja veiem que hi ha 9 primers camps que semblen ser els que tenen la informació sobre el pacient mostra. Mirem que contenen aquests camps a values:

```
# En mirem 10 en realitat, per comprovar que la primera columna és un index
head(DataValues_S013[,1:10])
```

```
##   X.1 SUBJECTS SURGERY AGE GENDER Group MEDDM_TO MEDCOL_TO MEDINF_TO MEDHTA_TO
## 1   1         1 by pass  27     F     1         0         0         0         1
## 2   2         2 by pass  19     F     2         0         0         0         0
## 3   3         3 by pass  42     F     1         0         0         0         0
## 4   4         4 by pass  37     F     2         0         0         0         0
## 5   5         5 tubular  42     F     1         0         0         0         0
## 6   6         6 by pass  24     F     2         0         0         0         0
```

Com suposavem, aquestes columnes corresponen a: 1. Un codi de pacient (que està repetit a les dues primeres columnes) 2. La cirurgia a la que va ser sotmès el pacient 3. L'edat i el gènere 4. Un grup de tractament 5. 4 tractaments que venen amb 0 i 1

Per veure que la resta són metabolits, mirem 5 columnes més

```
head(DataValues_S013[,11:15],10)
```

```
##   GLU_TO INS_TO HOMA_TO HBA1C_TO HBA1C.mmol.mol_TO
## 1    85  11.40   2.40      NA              NA
## 2    78  12.10   2.32      NA              NA
## 3    75   8.41   1.56    5.4          35.51
## 4    71  12.80   2.25    5.1          32.23
## 5    82   6.01   1.22    5.6          37.69
## 6    71   9.88   1.73    5.1          32.23
## 7    80   9.20   1.82    5.6          37.69
## 8    90   3.40   0.76    5.5          36.60
## 9    92   5.43   1.23    5.7          38.78
## 10   84   6.98   1.45    5.5          36.60
```

Obviament ja ho veiem i també podem veure que hi ha més d'un valor que no es té (NA)