

Descripción de los metadatos

Ana Gabriela Ponce González

2025-04-08

Descripción:

El dataset seleccionado para el análisis requerido en la prueba de evaluación continua n°01 fue el llamado *MetabotypingPaper*, disponible en el repositorio Nutrimetabolomics Lab. En la carpeta que contenía los datos había dos archivos de texto. Uno de ellos denominado DataValues_S013.csv, y el otro, DataInfo_S013.csv. El primero se corresponde con la matriz de expresión y el segundo, con los metadatos.

Centrándonos en los metadatos, estos describen las características clínicas y experimentales asociadas a cada paciente (muestra) de nuestro conjunto de datos.

```
# Cargamos los datos:
metadatos <- read.csv("DataInfo_S013.csv")
# Exploramos los datos:
head(metadatos, 10)
```

##	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_T0	MEDDM_T0	integer	dataDesc
## 7	MEDCOL_T0	MEDCOL_T0	integer	dataDesc
## 8	MEDINF_T0	MEDINF_T0	integer	dataDesc
## 9	MEDHTA_T0	MEDHTA_T0	integer	dataDesc
## 10	GLU_T0	GLU_T0	integer	dataDesc

Variables incluidas:

- **GENDER**: género del paciente (“male” o “female”).
- **AGE**: edad en años.
- **SURGERY**: tipo de cirugía bariátrica realizada (“by_pass” o “tubular”).
- **GROUP**: grupo experimental asignado según el tipo de cirugía y el tiempo de seguimiento.

Además de estas, que sirven para identificar al paciente, están las variables relacionadas con el tipo de muestras tomadas. Todas ellas se dividen en variable de T0 a T5 dependiendo del momento a lo largo del estudio en el que se tomaron. Algunas de ellas son:

- **BMI**: índice de masa corporal de los pacientes en el momento correspondiente.
- **LDL**: colesterol de baja densidad.

- **HDL:** colesterol de alta densidad. -**TG:** triglicéridos.

En general hay distintos tipos de marcadores medidos, desde el perfil lipídico, para evaluar el estado metabólico del paciente desde el punto de vista cardiovascular, hasta la concentración de cada aminoácido, probablemente para evaluar también el estado nutricional, entre otros.

Observaciones:

- El archivo contiene está formado por 695 observaciones de 4 variables. Concretamente, se describe el nombre de la variable y el tipo de la misma.
- Durante el desarrollo de la evaluación, los metadatos fueron limpiados para eliminar los valores nulos (VN).
- Se sustituyó el nombre en la variable “SURGERY” de “by pass” por “by__pass”.