

## Informe ADO-PEC1 Giuseppe Tardiolo

### *Punto 1: Selección y descarga del dataset.*

Para la prueba, he seleccionado el dataset “2018-MetabotypingPaper”, disponible en el repositorio de GitHub como se indica en el enunciado: <https://github.com/nutrimetabolomics/metaboData/>. Los datos fueron utilizados en el artículo de Palau-Rodriguez, M. et al. (2018), *Metabotypes of response to bariatric surgery independent of the magnitude of weight loss*. PloS one, 13(6), e0198214, <https://doi.org/10.1371/journal.pone.0198214>.

Este conjunto de datos resulta interesante porque proporciona información sobre las respuestas metabólicas individuales de pacientes que han pasado por cirugía bariátrica, sin depender únicamente del indicador tradicional de pérdida de peso. El conjunto de datos está compuesto por tres archivos principales: *DataInfo\_S013.csv*, archivo que contiene los metadatos; *DataValues\_S013.csv*, archivo que incluye valores clínicos y metabolómicos de 39 pacientes; *AAInformation\_S006.csv*, archivo que ofrece información adicional sobre los metabolitos.

El archivo *DataInfo\_S013.csv* describe diversas variables incluidas en el dataset seleccionado, como variables de sujetos y demográficas, variables clínicas y bioquímicas (incluyendo glucosa, insulina, colesterol, y biomarcadores inflamatorios, entre otros), variables que miden la concentración de aminoácidos y otros metabolitos específicos, así como otras moléculas que indican diferentes tipos de lípidos y metabolitos adicionales. El archivo *DataValues\_S013.csv* contiene valores de diversas mediciones bioquímicas, hematológicas y de aminoácidos para un grupo de sujetos identificados mediante códigos. Los datos incluyen detalles como el tipo de cirugía realizada, edad, género, grupos de medicación, y varias mediciones de glucosa, insulina, colesterol, triglicéridos y otros parámetros metabólicos, entre otros.

### *Punto 2. Creación de un contenedor del tipo SummarizedExperiment.*

Para este punto, he utilizado el paquete *SummarizedExperiment* de R para crear un contenedor que almacene tanto los datos clínicos y metabolómicos como los metadatos. El contenedor *SummarizedExperiment* facilita el manejo de conjuntos de datos complejos, permitiendo almacenar los datos y metadatos en un solo objeto. Una vez instalado y cargado el paquete, he procedido a crear un objeto *SummarizedExperiment*. Para esto, se necesita:

- los datos de expresión, es decir, la matriz principal de datos (*data\_values*), donde las filas representan las muestras y las columnas representan las variables;
- los metadatos de las filas, es decir, la información sobre las muestras (por ejemplo: edad, género, grupo, etc.);
- los metadatos de las columnas, es decir, la información sobre las variables (por ejemplo: tipos de mediciones). He verificado el objeto utilizando las funciones `dim()`, `colnames()` y `rownames()` para confirmar la correcta estructuración.

Al visualizar el objeto *SummarizedExperiment*, he observado su estructura y lo que representa cada elemento:

- la Clase, es decir, *SummarizedExperiment*;
- las Dimensiones, indicando 39 filas (muestras) y 695 columnas (variables);
- el Assay, que incluye un solo assay llamado *counts*, que contiene los datos metabólicos en forma de matriz;
- el RowData, que contiene las características de las muestras (como SUBJECTS, SURGERY, AGE, GENDER, y Group);

- y el ColData, que contiene la información de las variables (con detalles como X, VarName, varTpe, y Description).

### *Punto 3. Exploración del dataset.*

La exploración del dataset incluyó los siguientes pasos:

- un resumen estadístico básico del dataset con la función `summary()` en R;
- la detección de valores ausentes;
- a modo de ejemplo, he utilizado la función `hist()` para producir un histograma de la variable HDL\_T0 para visualizar su distribución;
- y, como ilustré en el script de R, he utilizado una matriz de correlación entre variables y un mapa de calor que ayuda a identificar posibles patrones y relaciones en los datos.

La mapa de calor de correlaciones obtenida ofrece una visión general de las relaciones entre las variables seleccionadas (primeras 50 columnas debido a la extensión del dataset). El mapa de calor muestra agrupamientos de variables con correlaciones positivas (en rojo) y negativas (en azul). Estos bloques indican que ciertos grupos de variables tienden a tener valores relacionados o a seguir patrones similares. Algunas variables se agrupan en bloques, lo que sugiere relaciones cercanas entre ellas. Este agrupamiento puede ayudar a identificar subconjuntos de variables que podrían estar vinculadas en términos generales. Además de las áreas de correlación alta (positiva o negativa), se observan áreas en tonos amarillos, que indican correlaciones cercanas a cero. Esto muestra que hay variables en el dataset que no están directamente relacionadas entre sí.

### *Punto 4. Informe.*

Este punto está representado por el presente informe.

### *Punto 5: Creación del Repositorio en GitHub.*

Como se indica en el enunciado, he creado un repositorio en GitHub llamado **Tardiolo-Giuseppe-PEC1** para almacenar los archivos, incluyendo:

- El informe en formato PDF;
- El objeto *SummarizedExperiment* en formato binario (`se_object_Giuseppe_Tardiolo.Rda`);
- El script del código R empleado en formato `.rmd` (R markdown), que incluye los datos y metadatos;
- Los archivos de datos y metadatos en formato texto (`.csv`).

Finalmente, la URL del repositorio es: <https://github.com/Ellendel84/Tardiolo-Giuseppe-PEC1.git>