

Asignatura: (6648) Bioestadística

Máster de Bioinformática

**Trabajo individual** de FAEDE+AEDM

**Fecha tope de entrega: 3 de diciembre 23:59:59**

Entrega del informe por email a través de la herramienta mensajes privados del AV en cualquiera de los formatos que permite la compilación, junto con el fichero de código (\*.R).

Importante: Utiliza el fichero *pone\_0187371\_s002.csv* adjunto al mensaje.

Este ejercicio está centrado en la búsqueda de perfiles de expresión génica en un conjunto de datos de microarrays de GEO(NCBI) que consta de 124 observaciones divididas en dos grupos/clases:

- control (59) y
- autism (65).

Las muestras de individuos autistas y de control se recogieron de personas de Phoenix. El grupo de cada observación se encuentra en el factor Class, columna 201 del fichero *pone\_0187371\_s002.csv*.

Cada estudiante tiene asignados 11 genes del fichero *pone\_0187371\_s002.csv* correspondientes a las columnas indicadas en la Tabla incluida al final de este enunciado.

Se pide

1. Estudio del conjunto de genes significativos

- Ordena las variables de acuerdo a su información. Obtén los estadísticos descriptivos y representa la expresión de los once genes, para cada grupo del factor Class.
- Estudia la normalidad de los valores de expresión de cada gen a través de los gráficos cuantil-cuantil y de los contrastes de normalidad adecuados.
- Contrasta la **igualdad de medias** de los valores de expresión para cada gen entre individuos autistas y de control. Interpreta los resultados. Explica tu consideración sobre la adecuación de la prueba paramétrica aplicada. Indica justificadamente qué genes se expresan diferencialmente

2. Realiza un Análisis de Componentes Principales.

Se sugiere el estudio de idoneidad del ACP, variabilidad de cada una de las componentes principales, determinación del número adecuado de componentes principales, la expresión de las componentes extraídas y su interpretación.

3. Representa conjuntamente las expresiones génicas (variables) y los individuos de la muestra (observaciones). Interpreta la nube de puntos.

#### 4. Determinación del número óptimo de conglomerados o clases

- Realiza un análisis jerárquico de clasificación no supervisado y encuentra la estrategia de enlace óptima. Obtén y comenta el correspondiente dendrograma. Incluye distintas estrategias de enlace, la comparación y la justificación de la elección.
- Determina el número de conglomerados adecuado.
- Obtén el número de individuos y de qué grupo están incluidos en cada uno de los conglomerados obtenidos.

#### 5. Realiza un análisis clúster no jerárquico de los individuos. Interpreta los resultados, y completa las siguientes cuestiones:

- ¿Cuántos individuos y de qué grupo están incluidos en cada uno de los conglomerados? Compara estos resultados con los del apartado 4.c.
- Los conglomerados de la partición obtenida ¿están bien diferenciados?

Se sugiere obtener la media, mínimo y máximo de las variables para cada conglomerado y comenta los resultados.

Por último, la inclusión de resultados estadísticos adicionales en uno de los apartados siguientes: 2, 4 o/y 5 supondrá hasta 1 punto de la puntuación total del ejercicio.

> `set.seed(1234); sample(1:190,size=23,replace=F)`

[1] 142 51 152 58 93 75 96 2 86 185 38 103 94 10 160 40 167 180 30 1 72 12 3

ALBALADEJO RIAD, NORA	142 – 152
BELMONTE MARTINEZ, JOSE MARIA	51 - 61
CABANES MADRID, JOSE ALBERTO	152 – 162
CANTON SANDOVAL, JOAQUIN	58 – 68
GARAY MAYOL, BEATRIZ	93 – 103
GONZALEZ-CONDE RUIZ, MIRIAM	75 – 85
HANNAOUI ANAAOUI, ADIL	96 – 106
HERNANDEZ HIDALGO, ALBERTO	2 – 12
LOPEZ CARBONELL, FRANCISCO JAVIER	86 – 96
LOPEZ FERNANDEZ, EVA MARIA	185 – 195
LOPEZ MARIN, JUAN PEDRO	38 – 48
LUCAS RUIZ, FERNANDO	103 - 113
MAS GOMEZ, JORGE	94 - 104
MORENO GONZALEZ, NOELIA	10 – 20
PESCATORE, GIULIA MARIA	160 – 170
RAMOS GONZALEZ, SARA	40 – 50
REDONDO ANIORTE, FRANCISCO JAVIER	167 – 177
RIQUELME GARCIA, ANDREA	180 – 190
RODRIGUEZ RODENAS, ALBA	30 – 40
SGAMBELLONE, NATALIA	1 – 11
TAHIRI ZAINANE, GHIZLANE	72 – 82
TURPIN GOMEZ, ASUNCION	12 - 22
VIDAL ORTEGA, ALICIA	3 – 13