



## 1. DATOS BÁSICOS DEL TFG:

**Título:** Técnicas de clustering para datos longitudinales

**Descripción general** (resumen y metodología):

En los estudios longitudinales en el ámbito clínico, donde se tienen medidas repetidas a lo largo del tiempo de un mismo individuo, uno de los objetivos principales es establecer grupos de pacientes con trayectorias similares que comparten patrones a lo largo del tiempo. En este contexto, las técnicas de clustering se usan frecuentemente para este propósito. No obstante, el análisis de datos longitudinales presenta diferentes retos desde el punto de vista metodológico, como dependencia temporal entre observaciones, presencia de datos faltantes o la posible no linealidad de las trayectorias. Estos factores limitan en gran medida el uso de métodos clásicos y requieren del uso de técnicas avanzadas diseñadas de forma específica para manejar y detectar patrones en este tipo de datos.

En este Trabajo de Fin de Grado se van a analizar los principales métodos de clustering en datos longitudinales y los paquetes de R que implementan estas metodologías. Además, se aplicarán algunas de estas metodologías a un conjunto de datos longitudinales reales de 3000 pacientes de Párkinson para establecer una estratificación de los mismos en base a sus trayectorias temporales.

**Tipología:** Estudio de casos, teóricos o prácticos, relacionados con la temática del Grado.

**Objetivos planteados:**

El objetivo principal de esta propuesta es revisar y aplicar las técnicas de estratificación para datos ómicos longitudinales. Para ello, se proponen los siguientes subobjetivos:

1. Hacer una revisión bibliográfica de métodos y paquetes de software de clustering para datos longitudinales
2. Analizar desde un marco teórico los métodos seleccionados.
3. Aplicar estos métodos a un estudio de pacientes de Párkinson con datos de expresión génica.
4. Caracterizar los subgrupos obtenidos con la información clínica disponible.

**Bibliografía básica:**

- Zihang Lu. Clustering Longitudinal Data: A Review of Methods and Software Packages. International Statistical Review. 2024
- Poewe, W., Seppi, K., Tanner, C.M., Halliday, G.M., Brundin, P., Volkman, J., Schrag, A.-E., and Lang, A.E. (2017). Parkinson disease. Nature Reviews Disease Primers 3, 17013.
- Hagenberg et al. longmixr: a tool for robust clustering of high-dimensional cross-sectional and longitudinal variables of mixed data types. Bioinformatics, 2024

**Recomendaciones y orientaciones para el estudiante:**

Se recomienda especialmente el trabajo a aquellos estudiantes con interés en estadística computacional y análisis de datos en biomedicina.

**Plazas:** 1

**2. DATOS DEL TUTOR/A:**

**Nombre y apellidos:** PEDRO MARÍA CARMONA SÁEZ

**Ámbito de conocimiento/Departamento:** ESTADÍSTICA E INVESTIGACIÓN OPERATIVA

**Correo electrónico:** pcarmona@ugr.es

**3. COTUTOR/A DE LA UGR (en su caso):**

**Nombre y apellidos:**

**Ámbito de conocimiento/Departamento:**

**Correo electrónico:**

**4. COTUTOR/A EXTERNO/A (en su caso):**

**Nombre y apellidos:**

**Correo electrónico:**

**Nombre de la empresa o institución:**

**Dirección postal:**

**Puesto del tutor en la empresa o institución:**

**Centro de convenio Externo:**

**5. DATOS DEL ESTUDIANTE:**

**Nombre y apellidos:**

**Correo electrónico:**