MÁSTER UNIVERSITARIO EN BIOINFORMÁTICA Y BIOESTADÍSTICA

PROGRAMARIO PARA EL ANÁLISIS DE DATOS - PEC4

Objetivo: Validar los conocimientos adquiridos en todos los módulos.
Fecha de entrega: 13 de enero de 2026, 12:00 PM.
Formato: *ApellidoEstudiante1_ApellidoEstudiante2_PAC4_SAD.pdf*
Entrega: REC (Registro de evaluación continua) del aula.
Comentarios previos a la realización de la PEC4:

1. El foro será el espacio para compartir motivaciones sobre la temática y formar grupos. Si se desea formar grupo con un/a compañero/a de otra aula, es necesario notificarlo por correo electrónico a los profesores o tutores implicados.

2. La práctica se realizará preferentemente **en grupos de 2 personas**, salvo en casos muy excepcionales y debidamente justificados.

3. Es necesario incluir el nombre de los estudiantes del grupo dentro de la PEC.

4. El formato de entrega debe ser el generado a partir de **RMarkdown** en formato **HTML** o **PDF**.

5. Es necesario incluir el código utilizado y comentar los resultados obtenidos.

6. En el contexto del grupo, las tareas deben organizarse de manera equitativa. Ambos miembros del grupo deben conocer la totalidad del trabajo. De forma opcional, se puede dejar constancia de la distribución del trabajo y la valoración del trabajo en grupo.

Enunciado

Esta PEC consiste en la elaboración de un informe de un caso práctico a vuestra elección que permita poner en práctica los conceptos principales del curso. A partir de una fuente de datos de un determinado caso o ámbito, debéis realizar un **estudio de análisis de datos detallado en R**. Si los datos utilizados son reales, debéis garantizar que son de uso libre o público.

Pautas del estudio

Sección 1. Contexto y objetivo del estudio. Datos (1 punto)

Debéis buscar un conjunto de datos relacionado con la Bioestadística y la Bioinformática, basándoos en los intereses profesionales o las preferencias del grupo. Podéis utilizar recursos como:

- http://www.bioinformatics.org/sms2/index.html
- https://hbiostat.org/data
- O también podéis utilizar otros recursos propios, siempre que los datos sean públicos.

Especificad la procedencia o el origen de los datos, incluid las referencias correspondientes y justificad por qué habéis elegido estos datos, indicando qué objetivos o preguntas queréis responder.

Sección 2. Prospección y preparación de los datos (2 puntos)

2.1 Descripción de los datos (1 punto) Utilizando R, mostrad y explicad el tipo de fichero que habéis importado y las variables que lo componen. Esta descripción debe incluir (basándose en conceptos del LAB1):

- Descripción general del conjunto de datos.
- Tipo y clasificación de las variables (numéricas, categóricas, etc.).
- Tamaño del conjunto de datos.
- Detección de valores nulos.
- Valoración del conjunto de datos (necesidad de transformaciones o presencia de inconsistencias).

Debéis incluir capturas de pantalla y las instrucciones en R utilizadas para importar y mostrar los datos.

2.2 Preguntas "objetivo" (1 punto) Plantead un mínimo de cuatro preguntas objetivo que den una idea de la información contenida en el conjunto de datos. El objetivo es obtener información a partir de determinados criterios, según variables o rangos de valores. Podéis basaros en el tipo de consultas realizadas en la PEC1 y utilizar, en alguno de los casos, la definición de funciones (tal como se trabaja en el LAB3).

Sección 3. Análisis exploratorio de los datos (2,5 puntos)

3.1 Análisis descriptivo y gráfico (1 punto) Realizad un análisis descriptivo de los datos (basado en conceptos del LAB2). Este estudio debe incluir un resumen paramétrico de los datos y diversas representaciones gráficas basadas en criterios determinados según los tipos de variables y el objetivo del estudio. El tipo de gráficos y los criterios quedan a vuestra elección.

3.2. Ejercicios de inferencia y simulación (1,5 puntos) a) Basándoos en los conceptos trabajados en el LAB3, definid una función en R que realice algún tipo de cálculo de interés en el contexto del conjunto de datos. **b)** Basándoos en los conceptos del LAB4 y

la PEC2, plantead un mínimo de tres enunciados que respondan a una cuestión de probabilidad. **c)** Incluid un mínimo de un enunciado que corresponda a un breve modelo de simulación. Si vuestro conjunto de datos no facilita este tipo de enunciados, podéis generar una o varias distribuciones basándoos en parámetros determinados definidos por vosotros, afines al contexto del estudio..

Sección 4. Modelos de aprendizaje automático (2,5 puntos)

Basándoos en los conceptos trabajados en el LAB5 y a partir de vuestro conjunto de datos y de los objetivos del estudio, evaluad qué tipo de modelos conviene realizar: **aprendizaje supervisado o no supervisado**.

- Justificad vuestra elección.
- Mostrad el detalle del estudio realizado.
- Incluid las representaciones gráficas correspondientes (por ejemplo, clústeres, si procede).

Sección 5. Visualización (1,5 puntos)

Utilizando la herramienta **Shiny** (basada en los conceptos trabajados en el LAB6), realizad una propuesta de visualización de datos del conjunto de datos trabajado, a partir de unos determinados criterios a vuestra elección.

Sección 6. Conclusiones (0,5 puntos)

A partir de todo el estudio realizado en esta práctica, realizad una valoración final. Para ello, podéis basaros en preguntas como:

- *¿Disponemos de conclusiones finales?*
- *¿Sería necesario realizar un análisis más avanzado?*
- *¿Faltan datos para obtener otro tipo de información como…?*
- *¿Se puede utilizar alguna de las conclusiones para tomar algún tipo de decisiones?*

En otro orden de cuestiones, valorad también el trabajo en grupo y la calidad del informe de análisis de datos generado.