

## MÁSTER UNIVERSITARIO EN BIOINFORMÁTICA Y BIOESTADÍSTICA

### SOFTWARE PARA EL ANÁLISIS DE DATOS (SAD)

PEC 3	
<b>Objetivos</b>	Validar los conocimientos adquiridos en todas las unidades
<b>Fecha y hora</b>	<b>Del 11 de diciembre del 2020 al 12 de enero de 2021</b>
<b>Formato</b>	<i>Nombre_Apellido_PEC3_SAD.pdf</i>
<b>Entrega</b>	REC (Registro de evaluación continua) del aula.
<b>Contenidos</b>	<ol style="list-style-type: none"><li>1. Escribid el nombre dentro de la PEC y enumerad las páginas.</li><li>2. Incluid el código usado y comentad los resultados.</li><li>3. Cread un índice de trabajo si lo creéis necesario.</li></ol>

### Instrucciones de la práctica

Esta práctica corresponde al trabajo de síntesis de todos los conceptos trabajados a lo largo del curso y se realizará preferentemente en **grupo de 2 personas**, excepto en casos excepcionales debidamente justificados. Cada miembro del grupo debéis enviar la solución final con el formato "Nombre\_Apellido\_PEC3\_SAD.pdf". También debéis entregar el fichero de código en R. La entrega se hará a través del REC al igual que el resto de PECs del curso. En el documento que entreguéis ha de aparecer el nombre de los dos miembros del grupo de trabajo.

Se deberá utilizar MARKDOWN y entregarlo en Word o Pdf usando Knitr en RStudio (para ello, deben utilizarse los contenidos trabajados en el LAB6, en especial la herramienta ggplot2).

**Nota:** El trabajo debe distribuirse de forma equitativa, aunque el documento ha de ser homogéneo, es decir, al final se ha de entregar un único trabajo con un mismo estilo. Aunque no es requisito indispensable, podéis dejar constancia en el documento de cuál ha sido la distribución y la valoración del trabajo en grupo.

## Enunciado

El objetivo de la PEC es elaborar un análisis estadístico en R de un caso práctico con datos de obtención propia (pertenecientes a un determinado ámbito), lo más ajustado a la realidad. A continuación, podéis encontrar una serie de pautas para su realización:

- 1) (1 punto) Buscar un conjunto de datos relacionados con la Bioestadística o la Bioinformática. Posibles recursos bibliográficos son aquellos presentados en la PEC1, por ejemplo: <http://www.bioinformatics.org/sms2/index.html>; o bien <http://biostat.mc.vanderbilt.edu/wiki/Main/DataSets>. También es posible utilizar fuentes propias o de interés, siempre teniendo en cuenta que sean datos públicos para los usuarios. Será necesario especificar la procedencia y la justificación de la elección de los datos.
- 2) (1 punto) Utilizando R, mostrar y explicar qué tipo de ficheros se han importado, las variables de estudio (tipo, clasificación, ...), así como todo aquello relevante para el estudio. Incluir capturas de pantalla y las instrucciones en R utilizadas para importar y mostrar los datos.
- 3) (2 puntos) Con la Sección 2 de la PEC1 como base, elaborar y analizar una serie de cuestiones, que ayuden a explorar y a familiarizarse mejor con los datos de estudio. Además, en algunos casos, puede utilizarse la definición de funciones y el lenguaje SQL estudiado en el LAB3.
- 4) (1 punto) Realizar un análisis descriptivo de los datos. El análisis debe incluir (tal y como aparece en la Sección 3 de la PEC1) un resumen paramétrico de los datos y su representación gráfica, que mejor defina y complemente cada una de dichas variables.
- 5) (1,5 punto) Complementando el apartado anterior, elaborar un análisis de regresión de dos conjuntos de variables (LAB2 y Ejercicio 6 de la PEC1). La elección de las variables, los resultados, así como su relación deben de estar correctamente justificada.
- 6) (1 punto) Realizar, a partir de los conceptos trabajados en el LAB4 y la PEC2, un estudio probabilístico (a elección propia) de al menos 3 de las variables, que ayude a esclarecer cuestiones de relevancia que se plantean en los ámbitos de acción estudiados.
- 7) (1,5 punto) Complementando el apartado anterior, elaborar un análisis ANOVA de dos conjuntos de variables (LAB5 y Ejercicio 6 de la PEC2). La elección de las variables, los resultados, así como su relación deben de estar correctamente justificada. Además, realizar un test clúster de las variables, y si existe un fuerte agrupamiento, elaborar un dendograma (LAB5).
- 8) (1 punto) A partir de los datos de origen y el estudio realizado (incluyendo todos los puntos anteriores), presentar un apartado de conclusiones. Esta sección debe incluir un resumen de los principales resultados obtenidos en apartados anteriores, que ayuden al lector a comprender el ámbito de estudio. Además, se valorará positivamente la coherencia de resultados y las justificaciones presentadas.