

MÁSTER UNIVERSITARIO EN BIOINFORMÁTICA Y BIOESTADÍSTICA

SOFTWARE PARA EL ANÁLISIS DE DATOS (SAD)

PEC3	
Objetivos	Validar los conocimientos adquiridos en todas las unidades
Fecha y hora	Del 17 de Diciembre de 2018 al 14 de Enero de 2019 12:00 PM
Formato	Apellido_Nombre_PEC3_SAD.pdf
Entrega	REC (Registro de evaluación continua) del aula.
Comentarios	<ol style="list-style-type: none">1. Escribid el nombre dentro de la PEC y enumerad las páginas.2. Incluid el código usado y comentad los resultados.3. Cread un índice de trabajo si lo creéis necesario.

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.

Nota.- En el contexto del grupo, debéis organizaros las tareas de forma lo más equitativa posible pero teniendo en cuenta que los dos miembros del grupo siempre han de conocer la totalidad del trabajo. Aunque no es un requisito indispensable, podéis dejar constancia en el documento de entrega de cual ha sido la distribución de trabajo, así como la valoración del trabajo en grupo.

Enunciado

Esta PEC consiste en trabajar un caso práctico lo más ajustado a la realidad que sea posible y os permita poner en práctica los conceptos principales del curso. En resumen, se trataría de, a partir de una colección de datos pertenecientes a un determinado ámbito, realizar un estudio estadístico con R.

Para ello, podéis seguir las pautas que se explican a continuación.

Sección 1 (8 puntos)

1. (1 punto) Buscad un conjunto de datos relacionados con la Bioestadística o Bioinformática. Para ello, podéis utilizar recursos conocidos de la PEC1, por ejemplo, como es el caso de <http://www.bioinformatics.org/sms2/index.html> o de <http://biostat.mc.vanderbilt.edu/wiki/Main/DataSets>. También podéis utilizar otros recursos propios que conozcáis o que sean de vuestro interés, y siempre teniendo en cuenta que sean datos públicos que podéis utilizar. Tenéis que explicar la procedencia de los datos así como incluir las referencias que correspondan y justificar porqué habéis elegido estos datos.
2. (1 punto) Utilizando R, mostrad y explicad qué tipo de fichero habéis importado y las variables que forman parte de él (tipo, clasificación,...), así como todo aquello que creáis relevante. Incluir capturas de pantalla y las instrucciones en R que habéis utilizado para importar y mostrar los datos.
3. (2 puntos) Realizad un mínimo de seis preguntas objetivo que den una idea de la información contenida en el conjunto de datos escogido. Para ello, podéis basaros en el tipo de consultas realizadas a la Sección 2 de la PEC1 y también utilizando, en alguno de los casos, la definición de funciones tal como se trabaja en el LAB3.
4. (1 punto) Realizad un análisis descriptivo de los datos. Este estudio debe incluir, como se vio en la Sección 3 de la PEC1, un resumen paramétrico de los datos y diversas representaciones gráficas de los mismos basadas en determinados criterios. Dejamos a vuestra elección el tipo de gráficos y los criterios utilizados.

5. (1 punto) Realizad, basándoos en los conceptos trabajados en el LAB4 y PEC2, un mínimo de tres cuestiones que respondan a una cuestión de probabilidad y un mínimo de una cuestión que corresponda a un breve modelo de simulación.
6. (1 punto) Realizad un breve análisis de regresión a partir de las variables que disponéis y utilizando el criterio que responda a alguna pregunta de interés que os hayáis planteado.
7. (1 punto) A partir de los datos de origen y el estudio realizado, haced una valoración final. Para ello, podéis basaros en las siguientes preguntas: "disponemos de conclusiones finales?", "sería necesario hacer un análisis más avanzado?", "faltan datos para obtener otro tipo de información como...?",....

Sección 2 (2 puntos)

A lo largo del curso se ha trabajado con datos cuyo origen era diverso pero, básicamente, correspondían a archivos de tipo texto o hojas de cálculo. En este ejercicio se os pide que realicéis un breve estudio acerca de cómo gestionar la información a partir de una base de datos. En particular, se pide:

- Seleccionar una base de datos de libre acceso y importad, desde Rstudio, estos datos. Mostrad el código utilizado y el resultado obtenido por pantalla.
- Realizad un par de consultas, desde Rstudio, a partir de estos datos y mostrad el código utilizado y resultado obtenido por pantalla.