**Curso IMIM-CIBERCV**

**Introducción al análisis y manejo de datos con el programa R**

**Profesorado:**

*Isaac Subirana* es licenciado en Ciencias y Técnicas Estadísticas por la Universitat Politècnica de Catalunya (2014) y Doctor en Estadística por la Universitat de Barcelona (UB) (2014). Desde 2007 es profesor asociado en el departamento de Estadística de la UB. Desde 2003 trabaja a tiempo completo como técnico investigador en el IMIM-Hospital del Mar y en el grupo CIBERCV CB16/11/00246.

*Andrea Toloba* es graduada en Matemáticas por la Universitat de Barcelona (2019). Desde 2020 trabaja como técnico investigador en el grupo REGICOR (IMIM-Hospital del Mar). Actualmente cursando un Máster en Estadística por la Universitat Politècnica de Catalunya.

**Organización:**

*Jaume Marrugat* MD, PhD FESC Responsable del grupo REGICOR en el IMIM- Hospital del Mar, responsable del Programa de Epidemiología Cardiovascular del CIBERCV y coordinador del grupo CB16/11/00229 de CIBERCV; y *Juan Sanchis* MD PhD Catedrático de Medicina (Cardiología). Universidad de Valencia, Jefe de la Unidad de Hemodinámica y Cardiología intervencionista. Hospital Clínico Universitario. Valencia. Jafe del Grupo de investigación en cardiología clínica perteneciente al Instituto de Investigación Sanitaria INCLIVA, y coordinador del grupo CB16/11/00420 de CIBERCV

**Descripción del curso:** El programa R es un programa *open* *source* y de libre acceso reconocido por la comunidad estadística como el más usado en la investigación biomédica, bioinformática y matemática. En este curso se enseñará cómo leer y manejar bases de datos, calcular estadísticos descriptivos, realizar pruebas de asociación o representar los datos gráficamente con R. El contenido del curso será eminentemente práctico para que el alumno pueda ir aprendiendo el lenguaje y funcionamiento del programa mediante multitud de ejemplos.

**Dirigido a:** Todo personal investigador, técnicos, residentes o personal investigador en formación con experiencia previa en análisis de datos con otros programas estadísticos, por ejemplo SPSS, y que quieran usar el programa R en su trabajo diario.

**Objetivo del curso:** Capacitar al personal asistente en la utilización del programa R con la más amplia variedad de técnicas estadísticas y análisis gráfico de datos.

**Número de alumnos:** El aforo ideal para el curso es de 20 alumnos, aunque se podrá extender a 25 dependiendo de la demanda.

**Temario:**

* Introducción al entorno de R: Instalación de R y de RStudio.
* Estadística univariada: medias, medianas, desviación estándar, percentiles, tablas de frecuencia, intervalos de confianza, normalidad, etc.
* Relación entre dos variables cualitativas: prueba de Chi al cuadrado, prueba exacta de Fisher, Odds Ratio, Riesgo Relativa, etc.
* Relación entre una variable dicotómica y una cuantitativa: prueba t de Student para datos independientes, prueba U de Man-Whitney, etc.
* Relación entre una variable cualitativa de más de 2 categorías y una cuantitativa: ANOVA, prueba de Kruskall-Wallis; correcciones por comparaciones múltiples.
* Relación entre una variable cualitativa entre dos situaciones diferentes: la prueba de Mac-Nemar.
* Relación entre una variable cuantitativa medida en dos situaciones diferentes: la prueba t de Student para datos apareados: la prueba de Wilcoxon.
* Relación entre dos variables cuantitativas: correlación de Pearson/Spearman, regresión lineal simple.
* Lectura i exportación de bases de datos: formato csv, Excel, Access, SPSS, Stata, …
* Transformaciones: agregar casos, recodificar variables, contar apariciones, identificar duplicados, ejemplos generales de transformaciones, de manejo de fechas y de variables de tipo texto.
* Manejo de bases de datos: añadir casos/variables, eliminar/renombrar/seleccionar variables, cambiar/crear etiquetas de valor o de variable, estratificar por grupos, seleccionar casos,…
* Definición y creación de objetos: vectores, *arrays*, matrices, *data frames* y listas.
* Programación: creación de funciones, bucles, y expresiones condicionales.
* Gráficos: gráficos de cajas, de dispersión, de barras, etc.
* Introducción al paquete de R “compareGroups” para crear tablas descriptivas.

**Lugar:** Telemático

**Fechas y horas del curso:** El curso tendrá una duración total de 20 horas de clases distribuidas en 10 clases de 2 horas a lo largo de 5 semanas (2 clases por semana). Además se considerarán 10 horas de trabajo personal mediante tareas a resolver. Martes y jueves de 15:00h a 17:00h. Fechas 4, 6, 11, 13, 18,20, 25, 27 de Mayo y 1 y 3 de Junio 2021.

**Inscripciones:**

Mandar un correo electrónico a yferrer@imim.es incluyendo el nombre y los dos apellidos, el DNI o equivalente, institución y departamento, grupo CIBER al que pertenece en su caso y opcionalmente la formación.

**El precio de la inscripción para miembros de la CIBER es de 150€ y para el resto de participantes de 300€.**

**Materiales:** El alumno tendrá a su disposición las diapositivas con el código y los archivos de datos para reproducir los ejemplos que se irán comentando y explicando en clase.

**Requisitos:** Es recomendable que el alumno tenga conocimientos de estadística básica y que haya usado algún otro software como SPSS o Excel para el análisis de datos. No es necesario que haya visto el programa R ya que se explicará desde cero.