



Become a useR, not a loseR

Breve introducción

Actualmente las exigencias para lograr la publicación de un artículo científico en el ámbito médico han crecido exponencialmente y es necesario el conocimiento de nuevas herramientas para la gestión, análisis de datos y creación de tablas y gráficos.

Este curso es eminentemente teórico-práctico diseñado en 10 sesiones + 2 opcionales para servir como una introducción al lenguaje de R en aquellos usuarios que desconocen completamente su funcionamiento (4 primeras sesiones) y como una herramienta más avanzada para realizar análisis complejos, programación y creación de tablas para publicaciones médicas (sesiones 5 a 10).

Descripción del curso

R se trata de uno de los lenguajes de programación más usados en la Estadística, la Biomedicina, la Bioinformática y para el análisis de datos en general. Proviendo del lenguaje de programación S, en las últimas dos décadas R ha mostrado una popularización de su uso, tal y como demuestran los análisis de tendencias de Google. De este modo, el entorno de computación de R se ha visto favorecido por la aparición de librerías que facilitan diversos aspectos de su funcionamiento como son la gestión, análisis y visualización de datos. R se ha convertido en el lenguaje preferido de la Biomedicina, y actualmente cuenta con una comunidad muy activa de usuarios.

El propósito básico de este curso es dotar de una base sólida al usuario en esta herramienta informática, que le permita ejecutar sus propios análisis y profundizar en aquellos aspectos que crea convenientes. El curso aborda de manera exhaustiva todos aquellos aspectos que permiten entender y aprender la programación en R, focalizándose mucho en conseguir una interiorización de las estructuras básicas de computación al usuario, las cuales son extensibles a otros programarios o entornos informáticos. La enseñanza de estas estructuras, no obstante, se contextualizan en diferentes tareas o actividades habituales en el campo de la investigación biomédica. El objetivo final del curso es que el usuario sea capaz de explotar un *dataset* de forma eficiente y correcta usando los recursos principales que ofrece R.

Objetivos

Objetivos generales

- Dotar al alumno de independencia en el uso de R, para que pueda profundizar de manera autónoma en los tipos de análisis que requiera.
- Que al final del curso el usuario sea capaz de explotar una base de datos sanitaria.

Objetivos específicos

- Conocer las estructuras básicas de datos en R, y ser capaz de manipularlas y realizar operaciones con ellas de manera eficiente.
- Saber manipular adecuadamente los *data-frames*, estructuras orientadas a representar las bases de datos tal y como las conocemos. Esto incluye ser capaz de cargar bases de datos, filtrar-las, realizar consultas y hacer breves análisis descriptivos.



- Conocer las estructuras básicas condicionales y de iteración en R. Ser capaz, por lo menos, de programar un bucle *for*.
- Saber la sintaxis básica para realizar funciones.
- Conocer los pasos, y funciones asociadas a cada paso, para una correcta explotación de una base de datos.

Conocimientos Previos

No se requiere de ningún conocimiento previo de tipo informático para la realización de este curso. Es recomendable, no obstante, tener conocimientos básicos de estadística (saber qué es un *t*-test, correlación o ANOVA) o de uso de hojas de cálculo (Excel).

Modalidad

Se trata de un curso online. Se realizará una clase a la semana de 2 horas aproximadamente. En estas clases se intercalarán módulos teóricos y prácticas. Cada semana se hará llegar un correo al alumnado con el link de Zoom (o similar) y el programa de cada sesión.

Es muy importante practicar con los ejercicios que se proporcionaran en cada clase para una correcta interiorización de los conceptos teóricos trabajados. El alumno, como mínimo, debería realizar entre 2 y 5 horas de trabajo propio a la semana.

Material teórico y práctico

El curso dispone de material propio que se facilitará al alumnado. El canal de difusión del material está por definir, pero se realizará vía web o correo electrónico.

Necesidades informáticas

Para la realización del curso se requiere de un portátil u ordenador de sobremesa. Se recomienda que por lo menos se disponga de 4GB de RAM. Dicho ordenador deberá tener instalado los programas R y RStudio, así como todas las librerías necesarias para cada clase. Durante las semanas anteriores al curso, se facilitará un vídeo explicando cómo se debe configurar el equipo para la correcta realización del curso.

Duración, calendario y horario

Se realizará una sesión de dos horas (aproximadamente) a la semana, esperándose realizar un mínimo de 10 sesiones (2 meses y medio). El curso se iniciará el día 16 de enero de 2024, de 15:30 a 17:30. El día de la semana y horario de las sesiones siempre será los martes de 15:30 a 17:30. Todas las clases se grabarán y estarán disponibles para su revisión en aquellos casos en los que la persona no pudiera conectarse uno de los días.



Contenido del curso

El curso se estructura en clases teóricas y prácticas, según lo especificado en las siguientes páginas. De manera general, existe un conjunto de módulos teóricos básicos que, a modo de analogía, nos enseñaran las vocales, consonantes y gramática básica del lenguaje de programación de R. Después seguiremos con un módulo más orientado al análisis y a la explotación de datos que, siguiendo con la analogía, nos permitirá construir frases con lo que habremos aprendido en los temas anteriores. A parte, intercaladas con los temas teóricos citados, se realizarán diferentes prácticas que serán contextualizadas especialmente, aunque no únicamente, en el análisis de datos biomédicos. Finalmente, en el plan docente se incluyen algunos temas y bloques opcionales, que serán trabajados o excluidos en función de los intereses del alumnado y del transcurso del resto de módulos.

Módulos teóricos básicos

Tema 1: Tipos de objetos

Tipos de datos en R:

Valores atómicos.

Otros valores: NA, NULL, Inf.

Vectores:

Definición y características.

Operaciones matemáticas y lógicas.

Subsetting.

Matrices:

Definición y características.

Subsetting y dimensiones.

Operaciones matemáticas y lógicas.

Factores:

Definición y características.

Subsetting.

Operaciones con factores.

Data-Frames:

Definición y características.

Subsetting basado en operaciones lógicas.

Operaciones básicas con data-frames.

NAs en las data-frames (gestión de valores perdidos).

Listas:

Definición y características.

Subsetting.

Módulos teóricos avanzados



Tema 2: Control de flujo

Bloques condicionales:

Concepto y uso.

Tipologías: if, else, else if.

Loops:

Bucle for.

Bucle While.

Next & Break.

Funciones apply:

Sintáxis básica y uso.

Breve introducción a otras tipologías de apply (lapply, tapply, sapply).

Tema 3: Las funciones en R

Sintaxis básica.

Evaluación Perezosa.

Complicando la cosa:

Combinación con bloques if().

Instancia funciones.

Funciones anidadas.

Try y TryCatch.

Tema 4: Manipulación de Data-Frames

Ejemplo de estudio.

Limpieza:

Importar datos y limpieza inicial de datos.

Gestión de valores perdidos (NAs).

Visualización inicial, detectar tipos.

Manipulación:

Cálculo de variables

Estructura de los datos

Unión de bases

Introducción a la imputación.

Análisis de tendencias:

Descriptivo.

Función compareGroups.

Visualización rápida de tendencias.

Tema 5: Regresión (***Bloque Opcional***).

Concepto y objetivos.

Introducción a un caso de estudio.

Uso de la regresión en R.

Formulación.

Introducción de variables: factores vs continuas.

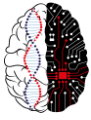
Manipulación de los outputs.

Como reportar la información.

Visualización.

Validación y revisión de modelos.

Módulos prácticos



Práctica 1: ¿Todo esto sirve para algo? ¿Cómo hacemos una correlación? Ejemplos de uso de vectores.

Práctica 2: ¿En que nos pueden ayudar las matrices y las operaciones de conjuntos a analizar unos resultados de una técnica ómica?

Práctica 3: ¿Cuál es el videojuego más vendido de la historia? Usando data-frames para responder a esta pregunta.

Práctica 4: ¿Cómo nos ayuda la iteración a hacernos más fácil nuestro día a día? ¿Cuáles son los inconvenientes de los tests de normalidad?

Práctica 5: Programar un análisis de expresión diferencial de proteínas con operaciones básicas de iteración. ¿Será muy difícil de realizar?

Práctica 6: ¿Cómo programamos el movimiento en un videojuego? Ejercicios básicos de funciones.

Práctica 7: Explotando los resultados de una base de datos (dataset todavía por elegir).

Organización de las sesiones (orientativo)

Módulos básicos

Sesión 0 (no se realiza clase, únicamente material).

- Instalación del programa.
- Instalación de librerías.

Sesión 1:

- Resolución de dudas en la instalación (o por correo).
- Introducción al curso.
- Tema 1-Parte 1.

Sesión 2:

- Práctica 1.
- Tema 1-Parte 2.

Sesión 3:

- Práctica 2.
- Tema 1-Parte 3.

Sesión 4:

- Práctica 3.

Módulos avanzados

Sesión 5:

- Tema 2.

Sesión 6:

- Práctica 4.
- Práctica 5 (conjuntamente).

Sesión 7:

- Tema 3.
- Inicio a tema 4 (si da tiempo).



Sesión 8:

- Práctica 6.

Sesión 9:

- Tema 4.

Sesión 10:

- Práctica 7.

Sesión 11 opcional:

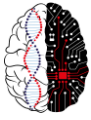
- Tema 5.

Sesión 12 opcional:

- Breve Introducción a *ggplot*.

FAQs

- 1- **¿Cuándo empieza el curso?** El 16 de enero (martes) de 15:30 a 17:30.
- 2- **¿Es un curso online?** Sí, es 100% online.
- 3- **¿Qué días se realizará?** El primer día es el 16. Después realizaremos una sesión semanal de 2 horas cada martes de 15:30 a 17:30.
- 4- **¿Cuántas sesiones se van a realizar?** Un total de 10 sesiones y dos opcionales, que las haremos o no en función de los intereses del alumnado, del progreso y de mi disponibilidad.
- 5- **¿Siempre será cada martes?** Sí. En principio no hay festivos que vayan a caer en martes en ese periodo, al menos del calendario nacional o Catalunya. Pero si alguna semana hay un problema ya iremos viendo.
- 6- **¿Qué pasa si una semana no vamos?** La enseñanza de R requiere constancia ya que es fácil perderse. Vamos a grabar las clases, así que quién no vaya un día podrá seguir el video.
- 7- **¿Cuánto cuesta?** Es gratuito. El curso lo he venido realizando de manera gratuita desde 2020. Y de momento no va cambiar.
- 8- **¿Hay certificado de asistencia?** Sí. Primer año que hay certificado de asistencia. Por favor, contactad con Lidia para más información al respecto.
- 9- **¿Habrá algún requisito para tener el certificado?** Todavía no se ha decidido, pero pondremos algún tipo de requisito de asistencia mínima.
- 10- **¿Cómo funciona el curso?** Hay sesiones teóricas en las que se imparte la clase mediante la programación directa compartiendo pantalla. Hay otras sesiones prácticas en las que el alumno ha realizado los ejercicios en su tiempo libre durante la semana. El modus operandi, en general, es muy práctico y enfocado al uso de R. Para más información de las sesiones, consultad el programa.
- 11- **¿Es suficiente con ir a las clases?** No. Para digerir la teoría hay que practicar. Es muy importante preocuparse por el R y practicar unas 3-5 horas semanales (como mínimo). Cuanto más se practique, pues mejor. La práctica consistirá en la lectura y ejecución del material teórico, así como la realización de ejercicios. Cualquier lenguaje de programación se aprende a base de práctica.
- 12- **¿Es un curso de estadística con R?** No. Es un curso de programación con R orientada a la biomedicina. Si bien se enseña a ejecutar análisis (por ejemplo, t-tests, correlaciones...), no se explican los fundamentos estadísticos de los mismos. También se enseña a explotar bases de datos, o los procesos implícitos en esto...pero se va al grano. No se explica la teoría detrás de estas técnicas. Se enseñan a hacer, pero siempre centrándonos mucho en la parte de programación en las clases teóricas. En las clases prácticas es donde lo contextualizamos en diversos campos, sobretodo la biomedicina.
- 13- **¿Vale la pena hacer un curso de R en programación? ¿Eso sirve de algo?** Depende de las necesidades de cada persona. Para mi tiene más sentido enseñar a usar la



herramienta de manera sólida y que después cada persona profundice en lo que necesita. No obstante, cada año hay personas que no lo van a usar en su trabajo pero que les despierta curiosidad, y a algunas les gusta y acaban profundizando en esto.

- 14- **¿Se termina el curso con una buena base de programación con R?** Sí. Si se siguen los ejercicios, clases y prácticas, el alumno estará capacitado para usar R y poder resolver sus dudas con las herramientas “on line” actualmente disponibles para los usuarios.
- 15- **¿Es muy difícil?** No, se iniciará por la parte más básica y tiene un diseño progresivo para adaptarse a cualquier perfil. La complejidad a la que se quiera llegar dependerá de cada persona. Está diseñado de manera que sea ameno y las clases son dinámicas.
- 16- **¿Se me da mal la informática, puedo hacer el curso?** Sí. Quizá descubres que no se te da tan mal.
- 17- **¿Se me da mal la estadística y las mates, puedo hacer el curso?** Sí. Quizá descubres que no se te da tan mal.
- 18- **¿Qué necesitamos para realizar el curso?** Un ordenador con internet y con RStudio. Las semanas antes del curso se enviará a los alumnos un video explicando como configurar el equipo. O, si lo preferís, le podéis pedir ayuda al equipo de IT de vuestro centro o a algún compañero/a que lo haya instalado antes (es fácil). Las dudas que existan se resolverán en la medida que lo permitan las clases.