

Introducción al lenguaje R + RStudio

I. ORIENTACIONES GENERALES

Fortalecimiento de capacidades del Departamento de Virología del INEI-ANLIS

II. ELEMENTOS DEL DISEÑO

Nombre de la actividad: Introducción al lenguaje R + RStudio

2. Fundamentación

El departamento de Virología del INEI requiere para el cumplimiento de sus funciones, ya sea como Laboratorios Nacionales de Referencia o como investigadores científicos, trabajar con datos con el fin de analizar y resumir la información y presentarla en informes técnicos, publicaciones científicas, reportes epidemiológicos, que, entre otros, contribuyan a la toma de decisiones en salud pública. El adecuado manejo de los sistemas de información requiere de recursos humanos calificados en el procesamiento y análisis de datos, capaces de identificar necesidades y problemas relevantes en la materia, y de lidiar con la complejidad creciente de los sistemas de información en salud. A su vez, es necesario dar respuesta a la demanda global de actualización en el análisis de datos, en paralelo a los adelantos tecnológicos y a la modernización de los procesos de trabajo. Actualmente es necesario dotar con estas competencias a los integrantes del Departamento de Virología, estandarizar sus procesos de trabajo aumentando su eficiencia, facilitar el intercambio de información útil y su visualización ante requerimientos externos al departamento, y mantener organizadamente el cúmulo de datos que genera y posee. También aplicar herramientas modernas como Quarto para la elaboración de informes técnicos o científicos, la creación de presentaciones gráficas interactivas en formato HTML, útiles para la construcción de salas de situación de salud, tanto coyunturales como de tendencias, así como también representar visualmente indicadores epidemiológicos y resultados científicos con precisión y claridad, identificar tendencias críticas en salud pública de forma oportuna y optimizar el flujo de trabajo mediante procesos reproducibles y automatizados. En este sentido, el Instituto Nacional de Epidemiología “Dr. Juan H. Jara” (INE) posee una amplia trayectoria en la formación de distintos profesionales de la salud pública en el área de epidemiología y ciencia de datos. Para continuar en este camino y dar respuesta a las demandas señaladas, se torna imprescindible plantear una capacitación que aborde los aspectos mencionados, promoviendo el uso de tecnologías innovadoras para el análisis y la gestión de información en

salud. La adopción de software gratuito y de código abierto como el lenguaje R y Quarto, un sistema de publicación técnica de código abierto, ofrece una solución moderna y eficiente para abordar las necesidades mencionadas. Estas herramientas permiten el tratamiento y manipulación de datos con alta reproducibilidad, una capacidad estadística avanzada y una forma de publicación versátil y adaptable a diversos contextos.

3. Contribución esperada

La actividad busca fortalecer la capacidad de trabajo con datos, desde los procesos de limpieza, hasta el análisis y comunicación de resultados. Asimismo, se espera potenciar la producción de informes técnicos y científicos, documentos estáticos y tableros dinámicos con información epidemiológica y científica, facilitando la toma de decisiones basadas en datos precisos y actualizados

4. Perfil del participante y requisitos

Perfil del participante: Integrantes del Departamento de Virología del INEI-ANLIS que deseen adquirir conocimientos para el procesamiento y análisis de datos con R/ RStudio.

Requisitos

Obligatorios: competencias intermedias en gestión de archivos y software en sistema operativo Windows. Utilización básica de Microsoft Excel. Conocimientos básicos de estadística descriptiva e inferencial.

Deseables: conocimientos básicos-intermedios de bases de datos y de lectura de idioma inglés.

Equipamiento necesario: Se recomienda contar con computadoras que tengan al menos procesadores I5, 8GB de Memoria RAM o superior y sistema operativo Windows 10/11 con conexión a internet, micrófono y parlantes para la realización de los sincrónicos y prácticas del curso.

5. Objetivos

Se espera que al finalizar este curso los participantes puedan:

- Comprender los fundamentos del lenguaje R
- Utilizar adecuadamente las herramientas de RStudio
- Importar y exportar bases de datos de diferentes formatos
- Entender y aplicar código de R base, tidyverse y otros paquetes específicos.
- Realizar análisis descriptivo utilizando funciones estadísticas
- Procesar y transformar datos utilizando funciones específicas
- Depurar, ordenar y unir tablas de datos.
- Realizar gráficos descriptivos
- Utilizar el lenguaje R para realizar análisis inferencial frecuentista paramétrico y no paramétrico.
- Producir informes y tableros mediante Quarto (HTML, Word, etc)

6. Contenidos

Unidad 1: Instalación e introducción al lenguaje R

1.1 Descarga e instalación de R, Rtools y RStudio

1.2 Introducción al lenguaje R: Scripts. Objetos. Funciones y argumentos. Librerías y dependencias. Tipos de datos. Estructuras de datos. Asignación. Errores y advertencias. Operadores. Tuberías. Filosofía tidyverse.

1.3 Rstudio: Característica del entorno de desarrollo integrado. Proyectos. Consola. Memoria. Historial. Ayuda. Paquetes. Atajos de teclado

1.4 Sintaxis en R. Scripts. Documentación. Hoja de estilo del lenguaje. Buenas prácticas

1.5 Lectura de datos. Texto plano, Excel y otros formatos específicos.

Unidad 2: Exploración, diagnóstico y limpieza de datos

2.1 Estructura de datos. Valores perdidos (NA)

2.2 Exploración y diagnóstico con paquetes skimr y dlookr

2.3 Manejo de duplicados. Limpieza de datos

2.4 Diagnóstico de valores perdidos. Paquete naniar

2.5 Exportación de datos limpios en diferentes formatos.

Unidad 3: Procesamiento de datos

3.1 Gestión de datos con tidyverse. Paquete dplyr.

3.2 Uniones de tablas de datos. Familia de join's y bind's

3.3 Datos “tidy” y “no-tidy”. Pivoteos. Paquete tidyr.

Unidad 4: Tratamiento de datos específicos

4.1 Variables de tiempo. Librería Lubridate.

4.2 Cadenas de caracteres y factores. Paquete stringr y forcats

Unidad 5: Estadística descriptiva

5.1 Cálculos: Proporciones, razones y tasas.

5.2 Resumen estadístico. Medidas de tendencia central, posición y dispersión.

5.3 Cálculos masivos. Función across() y rowwise()

5.4 Paquete rstatix y gtsummary.

Unidad 6: Inferencia estadística

6.1 Estimaciones con intervalos de confianza.

6.2 Test de bondad de ajuste. Normalidad. Homocedasticidad.

6.3 Test de hipótesis paramétricas y no paramétricas.

6.4 ANOVA. Test de comparaciones múltiples

Unidad 7: Visualización de datos

7.1 Gramática de gráficos con ggplot2. Gráficos estadísticos.

7.2 Capas geométricas de líneas, puntos, barras, histogramas y boxplots. Facetas

7.3 Pirámides poblacionales. Curvas epidémicas y corredores endémicos.

7.4 Exportación en diferentes formatos (raster y vectorial)

Unidad 8: Comunicar con RStudio

8.1 Quarto. Paquete tinytext.

8.2 Lenguaje de marcas markdown. Cabeceras YAML. Paquete flextable. Personalización estética

8.3 Documentos estáticos PDF y Word

8.4 Páginas HTML y tableros dinámicos

7. Estrategias metodológicas y recursos didácticos

Realización de trabajos prácticos (TP) semanales.

Se utilizará de apoyo un aula virtual alojada dentro de la plataforma educativa moodle de ANLIS. Actividades de desarrollo: cada semana se abordará una unidad, en un primer encuentro sincrónico el docente explicará los puntos temáticos, en un segundo encuentro se llevará a cabo la práctica relacionada. Las clases teóricas y prácticas serán virtuales, de una duración aproximada máxima de dos horas, mediante Zoom, que luego quedarán disponibles en la plataforma de aprendizaje. Cada unidad se acompañará de materiales de apoyo, prácticas y recursos en el aula virtual. Las actividades serán monitoreadas por el docente a través de la plataforma educativa, a la vez que ofrecerá apoyo a través de un foro de consultas.

8. Descripción de la modalidad

Virtual

9. Bibliografía para el participante

Hadley Wickham, Mine Çetinkaya-Rundel, and Garrett Grolemund (2023) R for Data Science (2e). Disponible en <https://r4ds.hadley.nz/>

Hadley Wickham, Danielle Navarro, and Thomas Lin Pedersen (2023) ggplot2: Elegant graphics for data analysis (3e). <https://ggplot2-book.org/>

Tipos de gráficos <https://www.data-to-viz.com/>

Neale Batra y otros (2022) R para epidemiología aplicada y salud pública. <https://epirhandbook.com/es/index.html>

10. Evaluación de los aprendizajes

Se requerirá la entrega y aprobación de 7 trabajos prácticos.

Los trabajos prácticos evaluativos parciales serán individuales y estarán relacionados a las unidades 1 a la 7. Se publicará una consigna en el aula virtual con puntos a resolver mediante código en lenguaje R que deberán entregar a la semana siguiente.

La evaluación final consistirá en un trabajo práctico grupal que consistirá en procesar datos crudos y obtener un producto final con una serie de resultados a visualizar con tablas y gráficos. Los pasos a considerar en el trabajo práctico final son:

Creación de un proyecto RStudio

Lectura de datos

Exploración y diagnóstico.

Formateo y depuración de datos crudos a datos limpios

Obtención de resúmenes descriptivos y creación de nuevas variables

Elaboración de tablas y gráficos

Creación de documento o tablero dinámico

11. Instrumentos para la evaluación

Instrumentos para la evaluación de los aprendizajes

Registro de Asistencia de las/os participantes. Grilla para la corrección de TPs. Grilla para la corrección del TP grupal final Instrumentos para la evaluación de la actividad Encuesta de satisfacción de la actividad de carácter anónimo

12. Perfil del instructor

Especialista en Sistemas – Especialista de Datos con orientación a la epidemiología y salud pública

13. Requisitos de asistencia y aprobación

Para la aprobación se requiere asistencia en el 80% de las clases sincrónicas (teóricas y prácticas). Aprobación de la totalidad de TP's planteados y del TP integrador final.

14. Duración en horas

El curso tendrá una duración total de 80 horas distribuidas a lo largo de 10 semanas.

15. Detalle de la duración

32 horas sincrónicas distribuidas en 16 encuentros de 2 horas cada uno (2 veces por semana durante 8 semanas) y 48 horas asincrónicas, calculadas 4 horas por semana durante las 8 semanas del curso para la lectura del material (1 hora), la

realización de los trabajos prácticos (3 horas) y la elaboración del TP integrador final (16 horas)

16. Lugar

Las clases virtuales se realizarán a través de la plataforma Zoom.

Se contará con un campus virtual para la publicación de material y recursos, consultas y entrega y corrección de TPs: <http://capitacion.anlis.gob.ar/>

17. Cronograma tentativo

Comienzo xx y finalización xx de 2026. Encuentros sincrónicos teóricos: Encuentros sincrónicos prácticos: Práctico integrador final:

18. Prestador

Christian Ballejo. Especialista de Datos del Servicio de Informática y Estadística del Departamento de Investigación Epidemiológica del Instituto Nacional de Epidemiología / Administración Nacional de Laboratorios e Institutos de Salud “Dr. Carlos G. Malbrán”.