

INTRODUCCIÓN A LA PROGRAMACIÓN Y A LA DESCRIPCIÓN Y VISUALIZACIÓN DE DATOS

INSTITUTOS/S: **INSTITUTO DE TECNOLOGÍA E INGENIERÍA**

CARRERA/S: **ESPECIALIZACIÓN EN ESTADÍSTICA APLICADA
EN SALUD Y EDUCACIÓN**

RESPONSABLE DE LA ASIGNATURA Y EQUIPO DOCENTE:
MARTIN PUSTILNIK (RESPONSABLE)

AÑO: 2025

CRÉDITOS: 5

CARGA HORARIA DE INTERACCIÓN PEDAGÓGICA: 54

CARGA HORARIA TOTAL: 175

1. Fundamentación

El presente curso se inscribe dentro del primer cuatrimestre de la carrera de Especialización en Estadística Aplicada en Salud y Educación. Este curso integra conocimientos de distintas ramas del saber, para profesionales formados en distintos campos del conocimiento, tales como como: conocimientos básicos de programación, pre-procesamiento de datos y estadística descriptiva.

Esta asignatura considera que las aplicaciones de la estadística descriptiva al campo de la salud y la educación constituyen un área de vacancia nacional y una oportunidad regional, que se desarrolla en un diálogo entre el sector público y el privado, así como entre la investigación básica, la aplicada, y su compleja traslación a la industria.

En particular, fortalecer la formación en estadística de sus egresados de grado en Salud Comunitaria y Educación para que puedan dedicarse a la investigación en sus respectivas áreas.

2. Propósitos y/u objetivos

El seminario busca que el estudiante adquiera conocimiento y competencias en el uso de herramientas informáticas disponibles para la producción de datos estadísticos de calidad tanto en las etapas de carga, como de procesamiento, de análisis y de visualización de datos.

Propósitos

Promover la comprensión de la naturaleza de los datos y tipos de datos.

Contribuir al análisis del problema desde el punto de vista de estadística descriptiva: ¿cómo pre-procesar los datos?, ¿cómo elegir qué gráfico utilizar? ¿cómo elegir el algoritmo que mejor se adecúa?

Identificar las métricas que permitan comparar los datos de acuerdo a la pregunta que se desea responder: ¿La distribución empírica de un conjunto de datos es similar o diferente a otro conjunto?

Objetivos

Son objetivos particulares que el estudiante pueda:

Instalar y manejar las herramientas básicas del entorno de programación R y su interfaz R-Studio, comprendiendo las estructuras de datos fundamentales y la creación de algoritmos simples para manipular datos.

Importar y exportar datos en múltiples formatos desde y hacia R, integrando bases de datos provenientes de diversas fuentes.

Preparar bases de datos para análisis estadísticos, realizando tareas de limpieza, transformación y recodificación de datos para asegurar su calidad y estructura adecuada antes del análisis.

Aplicar métodos de estadística descriptiva con R para describir y analizar datos.

Visualizar datos utilizando gráficos (histogramas, boxplots, entre otros) y personalizarlos utilizando las funciones de ggplot en R.

Realizar simulaciones y estudios de Montecarlo para estimar probabilidades y analizar resultados asintóticos.

Generar reportes reproducibles mediante R-Markdown para integrar código, resultados y texto explicativo en un solo archivo.

3. Contenidos mínimos:

Creación de bases de datos. Manipulación y visualización de datos. Introducción al lenguaje de programación R. Nociones elementales de programación Objetos en R: data frames y listas. Importación y exportación de un data frame desde distintos formatos de archivo. Análisis exploratorio y descriptivo. Medidas de posición y dispersión, histogramas, boxplots. Construcción y edición de gráficos. Semántica de gráficos y visualizaciones con ggplot. Producción de reportes mediante

R-Markdown. Extracción de muestras con y sin reposición. Estudios de simulación de Montecarlo para estimar probabilidades, describir la Ley de los Grandes Números y el Teorema Central del Límite, entre otros resultados asintóticos.

4. Carga Horaria

Este curso cuenta con una carga horaria de interacción teórica de 30 horas, más 26 horas de interacción práctica. Se consideran asimismo 119 horas de trabajo autónomo que, en total, dan una carga total de 175 horas, otorgando 7 créditos. En esta carga horaria se incluye la correspondiente a la instancia de evaluación.

<i>Créditos</i>	<i>Interacción pedagógica</i>	<i>Trabajo autónomo</i>	<i>TOTAL</i>
7	56	119	175

4.1. Trabajo autónomo de la/el estudiante

La carga horaria de trabajo autónomo prevista para este curso (119 horas) se reparte entre la lectura de la bibliografía y material didáctico obligatorios, la realización de la guía de ejercicios, la realización de los de los trabajos prácticos, así como la preparación para la evaluación final. El contenido se resume en la siguiente tabla:

Actividad	Carga horaria
Lectura de la bibliografía y material didáctico	29 hs
Realización de Trabajos prácticos domiciliarias	30 hs
Realización de Guía de ejercicios	30 hs
Preparación de la evaluación final.	30 hs

5. Programa analítico

Organización del contenido:

Las unidades temáticas se estructuran siguiendo una construcción gradual y espiralada de los conceptos tratados.

Unidad 1: Introducción a R y R-Studio

- Presentación de la familia de software R y R-Studio.
- Tipos de archivos y extensiones. Importación de datos.
- Ventanas del programa: datos, sintaxis y resultados.
- Construcción de algoritmos y procesos.
- Instalación y uso de paquetes.
- Creación, atributos y manipulación de objetos: vectores, factores, arreglos, matrices, data.frames, series temporales, listas.
- Producción de reportes mediante R-Markdown.

Unidad 2: Preparación de los datos

- Preparación de la base de datos para el tratamiento estadístico.
- Operaciones entre variables y selección de observaciones.
- Recodificaciones, renombre y transformaciones.

- Pegado de bases de datos (merge).
- Exportación e integración con otros softwares.

Unidad 3: Introducción a la estadística descriptiva con R

- Población y unidad de análisis.
- Tipos de variables y nivel de medición.
- Descripción de variables categóricas.
- Distribución de frecuencias.
- Tablas de doble entrada.
- Razones, tasas e índices.
- Representación de variables categóricas.
- Resumen de variables cuantitativas: medidas de tendencia central y de variación.

Unidad 4: Técnicas de visualización de datos con R

- Construcción y edición de gráficos.
- Semántica de gráficos y visualizaciones con ggplot.
- Histogramas.
- Boxplots.

Unidad 5. Simulación con R

- Extracción de muestras con y sin reposición.
- Estudios de simulación de Montecarlo para: estimar probabilidades, describir la Ley de los Grandes Números y el Teorema Central del Límite, entre otros resultados asintóticos.

6. Bibliografía y recursos

6.1 Bibliografía obligatoria

- Kelmansky, D. M. (2009). Capítulos 16 a 20 y 23 en: Estadística para todos. Buenos Aires: Ministerio de Educación, Instituto Nacional de Educación Tecnológica.
- Sitio web de la versión en español de "R for Data Science", de Hadley Wickham y Garrett Golemund: <https://es.r4ds.hadley.nz/>.
- Contento Rubio, M. (2019). Estadística con aplicaciones en R. Bogotá: Utadeo.
- Apezteguía, M. C., y Ferrario, J. (coord). (2019). Probabilidades y Estadística Análisis de datos. La Plata: EDULP.
- Ross, S. M. (2018). Introducción a la estadística. Barcelona: Reverte.
- Dalgaard, P. (2008). Capítulos 1 a 4 en: Introductory Statistics with R (2nd ed). New York: Springer.
- Moore, D. S., Mc Cabe, G. P. y Craig, B. A. (2014). Introduction to the Practice of Statistics. New York: W. H. Freeman and Company.

- Holcomb, Z. C., y Cox, K. S. (2022). Capítulos 1 a 3 en: Interpreting basic statistics: A workbook based on excerpts from journal articles (ninth edition). New York: Routledge Taylor & Francis.

6.2. Bibliografía optativa:

- Jorge Zappino J. (2020). Manual de estadística básica para no estadísticos. ISSN 2683-9644. CUINAP:
https://www.argentina.gob.ar/sites/default/files/cuinap_13_2020_0.pdf

7. Destinatarios/requisitos de ingreso:

Se requerirá a quienes se inscriban a este curso un título de grado universitario con título de grado en el área de salud o educación correspondiente a una carrera de cuatro (4) años de duración como mínimo, o graduado de otra universidad extranjera, con título equivalente.

8. Descripción de las actividades prácticas desarrolladas en la actividad curricular, indicando lugar donde se desarrollan, modalidad de supervisión y de evaluación.

El seminario se desarrolla en clases teórico-prácticas, en las que se presentan los temas conceptuales y teóricos de las distintas unidades de modo que le permita al estudiante comprender y ser capaz de abordar la resolución de la ejercitación práctica donde éstos se aplican. Las actividades prácticas son realizadas en la UNaHur y en instancias de trabajo autónomo fuera de la cursada. Tales prácticas consisten en la aplicación de las técnicas de procesamiento, descripción y visualización de datos afines a temas de salud y educación. Como actividades prácticas, los estudiantes programan los algoritmos necesarios para aplicar los procedimientos requeridos para los problemas que se tratan; también utilizan funciones ya incorporadas al R por sus desarrolladores.

Se utiliza un lenguaje de programación enfocado al análisis estadístico en un entorno amigable (R en R-Studio). Se propone el uso de software libre por sus reconocidas ventajas, como la libertad de uso y redistribución, la compatibilidad a largo plazo y el soporte existente en las múltiples comunidades en línea. Más allá de estas ventajas, R es un lenguaje de programación con actualización permanente en cuanto a librerías y

recursos en el área de estadística teórica y aplicada, y con comunidades de investigadores y usuarios profesionales que son muy activos en cuanto a su uso, desarrollo y popularización.

9. Condiciones de cursada y requisitos de aprobación

La cursada se desarrollará y aprobará de acuerdo a los siguientes requisitos:

9.1 Modalidad de evaluación

El curso contará con una evaluación final sobre contenidos teóricos y prácticos, en modalidad a libro abierto, que comprenderá la integración de conocimientos de todas las unidades teóricas tanto como de lo realizado en el trabajo práctico.

9.2 Aprobación de la cursada

Por un lado, la aprobación de las actividades curriculares bajo el régimen de regularidad requerirá una asistencia no inferior al setenta y cinco por ciento (75%) en las clases presenciales y la asistencia a la instancia de trabajo práctico. Asimismo, la evaluación final deberá aprobarse con 7 puntos o más, sobre un total de 10.

9.3 Acreditación del curso

La acreditación de los cursos de Posgrado, según nuevo reglamento

- a) Aprobó la asignatura (de 7 a 10 puntos)
- b) Reprobó la asignatura (0 a 6 puntos)
- c) Ausente

Se considerará ausente a aquel/lla estudiante que no cumpla con el porcentaje de asistencia o no se haya presentado a las instancias de evaluación pautadas en el Programa del curso.

10. Docente responsable del curso.

Martin Pustilnik es Lic. en Ciencias de la Computación por la Universidad de Buenos Aires y Especialista en Explotación de Datos y Descubrimiento del Conocimiento por la

misma universidad. Es profesor adjunto en la Universidad Nacional de Hurlingham. Es investigador en el Laboratorio de Investigación y Desarrollo Experimental en Computación (LIDEC) de la Universidad Nacional de Hurlingham.

Se adjunta el Currículum en formato CvAr