



## SÍLABO

### I. INFORMACIÓN GENERAL

<b>Nombre del curso</b>	: Fundamentos de R para Ciencias Sociales y Gestión Pública
<b>Modalidad</b>	: Sincrónica
<b>Número de horas</b>	: 16.5 horas
<b>Horario de clases</b>	: Sábado 10:00 - 11:30 am Domingo 8:30 - 11:30 am
<b>Docente</b>	: Christian Chiroque Ruiz
<b>Correo electrónico</b>	: christian.chiroque@pucp.edu.pe

### II. SUMILLA

El curso es de carácter teórico-práctico y tiene como objetivo brindar una introducción al uso del software estadístico R (a través del IDE R Studio) aplicado para la ciencia de datos en el marco de la gestión de políticas públicas o proyectos de investigación en ciencias sociales.

### III. PRESENTACIÓN

En las últimas dos décadas, el desarrollo tecnológico exponencial ha generado un aumento significativo en la cantidad y variedad de información disponible. Esto presenta nuevos desafíos y oportunidades para los tomadores de decisiones en la administración pública y los investigadores en diversos campos académicos. En este contexto, es imprescindible que los profesionales en ciencias sociales y gestión pública dominen no solo los conocimientos tradicionales de estadística, sino también las habilidades para utilizar eficientemente los avanzados recursos informáticos y técnicas computacionales actuales. Sumado a ello, la irrupción de herramientas como la inteligencia artificial generativa y el acceso a volúmenes masivos de datos, la capacidad de análisis y procesamiento de información se ha vuelto más crucial que nunca.

Este curso pretende brindar una introducción para el uso del software R desde la perspectiva de la ciencia de datos (data science). Para ello, se divide el curso en tres ejes de trabajo. El primero está orientado a estudiar las definiciones y los fundamentos de programación que sustentan el funcionamiento del software R y los distintos algoritmos que se utilizarán en el curso. El segundo eje está dirigido a brindar herramientas para la obtención, manipulación y transformación de los datos. Finalmente, el tercer eje está destinado a estudiar los fundamentos para la visualización de datos y elaboración de reportes analíticos.

#### IV. OBJETIVOS DE APRENDIZAJE

Al finalizar el curso el estudiante deberá comprender los fundamentos de programación requeridos para cada uno de los ciclos básicos de un proyecto típico de ciencia de datos en R, comprendiendo la importación, manipulación, transformación, visualización y comunicación de la información.

#### V. CONTENIDO DEL CURSO

1. Introducción a R y Principios Básicos de Programación
2. Análisis exploratorio de datos
3. Manipulación de datos
4. Visualización de datos
5. Transformación de datos

#### VI. TEMARIO

N°	TEMA	CONTENIDO
1	Introducción a R y Principios Básicos de Programación	<ul style="list-style-type: none"><li>• Presentación de R y RStudio: interfaz, paquetes, ayuda y recursos.</li><li>• Conceptos básicos de programación en R: tipos de datos, operadores, variables.</li><li>• Estructuras de datos en R: vectores, listas, data frames.</li><li>• Instalación y manejo de paquetes básicos y funciones.</li><li>• Diferencias entre Script (.r), R Markdown (.rmd) y Quarto (.qmd).</li><li>• Utilización de plantillas.</li></ul>
2	Análisis exploratorio de datos	<ul style="list-style-type: none"><li>• Importación y exportación de datos: lectura de CSV, Excel, y otros formatos. Lectura desde la nube.</li><li>• Exploración de datos: funciones para obtener resúmenes y descripciones básicas.</li><li>• Conceptos de análisis exploratorio de datos: Estadística descriptiva: medidas de tendencia central y dispersión.</li><li>• Creación de tablas simples y cruzadas con `gtsummary`.</li></ul>
3	Manipulación de datos	<ul style="list-style-type: none"><li>• Introducción al Tidyverse</li><li>• Manipulación de datos con dplyr: selección (select), filtrado (filter), ordenamiento (arrange), creación de nuevas variables (mutate) y resúmenes (summarise).</li><li>• Agrupación de datos y operaciones por grupo (group_by).</li><li>• Uso de pipes ( &gt;) para crear flujos de trabajo claros y legibles.</li></ul>
4	Visualización de datos	<ul style="list-style-type: none"><li>• Creación de gráficos básicos con ggplot2: histogramas, gráficos de barras, diagramas de cajas, gráficos de dispersión.</li><li>• Interpretación de gráficos: entender qué nos dice cada tipo de visualización sobre nuestros datos.</li><li>• Personalización de gráficos: ajustes de ejes, títulos, leyendas y temas.</li><li>• Uso de gráficos para comunicar resultados en el contexto de ciencias sociales y políticas públicas.</li></ul>

5	Transformación de tablas	<ul style="list-style-type: none"> <li>• Edición de tablas de datos: <i>Merge</i>, <i>appending</i> y <i>aggregating</i>.</li> <li>• Preparación de datos para análisis: creación de variables categóricas, codificación</li> <li>• Ejercicios prácticos.</li> </ul>
6	Trabajo Final	Presentación de proyecto utilizando todos los conceptos aprendidos

## VII. METODOLOGÍA

### A. ASPECTOS GENERALES

- **Sesiones sincrónicas:** Las sesiones prácticas se realizarán a través del programa ZOOM para la realización de videoconferencias, según el horario establecido. El docente enviará con anticipación un correo a todos los alumnos, el cual contendrá la convocatoria y el link para poder participar en la sesión.
- **Software requerido:** Se hará uso del lenguaje de programación [R](#) y del entorno de desarrollo integrado [R Studio](#) (de forma opcional podrán utilizar la plataforma web basado en la nube [R Studio Cloud](#)). Los recursos mencionados se encuentran dentro de la categoría de software libre, por lo que no tienen costo.
- **Contenido de la clase:** Durante la sesión se expondrá algunos puntos teóricos necesarios para el uso de las herramientas a analizar en el lenguaje de programación R. Así también, comprenderá la realización de ejercicios por parte de los alumnos con la dirección del docente. A fin de poder hacer seguimiento a lo realizado por los estudiantes se utilizará la función de “compartir pantalla”.
- **Atención de dudas y consultas:** El envío de comunicaciones al docente será a través del correo institucional de la universidad.

### B. FUENTE DE DATOS

- Se utilizarán diversas bases de datos para el análisis y los ejercicios de clase. Resaltan los siguientes:

N°	Fuente de data	Link
1	Kaggle: Your Machine Learning and Data Science Community	<a href="#">Link</a>
2	UCI Machine Learning Repository	<a href="#">Link</a>
3	Banco de datos del Banco Mundial	<a href="#">Link</a>
4	Índice de Democracia	<a href="#">Link</a>
5	Índice de Percepción de la Corrupción	<a href="#">Link</a>
6	Índice de Desarrollo Humano	<a href="#">Link</a>
7	Organización Mundial de la Salud – Banco de Datos	<a href="#">Link</a>
8	UNICEF – Datasets	<a href="#">Link</a>
9	Organización Internacional del Trabajo – Datasets	<a href="#">Link</a>

## VIII. SISTEMA DE EVALUACIÓN

La evaluación del curso está compuesta por lo siguiente:

Nº	Tipo de Evaluación	Ponderación sobre la nota final
1	Ejercicios prácticos	30%
2	Participación en clase	20%
3	Proyecto final	50%

### Ejercicios prácticos (30%):

- Los ejercicios prácticos son evaluaciones asincrónicas que se realizan a través de la plataforma PAIDEA, de forma individual.

### Participación en clase (20%):

- La participación en clase es evaluada considerando tanto la frecuencia como la calidad de las contribuciones de los estudiantes, de forma individual. Se realizará utilizando plataformas para realizar evaluaciones rápidas al finalizar cada semana.

### Trabajo grupal (50%):

- Se realizará de forma grupal.
- Cada grupo seleccionará un tema de interés y aplicará a una base de datos de su elección cada uno de los temas vistos en el curso: importación, manipulación y transformación de datos, visualización y elaboración de un reporte analítico.
- Al finalizar el curso cada grupo entregará un código en R Markdown (que deberá poder ser reproducido por el docente) así como un reporte (HTML).
- Importantes: para garantizar una evaluación justa y transparente, cada grupo deberá presentar, al final de su proyecto, un breve informe de contribuciones individuales (no más de 10 líneas). Este informe debe detallar las responsabilidades y tareas específicas asumidas por cada miembro del equipo. Si se diera el caso de que un miembro no ha participado de manera adecuada o ha habido una falta significativa de contribución, esto debe ser comunicado explícitamente en el informe.

## IX. BIBLIOGRAFÍA

Agresti, Alan y Christine Franklin (2013) *Statistics. The art and science of learning from data*. Pearson.

Cielen, Davy, Meysman A. y Mohamed A. (2016) *Introducing Data Science. Big Data, Machine learning and more, using Python Tools*. Manning, Shelter Island

R Project (2000) Introducción a R. Basado en *Notas sobre R: Un entorno de programación para Análisis de Datos y Gráficos*. [Link](#)

Urdinez, F. y Cruz, A. (Eds.) (2020). *Political Data Science Using R: A Practical Guide*. CRC Press.

Wickham, Hadley and Garrett Golemund (2016) *R for Data Science. Visualize, model, transform, tidy and import data*. O'reilly. [Link](#).

Wickham, Hadley (2019) *The tidyverse style guide*.

\_\_\_\_\_ (2016). *ggplot2: Elegant Graphics for Data Analysis*. useR. Springer.

\_\_\_\_\_ (2014). "Tidy Data." *The Journal of Statistical Software* 59.

Wickham, Stefan Milton Bache and Hadley (2014). *magrittr: A Forward-Pipe Operator for R*

## **X. Información adicional**

La evaluación de todos los trabajos contemplará el respeto de los derechos de autor. En este marco, cualquier indicio de plagio tendrá como consecuencia la nota cero. Esta medida es independiente del proceso disciplinario que la Secretaría Académica de la facultad estime iniciar según cada caso. Para obtener más información sobre el citado visitar el siguiente sitio web:

[www.pucp.edu.pe/documento/pucp/plagio.pdf](http://www.pucp.edu.pe/documento/pucp/plagio.pdf)