



SÍLABO

I. INFORMACIÓN GENERAL

Nombre del curso	: R intermedio para Ciencias Sociales y Gestión Pública
Modalidad	: Sincrónica
Número de horas	: 24 horas
Profesor del curso	: Christian Chiroque Ruiz
Correo electrónico	: christian.chiroque@pucp.edu.pe
Horario de clases	: Domingos de 2 pm a 5 pm

II. SUMILLA

El curso es de carácter teórico-práctico y tiene como objetivo profundizar en el uso del software estadístico R (a través del IDE R Studio) para el análisis de datos en el marco de políticas públicas o proyectos de investigación en ciencias sociales.

III. PRESENTACIÓN

En los últimos 20 años el creciente desarrollo tecnológico ha tenido su correlato en la disponibilidad de información que tienen tanto los tomadores de decisión dentro de la administración pública, como los investigadores en diversos campos académicos. Este cambio hace necesario que el profesional en ciencias sociales no sólo deba contar con conocimientos propios de estadística clásica (sea descriptiva o inferencial), sino también habilidades para utilizar recursos informáticos y técnicas computacionales dentro de procesos de investigación y toma de decisiones.

Este curso pretende ahondar en el uso del software R (a través del IDE R Studio) desde la perspectiva de la ciencia de datos o data science. Para ello, se divide el curso en cuatro ejes de trabajo: 1) la aplicación de técnicas de reducción de dimensiones; 2) la aplicación de técnicas de reconocimiento de patrones de datos; y 3) creación de dashboards interactivos para el análisis de datos; 3) introducción a la minería de texto.

IV. OBJETIVOS DE APRENDIZAJE

Al finalizar el curso el estudiante deberá estar en condiciones para afrontar cualquier proyecto de ciencia de datos con un nivel intermedio de dificultad, comprendiendo también tópicos como la reducción de dimensiones, reconocimiento de patrones de datos, la creación de plataformas web interactivas para análisis de datos y el manejo de data textual.

V. CONTENIDO DEL CURSO

1. Técnicas de reducción de dimensiones
2. Técnicas de reconocimiento de patrones de datos
3. Creación de aplicaciones web con Shiny
4. Introducción a la Minería de Texto

VI. CRONOGRAMA

Sesión N°	Tema	CONTENIDO
1	<i>Técnicas de reducción de dimensiones I</i>	<ul style="list-style-type: none">• Introducción a las técnicas de reducción de dimensiones.• Análisis de componentes principales
2	<i>Técnicas de reducción de dimensiones II</i>	<ul style="list-style-type: none">• Análisis factorial exploratorio.
3	Técnicas de reconocimiento de patrones de datos I	<ul style="list-style-type: none">• Introducción a técnicas no supervisadas y clustering.• Algoritmos de análisis clúster basados en partición: k-means (centroides); PAM y CLARA (medoides).• Algoritmos basados en densidad: DBSCAN.
4	Técnicas de reconocimiento de patrones de datos II	<ul style="list-style-type: none">• Introducción a la detección de anomalías.• Algoritmos de detección de anomalías: Isolation Trees e Isolation Forest.
5	Paquete Shiny y creación de aplicaciones web I	<ul style="list-style-type: none">• Introducción al paquete <i>Shiny</i> y <i>Shinydashboard</i>: antecedentes y aplicaciones.• Arquitectura básica del código <i>Shiny</i> para creación de aplicaciones (<i>ui</i> y <i>server</i>).• Detalle de los elementos de la interfaz de usuario.• Detalle de los elementos del servidor.
6	Paquete Shiny y creación de aplicaciones web II	<ul style="list-style-type: none">• Herramientas complementarias en la creación de Dashboards con Shiny. Uso de módulos y paneles.• Utilización de <i>Shinyapps</i> para alojar aplicaciones web en la nube.
7	Introducción a la minería de Texto	<ul style="list-style-type: none">• Análisis de texto y manipulación de strings (paquete <i>stringr</i> - <i>Tidyverse</i>): uso de expresiones regulares.• Ejercicios: Análisis de redes sociales (twitter) y discursos presidenciales.

VII. METODOLOGÍA

A. ASPECTOS GENERALES

- **Sesiones sincrónicas:** Las sesiones prácticas se realizarán a través del programa ZOOM para la realización de videoconferencias y tendrán una duración de 3 horas, 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](#)). También es necesario que el estudiante cree una cuenta en la plataforma de desarrollo colaborativo [Github](#). Los recursos mencionados se encuentran dentro de la categoría de software libre.
- **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 única y exclusivamente a través del correo institucional de la universidad.

VIII. SISTEMA DE EVALUACIÓN

La evaluación del curso está compuesta por lo siguiente:

Nº	Tipo de Evaluación	Ponderación sobre la nota final
1	Dos ejercicios prácticos	60%
2	Un trabajo grupal	40%

Ejercicios prácticos (60%):

- Los ejercicios prácticos son evaluaciones individuales asincrónicas que se realizan a través de la plataforma PAIDEA.
- El promedio de los dos ejercicios prácticos equivale al 60%.

Trabajo grupal (40%):

- El trabajo grupal equivale a 40% de la nota final.
- Cada grupo seleccionará un tema de interés e implementará un *dashboard* con el paquete Shiny. Al finalizar el curso cada grupo entregará un código (R Notebook) y deberá poder ser reproducido por el docente, así como el aplicativo operativo.

IX. BIBLIOGRAFÍA

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

Chiroque, Christian (2020) *Manual de R Studio para Ciencias Sociales*. Lima: Data Política. [Link](#)

King, Ronald S. (2015) *Cluster Analysis and Data Mining. An Introduction*.

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

Silge, Julia (2016) *Text Mining with R: A Tidy Approach*. [Link](#).

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

Wickham, Hadley and Garrett Grolemund (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

Conforme a los lineamientos establecidos por el Ministerio de Educación y la Superintendencia de Nacional de Educación Superior Universitaria (SUNEDU) dictados en el marco de la emergencia sanitaria para prevenir y controlar el COVID-19, la universidad ha decidido iniciar las clases bajo la modalidad virtual hasta que por disposición del gobierno y las autoridades competentes se pueda retornar a las clases de modo presencial. Esto involucra que los docentes puedan hacer los ajustes que resulten pertinentes al sílabo atendiendo al contexto en el que se imparten las clases.

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