

**Programa de estudios
ANÁLISIS AVANZADO DE DATOS I**

I. Identificación.

Código: SOC8224-01 02
Créditos: 5
Período Académico: 2º semestre 2024
Requisito: Estadística II
Duración: Semestral
Horario: Lunes 08:30-09:50 // Plenaria
Lunes 10:00-11:20 // Taller grupo 1
Lunes 11:30-12:50 // Taller grupo 2
Sala: Plenaria: Sala 202, Ejército 233
Talleres: Laboratorio de computación 2do piso,
Facultad de CCSS, Ejército 333
Asistencia: Obligatoria (70%)
Docente: Gabriel Sotomayor // gabriel.sotomayor@mail.udp.cl
Ayudantes: Patricio Alarcón // patricio.alarcon@mail.udp.cl
Francisca Hernández // francisca.hernandez_c@mail.udp.cl
Fernanda Hurtado // fernanda.hurtado@mail.udp.cl

II. Presentación

El curso aborda las principales técnicas para establecer relaciones entre variables. Se enfatizan la comprensión de procedimientos e interpretación de resultados, así como el uso de herramientas computacionales, en particular R y RStudio. El curso requiere de una formación básica previa en el análisis de datos estadísticos, tiene una orientación aplicada al análisis de datos multivariados en ciencias sociales.

III. Resultados de Aprendizaje

1. Resultado general: Entregar herramientas para que las personas puedan abordar el análisis de datos tanto de dependencia como de interdependencia y desde la estadística multivariante, elaborando y probando hipótesis y desarrollando modelos explicativos de determinados fenómenos de estudio. Las personas estudiantes aprenderán los métodos de análisis de datos más habituales, y realizarán prácticas en programas estadísticos, principalmente R.

El enfoque del curso será aplicado e irá dirigido a la utilización crítica de las técnicas estadísticas para explicar distintos problemas relacionados con las Ciencias Sociales. Se espera que las personas comprendan que la complejidad de los fenómenos estudiados por

las ciencias sociales obliga a que su análisis considere múltiples medidas para poderlos explicar en forma adecuada.

2. Resultados específicos:

Al final del curso las personas estudiantes deberán poder:

- Explicar distintos problemas sociales considerando sus múltiples factores, mediante el uso de análisis multivariado de datos.
- Utilizar el análisis de datos desde la estadística inferencial, elaborando y testeando hipótesis y desarrollando modelos explicativos de determinados fenómenos de estudio.
- Utilizar el análisis estadístico desde la interrelación entre múltiples variables y sus categorías.
- Identificar las situaciones en que se utilizan cada una las técnicas estadísticas multivariadas, justificar el uso de éstas y ser capaz de analizar e interpretar los resultados obtenidos.
- Demostrar una lectura crítica del análisis multivariado de datos en ciencias sociales.

IV. Contenidos

1) Uso de modelos estadísticos en Ciencias Sociales

2) Modelo de Regresión Lineal Múltiple

- Regresión lineal simple
- Ecuación de regresión múltiple
- Pruebas de hipótesis para los coeficientes
- Coeficiente de Determinación Ajustado
- Cálculo de la mejor ecuación de regresión múltiple
- Coeficientes “beta” estandarizados
- Variables ficticias o dummy
- Interpretación de los coeficientes
- Supuestos del modelo - Análisis de residuos
- Prácticas en programa estadístico R
- Interpretaciones

3) Modelos de Regresión Logística Binaria

- Introducción
- Modelo de regresión logística bivariado
- Modelo de regresión logística multivariado
- Evaluación del modelo
- Test de Wald - Interpretación de los coeficientes
- Requisitos y limitaciones
- Prácticas en programa estadístico R
- Interpretaciones

4) Análisis de Conglomerados

- Introducción
- ¿Qué es un conglomerado?
- Método jerárquico y Dendograma

- Método de K-medias
- Prácticas en programa en R Studio
- Interpretaciones

5) Análisis de Componentes Principales

- Introducción
- Análisis factorial y ACP
- La matriz de correlación
- Elección de los componentes principales
- Representación gráfica de los resultados
- Prácticas en programa estadístico R
- Interpretaciones

V. Metodología

- El curso se organizará en torno a sesiones plenarias y talleres prácticos. En las sesiones plenarias se abordarán los contenidos de las distintas técnicas estadísticas consideradas, mediante clases expositivas a cargo del docente. En los talleres prácticos se aplicarán dichas técnicas mediante el uso de R.
- Las/los ayudantes tendrán el rol de apoyar las sesiones de clases, especialmente durante los talleres prácticos. A su vez, realizarán ayudantías durante el semestre. Estas ayudantías serán definidas según disponibilidad horaria del curso. Durante las ayudantías, las personas estudiantes aprenderán a trabajar con bases de datos, realizando análisis y estimaciones de acuerdo con los tópicos vistos en clase, utilizando RStudio. Además, tendrán la función de revisar los trabajos prácticos, en conjunto y bajo la supervisión del profesor.
- Todos los materiales del curso estarán disponibles en la página <https://aadi2024.netlify.app/>.
- Las/los estudiantes debe cumplir con un porcentaje mínimo del 70% de asistencia para aprobar el curso. Se recomienda asistir a clases para manejar los distintos contenidos. Todos los contenidos vistos en clase serán evaluados en pruebas y trabajos prácticos (incluyendo aspectos que no aparezcan en las presentaciones en clases).

VI. Evaluaciones

Se utilizarán dos formas de evaluación: Pruebas solemnes y reportes de investigación:

- 2 pruebas solemnes (15% y 25% c/u, 40% total).
- 2 reportes de investigación (20% c/u, 40% total).
- 3 Tareas (6,6 c/u 20% total)

Pruebas solemnes: Pruebas escritas que evaluarán los contenidos vistos en clases, especialmente la interpretación de resultados de análisis estadísticos.

Reportes de investigación: Trabajos prácticos orientados a responder una pregunta de investigación específica, que derive de los intereses de las personas estudiantes. Los reportes se harán en grupos de 3 personas, y se realizarán fuera del horario de clases. Los informes tendrán que estar organizados a partir de la siguiente estructura:

- 1) Portada
- 2) Introducción: Problematicación, objetivo y pregunta de investigación.

- 3) Argumentos e hipótesis o expectativas
- 4) Datos, variables y métodos
- 5) Resultados
- 6) Conclusiones

Tareas: Tareas breves enfocadas en la aplicación de técnicas de análisis en R a partir de un problema de investigación dado.

Inasistencias a pruebas solemnes y reportes de investigación

Si una/un estudiante falta a una prueba solemne, puede presentar ante la Secretaría de Estudios un certificado médico dentro de los cinco días hábiles posteriores a dicha evaluación, para tener derecho a rendir una prueba recuperativa. Las/los estudiantes tendrán derecho a un máximo de una prueba recuperativa por curso. La prueba podrá ser rendida a fin de semestre, y considerará todos los contenidos vistos durante el curso. Las evaluaciones no rendidas tendrán nota 1.0.

En caso de tener certificado médico para la fecha de entrega de alguno de los reportes de investigación, se deberá conversar con la profesora una nueva fecha de entrega. De tener certificado médico, es responsabilidad del estudiante contactar a la profesora para definir la manera de proceder. Inasistencias o no entregas de trabajos sin certificado médico tendrán nota 1.

Examen

La nota que resulte del conjunto de evaluaciones parciales será la nota de presentación a examen. Si esta es 3,5 o superior, se podrá rendir el examen. En caso contrario, se considerará que la persona estudiante ha reprobado la asignatura, manteniéndose la nota de presentación como nota final.

La nota de eximición es de 5,5, siempre y cuando no se tengan notas bajo 4,0 en las pruebas solemnes. En ese caso, se considerará que la persona estudiante ha aprobado la asignatura, manteniéndose la nota de presentación como nota final.

La nota de presentación equivale al 70% del total, mientras que la nota en el EXAMEN al 30%. El examen consistirá en una evaluación escrita y práctica cuyos contenidos serán definidos en la última sesión de cátedra.

Recorrección

Según reglamento, las/los estudiantes tienen derecho a conocer las notas y la pauta de corrección de todas las evaluaciones en un plazo que no podrá exceder de 10 días hábiles contados desde la fecha desde que éstas fueran rendidas, como también a conocer el resultado de sus evaluaciones solemnes o principales, antes de rendir las siguientes de igual categoría en una misma asignatura.

Las/los estudiantes pueden imputar la corrección de cualquier evaluación escrita, en un plazo no superior a 5 días hábiles desde la entrega de la nota. Ello, a través de una carta dirigida profesor, que debe ir acompañada por la prueba, trabajo o informe evaluado, y, además, por las correcciones y comentarios que eventualmente haya entregado el docente. El académico deberá comunicar al estudiante el resultado de la corrección, fundamentando por escrito sus argumentos, en un plazo no superior a los 5 días hábiles.

VII. Bibliografía

VII.1. Bibliografía obligatoria.

- Cea D'Ancona, M. (2002). Análisis multivariable: teoría y práctica de la investigación Social. Madrid: Editorial Síntesis Sociológica.
- Camarero (2017). Regresión Logística: Fundamentos y aplicación a la investigación sociológica
- Darlington, R. B., & Hayes, A. F. (2017). Regression analysis and linear models: concepts, applications, and implementation. Guilford Press
- Wickham, H. (2017). R para Ciencia de Datos <https://es.r4ds.hadley.nz/index.html>
- Hair, J., Anderson, R., Tatham, R., & Black, W. (1999). Análisis multivariante. Madrid: Editorial Prentice Hall.
- Husson, F., Pagès, J., & Lê, S. (2012). Análisis de datos con R. Escuela Colombiana de Ingeniería Julio Garavito.
- Lizón, Á. (2006). Estadística y causalidad en la sociología empírica del XX. Papers, 80, 223-255. Universitat Autònoma de Barcelona, Departament de Sociologia.
- Moore (2010) Estadística aplicada básica. Barcelona: Antoni Bosch
- Wooldridge, J. M. (2008). Introducción a la econometría: un enfoque moderno. Paraninfo Cengage Learning.

VII.2. Bibliografía complementaria.

- Alboukadel, K. (2017). Practical Guide to Principal Component Methods in R. STHDA.com, 152.
- Darlington, R. B., & Hayes, A. F. (2017). Regression analysis and linear models: concepts, applications, and implementation. Guilford Press.
- Everitt, B., & Hothorn, T. (2011). An introduction to applied multivariate analysis with R. Springer Science & Business Media.
- Kassambara, A. (2017). Practical guide to cluster analysis in R: Unsupervised machine learning (Vol. 1). STHDA.

VIII. Cláusula Ética

La Escuela establece severas sanciones para las faltas graves que los estudiantes realicen, tales como plagio o falsificación de documentos, señalados en el Reglamento del Estudiante de Pregrado en el Título XIII de las Conductas Contrarias a los Principios Universitarios.

El Comité de Ética es la instancia responsable de revisar y resolver las situaciones que sean contrarias a los principios universitarios, en relación al desarrollo de conductas ilícitas que vician los procesos de evaluación. Son definidas como conductas que vician una evaluación, todos aquellos actos que implican un rompimiento de los estándares éticos que un estudiante debe resguardar en una evaluación, distorsionando la apreciación de los conocimientos y habilidades reales que el estudiante posee (por ejemplo, copia y uso indebido de documento y/o medios electrónicos, modificación o falsificación de documentos, plagio en trabajos escritos).

Se consideran como fuentes de información que se deben citar, los textos e imágenes en formato electrónico o en papel, en todas sus modalidades: libros, diarios, documentos de trabajo, artículos académicos, etc.

IX. Cronograma de trabajo

Fecha	Contenidos	Lecturas sugeridas	Hitos
12 de agosto	Introducción y Uso de modelos en ciencias sociales	<ul style="list-style-type: none"> Lizón, Á. (2006). Estadística y causalidad en la sociología empírica del XX. Papers, 80, 223-255. Universitat Autònoma de Barcelona, Departament de Sociologia. Esser, H. (2010). El problema de la "sociología de las variables". En J. A. Noguera (Ed.), Teoría sociológica analítica Madrid: Centro de Investigaciones Sociológicas. Moore (2010) 	
19 de agosto	Continuación de uso de modelos en ciencias sociales y estadística bivariada	Estadística aplicada básica. Barcelona: Antoni Bosch. Cap 2.	
26 de agosto	Regresión Lineal Simple	Moore (2010) Estadística aplicada básica. Barcelona: Antoni Bosch. Cap 2.	Pauta Tarea 1 (Gestión de datos y estadística bivariada)
2 de septiembre	Regresión Lineal Simple	Moore (2010) Estadística aplicada básica. Barcelona: Antoni Bosch. Cap 2.	Entrega Tarea 1 (Gestión de datos y estadística bivariada)
9 de septiembre	Prueba Solemne 1		
16 de septiembre	Feriado		
23 de septiembre	Regresión Lineal Multiple	<ul style="list-style-type: none"> Cea D'Ancona, M. (2002). Análisis multivariable: teoría y práctica de la investigación Social. Madrid: Editorial Síntesis Sociológica. Capítulo 2. Wooldridge, J. M. (2008). Introducción a la econometría: un enfoque moderno. Paraninfo Cengage Learning. Cap 3. 	
30 de septiembre	Regresión Lineal Multiple	<ul style="list-style-type: none"> Cea D'Ancona, M. (2002). Análisis multivariable: teoría y práctica de la investigación Social. Madrid: Editorial Síntesis Sociológica. Capítulo 2. Wooldridge, J. M. (2008). Introducción a la econometría: un enfoque moderno. Paraninfo Cengage Learning. Cap 3. 	Pauta Tarea 2 (Regresión lineal múltiple)
7 de octubre	Inferencia estadística y regresión	Moore (2010) Estadística aplicada básica. Barcelona: Antoni Bosch. Cap 7.	Entrega Tarea 2 (Regresión lineal múltiple)
14 de octubre	Regresión Logística I	<ul style="list-style-type: none"> Cea D'Ancona, M. (2002). Análisis multivariable: teoría y práctica de la investigación Social. Madrid: Editorial Síntesis Sociológica. Capítulo 2. Camarero (2017). Regresión Logística: Fundamentos y aplicación a la investigación sociológica 	Pauta Informe 1 (regresión lineal múltiple o logística)
21 de octubre	Regresión Logística II	<ul style="list-style-type: none"> Cea D'Ancona, M. (2002). Análisis multivariable: teoría y práctica de la investigación Social. Madrid: Editorial Síntesis Sociológica. Capítulo 2. Camarero (2017). Regresión Logística: Fundamentos y aplicación a la investigación sociológica 	

28 de octubre	Supuestos y chequeos de robustez	- Darlington & Hayes 2016 Cap16 Detecting and Managing Irregularities - Darlington & Hayes 2016 Cap12 Nonlinear relationships - Wooldridge cap 6 Temas adicionales	Entrega Informe 1 (regresión lineal múltiple o logística)
4 de noviembre	Análisis de Conglomerados (Método jerárquico)	• Cea D'Ancona, M. (2002). Análisis multivariable: teoría y práctica de la investigación Social. Madrid: Editorial Síntesis Sociológica. Capítulo 3. • Hair, J., Anderson, R., Tatham, R., & Black, W. (1999). Análisis multivariante. Madrid: Editorial Prentice Hall. Capítulo 9.	
11 de noviembre	Análisis de Conglomerados (Métodos de K-medias)	• Cea D'Ancona, M. (2002). Análisis multivariable: teoría y práctica de la investigación Social. Madrid: Editorial Síntesis Sociológica. Capítulo 3. • Hair, J., Anderson, R., Tatham, R., & Black, W. (1999). Análisis multivariante. Madrid: Editorial Prentice Hall. Capítulo 9.	Pauta Tarea 3 (conglomerados no jerárquicos)
18 de noviembre	Análisis de Componentes Principales	• Cea D'Ancona, M. (2002). Análisis multivariable: teoría y práctica de la investigación Social. Madrid: Editorial Síntesis Sociológica. Capítulo 5. • Husson, F., Pagès, J., & Lê, S. (2012). Análisis de datos con R. Escuela Colombiana de Ingeniería Julio Garavito. Capítulo 1.	Entrega Tarea 3 (conglomerados no jerárquicos) Pauta Informe 2 (Conglomerados o ACP)
25 de noviembre	Análisis de Componentes Principales	• Cea D'Ancona, M. (2002). Análisis multivariable: teoría y práctica de la investigación Social. Madrid: Editorial Síntesis Sociológica. Capítulo 5. • Husson, F., Pagès, J., & Lê, S. (2012). Análisis de datos con R. Escuela Colombiana de Ingeniería Julio Garavito. Capítulo 1.	
2 de diciembre	Prueba Solemne 2		
9 de diciembre	Prueba Recuperativa y taller de vinculación entre técnicas		Entrega Informe 2 (Conglomerados o ACP)
16 de diciembre	Examen		

X. NORMATIVA DE PREVENCIÓN Y SANCIÓN DE ACCIONES DE DISCRIMINACIÓN, VIOLENCIA SEXUAL Y DE GÉNERO DE LA UNIVERSIDAD DIEGO PORTALES.

Tal como señala esta normativa:

“La Universidad Diego Portales (UDP) promueve el pluralismo y la independencia crítica de quienes la integran y tienen como uno de sus principales objetivos el desarrollo integral de las personas que forman parte de ella y de la sociedad. En ese marco, la UDP promueve una convivencia reflexiva basada en el reconocimiento a la dignidad de las personas, la libertad de expresión, la libertad sexual, la identidad y expresión de género, y la autonomía.”

Por lo tanto: “Quienes integren la Universidad tienen la obligación de conocer esta normativa y los protocolos, sitios web y programas que de ella se deriven. Particularmente responsables de su implementación serán las autoridades de la Universidad, quienes tendrán la primera obligación de hacer frente a situaciones que deñen a sus integrantes. Para lograr estos objetivos, se debe realizar la correcta difusión y socialización de la normativa.”

Los detalles de la normativa pueden ser consultados en el siguiente enlace:

https://staticcl1.fidelizador.com/udportales/2022/enero/actualizacion-normativa/RESOLUCION_%20RECTORIA_%20N%3F03_%28APRUEBA_MODIF_NORMA_TIVA_VIOLENCIA%202022%29OK.pdf