

Clase 1: Introducción y uso de modelos en CS

Análisis Avanzado de Datos

Gabriel Sotomayor



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.



Resultados de Aprendizaje

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.



Resultados específicos

- 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.



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



Contenidos

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



Contenidos

4) Análisis de Conglomerados

- ¿Qué es un conglomerado?
- Método jerárquico y Dendograma
- Método de K-medias
- Prácticas en programa en R Studio
- Interpretaciones

5) Análisis de Componentes Principales

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



Bibliografía



Evaluaciones

- 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).

Primeras fechas

Tarea 1: 2 de septiembre

Prueba 1: 9 de Septiembre



Ayudantías

El curso tiene tres ayudantes:

- Patrio Alarcón patricio.alarcon@mail.udp.cl
- Fernanda Hurtado fernanda.hurtado@mail.udp.cl
- Francisca Hernández francisca.hernandez_c@mail.udp.cl

Están disponibles para responder las dudas que puedan tener a lo largo del curso, tanto estadísticas como de uso de software. Habrá sesiones de ayudantía cada 2 semanas aproximadamente, centradas en la aplicación de las técnicas que revisaremos en R. También les acompañarán en la realización de tareas y trabajos de investigación.



Página del Curso

<https://aadi2024.netlify.app/>

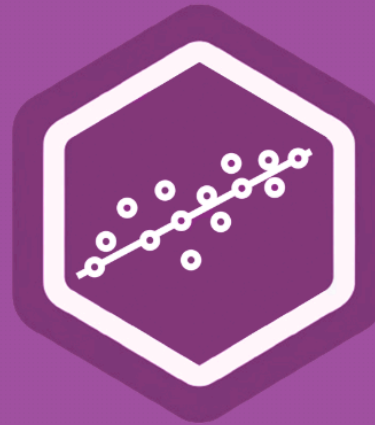


Organización ▾ Clases Prácticos Recursos

Análisis avanzado de datos I

Introducción a métodos multivariados para **Ciencias Sociales**







Profesor Gabriel Sotomayor • 2° Semestre 2024
Sociología • Universidad Diego Portales






Docente

 Gabriel Sotomayor López
 gabriel.sotomayor@mail.udp.cl
 @gsotomayorl

Ayudantes

 Patricio Alarcón
 patricio.alarcon@mail.udp.cl
 Fernanda Hurtado
 fernanda.hurtado@mail.udp.cl
 Francisca Hernández
 francisca.hernandez_c@mail.udp.cl

Detalles del curso

 Plenaria: Lunes 8:30-09:50 / Taller sección
1: Lunes 10:00-11:20 / Taller sección 2:
Lunes 11:30-12:50 /
 Ayudantía: Lunes 14:30-15:50
 Sala 603/ Laboratorio de Computación 2°
piso FCSH



Delegado/a de curso

Las comunicaciones del curso con el equipo docente para temas colectivos deberán gestionarse de manera centralizada mediante un/a delegado/a, especialmente considerando que hay estudiante de distintas generaciones. Esto es particularmente relevante para solicitudes respecto a evaluaciones.



Uso de modelos en Ciencias Sociales



Objetivo de la sesión

Reflexionar sobre el sentido del uso de estadística (multivariada) y modelos en Ciencias Sociales



¿Porqué usamos estadística en Ciencias Sociales?



¿Porque usar modelo (multivariados) en ciencias sociales?

Los modelos son formalismos lógicos o matemáticos que buscan describir la realidad. Nos permiten simultáneamente capturar la complejidad de la realidad social (y los datos con que contamos) y reducirla, de manera de hacerla inteligible: **producir conocimiento (vincular nuestros datos a un contexto teórico mayor)**.

Permiten:

- Formalizar, dando precisión y permitiendo poner teorías a prueba.
- Develar relaciones entre variables y mecanismo causales
- Predecir
- Simular



Sociología de las variables

Cuando utilizamos estadística, en particular multivariada caemos en la tentación de centrarnos en la técnica estadística por sobre la explicación sociológica.

Esser (2010) formaliza este problema llamándolo la **sociología de las variables**, en la cual se identifica una **variable dependiente (explanandum)** y se propone un conjunto de **variables independientes (explanans)** que podrían influir en ella. La explicación se considera lograda cuando se puede atribuir la **varianza** de la variable dependiente a los efectos de las variables independientes



Problemas de la Sociología de las Variables

Incompletitud: Las relaciones entre variables establecidas en un contexto pueden no ser aplicables en otros, revelando la falta de leyes sociológicas generales y estables. La SV, al intentar explicar fenómenos sociales, frecuentemente se queda en explicaciones ad hoc, lo que limita su alcance y efectividad.

Significado Variable: Las variables estructurales pueden tener significados diferentes según el contexto cultural o social. Este problema de equivalencia funcional implica que las mismas variables no siempre tienen el mismo impacto en diferentes escenarios, lo que dificulta la creación de explicaciones universales.



Problemas de la Sociología de las Variables

Interdependencia: Las estructuras sociales son procesos dinámicos donde las interacciones entre individuos y procesos son complejas. La SV no aborda adecuadamente cómo estas interdependencias afectan los resultados sociales, limitando la capacidad de la SV para explicar fenómenos complejos.

Falta de Sentido: La SV ignora el sentido subjetivo de las acciones individuales, centrándose solo en relaciones entre variables. Esto deja de lado la dimensión interpretativa crucial para una explicación sociológica completa, que considera las decisiones conscientes de los individuos.

Reduccionismo: Al reducir fenómenos sociales a simples relaciones entre variables, la SV pierde de vista la complejidad de las decisiones individuales y colectivas, y cómo estas influyen en los resultados sociales.



Explicación teórica y modelos estadísticos

“La consecuencia más importante de los problemas apuntados es una inversión de la perspectiva: los «modelos causales» sobre relaciones entre variables y la varianza «explicada» no son la explicación teórica que buscamos, sino, más bien, solo el explanandum de la «auténtica» explicación teórica conductual.” (Esser, 2010, p. 204).

Esto implica que se necesita contar con una **teoría de la acción social**.



Ejemplo del divorcio

En este caso vemos que hay una relación con forma de “hoz” del riesgo de divorciarse. Lo importante no es describir dicha relación, sino que poder plantear un mecanismo teórico relacionado con el comportamiento de los individuos que la explique.



¿Cómo se relacionan la SV con la tradición teórica sociológica?

Los fundamentos de la sociología de las variables se pueden encontrar, por ejemplo, en las ideas de Emile Durkheim (u otras tradiciones teóricas holistas), quien propuso tratar los “hechos sociales” como entidades independientes de las acciones individuales, buscando estabilidad en las estructuras y ambientes sociales (lo cual puede argumentarse que va perdiendo eficacia históricamente).

Giddens, en su rechazo al modelo de Durkheim, enfatizó la importancia de la agencia humana y la producción activa de la sociedad por sus miembros. Esto subraya un cambio hacia el análisis de las condiciones subjetivas y las decisiones individuales, lo que cuestiona las explicaciones puramente estructurales de la SV.

Es una forma en que aparece el problema de la relación entre **agencia y estructura**.



