

Introducción al curso

Sesión N° 1
Análisis avanzado de datos II

Profesor Gabriel Sotomayor López
Ayudantes Anaís Herrera y Nicolás Hernández

Contenidos

Presentación, resultados de aprendizaje y específicos

Contenidos, bibliografía y evaluaciones

Ayudantías

Contenidos

Página y delegado del curso

Objetivos de la sesión

¿Dónde se sitúa la estadística y el uso de software en el proceso de investigación?

Contenidos

Uso de software en ciencias sociales

¿Qué es R? y Rstudio

Bibliografía y otros recursos para profundizar

Introducción

El curso profundiza las principales técnicas multivariantes, que permiten analizar en forma conjunta la interacción de múltiples factores, lo cual da la posibilidad de estudiar de forma más compleja los problemas sociales.

Se enfatiza la comprensión de procedimientos e interpretación de resultados, así como el uso de herramientas computacionales, principalmente en R y RStudio. Se trabajará en la aplicación de los métodos sin necesidad de estudiar en profundidad los fundamentos matemáticos de todos ellos.

Al final del curso podrán realizar análisis descriptivo de los datos tanto primarios como secundarios y seleccionar la técnica estadística adecuada a utilizar cuando se tienen datos multivariantes.

Resultados de aprendizaje

Resultado general:

- Entregar herramientas al alumno para que pueda abordar el análisis de datos desde la estadística multivariante descriptiva e inferencial, elaborando y probando hipótesis y desarrollando modelos explicativos de determinados fenómenos de estudio.

Resultados de aprendizaje

Resultados específicos:

Manejar bases de datos complejas.

1. Plantear y estructurar adecuadamente un problema de análisis multivariante.
2. Identificar el tipo de técnica adecuada, o combinación de ellas, a aplicar en cada situación concreta.
3. Comprender las herramientas utilizadas en cada una de las técnicas estudiadas.
3. Aprender a utilizar los programas computacionales necesarios para aplicar los distintos modelos estudiados, particularmente R.
4. Interpretar los resultados del análisis.
5. Redactar correctamente un informe con los resultados del análisis.

Contenidos

Unidades	Descripción
U1: Análisis de datos	Análisis Univariado y gestión de los datos en R
	Calculo y análisis de deciles de ingreso: Caso Chile – Encuesta CASEN
U2: Formas funcionales de los Modelos de Regresión	Introducción
	Modelo Logarítmico bivariado
	Modelo Logarítmico Multivariado

Contenidos

Unidades	Descripción
U3: Análisis Factorial Exploratorio	Análisis factorial exploratorio y su aplicación en la investigación sociológica.
	El concepto de varianza y covarianza.
	Diferencias entre el análisis de componente principal y de factor común.
	Supuestos de la técnica.
	La extracción de factores comunes.

Contenidos

Unidades	Descripción
U3: Análisis Factorial Exploratorio	Métodos de extracción de factores, criterios de selección del número de factores y rotación.
	La matriz factorial y su interpretación. Evaluación del modelo factorial y cálculo de puntuaciones factoriales.
U4: Análisis Factorial Confirmatorio	Análisis factorial confirmatorio y su aplicación en la investigación sociológica.

Contenidos

Unidades	Descripción
U4: Análisis Factorial Confirmatorio	El análisis confirmatorio frente al análisis exploratorio.
	Supuestos de la técnica.
	Especificación e identificación del modelo. Estimación de parámetros, evaluación del ajuste y reespecificación.
	Fiabilidad y validez de los modelos de medida. Validez convergente y validez discriminante.

Contenidos

Unidades	Descripción
U4: Análisis Factorial Confirmatorio	Evaluación de la capacidad confirmatoria del modelo.
	Ejemplo de análisis factorial confirmatorio.
U5: Análisis de Sendero	Análisis de sendero y su aplicación en la investigación sociológica. Supuestos de la técnica.
	Especificación del modelo.
	Evaluación del modelo de medición y de la capacidad confirmatoria del modelo.

Contenidos

Unidades	Descripción
U5: Análisis de Sendero	Ejemplo de análisis de sendero.
U6: Modelos de Ecuaciones estructurales	Análisis de ecuaciones estructurales y su aplicación en la investigación sociológica. La obtención de un modelo de ecuaciones estructurales: la estructura del modelo.
	Supuestos de la técnica. Estimación del modelo empírico.

Contenidos

Unidades	Descripción
U6: Modelos de Ecuaciones estructurales	Evaluación del modelo.
	Introducción de modificaciones y/o conclusión del análisis. Evaluación de la capacidad confirmatoria del modelo.
	Ejemplos sobre modelamiento de ecuaciones estructurales.

Bibliografía

Queda pendiente insertar imagen

Evaluaciones

Evaluaciones

Queda pendiente insertar imagen

Ayudantías

El curso tiene dos ayudantes:

- Anais Herrera anais.herrera@mail.udp.cl
- Nicolás Hernández nicolas.hernandez_f@mail.udp.cl

Están disponibles para responder las dudas que puedan tener a lo largo del curso, tanto estadísticas como de uso de software.

Habrà sesiones de ayudantía cada 2 semanas, 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

Insertar imagen e hipervínculo

Delegado de curso

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

Objetivos de la sesión

- **Reflexionar sobre el uso de software estadístico en ciencias sociales.**
- **Introducir el uso de R y Rstudio para ciencias sociales.**

¿Donde se sitúa la estadística y el uso de software en el proceso de investigación?

Queda pendiente insertar imagen

Uso de Software en Ciencias Sociales

Progresivamente se ha generalizado el uso de software estadístico en ciencias sociales, abriendo grandes posibilidades de realizar análisis más complejos y facilitando su uso. Riesgo de falta de formación estadística y reflexividad (calidad del análisis)

Dos niveles de manejo del software:

1. Ejecución básica y correcta interpretación de las salidas estadísticas (necesario mínimo para utilizar la herramienta e interpretar otras investigaciones).
2. Manejo avanzado de software y procedimientos estadísticos. Nivel deseable para un uso reflexivo de las herramientas y lectura crítica de otras investigaciones.

¿Qué software existen y cuál utilizar?

Pendiente img de insertar x2

¿Qué es R?

R environment características:

R es un software y un lenguaje de programación, ambos de carácter libre, enfocado principalmente en el análisis y visualización de datos. Es desarrollado por The R Foundation for Statistical Computing.

Funciona con una lógica de objetos, sobre los que podemos trabajar con operadores y funciones. El hecho de que sea un lenguaje de programación nos permite programar funciones, por lo que es posible pasar de ser usuarios a programadores.

¿Qué es R?

Instalamos una versión básica y podemos agregar paquetes con nuevas funciones. Su interfaz es poco amigable, la solución: R Studio.

imagenes pendientes

¿Porqué aprender y usar R y RSTUDIO?

¿Porqué aprender y usar R y RSTUDIO?

imágenes pendientes

¿Por qué usar R?

¿Por qué usar R?

1. Es libre y gratis.
2. Por lo anterior ofrece grandes posibilidades de personalización de los análisis y amplitud de herramientas (en actualización).
3. Amplia y creciente comunidad de usuarios desarrollando funciones y resolviendo problemas en internet (la mayoría en inglés).
4. Su sintaxis es simple e intuitiva.

¿Porqué trabajar con sintaxis?

1. **Replicabilidad:** Permite a otros entender como construimos nuestros resultados y replicarlos, cada vez más importante en la ciencia.
2. **Eficiencia:** En condiciones “reales” de trabajo resulta más eficiente que trabajar con botones (aunque ahora pueda parecer contraintuitivo)
3. **Control:** Permite tener claridad de todas las etapas de nuestros análisis, cuestión que en otros software son realizadas por defecto. (Elousa, 2009 en Boccardo y Ruiz, 2018: 8-9)

Es una interfaz gráfica para R que nos permite tener mejor visualización y control de los procesos que realizamos.

Tiene 4 partes principales:

Editor de Sintaxis: formato de block de notas donde podemos construir nuestra sintaxis.

**Ambiente de trabajo
(Environment):** muestra
los objetos que hemos
creado (memoria temporal
– RAM)

Consola de R: Aquí es donde podemos ver los resultados de nuestros análisis, las operaciones ejecutadas y mensajes de error.

img pendiente para todas

Panel de visualización:
Aquí aparecen los gráficos
que ejecutemos, además
de tener la pestaña help y
nuestro directorio de
trabajo.

Instalar R y Rstudio

Instalar R

- Podemos encontrar R para Windows, Linux y Mac en <https://cloud.r-project.org/>
- Debemos descargar el paquete base, luego ejecutar el instalador, e instalarlo según las opciones predeterminadas
- Podemos encontrar R Studio en su sitio web, y bajar la versión gratis, que contiene todas las funciones <https://www.rstudio.com/products/rstudio/download/>
- Debemos fijarnos en elegir el sistema operativo correcto.
- Igualmente lo ejecutamos e instalamos según las opciones predeterminadas

El problema de Java x64

Para algunos análisis R se basa en Java, el problema es que nuestros computadores suelen instalar de manera automática la versión de 32 bits, porque los navegadores de internet funcionan (o funcionaban) en esta arquitectura, pero nuestro computadores y R funcionan a 64 bits. Para resolverlo debemos descargar e instalar la versión offline de 64 bits de Java

<https://www.java.com/es/download/manual.jsp>

imagen pendiente

Programación por objetos y funciones

Tipos de estructuras de datos en R Vector (vector): columna o fila de datos de un mismo tipo (una variable individual) Listas (list): Nos permiten agrupar objetos de distinto tipo. Matrices (matrix): arreglo de dos dimensiones de datos de un mismo tipo (conjunto de variables) Data.frame (base de datos): Matriz de datos en el que las columnas tienen asignado nombres, y que permite usar todo tipo de datos.

Tipos de variables en R Numérico (numeric; números con decimales) Números enteros (integer) Lógicos: Verdadero o Falso Variables cadena o texto (string / character) Factores (factor): para variables cualitativas (nominales)

Estructura básica de la sintaxis de R

¿Cómo continuar?

Dominar R es un camino de mediano plazo y es importante desarrollar habilidad de “hacking”, es decir: estar dispuesto a buscar respuestas de manera autónoma saber donde buscar la información utilizar nuevos datos y paquetes por su cuenta

Donde buscar:

- GOOGLE! (preferentemente en inglés) y youtube
- Pedir ayuda a otros usuarios
- Pestaña “help”
- **RDocumentation**

¿Cómo continuar?

Donde buscar:

- [Stackoverflow](#)
- [Stackexchange](#) dudas específicas de análisis de datos

Además hay múltiples recursos web para aprender R de manera profunda y guiada.

Cuestiones importantes que no revisamos: visualización de datos con ggplot2, el tidyverse (paquete dplyr), y profundizar en programación en R (loops y funciones), programación literada (Rmarkdown).

Introducción al curso

Sesión N° 1
Análisis avanzado de datos II

Profesor Gabriel Sotomayor López
Ayudantes Anaís Herrera y Nicolás Hernández