

61343 - Técnicas de análisis multivariante

Información del Plan Docente

Año académico: 2020/21

Asignatura: 61343 - Técnicas de análisis multivariante

Centro académico: 109 - Facultad de Economía y Empresa

Titulación: 525 - Máster Universitario en Economía

Créditos: 3.0

Curso: 1

Periodo de impartición: Segundo semestre

Clase de asignatura: Optativa

Materia: ---

1. Información Básica

1.1. Objetivos de la asignatura

La asignatura y sus resultados previstos responden a los siguientes planteamientos y objetivos:

El planteamiento de la asignatura es marcadamente instrumental dado que su objetivo es dotar al estudiante de un conjunto de herramientas estadísticas muy utilizadas en la realización de un análisis exploratorio multidimensional. Todos los temas se enfocarán desde un punto de vista práctico, utilizando diferentes conjuntos de datos para ilustrar las técnicas explicadas. Para ello se utilizará el entorno y lenguaje de programación R que integra multitud de paquetes que incrementan su capacidad y versatilidad. R dispone de funciones básicas relacionadas con los análisis descriptivos de datos, y de los modelos más complejos y actuales concernientes con los últimos avances en el campo de la estadística o el análisis de datos en áreas como la Economía.

El curso comienza con un tema dedicado a la **Introducción a R** en el que se presentan los comandos más habituales de manejo de datos, muy útiles para adquirir algunas nociones y conceptos básicos. A continuación, se detallan los tipos de datos que se pueden trabajar en R y, posteriormente, se explican comandos básicos para crear, borrar, definir y operar con los datos.

El siguiente tema titulado **Análisis inicial de datos** muestra la necesidad de examinar los datos previamente a la aplicación de cualquier técnica estadística, de esta forma se consigue un entendimiento básico de los mismos y de las relaciones existentes entre las variables analizadas. Este análisis proporciona métodos sencillos para organizar y preparar los datos, detectar fallos en el diseño y recogida de datos, tratamiento y evaluación de datos ausentes, identificación de casos atípicos y comprobación de los supuestos subyacentes en la mayor parte de las técnicas multivariantes. Comenzamos por un análisis estadístico unidimensional, cuyo tipo dependerá de la escala de medida de la variable analizada. Por ello se proporciona una guía de actuación sugiriéndole las representaciones gráficas y resúmenes numéricos más aconsejables para describir las características más importantes en cada caso. Una vez realizado el estudio unidimensional de cada variable por separado, el siguiente paso consiste en analizar la existencia de posibles relaciones entre ellas. Dicho estudio se realiza desde una óptica bidimensional y se analizan las tres situaciones generales que pueden presentarse. Así, se proporcionan las pautas a seguir cuando ambas variables son cualitativas, ambas son cuantitativas y cuando una es cuantitativa y la otra cualitativa.

El resto de los temas abordan ya técnicas multivariantes, por ello, en primer lugar, se presentan de una forma intuitiva, y se realiza una adecuada clasificación de las mismas para que el estudiante sea capaz de seleccionar la técnica que debería aplicarse en cada momento en función del problema que pretenda resolver. Una vez motivado el interés de estas técnicas mediante diferentes ejemplos, el tema 3 titulado **Análisis Factorial** presenta las dos técnicas más utilizadas en procesos de reducción de datos: el Análisis de Componentes Principales y el Análisis Factorial, indicando sus ventajas y sus inconvenientes. Ambas técnicas se utilizan para analizar interrelaciones entre un número elevado de variables métricas explicando dichas interrelaciones en términos de un número menor de variables denominadas componentes principales o factores.

El siguiente tema denominado **Análisis Clúster** busca clasificar a los individuos en grupos homogéneos internamente y heterogéneos externamente, analizando las diferencias existentes entre ellos. Es un ejemplo de técnica de clasificación no supervisada ya que los grupos son desconocidos a priori y es necesario derivarlos de las observaciones.

El último tema titulado **Análisis Discriminante** presenta una técnica de clasificación supervisada ya que las observaciones ya están previamente clasificadas en dos o más grupos, buscándose las razones que explican esa clasificación.

1.2. Contexto y sentido de la asignatura en la titulación

La asignatura tiene un carácter marcadamente instrumental y tiene como objetivo formar al estudiante en el manejo de técnicas cuantitativas multivariantes habitualmente utilizadas en la implementación de análisis exploratorio de datos multivariantes que constituye la primera fase de cualquier estudio estadístico-económico de carácter aplicado. Las técnicas que se estudian son, fundamentalmente, de dos tipos: el primer grupo (análisis de componentes principales y análisis factorial) buscan llevar a cabo un proceso de reducción de datos mediante la identificación y cálculo de factores latentes a los mismos cuya finalidad es expresar la información contenida en un conjunto de datos en términos de un

número menor de variables que representan las relaciones existentes entre las variables analizadas; el segundo grupo (análisis clúster) busca clasificar los objetos/individuos en grupos homogéneos internamente y heterogéneos externamente analizando las diferencias existentes entre los mismos. Todo ello redundará en una mejor comprensión de la información contenida en un conjunto de datos que puede ser relevante a la hora de plantear un análisis más formal de los mismos.

1.3.Recomendaciones para cursar la asignatura

Haber realizado un curso de introducción a la Estadística tanto descriptiva como inferencial así como un curso de introducción a la Econometría

2.Competencias y resultados de aprendizaje

2.1.Competencias

Al superar la asignatura, el estudiante será más competente para...

Al superar la asignatura, el estudiante será más competente para utilizar las herramientas estadísticas para extraer de los datos información relevante para elaborar y defender proyectos aplicados de carácter económico.

2.2.Resultados de aprendizaje

El estudiante, para superar esta asignatura, deberá demostrar los siguientes resultados...

Al finalizar el curso, el estudiante deberá ser capaz de:

- Realizar un análisis inicial de un conjunto de datos multivariantes
- Realizar un análisis de componentes principales e interpretar los resultados obtenidos
- Realizar un análisis factorial e interpretar los resultados obtenidos
- Relacionar el análisis factorial con el de componentes principales poniendo de manifiesto las similitudes y diferencias existentes entre ambas
- Llevar a cabo un proceso de clasificación de datos utilizando procedimientos jerárquicos aglomerativos
- Llevar a cabo un proceso de clasificación de datos utilizando procedimientos de particionamiento
- Diseñar procedimientos de clasificación que permitan discriminar entre grupos de observaciones de la forma más eficiente posible así como analizar las diferencias existentes entre ellos
- Validar e interpretar los resultados obtenidos en un procedimiento de clasificación

2.3.Importancia de los resultados de aprendizaje

Las técnicas estadísticas estudiadas en la asignatura dotarán al alumno de un conjunto de herramientas muy potentes para realizar un análisis exploratorio inicial de la información contenida en bases de datos económicos. De esta forma el alumno aprenderá a dotar de rigor científico a la resolución de problemas económicos mediante el análisis de datos empíricos que constituye un paso importante en la aplicación del método científico a la resolución de problemas..

3.Evaluación

3.1.Tipo de pruebas y su valor sobre la nota final y criterios de evaluación para cada prueba

El estudiante deberá demostrar que ha alcanzado los resultados de aprendizaje previstos mediante las siguientes actividades de evaluación

La evaluación será continua y, para ello, se realizará un trabajo para cada uno de los temas. Estos trabajos consistirán en ejercicios que se realizarán sobre diferentes bases de datos utilizando R. Más concretamente, en un primer trabajo los estudiantes realizarán algunos ejercicios sobre manipulación de datos, manejo de comandos básicos y elaboración de pequeños scripts en R. En el segundo y tercer trabajo resolverán ejercicios relacionados con un análisis exploratorio unidimensional y bidimensional, respectivamente, aplicados a una base de datos.

En los sucesivos trabajos realizarán informes sobre la aplicación de las distintas técnicas (análisis de componentes principales, análisis factorial, análisis clúster y análisis discriminante) a una base de datos. Para ello dispondrán de diferentes scripts que deberán modificar apropiadamente con el fin de que se adapten a los problemas concretos y a las bases de datos utilizadas.

La calificación final será un reparto de 10 puntos (calificación máxima) de forma equiponderada entre los diferentes trabajos realizados.

Nota: Está previsto que la evaluación se realice de manera presencial pero si las circunstancias sanitarias lo requieren, se realizará de manera semipresencial u online

Nota: Está previsto que la evaluación se realice de manera presencial pero si las circunstancias sanitarias lo requieren, se realizará de manera semipresencial u online

Los estudiantes que no superen este sistema de evaluación continua o los que opten por la evaluación global, realizarán un

examen final de evaluación que valore los conocimientos teóricos y prácticos de la materia mediante ejercicios en los que tendrán que responder a diferentes preguntas sobre distintas bases de datos. Las respuestas a las preguntas se realizarán mediante scripts utilizando el programa R, y redactando un informe con las conclusiones obtenidas. Está previsto que esta prueba se realice de manera presencial, pero si las circunstancias sanitarias lo requieren, se realizará de manera semipresencial u online. Si la evaluación se realizase de forma online, el estudiante podrá ser grabado, pudiendo éste ejercer sus derechos por el procedimiento indicado en el siguiente enlace sobre la CLÁUSULA INFORMATIVA REDUCIDA EN GESTIÓN DE GRABACIONES DE DOCENCIA:

https://protecciondatos.unizar.es/sites/protecciondatos.unizar.es/files/users/lop/dgdocencia_reducida.pdf.

Nota: Está previsto que la evaluación se realice de manera presencial pero si las circunstancias sanitarias lo requieren, se realizará de manera semipresencial u online

4. Metodología, actividades de aprendizaje, programa y recursos

4.1. Presentación metodológica general

El proceso de aprendizaje que se ha diseñado para esta asignatura se basa en lo siguiente:

El proceso de aprendizaje se basa en la combinación de la exposición, por parte del profesor, de las bases teóricas subyacentes a cada una de las técnicas explicadas, con su aplicación al estudio de casos prácticos realizados en el aula utilizando el paquete estadístico SPSS 22.0 y el programa R. Todo ello se realizará en un ambiente participativo en el que tanto el profesor como los estudiantes debatirán entre sí la interpretación de los resultados obtenidos lo cual aumentará el grado de aplicabilidad de las técnicas explicadas al hacerle más fácil al estudiante su asimilación a problemas de análisis de datos de carácter muy general.

Está previsto que las clases sean presenciales. No obstante, si fuese necesario por razones sanitarias, las clases podrán impartirse de forma semipresencial u online.

Está previsto que las clases sean presenciales. No obstante, si fuese necesario por razones sanitarias, las clases podrán impartirse de forma semipresencial u online.

4.2. Actividades de aprendizaje

El programa que se ofrece al estudiante para ayudarle a lograr los resultados previstos comprende las siguientes actividades...

El programa que se ofrece al estudiante para ayudarle a lograr los resultados previstos comprende las siguientes actividades

Actividad formativa	Nº Horas	% Presencialidad
Sesiones teórico-prácticas (50%-50%)	30	100%
Preparación de trabajos y estudio independiente del alumno	60	-----

4.3. Programa

Tema 0: Introducción a R (4 horas)

Tema 1: Análisis inicial de datos (8 horas)

Tema 2: Análisis Factorial (Análisis componentes principales) (6 horas)

Tema 3: Análisis Cluster (6 horas)

Tema 4: Análisis Discriminante (6 horas)

4.4. Planificación de las actividades de aprendizaje y calendario de fechas clave

Presentación de la asignatura: En la primera sesión del curso se explican de forma detallada los objetivos y contenidos de la asignatura, se plantea la metodología docente utilizada en las clases y se exponen los criterios de evaluación con nitidez. Así mismo, se dará una panorámica de las técnicas estadísticas multivariantes.

Sesiones teórico-prácticas: A lo largo del curso se realizarán 15 sesiones teórico-prácticas en un aula de informática en la que se explicarán las bases teóricas de las técnicas estadísticas explicadas en la asignatura para, a continuación, ilustrar su aplicación con el análisis de casos reales relacionados con el mundo de la economía. Para ello se utilizarán como herramientas el programa R.

Entrega de trabajos: Para superar el curso el alumno deberá realizar un trabajo para cada uno de los temas.

4.5. Bibliografía y recursos recomendados

- Afifi, A. A. Computer-aided multivariate analysis / A.A. Afifi and V. Clark . - 3rd. ed. reprinted London [etc.] : Chapman and Hall, 1998
- Análisis multivariante / Joseph F. Hair...[et al.] ; revisión técnica y compilación de las lecturas complementarias, Mónica Gómez Suárez ; traducción, Esme Prentice, Diego Cano . - 5ª ed., última reimp. Madrid [etc.] : Prentice Hall, D.L. 2008
- Sharma, Subhash. Applied Multivariate Techniques / Subhash Sharma New York : John Wiley and Sons, cop. 1996
- Peña Sánchez de Rivera, Daniel. Análisis de datos multivariantes / Daniel Peña . - [Reimp.] Madrid [etc.] : McGraw-Hill, D.L. 2010
- Uriel Jiménez, Ezequiel. Análisis multivariante aplicado : aplicaciones al marketing, investigación de mercados, economía, dirección de empresas y turismo / Ezequiel Uriel Jiménez, Joaquín Aldás Manzano . - [1ª ed.] Madrid : Thomson, D.L. 2005
- Fujikoshi, Yasunori. Multivariate statistics: high-dimensional and large-sample approximations / Yasunori Fujikoshi, Vladimir V. Ulyanov, Ryoichi Shimizu Hoboken. New Jersey (USA) : Wiley, 2010
- Mardia, K.V. Multivariate analysis / K.V. Mardia, J.T. Kent, J.M. Bibby . - Repr. London ; New York : Academic Press, 2003
- Ferrán Aranaz, Magdalena. SPSS para windows : Programación y análisis estadístico / Magdalena Ferrán Aranaz ; revisión técnica Antonio Vaquero Sanchez...[et al.] Madrid, [etc.] : McGraw-Hill, D.L. 1997
- Pérez López, César. Técnicas estadísticas multivariantes con SPSS / César Pérez López . - 1ª ed. Madrid : IberGarceta, D.L. 2009
- Visauta Vinacua, Bienvenido. Análisis estadístico con SPSS para Windows. Vol. 2, Estadística multivariante / Bienvenido Visauta Vinacua . - 2ª ed. Madrid[etc.] : McGraw-hill, D.L. 2003