Universidad Zaragoza

Curso Académico: 2021/22

27621 - Análisis de datos y técnicas multivariantes

Información del Plan Docente

Año académico: 2021/22
Asignatura: 27621 - Análisis de datos y técnicas multivariantes
Centro académico: 109 - Facultad de Economía y Empresa
Titulación: 450 - Graduado en Marketing e Investigación de Mercados
Créditos: 6.0
Curso: 3
Periodo de impartición: Primer semestre
Clase de asignatura: Obligatoria
Materia:

1. Información Básica

1.1. Objetivos de la asignatura

- La materia a impartir tiene un marcado carácter **práctico y cuantitativo,** proporcionando las herramientas estadísticas necesarias para analizar una base de datos multivariantes, haciendo especial hincapié en las técnicas de reducción y clasificación de datos muy utilizadas en Marketing e Investigación de Mercados.

  Su contenido está dividido en 4 bloques diferentes: 1) Introducción a R y R Commander, 2) Análisis inicial de un conjunto multivariante de datos, 3) Técnicas de reducción de datos y 4) Técnicas de clasificación.

  En el bloque ***Introducción a R y R Commander***, en primer lugar, se enseña a instalar el software libre R desde la página web: http://www.r-project.org/ y 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.

  En el segundo bloque denominado ***Análisis inicial de un conjunto multivariante de datos*** indicamos al estudiante 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.

  En el tercer bloque se comienza con las técnicas multivariantes, por ello, en primer lugar, se presentan de una forma intuitiva, realizando 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 bloque titulado ***Técnicas de reducción de datos*** presenta a los estudiantes 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 último bloque denominado ***Técnicas de clasificación*** busca clasificar a los individuos en grupos homogéneos internamente y heterogéneos externamente analizando las diferencias existentes entre ellos. Se indica que el Análisis Clúster es un ejemplo de técnica de clasificación no supervisada, mientras que el Análisis Discriminante lo es de una técnica de clasificación supervisada. En el Análisis Clúster los grupos son desconocidos a priori y es necesario derivarlos de las observaciones. En el Análisis Discriminante las observaciones ya están previamente clasificadas en dos o más grupos, buscándose las razones que explican esa clasificación y no la clasificación en sí.

  Todos los temas se enfocarán desde un punto de vista práctico, utilizando bases de datos para ilustrar las técnicas explicadas e incidiendo más en las ideas que subyacen detrás de las mismas que en el aparataje matemático que utilizan. 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 el marketing e investigación de mercados.

1.2. Contexto y sentido de la asignatura en la titulación

La asignatura tiene carácter obligatorio y se encuentra encuadrada en el módulo Investigación de Mercados, dentro de la materia Instrumentos. En dicha materia figura junto a Econometría, que también es obligatoria. Posteriormente se impartirán las asignaturas optativas Métodos Estadísticos en Investigación de Mercados y Sistemas de Información y Bases de Datos.

Con todas las asignaturas anteriormente mencionadas se proporciona al alumno un conjunto técnicas informáticas, econométricas y estadísticas, de tipo cualitativo y cuantitativo, muy útiles para desarrollar una investigación de mercados básica y obtener conclusiones que sean relevantes para la gestión empresarial

1.3. Recomendaciones para cursar la asignatura

La asignatura no tiene establecido ningún requisito previo. Sin embargo, es conveniente haber aprobado las asignaturas Matemáticas I, Matemáticas II, Estadística I y Estadística II para mejorar el aprendizaje de la misma y necesitar de menor esfuerzo.

Para poder alcanzar los objetivos de aprendizaje previstos es necesaria la asistencia a las clases, así como la **participación activa** en dichas clases. Es muy aconsejable el estudio continuado de la asignatura para facilitar la comprensión de la misma, así como la realización de las diversas actividades programadas.

Recursos web: Se empleará el ADD y, más concretamente, la plataforma MOODLE 2.

Las fechas de inicio y finalización de la asignatura, así como las fechas concretas de las actividades clave se fijarán de acuerdo al calendario académico y al horario establecido por el Centro.

La asignatura consta de 6 créditos ECTS (60 horas lectivas y 90 horas de tutorías y de trabajo personal del alumno). Las horas lectivas se reparten de la siguiente forma:

- 1,2 créditos de horas teórico-prácticas (30 horas lectivas)
- 1,2 créditos de horas teórico- prácticas (30 horas lectivas)

El periodo lectivo abarca 15 semanas y cada semana se imparten 4 horas de clases teórico-prácticas. Las fechas oficiales de los exámenes serán establecidas por el Centro y aparecerán publicadas con antelación suficiente, para todas las convocatorias.

A través del Anillo Digital Docente (ADD) el profesor irá informando puntualmente a los alumnos sobre la disponibilidad de presentaciones, casos prácticos, ejercicios, etc. relevantes sobre la materia para la realización de diferentes tareas. Es obligación del alumno conocer las actualizaciones que se van realizando en el ADD.

2. Competencias y resultados de aprendizaje

2.1. Competencias

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

Competencias Específicas:

- Conocer y aplicar las herramientas cualitativas y cuantitativas de análisis y diagnóstico para la investigación de mercados.
- Comprender las posibilidades de las TIC para la investigación de mercados.
- Desarrollar una investigación de mercados básica y obtener conclusiones para la gestión empresarial.
- Desarrollar, presentar y defender un proyecto en el ámbito de la Investigación de Mercados.

Competencias Transversales:

- Adquisición con rapidez de nuevos conocimientos.
- Búsqueda, análisis y síntesis de fuentes de información y datos.
- Búsqueda de nuevas ideas y soluciones.
- Detección de oportunidades.

- Capacidad para cuestionar de forma constructiva ideas propias o ajenas.
- Creatividad.

2.2. Resultados de aprendizaje

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

1. Manejar los comandos básicos de R para manipular y preparar la información de una base de datos.
2. Codificar los ítems de una encuesta del modo más apropiado en función del objetivo del estudio.
3. Realizar un análisis inicial unidimensional seleccionando la metodología más apropiada de acuerdo a los distintos tipos de las variables seleccionadas.
4. Realizar un análisis bidimensional para analizar las relaciones entre variables, seleccionando la metodología más apropiada de acuerdo a los distintos tipos de variables analizados.
5. Analizar los datos ausentes de una base de datos evaluando si existen sesgos por dicho motivo, poniendo los remedios oportunos.
6. Realizar un análisis de componentes principales e interpretar los resultados obtenidos.
7. Realizar un análisis factorial e interpretar los resultados obtenidos.
8. Relacionar el análisis factorial con el de componentes principales poniendo de manifiesto las similitudes y diferencias existentes entre ambas.
9. Llevar a cabo un proceso de clasificación de datos utilizando procedimientos jerárquicos aglomerativos.
10. Llevar a cabo un proceso de clasificación de datos utilizando procedimientos de particionamiento

2.3. Importancia de los resultados de aprendizaje

La asignatura tiene un doble objetivo para la formación del futuro profesional**. Un primer objetivo viene dado por el procesamiento de la información obtenida en un estudio de mercado**. Para ello se aprenden un conjunto de técnicas estadísticas multivariantes, habitualmente utilizadas en este tipo de estudios, muy útiles para llevar a cabo un análisis exploratorio de la información disponible, así como para realizar un proceso de reducción de datos y de obtención de grupos significativos que faciliten el diseño de políticas comerciales más adecuadas. Además, el uso de este tipo de técnicas, al cuantificar la información recogida y medir el grado de incertidumbre asociado a los patrones de comportamiento detectados en los datos, dotan de un mayor rigor científico a los procesos de tomas de decisiones dotando al estudiante de los criterios necesarios para entender y enjuiciar sus propios resultados o los proporcionados por otras fuentes. **El segundo objetivo consiste en aprender a realizar un informe detallado sobre la importancia de los resultados del análisis de una base de datos y las decisiones a tomar.**

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 previstos de aprendizaje en PRIMERA CONVOCATORIA por alguno de los dos siguientes sistemas de evaluación:

- ***Evaluación CONTINUA*** para los estudiantes que asistan regularmente a clase. Esta evaluación se llevará a cabo mediante la realización de 2 pruebas evaluables individuales (E1 y E2). Las dos pruebas evaluables se realizarán en las fechas habilitadas por el Centro para pruebas intermedias.

En la primera prueba evaluable (E1) se valorará la adquisición de conocimientos de R y de la realización de un análisis exploratorio uni y bidimensional de una base de datos. Esta prueba se realizará en torno a la mitad del semestre en la fecha habilitada y tendrá un peso en la nota final del 50%.

En la segunda prueba evaluable (E2) se valorará la aplicación de las técnicas estadísticas multivariantes explicadas en clase a un conjunto de datos reales y la posterior obtención de conclusiones. Esta prueba se realizará al acabar el semestre en las fechas habilitadas y tendrá un peso en la nota final del 50%.

Ambas pruebas evaluables consistirán en la realización de diferentes ejercicios en los que tendrán que responder a diferentes preguntas sobre distintas bases de datos.

- ***Evaluación GLOBAL*** para los estudiantes que no realicen la evaluación continua o que, habiéndola realizado, no la hayan superado, o que quieran mejorar su calificación (en este último caso prevalecerá la mejor calificación). En este caso el alumno realizará un examen final de evaluación (E) 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.

El sistema de evaluación continua sólo es aplicable a la primera convocatoria. En la segunda convocatoria, la evaluación será global.

Está previsto que estas pruebas se realicen de manera presencial, pero, en caso necesario, se realizarán online. En este último caso es importante destacar que, en cualquier prueba, el estudiante podrá ser grabado, pudiendo este ejercer sus derechos por el procedimiento indicado en https://protecciondatos.unizar.es/sites/protecciondatos.unizar.es/files/users/lopd/gdocencia_reducida.pdf

Criterios de valoración:

Las pruebas evaluables (E1, E2) serán puntuados en una escala de 0 a 10. El estudiante debe tener al menos un 3 en cada una de ellas para poder realizar la media.

La nota final deberá ser igual o mayor de 5 y se calculará como la media ponderada del 50% de la prueba E1 y el 50% de la prueba E2. Es decir, la fórmula aplicada será:

Nota final = 0,5*E1 + 0,5*E2

si el estudiante utiliza la modalidad de evaluación continua, o

Nota final = E

si se presenta a la evaluación global o acude a la segunda convocatoria.

4. Metodología, actividades de aprendizaje, programa y recursos

4.1. Presentación metodológica general

En la impartición de la asignatura se utilizarán diferentes métodos docentes en función de los objetivos marcados y las competencias desarrolladas. Así se emplearán clases magistrales para la presentación teórica de la asignatura, con el objetivo de analizar y desarrollar sus conceptos fundamentales. Sin embargo, y dado que la asignatura se enseñará en aulas de informática en su totalidad, se emplearán herramientas informáticas para complementar las explicaciones teóricas con su aplicación al análisis de bases de datos lo cual facilitará el aprendizaje práctico de las técnicas estudiadas.

El profesor actuará como tutor para ayudarles a resolver las preguntas que se planteen en cada tema. Como apoyo se utilizará la plataforma Moodle 2 en la que se publicarán materiales teóricos y prácticos de la asignatura, así como la información relevante para su desarrollo comenzando por la propia guía docente

4.2. Actividades de aprendizaje

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

- *Clases teórico-prácticas*: Se emplearán para desarrollar los conceptos y desarrollos de cada uno de los temas. En ellas se utilizarán técnicas expositivas, pero motivando la participación y discusión en clase. El profesor se apoyará en el ordenador para ilustrar el uso práctico de las técnicas explicadas, mediante su aplicación a la resolución de casos reales utilizando las Bases de Datos de la asignatura. Se recomienda la asistencia a clase, la participación y la toma de notas o aclaraciones a dicha presentación. Además, en todas ellas se utiliza el programa R para implementar las técnicas explicadas en la exposición.

- *Tutorías individuales y on line*: el estudiante podrá asistir a las horas programadas de tutoría para resolver dudas sobre la asignatura. En el caso de aquéllos estudiantes a los que las tutorías programadas coincidan con horas de clase de otras materias, podrán mandar un correo electrónico para concertar cita presencial o telemática.

Está previsto que la actividad docente se realice de manera presencial pero, en caso necesario, se realizarán online.

4.3. Programa

Bloque 1: Introducción a R y R Commander

Tema 1: Introducción a R

Tema 2: Introducción a R Commander

Bloque 2: Análisis inicial de un conjunto multivariante de datos

Tema 3: Análisis exploratorio unidimensional

Tema 4: Análisis exploratorio bidimensional

Bloque 3: Técnicas de reducción de datos

Tema 5: Análisis de Componentes Principales.

Tema 6: Análisis Factorial

Bloque 4: Técnicas de clasificación

Tema 7: Análisis Clúster

Tema 8: Análisis Discriminante

4.5. Bibliografía y recursos recomendados

[BB: Bibliografía básica / BC: Bibliografía complementaria]

- [BB] Análisis multivariante / Joseph F. Hair...[et al.] ; revisión técnica y compilación de las lecturas complementarias, Mónica Gómez Suárez ; traducció,Esme Prentice, Diego Cano . - 5ª ed., última reimp. Madrid [etc.] : Prentice Hall, D.L. 2008
- [BB] Hair, Joseph F.. Multivariate data analysis [recurso electrónico] / Joseph F. Hair, Jr. . - 7th ed. Harlow : Pearson Education Ltd., [2014]
- [BC] Guisande González, Cástor. Tratamiento de datos con R, STATISTICA y SPSS / Cástor Guisande González, Antonio Vaamonde Liste, Aldo Barreiro Felpeto Madrid : Díaz de Santos, D.L. 2011
- [BC] Johnson, Richard A.. Applied multivariate statistical analysis / Richard A. Johnson, Dean W. Wichern. . - 6th ed. Englewood, New Jersey : Prentice-Hall, 2007.
- [BC] Peña Sánchez de Rivera, Daniel. Análisis de datos multivariantes / Daniel Peña . - [Reimp.] Madrid [etc.] : McGraw-Hill, D.L. 2010
- [BC] Pérez López, César. Técnicas Estadísticas con SPSS / César Pérez López . - [1ª ed, reimp.] Madrid [etc.] : Prentice-Hall, D.L. 2004
- [BC] Pérez López, César. Técnicas Estadísticas con SPSS 12 : aplicaciones al análisis de datos / César Pérez López Madrid [etc.] : Prentice-Hall, D.L. 2005

Listado de URL

- Hair, J. Multivariate data analysis [recurso electrónico] / Joseph F. Hair, Jr. -- 7ª ed. -- Harlow : Pearson Education, [2014] [https://www.dawsonera.com.roble.unizar.es:9443/readonline/9781292035116]