

Grado en Marketing e Investigación de Mercados

27621 - Análisis de datos y técnicas multivariantes

Guía docente para el curso 2013 - 2014

Curso: 3, Semestre: 1, Créditos: 6.0

Información básica

Profesores

- **María Pilar Olave Rubio** polave@unizar.es

Recomendaciones para cursar esta asignatura

La asignatura no tiene establecido ningún requisito previo. Sin embargo, sería conveniente haber cursado y 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, tanto teóricas como prácticas, así como la participación activa en dichas clases. Es aconsejable el estudio continuado de la asignatura para facilitar la comprensión de la misma así como la realización de las diversas actividades programadas.

Actividades y fechas clave de la asignatura

Las fechas de inicio y finalización de la asignatura así como las fechas concretas de las actividades clave de la asignatura se fijarán de acuerdo al calendario académico y al horario establecido por el Centro, informándose de ello a los estudiantes a través del programa de la asignatura.

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. Además, las *tutorías optativas* se suelen realizar en grupos pequeños, utilizando bases de datos de interés en marketing y se responderá a preguntas planteadas por el profesor utilizando la metodología propia del Análisis Multivariante. Usualmente, se realizarán los Viernes de 13 a 15 horas.

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 en el ADD de lecturas, casos prácticos, ejercicios, noticias relevantes sobre la materia para la realización de diferentes tareas. Es obligación del alumno conocer las actualizaciones que se van realizando en el ADD.

Recursos Web

ADD de la asignatura

Inicio

Resultados de aprendizaje que definen la asignatura

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

1:

1. Estudiar los ítems de una encuesta para identificar la mejor información antes de analizar una base de datos.
2. Realizar un análisis exploratorio de una base de datos de acuerdo a los distintos tipos de variables analizados.
3. Analizar los datos ausentes de una base de datos evaluando si existen sesgos por dicho motivo y poniendo los remedios oportunos.
4. Realizar un análisis de regresión múltiple e interpretar los resultados.
5. Realizar un análisis de componentes principales y utilizar sus resultados para localizar atípicos multivariantes.
6. Realizar un análisis factorial e interpretar los resultados.
7. Realizar un análisis de tablas de contingencia e interpretar los perfiles obtenidos.
8. Realizar un análisis cluster utilizando métodos jerárquicos y no jerárquicos e interpretar los resultados obtenidos
9. Realizar un análisis factorial discriminante, validarla e interpretar los resultados obtenidos.

Introducción

Breve presentación de la asignatura

La información que maneja un experto en Análisis de Datos es de vital importancia para la toma óptima de decisiones. En la actualidad el aumento espectacular de la potencia de los ordenadores ha propiciado, en particular, la creación de GRANDES BASES DE DATOS en el mundo económico-empresarial (empresas de logística, comercio electrónico, prospección de mercados, análisis de encuestas, etc.).

La información contenida en dichas bases de datos suele ser de carácter MULTIDIMENSIONAL, lo que hace a las técnicas estadísticas multivariantes imprescindibles para tratar y procesar este tipo de información con la finalidad de interpretarla adecuadamente.

En este curso se pretende introducir al alumno en estas técnicas haciendo especial hincapié en su aspecto práctico. Se maneja el software estándar para este tipo de análisis.

Contexto y competencias

Sentido, contexto, relevancia y objetivos generales de la asignatura

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

La materia a impartir tiene un marcado carácter práctico y cuantitativo, proporcionando las herramientas estadísticas necesarias para analizar una base de datos multivariantes, utilizando técnicas tanto de carácter exploratorio como confirmatorio, haciendo especial hincapié en las técnicas estadísticas de reducción y clasificación de datos muy utilizadas en la Investigación de Mercados.

Su contenido está dividido en 6 bloques diferentes: introducción, análisis exploratorio de datos multivariantes, análisis de regresión múltiple, técnicas de reducción de datos, análisis de datos categóricos y técnicas de clasificación.

En el bloque introductorio se presenta la materia a impartir situándola del plan de estudios conectándola con lo estudiado en los cursos previos y se describe la programación docente de la misma así como los criterios de evaluación. Para ello se presentan las bases de datos a tratar en la asignatura y se plantean múltiples preguntas de acuerdo con su organigrama.

El segundo bloque describe las técnicas estadísticas más utilizadas para realizar un análisis exploratorio de datos prestando especial atención al problema de detección de datos atípicos así como al tratamiento de datos ausentes. El Análisis Multivariante propiamente dicho comienza partir del tercer bloque y contiene un bloque dedicado al análisis de regresión múltiple, un cuarto bloque a las técnicas de reducción de datos (Análisis de Componentes Principales y Análisis Factorial), otro bloque dedicado al análisis de datos categóricos y un último bloque dedicado a las técnicas de clasificación, tanto no supervisadas (Análisis Cluster) como supervisadas (Análisis Discriminante).

Todos los temas se enfocarán desde un punto de vista práctico, utilizando bases de datos seleccionadas 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 paquete estadístico SPSS 19.0 para el que la Universidad de Zaragoza tiene licencia de uso así como diversas bases de datos de diversas encuestas que ayuden al alumno una mejor comprensión y nivel de uso de las técnicas explicadas en clase.

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, económicas 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

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

1:

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.

Importancia de los resultados de aprendizaje que se obtienen en la asignatura:

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.

Evaluación

Actividades de evaluación

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

1:

En **PRIMERA CONVOCATORIA** se plantean dos sistemas de evaluación:

Sistema de Evaluación Continua. Esta evaluación constará de los siguientes elementos:

- **Tres pruebas** intermedias (P1, P2 y P3) de carácter teórico-práctico en las que resolverán diversas cuestiones y problemas prácticos de la materia y cuyo peso será el 60% de la calificación. Cada prueba tendrá un peso del 20% para la evaluación final.

- Una **evaluación final** (F) de carácter práctico en la que, dada una base de datos (de las planteadas en la asignatura) el alumno deberá responder a un conjunto de preguntas del ámbito de la Investigación de Mercados utilizando las técnicas más adecuadas de entre las explicadas en la asignatura, realizando el preceptivo informe en el ámbito de la investigación en la que se utilice dicha base. El peso de esta parte será un 40%.

Para superar la asignatura por este sistema se debe obtener una calificación mínima de 5 puntos sobre 10 además de haber obtenido una nota mínima de 3 puntos en cada una de las pruebas intermedias.

Aquellos alumnos que no hayan superado la asignatura por el sistema de evaluación continua, que no lo hayan realizado o que deseen mejorar su nota, disponen de:

Sistema de evaluación Global, consistente en una prueba a realizar en la fecha oficial marcada por el Centro. Dicha prueba constará de dos partes: una teórico-práctica (60%) y otra práctica (40%). La parte teórico-práctica consistirá en resolver un conjunto de cuestiones sobre los temas de la asignatura. En la parte práctica, el alumno deberá responder a un conjunto de preguntas formuladas sobre el contenido de una base de datos, utilizando las técnicas más adecuadas de entre las explicadas en la asignatura, realizando el preceptivo informe en el ámbito de la investigación en la que se utilice dicha base.

Para superar la asignatura por este sistema se debe obtener una calificación mínima de 5 puntos sobre 10

En **SEGUNDA CONVOCATORIA**, se realizará una prueba Global de 10 puntos en la fecha marcada por el Centro. Esta prueba se realizará siguiendo los mismos requisitos que los establecidos en la prueba global de la primera convocatoria.

Criterios de Valoración

Para aprobar la asignatura en la modalidad de **evaluación Continua** son de obligado cumplimiento las siguientes condiciones: a) obtener más de 5 puntos en la calificación total; y b) obtener una nota mínima de 3 puntos sobre 10 en cada una de las pruebas intermedias. En la modalidad Global se debe obtener más de 5 puntos en la calificación total; y obtener una nota mínima de 3 puntos sobre 10 en cada una de las partes que la componen.

Se valorará que el estudiante conozca, entienda y sea capaz de utilizar los conceptos estudiados en la asignatura. Además, se valorará la capacidad para integrar los conocimientos adquiridos.

En cualquier caso, tanto en las pruebas como en los trabajos prácticos se valorará el grado de acierto con que se responda a las preguntas planteadas, la ausencia de errores formales y la claridad y el orden en la presentación de resultados. Igualmente podrá valorarse la reflexión y el aporte de ideas que haya detrás de cada trabajo o intervención, así como la regularidad de la participación activa en el desarrollo de las clases.

Actividades y recursos

Presentación metodológica general

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

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

La práctica se desarrollará mediante la aplicación de las técnicas explicadas en cada bloque a las bases de datos más adecuadas en ese contexto. El profesor actuará como tutor para ayudarles a resolver las preguntas que se planteen en cada tema. Como ya se ha comentado, se realizan tutorías en grupo pequeños para que el alumno plantee sus comentarios y dudas de aprendizaje.

Como apoyo se utilizará la plataforma Blackboard 9.1 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.

Actividades de aprendizaje programadas (Se incluye programa)

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

1:

Clases teórico-prácticas: Se emplearán para desarrollar principalmente los conceptos y desarrollos teóricos 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. 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 paquete SPSS 19.0 para implementar las técnicas explicadas en la exposición más teórica.

Tutorías en grupo reducido: Se realizarán sesiones optativas de este tipo a lo largo del curso, en los horarios programados por el profesor de acuerdo con los estudiantes. En ellas se plantearán a los estudiantes la selección del método estadístico más adecuado para introducirlos en investigaciones concretas de mercados. Para tal fin se responderá a una batería de preguntas planteadas por el profesor utilizando la metodología propia del Análisis Multivariante.

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.

Además de las tutorías presenciales, el alumno podrá utilizar la plataforma del ADD.

Planificación y calendario

Calendario de sesiones presenciales y presentación de trabajos

Las actividades formativas y de evaluación se desarrollarán de acuerdo a la siguiente planificación:

Actividades	Horas Presenciales	Horas trabajo autónomo	Total carga estudiante
Clases teóricas	30	45	75
Clases prácticas	30	38	66
Tutorías en grupos reducidos	4		4
Pruebas intermedias	3		3
Examen final	2		2
TOTAL	68	83	150

El calendario orientativo de la asignatura, semana a semana, sería el siguiente:

Fecha	Tema	Método docente
1 ^a semana	Guía docente. <u>Método de trabajo</u> INTRODUCCION MULTIVARIANTE	Clase expositiva Clase expositiva
2 ^a semana	Tema 1 Tema 1	Clase teórico-práctica Práctica informática
3 ^a semana	Tema 1 Tema 1	Clase teórico-práctica Práctica informática
4 ^a semana	Prueba Intermedia 1 Tema 2	Prueba teórico-práctica Clase expositiva
5 ^a semana	Tema 2 Tema 2	Práctica Informática Clase teórico-práctica
6 ^a semana	Tema 2 Tema 3	Práctica Informática Clase expositiva
7 ^a semana	Tema 3 Tema 3	Práctica Informática Clase teórico-práctica
8 ^a semana	Tema 3 Prueba Intermedia 2	Práctica Informática Prueba teórico-práctica
9 ^a semana	Práctica tutorizada: Estudio de una Base Tema 4	Práctica Informática Clase expositiva
10 ^a semana	Tema 4 Tema 4	Práctica informática Práctica informática
11 ^a semana	Tema 5 Tema 5	Clase expositiva Práctica informática
12 ^a semana	Tema 5 Tema 5	Clase teórico-práctica Práctica informática
13 ^a semana	Prueba Intermedia 3 Tema 6	Prueba teórico-práctica Clase expositiva
14 ^a semana	Tema 6 Tema 6	Práctica informática Clase teórico-práctica
15 ^a semana	Tema 6 Práctica tutorizada: Plantear Proyecto Evaluación final	Práctica informática Práctica Informática Prueba práctica

PROGRAMA

PARTE 0: INTRODUCCION

TEMA 0: Introducción

Presentación de la asignatura: objetivos, programación docente, programa de la asignatura, evaluación, bibliografía. ¿Qué es el Análisis Multivariante? Estudio de una base de datos estándar en un análisis de mercados: definición de variables, importancia del análisis gráfico y de las técnicas estadísticas multivariantes para la toma de decisiones. Tipos de técnicas multivariantes. Etapas de un análisis de datos.

PARTE I: ANÁLISIS EXPLORATORIO DE DATOS

TEMA 1: Análisis Exploratorio de Datos (A.E.D.)

Objetivos e importancia del A.E.D. Etapas del A.E.D. Preparación de los datos. Examen gráfico de los datos: histograma, diagrama de tallos y hojas, polígono de frecuencias, boxplot, matriz de dispersión, perfiles multivariantes. Medidas descriptivas numéricas. Contrastes de normalidad: Q-Q plots, test K-S y test de Sapiro-Wilks, test de asimetría y curtosis. Tratamiento de outliers y datos missing: MCAR y MAR. Imputación de datos faltantes.

PARTE II: TÉCNICAS DE REDUCCIÓN DE DATOS

TEMA 2: Análisis de Regresión

Introducción. Modelo de regresión simple: método de los mínimos cuadrados. Modelo de regresión múltiple. Inferencia en la regresión lineal múltiple. Evaluación de las hipótesis del modelo: estudio de los residuos. Variables independientes cualitativas. Modelos de regresión con interacción. Regresión por pasos.

TEMA 3: Análisis Factorial

Variables aleatorias multivariantes y distribuciones multivariantes. Muestras multivariantes: matriz de datos, vector de medias, matriz de varianzas y covarianzas muestrales, matriz de correlaciones muestrales, espacio de objetos, espacio de variables. Técnicas de reducción de la dimensionalidad. El modelo del Análisis Factorial. Examen de la matriz de correlación. Extracción de factores: método de las componentes principales. Rotación de factores. Cálculo de puntuaciones factoriales. Interpretación de los resultados obtenidos. Detección de outliers. Codificación de encuestas: utilización del análisis factorial para la reducción de ítems.

PARTE III: ANÁLISIS DE DATOS CATEGÓRICOS

TEMA 4: Análisis de tablas de contingencia

Introducción. Análisis de encuestas con datos categóricos: tablas de contingencia. Perfiles marginales y condicionales. Interpretación y su importancia en una prospección de Mercados. Análisis de Muestras. Contrastes de independencia y homogeneidad.

PARTE IV: TÉCNICAS DE CLASIFICACIÓN

TEMA 5: Análisis Cluster

Introducción al problema de clasificación: técnicas supervisadas y no supervisadas. Medidas de semejanza y de distancia. Métodos jerárquicos: enlaces simple, completo, medio, método de Ward y método del centroide. Validación interna y externa de una solución. Elección del número de clusters. Interpretación de los resultados. Otros métodos: métodos de particionamiento (algoritmo de las k medias).

TEMA 6: Análisis Discriminante

Introducción: enfoque descriptivo y enfoque predictivo del análisis discriminante. Análisis Factorial Discriminante: criterio de discriminación de Fisher. Cálculo e interpretación de las funciones discriminantes canónicas. Funciones discriminantes y clasificación: caso de 2 grupos.

Referencias bibliográficas de la bibliografía recomendada

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