

240ST014 - Análisis de Datos de Transporte y Logística

Unidad responsable: 240 - ETSEIB - Escuela Técnica Superior de Ingeniería Industrial de Barcelona
Unidad que imparte: 715 - EIO - Departamento de Estadística e Investigación Operativa
Curso: 2019
Titulación: MÁSTER UNIVERSITARIO EN CADENA DE SUMINISTRO, TRANSPORTE Y MOVILIDAD (Plan 2014).
(Unidad docente Obligatoria)
MÁSTER UNIVERSITARIO EN INGENIERÍA INDUSTRIAL (Plan 2014). (Unidad docente Optativa)
Créditos ECTS: 5 Idiomas docencia: Castellano, Inglés

Profesorado

Responsable: Montero Mercadé, Lidia

Requisitos

Los estudiantes deben tener conocimientos previos de análisis matemático y algebra para asimilar los conceptos de probabilidad, distribuciones de probabilidad univariantes, series numéricas, notación matricial, funciones reales de una o más dimensiones, derivación e integración.

Los estudiantes han de saber programar en pseudocodi o en algún lenguaje de alto nivel.

Competencias de la titulación a las cuales contribuye la asignatura

Específicas:

CESCTM2. Diseñar procedimientos de recogida de datos de transporte que tengan en cuenta su especificidad, saber aplicar las técnicas adecuadas para tratarlos, analizarlos y extraer conclusiones para su uso adecuado en los modelos que los requieran.

CESCTM3. Diseñar y realizar estudios de análisis de la demanda, modelar la demanda y su estructuración según los distintos modelos de transporte.

CETM2. Comprensión y capacidad de cuantificación de las variables fundamentales de sistemas de transporte y de movilidad que determinan la seguridad, la calidad y la sostenibilidad de las infraestructuras de transporte y optimización del funcionamiento de estos sistemas.

240ST014 - Análisis de Datos de Transporte y Logística

Metodologías docentes

El aprendizaje del curso se compone de tres fases diferentes:

1. Adquisición de conocimientos específicos a través del estudio de la literatura y material proporcionado por los profesores.
2. La adquisición de habilidades en técnicas específicas de análisis de datos, explotación de la información y modelaje estadístico.
3. Integración de los conocimientos, habilidades y competencias (genéricas y específicas) para la resolución de casos prácticos cortos.

Las clases de teoría se exponen los fundamentos de las metodologías y técnicas de la asignatura.

Las clases de laboratorio sirven para aprender el uso de técnicas específicas para la resolución de problemas en el análisis de datos estadísticos mediante el uso de herramientas informáticas adecuadas, en este sentido, los estudiantes primero deben seguir y tomar notas sobre el análisis realizado por el profesor y luego resolver en el autoaprendizaje horas un caso de estudio corto similar, que se centra en los contenidos del bloque actual y siguiendo las pautas descritas en un cuestionario incluido en la descripción de las sesiones de laboratorio. El caso de estudio corto contenido en el cuestionario debe ser resuelto en no más de una semana o cuando indica el profesor durante la sesión de laboratorio. La evaluación formativa se llevará a cabo facilitando la corrección antes de la siguiente sesión de laboratorio, donde se incluirá una discusión sobre los problemas comunes encontrados por el profesor en los primeros 20 minutos.

Los casos de estudio cortos son desarrollados por los estudiantes en grupos durante horas de autoaprendizaje, sirven para poner en práctica los conocimientos, habilidades y competencias en la resolución de los casos todos ellos relacionados con Logística, Transporte y Movilidad. El Software R es la herramienta estadística seleccionada para el análisis de datos y el modelado. Software profesional común (TransCAD, EMME4, VISSUM) se presenta y se relacionan con las herramientas R en función de las disponibilidades de licencias docentes en la UPC.

Objetivos de aprendizaje de la asignatura

Aprender a hacer un informe sobre la calidad de los datos (perfil de los datos que faltan, detección de atípicos uni y bivariado). Recuperación de datos faltantes.

Aprender a utilizar e interpretar los conceptos fundamentales de la probabilidad y la estadística desde un punto de vista práctico cuando se utiliza el software estadístico R: evento aleatorio, población, muestra, variable aleatoria, al azar variables aleatorias continuas y discretas comunes. Punto y estimación del intervalo. Inferencia estadística computacional.

Aprender a analizar las bases de datos, incluyendo la descripción univariante numérica y gráfica, con herramientas descriptivas bivariadas y multivariadas. Determinación de las características significativas de los grupos de individuos.

Aprender a hacer un perfil para una respuesta, ya sea cuantitativa o cualitativa. Selección de características relevantes.

Aprender los principios básicos de la Clasificación: técnicas de clasificación jerárquica y vecinos K más cercano. Realizar y validar una propuesta de clasificación utilizando el software R.

Saber realizar el modelado de las respuestas numéricas por regresión: formulación, estimación e interpretación de modelos estadísticos utilizando el software R.

Conocer los indicadores de comparación de modelos de regresión generales y selección del mejor: Estadísticos de bondad del ajuste (R^2 , F-Test para los modelos anidados, AIC, BIC, etc)

Conocer los indicadores de diagnóstico de los modelos lineales generales con finalidades predictivas: valores atípicos y datos influyentes. Aplicar a los modelos de generación / atracción para las zonas de transporte (ZAT).

Saber realizar el modelado de opciones-discretos mediante los modelos lineales generalizados: formulación, estimación e interpretación de modelos estadísticos utilizando el software R.

Conocer los indicadores de comparación de modelos de respuesta binaria y selección del mejor: estadísticos de bondad del ajuste (Pearson X^2 , Test desviación para los modelos anidados, AIC, BIC, etc)

Conocer los indicadores de diagnóstico de los modelos lineales generalizados con finalidades predictivas.

Saber aplicar los modelos de respuesta binaria a la elección modal entre pares de zonas de transporte (ZAT). Agregada vs modelos desagregados.

Conocer los principios básicos de la teoría del muestreo: estimaciones por punto y por intervalo. Aprender cómo calcular

240ST014 - Análisis de Datos de Transporte y Logística

errores relativos y absolutos para los estimadores de medias, totales y proporciones en el muestreo aleatorio y el muestreo estratificado.

Horas totales de dedicación del estudiantado

Dedicación total: 125h	Horas grupo mediano:	30h	24.00%
	Horas grupo pequeño:	15h	12.00%
	Horas aprendizaje autónomo:	80h	64.00%

240ST014 - Análisis de Datos de Transporte y Logística

Contenidos

Bloque 1. Introducción al Análisis de Datos en Transporte y Logística

Dedicación: 6h

Grupo grande/Teoría: 2h
Grupo mediano/Prácticas: 1h
Aprendizaje autónomo: 3h

Descripción:

Introducción a las colecciones de datos comunes y encuestas en Logística, Transporte y Movilidad: encuestas domiciliarias, encuestas origen-destino, encuestas de cordón, las encuestas de preferencias reveladas y declarada. La recopilación de datos de tráfico: Los sensores de bucle inductivo y nuevas tecnologías (datos Bluetooth, datos de sensores magnéticos inalámbricos, etc.)

Actividades vinculadas:

Sesión Presencial y Presentación entorno estadístico R en Laboratorio

Objetivos específicos:

El objetivo de la asignatura es proporcionar a los estudiantes los conocimientos y habilidades para hacer frente a análisis exploratorio de datos y las necesidades de extracción de datos de las organizaciones y la práctica profesional en el campo del Transporte y la Logística. Es decir, para tomar ventaja de los datos almacenados por los agentes implicados y poderlas integrar en los sistemas automáticos para ayudar a la toma de decisiones y en la gestión de las operaciones de tráfico. La idea subyacente es que los datos son un tesoro para los interesados y a través de su explotación se convierte en información. El curso se desarrolla en base al estudio de caso. Se divide en cuatro áreas: Análisis Exploratorio de Datos e Interpretación- Descripción de resumen, Inferencia Estadística Computacional, Modelado y Predicción-Herramientas. Diseño de cuestionarios y Toma de Muestras de Diseño.

La asignatura proporciona una sólida formación en las técnicas para administrar, analizar, modelar y extraer conocimiento a partir de los conjuntos de datos masivos actuales, bases de datos, Internet, ..., así como en las técnicas para explotar ese conocimiento en el sector.

Bloque 2. Análisis Exploratorio de Datos

Dedicación: 17h 30m

Grupo grande/Teoría: 3h
Grupo mediano/Prácticas: 1h 30m
Aprendizaje autónomo: 13h

Descripción:

Análisis exploratorio de datos: herramientas numéricas y gráficas univariantes/bivariantes (características cuantitativas y cualitativas). Los datos que faltan: el perfil y su recuperación. La detección de valores atípicos univariantes y bivariantes. Medidas de asociación de datos multivariantes (correlación Pearson / Spearman) Ejemplo de datos masivos: Conteos de tráfico (recuperación de desaparecidos, detección de valores atípicos, filtrado)

Objetivos específicos:

Aprender a hacer un informe sobre la calidad de los datos (perfil de los datos que faltan, detección de los demás uni y bivariado). Falta de recuperación.

Aprender a analizar las bases de datos, incluyendo la descripción univariante numérica y gráfica, bivariado y multivariado con las herramientas de R. Determinación de las características significativas de los grupos de individuos.

240ST014 - Análisis de Datos de Transporte y Logística

<p>Bloque 3. Inferencia Estadística Computacional</p>	<p>Dedicación: 24h</p> <p>Grupo grande/Teoría: 2h Grupo mediano/Prácticas: 6h Aprendizaje autónomo: 16h</p>
<p>Descripción: Elementos estadísticos básicos utilizados en el transporte y la logística: distribuciones univariantes comunes (binomial, multinomial, de Poisson, exponencial, Weibull, gamma, (log) logística (log) normal, etc), con énfasis en los momentos y parámetros característicos (localización, escala y forma). Identificación. Inferencia estadística computacional para medias, varianzas y proporciones y las diferencias según grupos: paramétricos y non parametricos (Chi2, Anderson-Darling, Wilcoxon, Kruskal-Wallis, Barlett, etc).</p> <p>Objetivos específicos: Aprender a utilizar e interpretar los conceptos fundamentales de la probabilidad y la estadística desde un punto de vista práctico cuando se utiliza R : evento aleatorio, población, muestra, variable aleatoria, variables aleatorias continuas y discretas comunes. Punto y estimación del intervalo. Inferencia estadística computacional. Identificación de la distribución para datos univariantes.</p>	
<p>Bloque 4. Modelaje Estadístico por Regresión</p>	<p>Dedicación: 24h 30m</p> <p>Grupo grande/Teoría: 5h 30m Grupo mediano/Prácticas: 5h Aprendizaje autónomo: 14h</p>
<p>Descripción: Modelado mediante modelos de regresión múltiple. Estimación de mínimos cuadrados. Propiedades. La transformación de las variables. Las herramientas de diagnóstico y estadísticas: residuales, datos influyentes y valores atípicos. Modelo lineal general: la forma de introducir las variables cualitativas como variables explicativas - definición de variables ficticias. Efectos principales y las interacciones entre los factores y las covariables: interpretación y validación de modelos. Prueba F de Fisher para comparar modelos anidados. Ejemplo: Modelado de la casa al trabajo en hora punta de la mañana entre zonas de transporte (ZAT) con el método lm () y selección del mejor modelo con el método step() del software R, .</p> <p>Objetivos específicos: Saber modelar las respuestas numéricas: formulación, estimación e interpretación de modelos estadísticos utilizando el software R. Saber comparar los modelos lineales y selección del mejor: estadísticos de bondad del ajuste (R2, F-Test para los modelos anidados, AIC, BIC, etc) Saber hacer el diagnóstico de los modelos lineales generales: valores atípicos y datos influyentes. Saber hacer predicción con los modelos lineales generales. Saber aplicar a los modelos de generación / atracción para las zonas de transporte (ZAT).</p>	

240ST014 - Análisis de Datos de Transporte y Logística

<p>Bloque 5. Modelaje de Respuesta Binaria</p>	<p>Dedicación: 18h</p> <p>Grupo grande/Teoría: 4h Grupo mediano/Prácticas: 2h Aprendizaje autónomo: 12h</p>
<p>Descripción: Modelado de datos discretos binarios a través de modelos de regresión generalizada: función de enlace, la estimación ML, las propiedades, diagnosis y validación de modelos. Interpretación. Predicción: capacidad predictiva (ROC). Test de Devianza para comparar modelos anidados. Estudio de caso de la selección de modo entre los modos públicos y privados de acuerdo a características individuales glm() en R.</p> <p>Objetivos específicos: Conocer el modelado de los modelos lineales generalizados discretos: formulación, estimación e interpretación de modelos estadísticos utilizando el software R. Saber hacer diagnosis y validación de los modelos de respuesta binaria con las herramientas de R. Saber comparar modelos y hacer selección del mejor. Saber calcular e interpretar las medidas de Bondad de ajuste y los tests asociados (Pearson X2, Test desviación para los modelos anidados, AIC, BIC, etc) Saber hacer diagnóstico y validación de los modelos lineales generalizados para la respuesta binaria. Saber Aplicar los modelos de respuesta binaria a la elección modal entre pares de zonas de transporte (ZAT). Entender la diferencia y similitudes entre los modelos a partir de datos Agregados o Desagregados.</p>	
<p>Bloque 6. Introducción al Muestreo Estadístico</p>	<p>Dedicación: 12h</p> <p>Grupo grande/Teoría: 4h Aprendizaje autónomo: 8h</p>
<p>Descripción: Introducción a la teoría de muestreo: muestreo aleatorio y el muestreo estratificado. Punto y estimaciones de intervalo para los medios, totales y proporciones en un muestreo aleatorio. Selección del tamaño de la muestra para satisfacer errores absolutos / relativos en un muestreo aleatorio y estratificado. Ejemplo: Dimensionamiento de una encuesta de movilidad domiciliaria.</p> <p>Objetivos específicos: Conocer los principios básicos de la teoría del muestreo: estimaciones por punto y estimaciones de intervalo. Aprender cómo calcular errores relativos y absolutos para estimaciones de medias, totales y proporciones en el muestreo aleatorio y el muestreo estratificado.</p>	

240ST014 - Análisis de Datos de Transporte y Logística

<p>Bloque 7. Introducción a la Minería de Datos</p>	<p>Dedicación: 12h</p> <p>Grupo grande/Teoría: 2h Grupo mediano/Prácticas: 2h Aprendizaje autónomo: 8h</p>
<p>Descripción: Minería de datos de datos masivos: métodos útiles para la Logística y el Transporte. Clasificación: segmentación de la población de un área de estudio - Clasificación jerárquica con componentes R. Principio como una herramienta para la reducción de la dimensionalidad. Ejemplo: Encuesta de satisfacción de los usuarios del transporte de una red de autobuses</p> <p>Objetivos específicos: Saber cómo convertir los datos en información que es de utilidad para la toma de decisiones. Saber calcular Perfiles en R. Conocer estrategias para la reducción de la dimensionalidad. Saber hacer e interpretar la Clasificación Jerárquica no dirigida en R. Saber hacer e interpretar la Clasificación Dirigida (K-Means) en R.</p>	
<p>Evaluación: Parcial i Final</p>	<p>Dedicación: 11h</p> <p>Grupo grande/Teoría: 5h Aprendizaje autónomo: 6h</p>
<p>Descripción: Quiz y Examen Final</p> <p>Actividades vinculadas: El Quiz o Examen Parcial tiene lugar en la semana de parciales, a mitad de cuatrimestre, que programa la ETSEIB. La fecha depende de la programación de la ETSEIB no viene fijada por profesores, ni alumnos.</p>	

Sistema de calificación

La evaluación del curso integra las tres fases del proceso de aprendizaje: conocimientos, habilidades y competencias.

El conocimiento es evaluada por un cuestionario (prueba parcial) y el examen final (puntuaciones F1 y F2), en el medio y la semana pasada del curso.

Las habilidades y competencias se evalúan a partir de la entrega de las m prácticas ($m > 1$) con base en los casos de estudio cortos y relacionados con los contenidos del curso. Cada uno de los bloques, excepto el primero, podría implicar una práctica que los alumnos realizarán en grupo (como máximo 3 personas). El promedio de las m puntuaciones da la puntuación L. Los estudiantes tienen que cuantificar las horas dedicadas a resolver cada práctica y entregarlas según disponga la correspondiente Tarea en ATENEA. Feedback para la evaluación formativa por parte del profesor en no más de 10 días, antes de la siguiente sesión de laboratorio donde se discutirán los problemas y errores comunes en los primeros 20 minutos.

La nota final se obtiene un peso de las tres puntuaciones: $\text{Nota Final} = 0.65F + 0.35L$. Donde F es Max (F2, 0.3F1 + 0.7F2).

240ST014 - Análisis de Datos de Transporte y Logística

Normas de realización de las actividades

Los estudiantes podrán llevar calculadora, tablas estadísticas, etc. No se permite llevar las resoluciones de exámenes de años anteriores, pero si se pueden llevar las resoluciones de los casos de estudios disponibles en la web docente de la asignatura.

Bibliografía

Básica:

Washington, S.P. ; Karlaftis, M.G. ; Mannering, F.L. Statistical and Econometric methods for transportation data analysis. 2nd. Boca Raton: Chapman and Hall, 2011. ISBN 9781420082852.

Dalgaard, Peter. Introductory Statistics with R [en línea]. 2nd ed. New York: Springer, 2008 [Consulta: 05/10/2017]. Disponible a: <<http://dx.doi.org/10.1007/978-0-387-79054-1>>. ISBN 9780387790534.

Clairin, Rémy ; Brion, Philippe. Manual de Muestreo. Madrid: La Muralla, 2001. ISBN 8471337118.

Fox, John. Applied Regression Analysis and Generalized Linear Models. 2nd ed. Los Angeles: SAGE, 2008. ISBN 9780761930426.

Fox, John ; Weisber, Sanford. An R Companion to Applied Regression. 2nd ed. Thousands Oaks: SAGE, 2002. ISBN 9781412975148.

Ortúzar S., Juan de Dios; Willumsen, Luis G. Modelling transport. 4th ed. Chichester: John Wiley & Sons, 2011. ISBN 9780470760390.

Otros recursos:

Página web del curso:

- Planificación de la asignatura
- Notas relacionadas con los bloques de contenidos y diapositivas usadas en las clases teóricas.
- Descripción de las sesiones prácticas, los cuestionarios para cada bloque y casos de estudio.
- Casos de Estudio: los datos (formato MS-Excel y R) y descripción del contexto y la variable objetivo / s.
- Directrices para los casos de estudio que se presentan en forma de una lista de preguntas de guía para el análisis.
- Exámenes Finales y cuestionarios de los cursos anteriores.

Enlace web

Web Docent ADTL

<http://www-eio.upc.es/teaching/adtl/>

Material informático

ATENEA - Tasques

ATENEA dispondrà de Tareas para la entrega de los casos de estudio prácticos