240ST014 - Data Analysis of Transport and Logistics

Coordinating unit:	240 - ETSEIB - Barcelona School of Industrial Engineering
Teaching unit:	715 - EIO - Department of Statistics and Operations Research
Academic year:	2018
Degree:	MASTER'S DEGREE IN SUPPLY CHAIN, TRANSPORT AND MOBILITY MANAGEMENT (Syllabus 2014). (Teaching unit Compulsory) MASTER'S DEGREE IN STATISTICS AND OPERATIONS RESEARCH (Syllabus 2013). (Teaching unit Optional) MASTER'S DEGREE IN INDUSTRIAL ENGINEERING (Syllabus 2014). (Teaching unit Optional)
ECTS credits:	5 Teaching languages: Spanish, English

Teaching staff

Coordinator:	Montero Mercadé, Lídia

Requirements

Students must have sufficient knowledge of algebra and mathematical analysis in order to assimilate concepts regarding probability, univariant distribution of random variates, numerical series, matrix algebra, he functions of real variables in one or more dimensions, derivation and integration.

Student must have basic programming skills in pseucode or in a high level programming language

Degree competences to which the subject contributes

Specific:
   CESCTM2. Develop procedures for collecting transportation data that take into account their specificity, namely to apply appropriate treat, analyze and draw conclusions for appropriate use in models that require techniques.
   CESCTM3. Design and conduct studies demand analysis, demand modeling and structuring for different transport models.
   CETM2. Understanding and quantifying capacity fundamentals transport systems and mobility determine the safety, quality and sustainability of transport infrastructure and optimizing the operation of these systems.

240ST014 - Data Analysis of Transport and Logistics

Teaching methodology

Learning the course consists of three distinct phases:

1. Acquisition of specific knowledge through the study of literature and material provided by teachers.
2. The acquisition of skills in specific techniques of data analysis, exploitation of information and statistical modeling.
3. Integration of knowledge, skills and competencies (specific and generic) by solving short case studies.

Theory classes expose the foundations of methodologies and techniques of the subject.
The laboratory classes are intented to learn the use of specific techniques for problem solving in statistical data analysis, using appropriate informatics tools, in this sense, students first must follow and take notes about the analysis carried out by the teacher and then solve in the selflearning hours a similar one, that focuses on the current block/contents and with a questionnaire included in the description of laboratory sessions. The short case study described in the questionnaire has to be solved according to the questionnaire at most in 1 week as will be indicated by the lecturer during the lab session. Feedback will be provided before the next laboratory session, where a discussion about common problems encountered by the teacher will be place in the first 20 min.

The course case studies, where students are setttled in groups in selflearning hours, serves to put into practice the knowledge, skills and competences in solving a case provided by the teacher and related to Logistics, Transportation and Mobility. R software is the selected statistical tools for data analysis and modeling. Common professional software (TransCAD, EMME4, VISSUM) capabilities are presented and related to R tools.

Learning objectives of the subject

Learn how to make a report on data quality (missing data profile, univariate and bivariate outlier detection). Missing recovery.
Learn how to use and interpret fundamental concepts in probability and statistics from a practical point of view when using R statistical software: random event, population, sample, random variable, common continuous and discrete random variates. Point and interval estimate. Computational statistical inference.
Learn how to analyze databases including numeric and graphical univariate description, bivariate and multivariate tools. Determination of the significant characteristics of groups of individuals.
Learn how to make a profile for a target response, either quantitative or qualitative. Feature selection.
Learn the basic principles of Classification: hierarchical classification techniques and K-Nearest neighbors. Perform and validate a proposed classification using R software.
Know how to Model a numeric responses using general linear regression: formulation, estimation and interpretation of statistical models using R software.
Know how to interpret indicators on Model comparison and selection for general regression models: Goodness of fit statistics ($R2$, F-Test for nested models, AIC, BIC, etc)
Know how to validate general regression models: outliers and influential data.
Know to Apply general regression models to the generation/attraction of trip at Zones of Transport (ZAT).
Know how to Model discrete choices- by generalized linear models : formulation, estimation and interpretation of statistical models using R software.
Know how to interpret indicators on Model comparison and selection for generalized linear models: Goodness of fit statistics (X2 Pearson, Deviance Test for nested models, AIC, BIC, etc)
Know how to validate generalized linear models.
Know how to forecast in general linear models and binary choice generalized linear models using R.
Apply to the modal choice between pairs of Zones of Transport (ZAT). Aggregated vs Disaggregated models.
Know the basic principles of Sampling Theory: point and interval estimates. Learn how to compute relative vs absolute errors for estimates of means, totals and proportions in random sampling and stratified sampling.

240ST014 - Data Analysis of Transport and Logistics

Study load

Total learning time: 125h			
	Hours large group:	0h	0.00%
	Hours medium group:	30h	24.00%
	Hours small group:	15h	12.00%
	Guided activities:	0h	0.00%
	Self study:	80h	64.00%

240ST014 - Data Analysis of Transport and Logistics

Content

Block 1. Introduction to Data Analysis in Transportation and Logistics	Learning time: 6h Theory classes: 2h Practical classes: 1h Self study : 3h

Description:
Introduction to common data collections and surveys in Logistics, Transportation and Mobility: home-based surveys, O-D surveys, cordon surveys, stated and revealed preferences surveys. Traffic data collection: inductive loop sensors and new technologies (Bluetooth data, wireless magnetic sensor data, etc).

Related activities:
Theory class and Introduction to R in Laboratory

Specific objectives:
The purpose of the subject is to provide students with the knowledge and skills to cope with exploratory data analysis and data mining needs of organizations and professional practice in the field of Supply Chain, Transportation and Mobility. That is, to take advantage of the data stored by stakeholders to integrate automatic systems to aid decision making and traffic operations and management. The underlying idea is that data are a treasure for stakeholders and through its exploration becomes clear information contained in them.

The course is developed based on solving the problems of case studies. It is divided into four areas: Exploratory Data Analysis and Interpretation: summary description, Computational Statistical Inference, Modeling and Prediction-tools and Design of Questionnaires and Sampling Design.

The subject will give a solid background in the techniques to manage, analyze, model and extract knowledge from the current massive data sets, databases, Internet, …, as well as in the techniques to exploit that knowledge in the sector.

Block 2. Exploratory Data Analysis	Learning time: 17h 30m Theory classes: 3h Practical classes: 1h 30m Self study : 13h

Description:
Exploratory Data Analysis: numerical and graphical tools for univariant/bivariant data (quantitative and qualitative characteristics). Missing data: profile and recovery. Detection of univariant and bivariant outliers. Association measures for multivariant data (Pearson/Spearman correlation).Example of massive data: traffic counts (missing recovery, outlier detection, filtering)

Specific objectives:
Learn how to make a report on data quality (missing data profile, univariate and bivariate outlier detection). Missing recovery.
Learn how to analyze databases including numeric and graphical univariate description, bivariate and multivariate tools in R. Determination of the significant characteristics of groups of individuals.

240ST014 - Data Analysis of Transport and Logistics

Block 3. Computational Statistical Inference	Learning time: 24h
	Theory classes: 2h Practical classes: 6h Self study : 16h

Description:
Basic statistical elements used in transportation and logistics: common univariate distributions (binomial, multinomial, Poisson, exponential, Weibull, gamma, (log)logistic, (log)normal, etc) with emphasis in moments and characteristic parameters (location, scale and shape). Input Data Analysis. Computational statistical inference for means, proportions and variances according to groups: parametrics and nonparametrics tests for (Chi2, Anderson-Darling, Wilcoxon, Kruskal-Wallis, Barlett, etc).

Specific objectives:
Learn how to use and interpret fundamental concepts in probability and statistics from a practical point of view when using R statistical software: random event, population, sample, random variable, common continuous and discrete random variates. Point and interval estimate. Computational statistical inference. Input Data Analysis and Model Fitting.

Block 4. Statistical Modeling through Regression	Learning time: 24h 30m
	Theory classes: 5h 30m Practical classes: 5h Self study : 14h

Description:
Modeling through multiple regression models. Least square estimation. Properties. Conditions for optimal inferential properties and problems when properties lack. Transforming variables. Diagnostic tools and statistics: residuals, influent data and outliers. General Linear Model: how to introduce qualitative variables as explicative variables - definition of dummy variables. Main effects and interactions between factors and covariates: model interpretation and validation. F-Test to compare nested models. Example: Modeling home to work peak morning generation trips between Transportation Zones (ZAT) with R software, lm() method and step() for best model selection.

Specific objectives:
Modeling of numeric responses: formulation, estimation and interpretation of statistical models using R software.
Model comparison and selection: Goodness of fit statistics (R^2 , F-Test for nested models, AIC, BIC, etc)
Make diagnosis of general linear models: outliers and influential data.
Learn how to forecast a numeric target using a general linear model.
Learn how apply general linear models to the generation/attraction models between Zones of Transport (ZAT).

240ST014 - Data Analysis of Transport and Logistics

Block 5. Modeling Binary Response Data	Learning time: 18h
	Theory classes: 4h Practical classes: 2h Self study : 12h

Description:
Modeling binary discrete data through generalized regression models: link function role, ML estimation, properties, model validation and interpretation. Forescasting. ROC Curve. Deviance Test to compare nested models. Case study of mode selection between public and private modes according to trip and individual characteristics: glm() for binary family in R.

Specific objectives:
Learn how to Model discrete choices with generalized linear models : formulation, estimation and interpretation of statistical models using R software.
Learn how to Perform model comparison and best model selection.
Know how to compute with R Goodness of fit statistics and related tests (X2 Pearson, Deviance Test for nested models, AIC, BIC, etc)
Learn how to perform a Diagnosis of a generalized linear models for a discrete binary choice using R.
Learn how to Apply generalized linear models for binary choice to modal split between pairs of Zones of Transport (ZAT). Understand pros and cons of Aggregated vs Disaggregated data format for models.

Block 6. Introduction to Sampling Theory	Learning time: 12h
	Theory classes: 4h Self study : 8h

Description:
Introduction to Sampling theory: random sampling and stratified sampling. Point and interval estimates for means, totals and proportions in random sampling. Selection of sample size to satisfy absolute/relative errors in random and stratified sampling. Example: Setting a homebase survey sampling size.

Specific objectives:
Know the basic principles of Sampling Theory: point and interval estimates. Learn how to compute relative vs absolute error estimates for means, totals and proportions in random sampling and stratified sampling.

240ST014 - Data Analysis of Transport and Logistics

Block 7. Introductory Data Mining	Learning time: 12h
	Theory classes: 2h Practical classes: 2h Self study : 8h

Description:
Data Mining in massive data: Useful methods for Logistics and Transportation. Classification: segmentation of the population of a study area ? hierarchical classification with R. Principle components as a tool for reduction of dimensionality. Example: Satisfaction survey for transport users of a bus network

Specific objectives:
Knowing how to turn data into information that is of use for decision making.
Learn how to perform a Profiling in R.
Learn Reduction of dimensionality strategies.
Learn how to perform a Hierarchical classification in R.
Learn how to perform a K Means in R.

Assessment: Quizz and Final Exam	Learning time: 11h
	Theory classes: 5h Self study : 6h

Description:
Assessment: Quiz and Final Exam

Related activities:
The Quiz or Partial Exam takes place in the mid-semester week, which is programmed by ETSEIB. The date depends on ETSEIB and it is not set by teachers or students.

Qualification system

The evaluation of the course integrates the three phases of learning process: knowledge, skills and competencies.

The knowledge is assessed by one quiz (partial exam) and the final exam (F1 and F2 scores), in the middle and last week of the course.

The skills and competencies are assessed from the delivery of m practices (m>1) based on the short case studies and related to the contents of the course. Each of the blocks, except the first one, might involve a practice that students will perform in group (at most 3 persons). The average of the m scores comes out the L score. Students have to quantify the hours addressed to solve each practice and deliver it through ATENEA's DUE Task. Feedback for formative evaluation will be given by the lecturer at most in 10 days before the next laboratory session when common problems and mistakes will be discussed in the first 20 min.

The final grade will obtained weighing the three scores: Final Mark = 0.65F + 0.35L. Where F is Max(F2,0.3F1+0.7F2).

240ST014 - Data Analysis of Transport and Logistics

Regulations for carrying out activities

Students can carry calculator, statistical tables, etc.. Not allowed to carry resolutions of exams from previous years, but resolutions of case studies available on the ADTL Website are allowed.

Bibliography

Basic:

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 [on line]. 2nd ed. New York: Springer, 2008 [Consultation: 05/10/2017]. Available on: <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.

Others resources:
Website Course:
- Planning Course
- Lecture Notes and slides used in lectures.
- Description of the practice sessions, questionnaires for each block and case studies.
- Case Study: Data (Excel and MS-R) and description of the context and the target variable / s.
- Guidelines for case studies presented in the form of a list of questions to guide the analysis.
- Quizzes and final exams from previous years.

Hyperlink
Web Docent ADTL
http://www-eio.upc.es/teaching/adtl/

Computer material
ATENEA - Tasques
Tasks in ATENEA to deliver Assignments

Universitat Politècnica de Catalunya