AH2170 Transport Data collection and Analysis 7.5 credits

Transport Data collection and Analysis

Intended learning outcomes

This course aims to provide knowledge on data collection and analysis methods as well as selection and interpretation of appropriate statistical tests that are relevant to the solution of the studied problem.

The intended learning outcomes are as follow:

- Identify appropriate methods for transportation, traffic and spatial data collection.
- Understand transportation and geoinformation data needs
- Understand the role sampling the data collection
- Use descriptive statistics for the analysis and preparation of data
- Perform outlier analysis
- Perform statistical inference for hypothesis testing and interval estimations
- Specify and estimate linear regression models and discrete choice models
- Apply methods and interpret results using statistical software
- Design and perform stated-preference study
- Discuss and compare linear regression models and discrete choice models and their attributes

Course main content

- Transportation and geoinformation data needs
- Sampling and sample statistics.
- Descriptive statistics and outliers
- Hypothesis testing and confidence Intervals
- Linear regression and applications (in transport and traffic)
- Maximum estimation likelihood method and applications
- Other data analysis and model building methods

The content of the course is presented and trained in tutorials. Applications are in traffic studies, transport planning, safety studies and spatial analysis. Further training in field surveys and data analysis, model building and interpretation is carried out in the form of comprehensive project work.

The project covers all the major steps that have to be undertaken including report preparation, discussion of the results. The students will also present their results for discussion.

Eligibility

Bachelor’s degree in engineering, science, economics, planning or a similar degree, with at least 60 cr (ECTS) in mathematics, physics, statistics and/or computer science, as defined in the admission requirements for the Master’s programme in Transport and Geoinformation Technology. Together with documented proficiency in English corresponding to English B.

Literature

S. Washington, M. Karlaftis, F. Mannering, Statistical and Econometric Methods for Transportation Data Analysis, Second Edition (2011).

Supplementary literature is being made available.

Other useful books:

- M. Ben-Akiva, S. Lerman, Discrete Choice Analysis: Theory and Application to Travel Demand, MIT Press, 1987.
- J. de D. Ortúzar and L.G. Willumsen, Modelling Transport (2002).
- O'Flaherty (ed.), Transport Planning and Traffic Engineering, chapter 12-13, 1997.

Lab time will be used to a) elaborate on various points made during the lectures, b) expand upon theory covered in lectures, c) work on various exercises with the assistance of a teacher, and d) discuss issues related to the current project assignment. The lab sessions take place in laboratories equipped with computers. Students may also bring their laptops to the lab sessions.

Projects/Case Study Element

Lectures and Labs

It is expected that students participate actively in lectures. Active participation in labs is highly recommended. Lab time will be used to a) elaborate on various points made during the lectures, b) expand upon theory covered in lectures, c) work on various exercises with the assistance of a teacher, and d) discuss issues related to the current project assignment. The lab sessions take place in laboratories equipped with computers. Students may also bring their laptops to the lab sessions.

Projects

Students are expected to provide results (for both individual and team work) in both written and oral form as well as provide quality feedback to other students. There are 3 case studies with the following weights:

- Case Study 1 (individual): 25%
  In this case study, the student need to be able to demonstrate the ability to handle the observation data (travel time) collected from sensors in the street. Furthermore the student is required to provide descriptive analyses of the data, recognizing patterns, and infer reasons underlie the patterns. Furthermore, the student is required to provide analytical results and variance from the patterns, find correlations between variables, and draw appropriate conclusions. Through this work, the students are expected to demonstrate their understanding about hypotheses tests, comparing and analysing confidence intervals and the effects of sample size.
- Case Study 2 (team): 25%
  In this case study, the students are given mission to develop their own regression models given specific objectives. The students then need to reflect the model fitness based on different visual tests and check whether they have violated any basic principles of the regression model. Then using the given model, the students are expected to be able to do some forecasting analyses, presenting the results in written and plots, and suggest a quantitative error measurements that can be used for the given case.
- Case Study 3 (team): 50%
  In this case study, using the national travel survey, the students learns how to estimate multinomial logit model. The students need to provide basic descriptive analyses, formulate a-priori about the variable, check the correlation between variables, and report the important patterns among the variables. Then, using MATLAB, the students should propose their best two models, with justification of those selections. The students need to present and write a report based on their case study results.

Individual projects: Although we encourage student cooperation and discussion, each individual student must write a report in his or her own words, presenting his or her own analyses.

Team projects: Teams should be comprised of two students. Students may choose teams; however, the teaching staff reserves the right to change the composition of the teams.

ILOs and examination

Intended Learning Objectives	Exam	Case Study I	Case Study II	Case Study III
Identify appropriate methods for transportation, traffic and spatial data collection.	X			
Understand transportation and geoinformation data needs	X	X	X	
Understand the role sampling the data collection		X	X	X
Use descriptive statistics for the analysis and preparation of data	X	X	X	X
Perform outlier analysis	X			
Perform statistical inference for	X	X		X

hypothesis testing and interval estimations				
Specify and estimate linear regression models and discrete choice models	X		X	X
Apply methods and interpret results using statistical software		X	X	X
Design and perform stated-preference study	X			
Discuss and compare linear regression models and discrete choice models and their attributes	X			X

How the ILOs in particular are addressed through case studies' exercises?

Intended Learning Objectives	Case Study I	Case Study II	Case Study III
Identify appropriate methods for transportation, traffic and spatial data collection.			
Understand transportation and geoinformation data needs	Use different transport data to understand the movement pattern within the given transport network	Utilising partial video observation data to estimate travel time	
Understand the role sampling the data collection	Analyse different data from different sources and time slices and reflect the impacts	Develop two models based on different period of the observed data and analyse and compare the results	Develop a-priori hypotheses about variables which likely to be important in explaining the behaviours
Use descriptive statistics for the analysis and preparation of data	Estimate mean, mode median from a number of selected datasets, compare the plots and calculate variance-covariance matrices	Compare different observed and estimated variables, create a scatter plot, and discuss	Explore the data and report average values, standard deviations, and ranges of values for the variables
Perform outlier analysis			
Perform statistical inference for hypothesis testing and interval estimations	Compare the corresponding intervals and reflect on the effect of the sample size of the used datasets		Select the best model based on a priori hypotheses, statistics, and causal relationships.
Specify and estimate linear regression models and discrete choice models		Construct two different models based on the given data, analyse the results, and reflect	Formulate and estimate and present your best two model specifications
Apply methods and interpret results using statistical software	Use excel spreadsheet to do random sampling, descriptive analysis and statistical tests	Use Excel, MATLAB or any other software that can run regressions in order to produce expected results	Using MATLAB, estimate and present your best two model specifications, and interpret the estimation results of the two models.

Design and perform stated-preference study			
Discuss and compare linear regression models and discrete choice models and their attributes			Report model comparison, explain which one better and why

Grading Scheme:

In general, as this course is teaching analytical skill to solve the problems, the grading scheme is designed based on the student ability to demonstrate their understanding of the methods and solve the tested problems given. Thus, the exam and exercises are both graded independently according to the following criteria:

A : The student has presented solutions to all parts of the problem. The solutions are clearly motivated, correct and the results are discussed thoroughly and quantitatively. Minor obvious typos can be accepted.

C: The student's solutions treat most of the problem and is largely correct but may contain computational errors and lack motivation of a few steps. A qualitative discussion of the results is present. Faulty arguments and inconsistent results can be accepted to a minor degree.

E: The student's exam demonstrates a basic understanding of the major issues and concepts treated in the problem. The student has attempted to make proper progress towards a solution to the problem. A discussion at the basic level is present.

F: A grade F is given if the criteria for a grade E are not achieved.

Examination

- PRO1 - Project Assignments, 3.5, grading scale: A, B, C, D, E, FX, F
- TENA - Written Examination, 4.0, grading scale: A, B, C, D, E, FX, F

A mandatory written examination equivalent to 4.0 cr with grading scale A-F
A mandatory project assignment equivalent to 3.5 cr with grading scale A-F

The course then will have grading scale A-F, where the course grade will be determined by the grade on the written examination and the project work.

		PROJECT				
		A	B	C	D	E
EXAM	A	A	A	B	C	C
	B	B	B	C	D	D
	C	B	C	C	D	D
	D	C	C	D	D	E
	E	D	E	E	E	E

Teaching Staff

AK: Anders Karlström, amail@kth.se

FN: Fatemeh Naqavi, naqavi@kth.se
OV: Oskar Västberg, oskarbg@kth.se

Examiner

Anders Karlström, amail@kth.se