CEE598-TSA Transportation Safety Analysis

(9:00 - 10:20 a.m. MW, 2312 Newmark Civil Engineering Laboratory, or online)
Website: Illinois Canvas (https://canvas.illinois.edu/)

Instructor

Prof. Yanfeng Ouyang
1209 Newmark Civil Engineering Laboratory, (217)333-9858; yfouyang@illinois.edu
Office hour: 10:30-11:30 a.m. MW, or by appointment

Online Office Hour URL (for online students)
https://illinois.zoom.us/j/83126680278?pwd=dXpVNkE1QkVJemdZUm1aYXhZZGJxZz09
Meeting ID: 831 2668 0278; Password: 845818

Course Description

This graduate course aims to provide students with a general understanding of data-driven transportation safety analysis (which could involve cars, trucks, pedestrians, cyclists, trains, planes, marine vessels, etc.), basic econometric theories that support such analysis, and tools that can be used to analyze a broader spectrum of transportation data. The course will present a number of statistical and econometric modelling and estimation methods for continuous, discrete count, discrete outcome, and duration data. While examples will be drawn primarily from the context of transportation safety and related infrastructures, the methods presented will have a wide variety of applications in other civil engineering areas, as well as related behavioral, economics and marketing analyses. The course will emphasize model estimation and application, the underlying theory and model assumptions will be discussed to ensure that model limitations are properly understood.

Course Topics

- Course introduction; safety engineering basics, data analysis
- Current **state-of-practice in safety engineering**; safety performance function, empirical Bayesian analysis, and benefit cost analysis
- Review of estimators and properties, least squares regression; maximum likelihood estimation
- **Linear regression models**; basic assumptions; specification errors; instrumental variable and 2-stage regressions; simultaneous equation models
- **Count-data models** (Poisson regression; negative binomial; zero-inflated models) and application to traffic crash occurrence
- **Discrete choice/outcome models**; analysis of categorical data; application to traffic crash severity
- Properties and estimation of multinomial logit models; probit models
- Data sampling for logit models (stratified, cluster, choice-based, double, enriched, and exogenous sampling); self-selectivity bias
- Nested logit; generalized extreme value models
- Ordered probability models with fixed and random effects
- **Duration models**; parametric and nonparametric estimation, and application to safety duration or infrastructure life estimation
- Discrete/continuous models; censored and truncated data;

- **Random-parameter** (linear, count, discrete choice, and duration) models; fixed and random effects models for panel data; finite mixture models
- Related research topics and guest lectures

Prerequisites:
Basic background in probability, statistics and regression analysis.

Textbook
Required readings will include chapters from the following recommended textbook and various supplementary papers/articles.

- Washington, S. Karlaftis, M.G., Mannering, F. and Anastasopoulos, P. (2020) Statistical and Econometric Methods for Transportation Data Analysis, 3rd edition, Chapman & Hall/CRC.

Students are also encouraged to use the following references as needed.

- Dalpiaz, D. Applied Statistics with R. https://book.stat420.org/index.html
- Greene, W. (2019) Econometric Analysis (8th edition). Prentice Hall, NJ.

Assignments and Grading

Homework	20%
Term Project	40%
Mid-term Exam	35%
Classroom Participation	5%

Homework: Students will be graded on completion of about 6-7 sets of homework problems.

Mid-term Exam: One in-class open-book mid-term exam.

Term Project: Students are expected to finish a term project in groups (2-3 students) by the end of the semester. The focus of the project may be anything that involves new or existing statistical models that explore transportation safety data to reveal insights. To start, students will need to decide the scope of the project, the problem of interest, the types of models to explore, and the availability of data (e.g., what data to collect, and how to collect them).

Student groups will not be formed until the third week. Before that, students are expected to work independently to explore potential topics. They are also encouraged to talk with the instructor about initial ideas. Several in-class discussion sessions are scheduled during the semester where students will share their progress. At the end of the semester, each group will make a 30-minute presentation about their project.

Acknowledgments
This course has benefited from the input of colleagues and students. We thank Drs. Fred Mannering (U. South Florida) and Simon Washington (previously with UC Berkeley) for their teaching material and slides. Former graduate student Jingwen Yao (UIUC) helped format and compile some of the slides.

Tentative Schedule (subject to change)

Week	Date	Lecture Subject	Project
1	8/24/2026	Syllabus, overview and introduction	Introduction
	8/26/2026	Safety engineering practice, data-driven analysis	
2	8/31/2026	Safety performance function, empirical Bayesian, regression to the mean	
	9/2/2026	Guest Lecture (IDOT Safety Bureau Chief Stephane Seck)	
3	9/7/2026	No class - Labor Day	
	9/9/2026	Review of estimators and properties, least squares; maximum likelihood estimation	Form groups
4	9/14/2026	Linear regression models; basic assumptions; transformation	
	9/16/2026	Linear regression models; assumption violations; specification errors	
5	9/21/2026	Simultaneous equation models	
	9/23/2026	Instrumental variable and 2-stage regressions	1-page abstract
6	9/28/2026	Tutorial: R	
	9/30/2026	Homework review and discussion	
7	10/5/2026	Instrumental variable and 2-stage regressions	
	10/7/2026	Count-data models (Poisson regression; negative binomial)	
8	10/12/2026	Count-data models (zero-inflated models) and application to traffic crash occurrence	
	10/14/2026	Logistic regression model	
9	10/19/2026	Ordered probability (probit, logit) models	
	10/21/2026	Probit and multinomial logit models for crash severity	
10	10/26/2026	Probit and multinomial logit models and random utility theory	
	10/28/2026	Data sampling for logit models	
11	11/2/2026	**Mid-term Exam**	Progress report
	11/4/2026	Nested logit; generalized extreme value models	
12	11/9/2026	Duration models; censored and truncated data, tobit model	
	11/11/2026	Continuous + discrete models; self-selectivity	
13	11/16/2026	Panel data models (fixed- and random-effects)	
	11/18/2026	Ordered probability models with random effects	
14	11/23/2026	No class - Fall Break	
	11/25/2026	No class - Fall Break	
15	11/30/2026	Logit and count models with random effects	
	12/2/2026	Duration models with random effects	
16	12/7/2026	Guest Lectures (selected research projects)	
	12/9/2026	Team project presentations	Presentations
	12/10/2026	Reading Day	**Final report**