

# CE 392E: ACQUISITION & ANALYSIS OF TRANSPORT DATA

## Spring 2019(#15543)

Lectures: 12:30 – 2:00 pm, Mondays & Wednesdays ECJ 3.122

### I. Office Hours for Dr. Kara Kockelman

Mondays & Tuesdays 2-3:30 pm, 6.904 ECJ

Or, by appointment: 471-0210 (Office phone number) & [kkockelm@mail.utexas.edu](mailto:kkockelm@mail.utexas.edu)

### II. Prerequisites

Graduate students are not required to have satisfied any specific courses as prerequisites for this course. However, those outside the transportation engineering major should receive permission from the instructor before enrolling in this course. Moreover, all students should have had *at least one college course in probability or statistics* and should know how to perform ordinary least-squares regression before enrolling in this course (or be very handy with matrix algebra & such).

### III. Grading

For purposes of grading, the performance of students enrolled in this course will be assessed using the following scoring system:

Homeworks	25% to 30% of score/grade
Course Project #1	25%
Course Project #2	25% (20% for report + 5% for presentation)
Examination(s)	25% to 20%

Notes: (1) Pluses and minuses will be used in final course grades. (2) The instructor reserves the right to consider Class Participation & Quizzes in the evaluation of a student's performance in the course, where participation score is based on participation in class (including attendance. These items may contribute up to 10% of a student's grade, reducing the other percentages proportionally.

### IV. Examinations\*

Only one course examination (a significant “midterm”) is expected, towards the end of the semester, and this may take place partly outside of lecture hours (e.g., as a take-home exam). The UT-scheduled final examination slot (Wednesday, May 15, from 2 to 5 pm) may be used for student oral presentations of final-project work, if times during the final week of class do not work well.

Major Exam: Weds., April 17, 12:30-2:00 pm (or possibly given as a take-home exam, if the students prefer that)

\* The instructor reserves the right to periodically administer, grade, and use in student evaluation “pop”/unannounced *quizzes*. Students should come to class prepared to contribute to each class's lecture and discussion by staying up-to-date with homeworks and reading.

Make-up exams will *not* generally be given to any student. If a student is absent from a scheduled exam due to medical or other problems beyond her/his control and can plainly demonstrate this, the instructor can choose to give the student a completely different exam, additional assignments, and/or change the weighting of the student's various graded contributions.

### V. Homework Assignments

Homework problems will be assigned roughly every two weeks and must be handed in at the *beginning* of the period in which they are due. After this time, they will be considered late and given *no credit*. However, **all assigned problems must be completed** (within 3 weeks of their due date and at least one week before the final exam) or a student's participation score will be adversely affected.

## VI. Course Projects

Two major homework assignments, which we can refer to as “course projects”, will be assigned. **Project 1** is a team endeavor wherein all students compose, administer, and evaluate the results of a **long-distance business-trip-focused survey** of U.S. households to get a sense of how many such trips can be “saved”/shifted to less energy-intensive travel and meeting-participation options. **Project 2** involves independent investigation, discussion, and application of a distinctive analytical approach to Project 1's survey data (or other, more extensive and relevant data sets, if the student has access to such data, including data “scraped” from online application programming interfaces (APIs)). Such work will be followed by an oral presentation to the class (of roughly 10 minutes, with 5 minutes for questions & answers). Potential Project 2 methodological topics include machine learning techniques (like CNN, SVM, and decision trees or random forests), generation of synthetic populations, endogeneity issues (and solution methods), block-diagonal survey designs, factor analysis, structural equations modeling, Bayesian estimation, spatial econometric methods, multivariate categorical response modeling, and many others.

## VII. Texts and Course Notes

The required textbook for this course is S. Lohr's *Sampling: Design and Analysis* (Duxbury Press 1999), and copies of assigned chapters are available via a Course Packet for purchase from **Canopy Course Notes** via Jerome Kubala, 512-497-6662. This Course Packet also contains several chapters from Washington et al.'s *Statistical and Econometric Methods for Transportation Data Analysis* (Chapman & Hall 2003), for those wanting to read up on several valuable methods for statistical modeling (very handy for your course projects & our Methods discussions!).

Updated versions of lecture notes will be posted periodically online, for students to download. Any additional, required materials will be made available.

Other Reading Suggestions: The first edition of Richardson et al.'s *Survey Methods for Transport Planning* (Eucalyptus Press 1995) is of some interest, and hard copies can be borrowed from the instructor and/or accessed on-line via [www.TransportSurveyMethods.com.au](http://www.TransportSurveyMethods.com.au). (Note: Many transport survey papers are available via the TUTI website: [www.tuti.com.au](http://www.tuti.com.au).) Additional readings from TRB's Travel Survey Methods Committee's Manual (at [www.travelsurveymethods.org](http://www.travelsurveymethods.org)), Stopher's recent book (*Collecting, Managing, and Assessing Data Using Sample Surveys* 2011), and various journals may be assigned.

Since the course textbook does not cover all subjects the instructor will be teaching and does not include example problems, students are likely to need to consult other texts for further reading. Recommended econometrics texts (for data analysis methods) are Mannering et al.'s *Statistical and Econometric Methods for Transportation Data Analysis* (2003, Chapman & Hall) & W.H. Greene's *Econometric Analysis* (any edition, MacMillan). Good texts on the subject of sampling are W. Cochran's *Sampling Techniques* (Wiley 1963) and C. Särndal *et al.*'s rather advanced *Model Assisted Survey Sampling* (Springer 1992). Devore's *Probability & Statistics for Engineering & the Sciences* (an undergrad text) & J. Rice's *Mathematical Statistics and Data Analysis* are useful text for students who are “rusty” on their probability and/or basic statistics.

## VIII. Add/Drop Dates

From the 1st through the 12th class day, an undergraduate student can drop a course via the web and receive a refund, if eligible. From the 13th through the university's academic drop deadline, a student may Q drop a course with approval from the Dean, and departmental advisor. After the academic drop deadline has passed, a student may drop a course only with Dean's approval, and only for urgent, substantiated, non-academic reasons.

**For graduate students:** From the 1st through the 4th class day, graduate students can drop a course via the web and receive a refund. During the 5th through 12th class day, graduate students must initiate drops in the department that offers the course and receive a refund. After the 12th class day, no refund is given. No class can be added after the 12th class day. From the 13th through the 20th class day, an automatic Q is assigned with approval from the Graduate Advisor and the Graduate Dean. From the 21st class day through the last class day, graduate students can drop a class with permission from the instructor, Graduate Advisor, and the Graduate Dean. Students with 20-hr/week GRA/TA appointment or a fellowship may not drop below 9 hours.

## IX. Evaluation Plan

UT's Course/Instructor Survey form will be used as the basic evaluation tool. All students are encouraged to submit written comments during this survey. Other formal assessment opportunities are likely to arise mid-semester; and students are strongly encouraged to provide feedback at any time during the course, in person, via other students or anonymously, to the TA and/or the instructor.

## X. Other Information

1. The University of Texas at Austin provides, upon request, appropriate academic accommodations for qualified students with disabilities. For more information, contact the Division of Diversity and Community Engagement, Services for Students with Disabilities, 471-6259 (voice) or 232-2937 (video phone) or <http://www.utexas.edu/diversity/ddce/ssd>.
2. According to *The General Information Catalog* "a student who is absent from a class or examination for the observance of a religious holy day may complete the work missed within a reasonable time after the absence, if *proper notice* of the planned absence has been given". The deadline for proper notification of such an absence is the fifteenth day of the semester.
3. Students in CE392E are encouraged and authorized to work on homework assignments together and prepare for exams together. However, all written work handed in by a student is considered to be his/her own work, prepared without *unauthorized* assistance. To ensure your actions never compromise your and our class's integrity, please visit [http://deanofstudents.utexas.edu/sjs/acint\\_student.php](http://deanofstudents.utexas.edu/sjs/acint_student.php). Students who violate University rules on scholastic dishonesty (*e.g.*, anything which gives unfair academic advantage to a student) are subject to disciplinary penalties, including the possibility of failure in the course and/or dismissal from the University. Since such dishonesty harms the individual, all students, and the integrity of the University, policies on scholastic dishonesty will be strictly enforced. An "F" grade will be the recommended penalty in most cases of scholastic dishonesty. One should refer to the Student Judicial Services website at <http://deanofstudents.utexas.edu/sjs> to access the official University policies and procedures on scholastic dishonesty as well as further elaboration on what constitutes scholastic dishonesty. For further elaboration on what constitutes scholastic dishonesty see [http://deanofstudents.utexas.edu/sjs/scholdis\\_whatdis.php](http://deanofstudents.utexas.edu/sjs/scholdis_whatdis.php)
4. Math & statistics tutors and other learning assistance can be obtained via many resources (see <http://www.engr.utexas.edu/undergraduate/97/4668-tutoring-information> including the Academic Community Center at Jester West (see <http://www.engr.utexas.edu/undergraduate/services/tutoring/jester>).

## XI. Course Objectives, Academic/Learning Goals, Questions, Content, & Schedule

This course covers various aspects of transportation relating to the acquisition and analysis of transportation data. By the end of this course, students should be able to:

- design, develop, implement and evaluate *actual surveys* of stated & revealed travel behaviors; and
- identify & apply appropriate *analytical tools* for a variety of transport data types.

They also should be able to:

- reduce response & sampling *errors*;
- maximize response rates & *data quality*; and
- recognize subtleties in *variable definitions* (in order appropriately compute needed parameters).

To attain these objectives, we will systematically proceed through a series of topic modules in class, each with specific objectives, and students will undertake two major (and several relatively minor) homework assignments. Primary topics include experimental design and sampling, survey methods and data structure, hypothesis testing, and regression analyses. A tentative scheduling of the course topics is shown here.

### **LESSON TOPICS (+ Readings):**

#### **Topic 1. INTRODUCTION** (Lohr's Ch. 1; RAM Ch. 1 & 2) – 1 lecture

Overview of Course  
Study/Data Objectives  
Discussion of Case Study & Final Projects

*\*Case Study:* Austinites' Preferences for Self-Driving Cars & Dynamic Ride-sharing\* – 1/2 lecture

#### **Topic 2. DATA TYPES** – 2 lectures

Human Behavior (& Opinion) – Stated vs. Revealed, & Cognitive Illusions (Course Notes)

#### **Topic 3. SURVEY TYPES** – 1 lecture

Self-completion vs. Interviewer; Telephone vs. Intercept (RAM Ch. 3)  
Survey Method (PAPI, CATI, Web-based...) (RAM Ch. 3 + Course Notes)  
Passive Methods (cell-phone & GPS devices) (Course Notes)

#### **Topic 4. SURVEY DESIGN & ADMINISTRATION** – 3 lectures

Questionnaire Design: Instrument Format, Wording Choice, Question Ordering (Lohr 1.5 & RAM Ch. 5)  
Minimizing Non-Response: Reminders & Incentives (RAM 7.1-7.4)  
Sensitive Questions (Course Notes)  
Need for Pilot Surveys (RAM Ch. 6)

#### **Topic 5. PROBABILITY & STATISTICS REVIEW** (Washington et al. pp. 1-44, Lohr's App. B & Ch. 2, RAM 4.6, + Course Notes) – 2 lectures

Probability (mean & covariance calculations, independence & conditioning)  
Distributions (Bernoulli, geometric, Poisson; normal, Student's t, lognormal...)  
Combining Variables (mean & variance of functions of variables)  
Statistics (bias & precision, confidence intervals)  
Hypothesis Tests

#### **Topic 6. SAMPLING METHODS** (RAM Ch. 4 & Lohr Ch. 2, 4-7) – 6 lectures

SRS – with & w/out replacement (Lohr Ch. 2)  
Systematic (Lohr 5.6)  
Stratified Sampling: Single & Multi-Stage (Lohr Ch. 4)  
Cluster Sampling (Lohr Ch. 5 & 6)

Complex Surveys (Lohr Ch. 7)

Other (systematic sampling; double/two-phase samples, choice-based sampling, Bayesian sampling...)

**Topic 7. SURVEY ISSUES** – 4 lectures

Sample Size Calculations (RAM 4.6 & 4.6; Lohr 2.5, 4.5, 7.5 ...)

Non-response & Imputation (Lohr Ch. 8)

**Topic 8. DATA ANALYSIS & APPLICATION** – 5 lectures

UNIT WEIGHTS/EXPANSION FACTORS:

Regression methods (Lohr Ch. 11)

Iterative Proportional Fitting (estimating expansion factors over 2+ dimensions) (Course Notes)

PROBABILITY: Manipulation of Distributions: The Delta Method; Converting pdfs; Length-based sampling (Course Notes)

REGRESSION MODELS (OLS, WLS, Discrete Choice, Systems of Equations) (Course Notes & Washington et al. Ch. 3, 7, 10 & 11)

Note: Rice's & Greene's textbooks are valuable for further details on Topic 8 topics.

\* Review for Course **Examination** \* – 1 lecture

**Topic 9. OTHER TOPICS** – 2 lectures

Potential topics: SPATIAL STATISTICS, DATA PROCESSING & STORAGE, REMOTELY SENSED DATA, RANDOM NUMBER GENERATION, ETC.

**STUDENT PRESENTATIONS of Project #2 Work** – 2 lectures (or in final exam time slot)