Linear Models and Experimental Design
HUDM 5123
3 Credits, Fall 2019

Instructor: Bryan Keller, PhD   Office: GDH 456B
Class Meeting Time: M 11:00 am – 1:40 pm   Room: HM 433
Office Hours: T 12 – 2 pm; R 12 – 1 pm   keller4@tc.columbia.edu

TA: Aakash Kumar   Office Hours: W 3 – 5 pm   ask2229@tc.columbia.edu

----

Email is the best way to get in touch with me. If you send me a note in Canvas, my response may be delayed.

Overview:
This course provides an overview of experimental design and analysis from the perspective of the general linear modeling framework. Topics include the incremental F test for model comparisons, dummy and effect coding, single and multiple factor ANOVA and ANCOVA, analysis of categorical outcome data via generalized linear models, and repeated measures. The course includes lab time devoted to computer applications.

Prerequisite: HUDM 5122 or HUDM 5126 or another graduate-level course in regression analysis.

Required text:
- Fox, J. (2008). *Applied Regression Analysis and Generalized Linear Models* (2nd or 3rd ed.)

Other useful books (on reserve at Gottesman Library):
- (MD) Maxwell, S. E. & Delaney, H. D. (2004). *Designing Experiments and Analyzing Data: A Model Comparison Perspective* (3rd ed. – 2nd ed. OK too).
- Keppel G. & Wickens, T. D. (2004). *Design and Analysis: A Researcher's Handbook* (4th ed.).
- (HG) Hedeker & Gibbons (2006). *Longitudinal Data Analysis*.
- Field, A. (2013). *Discovering Statistics Using IBM SPSS Statistics* (4th ed.).
- (APA) Publication Manual of the APA (6th ed.)
- Fox, J. and Weisberg, S. (2011). *An R Companion to Applied Regression (2nd ed.)*.

Computers and software:
R and RStudio are both open source and freely available. Download R from the Comprehensive R Archive Network (CRAN) here: https://cran.r-project.org/; download RStudio from the RStudio homepage here: https://www.rstudio.com/. Some resources for more information on R:
- Stack Overflow. Searchable questions and answers.
- Quick-R. Resources and cheat sheet-style primers.
- Intro to R. Online resource that covers the basics with emphasis on statistical modeling.

Grading:
Grades will be determined by labs (30%, in aggregate), an in-class exam (35%), and a data analysis project (35%). Regular attendance is required. Numeric grades will be assigned as follows:

Average	Rare	[93,100]	[90,93)	[85,90)	[80,85)	[75,80)	[70,75)	[65,70)	[60,65)	[0,60)
Grade	A+	A	A-	B+	B	B-	C+	C	C-	F

For the full text of the grading symbols approved by Teachers College Faculty please refer to
http://www.tc.columbia.edu/policylibrary/Grading

Tentative Course schedule:

Class & Date	Topic	Readings
General Linear Models for Continuous Outcome Data-------------------------------------		
01. Sep 09	OLS Regression and Assumptions	Fox, ch 5 & 6
	Lab: Intro to R & OLS Regression	
02. Sep 16	Diagnostics for Linear Models	Fox, ch 11, 12, & 13
	Lab 01: OLS Regression Diagnostics	
03. Sep 23	Categorical Predictors	Fox, ch 7 & 8.1
	Lab 02: ANOVA: Dummy & Effect Coding	
04. Sep 30	Continuous Covariates	Fox, ch 8.4
	Lab 03: ANCOVA	
05. Oct 07	Modeling Interactions	Fox, ch 8.2
	Lab 04: Two-Way ANOVA	
06. Oct 14	Linear Contrasts	Fox, ch 8.5 & 9.1
	Lab: Same	
07. Oct 21	IN-CLASS EXAM	
08. Oct 28	Correcting for Multiple Comparisons	MD, ch 5
	Lab: Same	
Generalized Linear Models for Categorical Outcome Data-------------------------------		
09. Nov 04	Introduction to Categorical Outcome Data	
	Lab: Same (PROPOSAL DUE TODAY)	
10. Nov 11	Models for Dichotomous Data	Fox, ch 14.1
	Lab 05: Logistic Regression	
11. Nov 18	Models for Count Data	Fox, ch 15.1, 15.2, & 15.4
	Lab 06: Same	
Repeated Measures ANOVA and Alternatives---		
12. Nov 25	Introduction to Repeated Measures	HG, ch 1
	Lab: Same	
13. Dec 02	ANOVA Approaches to Repeated Measures	HG, ch 2
	Lab 07: Same	
14. Dec 09	Mixed Effects & Covariance Pattern Models	HG, ch 4, 6, & 7
	Lab 08: Same	
15. Dec 16	Final Project Q & A; Attendance Optional	

*Note: only numbered labs will be turned in for a grade.

Data-Analysis Project Description:

The goal of the project is to apply methods covered in the course to answer meaningful research questions with a real data set. Data that are appropriate for this project will have the following characteristics:

- a primary outcome variable (continuous, dichotomous, or count),
- moderate to large sample size, and at least two of the following:
    1. a categorical predictor for one-way ANOVA (ideally with units randomly assigned to groups, but OK if not),
    2. a continuous predictor that is correlated with the outcome for ANCOVA,
    3. a second categorical predictor for two-way ANOVA.

The research hypotheses should be formulated so that they line up with omnibus ANOVA/ANCOVA null hypothesis tests and follow-up tests using linear contrasts. These hypotheses should be clearly described. You will need to use some form of Type I error rate control for testing contrasts. It is easier to come up with meaningful research hypotheses with data you are familiar with, so I recommend using data you know. If you have never worked with data before, or if the data you are familiar with does not qualify (e.g., because you don't have permission to use it or it doesn't meet the requirements in some other way), use a publicly available database. Here are some links to publicly available resources:

- data.gov – open data from the US government
- PLOS ONE – open access journal that requires all data be made publicly available
- The National Center for Education Statisticsd, including ECLS
- The Youth Risk Behavior Survey, available from the CDC
- The Current Population Survey, available from the U.S. Census Bureau
- The Fragile Families & Child Wellbeing Study, available from Princeton University

APA style is required unless your field uses a different style. If you plan to use other than APA style for your paper, please notify me which style you will use when you submit the proposal.

Data-Analysis Project Key Dates:

- November 4 (Monday, by 11:59 PM) – **PROPOSAL** – Turn in a short (approx. 2 pages) paper describing your data set and primary research questions. Your description of the data set should include information on the number of cases, brief summaries of relevant variables, research hypotheses, and graphs or tables where appropriate. Describe the research questions and the statistical methods you will use to investigate them. Also discuss Type I error rate control strategy for the set of linear contrasts. **I will comment on your proposal with feedback and questions in Canvas. Please check for my comments and respond promptly on the same thread in Canvas.**
- December 17 (Tuesday, by 11:59 PM) – **FINAL PAPER** – A typical length for the final paper is about ten pages of double-spaced 12-point font, not including tables, figures, or appendix. Your paper should include
    1. an **introduction**,
    2. a **body** in which
        - research questions are clearly defined and linked to analyses and hypothesis tests,
        - methods used are described accurately, including formulas as needed,
        - assumptions and their tenability are discussed with evidence including plots and/or statistical tests,
        - results are described in text and with tables and plots (see APA Publication Manual checklists for formatting),
        - the meaning and interpretation of the results is discussed, and
    3. a **conclusion** that summarizes the research questions, results of the study and their essential implications, and limitations.
    4. a **references** section, and
    5. an **appendix** with commented computer code.

The Provost and Dean of the College in conjunction with the Faculty has adopted the following statements to be included on all Teachers College syllabi.

1. **Accommodations** – The College will make reasonable accommodations for persons with documented disabilities. Students are encouraged to contact the Office of Access and Services for Individuals with Disabilities (OASID) for information about registration. You can reach OASID by email at oasid@tc.columbia.edu, stop by 163 Thorndike Hall or call 212-678-3689. Services are available only to students who have registered and submit appropriate documentation. As your instructor, I am happy to discuss specific needs with you as well. Please report any access related concerns about instructional material to OASID and to me as your instructor.
2. **Incomplete Grades** – For the full text of the Incomplete Grade policy please refer to http://www.tc.columbia.edu/policylibrary/Incomplete Grades
3. **Student Responsibility for Monitoring TC email account** – Students are expected to monitor their TC email accounts. For the full text of the Student Responsibility for Monitoring TC email account please refer to http://www.tc.columbia.edu/policylibrary/Student Responsibility for Monitoring TC Email Account
4. **Religious Observance** – For the full text of the Religious Observance policy, please refer to http://www.tc.columbia.edu/policylibrary/provost/religious-observance/
5. **Sexual Harassment and Violence Reporting** – Teachers College is committed to maintaining a safe environment for students. Because of this commitment and because of federal and state regulations, we must advise you that if you tell any of your instructors about sexual harassment or gender-based misconduct involving a member of the campus community, your instructor is required to report this information to the Title IX Coordinator, Janice Robinson. She will treat this information as private, but will need to follow up with you and possibly look into the matter. The Ombuds officer for Gender-Based Misconduct is a confidential resource available for students, staff and faculty. "Gender-based misconduct" includes sexual assault, stalking, sexual harassment, dating violence, domestic violence, sexual exploitation, and gender-based harassment. For more information, see http://sexualrespect.columbia.edu/gender-based-misconduct-policy-students.

Emergency Plan:
TC is prepared for a wide range of emergencies. After declaring an emergency situation, the President/Provost will provide the community with critical information on procedures and available assistance. If travel to campus is not feasible, instructors will facilitate academic continuity through Canvas and other technologies, if possible.

1. It is the student's responsibility to ensure that they are set to receive email notifications from TC and communications from their instructor at their TC email address.
2. Within the first two sessions for the course, students are expected to review and be prepared to follow the instructions stated in the emergency plan.
3. The plan may consist of downloading or obtaining all available readings for the course or the instructor may provide other instructions.