Categorical Data Analysis (STAT 453/653) Fall 2019
Davidson Mathematics and Science (DMS), room 106
TR noon-1:15PM
3 credits

Instructor: Ilya Zaliapin **Office:** Davidson Math & Science (DMS), Room 221
Office hours: TR 11:00AM–11:50PM + by appointment
Phone: (775) 784-6077 **E-mail:** zal@unr.edu

Registered students will be able to access the course via Canvas (Web Campus)

Intro: Categorical data are routinely collected in social, geophysical, behavioral, biomedical, biological and agricultural sciences as well as in public health, marketing, education, and industrial quality control. The statistical inference for categorical data involves methods and approaches that differ (sometimes significantly) from those for quantitative random variables. Real-life examples, classical and modern statistical techniques, and numerical methods discussed in this class aim at answering the main question: How to make an efficient inference if our observations are categorical? A professional statistical package  will be used to apply the concepts discussed in the class to real data sets.

Required textbook:
Alan Agresti (2007) *An Introduction to Categorical Data Analysis,* 2nd ed., Wiley.

Optional reading:
Johnson, N. L., Kemp, A. W., and Kotz, S. (2005) *Univariate Discrete Distributions,* 3rd ed., Wiley.

Tentative Topics:
- Gathering and Organization of Categorical Data
- Discrete Distributions
- Basic Inference: Proportions, Odds Ratios, Fisher Exact Test, Chi-Square test
- Contingency Tables (two-way, multi-way, Simpson paradox)
- Cochran-Mantel-Haenszel test
- Generalized Linear Models (GLM)
- Logistic Regression Model (linear, non-linear, interactions, inference)

Student Learning Outcomes: Students will be able to
- Perform statistical inference for categorical data organized in contingency tables (including odds ratio, Fisher test, Chi-square test, inference for proportions, multiway contingency tables, Cohran-Mantel-Haenszel test).
- Build, analyze, and interpret generalized linear models for categorical data, including multivariate logistic regression model with linear and nonlinear terms.
- Perform categorical data analysis using professional statistical software and present results in technical reports and to professional audience.

Midterm: There will be one midterm preliminary scheduled on Thursday, October 10.

Final exam: Final exam will be given on Monday, December 16, 12:10-2:10PM.

Exam policy (for midterms and final): Closed books, closed notes. There will be **no make-ups** for exams, except legitimate medical reasons. In case of participating in

University-related activities or in any other special circumstances, contact instructor **in advance**.

Final group project (STAT 653 only): The project will consist of comprehensive theoretical and practical statistical analysis of a data set of your choice; it will result in a **project report** and a short **presentation** that will be delivered at the end of the semester to the class. The project should demonstrate that you are (i) well familiar with essential concepts, methods, and techniques studied in the class; (ii) able to use package R for analysis and report preparation; (iii) ready to present your findings to a professional audience. A detailed discussion of how to successfully complete the project will follow in class.

Statistical Lab is an integral part of the course. You will be given regular take-home assignments that require application of selected statistical techniques using the package R. The results should be presented in a form of illustrated reports (we will discuss the report writing in the class). The previous knowledge of R is not required (but definitely will make the class easier). R is a free software; the R-portal with downloads, manuals, FAQs, and much more is located at: http://www.r-project.org/. You are encouraged to discuss the Lab assignments with other students and instructor and can do them in groups, but your reports have to be written individually and demonstrate that you are able to perform the presented analysis independently.

Home works will be given approximately weekly. They consist of theoretical problems and should be done with pen and paper (i.e., no web system will be used).

Home work and Lab policies: You are encouraged to discuss HW and Lab assignments between each other and with instructor during office hours. The works must be prepared individually though. A home work is due on the date announced on the course web site, by the end of the class. E-mails with HW/Lab reports cannot be accepted. Late HW/Labs can be accepted under special circumstances (please discuss this with instructor) and given a maximum of half credit (50%).

Grading policy for STAT 653: Your letter grade for the course will be based on home works (20%), lab reports (20%), midterm (20%), final exam (20%), and group project (20%).

Grading policy for STAT 453: Your letter grade for the course will be based on home works (30%), lab reports (30%), midterm (20%), and final exam (20%).

Letter	A	A-	B+	B	B-	C+	C	D+	D	F
Min. Score	93%	90%	87%	83%	80%	77%	70%	67%	60%	0%

Prerequisites: STAT 352 or STAT 467/667.

Re-grading: If you find that your grade for exam, HW, Minitab assignment, or quiz is incorrect, contact instructor in person with a rational justification. Have the assignment that is being discussed with you. All re-grading requests must be submitted to instructor within one week after the discussed grade is announced; late requests will not be granted. The final decision about the new grade is made by the instructor.

Communication: Matters such as class cancellations, meeting times or room changes will be announced via the course web site (Canvas). It is a student responsibility to check the web announcements.

Graduate/Undergraduate levels: Graduate students will achieve deeper understanding of the material and will be offered sufficient opportunities for work at a higher academic level. This will be done by choosing different quality and quantity of assignments for homeworks and exams; in addition, a comprehensive data analysis project will be expected from graduate students.

Academic dishonesty will not be tolerated and will lead to an F grade. See http://www.unr.edu/stsv/acdispol.html

Class recording policy: Surreptitious or covert video-taping of class or unauthorized audio recording of class is prohibited by law and by Board of Regents policy. This class may be videotaped or audio recorded only with the written permission of the instructor. In order to accommodate students with disabilities, some students may have been given permission to record class lectures and discussions. Therefore, students should understand that their comments during class may be recorded.

Disability Services: Any student with a disability needing academic adjustments or accommodations is requested to speak with the Disability Resource Center (Thompson Building, Suite 101) as soon as possible to arrange for appropriate accommodations.

Academic success: Your student fees cover usage of the University Math Center (775) 784-4433, University Tutoring Center (775) 784-6801, and University Writing Center (775) 784-6030. These centers support your classroom learning; it is your responsibility to take advantage of their services. Keep in mind that seeking help outside of class is the sign of a responsible and successful student.

Title IX: The University of Nevada, Reno is committed to providing a safe learning and work environment for all. If you believe you have experienced discrimination, sexual harassment, sexual assault, domestic/dating violence, or stalking, whether on or off campus, or need information related to immigration concerns, please contact the University's Equal Opportunity & Title IX office at 775-784-1547. Resources and interim measures are available to assist you. For more information, please visit: https://www.unr.edu/equal-opportunity-title-ix

Tentative schedule (may change as class progresses)

Week	Tuesday	Topic	Thursday	Topic
1	Aug. 27	Intro, **1.1-1.4**	Aug. 29	Contingency tables**, 2.1**
2	Sep. 3	Difference of proportions, **2.2**	Sep. 5	Odds, odds ratio, **2.3**
3	Sep. 10	Chi-square tests, **2.4**	Sep. 12	Fisher exact test, **2.6**
4	Sep. 17	3-way tables, **2.7**	Sep. 19	3-way tables, **2.7**
5	Sep. 24	**Lab 1a:** CT, goodness-of-fit tests	Sep. 26	**Lab 1b:** CT, goodness-of-fit tests
6	Oct. 1	GLM, **3.1 – 3.2**	Oct. 3	Logistic regression, **4.1**
7	Oct. 8	**Lab 2:** Logistic regression, single explanatory variable	Oct. 10	**Midterm 1**
8	Oct. 15	Logistic regression: **4.2**	Oct. 17	Logistic regression: **4.3**
9	Oct. 22	Logistic regression: **4.4**	Oct. 24	Logistic regression: **4.4**
10	Oct. 29	**Lab 3:** Logistic regression, multiple explanatory variable	Oct. 31	Loglinear models: **7.**1 (2-way)
11	Nov. 5	**Lab 4a:** Logistic regression, non-linear terms	Nov. 7	**Lab 4b:** Logistic regression, non-linear terms
12	Nov. 12	Loglinear models: **7.3**	Nov. 14	Loglinear models: **7.3**
13	Nov. 19	Loglinear models: **7.1** (3-way)	Nov. 21	Independence graphs: **7.4**
14	Nov. 26	**Lab 5:** Loglinear models	Nov. 28	Thanksgiving
15	Dec. 3	**Lab 6:** Poisson models	Dec. 5	Matched pairs: **8.1-8.2**
16	Dec. 10	Project presentations		
Monday, Dec. 16, noon-2:10 – Final				