ED PSY 826: Analysis of Categorical Data

Instructor: Razia Azen
e-mail: azen@uwm.edu
- I do my very best to respond to email within 24 hours during regular business hours. If you do not get a response from me within a couple of work days, please feel free to follow up with a reminder. Note that I may not check and respond to email in the evenings or on weekends.
- Please do check the Canvas course page before emailing to see if the information is posted there. For example, if I already received some questions about a homework problem, I may have posted clarifications on Canvas. If you cannot find the information, do feel free to send me an email.

Office Hours: Please **contact me by e-mail to set up a time** to meet, as necessary.
- We can meet in person, online, or speak by phone.
- Please let me know at least a couple of days in advance of when you would like to meet so we can set up a mutually convenient meeting time and modality.

Course Objectives: This course will introduce students to techniques appropriate for analyzing categorical data (i.e., information measured in categories rather than on a continuous scale). The course will be taught from an applied perspective, and the main objectives are to enable students to:
- Identify the most appropriate analysis for research questions pertaining to categorical data.
- Use statistical software to conduct the data analysis and interpret the statistical results appropriately.
- Clearly communicate the results of the statistical analyses to address the research questions of interest.

By the end of the course students should be able to identify the categorical analysis procedure most appropriate to address a particular research question with a data set, carry out the analysis, and properly interpret (i.e., communicate) the results. The course will cover common categorical data analysis techniques that students may need to use in their own research or encounter in reading research articles.

Pre-requisites: A graduate-level statistics course (e.g., ED PSY 724) that included factorial ANOVA and multiple regression as well as statistical computing (e.g., SAS, SPSS, R). A full course on regression and/or ANOVA is highly recommended.

Textbook: Azen, R., & Walker, C. M. (2021). Categorical Data Analysis for the Behavioral and Social Sciences, 2nd edition, published by Routledge (Taylor & Francis).
- *Optional*: Agresti, A. (2007). An Introduction to Categorical Data Analysis, 2nd Edition. New York: Wiley. *This book is a bit more technical; Agresti is considered an expert in this field.*
- We may also discuss journal articles or other readings (or even data on current events) of relevance throughout the course.

Computing: Students will be required to use statistical software (e.g., SAS, SPSS, or R). Students are expected to already be proficient in at least one software package. The emphasis will be on demonstrating the use of SAS. Students who wish to use other software will need to do so on their own.
- Further software information and instructions can be accessed on the course Canvas site under "Computing Resources".

Course Web Site and Delivery: This course will make extensive use of its **Canvas** site.
- The Canvas site can be accessed at UWM Canvas Home and instructions/support for using Canvas are available at Student Canvas support.
- Students are expected to check the site on a regular basis for announcements, notes, assignments, discussions, and any other course information.
  - You can specify your preferred notification settings in Canvas. I recommend that you enable the notifications for the Announcements, at least.

- I will plan to post the materials, including **notes** (with and without a voice-over recording), no later than <u>Monday</u> each week. The notes documents will be up to date and may differ slightly from the notes in the recordings (if updates or corrections were needed since the recordings were made).
  - o You are expected to review the materials and attempt the provided **learning checks** (*not* counted towards the course grade) **prior** to the class meetings and before starting the homework assignments (which *will* count towards the course grade).
- We will meet on **Thursdays at 2:30pm**. As of now (August 2022), we plan to hold class meetings in person (in our assigned classroom, Enderis Hall 109). I hope that remote participation will be possible and that we are able to adapt as necessary and/or preferred.
  - o In the class meetings I will provide clarifications and students will have the opportunity to ask questions about any of the material (including assignments and learning checks) and discuss any other course information.

Course structure and student evaluation:

- *Homework:*
  - o Homework problem sets will generally be **assigned every week** (based on that week's material) and will be submitted on Canvas every 2-3 weeks.
  - o The assignments will typically be due on Thursdays, approximately every 2-3 weeks. Specific due dates will be posted on Canvas. (Note that assignment points may show up as 0 on Canvas but this will be updated once assignments are posted/graded.)
  - o <u>Late homework assignments</u> will be accepted with a **10% point-deduction** for each day they are late (in other words, you will lose 10% of the total number of points for the assignment for each day it is late, including weekend days). This penalty may be reduced or exempted if the reason for late submission is acceptable to the instructor, so *please communicate with me (the instructor) if there are extenuating circumstances and/or your homework will be late*. It is preferable that you submit an assignment that reflects **your best work** and is a bit late than one that does not reflect your best work but is on time.
  - o Homework assignments will consist of problems that are intended to give you some hands-on practice with the material; you are expected to clearly demonstrate and communicate your understanding in your answers (see also the **Homework Guidelines** document on Canvas).
  - o Each problem (or major component) will be graded on a three-point scale to indicate the level of accuracy and understanding reflected in it:

Points	Description
3	Complete, clear and correct.
2	Some mistakes and/or misconceptions, somewhat unclear or incomplete.
1	Many mistakes and/or major misconceptions, very unclear or incomplete.
0	Not done or barely attempted.

  - o Homework grades and comments are designed to provide you with feedback on the level of understanding conveyed in your assignment. If you find the feedback insufficient, it is **your responsibility** to make sure that you ask and understand it. Do not put off getting help if you don't know how to do a problem or do not understand the feedback you received. Better yet, if you are unsure how to do a problem, ask *before* the assignment is submitted. (Note that I will not "pre-grade" your answers but will be happy to clarify anything that is unclear to you.)
  - o Please feel free to post any questions (or helpful information for other students) in the **Discussions** area on Canvas.
  - o <u>A note on working with others</u>: The goal of the homework assignments is to make sure you are learning and understanding the material, so it would defeat the purpose if you get the right answer without really understanding why. Thus, while you are allowed to discuss the concepts on homework assignments, it is recommended that you complete the assignments on your own. In addition, **the work you submit must be your own**, and it is considered academic misconduct to submit anyone else's work (or words) as your own.

- *Course projects:*
  - o Purpose: The project is designed to give you the opportunity to apply the procedures we cover in the course to address research questions that are (potentially) of interest to you. You will need to obtain a data set, preferably of interest to you and/or from your field of study, that can be analyzed using the procedures covered in the course.
    - ▪ The data set should contain at least three categorical variables, one of which might be considered an outcome or response variable.
    - ▪ I plan to include a few progress checks to make sure that you have an appropriate data set for the project and resolve any issues you may have with it.
    - ▪ The data set for the course project can consist of data collected for another purpose (e.g., your research), but the projects should be original and should **not** consist of analyses conducted for another class or purpose.
  - o Groups: Given the size of the class, and to encourage interaction with classmates, these will be group projects with 2 students per group. Group members are expected to contribute equally to the project, and students may have an opportunity to evaluate the contributions of their project partner.
    - ▪ Feel free to try and find a partner through the Discussions section in Canvas.
    - ▪ Please let me know early on if you do not feel that you will be able to work with another student and will need to work on the project on your own.
    - ▪ There is a survey posted on Canvas where you can indicate/request a partner. Partners should be set by about the third week of the semester.
  - o Components: The project will have **two components**. The first component will be due around the middle of the semester and the second will be due at the end of the semester. The same data set can be used for both components (but this is optional).
    - ▪ Each component will generally require that you specify a relevant research question, carry out the appropriate analysis to address it, and submit a summary report.
    - ▪ Alternatively, students may choose a project topic that is more methodological or theoretical in nature (e.g., a simulation study to evaluate the performance of a relevant analytic method).
  - o Peer feedback/participation: Each student will be asked to comment and provide feedback on the first component for **at least two other projects**. The goal of this is to provide friendly, constructive feedback to your classmates so they can incorporate it into the second component. To encourage this participation, it will constitute a small part of the course grade.
  - o The projects will be graded based on the accuracy of the analyses as well as the clarity and level of understanding demonstrated in the written summary.
  - o Further details on all components will be posted on Canvas closer to the due dates.

- *Exams:*
  - o There will be a final exam ("take-home") that will be due during finals week.
  - o Students are expected to **complete the exam independently** and with the utmost regard for academic integrity (i.e., the exam should be treated as if it is in-class). Students will **not** be allowed to discuss the exam with anyone (other than the instructor) and violations of this will be treated as academic misconduct.

The weights assigned to each of these components will be:

Homework		**25%**
Projects:	**Project 1**	**10%**
	Peer feedback	**5%**
	Project 2	**25%**
Final exam		**35%**

There will be NO extra credit option in this course.

Using these weights (the homework assignments will be weighted by the number of points each assignment is worth), final scores (out of 100) will be computed and these will be converted to letter grades as follows:

A	A-	B+	B	B-	C+	C	C-	D+	D	D-	F
90-100	85-89	80-84	75-79	70-74	67-69	63-66	60-62	57-59	53-56	50-52	below 50

Academic Integrity:
- It is better to receive an honest grade on your own work than to risk committing plagiarism or academic misconduct. If you are unsure as to what is acceptable, or need help, please ask.
- Excerpts from UWM's Academic Integrity Online page:
  - Ask your instructor if you are unsure about how to complete an assignment or course requirement appropriately.
  - When collaborating with other students, collaborate for a better understanding of the material, not for answers. *Note that no collaboration of any kind is allowed on the exam in this course*
  - If you are unable to complete assignments, the instructor may be flexible with the deadlines, and receiving a late penalty is far better than academic dishonesty.
- There is additional information and resources for avoiding plagiarism at Plagiarism.org

Panther Community Health and Safety Standards: UWM has implemented health and safety protocols, taking into account recommendations by local, state, and national public health authorities, in response to the COVID-19 pandemic. As a member of our campus community, you are expected to abide by the Panther Interim COVID-Related Health & Safety Policy, which was developed in accordance with public health guidelines. These standards apply to anyone who is physically present on campus, UWM grounds, or participating in a UWM-sponsored activity:
- UWM recommends that all individuals visiting UWM facilities wear face coverings while indoors.
- UWM recommends getting vaccinated for COVID-19 and getting the most recent booster shot available to you.
- UWM requires that you check daily for COVID-19 symptoms and not come to campus if you are feeling sick. If you are feeling sick, get tested for COVID-19 and quarantine until symptoms subside. Use the CDC Quarantine and Isolation Calculator to determine next steps.
- If you test positive for COVID-19, UWM requires that you self-report at the Dean of Students Reporting Form. Use the CDC Quarantine and Isolation Calculator to determine next steps.

Additional details about student and employee expectations can be found on the UWM COVID-19 webpage.

University Policies: Policies regarding participation of students with disabilities, accommodations for religious observances, academic misconduct, student complaints, grade appeals, sexual harassment, attendance, assignment of a grade of "incomplete", etc., are available at this Syllabus Links document. *Students should review these policies at the start of the course*.

Time Investment: This will vary by student and by week, but my expectation is that students will spend a total of about 150 hours on the course (as would be the case for an in-person 3-credit course). This document provides a more detailed breakdown. This is an estimated workload and students will be assessed on their performance (as indicated in the syllabus), not on the time put into the course.

TENTATIVE SCHEDULE

Week	Date	Topic	Reading*
1	Sept. 8	Introduction; Scales of measurement	A&W: Chapter 1 *A: Sections 1.1-1.2, 11.1-11.2*
2	Sept. 15	Probability distributions	A&W: Chapter 2
3	Sept. 22	Inference for a proportion	A&W: Chapter 3 *A: Sections 1.3-1.4*
4	Sept. 29	Two-way contingency tables: odds-ratio	A&W: Sections 4.1-4.3 *A: Sections 2.1-2.3*
5	Oct. 6	Inference: tests of independence	A&W: Sections 4.4-4.8 *A: Sections 2.4, 2.6*
6	Oct. 13	Three-way contingency tables	A&W: Chapter 5 *A: Section 2.7*
7	Oct. 20	Three-way contingency tables, *continued* Generalized linear models	A&W: Chapter 5 *A: Section 2.7* A&W: Chapter 6 *A: Chapter 3*
8	Oct. 27	Log-linear models ➢ *Project 1 due*	A&W: Chapter 7 *A: Chapter 7*
9	Nov. 3	Log-linear models, *continued*	A&W: Chapter 7 *A: Chapter 7*
10	Nov. 10	Logistic regression (with continuous predictors)	A&W: Chapter 8 *A: Sections 4.1-4.2, 4.4-4.5, Chapter 5*
11	Nov. 17	Logistic regression (with categorical predictors)	A&W: Chapter 9 *A: Sections 4.3, 5.3.1, 7.3*
12	Nov. 24	*~ Thanksgiving break ~*	
13	Dec. 1	Multicategory logit models	A&W: Chapter 10 *A: Chapter 6*
14	Dec. 8	Review ➢ *Project 2 due*	
Finals week	Dec. 16	➢ *Final exam due*	

*Note: A&W = Azen & Walker (2021); *A* = Agresti (2007)