



FACULTY OF SOCIAL SCIENCES

International Laboratory for Applied Network Research

Applied Statistics with Network Analysis (MASNA) Program

Categorical Data Analysis Course Syllabus

Professor: TBD
Office: TBD
Office Hours: TBD
Email: TBD

Class Hours: TBD

Text & Materials

1. Required readings:

- Agresti, A. (2007). An Introduction to Categorical Data Analysis, 2nd Edition. New York: Wiley (primary text).

2. Optional readings:

- Agresti, A. (2013). Categorical Data Analysis, 3rd Edition. New York: Wiley (an advanced text).

There are some additional readings that you are responsible for throughout the semester; they will be handed out, as necessary, during the course.

3. **Software:** We will cover this course in SAS. You already have some fundamental knowledge of SAS, but our textbook by Agresti is written precisely for this course in SAS, so you will have a chance to enhance your SAS skills.

I. Course Summary

This course is designed to introduce basic concepts and common statistical models and analyses for categorical data; to provide enough theory, examples of applications in a variety of disciplines (especially in social and behavioral science); and practice using categorical techniques and computer software so that students can use these methods in their own research; to attain knowledge necessary to critically read research papers that use such methods.

II. Field of application and normative references

The program of the course builds up the minimum requirements to the knowledge and skills of a student and determines the content and types of classes and assessment. The program is designed for instructors of this discipline, teaching assistants and students of the Master's program "Applied Statistics with Network Analysis." The program is developed according to:

- Educational standards of NRU HSE for the master's programs in "Applied mathematics and informatics" field, approved by HSE Scientific Council in December of 2017
- University Academic Plan of NRU HSE for Master level education, developed in 2014.

III. Course Overview & Objectives

This class deals with the most fundamental regression models for binary, ordinal, nominal and count outcomes. While advances in software make it simple to estimate these models, post-estimation interpretation is difficult due to the nonlinearity of the models. The class begins by considering the general objectives for interpreting results from any regression model and then considers why achieving these objectives is more difficult when models are nonlinear. Basic concepts and notation are introduced in a quick review of the linear regression model. In this context the ideas of maximum likelihood estimation and identification are introduced. These ideas are the used to develop the binary logit and probit models. For these models numerous methods of interpretation are presented. Wald and LR tests are illustrated along with how to analyze complex samples. The binary model is extended to explore problems related to “nonlinear nonlinear” models and methods for comparing groups. Next, these ideas are extended to nominal outcomes, focusing on the multinomial logit model. The ordinal logit and probit models are then presented along with several less common models for ordinal outcomes. Restrictions imposed by ordinality are explored and ways to chose between ordinal and nominal models are considered. Finally, a series of models for count data, including Poisson regression, negative binomial regression, and zero modified models are presented. A major focus of the class is showing how to use SAS for sophisticated interpretation of nonlinear models. The course assumes familiarity with the linear regression model.

The Course develops the following competencies

Competencies	NC/NRU-HSE Code	Descriptors - the learning outcomes (the indicators of achievement)	Teaching forms and methods of that contribute to the development of a competence
<u>Systemic Competencies</u>			
1. Ability to independently learn new research methods, change the scientific and production profile of their activity.	CK-3	Can use new methods and techniques of network analysis, additional packages and tools, without direct supervision.	Lectures, readings, in-class exercises, data analysis projects
2. Ability to analyze, verify, evaluate the completeness of information in the process of their professional activities, to replenish and synthesize missing information, if necessary.	CK-6	Able to analyze, verify, evaluate the completeness of information, can integrate information found from various sources and compensate for lack of data by adjusting models.	Lectures; independent work.
3. Ability to generate fundamentally new ideas and products, to hold creativity and initiative.	CK-8	Can effectively generate new ideas and products in their professional activities.	Lectures, readings, in-class exercises, data analysis projects
<u>Instrumental Competencies</u>			
4. Ability to solve the tasks of professional activity as a part of	ПК-17	Can solve different tasks of professional activity.	Lectures, in-class exercises, data analysis

Competencies	NC/NRU-HSE Code	Descriptors - the learning outcomes (the indicators of achievement)	Teaching forms and methods of that contribute to the development of a competence
the research and production team in accordance with the profile of training, to communicate with experts in other subject areas.		Effectively communicate with experts from other areas.	projects

This course will emphasize advance preparation for each class period and will involve a high level of class participation. Often, experiential exercises and simulations will be used to illustrate key network analysis concepts. While I do not take attendance, missing classes on a regular basis will be detrimental to your learning.

Teaching Format

This is an interactive, participatory course. The course will be run mainly as a combination of lectures, small group discussions, and in-class exercises. In order to have good discussions of the course materials, students must come prepared for class. This means having **read the assigned reading materials before class and coming prepared to discuss the readings and ask questions.**

Lectures

Lectures are an integral part of this course, and they consist of two parts: the material I present in class and a slide deck with many more details (which we do not have time to go over in class). The slide deck, in many cases, is almost sufficient for you to master the material. In other words, if you read and understand *everything* in the slide deck, the assigned reading will only solidify the material further, but may not be required to do well in class. If you have issues with what is written in the slide deck, then the assigned reading material is essential – you must read and go through every minute detail, or you are highly unlikely to succeed.

IV. Requirements and Evaluation Criteria

Grades

Course grades will be computed as follows:

Course Element	% Towards Final Grade
Final Project	50%
<i>Take-home project</i>	50%
Participation and responsibility grade	50%
<i>Homework Assignments (5 x Varied points)</i>	30%
<i>In-Class Labs (9-10 x Varied points)</i>	20%
Extra credit	As assigned
Total	100%

Grading Scale, Rounding, and Curves:

Your grade is the grade you *earn*. There is no curve imposed in this class.

I prefer to hear your concerns about grading during the semester. **Do not wait until the end of the semester to see me regarding problems with course materials or your performance (it will be too late to address deficiencies at the end of the semester).** If you are aware that you must achieve a particular grade in this course, please see me during the first week of the course. This will allow me to alert you of deficiencies in your performance. There is nothing that either of us can do at the end of the course.

Extra-credit: I firmly believe that education does not start and end with the required course material. Very often, what you remember the most from the course does not come from a book, but perhaps from a wise comment by a peer, a clever exercise, or an article you've read in Wall Street Journal because you related it to the course you were taking, but would have missed otherwise. I encourage students to learn things outside of class. I will sometimes point you to the facts or readings that I find interesting, and may ask you to write a page or two, or lead a discussion on such a topic. Because these topics often come up after the grading system has been announced and we are well into the semester, I will assign extra-credit points to them. A few things need to be noted:

- You *do not need* extra credit to do well in this class. Please do not feel obligated to turn in extra-credit assignments you are not interested in doing.
- Extra-credit assignments are due on their due date, and cannot be made up for ANY reason (including university-approved absences).
- DO NOT ask me for extra-credit assignments at the end of the semester if you are a few points short of your desired grade. Take care of any potential shortfalls when the opportunities are provided.
- Extra-credit opportunities will be offered to everyone. No individual extra-credit assignments will be offered or provided for any reason.

Final Project

A project is essential for this course, so it is worth 50% of your grade. By the end of the third week of the course, you need to turn in a proposal describing the intended project before spring break. The proposal is to ensure that the project is acceptable for this course and it provides an opportunity for preliminary feedback and suggestions. A final paper describing the project is due no later than a month before the project deadline.

The range of possible projects is quite broad. The intent of the project is to provide an opportunity to apply the methods for categorical data analysis covered in class to your own research and effectively communicate the results. Projects will typically consist of analyses of data from research that you are currently involved in (e.g., masters or dissertation research, collaborative research projects, etc.).

Possible projects include (but not limited to):

- Use categorical methods to analyze data from your own research or research in which you are involved.
- Critique the use of procedures often used in your field or in a published research paper(s) and present more appropriate alternative analyses. Such a project should include a comparison of results obtained from the different types of analyses (e.g., using loglinear models rather than ANOVA).

- An in depth study of a procedure covered in class or one not covered in class with the use of categorical variables (e.g., latent class analysis, log multiplicative association models, correspondence analysis, random effects models for discrete response data), including an application of it to data.

Please feel free to talk to the instructor if you have hard time coming up with the project idea. Project submission is not limited to a certain length, but should include the following:

- Project purpose
- Research design
- Analysis (with all code)
- Interpretation, conclusion, and implications for research or practice

Homeworks

In this class, homeworks are essential for learning. Simply put, you CANNOT learn statistics by simply attending the class. Homeworks will be more along the lines of the real-life problems that you will have to solve in the future, and you will have a week after the topic was introduced in class to work on these. Homework assignments are handed out in class (during seminars) and will be available electronically. I strongly recommend that you do not wait until the due date to complete those, and work on the problems a few at a time throughout the assigned period.

Due dates for all homeworks are clearly stated in the syllabus. Late homeworks are not accepted for ANY REASON. All homeworks should be submitted to me via LMS.

In-class Labs

There will be a lab assignment in almost every seminar, depending on our progress. Since we will be learning SAS, and learning quickly, you will need to devote a substantial time to it. Seminar labs should help you with this task. At the end of the lab, you will submit your completed assignment for the day (or as much as you were able to complete) to me via LMS.

Rounding and grade calculation

Your final grade is an average of a cumulative grade and a final exam grade. Grades earned as percentages will translate into point grades in 10% increments: 10% - 1 point. 20% - 2 points, etc. Grades will be assigned as whole grades only. Grade rounding follows the standard mathematical averaging rules: 34.9% is a 3; 34.99999% is a 3, and only 35.0% is a 4. I know 1/10 of a percent is not a big deal, but please do not ask me to “bump” your grade 1/10 of a percent; earn it instead.

V. Tips for Success and Other Issues of Concern

Absences and Excuses

You are responsible for attending class. If you miss class, you are still responsible for everything covered in class, including announcements. Absences excuse you (the body) NOT ANY WORK THAT IS DUE, *even if excuse is documented*. Failure to turn in assignments on time will result in a loss of participation/responsibility points, and a zero on the assignment. Similarly, being absent does not excuse you from obtaining handouts and assignments that you may have missed. It is your responsibility to find out what you have missed and to make arrangements to obtain any handouts, assignments, etc. All work is due as stated in the course schedule. **Exception: religious observance absences as stated below.**

Religious Observance: In keeping with university policy, accommodations will be made for observance of religious holidays. I require that you request accommodations in advance by notifying me in writing.

Class preparation

Considering that class preparation is a personal matter, and that there is no one formula, the following are some generally recommended guidelines for most cases:

1. Read the assigned material quickly, noting the major issues and a general sense of the layout. Read to get a sense of what the chapter is about. Ask yourself how you can relate to the materials covered, and whether all the new terms make sense.
2. Reread the material carefully, annotating, highlighting and distinguishing important information, omissions, and questions raised by the reading.
3. Decide what the most important issues are. Write down questions you don't understand.
4. Discuss the chapter with others, before class if possible, to test out your ideas and further your understanding of the issues.
5. Prepare notes to guide your class participation, including: answering assigned questions, summary of the main issue(s), further questions raised by the reading, assumptions made by the chapter, your personal experiences, and possible approaches or solutions to any problems assigned with the chapter.
6. DO YOUR HOMEWORK and ATTEND SEMINARS TO COMPLETE LABS.

Virtual Office Hours – LMS Forum

This is a very large class, and I may not be able to answer all questions that are e-mailed to me on a timely basis. Moreover, I've found over the years that usually 90% of all questions from students are nearly identical – that is, if you do not understand something, chances are, most other people do not, either. Instead of repeating answers over and over via email, I will be answering all content-related questions in LMS forum (it is already open for the class, please feel free to start threads and ask questions).

Feel free to answer questions posted on the Forum. Not only that will provide an answer to your peers quicker (I may not be able to respond to Forum questions immediately), but I will also monitor your Forum activities and will provide **extra-credit** to the most active students who provide best answers.

Messages and Memos for Me

If you have any messages or specific requests for me, please submit them by e-mail or in typed format. Ensure that your message includes your name, a complete description of your concern, and a recommendation for resolution. Please note that I have a very busy inbox; usually, I answer emails only once a day, very early in the morning. So if you send me a question around 10 am, please do not expect to receive answer until the following day.

Stay Informed about Class Schedules & Policies

It is the student's responsibility to stay informed about class schedules and policies. The information you need is included on both the paper copy of the syllabus, and the online website pages. In addition, announcements will be made regularly in class and on website, and it is your responsibility to keep up with that information. If you are unclear about any policies or other information, please ask promptly. Don't wait and get an unpleasant surprise later.

Participation Ground Rules

In an effort to provide a classroom environment as conducive to learning as possible, the following ground rules should be observed:

1. *Confidentiality.* Concepts and ideas can be taken from the class and discussed freely. However, personal stories or issues raised by individuals are to be kept confidential and as the property of the class.

2. *Respectful Listening*. When differing with another participant's point of view, listen first before raising questions. When another participant raises a point we disagree with or find offensive, it is important to remember that the human being behind that question or comment deserves respect. Please freely utilize the concepts we'll learn in the second week of class.
3. *Participation*. Participants who tend to be quieter are encouraged to contribute to enhancing the learning process by sharing their perspectives and experiences. Those who are aware they are prone to monopolizing discussions are encouraged to self-monitor their behavior and make room for quieter students.
4. *No Zaps*. In keeping with the notion of respectful listening, "putting-down" others in class is discouraged. "Zapping" another person often serves to discourage open and honest exchange of ideas among the whole group.

Academic Honesty

Scholastic dishonesty, including cheating in exams or plagiarism, will be treated as a violation of university's regulations. As commonly defined, plagiarism consists of passing off as one's own the ideas, words, writings, etc., which belong to another. In accordance with this definition, you are committing plagiarism if you copy the work of another person and turn it in as your own, even if you should have the permission of that person. Plagiarism is one of the most severe forms of academic dishonesty and can be expected to result in appropriate consequences.

You must complete exams and assignments alone, except group exercises. You may use only the materials I specify to help you complete your work. If, due to a disability or extraordinary circumstances, you need special accommodations or help in completing course requirements, you must see the instructor **BEFORE** the exam or assignment due date.

Personally, I have no tolerance for cheating, regardless of the reason. **Simply put, don't do it.**

Copyright Notice

All handouts in this course are copyrighted, including all materials posted on the website for this course. "Handouts" refers to all materials generated for this class, which include but are not limited to the syllabus, class notes, quizzes, exams, lab problems, in-class materials, review sheets, and additional problem sets. You have the right to download materials from the course website for your own use during this class; however, because these materials are copyrighted, you do not have the right to copy the handouts for other purposes unless the instructor expressly grants permission.

Good luck and have fun!

VI. Course Content

The topics to be covered in course, with assigned readings and corresponding assignments, are listed below. They are subject to change with notice, and topics may not always be given in the order listed; please see the class schedule provided in the separate document. Each lecture will be accompanied by a corresponding lab (provided separately).

1. **Introduction to Categorical DA**
 - a. Categorical response data
 - b. Probability distributions for categorical data
 - c. Statistical inference for discrete data
2. **Contingency tables**
 - a. Probability structure for contingency tables
 - b. Comparing proportions with 2x2 tables

- c. The odds ratio
 - d. Tests for independence
 - e. Exact inference
 - f. Extension to three-way and larger tables
- 3. Generalized linear models**
- a. Components of a generalized linear model
 - b. GLM for binary and count data
 - c. Statistical inference and model checking
 - d. Fitting GLMs
- 4. Logistic regression**
- a. Interpreting the logistic regression model
 - b. Inference for logistic regression
 - c. Logistic regression with categorical predictors
 - d. Multiple logistic regression
 - e. Summarizing effects
 - f. Building and applying logistic regression models
 - g. Multicategory logit models
- 5. Loglinear models for contingency tables**
- a. Loglinear models for two-way and three-way tables
 - b. Inference for Loglinear models
 - c. The loglinear-logistic connection
 - d. Independence graphs and collapsibility
- 6. Models for matched pairs**
- a. Comparing dependent proportions
 - b. Logistic regression for matched pairs
 - c. Comparing margins of square contingency tables
 - d. Symmetry issues
- 7. Random effects: GL Mixed Models**
- a. Random effects modeling of clustered categorical data
 - b. Extensions to multinomial responses or multiple random effect terms
 - c. Hierarchical models
 - d. Final notes on fitting and inference