<div align="center">**Course syllabus «Categorical Data Analysis»**</div>

<div align="right">
Approved by
Programme Academic Council
Protocol Nr. 10 from 27.08.2019
</div>

Author	Dr. Tamas Rudas
Number of credits	4
Contact hours	48
Self-study hours	104
Course	1,2
Educational format	Without use of online course

I. Goals and Results of Mastering the Discipline; Prerequisites

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.

As a results, students should:

Know:
- the most fundamental regression models for binary, ordinal, nominal and count outcomes.
- the approaches to building the binary logit and probit models
- the foundation of multinomial logit models

Be able to:
- interpret the results of models with non-linear outcomes.
- work with major linear modeling programs, especially SAS, so that they can use them and interpret their output.
- develop and/or foster critical reviewing skills of published empirical research using applied statistical methods.
- to criticize constructively and determine existing issues with applied linear models in published work

Have:
- an understanding of the basic principles of binary models and lay the foundation for future learning in the area.
- the skill to meaningfully develop an appropriate model for the research question
- the skill to work with statistical software, required to analyze the data

Basic knowledge of introductory statistics are required for this course.

The basics of this discipline should be used in the following courses and activities:
•Generalized linear modeling

•all other program related courses

The course is strongly related and complementary to other compulsory courses provided in the first year (e.g. Contemporary Data Analysis, Research Seminar) and sets a crucial prerequisite for later courses and research projects as well as for the master thesis. The course gives students an important foundation to develop and conduct their own research as well as to evaluate research of others.

II. Content of the Course

Please note: the course is broken into topic sessions, some of which will take more than one scheduled class meeting. Actual progression of the course will be updated via class schedule, provided separately.

SESSION ONE: Introduction to Categorical data analysis

The first session will introduce the main concepts of behind categorical response data, probability distributions for categorical data, and statistical inference for discrete data.

SESSION TWO: Contingency tables

The session will discuss Probability structure for contingency tables, comparing proportions with 2x2 tables, the odds ratio, tests for independence, exact inference, and extensions to three-way and larger tables.

SESSION THREE: Generalized linear models

The session provides the theoretical basis and the derivation of the components of a generalized linear model, GLM for binary and count data, statistical inference and model checking, and fitting GLMs.

SESSION FOUR: Logistic regression

This sessions provides the framework for interpreting the logistic regression model, inference for logistic regression, logistic regression with categorical predictors, multiple logistic regression, summarizing effects, building and applying logistic regression models, and multicategory logit models.

SESSION FIVE: Loglinear models for contingency tables

This session covers loglinear models for two-way and three-way tables, inference for Loglinear models, the loglinear-logistic connection, and independence graphs and collapsibility.

SESSION SIX: Models for matched pairs

This session will focus on comparing dependent proportions, logistic regression for matched pairs, comparing margins of square contingency tables, and symmetry issues.

SESSION SEVEN: Random effects – GL Mixed Models

This session will look at random effects modeling of clustered categorical data, extensions to multinomial responses or multiple random effect terms, hierarchical models, and final notes on fitting and inference.

III.   Grading

Course grade will be completed as follows:

Course Element	% Towards Final Grade
Final Project	**50%**
Take-home project	*50%*
Participation and responsibility grade	**50%**
Homework Assignments (5 x Varied points)	*30%*
In-Class Labs (9-10 x Varied points)	*20%*
Total	100%

If the final grade is non-integer, it is rounded according to algebraic rules. If has a half (.5) at the end, we are rounding upward. Rounding of cumulative grades and other rounding issues are performed according to the HSE rules.

IV.   Grading Tools

This class contains several assignments that test student knowledge and understanding throughout the course.

<u>In-class Labs</u>
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.

<u>Homework assignments</u>
There will be several homework assignments that will provide additional hands-on practice for the concepts we've learned in class and practiced during the seminar. Homeworks will be assigned as needed throughout the semester. All homework submissions must be done by the stated deadline via the LMS system.

<u>Final project</u>
Final project is a take-home project in which students are asked to analyze an actual dataset with one or more categorical criteria. To do so, they must first define the research question, build and appropriate model, interpret the outcome, and make appropriate conclusions. This project should mimic an actual research paper and be no shorter than 15 pages in length, including references and software output.

V.   Resources

5.1 Main Literature

1. Agresti, Alan. Categorical Data Analysis, John Wiley & Sons, Incorporated, 2012. ProQuest Ebook Central, https://ebookcentral.proquest.com/lib/hselibrary-ebooks/detail.action?docID=1168529
2. Rudas, Tamás. *Lectures on Categorical Data Analysis*. Springer, 2018. URL https://proxylibrary.hse.ru:2176/book/10.1007/978-1-4939-7693-5. Springer Link.
3. Upton, Graham J. G.. Categorical Data Analysis by Example, John Wiley & Sons, Incorporated, 2016. ProQuest Ebook Central, https://ebookcentral.proquest.com/lib/hselibrary-ebooks/detail.action?docID=4722461.
4. New Developments in Categorical Data Analysis for the Social and Behavioral Sciences, edited by der Ark, L. Andries van, et al., Taylor & Francis Group, 2004. ProQuest Ebook Central, https://ebookcentral.proquest.com/lib/hselibrary-ebooks/detail.action?docID=227471.
5. Sutradhar, Brajendra C.. Longitudinal Categorical Data Analysis, Springer, 2014. ProQuest Ebook Central, https://ebookcentral.proquest.com/lib/hselibrary-ebooks/detail.action?docID=1966818.

5.2 Additional Literature

1. Fienberg, Stephen E. *The analysis of cross-classified categorical data*. Springer Science & Business Media, 2007. URL https://proxylibrary.hse.ru:2176/book/10.1007/978-0-387-72825-4. Springer Link.
2. Andersen, Erling B. *Introduction to the statistical analysis of categorical data*. Springer Science & Business Media, 2012. URL https://proxylibrary.hse.ru:2176/book/10.1007/978-3-642-59123-5. Springer Link.
3. Andersen, Erling B. *The statistical analysis of categorical data*. Springer Science & Business Media, 2012. URL https://proxylibrary.hse.ru:2176/book/10.1007/978-3-642-78817-8. Springer Link.

5.3 Software

№ п/п	Name	Access conditions
1.	MicrosoftWindows 7 Professional RUS MicrosoftWindows 10 MicrosoftWindows 8.1 Professional RUS	*From the university's internal network (contract)*
2.	Microsoft Office Professional Plus 2010	*From the university's internal network (contract)*
3.	R, R studio	*Open access. URL: https://www.r-project.org/*
4.	SAS	*Open access University Edition. URL: https://www.sas.com/ru_ru/software/university-edition.html*

5.3 Material and technical support

Classrooms for lectures on the discipline provide for the use and demonstration of thematic illustrations corresponding to the program of the discipline, consisting of:
– PC with Internet access (operating system, office software, antivirus software);
– multimedia projector with remote control.