Final Project Assignment

General Specification

The goal is to analyze **continuation rates** of students that have graduated from the bachelor's programs at the Faculty of Mathematics and Physics [MFF UK]. Most of the graduates continue in one of the master's programs offered by the Faculty, some of them go elsewhere.

The continuation is defined as a student's enrollment in any master's program of the same field of education (Physics, Computer Science, Mathematics, Teacher's Education) *immediately* after the completion of their bachelor's studies.

The data include information on all students who enrolled in any bachelor's program at MFF UK since 2005 and successfully graduated.

The Tasks

1. Analyze the development of continuation rates over time. Especially, determine whether there was an overall positive or negative trend in continuation rates across MFF UK.
2. Determine whether the continuation rates varied between the fields of education (Physics, Computer Science, Mathematics, Teacher's Education) and whether the time trends were the same in all fields.
3. Find out whether continuation rates depend on student's previous performance in bachelor's study, in particular whether weaker students were less likely to continue in the master's programs of the same field.

Datasets

The dataset `pmse2025.RData` is available in the R data format and can be downloaded from the SIS page of the course NMST412. It is necessary to log in and to be registered for NMST412, otherwise the data set is not available. The dataset includes the single dataframe `pmse`.

The variables are listed and their meaning is explained in the variable coding table at the end of the assignment.

Requirements

Create an electronic report in the pdf format summarizing the solution to the problem, the results and conclusions. The ideal length of the report is under 10 pages. You can write it in Czech, Slovak or English. **The requirements on the report** are as follows:

1. Describe the manipulation with the data set that preceded the analysis (variable transformations and recoding, treatment of missing values, omitted observations).
2. Include a concise descriptive analysis (descriptive tables, figures) targeted towards answering the questions of interest.

3. Describe the models used to address the objectives and explain how they were developed. Display the model formulas in the report (in mathematical notation, not in R code). Justify the appropriateness and validity of the models (can there be a better model than this? do all the assumptions hold? does the model make sense?). Provide p-values for testing the relevance of each term included in the models.

4. Interpret all the parameters of the models and provide their confidence intervals. Explain which of them will be used to address the questions of interest.

5. Provide *an explicit answer to each question of interest*. The answers must be formulated in lay language and evaluate the result both qualitatively (yes/no) and quantitatively (how much + uncertainty in this).

6. Include a short discussion of your approach to the analysis (its strengths and weaknesses) and of the meaning and reliability of the results.

7. Do not use AI tools for data analysis tasks of for writing parts of the project report. You can use AI tools, if you wish, to perform tedious tasks or to improve the language level of your report. If you do so, describe in the report where/how AI tools were used.

You are not asked to discover "the true" model as there is none. It suffices to come up with a reasonable model that makes sense and performs the task as long as you provide valid and relevant arguments to justify its choice.

Mail the report together with another file containing the **R code** used to perform data manipulations and to obtain the results included in the report.

Due date: Three working days before the scheduled project evaluation date.

Variable Coding Table

The dataframe `pmse` includes 13 variables and 3916 observations.

Variable	Variable Label	Variable Coding
id	Observation order	integer
field.bc	Field of education in bachelor's study	factor: 'Physics'; 'Computer Science'; 'Mathematics'; 'Teacher's Education'
bc.start.year	Academic year of start of bachelor's study*	numeric
bc.end.year	Academic year of bachelor's study graduation*	numeric
bc.dur	Bachelor's study duration in years	numeric
bc.cred	Total number of credits gained in bachelor's study	numeric
def.gr	Thesis defense grade	numeric: '1' = excellent; '2' = very good; 3 = 'good'
def.att	# attempts to pass the defense	numeric
oral1.gr	Final oral exam grade (part 1)	numeric: '1' = excellent; '2' = very good; 3 = 'good'
oral1.att	# attempts to pass part 1 of the final oral exam	numeric
oral2.gr	Final oral exam grade (part 2)†	numeric: '1' = excellent; '2' = very good; 3 = 'good'
oral2.att	# attempts to pass part 2 of the final oral exam†	numeric
field.mgr	Field of education in master's study	factor: 'Physics'; 'Computer Science'; 'Mathematics'; 'Teacher's Education'; missing = no enrollment

* The value N means the academic year N/(N+1).

† Some study programs had two parts of final oral exam. If there was only one part, this variable is missing.