



Exercise 7

Applied Longitudinal Data Analysis

Topic A: Political interest across the life course

Download the data-file POL.dta. The data is a data extract from the SOEP. The data is limited to respondents who participated in the survey at ages 18-25 in the years 2000-2020. It is organized in wide format. The outcome variable is political interest (POL: very strong, strong, not so much, not at all). Conduct a sequence analysis (see below).

Topic B: Trajectories in and out of AFD

Download the data-file AFD.dta. The data is a data extract from the SOEP. The data is limited to respondents who participated in the survey 2015-2020 and reported at least once that they leaned towards AFD. The outcome variable is party preference (PARTY: AFD, CDU, SPD, other, no preference). It is organized in wide format. Conduct a sequence analysis (see below).

Topic C: Maternal employment trajectories after first childbirth

Download the data-file KIDAGE.dta. The data is a data extract from the SOEP. The data is limited to female respondents who had at least one child. It includes the six years after first childbirth. It is organized in wide format. The outcome variable is employment status (EMP: In education, unemployment, employment, not in labour market). Conduct a sequence analysis (see below).

Topic D: Men's retirement trajectories

Download the data-file RETIRE.dta. The data is a data extract from the SOEP. The data is limited to male respondents who participated in the survey at ages 55-64 in the years 2000-2020. It is organized in wide format. The outcome variable is employment status (EMP: Unemployed, employed, not in labour market). Conduct a sequence analysis (see below).

Topic E: Self-employment trajectories

Download the data-file SELF.dta. The data is a data extract from the SOEP. The data is limited to respondents who participated in the survey 2016-20 and were self-employed at least once during that time. The outcome variable is employment status (SELF: Self-employed, employed full-time, employed part-time, in education, not in labour market/unemployed). Conduct a sequence analysis (see below).

Topic F: Family Trajectories after First Motherhood

Download the data-file FAM.dta. The data is a data extract from the SOEP. The data is limited to respondents who had a first child in the years 2016-20. The outcome variable is family status (Single, cohabiting, married). Conduct a sequence analysis (see below).

Procedure

- Preparation: Plot the sequences for the entire sample to get an idea of overall patterns.
- Preparation: Categorize your outcome variable in a meaningful manner.
- Analysis: Find the number of optimal clusters.
- Analysis: Plot the sequence index plots by clusters and select a suitable name for each cluster.
- Analysis: Describe cluster membership by suitable covariates (such as gender, birth cohort (e.g., born 1970 before or after 1970), region (East/West Germany), education (low, medium, high)). Plot the results.
- Analysis (Optional): Estimate a multinomial model that predicts cluster membership. Visualize the predicted values in a forest plot.
- Presentation: Prepare 3-5 slides that summarize your results and present findings to class (max. 5-10 minutes).