Exercise 4

Applied Longitudinal Data Analysis

Deadline: Upload your assignment by Monday (March 2nd), **5 p.m.** Upload **one** file only (.pdf). Include the R-code into the Appendix. Cite data and readings (and preferable R-packages). Label and number all figures and tables.

Exercise 4.1 (done in class)

a) Prepare the data as described in the Appendix. Provide the sample statistics by migration background. Use person-months as the unit of analysis. Describe the table in a concise manner.
b) Estimate a survival function by migration background. Describe patterns (very briefly).

Exercise 4.2 (done in class)

a) Formulate a testable hypothesis that deals with the impact of migration background on unemployment duration.
b) Estimate a Cox-model. Describe the model results in a concise manner. Do the analysis confirm your hypothesis?

MODEL01: SEX+MIG+AGE

Exercise 4.3 (done in class)

Estimate the following models. Present the hazard ratios from MODEL01-MODEL03 in one table (i.e., use the package "gtsummary").

a) Why did the hazard ratio for MIG change across models? Explain.
b) Did the inclusion of EDU in Model02 improve the model fit? Did the inclusion of health improve it as well?

MODEL02: SEX+MIG+AGE+EDU
MODEL03: SEX+MIG+AGE+EDU+HEALTH

Exercise 4.4

a) Formulate a testable hypothesis that deals with the impact of migration background and gender (in interaction) on unemployment duration. The readings on moodle may be helpful to buttress your argumentation. [around 150 words]
b) Estimate a Cox-model with migration in interaction with gender as the main covariates. Present the model results in a forest plot (i.e., use the command "ggforest"). Describe and interpret the results from the interaction. Do the results support your hypothesis?

MODEL04: AGE+EDU+ SEX:MIG

Exercise 4.5

a) Formulate a testable hypothesis on the gendered effect of family status on unemployment duration. The paper by Jacob and Kleinert (2014) may provide some ideas that may be helpful to generate a hypothesis (note however that our data does not include any partner characteristics).
b) Discuss and interpret the following model results against your hypothesis.

Model05: EDU+MIG+AGE+ SEX:FAM

Exercise 4.6

You have found a very strong effect of migration status (MIG) on unemployment duration in MODEL03. To what extent do you think is this effect causal? For example, which omitted variables could have biased your results? Why and how? [around 200 words]

Appendix

Preparation A: Download the following SOEP-data extracts (teaching version) from moodle:

- CALEN: Monthly employment biography
- FIX: File with time-constant covariates
- VARYING: File with time-varying covariates

Preparation B:

- Merge FIX and VARYING by the identifier (ID).
- Merge CALEN to that data by the identifiers (ID, SYEAR).

Preparation C:

- Use HEALTH01 and categorize this variable in a meaningful manner. Transfer this variable into a factor variable (HEALTH).
- Proceed by the same logic with the other variables (FAM, EDU, AGE, MIG).

CALEN

Variable	Variable labels	Value labels
BEGIN	Begin of spell	
END	End of spell	
EVENT	Censoring variable	0=censored 1=employment 2=education
SYEAR	Calendar year	

FIX

Variable	Variable labels	Value labels	New variable
SEX01	Sex of respondent	1=male 2=female	SEX
MIGBACK	Migration background	1=no migration background 2= migration background	MIG

VARYING

Variable	Variable labels	Value labels	New variable
SYEAR	Calendar year		
PGISCED97	Education (ISCED-1997)	0= in school 1= inadequately 2= general elementary 3= middle vocational 4= vocational + high school degree 5= higher vocational 6= higher education	EDU
PGFAMSTD	Marital status	1=married 2=married, but separated 3=single 4=divorced 5=widowed 6=partner in foreign country 7=registered partnership 8=registered partnership, but separated	FAM
HEALTH01	Health status	1=Big worries 2=Some worries 3=No worries	HEALTH
AGE01	Age continuous		AGE