

Prof. Dr. Stefanie Kley

Email: [stefanie.kley@uni-hamburg.de](mailto:stefanie.kley@uni-hamburg.de)

Allende Platz 1, room no. 344

Consulting hour: thursdays from 2 to 3 p.m., a.b.a.

**Paneldata Analysis**

**Assignment 3**

# Summer term 2023, event no. 24-504.31 M.A. Sociology, profile module empirical social research, 6 cp

Time and place: thursdays 10 to 12 a.m., WiWi 2043/2047

1. Goal is to describe the data, to estimate different models und to compare the estimates. Open **hd-data** and have a quick look at the data. Make sure that the data is sorted by person id and survey year.

Generate a new variable *num* as a running number within each person (use *\_n*). Generate a second new variable N that summarizes the rows within each person (use \_N). Move these new variables at top of the dataset (use *move* or *order*) and control the result und the data browser.

Label and summarize the new variables; the range of num is 1-14697, the range of N is 1-33.

2. It is not desirable to have single spells in our dataset. Drop all persons with only one observation. The number of cases deleted is 28.

Now the range of num is 1-14697, the range of N is 2-33.

Sort the data again by person id and survey year.

3. We want to analyse the transition to homeownership. Therefore, all persons in the dataset must not be homeowners at the beginning of the observation. Control this. The number of homeowners at the beginning of observation is 593.

Generate a new variable *flag*=1 for those who are homeowners in the first spell. Control the result; *flag* is 1 in \_\_\_\_\_\_\_\_\_\_\_ cases.

Within persons, replace all following spells with 1 if the first spell is 1. Control the result also in the browser (use *order flag num*). *flag* is 1 in \_\_\_\_\_\_\_\_\_\_\_ cases.

Drop all cases with *flag*==1. Sort the data again by person id and survey year.

The number of persons is now \_\_\_\_\_\_\_\_\_\_\_\_. The number of all spells is \_\_\_\_\_\_\_\_\_\_\_.

4. Describe the data based on the first or last observation for each person. For the ease of analyses, do not use weights.

How many persons are homeowners at the start of observation? \_\_\_\_\_\_\_\_\_\_\_\_\_\_

How many persons are homeowners at the end of observation? \_\_\_\_\_\_\_\_\_\_\_\_\_\_

5. We want to analyse at which age people become homeowners for the first time, based on the age of the household head (*agehh*). First, have a look at the age variable in the data browser (use *order*). Is it time variable (yes/no) \_\_\_\_\_\_\_? The age range in the data is \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_.

Declare the data as survival data with stset agehh, id(pid) failure(owner).

The number of subjects is \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_, the number of failures is \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_.

Why is the number of failures a bit different to the number of homeowners at the end of observation? \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

6. Type *stdescribe*. The mean number of records is \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_, the mean time at risk is \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_, the percentage of entries into homeownership is \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_.

7. Type *sts list, failure by(class) compare at(20 25 30 35 40).* Estimates of the Kaplan-Meier functions for the transition into first homeownership by age 40 are:

For skilled (blue-collar) workers: \_\_\_\_\_\_\_\_\_\_\_, and for white collar workers: \_\_\_\_\_\_\_\_\_\_\_.

8. Estimate a simple logit model, and estimate the same model with robust standard errors, by typing the option after the comma:

*logit owner agehh agesq sex labgro i.partner kidnr powner*

*logit owner agehh agesq sex labgro i.partner kidnr powner, vce(cluster pid)*

What is the difference? \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

9. Declare the data to be paneldata with *xtset pid syear*. Estimate the same model with *xtlogit* and the fixed-effects option. Type *help xtlogit* to learn how to set up this model.

For the following variables effects cannot be estimated \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_.

Why is this the case? \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

The number of cases in this model drops dramatically. Why is this the case?

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_