



## Assignment No. 4

**Out: 10 Jan 2019**

**Due by: 25 Jan 2019**

Solutions to the assignment should be sent to `gampe@demogr.mpg.de` with appropriate attachments. These are

- (1) a pdf-file containing your derivations/results/figures/interpretations and
- (2) a file with the R-program that you used for those problems where you have to use R.

If you add some comment lines in your programs, explaining what the lines of code are doing, this would be extremely helpful.

### Problem 1:

In residential mobility studies the immune-fraction model is often called mover-stayer model, expressing the idea that there are two groups of people with different attitudes toward moving (or not). Both the propensity to move at all as well as the length of stay (in case you move) will depend on individual (and possibly also neighborhood) characteristics.

The file `stay.txt` contains data on  $n = 720$  individuals about how long they stayed in their house or apartment (in years). The information in the dataset is:

**stay** length of stay (in years) until moving out or end of follow-up

**exit** 1=moved out, 0=no move until end of follow-up

**owner** home ownership: 1=yes, 0=no

**educ** education; three levels: L=low ( $\leq 8$  yrs) , M=medium (8-12 yrs) ,  
H=high ( $> 12$  yrs)

**age\_20** age-20 (in years) when moving in (e.g., `age_20=21` is `age=41`)

First assume that everybody eventually will move, i.e., there is no immune fraction. Run a Cox PH and interpret the results.

Then assume that there is an 'immune' fraction of people that will not move at all, and a susceptible group of individuals that will move sooner or later. Assume a proportional hazards model for the hazard of moving out and fit a immune-fraction model to these data (using the R-package `smcure`). Answer the following questions:

1. Which variables influence whether people move or stay? In which direction do these covariates act? Who is least likely of moving?
2. If people are in the group of 'movers', which characteristics influence their hazard of moving out?

If you compare with the original Cox model, are there differences in your conclusions if you allow for an immune fraction of stayers?

**Problem 2:**

The duration of doctoral studies depends on many things. Besides talent, effort, supervision, . . . , the following covariates are expected to influence the speed of progress toward completion of a doctorate:

**sex** M=male, F=female

**residence** P=permanent, T=temporary (only for purpose of education)

**field** nat=natural sciences, soc=social sciences and humanities, eng=engineering

**topic** N=new (unrelated to previous studies), S=same or similar to previous studies

The file `doctorate.txt` contains the duration (in months) for 372 successful PhD candidates. The duration was determined as time between fixing of the topic and successful defense.

1. Fit an accelerated failure time model to these data by using the R-package `smoothSurv`. Which of the covariates do have an influence on the duration of doctoral studies and how do they act?
2. Do the results support a LogNormal distribution for the duration of doctoral studies?
3. Which group of students experience, according to the regression results, on average the shortest and the longest duration?