Course work II:
Analysis of old-age mortality data

EHA September 1 - 25, 2015

1. The data set oldmort in *e*ha package contains contains life histories for people aged 60 and above in the years 1860?1880, that is, 21 years. The data come from the Demographic Data Base at UmeåUniversity, Umeå, Sweden, and cover the sawmill district of Sundsvall in the middle of Sweden. This was one of the largest sawmill districts in Europe in the late 19th century. The town Sundsvall is located in the district, which also contains a rural area, where farming was the main occupation.

2. Give the summary of all the variables.

3. Choice of time origin. Plot the Kaplan-Meier curves when the time origin is set at (a) the age at entry; (b) birth. Interpret the two curves. Analyses men and women separately.

4. What type(s) incompleteness is observed in this dataset? Which of the above curve takes into account the observed incompleteness?

5. Hypothesis 1: women live longer than men. Test this hypothesis using the logrank test. You may use *c*oxph function from *s*urvival package. How much lower is the mortality among women compared to men? Is this finding valid from birth or some other higher age?

6. Choosing the best parametric distribution: Fit paramteric regression model to oldmort data using the covariates sex, civ, and birthplace. Choose the best parametric distribution that fits the baseline hazard and also, graphically examine the same. Test whether birthplace has significant effect on mortality. The paramteric distributions to try in the regression models are: i) Weibull, ii) lognormal, iii) loglogistic, iv) Gompertz, and v) extreme value. Note: Because the minimum entry

age is 60 years, the parametric distribution is truncated (see slide on Left truncation and late entry, lecture notes on Survival likelihood and parameteric survival distributions). It is necessary to calculate new entry and exit age by subtracting 60 from the original entry and exit age before fitting the paramteric model. This does not matter for the Cox model though since the risk sets are calculated correctly with left truncation.

{Hint: You can perform the above analyses using phreg function from eha package. Please see the help for it first. See the model under which the likelihood is the maximum. To plot graphically the best model, you can fit a regression model with coxph using the specified covariates and then using the function check.dist with the output of the coxph and the best parameteric model.}

References

Broström G. *Event History Analysis with R.* CRC Press, Taylor & Francis Group (2012).