Exercise for survival analysis

Alessio Crippa

February 28, 2018

Survival analysis, Exercises

Consider now the Whitehall study, a large prospective cohort of 17,260 male British Civil Servants. *Lancet*, Volume 323, Issue 8384, 1984, Pages 1003–1006. During 10 years follow-up (165,612 person-years) we observed 1,670 deaths.

Questions

Data inspection

1. Read the **wh** data available at http://alecri.github.io/downloads/data/whitehall.csv

2. Get familiar with the data. How many observations and variables (which type) are in the dataset?

3. `all10` and `pyall10` are the variables indicating if a person died (`all10 = 1`) and the corresponding follow-up time. Describe the two variables. What is the mortality rate (x 10000)?

4. Create a survival object. Display the first 10 observation?

5. Estimate the survival function using the Kaplan–Meier method. Why there is no information about the survival time?

6. Estimate the 1th and 5th percentiles of survival times and interpret the results.

7. Assume an exponential distribution for time. Estimate the survival curve using the corresponding parametric model.

8. Consider the possible health inequalities among british civil servants depending on the `jobgrade`. What is the mortality rate in the different jobgrade categories?

9. Estimate the survival curves and test for possible differences.

10. Specify a Cox regression model to investigate the association between jobgrade and (log) rates of death, adjusted for age. Interpret the results.

11. Assuming the effect of age on the (log) rates of death can be approximated by a quadratic curve. Estimate and present the results from the corresponding Cox model.

12. Run a similar analysis as in 10. assuming an exponential distribution for the survival time. Interpret the results

13. Compare the predicted survival curves based on the estimated models in 10. and 12. for a 50 years-old man with Clerical as jobgrade.