Home assignment 1
Analysis of Survival Data, fall 2025

Criteria to get your home assignment approved:

1. Your report must be complete and handed in on time. See the document 'Home assignment information' for further information.
2. Follow the check list (separate document) to ensure that you follow all requirements to pass.
3. All calculations and tables are to be produced by SAS. Graphs can be produced by SAS or, if you wish, by using other software (as long as the data for the graphs is produced by SAS).
4. All assumptions made and all assumptions used must be clearly presented and checked if they hold (or discussed, if not possible to check). If they don't hold, discuss what implications this might have.
5. When there is more than one method/estimator/test to choose from, motivate your choice.
6. When presenting confidence intervals, state what the interval is intended to cover, and interpret the interval.
7. When performing hypothesis testing you must state what the test is intended to test. Always state hypotheses, significance level (motivate your choice), choice of test, *P*-value, decision (reject/not reject H_0), and conclusion.
8. If any of the requested analyses are inappropriate for your specific data, find a solution (don't perform an inappropriate analysis).
9. Include your commented SAS code in an appendix (Appendix 1).
10. Include a short statement on any use of generative AI (e.g. ChatGPT) in a second appendix (Appendix 2).

Information:

Your report (only one per group) needs to be electronically registered (pdf version) at Studium before the deadline.

All documents you need are to be found at Studium, modules "Home assignments" and "SAS examples".

This home assignment is based on the contents from lectures L1-L4.

If you pass this home assignment on the first try you get 2 bonus points to be added to the points you score on the final exam.

Data to be used:

Find your own data! Public data sets are available from hundreds of different websites, on virtually any topic.

See more information in the document/assignment *Data description*.

Task 1

Find (or create) some groups in your data that you are interested in comparing (at least three groups).

Plot the estimated survival probabilities for the time to event variable, for each of the groups (in one graph), and test if there is any difference in the risk of experiencing the event for the different groups. Choose only one test for this comparison. Compare the groups using PROC LIFETEST and the option STRATA. Include confidence bands in the plot (only one plot). Interpret your results.

Also present and interpret the median time to event for the different groups with confidence intervals in a table.

Task 2

Plot the cumulative hazard for the different groups (in one graph). Interpret the plot.

Task 3

Some researchers like to work with log(time) as the dependent variable to reduce the influence of large observation values.

Calculate a new variable, log(time), and test if there is any difference in time to event between the groups based on this new variable. This time you don't need to motivate choice of test or significance level, nor check any assumptions.

Compare to the results in Task 1 and comment upon the results.

Task 4

Find/create another grouping variable, dividing the observations into at least two groups.

Repeat task 1, but now stratified on this second grouping variable. I.e., compare the time to event for the same groups you used in task 1, adjusting for the second variable by stratifying the analysis on the second grouping variable. Present appropriate survival curves (without confidence bands) and hypothesis tests. Interpret your results.

Good luck!
/Inger