Survival Analysis: Exercises (Part 2)

This is the second part of a series on conducting Survival Analysis in R using Survival and Survminer. It is advised to first complete the first set of exercises (here) before attempting these, as there is a direct continuation of knowledge.

The second part of this series focuses on more complex and insightful methods through the semi-parametric Cox Proportional Hazards model. Through a Cox Proportional Hazards model, it is possible to model covariates in a semi-parametric fashion. The advantage of this modeling strategy is that it makes modeling the survival times possible, without knowing or specifying the underlying distribution.

Solutions to these exercises can be found here. It should be noted that these solutions are quite verbose with the intention of breaking down each stage in the modeling process. In real life, you will probably want to use lapply and functions to speed up the variable selection stage.

Exercise 1
Before any modeling can commence, let us just test a few variables to get a feel for their effects on survival times. Create survival objects for sex, ph.karno, and wt.loss. Hint: You’ll need to group wt.loss.

Exercise 2
Plot these using Survminer to look for differences in the group’s survival curves.
Exercise 3
To check that these three variables adhere to the proportional hazards assumption, we must plot a log-cumulative hazard plot for each of them. Within each plot, we are looking for “roughly” parallel lines between each group. If this criterion is met, then the variables adhere to the PH assumption and can then be modeled through a Cox PH model.

Exercise 4
With no huge differences in parallelism, we can begin model building. First, just time and status. This will be known as the null model.

Exercise 5
Calculate the -2*log-likelihood for the null model. This will be our baseline comparison value when fitting terms.

Exercise 6
Now fit a model with sex, ph.karno, and wt.loss individually and obtain -2*log-likelihood values for these models.

Exercise 7
Conduct a Chi-Squared test now on the change in -2*log-likelihood values for each of the single term models, ensuring you have the correct degrees of freedom specified.

Learn more about Model Evaluation in the online course Regression Machine Learning with R. In this course you will learn how to:

- Avoid model over-fitting using cross-validation for optimal parameter selection
- Explore maximum margin methods, such as the best penalty of error term support vector machines with linear and non-linear kernels
- And much more
**Exercise 8**
Given that every term is significant indicates that they are needed in the model, as they introduce a significant amount of information. The next step is to fit them in the presence of each other. To do this, fit a saturated model and remove each term to test for a significant change in \(-2\times\log\text{-likelihood}\) values.

**Exercise 9**
With every term being significant, one last thing to check is for any interactions. Feel free to try a few and test for a significant change in deviance compared to the saturated model.

**Exercise 10**
Write out your final Cox Proportional Hazards model.