PSTAT 175: Lab D: Proportional Hazards Model fitting and checking

You can use a level $\alpha = 0.05$ for all of the hypothesis tests.

1. The data set that we used in lecture `hern.txt` contains information about treatment participation for heroin addicts in two clinics. In this problem, I want you to use likelihood ratio tests.

   (a) One of the covariates in the data, `Prison`, is a indicator of whether the subject had served prison time. Fit a Cox PH model to test whether prison has a significant effect on the time spent in the clinic.

   (b) The column `Clinic` tells us which clinic the subject was at. Repeat your test regarding the effect of prison time, but control for the possible confounding effect of the clinic.

   (c) Use a complementary log-log plot to visualize whether the Cox PH assumption is appropriate for modeling the effect of the clinics. Describe you conclusions from looking at the plot.

   (d) Model the effect of prison when you treat the clinics as stratified confounding variables. What do you conclude? Describe in your own words the difference between this model and the model in part (b).

   (e) Perform a test to see if there is a significant interaction between the prison variable and the clinic variable. We still want to use a stratified model.What do you conclude? Explain what the interaction term means.

2. The data set `retire` has information on the life expectancy of individuals living in a senior care facility. We begin by modeling `time` column which is the survival time in months spent at the facility. The indicator column `death` will be used as our status variable. We would like to model the difference between men and women so there is a column `gender` which is 1 for men and 2 for women.

   (a) Use a Cox Proportional Hazards model to test whether there is a significant difference between men and women. Report the likelihood ratio statistic and the appropriate P value.

   (b) Fit another model that adjusts for the confounding variable `ageentry` which gives the age in months of the subject when they entered the facility. Use the `anova` function to calculate the appropriate likelihood ratio test. Do you come to the same conclusion as in part (a)? How do you explain any difference?

   (c) Fit a model with an interaction between age and gender. What do you conclude?

   (d) Plot complementary log-log plot comparing the effect of gender on the survival time. Do you think the proportional hazards assumption is reasonable for this model?

   (e) Explain clearly why we chose to use `ageentry` as our covariate and not `age` which is the age of the subject when the event occurred.

3. Using the same `retire` data set, I want to fit a Generalized Cox Model where the effect of gender is different before and after 48 months.

   (a) Use the `survSplit` function to construct a new data frame with additional rows that split the `time` variable into before and after 48 months.

   (b) Use `coxph` to model the effect of gender including a change of parameter at 48 months. Please include the age at entry if that is still appropriate. Use our likelihood ratio test to determine if gender is still significant in this model.

   (c) Give 95% confidence intervals for the hazard ratio for men before and after the 48 month cut off.

   (d) Would you conclude that it is important to consider a change in the effect of gender before and after 4 years in the retirement facility?