

ISyE 6404 Computer Project #1 (CP-1): Proportion-Hazards (PH)-Regression

All enrichment and computer projects should follow the guidelines stated in EP-1 for preparing and submitting your reports.

1. PH-Regression Data-Fit (70%)

- 1) Locate a *data set* in the field of your interest, e.g., eCommerce, medical study, drug development, supply-chain/logistics operations, to practice the PH-regression technique. That is, run PH-regression in R-software package to see what outputs it provides.
- 2) Use the *bootstrap method* to construct a 90% pointwise confidence interval (CI) of beta-coefficient(s) based on the PH-regression parameter-estimate(s). With the 10000 bootstrap samples you have the entire distribution of beta-estimate(s). Thus, you can use the methods below to construct CIs.
 - i) Calculate standard error (denoted as $s.e.(beta-est)$) of a beta-estimate from the 1000 bootstrapped beta-estimates. Use $[beta-est. \pm 1.645*s.e.(beta-est.)]$ as the large-sample CI.
 - ii) If the R-program provides 90% confidence interval from either normal or chi-square approximation based on **the large-sample theory**, it would be worthwhile to compare the results against the ones given above to examine what R-program is doing.
 - iii) If the distribution of the bootstrap samples of beta-estimate is not symmetric like normal, use $10000*0.05 = 500^{th}$ lower and upper percentiles as the second large-sample CI. This is the so-called Percentile-Bootstrap-CI.
 - iv) Apply the Bias Correction (BC) Method to obtain BC-Bootstrap-CI. See Section 15.3 textbook for the details of the BC-method.
 - v) A better way to construct the large-sample CI for situation in (iii) is to adjust the percentiles in two tails for having 100 bootstrap samples in total. That is, use a trial-and-error to locate these two tail-percentiles (there are better methods to find them) for a non-symmetric distribution. For example, one tail might have 700 bootstrap samples, and the other tail has 300 bootstrap samples. The locations of these two tails form the CI. Students can look into the so-called Highest-Probability-Density Confidence Interval (HPD-CI) for a better version of this idea.

- 3) Skim through the CP-1 reference given in Files> Projects> directory to write a half-to-one-page report for summarizing the work there.
- 4) Outline steps for implementing one of the studied procedure addressed in the reference. You DO NOT need to implement them, but describe how to do it.
- 5) Discuss what might be difference from the results getting in (1) and (4).

2. Large-p-Small-n Problems in PH-Regression (30%)

- 1) Search the Internet using the key words from the title given in Task #2.
- 2) Locate three publications with a section on real-life data analysis/example.
- 3) Briefly summarize (in an half-page report for one publication)
 - i) what is the goal of the study,
 - ii) key ideas in solving the problem, and
 - iii) what do you learn from their real-life data study.