Homework #2

Survival Analysis with Time-varying Covariates and Two-stage Models

Due: Mon, Feb 15, 2021 by 11:59pm

Total possible points: **105**; Max score: **100**

This assignment uses the "aort_new.csv" file in the "Homework" folder on Canvas, as in Homework #1.

As a reminder, the data set contains data from 500 patients who received a human tissue valve in the aortic position. Patients either received a subcoronary implantation (SI) or a root replacement (RR). These patients were followed over time and longitudinal aortic gradient measurements were collected at echo examinations. These patients are at risk of experiencing death following their surgery. The following columns are available in the data set:

- **id** : unique patient id
- **aort.grad**: aortic gradient
- **time**: visit date (years from surgery date)
- **event**: death status (1: dead, 0: patient is alive at end of follow-up)
- **survtime**: years from surgery date until last follow-up
- **oper**: categorical indicator of type of operation (SI or RR)
- **sex**: gender
- **age**: patient age at time of surgery

Note: It might be useful to begin by creating a square root transformation of aort.grad and adding it to the data set (*sqrt.aort.grad*).

Question 1: Survival analysis with a time-varying covariate

a) **(10 pts)** The "aort" data set is in a longitudinal format (multiple rows per patient, one row for each visit time). Create the start/stop time data set, where the beginning of a time interval represents a measurement time, and the end of the final time window represents the survival time. You also need a new status indicator which takes the value of 0 for all intervals except the last (per patient), where it is 1 if an event is observed, or 0 if a patient is censored. For this created data set, print the rows for Patients 1 and 2.

b) **(5 pts)** For Patient 3, create a stepped line plot of their square root aortic gradient. Use a vertical dashed line to indicate the patient's observed survival time.

c) **(20 pts)** Fit an extended Cox survival model with square root aortic gradient as a time-varying covariate, and operation type and sex as a time-independent baseline covariate. Interpret the coefficient estimates. How is the interpretation from this model different from a survival model that uses only the baseline values?

d) **(10 pts)** Comment on why the assumptions made by the extended Cox model may not be appropriate for this data set?

(Questions continued on next page)

Question 2: Two-stage model

a) **(15 pts)** Fit a mixed effects model for square root aortic gradient, with fixed effects of linear time, operation type, and sex, and random intercept and random linear slope for time. Interpret the coefficient estimates from this model.

b) **(15 pts)** Obtain subject-specific predictions from the mixed effects model (using the contributions from the random intercept and random slope) and use it as a time-varying covariate in a Cox survival model, which also includes operation type and sex as time-independent baseline covariates. Interpret the coefficient estimates from this model.

c) **(20 pts)** Use bootstrapping to compute the standard errors for the parameters in the Cox component of the two-stage model. How do these differ from the standard errors estimated in the model in (2)(b)? In general, why would you expect there to be differences between these two methods for inference?

d) **(10 pts)** Compare the estimates between the two-stage model and the time-varying covariate model in (1)(c). Comment on any differences and why these differences may exist.

(End of assignment)