Homework #1

Linear mixed effects and Survival Models

Due: Mon, Feb 8, 2021 by 11:59pm

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

This assignment uses the "aort.csv" file in the "Homework" folder on Canvas.

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

Question 1: Longitudinal Model

a) **(5 pts)** The number of measurements varies per subject. Calculate the number of measurements per subject and summarize it. Look at the distribution of aortic gradient, and explain why we should consider a square root transformation of this outcome. Create the column "sqrt.aort.grad" in the dataset with this transformation.

b) **(5 pts)** Plot the longitudinal profiles of *sqrt.aort.grad* for 5 subjects who during follow-up experience death and 5 subjects that do not. Describe any observed trends.

c) **(20 pts)** Fit a random intercept model with *sqrt.aort.grad* as the outcome, assuming a linear effect of time. Is there evidence that mean aortic gradient changes over time? Does this change vary by type of surgery? Is there evidence that there is variation in baseline aortic gradient between subjects? Perform the necessary hypothesis tests to justify your answers. Interpret the coefficient estimates from your final model.

d) **(10 pts)** Now, extend the model to allow for a random slope. How many more parameters have been estimated compared to the random intercept model? Is there evidence that the change in aortic gradient over time varies between subjects?

e) **(10 pts)** We have assumed linearity in the above models. Explore including a non-linear function of time (e.g., use polynomials or splines) in the fixed and random effects. Comment on the results. *(Don't go crazy here, just demonstrate that you have tried a more complex relationship in modeling the longitudinal trajectory of sqrt.aort.grad).*

f) **(5 pts)** Using the model in (1e), plot the estimated population-averaged trajectory and plot the predicted patient-specific trajectories for 10 patients on the same plot. Compare these lines and comment.
g) **(10 pts)** Obtain the variance-covariance and correlation matrices for the longitudinal outcomes from the models in 1c (random intercept), 1d (random intercept + random slope), 1e (non-linear functions) for Patient 2. Comment on the differences.

Question 2: Survival Model

Using the same data set, you will now be applying a survival model to explore the relationships between the baseline variables and survival. Create an appropriate individual-level data set for this analysis. *Note: Make sure that you are using the baseline value of the longitudinal variable.*

a) **(10 pts)** Plot the Kaplan-Meier survival curves:
    i. Overall
    ii. By operation type (SI, RR)
b) **(10 pts)** Using the results from 2(a)(ii),
    i. Estimate the predicted survival at 10 years for each operation type
    ii. By inspecting the plot, given an (approximate) estimate of the hazard ratio.
c) **(5 pts)** Categorize the baseline value of the longitudinal variable (aort.grad) into 4 equally sized groups. Plot the Kaplan-Meier curves by this newly created variable. What is the relationship between baseline aortic gradient and survival?
d) **(5 pts)** Fit a Cox model using "coxph" with operation type ("oper") as the only predictor. Interpret the model results.
e) **(5 pts)** Predict the survival curves from the Cox model and overlay these on the Kaplan-Meier curves in (a)(ii). Does the Cox model give a reasonable fit to the data?
f) **(5 pts)** As in Q1, apply a transformation to the baseline value of aort.grad (*sqrt.aort.grad*) and add this variable to the Cox model. Interpret the model results.