AA502 – Survival Analysis

Homework 3

Matthew Austin

Background

Several hurricanes struck the Gulf Coast, resulting in severe casualty and property damage. One of the major defenses against hurricanes is the coordination and maintenance of pump operations during a critical 48-hour period (four high tides). As employees for the Steering Committee of the Center for Risk Management, your team is tasked with conducting a survival analysis for the pump stations in the area.

Data

The `katrina` dataset contains information collected from 770 pump stations. Pump survival is denoted in the variable `survive`, which is an indicator for whether or not the pump survived the entirety of the storm. There are five potential failure conditions which take the following values in the variable `reason`:

- (0) no failure (this is equivalent to `survive = 1`)
- (1) `flood`: overflow or accumulation of water that submerges the pump station
- (2) `motor`: mechanical failure
- (3) `surge`: onshore gush of water typically associated with levee or structural damage
- (4) `jammed`: accumulation of trash or landslide materials

There are eight factors that may influence the survivability of the pump stations. Not all pumps have each characteristic, but some are available for maintenance or upgrade and denoted as such (where a value of `1` indicates the factor is already present and cannot be upgraded):

- `backup` pump (upgrade available): a redundant system used to protect the station from flooding when the main pump is not operating
- `bridge` crane (upgrade available): allows vertical access to equipment and protecting materials
- `servo` (upgrade available): servomechanism used to provide control of a desired operation through the Supervisory Council and Data Acquisition (SCADA) systems
- `trashrack` cleaner (upgrade available): protects hydraulic structures against the inlet of debris, vegetation, or trash
- `elevation` (maintenance available): elevation of the pump station; may be altered by one foot via maintenance
- `slope`: ravine slope surrounding the pump station
- `age`: difference between the pump's installation date and the current date
- `H1--H48`: pumping status reported by pump stations during a 48-hour emergency period (accuracy of pump status not guaranteed to be error-free)

Assignment

Provide a follow-up to your last report and a set of recommendations summarizing the findings from your analysis. In this assignment, you will model `motor` and `surge` failures together and treat all other failure reasons as censored. To do this in `SAS`, use `hour*reason(0, 1, 4)`. In `R`, do `Surv(time = hour, event = reason %in% c(2, 3))`.

- Create both an AFT model and a Cox regression model with the following variables: `backup, bridgecrane, servo, trashrack, elevation, slope, age`. Which of these models do you prefer? Why?
- Provide the coefficient estimates and standard errors for the Cox regression model and interpret some of them (at least one categorical and one continuous). Is there any evidence that any of these effects might not be constant over time? (For this assignment, you are only required to check the assumption, although you are welcome to correct it appropriately if you feel compelled to do so.)
- The Army Corps of Engineers believes that motor failure is more likely if the motor has been running for 12 consecutive hours prior to failure. Add this to your model using the `H1--H48` variables and describe its effect on failure. What is your conclusion? (This is **not** a simple task to implement computationally; think carefully about what you are doing!)

Questions/topics to know

- PH models: the response, how/why it relates to the event, interpreting estimates
- What are the differences between PH models and AFT models? What are some advantages and disadvantages of both? Are there any instances in which they're related?
- How/why is the Cox regression model different from a standard PH model, if at all?
- What are the different types of residuals in survival analysis? What are each of them used for?
- What are the assumptions of proportional hazards models? What are some ways to check them? What can you when they are violated?
- What is the difference between a time-dependent coefficient and a time-dependent predictor?