AA502 – Survival Analysis

Homework 2

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. For this analysis, you will only focus on one type of failure—`flood`, so you will use `reason` instead of `survive` as your status variable, treating everything that didn't fail due to flooding as censored.[1] In `SAS`, you can specify multiple censoring codes by separating them with commas: `hour*reason(0, 2, 3, 4)`. In `R`, use `Surv(time = hour, event = reason == 1)`. This new report should include the following information:

- Create an AFT model with the following variables: `backup, bridgecrane, servo, trashrack, elevation, slope, age`. Don't worry about interactions for now.
- Fit this model with the exponential, Weibull, log-normal, and log-logistic distributions and discuss the possible distribution of the data.
- Once you have chosen a distribution, provide the coefficient estimates and standard errors from your model and comment on anything that you find interesting. You must interpret a few of them (at least one categorical and one continuous).
- The Army Corps of Engineers only has enough time and money to upgrade 20 pumps. Choose an upgrade for 20 pumps (one upgrade per pump); you can decide this however you'd like, but explain why you've recommended those 20. You don't have to give one explanation for every individual pump, just state any overarching reasons for upgrading the particular set of pumps that you chose.

Questions/topics to know

- AFT models: the response, how/why it relates to the event, interpreting estimates
- What are the assumptions of AFT models? What are some ways to check them?
- Pros and cons of parametric vs. nonparametric methods
- General relationships (if any) among the common error distributions
- How to get predicted quantiles/survival times from AFT models (the general process; you don't have to memorize any specific equations)
- For your own practice, try to come up with some situations for which you might prefer one particular distribution over another
- Also try to come up with some examples for the different types/ways of censoring

[1] This can be done for all failure types simultaneously and is called **competing risks**, which we'll learn about at the end of the course.