DATA 596 Survival Analysis – SPRING 2024

Group Project #3 Details

OBJECTIVE: The high-level goal of this class is for students to develop real world skills to collect, wrangle, explore, and statistically model data utilizing a variety of survival analysis methods. This project provides student with an opportunity to demonstrate knowledge and growth in these areas by analyzing real data and presenting it to a broad audience.

Presentation requirements

- Due date of presentations: 4/24/24, 6:00pm in Canvas. Complete submissions include the presentation (ppt or pdf) and the R code used to implement the analysis. Name the files with the following naming convention: DATA596_SP24_PROJECT3_LASTNAME1-LASTNAME2-LASTNAME3-LASTNAME4.[pptx, Rmd, pdf, etc]
- Length of presentation: ~15 minutes, not including time for Q&A
- Each person on the team must present at least 1 slide during the team presentation.
- # of slides: variable, but your presentation must contain at least the following:

  - **(15%) BACKGROUND INFORMATION**
    - What information would be helpful to your audience who doesn't have background knowledge on this subject? Need context for the study and why it was done.
    - What was the goal of the study?

  - **(20%) DATA & METHODOLOGY**
    - How was the data collected?
    - What data was collected?
      - response variables
      - covariates and/or factors
    - What were the experimental/observational units?
    - What was the sample size?

  - **(50%) RESULTS/ANALYSIS**

- Must have:

    - At least <u>2</u> Kaplan Meier curves

    - If dataset has time to failure data: results of a semiparametric or parametric survival regression model that satisfies the assumptions of said model and is appropriate for the type of data being analyzed

        If dataset contains data paths, an appropriate degradation model must be fit to the data

    - Accompanying descriptive statistics at different levels of resolution (overall, broken down by subgroups, etc)

    - Diagnostics for your model that verify its goodness-of-fit. This can include visuals and/or statistical metrics.

  - Need clear explanation of statistical results in context of the original problem

- **(15%) CONCLUSIONS**

    - What were the main takeaways of the analysis?

    - What are the limitations of your analysis?

    - Future steps/analyses of interest?

- **OTHER:** Rmd or .R file of code used in analysis (may provide more than 1 file)

Grading rubric

Groups will be assigned a group-level grade according to the following breakdown:

Component	Grade %
Intro/background	15
Data & methodology	20
Results	50
Conclusion	15

A more detailed breakdown is found below:

Category	Full credit work fulfills the following:
Intro/background (**15**)	Sets the stage for a generalist audience & provides adequate context for the rest of the presentation. Make sure you provide the appropriate amount of detail for an audience that has more background knowledge than a novice but doesn't possess expert-level background knowledge.

Data & methodology (**20**)	Describes the data that was examined.
	
	• How was the data collected?
	• What data was collected?
	    • response variables (+units)
	    • covariates and/or factors
	• What were the experimental/observational units?
	• What subgroups did you examine?
	• What was the sample size?
	• How prevalent was censoring in your dataset? What type of censoring was present in your dataset?
	• Other meta data of interest?
	
	Did you choose to omit any data? If so, why? Some justifications may include data quality issues, lack of representativeness, etc.

Results (**50**)	Clearly and accurately communicates the results in context of and with minimal jargon. Must have: • The questions being answered • A clear story. Start with high-level statistics and provide more resolution as the presentation progresses. • Appropriate and clear data viz for the data provided. All aspects of the graph and presentation must be legible. • Appropriate descriptive statistics provided alongside the graphical results • *Summarized* statistical models with their interpretation in the original context. Don't give me a screenshot of your output. • Correct interpretation of model results and justification for the chosen model • Diagnostics showing appropriateness of the chosen model. • All slides and contents should be clear and concise.

	Must mention any caveats of the interpretation and/or potential violations of analysis assumptions
	All plots must be clearly labeled, contain a compelling title and axis labels, and be generated via Tidyverse functions.
Conclusion (15)	Connects the results of the previous section to the original objectives of the analysis.
	Lists any limitations of the analysis and future directions of your research.

The extent to which team members' contributions were equal will be assessed after the in-class presentations are completed. You will be sent a participation survey in which you will debrief on your contribution to the analysis. You will also be given an opportunity to provide feedback on your partner(s)' contribution to the project. Cases of inequity in workload will result in a penalty proportional to the inequity. For instance, if on a 2-person team it is shown that student A only contributed 1/3 of what student B did, then student A will receive a 66 for the participation grade. Recall that participation is worth 20% of your overall grade. Please ensure that you and your teammates are communicative and respectful throughout this group project.