DATA 599 Group Project 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: 2/9/23, 6:00pm in your group's Google folder. 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: DATA599_GROUPPROJECT1_LASTNAME1-LASTNAME2-LASTNAME3-LASTNAME4.pdf[or Rmd]
- Length of presentation: 10-15 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:
      - a Kaplan Meier curve
      - results of a Log Rank Test between 2+ groups
    - 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 question being answered • a Kaplan Meier curve (with correct interpretation)

	• a Log Rank Test between 2+ groups (with correct interpretation) Must mention any caveats of the interpretation and/or potential violations of Log rank test assumptions. All plots must be clearly labeled with a compelling title and axis labels.
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.

Although participation is not a dedicated category in the above rubric, students are expected to actively engage in the group project discussions and share the project workload equally. Violations will result in a penalty up to 30% of the student's overall participation score of the semester.