STAT 417: SURVIVAL ANALYSIS METHODS
California Polytechnic State University, San Luis Obispo

PROJECT

Phase	Percent	Due date
1: Team selection	2%	Fri, Feb 9 9AM
2: Data submission	18%	Tue, Feb 20 9AM
3: Final report	78%	Fri, Mar 16 9AM
4: Peer evaluation	2%	Fri, Mar 16 9AM
	100%	

Purpose

The purpose of this project is to provide you the opportunity to perform some hands-on analysis of time-to-event data using the methods learned in class, and implementing these methods using Minitab and R.

Seek help

If you have **any** questions or need **any** clarification at **any** point, please (1) come to office hours, (2) schedule an appointment to meet with me, or (3) post on the course discussion forum.

Project options

Option 1 Analyze an existing time to event data set. Be sure to include a thorough citation/reference for the data set.

Option 2 Your team may also perform an experiment or survey to collect time-to-event data. You may want to discuss your plan with your professor if you want to pursue this option. The project may be more enjoyable and interesting if you collect your own data for a research question that interests you.

How original is your data set? I encourage some digging and research on your own to find a data set that is not already available in Minitab and R, and that has not been already examined in a textbook. You can improve your project grade by analyzing a data set that has not been introduced in other textbooks or that is not available in any statistical package. The investigation should be based on your group work, not a summary of work completed elsewhere.

Note: You cannot use a data set that has been introduced in class.

Minimum data set requirements

Regardless of the option selected, that data set must contain:

1. A time-to-event variable

2. A censoring status indicator variable (right censoring only)

3. One categorical (explanatory) variable

4. One continuous (explanatory) variable

5. At least one other explanatory variable

6. The data set should contain enough observations to satisfy the following rule of thumb: 10 observations per explanatory variable in the original data set.

Writing guidelines

- When writing a paragraph or summary, do not refer to variable names; instead, refer to the meaning of the variable. For example, I might have a variable in data set called `prev_stats` which indicates if a student has had previous experience with survival analysis prior to STAT 417.

  - It would be *incorrect* to write:

    The percent of students with prev_stats is 35%.

  - It would be *correct* to write:

    The percent of students with previous experience in statistics is 35%.

- If you start a sentence with a number, you must spell it out.

  - It would be *incorrect* to write:

    64 people participated in the study.

  - It would be *correct* to write:

    Sixty-four people participated in the study.

- When writing, round numbers (including p-values) to an appropriate number of decimal places. (You don't need to worry about rounding in your R output.)

  - It would be *incorrect* to write:

    The average weight is 130.2384638 pounds.

  - It would be *correct* to write:

    The average weight is 130.2 pounds.

  - An appropriate number of decimals to round a p-value is between 2 to 4 decimals. For really small p-values you may write "the p-value is less than 0.01", if appropriate.

- The length of the body of the report should be between 5 to 10 pages double-spaced (including graphs and tables). (Three to five pages single spaced is also acceptable.) You can add on as many pages as you want in the Appendix.

- Please write clearly and coherently, using complete sentences in your report.

- All graphs and tables should be directly inserted in the body of the report, not piled at the end. Appearance is important.

- If you are using Rmarkdown, note that is does not automatically indicate misspelled words like Microsoft Word does. In order to execute a spell check in RStudio, submit `F7`. Please make sure you proof-read your writing carefully.

- **Please submit your final report as a pdf.** If you are compiling your report from Rmarkdown, you need to install MikTex in order to *knit to pdf.*

Assessment

Please review rubrics associated with each phase on PolyLearn for full grading guidelines.

- Meeting the minimum requirements as outlined

- Correctness of your interpretations

- Originality of the data set used

- Organization and professional presentation of your complete report

Phase 1: Team selection

- **Identify** 2 to 3 classmates to work with (groups should be of 3 to 4 students total). Include their full name and e-mail address in the project submission. If you need help finding a team, please let me know. Include this information in your Phase 1 submission. In addition, create your group on PolyLearn by clicking on `Select groups for project` in the Project content area of PolyLearn.

- **Designate** one team member to be the primary contact person.

  - This person is responsible for facilitating communication between team members.
  - This person is my primary contact in case I have questions for your team.

- Create your team **contract**: state three things that you think makes a group work well together that you all agree to do.

Phase 2: Data submission

Submit the data set you plan to use for your project. Include a brief description of the data set. If you are pursuing *Option 1* be sure to provide a full citation for the data origin. If you are pursuing *Option 2* be sure to provide the protocol used to collect the data. Lastly, include brief descriptions of the variables in the data set (e. g., a data dictionary). Please state which variables in your data set satisfy/correspond to the specifications listed in minimum data set requirements (both with variable names and descriptions).

Note: Every group must work independently of each other - all groups must select a different data set.

<hr/>

Phase 3: Final report

- **Introduction**: Provide a brief description of the data set you have selected, i.e. the variables, how the study was conducted, and any other details you can provide. Be sure to define the time-to-event variable.

- **Descriptive Summaries**: Use nonparametric methods to describe and compare survival experiences of the different groups in your data set. Examine and comment on the survival experience of all individuals in your data set, as well as the survival experiences of different groups of individuals. Use graphical displays such as Kaplan-Meier curves, and/or estimated hazard curves, and perform and report results of the log-rank test. In addition, compute various descriptive statistics of interest.

- **Regression Analysis**: Develop a Cox regression model to investigate the effects of predictors (explanatory variables) on the hazard rate. You need only present your final model, including, but not limited to, parameter estimates, standard errors, and p-values. Discuss whether model assumptions have been met and whether you removed any influential or outlying cases. Be sure to present the residual diagnostic plots to support your discussion. Lastly, for some key values of the predictors, present estimated hazard ratios and confidence intervals for the corresponding true hazard ratio. Provide interpretations of the hazard ratios.

- **Discussion**: Based on the results in the main report, discuss some insightful discoveries about the time-to-event data.

- **Conclusion**: Summarize your most important findings.

- **Appendix**: Include all `R` code used to conduct your analyses in the Appendix at the end of the report.

<hr/>

Phase 4: Peer evaluation

- All team members are expected to contribute equally to the project.

  - Two percent of your *individual* project grade comes from *completing* the peer evaluation form on PolyLearn.
  - After peer evaluations are assessed by the instructor, an *individual's* project grade may be adjusted by $\pm$ 20% from the overall group grade.

- Here is how you will assess **yourself** and your peers:

  Rate the individual on the following attributes according to the scale

  1 = Strongly disagree, 2 = Disagree, 3 = Agree, 4 = Strongly agree
  - ∗ Communicated well with group members (electronically or in person)
  - ∗ Willingly volunteered for or accepted assigned tasks

* Contributed positively to group discussions with useful ideas
* Completed work on time or made alternative arrangements
* Did work accurately and completely
* Contributed a fair share to the project
* Overall was a valuable member of the team