

# MA405: Survival Analysis

## Final Project (Nov 30 – Dec 21)

---

### Project Description:

This is a real data analysis project to be accomplished in groups. There are 28 students in total, it would be great to have 8 groups of size 3 and 1 group of size 4.

The [project schedule](#) is:

Task	Deadline	Remark
Finalize group member	<b>December 4, 11:59pm</b>	Report to TA by QQ
Submit project materials and give in-class presentation	<b>December 21, 9:00am</b>	Submit on Blackboard, <a href="#">group submission</a> .
Submit reviews of other groups' project and group members' performance	<b>December 25, 11:59pm</b>	Submit on Blackboard, <a href="#">individual submission</a> .

The following dataset is provided:

Dataset	Description
arrhythmia.csv	Data from a large study of patients with arrhythmia.

Please refer to the README file for detailed information. Perform comprehensive data analysis on any problem you would like to explore. You are welcomed to use any other dataset you found and make sure to clearly explain the origin of the dataset. For the data analysis, applying methods learned in this course is required and other methods not covered in this course are also encouraged. Make [presentation slides](#) and write a [project report](#) to present your data analysis and results.

### General Guidelines on Data Analysis:

In complex datasets, there are always a variety of interesting phenomena to explore: effects may be nonlinear, there may be interactions, proportional hazards assumption may not hold, there may be strange outliers... I encourage you to be creative in modeling. Blindly following automatic stepwise variable selection procedures can prevent you from seeing important relationships in the data.

Thinking about whether a model makes biological/scientific sense is also important. Don't feel compelled to only include significant terms in your model. If a variable should affect survival, but for whatever reason doesn't, that may be interesting to report as well.

## Guidelines on Submissions:

1. Form groups of size three or two (three is recommended) and report your group members to our TA through QQ by [December 4<sup>th</sup>](#). If you are unable to find a group, let the TA know as soon as possible. A group number will be assigned to each group.
2. Prepare your [presentation slides](#), [project report](#), [R code](#) (as well as a copy of the dataset if you use a dataset other than the one provided) in a zip file, name it “GroupXX\_final.zip” (XX is your group number). Submit the zip file on Blackboard, the deadline is [December 21<sup>st</sup> 9:00am](#). Each group will give an in-class presentation on December 21<sup>st</sup>.
3. After submission, each group’s submitted materials will be uploaded on Blackboard so that all groups are able to review other groups’ project. Submit your scores and comments on other groups’ project as well as group members’ performance on Blackboard by filling a form, the deadline is [December 25<sup>th</sup>](#).

### For the **in-class presentation**:

- Around 15-20 minutes in total. Each member has to present (at least 5 minutes).
- English is preferred but Chinese is also accepted.

### For the **project report**:

- Include abstract, introduction, exploratory data analysis, statistical modeling, conclusion, discussion, reference. This is a list of recommendations and please exercise your judgement on what should be included in the report.
- Control the report within 15 pages.
- English is preferred.

## Assessment:

- Instructor and TA’s review based on the presentation (30 marks)
  - Background and problem setup (5 marks)
  - Exploratory data analysis (5 marks)
  - Statistical modeling (10 marks)
  - Conclusion and discussion (5 marks)
  - Presentation skill (5 marks)
- Instructor and TA’s review based on the project report (30 marks)
  - Background and problem setup (5 marks)
  - Exploratory data analysis (5 marks)
  - Statistical modeling (10 marks)

- Conclusion and discussion (5 marks)
- Writing skill (5 marks)
- Other groups' review based on the project presentation and report (15 marks)
  - Assume  $n$  groups, each group ranks the others by assigning scores  $1, 2, \dots, n$ , no tie.
  - Groups will be stratified to 3 levels (high, middle, low) based on the aggregated score.
  - The high groups will get 15 marks, the middle groups will get 10 marks, the low groups will get 5 marks.
- Intra-group review (10 marks)
  - Each group member will give a score ranging from 0 to 10 to other group members.
  - The member with the highest average score will get 10 marks, the member with the lowest average score will get 6 marks, other members will get 8 marks.
- Review of other groups' projects and group members' performance (15 marks)
  - Finish all other groups' and group members' scoring (5 marks)
  - Provide concrete comments on other groups' projects and group members' performance (10 marks)

Each student will know the aggregated score from other groups' review and intra-group review. I would like to keep the raw scores confidential.