BIOS 7712: Final Project

Overview

This project is intended to expand your knowledge about joint modeling and to enhance your skills as a statistical researcher. For your final project, you may choose **ONE** of the following options:

(1) Reading and presenting on a **research paper** related to joint modeling
(2) Conducting a **simulation study** to compare joint models and two-stage models

See the details and requirements for both options below. Whichever option you select, your project should be completed individually. The final project is worth 30% of your final grade. Identify on Canvas the option you have chosen by **Feb 12, 2021**.

Option #1: Research Paper

1. Read paper

You will read a research paper related to joint modeling. You can select a paper from the list provided on Canvas, or can select your own paper and have it approved by the instructor by **Feb 12, 2021**.

2. Prepare a presentation

You will prepare and record a ~10-minute oral presentation (5-10 slides) describing the main ideas, methods, and results/conclusions in the paper that you read. The goal is to teach your fellow classmates about the paper you have read. The presentation should be for a statistical audience that has not read the paper. The presentation should be uploaded onto the Canvas discussion boards by **Wed, Feb 24, 2021, 11:59pm**.

3. Prepare a report

You will submit a report (no more than 2 single-spaced pages) that should:

1) Briefly summarize the paper (research question, methods, data, results) (~ 1 page)
2) Present a thoughtful critique of the paper (~ 1 page). Some possible things you can comment on are (but should not be limited to):
    a. Why did you choose this paper?
    b. What are the strengths of the paper?
    c. What are some of the limitations?
    d. What are some future directions for this research?

4. Participate in/Generate a discussion

You will be assigned to watch two other students' presentations. Post a **substantive** question/comment on each of their videos. Respond to any questions/comments that are posted on your video.

Expected Deliverables:

1) A 20-minute pre-recorded presentation **due Wed, Feb 24, 2021, 11:59pm – 40 pts**
2) Question/comment on presentations **due Fri, Mar 5, 2021, 11:59pm – 20 pts**
3) Final report (max 2 pages, single-spaced) **due Fri, Mar 5, 2021, 11:59pm – 40 pts**

Option #2: Simulation Study

1. Conduct simulation study

You will conduct a simulation study to compare the performance of the standard joint model and the two-stage methods that were discussed in class. You will compute the bias and standard errors of the coefficient estimates from these two methods.

2. Prepare a report

You will submit a report that should:

1) Identify your hypotheses and explain how the simulation study will address them
2) Succinctly describe the setup of your simulation study, including definitions of variables and data generating models
3) Describe the models that you fit
4) Present and interpret the simulation results
5) Give appropriate conclusions tying the results back to your hypotheses

Your report should not exceed 2 pages (single-spaced) and should include 1-2 relevant tables and/or figures showing the results of your simulation study. Your report should resemble the simulation study section of a research paper.

Expected Deliverables:

1) Final report (max 2 pages, single-spaced) **due Fri, Mar 5, 2021, 11:59pm – 70 pts**
2) Commented code to produce results **due Fri, Mar 5, 2021, 11:59pm – 30 pts**

Note: Your simulation study can take 1-2 hours to run, so make sure you start early!

Outline for the Simulation Study:

The true parameter values are given in Table 1 below and skeleton code for implementing the following simulation study is given on Canvas.

For 500 simulated data sets, we simulate data for 250 individuals using the following steps.

1. Simulate longitudinal marker values

a) Generate longitudinal covariate values based on the following model:
$y_{ij} = \beta_0 + \beta_1 t + b_{0i} + b_{1i} t + \epsilon_{ij} = m_i(t) + \epsilon_{ij}$, $\epsilon_i \sim N(0, \sigma^2)$ and $(b_{0i}, b_{1i})' \sim N(0, A)$ with diagonal variance-covariance matrix $A = \begin{bmatrix} A_{11} & 0 \\ 0 & A_{22} \end{bmatrix}$

b) Generate random inspection times for each individual from a Poisson process to achieve an average of one measurement per year, and a maximum of 10 measurements per person (including baseline).

2. Simulate survival and censoring times

c) Generate survival times from the following Cox model with constant baseline hazard $h_0(t) = 0.1$:
$h_i(t) = h_0(t)\exp\{\alpha m_i(t)\}$

d) Generate censoring times from an uniform distribution, such that the mean censoring time is 5 years, and the max follow-up times is 10 years.

3. Fit the models

e) Fit a random effects joint model and a two-stage model.
Note: For the joint model use *method= "weibull-PH-aGH"* for faster computation time. Only extract the estimates that you need.

4. Compute performance metrics

f) Using the results from all 500 simulated data sets, compute the following metrics:

a. Empirical bias
b. Asymptotic standard errors (except for the variance component estimates)
c. Empirical standard errors
d. Mean square error
e. Coverage rates for 95% confidence intervals (except for the variance component estimates)
f. Average model computation time

Table 1. True parameter values for Simulation Study

Parameter	True value
α	0.6
β_0	-0.2
β_1	0.3
σ	0.5
A_{11}	0.5
A_{22}	0.3