Daniel J. Frobish, PhD

Associate Professor

Department of Statistics

MAK A1164

Grand Valley State University

Allendale, MI 49401

Office: (616)-331-3028

Dept office: (616)-331-3355

Email: frobishd at gvsu dot edu

I am originally from Illinois, attending the University of Illinois for my bachelor's degree in math, Illinois State University for my master's degree in math (as well as teacher certification in math for grades 6-12), and Northern Illinois University for my PhD in statistics. I came to GVSU in 2006, and was tenured and promoted to associate professor in 2012.

The purpose of this website is solely for potential REU students to see if they would like to apply for the specific project that I would supervise. All other inquiries can be sent to my email address.

For application information and instructions for REU, please visit http://www.gvsu.edu/mathreu/

REU Project Description: My area of specialty is Survival Analysis, which is concerned with developing models for time to event variables, based on a list of potential predictor variables. Examples of time to event variables include:

- Time to death
- Time in remission from cancer
- Time to germination for a species of plant
- Time in foreclosure for residential properties
- Time until breakdown for a machine

For example, if we are interested in the length of time in remission from cancer, then we want to answer if this length of time is related to other explanatory variables, such as treatment group, gender, age, or other demographic variables. It is well established how to build models to answer these questions, when the number of explanatory variables is reasonably small, relative to the sample size.

However, when the number of explanatory variables is large, it is not clear what approaches should be taken. For example, we can now measure to what extent human genes are expressed by measuring how much protein each gene produces. Because there are many thousands of genes, there can be many more explanatory variables than sample size ("big data" are becoming more prevalent outside of genetics as well). More than a few techniques have been proposed to handle these situations in survival analysis, but little work has been done in terms of determining which ones perform well under various circumstances.
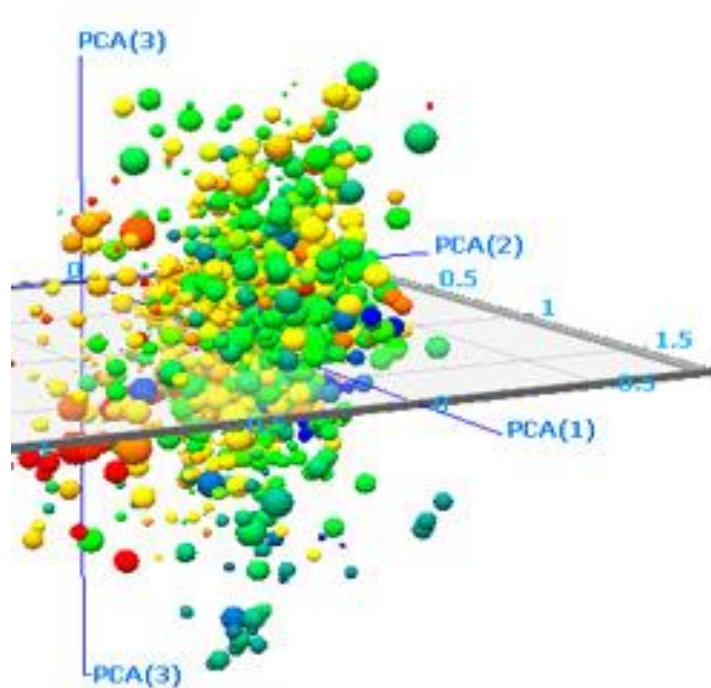
As an additional component, in survival analysis, we are interested in time to event, but in some cases, some individuals will not ever experience the event, even if they are followed indefinitely. This has obvious applications in medical fields, where a major goal is to cure diseases, and being able to use gene expression as inputs to this kind of analysis is extremely important. The simulation studies we will perform will also account for this possibility of cure, and there is nothing in the statistical literature about incorporating the potential for cure, when the data set has many thousands of potential explanatory variables.

The goal of this project will be to compare and evaluate, through simulations, several methods that have been proposed for dimension reduction in survival analysis, including principle components methods, partial least squares methods, and random forest methods, while allowing for the possibility of cure.

What this experience will involve for students:

- Learning background (survival analysis, dimension reduction methods)
- Learning how to use R if needed
- Programming simulations
- Evaluating simulations
- Preparing manuscript for submission to refereed journal for publication

Desirable qualifications of students:

- At least one statistics course at the college level that involves hypothesis testing and basic probability (probability density functions and cumulative distribution functions for continuous random variables).
- Experience programming in the statistical software R (helpful, but not required). https://cran.r-project.org/