

Reflection Nathan Dennis

In Autumn 2022 I decided to participate in the Statistics Directed Reading Program under the Survival Analysis Project. Me and another student were under the co-mentorship between Antonio Olivas and Anand Hemmady.

It was very interesting to learn about what survival analysis is and its many applications, mainly focused on the medical field. We began with the basics, such as some definitions and key terms for survival analysis. Two of the most important terms were survival time and censoring. Censoring was very interesting to me, as this is seen as incomplete data, yet is still useful in studies. In many situations I would simply remove incomplete data, seeing it as not useful. However, after learning more about censoring it was very interesting and unique how survival analysis can manipulate and use incomplete data.

We also discussed different functions within survival analysis, the two most important being the survival and hazard functions. The survival function maps the probability that an individual doesn't experience some event after a certain period of time, which is useful for graphing survival times. The hazard function was very interesting and unique, as it was measured through a potential per unit time, rather than a probability.

Tying into these functions we also discuss various graphs and models, the most important being the Kaplan Meier Curves and Cox PH (Proportional-Hazards) model. Kaplan Meier curves are useful in creating graphs to approximate the survival function. You can plot multiple graphs on one plot, observing differences between two or more graphs if necessary. This is very useful when comparing groups such as a treatment vs placebo group, seeing how different the survival functions may be. This graph takes advantage of censored data, which was very interesting and unique to me. The Cox PH model gives more numerical output than the Kaplan Meier Curves, allowing you to actually get numbers as to how different groups may be by calculating statistics such as the hazard ratio. Both the Kaplan Meier Curves and Cox PH model can relate to one another, one obtaining more visual results and the other numerical results.

We also discussed different topics in Survival Analysis, such as the Maximum Likelihood estimator for different distributions such as the Exponential or Weibull, which can be helpful in estimating the parameters of a distribution. Again, this process takes advantage of the incomplete censored data, something I found very unique in survival analysis.

Finally, we discussed the final project, where I analyzed simulated Covid-19 vaccines from the Moderna vaccine trials. I observed a significant difference in the placebo and vaccine groups, with the vaccine group having a greater vaccine efficacy, meaning it protects against Covid much better than the placebo. I was able to create Kaplan Meier Curves and use a Cox PH model to visualize and calculate some results.

Overall the DRP experience was very exciting and interesting, as I was able to learn more about a topic I most likely wouldn't have in a typical class. I was able to get a brief overview on how medical data is processed and analyzed, then even analyze some on my own.