Math 150 Survival Analysis Project

Carrie Saada

April 8, 2019

```
knitr::opts_chunk$set(message=FALSE, warning=FALSE, fig.height=4, fig.width=5,
                      fig.align = "center")
library(tidyverse)
library(broom)
library(survival)
library(survminer)
```

Outline of "Something New"

For my "something new", I will do a power analysis of the model. I will perform this analysis via simulation by randomly generating a dataset where the null hypothesis is false - there is a difference between the treatment groups. I think it will be useful to look at the power to distinguish between treatment and control, and between all 4 treatment groups (would we be able to distinguish IDV having an effect on survival in the ZDV + 3TC + IDV group, but not in the d4T + 3TC + IDV group?)

The power analysis of the model is relevant to this survival analysis model because it is useful to know whether our model can reject the null hypothesis when it is warranted. If the model is not very powerful and we fail to reject the null hypothesis, it may be a good idea to continue research with a larger sample size. Or, if the model is powerful and we fail to reject the null hypothesis, we can say that if there were a true difference, our model would probably have detected it, so it may not be worth continuing the research.

I'm going to use the class textbook and *The Analysis of Biological Data* by Whitlock and Schluter (2015) to learn about power and how it relates to sample and effect size. I will also read published articles such as Cohen 1992 adn Dorey 2011 (see working bibliography).

This will be challenging because I need to learn how to apply power analysis to an experiment that has already been completed, rather than to determine an ideal sample size. Additionally, I will need to learn how to do power analysis simulations for survival data - how do I generate realistic survival data? What variables do I include? How do I handle effect size? What about censoring?

Working Bibliography

Cohen, Jacob. "Statistical Power Analysis." Current Directions in Psychological Science, vol. 1, no. 3, June 1992, pp. 98–101. SAGE Journals, doi:10.1111/1467-8721.ep10768783.

Dorey, Frederick J. "In Brief: Statistics in Brief: Statistical Power: What Is It and When Should It Be Used?" Clinical Orthopaedics and Related Research, vol. 469, no. 2, Feb. 2011, pp. 619–20. PubMed Central, doi:10.1007/s11999-010-1435-0.

Kuiper, Shonda, and Jeffrey Sklar. Practicing Statistics : Guided Investigations for the Second Course. Pearson, 2013.

Whitlock, Michael, and Dolph Schluter. The Analysis of Biological Data. Second ed., Roberts and Company, 2015.

```
AD = read.csv("AIDSdata.csv")
```

The model is:

```r
model = coxph(Surv(time, censor) ~ tx*(sex + priorzdv) + raceth*sex + ivdrug*priorzdv + karnof + cd4, da
model
```

```
## Call:
## coxph(formula = Surv(time, censor) ~ tx * (sex + priorzdv) +
##     raceth * sex + ivdrug * priorzdv + karnof + cd4, data = AD)
##
##                      coef exp(coef)  se(coef)      z        p
## tx              -1.453900  0.233657  0.862695 -1.685 0.091931
## sex              2.469187 11.812844  0.866826  2.849 0.004392
## priorzdv        -0.015350  0.984767  0.010419 -1.473 0.140691
## raceth           1.670770  5.316262  0.625436  2.671 0.007554
## ivdrug          -0.855328  0.425144  0.337158 -2.537 0.011185
## karnof          -0.056287  0.945268  0.014591 -3.857 0.000115
## cd4             -0.015267  0.984849  0.003145 -4.854 1.21e-06
## tx:sex           1.392238  4.023845  0.666413  2.089 0.036694
## tx:priorzdv     -0.034642  0.965951  0.014195 -2.441 0.014666
## sex:raceth      -1.599834  0.201930  0.578075 -2.768 0.005648
## priorzdv:ivdrug  0.017395  1.017547  0.006697  2.597 0.009395
##
## Likelihood ratio test=104.2  on 11 df, p=< 2.2e-16
## n= 851, number of events= 69
```

P value of 0.000307816 that model is the same without tx (so model is not the same)