STA 635: Take-Home Final

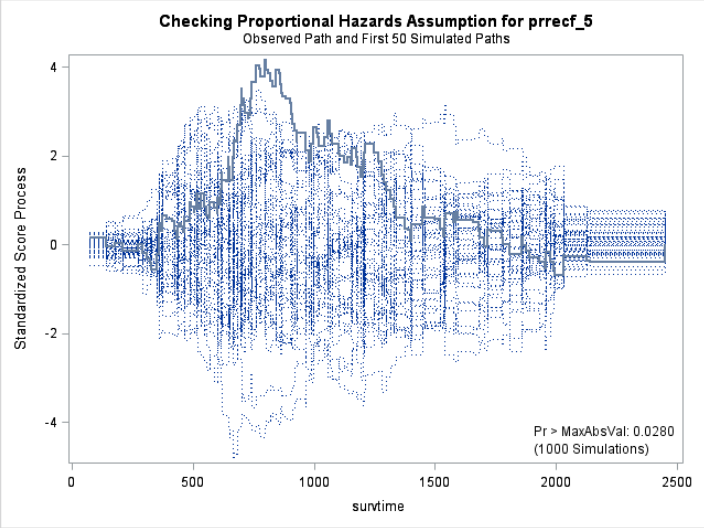
Michael Anderson

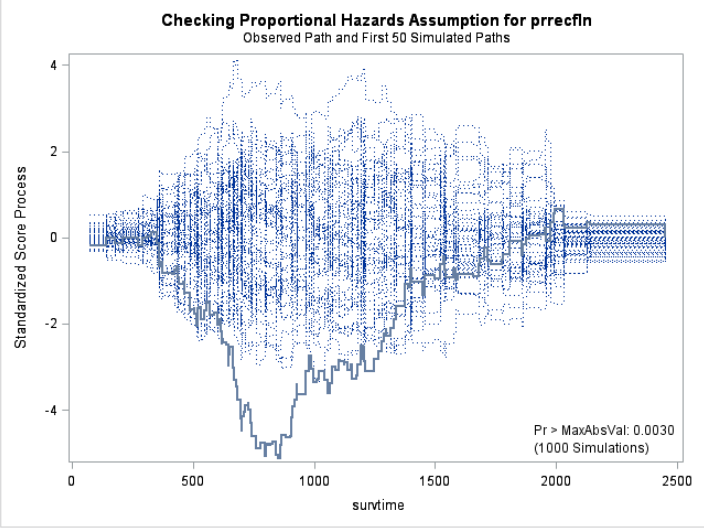
German Breast Cancer Study: Part 1

Our variable of interest in this data set is *survtime* (patient survival time) which is censored with the variable *censdead*. The "best subset" of predictors for patient survival time, using a stepwise selection algorithm with entry significance level 0.25 and exit significance level 0.10, is *nodes*, *prrec*, *size*, and *grade*, in order of their entry into the model. After we selected this "best subset", we next decided to test transformations of the predictors *size* and *prrec*. After using a fractional polynomial technique, with likelihood ratio tests, to check for significant transformations, we decided to include *prrec* in our model as a double transformation of $prrec^{-1/2}$ and $ln(prrec)$, with no transformation on size needed.

After that we checked if recoding *nodes* as a categorical variable would improve our model. We recoded it into three categories, with values of 1-3, 4-9, and 10+. Using this categorical *nodes* led to a significant likelihood ratio test, so we decided to keep it instead of the original variable. We then checked to see if any interactions between our original "best subset" variables would significantly improve the model, and the interaction between *size* and *nodes* did just that, so we added it to our list. Lastly, we checked if any of the original predictors not selected for our best subset would improve our model now that we have added transformations and interactions to it. We did another stepwise selection algorithm, this time with all the variables in our new model, plus the predictors not chosen before, and the algorithm chose to add *hormone* to the model. Also, *grade* was not deemed significant by the algorithm. So it was dropped, giving us our final model.

After getting to what we deemed to be an appropriate model, we decided to see if the model violated the proportional hazards assumption that is implicit in the Cox model we're using. Using the "ASSESS" statement in SAS for our model produced a test of the PH assumption for each predictor in our model, and it seems that *prrec* violates the PH assumption. Looking the plots below, we can see that the p-values on the tests for both transformations of *prrec* included in our model are low enough to be significant. This tells us that we should probably include some sort of time dependent covariate if we want to keep *prrec* in the model.





However, keeping *prrec* in the model for the time being, we next decided to include *censrec* (the censoring variable for recurrence) in our model to see if that would tell us something about the relationship between patient recurrence and survival. But, the variable was insignificant, so it was not included in the model, and no reliable inference could be made with regards to that relationship. Lastly, we wanted to see if the *hormone* treatment was still significant for those patients that did have a recurrence, so we created a new dataset of only women who had an observed recurrence ($censrec = 1$), then ran our original model on that subset to see if we could find a significant *hormone* effect. However, the variable was not significant in the model (but with a p-value of 0.088, it

is perhaps marginally significant). The estimate for the *hormone* coefficient was negative in this model, implying that the hormone treatment did seem to improve survival for patients after recurrence, but not significantly enough to meet the 0.05 standard.

German Breast Cancer Study: Part 2

Continuing with the last idea in part 1, we continued to look at just women who had a recurrence in the study, this time looking at time from recurrence to death (*survrec*). Starting off, we once again used a selection algorithm to choose our best subset, this time giving us a model with *prrec*, *size*, *age*, and *grade* in it. We then went on to use fractional polynomials to check for significant transformations of *prrec*, *size*, and *age*. However, none of the transformations were found to be significant enough to include in the model, using the likelihood ratio test. Next we tried to recode our continuous variables as categorical variables to see if that improved our explanatory power. We used three bracket categorizations for all of three variables, with cutoffs of 45 and 60 for *age*, 25 and 50 for *size*, and 10 and 40 for *prrec*. These cutoffs were created based on background information on the variables of interest, as well as the distribution of those variables in the data. However, once again, none of these transformations significantly improved our model, so we moved on. Next we checked for interactions between the four variables in our model, and the interaction between *age* and *size* was found to be significant, and it improved both AIC and BIC for our model, so we decided to keep it in. Lastly, we went back to original pool of predictors to see if any of the original variables could improve our model now that we've decided to include an interaction term. However, a selection algorithm gave us the exact same model, with none of the other predictors added to it and none of our predictors dropped. So, we decided to keep the model as is.

Cystic Fibrosis and Deoxyribonuclease

To start off, we had to do a little research into this dataset to fully comprehend the design of the experiment and what the variables were. After we felt like we better understood the study, we created a new dataset to handle all the different methods we wanted to use on the data (almost identical to the layout of the dataset given in the notes for recurrent events.) After we did that, we were able to run all of the different types of models we wanted to for this problem. First, we ran a simple model of time to first occurrence on the treatment and FEV values. In our dataset we created a variable named $diff$ which is the length of observation between events, so we used that variable where $enum = 1$ (where *enum* corresponds to which period of observation we are in, 1 for first event, 2 for second, etc.) The variable that denoted censored observations was *status*. Running this told us that the treatment variable, rx, was significant in improving time to first event, even after controlling for fev.

Next we wanted to see if the benefits of the treatment went beyond the first event time. Running the same model as before, this time just using the data for the second observation period ($enum = 2$) but leaving out those who censored in the first time period ($diff \neq 0$). Doing this showed us that both rx or fev were no longer significant factors for time to the second event. Next we used a few different methods to try to get a better idea of how the treatment affected time to recurrence for all events, while still controlling for FEV. We started by using an A-G model with a time interval response variable, and a single strata. So we essentially treated all events as a first event from different patients. Since we had to enter the response as an interval, we weren't able to use the jackknife method for standard errors, so we used the sandwich method instead. This gave us a significant rx variable (with a p-value of 0.0244). Next we tried a PWP model, with time intervals and stratifying on *enum* using all 5 possible observed event times. This gave us a p-value of 0.0440 on rx. Then we ran another PWP model, this time with time gaps instead of intervals, so we could use jackknife estimates for the standard errors. This model gave us a p-value of 0.0080 (the best since the time to first event model we originally ran).

The last method we tried for this dataset is WLW, which uses extended time gaps, essentially acting as if no previous event had ever occurred for each event, and looking at the length between time to entry and time to whatever numbered event is in question, for all observed events. The first one we used we folded the fourth and fifth event strata into the third, to simplify the model (basically giving first events, second events, and then higher order events.) Doing this gave us a p-value of 0.0080 again. Next we tried the same method, but this time with all five strata kept intact, and the p-value improved to 0.0037. Of all of the models we ran, the smallest p-value is on the original time to first event model. So it seems the drug is most effective at increasing the time to the first event. But, the significance of our other models (outside of the time to second event model) shows us that it is effective overall at improving time to events. Which model should be published depends on what the researchers

were most interesting in showing with this study. IF they were really only interested in improving time to first event, then just stick with the original model, since it's simpler, most significant, and has the best AIC of any of them. However, if they wanted to show that the drug helped improve time to recurrence for all events, then they should use one of the more robust models we ran, depending on what kind of interpretations and assumptions they want to make about the data. (Note: all code written for this part is included in the appendix)

Appendix

```
/*############################## Part 2: Sample Size Calculations ##############################*/


/*Problem 3*/

/*Part a*/
proc power;
twosamplesurvival
test=logrank
alpha=0.05
power=0.9
groupmedsurvtimes=(10 12.5)
accrualtime=0.1
followuptime=1000
npergroup=.;
run;

/*Part b*/
proc power;
twosamplesurvival
test=logrank
alpha=0.05
power=0.9
groupmedsurvtimes=(10 12.5)
accrualtime=6
followuptime=4
npergroup=.;
run;

/*Part c*/
proc power;
twosamplesurvival
test=logrank
alpha=0.05
power=0.9
groupmedsurvtimes=(10 12.5)
accrualtime=6
followuptime=4
grouplossexphazards=(0.0186 0.0186)
npergroup=.;
run;

/*Part d*/
proc power;
twosamplesurvival
test=logrank
```

```
alpha=0.05
power=0.9
groupmedsurvtimes=(10 12.5)
accrualtime=6
followuptime=4
grouplossexphazards=(0.0186 0.0186)
groupweights=(1 3)
ntotal=.;
run;

/*Problem 4*/

/*Part a*/
proc power;
twosamplesurvival
test=logrank
alpha=0.05
power=0.9
accrualtime=0.1
followuptime=1000
curve("Control")=(6):(0.8)
curve("Treatment")=(6):(0.9)
groupsurvival="Control"|"Treatment"
npergroup=.;
run;

proc power;
twosamplesurvival
test=gehan
alpha=0.05
power=0.9
accrualtime=0.1
followuptime=1000
curve("Control")=(6):(0.8)
curve("Treatment")=(6):(0.9)
groupsurvival="Control"|"Treatment"
npergroup=.;
run;

/*Part b*/
proc power;
twosamplesurvival
test=logrank
alpha=0.05
power=0.9
accrualtime=6
followuptime=4
curve("Control")=(6):(0.8)
curve("Treatment")=(6):(0.9)
groupsurvival="Control"|"Treatment"
grouplossexphazards=(0.0186 0.0186)
groupweights=(1 3)
ntotal=.;
```

```
run;

proc power;
twosamplesurvival
test=gehan
alpha=0.05
power=0.9
accrualtime=6
followuptime=4
curve("Control")=(6):(0.8)
curve("Treatment")=(6):(0.9)
groupsurvival="Control"|"Treatment"
grouplossexphazards=(0.0186 0.0186)
groupweights=(1 3)
ntotal=.;
run;

/*Problem 5*/

/*Part a*/
proc power;
twosamplesurvival
test=logrank
alpha=0.05
power=0.9
accrualtime=6
followuptime=4
curve("Control")=(5.5 7.6 8.7 10.1 14.3 18.2):(0.8 0.6 0.5 0.35 0.3 0.25)
refsurvival="Control"
hazardratio=0.75
grouplossexphazards=(0.0186 0.0186)
groupweights=(1 3)
ntotal=.;
run;

/*Part b*/
proc power;
twosamplesurvival
test=logrank
alpha=0.05
power=0.9
accrualtime=6
followuptime=4
curve("Control")=(5.5 7.6 8.7 10.1 14.3 18.2):(0.8 0.6 0.5 0.35 0.3 0.25)
curve("Treatment")=(5.5 7.6 8.7 10.1 14.3 18.2):(0.88 0.66 0.55 0.385 0.306 0.255)
groupsurvival="Control"|"Treatment"
grouplossexphazards=(0.0186 0.0186)
groupweights=(1 3)
ntotal=.;
run;
```

```
/*############################## Problem 8: Cystic Fibrosis ##############################*/

/*remaking the dataset*/
data mydata1;
set setone;
duration=time2-time1;
drop rx1 fev1 rx2 fev2 rx3 fev3;
run;

data mydata2;
set mydata1;
by id enum;
retain count 0 tstart tstop duration;
if first.id then
do;
count=0; tstart=time1; tstop=time2;
count=count+1; output;
end;
else
do;
tstart=time1; tstop=time2;
count=count+1; output;
end;
if (last.id and count<5) then
do
enum=(count+1) to 5; tstart=tstop; status=0; output;
end;
run;

data finaldata;
set mydata2;
diff=tstop-tstart;
if enum<=3 then enum3=enum;
else enum3=3;
drop duration count inst time1 time2 gap;
run;

/*Part a*/
proc phreg data=finaldata;
where enum=1;
class rx(ref="0");
model diff*status(0)=rx fev;
run;

/*Part b*/
proc phreg data=finaldata;
where enum=2 and diff ne 0;
class rx(ref="0");
model diff*status(0)=rx fev;
run;

/*Part c*/
```

```
/*A-G*/
proc phreg data=finaldata covs(aggregate);
class rx(ref="0");
model (tstart,tstop)*status(0)=rx fev;
where tstart<tstop;
id id;
run;

/*PWP-1*/
proc phreg data=finaldata covs(aggregate);
class rx(ref="0");
strata enum;
model (tstart,tstop)*status(0)=rx fev;
where tstart<tstop;
id id;
run;

/*PWP-2*/
proc surveyphreg data=finaldata varmethod=jackknife;
strata enum;
class rx(ref="0");
model diff*status(0)=rx fev;
cluster id;
where tstart<tstop;
run;

/*new dataset for WLW*/
data wlw;
set finaldata;
by id;
if first.id then newstop=tstop;
else newstop+diff;
run;

/*WLW-1*/
proc surveyphreg data=wlw varmethod=jackknife;
strata enum3;
class rx(ref="0");
model newstop*status(0)=rx fev;
cluster id;
run;

/*WLW-2*/
proc surveyphreg data=wlw varmethod=jackknife;
strata enum;
class rx(ref="0");
model newstop*status(0)=rx fev;
cluster id;
run;
```