

Biostatistics 209, Lab #3

1. Background

The purpose of this lab is to give you practice on checking proportional hazards assumption in Cox regression models.

2. Data

- Download the datasets `lab3-actg019_a.dta` and `lab3-pbc_a.dta` from the website (these are altered datasets from `actg019.dta` and `pbc.dta` that you used before!).

3. Checking the Cox Model for ZDV treatment in `lab3-actg019_a.dta`

Declare the data `lab3-actg019_a.dta` to be survival time data
`stset days, failure(cens)`

- a. Generate the Cox-KM plot and the log-minus-log KM plot by treatment (rx).

```
stcoxkm, by(rx) obslopts(recast(line) connect(stairstep) lcolor(blue))  
obs2opts(recast(line) c(stairstep) lc(red)) predlopts(recast(line)   
c(stairstep) lc(green)) pred2opts(recast(line) c(stairstep) lc(orange))
```

```
stphplot, by(rx) nonnegative nolntime plotlopts(recast(line)   
connect(stairstep)) plot2opts(recast(line) c(stairstep)) ytitle(Log Minus  
Log Survival) xtitle(Time) legend(order(1 "ZDV" 2 "Placebo"))  
scheme(sicolor)
```

Some options used to customize the graph:

<code>recast(line)</code>	gives no symbols at the estimate points.
<code>connect(stairstep)</code>	asks Stata to connect points using a step function.
<code>lcolor(red)</code>	specifies the line color.
<code>nolntime</code>	prevents Stata from taking the log of the x-axis,
<code>nonnegative</code>	specifies the log(-log) not -log(-log) plots,

- b. Does the rx HR appear proportional?
- c. Graph the log hazard ratio after fitting the Cox model and save the scaled Schoenfeld residuals. These are needed for subsequent plots and calculations. Note, by default, ZDV is set as the reference group for the variable rx.

```
stcox rx, sca(giveAname) (choose a name to be used later; no "*" b/c only 1 var)  
estat phtest, plot(rx) (this gives a running mean smoother, rather than lowess)
```

- d. Re-graph the log hazard ratio with a lowess smoother, and save the lowess values in the variable named "`smloghr`" (obviously you can call it something else).

```
lowess giveAname days, generate(smloghr) ytitle(rx Log Hazard Ratio)
```

- e. **(Question 3.1)** Look at the graph and the variable `smloghr` and fill in the following table.

```
gen smhr=exp(smloghr)
sort days
list days smloghr smhr if cens==1 & (days==95|days==181|days==362|days==540)
```

	log HR	HR
95 days (3 mos)		
181 days (6 mos)		
362 days (12 mos)		
540 days (18 mos)		

- f. **(Question 3.2)** Run the Schoenfeld test for the proportional hazards assumption. Is there evidence against that? Does this agree with the plots?

```
estat phtest, detail
```

- g. How would you summarize the effect of ZDV on progression of HIV?
- h. Consider the “stratification” approach to dealing with non-proportional hazards: run the log-rank test to conclude that the effect of ZDV is statistically significant and present the K-M plot to show its effect on free of HIV progression. Note that this approach does not provide a summary estimate of the ZDV effect (the primary predictor of the study). Also, why do we not use `stcox` here?
- i. **(Questions 3.3-3.4)** Explore the time-dependent covariate approach.

```
stset days, failure(cens) id(id) // define survival data with multiple
// observations per subject
stssplit grp, at(365) // split time variable (days) at 1 year (grp) that is,
// generate multiple rows for each subject; one for each timepoint up to
// and including the time of censoring or time of death
recode cens .=0 // recodes all newly generated rows to “censored” status
gen rx01=rx*(grp==0) // This set of commands generates 2 separate
gen rx1p=rx*(grp==365) // rx variables specific to each time interval;
// that is, rxxx only equals 1 if the patient is on placebo and the dataset row
// corresponds to period xx
sort id grp // sort data so prioritized by id and then by grp within id
list id _t0 _t cens rx grp rx01 rx1p in 1/10, sepby(id)
// take a look that the data seems ok
stcox rx??, nolog // fit Cox model
```

4. Checking the Cox Model for Cholesterol in lab3-pbc_a.dta

Declare the data lab3-pbc_a.dta to be survival time data

```
stset years, failure(status)
```

- a. **(Question 4.1)** Fit the Cox model for the effect of cholesterol and run the Schoenfeld test for the proportional hazards assumption. Does the evidence appear against the assumption?

```
stcox chol, sca(giveAname)      (Did you give the same name as before?)
estat phtest, detail
```

- b. Graph the log hazard ratio. What does the graph suggest? Do you have any concerns about the test?

```
lowess giveAname years, ytitle(cholesterol Log Hazard Ratio)
```

- c. **(Question 4.2)** Try excluding some potential influential points and then re-run the plot and test for the proportional hazards assumption. What do you conclude?

```
drop giveAname
stcox chol if years <= 12, sca(giveAname)
```