

Biostatistics Lab # 14, 4/14/05

0: BACKGROUND

Purpose: The purpose of this lab is to explore interaction and non-proportional effects in the Cox model.

Data: Two arms of ACTG 019 a study which compared zidovudine (AZT, ZDV) to placebo in a sample of asymptomatic HIV-infected subjects with CD4 count < 500 cells/mm³. (Volberding, PA, Lagakos, SW, Koch, MA et al. New Engl J Med 322:14 941-9). The failure outcome is time from enrollment to the development of an AIDS-defining infection or death (in days).

Event Time: days *Time to AIDS*

Event Type: cens (*0 Censoring, 1 AIDS*)

Predictors:

rx	Randomized Treatment	(1=ZDV,0=Placebo),
cd4	Baseline CD4 Count	(in cells per mm3)
strcd4	CD4 stratum	(0: CD4 <300 at baseline, 1: CD4=>300 at baseline)

1: DATA

- Download lab14-actg019.dta from the website <http://www.biostat.ucsf.edu/biostat209>
- 6 variables: id, days, cens, rx, cd4, strcd4
- Declare the data to be a `stset`
`stset days, failure(cens)`

2. COX MODELS

- Using the `sts graph` command to make Kaplan-Meier curves for the two treatment groups
`sts graph, by(rx)`
- You may find you like looking at the curves data better with the `failure` option.
- Does there appear to be a treatment effect? Try
`sts test rx`
- Fit Cox model with treatment:
`stcox rx`
- Write a sentence, as you would for a medical journal, summarizing the effect of ZDV treatment on HIV disease progression in these subjects.
- It is possible to calculate the HR with the ZDV arm as the comparison group using
`lincom -1*rx, hr`
- How would you write the sentence with ZDV subjects as the comparison group? Which sentence to you prefer?
- Fit Cox model with treatment and continuous CD4 count
`stcox rx cd4`
- What is estimated hazard ratio between the treatments? Does it differ from the unadjusted?
- Try adjusting by stratification on `strcd4`. This was the primary analysis done by the authors.
`stcox rx, strata(strcd4)`

3. INTERACTION

- Use an `if` statement to graph the Kaplan-Meier to compare the treatments `rx` in the high and then in the low CD4 strata.
`sts graph if strcd4==0, by(rx)`
`sts graph if strcd4==1, by(rx)`
- Based on the graphs, do you see evidence of interaction?
- Using the `if` statement, fit separate Cox models for `rx` for the lower and higher CD4 strata. Does this suggest interaction?
`stcox rx if strcd4==0`
`stcox rx if strcd4==1`

- Generate a product term and fit a Cox model for interaction. Is there a significant interaction?
- Based on this model, what are the hazard ratios for ZDV treatment in the CD4 strata?

4. CHECKING PROPORTIONALITY

Save the full dataset

```
preserve
```

Drop the high CD4 stratum

```
drop if strcd4==1
```

- Plot the log-minus-log KM plot by ZDV treatment
`stphplot, by(rx)`
- To make the plots like the ones in lecture, add the options (after the `by`)
 `nolntime nonegative recast(line) connect(stairstep)`
`nolntime` prevents Stata from taking the log of the x-axis,
`nonegative` specifies the `log(-log)` not `-log(-log)` plots,
`connect(stairstep)` asks Stata to connect points using a step function.
Finally, `recast(line)` gives no symbols at the transformed KM points.
- Does the `rx` effect appear proportional?
- Fit the Cox model and save the `schoenfeld` and `scaled schoenfeld` residuals. These are needed for subsequent calculations.
`stcox rx, sch(o*) sca(w*)`
- The Cox model was fit. And 2 new variables were created (`o1,w1`). The naming was taken from the `sch` and `sca` statements. The naming of the residuals is up to you.
- Test for proportional hazards
`stphtest, detail`
- The null hypothesis is that the variable obeys proportional hazards. Is there evidence against that? Does this agree with the plots?

5. SMOOTHING THE (LOG) HAZARD RATIO

It is possible to use the "scaled Schoenfeld residuals" to get a non-parametric estimate of the (natural) log-hazard ratio.

This is done simply by obtaining a lowess smooth of the residuals against time. The scaled Schoenfeld residuals are in the variable `w1`. Time is the variable `days`. The plot of the log hazard ratio is given by

```
twoway (lowess w1 days, sort), ytitle("Log Hazard Ratio") xtitle(Days)
```

The Y-axis is the log of the hazard ratio. If the data satisfied proportional hazards, then the smooth line should be flat. That would reflect a (log) hazard ratio that doesn't change with time. A negative log hazard ratio reflects a hazard ratio which is less than 1 (ZDV is protective). A positive hazard ratio occurs when progression is more frequent on ZDV.

Use the `display` command to obtain the hazard ratios associated with the log hazard ratio values of -2, -1, 0 and 0.5.

```
display exp(-2)
display exp(-1)
display exp(0)
display exp(0.5)
```

At what times does the curve seem to cross -2, -1, 0 and 0.5, respectively? What does this suggest about the effect of ZDV treatment with time?