

Math 150 - Methods in Biostatistics - Homework 10

your name here

Wednesday, April 12, 2023

Assignment Summary (Goals)

- building Cox PH models
- checking proportional hazards
- interpreting coefficients in the Cox PH model

Important

The data are in the files tab on Canvas (in a folder called “data”).

Q1. Collaborative Learning Describe one thing you learned from someone (a fellow student or mentor) in our class this week (it could be: content, logistical help, background material, R information, etc.) 1-3 sentences.

Q2. Chp 9, E11 Read text for description of data.

- check proportional hazard assumptions for treatment KM curves. Use `fun="cloglog"` inside the `ggsurvplot`. (We will cover checking PH in Tuesday’s class.)
- use all explanatory variables and likelihood ratio test to come up with the “best” model. [Note that the structure of the likelihood ratio test here is identical to the one we used in logistic regression! See: <http://st47s.com/Math150/Notes/survival-analysis.html#multcoxph>]
 - After you pipe the `coxph()` model into `glance()` you will see a column called `logLik`.
 - The test statistics is $G = 2 * (\log Lik_{biggermodel} - \log Lik_{smallermodel})$
 - `Type` and `Partners` are related and probably should not be thought of as numeric (instead they should be factor variables). That said, there are not 9 unique groups (3x3) so to use both, the six groups need to be defined:

```
fruitfly <- read_csv("https://pomona.box.com/shared/static/qnsl0sp0twdutz6azidxb5yt37boee7v", na="*") %>%
  mutate(Type_Partners = case_when(
    Partners == 0 ~ "0zero",
    Partners == 1 & Type == 0 ~ "1P0T",
    Partners == 8 & Type == 0 ~ "8P0T",
    Partners == 1 & Type == 1 ~ "1P1T",
    Partners == 8 & Type == 1 ~ "8P1T"
  ))
```

- The p-value will be calculated using a chisq distribution where the degrees of freedom are the number of extra coefficients which were estimated in the bigger model.
- using the final model, interpret each of the coefficients (in terms of hazard ratios). Don’t forget that when a model has multiple variables, the coefficient estimate will be interpreted while keeping all other variables constant.

Q3. Chp 9, E12 Read text for description of data.

- (a) check proportional hazard assumptions for treatment KM curves. Use `fun="cloglog"` inside the `ggsurvplot`. And/or use `cox.zph`. (We will cover checking PH in Tuesday's class.)
- (b) use all explanatory variables and likelihood ratio test to come up with the "best" model.
- (c) using the final model, interpret each of the coefficients (in terms of hazard ratios). Don't forget that when a model has multiple variables, the coefficient estimate will be interpreted while keeping all other variables constant.

```
VAlung <- read_csv("https://pomona.box.com/shared/static/r6hoo1gawopkt0526xvwz5f13245de",  
                  na="*")
```

```
praise()
```

```
## [1] "You are superior!"
```