Math 150 - Methods in Biostatistics - Homework 9

your name here

Wednesday, April 5, 2023

Assignment Summary (Goals)

- working with hazard functions as measures of survival (S(t) and h(t) are functions of each other!)
- working with cumulative hazard functions

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, E12 VA Lung Cancer Study (Lots to read in the text about the dataset.)

(a) Create a graph with both Kaplan-Meier curves to compare the survival time (use the variable `time`) for subjects with the standard and the test chemotherapy treatment. What do you observe about the survival probabilities for the groups of subjects?

(b) Conduct the log-rank test and the Wilcoxon test to compare the survival curves of both treatment groups. Interpret the results.

(c) It may be beneficial to incorporate health as a variable in the analysis. Patients with low Karnofsky scores are less healthy than patients with high Karnofsky scores. Create four groups with the `Veteran` data: `trt=1` and Karnofsky score low, `trt=1` and Karnofsky score high, `trt=2` and Karnofsky score low, and `trt=2` and Karnofsky score high. Recall that it is often best to keep sample sizes as equivalent as possible when you determine what is a low or high Karnofsky score. Create a Kaplan-Meier curve for each of the four groups. Conduct the log-rank test and the Wilcoxon test to compare the survival curves of the four groups. (While we have only discussed using these tests to compare two groups, they can easily be extended to more than two groups.) Did incorporating health into your analysis impact your conclusions? [The R syntax works like other modeling we have done, just add the explanatory variables after the tilde: `~ trt + karno2`.]

```
VAlung <- read_csv("https://pomona.box.com/shared/static/r6hoo1gawopkt0526xvwwze5fl3245de",
                   na="*") %>%
  mutate(karno2 = ifelse(karno <= 60, "low", "high"))
```

Q3. Chp 9, A45 I've included the R code to create a hazard curve (the R code came with your text book and is on Canvas). Note, however, as discussed, the hazard rate is extremely sensitive to each time interval. In lieu of looking at the hazard curve, it is often more informative to look at the cumulative hazard curve (see section 9.9). The values of the estimated hazard function can be seen in the cumulative hazard curve as the jumps at each time event.

You may use the code below (the function, below, is called `plot.haz()`) or you can use the code (see R code in the class notes, set `fun="cumhaz"`) to plot the cumulative hazard function using `ggsurvplot()`.

Use the software instructions provided to plot the estimated hazard rates for the college graduation data (see page 311).

```
plot.haz <- function(KM.obj,plot="TRUE") {
  ti <- summary(KM.obj)$time
  di <- summary(KM.obj)$n.event
  ni <- summary(KM.obj)$n.risk

  #Est Hazard Function
  est.haz <- 1:(length(ti))
  for (i in 1:(length(ti)-1))
    est.haz[i] <- di[i]/(ni[i]*(ti[i+1]-ti[i]))
  est.haz[length(ti)] <- est.haz[length(ti)-1]

  if (plot=="TRUE") {
    plot(ti,est.haz,type="s",xlab="Time", ylab="Hazard Rate",
         main=expression(paste(hat(h),(t)[KM])))
    }
  #return(list(est.haz=est.haz,time=ti))
}
```

```
grad <- read_csv("https://pomona.box.com/shared/static/yigpp4e8dvkyw9pf3f0c7o9nr0rt3k6m", na="*")
```

Q4. Chp 9, A46 Although the estimated hazard curve may not exhibit a distinguishable pattern, discuss some important features of the curve (see pg 311).

Q5. Chp 9, A47 Indicate periods of time during their college career when students are at their lowest and highest risk of graduating college. Does your answer match your common understanding of when students typically graduate from college? (see pg 311)

Q6. Chp 9, E9 Sketch hazard functions that would correspond to the following time-to-event random variables (You may want to do a little background research.)

(a) Lifetime of an individual measured from birth (don't assume anything about the health or demographics of this person).

(b) Time until death after surgery to remove a cancerous tumor.

Be sure to label the time axis, and mark time points appropriately. Briefly explain your reasons for any changes in the shape of the hazard function over time.

Q7. Chp 9, E10 The graphs displayed in Figure 9.19 (see pg 325) are population cumulative hazard functions for three distributions of the time-to-event random variable, T. For each one, sketch a possible corresponding hazard function h(t). Be sure to label the same time points on your sketches as are provided on the graphs of H(t).

```
praise()
```

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