Homework 5

Xinyu Zhang

Due date: Tuesday, December 11

This homework consists of questions proposed by the presenters. You are free to choose any five items that include at least one from each presenter.

1. Diagnostics for the Cox Model by **Tianjian Shi**:

   a. Write down the Schoenfeld residuals formula and briefly explain how it works.

   From notes we know that:

$$r_i = \Delta_i \left[X_i - \left\{ \frac{\sum_{j \in R(t_i)} X_j e^{X_j \hat{\beta}}}{\sum_{j \in R(t_i)} e^{X_j \hat{\beta}}} \right\} \right]$$

   This is same as the presentation:

$$r_k(\widehat{\beta}) = X_{(k)} - E(X_{(k)} \mid R_k)$$

   for uncensored observations and 0 for censored observations.

   Schoenfeld residuals are the individual terms of the score function in the maximum likelihood calculation, so they sum up to 0. Also the smaller each of them is, the better model it is. When the proportional hazard assumption holds, a plot of Schoenfeld residuals against event times will approximately scatter around 0.

   b. Describe the idea that Therneau et al. (1990) proposed to use martingale residuals in investigating the functional form of continuous covariate.

   When the covariate needs to be investigated is uncorrelated with other covariates, we know if we fit a model with only the other covariates then the residuals should be the only right side term that correlates with the investigating covariate. To move the appropriate effect from the residuals to the model, we plot the martingale residuals against the investigating covariate, determine their relationship then remove it.

2. Survival analysis through the eyes of a molecular biologist by **Kelly Daescu**:

   a. List the cons and pros for both the `ggplot2` and `ploty` packages.

   'ggplot2' is used by adding layers to the plot so it is easy to understand. 'plotly' is an interactive graph while 'ggplot2' is static. Of course we can use 'ggplot2' in a shiny app to make it dynamic, but its speed is reported slower than the web-based 'plotly' graphs when a few plots need to be generated at the same time.

   b. Interpret the Nelson-Aalen estimators presented on page 7.

   Both female and male have very close Nelson-Aalen cumulative hazard function before 10. After 10, male's hazard function behaves similar to what's before 10 while female's hazard function remains unchanged until 13 then increases rapidly and surpasses male at 15. The dip between 10 and 15 in the female's hazard function might because a local regression smoothing is used.

3. A brief introduction to `R Shiny` by **Jose Alfaro**:

   a. What is the purpose of the UI function in a Shiny app?

Copied from the presentation:

1. Controls the layout and appearance of the app through
2. Allows users to add text, images, and other HTML elements to the Shiny app and specify their locations
3. Also gives users the option to create tabs/panels to organize visuals and widgets

b. List three widgets and their uses.

Copied from the presentation:

1. Action Buttons - Creates a button with an initial value of 0 which increments by one each time it is pressed
2. Sliders - Used to select single values within a specified range (may be used to create intervals)
3. Check Boxes - Used to create checkboxes which specify logical values

4. Selecting a threshold for long-term survival probabilities and Kaplan-Meier estimator by **Patrick Thompson**:

   Suppose that in a laboratory experiment 10 mice are exposed to carcinogens. The experimenter decides to terminate the study after half of the mice are dead and to sacrifice the other half at that time. The survival times of the five expired mice are 4, 5, 8, 9, and 10 weeks. The survival data of the 10 mice are 4, 5, 8, 9, 10, 10^+, 10^+, 10^+, 10^+, and 10^+. Assuming that the failure of these mice follows an exponential distribution, estimate the survival rate λ and mean survival time μ.

$$\widehat{\lambda} = \frac{r}{\sum_{i=1}^{r} t_{(i)} + \sum_{i=r+1}^{n} t_{(i)}^{+}} = \frac{5}{4+5+8+9+6\times 10} = 0.0581$$

$$\widehat{\mu} = \frac{1}{\widehat{\lambda}} = \frac{1}{0.0581} = 17.2$$