Exam 1

Steven Chiou

Instructions

- Due date: **Tuesday, November 13**.
- This is a open resource exam, but you are not allowed to ask post exam questions online.
- You are not allowed to collaborate with classmates and/or people outside of class.
- Please circle or highlight your final answer.
- The toal possible point is 80.

Violation of this agreement will result in an **F** on this exam and it will be averaged in as a 0%.

---

1. Use the complete *WHAS100* dataset and gender as the group indicator, compute the log-rank statistic, Q, presented as Equation (9) on page 60 of note 2 with $\omega_i = 1$. Use Q to compute a p-value to test the null hypothesis of $H_o : S_0(t) = S_1(t)$. Do this **without a software package** (6 pts), and verify the calculation with `survdiff` (4 pts).

2. There are many ways to form a basis for survival curve comparison. Here are some:
   a. The numerator term in Q without the square:

$$D_1 = \sum_{i=1}^{D} \{d_{1i} - E(d_{1i})\}.$$

   b. The largest distance between the two curves:

$$D_2 = \max |S_1(t) - S_0(t)|.$$

   c. The difference between the median survival times:

$$D_3 = S_1^{-1}(0.5) - S_0^{-1}(0.5)$$

   d. The difference between the mean survival times:

$$D_4 = \int_0^{t_{(n)}} \{S_1(u) - S_0(u)\}\, du,$$

   where $t_{(n)}$ is the maximum observed survival time.

   Compute each of the above statistic for the $WHAS100$ dataset (5 pts $\times 4$).

3. The statistics computed in (2) do not provide meaningful interpretations when standing along. We will use a permutation approach to test for the null hypothesis of $H_o : S_0(t) = S_1(t)$ based on these statistics. The idea of a permutation test is simple. The general procedure can be summarized into the following steps:
   i. Compute the desired statistic based on the observed data; we will call this the observed statistic.
   ii. Permute the data under the null.
   iii. Compute the statistics for each possible permutation in Step ii.; we will call these permutation statistics.
   iv. Draw conclusion based on where the observed statistic stands among the permutation statistics.

   The statistics we computed in (2) are the observed statistics in Step i. If the null hypothesis of $H_o : S_0(t) = S_1(t)$ is true, then one can randomly shuffle the group indicator to generate different permutations (Step ii) and the statistics for these permutations should be similar (Step iii).

   a. (5 pts×4) Generate 5000 permutation and, for each of the permutation, compute the four statistics presented in (2). We will call the permutated statistics D^*_{1i}, D^*_{2i}, D^*_{3i} and D^*_{4i} for $i = 1, \ldots, 5000$. Create a histgram for these permutated statistics and print the `summary`.
   b. (5 pts×4) Compute the p-value based on these statistics by

$$p = 2 \cdot \frac{\min(N_1, N_2)}{5000},$$

   where $N_1 = \#\{D \geq D^*\}$, $N_2 = \#\{D \leq D^*\}$, and # means the "number of", e.g., N_1 is the number of these permutated statistics less than or equal to the observed statistic.

4. Another method to compare two survival curves is to consider a sign test. Suppose we have two groups of uncensored survival times:

   Males: $x_1, x_2, \ldots, x_{n_0}$.

   Females: $y_1, y_2, \ldots, y_{n_1}$.

   The sign test looks at the statistic

   $$U = \sum_{i=1}^{n_0} \sum_{j=1}^{n_1} \text{sgn}(x_i - y_j),$$

   where sgn$(\cdot)$ is the sign function. In the pretense of right censoring, survival times can not be compared directly and a modified version of $U = \sum_{i=1}^{n_0} \sum_{j=1}^{n_1} U_{ij}$ is considered, where

   $$U_{ij} = \begin{cases} 1 & \text{if } x_i > y_j, y_j \text{ is uncersored.} \\ -1 & \text{if } x_i < y_j, x_i \text{ is uncersored.} \\ 0 & \text{otherwise.} \end{cases}$$

   a. (5 pts) Compute U for the WHAS100 dataset.
   b. (5 pts) Create a histgram for these permutated statistics and print the `summary`. Obtain a permutation p-value based on 5000 permutations.