PubH 7450 Homework 2 (Spring 2023) Due date: March 04, 2023

Steffen Ventz

2023-02-23

Note: Please use an Rmarkdown document (Rmd and pdf) which includes your R code; show major steps in your derivations/calculations.

1) [23 Points]

Consider a (possible right-censored) dataset (T_i, δ_i) of size n, with $D \leq n$ unique event times $t_1 < \cdots t_D$. Let d_j, y_j be number of events and observations at risk for an event at event time $t_j, j = 1, \ldots, D$.

(i) [3 Points] You are given the KM-estimator $\hat{S}(t_j) = 0.82$ and $d_j = 2, y_j = 24$. Is the information sufficent to compute the (asymptotic) Variance $\widehat{Var}[\hat{S}(t_j)]$ for the estimator $\hat{S}(t_j)$? If yes, provide $\widehat{Var}[\hat{S}(t_j)]$, if no, why?

(ii)[10 Points] Given $\hat{S}(t_j)$ and $\widehat{Var}[\hat{S}(t_j)]$, use these two quantities to derive an approximate 95% confidence interval for $S(t_j)$ at the fixed time point t_j, based on the logit-transformation $T(S) = \log \frac{S}{1-S}$ for $S \in (0, 1)$. (Note: you will need to determine the variance of $T(\hat{S}(t_j))$ and confidence interval for $T(\hat{S}(t_j))$. Also note that the inverse of $y = T(S)$ is $T^{-1}(y) = 1/(1 + \exp\{-y\}))$.

(iii)[5 Points] Assume that the dataset contains also a binary variable $X \in \{0, 1\}$, for instance experimental vs control treatment. We estimate a KM for both groups and obtain estimates $\hat{S}_k(t_j), \widehat{Var}[\hat{S}_k(t_j)]$ for groups $k = 0, 1$. Suggest a two-sample two-sided hypothesis test for $H_0 : S_1(t_j) = S_0(t_j)$.

(iv)[5 Points] How would you compute a 95% confidence interval for the difference $S_1(t_j) - S_0(t_j)$? How is that different from the hypothesis test in (iii)?

2) [21 Points]

The data `kidney'' in the R package`KMsurv'' (see also section 1.4) report the times to first exit site infection (in months) of patients with renal insufficiency was reported. In the study 43 patients had a surgically placed catheter (Group 1) and 76 patients had a percutaneous placement of their catheter (Group 0).

(i) [3 Points] For each group plot the estimated survival function. Which technique seems better in delaying the time to infection?

(ii) [5 Points] Estimate the cumulative hazard rate for each group of patients. Provide a crude estimate of the hazard rate at 5 months after placement of the catheter in each group.

(iii) [5 Points] Find a 95% confidence interval for the mean time to first exit site infection restricted to 36 months for both groups.

(iv) [5 Points] Suggest and conduct a formal two-sample two-sided hypothesis test for the mean time to the first exit site infection restricted to 36 months for both groups. State the null hypothesis. Define an appropriate test statistics, and state the null hypothesis under the null hypothesis (Note: Think of this problem as an application of a two-sample z-test).

(v) [3 Points] Compute and plot 95% HW confidence bands for the survival functions of both groups.

3) [18 Points]

The data `kidtran'' in the R package`KMsurv'' (see also section 1.7) report survival times of 280 white females given a kidney transplant at the OSU transplant center.

(i) [3 Points] Estimate the distribution of the time to death, measured from transplant, for white female kidney transplant patients.

(ii) [3 Points] Find a 95% confidence interval, based on the linear transformation, for the probability a white female will survive at least 12 months (365 days) after transplantation.

(iii) [2 Points] Repeat ii using the log-transformed confidence interval.

(iv) [1 Points] Compare the intervals found in parts i–ii.

(v) [3 Points] Compute a 95% equal probability confidence band for the survival function of white females. Describe the results.

(vi) [3 Points] Find a 95% confidence interval for the mean survival time for white males and females restricted to 60 months (1825 days).

(vii) [3 Points] Find a 95% confidence interval for the difference in the mean survival time for white males and females (1825 days).