EMORY ROLLINS SCHOOL OF PUBLIC HEALTH | Department of Biostatistics and Bioinformatics

BIOS 522: Survival Analysis Methods

Homework - Weeks 6-8

Problem 1. Interaction model (1.5 points)

Suppose $h_i(t) = h_0(t)\exp(\beta_1 X_{1i} + \beta_2 X_{2i} + \beta_3 X_{3i})$ where $X_{1i} = 1$ for treated and $X_{1i} = 0$ for untreated, $X_{2i} = 1$ for older adults and $X_{2i} = 0$ for younger adults, and $X_{3i} = X_{1i} \times X_{2i}$ is their interaction.

What values of $\beta_1, \beta_2, \beta_3$ correspond to:

a. (0.5 points) Treatment hazard ratio is the same in older adults and younger adults.

$$\beta_3 = 0$$

b. (0.5 points) No treatment effect in older adults, but an effect in younger adults.

$$\beta_3 = -\beta_1$$

c. (0.5 points) No treatment effect.

$$\beta_1 = \beta_3 = 0$$

Problem 2. Examining proportional hazards (3.5 points)

This problem centers on Mazzaferro et al. (2009) *Lancet Oncology*; DOI:10.1016/S1470-2045(08)70284-5. Background is provided on the study.

Goal: Investigators sought to develop a prognostic model of survival for patients undergoing liver transplantation for certain types of hepatocellular carcinoma. In particular, they wanted to extend the prognostic model to patients who exceed the existing "Milan criteria" for liver transplantation (sicker patients).

Population: The study included 1556 patients who underwent liver transplantation for hepatocellular carcinoma at 36 centers in Europe, America, and Asia between 1984 and 2006. Data were collected retrospectively by a web-based survey completed by specialists from each center.

Outcome variable: The primary outcome was time from liver transplantation until death from any cause. Time was censored at the date of last follow-up assessment for patients who were still alive.

Covariates: Variables considered are size of the largest tumor in mm, number of tumors, and presence/absence of microvascular invasion. Because data were sparse when tumor size was more than 150 mm or number of tumors was greater than 15, higher values were truncated at these thresholds.

The authors fit a model using only the size of the largest tumor and the number of tumors as covariates. This model, which does not include microvascular invasion, is described in Figure 1, Table 2, and Figure 2A. The following questions will refer to this model only. To answer these questions, refer to the original paper posted on Canvas.

a. (0.5 points) The authors used a selection procedure to determine the functional form of the model. Report the types of terms included in the final model (e.g. linear, spline, interactions).

The model that uses only the size of the largest tumor and number of tumors has the form:

- Linear term for size of the largest tumor (truncated at 150mm)
- Cubic spline for number of tumors (truncated at 15)
- Linear-by-linear interaction for size-by-number

b. (0.5 points) What is the tumor size and number of tumors for the model's reference group?

3 tumors, largest tumor 35 mm.

c. (0.5 points) How was this reference group selected?

These are the median values.

d. (0.5 points) Why are three different hazard ratios reported in Table 2 for the effect of the size of the largest tumor on survival?

There is an interaction with number of tumors. The effect size varies by number of tumors, with a slightly larger hazard ratio for the effect of tumor size when there is only one tumor than when there are five tumors.

e. (0.5 points) Consider two patients in the study population. Patient A has 5 tumors, and the largest tumor is 50 mm. Patient B has 5 tumors, and the largest tumor is 22 mm. What is the hazard ratio comparing Patient A's survival to Patient B's survival?

1.30

f. (0.5 points) Describe the shape of the log hazard ratio (log relative hazard) for number of tumors when largest tumor size is fixed at 35 mm. Is the change in hazard for 1 → 2 tumors similar to the change in hazard for 9 → 10 tumors?

Hazard increases with number of tumors, but it increases fastest at below 3 tumors and then slows down. Thus, the change in hazard for 1 to 2 tumors is GREATER than the change in hazard for 9 to 10 tumors. Tumor number has a more dominant effect when there are fewer tumors overall.

g. What is the predicted 5-year survival for a patient with 1 tumor that is 65 mm large? Reporting a range is sufficient.

55-60% (uses Figure 2A).

Problem 3. WHAS data set (5 points)

Recall the Week 7 Computing Activity using the WHAS data set. Imagine you are working with a clinician collaborator who generated these data. The collaborator has asked you to prepare a preliminary report, with some basic plots and a fitted model. Submit your fitted model and report hazard ratios and 95% confidence intervals for the retained covariates.

For full points:

- (1 point) Report hazard ratios and 95% confidence intervals for the retained covariates;
- (1 point) Describe your model selection procedure and justify your choices;
- (1 point) Provide at least one plot used to assess the functional form of a covariate (e.g. a residual plot);
- (1 point) Provide at least one plot used to assess the proportional hazards assumption; and
- (1 point) Prepare this in a neat summary with code attached.