BIST P8110: Applied Regression II

11 - Cox Models:
Confounding, Effect Modification, and Model Comparison

Qixuan Chen

Department of Biostatistics
Columbia University

Assessment of Confounding

- Suppose we fit two models:

$$h(t, x_{i1}) = h_0(t)\exp(\beta_1 x_{i1})$$

  and

$$h(t, x_{i1}, x_{i2}) = h_0(t)\exp(\beta_1 x_{i1} + \beta_2 x_{i2})$$

- If X_2 is a confounder, β_1 in the two models differ. If they are about the same, we conclude that X_2 does not appear to be a confounder.

Assessment of Confounding

- The following index is usually used to check the amount of confounding

$$\Delta\hat{\beta}\% = 100\frac{\hat{\theta} - \hat{\beta}}{\hat{\beta}}$$

  - $\hat{\theta}$ denotes the estimate from the model that does not contain the potential confounder (the smaller model)
  - $\hat{\beta}$ denotes the estimate from the model that does include the potential confounder (the larger model)
  - The criterion for $\Delta\hat{\beta}\%$ is more driven by research experience than statistical thresholds. If no clinical guideline is available, an absolute value of 20 percent is used (page 134, Hosmer et al. 2008).

Example 4: Control of Confounding

- **EX4**: Fit a Cox model to estimate gender effect ($d = 1$ for female, 0 for male) by controlling for continuous age x
  - Model: $h(t, d, x, \beta) = h_0(t)\exp(\beta_1 d + \beta_2 x)$
  - $\text{HR}(t, d(1 \text{ vs } 0), x) = \frac{h(t, d=1, x)}{h(t, d=0, x)} = e^{\beta_1}$
  - $\text{HR}(t, d, 5\text{-yr}\uparrow) = \frac{h(t, d, x=a+5)}{h(t, d, x=a)} = e^{5\beta_2}$

Example 4: Control of Confounding

- **EX4**: Interpretation of HR
  - $\widehat{\text{HR}}(t, d(1\text{vs } 0), x) = e^{\hat{\beta}_1}$: The death rate among females is $e^{\hat{\beta}_1}$ times of that among males given age.
  - $\widehat{\text{HR}}(t, d = 1, 5\text{-yr} \uparrow) = e^{5\hat{\beta}_2}$: The death rate increases by $(e^{5\hat{\beta}_2} - 1) \times 100$ percent for every 5-year increase in age given gender.
  - The model with gender only $h(t, d, \theta) = h_0(t)e^{\theta_1 d}$ is usually referred to as the **crude model**, and the model with control of confounding $h(t, d, x, \beta) = h_0(t)e^{\beta_1 d + \beta_2 x}$ is usually referred to as the **adjusted model**.

Example 4: Control of Confounding

- **EX4**: Does age appear to confound the relationship between gender and survival?
  - The coefficients $\hat{\beta}_1$ for gender in the crude model and the adjusted model are estimated, respectively, as 0.555 and 0.149.
  - $\Delta\hat{\beta}\% = 100\frac{0.555-0.149}{0.149} = 272.4\%$
  - Since including age in the model has huge impact on the coefficient for gender, it does appear that age confounds the relationship between gender and survival.

Including Interaction Terms in the Model

- An interaction term effect (effect modification) exists between two variables if the effect of variable A varied by values of variable B. Variable B is called the effect modifier.
- To model the interaction, the main effects as well as their product must be included in the model.

Including Interaction Terms in the Model

- Once we make the decision that there is evidence that a covariate is an effect modifier, discussion of its role as a confounder is no longer relevant. The model with interaction terms is usually referred to as the **interaction model**.
- In real data analysis, we only keep interaction terms that are statistically significant. However, a confounder with a big change in the estimated coefficient can be clinically significant even if it is not statistically significant. In this case, this confounder is usually kept in the model.

Example 5: Interaction Terms

- **EX5**: If gender effect varies with age, then age is said to be an effect modifier. The simplest way to determine whether x is an effect modifier is to include the product term $x \times d$ in the model.
  - Model: $h(t, d, x, \beta) = h_0(t)\exp(\beta_1 d + \beta_2 x + \beta_3 dx)$
  - $\text{HR}(t, d(1 \text{ vs } 0), x = a) = e^{\beta_1 + \beta_3 a}$
  - $\text{HR}(t, d(1 \text{ vs } 0), x = a + 5) = e^{\beta_1 + \beta_3 (a+5)}$
  - $\text{HR}(t, d = 1, x(a + 5 \text{ vs } a)) = e^{5\beta_2 + 5\beta_3}$
  - $\text{HR}(t, d = 0, x(a + 5 \text{ vs } a)) = e^{5\beta_2}$

Example 5: Interaction Terms

- **EX5**: Test the significance of the interaction term.
  - **Hypothesis**: $H_0 : \beta_3 = 0$ vs. $H_\alpha : \beta_3 \neq 0$
  - **Wald test**: $z = \frac{\hat{\beta}_3}{\widehat{SE}(\hat{\beta}_3)} = \frac{-0.042}{0.024} = -1.75$
  - **p-value**: $Pr(|Z| \geq |-1.75|) = 0.08 > 0.05$
  - **Rejection rule**: fail to reject H_0 at $\alpha = 0.05$
  - **Conclusion**: there is insufficient evidence to conclude that age modifies the effect between gender and survival.

Model Selection in the Cox Model

Testing Significance of a Cox Model

- Test the overall significance of the model:
  $h(t, x) = h_0(t)exp(\beta_1 X_1 + \beta_2 X_2 + \cdots + \beta_p X_p)$
  - hypothesis: $H_0 : \beta_1 = \beta_2 = \cdots = \beta_p = 0$ vs. H_α : not H_0
  - the log partial likelihood ratio test is $G = 2[l_p(\hat{\beta}) - l_p(0)]$
  - under the null hypothesis that all p coefficients are simultaneously equal to zero, and when large sample conditions are met, G **follows a chi-square distribution with p degrees of freedom.**
  - p-value for the test G is $\Pr(\chi^2_p \geq G)$.

Comparison of Two Nested Models

- Compare two models:
  - model 1: $h(t, X) = h_0(t) exp(\beta_1 X_1 + \beta_2 X_2 + \cdots + \beta_p X_p)$
  - model 2:
    $h(t, X) = h_0(t) exp(\beta_1 X_1 + \beta_2 X_2 + \cdots + \beta_p X_p + \cdots + \beta_q X_q)$
  - hypothesis: $H_0 : \beta_{p+1} = \cdots = \beta_q = 0$ vs. H_α : not H_0
  - the log partial likelihood ratio test is $G = 2[l_q(\hat{\beta}) - l_p(\hat{\beta})]$
  - Under H_0 and with large sample size, G **follows a chi-square distribution with** $q - p$ **degrees of freedom.**
  - p-value for the test G is $\Pr(\chi^2_{q-p} \geq G)$.

General Survival Data Analysis Strategy

- Step 1: Kaplan-Meier survival curve, quartiles, and log-rank test for variables of interest; descriptive statistics for all covariates
- Step 2: Univariable cox proportional hazards models – fit a model with one variable at a time
- Step 3: Multivariable cox proportional hazards model with variables significant in the univariable analysis as well as any other variables not selected with this criterion but judged to be of clinical importance
- Step 4: Reduce the multivarible model by comparing models using log partial likelihood ratio test $2[l_{\text{large model}}(\hat{\beta}) - l_{\text{small model}}(\hat{\beta})]$. The variables that are neither statistically significant nor an important confounder are removed from the multivariable model
- Step 5: Examine the scale of the continuous covariates, transformation of the covariate may be needed (Section 5.2.1 Hosmer et al. 2008).
- Step 6: Add significant interaction terms
- Step 7: Model evaluation: checking model assumption, influential observations, and overall goodness of fit (Chapter 6 Hosmer et al. 2008).

Suggested Readings

- Chapter 2.5, 3.1-3.2, 3.3 pages 80-85, 4.1-4.4 (Hosmer et al.)
- Chapter 6.11-6.14, 7.1-7.5 (Dupont)