BIST P8110: Applied Regression II

10. Cox Models Estimation and Interpretation

Qixuan Chen

Department of Biostatistics
Columbia University

Multivariable Cox Models

- Suppose that the hazard depends on several covariates, say p covariates $(X_1, X_2, \ldots, X_p)$.
  - For now, we assume that the values of these covariates do not change over time (time-independent).
  - Let the vector x represent the values of the p covariates, $x^T = (x_1, x_2, \ldots, x_p)$.
  - Let $h_0(t)$ be the hazard function for an individual with $x^T = (0, 0, \ldots, 0)$.
- The Cox model can be written as:

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

Multivariable Cox Model

- The Cox model can be rewritten as:

$$\log\left\{\frac{h(t, x_i)}{h_0(t)}\right\} = \beta_1 x_{i1} + \beta_2 x_{i2} + \cdots + \beta_p x_{ip}$$

  - there is no constant term in the linear component. Why?
  - we make no assumption regarding the actual form of the baseline hazard function, $h_0(t)$.
  - HR = e^{β}

HR Estimation and Interpretation: One Covariate Only

HR Estimate and 95% CI

- The estimated hazard ratio is $e^{\hat{\beta}}$.
- A 95% CI for HR is estimated as

$$\exp(\hat{\beta} - 1.96\widehat{SE}(\hat{\beta})), \exp(\hat{\beta} + 1.96\widehat{SE}(\hat{\beta}))$$

Example 1: Dichotomous Covariate

- **EX1**: Fit a Cox model with binary x (1 = experimental group; 0 = control group).
  - Model: $h(t, x, \beta) = h_0(t)e^{x\beta}$
  - Hazard ratio: $HR = \frac{h(t,x=1,\beta)}{h(t,x=0,\beta)} = \frac{h_0(t)e^{\beta\times 1}}{h_0(t)e^{\beta\times 0}} = e^{\beta}$

Example 1: Dichotomous Covariate

- **Example 1.1** positive β estimate
  - $\hat{\beta} = 0.555, \widehat{SE}(\hat{\beta}) = 0.282$
  - $\widehat{HR} = e^{0.555} = 1.742$
  - the estimated 95% CI for HR:
    $e^{\hat{\beta} \pm 1.96 \times \widehat{SE}(\hat{\beta})} = e^{0.555 \pm 1.96 \times 0.282} = (1.003, 3.027)$
  - The HR can be interpreted as
    - The death rate among patients in the experimental group is 1.742 times that among patients in the control group throughout the study period, and this could be as little as 1.003 times or as much as 3.027 times with 95 percent confidence,
    - Or, the death rate among patients in the experimental group is 74.2 percent larger than that among patients in the control group throughout the study period, and it could be as little as 0.3 percent larger or as much as 202.7 percent larger with 95 percent confidence.

Example 1: Dichotomous Covariate

- **Example 1.2** negative β estimate
  - $\hat{\beta} = -0.684, \widehat{SE}(\hat{\beta}) = 0.215$
  - $\widehat{HR} = e^{-0.684} = 0.505$
  - the 95% CI for HR:
    $e^{\hat{\beta} \pm 1.96 \times \widehat{SE}(\hat{\beta})} = e^{-0.684 \pm 1.96 \times 0.215} = (0.331, 0.769)$
  - The HR can be interpreted as
    - The death rate among patients in the experimental group is 0.505 times that of patients in the control group throughout the study period, and this could be as little as 0.769 times or as much as 0.331 times with 95 percent confidence,
    - Or, the death rate among patients in the experimental treatment group is 49.5 percent less than that of patients in the control group throughout the study period, and this decrease could be as little as 23.1 percent or as much as 66.9 percent with 95 percent confidence.

Example 2: Continuous Covariate

- **EX2**: Fit a Cox model with continuous x = age (in years).
  - Model: $h(t, x, \beta) = h_0(t)e^{x\beta}$
  - $\hat{\beta} = 0.046$, $\widehat{SE}(\hat{\beta}) = 0.012$
  - For continuous covariates, the hazard ratio for the clinically interesting unit of change along with its CI are frequently reported, e.g. consider a 5-year increase in age:
    - $\widehat{HR}(\text{5-year increase in age}) = e^{5\times 0.046} = 1.259$
    - 95% CI = $e^{5\times 0.046 \pm 1.96\times 5\times 0.012} = (1.119, 1.416)$
    - Interpretation: the death rate increases by 25.9 percent for every 5-year increase in age, and this increase could be as little as 11.9 percent and as much as 41.6 percent with 95 percent confidence.

Example 3: Nominal Covariate

- **EX3**: the age of patients was categorized into four groups < 60, $[60, 70)$, $[70, 80)$, and ≥ 80.
  - With the method of reference cell coding, using the youngest age group (age < 60) as the reference group, three dummy variables were created:
    - “age_2" (1 = $60 \leq$ age < 70; 0 = otherwise)
    - “age_3" (1 = $70 \leq$ age < 80; 0 = otherwise)
    - “age_4" (1 = ≥ 80; 0 = otherwise)

Example 3: Nominal Covariate

- **EX3**: let $x_1 = $ age_2, $x_2 = $ age_3, and $x_3 = $ age_4.
  - Model: $h(t, x_1, x_2, x_3, \beta) = h_0(t)\exp(\beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3)$
  - HR(t, age group 2 vs. 1) = e^{β_1}
  - HR(t, age group 3 vs. 1) = e^{β_2}
  - HR(t, age group 4 vs. 1) = e^{β_3}
  - HR(t, age group 3 vs. 2) = $e^{\beta_2 - \beta_1}$
  - HR(t, age group 4 vs. 2) = $e^{\beta_3 - \beta_1}$
  - HR(t, age group 4 vs. 3) = $e^{\beta_3 - \beta_2}$

Example 3: Nominal Covariate

- **EX3**: The estimation and interpretation of HR for the first three group comparisons (2 vs 1, 3 vs 1, 4 vs 1) are similar to dichotomous covariates.

Example 3: Nominal Covariate

- **EX3**: The estimation of HR for the last three group comparisons (3 vs 2, 4 vs 2, 4 vs 3) needs some calculation.
  - Suppose that
    - $\hat{\beta}_2 = 0.986$, $\hat{\beta}_3 = 1.263$
    - $\widehat{Var}(\hat{\beta}_2) = 0.173$, $\widehat{Var}(\hat{\beta}_3) = 0.198$
    - $\widehat{Cov}(\hat{\beta}_2, \hat{\beta}_3) = 0.126$
  - $\widehat{HR}(t, \text{age group 4 vs 3}) = e^{\hat{\beta}_3 - \hat{\beta}_2} = e^{1.263 - 0.986} = 1.319$
  - The estimated variance of $\hat{\beta}_3 - \hat{\beta}_2$ is

$$\begin{aligned} \widehat{Var}(\hat{\beta}_3 - \hat{\beta}_2) &= \widehat{Var}(\hat{\beta}_3) + \widehat{Var}(\hat{\beta}_2) - 2\widehat{Cov}(\hat{\beta}_3, \hat{\beta}_2) \\ &= 0.198 + 0.173 - 2 \times 0.126 = 0.119 \end{aligned}$$

  - The estimated 95% CI is

$$\begin{aligned} e^{(\hat{\beta}_3 - \hat{\beta}_2) \pm 1.96 \times \widehat{SE}(\hat{\beta}_3 - \hat{\beta}_2)} &= e^{(1.263 - 0.986) \pm 1.96 \times \sqrt{0.119}} \\ &= (0.671, 2.594) \end{aligned}$$

Example 3: Nominal Covariate

- **EX3**: Test the significance of the difference in two age groups.
  - Hypothesis: $H_0 : \beta_3 - \beta_2 = 0$ vs. $H_\alpha : \beta_3 - \beta_2 \neq 0$
  - Test statistic: $z = \frac{\hat{\beta}_3 - \hat{\beta}_2}{\widehat{SE}(\hat{\beta}_3 - \hat{\beta}_2)} = \frac{1.263 - 0.986}{\sqrt{0.119}} = 0.803$
  - P-value = $Pr(|Z| \geq |0.803|) = 0.422$
  - Decision rule: p-value > 0.05, fail to reject H_0 at $\alpha = 0.05$.
  - Conclusion: There is insufficient evidence to conclude that the death rates differ between patients aged 70-80 and patients older than 80.

Example 3: Nominal Covariate

- the estimate of variance for $c_1\hat{\alpha} + c_2\hat{\gamma}$ is

$$\widehat{Var}(c_1\hat{\alpha} + c_2\hat{\gamma}) = c_1^2\widehat{Var}(\hat{\alpha}) + c_2^2\widehat{Var}(\hat{\gamma}) + 2c_1c_2\widehat{Cov}(\hat{\alpha}, \hat{\gamma})$$

- **Ex 3**: $c_1 = 1, c_2 = -1, \hat{\alpha} = \hat{\beta}_3, \hat{\gamma} = \hat{\beta}_2$

$$\begin{aligned}
&\widehat{Var}(\hat{\beta}_3 - \hat{\beta}_2) \\
=&\widehat{Var}(1\hat{\beta}_3 + (-1)\hat{\beta}_2) \\
=&1^2\widehat{Var}(\hat{\beta}_3) + (-1)^2\widehat{Var}(\hat{\beta}_2) + 2 \times 1 \times (-1)\widehat{Cov}(\hat{\beta}_3, \hat{\beta}_2) \\
=&\widehat{Var}(\hat{\beta}_3) + \widehat{Var}(\hat{\beta}_2) - 2\widehat{Cov}(\hat{\beta}_3, \hat{\beta}_2)
\end{aligned}$$