Chapter 4

Interpretation of a Fitted Proportional Hazards Regression Model

Categorical Covariates

- Suppose we are interested in the effect of a dichotomous covariate on the hazard function.
  - Examples:
    - Compare two treatments
    - Investigate the effect of gender on survival
- A simple Cox model:
  - $h(t|\beta,x) = h_0(t)\cdot e^{\beta x}$, where x = 0 (reference) or 1
  - The hazard ratio is

$$HR(t|1,0) = HR(t|x = 1 \; vs \; x = 0) = \frac{h_0(t)e^{\beta\cdot 1}}{h_0(t)e^{\beta\cdot 0}} = e^{\beta}$$

Categorical Covariates

- Say we have fit a model and estimated a hazard ratio,

$$\hat{\theta} = e^{\hat{\beta}}$$

How should we interpret this result?

- Interpretations
  - "The hazard for the event when x = 1 is $\hat{\theta}$-times the hazard when x = 0".
  - "When x = 1 the event occurs at $\hat{\theta}$-times the rate for when x = 0".
  - "x = 1 is associated with a [($\hat{\theta}$ - 1) x 100%] increase in risk of event as compared to x = 0".

Categorical Covariates

- Example:
  - Outcome: survival following cystectomy
  - Covariate: treatment assignment
    - Let X = 0 for Drug A, X = 1 for Drug B
  - Fit a Cox proportional hazards model and estimated the coefficient for X as $\hat{\beta}$ = 0.743.

$$e^{\widehat{\beta}} = e^{0.743} = 2.1$$

Categorical Covariates

- Interpretations:
  - The hazard of death for subjects taking Drug B is 2.1 times the hazard for those taking Drug A
  - Subjects taking Drug B die at 2.1 times the rate of those on Drug A
  - Taking Drug B is associated with a 110% increase in the risk of death, compared to Drug A
    - Percent increase: [(2.1 – 1) x 100%] = 110%

Categorical Covariates

- Sometimes may be helpful to look at results in the opposite direction.

$$\mathrm{HR}(\mathrm{t}|\mathrm{Drug\ A\ vs\ Drug\ B}) = \frac{e^{\beta\cdot 0}}{e^{\beta\cdot 1}} = e^{-\beta}$$

  - $e^{-\hat{\beta}} = e^{-0.743} = 0.48$

- Taking Drug A is associated with a 52% decrease in the risk of death
  - Percent Increase: [(0.48 – 1) x 100%] = -52%, or percent decrease = 52%

Categorical Covariates

- Often categorical covariates will have more than two levels.
  - If there are K levels, need K-1 dummy variables to investigate the effect of that covariate

$$\log(h(t|\beta,x)) = \log h_0(t) + \beta_1 X_1 + \beta_2 X_2 + \ldots + \beta_{K-1} X_{K-1}$$

  - To test the effect of the covariate, test

H_0: $\beta_1 = \beta_2 = \ldots \ \beta_{K-1} = 0$

H_1: Not H_0.

  - To compare level *i* of the covariate to level *j*

$$HR(t|level\ i\ to\ level\ j) = \frac{e^{\hat{\beta}_i}}{e^{\hat{\beta}_j}} = e^{\hat{\beta}_i - \hat{\beta}_j}$$

Categorical Covariates

- Example: Investigate the effect of stage of cancer on survival time
  - Here, Stage 1 is the reference level and dummy variables are created for the remaining three stages.

Original Variable	Dummy Variables		
Cancer Stage	Stage2	Stage3	Stage4
1	0	0	0
2	1	0	0
3	0	1	0
4	0	0	1

Categorical Covariates

- The model:

$$h(t|\beta, x) = h_0(t)e^{\beta_2 Stage2 + \beta_3 Stage3 + \beta_4 Stage4}$$

  - e^{β_2} is the hazard ratio for Stage 2 versus Stage 1
  - e^{β_3} is the hazard ratio for Stage 3 versus Stage 1
  - e^{β_4} is the hazard ratio for Stage 4 versus Stage 1
- How can we estimate a hazard ratio for Stage 4 versus Stage 3?
  - HR(t|4,3) = $\frac{h_0(t)e^{\beta_4}}{h_0(t)e^{\beta_3}} = e^{\beta_4 - \beta_3}$

Categorical Covariates

- To get a 100(1-α)% CI for this hazard ratio, need to find a CI for β_4 - β_3.
  - The maximum partial likelihood estimators, $\hat{\beta}_3$ and $\hat{\beta}_4$, are approximately normal for large samples.
    - $\hat{\beta}_4 - \hat{\beta}_3 \sim N(\beta_4 - \beta_3\ ,\ Var(\hat{\beta}_4) + Var(\hat{\beta}_3) - 2Cov(\hat{\beta}_4, \hat{\beta}_3)$
  - 100(1-α)% CI for β_4 - β_3 is

$$\left(\hat{\beta}_4 - \hat{\beta}_3\right) \pm z_{1-\alpha/2}\sqrt{\hat{V}\left(\hat{\beta}_4 - \hat{\beta}_3\right)}$$

Linear Contrasts

- Can also test the null hypothesis H_0: $\beta_4 = \beta_3$ against the alternative H_1: $\beta_4 \neq \beta_3$ using a Wald test:
  - The test statistic is

$$z = \frac{\hat{\beta}_4 - \hat{\beta}_3}{\sqrt{\hat{V}\left(\hat{\beta}_4 - \hat{\beta}_3\right)}}$$

  - Reject H_0 in favor of H_1 if $z < z_{\alpha/2}$ or $z > z_{1-\alpha/2}$
  - The two-sided p-value is $2 \cdot P(Z > |z|)$, where Z follows a standard normal distribution

- This test is an example of a linear contrast.

Linear Contrasts

- To test the hypotheses

$$H_0: c_1\beta_1 + c_2\beta_2 + \ldots c_p\beta_p = 0$$
$$H_1: c_1\beta_1 + c_2\beta_2 + \ldots c_p\beta_p \neq 0$$

  - The test statistic is

$$z = \frac{c^T\hat{\beta}}{\sqrt{c^T\hat{V}\left(\hat{\beta}\right)c}}$$

where $\hat{\beta} = (\hat{\beta}_1, \hat{\beta}_2, \ldots, \hat{\beta}_p)$, $c = (c_1, c_2, \ldots, c_p)$, and $\hat{V}$ is the estimated covariance matrix for $\hat{\beta}$

  - The two-sided p-value is $2 \cdot P(Z > |z|)$, where Z follows a standard normal distribution

Linear Contrasts

- The multivariate version of a linear contrast can also be used to test multiple hypotheses simultaneously.
  - Example: Want to test H_0: $\beta_3 = 0$, $\beta_4 = 0$ versus H_1: At least one of β_3, β_4 is nonzero.

- Multivariate linear contrast: to test the hypotheses

$$H_0: C\beta = \mathbf{0} \text{ vs } H_1: C\beta \neq \mathbf{0}$$

(where C is a $d \times p$ matrix of coefficients and $\mathbf{0}$ is a $d \times 1$ vector of zeros) or, equivalently,

$$H_0: c_{11}\beta_1 + \ldots + c_{1p}\beta_p = 0, \ldots, c_{d1}\beta_1 + \ldots + c_{dp}\beta_p = 0$$
$$H_1: \text{Not } H_0$$

Linear Contrasts

The test statistic is

$$z^2 = \left(C\hat{\beta}\right)^T \left(\hat{V}(\hat{\beta})\right)^{-1} \left(C\hat{\beta}\right)$$

- If H_0 is true then $z^2 \sim \chi_d^2$

Continuous Covariates

- Suppose we are interested in the effect of a continuous covariate on the hazard function.
  - Examples:
    - Investigate the effect of age on survival
    - Determine whether BMI is associated with the risk of an event
- A simple Cox model:
  - $h(t|\beta,x) = h_0(t)\cdot e^{\beta x}$
  - The hazard ratio associated with a one-unit increase in x is

$$HR(t|a+1,a) = HR(t|x=a+1 \; vs \; x=a)$$

$$= \frac{h_0(t)e^{\beta\cdot(a+1)}}{h_0(t)e^{\beta\cdot a}} = e^{\beta}$$

Continuous Covariates

- Example: x = log WBC

$$\mathrm{HR}(\mathrm{t}|3{,}2) = \frac{h_0(t)e^{3\cdot\beta}}{h_0(t)e^{2\cdot\beta}} = e^{\beta}$$

is the hazard ratio that corresponds to a 1-unit increase in log WBC

- $e^{\beta} > 1$: for every one-unit increase in x, the hazard rate increases by $(e^{\beta} - 1)$x100%

- $e^{\beta} < 1$: for every one-unit increase in x, the hazard rate decreases by $(1 - e^{\beta})$x100%

Continuous Covariates

- Can also calculate hazard ratios for larger intervals than one "unit"
  - *c* unit increase

$$\mathrm{HR}(\mathrm{t}|\mathrm{x}+\mathrm{c}, x) = \frac{h_0(t)e^{\beta(x+c)}}{h_0(t)e^{\beta x}} = e^{c\beta}$$

- Point estimate for the hazard ratio: $e^{c\hat{\beta}}$
- 100(1-α)% CI for the hazard ratio:
  - Find confidence interval for the log hazard ratio, cβ

$$(\beta_L, \beta_U) = c\hat{\beta} \pm z_{1-\alpha/2} \cdot \sqrt{c^2\hat{V}(\hat{\beta})}$$

  - Change back to the hazard ratio scale

$$\left(e^{\beta_L}, e^{\beta_U}\right)$$

Continuous Covariates

- This is equivalent to multiplying the endpoints of a confidence interval for β by c, and then exponentiating.
- Can also raise the endpoints of a confidence interval for e^{β} to the c^{th} power.

Multiple Covariate Models

- If we have several covariates that are significantly associated with survival time, we may want to determine whether confounding or interaction is present.
  - Interaction or effect modification: the effect of one covariate varies according to the value of another covariate.
    - Example: A new treatment has been shown to improve survival. You also notice that the treatment works better for younger patients than for older patients.
    - There is an interaction between the treatment and age.

Multiple Covariate Models

- Confounding variable: a variable that is associated both with a covariate and with the outcome. Failing to account for a confounding variable will lead to biased estimates
  - Example: A study found an association between birth order and the risk of being born with Down syndrome.
  - But the mother’s age is associated with both birth order (a woman giving birth to her 4th child is likely to be older than one giving birth to her 1st) and risk of Down syndrome.
  - When age was included as a covariate in the analysis, there was no significant relationship between birth order and risk of Down syndrome.
  - Maternal age is a confounder for the effect of birth order on risk of Down syndrome

Confounding Variables

- To determine whether a variable is a confounder, fit the model both with and without that variable.
  - If the coefficient estimates change substantially, the variable is a confounder.
- If the variable is a confounder, then we need to adjust for that variable when reporting our results.

Confounding Variables

- Model without the potential confounder:

$$\log(h(t|\theta,x_1) = \log(h_0(t)) + \theta\cdot x_1$$

- Model including the potential confounder:

$$\log(h(t|\beta,x) = \log(h_0(t)) + \beta_1\cdot x_1 + \beta_2\cdot x_2$$

- $e^{\widehat{\theta}}$ and $e^{\widehat{\beta}_1}$ are both estimates of the hazard ratio associated with x_1.
  - $e^{\widehat{\theta}}$ is called the crude or unadjusted hazard ratio
  - $e^{\widehat{\beta}_1}$ is the adjusted hazard ratio

Confounding Variables

- If the unadjusted and adjusted hazard ratios are similar, then confounding is not present
  - The measure of difference (done on the log scale):

$$\Delta\hat{\beta}_1\% = 100\% \cdot \frac{\hat{\theta} - \hat{\beta}_1}{\hat{\beta}_1}$$

$$100\% \text{ x } \frac{(\text{Unadj.} - \text{Adj.})}{\text{Adj.}}$$

  - Rule of thumb: if the percent change is more than 20%, confounding is present.

Interactions

- An interaction is present if the effect of a covariate depends on the value of another covariate.
- Can test for an interaction between two covariates by defining a new covariate that is the product of the two.
  - Can use the Wald test to assess whether the interaction is a significant predictor.
  - The partial likelihood ratio test can also be used.
- An interaction term should not be included in a model unless both of the main effects are also included

Interactions

- Interactions by variable type:
  - Between two categorical variables → categorical
  - Between two continuous variables → continuous
  - Between one categorical and one continuous variable → continuous

Interactions

- Example: Determine whether there is an interaction between gender and treatment effect for time to relapse for leukemia patients.
  - Treatment: $x_1 = \begin{cases} 0 \text{ if old treatment} \\ 1 \text{ if new treatment} \end{cases}$
  - Gender: $x_2 = \begin{cases} 0 \text{ if male} \\ 1 \text{ if female} \end{cases}$
  - Interaction term: $x_3 = x_1 \cdot x_2$. x_3 is 1 if subject is a woman receiving the new treatment, equals zero otherwise
- Fit the Cox model:

$$\log(h(t|\beta,x)) = \log(h_0(t)) + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3$$

Interactions

- HR for effect of treatment:
  - For men: HR(t | new vs old) = e^{β_1}
  - For women: HR(t | new vs old) = $e^{\beta_1+\beta_3}$
- Interpretation
  - If $\beta_3 = 0$, then the treatment works the same for men and women
  - $\beta_3 > 0$: HR for treatment is greater for women than for men.
  - $\beta_3 < 0$: HR for treatment is less for women than for men.

Interactions

- Example: Determine whether there is an interaction between gender and log WBC for time to relapse for leukemia patients.
- Fit the Cox model:

$$\log(h(t|\beta,x)) = \log(h_0(t)) + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3$$

  - Gender: $x_1 = \begin{cases} 1 \text{ if female} \\ 0 \text{ if male} \end{cases}$
  - x_2 = log WBC
  - Interaction term: $x_3 = x_1 \cdot x_2$

Interactions

- When an interaction involves a continuous covariate, can estimate the hazard ratio at specific values of the covariate.
  - Example: What is the hazard ratio associated with gender for subjects with log WBC = 3?

- The model

$$\log(h(t|\beta,x) = \log(h_0(t)) + \beta x$$

  is a linear function

  - Can plot the estimated log hazard against a continuous covariate to visualize the interaction effect