Sample Questions: Proportional Hazards Regression
Part One

STA312 Fall 2023. Copyright information is at the end of the last page.

1. For a proportional hazards regression model with hazard function $h_i(t|\boldsymbol{\beta}) = h_0(t)e^{\beta_0+\mathbf{x}_i^\top\boldsymbol{\beta}}$, the partial likelihood is

$$\mathrm{PL}(\boldsymbol{\beta}) = \prod_{i=1}^{D}\left(\frac{h_{(i)}(t_{(i)}|\boldsymbol{\beta})}{\sum_{j\in R_{(i)}} h_j(t_{(i)}|\boldsymbol{\beta})}\right) = \prod_{i=1}^{D}\left(\frac{e^{\mathbf{x}_{(i)}^\top\boldsymbol{\beta}}}{\sum_{j\in R_{(i)}} e^{\mathbf{x}_j^\top\boldsymbol{\beta}}}\right).$$

   (a) What happened to β_0? It cancelled, like this

$$PL(\beta) = \prod_{i=1}^{D} \frac{\cancel{h_0(t)}\cancel{e^{\beta_0}} e^{x^\top_{(i)}\beta}}{\sum_{j\in R_{(i)}} e^{x^\top_j\beta}\cancel{h_0(t)}\cancel{e^{\beta_0}}}$$

   (b) Write the log partial likelihood for a model with *one explanatory variable*. Differentiate and set to zero.

$$\ell = \log \prod_{i=1}^{D} \frac{e^{x_{(i)}\beta}}{\sum_{j\in R_{(i)}} e^{x_j\beta}} = \sum_{i=1}^{D}\left(\log e^{x_{(i)}\beta} - \log \sum_{j\in R_{(i)}} e^{x_j\beta}\right)$$

$$= \sum_{i=1}^{D} x_{(i)}\beta - \sum_{i=1}^{D} \log \sum_{j\in R_{(i)}} e^{x_j\beta}$$

$$= \sum_{i=1}^{D}\left(x_{(i)}\beta - \log \sum_{j\in R_{(i)}} e^{x_j\beta}\right) \qquad \sum_x x\, p(x)$$

$$\ell'(\beta) = \sum_{i=1}^{D}\left(x_{(i)} - \frac{\sum_{j\in R_{(i)}} e^{x_j\beta}x_j}{\sum_{j\in R_{(i)}} e^{x_j\beta}}\right)$$

$$= \sum_{i=1}^{D}\left(x_{(i)} - \sum_{j\in R_{(i)}} x_j \frac{e^{x_j\beta}}{\sum_{k\in R_{(i)}} e^{x_k\beta}}\right)$$

$$S(t) = e^{-\int_0^t h(y)\,dy}$$

2. Let $h(t) = h_0(t)\, e^{\mathbf{x}^\top \boldsymbol{\beta}}$. Show $S(t) = S_0(t)^{\exp\{\mathbf{x}_i^\top \boldsymbol{\beta}\}}$.

$$S_0(t) = e^{-\int_0^t h_0(y)\,dy}$$

$$S(t) = e^{-\int_0^t h(y)\,dy}$$

$$= e^{-\int_0^t h_0(y)\, e^{x^\top\beta}\,dy}$$

$$= e^{-e^{x^\top\beta}\int_0^t h_0(y)\,dy}$$

$$= \left(e^{-\int_0^t h_0(y)\,dy}\right)^{e^{x^\top\beta}}$$

$$= S_0(t)^{\exp\{x^\top\beta\}}$$

3. Adult volunteers who were unemployed were randomly assigned to either Job Training Program A, Job Training Program B, or a wait list control group. The response variable was time until employment. It might be censored because the person was still unemployed at the end of the study, or if they left the study for other reasons. Age is a covariate.

(a) Assuming a proportional hazards regression model, write the hazard function, denoting the length of time until employment by t. Denote age by a. There should be *no interactions* in the model, in case you know what that is. You do not need to say how your dummy variables are defined. You will do that in the next part. Complete the equation below.

$h(t) =$ $h_0(t)\, e^{\beta_1 a + \beta_2 d_1 + \beta_3 d_2}$

(b) In the table below, make columns showing how your dummy variables are defined. In the last column, write the hazard function $h(t)$ ~~and~~ for a particular vector of explanatory variable values $\mathbf{x}$, using the notation of your model from Question 3a above. If *symbols* for your dummy variables appear in the last column, the answer is wrong.

	d_1	d_2	$h(t)$
Wait List	0	0	$h_0(t)\, e^{\beta_1 a}$
Program A	1	0	$h_0(t)\, e^{\beta_1 a}\, e^{\beta_2}$
Program B	0	1	$h_0(t)\, e^{\beta_1 a}\, e^{\beta_3}$

(c) In the notation of your model, what is the risk of employment at time t for a 25-year-old participant on the wait list?

$h_0(t)\, e^{25\beta_1}$

(d) For a 60-year-old participant in Program A, the chances of finding a job at any time period are _________ times as great as the chances for a 60-year-old on the wait list. Answer in terms of the Greek letters from your model.

e^{β_2}

$= \dfrac{\cancel{h_0(t)\, e^{\beta_1 a}}\, e^{\beta_2}}{\cancel{h_0(t)\, e^{\beta_1 a}}}$

(e) For a 47-year-old ~~patient~~ in Program A, the hazard of finding a job at time t is _________ times as great as the hazard for a 47-year-old in Program B. Answer in terms of the Greek letters from your model.

$$e^{\beta_2} / e^{\beta_3} = e^{\beta_2 - \beta_3}$$

(f) You want to know whether, controlling for age, training program (A, B or neither) has any effect on time until employment. What is the null hypothesis? Answer in terms of the Greek letters from your model.

$$H_0 : \beta_2 = \beta_3 = 0$$

(g) That last question could be answered with either a large-sample likelihood ratio test, or a Wald test.

partial

i. Suppose you decided on a likelihood ratio test. Write the hazard function for the restricted model in the space below.

$h(t) = h_0(t) e^{\beta_1 a}$

ii. Suppose you decided on a Wald test. Write $H_0 : \mathbf{L}\boldsymbol{\theta} = \mathbf{0}$ in terms of specific matrices.

β

$$\begin{array}{cccc} L & \beta & = & 0 \end{array}$$

$$\begin{pmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} \beta_1 \\ \beta_2 \\ \beta_3 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}$$