Sample Questions: Weibull Regression

STA312 Spring 2019. Copyright information is at the end of the last page.

1. Let the failure time $t_i = \exp\{\beta_0 + \beta_1 x_{i,1} + \ldots + \beta_{p-1} x_{i,p-1}\} \cdot \epsilon_i^\sigma$. Show that if $x_{i,k}$ is increased by c units, t_i is multiplied by $e^{c\beta_k}$.

t_i^* = observation with x_{ik} increased by c

$$t_i^* = c\, t_i \iff c = \frac{t_i^*}{t_i}$$

$$= \frac{e^{\beta_0 + \beta_1 x_{i1} + \cdots + \beta_k (x_{ik} + c) + \cdots + \beta_{p-1} x_{i,p-1}} \times \cancel{\varepsilon_i^\sigma}}{e^{\beta_0 + \beta_1 x_{i1} + \cdots \beta_k x_{ik} + \cdots \beta_{p-1} x_{i,p-1}} \times \cancel{\varepsilon_i^\sigma}}$$

$$= \frac{\cancel{e^{\beta_0}}\, \cancel{e^{\beta_1 x_{i1}}} \cdots \cancel{e^{\beta_k x_{ik}}}\, e^{\beta_k c} \cdots \cancel{e^{\beta_{p-1} x_{i,p-1}}}}{\cancel{e^{\beta_0}}\, \cancel{e^{\beta_1 x_{i1}}} \cdots \cancel{e^{\beta_k x_{ik}}} \cdots \cancel{e^{\beta_{p-1} x_{i,p-1}}}}$$

$$= e^{c\beta_k}$$

$\varepsilon_i \sim Exp(1)$

2. Let $t_i = e^{\mu_i}\epsilon_i^\sigma$, where $-\infty < \mu_i < \infty$ and $\sigma > 0$. The idea is that $\mu_i = \beta_0 + \beta_1 x_{i,1} + \ldots + \beta_{p-1}x_{i,p-1}$.

(a) Derive the density of t_i.

$$f_i(t) = \frac{d}{dt}P(t_i \le t) = \frac{d}{dt}P(e^{\mu_i}\varepsilon_i^\sigma \le t)$$

$$= \frac{d}{dt}P(\varepsilon_i^\sigma \le t e^{-\mu_i}) = \frac{d}{dt}P(\varepsilon_i \le (te^{-\mu_i})^{\frac{1}{\sigma}})$$

$$= \frac{d}{dt}F_\varepsilon(t^{1/\sigma}e^{-\mu_i/\sigma})$$

$$= f_\varepsilon(t^{\frac{1}{\sigma}}e^{-\mu_i/\sigma})\cdot e^{-\frac{\mu_i}{\sigma}}\frac{1}{\sigma}t^{\frac{1}{\sigma}-1}$$

$$= \frac{1}{\sigma}e^{-\mu_i/\sigma}t^{\frac{1}{\sigma}-1}e^{-(e^{-\mu_i/\sigma}t^{1/\sigma})}$$

for $t \ge 0$

(b) Re-parameterizing by $\lambda_i = e^{-\mu_i}$ and $\alpha = 1/\sigma$, verify that t_i has a Weibull distribution.

$$f(t \mid \alpha, \lambda_i) = \alpha \lambda_i^{\alpha} t^{\alpha-1} e^{-(\lambda_i t)^{\alpha}}$$

$$= \alpha \lambda_i (\lambda_i t)^{\alpha-1} e^{-(\lambda_i t)^{\alpha}} \quad t \geq 0$$

Weibull density

$\alpha = \frac{1}{\sigma}, \quad \lambda_i = e^{-\mu_i}$

(c) The hazard function of a Weibull is $h(t) = \alpha\lambda^{\alpha}t^{\alpha-1}$. Going back to the (μ, σ) parameterization of the Weibull and substituting $\mu_i = \mathbf{x}_i^{\top}\boldsymbol{\beta}$, write the hazard function $h(t)$.

$$h(t) = \frac{1}{\sigma} e^{-\mu_i/\sigma} t^{1/\sigma - 1}$$

Q4 Show Weibull Regression has proportional hazards.

Setting $\mu_i = x_i^{\top}\beta$. Let x_1 & x_2 be the explanatory variable values for 2 individuals, $x_1 \neq x_2$

$$\frac{h_1(t)}{h_2(t)} = \frac{\cancel{\frac{1}{\sigma}} e^{-\frac{1}{\sigma}(x_1^{\top}\beta)} \cancel{t^{1/\sigma - 1}}}{\cancel{\frac{1}{\sigma}} e^{-\frac{1}{\sigma}(x_2^{\top}\beta)} \cancel{t^{\frac{1}{\sigma}-1}}}$$

Does not depend on t

$\alpha = \frac{1}{\sigma}, \lambda = e^{-\mu_i}$

(d) The expected value of a Weibull is $\Gamma(1+\frac{1}{\alpha})/\lambda$. Write this in the (μ, σ) parameterization, letting $\mu_i = \mathbf{x}_i^\top \boldsymbol{\beta}$.

$$E(T) = \Gamma(1+\sigma)/e^{-x_i^T\beta}$$

$$= e^{x_i^T\beta}\,\Gamma(1+\sigma) \checkmark$$

$$= e^{\beta_0+\beta_1 x_{i1}+\cdots+\beta_{p-1}x_{i,p-1}}\,\Gamma(1+\sigma)$$

(e) If $x_{i,k}$ is increased by one unit, the expected failure time is multiplied by ____.

$$e^{\beta_k}$$

(f) The median of a Weibull random variable is $\frac{[\log(2)]^{1/\alpha}}{\lambda}$. Write this in the (μ, σ) parameterization, letting $\mu_i = \mathbf{x}_i^\top \boldsymbol{\beta}$.

with $\sigma = \frac{1}{\alpha}$, $\lambda = e^{-\mu_i}$

Median $= e^{x_i^\top \beta} \log(2)^{\sigma}$

(g) If $x_{i,k}$ is increased by one unit, the median failure time is multiplied by ___.

e^{β_k}

3. For a particular form of cancer, the standard treatment is a combination of chemotherapy and radiation therapy. Both chemotherapy and radiation have serious side effects. Some patients may be so weakened by the treatment that they die from other things (such as infections) that are apparently unrelated to the cancer.

Volunteer patents who were considering no treatment at all were randomly assigned to one of three experimental conditions. They received either Chemotherapy only, Radiation only, or Both treatments. The response variable is survival time, which in some cases will be right-censored. Age is an important predictor of survival, and is used as a covariate.

(a) Write the (multiplicative) Weibull regression equation, denoting the length of time between diagnosis and death (call it survival time) for patient i by t_i. Denote age by x_i. 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.

$t_i = e^{x_i^T \beta} \cdot \varepsilon_i^{\sigma} = e^{\beta_0 + \beta_1 d_{i,1} + \beta_2 d_2 + \beta_3 x_i} \varepsilon_i^{\sigma}$

(b) In the table below, make columns showing how your dummy variables are defined. In the last column, write the expected survival time, 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.

$e^{(\beta_0 + \beta_1 d_1 + \beta_2 d_2 + \beta_3 x)} \Gamma(1+\sigma)$

	d_1	d_2	Expected Survival Time
Chemotherapy	1	0	$e^{\beta_0 + \beta_1 + \beta_3 x} \Gamma(1+\sigma)$
Radiation	0	1	$e^{\beta_0 + \beta_2 + \beta_3 x} \Gamma(1+\sigma)$
Both	0	0	$e^{\beta_0 + \beta_3 x} \Gamma(1+\sigma)$

(c) In the notation of your model, what is the expected survival time for a 25-year-old patient receiving both radiation and chemotherapy?

$e^{\beta_0 + 25\beta_3}$ $\cancel{g(\sigma)}$ $\Gamma(1+\sigma)$

~~(d) For a 60-year-old patient receiving radiation only, the expected survival time is _________ times as great as the expected survival time for a 60-year-old receiving both radiation and chemotherapy. Answer in terms of the Greek letters from your model.~~

(d) You want to produce a large-sample confidence interval for expected survival time, for a 25-year-old patient receiving both radiation and chemotherapy. You need to use the delta method.

i. What is the parameter vector $\boldsymbol{\theta}$? Give a general answer for your model.

$$\theta = (\beta_0, \beta_1, \beta_2, \beta_3, \sigma)$$

ii) $$g(\theta) = e^{\beta_0 + 25\beta_3}\,\Gamma(1+\sigma)$$

iii. What is $\dot{g}(\boldsymbol{\theta})$? $= \left(\frac{\partial g}{\partial \beta_0}, \frac{\partial g}{\partial \beta_1}, \frac{\partial g}{\partial \beta_2}, \frac{\partial g}{\partial \beta_3}, \frac{\partial g}{\partial \sigma}\right)$

$$= \left(e^{\beta_0 + 25\beta_3}\,\Gamma(1+\sigma),\ 0,\ 0,\ e^{\beta_0 + 25\beta_3}\,25\,\Gamma(1+\sigma),\ e^{\beta_0 + 25\beta_3}\,\Gamma'(1+\sigma)\right)$$

(e) For a 60-year-old patient receiving radiation only, the expected survival time is __________ times as great as the expected survival time for a 60-year-old receiving both radiation and chemotherapy. Answer in terms of the Greek letters from your model.

$$\frac{e^{\beta_0 + \beta_2 + \beta_3 \cancel{x}^{60}}\,\cancel{\Gamma(1+\sigma)}}{e^{\beta_0 + \beta_3 \cancel{x}^{60}}\,\cancel{\Gamma(1+\sigma)}} = e^{\beta_2}$$

(e) For a 47-year-old patient receiving radiation only, the expected survival time is _________ times as great as the expected survival time for a 47-year-old receiving chemotherapy only. Answer in terms of the Greek letters from your model.

$$\frac{e^{\beta_0+\beta_2+\beta_3 x}\,\Gamma(1+\sigma)}{e^{\beta_0+\beta_1+\beta_3 x}\,\Gamma(1+\sigma)} = e^{\beta_2-\beta_1}$$

(f) You want to know whether, controlling for age, experimental treatment (Chemotherapy, Radiation, or Both) has any effect on average survival time. What is the null hypothesis? Answer in terms of the Greek letters from your model.

$$H_0: \beta_1 = \beta_2 = 0$$

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

i. Suppose you decided on a likelihood ratio test. Write the multiplicative Weibull regression equation for the restricted model.

$t_i =$ $e^{\beta_0+\beta_3 x_i} \cdot \varepsilon_i^{\sigma}$

ii. Suppose you decided on a Wald test. Write the $\mathbf{L}$ matrix for $H_0 : \mathbf{L}\boldsymbol{\theta} = \mathbf{0}$.

$$\underbrace{\begin{pmatrix} 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \end{pmatrix}}_{L} \begin{pmatrix} \beta_0 \\ \beta_1 \\ \beta_2 \\ \beta_3 \\ \sigma \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}$$

(h) You want to know whether it is better for patients to get both radiation and chemotherapy, or just radiation. What is the null hypothesis? Answer in terms of the Greek letters from your model.

$$H_0: \beta_2 = 0$$

(i) You want to know whether it is better for patients to get both radiation and chemotherapy, or just chemotherapy. What is the null hypothesis? Answer in terms of the Greek letters from your model.

$$H_0: \beta_1 = 0$$

(j) You want to know whether it is better for patients to get just radiation or just chemotherapy. What is the null hypothesis? Answer in terms of the Greek letters from your model.

$$H_0: \beta_1 = \beta_2$$

4. Show that the multiplicative Weibull regression model has proportional hazards. Consider two patients with different vectors of explanatory variable values.

Done on P. 4
