Summative Quiz 3 (Simple Cox Regression) Solutions

1. What guides a researcher when deciding between using either linear, logistic, or Cox proportional hazards regression as an analysis tool?

Answer: The outcome variable type (continuous, binary, or time–to-event) for the particular analyses utilizing the regression model(s).

The following information will be necessary for questions 2 and 3:

The generic formulation of a multiple regression model relating an outcome (continuous, binary, number of events, or time-to-event) to a three-category predictor, health insurance type (none, public, private), is as follows:

$$LHS = intercept + \hat{\beta}_1 x_1 + \hat{\beta}_2 x_2$$

Where
LHS = "left hand side", and intercept = $\hat{\beta}_o or$ ln $(\hat{\lambda}_o(t))$.
x_1 = 1 if public insurance, 0 if not public insurance
x_2 = 1 if private insurance, 0 if not private insurance

2. What does the slope, $\hat{\beta}_1$, estimate?

Answer: The difference in the LHS values for persons with public insurance and the LHS for persons with no insurance.

Reasoning: Notice: x_1= 1, and x_2= 0 for those with public insurance: $LHS = intercept + \hat{\beta}_1$ and x_1= 0, and x_2= 0 for those with no insurance: $LHS = intercept$.
So the difference in the LHS estimate is $\hat{\beta}_1$. For regression types where the LHS is on the natural log scale (ln(odds) for logistic; ln(hazard) for Cox), this resulting difference can be exponentiated to get the corresponding ratio of interest (odds ratio for logistic, hazard ratio for Cox)

3. What is the null hypothesis for the test comparing the LHS value for person with private insurance to persons with no insurance?

Answer: H_o: β_2=0

Reasoning: The null value for slope values of any type of regression is 0. In logistic and Cox, when the slopes are exponentiated to compute (odds or hazard) ratios, the null value of the exponentiated slope is 1, *but on the slope scale the null value is always 0.*

4. Which of the following is a comparison of the LHS value for subjects with public insurance to subjects with private insurance?

Answer:

$$\hat{\beta}_1 - \hat{\beta}_2$$

Reasoning: Notice: x_1= 1, and x_2= 0 for those with public insurance:

$LHS = intercept + \hat{\beta}_1$

And: $x_1 = 0$, and $x_2 = 1$ for those with public insurance: $LHS = intercept + \hat{\beta}_2$

So the difference in the LHS estimate is $\hat{\beta}_1 - \hat{\beta}_2$.

<u>The following information may be relevant to questions 5-8.</u>

A 2018 article in the *New England Journal of Medicine* [1] reports the results from a randomized study designed to evaluate the association between following a Mediterranean diet plan (one of two types), and having a major cardiovascular event.

As per the researchers,

"In a multicenter trial in Spain, we assigned 7,447 participants (55 to 80 years of age, 57% women) who were at high cardiovascular risk, but with no cardiovascular disease at enroll-ment, to one of three diets: a Mediterranean diet supplemented with extra-virgin olive oil, a Mediterranean diet supplemented with mixed nuts, or a control diet (advice to reduce dietary fat). *The primary outcome was a major cardiovascular event (myocardial infarction, stroke, or death from cardiovascular causes).* The median follow-up period was 4.8 years."

The researchers used simple Cox regression to measure the association between the primary outcome in the follow-up period and diet type. The model used was:

$$\ln(hazard\ of\ primary\ outcome\ at\ time\ t; x_1) = \ln\left(\hat{\lambda}_o(t)\right) + \hat{\beta}_1 x_1 + \hat{\beta}_2 x_2$$

Where: $x_1 = 1$ if patients was randomized to receive a Mediterranean diet supplemented with extra-virgin olive oil, and 0 if not

$x_2 = 1$ if patients was randomized to receive a Mediterranean diet supplemented with nuts, and 0 if not

The estimates and standard errors for the regression slopes are: (yes, $\hat{\beta}_1 = \hat{\beta}_2$, this is not a typo)

$$\hat{\beta}_1 = -0.36;\ SE(\hat{\beta}_1) = 0.14$$
$$\hat{\beta}_2 = -0.36;\ SE(\hat{\beta}_2) = 0.14$$

5. What does the function $ln(\hat{\lambda}_o(t))$ represent in the Cox regression equation?

---

[1] Estruch R, et al. Primary Prevention of Cardiovascular Disease with a Mediterranean Diet Supplemented with Extra-Virgin Olive Oil or Nuts *New England Journal of Medicine* 378;25. (2018)

Answer: The ln(hazard) of the primary outcome in the control diet group as a function of time across the follow-up period.

6. What assumption did the researchers make in order to use Cox regression to quantify the relationship between the primary outcome and diet type?

   Answer: The ratio of the hazard of the primary outcome for each of the Mediterranean diet groups compared to the control group, and to each other, is constant over follow-up period.

   Reasoning: This is the proportional hazards assumption.

7. What is the estimated hazard ratio of the primary outcome for the Mediterranean diet group supplemented with nuts compared to the control diet group, and the 95% CI, based on the results from this research?

   Answer: 0.70 (0.53, 0.92)
   Reasoning: The slope, $\hat{\beta}_1 = -0.36$ is the estimated ln(hazard ratio) of the primary outcome for the PCI group compared to the control group. A 95% for the slope is given by $\hat{\beta}_1 \pm 2\widehat{SE}(\hat{\beta}_1) \rightarrow -0.36 \pm 2(0.14) \rightarrow (-0.64, -0.08)$.
   The estimated hazard ratio of the primary outcome Mediterranean diet group supplemented with nuts compared to the control diet group is $e^{\widehat{\beta}_1} = e^{-0.36} \approx\ 0.70$. The confidence interval for the hazard ratio can be computed by exponentiating the endpoints of the 95% CI for the slope:$(e^{-0.64}, e^{-0.08}) \rightarrow (0.53, 0.92)$.

8. What can be inferred from the fact that $\hat{\beta}_1 = \hat{\beta}_2$ in the given Cox regression result?

   Answer: There is no difference in the hazard of the primary outcome during the follow-up period between the Mediterranean diet groups supplemented with nuts and the Mediterranean diet group supplemented with olive oil.
   Reasoning: The ln(hazard ratios) of the primary outcome for each of the two Mediterranean diet groups compared to the control group are equal (both equal -0.36), and hence the respective hazard ratios are equal (both equal 0.7). As such, the difference in ln(hazards) and hence ln(hazard ratio) fo the Mediterranean diet group supplemented with nuts compared to the Mediterranean diet groups supplemented with olive oil is 0, and the resulting hazard ratio is 1.

9. The following results are based on the Mayo Clinic clinical trial for D-Penicillamine (DPCA) versus control in patients with primary biliary cirrhosis (PBC):

   Histologic stage is an ordinal categorical variable coded as 1-4 (higher values meaning more advanced progression of disease). Suppose this is initially modeled as a semi continuous variable in a Cox regression, resulting in the following model:

ln(*hazard of mortaility at time t: x*)= $\ln\left(\hat{\lambda}_o(t)\right) + 0.82x_1$, where x_1 = 1 to 4 for the four histologic stage levels. The standard error of the slope for x_1. ($\widehat{SE}\left(\hat{\beta}_1\right)$)) is 0.12.

Based on theses results, what is the estimated hazard ratio of death, and 95% CI for patients with a histologic stage of 4 compared to patients with a histologic stage of 1?

Answer: 11.70 (5.7, 24.0)

Reasoning: $e^{(3*0.82)}$ (($e^{3*(0.82 -2*.12)}$, $e^{3*(0.82 +2*.12)}$)

10. The following results are based on the Mayo Clinic clinical trial for D-Penicillamine (DPCA) versus control in patients with primary biliary cirrhosis (PBC):

 Now, suppose another Cox regression is fit with a categorical version of histologic stage as its predictor yielding the following results:

 ln(*hazard of mortaility at time t: x*)= $\ln\left(\hat{\lambda}_o(t)\right)$+1.6$x_1$+2.2$x_2$+3.1$x_3$, where x =1 for patients with stage 2 (and 0 if not), x =1 for patients with stage 3 (and 0 if not), and x =1 for patients with stage 4 (and 0 if not). The standard errors for the slopes of x, x, and xare 1, 1, and 1 respectively.

 Based on this model (with histologic stage as multi-categorical), what is the estimated hazard ratio of death, and 95% CI for patients with a histologic stage of 4 compared to patients with a histologic stage of 1?

 Answer: 22.1 (3.0, 164.0)

 Reasoning: $e^{(3.1)}$ (($e^{(3.1-2*1)}$, $e^{(3.1+2*1))}$)