Homework 2: BST 665: Survival Analysis

Devin Koehl

Spring Semester 2019

1 .A researcher is interested in studying risk factors for myocardial infarction (MI). Four hundred patients identified to be potentially at risk of MI were studied. Fit a Cox proportional hazards model to assess the effect of atrial fibrillation and blood pressure on time to MI. [Your model should only have main effects for atrial fibrillation and blood pressure]. Use the results of this model to answer the following questions.

A. Write the model you just fit using mathematical notation (i.e., using β coefficients). Be sure to define all variables used in the model, including any dummy variables.

$$logh(t|\beta, x) = logh_0(t) + \beta_{bp}x_{bp} + \beta_{af}x_{af}$$

Where bp = blood pressure and af = atrial fib. This model shows the two variables that are currently in our model, blood pressure and atrial fibulation. x_{af} will take on 0 if atrial fibulation is no, 1 if atrial fibulation is yes. x_{bp} will take on 0 if blood pressure is normal, and 1 if blood pressure is high.

Here I used the reference group as "No" for atrial fibulation and "normal" for blood pressure. The results are below:

Analysis of Maximum Likelihood Estimates

Parameter		DF	Parameter Estimate	Standard Error	Chi-Square	Pr > ChiSq	Hazard Ratio	95% Hazard Ratio Confidence Limits		Label
BP	High	1	0.23836	0.15605	2.3331	0.1266	1.269	0.935	1.723	BP High
AFib	Yes	1	0.96086	0.19479	24.3320	<.0001	2.614	1.784	3.829	AFib Yes

If we plug in the β values from the model we just fit, we can plug this in

$$logh(t|\beta, x) = logh_0(t) + (0.23836)x_{bp} + (0.96086)x_{af}$$

B. Report a hazard ratio comparing a patient with high blood pressure and atrial fibrillation to a patient with normal blood pressure and no atrial fibrillation. Provide a 95% confidence interval for this hazard ratio.

To do this, we can use a contrast statement in SAS to test this specific hypothesis:

Contrast Estimation and Testing Results by Row

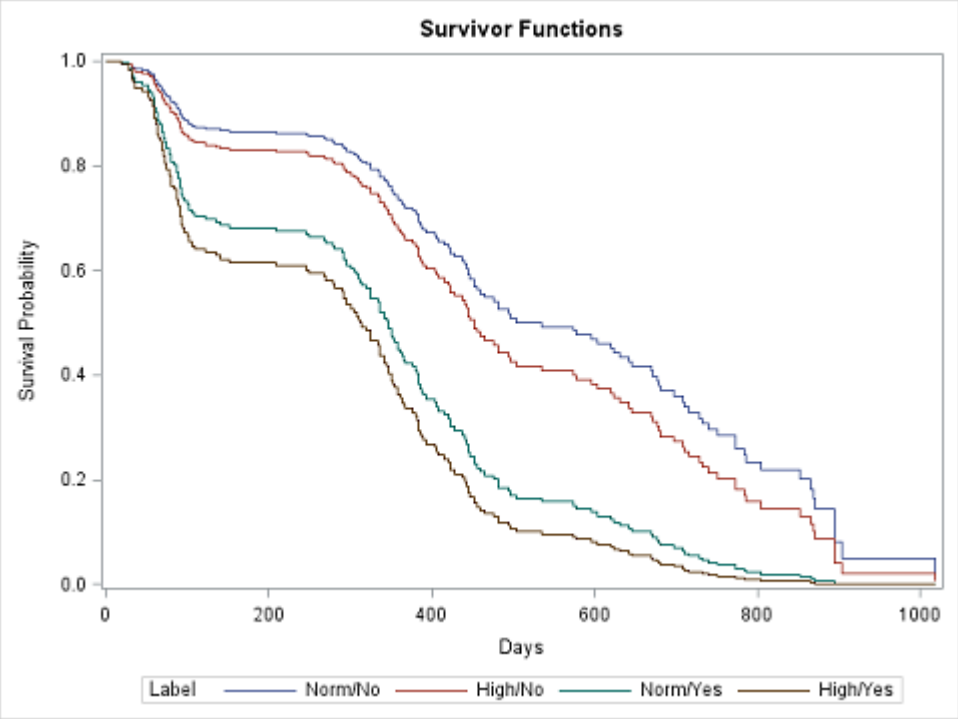
Contrast	Type	Row	Estimate	Standard Error	Alpha	Confidence Limits		Wald Chi-Square	Pr > ChiSq
High BP and Afib vs No BP and Afib	PARM	1	1.1992	0.2544	0.05	0.7006	1.6979	22.2169	<.0001

When we do this, the results are below:

From here we can see that a hazard ratio comparing these two patient groups is 1.1992 with 95% confidence intervals of (0.7006-1.6979) with a p-value of <0.0001.

C. On the same graph, plot the estimated survival functions for four hypothetical subjects

Using the baseline covariates statement in proc phreg, we can have the Cox model output the estimated survival functions for the hypothetical subjects:



What conclusions can we draw from this graph?

From the graph, we can see the patients with normal blood pressure and no atrial fibulation do best, while patients with high blood pressure and atrial fibulation do the worst.

2. Add BMI and an interaction between BMI and atrial fibrillation to the model used in Exercise 1. Use the results of this model to answer the following questions.

A. Does this model provide a significantly better fit to these data than the model used in Exercises 1-4? Use an appropriate hypothesis test to justify your answer. (Make sure you state the null and alternative hypotheses, level of significance used, test statistic, p-value and conclusion in terms of the problem).

Yes, this model provides a better fit. Examining the log-likelihoods of each model reveal this. In the reduced model we have a log-likelihood of 1640.165 and 1531.255 with the full model. We will use the partial likelihood ratio test to test which model is a better fit for the data.

H_0: The reduced model is a better fit and therefore the best fit model

H_A: The full model is a better fit and therefore the best fit model

The level of signifiance is α=0.05. The test statistic is 108.91 which is given by G = -full + reduced. The p-value for this turns out to be <0.001. Therefore, we reject the null hypothesis and conclude that the full model including BMI is a better fit for the data.

B. How much reduction in risk of death is associated with a 5-unit decrease in BMI for a person without atrial fibrillation? What about for a person with atrial fibrillation? Provide 95% confidence intervals for your answers.

First we if apply the units = 5 in phreg, we can show the increase. To get the decrease, we take 1/HR to go in the opposite direction.

Hazard Ratios for BMI			
Description	Point Estimate	95% Wald Confidence Limits	
BMI Unit=5 At AFib=No	3.107	2.531	3.816
BMI Unit=5 At AFib=Yes	1.772	1.168	2.689

For Afib = Yes:

First we take 1/1.772 and get a HR of 0.5643. To get the decrease we subtract (1-0.5643). This corresponds to a 44% decrease in risk. The 95% CI for the decrease are calculated by first taking 1/1.168 = 0.856 and 1/2.689 = 0.372. Now we subtract them both from 1. So the The 95% CI for the decrease in risk

is 14.4% to 62.8%.

For Afib = No:

First we take 1/3.107 and get a HR of 0.3219. To get the decrease we subtract (1-0.3219). This corresponds to a 68% decrease in risk. The 95% CI for the decrease are calculated by first taking $1/2.531 = 0.3951$ and $1/3.816 = 0.262$. Now we subtract them both from 1. So the The 95% CI for the decrease in risk is 60% to 74%.

C. Write 2-3 paragraphs summarizing the results of this study. Be sure to include an explanation of the interaction. Provide your interpretation of any results presented. Include figures as necessary.

In identifying risk factors for time in days to MI, two Cox Proportional Hazards model were fit. Data were analyzed using statstical software SAS 9.4 (Sas Insitute Inc., Cary, NC). A p-value of <0.05 was considered statistically significant. The risk factors in the first model were blood pressure, cateogorized as normal or high, and atrial fibulation, categorized as yes or no. After fitting the Cox Prorpotional Hazards model, the results showed that having atrial fibulation increased the risk of developing MI (HR 2.614, CI 1.784-3.829,$p<0.001$). Blood pressure did not come out as statistically significant in the model (HR 1.3, CI 0.935-1.723).

The second model assessed effect of Body Mass Index (BMI). After adjusting for a 5 unit change in BMI and the interaction between BMI and Atrial Fibulation, there was a significant interaction between Atrial Fibulation and BMI. In this model, it was shown that for every 5 unit increase of BMI with patients with Atrial Fibulation, the risk increased by 210% (HR 3.107, CI 2.531-3.816), while a 5 unit increase in BMI for patients with no atrial fibulation, the risk incresaed by only 17%. This demonstrates the strong effect of atrial fibulation on time to MI. Patients with atrial fibulation and higher BMI are at the most risk to develop MI.

3. A researcher is interested in identifying risk factors for multiple myeloma. He has fit a Cox proportional hazards model assessing the effects of age, protein level (g/dL), and hemoglobin on time to multiple myeloma diagnosis. Hemoglobin has been divided into three categories: low, medium, and high. Part of his SAS output is shown below.

A. Calculate the hazard ratio and 95% confidence interval for the hazard ratio associated with a four-unit increase in protein.

To calculate a 4 unit increase in protein, we multiply the parameter estimate by 4.

$e^{4*(-0.492)} = e^{-1.968} = 0.1397$. The HR is 0.1397.

Now to calculate the 95% confidence intervals, we use the the standard error along with the unit increase.

4*(-0.492) +/- 4*1.96(0.227)

4*(-0.492) + 4*1.96(0.227) = -0.18832 = $e^{-0.18832} = 0.8283$.

4*(-0.492) - 4*1.96(0.227) = -3.74768 = $e^{-3.74768} = 0.0236$.

Working this out, we have a HR and 95% confidence of HR 0.1397 (0.0236,0.8283)

B. Calculate all pairwise hazard ratios comparing subjects with high, medium, and low hemoglobin levels. Provide 95% confidence intervals all hazard ratios except the hazard ratio comparing a subject with high hemoglobin to one with medium hemoglobin.

If we were to calculate the pairwise comparisons, we are using the model with the levels of hemoglobin:

$logh(t|\beta, x) = logh_0(t) + \beta_{mh}x_{mh} + \beta_{hh}x_{hh}$

Where $x_{mh} = 1$ if medium hemoglobin and 0 otherwise, and $x_{hh} = 1$ if high hemoglobin and 0 otherwise.

If we want to do a pairwise comparison of medium vs. low, we have the following:

$\frac{logh_0(t)+e^{-0.490*1+-0.525*0}}{logh_0(t)+e^{-0.490*0+-0.525*0}} = e^{-0.490} = 0.612$.

The 95% confidence intervals for this would be using the β +/- 1.96*(SE), which is given to us in the table.

-0.490 +/- 1.96*0.200

-0.490 + 1.96 * 0.200 = $-0.098 = e^{-0.098} = 0.9066$

$-0.490 - 1.96 * 0.200 = -0.882 = e^{-0.882} = 0.413$

Therefore, the results of comparing medium vs low are a HR of 0.612 with CI (0.413,0.9066).

If we want to do a pairwise comparison of high vs. low, we have the following:

$\frac{logh_0(t)+e^{-0.490*0+-0.525*1}}{logh_0(t)+e^{-0.490*0+-0.525*0}} = e^{-0.525} = 0.591.$

The 95% confidence intervals for this would be

-0.591 +/- 1.96*0.196

$-0.525 + 1.96 * 0.196 = -0.14084 = e^{-0.14084} = 0.8686$

$-0.525 - 1.96 * 0.196 = -0.90916 = e^{-0.90916} = 0.40286$

Therefore, the results of comparing high vs low are a HR of 0.591 with CI (0.40286,0.8686).

If we want to do a pairwise comparison of high vs. medium, we have the following:

$\frac{logh_0(t)+e^{-0.490*0+-0.525*1}}{logh_0(t)+e^{-0.490*1+-0.525*0}} = \frac{e^{-0.490}}{e^{-0.525}}$

$\frac{e^{-0.490}}{e^{-0.525}} = e^{-0.490-(-0.525)} = e^{(-0.490+0.525)} = e^{0.035} = 1.035$

The results of comparing high vs medium show a HR of 1.035. The CI were not asked to be calculated in this problem. This concludes the pairwise comparisons for high vs low, high vs medium, and medium vs low.

C. Is hemoglobin significantly associated with the risk of multiple myeloma? Support your answer using a hypothesis test. (Make sure you state the null and alternative hypotheses, level of significance used, test statistic, p-value and conclusion in terms of the problem).

Yes, hemoglobin is significantly associated with the risk of multiple myeloma. We can see this from the overall p-value given in the type 3 test.

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

H_A: At least one of β_2 or β_3 is non-zero

The Wald Chi-square statistic is 9.63, with $\alpha = 0.05$. The p-value from the model is p=0.0081. We reject the null hypothesis and conclude that at least one of β_2 or β_3 is non-zero and hemoglobin is associated with the risk of multiple myeloma.

D. Help the researcher write the results section of his paper. Write 1-2 paragraphs describing the results of these analyses. Be sure to include your interpretation of the results and to take the stated goal of the study into consideration.

In identifying risk factors for multiple myeloma, a Cox Proportional Hazards model was fit. Data were analyzed using statstical software SAS 9.4 (Sas Insitute Inc., Cary, NC). A p-value of <0.05 was considered statistically significant. The covariates assesed in the model were age, protein, and hemoglobin. Hemoglobin was divided into low, medium, and high groupings. After adjusting for protein and hemoglobin levels, age came out at a risk factor associated with time to multiple myeloma (HR 1.04, CI 1.03-1.06, p=<0.001). For every one unit increase of age, the risk for multiple myeloma increased by 4%. A one unit increase in protein came out in the model as protective (HR 0.611, p=0.03). For every one unit increase in protein, the risk for multiple myeloma decreased by 39%. If we analyze a 4 unit increase in protein, this shows us that the risk for multiple myeloma decreases by even more, 86%. Hemoglobin came out significant in the model demonstrating the effect of high and medium levels of hemoglobin wth a decrease in risk. With high values of Hemoglobin, the risk decreases by 41%, while medium levels of hemoglobin resulted in a 39% decrease in risk. In conclusion, age, protein, and hemoglobin have a significant effect on for time to multiple myeloma.