P8110: Applied Regression II

Homework #5

[15 points]

NOTE: Answer the following questions using SAS. Cut and paste relevant SAS output to appropriate places in the texts of your solutions.

The "HW5data.csv" gives the time until death (in days) for 500 patients following hospital admission for acute myocardial infarction (MI). In this study, our goal is to investigate whether the rate of death is different between patients with first time MI and patients with MI recurrence. The number of days from hospital admission to last follow-up and the vital status at last follow-up were recorded. Other risk factors considered here include age, bmi, and year of the cohort study. The variables in the data file from left to right are:

ID	=	Patient ID
lenfol	=	Days from Hospital Admission Date to Date of Last Follow-up
fstat	=	Vital Status at Last Follow-up (**0** - Alive, **1** - Dead)
miord	=	MI order (**0** - First, **1** - Recurrent)
bmi	=	Body Mass Index
year	=	Cohort Year (**1** - 1997, **2** - 1999, **3** - 2001)
age_c	=	Age at Hospital Admission (**1** - age $<$ 60, **2** - 60 $\leq$ age $<$ 70, **3** - 70 $\leq$ age $<$ 80, **4** - age $\geq$ 80)

Definition of variables that will be used in the following Cox models:

$$X_1 = \begin{cases} 1, \text{if miord} = 1 \\ 0, \text{if miord} = 0 \end{cases}, \quad \text{the indicator of the MI order;}$$

$$X_2 = \begin{cases} 1, \text{if age\_c} = 2 \\ 0, \text{otherwise} \end{cases}, \quad X_3 = \begin{cases} 1, \text{if age\_c} = 3 \\ 0, \text{otherwise} \end{cases}, \quad X_4 = \begin{cases} 1, \text{if age\_c} = 4 \\ 0, \text{otherwise} \end{cases},$$

where X_2, X_3 and X_4 are the three dummy variables for the Age at Hospital Admission, and the first age group (age $<$ 60) is set as the reference group;

$$X_5 = \begin{cases} 1, \text{if year} = 2 \\ 0, \text{ otherwise} \end{cases}, \quad X_6 = \begin{cases} 1, \text{if year} = 3 \\ 0, \text{otherwise} \end{cases},$$ where X_5 and X_6 are the two dummy variables for Cohort year, and the year 1997 is set as the reference;

X_7 = value of BMI.

1. **Fit a Cox model with MI order as the only covariate (Model 1). Is the rate of death different between patients with first time MI and patients with MI recurrence, at a significant level of $\alpha = 0.05$? Write down the model. State**

the null and alternative hypothesis, test statistic (state which test was used), *p*-value, degrees of freedom, and conclusion. [**3 points**]

Model 1:

$$h(t, X_1, \beta_1) = h_0(t) \exp(\beta_1 X_1).$$

Hypotheses:

$$H_0 : \beta_1 = 0 \quad \text{v.s.} \quad H_\alpha : \beta_1 \neq 0$$

Test statistics and P-values : Any of the Wald, Likelihood Ratio or Score test could be used here. Based on the SAS output of Model 1, we have that

$$\text{Wald test:} \quad T_W = 9.4081, \quad \text{with 1 degree of freedom (d.f.)}$$

The corresponding p-value is $P(\chi_1^2 \geq 9.4081) = 0.0022$.

$$\text{Likelihood Ratio test:} \quad T_{LR} = 9.1431, \quad \text{with 1 d.f.}$$

The corresponding p-value is $P(\chi_1^2 \geq 9.1431) = 0.0025$.

$$\text{Score test:} \quad T_S = 9.5485, \quad \text{with 1 d.f.}$$

The corresponding p-value is $P(\chi_1^2 \geq 9.5485) = 0.002$.

Conclusion: According to the results of any of the three tests, we can reject H_0 at a significant level of $\alpha = 0.05$. Therefore, the MI order is a statistically significant predictor of the death rate following hospital admission for acute MI.

SAS output:

Testing Global Null Hypothesis: BETA=0			
Test	Chi-Square	DF	Pr > ChiSq
Likelihood Ratio	9.1431	1	0.0025
Score	9.5485	1	0.0020
Wald	9.4081	1	0.0022

2. **Fit a Cox model with MI order, age, bmi, and cohort year as covariates (Model 2). Is Model 2 significantly better than Model 1, at a significant level of $\alpha = 0.05$? Write down Model 2. State the null and alternative hypothesis, test statistic (state which test was used), *p*-value, degrees of freedom, and conclusion. [4 points]**

Model 2:

$$h(t, \mathbf{X}, \beta) = h_0(t) \exp\big(\beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \beta_4 X_4 + \beta_5 X_5 + \beta_6 X_6 + \beta_7 X_7\big),$$

where $\mathbf{X} = (X_1, X_2, X_3, X_4, X_5, X_6, X_7)$, $\beta = (\beta_1, \beta_2, \beta_3, \beta_4, \beta_5, \beta_6, \beta_7)$.

Hypotheses:

$$H_0 : \beta_2 = \beta_3 = \beta_4 = \beta_5 = \beta_6 = \beta_7 = 0 \ \text{ v.s. } \ H_\alpha : \text{Not all } \beta_i\text{'s in } H_0 \text{ are equal to zero}$$

Test statistic and P-value : We can use the Likelihood Ratio test here, since Model 1 is nested in Model 2. Based on the SAS output for model 2, we have that

$$T_{LR} = -2\Big(l_p\big(\hat{\beta}_{Model1}\big) - l_p\big(\hat{\beta}_{Model2}\big)\Big) = 2445.498 - 2305.777 = 139.721, \text{ with 6 d.f.}$$

The p-value is $P(\chi^2_6 \geq 139.721) < 0.0001$.

Conclusion: We can reject H_0 at the level of $\alpha = 0.05$, according to the Likelihood Ratio test for the nested models. Therefore, the effects of Age, BMI and Cohort year on the death rate are statistically significant, and thus they should be included in the Cox model to control for confounding or improve model fitting. Model 2 is significantly better than Model 1.

SAS output:

Model Fit Statistics

Criterion	Without Covariates	With Covariates
-2 LOG L	2454.641	2445.498
AIC	2454.641	2447.498
SBC	2454.641	2450.869

Model Fit Statistics

Criterion	Without Covariates	With Covariates
-2 LOG L	2454.641	2305.777
AIC	2454.641	2319.777
SBC	2454.641	2343.372

Figure 1: The Model Fit Statistics of Model 1 (left) and Model 2 (right)

3. **Fit a Cox model with MI order, age, bmi, cohort year, and the interaction between the MI order and age as covariates (Model 3). Write down Model 3. Hand calculate the hazards ratio and 95% CI between patients with first time MI and patients with MI recurrence at age 60-70. Also obtain the 95% CI using SAS (provide the SAS code).** [5 points]

Model 3:

$$
\begin{aligned}
h(t, \mathbf{X}, \beta) &= h_0(t) \exp\big(\beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \beta_4 X_4 + \beta_5 X_5 + \beta_6 X_6 + \beta_7 X_7 \\
&\quad + \beta_8 X_1 * X_2 + \beta_9 X_1 * X_3 + \beta_{10} X_1 * X_4\big),
\end{aligned}
$$

where $X_1 * X_2$ is the interaction term between the MI order and the first dummy variable of Age; $X_1 * X_3$ and $X_1 * X_4$ have similar interpretation;

$$
\begin{aligned}
\mathbf{X} &= (X_1, X_2, X_3, X_4, X_5, X_6, X_7, X_1 * X_2, X_1 * X_3, X_1 * X_4), \\
\text{and } \beta &= (\beta_1, \beta_2, \beta_3, \beta_4, \beta_5, \beta_6, \beta_7, \beta_8, \beta_9, \beta_{10}).
\end{aligned}
$$

Point estimate of the HR

$$
\begin{aligned}
& HR(\text{First time MI vs. MI recurrence among patients aged 60-70}) \\
= & \frac{h_0(t) \exp\big(\beta_1 \times 0 + \beta_2 \times 1 + \beta_3 \times 0 + \beta_4 \times 0 + \beta_5 x_5 + \beta_6 x_6 + \beta_7 x_7 + \beta_8 \times 0 + \beta_9 0 + \beta_{10} 0\big)}{h_0(t) \exp\big(\beta_1 \times 1 + \beta_2 \times 1 + \beta_3 \times 0 + \beta_4 \times 0 + \beta_5 x_5 + \beta_6 x_6 + \beta_7 x_7 + \beta_8 \times 1 + \beta_9 0 + \beta_{10} 0\big)} \\
= & \exp(-\beta_1 - \beta_8)
\end{aligned}
$$

The point estimate of this hazard ratio is

$$
\hat{HR} = \exp(-\hat{\beta}_1 - \hat{\beta}_8) = \exp(-1.09033 - 0.40491) = 0.224
$$

95% C.I. of the HR

$$
\begin{aligned}
Var(-\hat{\beta}_1 - \hat{\beta}_8) &= Var((-1) \times \hat{\beta}_1 + (-1) \times \hat{\beta}_8) \\
&= (-1)^2 Var(\hat{\beta}_1) + (-1)^2 Var(\hat{\beta}_8) + 2(-1)(-1) Cov(\hat{\beta}_1, \hat{\beta}_8) \\
&= Var(\hat{\beta}_1) + Var(\hat{\beta}_8) + 2Cov(\hat{\beta}_1, \hat{\beta}_8).
\end{aligned}
$$

thus, according to the estimated Variance-Covariance matrix of Model 3a,

$$
\begin{aligned}
\widehat{Var}(-\hat{\beta}_1 - \hat{\beta}_8) &= \widehat{Var}(\hat{\beta}_1) + \widehat{Var}(\hat{\beta}_8) + 2\widehat{Cov}(\hat{\beta}_1, \hat{\beta}_8) \\
&= 0.2302 + 0.4539 + 2 \times (-0.227) \\
&= 0.2301, \\
\hat{se}(\hat{\beta}_1 + \hat{\beta}_8) &= \sqrt{\widehat{Var}(\hat{\beta}_1 + \hat{\beta}_8)} \\
&= 0.4797.
\end{aligned}
$$

Then, the 95% C.I. of the HR can be estimated as:

$$\begin{aligned}\exp\Big(\big(-\hat{\beta}_1-\hat{\beta}_8\big)\pm 1.96\times \widehat{se}(-\hat{\beta}_1-\hat{\beta}_8)\Big) &= \exp\Big(-1.49524 \pm 1.96\times 0.4797\Big)\\ &= \big(\exp(-2.435452),\ \exp(-0.555028)\big)\\ &= \big(0.088,\ 0.574\big).\end{aligned}$$

Analysis of Maximum Likelihood Estimates											
Parameter			DF	Parameter Estimate	Standard Error	Chi-Square	Pr > ChiSq	Hazard Ratio	95% Hazard Ratio Confidence Limits		Label
miord	1		1	1.09033	0.47976	5.1649	0.0230	.	.	.	miord 1
age_c	2		1	0.20705	0.49322	0.1762	0.6746	.	.	.	age_c 2
age_c	3		1	1.90747	0.36561	27.2193	<.0001	.	.	.	age_c 3
age_c	4		1	2.36623	0.35174	45.2554	<.0001	.	.	.	age_c 4
year	2		1	0.20725	0.17375	1.4228	0.2329	1.230	0.875	1.729	year 2
year	3		1	0.64636	0.19217	11.3132	0.0008	1.909	1.310	2.782	year 3
bmi			1	-0.04925	0.01529	10.3695	0.0013	0.952	0.924	0.981	
age_c*miord	2	1	1	0.40491	0.67369	0.3612	0.5478	.	.	.	age_c 2 * miord 1
age_c*miord	3	1	1	-0.95485	0.54160	3.1083	0.0779	.	.	.	age_c 3 * miord 1
age_c*miord	4	1	1	-1.27263	0.51765	6.0440	0.0140	.	.	.	age_c 4 * miord 1

The 95% C.I. of the HR can also be obtained from SAS as follows:

$$\hat{HR} = (0.088, 0.574)$$

all pairs diff at 65-years-old: Hazard Ratios for miord			
Description	Point Estimate	95% Wald Confidence Limits	
miord 0 vs 1 At age_c=2	0.224	0.088	0.574

Interpretation: Among the patients aged between 60 and 70, the death rate of those with MI recurrence is estimated to be 4.4604 times that of those with the first MI, adjusted for the effects of Cohort year and BMI. Moreover, we are 95% confident that, among the patients aged between 60 and 70, the true death rate of the patients with MI recurrence could be as little as 1.741990 times or as much as 11.420980 times that of those with the first MI. (not required for this homework)

4. **Draw the survival curve for a patient with MI recurrence, who is older than 80, with BMI equal to 30, and in the 2001 cohort study, based on Model 3. What is the probability that this patient survives more than a year (365 days)? [3 points]**

The values of variables for this patient are: $t = 365,\ X_1 = 1,\ X_2 = X_3 = 0,\ X_4 = 1,\ X_5 = 0,\ X_6 = 1,\ X_7 = 30$. Then, based on the estimation of Model 3, the survival function at 365 days $S(365)$, which is equal to $S(363)$, can be estimated by $\widehat{S(363)} = 0.5418$. So the probability that the patient survives more than a year is estimated to be 0.5418.

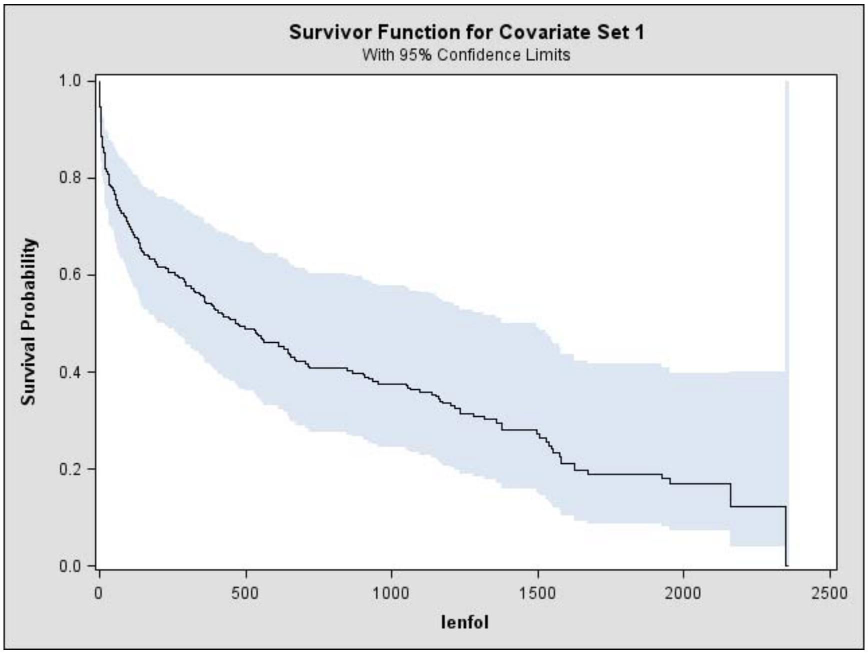

Figure 2: The survival curve

SAS code:

```
data hw5;
infile 'HW5data.csv' delimiter=',' missover dsd;
input id lenfol fstat miord bmi year age_c;
run;
proc phreg data=hw5;
model lenfol*fstat(0)=miord/alpha=0.05 risklimits ties=efron;
title 'Model 1: MI ORDER';
run;
proc phreg data=hw5;
class age_c(ref=first) year(ref=first) miord(ref=first)/param=ref;
model lenfol*fstat(0)=miord age_c year bmi/alpha=0.05 risklimits ties=efron;
title 'Model 2: MI ORDER+AGE+CohortYEAR+BMI';
run;
proc phreg data=hw5;
class age_c(ref=first) year(ref=first) miord(ref=first)/param=ref;
model lenfol*fstat(0)=miord age_c year bmi miord*age_c/alpha=0.05 risklimits covb ties=efron;
title 'Model 3: MI ORDER+AGE+CohortYEAR+BMI+miord*age_c';
run;
proc phreg data=hw5;
class age_c(ref=first) year(ref=first) miord(ref=first)/param=ref;
model lenfol*fstat(0)=miord age_c year bmi miord*age_c/alpha=0.05 risklimits covb ties=efron;
hazardratio 'all pairs diff at 65-years-old' miord / at (age_c="2") diff=ALL;
title 'Hazzard Ratio and 95% CI';
run;
data hw5_2;
input miord age_c year bmi;
cards;
1 4 3 30
;
ods graphics on;
proc phreg data=hw5 plots(cl)=s;
class age_c(ref=first) year(ref=first) miord(ref=first)/param=ref;
model lenfol*fstat(0)=miord age_c year bmi miord*age_c/alpha=0.05 risklimits ties=efron;
baseline covariates=hw5_2 out=pred survival=_all_;
title 'Prediction';
run;
ods graphics off;
proc print data=pred; run;
```