

## Biostatistics 209 Homework #1 Solutions

For Questions 1-3, use the dataset `pbcc.dta`.

**Question 1:** Fit a Cox proportional hazards model to the PBC data with terms for histology and bilirubin

```
. xi: stcox bilirub i.histol  
. est store A
```

```
No. of subjects =          312          Number of obs =          312  
No. of failures =           125  
Time at risk   = 1713.853528  
Log likelihood = -576.60724          LR chi2(4) =          126.75  
                                          Prob > chi2 =           0.0000
```

<code>_t</code>	Haz. Ratio	Std. Err.	z	P> z	[95% Conf. Interval]	
bilirub	1.158085	.0161909	10.50	0.000	1.126782	1.190257
<code>_Ihistol_2</code>	4.460183	4.60125	1.45	0.147	.5905127	33.68807
<code>_Ihistol_3</code>	6.602234	6.696725	1.86	0.063	.9042883	48.20309
<code>_Ihistol_4</code>	16.03493	16.20056	2.75	0.006	2.213465	116.1612

(a) Adjusting for bilirubin, does histology appear to be predictor of risk of death?

*It appears that way but it can't be tested for directly from the output above. There is a steady increase in the hazard of death with each increase in the grade of histology and histology grade 4 is statistically different from grade 1 ( $p < 0.01$ ).*

*The following is the overall Wald test to verify that.*

```
. testparm _Ihistol_*  
( 1)  _Ihistol_2 = 0  
( 2)  _Ihistol_3 = 0  
( 3)  _Ihistol_4 = 0  
      chi2( 3) =    36.86  
      Prob > chi2 =    0.0000
```

(b) Implement the likelihood ratio test to verify (a).

```
. xi: stcox bilirub  
. lrtest A  
Likelihood-ratio test          LR chi2(3) =          42.15  
(Assumption: B nested in A)  Prob > chi2 =           0.0000
```

*The likelihood ratio test (which is an overall) test finds that histology is a significant predictor of survival; however, it does not give us the direction of the association. We need to look at the hazard ratios to discern that high grades of histology are associated with higher risk of death.*

(c) Calculate the HR and its confidence interval of death comparing Grade 2 histology to Grade 1.

*This is available off the Stata output above. Subjects with grade 2 (adjusting for bilirubin) have 4.5 times the hazard of death compared to subjects with grade 1. The 95% confidence interval is from 0.6 to 33.7.*

(d) Calculate the HR and its confidence interval of death comparing Grade 3 histology to Grade 2.

```
. lincom _Ihistol_3 - _Ihistol_2, hr
```

_t	Haz. Ratio	Std. Err.	z	P> z	[95% Conf. Interval]
(1)	1.480261	.4372279	1.33	0.184	.829692 2.640946

*Subjects with grade 3 (adjusting for bilirubin) have 1.5 times the hazard of death compared to subjects with grade 2. The 95% confidence interval is from 0.8 to 2.6. It seems much worse to go from grade 1 to 2 than from 2 to 3.*

(e) Calculate the HR and its confidence interval of death comparing Grade 4 histology to Grade 3.

```
. lincom _Ihistol_4 - _Ihistol_3, hr
```

_t	Haz. Ratio	Std. Err.	z	P> z	[95% Conf. Interval]
(1)	2.428712	.4809663	4.48	0.000	1.64744 3.58049

*Subjects with grade 4 (adjusting for bilirubin) have 2.4 times the hazard of death compared to subjects with grade 3. The 95% confidence interval is from 1.6 to 3.6.*

(f) Calculate a test for trend and interpret the result.

```
. test -1* _Ihistol_2 + _Ihistol_3 + 3* _Ihistol_4=0
```

chi2( 1) = 8.18  
Prob > chi2 = 0.0042

*There is a significant trend toward shorter survival with increasing grade of histology ( $p < 0.01$ ). Note, the test for departure from a linear trend is not significant ( $p = 0.42$ , results not shown but you can do it as shown in Lab #1).*

**Question 2:** In the same model as you used in Question 1,

(a) Calculate the HR and its confidence interval of death for every standard deviation increase in bilirubin.

The hazard of death increases 1.94-fold for every standard deviation (SD) increase in bilirubin. The 95% confidence interval is from 1.71 to 2.20.

Note, you can get this by (1) finding the SD of bilirubin =4.530315 (e.g. `summ bilirub`) and doing `lincom 4.530315 *bilirub, hr`

Or (2) creating a standardized variable and re-run the Cox model:

```
. egen s_bili=std(bilirub)      (this standardizes a variable and saves it a new name)
. xi: stcox s_bili i.histol
```

(b) How many units (mg/dL) increase in bilirubin will result in a HR of death in the same magnitude as histology of Grade 2 compared to Grade 1?

The effect of a 10.18-point increase in bilirubin is a 4.46 fold increase in the hazard of death, which is the same magnitude as histology of grade 2 compared with grade 1.

Note, you can get this by (i) trying a series of `lincom tryvalue *bilirub, hr` until you get a value close to 4.46, e.g. from (a) you find a 4.53 unit increase will get you a HR just about 2 fold, so maybe try values like 10, 10.1, 10.2 ...etc.

Or (ii) you can solve it by using HR values from the model results above and calculating  $\log(4.4602)/\log(1.1581)=10.19$ , subject to rounding error!

Or (iii) use Stata to calculate the value:

```
. matrix coeff=e(b)          (this saves the log-HR coefficients to the variable coeff)
. matrix list coeff          (this displays the variables and their corresponding values)
coeff[1,4]
      bilirub  _Ihistol_2  _Ihistol_3  _Ihistol_4
y1      .14676752   1.4951898   1.8874081   2.7747692

. display coeff[1,2]/coeff[1,1]
10.18747
```

**Question 3:** In the same model as you used in Question 1, calculate the likelihood ratio test for the effect of bilirubin (adjusting for histology). How do the results compare to the Wald test (aka Z test)?

```
. xi: stcox i.histol
. lrtest A
Likelihood-ratio test          LR chi2(1) =      74.03
(Assumption: B nested in A)   Prob > chi2 =      0.0000
```

The likelihood ratio test for bilirubin, like the Wald test, is overwhelmingly significant ( $p < 0.001$ ).

For Question 4, use the dataset `actg019.dta`

**Question 4:** Problem 7.6 in the textbook. **Write out** the Cox model allowing for an interaction between ZDV treatment `rx` and the baseline CD4 cell counts `cd4`.

The Cox model is

$$h(\text{days} \mid \text{rx}, \text{cd4}, \text{interaction}) = h_0(\text{days}) * \exp(\beta_1 * \text{rx} + \beta_2 * \text{cd4} + \beta_3 * \text{interaction})$$

(a) **Express** the test of the null hypothesis of no interaction between CD4 and treatment in terms of the **parameters** of the model.

The test of the null hypothesis of no interaction is

$$H_0: \beta_3 = 0$$

(b) Again, using the **parameters** of the mode, what is the hazard ratio for a ZDV-treated subject with  $x$  CD4 cells compared with a placebo-treated subject with  $x$  CD4 cells?

First, you can find out  $\text{rx} = 1$  corresponds to the ZDV group by

```
. desc rx          (OR label dir)
. label list trt
```

Then the hazards for subjects on ZDV with CD4 equals to  $x$  is

$$h_0(\text{days}) * \exp(\beta_1 + \beta_2 * x + \beta_3 * x)$$

If they are on placebo, their hazards is

$$h_0(\text{days}) * \exp(\beta_2 * x)$$

The HR is then

$$\exp(\beta_1 + \beta_3 * x) = \exp(\beta_1) * \exp(\beta_3)^x = (\text{HR}_1) * (\text{HR}_3)^x$$

(c) Fit the model. Does there appear to be an interaction between treatment and CD4 stratum? If so, what is the interpretation?

```
. gen interaction=rx*cd4
. stcox rx cd4 interaction , nolog
```

```
No. of subjects =          880          Number of obs   =          880
No. of failures =           55
Time at risk    =        354872
Log likelihood   =   -311.88738          LR chi2(3)       =         39.03
                                          Prob > chi2      =         0.0000
```

_t	Haz. Ratio	Std. Err.	z	P> z	[95% Conf. Interval]
rx	1.835566	1.286131	0.87	0.386	.4648965 7.247423
cd4	.995174	.0014709	-3.27	0.001	.9922952 .9980612
interaction	.9940859	.0028374	-2.08	0.038	.9885401 .9996628

We do see a significant interaction ( $p=0.04$ ). The direction of the interaction is one where the treatment effect increases (which corresponds a decreasing HR) as the CD4 count increases.

(d) What are the hazard ratios for ZDV as compared to placebo for patients with 500, 109, and 50 CD4 cells, respectively?

The hazard ratios are 0.095 (a 90% reduction in risk due to ZDV), 0.96 (no real treatment effect) and 1.36 (an increase in risk on ZDV) for patients with 500, 109, and 50 CD4 cells, respectively. You can get these values from the following

```
. lincom rx + 500 * interaction, hr
```

```
( 1) rx + 500 interaction = 0
```

_t	Haz. Ratio	Std. Err.	z	P> z	[95% Conf. Interval]
(1)	.0945642	.0811772	-2.75	0.006	.0175803 .5086605

```
. lincom rx + 109 * interaction, hr
```

```
( 1) rx + 109 interaction = 0
```

_t	Haz. Ratio	Std. Err.	z	P> z	[95% Conf. Interval]
(1)	.9615602	.4271512	-0.09	0.930	.4025776 2.296695

```
. lincom rx + 50 * interaction, hr
```

```
( 1) rx + 50 interaction = 0
```

_t	Haz. Ratio	Std. Err.	z	P> z	[95% Conf. Interval]
(1)	1.364475	.7864717	0.54	0.590	.4409002 4.222706

Or plug in values in the HR formula in (b): The hazard ratios are

$$\exp(\log(1.8356) + \log(.9941)*500) = 1.8356 * .9941^{500} = 0.095,$$

$$\exp(\log(1.8356) + \log(.9941)*109) = 1.8356 * .9941^{109} = 0.963, \text{ and}$$

$$\exp(\log(1.8356) + \log(.9941)* 50) = 1.8356 * .9941^{50} = 1.365.$$