Biostatistics 209 Homework #1 Solutions

Question 1 [5 points]: Fit a Cox proportional hazards model to the PBC data with terms for histology and bilirubin

```
. stcox bilirub i.histol

No. of subjects =          312                     Number of obs    =       312
No. of failures =          125
Time at risk    =  1713.853528
                                                   LR chi2(4)       =    126.75
Log likelihood  =   -576.60724                     Prob > chi2      =    0.0000
------------------------------------------------------------------------------
          _t | Haz. Ratio   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
     bilirub |   1.158085   .0161909    10.50   0.000     1.126782    1.190257
      histol |
          2  |   4.460183    4.60125     1.45   0.147     .5905127    33.68807
          3  |   6.602234   6.696725     1.86   0.063     .9042883    48.20309
          4  |   16.03493   16.20056     2.75   0.006     2.213465    116.1612
------------------------------------------------------------------------------
```

(a) [2 points] Adjusting for bilirubin, does histology appear to be a predictor of risk of death? Also implement the likelihood ratio test to verify it and summarize the result of the test.

It appears that way but it can't be tested for directly from the output above. There is a 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 i.histol

 ( 1)  2.histol = 0
 ( 2)  3.histol = 0
 ( 3)  4.histol = 0

           chi2(  3) =   36.86
         Prob > chi2 =    0.0000
```

```
. est store A

. stcox bilirub

. lrtest A

Likelihood-ratio test                                 LR chi2(3)  =     42.15
(Assumption: . 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.

(b) [2 points] Calculate the **three** HRs and their confidence intervals of death: Grade **1** histology relative to Grade 2, Grade **2** relative to Grade 3, Grade **3** histology relative to Grade 4. Give a brief interpretation of the results.

```
. lincom -2.histol, hr
------------------------------------------------------------------------------
          _t | Haz. Ratio   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
         (1) |    .224206   .2312972    -1.45   0.147     .0296841    1.693444
------------------------------------------------------------------------------

. lincom  2.histol-3.histol, hr
------------------------------------------------------------------------------
          _t | Haz. Ratio   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
         (1) |   .6755567   .1995407    -1.33   0.184     .3786522    1.205267
------------------------------------------------------------------------------

. lincom  3.histol-4.histol, hr
------------------------------------------------------------------------------
          _t | Haz. Ratio   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
         (1) |   .4117409   .0815385    -4.48   0.000     .2792914    .6070023
------------------------------------------------------------------------------
```

Subjects with grade 1 (adjusting for bilirubin) have 78% decrease in the hazard of death compared to subjects with grade 2. The 95% confidence interval is from 97% decrease to 69% increase.
Subjects with grade 2 (adjusting for bilirubin) have 32% decrease in the hazard of death compared to subjects with grade 3. The 95% confidence interval is from 62% decrease to 21% increase. It seems worse to go from grade 1 to 2 than from 2 to 3.
Subjects with grade 3 (adjusting for bilirubin) have 59% decrease in the hazard of death compared to subjects with grade 4. The 95% confidence interval is from 72% decrease to 39% decrease.

(c) [1 point] Perform a test for trend and interpret the result, providing direction of effect.

New approach

```
. contrast p.histol

Contrasts of marginal linear predictions

Margins      : asbalanced

------------------------------------------------
             |         df        chi2     P>chi2
-------------+----------------------------------
      histol |
   (linear)  |          1        8.18     0.0042
(quadratic)  |          1        0.33     0.5629
    (cubic)  |          1        1.43     0.2325
      Joint  |          3       36.86     0.0000
```

```
-----------------------------------------------
-------------------------------------------------------------
             |   Contrast   Std. Err.     [95% Conf. Interval]
-------------+-----------------------------------------------
      histol |
   (linear)  |   .9745372   .3407443      .3066906    1.642384
(quadratic)  |  -.1519572   .2626469     -.6667357    .3628213
    (cubic)  |   .1786746   .1496485     -.1146311    .4719804
-------------------------------------------------------------
```

Old approach

```
. test -2.histol + 3.histol + 3* 4.histol=0

           chi2(  1) =    8.18
         Prob > chi2 =    0.0042
```

There is a statistically significant trend toward shorter survival with increasing grade of histology (p<0.01).

Question 2 [3 points]: In the same model as you used in Question 1,

(a) [1 point] Similar to 1(a), implement the likelihood ratio test for the effect of bilirubin (adjusting for histology). How do the results compare to the Wald (aka Z test)?

```
. stcox i.histol

. lrtest A

Likelihood-ratio test                                 LR chi2(1)  =     74.03
(Assumption: . nested in A)                           Prob > chi2 =    0.0000
```

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

(b) [1 point] 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)

. stcox s_bili  i.histol
```

(c) [1 point] 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 2(b) 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[1,5]
                      1b.          2.          3.          4.
      bilirub     histol      histol      histol      histol
y1   .14676752          0   1.4951898   1.8874081   2.7747692

. display coeff[1,3]/coeff[1,1]
10.18747
```

Question 3 [5 points]: Taken from VGSM: problem 6.8 (new edition) / 7.6 (old edition).

The Cox model allowing for an interaction between ZDV treatment `rx` and the baseline CD4 cell counts `cd4 is`:

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

(a) [1 point] **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

$$H_0: \beta_3 = 0 \quad or \quad HR_3 = \exp(\beta_3) = 1$$

(b) [1 point] Again, using the **parameters** of the model, 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 `rx` *=1 corresponds to the ZDV group by*

```
. desc rx        (or using label dir)

. label list trt                          0=Placebo, 1=ZDV
```

Then the hazards for subjects on ZDV with CD4 equals to $\boldsymbol{x}$ *is*

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

If they are on placebo, their hazard is

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

The HR is then

$$\exp(\beta_1 + \beta_3 * \boldsymbol{x}) = \exp(\beta_1) * \exp(\beta_3)^{\boldsymbol{x}} = (HR_1) * (HR_3)^{\boldsymbol{x}}$$

(c) [2 points] Fit the model. Does there appear to be an interaction between treatment and CD4 stratum? If so, what is the interpretation?

```
. stset days, failure(cens)

. stcox rx##c.cd4, nolog    (Alternatively, generate interaction by gen interaction=rx*cd4)

No. of subjects =          880                     Number of obs    =       880
No. of failures =           55
Time at risk    =       354872
                                                   LR chi2(3)       =     39.03
Log likelihood  =   -311.88738                     Prob > chi2      =    0.0000
------------------------------------------------------------------------------
          _t | Haz. Ratio   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
        1.rx |   1.835566   1.286131     0.87   0.386     .4648965    7.247423
         cd4 |    .995174   .0014709    -3.27   0.001     .9922952    .9980612
             |
    rx#c.cd4 |
          1  |   .9940859   .0028374    -2.08   0.038     .9885401    .9996628
------------------------------------------------------------------------------
```

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

(d) [1 point] 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 (a statistically non-significant increase in risk on ZDV) for patients with 500, 109, and 50 CD4 cells, respectively. You can get these values from the following

```
. lincom 1.rx + 500*1.rx#c.cd4, hr

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

. lincom 1.rx + 109*1.rx#c.cd4, hr

------------------------------------------------------------------------------
          _t | Haz. Ratio   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
         (1) |   .9615602   .4271512    -0.09   0.930     .4025776    2.29669
------------------------------------------------------------------------------
```

```
. lincom 1.rx + 50*1.rx#c.cd4, hr

------------------------------------------------------------------------------
          _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, and

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