**Biost 518: Applied Biostatistics II**

**Biost 515: Biostatistics II**

Emerson, Winter 2015

**Homework #4**

February 2, 2015

**Written problems:** To be submitted as a MS-Word compatible file to the class Catalyst dropbox by 9:30 am on Monday, February 9, 2014. See the instructions for peer grading of the homework that are posted on the web pages.

*On this (as all homeworks) Stata / R code and unedited Stata / R output is* ***TOTALLY*** *unacceptable. Instead, prepare a table of statistics gleaned from the Stata output. The table should be appropriate for inclusion in a scientific report, with all statistics rounded to a reasonable number of significant digits. (I am interested in how statistics are used to answer the scientific question.)*

***Unless explicitly told otherwise in the statement of the problem, in all problems requesting “statistical analyses” (either descriptive or inferential), you should present both***

* ***Methods: A brief sentence or paragraph describing the statistical methods you used. This should be using wording suitable for a scientific journal, though it might be a little more detailed. A reader should be able to reproduce your analysis. DO NOT PROVIDE Stata OR R CODE.***
* ***Inference: A paragraph providing full statistical inference in answer to the question. Please see the supplementary document relating to “Reporting Associations” for details.***

This homework investigates associations between death from any cause and age, sex, and serum bilirubin in a population of patients with primary biliary cirrhosis who were enrolled in a randomized clinical trial (RCT) of D-penicillamine. The data can be found on the class web page (follow the link to Datasets) in the file labeled liver.txt. Documentation is in the file liver.doc.

1. Provide suitable descriptive statistics pertinent to the scientific questions addressed in this homework.

**Methods:** Patients were enrolled at the Mayo Clinic for a randomized control trial of using D-penicillamine in treatment for primary biliary cirrhosis (PBC). Of the 418 subjects enrolled, 106 were not assigned to either treatment arm and were missing many data and have been excluded from all analyses. The 312 remaining subjects were categorized based on serum bilirubin levels (along a multiplicative scale, keeping the fold change in each group approximately constant). The levels were originally based on normal levels of 0.3 – 1.9 (as reported on nlm.nih.gov/medlinePlus); low normal (below 0.7 mg/dL), high normal (0.8 – 1.9 mg/dL), mild hyperbilirubinemia (2.0 – 4.6 mg/dL), moderate hyperbilirubinemia (4.7 – 9.3 mg/dL), and sever hyperbilirubinemia (greater than or equal to 9.3 mg/dL). Kaplan-meier curves were created for each stratum, and the number of days at the survival probability was 0.75 was calculated (25th %ile survival).

**Results:** Descriptive statistics on the 312 subjects that were randomized to a treatment arm in an RCT for D-penicillamine at the Mayo Clinic are presented in the table below. There were no missing data in the variables presented among these subjects. The mean age across the groups defined by serum bilirubin were relatively similar, with no clear trends. The percentage of females in each category decreased as serum bilirubin increased, except at the highest category where all subjects were female. Interestingly, although it was a randomized controlled trial, the percentage of subjects assigned to the D-penicillamine trial in the highest bilirubin group was well below 50%. The Kaplan Meier curves of survival in each of the serum bilirubin categories shows markedly decreased survival probabilities in the groups with the higher bilirubin levels. As a quick quantification of this effect, the time to 25th %ile survival (i.e. 0.75 survival probability) was drastically different for the group with lowest bilirubin (where survival probabilities didn’t reach 0.75 in over 4000 days of follow up) and the group with the highest levels (174 days).

|  |  |  |
| --- | --- | --- |
|  | **Subjects Categorized by Serum Bilirubin Levels** | **Across all levels** |
| **Serum bilirubin (mg/dL)** | **0.3 – 0.7** | **0.8 – 1.9** | **2.0 – 4.6** | **4.7 – 9.3** | **9.3 – 28.0** | **All** |
| **Number subjects** | 73 | 108 | 69 | 38 | 24 | 312 |
| **Age (y)** | 50.0 (9.6, 32 - 73) | 50.0 (10.6, 26 - 77) | 49.7 (11.3, 30 - 72) | 50.5 (11.8, 31 - 78) | 50.7 (10., 33 - 71) |

|  |
| --- |
| 50.0 (10.6, 26 - 78) |

 |
| **% Female** | 94.5 | 88.9 | 81.2 | 81.2 | 100 | 88.5 |
| **% Assigned to D-penicillamine** | 47.2 | 54.2 | 55.1 | 47.4 | 37.5 | 50.6 |
| **25th %ile Survival (days)** | NA | 2,689 | 999 | 980 | 174 | 1487 |

1. In prior homeworks using the Cardiovascular Health Study datasets, we were able to use logistic regression to investigate associations between mortality and various covariates. Why might such an approach not seem advisable with these data? (Consider the extent to which such analyses might be confounded and/or lack precision.)

**Answer:** In previous homeworks, we were often working with saturated or near-saturated models, i.e. our model had the same number of parameters as distinct groups and thus we could rest assured that the model was a good fit. In this data set, there are many more variables and these are slightly less biologically clear-cut. Many of the variables are surrogates of ill-defined phenomena (e.g. alkphos levels for liver health or serum bilirubin for liver function), and as such will have more variance and might be interacting with our data in more complicated ways. The fact that the overall rate of death is ~40% also means that there will tend to be relatively high variance in our response variable. Since logistic regression is modeling the odds ratio, having a p close to 0.5 (and thus a censoring rate close to 0.5) will be problematic for logistic regression.

1. Perform a statistical regression analysis evaluating an association between serum bilirubin and all-cause mortality by comparing the instantaneous risk (hazard) of death over the entire period of observation across groups defined by serum bilirubin modeled as a continuous variable.
	1. Include a full report of your inference about the association.

**Methods:**  As in question 1, the 312 subjects with complete data were analyzed in order to address whether there existed an association between serum bilirubin and all-cause mortality. A cox proportional hazards model with Huber-White sandwich estimator (robust standard error) was applied with death as the response variable and the predictor of interest being untransformed serum bilirubin. The z statistic (Wald test) was used to calculate a p-value with a threshold of 0.05, and Wald based confidence intervals were used to calculate the 95% CI.

**Results:** Of the 312 subjects included in this analysis, 125 were observed to die in the study period. From the Cox proportional hazards regression model, subjects had a relative 16.1% greater (hazard ratio = 1.161) instantaneous risk of dying for every increase of 1 mg/dL in serum bilirubin. This observation would not be atypical if the true increased risk of death was between 12.6% and 19.6% higher (95% CI hazard ratio 1.126 – 1.196) for every 1 mg/dL increase in serum bilirubin. Based on the two-sided p-value <0.001, we can reject the null hypothesis that there is no association between serum bilirubin levels and instantaneous risk of death (that the hazard ratio was constant across serum bilirubin levels), and we can say that the hazard ratio tends to increase for increasing serum bilirubin.

* 1. For each population defined by serum bilirubin value, compute the hazard ratio relative to a group having serum bilirubin of 1 mg/dL. (This will be used in problem 6). If *HR* is the hazard ratio (use the actual hazard ratio estimate) obtained from your regression model, this can be effected by the Stata code

gen fithrA = *HR ^ (bili* – 1)

It could also be computed by creating a centered bilirubin variable, and then using the Stata predict command

 gen cbili = bili – 1

stcox cbili

predict fithrA

1. Perform a statistical regression analysis evaluating an association between serum bilirubin and all-cause mortality by comparing the instantaneous risk (hazard) of death over the entire period of observation across groups defined by serum bilirubin modeled as a continuous logarithmically transformed variable.
	1. Why might this analysis be preferred *a priori?*

**Answer:** Serum bilirubin is a biomarker of liver function and based on the mechanism of its accumulation, it seems as if it will be working on a multiplicative scale, i.e. with a little liver malfunction you might have low levels of serum bilirubin, but as you decrease liver function you can get large accumulations such that fold change, rather than absolute difference defines progression. Because of this, it is more natural to talk about a “doubling” of serum bilirubin as opposed to an absolute “1 mg/dL” increase. In addition to this scientific reason, we will generally have more precision and be less worried about confounding after a log-transformation of our data.

* 1. Include a full report of your inference about the association.

**Methods:**  As in question 1, the 312 subjects with complete data were analyzed in order to address whether there existed an association between serum bilirubin and all-cause mortality. The serum bilirubin was first logarithmically transformed (base 2) and then a cox proportional hazards model with Huber-White sandwich estimator (robust standard error) was applied with death as the response variable and the predictor of interest being the log-transformed serum bilirubin. The z statistic (Wald test) was used to calculate a p-value with a threshold of 0.05, and Wald based confidence intervals were used to calculate the 95% CI.

**Results:** Of the 312 subjects included in this analysis, 125 were observed to die in the study period. From the Cox proportional hazards regression model, subjects had a hazard ratio of 2.122 (i.e. a relative 112.2% greater instantaneous risk of dying) for each doubling of serum bilirubin. This observation would not be atypical if the true hazard ratio was between 1.881 and 2.393 for every doubling in serum bilirubin (i.e. 95% CI is a relative 88.1% to 139.3% increase in instantaneous risk of death for each doubling of serum bilirubin). Based on the two-sided p-value <0.001, we can reject the null hypothesis that there is no association between serum bilirubin levels and instantaneous risk of death (that the hazard ratio was constant across serum bilirubin levels), and we can say that the hazard ratio tends to increase for increasing serum bilirubin.

* 1. For each population defined by serum bilirubin value, compute the hazard ratio relative to a group having serum LDL of 1 mg/dL. (This will be used in problem 6). If *HR* is the hazard ratio (use the actual hazard ratio estimate) obtained from your regression model, this can be effected by the Stata code

gen logbili = log(bili)

stcox logbili

fithrB = *HR ^ (logbili*)

(Note that the log(1) = 0 when using any base, so there is no need to rescale by the bilirubin values. Note also that you might want to use a different base in your logarithmic transformation in order to facilitate more natural reporting of effects.)

1. One approach to testing to see whether an association between the response and the predictor of interest is adequately modeled by an untransformed continuous variable is to add some other transformation to the model and see if that added covariate provides statistically significant improved “fit” of the data. In this case, we could test for “linearity” of the bilirubin association with the log hazard ratio by including both the untransformed and log transformed bilirubin. (Other alternatives might have been bilirubin and bilirubin squared, but in this case our *a priori* interest in the log bilirubin might drive us to the specified analysis.)
	1. Provide full inference related to the question of whether the association is linear.

**Methods:**  As in question 1, the 312 subjects with complete data were analyzed in order to address whether there existed an association between serum bilirubin and all-cause mortality. The serum bilirubin was first logarithmically transformed (base 2) and then a cox proportional hazards model with Huber-White sandwich estimator (robust standard error) was applied with death as the response variable and the predictor of interest being the log-transformed serum bilirubin with an additional covariate added, the untransformed serum bilirubin. The z statistic (Wald test) was used to calculate a p-value with a threshold of 0.05, and Wald based confidence intervals were used to calculate the 95% CI.

**Results:** Of the 312 subjects included in this analysis, 125 were observed to die in the study period. From the Cox proportional hazards regression model, subjects had a hazard ratio of 2.323 (i.e. a relative 132.3% greater instantaneous risk of dying) for each doubling of serum bilirubin. This observation would not be atypical if the true hazard ratio was between 1.839 and 2.936 for every doubling in serum bilirubin. Based on the two-sided p-value <0.001, we can reject the null hypothesis that there is no association between serum bilirubin levels and instantaneous risk of death (that the hazard ratio was constant across serum bilirubin levels), and we can say that the hazard ratio tends to increase for increasing serum bilirubin. To assess whether the additional untransformed bilirubin provided better “fit” for our model, a Wald test on whether the coefficient for this parameter was not zero. This test yielded a p-value of 0.324, indicating that we can not reject the null hypothesis that this added term has a non-zero slope (i.e. that the instantaneous risk of death is not modeled “well-enough” by the log transformed serum bilirubin). With a quick check of the model R2 to address the linearity of the association, we can see that adding the untransformed serum bilirubin as a covariate in this model produced an R2 that was only slightly higher (0.349 vs 0.347 for the log-transformed model alone), indicating that it didn’t do much for the linearity of fit.

* 1. Again, save the fitted values from this model by obtaining the estimated HRs relative to a group with bilirubin of 1 mg/dl. (This will be used in problem 6.)
1. Display a graph with the fitted hazard ratios from problems 3 - 5. Comment on any similarities or differences of the fitted values from the three models.

See the graph on the next page. All three models show increasing hazard as serum bilirubin increased. All three models are relatively consistent fitted hazard ratios for very low serum bilirubin (below 1 or 2 mg/dL), but the model from question #3 (i.e. untransformed bilirubin) quickly diverges from the other two models. The untransformed model shows a curvilinear increase in the hazard ratio as serum bilirubin increases. The other two models are more similar to each other, and show a relatively linear increase as serum bilirubin increases. The log-transformed model does appear more linear on these axes than the model that added the non-transformed term.

1. We are interested in considering analyses of the association between all cause mortality and serum bilirubin after adjustment for age and sex.
	1. What evidence is present in the data that would make you think that either sex or age might have confounded the association between death and bilirubin? (In real life, we would ideally decide whether to adjust for potential confounding in our pre-specified statistical analysis plan (SAP)).

**Answer:** The potential association of gender with serum bilirubin levels (POI) in our sample (as can be seen from the descriptive statistics table) satisfies one requirement of confounding. The propensity of females for more severe auto-immune diseases (one of the hypothesized mechanisms of PBC) might suggest that there is a causal relationship between sex and all-cause mortality that lies outside of the association with bilirubin. This second is hard to see in our data.

* 1. What evidence is present in the data that would make you think that either sex or age might have added precision to the analysis of the association between death and bilirubin? (In real life, we would ideally decide whether to adjust in our pre-specified SAP).

**Answer:** The key here is that precision variables can decrease standard errors. There is a wide range of ages across groups defined for serum bilirubin. Although there doesn’t seem to be a strong association between age and bilirubin based on this analysis, it is reasonable to believe that age is causally associated with death (in a manner independent and dependent on serum bilirubin levels). This causal association with response without association with POI in our sample suggests that we will gain precision by accounting for it.

* 1. Provide full inference regarding an association between death and bilirubin after adjustment for sex and age.

**Methods:**  As in question 1, the 312 subjects with complete data were analyzed in order to address whether there existed an association between serum bilirubin and all-cause mortality. The serum bilirubin was first logarithmically transformed (base 2) and then a cox proportional hazards model with Huber-White sandwich estimator (robust standard error) was applied with death as the response variable and the predictor of interest being the log-transformed serum bilirubin. Additional covariates added to this model were the subjects’ age at baseline and the sex. The z statistic (Wald test) was used to calculate a p-value with a threshold of 0.05, and Wald based confidence intervals were used to calculate the 95% CI for each coefficient.

**Results:** Of the 312 subjects included in this analysis, 125 were observed to die in the study period. From the Cox proportional hazards regression model, after accounting for age and sex, subjects had a hazard ratio of 2.109 for each doubling of serum bilirubin. This observation would not be atypical if the true hazard ratio was between 1.840 and 2.417 for every doubling in serum bilirubin. Based on the two-sided p-value <0.001, we can reject the null hypothesis that there is no association between serum bilirubin levels and instantaneous risk of death (that the hazard ratio was constant across serum bilirubin levels) after accounting for age and sex, and we can say that the hazard ratio tends to increase for increasing serum bilirubin. In addition, even as bilirubin and sex are held constant, the age of the subject is associated with their instantaneous risk of dying, as evidenced by a statistically significant non-zero coefficient in this model (p = 0.00016). In fact, the hazard ratio is 1.461 (i.e. the instantaneous risk of death increases a relative 46.1%) for every 10 year increase in age, after accounting for sex and serum bilirubin (95% CI 1.200 – 1.774 for every 10 year increase).

1. Note that in the above analyses, we completely ignored the intervention in the RCT? What impact could this have had on our results?

**Answer:** Since this is a randomized controlled trial, it would be nice to assume that within each subgroup there were exactly even proportions on subjects randomized to the treatment arm. Without stratified randomization, however, there is likely to be stochastic variations. If we expect D-penicallamine to produce an effect on all-cause mortality (hopefully strongly diminished), then even small variations can confound any analysis on bilirubin. As can be seen by table 1 of the descriptive statistics, there does seem to be non-uniform distribution of treatment groups across serum bilirubin levels.