Biost 518: Applied Biostatistics II
Emerson, Winter 2007

Homework #3 Key
January 31, 2007

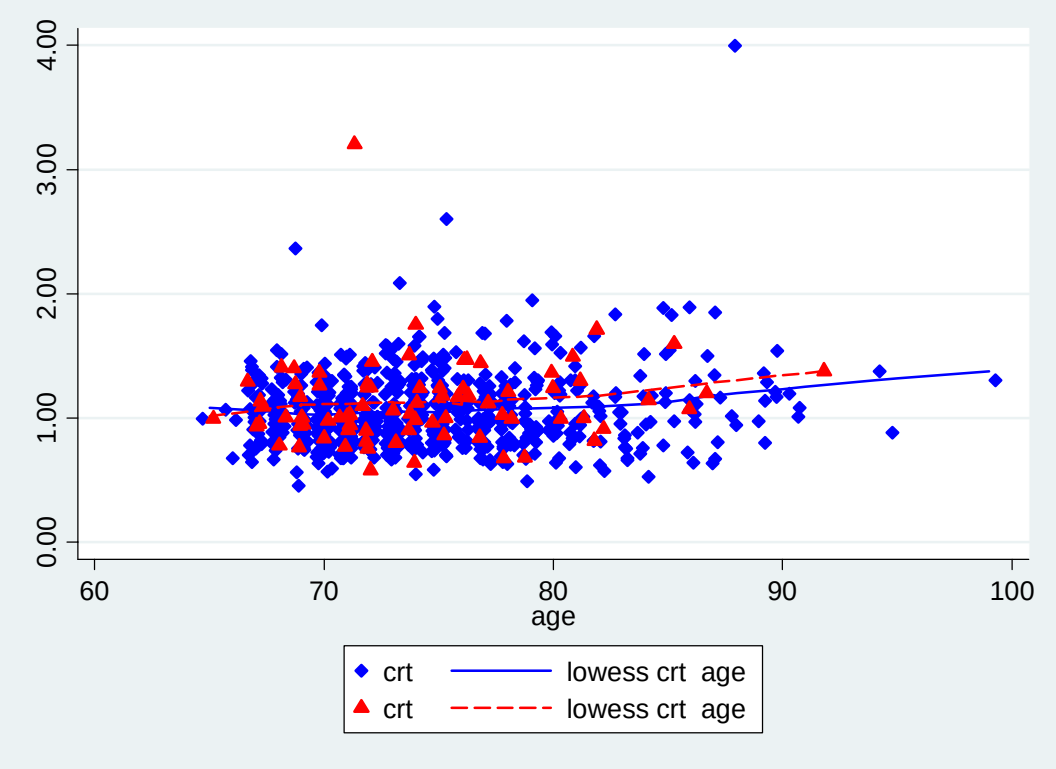
The following questions pertain to the dataset for MRI and cerebral atrophy stored as mri.txt on the class web page. For all questions involving statistical inference, provide estimates, confidence intervals, and P values in text suitable for a scientific journal.

1. We are interested in examining how mean creatinine levels vary by age and prevalence of diabetes.
    a. Provide suitable descriptive statistics regarding the distribution of creatinine levels by age and prevalence of diabetes.

Ans: From the following table and graph, we see a tendency for increasing mean creatinine levels with increasing age in both diabetics and nondiabetics. The trend appears a bit more consistent in diabetics. Diabetics tend to have slightly higher creatinine for each age group.

Table 1: Descriptive statistics for C reactive protein by sex and within age strata for each sex.

	N	Mean	SD	Min	25%ile	Median	75%ile	Max
Nondiabetes	654	1.06	0.30	0.50	0.90	1.00	1.20	4.00
65- 69 yo	99	1.05	0.26	0.50	0.90	1.00	1.20	2.40
70- 74 yo	275	1.04	0.25	0.60	0.80	1.00	1.20	2.10
75- 79 yo	168	1.05	0.30	0.50	0.80	1.00	1.20	2.60
80- 84 yo	71	1.04	0.29	0.50	0.80	1.00	1.20	1.80
85- 89 yo	32	1.25	0.61	0.60	0.95	1.10	1.45	4.00
90- 94 yo	7	1.23	0.17	1.00	1.10	1.20	1.40	1.50
95- 99 yo	2	1.10	0.28	0.90	0.90	1.10	1.30	1.30
Diabetes	79	1.13	0.34	0.60	0.90	1.10	1.30	3.20
65- 69 yo	16	1.07	0.19	0.80	0.95	1.00	1.25	1.40
70- 74 yo	30	1.13	0.47	0.60	0.90	1.00	1.30	3.20
75- 79 yo	19	1.11	0.23	0.70	1.00	1.20	1.20	1.50
80- 84 yo	10	1.20	0.29	0.80	1.00	1.20	1.40	1.70
85- 89 yo	3	1.30	0.26	1.10	1.10	1.20	1.60	1.60
90- 94 yo	1	1.40	.	1.40	1.40	1.40	1.40	1.40
95- 99 yo	0							



b. Perform an analysis to determine whether the mean creatinine levels differ across groups defined by prevalence of diabetes.

Ans: On average, the mean creatinine levels in diabetics is estimated to be 0.0744 mg/dl higher than in nondiabetics (95% CI 0.0054 mg/dl lower to 0..154 higher), a result that is not unexpected when there is no true difference between the sexes with respect to mean C reactive protein levels (P = 0.067).

(Note that I chose to report the analysis based on a t test which allows unequal variances, because that would probably be the more standard approach. It certainly would have been okay to report the analysis based on the linear regression with robust SE, had I chosen to use that a priori. It would not be okay to do both analyses, and then report the one I liked better. Of course, the two analyses will be quite similar, though not exactly the same due to the way the degrees of freedom are computed, as well as slight differences in the computation of the SE.)

c. Perform an analysis to determine whether there is a linear trend in mean creatinine levels by age.

Ans: On average, the mean creatinine level is estimated to be 0.0056 mg/dl higher for every year difference in age, with the older subjects having higher levels (95% CI 0.0005 mg/dl to 0.0106 higher). This result is not typical when there is no true difference with respect to mean creatinine levels across age groups (P = 0.032).

d. Perform an analysis to determine whether the mean creatinine levels differ across diabetes groups after adjustment for age.

Ans: On average, the mean creatinine levels in diabetics is estimated to be 0.0765 mg/dl higher than in nondiabetics of the same age (95% CI 0.0013 mg/dl lower to 0.154 mg/dl higher), a result that is not unexpected when there is no true difference between the diagnostic groups with respect to mean creatinine levels (P = 0.054).

e. Perform an analysis to determine whether there is a linear trend in mean creatinine levels by age after adjustment for prevalence of diabetes.

Ans: On average, the mean creatinine levels is estimated to be 0.0057 mg/dl higher for every year difference in age when comparing subjects of the same diabetes diagnostic group (95% CI 0.0006 mg/dl to 0.0107 mg/dl higher), with the older subjects tending toward higher levels. This result is more extreme than what would be expected when there is no true difference with respect to mean creatinine levels across age groups (P = 0.028).

f. Perform an analysis to determine whether there is a linear trend in mean creatinine levels by age in non-diabetics.

Ans: In nondiabetics, the mean creatinine level is estimated to average 0.0054 mg/dl higher for every year difference in age, with the older subjects having higher levels (95% CI 0.00004 mg/dl lower to 0.0109 mg/dl higher). Using a 5% level of significance, we can not rule out the possibility that there is no true difference with respect to mean creatinine levels across age groups (P = 0.052).

g. Perform an analysis to determine whether there is a linear trend in mean creatinine levels by age in diabetics.

Ans: In diabetics, the mean creatinine level is estimated to average 0.0076 mg/dl higher for every year difference in age, with the older subjects having higher levels (95% CI 0.0037 mg/dl lower to 0.0188 mg/dl higher). Using a 5% level of significance, we can not rule out the possibility that there is no true difference with respect to mean creatinine levels across age groups (P = 0.185).

h. Perform an analysis to test whether the results obtained in part g are statistically significantly different from those in part f. Interpret all parameters in the model used to answer this question, and relate those estimates to the parameter estimates obtained in parts f and g.

Ans: In nondiabetics, the mean creatinine level is estimated to average 0.0054 mg/dl higher for every year difference in age, with the older subjects having higher levels (95% CI 0.00004 mg/dl lower to 0.0109 mg/dl higher). In diabetics, the mean creatinine level is estimated to average 0.0076 mg/dl higher for every year difference in age, with the older subjects having higher levels (95% CI 0.0037 mg/dl lower to 0.0188 mg/dl higher). This difference between the diabetes diagnostic groups in age effects of 0.0021 (95% CI -0.0102 to 0.0144) is not beyond that which might be reasonably expected when the association between mean creatinine levels and age is the same for both diagnostic groups (P= 0.735).

(Note that I fit a regression model including an age-diabetes interaction in order to address this question. However, the estimates from that model are exactly the same as were estimated in the

"disaggregated" analyses of the age effect in each diagnostic group separately. Furthermore, the statistical significance of the interaction is nearly identical to that we would have obtained has we used the regression results for the stratified analyses to create a statistic testing the difference in the slopes. That is, in the regressions used to answer parts f and g, the slope estimates were approximately normally distributed. Hence we could estimate the standard error of the difference between the slope estimates by taking the square root of the sum of the squared standard errors from the two analyses. Fitting the regression model with the interaction let me avoid having to go to that trouble. But in the annotated Stata log I did this latter approach as well.)

i. Based on the results of the above analyses, is there evidence that age confounds the association between mean creatinine and diabetes diagnosis? Explain.

Ans: No. The estimates of the diabetes effect are very similar in the adjusted and unadjusted analyses. In linear regression, we can use this criterion to assess confounding. I do note that there is evidence that age is associated with creatinine level, but there is no important trend between age and diabetes in this sample. (see Stata output)

j. Based on the results of the above analyses, is there evidence that diabetes diagnosis confounds the association between mean creatinine and age? Explain.

Ans: No. The estimates of the age effect are very similar in the adjusted and unadjusted analyses. In linear regression, we can use this criterion to assess confounding. I do note that there is no evidence that diabetes is associated with an important difference in mean creatinine level, and there is no important trend between age and diabetes in this sample. (see Stata output)

k. Based on the results of the above analyses, is there evidence that age modifies the association between mean creatinine and diabetes diagnosis?

Ans: No. We answered this question in part h.

2. We are interested in examining how geometric mean creatinine levels vary by age and prevalence of diabetes.

    a. Perform an analysis to determine whether the geometric mean creatinine levels differ across diabetes groups after adjustment for age.

Ans: On average, the geometric mean creatinine levels in diabetics is estimated to be 7.08% higher than in nondiabetics of the same age (95% CI 0.92% to 13.6% higher), a result that is unexpected when there is no true difference between the diagnostic groups with respect to mean creatinine levels (P = 0.024).

3. We are interested in examining how the odds of having high creatinine levels (greater than 1.45) vary by age and prevalence of diabetes.

    a. Perform an analysis to determine whether the odds of high creatinine levels differ across diabetes groups after adjustment for age.

Ans: On average, the odds of high creatinine levels (> 1.5 mg/dl) in diabetics is estimated to be 30.1% higher than in nondiabetics of the same age (95% CI 41.0% lower to 187% higher), a result that is not unexpected when there is no true difference between the diagnostic groups with respect to the prevalence of high creatinine levels (P = 0.514).

b. Perform an analysis to determine whether there is a linear trend in the odds of high creatinine levels across age groups after adjusting for diabetes.

Ans: On average, the odds of high creatinine levels (> 1.5 mg/dl) is estimated to be 8% higher for each year difference in age when comparing individuals in the same diabetes diagnostic group (95% CI 3.89% to 12.4% higher), a result that is unexpected when there is no true difference between by age with respect to the prevalence of high creatinine levels (P < .0005).

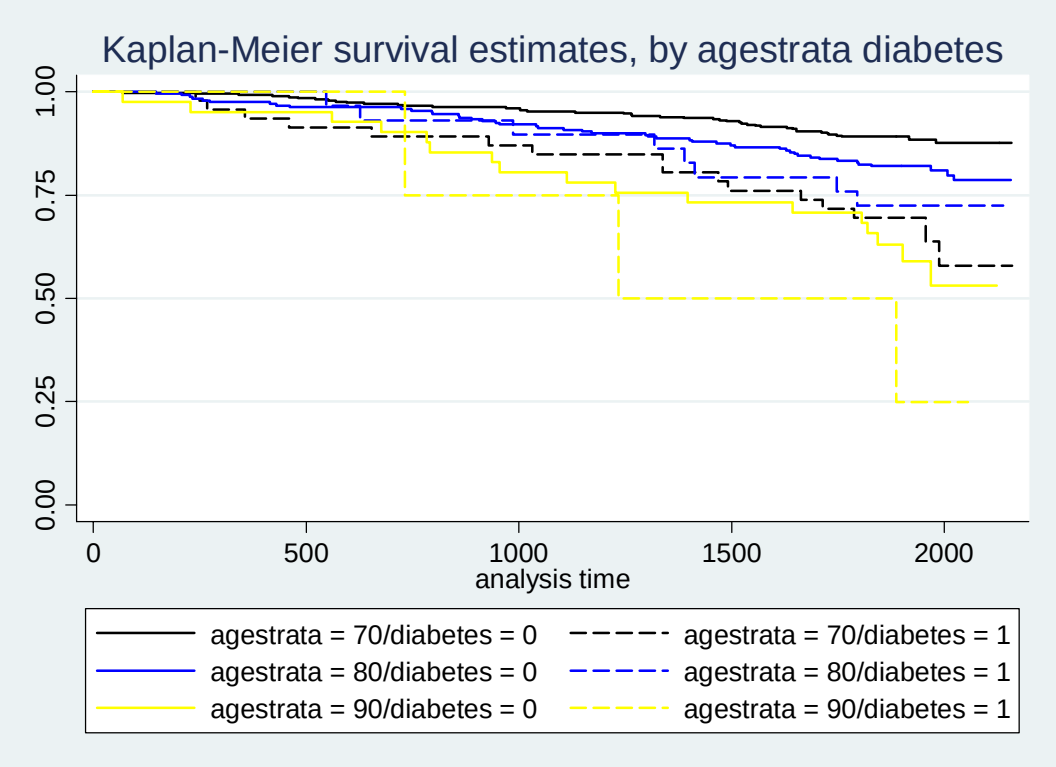
c. Perform and analysis to determine whether any linear trend found in part b is well described as a straight line in the log odds scale. That is, perform a test to see whether there is sufficient evidence in the data to suggest a nonlinear trend in the odds of high creatinine levels by age after adjustment for diabetes. (A typical approach is to consider the possibility of a curvilinear trend by fitting both age and a new variable equal to the square of age.)

Ans: Based on a logistic regression model that included both age and age squared, the evidence for a trend in log odds of high creatinine that deviated from a straight line relationship was right at the threshold for declaring statistical significance (P = 0.05).

4. We are interested in examining how the distribution of time to death differs across groups defined by creatinine level after adjustment for age and prevalence of diabetes.

a. Provide suitable descriptive statistics regarding the distribution of times to death across groups defined by creatinine levels, age, and prevalence of diabetes.

Ans: The following plots and table summarize the probability of survival according to diabetes, age, and serum creatinine. Evident are trends toward worse survival with increasing age, increasing creatinine, or diabetes diagnosis. ***(With more effort, and at the price of a very complicated graph, we could have simultaneously stratified on all three variables. We do this sometimes, but most times just let regression modeling sort out these issues.)***



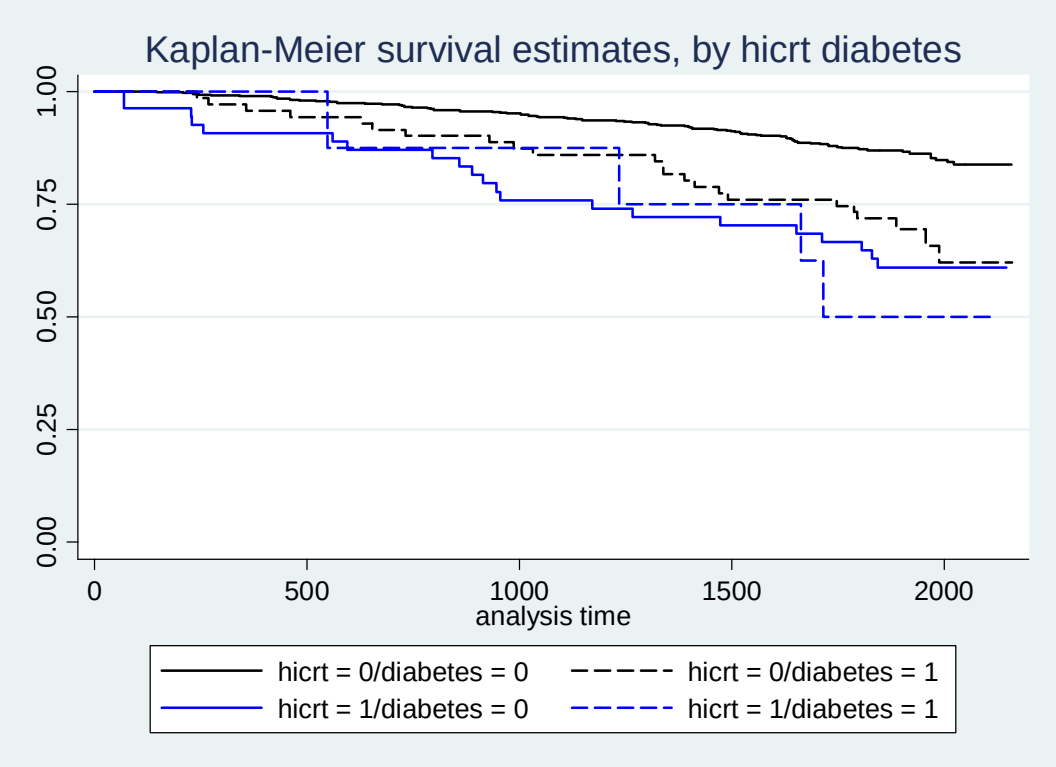


Table 2. Two- and four-year survival probabilities by diabetes diagnostic group and either age or creatinine level.

	Nondiabetics		Diabetics	
	2 yr Surv	**4 yr Surv**	**2 yr Surv**	**4 yr Surv**

Age 65-74	0.97	0.93	0.89	0.80
Age 75-84	0.96	0.88	0.93	0.79
Age 85-99	0.90	0.73	1.00	0.50
Crt < 1.45	0.97	0.92	0.92	0.79
Crt > 1.45	0.87	0.72	0.88	0.75

b. Perform an analysis to determine whether creatinine levels are associated with all cause mortality.

Ans: On average, the instantaneous risk of death is 4.00 times higher for each 1 mg/dl difference for groups differing in their serum creatinine levels (95% CI 2.74 to 5.82 times higher). This result is highly statistically significant (P < 0.0005), and thus is not typical of random observations in the absence of a true association between creatinine and survival.

c. Perform an analysis to determine whether creatinine levels are associated with all cause mortality after adjustment for age and diabetes diagnosis.

Ans: On average, the instantaneous risk of death is 3.40 times higher for each 1 mg/dl difference for groups differing in their serum creatinine levels but being of similar age and diabetes status (95% CI 2.30 to 5.01 times higher). This result is highly statistically significant (P < 0.0005), and thus is not typical of random observations in the absence of a true association between creatinine and survival after adjustment for age and diabetes.

d. Is there evidence that age and/or diabetes diagnosis confounds the association between mortality and creatinine level? Explain.

Ans: Yes. Diabetes and age are each associated with worse survival, and as seen in problems 1-3, age and diabetes are at least somewhat associated with creatinine level. We futher note that because the adjusted HR is closer to the null HR of 1.0 than is the unadjusted HR, there must be confounding. (This is a criterion that can be used with logistic regression or PH regression. Had the adjusted estimate been more extreme than the unadjusted estimate, we would not have been able to judge confounding by this criterion of comparing the unadjusted and adjusted estimates.)