**Biost 518: Applied Biostatistics II**

**Homework #1** January 6, 2014

1. There are total 735 observation in the study. The minimal observation time among those who didn’t die on the study (earliest censoring time) was at 5.002 years. Therefore, it is reasonable to dichotomize the time to death according to death within 5 years of study enrolment or death after 5 years.
2. Descriptive statistics were used to describe the distribution of some variables between group with death within 5 years and death after 5 year. They are continuous variables, such as age, weight, and LDL level; nominal variable- sex; and ordered categorical variables- congestive heart failure and stroke.

Descriptive statistics for selected variables within groups defined by 5 year all-cause mortality

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
|  |  | Number (missing data) | Mean (SD) |  | Median(IQR) / *Proportion*  | (Min, Max) |
| Death within 5 years | Age (years) | 121 (0) | 76.48 (6.17) |  | 75 (72, 81) | (67, 91) |
|  | Weight (Pounds) | 121 (0) | 159.12 (32.79) |  | 154 (139, 176) | (96, 264) |
|  | Somking (pack-years) | 120 (1) | 28.05 (36.04) |  | 18.375 (0, 46) | (0, 240) |
|  | Serum LDL (mg/dl) | 119 (2) | 118.70 (36.16) |  | 117 (96, 142) | (11, 227) |
|  | Sex (Male) | 121 (0) |  |  | *0.48* |  |
|  | Congestive heart failure (Yes) | 121 (0) |  |  | *0.35* |  |
|  | Coronary heart disease | 121 (0) | No |  | *0.62* |  |
| Angia |  | *0.14* |  |
| Myocardial infraction |  | *0.24* |  |
|  | Stroke | 121 (0) | No |  | *0.71* |  |
| Transient ischemic attack |  | *0.06* |  |
| Stroke |  | *0.23* |  |
| Death after 5 years | Age (years) | 614(0) | 74.19 (5.22) |  | 73 (71, 77) | (65, 99) |
|  | Weight (Pounds) | 614(0) | 160.11 (30.35) |  | 158.75 (138.5, 180) | (74, 258) |
|  | Somking (pack-years) | 614(0) | 17.95 (24.69) |  | 4.35 (0, 31.88) | (0, 180) |
|  | Serum LDL (mg/dl) | 606(8) | 127.20 (32.93) |  | 127 (103, 148) | (39, 247) |
|  | Sex (Male) | 614(0) |  |  | *0.47* |  |
|  | Congestive heart failure (Yes) | 614(0) |  |  | *0.039* |  |
|  | Coronary heart disease | 614(0) | No |  | *0.82* |  |
| Angia |  | *0.08* |  |
| Myocardial infraction |  | *0.10* |  |
|  | Stroke | 614(0) | No |  | *0.89* |  |
| Transient ischemic attack |  | *0.03* |  |
| Stroke |  | *0.08* |  |

1. An independent two-sample t test with unequal variance assumption was conducted to compare the mean LDL value between the group with death within 5 years and death after 5 year. The mean of LDL value of the group with death happened within 5 year was 118.70 mg/dl (standard deviation: 36.16). The mean of LDL value of the group with death happened after 5 year was 127.20 mg/dl (standard deviation: 32.93). There was a significant difference between the mean LDL with a 8.5 mg/dl larger in the group with death happened after 5 year. (95% CI: 1.44, 15.56; P value: 0.02)
2. After taking the logarithms of all the LDL values, an independent two-sample t test with unequal variance assumption was conducted to compare the means of those logarithms LDL value between the group with death within 5 years and death after 5 year. Finally, the means of the logartithms and the differences in the means and 95% CI of the logarithms were exponentiated. The mean of LDL value of the group with death happened within 5 year was 112.01 mg/dl (standard deviation: 1.46). The mean of LDL value of the group with death happened after 5 year was 122.83 mg/dl (standard deviation: 1.31). There was a significant difference between the mean LDL with a 2.51 mg/dl larger in the group with death happened after 5 year. (95% CI: 1.02, 1.18; P value: 0.01)
3. Five-year survival probability for high and low LDL was calculated from Kaplan-Meier Estimates. Five year survival probability for low LDL was 83.01% (95% CI: 79.81%-85.75%), while five year survival probability for high LDL was 86.92% (95% CI: 78.91% -92.03%). The difference in 5-year survival probability was 3.91% (95% CI: -0.0309; 0.1091). Since the confidence interval contained 0, we cannot find enough evidence to prove there was no association between serum LDL and 5 year all-cause mortality.
4. Chi-Squared test was used to evaluate the association between serum LDL and 5 year all-cause mortality. LDL values were dichotomized by the level of 160mg/dl. We can interpret the estimate as the odds of death within 5 years is 0.735 times higher for those who have high LDL levels as compared to those with low LDL levels. But the 95% CI was from 0.403 to 1.340 and the P value was 0.3139. Therefore the precision was not adequate to demonstrate that there was no association between serum LDL and 5 year all-cause mortality.
5. Log Rank test was used to detect an association between high serum LDL and 5 year all-cause mortality. Based on the two-sided P-value of 0.2249, we cannot find enough evidence to prove there was no association between serum LDL and 5 year all-cause mortality.
6. I would use two-sample t test that presumes unequal variances to compare mean LDL values between the group with death within 5 years and death after 5 year. Because this method treats LDL values as continuous variable, instead of dichotomizing this variable and treating as a categorical variable. In this way, this analysis preserved more information.