Applied Biostatistics II - Winter 2014 - Homework #1

1. From the table below we see that of the 602 censoring events in the data, all were recorded after 5 years of study enrollment. These observations are right censored, so we know that true times to death are at least 5 years. Hence, we can dichotomize observed time to death into groups by time to death within 5 years of study enrollment and time to death after at least 5 years of enrollment.

|  |  |  |
| --- | --- | --- |
|  | *Event*: | |
| **Time to event** | **Censoring** | **Death** |
| Less than 5 years | 0 | 121 |
| Greater than 5 years | 602 | 12 |

1. We are interested in whether there is an association between serum LDL and 5 year all-cause mortality in the study participants. To explore this possible association, we summarize the serum LDL, age, sex, weight, smoking history, as well as prior history of cardiovascular disease across the groups determined by 5-year mortality.

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **Variable** | **Group** | **N** | **Mean** | **Standard Deviation** | **Min** | **Max** |
| Serum LDL  (mg/dL) | All | 725 | 125.8 | 33.6 | 11 | 247 |
| Years < 5 | 119 | 118.7 | 36.2 | 11 | 227 |
| Years > 5 | 606 | 127.2 | 32.9 | 39 | 247 |
|  |  |  |  |  |  |  |
| Age  (Years) | All | 735 | 74.6 | 5.5 | 65 | 99 |
| Years < 5 | 121 | 76.5 | 6.2 | 67 | 91 |
| Years > 5 | 614 | 74.2 | 5.2 | 65 | 99 |
|  |  |  |  |  |  |  |
| Male  (%) | All | 735 | 49.8 | 50.0 | 0 | 1 |
| Years < 5 | 121 | 64.5 | 48.1 | 0 | 1 |
| Years > 5 | 614 | 46.9 | 50.0 | 0 | 1 |
|  |  |  |  |  |  |  |
| Weight  (Pounds) | All | 735 | 160.0 | 30.7 | 74 | 264 |
| Years < 5 | 121 | 159.1 | 32.8 | 96 | 264 |
| Years > 5 | 614 | 160.1 | 30.3 | 74 | 258 |
|  |  |  |  |  |  |  |
| Ex- or Current Smoker (%) | All | 735 | 56.3 | 49.6 | 0 | 1 |
| Years < 5 | 121 | 62.8 | 48.5 | 0 | 1 |
| Years > 5 | 614 | 55.0 | 49.8 | 0 | 1 |
|  |  |  |  |  |  |  |
| Congestive Heart Failure (%) | All | 735 | 5.6 | 23.0 | 0 | 1 |
| Years < 5 | 121 | 14.0 | 34.9 | 0 | 1 |
| Years > 5 | 614 | 3.9 | 19.4 | 0 | 1 |
|  |  |  |  |  |  |  |
| Coronary Heart Disease (%) | All | 735 | 21.1 | 40.8 | 0 | 1 |
| Years < 5 | 121 | 38.0 | 48.7 | 0 | 1 |
| Years > 5 | 614 | 17.8 | 38.2 | 0 | 1 |
|  |  |  |  |  |  |  |
| Stroke (%) | All | 735 | 13.5 | 34.2 | 0 | 1 |
| Years < 5 | 121 | 28.9 | 45.5 | 0 | 1 |
| Years > 5 | 614 | 10.4 | 30.6 | 0 | 1 |

We observe that participants who survived beyond 5 years have higher serum LDL on average than those not surviving at least 5 years. The average ages and weights of participants do not appear very different across groups, though the proportion of males is higher in participants with time to death less than 5 years. The proportion of participants with histories of smoking, congestive heart failure, coronary heart disease, or stroke are all higher in the group surviving less than 5 years. Additionally, we note that there are 10 individuals with missing serum LDL levels.

1. To test the hypothesis that there is an association between mean LDL and 5 year all-cause mortality, we performed a two-sample t-test for the difference in mean serum LDL between the groups based on 5-year mortality. The observed difference in mean serum LDL levels was -8.50 mg/dL. With a p-value of 0.0115 and a 95% confidence interval of -15.1 to -1.9 mg/dL, we have statistically significant evidence that mean serum LDL levels differ between participants who survived less than 5 years and those surviving at least 5 years. This result suggests that there is an association between mean LDL and 5 year all-cause mortality.
2. We next tested the hypothesis that the ratio of geometric mean serum LDL between the groups based on 5-year mortality was equal to one with another two-sample t-test. The observed ratio of geometric means was 0.91. With a p-value 0.0016 and 95% confidence interval 0.86 to 0.97, the ratio of geometric mean serum LDL differs significantly from one in our sample. This result again suggests that there is an association between mean LDL and 5 year all-cause mortality.
3. Next we compared the risk of high serum LDL (at least 160 mg/dL) between the two groups based on 5-year mortality and found that the risk difference was 0.03. With a 95% confidence interval from -0.04 to 0.10 and a p-value of 0.3753, we fail to detect an association between high serum LDL and 5-year all-cause mortality.
4. The odds ratio between the groups based on 5-year mortality for high serum LDL was estimated as 1.29 with a 95% (Woolf) confidence interval from 0.73 to 2.28 with a p-value of 0.3753. Thus we again fail to detect an association between high serum LDL and 5-year all-cause mortality.
5. We now compare instantaneous risk of death across the entire study time for groups based on high serum LDL (at least 160 mg/dL). This is equivalent to testing the hazard ratio. The log rank test was significant at the p-value < 0.0001 level, suggesting a possible association between high serum LDL and instantaneous risk of death.
6. Due to the presence of censoring in the time to death data, I would perform an analysis based on Kaplan-Meier survival methods. Rather than dichotomize the data by time to death, I would create two groups based on high serum LDL (at least 160 mg/dL) and low-to-normal serum LDL (less than 160 mg/dL) and compare their survival curves over the entire observation period. Testing for an association using this interpretation corresponds to a test for proportional hazards, such as the log rank test performed in problem 7.