Biostat 515/518 Homework 1

2014 Winter

**28/40**

Dichotomizing the time to death to within 5 years of study enrollment or more than five years from study enrollment is valid in this study. This is due to the fact that all deaths that were not observed (aka censored) occurred at least 5 years after that patient was enrolled (the date of their MRI). Patients who had not died on September 16, 1997 were censored, but the patient with the least number of days between his/her MRI and this date was enrolled for 1827 days, which is a couple days past five years. Therefore, dichotomizing in this way removes the issue of censored data since we know with certainty in which of the two groups every patient should belong.

**Question 2**:

Tables 1 and 2 below summarize descriptive statstics for our sample in this study. More participants had LDL levels below 160mg/dL than above that mark (618 vs 107). The proportion who died within 5 years wsa greater for those with high LDL in all sub-categires except for those who had been diagnosed with angina, myocardial infarction, or congestive heart failure.

Within both high and low LDL groups, those who died within five years were younger and smoked fewer pack-years, on average.

Table 1: Categorical Variables

|  |  |  |  |
| --- | --- | --- | --- |
|  |  | **High LDL (>160 mg/dL)** | **Low LDL (<160 mg/dL)** |
|  |   | **N** | **% Died within 5 Years** | **N** | **% Died within 5 Years** |
| Overall |  |  107 | 86.92% | 618 | 83.00% |
| Sex | **Males** | 45 | 80.00% | 315 | 78.41% |
|  | **Females** | 62 | 91.93% | 303 | 87.79% |
| CHD | **No Diagnosis** | 86 | 93.02% | 488 | 86.07% |
|  | **Diagnosis of Anigna** | 8 | 62.50% | 54 | 74.07% |
|  | **Diagnosis of Myocardial Infarction** | 13 | 61.54% | 76 | 69.73% |
| CHF | **No Diagnosis** | 104 | 88.46% | 581 | 84.51% |
|  | **Diagnosis of Congestive Heart Failure** | 3 | 33.33% | 37 | 59.46% |
| Stroke | **No Diagnosis** | 87 | 89.66% | 541 | 85.95% |
|  | **Diagnosis of a Transient Ischemic Attack** | 6 | 83.33% | 18 | 66.66% |
|  |  **Diagnosis of a Stroke** | 14 | 71.42% | 54 | 61.02% |

Table 2: Continuous Variables

|  |  |  |  |
| --- | --- | --- | --- |
|  |  | **High LDL (>160 mg/dL)** | **Low LDL (<160 mg/dL)** |
|  |  | Died within 5 Years | Survived 5 Years | Died within 5 Years | Survived 5 Years |
| Age (yrs) | Mean(sd) | 74.46(5.46) | 77.64(7.14) | 74.12(5.17) | 76.44(6.05) |
|  | Min-Max | 65-94 | 69-89 | 65-99 | 67-91 |
| Smoking History (Pack-years) | Mean(sd) | 16.66(23.58) | 27.55(28.50) | 18.22(24.93) | 28.03(37.20) |
|  | Min-Max | 0-102 | 0-78 | 0-180 | 0-240 |

4/4 for general table layout

2/3 for the choice of descriptive statistics

0/3 for discussion of finding

Did not mention about missing data (-1)

Total: 6/10

**Question 3**:

A two-sample t-test was performed to compare mean LDL levels of those who did and did not survive past 5 years of their MRI date. There is evidence that the mean LDL level is different between those who did and those who did not survive past five years after their MRI (p-value of 0.0186). Those who died had an average LDL 8.501mg/dL higher than those who survived. With 95% confidence the true difference in arithmetic mean is between 1.441 and 15.697 mg/dL.

**Question 4**:

A two-sample t-test was performed to compare the geometric mean LDL levels of those who did and did not survive past 5 years of their MRI date. This was accomplished by taking the log of the LDL levels, performing the test, and exponentiating both point the estimate and the confidence interval bounds. There is evidence that the geometric mean LDL level is different between those who did and those who did not survive past five years after their MRI (p-value of 0.0128). Those who died had n geometric average 1.0965mg/dL higher than those who survived. With 95% confidence the true difference in geometric mean is between 1.020 and 1.1787 mg/dL.

5/5 for performing an appropriate analysis

1/5 for reporting the association appropriately

Did not report the geometric means of each groups (-1)

Wrong report the point estimate(ratio of geometric mean) (-1)

Wrong report which of geometric mean of LDL between two groups is higher (-0.5)

Did not report whether the p-value is two-sided or one-sided(-0.5)

Wrong interpretation of CI (-1)

Total: 6/10

**Question 5**:

A two-sample t-test was performed to compare the probability of death within five years between those with high and low LDL levels, where “high” is defined to be >160mg/dL. There is no evidence that the probability of death is different for people with high LDL levels (>160mg/dL) versus low LDL (p-value of 0.2806). Those with high LDL levels had a probability of death within five years 0.039mg/dL higher than those with low levels of LDL. With 95% confidence the true difference in probability is between -0.0322 and 0.1103 mg/dL.

**Question 6**:

The probability of dying within 5 years for participants with high levels of LDL is estimated to be 0.8692, while the probability for the low LDL group is estimated to be 0.8301. This leads to odds of 6.643 and 4.886, respectively, and an odds ratio of 1.359. In other words, participant with high LDL levels (at least 160mg/dL) were 1.359 times as likely to die within five years as those with high LDL levels (at least 160mg/dL) were 1.359 times as likely to die within five years as those with low levels of LDL. An odds ratio test suggests that the true odds ratio is between 0.7349 and 1, with a 95% confidence level.

**Question 7**:

A Cox regression comparison of the instantaneous risk of death (hazard) between the high and low LDL groups yields a hazard ratio of 0.7179. There is no evidence that these instantaneous risks are different (p-value of 0.227). A 95% confidence interval suggests the true hazard ratio is between 0.4193 and 1.229.

5/5 for performing an appropriate analysis

3/5 for reporting the association appropriately

No interpretation of CI(-1)

Did not report whether the p-valu is two-sided or one-sided(-1)

Total: 8/10

**Question 8**:

To answer the question about an association between mortality and serum LDL, a priori I would have chosen to look at the hazard ratio, as it is the most efficient in the sense that we need not lose information by categorizing participants into those who died and those who did not die within five years. We are able the more precise measurements of time to death (or censoring).

Choose appropriate analysis (4)

Performed analysis that are valid (2)

Mention about losing information by dichotomizing a continuous variable (2)

Total: 8/10