Biost 515/518

HW 1

Scott Emerson

1/13/14

1. A new dummy variable was created that dichotomized observation time. Patients were recorded as 0 if they had an observation time/time to death of 5 years or less, and a 1 if they had an observation time of more than 5 years. A contingency table was constructed to look at the distribution of death for both observation time groups. The results are shown below.

|  |  |  |
| --- | --- | --- |
| Follow-up Time | Death Observed | Death not Observed |
| ≤ 5 Years | 121 | 0 |
| > 5 Years | 12 | 602 |

Note that all patients who had observation 5 years or less are not censored data points. Therefore, when preforming statistical analysis on the subset of observations with observation times within 5 years we can use standard statistical methods and not be concerned with censoring. However, if we were to preform analyses on the whole data set without dichotomizing time, we must use methods that account for censored data as the majority of patients who had follow-up times greater that 5 years are censored observations.

1. Arithmetic mean, standard deviation, percentiles, minimum and maximum were calculated for continuous variables of interest. Proportions were calculated for binary and categorical variables of interest. Smoking was treated as a binary variable, where patients with 0 pack-years of smoking were called non-smokers. Descriptive statistics were calculated for the whole sample, as well as by LDL subgroups. The two LDL subgroups, normal LDL and high LDL, were defined as patients with LDL serum levels less than 160 mg/dL and patients with LDL serum levels of 160 mg/dL or higher. Due to missing data points for the predictor of interest, LDL level, 10 observations were removed from the analysis. The table summarizing the results is presented on the next page.

A larger proportion of the patients in the normal LDL group died with in 5 years (17%) compared to higher LDL group (13%). Also, the normal LDL group had higher instances of Angina and Myocardial infarction diagnoses compared to patients with high LDL levels. However, more patients with high levels of LDL had history of stroke.

|  |  |
| --- | --- |
|  | **Table I. Descriptive Statistics by LDL Group Level** |
| Variables | All (N=725) | LDL < 260 mg/dL (N=618) | LDL ≥ 160 mg/dL (N=107) |
| Mean (SD) | Med. (IQR) | Min, Max | Mean (SD) | Med. (IQR) | Min, Max | Mean (SD) | Med. (IQR) | Min, Max |
| LDL Cholesterol Level (mg/dL) | 125.80 (33.60) | 125 (102, 147) | 11, 247 | 116.40 (25.73) | 118 (98, 137) | 11, 159 | 180.40 (18.26) | 175 (167, 188) | 160, 247 |
| Age (years) | 74.57 (5.45) | 74 (71, 78) | 65, 99 | 74.51 (5.39) | 73 (71, 78) | 65, 99 | 74.88 (5.77) | 74 (70, 78) | 65, 94 |
| Weight (lbs) | 160.00 (30.77) | 158 (74, 179) | 74, 264 | 159.40 (30.78) | 158 (138, 178) | 86, 264 | 162.70 (30.68) | 159 (143, 181) | 74, 257 |
| Death in 5 years | 0.16 | - | - | 0.17 | - | - | 0.13 | - | - |
| Sex | 0.50 | - | - | 0.51 | - | - | 0.42 | - | - |
| Smoker  | 0.44 | - | - | 0.44 | - | - | 0.46 | - | - |
| Congestive Heart Failure  | 0.06 |  |   | 0.06 |  |   | 0.03 |  |  |
| History of Heart Disease |  |  |   |   |  |   |   |  |  |
| No History | 0.79 | - | - | 0.79 | - | - | 0.80 | - | - |
| Angina Diagnosis | 0.09 | - | - | 0.09 | - | - | 0.07 | - | - |
| MI Diagnosis | 0.12 | - | - | 0.12 | - | - | 0.07 | - | - |
| History of Stroke |  |  |   |   |  |   |   |  |  |
| No History | 0.87 | - | - | 0.88 | - | - | 0.81 | - | - |
| Transient Ischemic Attack | 0.03 | - | - | 0.03 | - | - | 0.06 | - | - |
| Diagnosis of Stroke | 0.10 | - | - | 0.10 | - | - | 0.13 | - | - |

2)

1. A two-sample t-test with unequal variances was conducted to assess the association between LDL levels and 5-year all cause mortality. The mean LDL level was estimated for the two mortality groups. The difference in means was estimated, and a confidence interval for the difference in means was generated. A p-value was also generated to test the hypothesis that the difference in means is equal to zero.

The mean LDL level for the patients who had died within 5 years is 118.70 mg/dL. This is slightly lower than the mean LDL level for patients who had survived past 5 years, 127.20 mg/dL. The estimated difference in means (normal LDL minus high LDL) is 8.50. The 95% confidence interval for the true difference in means is (1.44, 15.56). A p-value of 0.02 was yield, and we therefore reject the null hypothesis that the mean LDL levels are equal across mortality groups. Based on the confidence interval, we see that is likely that mean LDL levels are lower for patients who have died within 5 years.

1. The geometric mean for LDL was calculated for each mortality group. The hypothesis that the geometric mean of LDL is equal across each group was tested using a two-sample t-test assuming unequal variances. To preform the t-test and estimate the geometric mean, the mean of the logged LDL data was taken. A 95% confidence interval for the true difference in mean LDL levels was also generated.

The geometric mean of LDL for patients who had had survived past years is 122.83 mg/dL. The geometric mean of LDL for patients who had died within 5 years was slightly lower, 112.01 mg/dL. The point estimate for the difference in geometric means is 1.10 mg/dL. The 95% confidence interval for the true difference in geometric means is (1.02, 1.18). The p-value yielded when testing for equality in geometric means across two groups is 0.01. We reject the null hypothesis that the geometric mean LDL is equal across mortality groups. Note the conclusion reached from this p-value agrees with the confidence interval.

1. The proportion of deaths was estimated for each LDL group by calculating the number of deaths out of total number of normal or high LDL patients. The risk difference (difference in proportions) was estimated. The 95% confidence interval for the risk difference was also calculated using the formula below, where p1 and p2 are the estimate proportions of deaths and n1 and n2 are the sample size for each LDL group. A chi-squared goodness of fit test was also preformed to assess the association between LDL group and all cause 5-year mortality.

$$RD\pm 1.96\* \sqrt{\frac{p\_{1}\left(1-p\_{1}\right)}{n\_{1}}+\frac{p\_{2}(1-p\_{2})}{n\_{2}}}$$

The proportion of deaths for the normal LDL group is 0.17. This is higher than the proportion of deaths for the high LDL group, which had a proportion of deaths of 0.13. The point estimate for the RD is 0.04, with a confidence interval of (-0.03, 0.11). Note this confidence interval contains 0, so we cannot say the proportion of deaths is unequal between LDL groups. The chi-squared test yielded a p-value of 0.31. We therefore cannot reject the null hypothesis that there is no association between LDL group level and mortality group.

1. The odds ratio comparing the odds of death for high LDL group to the normal LDL group was estimated forming the contingency table shown below. The odds ratio and corresponding 95% confidence interval for the odds ratio was computed using the formula below, where a, b, c, and d are the cell counts in the contingency table.

$$\frac{ad}{bc}\pm 1.96\*\sqrt{\frac{1}{a}+\frac{1}{b}+\frac{1}{c}+\frac{1}{c}}$$

The point estimate for the odds ratio is 0.74. That is, patients with high LDL levels have 0.74 times the odds of death compared to patients with normal LDL levels. The 95% confidence interval for the odds ratio is (0.40, 1.34). Note this interval contains one, so it is possible that the odds of death are equal for both LDL groups. This is in agreement with chi-squared test preformed in question 5 that also assess the association between LDL levels and mortality.

1. Censored data analyses were used to evaluate the association between LDL levels and mortality across the entire study period. Kaplan-Meier curves were generated to visually the survival over time for each LDL group. Confidence intervals for the survival curves were also generated. We estimated the hazard ratio using Cox Proportional Hazard regression. The hazard curves were assumed to be proportional, however this assumption may not be valid. A 95% confidence interval was also calculated for the hazard ratio. Finally, a log-rank test was performed to test they hypothesis that the survival curves (or, equivalently, hazard curves) were equal for all time points.

The Kaplan-Meier plot is shown below. Although the survival curves do overlap, particularly for early time points, survival for the high LDL group seems to be consistently higher compared to the normal LDL group. This difference is more apparent at later time points. However, the confidence intervals on the survival curves are overlapping so it is possible that the survival times are the same between LDL groups. The estimate yielded for the hazard ratio estimate 0.72. That is, the patients with high LDL have 0.72 times the risk of death compared to the patients with low LDL. The 95% confidence interval for the hazard ratio is (0.42, 1.23). Note this confidence interval contains 1, so it is possible that the LDL groups have equal risk of death. A p-value of 0.23 was yielded from the log-rank test. We, therefore, cannot reject the null hypothesis that the survival curves are unequal for all time points. This is consistent with our confidence interval generated for the hazard curve.



1. I would use have assessed the association between LDL levels and 5-year all cause mortality by using censored data analyses as in question 7. Although it was valid to dichotomize death, we do lose some information when treating death as a binary variable. When looking at the survival curves as in question 7, we can see how the relationship between LDL and death changes over time as well as the overall relationship.