BIOST 518 Winter 2014

ID 1391

1. It is valid to dichotomize the time to death according to death within 5 years of study enrolment or death after 5 years because all the censored subjects have observation times greater than 5 years (i.e. all subjects who died within 5 years are uncensored), as shown in the following table:

|  |  |  |  |
| --- | --- | --- | --- |
|  | Observation time < 5 years | Observation time > 5 years | Total |
| Censored | 0 | 602 | 602 |
| Uncensored | 121 | 12 | 133 |
| Total | 121 | 614 | 735 |

1. Methods: Descriptive statistics (N, mean, standard deviation, median, interquartile range, minimum, maximum) were constructed for the indicated variables, dichotomized by serum LDL levels (high serum LDL ≥ 160mg/dL, low serum LDL < 160mg/dL).

**Table 1: Descriptive Statistics of the association of serum LDL levels and 5-year all-cause Mortality Rates**

|  |  |  |
| --- | --- | --- |
|  | **Normal Serum LDL** | **High Serum LDL** |
| Age (mean, sd) | 74.51 (5.39) | 74.84 (5.78) |
| Weight (mean, sd) | 159.36 (30.78) | 163.09 (30.45) |
| **Sex**  *Male (%)*  *Female (%)* | 51  49 | 44  56 |
| **Smoking history**  *Never smoker (%)*  *Past/current smoker (%)* | 43  57 | 44  56 |
| **Prior history of CHD**  *No CHD (%)*  *Angina (%)*  *Myocardial infarction (%)* | 79  9  12 | 79  8  13 |
| **Congestive heart failure**  *No CHF (%)*  *CHF (%)* | 94  5 | 97  3 |
| **Stroke status**  *No Stroke (%)*  *Transient ischemic attack (%)*  *Stroke (%)* | 87  3  10 | 81  5  14 |
| **5-year mortality rate**  *Died in 5 years (%)*  *Survived in 5 years (%)* | 17  83 | 14  87 |

Inference: The mean age of subjects with high serum LDL and normal serum LDL are about the same. Subjects with high serum LDL have higher weights (163.09lb, sd = 30.45) compared to those with normal serum LDL (159.36lb, sd = 30.78). The sex breakdowns are approximately equal among those with normal serum LDL, while there are more females (56%) than males with high serum LDL. Subjects who have ever been smokers are of equal proportions between normal and high serum LDL levels. The proportions of subjects with past history of Cardiovascular Heart Disease (CHD) are about the same between serum LDL levels. There are more subjects with no history of stroke among normal serum LDL levels (87%) compared to high serum LDL levels (81%). Lastly, the five-year mortality rate is slightly higher among those with normal serum LDL levels (17%) compared to high serum LDL levels (14%).

1. Methods: A two-sample two-sided t-test with unequal variance was conducted. Denote μ1 as mean of serum LDL levels among those who survive within 5 years and μ2 as mean of serum LDL levels among those who die within 5 years, then

H0: μ1 – μ2 = 0

HA: μ1 – μ2 ≠ 0

Inference: The 95% confidence interval of the difference in mean serum LDL levels is (-15.56mg/dL, -1.44mg/dL). The two-sided p-value is 0.0186. With α = 0.05, we can reject the null hypothesis that the difference in means is zero. We conclude that there is a difference in mean serum LDL levels defined by vital status at 5 years, with those surviving having a higher mean serum LDL level (127.20mg/dL) compared to those who died (118.70mg/dL).

1. Methods: The log serum LDL levels based on vital status is calculated in STATA. Then a two-sample two-sided t-test with unequal variance was conducted on the difference in means. The difference and 95% confidence intervals for the geometric means are then calculated by exponentiating the log transformed numbers, since Y = exp(1/n\* Σ log(Xi)). Denote μ1 as geometric mean of serum LDL levels among those who survive within 5 years and μ2 as geometric mean of serum LDL levels among those who die within 5 years, then

H0: μ1/ μ2 = 1

HA: μ1/ μ2 ≠ 1

Inference: The 95% confidence interval of the ratio of geometric mean serum LDL levels is (0.85mg/dL, 0.98mg/dL). The two-sided p-value is 0.0128. With α = 0.05, we can reject the null hypothesis that the difference in means is zero. We conclude that there is a difference in mean serum LDL levels defined by vital status at 5 years, with those surviving having a higher mean serum LDL level (122.83mg/dL) compared to those who died (112.01mg/dL).

1. Methods: A chi-square test of independence is conducted on the distributions of 5-year mortality rate and dichotomized serum LDL levels (160mg/dL as cut-off point).

H0: Probability of death in 5 years and serum LDL are independent

HA: Probability of death in 5 years and serum LDL are not independent

Inference: The 95% binomial wald confidence interval for the risk difference is (-0.113, 0.028). The chi-square p-value is 0.2746. With α = 0.05, we cannot reject the null hypothesis that probability of death in 5 years and serum LDL are independent. In fact the probability of death in 5 years is higher among those with lower serum LDL.

1. Methods: The odds of death across groups whether subjects have high serum LDL are calculated, and a 95% confidence interval is constructed using the log odds. Denote OR = ad/bc as the odds of death of subjects with high serum LDL, then the woolf OR and 95% confidence intervals of OR are calculated in STATA using the following data:

|  |  |  |
| --- | --- | --- |
| Exposure/Outcome | N died in 5 years | N survive in 5 years |
| Serum LDL < 160mg/dL | 116 | 502 |
| Serum LDL ≥ 160mg/dL | 17 | 100 |

Inference: The odds of deaths of subjects with high serum LDL is 0.74. The Woolf 95% confidence interval is (0.42, 1.28). The chi-square p-value is 0.2746. With α = 0.05, we cannot reject the null hypothesis that probability of death in 5 years and serum LDL are independent.

1. Methods: A logrank test under the assumption of proportional hazards is conducted on the hazard curves of subjects with “normal” (<160mg/dL) and high (≥160mg/dL) serum LDL levels. Denote h1(t) as the hazard function of subjects with normal serum LDL levels and h2(t) as the hazard function of subjects with high serum LDL levels, then

H0: h1(t) = h2(t) for all t

HA: h1(t) ≠ h2(t) for all t

Inference: The two-sided p-value is 0.2664. With α = 0.05, we cannot reject the null hypothesis that the hazard functions are equal for all t. The graph of hazard rates is shown as follows.



From the above graph, it seems like subjects with normal and high serum LDL have similar hazards until about Day 600, then subjects with normal serum LDL have a higher probability of death.

1. Without any prescription of methods of analyses, I am inclined to perform the logrank test. Since the data is censored and dichotomizing serum LDL according to the 160mg/dL makes clinical sense, the logrank test is the natural test. However this requires the assumption of proportion hazards, which we may not have in the data. Another good candidate is the t-test on geometric means – this does not require any dichotomization of serum LDL levels, thus no issue of losing data here. Since we are interested in an association, not causation, it does not matter whether serum LDL levels or 5-year mortality is listed as the “predictor” or “outcome”.