1. The observations of time to death in this data are subject to (right) censoring. Nevertheless, problems 2 – 6 ask you to dichotomize the time to death according to death within 5 years of study enrolment or death after 5 years. Why is this valid? Provide descriptive statistics that support your answer.

**Ans: The observations of time is right censored data. To dichotomize the time to death within 5 years of study enrolment or death after 5 years could convert the right censored data to binary data which could be used to calculate the relative risk. The earliest censoring time was at 5.005, Hence analyses based on dichotomizing subjects with respect to 5 year survival or not are valid.**

1. Provide a suitable descriptive statistical analysis for selected variables in this dataset as might be presented in Table 1 of a manuscript exploring the association between serum LDL and 5 year all-cause mortality in the medical literature. In attention to the two variables of primary interest, you may restrict attention to age, sex, weight, smoking history, and prior history of cardiovascular disease (coronary heart disease (CHD), congestive heart failure (CHF), and stroke.

Ans:

|  |  |
| --- | --- |
| LDL | 5 year all-cause mortality(%) |
| Below 70 mg/dL | 40.9 |
| Below 100 mg/dL | 20 |
| 100-129 mg/dL | 19.2 |
| 130-159 mg/dL | 13.4 |
| 160-189 mg/dL | 11.4 |
| 190 mg/dL and above | 16.7 |

1. Perform a statistical analysis evaluating an association between serum LDL and 5 year all-cause mortality by comparing mean LDL values across groups defined by vital status at 5 years.

**Ans: The standard analysis to compare means of a continuous random variable across two groups is at test. H0: mean LDL value in participants who die within 5 years is equal to that in participants who did die within 5 years. H1: mean LDL value in participants who die within 5 years is lower equal to that in participants who did die within 5 years. A comparison of the two groups thus estimates that the mean ldl is 8.5 mg/dl higher among subjects who survive 5 years relative to those who die within 5 years. This observed difference is statistically different from 0 (P =0.0093< 0.0001), with a 95% confidence interval suggesting that the observed difference is what might be typically observed if the true difference between survivors and nonsurvivors was anywhere between 1.44 mg/dl and 15.56 mg/dl, with the survivors averaging higher ldl. We thus reject the null hypothesis of no association between survival time and ldl at study entry in favor of a trend toward higher mean ldl among subjects surviving the longer period of time.**

1. Perform a statistical analysis evaluating an association between serum LDL and 5 year all-cause mortality by comparing geometric mean LDL values across groups defined by vital status at 5 years.

**Ans: The geometric mean ldl is estimated to be 4.72mg/dl among subjects who die within 5 years of study entry and 4.81mg/dl among subjects who survive at least 4 years. A comparison of the two groups thus estimates that the ldl is 1.9% higher among subjects who survive 5 years relative to those who die within 5 years. This observed difference is statistically different from 0 (P < 0.05), with a 95% confidence interval suggesting that the observed difference is what might be typically observed if the true difference between survivors and nonsurvivors was such that the geometric mean for survivors was anywhere between 0.4% and 3.8% higher than that for nonsurvivors. We thus reject the null hypothesis of no association between survival time and ldl at study entry in favor of a trend toward higher ldl among subjects surviving the longer period of time.**

1. Perform a statistical analysis evaluating an association between serum LDL and 5 year all-cause mortality by comparing the probability of death within 5 years across groups defined by whether the subjects have high serum LDL (“high” = LDL > 160 mg/dL).

**Ans: For participants with ldl above 160mg/dL, 13out of 13 (12.6%) die within five years. For participants with ldl below 160mg/dL, 106 out of 622(17.0%) die after five years of study enrollent. Based on the chi squared test, the observed absolute difference of 4.4% in 5 year all-cause mortality is what might be expected to occur by chance in the absence of a true effect of high ldl (P = 0.2619; 95% CI for difference in 5 year all-cause mortality is an absolute -10.14% to 2.64 % higher 5 year all-cause mortality in the participants with ldl above 160mg/dL).**

|  |  |  |
| --- | --- | --- |
| ldl | death wihtin 5 years | total |
| yes | no |
| >=130mg/dL | 13 | 90 | 103 |
| <130mg/dL | 106 | 516 | 622 |
| total | 119 | 606 | 725 |

1. Perform a statistical analysis evaluating an association between serum LDL and 5 year all-cause mortality by comparing the odds of death within 5 years across groups defined by whether the subjects have high serum LDL (“high” = LDL > 160 mg/dL).

**Ans: Based on the chi squared test, the odds of dying within 5 years is estimated to be 29.7% lower (odds ratio 0.7031) for each 10 mg/dl increase in ldl levels. This observed difference is statistically different from an odds ratio of 1 (P =0.2619), with a 95% confidence interval suggesting that the observed odds ratio is what might be typically observed if the true odds of dying within 5 years was between 30.44% and 62.10% lower in subjects with ldl level above 160 mg/dL. We thus could reject the null hypothesis of no association between survival time and high ldl at study entry in favor of a trend toward higher odds of survival among subjects with higher ldl levels.**

1. Perform a statistical analysis evaluating an association between serum LDL and all-cause mortality over the entire period of observation of these subjects by comparing the instantaneous risk of death across groups defined by whether the subjects have high serum LDL (“high” = LDL > 160 mg/dL).

**Ans: Based on the Log-rank test, we reject (p<0.001) the null hypothesis of equal survival probabilities between subjects with ldl above 160 mg/dL and those with ldl below 160 mg/dL.**

1. Supposing I had not been so redundant (in a scientifically inappropriate manner) and so prescriptive about methods of detecting an association, what analysis would you have preferred *a priori* in order to answer the question about an association between mortality and serum LDL? Why?

**Ans:3 and 4 dichotomized the survival time distribution at 5 years and modeled the ldl level distribution continuously, while 5-7 dichotomized both the survival time distribution at 5 years and the ldl level. I prefer t o compare the instantaneous risk of death but do not dichotomize the ldl level.**