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.

***Methods:***

Summarize the censoring time and the minimum censoring time was found as 5.002 which is greater than 5. Therefore, dichotomize the time to death within 5 years and after 5 years is valid for the statistical methods on problems 2-6.

deathin5 :

|  |  |  |
| --- | --- | --- |
|  | Frequency | Proportion |
| Death within 5 years | 121 | 16.5% |
| Death after 5 years | 614 | 83.5% |

After dichotomize the time to death into a binary variable (deathin5) which is not a censoring data, we can perform the following analysis.

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.

***Methods:***

|  |  |  |  |
| --- | --- | --- | --- |
|  | **Death within 5 years****Mean (SD; Min Mdn Max; n)****n (%)** | **Death after 5 years** **Mean (SD; Min Mdn Max; n)****n (%)** | **All participants** **Mean (SD; Min Mdn Max; n)****n (%)** |
| **Age (years)** | 76.5 (6.2; 67 75 91; n=121) | 74.2 (5.2; 65 73 99; n=614) | 74.6 (5.5; 65 74 99; n=735) |
| **Weight (lbs)** | 159.1 (32.8; 96 154 264; n=121) | 160.1 (30.3; 74 158.8 258; n=614) | 159.9 (30.7; 74 158 264; n=735) |
| **Smoking history (pack a year)** | 28.0 (36.0; 0 18.4 240; n=120) | 17.9 (24.7; 0 4.35 180; n=614) | 19.6 (27.1; 0 6.5 240; n=734) |
| **Serum LDL (mg/dl)** | 118.7 (36.2; 11 117 227; n=119) | 127.2 (32.9; 39 127 247; n=606) | 125.8 (33.6; 11 125 247; n=725) |
| **Sex** |  |  |  |
|  Female | 43 (35.5) | 326 (53.1) | 369 (50.2) |
|  Male | 78 (64.5) | 288 (46.9) | 366 (49.8) |
| **CHD** |  |  |  |
|  No | 75 (62.0) | 505 (82.2) | 580 (78.9) |
|  Angina | 17 (14.0) | 47 (7.7) | 64 (8.7) |
|  Myocardial infraction | 29 (24.0) | 62 (10.1) | 91 (12.4) |
| **CHF** |  |  |  |
|  No | 104 (86.0) | 590 (96.1) | 694 (94.4) |
|  Yes | 17 (14.0) | 24 (3.9) | 41 (5.6) |
| **Stroke** |  |  |  |
|  No | 86 (71.1) | 550 (89.6) | 636 (86.5) |
|  Transient ischemic attack | 7 (5.8) | 17 (2.8) | 24 (3.3) |
|  Stroke | 28 (23.1) | 47 (7.6) | 75 (10.2) |

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.

***Methods:***

Two samples t test was performed to compare mean LDL values between participants who died in 5 years and participants who died after 5 years.

***Inference:***

The mean serum LDL of the participants who died within 5 years is 118.7 mg/dl (n=119) and the mean serum LDL of the participants died after 5 years is 127.2 mg/dl (n=606). The estimated mean serum LDL of the participants who died after 5 years is 8.5 mg/dl higher than the participants who died within 5 years. The observed data indicate the means serum LDL for two groups is statistically different from 0 (P=0.0186, 95% CI=(1.44, 15.6)). Therefore, we reject null hypothesis of no association between serum LDL values and 5 years survival time, and conclude that higher LDL values are associated with survival time exceed 5 years.

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.

***Methods***

Log transform on LDL data and perform two samples t test to compare geometric mean LDL values between participants who died in 5 years and participants who died after 5 years.

***Inference:***

The geometric mean serum LDL of the participants who died within 5 years is 112.0 mg/dl (n=119) and the geometric mean LDL of the participants died after 5 years is 122.8 mg/dl (n=606). The estimated geometric mean LDL of the participants who died after 5 years is 9.65% statistically higher than the participants who died within 5 years (P=0.0128, 95% CI 2.01% higher to 17.9% higher). Therefore, we reject null hypothesis of no association between serum LDL values and 5 years survival time, and conclude that higher LDL values are associated with the survival time exceed 5 years.

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).

***Methods:***

Dichotomize LDL in groups defined by whether the participants who have serum LDL higher than 160 mg/dl and perform Chi-Square test to compare the probability of death within 5 years across groups.

***Inference:***

The probabilities of death within 5 years for participants who have serum LDL higher than 160 mg/dl and for participants who have serum LDL less than or equal to 160 mg/dl are 14% and 17%, respectively. The observed probabilities of death within 5 years across groups are not statistically different (P=0.3753). The estimated risk ratio is 0.80 and 95% CI indicate the observed risk ratio is not surprising if the true risk ratio is between 0.49 and 1.31. Therefore, we can’t reject the null hypothesis of no association between serum LDL and 5 year all-cause mortality.

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).

***Methods:***

Dichotomize LDL in groups defined by whether the subjects have serum LDL higher than 160 mg/dl or not and perform odds ratio test to compare the odds of death within 5 years across groups.

***Inference***

The odds of death within 5 years for participants who have serum LDL higher than 160 mg/dl and for participants who have serum LDL less than or equal to 160 mg/dl are 0.16 and 0.20, respectively. The observed odds of death within 5 years across groups are not statistically different (P=0.3753). The estimated odds ratio is 0.77 and 95% CI indicate the observed risk ratio is not surprising if the true odds ratio is between 0.41 and 1.39. Therefore, we can’t reject the null hypothesis of no association between serum LDL and 5 year all-cause mortality.

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).

***Methods:***

Dichotomize LDL in groups defined by whether the participants have serum LDL higher than 160 mg/dl and perform a log-rank test to compare the instantaneous risk of death across groups.

***Inference:***

Based on P=0.2664, it indicates that the survival distribution across groups defined by whether the participants have high serum LDL is not statistically different. Therefore, we can’t reject the null hypothesis that the probabilities of survival are the same for the participants who serum LDL are higher than 160 mg/dl and for the participants who serum LDL are less than or equal to 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? Choose which one is the best method and explain why. scientific reason.

***Methods:***

I would prefer a log-rank test to compare the survival distribution across groups defined by whether the participants have serum LDL higher than 160 mg/dl. I would consider serum LDL values as the primary of interest and mortality as the response, since many research have suggested serum LDL values could be a risk factor for mortality. Therefore, investigate the mortality across groups defined by serum LDL values would be more appropriate.