**Question 1**

*Methods*:

Determine when the first censored data point occurred within the dataset, by finding the minimum value for time-to-death, where a death was not observed.

*Inference*:

Time to death can be dichotomized at the 4-year point for this dataset because the first censored data occurred after 4 years. In fact, the minimum time to follow-up for the censored observations was 1480 days, which is over 4 years.

**Question 2**

*Methods*:

The data was divided into four subsets based on serum CRP levels (low risk, average risk, high risk, and all levels). Descriptive statistics were calculated within each of these groups for age, gender, BMI, smoking status, cholesterol, and if they had a prior history of cardiovascular disease (angina, MI, TIA, stroke). Death within four years of study enrollment was also determined by creating a dichotomous variable from time to death, with 1460 days as the threshold; no data was censored before 1460 days. For the continuous variables (age, BMI and cholesterol) the mean, standard deviation, minimum and maximum values were calculated. For the binary variables (sex, smoking status, death within 4 years and prior history of cardiovascular disease), percentages were used.

*Inference*:

The following table provides the descriptive statistics for the 5000 subjects, based on their serum CRP levels as it relates to cardiovascular risk (low, average, high, all).

|  |  |
| --- | --- |
|  | **Serum C-Reactive Protein (CRP) Levels** |
|  | **< 1 mg/L****(N = 428)** | **1 – 3 mg/L****(N = 3330)** | **> 3 mg/L****(N = 1175)** | **All****(N = 5000)** |
| **Death w/in 4 years (%)** | 4.9% | 8.4% | 15.6% | 9.9% |
| **Age (years) 1** | 73.4 (5.80; 65 – 94) | 72.7 (5.52; 65 – 100) | 72.7 (5.58; 65 – 93) | 72.8 (5.60; 65 – 100) |
| **Male (%)** | 45.6% | 43.3% | 37.0% | 41.9% |
| **BMI 1** | 23.8 (3.64; 15.6 – 38.6) | 26.4 (4.31; 14.7 – 53.2) | 28.4 (5.463; 15.3 – 58.8) | 26.7 (4.74; 14.7 – 58.8) |
| **Smoker (%)** | 9.6% | 10.9% | 16.4% | 12.1% |
| **Cholesterol (mg/dl) 1** | 206 (40.53; 109 – 407) | 212.8 (38.57; 73 – 363) | 210.5 (40.39; 97 – 430) | 211.7 (39.29; 73.0 – 430) |
| **History of cardio. disease (%)** | 18.2% | 21.4% | 28.8% | 22.9% |

Note: 1 Mean (standard deviation; minimum – maximum)

Note that for BMI there was a total of 14 missing data points (12 for the 1 – 3 mg/L serum CRP group and 1 for the group with levels greater than 3 mg/L). For cholesterol there was a total of 47 missing data points (2 for the greater than 3 mg/L group and 1 for the less than 1 mg/L group).

**Question 3**

*Methods*:

The mean serum CRP levels were compared between subjects who died within 4 years and who survived at least 4 years. A t test assuming unequal variances was used. The results include a 95% confidence interval.

*Inference*:

The mean serum CRP level for the 405 subjects that died within 4 years was 5.376 mg/L. The mean serum CRP level for the 4505 subjects that survived at least 4 years was 3.422 mg/L. The t test revealed the 95% confidence interval for the difference was [1.210, 2.697], which was statistically significant at p < 0.0001. Therefore, there was a difference between mean serum CRP levels between those that survived and died, in which the survivors had, on average, lower levels.

**Question 4**

*Methods*:

The serum CRP levels were log transformed to compare the geometric mean levels between the subjects that survived and died at the 4 year mark. A t test assuming unequal variances was used, with *α* = 0.05. Since the dataset included 0’s for CRP levels, these were adjusted to allow for log transformation. Each 0 was replaced with the lowest, nonzero CRP level for that group. For each group, the lowest observed level was 1, therefore all zeros were replaced with 1.

*Inference*:

The geometric mean serum CRP level was 1.119 mg/L for the 405 subjects that died within 4 years. The geometric mean serum CRP level was 0.771 mg/L for the 4505 subjects that survived at least 4 years. The t test revealed a 95% confidence interval for the difference of [0.257, 0.438], with p < 0.0001. Therefore, there was a difference between the geometric mean levels, where the survivors had a lower mean CRP level.

**Question 5**

*Methods*:

A chi square test of independence was used to determine if there was a difference in the probabilities of death for subjects with serum CRP levels greater than 3 mg/L and ≤ 3 mg/L. A matrix was created with the number of subjects who survived and died at the four years in each of these two CRP level groups.

*Inference*:

For the 1175 subjects with serum CRP levels greater than 3 mg/L, 18.4% died within 4 years. For the remaining 3758 subjects with serum CRP levels less than 3 mg/L, 8.0% died within 4 years. Using a chi square test, there is a difference between these proportions for the probability of death for serum CRP levels greater than 3 mg/L and ≤ 3 mg/L. The test revealed that for p < 0.0001, the probability of death was higher for those with high risk levels of serum CRP.

**Question 6**

*Methods*:

The odds of dying within 4 years was calculated for serum CRP levels ≤ 3 mg/L and > 3 mg/L. The Fisher’s exact test was used, by comparing the odds ratio to 1.

*Inference*:

For the 1175 subjects with serum CRP levels greater than 3 mg/L, the odds of dying within 4 years was 0.087. For the remaining 3758 subjects with serum CRP levels less than 3 mg/L, the odds of death was 0.184. This odds ratio is 0.472. The Fisher’s exact test had a p value < 0.001, which suggests that we can reject the null hypothesis and conclude that serum CRP levels and 4 year mortality are associated.

**Question 7**

*Methods*:

A Kaplan-Meier survival curve was used to estimate the survival distribution between these two groups (high risk and less than high risk). From this, the two groups were tested using the logrank statistic.

*Inference*:



|  |  |
| --- | --- |
|  | Survival Probabilities (Kaplan-Meier) |
|  | CRP ≤ 3 mg/L | CRP > 3 mg/L |
| 1 years | 0.987 | 0.966 |
| 2 years | 0.972 | 0.928 |
| 3 years | 0.948 | 0.881 |
| 4 years | 0.920 | 0.843 |

The above graph and table are based on the computed Kaplan-Meier survival estimates. These suggest that the instantaneous risk of death is higher for those with high-risk serum CRP levels. The two-sided logrank test supports this, with p < 0.001. Therefore we can reject the null hypothesis.

**Question 8**

The above methods lost precision by transforming the continuous variable of time to death into a dichotomous variable, with the 4-year threshold. Perhaps a more appropriate method to detect association would have been to use regression. For this, the dependent variable could have been kept continuous and a linear regression model could have been used. However, if the 4-year threshold of survival was used as the dependent variable, a binary logit model could be used. In these methods of regression, we could account for other independent variables other than serum CRP levels, such as gender, age, smoking habits, and cardiovascular history.