Homework 1

1. Using 4 years to dichotomize time of death is valid, because the earliest censored observation occurs at 1480 days, a little more than 4 years. Therefore, at 4 years, we still have complete information for all individuals enrolled in the study.

Good. 5 points

1. **Methods:** Descriptive statistics are presented for groups defined by Serum C Reactive Protein (CRP) levels as well as the entire sample. Based on prior scientific knowledge provided by the Mayo Clinic, CRP groups are separated as follows: less than 1 mg/L, 1 – 3 mg/L, and above 3 mg/L. For continuous variables we present: mean (standard deviation; min – max). For binary variables we present percentages.

**Results:** We computed our analyses with data from 5000 individuals, of which 67 are missing data for CRP levels. Note that the effect on generalizability of omitting these patients cannot be quantified. Some of the individuals without missing CRP levels are missing data in other variables. Analyses for the variables other than our 2 prime variables of interest are computed by ignoring (omitting) any missing values.

Of the 4933 individuals we ran analyses on, 428 had low CRP levels (below 1 mg/L), 3330 had medium CRP levels (1 – 3 mg/L) and 1175 had high CRP levels (above 3 mg/L). The table below presents descriptive statistics for these 3 groups along with a column for all 4933 subjects for which we had CRP data. Individuals with higher CRP levels were more likely to be female. There was no clear trend with respect to age or cholesterol for the different CRP groups. There was a positive correlation between CRP level and BMI, smoker status, prior CVD and death within 4 years. The mortality rate for the different groups were: 4.9% for the low group, 8.4% for the medium group, 15.6% for the high CRP group and 9.8% overall.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|   |   |  Serum C Reactive Protein (CRP) |   |   |
|   | Below 1 mg/L | 1 - 3 mg/L | Above 3 mg/L | All Subjects |
| Male (%) | 45.6% | 43.3% | 37.0% | 42.0% |
| Age (yrs) | 73.5 (5.80; 65 - 94) | 72.7 (5.52; 65 - 100) | 72.7 (5.58; 65 - 93) | 72.8 (5.56; 65 - 100) |
| BMI  | 23.8 (3.64; 15.6 - 38.6) | 26.4 (4.31; 14.7 - 53.2) | 28.5 (5.46; 15.3 - 58.8) | 26.7 (4.72; 14.7 - 58.8) |
| Smoker (%) | 9.6% | 11.0% | 16.4% | 12.2% |
| Cholesterol (mg/dl) | 206 (40.5; 109 - 407) | 213 (38.6; 73 - 363) | 211 (40.4; 97 - 430) | 212 (39.2; 73 - 430) |
| Prior CVD (%) | 18.2% | 21.5% | 28.8% | 22.9% |
| Death w/in 4 years | 4.9% | 8.4% | 15.6% | 9.8% |

* For continuous variable we present mean (standard deviation; min – max)

Good, 10 points. It might help to present the ‘n’ for each CRP category in the table as well as the text, and to indicate the number missing for each variable as Scott did in his key.

1. **Methods:** We compared the mean CRP levels between the individuals who were alive after 4 years and the individuals who died within 4 years. The difference in the mean CRP levels was analyzed using a t test and 95% confidence interval; neither assuming equal variances.

**Results:** The mean CRP level was 3.4 mg/L for the 4449 individuals who survived beyond 4 years after enrollment in the study, and 5.4 mg/L for the 484 who died within 4 years of study enrollment. Based on our 95% confidence interval, our observed difference in mean CRP of -2.0 mg/L (alive – dead) would not be surprising if the true difference in population mean CRP between people who survived more than 4 years and those who died within 4 years was between -2.7 mg/L and -1.2 mg/L. Based on the results of our two sample t test without assuming equal variances at a significance level of .05, we reject the null hypothesis (two-sided p-value < .0001) that the mean CRP levels are equal for those who were alive after 4 years and those who died within 4 years. This data suggests that dying within 4 years is associated with higher CRP levels.

Good. 10 points.

1. **Methods:** We compared the geometric mean CRP levels between the individuals who were alive after 4 years and the individuals who died within 4 years. The difference in the geometric mean CRP levels was analyzed using a t test and 95% confidence interval; neither assuming equal variances.

**Results:** The geometric mean CRP level was 2.03 mg/L for the 4449 individuals who survived beyond 4 years after enrollment in the study, and 2.97 mg/L for the 484 who died within 4 years of study enrollment. Based on our 95% confidence interval, our observed difference in geometric mean CRP of -.94 mg/L (alive – dead) would not be surprising if the true difference in population geometric mean CRP between people who survived more than 4 years and those who died within 4 years was between -1.72 mg/L and -.17 mg/L. Based on the results of our two sample t test without assuming equal variances at a significance level of .05, we reject the null hypothesis (two-sided p-value = .017) that the geometric mean CRP levels are equal for those who were alive after 4 years and those who died within 4 years. This comparison suggests that dying within 4 years is associated with higher CRP levels.

7 points: -1 for not reporting how the geometric mean was calculated or how zeros were handled, -1 for reporting differences instead of ratios, and -1 for incorrect p-value.

1. **Methods:** We compared the proportions of individuals who died within 4 years between people with high CRP levels (above 3 mg/L) and everyone else (CRP less than or equal to 3 mg/L). We tested the differences in these proportions using the Pearson’s Chi-Squared test for independence at a .05 significance level. We also used the Wald statistic to compute a confidence interval for the difference in proportions of individuals who died within 4 years between the high and low CRP groups.

**Results:** 183/1175 (15.6%) of the individuals with high CRP (above 3 mg/L) died within 4 years of enrollment in the study, and 301/3758 (8.0%) of the individuals with CRP at or below 3 mg/L died within 4 years of enrollment. Based on our 95% confidence interval, our observed difference in mortality proportion of -7.6% (alive – dead) would not be surprising if the true difference in mortality proportion between people with high CRP and people with lower CRP was between -9.8% and -5.3%. Our chi-squared test for independence produced a p-value < .0001, and therefore, we reject the null hypothesis that CRP level and mortality proportion are independent. This data suggests that higher CRP levels are associated with higher mortality proportion.

10 points.

1. **Methods:**  We compared the odds of dying within 4 years of study enrollment between people with high CRP (above 3 mg/L) and everyone else (less than or equal to 3 mg/L). We tested the null hypothesis that the odds ratio was 1 with a chi squared test with .05 level of significance. We also produced a 95% confidence interval using the Wald statistic.

**Results:** For the 1175 people who had a high CRP (above 3 mg/L), the odds of dying within 4 years of enrollment was .184, while the 3758 people with lower CRP had an odds of dying of .087. Based on our 95% confidence interval, the observed odds ratio of 2.119 would not be unusual if the true population odds ratio was between 1.740 and 2.980. A Chi-squared test at significance level of .05 produced a p-value < .0001. Therefore, we reject the null hypothesis that the true odds ratio is 1, providing evidence that higher levels of CRP are associated with higher odds of death within 4 years.

 10 points

1. **Methods:** We analyzed overall mortality between the 1175 people with high CRP (above 3 mg/L) and the 3758 people with low CRP (at or below 3 mg/L) using Kaplan-Meier estimates. We tested the differences in these survival curves using a logrank test. We also computed the hazard ratio with a 95% confidence interval using cox proportional hazards regression.

**Results:** The graph and table below illustrate the Kaplan-Meier curves and survival probability estimates for people with high CRP and low CRP. As you can see, at all time points the probability for survival is greater for the people with lower CRP than people with high CRP. The instantaneous risk of death is estimated to be 68.7% higher for people with high CRP (above 3 mg/L). The corresponding hazard ratio of 1.687 would not be unusual if the true population hazard ratio was between 1.486 and 1.915, based on a 95% confidence interval. A two-sided logrank test yielded a p-value < .0001, suggesting that we reject the null hypothesis that the hazard ratio is 1. This data suggests that higher CRP levels are associated with lower probability of survival.

|  |
| --- |
|  Survival Probabilities (Kaplan-Meier) |
|   | CRP <= 3 | CRP > 3 |
| 1 year  | 0.987 | 0.966 |
| 2 years | 0.971 | 0.926 |
| 3 years | 0.948 | 0.881 |
| 4 years | 0.920 | 0.844 |
| 5 years | 0.884 | 0.800 |

What is going on with the thick bands in the survival plots? Are those clusters of censored data? It would be good to provide a key to interpret this for the reader.

10 points

1. I would have preferred using the Kaplan-Meier survival estimates, comparing survival curves across CRP levels that are scientifically predetermined. Here are some reasons why:
* Using Kaplan-Meier allows us to use all data, including data beyond the first censored observation.
* Plotting and summarizing the survival probabilities allows us to ascertain the scientific significance (not just statistical significance) in the difference between survival distributions.
* The hazard ratio confidence interval gives us a range of the expected increase/decrease in instantaneous risk of death between CRP levels.
* Kaplan-Meier is the only method we know of which can incorporate censored observations.

0 points. I think you make good points (and I agree with them), but I don’t see any that match Scott’s list of points or his conclusion.