Biostats HW#1

1.) The minimum follow up time for censored observations in this study is 1480 days, just over 4 years. Therefore, the vital status is known for each individual in the study at 4 years.

2.)

**Methods:**

 An indicator variable was created for death within 4 years after study enrollment (no patients were censored during the first 4 years). Descriptive statistics are presented for all subjects with a CRP value, as well as for groups defined by categorized CRP values that correspond to low, average, and high risk of heart disease (with categories < 1 mg/L representing low risk, 1-3 mg/L representing average risk, and > 3 mg/L representing high risk of heart disease). Descriptive statistics provided for continuous variables include the mean, standard deviation, minimum, and maximum. For binary variables, percentages are reported.

**Results:**

 Of the 5000 subjects included in the study, 67 do not have CRP values and therefore are omitted from the study. Of the remaining 4933 subjects, 6 are missing data on smoking status, 13 on BMI, 10 on systolic blood pressure, and 117 on ankle-arm blood pressure index. This missing data may introduce bias and affect generalizability of our results.

 Of the 4933 subjects with CRP results, 428 subjects had values less than 1 mg/dL, 3330 subjects had values between 1 mg/dL and 3 mg/dL inclusive, and 1175 subjects had values greater than 3 mg/dL. The table below presents the descriptive statistics for these groups and for all subjects with CRP values. As seen in the table, there is no clear difference in age between the groups. Subjects in the CRP > 3 mg/L group were more likely to be female than subjects in the other two groups. Subjects in the CRP < 1 mg/L group had lower mean cholesterol and systolic blood pressure compared to the other two groups. Mean BMI, prevalence of smoking, prevalence of prior cardiovascular disease, and incidence of death within 4 years of study enrollment were all higher in each successively higher CRP group (ie the highest interval group has larger values than the intermediate interval group, and the intermediate interval group has higher values than the lowest interval group). The mean ankle-arm blood pressure index is lower in each successively higher CRP group.

|  |  |
| --- | --- |
|  | **Serum C Reactive Protein (CRP)** |
|  | **< 1 mg/L****(n=428)** | **1-3 mg/L****(n=3330)** | **> 3 mg/L****(n=1175)** | **Any Level****(n=4933)** |
| **Male (%)** | 45.6% | 43.3% | 37.0% | 42.0% |
| **Age (yrs)1** | 73.5 (5.80; 65-94) | 72.7 (5.52; 65-100) | 72.7 (5.58; 65-93) | 72.8 (5.56; 65-100) |
| **BMI1** | 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) |
| **Cholesterol (mg/dL)1** | 206.0 (40.5; 109-407) | 212.8 (38.6; 73-363) | 210.5 (40.4; 97-430) | 211.7 (39.2; 73-430) |
| **Systolic BP1**  | 132.7 (21.3; 89-199) | 136.6 (21.8; 82-235) | 137.5 (21.8; 77-230) | 136.5 (21.8; 77-235) |
| **Ankle-arm BP index1** | 1.085 (0.155; 0.317-1.632) | 1.072 (0.170; 0.278-2.385) | 1.034 (0.191; 0.304-1.888) | 1.064 (0.175; 0.278-2.385) |
| **Smoker (%)** | 9.6% | 11.0% | 16.4% | 12.2% |
| **Prior CV disease2 (%)** | 18.2% | 21.4% | 28.7% | 22.9% |
| **Death w/in 4 yrs (%)** | 5.9% | 8.4% | 15.5% | 9.8% |

1Descriptive statistics presented are the mean (standard deviation; minimum-maximum)

2Prior CV (cardiovascular) disease includes any history of previous angina, myocardial infarction, transient ischemic attack, or stroke at study enrollment.

BP = blood pressure.

3.)

**Methods:**

To examine the difference in mean CRP values between subjects that died within 4 years of enrollment and subjects that survived at least 4 years, we performed a two sample t test allowing for unequal variances. 95% confidence intervals were also calculated allowing for unequal variances.

**Results:**

Among the 484 subjects that died within 4 years after study enrollment, mean serum CRP was 5.38 mg/L. Among the 4449 subjects that survived at least 4 years, mean serum CRP was 3.42 mg/L. The mean serum CRP level was 1.95 mg/L higher in subjects that died within 4 years (95% confidence interval 1.21 mg/L to 2.69 mg/L). If the true population difference in mean CRP is between 1.21 mg/L and 2.26 mg/L, our data would not be unusual. This observed difference is statistically significant at an alpha of 0.05 (p < 0.0001); therefore, we reject the null hypothesis that mean CRP levels are the same in each group in favor of the alternative hypothesis that death within 4 years is associated with higher mean serum CRP levels.

4.)

**Methods:**

To examine the difference in geometric mean CRP values between subjects that died within 4 years of enrollment and subjects that survived at least 4 years, we first log transformed CRP values. We then performed a two sample t test that allows for unequal variances on the log transformed values. 95% confidence intervals for difference in population means for log CRP were also calculated allowing for unequal variances. Estimates and confidence intervals were than exponentiated in order to obtain inference on the geometric mean.

**Results:**

Among the 484 subjects that died within 4 years after study enrollment, geometric mean serum CRP was 2.86 mg/L. Among the 4449 subjects that survived at least 4 years, geometric mean serum CRP was 1.87. The geometric mean serum CRP level was 0.654 mg/L higher in subjects that died within 4 years (95% confidence interval 0.590 mg/L to 0.725 mg/L). If the true population difference in geometric mean CRP is between 0.590 mg/L and 0.725 mg/L, our data would not be surprising. This observed difference is statistically significant at an alpha level of 0.05 (p < 0.0001); therefore, we reject the null hypothesis of no difference in geometric mean CRP between groups in favor of the null hypothesis that death within 4 years of study enrollment is associated with higher geometric mean CRP.

5.)

**Methods:**

To examine the difference in proportion of subjects dying within 4 years of study enrollment in subjects with CRP >3 mg/L compared with subjects with CRP <=3 mg/L, we performed a chi squared test of independence. 95% confidence intervals for the difference in population 4 year probability of death were calculated using Wald statistics.

**Results:**

Of the 1175 subjects with CRP > 3 mg/L, 15.6% died within 4 years of study enrollment. Of the 3758 subjects with CRP <= 3 mg/L, 8.0% died within 4 years of study enrollment. The proportion that died within 4 years was 7.6% higher in the “high” CRP group (95% confidence interval 5.3% to 9.8%). If the true population difference in probability of death within 4 years is between 5.3% and 9.8%, our data would not be surprising. The observed difference in proportion dying within 4 years is statistically significant at an alpha of 0.05 (p <0.0001); therefore, we reject the null hypothesis of no difference in probability of death within 4 years between groups in favor of the alternative hypothesis that “high” serum CRP levels are associated with a higher probability of death within 4 years after study enrollment.

6.)

**Methods:**

To compare the odds of dying within 4 years of study enrollment in subjects with serum CRP > 3 mg/L compared with subjects with serum CRP <= 3 mg/L, an odds ratio different from 1 was tested using the chi squared test of independence. 95% confidence intervals for the odds ratio were computed using the Cornfield method.

**Results:**

Of the 1175 subjects with CRP > 3 mg/L, the odds of death within 4 years of study enrollment is 0.184. Of the 3758 subjects with CRP <=3 mg/L, the odds of death within 4 years is 0.0871. The yields an odds ratio of 2.12 (95% confidence interval 1.74 to 2.58). If the true odds ratio in the population were between 1.74 and 2.58, our data would not be surprising. The observed odds ratio is statistically significant at an alpha of 0.05 (p < 0.0001); therefore, we reject the null hypothesis that the odds ratio is 1 in favor of the alternative hypothesis that “high” serum CRP levels are associated with a higher odds of death within 4 years after study enrollment.

7.)

**Methods:**

Using Kaplan-Meier methods, we estimated the survival distribution for 2 strata: those with CRP > 3 mg/L, and those with CRP <= 3 mg/L. We used the logrank statistic to test for differences in survival distributions between these two strata. The hazard ratio and 95% confidence interval were calculated using Cox proportional hazards regression.

**Results:**

The graph and table below show Kaplan-Meier survival estimates for 2 different groups: the 1175 subjects with CRP > 3 mg/L, and the 3758 subjects with CRP <=3 mg/L. As can be seen in the graph and table, the lower CRP group has a higher survival probability at every point in time. The instantaneous risk of death is estimated to be 68.7% higher for the higher CRP group (hazard ratio 1.687, 95% confidence interval 1.485 to 1.917). If the true hazard ratio in the population is between 1.485 and 1.917, our data would not be surprising. Based on a two sided logrank test p value of < 0.0001, we reject the null hypothesis of no difference in the probability of survival between the two groups in favor of the alternative hypothesis that higher serum CRP is associated with lower probability of survival.



|  |  |
| --- | --- |
|  | **Survival Probabilities (Kaplan Meier)**  |
|  | **CRP > 3 mg/L** | **CRP <= 3 mg/L** |
| **1 year** | 0.967 | 0.988 |
| **2 years** | 0.962 | 0.971 |
| **3 years** | 0.881 | 0.948 |
| **4 years** | 0.844 | 0.920 |
| **5 years** | 0.800 | 0.884 |
| **6 years** | 0.755 | 0.853 |
| **7 years** | 0.710 | 0.810 |
| **8 years** | 0.647 | 0.769 |

8.)

I prefer an analysis that conditions on CRP for 2 reasons: the blood measurement must occur before death, and it may be that intervention targeted a lowering CRP could change survival (this question cannot be answered in this current study, but establishing an association between CRP levels and survival in this study could lead to further studies aimed at intervening on CRP levels). I also prefer an analysis that does not split continuous variables into categories unless there is a clear reason for doing so because in dividing the variable into a few categories, you necessarily lose some information in the data. Finally, for laboratory data I prefer to analyze geometric means. Therefore, I prefer the analysis from question #4 that performs a two sample t test (that allows for unequal variances) of geometric mean CRP levels, comparing those that died within 4 years and those that survived at least 4 years.