Bios 515/518

Homework #1

1. If all participants enrolled in the study either survives or dies and no one is censored during the study, then we know all the information about the population within this defined time period. A subset of 735 participants is included in the MRI dataset. Each of these patients have a different amount of observation time depending on the date of the MRI procedure. Of these 735 participants, 121 participants were enrolled in the study for longer than five years and 614 were enrolled for less than or equal to five years.

|  |
| --- |
| Table 1. Descriptive Statistics of LDL, 5-year survival, and Risk Factors in the MRI and Cerebral Atrophy Study |
|  | 5 Year Survival = Yes | 5 Year Survival = No |
|  | High LDL n=100 | Low LDL n=502 | High LDL n=17 | Low LDL n=116 |
| Risk Factor | Mean/Proportion (SD) | Mean/Proportion (SD) | Mean/Proportion (SD) | Mean/Proportion (SD) |
| Sex=Male | 0.40 | 0.48 | 0.65 | 0.65 |
| Weight (lbs) | 163.7 (30.4) | 159.4 (30.3) | 159.4 (31.2) | 159.2 (33.1) |
| Smoking: # Pack Years  | 16.8 (23.6) | 18.2 (24.9) | 26.4 (27.6) | 27.3 (36.3) |
| CHD  | 0.25 | 0.28 | 0.88 | 0.54 |
| CHF | 0.02 | 0.04 | 0.12 | 0.14 |
| Stroke | 0.25 | 0.16 | 0.76 | 0.49 |

*Methods*: I did a single variable analysis of sex, weight, smoking history, and history of CHD, CHF, or stroke diagnosis stratified on LDL level and 5-year survival status.

*Inference*: Women were more likely than men to survive regardless of LDL status, 65% of the patients who did not survive for 5 years post-MRI were men. Body weight did vary across LDL level or survival status. Reported history of smoking was lower among patients with 5-year survival regardless of LDL level, mean 16.8 pack years (standard deviation 23.6) compared to 26.4 (S.D. 27.6) for high LDL. Finally, history of CHD, CHF, and stroke diagnosis was higher among patients who did not survive to five year observation period; in the high LDL subset 88% vs 25% where diagnosed with CHD, 12% vs 2% were diagnosed with CHF, and 76% vs 25% had a history of stroke, respectively.

3. *Methods*: I generated a new variable for LDL that describes the different levels of risk for a given LDL level. A Pearson’s chi-square of significance was used to test the hypothesis that the distribution of LDL level is independent of survival status. The alternative hypothesis is that the distribution of LDL level is not independent of survival status.

*Inference*: Given a chi-square value of 14.13, we reject the null hypothesis that LDL level is independent of survival status, p-value=0.015.

4. *Methods*: I took the natural log of the response variable, LDL, conducted a chi-square test, and back transformed the estimate to test the hypothesis that the distribution of the geometric mean of LDL is independent of survival status.

*Inference*: Given a chi-square value of 199.3, we reject the null hypothesis that LDL level is independent of survival status, p-value=0.004.

5. *Methods*: LDL level was dichotomized to compare high LDL levels (>=160 mg/dl) to normal LDL levels (<160 mg/dl). I performed a two-sample t-test testing the hypothesis that the patients with LDL levels higher than 160 mg/dl (high) have the same survival probability as patients with LDL levels less than 160 mg/dl (normal).

*Inference*: When we dichotomize LDL, we fail to reject the null hypothesis; the data is consistent with no difference between high and normal LDL levels and survival status.

6. *Methods*: LDL level was dichotomized to compare high LDL levels (>=160 mg/dl) to normal LDL levels (<160 mg/dl). The null hypothesis is that the odds of death will be the same in both the high and low LDL groups.

*Inference*: The odds of death among the patients with LDL levels higher than 160 mg/dl was 0.73 times the odds of death among patients with LDL levels less than 160 mg/dl, 95% CI (0.39, 1.30). We fail to reject the null hypothesis; there is not enough evidence to conclude that odds of death are different among patients with high serum LDL levels.

7. *Methods*: I performed a survival analysis on the data using observation time and death as the outcome variable. A cox proportional hazard ratio was calculated to test the null hypothesis that survival time will be the same for patients with serum LDL levels >= 160 mg/dl (high) and serum LDL levels <160 mg/dl (normal).

*Inference*: The hazard ratio is 0.74, 95% CI (0.44, 1.24) and a p-value of 0.267. We fail to reject that null hypothesis that there is a difference between the survival time between patients with high and normal LDL serum levels.

8. I would have chosen a simple logistic regression between the predictor of interest, serum LDL level and response variable, death. By using a binary predictor of interest, the output could be used in a clinical setting to better understand the relationship between serum LDL and survival outcomes. If serum levels of LDL higher than 160 mg/dl have a negative association with survival, then a clinical intervention may improve patient outcomes.