Homework 1

31/40

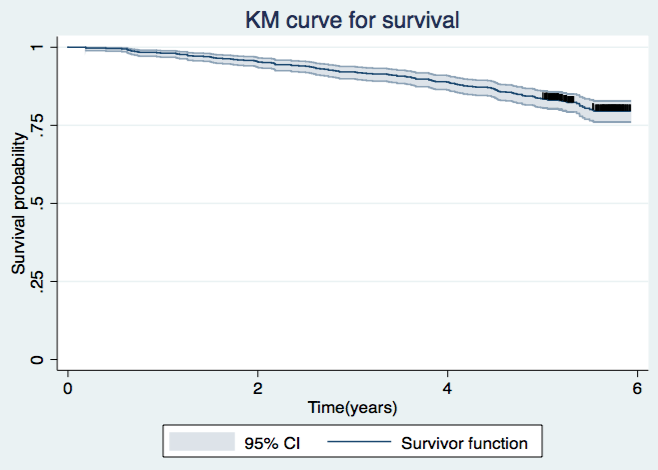
Question 1:

Some assumptions made with censored data is that is must be non-informative and a random sample of those at risk at the time of censoring.

I evaluated the Kaplan-Meier curve to determine if there were any censored data anywhere on the x-axis (time). Figure 1 illustrates censored data reported at approximately 5 years. At 5.002 years, two censored data were first identified.

It would be reasonable to evaluate mortality at 5 years after study enrollment since the first censored value was just after 5 years (5.002 years). Measuring mortality at 5 years is valid because we can’t be entirely sure why the patient was censored. It could be that the patient dropped out and lived a long life; conversely, the patient may have died, but was lost to follow-up. In total, there were 602 censored values.

Figure 1. Kaplan-Meier survival curve of deaths for the entire study period, 11 years.



Question 2:

The parameter of interest is LDL level which was categorized as: LDL < 100 and LDL >/= 100 mg/dL. The Outcome variable is death (or mortality) at 5 years.

In the descriptive analysis, patients who had an LDL less than 100 mg/dL had fewer deaths at 5 years compared to patients with LDL greater than or equal to 100 mg/dL (33 versus 86). Table 1 summarizes the number of deaths at 5 years and the baseline characteristics across all LDL groups.

A larger proportion of males died at 5 years in the LDL < 100 mg/dL group compred to the LDL >/= 100 mg/dL group (72.73% versus 61.63%). Among those patients with an LDL < 100 mg/dL and died at 5 years, there was a greater proportion with CHF compared to those patients with an LDL >/= 100 and died at 5 years (21.21% versus 11.63%).

Table 1. Descriptive analysis for patients with LDL <100 or >/= 100 mg/dL and mortality at 5 years.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | LDL < 100 mg/dL (N=165) | | LDL >/= 100 mg/dL (N=560) | |
|  | Death | No Death | Death | No Death |
| N | 33 | 132 | 86 | 474 |
| Age (mean, SD) | 75.79 (5.26) | 74.52 (5.55) | 76.88 (6.48) | 74.08 (5.11) |
| Weight, pounds | 162.05 (31.39) | 159.35 (31.60) | 158.35 (33.66) | 160.12 (30.03) |
| Sex (number, %\*) |  |  |  |  |
| Male | 24 (72.73) | 71 (53.79) | 53 (61.63) | 212 (44.73) |
| Female | 9 (27.27) | 61 (46.21) | 33 (38.37) | 262 (55.27) |
| Smoking history, pack years (mean, SD) | 24.62 (27.45) | 15.77 (22.89) | 29.26 (39.09) | 18.59 (25.20) |
| Smoking history categories (number, %) |  |  |  |  |
| 0 pack years | 13 (35.00) | 68 (51.52) | 32 (37.21) | 213 (44.94) |
| 1 or more pack years | 20 (65.00) | 64 (48.48) | 54 (62.79) | 261 (55.06) |
| Coronary heart disease (number, %) |  |  |  |  |
| No | 21 (63.64) | 114 (86.36) | 53 (61.63) | 386 (81.43) |
| Angina | 5 (15.15) | 9 (6.82) | 12 (13.95) | 36 (7.59) |
| Myocardial infarction | 7 (21.21) | 9 (6.82) | 21 (24.42) | 52 (10.97) |
| Congestive heart failure (number, %) |  |  |  |  |
| Yes | 7 (21.21) | 8 (6.06) | 10 (11.63) | 15 (3.16) |
| No | 26 (78.79) | 124 (93.94) | 76 (88.37) | 459 (96.84) |
| Stroke (number, %) |  |  |  |  |
| No | 25 (75.76) | 117 (88.64) | 60 (69.77) | 426 (89.87) |
| Ischemic attack | 2 (6.06) | 6 (4.55) | 5 (5.81) | 11 (2.32) |
| Stroke | 6 (18.18) | 9 (6.82) | 21 (24.42) | 37 (7.81) |

N, number

SD, standard deviation

\* Column percentages

4/4 for general table layout

2/3 for the choice of descriptive statistics

1/3 for discussion of finding

Did not write down the unit of each variable (-1)

Did not mention for potential confounding (-1)

Did not mention about missing data (-1)

Total: 7/10

Question 3:

We assume that the data is normally distributed. This is valid due to the large sample size. In order to compare the mean difference in LDL for those that experienced mortality at 5 years and those that do not, the Student’s t-test was used. Equal variances were not assumed. We performed a two-tailed test because we are not certain which group will have more or less deaths. Statistical significance was defined as P<0.05.

The following hypotheses were tested using the Student’s t-test, two-tailed:

H0: uA = uB, where uA denotes the average LDL for those who experienced mortality at 5 years and uB denotes the average LDL for those who did not experience mortality at 5 years.

Ha: uA != uB

Table 2. Statistical comparison of average LDL for patients who did or did not experience mortality at 5 years.

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  |  | LDL cholesterol level (mg/dL) | | | |
| Death at 5 years | N | Mean | SD | 95% CI | P-value |
| Yes | 119 | 118.70 | 36.16 | 112.13, 125.26 | 0.0186 |
| No | 606 | 127.20 | 32.93 | 124.57, 129.83 |

SD, standard deviation

CI, confidence interval

The difference in mean LDL between patients who died at 5 years and did not die was -8.50 (95% CI: -15.56, -1.44) mg/dL. A 95% CI suggests that this observation is no unusual if the true difference in mean LDL between patients who died and did not die were between -15.56 and 1.44 mg/dL. Because the P=0.0186, we reject the null hypothesis that there is no difference in mean LDL between patient who died and did not die at 5 years.

Question 4:

The predictor of interest is the geometric mean LDL and the outcome is mortality at 5 years. We performed a two-tailed Student t-test, and statistical significant was defined as P<0.05. Equal variances were not assumed. The hypotheses are:

H0: gmeanA = gmeanB, where gmeanA denotes the geometric mean of patients who died at 5 years and gmeanB denotes patients who did not die at 5 years.

Ha: gmeanA != gmeanB

To compare the differences in geometric mean LDL between patients who died or did not die at 5 years, we need to log-transform LDL.

We perform the t-test (equal variances not assumed) on the log(LDL) and exponentiated the mean and confidence intervals.

We get the following results:

For the group where patients died at 5 years, the geometric mean LDL was 112.01 mg/dL (95% CI: 104.54, 120.02); for the group where patients did not die at 5 years, the geometric mean LDL mg/dL was 122.83 (95% CI: 120.21, 125.50). Hence, the geometric mean LDL for patients who died at 5 years was 8.80% lower (95% CI: 3.43%, 13.89%) than for patients who did not die at 5 years (P=0.0016).

A 95% CI suggest that this observation is not unusual if the reduction in geometric mean LDL for those who died compared to those who did not die at 5 years were between 3.43% and 13.89%. Because the P=0.0016, we reject the null hypothesis that there is no difference in geometric mean LDL between patients who died and did not die at 5 years.

5/5 for performing an appropriate analysis

4/5 for reporting the association appropriately

Wrong interpretation of CI (-1)

Total: 9/10

Question 5:

We performed the chi-square test to evaluate the association between serum LDL (<160 or >/= 160 mg/dL) and 5-year all-cause mortality. The hypotheses are:

H0: uA – uB = 0, where uA is the risk of patients with LDL >/= 160 mg/dL who died at 5 years and uB is the risk of patient with LDL < 60 mg/dL who died at 5years. The null hypothesis is that the risk difference is 0.

Ha: uA – uB != 0

Among patients with LDL less than 160 mg/dL, there were 105 (16.99% out of 618) total deaths at 5 years. Among patients with LDL greater than or equal to 160 mg/dL, there were 14 (13.08% out of 107) total deaths at 5 years. The risk difference between those with LDL </= 160 versus LDL < 160 was 0.039 (95% CI: -0.031, 0.109). The P-value was 0.3139 (chi-square=1.01, df=1). Because the P-value is greater than .05, we do not have enough evidence to reject the null hypothesis that there is no difference in the number of deaths at 5 years across LDL groups.

Question 6:

We calculated an odds ratio to evaluate the association between deaths at 5 years across LDL groups (<160 and >/= 160 mg/dL). The hypotheses are:

H0: uA/uB = 1, where uA denotes the odds of death at 5 years for a patient with an LDL >/= 160 mg/dL and uB denotes the odds of death at 5 years for a patient with an LDL < 160 mg/dL.

Ha: uA/uB != 1

Patients with an LDL >/= 160 mg/dL had a lower odd or experiencing death at 5 years compared to patients with an LDL < 160 mg/dL (OR=0.74; 95% CI: 0.37, 1.36). The 95% CI suggests that this observation is not unusual if the true odds ratio for a patient with an LDL >/= 160 mg/dL being dead at 5 years compared to a patient with an LDL < 160 mg/dL being dead at 5 years were between 0.37 and 1.36. Because the P-value was 0.3139, we do not have enough evidence to reject the null hypothesis that there is no difference in odds of death at 5 years across the LDL groups.

Question 7:

We will use Cox Proportional Hazard model to answer this question. The hypotheses are:

H0: uA/uB = 1, where uA is the hazard of death for patients with LDL >/= 160 mg/dL and uB is the hazard of death for patients with LDL < 160 mg/dL. We assumed that the proportionality requirement holds for these curves. The hypotheses are:

Ha: uA/uB != 1

The HR = 0.72 (95% CI: 0.42, 1.23), P-value=0.2076.

Patients who had LDL >/= 160 mg/dL had a lower hazard of death compared to a patient with LDL < 160 mg/dL (HR=0.72; 95% CI: 0.42, 1.23). A 95% CI suggests that this observation is not unusual if the hazard ratio were between 0.42 and 1.23. Because the P-value is 0.2076, we do not have enough evidence to reject the null that there is no difference in hazard in death across the LDL groups.

5/5 for performing an appropriate analysis

4/5 for reporting the association appropriately

Wrong interpretation of CI(-1)

Total: 9/10

Question 8:

I would have preferred to use the Cox Proportional Hazard model to estimate the association between LDL and mortality. The outcome data is censored which would require us to use Kaplan-Meier plots to describe the summary data. In addition, the outcome is mortality which is dependent on time. Any other method would not have taken time into account. The Cox Proportional Hazard model could take time into account. The other benefit of the Cox Proportional Hazard model is that it can also control for potential confounders. We can also evaluate if there are effect modifiers because the Cox Proportional Hazard model is essentially a regression model.

Choose appropriate analysis (4)

Performed analysis that are vaild (2)

6/10