Homework 4

1. We model this problem using a proportional hazards regression model.
2. We use a proportional hazards regression model of the censored time to death on continuous measure of variable LDL so that we can allow each serum LDL level to have a distinct instantaneous risk of death. The problem did not specify how we should model LDL, so we use the untransformed values. The problem also did not specify whether we should use classical proportional hazards regression or proportional hazards regression with robust standard error estimates, so we use the latter since there is no reason to assume equal variance among two groups of subjects with different LDL levels. Taking into account the clinical significance of LDL levels, we consider levels of 10 mg/dL rather than 1 mg/dL.

|  |  |
| --- | --- |
| **Proportional Hazards Regression with Robust SE** |  |
| **t** | **Haz. Ratio** | **Std. Err.** | **z** | **P>|z|** | **[95% Conf.** | **Interval]** |
| **LDL** | 0.993 | 0.003 | -2.600 | 0.009 | 0.987 | 0.998 |

Comparing two groups with different LDL levels, the instantaneous risk of death is estimated to be 7.14% lower (hazards ratio 0.9286) for each 10 mg/dL difference in LDL level, with the group having the higher level of LDL tending toward a lower instantaneous risk of death. This observed difference is statistically different from an hazard ratio of 1 (p-value = 0.009), with a 95% confidence interval suggesting that the observed hazard ratio is what we might typically expect if the true instantaneous risk of dying was between 12.18% lower and 1.80% lower for each 10 mg/dL higher LDL level. Based on these results, we can reject the null hypothesis of no association between survival time and LDL at study entry in favor of a trend toward lower risk of death among subjects with higher LDL levels.



|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Summary of Survival Estimates** |  |  |  |  |
| **Time (mon.)** | **Beg. Total (subjects)** | **Total Fail (subjects)** | **Surv Fcn.** | **Std. Err** | **[95% Conf. Int.]** |
| **below 130 mg/dL**  |   |   |   |   |   |
| 12 | 387 | 7 | 0.9822 | 0.0067 | 0.963 | 0.9915 |
| 24 | 374 | 13 | 0.9491 | 0.0111 | 0.9222 | 0.9669 |
| 36 | 359 | 15 | 0.9109 | 0.0144 | 0.8782 | 0.9352 |
| 48 | 344 | 15 | 0.8728 | 0.0168 | 0.8356 | 0.902 |
| 60 | 318 | 26 | 0.8066 | 0.0199 | 0.764 | 0.8423 |
| **130 mg/dL - 160 mg/dL** |   |   |   |   |
| 12 | 221 | 5 | 0.9778 | 0.0098 | 0.9474 | 0.9907 |
| 24 | 216 | 5 | 0.9556 | 0.0137 | 0.919 | 0.9758 |
| 36 | 210 | 6 | 0.9289 | 0.0171 | 0.8865 | 0.9558 |
| 48 | 206 | 4 | 0.9111 | 0.019 | 0.8656 | 0.9417 |
| 60 | 197 | 9 | 0.8711 | 0.0223 | 0.8199 | 0.9086 |
| **over 160 mg/dL** |   |   |   |   |   |
| 12 | 0 | 0 | 1 |   |   |   |
| 24 | 106 | 2 | 0.9813 | 0.0131 | 0.9273 | 0.9953 |
| 36 | 103 | 3 | 0.9533 | 0.0204 | 0.8914 | 0.9803 |
| 48 | 98 | 5 | 0.9065 | 0.0281 | 0.8333 | 0.9486 |
| 60 | 94 | 4 | 0.8692 | 0.0326 | 0.7891 | 0.9203 |

Analysis of descriptive statistics: We categorize LDL into three classes: below 130 mg/dL, between 130 mg/dL and 160 mg/dL, and over 160 mg/dL and plot a Kaplan-Meier curve for the three groups. Since we are working with a censored data, Kaplan-Meier curve is a reliable descriptive statistic for our analysis of association between LDL and mortality rate. We also present a table with a summary of survival estimates by time (survival function), and analyzing both we note that survival rates tend to improve for higher levels of LDL.

1. We model this problem using a proportional hazards regression model on log-transformed data.
2. We use a proportional hazards regression model of the censored time to death on the log-transformed data of the continuous variable LDL so that we can allow each serum LDL level to have a distinct instantaneous risk of death. The problem did not specify whether we should use classical proportional hazards regression or proportional hazards regression with robust standard error estimates, so we use the latter since there is no reason to assume equal variance among two groups of subjects with different LDL levels. Taking into account the clinical significance of LDL levels, we consider levels of 10 mg/dL rather than 1 mg/dL.

|  |  |
| --- | --- |
| **Proportional Hazards Regression with Robust SE** |  |
| **t** | **Haz. Ratio** | **Std. Err.** | **z** | **P>|z|** | **[95% Conf.** | **Interval]** |
| **log(LDL)** | 0.438 | 0.087 | -4.170 | 0.000 | 0.297 | 0.645 |

Comparing two groups with different logarithmic LDL levels, we estimate that the instantaneous risk of dying is estimated to be 7.6% lower (hazard ratio 0.924) for each 10% difference in LDL level, with the group having higher level of LDL having lower instantaneous risk of death. This observed difference is statistically different from a hazards ratio of 1 (p-value of 0), with a 95% CI suggesting that the observed hazard ratio is what might be typically observed if the true instantaneous risk of death was anywhere between 10.9% and 4.1% lower for each 10% higher LDL level. We thus reject the null hypothesis of no association between time to death and LDL levels in favor of a trend toward lower risk of death for subjects with higher LDL levels.

For descriptive statistics, we can infer to the Kaplan-Meier curve from problem 1 since log-transformation of the data does not change our inferences regarding hazard ratios. This is due to the one-to-one relationship between our data and its log-transformed values.

1. We model this problem using a proportional hazards regression model.
2. We generate a variable ldlsqr which consists of the squared values of LDL, and use a proportional hazards regression model of the censored time to death on the continuous variable LDL. The problem did not specify whether we should use classical proportional hazards regression or proportional hazards regression with robust standard error estimates, so we use the latter since there is no reason to assume equal variance among two groups of subjects with different LDL levels.

|  |  |
| --- | --- |
| **Proportional Hazards Regression with Robust Standard Error** |  |
| **t** | **Haz. Ratio** | **Std. Err.** | **z** | **P>|z|** | **[95% Conf.** | **Interval]** |
| **LDL** | 0.97423 | 0.00953 | -2.67000 | 0.00800 | 0.95573 | 0.99309 |
| **LDL^2** | 1.00008 | 0.00004 | 1.92000 | 0.05500 | 1.00000 | 1.00015 |

Looking at the p-value of 0.055 for the squared LDL term, we find a statistically insignificant association at the 0.05 level (although by a thin margin) so we can conclude that there is lack of evidence of nonlinear association between LDL and mortality rate. To determine what this quadratic function predicts, we compute the estimated hazard ratios for each group relative to a LDL level of 0 mg/dL (although the meaning is clinically unimportant). Consistent to previous results, we predict higher levels of hazard for lower levels of LDL.

If we log-transform our data, our inferences are exactly the same regarding hazard ratios, so we would be able to draw the same inferences. This is because there is a 1 to 1 relationship between the data and its log-transformed values. It is also hard to define a clear linear relationship between LDL and mortality rate since the lack of a quadratic relationship does not imply the existence of a linear relationship.

1. Plotting the fitted hazard ratios from problems 1-3 versus LDL levels, we notice that all three models are consistent in that they tend to predict lower hazard ratios for higher measures of LDL. At the reference point of 160 mg/dL, all three fitted hazard ratios are equal to 1. For measures of LDL lower than 160 mg/dL, the hazard ratios tend to increase and for measures of LDL greater than 160 mg/dL, the hazard ratios tend to decrease. It is also notable that the quadratic fit tends to result in larger predicted centered hazard ratios (linear fit results in smallest fitted values). In any case, we can infer that all three fitted values of hazard ratios are consistent with the fact that higher LDL levels tend to result in lower hazards.