1. (a)

**Methods**

The survival distribution was estimated using the Kaplan-Meier estimates for various groups defined by serum LDL modeled as a continuous variable. To address the association between serum LDL and all-cause mortality, I consider three LDL strata’s; LDL less than or equal to 129 mg/dl, LDL within (130,160) mg/dl, and LDL greater than or equal to 160 mg/dl. The difference in survival distributions between these three groups is tested using Wald test. Cox proportional hazards regression with Huber-White sandwich estimator of the standard errors is used to compute the hazard ratio and the 95% CI.

**Results**

Of the 725 subjects with available measurements, 393 had serum LDL measurements less than or equal to 129 mg/dl, 225 had measurements between 130 mg/dl and 159 mg/dl inclusive, and 107 had measurements greater than or equal to 160 mg/dl. The following table and graph depicts Kaplan-Meier estimates of survival probability for the subjects under each of the three groups. When comparing two groups; 1 vs 2 and 2 vs 3, the instantaneous risk of death is estimated to be 23.68% lower for the high LDL group compared to the low LDL group. Based on a 95 % confidence interval, this observed hazard ratio of 0.7632 for the comparison of LDL groups 1-2 and 2-3, would not be judged unusual if the true hazard ratio were anywhere between .5895 to .9881. A two sided p-value of 0.040 suggests that we reject the null hypothesis of no association between serum LDL and survival time in favor of increased survival experience for subjects with high LDL.

|  |  |  |  |
| --- | --- | --- | --- |
|  | Survival probabilities (Kaplan-Meier) | | |
|  | LDL129 mg/dl | 130 mg/dlLDL159 mg/dl | LDL160 mg/dl |
| 1 year | 0.982 | 0.978 | 1.000 |
| 2 years | 0.949 | 0.956 | 0.981 |
| 3 years | 0.911 | 0.929 | 0.953 |
| 4 years | 0.873 | 0.911 | 0.907 |
| 5 years | 0.807 | 0.871 | 0.869 |



1. (a)

The survival distribution was estimated using the Kaplan-Meier estimates for various groups defined by serum LDL modeled as a log of continuous variable LDL. To address the association between serum log (LDL) and all-cause mortality, I consider three LDL strata’s; log(LDL) less than or equal to log(129) mg/dl, log(LDL) within (log(130), log(160)) mg/dl, and log(LDL) greater than or equal to log(160) mg/dl. The difference in survival distributions between these three groups is tested using Wald test. Cox proportional hazards regression with Huber-White sandwich estimator of the standard errors is used to compute the hazard ratio and the 95% CI.

The descriptive statistics are similar to those in 1 (a). In addition, by looking at the histogram of log (LDL), one of the observed LDL values (11mg/dl) appears to be an outlier.



Of the 725 subjects with available measurements, 393 had serum LDL measurements less than or equal to log (129) mg/dl, 225 had measurements between log (130) mg/dl and log (159) mg/dl inclusive, and 107 had measurements greater than or equal to log (160) mg/dl.

When comparing two groups with different LDL levels, the instantaneous risk of dying is estimated to be 43.62 % lower (hazard ratio 0.5638) for each two fold increase in LDL level, with the group having the higher level of LDL tending toward a lower instantaneous risk of death. This estimate is highly statistically significant (p<0.0001). A 95% confidence suggests that this observation is not unusual if a group that has LDL twice as high as another might have risk of death anywhere from 56.93% lower to 26.19% lower as group with the lower LDL. Thus we reject the null hypothesis of no association between LDL and risk of death in favor of reduced risk of death for subjects with high LDL.

5/5 for performing an appropriate analysis

4.5/5 for reporting the association appropriately

Did not report whether the p-value is two-sided or one-sided(-0.5)

Total: 9.5

1. To evaluate the association between serum LDL and all-cause mortality in this regression analysis, we consider both linear and quadratic modelling of LDL. Using Cox regression, we fit a model for both LDL and square (LDL). The descriptive statistics are similar to 1 a.

Although the p-value for the nonlinear term (HR 1.000) is non-significant at alpha =0.05 (two sided p-value=0.055) and a 95% CI (1.000, 1.0002), we can’t conclude linearity since it could have been nonlinear in a way that the quadratic polynomial could not detect.

The LDL linear term (HR .9983), is significant (two sided p-value=0.008) and a 95% CI (.9906, 1.006) though we can’t conclude linearity since the linear term is directly correlated with the quadratic term.

1. The graph below represents the predicted hazard ratios for the linear, log, and quadratic LDL levels.



For LDL levels less than 100 mg/dl and above 200mg/dl, the quadratic LDL transform demonstrates higher relative risk compared to both log transformed and the linear terms of LDL. For LDL values between 100 to 170 mg/dl (approximately), there is not a significant difference in relative risk between the three hazards. The minimum relative risk appears to occur at around 150 mg/dl LDL.