**#1279**

**HW01**:

**Q1**: The observation of time to death (histogram shown below) is right censored has a mean of 1804 days with a standard deviation of 392 days, minimum of 68 days to the first death occurring, and is right censored with a mean follow-up of ~11 years per the dataset description and max follow-up time of 2159 days. We can dichotomize this variable as deaths occurring before or after 5 years, or survival to 5 years (5 years ~1830 days), changing it to a binary variable to simplify statistical analysis. Also the restricted mean of the variable observation time is very close to 5 yrs (mean of obstime 1804 days, 5 yrs = 1830 days)

Points: 5



**Q2**:

Table 1: Baseline Characteristics of cases by LDL level

|  |  |  |  |
| --- | --- | --- | --- |
| **Characteristic** | **Normal LDL**(LDL <160 mg/dL) | **High LDL**(LDL>=160 mg/dL) | ***p-value*** |
| Age  | 74.5 (5.4) | 74.8 (5.7) | *0.57* |
| Sex (Male %) | 51% | 44% | *0.14* |
| Weight in lbs | 159 (30.7) | 163 (30) | *0.22* |
| Smoking |  |  |  |
| Pack years | 19.8 (27) | 18 (24) | *0.47* |
| Years since quit | 9.6 (14) | 9.7 (13.8) | *0.97* |
| CHD (%) | 33% | 34% | *0.90* |
| CHF (%) | 6% | 3% | *0.18* |
| Stroke (%) | 22% | 32% | *0.13* |
| 5 Year Survival(%) | 80% | 85% | *0.19* |

Points: 5

Table: Age: lack of unit. Lack of the following items: sample size, missing data for the continuous variables. No need of p-value -2

No discussion: -3

**Q3**:

To test an association between serum LDL means and 5 year survival, assuming a normal distribution of LDL (histogram of LDL below), we can perform a one sided ttest that allows for the possibility of unequal variance with the null hypothesis being that mean LDL would be lower in the group with death within 5 years, and the alternate hypothesis that mean LDL is lower in the group without death within 5 years. 95% confidence intervals for the difference in population means were similarly based on that same handling of variances.

Results: mean LDL in the group with the 606 subjects that survived to 5 years was 127.2 mg/dL (95% CI 124.517, 129.894) and the mean LDL in the group that did not survive to 5 years was 119.9 mg/dL (95% CI 114.13, 125.8559) with a resultant p value of 0.98, thus we fail to reject the null hypothesis.

Based on a 95% confidence interval computed with an allowance for unequal variances, this observed tendency of 8.50 mg/dL lower mean serum LDL among subjects dying earlier would not be judged unusual if the true difference population means were anywhere between a 1.44 mg/dL to 15.6 mg/dL lower mean LDL among subjects who die within 5 years. Using a t test that similarly allows for the

possibility of unequal variances, this observation is statistically significant at a 0.05 level of significance (two-sided P= 0.0186), and we can with high confidence reject the null hypothesis that the mean serum LDL levels are not different by vital status at 5 years in favor of a hypothesis that death within 5 years is associated with lower mean serum LDL.

We deduce from this that we do not have evidence to suggest that mean LDL values are higher in the group with death within 5 years.

Points: 4+2=6

Histogram of LDL values

**Q4**:

To test if there is an association between serum LDL geometric means and 5 year survival, we would first need to generate the log of the LDL values, then we perform the ttest that allows for the possibility of unequal variance to compared the differences in the mean of log transformed serum LDL levels

, then exponentiate all the estimates and CI.

Result: the geometric mean of LDL in the group with death within the first 5 years was 113.7 (95% CI 107-120.8), the geometric mean of LDL values in the group without death in the first 5 years was 122.8 (95% CI 120.1-125.5), the risk difference is 1.07 (95% CI 1.15-1.012) and the p value is 0.02 hence there is sufficient evidence to suggest that the geometric mean of LDL values in the group with death within 5 years is different than the geometric mean of LDL in the group without death within 5 years. And we note that the value is lower in the group with death within 5 years.

Points: 4+2=6 (lack of the size of the groups, didn’t mention the differences in the means of two groups, 95% CI, p value)

**Q5**:

We will compare the proportion of subjects dying within 5 years of study enrollment across two groups with high LDL levels (>=160 mg/dL) and normal (<160 mg/dL) Differences in the probability of death within 5 years were tested using Pearson’s chi squared test for independence. 95% confidence intervals for

the difference in population 5 year mortality probabilities were computed using Wald statistics.

 The resultant risk difference between the two groups is 0.04 (95% CI -0.021 to 0.103) with a p value of 0.225 hence we can infer that there was no difference in 5 year survival when we group the cases by LDL value >=160mg/dL or <160 mg/dL.

Points: 3+2=5 (lack of the size of the groups, didn’t mention proportion estimate of two groups, p value)

**Q6**:

We will compare the odds of survival beyond 5 years between groups based on LDL values of high ie >=160mg/dL or normal ie <160 mg/dL by performing a chi square test with odds ratio. 95% confidence intervals for the odds ratio was also computed using exact methods.

The odds of survival in the high LDL group was 1.39 times higher than the normal LDL group (95% CI 0.8 to 2.5) with p value of 0.22 suggesting no statistically significant increase in risk of death between both groups.

Points: 4+1=5 (lack of the size of the groups, didn’t mention odds/probability estimate of two groups, wrong 95% CI, and p value)

**Q7**:

To evaluate an association between serum LDL and all cause mortality over the entire period of the observation (two continuous variables) we would perform a pearson correlation which is 0.275 suggesting that there isn’t significant correlation between serum LDL and death

Points: 0 (See the key; wrong test and lack of the details)

**Q8**:

A ttest would have been sufficient to answer my specific scientific question of 5 year survival and LDL values, or if I wwere to dichotomize LDL values by high/normal, then a chi square test.

Points: 2

Total points: 34