Score: 39/75 . Overall: Far more detail is needed to get credit for the homework, especially in the methods and results sections. I find it useful to approach each problem as if you are preparing an analysis for a manuscript. Make sure you provide the reader with appropriate tables to go with the text and that you take the time to describe the table. Also, don’t forget the methods and results section as explained on the course website. It is helpful for the peer grader if you clearly mark these sections.

(date of MRI) or death after five years after enrollment because you do not lose any data in doing so. Specifically, all participants who were still alive at the end of the study (n=602) What percentage of participants is this? had been enrolled for > 5 years, so they are all in the death after 5 year group; **there were** **0 study participants** who were alive at the end of the study and had been **enrolled less than 5 years**. Further, there were 12 participants who died during the study after 5 years of survival, and 121 who had died within 5 years.

Score 4/5. Your wording of this answer is a little confusing. It is possible because all the participants lived for at least 5 years. Be sure to provide percentages along with raw numbers. This is the only way someone knows the magnitude of what you’re trying to explain.

2.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Participant surviving > 5 years post MRI (n=614) | | | | |
|  | LDL | Age (years) | Weight | Years smoking |
| Mean | 127.20 | 74.19 | 160.11 | 17.95 |
| Standard Deviation | 32.93 | 5.22 | 30.35 | 24.69 |
| Min | 39 | 65 | 74 | 0 |
| 25th %tile | 103 | 71 | 138.5 | 0 |
| 50th %tile | 127 | 73 | 158.75 | 4.35 |
| 75th %tile | 148 | 77 | 180 | 31.88 |
| Max | 247 | 99 | 258 | 180 |
|  | Proportion who have had Coronary heart disease | Proportion who have had Coronary Heart Failure | Proportion who have had a stroke | Proportion who are male |
|  | 0.28 | 0.04 | 0.18 | 0.47 |

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Participants surviving < 5 years post MRI (n=121) | | | | |
|  | LDL | Age (years?) | Weight | Years smoking |
| Mean | 118.70 | 76.48 | 159.12 | 28.05 |
| Standard Deviation | 36.16 | 6.17 | 32.79 | 36.04 |
| Min | 11 | 67 | 96 | 0 |
| 25th %tile | 96 | 72 | 139 | 0 |
| 50th %tile | 117 | 75 | 154 | 18.38 |
| 75th %tile | 142 | 81 | 176 | 46 |
| Max | 227 | 91 | 264 | 240 |
|  | Proportion who have had Coronary heart disease | Proportion who have had Coronary Heart Failure | Proportion who have had a stroke | Proportion who are male |
|  | 0.62 | 0.14 | 0.52 | 0.64 |

Score:

5/10

3/4 table format

2/3 choice of descriptive statistics

0/3 discussion of findings

Remember to state units. Age can be months, years, days, etc. The reader won’t know unless you tell them. Similarly with serologic samples. There may have been a better way to format the table. As written the Proportions of those with CHD, CHF, Stroke, etc looks as though they fall under the columns for the other predictors. Review scientific papers for a good way to format tables. For this problem, focus on the 1st or 2nd table in most papers. They provide great examples.

Also, never present a table without summarizing the contents in text.

3.

A two sample two-sided T test assuming unequal variances was performed on the mean serum LDL level for each group (defined by their 5 year vital status) to examine if the mean LDL level was different between the groups. So are the groups survived >5 years, did not survive > 5 years. The means were found to be statistically different at the 0.05 significance level (p-value=0.0186), with the group surviving greater than five years having an 8.5005 unit (actually state the units of LDL) higher mean LDL level (127.20 vs. 118.70), and the 95% confidence interval for the difference in means is (1.44, 15.56). This means that these results would be unsurprising if the true difference in means between these groups was between 1.44 and 15.56 mg/dL higher in the group that survived longer than five years.

Score:

8/10

4/5 analysis and description

4/5 Interpreting results

Unequal variance assumption is actually just not assuming the variances are equal, allowing for the possibility that the variances are unequal.

Once again state the actual units of LDL (mg/dL) vs. stating the “unit” changes.

4.

A two sample two-sided T test was performed on the geometric mean serum LDL level for each group (defined by their 5 year vital status) to examine if the geometric mean LDL level was different between the groups. The geometric means were found to be statistically different (same as above. State significance level) (p-value=0.0013), with the group surviving greater than five years having an 10.8 unit higher mean LDL level (122.82 vs. 112.01), and the 95% confidence interval for the difference in means is (4.22, 17.40). This means that these results would be unsurprising if the true difference in means between these groups was between 4.22 and 17.40 mg/dL higher in the group that survived longer than five years.

Score:

6/10

4/5 analysis and description

2/5 interpreting results

Comments same as for number 3. State significance level and provide units. State explicitly at least the null and what outputs you will evaluate to come to a conclusion. Because we log transformed the LDL variable when performing the 2-sample t-test we describe results in terms of ratios vs. differences.

5.

A Chi squared test was performed to examine the association between high LDL and probability of death within five years post MRI. The test failed to show that there was a statistically different probability of death amongst those with vs. those without high LDL (>=160 mg/dL) (p-value = 0.375) based on the observed data.

The group with high LDL had a .860 probability of death within five years, while the group without high LDL has a .832 probability of death within five years. The difference in probability of death was .028 with a 95% confidence interval of the difference in probabilities being (-.034, .099). This means that these results would be unsurprising if the true difference in probabilities of death between these groups was between -.034 and .099 higher in the high LDL group.

Score:

5/10

3/5 analysis and description

2/5 interpreting results

The methods sections should state exactly which groups are being compared. Ttest results are inaccurate.

6. A Chi squared test was performed to examine the association between high LDL and probability of death within five years post MRI. The test failed to show that there was a statistically different probability of death amongst those with vs. those without high LDL (>=160 mg/dL) (p-value = 0.375) based on the observed data. This is merely repeated from question 5. The odds ratio of the odds of death of the group with high LDL:the group without high LDL was 1.292, with a 95% confidence interval of this value being (.737, 2.265). While the high LDL group had a greater odds of dying before 5 years, there was not a statistically significant difference. This means that these results would be unsurprising if the true odds ratio was between .737 and 2.265.

Score:

5/10

3/5 analysis and description

2/5 interpreting results

Be more descriptive in your methods. Your responses are inaccurate. Refer to key for correct answer. Could’ve been an issue coding when creating high and low LDL groups.

7. A logrank test was performed to evaluate the association between LDL and mortality over the observed time period and by examining the observation time of death (days post MRI) and classification of high LDL (>=160 mg/dL). The p value of this test was 0.0435, indicating that high LDL is associated with instantaneous rate of death in the sampled population.

Score:

2/10

2/5 analysis and description

0/5 interpreting results

Provide more detail in terms of groups compared and analysis performed. Log rank test compares the distributions of the two groups. Remember you are not providing a graph or a table so you must explicitly state what the findings are and what conclusions you came to as a results.

8.

I would have performed a simple two sample, two-sided T-test to examine if the two groups (defined by their vital status at 5 years) had a statistically significant different mean LDL level. I would have doe this because it is a straightforward and easy way to test between our outcome (death within five years) and predictor of interest (LDL level). While we are examining by the outcome, this is a useful test because we defined it such to not have any censored data.

Score:

4/10

I interpreted your statement of the ttest being straightforward as it was a test you were comfortable with, knew how to do, and that was easily interpretable. You did not however hit on any of the other major points.