Biostats 518 – Homework 1

1. The nature of the dataset is such that those that died during the study have their death variable labeled 1. Those who did not die have their death variable labeled 0. The obstime variable counts the number of days between the first mri and time of death. Looking at the data, everyone with an obstime value of less than 365\*5 days has a death value of 1. And very few of the people with obstime values greater than 365\*5 days have a death value of 1.

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **Type** | **Min** | **1st Q** | **Median** | **Mean** | **3rd Q** | **Max** |
| **Obstime >= 365\*5** | 0.00 | 0.00 | 0.00 | 0.02 | 0.00 | 1.00 |
| **Obstime < 365\*5** | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 |

1. It looks like those who lived at least 5 years or more have a different profile than those who did not survive at least 5 years. Those who died within 5 years also seem to have a stronger smoking history (higher pack-years value) as well as much higher incidences of heart problems and stroke. There also seem to be slightly more males and slightly lower in weight (although these differences are negligible).
   1. **Methods**: Basic descriptive statistics. Two separate tables are presented for those who have survival times “greater than or equal to 5 years” and “less than 5 years.”
   2. Greater than 5 years

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **Type** | **Min** | **1st Q** | **Median** | **Mean** | **3rd Q** | **Max** |
| **Age** | 65.00 | 71.00 | 73.00 | 74.19 | 77.00 | 99.00 |
| **Weight** | 74.00 | 138.50 | 158.80 | 160.10 | 180.00 | 258.00 |
| **Male** | 0.00 | 0.00 | 0.00 | 0.46 | 1.00 | 1.00 |
| **Smoking History (pkyrs)** | 0.00 | 0.00 | 4.35 | 17.95 | 31.79 | 180.00 |
| **Coronary Heart Disease** | 0.00 | 0.00 | 0.00 | 0.27 | 0.00 | 2.00 |
| **Congestive Heart Failure** | 0.00 | 0.00 | 0.00 | 0.03 | 0.00 | 1.00 |
| **Stroke** | 0.00 | 0.00 | 0.00 | 0.18 | 0.00 | 0.00 |

* 1. Less than 5 years

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **Type** | **Min** | **1st Q** | **Median** | **Mean** | **3rd Q** | **Max** |
| **Age** | 67.00 | 72.00 | 75.00 | 76.48 | 81.00 | 91.00 |
| **Weight** | 96.00 | 139.00 | 154.00 | 159.10 | 176.00 | 264.00 |
| **Male** | 0.00 | 0.00 | 1.00 | 0.64 | 1.00 | 1.00 |
| **Smoking History (pkyrs)** | 0.00 | 0.00 | 18.38 | 28.05 | 46.00 | 240.00 |
| **Coronary Heart Disease** | 0.00 | 0.00 | 0.00 | 0.61 | 1.00 | 2.00 |
| **Congestive Heart Failure** | 0.00 | 0.00 | 0.00 | 0.14 | 0.00 | 1.00 |
| **Stroke** | 0.00 | 0.00 | 0.00 | 0.52 | 1.00 | 2.00 |

1. There seems to be a difference between the mean LDL values for the group that died within 5 years of the MRI and the group that did not die within 5 years of the MRI
   1. **Methods:** t-test

|  |  |
| --- | --- |
|  | **Value** |
| **Mean LDL < 5 years** | 118.70 |
| **Mean LDL > 5 years** | 127.20 |
| **T** | -2.38 |
| **Df** | 158.75 |
| **p-value** | 0.02 |

* 1. **Inference**: The difference between the means does not seem to equal zero.

1. It looks like the mean of the >5-year survival group dropped to 122 from 127 (comparing the geometric mean to the normal mean from the previous question). Age stayed the same and weight dropped a little bit (comparing geometric means to the means calculated in question 2).
   1. **Method:** geometric mean

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **<5 years survival** |  |  | **>5 years survival** | |
| **Type** | **Geometric Mean** |  | **Type** | **Geometric Mean** |
| **LDL** | 112.0114 |  | **LDL** | 122.8254 |
| **Age** | 76.23 |  | **Age** | 74.01 |
| **Weight** | 155.89 |  | **Weight** | 157.24 |
| **Male** | 0.00 |  | **Male** | 0.00 |
| **Smoking History (pkyrs)** | 0.00 |  | **Smoking History (pkyrs)** | 0.00 |
| **Coronoary Heart Disease** | 0.00 |  | **Coronoary Heart Disease** | 0.00 |
| **Congestive Heart Failure** | 0.00 |  | **Congestive Heart Failure** | 0.00 |
| **Stroke** | 0.00 |  | **Stroke** | 0.00 |

* 1. **Inference:** The >5-year survival group still seems to have a higher LDL cholesterol value compared to the <5-year survival group.

1. The >5 year survival group has a higher probability of having a “high” LDL value, but the number of people in that group is tiny and that could be throwing us off.
   1. **Method:** Basic Probability

|  |  |  |  |
| --- | --- | --- | --- |
|  | **Total** | **LDL > 160** | **Probability** |
| **>5 year** | 13.00 | 2.00 | 0.15 |
| **<5 year** | 735.00 | 103.00 | 0.14 |

* 1. **Inference:** The >5 year survival group seems to have a slightly higher probability at 15% compared to the <5 year survival group at a probability of 14%.

1. The odds ratio seems to show that the odds of having high LDL and dying during the study are lower than the other way around. The 95% confidence interval however, surpasses 1.0 for the high end of the interval; showing that it is still possible for high LDL to have a relationship with death during the study.
   1. **Method:** Odds ratio

|  |  |  |
| --- | --- | --- |
|  | **Died during study** | **Died after study** |
| **>5 years & high LDL** | 16 | 100 |
| **<5 years & high LDL** | 18 | 73 |
|  |  |  |
|  | **Values** |  |
| **Odds ratio** | 0.6489 |  |
| **95 % CI** | 0.3102 to 1.3572 |  |
| **z statistic** | 1.149 |  |
| **P** | 0.2507 |  |

* 1. **Inference:** The odds ratio is 0.64, which means that the odds of dying due to high LDL outside the course of the study is lower.

1. The risk ratio is not too different from the odds ratio we previously calculated; it’s a little higher. Odds ratio and relative risk are usually interchangeable if the disease at hand is rare enough, but it looks like the numbers are different for this example.
   1. **Method:** Calculate the risk ratio between groups

|  |  |  |
| --- | --- | --- |
|  | **Died during study** | **Died after study** |
| **>5 years & high LDL** | 16 | 100 |
| **<5 years & high LDL** | 18 | 73 |
|  |  |  |
|  | **Values** |  |
| **Relative risk** | 0.6973 |  |
| **95 % CI** | 0.3770 to 1.2898 |  |
| **z statistic** | 1.149 |  |
| **P** | 0.2506 |  |

* 1. **Inference**: It looks like there is a reduced risk of dying outside of the course of the study if one has a higher LDL. Specifically, 0.7X the risk.

1. I would have preferred a chi-squared test to see if the distribution of death is similar between individuals who fall into different LDL risk groups. Interestingly though, my point of view has changed now and I’d prefer a t-test to the chi-squared test. We learned in BIOSTATS 517 that it is inappropriate to use statistical tests to inform the use of other statistical tests however. In class, it was clarified that the intent of walking us through all of these redundant steps was to illustrate that there is more than one way to answer a question (with the implication that we are balancing the pros and cons of each approach).