Biostats 518 Homework #5

1. Question 1
   * 1. **Method:** Perform a logistic regression and then a Wald test to obtain a point estimate for odds ratio and then a 95% confidence interval for that odds ratio. I am also going to create a table of basic statistics:

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Race** | **mean** | **min** | **p50** | **max** | **sd** |
| **1 = White** | 0.0979021 | 0 | 0 | 1 | 0.2974423 |
| **2 = Black** | 0.1730769 | 0 | 0 | 1 | 0.3801458 |
| **3 = Asian** | 0.0638298 | 0 | 0 | 1 | 0.2470922 |
| **4 = Other** | 0.1666667 | 0 | 0 | 1 | 0.3892495 |
| **Total** | 0.107483 | 0 | 0 | 1 | 0.3099372 |

* + 1. **Inference:** The resulting equation from the logistic regression is E(diabetic state | race) = -2.3095 + 0.1434(race). The 95% CI for race is -.1639128 and .450804. The 95% CI for the intercept is -2.790673 and -1.82839. This means the odds of being diabetic is 2.1 times less when you are white before the MRI.
  1. The intercept means that people in general have are 2.3x less likely to be diagnosed with diabetes before the MRI.
     1. E(diabetic state | race) = -2.3095 + 0.1434(1) = -2.156
        1. Whites have a 2.1x lower odds of being diagnosed with diabetes before the MRI
     2. E(diabetic state | race) = -2.3095 + 0.1434(2) = -2.013
        1. Blacks have a 2x lower odds of being diagnosed with diabetes before the MRI
     3. E(diabetic state | race) = -2.3095 + 0.1434(3) = -1.869
        1. Asians have a 1.8x lower odds of being diagnosed with diabetes before the MRI
     4. E(diabetic state | race) = -2.3095 + 0.1434(4) = -1.726
        1. Those of “other races” have a 1.7x lower odds of being diagnosed with diabetes before the MRI
  2. We would conclude that the intercept value is significant at a alpha level of .05 and the value of race is not significant.
  3. The logistic model is actually the same. Whether blacks or whites are the reference group doesn’t make a difference. The point estimate for odds would change from 2.1x lower to 2x lower.
  4. We have the exact same model, so the results are identical to the answers from part (b)
  5. We would conclude that the intercept value is significant at a alpha level of .05 and the value of race is not significant.
  6. Realistically, we already know from other studies that race has a relationship with diabetes (some racial groups being more genetically predisposed than others). So, we know that if we only trust p-values, we can be mislead.

1. Question 2
   * 1. **Methods:** Create a table of descriptive statistics including a breakdown of LDL serum levels and mean, median, standard deviation, min, and max. Then I’m going to create a proportional hazard regression model to evaluate the odds of instantaneous death and use a Wald test to create a 95% CI.

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **ldlCTG** | **mean** | **p50** | **sd** | **min** | **max** |
| **0** | 0.4545455 | 0 | 0.5096472 | 0 | 1 |
| **70** | 0.1958042 | 0 | 0.3982133 | 0 | 1 |
| **100** | 0.1929825 | 0 | 0.3955077 | 0 | 1 |
| **130** | 0.1511111 | 0 | 0.3589557 | 0 | 1 |
| **160** | 0.1325301 | 0 | 0.3411274 | 0 | 1 |
| **190** | 0.1666667 | 0 | 0.3806935 | 0 | 1 |
| **Total** | 0.1806897 | 0 | 0.3850265 | 0 | 1 |

* + 1. **Inference:** The resulting cox regression equation is E(death | ldlCTG) = 1.0006(ldlCTG) + 2.89e18. The 95% CI for the ldlCTG coefficient is .9966 and 1.0046 and the 95% CI for the death coefficient is 0 and infinity. This equation is nonsensical. This indicates that the magnitudes likely differ between variabels in the model, throwing us off.
  1. The intercept is equal to 2.89e18, which basically means that everyone is alive. The value of the hazard ratio is equal to 1.0006, which basically means that the hazard of having high ldl serum levels increases linearly with risk. In the big picture, everything is overpowered by the intercept, so it doesn’t matter.
  2. I would compare the F-values or chi-square values between the models to evaluate how each one performed.
  3. E(death|ldlCTG) = 1.0006(ldlCTG) + 2.89e18
     1. E(death|ldlCTG) = 1.0006(0) + 2.89e18 = 2.89e18
     2. E(death|ldlCTG) = 1.0006(70) + 2.89e18 = 2.89e18
     3. E(death|ldlCTG) = 1.0006(100) + 2.89e18 = 2.89e18
     4. E(death|ldlCTG) = 1.0006(130) + 2.89e18 = 2.89e18
     5. E(death|ldlCTG) = 1.0006(160) + 2.89e18 = 2.89e18
     6. E(death|ldlCTG) = 1.0006(190) + 2.89e18 = 2.89e18
     7. E(death|ldlCTG) = 1.0006(250) + 2.89e18 = 2.89e18

1. Question 3
   * 1. **Methods:** We create a proportional hazard regression model to evaluate the odds of instantaneous death and use a Wald test to create a 95% CI. We will do this for LDL levels of 0, 70, 100, 130, 160, and 190.
     2. **Inference:** The hazard ratios seem to increase as LDL serum levels increase. This is basically saying that low and high LDL serum values are associated with higher risk of death. Those with an LDL level between 100 and 160 seem to be at lowest risk of death.
     3. 0 = 1.011128
     4. 70 = 1.003486
     5. 100 = .9996456
     6. 130 = .9960891
     7. 160 = 1.000603
     8. 190 = 1.03527
   1. Since we are dealing with hazard ratios, I would use the logrank test to compare all of the models to see which fits the data best.
      1. 0 = 1.011128\*(160) = 161.7804
      2. 70 = 1.003486\*(160) = 160.5577
      3. 100 = .9996456\*(160) = 159.9432
      4. 130 = .9960891\*(160) = 159.3742
      5. 160 = 1.000603\*(160) = 160.0964
      6. 190 = 1.03527\*(160) = 165.6432
2. Question 4
   1. Proportional hazard regression allows us to compare he hazard ratios of various groups while having right-censored data and avoiding Kaplan-meier calcuations. Linear regression provides us a way to compare mean differences, logistic regression allows us to compare odds of events, and poisson regression provides us a way to characterize relative risk of events.
   2. The models are essentially modeling the same data, but we are using surrogates in this model. We are still calculating the proportional hazard ratios between groups with respect to death.
   3. I would have chosen logistic regression a priori just because there are two outcomes, dead or alive. Now, I would go with proportional hazard regression since we get a more detailed breakdown of groups and their results.