BIOST536 Homework 3

Homework code: 3992

Due November 27 by 5 pm

1. Descriptive statistics:

|  |  |  |
| --- | --- | --- |
| **Table 1. Patient characteristics by nadir PSA category (normal or high)** | | |
| **Characteristics (means)** | **normal range (nadir psa< 4.0) n=33** | **high range (nadir psa >= 4.0) n=17** |
| **Pre-treatment PSA** | 527.4 | 938.4 |
| **Performance status** | 82.50 | 77.50 |
| **Bone scan score** | 2.375 | 2.813 |
| **Tumor grade** | 2.172 | 2.083 |
| **Age (years)** | 66.79 | 68.71 |
| **Time observed in remission (months)** | 36.27 | 13.29 |
| **Proportion relapsed within 24 months** | 0.2121 | 0.8824 |

We can see that there are some differences between the two groups: the proportion of the high-range nadir psa individuals that relapsed within 24 months does appear to be quite a lot higher than the proportion of normal range psa nadir individuals that relapsed in this time frame (and we can see that overall time in remission also appears to be quite a bit longer for individuals with normal psa). Other differences (aside from pre-treatment PSA) are less drastic and require more extensive analysis to determine if there are any real differences between the two groups.

**2a**. When comparing two groups with different nadir psa but similar bone scan score and performance status, the odds of relapse within 24 months is estimated to be 3.39% higher (odds ratio = 1.0339) for each 1 ng/ml increase in nadir psa. This observed difference is not statistically different from an odds ratio of 1 at the α=0.05 (p=0.476). The observed odds ratio would not be unusual if the true odds ratio were between 0.9433 and 1.133. We therefore cannot reject the null hypothesis of no association between relapse in 24 months and nadir psa after adjusting for bone scan score and performance status.

**2b**. When comparing two groups with different nadir psa but similar bone scan score and performance status, the odds of relapse within 24 months is estimated to be 8.54% higher for each 10% increase in nadir psa. This observed difference is statistically different from an odds ratio of 1 at the α=0.05 (p=0.007), with a 95% confidence interval suggesting that the observed odds ratio is what might be typically observed if the true odds of relapse within 24 months was between 2.29% and 15.17% higher for each 10% increase in nadir psa. We thus reject the null hypothesis of no association between relapse in 24 months and nadir psa after adjusting for bone scan score and performance status.

**2c**. When comparing two groups with different nadir psa but similar bone scan score and performance status, we observe that the relationship between nadir psa and relapse does not appear to be linear when examining a linear splines model with knots at 1, 4, and 16 ng/ml nadir psa. This observed departure from linearity is statistically significant at α=0.05 (p=0.0114) when we perform a Wald test of the splines simultaneously.

**2d**. Intercept interpretation:

For problem 2a (logistic regression on continuous, untransformed variable), the odds of relapse within 24 months for an individual with a nadir psa of zero, bone scan score of 1, and performance status of zero is 2.072.

For problem 2b (logistic regression on a log transformed variable), the odds of relapse within 24 months for an individual with a nadir psa of zero bone scan score of 1, and performance status of zero is 3.061.

For problem 2c (logistic regression using splines), the odds of relapse within 24 months for an individual with a nadir psa of zero, bone scan score of 1, and performance status of zero is 0.3809.

**3a**. Using a linear regression analysis, we observe that for individuals with similar bone scan scores and performance status, the mean nadir psa is 23.52 ng/ml higher in individuals who relapse within 24 months. This result is significantly different from zero at the α=0.05 (p=0.046), with a 95% confidence interval suggesting that these results would not be unusual if the true difference in nadir psa between individuals who relapse within 24 months and those who don’t is between 0.4765 and 46.56 ng/ml. We thus reject the null hypothesis that nadir psa does not differ between those who relapse and those who don’t while adjusting for bone scan score and performance status.

**3b**. From a linear regression analysis, for individuals with similar bone scan scores and performance status, we estimate that geometric mean nadir psa is 1365% ng/ml (on average) higher in individuals who relapse within 24 months and those who don’t. This result is statistically significant at the α=0.05 (p<0.0001), with a 95% confidence interval suggesting that these results would not be unusual if the true difference in geometric mean nadir psa is between 413.0% higher and 4516% higher for individuals who do relapse within 24 months. We thus reject the null hypothesis that geometric mean nadir psa does not differ between those who relapse and those who don’t while adjusting for bone scan score and performance status.

**4a**. Merits of each analysis:

1. Logistic regression with untransformed continuous variable: this is a very easy to interpret model
2. Logistic regression with a continuous, log transformed variable: not as easy to interpret as the untransformed variable, but seems to fit the data better
3. Logistic regression with splines: this is good for showing a clear departure from linearity in the regression model; however, it isn’t a good model to use for interpretation.
4. Linear regression with nadir psa as outcome, relapse as predictor: easiest model of all to interpret, though we’ve already seen that the data aren’t strictly linear so probably isn’t as good a fit
5. Linear regression with geometric mean nadir psa as outcome, relapse as predictor: difficult to interpret, though a better fit than model 4.

*A priori*, I would prefer model 2 to other models. A quick glance at the raw data shows that nadir psa is unlikely to be strictly linear (there are very tiny values and very large values with big leaps in between), so the log transformed nadir psa sounds more appropriate than an untransformed nadir psa as used in model 1. I would also discard model 3 because it can’t easily be used to interpret the association between relapse and nadir psa, only to indicate a departure from linearity. And I would prefer model 2 over models 4 and 5 because it seems more intuitive to analyze the data with relapse as our outcome, since in a clinical setting this is what we would be interested in predicting (even though at present, with these data, we don’t know which way the association goes for certain).

**4b**. Examining the lowest psa value post-therapy might not be the most useful measure to ascertain an association between psa and relapse. Even though in general we see that those with lowest nadir psa are less likely to relapse in 24 months, there is much variability, in which individuals with very low nadir psa often go on to relapse. Perhaps a better measure would be either a change in psa (difference between lowest and highest recorded in a certain time period after surgery), or a comparison with pre-surgery and/or baseline post-therapy measures of psa.