**Homework #3**

**Due: Wednesday, November 27, 2013 at 5:00 PM**

**1.**

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | **N(missing)** | **mean** | **sd** | **min** | **p25** | **p50** | **p75** | **max** |
| **no relapse in 2 yrs** |  |  |  |  |  |  |  |  |
| age (yrs) | 28 (0) | 66.7 | 5.84 | 58 | 63 | 66 | 70 | 81 |
| ps | 28 (0) | 84 | 9.6 | 50 | 80 | 80 | 90 | 100 |
| bss | 28 (0) | 2.3 | n/a | 1 | 2 | 3 | 3 | 3 |
| grade | 24 (4) | 2.1 | n/a | 1 | 1 | 2 | 3 | 3 |
| pretx | 23 (5) | 617.19 | 1252.080 | 4.8 | 45.0 | 100.0 | 387.0 | 4377.0 |
| nadir PSA | 28 (0) | 4.12 | 17.279 | 0.1 | 0.2 | 0.2 | 1.0 | 92.0 |
| **relapse in 2 yrs** |  |  |  |  |  |  |  |  |
| age (yrs) | 22 (0) | 68.4 | 5.68 | 61 | 64 | 68 | 71 | 86 |
| ps | 20 (2) | 77 | 11.8 | 50 | 70 | 80 | 80 | 100 |
| bss | 20 (2) | 2.8 | n/a | 2 | 3 | 3 | 3 | 3 |
| grade | 17 (3) | 2.2 | n/a | 1 | 2 | 2 | 3 | 3 |
| pretx | 20 (2) | 732.35 | 1357.341 | 25.0 | 69.5 | 174.0 | 530.0 | 4797.0 |
| nadir PSA | 22 (0) | 31.94 | 52.497 | 0.5 | 1.2 | 10.5 | 38.0 | 183.0 |
| **Total** |  |  |  |  |  |  |  |  |
| age (yrs) | 50 (0) | 67.4 | 5.77 | 58 | 63 | 66 | 70 | 86 |
| ps | 48 (2) | 81 | 11.1 | 50 | 80 | 80 | 90 | 100 |
| bss | 48 (2) | 2.5 | n/a | 1 | 2 | 3 | 3 | 3 |
| grade | 41 (9) | 2.1 | n/a | 1 | 2 | 2 | 3 | 3 |
| pretx | 43 (7) | 670.75 | 1287.638 | 4.8 | 46.0 | 127.0 | 429.0 | 4797.0 |
| nadir PSA | 50 (0) | 16.36 | 39.246 | 0.1 | 0.2 | 1.0 | 10.0 | 183.0 |

**2a. Output of logistic regression:**

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  | **OR** | **SE** | **p** | **95% CI LB** | **95% CI UB** |
| **nadir** | 1.034 | 0.024 | 0.156 | 0.987 | 1.083 |
| **bss** | 2.624 | 1.657 | 0.126 | 0.761 | 9.045 |
| **ps** | 0.952 | 0.031 | 0.138 | 0.892 | 1.016 |
| **\_cons** | 2.072 | 6.589 | 0.819 | 0.004 | 1051.108 |

Among individuals with similar bone scan scores and performance status scores, we estimate that for each point difference in lowest nadir PSA, the odds of relapse in two years is 1.034 times higher in the group with higher lowest nadir PSA, though this estimate is not statistically significant (P=0.156). This observation would not be unusual if the odds were 0.987 times as high or 1.083 times as high among the group with higher lowest nadir PSA.

**2b.**

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  | **Coef.** | **SE** | **p** | **95% CI LB** | **95% CI UB** |
| **log(nadir)** | 0.860 | 0.260 | 0.001 | 0.350 | 1.369 |
| **bss** | 0.852 | 0.820 | 0.299 | -0.755 | 2.460 |
| **ps** | -0.052 | 0.037 | 0.153 | -0.124 | 0.019 |
| **\_cons** | 1.119 | 3.725 | 0.764 | -6.182 | 8.419 |

Among individuals with similar bone scan score and performance status score, we estimate that for a two-fold increase in lowest nadir PSA, the odds of relapse in two years is 1.815 times higher in the group with the higher lowest nadir PSA and this estimate is statistically significant (P=0.001). This observation would not be unusual if the odds were 1.275 times higher or 2.58 times higher in the group with higher lowest PSA.

**2c. For linear splines below, knots at nadir values of 1, 4, and 16.**

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  | **OR** | **SE** | **p** | **95% CI LB** | **95% CI UB** |
| **nadir 1** | 29.617 | 56.154 | 0.074 | 0.721 | 1217.343 |
| **nadir 2** | 0.903 | 0.529 | 0.862 | 0.287 | 2.846 |
| **nadir 3** | 1.380 | 0.308 | 0.150 | 0.891 | 2.138 |
| **nadir 4** | 0.982 | 0.176 | 0.305 | 0.948 | 1.017 |
| **ps** | 0.937 | 0.039 | 0.112 | 0.864 | 1.015 |
| **bss** | 2.522 | 2.294 | 0.309 | 0.424 | 14.999 |
| **\_cons** | 0.507 | 2.052 | 0.867 | 0.000 | 1411.080 |

**Result of test for significance of all linear splines terms together:**

**Chi2 = 11.05**

**Prob>chi2 = 0.026**

Among groups with similar bone scan score and performance status scores, there is an association in the sample between lowest nadir PSA and odds of relapse within two years (P=0.026). For groups with similar bone scan score and performance status scores, and nadir PSA less than one, the odds of relapse in the group with nadir PSA one point higher is estimated to be 29.617 times the odds of relapse in two years among individuals in the lower PSA group. However, the comparison group for this estimate of odds (nadir PSA of 0) is outside our range of interest, and this estimate is therefore not very useful.

For similar circumstances with nadir PSA between 1 and 4, a group with a one-point increase in nadir PSA is estimated to have odds of relapse 0.903 times as high, which is not significant (P=0.862). This observation would not be unusual if the odds were 0.287 times as high or 2.846 times higher in the group with higher PSA. For similar circumstances, with nadir PSA between 4 and 16, the odds of relapse in a group with nadir PSA one point higher is estimated to be 1.380 times as high, which is not significant (P=0.150). This observation would not be unusual if the odds were 0.891 times as high or 2.138 times higher in the group with higher PSA. Under similar circumstances, with nadir PSA greater than 16, the estimated odds of relapse in a group with nadir PSA one point higher is estimated to be 0.982 times as high, which is not significant (0.305).

One reason for this paradox could be the inclusion of highly influential points in the model, which weight the splines regression towards the null much more than the overall regression. The maximum value for nadir PSA among individuals who did not relapse within two years is such a point. The point would be weighted more heavily in the splines regression for the slope between the last knot and the maximum nadir PSA value than in the overall linear regression, driving the odds to appear more similar across higher nadir PSA levels.

This regression could give some clues as to the threshold of nadir PSA levels that may indicate elevated odds of regressing within two years and offer values to analyze in future studies. From estimates in the analysis and related p-values, the range of nadir PSA values between 4 and 16 is estimated to be the area of strongest association, and may indicate the existence of threshold values on either bound of that spectrum.

**2d.**

**Part a: (If using the logit output rather than the logistic output used above)** The intercept in this model is an estimate of the log odds of relapse for a group with a bone scan score of 0, performance status score of 0, and nadir PSA of 0. Since this value is outside our range of interest (ex. the bone scan score is only values of 1, 2, and 3), this value on it’s own is not scientifically important.

**Part b:** The intercept in this model is also not scientifically useful because it is an estimate of the expected log odds of a group with a bone scan score of 0, performance status score of 0, and log(nadir) of 0, the latter of which is impossible.

**Part c:** Same as response to part a.

**3a.**

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  | **Coef.** | **SE** | **p** | **95% CI LB** | **95% CI UB** |
| **relap24** | 23.518 | 11.433 | 0.046 | 0.476 | 46.559 |
| **bss** | 6.846 | 4.689 | 0.151 | -2.604 | 16.295 |
| **ps** | -0.510 | 0.618 | 0.414 | -1.756 | 0.736 |
| **\_cons** | 31.028 | 53.122 | 0.562 | -76.033 | 138.089 |

I ran this linear regression to determine the association between the indicator of relapse within two years and expected nadir PSA values, where the indicator of relapse was the predictor of interest and nadir PSA values were the outcome. Bone scan score and performance status score were adjusted for to determine if the association between relapse within two years and mean nadir PSA values existed independently of these two factors. Robust standard errors were used.

In this case, the estimate of the coefficient for relap24 is the estimate of the average difference in nadir PSA values between two groups of individuals with the same bone scan score and performance status score. The estimate in this case is 23.518, indicating an estimated 23.518 point increase in PSA values in a group with relapse within 24 months compared with a similar group in terms of bone scan score and performance status score without relapse in 24 months. This value is significantly different from zero (P=0.046), and the observed difference would not have been uncommon in populations where the true difference was anywhere from 0.476 to 46.559.

**3b.**

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  | **Coef.** | **SE** | **p** | **95% CI LB** | **95% CI UB** |
| **relap24** | 2.614 | 0.593 | <0.001 | 1.418 | 3.810 |
| **bss** | 0.482 | 0.298 | 0.113 | -0.118 | 1.082 |
| **ps** | -0.007 | 0.028 | 0.795 | -0.063 | 0.049 |
| **\_cons** | -1.166 | 2.497 | 0.643 | -6.198 | 3.865 |

Using log nadir PSA as the outcome in this linear regression (using robust standard errors), between two groups with the same bone scan score and performance status score, the mean nadir PSA is estimated to be 13.654 times higher in a group with relapse within 24 hours compared to a group with no relapse within 24 months. The difference is significant (P<0.001) and would not have been unusual to see in a population with differences between 4.129 and 45.150 times higher in the group with relapse within 24 months.

**4a. Relative merits of each analysis:**

**Number 2, part a (logistic with continuous predictor nadir PSA, bivariate response relapse within 24 months):** This regression allows for an estimate of the ratio of odds of relapse for each interval increase in nadir PSA values. This regression can detect an overall first order trend in the association between nadir PSA and odds of relapse. A numerical difference (as opposed to a multiplicative difference) may be more useful for diagnostics and more easily interpretable. If the data itself has a purely linear relationship, this regression would fit the data well and offer better predictions of relative odds.

**Number 2, part b (logistic with log transformed continuous predictor PSA, bivariate response relapse within 24 months):** This regression allows for an estimate of the ratio of odds of relapse for each multiplicative increase in nadir PSA values. This regression would be useful if the expected data was curvilinear or there was a scientific reason to believe a log scale for the predictor made sense (ex. the predictor was actually measured on the log scale, like for pH). A multiplicative difference (as opposed to a numerical difference) may be more useful for diagnostics. It also downweights the importance of numerical differences among higher values and upweights numerical differences among lower values of the predictor (ex. a ten-fold increase in the predictor could pertain to the predictor going from 0.1 to 1 or 1 to 10). So, if sensitivity in lower values, with less sensitivity as values grow higher is of interest, this regression would be suitable.

**Number 2, part c (logistic with linear splines at three knots, bivariate response relapse within 24 months):** The main advantage of this regression is examination of associations between the predictor and outcome that cannot be fully explained by a linear or log transformed relationship. Particularly for this analysis, since thresholds are of interest, the relative placement of knots in the linear splines allows for inference into these associations within certain intervals on the predictor and the likelihood of certain knots being close to threshold values of the predictor.

**Number 3, part a (linear regression, nadir PSA response, relapse within 24 months as indicator predictor):** This regression compares mean nadir PSA values on a numerical difference scale between groups similar with respect to all other variables other than the POI (relapse). An advantage of this regression is that it alerts the investigator to any initial signs that there may be an association at all between nadir PSA values and relapse within 24 months. This regression seems ideal to use as an exploratory analysis to another study to encourage the start of this current study to examine the association with analysis used in 2a, because one does seem to exist in this model. This regression should be used if numerical differences among a wide range of PSA values are important (difference between 2 and 3 just as important as difference between 100 and 101).

**Number 3, part b (linear regression, log nadir PSA response, relapse within 24 months as indicator predictor):** This regression estimates multiplicative average differences between groups similar with respect to bone scale score and performance status score, but differing in their history of relapse within 24 months. An advantage of this regression is that it alerts the investigator to any initial signs that there may be an association at all between nadir PSA values and relapse within 24 months. This regression seems ideal to use as an exploratory analysis to another study to encourage the start of this current study to study the association with analysis used in 2b, because one does seem to exist in this model. This analysis should be used if numerical differences in smaller values are of more interest than the same difference among larger values (ex. the difference between 1 and 2 is more important than the difference between 100 and 101).

**A priori**…I would be most interested in using the regression in 2c, since previous studies have already shown the association between nadir PSA and relapse exists and investigation into a threshold value of PSA is now of greatest interest. A numerical rather than multiplicative difference seems to also be of more interest here given the range of nadir PSA values under consideration. However, given the distribution of nadir PSA values, more of a curvilinear relationship between nadir PSA and odds of remission might be more of interest, in which case I would choose the regression performed in 2b.

**4b.** One problem I could see with this study is the timing of nadir PSA measurements. Since the association of interest is the lowest nadir PSA measurement post-therapy and still being in remission after 24 months, the timing of the lowest nadir PSA measurement might be important. The length of time the hormonal therapy takes to treat the carcinoma isn’t specified, so the lowest value of nadir PSA could have occurred while the hormonal therapy was still in effect and the remission had not started, obscuring the apparent association.