Biostat 536 Homework 3

1. Descriptive statistics of data:

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | **No remission for >24mo** | | **Remission within 24mo** | |
| **Variable** | **Mean** | **Standard Deviation** | **Mean** | **Standard Deviation** |
| Nadir (of PSA levels after treatment in ng/mL) | 4.12 | 17.28 | 31.94 | 52.50 |
| Pretx (Pretreatment PSA level in ng/mL | 617.19 | 1252.08 | 732.35 | 1357.34 |
| PS (Ambulatory score [100=best]) | 83.93 | 9.56 | 76.50 | 11.82 |
| BSS (Metastasis score [1=least metastasis]) | 2.32 | 0.77 | 2.80 | 0.41 |
| Grade (Tumor grade [1=least aggressive]) | 2.08 | 0.83 | 2.24 | 0.75 |
| Age (years) | 66.71 | 5.84 | 68.36 | 5.68 |
| Obstime (Observed remission in months) | 42.07 | 12.05 | 11.14 | 6.40 |

1. A) After adjustment for bone scan and performance status, a logistic regression comparing the odds of relapse from prostate cancer for >24 months across groups defined by the nadir PSA level when modeled as a continuous, untransformed variable shows an odds ratio of 1.03, with a 95% confidence interval of 0.94-1.13. The p value for the model is an insignificant value of 0.07. If the model were significant, this would mean a 1.03 higher odds ratio for every addition unit in the measure of nadir PSA after treatment, measured in ng/mL.

B) After adjustment for bone scan and performance status, a logistic regression comparing the odds of relapse from prostate cancer for >24 months across groups defined by the nadir PSA level when modeled as a continuous, log-transformed variable shows an odds ratio of 2.36, with a 95% confidence interval of 1.27-4.40. The p value for the model is a significant value of 0.02. This means there is a 2.36 times higher odds ratio for every doubling in the unit measure of nadir PSA after treatment, measured in ng/mL.

C) After adjustment for bone scan and performance status, a logistic regression comparing the odds of relapse from prostate cancer for >24 months across groups defined by the nadir PSA level when modeled untransformed as a linear spline with knots at 1, 4 & 16ng/mL results in a model that has overall significance at p=0.02. The odds ratio for one additional unit in nadir PSA between 0ng/mL and 1ng/mL is 29.62 with a 95% confidence interval of 1.36-645.63, the odds ratio for one additional unit in nadir PSA between 1ng/mL and 4ng/mL is 0.90 with a 95% confidence interval of 0.33-2.50, the odds ratio for one additional unit in nadir PSA between 4ng/mL and 16ng/mL is 1.38 with a 95% confidence interval of 0.94-2.02, and the odds ratio for one additional unit in nadir PSA above 16ng/mL is 0.98 with a 95% confidence interval of 0.96-1.00.

D) After recoding the bone scan score so that a 0 value is permissible, the interpretations of the intercepts are:

2a – an odds ratio for remission of 5.44 is found for men with a nadir PSA value of 0, a 0 (healthiest measure) on bone scan, and an ambulatory score of 0.

2b - an odds ratio for remission of 7.18 is found for men with a nadir PSA value of 0, a 0 (healthiest measure) on bone scan, and an ambulatory score of 0.

2c – an odds ratio for remission of 1.28 is found for men with a nadir PSA value of 0, a 0 (healthiest measure) on bone scan, and an ambulatory score of 0.

1. A) A linear regression comparing nadir PSA with relapse and corrected for bone scan score and performance status shows that those who relapse have a 23.5 unit higher arithmetic mean value of PSA, measured in ng/mL, with a 95% confidence interval of 0.48-46.56. The p value for this model, however, at 0.07, does not reach significance.

B) A linear regression comparing log nadir PSA with relapse and corrected for bone scan score and performance status shows that those who relapse have a 2.61 unit higher geometric mean value of PSA, measured in ng/mL, with a 95% confidence interval of 1.42-3.81. The p value for this model is <0.0001.

1. A) Relative merits of each approach:

The approaches in question 2 are more desirable if the question of interest is the odds of remission given a nadir PSA value. The approaches in question 3 are more desirable if the question of interest is the difference of mean PSA value given remission/no remission. Beyond that:

2a – This approach is desirable if it is expected that the response of the odds of remission vs. the nadir PSA values will follow a linear relationship.

2b – This approach is desirable if it is expected that the response of the odds of remission vs. the nadir PSA values will follow an exponential relationship.

2c – This approach is desirable if it is expected that the response of the odds of remission vs. the nadir PSA values will follow different linear relationships within different ranges of PSA (specifically 0-1, 1-4, 1-16 & >16).

3a – This approach is desirable if it is expected that the values recorded for nadir PSA value are numerous enough to be representative

3b – This approach is desirable if it is expected that there are large values in a portion of the PSA dataset that would contribute disproportionate weight to the resultant means. Based on the difference calculated between 3a and 3b, and in examining the PSA data, this does appear to be the case.

A priori, it would depend on the question I wanted to ask.

-If I were a clinician wanting to make a diagnostic model, I would choose one of the approaches in question 2 – likely either based on an exponential relationship or using splines. I would make this decision based on what research has already been done on nadir PSA values and remission outcomes. Barring any extant evidence to that effect, I would probably go with an exponential relationship (2b).

-On the other hand, if I were a researcher trying to determine whether there was actually a difference in nadir PSA by remission status, I would likely use a geometric mean approach to avoid any outliers throwing off my result (3b).

B) The problem is that the observed time in remission is quite different between those in remission and those not in remission. The average observed remission time for no remission >24mo was ~42 months, and the average observed remission time for remission within 24mo was ~11. The window of time during which each group is susceptible to remission needs to be approximately the same in order for the two groups to be comparable. The measure should have been to follow all patients for 24 months and to record remission vs. not during that period. See table below, copied from page 1:

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | **No remission for >24mo** | | **Remission within 24mo** | |
| **Variable** | **Mean** | **Standard Deviation** | **Mean** | **Standard Deviation** |
| Nadir (of PSA levels after treatment in ng/mL) | 4.12 | 17.28 | 31.94 | 52.50 |
| Pretx (Pretreatment PSA level in ng/mL | 617.19 | 1252.08 | 732.35 | 1357.34 |
| PS (Ambulatory score [100=best]) | 83.93 | 9.56 | 76.50 | 11.82 |
| BSS (Metastasis score [1=least metastasis]) | 2.32 | 0.77 | 2.80 | 0.41 |
| Grade (Tumor grade [1=least aggressive]) | 2.08 | 0.83 | 2.24 | 0.75 |
| Age (years) | 66.71 | 5.84 | 68.36 | 5.68 |
| Obstime (Observed remission in months) | 42.07 | 12.05 | 11.14 | 6.40 |