BIOS 665: Problem Set 2

Assigned: September 12, 2017

Due: September 28, 2017

Reminder: For all hypothesis tests, please state the method, the null hypothesis, the test statistic, the p-value, and the interpretation of the test using a two-sided significance level of 0.05, unless otherwise stated.

Helpful hints: For estimates and tests, simply copying and pasting SAS output without any commentary will not earn full credit, especially on exams. Highlighting is not considered commentary. However, commentary can be as simple as: The 95% CI for the odds ratio is (\_\_\_, \_\_\_\_).

I have followed the Honor Code. Signed: \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

1. The following data arise from a health policy study that included interviews with subjects from both rural and urban geographic regions. Use logistic regression to describe the relationship of favorable opinion response (vs. unfavorable response) to stress and residence. Use reference groups of “Urban” for residence and “Low” for stress.

|  |  |  |  |
| --- | --- | --- | --- |
| Residence | Stress | Favorable Unfavorable | Total |
| Urban  Urban  Urban | Low  Medium  High | 64 20  76 50  122 100 | 84  126  222 |
| Total |  | 262 170 | 432 |
| Rural  Rural  Rural | Low  Medium  High | 55 30  68 60  115 90 | 85  128  205 |
| Total |  | 238 180 | 418 |

1. State the assumptions for using logistic regression to model these data. Mathematically specify the variables used in the model, and interpret all parameters.
2. Provide a quantity that expresses the effect of high stress (as compared to low stress) on favorable response (vs. unfavorable response), and provide a 99% two-sided confidence interval for this quantity. How would you estimate this quantity and its 99% confidence interval by hand, given the computer output? Show your calculations.
3. Formally test the hypothesis that residence has no effect on opinion response.
4. Provide predicted probabilities for favorable response for each of the following:
   1. An individual from an urban area with low stress.
   2. An individual from a rural area with medium stress.
5. Test the hypothesis that the model fit is adequate. Briefly justify your choice for the test.
6. The following results for two logistic regression models which were fit to describe the relationship between the probability of no headache pain (versus otherwise) at one hour post-treatment and explanatory variables for treatment (placebo, high dose, low dose), baseline severity (moderate, severe), and (for Model 2) their interaction are presented below. The results for the two models include estimated parameters, their standard errors, and the maximized log-likelihood for the fit of the model.

|  |  |  |
| --- | --- | --- |
|  | **Model 1** | **Model 2** |
| Parameter | Estimate Standard Error | Estimate Standard Error |
| Intercept | -1.642 0.395 | -1.524 0.489 |
| Severe (Baseline Severity) | -1.498 0.364 | -1.913 1.168 |
| High dose | 2.236 0.512 | 1.900 0.567 |
| Low dose | 1.113 0.386 | 1.131 0.567 |
| Severe\*High | NA NA | 0.860 1.189 |
| Severe\*Low | NA NA | 0.061 1.218 |
|  |  |  |
| Log-likelihood | -109.674 | -107.897 |

\* Explanatory variables equal 1 if the category applies, and 0 if otherwise.

\*\* NA = Not Applicable

1. For Model 1, state the relevant assumptions for the application of logistic regression to these data. Specify the mathematical structure of the model, including mathematical definitions for all explanatory variables.
2. Calculate the odds for each of the following conditions. Please provide numeric solutions.

|  |  |  |  |
| --- | --- | --- | --- |
| **Baseline Severity** | **Dose** | **Model 1** | **Model 2** |
| Severe | Placebo |  |  |
| Severe | Low |  |  |
| Severe | High |  |  |
| Moderate | High |  |  |

1. Numerically calculate the odds ratios of no headache (vs. otherwise) for each of the following conditions:

|  |  |  |
| --- | --- | --- |
|  | **Model 1** | **Model 2** |
| Odds ratio for severe to moderate, for placebo |  |  |
| Odds ratio for severe to moderate, for low dose |  |  |
| Odds ratio for high dose to placebo, for moderate baseline |  |  |
| Odds ratio for high dose to placebo, for severe baseline |  |  |

1. Comment on the calculations in c), particularly in regard to the comparisons between rows 1 and 2, and between rows 3 and 4.
2. How would you interpret the parameter corresponding to low dose in Model 1?
3. For Model 1, provide a 90% confidence interval and estimate for the odds ratio of no headache pain (vs. otherwise) for high dose versus placebo, controlling for baseline severity.
4. For Model 2, provide a predicted probability of no headache pain for an individual on low dose who had severe pain at baseline. Please provide a numeric solution.
5. Through the results for Models 1 and 2, apply a statistical test at the α = 0.05 significance level to assess the hypothesis that Model 1 has satisfactory goodness of fit in the sense that any association between baseline severity and probability of no headache pain at one hour post-treatment is homogeneous across the placebo, low dose, and high dose groups.
6. Consider the following data from a study that evaluated the relationship between dose and dichotomous response concerning pain relief.

|  |  |  |
| --- | --- | --- |
| Dose | Favorable Unfavorable | Sample size |
| 1 mg  10 mg  100 mg | 21 39  24 36  42 18 | 60  60  60 |

* 1. Use logistic regression to describe the relationship between favorable response (vs. unfavorable response) and dose.
     1. Mathematically specify the model, treating dose as categorical, and interpret the parameters. Use dose=1 mg as the reference group.
     2. Now, re-specify the model mathematically, treating dose as continuous and using a log10 transformation. Interpret these model parameters.
     3. ~~Provide estimates and 95% Fiducial Limits for the dose values corresponding to ED25, ED50, and ED75. In other words, provide estimates and 95% confidence limits for the dose values which produce a response with 0.25, 0.50, and 0.75 probabilities, respectively.~~
  2. ~~Use probit analysis to describe the relationship between favorable response (vs. unfavorable response) and dose.~~
     1. ~~Mathematically specify the model, treating dose as continuous using a log10 transformation. What types of interpretable quantities correspond to either the parameters or functions of the parameters?~~
     2. ~~Provide estimates and 95% Fiducial Limits for the dose values corresponding to ED25, ED50, and ED75. In other words, provide estimates and 95% confidence limits for the dose values which produce a response with 0.25, 0.50, and 0.75 probabilities, respectively.~~
  3. ~~Briefly compare and contrast your results from part (a)(iii) and part (b)(ii).~~