ERHS 642 Logistic Regression Spring 2016

Midterm Exam

Discussing any part of the exam with others is not permitted.
Students caught cheating or attempting to cheat will be reported to the CSU Conflict Resolution and Student Conduct Services (CRSCS) office and will receive a grade of 0 on the exam.

Consider Adolescent Placement Study (APS) data set described in the text on pages 26/27.

1. (1 point) Create a permanent SAS data set.

2. (1 point) Create a new dichotomous outcome variable, PLACE2, as follows:

   0 = Outpatient or day treatment
   1 = Intermediate residential or residential

3. (4 points) For the variable LOS (length of hospitalization), obtain separately for those with PLACE2=0 and those with PLACE2=1
   - Mean
   - Standard deviation
   - Quartiles
   - Minimum and maximum
   - 5 lowest and 5 highest values

4. (4 points) Obtain the predicted values for LOS. Graph
   - The predicted values vs PLACE2; and
   - The scatterplot of LOS vs. PLACE2

   on the same set of axes.

5. (2 points) Show that LOS is a significant predictor of PLACE2.

6. (4 points) Use spline effect plots to determine if LOS should be transformed. (Select knots and connections you find helpful.)

7. (4 points) Use the fp method to determine the best scale for LOS. Show and describe/explain all results.

8. (3 points) Based on your results in questions 6 and 7, how would you model LOS? Explain!
   Hint: Try out your idea(s) to see if they actually work.

9. (3 points) In a model containing only DANGER as the independent variable, is DANGER a significant predictor of PLACE2? Show a table containing the maximum likelihood estimates of the model coefficients and their p-values.

10. (3 points) In a model containing all study variables as independent variables (except ID and the outcome variables), is DANGER a significant predictor of PLACE2? (For simplicity, keep all continuous variables linear.) Show a table containing the maximum likelihood estimates of the model coefficients and their p-values. Explain why your results for DANGER are different from the results in question 9.

11. (3 points) Create a new variable as follows:

    NEURO_D=0 if NEURO=0, 1 or 2
    NEURO_D=1 if NEURO=3

    Based on the results in question 10 and other considerations, why does it make sense to combine NEURO categories 0, 1 and 2?

12. (2 points) In the model you created in question 10, replace the 4-category NEURO variable with the dichotomous variable NEURO_D you created in question 11. Show a table containing the maximum likelihood estimates of the model coefficients and their p-values. Is NEURO_D statistically significant?

13. Start with the model you created in question 12. Remove all variables that are non-significant at the 0.05 level.
    a. (2 points) Show a table of the reduced model containing the maximum likelihood estimates of the model coefficients and their p-values.
    b. (4 points) Determine if the full model from question 12 is significantly better than the reduced model.

14. (4 points) Compare the ORs from the models in questions 12 and 13 and determine if there is evidence of confounding.

15. Start with the reduced model you created in question 13.
    a. (3 points) Determine if NEURO_D is a multiplicative effect modifier of any of the other model covariates (use the 0.1 level of significance)
    b. (3 points) Notice that inclusion of the interaction between NEURO_D and CUSTD results in quasicomplete separation. Explain why this happens and show results to back up your conclusion.

16. (4 points) Start with the reduced model you created in question 13 and add the interaction between NEURO_D and RACE. Present meaningful ORs and 95% CIs for the model covariates.

17. (3 points) Interpret the ORs and 95% CIs from question 16.

18. (3 points) Based on earlier analyses in this exam, is your conclusion for LOS correct? Explain!

YOU MUST SHOW ALL RELEVANT MODEL OUTPUT
AND ATTACH YOUR SAS PROGRAM.