BIOS 507 HOMEWORK 7

Due 4/14/2025 by 11:59pm

Directions: Complete all questions. Any R or SAS code used should be attached at the end of the homework. Collaboration is encouraged, but the final product must be your own work.

Problem 1

A study was carried out to identify risk factors associated with giving birth to a low birth weight baby (weight less than 2500 grams). Use the low birth weight data, `LBW.csv`. for this analysis. A description of the relevant variables in the dataset is given below.

lbw Low birth weight (1: yes, 0: no)

age age of the mother in years

race Race (1: white, 2: black, 3: other)

smk smoking status (1: smoker, 0: non-smoker)

Consider the variable `lbw` as the outcome of interest for the following analysis.

1. Conduct an exploratory analysis to assess the bivariate association of the outcome variable with age, race, and smoking status. **Write a summary of your findings.**

2. Fit a logistic regression model including age, race, and smoking status as explanatory variables. Use "non-smoker" and "white" as reference categories. Based on this model, estimate the following odds ratios (OR) and the 95% confidence intervals:

   - OR comparing a smoker and a non-smoker. Does the choice of race affect the OR?

- OR associated with an increase of 5 years in the mother's age (separate OR for each race).

3. Fit a new logistic regression model that includes the variables in the model from (b) and the interaction of smoking status and race. Based on this model, estimate the following odds ratios (OR) and the 95% confidence intervals:

   - OR comparing a smoker and a non-smoker. Does the choice of race affect the OR?

   - OR associated with an increase of 5 years in the mother's age (separate OR for each race).

4. Why are your estimates different between Part 2 and 3?

5. Conduct goodness of fit tests for your models. What are your conclusions?

6. Construct the ROC curves for your models.

Problem 2

A psychologist conducted a study to examine the nature of the relation, if any, between an employee's emotional stability (X) and the employee's ability to perform in a task group (Y). Emotional stability was measured by a written test for which the higher the score, the greater emotional stability. Ability to perform in a task group ($Y = 1$ if able, $Y = 0$ if unable) was evaluated by the supervisor. The results for 27 employees were recorded and are available on canvas under `performance_ability_data.csv`. Use this data to answer the following questions (adapted from your textbook):

1. Write down the postulated model in two forms: (1) in terms of the probabilities, (2) in terms of the log odds

2. Find the MLE for β_0 and β_1. Write out the fitted response function in terms of the log odds.

3. Plot the data and overlay the fitted response function. Note that your x-axis should be emotional stability and the y-axis should be the estimated probability of ability to perform in a task group.

4. Obtain $e^{\hat{\beta}_1}$ and interpret this value.

5. Give the estimated probability that employees with an emotional stability test score of 550 will be able to perform in a task group.

6. Give a rough estimate of the emotional stability test score for which 70% of the employees with that test score are expected to be able to perform in a task group.

7. Which goodness of fit tests would be acceptable for this data?