HW6

INSTRUCTIONS

In this homework, we will be working on analyzing the association between *seatbelt use* to determine if it varies according to sex, age, and BMI using the BRFSS.48K.csv. dataset. Categorize seatbelt use into always/sometimes vs. never, excluding individuals with other responses. Submit on blackboard. Don't forget to use promising practices.

0. Conduct any data import and library steps needed to complete this homework.

1. Make a new seatbelt variable categorizing seatbelt use into always/nearly always vs. never/sometimes with always/nearly always being coded as 1 for logistic models and never/sometimes being coded as 0. Exclude individuals with other responses. Check how many you have in each seatbelt category in your newly created variable and make sure that the number is correct.

2. An important first step in any regression modeling exercise is to know your data. One of the components of knowing your data includes data visualization and creating bivariate tables. Make plots to visually examine seat belt use by the continuous variables BMI and age. For categorical variables, calculate the percentage of individuals in each category who are seat belt wearers.

3. Check whether running age and bmi as continous variables is appropriate for your regression models.

4. You make the decision based on these results to model age as a categorical variable. Create ~10 year categories for age (18-29, 30-39, 40-49, 50-59, 60-69, 70-80) to use in your logistic regression models.

5. Run a univariate logistic regression model for each risk factor of interest (age category, bmi, sex, and calculate ORs and 95% CIs).

6. Run a multivariate logistic regression model, exclude individuals who reported "Refused" for sex (note: the reason for this is because there are so few and the models don't like levels of a variables with few numbers).

7. Determine the top 5 influential observations using a Cook's Distance plot.

8. a. Determine the percentage of seatbelt wearers/sometimes wearers vs. non-wearers predicted by your model. b. From these numbers calculate and report the sensitivity and specificity of your model.

9. Exclude the top 5 influential observations and compare Betas between models with and without these observations.

10. Interpret your results in one paragraph.