**Exercise 4**

The data for the first part of this exercise (logistic regression) are from a substance abuse study (n=244) completed in New England examining factors related to resuming substance use after an acute detoxification.  The study actually randomized patients to a behavioral intervention, but we are not using that variable but rather another non-randomized treatment.  We selected a small subset of the variables to use logistic regression.  Some of the variables were collapsed and their names were changed to make them easier to understand and more relevant.  Like we have done in previous exercises, start by examining the dataset for univariate and bivariable distributions.

Variables

|  |  |  |
| --- | --- | --- |
| ID |  |  |
| substuse | If participant engaged in substance abuse during follow-up 0=no, 1=yes | Dichotomous |
| age | Age in years | Continuous |
| female | Female = 1, male=0 | Dichotomous |
| homeless | Slept on the street or in a shelter 1+ nights in past 6 months | Dichotomous |
| socsupp | Social Support scale (0 – 13.5) | Continuous |
| racebl | Race (Black or African American= 1, White = 0) | Dichotomous |
| satreat | Prior history of substance abuse treatment | Dichotomous |

**This time, instead of an analysis report, please address the following items:**

1. Create a table of odds ratios and respective 95% confidence intervals for the odds ratio of substance abuse at follow up among men vs. women, homeless or not, prior history of substance abuse treatment vs. none. Do not use a full model to calculate the ORs but use 1 model for each comparison.
2. Calculate the odds ratio associated with a 1-year increase in age.
3. Create a variable representing a 5-year increase in age. Calculate the odds ratio associated with a 5-year increase in age.
4. Create a variable for social support that is dichotomous and represents a mean split so that you are comparing the odds ratio among people with the lowest to highest social support. Report that odds ratio with a 95% CI.
5. Now create a full model that includes 5-year age increments, sex, homelessness, prior history of substance abuse treatment, race, and low vs. high social support. Report these odds ratios in a table. Identify and interpret any differences and similarities between the individual models and the combined model.