

BIOS 6312 - Modern Regression Analysis
Spring 2021
Lab #4

Objective: To investigate the effectiveness of REACH/ REACH + FAMS on hemoglobin A1c in a population of adults with uncontrolled diabetes. Please be sure to download the documentation for the REACH study.

Data pre-processing:

1. Load the REACH data set into Stata.
2. Create a three-level treatment variable.
3. Using the levels from the REACH data dictionary, relabel the raceeth variable: 0 = White, 1 = Black, 2 = Hispanic, 3 = Asian, 4 = Other.

Primary data analysis:

1. Center the baseline A1c variable.
2. Fit a linear regression model to quantify the association between 6-month A1c and REACH treatment, controlling for centered baseline A1c.

$$E[\text{a1c6}|\text{a1c0}_c, \text{REACH}] = \beta_0 + \beta_1 \text{a1c0}_c + \beta_2 \text{REACH} \quad (1)$$

3. Fit a linear regression model to quantify the association between 6-month A1c and treatment assignment, controlling for centered baseline A1c.

$$E[\text{a1c6}|\text{a1c0}_c, \text{treat3}] = \beta_0 + \beta_1 \text{a1c0}_c + \beta_2 I(\text{treat3} = \text{"REACH"}) + \beta_3 I(\text{treat3} = \text{"REACH + FAMS"}) \quad (2)$$

4. If asked about the association between 6-month HbA1c and the REACH treatment alone, which coefficient(s) would you turn to? How would you interpret that coefficient?
5. Would the variable representing centered baseline A1c be considered a confounder or a precision variable? Fit model 1 again, this time excluding baseline A1c. What differences do you notice?

$$E[\text{a1c6}|\text{a1c0}_c, \text{treat3}] = \beta_0^* + \beta_1^* I(\text{treat3} = \text{"REACH"}) + \beta_2^* I(\text{treat3} = \text{"REACH + FAMS"}) \quad (3)$$

6. Test for an overall treatment effect in model (2).

7. Fit a linear regression model to quantify the association between 6-month A1c and treatment assignment, controlling for centered baseline A1c, age, race, and gender.

$$\begin{aligned} E[\text{a1c6}|\text{a1c0}_c, \text{treat3}, \text{age}, \text{race}, \text{gender}] = & \beta_0 + \beta_1 \text{a1c0}_c + \beta_2 I(\text{treat3} = \text{"REACH"}) \\ & + \beta_3 I(\text{treat3} = \text{"REACH + FAMS"}) \\ & + \beta_4 \text{age} + \beta_5 I(\text{race} = \text{"Black"}) \\ & + \beta_6 I(\text{race} = \text{"Hispanic"}) + \beta_7 I(\text{race} = \text{"Asian"}) \\ & + \beta_8 I(\text{race} = \text{"Other"}) + \beta_9 \text{gender} \end{aligned} \quad (4)$$

8. Test for an overall treatment effect in model (4).
9. Develop a prediction interval for 6-month A1c among 60-year old, Black males with a baseline A1c of 7.2 who received REACH+FAMS.
10. Generate a residual-versus-predictor plot for each predictor in the model. Provide an interpretation of the residual-versus-age plot.

List of useful Stata commands:

- import delimited
- label values
- predict
- replace
- generate/ egen
- lowess
- group
- regress, robust
- label define
- testparm