Bootstrap

You plan to study the marriage behavior for the US Black women in the Midwest region. The data you will use is a sample from the March 2009 US Current Population Survey. The data file and description can be downloaded from Canvas. We use the model

$$\Pr(married = 1|x_i) = \Phi(x_i^{'}\beta)$$

where x_i is a vector of observable characteristics for individual i, including age, age square and education. That is $x_i^{'}\beta = \beta_0 + \beta_1 age + \beta_2 age^2 + \beta_3 educ$.

1. Read the data description file and construct the sample. You should have 433 observations in you data. Define a binary variable "mar" as 1 if married and 0 otherwise from the variable "marital".

2. Specify the ML function and estimate the Probit model.

3. Explain how you can calculate the bootstrap standard error.

4. Now you want to test the hypothesis that the level of education does not have any effect on the marriage decision. You are asked to use different methods:

   (a) t-test using both ML and Bootstrapped standard errors.

   (b) Confidence interval method using

      i. the percentile method

      ii. the percentile t-method.

   (c) What the difference between these two different confidence interval? Which one you would prefer and why?

5. What will be the bias corrected estimate for the parameter for age?

6. We want to know at which age the women will be most likely to marry.

   (d) Can you propose a estimator for this purpose? Implement your estimator using our data.

   (e) How can you obtain standard error for your proposed estimator? (hint: the Delta method and the Bootstrap method).

   (f) Test the hypothesis that this age is 40.

   (g) Can you present a graphic representation of the distribution of your proposed estimator?