**Econometrics 4400 HW 2**

*You are welcome to work in teams, but please: (1) turn in your own write up (i.e. in your own words) and computer code, (2) acknowledge with whom you worked at the top of your assignment. Please print out your code and attach it to your answers. Please submit a physical copy in class unless extenuating circumstances prevent your attendance.* ***I will grant 5% extra credit to your grade for this HW if you A) turn in a physical copy of your homework that is B)******completely typed out including any equations with R code attached****.* ***Under no circumstances will I accept late homework.***

*Some questions may have multiple possible answers.* ***Justify your claims for full credit****.*

1. You will use a dataset from a randomized experiment conducted by Marianne Bertrand and Sendhil Mullainathan.[[1]](#endnote-1) They sent fake resumes out to employers in response to job openings. The resumes differ by numerous characteristics, including names, education levels, and number of previous jobs. The researchers wanted to test whether resumes with “black-sounding” names received fewer interview callbacks.

First, download the dataset bm.Rda from Carmen under the second homework, and open it in RStudio. The datset contains the following variables:

* education (0 = not reported, 1 = HS dropout, 2 = HS graduate, 3 = some college, 4 = college and above)
* ofjobs (number of jobs listed on resume)
* yearsexp (number of years of work experience listed on resume)
* computerskills (whether the resume mentions computer skills, 1 = yes, 0 = no)
* female (1 = yes, 0 = no)
* black (1 = yes, 0 = no)
* call (whether the resume received an interview, 1 = yes, 0 = no)
  1. Calculate the average difference in the callback rate (call variable) between individuals with “black-sounding” names (the “treatment” group, for whom black=1) and those without (the “control” group, for whom black=0).
  2. Conduct t-tests for the equality of means between treatment and control groups for the education, ofjobs, yearsexp, and computerskills variables. Report the t statistic for each t-test and distinguish at what levels (0.10, 0.05, and 0.01) we can reject the null that the means are equal.
  3. Based on your results from part **c**, would you say that the treatment and control groups in this experiment are balanced? Explain, specifically referencing your results.
  4. Now estimate the following simple regression model: . Report the estimate   as well as the standard error. Report the t-statistic and the associated p-value. At what levels is it significant?
  5. Now estimate the following multiple regression model with control variables: . Report the estimate   as well as the standard error. Report the t-statistic and the associated p-value. At what levels is it significant? At what levels is each of the control variables significant at?
  6. Conduct a hypothesis test for the joint exclusion restriction: . Report the appropriate test statistic and the associated p-value, and distinguish at what levels we can reject the null.
  7. Compare the coefficient estimates on the treatment variable from parts **d** and **e**. Does the coefficient estimate from the multiple regression appear to be very different from that of the short regression? Why or why not? Explain clearly, referencing results from previous parts of question 1.
  8. If your answers for part **b** were different (i.e. if you found significantly different means suppose they were insignificant and vice versa), how would you expect your answers for parts **c,** **f,** and **g** to change?

1. Bertrand, Marianne and Sendhil Mullainathan. “Are Emily and Greg more employable than Lakisha and Jamal? A field experiment on labor market discrimination.” *The American Economic Review* 94, no. 4 (2004): 991-1013 [↑](#endnote-ref-1)