Due Date: December 21st, 2021 at 11:59PM

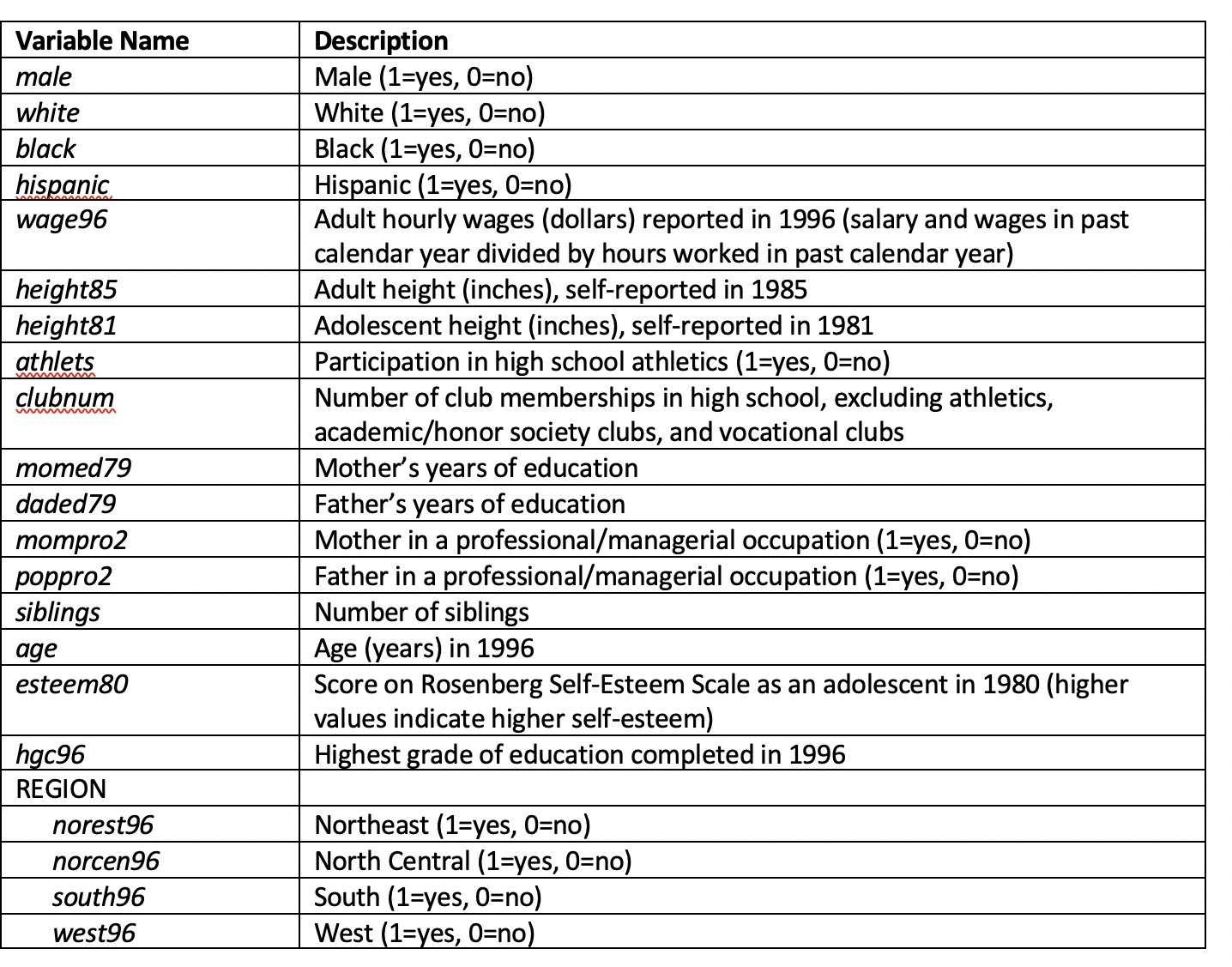
# Instructions

You should turn in your solution write-up on ELMS before 11:59pm EST on December 21st.

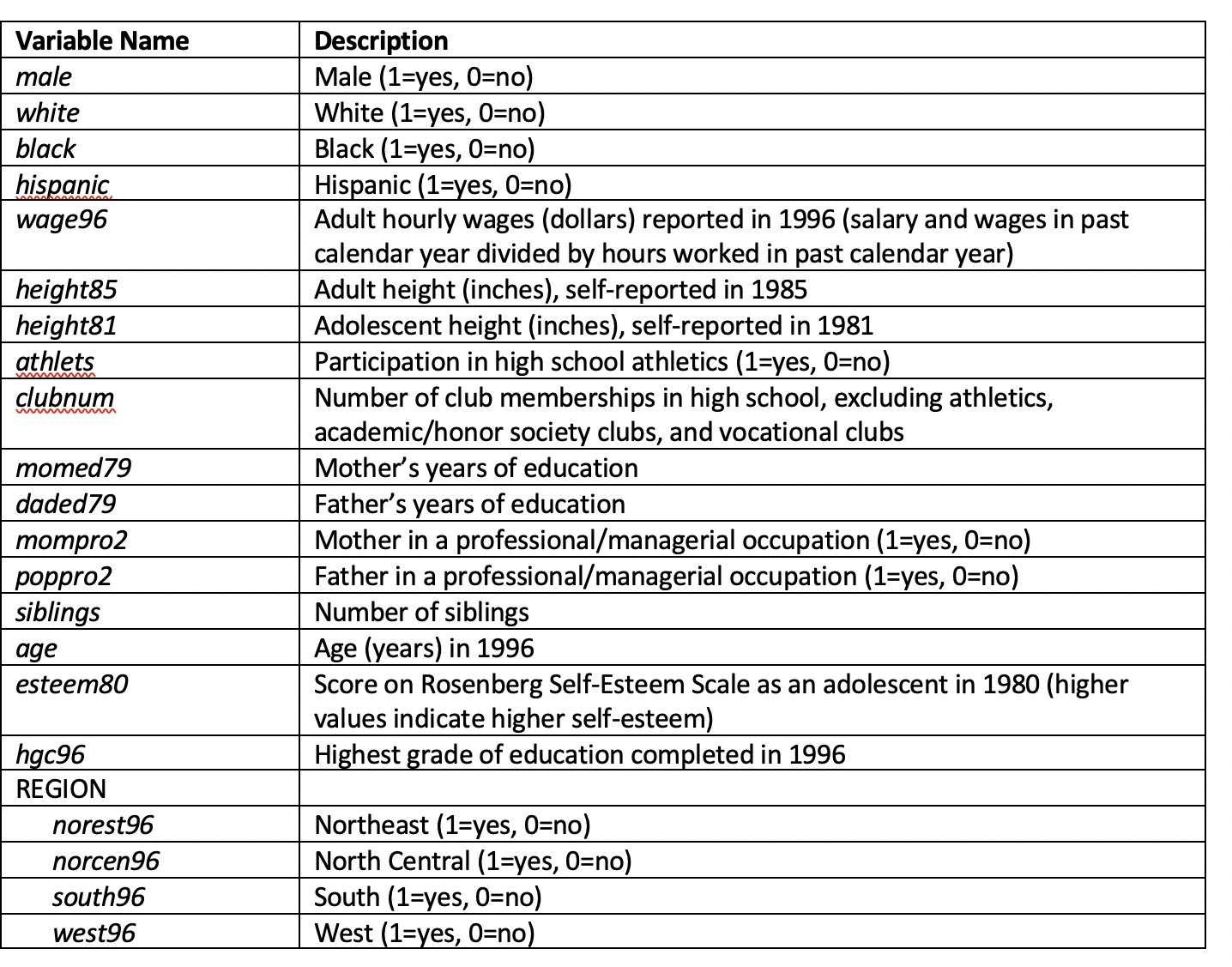
* The solution to each question should be presented with tables and figures included in the document and referenced within the text. I should not have to guess which figure/table you are talking about.
* Your write-up should explain clearly what you did. You should assume in your write-up that I cannot understand code, so you must describe and interpret the results in words, and, when asked, present results in nicely formatted tables and figures, not just R, Excel, or STATA output. The instructions are meant to ensure that you create your own tables by extracting relevant information from the computer programs rather than just pasting output.
* Your solution to each of the questions should start on a new page.
* You can use any available function in R, Stata, or Excel to solve each of the problems. If you cannot successfully write code that works, describe in your write-up what you were attempting to do and explain how you would have incorporated these results into your solution.
* Your submission on ELMS should also include your code.
* Answer any **three** of the following six questions. If you answer more than three of the questions, only the first three will be graded. In your write-up, make sure that you clearly indicate which questions you are answering. Each question is equally weighted.

Go on the internet and find a time-series dataset that exhibits autocorrelation. Note that you may **not** use GDP data (nominal or real) in answering this question. You must use data other than GDP.

1. Provide the reference/link for the dataset. If a link cannot be provided, give details on how to get the data (I should be able to find the data, otherwise, you will lose points here).
2. Show that there is autocorrelation in the data. Interpret/explain your finding.
3. Transform the data to address the autocorrelation. Explain what you did and why this will/should fix the autocorrelation.
4. Show that you have fixed the autocorrelation. How do you know that your time-series no longer exhibits autocorrelation?

Note that the data for this questions is called “problem 2”. In a 2004 article published in the Journal of Political Economy, Nicola Persico, Andrew Postlewaite, and Dan Silverman analyzed data from the National Longitudinal Survey of Youth (NLSY) 1979 cohort to assess the relationship between height and wages for white men. The NLSY is a nationally representative sample of 12,686 young men and women who were 14-22 years old when they were first surveyed in 1979. These individuals were interviewed annually through 1994 and biannually since then. This data set selected study variables based on the NLSY data. The following is the codebook for this question:

1. Estimate two OLS regression models: one in which adult wages is regressed on adult height for all respondents, the other in which adult wages is regressed on adult height and adolescent height for all respondents. Discuss differences across the two models. Explain why the coefficient on adult height changed.
2. Assess the multicollinearity of the two height variables.
3. Notice that IQ is omitted from the model. Is this a problem? Why or why not?
4. You’re the boss! Use the data in this file to estimate a model that you think sheds light on an interesting relationship. The specification decisions include deciding whether to limit the sample and what variables to include. Report only a single additional specification. Describe in no more than two paragraphs why this is an interesting way to assess the data.



1. Estimate an OLS regression model with adult wages as the dependent variable and adult height, adolescent height, and a dummy variable for males as the independent variables. Does controlling for gender affect the results?
2. Generate a female dummy variable. Estimate a model with **both** a male dummy variable and a female dummy variable. What happens? Why?
3. Reestimate the model from part (a) separately for males and females. Do these results differ from the model in which male was included as a dummy variable? Why or why not?
4. Estimate a model in which adult wages is the dependent variable and there are controls for adult and adolescent height in addition to dummy variable interactions of male times each of the two height variables. Compare the results to the results from part (c).

in which people also vote for state and national offices: turnout will be higher in on-cycle elections, and teachers and teachers unions will have relatively less influence.

From 2003 to 2006, all districts in the sample elected their school board members off-cycle. A change in state policies in 2006 led some, but not all, districts to elect their school board members on-cycle from 2007 onward. The districts that switched then stayed switched for the period from 2007 to 2009, and no other district switched.

1. Estimate the pooled model of *LnAvgSalaryi,t* = *β*0 + *β*1*OnCyclei,t* + *ϵi,t*. Discuss whether there is potential bias here. Consider in particular the possibility that teachers unions are most able to get off-cycle elections in districts where they are strongest. Could such a situation create bias? Explain why or why not.
2. Run a one-way fixed effects model in which the fixed effect relates to individual school districts. Interpret the results and explain whether this model accounts for time trends that could affect all districts.
3. Now use a two-way fixed effects model to estimate a difference-in-difference approach (note - here, you are **not** doing a difference-in-difference model, rather you are using two-way fixed effects to estimate a difference-in-difference model so all that you need to do here is a panel with two-way fixed effects). Interpret the results and explain whether this model accounts for (i) differences in preexisting attributes of the switcher districts and nonswitcher districts and (ii) differences in the post-switch years that affected all districts regardless of whether they switched.
4. Before we analyze the experimental data, let’s suppose we were to conduct an observational study of access based on a sample of Americans and we ran a regression in which the dependent variable indicates having met with a member of Congress and the independent variable was whether the individual donated money to a member of Congress. Would there be concerns about endogeneity? If so, why?
5. Use a probit model to estimate the effect of the donor treatment condition on probability of meeting with a member of Congress. Interpret the results.
6. What factors are missing from the model? What does this omission mean for our results?
7. Use a linear probability model (LPM) to make your estimate. Interpret the results. Assess the correlation of the fitted values from the probit and LPM models.

were told that they would be interviewed again later. The program they studied aired in California prior to the vote on Proposition 209, a controversial proposition relating to affirmative action.

1. Estimate a bivariate OLS model in which the information the respondent has about Proposition 209 is the dependent variable and whether the person watched the program is the independent variable. Comment on the results, especially whether and how they might be biased.
2. Estimate the model in part (a), but now include measures of political interest, newspaper reading, and education. Are the results different? Have we defeated endogeneity?
3. Why might the assignment variable be a good instrument for watching the program? What test or tests can we run?
4. Estimate a 2SLS model from using assignment to the treatment group as an instrument for whether a given respondent watched the program. Use the additional independent variables from part (b).
5. What do the 2SLS results suggest about the effect of watching the program on information levels? Compare the results to those in part (b). Have we defeated endogeneity?