

Homework 2
Econ 322: Econometrics
Summer 2019

Upload to Sakai by **Wednesday August 14** Uploaded to Sakai by 5:00pm

Please upload one .pdf document with your responses. **Name that document FirstnameLastname_Econ322HW2.pdf**. Also upload your .R command file separately. **Failure to follow the naming convention or upload a .pdf will result in the loss of points**

You may work in groups of 2-3 students, but you must do the write-up on your own. ***Put names of all group members on Homework.*** Groups may not turn in identical write-ups.

Determinants of Weight

The following questions are based on a sample of respondents in New Jersey in 2015 from the CDC's Behavioral Risk Factor Surveillance System, which is an excellent source for individual-level health data. For more on this data, see: http://www.cdc.gov/brfss/annual_data/annual_2015.html. Note I took a 1000 person random sample from amongst the New Jersey respondents. The codebook is very important and also available to download.

- 1) You want to run a regression of weight on age (as a quadratic), gender, income level, whether you are married, and height.

Data set up notes: Given the variables that you have in the dataset, you need to create age squared in order to do this, you can do this via something like

`data$agesq=data$age*data$age`. (where data is whatever you called your data, and agesq is just a new variable name, you can call it something else if you like). Also, right now, R will interpret sex as two numbers 1 and 2, but you can tell R that sex is a factor variable, so that it knows that it has mutually exclusive categories. You can do this via something like `data$sex.f=factor(sex)`. (where data is the name of your data, and sex.f is a new variable name, you can call it something else).

Helpful hint: run `options(scipen = 999)` to turn off scientific notation and make things easier to read

Can you include all 3 income levels in your regression? Why or why not?

Go ahead and run your regression in R (use lowincome as the reference group for income levels, don't forget to include your sex factor variable instead of just sex, don't forget to use the commands after for heteroskedasticity-robust standard errors which we went over in class in R-tutorial 2).

- 2) Create a table that shows these results, put heteroskedasticity-robust standard errors in

parenthesis under the coefficients. Put appropriate stars on the coefficients to indicate statistical significance (***) for $p < 0.01$, ** for $p < 0.05$, * for $p < 0.1$). Include the R^2 and the number of observations. See the sample table at the end of this assignment. You need to add an appropriate title and table notes, so that the reader can understand the table. Use actual words in the table, not variable names.

- 3) Questions about these results (for any questions about “significance” you can default to the 5% level or state a different level that you choose in your sentence):
 - a. Is there a quadratic relationship between age and weight? Is it significant? How do you know?
 - b. How do you interpret the coefficient on your gender variable? Are men significantly heavier than women?
 - c. What is the 95% confidence interval for the association between being married and weight? What does this mean?
 - d. Is income significantly associated with weight? How do you know? (which test statistic did you use and why?)
 - e. Do middle income people weigh significantly more than high income people? How do you know? (note you need to run an additional command)
 - f. How confident can you be that height has a statistically significant association with weight.
 - g. How much of the sample variation in weight can you explain with this regression?
- 4) Now you want to add race into your regression and education level into your regression. Run your new regression in R, include all the variables from the previous question AND these new ones. Add a second column to your table that includes these new results.
 - a. Does this new regression explain more of the sample variation in weight?
 - b. Is race significantly associated with weight? How do you know? (which test statistic did you use and why? See the Joint hypothesis testing R tutorial we did in class)
 - c. Is education level significantly associated with weight? How do you know? (which test statistic did you use and why?)
- 5) Run the regression from 4) again, but do not include height. Include this as the 3rd column in your table.
 - a. Do any of your coefficients change a lot? Choose the one that changed the most and tell an omitted variable bias story to explain the difference in the two results

Sample Table Format:

Table: TITLE

	(1)	(2)	(3)
Variable 1	Coefficient1 (std error1)	Coefficient1 (std error1)	
Variable 2	Coefficient2 (std error2)	Coefficient2 (std error2)	
Variable 3	Coefficient3 (std error3)	Coefficient3 (std error3)	
Variable 4		Coefficient4 (stderror4)	
		Etc...	
Constant	Constant1	Constant2	
(leave blank line above # of obs)			
Number of Observations	N1	N2	
R^2	$[R^2_1]$	$[R^2_2]$	

Notes: [Any important Notes to help the reader understand the table]