Homework 3

Homework 3 is due May 7th at 11:59 PM. All submissions should be in PDF format and be submitted through Gradescope. Note: You need to tag the individuals problems in Gradescope according to the instructions. Gradescope will allow you to highlight, for example, the response from Question 1 and then mark it as question 1.

Create a folder called "hw3" in a convenient location on your computer. Download **county_characteristics.dta** which is available in Canvas and put the file in your "hw3" folder.

Download the do-file **hw3_blank.do**. This do-file has a few hints to get you started. You can continue to fill in this do file to complete your homework.

The first thing to do is to change your working directory to the folder named "hw3". If your data is in the correct place, and you have correctly changed your working directory, you should now be able to type "use county_characteristics.dta, clear" to load the data.

The data for this homework comes from Opportunity Insights, and is a county-level dataset that has characteristics of different counties. The codebook for the data, which will give you some details about the variables in the dataset, is on the homework page.

1. In this homework, we will explore how rent within a county is associated with population density. There is a variable in the dataset called rent_twobed2015 that reports the average rent for a two-bedroom apartment in the corresponding county.

   What is the average, max, and min of this variable? Paste both the code used to generate these statistics and report the statistics themselves. (5 points)

   ```
   sum rent_twobed2015

   average = 684.8963; min = 172.1595; max = 2085.227
   ```

2. The variable fips_county is a unique identifier that identifies counties in the dataset. For example, the fips_county code for San Diego is equal to “06073”.

   What is the average rent of a two bedroom apartment in San Diego? Paste the code that computes this as well as the average rent. (3 points)

```
sum rent_twobed2015 if fips_county == "06073"

average = 1441.758
```

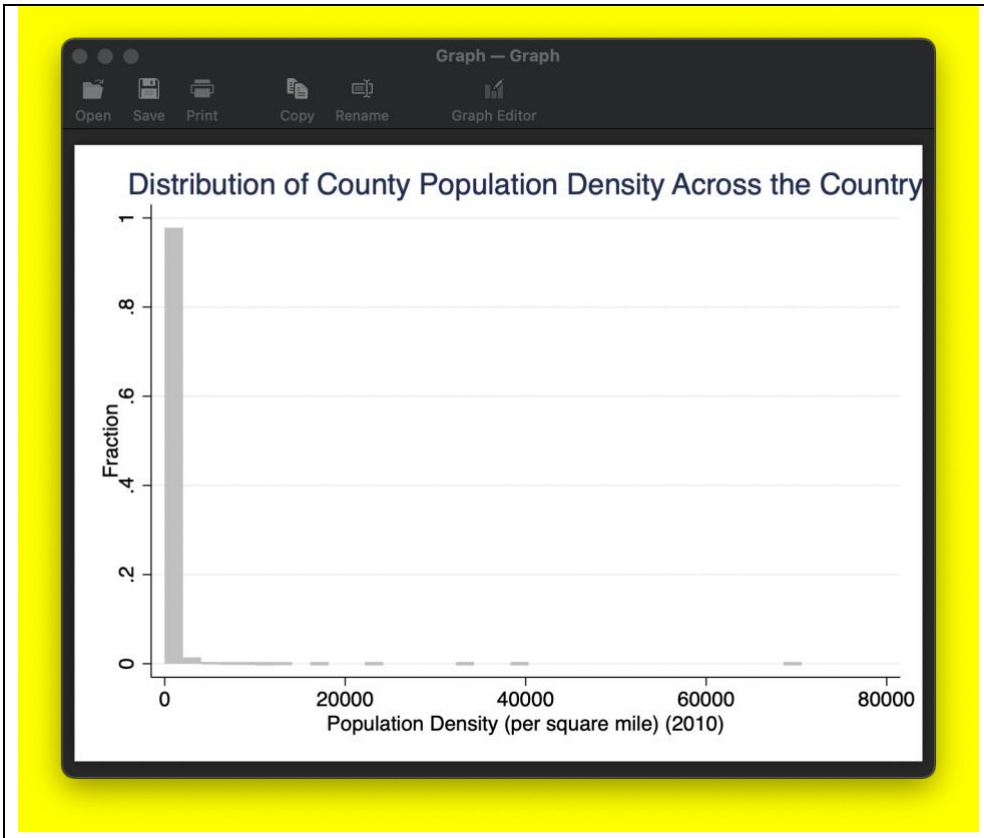
3. A. We are going to explore the relationship between average two-bedroom rent and population density (measured by the variable popdensity2010, which is equal to the number of residents per square mile).

   First, what is the population density for San Diego county? Paste the code below and report the population density. (3 points)

```
summarize popdensity2010 if fips_county == "06073"

population density of SD = 735.8169
```

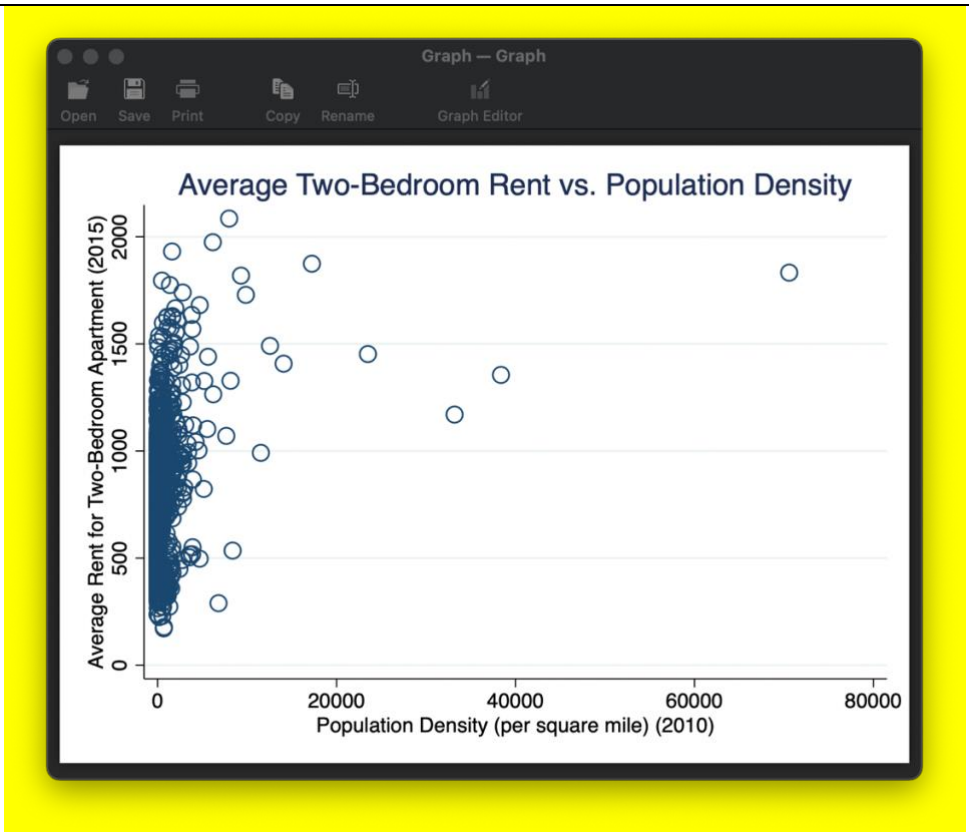
   B. Construct a histogram of popdensity2010 with fraction of observations on the y-axis. Make sure to label axes. Paste an image of the histogram below. (5 points)



C. Briefly comment on the distribution. What are some conclusions about the popdensity2010 variable that you can conclude from this figure? (5 points)

As we can see, approximately 95% of the counties have a population density that falls within the first bin. This tells us a few things: (1) the range for population density is big, (2) very few counties have high population densities per square mile, and (3) we're going to need to normalize the data if we want to understand it.

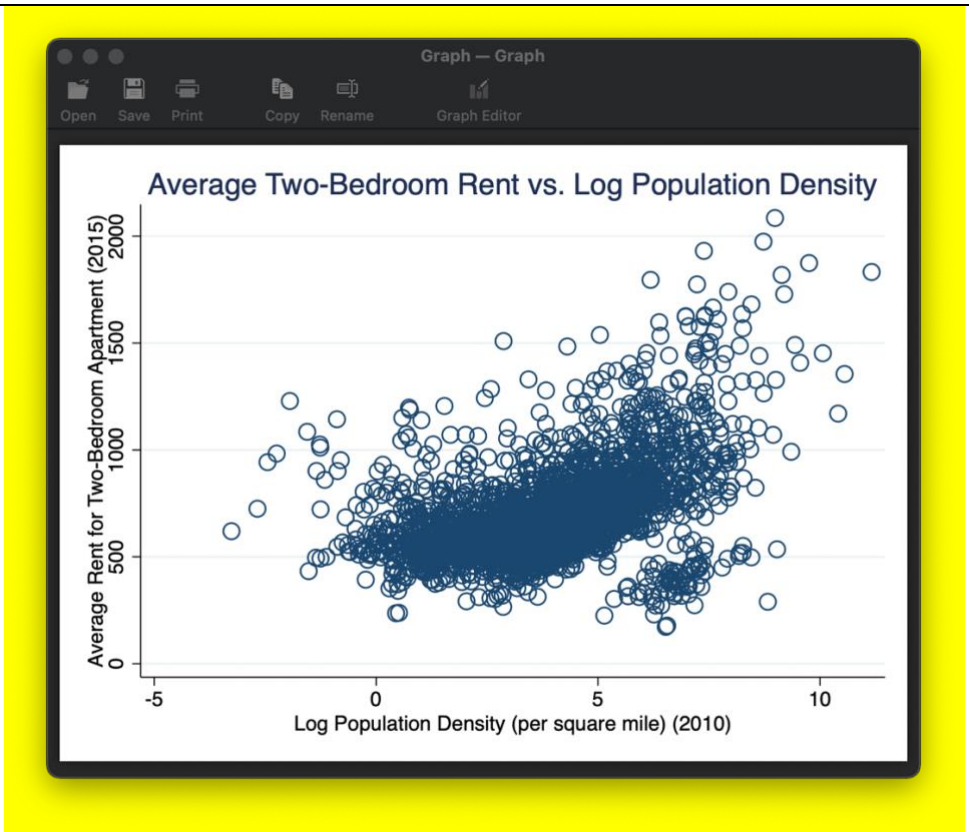
4. To begin exploring the relationship between average two-bedroom rent and population density, make a scatterplot with rent_twobed2015 on the y-axis and popdensity2010 on the x-axis. Make sure to label axes appropriately. Paste the scatterplot below. (5 points)



5. A. Create a new variable called log_popdensity2010, which is equal to the log of popdensity2010. Paste the code that generates this variable below. (3 points)

```
gen log_popdensity2010 = ln(popdensity2010)
```

B. Make a scatterplot with rent_twobed2015 on the y-axis and log_popdensity2010 on the x-axis. Paste a screenshot of your scatterplot below. (5 points)



C. Does this scatterplot more clearly display the relationship between rent and population density. Explain why or why not. (5 points)

Yes, this scatterplot does present a clearer picture of the relationship between rent & population density. After normalizing the Population Density data, we can see that there appears to be a positive correlation between rent for a two-bedroom apartment and the population density in a county.

6. A. Now we are going to estimate a linear regression. Recall from lecture that for a linear regression, we fit a model of the form $Y_i = \beta_0 + \beta_1 * X$, where we estimate the parameters β_0 (the intercept) and β_1 (the slope coefficient). Estimate a linear regression in Stata where rent_twobed15 is the Y-variable (or dependent variable) and popdensity2010 is the X-variable (or independent variable). Therefore, we are modeling rent as a function of the population density. Paste the code the runs this regression below. (5 points)

```
reg rent_twobed2015 popdensity2010
```

B. Interpret the slope coefficient. What does it tell us about how a change in population density is associated with a change in the average rent of a two-bedroom apartment? Make sure to be specific about the units, size, and direction of this relationship. (5 points)

For every 1 unit increase in Population Density per square mile, we expect the Average Rent of a Two-Bedroom Apartment will go up by .035 dollars.

C. Is this relationship statistically significant at the 5 percent level? How can you tell? (5 points)

The p-value that is produced from running this regression is 0.000, which suggests that this is indeed statistically significant at the 5% level. Because $p < 0.05$, this means that there's a less than 5% chance that this relationship is random.

D. Compute the predicted value of rent_twobed2015 for a county with a population density equal to 1000. Report the predicted value below. (4 points)

710.07878

E. Compute the predicted rent_twobed2015 for a county with a population density equal to 2000. Report the predicted value below. (4 points)

745.57243

7. A. Now we are going to estimate a linear regression with log_popdensity2010 as our X-variable instead of popdensity2010. Our Y-variable is still rent_twobed2015. Paste the code the runs this regression below. (3 points)

```
reg rent_twobed2015 log_popdensity2010
```

B. Interpret the slope coefficient. What does it tell us about how a change in log population density is associated with a change in the average rent of a two-bedroom apartment. Make sure to be specific about the units, size, and direction of this relationship. (5 points)

For every 1 unit increase in Log Population Density per square mile, we expect the Average Rent of a Two-Bedroom Apartment will go up by 53.57 dollars (log = 3.98 dollars).

C. Compute the predicted value of rent_twobed2015 for a county with a population density equal to 1000. Report the predicted value below. (4 points)

845.28989

D. Compute the predicted rent_twobed2015 for a county with a population density equal to 2000. Report the predicted value below. (4 points)

882.42375

8. Make a scatterplot with log_popdensity2010 on the x-axis, rent_twobed2015 on the y-axis and the regression line. You can do this either by following the instructions in the slides to add a best-fit line, or by using the command “lfit”. Type “help twoway lfit” to read about this command. Paste the scatterplot with a regression line below. (5 points)

9. (17 points)

A. Pick another variable in the dataset that you find interesting and think may be related to rent in the county. Write a paragraph that starts by explaining:

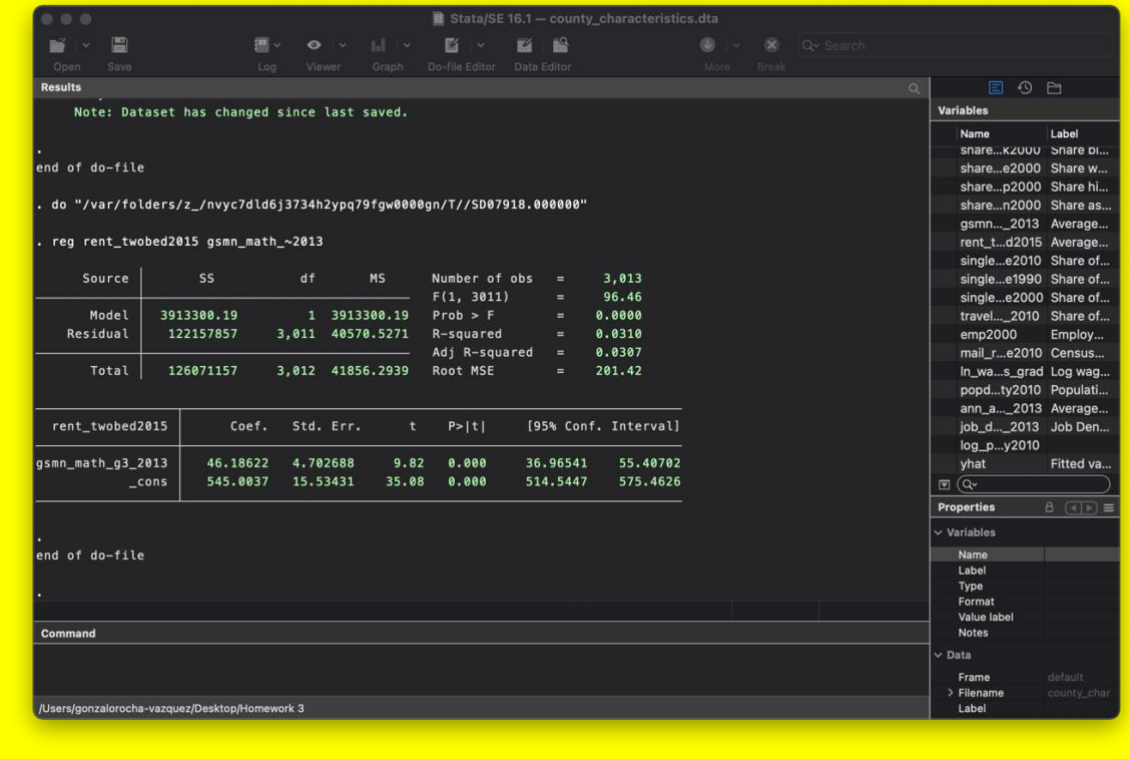
(1) What variable you are choosing to study

(2) Why you think it may be related to rental prices.

I'd like to explore the effect that quality of schools (IV) has on rent prices (DV). In order to measure quality of schools, I will utilize the gsmn_math_~2013 variable which measures Average School District Level Standardized Test Scores in 3rd Grade in 2013. The rent_twobed2015 variable, which measures Average Rent for Two-Bedroom Apartment in 2015, will continue to be used for rental prices. I believe that this will have an effect on rental prices because the quality of schools in an area is one of the main selling-points within real-estate. Many parents prefer to send their kids to high-quality

schools if possible, many of them are even willing to commute to schools outside of their district to do so. Rental Prices would be higher in these areas since realtors and landlords could raise their prices to profit from the proximity to high-quality schools.

B. Estimate a linear regression between rent and the variable you have chosen. Paste results of your regression table (i.e. take a screenshot of the regression output)



C. In a paragraph, answer the following questions about the regression you just ran:

(1)What is the interpretation of the slope coefficient?

(2) Is the relationship statistically significant at the 5 percent level or not? (Explain how you know).

(3) Is the result consistent with your original hypothesis? (note: it does not matter if your initial hypothesis is incorrect, just explain how the results relate to your hypothesis).

Our slope coefficient can be interpreted as, “For every 1 unit increase in Average School District Level Standardized Test Scores in 3rd Grade, we expect the Average Rent for Two-Bedroom Apartment to increase by 46.19 dollars”. Given that our p-value was 0.000, and a statistical significance at the 5 percent level requires $p > 0.05$, we can conclude that this is statistically significant at the 5% level. This result does provide support for my hypothesis, suggesting that rental prices are influenced by the quality of schools in the area. Specifically, it suggests a positive relationship between the two: higher quality of schools → higher rental prices.