Instructions: You will be doing the problems in this assignment by hand and in Stata. For the problems by hand, show your work for each step. For any Stata questions, when asked for graphs, save the graph as a .png file and paste the image into the appropriate section of the word document. Paste your code that produced the graph below the graph. When asked for output, you can use a screen capture of the Stata output (saved as a .png and pasted into the homework document). Paste the code used to create the output under the screenshot.

Dataset 1. Annual salaries and education levels for 13 randomly selected adults.

id	wage	educ
1438	548	17
3479	275	14
4934	275	8
2169	346	15
2628	750	10
2633	865	13
233	325	12
2532	649	14
5107	462	14
2186	921	13
4418	1026	13
3237	338	11
1987	865	12

Note: id = random identification number of respondent, wage = annual salary in 100s of dollars, educ = years of education completed.

Part I. The mayor of a small town is wondering whether to invest in more education-related programs and wants to know the link between education and earnings. A random sample of residents' earnings and years of education is included in dataset 1. Use Dataset 1 to answer the following questions (show your work used to calculate the relationships and statistics asked for).

1. Find the line of best fit for the relationship between wages and education. (HINT: The formula for the simple linear regression line is in Stata Handout 9).

2. Using the line of best fit, calculate the standard error for the estimated relationship between education and wages.

3. Is the relationship between education and wages significant? (HINT: you should calculate the t-score and p-value for your estimated slope for your answer).

4. What proportion of the variation in wages can be explained by years of education?

Part II. A mayor is considering implementing stronger noise ordinances. This is due, in part, because of the mayor's suspicion that both physical and mental health outcomes are shaped by sleep and the noise ordinance might prevent sleep disruptions. Use dataset 2 and Stata to answer the following questions.

1. Show a scatter plot with a best fit line and 95% confidence interval for the effect of hours of sleep on BMI.

2. Create the same plot as question 1 for the effect of hours of sleep on the mental health index.

3. Estimate a simple linear regression of the relationship between hours of sleep and BMI.

4. What is the 95% confidence interval of the effect of each hour of sleep on BMI?

5. Is the relationship between sleep and BMI statistically significant?

6. How much of the variation in BMI is explained by sleep?

7. Estimate a simple linear regression of the relationship between hours of sleep and mental health.

8. Is the relationship between sleep and mental health statically significant?

9. How much of the variation in mental health is explained by sleep?

Extra credit

1. Estimate a multiple regression model of mental health status as a function of sleep and poverty.

2. How does the effect of sleep on mental health change after controlling for poverty?

3. Estimate a multiple regression model of BMI as a function of sleep and the consumer price index (cpi2009).

4. Is the relationship between sleep and BMI statistically significant after accounting for the cost of consumer goods in the area?

5. Now add a control for poverty to your model of BMI.

6. How did the estimated relationship between sleep and BMI change after accounting for CPI and poverty status?