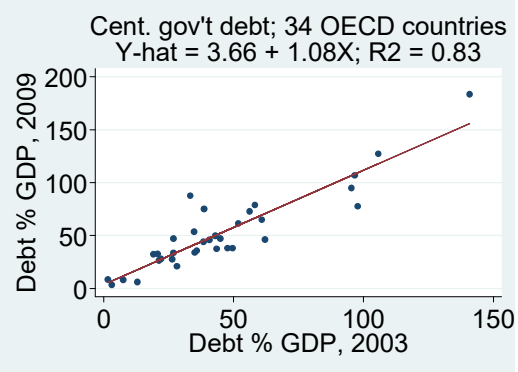
Homework 18: ECO220Y

Required Exercises: Chapter 18: 1, 2; Chapter 19: 22, 23, 25, 39, 45

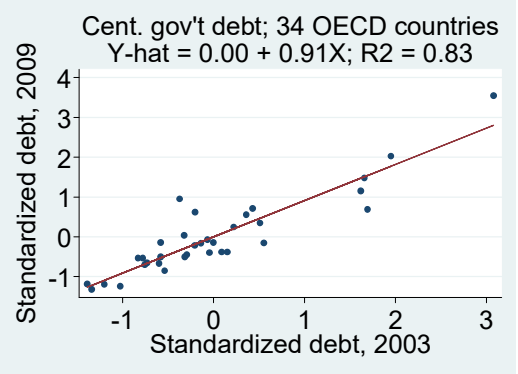
Required Problems:

(1) Review your course materials for weeks 4, 5, and 6 (i.e. Chapters 6 and 7, Lectures 4 – 6, Homework 4 – 6, and "Logarithms in Regression Analysis with Asiaphoria"). Afterwards, to test your understanding, try these:

(a) Consider data on central government debt downloaded from the OEDC website (http://stats.oecd.org/Index.aspx?DatasetCode=GOV_DEBT). Using the scatter diagram and OLS estimates to the right, fully interpret the intercept, slope and R^2.

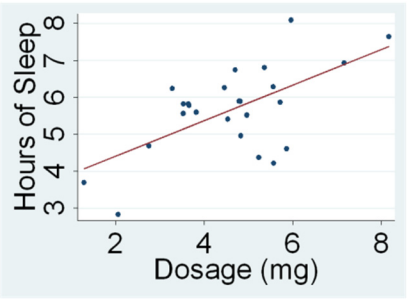


(b) Continuing with Part **(a)** consider standardized versions of these two variables. Results are shown to the right. Is it a coincidence that $(0.91)^2 = 0.83$? How do you interpret 0.91?



(c) Recall the drug (sleeping aid) trial example considered in Lectures 4 and 5. To refresh your memory below are the raw data, summary statistics, scatter diagram, and OLS line. How do you obtain the OLS intercept and slope (show calculations)? Why are these data experimental? How do you interpret the intercept and slope?

	Dosage (mg) x_i	Hours of Sleep y_i
1	5.9	4.6
2	3.5	5.8
3	7.2	6.9
4	3.6	5.8
…	…	…
25	8.2	7.6
X-bar = 4.61		Y-bar = 5.66
s_x = 1.51		s_y = 1.18
s_{xy} = 1.09		



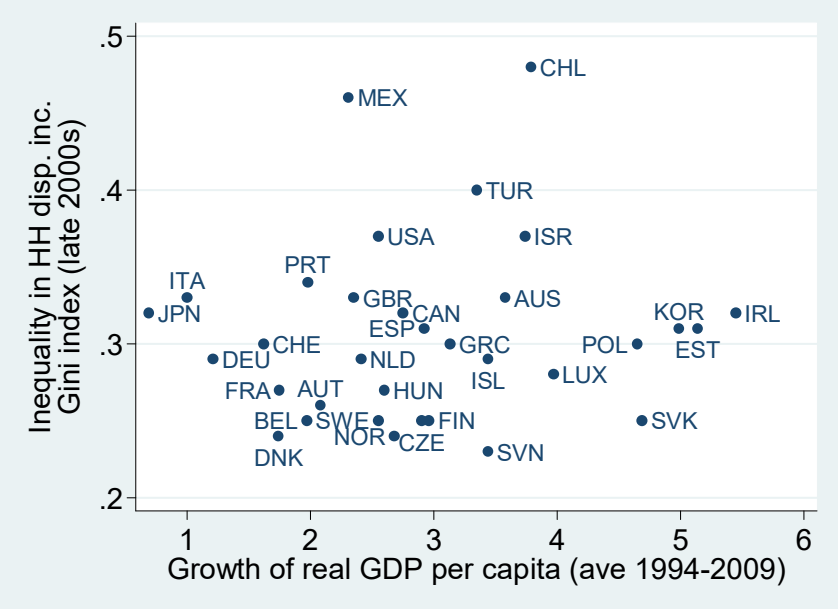
$$\widehat{Sleep}_i = 3.44 + 0.48\ Dosage_i$$

(2) Consider again OECD countries' debt as a percent of GDP in 2003 versus 2009 shown in parts **(a)** and **(b)** above. Go through each of the SIX assumptions discussed in Lecture 18 and explain why each one holds or is violated in that case.

(3) Consider a 2012 OECD report "Reducing income inequality while boosting economic growth: Can it be done?" The report is publically available: https://www.oecd.org/eco/growth/49421421.pdf. Here are a few excerpts:

– "There is growing consensus that assessments of economic performance should not focus solely on overall income growth, but also take into account income distribution." (OECD, 2012, p. 182)

– "Despite a vast theoretical literature on the link between inequality and growth, no general consensus has emerged and the empirical evidence is rather inconclusive. A simple scatter plot of inequality and growth also shows no link." (OECD, 2012, p. 194)



Open the report and scroll to page 194 where there is a scatter diagram titled "Figure 5.9. There is no simple link between inequality and growth." Further, the *StatLink* below the figure allows you to immediately download these data and analyze them yourself in Excel (or other software like Stata). Above is a Stata graph I drew with these data. I also used Stata to compute the coefficient of correlation $r = 0.0828$ and the coefficient of determination $R^2 = 0.0069$.

(a) Are these data observational? What can you conclude from the very low correlation (and the correspondingly very low R^2)? *Does no correlation mean no causation?*

(b) This is a particularly tricky example because there can be reverse causation. In other words, it is not clear what's the x-variable and what's the y-variable. Inequality can affect growth and growth can affect inequality. Of course when computing the coefficient of correlation and the coefficient of determination it makes no difference which is x and which is y. However, when estimating a regression line it *does* matter. Here are the results from Stata:

OLS: $\widehat{inequality} = 0.294 + .004\ GDP_{growth}$
OLS: $\widehat{GDP_{growth}} = 2.407 + 1.676\ inequality$

If you algebraically rewrite either (i.e. solve for x in terms of y): do they match? Explain.

(c) Give some specific examples of lurking variables (i.e. confounding, unobserved, omitted variables) relevant in this context.

(4) Practice key concepts from the Fall term, which are *highly* relevant again, with some old test questions.

(a) Solve questions (1), (2) and (6) on Term Test #1, November 2016:
http://homes.chass.utoronto.ca/~murdockj/eco220/TT220_1_NOV16.pdf.

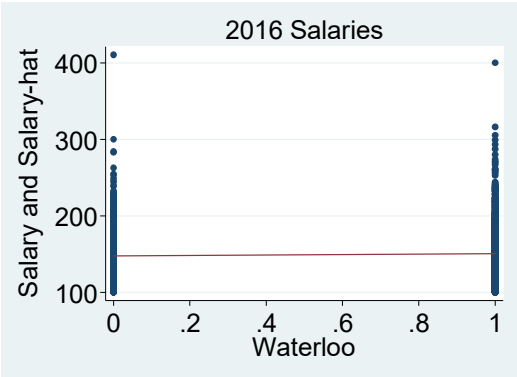
(b) Solve questions (1), (4) and (6) on Term Test #2, November 2017:
http://homes.chass.utoronto.ca/~murdockj/eco220/TT220_2_NOV17.pdf

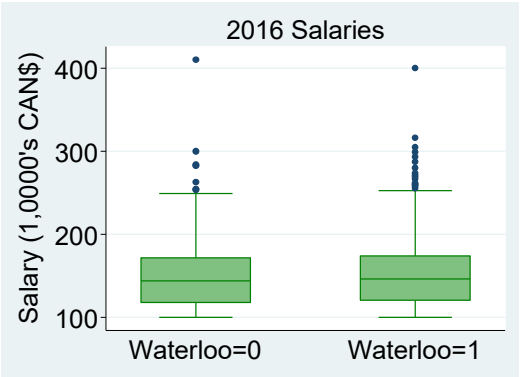
(c) Re-solve questions (2), (3), (4) and (6) on Term Test #2, November 2018:
http://homes.chass.utoronto.ca/~murdockj/eco220/TT220_2_NOV18.pdf

(5) Consider the 2017 disclosure of Ontario salary data and comparing the 2016 salaries for the University of Waterloo versus Ryerson University. The table below provides some basic summary statistics.

Employer	n	Mean ($1,000s)	S.d. ($1,000s)
Ryerson University	1,189	147.7342	34.97342
University of Waterloo	1,357	150.5722	36.70673

Consider also the graphs below and the OLS results.





OLS Results: Salary-hat = 147.7342 + 2.838028*Waterloo; n = 2,546; R-squared = 0.0016; s_e = 35.908

(a) Noting the diagrams and the OLS results, what is the interpretation of the OLS regression line?

(b) What is the interpretation of the value of the R-squared?

(c) What is the interpretation of the value of the s_e?

(d) Consider approaching this question as a difference in means (i.e. using the methods learned in Chapter 14). Using the summary statistics provided in the table for Ryerson and Waterloo, compute the pooled variance – $s_p^2 = \frac{(n_1-1)s_1^2+(n_2-1)s_2^2}{n_1+n_2-2}$ – for the special case where you assume the population variances are equal (Section 14.5 of the textbook). Notice that $\sqrt{s_p^2}$ matches the value of the s_e in the OLS results. Explain how this makes sense in light of the homoscedasticity (i.e. equal variances) assumption of OLS.

(e) Is there a notable violation of the homoscedasticity assumption?

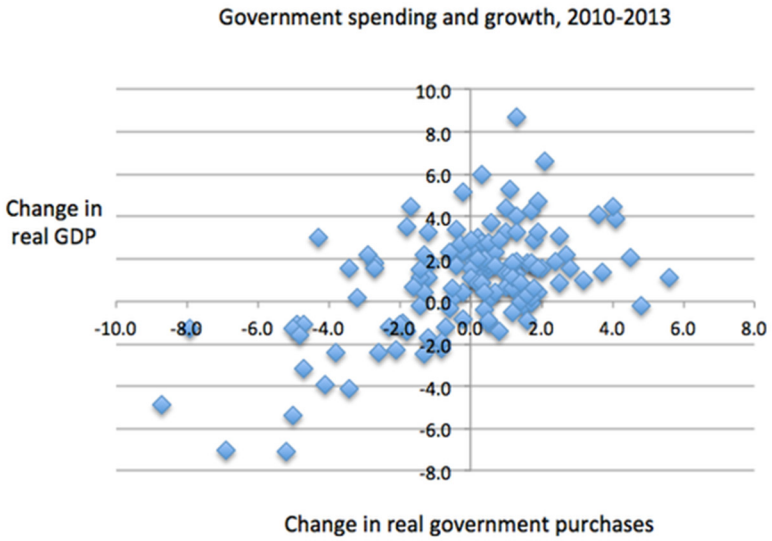
(f) What would the OLS regression results be if instead of a dummy variable for Waterloo we included a dummy variable for Ryerson?

(g) If we look at the 1,653 employees at York University, the mean salary is $154,762.3 and the standard deviation of salary is $42,017.82. If you ran a regression using the Waterloo and York employees, what would the OLS regression coefficients be for: Salary-hat = b0 + b1*Waterloo? Assume that salary is measured in dollars (not thousands of dollars like the previous parts to this question).

(6) In January 2015 one of Paul Krugman's columns in *The New York Times* is directly relevant to our course: "The Record of Austerity" posed on January 6, 2015 (http://krugman.blogs.nytimes.com/2015/01/06/the-record-of-austerity/). The column (and included graph) are reproduced below. It is a great opportunity to review concepts first discussed in Lectures 4 – 6 and the new concepts in Lecture 18 (and, in one case, Lecture 19).

> Let me jump right in here. How many people, I wonder — even among economists who have eagerly taken sides in the austerity debate — have a sense of what the overall picture looks like since the great turn to austerity in 2010? I don't mean what happened in country X in year Y, which you imagine supports your position; I mean the overall shape of events across many countries and multiple years.

Well, here's a quick and easy picture. I've taken annual data on the growth of real GDP and of government purchases from Eurostat, using every country for which data are available 2010-2013. I was tempted to edit out minor countries like Malta, but decided to do this as cleanly as possible. What we get are 33 countries for 4 years, 132 observations. And they look like this (bear in mind that these are percentage changes, so you can't read the slope of a trend line as a multiplier):



Does this picture make you think that Keynesian economics is nonsense? You can, if you like, argue that it's a spurious correlation for some reason. But surely the raw observations are consistent with the view that in depressed economies, cutting government spending hurts growth.

Of course, the fit isn't perfect. In fact, the R-squared is only 0.31. That's because in economics as in life, and as the bumper stickers don't quite say, stuff happens. And that is why we have statistics. Government spending only explains part of the variation in growth, but the t-statistic is 7.7; for the uninitiated, anything over around 2 is statistically significant at the 95 percent level.

As I said, you can, if you like, try to argue that this relationship is spurious, maybe not causal. But one form of argument that is really illegitimate is to comb through the data, pick out outliers, and claiming that the existence of these outliers — because stuff does, in fact, happen — disproves Keynesian logic. Unfortunately, you see a lot of that, including from economists who really should know better.

(a) What is the fundamental research question addressed in this column?

(b) Are the underlying data time series, panel (longitudinal) or cross sectional?

(c) If the x and y variables were both standardized, what would be the slope of the OLS line? Interpret it.

(d) Is the OLS slope economically significant?

(e) Is the OLS slope statistically significant? Significant? (Note: This refers to Lecture 19 material, but you can figure it out. t test statistics are evaluated like always. The hypotheses are about the slope: $H_0: \beta = 0$ and $H_1: \beta \neq 0$. As suggested in the article, this hypothesis test is conducted by using a t test statistic and the article gives the value of this statistic, which can be used to answer.)

(f) What is the *full* interpretation of the R^2?

(g) Go through all six underlying assumptions for OLS (see Lecture 18) and assess each in this specific example. (For Assumption #6, recall that spurious correlation is another way that people refer to endogeneity bias.)