ECON 256 – Correlations Regressions Professor Jack Rossbach

Problem 1 For this exercise, we will go back to the two data series

Data Series	Observation 1	Observation 2	Observation 3
X	15	15	30
Y	10	25	4

In last worksheet, we found that the correlation between X and Y was -0.72, indicating a relatively strong (although only because we have a small number of observations) linear fit.

Q1.1) Running an OLS regression returns an estimated slope coefficient of -0.9 and an estimated intercept of 31. That means we have the equation

$$y_i = 31 - 0.9x_i + \epsilon_i$$

What is the predicted value of Y when $X = 30$? What is the error for observation 3?

$$\hat{Y} = 31 - 0.9 \times 30 = 31 - 27 = 4$$

$$\epsilon_4 = y_4 - (31 - 0.9x_4) = 4 - (31 - 0.9 \times 30) = 4 - (31 - 27) = 4 - (4) = 0$$

Q1.2) Suppose X was to decrease from 15 to 5. How much would Y change, and in what direction?

$$\Delta Y = \beta(X' - X) = -0.9 \times (5 - 15) = -0.9 \times (-10) = 9 \Rightarrow Y \text{ will increase by 9}$$

Be careful with the minus sign in front of the β, and be careful to put the new value (5) first.

Q1.3) The standard error for this α estimate is 18.37 and the standard error for the β estimate is 0.87. Calculate approximate 95% confidence intervals for each estimate by adding/subtracting twice the standard error from each point estimate.

95 % confidence interval for α is approximately

$$[31 - (2 \times 18.37), 31 + (2 \times 18.37)] = [31 - 36.74, 31 + 36.74] = [-5.74, 67.74]$$

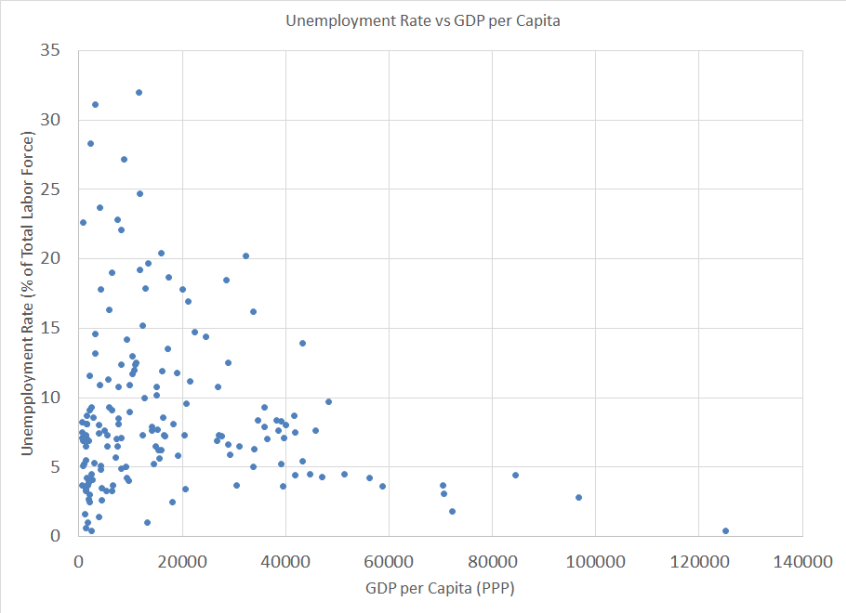
95 % confidence interval for β is approximately

$$[-0.9 - (2 \times 0.87), -0.9 + (2 \times 0.87)] = [-0.9 - 1.74, -0.9 + 1.74] = [-2.64, 0.84]$$

Q1.4) Are our estimated coefficients statistically significant at a significance level of 0.05?

No, because both confidence intervals contain zero.

Problem 2 Below is the Unemployment Rate vs GDP per Capita (PPP) in 2010 for all countries with data.



Last time we discovered the following correlations between GDP per Capita and the Unemployment Rate

Correlation for all Countries:
-0.165

Correlation for Rich Countries:
0.228

Correlation for Poor Countries:
-0.447

Cutoff between rich and poor was GDP per Capita of \$10k.

Q2.1) Fitting a regression line between GDP per Capita and Unemployment for rich countries we have the following best fit line:

$$\text{GDP} = 38847 - 1731 \times \text{Unemp}$$

What is the expected % change in GDP per Capita from decreasing Unemployment Rate from 10.5 to 10?

$$\%\ \text{Change in GDP} \equiv 100 \times \left(\frac{\text{GDP}'}{\text{GDP}} - 1\right)$$

$$\text{GDP} = 38847 - 1731 \times \text{Unemp} = 38847 - 1731 \times \mathbf{10.5} = 38847 - 18175.5 = 20671.5$$

$$\text{GDP}' = 38847 - 1731 \times \text{Unemp}' = 38847 - 1731 \times \mathbf{10} = 38847 - 17310 = 21537$$

$$\%\text{Change in GDP} = 100 \times \left(\frac{21537}{20671.5} - 1\right) = 100 \times (1.042 - 1) = 100 \times 0.042 = 4.2\ \%$$

Q2.2) We can also fit a regression line between GDP per Capita and Unemployment after taking logs:

$$\log \text{GDP} = 11.0861 - 0.4644 \times \log \text{Unemp}$$

The interpretation of the coefficient is how much a 1% increase in Unemployment increases GDP in %. An change from 10.5 to 10 is roughly a 5% decrease (percentage, **not** percentage points). According to this regression, how much should the % change in GDP be?

$$\%\text{Change in GDP} \approx -5 \times -0.4664 \approx 2.32$$

Get different answers, because the regression is not the exact same. The form that we run the regression in matters.

Additional Notes: For Q2.1 I wrote an incorrect best fit line. It should be

$$\text{GDP} = \mathbf{46372} - 1731 \times \text{Unemp}$$

With the correct best fit line we have the expected % change in GDP per Capita from decreasing Unemployment Rate from 10.5 to 10 is

$$\%\ \text{Change in GDP} \equiv 100 \times \left(\frac{\text{GDP}'}{\text{GDP}} - 1\right)$$

$$\text{GDP} = \mathbf{46372} - 1731 \times \text{Unemp} = \mathbf{46372} - 1731 \times \mathbf{10.5} = \mathbf{46372} - 18175.5 = 28196.5$$

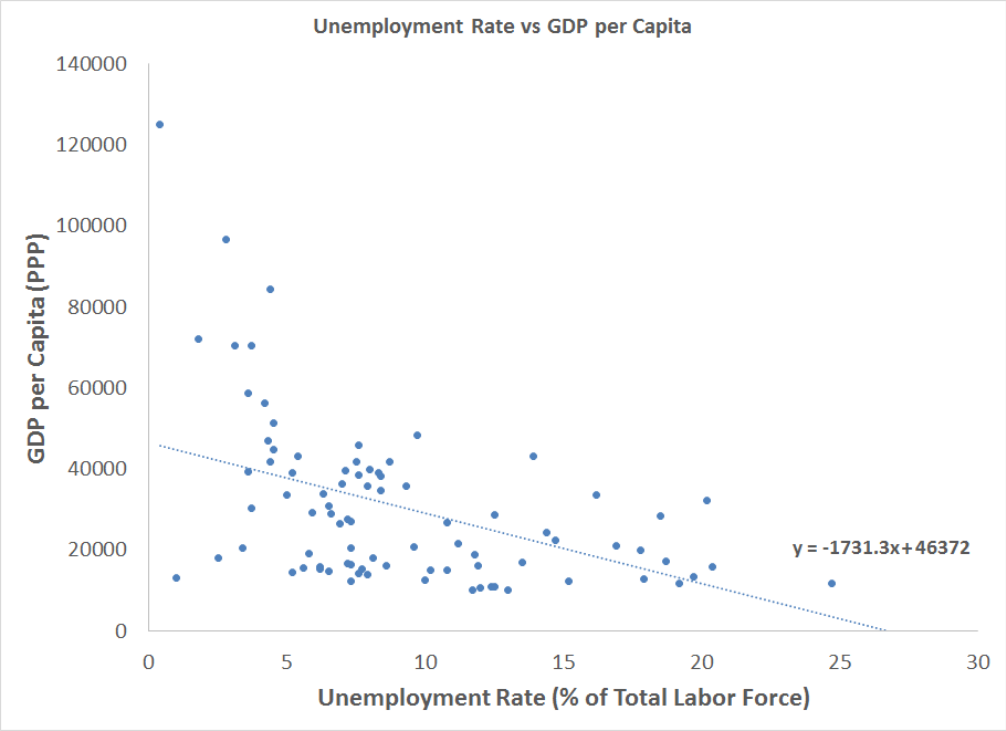
$$\text{GDP}' = \mathbf{46372} - 1731 \times \text{Unemp}' = \mathbf{46372} - 1731 \times \mathbf{10} = \mathbf{46372} - 17310 = 29062$$

$$\%\text{Change in GDP} = 100 \times \left(\frac{29062}{28196.5} - 1\right) = 100 \times (1.031 - 1) = 100 \times 0.031 = 3.1\ \%$$

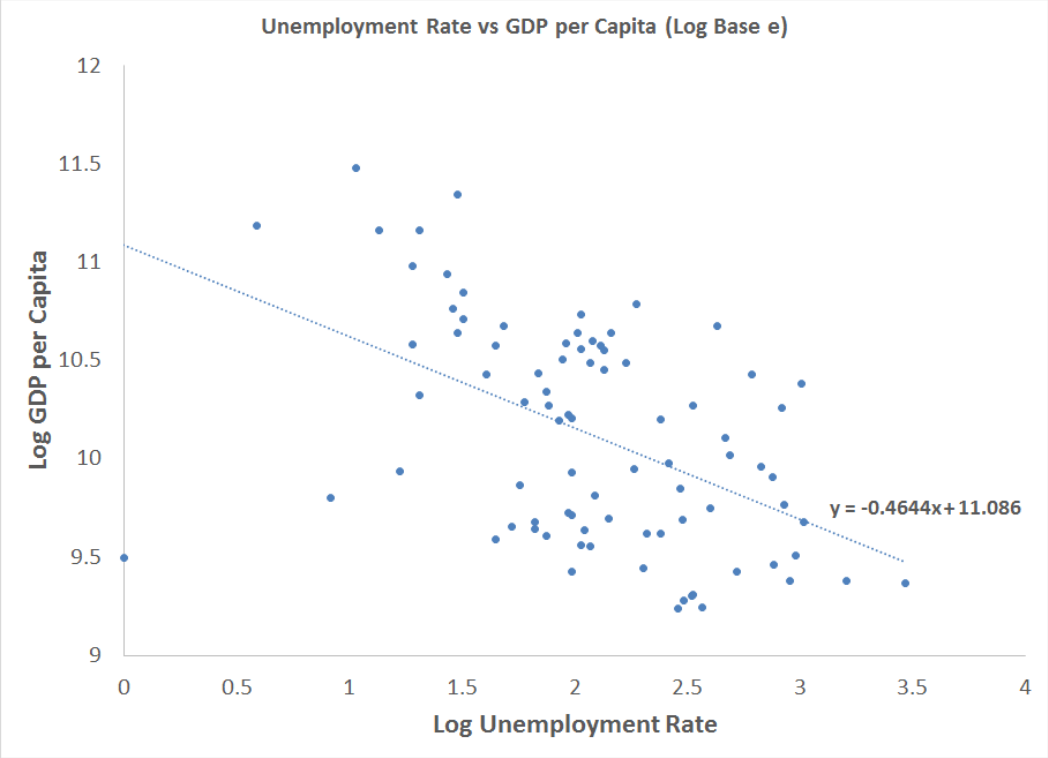
The reason I wanted to point out this typo is because I posted R code on my website for doing the regression and solving this problem on the computer, and this way the answers will align.

Additional Notes 2: Below are two charts. One showing the data and regression line for the data without any transformations other than selecting only countries with GDP per capita over $10,000. The second chart has both data series logged (natural log, base e). In the first chart we can see that errors are clearly bigger for countries with low unemployment rates. This is a problem, and can lead to incorrectly estimated regression coefficients. Taking logs seems to partially address this problem.

Linear Regression Fit



Log Regression Fit


Unemployment Rate vs GDP per Capita (Log Base e)
Log GDP per Capita
Log Unemployment Rate
y = -0.4644x + 11.086
12
11.5
11
10.5
10
9.5
9
0
0.5
1
1.5
2
2.5
3
3.5
4