Introductory Econometrics
Tutorial 4 Solutions

<u>PART A:</u> To be done before you attend the tutorial. The tutors will ask you questions based on this part and that will be the basis for your participation point. The solutions will be made available at the end of the week.

In this exercise, you need to continue working with WDI2019.xlsx data set that you used in tutorial 2.

1. This part is data cleaning, a very important step in the real world because in the real world data sets are messy. If you have any problems, please go to consultation hours of any member of the teaching team and sort it out.

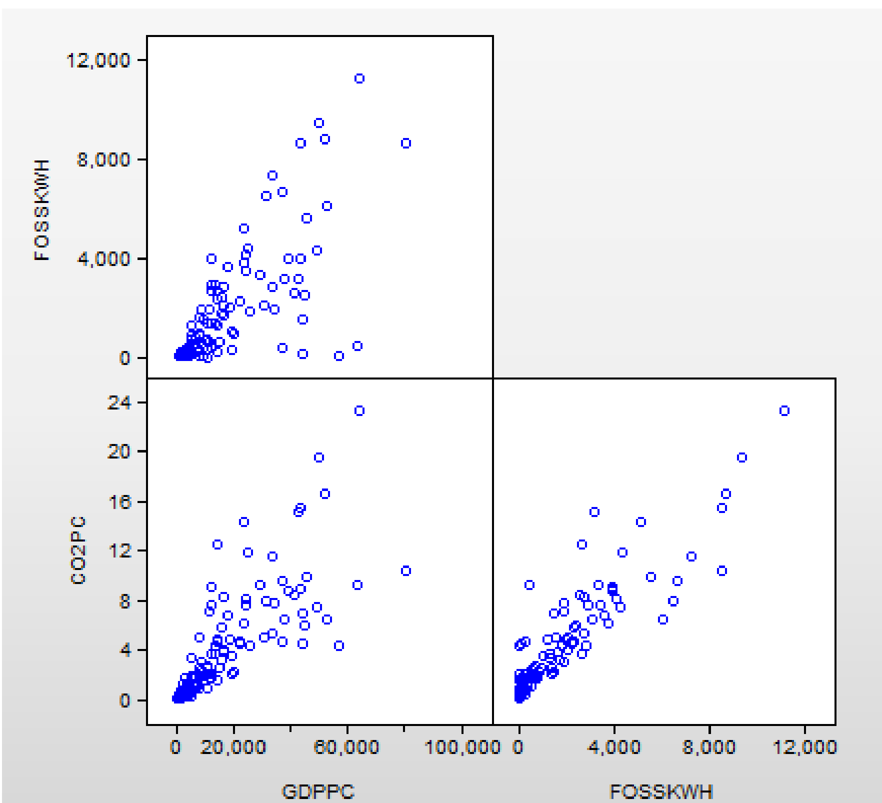
2. Same as 1.

3. Practice with the use of VLookup function in excel.

4. The top 5 countries in terms of electricity consumption per capita generated by fossil fuels are:

1	United Arab Emirates	11233
2	Saudi Arabia	9444
3	United States	8761
4	Singapore	8585
5	Australia	8576

5. All pairwise relationships are positive, which makes sense. The relationship between co2pc and fosskwh seem to be stronger than others, which also makes sense because electricity generation from fossil fuels produces CO_2 directly.



6. Sample correlation coefficients verify the information in the scatter plots.

Correlation			
	GDPPC	FOSSKWH	CO2PC
GDPPC	1.000000	0.715142	0.758696
FOSSKWH	0.715142	1.000000	0.891776
CO2PC	0.758696	0.891776	1.000000

7. Summary statistics based on a common sample:

	GDPPC	FOSSKWH	CO2PC
Mean	18828.16	1948.538	4.563131
Median	12715.97	1046.338	3.243288
Maximum	80305.45	11232.70	23.30202
Minimum	725.7301	0.000000	0.063369
Std. Dev.	17154.90	2374.063	4.524943
Skewness	1.227915	1.772917	1.589334
Kurtosis	3.907374	5.974838	5.849857
Jarque-Bera	29.98813	93.72385	79.73708
Probability	0.000000	0.000000	0.000000
Sum	1976957.	204596.5	479.1287
Sum Sq. Dev.	3.06E+10	5.86E+08	2129.412
Observations	105	105	105

PART B: You do not need to hand this part in. It will be covered in the tutorial. It is still a good idea to attempt these questions before the tutorial.

1. (*Post-multiplying a matrix by a vector produces a linear combination of the columns of the matrix):* Let

$$\mathbf{X} = \begin{bmatrix} 1 & 3 \\ 1 & 2 \\ 1 & 2 \\ 1 & 1 \end{bmatrix}$$

and

$$\widehat{\boldsymbol{\beta}} = \begin{bmatrix} 0.7 \\ 0.2 \end{bmatrix}.$$

Compute $\widehat{\mathbf{y}} = \mathbf{X}\widehat{\boldsymbol{\beta}}$, and show that the result is 0.7 times the first column of $\mathbf{X}$ plus 0.2 times the second column of $\mathbf{X}$.

- $\widehat{\mathbf{y}} = \begin{bmatrix} 1 & 3 \\ 1 & 2 \\ 1 & 2 \\ 1 & 1 \end{bmatrix} \begin{bmatrix} 0.7 \\ 0.2 \end{bmatrix} = \begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \end{bmatrix} \times 0.7 + \begin{bmatrix} 3 \\ 2 \\ 2 \\ 1 \end{bmatrix} \times 0.2 = \begin{bmatrix} 1.3 \\ 1.1 \\ 1.1 \\ 0.9 \end{bmatrix}$

2. Let's generalise the result in question 1. Suppose

$$\underset{n\times 3}{\mathbf{X}} = \begin{bmatrix} x_{11} & x_{12} & x_{13} \\ x_{21} & x_{22} & x_{23} \\ \vdots & & \vdots \\ x_{n1} & x_{n2} & x_{n3} \end{bmatrix}$$

and

$$\underset{3\times 1}{\widehat{\boldsymbol{\beta}}} = \begin{bmatrix} \hat{\beta}_1 \\ \hat{\beta}_2 \\ \hat{\beta}_3 \end{bmatrix}$$

Show that $\widehat{\mathbf{y}}= \mathbf{X}\widehat{\boldsymbol{\beta}}$ is an $n \times 1$ vector which is a linear combination (a weighted sum) of columns of $\mathbf{X}$ with weights given by the elements of $\widehat{\boldsymbol{\beta}}$. That is:

$$\widehat{\mathbf{y}}= \text{first column of } \mathbf{X}\times\hat{\beta}_1 + \text{second column of } \mathbf{X}\times\hat{\beta}_2 + \text{third column of } \mathbf{X}\times\hat{\beta}_3$$

In fact this is not specific to $\mathbf{X}$ having 3 columns. It is true for any $n \times k$ matrix $\mathbf{X}$ and $k \times 1$ vector $\widehat{\boldsymbol{\beta}}$.

- $\widehat{\mathbf{y}}= \begin{bmatrix} x_{11} & x_{12} & x_{13} \\ x_{21} & x_{22} & x_{23} \\ \vdots & & \vdots \\ x_{n1} & x_{n2} & x_{n3} \end{bmatrix} \begin{bmatrix} \hat{\beta}_1 \\ \hat{\beta}_2 \\ \hat{\beta}_3 \end{bmatrix} = \begin{bmatrix} x_{11}\hat{\beta}_1 + x_{12}\hat{\beta}_2 + x_{13}\hat{\beta}_3 \\ x_{21}\hat{\beta}_1 + x_{22}\hat{\beta}_2 + x_{23}\hat{\beta}_3 \\ \vdots \\ x_{n1}\hat{\beta}_1 + x_{n2}\hat{\beta}_2 + x_{n3}\hat{\beta}_3 \end{bmatrix} = \begin{bmatrix} x_{11} \\ x_{21} \\ \vdots \\ x_{n1} \end{bmatrix} \times\hat{\beta}_1 + \begin{bmatrix} x_{12} \\ x_{22} \\ \vdots \\ x_{n2} \end{bmatrix} \times\hat{\beta}_2 + \begin{bmatrix} x_{13} \\ x_{23} \\ \vdots \\ x_{n3} \end{bmatrix} \times\hat{\beta}_3$

3. *(Regression on a constant only - one way to compute the sample mean and sample variance of any variable using regression)*

$$y_i = \beta_0 + u_i, \qquad i = 1, \ldots, n.$$

(a) Use the OLS formula $\widehat{\boldsymbol{\beta}} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y}$ and show that the OLS estimator of the constant term in this case is the sample average of the dependent variable.

$$\begin{aligned} \mathbf{X}'\mathbf{X} &= \begin{bmatrix} 1 & 1 & \cdots & 1 \end{bmatrix} \begin{bmatrix} 1 \\ 1 \\ \vdots \\ 1 \end{bmatrix} = n \Longrightarrow (\mathbf{X}'\mathbf{X})^{-1} = \frac{1}{n} \\ \mathbf{X}'\mathbf{y} &= \begin{bmatrix} 1 & 1 & \cdots & 1 \end{bmatrix} \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix} = \sum_{i=1}^{n} y_i \\ &\Longrightarrow \widehat{\boldsymbol{\beta}} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y} = \frac{1}{n}\sum_{i=1}^{n} y_i = \bar{y} \end{aligned}$$

(b) What is $\hat{u}_i$ (the residual for each observation) in this case? Show that $\mathbf{X}'\hat{\mathbf{u}} = 0$ in this case is the same as $\sum_{i=1}^{n}(y_i - \bar{y}) = 0$, a result that we knew already.

$$\hat{u}_i = y_i - \hat{\beta}_0 = y_i - \bar{y}$$

which is the deviation of each observation from the sample mean.

$$\begin{aligned}
\mathbf{X}'\hat{\mathbf{u}} &= \begin{bmatrix} 1 & 1 & \cdots & 1 \end{bmatrix} \begin{bmatrix} y_1 - \bar{y} \\ y_2 - \bar{y} \\ \vdots \\ y_n - \bar{y} \end{bmatrix} \\
&= \sum_{i=1}^{n} (y_i - \bar{y}) = 0
\end{aligned}$$

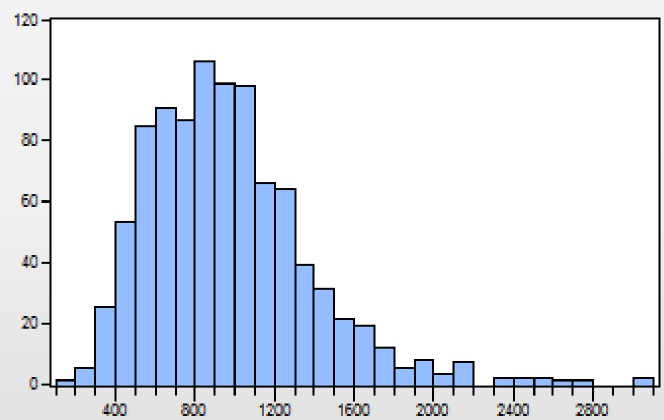
which we knew already because $\sum_{i=1}^{n} (y_i - \bar{y}) = \sum_{i=1}^{n} y_i - \sum_{i=1}^{n} \bar{y} = n\bar{y} - n\bar{y} = 0$.

(c) What is $\hat{\sigma}^2$ the estimator of the variance of u_i in this case? What is the standard error of regression in this case?

$$\hat{\sigma}^2 = \frac{1}{n-1} \sum_{i=1}^{n} \hat{u}_i^2 = \frac{1}{n-1} \sum_{i=1}^{n} (y_i - \bar{y})^2$$

which is the sample variance of y. The standard error of regression is the square root of this, which would be the sample standard deviation of y. Hence, this gives one way of getting sample mean and sample standard deviation using a regression. (This also automatically produces the standard error for the sample mean.)

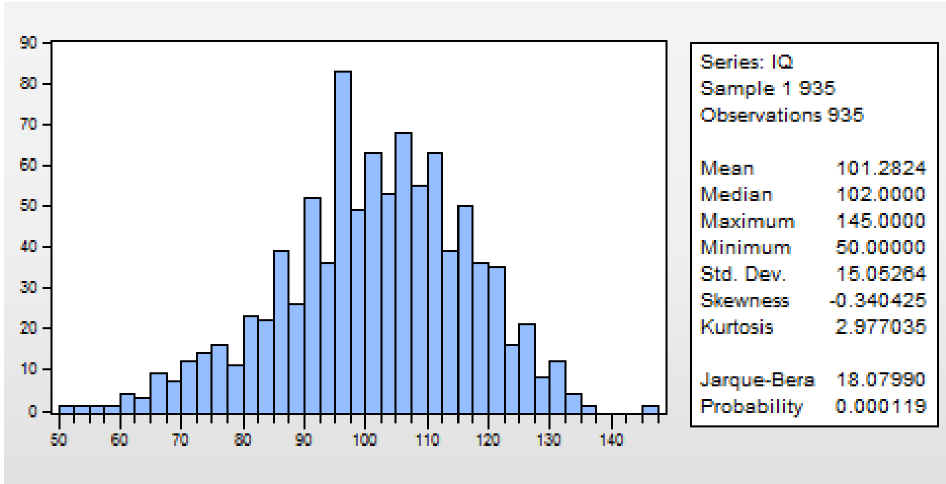
4. This question is based on question C4 in Chapter 2 of the textbook. The dependent variable is *wage* and the independent variable is IQ.
Preliminary analysis of the data (not asked in the question, but a step that we usually take before estimating a regression equation: "Look" at these variables (meaning that examine their histograms, summary statistics, scatter plot of wage against IQ, sample correlation coefficient between wage and IQ, and summarise your insights from just looking at data through these views).



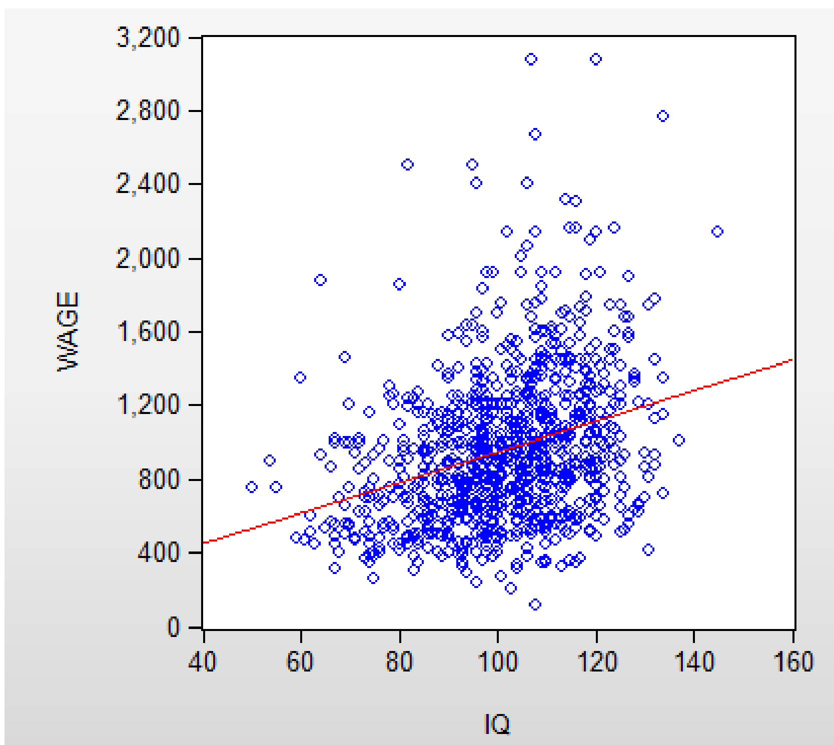
Series: WAGE	
Sample 1 935	
Observations 935	
Mean	957.9455
Median	905.0000
Maximum	3078.000
Minimum	115.0000
Std. Dev.	404.3608
Skewness	1.199259
Kurtosis	5.696661
Jarque-Bera	507.4271
Probability	0.000000

Wage is positively skewed. The sample mean is \$958, but because of this skewness the median \$905 is a better indicator of the central tendency than the mean. Need to make sure that the one outlying observation with \$3078 wage does not affect the analysis.

IQ seems to be representative, with mean close to 100 and standard deviation close to 15. There is a very smart person with IQ of 145, so need to ensure that it does not influence the analysis too much.



The scatter plot of wage against IQ, in particular with the regression line included, shows that the relationship between wage and IQ, although seems positive (sample correlation coefficient is 0.309), is not linear. The variation of wage around the regression line seems to be higher at higher IQs.



Sample: 1 935
Included observations: 935

Correlation	IQ	WAGE
IQ	1.000000	
WAGE	0.309088	1.000000

(a) *Verify question 3 with real data:* Run a regression of *wage* on a constant only. Verify that the OLS estimate of the intercept is the sample mean of *wage* and the standard error of regression is the sample standard deviation of *wage*.

Dependent Variable: WAGE
Method: Least Squares
Sample: 1 935
Included observations: 935

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	957.9455	13.22401	72.43985	0.0000

R-squared	0.000000	Mean dependent var	957.9455
Adjusted R-squared	0.000000	S.D. dependent var	404.3608
S.E. of regression	404.3608	Akaike info criterion	14.84356

(b) *Estimation, interpretation of the slope coefficient and R^2 of the regression*: Estimate a simple regression model where a one-point increase in *IQ* changes *wage* by a constant dollar amount. Use this model to find the predicted increase in *wage* for an increase in *IQ* of 15 points. Does *IQ* explain most of the variation in *wage*? What is the relationship between the R^2 of this regression and the sample correlation coefficient between wage and IQ? Name your estimated equation **eq01.** Save the residuals of this regression in a variable called **uhat01.** points. Does *IQ* explain most of the variation in *wage*? Name your estimated equation **eq01.** Save the residuals of this regression in a variable called **uhat01.**

- Dependent Variable: WAGE
  Method: Least Squares
  Date: 07/28/16 Time: 12:46
  Sample: 1 935
  Included observations: 935

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	116.9916	85.64153	1.366061	0.1722
IQ	8.303064	0.836395	9.927203	0.0000

R-squared	0.095535	Mean dependent var	957.9455
Adjusted R-squared	0.094566	S.D. dependent var	404.3608
S.E. of regression	384.7667	Akaike info criterion	14.74529
Sum squared resid	1.38E+08	Schwarz criterion	14.75564
Log likelihood	-6891.422	Hannan-Quinn criter.	14.74924
F-statistic	98.54936	Durbin-Watson stat	1.802114
Prob(F-statistic)	0.000000		

If *IQ* increases by 15 points (*i.e.* one standard deviation), the predicted increase in wage is $8.303 \times 15 = 124.545$ (get students to use their calculators).
The R^2 is very low, so there is a lot of unexplained variation. Note that in a regression with only one explanatory variable, R^2 is the square of sample correlation coefficient between the dependent variable and the explanatory variable. Ask them to verify that on their calculators $0.309088^2 = 0.095535$.

(c) Interpretation of the intercept: What does the intercept in eq01 mean? Now, run a regression of wage on a constant and (IQ-100), and name it eq02. Compare the results with your results in eq01 and note all similarities and differences. Save the residuals of this regression in a variable called uhat02. Open uhat01 and uhat02 side by side and see if they are different. What is the interpretation of the intercept in eq02?

- Intercept in eq01 does not have any meaningful interpretation because there is no individual with IQ of zero.

Dependent Variable: WAGE
Method: Least Squares
Date: 07/28/16 Time: 12:48
Sample: 1 935
Included observations: 935

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	947.2980	12.62885	75.01066	0.0000
IQ-100	8.303064	0.836395	9.927203	0.0000

R-squared	0.095535	Mean dependent var	957.9455
Adjusted R-squared	0.094566	S.D. dependent var	404.3608
S.E. of regression	384.7667	Akaike info criterion	14.74529
Sum squared resid	1.38E+08	Schwarz criterion	14.75564
Log likelihood	-6891.422	Hannan-Quinn criter.	14.74924
F-statistic	98.54936	Durbin-Watson stat	1.802114
Prob(F-statistic)	0.000000		

Only the intercept has changed. The estimate of the slope, its standard error, R-squared and the standard error of regression are all exactly the same as in eq01. The intercept, however, is now meaningful. It shows the predicted wage for a person with IQ of 100, i.e. average IQ.

(d) Discuss what you learned from this exercise.

- Among other things, we learn that if we subtract a constant from one of the explanatory variables, only the OLS estimator of the intercept will change. Everything else, in particular the estimate of the slope, its standard error, the residuals and the predicted wage for all observations will be exactly the same as before. Geometrically, this is because when we add or subtract a multiple of one column to another column of a matrix $\mathbf{X}$ its column space does not change (we are sliding one vector up or down another vector). So, the orthogonal projection of wage on this space will stay the same as before. We can also predict how the intercept will change because:

$$\begin{bmatrix} 1 & IQ_1 \\ 1 & IQ_2 \\ \vdots & \vdots \\ 1 & IQ_n \end{bmatrix} \begin{bmatrix} b_1 \\ b_2 \end{bmatrix} = \begin{bmatrix} 1 & IQ_1 - 100 \\ 1 & IQ_2 - 100 \\ \vdots & \vdots \\ 1 & IQ_n - 100 \end{bmatrix} \begin{bmatrix} b_1 + 100b_2 \\ b_2 \end{bmatrix}$$

for any $\begin{bmatrix} b_1 \\ b_2 \end{bmatrix}$. So, by observing eq01, we could have guessed that the estimate of intercept in eq02 was going to be $116.9916 + 100 \times 8.303064 = 947.298$ exactly!