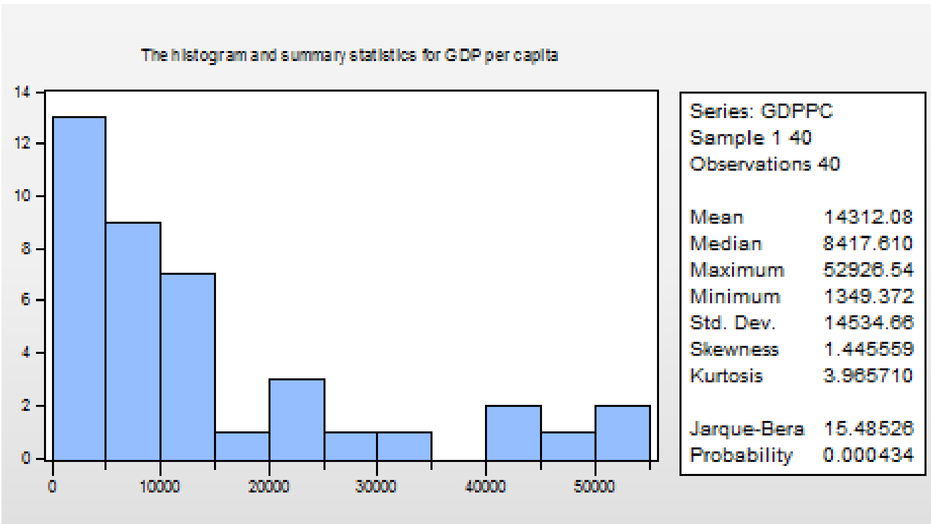
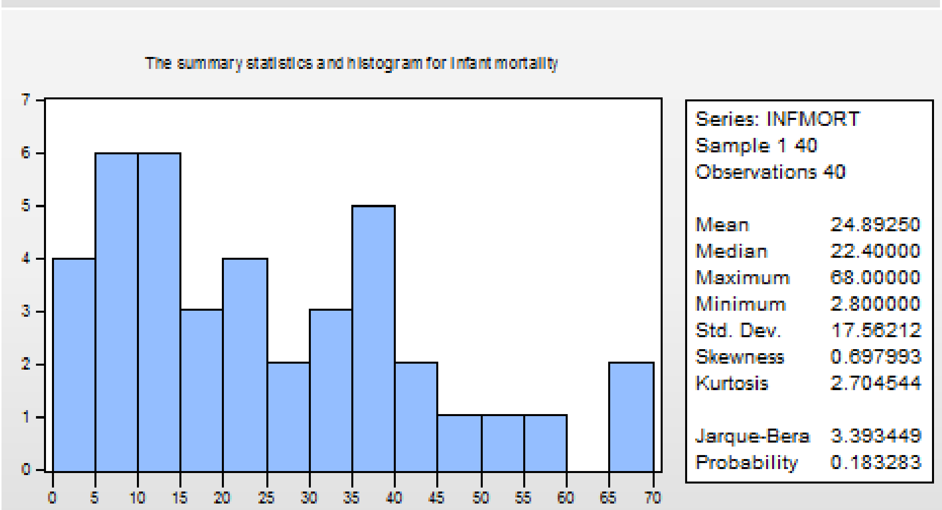
Introductory Econometrics

Tutorial 2

<u>**PART A:**</u> **To be done before you attend the tutorial, and handed in to your tutor at the beginning of the tutorial to earn 1 point. You may write the answers by hand. The solutions will be made available at the end of the week.**

1. Obtain the summary statistics of GDPPC and INFMORT in EViews, print those and hand them to your tutor in your tutorial to obtain the participation point for week 2. Save the EViews workfile.





END OF PART A.

PART B: To be done in the tutorial.

Question 1: *Important topics of the first year statistics: Histograms, scatter plots, confidence intervals + remembering logarithms*: Open the EViews workfile that you generated in Part A. If you have not saved it or have difficulty finding it, download "A random sample of countries.wf1" from the moodle site and use that.

1. Obtain the summary statistic and histogram of GDPPC. Discuss what you can learn from the histogram and summary statistics. If you had a different set of 40 countries, would the summary statistics be the same?

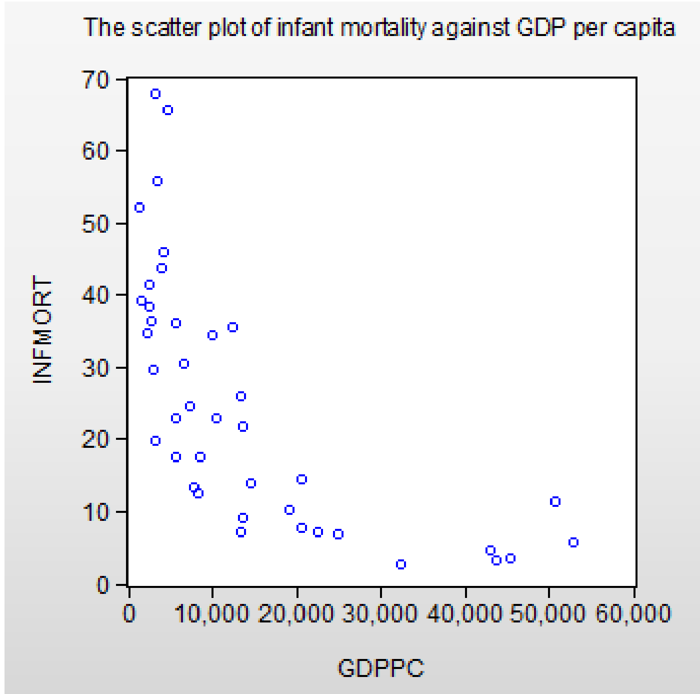
   Answer: Things that they may notice: The sample distribution is not symmetric (and hence not normal). The shape and also the fact that the mean is much higher than the median shows that the distribution is skewed to the right. Huge inequality among these 40 countries. The more descriptive the explanation, the better (i.e. instead of just reading that the minimum is 1349.372, it is better if they say the minimum GDP per capita is \$1349, which is less than 1/6 of the median in this sample, or which translates to \$3.70 per day per person. In contrast the maximum is 6 times as large as the median, which translates to \$145 per day per person. Please encourage them to try to "feel" the numbers. These are about GDP per capita, so each person should be able to feel what \$3.70 a day feels like. - Make sure that everyone understands that sample average is a random variable and sample mean will be a different number for a different set of 40 countries.

2. Using the sample average and the sample standard deviation, compute the 95% confidence interval for the population mean of GDPPC (if you have forgotten the formula for the confidence interval for the population mean, refer to equation (11.1) on page 66 of ETC1000 lecture notes that are on moodle (under Extra Study Materials related to tutorial 1), or equation [C.23] in Appendix C of the textbook. It is OK to use the rule of thumb given in equation [C.26] of the textbook. Explain what this confidence interval shows. The average GDPPC of *all* countries in WDI data base in 2015 is \$17406. Does your confidence interval contain this value?

$$\begin{aligned} & \text{Using } \pm 2sd \text{ rule of thumb, the 95\% confidence interval is:} \\ & \left[14312.08 - 2 * 14534.66/\sqrt{40},\ 14312.08 - 2 * 14534.66/\sqrt{40}\right] \\ = & \ [9716, 18908] \end{aligned}$$

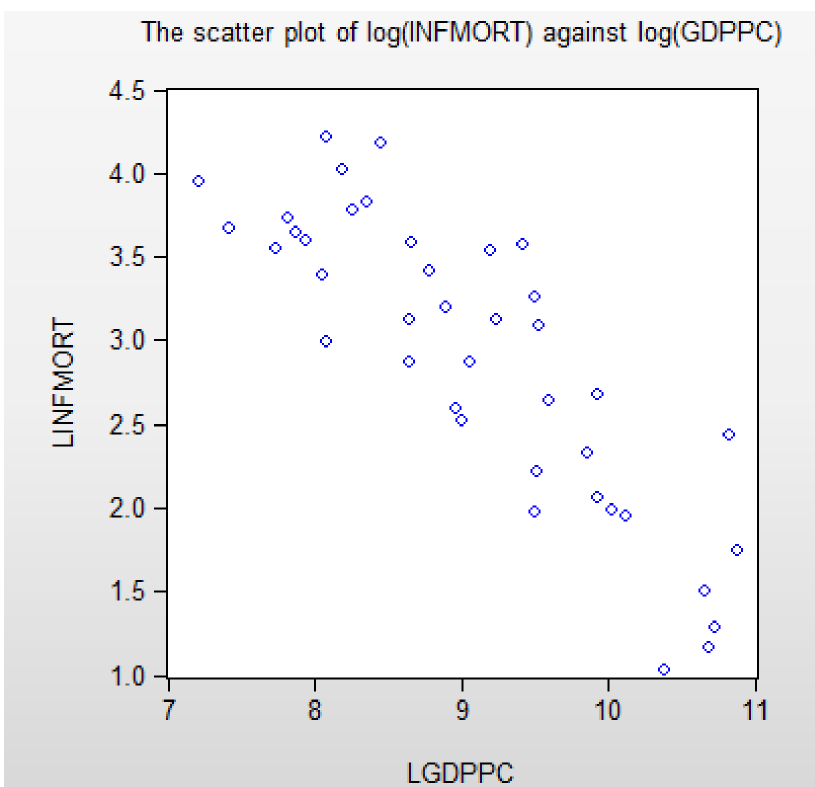
   It means that there is a 95% chance that this interval contains the population mean of GDPPC. Try to explain that as we obtain different samples, these bounds move around, whereas population mean is a constant number. As a result, saying that "95% of the times population mean is between 9716 and 18908" is an incorrect interpretation. Some may have to think about it in their own time to understand this, and each student learns this at their own time (hopefully). Here, this interval contains \$17406, which is the mean GDPPC of all countries listed in WDI. Some students might legitimately object to using a formula for confidence interval that is based on normality of the parent population, which in this case is clearly not normal. Please acknowledge that this is a legitimate and excellent question, but let them know that later in the unit we will learn that with large enough samples, the distribution of the sample mean is close to normal even if the parent population is not normal. (You may even want to replay the sampling distribution applet that is in the second week's lecture slides).

3. We want to explore the association between infant mortality and GDP per capita. What kind of a graph can give us an insight into the nature of this relationship? Based on this graph, are infant mortality and GDP per capita positively or negatively correlated? Is their relationship linear?



The plot shows that they are negatively correlated, which makes sense. Also, the relationship is not linear.

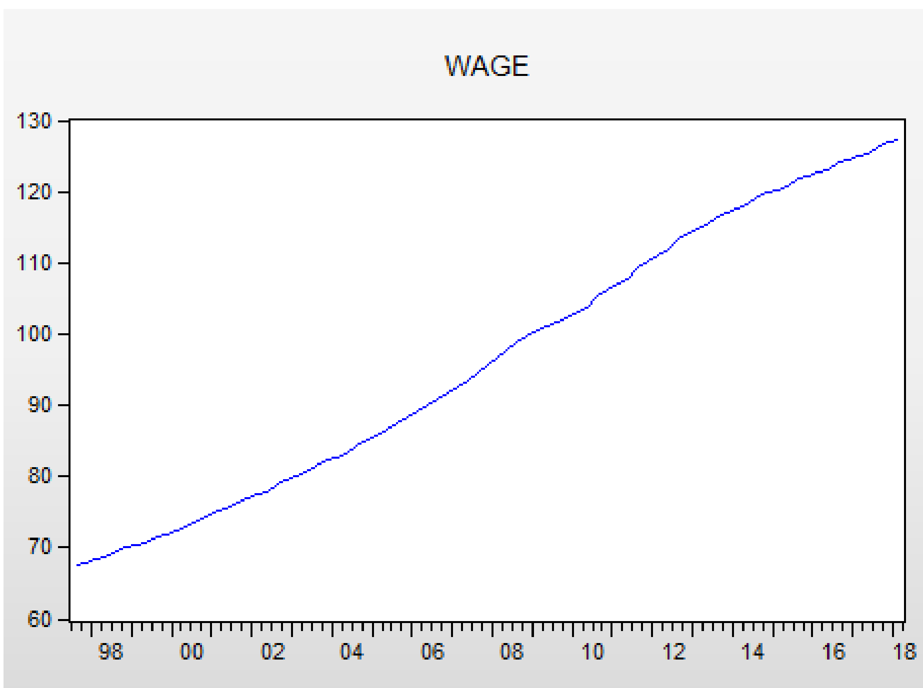
4. Generate the logarithmic transformation of GDPPC and INFMORT. Use the label LGDPPC for the natural logarithm of GDPPC and LINFMORT for the natural logarithm of INFMORT. Note: In EViews, log(X) calculates the natural logarithm of X. EViews does not recognise ln(X). Explore the association between LINFMORT and LGDPPC with the aid of an appropriate graph. Compare and contrast it with what you got in the previous part.



In contrast with the previous scatter plot, this one shows that the relationship between log(INFMORT) and log(GDPPC) is linear (or a linear model will be more appropriate after the variables have been transformed). You can explain that this is a very useful transformation, and one way that we can accommodate nonlinear relationships using tools that we already know (like scatter plots) and ones we learn in this unit.

Question 2: *Working with time series: plots, trends, seasonality, growth rate (log-returns):* Download hourly wages from www.ausmacrodata.org-> Categories -> wage price index -> the first series that shows up -> Download CSV. Open the CSV file, and tidy up the data set, i.e. only keep the first two columns and delete everything else, and give a better name for the second column, such as WAGE. Save the CSV file and read it in EViews. Note that EViews automatically realises that you have quarterly data.
About ausmacrodata.org: This website was created by me and two other staff members and an honours student, with the financial support of the Australian Research Council. It automatically updates itself at 3 am every day by crawling over the ABS and Reserve Bank data websites. Students may like this story because it tells them how close they are to being able to produce something that can be used by all researchers who want to work on Australian macro economy.

1. Plot the WAGE series (plotting a time series means producing a line plot of the series in which the x-axis is time. In EViews, clicking on a series opens a window that shows the value of the series in a spreadsheet. This window has a menu bar. Under View is Graph. And the default for Graph is a line plot). What can we learn from this plot?



   The plot shows that hourly wage has been growing steadily in this period (positive or upward trend). Explain this means high persistence (each observation very close to observation before it). Although not so clear, some may see accelerating growth in the first half and decelerating growth towards the end.

2. If we were interested in forecasting hourly wage in the next period, would sample average be a good forecast? Suggest more appropriate forecasts.

   No, sample mean will not be of much use here. The population mean is changing over time and increasing with time. We can use the last observation as the forecast for the next period, but we can do better than that. For example, we can compute the average quarter on quarter growth of wage over the sample period, and assumes that wages grow by that amount next quarter and make a better forecast based on the end quarter and this growth rate. We can even do better than that given what we learn in part 3.

3. In most financial or economic time series, trend is so dominant that it is the only thing that we can immediately see and all other aspects of the time series are dwarfed by its trend. To see other aspects of the series we have to remove its trend. One way is to make a model with time as an explanatory variable (we will do this later in the course). Another way is to compute the growth rate, in this case $g_t = 100 \times \frac{wage_t - wage_{t-1}}{wage_{t-1}}$ (multiplication by 100 is just to express it in percentage points). A more prevalent way of calculating the growth rate, in particular in finance, is to use what is known as "log-returns"

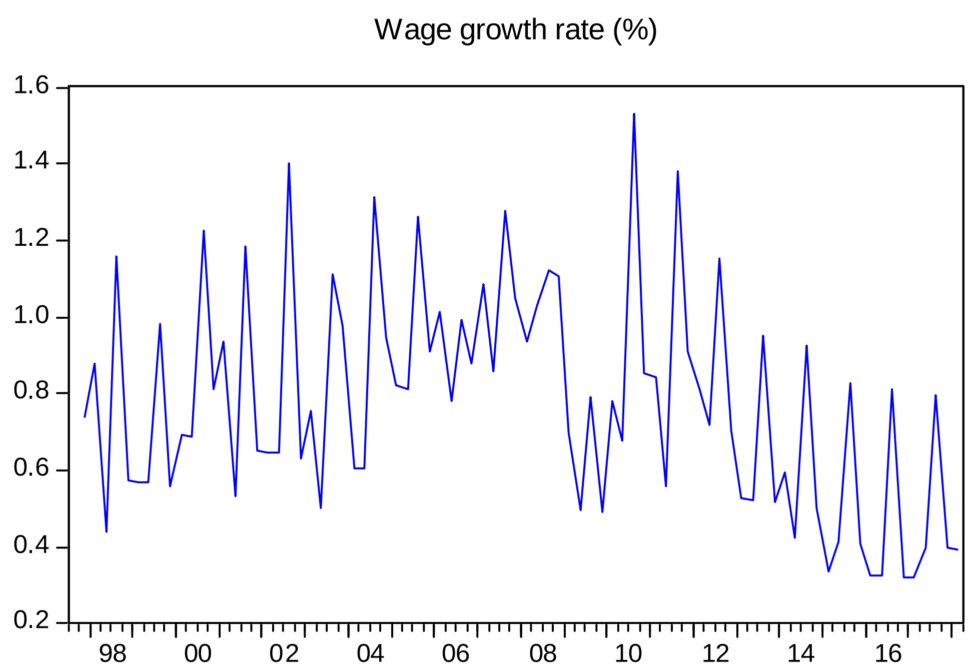
$$g_t = 100 \times \Delta \log(wage_t) = 100 \times (\log(wage_t) - \log(wage_{t-1}))$$

These two methods of calculating the growth rate produce values that are close to each other as long as growth rate is less than 10% in absolute value. EViews has a built in function 'dlog(X)' that computes the difference of logarithm of X. Generate the growth rate of hourly wage using the log-returns formula. Open this series. Why is the first value of this series NA?

	Modified: 9/01/1997 3/01/2018 // g=100*dlog(wage)
9/01/1997	NA
12/01/1997	0.739102
3/01/1998	0.879771
6/01/1998	0.437000
9/01/1998	1.156082

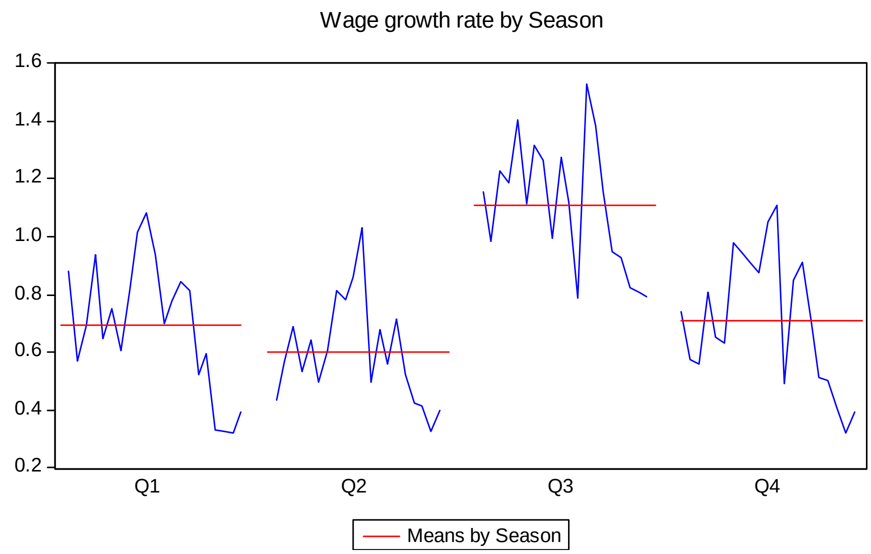
Because to compute $100 \times (\log(wage_t) - \log(wage_{t-1}))$ when t is the beginning of the sample (1997 Q3), we need wage data for the previous period (1997 Q2), which we do not have.

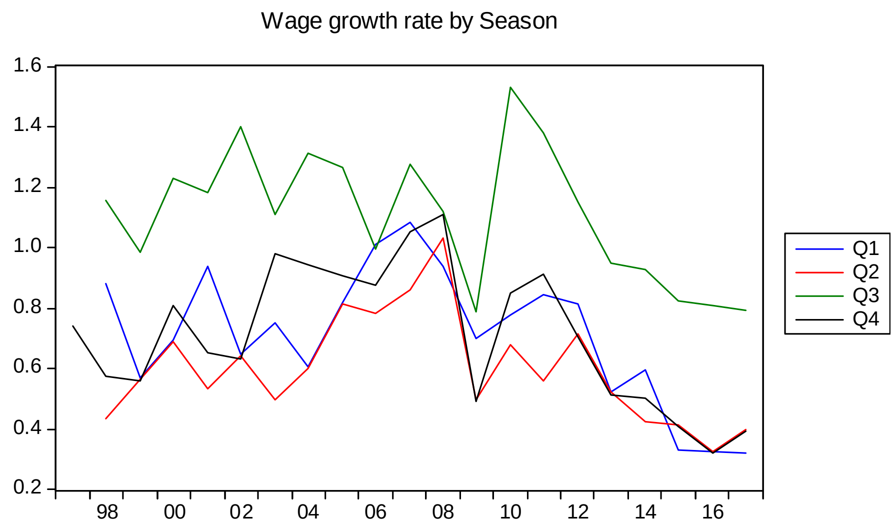
4. Plot the growth rate of wage. What does this plot tell you?



There is a clear seasonality (intra-year pattern) in the data. If they hover the mouse over the peaks, they can see that most (if not all) correspond to the 3rd quarter of each year. This may be due to the fact that most people's wage increases become effective at the beginning of the financial year. This is my guess.

5. Look at the seasonal plots of the growth rate of wage. To produce seasonal plots in EViews, in the series window, View -> Graph, and choose Seasonal Graph (the last option under Graph type). There are two seasonal plots: one that plots each season in a different panel side by side and shows average growth rate for each season. Another type shows four line plots, one for each season, overlayed on one graph. Look at both plots and discuss what you learn from these plots.





Both plots show that wage growth in Q3 is higher than in other quarters.

6. Look at the time series plot of the growth rate of wage again. Mentally adjust for seasonal variation. Do you see that the wage growth has been declining since 2010? Should that by itself worry us? What else do we need if we are worried about the value of one hour of work?

The point here is that these are nominal wages. The real value of work depends on how much we can buy with our wages. If nominal wage growth is lower but inflation is lower as well, the real wage can be lower, steady or higher depending on the relative magnitudes of nominal growth rate and inflation rate. So, we need to look at the inflation rate to determine what has happened to the real wage growth.