**Tutorial 6**

Software required:   Excel

Load the Excel add-in "Analysis ToolPak" following the procedure in http://www.excel-easy.com/data-analysis/analysis-toolpak.html

Qn 1

The Excel file "Final Marks.xlsx" contains the final marks of an EE course with 95 students.

a)  Find the (population) mean, median, mode, standard deviation, and variance.

(The relevant functions are: AVERAGE, MEDIAN, MODE, STDEV.P STDEV.S, VAR.P, VAR.S)

What is the difference between the .P and .S versions?

b)  Use the mark-to-grade conversion table in the Appendix. Plot the distribution of grades in a histogram following the procedure in http://www.excel-easy.com/examples/histogram.html Give it the proper labels, legends and title.

c)  Compute the skew of the distribution using the Excel function SKEW (http://www.excelfunctions.net/Excel-Skew-Function.html)

(Note:  You should use the raw data when computing SKEW)

Do you think the distribution is normal?

(A rule of thumb is to compute the skew of the distribution. If the absolute value of skew > $2\sqrt{6/n}$, then the skew is significant and the distribution is not normal)

d)   Whether a distribution is normal can be determined by a "normality test".  Give an example of normality test.  (Search the web for the answer.)

Qn 2

Take a random sample of 10 students.

a)  Find the (sample) mean, median, mode standard deviation, and variance.
b)  Plot the distribution of grade in a histogram.  Is there any difference compared with the shape in Qn 1 b)?  Discuss.

Qn 3

Download the closing price of the Hong Kong stock 0002.HK (CLP Holdings Limited), 0005.HK (HSBC Holdings plc), 0011.HK (Hang Seng Bank) from 1$^{st}$ January 2016 to 31$^{st}$ January 2016. The data can be found fromYahoo Finance (https://finance.yahoo.com/).

An example of downloading data from Yahoo Finance:
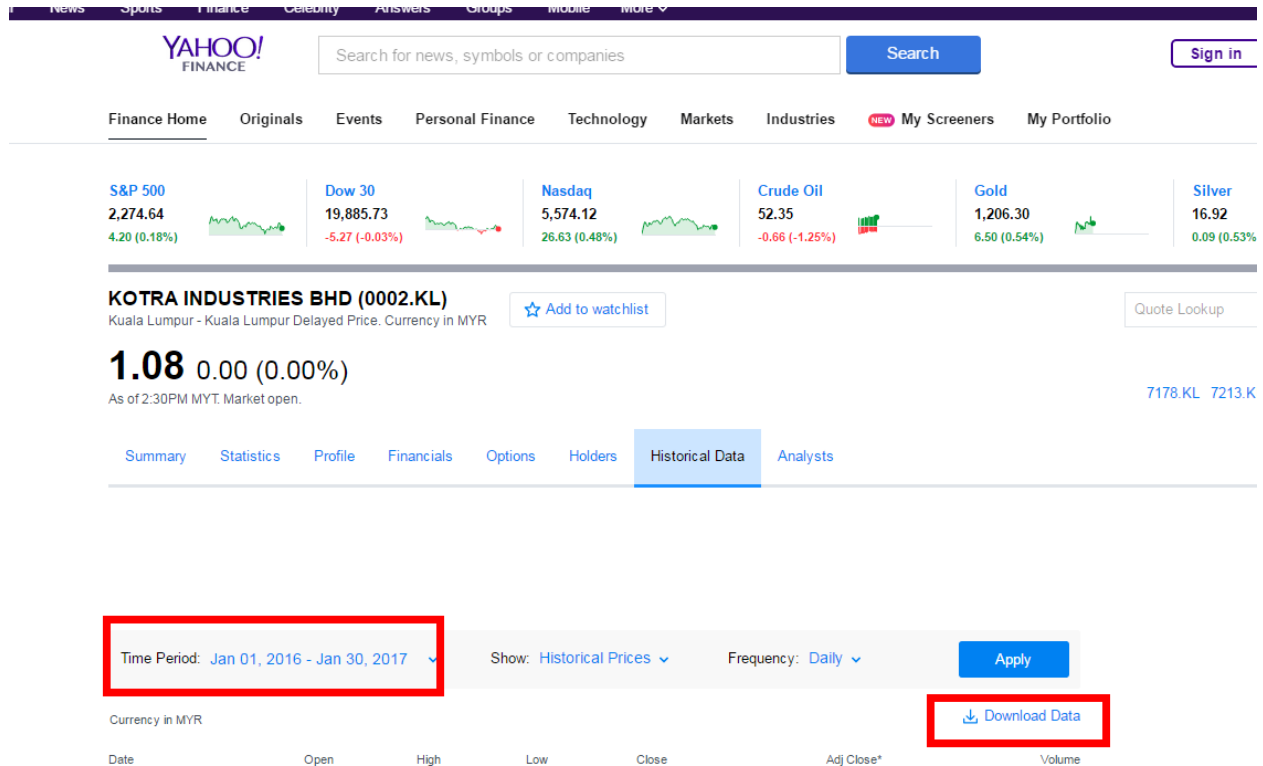


The price data is also available in the Excel file "Price_data.xlsx"

a) Plot the three time series data in one graph. (Use insert button →line chart →select line chat type)

b) Compute the Pearson coefficient between 0005 and 0011 using the Excel function Pearson.

c) Compute the Pearson coefficient between 0002 and 0005.

d) Discuss the meaning of the values of the Pearson coefficients computed in b) and c). What are the assumptions behind?

Qn 4

Plot the Poisson distribution with $\lambda = np$ and the binomial distribution $(n, p)$ when $n$ is large and $p$ is small. (For example, use $n = 2400, p = 0.001$)

What do you observe? Does it agree with theoretical predictions? Why?

Note:

A binomial distribution can be plotted by first using the Excel function BINOMDIST to give the values at different $x$. Then plot a bar chart.

Likewise, a Poisson distribution can be plotted using the Excel function POISSON.

You can also plot other probability distributions using Excel. See

http://www.icse.xyz/msor/psme/standard.html

**Appendix**

Mark-to-Grade Conversion Table

| Final Marks | Letter Grade |
|---|---|
| 80-100 | A+ |
| 75-79 | A |
| 70-74 | A- |
| 65-69 | B+ |
| 60-64 | B |
| 55-59 | B- |
| 50-54 | C+ |
| 45-49 | C |
| 40-44 | C- |

| | |
|---|---|
| 30-39 | D |
| 29 or below | F |