Student Name: METU Dept. Of Business Administration
Student Number: BA4318 Spring 2019

TEST 02 / 17.05.2019

Part 2 (70 Points, take-home)

You must take Part 1 of the exam in order to submit this part

**Question 3 (Data processing and analysis, 70 points)**

Imagine, as a junior analyst in one of the big consultancy companies, you are conducting a strategic management projects and you need to analyse the relationships of some products that ALLGOODS IMPORT EXPORT company trades in.

You are given the task of analysing a particular country where ALLGOODS sells cocoa, coffee, copper, cotton, orange juice, and sugar in bulk. The unit of sales is million kilograms. You recognize some patterns in sales and try to find out why. Then one of your colleagues suggest that some of these producst may be substitutes or complementaries.

After some googleing, you brush up your ECON101 knowledge such as:

- "The demand for substitute products shows a negative correlation. That is, consumption of one product reduces or replaces the need for the other. However, the demand and pricing of substitute products exhibit a positive correlation. It means if the price of one product increases, the demand for the other increases. "
- "Two goods that are substitutes show a positive cross elasticity. It means that as the price of product x rises, the demand for the other product rises"
- "Two goods that complement each other exhibit negative cross elasticity. The increase in the price of one product causes a drop in the quantity demanded of the other product."
- "Two goods that are neither complementary nor substitute and are independent of each other show zero cross elasticity. The change in the price of one product does not affect the other product pricing, and it remains constant."
- "Imperfect substitute products are the ones that although they can be replaced/substituted with each other, there is a probability there are those who will stick to the other product regardless of other factors."

You ask for sales data and you get the attached file for more than one year of sales trade data in the commodity exchanges of that particular country, including price and volume.

You also Google for how to calculate correlations and you are amazed to see that a DataFrame object simply gives simple correlations a covariances by

```
print( df.corr() )

print( df.cov() )
```

For any detailed analysis, such as simple linear regression, there is the scikit-learn library, but that needs a bit of reading

> https://scikit-learn.org/stable/modules/linear_model.html

Your tasks:

1. Read data from multiple files, and **merge necessary columns into a single dataframe**. (20 points)
2. Create and print simple correlation matrix to identify perfect substitute and complementtary products. (30 points)
3. Conduct linear regression to analyse imperfect substitute and complementary products. (20 points).

**Good Luck and Have a Great Summer!**