

Exercise class 5

(week 12)

Introduction to Programming and Numerical Analysis

Class 4 and 8

Rosa Haslund Meyer
Spring 2024

KØBENHAVNS UNIVERSITET



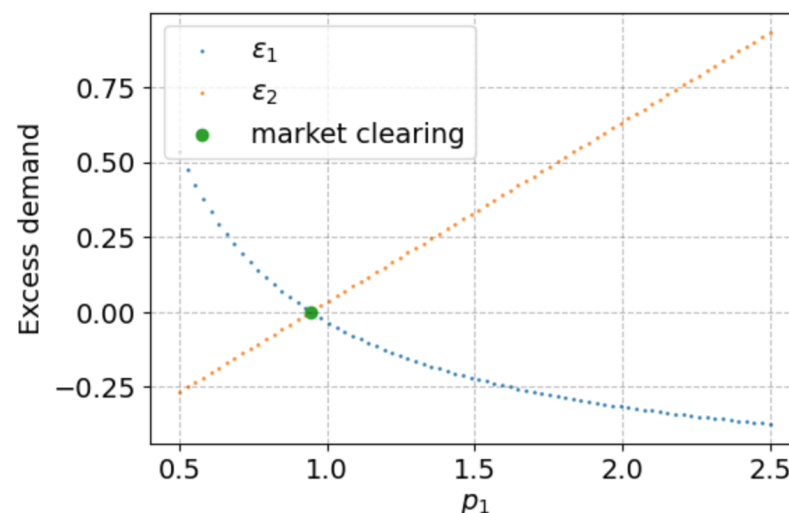
Short comment on Inaugural Project

Data basics and analysis in Python

Problem set 3. Loading and structuring data from Denmark Statistics

Answer to question regarding Q2/3 in the Inaugural Project

Some of you noticed that when plotting excess demand for good 1 and 2 as functions of p_1 , one of the functions were linear while the other was a concave up, decreasing function:



Even though it's not strictly necessary to draw a plot like this, you might see it when doing peer-feedback – as it's a nice and good way of illustrating the market clearing price.

So, what is the economic intuition behind this?

For ϵ_2 (the linear, increasing function) there's only a direct income effect on demand for good 2, as p_1 only effects the income but not the price of good 2.

Therefore, any change in demand for good 2 is solely due to the change in income caused by the change in the price of good 1.

Direct income effect is the only factor influencing the demand for good 2 as responses to changes in the price of good 1.

This implies that for every unit increase/decrease in the price of good 1, there would be a consistent (linear) increase/decrease in the demand for good 2.

So, what is the economic intuition behind this?

For ϵ_1 (the concave up, decreasing function) both the price effect and income effect influences the demand. So, if p_1 increases, good 1 becomes more expensive and demand decreases. However, part of this effect is mitigated/averted by the increase in income, as both individuals sell the good.

This means that if the price of good 1 (p_1) increases, two effects occur:

1. Price effect: As the price of good 1 (p_1) increases, good 1 becomes more expensive relative to good 2, leading consumers to buy less of it. This results in a decrease in demand for good 1.
1. Income effect: However, part of this decrease in demand is mitigated by the increase in income. When the price of good 1 (p_1) increases, consumers may experience a decrease in purchasing power. But if their income also increases, it partially offsets the decrease in purchasing power, allowing them to still afford to buy some quantity of good 1.

Working with data in Python

The **Pandas** package is essential when working with data in Python!

Learning how to use the Pandas package can seem boring and be quite syntax heavy. It's important that you're **willing to experiment** to get an understanding of how your code changes your data. Looking at answers is okay, as Pandas syntax takes a lot of time to learn!

Focus on:

- Accessing and manipulating (referencing) data – including slicing and the `.loc`-operator
- Copies vs. views – copies creates new independent objects, while views provide different perspectives or references to the same underlying data

Make sure to understand the difference in making a copy of a DataFrame and making modifications *in place*. Use the keyword argument `inplace=True`.

Some useful Pandas functions for you to experiment with!

Useful Pandas functions:

- `df.tail()` – shows last five rows
- `df.sample(n)` – displays the random n number of rows in the sample data
- `df.shape` – displays the sample data's rows and columns (dimensions)
- `df.describe()` – basic statistics of each column of the sample data
- `df.info()` – information about the various data types used and the non-null count of each column
- `df['income']` – select some column
- `df.iloc[0]` – select a particular row based on its index
- `df[['id', 'income']]` – select multiple columns given
- ... and many more

Data cleaning:

- `df.isnull()` – identify the missing values in your DataFrame – alias for `df.isna()`
- `df.dropna()` – remove the rows containing missing values in any column or vice versa
- ... and many more

Problem set 3

You're going to be working with the data basics of data analysis by loading and cleaning a dataset from Statistics Denmark.

The tasks will teach you to load, delete and rename variables, reshape, selecting subsets of the given data etc. Most task are short but remember to understand the syntax of different operations and experiment with them!

I'll advise you to look at the dataset form Statistics Denmark, before coding. In general, always get an overview with the data you're working with before coding!

As for the problem, make sure to follow the given structure – if it's done correctly, you should be able to run the cell with the call `plot_timeseries()` on your data.

Comments on task 10

Task 10 might be a bit tricky as the error codes are caught by the try-statement. If you're having problems, try to run one line at a time.

There's a few bugs in task 10 in the pre-written try-block, that will cause errors even though your code is correct... correct the following:

- The code refers to a variable in the dataset called 'con'. However, you were earlier asked to change the variable name 'consumption' to 'cons' which is the variable the code is trying to access. You can either change the 'con's in the try-block to 'cons' or go back and rename the variable consumption 'cons'.
- At some point in the try-block, `dt['inc'].values()` and `dt['con'].values()` are called. As `.values` is an attribute, not a method, you should not use parentheses.

Note: Instead of `.values`, I would always use the method `.to_numpy()` – preserves metadata such as index, column labels and data types which results in a more complete representation!

Time for exercises

Problem set 3:

- Task 1-9: Creating and modifying a Pandas DataFrame
 1. Variables
 1. Summary statistics
 2. Indexing
 3. Dropping
 4. Renaming...
- Task 10: Functions in relation to Pandas DataFrames
- Problem 11: Loading, cleaning and structuring data resulting in an interactive plot

Next time...

Video lectures:

- Combining datasets
- Fetching data
- Split-apply-combine

Exercises – Problem set 4. Analysing data:

- ... more data analysing!