Økonometri I

Problem set 1

Agenda
Introduction to class teacher
Introduction to problem set 1
Group work
Recap of group work
Introduction to Stata
Stata exercises

An econometric analysis of Engel curves for Canadian households
The aim of problem sets 1 and 2 is to carry out an empirical analysis of Engel curves. As you know from Micro A, Engel curves describe how demand for individual products or a group of products depend on income at given prices. For this analysis, we will use a data set containing information on a large number of Canadian households. The data file can be downloaded from the course website on Absalon.

The econometric analysis includes several steps. Problem set 1 focuses on selecting, transforming and describing the relevant variables for the analysis. In Problem set 2, we estimate Engel curves using regression analysis and relate the results to economic theory.

Data
The data are taken from the Canadian Consumer Expenditure Survey. Selected households from 5 regions of Canada have kept a diary of all their expenses for a period of two weeks in 1992. These expenses have been combined in a number of categories. Our data contains the following 11 expenditure categories:

fath:	Food (food at home)
rest:	Restaurant
hhop:	Phone, cleaning, childcare, electricity, water (household operations)
wcloth:	Women's clothing
mcloth:	Men's clothing
caruse:	Consumption for care use
tran:	Transportation
care:	Medicine, doctor and dentist
recr:	Recreation
tob:	Tobacco
alc:	Alcohol

In the data set, the variables are named such that *xcategory* refers to expenditures in that particular *category*. E.g., the *xfath*-variable indicates expenditures on food. In problem set 1, we will focus on an analysis of food, alcohol and men's and women's clothing.

The data set also contains some demographic information. We will use the variable $dmale = 1$ if the person is male, $dmale = 0$ if the person is female.

To make the analysis as simple as possible, the data set is restricted to include observations of a cross-section collected in 1992. The cross-section contains only singles without children (i.e., household with only one member).

Group work

Discuss the questions below in groups of max 3 people. Each group should first choose a spokesperson who reports for the group in the subsequent class discussion. The teacher is present and can help clarify issues as you go along.

A crucial step in an econometric analysis is to "translate" economic theory in order to relate it to data and empirical measures. Among other things, this requires you to consider which variables you want to use in the analysis, and whether to transform them or construct new variables.

Assume that we want to analyze the correlation between the demand for alcohol and household income.

1. In this exercise we rely on demand theory. Identify the relevant variables from demand theory.

2. Which information in the data can be used as empirical measures of the theoretical variables? Discuss why we can not estimate price effect with the current data set.

3. In the empirical analysis, total consumption is often used instead of income. Consider why. [Hint: Observe the products included in the data set and think about which products have been omitted.]

4. Why can it be beneficial to analyze data for households with only one member?

STATA exercise: Descriptive analysis
After having identified the relevant variables, it is time to take a look at the data. A descriptive analysis should be the first step in any empirical analysis. The aim is both to get an overview of the structure of data and also to check whether the data "looks reasonable". Often, the data is put together from multiple sources or is drawn from a larger data set. It therefore makes sense to look whether all relevant observations and variables are included (e.g. are there both men and women in our dataset? Is there information on all relevant consumption categories? Are there observations with errors such as negative consumption values?). This initial descriptive analysis is the focus of the first STATA exercise.

We will discuss the results of the STATA exercise next week. Therefore, write down your observations as you move along and save the required output and graphs.

One of the groups is asked to hand in a brief (max 1 page) written response to question 6. The answer should be uploaded to Absalon 24 hours before next week's class.

1. Load data. Data is in the file opg1.dta which can be downloaded from the course website on Absalon. The following commands can be used to create a do-file that loads the data into STATA. In this example, the data opg1.dta is located in the folder `c:\work\` (or `/work` if you are using Mac):

```
clear all
cd "c:\work\"
use opg1.dta
```

2. Examine the data. You can get an overview of the data by using the command `describe` or by clicking on `Data Browser`.

    How many households are there in the data set? What information is provided for each household? Are there households with expenditures for both women's and men's wear, and is it realistic (the purpose of this question is to check whether the data contains error observations)? [Hint: use `count if mcloth>0 & wcloth>0`]

3. Construct a variable for total expenditures. The following piece of code may help:

    generate xtot = xfath + xrest + ... + xwcloth

4. Construct expenditure shares for food, clothes and alcohol:

    generate wfath = xfath/xtot

    ...

    Why is it easier (and more meaningful) to work with expenditure share of a product rather than spending in absolute terms?

5. Explore the data using scatter plots. Plot the expenditure share on food against total expenditures for men and women, respectively. Do the same for expenditure shares on clothes and alcohol. The following code can be used (see next page):

    scatter wfath xtot if dmale==1

    Do you generally see a positive, negative or no correlation between the expenditure share on food (clothing, alcohol) and total expenditures?

6. Carry out a brief descriptive analysis of the expenditure shares for each of the analyzed product categories and of total consumption for each sex. You can use the summarize command:

    summarize xfath wfath ...

This command allows the use of the `bysort` option, thus allowing you to produce summary statistics by e.g. gender (dmale):

bysort dmale: summarize xfath wfath ...

`summarize` calculates number of observations, average, standard deviation, minimum and maximum for each variable. To calculate the median, you have to use `summarize xfath wfath ..., detail`. Set up a table summarizing the results from the descriptive analysis. Comment briefly on the table while you highlight main features of the data, and possibly error observations.

Homework
Finish the STATA exercises.