

## ECON 1291: Excel tutoring

Department of Economics  
Northeastern University  
Fall 2020

The aim of this tutoring is to familiarize you with excel so that you can use it in this class. During the tutoring you will learn to calculate means, standard deviations, minimum and maximum values, correlation, and to do a scatter plot (PS1, problem 4), to run a bivariate regressions using LINEST (PS2, problem 2; PS5; PS7). Hopefully what you will learn today and then master during the course will be valuable tools you will be able to use in other settings.

You are helping the Ministry of Health evaluate the effectiveness of a health program targeting poor households. The main objective of this program is to reduce the expenditure of low-income households in rural areas for health-related issues. A pilot has been carried out in several communities, and you have been asked to assess whether it actually generated a reduction in health expenditures for the beneficiary families.

The data you need to conduct the analysis is in the spreadsheet ECON1291excel tutoring.xlsx. The outcome variable is monthly per capita health expenditure per household. The database also contains many control variables at the household level (e.g. age, education, household size, poverty index). You have data for both eligible households (eligible=1) and not eligible households (eligible=0), in a treatment community (treatcom=1) and a control community (treatcom=0). For each household, you have data on their health expenditure before the start of the program (round=0) and a year after (round=1). The data set contains the following variables:

- **hhid**: a unique ID number for each household (constant across rounds)
- **round**: the year for which you have data (either 0 or 1 )
- **local**: community identifier
- **hhe**: the household health expenditure
- **treatcom**: a dummy variable equal to one for the treatment community
- **agehh**: years of education of the household head
- **hhsize**: household size at baseline
- **pscore**: the household poverty index
- **takeup**: a dummy variable equal to one if the household participated in the program

Use the data set in Excel to answer the following questions.

1. Make a table that reports the means, standard deviations, and minimum and maximum values of all the household health expenditure, household head education and age, household size and poverty index in the dataset.
  - (a) How many observations are in the dataset?
  - (b) How many households are in the dataset?
  - (c) How many localities are in the dataset?
  - (d) How many localities are in the treatment group? And how many in the control?
  - (e) How many household are in the treatment locality?
  - (f) How many household are in the control locality?
2. Create a new variable called *eligible*, which is equal to one if the household has a poverty index lower or equal to 750.
  - (a) How many households in the treatment locality are eligible to be in the program?
  - (b) What fraction of households in the treatment locality are eligible to be in the program?
  - (c) Consider now the variable *takeup*: how many *eligible* household in the treatment locality participated in the program?
  - (d) What fraction of households in the treatment locality participated in the program?
3. Make a table in which you report:
  - (a) the household health expenditure (*hhe*);
  - (b) the level of education of the household head (*educhh*);
  - (c) the household size (*hhsz*);
  - (d) the poverty index (*pscore*)at baseline (*round=0*) for households in treatment and control localities separately.
4. Make a table in which you report:
  - (a) the household health expenditure (*hhe*);
  - (b) the level of education of the household head (*educhh*);
  - (c) the household size (*hhsz*);
  - (d) the poverty index (*pscore*)at baseline (*round=0*) for *eligible* households in treatment and control localities separately.

5. Make a scatter plot showing the relationship between household health expenditure at baseline and the household poverty index.
  - (a) Paste the image (of the scatter plot) into your problem set.
6. Calculate the correlation between household health expenditure and education of the household head at baseline. Is the correlation positive or negative?
7. Estimate of the program's impact by regressing the health expenditure (hhe) variable on the `treatcom` variable using data from the year in which the intervention took place (round=1). Paste your Excel results into your write-up of the answers.
  - (a) How large is the estimated impact of the program?
8. Estimate of the program's impact by regressing the health expenditure (hhe) variable on the `treatcom` variable using data from the year in which the intervention took place (round=1) AND controlling for the level of education of the household head (`educhh`), the household size (`hhsz`), and the poverty index (`pscore`). Paste your Excel results into your write-up of the answers.