Date: 13 August 2018

Re: New joiners STATA exercise

# The dataset

* + 1. The original version of the dataset is available here: <https://sites.google.com/site/frankverbo/data-and-software/data-set-on-the-european-car-market> . However, as detailed below, **please use the version of the dataset which have been provided in the email.**
    2. This is a European car market dataset containing information on sales, prices and characteristics of the car models sold in Belgium, France, Germany, Italy and the U.K. during 1970-1999. The data are combined with macro variables such as exchange rates, GDP, population and price indices.
    3. This dataset is in the public domain and should only be used internally and for the purpose of this exercise.

# Please find below the tasks you should complete

# Importing and explore the data

* + - 1. You have been provided with an excel spreadsheet *“cars\_test\_without99”*. Please import the excel spreadsheet in STATA.
      2. The dataset should include year 1970-1999. The variable *“ye”* indicates the year the car was sold. Please ensure all the years are in the dataset and document your checks.
      3. Please append year 1999 using the dta file you have been provided i.e. *“cars\_test\_99”.*
      4. Please explore the dataset and provide documentation of your checks. For example, you may want to investigate the format of each variable in the dataset and summarise the data using descriptive statistics.

# Cleaning of the dataset

* + - 1. The variable *“ye”* indicates the year the car was sold. The variable only contains the last two digits of the year. For example, for year 1983, the variable only contains the digits 83. Please modify the variable to include the four digits (note you may need to change the variable format before doing this). Further, rename the variable to *“year”.*
      2. The variable *“pr”* indicates the price at which the car was sold. Using the exchange rate and macroeconomic variables included in the dataset please construct the following adjusted price variables:

Price relative to per capital income

Price in common currency

Price in exporter currency

Price in log

* + - 1. Please drop all the observations where the variable “zcode” is equal to 17.
      2. Please generate the following dummy variables:

1 if the car is a luxury car

1 if a car is a compact car

1 is the brand id Alfa Romeo and the class is Luxury.

* + - 1. Please remove all the commas in the variable “model” and replace them with a “/”.
      2. Please provide a dataset unique identifier starting from 1. In other words, a variable which is goes to 1 to total number of observations.
      3. Please generate a variable which indicates the total number of observations in the dataset.
      4. Please generate a variable which indicates, for each brand, the total number of observations (hint – use bysort)
      5. Please generate a variable which indicates, for each brand, the average price (hint use bysort and egen)

# Producing tables

The purpose of this section is to create summary tables. For each table created below, please export the table in excel spreadsheets.

### Hint – For the next tables use collapse command

* + - 1. Please create a summary table indicating for each car class, the average price, the average weight and the average length in the data.
      2. Please create a summary table indicating, for each combination of country and brand, the number of observations in the dataset (e.g. the number of observations of Alfa Romeo sold in Belgium etc)

### Hint – For the next tables use a combination of reshape and collapse commands

* + - 1. Please create a table indicating the total car prices for each combination of brand and market as in the format below (the table below is a simple example with two countries and two brands).

|  |  |  |
| --- | --- | --- |
|  | Italy | Uk |
| Fiat | [total price of Fiat cars sold in Italy] | [Total price of Fiat cars sold in Italy] |
| Opel | [Total price of Opel cars sold in Italy] | [Total price of Opel cars sold in UK] |

* + - 1. Please create the table as above but using the quantity instead of price.

# Merging lookups

You have been provided with the file “lookup\_radio\_model” which indicates whether each car model includes a Stereo or not. Please merge the lookup into the data using the appropriate command and investigate which observations are not merged correctly.

# Annex 1 – Variable explanation

|  |  |
| --- | --- |
| **Variable** | **Label/Explanation** |
| ye | year |
| ma | market |
| co | model code |
| zcode | alternative model code |
| brd | brand code |
| type | name of brand and model |
| brand | name of brand and model |
| model | name of model |
| org | origin code (demand side, country with which consumers associate model) |
| loc | location code (production side, country where the car was produced) |
| cla | class or segment code |
| home | domesti car dummy |
| frm | firm code |
| qu | sales (numer of new cars registration) |
| cy | cylinder volume of displacement (in cc) |
| hp | horsepower (in Kw) |
| we | weight (in kg) |
| pl | places (number, not reliable variable) |
| do | doors (number, not reliable variable) |
| le | length (in cm) |
| wi | width (in cm) |
| he | height (in cm) |
| li1 | measure 1 for fuel efficiency |
| li2 | measure 2 for fuel efficiency |
| li3 | measure 3 for fuel efficiency |
| li | average of li1 li2 and li3 |
| sp | maximum speed (km/hour) |
| ac | time to accelleration |
| pr | price (in destination currency) |
| avexr | average exchange rate of exported country |
| avdexr | average exchange rate of destination country |
| pop | population |
| ngdp | nominal gross domestic product of destination country (destination currency) |