

# Data analysis using Stata

## Problem 4

Marc Sangnier - marc.sangnier@univ-amu.fr

### To do

1. Use Stata *auto* example data set (`sysuse auto`). This data set contains information from an US car reseller.
2. Estimate, via ordinary least squares, the relationship between a car's price and its following characteristics: weight, length and fuel efficiency (*mpg*).
3. Estimate the same model using log-transformed variables. Compare the two models and select the one that performs best. Comment estimates of the selected model.
4. Add to the selected model a variable that indicates whether each car was produced in the US or abroad. Comment.
5. Predict the resell price of a car that is produced abroad and that has the following characteristics: weight = 785kg, length = 3.521m, fuel efficiency = 16.6Km/L.
6. Draw the distribution of the previous regression's residuals and compare it to the normal distribution. Comment.

### Guidelines

- A single do-file must be delivered.
- Answers to questions must be given as comments in the do-file next to the relevant command(s).
- The do-file must not make use of any directories. If files are created, they must be deleted by the do-file itself once it terminates (look at the `erase` command).
- The do-file must be named using the following structure:  
*stata\_problem\_4\_yourname1[\_yourname2]*
- Send your do-file to marc.sangnier@univ-amu.fr. The object of the mail must be "*stata\_problem\_4\_yourname1[\_yourname2]*".