Advanced Econometrics: Homework 1

2022

Instructions

- Form groups of three yourself. Write down your names in the following spreadsheet:
  `https://docs.google.com/spreadsheets/d/17xP3x0MhQ-eXgYdQpNYWl94k0c_mZjkhhxwqffaUeeU/edit?usp=sharing`
- **The solutions should be sent to lenka.nechvatalova@fsv.cuni.cz, with the following subject: 'AE HW1 2022: Group 99 surname1, surname2, surname3'**
- Send me the solution as **one Jupyter notebook** (.ipynb), named **'AE_HW1_group99_surname1_surname2_surname3.ipynb'** which should contain the main analysis together with commented code. Do not forget to use full potential of Jupyter notebooks, show graphs and results as output of your code, write the reasoning and other text in markdown cells (which supports headers, Latex equations, pictures, etc.). It is also possible to send me one pdf file with the analysis and one R script (following the naming convention).
- The empirical problems do not necessarily have a unique solution in terms of numbers, you are assessed based on the execution of the analysis not on the right numbers that you should get from the output. The emphasis is put mainly on the meaningful presentation and the extent of your knowledge.
- The problem set is due by 23:59 **November 9, 2022**. A late submission automatically means 0 points.
- If you have any questions concerning the homework, do contact me by mail and we can set up a consultation. Do it rather sooner than later, I won't give any consultation concerning the homework after November 7.

Problem 1

Suppose we have three groups (A, B, C) in the population. The probability of being in each group is listed in the table below.

Group	A	B	C
probability	θ^2	$2\theta(1-\theta)$	$(1-\theta)^2$

Suppose we obtain random sample where k_1 people are in group A, k_2 in group B and k_3 in group C. Find out the maximum likelihood estimator of θ, $\hat{\theta}$. Calculate the estimate when $k_1 = 543$, $k_2 = 248$ and $k_3 = 710$.

Problem 2

You are provided data 'vietnam_dataset.Rdata'. The dataset contains the results of World Banks 1997 Vietnam Living Standards Survey. The following variables in the dataset are of interest:

- *lhhexp1:* log of total expenditure over the year,
- *lhhex12m:* log of expenditures on health care,

- *farm:* dummy for living on a farm,
- *urban98:* dummy for living in an urban area,
- *age:* age of respondent,
- *sex:* sex of respondent.

Do the following:

(a) Describe the data statistically. Make some descriptive plots.

(b) Fit a linear model with *lhhex12m* as a dependent variable and the rest of the variables as regressors, include constant. Describe and interpret the results.

(c) Fit quantile regression of the same model specification for quantiles from 5% to 95% by 10% steps and plot the results. Interpret the results.

(d) With 95% confidence level for which quantiles are your coefficients different from OLS coefficients? (Do this visually from the pictures, the results are obvious from them.)

(e) Discuss what information did you get from the quantile regression. Contrast the linear and quantile regression.

(f) We know that quantile regression does not work by considering only some portion of the data (i.e. one decile). Demonstrate (if it is indeed the case) that splitting the data into deciles (based on the response variable) and performing linear regression for each decile separately (incorrect method) leads to different "decile" betas compared to the (correct) betas from quantile regression. For simplicity, consider only one independent variable, *lhhexp1* and intercept. Plot betas from both methods the same way as in the case of quantile regression (confidence intervals are not necessary). Reason why there is a difference between the two methods.

Problem 3

Gamma distribution has the following density:

$$f(x) = \frac{1}{\Gamma(k)\theta^k} x^{k-1} e^{-\frac{x}{\theta}}$$

(a) You are provided with sample data generated by the gamma distribution in 'gamma_data.Rdata'. Obtain maximum likelihood estimates of k and θ. Report estimates and standard errors.

(b) Exponential distribution is a special case of gamma distribution. Specify when this is true. Estimate this restricted model and comment on the results. Do you think the data comes from exponential distribution?

(c) Carry out Wald test, likelihood ratio test and Lagrange multiplier test to test the null hypothesis that the data come from exponential distribution. Interpret results of these tests.