

Problem Set 2 - IV

Professor: Chiara Monfardini;

Teaching assistant: Vito Stefano Bramante

- This problem set was sent on **March 23, 2022**
- Due date: **March 30, 2022** at **midnight**
- E-mail your assignment at vitostefano.bramante2@unibo.it no later than **midnight**. Assignments sent after the deadline will not be graded (no exceptions).
- You must attach a single zip file containing: (i) a pdf answer sheet; (ii) the Stata log-file; (iii) the Stata do-file. The pdf file should be **no longer than 5 pages (w/out tables)**.
- Each table and graph in the pdf file should be fully reproducible. **By simply running the .do file** I should be able to reproduce the **exact** table or graph you are showing in the pdf, including title, variable names and numbers. In the .do file you have to signal clearly which chunk of code reproduces which table or which chart. **Doing so will guarantee you up to 2 bonus points on each problem set.**
- Name the zip file (and each file) as *surname1_surname2_surname3.zip*; remember to write the name, surname and student number of each student in the answer sheet.
- The grade for the 3 assignments will be 40% of the final grade.
- Please follow carefully the instructions detailed above. **Any misconduct will negatively impact the grading of the assignment.**

Consider the data set *lifesat.dta*, sampled from the *Survey of Health, Ageing and Retirement* in Europe via [SHARE project](#)¹. The data set contains socio-demographic and health-related information about several thousands of individuals aged 50 and over from many European countries and Israel. Individuals were surveyed across several years. The data set contains three waves: 2011, 2013 and 2015. The aim of the problem set is to understand and estimate the relationship between happiness and income. The outcome variable is self-perceived life satisfaction, *life_sat*, which takes values 1 if the interviewed is really satisfied with her/his life and 0 otherwise (the question is about general life satisfaction and not current feelings)². The main explanatory variable is total household income at the moment of the interview (measured in tens of thousands Euro), *income*. The binary outcome variable model is:

$$life_sat_{i,t} = \alpha + \gamma income_{i,t} + \mathbf{x}_{i,t}'\beta + \lambda_t + \epsilon_{i,t} \quad (1)$$

where i stands for the individual and t for the year of interview. \mathbf{x}_{it} is a $(K \times 1)$ vector of socio-demographic and health-related variables; while λ_t is a (2×1) vector containing wave dummies.

¹A. Borsch-Supan. Survey of Health, Ageing and Retirement in Europe (SHARE) Wave 6. Release version: 6.0.0. SHARE-ERIC. Data set. DOI: 10.6103/SHARE.w4.500. 2016d

²The binary variable is built from a life satisfaction index which takes on a 10-point scale, where 0 is a completely dissatisfied person and 10 if completely satisfied. *life_sat* takes value 1 if the original index takes value 7, 8, 9 or 10.

During the analysis:

- consider a 10% significance level in hypothesis testing
- Choose carefully the variables to include in the vector \mathbf{x}_{it} ; there must be a sound explanation behind your choice. Some “must-have” variables are: *age* (possibly including a non-linear relationship), *gender*, *yedu*, *mstat*, *hstatus* and *gali*. Please refer to the Codebook in the Appendix for more information about variables available in the data set.
- You can transform variables as you wish, always explaining the reasons behind the choice.
- For example, you can turn qualitative variables into dummies, and use the latter in your model.
- Whenever you have finalized a regression model, stick with it without changes throughout the problem set.

Part 1: LPM

1. Estimate Model 1 using OLS. Is this model appropriate? Justify your answer by analyzing the predicted probabilities.
2. How would you interpret the estimated values of γ ? Comment the results.
3. Comment magnitude, sign and significance of gender, age and the other variables estimated coefficients. Are your estimates reasonable?

Part 2: Probit estimation

Now consider the following model:

$$Pr[life_{sat_{i,t}} = 1 | income_{i,t}, \mathbf{x}_{i,t}, \lambda_t] = \Phi(\alpha + \gamma income_{i,t} + \lambda_t \iota) \quad (2)$$

where $\Phi(\cdot)$ is the cumulative Standard Normal distribution function.

1. Estimate Model 2 using ML. Are the regressors jointly significant? How can you interpret the estimated coefficients? Can you compare these coefficients with the ones obtained via LPM?
2. Using both a Wald test and a Likelihood Ratio test, test for the joint significance of the trend dummies coefficients. Comment and compare your results. Are the conclusions similar?
3. Use the command *margins* to compute both the average partial effects (APE) and the partial effects at the average (PEA) of all explanatory variables. Explain the difference between the two.
4. Compute manually the partial effect at the average of the continuous variable *income* and interpret your findings, comparing it with the coefficient estimated by OLS.
5. Compute manually the partial effect at the average of *gali* and interpret your findings, comparing it with the coefficient of *gali* estimated by OLS
6. Which is the percent correctly predicted of this model in the estimation sample? Compute this indicator using a threshold equal to 0.5 and using as threshold the sample average of *life_sat*. Do you find that the second value is adequate?

Codebook

- *id*: person identifier
- *year*: year identifier
- *gender*: 1 if male, 2 if female
- *age*: age
- *yedu*: years of education
- *hstatus*: self-perceived health status; 1 if poor health, 2 if fair, 3 if good, 4 if very good, 5 if excellent
- *income*: total household income (tens of thousands)
- *mstat*: marital status; 1 if married and living with spouse, 2 if registered partnership, 3 if married and not living with spouse, 4 if never married, 5 if divorced, 6 if widowed *nchild* (*ngrchild*): number of children (grandchildren)
- *htype*: household type single: 1 if single, 0 if couple
- *nursinghome*: 1 if living in a nursing home, 0 if not
- *otrf* : 1 if owner of the house where he/she lives, 2 if member of a cooperative, 3 if tenant, 4 if subtenant, 5 if rent free
- *fahc* (*fohc*): amount spent on food at (outside) home
- *hprf* : annual home produced consumption
- *thexp*: total household expenditure
- *hnetw*: household net worth
- *gali*: 1 if limitation with activities, 0 if not
- *bmi*: Body Mass Index
- *esmoked*: 1 if ever smoked daily, 0 if not
- *phinact*: 1 if physically inactive, 0 if not
- *doctor*: seen/talked to medical doctor
- *hospital*: 1 if hospital in the last 12 months, 0 if not
- *cjs*: current job situation; 1 if retired, 2 if employed/self-employed, 3 if unemployed, 4 if permanently sick, 5 if homemaker, 6 if other