Exercise sheet 3 – November 12, 2021

Please submit your solution electronically until November 19, 2021, 23:59
Send your script-file to jerke@soziologie.uzh.ch

Notes:

- **Please sufficiently comment your script and structure it according to the different tasks.**
- **Whenever a task asks for an explicit answer, please write down your answer directly in the script within a comment.**
- **Make sure that you fully document your solution in your script. If you give an answer, but there is no code to clearly reconstruct how the answer was determined, the answer cannot be counted.**
- **The tasks vary in difficulty. For some of them you may have to combine commands in a new way or have to look in the documentation of the respective libraries.**
- **Most exercise sheets will contain bonus questions, providing the possibility to obtain extra points.**
- **The solution of this sheet will be published on OLAT after the submission deadline expires.**

1. We will be again working with the voting data set. Therefore, load it with Python and repeat the data cleaning steps from the script that we used in the lecture on logit regression and from the last exercise.

2. This time, we are interested in why people didn't vote. First, let's have a look at our variable of interest:
    a. How many people didn't vote?
    b. Compare the socio-demographic background of voters and non-voters:
        i. Do they differ in their age?
        ii. What is the gender distribution for voters and non-voters?
        iii. Do voters and non-voters differ in their share of lower and higher education?

3. Run an initial logistic regression to investigate whether there is a relationship between socio-demographic features of the respondents and their decision to vote.
    a. Which of the coefficients have a significant effect on non-voting behavior?
    b. Interpret the coefficients.
    c. Calculate the odds ratios for the coefficients.
    d. Demonstrate for a coefficient of your choice how the interpretation changes when we refer to the odds ratios.
    e. We can infer the direction of the effect from the raw coefficients by looking at their sign. How can we infer the direction of the effect from the odds ratios?

4. As we learned in the lecture it is better to use marginal effects when interpreting the coefficients.
    a. Calculate the AME and MEM for the previous regression model.
    b. Interpret the AME for the effect of age and education.

5. The decision to vote or not might well be related to the general interest in politics. We can measure the interest in politics by question q60 (“Interest in politics”) and q7 (“Politics in the media”).
    a. Add these variables to the previous regression model. Based on the Pseudo R-square, how does the explanatory strength change?
    b. How would you describe the effect of interest in politics on the voting behavior?
    c. [Bonus] Compare the AIC of the two regression models. Which model performs better?