**Universidade de São Paulo**

**Faculdade de Filosofia, Letras e Ciências Humanas**

Departamento de Ciência Política

FLP-0468 & FLS-6183

2º semestre / 2019

**Lab 7/ Class 8**

In this problem set, we will test the linearity assumptions related to interaction models.

In the article, “A Simple Multivariate Test for Asymmetric Hypothesis” by Clark, Gilligan and Golder, Duverger’s theory about the factors that determine the number of parties in a democracy is explored using an interaction model. Figure 1 depicts the relationship between district magnitude and the effective number of parties. Figure 2 depicts the relationship between ethnic fragmentation and the effective number of parties.

Figure 1



Figure 2



In the article, the authors argue that the hypothesis that the effect of the electoral district size on the number of legislative parties should be treated as asymmetric due to the fact the explanatory variable is not necessary and sufficient. Based on data and the replication of the test of Duverger’s theory developed by Amorim Neto and Cox (1997), the authors estimate the conditional effect that social cleavages has on the marginal effect of district size on the number of parties using an interaction model.

Thus, the model tested is:

$$Multipartism= α+ β\_{1}ElectoralDistrictSize+ β\_{2}SocialHeterogeneity+ β\_{3}ElectoralDistrictSize\*SocialHeterogeneity+ μ$$

Previously, we have reviewed a few rules about the proper interpretation of interaction models, such as the need to include of all constitutive terms in the estimated model and the analyses of marginal effects. In a recent paper, Hainmueller, Mummolo and Xu (2018) highlight that researchers need to check the linear interactive effect (LIE) assumption. In addition, they argue that researchers must pay attention while generalizing findings to the distribution of the observations. In this lab, our goal is to learn how to identify such problems, using the techniques presented by the authors.

The variables that are going to be used are the following and the data set is the original data set used by Cox and Amorim Neto (1997):

|  |  |
| --- | --- |
| **Variable Names** |  |
| Enps | Effective Legislative Parties – Multipartism |
| Eneth | Social heterogeneity |
| ml | Electoral district size |
| lnml | Log Median Electoral district size |

**Part 1 – X and Z are dummy variables**

$$Multipartism= α+ β\_{1}ElectoralDistrictSize+ β\_{2}SocialHeterogeneity+ β\_{3}ElectoralDistrictSize\*SocialHeterogeneity+ μ$$

where:

Multipartism is a continuous variable that measures the number of effective legislative parties.

Electoral District Size is a continuous variable that measures the amount of candidates that might be elected in each electoral district for legislative elections.

Social heterogeneity is a continuous variable that measure how many ethnic groups there are in each country of the sample.

1. Please estimate an *interaction* regression model based on the dummy variables you created in the do file and discuss the results.

To do so, please use the margins command in Stata carefully identifying the variables as categorical variables by inserting “i.” preceding each variable. What does the regression tell us about the effect of social heterogeneity on the relationship between electoral district size and the number of effective parties elected describe by Duverger? Please do not forget to report your Stata output and to write the regression equation generated as an equation.



1. Now, let us examine the model’s predictions on the effective number of legislative parties depending on both social heterogeneity and district type. Please fill in the following table with the coefficients and confidence intervals and include the commands in your do-file.

Table 1. Predicted Number of Parties Conditional on Social Heterogeneity and District Magnitude (Dummy Variables Case)

|  |  |  |
| --- | --- | --- |
| Social Heterogeneity  | Single-Member Districts = 0 | Multi-Member Districts = 1 |
| Heterogeneous = 1 |   |   |
| Homogenous = 0 |   |   |

c. Based on the confidence intervals you estimated, what can we conclude about the difference between Single-Member Districts versus Multi-Member Districts and how these are influenced by social heterogeneity. Use margins command to visualize graphically the predictions.

**Part 2. X is a dichotomous variable and Z is a categorical variable**

1. In this section of the lab, we will use Multipartism and Electoral district size variables in their continuous form while Social Heterogeneity remains as a dummy variable. Please summarize both continuous variables conditional on social heterogeneity. Can you notice any difference in both variables between groups?

1. Draw a scatterplot between Multipartism and Electoral district size (on the x-axis) and analyze the figure. In the commands in the do file, you will find two commands that we have not used yet in this course (lowess and lfit). You should search for what these commands are doing in order to understand the figure.

What do each of the lines describe and how does this relate to the distribution of the data shown by the scatterplot?

1. Now draw scatterplots of both continuous variables Multipartism and Electoral district size (on the x-axis) conditional on Social Heterogeneity variable. Are there are any differences between groups? Please describe the results and your conclusion.
2. Please estimate the regression model with the interaction as described above using the following command:

regress enps c.lnml##i.heterogeneity

Describe the results.

1. Using the script provided in the do file, construct a figure to analyze if the predicted number of effective legislative parties is determined on district size conditional on whether societies are homogeneous or heterogeneous.
2. Considering the results found before in the exercise a, b and c, how do your conclusions about the interaction regression change?

**Part 3: X is a continuous variable and Z is a categorical variable**

1. Describe the social heterogeneity variable and plot a kernel density graph.
2. In order to analyze the relation between the number of effective legislative parties and the district size, Hainmueller, Mummolo and Xu (2018) propose to divide the moderator variable according to its percentiles. In this case, ethnic group, is the moderator variable. In the do file, you will find the command used previously, but by grouping them into three categories. Please describe the results, comparing with what was found in part 2.
3. We will now examine the effect of district size on the number of effective legislative parties and whether the effect of district size on the number of parties is conditional on social heterogeneity when this characteristic is measured as a continuous variable. Please estimate the regression model and discuss the results.
4. Using the margins command, present the predicted values of the marginal effect of district magnitude on the number of parties conditional on social heterogeneity. Please interpret the results.
5. Using the margins command, present the predicted values of the marginal effect of social heterogeneity on the number of parties conditional on district magnitude. Please interpret the results.
6. Based on the assigned reading for this class on the interflex command, can you develop your own implementation of this command to more clearly show your findings? Please be sure to include your commands in the do file.
7. Does your interpretation of the hypothesis test of the interaction between social heterogeneity and district magnitude on the number of parties change from your earlier conclusions in Part I and II?