**Universidade de São Paulo**

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

Departamento de Ciência Política

FLP-0468 & FLS-6183

2º semestre / 2019

**Lab 6/ Class 7**

**Multiple Regression Models with Interactions**

In this problem set, we will explore multiple regression models that permit us to test interactive hypotheses. The general form of the interaction model we will study can be summarized as:



We will examine hypothesis testing in four cases:

1. X and Z are dummy variables;
2. X is a dummy and Z is a categorical variable;
3. X is a continuous and Z is a categorical variable; and,
4. X and Z are continuous variables.

This model imposes very different assumptions from the multiple regression model in which we only impose that the effects of the explanatory are additive.

Our goal is to understand how feminism and income influence Hillary Clinton’s Thermostat rating and whether the effect of these variables is conditional on an interaction that helps us to understand how they influence Hillary’s thermostatic rating:



The data set: *nes1996\_modif.dta.* For more information on the variables, see the *codebook* in Moodle.

|  |  |
| --- | --- |
| **Variable Names in Original Data Set** | **Transformed Variable Names** |
| v960281 | hillary (*Hillary*) |
| v960701 | faminc (*Income*) |
| v961039 | wmnmvt (*Women’s Movement*) |

Here are some very basic summary statistics.

****

This lab will also familiarize you with the commands in Stata that are commonly employed to estimate and analyze interactions. For further information, please familiarize yourself with the margins and marginsplot help files in Stata. In the final part of the lab, we will also use a user-written program (interflex) that will be especially helpful in analyzing non-linear interactions, which is the subject of our next class and lab.

**1. X and Z are dummy variables**

The general model can be summarized as:



where:

*Hillary Thermometer* is a discrete variable that measures the respondent’s opinion of Hillary Clinton ranging between a minimum of 0 (unfavorable opinion) and a maximum of 100 (favorable opinion);

*Women's Movement* is a discrete variable that measures the respondent’s opinion of the women’s movement ranging between a minimum of 0 (unfavorable opinion) and a maximum of 100 (favorable opinion); and,

*Income* is the categorical variable of the respondent’s 1995 family income. The categories range from 0 (lowest income group) to 24 (highest income group).

1. Please create two dummy variables for *Women's Movement* and *Income* so that each dummy measures who is a *Feminist* and who has *High Income*. What criteria did you decide to use to choose the cutoff measure for each measure? Please justify your coding decision based on the concepts that we seek to measure and test.
2. Please estimate an *additive* regression model based on the variables you created in (a) and discuss the results. (For example, regress hillary feministdummy hiincdummy). What does the regression suggest about how income and feminist sentiments influence how individual’s feel about Hillary Clinton? Please do not forget to report your Stata output and to write the regression equation generated as an equation.
3. Please estimate an *interaction* regression model based on the variables you created in (a) 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. (For example, regress hillary i.feministdummy##i.hiincdummy). What does the regression suggest about how income and feminist sentiments influence how individual’s feel about Hillary Clinton? Please do not forget to report your Stata output and to write the regression equation generated as an equation.
4. Let’s now construct a table to compare the predicted effects of high and low income and feminism on Hillary’s rating. Based on the regression estimated in (c), please fill in the blanks with the combination of coefficients we will use to calculate the average predicted effect.

Table 1. Predicted *Average Effects* for Hillary

|  |  |  |
| --- | --- | --- |
|  | Anti- *Feminist (hiwmnmvtdummy=0)* |  *Feminist (hiwmnmvtdummy =1)* |
| *Low Income* (hi*famincdummy=0*) |  |  |
| *High Income* (hi*famincdummy=1*) |  |  |

1. In Stata, we can use the margins command immediately after estimating a regression to obtain the predicted values given specific values of the explanatory variables. In this case, we will run the margins command in Stata specifying “margins i.feministdummy##i.hiincdummy, atmeans.”

 Based on this command, please fill in the table below with the point and 95% confidence interval estimates.

Table 2. Predicted *Average Effect*s for Hillary

|  |  |  |
| --- | --- | --- |
|  | Non- *Feminist (hiwmnmvtdummy=0)* |  *Feminist (hiwmnmvtdummy =1)* |
| *Low Income* (hi*famincdummy=0*) |  |  |
| *High Income* (hi*famincdummy=1*) |  |  |

1. Based on the confidence intervals you estimated, what can we conclude about the difference between High Income versus Low Income and how these are influenced by feminism?

**II. X is a dichotomous variable and Z is a categorical variable**

1. We will now examine the effect of income and how it is influenced by feminism when feminism is a dichotomous measure and income is treated as a categorical variable. Please estimate the regression model and discuss the results using the following command:

regress hillary i.wmnmvtdummy##i.faminc

1. Using the script provided in the do file, construct a graph to analyze the average predicted effect of income influences how feminist and non-feminists rate Hillary.
2. In the graph in (b) comparing the average predicted effect for a given income level given a person is a feminist or anti-feminist, it is difficult to discern if there are statistical differences as income varies. Use the commands provided in the do file to construct a slightly different plot. How do you interpret this plot?

**III. X is a continuous variable and Z is a categorical variable**

1. We will now examine the effect of income and how it is influenced by feminism when feminism is a continous measure and income is treated as if it was a categorical variable. Please estimate the regression model and discuss the results.
2. Using the commands below, we can compare the average predicted effect for low, medium and high-income level for a person across variations in their feminist score ranging from 0 to 100 increasing at 20 points per level. Run the following script and look at the results. How do you interpret these results?

regress hillary c.wmnmvt##i.faminc

margins, dydx(wmnmvt) at(faminc=(1(7)24) wmnmvt=(0(20)100))

**IV. X and Z are continuous variables**

1. We will now examine the effect of income and how it is influenced by feminism when feminism is a continuous measure. Thinking about marginal effects, we can summarize the partial derivatives for each variable in the additive and interaction models. For further information, please see the tables posted by Matt Golder in his website regarding interactions which also explain interactions employing partial derivatives). These models have an important theoretical implications. Please explain the theoretical implications and differences between both models, and then estimate both models. What do the results suggest in each case?









1. We are now going to use the interflex program, to calculate the marginal effects as proposed by Brambor, Clark and Golder (2006). We will look at both cases below. Please interpret both figures and the implications for hypothesis testing.

Case 1) The marginal effect of *Income* on *Hillary* (the dependent variable) conditional on the *Women’s Movement.*

Case 2) The marginal effect of the *Women’s Movement* on *Hillary* (the dependent variable) conditional on *Income.*

**IV.** With respect to the results obtained in the earlier results, please discuss if wnmvt and family income are necessary, sufficient, or independent based on table 3 (Clark, Gilligan e Golder, 2006: 322).

**V.** Based on the findings from the tests employed in the preceding exercises, please write up an abstract of the paper that would best capture your findings. Your abstract should include something along the lines of the following:

“In this paper, we test a theory in which X *increases (decreases)* Y. We also test whether the effect of X on Y is conditional on Z.”