

**Problem Set #4 – Interactions**

In this problem set, we will explore models with interactions. The general form of the model we will study can be summarized as:

$$y_i = \alpha_i + \beta_1 x_i + \beta_2 z_i + \beta_3 x_i z_i + u_i$$

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.

Our goal is to explore the example discussed in the last class and specifically to understand how feminism and income influence Hillary Clinton's Thermostat rating and whether these variables have an interaction that helps us to understand how they influence Hillary's thermostatic rating:

$$\begin{aligned} \text{HillaryThermometer} = & \alpha + \beta_1 \text{Women's Movement} + \beta_2 \text{Income} \\ & + \beta_3 (\text{Women's Movement} \times \text{Income}) + u \end{aligned}$$

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

Variable Names in Original Data Set	Transformed Variable Names
v960281	hillary ( <i>Hillary</i> )
v960701	faminc ( <i>Income</i> )
v961039	wmmvmt ( <i>Women's Movement</i> )

Here are some very basic summary statistics.

variable	mean	sd	p10	p25	p50	p75	p90
hillary	52.8089	29.84971	0	30	60	70	85
wmmvmt	63.73674	21.53509	41	51	61	81	86
faminc	15.02621	6.340515	4	11	16	20	22

1. X and Z are dummy variables

The general model can be summarized as:

$$HillaryThermometer = \alpha + \beta_1 Women's\ Movement + \beta_2 Income + \beta_3 (Women's\ Movement \times Income) + u$$

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 category of the respondent's 1995 family income. The categories range from 0 (lowest income group) to 24 (highest income group).

- a) 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 (hint: think about what we are trying to measure).?
- b) Please estimate the regression model based on the variables you created in (a) and discuss the results. (Hint: Using the margins command, you might want to estimate the following: regress hillary i.feministdummy##i.hiincdummy).
- c) Let's now construct a table to analyze the predicted effects of high and low income and feminism on Hillary's rating. Based on the regression, 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 (hiwmmvtdummy=0)	Feminist (hiwmmvtdummy =1)
Low Income (hifamincdummy=0)		
High Income (hifamincdummy=1)		

- d) Fill in the table below with the point and 95% confidence interval estimates.

Table 2. Predicted Average Effects for Hillary

	Anti-Feminist	Feminist
Low Income		
High Income		

- e) 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

- a) 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 if it was a categorical variable. Please estimate the regression model and discuss the results using the following command:

```
regress hillary i.wnmvtdummy##i.faminc
```

- b) Using the script provided in the do file, construct a graph to analyze the average predicted effect of income influences how feminist and anti-feminists rate Hillary.
- c) 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

- a) We will now examine the effect of income and how it is influenced by feminism when feminism is a continuous measure and income is treated as if it was a categorical variable. Please estimate the regression model and discuss the results.
- b) 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.wnmvt##i.faminc
```

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

## III. X and Z are continuous variables

- a.) We will now examine the effect of income and how it is influenced by feminism when feminism is a continuous measure. Thinking about the marginal effects, please summarize them as we discussed in class thinking about partial derivatives (hint: See the slides from the last lecture and the tables posted by Matt Golder in his website).

$$\begin{aligned} \text{Hillary Thermometer} = & \alpha + \beta_1 \text{Women's Movement} + \beta_2 \text{Income} \\ & + \beta_3 (\text{Women's Movement} \times \text{Income}) + u \end{aligned}$$

- b.) 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 analysis above, write up an abstract of the paper that you could write. Your abstract should include something along the lines of the following:

“In this paper, I develop a theory in which  $X$  *increases* (*decreases*)  $Y$ . I argue that the effect of  $X$  on  $Y$  is conditional on  $Z$ .”