

## POL 681 Problem Set 5: Statistical Interactions and Graphs—ANSWER KEY

Objective: The objective of this problem set is to give you experience in working with interaction terms in a linear regression model. You will be expected to correctly estimate the parameters of interest and correctly interpret them (including interpreting the uncertainty about the parameter estimates.)

Directions: Please use the dataset `hispanic.dta` to address the following questions. You will need to use `Stata`. Please turn in `Stata` log files and all graphs. Your answers should be wordprocessed and you should interpret your results using substantive language. Do not use the variable name stubs in your interpretation (i.e. don't refer to the dependent variable as "HISPTHRM"). This problem set is due 1 week from this Tuesday.

In the dataset `hispanic.dta`, there are four five variables of interest. The first, denoted in the data as `hispthrm` is the NES 101-point "feeling thermometer" question for Hispanics. The scale ranges from 0, denoting least favorably evaluated, to 100, denoting most favorably evaluated. For this problem set, this variable will be the dependent variable.

The variable denoted as `immthrm` is an NES 101-point feeling thermometer question for immigrants. The scale ranges from 0, denoting least favorably evaluated, to 100, denoting most favorably evaluated.

The variable denoted as `prej` is a scale derived from the NES measuring prejudicial attitudes. The scale ranges from 7 (denoting most prejudicial attitudes) to 30 (denoting least prejudicial attitudes).

The variable denoted as `hispa` denotes the percentage of Hispanics residing in the congressional district in which the survey respondent lives (data are from U.S. Census Bureau).

The variable `white` is a dummy variable denoting whether or not the respondent is white and non-Hispanic. A zero denotes the respondent is either Hispanic, African-American, Asian, Native American, or is in some other non-white racial or ethnic group. Hence, a 0 implies the respondent is from a racial or ethnic minority group.

A researcher is interested in the following questions. Please answer them by providing analysis of the data using `Stata`.

1. Do white attitudes toward Hispanics significantly differ from minority respondent attitudes toward Hispanics? Why or not? (8 points)

**This question requires you to estimate a simple bivariate regression model treating the Hispanic feeling thermometer as the dependent variable and the binary variable denoting white, non-Hispanic as the independent variable. The results from this model are shown in the log file for question 1. The interpretation is that whites, on average, give an evaluation of Hispanics that is about 13 "degrees cooler" than for non-whites. Because this estimate is statistically different from 0, we can conclude (based on this model) that white attitudes toward Hispanics is significantly different from minority attitudes.**

2. Does there seem to be a connection between attitudes toward immigrants and attitudes toward Hispanics after controlling for race/ethnicity of the survey respondent? Why or why not? (8 points)

**To answer this question, you would need to reestimate the previous model but add to the model the independent variable denoting attitudes toward immigrants. The `Stata` output for this model is shown in the log file for question 2. Based on the model, after controlling for the race of the respondent, we find a significant and positive relationship between attitudes toward immigrants and attitudes toward hispanics. Specifically, as the feeling thermometer**

for immigrants increases by “1 degree,” the expected Hispanic feeling thermometer increases by about .5 degrees. The relationship is positive and statistically different from 0.

3. Provide a graphical summary of the relationship found in question 2. How is this figure interpreted? How does this graph illustrate the additivity property? (6 points)

The graph of this relationship is shown in Figure 1. The graph illustrates the positive slope between Hispanic evaluations and immigrant evaluations. The difference between groups, which is derived from the dummy variable, suggests that nonwhites’ (the higher of the two intercepts) attitudes are more positive than attitudes of whites. The additivity property is illustrated by the fact that the slope, or the rate of change, between Hispanic evaluations and immigrant evaluations is constant for both whites and non-whites. The only difference between the two groups is attributable to the offset between the two groups (which is given by the dummy term). The Stata code to generate this graph is given in the log file for question 3.

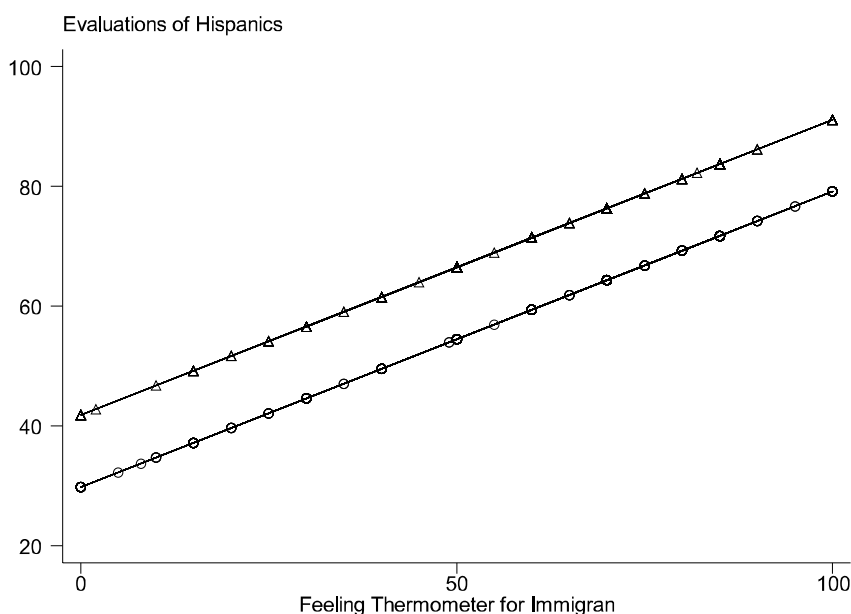


Figure 1: *Plot of Model Estimated in Question 2.*

4. Suppose that instead of additivity, the researcher conjectured that the impact of the evaluation of immigrants was conditional on whether or not the respondent was white (non-Hispanic) or non-white (or white Hispanic). Construct and estimate a model that tests this proposition. How is this model substantively interpreted? (10 points)

If one believed the relationship between Hispanic attitudes and immigrant attitudes was moderated by race of the respondent, then one may be led to the consideration of a product term (or an interaction term). This would be done by interacting the race of the respondent with attitudes toward immigrants. In the log file, the regression model (and code) is given for Question 4. We find that the interaction term is significantly different from 0, thus providing some evidence for a “conditional” relationship. The interaction term suggests the relationship between race and immigrant attitudes on Hispanic evaluations is multiplicative. Specifically, we see that the slope of the relationship between immigrant attitudes and Hispanic attitudes is steeper for whites. The conditional slope for non-whites is about .40; the conditional slope for whites is about .62.

5. Provide a graphical summary of the relationship found in question 4. How is this figure interpreted? How does this graph illustrate *non-additivity*? (6 points)

The graph of this relationship is shown in Figure 2. The graph illustrates the positive slope between Hispanic evaluations and immigrant evaluations. The graph also illustrates the point made above: the pitch, or steepness of the slope between immigration evaluations and Hispanic evaluations is greater for whites than nonwhites. This illustrates the conditional nature of the slope: the slope is conditional on respondent race. Non-additivity is illustrated here because of the nonconstant slopes across the two groups. The code to generate this graph is given in the log file for question 5.

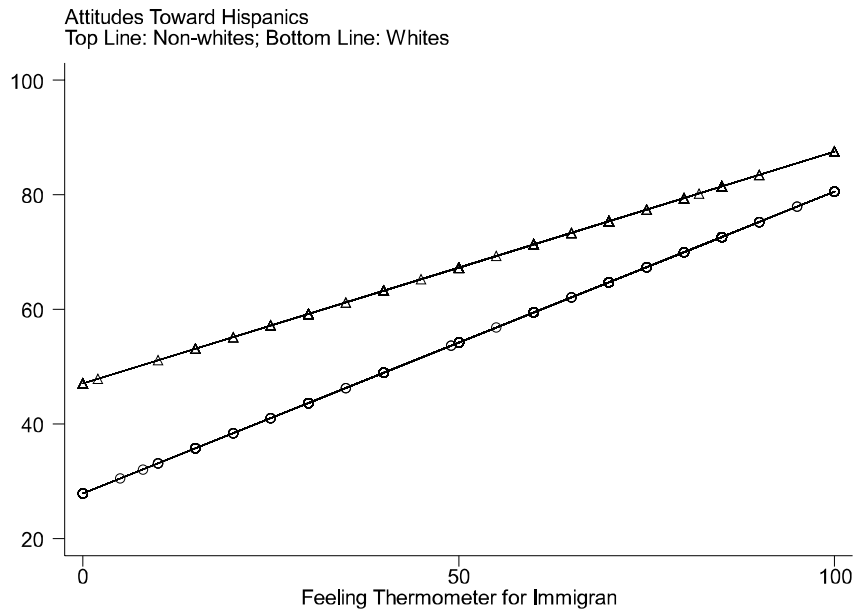


Figure 2: *Plot of Model Estimated in Question 4.*

For the next set of questions, omit the `white` and `immthrm` variables from your analysis.

6. Suppose that our researcher believes that one's level of prejudicial attitudes is multiplicatively related to the number of hispanics residing in close proximity to the respondent? The theory leading to this expectation is that as one is surrounded by more members of the minority group, negative stereotypes may be counteracted due to exposure. Estimate and interpret a model that tests this proposition. How do you substantively interpret this model? Provide me with any graphs or tables of predicted values you think you need in order to interpret this model. (20 points)

To evaluate a multiplicative relationship between prejudice, Hispanic population, and attitudes toward Hispanics, the regression model shown in the log file under question 6 must be estimated. To estimate this, an interaction term between the prejudice scale and the percent Hispanic population variable must be created. The significance ( $p < .10$ , two-tail test) of the interaction term suggests there is in fact an interaction between the two variables. Substantively, this implies that the relationship between Hispanic attitudes and prejudicial attitudes is conditional on the Hispanic population residing in the congressional district. Similarly, the relationship between the distribution of the Hispanic population and Hispanic attitudes is conditional on one's prejudicial attitudes. There are no main effects: the slopes are condi-

tional on the other covariate. To substantively interpret this model, use of graphical methods may be the best way to go. In the Stata log file, I give the code to generate the conditional relationships found in the interactive model. In Figure 3, I graph the slope of the Hispanic feeling thermometer on prejudice, conditional on Hispanic population. The three lines in the graph correspond to setting the prejudice scale to its minimum value (7), its maximum value (30), and its mean value (21). The “dots” represent the actual predicted values. In the Stata log, the variable “r1” denotes the actual predicted values; the variable “rpmax” denotes the case where the prejudice scale is set to 30; “rpmin” denotes the case where the prejudice scale is set to 7; “rpmean” denotes the case where the prejudice scale is set to its mean. The relationship shows that the conditional effect of the Hispanic population is to increase the slope between Hispanic evaluations and prejudicial attitudes.

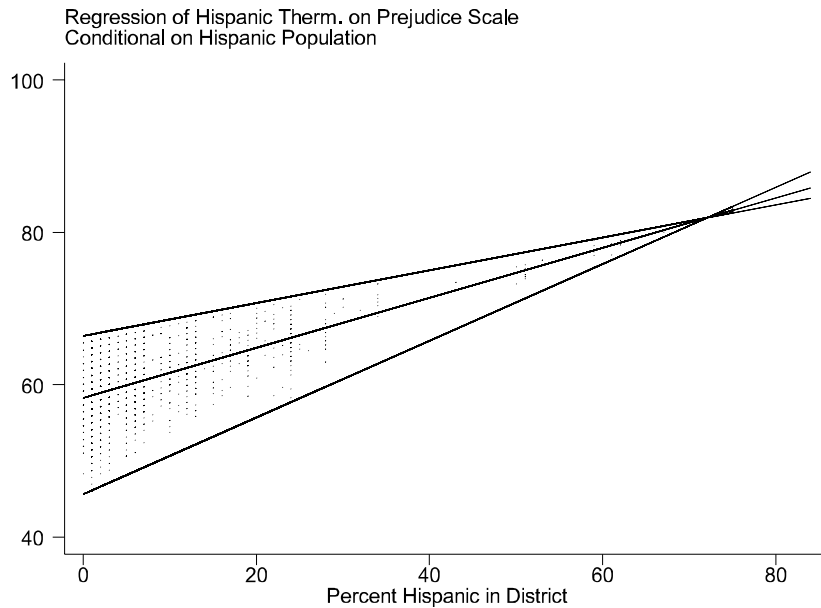


Figure 3: Relationship between Hispanic evaluations and the Prejudice Scale, conditional on the distribution of the Hispanic Population in the Congressional District.

The relationship between Hispanic attitudes and the Hispanic population conditional on the prejudice scale is shown in Figure 4. This relationship shows that as attitudes become more tolerant (less prejudicial), the slope between Hispanic evaluations and the distribution of the Hispanic population increases. The three lines on the graph correspond to the approximate maximum, minimum, and mean values of the percent Hispanic population variable. The dots denote the actual predicted values. The Stata log file gives the codes used to generate these predicted values and graphs. This information is given for question 6.

7. Because the slope estimates are conditional in this model, so are the standard errors. Are these slope estimates statistically different from 0 across the full range of the interaction term? Provide concise evidence to answer this question. (15 points)

(Note that concise evidence could come in the form of computing the  $t$ -ratios for the slope of hispthrm on prej conditional on hispa; compute the  $t$ -ratios for the slope of hispthrm on hispa conditional on prej and then graphing them. Also, graphing the 95 percent confidence intervals might be useful. How you choose to report your evidence is up to you; just make sure I can follow it!)

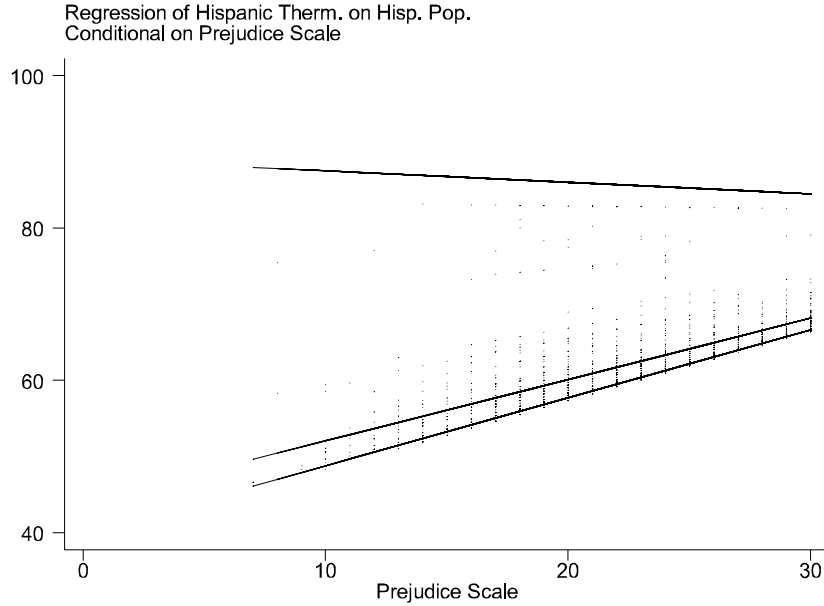


Figure 4: *Relationship between Hispanic evaluations and the percent Hispanic population, conditional on the Prejudice Scale.*

To determine if the interaction effects are significant over the full range of the term, I computed the conditional slopes and the standard errors for the conditional slopes and graphed the  $t$  ratios based on them. The graphs are shown in Figure 5. The Stata log file for question 7 gives the code for computing these quantities. For the slope of the Hispanic thermometer on prejudice, conditional on percent Hispanic, the interaction effect is significant over the full range of the prejudice scale (the lowest  $t$  ratio is well above the critical  $t$ ). For the other conditional slope, the interaction term is insignificant for larger values on the Hispanic population variable. This is the case because there are very few observations at the higher end of the Hispanic population variable.

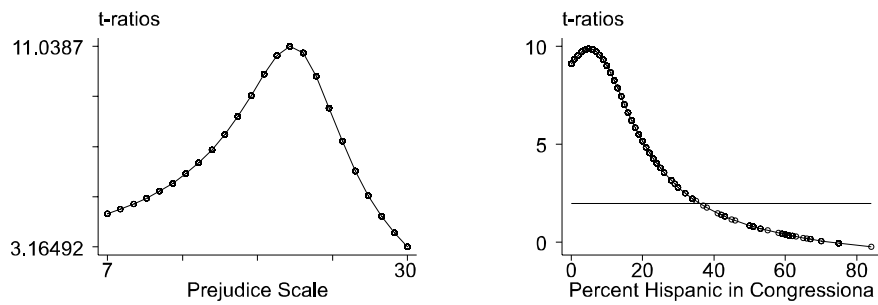


Figure 5: *This figure gives the t-ratios for the sets of conditional slopes. The first panel gives the conditional slope for the regression of Hispanic thermometer on the prejudice scale, conditional on Hispanic population. The second panel gives the conditional slope for the regression of the Hispanic thermometer on the percent Hispanic, conditional on the prejudice scale.*