# Assignment 6: Dummy Variables and Interaction Terms

Due Monday, November 25, 2024

Insert Name Here

## Instructions

Answer the questions within this document. Keep the questions included below. Make sure to answer the questions in sufficient detail.

- Report all regression results in table form using **texreg**.
- Submit your rendered .pdf in Blackboard.

## Number 1: Nominal Independent Variables

Let's say you're interested in country commitments to education funding. You decide to explore whether there are differences in these commitments across regime type. Using the "world" data, you're interested in the following variables:

- **ciaedex**: This is the dependent variable. It's percent of GDP spent on education.
- **dem_level4**: Regime type (your independent variable of interest). 1=authoritarian, 2=full democracy, 3=hybrid, and 4=part democracy.
- **pop_age**: control variable, median age in years.
- **econ_compete**: control variable, global economic competitiveness; higher values more competitive (ranges from 2.85 to 5.67)

A. Use ggplot to generate a bar graph showing the average of the dependent variable as a function of regime type. Note that this should be a pre-estimation, bivariate plot. Interpret the results.

B. Estimate a multiple regression model of percent of GDP spent on education as a function of regime type and the two control variables. Remember to specify regime type as a *nominal ("dummied out") variable*. Interpret your results.

C. Now generate a post-estimation graph (using the margins package) to show the effect of regime type while controlling for age and economic competition. Interpret the graph.

## Number 2: Interactions I

In this problem, you'll use the "int1" data.

- **feminist**: Dependent variable; 101 point thermometer score for respondents' feelings toward feminists and the feminist movement; higher values indicate warmer feelings toward feminists.
- **pid**: 7 point party identification scale, where 0=strong republican, 1=weak republican, 2=independent republican, 3=independent, 4=independent democrat, 5=weak democrat, and 6=strong democrat.
- **gender**: 0=men and 1=women.

A. Estimate the following regression:

$$feminist_i = \beta_0 + \beta_1 pid_i + \beta_2 gender_i + u_i$$

Interpret the results.

B. Write out two regression equations, one for men and one for women. What's the intercept for men, and what's the intercept for women?

C. Generate a graph (using the margins package) showing the effect of party id on feminist evaluations for men and women separately. Interpret this graph. What assumption does this regression make about the gender gap in feminist evaluations? Do you think it's plausible? Why or why not?

D. Now estimate the following model, which posits an interaction between gender and party id:**

$$feminist_i = \beta_0 + \beta_1 pid_i + \beta_2 gender_i + \beta_3 pid_i \times gender_i + u_i$$

Interpret the results.

E. Write out the separate regression equations for men and women. What's the intercept for men, and what's the intercept for women? What's the slope (for pid) for men, and what's the slope (for pid) for women?

F. Generate a graph (using margins) showing the effect of party id on feminist evaluations for men and women separately. Interpret this graph. Discuss conditional effects of *both* party id and gender.

G. Now, produce a graph showing how the marginal effect of gender (the gender gap) on feminist evaluations changes as a function of party id. The y-axis is the marginal effect of gender, and the x-axis is party id. Include confidence intervals in your graph so that you can specifically assess the conditions under which gender exhibits a statistically significant effect on feminist evaluations. Interpret the graph and discuss the substantive implications of these findings.

## Number 3: Interactions II

In this problem, you'll use the "int2" data.

- **congress**: Dependent variable; 101-point thermometer scale, where larger values indicate warmer feelings toward Congress.
- **effic**: external political efficacy (extent to which one feels like their voice influences government); a 4 point scale: 0=low efficacy, 1=medium low, 2=medium high, and 3=high.
- **age**: age in years, ranges from 18-97

- **educ**: 7-point scale; higher values mean more educated.

A. Estimate the following regression model:

$$congress_i = \beta_0 + \beta_1 effic_i + \beta_2 educ_i + \beta_3 age_i + \beta_4 effic_i \times age_i + u_i$$

Interpret your results in sufficient detail.

B. Generate a Brambor et al. graph showing how the marginal effect of efficacy changes as a function of age. Interpret the graph and discuss the substantive implications of these findings.