19 Lecture 19: March 6

Last time

- HW1 review
- Dummy-Variable regression

Today

- Interactions
- Git tag demo
- Lab session review

Interactions

Two explanatory variables are said to interact in determining a response variable when the partial effect of one depends on the value of the other. Consider the hypothetical data shown in Figure 19.1.

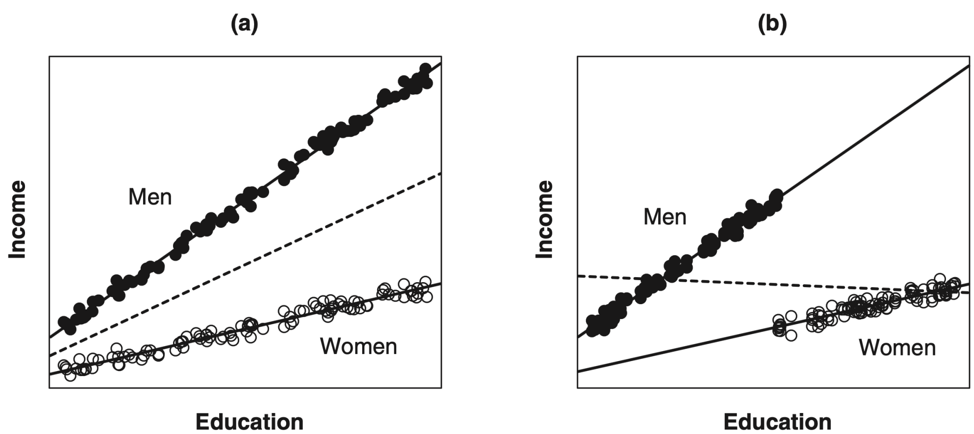

Figure 19.1: Idealized data representing the relationship between income and education for populations of men (filled circles) and women (open circles). In (a), there is no relationship between education and gender; in (b), women have a higher average level of education than men. In both (a) and (b), the within-gender (i.e., partial) regressions (solid lines) are not parallel. The slope for men is greater than the slope for women, and consequently education and gender interact in affecting income. In each graph, the overall regression of income on education (ignoring gender) is given by the broken line. JF Figure 7.7.

It is apparent in both Figure 19.1 (a) and (b) the within-gender regressions of income on education are not parallel: In both cases, the slope for men is larger than the slope for women.

Modeling interactions

We accommodate the interaction of education and gender by:

$$Y_i = \beta_0 + \beta_1 X_i + \beta_2 D_i + \beta_3 (X_i D_i) + \epsilon_i$$

where we introduce the <u>interaction regressor</u> XD into the regression equation. For women, the model becomes

$$\begin{aligned} Y_i &= \beta_0 + \beta_1 X_i + \beta_2 \cdot 0 + \beta_3 (X_i \cdot 0) + \epsilon_i \\ &= \beta_0 + \beta_1 X_i + \epsilon_i \end{aligned}$$

and for men

$$\begin{aligned} Y_i &= \beta_0 + \beta_1 X_i + \beta_2 \cdot 1 + \beta_3 (X_i \cdot 1) + \epsilon_i \\ &= (\beta_0 + \beta_2) + (\beta_1 + \beta_3) X_i + \epsilon_i \end{aligned}$$

The parameters β_0 and β_1 are, respectively, the intercept and slope for the regression of income on education among women (the baseline category for gender); β_2 gives the difference in intercepts between the male and female groups; and β_3 gives the difference in slopes between the two groups.

Usual guidance: Models that include an interaction between two predictors should also include the individual predictors by themselves regardless of the statistical significance of the associated β's.

Test for the interaction

We can simply test the hypothesis $H_0 : \beta_3 = 0$ and construct the test statistic $t = \frac{\hat{\beta}_i - 0}{\widehat{SE}(\hat{\beta}_i)} \sim t_{n-4}$ $(p = 3)$.

Interactions with multi-level factor

We can easily extend the method for modeling interactions by forming product regressors to multi-level factors, to several factors, and to several quantitative explanatory variables. Using the occupational prestige example, the occupational type could possibly interact both with income (X_1) and with education (X_2):

$$\begin{aligned} Y_i = \beta_0 &+ \beta_1 X_{i1} + \beta_2 X_{i2} + \gamma_1 D_{i1} + \gamma_2 D_{i2} \\ &+ \delta_{11} X_{i1} D_{i1} + \delta_{12} X_{i1} D_{i2} + \delta_{21} X_{i2} D_{i1} + \delta_{22} X_{i2} D_{i2} + \epsilon_i \end{aligned}$$

The model therefore permits different intercepts and slopes for the three types of occupations:

$$\begin{array}{llrrrr} \text{Professional:} & Y_i = & (\beta_0 + \gamma_1) + & (\beta_1 + \delta_{11}) X_{i1} + & (\beta_2 + \delta_{21}) X_{i2} + & \epsilon_i \\ \text{White collar:} & Y_i = & (\beta_0 + \gamma_2) + & (\beta_1 + \delta_{12}) X_{i1} + & (\beta_2 + \delta_{22}) X_{i2} + & \epsilon_i \\ \text{Blue collar:} & Y_i = & \beta_0 + & \beta_1 X_{i1} + & \beta_2 X_{i2} + & \epsilon_i \end{array}$$