EPI 809

Introducing Interactions in Multiple Linear Regression Models

The Additive Linear Model (no interactions)

$$y_i = \beta_0 + Fe_i\beta_1 + S_i\beta_2 + Edu_i\beta_3 + Exp_i\beta_4 + \varepsilon_i$$

Response
(Hourly Wage)

Intercept

Two Dummy Var.
(Female, South)

Two Continuous Covariates
(Education/Experience)

Error Term

Parameter Interpretation

Discrete:

- β_1: gender Gap (Female-Male)
- β_2: regional Gap (South-North)

Continuous:

- β_3: returns to education.
- β_4: returns to experience.

Additive model: the effects of each of the predictors does not depend on the level of other predictor (i.e., the model is linear on the predictors).

For instance, returns to education does not depend on gender or on years of experience.

Next Goal:

To extend the linear model to allow for the effect of one variable to depend on the level of another variable.

For instance we may want to allow for :

(a) the gender gap to be different across regions
(discrete×discrete)

(b) returns to education to be different between male and female
(discrete×continous)

(c) returns to experience to depend on the level of education
(continous×continous)

Finally, a simple extension of the above (c) will take us to polynomial regressions.

Discrete × Discrete

Example:

Female-Male wage gap may vary across ethnic groups.

Discrete Variables (Additive Model)

Additive Model
(no interaction)

$$y_i = \beta_0 + Fe_i\beta_1 + S_i\beta_2 + \varepsilon_i$$

Prediction Equations

	Male	Female
North	β_0	$\beta_0 + \beta_1$
South	$\beta_0 + \beta_2$	$\beta_0 + \beta_1 + \beta_2$

Gender Gap: β_1 (both in south and north)

Regional Gap: β_2 (both male and female)

What data tell us?

Average Hourly Wage by Region and Sex

Race	Sex	U$S/Hour
North	Female	8.49
	Male	10.45
South	Female	6.29
	Male	9.13

South-North Gap By Gender

	Female	Male
South-North	-2.2	-1.32

Gender Gap (Female-Male)

	U$S/Hour
South	-2.84
North	-1.96

Average Hourly Wage by Race and Sex

Race	Sex	U$S/Hour
Hispanic	Female	5.8
	Male	9.21
Non-White-Non-Hisp	Female	7.49
	Male	8.46
White	Female	8.05
	Male	10.37

Gender Gap (Female-Male)

	U$S/Hour
Hispanic	-3.41
Non-White-Non-Hisp	-0.97
White	-2.32

Race Gap By Sex

	Female	Male
Hisp-White	-2.25	-1.16
NW-NH – White	-0.56	-1.91
Hisp- NW-NH	-1.69	0.75

What do you think? Will the additive model work?

Discrete By Discrete

Additive Model
(no interaction)

$$y_i = \beta_0 + Fe_i\beta_1 + S_i\beta_2 + \varepsilon_i$$

Consider Now

$$y_i = \boldsymbol{\beta}_0 + Fe_i\boldsymbol{\beta}_1 + S_i\boldsymbol{\beta}_2 + \left(Fe_i \times S_i\right)\boldsymbol{\beta}_3 + \boldsymbol{\varepsilon}_i$$

Prediction Equations

	Male	**Female**
North	β_0	$\beta_0 + \beta_1$
South	$\beta_0 + \beta_2$	$\beta_0 + \beta_1 + \beta_2$

	Male	**Female**
North	β_0	$\beta_0 + \beta_1$
South	$\beta_0 + \beta_2$	$\beta_0 + \beta_1 + \beta_2 + \beta_3$

Gender Gap: β_1 (in both regions)

Regional Gap: β_2 (in both genders)

Gender Gap: β_1 (North) ; $\beta_1 + \beta_3$ (South)

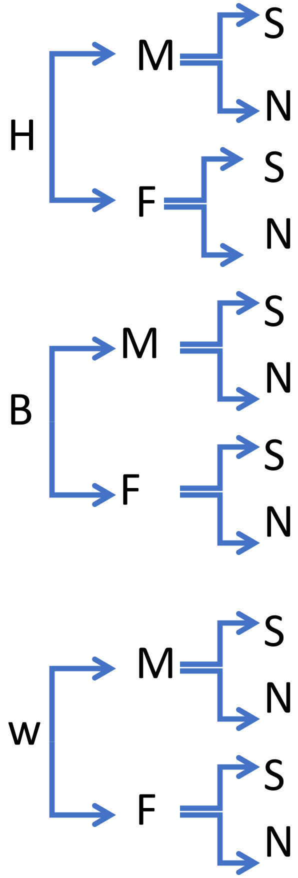
Regional Gap: β_2 (Male) ; $\beta_2 + \beta_3$(Female)

Approach: we grow our model sequentially (forward approach)

- Fit the additive model (sex + female)
- Create the Dummy variable sex×region.
- Add the dummy variables to the model (sex + region + sex×region)
- Compare R-square, what about F-test? (Long Versus Short Regression)

(Example)

3-Way Interactions



If we have 1 factors with 3 levels, and 2 factors with 2 levels, we can estimate up to 3x2x2=12 means.

The additive model has 5 parameters

$$y_i = \beta_0 + BL_i\beta_1 + HP_i\beta_2 + S_i\beta_3 + Fe_i\beta_4 + \varepsilon_i$$

If we add 1st order interactions

$$\begin{aligned} y_i = {} & \beta_0 + BL_i\beta_1 + HP_i\beta_2 + S_i\beta_3 + Fe_i\beta_4 \\ & + \left(BL_i \times S\right)\beta_5 + \left(BL_i \times Fe_i\right)\beta_5 \\ & + \left(HP_i \times S\right)\beta_7 + \left(HP_i \times Fe_i\right)\beta_8 \\ & + \left(Fe_i \times S\right)\beta_9 + \varepsilon_i \end{aligned}$$

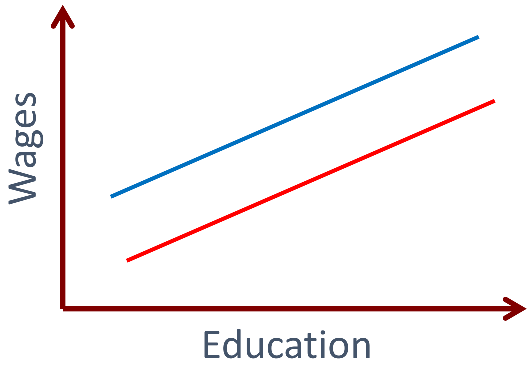
We could add 2nd order interactions (BLxSxFe) (HPxSxFe) . Such a model will have 12 parameters; therefore we will be fitting the 12 means without restrictions.

Discrete × Continuous

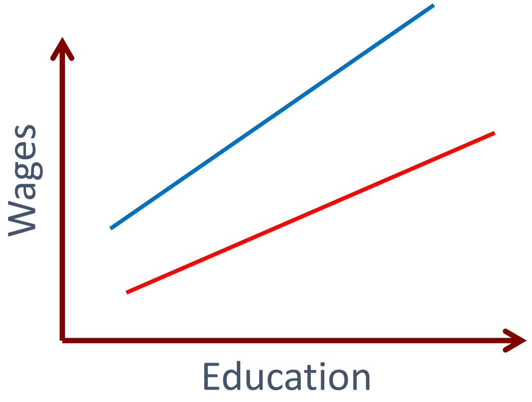
Example. Are returns to education sex-specific?

Possible Cases

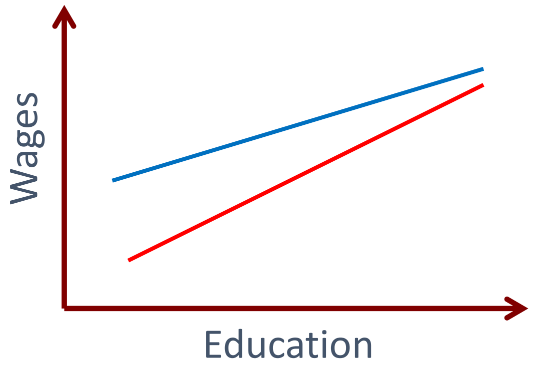
No-Interaction



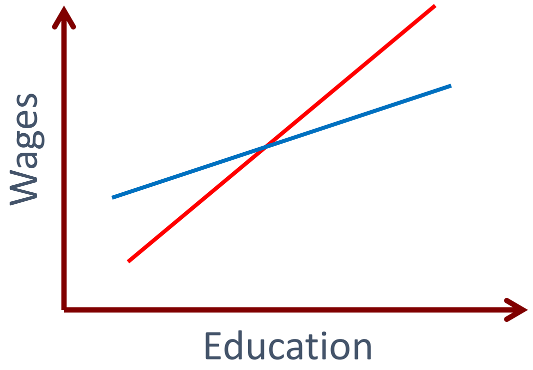
Interaction (I)



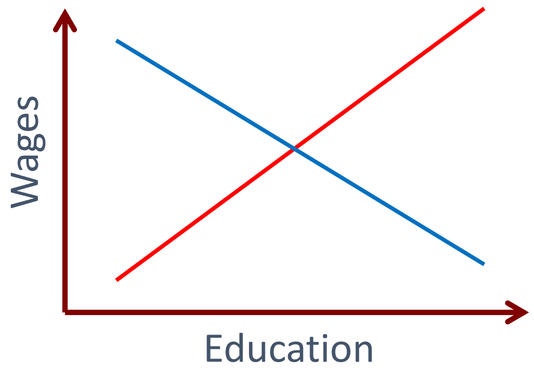
Interaction (II)



Interaction (Cross-over)



Interaction (Cross-over)



What‘s New?

$\Rightarrow$ In the additive model, slopes (returns to education) are the same for all subjects. We now want to make the slopes vary across groups (e.g., gender-specific).

$\Rightarrow$ Here is a way of having group-specific slopes

Additive Model

$$y_i = \beta_0 + Fe_i\beta_1 + Edu_i\beta_2 + \varepsilon_i$$

Model With Interaction

$$y_i = \beta_0 + Fe_i\beta_1 + Edu_i\beta_2 + (Fe_i \times Edu_i)\beta_3 + \varepsilon_i$$

Prediction Equations

Additive Model:

$$\hat{y}_i | Male = \beta_0 + Edu_i\beta_2$$

$$\hat{y}_i | Female = (\beta_0 + \beta_1) + Edu_i\beta_2$$

Model With Interaction:

$$\hat{y}_i | Male = \beta_0 + Edu_i\beta_2$$

$$\hat{y}_i | Female = (\beta_0 + \beta_1) + Edu_i\beta_2 + Edu_i\beta_3$$
$$= (\beta_0 + \beta_1) + Edu_i \times (\beta_2 + \beta_3)$$

Returns to education (differentiate the Pred. Eqs. with respect to education)

Additive Model:

β_2 (Male and Female)

Model With Interaction:

β_2 (Male)

$\beta_2 + \beta_3$ (Female)

Example

Setting: we grow our model sequentially (forward approach)

- Fit the additive model (Female+Education)
- Create the contrast Female×Education
- Add Female×Education to the model (Female+Education+ Female×Education)
- Compare R-square,
- Evaluate your estimate (magnitude, sign)
- Interpret your estimate
- Is your estimate of the interactions statistically different than zero?

Discussion

- We can model group-specific slopes by introducing as predictors in our model the product of the continuous predictor of interest (e.g., education) and the dummy variables that define the groups (e.g., female).

- The significance of the interaction term can be tested based on the p-value associated to the estimated effect of the interaction.

- So far we have covered

  Discrete × Discrete and

  Discrete × Continuous

- The same principles can be used to model interactions between two or more continuous variables (e.g. polynomial regressions).

Questions?