Interaction models
With dummy variables
With multiple category variables
With continuous variables

Multiple regression: Interaction effects

Johan A. Elkink
School of Politics & International Relations
University College Dublin

21 October 2019

Interaction models
With dummy variables
With multiple category variables
With continuous variables

1 Interaction models

With dummy variables

With multiple category variables

With continuous variables

Interaction models
With dummy variables
With multiple category variables
With continuous variables

Outline

1 Interaction models

With dummy variables

With multiple category variables

With continuous variables

Interaction models
With dummy variables
With multiple category variables
With continuous variables

Outline

1 Interaction models

With dummy variables

With multiple category variables

With continuous variables

Interaction models
With dummy variables
With multiple category variables
With continuous variables

Interactions

So far, we have only been adding variables in an **additive model**.

Imagine, however, that the relation between X and Y would depend on the group—e.g. the effect of ability on income is greater for those with a degree than those without a degree.

We call this an interaction effect, we have to **interact** the variable X with D, for example:

$$y_i = \beta_1 + \beta_2 x_i + \beta_3 d_i + \beta_4 x_i d_i.$$

Interaction models
With dummy variables
With multiple category variables
With continuous variables

Interaction with dummy variables

Model 4: $y_i = \beta_1 + \beta_2 x_i + \beta_3 d_i + \beta_4 x_i d_i$, where D is a dummy variable and X is continuous. Here there are two scenarios:

$d_i = 0$:

$$y_i = \beta_1 + \beta_2 x_i + \beta_3 \cdot 0 + \beta_4 x_i \cdot 0 = \beta_1 + \beta_2 x_i$$

and we have an intercept $\hat{\beta}_1$ and a slope coefficient $\hat{\beta}_2$ for the group where $D = 0$.

$d_i = 1$:

$$y_i = \beta_1 + \beta_2 x_i + \beta_3 \cdot 1 + \beta_4 x_i \cdot 1 = (\beta_1 + \beta_3) + (\beta_2 + \beta_4) x_i$$

and we have an intercept $\hat{\beta}_1 + \hat{\beta}_3$ and a slope coefficient $\hat{\beta}_2 + \hat{\beta}_4$ for the group where $D = 1$.

Interaction models
With dummy variables
With multiple category variables
With continuous variables

Including component variables

Note that this also shows the importance of including the component variables that make up the interaction. E.g.:

$$y_i = \beta_1 + \beta_2 d_i + \beta_3 x_i d_i,$$

where we exclude the variable X by itself, we would have:

$d_i = 0$:

$$y_i = \beta_1 + \beta_2 \cdot 0 + \beta_3 x_i \cdot 0 = \beta_1$$

and we have an intercept $\hat{\beta}_1$ and a slope coefficient 0 (!) for the group where $D = 0$.

$d_i = 1$:

$$y_i = \beta_1 + \beta_2 \cdot 1 + \beta_3 x_i \cdot 1 = (\beta_1 + \beta_2) + \beta_3 x_i$$

and we have an intercept $\hat{\beta}_1 + \hat{\beta}_2$ and a slope coefficient $\hat{\beta}_3$ for the group where $D = 1$.

So we **arbitrarily fix one slope** to zero.

Including component variables

Or similarly:

$$y_i = \beta_1 + \beta_2 x_i + \beta_3 x_i d_i,$$

where we exclude the dummy variable D by itself:

$d_i = 0$:

$$y_i = \beta_1 + \beta_2 x_i + \beta_3 x_i \cdot 0 = \beta_1 + \beta_2 x_i$$

and we have an intercept $\hat{\beta}_1$ and a slope coefficient $\hat{\beta}_2$ for the group where $D = 0$.

$d_i = 1$:

$$y_i = \beta_1 + \beta_2 x_i + \beta_3 x_i \cdot 1 = \beta_1 + (\beta_2 + \beta_3) x_i$$

and we have an intercept $\hat{\beta}_1$ and a slope coefficient $\hat{\beta}_2 + \hat{\beta}_3$ for the group where $D = 1$.

So we **fix the value of Y to be identical** for the two groups at the **arbitrary point of $X = 0$**.

Interaction models
With dummy variables
With multiple category variables
With continuous variables

Interaction models and t-tests

$$y_i = \beta_1 + \beta_2 x_i + \beta_3 d_i + \beta_4 x_i d_i$$

So we can think of the following t-tests:

$H_0 : \beta_2 = 0$, so under the null, the slope of the line is zero, *for the group where* $D = 0$.

$H_0 : \beta_3 = 0$, so under the null, the two groups have the same intercept.

In a regression with an interaction with a dummy variable, the t-test for that coefficient tests whether, given the other variables in the model, the slope for the two groups differ.

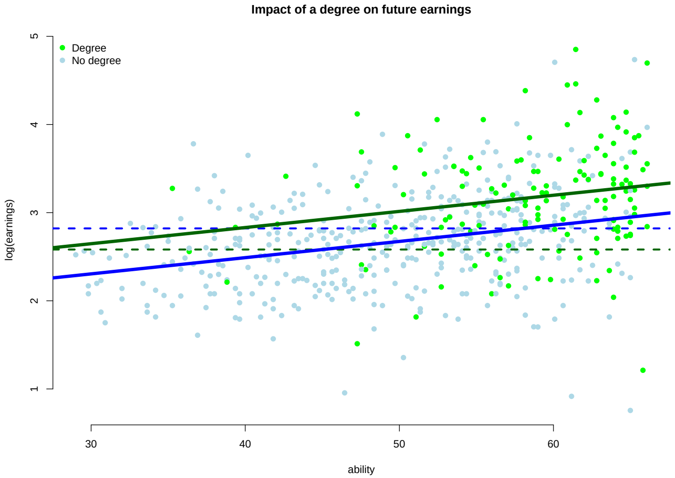
$H_0 : \beta_4 = 0$, so under the null, the two groups have the same slope between X and Y.

Interaction models
With dummy variables
With multiple category variables
With continuous variables

Example: degree and earnings

degree	0.340***	0.345
	(0.058)	(0.428)
ability	0.018***	0.018***
	(0.003)	(0.003)
degree × ability		−0.0001
		(0.007)
intercept	1.754***	1.753***
	(0.140)	(0.153)
N	540	540
R^2	0.204	0.204
Adjusted R^2	0.201	0.200
Residual Std. Error	0.531	0.531
F-Statistic	68.882***	45.836***
Note:	*p<0.1; **p<0.05; ***p<0.01	

Example: degree and earnings

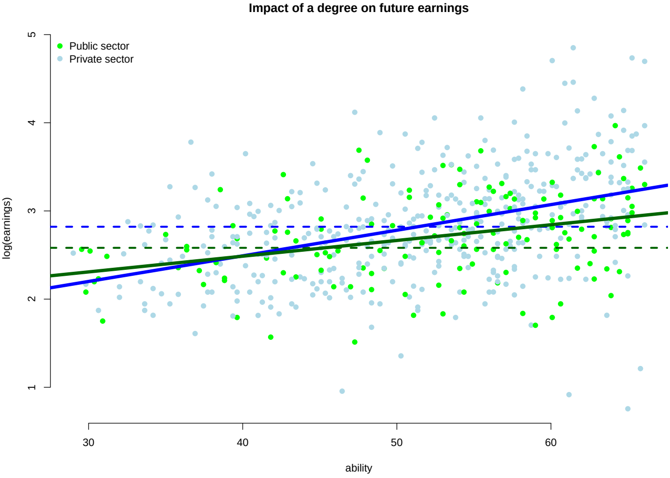


$$\log(earnings_i) = \beta_1 + \beta_2 degree_i + \beta_3 ability_i + \beta_4 degree_i \cdot ability_i$$

Example: public sector and earnings

	(1)	(2)
publicSector	-0.141^{***}	0.445
	(0.053)	(0.300)
ability	0.026^{***}	0.029^{***}
	(0.003)	(0.003)
publicSector × ability		-0.011^{**}
		(0.006)
intercept	1.496^{***}	1.329^{***}
	(0.135)	(0.159)
N	540	540
R^2	0.163	0.169
Adjusted R^2	0.160	0.165
Residual Std. Error	0.544	0.543
F-Statistic	52.418^{***}	36.444^{***}
Note:	*p<0.1; **p<0.05; ***p<0.01	

Example: public sector and earnings



$$\log(earnings_i) = \beta_1 + \beta_2 publicSector_i + \beta_3 ability_i + \beta_4 publicSector_i \cdot ability_i$$

Outline

1 Interaction models

With dummy variables

With multiple category variables

With continuous variables

Interaction models
With dummy variables
With multiple category variables
With continuous variables

Example: race and earnings

$$\log(earnings_i) = \beta_1 + \beta_2 black_i + \beta_3 hisp_i + \beta_4 ability_i + \beta_5 black_i \cdot ability_i + \beta_6 hisp_i \cdot ability_i,$$

Whites: $\log(earnings_i) = \beta_1 + \beta_4 ability_i$

Blacks: $\log(earnings_i) = (\beta_1 + \beta_2) + (\beta_4 + \beta_5) ability_i$

Hispanics: $\log(earnings_i) = (\beta_1 + \beta_3) + (\beta_4 + \beta_6) ability_i$

So β_2 and β_3 are differences in intercepts, relative to whites; β_5 and β_6 are differences in slopes, relative to whites and t-tests test whether intercepts or slopes, respectively, differ.

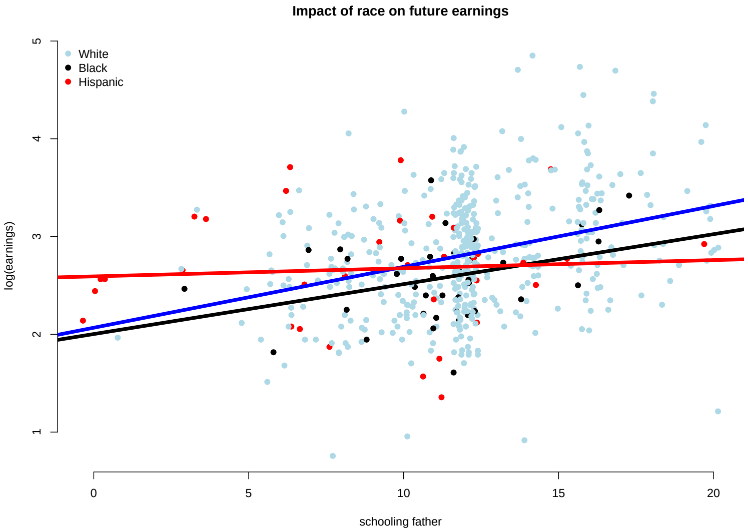
Interaction models
With dummy variables
With multiple category variables
With continuous variables

Example: race and earnings

ethblack	-0.198^{**}	-0.065
	(0.094)	(0.395)
ethhisp	0.022	0.525^{**}
	(0.103)	(0.229)
schoolingFather	0.054^{***}	0.062^{***}
	(0.008)	(0.008)
ethblack × schoolingFather		-0.011
		(0.034)
ethhisp × schoolingFather		-0.054^{**}
		(0.022)
intercept	2.164^{***}	2.067^{***}
	(0.095)	(0.104)
N	540	540
R^2	0.101	0.111
Adjusted R^2	0.096	0.103
Residual Std. Error	0.565	0.563
F-Statistic	20.011^{***}	13.312^{***}
Note:	*p<0.1; **p<0.05; ***p<0.01	

Interaction models
With dummy variables
With multiple category variables
With continuous variables

Example: race and earnings



Interaction models
With dummy variables
With multiple category variables
With continuous variables

Outline

1 Interaction models

With dummy variables

With multiple category variables

With continuous variables

Interaction models
With dummy variables
With multiple category variables
With continuous variables

Interactions between continuous variables

It is possible to interact two continuous variables. Here you expect the effects of X on Y to gradually change as some third variable Z changes.

$$y_i = \beta_1 + \beta_2 x_i + \beta_3 z_i + \beta_4 x_i z_i,$$

so when we take X as the key independent variable, we have:

Intercept:	$\beta_1 + \beta_3 z_i$
Slope:	$\beta_2 + \beta_4 z_i$

Both intercept and slope change with Z. These types of models are typically somewhat difficult to interpret and there is no statistical difference between whether the slope between X and Y varies for different values of Z, or the slope between Z and Y varies for different values of X. It requires a strong theory on causal relations to be able to make sense of the results.