Advanced Grad Stats
Interaction Effects

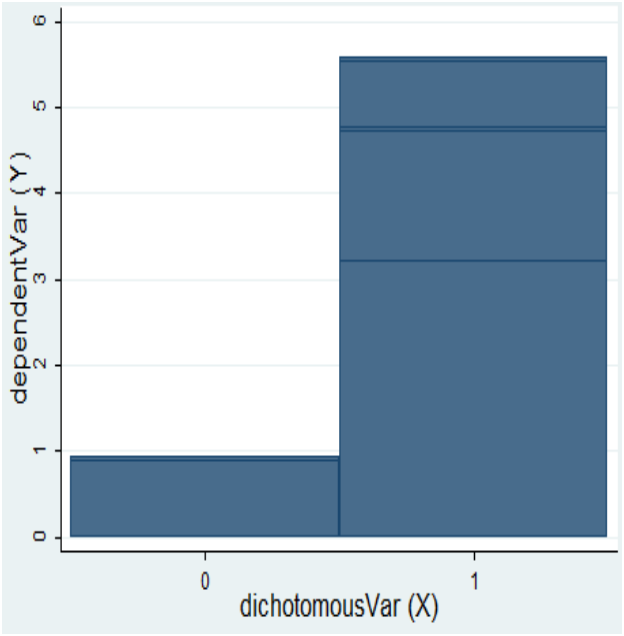
Evidence from 12 Latin American Countries

Lab 7

Diego F. Leal

www.diegoleal.info

A Dichotomous/Categorical Independent Variable




```
. reg dependentVar i1.dichotomousVar
```

Source	SS	df	MS
Model	45.6895343	1	45.6895343
Residual	636.946318	998	.638222764
Total	682.635852	999	.683319172

Number of obs	= 1000
F(1, 998)	= 71.59
Prob > F	= 0.0000
R-squared	= 0.0669
Adj R-squared	= 0.0660
Root MSE	= .79889

dependentVar	Coef.	Std. Err.	t	P>|t|	[95% Conf.	Interval]
1.dichotomou~r	2.263343	.2675026	8.46	0.000	1.73841	2.788275
_cons	.7485708	.2662962	2.81	0.005	.2260062	1.271135

Here a “unit increase” represents switching from one category to another. The coefficient is the average difference in Y between the category for which X = 0 (the reference group) and the category for which X = 1 (the comparison group)

A Dichotomous/Categorical Independent Variable

```
. reg dependentVar i1.dichotomousVar
```

Source	SS	df	MS
Model	45.6895343	1	45.6895343
Residual	636.946318	998	.638222764
Total	682.635852	999	.683319172

Number of obs	= 1000
F(1, 998)	= 71.59
Prob > F	= 0.0000
R-squared	= 0.0669
Adj R-squared	= 0.0660
Root MSE	= .79889

dependentVar	Coef.	Std. Err.	t	P>|t|	[95% Conf.	Interval]
1.dichotomou~r	2.263343	.2675026	8.46	0.000	1.73841	2.788275
_cons	.7485708	.2662962	2.81	0.005	.2260062	1.271135

Summary statistics: mean
by categories of: dichotomousVar

dichotomousVar	depend~r
0	.7485708
1	3.011913
Total	2.991543

Therefore, compared to the reference group, we would expect people in the comparison group to score 2.26 units higher on the dependent variable, on average.

Additive Model

This is our initial model:

$$E(I) = \alpha + \beta 1\ \text{Education} + \beta 2\ \text{Urban} + \epsilon$$

Where,

I = respondent's wealth

α = slope

β1 = Respondent's education → Main effect 1

β2 = Respondent's area of residence (urban = 1)→ Main effect 2

ϵ = error

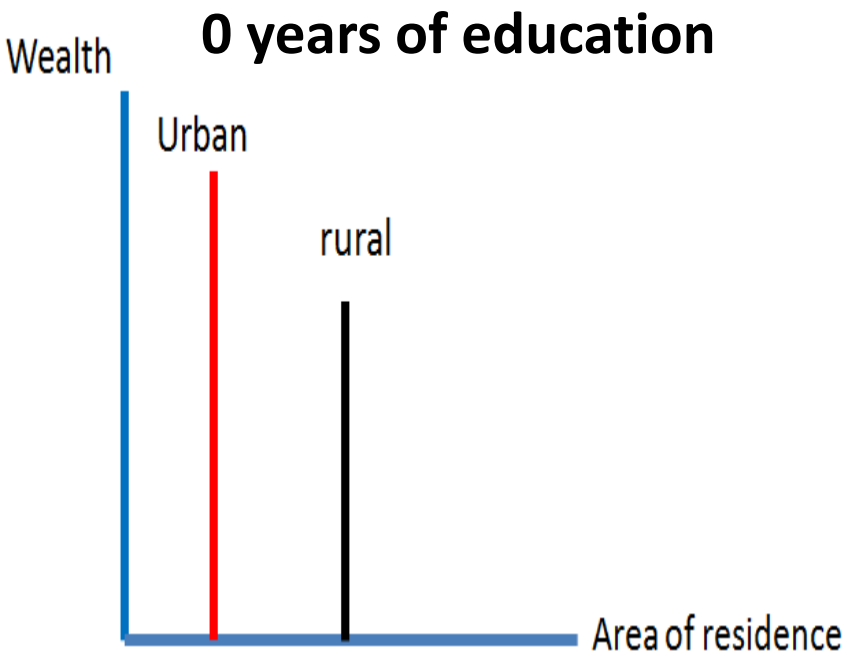
Additive Model: Stata Output, Evidence from Latin America

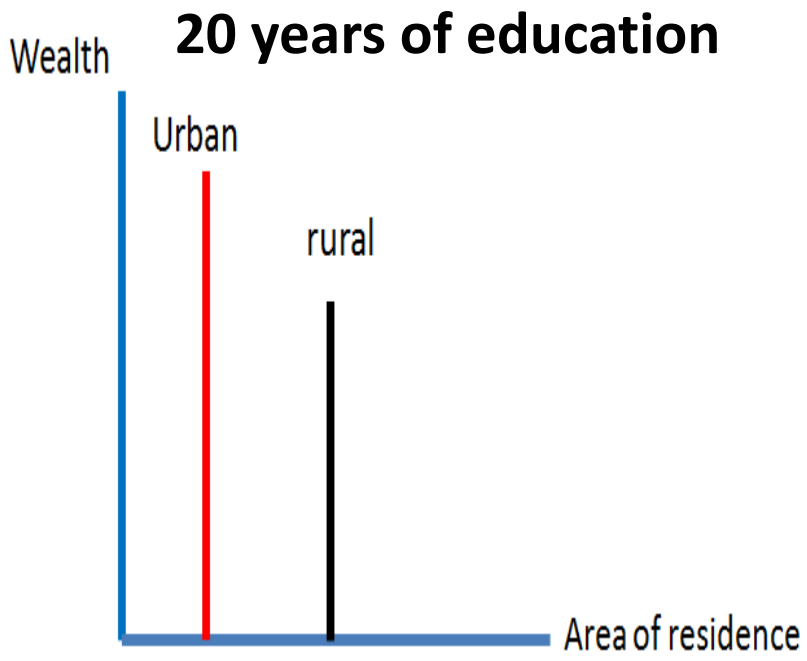
```
. xi: svy: reg wealthIndex i.urban education
i.urban           _Iurban_0-1         (naturally coded; _Iurban_0 omitted)
(running regress on estimation sample)

Survey: Linear regression

Number of strata   =         6                  Number of obs      =      8972
Number of PSUs     =       125                  Population size    =      8972
                                                Design df          =       119
                                                F(   2,    118)    =    343.34
                                                Prob > F           =    0.0000
                                                R-squared          =    0.3173
```

wealthIndex	Coef.	Linearized Std. Err.	t	P>|t|	[95% Conf.	Interval]
_Iurban_1	.2116932	.0138543	15.28	0.000	.1842603	.2391261
education	.029278	.0012547	23.33	0.000	.0267935	.0317624
_cons	-.1211798	.0121518	-9.97	0.000	-.1452415	-.0971181

Model Assumptions

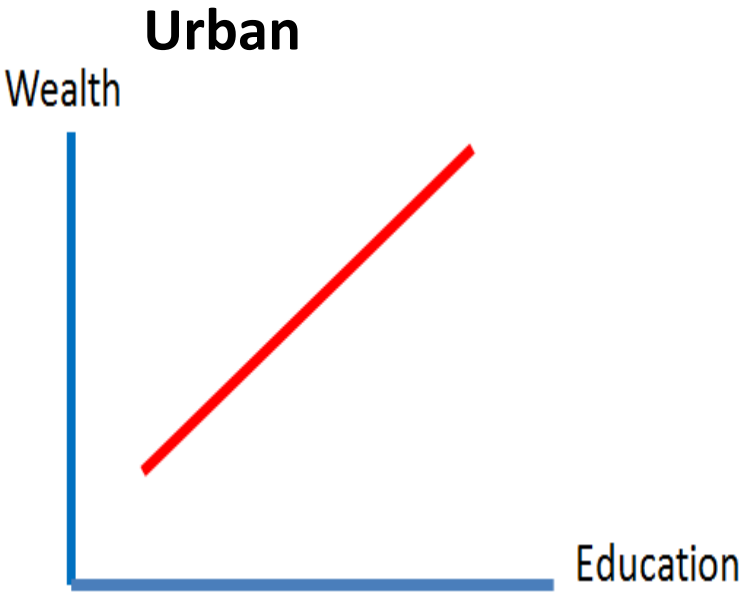
Model assumes the difference in wealth between rural and urban residents is constant across **all** levels of education

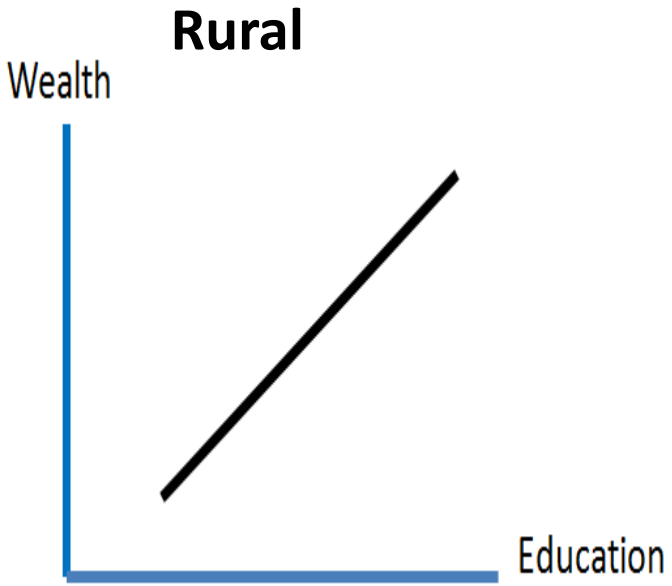




Model Assumptions

Model assumes the slope of education is the same for urban and rural residents.





Non-Additive Way of Thinking

Slope line for education could depend on the group (rural vs. urban residents) one is analyzing

Categorical variable as the moderator

Difference in income between urban and rural residents could depend on the amount of education people have

Continuous variable as the moderator

Definition

An interaction exists when the relationship between X and Y changes (in magnitud or direction) when examined at different levels of a third variable

Non-Additive Model

This is our initial model:

$$E(I) = \alpha + \beta 1\ \text{Education} + \beta 2\ \text{Urban} + \beta 3\ \text{Urban*Education} + \epsilon$$

Where,

I = respondent's wealth

α = slope

β1 = Respondent's education → conditional effect 1

β2 = Respondent's area of residence (urban = 1)→ conditional effect 2

β3 = Respondent's area of residence (urban = 1) * respondent's education → cross-product term

ϵ = error

Something to Consider First

```
. xi: svy: reg wealthIndex education if urban == 0
(running regress on estimation sample)

Survey: Linear regression

Number of strata   =         5                  Number of obs      =      2349
Number of PSUs     =        70                  Population size    =      2349
                                                Design df          =        65
                                                F(   1,     65)    =     61.47
                                                Prob > F           =    0.0000
                                                R-squared          =    0.0737
```

wealthIndex	Linearized Coef.	Std. Err.	t	P>|t|	[95% Conf.	Interval]
education	.0109204	.0013929	7.84	0.000	.0081386	.0137023
_cons	.0030168	.0086096	0.35	0.727	-.0141777	.0202113

Rural residents' education effect (slope) = **0.0109**

Rural residents' slope is not as steeper as urban residents' slope

```
. xi: svy: reg wealthIndex education if urban == 1
(running regress on estimation sample)

Survey: Linear regression

Number of strata   =         6                  Number of obs      =      6623
Number of PSUs     =       124                  Population size    =      6623
                                                Design df          =       118
                                                F(   1,    118)    =    640.78
                                                Prob > F           =    0.0000
                                                R-squared          =    0.2102
```

wealthIndex	Linearized Coef.	Std. Err.	t	P>|t|	[95% Conf.	Interval]
education	.0349324	.00138	25.31	0.000	.0321996	.0376651
_cons	.0336815	.0169335	1.99	0.049	.0001485	.0672144

Urban residents' education effect (slope) = **0.0349**

Wealth returns to education are stronger for urban residents

Interpreting the Results

```
. xi: svy: reg wealthIndex i.urban education urban_education
i.urban            _Iurban_0-1          (naturally coded; _Iurban_0 omitted)
(running regress on estimation sample)

Survey: Linear regression

Number of strata   =         6                  Number of obs      =      8972
Number of PSUs     =       125                  Population size    =      8972
                                                Design df          =       119
                                                F(   3,     117)   =    333.75
                                                Prob > F           =    0.0000
                                                R-squared          =    0.3352
```

wealthIndex	Linearized Coef.	Std. Err.	t	P>|t|	[95% Conf.	Interval]
_Iurban_1	.0306647	.0176263	1.74	0.084	-.0042373	.0655666
education	.0109204	.0014191	7.70	0.000	.0081105	.0137304
urban_education	.0240119	.0019381	12.39	0.000	.0201743	.0278496
_cons	.0030168	.0085909	0.35	0.726	-.0139941	.0200277

- **Education:** effect of education for rural residents (reference group)
- **_Iurban_1:** predicted income gap between rural and urban residents when education is zero in both groups
- **Urban_education**: part of education effect for urban residents

The Math

$E(Y) = \alpha + \beta 1\,(\text{Education}) + \beta 2\,(\text{Urban}) + \beta 3\,(\text{Urban*Education}) + \epsilon$

$E(I) = .003 + .01\,(\text{Education}) + .03\,(\text{Urban}) + .02\,(\text{Urban*Education})$

Substitute 1, 2 & 3 as possible values for education and the interaction; and 0 for urban

$E(I) = .003 + .01\,(1) + .03\,(0) + .02\,(0) = 0.4$

$E(I) = .003 + .01\,(2) + .03\,(0) + .02\,(0) = 0.5$

$E(I) = .003 + .01\,(3) + .03\,(0) + .02\,(0) = 0.6$

The difference between predicted values is (e.g. 0.5 - 0.4) =.01

So, .01 is the change in slope for rural residents → increment in wealth for rural risdents for one unit increase in education

The Math

$E(Y) = \alpha + \beta 1\,(\text{Education}) + \beta 2\,(\text{Urban}) + \beta 3\,(\text{Urban*Education}) + \epsilon$

$E(I) = .003 + .01\,(\text{Education}) + .03\,(\text{Urban}) + .02\,(\text{Urban*Education})$

Substitude 1, 2 & 3 for education; 1 urban; and 1, 2 & 3 for the interaction

$E(I) = .003 + .01\,(1) + .03\,(1) + .02\,(1) = 0.9$

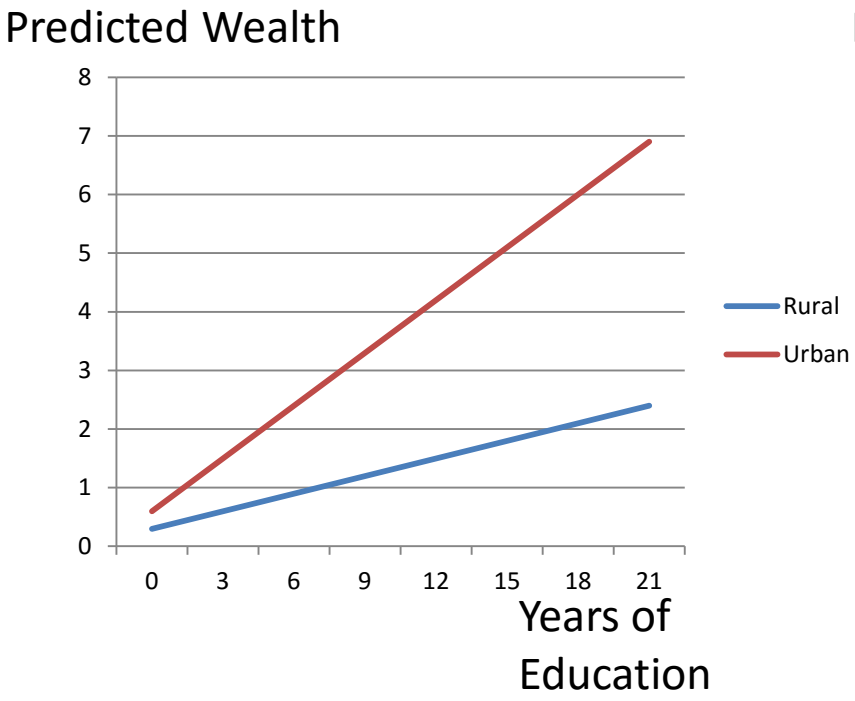
$E(I) = .003 + .01\,(2) + .03\,(1) + .02\,(2) = 1.2$

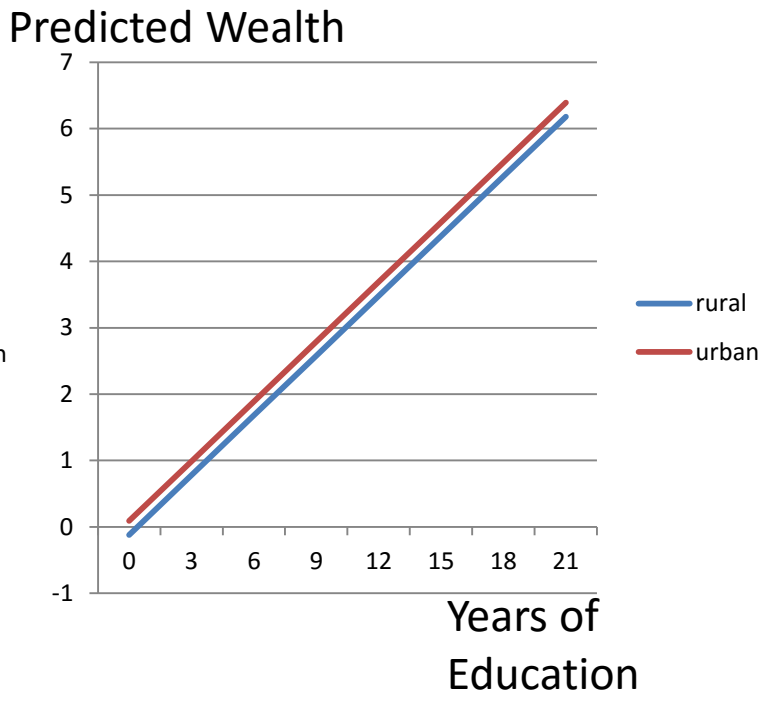
$E(I) = .003 + .01\,(3) + .03\,(1) + .02\,(3) = 1.5$

The difference between predicted values is (e.g. 1.2 - 0.9) = 0.3

So, .03 is the change in slope for urban residents → increment in wealth for urban residents for one unit increase in education

A Graphic Result: Non-Additive vs. Additive





Conclusion

When people have more education, the urban vs. rural gap in wealth is large.

If you live in a rural area, it does not pay that much to get more education (when compared to the situation in urban areas).

Given the size of the coefficients and their signs, a positive coefficient for the interaction means both slopes for area of residence are positive.