Applied Statistics and Econometrics

Giuseppe Ragusa
Luiss University

Lecture 10: Nonlinear Regression Functions: Interactions between independent variables

Interaction between independent variables

- Perhaps a class size reduction is more effective in some circumstances than in others...
- Perhaps smaller classes help more if there are many English learners, who need individual attention
- That is,

$$\frac{\Delta testscore}{\Delta str}$$

  might depend on *English* (the effect of class size may be different depending on the fraction of english learners in the given school)

Interaction between independent variables

- More generally, the effect of

$$\frac{\Delta Y}{\Delta X_1}$$

  might depend on X_2

  - How to model such "**interactions**" between X_1 and X_2?
  - We first consider binary X's, then continuous X's

Interaction between two binary variables

$$Y_i = \beta_0 + \beta_1 D_{1i} + \beta_2 D_{2i} + u_i$$

- D_{1i} and D_{2i} are binary
- β_1 is the effect of changing $D_1 = 0$ to $D_1 = 1$. In this specification, this effect does not depend on the value of D_2
- To allow the effect of changing D_1 to depend on D_2, include the "**interaction term**"

$$D_{1i} \times D_{2i}$$

as a regressor:

$$Y_i = \beta_0 + \beta_1 D_{1i} + \beta_2 D_{2i} + \beta_3 (D_{1i} \times D_{2i}) + u_i$$

Interpreting the coefficient

$$Y_i = \beta_0 + \beta_1 D_{1i} + \beta_2 D_{2i} + \beta_3 (D_{1i} \times D_{2i}) + u_i$$

- General rule: compare the various cases
  - $\underbrace{E(Y_i|D_{1i}=0, D_{2i}=d_2) = \beta_0 + \beta_2 d_2}_{(b)}$
  - $\underbrace{E(Y_i|D_{1i}=1, D_{2i}=d_2) = \beta_0 + \beta_1 + \beta_2 d_2 + \beta_3 d_2}_{(a)}$
- (b) - (a):

$$E(Y_i|D_{1i}=1, D_{2i}=d_2) - E(Y_i|D_{1i}=0, D_{2i}=d_2) = \beta_1 + \beta_3 d_2$$

The effect of D1 depends on d2 (what we wanted)

$$\beta_3 = \text{ increment to the effect of } D_1\text{, when } D_2 = 1$$

Example: TestScore, STR, English learners

Let

$$histr = \begin{cases} 0 & \text{if } str < 20 \\ 1 & \text{if } str \geqslant 20 \end{cases}, \qquad hienglish = \begin{cases} 0 & \text{if } english < 10 \\ 1 & \text{if } english \geqslant 10 \end{cases}$$

```r
## We create the dummy veriables using 'transform'
CASchools <- transform(CASchools, histr = ifelse(str >= 20, 1, 0),
    hienglish = ifelse(english >= 10, 1, 0))
## 'ifelse(cond, a, b)' takes a condition, eg. histr>=20, and
## return 'a' if the condition is true, 'b' otherwise
```

Example: TestScore, STR, English learners

```
lm1 <- lm(testscore ~ histr + hienglish + I(histr * hienglish), data = CASchools)
summary_rob(lm1)
```

```
##
## Coefficients:
##                       Estimate Std. Error z value Pr(>|z|)
## (Intercept)             664.14       1.39  478.46  < 2e-16
## histr                    -1.91       1.93   -0.99     0.32
## hienglish               -18.32       2.33   -7.85  4.3e-15
## I(histr * hienglish)     -3.26       3.12   -1.05     0.30
## ---
## Heteroskadasticity robust standard errors used
##
## Residual standard error: 16 on 416 degrees of freedom
## Multiple R-squared:  0.295,Adjusted R-squared:  0.29
## F-statistic:  179 on 3 and Inf DF,  p-value: <2e-16
```

Example: TestScore, STR, English learners

$$testscore = \underset{(1.4)}{664.1} - \underset{(1.9)}{1.9}\, histr - \underset{(2.3)}{18.3}\, hienglish - \underset{(3.1)}{3.3}\, histr * hienglish$$

- "Effect" of *histr* when $hienglish = 0$ is –1.9
- "Effect" of *histr* when $hienglish = 1$ is –1.9–3.3 = –5.2
- Class size reduction is estimated to have a **bigger** effect when the percent of English learners is **large**
- This interaction isn't statistically significant: $t = -3.3/3.1 \approx 1$

Example: testscore, str, english learners

$$histr = \begin{cases} 0 & \text{if } str < 20 \\ 1 & \text{if } str \geqslant 20 \end{cases}, \qquad hienglish = \begin{cases} 0 & \text{if } english < 10 \\ 1 & \text{if } english \geqslant 10 \end{cases}$$

$$testscore = \underset{(1.4)}{664.1} - \underset{(1.9)}{1.9}\, histr - \underset{(2.3)}{18.3}\, hienglish - \underset{(3.1)}{3.3}\, histr * hienglish$$

Can you relate these coefficients to the following table of group ("cell") means?

	low str	high str
low english	664.14	662.24
high english	645.83	640.66

Interactions between continuous and binary variables

$$Y_i = \beta_0 + \beta_1 D_i + \beta_2 X_i + u_i$$

- D_i is binary, X is continuous
- As specified above, the effect on Y of X (holding constant D) $= \beta_2$, which does not depend on D
- To allow the effect of X to depend on D, include the "interaction term"

$$D_i \times X_i$$

as a regressor:

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

Binary-continuous interactions: the two regression lines

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

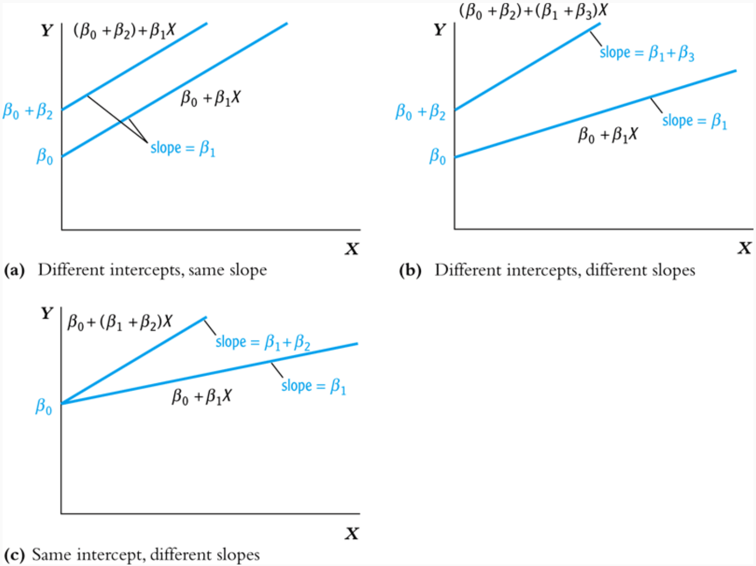
- Observations with $D_i = 0$ (the "$D = 0$" group):

$$Y_i = \beta_0 + \beta_2 X_i + u_i$$

- Observations with $D_i = 1$ (the "$D = 1$" group):

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

Binary-continuous interactions, ctd.



Interpreting the coefficients

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

- General rule: compare the various cases

$$\underbrace{E[Y_i|X_i = x + \Delta x, D = d] = \beta_0 + \beta_1 d + \beta_2 (x + \Delta x) + \beta_3 [d \times (x + \Delta x)]}_{(b)}$$

$$\underbrace{E[Y_i|X_i = x, D = d] = \beta_0 + \beta_1 d + \beta_2 x + \beta_3 (d \times x)}_{(a)}$$

- subtract (a)–(b):

$$E[Y_i|X_i = x + \Delta x, D = d] - E[Y_i|X_i = x, D = d] = \beta_2 + \beta_3 d$$

- The effect of X depends on D (what we wanted)

$$\beta_3 = \text{increment to the effect of } X, \text{ when } D = 1$$

Example: testscore, str, hienglish (=1 if english$\geq$10)

```
lm2 <- lm(testscore ~ str + hienglish + I(str * hienglish), data = CASchools)
summary_rob(lm2)
```

```
##
## Coefficients:
##                     Estimate Std. Error z value Pr(>|z|)
## (Intercept)          682.246     11.868   57.49   <2e-16
## str                   -0.968      0.589   -1.64     0.10
## hienglish              5.639     19.515    0.29     0.77
## I(str * hienglish)    -1.277      0.967   -1.32     0.19
## ---
## Heteroskadasticity robust standard errors used
##
## Residual standard error: 16 on 416 degrees of freedom
## Multiple R-squared:  0.31,Adjusted R-squared:  0.305
## F-statistic:  191 on 3 and Inf DF,  p-value: <2e-16
```

Example: testscore, str, hienglish (=1 if english≥10)

$$\underset{(11.87)}{testscore = 682.25} - \underset{(0.59)}{0.97}\, str + \underset{(19.52)}{5.64}\, hienglish - \underset{(0.97)}{1.28}\, str * hienglish$$

- When $hienglish = 0$

$$\widehat{testscore} = 682.2 - 0.97 \times str$$

- When $hienglish = 1$

$$\widehat{testscore} = 682.2 - 0.97 \times str + 5.6 - 128 \times str$$
$$= 687.8 - 2.25 \times str$$

- Two regression lines: one for each *hienglish* group.
- Class size reduction is estimated to have a **larger** effect when the percent of English learners is **large**.

Example, ctd: testing hypotheses

$$testscore = \underset{(11.87)}{682.25} - \underset{(0.59)}{0.97}\, str + \underset{(19.52)}{5.64}\; hienglish - \underset{(0.97)}{1.28}\, str * hienglish$$

- The two regression lines have the same slope ⇔ the coefficient on $str \times \times hienglish$ is zero:

$$t = –1.28/0.97 = –1.32$$

- The two regression lines have the same intercept ⇔ the coefficient on *hienglish* is zero:

$$t = –5.6/19.5 = 0.29$$

- The two regression lines are the same ⇔ population coefficient on $hienglish = 0$ and population coefficient on $str \times hienglish = 0$:

$$W = 89.94, \quad (p - value < .001)$$

- We reject the joint hypothesis but neither individual hypothesis (how can this be?)

Interactions between two continuous variables

$$Y_i = \beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + u_i$$

- X_1, X_2 are continuous
- As specified, the effect of X_1 doesn't depend on X_2
- As specified, the effect of X_2 doesn't depend on X_1
- To allow the effect of X_1 to depend on X_2, include the "interaction term" $X_{1i} \times X_{2i}$ as a regressor:

$$Y_i = \beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + \beta_3 X_{1i} \times X_{2i} + u_i$$

Interpreting the coefficients:

$$Y_i = \beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + \beta_3 X_{1i} \times X_{2i} + u_i$$

- General rule: compare the various cases

$$\underbrace{E[Y_i|X_{1i} = x_1 + \Delta x_1, X_{2i} = x_2] = \beta_0 + \beta_1(x_1 + \Delta x_1) + \beta_2 x_2 + \beta_3[(x_1 + \Delta x_1) \times x_2]}_{(b)}$$

$$\underbrace{E[Y_i|X_{1i} = x_1, X_{2i} = x_2] = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_1 \times x_2}_{(a)}$$

- subtract (a)–(b):

$$E[Y_i|X_{1i} = x_1 + \Delta x_1, X_{2i} = x_2] - E[Y_i|X_{1i} = x_1, X_{2i} = x_2] = \beta_2 + \beta_3 x_2$$

- The effect of X_1 depends on X_2 (what we wanted)

$$\beta_3 = \text{increment to the effect of } X_1 \text{from a unit change in } X_2$$

Example: testscore, str, english

```
lm3 <- lm(testscore ~ str + english + I(english * str), data = CASchools)
summary_rob(lm3)
```

```
##
## Coefficients:
##                   Estimate Std. Error z value Pr(>|z|)
## (Intercept)      686.33853   11.75935   58.37   <2e-16
## str               -1.11702    0.58751   -1.90    0.057
## english           -0.67291    0.37412   -1.80    0.072
## I(english * str)   0.00116    0.01854    0.06    0.950
## ---
## Heteroskadasticity robust standard errors used
##
## Residual standard error: 14 on 416 degrees of freedom
## Multiple R-squared:  0.426,Adjusted R-squared:  0.422
## F-statistic:  465 on 3 and Inf DF,  p-value: <2e-16
```

Example: testscore, str, english

$$\mathit{testscore} = \underset{(11.759)}{686.339} - \underset{(0.588)}{1.117}\,\mathit{str} - \underset{(0.374)}{0.673}\,\mathit{english} + \underset{(0.019)}{0.001}\,\mathit{english} * \mathit{str}$$

The estimated effect of class size reduction is nonlinear because the size of the effect itself depends on *english*:

english	$\frac{\Delta \mathit{testscore}}{\Delta \mathit{str}}$
0	$-1.12 + 0 * 0.00 = -1.12$
10%	$-1.12 + 10 * 0.00 = -1.11$
30%	$-1.12 + 30 * 0.00 = -1.08$
...	
90%	$-1.12 + 90 * 0.00 = -1.01$

Application I: Nonlinear Effects on Test Scores of the Student-Teacher Ratio

Nonlinear specifications let us examine more nuanced questions about the *testscore* – *str* relation, such as:

1. Are there nonlinear effects of class size reduction on test scores? (Does a reduction from 35 to 30 have same effect as a reduction from 20 to 15?)
2. Are there nonlinear interactions between *english* and *str*? (Are small classes more effective when there are many English learners?)

Strategy for Question #1 (different effects for different *str*?)

- Estimate linear and nonlinear functions of STR, holding constant relevant demographic variables
  - *english*
  - *income*(remember the nonlinear $testscore - income$ relation!)
  - *lunch* (fraction on free/subsidized lunch)
- See whether adding the nonlinear terms makes an "economically important" quantitative difference ("economic" or "real-world" importance is different than statistically significant)
- Test for whether the nonlinear terms are significant

Strategy for Question #2 (interactions between *english* and *str*?)

- Estimate linear and nonlinear functions of STR, interacted with English.
- If the specification is nonlinear (with str, str^2, str^3), then you need to add interactions with all the terms so that the entire functional form can be different, depending on the level of *english*.
- We will use a binary-continuous interaction specification by adding $hienglish \times str$, $hienglish \times str^2$, and $hienglish \times str^3$

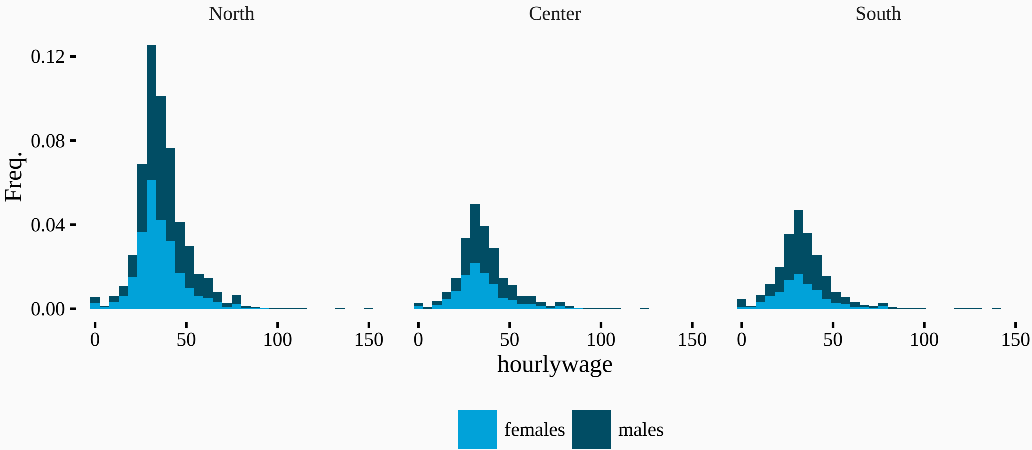
Regressions

	Dependent variable:						
	testscore						
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
str	-1.000^{***}	-0.730^{***}	-0.970^{*}	-0.530^{*}	64.000^{**}	84.000^{***}	65.000^{**}
	(0.240)	(0.230)	(0.540)	(0.300)	(25.000)	(30.000)	(25.000)
english	-0.120^{***}	-0.180^{***}					-0.170^{***}
	(0.032)	(0.032)					(0.032)
str^2					-3.400^{***}	-4.400^{***}	-3.500^{***}
					(1.300)	(1.500)	(1.300)
str^3					0.059^{***}	0.075^{***}	0.060^{***}
					(0.022)	(0.025)	(0.022)
lunch	-0.550^{***}	-0.400^{***}		-0.410^{***}	-0.420^{***}	-0.420^{***}	-0.400^{***}
	(0.022)	(0.030)		(0.029)	(0.028)	(0.029)	(0.030)
$\log(income)$		12.000^{***}		12.000^{***}	12.000^{***}	12.000^{***}	12.000^{***}
		(1.700)		(1.800)	(1.700)	(1.800)	(1.700)
$english \geq 20$			5.600	5.500	-5.500^{***}	816.000^{*}	
			(17.000)	(9.100)	(1.000)	(435.000)	
$str \times (english \geq 20)$			-1.300	-0.580		-123.000^{*}	
			(0.840)	(0.470)		(66.000)	
$str^2 \times (english \geq 20)$						6.100^{*}	
						(3.400)	
$str^3 \times (english \geq 20)$						-0.100^{*}	
						(0.056)	
Constant	700.000^{***}	659.000^{***}	682.000^{***}	654.000^{***}	252.000	122.000	245.000
	(4.700)	(7.700)	(11.000)	(8.900)	(166.000)	(192.000)	(166.000)
Observations	420	420	420	420	420	420	420
R^2	0.780	0.800	0.310	0.800	0.800	0.800	0.800
Adjusted R^2	0.770	0.790	0.300	0.800	0.800	0.800	0.800

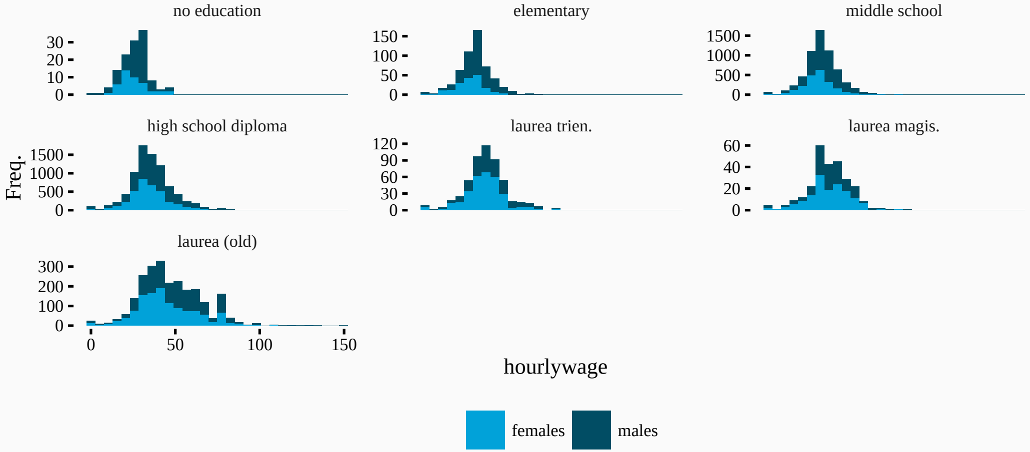
Application II: Italian earnings

- Earnings functions are one of the most investigated relationships in economics.
- Earnings functions relate the logarithm of earnings to a series of explanatory variables such as education, work experience, gender, race, etc.
- We use data on a sample of Italian workers aged between 15 and 64. The sample is from the Labor Force Survey from ISTAT which is the most comprehensive Italian survey.

Italian Labor Survey



Italian Labor Survey



Regression I

```
##
## Coefficients:
##             Estimate Std. Error z value Pr(>|z|)
## (Intercept)  2.57318    0.06205    41.5   <2e-16
## educ         0.03514    0.00110    32.0   <2e-16
## male         0.10698    0.00756    14.1   <2e-16
## ---
## Heteroskadasticity robust standard errors used
##
## Residual standard error: 0.5 on 18051 degrees of freedom
## Multiple R-squared:  0.121,Adjusted R-squared:  0.12
## F-statistic: 2.11e+03 on 16 and Inf DF,  p-value: <2e-16
## ---
## Factors not reported: CLETAS RIP3
```

Regression II

```
lm1 <- lm(I(log(hourlywage)) ~ TISTUD + CLETAS + male + RIP3, data = lfs)
summary_rob(lm1, omit_factor = TRUE)
```

```
##
## Coefficients:
##             Estimate Std. Error z value Pr(>|z|)
## (Intercept)  2.67867    0.07721    34.7   <2e-16
## male         0.10627    0.00757    14.0   <2e-16
## ---
## Heteroskadasticity robust standard errors used
##
## Residual standard error: 0.5 on 18046 degrees of freedom
## Multiple R-squared:  0.124,Adjusted R-squared:  0.123
## F-statistic: 2.2e+03 on 21 and Inf DF,  p-value: <2e-16
## ---
## Factors not reported: TISTUD CLETAS RIP3
```

Regression III

```
lm3 <- lm(I(log(hourlywage)) ~ educ * male + CLETAS + male + RIP3,
    data = lfs)
summary_rob(lm3, omit_factor = TRUE)
```

```
##
## Coefficients:
##             Estimate Std. Error z value Pr(>|z|)
## (Intercept)  2.50813    0.06440   38.95  < 2e-16
## educ         0.04038    0.00162   24.93  < 2e-16
## male         0.21462    0.02628    8.17  3.1e-16
## educ:male   -0.00910    0.00216   -4.21  2.6e-05
## ---
## Heteroskadasticity robust standard errors used
##
## Residual standard error: 0.5 on 18050 degrees of freedom
## Multiple R-squared:  0.122,Adjusted R-squared:  0.121
## F-statistic: 2.14e+03 on 17 and Inf DF,  p-value: <2e-16
## ---
## Factors not reported: CLETAS RIP3
```

Regression IV

```
lm4 <- lm(I(log(hourlywage)) ~ educ * RIP3 + CLETAS + male, data = lfs)
summary_rob(lm4, omit_factor = TRUE)
```

```
##
## Coefficients:
##              Estimate Std. Error z value Pr(>|z|)
## (Intercept) 2.596689   0.062803   41.35   <2e-16
## educ        0.033042   0.001359   24.31   <2e-16
## male        0.107476   0.007554   14.23   <2e-16
## educ:Center 0.000501   0.002596    0.19    0.847
## educ:South  0.008081   0.002882    2.80    0.005
## ---
## Heteroskadasticity robust standard errors used
##
## Residual standard error: 0.5 on 18049 degrees of freedom
## Multiple R-squared:  0.121,Adjusted R-squared:  0.121
## F-statistic: 2.12e+03 on 18 and Inf DF,  p-value: <2e-16
## ---
## Factors not reported: RIP3 CLETAS
```

Regression V

```
lm5 <- lm(I(log(hourlywage)) ~ REG + educ + male, data = lfs)
summary_rob(lm5, omit_factor = TRUE)
```

```
##
## Coefficients:
##             Estimate Std. Error z value Pr(>|z|)
## (Intercept)  3.08513    0.01785   172.9   <2e-16
## educ         0.03109    0.00111    28.0   <2e-16
## male         0.09737    0.00782    12.4   <2e-16
## ---
## Heteroskadasticity robust standard errors used
##
## Residual standard error: 0.52 on 18046 degrees of freedom
## Multiple R-squared:  0.0612,Adjusted R-squared:  0.0601
## F-statistic: 1.23e+03 on 21 and Inf DF,  p-value: <2e-16
## ---
## Factors not reported: REG
```

Example: IQ and Education

Question: Do education and IQ have an interactive effect in the log(*wage*) equation?

Data: M. Blackburn and D. Neumark (1992), "Unobserved Ability, Efficiency Wages, and Interindustry Wage Differentials," Quarterly Journal of Economics 107, 1421-1436.

```
##    variable                             label
## 1      wage                  monthly earnings
## 2     hours              average weekly hours
## 3        IQ                          IQ score
## 4       KWW knowledge of world work score
## 5      educ                years of education
## 6     exper           years of work experience
## 7    tenure     years with current employer
## 8       age                      age in years
## 9   married                      =1 if married
## 10    black                        =1 if black
## 11    south                =1 if live in south
## 12    urban                 =1 if live in SMSA
## 13     sibs                 number of siblings
## 14  brthord                        birth order
## 15    meduc                 mother's education
## 16    feduc                 father's education
```

IQ and Education

Table 1:

Statistic	N	Mean	St. Dev.	Min	Max
wage	935	958.000	404.000	115	3,078
hours	935	44.000	7.200	20	80
IQ	935	101.000	15.000	50	145
KWW	935	36.000	7.600	12	56
educ	935	13.000	2.200	9	18
exper	935	12.000	4.400	1	23
tenure	935	7.200	5.100	0	22
age	935	33.000	3.100	28	38
married	935	0.890	0.310	0	1
black	935	0.130	0.340	0	1
south	935	0.340	0.470	0	1
urban	935	0.720	0.450	0	1
sibs	935	2.900	2.300	0	14
brthord	852	2.300	1.600	1	10
meduc	857	11.000	2.900	0	18
feduc	741	10.000	3.300	0	18

IQ and Education, ctd.

We estimate

$$lwage = \beta_0 + \beta_1 educ + \beta_2 IQ + \beta_3 educ \cdot IQ + \beta_4 exper + \beta_5 hours + u$$

- What is the interpretation of the coefficients?
- What do you think the signs are going to be?
- What is the economic intuition?

IQ and Education, ctd

```
lm1 <- lm(I(log(wage)) ~ educ + educ * IQ + exper + hours, data = wage2)
summary_rob(lm1)
```

```
##
## Coefficients:
##              Estimate Std. Error z value Pr(>|z|)
## (Intercept)  5.226658   0.551609    9.48  < 2e-16
## educ         0.069744   0.042495    1.64    0.101
## IQ           0.007326   0.005278    1.39    0.165
## exper        0.019400   0.003245    5.98  2.3e-09
## hours       -0.004553   0.002068   -2.20    0.028
## educ:IQ     -0.000111   0.000401   -0.28    0.782
## ---
## Heteroskadasticity robust standard errors used
##
## Residual standard error: 0.39 on 929 degrees of freedom
## Multiple R-squared:  0.168,Adjusted R-squared:  0.164
## F-statistic:  208 on 5 and Inf DF,  p-value: <2e-16
```

IQ and Education, ctd

```
lm2 <- lm(I(log(wage)) ~ educ + IQ + educ:I(IQ - mean(IQ)) + exper +
    hours, data = data)
summary_rob(lm2)
```

```
##
## Coefficients:
##                        Estimate Std. Error z value Pr(>|z|)
## (Intercept)            5.226658   0.551609    9.48  < 2e-16
## educ                   0.058501   0.007834    7.47  8.2e-14
## IQ                     0.007326   0.005278    1.39    0.165
## exper                  0.019400   0.003245    5.98  2.3e-09
## hours                 -0.004553   0.002068   -2.20    0.028
## educ:I(IQ - mean(IQ)) -0.000111   0.000401   -0.28    0.782
## ---
## Heteroskadasticity robust standard errors used
##
## Residual standard error: 0.39 on 929 degrees of freedom
## Multiple R-squared:  0.168,Adjusted R-squared:  0.164
## F-statistic:  208 on 5 and Inf DF,  p-value: <2e-16
```

Summary: Nonlinear Regression Functions

- Using functions of the independent variables such as $\ln(X)$ or $X_1 \times X_2$, allows recasting a large family of nonlinear regression functions as multiple regression.
- Estimation and inference proceed in the same way as in the linear multiple regression model.
- Interpretation of the coefficients is model-specific, but the general rule is to compute effects by comparing different cases (different value of the original X's)
- Many nonlinear specifications are possible, so you must use judgment:
  - What nonlinear effect you want to analyze?
  - What makes sense in your application?