

stata_conference_2020_porto

January 25, 2020

1 2020 Portugal Stata Conference | Porto | 25 January

2 Stata sample exercise using Jupyter Notebook

M Portela

2.1 1. Configuration

```
[140]: %set execution_mode console
```

```
[113]: %set graph_format svg --permanently
```

```
[61]: %set autocomplete_closing_symbol True
```

2.2 2. Change working directory

```
[29]: cd "/Users/miguelportela/Dropbox/statafep2020/workshop_25/apresentacoes/  
→jupyter_notebook/stata_full_exercise/logs"
```

```
/Users/miguelportela/Dropbox/statafep2020/workshop_25/apresentacoes/jupyter_note  
book/stata_full_exercise/logs
```

2.3 3. Some *magics*

```
[27]: %%help magics
```

```
[28]: %%help histogram
```

2.4 4. Read data

```
[64]: //webuse nlswork  
//save ../data/nlswork, replace
```

```
[92]: use ../data/nlswork, clear
```

(National Longitudinal Survey. Young Women 14-26 years of age in 1968)

2.5 5. Explore the data

```
[63]: %head 5
```

```
+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+
| idcode  year  birth_yr  age  race  msp  nev_mar  grade  collgrad  not_smsa  |
+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+
1. |      1    70      51   18  black    0        1    12        0        0
2. |      1    71      51   19  black    1        0    12        0        0
3. |      1    72      51   20  black    1        0    12        0        0
4. |      1    73      51   21  black    1        0    12        0        0
5. |      1    75      51   23  black    1        0    12        0        0
+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+
```

```
[86]: codebook, compact
```

Variable	Obs	Unique	Mean	Min	Max	Label
idcode	28534	4711	2601.284	1	5159	NLS ID
year	28534	15	77.95865	68	88	interview year
birth_yr	28534	14	48.08509	41	54	birth year
age	28510	33	29.04511	14	46	age in current year
race	28534	3	1.303392	1	3	race
msp	28518	2	.6029175	0	1	1 if married, spouse present
nev_mar	28518	2	.2296795	0	1	1 if never married
grade	28532	19	12.53259	0	18	current grade completed
collgrad	28534	2	.1680451	0	1	1 if college graduate
not_smsa	28526	2	.2824441	0	1	1 if not SMSA
c_city	28526	2	.357218	0	1	1 if central city
south	28526	2	.4095562	0	1	1 if south
ind_code	28193	12	7.692973	1	12	industry of employment
occ_code	28413	13	4.777672	1	13	occupation
union	19238	2	.2344319	0	1	1 if union
wks_ue	22830	61	2.548095	0	76	weeks unemployed last year
ttl_exp	28534	4744	6.215316	0	28.88461	total work experience
tenure	28101	270	3.123836	0	25.91667	job tenure, in years
hours	28467	85	36.55956	1	168	usual hours worked
wks_work	27831	105	53.98933	0	104	weeks worked last year
ln_wage	28534	8173	1.674907	0	5.263916	ln(wage/GNP deflator)

2.5.1 5.1 Transform variables

```
[93]: ren ttl_exp exper
      gen exper2 = exper^2
```

```
[95]: label var exper "Experience sqr."
```

```
[97]: gen wage = exp(ln_wage)
```

2.5.2 5.2 Summary statistics

```
[82]: sum ln_wage, detail
```

ln(wage/GNP deflator)				

	Percentiles	Smallest		
1%	.4135621	0		
5%	.9927515	0		
10%	1.166102	.0044871	Obs	28,534
25%	1.361496	.0044871	Sum of Wgt.	28,534
50%	1.640541		Mean	1.674907
		Largest	Std. Dev.	.4780935
75%	1.964083	4.349226		
90%	2.27569	4.49981	Variance	.2285734
95%	2.456406	4.828314	Skewness	.3278257
99%	2.956471	5.263916	Kurtosis	4.666831

```
[88]: tab south union
```

		1 if union		
1 if south		0	1	Total

	0	8,090	3,169	11,259
	1	6,637	1,339	7,976

Total		14,727	4,508	19,235

```
[99]: tabstat wage,by(union) stat(N mean sd p1 p50 p99)
```

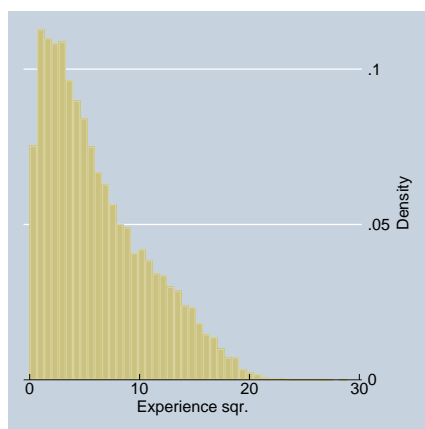
```
Summary for variables: wage
      by categories of: union (1 if union)
```

union	N	mean	sd	p1	p50	p99
0	14728	6.148258	3.4076	1.680601	5.351879	18.19645
1	4510	7.602447	4.764265	2.783964	6.810506	22.20554
Total	19238	6.489166	3.819589	1.799999	5.642454	19.35587

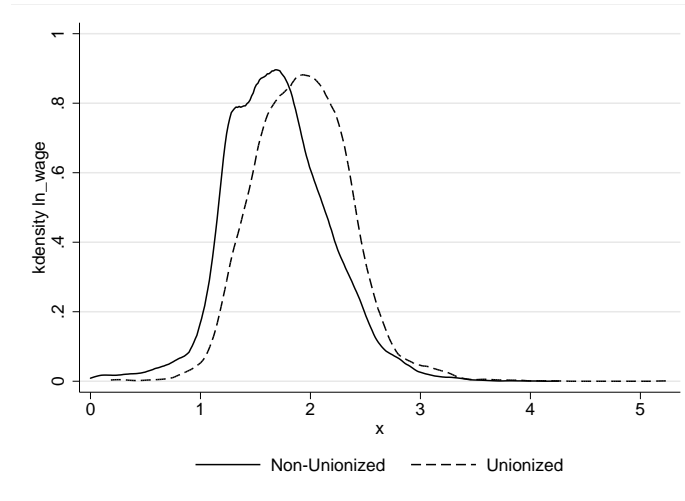
2.5.3 5.3 Graphs

```
[144]: histogram exper,scheme(economist)
```

```
(bin=44, start=0, width=.65646852)
```



```
[115]: twoway (kdensity ln_wage if union == 0) (kdensity ln_wage if union == 1),scheme(sj) graphregion(color(white))
        ↪ legen(region(color(white)) order(1 ↪
        ↪ "Non-Unionized" 2 "Unionized"))
```



```
[122]: quiet eststo m1: regress ln_wage union
```

```
[129]: quiet eststo m2: regress ln_wage collgrad exper exper2 union
```

```
[137]: quiet eststo m3: regress ln_wage collgrad exper exper2 union south
```

2.6 6. Regression analysis

The model to be estimated is defined by

$$\text{Ln_Wage}_i = \beta_0 + \beta_1 \text{Union}_i + \beta_2 \text{Collgrad}_i + \beta_3 \text{Exper}_i + \beta_4 \text{Exper}_i^2 + \varepsilon_i$$

where Ln_Wage_i stands for log wage, ...

```
[139]: esttab m1 m2 m3, b(%4.3f) se(%4.3f) nocons r2 aic(%5.0f)
```

	(1)	(2)	(3)
	ln_wage	ln_wage	ln_wage
union	0.224*** (0.008)	0.176*** (0.007)	0.150*** (0.007)
collgrad		0.359*** (0.007)	0.352*** (0.007)
exper		0.044*** (0.002)	0.044*** (0.002)
exper2		-0.001*** (0.000)	-0.001*** (0.000)

south			-0.170*** (0.006)
-------	--	--	----------------------

N	19238	19238	19235
R-sq	0.041	0.268	0.300
AIC	24558	19367	18516

Standard errors in parentheses

* p<0.05, ** p<0.01, *** p<0.001

Estimations from model (3) indicate that, conditional on Education, Experience and Location, Unionized workers have a premium of about 15%.

[]: