

Data – wage2.dta

. des wage lwage educ sibs

variable name	type	format	label	variable label
---------------	------	--------	-------	----------------

wage	int	%9.0g		monthly earnings
lwage	float	%9.0g		natural log of wage
educ	byte	%9.0g		years of education
sibs	byte	%9.0g		number of siblings

. reg lwage educ

Source	SS	df	MS	Number of obs =	935
-----+-----				F(1, 933) =	100.70
Model	16.1377042	1	16.1377042	Prob > F	= 0.0000
Residual	149.518579	933	.160255712	R-squared	= 0.0974
-----+-----				Adj R-squared =	0.0964
Total	165.656283	934	.177362188	Root MSE	= .40032

lwage	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]	
-----+-----						
educ	.0598392	.0059631	10.03	0.000	.0481366	.0715418
_cons	5.973063	.0813737	73.40	0.000	5.813366	6.132759

(OLS) Return to education = 5.98% (t=10.03)

. reg educ sibs

Source	SS	df	MS	Number of obs =	935
-----+-----				F(1, 933) =	56.67
Model	258.055048	1	258.055048	Prob > F	= 0.0000
Residual	4248.7642	933	4.55387374	R-squared	= 0.0573
-----+-----				Adj R-squared =	0.0562
Total	4506.81925	934	4.82528828	Root MSE	= 2.134

educ	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]	
------	-------	-----------	---	------	----------------------	--

```
-----+-----
sibs | -0.2279164 .0302768 -7.53 0.000 -.287335 -.1684979
_cons | 14.13879 .1131382 124.97 0.000 13.91676 14.36083
-----+-----
```

(implication) Therefore, educ and sibs are negatively correlated. Use sibs as an IV for educ

. reg lwage educ (sibs)

Instrumental variables (2SLS) regression

```
-----+-----
Source |    SS    df    MS                Number of obs =   935
-----+-----                F( 1, 933) = 21.59
Model | -1.51973315    1 -1.51973315    Prob > F   = 0.0000
Residual | 167.176016   933 .179181154    R-squared   = .
-----+-----                Adj R-squared = .
Total | 165.656283   934 .177362188    Root MSE   = .4233
```

```
-----+-----
lwage |    Coef.  Std. Err.    t  P>|t|  [95% Conf. Interval]
-----+-----
educ | .1224326  .0263506    4.65 0.000  .0707194  .1741459
_cons | 5.130026  .3551712   14.44 0.000  4.432999  5.827053
-----+-----
```

(OLS) Return to education = 5.98% (t=10.03)

(IV) Return to education = 12.24% (t=4.65) - double

. reg lwage educ (sibs), robust

IV (2SLS) regression with robust standard errors Number of obs = 935

```
F( 1, 933) = 24.80
Prob > F   = 0.0000
R-squared   = .
Root MSE   = .4233
```

	Robust		t	P> t	[95% Conf. Interval]	
lwage	Coef.	Std. Err.				
educ	.1224326	.0245865	4.98	0.000	.0741814	.1706839
_cons	5.130026	.330732	15.51	0.000	4.480961	5.779091

- Multiple Regression and IV

. des lwage educ exper tenure married black south urban

variable name	type	format	label	variable label
lwage	float	%9.0g		natural log of wage
educ	byte	%9.0g		years of education
exper	byte	%9.0g		years of work experience
tenure	byte	%9.0g		years with current employer
married	byte	%9.0g		=1 if married
black	byte	%9.0g		=1 if black
south	byte	%9.0g		=1 if live in south
urban	byte	%9.0g		=1 if live in SMSA

. reg lwage educ exper tenure married black south urban

Source	SS	df	MS	Number of obs =	935
				F(7, 927) =	44.75
Model	41.8377619	7	5.97682312	Prob > F	= 0.0000
Residual	123.818521	927	.133569063	R-squared	= 0.2526
				Adj R-squared =	0.2469
Total	165.656283	934	.177362188	Root MSE	= .36547

lwage	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]	
educ	.0654307	.0062504	10.47	0.000	.0531642	.0776973

exper		.014043	.0031852	4.41	0.000	.007792	.020294
tenure		.0117473	.002453	4.79	0.000	.0069333	.0165613
married		.1994171	.0390502	5.11	0.000	.1227801	.276054
black		-.1883499	.0376666	-5.00	0.000	-.2622717	-.1144281
south		-.0909036	.0262485	-3.46	0.001	-.142417	-.0393903
urban		.1839121	.0269583	6.82	0.000	.1310056	.2368185
_cons		5.395497	.113225	47.65	0.000	5.17329	5.617704

Simple Linear Regression

(OLS) Return to education = 5.98% (t=10.03)

(IV) Return to education = 12.24% (t=4.65) - double

Multiple Linear Regression

(OLS) Return to education = 6.54% (t=10.47)

. reg educ sibs exper tenure married black south urban

Source		SS	df	MS	Number of obs =	935
					F(7, 927) =	48.83
Model		1214.11526	7	173.445037	Prob > F =	0.0000
Residual		3292.70399	927	3.55199999	R-squared =	0.2694
					Adj R-squared =	0.2639
Total		4506.81925	934	4.82528828	Root MSE =	1.8847

educ		Coef.	Std. Err.	t	P> t	[95% Conf. Interval]
sibs		-.1679454	.0281722	-5.96	0.000	-.223234 - .1126567
exper		-.2252668	.0146799	-15.35	0.000	-.2540766 - .1964571
tenure		.0257891	.0126233	2.04	0.041	.0010155 .0505626
married		-.1261701	.2013338	-0.63	0.531	-.5212929 .2689527
black		-.5878081	.2009304	-2.93	0.004	-.9821393 -.1934769
south		-.2095298	.1352271	-1.55	0.122	-.4749165 .055857
urban		.2352623	.1389367	1.69	0.091	-.0374046 .5079293
_cons		16.47151	.281876	58.44	0.000	15.91832 17.0247

**. reg lwage educ exper tenure married black south urban (sibs exper
tenure married black south urban)**

Instrumental variables (2SLS) regression

Source	SS	df	MS	Number of obs =	935
-----+-----				F(7, 927) =	29.76
Model	40.6135431	7	5.80193472	Prob > F	= 0.0000
Residual	125.04274	927	.134889687	R-squared	= 0.2452
-----+-----				Adj R-squared =	0.2395
Total	165.656283	934	.177362188	Root MSE	= .36727

lwage	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]	
-----+-----						
educ	.0843535	.0326893	2.58	0.010	.0201999	.148507
exper	.0183924	.0080385	2.29	0.022	.0026168	.0341681
tenure	.0112142	.0026255	4.27	0.000	.0060615	.0163668
married	.2019726	.0394812	5.12	0.000	.1244898	.2794554
black	-.1704875	.0484751	-3.52	0.000	-.2656211	-.075354
south	-.087178	.0271236	-3.21	0.001	-.1404088	-.0339472
urban	.178594	.0285521	6.26	0.000	.1225597	.2346282
_cons	5.092169	.5266761	9.67	0.000	4.058553	6.125785

Simple Linear Regression

(OLS) Return to education = 5.98% (t=10.03)

(IV) Return to education = 12.24% (t=4.65) - double

Multiple Linear Regression

(OLS) Return to education = 6.54% (t=10.47)

(IV) Return to education = 8.44% (t=2.58)

. reg lwage educ exper tenure married black south urban (sibs exper tenure married black south urban), robust

IV (2SLS) regression with robust standard errors

Number of obs = 935

F(7, 927) = 32.87
 Prob > F = 0.0000
 R-squared = 0.2452
 Root MSE = .36727

	Robust					
lwage	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]	
educ	.0843535	.030497	2.77	0.006	.0245023	.1442047
exper	.0183924	.0078274	2.35	0.019	.003031	.0337539
tenure	.0112142	.0026446	4.24	0.000	.006024	.0164043
married	.2019726	.0405929	4.98	0.000	.1223079	.2816372
black	-.1704875	.0470413	-3.62	0.000	-.2628073	-.0781678
south	-.087178	.0285618	-3.05	0.002	-.1432312	-.0311247
urban	.178594	.0288014	6.20	0.000	.1220705	.2351174
_cons	5.092169	.4960079	10.27	0.000	4.118741	6.065598

- Two Stage Least Squares

Data – card.dta

. des lwage educ nearc4 nearc2 exper expersq black smsa south smsa66 reg662-reg669

variable name	type	format	label	variable label
lwage	float	%9.0g		log(wage)
educ	byte	%9.0g		years of schooling, 1976
nearc4	byte	%9.0g		=1 if near 4 yr college, 1966
nearc2	byte	%9.0g		=1 if near 2 yr college, 1966
exper	byte	%9.0g		age - educ - 6
expersq	int	%9.0g		exper^2
black	byte	%9.0g		=1 if black
smsa	byte	%9.0g		=1 in in SMSA, 1976

```

south      byte %9.0g      =1 if in south, 1976
smsa66     byte %9.0g      =1 if in SMSA, 1966
reg662     byte %9.0g      =1 for region 2, 1966
reg663     byte %9.0g      =1 for region 3, 1966
reg664     byte %9.0g      =1 for region 4, 1966
reg665     byte %9.0g      =1 for region 5, 1966
reg666     byte %9.0g      =1 for region 6, 1966
reg667     byte %9.0g      =1 for region 7, 1966
reg668     byte %9.0g      =1 for region 8, 1966
reg669     byte %9.0g      =1 for region 9, 1966

```

. reg lwage educ exper expersq black smsa south smsa66 reg662-reg669

```

Source |      SS      df      MS      Number of obs = 3010
-----+-----
Model | 177.695591   15 11.8463727      F( 15, 2994) = 85.48
Residual | 414.946054 2994 .138592536      Prob > F      = 0.0000
-----+-----
Total | 592.641645 3009 .196956346      R-squared     = 0.2998
                                           Adj R-squared = 0.2963
                                           Root MSE    = .37228

```

```

-----
lwage |      Coef.   Std. Err.      t    P>|t|   [95% Conf. Interval]
-----+-----
educ |   .0746933   .0034983    21.35  0.000   .0678339   .0815527
exper |   .084832   .0066242    12.81  0.000   .0718435   .0978205
expersq | -.002287   .0003166    -7.22  0.000  -.0029079  -.0016662
black | -.1990123   .0182483   -10.91  0.000  -.2347927  -.1632318
smsa |   .1363845   .0201005     6.79  0.000   .0969724   .1757967
south | -.147955   .0259799    -5.69  0.000  -.1988952  -.0970148
smsa66 | .0262417   .0194477     1.35  0.177  -.0118905   .0643739
reg662 | .0963672   .0358979     2.68  0.007   .0259801   .1667542
reg663 | .14454     .0351244     4.12  0.000   .0756696   .2134105
reg664 | .0550756   .0416573     1.32  0.186  -.0266043   .1367554
reg665 | .1280248   .0418395     3.06  0.002   .0459878   .2100618
reg666 | .1405174   .0452469     3.11  0.002   .0517992   .2292356
reg667 | .117981    .0448025     2.63  0.008   .0301343   .2058277
reg668 | -.0564361   .0512579    -1.10  0.271  -.1569404   .0440682
reg669 | .1185698   .0388301     3.05  0.002   .0424335   .194706
_cons | 4.620807   .0742327    62.25  0.000   4.475254   4.766359
-----

```

Multiple Linear Regression

(OLS) Return to education = 7.47% (t=21.35)

```
. reg educ nearc4 nearc2 exper expersq black smsa south smsa66 reg662-
reg669
```

Source	SS	df	MS	Number of obs = 3010		
-----+-----				F(16, 2993) = 170.99		
Model	10297.1164	16	643.569774	Prob > F = 0.0000		
Residual	11264.9637	2993	3.76377002	R-squared = 0.4776		
-----+-----				Adj R-squared = 0.4748		
Total	21562.0801	3009	7.16586243	Root MSE = 1.94		

educ	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]	
-----+-----						
nearc4	.3205819	.0878425	3.65	0.000	.148344	.4928197
nearc2	.1229986	.0774256	1.59	0.112	-.0288142	.2748114
exper	-.4122915	.0336914	-12.24	0.000	-.4783521	-.3462309
expersq	.0008479	.00165	0.51	0.607	-.0023874	.0040832
black	-.9451729	.0939073	-10.06	0.000	-1.129302	-.7610434
smsa	.4013708	.1047858	3.83	0.000	.1959113	.6068303
south	-.0419115	.1355316	-0.31	0.757	-.3076561	.2238331
smsa66	.0000782	.1069445	0.00	0.999	-.2096139	.2097704
reg662	-.1002481	.1875618	-0.53	0.593	-.4680113	.2675151
reg663	-.0214286	.1833737	-0.12	0.907	-.3809798	.3381226
reg664	.1310678	.2173736	0.60	0.547	-.295149	.5572846
reg665	-.2683558	.2183813	-1.23	0.219	-.6965485	.1598369
reg666	-.3334436	.2377938	-1.40	0.161	-.7996995	.1328122
reg667	-.2087488	.2343833	-0.89	0.373	-.6683174	.2508198
reg668	.5507871	.2679423	2.06	0.040	.0254175	1.076157
reg669	.1687829	.2040832	0.83	0.408	-.2313747	.5689404
_cons	16.60428	.2415174	68.75	0.000	16.13072	17.07783

```
. test nearc4 nearc2
```

(1) nearc4 = 0

(2) nearc2 = 0

F(2, 2993) = 7.89
 Prob > F = 0.0004

(Comments) p-value is 0.0004, so at least one variable is partially correlated with “educ”. “nearc4” is more important.

. corr nearc2 nearc4
(obs=3010)

```

      | nearc2 nearc4
-----+-----
nearc2 | 1.0000
nearc4 | 0.1191 1.0000

```

. reg lwage educ exper expersq black smsa south smsa66 reg662-reg669
(nearc4 nearc2 exper expersq black smsa south smsa66 reg662-reg669)

Instrumental variables (2SLS) regression

```

Source |   SS   df   MS           Number of obs = 3010
-----+-----           F( 15, 2994) = 47.07
Model | 100.869   15 6.72459998       Prob > F   = 0.0000
Residual | 491.772645 2994 .16425272       R-squared  = 0.1702
-----+-----           Adj R-squared = 0.1660
Total | 592.641645 3009 .196956346       Root MSE   = .40528

```

```

-----+-----
lwage |   Coef.  Std. Err.   t  P>|t|   [95% Conf. Interval]
-----+-----
educ | .1570594 .0525782   2.99 0.003   .0539662 .2601525
exper | .1188149 .0228061   5.21 0.000   .0740977 .163532
expersq | -.0023565 .0003475  -6.78 0.000  -.0030379 -.0016751
black | -.1232778 .05215   -2.36 0.018  -.2255313 -.0210243
smsa | .100753 .0315193   3.20 0.001   .0389512 .1625548

```

south		-.1431945	.0284448	-5.03	0.000	-.1989678	-.0874212
smsa66		.0150626	.022336	0.67	0.500	-.0287328	.058858
reg662		.1027473	.0392906	2.62	0.009	.025708	.1797867
reg663		.1499316	.0383918	3.91	0.000	.0746546	.2252086
reg664		.0475676	.0456013	1.04	0.297	-.0418454	.1369806
reg665		.1544801	.0485628	3.18	0.001	.0592603	.2497
reg666		.1729728	.0534164	3.24	0.001	.0682362	.2777094
reg667		.1420356	.0511219	2.78	0.005	.041798	.2422731
reg668		-.0950611	.0609801	-1.56	0.119	-.2146281	.024506
reg669		.102976	.0434224	2.37	0.018	.0178353	.1881167
_cons		3.236711	.8849118	3.66	0.000	1.501614	4.971807

Multiple Linear Regression

(OLS) Return to education = 7.47% (t=21.35)

(IV) Return to education = 15.71% (t=2.99)