

name: <unnamed>
log: C:\Users\XuQi\Desktop\第四章.smcl
log type: smcl
opened on: 15 Jul 2024, 10:08:11

```
. do "C:\Users\XuQi\AppData\Local\Temp\STD2670_000000.tmp"  
  
. use "C:\Users\XuQi\Desktop\cfps2010.dta", clear  
  
. *查看变量  
. describe
```

Contains data from C:\Users\XuQi\Desktop\cfps2010.dta
Observations: 4,137
Variables: 12 5 Aug 2022 19:25

Variable name	Storage type	Display format	Value label	Variable label
pid	double	%12.0g	pid	个人id
provcd	double	%24.0g	provcd	省国标码
gender	double	%12.0g	gender	性别
age	float	%9.0g		年龄
age2	float	%9.0g		年龄平方
age3	float	%9.0g		年龄三次方
lninc	float	%9.0g		收入对数
college	double	%9.0g	yesorno	是否上大学
hukou	double	%12.0g	hukou	3岁时户口性质
sibling	float	%9.0g	yesorno	是否独生子女
race	double	%9.0g	yesorno	是否汉族
fmedu	float	%9.0g	fmedu	父母是否上过高中

Sorted by:

```
.  
. *一元线性回归  
. reg lninc college
```

Source	SS	df	MS	Number of obs	=	4,137
Model	671.883317	1	671.883317	F(1, 4135)	=	508.22
Residual	5466.56371	4,135	1.32202266	Prob > F	=	0.0000
				R-squared	=	0.1095
				Adj R-squared	=	0.1092
Total	6138.44703	4,136	1.48415063	Root MSE	=	1.1498

lninc	Coefficient	Std. err.	t	P> t	[95% conf. interval]	
college	.823612	.0365338	22.54	0.000	.7519861	.895238
_cons	9.353189	.0230235	406.25	0.000	9.308051	9.398327

```
.  
. reg lninc college, vce(robust) // 异方差稳健标准误
```

Linear regression	Number of obs	=	4,137
	F(1, 4135)	=	582.05
	Prob > F	=	0.0000
	R-squared	=	0.1095
	Root MSE	=	1.1498

lninc	Coefficient	Robust std. err.	t	P> t	[95% conf. interval]	
college	.823612	.0341385	24.13	0.000	.7566823	.8905418
_cons	9.353189	.0256882	364.10	0.000	9.302826	9.403552

```
. reg lninc college, vce(cluster provcd) //聚类稳健标准误
```

Linear regression

Number of obs	=	4,137
F(1, 24)	=	271.17
Prob > F	=	0.0000
R-squared	=	0.1095
Root MSE	=	1.1498

(Std. err. adjusted for 25 clusters in provcd)

lninc	Coefficient	Robust std. err.	t	P> t	[95% conf. interval]	
college	.823612	.0500155	16.47	0.000	.7203851	.926839
_cons	9.353189	.1084703	86.23	0.000	9.129317	9.577061

```
.  
. *多元线性回归  
. reg lninc college hukou, vce(cluster provcd)
```

Linear regression

Number of obs	=	4,137
F(2, 24)	=	219.69
Prob > F	=	0.0000
R-squared	=	0.1169
Root MSE	=	1.1451

(Std. err. adjusted for 25 clusters in provcd)

lninc	Coefficient	Robust std. err.	t	P> t	[95% conf. interval]	
college	.798281	.0451727	17.67	0.000	.7050491	.891513
hukou	.2155333	.0611855	3.52	0.002	.0892526	.3418141
_cons	9.27541	.0993474	93.36	0.000	9.070367	9.480453

```
.  
. *理解偏回归系数  
. reg college hukou, vce(cluster provcd)
```

Linear regression

Number of obs	=	4,137
F(1, 24)	=	27.84
Prob > F	=	0.0000
R-squared	=	0.0137
Root MSE	=	.48606

(Std. err. adjusted for 25 clusters in provcd)

college	Coefficient	Robust std. err.	t	P> t	[95% conf. interval]	
hukou	.116539	.0220861	5.28	0.000	.0709555	.1621224
_cons	.3496532	.0161816	21.61	0.000	.3162561	.3830504

```
. predict e, residuals
```

```
. reg lninc e, vce(cluster provcd)
```

Linear regression

Number of obs	=	4,137
F(1, 24)	=	321.05
Prob > F	=	0.0000
R-squared	=	0.1014
Root MSE	=	1.155

(Std. err. adjusted for 25 clusters in provcd)

lninc	Coefficient	Robust std. err.	t	P> t	[95% conf. interval]	
e	.798281	.044552	17.92	0.000	.7063302	.8902318
_cons	9.680285	.1232655	78.53	0.000	9.425877	9.934692

```
.
. *遗漏变量公式
. dis .823612-.798281
.025331
```

```
.
. reg hukou college, vce(cluster provcd)
```

Linear regression

Number of obs = 4,137

F(1, 24) = 33.17

Prob > F = 0.0000

R-squared = 0.0137

Root MSE = .48812

(Std. err. adjusted for 25 clusters in provcd)

hukou	Coefficient	Robust std. err.	t	P> t	[95% conf. interval]	
college	.1175271	.0204066	5.76	0.000	.0754099	.1596443
_cons	.3608661	.0559186	6.45	0.000	.2454558	.4762764

```
. dis .1175271*.2155333
.025331
```

```
.
.
. *纳入更多控制变量
. reg lninc college hukou##i.age, vce(cluster provcd)
```

Linear regression

Number of obs = 4,137

F(23, 24) = .

Prob > F = .

R-squared = 0.1501

Root MSE = 1.1316

(Std. err. adjusted for 25 clusters in provcd)

lninc	Coefficient	Robust std. err.	t	P> t	[95% conf. interval]	
college	.7477749	.0544757	13.73	0.000	.6353425	.8602073
hukou 城镇户口	-.0787305	.1418798	-0.55	0.584	-.3715559	.2140949
age						
26	.3980867	.1481415	2.69	0.013	.0923376	.7038358
27	.2684265	.2134622	1.26	0.221	-.1721377	.7089908
28	.4034527	.1960134	2.06	0.051	-.0010991	.8080045
29	.3063752	.2061596	1.49	0.150	-.1191172	.7318677
30	.808574	.138855	5.82	0.000	.5219914	1.095157
31	.6390432	.1597057	4.00	0.001	.3094269	.9686596
32	.3162257	.1835314	1.72	0.098	-.0625646	.6950159
33	.5423168	.1908591	2.84	0.009	.148403	.9362305
34	.4480384	.1642828	2.73	0.012	.1089753	.7871016
35	.4871379	.1332159	3.66	0.001	.2121937	.762082
36	.7649181	.1431828	5.34	0.000	.4694033	1.060433
37	.5898895	.1644897	3.59	0.001	.2503994	.9293797
38	.4446396	.1414918	3.14	0.004	.1526149	.7366642
39	.3132889	.2152405	1.46	0.158	-.1309457	.7575234
40	.6295174	.1203586	5.23	0.000	.3811094	.8779254
41	.7413657	.1649781	4.49	0.000	.4008676	1.081864
42	.6440224	.1854795	3.47	0.002	.2612115	1.026833
43	.1100458	.261421	0.42	0.678	-.4295006	.6495921
44	.3310236	.1953046	1.69	0.103	-.0720652	.7341124
45	.2092177	.2516651	0.83	0.414	-.3101935	.7286288
46	.5989884	.1833996	3.27	0.003	.2204702	.9775067
47	.2833783	.1619895	1.75	0.093	-.0509517	.6177083
48	.2354062	.1468512	1.60	0.122	-.0676798	.5384923
49	.4378188	.164764	2.66	0.014	.0977627	.7778749
50	.0167503	.1750601	0.10	0.925	-.344556	.3780566
51	-.1610867	.1851799	-0.87	0.393	-.5432792	.2211059
52	-.1302052	.286344	-0.45	0.653	-.7211901	.4607797
53	-.0033811	.2062946	-0.02	0.987	-.4291521	.42239
54	.2233507	.1566927	1.43	0.167	-.1000471	.5467486
55	.216386	.2374876	0.91	0.371	-.2737643	.7065363

hukou#age						
城镇户口#26	-.0602352	.2412837	-0.25	0.805	-.5582202	.4377497
城镇户口#27	.4551441	.2638459	1.73	0.097	-.0894072	.9996953
城镇户口#28	.3544994	.2567836	1.38	0.180	-.1754758	.8844747
城镇户口#29	.3823274	.297972	1.28	0.212	-.2326564	.9973113
城镇户口#30	-.0193147	.2555672	-0.08	0.940	-.5467794	.50815
城镇户口#31	-.0932038	.2561528	-0.36	0.719	-.6218772	.4354696
城镇户口#32	.6327651	.2259055	2.80	0.010	.166519	1.099011
城镇户口#33	.1374474	.2483846	0.55	0.585	-.3751932	.650088
城镇户口#34	.4132721	.2550715	1.62	0.118	-.1131696	.9397139
城镇户口#35	.2141665	.1854107	1.16	0.259	-.1685025	.5968354
城镇户口#36	.0205633	.2148722	0.10	0.925	-.4229113	.4640378
城镇户口#37	-.0448259	.2031256	-0.22	0.827	-.4640566	.3744048
城镇户口#38	.3187681	.1745486	1.83	0.080	-.0414825	.6790186
城镇户口#39	.4006106	.234722	1.71	0.101	-.0838317	.8850529
城镇户口#40	.0135196	.2054592	0.07	0.948	-.4105273	.4375666
城镇户口#41	-.0207122	.2451003	-0.08	0.933	-.5265743	.4851499
城镇户口#42	.0024452	.2642466	0.01	0.993	-.5429331	.5478234
城镇户口#43	.8893687	.3581507	2.48	0.020	.1501819	1.628555
城镇户口#44	.3810794	.2435123	1.56	0.131	-.1215052	.883664
城镇户口#45	.5548028	.3701405	1.50	0.147	-.2091297	1.318735
城镇户口#46	-.0202823	.2068561	-0.10	0.923	-.4472122	.4066477
城镇户口#47	.1388032	.3315924	0.42	0.679	-.5455699	.8231762
城镇户口#48	.4089894	.2048096	2.00	0.057	-.0137169	.8316956
城镇户口#49	.3254708	.2237488	1.45	0.159	-.1363241	.7872657
城镇户口#50	.65057	.2474684	2.63	0.015	.1398203	1.16132
城镇户口#51	.7067489	.3553651	1.99	0.058	-.0266886	1.440186
城镇户口#52	.7470444	.3445403	2.17	0.040	.0359482	1.458141
城镇户口#53	.694888	.2878366	2.41	0.024	.1008224	1.288954
城镇户口#54	.4754364	.2742435	1.73	0.096	-.0905744	1.041447
城镇户口#55	.2745935	.3227371	0.85	0.403	-.3915031	.9406901
_cons	8.937311	.1616701	55.28	0.000	8.60364	9.270982

. estat ic

Akaike's information criterion and Bayesian information criterion

Model	N	ll(null)	ll(model)	df	AIC	BIC
.	4,137	-6686.381	-6349.885	24	12747.77	12899.63

Note: BIC uses N = number of observations. See [\[R\] IC note](#).

.
. reg lninc college hukou##(c.age c.age2 c.age3), vce(cluster provcd)

Linear regression	Number of obs	=	4,137
	F(8, 24)	=	88.50
	Prob > F	=	0.0000
	R-squared	=	0.1331
	Root MSE	=	1.1354

(Std. err. adjusted for 25 clusters in provcd)

lninc	Coefficient	Robust std. err.	t	P> t	[95% conf. interval]	
college	.7547539	.0501103	15.06	0.000	.6513314	.8581763
hukou						
城镇户口	-.5628695	3.164092	-0.18	0.860	-7.093234	5.967495
age	.6604027	.2081753	3.17	0.004	.23075	1.090055
age2	-.0152886	.005478	-2.79	0.010	-.0265947	-.0039824
age3	.0001114	.0000469	2.38	0.026	.0000147	.0002082
hukou#c.age						
城镇户口	.0697643	.2557794	0.27	0.787	-.4581385	.5976671
hukou#c.age2						
城镇户口	-.0023666	.0067435	-0.35	0.729	-.0162844	.0115512
hukou#c.age3						
城镇户口	.0000268	.0000577	0.46	0.646	-.0000923	.0001459

_cons	.358661	2.584047	0.14	0.891	-4.974549	5.691871
-------	---------	----------	------	-------	-----------	----------

. estat ic

Akaike's information criterion and Bayesian information criterion

Model	N	ll(null)	ll(model)	df	AIC	BIC
.	4,137	-6686.381	-6390.917	9	12799.83	12856.78

Note: BIC uses N = number of observations. See [\[R\] IC note](#).

.
. reg lninc college hukou age age2 age3, vce(cluster provcd)

Linear regression	Number of obs	=	4,137
	F(5, 24)	=	130.17
	Prob > F	=	0.0000
	R-squared	=	0.1304
	Root MSE	=	1.1367

(Std. err. adjusted for 25 clusters in provcd)

lninc	Coefficient	Robust std. err.	t	P> t	[95% conf. interval]	
college	.7483616	.0474037	15.79	0.000	.6505252	.8461981
hukou	.2233219	.0657661	3.40	0.002	.0875874	.3590565
age	.6884785	.143905	4.78	0.000	.3914732	.9854837
age2	-.0162254	.0037903	-4.28	0.000	-.0240482	-.0084026
age3	.000122	.0000326	3.74	0.001	.0000546	.0001893
_cons	.0389147	1.813509	0.02	0.983	-3.703984	3.781814

. estat ic

Akaike's information criterion and Bayesian information criterion

Model	N	ll(null)	ll(model)	df	AIC	BIC
.	4,137	-6686.381	-6397.289	6	12806.58	12844.54

Note: BIC uses N = number of observations. See [\[R\] IC note](#).

.
. reg lninc college hukou age age2 age3 gender race sibling i.fmedu, vce(cluster provcd)

Linear regression	Number of obs	=	4,137
	F(10, 24)	=	241.99
	Prob > F	=	0.0000
	R-squared	=	0.1601
	Root MSE	=	1.1178

(Std. err. adjusted for 25 clusters in provcd)

lninc	Coefficient	Robust std. err.	t	P> t	[95% conf. interval]	
college	.7418545	.0426899	17.38	0.000	.6537468	.8299622
hukou	.2232857	.0579839	3.85	0.001	.1036129	.3429585
age	.6743558	.164432	4.10	0.000	.3349849	1.013727
age2	-.0156173	.0042856	-3.64	0.001	-.0244623	-.0067723
age3	.0001151	.0000365	3.16	0.004	.0000398	.0001904
gender	.3943385	.0417227	9.45	0.000	.3082271	.4804498
race	.1793763	.1014295	1.77	0.090	-.029964	.3887166
sibling	.1837219	.0792935	2.32	0.029	.0200681	.3473757
fmedu						
是	.025374	.0357401	0.71	0.485	-.0483899	.0991379
缺失	.0073327	.0397113	0.18	0.855	-.0746275	.0892929
_cons	-.3450864	2.094658	-0.16	0.871	-4.668247	3.978074

```
.
. *分解回归系数
. reg lninc college if hukou==0, vce(cluster provcd)
```

Linear regression

Number of obs	=	2,451
F(1, 24)	=	477.21
Prob > F	=	0.0000
R-squared	=	0.1018
Root MSE	=	1.238

(Std. err. adjusted for 25 clusters in provcd)

lninc	Coefficient	Robust std. err.	t	P> t	[95% conf. interval]	
college	.873513	.0399865	21.85	0.000	.7909849	.9560412
_cons	9.249105	.1050262	88.06	0.000	9.032342	9.465869

```
. reg lninc college if hukou==1, vce(cluster provcd)
```

Linear regression

Number of obs	=	1,686
F(1, 24)	=	63.29
Prob > F	=	0.0000
R-squared	=	0.1097
Root MSE	=	.99279

(Std. err. adjusted for 25 clusters in provcd)

lninc	Coefficient	Robust std. err.	t	P> t	[95% conf. interval]	
college	.6983452	.0877811	7.96	0.000	.5171739	.8795166
_cons	9.537533	.1003993	95.00	0.000	9.330319	9.744747

```
.
. tab hukou
```

3岁时户 口性质	Freq.	Percent	Cum.
农村户口	2,451	59.25	59.25
城镇户口	1,686	40.75	100.00
Total	4,137	100.00	

```
. tab college if hukou==0
```

是否上 大学	Freq.	Percent	Cum.
否	1,594	65.03	65.03
是	857	34.97	100.00
Total	2,451	100.00	

```
. tab college if hukou==1
```

是否上 大学	Freq.	Percent	Cum.
否	900	53.38	53.38
是	786	46.62	100.00
Total	1,686	100.00	

Iteration 0: EE criterion = **2.620e-24**
Iteration 1: EE criterion = **1.079e-29**

Treatment-effects estimation	Number of obs	=	4,137
Estimator	: regression adjustment		
Outcome model	: linear		
Treatment model	: none		

lninc	Coefficient	Robust std. err.	z	P> z	[95% conf. interval]	
ATET college (否 vs 是)	-.866173	.0405392	-21.37	0.000	-.9456283	-.7867176
P0mean college 是	10.21936	.0317611	321.76	0.000	10.15711	10.28161

```
. *手动实现
. reg lninc hukou age age2 age3 gender race sibling i.fmedu if college==0
```

Source	SS	df	MS	Number of obs	=	2,494
				F(9, 2484)	=	21.72
Model	299.291704	9	33.2546338	Prob > F	=	0.0000
Residual	3803.21617	2,484	1.53108541	R-squared	=	0.0730
				Adj R-squared	=	0.0696
Total	4102.50787	2,493	1.64561086	Root MSE	=	1.2374

lninc	Coefficient	Std. err.	t	P> t	[95% conf. interval]	
hukou	.3400725	.0536769	6.34	0.000	.2348163	.4453286
age	.5876334	.2255975	2.60	0.009	.1452547	1.030012
age2	-.0140172	.0056768	-2.47	0.014	-.0251491	-.0028854
age3	.000105	.0000465	2.26	0.024	.0000138	.0001963
gender	.5032089	.0511487	9.84	0.000	.4029104	.6035073
race	.1929442	.1168315	1.65	0.099	-.0361529	.4220413
sibling	.2414546	.0893476	2.70	0.007	.0662513	.416658
fmedu						
是	.0178976	.0790321	0.23	0.821	-.137078	.1728732
缺失	.0028774	.0613876	0.05	0.963	-.1174987	.1232535
_cons	1.072145	2.916684	0.37	0.713	-4.647238	6.791527

```
. predict y0hat if college==1
(option xb assumed; fitted values)
(2,494 missing values generated)
```

```
. reg lninc hukou age age2 age3 gender race sibling i.fmedu if college==1
```

Source	SS	df	MS	Number of obs	=	1,643
				F(9, 1633)	=	12.17
Model	85.7486412	9	9.5276268	Prob > F	=	0.0000
Residual	1278.30719	1,633	.782796812	R-squared	=	0.0629
				Adj R-squared	=	0.0577
Total	1364.05584	1,642	.83072828	Root MSE	=	.88476

lninc	Coefficient	Std. err.	t	P> t	[95% conf. interval]	
hukou	.0484465	.0485173	1.00	0.318	-.0467162	.1436091
age	.6205456	.1870398	3.32	0.001	.2536824	.9874087
age2	-.0143115	.0049184	-2.91	0.004	-.0239586	-.0046644
age3	.0001088	.000042	2.59	0.010	.0000264	.0001912
gender	.2153348	.044547	4.83	0.000	.1279596	.30271
race	.1548971	.0931465	1.66	0.097	-.0278021	.3375962
sibling	.2273577	.0611421	3.72	0.000	.1074325	.347283
fmedu						
是	.091336	.0511668	1.79	0.074	-.0090234	.1916954
缺失	.0564651	.0651018	0.87	0.386	-.0712268	.184157
_cons	1.06923	2.310655	0.46	0.644	-3.462931	5.60139


```
. predict y1hat if college==0
(option xb assumed; fitted values)
(1,643 missing values generated)

.
. gen y0=lninc if college==0
(1,643 missing values generated)

. replace y0=y0hat if college==1
(1,643 real changes made)

.
. gen y1=lninc if college==1
(2,494 missing values generated)

. replace y1=y1hat if college==0
(2,494 real changes made)

.
. gen effect=y1-y0

. tab college, sum(effect)
```

是否上大学	Summary of effect		Freq.
	Mean	Std. dev.	
否	.86617296	1.2711443	2,494
	.6711883	.92538643	1,643
Total	.78873525	1.1502225	4,137

```
.
.
.
end of do-file

. log close
name: <unnamed>
log: C:\Users\XuQi\Desktop\第四章.smcl
log type: smcl
closed on: 15 Jul 2024, 10:08:33
```
