*统计描述（表 1-第六列数据）
sum schyear highqua age married earning own_exp full self

Variable	Obs	Mean	Std. Dev.	Min	Max
schyear	415	13.19277	3.036246	10	24
highqua	428	14.10981	2.50051	10	17
age	428	42.47664	10.04938	21	59
married	428	.6004673	.4903755	0	1
earning	428	10.02719	9.123297	.2403846	96.84375
own_exp	399	11.90476	9.153682	0	55
full	426	.6079812	.4887749	0	1
self	421	.0332542	.1795128	0	1

*定义每小时工资对数和年龄平方
gen lnearn=ln(earning)

gen agesq=age*age

*做线性回归 工资对数关于教育年限、年龄的二元二次项回归（表 2-第二列数据）
reg lnearn highqua age agesq

Source	SS	df	MS
Model	20.7258534	3	6.9086178
Residual	118.492426	424	.279463268
Total	139.218279	427	.326038124

Number of obs	= 428
F(3, 424)	= 24.72
Prob > F	= 0.0000
R-squared	= 0.1489
Adj R-squared	= 0.1429
Root MSE	= .52864

lnearn	Coef.	Std. Err.	t	P>|t|	[95% Conf.	Interval]
highqua	.0767543	.0105917	7.25	0.000	.0559355	.0975731
age	.0778154	.0213949	3.64	0.000	.0357622	.1198687
agesq	-.0009675	.0002658	-3.64	0.000	-.0014899	-.0004451
_cons	-.4282208	.4347756	-0.98	0.325	-1.282805	.4263631

*工具变量回归（原来的教育年限被 twihigh 取代）（表 2—第三列数据）
ivreg lnearn (highqua=twihigh) age agesq

```
Instrumental variables (2SLS) regression
```

Source	SS	df	MS		
Model	20.4445064	3	6.81483547	Number of obs	= 428
Residual	118.773773	424	.280126822	F(3, 424)	= 16.40
				Prob > F	= 0.0000
				R-squared	= 0.1469
				Adj R-squared	= 0.1408
Total	139.218279	427	.326038124	Root MSE	= .52927

lnearn	Coef.	Std. Err.	t	P>|t|	[95% Conf.	Interval]
highqua	.0873817	.0166363	5.25	0.000	.0546818	.1200815
age	.0764781	.0214809	3.56	0.000	.0342558	.1187005
agesq	-.0009428	.0002677	-3.52	0.000	-.0014691	-.0004165
_cons	-.5684209	.4669861	-1.22	0.224	-1.486317	.3494751

```
Instrumented:  highqua
Instruments:   age agesq twihigh
```

*把数据对应同一家庭的两行双胞胎数据合成一行

reshape wide earning-agesq, i(family) j(twinno)

```
(note: j = 1 2)

Data                               long   ->   wide
-----------------------------------------------------------------------------
Number of obs.                      428   ->    214
Number of variables                  20   ->     37
j variable (2 values)           twinno   ->   (dropped)
xij variables:
                                earning   ->   earning1 earning2
                                schyear   ->   schyear1 schyear2
                                highqua   ->   highqua1 highqua2
                                twihigh   ->   twihigh1 twihigh2
                                    age   ->   age1 age2
                                LNandSE   ->   LNandSE1 LNandSE2
                                   part   ->   part1 part2
                                   full   ->   full1 full2
                                   self   ->   self1 self2
                                married   ->   married1 married2
                                own_exp   ->   own_exp1 own_exp2
                                bweight   ->   bweight1 bweight2
                                exp_par   ->   exp_par1 exp_par2
                                 parted   ->   parted1 parted2
                                   sm16   ->   sm161 sm162
                                   sm18   ->   sm181 sm182
                                 lnearn   ->   lnearn1 lnearn2
                                  agesq   ->   agesq1 agesq2
```

\

	family	earning1	schyear1	highqua1	twihigh1	age1	LNandSE1	part1	full1	self1	married1	own_exp1	bweight1	exp_par1	parted1	sm161	sm181	lnearn1	agesq1	earning2	schyear2	highqua2	twihigh2	age2
1	1	4.659763	11	12	12	58	1	1	0	0	0	14	92	.	.	1	1	1.538965	3364	12.98077	11	12	12	58
2	2	8.835759	12	11	11	30	0	0	1	0	0	10	81	.	.	0	1	2.178807	900	4.221893	12	11	11	30

*定义差分

gen dlnearn = lnearn1-lnearn2

gen dhigh = highqua1-highqua2

gen dtwihi=twihigh1-twihigh2

*做线性回归 工资对数的差分关于教育年限的差分（无常数项）（表 2-第四列）
reg dlnearn dhigh, noc

Source	SS	df	MS	Number of obs =	214
				F(1, 213) =	3.04
Model	1.43564569	1	1.43564569	Prob > F =	0.0826
Residual	100.55228	213	.472076434	R-squared =	0.0141
				Adj R-squared =	0.0094
Total	101.987926	214	.476579094	Root MSE =	.68708

dlnearn	Coef.	Std. Err.	t	P>|t|	[95% Conf.	Interval]
dhigh	.0393535	.0225666	1.74	0.083	-.0051289	.083836

*工具变量线性回归 (同上把 dhigh 用 dtwini 取代)（无常数项）（表 2-第五列）

Source	SS	df	MS	Number of obs =	214
				F(1, 213) =	.
Model	.096383507	1	.096383507	Prob > F =	.
Residual	101.891543	213	.47836405	R-squared =	.
				Adj R-squared =	.
Total	101.987926	214	.476579094	Root MSE =	.69164

dlnearn	Coef.	Std. Err.	t	P>|t|	[95% Conf.	Interval]
dhigh	.0773631	.0330598	2.34	0.020	.0121968	.1425294

Instrumented: dhigh
Instruments: dtwihi

*定义其他变量的差分
gen dpart = part1-part2
gen dmarried = married1-married2
gen dLNandSE = LNandSE1-LNandSE2
gen down_exp = own_exp1-own_exp2

*删去差分数据都为 0 的样本，即双胞胎数据在变量上来看完全一样的删去（删去 27 个）
drop if dlnearn==.|dhigh==.|dtwihi==.|dpart==.|dmarried==.|dln==.|down_exp==

*把一行数据拆分成两行，即恢复成原来的初始样子（样本数从 428 变为 374）
reshape long

*做线性回归 工资对数的多元线性回归（表 2-第六列）
reg lnearn highqua age agesq LNandSE part married own_exp

Source	SS	df	MS	Number of obs =	374
				F(7, 366) =	13.92
Model	25.1765463	7	3.59664948	Prob > F =	0.0000
Residual	94.5919191	366	.258447867	R-squared =	0.2102
				Adj R-squared =	0.1951
Total	119.768465	373	.321095082	Root MSE =	.50838

lnearn	Coef.	Std. Err.	t	P>|t|	[95% Conf.	Interval]
highqua	.0730804	.0110664	6.60	0.000	.0513186	.0948421
age	.059138	.024276	2.44	0.015	.0114001	.1068759
agesq	-.0008246	.0002934	-2.81	0.005	-.0014015	-.0002477
LNandSE	.0649138	.0543932	1.19	0.233	-.0420486	.1718762
part	-.0911897	.0645726	-1.41	0.159	-.2181695	.0357902
married	-.0039644	.0592855	-0.07	0.947	-.1205474	.1126186
own_exp	.0121126	.0032927	3.68	0.000	.0056376	.0185877
_cons	.0185805	.487682	0.04	0.970	-.9404298	.9775908

*工具变量线性回归 （同上把 highqua 用 twihigh 取代，对教育年限进行修正）（表 2-第七列）

ivreg lnearn (highqua=twihigh) age agesq LNandSE part married own_exp

Source	SS	df	MS	Number of obs =	374
				F(7, 366) =	10.40
Model	25.1501269	7	3.59287527	Prob > F =	0.0000
Residual	94.6183386	366	.258520051	R-squared =	0.2100
				Adj R-squared =	0.1949
Total	119.768465	373	.321095082	Root MSE =	.50845

lnearn	Coef.	Std. Err.	t	P>|t|	[95% Conf.	Interval]
highqua	.0766186	.0175752	4.36	0.000	.0420574	.1111797
age	.0585905	.0243711	2.40	0.017	.0106655	.1065154
agesq	-.000815	.0002958	-2.76	0.006	-.0013965	-.0002334
LNandSE	.0660656	.054582	1.21	0.227	-.0412681	.1733993
part	-.0887173	.0652824	-1.36	0.175	-.217093	.0396584
married	-.0041981	.0593006	-0.07	0.944	-.1208109	.1124146
own_exp	.0119957	.003324	3.61	0.000	.0054592	.0185321
_cons	-.026319	.5176053	-0.05	0.959	-1.044173	.9915346

*做关于工资对数差分的多元线性回归（无常数项）（表 2-第八列）

reg dlnearn dhigh dLNandSE dpart dmarried down_exp, noc

Source	SS	df	MS	Number of obs =	374
				F(5, 369) =	2.13
Model	5.02484483	5	1.00496897	Prob > F =	0.0610
Residual	173.919415	369	.471326329	R-squared =	0.0281
				Adj R-squared =	0.0149
Total	178.94426	374	.478460589	Root MSE =	.68653

dlnearn	Coef.	Std. Err.	t	P>|t|	[95% Conf.	Interval]
dhigh	.0394362	.0171575	2.30	0.022	.0056975	.0731748
dLNandSE	.1099124	.0854599	1.29	0.199	-.0581372	.2779619
dpart	-.099367	.0683882	-1.45	0.147	-.2338464	.0351124
dmarried	-.0531351	.0639129	-0.83	0.406	-.1788143	.0725441
down_exp	-.0014335	.0039539	-0.36	0.717	-.0092085	.0063415

*做工具变量线性回归（同上把 dhigh 用 dtwihi 取代，修正教育年限差分数据）（无常数项）（表 2-第九列）

ivreg dlnearn (dhigh=dtwihi) dLNandSE dpart dmarried down_exp, noc

Source	SS	df	MS	Number of obs =	374
Model	2.10415712	5	.420831423	F(5, 369) =	.
Residual	176.840103	369	.479241472	Prob > F =	.
				R-squared =	.
				Adj R-squared =	.
Total	178.94426	374	.478460589	Root MSE =	.69227

dlnearn	Coef.	Std. Err.	t	P>|t|	[95% Conf.	Interval]
dhigh	.0821467	.0252602	3.25	0.001	.0324746	.1318187
dLNandSE	.124242	.0863955	1.44	0.151	-.0456472	.2941312
dpart	-.101874	.0689685	-1.48	0.140	-.2374945	.0337466
dmarried	-.0475976	.0644915	-0.74	0.461	-.1744146	.0792193
down_exp	.0005013	.0040732	0.12	0.902	-.0075083	.0085109

Instrumented: dhigh
Instruments: dLNandSE dpart dmarried down_exp dtwihi

*清理数据

Clear

*定义每小时工资对数和年龄平方
gen lnearn=ln(earning)

gen agesq=age*age

*把数据对应同一家庭的两行双胞胎数据合成一行
reshape wide earning-agesq, i(family) j(twinno)

*定义差分
gen dweight = bweight1 - bweight2
gen aveweigh = (bweight1 + bweight2)/2
gen dmarry = married1 - married2
gen avemarry = (married1 + married2)/2
gen dself = self1 - self2
gen aveself = (self1+self2)/2
gen dhigh = highqua1 - highqua2
gen avehigh = (highqua1+highqua2)/2
gen dpart = part1 - part2
gen avepart = (part1+part2)/2
gen dpartexp = exp_par1 - exp_par2
gen avepexp = (exp_par1 + exp_par2)/2
gen aveped = (parted1+parted2)/2
gen dped = parted1 - parted2
gen avesm16 = (sm161+sm162)/2
gen dsm16 = sm161-sm162
gen avesm18 = (sm181+sm182)/2
gen dsm18 = sm181-sm182

*把一行数据拆分成两行，即恢复成原来的初始样子（样本数依旧 428，没有删去）
reshape long

*多变量协方差矩阵（10%显著性）（表 3-1-6 行的第一列数据）
pwcorr avehigh avemarry aveself avepart avepexp aveped aveweigh, sig st(10)

	avehigh	avemarry	aveself	avepart	avepexp	aveped	aveweigh
avehigh	1.0000						
avemarry	-0.1279*	1.0000					
	0.0081						
aveself	-0.0876*	0.0654	1.0000				
	0.0751	0.1842					
avepart	-0.2067*	0.2927*	-0.0441	1.0000			
	0.0000	0.0000	0.3709				
avepexp	-0.2124*	0.2971*	-0.0162	0.2379*	1.0000		
	0.0034	0.0000	0.8282	0.0010			
aveped	0.4908*	-0.0170	0.1900*	0.0117	0.0503	1.0000	
	0.0000	0.8133	0.0083	0.8711	0.5414		
aveweigh	0.2153*	-0.1188*	-0.0517	-0.1169*	-0.0409	0.2642*	1.0000
	0.0001	0.0295	0.3521	0.0322	0.6146	0.0007	

*同上（5%显著性）
pwcorr avehigh avemarry aveself avepart avepexp aveped aveweigh, sig st(5)
* 同上（1%显著性）
pwcorr avehigh avemarry aveself avepart avepexp aveped aveweigh, sig st(1)

*多变量协方差矩阵（差分）（10%显著性）（表 3-1-6 行的第二列数据）
pwcorr dhigh dmarry dself dpart dpartexp dped dweight, sig st(10)

	dhigh	dmarry	dself	dpart	dpartexp	dped	dweight
dhigh	1.0000						
dmarry	-0.0310 0.5221	1.0000					
dself	-0.0300 0.5430	0.1082* 0.0277	1.0000				
dpart	0.0379 0.4341	0.1054* 0.0293	-0.1845* 0.0002	1.0000			
dpartexp	-0.0093 0.8992	0.0733 0.3175	0.0090 0.9043	-0.0457 0.5338	1.0000		
dped	0.0305 0.6708	0.0374 0.6031	-0.1247* 0.0849	0.0274 0.7032	-0.1529* 0.0618	1.0000	
dweight	-0.0765 0.1617	-0.1062* 0.0518	0.0315 0.5704	0.0052 0.9245	0.1768* 0.0283	0.0812 0.3044	1.0000

*同上（5%显著性）
pwcorr dhigh dmarry dself dpart dpartexp dped dweight, sig st(5)
* 同上（1%显著性）
pwcorr dhigh dmarry dself dpart dpartexp dped dweight, sig st(1)

*多变量协方差矩阵（1%显著性）（表 4-第一列数据）
pwcorr avehigh avesm16 avesm18, sig st(1)

	avehigh	avesm16	avesm18
avehigh	1.0000		
avesm16	-0.2680* 0.0000	1.0000	
avesm18	-0.2699* 0.0000	0.7631* 0.0000	1.0000

*多变量协方差矩阵（差分）（10%显著性）（表 4-第二列数据）
pwcorr dhigh dsm16 dsm18, sig st(10)

	dhigh	dsm16	dsm18
dhigh	1.0000		
dsm16	-0.0241 0.6184	1.0000	
dsm18	-0.0541 0.2643	0.5538* 0.0000	1.0000

*同上（5%显著性）
pwcorr dhigh dsm16 dsm18, sig st(5)
* 同上（1%显著性）
pwcorr dhigh dsm16 dsm18, sig st(1)

*做线性回归 工资对数关于教育年限、年龄的二元二次项回归（表五-第一列数据）
reg lnearn highqua age agesq

Source	SS	df	MS
Model	20.7258534	3	6.9086178
Residual	118.492426	424	.279463268
Total	139.218279	427	.326038124

Number of obs = 428
F(3, 424) = 24.72
Prob > F = 0.0000
R-squared = 0.1489
Adj R-squared = 0.1429
Root MSE = .52864

lnearn	Coef.	Std. Err.	t	P>|t|	[95% Conf.	Interval]
highqua	.0767543	.0105917	7.25	0.000	.0559355	.0975731
age	.0778154	.0213949	3.64	0.000	.0357622	.1198687
agesq	-.0009675	.0002658	-3.64	0.000	-.0014899	-.0004451
_cons	-.4282208	.4347756	-0.98	0.325	-1.282805	.4263631

*同上，工具变量线性回归（把 highqua 用 sm16 取代）（表五-第二列数据）
ivreg lnearn (highqua=sm16) age agesq

Source	SS	df	MS	Number of obs =	428
				F(3, 424) =	9.11
Model	17.8956714	3	5.96522381	Prob > F =	0.0000
Residual	121.322608	424	.286138225	R-squared =	0.1285
				Adj R-squared =	0.1224
Total	139.218279	427	.326038124	Root MSE =	.53492

lnearn	Coef.	Std. Err.	t	P>|t|	[95% Conf.	Interval]
highqua	.1104606	.044446	2.49	0.013	.0230987	.1978225
age	.073574	.0223189	3.30	0.001	.0297045	.1174436
agesq	-.0008892	.000287	-3.10	0.002	-.0014533	-.0003251
_cons	-.8728873	.7192767	-1.21	0.226	-2.286679	.5409049

*同上，工具变量线性回归（把 highqua 用 sm18 取代）（表五-第三列数据）
ivreg lnearn (highqua=sm18) age agesq

Source	SS	df	MS	Number of obs =	428
				F(3, 424) =	8.91
Model	18.9371733	3	6.31239109	Prob > F =	0.0000
Residual	120.281106	424	.283681853	R-squared =	0.1360
				Adj R-squared =	0.1299
Total	139.218279	427	.326038124	Root MSE =	.53262

lnearn	Coef.	Std. Err.	t	P>|t|	[95% Conf.	Interval]
highqua	.1035503	.0445464	2.32	0.021	.015991	.1911096
age	.0744436	.0222322	3.35	0.001	.0307446	.1181426
agesq	-.0009052	.000286	-3.17	0.002	-.0014674	-.0003431
_cons	-.7817242	.7193222	-1.09	0.278	-2.195606	.6321574

log close
clear