

Quant I – Fall 2013
Controlling for a Variable:
X: Education, Y: Income and Z: Father's Education

```
use "GSS7212_R3.DTA", clear
```

```
. tab realinc if sex == 1  
*note: limiting to males
```

family income in constant \$	Freq.	Percent	Cum.
245	12	0.05	0.05
259	14	0.06	0.11
267.75	8	0.03	0.15
284.25	13	0.06	0.20
301.9	17	0.07	0.28
312.85	16	0.07	0.35
333	7	0.03	0.38
345	19	0.08	0.46
363	8	0.03	0.49
382	8	0.03	0.53
393	5	0.02	0.55
418	3	0.01	0.56
444	1	0.00	0.57
...			
91587	62	0.27	92.47
92858	65	0.28	92.76
93210	44	0.19	92.95
94738	59	0.26	93.20
94853	181	0.79	93.99
99257	88	0.38	94.37
99956	139	0.60	94.97
99988	79	0.34	95.32
102084	58	0.25	95.57
109355	21	0.09	95.66
110160	81	0.35	96.01
110895	68	0.30	96.31
115841	71	0.31	96.61
119606.1	64	0.28	96.89
128141	37	0.16	97.05
128434.6	154	0.67	97.72
137237.8	121	0.53	98.25
141038	95	0.41	98.66
143778	19	0.08	98.74
144502.7	127	0.55	99.29
146153.7	66	0.29	99.58
155140	76	0.33	99.91
162607	21	0.09	100.00
Total	23,042	100.00	

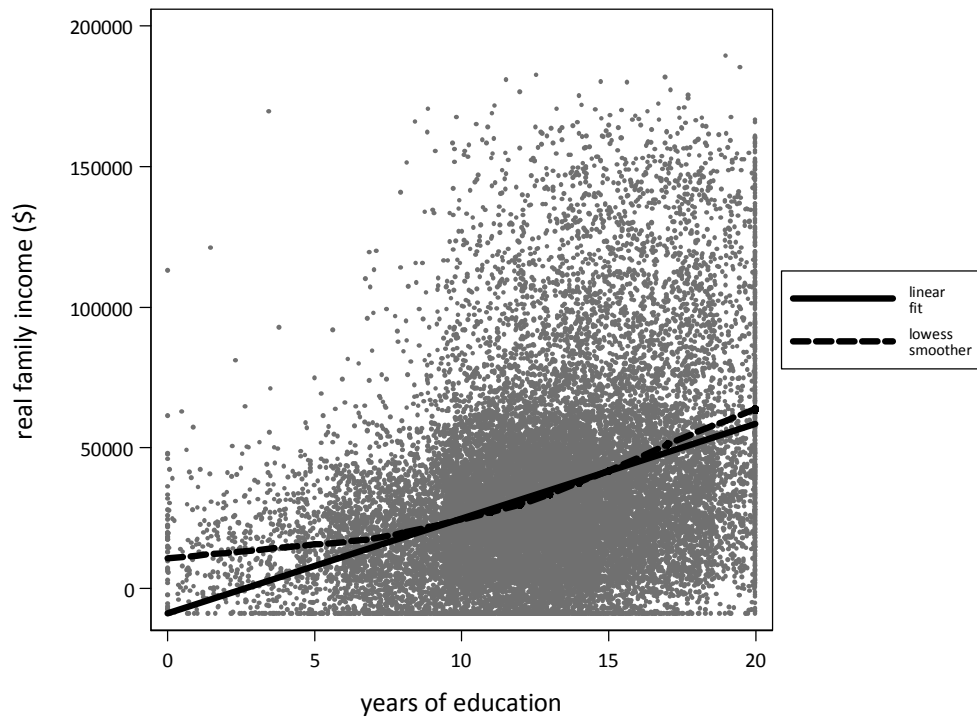
```
. table educ [pw=wtssall] if sex == 1,c(mean realinc)
```

highest year of school completed	mean(realinc)
0	11249.46173
1	15717.5519
2	19277.20657
3	13516.19767
4	14818.9965
5	17199.89956
6	15509.03308
7	19051.1232
8	21364.94477
9	21832.42802
10	24811.19666
11	27352.98068
12	32599.01844
13	35594.14105
14	39814.4767
15	38056.38513
16	52124.11244
17	52730.12397
18	59423.68176
19	66262.5068
20	66179.29469
dk	21622.38492
na	35446.62074

```

twoway (scatter realinc educ, msize(tiny) mc(gs7) jitter(20)) (lfit realinc
educ, lw(thick) lc(black)) (lowess realinc educ, lc(black) lp(dash) lw(thick))
if sex==1, legend(pos(3) order(2 3) label(2 "linear" "fit") label(3 "lowess"
"smoother")) cols(1) size(vsmall)) ytitle("real family income ($)")
xtitle("years of education")

```

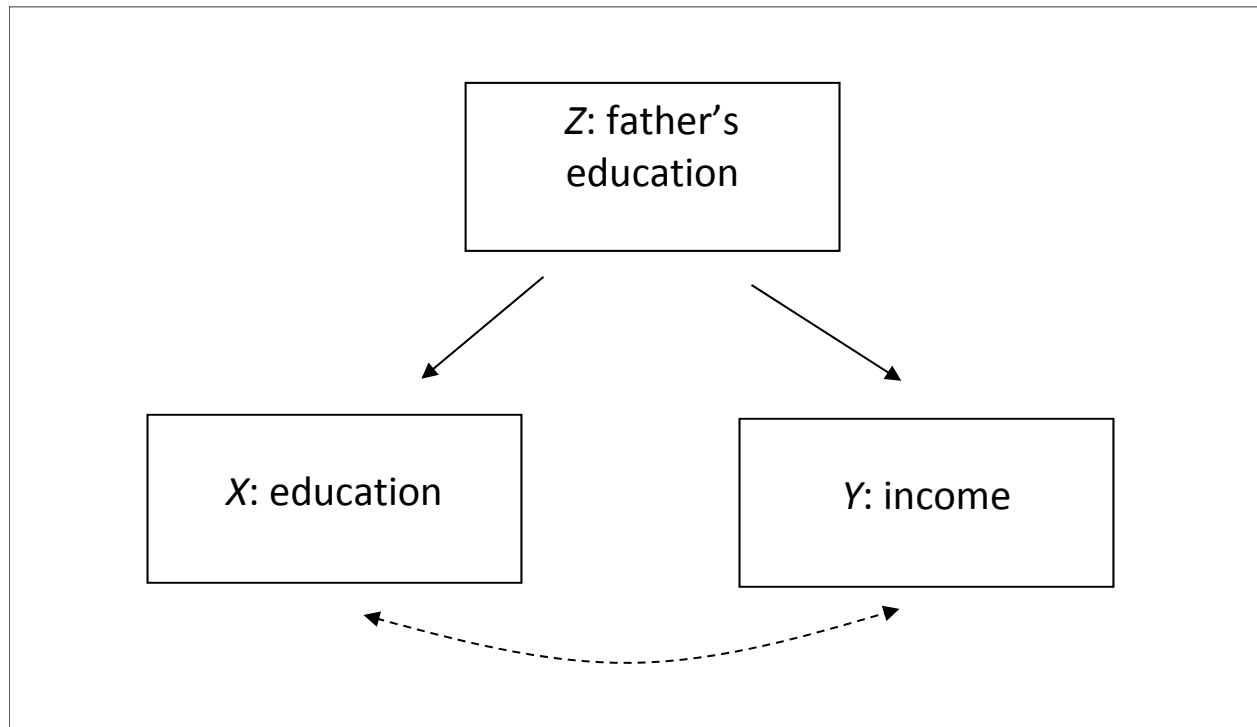


```
. reg realinc educ if sex==1
```

Source	SS	df	MS	Number of obs = 23009		
Model	2.9652e+12	1	2.9652e+12	F(1, 23007) = 3795.81		
Residual	1.7973e+13	23007	781187009	Prob > F = 0.0000		
Total	2.0938e+13	23008	910031756	R-squared = 0.1416		
				Adj R-squared = 0.1416		
				Root MSE = 27950		

realinc	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]	
educ	3379.168	54.84755	61.61	0.000	3271.663	3486.673
_cons	-9003.901	732.8073	-12.29	0.000	-10440.25	-7567.55

A POTENTIAL CONFOUND



```
. tab paeduc if sex==1
```

highest year school completed, father	Freq.	Percent	Cum.
0	526	2.88	2.88
1	55	0.30	3.18
2	159	0.87	4.05
3	519	2.84	6.90
4	494	2.71	9.61
5	449	2.46	12.07
6	1,031	5.65	17.72
7	512	2.81	20.52
8	2,632	14.42	34.94
9	616	3.38	38.32
10	834	4.57	42.89
11	523	2.87	45.75
12	5,057	27.71	73.46
13	505	2.77	76.23
14	1,134	6.21	82.44
15	214	1.17	83.62
16	1,762	9.65	93.27
17	191	1.05	94.32
18	465	2.55	96.87
19	152	0.83	97.70
20	420	2.30	100.00
Total	18,250	100.00	

```
. pwcorr educ realinc paeduc if sex==1
```

	educ	realinc	paeduc
educ	1.0000		
realinc	0.3763	1.0000	
paeduc	0.4788	0.2045	1.0000

```
*recode paeduc:
```

```
. recode paeduc (0/8=1) (9/11=2) (12=3) (13/15=4) (16=5) (17/20=6),
gen(paeduc_rec)
(40061 differences between paeduc and paeduc_rec)
```

```
. label def paeduc_rec 1 "8th gr or less" 2 "HS dropout" 3 "HS dip" 4 "some
coll" 5 "BA" 6 "post grad"
```

```
. label values paeduc_rec paeduc_rec
```

```
. tab paeduc_rec
```

RECODE of paeduc (highest year school completed, father)	Freq.	Percent	Cum.
8th gr or less	14,276	35.54	35.54
HS dropout	4,481	11.15	46.69
HS dip	11,212	27.91	74.60
some coll	4,007	9.97	84.57
BA	3,651	9.09	93.66
post grad	2,546	6.34	100.00
Total	40,173	100.00	

CONTROLLING FOR Z: FEWEST ASSUMPTIONS

- no linearity
- different xy relationship for different values of z

```
twoway (scatter realinc educ, jitter(20) msize(tiny) mc(gs7)) (lowess realinc  
educ, lw(thick) lc(black)) if paeduc_rec==1 & sex==1, title("Father's  
Education:" "8th grade or less") aspect(1) legend(off) ytitle("real family  
income ($)") xtitle(education)
```

```
graph save "paeduc1.gph", replace
```

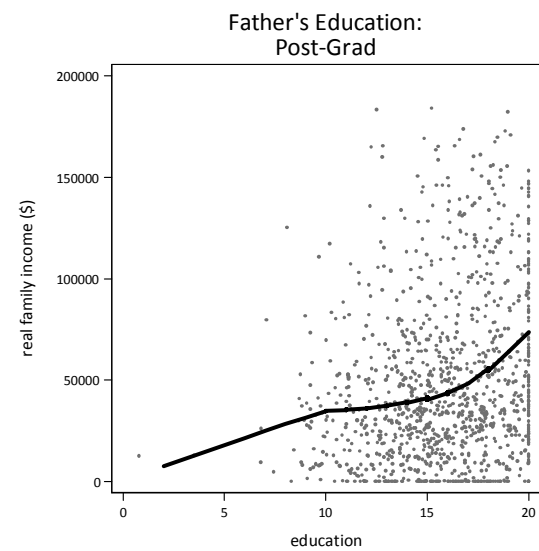
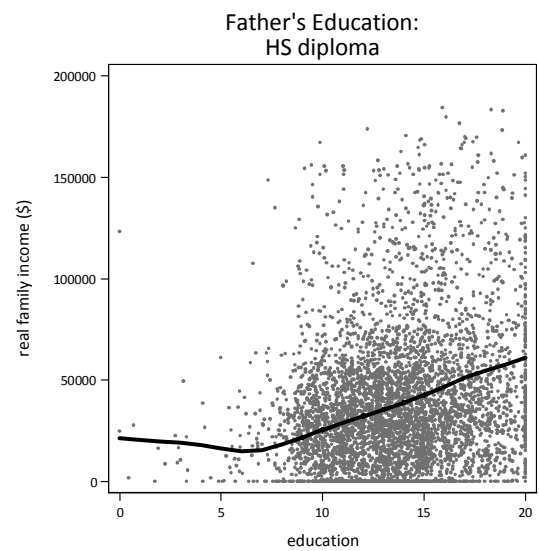
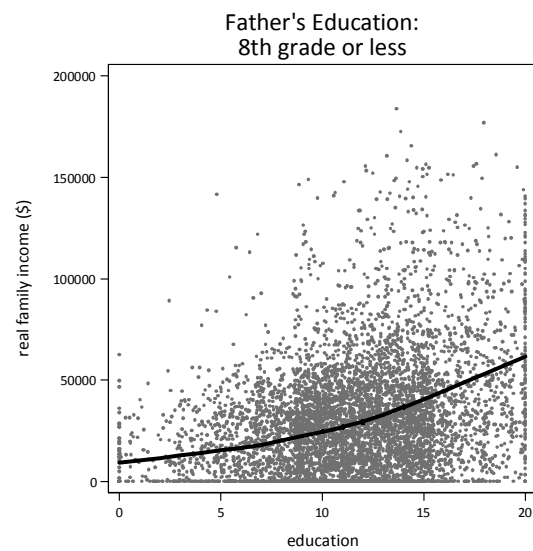
```
twoway (scatter realinc educ, jitter(20) msize(tiny) mc(gs7)) (lowess realinc  
educ, lw(thick) lc(black)) if paeduc_rec==3 & sex==1, title("Father's  
Education:" "HS diploma") aspect(1) legend(off) ytitle("real family income  
($)") xtitle(education)
```

```
graph save "paeduc3.gph", replace
```

```
twoway (scatter realinc educ, jitter(20) msize(tiny) mc(gs7)) (lowess realinc  
educ, lw(thick) lc(black)) if paeduc_rec==6 & sex==1, title("Father's  
Education:" "Post-Grad") aspect(1) legend(off) ytitle("real family income  
($)") xtitle(education)
```

```
graph save "paeduc5.gph", replace
```

```
graph combine "paeduc1.gph" "paeduc3.gph" "paeduc6.gph", colfirst ycommon  
xcommon rows(1) ysize(4) xsize(10)
```



CONTROLLING FOR Z: MORE ASSUMPTIONS, BUT NOT TYPICAL APPROACH

$$realincome = \beta_{0J} + \beta_{1J}(educyears) + u, J = 1, \dots, 6$$

- linearity
- xy (linear) relationship varies with different values of z

. by paeduc_rec, sort: reg realinc educ if sex==1

-> paeduc_rec = 8th gr or less

Source	SS	df	MS	Number of obs	=	5933
Model	5.8475e+11	1	5.8475e+11	F(1, 5931)	=	1074.45
Residual	3.2279e+12	5931	544234563	Prob > F	=	0.0000
				R-squared	=	0.1534
				Adj R-squared	=	0.1532
Total	3.8126e+12	5932	642718505	Root MSE	=	23329

realinc	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]
educ	2839.174	86.6163	32.78	0.000	2669.375 3008.974
_cons	-2984.487	1056.329	-2.83	0.005	-5055.276 -913.6974

-> paeduc_rec = HS dropout

Source	SS	df	MS	Number of obs	=	1855
Model	2.1853e+11	1	2.1853e+11	F(1, 1853)	=	296.18
Residual	1.3672e+12	1853	737853646	Prob > F	=	0.0000
				R-squared	=	0.1378
				Adj R-squared	=	0.1373
Total	1.5858e+12	1854	855327518	Root MSE	=	27163

realinc	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]
educ	4122.759	239.5593	17.21	0.000	3652.924 4592.593
_cons	-18406.75	3213.193	-5.73	0.000	-24708.61 -12104.89

-> paeduc_rec = HS dip

Source	SS	df	MS	Number of obs = 4691			
Model	3.6785e+11	1	3.6785e+11	F(1, 4689)	=	429.53	
Residual	4.0156e+12	4689	856386400	Prob > F	=	0.0000	
				R-squared	=	0.0839	
				Adj R-squared	=	0.0837	
Total	4.3834e+12	4690	934635773	Root MSE	=	29264	

realinc	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]	
educ	3442.621	166.1082	20.73	0.000	3116.97	3768.271
_cons	-8751.409	2320.572	-3.77	0.000	-13300.82	-4201.997

-> paeduc_rec = some coll

Source	SS	df	MS	Number of obs = 1729			
Model	1.5135e+11	1	1.5135e+11	F(1, 1727)	=	144.89	
Residual	1.8039e+12	1727	1.0446e+09	Prob > F	=	0.0000	
				R-squared	=	0.0774	
				Adj R-squared	=	0.0769	
Total	1.9553e+12	1728	1.1315e+09	Root MSE	=	32320	

realinc	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]	
educ	3827.571	317.9813	12.04	0.000	3203.902	4451.24
_cons	-13570.78	4693.648	-2.89	0.004	-22776.62	-4364.948

-> paeduc_rec = BA

Source	SS	df	MS	Number of obs = 1615			
Model	1.1590e+11	1	1.1590e+11	F(1, 1613)	=	88.86	
Residual	2.1038e+12	1613	1.3043e+09	Prob > F	=	0.0000	
				R-squared	=	0.0522	
				Adj R-squared	=	0.0516	
Total	2.2197e+12	1614	1.3753e+09	Root MSE	=	36115	

realinc	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]	
educ	3396.117	360.2629	9.43	0.000	2689.484	4102.749
_cons	-4869.59	5577.424	-0.87	0.383	-15809.35	6070.169

-> paeduc_rec = post grad

Source	SS	df	MS	Number of obs = 1121		
Model	1.3336e+11	1	1.3336e+11	F(1, 1119) = 93.68		
Residual	1.5928e+12	1119	1.4235e+09	Prob > F = 0.0000		
Total	1.7262e+12	1120	1.5412e+09	R-squared = 0.0773		
				Adj R-squared = 0.0764		
				Root MSE = 37729		

realinc	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]	
educ	4242.855	438.3537	9.68	0.000	3382.767	5102.942
_cons	-20521.43	7098.973	-2.89	0.004	-34450.23	-6592.638

CONTROLLING FOR Z: TYPICAL APPROACH

$$realincome = \beta_0 + \beta_1(educyears) + \beta_2(paeduc) + u$$

- **linearity**
- **xy (linear) relationship is constant across different values of z**
- treat paeduc as interval-level variable

```
. reg realinc educ paeduc if sex==1
```

Source	SS	df	MS	Number of obs = 16944		
Model	2.1429e+12	2	1.0715e+12	F(2, 16941) = 1282.24		
Residual	1.4156e+13	16941	835615329	Prob > F = 0.0000		
Total	1.6299e+13	16943	961995210	R-squared = 0.1315		
				Adj R-squared = 0.1314		
				Root MSE = 28907		

realinc	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]	
educ	3259.935	77.89661	41.85	0.000	3107.25	3412.621
paeduc	301.6454	57.71808	5.23	0.000	188.512	414.7789
_cons	-9596.13	948.1955	-10.12	0.000	-11454.69	-7737.568