

```
use cex8_06.dta
```

The data set has, among other things, information on individual earnings, gender age and years of education

Suppose there is an interest in trying to establish if the local area's human capital level has externalities for individual earnings

You may be tempted (though beware of endogeneity) to regress individual earnings on a measure of the average level of years of education in the local region

To obtain regional averages use the following commands

```
egen reged=mean(yearsed),by(region)
tab reged
```

```
xi:reg hourpay sex yearsed age london reged
```

Source	SS	df	MS	Number of obs = 10016		
Model	55802.9557	5	11160.5911	F(5, 10010)	=	615.74
Residual	181434.924	10010	18.125367	Prob > F	=	0.0000
-----				R-squared	=	0.2352
-----				Adj R-squared	=	0.2348
Total	237237.88	10015	23.6882556	Root MSE	=	4.2574

hourpay	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]	
sex	-3.108907	.0853764	-36.41	0.000	-3.276262	-2.941553
yearsied	.4913566	.0138883	35.38	0.000	.4641328	.5185805
age	.0684965	.004346	15.76	0.000	.0599775	.0770156
london	1.264466	.1793757	7.05	0.000	.9128535	1.616078
reged	.3149518	.0671497	4.69	0.000	.1833249	.4465786
_cons	-.172124	.8643492	-0.20	0.842	-1.866422	1.522174

However this regression fails to account for the effects of clustering on the regional average variable (there are only 20 regions in the data set and hence only 20 points of variation in this variable). So always use the cluster option

```
. xi:reg hourpay sex yearsied age london reged, cluster(region)
```

```
Linear regression
```

				Number of obs = 10016		
				F(5, 18)	=	327.62
				Prob > F	=	0.0000
				R-squared	=	0.2352
				Root MSE	=	4.2574
(Std. Err. adjusted for 19 clusters in region)						

hourpay	Coef.	Robust Std. Err.	t	P> t	[95% Conf. Interval]	
sex	-3.108907	.2440393	-12.74	0.000	-3.621615	-2.5962
yearsied	.4913566	.0385925	12.73	0.000	.4102768	.5724364
age	.0684965	.0105139	6.51	0.000	.0464078	.0905853
london	1.264466	.4691776	2.70	0.015	.2787603	2.250171
reged	.3149518	.1661497	1.90	0.074	-.0341157	.6640192
_cons	-.172124	2.261663	-0.08	0.940	-4.923701	4.579453