

An example of residual analysis: Study of birth rates

Key variables:

gnpcap as people have more money, they choose to have fewer children
 -- they invest more in child "quality"
 urban people in urban areas also choose to have fewer children

Both of these variables are reflections of the quantity-quality tradeoff hypothesis

Promising variables:

infmort as children are less likely to die, people need to bear fewer in order to have a certain
 number of surviving children
 school2 more educated people have fewer children

ANALYSIS OF KEY VARIABLES

```
. use nations
. regress birth gnpcap urban
```

Source	SS	df	MS	Number of obs =	108
Model	10796.488	2	5398.24399	F(2, 105) =	64.22
Residual	8825.5861	105	84.053201	Prob > F =	0.0000
Total	19622.0741	107	183.38387	R-squared =	0.5502
				Adj R-squared =	0.5417
				Root MSE =	9.1681

birth	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]
gnpcap	-.000842	.0002637	-3.193	0.002	-.0013649 -.0003191
urban	-.2823184	.0462191	-6.108	0.000	-.3739624 -.1906744
_cons	48.85603	1.986909	24.589	0.000	44.91635 52.7957

ANALYSIS OF PROMISING VARIABLES

The first step is to computer residuals -- we need to purge birth, infmort, and school2 of the effects of gnpcap and urban. I've called the resulting residuals blresid, infm, and sc2m

```
. predict blresid, resid
(1 missing value generated)
```

* blresid will be the dependent variable in the analysis of school and infant mortality, but first THOSE variables need to be purged of their relationship to gnpcap and urban

```
. regress school2 gnpcap urban
```

Source	SS	df	MS	Number of obs =	103
Model	62012.9189	2	31006.4594	F(2, 100) =	118.05
				Prob > F =	0.0000

```
Residual | 26264.4598 100 262.644598
-----+-----
Total | 88277.3786 102 865.464496
```

R-squared = 0.7025
Adj R-squared = 0.6965
Root MSE = 16.206

```
-----+-----
school2 |      Coef.  Std. Err.      t    P>|t|      [95% Conf. Interval]
-----+-----
gnpcap |   .0020547   .0004987    4.120  0.000   .0010652   .0030441
urban  |   .6889479   .0838135    8.220  0.000   .5226644   .8552315
_cons  |   7.577034   3.575473    2.119  0.037   .4833982  14.67067
-----+-----
```

```
. predict sc2m, resid
(6 missing values generated)
```

```
. regress infmort gnpcap urban
```

```
Source |      SS      df      MS                Number of obs =      108
-----+-----+-----+-----                F( 2, 105) =      76.79
Model | 150958.509    2   75479.2543            Prob > F      = 0.0000
Residual | 103201.816  105  982.874433            R-squared     = 0.5939
-----+-----+-----+-----            Adj R-squared = 0.5862
Total | 254160.324  107 2375.33013            Root MSE     = 31.351
-----+-----+-----+-----
```

```
-----+-----
infmort |      Coef.  Std. Err.      t    P>|t|      [95% Conf. Interval]
-----+-----+-----+-----
gnpcap |  -.0024474   .0009018   -2.714  0.008   -.0042355  -.0006593
urban  | -1.160319   .1580497   -7.341  0.000   -1.473702  -.8469357
_cons  | 131.2357    6.794381   19.315  0.000   117.7637   144.7077
-----+-----+-----+-----
```

```
. predict infm, resid
(1 missing value generated)
```

. * we now have three variables that are purged of the effects of urban and gnpcap. We do residual analysis

```
. regress blresid sc2m infm
```

```
Source |      SS      df      MS                Number of obs =      103
-----+-----+-----+-----                F( 2, 100) =      67.90
Model | 4765.04887    2   2382.52443            Prob > F      = 0.0000
Residual | 3508.65027  100  35.0865027            R-squared     = 0.5759
-----+-----+-----+-----            Adj R-squared = 0.5674
Total | 8273.69913  102  81.1146974            Root MSE     = 5.9234
-----+-----+-----+-----
```

```
-----+-----
blresid |      Coef.  Std. Err.      t    P>|t|      [95% Conf. Interval]
-----+-----+-----+-----
sc2m |  -.1660183   .047051   -3.528  0.001   -.2593661  -.0726705
infm |   .1575793   .0245993    6.406  0.000   .108775   .2063836
_cons |  -.1315096   .5842425   -0.225  0.822   -1.29063   1.027611
-----+-----+-----+-----
```

```
. predict b2resid, resid
(6 missing values generated)
```

These results are VERY different from those we would obtain if we had simply put all variables into a single model. In that case, neither gnpcap nor urban are significant!

```
. regress birth gnpcap infmort school2 urban
```

Source	SS	df	MS			
Model	15079.7314	4	3769.93285	Number of obs =	103	
Residual	3506.24917	98	35.7780528	F(4, 98) =	105.37	
Total	18585.9806	102	182.215496	Prob > F =	0.0000	
				R-squared =	0.8113	
				Adj R-squared =	0.8036	
				Root MSE =	5.9815	

birth	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]	
gnpcap	-.0000731	.0001991	-0.367	0.714	-.0004681	.000322
infmort	.1576034	.0248464	6.343	0.000	.1082965	.2069103
school2	-.1659893	.0475169	-3.493	0.001	-.2602849	-.0716936
urban	.007387	.0408106	0.181	0.857	-.0736003	.0883744
_cons	29.52888	3.695856	7.990	0.000	22.19457	36.86318