Stata How To: Controlling for Variables, a.k.a. Regressions with Fixed Effects
Wellesley College
Revised November 2012

When you’re working with panel data, you often need to run your regressions with fixed effects. This means that you want to control for a particular categorical variable or variables using a set of dummies. For example, if you have state panel data then you need dummy variables for \(N(\text{states})-1\), if you have individual panel data then you need dummy variables for \(N(\text{individuals})-1\), etc.

There are several ways to run regressions with fixed effects in Stata. Think carefully about the data and the assumptions underlying each of these approaches to decide which option you should use. We’ll start with the most basic option and work towards the most complex option:

**Option 1: Create Dummy Variables for Each Category of a Variable**

You can manually create dummy variables (in this case for geographic regions, as denoted by the region variable):

```
    tab region, missing //always tab the variable first and check for missing values!
    gen regionNE = region=="Northeast"
    gen regionNW = region=="Northwest"
    // etc
```

However, *this would be a waste of your time* because with one easy command you can tell Stata to create the dummies for you! There are 2 ways to do this, each producing (almost) the same result:

```
    tab region, gen(regiondum) /*creates new variables, 1 for each category of the region
                                 variable (regiondum1, regiondum2, etc) */
    xi i.region /* creates new variables, 1 variable for all but 1 category of the
                   region variable (_Iregion_1, _Iregion_2, etc) */
```

You then include these variables in your regression:

```
    reg Y X1 X2 regiondum1 regiondum2 regiondumN //remember: omit 1 dummy variable!
    reg Y X1 X2 _Iregion* /*The xi command automatically creates 1 dummy variable less than
                           total number of categories, so you don't have to worry about omitting
                           1 dummy variable */
```

Tip: In the command above, the asterisk (*) is a wildcard that represents many characters in the variable name. So _Iregion* tells Stata to include all variables beginning with “_Iregion” in the regression. Think carefully before using wildcards, though, because Stata will include all variables that begin or end with the specified prefix.

Even faster, you can create the dummy variables and run the regression using one command:

```
    xi: regress Y X1 X2 i.region
```

**Option 2: areg**

Alternatively to creating a bunch of dummies using the ‘brute force’ method, you can also implement a fixed effects regression using the `areg` command. The only difference is `areg` does not display the coefficients for each dummy variable. So:

```
    areg Y X1 X2, absorb(region)
```

is the same as:

```
    xi: regress Y X1 X2 i.region
```

Note that you can only ‘absorb’ one variable. If you need to control for multiple variables, you can use the `xi` command to create them and then include the `_lcategory` variables in your regression with `areg`. 
Option 3: xtreg

There is yet a third way to create fixed-effects linear models for panel data. Additionally, xtreg can be used to create between-effects and random-effects regressions. Running xtreg with fixed effects is the same as including a dummy variable for each category minus 1 (e.g., region):

```
xtrg Y X1 X2 X3 X4, i(region) fe  // fe option denotes fixed effects
```

which is the same as
```
xtri: regress Y X1 X2 X3 X4 i.region
```

which is the same as
```
areg Y X1 X2 X3 X4, absorb(region)
```

The main difference between xtreg (with the fe option) and areg is that areg’s overall $F$ test is for the model including the dummy variables (even though we absorbed them and do not see the estimated coefficients), while xtreg output includes the overall $F$ test only of those terms in which we are interested (the ones that are listed in the output) instead of showing the $F$ test including the absorbed terms.

To create between-effects or random-effects models, you change the option at the end of the command:

```
xtrg Y X1 X2 X3 X4, i(region) be  // be option denotes between effects
```
```
xtrg Y X1 X2 X3 X4, i(region) re  // re option denotes random effects
```

xtreg has many other options, which you can read more about in the help (type help xtreg).