

## Basic Panel Data Commands in STATA

Panel data refers to data that follows a cross section over time—for example, a sample of individuals surveyed repeatedly for a number of years or data for all 50 states for all Census years.

- `reshape` There are many ways to organize panel data. Data with one observation for each cross section and time period is called the “long” form of the data:

state	year	unem
1	1980	.052
2	1980	.074
3	1980	.065
4	1980	.031
1	1990	.081
2	1990	.032
3	1990	.045
4	1990	.043

Data with one observation for each cross section unit is called the “wide” form of the data:

state	unem1980	unem1990
1	.052	.081
2	.074	.032
3	.065	.045
4	.031	.043

Stata can easily go back and forth between the two types using the `reshape` command.

```
reshape wide unemrate taxrate, i(state) j(year)
reshape long unemrate taxrate, i(state) j(year)
```

This is a fairly complex command to use, so read the manual documentation before doing it. Make sure that if your data is in the “wide” form, that the variable stem name (unem in the example above) is consistent across years and the year suffix is consistent. (For example, if your year suffix is 98, 99, 00, Stata will put 00 as a year before 99.)

A few more useful panel data commands to look up:

- The `by:` construction. We covered this before, but you will use it a lot with panels.
- `collapse:` makes a dataset of summary data statistics. For example, you can take a dataset of individual level data and collapse it into mean statistics by state.  
`collapse (mean) income (median) medinc=income (sum) population, by(year)`
- `egen:` Extensions to generate. We covered this before, but you will use it a lot with panels.

There are 4 options for doing FIXED EFFECT models in STATA.

Suppose data consist of a panel of 50 states observed over time.

1. Make the demeaning transformation (no reason to do this—just illustrating the commands)

```
egen avg_wage = mean(wage), by(state)
gen delt_wage = wage - avg_wage
egen avg_exp = mean(exp), by(state)
gen delt_exp = exp - avg_exp

reg delta_wage delta_exp
```

2. Use the `xi` prefix

```
xi: reg wage experience education i.state
```

This will give you output with all of the state fixed effect coefficients reported. You will notice in your variable list that STATA has added the set of generated dummy variables. Options are available to control which category is omitted.

3. Use the `absorb` command to run the same regression as in (2) but suppressing the output for the individual dummy variables

```
areg wage experience education, absorb(state)
```

4. Use STATA's panel regression command `xtreg`. Note that all the documentation on XT commands is in a separate manual.

`iis state` declares the cross sectional units are indicated by the variable *state*

`tis year` declares time periods are indicated by *year*.

Or use `tsset panelvar timevar` (so following this example `tsset state year`) to declare your data to be a panel. There are a lot of options on this—check the help menu.

After you let STATA know how the data is organized you can use the `xtreg` command.

```
xtreg wage experience education , fe
```

Note that this is the same command to use for random effects estimators, just with the `re` option instead of `fe`

These options are all equivalent in terms of the coefficient estimates. The advantage of creating the dummy variables explicitly is that sometimes you actually want to examine their values—for example, you might be interested in looking at fixed effects for particular states and see if they make sense.

However, often you DON'T want all that info—suppose you are looking at NLSY that follows thousands of people over time. You don't care about Joe Bob's fixed effect and don't want it in the output. Use `xtreg` or `areg` in that case. `areg` is somewhat faster in older versions of STATA.

To be clear, dummy variables *are* included in the `xtreg` and `areg` regressions, *but they will be suppressed in the output*.