Entering Factor variables in regressions in Stata

Factor variables are a way to quickly enter dummy variables or interactions in a regression model in stata without creating new variables first. They have an added benefit in that post-estimation commands better "understand" the components of your model.

There are two main prefixes that you can use with factor variables:

i. designates a categorical variable

c. designates a continuous variable

For example, if you had a race variable with 4 categories (1=black, 2=white, 3=Asian, 4=other), you could not enter that race variable directly into your equation. You'd first have to create four dummy variables and enter three of those into your equation, but the two commands below would be identical:

```
. reg del white Asian other
```

```
. reg del i.race
```

If you preferred to omit a different category, you can do so with the ib. prefix. To omit the third category, for example:

```
.reg del ib3.race
```

On the other hand, if you have a continuous variable, you can enter c. before it to designate it as such. Alone, this doesn't change anything. The following two would be identical, where sbond is a continuous measure of school bonding:

```
.reg del sbonds
```

```
.reg del c.sbonds
```

But the c. prefix can come in handy when you want to enter interaction terms. For example, if you wanted to test the interaction of race and school bonds, you could enter the following and the main effects and all interaction terms would be automatically created:

```
.reg del i.race##c.sbonds
```

Note: if you only include one hash mark (#) only the interaction terms are included. This is usually not what you want.

Finally, if you wanted to enter squared terms, you can do so using the c. prefix as well:

```
. reg del c.sbonds##c.sbonds
```

You can even specify complicated multiway interactions along with squared terms. Just know what you're doing:

```
.reg del c.sbonds##c.sbonds##i.race##i.male
```

For further information: http://www.stata.com/help.cgi?fvvarlist