

LOOPING OVER ALL COMBINATIONS OF A VARIABLE LIST

A researcher may want to run some analysis on all combinations of a specific list of variables. This may be of interest, for example, in order to select the combination of controls to use in a model out of a larger pool of candidate variables based on a criteria of interest (e.g., R-squared).

The command `tuples` provides a simple and flexible way to loop over all combinations of a list of items – in this case, a list of variables.

Before we start, install the command:

```
ssc install tuples
```

Suppose we have an outcome of interest (y) and three candidate covariates (x_1 , x_2 and x_3). Suppose we wish to select the combination of these covariates that maximizes the adjusted R-squared. The first step involves the following syntax:

```
tuples x1 x2 x3
```

This produces $2^n - 1$ local macros (n denoting the number of elements in the list—3 in this case), each of which contains one possible combination of the elements we supplied to the list. These local macros are named “`tuples1`”, “`tuples2`”, ..., “`tuples7`”. To illustrate, the following list shows the mapping between each of these macros and the string of characters they contain.

```
tuples1 : x1  
tuples2 : x2  
tuples3 : x3  
tuples4: x2 x3  
tuples5: x1 x3  
tuples6: x1 x2  
tuples7: x1 x2 x3
```

The number of macros produced is stored in the local macro “`ntuples`” (in this case, 7). Now that each possible combination of x_1 , x_2 and x_3 has been stored in the locals, we can call each of these in a loop and obtain the adjusted R-squared from each combination of covariates. We can then output the variables used in each specification and the corresponding adjusted R-squared to a csv file.

```
file open output using "${output}", write replace  
file write output "Model, Adjusted R-squared" _n
```

Use the `tuples` command on the list of candidate variables:

```
tuples x1 x2 x3
```

Loop over the number of combinations (and hence the number of macros containing each combination), stored in the local ``ntuples'`:

```
forvalues i = 1/`ntuples' {  
reg y `tuple`i' // Regress y on each combination (stored in `tuple`i')  
file write output "`tuple`i'", `e(r2_a)'" _n  
}  
capture file close output // Output results to a CSV file:
```

There may be instances where we wish to consider multiple variables to be one single element. For example, consider the list of covariates `x1_1`, `x1_2`, `x2`, `x3`. Suppose that we wish to loop over all combinations of these covariates, but that `x1_1` and `x1_2` must always be together. We can “bind” these two covariates together by using double quotes so that the tuples command treats “`x1_1 x1_2`” as a single element. The tuples syntax would now be:

```
tuples "x1_1 x1_2" x2 x3
```

For more details about the tuples command type `help tuples` after installing it.

References:

Joseph N. Luchman & Daniel Klein & Nicholas J. Cox, 2006. "TUPLES: Stata module for selecting all possible tuples from a list," Statistical Software Components S456797, Boston College Department of Economics, revised 17 Feb 2019.