

Tips and Tricks: Loops and Dupes

Jen Cocohoba, Pharm.D., MAS
Health Sciences Associate Clinical Professor
UCSF School of Pharmacy

NOTE: tips tested for STATA v.12, though most info will be relevant for STATA 13

Programming loops

Table 1

	Men	Women	P-value
Age, mean (SD)			
# Reported adverse effects, mean (SD)			
Adverse effect severity index, mean(SD)			

- Example
 - Determine whether age, number of side effects, and scaled severity of side effects differ by gender
- Start programming...


```
ttest age, by(sex)
ttest numsidefx, by(sex)
ttest severity, by(sex)
```

Simple Loop Syntax

```
foreach var in variable1 variable2 {
  firstcommand `var'
  secondcommand `var'
}
```

Diagram labels: "List of variables" points to the list of variables in the loop header. "Loop begin" points to the opening brace. "Loop end" points to the closing brace. "Perform these commands, replacing the 'var' with the variables in the list. NOTE the special apostrophe marks (the first one lies below the ~ on the keyboard, the other is a normal apostrophe)" points to the backtick-quoted variable names.

* NOTES

1. Open brace must appear on the same line as the foreach command.
2. Nothing may follow the open brace (except for comments)
3. The first command must be on a separate line
4. The close brace must be on its own line

Simple Loop Syntax

```
foreach var in age numsidefx severity {
  ttest `var', by(gender)
}
```

Diagram labels: "Variables" points to the list of variables in the loop header. "Loop begin" points to the opening brace. "Loop end" points to the closing brace. "Perform this command, replacing the generic placeholder 'var' with variables I specified in my list. NOTE the special apostrophe marks (the first one lies below the ~ on the keyboard, the other is a normal apostrophe)" points to the backtick-quoted variable name.

* NOTES

1. Open brace must appear on the same line as the foreach command.
2. Nothing may follow the open brace (except for comments)
3. The first command must be on a separate line
4. The close brace must be on its own line

Loops in do files - examples

```
***** Frequencies for Table 1 - baseline characteristics *****
/* Get proportions of categorical variables and estimates of missing */
foreach var in agecat afamer highschool income employed insurtype {
  tab `var' sex, missing col chi2
}

/* Get means, standard deviations, and test for differences of continuous variables */
foreach var in ageatvisit cd4 log10v1 {
  bysort sex: sum `var', detail
  ttest `var', by(sex)
}

***** Table 2 Odds Ratios *****
/* Get all of the univariate odds ratios for important factors with guideline not recommended
regimens */
foreach var in highschool income employed drugcover depressed {
  xi: logistic guideline i.`var'
}
```

Handling Duplicate Observations



- Just want to find the duplicates?

```
duplicates list
variable1 variable2
```

Or...

```
duplicates report
variable1 variable2
```

or...

```
duplicates tag variable1
variable2, gen(newvar)
```

Handling Duplicate Observations

Usual goal is to either find the duplicates or get rid of them ... or both

Method 1: 2 step process of tagging then dropping

```
duplicates tag variable1 variable2, gen(newvar)  
duplicates drop if newvar==1
```

Method 2: just dropping

```
duplicates drop variable1 variable2, force
```