SHAPE DATA

In Stata

YouTube Channel

Project 1

Rochester Econometrics, Stata for Undergraduates

Each video 5-10 minutes:

- Workflow, Data sets, Data types, Using 'by', Tests, Graphs, Regression, Storing information, Combing data sets, Utilities, Help, Log, & .do files, Exploring data.

Overview

Project 1

Why and when to use each tools.

- Reshape   Combine & Separate; Vars & Obs
- Merge   Combine; Datasets, Add Vars & Obs

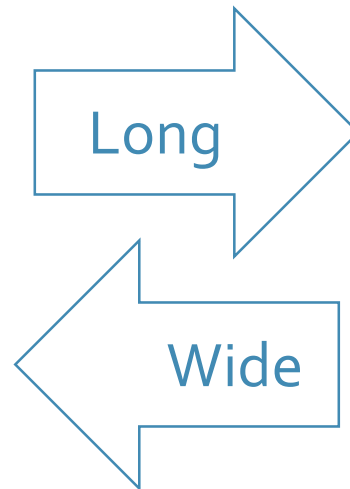- Append   Combine; Datasets, Add Vars
- Collapse   Combine; Obs

Variables (Cols), Observations (Rows)

Reshape

- reshape **long** *inc* , **i**(*id*) **j**(*year*)

(wide form)

i		.... x_ij		
id	sex	inc80	inc81	inc82
1	0	5000	5500	6000
2	1	2000	2200	3300
3	0	3000	2000	1000

Long

Wide

(long form)

i	j		x_ij
id	year	sex	inc
1	80	0	5000
1	81	0	5500
1	82	0	6000
2	80	1	2000
2	81	1	2200
2	82	1	3300
3	80	0	3000
3	81	0	2000
3	82	0	1000

Reshape Long   Convert data from wide to long

- **reshape long** *stub***, i(** *id* **) j(** *suffix* **)** [options]

Options	Descriptions
long	reshape data from **wide** to **long**
stub	variable to be converted
i(*id* **)**	unique id for records in **wide** format
j(*suffix* **)**	*stub* suffix **given to** *suffix* variable

Example   **References**

UCLA   Stata

```
              long                              wide
        ┌─────────────┐                    ┌──────────────────┐
        │ i  j   stub │                    │ i  stub1  stub2  │
        ├─────────────┤                    ├──────────────────┤
        │ 1  1   4.1  │      reshape       │ 1   4.1    4.5   │
        │ 1  2   4.5  │   <──────────>     │ 2   3.3    3.0   │
        │ 2  1   3.3  │                    └──────────────────┘
        │ 2  2   3.0  │
        └─────────────┘
```

Reshape Wide   Convert data from long to wide

- **reshape** **wide** *stub***,** **i(** *id* **)** **j(** *suffix* **)** [options]

Options	Descriptions
wide	reshape data from **long** to **wide**
stub	variable to be converted
i(*id* **)**	unique id for records in **wide** format
j(*suffix* **)**	*stub* suffix **given to** *suffix* variable

Example

UCLA

References

Stata
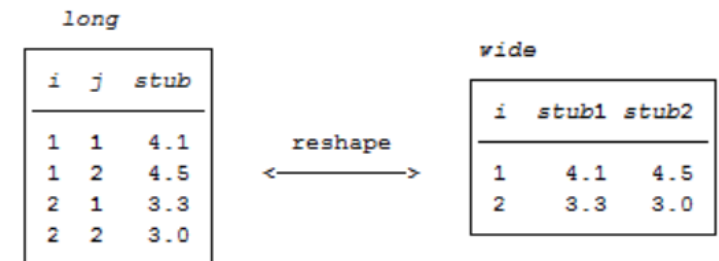
Append

- **append** using *filename* [*filename* …] [, options]

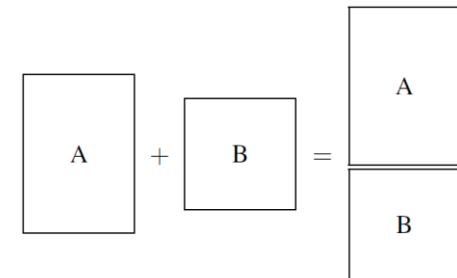| Options	Descriptions
generate(newvar)	newvar marks source of resulting observations
keep(varlist)	keep specified variables from appending dataset(s
force	append string to numeric & vice versa without error

Example

UCLA

References

Princeton   Stata

Merge

- **merge** 1:1 *varlist* using *filename* [, options]

master : using *filename*	**Merge** on specified key variables in *varlist*
1:1	One-to-One
1:m	One-to-Many
m:1	Many-to-One
m:m	Many-to-Many

Example

UCLA

References

Princeton   Stata

| A | + | B | = | A | B |
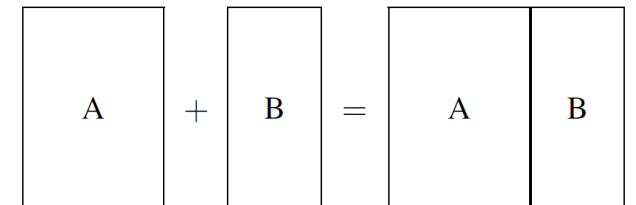
Collapse
Make dataset of summary statistics

- **collapse** [(*stat*)] *varlist* [, by(*varlist*)]

Options	Descriptions
varlist	Calc (*stat*) on *varlist*, default (*mean*)
stat	Mean, Median, Percentiles, SDev, Sums, Count, Min, Max, Range, First, Last values
by(*varlist*)	Calc stats over group
Pro Tip	Convert TS data to lower frequency

Example   **References**

UCLA   TimeSeries   Other   Stata

Excel Data Tools

2 Excel Tricks 2 Also Remember!

- VLOOKUP,   Link.
- PIVOT TABLES,   Link1, Link2.