Combining datasets

You have two datasets that you wish to combine. Below, we will draw a dataset as a box where, in the box, the variables go across and the observations go down.

See [D] `append` if you want to combine datasets vertically:

```
A + B = A
B
```

`append` adds observations to the existing variables. That is an oversimplification because `append` does not require that the datasets have the same variables. `append` is appropriate, for instance, when you have data on hospital patients and then receive data on more patients.

See [D] `merge` if you want to combine datasets horizontally:

```
A + B = A
B
```

`merge` adds variables to the existing observations. That is an oversimplification because `merge` does not require that the datasets have the same observations. `merge` is appropriate, for instance, when you have data on survey respondents and then receive data on part 2 of the questionnaire.

See [D] `joinby` when you want to combine datasets horizontally but form all pairwise combinations within group:

```
A × B = A
B
```

`joinby` is similar to `merge` but forms all combinations of the observations where it makes sense. `joinby` would be appropriate, for instance, where A contained data on parents and B contained data on their children. `joinby familyid` would form a dataset of each parent joined with each of his or her children.
Also see [D] cross for a less frequently used command that forms every pairwise combination of two datasets.

See Mitchell (2010, chap. 6) for more information on combining datasets in Stata.

References


