

## Notes on Panel Datasets

The pre-made panel datasets available for download have a common format which can be summarized as follows:

- All panel datasets contain a variable "**panel**" that allows easy identification of panel observations
- Some panel datasets contain a variable "**eligibility**" showing the status of each establishment surveyed in the previous round.
- All panel datasets contain the original panel ID variable that is matched across years. This variable differs from country to country but is always one of the first 5 variables in a dataset. (Examples: idstd\_2003, id2006 eec\_panel, codeentr, interview\_no, serial, etc)
- Wherever possible variables in the older rounds have been matched to variables in the newest round.
- Any variables that could not be matched are retained in their original form.
- Due to methodological changes and evolution of the survey instrument it is not possible to match all variables in the datasets.
- It is recommended that users thoroughly familiarize themselves with the questionnaires from each of the years contained in the dataset before proceeding with analysis.
- Most surveys conducted after 2006 use stratified sampling and contain weights based on this information. Prior surveys are likely to contain no information regarding weights.

The panel ID and year are already set in the data (xtset):

- panel variable: panelid (unbalanced)
- time variable: year

To understand more about the panel structure, please type in Stata: **tab panel year**

Panel: Firm interviewed in these years	Year of survey		Total
	2006	2010	
2010 only	0	320	320
2006 only	334	0	334
2006 and 2010	287	287	574
Total	621	607	1,228

In this example from Uruguay, the panel variable is not overlapped between categories. The panel label "2006 and 2010" indicates firms interviewed in both 2006 and 2010 only. The 2006 only and 2010 only firms in the other two labels are exclusively different.