The Stata Journal (2007)
7, Number 2, pp. 249–265

Speaking Stata: Identifying spells

Nicholas J. Cox
Durham University
Durham City, UK
n.j.cox@durham.ac.uk

Abstract. Spells in time series (and more generally in any kind of one-dimensional series) may be defined as sequences of observations that are homogeneous in some sense. For example, a categorical variable may remain in the same state, or values of a measured variable may satisfy the same true–false condition. Devices for working with spells in Stata include marking the start of each spell with indicator variables and tagging spells with integer codes. Panel data are easy to handle with the `by:` prefix. Some kinds of spell identification require two passes through the data, as when only spells of some minimum length are of interest or short gaps are tolerable within spells. Many questions concerning spells are easy to answer given careful use of `by:` and appropriate sort order, selection of just 1 observation from each panel or spell, and appreciation of the many functions written for `egen`. Gaps before, between, and after spells can also be important, and I suggest a convention for handling them.

Keywords: dm0029, spells, runs, time series, data management

