The Stata Journal (2004)
4, Number 2, pp. 221–222

Stata tip 8: Splitting time-span records with categorical time-varying covariates

Ben Jann, ETH Zürich, Switzerland
jann@soz.gess.ethz.ch

In survival analysis, time-varying covariates are often handled by the method of episode splitting. The `stsplit` command does this procedure very well, especially in the case of continuous time-varying variables such as age or time in study. Quite often, however, we are interested in evaluating the effect of a change in some kind of categorical status or the occurrence of some secondary event. For example, we might be interested in the effect of the birth of a child on the risk of divorce or the effect of having completed further education on the chances of upward occupational mobility.

In such situations, the creation of splits might appear to be more complicated, and `stsplit` does not seem to be of much help, at least judging from the rather complicated examples provided with [ST] **stset** (*Final example: Stanford heart transplant data*) and [ST] **stsplit** (*Example 3: Explanatory variables that change with time*). Fortunately, the procedure is simpler than it appears.

Consider the Stanford heart transplant data used in examples for [ST] **stset** and [ST] **stsplit**:

```
. use http://www.stata-press.com/data/r8/stanford, clear
(Heart transplant data)

. list id transplant wait stime died if id==44 | id==16
```

	id	transp~t	wait	stime	died
33.	44	0	0	40	1
34.	16	1	20	43	1

The goal here is to split the single time-span records into episodes before and after transplantation (e.g., to split case 16 at time 20). This can easily be achieved by splitting "at 0" "after `wait`", the time of transplantation. Note that, if no transplantation was carried out at all, `wait` should be recoded to a value larger than the observed maximum episode duration (maximum of `stime`) before the `stsplit` command is applied:

```
. replace wait = 10000 if wait == 0
(34 real changes made)

. stset stime, failure(died) id(id)
  (output omitted)

. stsplit posttran, after(wait) at(0)
(69 observations (episodes) created)

. replace posttran = posttran + 1
(172 real changes made)
```

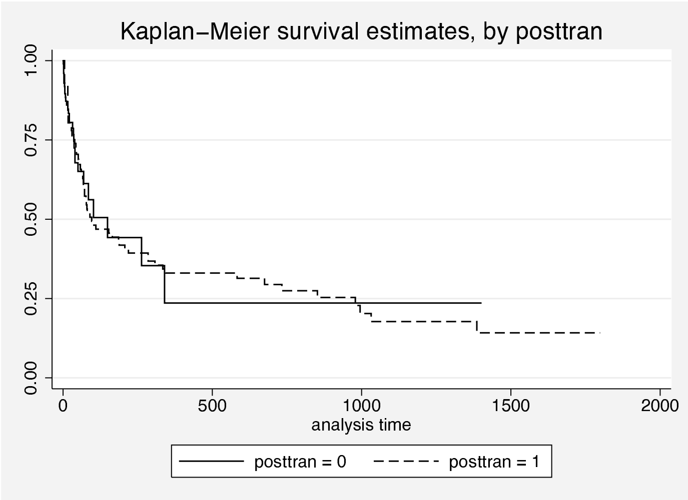
 st0066

```
. list id _t0 _t posttran if id == 44 | id == 16
```

	id	_t0	_t	posttran
23.	16	0	20	0
24.	16	20	43	1
70.	44	0	40	0

It is now possible to evaluate the effect of transplantation on survival time using `streg`, for example, or to plot survivor functions with time-dependent group membership:

```
. sts graph, by(posttran)
         failure _d:  died
   analysis time _t:  stime
                 id:  id
```



Note that the situation is even simpler if a Cox proportional hazards model is to be fitted. As explained in [ST] **stsplit**, the partial likelihood estimator takes only the times at which failures occur into account. Thus, in the context of Cox regression, the following code would do:

```
. use http://www.stata-press.com/data/r8/stanford, clear
. stset stime, failure(died) id(id)
. stsplit, at(failures)
. generate posttran = wait<_t & wait!=0
. stjoin
. stcox age posttran surgery year
```