SCOP: survival analysis

# Desired format

## convert snpashot data

**Snapshot data** = observations depict an instance (status of a given subject at a certain point in time) in time rather than an interval.

* Snpaspan convert it to survival-time data before proceeding

snapspan *id tvar evar*, generate(*time0*)

* + *idvar* records the subject id
  + *timevar* records the end time of the snapshot;
  + *varlist* are the “event” variables, meaning that they occur at the instant of timevar (=intstantaneous variable). Any variables in our data that are left unspecified are treated as enduring variables. varlist can also include retrospective variables that are to apply to the time span ending at the time of the current snapshot. For the enduring variables, we will obtain their values from the first observation of the pair; For the instantaneous variables, we will obtain the value from the second observation of the pair:

|  |
| --- |
| gen bankrupt = 1 if (year+1) == annee\_radiation |(year+2) == annee\_radiation | (year+3) == annee\_radiation  replace bankrupt = 0 if bankrupt == .  label variable bankrupt "1 if SCOP has known bankruptcy, 0 otherwise"    tostring year, replace  gen time1 = date(year, "Y")  format time1 %td    \*snapspan idvar time\_var instantaneous\_vars, generate(new\_begin\_date)  snapspan id time1 bankrupt salaries societaires\_salaries societaires capital\_social\_ou\_individuel K tota\_capital\_salaris\_associs dividend age lnK LS KLS CO COKLS SMALL MEDIUM LARGE, generate (time0) replace |

# Declare data as survival-time data

stset *timevar*, id(*idvar*) failure(*failvar*==*numlist*) origin() time0() exit() scale()

* + Main
    - Timevar: survival time (or a date) of the event/censoring time.
    - id(*idvar*): multiple-record ID variable
    - failure(*failvar==numlist*): failure event
  + Options
    - origin(): define when a subject becomes at risk
    - scale(#) rescale time value
    - enter() specify when subject first enters study, came under observation
    - exit() specify when the latest time at which the subject is at risk. The default is exit(failure), i.e. the subject is removed from the risk set after their event.
    - Time0(): time0 refers to the beginning time (recorded in time units) of a record. If the dataset does not contain the beginning time for each record, subsequent records are assumed to begin where previous records ended.

## Truncation and censor

In survival analysis, truncation occurs when subjects are observed only if their failure times fall within a certain observational period of a study. Censoring, on the other hand, occurs when subjects are observed for the whole duration of a study, but the exact times of their failures are not known; it is known only that their failures occurred within a certain time span.

### Truncation

For absorbing events such as death, truncation is defined as a period over which the subject was not observed but is, a posteriori, known not to have failed. **Truncation** causes the statistical difficulty that, had the subject failed, he or she would never have been observed.

**Left-truncatio**n usually arises because we encounter a subject who came at risk some time ago. The subject's subsequent survival can be analyzed, but we do not want to make too much of the fact that the subject survived until the point we encountered him. Left-truncation occurs when subjects come under observation only if their failure times exceed some time t1. It is only because they did not fail before t1 that we even knew about their existence. Left-truncation differs from left-censoring in that, in the censored case, we know that the subject failed before time t1, but we just do not know exactly when.

**Right-truncation** occurs when subjects come under observation only if their failure times do not exceed some time tr. Right-truncated data typically occur in registries. For example, a cancer registry includes only subjects who developed a cancer by a certain time, and thus survival data from this registry will be right-truncated.

**Interval-truncation (gaps):** a subject who disappears for a while but then reports back to the study, causing a gap in our follow-up. The statistical issue is that, had the subject died, he or she never could have reported back. The solution to this is to remove the person from the risk pool during the observational gap.

To determine that you have gaps, your data must provide starting and ending times for each record. Most datasets provide only ending times, making it impossible to know that you have gaps. You use time0() to specify the beginning times of records. time0() specifies a mechanical aspect of interpreting the records in the dataset, namely, the beginning of the period spanned by each record.

## Censor

Censoring is defined as when the failure event occurs and the subject is not under observation. An observation is censored when the exact time of failure is not known, and it is uncensored when the exact time of failure is known.

**Right-censoring**: the subject participates in the study for a time and, thereafter, is no longer observed; the time of failure is not known; it is merely known that the failure occurred after tr. If a patient survives until the end of a study, the patient’s time of death is right-censored. In common usage, censored without a modifier means right-censored.

An observation is **left-censored** when the exact time of failure is not known; it is merely known that the failure occurred before tl. Suppose that the event of interest is becoming employed. If a subject is already employed when first interviewed, his outcome is left-censored.

An observation is **interval-censored** when the time of failure is not known; it is merely known that the failure occurred after tl but before tr. Suppose that the event of interest is an onset of breast cancer. Patients are assessed periodically during their yearly checkups. The actual time of the onset of the disease, if present, is rarely known. Often, it is only known that the disease happened between the last and the current checkups. The time to the onset of breast cancer is the interval-censored.

|  |
| --- |
| //we have: right-censored, left truncation (🡪origin), interval truncation (🡪 time0())  stset time1, time0(time0) failure(bankrupt==1) id(id) origin(annee\_creation)  PROBLEM: time1 is no correct. For year == 2006, should be 31/12/2006 |

# Model

Quelles variables pour mesurer la participation ?

* LS : d’office, la participation à la gouvernance est LA variable pour définir une coop
* KLS :
* LS\*LS : pour la forme linéaire ou non

Comment prendre en compte la différence de taille ? du secteur ?

# Literature

Burdin 2014 : pourquoi ces variables la ? problème d’endogénéité ?

Se baser sur la litt pour définir si on prend « LS » ou si on prend « nombre d’employés et L »

Choix de la litt mobilisée : litt classique coopérative ? étendre avec la théorie sur les communs ?

Élargir la litt sur la participation plus large (ex : Allemagne ou les travailleurs ont plus participation, mécanismes de participation dans la gouvernance).

Extensions d’explication : autre forme de participation qui permet de reculer la faillite ??