Max Planck Odense Center on the Biodemography of Aging

University of Southern Denmark

Cure models in survival analysis

Virginia Zarulli - Adam Lenart

MaxO - SDU

Introduction

- **Aim:** Cure models aim to estimate the proportion of individuals who are cured (no longer at risk) from experiencing the event.
- **Assumption:** (i) a fraction of individuals will never experience the event, (ii) excess mortality of individuals with a disease compared to a background population after some time might vanish.

Cure models

$$S(t) = S^*(t)R(t),$$

where $S(t)$ is the all-cause survival function, $S^*(t)$ is the background (expected) survival function and $R(t)$ the relative survival function.

$$h(t) = h^*(t) + \lambda(t),$$

where $h(t)$ is the all-cause hazard, $h^*(t)$ the background risk and $\lambda(t)$ the excess mortality risk associated with the variable of interest.

Two usual fields of usage:

- Special data sets: When you can assume that background risk of failure is 0 (for example estimating cure from cancer for children), i.e. $h^*(t) = 0$
- Population-based data sets: When we cannot be certain about the cause of death, we only know that a person is infected with a disease and whether is alive or not at the end of follow up period. In this case, we have to include background risk for all-cause mortality from an external source (e.g., official information, Human Mortality Database etc.).

Types of cure models

Two types of cure models:

- **Mixture cure models:** a fraction (π) of the patients are cured and are not at risk of experiencing the event, while the others remain uncured ($1 - \pi$)
- **Non-mixture cure models:** cumulative hazard converges to an asymptote, hence some of the individuals will not experience the event.

Mixture models

$$S(t) = S^*(t)\left[\pi + (1-\pi)S_u(t)\right],$$

where $S_u(t)$ si the survival function for the uncured. For $S^*(t) = 1$, this model is called the standard cure model. On the hazard scale:

$$h(t) = h^*(t) + \frac{(1-\pi)f_u(t)}{\pi + (1-\pi)S_u(t)},$$

where $f_u(t)$ is the probability density function of the uncured.

Non-mixture model

$$S(t) = S^*(t)\pi^{1-S_z(t)},$$

where S_z is usually a parametric distribution that changes the probability being cured or uncured at time t.

$$h(t) = h^* t - \log(\pi) f_z(t),$$

where $f_z(t)$ is the probability density function for $S_z(t)$.
Split-time: we can also specify, that cure doesn't start from time 0, but only at a later time point and before that a parametric distribution is fitted (similar to `streg`).

Cure models in Stata

We have to install the `strsmix` and `strsnmix` commands:

1. `findit lambert cure`
2. Click on the link to the Stata journal and install.

- `strsmix` *`covariates`*`, distribution(...) link(...) bhazard(...)`
- `strsnmix` *`covariates`*`, distribution(...) link(...) bhazard(...) split(...) earlydist(...)`

Cure models in Stata

- `distribution:` weibull, lognormal, gamma, weibweib, weibexp
- `link:` identity, logistic, loglog
- `bhazard:` a variable with the background risk (can be a vector of 0's)
- `split:` the splitting time point available for non-mixture models
- `earlydist:` weibull or exponential, the parametric regression that is fit before the splitting time.

Exercise

Estimate the cured population of colon cancer patients, based on 10-year survival from the Finnish cancer registry.

1. Install strsmix and strsnmix commands.
2. Read `colon.dta` in.
3. Look at stage, agegrp and subsite as interesting-looking covariates. (`tab ..., sum(...)`
4. Declare survival model: survival time is in months, scale it by 12 to get years using option `scale()` and as we are interested only in 10-year survival, right censor everybody after 120 months using option `exit(time 120)`. Declare that status 1 and 2 are both counted as events.
5. Generate dummy variables for cancer stages and age groups.
6. Look at background mortality rate.

Exercise contd.

7. Start with a mixture model (weibull with identity link), include covariates (remove those that are not significant)
8. Check if it looks better with a non-mixture model.
9. Plot predictions for the hazard of distant and localized stages for old elderly patients
   - e.g. `predict disthaz if o_eld==1 & distant==1, hazard`
   - e.g. `predict disthaz if o_eld==1 & distant==1, hazard uncured`
10. Create interaction variables of old elderly and distant and localized cancer stages
11. Run model again.

Exercise contd.

12. Predict median survival for the uncured.
    - `predict med, centile centval(.5)`
13. Maybe the people who fail at a low survival time are different from the others, check it with a split time design at time 0.5. You can use exponential hazard for this short time period