Description Remarks and examples References Also see

Description

Cure models, or split-population models, are used to model survival data where a fraction of the population will never experience a failure. Mixture cure models represent the population as a combination of two types of individuals: a short-term survivor (noncured) group and a long-term survivor (cured) group. These models allow us to detect covariates associated with class membership (being cured or not) and to investigate the impact of covariates on the hazard for the noncured group as well.

In this example, we demonstrate how to fit a cure model as a two-component FMM with one component being a parametric survival model and one component being a point mass density that represents the cured group.

Remarks and examples

Implantation of intraocular lenses is a common surgery used to treat cataracts. One possible complication after this procedure is calcification of the lenses. Some patients will experience calcification during the follow-up period and some will not. Just because patients have not experienced calcification during the follow-up period does not mean that they truly are cured. It is still possible that they might experience calcification after the follow-up period ends. Thus, the cured group must be considered right-censored, with some individuals not observed to have calcification possibly belonging to this group.

In the language of FMM, we have two latent groups: a cured group and a noncured group. We know that patients who experience calcification are members of the noncured group. We do not know which group that patients who remain healthy belong to. That is, some of the patients we observe as healthy are truly cured, whereas others are members of the noncured group who are right-censored because they happened to not experience calcification during the study.

With a mixture cure model, we can predict the probability that an individual who did not experience calcification during the study is noncured. Let π be the probability of being in the noncured group, and let $S_1(t)$ be the survivor function for the noncured group. For the noncured group, the time to failure is modeled with a parametric distribution accounting for right-censoring, such as exponential or Weibull. If we let S(t) be the probability of not failing before time t for an individual in the population, our model is

$$S(t) = (1 - \pi) + \pi S_1(t)$$

To illustrate the model, we use the artificial dataset, lenses.dta, with some of the characteristics of the calcification study described in Ma (2009). About 46% of the patients did not have postsurgery calcification of lenses during the follow-up period. We will predict how many of those are likely to have calcification after the follow-up period.

In our model, $S_1(t)$ is a Weibull using a proportional hazards parameterization. The covariates of interest are patient's sex (sex), patient's age at implantation divided by 10 (age10), and incision length (inclength).

The variable fail in the dataset contains an indicator for failure (occurrence of calcification). When fail = 1, we know that an individual belongs to the noncured group. When fail = 0, the individual is observed as healthy, but we cannot say they are a member of the cured group.

We first use stset to declare the data to be survival-time data. We specify t as the variable that contains analysis time and fail as the variable that indicates failure; see [ST] stset for details.

```
. use https://www.stata-press.com/data/r19/lenses
(Simulated calcification data)
. stset t, failure(fail)
Survival-time data settings
        Failure event: fail!=0 & fail<.
Observed time interval: (0, t]
    Exit on or before: failure
        770 total observations
         0 exclusions
       770 observations remaining, representing
        415 failures in single-record/single-failure data
20,133.467 total analysis time at risk and under observation
                                                At risk from t =
                                                                         0
                                                                         0
                                     Earliest observed entry t =
                                          Last observed exit t =
                                                                        36
```

To model time-to-calcification in the noncured group, we fit a Weibull model for right-censored data where the dependent variable is the time variable. This includes the patients observed as noncured and those who appear healthy. To model the probability of being cured, we use a point mass density at fail = 0 because this indicates that calcification was not observed. See [FMM] **fmm: pointmass** for details about the point mass distribution.

. imm: (point	mass Iall) (su	teg increng	un r.ser	agei0,	distinution(we	IDUII))
(iteration log of	mitted)					
Finite mixture Log likelihood	e model d = -1980.1495				Number of (obs = 770
	Coefficient	Std. err.	Z	P> z	[95% conf.	interval]
1.Class	(base outcome)					
2.Class cons	1.01863	.2703434	3.77	0.000	.4887664	1.548493
Class: 2 Response: _t Model: stre	eg, dist(weibull)			No. of failures = 415 Time at risk = 20,133.47		
	eg, dist(weibu			111	ne at risk = .	20,133.47
	Coefficient	Std. err.	Z	P> z	[95% conf. :	20,133.47
_t inclength	Coefficient 5922698	Std. err.	z -2.60	P> z	[95% conf. :	interval]
_t inclength sex male age10 _cons	Coefficient 5922698 .3314051 .1600672 -4.939691	Std. err. .2273662 .1259957 .032798 .940024	z -2.60 2.63 4.88 -5.25	P> z 0.009 0.009 0.000 0.000	[95% conf. : -1.037899 - .0844581 .0957843 -6.782104 -	1466402 .5783522 .2243502 -3.097278

The first table in the output shows the estimated coefficient on the logit scale for the class 2 (noncured group) probability. This probability is 0.73 [invlogit(1.019) \approx 0.73], which implies that the probability of being in the cured group is 0.27.

The second table presents the results for the Weibull regression model for the noncured group. We see that longer incisions decrease the hazard of calcification, while being male and being older increase the hazard of calcification.

We may want to know the probability that patients who have not experienced calcification will do so in the future. We can predict the posterior probability of being in class 2. We list the first 10 patients for the cured group.

. predict pprob2, classposterior class(2.Class)

```
. sort fail, stable
```

. list fail pprob2 in 1/10

	fail	pprob2
1. 2. 3. 4. 5.	0 0 0 0	.2569577 .4447927 .3233174 .4677424 .4549083
6. 7. 8. 9. 10.	0 0 0 0 0	.4183038 .3161573 .4540032 .2782425 .5745969

We see that the posterior probability of having calcification in the future is over 50% for the last patient.

We generate an indicator variable prfail that takes on value 1 if the posterior probability of calcification is greater than 50% and zero otherwise. We construct a classification table where we tabulate our variable against the indicator of failure fail.

```
. generate prfail = pprob2 > .5
. tabulate prfail fail
                failed=1, didn't
                     fail=0
                                           Total
   prfail
                      Λ
                                  1
         0
                    257
                                  0
                                              257
         1
                     98
                                415
                                              513
     Total
                    355
                                415
                                              770
```

Out of 355 individuals who did not experience calcification during the study, we estimate that 98 are more likely than not to have calcification in the future.

References

Lambert, P. C. 2007. Modeling of the cure fraction in survival studies. *Stata Journal* 7: 351–375. Ma, S. 2009. Cure model with current status data. *Statistica Sinica* 19: 233–249.

Also see

[FMM] **fmm** — Finite mixture models using the fmm prefix

[FMM] fmm postestimation — Postestimation tools for fmm

Stata, Stata Press, and Mata are registered trademarks of StataCorp LLC. Stata and Stata Press are registered trademarks with the World Intellectual Property Organization of the United Nations. StataNow and NetCourseNow are trademarks of StataCorp LLC. Other brand and product names are registered trademarks or trademarks of their respective companies. Copyright © 1985–2025 StataCorp LLC, College Station, TX, USA. All rights reserved.



For suggested citations, see the FAQ on citing Stata documentation.