

区間打ち切り Cox モデル

キーワード

- セミパラメトリック生存モデリング
- 左打ち切り、右打ち切り、間隔打ち切り
- 現在の状態と一般的な間隔打ち切りデータ
- 層別推定
- ベースラインハザードの 2 つの推定量
- 標準誤差の 2 つの推定量
- 生存グラフ、累積ハザードグラフ、ハザード関数グラフ
- 残差の診断
- 比例ハザード仮定をグラフで確認

概要

- 癌が再発した時間やコロナウイルス感染が発生した時間は、正確には分かりません。分かっているのは、検査で陽性反応の出た時間とその前の検査で陰性であった時間の間のどこかで発病したということです。これが間隔打ち切りと左打ち切りの time-to-event データの例です。機械部品の故障時間や故障期間にも利用できます。
- 区間打ち切りのイベント-時間データは、医学、疫学、経済、金融、社会学など、多くの分野で扱われます。区間打ち切りを無視すると、推定に偏りが生じます。
- セミパラメトリック Cox 比例ハザード回帰モデルは、打ち切りなしまたは右打ち切りのイベント-時間データを解析するためによく使用されます。
- 右打ち切りデータと同様に、Cox モデルは区間打ち切りデータにも使用されます。セミパラメトリック推定を使用し、ベースラインハザード関数にパラメトリック仮定を必要としないからです。また、イベントの発生率が低い場合、指数回帰パラメーターで相対リスクの対数を近似できます。
- 区間打ち切りイベント-時間データのセミパラメトリック推定は、イベント時間が正確に観測されていないので、難しいことに挑戦することになります。したがって、このようなデータの“セミパラメトリック”モデリングは、ベースラインハザード関数をスプライン法や区分指数モデルで表す方法がよく使われてきました。
- **stintcox** コマンドが実装され、近年の方法論である純正の区間打ち切りイベント-時間データのセミパラメトリックモデリングが利用できるようになりました。

推定

- Zeng, Mao, and Lin (2016)は、タイの薬物使用者のコホート研究において、HIVに感染するまでの時間を調べました。例として、この論文のデータを使用します。
- データセットには、1124の被験者データが含まれています。被験者は、最初はHVI検査結果は陰性でした。その後、約4か月おきの血液検査を行い、被験者のHIV-1血清反応陽性評価しました。
- 定期的に検査を実施したため、HIV感染が起こった正確な時間は観測できませんが、前回の検査時(変数 `ltime`)と最新の検査時(変数 `rtime`)の間のどこかの時点で感染が起こったということが分かります。
- ここで、HIVに感染するまでの時間に影響する因子を特定したいと思います。調査する因子は、被験者の観測開始時の年齢(`age`)、性別(`male`)、注射針の共用の経歴(`needle`)、薬物使用の経歴(`inject`)、観測開始時に収監されていたかどうか(`jail`)です。
- 次のコマンドを入力してデータを入手し、内容を確認します。

```
. use https://www.stata-press.com/data/r17/idu
. describe
```

```
Contains data from https://www.stata-press.com/data/r17/idu.dta
Observations:      1,124      Modified Bangkok IDU Preparatory Study
Variables:         8         15 Dec 2020 13:34
                        (_dta has notes)
```

Variable name	Storage type	Display format	Value label	Variable label
<code>age</code>	byte	%8.0g		Age (in years)
<code>male</code>	byte	%8.0g	yesno	Male
<code>needle</code>	byte	%8.0g	yesno	Shared needles
<code>jail</code>	byte	%8.0g	yesno	Imprisoned
<code>inject</code>	byte	%8.0g	yesno	Injected drugs before recruitment
<code>ltime</code>	double	%10.0g		Last time seronegative for HIV-1
<code>rtime</code>	double	%10.0g		First time seropositive for HIV-1
<code>age_mean</code>	double	%10.0g		Centered age (in years)

- 上記の因子による HIV に感染した時間について、Cox 比例ハザードモデルをフィットさせます。ベースラインハザード関数をより妥当なものにするために、年齢(平均値からの差)変数 `age_mean` を使います。

```
. stintcox age_mean i.male i.needle i.inject i.jail, interval(ltime rtime)
```

note: using adaptive step size to compute derivatives.

Performing EM optimization (showing every 100 iterations):

```
Iteration 0:   log likelihood = -1086.2564
Iteration 100: log likelihood = -597.65634
Iteration 200: log likelihood = -597.57555
Iteration 295: log likelihood = -597.56443
```

Computing standard errors: done

```
Interval-censored Cox regression      Number of obs   = 1,124
Baseline hazard: Reduced intervals    Uncensored      =    0
                                       Left-censored   =   41
                                       Right-censored  =  991
                                       Interval-cens.  =   92

                                       Wald chi2(5)    =  17.10
Log likelihood = -597.56443           Prob > chi2     =  0.0043
```

	Haz. ratio	OPG std. err.	z	P> z	[95% conf. interval]	
age_mean	.9684341	.0126552	-2.45	0.014	.9439452	.9935582
male						
Yes	.6846949	.1855907	-1.40	0.162	.4025073	1.164717
needle						
Yes	1.275912	.2279038	1.36	0.173	.8990401	1.810768
inject						
Yes	1.250154	.2414221	1.16	0.248	.8562184	1.825334
jail						
Yes	1.567244	.3473972	2.03	0.043	1.014982	2.419998

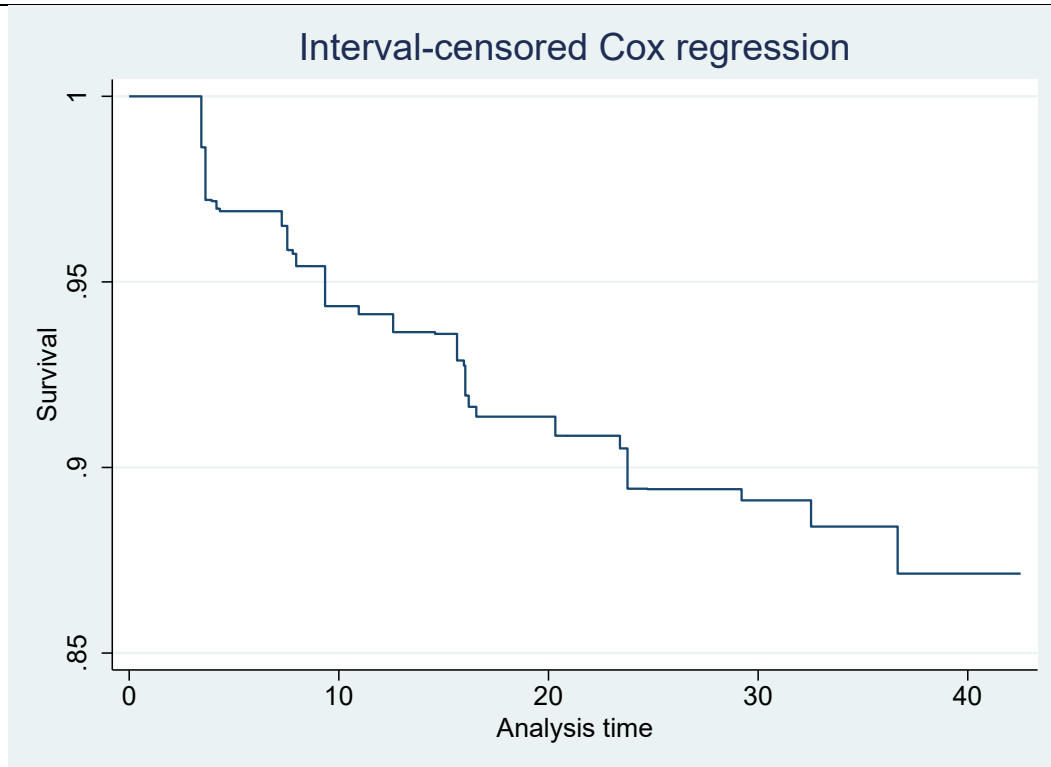
Note: Standard-error estimates may be **more variable** for small datasets and datasets with low proportions of interval-censored observations.

- 調査参加時の年齢は、HIV 感染リスクが低いことと関連していると分かります。また、観測開始時に収監されていることと、HIV 感染リスクが高くなることに関連があることが分かります。他の因子については、統計的に有意な関連性がみられません。

生存関数グラフの描画

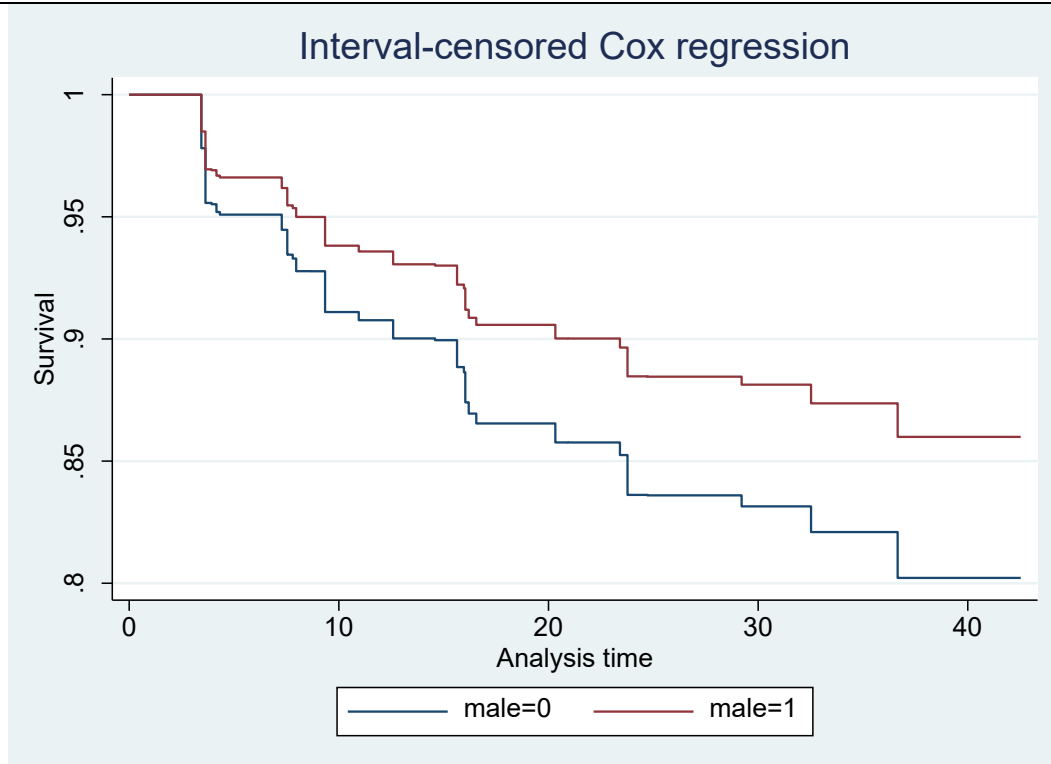
- `stcurve` コマンドとそのオプションを使って、全ての共変量を 0 に設定したベースライン生存曲線を描画します。

```
. stcurve, survival at((zero) _all)
```



- 下記のように、共変量を任意の値に設定して生存関数を描くことができます。

```
. stcurve, survival at(male=(0 1))
```



- 女性のベースラインハザード関数が男性のものとは異なることを仮定した場合は、`strata()` オプションを使用して層別 Cox 比例ハザードモデルをフィットさせることができます。

```
. stintcox age_mean i.needle i.inject i.jail, interval(ltime rtime)
strata(male)
```

note: using adaptive step size to compute derivatives.

Performing EM optimization (showing every 100 iterations):

```
Iteration 0: log likelihood = -1087.0536
Iteration 100: log likelihood = -585.59848
Iteration 200: log likelihood = -585.53143
Iteration 282: log likelihood = -585.5222
```

Computing standard errors: done

Stratified interval-censored Cox regression
Baseline hazard: Reduced intervals

```
Strata variable: male                                Number of obs   = 1,124
                                                    Uncensored     = 0
                                                    Left-censored  = 41
                                                    Right-censored = 991
                                                    Interval-cens. = 92

                                                    Wald chi2(4)   = 14.84
Log likelihood = -585.5222                        Prob > chi2    = 0.0051
```

	Haz. ratio	OPG std. err.	z	P> z	[95% conf. interval]	
age_mean	.9682508	.0126326	-2.47	0.013	.9438052	.9933295
needle						
Yes	1.276222	.2270302	1.37	0.170	.9005422	1.808625
inject						
Yes	1.245357	.2393768	1.14	0.254	.8544367	1.815131
jail						
Yes	1.57314	.3490687	2.04	0.041	1.018337	2.430205

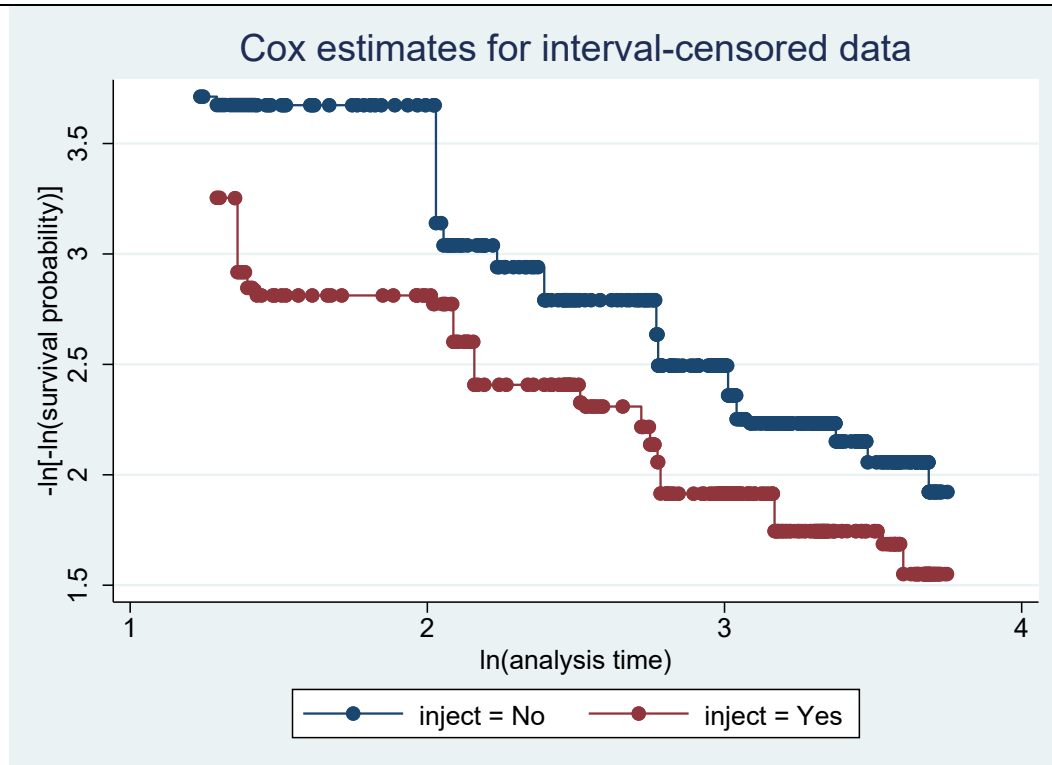
Note: Standard-error estimates may be [more variable](#) for small datasets and datasets with low proportions of interval-censored observations.

- ここでの結果は、層別でない場合の結果と同様です。

比例ハザード仮定の確認

- 「ハザード比は時間経過に対して一定である」という Cox 比例ハザードモデルの推定の基礎となる仮定の妥当性を評価することが重要です。Stata17 には、比例ハザード仮定を評価するための 2 種類のグラフ機能が搭載されています。
- Cox モデルが単一カテゴリの共変量の場合は、`stintphplot` コマンドを使います。このコマンドは、各カテゴリについて、X 軸：In（解析時間）に対して Y 軸： $-\ln\{-\ln(\text{生存率})\}$ の曲線を描きます。プロットの変化がカテゴリ間で平行の場合は、比例ハザード仮定に反していません。
- `stintcoxnp` コマンドを使用すると、ノンパラメトリック最尤推定の生存曲線対 Cox 予測生存曲線をカテゴリごとにプロットすることができます。
- 上記の例のように Cox モデルに複数の共変量が含まれている場合、比例ハザード仮定の検定に適しているのは `stintphplot` コマンドのみです。この場合、`adjustfor()` オプションを使用してください。
- 例えば、変数 `inject` の比例ハザード仮定を検定する場合は、下記のように他のすべての共変量を `adjustfor()` オプションの中に入れます。

```
. stintphplot, interval(ltime rtime) by(inject) adjustfor(age_mean i.male i.needle i.jail)
```



- `adjustfor()` オプションに他のすべての共変量を入れることで分割された Cox モデルは、`inject` の各レベルにフィットしています。そして、2つのプロットはほぼ平行で

す。これは、カテゴリ変数 **inject** についての比例ハザード仮定が反していないことを意味します。

引用文献

Zeng, D., L. Mao, and D. Lin. 2016. Maximum likelihood estimation for semiparametric transformation models with interval-censored data. *Biometrika* 103: 253–271.