

医学統計勉強会

東北大学病院循環器内科・東北大学臨床研究推進センター 共催

東北大学大学院医学系研究科EBM開発学寄附講座

宮田 敏

生存時間解析 生存曲線, Cox比例ハザードモデル

生存時間解析 (survival time analysis) では, 基準となるある時点から, 目的となるイベントの発生までの時間を解析する. 例えば, ある疾患の登録研究において, 登録時から疾患発症, 死亡, 入院などのイベント発生までの時間が解析対象となる.

生存時間解析では, 対象となる事象を**イベント (event)**, **エンドポイント (end point)**, **結果 (outcome)**, 解析対象の時間を **生存時間 (survival time, failure time)** などと呼ぶ. 生存時間を説明する変数は, **説明変数**, **独立変数**, **共変量 (covariate)**, **危険因子 (risk factor)**, **予後因子 (prognostic factor)** などと呼ばれる.

打ち切り (censor(ing))

生存時間解析では、イベントの発生が観察出来ず、正確な生存時間が不明となる場合がある。

- 観察期間終了時までイベントが起こらなかった。
- 同意撤回，通院中止などによる打ち切り(withdraw)
- 行方不明，追跡不能 (lost to follow-up)

このような状況を**打ち切り (censor)**が生じたと言う。生存時間データは、**生存時間 (観察期間)**と**打ち切りの有無**の二つの情報のペアとして記録される。

打ち切りが生じた場合，生存時間は不明となるが「観察期間中にイベントが起こらなかった」という情報は手に入る。

生存時間解析の基本概念

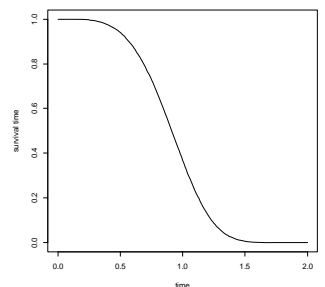
確率変数 T を，あるイベントが起こるまでの時間を表すとする。

$f(t)$: 確率密度関数, $F(t) = P(T \leq t)$: 分布関数

定義:

生存関数: $S(t) = P(T > t) = 1 - F(t) = \int_t^{\infty} f(x)dx,$

$$f(t) = -\frac{d}{dt} S(t).$$



生存関数は、 $t=0$ のとき $S(0)=1.0$ となり、 $t=\infty$ のとき0に収束する単調減少 (非増加) 関数。

生存時間解析の基本概念（続き）

定義: **ハザード生存関数**:

$$h(t) = \lim_{\Delta t \rightarrow 0} \frac{P(t \leq T < t + \Delta t | T \geq t)}{\Delta t} = \frac{f(t)}{S(t)} = -\frac{d}{dt} \log[S(t)].$$

$$H(t) = \int_0^t h(x) dx = -\log[S(t)].$$

ハザード関数は、t時点までイベントが起こらなかつたとき、続くt時点以降でイベントが起こる瞬間的な確率。ハザード関数は必ず非負の関数。

Tに対して、 $f(t), F(t), S(t), h(t), H(t)$ は一意的に定まる。

生存時間解析の基本概念（続き）

定義: **累積ハザード生存関数**:

$$H(t) = \int_0^t h(x) dx = -\log[S(t)],$$

$$S(t) = e^{-H(t)} = \exp\left[-\int_0^t h(x) dx\right] = \exp[-H(t)].$$

ここで、 $H(t) = \int_0^t h(x) dx$ は累積ハザード。確率変数Tに対しては確率密度関数 $f(t)$ 、累積分布関数 $F(t)$ 、生存関数 $S(t)$ 、ハザード関数 $h(t)$ 、累積ハザード関数 $H(t)$ が一意的に定まります。

生存曲線

例として、以下の生存時間データを考える。

2, 4, 4, 5, 7, 8, 10

このとき、生存時間曲線は以下のように作られる。

1. 観察開始時0日から2日目まで、イベントはなし

$$S(t) = 1.0, 0 \leq t \leq 2$$

2. 2日目で1件イベントが発生（死亡）したので、2日目が過ぎた瞬間を2+0とすると、

$$S(2+0) = 6/7$$

3. 2+0日目から4日目まではイベントはなし

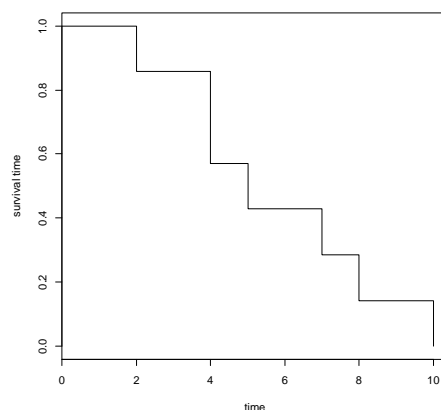
$$S(t) = 6/7, 2 < t \leq 4$$

生存曲線（続き）

4. 4日目には2件イベントが発生。このように同じ時点で複数のイベントが起こることをタイ (tie) という。4日目が過ぎた瞬間の4+0では、

$$S(4+0) = 4/7$$

5. これ以降も同様。



ここで4日目の状況を考える。4日目まで6件生存中で、4日目にタイのイベント2件が起こり、生存率は4/7となった。

$$s(4+0) = \frac{4}{7} = \frac{4}{6} \times \frac{6}{7}$$

上の式の最右辺は、

(4日目に生存中の個体のうち、4日目を過ぎた瞬間生存している個体の割合) × (4日目までの生存率)

と解釈出来る。打ち切りがある場合も、実はこの定義を踏襲する。

打ち切りのある場合を考えるため、元のデータの7日目と10日目が「打ち切り」であるとする。

2, 4, 4, 5, 7+, 8, 10+

1. 5日目までに4件のイベント発生。5日目までの生存率=3/7.

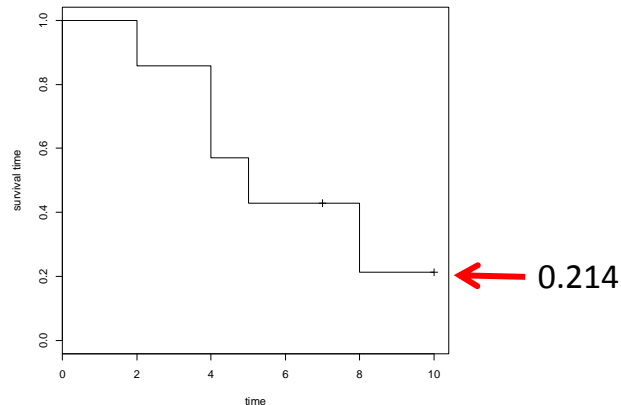
2. 7日目に最初の打ち切り。打ち切り例の本当の生存時間は「7以上」ということ以外は不明。

3. 5日目が過ぎた瞬間生存しているのは3件だが、打ち切りがあったため、次のイベントのある8日目に生存中の個体は2件に減る。

8日目の生存率は,

(8日目に生存中の個体のうち, 8日目を過ぎた瞬間生存している個体の割合) × (8日目までの生存率)
= $(1/2) \times (3/7)$
= $3/14 = 1.5/7 = 0.214$

なお, イベント発生時の個体の数を **number at risk** と呼ぶ。



Kaplan-Meier (product-limit) 推定量

生存時間: $0 = t_0 < t_1 < \dots < t_m$

イベント数: $d_j \geq 1, j = 1, \dots, m$

打ち切り数: $w_j \geq 0,$

Number at risk: $n_j, n_{j+1} = n_j - (d_j + w_j)$

Kaplan-Meier (product-limit) 推定量

$$S(t) = \frac{n_j - d_j}{n_j} S(t_{j-1} + 0) = \frac{n_j - d_j}{n_j} \frac{n_{j-1} - d_{j-1}}{n_{j-1}} S(t_{j-2} + 0) = \dots = \prod_{1 \leq i \leq j} \left(\frac{n_i - d_i}{n_i} \right)$$

ただし, 第2項のjは $t_j < t$ となる最大のj.

2群の生存関数の比較：Log-rank検定

$$H_0 : S_1(t) = S_2(t), \forall t$$

$$H_1 : S_1(t) \neq S_2(t)$$

帰無仮説が正しいとき，2群を併せたデータのイベントごとに以下の2×2分割表を考える。

	死亡数	生存数	合計
治療群	d_j		n_j
対照群			$N_j - n_j$
合計	D_j	$N_j - D_j$	N_j

検定統計量

$$Z^2 = \frac{(D - E)^2}{V} \sim \chi(1) \text{ under } H_0, \text{ where } D = \sum d_j, E = \sum E_j, V = \sum V_j.$$

2013/10/17

東北大学 医学統計勉強会

13

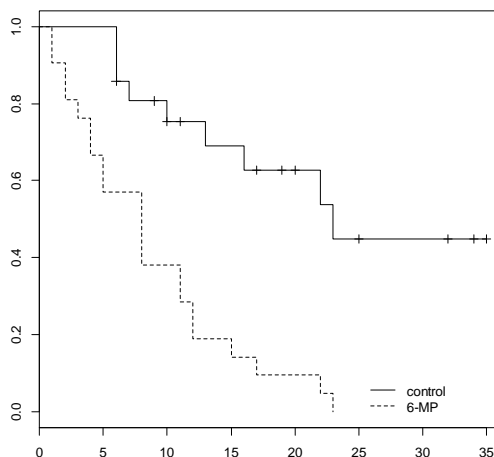
例：白血病患者に対する寛解期間の臨床比較試験

time: resimen time in weeks

cens: censoring, 0/1

treat: treatment, control or 6-MP (6-mercaptopurine)

pair	time	cens	treat
1	1	1	control
1	10	1	6-MP
2	22	1	control
2	7	1	6-MP
3	3	1	control
3	32	0	6-MP



Log-ran test p-value = 4.17×10^{-5}

Gehan, E.A. (1965) A generalized Wilcoxon test for comparing arbitrarily single-censored samples. *Biometrika* **52**, 203–233.

2013/10/17

東北大学 医学統計勉強会

14

Log-rank検定の特徴と問題点

Log-rank検定は時間に依存しない: Kaplan-Meier推定量による生存曲線の印象と、log-rank検定の結果が一致しないことがある。

Log-rank検定は単変量解析である: log-rank検定は、単一の因子によって群を場合分けし、群間で生存関数に有意差がないかを検定。生存時間に影響する他の因子がある場合、log-rank検定は不適切。
⇒ 層別log-rank検定, Cox比例ハザードモデル。

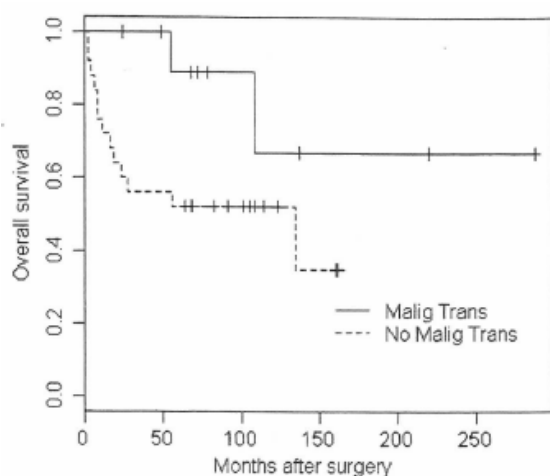
Log-rank検定以外に、**一般化Wilcoxon検定**がある。

2013/10/17

東北大学 医学統計勉強会

15

例: **一般化Wilcoxon検定** Malignant transformation of breast fibroadenoma to malignant phyllodes tumor



一般化Wilcoxon検定は、log-rank検定と同じ状況で用いられるが、より早い時点での生存曲線の差に対して検出力が高い

log-rank test $p = 0.0551$,
Peto and Peto's test $p = 0.0409$

Abe, M. et al (2011) Malignant transformation of breast fibroadenoma to malignant phyllodes tumor: long-term outcome of 36 malignant phyllodes tumors, Breast Cancer (2011) 18:268-272

2013/10/17

東北大学 医学統計勉強会

16

層別Log-rank検定

Log-rank検定で2群を比較する際、群を分ける要因以外に、**アウトカムに影響を与える因子**が存在する場合がある。⇒ log-rank検定は**不適切**。

危険因子によってデータがいくつかの層 (strata = subgroup) に分けられるとき、層ごとに分割表を集計し、後で全体を合成する ⇒ **層別log-rank検定**

- 層の分割が、恣意的になりがち。
- 多数の層に分割すると、層ごとのサンプル数が少なくなる。⇒ **Cox比例ハザードモデル**

Cox比例ハザードモデル

Cox比例ハザードモデル (Cox proportional hazard model) は、多変量の生存時間解析モデル。

$$h(t) = \lim_{\Delta t \rightarrow 0} \frac{P(t \leq T < t + \Delta t | T \geq t)}{\Delta t} = \frac{f(t)}{S(t)} = -\frac{d}{dt} \log[S(t)]$$

Cox比例ハザードモデル：

$$h(t | x) = h_0(t) \exp\{\beta_0 + \beta_1 x_1 + \dots + \beta_k x_k\}$$

$h_0(t)$: **ベースラインハザード** (baseline hazard)

$\exp\{\beta_0 + \beta_1 x_1 + \dots + \beta_k x_k\}$: **相対危険度関数** (relative risk function)

比例ハザード性: Cox比例ハザードモデルでは、時間に依存する部分 (baseline hazard) と、共変量に依存する部分 (relative risk) が分割されている。このため、異なるハザード関数の比が比例関係にあり、**ハザード比は共変量のみ**に依存し、**時間に依存しない**。

$$h(t | x') = \exp\{\beta'(x' - x)\}h(t | x)$$

対数線形性: Cox比例ハザードモデルは、相対危険度関数を通じて、共変量にはその一次式のみに依存する。 x_1 のみが1単位変化したときのハザード比は、

$$\frac{h(t | x')}{h(t | x)} = \frac{h_0(t)\exp\{\beta_0 + \beta_1(x_1 + 1) + \dots + \beta_k x'_k\}}{h_0(t)\exp\{\beta_0 + \beta_1 x_1 + \dots + \beta_k x_k\}} = e^{\beta_1}$$

Cox比例ハザードモデルの推定

Cox比例ハザードモデルは、partial likelihood method (部分尤度法)によって推定される。最尤推定量の性質に従い、信頼区間の構成、仮説検定が行われる。

	coef	exp(coef)	se(coef)	z	Pr(> z)	
treatcontrol	1.5721	4.8169	0.4124	3.812	0.000138	***
	exp(coef)	exp(-coef)	lower 0.95	upper 0.95		
treatcontrol	4.817	0.2076	2.147	10.81		
Likelihood ratio test= 16.35 on 1 df, p=5.261e-05						
Wald test = 14.53 on 1 df, p=0.0001378						
Score (logrank) test = 17.25 on 1 df, p=3.283e-05						

Cox比例ハザードモデル:推定と検定のチェックポイント

1. パラメターの推定値 $\beta_1 = 1.5721$
2. パラメターの有意性検定のp値 $p = 0.000138$
3. ハザード比 $e^{\beta_1} = 4.8169$
4. ハザード比の信頼区間 (2.147, 10.81)
5. モデルの適合度検定 (以下の三通りあるが、気にしない)

Likelihood ratio test= 16.35 on 1 df, p=5.261e-05
Wald test = 14.53 on 1 df, p=0.0001378
Score (logrank) test = 17.25 on 1 df, p=3.283e-05

Coxモデルの比例ハザード性の検証

Cox比例ハザードモデルは、ハザード比が共変量のみ
に依存し特定の時点に依存しない、という**比例ハザード性の仮定**を大前提にしている。

Coxモデルにおいて、もし比例ハザード性が満たされて
いれば以下の関係が成り立つことが知られている。

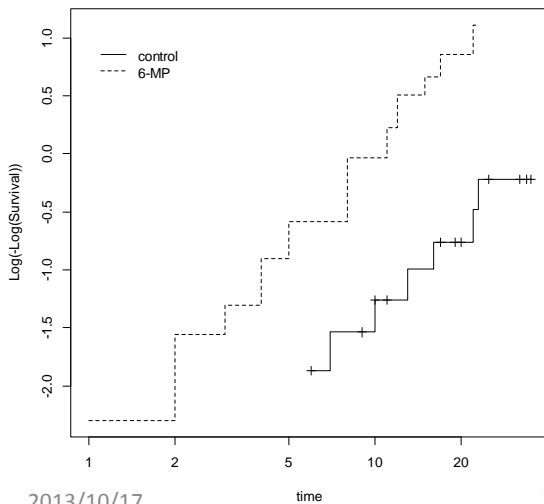
$$\log\{-\log S(t)\} = \exp\{-\log S_0(t)\} + \{\beta_0 + \beta_1 x_1 + \dots + \beta_k x_k\}$$

これは、任意の時点tに対しての値は平行に移動する
ことを意味している。

補対数-対数プロット (log-log plot)

$$\log\{-\log S(t)\} = \exp\{-\log S_0(t)\} + \{\beta_0 + \beta_1 x_1 + \dots + \beta_k x_k\}$$

を検証するため、縦軸に $\log\{-\log S(t)\}$ 横軸に $\log S(t)$ をプロットする。もし、比例ハザード性が成り立っているならば、層ごとにプロットは平行になるはずである。



Gehanの白血病データのlog-log plot

6-MPによる治療群と、対照群でlog-log plotを描いた。プロットは十分に平行であり、比例ハザード性は満たされていると考えられる。

2013/10/17

time

東北大学 医学統計勉強会

23

非比例ハザード性への対処

比例ハザード性の仮定が破綻している場合、以下の理由が考えられる。

1. データ集団の中に、時間への依存の仕方が異なる、複数のハザード関数が存在する。

⇒ 層別Cox比例ハザードモデル

2. 共変量が時間によって一定ではない。

⇒ 時間依存型共変量

3. 対数線形性が破綻しており、非線形な構造が存在する。

⇒ Coxモデルの線形項に、非線形変換を導入した加法型モデルを検討する。

2013/10/17

東北大学 医学統計勉強会

24

Take Home Message

1. 生存時間解析
2. 生存時間解析の基本概念
3. 生存率とKaplan-Meier推定量
4. 2群間の生存関数の差の検定
5. Cox比例ハザードモデル
6. Cox比例ハザードモデルの推定と検定
7. Coxモデルの比例ハザード性の検証
8. 非比例ハザード性への対処

以上