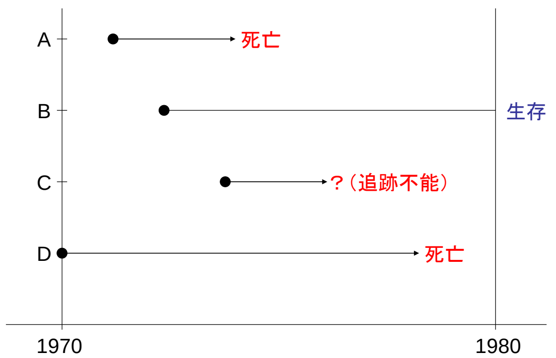
S-PLUSによる生存時間解析入門

大阪電気通信大学　情報通信工学部
辻谷将明
e-mail:ekaaf900@ricv.zaq.ne.jp

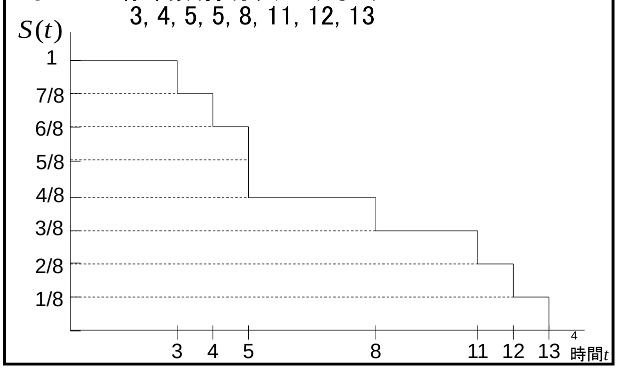
講義内容



観察期間(1970年～1980年)

生存率とは(古川・丹後,1993,10.1節)

- ある抗癌剤治療を受けた患者(8例)の、投与時点から死亡までの生存年数（打切りデータなし）
  3, 4, 5, 5, 8, 11, 12, 13



$S(t)$:(実)時間 t での生存率

1) 3年目までは全員が生存: $S(t)=1.0 \quad (0 \le t \le 3)$
2) 3年目で1人死亡したので、3年目を過ぎた瞬間 $(3+0)$ の生存率: $S(3+0)=7/8$
3) 3年目を過ぎた瞬間から4年目までは、残りの人は全員生存: $S(t)=7/8 \quad (3<t\le 4)$
4) 4年目で1人死亡したので、4年目を過ぎた瞬間の生存率: $S(4+0)=6/8$
5) 4年目を過ぎた瞬間から5年目までは、残りの人は全員生存: $S(t)=6/8 \quad (4<t\le 5)$
6) 5年目で2人死亡したので、5年目を過ぎた瞬間の生存率: $S(5+0)=4/8$
7) 5年目を過ぎた瞬間から8年目までは、他の人は死亡していない: $S(t)=4/8 \quad (5<t\le 8)$

以下同様

$$S(t)=\begin{cases} 3/8 & (8<t\le 11) \\ 2/8 & (11<t\le 12) \\ 1/8 & (12<t\le 13) \end{cases}$$

- $S(5+0)=\boxed{\frac{4}{8}}=\boxed{\frac{\cancel{6}}{8}}\times\boxed{\frac{4}{\cancel{6}}}$, 　$\frac{4}{6} = 1-\frac{2}{7-1}$

5年目を過ぎた瞬間の生存率　4年目の生存率　5年目に2人死亡　4年目までの生存数　区間[4,5)での死亡数

→打切りがある場合（例えば、8年目が打切りデータ）

3, 4, 5, 5, 8*, 11, 12, 13

- $S(11+0)=\boxed{\frac{1}{3}}=\boxed{\frac{4}{8}}\times\left(1-\frac{1}{6-2-1}\right)$

11年目を過ぎた瞬間の生存率　5年目の生存率　11年目に1死亡　5年目までの生存数　[5,11)での死亡数　[5,11)での打切り数

1. 生存時間解析とは

1.1 生存時間関数

生存時間関数 $S(t)=\Pr\{T\ge t\}$: t まで生存する確率

（累積確率）分布関数 $F(t)=1-S(t)$

ハザード（瞬間死亡率）

t の直前まで生存した人が次の Δt の期間に死亡する条件付き確率

$$\Pr\left\{t\le T<t+\Delta t \middle| T\ge t\right\}=\frac{\Pr\{T<t+\Delta t\}-\Pr\{T\ge t\}}{\Pr\{T\ge t\}}$$

Δtに依存する

単位時間当りの平均死亡率 : $\dfrac{\Pr\left\{t\le T<t+\Delta t \middle| T\ge t\right\}}{\Delta t}$

$\Delta t\to 0$のとき$\Pr\left\{t\le T<t+\Delta t \middle| T\ge t\right\}\to 0$

時間 t におけるハザード $h(t)=\lim_{\Delta t\to 0}\dfrac{\Pr\left\{t\le T<t+\Delta t \middle| T\ge t\right\}}{\Delta t}$:

t まで生存した人のうち、$t+\Delta t$ までに死ぬ人の割合を、単位時間当りの量に換算し、$\Delta t\to 0$ としたときの極限値

$$h(t)=\lim_{\Delta t\to 0}\frac{S(t)-S(t+\Delta t)}{S(t)\Delta t}$$

$$=\frac{1}{S(t)}\lim_{\Delta t\to 0}\frac{F(t+\Delta t)-F(t)}{\Delta t}$$

$$=\frac{f(t)}{S(t)}$$

$f(t)=\dfrac{dF(t)}{dt}$

確率密度関数

$\Pr\{B|A\}=\dfrac{\Pr\{A\cap B\}}{\Pr\{A\}}$ より

$$\Pr\left\{t\le T<t+\Delta t \middle| T\ge t\right\}=\frac{\Pr\left\{(t\le T<t+\Delta t)\cap(T\ge t)\right\}}{\Pr\{T\ge t\}}=\frac{\Pr\{T<t+\Delta t\}-\Pr\{T<t\}}{\Pr\{T\ge t\}}=\frac{\{1-S(t+\Delta t)\}-\{1-S(t)\}}{S(t)}$$

1.2 打切り標本

目的変数 : エンドポイント、結果(outcome)、主変数(primary variable)

独立変数 : 共変量(covariate)、危険因子(risk factor)、予後因子(prognostic factor)

データ＝（イベント発生までの観察期間、打切りか否か、共変量の値）

打切りの3つのタイプ(Putter et al., 2007)

- 研究(治験)の終了 : 打切りと死亡は独立
- 追跡不能
  - 生存関数を過小推定(downward bias) : 患者自身で完治したと判断して治療打切り
  - 生存関数を過大推定(upward bias) : 余命が少ないため，古里へ帰郷
- 競合リスク
  研究対象以外の死因が発生

Gehanの白血病データ（単位:週）

対照群

1	1	2	2	3	4	4	5	5	8	8
8	8	11	11	12	12	15	17	22	23	

6-MP投与群

6*	6	6	6	7	9*	10*	10	11*	13	16
17*	19*	20*	22	23	25*	32*	32*	34*	35*	

* : 打切り

2. Kaplan-Meier(KM)法

生存率曲線(survival curve)

生存時間分布 $S(t)$のグラフで、観測開始時は1で、その後、観測死亡時ごとに減少する段階関数

死亡発生(実)時間 $0 < t_1 < t_2 < \cdots < t_j < \cdots$

$t_j < t \le t_{j+1}$ の t では

- $\hat{S}(t) = \prod_{i=1}^{j}\left(1 - \frac{d_i}{n_i}\right)$　　t_jで始まる時間区間で生存する確率の推定量

- 推定標準誤差(SE:Standard Error)

Greenwoodの公式

$$SE(t) = \hat{S}(t_j)\sqrt{\sum_{i=1}^{j}\left\{\frac{d_i}{n_i(n_i - d_i)}\right\}}$$

→95%信頼区間　$\hat{S}(t) \pm 1.96SE(t)$

(注)この値は1を超えたり、負になることがある

6-MP投与群

- $t = 6$ で3例死亡 $\rightarrow \hat{S}(6) = 1 - \frac{3}{21} = 0.8571$

t=7での死亡数

- $t = 7$で1例死亡 $\rightarrow \hat{S}(7) = \hat{S}(6) \times \left(1 - \frac{1}{17}\right) = 0.8067$

$$17 = 21 - 3 - 1$$

t=6までの生存数

区間[6,7)での死亡数

区間[6,7)での打切り数

t=10での死亡数

- $t = 10$で1例死亡 $\rightarrow \hat{S}(10) = \hat{S}(7) \times \left(1 - \frac{1}{15}\right) = 0.7529$

$$15 = 17 - 1 - 1$$

t=7までの生存数

区間[7,10)での死亡数

区間[7,10)での打切り数

6-MP投与群

- $t = 6$ で3例死亡 $\rightarrow \hat{S}(6) = 1 - \frac{3}{21} = 0.8571$
- $t = 7$ で1例死亡 $\rightarrow \hat{S}(7) = \hat{S}(6) \times (1 - \frac{1}{17}) = 0.8067$

死亡　打切り　$17 = 21 - 3 - 1$

- $t = 10$で1例死亡 $\rightarrow \hat{S}(10) = \hat{S}(7) \times (1 - \frac{1}{15}) = 0.7529$

死亡　打切り　$15 = 17 - 1 - 1$

- $t = 13$で1例死亡 $\rightarrow \hat{S}(13) = \hat{S}(10) \times (1 - \frac{1}{12}) = 0.6902$

死亡　打切り　$12 = 15 - 1 - 2$

- $t = 16$で1例死亡 $\rightarrow \hat{S}(16) = \hat{S}(13) \times (1 - \frac{1}{11}) = 0.6275$

死亡　$11 = 12 - 1$

- $t = 22$で1例死亡 $\rightarrow \hat{S}(22) = \hat{S}(16) \times (1 - \frac{1}{7}) = 0.5379$

死亡　打切り　$7 = 11 - 1 - 3$

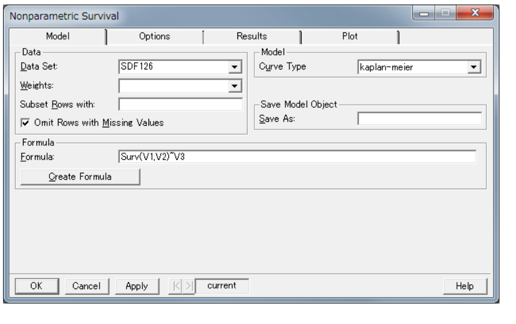
- $t = 23$で1例死亡 $\rightarrow \hat{S}(23) = \hat{S}(22) \times (1 - \frac{1}{6}) = 0.4483$

表1　6-MP投与群

時点	n_i	d_i	$\frac{d_i}{n_i(n_i - d_i)}$	$\sum_i\left\{\frac{d_i}{n_i(n_i - d_i)}\right\}$	$\hat{S}(t)$	SE	信頼区間 $\hat{S} - SE$	信頼区間 $\hat{S} + SE$	修正信頼区間 $\hat{S}(t)^q$	修正信頼区間 $\hat{S}(t)^{p}$
6	21	3	0.0079	0.0079	0.8571	0.0762	0.7809	0.9333	0.6204	0.9514
7	17	1	0.0037	0.0116	0.8067	0.0869	0.7198	0.8936	0.5633	0.9228
10	15	1	0.0048	0.0164	0.7529	0.0964	0.6565	0.8493	0.5030	0.8894
13	12	1	0.0076	0.0240	0.6902	0.1069	0.5833	0.7971	0.4314	0.8492
16	11	1	0.0091	0.0331	0.6275	0.1142	0.5133	0.7417	0.3672	0.8051
22	7	1	0.0238	0.0569	0.5379	0.1283	0.4096	0.6662	0.2677	0.7470
23	6	1	0.0333	0.0902	0.4482	0.1346	0.3136	0.5828	0.1880	0.6802

$$0.1069 = 0.6902 \times \sqrt{0.0240}$$

$$0.5833 = 0.6902 - 0.1069$$

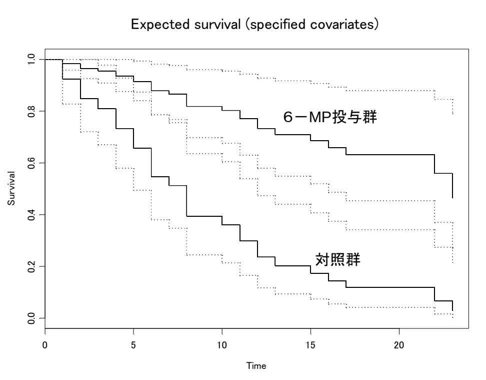


V3=0(対照群)

time	n.risk	n.event	survival	std.err	lower 95% CI	upper 95% CI
1	21	2	0.9048	0.0641	0.78754	1.000
2	19	2	0.8095	0.0857	0.65785	0.996
3	17	1	0.7619	0.0929	0.59988	0.968
4	16	2	0.6667	0.1029	0.49268	0.902
5	14	2	0.5714	0.1080	0.39455	0.828
8	12	4	0.3810	0.1060	0.22085	0.657
11	8	2	0.2857	0.0986	0.14529	0.562
12	6	2	0.1905	0.0857	0.07887	0.460
15	4	1	0.1429	0.0764	0.05011	0.407
17	3	1	0.0952	0.0641	0.02549	0.356
22	2	1	0.0476	0.0465	0.00703	0.322

V3=1(6-MP投与群)

time	n.risk	n.event	survival	std.err	lower 95% CI	upper 95% CI
6	21	3	0.857	0.0764	0.720	1.000
7	17	1	0.807	0.0869	0.653	0.996
9	16	0	0.807	0.0869	0.653	0.996
10	15	1	0.753	0.0963	0.586	0.968
11	13	0	0.753	0.0963	0.586	0.968
13	12	1	0.690	0.1068	0.510	0.935
16	11	1	0.627	0.1141	0.439	0.896
17	10	0	0.627	0.1141	0.439	0.896
19	9	0	0.627	0.1141	0.439	0.896
20	8	0	0.627	0.1141	0.439	0.896
22	7	1	0.538	0.1282	0.337	0.858
23	6	1	0.448	0.1346	0.249	0.807
25	5	0	0.448	0.1346	0.249	0.807
32	4	0	0.448	0.1346	0.249	0.807
34	2	0	0.448	0.1346	0.249	0.807
35	1	0	0.448	0.1346	0.249	0.807



生存関数のプロット

3. Cox回帰とその応用

Coxの比例ハザードモデルの流れ

ハザード関数を求める

↓

ハザード関数の妥当性をチェック

↓

回帰係数の検定

↓

生存関数の推定

＜参考＞回帰分析とは

表1は中学生男子15人のボール投げ，握力，身長の測定値を示す．ボール投げ(y)が，基礎的な体力をあらわす握力(X_1)，身長(X_2)の2つの変数で，どの程度説明されるか．

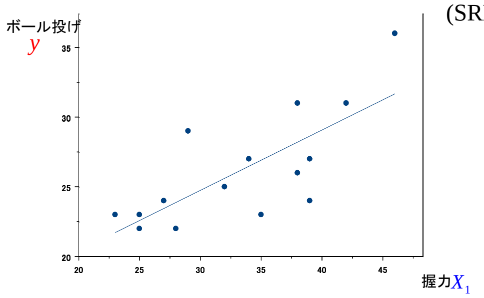
表1 中学生15人の体力測定値

生徒番号	目的変数	説明変数	
	ボール投げ(m)(y)	握力(kg)(X_1)	身長(cm)(X_2)
1	22	28	146
2	36	46	169
3	24	39	160
4	22	25	156
5	27	34	161
6	29	29	168
7	26	38	154
8	23	23	153
9	31	42	160
10	24	27	152
11	23	35	155
12	27	39	154
13	31	38	157
14	25	32	162
15	23	25	142

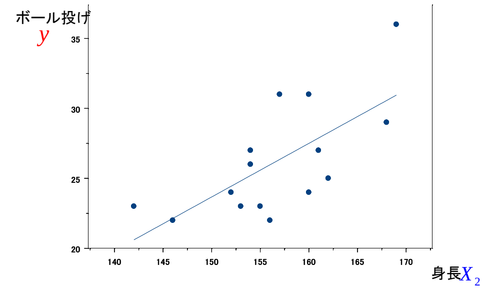
単回帰分析

単回帰分析を行い，空白の中に正しい数字を書け．

(SREG)



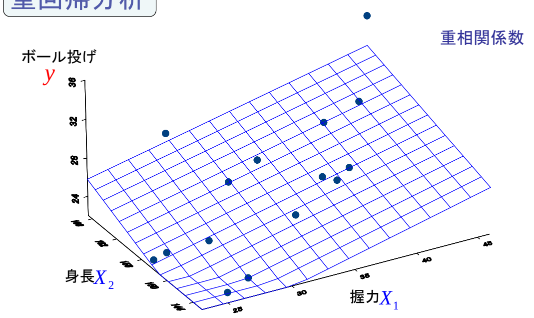
$y = \boxed{} + \boxed{}\ X_1$ 相関係数= $\boxed{}$



$y = \boxed{} + \boxed{}\ X_2$ 相関係数= $\boxed{}$

重回帰分析

重相関係数



$y = \boxed{} + \boxed{}\ X_1 + \boxed{}\ X_2$

↑ボール投げ　↑握力　↑身長

3.1 比例ハザードモデル

- 個体の生存時間に影響を与える因子（共変量）

  →背景因子、予後因子(prognostic factor)

  $\mathbf{z} = \left(z_1, z_2, \cdots, z_p\right)$ ：$\mathbf{z}$ の値は時間に依存しない

- $h\left(t \middle| \mathbf{z}\right)$ ：共変量 $\mathbf{z}$ をもつ個体のハザード

$\mathbf{z}$ の関数

$$h\left(t \middle| \mathbf{z}\right) = \underbrace{h_0\left(t\right)}_{\text{基準(ベースライン)ハザード}} \underbrace{r\left(\mathbf{z}\right)}_{\text{相対リスク(relative risk)}}$$ ：比例ハザードモデル

回帰係数

- $$h\left(t\right) = h_0\left(t\right)\exp\left(\underbrace{\boldsymbol{\beta}^T\mathbf{z}}_{\text{予後指数(prognostic index)}}\right)$$

$$= h_0\left(t\right)\exp\left(\beta_1 z_1 + \beta_2 z_2 + \cdots + \beta_p z_p\right)$$

特に $z_1 = z_2 = \cdots = z_p = 0$ のとき

$$h\left(t\right) = h_0\left(t\right)$$ ：基準（ベースライン）ハザード関数

- 共変量1個の場合

$$h\left(t\right) = h_0\left(t\right)\exp\left(\beta_1 z_1\right)$$

$$z_1 = \begin{cases} 0 \text{ ：ﾌﾟﾗｾﾎﾞ} \to h_p\left(t\right) = h_0\left(t\right) & \text{：ﾌﾟﾗｾﾎﾞのﾊｻﾞｰﾄﾞ関数} \\ 1 \text{ ：実薬} \to h_a\left(t\right) = h_0\left(t\right)\exp\left(\beta_1\right) \end{cases}$$

$$\frac{h_a\left(t\right)}{h_p\left(t\right)} = \exp\left(\beta_1\right) \quad (1)$$ ：薬剤間のハザード比

$$\begin{cases}\hat{\beta} = -1.760 \\ SE\left(\hat{\beta}\right) = 1.184\end{cases}$$

●プラセボに対する実薬のハザード比

$$\begin{cases}\text{推定値}:\ \exp(-1.760) = 0.172 \\ \text{95\%信頼区間}:\ \exp(-1.760 \pm 1.96 \times 1.184)\end{cases}$$

$$= [0.017, 1.752]$$

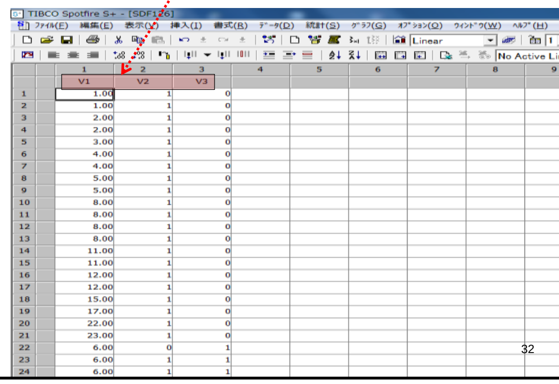
●帰無仮説$H_0 : \beta = 0$

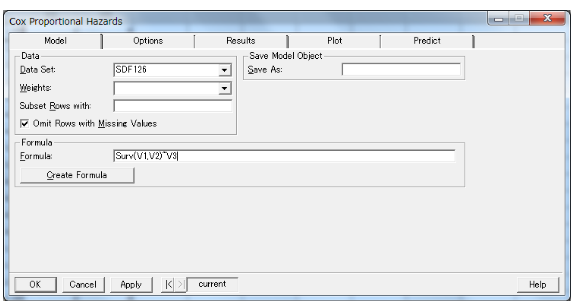
Wald検定 $(-1.760/1.184)^2 = 2.210$

$\rightarrow p\text{値} = 0.137$

●$AIC = -2 \times (\text{部分対数尤度}) + 2 \times (\text{共変量の個数})$
$= 169.485 + 2 \times 2 = 171.485$

変数名を指定しなければ自動的にV1,V2,V3となる





*** Cox Proportional Hazards ***

Call:
coxph(formula = Surv(V1, V2) ~ V3, data = SDF4,
na.action = na.exclude, method
= "efron", robust = FALSE)

n= 7

$\hat{\beta}$ $\boxed{-1.93/1.2}$

	coef	exp(coef)	se(coef)	z	p
V3	$\boxed{-1.93}$	0.145	1.2	$\boxed{-1.6}$	0.11

	exp(coef)	exp(-coef)	lower .95	upper .95
V3	0.145	6.9	0.0137	1.53

Rsquare= 0.352 (max possible= 0.854)
Likelihood ratio test= 3.03 on 1 df, p=0.0816
Wald test = 2.57 on 1 df, p=0.109
Score (logrank) test = 3.2 on 1 df, p=0.0735

Coxの比例ハザードモデルの流れ

ハザード関数を求める

↓

ハザード関数の妥当性をチェック

↓

回帰係数の検定

↓

生存関数の推定

[実例2(Collett,1994)] 前立腺癌(ステージⅢ)

生存時間(月)	打切り(0)	処置法	年齢(歳)	SHG (gm/100ml)	腫瘍のサイズ (cm²)	Gleason指数
TIME	CENSOR	TREAT	AGE	SHG	SIZE	INDEX
65	0	0	67	13.4	34	8
61	0	1	60	14.6	4	10
.	.	.	.	.	.	.
14	1	0	73	12.4	18	11
.	.	.	.	.	.	.
62	0	0	63	13.2	3	8

処置法:0→ﾌﾟﾗｾﾎﾞ、1→DES(ｼﾞｴﾁﾙｽﾁﾙﾍﾞｽﾄﾛｰﾙ;合成女性ﾎﾙﾓﾝ)1.0mg投与

SHG:血清ヘモグロビンのレベル

Gleason指数:腫瘍ステージとgradeを結合した指数(腫瘍が進行するとこの値は大きくなる)

変数選択

```
library(MASS)
train <- read.table("G:¥¥train.txt",header=T)
model<-survreg(Surv(TIME,CENSOR)~.,data=train)
stepAIC(model)
```

デフォルト:変数減少法

c.f.,変数減増法
stepAIC(model,direction="both")


```
Start:  AIC= 33.95
 Surv(TIME, CENSOR) ~ TREAT + AGE + SHG + SIZE + INDEX
        Df     AIC
  - AGE  1 32.04660
  - SHG  1 32.07095
- TREAT  1 33.06414
 <none> NA 33.95347
 - SIZE  1 36.33470
- INDEX  1 38.35046

Step:  AIC= 32.05
 Surv(TIME, CENSOR) ~ TREAT + SHG + SIZE + INDEX
        Df     AIC
  - SHG  1 30.13838
- TREAT  1 31.09420
 <none> NA 32.04660
 - SIZE  1 34.58479
- INDEX  1 36.44962
```



```
Step:  AIC= 30.14
 Surv(TIME, CENSOR) ~ TREAT + SIZE + INDEX
        Df     AIC
- TREAT  1 29.30167
 <none> NA 30.13838
 - SIZE  1 33.33638
- INDEX  1 34.46777
Step:  AIC= 29.3
 Surv(TIME, CENSOR) ~ SIZE + INDEX
        Df     AIC
 <none> NA 29.30167
- INDEX  1 33.95789
 - SIZE  1 34.37658
Call:
survreg(formula = Surv(TIME, CENSOR) ~ SIZE + INDEX, data = train)
Coefficients:
 (Intercept)       SIZE      INDEX
   8.215146 -0.04460917 -0.2897448
Dispersion (scale) = 0.3678091
Degrees of Freedom: 38 Total; 34 Residual
-2*Log-Likelihood: 21.30167
```


c.f.,全共変量を用いた場合

Analysis of Maximum Likelihood Estimates

Variable	DF	Parameter Estimate	Standard Error	Chi-Square	Pr> ChiSq	Hazard Ratio
TREAT	1	-1.182	1.210	0.954	0.329	0.307
AGE	1	0.044	0.072	0.373	0.541	1.045
SHG	1	-0.022	0.453	0.002	0.961	0.978
SIZE	1	0.094	0.052	3.254	0.071	1.099
INDEX	1	0.723	0.350	4.273	0.039	2.061

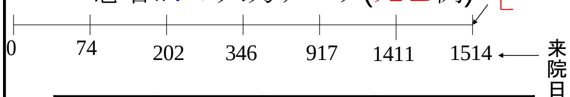
4. 周辺のモデル
(Kalbfleisch et al., 2002,6章;Lee et. Al., 2003,13章)
4.1時間依存型モデル(Collett, 2003)

患者#	生存時間(日)	打切り(=0)	ﾋﾞﾘﾙﾋﾞﾝ値:初診時
1	281	1	3.2
2	604	0	3.1
3	457	1	2.2
・	・	・	・
7	1514	1	2.4
・	・	・	・
12	1071	1	3.1

ﾋﾞﾘﾙﾋﾞﾝ(*mg/dl*):胆汁に含まれる赤黄色の色素(胆汁色素)

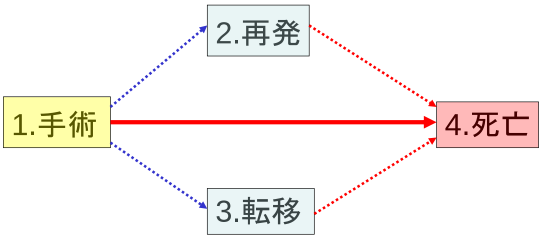
t_j	δ	ln(bil)	尤度
182	0	2.4	
281	1	3.2	$e^{3.2\beta}/(e^{3.2\beta}+e^{2.8\beta}+e^{3.9\beta}+\cdots+e^{2.3\beta}+e^{2.4\beta})$
341	0	2.8	
384	1	3.9	$e^{3.9\beta}/(e^{3.9\beta}+e^{2.2\beta}+e^{3.1\beta}+\cdot\cdot+e^{2.3\beta}+e^{2.4\beta})$
457	1	2.2	$e^{2.2\beta}/(e^{2.2\beta}+e^{3.1\beta}+e^{3.8\beta}+\cdot\cdot+e^{2.3\beta}+e^{2.4\beta})$
604	0	3.1	
814	1	3.8	$e^{3.8\beta}/(e^{3.8\beta}+e^{2.4\beta}+e^{3.1\beta}+\cdot\cdot+e^{2.3\beta}+e^{2.4\beta})$
842	1	2.4	$e^{2.4\beta}/(e^{2.4\beta}+e^{3.1\beta}+e^{2.5\beta}+e^{2.3\beta}+e^{2.4\beta})$
1071	1	3.1	$e^{3.1\beta}/(e^{3.1\beta}+e^{2.5\beta}+e^{2.3\beta}+e^{2.4\beta})$
1121	1	2.5	$e^{2.5\beta}/(e^{2.5\beta}+e^{2.3\beta}+e^{2.4\beta})$
1411	0	2.3	
1514	1	2.4	$e^{2.4\beta}/e^{2.4\beta}$

患者#7の入力データ(死亡例)

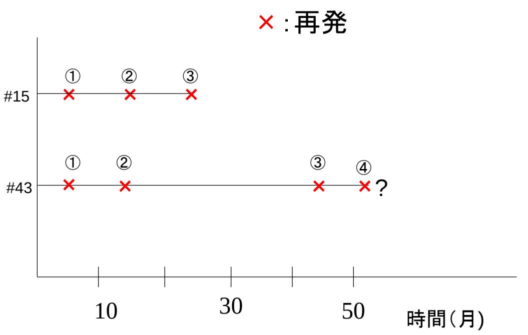


Start	Stop	打切り	*ln*(ビリルビン値)
0	74	0	2.4
74	202	0	2.9
202	346	0	3.0
346	917	0	3.0
917	1411	0	3.9
1411	1514	1	5.1

4.2 Multi-stateモデル(Lee et al., 2003,13.4節)



膀胱癌データ



患者#	追跡時間	個数(初期値)	サイズ(初期値)	処置グループ	再発時間 1	再発時間 2	再発時間 3	再発時間 4
#15	25	3	1	1	3	15	25	
#43	53	1	3	1	3	15	46	51
#86	59	1	3	2				

時間依存型

患者#	Start	Stop	Cens	Num	Size	Treat	Clinic Visit
.	.	.	.	.	.	.	.
15	0	3	1	1	1	1	1
15	3	15	1	1	1	1	2
15	15	25	1	1	1	1	3
.	.	.	.	.	.	.	.
43	0	3	1	1	3	1	1
43	3	15	1	1	3	1	2
43	15	46	1	1	3	1	3
43	46	51	1	1	3	1	4
43	51	53	0	1	3	1	5
.	.	.	.	.	.	.	.

4.3 競合リスクモデル

例

- 喫煙習慣と肺癌の疫学調査

死因：肺癌，その他の癌，その他の疾患，事故

エンドポイント　　競合リスク(“打切り”として扱う)

→ 4つの死因は(確率的に)独立ではない

(肺癌のリスクの高い人は，他の癌のリスクも高い)

打切りの3つのタイプ(Putter et al., 2007)

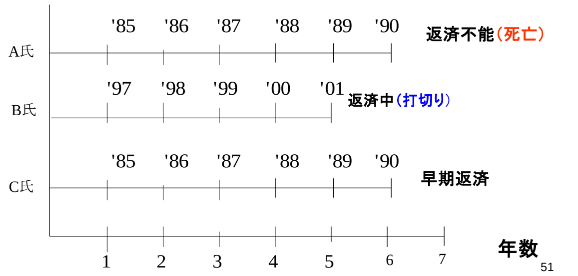
- 研究(治験)の終了：打切りと死亡は独立
- 追跡不能
  - 生存関数を過小推定(downward bias)：
    患者自身で完治したと判断して治療打切り
  - 生存関数を過大推定(upward bias)：
    余命が少ないため，古里へ帰郷
    終末医療のため転院
- 競合リスク
  研究対象以外の死因が発生

マウス発癌性データ(単位:日数)
(Kalbfleisch et. al., 2002, pp.257-259)

共変量	死因 (300radの放射後)		
	胸腺リンパ腫	網状組織細胞肉腫	その他
対照群	159,189,191,….,428,432	317,318,399,…, 748,753	40,42,51,…, 761,763
無菌群	158,192,193, …,707,800	430,590,606,…, 821,986	136,246,…, 1015,1019

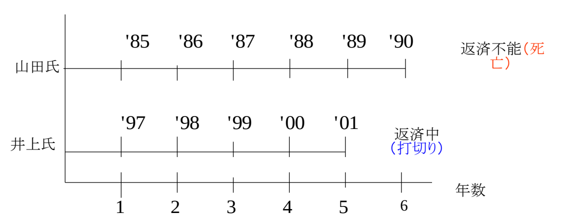
個人ローンデータ(Tsujitani,Baesense,2012)

1984年から2001年まで観測
●A氏：6年目に返済不能(死亡)
●B氏：5年目でも返済中(打切り)
●C氏：6年目で早期返済



個人ローン(生存時間解析)

1984年から2001年まで観測
●山田：6年目に返済不能(死亡)
●井上：5年目でも返済中(打切り)



個人ローンデータ

打切り= {1:返済不能, 0:その他}

顧客 #	生存 時間	打切り (=0)	共変量			
			早期返済	年齢	ローン期間	ローン目的
1	11	0	0	23	12	2
.	.	.	.	.	.	.
1813	16	1	0	44	36	3
.	.	.	.	.	.	.
5000	3	0	1	39	6	3

ﾛｰﾝ目的 {
1:Lowﾘｽｸ：ﾊﾈﾑｰﾝ, 台所用品,…
2:Mediumﾘｽｸ：家の改築, 車購入,…
3:Highﾘｽｸ：再融資、追加的な抵当,…
}

参考文献

[1]赤澤宏平，柳川尭(2010):ｻﾊﾞｲﾊﾞﾙﾃﾞｰﾀの解析，近代科学社.
[5]Cantor,A.(1997):Extending SAS Survival Analysis Techniques for Medical Research, SAS Publisher
[6]Kalbfleisch,J.D.,and Prentice,R.L.(2002):The Statistical Analysis of Failure Time Data, John Wiley, New York.
[7]J.P.ｸﾗｲﾝ，M.L.ﾒｼｭﾍﾞﾙｶﾞｰ著、打波守 訳(2009):生存時間解析，ｼｭﾌﾟﾘﾝｶﾞｰｼﾞｬﾊﾟﾝ
[8]古川俊之，丹後俊郎(1993):新版 医学への統計学、朝倉書店
[9]中村剛(2001):Cox比例ハザードモデル、朝倉書店
[10]大橋靖雄,浜田知久馬(1995):生存時間解析、東大出版
[11]丹後俊郎(2000):統計ﾓﾃﾞﾙ入門，朝倉書店
[12]辻谷将明，竹澤邦夫(2009):Rで学ぶﾃﾞｰﾀｻｲｴﾝｽ ﾏｼﾝﾗｰﾆﾝｸﾞ、共立出版
[13]辻谷将明，和田武夫(2012):Rで学ぶ確率・統計，共立出版
[14]Tsujitani,M. and Baesense,B.(2012):Survival analysis for personal loan data using generalized additive models, Behaviormetrika,Vol.39,pp.9-23.