多変量解析法—⑥

生命表解析と比例ハザードモデル

高木廣文（聖路加看護大学）
服部芳明（日本原子力事業㈱総合研究所システム解析部）

Ⅰ．はじめに

生物は特定の時間が経過すると死に至るし，機械は故障する．生体の死や機械の故障などの非可逆的変化（failure）についての特定のモデルは，一般に寿命モデルと呼ばれている．また，時間経過による個体の生死の状況などと各種の要因との関連についての分析方法は，生命表解析（life table analysis）として知られている．

寿命モデルには，生存時間の関数として指数分布，ワイブル分布，ガンマ分布，対数正規分布などを仮定することが多いが，ここではそのような仮定を設けずに，各種の要因の寿命に対する影響を分析するのに極めて有用であるCoxの比例ハザードモデル（proportional hazard model）について説明する．また，生命表解析の基本的考えや用語を理解するために，Kaplan-Meierの方法について簡単な説明を加えることにする．

Ⅱ．生命表解析の基本的方法[1,2]

特定の対象について，観測を開始した時点から非可逆的変化（死亡，故障など）が生起するまでの時間を確率変数 T で表そう．時間 t までの非可逆的変化の生起確率を $Pr(T \leq t)$ とすると，これは T の確率分布関数 $F(t)$ となる．

T の確率密度関数を $f(t)$ として，

$$\lambda(t)=\left\{\frac{dF(t)}{dt}\right\}\Big/\{1-F(t)\}=\frac{f(t)}{1-F(t)} \qquad (1)$$

を定めると，$\lambda(t)$ は時間 t まで生存した場合の，t における瞬間の致命率（瞬間故障率）を与えるものであり，ハザード関数（hazard function）と呼ばれる．また，

$$S(t)=1-F(t) \qquad (2)$$

とおくと，$S(t)$ は時間 t までの生存確率 $Pr(T<t)$ を表し，生存関数（survival function）と呼ばれる．

時間 t_i において，n_i 個体中の1例が死亡した場合を考える．$t_i<t$ となる i の集合を $v(t)=\{i|t_i<t\}$ とすると，生存関数の推定値 $S(t)$ は，

$$S(t)=\prod_{i\in v(t)}\left(1-\frac{1}{n_i}\right) \qquad (3)$$

により求められる．実際には，時間は離散的に観測されるため，時間 t_i での死亡数は2以上のこともあり（タイ tie のある場合），ここでは d_i 例あるとしよう．この場合，$S(t)$ の推定値は，

$$S(t)=\prod_{i\in v(t)}\left(1-\frac{d_i}{n_i}\right) \qquad (4)$$

となる．また，$S(t)$ の分散は，

$$Var\{S(t)\}=S^2(t)\sum_{i\in v(t)}\frac{d_i}{n_i(n_i-d_i)} \qquad (5)$$

と近似される．

上記の方法が，Kaplan-Meierの方法と呼ばれるものである．なお，実際の分析では観測はある期間内で行われるため，観測期間終了時に生存している対象者は，その時点で観測を打切ることになる．そのような対象者は打切り標本（truncated observations）と呼ばれる．また，特定の時点までは生存が確認できたが，それ以降は調査不能となった例（censored）もありえるが，両者共に分析に含めるのが普通である．ただし，観測打切りの理由が余命に大きく影響しないことが前提となる．

Ⅲ．比例ハザードモデル[2]

ある対象の寿命が，いくつかの変数の影響を受けるものと考えられる場合，それらの影響を分析できるような寿命モデルが必要とされる．Coxの比例ハザードモデルは，(1)式によるハザード関数 $\lambda(t)$ が，

p 個の説明変数 $x_1, x_2, \cdots, x_p$ の線形関数の指数関数に比例すると仮定したモデルである．すなわち，

$$\lambda(t|\boldsymbol{x}, \boldsymbol{\beta}) = \lambda_0(t) \cdot \exp(\boldsymbol{x}'\boldsymbol{\beta}) \tag{6}$$

とするモデルである．ここで，$\boldsymbol{x} = (x_1, x_2, \cdots, x_p)'$，$\boldsymbol{\beta}$ はパラメータベクトル $(\beta_1, \beta_2, \cdots, \beta_p)'$ である．また，$\lambda_0(t)$ は時間 t の関数であるが，後述するように分析上は $\lambda_0(t)$ を特定する必要はない．以下にパラメータの推定方法を，いくつかの場合に分けて説明する．

(1) 時点が連続的に観測される場合

対象数 n の観測データにおいて，k 個の時点 $t_{(1)} < t_{(2)} < \cdots < t_{(k)}$ $(k \leqq n)$ で死亡 (failure) が観測されたとしよう．観測時点が連続であれば，同一時点での死亡は1例のみとなりタイは存在しない．

時点 $t_{(i)}$ に死亡した個体 (i) の p 個の説明変数の組を p 次元ベクトル $\boldsymbol{x}_{(i)}$ とし，時点 $t_{(i)}$ まで生存が観測された個体の集合を，リスク集合 (risk set) と呼び $R(t_{(i)})$ とする．また，$R(t_{(i)})$ の要素 j の説明変数ベクトルを $\boldsymbol{x}_j$ とする．

時点 $t_{(i)}$ に死亡した個体のハザード関数値は，

$$\lambda(t_{(i)}|\boldsymbol{x}, \boldsymbol{\beta}) = \lambda_0(t_{(i)}) \cdot \exp(\boldsymbol{x}'_{(i)}\boldsymbol{\beta}) \tag{7}$$

となる．また，リスク集合の各個体についてのハザード関数値の和は，

$$\sum_{j \in R(t_{(i)})} \lambda_0(t_{(i)}) \cdot \exp(\boldsymbol{x}'_j\boldsymbol{\beta}) \tag{8}$$

となる．従って，時点 $t_{(i)}$ においてリスク集合 $R(t_{(i)})$ のうち，個体 (i) が死亡するという条件付確率 $l_{(i)}(\boldsymbol{\beta})$ は，

$$l_{(i)}(\boldsymbol{\beta}) = \frac{\exp(\boldsymbol{x}'_{(i)}\boldsymbol{\beta})}{\sum_{j \in R(t_{(i)})} \exp(\boldsymbol{x}'_j\boldsymbol{\beta})} \tag{9}$$

となり，$\lambda_0(t_{(i)})$ によらないことがわかる．

観測された全時点における上式による条件付確率の積は，モデルの全尤度ではなく，$\boldsymbol{\beta}$ についての偏尤度 (parial likelihood) を与える．対数偏尤度 $l(\boldsymbol{\beta})$ は，

$$\begin{aligned} l(\boldsymbol{\beta}) &= \sum_{i=1}^{k} \ln\, l_{(i)}(\boldsymbol{\beta}) \\ &= \sum_{i=1}^{k} [\boldsymbol{x}'_{(i)}\boldsymbol{\beta} - \ln\{\sum_{j \in R(t_{(i)})} \exp(\boldsymbol{x}'_j\boldsymbol{\beta})\}] \end{aligned} \tag{10}$$

となる．(10)式による $l(\boldsymbol{\beta})$ を最大にするような $\hat{\boldsymbol{\beta}}$ を，ニュートン・ラフソン法などの数値解析により求めればよいが，タイのある場合を含む後述の方法を用いる方がよいだろう．

(2) 時点が離散的に観測される場合

一般に観測時点は離散的であり，同一時点での死亡例が複数（タイ）となることが多い．

時点 t_i で d_i 例の死亡があり，それぞれの個体を $i_1, i_2, \cdots, i_{di}$ とし，各個体の順列の集合を Q_i とする．Q_i の要素は $d_i!$ 個の順列よりなる．集合 Q_i の任意の要素，すなわち順列 $(i_1), (i_2), \cdots, (i_{di})$ を U_i とする．順列 U_i の r 番目より前にある個体の集合 $\{U_1, U_2, \cdots, U_{r-1}\}$ を，リスク集合 $R(t_{(i)})$ の要素から除いた集合を $R(t_{(i)}, U_{i(r)})$ とする．すなわち，

$$R(t_{(i)}, U_{i(r)}) = R(t_{(i)}) - \{U_1, U_2, \cdots, U_{r-1}\} \tag{11}$$

とすると，集合 $R(t_{(i)}, U_{i(r)})$ は d_i 例の死亡の順序を U_i とした場合，r 番目の個体が死亡した時点でのリスク集合となる．死亡順序について，可能な全ての場合を考慮してモデルの偏尤度を求めると，その対数偏尤度 $l(\boldsymbol{\beta})$ は，

$$l(\boldsymbol{\beta}) = \sum_{i=1}^{k} [\boldsymbol{s}'_i\boldsymbol{\beta} + \ln \sum_{i=1}^{k} \{\prod_{r=1}^{di} \sum_{j \in R(t_{(i)}, U_{i(r)})} \exp(\boldsymbol{x}'_j\boldsymbol{\beta})\}^{-1}] \tag{12}$$

となる．ここで，$\boldsymbol{s}_i$ は d_i 死亡例の各説明変数ベクトル $\{\boldsymbol{x}_{(i,1)}, \boldsymbol{x}_{(i,2)}, \cdots, \boldsymbol{x}_{(i,di)}\}$ の和である．すなわち，

$$\boldsymbol{s}_i = \boldsymbol{x}_{(i,1)} + \boldsymbol{x}_{(i,2)} + \cdots + \boldsymbol{x}_{(i,di)} \tag{13}$$

とする．

時点 t_i における死亡例数 d_i がリスク集合の大きさに比べて小さい場合には，以下の近似式を用いることができる．すなわち，対数偏尤度関数 $l(\boldsymbol{\beta})$ は，

$$l(\boldsymbol{\beta}) = \sum_{i=1}^{k} [\boldsymbol{s}'_i\boldsymbol{\beta} - d_i \ln\{\sum_{j \in R(t_{(i)})} \exp(\boldsymbol{x}'_j\boldsymbol{\beta})\}] \tag{14}$$

により近似される．(14)式において $d_i = 1$ とすれば，タイのない場合の(10)式に一致する．「多変量解析ハンドブック」による統計学パッケージ HALBAU[2] では，(14)式による偏尤度を用いて，ニュートン・ラフソン法によりパラメータ $\boldsymbol{\beta}$ を推定している．ニュートン・ラフソン法などの非線形なモデルに対する数値解析法の詳細は成書に譲るが[3]，計算のための1階偏微分と2階偏微分は以下のようになる．なお，以下の式では，$w_j = \exp(\boldsymbol{x}'_j\boldsymbol{\beta})$ とし，$j \in R(t_{(i)})$ は j と略記する．

$$\frac{\partial l(\boldsymbol{\beta})}{\partial \boldsymbol{\beta}} = \sum_{i=1}^{k} \{\boldsymbol{s}_i - d_i \sum_j w_j\boldsymbol{x}_j / \sum_j w_j\} \tag{15}$$

および，

$$\begin{aligned} \frac{\partial^2 l(\boldsymbol{\beta})}{\partial \boldsymbol{\beta}^2} = -\sum_{i=1}^{k} d_i \{&\sum_j w_j\boldsymbol{x}_j\boldsymbol{x}'_j / \sum_j w_j \\ &- (\sum_j w_j\boldsymbol{x}_j)(\sum_j w_j\boldsymbol{x}_j)' / (\sum_j w_j)^2\} \end{aligned} \tag{16}$$

となる．なお，パラメータの推定値 $\hat{\boldsymbol{\beta}}$ の分散は以下のようにして推定される．フィシャー情報行列を，

$$I(\hat{\boldsymbol{\beta}}) = -\left(\frac{\partial^2 l(\boldsymbol{\beta})}{\partial \boldsymbol{\beta}^2}\right)_{\boldsymbol{\beta} = \hat{\boldsymbol{\beta}}} \tag{17}$$

としたとき，その逆行列 $\{I(\hat{\beta})\}^{-1}$ の対角成分を I^{ii} とすると $(i=1,2,\cdots,p)$，$\hat{\beta}_i$ の分散は，

$$\mathrm{Var}(\hat{\beta}_i)=I^{ii} \quad (i=1,2,\cdots,p) \tag{18}$$

により近似される．

(3) 生存関数の推定

比例ハザードモデルを使用する主な目的は，寿命に関する各種要因の影響を分析することにある．時間経過による生存率の変化に興味があるのなら，生存関数の推定には Kaplan-Meier 法を用いればよい．しかし，いくつかの説明変数の特定の値について生存関数を推定したい場合は，以下のような方法を用いることができる．

死亡が観測された時点 $t_{(i)}$ において，説明変数ベクトル $\boldsymbol{x}$ が $\mathbf{0}$ の場合，その死亡時刻 T が $t_{(i-1)}<T\leqq t_{(i)}$ である条件付確率を，

$$Pr(t_{(i-1)}<T\leqq t_{(i)}|\boldsymbol{x}=\mathbf{0},\boldsymbol{\beta})=1-\alpha_i \tag{19}$$

とする．説明変数ベクトルが $\boldsymbol{x}$(平均値ベクトルを用いることが多い) の場合の生存関数は，$w=\exp(\boldsymbol{x}'\boldsymbol{\beta})$ としたとき，

$$S(t_{(i)}|\boldsymbol{x},\boldsymbol{\beta})=\prod_{j=0}^{i-1}\alpha_i^{w} \quad (\text{ただし，}\alpha_0=1) \tag{20}$$

となる．なお，上記の生存関数は $t_{(i-1)}\leqq T<t_{(i)}$ なる期間での推定値を与えるものである．時点 $t_{(i)}$ での死亡例の集合を $D(t_{(i)})$ とし，それ以外の対象の集合を $R(t_{(i)})$ とする．すなわち，$A(t_{(i)})=R(t_{(i)})-D(t_{(i)})$ とする．

対数尤度関数は，$w_j=\exp(\boldsymbol{x}_j'\boldsymbol{\beta})$ とすると，

$$l(\boldsymbol{\beta},\boldsymbol{\alpha})=\sum_{i=1}^{k}\{\sum_{j\in D(t_{(i)})}\ln(1-\alpha_i^{w_j})+\ln\alpha_i\sum_{j\in A(t_{(i)})}w_j\} \tag{21}$$

となる．ニュートン・ラフソン法により $\boldsymbol{\alpha}$ を推定するためには，(21)式による対数尤度関数の 1 階偏微分と 2 階偏微分を求める必要がある．

$$\frac{\partial l(\boldsymbol{\beta},\boldsymbol{\alpha})}{\partial\alpha_i}=-\{\sum_{j\in D(t_{(i)})}w_j/(1-\alpha_i^{w_j})-\sum_{j\in R(t_{(i)})}w_j\}/\alpha_i \tag{22}$$

および，

$$\frac{\partial^2 l(\boldsymbol{\beta},\boldsymbol{\alpha})}{\partial\alpha_i^{\,2}}=-\frac{1}{\alpha_i}\cdot\frac{\partial l(\boldsymbol{\beta},\boldsymbol{\alpha})}{\partial\alpha_i}-\frac{1}{\alpha_i^{\,2}}\sum_{j\in D(t_{(i)})}\left(\frac{w_j}{1-\alpha_i^{w_j}}\right)^2\alpha_i^{w_j} \tag{23}$$

となる．なお，2 階偏微分の式では，$\alpha_l\,(l\neq i)$ の項は消えてしまう．

IV. 例　　題

比例ハザードモデルのデータの形式や，それから得られる分析結果について理解するために，簡単なモデルを仮定して100例のデータを生成したものを表 1 に示した．なお，表 1 は死亡や打切りなどの観測時点によって，データをソートしたものである．

表 1　生存期間と説明変数のデータ（架空例）

生死	期間	X1	X2	X3	生死	期間	X1	X2	X3	生死	期間	X1	X2	X3
1	0.2	8	8	6	0	3.4	3	1	8	1	6.5	8	7	6
0	0.4	2	3	4	0	3.5	6	4	3	1	6.6	7	8	1
0	0.5	4	3	3	1	3.6	3	7	1	1	6.6	7	1	2
0	0.8	1	3	5	1	3.6	2	3	0	1	7.1	7	8	7
0	0.8	7	0	5	1	3.6	2	1	0	1	7.2	2	0	4
1	0.9	7	3	7	0	3.7	1	8	2	1	7.5	0	4	2
1	1.0	8	9	2	1	3.8	9	1	8	1	7.8	8	0	0
1	1.1	8	5	1	0	3.9	3	9	9	1	7.8	2	2	9
0	1.4	7	4	1	1	4.1	4	6	9	0	7.9	1	4	8
1	1.5	5	4	9	1	4.1	6	7	0	1	8.0	7	7	6
0	1.5	4	1	4	1	4.3	9	5	2	1	8.0	1	6	3
1	1.6	4	7	3	1	4.3	8	0	1	1	8.0	4	1	7
0	1.6	1	4	8	1	4.4	5	5	2	1	8.1	0	1	0
0	1.7	2	7	4	1	4.6	7	6	3	1	8.3	0	7	3
1	2.1	1	8	7	1	4.6	7	7	3	1	8.5	5	9	6
0	2.1	1	9	8	1	4.6	4	7	2	1	9.2	8	0	3
1	2.2	7	3	2	1	4.6	9	3	2	1	9.3	1	0	0
0	2.2	9	0	9	0	4.8	2	5	1	1	9.4	3	7	2
1	2.3	0	7	8	1	4.8	1	4	1	1	9.5	5	5	5
1	2.7	9	8	6	1	4.8	5	6	8	1	9.7	0	3	8
1	2.7	9	9	9	1	4.9	0	9	0	1	9.8	2	6	8
0	2.8	5	3	4	1	5.0	8	8	4	1	10.8	9	1	6
1	3.0	3	8	6	1	5.1	7	0	7	1	11.4	1	2	9
1	3.0	9	8	1	1	5.1	8	6	8	1	11.5	5	4	5
1	3.1	7	0	3	1	5.1	5	2	9	0	11.5	0	5	8
1	3.1	1	6	5	1	5.3	4	8	4	1	11.8	9	0	8
1	3.1	7	4	8	0	5.3	3	6	2	0	12.8	0	1	6
1	3.1	8	7	2	1	5.5	9	1	0	0	14.0	2	0	6
1	3.2	9	2	4	1	5.7	5	7	4	1	15.0	8	1	8
1	3.2	7	5	4	1	5.7	5	8	7	0	16.0	1	6	9
0	3.4	7	6	0	1	6.1	5	2	4	1	16.2	5	1	7
1	3.4	5	9	6	1	6.3	0	5	1	1	18.6	3	0	6
1	3.4	4	8	9	0	6.4	2	1	5					
1	3.4	6	5	8	1	6.4	3	4	1					

(注)　生死：1 は死亡，0 はその時点で打切り．

表 1 において，生死が 1 ならば死亡を，0 ならば打切り例を表し，期間にそれが発生した時点を示した．$X1, X2, X3$ は説明変数であり，0 ～ 9 の離散値をとる．モデルとして仮定したパラメータベクトル $\boldsymbol{\beta}$ は，$(0.15, 0.1, -0.1)$ である．したがって，変数 $X1$ と $X2$ はハザードを高めるように，変数 $X3$ は低めるように設定されている．また，各時点は 0.1 の間隔とし，説明変数ベクトルが $\mathbf{0}$ の場合，その期間内の

表2　比例ハザードモデルの分析結果

変数名	係数	標準誤差	t 値	F 値	限界確率
$X1$	1.544926D−01	4.161957D−02	3.712	13.779	0.00034
$X2$	1.686166D−01	4.183922D−02	4.030	16.242	0.00011
$X3$	−1.177692D−01	4.571098D−02	−2.576	6.638	0.01151
最大対数尤度	−2.568520D+02				
AIC	5.197040D+02				

生存率は0.9998とした．

各例について説明変数ベクトルの値を一様乱数により定め，各期間内の生存率が乱数の値よりも小さければ死亡とした．さらに，生存の場合にはもう一度乱数を発生させ，その値が0.996以上ならばその時点で打切り例とした．データを発生させたルーチンをプログラム1に示したので，詳細はそちらを参考にして欲しい．

プログラム1

```
100 'プログラム1　生存データの発生ルーチン
110 :
120 DEFINT I-L:DEFDBL C,H,P,T,X-Z
130 DEF FNR(X,Y,Z)=EXP(.15#X+.1#Y-.1#Z)
140 OPEN "B:SUV.DAT" FOR OUTPUT AS #1
150 PRINT #1,"100.6":PRINT #1,"No,Criteria,Time,X1,X2,X3"
160 A=VAL(TIME$)+VAL(MID$(TIME$,4))+VAL(MID$(TIME$,7))
170 RANDOMIZE A:P=.9998#
180 FOR I=1 TO 100:T=.1#:H=1#
190   S=32767#RND(1):RANDOMIZE S
200   GOSUB #XX.DOSE:C=FNR(X,Y,Z):C=P^C:IF C>=1 THEN C=P
210     H=H#C
220     IF RND(1)>H THEN ID=1:GOTO 260
230     IF RND(1)>=.996 THEN ID=0:GOTO 260
240     T=T+.1#
250     GOTO 210
260   PRINT #1,USING"###_,#_.##.#_,#_.#_.#";I,ID,T,X,Y,Z
270 NEXT I
280 CLOSE:END
290 :
300 #XX.DOSE
310   X=INT(RND(1)#10)
320   Y=INT(RND(1)#10)
330   Z=INT(RND(1)#10)
340 RETURN
```

表2は多変量解析パッケージHALBAUを用いて，各変数のパラメータを推定した結果である．表2中の「係数」に各変数のパラメータのニュートン・ラフソン法による推定値を示した．また，「標準誤差」は(17)，(18)式によるフィッシャー情報行列を用いて求めた分散の平方根である．「t 値」は各パラメータが0か否かの検定のためのものであり，「F 値」はこの場合は t 値の2乗である．目安として F 値が2以上ならば，標本数が多い場合にはその変数は比例ハザードモデルに何等かの寄与をしていると考えてよい場合が多い．ただし，検定結果に対する正確な有意水準は「限界確率」によるべきだろう．限界確率が0.05以下ならば，その変数のパラメータは0でないと5％以下の有意水準でいえる．「最大対数尤度」は(14)式に推定されたパラメータを代入し，データに対して計算したものであり，モデルのデータへの適合度を比較する場合に使用することもあり，HALBAUでは標準出力となっている．「AIC」はモデルの適合度を示すためのものであり，ここでは説明変数の追加や削除の影響を調べるために出力される．詳しくは「多変量解析ハンドブック」の該当する箇所を参照して欲しい．

表2から，変数 $X1, X2, X3$ のパラメータの推定値はそれぞれ，0.154, 0.169, −0.118であり，検定結果はすべて高度に有意であることがわかる．しかし，$X2$ の推定値が仮定したモデルでの値0.1よりかなり大きく推定されている．ただし，表2に示した各変数のパラメータの95％信頼区間は，およそ“推定値± 2 ×標準誤差”となるので，$X2$ でもその95％信頼区間は(0.085, 0.252)となり，真の値0.1を含んでいる．推定値がやや偏よりをもつ原因として，打切り例が24例とかなり多く，死亡例と打切り例では変数の分布に差があることがあげられる（必ずしも有意差があるということではない)．これには使用した乱数の問題もあるが，1度だけのシミュレーションでは結果にこの程度の偏りが生じるのはそれほど稀なことではない．この点から，通常の分析はただ1度のデータ収集に基づいて行われるので，かりに無作為抽出法により標本を選んだとしても，調査方法や打切り例の取り扱いに関しては，かなり慎重に吟味する必要があるものといえよう．ここでは紙面の都合上，各変数の単純集計や相関などについて触れていないが，実際の分析では多変量解析を行う前に，必ず変数の分布などを調べ，データの特性をよく把握しておくべきである．

図1は生存関数の変化を示したものである．比例ハザードモデルでは，各変数の平均値を用いて(20)式により各時点の生存関数を推定した．また，(4)式のKaplan-Meier法による生存関数の推移を併せて表示した．すでに述べたように2つの方法は，分析の目的が異なるので単純な比較はあまり意味がない．

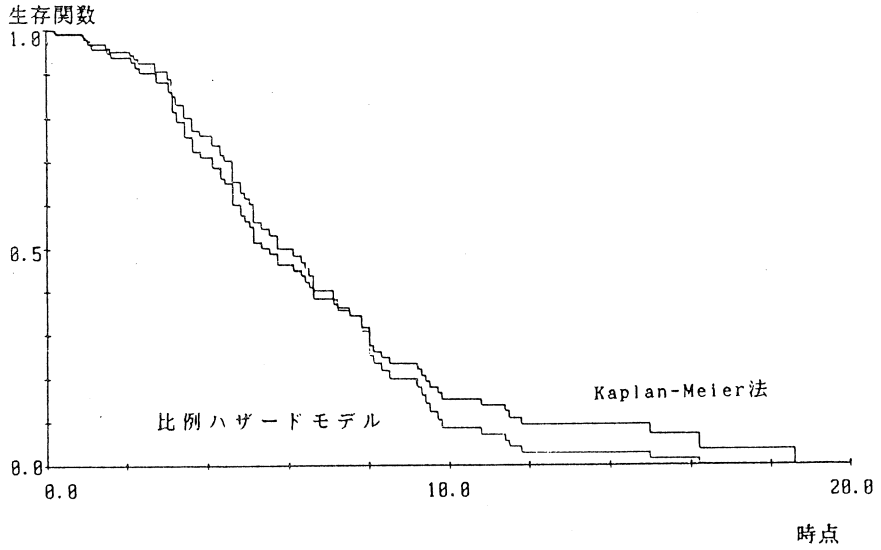

図1　生存関数の推移

図1から，2つの生存関数の推移は比較的よく符合しているものといえよう．ただし，期間が約8.0までは比例ハザードモデルによる生存関数の値の方がKaplan-Meier法によるものより大きく，その後小さくなるということが判る．また，期間が長くなるにつれ両者の差が大きくなるように見えるが，これは例数がその辺りでは極めて小さく，生存関数の推定値が不安定であることもその原因の一つと考えられる．

V．おわりに

生命表解析の基本概念について簡単な説明を行い，寿命に影響すると考えられる変数がある場合に，比例ハザードモデルによりその影響の程度を分析する方法についても解説を加えた．寿命に関する分析方法は，ここで取り上げた以外にも多数あるので，自分のデータの特性をよく把握し，然るべきモデルにより分析を行って欲しい．また，多変量解析の手法を用いる場合には，必ず基本的解析をあらかじめ行って欲しい．すなわち，度数分布や平均値・分散などを求めること，クロス集計や相関係数・相関図などを求めることなどは，分析の基本である．

これまでに，多変量解析パッケージHALBAUにバグがあるのではないかという抗議や質問が多数寄せられている．それらの問い合わせの大部分は，プログラムのバグではなく，データをよく吟味せずに分析を行い，計算途中でエラーを起こしたものであった．多変量解析は理論をある程度理解していないと，分析に使用できないデータもあり，データを理解するという点からも，基本統計による分析を必ず実行して欲しい．そのために，HALBAUには基本統計のためのルーチンを付録として提供している．何でも多変量解析というのではなく，分析にはそれなりの手順があることを理解すべきだろう．

参考文献

1)　富永祐民：治療効果判定のための実用統計学―生命表法の解説―，蟹書房，1980
2)　柳井晴夫，高木廣文　編著：多変量解析ハンドブック，160-182，現代数学社，1986
3)　Kowalik, J. and Osborne, M.R.: Methods for Uncontrained Optimization Problems, 1968, Am. Elsevier Pub. Com., Inc.（山本善之，小山建夫共訳：非線形最適化問題　制約条件のない最適化の手法，1970，培風館）

（たかぎ　ひろぶみ／はっとり　よしあき）