

# AMCoR

Asahikawa Medical College Repository <http://amcor.asahikawa-med.ac.jp/>

北海道放射線技術雑誌 (1990.03) 50号:32～34.

パーソナルコンピューターを用いた生存率曲線の作成

平田良昭, 西部茂美, 菊池雄三

## 《ノート》

## パーソナルコンピュータを用いた生存率曲線の作成

平田良昭\* 西部茂美\* 菊池雄三\*\*

**要旨** 市販のアプリケーションソフトにより作成されたデータファイルからデータを読み込んでKaplan-Meierによる計算法で生存率を算出し、それをグラフ表示するプログラムを作成した。

プログラムの概要については、各アプリケーション毎のデータファイルを変数として読み込み、生存率計算式の必要項目をもとめる。ここでは各患者毎にグループ 1、観察開始年月日 6、追跡年月日 6、追跡結果 1の4項目計18バイトにデータを分割している。観察開始年月日と追跡年月日、死亡している場合には死亡年月日の差により観察期間が計算される。グループ毎に振り分けられた患者の生存期間を短い順に並べ替える。次に計算式により生存率とその時の標準偏差をもとめる。最後に得られた生存率をプリンターに出力し、CRTにはグラフ表示を行う。

従来、手計算で数10分の計算時間を要していたが、データが完備されていれば生存率計算プログラムの作成により数秒で結果が得られるようになった。

北放技術誌 50:32-34, 1990

## I. 緒言

近年、パーソナルコンピュータを用いた放射線治療患者のデータベースが作成され種々の治療効果の解析が行われている。中でもKaplan-Meierの方法による生存率の計算は標本数が比較的小さくても、正確な生存率が得られる方法として広く使用されている事は周知の通りである。またこの方法は打ち切り例が表示されるため、残りの生存例が確認される方法なので便利である。しかし市販のアプリケーションソフトを利用してデータベースを構築してもビジネス用途以外の特殊なグラフィックの出力についてはまだ充分とは言えない。今回市販のアプリケーションソフトにより作成されたデータファイルからデータを読み込んでKaplan-Meierによる計算法で生存率を算出し、それをグラフ表示するプログラムを作成したので報告する。

## II. 目的

従来、手計算で数10分間の計算時間を要していたのを、短時間で且つ正確に結果を得られるようにするため(省力化)、又同一OS下において構築されている放

射線治療患者情報の有効利用を計るため(リレーション操作)に作成した次第である。

## III. 使用機器、ハードウェアおよびソフトウェア

CPU: NEC PC-9801RA5 (5 inch×2+40 M hard disk)

CRT: PC-KD853 (640×400×2)

PRINTER: PC-PR201H3

APPLICATION SOFTWARE:

N88BASIC (MS-DOS版) 日本電気(株)

dBASE3 PLUS 日本アッシュトンテイト(株)

PC-PAL (株)大塚商会

一太郎 (株)ジャストシステム

## IV. Kaplan-Meierの計算式について

$$S(t) = \frac{(n-1)}{(n-1+1)} \cdot \frac{(n-2)}{(n-2+1)} \cdots \frac{(n-r)}{(n-r+1)}$$

$S(t)$ : 時間  $t$  まで生存期間が短い順に  $r$  番目の症例が生存する確率

$n$ : 全症例数

$$\text{Var}[S(t)] = S(t)^2 \times \left\{ \sum_{i=1}^r \frac{1}{(n-i)(n-i+1)} \right\}$$

$\text{Var}[S(t)]$ :  $S(t)$  の分散

\* 旭川医科大学附属病院放射線部

\*\* 旭川医科大学医学部放射線医学講座

(平成2年1月18日受理)

この方法は観察開始期間から死亡が発生する毎に一例ずつ死亡時点までの生存率を計算するものである。ここで、 $S(t)$ は時間  $t$  までの生存期間が短い順に  $r$  番目の症例が生ずる確率である。 $r$  番目までの打ち切り例はこの式の  $r$  に 1 を加えて  $S(t)$  を計算する。又  $S(t)$  の分散は下の式で計算され、標準偏差はこの値の平方根をとった値となる。

V. 方法及び結果

プログラムの概要について以下に示す。各アプリケーション毎のデータファイルを変数として読み込み、先に示した計算式の項目をもとめる。ここでは各患者毎にグループ 1、観察開始年月日 6、追跡年月日 6、追跡結果 1 の 4 項目計 18 バイトにデータを分割している。観察開始年月日と追跡年月日、死亡している場合には死亡年月日の差により観察期間が計算される。グループ毎に振り分けられた患者の生存期間を短い順に並べ替える。次に前述の計算式により生存率とその時の標準偏差をもとめる。最後に得られた生存率をプリンターに出力し、CRTにはグラフ表示を行う。

アプリケーションソフトにより作成されたデータフォーマットについてはつぎの通りである (Table 1)。

Basicシーケンシャルファイル	18バイト/1データで順に書き込まれる
Database A	255バイト/1データで書き込まれ初めの255バイトはシステムで使用し次の255バイトよりデータを書き込んでいる
Database B	255バイト/1データで書き込まれ初めの30バイトはシステムで使用し次の31バイトよりデータを書き込んでいる
Wordprocessor	Basicシーケンシャルファイルと同じ

Table 1 アプリケーションソフトによるデータフォーマットの違い

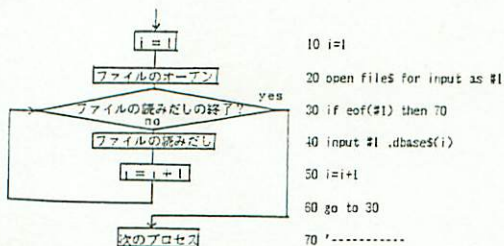


Fig. 1 Basic によるシーケンシャルファイルの読みだし

ここでは BASIC 文によるシーケンシャルファイルのアクセスする方法を利用している。Fig. 1 に示すように順々にファイルを読み込んでいく。しかし、この方法だけではアプリケーションソフトのフォーマットの形式により、ファイルの先頭に本プログラムに使用しないデータが入っているため、これらのデータを読み込んだ後、文字列操作関数により必要な項目にデータを振り分けるようにしなければならない。Data Base A の場合には先頭ファイル 255 バイトはアプリケーションの方で使用し、本プログラムでは使用しない。Data Base B では先頭ファイルの 30 バイトがアプリケーションで使用している。又、これらのデータファイルの読み込みは 255 バイトずつ行われ、患者 1 人分のデータが 2 つのファイルにまたがる事があり、これにも文字列操作が必要である。通常リレーショナルデータベースの場合、患者名、住所等、計算式に不要な項目が含まれているので、リレーショナル操作でいうところの射影を用いて必要なデータ項目のみに凝縮する処理を行う。ワープロソフトでは BASIC 文と同じようにそのままデータを読み込んでデブロッキングすればよい。

デブロックの方法について以下に述べる。

- (1) 1 レコード目の 255 バイト中何バイト目からデータとなっているかチェックする。
- (2) それが決まればそこから 18 バイトづつ文字列データにデブロックする。
- (3) 1 レコード目と 2 レコード目にまたがるデータをチェックする。
- (4) 2 レコード目では 1 レコード目とまたがったデータを除いて同じく 18 バイトづつデブロックしていく。
- (5) 最後のレコードまで (4) を繰り返す。

生存率曲線と結果の出力例を示す (Table 2, Fig. 2)。データがフロッピーディスク内に完備されていれば、数秒で CRT 及びプリンターに出力可能である。

VI. 考 察

データベースは蓄積されたデータが統計処理より何かの傾向を推論したり集計したりして、ある結果を導き出す手段に用いられてこそ有効に生きてくる。今回我々は同一 OS であるところの MS-DOS の下でプログラムを作成したが、データのフォーマットが異なったり、DOS が異なったりして、蓄積されたデー

グループコード : 1

No	生存期間 (カ月)	生死 (0:死 1:生存)	生存率	標準誤差
1	2	0	0.5375	0.3605
2	2	0	0.5750	0.3827
3	2	1	0.8750	0.3827
4	2	1	0.8750	0.3827
5	3	1	0.8750	0.3827
6	3	1	0.8750	0.3827
7	3	1	0.8750	0.3827
8	5	1	0.8750	0.3827
9	5	1	0.8750	0.3827
10	7	0	0.7500	0.1357
11	11	1	0.7500	0.1357
12	11	1	0.7500	0.1357
13	12	0	0.5625	0.1916
14	15	1	0.5625	0.1916
15	17	1	0.5625	0.1916
16	20	1	0.5625	0.3000

グループコード : 2

No	生存期間 (カ月)	生死 (0:死 1:生存)	生存率	標準誤差
1	2	0	0.9286	0.3688
2	2	0	0.8571	0.3935
3	2	0	0.7857	0.1097
4	2	0	0.7143	0.1207
5	3	1	0.7143	0.1207
6	3	1	0.7143	0.1207
7	3	0	0.6250	0.1347
8	4	1	0.6250	0.1347
9	5	0	0.5208	0.1471
10	8	1	0.5208	0.1471
11	8	0	0.3906	0.1578
12	10	1	0.3906	0.1578
13	12	0	0.1953	0.1590
14	14	1	0.1953	0.3000

Table 2. プリンターへの生存率データの出力例

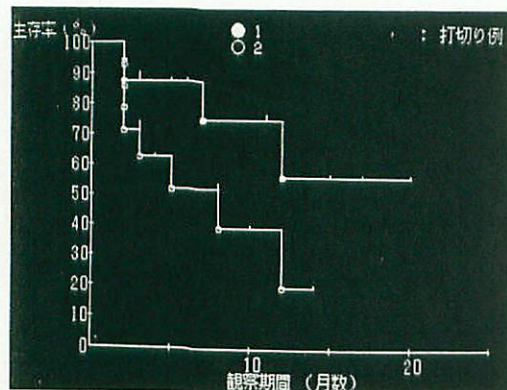


Fig. 2 CRT 上への生存率曲線の表示例

タの利用法に制限がついてはデータが活きることもまた難しい。ユーザーが独自に有しているデータが、市販のアプリケーションでファイルコンバーターを用いる事無しに使用できるようになることが望ましい。データが少ない場合にはプリントアウトして目で内容を確認できるのであるが大量のデータをそれにより確認するのは困難なので持っているデータをそのままアクセスすることにより、要求される目的が達成されるソフトウェアが理想的であろう。

## VII. 結 論

従来、手計算で数10分の計算時間を要していたが、データが完備されていれば生存率計算プログラムの作成により数秒で結果が得られるようになった。現在では同じOSの下でのワープロ、データベースソフト、表計算ソフト等各種アプリケーションソフトの間では相互のデータファイル変換ソフトが付属している事が多く、これらの利用によりデータの有効利用が計られる。

今後は、更に蓄積された患者情報の有効なデータ処理法について検討していきたい。

## 文 献

- 1) 富永祐民: 治療効果判定のための実用統計学, 90-94, 蟹書房
- 2) PC-9801RAN88日本語 BASIC (86) ユーザーズマニュアル
- 3) PC-9801RAN88日本語 BASIC (86) リファレンスマニュアル