研 究 ノ ー ト

特定死因の死亡公算算出近似式の評価

大 場 保

1. はじめに

筆者は本誌において特定死因の死亡公算を求める近似式の評価を既に報告したが[1],近似式は他にもいくつかあるため，本稿ではそれらについても評価を行うとともに，誤差の発生する傾向についてさらに詳しく検討を行う．また，真の値を求める計算方法についても，前回報告できなかったため併せて報告する．また，死亡確率，死亡公算およびp, q_i, Q_i等の呼び方と記号の意味は，特に断わらない限り前報告と同様である．

2. 方 法

（再評価の対象とする近似式）

ここで再評価を行なうのは，Jhonson and Johnson が近似式どうしで比較しあって評価した次の5つの近似式である[2]．このうち，式(1)と(2)は前報告で評価を行なったものであるが，他の近似式ともあわせて比較する意味から再び取り上げた．

$$q_{1_i} = 1 - p^{\frac{Q_i}{q}} \tag{1}$$

$$q_{2_i} = 1 - \frac{1}{2} \cdot \left((2 - Q_{-i} + Q_i) - \sqrt{(2 - Q_{-i} + Q_i)^2 - 8Q_i}\right) \tag{2}$$

$$q_{3_i} = Q_i \cdot \frac{1 - \frac{Q_{-i}}{2}}{1 - Q_{-i}} \tag{3}$$

$$q_{4_i} = \frac{Q_i}{1 - \frac{Q_{-i}}{2}} \tag{4}$$

$$q_{5_i} = Q_i \cdot \frac{1 - \frac{Q_i}{2}}{1 - \frac{q}{2}} \tag{5}$$

ただし，q_{l_i}（$l = 1, 2, 3, 4, 5$）は，式(l)による死因iの死亡公算q_iの近似値である．

---

1) 大場　保，「特定死因を除いた場合の死亡確率計算に関する考察」，『人口問題研究』，187号，1988年．

2) R. C. Elandt-Johnson and N. L. Johnson, *Survival Models and Data Analysis* (New York : Wiley-Interscience, 1980), pp.312－17.

（死亡公算の考え方と $\{q_i\}$ より $\{Q_i\}$ を求める方法と $\{Q_i\}$ より $\{q_i\}$ を求める方法）

はじめに，死亡公算の考え方と $\{q_i\}$ から $\{Q_i\}$ を求める方法について説明する．

まず，各死因は確率的に独立であると仮定される．これをわかりやすく例えれば，「あるひとりの人のある期間の生死を決定する神様は，天国に死因の数だけクジの箱を持っていて，各死因の当たりクジには死亡する日付が書いてある．神様はその期間の始めにすべての箱のクジを引き，一つも当たらなかったらその人はその期間生き残れる．どれか一つのクジに当たれば，その箱の死因によりクジの日付の日に死亡する．複数の当たりクジを引いたなら，日付の早い方のクジに従って死亡する．また，各死因の当たりクジにかかれた日付の出現頻度は，期間を通して一定であるものとする．（厳密には，各クジがあたる確率 q_i と総死亡確率 q の比が期間を通して一定であるとする．）」となる．この場合，m 個の当たりクジが出た場合は，$1/m$ の確率でそれぞれの当たりクジ死因で死亡することになる．

このように考えると総生残確率 p は，

$$p = p_1 \cdot p_2 \cdots p_n = (1-q_1)\cdot(1-q_2)\cdots(1-q_n) \tag{6}$$

として求まる．ここで，n は死因の数，p_i は i 番目の死因のクジがはずれる確率（生残公算），q_i は i 番目の死因のクジが当たる確率（死亡公算）である．一方，現実に観測される死因別死亡数を期間初頭の母集団人口数で除して推定される死因別死亡確率 Q_i と全生残確率 p とには，

$$p = 1-q = 1-(Q_1+Q_2+\cdots Q_n) = 1-(Q_1+Q_{-1}) \tag{7}$$

という関係が成立する．q は，全死因による総死亡確率である．次に，$\{q_i\}$ と $\{Q_i\}$ の関係を式で表わすこととする．$f(i,m)$ を死因 i を含み m 個のクジが当たる確率とすると，

$$Q_i = f(i,1) + \frac{1}{2} f(i,2) + \cdots + \frac{1}{n} f(i,n) \tag{8}$$

と表わせる．$f(i,m)$ は，

$$1 = \prod_{j=1}^{n} (p_j + q_j) \tag{9}$$

の右辺を展開した時の q_i を含み m 個の q_j を含んだ項の総和である．そこでこれを求めることを考える．

さて，

$$\begin{aligned}
\prod_{j=1}^{n}(p_j+q_j) = p_1 p_2 \cdots p_n &+ q_1 p_2 \cdots p_n && + q_1 q_2 p_3 p_4 \cdots p_n + \cdots + q_1 q_2 \cdots q_n \\
&+ p_1 q_2 \cdots p_n && + q_1 p_2 q_3 p_4 \cdots p_n \\
&\quad \vdots \quad \ddots && \quad \vdots \quad \ddots \\
&+ p_1 p_2 \cdots q_n && + q_1 p_2 p_3 p_4 \cdots q_n \\
& && + p_1 q_2 q_3 p_4 \cdots p_n \\
& && + p_1 q_2 p_3 q_4 \cdots p_n \\
& && \quad \vdots \quad \ddots \\
& && + p_1 q_2 p_3 p_4 \cdots q_n \\
& && \quad \vdots \\
& && + p_1 p_2 \cdots q_{n-2} q_{n-1} p_n \\
& && + p_1 p_2 \cdots q_{n-2} p_{n-1} q_n \\
& && + p_1 q_2 \cdots p_{n-2} q_{n-1} q_n
\end{aligned} \tag{10}$$

であるが，n個の死因から，死因iを除いた他の死因，即ち，死因1,死因2,…,死因$i-1$，死因$i+1$，…，死因nに番号を付け直したもの，死因k_1，死因k_2，…，死因k_{n-1}についてくじの当りはずれの組合せを同様に考えたうえで両辺にq_iを乗ずれば，

$$
\begin{array}{llll}
 & m=1 & m=2 & m=3 \qquad\qquad\qquad m=n \\
 & \Downarrow & \Downarrow & \Downarrow \qquad\qquad\qquad\quad \Downarrow \\
q_i \prod_{j=1}^{n-1}(p_{k_j}+q_{k_j}) = & q_i p_{k_1} p_{k_2} \cdots p_{k_{n-1}} & + q_i q_{k_1} p_{k_2} \cdots p_{k_{n-1}} & + q_i q_{k_1} q_{k_2} p_{k_3} p_{k_4} \cdots p_{k_{n-1}} \quad + \cdots + q_i q_{k_1} q_{k_2} \cdots q_{k_{n-1}} \\
 & & + q_i p_{k_1} q_{k_2} \cdots p_{k_{n-1}} & + q_i q_{k_1} p_{k_2} q_{k_3} p_{k_4} \cdots p_{k_{n-1}} \\
 & & \vdots \quad \vdots \quad \vdots \quad \ddots \quad \vdots & \vdots \quad \vdots \quad \vdots \quad \vdots \quad \vdots \quad \ddots \quad \vdots \\
 & & + q_i p_{k_1} p_{k_2} \cdots q_{k_{n-1}} & + q_i q_{k_1} p_{k_2} p_{k_3} p_{k_4} \cdots q_{k_{n-1}} \\
 & & & + q_i p_{k_1} q_{k_2} q_{k_3} p_{k_4} \cdots p_{k_{n-1}} \\
 & & & + q_i p_{k_1} q_{k_2} p_{k_3} q_{k_4} \cdots p_{k_{n-1}} \\
 & & & \vdots \quad \vdots \quad \vdots \quad \vdots \quad \vdots \quad \ddots \quad \vdots \\
 & & & + q_i p_{k_1} q_{k_2} p_{k_3} p_{k_4} \cdots q_{k_{n-1}} \\
 & & & \quad \vdots \\
 & & & + q_i p_{k_1} p_{k_2} \cdots q_{k_{n-3}} q_{k_{n-2}} p_{k_{n-1}} \\
 & & & + q_i p_{k_1} p_{k_2} \cdots q_{k_{n-3}} p_{k_{n-2}} q_{k_{n-1}} \\
 & & & + q_i p_{k_1} p_{k_2} \cdots p_{k_{n-3}} q_{k_{n-2}} q_{k_{n-1}}
\end{array}
\qquad (11)
$$

となるから，$f(i,1)$，$f(i,2)$，$f(i,3)$，…は，

$$
f(i,1) = q_i p_{k_1} p_{k_2} \cdots p_{k_{n-1}},
$$

$$
\begin{aligned}
f(i,2) = {} & q_i q_{k_1} p_{k_2} \cdots p_{k_{n-1}} \\
& + q_i p_{k_1} q_{k_2} \cdots p_{k_{n-1}} \\
& \quad \vdots \quad \vdots \quad \vdots \quad \ddots \quad \vdots \\
& + q_i p_{k_1} p_{k_2} \cdots q_{k_{n-1}},
\end{aligned}
$$

$$
\begin{aligned}
f(i,3) = {} & q_i q_{k_1} q_{k_2} p_{k_3} p_{k_4} \cdots p_{k_{n-1}} \\
& + q_i q_{k_1} p_{k_2} q_{k_3} p_{k_4} \cdots p_{k_{n-1}} \\
& \quad \vdots \quad \vdots \quad \vdots \quad \vdots \quad \vdots \quad \ddots \quad \vdots \\
& + q_i q_{k_1} p_{k_2} p_{k_3} p_{k_4} \cdots q_{k_{n-1}} \\
& + q_i p_{k_1} q_{k_2} q_{k_3} p_{k_4} \cdots p_{k_{n-1}} \\
& + q_i p_{k_1} q_{k_2} p_{k_3} q_{k_4} \cdots p_{k_{n-1}} \\
& \quad \vdots \quad \vdots \quad \vdots \quad \vdots \quad \vdots \quad \ddots \quad \vdots \\
& + q_i p_{k_1} q_{k_2} p_{k_3} p_{k_4} \cdots q_{k_{n-1}} \\
& \qquad \vdots \\
& + q_i p_{k_1} p_{k_2} \cdots q_{k_{n-3}} q_{k_{n-1}} p_{k_{n-1}} \\
& + q_i p_{k_1} p_{k_2} \cdots q_{k_{n-3}} p_{k_{n-1}} q_{k_{n-1}} \\
& + q_i p_{k_1} p_{k_2} \cdots p_{k_{n-3}} q_{k_{n-1}} q_{k_{n-1}},
\end{aligned}
$$

$$\vdots$$

となる．ここで，$p_1 p_2 \cdots p_n = p$であるから，

$$f(i,1)=p\cdot\frac{q_i}{p_i},$$

$$f(i,2)=p\cdot\frac{q_i}{p_i}\sum_{j=1}^{n-1}\left(\frac{q_{k_j}}{p_{k_j}}\right),$$

$$f(i,3)=p\cdot\frac{q_i}{p_i}\sum_{j_1=1}^{n-2}\left(\frac{q_{k_{j_1}}}{p_{k_{j_1}}}\sum_{j_2=j_1+1}^{n-1}\left(\frac{q_{k_{j_2}}}{p_{k_{j_2}}}\right)\right),$$

$$\vdots$$

以下同様にして,

$$f(i,m)=p\cdot\frac{q_i}{p_i}\cdot\sum_{j_1=1}^{n-m+1}\left(\frac{q_{k_{j_1}}}{p_{k_{j_1}}}\cdot\sum_{j_2=j_1+1}^{n-m+2}\left(\frac{q_{k_{j_2}}}{p_{k_{j_2}}}\cdots\sum_{j_{m-2}=j_{m-3}+1}^{n-2}\left(\frac{q_{k_{j_{m-2}}}}{p_{k_{j_{m-2}}}}\cdot\sum_{j_{m-1}=j_{m-2}+1}^{n-1}\left(\frac{q_{k_{j_{m-1}}}}{p_{k_{j_{m-1}}}}\right)\right)\cdots\right)\right) \quad (12)$$

が得られる.

次に, $\{Q_i\}$から$\{q_i\}$を求める方法について説明する.

死因数nが4以下であれば, (12) 式のn個の連立方程式を代数学的に解き, $\{Q_i\}$から$\{q_i\}$を求めることも可能であり, Jhonson and Johnson は$n=2$としたときについての評価を行なっている ((2)式による値). しかしながら, nが5以上の場合は$\{Q_i\}$から$\{q_i\}$を代数学的に求められないため, $\{q_i\}$を与えて$\{Q_i\}$を求めることを繰り返し行うことにより本来の$\{Q_i\}$を導くような$\{q_i\}$を求めるほかない.

実際の計算手順を, 以下に示す.

(1) $\{q_i\}$の推定値$\{\hat{q}_i\}$の初期値として$\{q_{1_i}\}$(他のものでもよい) を与える.

(2) $\{\hat{q}_i\}$を用いて (12) 式より$\{Q_i\}$の暫定値$\{\hat{Q}_i\}$を求める.

(3) 新たな$\{q_i\}$の推定値として$\{\hat{q}_i\cdot(Q_i/\hat{Q}_i)\}$を与える.

(4) 必要な有効桁数が得られるまで(2)と(3)を繰り返す.

以上により$\{q_i\}$を求めることができる.

各近似式を評価するにあたっては, 死因別死亡確率$\{Q_i\}$(すべての死因別死亡確率の列)から死亡公算$\{q_i\}$(すべての死因別死亡公算の列) の真の値は直接には求められないため, 前報告同様のモデル計算として適当な条件を満たす$\{q_i\}$を与えて, これより$\{Q_i\}$を求め, この$\{Q_i\}$より各近似式を用いて推定される$\{q_{l_i}\}$を$\{q_i\}$と比較した.

次に, モデル計算の方法を説明する. ここでは以下に示す2通りの方法により$\{q_i\}$を設定し, この条件の下に$\{Q_i\}$を決定し, このような$\{Q_i\}$をもたらすような$\{q_i\}$を反復計算によって求めた. この$\{Q_i\}$から, 各近似式を用いて$\{q_{1_i}\}$, $\{q_{2_i}\}$, $\{q_{3_i}\}$, $\{q_{4_i}\}$, $\{q_{5_i}\}$を求め, $\{q_i\}$との比較を行った.

(方法1)

$q=0.2$
$Q_1\neq Q_2=Q_3=\cdots=Q_n$
$Q_1:Q_{-1}(=Q_2+Q_3+\cdots+Q_n)=0.5,\ 2$
(死因1の占める割合が, 1/3, 2/3)
$n=6,\ 11,\ 21$

(方法2)

$q=0.2$
$\{q_i\}$は公比0.75の等比数列
$n=6,\ 21$

3. 結 果

（方法１）

Q_1: Q_{-1} = 0.5，2 としたときの $\{q_{1_i}\}$〜$\{q_{5_i}\}$の誤差を図 1，2 に示す．q_i に対する誤差として q_{l_i} を q_i で除してさらに 1 を減じた値を用い，これが死因別死亡確率 { Q_i } の大きさ（常用対数をとってある）とともにどのように変化するかを示してある．これは，方法 2 においても同様である．また，方法 1 においては q_{l_2}〜q_{l_n}（ l =1, 2, 3, 4, 5 ）の値はすべて等しいので，q_{l_2}のみを示し，nが等しいものを直線で結んである．

図1　q = 0.2，Q_1 : $Q-1$ = 0.5 としたときの q_{1_i} 〜 q_{5_i}（ i = 1，2〜n）の誤差

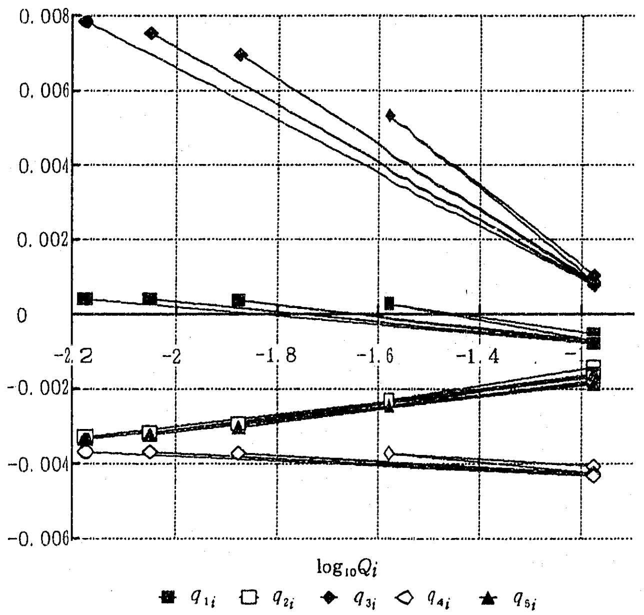
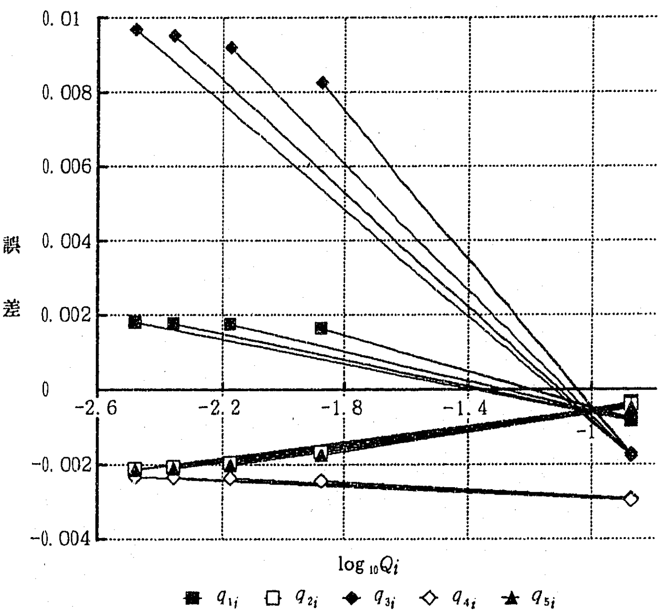


図2　q = 0.2，Q_1 : $Q-1$ = 2.0 としたときの q_{1_i} 〜 q_{5_i}（ i=1，2〜n）の誤差



まず，$\{q_{1_i}\}$であるが，これは，おおむね他の近似式より誤差が少ないことが明らかとなった．死因 1 の占める割合が，⅓，½，⅔と変化するに従い，q_{1_1}の誤差は大した変化は見られぬもののqq_{1_2}〜q_{1_n}は誤差が大きく減少した．次に誤差の小さい近似式は，$\{q_{2_i}\}$であるが，$\{q_{2_i}\}$および$\{q_{5_i}\}$の誤差は，互いにほぼ同様の変化を示し，Q_iが増加するに従い誤差が小さくなった．Q_1: Q_{-1} =0.5 のときには，q_{2_1}, q_{5_1}ともにq_{1_1}より小さい誤差を示した．方法 1, 方法 2 を通して，{ q_{1_i} } より小さい誤差が見られたのはこのときだけである．また，これらの近似値の誤差は，Q_iが小さくなるほど絶対値が増大した．{ q_{3_i} }の誤差は，{ Q_i } が 0.03 より小さいとき（q_{3_2}）には，他の近似式の誤差のいずれのものより大きな値を示した．特に，{ Q_i } が 0 に近いときほど誤差は大きく，1 %近い誤差が見られた．一方，q_{3_1}の誤差は，比較的小さかった．{ q_{4_i} }の誤差は，{ Q_i }が 0.03 より小さいときには，{ q_{3_i} } の次に大きな値を示したが，q_{4_1}は，同じ { Q_i } による他の近似値に比べて大きな誤差を示した．{ q_{4_i} }だけは，{ Q_i } が増加するに従い誤差が増加した．

（方法 2 ）

n = 6, 21としたときの $\{q_i\}$に対する $\{q_{1_i}\}$〜$\{q_{5_i}\}$の誤差を n の値に応じて図 3，4 に示す．

いずれの場合にも誤差の大きさの傾向は共通している．全ての場合を通して $\{q_{1_i}\}$の誤差が最も小さく，次に{ q_{2_i} } および{ q_{5_i} }が小さい．{ Q_i } が 0.03 より小さいところでは{ q_{3_i} }が最も大きな誤差を示している．{ q_{4_i} }は，Q_i が 0 に近いところでは{ q_{2_i} }および{ q_{1_i} } とほとんど同じ大きさの誤差を示すものの，{ Q_i } が 0.04 より大きいところでは，最も大きな誤差を示した．

図3　$q=0.2$，$n=6$，$\{q_i\}$を公比0.75の等比数列としたときのq_{1_i}～q_{5_i}の誤差

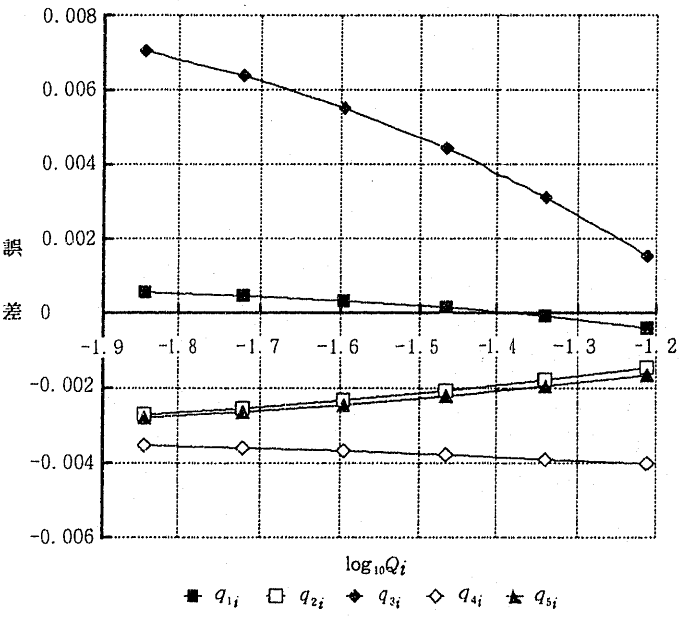
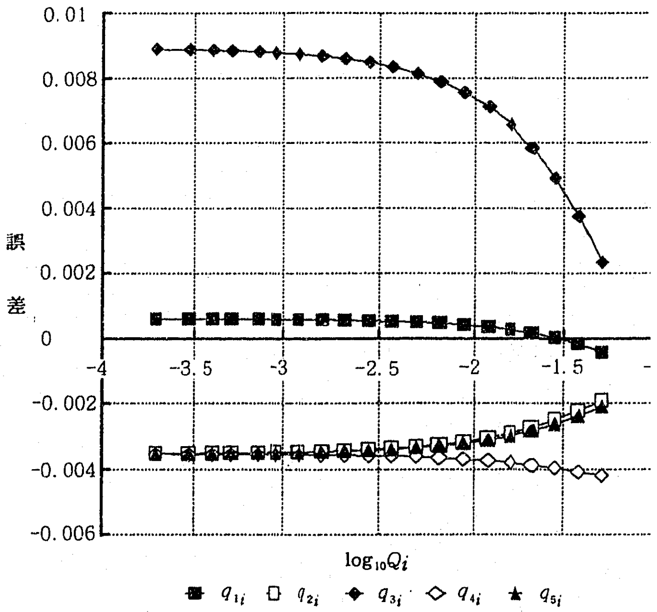


図4　$q=0.2$，$n=21$，$\{q_i\}$を公比0.75の等比数列としたときのq_{1_i}～q_{5_i}の誤差



4. 考　察

$\{q_{1_i}\}$～$\{q_{5_i}\}$は，$\{q_i\}$との間に明らかに一定の法則にしたがった差を生じている．しかしながら，これはそのときそのときの条件により微妙に異なっている．例えば，$\{q_{1_i}\}$の誤差を示す点を結んだ折れ線がQ_i軸と交わる点は条件によって前後に変動している．従って，これを簡単な直線などで補正するのは容易ではない．

それぞれの近似式についてみると，$\{q_{1_i}\}$は，全体を通して小さい誤差を示した．ただし，Q_iが0.1前後より大きいところでは$\{q_{2_i}\}$，$\{q_{3_i}\}$，$\{q_{5_i}\}$より大きな誤差を示した．$\{q_{2_i}\}$と$\{q_{5_i}\}$はほとんど同じ誤差を示した．

これらのことから，人間の生命表における死因別死亡公算についての問題を扱う場合は，1年間の期間においては0.1より十分に小さい場合が多いと考えられることから，$\{q_{1_i}\}$を用いるのが最も望ましいことが明らかとなった．計算の簡略化を望むのであれば，$\{q_{5_i}\}$は，きわめて単純な式により比較的誤差の少ない近似値をあたえるものといえる．しかしながら，死因別死亡確率が0.1を超えるような場合は，$\{q_{1_i}\}$はもはや最適な近似式とはいいがたい．このときは，$\{q_{2_i}\}$が最も適当であり，$\{q_{5_i}\}$がそれに準ずるといえる．いずれの近似式においても誤差の最大値は高々1％であった．

5. まとめ

本研究は，特定死因の死亡公算の真の値を算出する方法を示した．次に，これを用いてこれまで近似値どうしの比較によってのみでしか評価されていなかった特定死因の死亡公算の推定値算出方法に対して真の値と比較することによる評価を行なった．その結果，総死亡確率が0.2の場合，

(1)　いずれの近似式の誤差も高々1％以内である，

(2)　一番誤差の少ない近似値を与えることが多かったのは，(1)式によるものである，

(3)　死因別死亡確率が0.1を超える場合には，(2)式の誤差が最も小さい，

(4)　計算が簡単であるにもかかわらず，比較的よい近似を示したのは(5)式である，

ことが明らかとなった．