研究ノート

特定死因を除いた場合の死亡確率計算に関する考察

大場　　保

1. はじめに

特定死因を除去した場合の死亡確率を計算する際には死因別死亡公算（後述）を算出する必要があるが，これには Greville[1]の示した計算式を用いる場合（以下，Greville 法と略す）が多い．一方，Greville 法は近似式であって理論的真値を求めるものではないが，長い間その評価はされなかった．32年後にこの方法の評価を試みたのは Johnson and Johnson[2]である．彼らは，Greville 法，全死因を2死因に分類した場合に理論的真値として求まる死亡公算を多死因の場合の推定値とする方法（以下，ここでは2分法と略す），ほか2方法を比べ，全死因による死亡率が小さいときはいずれの方法も同様の値を与えることを明らかにした．彼らが比較したものは，いずれも死因数に依存しない推定式であるが，これらが死因数の変化に伴い真の値とどう異なるのかは全く明らかにされていない．

そこで本研究では，死因の数を増やした場合の真の死亡公算に対する，見かけの死因別死亡率[3]の違い，Greville 法及び2分法によって推定される死亡公算の誤差についての評価を行ない，各方法の長所，短所を明らかにすることを試みた．

2. 方　法

はじめに，死因別死亡公算の考え方について説明しておこう。簡単のため，各死因は確率的に独立であるとする（仮定Ⅰ）．即ち，ある死因による死亡確率が何らかの理由で変化しても他の死因の死亡確率には変化を与えないということである．これをわかりやすくたとえれば，「ある人のある期間の生死を決定する神様は，天国で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．期間の初めにすべてのクジを引き，ひとつも当たらなければ生残，複数の当たりクジが出た場合は日付の早い死因で死亡する．」ということである．また，各死因の当たりクジに書かれた日付の出現頻度は，期間中一定であるとする（仮定Ⅱ）．このように考えると，m個の当たりクジが出た場合は，$\frac{1}{m}$の確率で各当たりクジ死亡で死亡することになる．

従って，生残確率pは，

(1)　$p = p_1 \cdot p_2 \cdot \cdots \cdot p_n$

1) T. N. E. Greville, "Mortality Tables Analyzed by Cause of Death", *Rec. Am. Inst. Actuar.*, 37 (1948), pp.283−294.

2) R. C. Elandt-Johnson and N. L. Johnson, *Survival Models and Data Analysis* (New York: Wiley-Interscience, 1980), pp.312−317.

3) 例えば，ある集団の観察期間初期人口数が100で，うち50人が観察期間中に交通事故により死亡，10人が癌により死亡し，他の死因による死亡はなかったとしよう．この場合，死因別死亡確率はそれぞれ0.5，0.1であるが，癌が特効薬の発明により死ぬ病でなくなったとすると，癌で死ぬことになっていた10人は交通事故でも死ぬことはなかったろうか．おそらく10人中5人位は交通事故により死亡するであろう．従って交通事故により死亡する危険性（死亡公算）は0.55程度，癌のそれは同様に0.15程度だったと考えられる．即ち，死因別死亡数としては現われてこない，各死因間の重複部分に隠蔽された0.55−0.5=0.05，0.15−0.1=0.05という死亡確率があったと考えられる．

$= (1-q_1)\cdot(1-q_2)\cdot\cdots\cdot(1-q_n)$

として求まる．ここで，n は死因数，p_i 及び q_i は i 番目の死因による生残確率，死亡確率（クジの当たる確率）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確率 Q_i については，

(2) $p = 1-q$

$= 1-(Q_1+Q_2+\cdots+Q_n)$

という関係が成立する．q は全死因による死亡確率である．

ここで q_i と Q_i がまぎらわしいため，水島[4]にならって，以下 q_i を死因 i の死亡公算，同様に p_i（$=1-q_i$）を生残公算と呼ぶこととする．

次に，q_i と Q_i の関係であるが，$f(i, m)$ を死因 i を含み m 個のクジが当たる確率とすると，仮定IIを考慮して，

(3) $Q_i = f(i,1)+\dfrac{1}{2}f(i,2)+\cdots+\dfrac{1}{n}f(i,n)$

と表せる．$f(i, m)$ は，

(4) $1=\prod_{j=1}^{n}(p_j+q_j)$

の右辺を展開したときの q_i を含み m 個の q_j を含んだ項の総和であるから，

(5) $f(i,m)=p_1\cdot p_2\cdot\cdots\cdot p_n\cdot\dfrac{q_i}{p_i}\cdot\sum_{a=1}^{n-m+1}\left(\dfrac{q_{i_a}}{p_{i_a}}\cdot\sum_{b=a+1}^{n-m+2}\left(\dfrac{q_{i_b}}{p_{i_b}}\cdot\cdots\cdot\sum_{y=x+1}^{n-2}\left(\dfrac{q_{i_y}}{p_{i_y}}\cdot\sum_{z=y+1}^{n-1}\cdot\dfrac{q_{i_z}}{p_{i_z}}\right)\cdots\right)\right)$

である．ここで q_{i_k} は，n 個の q_j の列

$q_1,\ q_2,\ \cdots,\ q_{i-1},\ q_i,\ q_{i+1},\ \cdots,\ q_n$

から q_i を除いた列

$q_1,\ q_2,\ \cdots,\ q_{i-1},\ q_{i+1},\ \cdots,\ q_n$

に順番をつけなおした列

$q_{i_1},\ q_{i_2},\ \cdots,\ q_{i_k},\ \cdots,\ q_{i_{n-1}}$

の第 k 項である．

死因数 n が 4 以下であれば，(3)式の n 個の連立方程式を解き，観測値から得られた Q_i から q_i を求めることも可能であり，Johnson and Johnson もこの方法によって $n=2$ のときの値を求めている．しかしながら n が 5 以上の場合は，Q_i から q_i を代数学的には求められない．ここでは，精度を評価するためのモデル計算であるから，一定の条件の下に q_i を設定して Q_i を求めた．

次に，Greville 法について説明する．これは，

(6) $p_i' = p^{\frac{Q_i}{q}}$，$q_i' = 1-p_i'$

として求めるものである．ここで p_i' は Greville 法による死因 i の生残公算，q_i' も同様である[5]．

本研究では，(i) n 個の q_i を設定し，(ii) (3)式及び(5)式より Q_i を算出，(iii) Q_i より q_i' を算出，(iv) Q_i より 2 分法による死亡公算 q_{2i} を算出し，比較を行った．ただし，q_{2i} は $n=2$ として(3)式の 2 次連立方程式を解き，

(7) $q_{2_i}=(Q_i-Q_{-i}+2-\sqrt{(Q_i-Q_{-i}+2)^2-8Q_i})/2$ として求まるものである．（$Q_{-i}=q$

4) 水島治夫，『生命表の研究』，生命保険文化研究所，1963．

5) Greville の論文では，Q_i ではなく，期間の総死亡数 D 及び死因別死亡数 D_i を用いて $p_i' = p^{\frac{D_i}{D}}$ としている．

$-Q_i$ である．）

q_i の設定は以下のようにした．

1．一つの死因に注目し，その死亡確率が他の死因の死亡確率を含めた全体の死亡確率に占める割合の変化及び死因数の変化に従って誤差がどう変化するかを調べるために，

$$\begin{cases} q = 0.2, \\ Q_1 \neq Q_2 = Q_3 = \cdots = Q_n, \\ Q_1 : Q_{-1}\left(= \sum_{i=2}^{n} Q_i \right) = \frac{1}{2},\ 1,\ 2, \\ n = 6,\ 11,\ 16 \end{cases}$$

として反復計算により q_i を定めた．

2．各死因の死亡確率がすべて異なった場合に，n が増えるにつれて誤差がどう変化するかを調べるため，

$$\begin{cases} q = 0.2, \\ \{ q_i \} \text{ は公比0.75の等比数列}, \\ n = 5,\ 10,\ 20 \end{cases}$$

として1.同様反復計算した．

3．上記1. 2.について q を増減した．

計算は，8087付のPC－9801m2によって行い，C言語によりプログラムを作成し，すべて倍精度実数を用いた．

3．結　果

1．$Q_1 : Q_{-1} = \frac{1}{2}$，1，2の場合の結果を表1に示した．$q_i$ に対する Q_i，q_i'，q_{2_i} の誤差を示すため，それぞれを q_i で除してさらに1を減じて示してある．これは2.でも同様である．

まず，Q_i についてみると，q_i に対して－9.99～－3.61％の誤差が出た．この絶対値は n が増える

表1．$Q_1 : Q_{-1} = 0.5,\ 1,\ 2,\ n = 6,\ 11,\ 16$ とした場合の q_i に対する Q_i，q_i'，q_{2_i} の誤差

$Q_1 : Q_{-1}$	n	i	q_i	Q_i	Q_i/q_i-1	q_i'	q_i'/q_i-1	q_i	q_{2_i}/q_i-1
0.5	6	1	0.0717	0.0667	-7.05E-2	0.0717	-5.33E-4	0.0716	-1 42E-3
		2- 6	0.0293	0.0267	-9.01E-2	0.0293	2.73E-4	0.0292	-2.32E-3
	11	1	0.0717	0.0667	-7.06E-2	0.0717	-7.10E-4	0.0716	-1.60E-3
		2-11	0.0148	0.0133	-9.67E-2	0.0148	3.66E-4	0.0147	-2.95E-3
	16	1	0.0717	0.0667	-7.07E-2	0.0717	-7.69E-4	0.0716	-1.66E-3
		2-16	0.0099	0.0089	-9.89E-2	0.0099	3.98E-4	0.0098	-3.18E-3
1.0	6	1	0.1057	0.1000	-5.35E-2	0.1056	-7.84E-4	0.1056	-7.84E-4
		2- 6	0.0220	0.0200	-9.29E-2	0.0221	8.21E-4	0.0220	-2.13E-3
	11	1	0.1057	0.1000	-5.36E-2	0.1056	-8.82E-4	0.1056	-8.82E-4
		2-11	0.0111	0.0100	-9.79E-2	0.0111	9.29E-4	0.0111	-2.59E-3
	16	1	0.1057	0.1000	-5.37E-2	0.1056	-9.14E-4	0.1056	-9.14E-4
		2-16	0.0074	0.0067	-9.95E-2	0.0074	9.65E-4	0.0074	-2.75E-3
2.0	6	1	0.1383	0.1333	-3.61E-2	0.1382	-7.68E-4	0.1383	-3.41E-4
		2- 6	0.0147	0.0133	-9.55E-2	0.0148	1.65E-3	0.0147	-1.68E-3
	11	1	0.1383	0.1333	-3.62E-2	0.1382	-8.11E-4	0.1383	-3.84E-4
		2-11	0.0074	0.0067	-9.88E-2	0.0074	1.75E-3	0.0074	-1.97E-3
	16	1	0.1383	0.1333	-3.62E-2	0.1382	-8.25E-4	0.1383	-3.98E-4
		2-16	0.0049	0.0044	-9.99E-2	0.0049	1.78E-3	0.0049	-2.08E-3

備考）7.05E-2は 7.05×10^{-2} を表す．

とともに増加した．また，Q_1 の誤差は，Q_1 が q（$=\sum Q_i$）に占める割合が $\frac{1}{3}$，$\frac{1}{2}$，$\frac{2}{3}$（Q_1：$Q_{-1}=\frac{1}{2}$，1，2）と増えるに従い減少し，$Q_2 \sim Q_n$ でも同様であった．

次にq_i' と q_i の比較であるが，誤差の幅は－0.0914～＋0.165％であった．また，q_1'，q_2'，～ q_n' いずれも n が増加するにつれて誤差が増加した．q_1'の誤差は負であり，一方，$q_2' \sim q_n'$ のそれは正であった．

q_{2_i} はすべての負の誤差を示し，その幅は－0.318～－0.0341％であった．これを q_i' と比べると同一であったのは，Q_1：$Q_{-1}=1$，$i=1$のときであり，q_{2_i}の方が小さい誤差を示したのは，Q_1：$Q_{-1}=\frac{1}{2}$，$i=1$のときのみであった．その他の場合はすべて q_{2_i} の方が大きい誤差を示した．しかしながら，2つの誤差の比（$q_i'/q-1$）/（$q_{2_i}/q-1$）をとってみると，n が6，11，16と増加するに従い－0.118，－0.124，－0.125（Q_1：$Q_{-1}=\frac{1}{2}$，$i=2 \sim n$のとき）となり，誤差の絶対値は近づく傾向がみられた．

表2　{q_i}を公比0.75の等比数列として与えた場合の q_i に対する Q_i，q_i'，q_{2_i} の誤差

n	i	q_i	Q_i	Q_i/q_i-1	q_i'	q_i'/q_i-1	q_i	q_{2_i}/q_i-1
5	1	0.0713	0.0662	-7.06E-2	0.0713	-3.92E-4	0.0712	-1.29E-3
	2	0.0535	0.0492	-7.89E-2	0.0535	-5.24E-5	0.0534	-1.60E-3
	3	0.0401	0.0367	-8.51E-2	0.0401	2.02E-4	0.0400	-1.90E-3
	4	0.0301	0.0274	-8.96E-2	0.0301	3.94E-4	0.0300	-2.16E-3
	5	0.0226	0.0205	-9.30E-2	0.0226	5.37E-4	0.0225	-2.39E-3
10	1	0.0581	0.0536	-7.71E-2	0.0580	-4.29E-4	0.0580	-1.80E-3
	2	0.0436	0.0399	-8.38E-2	0.0435	-1.55E-4	0.0435	-2.10E-3
	3	0.0327	0.0298	-8.87E-2	0.0327	5.13E-5	0.0326	-2.38E-3
	4	0.0245	0.0222	-9.24E-2	0.0245	2.06E-4	0.0244	-2.62E-3
	5	0.0184	0.0166	-9.51E-2	0.0184	3.22E-4	0.0183	-2.81E-3
	6	0.0138	0.0124	-9.71E-2	0.0138	4.09E-4	0.0137	-2.96E-3
	7	0.0103	0.0093	-9.86E-2	0.0103	4.74E-4	0.0103	-3.08E-3
	8	0.0078	0.0070	-9.98E-2	0.0078	5.23E-4	0.0077	-3.17E-3
	9	0.0058	0.0052	-1.01E-1	0.0058	5.59E-4	0.0058	-3.24E-3
	10	0.0044	0.0039	-1.01E-1	0.0044	5.87E-4	0.0043	-3.30E-3
20	1	0.0551	0.0507	-7.85E-2	0.0550	-4.37E-4	0.0550	-1.92E-3
	2	0.0413	0.0378	-8.48E-2	0.0413	-1.78E-4	0.0412	-2.22E-3
	3	0.0310	0.0282	-8.95E-2	0.0310	1.73E-5	0.0309	-2.50E-3
	4	0.0232	0.0211	-9.30E-2	0.0232	1.64E-4	0.0232	-2.72E-3
	5	0.0174	0.0158	-9.56E-2	0.0174	2.73E-4	0.0174	-2.91E-3
	6	0.0131	0.0118	-9.75E-2	0.0131	3.55E-4	0.0130	-3.05E-3
	7	0.0098	0.0088	-9.89E-2	0.0098	4.17E-4	0.0098	-3.17E-3
	8	0.0074	0.0066	-1.00E-1	0.0074	4.63E-4	0.0073	-3.26E-3
	9	0.0055	0.0050	-1.01E-1	0.0055	4.98E-4	0.0055	-3.32E-3
	10	0.0041	0.0037	-1.01E-1	0.0041	5.24E-4	0.0041	-3.37E-3
	11	0.0031	0.0028	-1.02E-1	0.0031	5.44E-4	0.0031	-3.41E-3
	12	0.0023	0.0021	-1.02E-1	0.0023	5.58E-4	0.0023	-3.44E-3
	13	0.0017	0.0016	-1.02E-1	0.0017	5.69E-4	0.0017	-3.46E-3
	14	0.0013	0.0012	-1.03E-1	0.0013	5.78E-4	0.0013	-3.48E-3
	15	0.0010	0.0009	-1.03E-1	0.0010	5.84E-4	0.0010	-3.49E-3
	16	0.0007	0.0007	-1.03E-1	0.0007	5.88E-4	0.0007	-3.50E-3
	17	0.0006	0.0005	-1.03E-1	0.0006	5.92E-4	0.0006	-3.51E-3
	18	0.0004	0.0004	-1.03E-1	0.0004	5.95E-4	0.0004	-3.51E-3
	19	0.0003	0.0003	-1.03E-1	0.0003	5.96E-4	0.0003	-3.52E-3
	20	0.0002	0.0002	-1.03E-1	0.0002	5.98E-4	0.0002	-3.52E-3

備考）7.05E-2は 7.05×10^{-2} を表わす．

2.　Q_i，q_i'，q_{2_i}のすべてにおいて，nの増加とともに誤差の増加がみられた（表２）．

Q_iについてみれば，誤差は$-7.06\sim-10.3$％の範囲であり，Q_iが小さくなるに従い誤差は増大した．

q_i'では，誤差は$-0.0437\sim+0.0598$％の範囲であり，q_1'とq_2'が正，q_3'以後は正であった。$q_3'\sim q_n'$ではQ_iが小さくなるほど誤差は大きくなった．

q_{2_i}は，1.の場合と同様すべて負であり，$-0.352\sim0.129$％の範囲であった.各nの中では,Q_iが小さくなるほど誤差が大きくなった．また，すべてのiにおいてq_i'より誤差が大きかった．

3.　簡単のため結果の表は省略する．1.及び2.で見いだされた誤差の傾向は変わらなかった．一方，Q_i，q_i'，q_{2i}のいずれにおいても，qが大きくなるに従い誤差が増大し，逆に小さくなれば著しく減少した．例えば，1.の条件でQ_1：$Q_{-1}=2$，$n=6$のとき，$q=0.1$，0.2，0.5と増加させると,Q_1の誤差は-3.42，-7.05，-19.6％と増大し，q_i'でも-0.0121，-0.0533，-0.473％，q_{2_i}では-0.0323，-0.142，-1.26％となった．

4. 考　察

Q_iがすべて負の誤差を示したのは，複数の当たりクジが出た場合に，日付が２番目以後のものは隠れてしまうことによる．q_{2_i}がすべて負の誤差であったのも同じ理由によるものである．即ち，１つの死因とその他の死因に分けた場合，その他でまとめられた各死因間の重複部分が隠される．従ってqに占める割合の大きい死因ほど，その他の部分に隠される部分が小さくなるため誤差が小さい．

さて，Johnson and Johnson は死因数が２のとき，(7)式で求めた死亡公算と Greville法の比較を行ったが，本研究の結果によると２より多い死因を，ある特定のものとその他に分けて特定死因の死亡公算を２分法で求めた場合の誤差はGreville法よりほとんどの場合において大きいことが判明した．実際に２分法が用いられるのは，死因の数が全部で２である場合はほとんどなく，多くの死因の中でのある特定死因とその他の死因という分類を行った場合であろう．q_{2_i}の誤差がq_i'のそれと等しかったのは，1.でQ_1：$Q_{-1}=1$，$i=1$のときであり，これは$Q_1=Q_{-1}=\frac{q}{2}$を(6)式及び(7)式に代入すれば$q_1'=q_{2_1}=1-\sqrt{p}$となるからである．$Q_1=\frac{2}{3}q>\frac{q}{2}$，$i=1$のときには$q_i'$の誤差の方が$q_{2_i}$のそれより大きかった．

これらのことを考慮・類推すると，$Q_i\leq\frac{q}{2}$ならばq_iをq_i'で推定し，$Q_i>\frac{q}{2}$ならばq_iをq_{2_i}で推定するのが望ましい．多くの場合は$Q_i\leq\frac{q}{2}$であるので，特定死因を除去した場合の死亡確率[6)]を推定するにはGreville法が適当であろう．２分法が有利な場合としては，特定死因以外の死因が除かれた場合の死亡確率を推定する場合が考えられる．

また，Greville 法の利点は計算の簡便さにもある．死因数が20程度であれば，本研究で用いた反復法の応用によってQ_jからq_iを求めるために要す時間は数時間であろうが，死因が１つ増えれば計算時間は約２倍になるため，それ以上の死因数に対応することは困難である．従って，死亡頻度の大きいものから20死因までとその他の死因という程度で済むのであれば反復法の応用により死亡公算を求めることが可能であり，最も望ましい手段であるが，死因数がさらに多い場合や簡便さが必要とされるならば，Greville 法によって推定するのがよいだろう．

6）特定死因iを除去した場合の死亡確率q_{-i}は，死亡公算q_iを用いて，

$$q_{-i}=1-p_{-i}=1-\frac{p}{1-q_i}$$

として求められる．

5. 結　論

本研究は，特定死因を除去した場合の死亡確率計算法を評価するために，全死因による死亡率 q を一定として，死因数 n が2より大きい場合について，各死因の死亡公算 q_i，死亡率 Q_i，Greville 法にによって推定される死亡公算 q_i'，2分法によって推定される死亡公算 q_{2_i} についてモデル計算による比較を行った．その結果，

1. $n \leq 20$程度ならば，反復法の応用により Q_i から q_i を求めてしまう方法が望ましい．
2. $Q_i \leq \frac{q}{2}$ ならば，q_i' の方が q_{2_i} よりも小さい誤差を示した．従って，$n > 20$あるいは計算に簡便さが要求される場合には，Greville 法の方が2分法より望ましい．
3. 上記2.の条件の下で，逆に特定死因以外の死因がまとめて除かれたとした場合の死亡確率を求めるには，2分法が望ましい．

ことが明らかとなった．