都道府県別生命表死亡率のワイブル分布へのあてはめ

石井(イシイ)　太(フトシ)*

目的　低年齢死亡率のワイブル分布への適合状況に関して完全生命表を用いて評価を行い，次に，都道府県別生命表作成の観点から，実際のデータを用いてワイブル分布のパラメータを得る方法について検討する。

方法　完全生命表のデータを用いて，年齢の対数値と死力の対数値が，いかなる年齢範囲において線形関係にあるか観察し，その後，パラメータ推定を行ったワイブル分布関数値を用いた死亡率とオリジナルの死亡率を比較する。次に，人口規模の異なる幾つかの県について，この方法を用いて実際に死亡率推定を行い，死亡率の動きについて検討する。

結果　完全生命表のデータを用いて，年齢の対数値と死力の対数値が，いかなる年齢範囲において線形関係にあるか観察したところ，おおむね10歳程度までについて，両者の間に線形関係が認められた。また，パラメータ推定を行ったワイブル分布関数値による死亡率とオリジナルの死亡率を比較したところ，10歳までの範囲においてほぼオリジナルの死亡率を再現することができていることが確認された。人口規模の異なる幾つかの県について，この方法を用いて実際に死亡率推定を行ったところ，ワイブル分布によるあてはめを行った場合の方が，より安定した死亡率の動きを示すことが確認された。

結論　都道府県別生命表作成における低年齢死亡率推定については，従来の方法と比較して，ワイブル分布関数を用いたパラメータ推定による方法がより安定性が高く，好ましいものであることが結論づけられる。

キーワード　生命表，都道府県，死亡率，ワイブル分布，パラメータ推定

Ⅰ　緒　　言

厚生労働省において5年ごとに作成してきている都道府県別生命表では，対象となる人口集団が全国単位の生命表に比較して小さい。このため，死亡率を作成するにあたり，安定性を確保する観点から幾つかの工夫がなされてきている。第一に，国勢調査年の前後年を含め，3年分の死亡データを用いていることである。第二に，5歳以上の死亡率作成にあたっては，いったん5歳階級ごとにまとめた粗死亡率を作成し，その後，死力が4次多項式で近似できるとの仮定に基づき，前後の階級値を利用した補間法を用いて各歳に分解する手法を用いて推定を行っていることである。しかし，5歳未満の死亡率については，死力の様相が5歳以上とは異なると考えられることなどから，各歳単位（0歳はより細分化されている）で死亡率推定を行ってきている[1)]。ところが，3年分の死亡データを用いているとはいえ，都道府県別では人口規模が小さいため，各歳単位での死亡率は安定性に欠ける面があることは否めない。

そこで，平成12年都道府県別生命表の作成にあたっては，2～4歳の死亡率作成に際し，低

*厚生労働省大臣官房統計情報部企画課審査解析室長補佐

年齢の死亡の挙動が，寿命データをよく説明する分布であるワイブル分布に適合するものとして行った。本稿では，低年齢死亡率のワイブル分布への適合状況に関して完全生命表を用いて評価を行い，次に，都道府県別生命表作成の観点から，実際のデータを用いてワイブル分布のパラメータを得る方法に関しての考え方とその結果について述べる。

II 方　　法

死亡の年齢別パターンを関数によって近似することについては各種の研究や応用例がある。例えば，完全生命表など，厚生労働省で作成している各種生命表の高齢部の死亡率推定においては，死力がGomperts-Makehamの法則に従うものとし，パラメータ推定を行うことにより算定を行っている[2]。具体的には，死力μ_xが，

$$\mu_x=\alpha+\beta e^{\gamma x}$$

という関数によってその年齢別パターンが表されるものと考え，実際のデータからこれらのパラメータ推定を行い，死亡率を推定しているわけである。このような研究の中で，死亡の年齢別パターンを「ワイブル分布」を用いて表現するものも見られる[3)4)]。

ワイブル分布とは，確率密度関数が，正のパラメータα，βを用い，

$$f(x)=\frac{\beta}{\alpha}\left(\frac{x}{\alpha}\right)^{\beta-1}exp\left(-\left(\frac{x}{\alpha}\right)^{\beta}\right)$$

で表される確率分布である（$x>0$）。ここで，αは尺度母数，βは形状母数とよばれ，それぞれ分布の広がり，分布の形状を表すパラメータとなっている。

いま，死力μ_xが，パラメータθ，cを用いて，

$$\mu_x=\frac{c}{\theta^c}x^{c-1}\quad(c>0)\quad\cdots\cdots\cdots\cdots\cdots\cdots\quad ①$$

と表されるとする。このとき，死亡の確率密度関数（＝生命表上の死亡数を連続化したもの）はワイブル分布で表される。このことを示すため，①の死力を積分し，x歳における生存率${}_xp_0$を求めると，

$$ {}_xp_0=exp\left(-\int_0^x\mu_t dt\right)=exp\left(-\int_0^x\frac{c}{\theta_c}t^{c-1}dt\right)$$

$$=exp\left\{-\left(\frac{x}{\theta}\right)^c\right\}\quad\cdots\cdots\cdots\cdots\cdots\cdots\quad ②$$

となる。そこで，これをマイナスしてxで微分すると，

$$-\frac{d_xp_0}{dx}=\frac{c}{\theta}\left(\frac{x}{\theta}\right)^{c-1}exp\left\{-\left(\frac{x}{\theta}\right)^c\right\}$$

という式が得られ，死亡の確率密度関数が，尺度母数をθ，形状母数をcとしたワイブル分布で表されることが示される[5)]。

さて，実際に死亡の確率密度関数をワイブル分布にあてはめるに際しては，「どのデータ範囲からどの関数にあてはめを行うのか」を決定する必要がある。

まず，関数のあてはめ範囲を考察するため，完全生命表の死力データの観察を行う。①式の両辺の対数をとることにより，

$$log\mu_x=log\frac{c}{\theta^c}+(c-1)\,logx$$

という関係が得られる。これから，一本のワイブル分布関数で死力が表される範囲内では，年齢の対数値である$logx$と，死力の対数値である$log\mu_x$は線形関係にあることになる。そこで，完全生命表（第17～19回）を用い，いかなる年齢範囲について，ワイブル分布での近似が行えるかを検討する。検討の結果は「III　結果」において示す。

一方，実際のあてはめにあたっては，死力を直接得ることはできないため，${}_nq_x$を用いてワイブル分布のパラメータを推定することが必要となる。そこで，②式を利用して実際のパラメータ推定を行う方法を考察する。

②式の両辺の対数をとって，両辺をマイナスし，さらに両辺の対数をとることにより，

$$log(-log_xp_0)=log\left(-log\ exp\left(-\left(\frac{x}{\theta}\right)^c\right)\right)$$

$$=clogx-clog\theta\quad\cdots\cdots\cdots\cdots\quad ③$$

という式が得られる。これにより，${}_nq_x$から$log(-log_xp_0)$を作成し，$logx$との回帰を行ってパラメータを推定することが可能であることがわかる。そこで，ワイブル分布への近似が行え

図1　年齢の対数値と死力の対数値の関係(第17～19回生命表)男

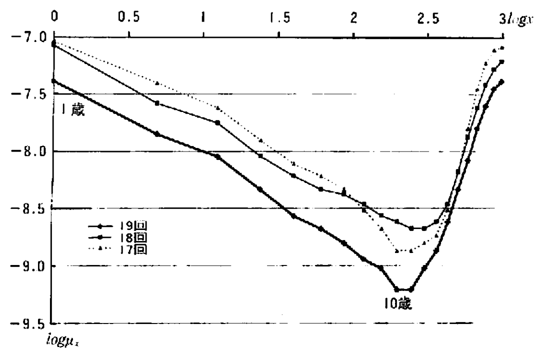
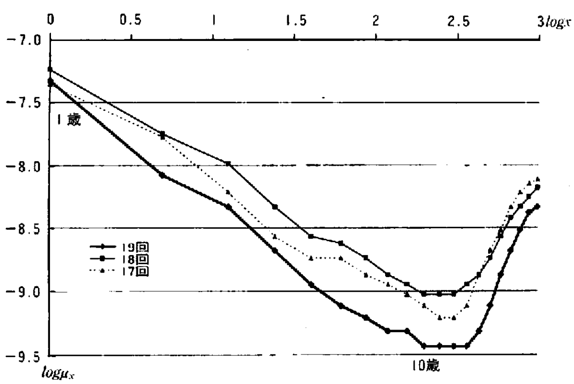


図2　年齢の対数値と死力の対数値の関係(第17～19回生命表)女



る範囲と，都道府県別生命表作成に使用可能なデータとの関係から，あてはめに使用するデータ等を考慮した。この方法に基づく関数あてはめの適合度に関しては，完全生命表（第17～19回）を用いて同様のデータを作成し，オリジナルの死亡率がどこまで再現できるかについて確認を行う。この結果は「III　結果」において示す。

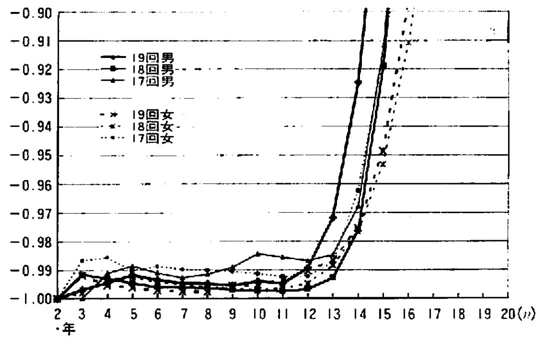
最後に，実際の都道府県別の死亡率データを用い，ワイブル分布によるあてはめを用いた場合の死亡率を，用いない場合のそれと比較することとする。

III　結　　果

最初に，ワイブル分布のあてはめ範囲を検討するため，年齢の対数値$logx$と死力の対数値$log\mu_x$が，いかなる年齢範囲において線形関係にあるか，完全生命表（第17～19回）のデータを用いてグラフにしたものが，図1，2である。

このグラフを見ると，男女ともおおむね10歳程度までについて，両者の間に線形関係が認められる。これをさらに詳しくみるため，1歳～n歳に関して，年齢の対数値と死力の対数値の相関係数をとったものが図3である。これによれば，相関係数は10歳前後までは−0.98未満の値をとっているが，10歳を超えると急激に増加していくことがわかる。都道府県別生命表では5歳以上について5歳階級で粗死亡率を算定していることも考えあわせれば，10歳を一つの区切りとし，そこまでが一本のワイブル関数であて

図3　1歳～n歳までの年齢対数値と死力対数値の相関係数



はめられると考えてパラメータ推定を行うことが望ましいと考えられる。

さて，以上から，あてはめの範囲に関しては，10歳までのデータを用いることとなったわけであるが，死力は実際のデータから直接得られないため，実際のデータからパラメータ推定を行う方法の確立が必要となる。

「II　方法」において，③という形で${}_nq_x$から$log(-log_xp_0)$を作成し，$logx$との回帰を行ってパラメータを推定する方法を述べた。そこで，実際に得られるデータとの関係から${}_1p_0$，${}_1p_1$，${}_3p_2$，${}_5p_5$を作成し，パラメータ推定を行ってワイブル分布関数にあてはめを行うことを考える。

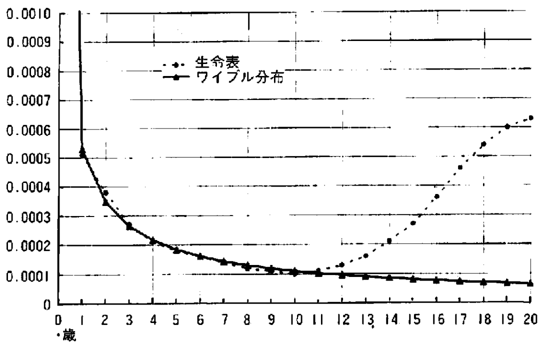
この手法による死亡率のあてはまり度合を検証するため，まず，完全生命表を用いて，この手法によるパラメータ推定を行うことにする。

表1は，第19回生命表（男）のデータによる基礎数値を表したものである。${}_1p_0$，${}_1p_1$，${}_3p_2$，${}_5p_5$に基づいて$logx$，$log(-log_xp_0)$が算定される。これをグラフにしたものが，図4である。

表1　第19回生命表（男）による基礎数値

x	${}_{n}p_{x-n}$		${}_{x}p_0$		$logx$	$log(-log({}_{x}p_0))$
1	${}_{1}p_0$	0.99655	${}_{1}p_0$	0.99655	0.000	−5.668
2	${}_{1}p_1$	0.99949	${}_{2}p_0$	0.99604	0.693	−5.530
5	${}_{3}p_2$	0.99914	${}_{5}p_0$	0.99519	1.609	−5.334
10	${}_{5}p_5$	0.99929	${}_{10}p_0$	0.99448	2.303	−5.196

図5　第19回生命表（男）死亡率のワイブル分布へのあてはめ



これでみてとれるように，両者はこの範囲内でほぼ線形関係にあり，決定係数も0.9997と非常に高いものとなっている。この結果から，パラメータについては，$\theta=9.00\times10^{11}$，$c=0.206$と推定がされる。

この推定されたパラメータによるワイブル分布関数値を用いた死亡率と，オリジナルの死亡率を比較したものが，図5である。これを見ると，10歳までの範囲においてほぼオリジナルの死亡率を再現することができていることが確認される。

このような検討を経て，今回の都道府県別生命表の2～4歳における死亡率の算定にあたっては，各都道府県等における${}_{1}p_0$，${}_{1}p_1$，${}_{3}p_2$，${}_{5}p_5$のデータを基に②式を利用し，最小二乗法によるパラメータ推定を行ったところである。

そこで，次に，実際の都道府県別データにあてはめた場合の結果をみてみよう。ここでは，人口規模の異なる県をいくつかみてみるため，人口規模が比較的大きい埼玉県，中程度の鹿児島県，比較的小さい島根県の3県を例にとり，男の死亡率の推定状況についてみてみることとする。

人口規模が比較的大きい埼玉県について，死亡率の推定状況をグラフにしたものが，図6で

図4　$logx$と$log(-log({}_{x}p_0))$の関係

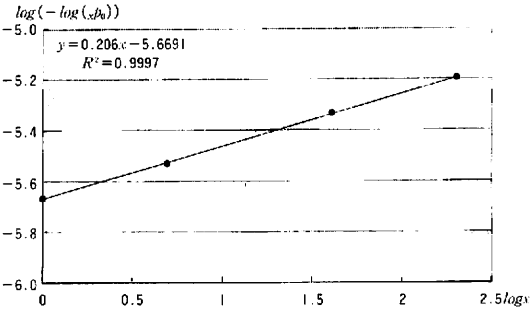
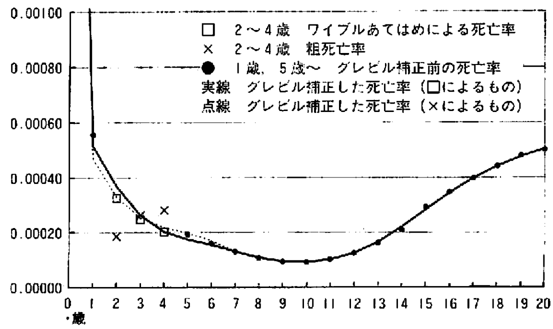


図6　ワイブルあてはめの有無による死亡率の比較　埼玉県　男



ある。2～4歳における粗死亡率は図の×印で示されているとおりであるが，2歳から4歳にかけて上昇する動きを示しており，このような比較的人口規模のある県において3年分の死亡データを基に死亡率を算出しても，各歳ごとに粗死亡率を見るとばらつきが生じてしまうことがわかる。一方，ワイブル分布のあてはめに基づく死亡率は□印で示されているものであり，データを3歳分まとめた効果と，前後の死亡率を用いて関数あてはめを行った効果があいまって，グレビル補正を行う前においても安定した動きに補正されていることがわかる。平成12年都道府県別生命表では，階級と階級の間のギャップを埋める観点からグレビル補正を行っており，実際の生命表に使われた死亡率は●と□のデータにグレビル補正を行った実線となっている。一方，点線で示したものは，ワイブルあてはめを行わないデータである×を用いてグレビル補正を行ったものである。グレビル補正のみでも平滑化効果は得られるが，粗死亡率の動きに影響され，1，2歳のデータがより低く，5，6歳のデータがより高く補正されてしまってい

図7　ワイブルあてはめの有無による死亡率の比較　鹿児島県　男

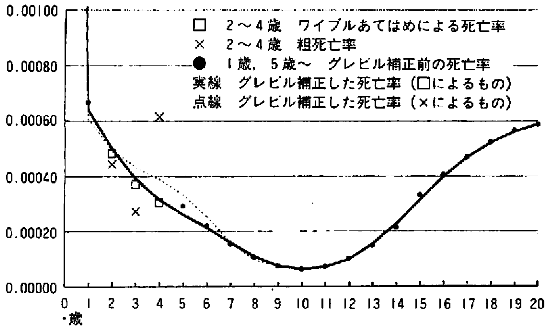
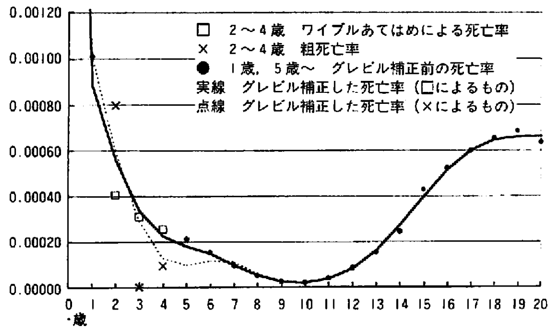


図8　ワイブルあてはめの有無による死亡率の比較　島根県　男



る状況が現れている。

人口規模が中程度の鹿児島県でも，×印で示される粗死亡率を見ると4歳のデータがかなり高い値となっており，データのブレがみられる。一方，□印のあてはめデータはこのブレを修正し，前後との連続性がより強いものとなっている。ここでも，仮に2～4歳に各歳別の粗死亡率を使用してグレビル補正を行うと，4歳の死亡率がその前後の死亡率を高めに引き上げることにより自然な動きとならないことが点線のグラフからみてとれる（図7）。

人口規模の小さい島根県のデータでは粗死亡率のブレはかなり大きいものとなっているが，ワイブルあてはめに基づく死亡率は安定した動きを示している。ただ，1歳の死亡率がかなり高い値となっていることがグレビル補正にも影響を与え，実線データと□印に乖離はあるものの，粗死亡率を用いてグレビル補正した点線よりは自然な動きとなっており，ワイブル分布をあてはめたことによる効果が現れている（図8）。

Ⅳ 考　　察

本稿では，都道府県別生命表作成の観点から，ワイブル分布へのあてはめを用いたパラメータ推定を用いて低年齢死亡率を作成する方法について検討を行った。検討によって得られた結果をまとめると，以下のとおりである。

(1)　完全生命表（第17～19回）のデータを用いて，年齢の対数値 $log x$ と死力の対数値 $log \mu_x$ が，いかなる年齢範囲において線形関係にあるか観察したところ，おおむね10歳程度までについて，両者の間に線形関係が認められた。

(2)　$_nq_x$ から $log(-log_x p_0)$ を作成し，$log x$ との回帰を行ってパラメータを推定する方法に基づき，完全生命表の $_1p_0$，$_1p_1$，$_3p_2$，$_5p_5$ を用いて，パラメータ推定を行ったワイブル分布関数値を用いた死亡率と，オリジナルの死亡率を比較したところ，10歳までの範囲においてほぼオリジナルの死亡率を再現することができていることが確認された。

(3)　人口規模の異なる幾つかの県について，この方法を用いて実際に死亡率推定を行ったところ，ワイブル分布によるあてはめを行った場合の方が，より安定した死亡率の動きを示すことが確認された。

以上の検討結果に基づき，都道府県別生命表作成における低年齢死亡率推定については，従来の方法と比較してワイブル分布関数を用いたパラメータ推定による方法がより安定性が高く，好ましいものであるということが結論づけられる。

（本稿の中で意見にわたる部分は筆者の個人的見解である）

文　　献

1）厚生省大臣官房統計情報部．平成7年都道府県別生命表．1998.
2）厚生労働省大臣官房統計情報部．第19回生命表．2002.
3）古川俊之．寿命の数理．東京：朝倉書店，1996；72-93.
4）W. Nelson. 寿命データの解析．東京：日科技連出版社，1988；34-6.
5）Survival Models and Data Analysis. Regina C. Elandt-Johnson, Norman L. Johnson. New York：John Wiley & Sons, Inc., 1980；181-224.