

## 論文

ガウス回帰による各種がんの年齢別罹患率  
…… 福島県甲状腺がんの被曝罹患人口の下算を主題に ……平 井 孝 治\*  
松 本 祐 輔\*\*

## 要旨

原子力災害による被曝がんについて、チェルノブイリから 31 年経った現在でも、依然として明確にされていない。疫学界などでは、年齢に関し『離散表示』の年齢階級別罹患率で各種のがんに対処している。我々は、〈加重した残差平方〉を用いて、これを『連続表示』の年齢別罹患率に回帰することを試みた。

その結果、各種がんの罹患率がガウス型の年齢密度関数になることを解明し、年齢別に分析できるようにした。今まで調べた範囲では、実際値と回帰値の修正済み相関係数は「ガウス近似の閾値 0.9920」を超えている。このガウス型の関数近似を使えば、年齢別人口から〈生涯罹患人口〉も算定できる。

他方、福島県の甲状腺がんの年齢別罹患率は、移動平均で  $A_0(2) \sim A_0(19)$  まで既知であるが、20 歳以降は不明である。しかし、ガウス型の年齢密度関数を使い、論理的に「下算の限界」を整備すれば、右側のガウス関数が求められる。かくして、 $\lambda^{0.5}$  接続を用いて〈被曝に由来する甲状腺がん罹患人口〉が下算できた。

## キーワード

甲状腺がん先行検査、被曝時年齢、受診時年齢、年齢階級別罹患率、年齢別罹患率、ガウス回帰、罹患率の年齢密度関数、ガウス近似の閾値、超近似の閾値、生涯罹患、福島原発事故、被曝由来の罹患人口、 $\chi$ ope 制約、知見水準、重回帰のレベル区分

## 目 次

はじめに

§ 1 自然罹患率  $B_0(x)$  と  $C_0(x)$  のガウス回帰

§ 2 一般がん種の罹患率回帰のために

§ 3 福島の甲状腺がん、県民罹患率  $A(x)$  の解明§ 4 累積罹患  $A(k, n)$  と生涯罹患人口  $\Psi_A(y, z)$ 

おわりに

\* 元立命館大学経営学部教授

\*\* ソシオアート株式会社（在、三重県津市）代表取締役

## は じ め に

東京電力福島第一原発による原子力災害後の甲状腺がん罹患率と、自然界にもともと存在する罹患率との間に、有意な差があるか否か論争されている。疫学上、がんの罹患率は年間 10 万人あたり何件見つかるかで表記する習がある。そこでこの論文では、年齢  $x$  を用いて、前者のそれを「県民罹患率  $A_0(x)$ 」<sup>1)</sup>、後者のそれを「自然罹患率  $B_0(x)$ 」と表記する。

さらに、疫学界では平準化のため、0～4 歳から始め五歳刻みで、年齢階級別の年間罹患率を算定するしきたりがある。筆者らは、河宮信郎氏の主宰するゴフマン研究会で研究を進め<sup>2)</sup>、それを Gauss 型の年齢密度関数  $F(x)$  に回帰すると、実際値と回帰値の修正済み相関係数  $q$  が、調べた範囲では、「G 近似の閾値」である 0.9920 以上になることが判った。

この発見と計算手法の開発によって、後述する「Double Gauss 接続法」の洗い替え Aに準ずる手続きで県民罹患率  $A(x)$  を求め、事故当時  $k$  歳だった人の生涯罹患  $A(k, 85)$  が算定できるようになった。その結果、事故当時 (2011 年 3 月) 頃に生まれたばかりの  $k = 0$  歳児のうち、少なくとも 4.79% は罹患を免れず、それは自然罹患の 6.90 倍に達する。さらに、被曝由来の罹患人口<sup>3)</sup> が  $\Psi_{A-B}(0, 84) = 40,464$  人以上になることが判明した。

我々の発見と開発による『生涯罹患人口』の概念と算定法を適用すると、取り上げたがん種が何であれ、「コホート (地域など属性を特定した集団)」の年齢別人口  $P_1(x)$  と、当該がん種の年齢階級別罹患率さえ判れば、例えば、県民罹患人口と対応する自然罹患人口を計算し、その差分として被曝罹患人口を「可及的に少なく推定する下算」も容易にできる。

§ 1 自然罹患率  $B_0(x)$  と  $C_0(x)$  のガウス回帰

本論では、甲状腺がんの年齢階級別の年間罹患率を、例えば 5～9 歳の場合なら「7 歳における年間罹患率」とみなし、五歳刻みの中央年齢を用いて  $B_0(7)$  のように記すことにする。

次の図表 0 で、 $A_0(x)$  は福島県民の実際値、 $B_0(x)$  はがん研データ<sup>3A)</sup> の 2007 年と 2008 年の平均、 $C_0(x)$  はがん研データの 1998 年～2007 年の平均で、 $D_0(x)$  は説明用の Dummy の

1) 本論では、原発事故当時、福島県に住んでいた人々のその後の罹患率を「県民罹患率」と称し、 $A_0(x)$  と記す。そのデータソースは、福島県の『甲状腺先行検査』【平成 27 年追補版】である。

それから自然罹患率を控除した率を「被曝罹患率」と称する。

2) 元中京大学経済学部教授。資料入手は蔵田計成氏、理論は河宮氏、計算は筆者平井、年齢別罹患率の算定用アルゴリズムは筆者松本が、それぞれ主担で研究してきた。

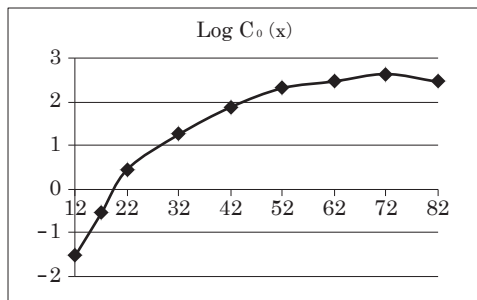
3) 福島県民の災害直近時の年齢別人口は、総務省統計局の『平成 22 年国勢調査』による。

3A) 「がん研データ」のデータソースは、国立がん研究センターの『Cancer Incidence』(1975-2008) で、男女別の母数が明記されてなかったから、単純平均で罹患率を算出した。

罹患率である。このうち、 $C_0(x)$  の  $x = 12 \sim 82$  歳の自然対数をグラフに示すと、図表 1 のようになる。

図表 0 罹患率の実際値（十万人あたり）

年齢 x	$A_0(x)$	$B_0(x)$	$C_0(x)$	$D_0(x)$
2	4.54E-05	4.54E-05	0.020	0.037
7	0.42	0.075	0.02	0.55
12	5.78	0.15	0.22	0.15
17	30.90	0.85	0.58	0.60
22		2.05	1.58	2.19
32		4.50	3.53	3.53
42		9.80	6.59	9.66
52		11.93	10.20	10.36
62		15.75	11.90	14.61
72		16.10	14.10	14.10
82		12.30	11.80	12.50
92				10.00



図表 1 がん研データの自然対数グラフ

図表 0-A 福島県の甲状腺がん「県民罹患率」

受診時 年齢	平準化 罹患率 $A_0(x)$	罹患 者数	算入 罹患 者数
2	0.00005		
3	0.00006		
4	0.00009		
5	0.00018		
6	0.44		
7	0.42		
8	0.40	1	1
9	1.93		
10	2.22		
11	3.95	4	4
12	5.78	1	1
13	10.96	6	6
14	12.63	5	5
15	16.76	14	14
16	22.08	8	8
17	30.90	11	11
18	36.79	17	13.6
19	45.85	22	13.2
20		19	7.6
21		8	1.6
歳	計	116	86.0

a) 「4.54E-05」は  $4.54 \times 10^{-5}$  の意で、本当は「0」であった。しかし、それでは逆数や対数が取れない。だが、世界中の 5 歳（約 2 億人）で 0.363 人は事実上 0 人である。

b)  $A_0(4)$  は  $A_0(2)$  と  $A_0(6)$  の、 $A_0(3)$  は  $A_0(2)$  と  $A_0(4)$  の、 $A_0(5)$  は  $A_0(4)$  と  $A_0(6)$  の調和平均である。

d) 表 0 の  $D_0(x)$  は Simulation 用の仮想の実際値。 $D_0(27) = 2.05$  が欠けている。

対数を取ると放物線のように見えるということは、正規分布のような Gauss 型関数に従うことを強く示唆している。ただし横軸は確率変数ではなく、年齢変数である点が大きく違うが、ピーク時年齢を  $\mu$ 、年齢偏差を  $\sigma$ 、年間罹患率のピーク値を  $\gamma$  で示すと、次のような数式で表現できるということになる。

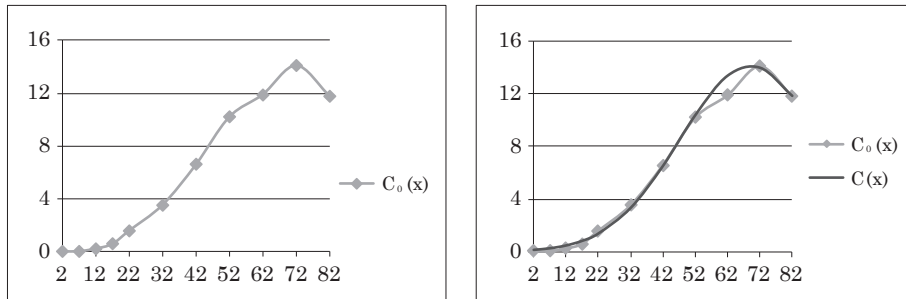
$$C(x) = \gamma \text{Exp} \{ -0.5 ((x - \mu) / \sigma)^2 \} \quad \cdots \cdots \textcircled{1}$$

この種のタイプは、実際値の対数を目的変数とし、年齢  $x$  と、 $x^2$  を説明変数として、重回帰するのが常道であるが、これを本論文では「対数回帰」ということにする。さすると、Exp 戻しで、修正済み相関係数が  $q = 0.9281$  となり、付録 I に示す重回帰レベル「代用 S」の 0.9160 にはいたるが、近似の閾値 0.9900 にはおよばない。なお、説明変数の数  $\chi$  に関する制約については、付録 II の脚注「 $\chi$  ope 制約の特約条項」を参照されたい。

そこで、年齢ごとに実際値と回帰値の残差平方を加重した和を最小化する『加重 G 回帰』によって、①式のピーク時年齢  $\mu$  と年齢偏差  $\sigma$  とピーク値  $\gamma$  を求めることにする。

すると、自然罹患率  $C_0(x)$  は次頁に示すフローチャートのわずか Step ② で決着がつき、回帰式は、 $C(x) = 14.10 \text{ Exp} \{-0.5 ((x-69.2)/21.8)^2\}$  ……② となった。この結果をグラフに示したのが、次の図表 2 である。

図表 2 甲状腺がんの自然罹患率  $C_0(x)$  (がん研データ, 1998 ~ 2007 年の平均)



このグラフから、 $C_0(62) = 11.90$  だけは外乱サンプルだが、図のようにガウス関数  $C(x)$  は  $C_0(x)$  を見事に回帰していることが解る。さらに、②式は、甲状腺がんの自然罹患率が、69.2 歳でピークを迎え、ピーク値が十万人あたり 14.10 人になることを示している。

ガウス回帰の場合、ピーク時年齢  $\mu$  と年齢偏差  $\sigma$  が説明変数で、ピーク値  $\gamma$  は定数項に相当するので、回帰の自由度 (説明変数の数) は  $\chi = 2$  ということになる。この場合、修正前の相関係数  $r = 0.9969$  を自由度で修正すると  $q = 0.9961$  となり、重回帰のレベルは筆者が定義する「超近似  $q \geq 0.9960$ 」に達している<sup>4)</sup>。後述するように、一般にがん罹患率は左関数と右関数の合成だが、この場合は、左右がほとんど重なっているとみなせる。

なお、通常重回帰では、実際値の平均＝回帰値の平均 や、全変動＝回帰変動＋残差変動 が成立するが、ガウス回帰では成立しない。それゆえ、説明変数の  $t$  値、 $p$  値 や 回帰 (モデル) 式全体の  $P$  値を論じても特段の意味は無い。

罹患率の回帰においては、 $\mu$  や  $\sigma$  は 10 倍した整数解で求めた後に 10 で割って少数第一位にするが、以下、これを本論では「整数解で少数第一位に丸める」と表現する。同様に、 $\gamma$  についても整数解で少数第二位に丸めることにする。なお、本論文の有効数字は 4 桁であるが、この丸めの影響による相対誤差は、最大で 0.0001 である。

4) 不自然でない程度に高めに計算することを「上算」というが、ここでは、 $C_0(62)$  を 11.90 人から 13.30 人と、自然罹患率を上算している。また、後述の  $B_0(62)$  も、17.75 人から 17.78 人に上算している。ただし、相関係数  $r$  は、元の実際値に戻して算定する必要がある。

上算の反対は下算だが、県民罹患率の方は、可及的に小さくなるよう「下算」に努めた。

### Double Gauss 接続法のフローチャート

修正済み相関係数  $q$  は少数第四位まで求め、その少数 4 桁を「Point」と称する。

実際値と回帰値の修正前相関係数  $r$  を、 $q^2 = 1 - (n-1)(1-r^2)/(n-1-\chi)$  で修正。

ガウス回帰の場合、説明変数の数は  $\chi = 2$  である。以下、 $FF(x)$  は 仮関数 の意。

- ①  $2 \sim X_0$ （元々の最高齢）で 加重 G 回帰して、 $FF_{右}(x)$  とする。
- ② 修正済みで  $q \geq 0.9960$ （超近似）なら、 $F(x) := FF_{右}(x)$  として End。 ex. C(x)  
 $q < 0.9800$  なら、Halt。

どちらでもない場合は、 $x = X_0$  のまま 事前手続き に進む。

事前手続き

- ③  $2 \sim x$ （現時点の最高齢）まで 単純 G 回帰して、 $FF_{左}(x)$  とする。
- ④ 左関数と右関数が 2 点で交わり、かつ、二つの回帰値の  
 乖離幅が  $0.05 \gamma_{右}$  以上の内包 Sample が一つでもあれば ……\* ex. B(x) at 52 歳  
 最も左内包サンプルのすぐ左隣りサンプルの年齢を「左端年齢  $X_{左}$ 」とし、最も右内包サン  
 プルのすぐ右隣りサンプルの年齢を「右端年齢  $X_{右}$ 」として、洗い替え A に進む。
- ⑤ さもないときは一つ若齢化し、新規の  $x$  につき  $x \geq 32$  歳なら③に戻る。 ex. D(x)  
 $x \leq 22$  歳になったら、元々の最高齢  $X_0$  で 索唯一交点 に進む。

索唯一交点

- ⑥  $2 \sim x$ （現時点の最高齢）まで 単純 G 回帰して、新規に  $FF_{左}(x)$  とする。
- ⑦ 左右の関数が 1 点でのみ交わり、かつ、Step ④の「\*状況」にあれば、  
 $FF_{左}(0) > FF_{右}(0)$  を確認し、交点を超えない最大年齢を  $X_{左}$  とする。さらに、  
 $X_{右} = X_{左} + 10$  として、洗い替え B に進む。 ex. E(x) at 38.5 歳,  $X_{左} = 38$
- ⑧ さもないときは一つ若齢化し、新規の  $x$  につき  $x \geq 32$  歳なら⑥に戻る。  
 $x \leq 22$  歳になったら、 $F(x) :=$  ②の  $FF_{右}(x)$  として Single Gauss。

事後手続き

- ⑨ この  $F(x)$  でも、修正済みの相関係数が「G 近似」0.9920 を超えているなら Accept。  
 「近似の閾値」 $0.9900 \leq q < 0.9920$  なら Pending。  $q < 0.9900$  なら Reject。

注 a) 以上の接続手続きで、自由度修正済みの  $q$  は、4 桁表示で 100 Point Up までありうる。

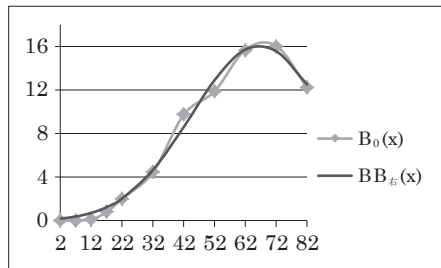
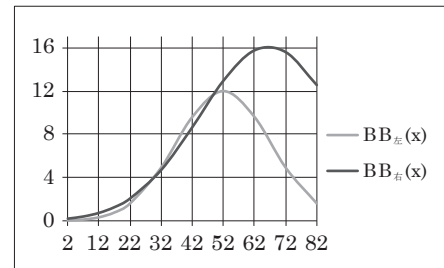
注 b) 筆者は従来から  $q \geq 0.9900$  を、付録 I で示すように「近似の閾値」としてきたが、Gauss 回帰では  
 実際値平均＝回帰値平均 や、全変動＝回帰変動＋残差変動が成立しないので、Gauss 関数による回帰では、  
 幾分厳しくするため、G 近似  $\geq 0.9920$  を設けた。

次のがん研データの 2007 年と 2008 年の平均である  $B_0(x)$  であるが、本論では、福島県の  
 甲状腺がん罹患率  $A_0(x)$  と比較する 規範となる自然罹患率 としては、 $B(100) < C(100)$  など  
 の事情もあり、保守的立場から  $B_0(x)$  を用いる。そこで  $B_0(x)$  を加重 G 回帰すると、

$$BB_{右}(x) = 16.16 \text{ Exp} \{-0.5((x-66.5)/21.7)^2\} \quad \cdots \cdots ③ \quad \text{となった。}$$

$B_0(x)$  と  $BB_{右}(x)$  の修正済み相関係数は  $q = 0.9954$  で、重回帰レベルの「G 近似 0.9920」をクリアしてるが、超近似の 0.9960 にはわずかに及ばない。そこで前頁のフローチャートに従い、「事前手続き」として、単純 G 回帰で  $BB_{左}(x)$  を求めたら、次のようになった。

$$BB_{左}(x) = 11.98 \text{ Exp} \{-0.5((x-51.9)/14.9)^2\} \quad \cdots \cdots ④$$

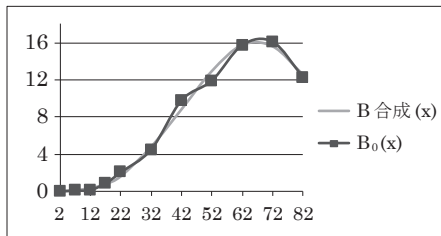
図表 3-A  $B_0(x)$  と加重 G 回帰  $BB_{右}(x)$ 図表 3-B 単純  $BB_{左}(x)$  と加重  $BB_{右}(x)$ 

この「事前手続き」で、図表 3-B のグラフのように、 $BB_{左}(x)$  の曲線が二つのサンプル B (32) と B (42) を内包しつつ、 $BB_{右}(x)$  曲線から横にはみ出すようになる。本論ではこれら内包されたすぐ左サンプルの年齢を「左端年齢  $X_{左} = 22$ 」、すぐ右サンプルの年齢を「右端年齢  $X_{右} = 52$ 」と称する。そして、フローチャートに従い「洗い替え A」に入り、左右のガウス関数を洗い替えしたが、左右とも、仮関数③、④式と何んら変わらなかった。

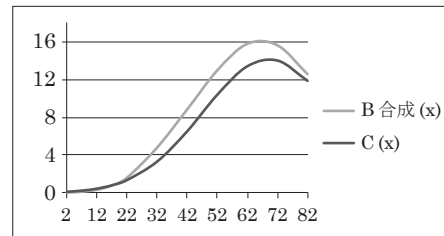
よって、 $X_{左} = 22 \sim X_{右} = 52$  歳で左右のガウス関数を、 $\lambda^{0.5}$  平均を用いて接続すると、 $B(x) = (1-\lambda^{0.5}) B_{左}(x) + \lambda^{0.5} B_{右}(x) \cdots \cdots ⑤$  となる。

ここに、 $\lambda$  は年齢  $x$  の関数で、 $\lambda(x) = (x - X_{左}) / (X_{右} - X_{左}) \cdots \cdots ⑥$  である。

このように、左右を接続した関数が、連続かつ微分可能であることは自明である。合成した  $B(x)$  から得られる回帰値と実際値  $B_0(x)$  の修正済み相関係数は  $q = 0.9962$  と、従前から

図表 4-A  $B_0(x)$  と  $\lambda^{0.5}$  接続した  $B(x)$ 

図表 4-B G 回帰した自然罹患率 B と C



8 ポイント上昇し、「超近似の閾値 0.9960」を超えるに至った。

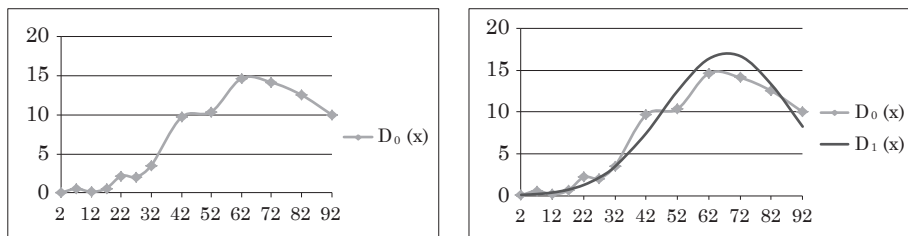
このように、がんの年間罹患率は左右二つのガウス関数の合成で近似できるということを発見したのが、河宮氏と筆者平井で、それを **Algorithm** に乗せ、いかなるがん種についても計算できるプログラムを組んだのが筆者松本である。なお、罹患率の年齢分布の構造に関するゴフマンと河宮（ゴフマン・タンプリン『原子力公害』（明石書店）の訳者）による理論的洞察が、上記事実を発見する基盤となった<sup>5)</sup>。

図表 4-B に示した、甲状腺がんに関する 1998 ～ 2007 年の十年分の平均  $C(x)$  と、2007 年、2008 年の平均から得られる  $B(x)$  とを、比較してすぐ気付く知見が二、三ある。 $C$  のピーク時年齢が 69.2 歳に対し、 $B$  のそれは 66.5 歳であることから、〈罹患の若齢化〉が見られること。今一つは、罹患のピーク値が  $\gamma = 14.10$  人から  $\beta = 16.16$  人へと〈罹患率の増加傾向〉があること。この二つの知見が他のがん種についても妥当するようである。

## § 2 一般がん種の罹患率回帰のために

甲状腺がん以外のがん種罹患率のガウス回帰を論ずるため、この節では人工的に創作した仮想の罹患率  $D_0(x)$  を用いることにする。まずは、対数回帰したグラフ図表 5 と、対数回帰値  $\text{Log } D(x)$  を指数関数  $\text{Exp}$  で元の罹患率に戻した  $D_1(x)$  の数値図表 6 とを表示しておく。

図表 5 仮想の罹患率  $D_0(x)$  と、対数回帰値を「Exp 戻し」した  $D_1(x)$  のグラフ



この対数回帰によって、当該回帰式は付録Ⅱに示すモデルの「判定水準」には達しているが、修正済み相関係数は  $q = 0.9736$  なので、近似の閾値 0.9900 には遠く及ばない。

五歳刻みの年齢階級別罹患率がガウス型の年齢分布かも知れないと思い、対数回帰したとしても、解析の結果が上の図表 5 右のようなグラフになっていたら、実際値と回帰値が近似しているとはいいがたい。しかし、 $D_0(x)$  のような一見ガタガタな三角分布であっても、河宮氏の理論と筆者の計算手法によれば、ガウス関数で近似できる。

5) 河宮信郎，蔵田計成，平井孝治，「がん罹患率への分析視座，年齢階級別統計から年齢別解析へ」『科学』（岩波書店）に投稿中。

図表 6 仮定の罹患率  $D_0(x)$ 、対数回帰分析の数値結果

分析精度

説明係数 $R^2$	0.9291
修正済み説明係数 $Q^2$	0.9150
修正済み相関係数 $q$	0.9565

相係 q の戻し計算		回帰レベル
説係 Q <sup>2</sup>	0.9478	代用 T
相係 q	0.9736	≤ 99.00 %

対数回帰値の戻し計算

		Exp 戻し	自然対数	
年齢	$D_0(x)$	$D_1(x)$	実績値	回帰値
2	0.037	0.07	-3.30	-2.61
7	0.55	0.16	-0.60	-1.81
12	0.15	0.34	-1.90	-1.07
17	0.60	0.67	-0.51	-0.40
22	2.19	1.24	0.78	0.22
27	2.05	2.15	0.72	0.76
32	3.53	3.48	1.26	1.25
42	9.66	7.57	2.27	2.02
52	10.36	12.77	2.34	2.55
62	14.61	16.68	2.68	2.81
72	14.10	16.89	2.65	2.83
82	12.50	13.26	2.53	2.58
92	10.00	8.07	2.30	2.09

分散分析表

変動	偏差平方和	自由度	不偏分散	分散比	p 値	判定
全変動	44.60	12				
回帰変動	41.44	2	20.72	65.55	1.8E-06	[☆]
残差変動	3.16	10	0.32			

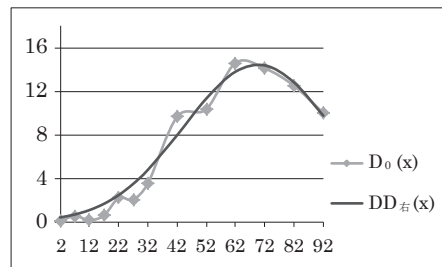
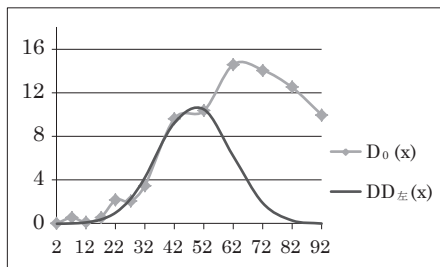
重回帰式

	回帰係数	標準回帰係数	F 値	t 値	p 値	判定
$x^2$	-0.001	-1.859	30.9	-5.56	2.4E-04	[****]
年齢 $x$	0.172	2.642	62.3	7.89	1.3E-05	[*****]
定数項	-2.952		55.4		2.2E-05	[*****]

我々の開発した手法では、先ず 2 歳から最高齢の 92 歳までの残差平方に加重をかけ、その和が最小化になるようガウス回帰の三つのパラメータ、即ち、ピーク時年齢  $\mu$  と年齢偏差  $\sigma$  とピーク値  $\delta$  を求めて、仮の  $DD_{右}(x)$  とする。すると次のような ⑦式になる。

$$DD_{右}(x) = 14.44 \text{ Exp} \{ -0.5((x-69.6)/25.3)^2 \} \quad \cdots \cdots \textcircled{7}$$

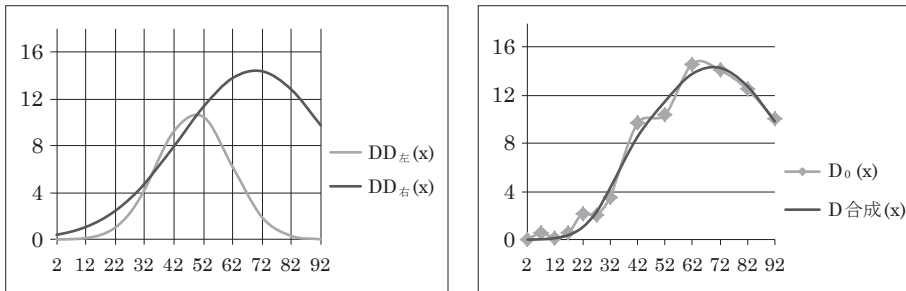
この加重 G 回帰だけで、修正済み相関係数は  $q = 0.9879$  と実に 143 Point も上昇するが、それでもまだ近似の閾値 0.9900 には及ばない。これをグラフにすると、図表 7 の右側のようになり、対数回帰との優劣は歴然としてくる。

図表 7  $D(x)$  の仮の左関数と仮の右関数

今求めた仮の右関数は前述のフローチャート Step ②の超近似 0.9960 にとどいてないので、「事前手続き」に入り、Step ⑤の  $x = 52$  歳に至る。式は省くが、単純 G 回帰により上の図表 7 の左に示すような仮の左関数  $DD_{左}(x)$  が得られ、次の図表 8 左のグラフに見られるように、サンプル  $D_0(42) = 9.66$  だけが内包されるようになる。



図表 8 洗い替えた左右の関数とそれを  $\lambda^{0.5}$  平均で接続した  $D(x)$  のグラフ



そこで  $X_{左} = 32$  歳と  $X_{右} = 52$  歳を用いて、左右の関数を洗い替えすることになる。即ち、新規に  $2 \sim X_{右}$  までの単純 G 回帰で  $D_{左}(x)$  を、 $X_{左} \sim$  最高齢までの加重 G 回帰で  $D_{右}(x)$  を、算定する。その結果、 $\lambda^{0.5}$  平均をへて最終的に、次の年齢密度関数  $D(x) \cdots \textcircled{8}$  が得られる。 $\delta = 0.05^{0.5}$  として、

$$D(x) = \begin{cases} 10.81 \text{ Exp} \{ -0.5((x-48.9)/12.4)^2 \} & \text{if } x \leq 32 \\ (1 - \delta(x-32)^{0.5}) D_{左}(x) + \delta(x-32)^{0.5} D_{右}(x) & \text{if } 32 \leq x \leq 52 \\ 14.35 \text{ Exp} \{ -0.5((x-69.4)/26.0)^2 \} & \text{if } x \geq 52 \end{cases}$$

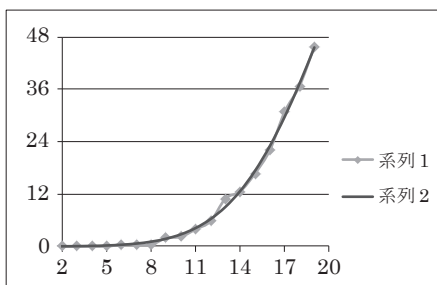
実際値  $D_0(x)$  とこの合成した  $D(x)$  で得られる回帰値、計 13 サンプルの自由度修正済み相関係数は  $q = 0.9920$  となり、「G 近似」と一致し、⑦式をさらに 41 ポイントも上回るようになった。対数回帰から見ると、実に 184 Point もの Up である。このように、

**加重 G 回帰を用いると、がん罹患率が左右のガウス関数の  $\lambda^{0.5}$  接続で近似できる。**

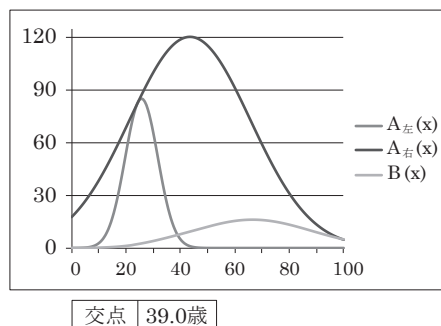
### § 3 福島甲状腺がん、県民罹患率 $A(x)$ の解明

我々の研究の動機は、被曝後の福島の甲状腺がん罹患率にある。そこで、まずは図表 0-A にある実際値  $A_0(2) \sim A_0(19)$  をグラフで観察して視ることにする。

図表 9-A 県民罹患率  $A_{左}(x)$  の左尾



図表 9-B 24.3 歳で左関数に接する Case I



図表 9-A のグラフを見ると、指数関数のようにも見えるが、普通なら「六十歳代ぐらいでピークを迎え、それより高齢ではがん細胞への変異が鈍化する」……\*\* ので、幾つになっても罹患率が指数状に伸びることなどあり得ない。だとすると、当該グラフはガウス関数の一部と見るのが妥当である。即ち、今まで論じてきた左関数のレフトテールである。

福島県の実際値  $A_0(x)$  は  $x = 2 \sim 19$  歳までしかないが、これを単純 G 回帰した左関数は次式ようになる。

$$A_{\text{左}}(x) = 85.12 \text{ Exp} \{ -0.5((x-25.7)/6.0)^2 \} \quad \cdots \cdots \textcircled{10}_{\text{左}}$$

この $\textcircled{10}_{\text{左}}$ 式による回帰値と実際値の自由度修正済み相関係数は  $q = 0.9988$  で、筆者たちが見かけた生涯最高の近似である。これほど厳格な近似は、〈被曝時年齢を受診時年齢に変換する河宮・蔵田プロセス<sup>6)</sup>の妥当性〉と、〈ガウス型の年齢密度関数のがんの年間罹患率に適合〉することの、まぎれもない証左である。

しかし、このガウス関数は、図表 9-B のグラフに見られるように、39.0 歳から自然罹患率  $B(x)$  を下回り、右側の G 関数が存在しないことには理屈に合わない。福島県当局も国立がん研究センターも、事故当時 18 歳超の県民の標本調査を実施していないため、このままでは加重をかけた右関数  $A_{\text{右}}(X)$  が求められない。

右関数を推計するには、県民罹患率を可及的に小さく見積もる 下算の限界を整備しておく必要がある。そこで、年齢ごとの乖離幅  $\triangle(x) = A(x) - B(x)$  を導入する。

< a > 自明の理ではあるが、 $A(x)$  は自然罹患率を下回らない。具体的には  $20 \leq x \leq 99$  で、 $\triangle(x) \geq 0$ 。ただし、 $-0.20 \leq \triangle(100) < 0.00$  を境界条件<sup>6A)</sup>とする。

< b > 当然、 $\mu_{\text{右}}(> 43.5)$  より若齢期では増加関数である。これを次の無理のない条件  $20 \leq x \leq 43$  において<sup>6B)</sup>、 $1.001 \leq \text{Min} \{ A(x) / A(x-1) \} \leq 1.004 \cdots \textcircled{10}'$  で代替する。この比率 1.001 は、 $\lambda^{0.5}$  接続のための『遊び』であるとともに、接続した後のグラフが、「左関数から右関数に、見た目平滑に移行する」ためでもある。

これらの条件の下に左右の関数を  $\lambda^1$  接続し、

< c > 20 歳から 99 歳までの乖離幅  $\triangle(x)$  の和  $\Sigma \triangle(x)$  を最小化すると、下算の限界に至る。

0 ~ 19 歳では  $A_{\text{左}}(x)$  に限る ことを前提に、上記の三条件を満たす右関数のうち、( $\lambda$  接続した)  $\Sigma \triangle(x)$  が最小になるものが下算した  $A_{\text{右}}(x)$  の Unique Solution である。

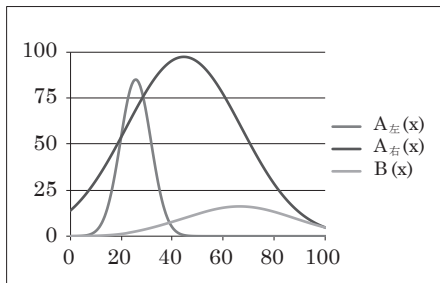
6) 脚注 5) の論文「がん罹患率への分析視座、年齢階級別統計から年齢別解析へ」の付録 O。

6A) この頁の冒頭に記した「……\*\*事実」より、高々齢では罹患率が下がるので、このように想定した。

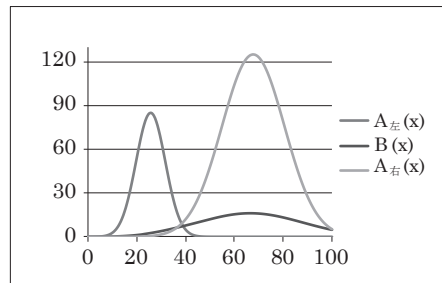
6B) 接する場合のピーク時年齢  $\mu$  が 43.5 歳なので、その他の Case では  $\mu_{\text{右}} > 43.5$ 。

右関数の類型としては、左関数と接する場合（Case I）、それより右関数が右の方にシフトして 2 点で交わる場合（Case II）、さらに右側にシフトして左関数と 1 点でのみ交わる場合（Case III）の三つである。Case I は既に図表 9-B に示しているの、残り二つの Case を、次に図で例示する。

図表 9-C 右関数が県民罹患率  $A_{\text{県}}(x)$  と 2 点で交わる Case II



図表 9-D 右関数が 38.4 歳でのみ左関数と交わる Case III



まずは Case I だが、図表 9-B のように 24.3 歳で接する場合、ガウス式 ( $\mu = 43.5$ ) は省くが、左関数と  $\lambda^1$  接続すると、 $\Sigma \angle(x) = 4,897$  となる。

次に Case II であるが、前述の「事前手続き」で左関数と 2 点で交わる右関数を想定し、20 歳以降の罹患率を下算する。二つの関数が交わる左端年齢  $X_{\text{左}}$  は、高齢化すればするほど、条件  $\langle a \rangle$  と  $\langle b \rangle$  のために  $\Sigma \angle(x)$  が大きくなる。よって、下算のためには  $X_{\text{左}} = 19$  歳よりない。

他方、実際値が存在しない中で、右端年齢  $X_{\text{右}}$  を如何に設定するかが問われる。そこで  $X_{\text{右}} = 20, 21$  と一歳刻みに増やしていくと、( $\lambda^1$  接続の段階で)  $\Sigma \angle(x)$  が減っていき、27 歳で 5,839、28 歳で 4,631、29 歳で 3,978 になるが、 $x = 30$  歳から以降では、⑩' 式の差分比が 0.9956 より小さくないと、右側の交点が存在しない。従って Case II で、下算の限界を究める右端年齢は  $X_{\text{右}} = 29$  歳ということになる。

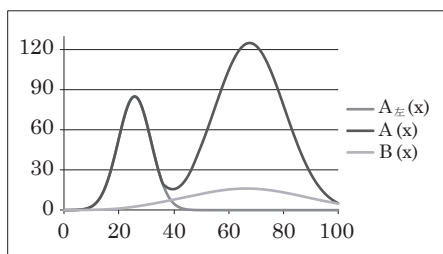
最後に Case III であるが、この場合には、既述の三条件の他にも、 $A_{\text{左}}(0) > A_{\text{右}}(0)$  が不可欠である。これを充足するのが、例えば、図表 9-D の (接続前の) 仮関数で、次の如くである。

$$AA_{\text{右}}(x) = 125.20 \text{ Exp} \{ -0.5((x-67.7)/12.8)^2 \} \quad \cdots \cdots \textcircled{9}$$

$\Sigma \angle(x)$  が最小になるよう、左右の関数を 36 ~ 48 歳まで 2 次式で接続すると、図表 9-E のようなグラフになる (この場合、 $\Sigma \angle(x) = 4,219$ )。しかしこれは明らかに条件  $\langle b \rangle$  に反する。そこで条件と整合するよう、34.6 歳で微係数が 0 になる 4 次式を使って左右のピークに綱を架け渡したと、図表 9-F に示すごとく、 $\Sigma \angle(x)$  が大きくなる (図表 9-F の場合  $\Sigma \angle(x) = 5,999$ )。かくして、下算のためには Case III は断念するよりない。

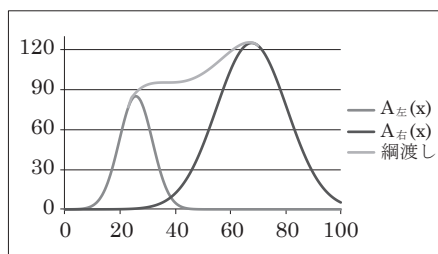
図表 9-E Case III の 2 次式接続

$$0.316(x-39.6)^2 + 15.7$$



図表 9-F Case III の「網渡し」

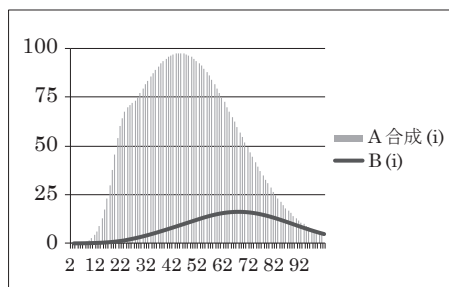
$$-0.00011x^4 + 0.0206x^3 - 1.38x^2 + 39.8x - 325.23$$



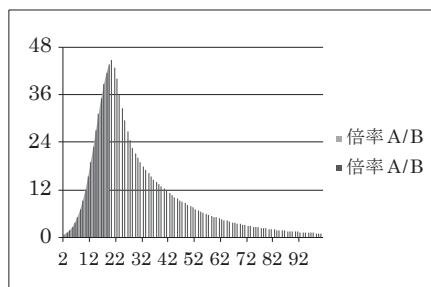
こうして、左端年齢  $X_{左} = 19$  歳、右端年齢  $X_{右} = 29$  歳の 2 点で左関数と交わる G 関数が、求める下算の限界を究めた右関数ということになる。先に⑩<sub>左</sub>式で示した  $A_{左}(i)$  と  $\lambda^{0.5}$  接続すると、下算した県民罹患率  $A(i)$  は、次のような式になる。 $\alpha = 0.1^{0.5}$  として、……⑩

$$A(i) = \begin{cases} 85.12 \text{ Exp} \{-0.5((i-25.7)/6.0)^2\} & \text{if } i \leq 19 \\ (1-\alpha(i-19)^{0.5})A_{左}(i) + \alpha(i-19)^{0.5}A_{右}(i) & \text{if } 19 \leq i \leq 29 \\ 97.45 \text{ Exp} \{-0.5((i-44.6)/22.6)^2\} & \text{if } i \geq 29 \end{cases}$$

この  $A(i)$  を 2 ～ 100 歳まで一歳刻みにプロットしたのが図表 10 左側の細棒グラフで、グラフエリアの左肩（左右の接続部）は、下算した「傷跡」である。また、下の方に「這いつくばっている」曲線は自然罹患率  $B(i)$  である。更にまた、右側の細棒グラフは県民罹患率  $A(i)$  を自然罹患率  $B(i)$  で割った倍率で、塗りつぶし部分は 2 歳から 19 歳までの実績を反映したものである。なお、 $\lambda^{0.5}$  接続した結果、 $\Sigma \Delta(x)$  は 3,960 と 18 下がった。

図表 10 県民罹患率  $A(i)$  の細棒グラフと、自然罹患率との倍率  $A/B$ 

Max	44.6 歳	97.45 人
-----	--------	---------



Max	20.0 歳	44.68 倍
-----	--------	---------

最大値は  $\mu_{右} = 44.6 (\pm 0.5)$  歳で、10 万人あたり  $\alpha_{右} = 97.45$  人である。

#### § 4 累積罹患 $A(k, n)$ と生涯罹患人口 $\Psi_A(y, z)$

がん研データの年齢階級別罹患率は 80 ～ 84 歳までであり、かつまた、平均寿命に近いこ

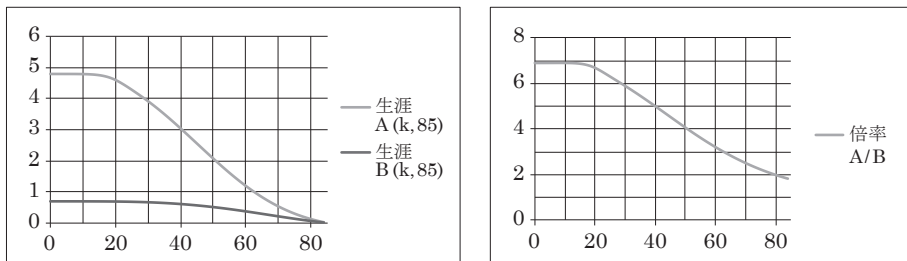
とから、事故当時福島県に住んでいた人々が満 85 歳直前まで生存すると仮定する。すると、当時満 k 歳だった人の n 歳直前までの累積罹患は、次の⑪式で表せる。

$$\text{累積罹患 } (0 \leq k \leq n \leq 85) : A(k, n) = \sum_{i=k}^{n-1} A(i) \cdots \cdots \text{⑪}$$

我々は、年間罹患率  $A(i)$  をガウス関数で近似したので、「累積罹患」や「生涯罹患」も年齢で積分して定義するのが妥当である。しかし、年齢別人口  $P_1(k)$  が高々一歳刻みの年齢に関し離散型なので、有効数字が少数第二位まで信頼できる Summation で定義する。 $n = 85$  のときは生涯罹患だが、自然罹患の方も同様に生涯罹患  $B(k, 85)$  が計算できる。

一般に、⑪式の累積罹患  $F(k, n)$  は、『満 k 歳まで罹患してない人が、それ以降満 n 歳になるまでに罹患する率』を含意している。そこで、福島県民の生涯罹患  $A(k, 85)$  と、自然罹患のそれ  $B(k, 85)$  を、図表 11 にグラフで示すことにする。

図表 11 十万人当たり 当時満 k 歳の人々の生涯罹患  $A(k, 85)$  と自然罹患  $B(k, 85)$  との倍率



4.79 千人、0.695 千人 このグラフの値は「%」でもある。 6.898 倍 Max は 8 歳の 6.901 倍

事故当時、生まれたばかりの  $k = 0$  歳児は 10 万人中、生涯に渡って少なくとも 4,791 人、率にして 4.79% が甲状腺がん罹患し、それは、自然罹患の 0.695% の実に 6.90 倍にもなる。下算の限界に挑んできたので、実際には 10.87 倍になるものと思われる。さらに、当時 19 歳以上の人々の被曝がんは等閑視されているが、例えば、事故当時  $k = 40$  歳だった人でさえ、生涯罹患は 3.04% もあり、自然罹患の 5.00 倍にのぼる。

かくして、Cohort（この場合は事故当時の福島県民）の一歳刻みの人口  $P_1(k)$  を特定すれば、年齢別の生涯罹患人数は  $P_1(k) A(k, 85) \cdots \text{⑫}$  となる。さらに、この人数の Summation

$$y \text{ 歳} \sim z \text{ 歳までの生涯罹患人口} : \Psi_A(y, z) = \sum_{k=y}^z P_1(k) A(k, 85) = \sum_{k=y}^z P_1(k) \sum_{i=k}^{85-1} A(i) \cdots \cdots \text{⑬}$$

は、「当時 y 歳 ~ z 歳の世代が 85 歳直前までに罹患する人口」を含意する。従って  $y \leq z \leq 84$  のとき、被曝に由来する世代別生涯罹患人口  $\Psi_{A-B}(y, z)$  は、次の式で計算できる。

$$\Psi_{A-B}(y, z) = \sum_{k=y}^z P_1(k) A(k, 85) - \sum_{k=y}^z P_1(k) B(k, 85) = \sum_{k=y}^z P_1(k) \{A(k, 85) - B(k, 85)\} \cdots \cdots \text{⑭}$$

2010 年の国勢調査から福島県の年齢別人口  $P_1(k)$  を拾うと、当時 18 歳以下の生涯罹患人口は  $\Psi_{A-B}(0, 18) = 14,989$  人である。さらに、県民の被曝罹患人口は  $\Psi_{A-B}(0, 84) = 40,464$  人となる。その他、 $\Psi_{A-B}(20, 39)$  などを計算すると、世代間差異など、被曝に由来する罹患人口について議論できるが、詳細は河宮氏と共同執筆中の論文(仮題)「Cancer Incidence Rate (Age Function) and the Resultant Health Situations by Cohort」で展開する。

⑬式で解るように、年間罹患率が年齢別でないと、生涯罹患人口  $\Psi_F(y, z)$  が計算できず、従って、被害の全貌も明らかにできない。ベラルーシなど、世界中を検索しても、これを論じた跡がないのは、罹患率を一歳刻みで計算する作法が無いからだと思われる。このように、各種がんについて、世代別生涯罹患人口  $\Psi_F(y, z)$  を計算することには大義がある。

この節に限らず、本論文では甲状腺はもとより、どんながん種にも敷衍できる重要な概念を反映した諸々の関数を紹介してきた。これらの関数と Gofman の「絶対過剰罹患」との関連については、文末の付録 0「関数の定義と整合性の吟味」を参照してもらいたい。

稠密な議論には、ガウス回帰によって年間罹患率を密度関数に変換することが不可欠だが、概算でよければ、起点の年齢階級別罹患率でも「それなりの生涯罹患」が計算できる。その算法については、文末の Annex『補間による「年齢階級別罹患率から生涯罹患へ」の概算法』を参照して頂きたい。仄聞するところによると、国や県当局などは「甲状腺がんの罹患者が多いのは、被曝ではなく過剰診断のせいだ」とのよし。関係諸機関には「歴史の検証に耐えられる診たてと施策」を要望したいものである。幸いなことに、甲状腺がんは早期に対処さえすれば、死にいたる病ではないのですから。

## お わ り に

この研究のきっかけは、2015 年の秋にゴフマン研究会の会員である蔵田計成氏が、福島県の年齢階級別罹患率を算定し、自然罹患のそれと比較するため、被曝時年齢を受診時年齢に変換する手法を、同研究会を主宰する河宮信郎氏に相談したことに始まる。かくして開発された手法〈河宮・蔵田プロセス〉を経て、年間罹患率の実値  $A_0(7)$ ,  $A_0(12)$ ,  $A_0(17)$  を計算し、対応する自然罹患率  $B_0(x)$  の何倍になるかを論じておられた。その後、エントロピー学会でかねてから河宮氏と親交のある筆者平井や、Algorithm とプログラミングに長けた筆者松本も該研究会に参加することになった。

当初は、県民罹患率と自然罹患率との差が有意であることを、t 検定や  $\chi^2$  検定で示せば充

分だと考えていた。しかし井野博満氏<sup>7)</sup>から当該が事実上の悉皆調査であると知らされ、検定では意味がないことが判った。検定が駄目なら通常は回帰分析でモデル式を作って検討する。ところが、説明変数の数  $x$  は、サンプル数を  $n$  とすると  $x \leq 0.2 (n-1)$  なる「 $x$  ope 制約」(付録Ⅱの脚注参照)があり、3 対のサンプルでは如何ともしがたい。

そこで、サンプル数不足を補う手が無いかと、河宮氏と共に検討していたところ、 $A_0(x)$ を一歳ごとの移動平均で求めたら、罹患率が年齢密度関数になることに気付いた。そこで、各種がんのがん研データ（の対数）をグラフにとると、ほぼ放物線になっていた。ここまで来ると、がんの「年間罹患率はガウス型の年齢密度関数で近似できる」との確信に至る。

しかし、通常の変数解析では罹患率を「高い分解能で関数近似する」など思いもよらない。思うに、統計処理に精通した疫学者たちも、既に我々と同じバリアに遭遇されたであろう。この障壁を超えられたのは、「ゴフマンの被曝罹患に関する洞察」に対する河宮氏の解釈であった。彼の解釈から、比較的若齢期のガウス型密度関数と、比較的高齢期のそれを接続させる罹患率モデル式の試作を繰り返し、実際値と回帰値の自由度修正済み相関係数が『G 近似の閾値 0.9920』を超えるに至った。

なお本文中、同じ近似でも「近似の閾値 0.9900」と「G 近似 0.9920」,「超近似 0.9960」を使い分けてきたが、見た目の程度の近似なのかは、付録Ⅲのグラフを参照されたい。

河宮信郎氏を含む我々筆者の研究成果を要約すると、次の三点になる。

- (1) 「がんの年間罹患率はガウス型の年齢密度関数に従う」という事実の発見。
- (2) 年齢階級別罹患率をガウス型の密度関数で近似する計算手法の開発。
- (3) それに基づく累積罹患  $A(k, n)$  と、被曝由来の罹患人口  $\Psi_{A-B}(0, 84)$  の予測。

実をいうと、「加重した残差平方」の加重のかけかたについて、本論文では秘匿している。というのも、河宮氏と共に、発見した事実と開発した計算手法を、世界中に伝える論文を企図しているからである。この執筆中の論文は科学雑誌『Nature』かまたは、医学系の専門誌『LANCET』に投稿を予定している。ご理解いただきたい。

最後になりましたが、「加重をかけてガウス関数に回帰する」ことに思い至ったのは、これまで筆者の思索に磨きをかけて下さった方々のお蔭だと思います。なかでも、筆者平井の小学校時代の恩師福井基一先生には学恩の念を禁じえません。今一人は筆者平井の長年の友人であり、筆者松本の親しい知人でもある田原孝氏<sup>8)</sup>で、ここに感謝の意を表します。また、我々の拙稿を快く受理して頂いた本誌関係者にも、紙上を借りて厚くお礼を申し上げます。

7) 元東京大学工学研究科教授，エントロピー学会会員

8) 元日本福祉大学福祉経営学部教授，精神科医



## Annex『補間による「年齢階級別罹患率から生涯罹患へ」の概算法』

疫学界では生涯罹患人口  $\Psi_c(y, z)$  を計算するしきたりがないように見受けられる。凹凸を平準化するために、五歳ごとにブロック化して年齢階級別罹患率を算出するが、これを一歳ずつ移動して計算すれば年齢別罹患率が求まり、ひいては累積罹患も計算できる。しかし国立がん研究センターからデータの提供を受けないと出来るものではない。本文で紹介した年齢密度関数は年齢に関して「連続表示」だが、「離散表示」である公表データからでも年齢別罹患率がそれなりに概算できるので、以下にその手法を紹介する。

本文中で紹介した甲状腺がんの十年分の平均である  $C_0(x)$  は、2007 年と 2008 年の年間罹患率の平均  $B_0(x)$  より均整が取れているので、これを例に取る。本文中の記述法と同様に、例えば、30～34 歳の年齢階級別の年間罹患率は、 $x = 32$  歳のそれとみなし、中央年齢を用いて  $C_0(32)$  と記す。当該データを、図表 0 から次に再掲しておく。

図表 12 自然罹患率  $C_0(x)$  の五歳階級別罹患率 (再掲)

$C_0(x)$	$C_0(2)$	$C_0(7)$	$C_0(12)$	$C_0(17)$	$C_0(22)$	$C_0(27)$	$C_0(32)$	$C_0(37)$	$C_0(42)$
罹患率	0.02	0.02	0.22	0.58	1.58		3.53		6.59

$C_0(x)$	$C_0(42)$	$C_0(47)$	$C_0(52)$	$C_0(57)$	$C_0(62)$	$C_0(67)$	$C_0(72)$	$C_0(77)$	$C_0(82)$
罹患率	6.59		10.20		11.90		14.10		11.80

この図表 12 から分かるように、生涯罹患  $C_0(k, 85)$  を算定するため、まずは、公表データにはない六つの年間罹患率を推計する必要がある。証明は省くが、 $n \geq 1$  を前提に、 $0 < a \leq b$  で、 $0 \leq \lambda \leq 1$  なら、次の不等式が成立する。

$$a \leq \frac{1}{\left\{ \frac{(1-\lambda)}{a} + \frac{\lambda}{b} \right\}} \leq \left\{ (1-\lambda) \sqrt[n]{a} + \lambda \sqrt[n]{b} \right\}^n \leq (1-\lambda)a + \lambda b \leq ((1-\lambda)a^n + \lambda b^n)^{\frac{1}{n}} \leq b \quad \cdots \cdots (15)$$

$n = 2$  で  $\lambda = 0.5$  なら、左から順に調和平均、平方根平均、算術平均、二乗平均であるが、ガウス風の曲線の場合、補間 (ないし補正) するには  $n = 2.5$  がよく適合する。

まず、グラフの頂点付近にある  $C_0(62) = 11.90$  がイレギュラーなので、3 次式を用いて補正しておく。 $C_0(62) = 11.90 \Rightarrow C_0(62) = 13.135$  (ちなみに②式によると、 $C(62) = 13.352$ )

$C_0(27)$  や  $C_0(37)$  がレフトテールなら調和平均が似合うのだが、どうやら凹部なので、0.4 乗平均を用いるのが妥当である。よって、

$$C_0(27) = (0.5(C_0(22))^{0.4} + 0.5(C_0(32))^{0.4})^{2.5} = 2.439$$

$$C_0(37) = (0.5(B_0(32))^{0.4} + 0.5(B_0(42))^{0.4})^{2.5} = 4.918 \quad \text{となる。}$$



以下、補間した罹患率は少数第三位に丸めて表記することにする。

$C_0(47)$  は、変曲点付近にあるようだから、算術平均で差支えない。

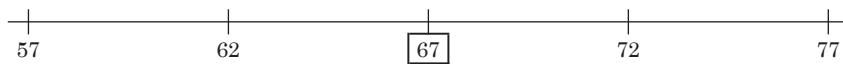
$$C_0(47) = 0.5 C_0(42) + 0.5 C_0(52) = 8.395$$

$C_0(57)$  や  $C_0(77)$  は左右の凸部にあるようだから、2.5 乗平均を用いるのが妥当する。

$$C_0(57) = (0.5(C_0(52))^{2.5} + 0.5(C_0(62))^{2.5})^{0.4} = 11.805$$

$$C_0(77) = (0.5(C_0(72))^{2.5} + 0.5(C_0(82))^{2.5})^{0.4} = 13.026$$

$C_0(67)$  は頂点付近にあるようだから平均律では補間できない。そこで次のような数直線を想定し、当該の右側に 2 サンプルあれば 3 次式で、1 サンプルしか無ければ 2 次式で補間する。この場合は、今補間したばかりの  $B_0(57)$  や  $B_0(77)$  も動員して、3 次式で補間するが、後の手間を省くため、63～71 歳まで一気に補間することにする。



その結果は次のとおりである。

Ca (x)	Ca (63)	Ca (64)	Ca (65)	Ca (66)	Ca (67)	Ca (68)	Ca (69)	Ca (70)	Ca (71)
罹患率	13.359	13.562	13.741	13.894	14.018	14.110	14.167	14.186	14.165

以上で五歳刻みの年齢階級別罹患率が補間できたので、生涯罹患は、次のようになる。

$$\Delta k = 5 \text{ として, } C_0(0, 85) = \sum_{k=0}^{16} C_0(5k+2) \Delta k = 581.88 \cdots \textcircled{16}$$

さらに、コホートの五歳刻みの人口  $P_5(k)$  が入手できれば、粗くはあるが、 $y \sim z$  歳の人が 85 歳直前までに罹患する生涯罹患人口  $\Psi_c(y, z)$  も、五歳刻みなら算定することができる。

だが、それなりの累積罹患を求めようとするなら、一歳刻みの年齢別罹患率が必要となる。そのためには、0～84 歳まで罹患率を補間する必要がある。2～7 歳、……、77～82 歳までは、まずは図表 13 のように、部位を見極めたうえで、各部位に適合する平均律を用いて、 $\lambda = 0.2, 0.4, 0.6, 0.8$  と補間すればよい。

図表 13 曲線  $C_0(x)$  における各年齢階級の部位

年齢階級	2～7 歳	7～12 歳	12～17 歳	17～22 歳	22～27 歳	27～32 歳	32～37 歳	37～42 歳
部位	テール	テール	凹	凹	凹	凹	凹	凹

年齢階級	42～47 歳	47～52 歳	52～57 歳	57～62 歳	62～72 歳	72～77 歳	77～82 歳
部位	直線	凸	凸	凸	頂点付近	凸	凸

問題は「0, 1 歳」と、「83, 84 歳」の補間である。前者の場合は、 $C_0(2)$  と  $C_0(7)$  での補間

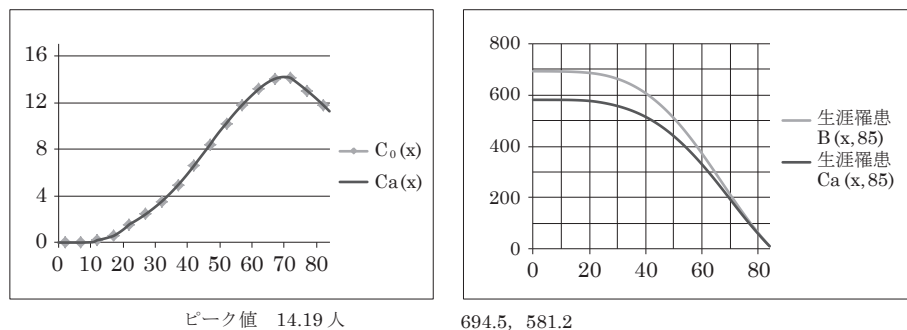
を,  $\lambda = -0.4, -0.2, 0, 0.2, 0.4, \dots$  と左側に 2 歳分外延して調和平均を取ればよい。  
しかし,  $C_0(x)$  の場合は  $C_0(2) = C_0(7) = 0.02$  なので, 計算するまでもない。

後者の場合は,  $C_0(77)$  と  $C_0(82)$  での補間を,  $\lambda = 0.2, \dots, 1, 1.2, 1.4$  と右側に 2 歳分外延して 2.5 乗平均を取れば足る。その結果は次のとおりである。

$C_a(x)$	$C_a(77)$	$C_a(78)$	$C_a(79)$	$C_a(80)$	$C_a(81)$	$C_a(82)$	$C_a(83)$	$C_a(84)$
罹患率	13.026	12.795	12.557	12.313	12.060	11.800	11.531	11.252

このように補間して得られたのを「補間済み年齢別罹患率  $C_a(x)$ 」と称することにし, 実際値  $C_0(x)$  とともにグラフで図示すると, 図表 14 左のようになる。

図表 14 十万人当たりの補間済み年齢別罹患率  $C_a(x)$  と生涯罹患  $C_a(x, 85)$



さらに, 本文中の⑪式のように, 累積罹患は次のようになる。

$$k \text{ 歳から } n \text{ 歳になる直前までの累積罹患} : C_a(k, n) = \sum_{i=k}^{n-1} C_a(i) \quad \dots\dots ⑰$$

$n$  に平均寿命と仮定する 85 歳直前を代入した生涯罹患  $C_a(x, 85)$  を, 同じ自然罹患である既述の  $B(x, 85)$  とともに図表 14 右のグラフで示しておく。

⑯式の生涯罹患は  $C_0(0, 85) = 581.88$  人であったが, この度は  $C_a(0, 85) = 581.22$  人となった。この生涯罹患の差は五歳刻みと一歳刻みの違いに由来する。さらに, コホートの一歳刻みの人口  $P_1(k)$  が手に入れば, 年齢別の生涯罹患人数  $P_1(k) C_a(k, 85) \quad \dots\dots ⑱$  も計算できることになり, ひいては, 世代別の生涯罹患人口  $\Psi_c(y, z)$  も算定できる。

以上, 生涯罹患の便宜的な概算法を紹介してきた。このようにして得た「補間済み年齢別罹患率」は連続まがいではあるが, 積分不能である。また, 相関係数  $r$  が算定できないので, 「近似の精度」も評価できない。それゆえ, 「補間済み」から得られた「生涯罹患人口」も, 累加する過程で誤謬が緩和されるとはいえ, その利用については注意を要する。

付録 0

関数の定義と整合性の吟味

以下では Inci. Cumu. Rad- は、 各々 Incidence, Cumulative, Radiation (or Radiational) の略

	和 英 名 称 など	Label	意味論的定義	定義式 and/or 同義式	次元
①	年齢別罹患率	$F(x)$	Incidence Rate by Age	満 $x$ 歳の回帰罹患率	$10^{-5}/y$
②	累積罹患 Cumu. Inci. from $k$ to $(n-1)$ Years of Age	$F(k, n)$	満 $k$ 歳直後から満 $n$ 歳になる直前までの累積罹患	$0 \leq k \leq n$	$10^{-5}$
③	生涯罹患 Life-Span Incidence after $k$ Years of Age	$F(k, 85)$	満 $k$ 歳直後から満 85 歳になる直前までの累積罹患	$\sum_{i=k}^{n-1} F(i) \Leftrightarrow \int_k^n F(t) dt$ $n \leq k \Rightarrow F(k, n) = 0$	
④	ゴフマン・河宮の被曝履歴関数 Historical Incidence Affected by Rad-Exposure 通称「Butterfly Function」、 $\tau$ : 被曝時年齢 $N(0, \tau+1)$ は自然罹患の $0 \sim \tau$ の累積罹患	Life- $\Phi(\tau, x)$	$\Phi(\tau, x) = \begin{cases} \int_0^x N(t) dt & \text{if } x \leq \tau \\ N(0, \tau+1) + \int_\tau^x A(t) dt & \text{if } x \geq \tau \end{cases}$		$10^{-5}$
⑤	被曝誘発罹患率 Rad-Induced Incidence Rate $\Phi(\tau, x) - N(0, x+1)$ の偏微分 $\tau$ : Age at Radiation Exposure if $x \leq \tau$ then $\varphi_\tau(x) = 0$	$\varphi_\tau(x)$ $\tau \leq x$	$= \frac{d}{dx} \{ \Phi(\tau, x) - N(0, x+1) \} = \frac{d}{dx} \int_\tau^x \{ A(t) - N(t) \} dt$ $= [A(t) - N(t)]_\tau^x = \{ A(x) - N(x) \} - \{ A(\tau) - N(\tau) \}$ $= A(x) - N(x) \dots\dots$ (連続表示)		$10^{-5}/y$
⑥	Gofman の絶対過剰罹患率 Absolute Excess Incidence Rate	$\varphi_\tau(j)$ $\tau \leq j$	$= DK(\tau, j - \tau)N(j) \dots\dots$ (離散表示) ここに、定数 $D$ は Cohort の各人が $\tau$ 歳で一様に浴びた線量 $K(*, *)$ はリスク係数		
⑦	年齢別累積罹患人数 Cumu. Cohort Inci. for Ages from $k$ to $(n-1)$ Years		満 $k$ の人が満 $n$ 歳になる直前までの累積罹患人数	$P_1(k) F(k, n)$ ここに $P_1(k)$ は該コホートの年齢別人口 ( $10^5 \text{ person}$ )	person
⑧	世代別生涯罹患人口 当歳から 85 歳になる直前まで	Life-Span Cohort Incidence after $y \sim z$ Years of Age		$\Psi_F(y, z) = \sum_{k=y}^z P_1(k) F(k, 85)$	person
⑨	被曝由来の生涯罹患人口 Cumulative Rad-Origin Incidence in the Life of Rad-Exposed Cohorts at $0 \sim 84$ years of Age		$\sum_{\tau=0}^{84} P_1(\tau) \{ \Phi(\tau, 85) - C(0, 85) \} = \sum_{\tau=0}^{84} P_1(\tau) \int_0^{85} \varphi_\tau(t) dt$ $= \sum_{\tau=0}^{84} P_1(\tau) \int_\tau^{85} \varphi_\tau(t) dt \Leftrightarrow \sum_{\tau=0}^{84} P_1(\tau) \sum_{j=\tau}^{85-1} \{ A(j) - C(j) \}$		person
仮定	当該 Cohort に属するどの年齢も満 85 歳になる直前まで生存する		$= \sum_{\tau=0}^{84} P_1(\tau) \{ A(\tau, 85) - C(\tau, 85) \} = \Psi_{A-C}(0, 84)$		
⑩	Gofman の絶対過剰罹患人口 $0 \sim$ 満 84 歳の DKS S : Cumu. Spontaneous Incidence よって、 $DKS = \Psi_{A-C}(0, 84)$		$\sum_{\tau=0}^{84} P_1(\tau) \sum_{j=\tau}^{84} DK(\tau, j - \tau) C(j) = \sum_{\tau=0}^{84} P_1(\tau) \sum_{j=\tau}^{84} \{ A(j) - C(j) \}$ $= \sum_{\tau=0}^{84} P_1(\tau) \{ A(\tau, 85) - C(\tau, 85) \} = \Psi_A(0, 84) - \Psi_C(0, 84)$		
参考	$S = \sum_{\tau=0}^{84} P_1(\tau) \sum_{j=\tau}^{84} C(j), KS = \sum_{\tau=0}^{84} P_1(\tau) \sum_{j=\tau}^{84} K(\tau, j - \tau) C(j)$ , 集団被曝線量の総計 : $DP = \sum_{\tau=0}^{84} DP_1(\tau)$				

G-G Approach Gap : リスク係数  $K(\tau, j - \tau)$  は自然罹患率  $C(j)$  と併用しないと、数式中、単独では使えない。

## 付録 I

五つの相関レベルと 10 区分した重回帰レベルの始点 S と標点 T<sup>4)</sup>

(2016 年 10 月末改定)

レベル名		自由度修正済み		始点 S と 標点 T
		相関係数 q	説明係数 Q <sup>2</sup>	
無 相 関	無 為	0.155 i	−0.0240	無為 S
		0.076 i	−0.0058	無為 T
	無 策	0.000	0.0000	無策 S
		0.074	0.0055	無策 T
弱 相 関	微 弱	0.146	0.0213	微弱 S
		0.216	0.0467	微弱 T
	若 干	0.283	0.0801	若干 S
		0.348	0.1211	若干 T
中 相 関	参 考	0.411	0.1689	参考 S
		0.472	0.2228	参考 T
	説 明	0.530	0.2809	説明 S
		0.586	0.3434	説明 T
強 相 関	実 証	0.640	0.4096	実証 S
		0.692	0.4789	実証 T
	確 証	0.741	0.5491	確証 S
		0.788	0.6209	確証 T
高 相 関	実 用	0.833	0.6939	実用 S
		0.876	0.7674	実用 T
	代 用	0.916	0.8391	代用 S
		0.954	0.9101	代用 T
		0.990	0.9801	近 似

注 0) Gauss 型関数に回帰する場合、修正済み相係が Gauss 近似 0.992 を超えるのが望ましい。

注 1) 自由度修正済み相関係数 q は、修正済みの説明係数 Q<sup>2</sup> の平方根で定義される。ただし、Q<sup>2</sup> が負のときは、虚数『i』を用いる。

単発の相関係数 r を自由度で修正するには、 $q^2 = 1 - (n-1)(1-r^2)/(n-2)$  によるが、q の符号は r と一致させる。この付録では負の相関もその絶対値で区分表示している。

注 2) 修正済み相関 q のレベル区分は、「無相関」の始点 0.155 i (−0.155) から「近似の閾値」0.990 までを 10 区分したもので、「無相関」の分岐線 (Dividing Line) 0.000 と発表水準の始点 0.530 と、0.990 を礎点に、レベル始点の相関係数列 {q<sub>i</sub>} の階差数列が公差 −0.009 の等差数列になっている。

注 3) q の二乗である修正済み説明係数 Q<sup>2</sup> を、少数第四位に丸めて表示している。それによる誤差の平均はわずか 17.20 μ である。

注 4) 各レベル名末尾の「S」はレベルの始点を、「T」は当該レベルの回帰を代表する「標」(Triangulation Point) を示している。よって、例えば、「説明」の下限は Q<sup>2</sup> = 0.2809 で、上界は Q<sup>2</sup> = 0.4096 である。

注 5) サンプル数は n ≥ 6 で、説明変数の数 x は  $x \leq (n'-1)/5$  を限度とする。5% ルールによって除外した後のサンプル数 n' は、元々のサンプル数 n に対して、n' ≥ 0.95 n にする。

最小の n = 6 の場合、原点周辺の正六角形の六点のように、Q<sup>2</sup> が −0.2500 になりうる。

注 6) 「高相関」とは、目的変数 φ と説明変数群 x (カイ) とが理論的な関係にないとき起りえない程度に高いレベルの相関関係の意である。

枠外相関	人為 論外	0.500 i	-0.2500	人為 S
		0.322 i	-0.1037	論外 S
		0.155 i	-0.0240	無為 S

付録Ⅱ

重回帰，三種のモデル水準<sup>φ)</sup>と諸元の規格

$$\chi \leq (n' - 1) / 5$$

要件厳格度		項 目	発表水準	知見水準	判定水準	$\alpha$
◎	①	自由度修正済み重相関係数 $q^{\varepsilon)}$	$\geq 0.530$	$\geq 0.640$	$\geq 0.833$	
◎	②	符号マルチコ変数 <sup>β)</sup> の数 $m$	$m \leq 1$	$m = 0$	$m = 0$	
○	③	各説明変数の $t$ 値の絶対値 $ t_i $	$\geq 1.640$	$\geq 1.960$	$\geq 2.575$	$\gamma)$
○	④	各説明変数の $p$ 値 $p_i$	$\leq 10.10\%$	$\leq 5.00\%$	$\leq 1.00\%$	$\gamma)$
△	⑤	各説明変数の標準回帰係数 $ \beta_i $	$\geq 0.081$	$\geq 0.100$	$\geq 0.128$	$\delta)$
参考	⑥	回帰式全体の $P$ 値 $\leq$	9.999 E-12	9.999 E-18	9.999 E-30	
*)	**)	定数項 $\alpha_0$ の $p$ 値, $t$ 値については, 無制約	例えば, E-18 は 10 の-18乗の意			

注\*) ○印は, 充足することが求められる。△印は, ほぼ充足していれば足る。

注\*\*) ①～③を満たせば, ④～⑥はほとんど自動的に成立する。

注α) 発表水準は外部に報告する際の最低基準。知見水準は重回帰モデルが知見をもたらす基準。判定水準は疾病など, 科学的な判定に使えるモデルの基準。ただし, いずれの水準も 5% ルールによる外乱サンプルの除外を前提に, 削除後の説明変数の数  $\chi$  は,  $0.2 (n' - 1)$  が限度である。ここに,  $n'$  は, 除外後のサンプル数である。

注β) 当該説明変数の回帰係数  $\alpha_i$  が, 目的変数  $\phi$  との相関係数  $r_i$  と, 符号が異なるマルチコ変数。

注γ)  $z$  検定における  $t$  値を 1.64, 1.96, 2.575にとると, 両尾の有意水準  $\alpha$  は, 各々 10.10%, 5.00%, 1.00%になる。また,  $p$  値マークには「10.1% 以下, 5.0% 以下, 1.0% 未満」に, 各々「\*, \*\*, \*\*\*」を使用する。

注δ) 標準回帰係数  $\beta_i$  の絶対値について。表示した下限は望まれる基準なので, 0.002 程度の緩和なら許容される範囲である。それゆえ⑤は準要件の「△」としている。

注ε) 上記三つのモデル水準の始点  $q$  は, 各々相関レベル「説明」, 「実証」, 「実用」の始点である。

注φ)  $\phi$  個の変数群  $\Psi$  (プサイ) を主成分得点などで1次結合して, 目的変数  $\phi$  を合成した場合, 説明変数群  $\chi$  (カイ) に対し, 当該回帰式を  $\phi = F(\chi)$  と略記することがある。

注χ) 説明変数の削除に先立ち, 対数を取るなど, 当該変数を順序保存変換して分析すると, モデルの改善が希にみられる。例えば,  $\chi$  に貨幣と物量が混在している際などは, 試みる価値がある。

☆<sub>1</sub>  $\chi$  ope 制約の特約条項

元々のサンプル数  $n$  に対し除外後のサンプル数を  $n' \geq 0.95 n$  に, 説変の数を  $\chi \leq 0.2 (n' - 1)$  に限ることを  $\chi$  ope 制約という。しかし, この制約から  $n \leq 20$  のとき, 外乱サンプルが除外出来ず, また, 説明変数の数も  $\chi \leq 3$  なので, 重回帰分析で得られるべき知見を失う恐れがある。

そこで, サンプルの一つに限ってだけ目的変数の実際値を都合のいい値に変更することを許す条項。これによって,  $n \leq 20$  でも知見の獲得が期待できる。

☆<sub>2</sub> アンケートの感性的回答を目変や説変にする場合, 選択肢に Worst から Best ( $p, q, \dots$ )へ, 次の加重係数を付値すれば,  $Q^2$  の改善がありうる。

三択	$(-6p + q + 6r) / 6$
四択	$(-9p - 2q + 4r + 9s) / 9$
五択	$(-40p - 17q + 4r + 23s + 40t) / 40$
六択	$(-75p - 41q - 9r + 21s + 49t + 75u) / 75$

例えば五択の場合, Worst には-1を, 中立選択肢には0.1を, Best には+1を付値する。この加重係数に関する詳細は文献「判定値」を参照。

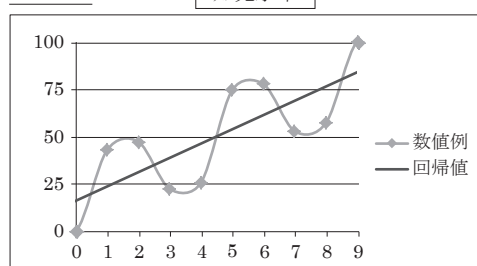
## 付録Ⅲ 『知見水準と近似』の数値例とグラフ

付録Ⅰにある重回帰のレベルは、修正済み相関係数  $q$  を少数第三位まで使って定義し、25 区分している。そこで数値例も「整数解で少数第一位に丸める」ことによって、「論外 S の  $q = 0.322 i$ 」から「近似の閾値  $q = 0.990$ 」まで、誤差  $10 \mu$  以内で作成している。

以下の数値例は、 $F(0) = 0, \dots, F(9) = 100$  の計 10 サンプルで、かつ、座標  $(4.5, 50)$  で点対称になっている。今回、ガウス回帰のために、「G 近似」と「超近似」を設けたが、どの程度の“近似”か、付録Ⅱの「知見水準」と比較し、見た目にも実感できるように、それぞれの数値例とグラフを示しておく。

グラフ 1

知見水準



実証 S

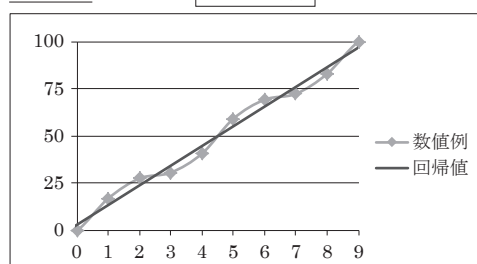
 $q=0.640$   $Q^2=0.40960$   $e=9.1\mu$ 

X th	数値例	回帰値
0	0	18.7
1	47.0	25.6
2	53.0	32.6
3	20.0	39.6
4	22.0	46.5
5	78.0	53.5
6	80.0	60.4
7	47.0	67.4
8	53.0	74.4
9	100	81.3
志向 a	腰高 b	残差 $e_\sigma$
6.96	18.69	21.00

	残差 Max	$\sigma$ 倍 k
	24.52	10.616
	XY 平均	共分散
	282.4	57.40
	X	Y
V	8.25	840
	2.87	28.99
σ	相関係数	説明係数
	0.6894	0.47521
事前	0.6400	0.40961
事後	目標との差	9.06
	$e_{\sigma} - b$	目標との差
	2.310	9.06

グラフ 2

Gauss 近似



G 近似

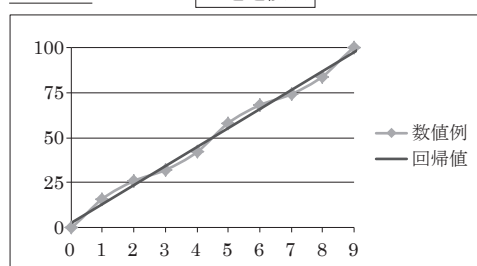
 $q=0.992$   $Q^2=0.98406$   $e=0.2\mu$ 

X th	数値例	回帰値
0	0	3.1
1	17.0	13.5
2	27.6	23.9
3	30.7	34.4
4	40.8	44.8
5	59.2	55.2
6	69.3	65.6
7	72.4	76.1
8	83.0	86.5
9	100	96.9
志向 a	腰高 b	残差 $e_\sigma$
10.43	3.09	3.59

	残差 Max	$\sigma$ 倍 k
	3.99	7.910
	XY 平均	共分散
	311.0	86.01
	X	Y
V	8.25	910
	2.87	30.16
$\sigma$	相関係数	説明係数
	0.9929	0.98583
事前	0.9920	0.98406
事後	目標との差	-0.22
	$e_\sigma - b$	目標との差
	0.504	-0.22

グラフ 3

超近似



超近似

 $q=0.996$   $Q^2=0.99202$   $e=3.7\mu$ 

X th	数値例	回帰値
0	0	2.5
1	16.0	13.1
2	25.9	23.6
3	32.0	34.2
4	41.9	44.7
5	58.1	55.3
6	68.0	65.8
7	74.1	76.4
8	84.0	86.9
9	100	97.5
志向 a	腰高 b	残差 $e_\sigma$
10.55	2.51	2.56

	残差 Max	$\sigma$ 倍 k
	2.93	59.623
	XY 平均	共分散
	312.1	87.06
	X	Y
V $\sigma$	8.25	925
	2.87	30.42
事前 事後	相関係数	説明係数
	0.9964	0.99291
	0.9960	0.99202
	目標との差	3.74
	$e_\sigma - b$	目標との差
	0.049	3.74

# 付録Ⅳ「プログラミング言語 R による実装」

（主要なロジックのみ記載）

```
library(R6)

Modrap <- R6Class("Modrap",
  private = list(),
  public = list(
    # フィールド
    x = NULL,
    y = NULL,
    log_y = NULL,

    nc = NULL, # 説明変数の数
    nr = NULL, # 行数

    n_try = NULL, # 繰り返し回数
    n_para = NULL, # パラメータの数

    solutions = NULL, # 全ての計算結果
    solution = NULL, # 一番成績の良い計算結果
    sol_par = NULL, # 計算結果の係数
    attrrs = NULL, # 結果の属性

    # メソッド
    # 初期化 (コンストラクタ)
    initialize = function(v_x = NA, v_y = NA){
      self$x <- v_x
      self$y <- v_y
      self$log_y <- log(y)

      self$nc <- 2 # 説明変数の数 (ガウス近似では説明変数の個数は2とする)
      self$nr <- length(x) # 行数

      self$n_try <- 100
      self$n_para <- 3
    },

    # 求める関数
    gaussian = function(x = NA, param = NA) {
      exp(-0.5 * ((x - param[2]) / param[1])^2) * param[3]
    },
    # parabola = function(x, param) {
    #   param[1] * (x - param[2])^2 + param[3]
    # },

    # mode="MOD" とすると荷重残差で計算
    calc = function(mode=F){
      # 残差平方和 (residual sum of squares)
      RSS <- function(param) {
        yhat <- self$gaussian(self$x, param)
        sum( (self$y-yhat)^2 )
      }
      # 加重ガウス回帰：ある種の残差平方和 (kind of residual sum of squares)
      RSS_MOD <- function(param) {
        yhat <- self$gaussian(self$x, param)
        sum( BLACK_BOX )
      }

      # 結果の入るリスト
      res_list <- list()
      res_mtx <- matrix(ncol=9, byrow=TRUE)

      count <- 1
      for(i in 1:self$n_try){
        # 係数の初期値ベクトル
        para_list <- runif(self$n_para) * 10 #runif は 1 以下の乱数を返すので、初期値の幅を大きく取るために 10 倍

        # 最適化計算
        if (mode == "RSS_MOD") {
          res <- optim( para_list, RSS_MOD )
        } else {
          res <- optim( para_list, RSS )
        }

        # 相関係数を計算
        corr <- cor(self$y, self$gaussian(self$x, res$par))

        # 結果のリスト
        res_list <- c(para_list, res$par, res$value, res$convergence, corr)
      }
    }
  )
}
```

```
# 行列に追加
res_mtx <- rbind(res_mtx, res_list)

count <- count + 1
}
# 結果をデータフレームに
res_data <- data.frame(st_para1=res_mtx[,1], st_
para2=res_mtx[,2], st_para2=res_mtx[,3], end_
para1=res_mtx[,4], end_para2=res_mtx[,5], end_
para3=res_mtx[,6], value=res_mtx[,7], convergence=res_
mtx[,8], cor=res_mtx[,9])

# 相関係数で降順ソートして、最も相関係数の高い行を抜き出す
self$solutions <- res_data[order(res_data$cor, res_
data$convergence, decreasing = TRUE),]
self$solution <- self$solutions[1,]
self$sol_par <- c(self$solution$end_para1,
self$solution$end_para2, self$solution$end_para3)

# 自由度修正済み説明係数 Q2
Q2 <- 1 - (1-self$solution$cor^2) * (self$nr-1)/(self$nr-
self$nc-1)

# 自由度修正済み相関係数 q
q <- Q2 ^ 0.5

# 解析結果の属性値
self$attrs <- data.frame(cor=self$solution$cor, corr=(1-
self$log_y^0.5), q2=Q2, q=q)
}
)

LoopA <- R6Class("LoopA",
  private = list(),
  public = list(
    # フィールド
    x = NULL, # 計算対象データの x ベクトル
    y = NULL, # 計算対象データの y ベクトル
    func_right = NULL, # 右関数
    func_left = NULL, # 左関数
    current_limit = NULL, # 現時点の最高年齢 (計算対象年齢の上限)
    roots = NULL, # 左右関数の交点の x 座標

    x_limit = NULL, # ループ時にこの年齢を計算の下限とする
    include_points = NULL, # 内包点の x 座標
    include_points_count = NULL, # 内包点の数
    dif_right_left = NULL, # 左右関数の差のベクトル

    # メソッド
    # 初期化 (コンストラクタ)
    initialize = function(v_x = NA, v_y = NA, modrap_right =
NA){
      library(rootSolve)
      self$x <- v_x
      self$y <- v_y
      self$func_right <- modrap_right
      self$current_limit <- max(self$x) # 初期値は元データの最
      高齢

      self$x_limit <- 32 # この年齢より小さくなったら繰り返し計
      算やめ
    },

    setLimit = function(limit) {
      self$current_limit <- limit
    },

    # calc を繰り返し呼んで左関数を求める
    rep_calc = function() {
      result <- 0

      while (result < 0.8) {
        # 計算
        result <- self$calc()
        if(result == F){
          #2 点で交わらない場合
          break
        } else {
          if(result >= 0.8){
            # 内包点の値と回帰値の差の最大値が 0.8 以上
            return(T)
          } else {

```



```

#current_limitを更新
self$current_limit <- max( self$x[self$x <
self$current_limit] ) # 現在の上限より小さい数に限定したベ
クトルの最大値

#self$x_limitより小さくなら繰り返し計算やめ
if(self$current_limit < self$x_limit){
  print("LoopBに進む")
  break
}
}
}
}
return(F) # 収束しなかった場合は False を返す
},

# 左関数を求めるための計算
calc = function() {
  # 現時点の最高齢までの値のベクトルを作る
  x_lim <- self$x[self$x <= self$current_limit]
  y_lim <- self$y[(1:length(self$x))[self$x <= self$current_
limit]]

  # 最大値までの値で左関数を計算
  self$func_left <- Modrap$new(x_lim, y_lim)
  self$func_left$calc("RSS")

  # 交点を求める ( 左関数、右関数の差が 0 となる点 )
  f_diff <- function(x){
    self$func_right$gaussian(x, self$func_right$sol_par)
    - self$func_left$gaussian(x, self$func_left$sol_par)
  }
  self$roots <- uniroot.all(f_diff, c(0,100))

  # 内包点を調べる
  if(length(self$roots)==2) { # 交点が 2 箇所
    self$roots <- sort(self$roots) # 昇順で並べ替え
    x_span <- self$x[self$roots[1]<self$x &
self$x<self$roots[2]] # 交点の間の x 座標を返す
    x_span <- sort(x_span)
    include_points_count <- length(x_span)

    dif_right_left <- c()
    for (i in x_span) {
      # 内包点に対応する回帰値を求める
      left_value <- self$func_left$gaussian(i, self$func_
left$sol_par)
      right_value <- self$func_left$gaussian(i, self$func_
right$sol_par)
      # 内包点の x 座標に対応する y 座標の値
      index <- (1:length(x_lim))[x_lim == i]
      observed_value <- y_lim[index]
      # 差をとってベクトルに入れる
      dif_right_left <- c(dif_right_left, abs(left_value - right_
value)) # 左右関数の差の絶対値
    }
    # 左右関数の差をとり、その最大値が 0.8 以上であれば、そ
の値を記録
    if(max(dif_right_left) >= 0.8){
      self$dif_right_left <- dif_right_left
      self$include_points <- x_span
      self$include_points_count <- length(x_span) # 内包点
の個数 ( 交点の間の x 座標の個数 )
      cat( sprintf(" 左右関数の差が 0.8 以上になりました。( 最
高齢 %2.0f で計算 / 内包点の数 : %2.0f / 最大値 : %3.5f) \
n", self$current_limit, include_points_count, max(dif_
right_left) ) )
    } else{
      # print(dif_right_left)
      cat( sprintf(" 左右関数の差の最大値が 0.8 以上になりませ
ん。( 最高齢 %2.0f で計算 / 内包点の数 : %2.0f / 最大値 :
%3.5f) \n", self$current_limit, include_points_count,
max(dif_right_left) ) )
    }
    return( max(dif_right_left) ) # 左右関数の差を返す
  }

} else{
  cat( sprintf("2 点で交わりません。交点は %1.0f 点です。 \
n", length(self$roots)) )
  return( F )
}
}
}
)

```

```

#CSV からデータを読み込む
#data_path <- "~/dev/modrap/data.csv"
data_path <- "~/dev/modrap/data2.csv"
data <- read.table(data_path, header=T, sep=",")
x <- data$x
y <- data$y

# 元データのデータをプロット
plot(x, y, ylim=c(0, 16), xlim=c(0, 100), type = "b")

# 計算
# 右関数
FFr <- Modrap$new(x, y)
FFr$calc("RSS_MOD")
curve(FFr$gaussian(x, FFr$sol_par), 2, 92, add=TRUE,
col="green")

if(FFr$attrs$q < 0.9800){
  # 自由度修正済み相関係数が 0.9800 より小さい
  # 処理中断
  print(" 自由度修正済み相関係数が 0.9800 より小さいため処理を
中断しました ")
} else if (FFr$attrs$q >= 0.9960){
  # 自由度修正済み相関係数が 0.9960 以上 ( 超近似 )
  # Single Gauss
  print("- Single Gauss -")
  FFr$render("green")
} else {
  # LoopA
  print("- LoopA -")
  loopA <- LoopA$new(x, y, FFr)

  if( loopA$rep_calc() ){
    # 左関数
    FFI <- loopA$func_left
    curve(FFI$gaussian(x, FFI$sol_par), 2, 92, add=TRUE,
col="red")

    # 接続の左右端を求める
    include_point_right <- max(loopA$include_points) # 内包点
x 座標の最大値 ( 内包点右端 )
    include_point_left <- min(loopA$include_points) # 内包点 x
座標の最小値 ( 内包点左端 )

    include_point_index_right <- (1:length(x))[x == include_
point_right]
    include_point_index_left <- (1:length(x))[x == include_
point_left]
    limit <- c(x[include_point_index_left - 1], x[include_point_
index_right + 1])

    # 接続区間と接続区間の関数を求める
    Con <- ConnectionA$new(x, y, limit)

    # 接続関数を全区間でプロット
    curve(Con$ramda_average(x), 2, 92, add=TRUE,
col="blue")

    # 接続した関数をプロット
    FFc <- Combine$new(limit_left=Con$limit_left, limit_
right=Con$limit_right, ffi=FFI, ffr=FFr, ffc=Con)
    #r=0.9939328

    corr <- cor(y, FFc$combined(x))
    Q2 <- 1 - (1-corr^2) * (length(x)-1)/(length(x)-2-1)
    q <- Q2 ^ 0.5
    print(corr)
    print(Q2)
    print(q)

    # 合成関数をプロット
    curve(FFc$combined(x), 2, 92, add=TRUE, col="navy",
lwd=4)
  }

  # 交点をプロット
  par(new=T)
  plot(loopA$roots, FFr$gaussian(loopA$roots, FFr$sol_par),
col="blue", ylim=c(0, 16), xlim=c(0, 100))
}

```