

Coxの生命表回帰モデルについて

○服部 芳明*, 柳井 晴夫**, 高木 広文***, 前田 和甫*
(東京大学医学部*, 千葉大学人文学部**, U.S. N.I.E.H.S. BIOMETRY BRANCH***)

追跡調査データの解析においては, ある事象が対象に起こるか否かだけでなく事象の起こる時刻が問題となることが多い。医学分野のこの種の解析では, 対象とした個体の死亡や特定の症状の発症または回復などのfailureの起こる時刻を, 追跡開始時または初診時における対象者の検査成績等から推定する必要があることがあり, 一般には解析法として, 生命表分析等の方法がよく用いられる。Coxはこのようなデータの解析に, 時間の関数と説明変数の指数関数の積でhazard functionを表わしたモデルの適用を提唱した¹⁾。本研究は, このCoxの回帰モデルの有効性について, シミュレーションを用いて検討を行なった。

<方法>

survival function を $F(t)$ とし, hazard function を $\lambda(t)$ とする ($\lambda(t) = -(dF(t)/dt)/F(t)$) $\lambda(t)$ のモデルとしてCoxは次式を用いた。

$$\lambda(t) = \exp(\mathbf{Z}'\mathbf{b}) \lambda_0(t) \cdots (1)$$

ここで, $\mathbf{Z}$ は説明変数のベクトル, $\mathbf{b}$ はそれらの変数に対するパラメータのベクトルである。

時刻 $t_{(1)}, t_{(2)}, \dots, t_{(n)}$ において failure が観測され, $t_{(i)}$ における failure の個体数を $m_{(i)}$ とする。また, $t_{(i)}$ において failure の起こる可能性のある個体の集合 (risk set) を $R(t_{(i)})$ とし, j 番目の個体にいての観測ベクトルを $\mathbf{Z}_j$ とすると, 総ての $t_{(i)}$ における対数尤度の和 $L(\mathbf{b})$ を, Kalbfleisch と Prentice は次の式で表わした²⁾。

$$L(\mathbf{b}) = \sum_{i=1}^n S_{(i)}' \mathbf{b} - \sum_{i=1}^n m_{(i)} \log \left\{ \sum_{j \in R(t_{(i)})} \exp(\mathbf{Z}_j' \mathbf{b}) \right\} \cdots (2)$$

ここで, $S_{(i)}$ は $t_{(i)}$ において failure の起こった個体の $\mathbf{Z}_j$ の総和とする。Coxによる離散時刻での尤度の式は, $t_{(i)}$ における $R(t_{(i)})$ の要素数 $m_{(i)}$ の総ての部分集合についての計算が必要とされ, 対象例数の多い時には計算機の桁あふれや実行時間の点から考えて実用的でなく, (2) の式を用いることにした。 $\partial L(\mathbf{b}) / \partial \mathbf{b} = 0$ を Newton 法で解いて, $\mathbf{b}$ の最尤推定値 $\hat{\mathbf{b}}$ を求め, Fisher の情報量行列から標準誤差を推定した。

まず, テストデータを2組作り, 係数 $\mathbf{b}$ の再現性をみた。 $\mathbf{Z}$ は TABLE 1, 2 に示した相関係数の6次元のベクトルで, 各変数とも標準正規分布に従う512例とした。hazard function として (3) 式で定義される指数 Weibull モデルを用い³⁾, failure のシミュレーションを行なった。

$$\lambda(t) = KC \exp(\mathbf{Z}'\mathbf{b}) t^{C-1} \cdots (3)$$

定数の値は, $\mathbf{b} = (0.4, 0.4, 0.2, 0.2, 0.0, 0)$, $K = 10^{-3}$, $C = 1.5$ を与え $t = 1.2, \dots, 240$ として failure の起こる時刻を定め, テストデータとし, $\hat{\mathbf{b}}$ の推定を行なった。

次に, 説明変数の内部相関が推定値とその標準誤差に与える影響をみるために, $\mathbf{Z}$ を2次

TABLE 1 Correlations of 6 Explanatory Variables of the Test Data 1

	1	2	3	4	5	6
1	1.000	0.038	0.400	-0.099	0.279	-0.872
2	0.038	1.000	0.688	-0.506	0.463	0.192
3	0.400	0.688	1.000	-0.594	0.607	0.034
4	-0.099	-0.506	-0.594	1.000	0.024	-0.295
5	0.279	0.463	0.607	0.024	1.000	-0.130
6	-0.872	0.192	0.034	-0.295	-0.130	1.000

TABLE 2 Correlations of 6 Explanatory Variables of the Test Data 2

	1	2	3	4	5	6
1	1.000	0.000	-0.200	0.200	0.250	0.150
2	0.000	1.000	0.100	0.300	0.200	0.200
3	-0.200	0.100	1.000	-0.100	0.200	-0.300
4	0.200	0.300	-0.100	1.000	0.300	0.050
5	0.250	0.200	0.200	0.300	1.000	-0.200
6	0.150	0.200	-0.300	0.050	-0.200	1.000

元で, 各々変数が標準正規分布に従うとし, TABLE 5 に示す10通りの相関係数について, それぞれの場合100例のデータの組を10組ずつ発生させ, (3) 式で failure の時刻を定めた。定数は, $\mathbf{b} = (0.5, 0)$, $K = 5 \times 10^{-3}$, $C = 1.5$ とし, $t = 1.2, \dots, 60$ について発生させたデータで $\hat{\mathbf{b}}$ を推定した。

<結果>

6つの説明変数による2通りのテストデータの $\hat{\mathbf{b}}$ の推定値, 標準誤差, t 値は TABLE 3, 4 に示した通りである。テストデータ2の方がテストデータ1より, 仮定した値に推定値が近く標準誤差も小さい。テストデータ2では推定値は仮定した $\mathbf{b}$ の値の付近にあり, テストデータ1でも第1変数と第6変数の係数以外はほぼ仮定した値の近くに推定値がある。

相関係数を10通りに定めた各々10組のデータについての $\hat{\mathbf{b}}$ の推定値の平均値と標準偏差, 及び, 標準誤差の推定値の平均値と標準偏差を TABLE 5 に示した。2つの説明変数間の相関が高くなるにつれて, $\hat{\mathbf{b}}$ の推定値の標準偏差と標準誤差の平均値は大きくなる傾向がみられるが, $\hat{\mathbf{b}}$ の推定値の平均値には一定の傾向がみられなかった。

<考察>

6変数によるテストデータでは, テストデータ1の内部相関が高かったため, 係数の推定値の誤差が大きくなったと考えられる。テストデータ2の内部相関は比較的強く, 推定値の精度がよかった。一般の場合には, 相関はテストデータ1の場合ほど高くなく, 相関の高い場合は変数選択を行ない得るので, この方法は実用的と考える。

10通りの相関係数によるデータの標準誤差の比較による結果, 100例前後では相関係数0.5以下の場合, 推定値の標準誤差は大体変わらないように見える。例数が多ければ相関がもっと高くても推定の精度は低くならないと考えられる。しかし, 変数が多くなると説明変数間の相関は低くなければ推定の精度は低くなると推定される。

このCoxの方法は, 現在, 広く用いられているロジスティック・モデルと異なり, 事象の生起の有無だけでなくその時刻についても解析が可能である。今回は行なわなかったが, $\hat{\mathbf{b}}$ から hazard function を求める方法もCoxは提唱している。指数 Weibull モデルを用いた解析の場合は $\hat{\mathbf{b}}$ と Weibull 分布のパラメータ C との同時推定が可能であるが, Coxの方法では hazard function は Weibull 分布のそれである必要がないことが利点であろう。

TABLE 3 Estimated Coefficients of the Test Data 1

	COEFFICIENT	S.E.	t-VALUE
1	0.0898	0.3581	0.25
2	0.3854	0.0965	4.00
3	0.3418	0.1962	1.74
4	0.1363	0.0971	1.40
5	-0.0439	0.0770	-0.57
6	-0.3581	0.3301	-1.09

TABLE 4 Estimated Coefficients of the Test Data 2

	COEFFICIENT	S.E.	t-VALUE
1	0.3897	0.0485	8.04
2	0.3564	0.0471	7.57
3	0.2949	0.0533	5.53
4	0.1947	0.0506	3.85
5	-0.0226	0.0529	-0.43
6	-0.0349	0.0487	-0.72

TABLE 5 Means and Standard Deviations of 10 Trials of Estimated Coefficients and Standard Errors of 2 Variables

r	coefficient of var.1				coefficient of var.2			
	estim.value	s.e.	mean	s.d.	estim.value	s.e.	mean	s.d.
0.01	0.470	0.093	0.117	0.012	0.007	0.120	0.112	0.015
0.1	0.514	0.102	0.121	0.009	-0.009	0.094	0.114	0.014
0.2	0.434	0.089	0.118	0.011	-0.008	0.066	0.110	0.009
0.3	0.529	0.129	0.128	0.015	0.053	0.069	0.112	0.004
0.4	0.575	0.148	0.130	0.011	-0.013	0.113	0.117	0.011
0.5	0.551	0.107	0.135	0.018	-0.015	0.104	0.124	0.016
0.6	0.470	0.137	0.142	0.016	0.008	0.158	0.137	0.013
0.7	0.483	0.133	0.167	0.010	0.012	0.157	0.157	0.013
0.8	0.569	0.185	0.188	0.018	-0.085	0.182	0.180	0.015
0.9	0.479	0.250	0.256	0.023	0.080	0.213	0.250	0.022

<文献>

- 1) D.R.Cox: J.R.Stat.Soc., Series B, 34, 187-202 (1972)
- 2) J.Kalbfleisch and R.L.Prentice: J.R.Stat.Soc., Series B, 34, 215-216 (1972)
- 3) A.R.Dyer: J.Chron.Dis., 28, 565-578 (1975)

肝がん・肝硬変死亡と関連要因に関する重回帰分析

—山梨県における市町村別資料を使用して—
・稲葉 裕*, 高木 慶文**, 柳井 晴夫*** (順天堂・医*, 東大・医**, 千葉大・人文***)

【目的】 山梨県は東日本において、肝がん・肝硬変の死亡率が著しく高いことが知られている。しかし、山梨県内のこれらの肝疾患死亡率の市町村別分布をみると、決して一様ではなく、特に日本住血吸虫症（以下日住症という）の過去の受療率との間に、肝がん・肝硬変死亡率と比較的高い正の相関が認められることとすでに報告した^(1,2)。その後、これらの肝疾患の発生に関連すると考えられるHBs抗原の市町村分布がより明確になり、前述のデータの見直しも必要とされてきたので、今回、あらためて、肝がん・肝硬変死亡率に影響を与えていると考えられる諸指標を選び、重回帰分析の手法を用いて、再検討を試みた。

【方法】 肝がん・肝硬変死亡率は、昭和43~47年の平均値、昭和45年の全国人口を基準人口として、年令訂正死亡率を間接法により求めたものである。説明変数としては、表1に示す21の環境要因、死因別死亡率などより上げた。計算は東大大型計算機センターのSPSS-V6、重回帰分析（一括多重投入法）⁽³⁾によった。

【結果および考察】 表1に示すように、肝硬変死亡率は、男女とも日住症との地域相関がきわめて強く、重回帰分析の結果でも、標準化偏回帰係数（以下 β ）が男0.514、女0.954、F値が男11.522、女57.992と非常に高値を示した。（重相関係数は男0.832、女0.888）このことから、地域相関がみられる限り、肝硬変死亡率は、日住症受療率でよく説明できると考えられる。もちろん、これをただちに、医学的・生物学的に、日住症が肝硬変の主たる病因であることの証明に結びつけることはできないが、強い関連の存在を示唆するものといえる。

一方、肝がん死亡率に関しては、表2に示すように、64市町村で見ると、 β の比較的大きい指標は、男では高血圧死亡（脳卒中を含む）、流入人口、薬局数（正）や出生率、診療所数（負）であり、女では、第一次～第三次産業人口の割合、高血圧死亡（正）、胃がん死亡、出生率（負）である。これらで関連があると考えられてきたHBs抗原や日住症の数値はあまり高くなく、女ではむしろ負

表1 肝がん・肝硬変死亡率と関連諸指標の相関係数

	肝がん		肝硬変	
	男	女	男	女
HBs抗原	0.153	-0.001	0.023	0.130
第一次産業	0.074	0.168	0.150	0.162
第二次産業	-0.146	-0.040	-0.106	-0.018
第三次産業	0.042	-0.235	-0.129	-0.251
出生	-0.035	-0.147	0.208	0.036
死亡	0.033	0.165	0.120	0.069
人口	-0.058	-0.138	-0.346	-0.358
上水道普及率	-0.087	-0.152	-0.037	0.032
乳がん死亡	0.056	-0.089	-0.036	-0.164
医師	0.078	-0.029	0.162	0.035
診療所	-0.147	-0.152	-0.113	0.048
薬局	0.176	-0.086	-0.194	-0.185
看護婦	-0.044	-0.125	-0.063	-0.082
高血圧死亡	0.012	0.033	-0.014	-0.029
胃癌	-0.014	-0.021	-0.017	-0.013
胃がん	-0.030	0.018	0.016	-0.051
肺がん	-0.017	-0.059	0.081	0.027
子宮がん	0.027	0.038	0.048	0.036
結核	-0.052	-0.017	-0.129	-0.113
流入人口	0.310	0.059	0.526	0.422
日本住血吸虫症	0.303	0.083	0.674	0.849

(山梨県64市町村について)

の値をとっている。重相関係数もあり高い値ではない。

ところで、山梨県の市町村別人口数には、かなり差があり、使用している指標がすべて人口あたりについての数値であるため、人口数の少ない所では偏りが生じやすい。そこで男女いづれかの人口が2000人以下の町村について、計算から除外し、残り45市町村に関する分析を行なってみた。結果は表2の右欄に示すとおりである。すなわち、男女とも第一次～第三次産業人口の割合が、大きな β 値を示し、男ではそれらに次いで高血圧死亡（正）、胃がん死

表2 山梨県肝がん重回帰分析の標準化偏回帰係数

	64市町村		45市町村	
	男	女	男	女
HBs抗原	0.271	-0.003	-0.020	0.517
第一次産業	0.091	5.528	1.744	14.407
第二次産業	-0.181	3.697	0.939	8.809
第三次産業	0.121	2.832	0.820	8.181
出生	-0.400	-0.486	-0.182	-0.400
死亡	-0.095	0.254	0.057	0.133
人口	0.170	-0.143	0.287	-0.012
上水道普及率	0.025	0.151	0.197	0.081
乳がん死亡	0.070	-0.076	-0.013	0.348
医師	0.068	0.265	-0.142	0.299
診療所	-0.333	-0.327	-0.070	-0.097
薬局	0.335	-0.024	0.242	-0.153
看護婦	-0.193	-0.253	0.053	-0.509
高血圧死亡	0.520	0.653	0.813	0.290
胃癌	0.107	0.302	-0.173	0.372
胃がん	-0.238	-0.538	-0.754	0.147
肺がん	-0.088	-0.178	-0.219	-0.321
子宮がん	-0.091	-0.111	0.329	-0.319
結核	-0.082	-0.103	-0.235	-0.094
流入人口	0.427	0.313	0.147	0.136
日本住血吸虫症	0.285	-0.290	0.403	0.087
重相関係数	0.662	0.559	0.701	0.754

亡（負）が大きく、女では、HBs抗原（正）、看護婦（負）が大きな値を示す。産業人口の占める比率が大きくなった背景には、人口の少ない町村における産業の偏り（第一次産業が多い町村とされる）が、とり除けられた影響も無視できない。しかし、単純相関では互いに逆符号の相関を示す第一次産業と第二次産業が、同じ符号の高い β 値を示すことは興味深い。肝がんと職業との関係は、今後詳しく調べてみる価値があるように思う。またHBs抗原は、 β 値では4番目の数値になっているが、F値では最も高く（5.606）、市町村（女）別の分布でみるかぎり、女の肝がん死亡に関連しているようにみえる。HBs抗原の分布は、県の血液センターの献血者についての資料であり、男女の区別がされていないため、厳密な意味での市町村分布を表わしているとは言えない面がある。今後、変数追加法による分析などによってさらに検討を加えていく予定である。

【結論】 山梨県の市町村を単位とした肝がん・肝硬変と関連要因の重回帰分析を行った。肝硬変に関しては、これらと同様、日本住血吸虫症の影響が非常に強いことが確認された。肝がんに関しては、産業人口割合の影響が強く、職業との関係も追究する必要性が認められた。HBs抗原については、女の肝がんへの影響が認められた。

【文献】

- 1) 稲葉 裕, 丸地 信弘 et al., 日本公衛誌 24(8) 465-473, 1977
- 2) 稲葉 裕, 高橋 月容 et al., 日本公衛誌 24(12) 811-815, 1977
- 3) 三宅 一郎, 中野 嘉弘 et al., SPSS統計パッケージ, 東洋経済新報社, 1977
- 4) 神代 正道, 中島 敏郎 et al., 肝臓 15(12) 771-778, 1974

(なおこの研究の一部は、厚生省がん研究助成金「がんの疫学的研究」班の補助を受けた。)

追加・訂正資料

肝がん・肝硬変死亡と関連要因に関する重回帰分析

箱景祐, 高木廣文, 柳井晴夫

(追加)

図1

山梨県市町村別重回帰分析 肝がん (男) 64市町村

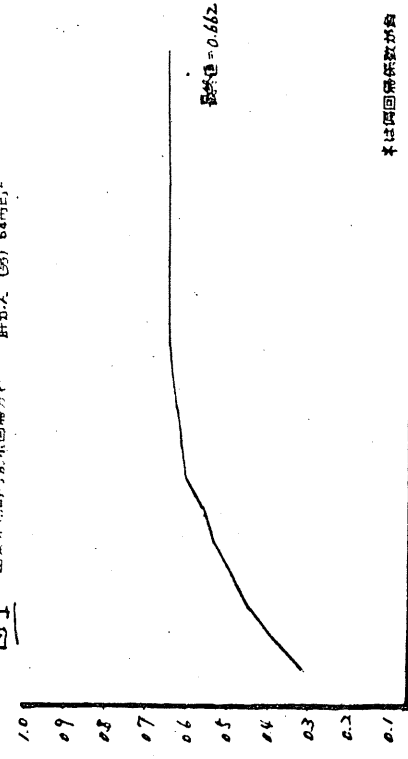


図3

山梨県市町村別重回帰分析 肝がん (女) 64市町村

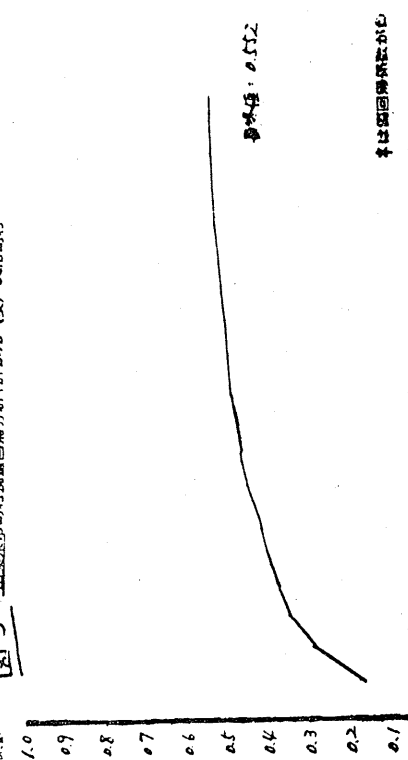


図2

山梨県市町村別重回帰分析 肝がん (男) 45市町村

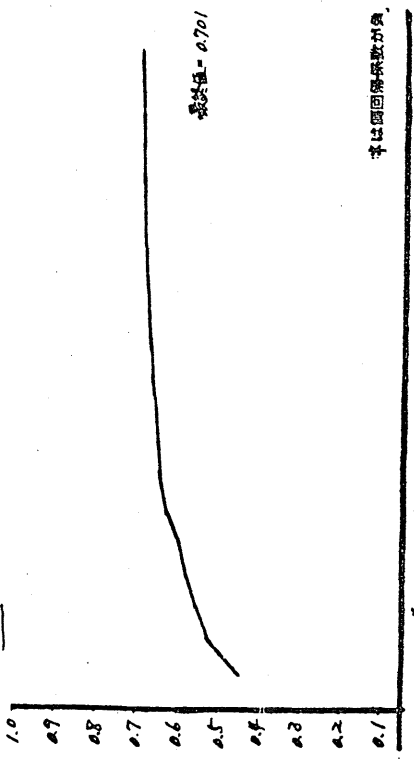


図4

山梨県市町村別重回帰分析 肝がん (女) 45市町村

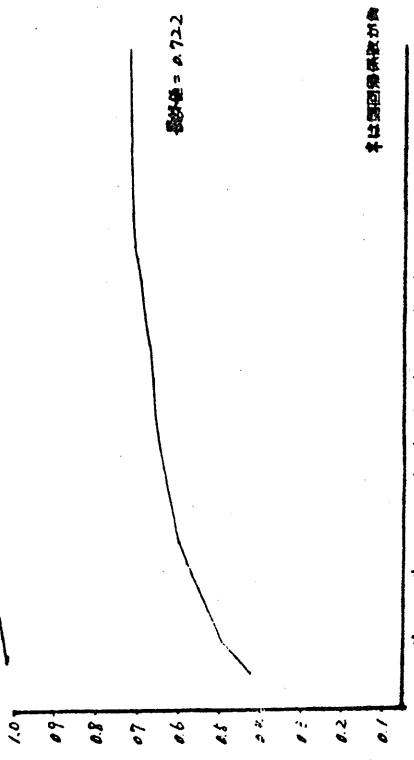


表2

山梨県肝がん重回帰分析の標準化偏帰係数

	64市町村		45市町村	
	男	女	男	女
人口	0.271	0.007	-0.016	0.556
第2次産業	-0.115	0.715	0.227	-0.731
出生	-0.319	0.470	-0.027	-0.834
死亡	-0.400	-0.478	-0.182	-0.401
乳がん	-0.094	0.268	0.062	0.181
子宮がん	0.169	-0.155	0.293	0.047
高齢者	0.025	0.134	0.198	-0.072
医師	0.070	-0.065	-0.009	0.388
肝がん	0.067	0.233	-0.144	0.280
肝硬変	-0.333	-0.325	-0.063	-0.029
上水道	0.335	-0.032	0.230	-0.273
下水道	-0.192	-0.230	0.041	-0.629
乳がん	0.524	0.730	0.805	0.204
子宮がん	0.107	0.312	-0.165	0.453
高齢者	-0.239	-0.556	-0.752	0.165
医師	-0.088	-0.178	-0.225	-0.383
肝がん	-0.094	-0.177	0.332	-0.287
肝硬変	-0.083	-0.145	-0.244	-0.188
乳がん	0.426	0.295	0.150	0.170
子宮がん	0.285	-0.276	0.396	0.011
最終値	0.662	0.552	0.701	0.722

表2. 山梨県肝がん重回帰分析の標準化偏帰係数

(訂正)