

サポートベクター回帰における感度分析による 変数選択の有効性の検証

—都道府県別全死因死亡率の影響要因の分析—

田辺 和俊¹・鈴木 孝弘²

(受付 2019 年 6 月 1 日；改訂 8 月 23 日；採択 12 月 26 日)

要 旨

本研究では、筆者らが開発したサポートベクター回帰における変数選択としての感度分析法を都道府県別死亡率の要因分析に適用し、その有効性の検証を行った。わが国では少子高齢化社会の次に到来するのが多死社会とされ、医師不足、終末期医療、孤独死、死に場所難民、放置空き家等、さまざまな社会問題の発生が危惧されている。地域の自治体が多死化対策を行うためには、多数の要因の中から死亡率に重大な影響を与える要因の解明とその相対的影響度の推定が重要になる。そこで、各都道府県の全死因年齢調整死亡率を目的変数、その影響要因の候補として生活習慣、医療・福祉、社会・経済の 3 分野の指標 56 種を説明変数とし、サポートベクター回帰により解析し、感度分析により影響要因を探索した。その結果、都道府県別死亡率を統計的に有意な精度で再現する 11 種の影響要因が得られ、喫煙率や高齢単身率だけでなく、これまで未検証の社会福祉士数が死亡率に大きな影響を与えることを見出した。また、地域の多死化対策への試みとして、死亡率が最高の青森県について、死亡率が最小の長野県との対比において影響要因の指標値に基づいて死亡率低減策を考察した。以上の結果から、広範な分野の多数の説明変数の中から感度分析により影響要因を探索する解析手法の有効性を実証した。

キーワード：サポートベクター回帰、変数選択、感度分析、全死因死亡率、影響要因。

1. はじめに

わが国では平均寿命が延び続ける一方で少子化に歯止めがかかっていないが、少子高齢化社会の次に到来する問題として危惧されているのが多死社会である。「平成 29 年版高齢社会白書」によれば、2010 年に約 120 万人であった年間死亡者数は、団塊世代が 80 歳代後半となる 2030 年代には 160 万人を超え、2040 年には最多の 168 万人に達すると予想されている。その影響として、医師不足、終末期医療、孤独死、死に場所難民、放置空き家等、さまざまな社会問題の発生が危惧されている。そのため、国は 2025 年をめどに地域包括ケア体制の構築に取り組み始めた。このシステムでは、本格的な高齢社会において地域全体で高齢者を支えるため、保健・医療・福祉関係者や、地域の各種団体や住民が連携し、介護、予防、医療、生活支援、住まいのサービスを一体化して提供していく地域づくりに取り組むとしている。

¹ 東洋大学 現代社会総合研究所：〒112-8606 東京都文京区白山 5-28-20

² 東洋大学 経済学部：〒112-8606 東京都文京区白山 5-28-20

ところで、20 世紀後半以降、人類の健康や病気が遺伝等の先天的要因だけでなく、経済、社会、文化、環境等の後天的要因によっても影響されることが広く認められるようになった。WHO の報告書 (2016) “The determinants of health” は健康格差の社会的影響要因として社会格差、ストレス、社会的排除、労働、失業、社会的支援等の 10 項目を挙げている。わが国の健康格差については、所得や失業等の経済・労働要因、学歴等の教育要因、病院や医師等の医療資源要因の影響が大きいとされている (福田・今井, 2007; Kagamimori et al., 2009; 日本公衆衛生学会, 2011; 橋本, 2012; 近藤, 2014)。日本学術会議は健康の社会格差の現状に関する報告書 (2011) を発表し、保健医療福祉政策に健康格差を考慮するよう提言した。

このように死亡率には非常に多くの要因が影響すると考えられるため、国や地域自治体が多死化対策を行うためには、多数の要因の中から死亡率に重大な影響を与える要因の解明とその相対的影響度の推定が重要になる。その手法として、地域の死亡率のデータを目的変数、いくつかの指標を説明変数として重回帰分析する生態学的分析手法 (Rothman, 2012) がある。この手法を用いて、これまでは特定地域の死亡率を解析した研究は多いが、地域別の死亡率について要因の解明を行った研究は少ない。

たとえば、Araki and Murata (1986) は都道府県別の全死亡率について学歴、所得、失業率等を説明変数に用いて重回帰分析した。Fukuda et al. (2004) は全国の市区町村の全死亡率を対象とし、説明変数に失業率、学歴、所得等を用いて解析した。大坪 他 (2009) は医師数、病床数、高齢世帯数等を説明変数に用いて解析した。しかし、これまでの研究では特定の分野の比較的小数の説明変数を用いて線形回帰分析 (OLS) により解析しているため、回帰決定係数が低く、死亡率の要因解明は未だ十分ではない。

本研究では、地域の多死化対策に資する情報提供を目的として、都道府県別の死亡率を目的変数とし、多種多様な指標を説明変数として非線形重回帰分析により解析し、死亡率の影響要因を解明する実証研究を試みた。本研究の意義は、先行研究での説明変数の不足を考慮して広範囲の多種多様な説明変数の中から影響要因を探索したこと、先行研究で多用されている線形重回帰分析の不十分さを改善するために、非線形回帰分析の 1 手法であるサポートベクター回帰 (SVR) を適用したこと、および多数の説明変数の中から影響要因を探索する変数選択の手法として、筆者らが開発した感度分析法を適用しその有効性を検証したことの 3 点である。

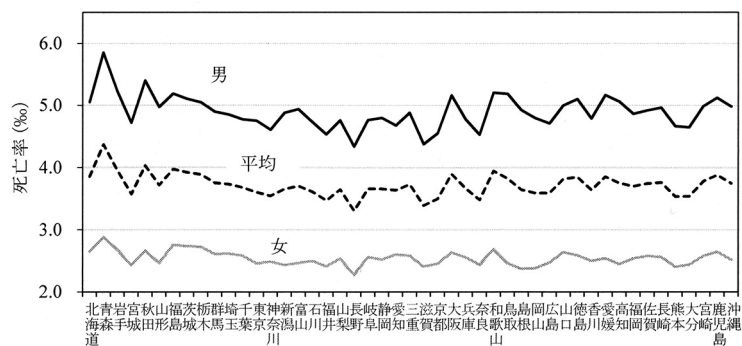
2. データと方法

2.1 死亡率データ

死亡率の影響要因を重回帰分析によって解明する手法を設計する際、3 つの検討すべき点がある。第 1 は目的変数に用いる死亡率の内容である。「平成 28 年人口動態統計」による日本人の死因は表 1 のように男女で異なり、また、年齢階層でも異なる。これらの死因の要因として喫煙、飲酒、肥満等の生活習慣の影響が大きいことが判明している。そのため、疾患別の死亡率について、生活習慣指標等を説明変数に用いて回帰分析を行った研究は多い。たとえば、筆者らは肺がん死亡率の解析から喫煙率や脂質摂取量等が重大要因であることを見出した (田辺 他, 2016)。

しかし、地域自治体にとってはそのような多種類の死因に関する影響要因の情報は多死化対策の観点からは実施上の困難さがあり、むしろ、できるだけ簡潔かつ包括的な情報の方が施策には有用であると考えられる。そこで、目的変数には「平成 29 年人口動態調査特殊報告」による 2015 年の都道府県別の全死因年齢調整死亡率 (以下、死亡率) の男女平均を用いることにした。死因別の死亡率でなく全死因の死亡率の採用については、死因により影響要因に違いがあると考えられることから、また、死亡率の男女平均の採用については、男女間で死因に違いが

	男女合計		男		女	
	死因	割合(%)	死因	割合(%)	死因	割合(%)
1	悪性新生物	28.5	悪性新生物	32.6	悪性新生物	24.2
2	心疾患	15.1	心疾患	13.8	心疾患	16.5
3	肺炎	9.1	肺炎	9.7	老衰	11.0
4	脳血管疾患	8.4	脳血管疾患	7.8	脳血管疾患	8.9
5	老衰	7.1	老衰	3.4	肺炎	8.5
6	不慮の事故	2.9	不慮の事故	3.3	不慮の事故	2.6
7	腎不全	1.9	自殺	2.2	腎不全	2.0
8	自殺	1.6	慢性閉塞性肺疾患	1.9	大動脈瘤及び解離	1.4
9	大動脈瘤及び解離	1.4	腎不全	1.8	血管性等の認知症	1.3
10	肝疾患	1.2	肝疾患	1.5	アルツハイマー病	1.3



あることから、疫学的観点からは異論があると思われる。しかし、本研究では地域の多死化対策に有用な要因を探ることが目的であるため、男女別の全死因年齢調整死亡率の単純平均を採用した。

男女平均死亡率の地域差をみると、青森県が4.37%と突出して高く、最少(3.31%)の長野県との差は大きい。地域別の傾向では、宮城県を除く東北5県の死亡率が全般的に高く、隣接の北関東も高い。一方、南関東と中部地方は全国平均を下回るが、西日本は全国平均を超える府県が多い。また、近畿地方のように死亡率が高い府県と低い府県が隣接している地域もある。このように複雑な地域格差は多種多様な要因が死亡率に関与している可能性を示唆する。

重回帰分析の第2の検討点は説明変数の選定であり、一般に2つのアプローチがある。1つは目的変数に関する理論または仮説を立て、それに基づく比較的少数の説明変数を設定し、多くの場合、説明変数を固定して重回帰分析を行う理論主導アプローチである。この手法の目的は、設定した理論や仮説が目的変数データをどの程度説明するかの検証に重点が置かれる。しかし、研究者により理論や仮説が異なるため、解析で得られた結果が異なることが多い。また、

説明変数が少数なため、一般に回帰分析の決定係数が低く、目的変数データを十分再現できないことが多い。

一方、これと逆に、目的変数との関連にはこだわらず、入手可能な多数の説明変数を用いて解析し、その中から変数選択法を用いて目的変数を再現する必要最小限の説明変数を探索するデータマイニング的アプローチが考えられる。この手法は、目的変数に関する原理や機構が未解明であるが、社会政策上の要請から目的変数の予測システムの開発が急務であるような問題の場合に採用される。この問題例としては化学物質の発がん性予測を検証した筆者らの研究(Tanabe et al., 2010)がある。したがって、この手法の目的は解析に用いた既存データだけでなく、未知データに対しても高精度の結果を与える予測システムを開発することであり、有効な説明変数をいかに探索するかに力点が置かれる。しかし、得られた影響要因のすべてについて目的変数との関連性を必ずしも説明できない場合がありうる。

本研究ではこれら2つのアプローチの中間的手法を取る。すなわち、死亡率との関連が想定され、これまで検証されてきた多数の指標を説明変数に用いて重回帰分析により一括解析し、それらの中から変数選択法により死亡率の影響要因を探索する実証的手法である。そこで、説明変数としてこれまで検証された指標は100種以上に上るが、本研究ではそれらの中で検証例が多いものとして、喫煙・摂取食品・肥満等の生活習慣指標、医師・保健師・健診等の医療・福祉指標、所得・学歴・就業等の社会・経済指標を選定した。さらに、健康への効果が報告されているが、これまで回帰分析の説明変数として検証されていない数種の新規指標を追加した。

以上の方法により選定した説明変数56種を表2に示す。これらの都道府県別データは死亡率のデータが2015年時点であるため、各種政府統計の2010年～2015年の調査データを使用した。総数が多いため、それらの指標の記述統計量の一部のみ表3に示したが、分布の非対称性が高く、線形回帰(OLS)では統計的に有意な結果を得ることが困難と推測される変数が多い。なお、指標の単位が異なり、また下記の感度分析のために、各指標は最小と最大が0と1になるよう正規化して解析に用いた。

これらの説明変数を採用した理由を簡潔に記すと、生活習慣分野では、死亡率の中で大きな割合を占めるがんや、心・脳血管疾患(表1)等の主因とされている喫煙、飲酒、塩分、肥満、高血圧、ストレスなどを用いた。さらに、これらの疾患予防に効果が期待される歩数、スポーツなども追加した。医療・福祉分野では、死亡率の低下が期待できる医療費、医師、保健師、健診、がん検診(胃・肺・大腸・子宮・乳がんの5種の平均受診率)を採用した。さらに、特に高齢者の健康に貢献すると思われる指標として、老人医療費、訪問介護員、老人ホーム、介護施設などを採り上げた。社会・経済分野では、高齢ほど死亡率が高いことから、高齢者に関連する指標を採用した。失業者や貧困者は健康意識が低く死亡率が高いとされることから、所得や就業に関連する指標を採用した。また、多雪地域は多種の面での住民サービスの低下が死亡率の向上を招くと考えられることから、年間の降雪日を採用し、また、長野県のように山間地域が多い県民は足腰が鍛えられることで死亡率が低いとされていることから、標高を追加した。さらに、人口集中地区ほど医療や福祉が充実し、死亡率が低いと考えられることから、人口密度や都市化率を追加した。

2.3 解析方法

死亡率の影響要因探索における第3の検討事項は重回帰分析の手法である。死亡率要因分析の先行研究では、最も一般的な線形重回帰分析(OLS)が多用されてきた。しかし、表3に示すように、本研究の説明変数の中には、死亡率に対して非線形関係にあるものが多いため、OLSでは統計的に有意な結果を得ることが困難と思われる。

先行研究では、このような非線形関係に対処するために、一部の説明変数の2乗の項の追加

表 2. 説明変数の定義、およびデータの出所。出所：国生：国民生活基礎調査（2013）、国健：国民健康栄養調査（2012）、特健：特定健康診査（2010）、社統：社会生活統計指標（2013）、家計：家計調査（2013）、社基：社会生活基本調査（2014）、衛行：衛生行政報告（2014）、社福：社会福祉振興・試験センター（2015）、社人：社会・人口統計体系（2014）、筆者：田辺・鈴木（2018）、賃金：賃金構造基本統計調査（2010）。

分野	説明変数	定義(単位)	出所
生活習慣	喫煙	喫煙者の割合(%)	国生
	飲酒	飲酒者の割合(%)	国生
	エネルギー	1日の平均エネルギー摂取量(Kcal)	国健
	脂質	1日の平均脂質摂取量(g/日)	国健
	野菜	1日の平均野菜摂取量(g/日)	国健
	塩分	1日の平均食塩摂取量(g/日)	国健
	規則食事	規則正しく食事している人の割合(%)	国生
	バランス食事	バランスの取れた食事をしている人の割合(%)	国生
	薄味食事	薄味の食事をしている人の割合(%)	国生
	適量食事	食べ過ぎないようにしている人の割合(%)	国生
	健康無関心	日頃健康のために特に何もしない人の割合(%)	国生
	肥満	BMI (Body Mass Index) の平均(kg/m ²)	国健
	高血圧	収縮期血圧 140mmHg 以上の該当者の割合(%)	特健
	メタボリックシンドローム	メタボリックシンドローム該当者の割合(%)	特健
	ストレス	悩みやストレスのある人の割合(%)	国生
	歩数	1日の平均歩数(歩)	国健
	スポーツ	スポーツの年間行動者率(%)	社統
	交際	世帯当たりの交際費年間支出金額(千円)	家計
	趣味	趣味・娯楽の高齢行動者の割合(%)	社基
医療・福祉	保健医療費	世帯当たりの保健医療費年間支出金額(千円)	家計
	老人医療費	被保険者当たりの後期高齢者医療費(千円)	社統
	老人福祉費	高齢人口当たりの老人福祉費(千円)	社統
	医師	人口当たりの医療施設に従事する医師数(人)	社統
	保健師	人口当たりの保健師数(人)	衛行
	社会福祉士	人口当たりの社会福祉士数(人)	社福
	介護福祉士	人口当たりの介護福祉士数(人)	社福
	訪問介護員	人口当たりの訪問介護員数(人)	社人
	老人ホーム	高齢人口当たりの老人ホーム数(所)	社統
	老人福祉センター	高齢人口当たりの老人福祉センター数(所)	社統
	介護老人福祉施設	高齢人口当たりの介護老人福祉施設数(所)	社統
	介護療養型医療施設	高齢人口当たりの介護療養型医療施設数(所)	社統
	高齢生活保護	高齢人口当たりの生活保護高齢者数(人)	社統
	健診	健診の受診率(%)	国生
	がん検診	5種のがん検診の平均受診率(%)	国生
社会・経済	高齢世帯	高齢者のいる世帯の割合(%)	社統
	単身世帯	単身世帯の割合(%)	社統
	高齢単身	高齢単身世帯の割合(%)	社統
	学歴	平均就学年数(年)	社統
	社会教育費	人口当たりの社会教育費(千円)	社統
	公民館	人口当たりの公民館数(館)	社統
	高齢者講座	人口当たりの高齢者講座数(講座)	社統
	高齢ボランティア	高齢ボランティア活動者の割合(%)	社基
	家計収入	世帯の平均月間純収入(千円)	家計
	貯蓄	世帯の貯蓄現在高(千円)	社統
	負債	貯蓄額に対する負債額の比率	社統
	就業	就業者の割合(%)	社統
	高齢就業	高齢就業者の割合(%)	社統
	失業	完全失業率(%)	社統
	貧困率	貧困層の比率を示す指数(%)	筆者
	残業	超過労働時間(時間)	賃金
	標高	都道府県所在地の標高(m)	社統
	気温	年間の平均気温(℃)	社統
	日照	年間の日照時間(時間)	社統
	降雪	年間の平均降雪日数(日)	社統
	人口密度	可住地面積当たりの人口密度(人/km ²)	社統
	都市化	人口集中地区人口割合(%)	社統

表 3. 説明変数の記述統計. *死亡率との単相関係数.

分野	説明変数	最小	平均	最大	標準偏差	歪度	相関係数*
生活習慣	喫煙	163	20.5	266	2.1	0.62	0.474
	飲酒	22.8	26.4	31.2	1.9	0.31	0.182
	エネルギー	1,843	2,059	2,143	46.2	-2.54	-0.140
	脂質	52.0	58.8	64.0	2.5	-0.27	-0.315
	野菜	233	280	344	21	0.67	0.024
	塩分	8.6	10.0	10.9	0.5	-0.36	0.078
	規則食事	44.0	52.6	58.9	3.2	-0.45	-0.043
	バランス食事	30.8	35.9	41.7	2.4	-0.08	-0.384
	薄味食事	26.0	31.6	35.9	2.2	-0.25	0.062
	適量食事	32.5	38.4	41.5	1.8	-0.86	-0.005
	健康無関心	13.2	16.0	18.1	1.2	-0.51	0.333
	肥満	22.7	23.3	24.4	0.4	0.91	0.374
	高血圧	16.3	19.6	24.2	1.9	0.70	0.389
	メタボリックシンドローム	12.8	14.7	17.7	1.1	0.66	0.332
	ストレス	44.0	47.5	51.1	1.8	0.17	-0.390
	歩数	5,743	7,034	8,048	514	0.14	-0.354
	スポーツ	11.2	27.0	49.9	8.1	0.30	-0.250
	交際	81	140	185	24	-0.19	-0.231
	趣味	22.2	26.3	30.7	2.3	-0.21	-0.544
医療・福祉	保健医療費	81	124	167	19	0.75	-0.169
	老人医療費	745	926	1,182	108	0.39	-0.096
	老人福祉費	154	210	255	24	-0.28	0.150
	医師	223	309	423	45	0.54	-0.156
	保健師	21.0	43.0	64.4	10.8	-0.22	-0.034
	社会福祉士	109	157	216	24	0.26	-0.547
	介護福祉士	756	1,275	1,772	255	-0.07	0.306
	訪問介護員	30	74	179	34	1.14	0.012
	老人ホーム	35	65	143	23	1.17	0.283
	老人福祉センター	1.0	6.9	13.2	2.9	0.48	0.074
	介護老人福祉施設	1.0	5.0	16.1	3.9	1.28	-0.015
	介護療養医療施設	12.9	23.1	36.0	4.5	0.24	0.201
	高齢者生活保護	6.2	23.9	59.7	12.5	1.10	0.283
	健診	54.9	62.4	70.0	3.7	-0.09	-0.257
	がん検診	23.3	31.3	43.0	4.0	0.54	-0.030
社会・経済	高齢世帯	30.9	44.8	55.8	5.3	-0.44	0.253
	単身世帯	25.5	31.5	47.3	4.1	1.28	-0.013
	高齢単身	8.3	11.4	16.5	1.9	0.71	0.327
	学歴	8.9	9.8	10.7	0.4	-0.06	-0.121
	社会教育費	6.0	12.2	26.3	4.3	1.51	-0.113
	公民館	6.1	188	724	146	1.58	-0.104
	高齢者講座	20	449	1,225	304	0.64	0.151
	高齢ボランティア	14.1	24.2	33.5	4.1	-0.30	-0.486
	家計収入	420	531	633	57	-0.13	-0.232
	貯蓄	5,747	14,497	19,669	3,082	-0.56	-0.446
	負債	0.16	0.33	0.85	0.11	2.54	0.212
	就業	50	55	59	2.6	-0.26	-0.184
	高齢就業	17.9	22.7	28.7	2.1	0.21	-0.200
	失業	2.9	4.2	6.3	0.7	0.49	0.404
	貧困率	10.2	12.9	16.9	1.4	0.72	0.295
	残業	2.1	11.3	40.0	9.3	2.00	-0.112
	標高	2.4	41.9	371	73	3.01	-0.265
	気温	9.3	16.2	24.1	2.4	-0.22	-0.234
	日照	1,622	1,921	2,188	160	-0.02	-0.174
	降雪	0.0	26.2	114	27.8	1.54	0.347
	人口密度	239	1,357	9,605	1,757	3.41	-0.166
	都市化	24.2	52.5	98.4	18.9	0.90	-0.200

や、対数変換を行った論文がある。また、説明変数間の交絡効果に対処するために2変数の積の項を追加して解析した論文がある。さらに、説明変数間に相関の高い組がある場合、OLSでは多重共線性効果により回帰分析が不安定になるため、高相関の組の一方を削除して解析した論文がある。さらに、本研究のように説明変数が目的変数のデータ数より多い場合、OLSではそのままでの解析は不可能なため、主成分分析等の手法を用いて説明変数を減らして解析した論文がある。しかし、これらの対処はad hoc的なものであり、完全な解決策とはいえない。

本研究では、これらの問題を解決するために非線形重回帰分析の手法としてサポートベクターマシン(SVM)(大北, 2005; 小野田, 2007; 阿部, 2011)を適用した。SVMは、説明変数の数値に対してカーネルと呼ぶ非線形関数を用いて学習パターンを別の空間(超平面)に写像し、その空間で線形回帰を行う。この操作により、説明変数の元の数値での非線形回帰が可能になり、目的変数と説明変数の間の任意の関係に対して高精度の回帰結果が得られる。さらに、SVMでは、①説明変数間の交絡効果が予想される場合でもこの効果が自動的に考慮されるため、変数の積の項の追加は不要、②変数間に強い相関がある場合でも解析可能であり、多重共線性問題は生じない、③目的変数のデータ数以上の説明変数を用いた解析が可能、などの利点がある。以上のSVMの多くの利点はカーネル回帰という手法の採用によるものである(赤穂, 2008)。

SVMのソフトウェアはLIBSVM ver. 3.11(Chang and Lin, 2011)の回帰機能(SVR)を、カーネル関数はradial basis function(RBF)を用いた。多数の説明変数の中から影響要因を探索するためにはSVRモデルと説明変数の最適化を行う必要がある。本研究では、前者については、LIBSVMの ϵ SVRのオプションパラメータのうち、最も重要な g (RBF関数 $\exp(-g*|u-v|^2)$ における係数)、 c (cost)および p (終了のしきい値)(これらのパラメータの詳細についてはChang and Lin, 2011 参照)の最適化を1個抜き交差検証法(LOOCVT)により行った。

後者に関しては、回帰分析では一般に説明変数の中に有効でないものがあると過学習状態に陥り、学習データに対する誤差は減少するが、予測データについての誤差は増大するため、必要最小限の説明変数を抽出する変数選択が必要である。本研究では迅速な変数選択法として感度分析法を採用した。この感度分析法は、目的変数に対する各説明変数の感度を計算し、感度の低い変数を順次削除しながらSVRモデルを最適化し、目的変数の予測値と実測値の平均二乗誤差が最小となる組み合わせを探索する方法である。筆者らはこの感度分析法の有用性を別のタイプの問題において検証した(Tanabe et al., 2013)が、本研究ではデータの特性が異なる問題において有用性をさらに詳細に検証した¹⁾。

この感度分析法の具体的な方法は、感度を推定する説明変数以外の変数はそれぞれ平均に固定し、当該変数のみ実際の数値に設定したときの死亡率のSVR出力値の変化をその変数の感度として算出する。その際、多くの説明変数の場合にSVRの出力値は説明変数の実測値にほぼ線形に変化することが認められるため、実測値と出力値の線形回帰から算出される回帰直線の傾きをその説明変数の感度として推定する。

そこで、以下の手順により影響要因の探索を行った(操作の流れを図2に示す)。

- ① 1個抜き交差検証法(LOOCVT)を用いてSVRモデルを最適化するために、LIBSVMのCVT機能パラメータ v を47(都道府県の数)に設定し、47行57列の学習データを用いてSVRのモデルパラメータ(g, c, p)をそれぞれ0.001~1000の範囲でグリッドサーチし、全都道府県について死亡率の予測値と実測値との平均誤差(RMSE)が最小になる最適条件を探索する。
- ② 各説明変数の感度を求めるために、当該変数は実際の数値のまま、その他の変数は全都道府県の平均値に設定したデータセットを最適モデルに入力して出力値を求め、当該変数の実測値を説明変数、出力値を目的変数とする単回帰分析を行い、回帰直線の傾きをその変

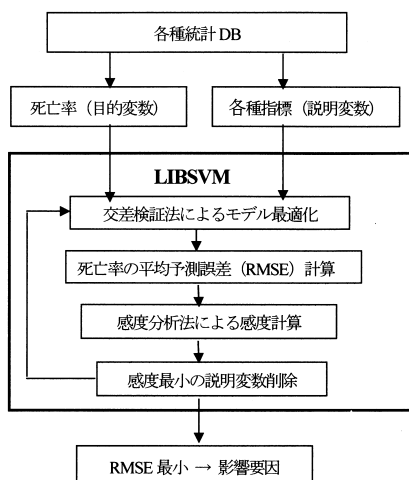


図 2. 死亡率の影響要因探索の流れ図.

数の感度とする.

- ③ 全説明変数の中で感度の絶対値の最も小さい変数を取り除き, ①②の操作を繰り返し, RMSE が最小になる説明変数の組み合わせを死亡率の影響要因とする.

3. 結果

以上の方法により, 56 種の説明変数の中から影響要因を探索した結果, 11 種の変数を用いた場合に死亡率の予測値と実測値との RMSE が最小となり, その前後で RMSE は増加した. この 11 種の変数による最適モデルの自由度調整済回帰決定係数(AR^2)は 0.740 であり, 危険率 1% で有意と判定される.

この SVR の結果と比較するために, 線形回帰分析(OLS)を行った. 先ず, 56 種の説明変数のデータについて, 非線形性に対処するために対数変換を行い, さらに最小 0, 最大 1 に正規化した. 次に, これら 56 種のデータを 28 種の 2 群にランダムに分割し, それぞれの群で OLS を行い, その結果から各群で有意性の低い(p -値の高い)各 8 変数を取り除いた. さらに, 両群を合体させた 40 種の説明変数により OLS を行い, 変数減少法による変数選択を実行した(使用ソフトは(株)エスミの Excel 多変量解析 ver.7.0). その結果, AR^2 は 0.677 となり, SVR とほぼ同程度の回帰結果が得られた. しかし, この際に選択された変数は日照, 健康無関心, 介護福祉士, 趣味, 高齢ボランティア, 歩数, 標高, 保健師, 老人福祉費, 高齢者講座, 貧困率, 公民館, 老人福祉センター, 介護老人福祉施設の 14 種であり, SVR で寄与率の上位になった喫煙や高齢単身などが選択されないのは疑問である. さらに, このうちの日照, 介護福祉士, 歩数, 標高, 高齢者講座, 貧困率, 老人福祉センター, 介護老人福祉施設の 8 変数については, 偏回帰係数の符号が想定とは逆となった. ここで, 得られた有意な変数の偏回帰係数の符号に関して想定との整合性を判定したが, 筆者らは先行研究や疫学調査などの結果から, 各説明変数について死亡率に及ぼす影響の方向を想定している.

以上の結果から OLS の結果は受け入れできないと判定され, 一方, SVR では回帰決定係数が低く, また 11 種の影響要因の感度の符号が筆者らの想定と整合することから, 本研究の SVR と感度分析の有効性, およびその解析による要因の信頼性が実証できたと判断される.

表 4. 影響要因の内訳, および死亡率に対する感度と寄与率.

	影響要因	分野	感度			寄与率 (%)
			危険要因	抑制要因	(誤差)	
1	社会福祉士	医療・福祉	0.211	-0.231	(0.032)	20.2
2	喫煙	生活習慣			(0.031)	16.8
3	高齢单身	社会・経済			(0.022)	11.9
4	がん検診	医療・福祉	0.153	-0.157	(0.021)	9.3
5	健康無関心	生活習慣			(0.019)	8.9
6	塩分	生活習慣			(0.018)	7.4
7	保健師	医療・福祉	0.140	-0.139	(0.018)	7.3
8	肥満	生活習慣			(0.018)	4.7
9	趣味	生活習慣			(0.017)	4.7
10	交際	生活習慣	0.112	-0.109	(0.017)	4.5
11	都市化	社会・経済			(0.016)	4.2

次に, 得られた各要因の死亡率への相対的影響度について考察するために, 影響要因 i の感度 S_i (前頁の操作②で求めたもの) から式

$$(3.1) \quad C_i(\%) = \frac{S_i^2}{\sum_{i=1}^{11} S_i^2} \times 100$$

により死亡率に対する寄与率 C_i を推定した. 影響要因の内訳, 死亡率に対する感度とその推定誤差, および寄与率を表 4 に示す. 得られた影響要因の感度の誤差は感度分析過程における感度値のばらつきから推定したが, その大きさを見ると, 感度はいずれも信頼できる精度で推定されていると考えられる.

また, 感度分析で得られた各要因の感度は, 他の変数は固定し, 当該変数のみ数値を変化させた時の死亡率の変化から求めたので, 死亡率に対する当該要因の正味の影響度を表わしている. したがって, 感度が正の 5 種の要因 (喫煙, 高齢单身, 健康無関心, 塩分, 肥満) は死亡率増加に寄与する危険要因であり, 感度が負の 6 種の要因 (社会福祉士, がん検診, 保健師, 趣味, 交際, 都市化) は死亡率減少に寄与する抑制要因であると解釈できる.

4. 考察

4.1 変数選択における感度の符号について

表 4 に示した 11 種の影響要因のうち, 感度が正の危険要因 5 種と, 感度が負の抑制要因 6 種についての結果は, 各要因の死亡率に対する影響の予測方向と整合するが, この点について補足しておく. 本研究では, 図 2 に示したように, 感度分析による変数選択において感度最小の変数を順次削除していき, 死亡率の平均予測誤差 (RMSE) が最小になったときの説明変数を死亡率の影響要因としたが, 実際の削除すべき説明変数の選択においては, 感度の数値だけでなく, その符号も考慮している. これは通常の線形重回帰分析 (OLS) においても遭遇する問題であるが, 実例でこの問題に対する筆者らの対処法を示そう.

本研究で使用した都道府県別の死亡率データを目的変数とし, 簡単のために, 説明変数は喫煙率と魚介類摂取量の 2 変数のみとする. OLS では

$$(4.1) \quad \text{死亡率} = a \times \text{喫煙率} + b \times \text{魚介類摂取量}$$

として, 係数 a と b を最小二乗法で計算すると, a, b は正の数値になる. また, 本研究で用いた SVR の感度分析の結果でも a, b とも符号は正になる.

周知のように喫煙は肺がんや心疾患, 脳血管疾患などの重大疾患 (表 1) の重要原因であると

されることから、係数 a が正符号になることは妥当である。一方、係数 b の符号については、死亡率国内 1 位の青森県は魚介類の摂取量も 1 位で、死亡率最下位の長野県の摂取量が低く、47 都道府県での死亡率と魚介類摂取量との相関係数が 0.418 と正符号であることから、係数 b も正符号になることは妥当な結果であるようにみえる。しかし、魚介類に多く含まれる DHA や EPA などの栄養成分は、悪玉コレステロールや中性脂肪を下げ、動脈硬化の予防に有効とされていることから、魚介類の摂取は死亡率の低下に働く可能性が高いと考えられている。したがって、OLS と SVR での係数 b の符号が正であるという結果は想定外であり、説明変数を多数に増やしてもこの結論は変わらない。

このような回帰係数の符号が想定と異なる場合、OLS の解説書などでは多重共線性問題として、説明変数間の相関分析を行い、高相関の組の一方を削除する処理を勧めていることが多い。しかし、47 都道府県での喫煙率と魚介類摂取量との相関係数は 0.352 と低く、多重共線性問題ではないと考えられる。そこで本研究では、表 2 に挙げた説明変数について予備的な SVR の感度分析を行い、想定とは異符号の説明変数は不適当(厳密に言えば、本研究のような都道府県別の死亡率の影響要因を探索する場合には)として削除した。このような前処理を施すことにより本研究で得られた影響要因 11 種に関しては、栄養学、疫学、社会学などの観点から、危険要因であるか、抑制要因であるかの点で問題がない結果を得ることができたと考えられる。

4.2 感度と相関係数

影響要因を含む全説明変数について、死亡率との単相関係数と、死亡率に対する感度との散布図を図 3 に示す。この図をみると、全説明変数の中で感度が高いものとして選ばれた影響要因の中には相関係数の絶対値が低いものがある。また逆に、説明変数の中には相関係数が高いにもかかわらず、感度が低く影響要因にならなかったものが多数ある。さらに、感度と相関係数が異符号の変数も多数ある。これらの結果は、感度が死亡率に対する説明変数個々の正味の感度を示すのに対し、死亡率との相関係数には他の説明変数の寄与が含まれているためであるが、先行研究における解析に関して 2 つの問題点を示唆する。

第 1 は死亡率との相関分析に基づく説明変数の影響度の考察である。これまでは都道府県の死亡率との相関係数に基づいて各説明変数の影響度を分析した研究が多い。しかし、本研究の結果はこのような相関分析で得られた先行研究の結果には疑問があることを示唆する。第 2 は説明変数の選定であり、死亡率と相関の高い指標を説明変数に選び重回帰分析を行った先行研究が多い。しかし、このような方法では選定された説明変数以外に死亡率に大きな影響を与え

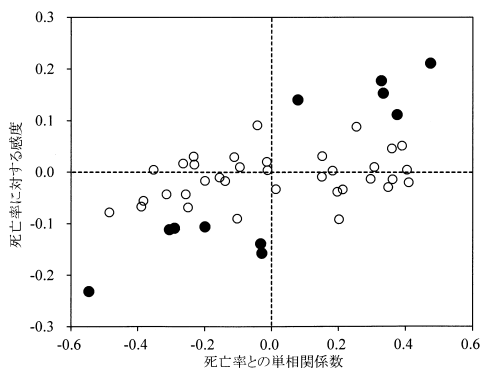


図 3. 全説明変数の死亡率との単相関係数 vs 感度の散布図(●: 影響要因, ○: 非影響要因)。

る変数を見落とす可能性がある。たとえば、本研究で影響要因となったがん検診は死亡率との相関係数が -0.030 、保健師は -0.034 と非常に低いため、それらを説明変数として見落とす可能性が高いと考えられる。

以上を総括すると、先行研究では説明変数の選定や死亡率との影響度に関する考察において疑問がある場合が考えられる。これに対し、本研究では死亡率に対する正味の感度に基づいて広範な分野の多数の説明変数の中から探索した影響要因が死亡率を統計的に有意な精度で再現する結果が得られたことから、本研究の解析法の有効性が実証されたと考えられる。

4.3 影響要因に関する全般的考察

表4の11種の影響要因を分野別に分類すると、生活習慣分野が6要因、医療・福祉分野が3要因、社会・経済分野が2要因であり、影響要因が3分野にまたがっている。このことは多くの先行研究のように、特定の分野の比較的少数の説明変数を用いた解析では、その他の分野の影響が大きい説明変数が欠落している可能性があり、得られた死亡率の影響要因について信頼性に疑問があることを示唆する。たとえば、Fukuda et al. (2004)は失業率、学歴、所得等を中心とする社会経済指標を、大坪 他 (2009)は医師数、病床数等の医療資源指標を、また、林 他 (2009)は喫煙、飲酒、食塩摂取量等の健康・栄養関連指標をそれぞれ用いて解析しているが、これらの結果はいずれも回帰係数が低く、それらの結果には疑問の余地がある。これに対し、本研究では広範な分野の多数の説明変数を用いて非線形重回帰により解析したことで、影響要因に関して先行研究より信頼性の高い結果が得られたと考えられる。

影響要因の寄与率を分野別に合計すると、生活習慣 47.0%、医療・福祉 36.9%、社会・経済 16.1%となる。海外では死亡率や平均寿命等の健康に関連する社会経済的要因はSES (Socio-Economic Status)と呼ばれ、中でも三大要因とされる学歴、所得、職業と健康との関係が多く研究されている (Adler and Newman, 2002)。日本人については所得、失業、学歴等の経済・社会要因、病院や医師等の医療要因の影響が大きいとされている (福田・今井, 2007; Kagamimori et al., 2009; 日本公衆衛生学会, 2011; 橋本, 2012; 近藤, 2014)。しかし、これまでは本研究のように広い分野の多数の説明変数を用いて解析した研究はない。

4.4 生活習慣分野の影響要因

影響要因を分野別に考察すると、生活習慣分野では、上記のように人間の死亡には生活習慣が大きな影響を与えるとされていることから、喫煙、飲酒等の19種の説明変数を挙げた。その内、影響要因となったのは喫煙、健康無関心、塩分、肥満、趣味、交際の6種であり、その寄与率の合計は3分野の中で最大である。この結果に対し、これまで喫煙や飲酒等の生活習慣の影響を調べた研究はあるが、上記のように説明変数の分野が限定的な研究がほとんどのため、生活習慣要因の影響が全体の中でどの程度を占めているかを明らかにしたものはない。

本研究で見出された影響要因の中でも、喫煙が寄与率 16.8%で影響要因2位となったことが注目される。この結果は日本人の死因の上位を占めるがん、心疾患、肺炎、脳血管疾患(表1)のいずれにも喫煙が悪影響を及ぼすとされていることによく対応している。死亡率に対する喫煙の影響を調べた先行研究としては、林 他 (2009)は都道府県別死亡率について喫煙、飲酒、食品摂取量等の健康・栄養関連指標を用いて解析し、歩数や塩分は影響要因になるが、喫煙はならないとした。片山他(2012)は都道府県別乳がん死亡率について喫煙、ボランティア活動、教育施設等の説明変数を用いて解析し、喫煙の影響が大きいことを示した。しかし、本研究のように都道府県別全死因死亡率について多種多様な説明変数を用いて解析し、喫煙の影響が大きいことを示した先行研究はない。

しかし、喫煙に関してはその単独要因としての影響だけでなく、他の要因と複合した関連要

因としての影響の大きさが注目される。すなわち、喫煙、健康無関心(要因 5 位)、塩分(6 位)、肥満(8 位)および医療・福祉分野のがん検診(4 位)を統合すると健康関心要因とみなすことができ、これら 5 要因の寄与率の合計は 47.2% に達する。このうちで、健康無関心者のデータは「国民生活基礎調査」において、日頃健康のために実行している事柄は何か、という質問に対し、規則正しく食事している、適度に運動している、十分に睡眠している、煙草を吸わない、などの項目には回答せず、特に何もしていない、と回答した人の割合として集計した。一方、これまで健康への関心度と健康指標との関連を調べた研究としては、森本・丸山(2001)は地域住民や高齢者等を対象に喫煙、飲酒、運動などの生活習慣と各種疾患との相関を解析し、健康意識の低い人ほど慢性疾患の発症率が高いことを示した。また、熊谷 他(2008)は地域住民を対象として主観的精神健康度と身体健康度の関連を解析した。しかし、都道府県別死亡率に対する健康関心度の影響を解析した先行研究はない。

4.5 医療・福祉分野の影響要因

医療・福祉分野については、直接的に人の死亡に関連することから、医療費、医療施設数・従事者数等の医療資源指標を中心に 15 種の説明変数を取り上げた。その結果、影響要因となったのは社会福祉士、がん検診、保健師の 3 種であり、その寄与率の合計は 36.9% で、死亡率に大きな影響を与えることを示している。先行研究でも医療分野の指標が死亡率に対して影響要因になることを示した研究が多い。大坪 他(2009)は全国の市区町村を対象として各種医療資源等の指標を説明変数に用いて重回帰分析を行い、医師数や救急告示病院数が死亡率に有意な影響を与えることを示した。また、五十嵐 他(2014)は全死因の都道府県別自宅死亡割合と医療社会的指標の相関を分析し、病床数が少なく、在宅医療へのアクセスが良い都道府県で自宅死亡割合が高い傾向にあることを示した。しかし、先行研究では用いられた説明変数の範囲が限定的であるため、医療・福祉分野の影響要因が死亡率にどの程度の寄与を及ぼすかについては明らかにされていない。

本研究の解析で得られた影響要因 4 位のがん検診が寄与率 9.3% で死亡率に影響しているという結果は、表 1 のように現在のわが国ではがんによる死亡が全死因の 28.5% を占めている実態に対応しており、がん検診の有効性を示している。蔵満 他(2014)は北海道の 179 市町村の健康寿命について保健医療環境や社会経済的指標を説明変数に用いて重回帰分析を行い、肺がん検診率が健康寿命の延伸に関連していることを示した。しかし、全都道府県の全死因死亡率に対するがん検診の効果を検証した先行研究はない。

しかし、本研究の結果の中でもっとも注目すべきは、これまで検証されていなかった社会福祉士数が寄与率 20.2% で第 1 位の影響要因になり、各種要因の中で死亡率に対してもっとも大きな影響を与える結果が得られたことである。2006 年に改正された介護保険法により全国に設置された地域包括支援センターには、保健師、社会福祉士、介護福祉士等主任ケアマネジャーの 3 職種が配置されている。社会福祉士は医療・福祉等の施設において病气やケガなどにより日常生活を送るのが難しい人に対して助言や指導を行う職業専門職であり、その活動が地域住民の健康に機能している状況については多くの報告がある。しかし、これまで社会福祉士が地域の健康指標の改善に及ぼす影響を統計的に検証した研究はない。蔵満 他(2014)は北海道の 179 市町村の健康寿命について保健医療環境や社会経済的指標を説明変数に用いて重回帰分析を行い、保健師が健康寿命の延伸に関連していることを示したが、社会福祉士は検証していない。この要因についてはこれまで実証例がないため、今後の検証研究が待たれる。

4.6 社会・経済分野の影響要因

社会・経済分野では 3 分野の中で最多の 21 種の説明変数を用いて解析した結果、高齢単身

と都市化の2種が影響要因で、寄与率16.1%で死亡率に影響していることが見出された。この中で注目すべきは、高齢単身が3位の11.9%という高率で死亡率に影響している結果が得られたことである。高齢単身の割合と死亡率等との関連を解析した先行研究としては、尾形他(2011)は慢性透析患者数と高齢単身者数や病院数等との関連を解析し、吉本他(2011)は救急搬送時の転倒数と高齢単身率との関連を解析している。しかし、本研究のように都道府県別死亡率と高齢単身世帯数との関連を解析した先行研究はない。

この高齢単身についても、その単独要因としての影響だけでなく、他の要因と複合した関連要因としての影響の大きさが注目される。すなわち、生活習慣分野の趣味、交際、および社会・経済分野の都市化との統合であり、これらに高齢単身を加えた4要因は死亡率に対して共通する関連性があり、それは社会との関連性・孤立性である。「高齢社会白書」によれば、近年の日本では職場、地域、家庭における人間関係が希薄化し、社会的に孤立する人が増え、健康への影響が懸念されている(近藤, 2004; 内閣府, 2018)。藤本他(2011)はこのような社会的孤立と健康との関連性について実証的研究を行い、社会・経済分野の三大要因である所得、学歴、職業だけでなく、社会的孤立も健康に影響している可能性があることを示した。また、安梅他(2006)は農村に居住する高齢者について社会とのかかわり状況を示す社会関連性指標を用いて死亡率との関連を回帰分析し、活動参加、趣味、積極性等の社会関連性が乏しいと死亡率が高いという結果を得た。しかし、本研究で全都道府県の死亡率の解析から得られた社会連携関連4要因の寄与率の合計が25.3%に達することから、社会的孤立が健康に大きく影響していることを実証したのは本研究が初めてである。

以上を総括すると、本研究では47都道府県別の死亡率について56種の説明変数を用いて非線形重回帰分析した結果、死亡率を統計的に有意な精度で再現する11種の影響要因が得られ、その内でも、これまで未検証の社会福祉士が死亡率にもっとも大きな影響を与えること、健康関心要因や社会関連要因といった複合要因も大きな影響を与えること等、新たな結果が得られ、広範な分野の多数の説明変数の中から感度分析により影響要因を探索する解析手法の有効性を実証した。

4.7 地域の多死化対策の試論

最後に、本研究の結果に基づいて地域の多死化対策に資する情報提供の可能性を考察してみる。対象地域として、国内で死亡率が長年、最高の青森県と、近年、最低の長野県を取り上げると、両県での影響要因11種の国内順位は表5のようになる。ここで各要因の危険度順位は、感度が正の危険要因については指標値の降順、符号が負の抑制要因については昇順の順位である。この表を見ると、長野県は塩分摂取量以外は全体的に危険度順位が低い要因が多く、死亡率の全国最低に対応している。これに対し、青森県は危険度順位の高い要因が多く、特に寄与率の高い社会福祉士と喫煙の危険度順位の高さが青森県の死亡率の国内最高に大きく関与していることが明らかである。

青森県では現在、健康づくりの推進や自殺対策などの短命県返上への模索を実施しているが、本研究の影響要因の結果を基に青森県の死亡率改善策を考えてみると、まず、社会福祉士の増員と、喫煙率や塩分摂取を中心とする健康意識の向上が考えられる。青森県の社会福祉士数(人口当たり)は全国で3番目に少ない123人であり、これは最多の新潟県の6割弱にとどまる。社会福祉士の充実が青森県のみならず、全国の自治体の死亡率低減に費用対効果の高い対策であると考えられる。また、喫煙率24.9%は北海道の26.6%に次ぐ全国2位であり、最下位の奈良県の16.3%とは大きな開きがある。

青森県は降雪日数が2位の雪国県であり、屋外活動率の低さなどが肥満(5位)につながり、飲酒率1位、喫煙率2位、塩分摂取6位などの健康意識の低さ、無関心につながり、最終的に

表 5. 影響要因の指標値の最小、最大と該当県、および青森県と長野県の指標値と危険度順位. *危険度順位：危険要因は指標値の降順，抑制要因は指標値の昇順.

影響要因	影響	最小		最大		青森県		長野県	
		指標値	該当県	指標値	該当県	指標値	順位*	指標値	順位*
社会福祉士	抑制	109	茨城	216	新潟	123	3	162	28
喫煙	危険	16.3	奈良	26.6	北海道	24.9	2	19.3	31
高齢单身	危険	8.3	滋賀	16.5	高知	12.1	14	10.5	29
がん検診	抑制	23.3	大阪	43.0	山形	31.9	28	35.4	40
健康無関心	危険	13.2	長野	18.1	青森	18.1	1	13.2	47
塩分	危険	8.6	沖縄	10.9	長野	10.5	6	10.9	1
保健師	抑制	21.0	神奈川	64.4	山梨	41.9	18	62.2	45
肥満	危険	22.7	京都	24.4	福島	23.9	5	23.4	17
趣味	抑制	22.2	青森	30.7	奈良	22.2	1	27.0	28
交際	抑制	81	青森	185	高知	81	1	178	45
都市化	抑制	24.2	島根	98.4	東京	46.6	23	34.2	7

死亡率の高さを招いていることが推測される。そこで、青森県の根本的な死亡率低減対策には健康意識の向上が重要であると考えられ、そのための一つの方策として第一次産業からの産業構造の転換を提案する。同県が雪国県である点を考慮すると、IoT、ビッグデータ、AI など ICT(情報通信技術)産業の推進が有効と思われる。山形県なども同様の施策を進めているが、その見本になるのは自然条件が似ている北欧のフィンランドとエストニアであり、両国の新興 IT 企業は世界に進出している。青森県内の大学や高専の情報関係学部と奨学金制度の拡充強化や IT 企業の誘致に成功すれば、教育水準の改善が図られ、これらを通じて健康意識の向上、ひいては死亡率の低下が期待できよう。

5. 結論

本研究では、地域の多死化対策に有用な情報を提供することを目的に、都道府県の全死因死亡率の男女平均値を目的変数とし、生活習慣、医療・福祉、社会・経済の 3 分野の指標 56 種を説明変数に用い、サポートベクター回帰により影響要因を探索した。その結果、社会福祉士、喫煙、高齢单身、がん検診、健康無関心、塩分、保健師、肥満、趣味、交際、都市化の 11 種が見出された。このうち、喫煙を中心とする健康関心関連要因、高齢单身を中心の社会連携関連要因、およびこれまで未検証の社会福祉士要因が死亡率に大きな影響を与えることを明らかにした。これらの結果から、非線形重回帰分析における変数選択としての感度分析法により影響要因を探索する解析手法の有効性を実証した。また、地域の多死化対策への試みとして、死亡率が最高の青森県について死亡率低減策を検討した。

今後の課題として、本研究で用いた手法は生態学的研究であるため、生態学的誤謬(Ecological Fallacy)の問題がある。すなわち、都道府県別の死亡率の解析から得られた要因は個人の死因に関連付けられるものではなく、都道府県の死亡率差を説明するものにすぎない。地域の多死化対策により有用な情報を得るためには、時系列データや個人単位のマイクロデータ等の各種データを利用した総合的な解析を行う必要がある。

また、本研究で見出された 11 種の影響要因はそれらが単独に死亡率に影響するだけでなく、喫煙を中心とする健康関心関連要因、および高齢单身を中心の社会連携関連要因のように、幾つかの要因が統合して死亡率に影響している可能性があることが推測された。このような階層構造を介した健康指標への影響分析に関しては、共分散構造分析手法やマルチラベルモデルを用いたアプローチがあり、これらの手法を用いた死亡率の要因分析は今後の課題である。

以上、本研究では都道府県単位の死亡率について多種多様な説明変数の中から影響要因を探

索した結果、都道府県の死亡率の影響要因を見出すことができ、地域の多死化対策やコホート研究等に役立つ参考情報が得られた。しかし、地域の多死化対策にさらに密接した情報を取得するためには、都道府県内の市区町村単位の死亡率について同様の解析を行うことで、より地域の実態に即した情報が得られることを付記しておく。

注.

- 1) 筆者らがサポートベクター回帰における変数選択のための感度分析法を開発、検証した既報(Tanabe et al., 2013)においては、使用データの特徴の点において本研究とは異なっており、このことが本研究における感度分析法の検証の理由である。すなわち、既報においては化学物質の発がん性予測システムを開発するために、1500種以上の化学物質発がん性データを目的変数としてサポートベクター回帰による影響要因の探索を試みた。しかし、現時点では化学物質の発がん性に関しては原理や機構が未解明であるため、入手可能な1500種以上の説明変数の中から感度分析による変数選択法を用いて影響要因を探索した。その際、採用した説明変数の大半は連続変数ではなく離散変数であり、しかも、その数値0, 1, 2, ... に対して該当化学物質の頻度がたとえば1000, 100, 10, 1, 0, 0, ... のように極端に非対称な分布形状をしている変数が多かった。これに対して、本研究で検証した都道府県別の全死因死亡率に対する説明変数はすべて連続変数であり、多少の非対称分布を示す変数もあるが、大半の変数は対称性の高い分布を示している。筆者らとしてはこのようなデータに対して感度分析法による変数選択の有効性を検証したいと考えたのが本研究における感度分析法の検証の理由である。

参 考 文 献

- 阿部重夫 (2011). 『パターン認識のためのサポートベクトルマシン入門』, 森北出版, 東京.
- Adler, N. E. and Newman, K. (2002). Socioeconomic disparities in health: Pathways and policies, *Health Affairs*, **21**(3), 60-76.
- 赤穂昭太郎 (2008). 『カーネル多変量解析非線形データ解析の新しい展開』, 岩波書店, 東京.
- 安梅勅江, 篠原亮次, 杉澤悠圭, 伊藤澄雄 (2006). 高齢者の社会関連性と生命予後: 社会関連性指標と7年間の死亡率の関係, *日本公衆衛生雑誌*, **53**(9), 681-687.
- Araki, S. and Murata, K. (1986). Social life factors affecting the mortality of total Japanese population, *Social Science and Medicine*, **23**, 1163-1169.
- Chang, C.-C. and Lin, C.-J. (2011). LIBSVM-a library for support vector machines, <http://www.csie.ntu.edu.tw/~cjlin/libsvm/>.
- 賃金構造基本統計調査 (2010). <https://www.e-stat.go.jp/stat-search/files?page=1&toukei=00450091&tstat=000001011429>.
- 衛生行政報告 (2014). <https://www.mhlw.go.jp/toukei/saikin/hw/eisei/16/>.
- 藤本健太郎, 濱野強, 藤澤由和 (2011). 社会的孤立と健康, *エストレーラ*, (205), 9-14.
- 福田吉治, 今井博久 (2007). 日本における「健康格差」研究の現状, *保健医療科学*, **56**(2), 56-62.
- Fukuda, Y., Nakamura, K. and Takano, T. (2004). Wide range of socioeconomic factors associated with mortality among cities in Japan, *Health Promotion International*, **19**, 177-187.
- 橋本英樹 (2012). 健康格差の実証研究—方法論的課題と展望, *医療と社会*, **22**(1), 5-17.
- 林美美, 横山徹爾, 吉池信男 (2009). 都道府県別にみた健康・栄養関連指標の状況と総死亡および疾患別死亡率, *日本公衆衛生雑誌*, **56**, 633-644.
- 五十嵐美幸, 佐藤一樹, 清水恵, 菅野雄介, 菅野喜久子, 川原礼子, 宮下光令 (2014). がん死亡および全死

- 因の都道府県別自宅死亡割合と医療社会的指標の地域相関分析, *Palliative Care Research*, **9**(2), 114-21.
- 人口動態調査特殊報告 (2017). <https://www.e-stat.go.jp/stat-search/files?page=1&toukei=00450013&tstat=000001102115>.
- 人口動態統計 (2016). <https://www.mhlw.go.jp/toukei/saikin/hw/jinkou/kakutei16/index.html>.
- Kagamimori, S., Gaina, A. and Nasermoaddeli, A. (2009). Socioeconomic status and health in the Japanese population, *Social Science and Medicine*, **68**, 2152-2160.
- 家計調査 (2013). https://www.e-stat.go.jp/stat-search/files?page=1&layout=datalist&toukei=00200561&tstat=000000330001&cycle=7&year=20170&month=0&tclass1=000000330001&tclass2=000000330019&tclass3=000000330020&result_back=1.
- 片山佳代子, 助友裕子, 黒沢美智子, 横山和仁, 岡本直幸, 稲葉裕 (2012). 都道府県別乳がん死亡率と教育系ファシリティとの関連: ソーシャル・キャピタルの視点から, 厚生指標, **59**, 26-34.
- 国民健康栄養調査 (2012). https://www.e-stat.go.jp/stat-search/files?page=1&layout=datalist&toukei=00450171&kikan=00450&tstat=000001041744&cycle=7&tclass1=000001111535&survey=%E5%81%A5%E5%BA%B7&result_page=1&second2=1.
- 国民生活基礎調査 (2013). <https://www.e-stat.go.jp/stat-search/files?page=1&layout=datalist&toukei=00450061&tstat=000001031016&cycle=7&tclass1=000001064999&tclass2=000001065383&tclass3=000001065392&second2=1>.
- 近藤克則 (2004). 人間関係と健康, 公衆衛生, **68**(3), 224-228.
- 近藤克則 (2014). 健康の社会的決定要因と医療経済・政策学, 医療経済研究, **26**(2), 81-99.
- 熊谷幸恵, 森岡郁晴, 吉益光一, 富田容枝, 宮井信行, 宮下和久 (2008). 主観的な精神健康度と身体健康度, 社会生活満足度および生きがい度との関連性: 性およびライフステージによる検討, 日本衛生学雑誌, **63**, 636-641.
- 蔵満美奈, 木村宣哉, 藤田直人, 河原田まり子 (2014). 健康寿命の延伸と地域保健活動との関連—北海道市町村を対象にした生態学的研究—, 日本公衆衛生看護学会誌, **2**(1), 20-28.
- 森本兼義, 丸山総一郎 (2001). ライフスタイルと心身の健康, 心身医学, **41**, 241-251.
- 内閣府 (2018). 平成 29 年版高齢社会白書, https://www8.cao.go.jp/kourei/whitepaper/w-2017/zenbun/29pdf_index.html.
- 日本学術会議パブリックヘルス科学分科会 (2011). わが国の健康の社会格差の現状理解とその改善に向けて, <http://www.scj.go.jp/ja/info/kohyo/pdf/kohyo-21-t133-7.pdf>.
- 日本公衆衛生学会公衆衛生モニタリング・レポート委員会 (2011). 高齢者における社会経済的要因と健康, 日本公衆衛生雑誌, **58**(7), 564-568.
- 尾形聡, 西慎一, 若井建治, 井関邦敏, 椿原美治 (2011). 慢性透析患者の地域差の要因, 日本透析医学会雑誌, **44**, 681-688.
- 小野田崇 (2007). 『サポートベクターマシン』, オーム社, 東京.
- 大北剛(訳) (2005). 『サポートベクターマシン入門』, 共立出版, 東京.
- 大坪浩一, 山岡和枝, 横山徹爾, 高橋邦彦, 丹後俊郎 (2009). 標準化死亡比の経験的ベイズ推定量に基づく医療資源と死亡との関連: 全国の市区町村を対象として, 日本公衆衛生雑誌, **56**(2), 101-110.
- Rothman, K. J. (2012). *Epidemiology — An Introduction*, Oxford University Press, Oxford.
- 社会福祉振興・試験センター (2015). 社会福祉士・介護福祉士・精神保健福祉士の都道府県別登録者数, http://www.sssc.or.jp/touroku/pdf/pdf_t04_t04_3.pdf.
- 社会・人口統計体系 (2014). <https://www.e-stat.go.jp/regional-statistics/ssdsviiew>.
- 社会生活基本調査 (2014). <https://www.e-stat.go.jp/stat-search/files?page=1&toukei=00200533&tstat=000001095335>.
- 社会生活統計指標—都道府県の指標 (2013). <https://www.e-stat.go.jp/stat-search/files?page=1&layout=datalist&toukei=00200502&tstat=000001095336&cycle=0&tclass1=000001095337>.
- 田辺和俊, 鈴木孝弘 (2018). 都道府県の相対的貧困率の計測と要因分析, 日本労働研究雑誌, **60**(2-3),

45-58.

- Tanabe, K., Lučić, B., Amić, D., Kurita, T., Kaihara, M., Onodera, N. and Suzuki, T. (2010). Prediction of carcinogenicity for diverse chemicals based on substructure grouping and SVM modeling, *Molecular Diversity*, **14**(4), 789-802.
- Tanabe, K., Kurita, T., Nishida, K., Lučić, B., Amić, D. and Suzuki, T. (2013). Improvement of carcinogenicity prediction performances based on sensitivity analysis in variable selection of SVM models, *SAR and QSAR in Environmental Research*, **24**(7), 565-580.
- 田辺和俊, 鈴木孝弘, 中川晋一 (2016). サポートベクター回帰による都道府県別肺がん死亡率の関連要因に関する検討, 保健医療科学, **65**(6), 598-610.
- 特定健康診査 (2010). <https://www.mhlw.go.jp/bunya/shakaihoshou/iryouseido01/info02a-2.html>.
- World Health Organization (2016). The determinants of health, <http://www.who.int/hia/evidence/doh/en/>.
- 吉本好延, 三木章江, 浜岡克伺, 大山幸綱, 河野淑子, 佐藤厚 (2011). 救急搬送における高齢者の転倒の標準化発生比と社会経済状態の関連, 日本公衆衛生雑誌, **58**, 183-189.

Verification of the Effectiveness of Sensitivity Analysis as a Variable Selection in Support Vector Regression

—Analysis of Factors Affecting Prefectural All-cause Mortality Rates—

Kazutoshi Tanabe¹ and Takahiro Suzuki²

¹Institute of Social Sciences, Toyo University

²Department of Economics, Toyo University

The sensitivity analysis method as a variable selection in support vector regression has been applied to the search of factors affecting prefectural all-cause mortality rates, and its effectiveness has been verified. In Japan, a multi-death society will come after an aging society with a declining birthrate, and various social problems such as shortage of doctors, end-of-life care, lonely death, death place refugees, vacant houses are feared. In order for local governments to take countermeasures against mortality, it is important to clarify the factors that have a significant effect on mortality among many factors and estimate their relative impact. All-cause age-adjusted mortality rates of 47 prefectures were used as an objective variable, and 56 indices in lifestyle habit, medical care/welfare, and society/economy fields were employed as potential explanatory variables. Factors related to the mortality rates were searched by applying a support vector regression technique to these data, and their sensitivities to the mortality rates were estimated by employing a sensitivity analysis method. Eleven kinds of factors which reproduce the observed mortalities of 47 prefectures with an accuracy of statistical significance level were obtained. It is found that rates of social workers not examined in previous studies as well as those of smoking habit and of elderly singles highly contribute to the observed mortality rates. Countermeasures for decreasing the mortality rate of Aomori prefecture showing the highest rate in Japan were proposed on the basis of the affecting factors obtained in this study. From these results, the effectiveness of the sensitivity analysis method as a variable selection in support vector regression has been demonstrated.