

特集Ⅱ：長寿革命に係る人口学的観点からの総合的研究（その2）

平均余命の差異の各種要因の探索

—順位区分尺度によるデータ変換と縮約に基づく方法の利用—

井川孝之*

本稿では、都道府県別の男女別・年層別平均余命と各種政府統計・調査データを用い、カーネル密度推定による累積確率密度に応じた順位区分尺度によるデータ変換を施した上で、主成分分析・カーネル主成分分析、並びに、相関分析・重回帰分析を行うことにより、平均余命の差異の各種要因を探索する。主成分・カーネル主成分の第1主成分は、各カテゴリーとも概ね同様となり、データ変換が外れ値への対処となることに加え、線形の成分が抽出されることが確認できる。第2主成分以降については、主成分とカーネル主成分に差違がみられる場合がある。重回帰分析では、第2主成分以降を含むカーネル主成分を説明変数とした場合の適合度が良好となる場合が確認された。これらの主成分・カーネル主成分、及び、重回帰分析を通じ、平均余命の差異の要因として、結婚・離婚や世帯の状況、医療の充実等の関連がわかった。

キーワード：平均余命，社会経済要因，データ変換，カーネル主成分分析，地域相関分析

I. はじめに

1. 背景・目的

少子化・長寿高齢化が進行し、国立社会保障・人口問題研究所が作成している日本の将来推計人口（平成29年推計）によれば、団塊の世代（1947～1949年生れ）が75歳以降の後期高齢者となる2024年には、65歳以上人口が占める割合を表す高齢化率が29.8%となる見通しである。高齢化率は、その後も上昇し、2040年には35.3%となり、65歳以上の高齢人口は3,900万人程度で推移するが、高齢化率はさらに上昇し、38%を上回る水準になると見込まれている。一方で、足元の出生数は2019年865,239人、2020年840,835人と100万人を下回り、生産年齢人口は2021年1月1日時点で7,437万人（総務省統計局人口推計）となっているが、2040年には5,978万人まで減少することが見込まれている。このような状況の下、社会経済が人口動態に及ぼす影響や人口構造について分析し、社会保障制度の設計・運営やその他の各種政策立案、地域施策・経営等に応用して行くことが望まれる。

社会保障制度における給付やリスク負担については、全年層が社会経済に関しどのよう

* 明治大学大学院先端数理科学研究科兼任講師

な影響を受けているかを明らかにし、これを考慮すべきである。人口に関わる指標のうち平均余命は、全年層に関わる指標の1つであり、長寿化・高齢化に関わる指標でもあることから、本稿では、平均余命を分析の対象として取り上げる。平均余命には地域差がみられることから、都道府県別の社会経済・自然環境に関する各種データを基に、データ変換と主成分・カーネル主成分を利用し、平均余命の差異を生じさせる社会経済や自然環境等の各種要因を探索する。なお、2020年に始まった新型コロナウイルス感染拡大による死亡率への影響が考えられるため、十分データが蓄積されていない状況にも鑑み、2020年1月より前のデータに基づき分析する。

2. 先行研究

日本の死亡率の差異の社会経済要因は、都道府県別の生命表や年齢調整死亡率を用いた地域間差異の分析や小規模な疫学的調査等を通じて行われてきた。堀内（2010）は、日本人の死亡リスクの社会経済的差異の研究は、大きな標本サイズと長期にわたる全国的に抽出された個人レベルのパネル調査データが望ましいとしつつ、米国や西欧諸国と比較し、このようなデータがかなり不足している状況を述べている。実際、日本の死亡率の地域差と個人差の双方を対象とした研究は多くない。

都道府県別の平均余命や健康寿命といくつかの指標については、相関分析や重回帰分析、主成分分析により要因を探索する実証研究が行われている（坂井 1986, 鈴木 2003, 北島・太田 2004, 高・梯 2006, 豊田 2011）。これらの研究では、社会経済要因のみならず、栄養や生活習慣、気候等が取り上げられているが、いずれも、焦点をあてた要因について分析したものとなっている。また、年齢別・死因別の寄与により、都道府県の地域差を分析した研究がある（仲都留・大西 2008, 竹内・關 2013）。山本他（2009）は、都道府県別の新国民生活指標により、男女別の平均寿命の地域格差を生じさせる要因を、単相関分析と重回帰分析を用いて探索している。石井（2015）は、日本版死亡データベース（Japan Mortality Database, 以下, JMD）の人口分析への応用を示した中で、都道府県別の死因別死亡確率の推移を調べ、階層的クラスター分析を実施している。井川（2017）は、構造方程式モデリングやlasso回帰の手法を用いて、都道府県別の平均余命の差異の要因を探索しモデリングしている。

田辺・鈴木（2015）は、先行研究が取り扱っている社会経済等の要因の説明変数が少数で限定的であり、決定係数が低く、統計的に有意な結果が得られていないこと、及び、平均寿命と各要因の非線形の関係性にも留意し、サポートベクターマシーン（SVM）を用いた非線形回帰の手法を用いて、都道府県別の平均寿命及び健康寿命の差異の要因について分析している。田辺・鈴木（2020）では、SVMによる回帰の感度分析による変数選択の有効性の検証手法を提案し、都道府県別全死因死亡率の影響要因について分析している。都道府県の数より多い説明変数を基に分析しており、モデルパラメータの推定は交差検証法（クロスバリデーション）を用いたグリッドサーチにより実施している。SVMは、機械学習の手法の1つであり、非線形関係も取り扱える長所があるが、過剰適合の課題もあ

ることが指摘されている。

海外の平均余命の差異に関する社会経済要因については、1990年代に入り、Wilkinson (1992) が各国における国内所得分布と平均寿命との相関関係を示している。Kennedy et al. (1996) は、米国の州別所得分布と年齢調整死亡率との相関関係を示した。これらは、所得再分配等の政策立案やその後の研究に影響を及ぼし、我が国についても様々な長寿・健康の要因分析が行われている。

平均余命に関連し、健康状態の社会経済要因について研究したものがある。近藤 (2005) は、社会経済状態から健康状態 (健康・疾患・死亡) に至るまでに介在する因子を掲げ、それぞれの因果関係 (プロセス) が実証されつつあるとしている。年齢・性別・遺伝子等の個人固有の因子の他、生活習慣 (ライフスタイル) については、喫煙・飲酒・運動・食生活・体重等に関わるもの、人間関係については、婚姻状態・社会参加・社会ネットワーク・社会サポート等を個人の社会経済的因子として挙げている。これに地球環境や職場・コミュニティ等を環境としての社会の因子として加え、3つの階層が健康に影響を及ぼす因子としている。この他、健康の地域差に関し、社会経済、自然環境、文化等の要因を取り上げている研究として、山口・梯 (2001)、Shibuya et al. (2002)、Fukuda et al. (2005)、川上他 (2006)、京都大学 (2007)、福田・今井 (2007)、Kagamimori et al. (2009)、七田 (2010)、杉澤 (2012)、橋本 (2012)、長野県健康長寿プロジェクト・研究事業 研究チーム (2015) 等がある。

上述の先行研究のうち都道府県別の地域差を調べたものについては、地域差と個人差の双方を取り扱ったパネルデータ等を利用した研究を除き、機械学習の手法を用いた田辺・鈴木 (2015)、田辺・鈴木 (2020) 以外は、関心のある比較的少数の項目 (少なくとも都道府県の数よりは少ない) を対象としており、他の可能性のある項目までは考慮されていない。また、都道府県別の各種データには、外れ値の可能性のあるものが含まれるが、これについての議論は特になされていないか、あるいは、十分に議論されていない。機械学習の手法については、前述の通り、非線形関係も取り扱える長所があるが、過剰適合の課題もあることが指摘されている。

本稿では、このような先行研究の課題を踏まえ、都道府県の数より多いより広範な都道府県別の各種データを用い、外れ値や過剰適合に対処しながら線形関係と非線形関係の特徴をデータ変換と主成分・カーネル主成分、並びに、これらの主成分による重回帰分析により把握し、その結果を通じて平均余命の差異の各種要因を探索する。

II. 使用するデータ

地域別の平均余命の差異の要因を探索するため、以下の通り、平均余命のデータと各種要因に関わる政府統計等のデータを使用する。

1. 平均余命のデータ

都道府県別の平均余命の差異の社会経済要因を探索するため、国立社会保障・人口問題研究所が策定している前述の JMD を用いる。JMD は、国際的な死亡データベース Human Mortality Database (HMD) の方法論を基礎としているが、都道府県別の死亡データも提供しており、地域別死亡状況の分析も可能となっている特色がある。

厚生労働省は、5年毎に実施される国勢調査に基づき、都道府県別生命表及び市区町村別生命表を作成している。2022年8月28日時点で公表されている最新の都道府県別生命表は、2015年の国勢調査に基づき作成されたものであるが、上述の JMD より時点がやや旧くなっている。また、都道府県別生命表は、1965年より継続して作成されているが、5年毎となっている。このようなことを踏まえ、JMD の都道府県別データを使用している。

分析に用いた JMD の平均余命は、都道府県別の生命表のうち、5歳×5年（センサス中心）の区分名2018-2019年の男女別の5歳階級別の年齢の平均余命である。区分名2018-2019年は当該データが提供された時点の名称で、2015年国勢調査による人口と2015年以降の死亡数・出生数に基づき、JMD の方法により算出されたものである。

2. 各種要因に関わる統計データ

平均余命の差異の各種要因を探索するため、各都道府県について継続してデータを提供している政府統計・調査を使用する。具体的には、先行研究で用いられている政府統計・データも踏まえながら、より多くのデータを取り扱えるよう、表1に掲げる統計・調査を利用する。なお、データの基準日は、平均余命より前の直近の基準日時点としている。

表1 分析に使用する政府統計・調査名

番号	統計・調査名	年	概要
1	社会経済統計指標	2020 2021 公表	地域別統計データベース*に含まれる以下の各政府統計・調査 <ul style="list-style-type: none"> ・国勢調査（総務省統計局） ・人口推計（総務省統計局） ・人口動態調査（厚生労働省） ・県民経済計算（内閣府） ・市町村税課税状況等の調（総務省） ・小売物価統計調査（総務省統計局） ・就業構造基本調査（総務省統計局） ・賃金構造基本統計調査（厚生労働省） ・家計調査（総務省統計局） ・貸出・預金動向（日本銀行） ・全国消費実態調査（総務省統計局） ・学校基本調査（文部科学省） ・地方教育費調査（文部科学省） ・社会福祉施設等調査（厚生労働省） ・内閣府子ども・子育て本部（内閣府）

番号	統計・調査名	年	概要
			<ul style="list-style-type: none"> ・全国都道府県市区町村別面積調（国土交通省国土地理院） ・住宅・土地統計調査（総務省統計局） ・下水道施設等実態調査（公益社団法人日本水道協会） ・火災年報（総務省消防庁） ・消防白書（総務省消防庁） ・交通統計（警察庁） ・犯罪統計（警察庁） ・公害苦情調査（公害等調整委員会） ・病院報告（厚生労働省） ・医師・歯科医師・薬剤師調査（厚生労働省） ・衛生行政報告例（厚生労働省） ・被保護者調査（厚生労働省） ・国民医療費（厚生労働省） ・後期高齢者医療事業年報（厚生労働省） ・過去の気象データ（気象庁）
2	国民生活基礎調査	2016	通院率，有訴率，健診受診，心の悩み，飲酒，喫煙，睡眠
3	国民健康栄養調査	2016	野菜摂取量，食塩摂取量，歩数
4	特定健診調査	2015	メタボリックシンドローム該当者割合，メタボリックシンドローム予備群者割合
5	社会生活基本調査	2016	ボランティア参加率，趣味娯楽行動割合
6	社会福祉振興試験センター	2015	社会福祉士，介護福祉士
7	全国家計構造調査	2014	ジニ係数（地域別年間収入・総世帯）

* <https://www.e-stat.go.jp/regional-statistics/ssdsview/>

Ⅲ. 方法

前述の通り，既存の統計手法を用いた先行研究では，関心のある比較的少数のデータを用いた重回帰分析が行われてきたが，必ずしも決定係数は高くない。また，外れ値については特に議論されていないか，あるいは，十分に議論されていない。田辺・鈴木（2015），田辺・鈴木（2020）は，都道府県の数を上回る変数を用い，機械学習の手法であるSVMによる非線形回帰により要因を探索しているが，都道府県数は多くなく，抽出された変数は平均余命に過剰適合してしまう可能性がある。

本稿では，都道府県数を上回る多数種類のデータを前提とし，外れ値や過剰適合に対処するため，以下で述べる通り，データ変換を施した上で主成分を算出して相関分析と重回帰分析を実施し，平均余命の差異の要因を探索する方法を利用する。

1. データの性格とアプローチ

要因探索のため用いている政府統計・調査は，毎年又は数年に1度の頻度で実施されて

おり、データの項目数等も、所謂ビッグデータと呼ばれるものよりはるかに少ない。都道府県の数多くないことから、外れ値として取り扱うことが適当なデータが含まれる可能性があり、また、上述の非線形回帰における過剰適合が生ずる可能性もある。

このような性格を持つデータから、都道府県間の平均余命の差異の要因を探索するため、本稿では、平均余命の差異に関わる社会経済等の特徴を表すシグナルを抽出するアプローチを考える。具体的には、使用するデータを各カテゴリー分野に分類し、外れ値や過剰適合に対処するため、後述の順位区分尺度によるデータ変換を施し、その上で、主成分及びカーネル主成分を算出し、これらの主成分と平均余命との相関分析と重回帰分析を実施することにより、都道府県間の平均余命の差異の要因を探索する方法である。当該方法の有効性については、後述の重回帰分析の適合度等の結果により確認する。

2. データ選定とカテゴリー分類

利用可能な政府統計・調査のうち、先行研究で取り上げられているものを始め、関係性の可能性のあるデータを表1に掲載されるものから選択する¹⁾。選択したデータは、表2に掲載した本稿で用いるカテゴリー分野に分類する(表3参照)。当該カテゴリー分野は、多くのデータを参照している社会経済統計指標の分類や先行研究等の分類を踏まえ、関連性の可能性が考えられる8分野を設定している。カテゴリー分野の数は、重回帰分析を実施することにも鑑み、都道府県の数踏まえながら8としている。

表2 カテゴリー分野

番号	統計・指標又は文献	分野数	カテゴリー分野
1	社会経済統計指標	13	人口世帯, 自然環境, 経済基盤, 行政基盤, 教育, 労働, 文化スポーツ, 居住, 健康医療, 福祉社会保障, 安全, 家計, 生活時間
2	田辺・鈴木(2020)	3	生活習慣, 医療・福祉, 社会・経済
3	Fukuda et al.(2004)	8	経済, 教育, 生活環境, 植生・都市空間, 交通, 予防活動, 医療, 人口統計
4	新国民生活指標 (PLI, 1992-99)	8	住む, 費やす, 働く, 育てる, 癒す, 遊ぶ, 学ぶ, 交わる
5	本稿	8	経済, 雇用・労働, 家計, 教育・世帯, 都市・人口, 医療・福祉, 健康・生活, 自然環境

1) 健康・生活に関わるデータとして、健康寿命に関わる各種期間があるが、健康の捉え方や算出方法に複数の定義があり、また、平均余命の基礎となっている生命表自体と関連性が深いため、本稿では、対象外と整理している。実際には、健康状態と平均余命は、何らかの関連性があるものと考えられる。

番号	カテゴリー分野	年	データ項目
7	医療・福祉	2016	有訴者率（人口千人当たり）*
		2016	通院率（人口千人当たり）*
		2015	一般病院の1日平均外来患者数（人口10万人当たり）
		2016	医療施設に従事する医師数（人口10万人当たり）
		2016	医療施設に従事する看護師及び准看護師数（人口10万人当たり）
		2016	保健師数（人口10万人当たり）
		2015	一般病院平均在院日数
		2015	生活保護被保護実世帯数（月平均一般世帯千世帯当たり）
		2015	1人当たり国民医療費
		2015	後期高齢者医療費（被保険者1人当たり）
		2015	社会福祉士登録者数（人口千人当たり）
		2015	介護福祉士登録者数（人口千人当たり）
		8	健康・生活
2016	野菜類摂取量の平均値（女20歳以上）*		
2016	食塩摂取量の平均値（男20歳以上）*		
2016	食塩摂取量の平均値（女20歳以上）*		
2016	歩数の平均値（男20-64歳）*		
2016	歩数の平均値（女20-64歳）*		
2016	飲酒（男女20歳以上）（割合）*		
2016	喫煙（男女20歳以上）（割合）*		
2016	健診受診（男女20歳以上）（割合）*		
2016	睡眠時間（男女12歳以上）*		
2016	心の悩み（男女12歳以上）（割合）*		
2016	ボランティア活動の行動者率		
2016	趣味・娯楽の行動者率		
2015	メタボリックシンドローム該当者割合		
2015	メタボリックシンドローム予備群者割合		
9	自然環境	2015	年平均気温
		2015	最高気温（日最高気温の月平均の最高値）
		2015	最低気温（日最低気温の月平均の最低値）
		2016	日照時間（年間）
		2015	年間降水量（年間）

* 2016年の熊本県の観察値はないため、2013年の都道府県平均からの偏差に基づき補整

3. データ変換

各項目について、都道府県別のデータ x_i の最大値と最小値をそれぞれ x_{max} , x_{min} とし、 x_i の度数分布を基にカーネル密度推定し得られた累積密度関数 $F(x_i)$ を前提として、区分数 K に応じ、以下の通りデータ変換する。カーネル密度推定は、正規分布²⁾を用い、バンド幅は漸近的な確率密度関数の平均2乗誤差の合計を最小化するように設定する。なお、以下の区分数 K は、最適なものを探索し設定することが可能であるが、ここでは、都道

2) 他の確率分布に基づく累積確率密度関数とすることも可能である。

府県の数も踏まえ $K=6$ の場合について示す.

$1 \leq n \leq K$ について, 以下を満たすとき, $x_i \rightarrow n$ ³⁾

$$F(x_{min}) + [F(x_{max}) - F(x_{min})] \times (n-1)/K \leq F(x_i) < F(x_{min}) + [F(x_{max}) - F(x_{min})] \times n/K \quad (\text{III}-1)$$

4. 主成分分析・カーネル主成分分析

各カテゴリー分類のデータを縮約するため, 主成分とカーネル主成分を算出する. ここで, 主成分は, 相関係数行列に基づく主成分分析で算出される主成分得点である. カーネル主成分は, 都道府県 i, j のデータベクトルを X_i, X_j として, 広く用いられている以下の Gauss カーネルを前提とし算出されるカーネル主成分分析における主成分得点とする. 式 (III-2) における $\|X_i - X_j\|$ は, データベクトル間の距離を表している. カーネル主成分の数式を用いた定義やパラメータ β の設定については, 末尾の付録を参照されたい.

$$k(X_i, X_j) = \exp(-\beta \|X_i - X_j\|^2) \quad (\text{III}-2)$$

5. 相関分析と重回帰分析

都道府県別の男女別の各年層の平均余命と各カテゴリーの主成分・カーネル主成分との相関分析を実施する. また, 男女別の各年層の平均余命を目的変数とし, 各カテゴリー分野の第1から第3主成分の主成分とカーネル主成分を説明変数の候補として赤池情報量規準 (AIC) に基づきステップワイズにより説明変数を選択する重回帰分析を実施する.

IV. 結果

前章で述べた方法により, データ変換の上, 各カテゴリーの主成分・カーネル主成分を算出し, その結果に基づき, 相関分析及び重回帰分析を実施した. 結果は以下の通りである.

1. 各カテゴリー分野の主成分とカーネル主成分

(1) カテゴリー分野: 経済

表3に掲げる経済分野の各データについて標準化し, 横軸をデータ項目, 縦軸を都道府県とし, 最大値・最小値の範囲で描画すると図1の左図のようになる. これに対し, 式 (III-1) に基づき変換したデータについて, 同様に描画すると, 図1の右図の通りとなる. 変換前は, 最大値近辺のデータは色濃く表示されているが, その他の水準のデータについ

3) x_{max} については, $x_{max} \rightarrow K$ とする.

ては、中間色のグレーの多少濃淡があるような状態になっている。一方、変換後では、濃淡が変換前より明確に表示された状態となっている⁴⁾。

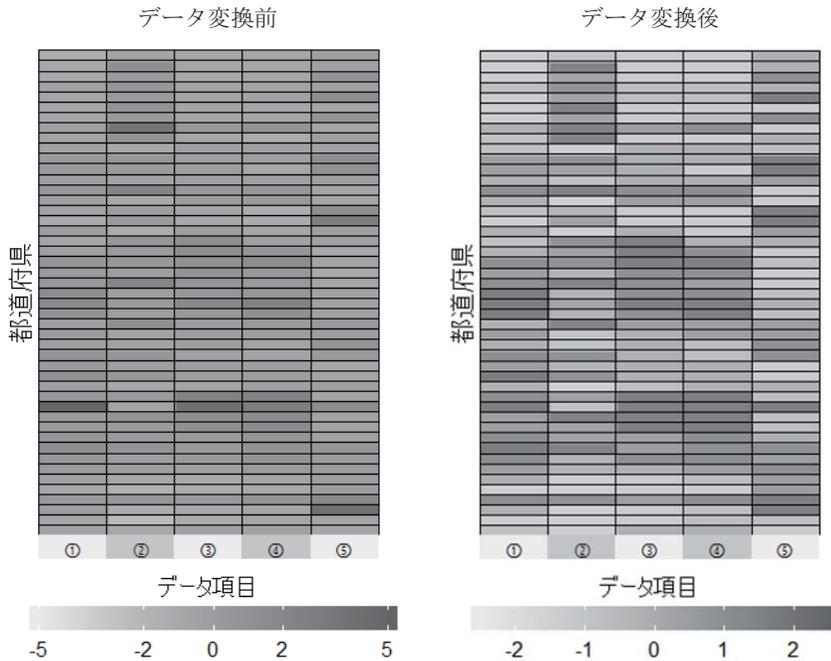


図1 経済分野（データ変換前・後）

（注1）横軸のデータ項目①～⑤は、表3に掲げる経済分野の以下の各項目を表す。データ変換前後とも、標準化した数値をヒートマップ形式にて表示している。

- ① 1人当たり県民所得
- ② 消費者物価指数対前年変化率（総合）
- ③ 課税対象所得（納税義務者1人当たり）
- ④ 財政力指数（都道府県財政）
- ⑤ 実質収支比率（都道府県財政）

（注2）縦軸の都道府県は、下から都道府県コード（全国地方公共団体コードの上2桁の数値）の順に掲載している。すなわち下から、北海道、青森県、岩手県、宮城県、秋田県、山形県、福島県、茨城県、栃木県、群馬県、埼玉県、千葉県、東京都、神奈川県、新潟県、富山県、石川県、福井県、山梨県、長野県、岐阜県、静岡県、愛知県、三重県、滋賀県、京都府、大阪府、兵庫県、奈良県、和歌山県、鳥取県、島根県、岡山県、広島県、山口県、徳島県、香川県、愛媛県、高知県、福岡県、佐賀県、長崎県、熊本県、大分県、宮崎県、鹿児島県、沖縄県と並んでいる。

次に、変換後のデータについて、主成分・カーネル主成分⁵⁾を算出した。主成分の累積寄与率は、第3主成分までで90%以上となっている。主成分負荷量をみると、主成分・カーネル主成分の双方について、第1主成分は1人当たり県民所得、課税対象所得、財政力指数、第2主成分は消費者物価指数、第3主成分は実質収支比率の相関が高く⁶⁾なっている。

4) 後述の各カテゴリー分野のデータ変換についても、同様の状態が確認される。

5) Gaussカーネルのパラメータ β は、オーダー別(1, 0.1, 0.01)について調べ、0.1の場合を示している。いくつかのカテゴリー分野では、0.01の場合についても補足する。

6) ここでは、相関係数が概ね0.6以上の場合を相関が高いという。

図2の経済分野において、主成分・カーネル主成分の第1から第3主成分までを比較⁷⁾している。第1主成分はほぼ同形、第3主成分は類似の形となっており⁸⁾、これらのカーネル主成分に線形と同様の成分が含まれることがわかる。

(2) カテゴリー分野：雇用・労働

表3に掲げる雇用・労働分野のデータについて、変換後のデータで主成分・カーネル主成分を算出すると、主成分の累積寄与率は、第3主成分までで85%以上となる。主成分負荷量をみると、主成分・カーネル主成分について、第1主成分は完全失業率、高齢者就業割合、労働力人口比率、第2主成分は超過労働時間の相関が高くなっている。第3主成分は完全失業率と弱い相関がある。

図2の雇用・労働分野において、主成分とカーネル主成分を比較している。第1主成分と第2主成分はほぼ同形、第3主成分は異なる形となっている。

(3) カテゴリー分野：家計

表3に掲げる家計分野のデータについて、変換後のデータで主成分・カーネル主成分を算出すると、主成分の累積寄与率は、第3主成分までで70%以上となる。主成分負荷量をみると、主成分・カーネル主成分について、第1主成分は国内銀行預金残高、1世帯当たり実収入、世帯主収入、貯蓄現在高、第2主成分は教育費割合、負債現在高の相関が高くなっている。第3主成分については、国内銀行預金残高の相関が高い。

図2の家計分野において、主成分とカーネル主成分を比較している。第1主成分と第2主成分はほぼ同形、第3主成分も若干異なる部分があるものの概ね同じ形となっている。

(4) カテゴリー分野：教育・世帯

表3に掲げる教育・世帯分野のデータについて、変換後のデータで主成分・カーネル主成分を算出すると、主成分の累積寄与率は、第3主成分までで80%以上となる。主成分負荷量をみると、主成分・カーネル主成分について、第1主成分は婚姻率、離婚率、一般世帯の平均人員数、共働き世帯割合、単独世帯割合、第2主成分は最終学歴が大学・大学院卒の者の割合、高等学校卒業者の進学率、高等学校教育費、婚姻率、高齢単身世帯の割合、第3主成分は高齢夫婦のみの世帯割合の相関が高くなっている。

図2の教育・世帯分野において、主成分とカーネル主成分を比較している。第1主成分と第2主成分はほぼ同形、第3主成分は異なる部分もあるが類似の形となっている。

(5) カテゴリー分野：都市・人口

表3に掲げる都市・人口分野のデータについて、変換後のデータで主成分・カーネル主

7) これ以降示す各カテゴリー分野の主成分・カーネル主成分の比較においては、形状の比較のため、正負の符号やスケールを適宜調整している。

8) 変換前のデータで主成分を算出すると、カーネル主成分は主成分と異なる形になる。

成分を算出すると、主成分の累積寄与率は、第3主成分までで65%以上となる。主成分負荷量をみると、主成分・カーネル主成分について、第1主成分はDID人口比率、人口密度、65歳以上人口割合、生産年齢人口比率、保育所等数、認定こども園数、最低居住面積水準以上世帯割合、下水道普及率、刑法犯認知件数、第2主成分は年少人口比率、交通事故発生件数、第3主成分は火災出火件数の相関が高くなっている。カーネル主成分の第3主成分は、年少人口比率の相関も高い。

図2の都市・人口分野において、主成分とカーネル主成分の比較をしている。第1主成分はほぼ同形、第2主成分と第3主成分は異なる部分もあるが類似の形となっている⁹⁾。

(6) カテゴリー分野：医療・福祉

表3に掲げる医療・福祉分野のデータについて、変換後のデータで主成分・カーネル主成分を算出すると、主成分の累積寄与率は、第3主成分までで70%以上となる。主成分負荷量をみると、主成分・カーネル主成分について、第1主成分は一般病院平均外来患者数、医療施設従事医師数、医療施設従事看護師・准看護師数、一般病院平均在院日数、1人当たり国民医療費、後期高齢者医療費、介護福祉士登録者数、第2主成分は有訴率、保健師数、生活保護世帯数の相関が高くなっている。第3主成分は、主成分について通院率の相関が高い。

図2の医療・福祉分野において、主成分とカーネル主成分を比較している。第1主成分と第2主成分はほぼ同形、第3主成分は異なる形となっている。

(7) カテゴリー分野：健康・生活

表3に掲げる健康・生活分野のデータについて、変換後のデータで主成分・カーネル主成分を算出すると、主成分の累積寄与率は、第3主成分までで60%以上となる。主成分負荷量をみると、主成分・カーネル主成分について、第1主成分は野菜摂取量、食塩摂取量、歩数、健診受診、睡眠時間、第2主成分は趣味娯楽行動割合、第3主成分はメタボリックシンドローム予備群者割合の相関が高くなっている。主成分の第3主成分は、ボランティア参加率の相関も高い。

図2の健康・生活分野において、主成分とカーネル主成分を比較している。第1主成分は一部異なる部分もあるが概ね同形、第2主成分と第3主成分は異なる部分があるが類似の形となっている¹⁰⁾。

(8) カテゴリー分野：自然環境

表3に掲げる自然環境分野のデータについて、変換後のデータで主成分・カーネル主成分を算出すると、主成分の累積寄与率は、第3主成分までで85%以上となる。主成分負荷量をみると、主成分・カーネル主成分について、第1主成分は年平均気温、最高気温、最

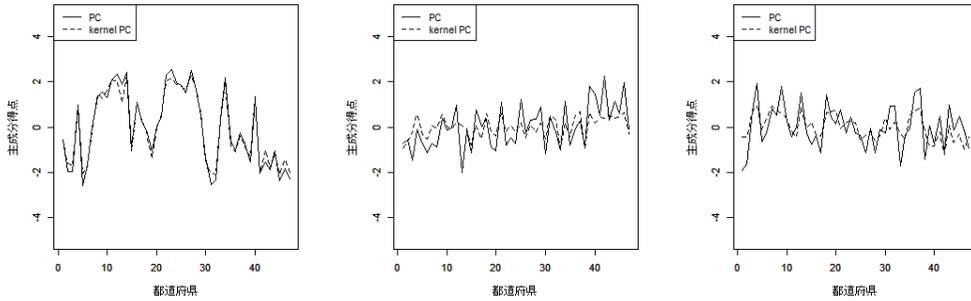
9) Gauss カーネルのパラメータ β が0.01の場合は、第1から第3主成分まで全てほぼ同形となる。

10) Gauss カーネルのパラメータ β が0.01の場合は、第1から第3主成分まで全てほぼ同形となる。

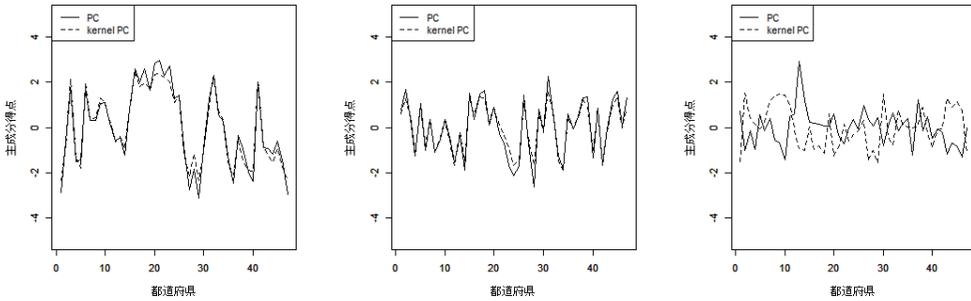
低気温，年間降水量，第2主成分は日照時間の相関が高くなっている．第3主成分については，最高気温の相関が比較的高い．

図2の自然環境分野において，主成分とカーネル主成分を比較している．第1主成分と第2主成分はほぼ同形，第3主成分は異なる部分があるが類似の形となっている．

<経済分野>



<雇用・労働分野>



<家計分野>

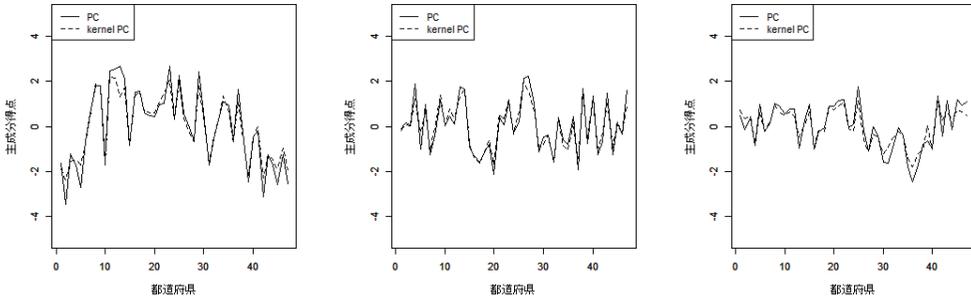
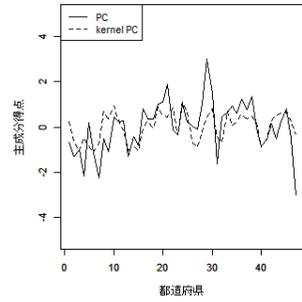
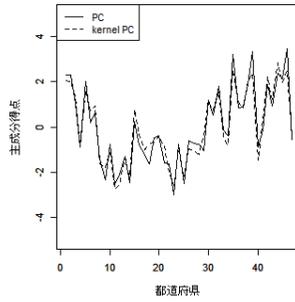
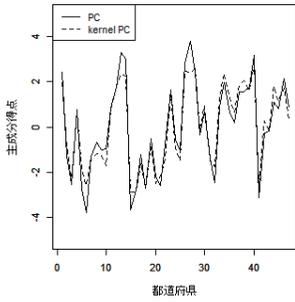
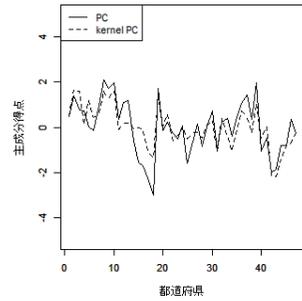
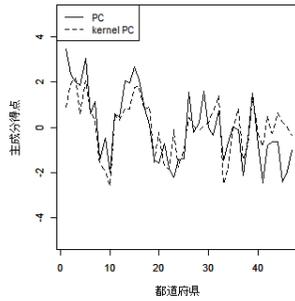
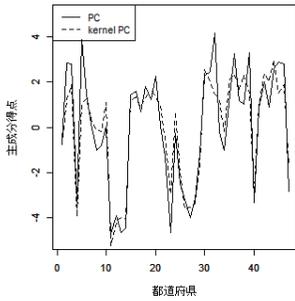


図2 各分野の主成分とカーネル主成分の比較（データ変換後）

<教育・世帯分野>



<都市・人口分野>



<医療・福祉分野>

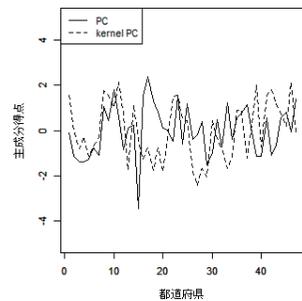
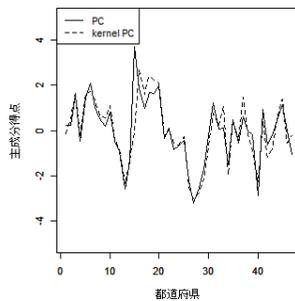
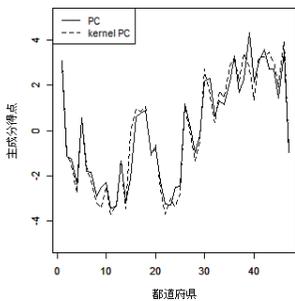
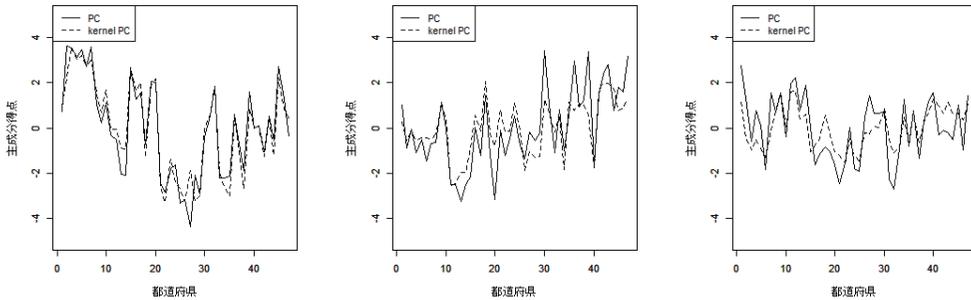


図2 主成分とカーネル主成分の比較 (各分野・データ変換後) (続き)

<健康・生活分野>



<自然環境分野>

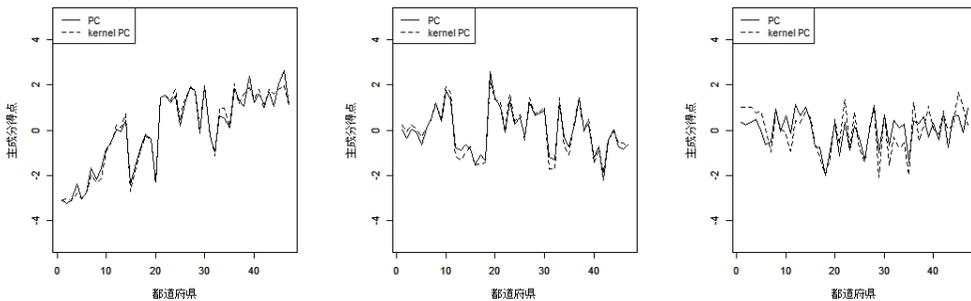


図2 主成分とカーネル主成分の比較（各分野・データ変換後）（続き）

（注1）各図の凡例のPCは主成分，kernel PCはカーネル主成分を表す。

（注2）各図の横軸の都道府県は、左から都道府県コード（全国地方公共団体コードの上2桁の数値）の順に掲載している。すなわち左から、北海道、青森県、岩手県、宮城県、秋田県、山形県、福島県、茨城県、栃木県、群馬県、埼玉県、千葉県、東京都、神奈川県、新潟県、富山県、石川県、福井県、山梨県、長野県、岐阜県、静岡県、愛知県、三重県、滋賀県、京都府、大阪府、兵庫県、奈良県、和歌山県、鳥取県、島根県、岡山県、広島県、山口県、徳島県、香川県、愛媛県、高知県、福岡県、佐賀県、長崎県、熊本県、大分県、宮崎県、鹿児島県、沖縄県と並んでいる。

2. 相関分析

前節で述べた各カテゴリー分野の第1から第3主成分（主成分・カーネル主成分）と、JMDの5つの年層（0歳，20-24歳，40-44歳，60-64歳，80-84歳）の平均余命との相関分析を実施した結果は、表4～7の通りである¹¹⁾。平均余命は、式（Ⅲ-1）によるデータ変換を施している。

11) 脚注7で述べた通り、前掲の主成分・カーネル主成分の比較においては、比較のため、主成分の正負の符号を適宜調整しているが、表4～7においては、正負の符号はそのまま表示している。

表4 平均余命（男）と各カテゴリー分野の主成分との相関

年齢区分	主成分	経済	雇用・労働	家計	教育・世帯	都市・人口	医療・福祉	健康・生活	自然環境
0	第1	0.40	0.28	0.48	0.08	-0.31	-0.12	-0.45	0.21
	第2	0.09	-0.16	-0.01	-0.48	-0.14	-0.34	-0.34	0.08
	第3	0.11	0.30	0.04	0.38	-0.43	-0.66	-0.32	-0.31
20-24	第1	0.42	0.28	0.48	0.08	-0.30	-0.11	-0.45	0.22
	第2	0.13	-0.18	0.00	-0.46	-0.12	-0.37	-0.35	0.03
	第3	0.09	0.26	0.05	0.39	-0.47	-0.64	-0.33	-0.32
40-44	第1	0.37	0.27	0.45	0.03	-0.27	-0.11	-0.36	0.14
	第2	0.10	-0.14	-0.07	-0.43	-0.11	-0.28	-0.36	0.01
	第3	0.12	0.28	0.01	0.31	-0.43	-0.62	-0.34	-0.30
60-64	第1	0.25	0.16	0.31	0.08	-0.20	-0.01	-0.30	0.23
	第2	0.21	-0.11	-0.07	-0.32	-0.14	-0.25	-0.21	0.02
	第3	0.09	0.21	0.10	0.30	-0.38	-0.54	-0.28	-0.19
80-84	第1	-0.06	-0.23	-0.01	0.40	-0.14	0.27	-0.23	0.36
	第2	0.18	0.07	0.05	0.06	-0.18	-0.29	-0.01	0.01
	第3	0.07	0.33	0.02	0.08	-0.06	-0.28	-0.02	0.09

(注) 上表において、太斜字は1%水準で有意であることを示す。

表5 平均余命（男）と各カテゴリー分野のカーネル主成分との相関

年齢区分	主成分	経済	雇用・労働	家計	教育・世帯	都市・人口	医療・福祉	健康・生活	自然環境
0	第1	-0.38	-0.27	0.47	-0.05	-0.25	-0.08	-0.44	-0.26
	第2	0.14	-0.13	-0.01	-0.44	-0.21	-0.26	0.17	0.00
	第3	0.03	0.39	-0.05	-0.17	-0.43	-0.69	-0.20	0.45
20-24	第1	-0.40	-0.27	0.47	-0.06	-0.25	-0.07	-0.45	-0.26
	第2	0.15	-0.15	0.00	-0.43	-0.18	-0.30	0.18	0.04
	第3	-0.02	0.38	-0.04	-0.15	-0.47	-0.65	-0.20	0.43
40-44	第1	-0.34	-0.27	0.43	-0.03	-0.23	-0.07	-0.38	-0.18
	第2	0.14	-0.12	-0.08	-0.39	-0.18	-0.20	0.18	0.06
	第3	0.04	0.35	-0.07	-0.14	-0.44	-0.66	-0.23	0.45
60-64	第1	-0.23	-0.16	0.32	-0.08	-0.19	0.01	-0.33	-0.26
	第2	0.21	-0.10	-0.08	-0.29	-0.17	-0.18	0.10	0.05
	第3	-0.07	0.32	0.00	-0.22	-0.43	-0.58	-0.13	0.32
80-84	第1	0.08	0.23	-0.05	-0.40	-0.15	0.29	-0.24	-0.38
	第2	0.20	0.04	0.02	0.02	-0.14	-0.27	0.07	0.08
	第3	-0.14	0.25	-0.01	-0.11	-0.32	-0.30	0.14	0.10

(注) 上表において、太斜字は1%水準で有意であることを示す。

表6 平均余命（女）と各カテゴリー分野の主成分との相関

年齢区分	主成分	経済	雇用・労働	家計	教育・世帯	都市・人口	医療・福祉	健康・生活	自然環境
0	第1	0.03	0.13	0.13	0.03	-0.04	0.26	-0.27	0.27
	第2	0.23	0.02	-0.13	-0.15	-0.08	-0.22	-0.07	-0.20
	第3	-0.02	0.25	-0.03	0.20	-0.53	-0.54	-0.37	-0.20
20-24	第1	-0.01	0.11	0.09	0.04	-0.02	0.29	-0.23	0.28
	第2	0.27	0.06	-0.10	-0.11	-0.09	-0.19	-0.06	-0.24
	第3	-0.03	0.19	0.02	0.18	-0.55	-0.50	-0.34	-0.20
40-44	第1	-0.04	0.10	0.08	0.06	-0.01	0.31	-0.24	0.30
	第2	0.27	0.08	-0.12	-0.09	-0.13	-0.20	-0.05	-0.22
	第3	0.01	0.21	-0.01	0.16	-0.53	-0.52	-0.36	-0.19
60-64	第1	-0.12	0.04	-0.03	0.05	0.01	0.35	-0.15	0.23
	第2	0.28	0.14	-0.11	-0.03	-0.05	-0.15	-0.02	-0.32
	第3	-0.05	0.27	0.02	0.06	-0.53	-0.43	-0.28	-0.20
80-84	第1	-0.25	-0.10	-0.18	0.21	0.07	0.52	-0.14	0.25
	第2	0.23	0.23	-0.04	0.19	-0.07	-0.16	0.11	-0.28
	第3	-0.11	0.25	-0.04	-0.01	-0.39	-0.34	-0.16	-0.08

(注) 上表において、太斜字は1%水準で有意であることを示す。

表7 平均余命（女）と各カテゴリー分野のカーネル主成分との相関

年齢区分	主成分	経済	雇用・労働	家計	教育・世帯	都市・人口	医療・福祉	健康・生活	自然環境
0	第1	-0.02	-0.13	0.13	-0.03	-0.04	0.28	-0.28	-0.29
	第2	0.15	0.02	-0.18	-0.11	-0.01	-0.15	-0.02	0.27
	第3	-0.20	0.41	-0.10	-0.08	-0.53	-0.62	-0.14	0.32
20-24	第1	0.03	-0.11	0.09	-0.04	-0.03	0.31	-0.23	-0.30
	第2	0.16	0.05	-0.16	-0.08	0.02	-0.14	-0.01	0.30
	第3	-0.25	0.38	-0.05	-0.06	-0.56	-0.57	-0.10	0.29
40-44	第1	0.05	-0.09	0.07	-0.07	-0.02	0.33	-0.25	-0.32
	第2	0.20	0.07	-0.17	-0.06	0.00	-0.14	-0.02	0.29
	第3	-0.23	0.38	-0.07	-0.06	-0.57	-0.58	-0.11	0.31
60-64	第1	0.12	-0.04	-0.03	-0.05	-0.02	0.36	-0.14	-0.24
	第2	0.16	0.11	-0.15	-0.01	0.08	-0.10	-0.02	0.38
	第3	-0.29	0.40	-0.03	0.03	-0.55	-0.50	-0.05	0.30
80-84	第1	0.26	0.10	-0.20	-0.21	0.06	0.54	-0.13	-0.26
	第2	0.10	0.20	-0.11	0.19	0.07	-0.13	-0.09	0.34
	第3	-0.32	0.39	-0.05	0.02	-0.52	-0.36	0.05	0.18

(注) 上表において、太斜字は1%水準で有意であることを示す。

主成分・カーネル主成分の相関を比較すると、第1主成分については、概ね同程度の水準であるが、第2主成分と第3主成分については、主成分とカーネル主成分の異なる部分の影響で相違している場合がある。

男について、0歳の平均余命（平均寿命）との相関係数が1.0%水準で有意となるのは、主成分・カーネル主成分の双方において、第1主成分は、経済、家計、健康・生活の3分野である。第2主成分は、教育・世帯、第3主成分は、都市・人口、医療・福祉となっている。雇用・労働については、0歳、20-24歳において、カーネル主成分の第3主成分が有意となっている。自然環境については、0歳、20-24歳、40-44歳において、カーネル主成分の第3主成分が有意となっており、80-84歳において、カーネル主成分の第1主成分が有意となっている。

女について、0歳の平均余命（平均寿命）との相関係数が1.0%水準で有意となるのは、主成分・カーネル主成分の双方において、都市・人口の第3主成分、医療・福祉の0歳、20-24歳、40-44歳、60-64歳の第3主成分と80-84歳の第1主成分である。カーネル主成分のみ、雇用・労働の第3主成分と自然環境の60-64歳の第2主成分の相関係数が有意である。

3. 重回帰分析

重回帰分析の説明変数の候補として、第1主成分が挙げられるが、多重共線性が生じる可能性がある。表8は、各カテゴリー分野の主成分の第1主成分間の相関である。経済と家計、経済と都市・人口は大きさが0.7以上の相関となっており、その他のカテゴリー分野間についても相関が高い場合があることがわかる。

表8 各カテゴリー分野の主成分の第1主成分の相関

	経済	雇用・労働	家計	教育・世帯	都市・人口	医療・福祉	健康・生活	自然環境
経済	1.00							
雇用・労働	0.09	1.00						
家計	0.71	0.32	1.00					
教育・世帯	0.41	-0.66	0.12	1.00				
都市・人口	-0.80	0.17	-0.54	-0.56	1.00			
医療・福祉	-0.61	-0.36	-0.51	0.14	0.55	1.00		
健康・生活	-0.58	0.14	-0.49	-0.50	0.55	0.03	1.00	
自然環境	0.10	-0.25	0.06	0.51	-0.14	0.37	-0.60	1.00

このため、第1から第3までの主成分の中からAICを基にステップワイズにより説明変数を選択する重回帰分析を実施した。表9は、男女別・各年層別の自由度調整済み決定係数（以下、決定係数）とAICの結果を表示している。第1主成分のみを説明変数の候補として選択した場合は、第1から第3までの主成分を候補として選択した場合と比較し、適合度が低い。第1から第3までの主成分を候補とした場合に適合度が良好となるのは、

男の場合、0歳、20-24歳については主成分、60-64歳、80-84歳についてはカーネル主成分である。40-44歳については、決定係数でみるとカーネル主成分、AICでみると主成分の方が良好である。女の場合、全年層でカーネル主成分の方が良好となっている。これらの結果より、適合度は、主成分が良好な場合とカーネル主成分が良好な場合の双方があることが確認できる。

表9 重回帰分析の結果（決定係数・AIC）

性別	年齢区分	決定係数				AIC			
		第1主成分のみを候補としたステップワイズ		第1から第3主成分を候補としたステップワイズ		第1主成分のみを候補としたステップワイズ		第1から第3主成分を候補としたステップワイズ	
		主成分	カーネル主成分	主成分	カーネル主成分	主成分	カーネル主成分	主成分	カーネル主成分
男	0	0.288	0.331	0.718	0.665	-13.1	-14.2	-49.3	-43.3
	20-24	0.296	0.347	0.735	0.680	-13.6	-15.3	-52.3	-42.8
	40-44	0.241	0.272	0.602	0.628	-8.3	-10.2	-36.9	-36.3
	60-64	0.098	0.147	0.419	0.570	-1.1	-2.8	-14.7	-31.6
	80-84	0.197	0.240	0.542	0.609	-6.5	-7.3	-26.5	-33.3
女	0	0.228	0.271	0.534	0.633	-8.4	-11.1	-27.9	-37.0
	20-24	0.227	0.288	0.542	0.635	-8.3	-12.1	-25.9	-36.6
	40-44	0.237	0.312	0.593	0.694	-8.9	-12.9	-32.1	-46.3
	60-64	0.231	0.299	0.562	0.625	-8.5	-12.9	-30.1	-35.3
	80-84	0.401	0.437	0.617	0.719	-19.4	-22.3	-35.7	-46.3

表10は、男女・各年層別の1%水準で有意な説明変数を示している。選択された変数は、平均余命の差異の要因を探索するための手掛かりとなる情報を提供する¹²⁾が、上述の通り、多重共線性がみられるため、選択されなかった主成分・カーネル主成分にも留意する必要がある。

12) 岩崎（2021）では、重回帰分析の目的が記述の場合、ある説明変数の偏回帰係数は、それ以外の説明変数が説明した残りの中で当該説明変数が説明できる度合いである旨、述べられている。

表10 重回帰分析の結果（1%水準で有意な説明変数）

性別	年齢区分	主成分/ カーネル 主成分	有意な説明変数		有意性	性別	年齢区分	主成分/ カーネル 主成分	有意な説明変数		有意性		
			第1	第3					第1	第3			
男	0	主成分	第1	教育・世帯	***	女	0	主成分	第1	医療・福祉	**		
			第3	教育・世帯	***				第2	教育・世帯	**		
			第3	都市・人口	**			カーネル 主成分	第1	都市・人口	**		
		第3	医療・福祉	**	第3				教育・世帯	**			
		第2	健康・生活	**	第3				医療・福祉	***			
		20-24	カーネル 主成分	第3	教育・世帯			**	20-24	主成分	第1	医療・福祉	***
	第3			都市・人口	***		第2	教育・世帯			**		
	第3			医療・福祉	***		第3	健康・生活			**		
	20-24		主成分	第1	教育・世帯		***	40-44		カーネル 主成分	第3	教育・世帯	**
				第3	経済		**				第3	医療・福祉	***
				第3	教育・世帯		**				主成分	第1	教育・世帯
		第3	都市・人口	**	第1		医療・福祉		**				
		第3	都市・人口	**	第2	教育・世帯	**						
		40-44	カーネル 主成分	第3	教育・世帯	**	40-44		カーネル 主成分	第3	健康・生活	**	
	第3			都市・人口	**	第1		経済		***			
	第3			医療・福祉	***	第1		都市・人口		**			
	40-44		主成分	第2	健康・生活	***		60-64	主成分	第3	教育・世帯	**	
				第3	都市・人口	***				第3	医療・福祉	***	
				第3	医療・福祉	***				第1	医療・福祉	***	
		カーネル 主成分	第3	教育・世帯	**	第1	健康・生活			**			
			第3	医療・福祉	***	第2	教育・世帯			***			
			第3	雇用・労働	**	第3	家計			**			
	60-64	主成分	第2	教育・世帯	**	60-64	カーネル 主成分	第3	健康・生活	***			
			第3	健康・生活	**			第1	経済	**			
第3			教育・世帯	***	第1			健康・生活	**				
カーネル 主成分		第3	都市・人口	**	80-84		主成分	第2	自然環境	**			
		第3	都市・人口	**				第3	医療・福祉	***			
		第3	医療・福祉	***				第1	教育・世帯	***			
80-84	主成分	第1	教育・世帯	***		80-84	カーネル 主成分	第1	医療・福祉	***			
		第3	家計	**				第2	教育・世帯	**			
		第1	経済	**				第3	雇用・労働	**			
	カーネル 主成分	第1	教育・世帯	**	主成分		第3	家計	**				
		第3	家計	**			第3	教育・世帯	**				
		第3	教育・世帯	**			第3	医療・福祉	**				
カーネル 主成分	第3	医療・福祉	***	主成分	第3	医療・福祉	**						

(注) ***, **は、それぞれ0.1%, 1%水準で有意であることを示す。

V. 考察

1. 相関分析

データを各カテゴリー分野に分類し、順位区分尺度によるデータ変換を施した上で、各分野の主成分とカーネル主成分¹³⁾を算出した。データ変換を施した場合、施さない場合と異なり、カーネル主成分は、各分野において第1主成分について概ね同様となり、また、分野によっては第2主成分、第3主成分も概ね同様となる場合があった。このことから、データ変換は外れ値への対処となることに加え、データ変換を施した上で算出したカーネル主成分は変換前のデータに過剰適合することなく一部は主成分と同様に算出され線形成分が抽出されているものと考えられる。

平均余命と各分野の主成分・カーネル主成分との相関分析を実施することにより、これらの基本的な関係を捉えることができる。前章で示した結果では、男については、若年・中年層において相関が有意となる分野が多くみられたが、女については、全年層において相関が有意となる分野は限られている。なお、相関係数は、直接的な因果関係を表しているとは限らないことには、留意が必要である。

多くのデータ項目が平均余命と関連している可能性があるが、主成分・カーネル主成分に縮約することにより、各種要因の探索が可能となる。後述の重回帰分析では、主成分・カーネル主成分を説明変数の候補として、ステップワイズにより変数選択を行い、選択された説明変数である主成分又はカーネル主成分を通じて平均余命の差異の各種要因を探索する。

2. 重回帰分析

男女・各年層の平均余命について、主成分又はカーネル主成分を説明変数の候補として、AICを基準としたステップワイズにより変数選択する重回帰分析を行った。説明変数の候補は、第1主成分のみとした場合、適合度は良好でないが、第1から第3主成分を候補とした場合、いずれのケースも適合度が良好な結果が得られる。このことは、選択された第2・第3主成分に平均余命の差異の要因に関する情報が含まれていることを示唆している。

男の場合、0歳、20-24歳については主成分を説明変数、40-44歳、60-64歳、80-84歳についてはカーネル主成分を説明変数とする方が決定係数でみた適合度の結果が良好である。女の場合、全年層において、カーネル主成分を説明変数とする方が適合度の結果が良好である。このように、主成分が良好な場合とカーネル主成分が良好な場合の双方がみられる。

表10は、男女・各年層について、第1から第3主成分を候補とした場合の1%水準で有意となる説明変数を抽出した結果であり、男女・各年層の結果には同じ主成分又はカーネ

13) Gauss カーネル以外のカーネルを用いる方法も考えられる。

ル主成分が説明変数として選択されている場合があることがわかる。多重共線性があり、選択されなかった主成分又はカーネル主成分も平均余命の差異に影響している可能性があることに留意が必要であるが、ここでは選択された主成分又はカーネル主成分と関連のあるデータ項目を詳しく調べる。

表11と表12は、それぞれ、上述の重回帰分析において選択された0.1%水準で有意な主成分、カーネル主成分について、関連データ項目とその意味、及び、関連データ項目の平均余命への影響について記載したものである。主成分又はカーネル主成分と各項目のデータとの関連は、それらの相関を表す主成分負荷量を通じ把握することが可能であり、表11と表12には、主成分負荷量の大きさが0.6以上の項目（全て0.6を下回る場合は、最大となる項目、又は、カーネル主成分分析の基になっている各項目のデータの2乗との相関が0.6以上の項目）を関連データ項目として記載している。複数の性別・年層について0.1%水準で有意となるのは、教育・世帯分野の主成分の第1主成分、医療・福祉分野の主成分の第1主成分とカーネル主成分の第3主成分である。教育・世帯分野の主成分の第1主成分は、結婚や離婚の多さ、世帯人員数の少なさ、共働きの少なさ等を表しており、平均余命の長短との因果関係の解釈には困難を伴うが、対象となる性別・年層がどの程度社会的な支援を得ることができストレスの少ない生活を送ることができる世帯環境かを意味しているものとの捉え方もあり得るかも知れない。教育・世帯分野の主成分の第1主成分は、男80-84歳についてマイナス、男0歳・20-24歳、女80-84歳についてプラスの平均余命に対する影響が考えられる。医療・福祉分野の主成分の第1主成分は、医療機関等における医療関係者の多さ、医療機関の利用度合、医療費の大きさ等を表しており、医療の充実や利用の度合を表しているものと想定される。医療・福祉分野の主成分の第1主成分は、女20-24歳・60-64歳・80-84歳についてプラスの平均余命に対する影響が考えられる。医療・福祉分野のカーネル主成分の第3主成分は、病気の自覚症状の無さ等を表しており、女の80-84歳を除く男女の全年層において、マイナスの平均余命に対する影響が考えられる。但し、医療・福祉分野のカーネル主成分の第3主成分については、主成分負荷量の大きさが最大のもので0.3を少し超える程度となっており、非線形成分が少なからず内包されていることが考えられる。このため、カーネル主成分分析の基になっている各項目のデータの累乗との相関を調べたところ、データの2乗について、医療施設に従事する看護師・准看護師数、及び、1人当たり国民医療費について0.6以上となった。このことから、これらの項目については、平均からの偏差が大きいと平均余命にプラスの関連があるものと考えられる。すなわち、単位人口当たりの看護師・准看護師数や1人当たり国民医療費が大きく平均からの偏差が大きい場合は、医療の充実により平均余命が長くなるものと想定される。逆に、単位人口当たりの看護師・准看護師数や1人当たり国民医療費が小さく平均からの偏差が大きい場合は、解釈が難しいが、高齢化があまり進んでおらず医療をそれ程必要としない若年層が多く、高齢者が受ける医療は充実しており平均余命にプラスの影響を及ぼしている状況等の可能性が考えられる。

表11 0.1%水準で有意な主成分と関連データ項目・考えられる意味

分野 / 主成分	関連データ項目 (主成分負荷量の符号)	主成分の意味 / 関連データ項目の 平均余命への影響	有意となる 性別・年層 (偏回帰係数の符号)
教育・世帯 主成分の 第1主成分	婚姻率 (+), 離婚率 (+), 一般世帯の平均人員数 (-), 共働き世帯割合 (-), 単独世帯割合 (+)	結婚や離婚の多さ, 世帯人員数の少なさ, 共働きの少なさ 男・80-84歳においてマイナス, それ以外はプラスの影響	男: 0歳 (+) 20-24歳 (+) 80-84歳 (-) 女: 80-84歳 (+)
教育・世帯 主成分の 第2主成分	最終学歴が大学・大学院卒の者の割合 (-), 高等学校卒業者の進学率 (-), 高等学校教育費 (+), 婚姻率 (-), 高齢単身世帯の割合 (+)	学歴の低さや進学のし難さ 女・60-64歳において, マイナスの影響	女: 60-64歳 (-)
教育・世帯 主成分の 第3主成分	高齢夫婦のみの世帯割合 (+)	高齢夫婦のみ世帯の多さの状況 男・0歳において, プラスの影響	男: 0歳 (+)
都市・人口 主成分の 第3主成分	火災出火件数 (+)	居住地の火災出火の状況 男・0歳において, マイナスの影響	男: 40-44歳 (-)
医療・福祉 主成分の 第1主成分	一般病院平均外来患者数 (+), 医療施設従事医師数 (+), 医療施設従事看護師・准看護師数 (+), 一般病院平均在院日数 (+), 1人当たり国民医療費 (+), 後期高齢者医療費 (+), 介護福祉士登録者数 (+)	医療機関等における医療関係者の多さ, 医療機関の利用度, 医療費の大きさ 女・20-24歳, 60-64歳, 80-84歳において, プラスの影響	女: 20-24歳 (+) 60-64歳 (+) 80-84歳 (+)
医療・福祉 主成分の 第3主成分	通院率 (-)	通院の少なさの状況 男・40-44歳において, マイナスの影響	男: 40-44歳 (-)
健康・生活 主成分の 第2主成分	趣味娯楽行動割合 (-)	趣味娯楽行動の少なさの状況 男・40-44歳において, マイナスの影響	男: 40-44歳 (-)
健康・生活 主成分の 第3主成分	ボランティア参加率 (-), メタボリックシンドローム予備群者割合 (+)	ボランティア活動の少なさや不健康リスクの多さ 女・60-64歳において, マイナスの影響	女: 60-64歳 (-)

表12 0.1%水準で有意なカーネル主成分と関連データ項目・考えられる意味

分野 / 主成分	関連データ項目 (主成分負荷量の符号, 大きさが 0.6より小さい場合は 最大値, 又は, 0.6以上の 2次系列との相関)	カーネル主成分の意味 / 関連データ項目の 平均余命への影響	有意となる 性別・年層 (偏回帰係数の符号)
経済 カーネル主成分の 第1主成分	1人当たり県民所得(-), 課税対象所得(-), 財政力指数(-)	経済状況の低調度 女・40-44歳において, プラスの影響	女: 40-44歳(+)
教育・世帯 カーネル主成分の 第3主成分	高齢夫婦のみの世帯割合(-)	高齢夫婦のみ世帯の少なさの状況 男・60-64歳において, マイナスの影響	男: 60-64歳(-)
都市・人口 カーネル主成分の 第3主成分	年少人口比率(-0.58)	年少人口が占める割合の少なさ 男・0歳において, マイナスの影響	男: 0歳(-)
医療・福祉 カーネル主成分の 第3主成分	有訴率(-0.32) <各項目のデータの2乗との相関> 医療施設に従事する看護師・准看護師数(+), 1人当たり国民医療費(+)	病気の自覚症状の無さ 女の80-84歳を除く男女・全年層において, マイナスの影響 <各項目のデータの2乗との相関から考えられるカーネル主成分の意味> 看護師・准看護師数や1人当たり国民医療費の偏差の関連性	男: 0歳(-) 20-24歳(-) 40-44歳(-) 60-64歳(-) 80-84歳(-) 女: 0歳(-) 20-24歳(-) 40-44歳(-) 60-64歳(-)

3. 実務における利用

利用した方法による各分野の主成分・カーネル主成分や関連データ項目を, 都道府県別の平均余命(男女別・年層別)との相関関係や重回帰分析の結果と併せて関連指標として提示し, 各地域の人口に関わる運営や諸政策の立案の参考とする対応が考えられる。また, これらの主成分・カーネル主成分を通じて, 各分野の主成分又はカーネル主成分の関連データ項目が変動した場合の各年層の平均余命への影響を調べることも可能となる。直近の平均余命の差異の要因分析に留まらず, 社会経済等の要因を踏まえた中長期的な将来死亡率の変動を評価するには, より多くの種類・数のデータを用いて分析し, モデリングするこ

とが必要となろう。このような様々な要因を踏まえた将来死亡率の推計や不確実性の評価は、地域や国の人口構造の分析、社会保障・健康等の政策立案、公的年金・私的年金の財政や保険・共済の財務の運営に活用することが可能である^{14, 15)}。

VI. 結び

都道府県間の平均余命の差異の要因を、関連の可能性のある政府統計・調査データを複数のカテゴリー分野に分類し、順位区分尺度によりデータ変換した上で主成分とカーネル主成分を算出し、それを基に、相関分析と重回帰分析を実施する方法を利用し、要因を探索した。カテゴリー分野の分類や各分野のデータ項目は、モニタリングしながら、その時々にも最適な分類とし項目を抽出する対応が望ましい。また、新型コロナウイルスの感染拡大は、経済や雇用、医療・福祉機関や国民生活の様々な面に影響を及ぼしており、こうした社会経済の変化の中で、どのように人口が変動し将来的に推移して行くか、不確実性も考慮しながら関連する内容も含めて分析を実施して行くことが必要であろう。

謝辞

本稿執筆にあたり、匿名査読者、日本人口学会第74回大会（2022年）自由論題 E-1セッション参加者、並びに、社人研一般会計プロジェクト「超長寿社会における人口・経済・社会のモデリングと総合分析」の委員各位より、複数回にわたり多くの貴重な御意見・コメントをいただいたことに、心より感謝を申し上げます。

（2022年9月27日査読終了）

参考文献

- Bonnar, S., Curtis, L., Leon-Ledesma, M., Oberoi, J., Rybczynski, K., and Zhou, M. (2018) "Population Structure and Asset Values," *International Congress of Actuaries 2018, Berlin*.
- Fukuda, Y., Nakamura, K., and Takano, T. (2005) "Accumulation of health risk behaviours is associated with lower socioeconomic status and women's urban residence: a multilevel analysis in Japan," *BMC Public Health*, Vol. 5, 53.
- Fukuda, Y., Nakamura, K., and Takano, T. (2004) "Wide range of socioeconomic factors associated with mortality among cities in Japan," *Health Promotion International*, Vol. 19, pp.177-187.
- Kagamimori, S., Gaina, A., and Nasermoaddeli, A. (2009) "Socioeconomic Status and Health in the Japanese Population," *Social Science & Medicine*, Vol. 68, pp.2152-2160.
- Kennedy, B. P., Kawachi, I., and Prothrow-Smith, D. (1996) "Income Distribution and Mortality: Cross

14) 佐藤他（2021）、佐藤他（2020）や佐藤（2020）は、我が国の公的年金の財政検証における経済前提とマクロ計量モデルについて取り扱っており、Bonnar, S. et al. (2018) は、人口構造と資産価格のモデルをカナダの年金に適用し考察している。

15) 広井（2019）、嶺（2019）は、AI手法による人口・社会・経済の予測を用いて、持続可能な政策に関わる提言を行っている。

- Sectional Ecological Study of the Robin Hood Index in the United States," *British Medical Journal*, Vol. 312, pp.1004-1007.
- Shibuya K., H. Hashimoto, and E. Yano (2002) "Individual Income, Income Distribution, and Self Rated Health in Japan: Cross Sectional Analysis of Nationally Representative Sample," *British Medical Journal*, Vol. 324, pp.16-19.
- Wilkinson, R. G. (1992) "Income Distribution and Life Expectancy," *British Medical Journal*, Vol. 304, pp.165-168.
- 井川孝之 (2017) 「平均余命の地域差を表す社会・経済指標と死亡率推計」『国立社会保障・人口問題研究所 2014～2016年度人口問題プロジェクト研究 長寿化・高齢化の総合的分析及びそれらが社会保障等の経済社会構造に及ぼす人口学的影響に関する研究－第3報告書－』, pp.155-192.
- 石井太 (2015) 「日本版死亡データベースの人口分析への応用」『人口問題研究』第71巻, 第2号, pp.141-155.
- 岩崎学 (2021) 「統計的因果推論の視点による重回帰分析」『日本統計学会誌』第50巻, 第2号, pp.363-379.
- 川上憲人・小林廉毅・橋本英樹 (2006) 『社会格差と健康－社会疫学からのアプローチ』東京大学出版会.
- 北島晴美・太田節子 (2004) 「都道府県別平均寿命の分布の変遷と気候の影響」『信州大学山地水環境教育研究センター研究報告』第3号, pp.53-75.
- 京都大学 (2007) 『健康と経済社会的属性との関係に関する調査研究報告書』.
- 高俊珂・梯正之 (2006) 「都道府県別の平均寿命と社会・経済指標および栄養指標との関連性」『広島大学保健学ジャーナル』第5巻, 第2号, pp.62-69.
- 国立社会保障・人口問題研究所 (2017) 『日本の将来推計人口 (平成29年推計)』.
- 国立社会保障・人口問題研究所 「日本版死亡データベース」,
<https://www.ipss.go.jp/p-toukei/JMD/index.asp> (2021年11月3日ダウンロード).
- 近藤克則 (2005) 『健康格差社会－何が心と健康を蝕むのか』医学書院.
- 坂井博通 (1986) 「日本人の60歳時平均余命と社会経済的要因の関連に関する一考察」『人口問題研究』第180号, pp.46-51.
- 佐藤格・石井太・増田幹人 (2020) 「2019年財政検証における経済前提と整合的なマクロ計量モデル開発のための予備的研究」『国立社会保障・人口問題研究所2017～2019年度人口問題プロジェクト研究 長寿革命に係る人口学的観点からの総合的研究－第3報告書－』, pp.89-99.
- 佐藤格 (2020) 「経済前提と財政検証」『社会保障研究』第4巻, 第4号, pp.445-459.
- 佐藤格・石井太・増田幹人 (2021) 「マクロ計量モデルにおける賃金率・利子率の決定方法の整理」『国立社会保障・人口問題研究所2020～2022年度人口問題プロジェクト研究 超長寿社会における人口・経済・社会のモデリングと総合分析－第1報告書－』, pp.51-56.
- 七田恵子 (2010) 「長野県高齢者の健康に関する指標の検討」『佐久大学看護研究雑誌』第2巻, 第1号, pp.51-58.
- 杉澤秀博 (2012) 「健康の社会的決定要因としての社会関係：概念と研究の到達点の整理」『季刊 社会保障研究』第48巻, 第3号, pp.252-265.
- 鈴木健二 (2003) 「各種社会指標と都道府県別生命表の関係」『厚生指標』第50巻, 第5号, pp.30-35.
- 竹内光・關雅夫 (2013) 「平成22年度都道府県別生命表における平均寿命の地域差分析」『厚生指標』第60巻, 第16号, pp.32-39.
- 田辺和俊・鈴木孝弘 (2020) 「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－」『統計数理』第68巻, 第1号, pp.175-192.
- 田辺和俊・鈴木孝弘 (2015) 「平均寿命および健康寿命の都道府県格差の解析」『季刊 社会保障研究』第51巻, 第2号, pp.198-210.
- 豊田哲也 (2011) 「都道府県別に見た世帯所得の分布と平均寿命の変化－地域の所得格差は健康を損なうか－」『徳島大学総合科学部 人間科学研究』第19巻, pp.87-100.
- 仲都留隆・大西雄基 (2008) 「都道府県別生命表による平均寿命の地域差分析」『厚生指標』第55巻, 第5号, pp.44-53.
- 長野県健康長寿プロジェクト・研究事業 研究チーム (2015) 『長野県健康長寿プロジェクト・研究事業報告書～長野県健康長寿の要因分析～』.
- 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』第22巻, 第1号, pp.5-17.

- 広井良典（2019）『人口減少社会のデザイン』東洋経済新報社。
- 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』第56巻，第2号，pp.56-62.
- 堀内四郎（2010）「日本人の寿命伸長：要因と展望」『人口問題研究』第66巻，第3号，pp.40-49.
- 嶺竜治（2019）「持続可能な未来の実現に資する「政策提言AI」」『日立評論』第101巻，第3号，pp.386-391.
- 山口扶弥・梯正之（2001）「高齢者の平均自立期間および要介護期間に関連する諸要因の分析」『人口問題研究』第57巻，第4号，pp.51-67.
- 山本正治・土屋康雄・遠藤和男・斉藤トシ子・大西秀明・高橋榮明（2009）「都道府県別新国民生活指標と平均寿命との関連性について」『新潟医療福祉学会誌』第9巻，第2号，pp.5-9.

付録 カーネル主成分分析

主成分分析では，分散共分散行列又は相関係数行列の固有値問題を解くことにより，主成分を算出するが，カーネル主成分分析では，各データより算出したカーネルによる以下のグラム行列（都道府県別データであれば， n は都道府県の数）の固有値問題を解くことにより，カーネル主成分を算出する．

$$K = \begin{bmatrix} k(X_1, X_1) & \cdots & k(X_1, X_n) \\ \vdots & \ddots & \vdots \\ k(X_n, X_1) & \cdots & k(X_n, X_n) \end{bmatrix}$$

式（Ⅲ-2）の Gauss カーネルのパラメータ β の設定は，データの誤差の分散がある程度把握できる場合はそれを基に設定する方法や，1つの標本を抜いた場合の推定値との誤差を表す one-leave-out クロスバリデーション誤差を最小にする値とする方法等があるが，どの方法を用いるかは必ずしも明確でない． β が大きいほど，標本誤差は小さくなるが，one-leave-out クロスバリデーション誤差は大きくなる．本稿では， β をオーダー別（1, 0.1, 0.01）に調べ，0.1と0.01の場合について記載している．

Exploring Factors on the Difference of Life Expectancy between the Japanese Prefectures: the Application of the Method Using the Data Transformation by Ordinal Scale and Contraction

IGAWA Takayuki

In this paper, we explore the factors on the difference of life expectancy by age and gender between the Japanese prefectures using the data transformation method by ordinal scale and kernel cumulative probability distribution, the principal component and the kernel principal component based on the data from the Japanese Mortality Database, the government statistics and another statistical surveys. We conduct correlation analysis and regression analysis with these principal components by categories. We also consider the factors in relation to life expectancies based on the analysis.

In the principal component analysis, we confirm that the first principal component and the first kernel principal component are similar. As for the second and more order, some principal components are different from corresponding kernel principal components. In the regression analysis using the selection of explanation variables by stepwise, we can confirm that there are two cases, one of which is that the regression by the principal components as explanation variables has better fit, and the other of which is that the regression by kernel principal components as explanation variables has better fit. Through the principal components, the kernel principal components and the regression analysis, we find that some of the factors on the difference of life expectancy are related to marriage, divorce, household status and medical care.

Key words: life expectancy, socio-economic factors, data transformation, kernel principal component analysis, local area correlation analysis