

投稿：論文

## 機械学習による都道府県別医療費の分析

須田 茂夫\*

## 抄 録

本研究では、機械学習の手法であるランダムフォレストおよび回帰木により26の指標を用いて分析を行い1人当たり医療費の水準によりグループ分けを行った。その結果、医療提供体制である看護師数、病床数、医療機関数を基準に5つのグループに分かれた。

グループ別の平均値の比較により、医療費の高い地域ほど医師数、看護師数も多いが、それ以上に病床数が多いため、病床あたりの配置人員は手薄になっており、平均在院日数が長くなっていることが確認された。

また、1人当たり医療費の高い地域ほど介護サービス受給者数が多い傾向があるが、最も医療費の高いグループでは介護福祉施設定員およびサービス受給者数は中程度であった。そのほか、特定健康診査受診率、在宅死亡率が高いほど1人当たり医療費が低くなっていることが確認された。

キーワード：医療費の地域差、看護師、医療の供給体制、ランダムフォレスト、回帰木

社会保障研究 2018, vol.3, no.3, pp.403-415.

## I はじめに

財政健全化が重要課題とされるなか、医療費は年間約1兆円の増加が見込まれており、その適正化が求められている。都道府県別の1人当たり医療費は2倍近い差があり年齢構成を調整しても1.5倍程度の差となっており、このような地域差を縮小させることが医療費の適正化の目標とされている<sup>1)</sup>。また、医療費の地域差には「西高東低」という傾向があることも以前から知られており、全国で同一の診療報酬、医療技術の普及や制度の差がないと仮定されるにもかかわらず生じている地域

差とその傾向の要因については、先行研究においてさまざまな検証がなされてきた。

一般に医療費に影響を与えるものとしては、長期の時系列分析では高齢化や医療技術の進歩、クロスセクション分析による国際比較では所得水準などがあげられてきたが、国内における医療費の地域差は医療の供給体制、患者の受療行動等に起因すると考えられてきた〔郡司 (2001)〕。

特に、医療費と同様に地域差が大きく西高東低の傾向もあることから人口当たりの病床数の影響が大きいと考えられ〔府川 (2006)〕、前田 (1983) 以来、人口当たりの病床数が多くの研究で取り上げられてきた。同様に医療の供給体制に関するも

\* 日本経済政策学会、公共選択学会 会員

<sup>1)</sup> 「経済・財政再生アクション・プログラム2016」(経済財政諮問会議)。

のとしては、医療提供者間の競争が激しいこと〔中西(2001)〕、適用の判断基準が厳格でない薬剤など〔池田・望月(2001)〕、医療機関の属性(開設者、病床数、看護体制)によるもの〔泉田(2001)〕、医療における大きな設備投資がさらなる需要を誘発させていること〔郡司(2001)〕、また、患者の行動としては大規模な医療施設が選好されること〔青木(2001)〕などが指摘されている。

医療の供給体制以外の要因を検討したものとして、保健師数、保険事業費など保健行政について〔足立・赤井・植松(2012)〕、介護施設、健診受診率等について〔罫(2013)、印南(2016)〕がある。

これらの研究の前提となる有力な考え方は、医療提供者と患者との間の情報の非対称性を考慮した供給者誘発需要仮説である。誘発需要の生じるメカニズムにはいくつかの説があるとともに多くの実証研究があるが技術的な問題もあり現在も確定した結論はない<sup>2)</sup>が、実際の医療政策においても医療供給者の行動、提供体制に着目した施策が実施されてきた。代表的なものとして、公定価格のもとでの出来高払い制における過剰診療、老人医療費の無償化政策時代における病床の増加等が問題とされ、包括払い制の導入、病床規制等があげられる。

現在、2014年に制定された「医療介護総合確保推進法」により、各都道府県において「地域医療圏構想」の策定が進められ、病床の機能分化・連携の推進により病床数を適正水準にし、医療従事者の確保・育成することが重要課題とされている。

先行研究における分析方法としてクロスセクションによる回帰分析では、サンプル数が限られているため、特定の論点から導かれた少数の要因の影響を検証するものとなっており、多くの要因の影響力の大きさを総合的に比較するものとはなっていない。また、印南(2016)では、多重共

線性的問題があるため病床数と医師数のように相関が高い変数を同時に検証することも不可能であったことを指摘している。

時系列要素を含むパネルデータでは、観測数が増えるためより多様な要因を考慮した検証が可能であり、病床規制、診療報酬改定など制度変更による政策の効果も検証可能となる反面、現時点の問題点を把握するには適切とは言えない部分があると考えられる。また、集計情報を用いた医療需要分析は異時点間の推計結果が不安定になるという問題が指摘されている〔橋本・泉田(2011)〕。

本稿では、機械学習ないしはデータマイニングと呼ばれる手法である回帰木およびランダムフォレストを用いて分析を行う。

回帰木は、説明変数を値や範囲などで分割させ予測や判別のルールを構築し、「樹木」によって視覚的に表現する手法であり、分類木と併せて決定木と呼ばれるものである<sup>3)</sup>。この方式の統計学的基礎は、層別解析であり、これを多層的に適用するものである。樹木構造を成長させるための分岐点を算出する方法は種々のものがあるが、回帰木の場合は平方和の分解を利用する。

目的変数の偏差平方和を

$$DS = \sum_{i=1}^N (Y_i - \bar{Y})^2 \quad (1-1)$$

とする。ここでNはデータ数、 $\bar{Y}$ は平均値である。

任意の説明変数により2分岐させ、右に $N^R$ 個、左に $N^L$ 個のデータに分割し、それぞれの偏差平方和を求める。

$$DS_{TR} = \sum_{i=1}^{N^R} (Y_i - \bar{Y}_R)^2 \quad (1-2)$$

$$DS_{TL} = \sum_{i=1}^{N^L} (Y_i - \bar{Y}_L)^2 \quad (1-3)$$

これらより、元の偏差平方和と分岐先の平方和の

<sup>2)</sup> 井伊・別所(2006)、橋本・泉田(2011)。

<sup>3)</sup> 回帰木と分類木は、従来の多変量解析における重回帰分析と判別分析に相当するものであり、基準となる変数が連続変数と質的変数の場合に対応している。したがって、連続変数であるデータを分類処理したものは分類木ではなく回帰木となる。

合計の差を求める。

$$DS_{TB} = DS - (DS_{TR} + DS_{TL}) \quad (1-4)$$

これは、回帰木ではノード間平方和と呼ばれるものであり（分散分析では級間平方和）、説明変数による目的変数の分離の基準とされるものである。これを説明変数ごとに計算し、最大になるものにより分岐を進め、回帰木を成長させるものである。

この分岐を進め、どこまで細分化する（樹木を成長させる）かについて、回帰木では「交差妥当化」という方法により、最適な水準を決定する。これは、枝が複雑なほど詳細に特徴を捉えられるが、予測・分類の精度は低くなる（本稿の場合、1人当たり医療費により分割したグループ間の順位の子となる部分が多くなる。）というトレードオフ関係を利用したものである。

このような回帰木の特徴としてあげられる第一の点は、目的変数に対してグループ化が行われることである。同様な多変量データを用いたものとしては、クラスター分析等があるが、説明変数による類似度の高いものをグループ化しているものである。このため、回帰木では政策目標とされる変数によりグループ化することが可能であるのに対し、クラスター分析では政策目標となる変数の平均値等を事後的に算出、比較するものでしかない。

第二の特徴は、分岐が説明変数により直接行われ、グループ間の相違が明確である点である。機械学習の各種手法においては、「人口知能の闇」という言葉に示されるように判別・分類の方法が複雑・抽象度が高いため<sup>4)</sup>、解釈が難しいという問題が指摘されているが、この点で回帰木の明確さは大きな利点である。

第三の特徴は、分岐が進む中で、次に分岐の基準となる変数がグループにより異なることを示す点<sup>5)</sup>が大きな特徴である。この点は、本稿の分析

におけるように説明変数が互いに相関していると考えられるものの関係に踏み込んだ解釈が可能となる。

次に、ランダムフォレストは集団学習という必ずしも精度が高くない学習機の結果を統合・組み合わせることで精度を向上させる手法のひとつである。ベースとなるのは回帰木であるが、説明変数をランダムに抽出したデータセットを複数作り、それらについて回帰木を実行した結果から変数の重要度（寄与度）を算出するものである。

今回用いた回帰木等のデータマイニングと呼ばれる手法の一般的特徴ないし従来の統計手法の観点からすれば問題点としてあげられるのが、仮説検定手法がなく、結果の一意性・最適性が保証されていないという点である。回帰木では、この点について、先述の交差妥当化処理という、最適な分割数を決定する手法を備えている。また、ランダムフォレストによる変数選択は、回帰分析における赤池情報量基準による変数選択に相当するものと考えられる。本稿では、これらを併用することにより結果の妥当性・最適性を確保しようとするものである。

本稿では、まずランダムフォレストを実行し変数選択した後、回帰木を実行する。それにより作成される都道府県別の1人当たり医療費の水準のグループについて平均値による比較・検討を行う。この方法は、回帰木では、多重共線性が起こりにくいとされているため、これまで同時には扱うことのできなかった変数（医師数、看護師数、病床数等）を同時に用いての比較・検討が可能になる。

## II 分析

### 1 使用するデータ

本稿では、以下の変数の2014年（平成26年）分該当のクロスセクション・データにより分析を行った。

<sup>4)</sup> 例としてクラスター分析では、説明変数で構成される多次元空間上の距離によりグループ化される。

<sup>5)</sup> 回帰木の有効性が注目されるようになったものとして、住宅価格の分析があり、「小学校までの距離」といった個々の要素が、物件の規模（部屋数）により価格に対する影響が異なる点を示せることである。

目的変数は、都道府県ごとの医療費を検証するものとして、先行研究にならない地域保険である国民健康保険の厚生労働省「医療費の地域差分析」の基礎データを用いた。同データのうち「市町村国民健康保険」および「後期高齢者医療制度」と両方を合わせた「1人当たり実績医療費」（以下1人当たり医療費）の値を用いた。先行研究では、後期高齢（老人医療費）を対象に分析したものも多いが、本稿が対象とした2014年は、7対1入院基本料届出病床数が非常に多いことが特徴であり、看護配置基準が高い病床ほど後期高齢者の比率が低いことが確認されているため<sup>6)</sup>、目的変数は後期高齢に絞らず全年齢を対象としたものである。

説明変数は表1の26の指標を用いた。医療の供給体制に関するものとして、医療従事者（すべて

人口10万人当たり）である医師数、歯科医師数、薬剤師数を用いた。また、医療従事者のうち看護職員として看護師数、准看護師数、総看護師数（看護師+准看護師）、総看護師数に対する看護師の比率を用いた。看護職員に関して合成指標を用意したのは、本稿における分析において留意すべき看護師と准看護師との相違点として、7対1～13対1の高い入院基本料を算定する際に、看護職員の70%以上が看護師であることが要件となっているためである。

保健師も医療従事者であるが、後述の予防関連、健康指導の従事者としての側面もある。

医療施設関連のものとしては、病床数、医療機関数（いずれも病院と診療所を合わせたもので人口10万人当たり）を用いた。また、供給者誘発需

表1 説明変数一覧

変 数 名	出 典
医師数（人口10万当たり） 歯科医師数（人口10万当たり） 薬剤師数（人口10万当たり）	医師・歯科医師・薬剤師調査（厚生労働省）
保健師数（人口10万当たり） 看護師数（人口10万当たり） 准看護師（人口10万当たり） 総看護師数（人口10万当たり） 総看護師数に対する看護師の比率	衛生行政報告例（厚生労働省）  * *
病床数（人口10万当たり） 医療機関数（人口10万当たり）	医療施設調査（厚生労働省）
病床あたり医師数 病床あたり総看護師数 病床あたり看護師数	* * *
介護サービス受給者総数（人口10万当たり） 介護施設サービス受給者総数（人口10万当たり） 介護福祉施設サービス受給者総数（人口10万当たり）	介護給付費実態調査報告（厚生労働省）
介護老人福祉施設定員（人口10万当たり）	介護サービス施設・事業所調査（厚生労働省）
特定健康診査受診率 特定保健指導実施率 メタボリックシンドローム該当者割合	特定健康診査・特定保健指導に関するデータ（厚生労働省）
老人クラブ会員加入率（対老年人口） 民生委員相談件数（健康・保健医療：人口10万当たり）	福祉行政報告例（厚生労働省）
在宅死亡率	人口動態調査（厚生労働省）
65歳以上人口（千人）	人口推計（総務省）
1人当たり県民所得（千円）	県民経済計算（内閣府）
高齢者就業割合（65歳以上）対老年人口	社会生活統計指標－都道府県の指標－（総務省）

注1：出典欄の\*は他指標より算出したもの。

注2：介護関連と民生委員相談件数は人口当たりとするため、総人口（人口推計（総務省））を用いた。

<sup>6)</sup> 厚生労働省（2017）p.7。

要仮説の観点から、医療施設当たりの供給体制の手厚さを示すものとして、病床当たりの医師数、総看護師数、看護師数を用いた。

医療の供給体制以外のものとして、介護サービスの受給状況を示す介護サービス受給者総数、入院医療と関連が高いと想定される介護施設サービス受給者総数、介護福祉施設サービス受給数、介護福祉施設定員を用いた。予防、健康指導関連のものとして特定健康診査受診率、特定保健指導実施率、メタボリックシンドローム該当者数を用いた。健康な状態の維持に対し、影響を与える可能性があるソーシャルキャピタルを示すものとして、老人クラブ会員数、民生委員相談件数（福祉関連）を用いた。そのほか、社会・経済関連の指標として、在宅死亡率、65歳以上人口、1人当たり県民所得、高齢者就業割合（65歳以上）を用いた。

## 2 ランダムフォレストおよび回帰木による分析

本稿で用いたソフトウェアは「R version 2.15.1 (2012-06-22)」であり、決定木 (mvp) およびランダムフォレスト (randomForest) のパッケージを使用した。先に述べたように、ランダムフォレストと回帰木による以下の手順により分析をおこなった<sup>7)</sup>。

(1) すべての説明変数を用いランダムフォレストを実行、その際に生成する回帰木の樹木数は、システムの規定値である500を用いた。生成する樹木数の妥当性を検証する方法として、樹木数に対する平均2乗残差の挙動を示すグラフにより確認することとされており、本稿の分析では、約30本のあたりまで平均2乗残差が急減し、その後若干増えた後、約90本以降は定常状態となったので、生成回数は十分であると判断した。

(2) ランダムフォレストで重要度が高いことが示された変数により分割数の制限は設けず回帰木を行った。

(3) 「交差妥当化」確認は回帰木の枝の数とエラーの度合いを示すグラフをにより行うこととされているが、(2)の結果よりグラフを作成、確認したところ分割数は5が最適であると判断した。

(4) 分割数の制限を設定して回帰木分析を行った。

上記(2)～(4)を繰り返し、回帰木の分岐の状況から、分岐のために使用する変数の候補数は5つで十分であると判断し、ランダムフォレストでの上位5個のデータを用いたものを最終の分析結果とした。

(5) 回帰木分析により分岐したグループごとに各説明変数について平均値を算出した。

また、平均値は回帰木で用いた説明変数に加え、「医療費の地域差分析」にある平均在院日数および「市町村国民健康保険」と「後期高齢者医療制度」別の医療費の3要素（受診率、1件当たり日数、1日当たり医療費）と、病床機能報告制度による報告結果<sup>8)</sup>より、高度急性期、急性期、回復期、慢性期の全体に占める比率（%）も用いた。

## III 結果

### 1 ランダムフォレストによる分析結果

図1は、ランダムフォレストにより算出された重要度をグラフ化したものである。「%IncMSE」とは、ランダムに抽出されたデータにより作成した回帰木に対し、抽出されなかったデータを当てはめた結果より算出される特徴量であり、「IncNodePurity」とは、枝を分岐させていく際の平均2乗誤差により算出された特徴量である。前者は試行処理をするため計算時間がかかるのに対し、後者は回帰木を作成する際の値を用いるので計算時間は短い精度は劣るとされている。

算出方法の違いにより結果は異なるが、重要度の高い上位5指標が、総看護師数、病床数、看護師

<sup>7)</sup> 回帰木、ランダムフォレストの原理や用語の詳細、分析方法については、金（2007）、下川他（2013）、豊田（2008）等を参照されたい。

<sup>8)</sup> 「病床機能報告制度による報告結果（2014年（平成26年）7月1日時点の医療機能別の病床数）について」平成27年6月15日 医療・介護情報の活用による改革の推進に関する専門調査会 資料2-1 p.33（2017年11月27日最終確認）。

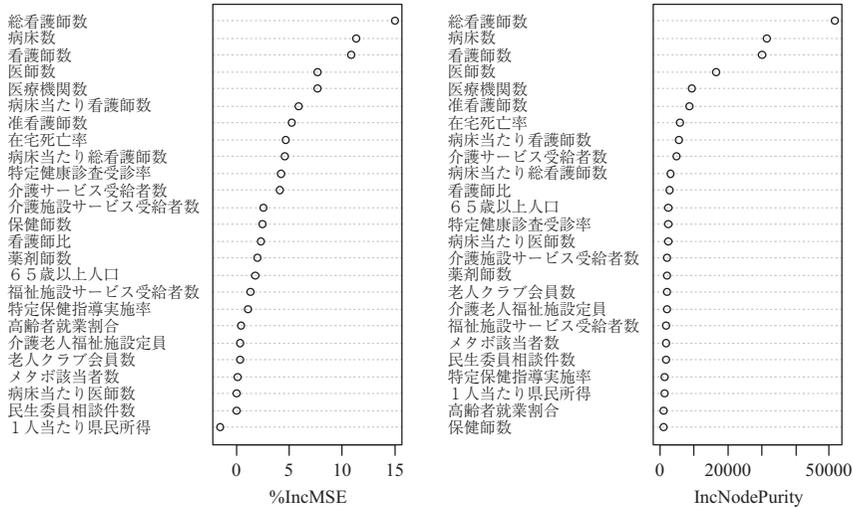


図1 ランダムフォレストによる変数重要度

数、医師数、医療機関数となり、すべて医療の供給体制に関するものとなった。これらの変数は、1人当たり医療費に対する単回帰による決定変数も高く、総看護師数 (0.84)、病床数 (0.84)、看護師数 (0.86)、医師数 (0.67)、医療機関数 (0.65) となっており、ランダムフォレストの結果もこれを反映したものとなっている。上位5指標については、相互に相関係数の高いものが多く、特に高いものは、総看護師数と看護師数 (0.957)、総看護師数と病床数 (0.956)、看護師数と病床数 (0.915)、医師数と医療機関数 (0.808)、医師数と看護師数 (0.606) であり、最低は医療機関数と病床数 (0.554) であった。医療の供給体制以外では、在宅死亡率、介護サービス受給者総数、特定健康診査受診率、民生委員相談件数などが高いものとなった。

2 回帰木による分析結果

ランダムフォレストの結果の上位5個を説明変数とし、回帰木を行った結果が図2である。全体が5つのグループに分割され、図下部にはそれぞれの1人当たり医療費 (単位 千円) に対する箱ひげ図が表示されている。

分岐の状況を確認すると、最初に総看護師数で

2つに分岐し、次にそれぞれのグループで病床数と看護師数で分岐し、看護師数で分けられたグループでは、さらに医療機関数で分岐している。ランダムフォレストで影響が大きい3つの変数がまず分岐の基準となり、次が医療機関となった。医師数はランダムフォレストの結果や決定係数では、医療機関数よりも影響がやや大きいと考えられるが、最後の分岐点があるグループ2、3のみのデータに対しては、決定係数が医師数 (0.43)、医療機関数 (0.59) であり、このことが影響し分岐点とはならなかったと考えられる。最初に2つに分岐した段階では宮崎県を除き、一人当たり医療費の高低の2グループに分割されている。さらに分岐した先では、全体での順位とグループの順位を比較すると、入れ子になっている部分が生じ、完全なレンジ分けにはなっていないが、箱ひげ図の「箱」の部分はグループごとに異なる範囲に収まっており、先述の「交差妥当化」処理の結果、適切な回帰木の枝の剪定が出来たものと考えられる。

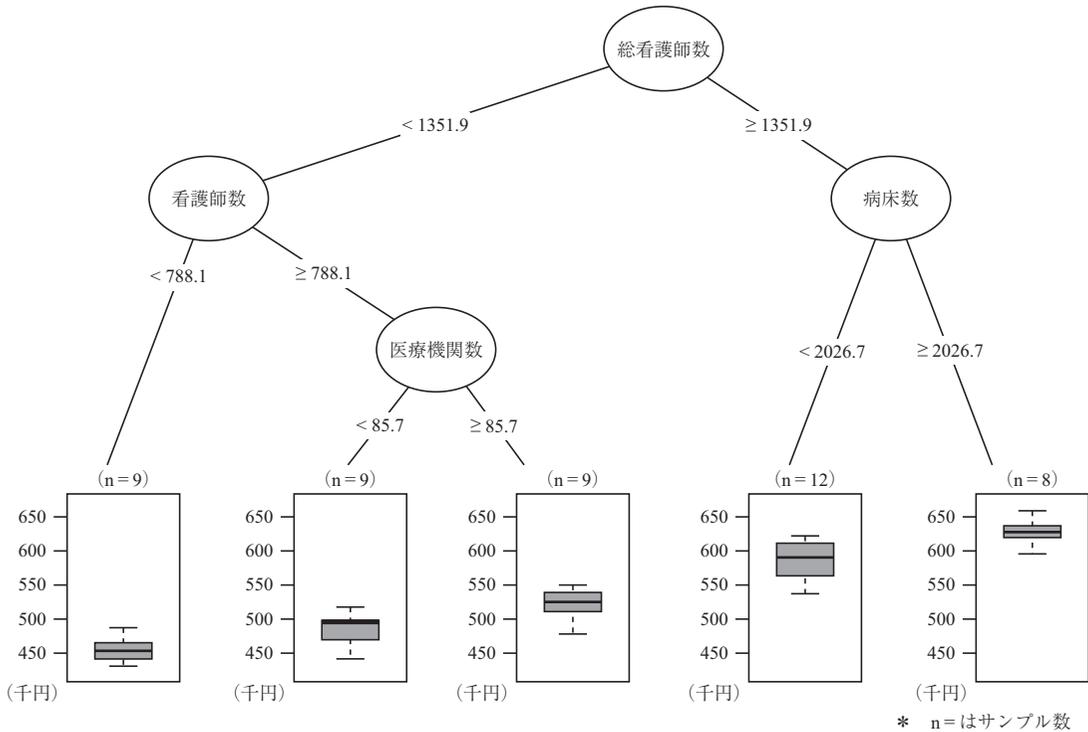


図2 1人当たり医療費の回帰木

### 3 グループごとの比較

表2は、回帰木により5個に分割されたグループを1人当たり医療費の低い順に1~5とし、グループ間の各変数の平均値をまとめたものである。グループ間の最下位と最上位の比率は1.37倍であり、グループ間の差は、回帰木の最初の分岐点である上位2つと下位3つの間で大きくなっている。グループ別の都道府県を見ると、秋田、北海道を除く関東、東北の各地域がグループ1および2となり、グループ5は四国、九州の各地域となり、いわゆる「西高東低」と呼ばれる傾向が当てはまっている。

人口10万あたりの従事者数は、総看護師数と保健師数がグループ2と3の間で逆転しているが、そのほかは医療費の高い地域ほど大きい値となった。

これとは逆に医療施設あたりの供給体制の手厚さを示す総看護師に対する看護師の比率および病

床当たりの各種人数は、医療費の高いグループ4、5で低い傾向があるがグループ1~3の間では、順位の比例・反比例の明確な関係はなかった。

平均在院日数は、市町村国保のグループ2と3が逆転している以外は、1人当たり医療費が高い地域ほど長いものとなっている。病床機能報告における病床種別を見ると、配置人員の低いグループ4、5では慢性期病床の比率が高くなっている。また、グループ1と2、3を比較すると病床当たりの看護師数は同程度であるがグループ1の医師数が多くなっている。また、グループ1の高度急性期の比率が5%高いが、逆に急性期病床が3~4%低くなっている。

医療費の3要素についてみると、受診率は「外来+調剤・後期高齢」以外は、医療費の高い地域ほど高く、差も大きなものとなった。1件当たり日数はグループ1~3では順位の逆転が生じているが、グループ4、5は順当に長くなっている。1日当

表2 グループ別平均値

グループ	1	2	3	4	5	
一人当たり医療費（千円）	456.77	487.01	522.09	585.63	627.02	
該当都道府県	宮城県 茨城県 栃木県 埼玉県 千葉県 東京都 神奈川県 静岡県 愛知県	青森県 岩手県 福島県 群馬県 新潟県 長野県 岐阜県 滋賀県 沖縄県	秋田県 山形県 山梨県 三重県 京都府 大阪府 兵庫県 奈良県 和歌山県	北海道 富山県 石川県 福井県 鳥取県 島根県 岡山県 広島県 香川県 愛媛県 福岡県 宮崎県	山口県 徳島県 高知県 佐賀県 長崎県 熊本県 大分県 鹿児島県	
分岐基準	総看護師数（人口10万当たり） 看護師数（人口10万当たり） 病床数（人口10万当たり） 医療機関数（人口10万当たり）	909.6 701.8 1034.5 74.8	1205.4 898.7 1318.2 77.9	1159.3 900.3 1360.8 95.7	1479.6 1084.2 1726.7 92.4	1732.6 1183.0 2259.2 100.7
医療提供体制	医師数（人口10万当たり） 保健師数（人口10万当たり） 総看護師数に対する看護師の比率 病床あたり医師数 病床あたり総看護師数 病床あたり看護師数 平均在院日数（市町村国保＋後期高齢） 平均在院日数（市町村国保） 平均在院日数（後期高齢）	214.4 33.6 0.78 0.21 0.88 0.68 34.0 29.1 38.0	215.7 50.8 0.75 0.17 0.92 0.69 38.0 34.4 40.5	251.9 42.9 0.78 0.19 0.85 0.66 38.1 33.2 41.3	273.0 51.5 0.73 0.16 0.86 0.63 42.9 37.6 46.1	284.6 54.5 0.68 0.13 0.77 0.53 50.0 44.7 52.9
病床報告	高度急性期（％） 急性期（％） 回復期（％） 慢性期（％）	0.19 0.48 0.08 0.25	0.14 0.52 0.09 0.25	0.14 0.51 0.09 0.27	0.15 0.44 0.09 0.31	0.09 0.42 0.12 0.37
医療費の三要素	受診率（入院・市町村国保） 受診率（外来＋調剤・市町村国保） 受診率（入院・後期高齢） 受診率（外来＋調剤・後期高齢） 1件当たり日数（入院・市町村国保） 1件当たり日数（外来＋調剤・市町村国保） 1件当たり日数（入院・後期高齢） 1件当たり日数（外来＋調剤・後期高齢） 1日当たり医療費（入院・市町村国保） 1日当たり医療費（外来＋調剤・市町村国保） 1日当たり医療費（入院・後期高齢） 1日当たり医療費（外来＋調剤・後期高齢）	0.19 8.08 0.69 15.81 15.09 1.59 17.12 1.83 37163 13400 32524 14530	0.23 8.11 0.76 15.19 16.34 1.56 17.59 1.75 33557 13707 30950 14815	0.23 8.62 0.78 16.13 16.08 1.62 17.76 1.91 34962 13059 31475 13804	0.29 8.56 0.95 15.45 17.03 1.65 18.54 1.97 32286 13367 28935 14018	0.33 8.81 1.13 16.27 18.33 1.70 19.50 2.05 28809 12736 26694 12813
介護関連	介護サービス受給者総数（人口10万当たり） 介護施設サービス受給者総数（人口10万当たり） 介護福祉施設サービス受給者総数（人口10万当たり） 介護老人福祉施設定員（人口10万当たり）	3231.7 833.0 434.3 354.6	4161.0 1111.8 562.6 449.7	4335.0 1099.0 586.0 471.6	4395.0 1220.2 602.8 480.7	4563.7 1286.2 578.6 457.8
特定	特定健康診査受診率 特定保健指導実施率 メタボリックシンドローム該当者割合	0.52 0.17 0.15	0.49 0.22 0.15	0.47 0.19 0.14	0.46 0.22 0.15	0.45 0.25 0.15
社会・経済等	老人クラブ会員加入率（対老年人口） 民生委員相談件数（健康・保健医療：人口10万当たり） 在宅死亡率 65歳以上人口 1人当たり県民所得（千円） 高齢者就業割合（65歳以上）対老年人口	18.4 420 13.88 0.25 3240.0 21.21	20.4 312 12.33 0.27 2729.1 20.52	19.1 245 13.50 0.28 2801.5 19.94	19.1 266 10.77 0.28 2736.8 20.05	9.7 171 9.30 0.29 2598.9 19.03

たり医療費は、ほかに比べグループ間での順位の逆転が多いが、医療費の低い方の3グループが高くなっている。

介護サービスについては、受給者総数および施設サービス受給者総数は医療費の高い地域ほど多いが、福祉施設サービス受給者数と老人福祉施設定員は、グループ1~4では順当に多くなっているが、グループ5だけは中程度と低くなっている。

特定健診では、受診率が高いほど医療費が低くなっているが、保健指導実施率は逆の傾向がある。メタボリックシンドロームの該当者割合はグループ間の差は特になかった。

ソーシャルキャピタルを示す変数では、老人クラブ会員加入率ではグループ5だけが特に低く、民生委員相談件数が多いほど医療費が低くなっている。

社会・経済関連指標では、在宅死亡率が高いほど、65歳以上人口が低く高齢者就業割合が高いほど医療費が低いものとなり、1人当たり県民所得はグループ2と3が逆転している以外は高いほど医療費は低いものとなった。

また、65歳以上人口は、グループ1が低いが、ほかのグループ間は同水準となった。1人当たり県民所得は、グループ1と5が、1人当たり医療費とは逆に高低が目立つ結果となった。

#### IV 考察

これまでの分析により、以下の傾向が確認できた。図1では、総看護師数の影響が大きいものとなった。先行研究では、医療の供給体制の指標として病床数、医師数が用いられてきたが、診療報酬の算定において高い看護基準が新たに設けられたことが本稿での分析結果に影響していると考えられる。

図2の回帰木による分析において、最初に総看護師数で分岐した2つのグループにおいて、医療費の高い方では病床数、低い方では看護師数が医療費に与える影響が大きくなることを示唆している。なお、この3つの変数は相互に高い相関があるが、総看護師数のみを用いて回帰木分析を実行

したところ、2つに分岐した下位の分岐は医療費の水準によるグループ分けは適切に行えず、2つのグループでは、医療費の決定要因が異なるものであると考えられる。

表2においては、病床あたりの医師数、看護師数等は医療費の高い地域ほど小さい値となった。このことは医療費の高い地域ほど施設は多いが配置人員は手薄になっていることを示している。

平均在院日数は、これに対応し医療費の高い地域ほど長くなる傾向が確認できた。また、病床報告の数値を見ると、医療費の高い2つのグループは、回復期、慢性期の比率が高くなっている。ほかの3つのグループで比較すると回復期、慢性期は、ほぼ同水準であり、高度急性期と急性期の組み合わせで見ると、医療費の最も低いグループ1だけが高度急性期の比率が高くなっており、平均在院日数もグループ1が、ほかの2つのグループよりも短いものとなっている。

医療費の3要素についてみると、受診率のうち、「外来+調剤」は患者の自発的意思によるものと考えられるが、医療機関が多いことにより増大するというアクセスコスト低下説もあり、医療機関数はグループ1~5へと多くなっているが、受診率も順に高いものとなっている。「入院」については供給者からの誘発需要が生じる可能性が高いと考えられ、グループ1と5の差も大きなものとなっている。逆に、1日当たり医療費については診療密度の高いグループほど高い傾向がある。

介護サービスでは、介護福祉サービスの受給者数と介護福祉施設定員が全体の傾向とは別に、グループ5が低くなっている。介護サービスの提供は訪問・通所でも行われており、介護サービスの提供体制の状況については、本稿の検討対象外であるが、平均在院日数や病床報告に慢性期の比率が高いことを考慮すると、ほかの地域では介護施設で対応しているケースを医療施設で対応している状況が生じている可能性が考えられる。

これらの点を医療機関の経営戦略という点からみると、病床数を増加させることは現行制度上、困難であり、収益を向上させるためには看護配置を高くするということが一つの選択となる。医療

費の低い3グループでは、人口10万人あたりの総看護師数等は低い、病床あたりの配置人員が高く、また高い看護配置基準を適用するための必要から看護師の総看護師に対する比率が高くなっている理由であると推察される。また、この3グループは高度急性期と急性期の合計は、ほぼ同一であるが、グループ1が病床当たりの医師数が多く高度急性期の病床も多くなっている。医療費の三要素からみてもでは、入院については、1日あたりの医療費が高いが受診率、1件当たり日数は低くなっている。他方、医療費の高い2つのグループでは、総看護師数が多いが看護師の比率は低く、医療費の高い地域のほうが病床数も多く、療養系の病床が選択されていることが推察される。

予防については、特定健診受診率、民生委員相談件数は高いほど医療費は低く、これらの施策が一定の効果があると考えられる。生活習慣という点では、就業や老人クラブへの参加など、何らかの活動を行っていることが医療費を抑制する効果があると考えられる。

先に述べたように、本稿が対象とした2014年は、7対1入院基本料届出病床数がピークであったが、これは2006年（平成18年）4月に病床の機能分化を進める政策の一つとして創設されたものであるが、ほかの入院基本料と比較してかなり高いものであったため、当初は約4.5万床であったものが急速に増え、本稿が対象とした2014年には38万床を超え、その後は若干減少し2017年には36万床を割り込んでいる<sup>9)</sup>。Noguchi（2015）では、この変更の効果が政策当局の想定範囲を超え、医療機関が急速に高い看護配置の病床を増床したことを指摘している。

この入院基本料が導入される以前のデータを用い本稿と同様な分析をしたものとして、「平成19

年版 厚生労働白書<sup>10)</sup>があり、クラスター分析による47都道府県を10のグループに分け1人当たり老人医療費について検証している（使用データは2000～2005年）。この分析結果と比較を行ったところ、医療費の高い順の6グループ（21府県）は、大阪、京都、沖縄の3府県を除き本稿での医療費の高いグループ4、5に該当した。医療費の低い順の3グループ（22都県）は本稿での医療費の低いグループ1～3に該当し、3に該当した県は、全体の病床数が減少する中で、その数が横ばいになっている傾向がみられた。

2006年（平成18年）について、本稿で用いた上位5指標のみの簡略した分析をおこなったところ、同様に総看護師数で2分（上位は18県）され、次に医療機関数、医師数で分岐した<sup>11)</sup>。

1人あたり医療費に対して、単回帰係数を2006年と比較すると総看護師数、看護師数、病床数は係数の値が大きくなり、医師数は小さく、医療機関数は同じであった<sup>12)</sup>。

これらのことから、本稿における看護師総数で2分割される上位約20府県が医療費の高い傾向にあることは従前からの傾向であったと考えられ、看護基準の変更により影響度が大きくなったものと考えられる。

2つの分析について医師数、医療機関数の影響は目立たない結果となり、供給者誘発需要の観点からすればやや以外なものであるが、前述の看護基準の変更、過去におけるバブル期、病床規制前の駆け込み増床等、医療機関が経営戦略として行った行動からすれば、病床数、看護師数の医療費に与える影響が大きいことも妥当性があると考えられる。看護師については、以前から地域的偏在〔大石他（2001）〕と、その背景として、看護師養成機関の地域的偏在〔大石（2009）〕が指摘され

<sup>9)</sup> 厚生労働省（2017）p.14。

<sup>10)</sup> この分析は、みずほ情報総研(株)が2007年に行った「保健医療をめぐる現状とその要因の地域間格差に関する統計的分析」によるものである。

<sup>11)</sup> 1人あたり医療費が高い順に、①総看護師数 $\geq$ 1166.5、②総看護師数 $<$ 1166.5&医療機関数 $\geq$ 68.25&医師数 $\geq$ 211.25③総看護師数 $<$ 1166.5&医療機関数 $\geq$ 68.25&医師数 $<$ 211.25、④総看護師数 $<$ 1166.5&医療機関数 $<$ 68.25の4グループとなった。

<sup>12)</sup> 2006年から2014年の変化は、総看護師数（0.80 $\rightarrow$ 0.83）、看護師数（0.81 $\rightarrow$ 0.86）、病床数（0.81 $\rightarrow$ 0.84）、医師数（0.70 $\rightarrow$ 0.67）、医療機関数（0.657 $\rightarrow$ 0.659）となった。

ている<sup>13)</sup>。医師に比べれば地域間での人材の移動性が低いと考えられ、医療提供体制のなかでは調整が進みにくい部分であるとも推察される。

## V 結語

ランダムフォレストによる分析では、医療の提供体制が医療費の決定要因として重要度が高いことが示唆された。1人当たり医療費を目的変数とした回帰木によるグループ分けでは、グループごとの各変数の平均値が、比較的明瞭に1人当たり医療費に対し比例・反比例の傾向を示している。また、相互に相関が高い、医師数、病床数、看護師数を同時に取り扱ったことにより、人口当たりの医療提供体制が小さく、医療施設当たりの人員の配置密度が高いほど医療費が低いことが確認できた。

現在、「行政情報の見える化」、「エビデンスに基づく政策」といった観点から各種行政データが整備され、積極的な利用が進められている。これらを用いたランキング情報や機械的な等分割によるグループ間の比較よりは一歩進んだ政策の目標となる情報を提供する手段として、本稿で用いたランダムフォレストおよび回帰木の有用性は非常に大きいと考えられる。

しかしながら、本稿の分析で用いたには単年度のデータであり、グループ間の平均値の比較であるため、因果関係の特定など一般化しうるものではないので、分析方法の改善とともに、さらなる検証を重ねていく必要があると考えられる。

(平成29年12月投稿受理)

(平成30年7月採用決定)

## 参考文献

- 青木 研 (2001)「患者の行動による地域差」地域差研究会 (編)『医療費の地域差』東洋経済新報社。  
 足立泰美・赤井伸郎・植松利夫 (2012)「保健行政における医療費削減効果」、『季刊社会保障研究』, Vol.48 (3), pp.334-348。  
 井伊雅子・別所俊一郎 (2006)「医療の基礎的実証分析

- と政策：サーベイ」『フィナンシャルレビュー』, 2006年 (1), pp.117-156。  
 池田俊也・望月真弓 (2001)「薬剤費の地域差」地域差研究会 (編)『医療費の地域差』東洋経済新報社。  
 泉田信行 (2001)「医療機関の行動による地域差」地域差研究会 (編)『医療費の地域差』東洋経済新報社。  
 印南一路 (2016)「医療費と医療政策に関わる諸問題」政策シンクネット  
<http://thinknet.org/theme03/2016053016.html> (2017年11月27日最終確認)。  
 大石杉乃・内藤雅子・根岸龍雄 (2001)「看護従事者に地域差が生じた要因」『民族衛生』, Vol.67 (5), pp.206-218。  
 大石杉乃 (2009)「1974年から2006年における看護師・准看護師養成状況の地域差と経年的変化」『東京慈恵会医科大学雑誌』, Vol.124 (5), pp.201-212。  
 金明 哲 (2007)『Rによるデータサイエンス』森北出版株式会社。  
 郡司篤晃 (2001)「医療費の地域差研究とその意義」地域差研究会 (編)『医療費の地域差』東洋経済新報社。  
 厚生労働省 (2007)「平成19年版厚生労働白書」。  
 ——— (2014)「第1回看護職員受給見直しに関する検討会 平成26年12月1日 資料3-1」,  
<http://www.mhlw.go.jp/file/05-Shingikai-10801000-Iseikyoku-Soumuka/0000072895.pdf> (2018年5月25日最終確認)。  
 ——— (2017)「中医協・診療報酬調査専門組織 (入院医療等の調査・評価分科会) 平成29年6月21日 資料 (入-1)」,  
<http://www.mhlw.go.jp/file/05-Shingikai-12404000-Hokenkyoku-Iryouka/0000168356.pdf> (2018年5月25日最終確認)。  
 下川敏雄・杉本知之・後藤昌司 (2013)『樹木構造接近法』共立出版。  
 豊田秀樹 (編著) (2008)『データマイニング入門』東京図書。  
 中西悟志 (2001)「医療施設の競争と医療費の地域間格差」地域差研究会 (編)『医療費の地域差』東洋経済新報社。  
 橋本英樹・泉田信行 (編) (2011) (編著)『医療経済学講義』東京大学出版会。  
 府川哲夫 (2006)「少子高齢化と医療費をめぐる論点と実証研究」西村周三・田中滋・遠藤久夫 (編著)『医療経済学の基礎理論と論点』勁草書房。  
 前田信雄 (1983)「高齢者の入院医療費の高低に関する研究 - 地域差の分析 -」, 『季刊社会保障研究』, Vol.19 (1), pp.56-69。  
 罇 淳子 (2013)「老人医療費の都道府県格差に及ぼす要因の検討 - 老人医療費の多寡によるグループ分け

<sup>13)</sup> 厚労省 (2014) p.11, p.13人口当たり看護師数の多い地域ほど、養成機関の卒業生が多いため県内への就業割合が低い傾向が確認できる。

からみた分析-」, 『新潟青陵学会誌』, Vol.6, No1,  
pp.1-11。  
Noguchi, Haruko (2015) “How does the Price Regulation  
Policy Impact on Patient-Nurse Ratios and the Length

of Hospital Stays in Japanese Hospitals?”, *Asian  
Economic Policy Review*, Vol.10 (2), pp.301-323。

(すだ・しげお)

## **Analysis of Prefectural Medical Expenses Using a Machine Learning Approach**

Shigeo SUDA \*

### Abstract

In this research, Random Forest and Regression tree which are the techniques of machine learning analyzed using the index of 26, and the level of medical expenses performed the group division per person.

As a result, it divided into five groups on the basis of the number of nurses, the number of sickbeds, and the number of medical institutions which are medical care provision system.

With comparing the average value by each group, the higher the medical expenses is, the more the number of doctors and the number of nurses are relatively more, but because the number of beds is more than that, the number of persons placed per bed is narrow, It was confirmed that the number of days in the hospital was longer.

Moreover, although there was a tendency with many long-term-care-service recipients per person as the high area of medical expenses, in the group with the highest medical expenses, the Home-based care support services capacity and the number of service recipients were degrees in the middle.

In addition, it was confirmed that medical expenses are low per person, if Specified Health Checkups consultation rate and home mortality rate were high.

Keywords : Regional Difference of Medical Expenses, Nurse, Medical Supply Organization, Random Forest, Regression Tree

---

\* Member, Japan Economic Policy Association, The Japan Public Choice Society