

# 都道府県別に見た平均寿命に関する統計的分析

2012SE202 大野慎也

指導教員：木村美善

## 1 はじめに

平均寿命は戦後から著しく伸びているが、その理由や長寿の国となった秘訣はどこにあるのか探してみたいと思い、本研究のテーマにした。そして男女の共通点と相違点を明らかにするために男女別に平均寿命の決定要因について分析を行うことが本研究の目的である。

## 2 データについて

国立社会保障・人口問題研究所の2014年度の表12-36、表12-37([2]参照)、政府統計の総合窓口の平成22年度都道府県別生命表([3]参照)、都道府県別統計とランキングで見る県民生([4]参照)、都道府県別健康寿命の要約([5]参照)のデータを用いた。

県のデータ番号は北から順に1:北海道, 2:青森, ..., 47:沖縄とした。また、本研究における平均寿命とは0歳児における平均余命のことをいう。

## 3 変数について

男性の場合は  $x_1$ :平均寿命,  $x_2$ :人口,  $x_3$ :完全失業者率,  $x_4$ :平均給料,  $x_5$ :野菜摂取量,  $x_6$ :食塩の摂取量,  $x_7$ :年間雪日数,  $x_8$ :日照時間,  $x_9$ :高齢者数,  $x_{10}$ :歩数,  $x_{11}$ :喫煙率,  $x_{12}$ :飲酒率,  $x_{13}$ :1人あたりの県民所得,  $x_{14}$ :悪性新生物,  $x_{15}$ :心疾患,  $x_{16}$ :脳血管疾患,  $x_{17}$ :肺炎,  $x_{18}$ :生涯未婚率,  $x_{19}$ :初婚年齢,  $x_{20}$ :肥満率とする。

女性の場合は  $x_1$  から  $x_{10}$  までは男性と同じとし,  $x_{11}$ :1人あたりの県民所得,  $x_{12}$ :悪性新生物,  $x_{13}$ :心疾患,  $x_{14}$ :脳血管疾患,  $x_{15}$ :肺炎,  $x_{16}$ :生涯未婚率,  $x_{17}$ :自殺者数,  $x_{18}$ :初婚年齢,  $x_{19}$ :1人あたりの酒消費量,  $x_{20}$ :喫煙率とする。

## 4 重回帰分析

### 4.1 男性の平均寿命の分析

目的変数  $y$  を平均寿命  $x_1$ , 説明変数を  $x_2, \dots, x_{20}$  として重回帰分析を行った。VIFと相関行列の固有値から多重共線性が見られた。ステップワイズ法により変数選択を行った結果、VIFと相関行列の固有値の値から多重共線性の疑いはなく、決定係数は0.417、調整済み決定係数は0.338となり、男子よりは十分に説明できていないと感じた。LMS推定量を用いて残差分析をした結果、外れ値が存在した。外れ値である5番(秋田県)は他県に比べ人口と日照時間が低めであり、28番(兵庫県)は他県に比べ日照時間がやや長いところから外れ値になったと考えられる。

平均寿命に関しては、野菜を多く摂取する、初婚年齢が遅いほうが男性の平均寿命は延びることがデータから読み取れる。野菜の摂取量に関しては、常々野菜は健康に良い、

表1 男性の回帰分析結果

変数	係数	t 値	p 値
定数項	75.497	17.99	$2 \times 10^{-16}$
$x_3$	-7.8510	-1.520	0.137390
$x_5$	0.0114	3.964	0.000346
$x_6$	-0.4654	-4.340	0.000115
$x_{10}$	0.0003	2.675	0.011293
$x_{13}$	-0.0005	-1.335	0.190468
$x_{14}$	-0.9677	-3.166	0.000319
$x_{15}$	-1.0093	-3.238	0.002635
$x_{16}$	-1.4005	-2.444	0.019713
$x_{17}$	-1.1389	-1.741	0.090539
$x_{18}$	-6.2777	-1.887	0.067508
$x_{19}$	0.4651	3.173	0.003133

体に良いといわれるところからも正の相関になることがわかった。初婚年齢に関しては、幸福度の観点から遅いほうが生活面で安定していて正の相関になると考えられる。また、負の相関については、食塩の摂取量、心疾患が大きな影響を及ぼしている。食塩の摂取量が増えると心疾患や他の病気につながることから平均寿命が下がるので負の相関になることも納得がいく。

### 4.2 女性の平均寿命の分析

目的変数  $y$  を平均寿命  $x_1$ , 説明変数を  $x_2, \dots, x_{20}$  として重回帰分析を行った。VIFと相関行列の固有値から多重共線性が見られた。ステップワイズ法により変数選択を行った結果、VIFと相関行列の固有値の値から多重共線性の疑いはなく、決定係数は0.417、調整済み決定係数は0.338となり、男子よりは十分に説明できていないと感じた。LMS推定量を用いて残差分析をした結果、外れ値が存在した。外れ値である5番(秋田県)は他県に比べ人口と日照時間が低くなったので外れ値になったと考えられる。

表2 女性の回帰分析結果

変数	係数	t 値	p 値
定数項	89.14	70.41	$2 \times 10^{-16}$
$x_3$	-16.62	-2.472	0.01781
$x_4$	0.0009	0.490	0.62769
$x_5$	0.0084	2.987	0.00479
$x_8$	-0.0007	-2.079	0.04213
$x_{13}$	-0.8676	-2.248	0.03013
$x_{14}$	-2.018	-3.378	0.00164

平均寿命に関しては男性と同様に女性も野菜を多く摂取するほうが平均寿命が延びることがデータから読み取れる。野菜の摂取量に関しては、女性の場合も男性と同様な理由から正の相関になると考えられる。また、負の相関については、心疾患、脳血管疾患が大きな影響を及ぼしており、これらの病気が増えるほど平均寿命が下がっていく傾向にある。さらに、女性は日照時間や完全失業率が負に影響を及ぼしている。女性は日焼け止めをして、肌を守ることから日照時間が多いほど体に害を及ぼすので、データから負の相関になることにも納得がいく。完全失業率が平均寿命に大きくマイナスの影響を及ぼすところが男性との違いである。このことから、意外にも男性よりも女性の方が仕事の職を失うことが痛手になっているのではないかと思った。また、決定係数の値からみると男性の場合ほどよくは説明ができていないと思った。

## 5 加重回帰分析

通常の重回帰分析では都道府県の人口によるばらつきの大さきの違いの影響が考えられていないので、加重回帰分析を行う。([1] 参照)

### 5.1 男性の平均寿命の分析

通常の重回帰分析と同様の変数が平均寿命に大きく影響を与えるという点では変わりがなかった。また、決定係数は 0.819 から 0.823 とあまり変化はなかった。しかし、選ばれた変数や決定係数がわずかであるが変化した。また、外れ値が 4 つになり、いずれも他県より人口が少ない県であるということからも都道府県の人口によるばらつきがあったと考えられるので、加重回帰分析の方がしっかりと分析ができていないといえる。

### 5.2 女性の平均寿命の分析

男子同様、決定係数の値からも通常の重回帰分析とそれほど大きな違いはみられなかった。しかし、外れ値に変化があり、13 番の東京が入った。東京は人口が他県に比べると爆発的に多いところである。このことから都道府県の人口によるばらつきの大さが影響していると考えられる。

## 6 主成分分析

変数間の関係を見るために全変数を用いて主成分分析を行い、平均寿命に関する考察を行う。男女いずれも第 6 主成分までの累積寄与率がそれぞれ 78 %、80 % であるから第 6 主成分までの解釈を行った。男性についての各主成分の解釈としては、給料が多ければ多いほど平均寿命は伸びると考えられ、病気が関与すればするほど平均寿命に対して負となった。女性についても平均寿命に対して負となる影響は同様なことが言える。主成分分析を通して、男女いずれも平均寿命に対して病気の発生および生涯未婚率、完全失業率といった病気の発生する要因が平均寿命に対して負となることがわかったので、これらの病気の要因に注意し、病気を減らすことが平均寿命をあげるために必要であ

ることがわかる。

## 7 クラスタ分析

第 6 主成分までの主成分得点を用いて ward 法で、クラスタ分析を行った。男性の結果は、第 1 群では、東北地方や北陸地方の都道府県、長野といった県など年間雪日数が多い群であると考えられる。第 2 群ではの都道府県は東京、千葉、埼玉、愛知を含む首都圏であるので、人口のばらつきが大きい群と考えられる。第 3 群では、四国地方や兵庫、奈良など高齢者の人口のばらつきが多い県だと考えられる。女性についても同様の分析を行ったが、分析結果は人口に関する群、雪日数が多い群と少ない群とに分けられた。

## 8 まとめ

今回重回帰分析を男性の場合と女性の場合の 2 種類で行ったが、決定係数が高いことから男子の平均寿命に関してはうまく説明できていると感じている。女子に関しては決定係数の値からも男子ほどうまく平均寿命を説明できていない気がする。また、外れ値を考えると加重回帰分析を行った分析が一番うまく説明できている。主成分分析では、平均寿命に対して負の相関を及ぼすのは何かを説明できたように思う。クラスタ分析では似ている都道府県同士をうまく分けることができたと感じている。

## 9 おわりに

私は本研究をするにあたり、平均寿命に関しては高齢者の人口が多いと平均寿命は上がると思っていた。しかし、研究を通じて、平均寿命と高齢者の人口とは関係がないということがわかり、関係があるのは食文化に関係のあるものであった。共通要因としては食文化が共通しており、その他には共通するものがなかった。研究前には気づかなかった要因に気づけてよかったと思っている。本研究の結果を参考にして、これからの人生においてどうしたら更に長生きできるかを考えたい。

## 参考文献

- [1] Chatterjee, S. and B. Price : *Regression Analysis by Example*, John Wiley & Sons, New York, 1977. (加納 悟・佐和隆光訳: 回帰分析の実際, 新曜社, 1981. )
- [2] 国立社会保障・人口問題研究所, 2014 年度人工統計集 <http://www.stat.go.jp>
- [3] 政府統計の総合窓口平成 22 年度都道府県別生命表 <http://www.ipss.go.jp/>
- [4] 都道府県別統計とランキングで見る県民性 <http://todo-ran.com/>
- [5] 都道府県別健康寿命の要約 <http://human.cc.hirosaki-u.ac.jp/>