

# 都道府県別の新型コロナウイルス感染者数に関する統計的分析

2017SS072 田島海渡

指導教員：白石高章

## 1 はじめに

今年に入り、毎日のように新型コロナウイルスに関するニュースが報道されている。新型コロナウイルスの感染防止策として、私自身も大学での講義や就職活動、アルバイト先などの場で大きな影響を受けている。このようなニュースでの報道や自分が過ごしている環境から新型コロナウイルスに関心を持った。そこで、全国の新型コロナウイルス感染者数と様々なデータとの関係を分析することによって、新型コロナウイルス感染者数が多い県にはどのような特徴があるのか考察を行う。

## 2 データと分析方法について

データは、令和2年1月から11月における新型コロナウイルス感染者数 ([1]) と気象庁のデータなどの都道府県ごとの様々なデータから引用した。各変数は、 $x_1$ (新型コロナウイルス感染者数)、 $x_2$ (退院者数)、 $x_3$ (死亡者数)、 $x_4$ (PCR検査人数)、 $x_5$ (生活習慣病による死亡者数)、 $x_6$ (がん死亡率)、 $x_7$ (GDP)、 $x_8$ (血液型 A 型の割合)、 $x_9$ (血液型 O 型の割合)、 $x_{10}$ (血液型 B 型の割合)、 $x_{11}$ (血液型 AB 型の割合)、 $x_{12}$ (タバコの消費量)、 $x_{13}$ (お酒の消費量)、 $x_{14}$ (スポーツ施設の数)、 $x_{15}$ (公共交通機関での通勤・通学率)、 $x_{16}$ (自家用車での通勤・通学率)、 $x_{17}$ (平均気温)、 $x_{18}$ (降水量)、 $x_{19}$ (平均湿度)、 $x_{20}$ (人口密度) としている。また、分析で扱うデータは人口 10 万人あたりのデータを使用する。

分析方法は、相関係数、クラスター分析、因子分析を用いて行った。人口密度については順位相関係数も求めた。分析は主に金 [2] を参考に行った。

## 3 相関係数

新型コロナウイルスと様々なデータとの相関係数を調べ、新型コロナウイルス感染者数との関連性について調べる。ただし、各データに対して無相関検定を行い  $P$  値 が棄却されたデータに対して考察を行う。

### 3.1 相関係数での考察

$x_1$  と負の相関が高かった項目に  $x_5$  があった。このことから、生活習慣病にかかる人は新型コロナウイルスに対する免疫力が高いのではないかと考えられる。感染者数と正の相関が見られたのは、 $x_7$  であった。経済力が高い地域ほど感染が広まりやすいということが分かる。また、 $x_{15}$  が 0.663 と強い正の相関、 $x_{16}$  が  $-0.688$  と強い負の相関を示したことから人が集まりやすい場で感染するということや  $x_{14}$  からは運動することで感染に対する免疫力が上がるのではないかと考えられる。

### 3.2 順位相関係数での考察

$x_1$  と  $x_{20}$  の相関を調べたところ値は 0.733、順位相関係数は 0.674 であった。どちらの相関係数を見ても 0.5 以上の正の相関があることから人が集まる地域ほど感染リスクが高まるということが分かった。

## 4 クラスタ分析

データを標準化してクラスター分析をウォード法を用いて行った。

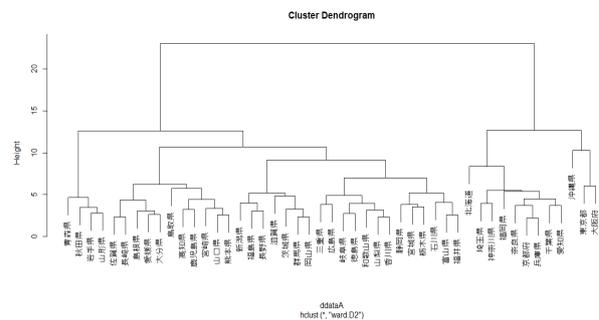


図 1 クラスタ分析

### 4.1 群の特徴

- 第 1 群: 感染者数が最も少ない都道府県
- 第 2 群: 人口密度が低い都道府県
- 第 3 群: 健康に対する意識が高い都道府県
- 第 4 群: 地域ごとの感染対策が必要な都道府県
- 第 5 群: 感染者数が最も多い都道府県

### 4.2 考察

第 1 群: (青森県, 秋田県など)

この群の特徴は、感染者数が最も少ないことが挙げられる。含まれている都道府県は、全て東北地方に属する県であり公共交通機関での通勤・通学率も最も低い。このため通勤通学による人の集まりが少なく感染リスクも低いと考えられる。

第 2 群: (佐賀県, 長崎県など)

この群に含まれる県は全て西日本に属している。この群の主な特徴は人口密度の項目がほかの群と比べると低いという点が挙げられる。このことから、感染者数も少ないと考えられる。一方で PCR 検査人数も少ないため検査を増やすことで陽性者数が増える恐れもあると考えられる。

第 3 群: (新潟県, 福島県など)

この群の主な特徴は、がん死亡率、お酒の消費量、タバコの消費量が低いということであり、健康志向の人たちが住む地域であると考えられる。また、第 3 群は大都市と隣接し

ている県も含まれているものの感染者数は群の中で3番目ということから感染対策がよくされている都道府県の集まりであると考えられる。

第4群: (北海道, 埼玉県など)

生活習慣病による死亡者数が少ないものの感染者数は2番目の多さである。個人は規則正しい生活を送っているが多くの人と接する場での感染対策を強める必要があるのではないかと考えられる。

第5群: (沖縄県, 東京都など)

この群は, 感染者数が最も多い。他にも, 公共交通機関での通勤・通学率, 人口密度も一番多いという特徴がある。また, この群に含まれる都道府県は観光地としても栄えており多くの人と接する機会があると考えられるため感染リスクも高くなると考えられる。

## 5 因子分析

分析結果や  $P$  値 から因子数を5と設定し, 因子軸の回転方法はバリマックス回転で行った。分析結果から血液型に関する変数は除いて分析を行った。因子負荷量を求めた結果が表1である。次にデータを分析した結果を示す。

表1 因子負荷量

変数	Factor1	Factor2	Factor3	Factor4	Factor5
$x_1$	0.433	0.853		0.171	0.224
$x_2$	0.426	0.856		0.156	0.231
$x_3$	0.233	0.816		0.171	
$x_4$	0.395	0.700			0.199
$x_5$	-0.555	-0.636		0.495	-0.131
$x_6$			-0.139	0.546	
$x_7$	0.521	0.348	-0.116	0.132	0.756
$x_{12}$		0.328		-0.103	
$x_{13}$		0.225		0.714	0.299
$x_{14}$	-0.757	-0.465		0.236	-0.174
$x_{15}$	0.883	0.276	-0.189		0.209
$x_{16}$	-0.935	-0.258		-0.170	-0.160
$x_{17}$	0.204	0.291	0.905	-0.218	
$x_{18}$	-0.127		0.679		
$x_{19}$	-0.398		-0.176	0.190	
$x_{20}$	0.617	0.361		0.181	0.573

第1因子は,  $x_{15}, x_{20}, x_{16}, x_{14}$  の相関が高い。第1因子は人の集まり具合を示していると考察できる。関係が深い都道府県は神奈川県, 京都府, 大阪府であることがわかった。この3府県は人口密度が上位であることが分かる。また, 3府県は鉄道数も多く1日あたりの同じ駅での乗降客数が多い。このことが原因で人口が集中している場所での感染が発生していると考えられる。

第2因子は,  $x_1, x_2, x_3, x_4$  の相関が高い。第2因子はPCR検査人数が増えると新型コロナウイルス感染者数が増えていることを示す。また, 感染者数の多い地域ほど退院者数, 死亡者数も同様に増加していることが考えることができる。関係が深い都道府県は沖縄県, 北海道, 東京都, 大阪府である。日本の新型コロナウイルスによる死亡者数は

少ないが4都道府県の共通点として10万人あたりの死亡者数が3人から4人という点が挙げられ全国的にみると多い。この4都道府県が第2因子に含まれている理由として大都市や観光地として利用されている地域でのPCR検査の場が多く感染者数もしっかりと把握できているからだと考えられる。

第3因子は  $x_{17}, x_{18}$  の相関が高い。第3因子は天候を示している。関係が深い都道府県は沖縄県, 鹿児島県, 北海道であることがわかった。沖縄県と鹿児島県は, 平均気温も全国的に見て高く, 北海道は平均気温が一番低い。このことから, 天候による激しい温度や湿度の変化は新型コロナウイルス感染の原因の1つではないかと考えられる。

第4因子は  $x_{13}, x_6, x_5$  の相関が高い。第4因子は不健康な生活を示している。関係が深い都道府県は秋田県, 北海道, 大阪府であることがわかった。この3道府県は, お酒の消費量が多いという共通点がある。また, 秋田県は生活習慣病による死亡者数も多い。このことが原因でこれらの県は第4因子に含まれているのではないかと考えられる。

第5因子は  $x_7, x_{20}, x_{13}$  の相関が高い。第5因子は飲み屋での人の多さを示していると考察できる。関係が深い都道府県は東京都, 愛知県であることがわかった。特に関係が深い都道府県は東京都であり, 東京都の緊急事態宣言後の人の動きについて調べてみたところ夜間における飲食店の利用が多いことがわかった。これが原因で長時間多くの人々が接する機会が増えていたと考えられる。愛知県についても同様に考えられる。

## 6 考察

本研究を行った結果, 新型コロナウイルス感染者数は都道府県によって偏っているといえる。新型コロナウイルス感染者数とPCR検査人数の関係が深いことからPCR検査を受けられる地域ほど新型コロナウイルス感染者数を把握することができる一方でPCR検査人数が少ない地域ほど新型コロナウイルスの感染者数が少ない地域と判断できてしまう。そのような地域ほどPCR検査を受けていない人が新型コロナウイルスに感染している可能性があり他の地域への感染拡大の恐れがあるため47都道府県すべての地域で検査を受けられる環境を増やしていく必要がある。

## 7 おわりに

本研究を通して感染の要因は人口密度だけではなく, 都道府県ごとに様々な要因があると考察することができた。感染者の数を増やさないためにも各都道府県に応じた感染対策が必要であるように思われる。

## 参考文献

- [1] 東洋経済  
<http://toyokeizai.net/sp/visual/tko/covid19/>  
 (最終閲覧日: 2020年12月6日)
- [2] 金明哲: 「Rによるデータサイエンス データ解析の基礎から最新手法まで」。森北出版, 東京, 2007.