

インフルエンザの感染率に関する研究

2005MM073 鈴木康太

指導教員：松田眞一

1 はじめに

対策を立てることができるにも関わらず毎年流行する最も有名な感染症としてインフルエンザがあるが、その感染率について研究しようと考えた。

2 データについて

国立感染症情報センター、各県の感染症情報センターのインフルエンザ情報から感染者数を割り出し、曲線を描いた。使用データは2005年度、2006年度、2007年度の三年間分、かつ一週間ごとに分割されたものを用いた。(web[1],web[2] 参照)

3 曲線の近似式

曲線を近似するにあたり、形状からシグモイド曲線で近似することが最適だと考え、その中でもゴンペルツ曲線を近似する方法で近似した。

近似式を計算し、比較したところ曲線の後半部分で実測値との誤差が大きくなることがわかった。その点を解決するための方法論を以後進めていく。

3.1 問題点の解決法

後半部分で誤差が大きく出た理由をまず考えた。回帰計算を行う時点での二変数で散布図を出してみると以下ようになった。

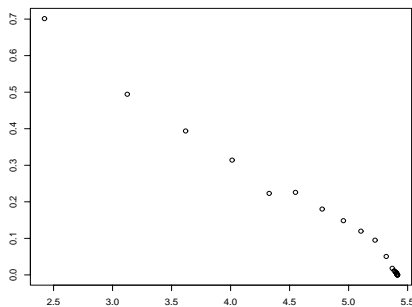


図1 関東2006年散布図

傾きが同じ一つの直線になっておらずやはり一つの式で後半部分まで正確に近似することは困難にみえた。この曲線を正確に近似するため、曲線を前半部分、後半部分に分け、それぞれについて近似式を求めた。前半部分と後半部分の分け方にはすべての都道府県で同じ基準として一番最初に散布図をみて数値が上がったところまでを前半部分としその後を後半部分とした。

3.2 前半後半に分けて近似式を出した結果

ここで扱うデータには東京、神奈川、千葉、埼玉、茨城、群馬、栃木の七つの都道府県の合算データを選んだ。

まず、関東地方を選んだ理由としては最も人口の集中している地域であり、データ量が多いと考えたためであり、合算で解析した理由は他県からの感染を含めて考えることが可能になる点、インフルエンザ感染者発生時からすぐのデータの暴れる時期を合算で扱えば振幅を抑えられる点である。さらに一般化するため開始地点も揃えるため合算するすべての都道府県の一週間の感染者数が二桁になった週からにした。

3.3 曲線の近似についてのまとめ

この結果各都道府県のデータを合算で考え、前半と後半に分けて近似式を出すことで正確に近似することができた。では実際に感染者数の予測は可能だろうか。全国で最も早く感染者の出た三、四つの都道府県の合算データを用い、それらの最終累計感染者数の人口に対する割合と全国の累計感染者数の人口に対する割合を三年分比較してみた。これが近い値の場合発生が早い都道府県のデータを参考にすればその年の全国の感染者数を推せるはずである。

4 日本全体の感染者割合について

過去三年分のそれぞれの感染者数、全人口に対する割合は以下の表である。

	累積感染者人数	全人口に対する割合
2007年	662881	0.005187
2006年	1074505	0.008407
2005年	1023100	0.008005

4.1 三年間の感染者割合について

三県の誤差平均値は2007年0.005075、2006年0.008737、2005年0.0075となった。全国と比べて違いも少ない。

4.2 この方法のまとめ

三県まで見れば全国のデータに対して高低は出るものの平均をとることで精度の高い結果を得ることができた。しかし、これらのデータではそれぞれの都道府県のインフルエンザ感染者がいなくなるまでの数を用いているため、精度という点では良いが早い時期での予想には使えない。

5 クラスタ分析

これまでの結果から傾きに最も違いが出ることが分かったが、それはすべての地域が全く独立なものなのだろうか。ここで、日本を北海道、東北、関東、中部、関西、中国、四国、九州の八ブロックに分け、傾きに関してクラスタ分析で解析した。二年分のデータを用いて、かつ、前半後半に分けたそれぞれを一つのデータにした。複数

の都道府県の合算で考えることで曲線序盤の不安定な区間を安定させ、最大でも二つの曲線の複合と考えることができるため前半と後半の二つに分けることに決めた。

一つのデータには一つの地域の前半の傾きを表す変数と後半部分の傾きを表す変数の二つを用い、八ブロックそれぞれに対し二年間分のデータを使った。

5.1 クラスタ分析の結果

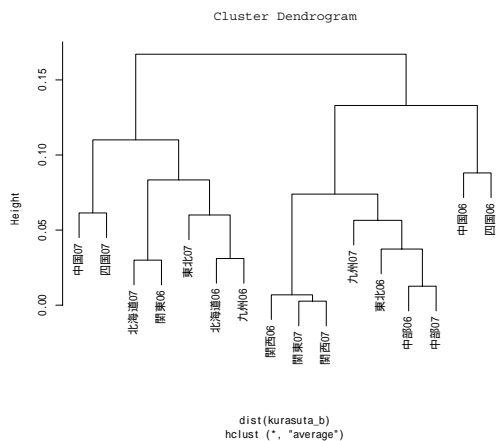


図 2 クラスタ分析の結果

図 2 を左から 4 群に分けた。

第一群 傾きが緩やかで伸び率の低いグループ。四国 2007 と中国 2007 で構成されているが四国地方と中国地方は距離も近く、同じ年では同じような伸び方を見せるのではないだろうか。

第二群 第一群ほどではないが傾きが緩やかで伸び率の低いグループ。該当するのは北海道の二年分と関東の 2006、東北 2007、九州 2006 である。このグループの特徴は北海道が二年分入っていることと関東や関西、中部等の都市圏が一つしか入っていないことであり、やはり都市圏に比べ北海道は他県からの影響が少ないため感染率は低いことがわかる。

第三群 傾きが比較的大きく感染率の高いグループ。このグループには関東、関西、中部と人口の多い地域が固まっており、都市圏では県内の感染だけでなく他県からの影響が出てくるため感染者数の伸びは大きくなることがわかる。

第四群 最も傾きが大きく感染者数の伸びが大きいグループ。該当するのは四国 2006 と中国 2006 であり、年が共通している点からやはり中国地方と四国地方は同じ年では同じような伸びを見せることが再確認できる。また四国地方と中国地方は最も伸び率の大きいグループと最も伸び率の小さいグループに独立で入ったが、特に四国に関しては大流行になった場合都市圏以上の伸びを見ることがわかった。周りを海に囲まれていることが影響しているのではないだろうか。

5.2 クラスタ分析のまとめ

まとめとして、前章の結果から 2006 年度と 2007 年度では日本全体で考えると 2007 年度のほうが全体の感染

者数は少なかったが第一群と第四群を見てわかるように四国地方と中国地方は日本全体の状況に影響されることがわかる。また、東北地方についても 2006 年度と 2007 年度の日本全体の状況と同じ結果になった。なお、関東地方の 2006 年度が第二群に入っている理由としてデータを取る段階で合算する都道府県のすべてで感染者数が二桁になった時点を開始点としているが最も人口の多い地域であるため多少の誤りが出た可能性が考えられる。次に 2006 年度と 2007 年度の日本全体の状況と比べ逆になっている九州に関して考察する。日本全体として 2006 年度は九州以外が一月前後に注意報を発令したのに対し九州だけが十一月中に注意報を出しており、他地方に比べ流行の始まりは二か月以上も早かった。そのため九州の中でのみゆるやかに感染者数が増え日本全体で大流行している時期には既に感染した人口が多かったため増加の割合が日本全体と逆になったと考えられる。2007 度に関しては 2006 年と逆に九州地方が日本全体でもっとも遅く九州以外全域で広がってから一か月程度遅れていた。日本で最後に、特に周りに比べ短い時間で感染者が増えたため傾きが急になり日本全体と逆になったと考えた。また、九州地方の最終的な累計感染者数は日本全体の傾向と同じように 2007 年に比べ 2006 年が多くなっているため、発生の条件が標準的な場合は四国や中国、東北地方と同じ傾向があることがわかった。

6 まとめ

感染者数の最終予測には前半部分のデータだけでは難しいことがわかった。理由としては複数の曲線が重なり合っているためであり後半部分がどのようになるかは予測できないせいである。そのためまず曲線の内部構造を理解した結果、曲線を前半部分と後半部分にわけてそれぞれについて近似式を求めることで正確に曲線を近似できた。ここで感染者数の伸びに絞った場合、地域差は現れるのかどうかについてクラスタ分析を用いて解析すると北海道、都市圏、その他にうまく分かれた。やはり感染症であるため人口や人口密度が感染率に大きな影響を及ぼしていることが確認できた。

7 おわりに

これまでの結果から、不確定要素の多くあったインフルエンザについて確定的なものを見つけることができた。データの扱い方に同じ基準を設けた結果、都市圏とその他に分かれたが四国、中国地方の特異性がここまで大きく出ると考えてはいなかった。この点でも理解が深まった。やはり人と人との距離が感染症であるため重要な部分であるが四国、中国地方等予測を超える部分を見つけることができ有意義な研究であったと感じた。

参考文献

- [1] 各都道府県の感染症情報センター。
- [2] 国立感染症研究所 感染症情報センター、
<http://hasseidoko.mhlw.go.jp/Hasseidoko/Levelmap/flu/index.html>