

地域と購入車種に関する統計的分析

2012SE118 小島修兄

指導教員：木村美善

1 はじめに

私が本研究をはじめるときかけとなったことが3つある。一つ目は自分自身が自動車が好きであるということ。二つ目は自動車産業が日本経済に大きな影響を与えていること。三つ目は自分自身が自動車メーカーに就職することである。そして自動車購入台数に何が一番影響を及ぼしているのかを知りたいと考えた。さらに都道府県間の違いを統計的に分析することで就職後の自分の役に立ち、そしてまた社会の役にも立つのではないかと思う。

2 データについて

本研究では、「一般財団法人自動車検査登録情報協会」のホームページ ([4], [5], [6], [7], [8] 参照) から H26 年度の各都道府県別、車種別 (普通車, 軽自動車, 貨物車) の「自動車保有台数」を得た。次の順に変数名をつける, x_2 : 公示地価平均 (m^2), x_3 : 年平均気温, x_4 : 年間降水量 (mm), x_5 : 最低賃金時間額 (円), x_6 : 平均年齢, x_7 : 平均寿命 (男), x_8 : 平均寿命 (女), x_9 : 生活保護受給者率 (%), x_{10} : 平均年収 (円), x_{11} : 農業・林業就業者率 (%), x_{12} : 漁業就業者率 (%), x_{13} : 附置義務駐車場 (台数), x_{14} : コンビニ数, x_{15} : 免許保有率 (%), x_{16} : 駅数 ($/100km^2$) x_{17} : 空港設置数, x_{18} : 高速道路 IC 数, x_{19} : インターネット利用率 (%) とする。都道府県の番号は 1:北海道, 2:青森, ..., 47:沖縄とする。また, 人口データは重みをつけるときのみ使用した。

3 車種別保有数の重回帰分析

y (普通車, 軽自動車, 貨物車) を目的変数とし, x_2, \dots, x_{19} を説明変数として重回帰分析を行い, 自動車の保有数がどのような変数に影響を受けているのかを考察していく。人口を説明変数に入れて解析したところ, すべての車種の決定係数が約 99% となり意味がないので外して行う。また等分散性ではないので, 加重回帰分析を行う。([1] 参照)

4 普通車の加重回帰分析

加重回帰分析を行う際に, 変数選択法 (ステップワイズ法)([2], [3] 参照) を用いて変数選択を行う。まずは, 47 都道府県全てのデータを用いて, 回帰分析を行う。

4.1 分析結果

決定係数は 0.913, 自由度調整済み決定係数は 0.895 となった。AIC を最小にする変数として, $x_4, x_6, x_7, x_8, x_{12}, x_{14}, x_{17}, x_{18}$ が選択された。普通自動車の保有台数が多い地域は, 東京や大阪のような都会過ぎるところよりは, 人口もある程度多い地域でかつ田舎の地域であるこ

表 1 普通車の加重回帰分析結果

変数	推定値	t 値	p 値
intercept	1.323×10^7	2.447	0.19×10^{-1}
x_4	6.563×10	1.324	1.93×10^{-1}
x_6	-5.177×10^4	-2.629	0.12×10^{-1}
x_7	9.076×10^4	1.325	1.93×10^{-1}
x_8	-2.119×10^5	-2.165	0.37×10^{-1}
x_{12}	9.605×10^4	1.352	1.84×10^{-1}
x_{14}	3.614×10^2	10.680	5.31×10^{-13}
x_{17}	-1.788×10^4	-1.514	1.38×10^{-1}
x_{18}	1.563×10^4	6.028	5.20×10^{-7}

とがわかった。

5 軽自動車の加重回帰分析

普通車と同様に分析していく。

5.1 分析結果

表 2 軽自動車の加重回帰分析結果

変数	推定値	t 値	p 値
intercept	1.7071×10^7	2.230	0.0317
x_3	4.638×10^4	3.466	0.13×10^{-2}
x_6	-4.429×10^4	-2.295	0.27×10^{-1}
x_8	-9.401×10^4	-1.740	0.90×10^{-1}
x_{14}	1.133×10^2	2.600	0.13×10^{-1}
x_{15}	-1.351×10^4	-1.498	1.42×10^{-1}
x_{16}	-1.3524×10^4	-1.633	1.11×10^{-1}
x_{17}	-1.705×10^4	-1.982	0.55×10^{-1}
x_{18}	1.972×10^4	8.181	6.61×10^{-10}

決定係数は 0.827, 自由度調整済み決定係数は 0.790 となった。AIC を最小にする変数として, $x_3, x_6, x_8, x_{14}, x_{15}, x_{16}, x_{17}, x_{18}$ が選択された。軽自動車の保有台数が多い地域は, 都会の地域よりは比較的田舎の地域であり, 若い女性の多い地域であることがわかった。

6 貨物車の加重回帰分析

上記と同様に分析を行なう。

6.1 分析結果

決定係数は 0.871, 自由度調整済み決定係数は 0.848 となった。AIC を最小にする変数として, $x_3, x_6, x_8, x_{10}, x_{11}, x_{14}, x_{18}$ が選択された。貨物車の所有台数が高い地域は農業・林業が盛んで高速道路を整備できるような広い土地がある田舎の地域であることがわかった。

表3 貨物車の加重回帰分析結果

変数	推定値	t 値	p 値
intercept	4.112×10^6	1.987	0.54×10^{-1}
x_3	9.798×10^3	1.869	0.69×10^{-1}
x_6	-1.679×10^4	-1.994	0.53×10^{-1}
x_8	-4.310×10^4	-1.834	0.74×10^{-1}
x_{10}	5.253×10^{-2}	1.473	1.49×10^{-1}
x_{11}	1.101×10^4	1.990	0.54×10^{-1}
x_{14}	7.968×10	6.876	3.18×10^{-8}
x_{18}	8.015×10^3	8.245	4.47×10^{-10}

7 都道府県間の違いを分析

7.1 データについて

普通車の重回帰分析で変数選択をして残った所有台数と関係性が高い説明変数について主成分分析 [2], [3] を行った。第4主成分までの累積寄与率が78%以上になったので第4主成分まで分析した。

7.2 分析結果

表4 主成分負荷量

	降水量	年齢	寿命(男)	寿命(女)
PC1	0.089	0.745	-0.777	-0.424
PC2	-0.460	-0.249	-0.478	-0.766
PC3	0.390	-0.249	-0.067	0.261
PC4	-0.562	0.304	0.243	0.296

	漁業者率	コンビニ数	空港数	IC数
PC1	0.737	-0.568	0.086	-0.191
PC2	-0.092	0.582	0.105	0.422
PC3	0.397	0.260	0.931	0.186
PC4	0.229	-0.139	0.063	0.186

第1主成分は、その地域に住む人が漁業にどれだけ携わっているかを示す尺度である。第2主成分は、その地域の生活設備等の住みやすさを示す尺度である。第3主成分は、どのようなロケーション・場所にあるのかといった地理的な尺度である。第4主成分は、その地域の天候を示す尺度である。

8 クラスタ分析

8.1 データについて

本研究では、47都道府県と重回帰分析で残った8つの変数をデータとした。主成分得点を用いてウォード法 ([2] 参照) で分類した。距離を10で切り5群に分類した。

8.2 分析結果

クラスタ分析によって分類された5群は以下の通りである。

・第一群 :15(山梨県), 16(新潟県), 17(富山県), 18(石川県), 21(岐阜県), 22(静岡県), 39(高知県), 41(佐賀県),

45(宮崎県), 46(鹿児島県)

・第二群 : 4(宮城県), 6(山形県), 14(神奈川県), 19(長野県), 20(福井県), 24(三重県), 25(滋賀県), 26(京都府), 28(奈良県), 32(島根県), 33(岡山県), 34(広島県), 36(徳島県), 37(香川県), 38(愛媛県), 43(熊本県), 44(大分県)

・第三群 : 2(青森県), 3(岩手県), 5(秋田県), 29(和歌山県), 31(鳥取県), 35(山口県), 42(長崎県)

・第四群 : 7(福島県), 8(茨城県), 9(栃木県), 10(群馬県), 11(埼玉県), 12(千葉県), 23(愛知県), 27(大阪府), 30(兵庫県), 40(福岡県)

・第五群 : 1(北海道), 13(東京), 47(沖縄県)

8.3 考察

第一群は、年間の降水量が多いが他の変数はあまり差がない数値となった。第二群は、降水量が一番低く、平均寿命(男)(女)が最も高い群となった、このことから長寿の群といえる。第三群は、漁業就業率が最も高く、平均寿命、コンビニ数が低い群であった。漁港が盛んな群であるといえる。第四群は、高速道路IC数が、最も多い群で、漁業就業率、空港設置数が一番少ない群であり他の変数も低めの数値であった。第五群は、コンビニ数、空港設置数が一番多い群であり、平均年齢が若い群であった。また他の変数も比較的高い数値であった。

9 おわりに

本研究を終えて、自動車の購入台数には収入に関わる要因が影響をしていることが改めてわかった。また、地域間で群分けの結果を得ることができた。本研究で得た結果を今後の仕事に役立てていきたいと思う。

参考文献

- [1] Chatterjee, S. and B. Price: Regression Analysis by Example, Jhon 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] 中村 永友: 多次元データ解析法, 共立出版, 2009.
- [3] 永田 靖・棟近 雅彦: 多変量解析法入門, サイエンス社, 2001.
- [4] 自動車保有台数統計データ
<http://www.airia.or.jp/number/index.html>
- [5] 統計局ホームページ
<http://www.stat.go.jp/data/ssds/5.htm>
- [6] 土地代データ
<http://www.tochidai.info/>
- [7] 都道府県データランキング
<http://uub.jp/pdr/t/c.html>
- [8] 全国軽自動車協会連合会
<http://www.zenkeijikyo.or.jp/statistics/>