

感性情報処理 レポート 3

秋田 一樹 (CRb 29119001)

2019 年 7 月 29 日

1 今回の分析テーマ

今回の課題は”名古屋市の統計データを重回帰分析し、分析せよ”といった課題であり、オリジナルティを出すこととの指定があった。私にとって身近なものをテーマに選択しようと思ったところ、私は塾のアルバイトをしていることから、教育のテーマを選択しようと思い立った。

2 使用するデータの選択

課題で指定されたサイトにある教育の欄には、大きく分けて在籍生徒数と学校数が載っていたが、在籍生徒数は人口欄との関連性が強く出そうであったため、あえて学校数 (各区分) を目的変数として分析を行った。

初めに説明変数として区別面積, 区別世帯数, 区別人口を用意して単回帰分析を行ってみたところ、

- 目的変数 : 区別小学校, 中学校, 高校の総数
- 区別面積による単回帰分析 $y = 14.13 + 0.511x$, $R^2 = 0.637$
- 区別世帯数による単回帰分析 $y = -0.0493 + 0.000357x$, $R^2 = 0.823$
- 区別人口による単回帰分析 $y = 3.55 + 0.000145x$, $R^2 = 0.935$ という結果が出た。

この結果から、区別小学校, 中学校, 高校の総数についても、おおよそ人口だけで説明できてしまうことがわかった。横軸に人口、縦軸に学校総数をとった式を以下の図 1 に示す。

もちろんこれでは重回帰分析をする必要性がなくなってしまう。そこで注目したのは、高校とその他の値との相関係数である。高校数とそれぞれの相関係数は次のようになっていた。

$r_{\text{高校数, 面積}} = -0.389$, $r_{\text{高校数, 世帯数}} = -0.122$, $r_{\text{高校数, 人口}} = -0.219$, $r_{\text{高校数, 小学校数}} = -0.325$, $r_{\text{高校数, 中学校数}} = 0.400$ というように、高校数は他の変数との関係が特殊のように感じた。そのため、高校の数を目的変数として重回帰分析を行った。

高校数を目的変数とする時の問題点は、区別高校数は平均が 4 未満と小さい数であることである。その問題を解決するため、これ以降はデータを標準化してから分析することとする。

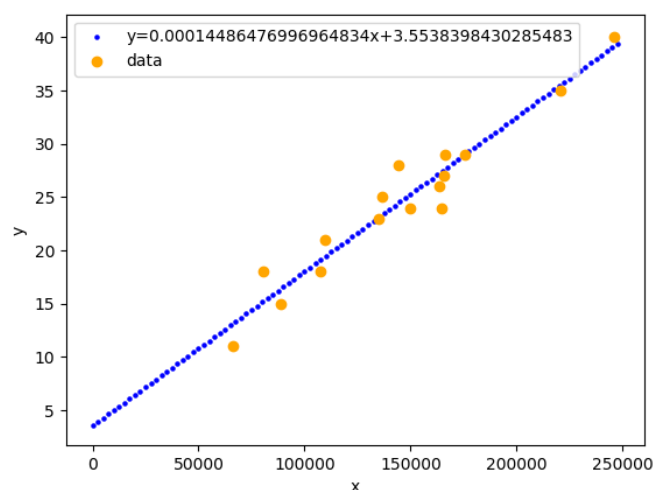


図 1 単回帰分析結果 (x:人口, y:学校数)

3 重回帰分析

目的変数を区別高校数とするが、全章で扱った面積、人口、世帯数といった基礎的なデータは、主成分分析を行って1つの変数としてまとめて扱おうとした。しかし、主成分分析を行って得られた主成分得点では目的変数と関連するものは得られなかった。そこで、“人口とは関係が見られなかったが、人口密度との関係はある”という仮説を立てて、人口密度を用いて分析を行うことにした。

重回帰分析に用いる2つ目の変数は、小学校数、中学校数を用いて作成することにした。もう一度相関係数に着目してみると、 $r_{\text{高校数}, \text{小学校数}} = -0.325$, $r_{\text{高校数}, \text{中学校数}} = 0.400$ というように、符号が反転しているが、 $r_{\text{高校数}, \text{小学校数}}$, $r_{\text{高校数}, \text{中学校数}}$ は同程度の相関があることが見られた。これを参考にし、中学校数/(小学校数 + 中学校数), 小, 中学校数に対する中学校数の割合を指標にできるのではないかと考えた。

以上より、

目的変数：区別高校数 (y)、説明変数：人口密度 (x_1), 小, 中学校数に対する中学校数の割合 (x_2) をそれぞれの変数で標準化し、重回帰分析をして得られた結果は、

$y = 3.94 + 0.295x_1 + 1.73x_2$ となった。この結果を三次元空間上にプロットしたものを以下の図 2 に示す。

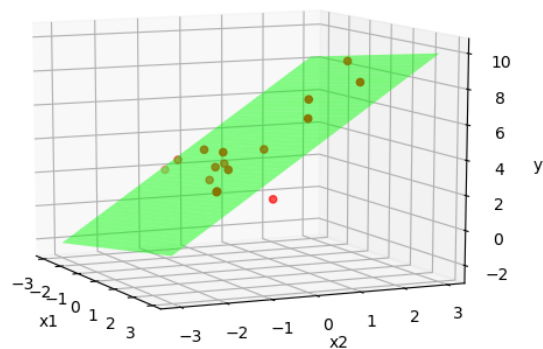


図2 重回帰分析結果 1(x1:人口密度, x2:小学校に対する中学校数の割合, y:高校数)

重回帰分析の結果として、グラフ以上に自由度調整済み寄与率が 0.649 と低い結果になっていた。その要因として、1 点だけ大きく外れてしまっている点があることがわかる。この点は熱田区であり、この点を除いて重回帰分析をして得られたグラフを以下の図 3 に示す。

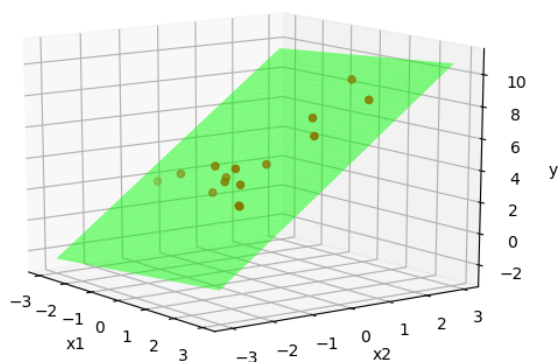


図3 重回帰分析結果 2(熱田区データ削除)

この時の寄与率は R^2 は 0.898, 自由度調整済み寄与率は R^{*2} は 0.882 まで上がった。

4 考察

以上の結果より、名古屋市のデータでは熱田区を除けば、区別の人口密度と小, 中学校に対する中学校数の割合から、区別高校数を表す式を求めることができた。今回は熱田区を除外したが、熱田区に高校数が少ないことにも何か要因があるはずである。熱田区は他の区と比べ、面積を考慮しても高校だけでなく、小学校, 中学校の数も少ない。予想される要因としては、熱田神宮や金山駅といった大規模な施設があるということが挙げられる。

5 使用したデータ

- 2018-02-05 名古屋市各区分 面積, 人口 (総数), 世帯数
- 2018-18-02 名古屋市各区分 学校数 (小学校, 中学校, 高校)