

# Lasso 回帰による幸福度分析 —長久手市の幸福度調査を事例として—

内 田 俊 博

## 要 旨

本研究は、2016年に実施された愛知県長久手市の「ながくて幸せ実感アンケート」の個票データを用いて、幸福度に影響を与える要因を Lasso 回帰を用いて特定することを目的とする。分析の結果、概ね先行研究と同様の結果が得られたが、「家計の状況（所得・消費）の満足度」の推定値が全変数の中で2番目に大きな値となった一方、世帯収入の推定値は0となった。この結果は、世帯年収がそれ程高くなくても「家計の状況」に満足している人も多くいて、幸福度を予測するにあたって重要なのは、所得そのものではなくどれだけ所得に満足しているかである、と解釈できる。

## 1. 序論

幸福の経済学とは、主観的な幸福度を実証分析の変数として用いる経済学研究全般を指すものである (Clark, 2018)。その中でも特に、目的変数に幸福度を用いて幸福度に影響を与える要因を回帰分析する研究は、幸福の経済学の初期の段階から数多く研究されてきている (Clark, 2018)。個票データを用いた幸福の経済学研究では、幸福度に影響を与えると思われる個人属性が説明変数として用いられる。しかし、どのような要因が幸福度に影響を与えるのかを経済理論から導くことは難しいため、回帰分析に用いる変数の選択は、先行研究とともに研究者の主観に基づいている部分が強いと言える。このような変数選択の方法には2つの問題点がある。第一に、変数選択が既存研究に基づいているため、幸福度に影響を与える新たな変数を見つけること難しい点である。第二に、説明変数の候補には数多くの変数が考えられるが、それらの中から実際に幸福度の予測能力が高い変数を選び出す方法が、客観的かつシステムティックでない点である。

本研究では、Lasso 回帰 (least absolute shrinkage and selection operator) を用いることで、

上記2点の克服を目指す。Lasso 回帰 (Tibshirani, 1996) は、過学習を抑えるために L1 正則化を取り入れた線形回帰モデルであり、目的変数の予測能力が低い変数を回帰モデルに多く含めて推定を行っても、それらの回帰係数が 0 になりやすい性質を持つ。そのため、多くの説明変数の中から予測能力の高い変数を明確に選び出すことを容易にする。また、回帰係数の推定を通して客観的かつシステムティックに説明変数の選択を行うことができる。先にも述べたように、幸福度に影響を与える要因は基本的に経済理論から導くことは難しいため、予測能力の高い機械学習の推定手法を用いることは、研究目的に照らし合わせて理にかなっていると言える。本研究では、具体的には2016年に実施された愛知県長久手市の「ながくて幸せ実感アンケート」の個票データを用いて、幸福度に影響を与える要因を Lasso 回帰を用いて特定することを試みる。

本論文の構成は以下のようになっている。次節では本研究で用いるデータ及び分析手法について説明する。第3節では結果について説明し、第4節では本研究で明らかになったことをまとめる。

## 2. データ及び分析手法

愛知県長久手市では、地域づくりの一環として2014年度に「ながくて幸せ実感アンケート」、更に2016年度には「第2回ながくて幸せ実感アンケート」を実施し、両アンケート調査の調査票・報告書・および個票データをホームページ上で公開している<sup>1</sup>。本研究では、ホームページ上で公開されている「第2回ながくて幸せ実感アンケート」の個票データを使用する。第2回ながくて幸せ実感アンケートでは、住民基本台帳から無作為に選んだ長久手市内在住の満18歳以上の市民、および第1回のアンケートで次回もアンケートに協力してもよい回答した市民を合わせた5000人に対して調査票を郵送し、回収も郵送で行った。調査は2016年12月から2017年1月にかけて行われ、有効回収数は2569、有効回収率は51.4%であった。

本研究では、幸福度に影響を与える要因を特定するために、Lasso 回帰を用いる。Lasso 回帰は、回帰式そのものは通常の重回帰分析と同様である。異なるのは損失関数であり、通常の重回帰分析では平均二乗誤差 (MSE: Mean Squared Error) を最小化するパラメータの値を推定するのに対し、Lasso 回帰では正則化項を加えた

$$\text{MSE} + \alpha \sum_{i=1}^k |w_i|$$

---

1 第2回ながくて幸せ実感アンケート調査結果 (平成28年度実施)  
<https://www.city.nagakute.lg.jp/keiei/siawasenomonosashi/anketo2016.html>

を最小化するように推定する。ここで、 $w_i$  は回帰係数である。

Lasso 回帰の目的変数には、以下の質問によって測られる幸福度を用いた。『あなたは現在幸せですか。「とても幸せ」を10点、「とても不幸」を0点とすると、何点くらいになると思いますか。』

一方説明変数であるが、本研究の目的に照らし合わせて、原則としてアンケート調査票に含まれる全ての質問を説明変数として用いることを試みた。ただし、以下の基準によりいくつかの変数は排除した。(1) 欠損値の多い変数、(2) 目的変数以外で、回答者の幸福度を直接測っていると判断される変数。たとえば、(2) に当てはまる例としては、生活全般についてどの程度満足しているか尋ねている質問が挙げられる。以上により、最終的に分析に使用した説明変数の数は55となった。また、欠損値については削除し、最終的なサンプルサイズは1941となった<sup>2</sup>。Lasso 回帰は Python (scikit-learn) を用いて行い、全ての説明変数について平均0、分散1となるように標準化を行った。正則化パラメータ  $\alpha$  の値はグリッドサーチにより決定した<sup>3</sup>。

### 3. 結果

表1にはLasso回帰により得られた推定値を、絶対値の大きな順に上位10まで示してある。これらの変数は、大きく以下のように分類することができる。

表1 推定値(絶対値)の大きい説明変数

| 変数  | 推定値   |
|---|-------|
| 家族関係(子育て、教育、夫婦、父母)の満足度                      | 0.477 |
| 家計の状況(所得・消費)の満足度                            | 0.270 |
| 健康状況の満足度                                    | 0.209 |
| 配偶者の有無                                      | 0.171 |
| 仕事や趣味、社会貢献などの生きがいの満足度                       | 0.163 |
| ストレスを発散する場や機会、精神的なやすらぎの場があるか                | 0.112 |
| 長久手に愛着を感じているか                               | 0.106 |
| 長久手を住みよいまちだと思うか                             | 0.095 |
| 困ったときに頼りになる(悩みを相談したり助けてと言ったりできる)知人・友人が市内にいる | 0.074 |
| 地域のまち並み(景観・風景)はきれいだと思うか                     | 0.036 |

\* 推定値の大きさを直接比較できるように、説明変数は全て標準化して回帰分析を行った。

2 世帯収入については無回答による欠損値が多かったが、経済学的に非常に重要な変数のため、削除せずに分析に使用した。最終的なサンプルサイズは2569から1941に減っているが、削除した観察値の大半は世帯収入が欠損していたためである。

3 グリッドサーチによる値は  $\alpha = 0.1$  となった。

### (1) 人間関係

「家族関係の満足度」, 「配偶者の有無」, 「困ったとき頼りになる知人・友人が市内にいる」の3つの変数は, 良好な人間関係を示していると解釈できる。「家族関係の満足度」の推定値は全ての説明変数の中で1番大きく0.477であり, 2番目に大きな「家計の状況の満足度」の推定値0.270を大きく上回っている。また, 「配偶者の有無」も0.171と4番目に大きく, 良好な人間関係は幸福度に大きくプラスに働くことを示している。

### (2) 所得の満足度

「家計の状況(所得・消費)の満足度」は全説明変数の中で2番目に大きく, 0.270となっている。先行研究では, クロスセクションデータの場合, 所得水準の高さは幸福度の高さとの正の相関が見られるとの結果が多く(Clark, 2018), 一見すると本研究も先行研究と同様の結果が得られたように思われる。しかしながら, 本研究では「家計の状況(所得・消費)の満足度」とは別に, 所得そのものを表す「世帯年収」という説明変数もあり, こちらの推定値は0となっている。また, 「家計の状況(所得・消費)の満足度」と「世帯年収」の相関係数は0.384とそれ程大きくないため, 前者の説明変数は所得そのものとは別の側面を表していると考えられる。

一般的に, 世帯年収が高いほど「家計の状況(所得・消費)の満足度」は高くなるかと予想できる。しかし, 相関係数が0.384であることから, 世帯年収がそれ程高くなくても「家計の状況」に満足している回答者が多くいると考えられる。そして本研究の結果は, 幸福度を予測するにあたって重要なのは, 所得そのものではなくどれだけ所得に満足しているかである, と解釈できる。所得の満足度を説明変数に加えた研究は筆者の知る限りはじめてであり, また所得(世帯年収)に関する推定の結果も先行研究とは異なる重要な示唆が得られた。これらの結果については, 今後, 他のデータによる検証が待たれる。

### (3) 健康状況

「健康状況の満足度」及び「ストレスを発散する場や機会, 精神的なやすらぎの場があるか」の2変数は, 肉体的・精神的な健康状況を表していると解釈できる。「健康状況の満足度」の推定値は3番目に大きく, 健康状況が幸福度に大きく影響するという結果は先行研究と同様の結果となった。

#### (4) 生きがい

「仕事や趣味、社会貢献などの生きがいの満足度」は、質問の通り生きがいを表していると言える。ただし、生きがいがあることが幸福な状態であるとも定義できる点に加え、個人の生きがいそのものを政策的に変えることは難しいため、この結果を政策として活かすことは難しい部分がある。

#### (5) 住環境

「長久手に愛着を感じているか」、「長久手を住みよいまちだと思うか」、及び「地域のまち並み（景観・風景）はきれいだと思うか」の3変数は、地域の住環境を表していると解釈できる。

ここまでは推定値の大きな説明変数について説明したが、以下では主な説明変数の中から推定値が小さくなったものを取り上げる。年齢の影響については様々な研究結果が報告されているが、一般的にはU shape、すなわち40代の中年の幸福度が下がる例が多くなっている（Frey, 2018）。本研究では、年齢の影響は小さく、 $-0.035$ となった。他の説明変数の大きさと比較すると、年齢の影響はほぼ無視できると言える<sup>4</sup>。性別については、本研究では推定値は0となった。先行研究では女性の方が有意に幸福度が高くなることが多いが（Clark, 2018）、本研究では異なる結果となった。また上で説明したように、世帯年収の推定値は0となった。

## 4. 結論

本研究では、愛知県長久手市の「ながくて幸せ実感アンケート」の個票データを用いて、幸福度に影響を与える要因をLasso回帰を用いて特定することを試みた。概ね先行研究と同様の結果が得られたが、重要と思われる新たな知見が何点か得られた。一番重要な結果は、世帯収入の推定値が0となったことであろう。「家計の状況（所得・消費）の満足度」も説明変数として加えた結果、こちらの変数の推定値は全変数の中で2番目に大きな値となった。世帯年収がそれ程高くなくても「家計の状況」に満足している人も多くいると考えられ、幸福度を予測するにあたって重要なのは、所得そのものではな

---

4 世帯収入の二乗項も加えたLasso回帰も行ったが、その場合も年齢の影響はほぼ無視できる結果となった。また、先行研究のようなU shapeの結果は得られなかった。

くどれだけ所得に満足しているかである、と解釈できる。また、年齢や性別という重要な個人属性の影響がほぼ0になった。これらの知見は、今後別のデータセットを用いて更に検証していくことが期待される。

分析手法の観点から見た本研究の貢献は、幸福の経済分析に Lasso 回帰を用いたことである。Lasso 回帰は、回帰係数の推定を通して客観的かつシステムティックに説明変数の選択を行うことを可能にする。幸福度に影響を与える要因は基本的に経済理論から導くことは難しいため、予測能力の高い機械学習の推定手法を用いることは、研究目的に照らし合わせて理にかなっていると言える。本研究が同様の問題意識を持つ他の幸福の経済学研究の助けになれば幸いである。

## 引用文献

- Clark, A. E. (2018) “Four decades of the economics of happiness: Where next?” *The Review of Income and Wealth*, Vol. 64(2), pp.245-269.
- Frey, B. S. (2018) *Economics of Happiness*, Basel: Springer.
- Tibshirani, R. (1996) “Regression shrinkage and selection via the lasso”, *Journal of the Royal Statistical Society*, Vol. 58(1), pp.267–288.