

幸福度要因の LASSO分析

中京大学 内田ゼミ

伊藤綾

河合千里

四井友偉

諏訪祥輝

田中毎樹

土田純己

幸福の経済学とは

経済学において、

“人々の幸福を実現するにはどうしたらいいか”

を考える学問



幸福の経済学の始まり

1974年に経済学者リチャード・イースタリンによって発表された、「イースタリン・パラドックス」という現象が指摘されてから本格的に研究されるようになった。

イースタリン・パラドックスとは、普通、所得が増えれば個人の感じる幸福感は増大すると考えられているが、実際には所得が増えても幸福感が一定で、ほとんど変化しないという現象。

幸福感を決めるものとして考えられるのは？



性別



年齢



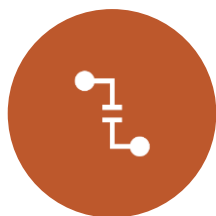
健康



結婚



学歴



人とのつながり



年収



所得

リサーチクエスト

幸福度にどのような行動がどのような影響を与えているのか



データ

今回は2012年2月～4月に実施された日本語版 General Social Surveysを東京大学社会科学研究所の許可を取って使用。

面接調査と留置調査を併用し、2332のサンプルから分析する。(有効回収率59.1%)

分析手法

LASSO回帰の特徴

今回はPythonのscikitlearnを使ってLASSO回帰分析を行う。

LASSO回帰とは、正則化された線形回帰の一つ。最小二乗コスト関数に対して重みの合計を足したもの。

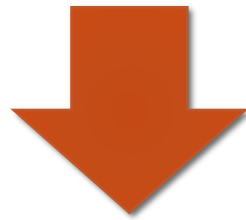
不要と判断される説明変数の重みが0になる

→→モデル構築においてくつかの変数が無視される

→→モデル解釈しやすくなる！！

LASSOの強み

- ・通常の分析だと限られた変数で分析するので、研究者が関係ありそうな変数をいくつか主観で決める

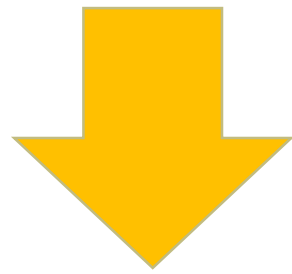


LASSOの場合

- ・不要と判断される説明変数の重みは0となるので変数を絞る必要がない
※今回は54の変数を使用した

以上のことから従来よりも客観的な分析が可能と言える！

最小二乗関数 $\min \sum (y - f(x, w))^2$



+ 正則化項

LASSO $\min \sum (y - f(x, w))^2 + \alpha \|w\|_1$

重回帰分析の特徴

ある変数Yの変動を別の変数Xの変動により説明・予測・影響関係を検討するための手法。

変数がm個ある時・・・

$$Y = W_1 \times X_1 + \dots + W_m \times X_m$$

(Y: 目的変数 X: 説明変数 W: 偏回帰係数)



正則化

LASSO回帰と重回帰分析違いは正則化。

- ・係数と誤差項の総和を小さくする。
 - ・正則化とは、過学習を防ぎ汎化性能(未知のデータへの対応能力)を高めるため手法。
 - ・過学習・・・データの中の一部の例外的なデータに過度に適用したモデルが構築されること。
- 一部のデータだけに最適化されてしまい汎用性がない

$$\min f(x) + \alpha \sum_{i=1}^n |w_i|$$

Pythonのscikitlearnを使ってLASSO回帰分析を行う。

LASSO回帰は α によって変わる。

→→最適な α を見つけるためにGridsearchを使い
mean_squared_errorを最小にする。

Gridsearch・・・モデルの精度を向上させる手法

変数変換についてのまとめ4点

性別、就労経験、結婚状況、子供の数、家族の人数、居住地域、喫煙経験、居住形態、消費電力を減らす努力のこれら全てについてダミー変数化。

ダミー変数化とは、数字ではないデータを数字に変換する手法のこと。

例えば就労経験は「ある」か「ない」、結婚状況は「結婚している」か「結婚していない」、子供の数は「子供がいる」か「子供がいない」といったように変換する。

①ダミー変数化表 1

性別	就労経験	結婚状況	子どもの数	家族の人数
男性=0	仕事を持っていない=0	結婚していない、 同棲していない=0	子どもがいない =0	ひとり=0
女性=1	仕事を持っている=1	結婚している=1	子どもがいる=1	家族と暮らしている=1

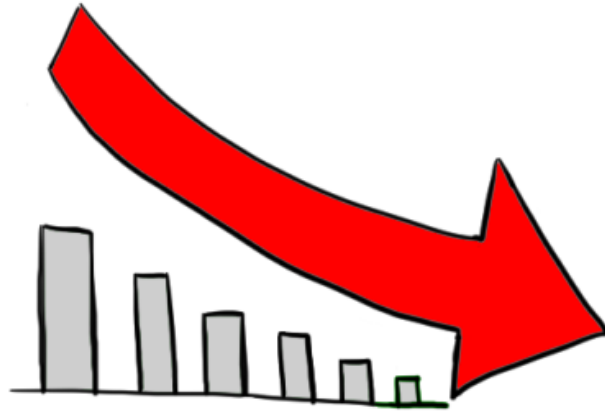
①ダミー変数化表 2

居住地域の規模	喫煙経験	居住形態	消費電力を減らす工夫
町村、農村、漁村 =0	喫煙していない =0	持ち家がない=0	工夫していない =0
都心部=1	喫煙している=1	持ち家がある=1	工夫している=1

②サンプルの削除

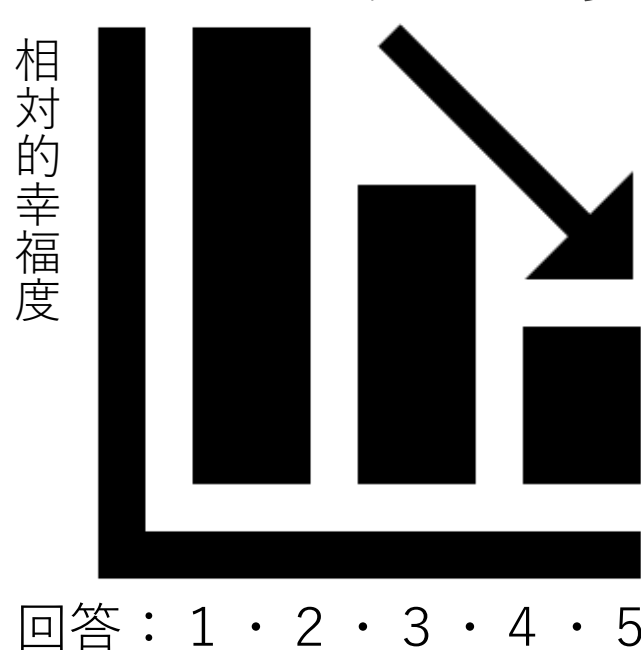
一つ一つの説明変数について無回答のサンプルを削除。

reduce

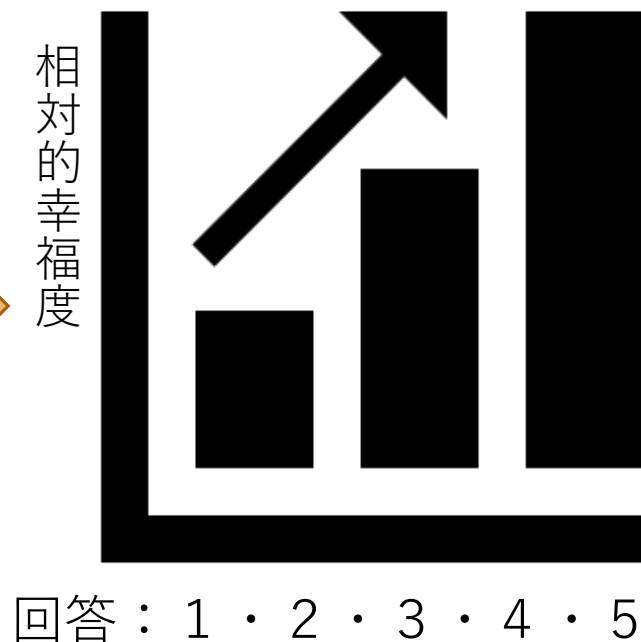


③逆変換

世帯年収、頻度、健康状態、賛成度合い、不安など
「相対的な幸福度」と回答の数字について大小関係を対応
させるように逆変換。



少ない、低いほど数字が小さく、
多い、高いほど数字が大きくな
るようにする

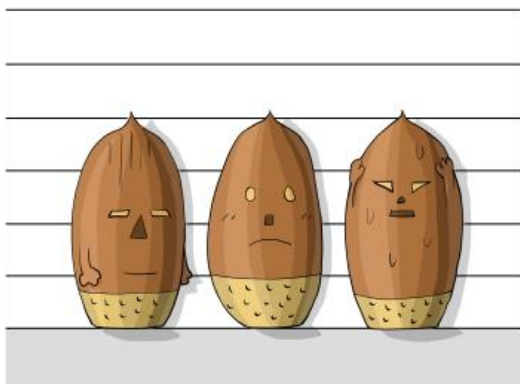


④標準化

多様な変数を比較できるように標準化。

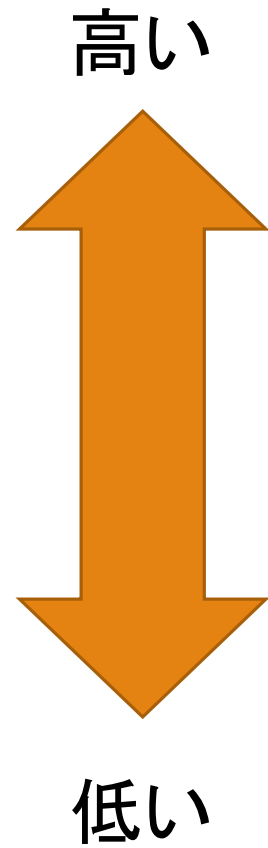
$$Z_i = \frac{x_i - \mu}{\sigma}$$

(Z =標準化変数 x =各データ μ =平均 σ =標準偏差)



結果

係数が正であった変数



0.252	健康状態
0.079	結婚状況
0.076	家族と一緒に夕食
0.076	子供の人数
0.048	親しい友人の数
0.037	友人との会食
0.036	周囲の人はおおむね友好的
0.035	人間の本性
0.035	互いに気にかけている
0.035	性別
0.031	ドライブ
0.018	テレビゲーム
0.013	園芸
0.010	家族人数(本人除く)
0.002	1泊以上の旅行頻度
0.0003	映画鑑賞

係数が正の変数からわかること

- **健康**が圧倒的に関係している
- **結婚、家庭**に関わるのが上位である
- 人間関係がその次
- 最後に趣味系がある



係数が負であった変数

- 0.168 NOFUTR 希望のなさ
- 0.103 AXECNSF 将来の経済的不安
- 0.032 NOGOAL 目標は達成できない
- 0.022 OP5RADCA 政治意識
- 0.013 XTRAUM5Y トラウマの回数
- 0.010 DORL 信仰する宗教の有無

係数が負の変数からわかること

- ・負の関係があったものは、ほとんどが現実の状況というより、**気持ちや精神面**と関係したもの。
- ・「希望のなさ」や「将来の経済的不安」といった、**未来に対しての不安**が、大きく関係している。



係数が0になったもの(抜粋)

ZHSINCM

世帯年収

DOSMOKEX

喫煙経験

DO7DRINK

飲酒頻度



係数が0のことからわかったこと

SZHSINCM 世帯年収

→「将来への経済的不安」と幸福度とは負の関係があったが、「世帯年収」とは関係がなかった。

DOSMOKEX 喫煙経験、DO7DRINK 飲酒頻度

→「健康状態」は幸福度と大きくかかわっているが、「喫煙経験」と「飲酒頻度」とは関係がなかった。

LASSOと重回帰分析

LASSOでは重回帰式の精度を表すp値が表せない。



重回帰分析では表せるので重回帰分析をする。

P値・・・0に近づくほど統計的な意味あり

0.05より大きくなるとたまたまな可能性が高い。

重回帰分析

既存研究で重要と言われている5つの変数

- ・世帯年収
- ・年齢
- ・性別
- ・健康状態
- ・結婚状況

比較

既存研究の5つ＋LASSO分析の結果関係の大きかった変数

- ・家族と一緒に夕食
- ・子どもの人数
- ・親しい友人の数
- ・友人との会食
- ・周囲の人が概ね友好的である

重回帰分析の結果1

	推定値	p値
年齢	0.0015	0.298
世帯年収	0.0227**	0.003
性別	0.2267**	0.000
結婚状況	0.4912**	0.000
健康状態	0.3495**	0.000

重回帰分析の結果2

	推定値	p値
年齢	0.0001	0.924
世帯年収	0.0075	0.316
子供の人数	0.1673*	0.014
友人との会食	0.0469**	0.009
性別	0.1420**	0.001
結婚状況	0.3338**	0.000
健康状態	0.3068**	0.000
家族と一緒に夕食	0.0518**	0.000
親しい友人の数	0.0768**	0.000
周囲の人はおおむね友好的だ	0.2268**	0.000

重回帰分析の結果3

年齢 今回の分析では関係性が見られない

世帯年収 既存研究で重要な5つの変数 →→→ 関係性有
既存研究5つ＋関係性があるとされた変数
→→→ 関係性小さくなる！！

結論

- ・幸せになるために**健康**と**人間関係**を良好に保つように日々の生活を送るのがよい。
- ・**精神的な不安**が幸福度と大きくかかわっている。
- ・世帯年収など幸福度と大きく関係していそうな変数とは、今回は関係性が見られなかった。

