1ZH-03

シビックプライド醸成に繋がる住民価値の掘り起こしと貢献度の検証に関する研究

2023.3.2

東京国際工科専門職大学 情報工学科

高田 晃希　黒羽 晟　山本 裕

株式会社 百代

橋本 沙也加　橋本 尚子　岡田 ゆかり

東京国際工科専門職大学
International Professional University of Technology in Tokyo

目次

1. 研究の背景

社会課題

◆近年、日本は少子高齢化に伴って東京一極集中が進み、地方の過疎化・高齢化が進行している。

◆地方公共団体には、こうした地方の衰退を改善したいという課題がある。

シビックプライドの醸成と研究目的

◆「市民の当事者意識に基づく、都市に対しての誇り」[1]という意味の「シビックプライド」という思想に着目。

◆シビックプライドを醸成することにより、地方住民が自治行政を推進し、持続可能なまちづくりが実現できる。
本研究は、シビックプライドの醸成のための施策を地方公共団体へ提案することを目的とする。

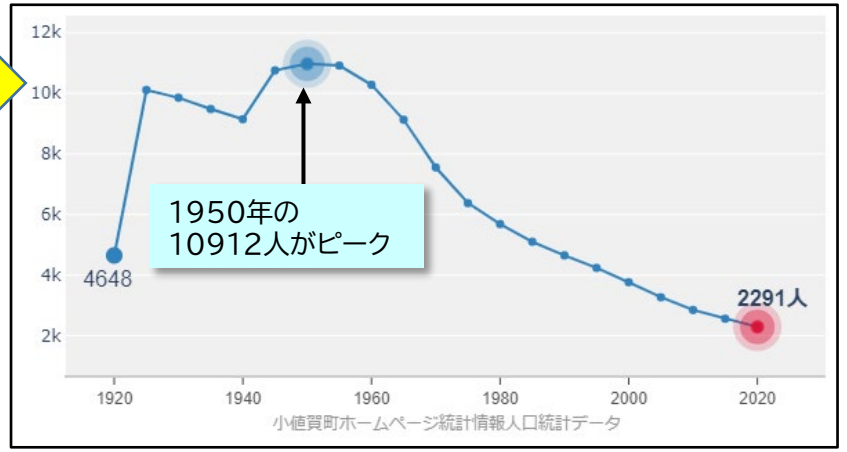
地方課題の解決を1事例とした「潜在的住民価値発掘モデル」

◆長崎県小値賀町の事例を研究対象とさせて頂き、シビックプライド醸成施策導出のための「潜在的な住民価値発掘モデル」を構築、小値賀町を含めた多くの地域での訴求を図るため「潜在的な住民価値発掘モデル」を汎用化したい。

引用：小値賀町ホームページ/町政/総合計画
「少子高齢化、過疎化が進む中で、小値賀町のまちづくりはますます厳しさを増していくことが予想されるため、**町民一人ひとりが誇りと希望をもてる**まちづくりを進めていく必要があります。」

[小値賀町総人口推移グラフ]
1920年～2020年
5年ごとに総人口をプロット

人口が1950年から2020年まで単調減少している。
基幹産業の人口も年々減少。
1990 年には老年人口が年少人口を、更に 2015 年には生産年齢人口を逆転。



2. 研究の概要 -課題と研究概要

◆類似研究の調査による本研究の新規性：アンケート（選択回答＋自由記述回答）から潜在的な住民価値を発掘し、施策を導出する事例は少ない。 参照23ページ

【研究の目的詳細】

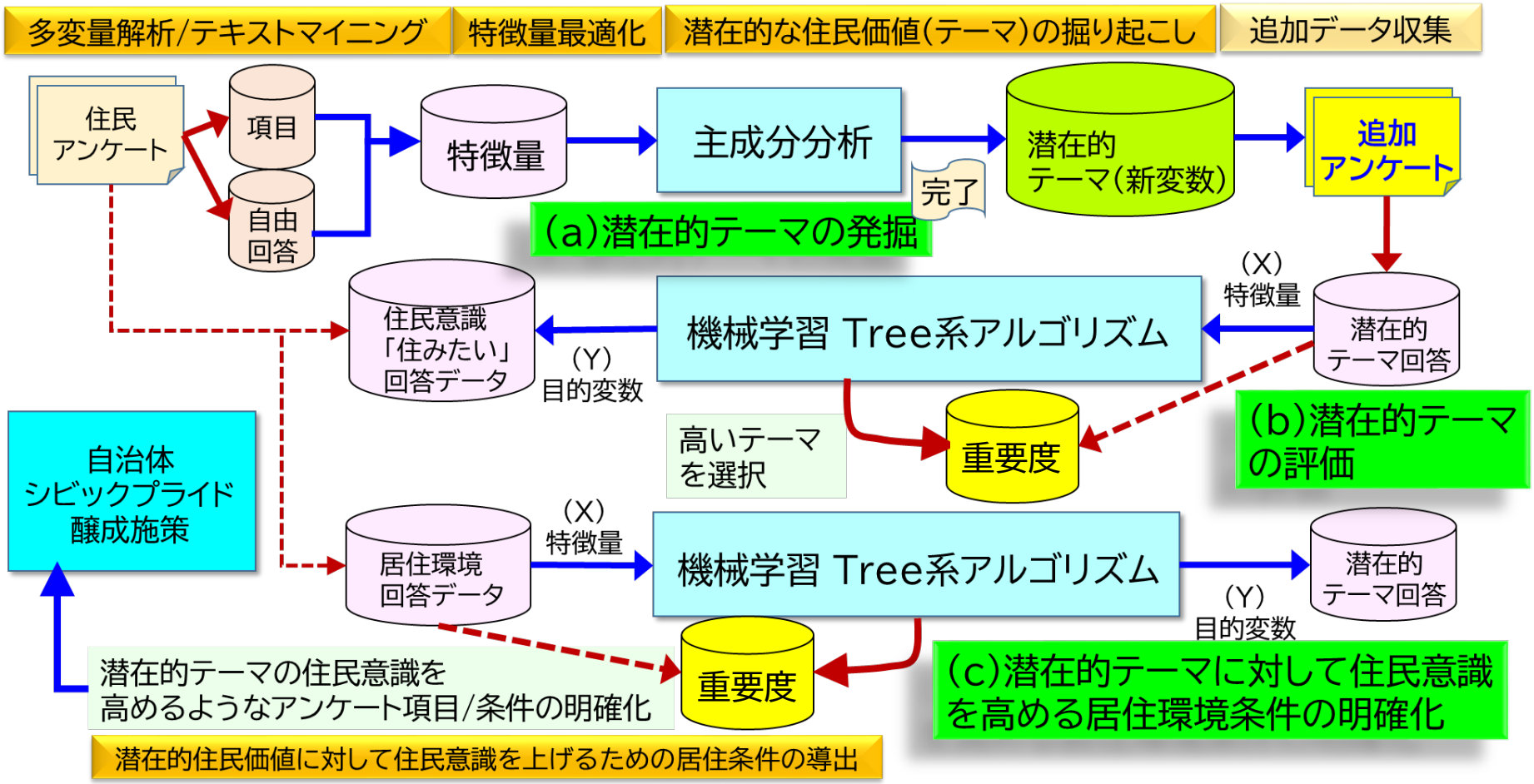
（1）新住民価値の明確化：「小値賀町第4次総合計画アンケート」結果を分析。
AIを用いて**シビックプライド醸成に寄与する潜在的な「新住民価値」の発掘**を行う。

（2）新住民価値創出のための施策提案：
発掘した「新住民価値」を向上させるために、（アンケートの項目に紐づく）**どういう環境要件が重要か**を明確にする。
明確にした環境要件から、具体的な施策案を検討・提案する。 →汎用化

目的	研究プロセス		概要 今回発表の範囲
（1）新住民価値の明確化	（a）潜在的テーマの発掘	（a-1）特徴量選択	アンケート結果から潜在的な住民価値発掘の「ための主成分分析の精度向上のための特徴量選択。 数量化Ⅱ類、相関分析、テキストマイニング、機械学習
		（a-2）新住民価値の概念発掘	主成分の意味付け（主成分分析）とアンケート追加
	（b）潜在的テーマの評価		追加アンケート結果の分析 （追加アンケート結果を（説明変数（X））とした住み易い（目的変数（y））に対する重要度）
（2）新住民価値創出のための施策提案	（c）潜在的テーマに対して住民意識を高める居住環境条件の明確化		追加アンケート結果の住民意識を高めるような、居住環境条件を導出する （追加アンケート結果を（目的変数（y））とし他アンケート回答を（説明変数（X））とした場合の（X）の重要度評価）

2. 研究の概要 -主な研究アプローチ

・本研究のプロセスとして下記(a)(b)(c)の3つの課題を通して目的を達成する。
・今回の発表対象は(a)(2022年度完了)。(b)(c)は推進中(2023年度完了予定)



研究プロセス(a)→(b)→(c)

(a) 小値賀町住民アンケート(2018年度)から潜在的な新住民価値を発掘。(2022年度)
※(a)の概要は次ページ

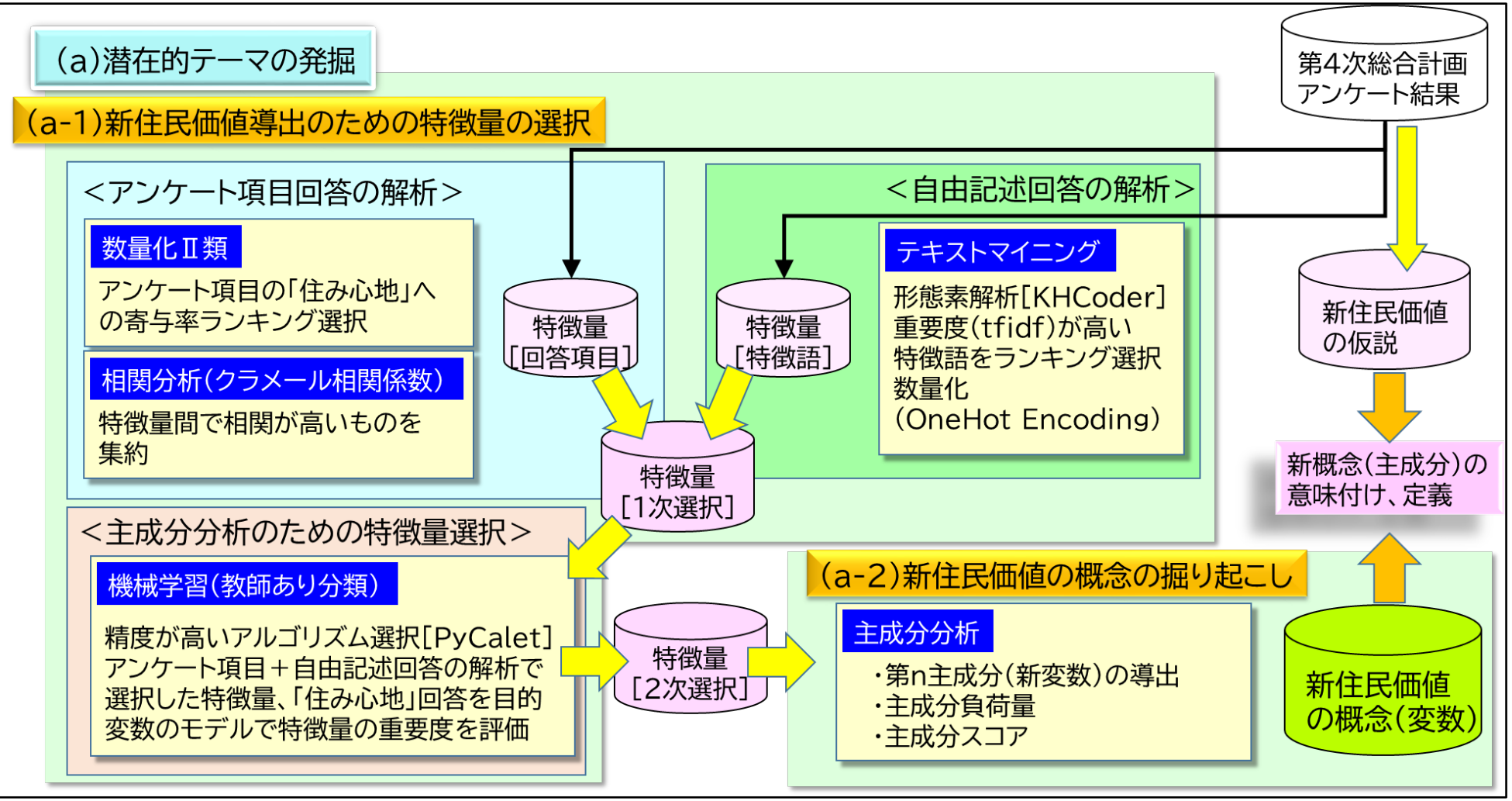
(b) 導出した新住民価値(潜在的テーマ)を新項目として2023年度のアンケートに追加して頂いた。
住み心地回答に対する新項目回答の重要度が高いかどうかを検証。
((a)の仮説検証)

(c) 2023年度アンケートを対象として新項目回答に対する他アンケート項目の重要度を評価。

重要度が高い特徴量をシビックプライドの醸成に寄与するものとし、施策提案。

2. 研究の概要 -主な研究アプローチ

今回発表の対象となる(a)の概要・プロセス



(a)は2つのフェーズに分かれる。
大きな指針は新住民価値を発掘すること。

(a-1)アンケートから特徴量選択。

(a-2)選択した特徴量から
主成分分析で新住民価値を発掘。

(a-1)特徴量選択の意味

新住民価値はシビックプライド醸成に寄与するものである必要がある。

そのため、アンケートからシビックプライド醸成に寄与する特徴量を選別・選択してから、
(a-2)新住民価値発掘を行う。

3. 新住民価値導出のための特徴量の選択

(1)数量化Ⅱ類

アンケート回答項目を説明変数、アンケート回答項目の選択項目をカテゴリと定義し、
モデル式のカテゴリ係数を算出する。

数量化2類の詳細はP32

特徴量選択の意義はP28

数量化Ⅱ類の適用例

住民アンケート

1. 住み心地はどうか？
①良い ②悪い

目的変数

2. 年齢
①10～30代 ②40～60代
③70～90代

3. 交通の満足度
①満足 ②やや満足 ③やや不満
④不満

4. 一人暮らしか？
①はい ②いいえ

モデル式

$$y = (0.231x_{11} + 0.913x_{12} - 0.872x_{13}) + (1.013x_{21} + 0.447x_{22} - 0.012x_{23} - 0.526x_{24}) + (0.202x_{31} - 0.149x_{32})$$

	説明変数	年齢			交通の満足度				一人暮らしか？	
	カテゴリ係数	0.231	0.913	-0.872	1.013	0.447	0.012	-0.526	0.202	-0.149
	群	10～30	40～60	70～90	満足	やや満足	やや不満	不満	はい	いいえ
住み心地良い	1	0	1	0	1	0	0	0	1	0
	1	1	0	0	0	1	0	0	0	1
	1	0	1	0	1	0	0	0	0	1
	1	0	1	0	0	1	0	0	0	1
住み心地悪い	2	1	0	0	0	1	0	1	1	0
	2	0	0	1	0	0	1	0	0	1
	2	0	0	1	0	0	0	1	1	0

3. 新住民価値導出のための特徴量の選択

(1)数量化Ⅱ類

相関比(「住みたい」「住みたくない」の2群が離れている度合)が、**最大**になるように**各特徴量(アンケート項目)の重み(モデル式のカテゴリ係数)**を算出。
→「住みたい」「住みたくない」を判別するための寄与率(カテゴリ係数)
以下はモデル式からレンジを算出する例

目的変数 y ＝ 説明変数:年齢 $(0.231x_{11} + 0.913x_{12} - 0.872x_{13})$ ＋ 説明変数:交通満足度 $(1.013x_{21} + 0.447x_{22} - 0.012x_{23} - 0.526x_{24})$ ＋ 説明変数:一人暮らしか $(0.202x_{31} - 0.149x_{32})$

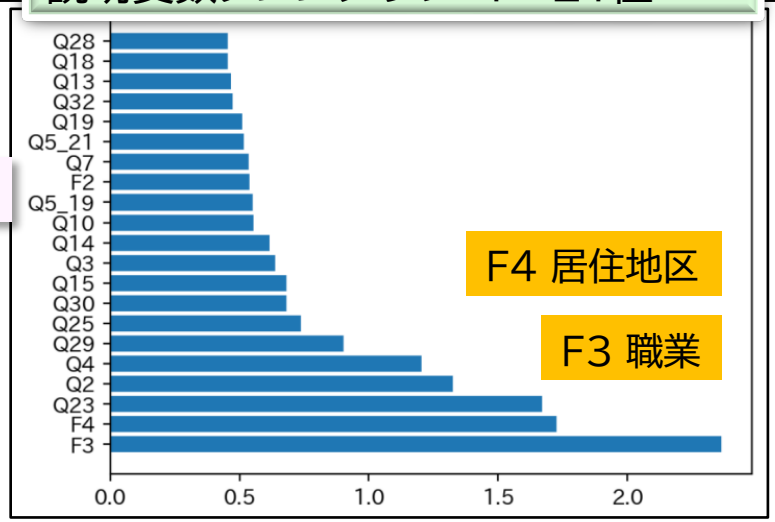
最大値 ＝ 0.913
最小値 ＝ −0.873
レンジ ＝ 0.913−(0.873)
＝ 目的変数に対する年齢の重要度

数量化Ⅱ類結果から特徴量を選択

◆特徴量を選択するため、住み心地(目的変数)に対して重要な説明変数の選択をする。

◆そこで、説明変数内において、カテゴリ係数の最大値と最小値の差で与えられる**レンジ**を重要度を示す尺度として利用し、レンジ0.15以下である12個の説明変数を削除

説明変数レンジグラフ　1~21位



ランキング詳細は29ページ

結果評価

◆職業に関して「農業」のカテゴリ係数の絶対値が最も大きい。

◆レンジランキング2位の居住地区は地区によって
「全員が住み心地が良い」「10%が住み心地が悪い」などと意識が異なる。

◆3位は訪島者からの協力金で、アンケート回答の自由記述を眺めたところ、世界文化遺産に認定された環境を島としてどう利用するかが、住民の関心事の1部であることが認識できた。

3. 新住民価値導出のための特徴量の選択

(2)相関分析(クラメール連関係数)

・多重共線性(マルチコ)を排除するために、特徴量間の相関を調べる。
・アンケートデータの特徴量は全て質的変数なので、相関係数を用いることはできない。
・クラメール連関係数という質的データと質的データの相関を求める尺度を利用する。

・クラメール連関係数はクロス集計表の関連性の度合いから算出される。

相関分析(クラメール連関係数)結果から特徴量を選択

◆**クラメール連関係数が0.5以上のもの**をマルチコの変量と認識する[3]

◆集約した変量は9変量（他アンケート項目と相関が強い項目を削除、代替変数があるため）

アンケート#	内容サマリ
Q5-1-4	水道・下水道の整備
Q5-1-7	がけ崩れや危険箇所対策
Q5-1-19	小学校・中学校の教育内容
Q5-1-21	生涯スポーツ
F 職業複数	F自営業-農業(複数回答型)

アンケート#	内容サマリ
Q5-1-25	歴史・文化や自然景観など、町の資源活用
Q16-1	野崎島は「世界文化遺産」に登録、小値賀町域の一部は国の重要文化的景観に選定。地域の景観を守るためにどんな協力ができるか？
Q23－1	問22で訪島する方から協力金、税収を設定する場合に、1名あたり徴収する妥当な金額
Q27	あなたは配食サービスを利用していますか

3. 新住民価値導出のための特徴量の選択

(3)自由記述回答からの特徴語抽出

テキストマイニングの目的

◆アンケート項目回答だけでなく、自由記述回答も含めた特徴量から新住民価値を掘り起こしたい。

◆自由記述回答から特徴量として特徴的な語(特徴語)を抽出するためにテキストマイニングを行う。

テキストマイニングの手法

◆自由記述回答から特徴語をマイニングするため、文書群において単語がどのくらい特徴的かを表す指標である**TF-IDF**[4]を利用する。

◆文書群の中で特徴的な語ほど、**TF-IDF**の値が高い。

TF-IDF
TF : Term Frequency 単語頻度
IDF : Inverse Document Frequency 逆文書頻度

$$\mathrm{TF} = \frac{\text{単語}t\text{の出現回数}}{\text{文書内の総単語数}} \qquad \mathrm{IDF} = \log\frac{\text{総文書数}}{\text{単語}t\text{を含む文書の数}}$$

$$\textbf{TF-IDF} = \mathrm{TF} \times \mathrm{IDF} = \text{単語使用頻度} \times \text{単語レア度}$$

3. 新住民価値導出のための特徴量の選択

(3)自由記述回答からの特徴語抽出

※KH Coderとはテキストマイニングのためのフリーソフト

テキストマイニングの具体的手法(例)

回答者	文書群
1	文書1
2	文書2
3	文書3
4	文書4

※KH Coder を使って形態素解析

回答者	単語数	単語1	単語2	単語3
1	4	3	1	0
2	21	10	2	9
3	13	0	1	12
4	20	4	10	6

TF-IDF算出

単語1TF-IDF	単語2TF-IDF	単語3TF-IDF
0.21576	0	0
0.13699	0	0.12329
0	0	0.26555
0.05753	0	0.08631

[TF-IDFの算出例]

・回答者2が使用した単語1のTF-IDFを求める。

総文書数 = 4

回答者2の単語1の使用数 = 10

回答者2の使用単語数 = 21

単語1を含む文書数 = 3

$$TF\text{-}IDF = TF \times IDF$$

$$= \frac{\text{回答者}_2\text{単語}_1\text{の使用数}}{\text{回答者}_2\text{の使用単語数}} \times \log\left(\frac{\text{総文書数}}{\text{単語}_1\text{を含む文書数}}\right)$$

$$= 10/21 \times \log(4/3) = 0.13699$$

3. 新住民価値導出のための特徴量の選択

出現単語ごとのTF-IDF一覧

高血圧	回答	ない	円	タクシー	糖尿
0	0	0	0	0	0
0	0	0	0	0	0
0	0	0	0	0	0
0	0	0	0	0	0
0	0	0	0	0	0
0	0	0	0	0	0
0	0	0	0	0	0
0	0	0	0	0	0
0	0	0	0	0	0
0	0	0	0	0	0
0	0	0	0	0	0
0	0	0	0	0.310235	0
0	0	0	0	0	0
0	0	0	0	0	0
0	0	0	0	0	0
0	0	0	0	0	0
0	0	0	0	0	0
0	0	0	0	0	0
0	0	0	0	0	0
0	0	0	0	0	0
0	0	0	0	0	0
0	0	0	0	0	0

TF-IDF値分散ランキング

順位	出現単語	tfidf値の分散
1	高血圧	0.03480
2	小児科	0.02343
3	賃金	0.01849
4	人材	0.01532
5	カロリー	0.01316
6	回答	0.01295
7	円	0.00813
8	糖尿	0.00743
9	制限	0.00582
10	塩	0.00564

TF-IDF値の分散値が高い特徴語を含む文書内容

#	特徴語	質問#	件数	回答内容の概要
1	高血圧	Q31-2	8	医師や栄養士から高血圧の指導
2	小児科	Q5-2他	3	診療所に小児科が欲しい
3	賃金	Q33他	10	低賃金、生活不安、所得底上げ、人材との関係
4	人材	Q33他	13	基幹産業、観光産業の振興、雇用促進、人材育成
5	カロリー	Q31-2	2	カロリー計画、制限

・絞り込み検索結果のランキングの特徴語の総合的な分析が必要
・特徴語の選択基準として、(TF-IDFの分散値ランキング50)＋(TF-IDFの合計値ランキング50)の結果から、重複排除

76語を特徴語として抽出

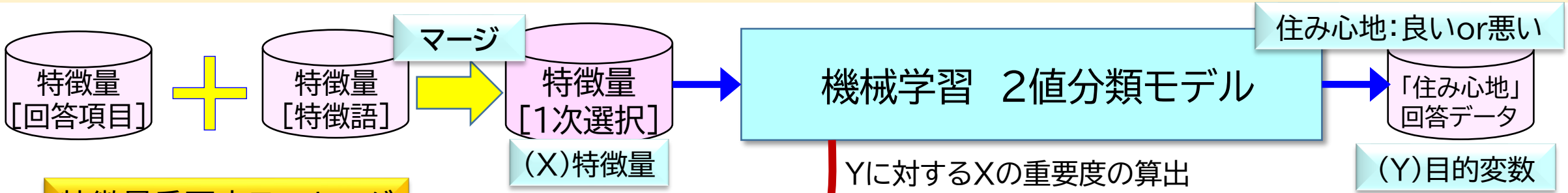
ない.1	する	高血圧	回答	ない	円	タクシー	思う	糖尿	ある	塩	わかる	小児科	カロリー	減	なる	ほしい.1	
町	高齢	できる	いる	行く	子供	小値賀	人	増加	もっと	子ども	多い	前	人口	参加	制限	観光	
出来る	施設	町民	分かる	医療	行う	賃金	人材	進学	計画	減	保全	ナイ	流れ	中途半端			
しかた	DM	任せる	なんとも	環境	特に	身体	食事	IT	悪化	すべて	転勤	企業	誘致	葬儀	予備	納税	
ふるさと	船	用事	まかなう	クラウドファンディング				住宅	血糖	整備	訪れる	抜	余裕				

単語ごとにTF-IDF値の分散を算出しランキング。
平成30年度アンケートの全回答者(1738名)
ランキングの順位は、
全回答者の全文書の中で、単語のTF-IDF値の分散値及び合計値を指標とした。(上記は分散値が高いランキングを例として示す)

[特徴語を含む文書の傾向分析]
・健康/医療に関する懸念・不安　・人口増加/減少に関する懸念
・生活インフラ・交通　・賃金と生活不安　・産業・雇用促進・人材育成

[分析上の課題]
・「全体」「住みたい」「住みたくない」グループ比較
・TF-IDF値が高い回答者の回答深耕　・共起度分析による気づき
・ストップワードを活用した特徴語の絞り込み　・深層学習による感情分析など

4.機械学習モデルの適用と特徴量の評価



特徴量重要度ランキング

記号	アンケート内容	選択回答
Q2_1	住み続けたさ	#1住み続けたい
F4	居住地区	無記載
Q3-1-4	住み続けたい理由	#4土地と家あり
Q3-1-6	住み続けたい理由	#6自然環境良い
F2	年代	無記載
Q3-1-2	住み続けたい理由	#2ずっと住んでいる
Q2_4	住み続けたさ	#4どちらとも言えない
Q5-1-26_1	日常買い物利便性	#1満足していない
Q3-1-1	住み続けたい理由	#1町に愛着あり
Q2_5	住み続けたさ	#5未回答
Q5-1-1_4	本土との海上交通	#4未回答
Q7-1-1_23	町の課題	#23未回答
Q3-1-14	住み続けたい理由	#14近所付き合いに満足
Q2_2	住み続けたさ	#2将来他に移りたい
Q6-1_5	定住人口問題	#5未回答
Q4-1_5	将来他に移りたい	#5買い物通院に不便
Q5-1-28_1	働く場の確保	#1満足していない
Q26-3	1人暮らしか	#3未回答

機械学習モデルによる特徴量評価の目的

◆選択回答項目と自由記述回答をマージしたデータから統合的に特徴量を評価することを目的とする。

手法・評価

◆住み心地を分類する上で重要度が高かった特徴量を算出し、降順にランキング。

◆ランキングから重要度0.004以上の49個の特徴量を主成分分析に入力する特徴量として選択した。

◆生活基盤(交通、経済的基盤)の充実が住み心地に大きく影響。

◆居住地区・年代も住み心地に影響あり。居住地区・年代を軸にした掘り下げが課題。

主成分分析(1) 方法

指針 ： 主成分を潜在的な新住民価値として、主成分分析を行う。

主成分分析の
一般的な説明はp32

◆主成分分析に入力する特徴量の最適化

・前プロセスでの機械学習(ランダムフォレスト)にて、Q1に対する影響度が0.004以上の特徴量(49個)を選択。

・主成分を意味付けする際に活用する「主成分負荷量」の精度を考慮し、曖昧な意味付けとなる「未回答」が含まれるカラム(780)を削除。

◆主成分分析の実行

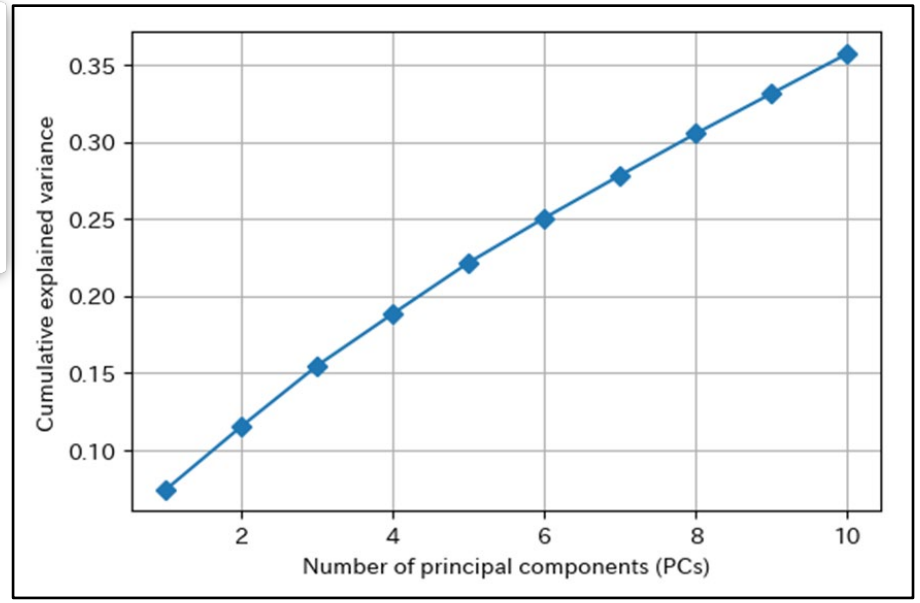
・F2 から Q26 までのフィールドの値を要素とする 58×958 の行列を標準化し、第10主成分まで求めた。

・pca1(第1主成分)及びpca2(第2主成分)を使って新変量を意味付け。

・第1～第4主成分での累積寄与率は20%程度。

主成分	寄与率
PC1	0.074565
PC2	0.041046
PC3	0.039408
PC4	0.033555
PC5	0.033219

累積寄与率グラフ
縦軸:累積寄与率
横軸:主成分数



寄与率の定義式
(λは第i主成分に対応する分散)

$$\frac{\lambda_k}{\sum_{i=1}^{d}\lambda_i} = \frac{\lambda_k}{\lambda_1+\lambda_2+\cdots+\lambda_d}$$

主成分分析(2) 結果概要と主成分負荷量

主成分分析でPC(主成分)導出

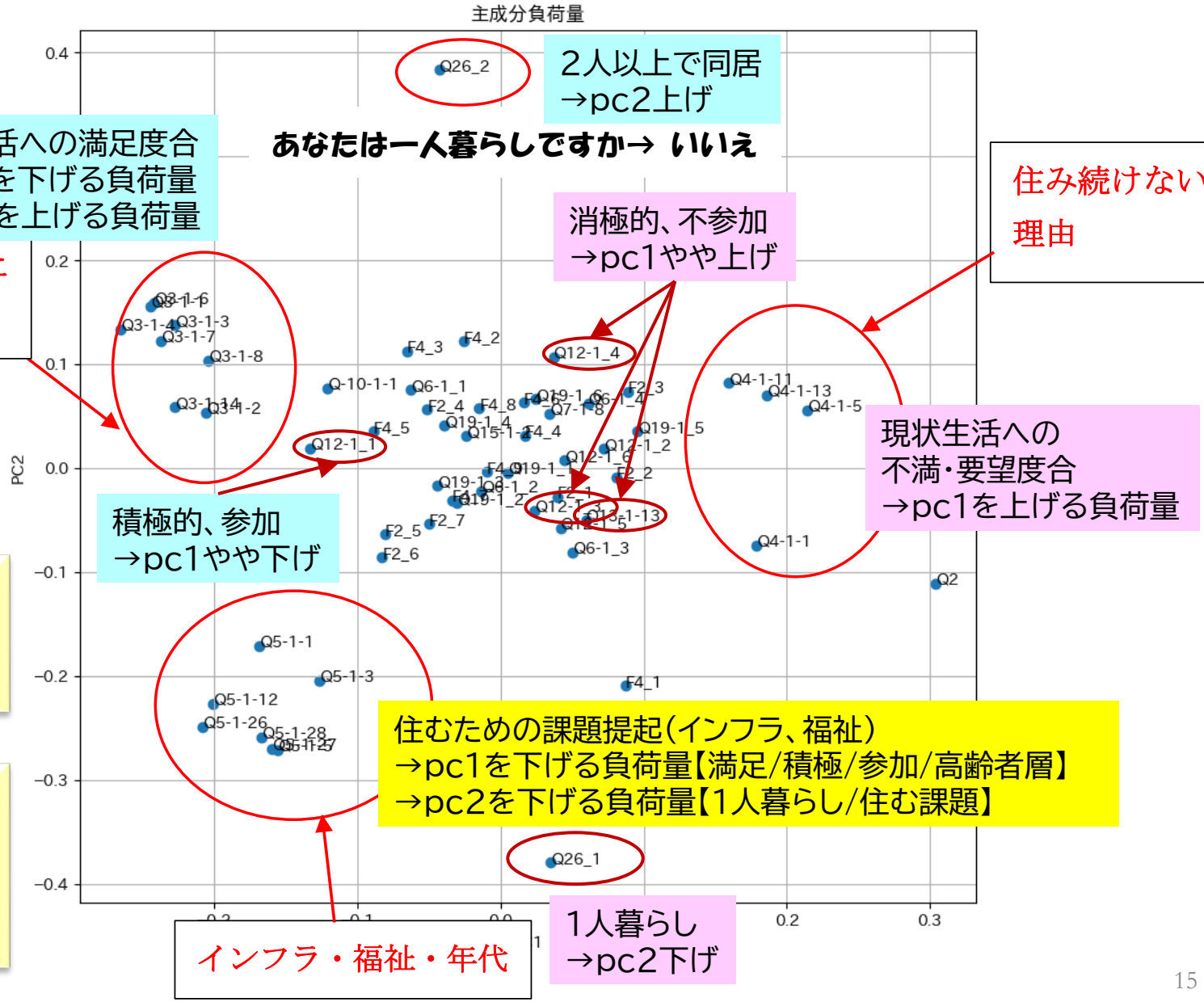
選択した特徴量(アンケート回答項目)の負荷量

$=Q_1, Q_2, Q_3, \ldots..,Q_n$

PC(主成分)$=(Q_1, Q_2, Q_3, \ldots..,Q_n)$

主成分1(pc1)：**未来改善(不)志向度**
(大)現状環境に不満、やや消極的、不参加
(小)現状環境に満足、やや積極的、参加意識

主成分2(pc2)：**家族居住/居住継続意向**
(大)2人以上居住、住み続けたい
(小)1人暮らし、
住むための課題提起(インフラ、福祉)

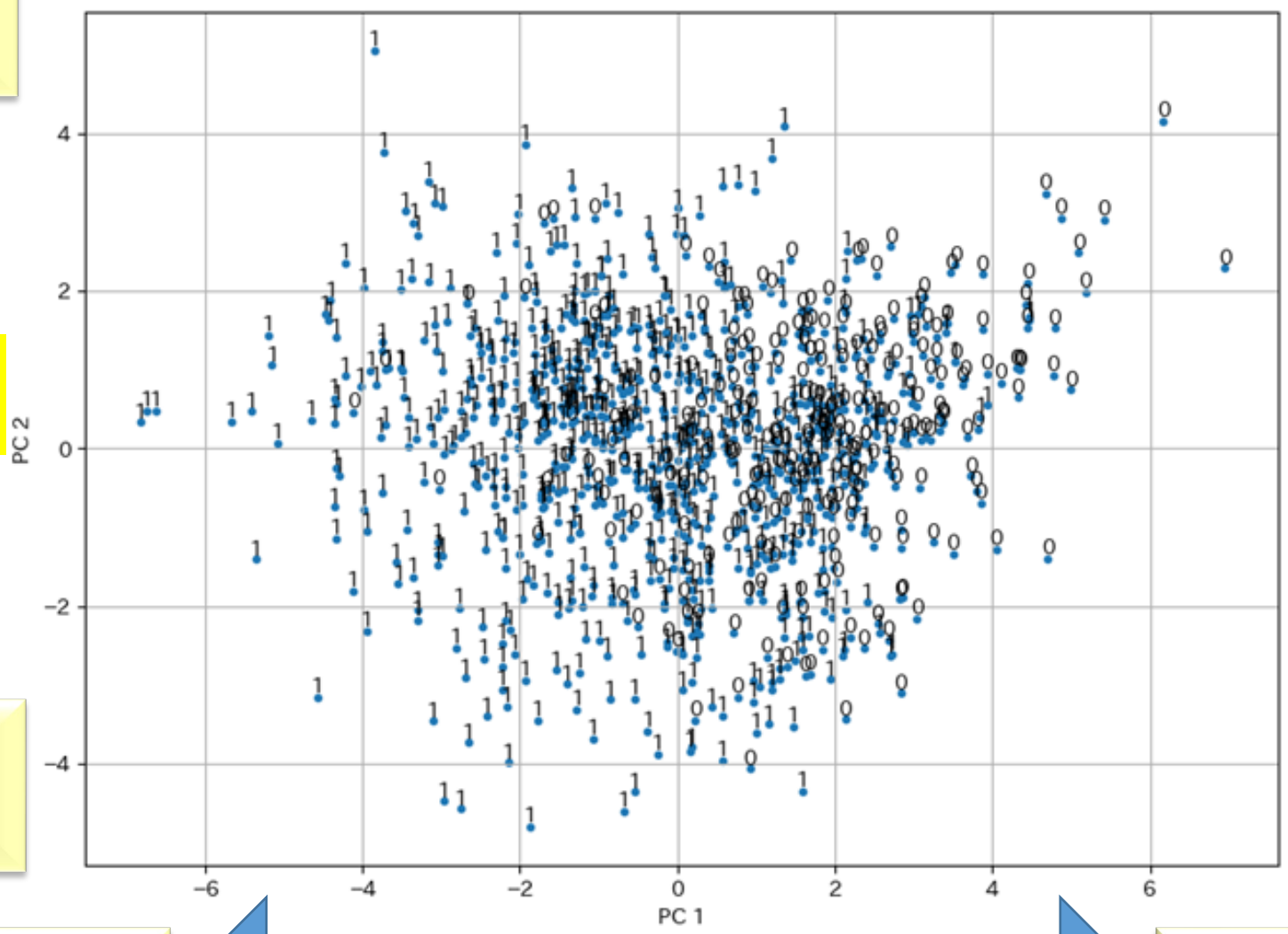


主成分分析(3) 主成分得点

2人以上で居住
住み続けたい

家族居住(同居)/
居住継続意向度

1人暮らし、
住むための課題
(インフラ、福祉)



主成分得点

主成分に対して
(1)Q1「住みごごち」
◎「1」住みごごちが良い人は
左のエリア
→住みごごちが良い人は
積極的、参加
◎「0」住みごごちが良くない人は
右のエリア
→住みごごちが良くない人は
消極的、不参加

(2)pc1,pc2に対して総合的な
一様な分布

現状環境に満足
やや積極的、参加

未来改善(不)志向度

現状環境に不満、
やや消極的、不参加

主成分分析(4) 主成分内容(pc1/pc2)

導出した主成分(pc1,pc2)の内容について

		pc1	pc2
合成変量の意味		未来改善(不)志向度	家族居住/居住継続意向
合成変量の傾向	主成分負荷量(正)	・現状の生活環境不満 ・まちづくり参加意欲小 ・やや消極的 ・人口・就業・文化遺産活用に一部課題提起 ・地域特性 (笛吹郷、中村郷、斑島郷) ・年代特性 (40～50代、10代～30代)	・2人以上で居住 ・住み続けたい意向が強い ・地域特性 (前方郷、柳郷) ・年代特性 (40～50歳代)
	主成分負荷量(負)	・現状の生活環境満足 ・まちづくり参加意欲大 ・積極性 ・生活インフラ/環境、コミュニティ改善意向 ・未来の環境に対しての見解意識 (コミュニティ、まちづくり、産業振興、文化遺産利活用、人口問題) ・地域特性 (浜津郷、柳郷、前方郷、黒島郷、大島郷、納島郷) ・年代特性 (60代～90代)	・1人暮らし ・居住環境への課題提起多い(インフラ、福祉) ・地域特性(斑島郷、大島郷)

6. 今後の3つの課題

課題(1)：主成分分析における第2主成分までの累積寄与率が11.6％であり、新住民価値である主成分の寄与率が低いことである。

残り2つの課題(2), (3)は次ページ

課題(1)主成分の寄与率向上のための改善案

◆数量化Ⅱ類で説明変数を選択するためのレンジ閾値をチューニングする。

◆テキストマイニングで抽出した特徴語データをダミー変数化せずにそのままTF-IDF値を使用する。

◆相関分析では、相関が高い2変数を無作為に片方削除する方法ではなく、目的変数との相関が強い変数の方を削除する。

◆機械学習モデルに入力するデータの分割法をホールドアウト法ではなく、交差検証法を用いる。

◆機械学習モデルを手動で選び、複数のモデルで特徴量選択を行い、主成分分析の結果を比較する。

6. 今後の3つの課題

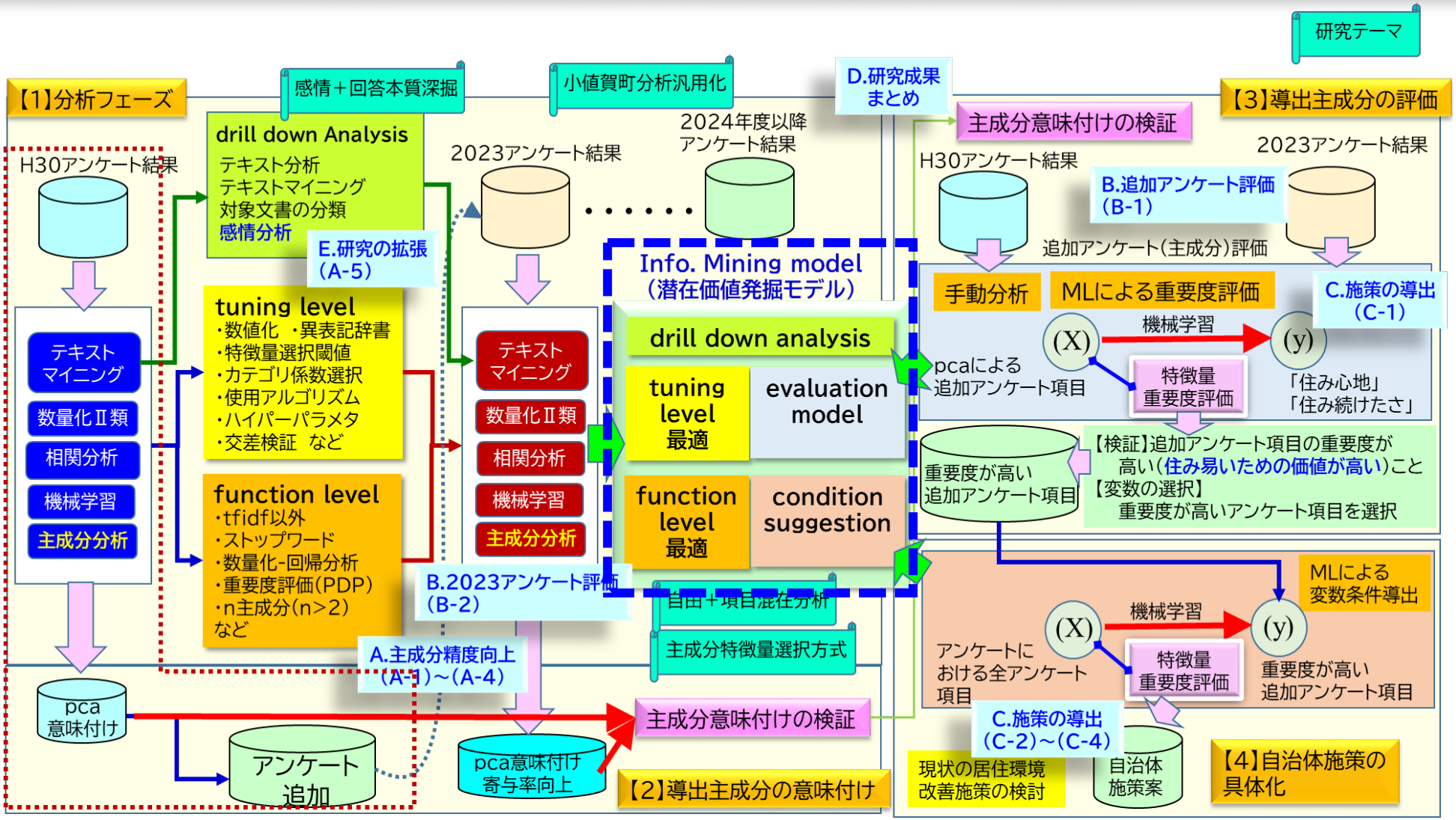
・課題(2)：主成分分析で発掘した潜在的テーマが「新住民価値」として重要度が高いかどうかを追加アンケートで検証する(下記(b))
・課題(3)：発掘した潜在的テーマに関して、住民意識を高めるための居住環境条件の改善提案(下記(C))

課題(2),(3)の概要

目的	研究プロセス		概要
(1)新住民価値の明確化	(a)潜在的テーマの発掘	(a-1)特徴量選択	アンケート結果から潜在的な住民価値発掘の「ための主成分分析の精度向上のための特徴量選択。 数量化Ⅱ類、相関分析、テキストマイニング、機械学習
		(a-2)新住民価値の概念発掘	主成分の意味付け(主成分分析)とアンケート追加
	(b)潜在的テーマの評価		追加アンケート結果の分析 (追加アンケート結果を(説明変数(X))とした住み易い(目的変数(y))に対する重要度)
(2)新住民価値創出のための施策提案	(c)潜在的テーマに対して住民意識を高める居住環境条件の明確化		追加アンケート結果の住民意識を高めるような、居住環境条件を導出する (追加アンケート結果を(目的変数(y))とし他アンケート回答を(説明変数(X))とした場合の(X)の重要度評価)

7. 今後の方針

今後シビックプライド醸成のための「潜在的な住民価値発掘モデル」(Info.Minning model)の汎用化を探求する



Info.Mining model の構想案

(1)tuning/function level での導出主成分の精度向上 (【2】導出主成分の意味付け)

(2)導出主成分の評価プロセス (【3】導出主成分の評価)

(3)導出主成分が住民価値になるための施策導出 (【4】自治体施策の具体化)

(4)drill down analysis テキストマイニング、感情分析などを活用してアンケート文書の真意を深掘りする

参考文献

[1] 読売広告都市生活研究局(著):シビックプライド-都市のコミュニケーションをデザインする

[2] Hotelling, H. : “Analysis of a complex of statistical variables into principal components”. *Journal of Educational Psychology* **24**: 417–441, 498–520,(1933).

[3] 菅 民郎, 藤越 康祝(著): 質的データの判別分析 数量化Ⅱ類.

[4] 佐藤浩輔：島根大学人間科学部2019.07.13,応用心理学研究Ⅰ,テキストマイニング講義資料,https://www.slideshare.net/cos039840935/ss-155407947.

[5] scikit-learn公式,sklearn.ensemble.RandomForestClassifier: https://scikit-learn.org/stable/modules/generated/sklearn.ensemble.RandomForestClassifier.html

https://www.soumu.go.jp/main_content/000629037.pdf

既存研究との関連性について -研究背景と課題

【研究の背景】
　各自治体において、シビックプライドの改善の取り組みが活発になっている昨今、現在から未来へ、新しい住民価値を創出する各種検討が推進されている。取組の殆ど、住民アンケートをベースに地域の特性（衣食住、医療、福祉、生活基盤、産業など）を把握して、各自治体の担当者が回答データを分析して施策を企画・立案する実情が多い。

【研究の対象となる課題】
　居住に関する既存の住民アンケート（アンケート項目選択＋自由記述の形式）から、シビックプライドに繋がる住民意識・価値を高めるための潜在的なテーマを導出する、分析・対策プロセスが確立できていないという課題を認識する。

本研究では、既存の住民アンケート結果（アンケート項目＋自由記述の形式）から、潜在的な新しい「住民価値を高める」テーマを導出し、そのテーマが住みたい意識をどう改善でき、意識改善するための具体的な居住環境の改善とは何なのかを、導出・評価するプロセスを明確にする。また、そのプロセスの汎用化に繋げる研究活動とする。

（a）選択回答アンケート項目分析

住民アンケートをベースに傾向分析を実施し、傾向把握のための分類などを目的とした、主成分分析を活用し導出した新しい軸を使ったデータ解析などの例がある。（2）（3）

選択式アンケート回答結果の多変量解析方法において、潜在的な住民価値と施策を導くような分析手法は明らかにされていない

（b）選択回答＋自由回答分析

選択式アンケート項目と自由記述回答の関連性を評価するテキストマイニング事例がある。（4）（5）（6）

選択式アンケートと自由回答アンケートの両方を重みづけ評価を行い、アンケート全体から潜在的な住民価値を導く分析手法は明らかにされていない

（c）シビックプライドを高める活動

都市環境に対する価値の住民アンケートからの因子分析などによる地域に対する意識を明確化するなどの例がある。（6）

自治体既存のアンケートを使い、居住環境全般にわたる、住民価値向上の施策を導出するための潜在価値を発見する手法は明らかでない

既存研究との関連性について -類似研究の調査(1/3)

【(a)選択回答アンケート項目分析】

(1) 江崎雄治:居住環境から見た住民の価値意識, 地理学評論, 68A-3 168-179 1995,
https://www.jstage.jst.go.jp/article/grj1984a/68/3/68_3_168/_pdf/-char/ja .
<概要, 手法>居住環境の各要素に対して、住民が無意識のうちに与えている「重み」を抽出する。目的変数を居住環境の総合的な満足度、説明変数を個別要素に対する満足度として重回帰分析をし、得られた偏回帰係数を各要素に与えられた「重み」とする。
<本研究との差異>アンケートから重要度(重み)を抽出する観点は類似しているが、そこから「潜在的な住民価値」を導出する点が異なる。

(2) 加藤潤三:地域コミュニティに対する住民の価値を測定する―『コミュニティ価値』尺度の作成と検討― , 立命館産業社会論集, 55巻3号, 55 - 66, 2019-12
http://doi.org/10.34382/00012886.
<概要, 手法>住民が「コミュニティ価値として重視していると考えられる15の諸要素」を100点配分するアンケート調査(全国11ブロック人口比に応じた651名Web調査)から重要度判断。上記をクラスタ分析で5クラスタに分類、別アンケート項目の「コミュニティに対する態度(コミュニティ意識・コミュニティ感覚、および行動(住民参加)」との関連性(相関)、『コミュニティ価値』を量的に測定する尺度を作成。
要素項目⇔意識、感覚、行動回答項目との相関分析.
<本研究との差異>既存の自治体アンケートを元に住民価値を分析するところ(本研究)が差異。当該研究は、地域コミュニティ価値と意識に関する研究目的のアンケートを元に分析。自由回答とアンケート項目回答を含めた全体アンケート結果から潜在価値を発見する(本研究)ところが差異。

(3)ICTの活用実態と地域活性化との相関関係の把握への主成分分析の活用
https://www.soumu.go.jp/johotsusintokei//linkdata/other034_200803_hokoku.pdf
<概要, 手法> 自治体ごとのICT分野別活用指標(アンケート)から、ICTシステム間の関係の強さなどの構造を示す新たな指標を、主成分分析を用いて導出する(新指標:福祉軸、地域活性化軸)
→ICTシステムがどの分野に使われている傾向が強いかを、新しい軸を設定して傾向分析する
<本研究との差異>当該研究はICT活用実態に特化しているものの主成分分析を用いて新たな指標を導出しているアプローチは同じで、主成分分析で導出する価値の精度を高めるために、主成分分析の説明変数を選択している。
(機械学習)部分と、自由回答をアンケート項目回答を併せてアンケート結果の重要度を測っているところが差異。

既存研究との関連性について-類似研究の調査(2/3)

【(b)選択回答＋自由回答分析】

(4)テキストマイニングによる「市民の声」の分析
https://www.jstage.jst.go.jp/article/jichitaigaku/28/2/28_42/_pdf/-char/ja
https://www.jkk-labo.jp/wp-content/uploads/2016/11/textmining_shiryo.pdf

(5)自由記述データを用いたテキストマイニングによる都市のイメージ分析
https://www.jstage.jst.go.jp/article/jscejipm/68/5/68_I_315/_pdf/-char/en

(6)テキストマイニングを用いた自由記述データの有効活用に関する研究
http://library.jsce.or.jp/jsce/open/00039/201306_no47/pdf/398.pdf

〈概要・手法〉
・コレスポンデンス分析とクラスタ分析や共起度を活用し、アンケート項目と自由記述回答の関連性を評価するテキストマイニング事例(4)(5)(6)
・アンケート項目に関してコレスポンデンス分析で絞り込んだ課題に関連する「改善」に関する特徴的な語をキーにして課題を深掘りするアプローチ(4)
・コレスポンデンス分析＋クラスタ分析の結果と、共起度分析の結果により、テキストマイニングの結果導出した語の重要度の定量評価をするアプローチ(5)
・アンケート項目回答結果の分析プロセスにおいて、テキストマイニングで補完するアプローチ(6)

〈本研究との差異〉
アンケート項目と自由記述回答全体から、回答における重要度が高い特徴量（特徴語）を選択、主成分を導出することで潜在的な住民価値を見つけ出す手法を明らかにしようとしている部分が差異

既存研究との関連性について -類似研究の調査(3/3)

【(c)シビックプライドを高める活動】

シビックプライドの意義としては「地域プライド」として2005年に国土交通省と文部科学省，文化庁が共同まとめた
『地域プライド創発による地域づくりのあり方に関する調査』(平成18年3月)の中で定義されている
https://www.mlit.go.jp/kokudokeikaku/souhatu/h17seika/6pride/06_syu.pdf

(7) 143 都市環境はいかにシビックプライドを高めるか　-今治市を事例とした実証分析,公益社団法人 日本都市計画学会 都市計画論文集 Vol.52 No.3 2017年10月

〈概要・手法〉
アンケート調査の結果に対して、因子分析と共分散構造分析を用いて、シビックプライドの因子及び、都市環境の評価がシビックプライドに及ぼす影響構造を明らかにする。都市が構築した環境(xxx公園、xxx商店街など)に対する価値の住民アンケートからの因子分析による、住民の潜在意識から地域に対する気持ちの明確化を実施。各因子の各アンケート項目(地域愛着(選好)、地域愛着(感情)などのシビックプライド尺度)に対する因子負荷量評価で意味付けを実施。そこからの潜在意識の推定を実施。

〈本研究との差異〉
一部の都市環境に関して、地域に対する意識・感情についてのアンケート結果を因子分析して潜在意識の推定を行うアプローチは同様のものであるが、既存のアンケートから潜在的な価値をあぶり出し、その価値に対する意識の評価と意識を高めるための分析手法は明らかにされていない。本研究では、既存アンケートから主成分分析を用いて、住民価値に繋がる潜在的な住民価値をあぶり出し、評価を行って、当該価値に対する住民意識が高まるような居住条件を導出する手法を明らかにするものである。

3. 新住民価値導出のための特徴量の選択

特徴量選択の意義とプロセス

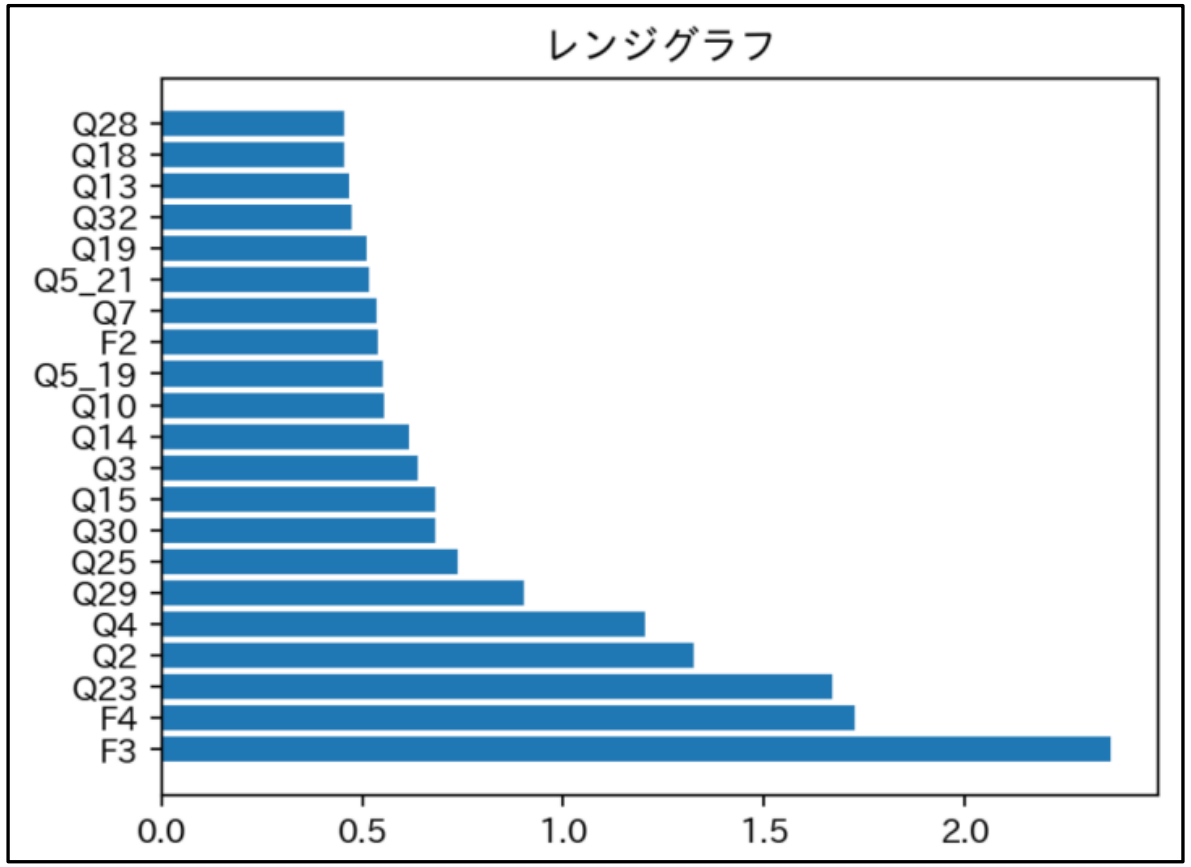
【特徴量選択の目的】
・新住民価値の導出方法として主成分分析を活用する。
・新住民価値はシビックプライド醸成に寄与するベクトルになることが必要なため、入力する説明変数（特徴量）は「住み心地」回答に影響度・重要度が高いアンケート項目および自由記述回答を選択する。

目的/プロセス	手法	手法の概要
アンケート項目回答の解析（特徴量選択）（多重共線性排除）	数量化Ⅱ類	「住み心地」の群を最も分離するアンケート項目（特徴量）の重み付け、特徴量の重み（「住み心地」に対する寄与率）評価、寄与率ランキングでアンケート回答項目を選択（カテゴリウェイト/レンジが0.15以上）
	相関分析	数量化Ⅱ類で選択した特徴量に対してクロス集計を行い、変数間のクラメール連関係数を評価
アンケート自由記述回答の解析（特徴語選択）	テキストマイニング	形態素解析[KH Coder]で、特徴語のTF-IDFを評価（コーパスごとに）、特徴語ごとのTF-IDF値ランキング50を選択。選択した特徴語を数量化（One-hot encoding）
主成分分析のための特徴量選択	機械学習（教師あり分類）	・数量化Ⅱ類で選択した特徴量とテキストマイニングで選択した特徴語をマージ。機械学習に入力する特徴量を生成。[1次選択] ・精度が高いアルゴリズムを評価選択（PyCalet）。 ・「住み心地」回答変数を目的変数、上記を説明変数とし機械学習モデルで学習/予測/精度評価。特徴量の重要度を評価し50個の特徴量を選択[2次選択]

3. 新住民価値導出のための特徴量の選択

(2)数量化Ⅱ類

説明変数レンジランキング　1～21位

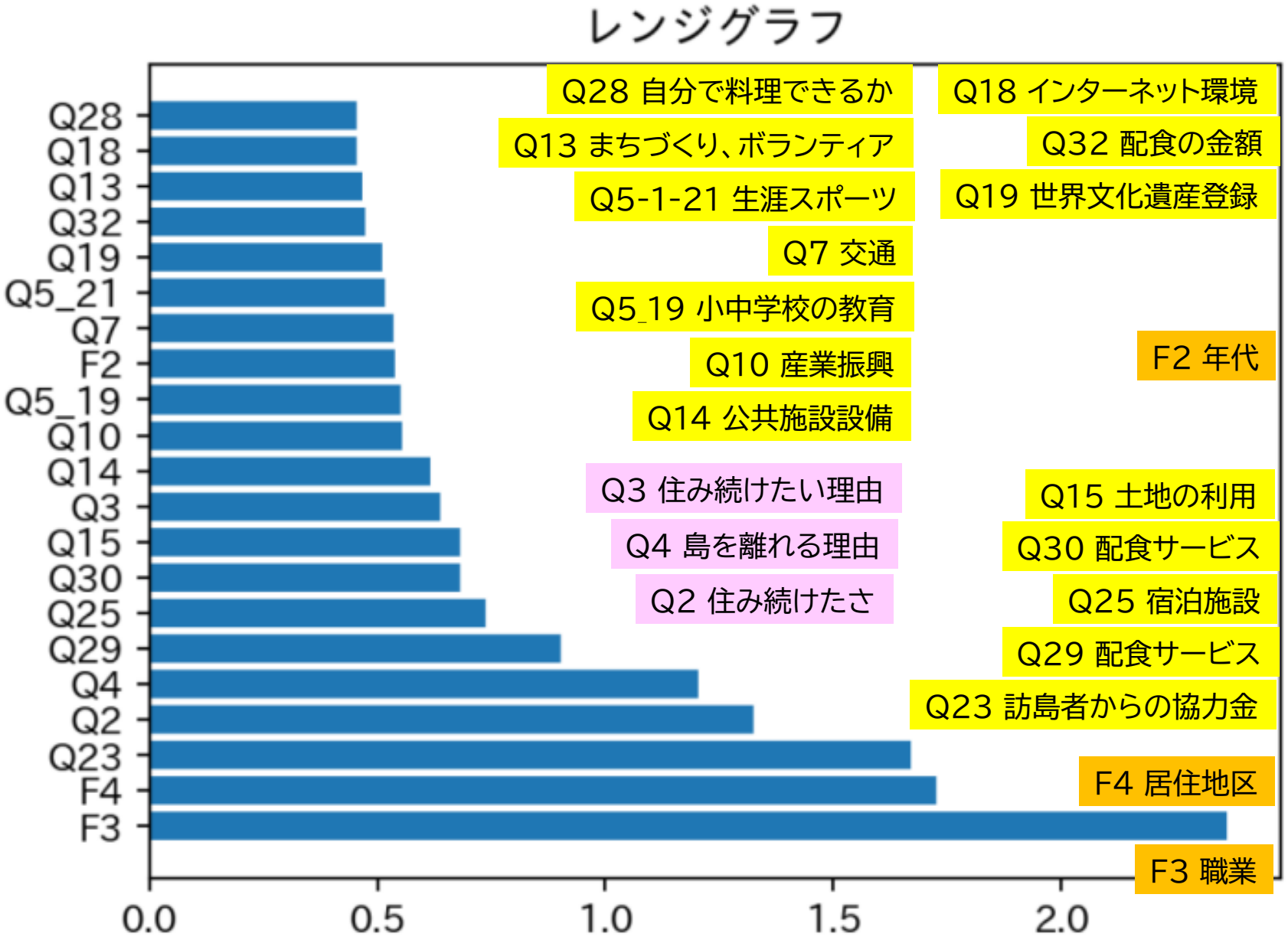


職業が最も重要度が高い

Q5-1-21 生涯スポーツ
Q7 交通
F2 年代
Q5_19 小中学校の教育
Q10 産業振興
Q14 公共施設設備
Q3 住み続けたい理由
Q15 土地の利用
Q30 配食サービス
Q25 宿泊施設
Q29 配食サービス
Q4 島を離れる理由
Q2 住み続けたさ
Q23 訪島者からの協力金
F4 居住地区
F3 職業

Q28 自分で料理できるか
Q18 インターネット環境
Q13 まちづくり、ボランティア
Q32 配食の金額
Q19 世界文化遺産登録

図2　数量化Ⅱ類レンジランキング



多変量解析手法の説明

主成分分析

・教師なし学習の一つ。ある多変量データからそのデータを要約してくれる潜在的なベクトル(主成分)を導出できる。
・主成分分析は、ベクトルにデータの分散を射影し、分散を最大化するようにそのベクトル(主成分)を決定する
・主成分は、データ全体の情報(分散)を豊富に表現されたベクトルであり、多変量データをそのベクトルで要約することができる。
・最も大きい分散が射影されたベクトルが第一主成分、次に大きい分散が射影されたベクトルが第二主成分である。

数量化Ⅱ類

・群データで与えられる目的変数と質的データで与えられる説明変数との関係をモデル式で表し、モデル式によって、説明変数と目的変数との関連性を明らかにする手法。
・これまで、アンケート調査の分析などに用いられてきた解析手法である。

クラメール連関係数

・質的データと質的データの相関を求める尺度。クラメール連関係数はクロス集計表の関連性の度合いから算出

TF-IDFの説明

TF-IDF

◆TF-IDFは自由記述回答から特徴語をマイニングするため、文書群において単語の特徴度合を表す指標である

◆文書群の中で特徴的な語ほど、TF-IDFの値が高い。

情報理論において確率pで生起する事象が起こったことを知ったときに得られる情報量をI(p)=-logpで表す。
IDFは情報量と同じ気持ちである。

IDFの場合すべての文書が同様に確からしいとは限らない。しかし、TF-IDFは理論的な基礎ははっきりしないが、有用なため、テキストマイニングや検索エンジンなどで幅広く使われている。

TF-IDF
TF : Term Frequency 単語頻度
それぞれの文書について、その単語が出てくる程度

IDF : Inverse Document Frequency 逆文書頻度
全体の文書のうち、その単語を含む文書の程度(の逆数)、複数の文書に出現する単語ほど特徴的でない

$$\mathrm{TF} = \frac{\text{単語}t\text{の出現回数}}{\text{文書内の総単語数}} \qquad \mathrm{IDF} = \log\frac{\text{総文書数}}{\text{単語}t\text{を含む文書の数}}$$

$$\text{TF-IDF} = \mathrm{TF} \times \mathrm{IDF} = \text{単語使用頻度} \times \text{単語レア度}$$

ここから検討資料

（1）相関分析の概要と相関の種類

引用:[8]

相関分析

・2変数間に関係性（相関）があるか、どのような関係性があるかを分析（**相関分析**）
・相関図（散布図）で見える化

<ある年のプロ野球一軍で50試合以上出場の選手データ>

選手No	打率	100m走	視力	血液型	甲子園出場経験
1	0.2253	12.0	1.11	B型	有り
2	0.1927	13.1	0.93	O型	無し
3	0.1644	12.8	0.93	AB型	無し
:	:	:	:	:	:
98	0.3459	12.4	1.34	A型	有り
99	0.1784	12.5	0.96	A型	無し
100	0.1888	12.4	0.81	AB型	有り
	量的データ	量的データ	量的データ	質的データ	質的データ

有り:1
無し: 0

量的データ

【分析の目的】（例）
・足が速い選手、視力が良い選手ほど、打率は良いか（100m走、視力と打率の関係）
・特定の血液型や甲子園出場経験者が、打率が良い傾向があるか。

【種類】
（1）（単）相関係数：
「量的データ」⇔「量的データ」の関係を明らかにする

（2）相関比：
「量的データ」⇔「質的データ」の関係を明らかにする

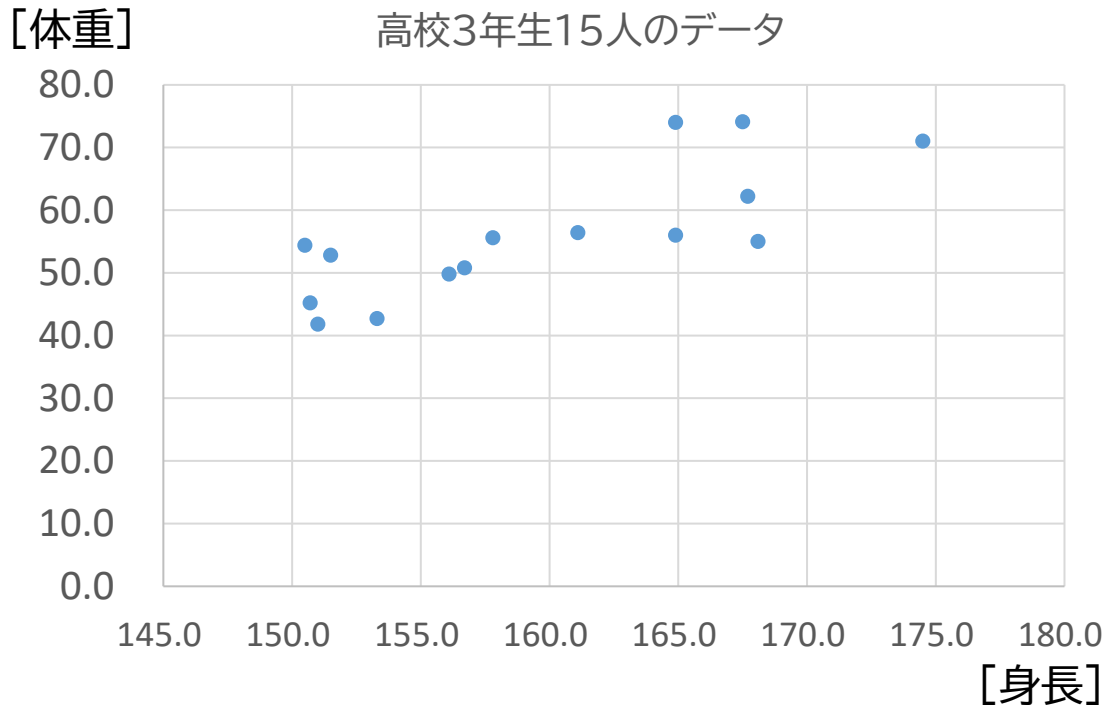
（3）クラメール連関係数：
「質的データ」⇔「質的データ」の関係を明らかに

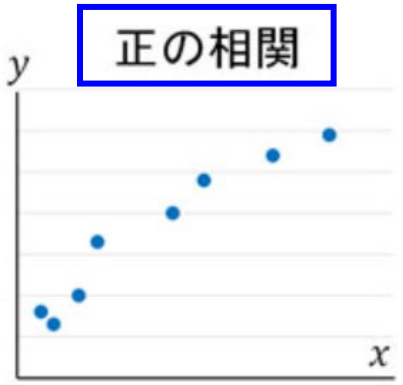
相関図（散布図）について

相関図（散布図）

・2変量の関係を視覚的に表現する。2変量をx軸、y軸にとりそれぞれのデータ点をプロット（複数の変量の資料があるとき、まず2変量相互の関係を調べる）

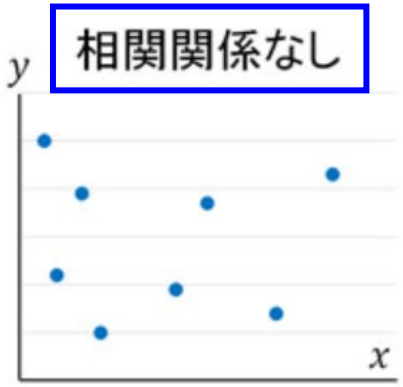
・2変数間に関係性（相関）があるか、どのような関係性があるかを分析（**相関分析**）





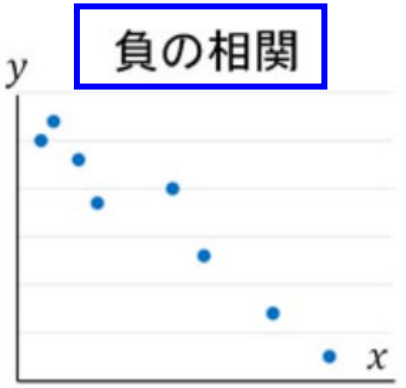
変量xが増加
→変量yも増加

（例）身長と体重



変量x、変量yに
特筆すべき
関係はない

（例）年収と体重



変量xが増加
→変量yは減少

（例）遊ぶ時間と
受験の合格率

（2）相関係数について

相関係数

ピアソンの
積率相関係数

$$r_{xy} = \frac{\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^{n}(x_i - \bar{x})^2}\sqrt{\sum_{i=1}^{n}(y_i - \bar{y})^2}}$$

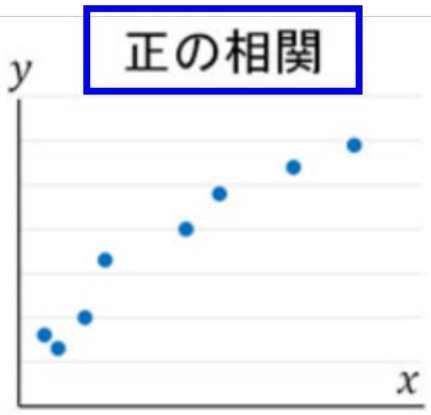
・共分散は各変数のスケールや単位によって値が変わる。スケールや単位の影響を受けずに、より客観的な相関関係の指標として、2つの変数の関係を数値化する。
・共分散をそれぞれ(x,y)の標準偏差の積で割った値

rxy: 相関係数
n: データの数
xi,yi: データの値
Sx,Sy: x,yの偏差平方和
$\bar{x}$, $\bar{y}$ ：xおよびyの平均値

$$r_{xy} = \frac{S_{xy}}{\sqrt{S_x}\sqrt{S_y}}$$

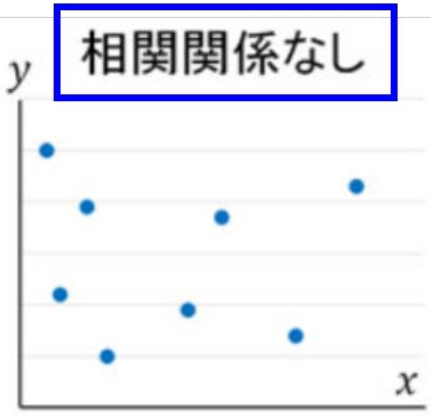
$-1 \leqq r_{xy} \leqq 1$

rxyの値は、
1に近いほど正の相関が強く
-1に近いほど負の相関が強い、
0に近い場合は無相関



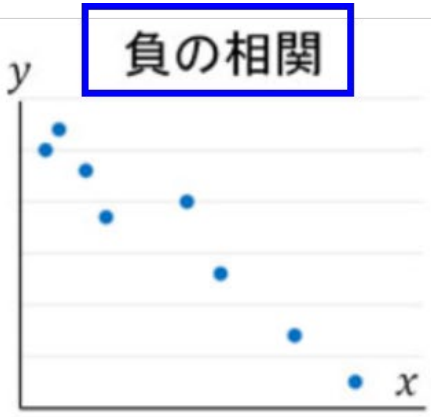
正の相関

rxy が1に近い



相関関係なし

rxy ≒ 0



負の相関

rxy が-1に近い

(例題3-1)
アイスクリームの支出と気温の相関係数
(corrcoef())

temp and Ice cream expenditures

(演習3-1) load_boston
RM(住居の平均部屋数)/MEDV(住宅価格の中央値)

[for exercises] correlation analysis for boston

(2)相関係数について　相関係数の求め方

(例)学生の身長と体重のデータ

学生	体重 xi	身長 yi	偏差 (xi-xave)	偏差 (yi-yave)	偏差平方 (xi-xave)^2	偏差平方 (yi-yave)^2	共分散 (xi-xave)(yi-yave)
A	45	146	-5	-4	25	16	20
B	46	145	-4	-5	16	25	20
C	47	147	-3	-3	9	9	9
D	49	149	-1	-1	1	1	1
E	48	151	-2	1	4	1	-2
F	51	149	1	-1	1	1	-1
G	52	151	2	1	4	1	2
H	53	154	3	4	9	16	12
I	54	153	4	3	16	9	12
J	55	155	5	5	25	25	25
平均	50	150	0	0	110	104	98

偏差平方和　　共分散

$$r_{xy} = \frac{\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^{n}(x_i - \bar{x})^2}\sqrt{\sum_{i=1}^{n}(y_i - \bar{y})^2}}$$

(例)
rxy=
x,yの共分散/√(xの偏差平方和×yの偏差平方和)

rxy = 98/(√110*√104)=0.91625

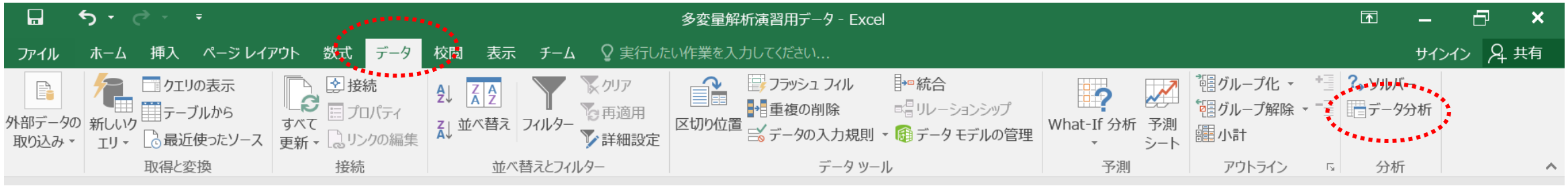
<例題3-2> 相関係数の求め方(Excel)

＜相関係数の評価＞強弱の違いがあるが、2変数間には大抵、相関関係がある。強い相関があるかどうかが重要。

r（相関係数）≧ 0.80	強い正の相関関係
0.50 ≦ r（相関係数）< 0.80	正の相関関係
0.30 ≦ r（相関係数）< 0.50	弱い正の相関関係
0 < r（相関係数）< 0.30	相関がない

r（相関係数）≦ -0.80	強い負の相関関係
-0.80 < r（相関係数）≦ -0.50	負の相関関係
-0.50 < r（相関係数）≦ -0.30	弱い負の相関関係
-0.30 < r（相関係数）< 0	相関がない

Excelのデータ分析について



＜データ分析のインストール＞
・「データ」リボンの中に「データ分析」のメニューがない場合、アドインのインストール作業が必要
・「ファイル」＞「オプション」＞「アドイン」＞データ分析を選択

【例題】

Excel 相関係数の求め方

(3)相関比

相関比

・「量的データ」と「質的データ」の関係を明らかにする統計手法
・判別分析（複数群に分けられた資料から群分けの明確な基準を決め、変量の関係を調べる）に必要な2群の離れ具合、各群のまとまり具合を示す指標。
・値は0～1の間をとり、1に近づくほど2変数は相関関係がある

＜ある企業の採用試験データ＞

不合格		合格	
番号	得点	番号	得点
1	80	11	98
2	82	12	96
3	84	13	90
4	82	14	89
5	87	15	86
6	84	16	95
7	91	17	95
8	85	18	92
9	81	19	88
10	85	20	94
人数	10	人数	10

・得点の中央付近で合格/不合格が混じっている
・「どのくらいの混じり具合なのか」を表す指標

＜2群P,Qを対象にした変量zについての資料＞

個体名	変量z	群	個体名	変量z	群
1	z1	P	m+1	z(m+1)	Q
2	z2	P	m+2	z(m+2)	Q
…	…	…	…	…	…
m	zm	P	n	zn	Q

群P：　個体番号1～mまでの（np個=m個）のデータが所属

群Q：　個体番号(m+1)～nまでの（nq個=n-m個）のデータが所属

全変動(ST)(偏差平方和)：　群間変動(SB)　群内変動(SW)

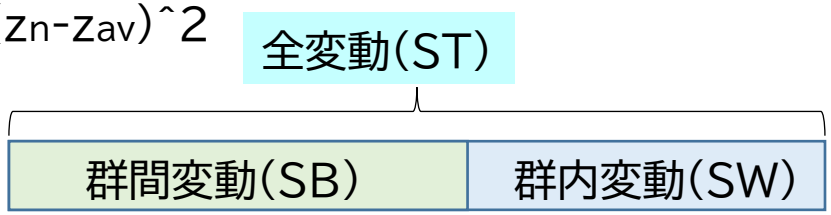
$S_T = (z_1-z_{av})^2+\cdots+(z_i-z_{av})^2+\cdots+(z_m-z_{av})^2 + (z_{m+1}-z_{av})^2+\cdots+(z_j-z_{av})^2+\cdots+(z_n-z_{av})^2$

z_{av}：　zの平均値
z_{pav}：Pに属するzの平均値
z_{qav}：Qに属するzの平均値

$ST = SB + SW$

$SB = np(z_{pav}-z_{av})^2 + nq(z_{qav}-z_{av})^2$

$SW = (z_1-z_{pav})^2+\cdots+(z_m-z_{pav})^2 + (z_{m+1}-z_{qav})^2 + \cdots + (z_n-z_{qav})^2$



(3)相関比

群間変動(SB)

群内変動(SW)

zav： zの平均値
zpav：Pに属するzの平均値
zqav：Qに属するzの平均値

ST(全変動) = SB(群間変動) + SW(群内変動)

SB = np(zpav-zav)^2 + nq(zqav-zav)^2

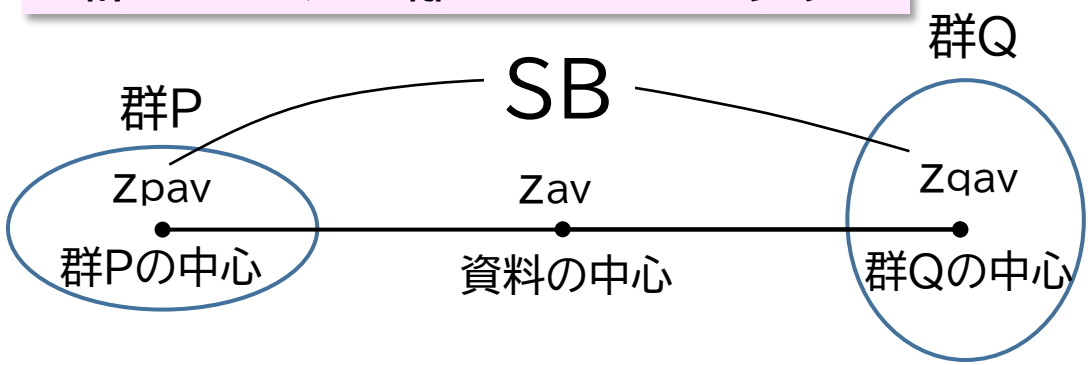
zpav-zav、zqav-zav：

群Pの平均値zpavと全体平均値zavの差
群Qの平均値zqavと全体平均値zavの差

np(zpav-zav)^2、nq(zqav-zav)^2：

群P全体がどれだけ資料の中心から離れているか
群Q全体がどれだけ資料の中心から離れているか

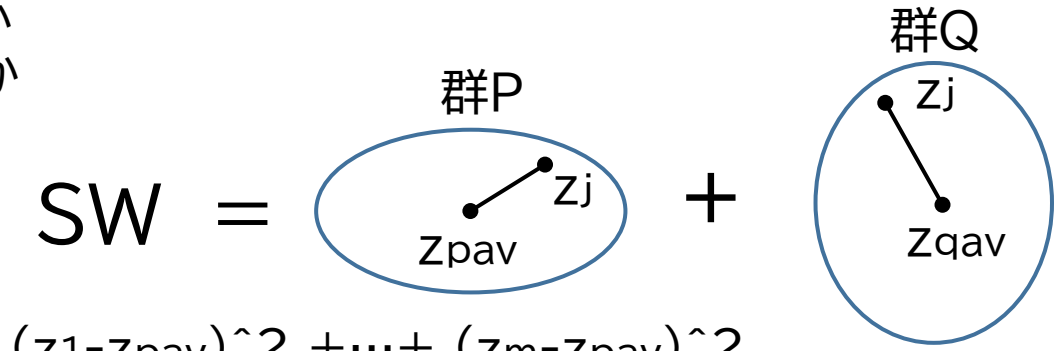
2群がどのくらい離れているかを表す



SW = (z1-zpav)^2 +…+ (zm-zpav)^2 + (zm+1-zqav)^2 + … + (zn-zqav)^2

群Pの中の変動：(z1-zpav)^2 +…+ (zm-zpav)^2
群Qの中の変動：(zm+1-zqav)^2 + … + (zn-zqav)^2

群内の「散らばり具合(まとまり具合)」を表す



(z1-zpav)^2 +…+ (zm-zpav)^2

(zm+1-zqav)^2 + … + (zn-zqav)^2

(3)相関比

相関比　全変動(ST)に占める群間変動の割合

$$\eta^2 = SB / ST \quad (0 \leqq \eta^2 \leqq 1)$$

$$\text{相関比}\eta^2 = \frac{\text{群間変動(SB)}}{\underbrace{\text{群間変動(SB)}\quad\text{群内変動(SW)}}_{\text{全変動(ST)}}}$$

SB：群間変動
SW：群内変動
ST：全変動

・相関比が1に近いとき：
　　STに占めるSBの比率が大きい　⇒　SWは小さい、各群は固まる：2群は分離
・相関比が0に近いとき：
　　STに占めるSBの比率が小さい　⇒　SWは大きい、各群は広がる：2群は重なる

＜相関比の評価＞0～1の間をとり、1に近づくほど、2変数は相関関係がある

η(相関比) ≧ 0.50	強い関連性がある
0.25 ≦ r (相関比) < 0.50	関連性がある
0.10 ≦ r (相関比) < 0.25	弱い関連性がある
0 < r (相関係数) < 0.10	関連性がない

Excel　実習26

(4)クラメール連関係数

クラメール連関係数

・「質的データ」と「質的データ」の関係を明らかにする統計手法
・クロス集計表から求めた期待度数と、クロス集計表の実測度数の一致度の合計値（χ２：カイ二乗）から算出

$$r = \sqrt{\frac{\chi^2}{n(k-1)}}$$

χ^2 ＝カイ2乗値
n ＝サンプルサイズ
k ＝カテゴリー数（2つのうち小さいほう）

χ2値を求める

<例:ある中学校のクリスマスプレゼントでゲーム機購入があるか>

	中学学年別	ゲーム機購入予定
A	1年	ある
B	1年	ある
C	1年	ない
D	2年	ある
E	2年	ある
F	2年	ある
G	2年	ない
H	3年	ある
I	3年	ない
J	3年	ない

<クロス集計表>

	ある	ない	横計
1年	2	1	3
2年	3	1	4
3年	1	2	3
縦計	6	4	10

実測度数

	ある	ない	横計
1年	6×3/10=1.8	4×3/10=1.2	3
2年	6×4/10=2.4	4×4/10=1.6	4
3年	6×3/10=1.8	4×3/10=1.2	3
縦計	6.0	4.0	10

期待度数

[該当する列の縦計]
×[該当する行の横計]
÷[全数]

(4)クラメール連関係数　χ^2値(カイ2乗値)から求める

クラメール連関係数

$$r = \sqrt{\frac{\chi^2}{n(k-1)}}$$

χ^2 =カイ2乗値
n =サンプルサイズ (全人数)　k =カテゴリー数(2つのうち小さいほう)

χ^2値を求める

実測度数

	ある	ない	横計
1年	2	1	3
2年	3	1	4
3年	1	2	3
縦計	6	4	10

$$\frac{(\text{実測度数}-\text{期待度数})^2}{\text{期待度数}}$$

期待度数

	ある	ない	横計
1年	1.8	1.2	3
2年	2.4	1.6	4
3年	1.8	1.2	3
縦計	6.0	4.0	10

	ある	ない
1年	0.022	0.033
2年	0.150	0.225
3年	0.356	0.533

$\sum$ 合計値 = 1.319

χ^2値 = 1.319

n(k-1) = 10*(2-1) = 10
クラメール連関係数 = 0.363

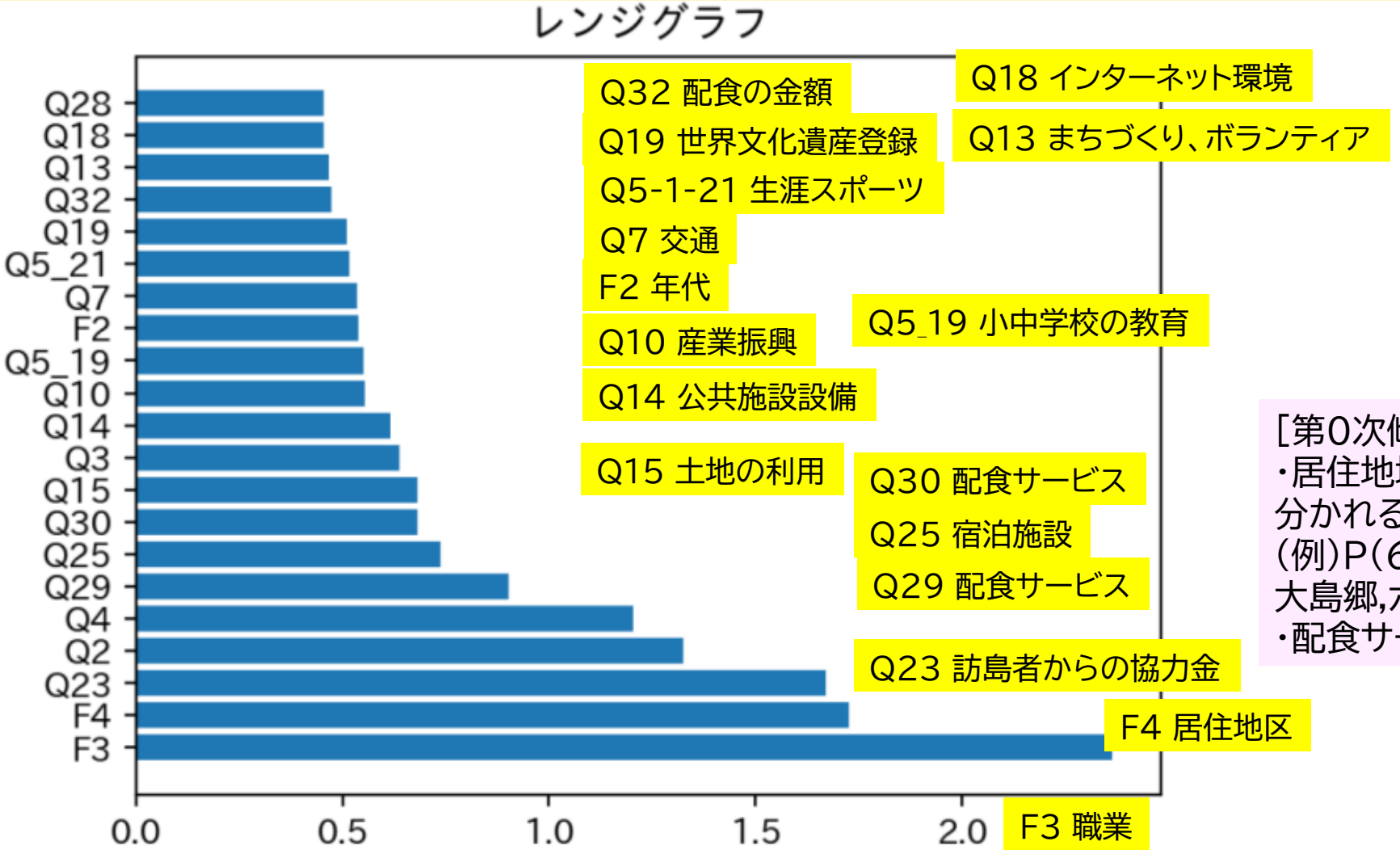
＜クラメール連関係数の目途＞

r ≧ 0.50	強い関連性がある
0.25 ≦ r ＜ 0.50	関連性がある
0.10 ≦ r ＜ 0.25	弱い関連性がある
0 ＜ r ＜ 0.10	関連性がない

Excel　クラメール連関係数

分析結果(0次) 数量化Ⅱ類

以下を分析対象とした、「住みたい」「住みたくない」回答に対する、影響度が大きいアンケート項目ランキング
・平成30年度アンケートの全回答者(1738名)
・ランキングの順位は、数量化Ⅱ類のカテゴリウェイトの値の昇順とした



[分析課題]
・影響度が「住みたい」「住みたくない」にどう影響しているかの分析
(サンプルスコアの傾向、カテゴリウェイトの値の評価)
・影響度が高い人のカテゴリの具体値の調査・分析
(住みたい人はどういう人か)

[第0次傾向分析]
・居住地域により「住みたい」「住みたくない」が分かれる。(Positive=P,Negative=N)
(例)P(68～81%一部除)、N(3～10%)
大島郷,六島郷(Nなし)、納島郷(P25%)
・配食サービス、宿泊施設などのインフラ周り

分析結果(0次) 数量化Ⅱ類 居住地区分析

F4値	内容	回答数	Q1住みごごち						数				比率（%）		
			1	2	3	4	5	6	P(1or2)	P以外	N(4or5)	全体	P(1or2)	P以外	N(4or5)
1	笛吹郷	805	170	393	139	61	18	24	563	242	79	805	69.9		9.8
2	前方郷	232	39	122	42	16	3	10	161	71	19	232	69.4		8.2
3	柳郷	104	17	55	20	6	0	6	72	32	6	104	69.2		5.8
4	中村郷	98	20	52	13	9	1	3	72	26	10	98	73.5		10.2
5	浜津郷	170	44	81	28	11	2	4	125	45	13	170	73.5		7.6
6	斑島郷	111	17	58	19	9	2	6	75	36	11	111	67.6		9.9
7	黒島郷	40	7	24	8	1	0	0	31	9	1	40	77.5		2.5
8	大島郷	47	14	24	9	0	0	0	38	9	0	47	80.9		0.0
9	納島郷	12	1	2	8	1	0	0	3	9	1	12	25.0		8.3
10	六島郷	3	1	2	0	0	0	0	3	0	0	3	100.0		0.0
11	未回答	116	4	9	9	5	2	87	13	103	7	116	11.2		6.0
合計		1738	334	822	295	119	28	140	1156	582	147	1738	66.5		8.5
★未回答除く			330	813	286	114	26	53	1143	479	140	1622	70.5		8.6

(1) 全体:
居住地区未回答含:P(1or2) 66.5%、N(4or5) 8.5%
居住地区未回答除く:P(1or2) 70.5%、N(4or5) 8.6%

(2) 地区別:Pベスト(大島郷)80.9、Pワースト(納島郷)25.0
Nベスト(黒島郷)2.5、Nワースト(中村郷)10.2

Q1. あなたは、小値賀町の住み心地についてどう思いますか？
1. とても住みよい 2. まぁまぁ住みよい 3. どちらとも言えない
4. やや住みにくい 5. とても住みにくい 6. 未回答

Pベスト	P回答比率(その郷全体回答数のうち、Q1の#1または#2の回答数の比率）のうち、最もP回答比率が高い郷
Pワースト	P回答比率(その郷全体回答数のうち、Q1の#1または#2の回答数の比率）のうち、最もP回答比率が低い郷
Nベスト	N回答比率(その郷全体回答数のうち、Q1の#4または#5の回答数の比率）のうち、**最もN回答比率が低い郷**
Nワースト	N回答比率(その郷全体回答数のうち、Q1の#4または#5の回答数の比率）のうち、**最もN回答比率が高い郷**

特徴量1次選択　数量化Ⅱ類/相関分析

・「住みたい」回答への寄与率が高い、その他アンケート項目回答（特徴量）の寄与率（カテゴリスコア）を算出 **[数量化Ⅱ類]**
→寄与率が低い特徴量を削減
・特徴量の説明力を上げ、予測精度を高めるため、マルチコ（多重共線性）の変量を集約 **[相関分析/クラメール連関係数]**
→クラメール連関係数が高い特徴量を集約

数量化Ⅱ類結果から特徴量を選択

◆**相関比**（「住みたい」「住みたくない」の2群が離れている度合）が、**最大**になるように
各特徴量（アンケート項目）の重み（寄与率：カテゴリウィト） を算出。
→「住みたい」「住みたくない」を判別するための寄与率（カテゴリウェイト）

◆寄与率ランキングでアンケート項目を選択：
特徴量選択：カテゴリウェイトのレンジ（max-minの差）が0.15以下を削除　［寄与率が低い］

◆削除したカテゴリ（寄与率が低いアンケート項目）：9項目

アンケート#	内容サマリ
Q5-1-2	島内交通（バス・渡船）
Q5-1-8	防火対策
Q5-1-9	街灯など、夜間歩行対策
Q5-1-15	こども園サービス
Q5-1-17	診療所の医療サービス

アンケート#	内容サマリ
Q17	自宅のインターネット利用状況及び今後の利用意向
Q20	あなたは野崎島を過去3年間の内に何回訪れたことがありますか
Q21-1	問20で「1　ない」と回答した方。野崎島を訪れたことがない理由は
Q24	野崎島が世界文化遺産登録となり、観光客の増加が見込まれる。宿泊環境の充実に関してどう考えるか。

特徴量1次選択　数量化Ⅱ類/相関分析

・「住みたい」回答への寄与率が高い、その他アンケート項目回答（特徴量）の寄与率（カテゴリスコア）を算出 **[数量化Ⅱ類]**
→寄与率が低い特徴量を削減
・特徴量の説明力を上げ、予測精度を高めるため、マルチコ（多重共線性）の変量を集約 **[相関分析/クラメール連関係数]**
→クラメール連関係数が高い特徴量を集約

相関分析（クラメール連関係数）結果から特徴量を選択

◆マルチコの排除のため、アンケート項目間の相関性が高いものを排除する。

◆アンケート全項目同士のクロスマトリクスから、項目同士の**クラメール連関係数**を算出

◆**クラメール連関係数が0.5以上のもの**をマルチコの変量と認識する

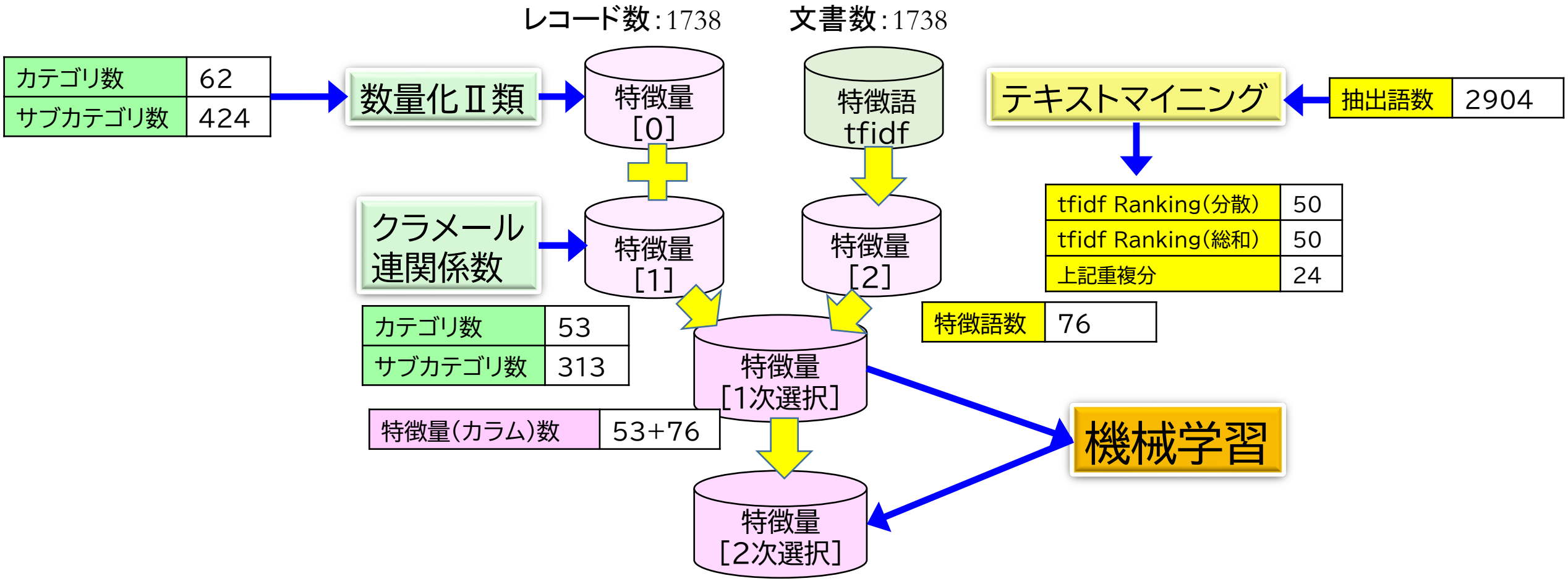
◆集約した変量は9変量（他アンケート項目と相関が強い項目を削除、代替変数があるため）

アンケート#	内容サマリ
Q5-1-4	水道・下水道の整備
Q5-1-7	がけ崩れや危険箇所対策
Q5-1-19	小学校・中学校の教育内容
Q5-1-21	生涯スポーツ
F 職業複数	F自営業-農業（複数回答型）

アンケート#	内容サマリ
Q5-1-25	歴史・文化や自然景観など、町の資源活用
Q16-1	野崎島は「世界文化遺産」に登録、小値賀町域の一部は国の重要文化的景観に選定。地域の景観を守るためにどんな協力ができるか？
Q23－1	問22で訪島する方から協力金、税収を設定する場合に、1名あたり徴収する妥当な金額
Q27	あなたは配食サービスを利用していますか

特徴量1次選択　特徴量の絞り込み

[数量化Ⅱ類] (a)特徴量(入力): カテゴリ62項目/サブカテゴリ424項目、(b)特徴量(絞込):カテゴリ53項目
[クラメール連関係数]　削減する特徴量(マルチコ排除):カテゴリ9項目、(c)特徴量(マルチコ排除後): サブカテゴリ313項目
[テキストマイニング]　特徴語を 2904(形態素解析の結果) → 特徴語 76語を選択 (tfidf値ランキングより)



特徴量2次選択 MLモデル実行と特徴量の重要度

・「住みたい」に対する寄与率が高いアンケート項目[数量化Ⅱ類/クラメール連関係数]と、自由記述回答のうち重要な特徴語[テキストマニニング]をマージし、機械学習モデルの入力とする
・「住みたい」回答の予測を行う機械学習モデルを構築、学習・予測・特徴量の重要度評価を行う。[特徴量の第2次選択]
・選択した特徴量を「主成分分析」へ入力する

機械学習モデル実行

・アンケート項目のうち「住みたい」に対する寄与率が高い項目[数量化Ⅱ類/クラメール連関係数]と、
自由記述回答のうち重要な特徴語[テキストマニニング]を選択してマージした結果を特徴量として、
「住みたい」を予測する機械学習モデルを構築する
・構築した機械学習モデルでの予測結果、重要度が高い特徴量を選択して、「主成分分析」への入力とする
[特徴量の第2次選択]

[実行プロセス」
(1)pycaretの実行:
・説明変数からQ1(住み心地)を分類する上で最も精度が高いモデルをPyCaretで自動選択。

(2)最も精度が高いモデルで予測を実行した結果、特徴量の重要度評価を実施(feature_importance、SHAP)

(3)特徴量の重要度ランキング(50程度)から、特徴量を2次選択。

Pycaret:
・Pythonのオープンソースのローコード機械学習ライブラリ。仮説から考察までのサイクルタイムを短縮することを目的としている。

・主要な学習モデルの精度を比較、精度が高い順に並べてくれる。
ランキングはpycaretが自動で出力。
accuracyやf値、適合率、rocなどの指標の%が大きい順に自動でソート

特徴量2次選択　MLモデル実行と特徴量の重要度

機械学習モデルでの予測

(1)Pycaretでの精度評価:
・ランダムフォレストがベストモデル
・accuracy(正解率)、適合率、再現率、f値、AUC共に
8割を超える精度を確認。

	Model	Accuracy	AUC	Recall	Prec.	F1	Kappa	MCC	TT (Sec)
rf	Random Forest Classifier	0.8018	0.8628	0.9239	0.8056	0.8602	0.5247	0.5419	0.557
ada	Ada Boost Classifier	0.7961	0.8421	0.8814	0.8227	0.8502	0.5311	0.5377	0.224
et	Extra Trees Classifier	0.7952	0.8578	0.9027	0.8098	0.8532	0.5173	0.5281	0.562
lightgbm	Light Gradient Boosting Machine	0.7944	0.8544	0.8852	0.8184	0.8501	0.5239	0.5299	0.215
lr	Logistic Regression	0.7903	0.8290	0.8665	0.8247	0.8446	0.5220	0.5254	0.465
gbc	Gradient Boosting Classifier	0.7895	0.8451	0.8815	0.8147	0.8465	0.5128	0.5182	0.632
ridge	Ridge Classifier	0.7771	0.0000	0.8540	0.8164	0.8341	0.4938	0.4972	0.033
lda	Linear Discriminant Analysis	0.7688	0.8067	0.8465	0.8116	0.8277	0.4754	0.4794	0.143
svm	SVM - Linear Kernel	0.7624	0.0000	0.8364	0.8198	0.8206	0.4587	0.4782	0.064
knn	K Neighbors Classifier	0.7393	0.7669	0.8404	0.7822	0.8097	0.3972	0.4012	0.240
dt	Decision Tree Classifier	0.7196	0.6930	0.7769	0.7942	0.7850	0.3819	0.3831	0.048
dummy	Dummy Classifier	0.6595	0.5000	1.0000	0.6595	0.7948	0.0000	0.0000	0.022
nb	Naive Bayes	0.5535	0.7574	0.3916	0.8530	0.5346	0.2069	0.2678	0.030
qda	Quadratic Discriminant Analysis	0.3454	0.4914	0.0337	0.5848	0.0630	-0.0119	-0.0386	0.131

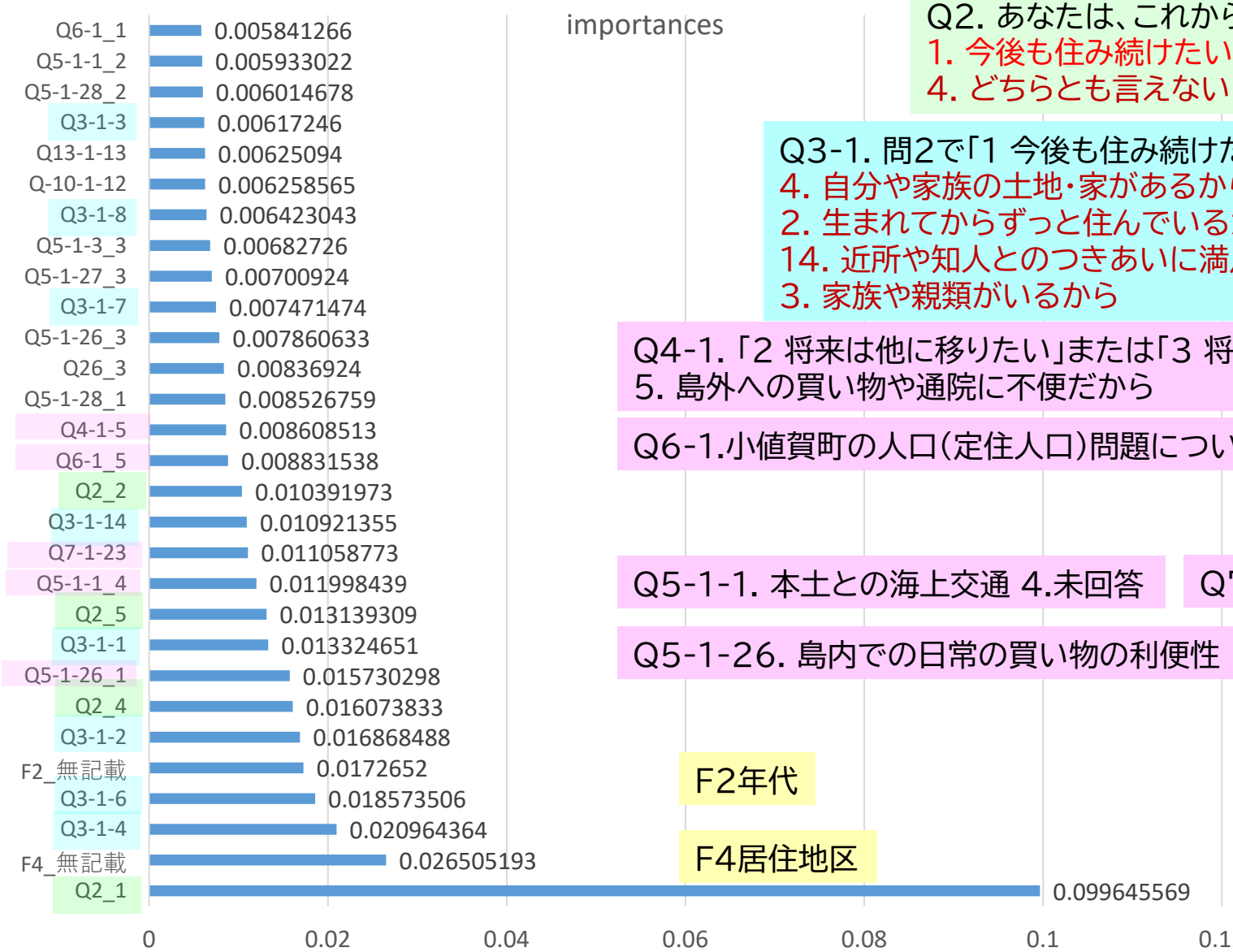
(2)ランダムフォレスト実行

```
x_train, x_test, y_train, y_test =
 train_test_split(x, y, test_size=0.3, shuffle = True)
clf  = RandomForestClassifier
(n_estimators = 100 ,max_depth=10)
```

特徴量2次選択　MLモデル実行と特徴量の重要度

特徴量の重要度　(1) feature_importance

ランダムフォレスト実行での特徴量重要度



Q2. あなたは、これからも小値賀町に住み続けたいと思いますか？
1. 今後も住み続けたい　2. 将来は他に移りたい　3. 将来は他に移らざるを得ない
4. どちらとも言えない　5. 未回答

Q3-1. 問2で「1 今後も住み続けたい」とお答えの方におたずねします。その理由は何ですか？
4. 自分や家族の土地・家があるから　6. 自然環境がよいから
2. 生まれてからずっと住んでいるから　1. 町に愛着を感じているから
14. 近所や知人とのつきあいに満足しているから　7. 治安がよいから　8. 災害が少ないから
3. 家族や親類がいるから

Q4-1.「2 将来は他に移りたい」または「3 将来は他に移らざるを得ない」理由
5. 島外への買い物や通院に不便だから

Q6-1.小値賀町の人口(定住人口)問題について 5.未回答

Q5-1-1. 本土との海上交通 4.未回答

Q7-1.特に優先して取り組んでほしいと思う課題　23.未回答

Q5-1-26. 島内での日常の買い物の利便性　1. 満足していない

F2年代

F4居住地区

「日常の交通の利便性」の重要度が高い

仮説提唱(1)

テキストマイニング結果からの仮説

＜tfidfのトップ50の分析＞

◆【ワード】(健康) カロリー、高血圧、糖尿、 (インフラ)小児科、設備、環境 → 医療や健康に関する設備
◆居住地域(F4)と「住みごごち」(Q1)とのクロスマトリクス:
ポジティブ(Yes)な内容がほとんど。ネガティブ(No)回答は特徴的な傾向 ⇒ YesとNoの割合の差分を評価
(Yes:回答1or2、No:回答4～5)

個数 / F4	Q1										
F4	1	2	3	4	5	6	(空白)	総計		yes	no	
1	170	393	139	61	18	24		805		0.699	0.098	0.60124
2	39	122	42	16	3	10		232		0.694	0.082	0.61207
3	17	55	20	6		6		104		0.692	0.058	0.63462
4	20	52	13	9	1	3		98		0.735	0.102	0.63265
5	44	81	28	11	2	4		170		0.735	0.076	0.65882
6	17	58	19	9	2	6		111		0.676	0.099	0.57658
7	7	24	8	1				40		0.775	0.025	0.75
8	14	24	9					47		0.809	0	0.80851
9	1	2	8	1				12		0.25	0.083	0.16667
10	1	2						3		1	0	1
11	4	9	9	5	2	87		116		0.112	0.06	0.05172
(空白)												
総計	334	822	295	119	28	140		1738		0.665	0.085	0.58055

人口が少ない地域を中心に
医療・健康設備の不足が
発生しているのではないか

仮説提唱(2)

数量化Ⅱ類結果からの仮説

人口の少ない地域に医療設備やアクセスを実現する手段
島全体の情報管理、交換の効率を上げる取り組み

・カテゴリウェイトが最高の職業(F3-1)と「住みごごち」(Q1)とのクロスマトリクス:
　居住地域の場合と同様に、YesとNoの割合を評価 (Yes:回答1or2、No:回答4～5)
・[Yes]最高は 1.農業、[No]が高いのは 5.会社員、4.公務員
・自由回答の有効性から、農業、公務員の自由解答欄を参照してみた
・[農業]:
　- 将来的に良くなる提案型の回答(将来を見た町作りが必要、離島小値賀町しかできない様なイベントを年に1回実施したらどうか)
　- 現状の問題を訴えている人は全体的に見て少なく、特別他と比べて強く訴えている要望はなかった
[公務員]:
　- 情報に関する要望が多い(産業のIT化を進めてほしい、うまく情報が伝わらないことが多い)

・公的機関のIT化遅れ
・町全体の管理のためのマンパワー不足？からのIT化要望

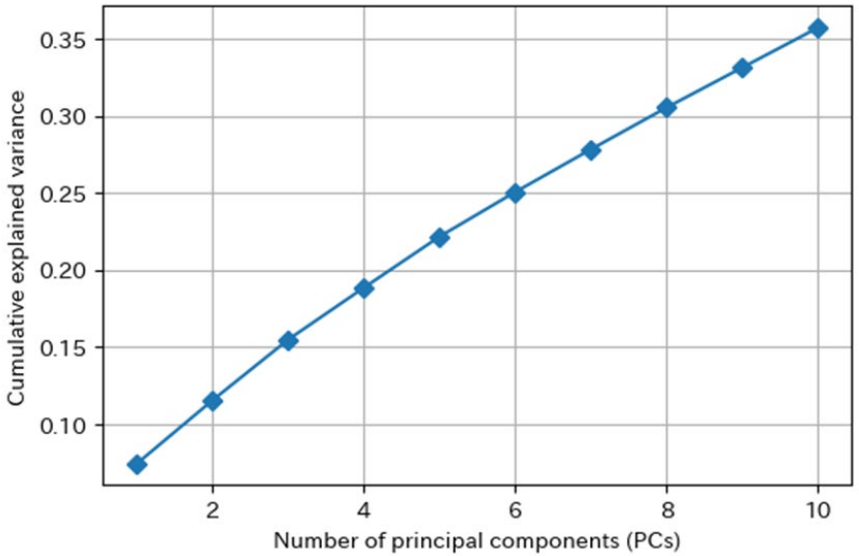
合計 / Q1	Q1										
F3-1	1	2	3	4	5	6	(空白)	総計		yes	no
農業	1	65	256	111	36	15	48		531	①	0.604519774	0.096045198
漁業	2	26	86	54	20		18		204		0.549019608	0.098039216
自営業	3	39	156	78	32	15	24		344	②	0.566860465	0.136627907
公務員	4	18	116	75	48	10	6		273		0.490842491	0.212454212 ②
会社員	5	16	188	132	108	35	24		503		0.4055666	0.284294235 ①
主婦	6	36	280	132	76	30	36		590		0.53559322	0.179661017 ③
高校生	7	5	44	24	8		6		87	③	0.563218391	0.091954023
無職	8	103	430	198	104	25	144		1004		0.530876494	0.128486056
	9	8	44	33	20		6		111		0.468468468	0.18018018
	10	6	14	24	16	10	528		598		0.033444816	0.043478261
(空白)												
総計		322	1614	861	468	140	840		4245		0.45606596	0.143227326

主成分分析(1) 方法

◆主成分分析に入力する特徴量の最適化:
・前プロセスでの機械学習(ランダムフォレスト)にて、Q1に対する影響度が0.004以上の特徴量(50個)を選択
・合成変量(主成分)を意味付けする際に活用する「主成分負荷量」の精度を考慮し、曖昧な意味付けとなる「未回答」が含まれるレコード(780)を削除
・数量化Ⅱ類及び機械学習のプロセスで未実施の特徴量の数量化(One-hot encoding)を実施
(対象カラム: F2、F4、Q6-1、Q12-1、Q19-1、Q26)

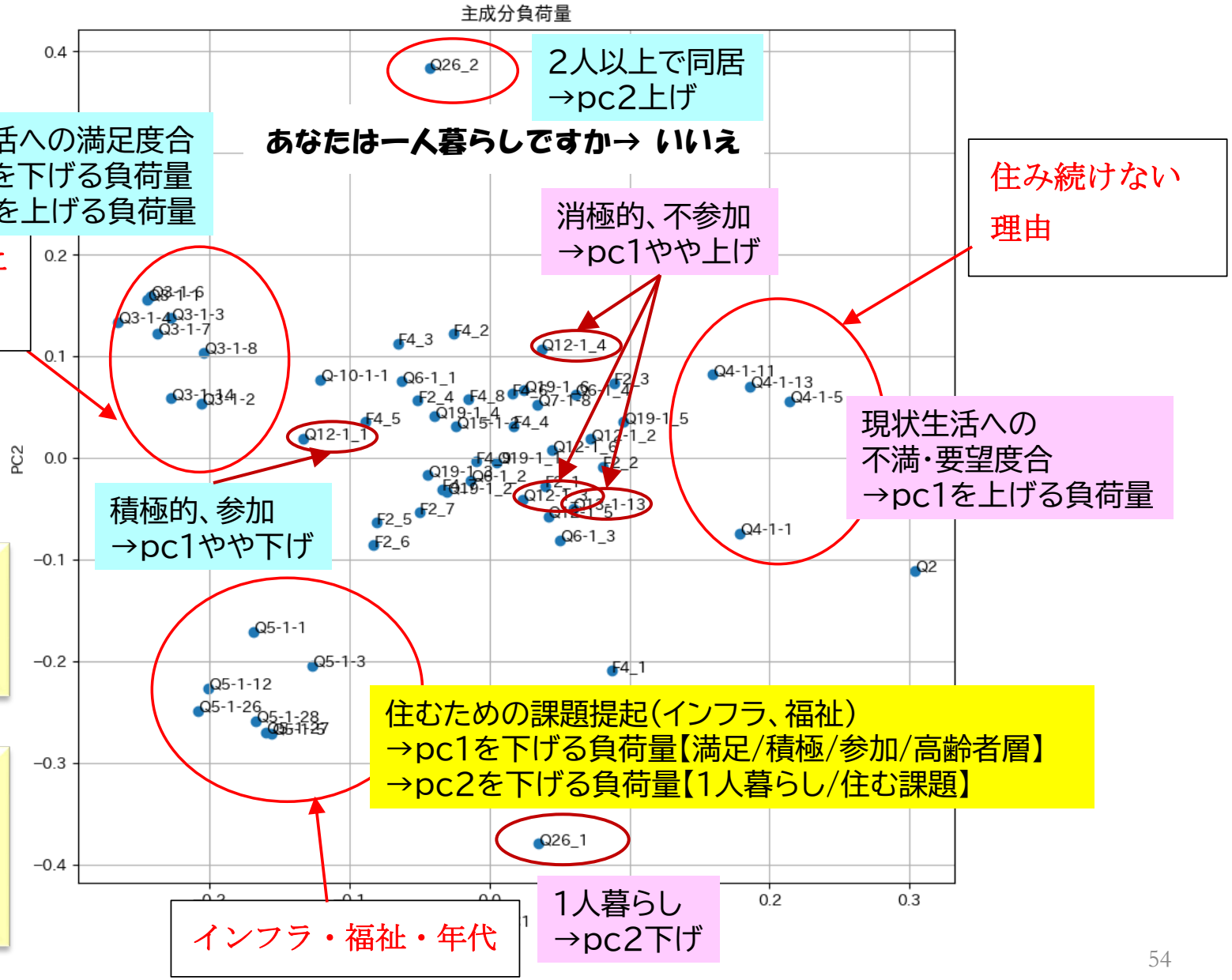
◆主成分分析の実行:
・F2 から Q26 までのフィールドの値を要素とする 58×958 の行列を標準化し、sklearnのpca.fit_transform()を適用して主成分を求めた
・算出する主成分数を10(pcaのn_components)とし、pca1(第1主成分)及びpca2(第2主成分)を使って新変量を意味付け
・第1～第4主成分での累積寄与率は20%程度

	寄与率
PC1	0.074565
PC2	0.041046
PC3	0.039408
PC4	0.033555
PC5	0.033219
PC6	0.028841
PC7	0.028009
PC8	0.026732
PC9	0.026346
PC10	0.025278



主成分分析(2) 結果概要と主成分負荷量

導出した主成分(pc1,pc2)と影響度が高い特徴量(アンケート項目)について



主成分1(pc1): **未来改善(不)志向度**
(大)現状環境に不満、やや消極的、不参加
(小)現状環境に満足、やや積極的、参加意識

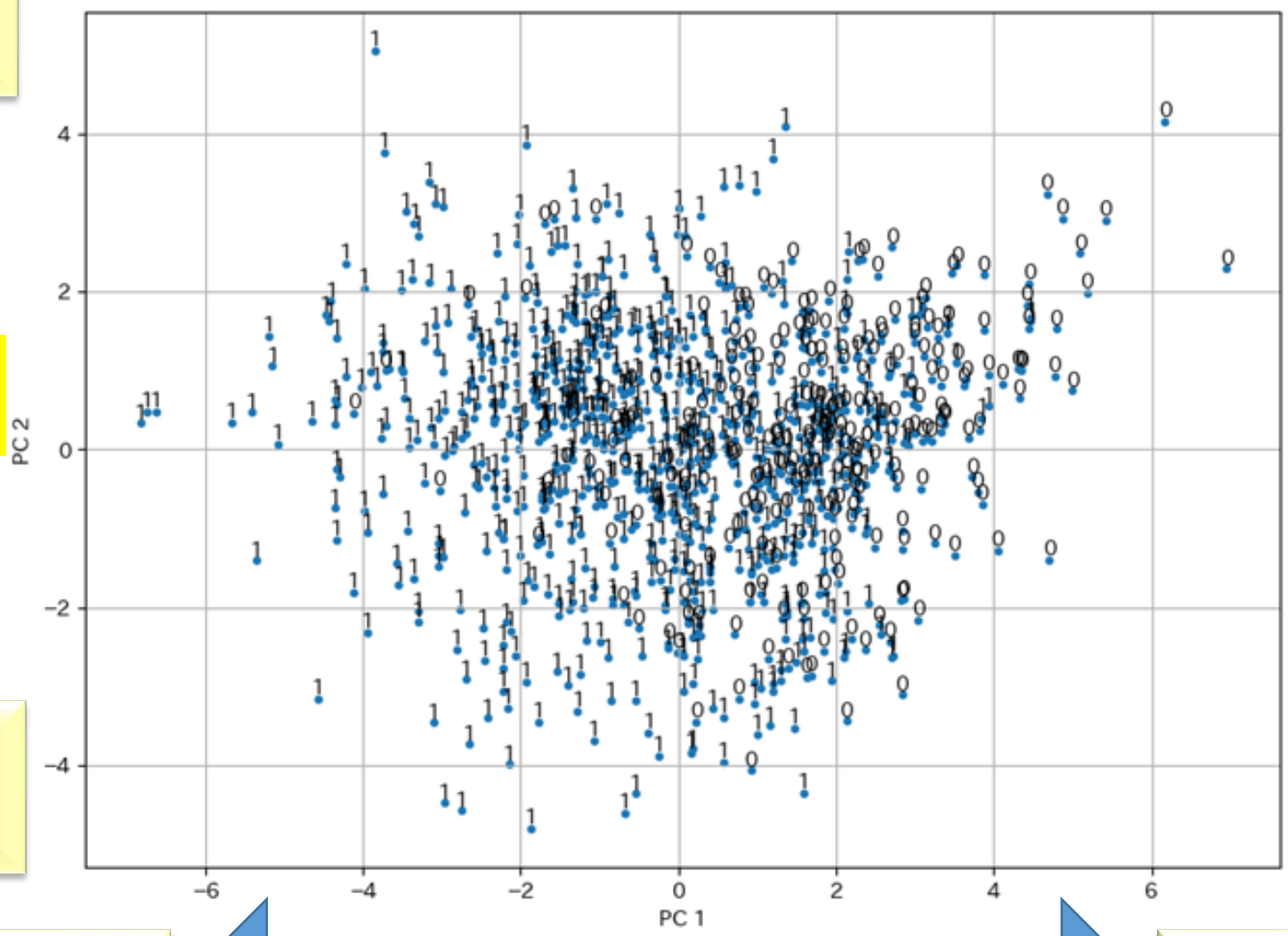
主成分2(pc2): **家族居住/居住継続意向**
(大)2人以上居住、住み続けたい
(小)1人暮らし、
住むための課題提起(インフラ、福祉)

主成分分析(3) 主成分得点

2人以上で居住
住み続けたい

家族居住(同居)/
居住継続意向度

1人暮らし、
住むための課題
(インフラ、福祉)



主成分得点

主成分に対して
(1)Q1「住みごごち」
◎「1」住みごごちが良い人は
左のエリア
→住みごごちが良い人は
積極的、参加
◎「0」住みごごちが良くない人は
右のエリア
→住みごごちが良くない人は
消極的、不参加

(2)pc1,pc2に対して総合的な
一様な分布

現状環境に満足
やや積極的、参加

未来改善(不)志向度

現状環境に不満、
やや消極的、不参加

主成分分析(4) 主成分内容(pc1/pc2)

導出した主成分(pc1,pc2)の内容について

		pc1	pc2
合成変量の意味		未来改善(不)志向度	家族居住/居住継続意向
合成変量の傾向	主成分負荷量(正)	・現状の生活環境不満 ・まちづくり参加意欲小 ・やや消極的 ・人口・就業・文化遺産活用に一部課題提起 ・地域特性 (笛吹郷、中村郷、斑島郷) ・年代特性 (40～50代、10代～30代)	・2人以上で居住 ・住み続けたい意向が強い ・地域特性 (前方郷、柳郷) ・年代特性 (40～50歳代)
	主成分負荷量(負)	・現状の生活環境満足 ・まちづくり参加意欲大 ・積極性 ・生活インフラ/環境、コミュニティ改善意向 ・未来の環境に対しての見解意識 (コミュニティ、まちづくり、産業振興、文化遺産利活用、人口問題) ・地域特性 (浜津郷、柳郷、前方郷、黒島郷、大島郷、納島郷) ・年代特性 (60代～90代)	・1人暮らし ・居住環境への課題提起多い(インフラ、福祉) ・地域特性(斑島郷、大島郷)

主成分分析(5) 主成分の内容(pca1)

主成分1(pc1): 未来改善(不)志向度

【合成変量(正/大)の傾向】
・現状の生活環境不満、まちづくり参加意欲小
・消極的
・人口・就業・文化遺産活用に一部課題提起

<詳細>
・他に移りたい志向が強い40～50代、10～30代
・現状生活環境への不満・要望高い
・若い世代、一人暮らし
・住民まちづくりは、やや消極的
・居住地区:笛吹郷、中村郷、斑島郷

【合成変量(負/小)の傾向】
・現状の生活環境満足、まちづくり参加意欲大
・積極的
・生活インフラ/環境、コミュニティ改善意向
・未来の環境に対しての見解意識
(コミュニティ、まちづくり、産業振興、文化遺産利活用、人口問題)

<詳細>
・住み続けたい志向が強い60代以降
・現状生活環境に満足、提案型、家族/親類と同居
・日々の生活基盤を大切にしたい
・高齢化や、高齢に伴う障害、障がい者福祉の懸念
・積極的な自治体活動への参画を志向
・居住地区:浜津郷、柳郷、前方郷、黒島郷、大島郷、納島郷
・対人関係は良好

主成分分析(5) 主成分の内容(pca1)

主成分1(pc1)：未来改善(不)志向度

【合成変量(正/大)の傾向】
・現状の生活環境不満、まちづくり参加意欲小
・消極的
・人口・就業・文化遺産活用に一部課題提起

【合成変量(負/小)の傾向】
・現状の生活環境満足、まちづくり参加意欲大
・積極的
・生活インフラ/環境、コミュニティ改善意向
未来の環境に対しての見解意識

【共通項として登場するパラメタ】

(A)買い物/交通の利便性改善
(B)福祉対策
(C)働く場の確保
(D)まちづくりへの住民参加
(E)定住人口問題
(F)世界文化遺産の利活用
(G)生活コミュニティ

[課題提起]
[課題提起]
[課題提起]
[積極参加]
[定住人口増大]
[具体的提案(観光目的)]
[家族親類同居]

[現状不便]
[現状不満]
[新たな、不満]
[理解/消極的]
[交流人口増大]
[具体的提案/期待無し]
[身寄無/1人暮らし]

主成分分析のプロセスの中で
取り組むべき重要なテーマ　＜別ファイル＞

主成分分析の結果を受けての施策(案)

1. pc1(未来改善志向度)/pc2(家族居住/居住継続意向)を上げるために:

(1)若い世代、生産世代に魅力ある小値賀のプロジェクトを発掘 (自治体主導+若者)【若者のモチベーション】
→発見、興味、若者・生産世代のアイディアを発掘し活かすための試行する

(2)小値賀、みらい発見!プロジェクト (住民主導、若者+ベテラン)【参加意識の向上、積極性の醸成】
<企画> 住民主体に持ち込む
自治体主催で、現在進めているテーマ、今後小値賀町が発展(人工増加、生活インフラ充実、産業振興)するためのテーマを
50代以降も含めた分かち合いの場から、一大ブレストを若者主体で実施する
他自治体との連携などもイベントを交えて積極的に。

<試行>企画で出てきたアイディアの試行

(3)主成分分析で把握した重要度の高い以下の課題対策:
(a)買い物/交通への不満 (b)生活風習・慣習の改善 (c)一人暮らし、家族は町外 (d)生活支援サービス(福祉・医療など)
(e)人口問題(定住人口)交流人口を増やすべき (f)働く場の確保 (g)世界文化遺産の利活用:野崎島の歴史の教育的素材としての活用

(4)1人暮らしが生きがいを感じる生活基盤や環境を整備

アンケート項目について

＜目的＞
・pc1、pc2の意味付けの仮説は間違っていないかを検証
・いづれにせよ影響度/重要度の高い項目を深掘り
→潜在テーマがさらに浮き彫りに

1．主成分分析の項目

・まちづくりへの参加意欲
・自治体のまちづくりイベントに積極的/消極的、その理由
・住民自身のまちづくりイベントに関して、やる/やらない、その理由
・今後小値賀町が発展（人工増加、生活インフラ充実、産業振興）するためのテーマ

2．主成分分析時に出てきた寄与率が高い項目

3．機械学習、重要度が高い項目

主成分分析の仮説(pca1)の検証について

【分析(pc1)】①(負)まちづくり参加意欲[大]生活環境[満足] vs ②(正)まちづくり参加意欲[小]生活環境[不満]　の①②に2極化。
①は60代以上、②は生産人口/若手が負/正の主成分負荷量(絶対値)が大きい

【仮説】(a)若手が言いにくい、活躍しにくい環境である
　　(b)若手が活躍する環境、若手を主体に町民全員で取り組む環境を創ることが重要で将来の価値に繋がる

→【アンケート】問10(1～5)の結果を測り(a)(b)を検証する。
(a)若手と問10の相関。[検証1]
(a)(b)問10の重要度。問10-1,2,3,4,5を特徴量(xi)、問5or問6の住みごごちを目的変数(y)とした機械学習モデル予測時の(xi)の重要度評価 [検証2]

→【仮説検証結果からの施策導出(仮説)】
確かに(a)&(b)である(潜在的な課題)。では、この課題に対して住民意識を高める条件を別のアンケート項目の重要度から導出する。

　重要度が高い問10-1,2,3,4,5を目的変数(yj)とした時の、前回アンケート各項目＋特徴語(xj)のうち重要度が高く、(yi)の値が1のpredict plobaを
　大きくなるような(xj)の条件を求める　[施策案の導出-シビックプライド醸成施策]　　★仮説として次回のアンケートでより具体的な施策価値を検証する

(1)pca1の仮説→2022年度アンケート結果からの検証→施策導出(仮説)→施策有効性の確認(検証)[次回追加アンケート]
(2)pca2の仮説→2022年度アンケート結果からの検証→施策導出(仮説)→施策有効性の確認(検証)[次回追加アンケート]
(3)pcanの寄与度向上→追加潜在価値(課題)の導出→次回追加アンケート(今回追加アンケートから明示的に検証できる可能性もあり)
→次回追加アンケート結果からの検証→施策導出(仮説)→施策有効性の確認(検証)[次々回追加アンケート]
(4)2022年度追加アンケート結果からの主成分分析(今回実施した内容と比較評価)による新潜在価値導出
(5)感情分析＋テキストマイニングからの改善提案(平成30年度、2022年度結果と差異分析)

主成分分析の仮説(pca2)の検証について （作成中）

【分析(pc2)】

==以下pc1のもの
①(負)まちづくり参加意欲[大]生活環境[満足] vs ②(正)まちづくり参加意欲[小]生活環境[不満]　の①②に2極化。
①は60代以上、②は生産人口/若手が負/正の主成分負荷量(絶対値)が大きい

【仮説】(a)若手が言いにくい、活躍しにくい環境である
　　(b)若手が活躍する環境、若手を主体に町民全員で取り組む環境を創ることが重要で将来の価値に繋がる

→【アンケート】問10(1～5)の結果を測り(a)(b)を検証する。
(a)若手と問10の相関。[検証1]
(a)(b)問10の重要度。問10-1,2,3,4,5を特徴量(xi)、問5or問6の住みごごちを目的変数(y)とした機械学習モデル予測時の(xi)の重要度評価 [検証2]

→【仮説検証結果からの施策導出(仮説)】
確かに(a)&(b)である(潜在的な課題)。では、この課題に対して住民意識を高める条件を別のアンケート項目の重要度から導出する。

重要度が高い問10-1,2,3,4,5を目的変数(yj)とした時の、前回アンケート各項目＋特徴語(xj)のうち重要度が高く、(yi)の値が1のpredict plobaを
大きくなるような(xj)の条件を求める　[施策案の導出-シビックプライド醸成施策]　　★仮説として次回のアンケートでより具体的な施策価値を検証する

2023年度以降の全体ソリューションアーキテクチャ(概念図)

研究テーマ

【1】分析フェーズ

感情＋回答本質深掘

小値賀町分析汎用化

D.研究成果まとめ

【3】導出主成分の評価

H30アンケート結果

drill down Analysis
テキスト分析
テキストマイニング
対象文書の分類
感情分析

2023アンケート結果

2024年度以降
アンケート結果

E.研究の拡張
(A-5)

Info. Mining model
(潜在価値発掘モデル)

tuning level
・数値化　・異表記辞書
・特徴量選択閾値
・カテゴリ係数選択
・使用アルゴリズム
・ハイパーパラメタ
・交差検証　など

function level
・tfidf以外
・ストップワード
・数量化-回帰分析
・重要度評価(PDP)
・n主成分(n>2)
など

テキスト
マイニング
数量化Ⅱ類
相関分析
機械学習
主成分分析

テキスト
マイニング
数量化Ⅱ類
相関分析
機械学習
主成分分析

drill down analysis

tuning level 最適

evaluation model

function level 最適

condition suggestion

B.2023アンケート評価
(B-2)

自由＋項目混在分析

主成分特徴量選択方式

A.主成分精度向上
(A-1)～(A-4)

pca
意味付け

主成分意味付けの検証

アンケート
追加

pca意味付け
寄与率向上

【2】導出主成分の意味付け

主成分意味付けの検証

H30アンケート結果

B.追加アンケート評価
(B-1)

2023アンケート結果

追加アンケート(主成分)評価

手動分析

MLによる重要度評価

C.施策の導出
(C-1)

(X)

機械学習

(y)

pcaによる
追加アンケート項目

特徴量
重要度評価

「住み心地」
「住み続けたさ」

重要度が高い
追加アンケート項目

【検証】追加アンケート項目の重要度が
高い(住み易いための価値が高い)こと
【変数の選択】
重要度が高いアンケート項目を選択

MLによる
変数条件導出

(X)

機械学習

(y)

アンケートに
おける全アンケート
項目

特徴量
重要度評価

重要度が高い
追加アンケート項目

C.施策の導出
(C-2)～(C-4)

現状の居住環境
改善施策の検討

自治体
施策案

【4】自治体施策の
具体化

アンケート回答の本質の掘り起こし -能動的発言からの本音発見

(a)アンケート選択回答方式と自由回答方式の両方の結果により回答内容の本質を知る
(b)回答に設問者が想定していない、隠れた本質的に言いたいことを知り対策を練る

例えば、
「xxxが良くない」「yyyして欲しい」という回答への対策を考えるときに
(A)単なるクレームで言っているのか （受動的な発言）
(B)現状から前向きな提案で言っているのか （能動的な発言）
によって、対策「xxxを良くする」「yyyして欲しいことをする」という対策が回答者の価値に繋がるか繋がらないかが変わってくる。
ネガティブな意見を評価するときにどうとらえるか
アンケートにより対策を講じる手法を執るときに、上記(A)(B)のどちらなのかを
見抜くことは対策方法を検討するときに重要な手がかりとなる。
前向きでない人は、面倒くさい自由回答は書かない。書いても短い。
でも書いてあるのは面倒くさくても書くのでそこに意思が現れている可能性が高い。
なので、書かれた文字回答が「能動的な発言」なのか「受動的な発言」なのかを
見抜いて回答内容が「クレーム」なのか「提案なのか」を見抜くことは意味がある。
テキスト全体分析ルール＋「感情分析（ネガポジ分析）」により、上記を見抜けないだろうか。

(1)マクロな視点で回答者がどういう意志でいるかを分類し（選択回答方式＋自由回答方式の分析）
(2)分類した回答者が本当に言いたいことを深掘りする（自由回答方式の分析深耕）

→(1)で埋もれた、潜在的に回答者が言いたいことを(2)で知る
まず(1)および(2)の、特徴量及び特徴的な言葉を軸にして、潜在的に言いたいことを予測する（主成分分析）【仮説1】
【仮説】立証のための追加質問（潜在的に言いたかったことはこれだよね）を用意して、その回答内容から検証する。
（回答結果のML重要度の評価。今のアプローチ）【仮説1の検証】
果たしてこの「潜在的に言いたかったこと」が本当なのか、さらに具体的に設問者が想定できない潜在的な言いたいことを深掘りする。(2)で自由回答を掘り下げると言いたいことが具体的な浮かび上がる可能性が高い。【仮説2】
(2)【仮説2】の【検証】方法が難しい。本当に言いたいことを予測しても、裏が取れない。

テキストマイニング手法を
アンケートに適用することに関して

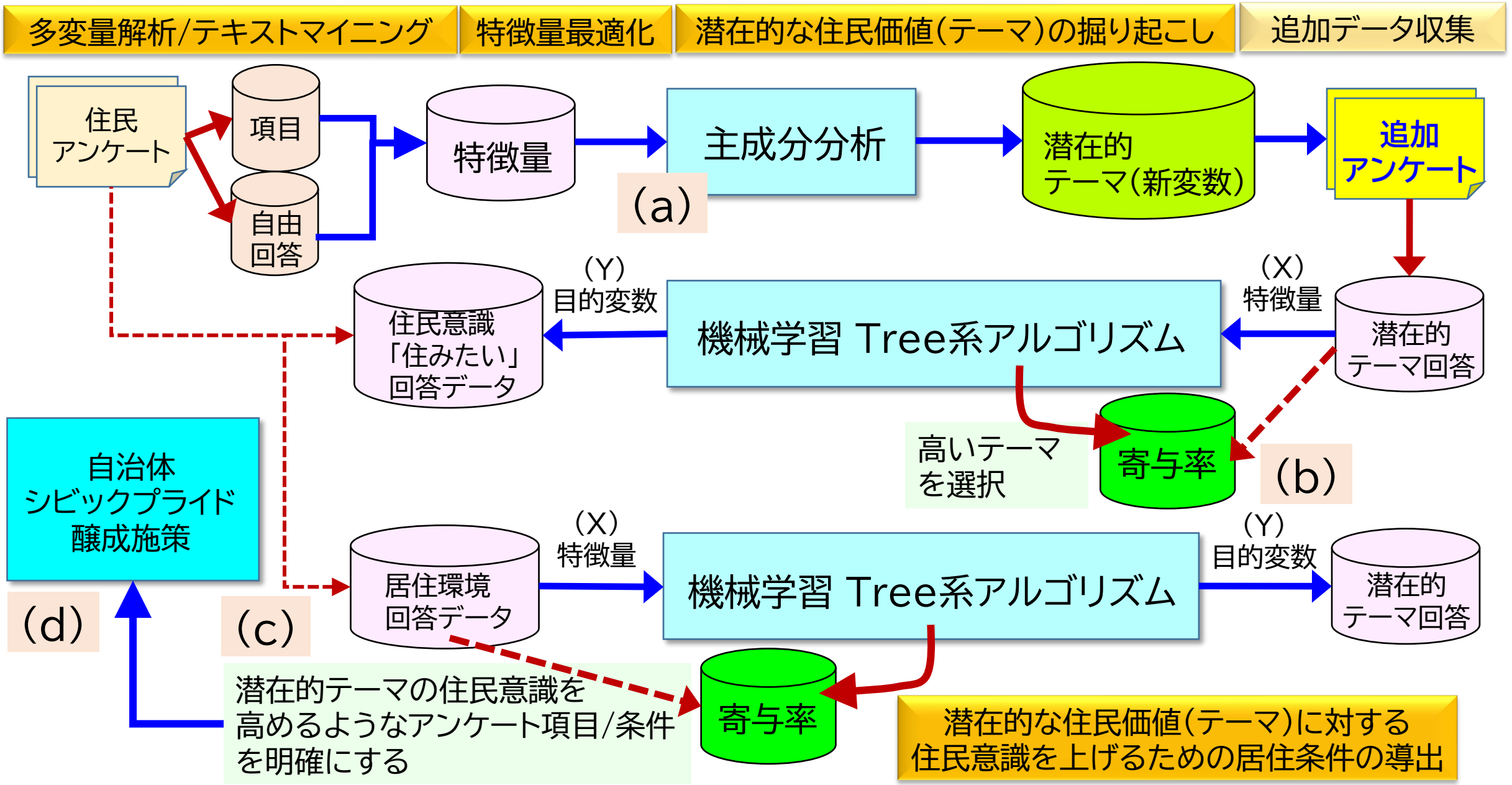
<pro>
回答者の本音を引き出すには、自由回答欄が有効。アンケートの回答者の感想や心境は、アンケートの作成者の想定を超えるもの。
アンケート用紙に「思うことを自由に記してください」と書いておくだけで、回答者は、作成者が予想しなかった指摘をしてくれるはずです。
そのため回答者は、自由回答欄には、よいことも悪いことも正直に書きます。コストをかけてアンケートを実施する目的は、消費者や顧客の本音を引き出すことです。自由回答欄には本音が多く記載されているので、気づきの宝庫になります。選択肢の場合、回答者は無意識に嘘を答えてしまうことがあります。
想定していなかったことを指摘してもらえる
自由回答欄に、アンケート作成者がまったく想定していなかったことを書く人がいます。
https://kotodori.jp/user-research/analytics/questionnaire-free-answer/

<con>
自由記述欄に書かれた内容がその人の本音である可能性は極めて低い？
テキストマイニングに適しているのは『能動的テキストデータ』であり、『受動的テキストデータ』には向いていません。アンケート自由記述回答は『受動的テキストデータ』
能動的テキストデータは、伝える側が『相手に伝わるように自分の気持ちを言語化している』ので、それをテキストマイニングで分析すれば、何かしら得られるものがあると思います。コールセンタなど
アンケートの自由記述欄に書かれるものは、いわゆる『書けと言われたから書いた』データ(=受動的テキストデータ)なので、回答者がきちんと気持ちを言語化している可能性は低いのです。
https://lactivator.net/2019/07/03/free_description/

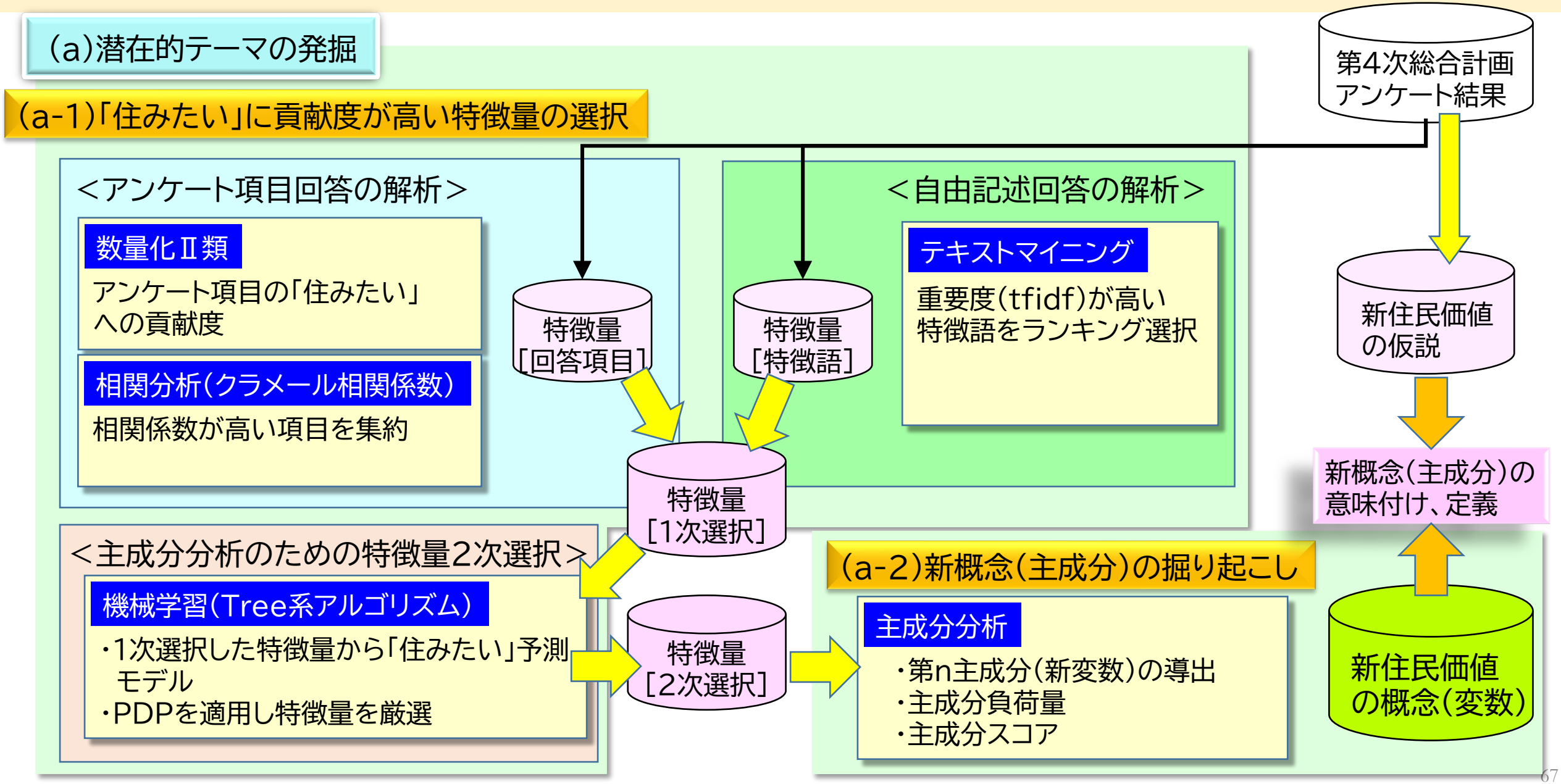
検討資料（科研費申請の図、紀要の図）

x_train, x_test, y_train, y_test = train_test_split(x, y, test_size=0.3, shuffle = True) clf = RandomForestClassifier(n_estimators = 100 ,max_depth=10)

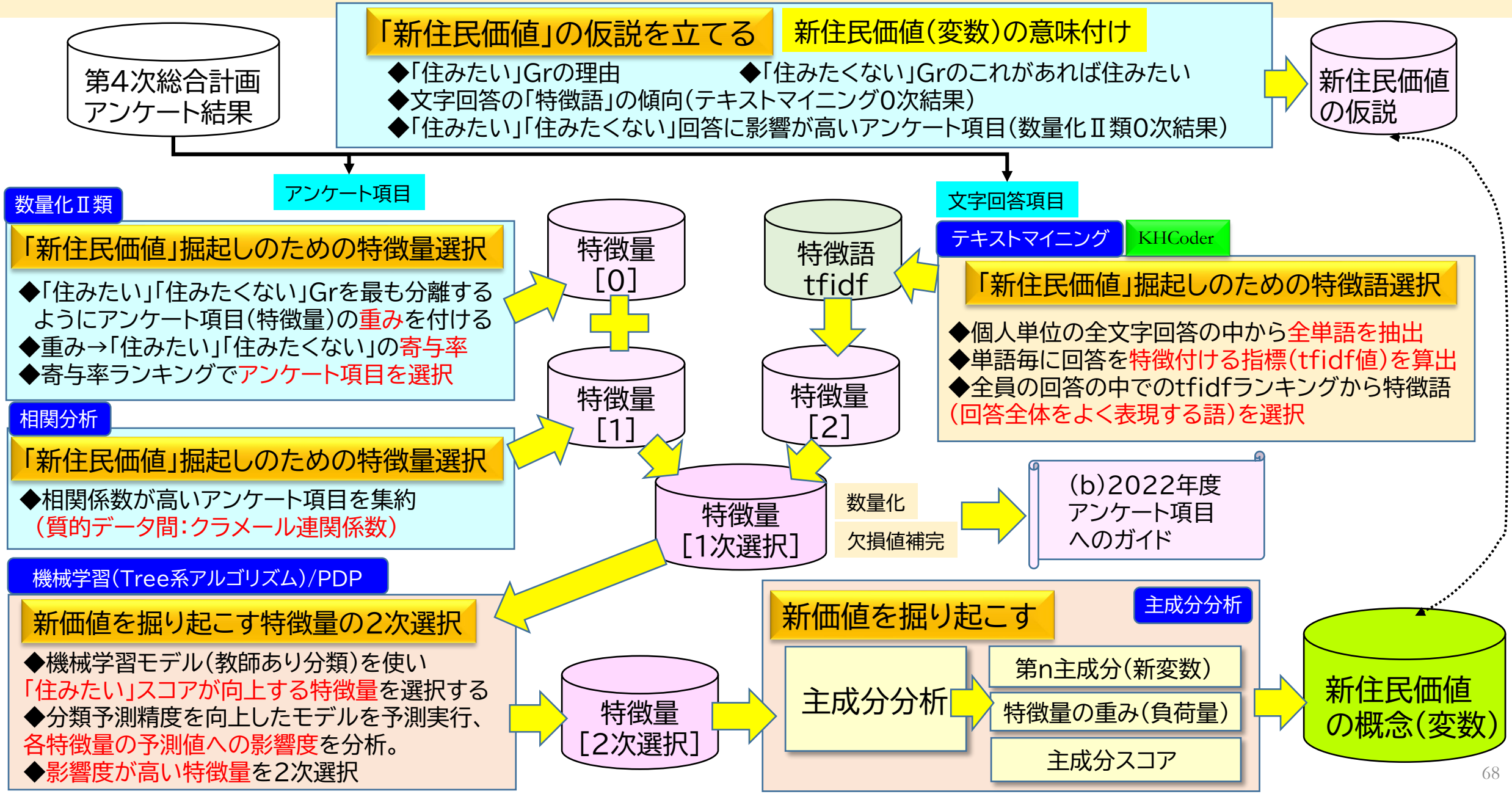
(1)全体プロセス　科研費図



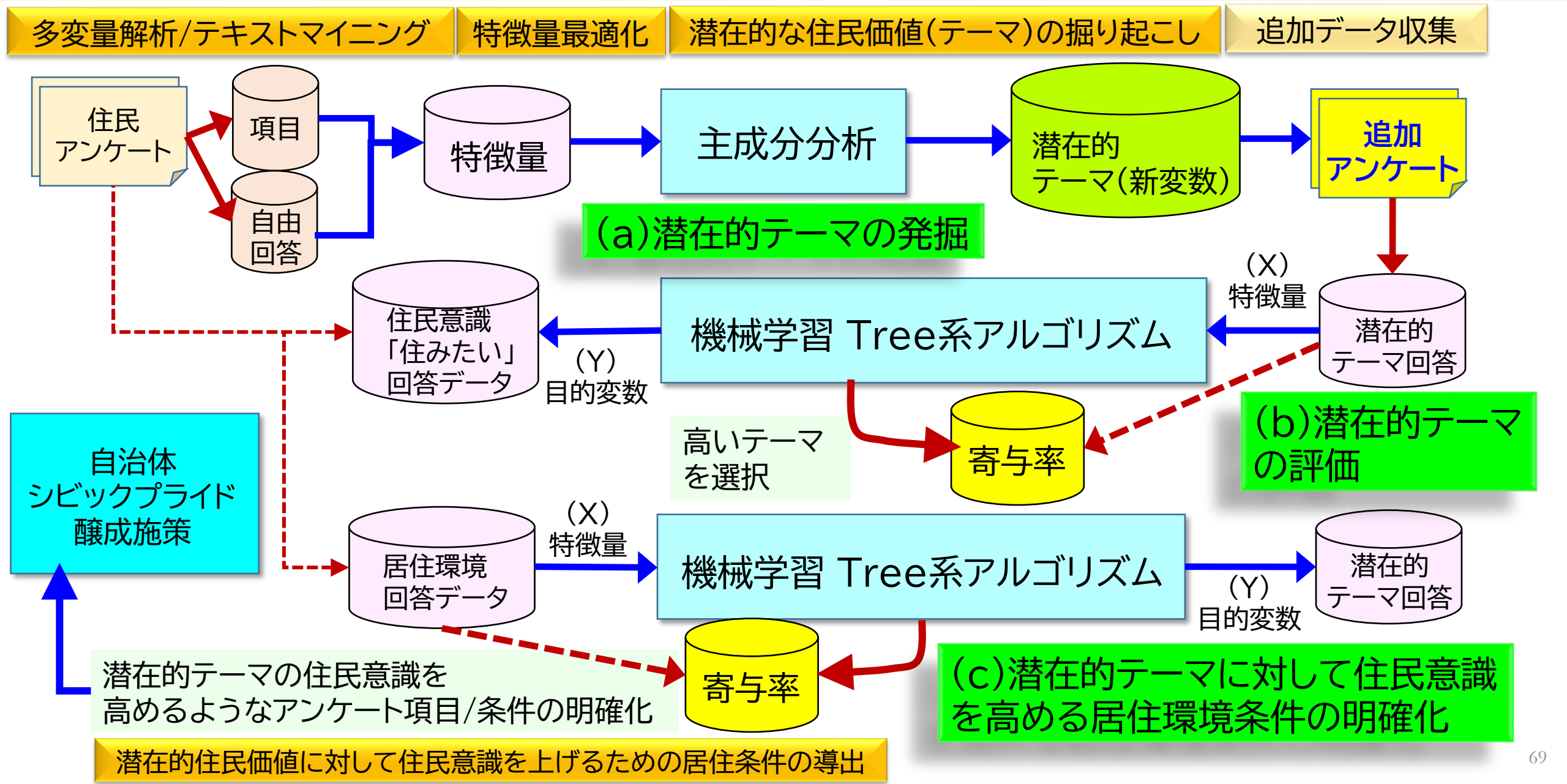
(1)新住民価値の明確化　科研費図



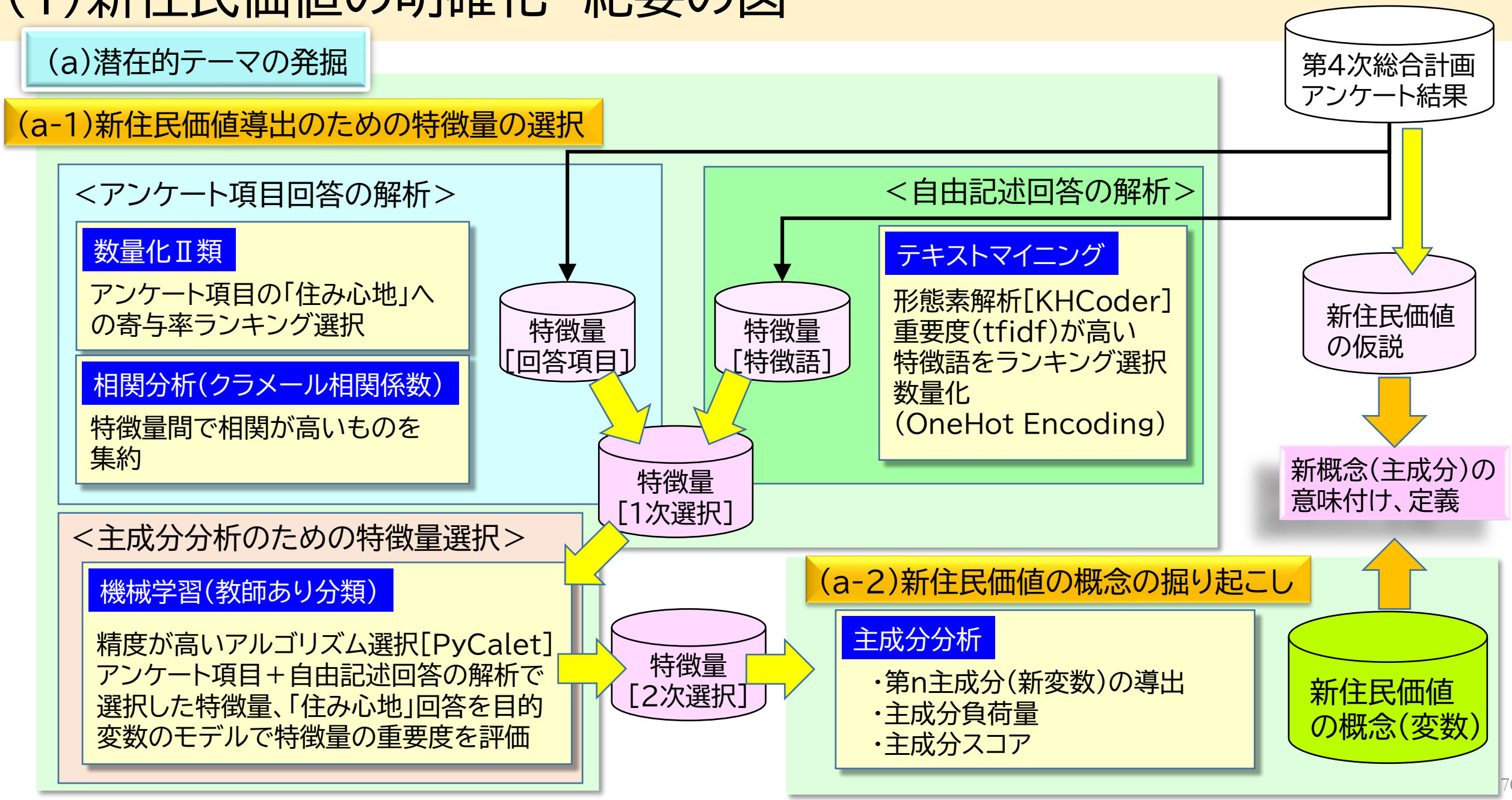
(1)新住民価値の明確化　科研費図



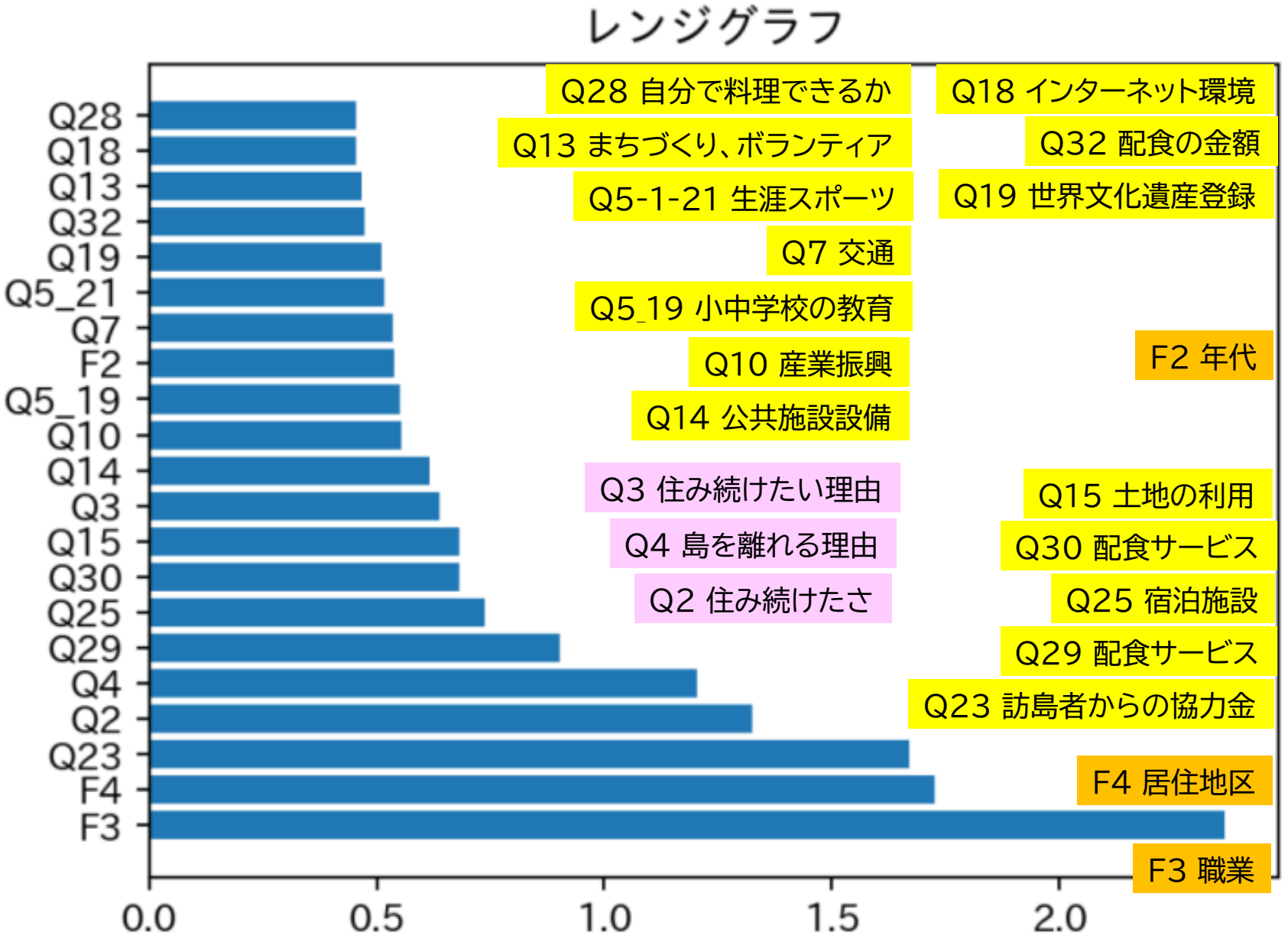
(1)全体概念図　紀要の図


多変量解析/テキストマイニング
特徴量最適化
潜在的な住民価値(テーマ)の掘り起こし
追加データ収集
住民
アンケート
項目
自由
回答
特徴量
主成分分析
潜在的
テーマ(新変数)
追加
アンケート
(a)潜在的テーマの発掘
(X)
特徴量
住民意識
「住みたい」
回答データ
(Y)
目的変数
機械学習 Tree系アルゴリズム
潜在的
テーマ回答
(b)潜在的テーマ
の評価
高いテーマ
を選択
寄与率
自治体
シビックプライド
醸成施策
(X)
特徴量
居住環境
回答データ
機械学習 Tree系アルゴリズム
(Y)
目的変数
潜在的
テーマ回答
寄与率
(c)潜在的テーマに対して住民意識
を高める居住環境条件の明確化
潜在的テーマの住民意識を
高めるようなアンケート項目/条件の明確化
潜在的住民価値に対して住民意識を上げるための居住条件の導出

(1)新住民価値の明確化　紀要の図



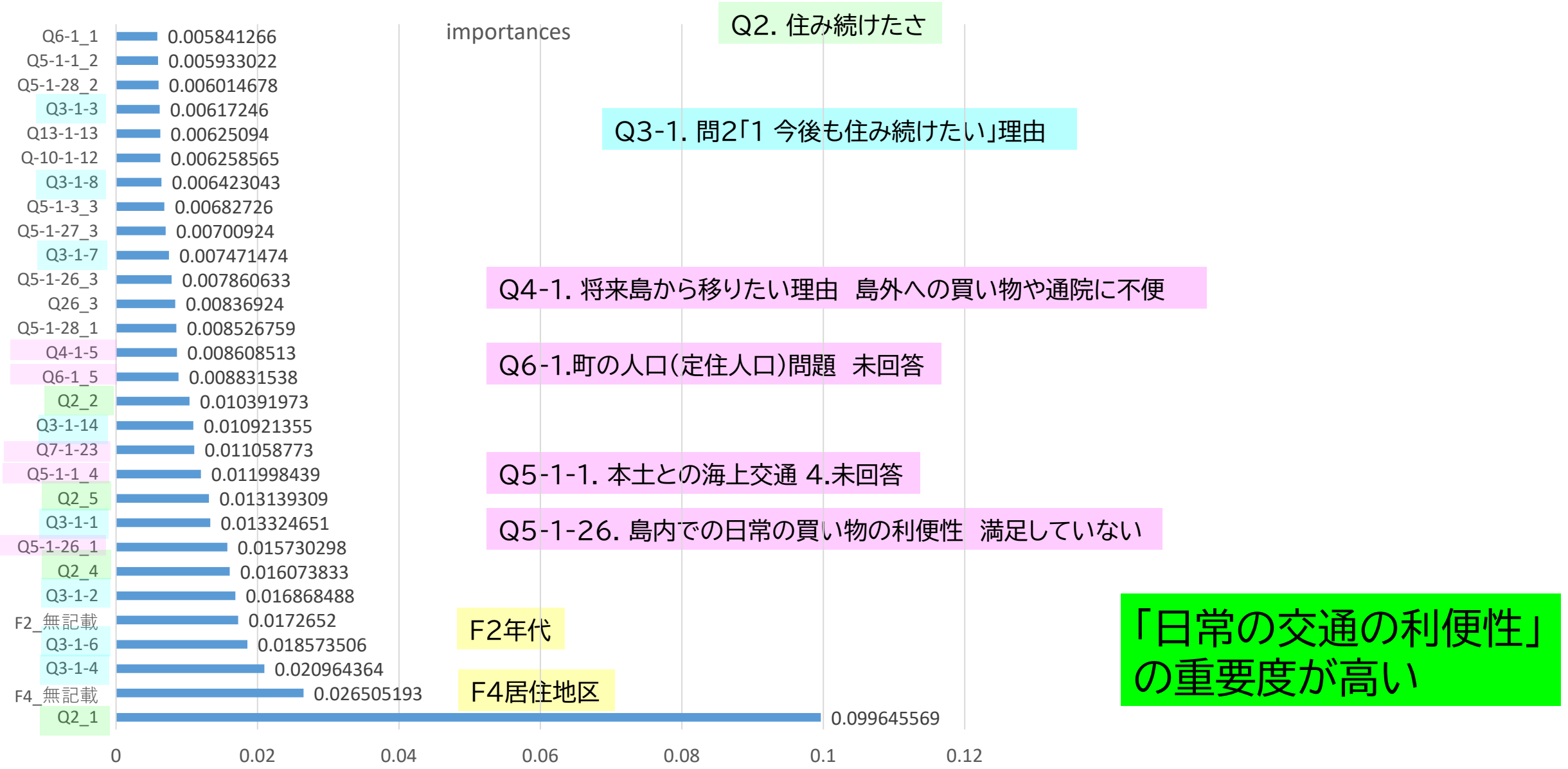
(1)新住民価値の明確化　紀要の図　数量化Ⅱ類レンジランキング



(1)新住民価値の明確化　紀要の図　テキストマイニング

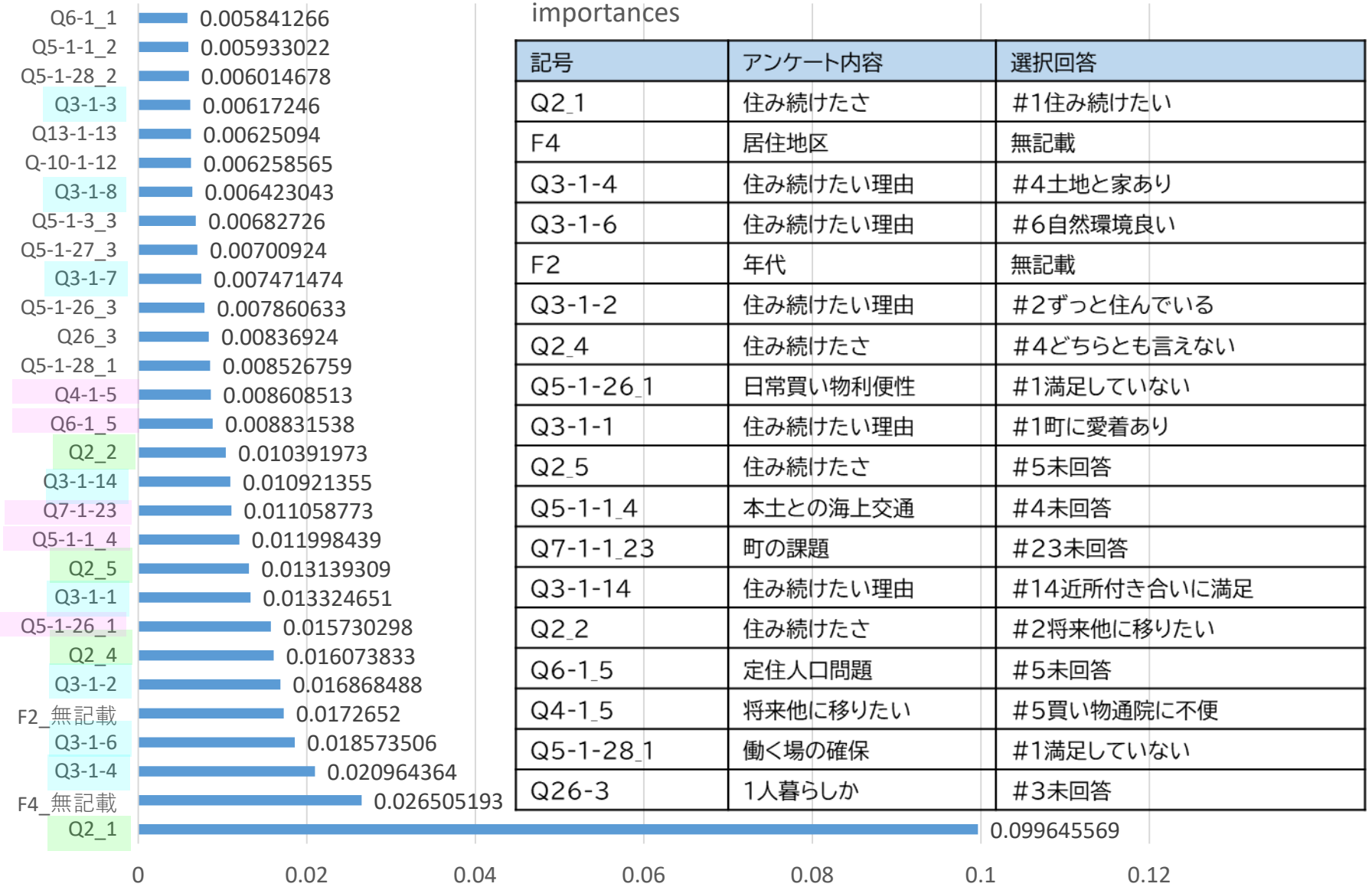
ない.1	する	高血圧	回答	ない	円	タクシー	思う	糖尿	ある	塩	わかる	小児科	カロリー	減	なる	ほしい.1		
町	高齢	できる	いる	行く	子供	小値賀	人	増加	もっと	子ども	多い	前	人口	参加	制限	観光		
出来る	施設	町民	分かる	医療	行う	賃金	人材	進学	計画	減	保全	ナイ	流れ	中途半端				
しかた	DM	任せる	なんとも	環境	特に	身体	食事	IT	悪化	すべて	転勤	企業	誘致	葬儀	予備	納税		
ふるさと	船	用事	まかなう	クラウドファンディング				住宅	血糖	整備	訪れる	抜	余裕					

(1)新住民価値の明確化　紀要の図　機械学習の重要度



「日常の交通の利便性」の重要度が高い

(1)新住民価値の明確化　紀要の図　機械学習の重要度



記号	アンケート内容	選択回答
Q2_1	住み続けたさ	#1住み続けたい
F4	居住地区	無記載
Q3-1-4	住み続けたい理由	#4土地と家あり
Q3-1-6	住み続けたい理由	#6自然環境良い
F2	年代	無記載
Q3-1-2	住み続けたい理由	#2ずっと住んでいる
Q2_4	住み続けたさ	#4どちらとも言えない
Q5-1-26_1	日常買い物利便性	#1満足していない
Q3-1-1	住み続けたい理由	#1町に愛着あり
Q2_5	住み続けたさ	#5未回答
Q5-1-1_4	本土との海上交通	#4未回答
Q7-1-1_23	町の課題	#23未回答
Q3-1-14	住み続けたい理由	#14近所付き合いに満足
Q2_2	住み続けたさ	#2将来他に移りたい
Q6-1_5	定住人口問題	#5未回答
Q4-1_5	将来他に移りたい	#5買い物通院に不便
Q5-1-28_1	働く場の確保	#1満足していない
Q26-3	1人暮らしか	#3未回答

「日常の交通の利便性」の重要度が高い

(1)新住民価値の明確化　紀要の図　機械学習の重要度

記号	アンケート内容	選択回答
Q2_1	住み続けたさ	#1住み続けたい
F4	居住地区	無記載
Q3-1-4	住み続けたい理由	#4土地と家あり
Q3-1-6	住み続けたい理由	#6自然環境良い
F2	年代	無記載
Q3-1-2	住み続けたい理由	#2ずっと住んでいる
Q2_4	住み続けたさ	#4どちらとも言えない
Q5-1-26_1	日常買い物利便性	#1満足していない
Q3-1-1	住み続けたい理由	#1町に愛着あり
Q2_5	住み続けたさ	#5未回答
Q5-1-1_4	本土との海上交通	#4未回答
Q7-1-1_23	町の課題	#23未回答
Q3-1-14	住み続けたい理由	#14近所付き合いに満足
Q2_2	住み続けたさ	#2将来他に移りたい
Q6-1_5	定住人口問題	#5未回答
Q4-1_5	将来他に移りたい	#5買い物通院に不便
Q5-1-28_1	働く場の確保	#1満足していない
Q26-3	1人暮らしか	#3未回答

Q5-1-26. 島内での日常の買い物の利便性　1. 満足していない

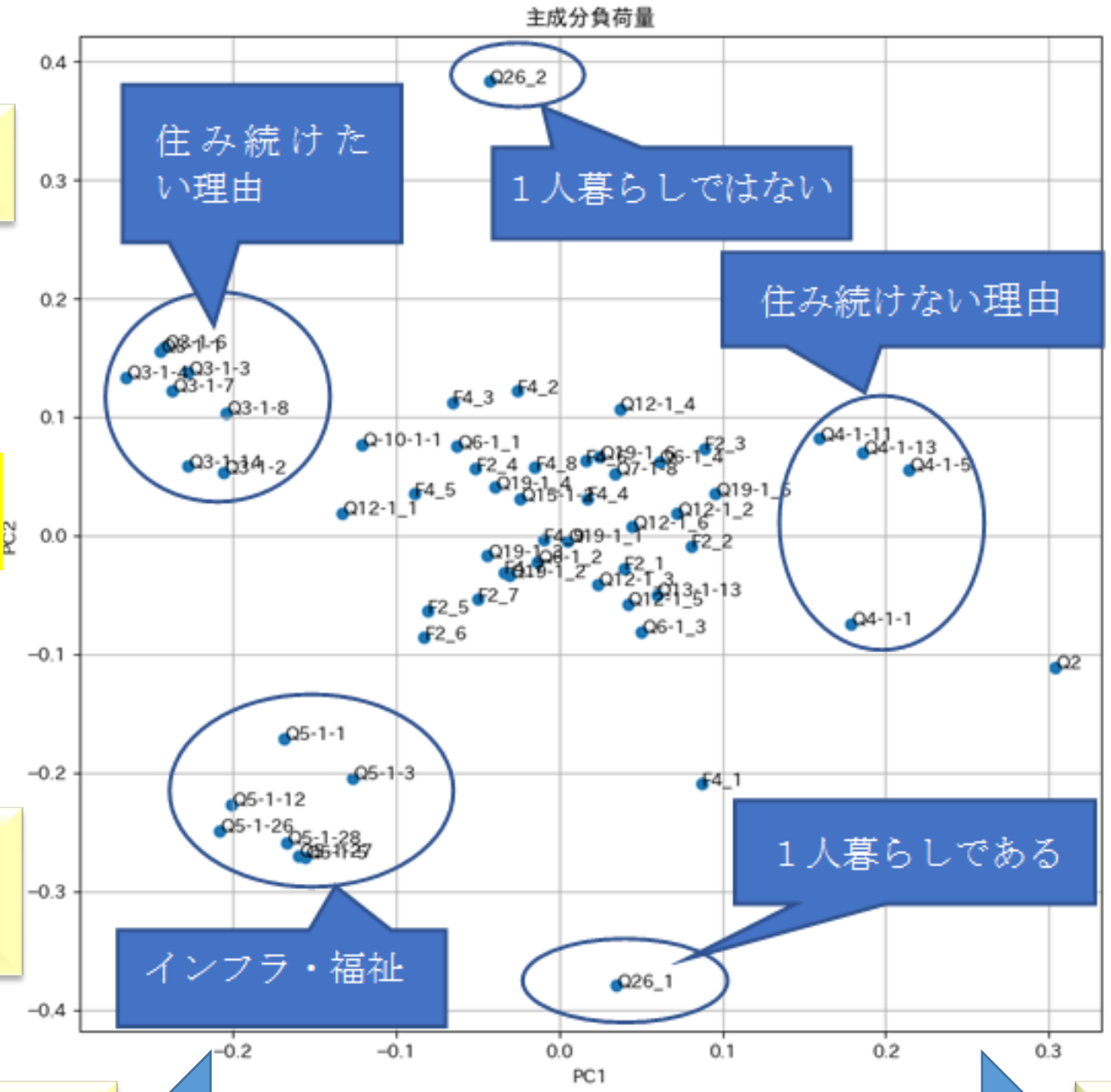
Q2. あなたは、これからも小値賀町に住み続けたいと思いますか？
1. 今後も住み続けたい　2. 将来は他に移りたい　3. 将来は他に移らざるを得
4. どちらとも言えない　5. 未回答

Q3-1. 問2で「1 今後も住み続けたい」とお答えの方におたずねします。その
4. 自分や家族の土地・家があるから　6. 自然環境がよいから
2. 生まれてからずっと住んでいるから　1. 町に愛着を感じているから
14. 近所や知人とのつきあいに満足しているから　7. 治安がよいから　8.
3. 家族や親類がいるから

2人以上で居住
住み続けたい

家族居住(同居)/
居住継続意向度

1人暮らし、
住むための課題
(インフラ、福祉)



	寄与率
PC1	0.074565
PC2	0.041046
PC3	0.039408
PC4	0.033555
PC5	0.033219
PC6	0.028841
PC7	0.028009
PC8	0.026732
PC9	0.026346
PC10	0.025278

現状環境に満足
やや積極的、参加

未来改善(不)志向度

現状環境に不満、
やや消極的、不参加

2. 研究の概要 -本研究で取り組む課題

論文の2.1の内容を、以下の部材を使って整理する？

（1）新住民価値の明確化：
第4次総合計画アンケート結果を分析し、AIを用いて**潜在的な「新住民価値」の掘り起こし**を行う

（2）新住民価値創出のための施策提案：
掘り起こした「新住民価値」を向上させるために、（アンケートの項目に紐づく）**どういう環境要件が重要か**を明確にする。明確にした環境要件から、具体的な施策案を検討・提案する

（1）新住民価値の明確化	（a）新住民価値の掘り起こし	・第4次総合計画アンケート結果を分析し、潜在的な「新住民価値」をAIで掘り起こす ・精度を上げるため、AIに入力する特徴量（アンケート項目・特徴語）を選択する
	（b）2022年度アンケート項目	（a）の分析の中での第1次分析結果を基に、住民意識（住みやすい、住み続けたい）を把握するのに重要なアンケート項目を明確化し、本年度のアンケート項目に反映する
（2）新住民価値を高める施策のご提案	（c）新住民価値条件の明確化	・「新住民価値（y）」を高めるような、現在の環境条件を明確にする。 ・環境条件に応じた施策案をご提案する 新住民価値（y）の教師データが必要
	（d）新住民価値での住民意識の検証	・「新住民価値（X）」により、住民意識（住みやすい、住み続けたい）を予測（検証）する （「新住民価値」により、住民意識はどう予測できるか）
	（e）新住民価値アンケート具体化と分析	（c）（d）より「新住民価値」を掘り起こすために重要なアンケート項目を明確化する