

ベースとして【個人特性】と【追跡】のグループを考える。それらについて、今後の解析に支障のないと考えられる範囲で可能な限りセルの併合（まるめの処理）を行い、【個人特性】については2パターン、【追跡】については4パターンのサブグループを定義し、それらの分類数 $K$ とユニークセル数 $S_1$ を求めた。

#### [検討4]

21のアンケートカテゴリに対し①アンケートカテゴリのみ、②【個人特性】とアンケートカテゴリをそれぞれ組み合わせた場合、③【追跡】とアンケートカテゴリをそれぞれ組み合わせた場合、④【個人特性】【追跡】の組み合わせに各アンケートカテゴリを組み合わせた場合、の全ての場合における分類数 $K$ とユニークセル数 $S_1$ を求めた。

### C. 結果

[検討1]より、単体の変数で一意である個体が存在するのは、「v0502 (10年観察終了日)」「v0600 (死因 ICD-9 コード 4 桁)」「v1200 (身長 (cm))」「v1201 (体重 (kg))」「v1610 (初経年齢)」「v1612 (自然閉経年齢)」「v1613 (手術閉経年齢)」「v1615 (出産人数)」「v1616 (初産年齢)」「v2101 (喫煙開始年齢)」「v2102 (喫煙本数/日)」「v2103 (禁煙年齢)」「v2801 (転入何年前か)」「v2940 (最も長く就いた仕事)」「v2950 (従事年数)」の15変数であった。(表2)

[検討2]より、204の全ての変数を組み合わせた場合に一意となる個体の数は100,605であった。

[検討3]より、性別×年齢×居住地の情報からなる【個人特性】グループにおいて、まるめの処理を行わない「個人特性1」では分類数673、ユニークセル数19であったのに対し、年齢を5歳階級とし85歳以上はまとめた「個人特性2」では、分類数120、ユニークセル数は0と、一意性が消失

した(表3)。

追跡に関する日付×転帰×死因からなる【追跡】グループでは、処理を行わない「追跡1」では分類数20,176、ユニークセル数16,631であったのに対し、まるめの処理として、死因ICD-9コードを3桁までとする、かつ日付を月までにする(「追跡2」)ことによりユニークセル数は約半分、同じく死因コード3桁かつ日付を追跡期間(単位:月)でみる(「追跡3」)ことによりさらに半分になり、一意性は減少した。さらに死因情報を除いて日付を追跡期間(単位:月)で見た場合(「追跡4」)では分類数が243、ユニークセル数が0になり一意性が消失した(表3)。

[検討4] ①~④の組み合わせから得られた329パターンについて、分類数、ユニークセル数、分類数に占めるユニークセル数の割合 $S_1/K$ を示した(表4)。

また分類数 $K$ を横軸、ユニークセル数 $S_1$ を縦軸にその分布を示した(図1)。さらに、分類数 $K$ を横軸、分類数に占めるユニークセル数の割合 $S_1/K$ を縦軸にその分布を示した(図2)。分類数が小さい時には分類数に占めるユニークセル数の割合も80%以下に分布するが、分類数の増加とともにユニークセルの割合が急増し、概ね分類数が20,000を超えると80%以上に分布した。すなわち、100,629例全体に対して16,000例程度(16%程度)が一意性のある個体数となり、分類数の増加に比例して、一意性のある個体数が増加した。

### D. 考察

各変数のユニークセル数の確認より、一意性には、変数v0501(10年観察終了日)のように、分類数が大きいことでそれぞれに振り分けられる個体数が少なくなるため生じる一意と、変数v1615(出産人数)において出産人数が20人というように、疫学的にまれな属性の個体が存在したために生じる一意の大きく2パターンが考えられた。前者に

対しては例えば日付データを月までにするなどにより分類数を減らすことで一意性を減少させることが可能であり、後者に対しては一定値以上（以下）については直接表示せず、無限までの片側区間で表示するといった方法により一意性の減少が図られる。

しかしながら今回すべての変数を組み合わせた場合の一意である個体の数は 100,605 であり、これは全レコード数の約 99.98%にあたる。このように大規模なコホートデータにおいては、変数が多くなる（質問項目が多い）ことによる一意性は容易に避けられるものではない。また、本研究に利用した 10 万人規模のデータであるからまるめ処理などによりある程度の一意性の減少がみられるが、規模が小さくなると一意性が上がる可能性も高い。

分類数とユニークセル数の関係から、コホートデータにおいて、変数が増えるほど分類数は増大し、概ね分類数が 20,000 を超えると一意である個体の数も分類数の 80%以上に分布した。一意性を上げないためには、一つのファイルに含む項目数を増やさないう、ファイルを分けて保管することなどが考えられるが、通常、一意性があるものとの前提で対応する必要がある。

死因に関しては、簡単分類を参考とした丸めの方法なども検討する必要がある。

## E. 結論

三府県コホートデータより、各変数、全変数あるいはいくつかの変数の組合せごとに一意性を検討した。三府県コホートデータのような 10 万人規模のデータの場合、分類数が概ね 20,000 を超えると一意性のある個体数は分類数の 80%以上となり、一意性があるものとの前提で対応を考える必要がある。

## F. 研究発表

1. 論文発表
2. 学会発表

いずれもなし

## G. 知的財産権の出願・登録状況

（予定を含む。）

1. 特許取得
2. 実用新案登録
3. その他

いずれもなし

表1 対象となる変数について(グレーの変数は解析対象外)

変数名	属性	内容	定義の相違	凡例など	備考
v0000	数値	通しID	共通	1-100629	府県(v0400)、自治体(v0403)、地域固有ID(v0001)のソート順
v0001	数値	地域固有ID	府県間で異なる	(欠損なし)	
v0100	数値	観察期間-負 フラグ	共通	1: Yes (欠損=)	解析時除外。 追跡開始日(v0500) > 観察終了日(v0521)。 愛知都市地区のみ14例。
v0200	数値	性別	共通	0: 男, 1: 女 (欠損なし)	
v0300	数値	追跡開始年齢	共通	40-99 (欠損なし)	
v0400	数値	府県番号	共通	1: 宮城, 2: 愛知, 3: 大阪	
v0401	数値	地区番号	共通	1: 都市地区, 0: 対照地区	
v0402	数値	地区通し番号	共通	1: 宮城対照, 2: 宮城都市, 3: 愛知対照, 4: 愛知都市, 5: 大阪対照, 6: 大阪都市	
v0403	数値	自治体番号	府県間で異なる	宮城: 1: 仙台, 2: 蒲谷, 3: 田尻; 愛知: 1: 名古屋, 2: 犬山; 大阪: 1: 東成, 2: 能勢, 3: 河南, 4: 熊取	
v0500	SAS日付	追跡開始日	共通	(欠損なし)	
v0501	SAS日付	10年コホート追跡終了日(地域ごとの定義)	共通(ただし日付は地区間で異なる)	宮城: 1999/12/31, 愛知都市地区: 1995/10/31, 愛知対照地区: 1995/8/30, 大阪東成: 1994/10/31, 大阪能勢: 1993/1/31, 大阪河南: 1994/11/30, 大阪熊取: 1995/1/31	
v0502	SAS日付	10年観察終了日(個人ごとの定義)	共通	コホート10年追跡終了日, 転出日, 死亡日のうち一番早い日(欠損なし)	
v0503	数値	10年転帰	共通	3: コホート追跡終了日において生存, 6: コホート追跡終了日までに死亡, 9: コホート追跡終了日までに転出(欠損なし)	
v0511	SAS日付	15年コホート追跡終了日(地域ごとの定義)	共通(ただし日付は地区間で異なる)	宮城: 1999/12/31, 愛知都市地区: 2000/10/31, 愛知対照地区: 2000/6/30, 大阪東成: 1999/10/31, 大阪能勢: 1998/1/31, 大阪河南: 1999/11/30, 大阪熊取: 2000/1/31 (欠損なし)	
v0512	SAS日付	15年観察終了日(個人ごとの定義)	共通	コホート15年追跡終了日, 転出日, 死亡日のうち一番早い日(欠損なし)	
v0513	数値	15年転帰	共通	3: コホート追跡終了日において生存, 6: コホート追跡終了日までに死亡, 9: コホート追跡終了日までに転出(欠損なし)	
v0520	数値	転帰(受領データマ)	共通	3: 生存, 6: 死亡, 9: 転出(欠損なし)	各地域の最終確認時の転帰。解析時は追跡終了日定義の上調整が必要。 愛知は転出(v0520=9)で死因ありのデータが都市地区277例、対照地区73例あり(転出として扱い、死因は無視(若井先生確認済))。
v0521	SAS日付	観察終了日(個人ごとの定義: 受領データマ)	共通	コホート追跡終了日による定義をする前の個人ごとの観察終了日。	愛知のみコホート追跡終了日以後の日付あり(2000/12/31まで)。
v0600	数値	死因ICD-9コード4桁	共通	ICD-9(例: 38.9, 150.1)	基本となる死因変数(ただし、宮城の15年死亡はv0601, v0602とOR結合が必要)。 10年死亡の死因別死亡例はこの変数のみで同定可。
v0601	数値	死因ICD-9コード4桁(宮城15年死亡)	宮城のみ	ICD-9(例: 38.9, 150.1)	宮城15年死亡の死因はv0600(ICD-9)、v0601(ICD-9)、v0602(ICD-10)の3変数に入っている(これら3変数間に重複なし)。 15年死亡の死因別死亡同定にはこれら3変数のOR結合が必要。
v0602	文字	死因ICD-10コード4桁(宮城15年死亡)	宮城のみ	ICD-10(例: C030, I219)	
v0610	文字	死因ICD-9コード	府県間で異なる	ICD-9(元データのママのため、形式が異なる)。	v0600の数値化前データ
v0611	文字	死因ICD-9またはICD-10コード(宮城15年死亡)	宮城のみ	ICD-9とICD-10が混在(元データのママ)	v0601, v0602のICD-9とICD-10に分ける前のデータ
v0620	数値	10年死因不明フラグ	共通	1: Yes (欠損=)	v0600=、かつ v0611="" かつ 10年死亡(v0503=6)。 10年死亡例中19例あり(すべて宮城)。
v0621	数値	15年死因不明フラグ	共通	1: Yes (欠損=)	v0600=、かつ v0611="" かつ 15年死亡(v0513=6)。 15年死亡例中29例あり(宮城28例、大阪東成1例)。
v1000	数値	自覚症状あるか 咳	共通	1: Yes (欠損=)	明示回答ありなら1, それ以外は欠損。 宮城、大阪熊取はv1011と矛盾ありママ。
v1001	数値	自覚症状あるか 痰	共通	1: Yes (欠損=)	明示回答ありなら1, それ以外は欠損。 宮城、大阪熊取はv1011と矛盾ありママ。
v1002	数値	自覚症状あるか 血痰	共通	1: Yes (欠損=)	明示回答ありなら1, それ以外は欠損。 宮城、大阪熊取はv1011と矛盾ありママ。
v1003	数値	自覚症状あるか 動悸	共通	1: Yes (欠損=)	明示回答ありなら1, それ以外は欠損。 宮城、大阪熊取はv1011と矛盾ありママ。
v1004	数値	自覚症状あるか 息切れ	共通	1: Yes (欠損=)	大阪東成は欠損。 明示回答ありなら1, それ以外は欠損。 宮城、大阪熊取はv1011と矛盾ありママ。
v1005	数値	自覚症状あるか 食欲不振	共通	1: Yes (欠損=)	大阪東成は欠損。 明示回答ありなら1, それ以外は欠損。 宮城、大阪熊取はv1011と矛盾ありママ。
v1006	数値	自覚症状あるか 便秘	共通	1: Yes (欠損=)	大阪東成は欠損。 明示回答ありなら1, それ以外は欠損。 宮城、大阪熊取はv1011と矛盾ありママ。
v1007	数値	自覚症状あるか 不眠	共通	1: Yes (欠損=)	愛知、大阪熊取、大阪河南は欠損。 明示回答ありなら1, それ以外は欠損。
v1008	数値	自覚症状あるか やせてきた	共通	1: Yes (欠損=)	明示回答ありなら1, それ以外は欠損。 宮城、大阪熊取はv1011と矛盾ありママ。
v1009	数値	自覚症状あるか 疲れやすい	共通	1: Yes (欠損=)	明示回答ありなら1, それ以外は欠損。 宮城、大阪熊取はv1011と矛盾ありママ。
v1010	数値	自覚症状あるか 喘息の発作	共通	1: Yes (欠損=)	宮城、大阪東成は欠損。 明示回答ありなら1, それ以外は欠損。
v1011	数値	自覚症状あるか なし	共通	1: Yes (欠損=)	「なし」回答ありの場合1, それ以外は欠損。 宮城、大阪熊取はv1000-v1009と矛盾ありママ。
v1100	数値	既往 高血圧	一部相違	1: 現在あり, 2: 以前あった, 3: 今までなし(欠損=)	宮城、大阪東成、大阪河南は疾病番号選択のみのため、回答ありを便宜上1に割り振り回答なしは欠損。 宮城、大阪熊取はv1120との矛盾ママ。
v1101	数値	既往 心臓病	一部相違	1: 現在あり, 2: 以前あった, 3: 今までなし(欠損=)	宮城、大阪東成、大阪河南は疾病番号選択のみのため、回答ありを便宜上1に割り振り回答なしは欠損。 宮城、大阪熊取はv1120との矛盾ママ。
v1102	数値	既往 脳卒中	一部相違	1: 現在あり, 2: 以前あった, 3: 今までなし(欠損=)	宮城、大阪東成、大阪河南は疾病番号選択のみのため、回答ありを便宜上1に割り振り回答なしは欠損。 宮城、大阪熊取はv1120との矛盾ママ。
v1103	数値	既往 糖尿病	一部相違	1: 現在あり, 2: 以前あった, 3: 今までなし(欠損=)	宮城、大阪東成、大阪河南は疾病番号選択のみのため、回答ありを便宜上1に割り振り回答なしは欠損。 宮城、大阪熊取はv1120との矛盾ママ。
v1104	数値	既往 肝臓病	一部相違	1: 現在あり, 2: 以前あった, 3: 今までなし(欠損=)	宮城、大阪東成、大阪河南は疾病番号選択のみのため、回答ありを便宜上1に割り振り回答なしは欠損。 宮城、大阪熊取はv1120との矛盾ママ。
v1105	数値	既往 腎臓病	一部相違	1: 現在あり, 2: 以前あった, 3: 今までなし(欠損=)	宮城、大阪東成、大阪河南は欠損。

変数名	属性	内容	定義の相違	凡例など	備考
v1106	数値	既往 胃十二指腸潰瘍	一部相違	1: 現在あり, 2: 以前あった, 3: 今までなし (欠損=)	大阪能勢は欠損。 宮城、大阪東成、大阪河内は疾病番号選択のみのため、回答ありを便宜上1に割り振り回答なしは欠損。 宮城、大阪能取はv1120との矛盾ママ。
v1107	数値	既往 胆石	一部相違	1: 現在あり, 2: 以前あった, 3: 今までなし (欠損=)	大阪東成、大阪能勢、大阪河内は欠損。 宮城、大阪河内は疾病番号選択のみのため、回答ありを便宜上1に割り振り回答なしは欠損。 宮城、大阪能取はv1120との矛盾ママ。
v1108	数値	既往 虫垂炎	一部相違	1: Yes (欠損=)	宮城のみ。 疾病番号選択のみのため、回答ありを便宜上1に割り振り回答なしは欠損。
v1109	数値	既往 結核ろく膜炎	一部相違	1: 現在あり, 2: 以前あった, 3: 今までなし (欠損=)	宮城、大阪東成、大阪河内は疾病番号選択のみのため、回答ありを便宜上1に割り振り回答なしは欠損。 大阪能勢は結核と肋膜炎の結合(選択肢1,2は1優先で結核と肋膜炎のOR結合、選択肢3は1,2でない場合に結核と肋膜炎のAND結合)。 宮城、大阪能取はv1120との矛盾ママ。
v1110	数値	既往 肺炎	一部相違	1: 現在あり, 2: 以前あった, 3: 今までなし (欠損=)	宮城、大阪東成、大阪河内は疾病番号選択のみのため、回答ありを便宜上1に割り振り回答なしは欠損。 宮城、大阪能取はv1120との矛盾ママ。
v1111	数値	既往 喘息	一部相違	1: 現在あり, 2: 以前あった, 3: 今までなし (欠損=)	宮城、大阪東成、大阪河内は疾病番号選択のみのため、回答ありを便宜上1に割り振り回答なしは欠損。 宮城、大阪能取はv1120との矛盾ママ。
v1112	数値	既往 慢性気管支炎	一部相違	1: 現在あり, 2: 以前あった, 3: 今までなし (欠損=)	宮城、大阪東成、大阪河内は疾病番号選択のみのため、回答ありを便宜上1に割り振り回答なしは欠損。 宮城、大阪能取はv1120との矛盾ママ。
v1113	数値	既往 肺炎腫	一部相違	1: 現在あり, 2: 以前あった, 3: 今までなし (欠損=)	宮城、大阪東成、大阪河内は疾病番号選択のみのため、回答ありを便宜上1に割り振り回答なしは欠損。 宮城、大阪能取はv1120との矛盾ママ。
v1114	数値	既往 じん肺	一部相違	1: 現在あり, 2: 以前あった, 3: 今までなし (欠損=)	宮城、大阪東成、大阪河内は疾病番号選択のみのため、回答ありを便宜上1に割り振り回答なしは欠損。 宮城、大阪能取はv1120との矛盾ママ。
v1115	数値	既往 貧血	共通	1: Yes (欠損=)	大阪東成、大阪河内のみ。 疾病番号選択のみのため、回答ありを便宜上1に割り振り回答なしは欠損。
v1116	数値	既往 アレルギー性鼻炎	共通	1: 現在あり, 2: 以前あった, 3: 今までなし (欠損=)	大阪能勢のみ。
v1117	数値	既往 尋麻疹	共通	1: 現在あり, 2: 以前あった, 3: 今までなし (欠損=)	大阪能勢のみ。
v1118	数値	既往 湿疹・皮膚炎	共通	1: 現在あり, 2: 以前あった, 3: 今までなし (欠損=)	大阪能勢のみ。
v1119	数値	既往 痛風	共通	1: Yes (欠損=)	宮城のみ
v1120	数値	既往 なし	共通	1: Yes (欠損=)	「なし」回答ありの場合1、それ以外は欠損。 大阪能勢欠損。 宮城、大阪能取はv1100~v1119との矛盾ママ。
v1200	数値	身長 (cm)	共通	(欠損=)	外れ値ママ
v1201	数値	体重 (kg)	共通	(欠損=)	外れ値ママ
v1202	数値	BMI連続変数	共通	(欠損=)	$v1201/v1200 \times 10000$ で計算。v1200またはv1201が0、999.9、欠損のいずれかの場合は欠損。
v1300	数値	胸部レントゲン検査受診有無	一部相違	1: Yes, 0: No (欠損=)	宮城、愛知、大阪能取は「胸部のレントゲン検査(結核検診)」。 宮城は「昨年(昭和58年)1年間に」、大阪東成、大阪河内は「この1年のあいだに」、それ以外は単に「受けたことがありますか?」と聞いて下位で時期質問。
v1301	数値	胸部レントゲン検査受診有無(1年以内)	一部相違	1: Yes (欠損=)	宮城、大阪東成、大阪河内はv1300=1、愛知はv1300=1かつ昭和60年受診、大阪能勢はv1300=1かつ昭和57年度受診、大阪能取はv1300=1かつ昭和59年受診。 注: ベースライン調査は、宮城は昭和59年1~2月、愛知名古屋は昭和60年10~11月、愛知大山は昭和60年7~8月、大阪能勢は昭和58年2月、大阪能取は昭和60年2~3月、大阪河内は昭和59年11~12月、大阪東成は昭和59年10~11月に実施。
v1302	数値	胸部レントゲン検査受診有無(1年より前~2年以内)	一部相違	1: Yes (欠損=)	宮城、大阪東成、大阪河内は欠損。愛知はv1300=1かつ昭和59年受診、大阪能勢はv1300=1かつ昭和56年度受診、大阪能取はv1300=1かつ昭和58年受診。 注: ベースライン調査は、宮城は昭和59年1~2月、愛知名古屋は昭和60年10~11月、愛知大山は昭和60年7~8月、大阪能勢は昭和58年2月、大阪能取は昭和60年2~3月、大阪河内は昭和59年11~12月、大阪東成は昭和59年10~11月に実施。
v1303	数値	胸部レントゲン検査受診有無(2年より前)	一部相違	1: Yes (欠損=)	宮城、大阪東成、大阪河内は欠損。愛知はv1300=1かつ昭和59年より前受診、大阪能勢はv1300=1かつ昭和56年度より前受診、大阪能取はv1300=1かつ昭和58年より前受診。 注: ベースライン調査は、宮城は昭和59年1~2月、愛知名古屋は昭和60年10~11月、愛知大山は昭和60年7~8月、大阪能勢は昭和58年2月、大阪能取は昭和60年2~3月、大阪河内は昭和59年11~12月、大阪東成は昭和59年10~11月に実施。
v1304	数値	胸部レントゲン検査受診場所(住民健診)	一部相違	1: Yes (欠損=)	大阪能取は「集団検診で(例えば成人病予防協会検診)」、大阪東成は欠損。 v1300=1かつ当該場所の受診有無・受診年いずれか回答ありなら1。
v1305	数値	胸部レントゲン検査受診場所(保健所・保健センター)	一部相違	1: Yes (欠損=)	大阪東成は「保健所や小学校の市民検診」。 v1300=1かつ当該場所の受診有無・受診年いずれか回答ありなら1。
v1306	数値	胸部レントゲン検査受診場所(職場)	一部相違	1: Yes (欠損=)	v1300=1かつ当該場所の受診有無・受診年いずれか回答ありなら1。
v1307	数値	胸部レントゲン検査受診場所(病院など)	一部相違	1: Yes (欠損=)	v1300=1かつ当該場所の受診有無・受診年いずれか回答ありなら1。
v1308	数値	胸部レントゲン検査受診場所(農協)	一部相違	1: Yes (欠損=)	大阪能取と大阪河内のみ。 v1300=1かつ当該場所の受診有無・受診年いずれか回答ありなら1。
v1309	数値	胸部レントゲン検査受診場所(その他)	一部相違	1: Yes (欠損=)	v1300=1かつ当該場所の受診有無・受診年いずれか回答ありなら1。

変数名	属性	内容	定義の相違	凡例など	備考
v1400	数値	胃レントゲン検査受診有無	一部相違	1: Yes, 0: No (欠損=)	宮城、愛知対照、大阪熊取は「胃のレントゲン検査(胃腸病検査)」; 大阪東成、大阪河南は「胃(ガン)検査」。 宮城は「昨年(昭和58年)1年間に」、大阪東成、大阪河南は「この1年のあいだに」、それ以外は単に「受けたことがありますか」と聞いて下位で時期質問。
v1401	数値	胃レントゲン検査受診有無(1年以内)	一部相違	1: Yes (欠損=)	宮城、大阪東成、大阪河南はv1400=1、愛知はv1400=1かつ昭和60年受診、大阪熊取はv1400=1かつ昭和57年度受診、大阪熊取はv1400=1かつ昭和59年受診。 注: ベースライン調査は、宮城は昭和59年1~2月、愛知名古屋は昭和60年10~11月、愛知犬山は昭和60年7~8月、大阪熊取は昭和58年2月、大阪熊取は昭和60年2~3月、大阪河南は昭和59年11~12月、大阪東成は昭和59年10~11月に実施。
v1402	数値	胃レントゲン検査受診有無(1年より前~2年以内)	一部相違	1: Yes (欠損=)	宮城、大阪東成、大阪河南は欠損。愛知はv1400=1かつ昭和59年受診、大阪熊取はv1400=1かつ昭和56年度受診、大阪熊取はv1400=1かつ昭和59年受診。 注: ベースライン調査は、宮城は昭和59年1~2月、愛知名古屋は昭和60年10~11月、愛知犬山は昭和60年7~8月、大阪熊取は昭和58年2月、大阪熊取は昭和60年2~3月、大阪河南は昭和59年11~12月、大阪東成は昭和59年10~11月に実施。
v1403	数値	胃レントゲン検査受診有無(2年より前)	一部相違	1: Yes (欠損=)	宮城、大阪東成、大阪河南は欠損。愛知はv1400=1かつ昭和59年より前受診、大阪熊取はv1400=1かつ昭和56年度より前受診、大阪熊取はv1400=1かつ昭和58年より前受診。 注: ベースライン調査は、宮城は昭和59年1~2月、愛知名古屋は昭和60年10~11月、愛知犬山は昭和60年7~8月、大阪熊取は昭和58年2月、大阪熊取は昭和60年2~3月、大阪河南は昭和59年11~12月、大阪東成は昭和59年10~11月に実施。
v1404	数値	胃レントゲン検査受診場所(住民健診)	一部相違	1: Yes (欠損=)	大阪熊取は「集団検診で(例えば成人病予防協会検診車)」; 愛知名古屋は「市医師会健診センターまたは休日急病診療所(検診車)」。 v1400=1かつ当該場所の受診有無・受診年いずれか回答ありなら1。
v1405	数値	胃レントゲン検査受診場所(保健所・保健センター)	一部相違	1: Yes (欠損=)	愛知名古屋と大阪東成のみ。 v1400=1かつ当該場所の受診有無・受診年いずれか回答ありなら1。
v1406	数値	胃レントゲン検査受診場所(職場)	一部相違	1: Yes (欠損=)	v1400=1かつ当該場所の受診有無・受診年いずれか回答ありなら1。
v1407	数値	胃レントゲン検査受診場所(病院など)	一部相違	1: Yes (欠損=)	v1400=1かつ当該場所の受診有無・受診年いずれか回答ありなら1。
v1408	数値	胃レントゲン検査受診場所(農協)	一部相違	1: Yes (欠損=)	大阪熊取と大阪河南のみ。 v1400=1かつ当該場所の受診有無・受診年いずれか回答ありなら1。
v1409	数値	胃レントゲン検査受診場所(その他)	一部相違	1: Yes (欠損=)	v1400=1かつ当該場所の受診有無・受診年いずれか回答ありなら1。
v1410	数値	胃レントゲン検査受診回数(5年間)	共通	(欠損=)	大阪東成、大阪熊取、大阪河南は欠損。
v1500	数値	健康診査または血圧検査受診有無	一部相違	1: Yes, 0: No (欠損=)	愛知は欠損。 宮城、大阪熊取、大阪熊取は「血圧の検査」、大阪東成、大阪河南は「健康診査(血圧測定や尿検査)」。 宮城は「昨年(昭和58年)1年間に」、大阪東成、大阪河南は「この1年のあいだに」、それ以外は単に「受けたことがありますか」と聞いて下位で時期質問。
v1501	数値	健康診査または血圧検査受診有無(1年以内)	一部相違	1: Yes (欠損=)	愛知は欠損。 宮城、大阪東成、大阪河南はv1500=1、大阪熊取はv1500=1かつ昭和57年度受診、大阪熊取はv1500=1かつ昭和59年受診。 注: ベースライン調査は、宮城は昭和59年1~2月、愛知名古屋は昭和60年10~11月、愛知犬山は昭和60年7~8月、大阪熊取は昭和58年2月、大阪熊取は昭和60年2~3月、大阪河南は昭和59年11~12月、大阪東成は昭和59年10~11月に実施。
v1502	数値	健康診査または血圧検査受診有無(1年より前~2年以内)	一部相違	1: Yes (欠損=)	宮城、愛知、大阪東成、大阪河南は欠損。 大阪熊取はv1500=1かつ昭和56年度受診、大阪熊取はv1500=1かつ昭和58年受診。 注: ベースライン調査は、宮城は昭和59年1~2月、愛知名古屋は昭和60年10~11月、愛知犬山は昭和60年7~8月、大阪熊取は昭和58年2月、大阪熊取は昭和60年2~3月、大阪河南は昭和59年11~12月、大阪東成は昭和59年10~11月に実施。
v1503	数値	健康診査または血圧検査受診有無(2年より前)	一部相違	1: Yes (欠損=)	宮城、愛知、大阪東成、大阪河南は欠損。 大阪熊取はv1500=1かつ昭和56年度より前受診、大阪熊取はv1500=1かつ昭和59年より前受診。 注: ベースライン調査は、宮城は昭和59年1~2月、愛知名古屋は昭和60年10~11月、愛知犬山は昭和60年7~8月、大阪熊取は昭和58年2月、大阪熊取は昭和60年2~3月、大阪河南は昭和59年11~12月、大阪東成は昭和59年10~11月に実施。
v1504	数値	健康診査または血圧検査受診場所(住民健診)	一部相違	1: Yes (欠損=)	大阪東成は「地区健康相談」。 v1500=1かつ当該場所の受診有無・受診年いずれか回答ありなら1。
v1505	数値	健康診査または血圧検査受診場所(保健所・保健センター)	一部相違	1: Yes (欠損=)	大阪東成は「保健所や小学校の市民検診」。 v1500=1かつ当該場所の受診有無・受診年いずれか回答ありなら1。
v1506	数値	健康診査または血圧検査受診場所(職場)	一部相違	1: Yes (欠損=)	v1500=1かつ当該場所の受診有無・受診年いずれか回答ありなら1。
v1507	数値	健康診査または血圧検査受診場所(病院など)	一部相違	1: Yes (欠損=)	v1500=1かつ当該場所の受診有無・受診年いずれか回答ありなら1。
v1508	数値	健康診査または血圧検査受診場所(農協)	一部相違	1: Yes (欠損=)	大阪熊取と大阪河南のみ。 v1500=1かつ当該場所の受診有無・受診年いずれか回答ありなら1。
v1509	数値	健康診査または血圧検査受診場所(その他)	一部相違	1: Yes (欠損=)	v1500=1かつ当該場所の受診有無・受診年いずれか回答ありなら1。

変数名	属性	内容	定義の相違	凡例など	備考
v1600	数値	子宮がん検診受診有無	一部相違	1: Yes, 0: No (欠損=)	女性のみ定義。宮城は「昨年(昭和58年)1年間」、愛知は「昨年(昭和59年)1年間に」、大阪東成、大阪河内は「この年のあいだに」、それ以外は単に「受けたことがありますか」と聞いて下位で時期質問。
v1601	数値	子宮がん検診受診有無(昨年)	一部相違	1: Yes (欠損=)	女性のみ定義。宮城、愛知、大阪東成、大阪河内はv1600=1。大阪能勢はv1600=1かつ昭和57年度受診、大阪熊取はv1600=1かつ昭和59年受診。 注: ベースライン調査は、宮城は昭和59年1~2月、愛知名古屋は昭和60年10~11月、愛知犬山は昭和60年7~8月、大阪能勢は昭和58年2月、大阪熊取は昭和60年2~3月、大阪河内は昭和59年11~12月、大阪東成は昭和59年10~11月に実施。
v1602	数値	子宮がん検診受診有無(一昨年)	一部相違	1: Yes (欠損=)	女性のみ定義。宮城、愛知、大阪東成、大阪河内は欠損。大阪能勢はv1600=1かつ昭和56年度受診、大阪熊取はv1600=1かつ昭和58年受診。 注: ベースライン調査は、宮城は昭和59年1~2月、愛知名古屋は昭和60年10~11月、愛知犬山は昭和60年7~8月、大阪能勢は昭和58年2月、大阪熊取は昭和60年2~3月、大阪河内は昭和59年11~12月、大阪東成は昭和59年10~11月に実施。
v1603	数値	子宮がん検診受診有無(一昨年より前)	一部相違	1: Yes (欠損=)	女性のみ定義。宮城、愛知、大阪東成、大阪河内はv1600と同義。大阪能勢はv1600=1かつ昭和56年度より前受診、大阪熊取はv1600=1かつ昭和58年より前受診。 注: ベースライン調査は、宮城は昭和59年1~2月、愛知名古屋は昭和60年10~11月、愛知犬山は昭和60年7~8月、大阪能勢は昭和58年2月、大阪熊取は昭和60年2~3月、大阪河内は昭和59年11~12月、大阪東成は昭和59年10~11月に実施。
v1604	数値	子宮がん検診受診場所(検診車・住民検診)	一部相違	1: Yes (欠損=)	女性のみ定義。大阪東成、大阪能勢、大阪熊取は欠損。大阪河内は「河内町が行った集団検診」。宮城、愛知は「検診車」。v1600=1かつ当該場所の受診有無・受診年いずれか回答ありなら1。
v1605	数値	子宮がん検診受診場所(保健所・保健センター)	一部相違	1: Yes (欠損=)	女性のみ定義。大阪東成は「保健所の検診車」。宮城と愛知は欠損。v1600=1かつ当該場所の受診有無・受診年いずれか回答ありなら1。
v1606	数値	子宮がん検診受診場所(職場)	一部相違	1: Yes (欠損=)	女性のみ定義。v1600=1かつ当該場所の受診有無・受診年いずれか回答ありなら1。
v1607	数値	子宮がん検診受診場所(病院など)	一部相違	1: Yes (欠損=)	女性のみ定義。v1600=1かつ当該場所の受診有無・受診年いずれか回答ありなら1。
v1608	数値	子宮がん検診受診場所(東成母子会自己採取)	一部相違	1: Yes (欠損=)	女性のみ定義。大阪東成のみ。v1600=1かつ当該場所の受診有無・受診年いずれか回答ありなら1。
v1609	数値	子宮がん検診受診場所(その他)	一部相違	1: Yes (欠損=)	女性のみ定義。v1600=1かつ当該場所の受診有無・受診年いずれか回答ありなら1。
v1610	数値	初経年齢	共通	(欠損=)	女性のみ定義。大阪は欠損。
v1611	数値	現在月経有り	共通	1: Yes, 0: No (欠損=)	女性のみ定義。大阪は欠損。
v1612	数値	自然閉経年齢	共通	(欠損=)	女性のみ定義。大阪は欠損。V1611に関わらず定義。宮城はv1613との重複回答あり。
v1613	数値	手術閉経年齢	共通	(欠損=)	女性のみ定義。大阪は欠損。V1611に関わらず定義。宮城はv1613との重複回答あり。
v1614	数値	出産経験あるか	共通	1: Yes, 0: No (欠損=)	女性のみ定義。
v1615	数値	出産人数	共通	(欠損=)	愛知都市地区は欠損。女性かつ「出産経験=なし以外」は定義(出産経験不明例の出産人数ママ、出産あり例の0人ママ)。女性かつ「出産経験=なし」は0人。
v1616	数値	初産年齢	共通	(欠損=)	愛知都市地区は欠損。女性かつ「出産経験=なし以外」は定義(出産経験不明例の出産人数ママ、出産あり例の0人ママ)。女性かつ「出産経験=なし」は0人。
v1700	数値	健康保険カテゴリ	一部相違	1: 国民健康保険, 2: 健康保険(政府管掌), 3: 健康保険(組合等), 4: 共済, 5: その他の保険, 6: 健康保険(政府管掌・組合), 9: 不明 (欠損=)	選択肢6は大阪対照地区のみ(2と3は欠損)。
v1701	数値	健康保険本人・家族別	共通	1: 本人, 2: 世帯主, 3: 家族 (欠損=)	
v1800	数値	ご飯1日何杯	共通	(欠損=)	宮城と愛知は2杯。大阪は1杯。
v1801	数値	パン摂取頻度カテゴリ	共通	1: 食べない, 2: 月に1-2日, 3: 週に1-2日, 4: 週に3-4日, 5: ほとんど毎日 (欠損=)	
v1802	数値	肉摂取頻度カテゴリ	共通	1: 食べない, 2: 月に1-2日, 3: 週に1-2日, 4: 週に3-4日, 5: ほとんど毎日 (欠損=)	
v1803	数値	魚類摂取頻度カテゴリ	共通	1: 食べない, 2: 月に1-2日, 3: 週に1-2日, 4: 週に3-4日, 5: ほとんど毎日 (欠損=)	
v1804	数値	卵摂取頻度カテゴリ	共通	1: 食べない, 2: 月に1-2日, 3: 週に1-2日, 4: 週に3-4日, 5: ほとんど毎日 (欠損=)	
v1805	数値	牛乳摂取頻度カテゴリ	共通	1: 食べない, 2: 月に1-2日, 3: 週に1-2日, 4: 週に3-4日, 5: ほとんど毎日 (欠損=)	
v1806	数値	緑黄色野菜摂取頻度カテゴリ	共通	1: 食べない, 2: 月に1-2日, 3: 週に1-2日, 4: 週に3-4日, 5: ほとんど毎日 (欠損=)	
v1807	数値	その他の野菜摂取頻度カテゴリ	共通	1: 食べない, 2: 月に1-2日, 3: 週に1-2日, 4: 週に3-4日, 5: ほとんど毎日 (欠損=)	
v1808	数値	果物摂取頻度カテゴリ	共通	1: 食べない, 2: 月に1-2日, 3: 週に1-2日, 4: 週に3-4日, 5: ほとんど毎日 (欠損=)	
v1809	数値	みそ汁摂取頻度カテゴリ	共通	1: 食べない, 2: 月に1-2日, 3: 週に1-2日, 4: 週に3-4日, 5: ほとんど毎日 (欠損=)	
v1810	数値	漬物摂取頻度カテゴリ	共通	1: 食べない, 2: 月に1-2日, 3: 週に1-2日, 4: 週に3-4日, 5: ほとんど毎日 (欠損=)	
v1811	数値	インスタント食品摂取頻度カテゴリ	共通	1: 食べない, 2: 月に1-2日, 3: 週に1-2日, 4: 週に3-4日, 5: ほとんど毎日 (欠損=)	宮城は欠損。
v1900	数値	緑茶摂取頻度カテゴリ	共通	1: 飲まない, 2: ときどき, 3: 1日に1-2杯, 4: 1日に3-4杯, 5: 1日5杯以上 (欠損=)	大阪能勢は「1.ほとんど飲まない」
v1901	数値	紅茶摂取頻度カテゴリ	共通	1: 飲まない, 2: ときどき, 3: 1日に1-2杯, 4: 1日に3-4杯, 5: 1日5杯以上 (欠損=)	大阪能勢は「1.ほとんど飲まない」
v1902	数値	コーヒー(非インスタント)摂取頻度カテゴリ	共通	1: 飲まない, 2: ときどき, 3: 1日に1-2杯, 4: 1日に3-4杯, 5: 1日5杯以上 (欠損=)	大阪能勢は「1.ほとんど飲まない」 宮城は欠損。
v1903	数値	インスタントコーヒー摂取頻度カテゴリ	共通	1: 飲まない, 2: ときどき, 3: 1日に1-2杯, 4: 1日に3-4杯, 5: 1日5杯以上 (欠損=)	大阪能勢は「1.ほとんど飲まない」 宮城は欠損。
v1904	数値	コーヒー(豆またはインスタント)摂取頻度カテゴリ	一部相違	1: 飲まない, 2: ときどき, 3: 1日に1-2杯, 4: 1日に3-4杯, 5: 1日5杯以上 (欠損=)	宮城のみオリジナル変数。 他はv1902とv1903のOR結合(摂取多い選択肢優先で、選択肢1はAND結合)で定義。
v2000	数値	飲酒習慣カテゴリ	一部相違	1: ほとんど毎日, 2: ときどき, 3: 集まりのときだけ, 4: 以前はよく飲んだが最近飲まない, 5: 以前から飲まない (欠損=)	大阪能勢は選択肢4はなし
v2100	数値	喫煙習慣カテゴリ	一部相違	1: 吸っている, 2: やめた, 3: 吸わない, 4: 時々吸っている (欠損=)	選択肢4は大阪能勢のみ。
v2101	数値	喫煙開始年齢 連続変数	共通	(欠損=)	宮城、大阪東成、大阪熊取は生涯喫煙以外の回答ママ。 年齢との比較など論理チェックせず(10未満あり)。 99はママ(愛知)。

変数名	属性	内容	定義の相違	凡例など	備考
v2102	数値	喫煙本数/日 連続変数	共通	(欠損=)	99はママ(愛知、大阪)。 宮城、大阪東成、大阪熊取は生涯喫煙以外の回答ママ。
v2103	数値	禁煙年齢 連続変数	共通	(欠損=)	年齢との比較など論理チェックせず。 宮城、大阪東成、大阪熊取は過去喫煙以外の回答ママ。 99はママ(愛知)。
v2104	数値	吸い込むか	共通	1: Yes, 0: No (欠損=)	宮城、大阪東成、大阪熊取は生涯喫煙以外の回答ママ。
v2200	数値	父親の既往 脳卒中	共通	1: Yes (欠損=)	
v2201	数値	父親の既往 心臓病	共通	1: Yes (欠損=)	
v2202	数値	父親の既往 がん	共通	1: Yes (欠損=)	
v2203	数値	父親の既往 上記3つなし	共通	1: Yes (欠損=)	宮城、大阪熊取はv2200, v2201, v2202との矛盾ママ。
v2204	数値	父親の既往 知らない	共通	1: Yes (欠損=)	大阪熊取は欠損。宮城、大阪熊取はv2200, v2201, v2202との矛盾ママ。
v2205	数値	父親の既往 がん 食道	共通	1: Yes (欠損=)	宮城、大阪東成、大阪熊取はv2202との矛盾ママ。
v2206	数値	父親の既往 がん 胃	共通	1: Yes (欠損=)	宮城、大阪東成、大阪熊取はv2202との矛盾ママ。
v2207	数値	父親の既往 がん 大腸	共通	1: Yes (欠損=)	大阪熊取は欠損。宮城、大阪東成、大阪熊取はv2202との矛盾ママ。
v2208	数値	父親の既往 がん 肝臓	共通	1: Yes (欠損=)	大阪東成、宮城、大阪熊取はv2202との矛盾ママ。
v2209	数値	父親の既往 がん 肺	共通	1: Yes (欠損=)	大阪東成、宮城、大阪熊取はv2202との矛盾ママ。
v2210	数値	父親の既往 がん 膀胱	共通	1: Yes (欠損=)	大阪東成、宮城、大阪熊取はv2202との矛盾ママ。
v2211	数値	父親の既往 がん 白血病	共通	1: Yes (欠損=)	大阪東成、宮城、大阪熊取はv2202との矛盾ママ。
v2212	数値	父親の既往 がん その他	共通	1: Yes (欠損=)	大阪東成、宮城、大阪熊取はv2202との矛盾ママ。
v2300	数値	母親の既往 脳卒中	共通	1: Yes (欠損=)	
v2301	数値	母親の既往 心臓病	共通	1: Yes (欠損=)	
v2302	数値	母親の既往 がん	共通	1: Yes (欠損=)	
v2303	数値	母親の既往 上記3つなし	共通	1: Yes (欠損=)	宮城、大阪熊取はv2300, v2301, v2302との矛盾ママ。
v2304	数値	母親の既往 知らない	共通	1: Yes (欠損=)	大阪熊取は欠損。宮城、大阪熊取はv2300, v2301, v2302との矛盾ママ。
v2305	数値	母親の既往 がん 食道	共通	1: Yes (欠損=)	宮城、大阪東成、大阪熊取はv2302との矛盾ママ。
v2306	数値	母親の既往 がん 胃	共通	1: Yes (欠損=)	宮城、大阪東成、大阪熊取はv2302との矛盾ママ。
v2307	数値	母親の既往 がん 大腸	共通	1: Yes (欠損=)	大阪熊取は欠損。宮城、大阪東成、大阪熊取はv2302との矛盾ママ。
v2308	数値	母親の既往 がん 肝臓	共通	1: Yes (欠損=)	宮城、大阪東成、大阪熊取はv2302との矛盾ママ。
v2309	数値	母親の既往 がん 肺	共通	1: Yes (欠損=)	宮城、大阪東成、大阪熊取はv2302との矛盾ママ。
v2310	数値	母親の既往 がん 膀胱	共通	1: Yes (欠損=)	大阪熊取は欠損。宮城、大阪東成、大阪熊取はv2302との矛盾ママ。
v2311	数値	母親の既往 がん 白血病	共通	1: Yes (欠損=)	宮城、大阪東成、大阪熊取はv2302との矛盾ママ。
v2312	数値	母親の既往 がん 子宮	共通	1: Yes (欠損=)	宮城、大阪東成、大阪熊取はv2302との矛盾ママ。
v2313	数値	母親の既往 がん 乳房	共通	1: Yes (欠損=)	宮城、大阪東成、大阪熊取はv2302との矛盾ママ。
v2314	数値	母親の既往 がん その他	共通	1: Yes (欠損=)	宮城、大阪東成、大阪熊取はv2302との矛盾ママ。
v2400	数値	小学生の頃の父親喫煙有無	一部相違	1: 吸っている, 2: やめた, 3: 吸わない, 4: 知らない(欠損=)	大阪熊取は時期を特定せず「お父さん(実父)はすべておられますか」 選択肢2は能勢のみ。愛知は受領データ選択肢1(吸う)2(吸わない)0だったのでそれぞれ1.3.(欠損)に分配。
v2401	数値	小学生の頃の母親喫煙有無	一部相違	1: 吸っている, 2: やめた, 3: 吸わない, 4: 知らない(欠損=)	大阪熊取は時期を特定せず「お母さん(実母)はすべておられますか」 選択肢2は能勢のみ。愛知は受領データ選択肢1(吸う)2(吸わない)0だったのでそれぞれ1.3.(欠損)に分配。
v2500	数値	現在同居家族喫煙あり	共通	1: Yes, 0: No (欠損=)	
v2501	数値	現在夫喫煙あり	共通	1: Yes (欠損=)	女性のみ定義。宮城、大阪東成、大阪熊取はv2500=1以外でも回答あり。
v2502	数値	現在妻喫煙あり	共通	1: Yes (欠損=)	男性のみ定義。宮城、大阪東成、大阪熊取はv2500=1以外でも回答あり。
v2503	数値	現在父喫煙あり	共通	1: Yes (欠損=)	宮城、大阪東成、大阪熊取はv2500=1以外でも回答あり。
v2504	数値	現在母喫煙あり	共通	1: Yes (欠損=)	宮城、大阪東成、大阪熊取はv2500=1以外でも回答あり。
v2505	数値	現在子喫煙あり	共通	1: Yes (欠損=)	宮城、大阪東成、大阪熊取はv2500=1以外でも回答あり。
v2506	数値	現在その他喫煙あり	共通	1: Yes (欠損=)	宮城、大阪東成、大阪熊取はv2500=1以外でも回答あり。
v2600	数値	暖房 なし	共通	1: Yes (欠損=)	
v2601	数値	暖房 電気(複数回答)	共通	1: Yes (欠損=)	宮城、大阪熊取はv2600=1でも回答あり。
v2602	数値	暖房 エアコン(複数回答)	共通	1: Yes (欠損=)	宮城、大阪熊取はv2600=1でも回答あり。
v2603	数値	暖房 クリーンヒーター(複数回答)	共通	1: Yes (欠損=)	宮城、大阪熊取はv2600=1でも回答あり。
v2604	数値	暖房 煤突きストーブ(複数回答)	共通	1: Yes (欠損=)	宮城、大阪熊取はv2600=1でも回答あり。
v2605	数値	暖房 煤突なしストーブ(複数回答)	共通	1: Yes (欠損=)	宮城、大阪熊取はv2600=1でも回答あり。
v2606	数値	暖房 炭(複数回答)	共通	1: Yes (欠損=)	宮城、大阪熊取はv2600=1でも回答あり。
v2700	数値	住居が火の通る道路に面しているか	共通	1: Yes, 0: No (欠損=)	
v2800	数値	生まれてからいまのところに住んでいるか	共通	1: 生まれてから住んでいる, 0: 他の市区町村から引っ越し(欠損=)	
v2801	数値	転入何年前か	一部相違	(欠損=)	愛知は「居住年数(計算)」として受領データセットにあり。 年齢との比較など論理チェックせず0以上ならママ。 大阪は欠損。
v2802	数値	転入年の元号	共通	1: 明治, 2: 大正, 3: 昭和(欠損=)	宮城は欠損。v2800, v2803との整合性チェックせずママ。
v2803	数値	転入年(和暦)	共通	(欠損=)	宮城は欠損。v2800, v2802との整合性チェックせず年数1以上ならママ。
v2804	数値	転入年(西暦)	共通	(欠損=)	宮城は欠損。v2802, v2803から単純計算(実在元号かつ調査年以前の場合のみ)。
v2901	数値	職業 病院・診療所	一部相違	1: Yes (欠損=)	大阪熊取は欠損。宮城は2段階質問形式の下位選択肢(場所一病院・診療所・歯科診療所・歯科技工所)。
v2902	数値	職業 美容院・理髪店	一部相違	1: Yes (欠損=)	大阪熊取は欠損。宮城は2段階質問形式の下位選択肢(場所一美容院・理髪店)。
v2903	数値	職業 紡績・織布・タオル工場	一部相違	1: Yes (欠損=)	大阪熊取は「本綿・亜麻・大麻の工場」。宮城は2段階質問形式の下位選択肢(物じん一繊維)。
v2904	数値	職業 出版・印刷工場	一部相違	1: Yes (欠損=)	宮城は2段階質問形式の下位選択肢(現場一印刷工場)。
v2905	数値	職業 メッキ作業所	一部相違	1: Yes (欠損=)	大阪熊取は欠損。宮城は2段階質問形式の下位選択肢(現場一メッキ作業所)。
v2906	数値	職業 運輸業務	一部相違	1: Yes (欠損=)	大阪熊取は欠損。宮城は2段階質問形式の第1問。
v2907	数値	職業 鉄鋼工場	一部相違	1: Yes (欠損=)	宮城は2段階質問形式の下位選択肢(現場一鉄鋼工場)。
v2908	数値	職業 金属製品工場	一部相違	1: Yes (欠損=)	大阪熊取は欠損。 宮城は2段階質問形式の下位選択肢(物じん一金属研削)。
v2909	数値	職業 鋳物工場	一部相違	1: Yes (欠損=)	宮城は2段階質問形式の下位選択肢(物じん一鋳物(サンドブラストなど))。
v2910	数値	職業 陶磁器・ガラス工場	一部相違	1: Yes (欠損=)	宮城は2段階質問形式の下位選択肢(物じん一陶磁器・製造一ガラス製品)。
v2911	数値	職業 プラスチック加工	一部相違	1: Yes (欠損=)	宮城は2段階質問形式の下位選択肢(製造一プラスチック製品)。
v2912	数値	職業 ゴム製品製造	一部相違	1: Yes (欠損=)	宮城は2段階質問形式の下位選択肢(製造一ゴム製品)。
v2913	数値	職業 耐熱・保温作業	一部相違	1: Yes (欠損=)	宮城は2段階質問形式の下位選択肢(現場一耐熱保温作業現場)。
v2914	数値	職業 造船所	一部相違	1: Yes (欠損=)	宮城は2段階質問形式の下位選択肢(現場一造船所)。

変数名	属性	内容	定義の相違	凡例など	備考
v2915	数値	職業 部品解体・再生工場	一部相違	1: Yes (欠損=)	宮城は2段階質問形式の下位選択肢(現場→部品解体・再生現場)。
v2916	数値	職業 建設・解体現場	一部相違	1: Yes (欠損=)	宮城は2段階質問形式の下位選択肢(現場→建設・解体現場)。
v2917	数値	職業 石油精製・石油製品の製造	一部相違	1: Yes (欠損=)	宮城は2段階質問形式の下位選択肢(現場→石油精製所、製造→石油製品)。
v2918	数値	職業 ビッチ・タール・アスファルト製造	一部相違	1: Yes (欠損=)	宮城は2段階質問形式の下位選択肢(製造→ビッチ・タール・アスファルト)。
v2919	数値	職業 バッテリー・電池製造	一部相違	1: Yes (欠損=)	宮城は2段階質問形式の下位選択肢(製造→バッテリー・電池)。
v2920	数値	職業 セメント・生コン製品製造	一部相違	1: Yes (欠損=)	宮城は2段階質問形式の下位選択肢(扱う仕事→セメント・生コン)。
v2921	数値	職業 石綿工場	一部相違	1: Yes (欠損=)	宮城は2段階質問形式の下位選択肢(扱う仕事→石綿または石綿含有物質)。
v2922	数値	職業 ランドリー・洗濯工場	一部相違	1: Yes (欠損=)	宮城は2段階質問形式の下位選択肢(場所→ランドリー・クリーニング)。
v2923	数値	職業 土石採取・石材加工	一部相違	1: Yes (欠損=)	宮城は2段階質問形式の下位選択肢(粉じん→石材、採土・採石)。
v2924	数値	職業 自動車修理・整備	一部相違	1: Yes (欠損=)	大阪能勢は欠損。宮城は2段階質問形式の下位選択肢(現場→自動車修理作業所)。
v2925	数値	職業 写真現像所	一部相違	1: Yes (欠損=)	大阪能勢は欠損。宮城は2段階質問形式の下位選択肢(場所→写真現像所)。
v2926	数値	職業 炭坑・その他の鉱山、鉱石の精錬	一部相違	1: Yes (欠損=)	宮城は2段階質問形式の第1問(鉱業)。
v2927	数値	職業 サービス業(公務員も含む)	共通	1: Yes (欠損=)	宮城、大阪能勢は欠損。
v2928	数値	職業 林業	一部相違	1: Yes (欠損=)	大阪東成、大阪能勢は欠損。宮城は2段階質問形式の第1問。
v2929	数値	職業 農業(畜産・果樹園芸も含む)	一部相違	1: Yes (欠損=)	大阪能勢は欠損。宮城は2段階質問形式の第1問。
v2930	数値	職業 皮革製品製造	一部相違	1: Yes (欠損=)	宮城、大阪能勢のみ。宮城は2段階質問形式の下位選択肢(製造→皮革製品)。
v2931	数値	職業 殺虫剤・除草剤製造	一部相違	1: Yes (欠損=)	宮城、大阪能勢のみ。宮城は2段階質問形式の下位選択肢(製造→殺虫剤・除草剤・殺虫剤)。
v2932	数値	職業 家具製造	一部相違	1: Yes (欠損=)	宮城、大阪能勢のみ。宮城は2段階質問形式の下位選択肢(製造→家具)。
v2933	数値	職業 卸売業・小売業	共通	1: Yes (欠損=)	大阪東成のみ。
v2934	数値	職業 有機溶剤	共通	1: Yes (欠損=)	宮城のみ。
v2935	数値	職業 放射線・放射性物質	共通	1: Yes (欠損=)	宮城のみ。
v2936	数値	職業 ベリリウム	共通	1: Yes (欠損=)	宮城のみ。
v2937	数値	職業 タルク(滑石)	共通	1: Yes (欠損=)	宮城のみ。
v2938	数値	職業 染料	共通	1: Yes (欠損=)	宮城のみ。
v2939	数値	職業 なし	共通	1: Yes (欠損=)	宮城欠損。v2901→v2932と矛盾ありママ(大阪能勢、愛知)。
v2940	数値	最も長く就いた仕事	一部相違	(欠損=)	3桁(愛知のみ2桁)コードそのまま読み(0、99、999もママ)。コードの出典は不明。
v2941	数値	最も長く就いた仕事	一部相違	(欠損=)	大阪能勢のみ。4桁コードそのまま読み。 大阪東成はすべて0のため欠損に。コードの出典は不明。
v2950	数値	従事年数	共通	(欠損=)	0以上はそのまま読み(0、99もママ)。



表2 各変数の分類数とユニークセル数、カテゴリとグループについて

グループ	カテゴリ名	変数名	内容	分類数	ユニークセル数	
個人特性	性別	v0200	性別	2		
	年齢	v0300	追跡開始年齢(40-99)	60		
	居住地	v0402	地区通し番号	6		
追跡	日付	v0500	追跡開始日	7		
		v0501	10年コホート追跡終了日(地域ごとの定義)	7		
		v0502	10年観察終了日(個人ごとの定義)	12651	304	
	転帰	v0503	10年転帰	3		
	死因	v0600	死因ICD-9コード4桁	836	311	
自覚症状	自覚症状	v1000	自覚症状あるか 咳	2		
		v1001	自覚症状あるか 痰	2		
		v1002	自覚症状あるか 血痰	2		
		v1003	自覚症状あるか 動悸	2		
		v1004	自覚症状あるか 息切れ	2		
		v1005	自覚症状あるか 食欲不振	2		
		v1006	自覚症状あるか 便秘	2		
		v1007	自覚症状あるか 不眠	2		
		v1008	自覚症状あるか やせてきた	2		
		v1009	自覚症状あるか 疲れやすい	2		
		v1010	自覚症状あるか 喘息の発作	2		
	v1011	自覚症状あるか なし	4			
	既往	既往	v1100	既往 高血圧	4	
			v1101	既往 心臓病	4	
			v1102	既往 脳卒中	4	
			v1103	既往 糖尿病	4	
			v1104	既往 肝臓病	4	
			v1105	既往 腎臓病	4	
			v1106	既往 胃十二指腸潰瘍	4	
			v1107	既往 胆石	4	
			v1108	既往 虫垂炎	2	
			v1109	既往 結核ろく膜炎	4	
			v1110	既往 肺炎	4	
			v1111	既往 喘息	4	
			v1112	既往 慢性気管支炎	4	
			v1113	既往 肺気腫	4	
			v1114	既往 じん肺	4	
			v1115	既往 貧血	2	
			v1116	既往 アレルギー性鼻炎	4	
			v1117	既往 蕁麻疹	4	
			v1118	既往 湿疹・皮膚炎	4	
			v1119	既往 痛風	2	
	v1120	既往 なし	2			
	身長・体重	身長・体重	v1200	身長(cm)	489	97
			v1201	体重(kg)	433	126
	まるめの身長・体重	まるめの身長・体重	v1200	身長(cm)※小数点切り捨て	89	6
			v1201	体重(kg)※小数点以下切り捨て	100	19
	胸部レントゲン	胸部レントゲン	v1300	胸部レントゲン検査受診有無	3	
			v1301	胸部レントゲン検査受診有無(1年以内)	2	
			v1302	胸部レントゲン検査受診有無(1年より前~2年以内)	2	
			v1303	胸部レントゲン検査受診有無(2年より前)	2	
			v1304	胸部レントゲン検査受診場所(住民健診)	2	
			v1305	胸部レントゲン検査受診場所(保健所・保健センター)	2	
			v1306	胸部レントゲン検査受診場所(職場)	2	
			v1307	胸部レントゲン検査受診場所(病院など)	2	
			v1308	胸部レントゲン検査受診場所(農協)	2	
			v1309	胸部レントゲン検査受診場所(その他)	2	
	胃レントゲン	胃レントゲン	v1400	胃レントゲン検査受診有無	3	
			v1401	胃レントゲン検査受診有無(1年以内)	2	
			v1402	胃レントゲン検査受診有無(1年より前~2年以内)	2	
			v1403	胃レントゲン検査受診有無(2年より前)	2	

グループ	カテゴリ名	変数名	内容	分類数	ユニークセル数
	胃レントゲン	v1404	胃レントゲン検査受診場所(住民健診)	2	
		v1405	胃レントゲン検査受診場所(保健所・保健センター)	2	
		v1406	胃レントゲン検査受診場所(職場)	2	
		v1407	胃レントゲン検査受診場所(病院など)	2	
		v1408	胃レントゲン検査受診場所(農協)	2	
		v1409	胃レントゲン検査受診場所(その他)	2	
		v1410	胃レントゲン検査受診回数(5年間)	16	
	健康診査	v1500	健康診査または血圧検査受診有無	3	
		v1501	健康診査または血圧検査受診有無(1年以内)	2	
		v1502	健康診査または血圧検査受診有無(1年より前～2年以内)	2	
		v1503	健康診査または血圧検査受診有無(2年より前)	2	
		v1504	健康診査または血圧検査受診場所(住民健診)	2	
		v1505	健康診査または血圧検査受診場所(保健所・保健センター)	2	
		v1506	健康診査または血圧検査受診場所(職場)	2	
		v1507	健康診査または血圧検査受診場所(病院など)	2	
		v1508	健康診査または血圧検査受診場所(農協)	2	
		v1509	健康診査または血圧検査受診場所(その他)	2	
	子宮がん	v1600	子宮がん検診受診有無	3	
		v1601	子宮がん検診受診有無(昨年)	2	
		v1602	子宮がん検診受診有無(一昨年)	2	
		v1603	子宮がん検診受診有無(一昨年より前)	2	
		v1604	子宮がん検診受診場所(検診車・住民検診)	2	
		v1605	子宮がん検診受診場所(保健所・保健センター)	2	
		v1606	子宮がん検診受診場所(職場)	2	
		v1607	子宮がん検診受診場所(病院など)	2	
		v1608	子宮がん検診受診場所(東成母子会自己採取)	2	
		v1609	子宮がん検診受診場所(その他)	2	
	出産・月経	v1610	初経年齢	17	2
		v1611	現在月経有り	3	
		v1612	自然閉経年齢	51	4
		v1613	手術閉経年齢	44	6
		v1614	出産経験あるか	3	
		v1615	出産人数	19	3
		v1616	初産年齢	41	4
	健康保険	v1700	健康保険カテゴリ	8	
		v1701	健康保険本人・家族別	3	
	食習慣	v1800	ご飯1日何杯	16	
		v1801	パン摂取頻度カテゴリ	6	
		v1802	肉摂取頻度カテゴリ	6	
v1803		魚類摂取頻度カテゴリ	6		
v1804		卵摂取頻度カテゴリ	6		
v1805		牛乳摂取頻度カテゴリ	6		
v1806		緑黄色野菜摂取頻度カテゴリ	6		
v1807		その他の野菜摂取頻度カテゴリ	6		
v1808		果物摂取頻度カテゴリ	6		
v1809		みそ汁摂取頻度カテゴリ	6		
v1810		漬物摂取頻度カテゴリ	6		
v1811	インスタント食品摂取頻度カテゴリ	6			
お茶・コーヒー	v1900	緑茶摂取頻度カテゴリ	6		
	v1901	紅茶摂取頻度カテゴリ	6		
	v1902	コーヒー(非インスタント)摂取頻度カテゴリ	6		
	v1903	インスタントコーヒー摂取頻度カテゴリ	6		
	v1904	コーヒー(豆またはインスタント)摂取頻度カテゴリ	6		
飲酒	v2000	飲酒習慣カテゴリ	6		
喫煙	v2100	喫煙習慣カテゴリ	5		
	v2101	喫煙開始年齢	83	11	
	v2102	喫煙本数/日	54	5	
	v2103	禁煙年齢	76	3	
	v2104	吸い込むか	3		
	v2200	父親の既往 脳卒中	2		

グループ	カテゴリ名	変数名	内容	分類数	ユニークセル数	
父親の既往	v2201	父親の既往	心臓病	2		
	v2202	父親の既往	がん	2		
	v2203	父親の既往	上記3つなし	2		
	v2204	父親の既往	知らない	2		
	v2205	父親の既往	がん 食道	2		
	v2206	父親の既往	がん 胃	2		
	v2207	父親の既往	がん 大腸	2		
	v2208	父親の既往	がん 肝臓	2		
	v2209	父親の既往	がん 肺	2		
	v2210	父親の既往	がん 膀胱	2		
	v2211	父親の既往	がん 白血病	2		
	v2212	父親の既往	がん その他	2		
	母親の既往	v2300	母親の既往	脳卒中	2	
		v2301	母親の既往	心臓病	2	
		v2302	母親の既往	がん	2	
v2303		母親の既往	上記3つなし	2		
v2304		母親の既往	知らない	2		
v2305		母親の既往	がん 食道	2		
v2306		母親の既往	がん 胃	2		
v2307		母親の既往	がん 大腸	2		
v2308		母親の既往	がん 肝臓	2		
v2309		母親の既往	がん 肺	2		
v2310		母親の既往	がん 膀胱	2		
v2311		母親の既往	がん 白血病	2		
v2312		母親の既往	がん 子宮	2		
v2313		母親の既往	がん 乳房	2		
v2314	母親の既往	がん その他	2			
周囲の喫煙	v2400	小学生の頃の父親喫煙	有無	5		
	v2401	小学生の頃の母親喫煙	有無	5		
	v2500	現在同居家族喫煙	あり	3		
	v2501	現在夫喫煙	あり	2		
	v2502	現在妻喫煙	あり	2		
	v2503	現在父喫煙	あり	2		
	v2504	現在母喫煙	あり	2		
	v2505	現在子喫煙	あり	2		
v2506	現在その他喫煙	あり	2			
暖房	v2600	暖房	なし	2		
	v2601	暖房	電気(複数回答)	2		
	v2602	暖房	エアコン(複数回答)	2		
	v2603	暖房	クリーンヒーター(複数回答)	2		
	v2604	暖房	煙突付きストーブ(複数回答)	2		
	v2605	暖房	煙突なしストーブ(複数回答)	2		
	v2606	暖房	炭(複数回答)	2		
住居	v2700	住居がバスの通る道路に面しているか	3			
転居	v2800	生まれてからいまのところに住んでいるか	3			
	v2801	転入何年前か	101	5		
職業	v2901	職業	病院・診療所	2		
	v2902	職業	美容院・理髪店	2		
	v2903	職業	紡績・綿布・タオル工場	2		
	v2904	職業	出版・印刷工場	2		
	v2905	職業	メッキ作業所	2		
	v2906	職業	運輸業務	2		
	v2907	職業	鉄鋼工場	2		
	v2908	職業	金属製品工場	2		
	v2909	職業	鋳物工場	2		
	v2910	職業	陶磁器・ガラス工場	2		
	v2911	職業	プラスチック加工	2		
	v2912	職業	ゴム製品製造	2		
	v2913	職業	耐熱・保温作業	2		
	v2914	職業	造船所	2		
	v2915	職業	部品解体・再生工場	2		

グループ	カテゴリ名	変数名	内容	分類数	ユニークセル数
	職業	v2916	職業 建設・解体現場	2	
		v2917	職業 石油精製・石油製品の製造	2	
		v2918	職業 ピッチ・タール・アスファルト製造	2	
		v2919	職業 バッテリー・電池製造	2	
		v2920	職業 セメント・生コン製品製造	2	
		v2921	職業 石綿工場	2	
		v2922	職業 ランドリー・洗濯工場	2	
		v2923	職業 土石採取・石材加工	2	
		v2924	職業 自動車修理・整備	2	
		v2925	職業 写真現像所	2	
		v2926	職業 炭坑・その他の鉱山、鉱石の精錬	2	
		v2927	職業 サービス業(公務員も含む)	2	
		v2928	職業 林業	2	
		v2929	職業 農業(畜産・果樹園芸も含む)	2	
		v2930	職業 皮革製品製造	2	
		v2931	職業 殺虫剤・枯草剤製造	2	
		v2932	職業 家具製造	2	
		v2933	職業 卸売業・小売業	2	
		v2934	職業 有機溶剤	2	
		v2935	職業 放射線・放射性物質	2	
		v2936	職業 ベリリウム	2	
		v2937	職業 タルク(滑石)	2	
		v2938	職業 染料	2	
		v2939	職業 なし	2	
		v2940	最も長く就いた仕事	583	62
	v2950	従事年数	81	7	

表3 ベースライングループ(【個人特性】【追跡】)に対するまるめの処理について

個人特性	処理内容	分類数	ユニークセル数
1	処理なし(性別×年齢×居住地)	673	19
2	年齢5歳階級かつ85歳以上まるめ	120	0

追跡	処理内容	分類数	ユニークセル数
1	処理なし(日付:日まで×転帰×死因ICD-9コード4桁)	20176	16631
2	日付:月まで×転帰×死因ICD-9コード3桁	11398	7936
3	追跡期間(月)×転帰×死因ICD-9コード3桁	6225	3654
4	追跡期間(月)×転帰×(死因削除)	243	0

表4 アンケートカテゴリの組み合わせによって得られた各パターンにおける分類数とユニークセル数

アンケートカテゴリ	個人特性	追跡	分類数 $K$	ユニークセル数 $S_2$	$S_2 / K$
なし(ベースラインのみ)	0	1	20176	16631	0.82
なし(ベースラインのみ)	0	2	11398	7936	0.70
なし(ベースラインのみ)	0	3	6225	3654	0.59
なし(ベースラインのみ)	0	4	243	0	0.00
なし(ベースラインのみ)	1	0	673	19	0.03
なし(ベースラインのみ)	1	1	31208	29062	0.93
なし(ベースラインのみ)	1	2	28383	24606	0.87
なし(ベースラインのみ)	1	3	27949	24168	0.86
なし(ベースラインのみ)	1	4	22743	18272	0.80
なし(ベースラインのみ)	2	0	120	0	0.00
なし(ベースラインのみ)	2	1	27899	25715	0.92
なし(ベースラインのみ)	2	2	22025	17925	0.81
なし(ベースラインのみ)	2	3	21475	17273	0.80
なし(ベースラインのみ)	2	4	13749	7258	0.53
自覚症状	0	0	660	188	0.28
自覚症状	0	1	28210	24958	0.88
自覚症状	0	2	22813	18730	0.82
自覚症状	0	3	17440	13779	0.79
自覚症状	0	4	8846	5484	0.62
自覚症状	1	0	19239	11191	0.58
自覚症状	1	1	46922	40305	0.86
自覚症状	1	2	46049	38839	0.84
自覚症状	1	3	44772	37992	0.85
自覚症状	1	4	41003	33822	0.82
自覚症状	2	0	9253	4740	0.51
自覚症状	2	1	38234	33756	0.88
自覚症状	2	2	35941	30299	0.84
自覚症状	2	3	35103	29635	0.84
自覚症状	2	4	30759	24836	0.81
既往	0	0	3379	2106	0.62
既往	0	1	30350	26980	0.89
既往	0	2	25222	20994	0.83
既往	0	3	21390	17371	0.81
既往	0	4	12964	8989	0.69
既往	1	0	23102	15583	0.67
既往	1	1	48119	42161	0.88
既往	1	2	47305	40782	0.86
既往	1	3	46844	40499	0.86
既往	1	4	43492	36711	0.84
既往	2	0	13350	8639	0.65
既往	2	1	40149	35853	0.89
既往	2	2	38057	32690	0.86
既往	2	3	37753	32397	0.86
既往	2	4	33648	27715	0.82
身長・体重	0	0	7635	4020	0.53
身長・体重	0	1	46327	38730	0.84
身長・体重	0	2	45863	37892	0.83
身長・体重	0	3	37393	32621	0.87
身長・体重	0	4	31829	26180	0.82
身長・体重	1	0	77778	62845	0.81
身長・体重	1	1	89452	81268	0.91
身長・体重	1	2	89429	81222	0.91
身長・体重	1	3	87748	78678	0.90
身長・体重	1	4	86625	76847	0.89
身長・体重	2	0	49270	30959	0.63
身長・体重	2	1	72093	59609	0.83
身長・体重	2	2	71999	59427	0.83

アンケートカテゴリ	個人特性	追跡	分類数 $K$	ユニークセル数 $S_1$	$S_1 / K$
身長・体重	2	3	69315	56671	0.82
身長・体重	2	4	66804	53244	0.80
まるめ身長・体重	0	0	5141	595	0.12
まるめ身長・体重	0	1	41602	34719	0.83
まるめ身長・体重	0	2	41023	33671	0.82
まるめ身長・体重	0	3	32987	28977	0.88
まるめ身長・体重	0	4	27059	22286	0.82
まるめ身長・体重	1	0	74249	57346	0.77
まるめ身長・体重	1	1	87561	78153	0.89
まるめ身長・体重	1	2	87528	78087	0.89
まるめ身長・体重	1	3	85794	75499	0.88
まるめ身長・体重	1	4	84559	73490	0.87
まるめ身長・体重	2	0	42689	23044	0.54
まるめ身長・体重	2	1	67968	54296	0.80
まるめ身長・体重	2	2	67850	54068	0.80
まるめ身長・体重	2	3	65128	51273	0.79
まるめ身長・体重	2	4	62388	47570	0.76
胸部レントゲン	0	0	205	30	0.15
胸部レントゲン	0	1	27170	24524	0.90
胸部レントゲン	0	2	20850	16390	0.79
胸部レントゲン	0	3	15898	11684	0.73
胸部レントゲン	0	4	6042	2754	0.46
胸部レントゲン	1	0	11866	4686	0.39
胸部レントゲン	1	1	42604	35904	0.84
胸部レントゲン	1	2	41898	34698	0.83
胸部レントゲン	1	3	40854	34007	0.83
胸部レントゲン	1	4	36615	29235	0.80
胸部レントゲン	2	0	4197	1324	0.32
胸部レントゲン	2	1	34563	30806	0.89
胸部レントゲン	2	2	32592	27701	0.85
胸部レントゲン	2	3	32008	27274	0.85
胸部レントゲン	2	4	26646	20947	0.79
胃レントゲン	0	0	569	152	0.27
胃レントゲン	0	1	26930	24062	0.89
胃レントゲン	0	2	20869	16530	0.79
胃レントゲン	0	3	16876	12716	0.75
胃レントゲン	0	4	8021	4532	0.57
胃レントゲン	1	0	15394	6708	0.44
胃レントゲン	1	1	45546	37820	0.83
胃レントゲン	1	2	44785	36449	0.81
胃レントゲン	1	3	44229	36155	0.82
胃レントゲン	1	4	40035	31532	0.79
胃レントゲン	2	0	5854	2198	0.38
胃レントゲン	2	1	36032	31512	0.87
胃レントゲン	2	2	33830	28004	0.83
胃レントゲン	2	3	33513	27733	0.83
胃レントゲン	2	4	28322	21655	0.76
健康診査	0	0	187	47	0.25
健康診査	0	1	22934	19579	0.85
健康診査	0	2	16044	12350	0.77
健康診査	0	3	12250	8708	0.71
健康診査	0	4	3334	1530	0.46
健康診査	1	0	6142	2157	0.35
健康診査	1	1	37800	32898	0.87
健康診査	1	2	35450	29225	0.82
健康診査	1	3	34319	28474	0.83
健康診査	1	4	29709	23436	0.79
健康診査	2	0	2120	718	0.34

アンケートカテゴリ	個人特性	追跡	分類数 $K$	ユニークセル数 $S_{\pm}$	$S_{\pm} / K$
健康診査	2	1	31509	28241	0.90
健康診査	2	2	26809	22125	0.83
健康診査	2	3	26189	21693	0.83
健康診査	2	4	20129	14188	0.70
子宮がん検診	0	0	75	21	0.28
子宮がん検診	0	1	24116	21309	0.88
子宮がん検診	0	2	15129	11206	0.74
子宮がん検診	0	3	9278	6007	0.65
子宮がん検診	0	4	1565	634	0.41
子宮がん検診	1	0	2559	631	0.25
子宮がん検診	1	1	33659	30388	0.90
子宮がん検診	1	2	31304	26620	0.85
子宮がん検診	1	3	30644	26126	0.85
子宮がん検診	1	4	25523	20354	0.80
子宮がん検診	2	0	701	130	0.19
子宮がん検診	2	1	29240	26747	0.91
子宮がん検診	2	2	24153	19832	0.82
子宮がん検診	2	3	23584	19249	0.82
子宮がん検診	2	4	16205	9765	0.60
出産・月経	0	0	13244	8858	0.67
出産・月経	0	1	40254	34928	0.87
出産・月経	0	2	34942	28745	0.82
出産・月経	0	3	29657	24001	0.81
出産・月経	0	4	23610	19102	0.81
出産・月経	1	0	34095	27894	0.82
出産・月経	1	1	57528	52122	0.91
出産・月経	1	2	55953	49674	0.89
出産・月経	1	3	54970	48858	0.89
出産・月経	1	4	51230	44430	0.87
出産・月経	2	0	25093	19856	0.79
出産・月経	2	1	49370	44319	0.90
出産・月経	2	2	46195	40124	0.87
出産・月経	2	3	45490	39569	0.87
出産・月経	2	4	39817	32323	0.81
健康保険	0	0	21	0	0.00
健康保険	0	1	26358	23781	0.90
健康保険	0	2	19154	14960	0.78
健康保険	0	3	13624	9783	0.72
健康保険	0	4	3738	823	0.22
健康保険	1	0	6968	1747	0.25
健康保険	1	1	38654	33266	0.86
健康保険	1	2	37554	31429	0.84
健康保険	1	3	36619	30929	0.84
健康保険	1	4	32121	26087	0.81
健康保険	2	0	1809	211	0.12
健康保険	2	1	32318	29366	0.91
健康保険	2	2	29553	25415	0.86
健康保険	2	3	28978	24666	0.85
健康保険	2	4	23581	18313	0.78
食習慣	0	0	86272	80565	0.93
食習慣	0	1	93495	90776	0.97
食習慣	0	2	93409	90630	0.97
食習慣	0	3	91710	88060	0.96
食習慣	0	4	90834	86815	0.96
食習慣	1	0	97840	97023	0.99
食習慣	1	1	98868	98344	0.99
食習慣	1	2	98865	98338	0.99
食習慣	1	3	98847	98311	0.99



アンケートカテゴリ	個人特性	追跡	分類数 $K$	ユニークセル数 $S_{\pm}$	$S_{\pm} / K$
食習慣	1	4	98771	98212	0.99
食習慣	2	0	96197	94918	0.99
食習慣	2	1	97838	97009	0.99
食習慣	2	2	97830	96994	0.99
食習慣	2	3	97787	96935	0.99
食習慣	2	4	97645	96739	0.99
お茶・コーヒー	0	0	1091	166	0.15
お茶・コーヒー	0	1	33906	30188	0.89
お茶・コーヒー	0	2	31810	26766	0.84
お茶・コーヒー	0	3	26948	22307	0.83
お茶・コーヒー	0	4	18879	13105	0.69
お茶・コーヒー	1	0	44453	25681	0.58
お茶・コーヒー	1	1	68962	56277	0.82
お茶・コーヒー	1	2	68873	56108	0.81
お茶・コーヒー	1	3	66621	53463	0.80
お茶・コーヒー	1	4	64047	49928	0.78
お茶・コーヒー	2	0	19782	8390	0.42
お茶・コーヒー	2	1	50541	40776	0.81
お茶・コーヒー	2	2	50171	40109	0.80
お茶・コーヒー	2	3	48043	38528	0.80
お茶・コーヒー	2	4	44114	33893	0.77
飲酒	0	0	6	0	0.00
飲酒	0	1	25032	22273	0.89
飲酒	0	2	17242	13277	0.77
飲酒	0	3	11223	7788	0.69
飲酒	0	4	1449	16	0.01
飲酒	1	0	3568	260	0.07
飲酒	1	1	35788	31544	0.88
飲酒	1	2	34463	29328	0.85
飲酒	1	3	33655	28974	0.86
飲酒	1	4	28907	23916	0.83
飲酒	2	0	720	2	0.00
飲酒	2	1	30930	28571	0.92
飲酒	2	2	27606	23629	0.86
飲酒	2	3	27148	23220	0.86
飲酒	2	4	21531	16536	0.77
喫煙	0	0	8565	5387	0.63
喫煙	0	1	38375	33784	0.88
喫煙	0	2	34045	28874	0.85
喫煙	0	3	26623	22025	0.83
喫煙	0	4	19744	15687	0.79
喫煙	1	0	34647	28606	0.83
喫煙	1	1	55675	51132	0.92
喫煙	1	2	54780	49665	0.91
喫煙	1	3	53865	48821	0.91
喫煙	1	4	50956	45474	0.89
喫煙	2	0	25435	19857	0.78
喫煙	2	1	48620	44151	0.91
喫煙	2	2	46431	41019	0.88
喫煙	2	3	45597	40191	0.88
喫煙	2	4	41651	35328	0.85
父親の既往	0	0	194	62	0.32
父親の既往	0	1	26090	23110	0.89
父親の既往	0	2	19421	15246	0.79
父親の既往	0	3	13363	9592	0.72
父親の既往	0	4	3683	1526	0.41
父親の既往	1	0	9161	3338	0.36
父親の既往	1	1	40929	34889	0.85

アンケートカテゴリ	個人特性	追跡	分類数 $K$	ユニークセル数 $S_{\pm}$	$S_{\pm} / K$
父親の既往	1	2	39970	33248	0.83
父親の既往	1	3	38910	32727	0.84
父親の既往	1	4	34396	27956	0.81
父親の既往	2	0	3138	1046	0.33
父親の既往	2	1	33859	30533	0.90
父親の既往	2	2	31277	26617	0.85
父親の既往	2	3	30693	26124	0.85
父親の既往	2	4	25584	20329	0.79
母親の既往	0	0	254	108	0.43
母親の既往	0	1	25860	22806	0.88
母親の既往	0	2	19176	15090	0.79
母親の既往	0	3	13162	9433	0.72
母親の既往	0	4	3644	1547	0.42
母親の既往	1	0	9311	3708	0.40
母親の既往	1	1	40892	35164	0.86
母親の既往	1	2	39775	33333	0.84
母親の既往	1	3	38742	32761	0.85
母親の既往	1	4	34301	28071	0.82
母親の既往	2	0	3354	1213	0.36
母親の既往	2	1	33869	30478	0.90
母親の既往	2	2	31135	26465	0.85
母親の既往	2	3	30542	25990	0.85
母親の既往	2	4	25412	20105	0.79
周囲の喫煙	0	0	674	186	0.28
周囲の喫煙	0	1	29631	26293	0.89
周囲の喫煙	0	2	24687	20394	0.83
周囲の喫煙	0	3	18932	14986	0.79
周囲の喫煙	0	4	9726	5446	0.56
周囲の喫煙	1	0	22188	11660	0.53
周囲の喫煙	1	1	50873	42545	0.84
周囲の喫煙	1	2	49980	41021	0.82
周囲の喫煙	1	3	48343	39757	0.82
周囲の喫煙	1	4	44604	35468	0.80
周囲の喫煙	2	0	9367	3920	0.42
周囲の喫煙	2	1	39840	34393	0.86
周囲の喫煙	2	2	37705	31246	0.83
周囲の喫煙	2	3	36672	30511	0.83
周囲の喫煙	2	4	32183	25553	0.79
暖房	0	0	85	8	0.09
暖房	0	1	25412	21678	0.85
暖房	0	2	19794	15669	0.79
暖房	0	3	14414	10742	0.75
暖房	0	4	4511	1502	0.33
暖房	1	0	11429	3948	0.35
暖房	1	1	42768	35719	0.84
暖房	1	2	41922	34249	0.82
暖房	1	3	40568	33398	0.82
暖房	1	4	36310	28806	0.79
暖房	2	0	3657	953	0.26
暖房	2	1	34553	30791	0.89
暖房	2	2	32204	27108	0.84
暖房	2	3	31478	26609	0.85
暖房	2	4	26480	20934	0.79
住居	0	0	3	0	0.00
住居	0	1	21647	18061	0.83
住居	0	2	13924	10347	0.74
住居	0	3	8484	5400	0.64
住居	0	4	702	39	0.06

アンケートカテゴリ	個人特性	追跡	分類数 $K$	ユニークセル数 $S_{\pm}$	$S_{\pm} / K$
住居	1	0	1820	144	0.08
住居	1	1	33221	30306	0.91
住居	1	2	31053	26827	0.86
住居	1	3	30486	26449	0.87
住居	1	4	25656	21295	0.83
住居	2	0	358	3	0.01
住居	2	1	29341	27117	0.92
住居	2	2	24577	20563	0.84
住居	2	3	24074	20011	0.83
住居	2	4	17735	11937	0.67
転居	0	0	265	28	0.11
転居	0	1	26289	22697	0.86
転居	0	2	21455	16833	0.78
転居	0	3	18103	13535	0.75
転居	0	4	9982	5818	0.58
転居	1	0	16154	6664	0.41
転居	1	1	45545	37744	0.83
転居	1	2	44982	36711	0.82
転居	1	3	44409	36331	0.82
転居	1	4	40159	31538	0.79
転居	2	0	5916	1613	0.27
転居	2	1	36150	31406	0.87
転居	2	2	34406	28553	0.83
転居	2	3	33882	27991	0.83
転居	2	4	28432	21553	0.76
職業	0	0	24650	17061	0.69
職業	0	1	52980	46356	0.87
職業	0	2	50850	43361	0.85
職業	0	3	47419	40044	0.84
職業	0	4	41266	32987	0.80
職業	1	0	61604	53275	0.86
職業	1	1	76300	70725	0.93
職業	1	2	76093	70360	0.92
職業	1	3	75830	69995	0.92
職業	1	4	73924	67682	0.92
職業	2	0	47907	38873	0.81
職業	2	1	67087	60757	0.91
職業	2	2	66495	59767	0.90
職業	2	3	66050	59204	0.90
職業	2	4	63275	55725	0.88

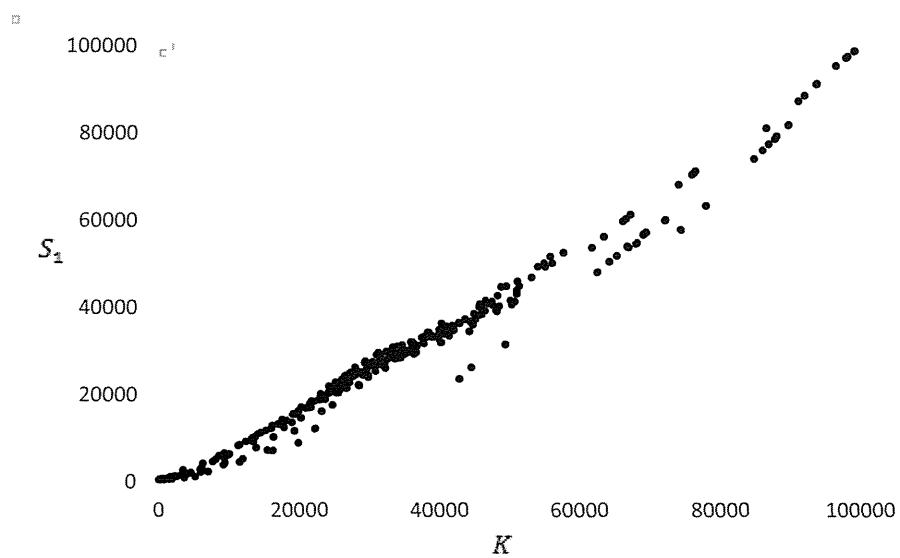


図1 分類数( $K$ )とユニークセル数( $S_1$ )の関係

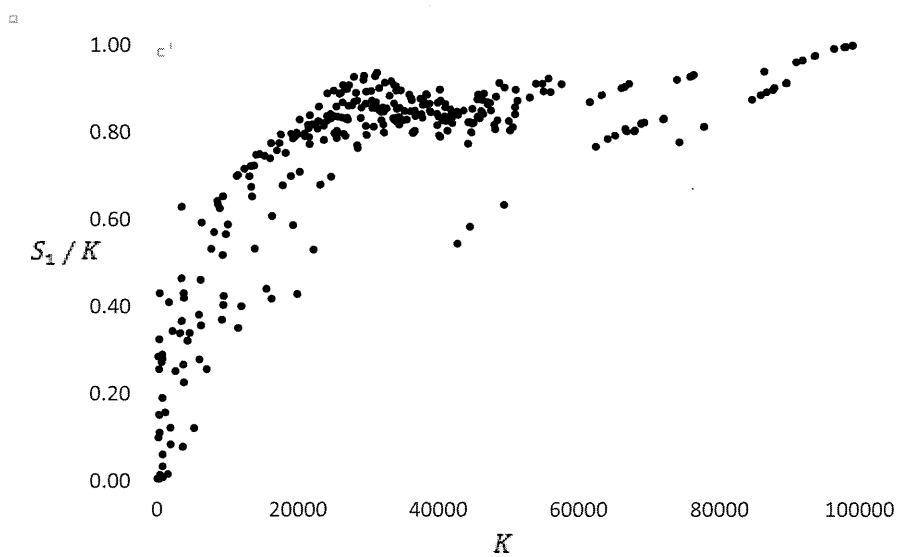


図2 分類数( $K$ )と分類数に占めるユニークセル数の割合( $S_1/K$ )の関係