

現在提供している
匿名データ (A)

より匿名性の高い
匿名データ (B)

現在提供してい「匿名データ (A)」
を基に、そこにさらに、

- ・リサンプリング
- ・リコードイング
- ・トップ・ボトムコーディング
- ・出現頻度が低いレコードの削除

などの手法を適用して、より匿名性の高いデータ (B)を作成

(※まずは搅乱手法を用いない簡単な作成方法について検討していく)

(※各種学会で中間結果の報告を予定)

- 「R2国勢調査」の使用を検討
 - ・項目数がそれほど多くない
 - ・カテゴリー変数が多い
- 教育利用の方法についても、
合わせて検討していく

■ 令和2年（2020年）国勢調査の匿名データの項目名一覧

※秘匿性の高い（情報量の少ない）匿名データの教育利用の方法についても合わせて検討

項目番号	変数名	項目名	項目番号	変数名	項目名
1	V0001	政府統計コード	16	V0016	世帯に関する事項_3世代世帯か否か
2	V0002	管理コード	17	V0017	個人に関する事項_世帯主との続き柄
3	V0003	調査年月（西暦）	18	V0018	個人に関する事項_男女の別
4	V0004	都道府県	19	V0019	個人に関する事項_年齢
5	V0005	市区町村	20	V0020	個人に関する事項_配偶者の有無
6	V0006	レコード一連番号	21	V0021	個人に関する事項_国籍
7	V0007	世帯一連番号	22	V0022	個人に関する事項_居住期間
8	V0008	世帯員番号	23	V0023	個人に関する事項_5年前の常住地
9	V0009	世帯に関する事項_世帯の種類	24	V0024	個人に関する事項_教育状態
10	V0010	世帯に関する事項_住宅の建て方	25	V0025	個人に関する事項_労働力状態
11	V0011	世帯に関する事項_建物全体の階数	26	V0026	個人に関する事項_利用交通手段
12	V0012	世帯に関する事項_世帯が住んでいる階	27	V0027	個人に関する事項_従業地・通学地
13	V0013	世帯に関する事項_住居の種類・住宅の所有の関係	28	V0028	個人に関する事項_従業上の地位
14	V0014	世帯に関する事項_世帯人員	29	V0029	個人に関する事項_産業（大分類）
15	V0015	世帯に関する事項_世帯の家族類型	30	V0030	個人に関する事項_職業（大分類）

- 有用性に関する指標：
 - ・（元の匿名データと新たな匿名データとの）分布間の距離に関する各種指標
- 秘匿性に関する指標：
 - ・各変数の組合せに関する3～5重クロス集計
⇒度数が1となるセルが生ずるもの割合を指標とする

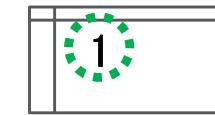
■ 秘匿性に関する指標のイメージ

より匿名性の高い
匿名データ (B)



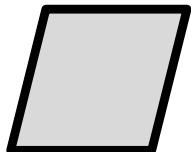
様々な項目の組合せで
クロス集計表を作成

それらのクロス集計表の中で
度数1のセルが含まれる表の
全体に占める割合
を秘匿性の指標とする



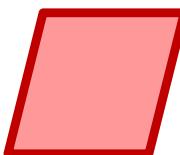
■ 匿名データ作成・検証のイメージ

現在提供している
匿名データ (A)

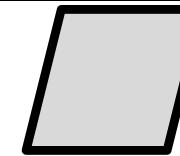


- リサンプリング
- リコーディング
- トップ・ボトムコーディング
- 出現頻度が低いレコードの削除

より匿名性の高い
匿名データ (B)

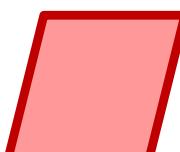


現在提供している
匿名データ (A)



秘匿性の指標①

結果の比較
(従来の匿名データよりも
秘匿性の指標の値が
○○だけ低くなっただけ)



秘匿性の指標②

より匿名性の高い
匿名データ (B)