

JSTAR データセット 個人特定化防止のための統計的処理

JSTAR Data Set は、研究目的のみに利用が認められるものであるが、研究目的の利用であっても、調査対象者の特定化を不可能なものにするために、データセットに対して以下に示す個人情報秘匿の為の処理を施す。秘匿処理は、大きく分類して次の 4 種類の情報に関して行われる。

1. 居住地、地域情報へのアクセス制限
2. 本人、家族属性情報に対するトップコード処理と制限
3. 所得、支出、資産等の経済変数に関するトップコード処理
4. 医療・介護給付情報(レセプトデータ)へのアクセス制限と集計処理

1. 居住地、地域情報へのアクセス制限について

本調査のデータセットには、調査対象者の居住地、地域情報が含まれるが、当然ながら本情報は、調査対象者個人の特定化につながる可能性が高いという意味において、最も重要な調査項目となる。したがって、本調査データセット利用審査基準に定められており、研究上の特別の理由が経済産業研究所から認められない限り、調査対象の居住地、地域情報は利用することができない。

2. 本人、家族情報に対するトップコーディング処理と制限について

本調査のデータセットには、調査対象者本人の属性だけではなく、その家族関係についての一定の情報も同時に含まれている。特に、調査対象者が特異な属性や家族構成を有する場合、これらの情報を相互参照することで個人の特定につながってしまう可能性がある。この可能性を排除するために、調査対象者本人およびその家族の属性に関する情報に関して制限とトップコード処理を施す。これらの処理が施された情報については、研究上の特別の理由が経済産業研究所で認められない限り、利用することはできない。具体的な処理内容については、次々項の表に示す。

3. 所得、支出、資産等の経済的情報に関するトップコード処理

本調査の対象者は 50 歳以上の高齢世帯であるが、一般に高齢世帯の多様性は若年世帯に比べて大きく、特に所得や資産といった経済的属性について、それは著しく現れる。この性質を利用すれば、特に著しく高い所得であったり、多大な資産を有する個人・世帯については、その他の情報と相互参照することで個人が特定される可能性が生じる。この可能性を排除するために、所得、支出、資産に関連する調査項目についてはトップコード処理を施すこととする。これらのトップコードされた項目については、研究上の特別の理由が

経済産業研究所から認められない限り、利用することはできない。具体的にトップコードされる調査項目については、次々項の表に示す。

4. 医療・介護給付情報(レセプトデータ)へのアクセス制限と集計処理

本調査においては、各調査地域の自治体の協力を得て、調査対象者の同意のもとに、国民健康保険および介護保険のレセプトデータを収集しており、これを突合することで調査結果の一部として含むものとなっている。

本来のレセプト情報には調査対象者の医療・介護給付情報に加え、その給付がなされた医療機関や介護保険事業所の特定につながる医療機関番号および介護保険事業所番号も含まれている。そこで、医療機関番号および介護保険事業所番号については、いかなる機密保護レベルにおける研究利用であった場合でも、番号そのものを研究者が使用することは認められないものとする。

ただし、機密保護レベル **Ultra High** もしくは機密保護レベル **Very High** の環境下においては、その申請研究内容において経済産業研究所が研究上の十分な理由があると認めた場合にのみ、医療機関番号および介護保険事業所番号を通し番号で置き換え、機関種、所在地、病床数等の、それぞれの研究に必要と認められる医療機関および介護保険事業所の属性情報を ID 項目化して加工した情報のみを利用可能とする。機密保護レベル **High** の環境下での本調査の利用を認められた研究者等には、機関番号および介護保険事業所番号を除外し通し番号も付帯せず、さらには支出・給付費情報等をサービスの種別（医療であれば入院・外来、介護であれば在宅・通所・施設サービスなど）の利用量（点数・金額など）を一定の期間についてまとめた集計情報のみを利用可能な情報として認める。

5. 秘匿処理される調査項目と処理内容

5.1 居住地・地域情報

調査項目	処理内容
居住地情報	機密保護レベル Ultra High での使用が承認された場合のみ利用可能
地域情報	機密保護レベル Very High 以上での使用が承認された場合のみ利用可能

5.2 本人・家族情報

調査項目	処理内容
本人生年月	月次を秘匿化し、生年のみを利用可能とする
配偶者生年月	月次を秘匿化し、生年のみを利用可能とする
本人学歴	特殊な学歴について具体的内容を秘匿化
配偶者学歴	特殊な学歴について具体的内容を秘匿化
結婚年月	月次を秘匿化し、年次のみを利用可能とする
配偶者死別年月	月次を秘匿化し、年次のみを利用可能とする
離婚年月	月次を秘匿化し、年次のみを利用可能とする
子ども人数	上位 3%以上の子ども人数の対象者をトップコード処理
子ども生年月	月次を秘匿化し、年次のみを利用可能とする
子どもの学歴	特殊な学歴について具体的内容を秘匿化
父親の年齢	上位 3%以上の年齢の父親の年齢に対してトップコード処理
母親の年齢	上位 3%以上の年齢の母親の年齢に対してトップコード処理
義理の父親の年齢	上位 3%以上の年齢の義理の父親の年齢に対してトップコード処理
義理の母親の年齢	上位 3%以上の年齢の義理の母親の年齢に対してトップコード処理
現在の仕事の種類	特殊な仕事の種類について具体的内容を秘匿化
現在の仕事の内容	具体的内容を秘匿化
勤め先の事業内容	具体的内容を秘匿化
勤め先の従業員数	上位 3%以上の従業員数についてトップコード処理
54 歳時点における仕事の種類	特殊な仕事の種類について具体的内容を秘匿化
54 歳時点における仕事の内容	具体的内容を秘匿化
54 歳時点における事業の内容	具体的内容を秘匿化
54 歳時点における勤め先の従業員数	上位 3%以上の従業員数についてトップコード処理
配偶者の現在の仕事の種類	特殊な仕事の種類について具体的内容を秘匿化
配偶者の現在の仕事の内容	具体的内容を秘匿化

配偶者の勤め先の事業内容	具体的内容を秘匿化
配偶者の勤め先の従業員数	上位 3%以上の従業員数についてトップコード処理

5.3 経済的情報

調査項目	処理内容
可処分所得	上位 3%以上の可処分所得額に対してトップコード処理
税、社会保険料合計額	上位 3%以上の税、社会保険料合計額に対してトップコード処理
納税額	上位 3%以上の納税額に対してトップコード処理
社会保険料納付額	上位 3%以上の社会保険料納付額に対してトップコード処理
配偶者の可処分所得	上位 3%以上の可処分所得額に対してトップコード処理
配偶者の税、社会保険料合計額	上位 3%以上の税、社会保険料合計額に対してトップコード処理
配偶者の納税額	上位 3%以上の納税額に対してトップコード処理
配偶者の社会保険料納付額	上位 3%以上の社会保険料納付額に対してトップコード処理
預貯金額	上位 3%以上の預貯金額に対してトップコード処理
保有債券額	上位 3%以上の保有債券額に対してトップコード処理
保有株式額	上位 3%以上の保有株式額に対してトップコード処理
配偶者の預貯金額	上位 3%以上の預貯金額に対してトップコード処理
配偶者の保有債券額	上位 3%以上の保有債券額に対してトップコード処理
配偶者の保有株式額	上位 3%以上の保有株式額に対してトップコード処理
現在の時給額	上位 3%以上の時給額に対してトップコード処理
現在の日給額	上位 3%以上の日給額に対してトップコード処理
現在の月給額	上位 3%以上の月給額に対してトップコード処理
現在の年間賞与額	上位 3%以上の年間賞与額に対してトップコード処理
現在の年間事業所得	上位 3%以上の年間事業所得額に対してトップコード処理
54 歳時点における年収	上位 3%以上の 54 歳時点年収額に対してトップコード処理
公的年金受給額	上位 3%以上の公的年金受給額に対してトップコード処理
公的年金受給予想額	上位 3%以上の公的年金受給予想額に対してトップコード処理
配偶者の公的年金受給額	上位 3%以上の公的年金受給額に対してトップコード処理
配偶者の公的年金受給予想額	上位 3%以上の公的年金受給予想額に対してトップコード処理
個人年金受給額および受給予定額	上位 3%以上の個人年金受給額に対してトップコード処理
1 カ月の食費	上位 3%以上の食費額に対してトップコード処理
1 カ月の外食費	上位 3%以上の外食費額に対してトップコード処理
1 カ月の生活費	上位 3%以上の生活費額に対してトップコード処理
1 年間の耐久消費財購入費	上位 3%以上の耐久消費財購入額に対してトップコード処理
保有自動車価格	上位 3%以上の自動車価格に対してトップコード処理

保有住居築年数	上位 3%以上の築年数に対してトップコード処理
保有住居面積	上位 3%以上の面積に対してトップコード処理
保有住居売却予想額	上位 3%以上の予想額に対してトップコード処理
住宅ローン残高	上位 3%以上のローン残高に対してトップコード処理
住居以外の保有不動産売却予想額	上位 3%以上の売却予想額に対してトップコード処理
負債残高	上位 3%以上の負債残高に対してトップコード処理
1年後の目標貯蓄残高	上位 3%以上の目標貯蓄残高に対してトップコード処理
5年後の目標貯蓄残高	上位 3%以上の目標貯蓄残高に対してトップコード処理
最終的な目標貯蓄残高	上位 3%以上の目標貯蓄残高に対してトップコード処理
受取済み生前贈与・遺産額	上位 3%以上の生前贈与・遺産額に対してトップコード処理
受取予定の生前贈与・遺産額	上位 3%以上の生前贈与・遺産額に対してトップコード処理
残す予定の生前贈与・遺産額	上位 3%以上の生前贈与・遺産額に対してトップコード処理
公的医療保険料	上位 3%以上の保険料額に対してトップコード処理
私的健康保険料	上位 3%以上の保険料額に対してトップコード処理
介護保険料	上位 3%以上の保険料額に対してトップコード処理

5.3 レセプト情報

調査項目	処理内容
医療機関番号_番号	機密保護レベル Ultra High もしくは機密保護レベル Very High での使用が承認された場合のみ、別の通し番号に置き換えた形式で利用可能
介護保険事業所番号_番号	機密保護レベル Ultra High もしくは機密保護レベル Very High での使用が承認された場合のみ、別の通し番号に置き換えた形式で利用可能
医療機関および介護保険事業所の属性にかかわる情報	機密保護レベル Ultra High もしくは機密保護レベル Very High での使用が承認された場合のみ、研究上必要と認められた属性に関して ID 化した形式で利用可能
医療・介護の支出・給付費の情報	機密保護レベル High においては主要項目および一定期間について、集計された数値のみ利用可能

6. 標本の無作為抽出による個人情報秘匿処理

機密保護レベル High における研究利用が承認された研究者には、全サンプルから無作為に抽出された 9 割のデータが利用可能となる。