

(案)

賃金構造基本統計調査 匿名データの作成方針

1 基本的な考え方

本調査の匿名データ化については、「匿名データの作成・提供に係るガイドライン」（平成 21 年 2 月 17 日総務省政策統括官（統計基準担当）決定）の匿名化処理基準に準拠しつつ、本調査の特徴を踏まえ、所要の匿名化処理を講じる。

なお、匿名化処理基準は、平成 30 年賃金構造基本統計調査を基に作成し、作成する他の年次（平成 27 年～29 年、令和元年）については、これに準じて作成するものとする。

2 作成するデータの構成概要

賃金構造基本統計調査の個人票の労働者をレコードとし、当該労働者が属する事業所の情報を付加し、以下の匿名データを作成する。

（平成 30 年の場合）

調査本体の標本の大きさ	リサンプリング率	匿名データの標本の大きさ
約 130 万人 （約 5.7 万事業所）	約 40%	約 52 万人

※ 括弧内は、事業所数

3 適用する匿名化処理

本調査では、調査票情報に対して、以下の匿名化処理を適用する。

(1) リサンプリング

リサンプリングは、労働者を等確率に抽出し、抽出率が約 40%となるようにする。

(2) 識別情報

以下の匿名化処理を講じるほか、出現数が少なく個体識別リスクが高いレコードは削除する（詳細は、「匿名データの審査表」を参照）。

ア 地域情報の削除

他の項目（産業大分類等）との組み合わせにおいて、度数が少なく特定のリスクが高いことから削除する。

イ 事業所一連番号の削除

同一事業所の労働者であることが特定できることから削除する。

ウ 産業大分類「C：鉱業，採石業，砂利採取業」に属する労働者の削除

事業所数が少ないため削除する。

エ 初任給額及び採用人員の削除

採用情報の情報と結びつけることで、事業所の特定につながるため削除する。
（他の個人項目から算出する方法を提示）

オ 公営の事業所の削除

事業所数が少ないため削除する。

カ 労働者が限定されている項目の提供不可

産業分類及び事業所規模により記入する労働者が限定されている項目は提供しない。

キ 集計用乗率（復元倍率）

事業所抽出率に労働者抽出率を乗じたものを提供することとするが、事業所規模の大きな事業所については、一部再付与する。

ク その他

これら以外にも、個体識別リスクの低減のため、レコードの削除、トップ又はボトムコーディング、リコーディング、乱数によるレコード順の並び変え等、必要な措置を行う。

また、トップ又はボトムコーディングに当たっては、利用者の利便性を考慮し、実際の結果との乖離を最小限にするとともに、統計調査の本体集計の結果表章に用いられる分類を参考とする（しきい値 0.5%）。

更に、匿名データの作成・確認表を作成する際に、組み合わせ等により個体識別リスクが高いレコードが出現した場合は、今回追加する匿名化処理基準に準じ、削除等の匿名化処理を行う。