

第 8 回匿名データ作成方法ワーキンググループ

貸金構造基本統計調査の
匿名データの検討について（案）

抽出方法の概要 (厚生労働省HP等参照)

(1) 調査対象事業所等の抽出方法

- 母集団は、総務省の「事業所母集団データベース」
- 事業所を第1次抽出単位、労働者を第2次抽出単位とする層化二段抽出法
- 事業所を都道府県(47)×産業中分類(80)×事業所規模(8)別に層化
- 上記で決定した数の調査対象事業所を、母集団から層ごとに無作為に抽出
- 調査対象労働者は、各調査対象事業所が、層ごとの労働者抽出率に基づいて定まる数を無作為に抽出
- 目標精度は、常用労働者の1人平均所定内給与額について設定し、基本的に、都道府県、表章産業及び企業規模別の標準誤差率を5%以内に定めている。

(2) 抽出率

- 事業所抽出率は都道府県、産業及び事業所規模別に定めている。
- 労働者の抽出率は、100人以上の事業所については産業及び事業所規模別に、100人未満の事業所については事業所規模別に定めている。

(3) 復元倍率(乗率)

- 第1次抽出率の逆数及び第2次抽出率の逆数を乗じたもの

リサンプリング（案1）

- 『労働者』を無作為に抽出
- 方法
 - 調査票情報のソート
 - ・ 地域（県番号）> 産業（中分類）> 事業所規模> 企業規模> 事業所一連番号> 労働者番号
 - 全て同じリサンプリング率により、単純無作為に抽出
 - 復元倍率は、本体集計時のものに、リサンプリング率の逆数を乗じて作成
- メリット
 - 方法が単純なため、作業コストが低い
- デメリット
 - 同じ復元倍率を持つ常用労働者の数から、事業所規模が類推されるリスクがある
 - 特に、単一事業所企業の場合、グルーピング前の企業規模が類推されるリスクもある
- 留意点
 - 悉皆の事業所のうち、母集団一意二意となるものは削除することになるが、有用性に問題はないか
 - 復元倍率が同一の労働者は、同一事業所の者と特定されるリスクがある
 - ・ 抽出前後の労働者数により、復元倍率の調整をするか
 - 事業所規模別
 - ・ 1,000人以上の事業所については、悉皆層では、特定リスクが高い（対処が必要か？）
 - ・ 1,000人未満の事業所については事業所数の多さ等のため、復元倍率での類推は難しいと思われる

[補足]リサンプリング（案1）：作成方法の概要

1. 調査票情報のソート

- 地域（県番号）＞産業（中分類）＞事業所規模＞企業規模＞事業所一連番号＞労働者番号

2. 労働者の抽出

リサンプリング率に合わせて、等確率で抽出

3. 復元倍率の再作成

本体集計時のものに、リサンプリング率の逆数を乗じて作成

リサンプリング（案2）

- 『労働者』を産業分類別、事業所規模別に抽出

- 方法

- 調査票情報のソート

- 地域（県番号）>産業（中分類）>事業所規模>企業規模>事業所一連番号>労働者番号

- 提供する事業所規模1,000人以上の復元倍率を産業大分類ごとに均一化するために、以下の2つの条件を満たす事業所を対象して調整係数を作成する

- ① 事業所規模1,000人以上

- ② 悉皆層

- 事業所規模1,000人以上は3階層に分けられ、それぞれにおいて常用労働者の抽出率が異なる

- ✓ 1,000~4,999人、5,000~14,999人、15,000人以上

- 第2次抽出率が一番小さな15,000人以上にあわせて、1,000~4,999人、5,000~14,999人について抽出率を調整する

$$\text{事業所ごとの第2次抽出率を調整する係数} = \frac{\text{15,000人以上の抽出率}}{\text{1,000~4,999人（又は、5,000~14,999人）の抽出率}}$$

- リサンプリング率をこの係数で調整する（第1次抽出率にも留意）

- 上記①及び②以外の事業所については、（案1）と同様に労働者を等確率無作為に抽出する

- メリット

- 同じ復元倍率より、同一事業所（特に単一事業所企業の規模）を類推させない

- デメリット

- 作業が複雑である

- 留意点

- 悉皆層の事業所のうち、母集団一意二意となるものは削除することになるが、有用性に問題はないか

- 事業所規模15,000人以上の第2次抽出率は、全産業とも最小である

[補足] リサンプリング（案2）：作成方法の概要

1. 調査票情報のソート

- 地域（県番号） > 産業（中分類） > 事業所規模 > 企業規模 > 事業所一連番号 > 労働者番号

2. 労働者の抽出

- ① 事業所規模1,000人以上、かつ、悉皆層の事業所
 - i. 産業中分類別、提供する事業所規模別の事業所一連番号ごとに設定したリサンプリング率での抽出前後の労働者数の度数表を作成
 - ii. 産業大分類ごとに労働者の抽出率が均一になるように、抽出率が事業所規模15,000人以上のものとなる（に近くなる）ように、調査時の労働者の抽出率を調整する係数を事業所ごとに作成（iで作成した度数表も考慮する）
 - iii. 上記iiの係数で調整した労働者の抽出率及びリサンプリング率に合わせて、等確率で抽出
- ② ①以外の事業所
 - リサンプリング率に合わせて、等確率で抽出

3. 復元倍率の再作成

- ① 事業所規模1,000人以上、かつ、悉皆層の事業所
本体集計時のものに、調整されたリサンプリング率の逆数を乗じて作成
- ② ①以外の事業所
本体集計時のものに、リサンプリング率の逆数を乗じて作成

リサンプリング（案3）

- 『事業所』を産業分類別、事業所規模別に抽出（その後、労働者を抽出）
- 方法
 - 調査票情報のソート
 - 地域（県番号）>産業（中分類）>事業所規模>企業規模>事業所一連番号>労働者番号
 - 産業中分類ごとの調査対象事業所数にリサンプリング率で乗じた値
 - ① 1,000人未満
 - i. リサンプリング事業所数 1以上
設定したリサンプリング率で等確率で産業中分類ごとに抽出し、当該事業所の労働者は全て対象
 - ii. リサンプリング事業所数 1未満
リサンプリングの対象としない
 - ② 1,000人以上
 - ✓ 産業中分類ごとに労働者の抽出率が均一になるように、常用労働者の抽出率が事業所規模15,000人以上のものとなる（に近くなる）ように、調査時の労働者の抽出率を調整する係数を事業所ごとに作成（案2と同様）
 - i. リサンプリング事業所数 1以上
設定したリサンプリング率で等確率で産業中分類ごとに抽出し、当該事業所の労働者は全て対象
 - ii. リサンプリング事業所数 1未満
悉皆層が含まれているかなど、確認を行い提供の可否を検討する復元倍率もリサンプリング率及び調整係数から作成
- メリット
 - 方法が単純なため、作業コストが低い
- デメリット
 - 悉皆層の事業所は、規模の大きい事業所（特に単一事業所企業）の特定のリスクがある
- 留意点
 - 悉皆層の事業所のうち、母集団一意二意となるものは削除することになるが、有用性に問題はないか
 - 事業所規模15,000人以上の第2次抽出率は、全産業とも最小

[補足]リサンプリング（案3）：作成方法の概要

1. 調査票情報のソート

- 地域(県番号)＞産業(中分類)＞事業所規模＞企業規模＞事業所一連番号＞労働者番号

2. 事業所の抽出

- 産業中分類別、事業所規模別の事業所一連番号ごとに労働者数の度数表を作成
- 産業中分類ごとの調査対象事業所数にリサンプリング率で乗じる

① 1,000人未満

i. リサンプリング事業所数 1以上

- ✓ 設定したリサンプリング率で等確率で、産業中分類ごとに抽出
- ✓ リサンプリングされた事業所の労働者は全て対象

ii. リサンプリング事業所数 1未満

リサンプリングの対象としない

② 1,000人以上

産業大分類ごとに労働者の抽出率が均一になるように、常用労働者の抽出率が事業所規模15,000人以上のものとなる（に近くなる）ように、調査時の労働者の抽出率を調整する係数を事業所ごとに作成（案2と同様）

i. リサンプリング事業所数 1以上

- ✓ 設定したリサンプリング率で等確率で、産業中分類ごとに抽出
- ✓ 上記の係数で調整した労働者の抽出率及びリサンプリング率に合わせて、等確率で抽出

ii. リサンプリング事業所数 1未満

作成した度数表、悉皆層が含まれているかなど、確認を行い提供の可否を検討する

3. 復元倍率の再作成

- 上記2.①の事業所

本体集計時のものに、リサンプリング率の逆数を乗じて作成

- 上記2.②で抽出対象となった事業所

本体集計時のものに、調整されたリサンプリング率の逆数を乗じて作成

今後について

●次回の作成方法WG（第9回）

➤日時

- 4月上旬以降を予定

➤議題

- リサンプリング方法及び提供する項目の匿名化処理について、WG案を策定する

※状況に応じて、度数表等の再作成を行い、再検討を行う