

第 9 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）

【整理①】 提供する事業所事項項目について

- 地域 → 情報の削除（提供しない（地域区分なし（全国）））
 - 事業所一連番号 → 情報の削除（提供しない）
 - 産業分類 → 大分類
C 鉱業，採石業，砂利採取業 → レコード削除対象
 - 企業規模
5 区分 → 5～29人、30～99人、100～499人、500～999人、1,000人以上
 - 初任給額 → 情報の削除（提供しない）
計算方法を提示（個人項目から算出）
 - 加工項目
 - 民・公区分 → 公営をレコード削除（民営のみを提供）
 - 復元倍率は、リサンプリング方法により検討
- ※ 地域、産業大分類、従業員規模により母集団（経済センサス等の当該年の情報）と照合し、一意二意となるレコード削除

【整理②】 提供する個人事項項目について

(匿名化処理する項目のみ (トップコーディング及びボトムコーディングのしきい値は、0.5%))

●年齢

➤ グループニング → 5歳階級 (ただし、24歳以下は、15～17歳、18～19歳、20～21歳、22～24歳) で提供)

➤ トップコーディング → 75歳以上

●労働者の種類 → 情報の削除 (提供しない)

●役職番号 → 情報の削除 (短時間労働者は、提供しない)

●職種番号

➤ グループニング → 7区分

●在留資格番号

➤ グループニング → 6区分

【整理③】 提供する個人事項項目について (続き)

(匿名化処理する項目のみ (トップコーディング及びボトムコーディングのしきい値は、0.5%))

項目	匿名化の方法	対象	
		一般労働者	短時間労働者
● 勤続年数	トップコーディング	44年以上	43年以上
● 実労働日数	ボトムコーディング	13日以下 ⇒不要か?	なし
	トップコーディング	28日以上	28日以上
● 所定内実労働時間数	ボトムコーディング	101時間以下 ⇒不要か?	なし
	トップコーディング	215時間以上	176時間以上
● 超過実労働時間数	トップコーディング	83時間以上	40時間以上
● きまって支給する現金 給与額	トップコーディング	96万円以上 ⇒特異値として処理か?	37万円以上 ⇒特異値として処理か?
● 超過労働給与額	トップコーディング	20万円以上 ⇒特異値として処理か?	5万円以上 ⇒特異値として処理か?
● 昨年1年間の賞与、期 末手当等特別給与額	トップコーディング	546万円以上 ⇒特異値として処理か?	118万円以上 ⇒特異値として処理か?

● 通勤手当、精皆勤手当、家族手当 → 情報の削除 (提供しない)

リサンプリング

- リサンプリング率 → 40%

- 方法は、以下の3案

- (案1) 『労働者』を無作為に抽出

- 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

- (案2) 『労働者』を産業分類別、事業所規模別に抽出

- 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、事業所規模1,000人以上及び悉皆層に留意
- 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
- リサンプリング率をこの係数で調整 (第1次抽出率にも留意)

- (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)

- 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に対して1をしきい値とし、事業所規模 (1,000人未満、1,000以上) により、リサンプリングの対象か否かを判断
- リサンプリング率は、案2と同様

- 抽出後の歪みの少ない方法をベースに検討

今後について

●次回の作成方法WG（第10回）

➤日時

- 5月中旬以降を予定

➤議題

- WG案のとりまとめ

※状況に応じて、度数表等の再作成を行い、再検討を行う