

第11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）

リサンプリング（1）

- リサンプリング率は「40%」とし、労働者を等確率で抽出
- 前回までに提示した案の問題点
 - （案1）『労働者』を無作為に抽出
 - 復元倍率から事業所規模が推測され、特定化されるリスク
 - （案2）『労働者』を産業分類別、事業所規模別に抽出
 - （第2次抽出率が一番小さな事業所規模15,000人以上にあわせて、1,000人以上の抽出率を調整）
 - 復元倍率については、種類が多く、値の幅も大きいため、最大値に合わせてと抽出される労働者数が少なくなる
- ➔ 事業所規模1,000人以上の復元倍率を産業大分類ごとにできるだけ均一化（クラスター化）し、事業所を特定されるリスクに対処する
- ➔ 同値の復元倍率に対する事業所数を3以上に集約する
 - ※ 復元倍率は、以下のいずれにかにより算出
 - 事業所抽出率 × 常用労働者抽出率(雇用形態1~4)
 - 事業所抽出率 × 臨時労働者抽出率(雇用形態5)
 - 事業所抽出率 × [事業所の労働者数 / 個人票の労働者数]

リサンプリング（2）

- 事業所規模1,000人以上の復元倍率を産業大分類ごとにできるだけ均一化（クラスター化）
 - 同値の復元倍率に対する事業所数により、以下の手順で復元倍率を調整（3事業所以上に集約する）

- ① 同値の復元倍率に対して、1及び2事業所数について（調整案②(第10回)）
 - 集約する事業所数の合計が3以上とする（原則、1、2事業所では行わない）
 - 「集約する復元倍率」を基に範囲内を設定し、その範囲内の事業所を対象とし、復元倍率等を調整する※1

を考慮し、以下のいずれかの方法で行う

- ✓ その1：「集約する復元倍率」は、設定した範囲内（近傍）の労働者数が最も多い値とする（労働者数はそのまま）
- ✓ その2：「集約する復元倍率」を上限とした範囲を設定し、調整した労働者数を算出する

$$\text{（調整した労働者数）} = \text{（調整対象の復元倍率の労働者数）} \times \frac{\text{（調整対象の復元倍率）}}{\text{（集約する復元倍率）}}$$

- ✓ その3：「集約する復元倍率」より設定した範囲内において、復元倍率を以下により再算出する（労働者数はそのまま）

$$\text{（再算出した復元倍率）} = \frac{\sum \{(\text{復元倍率}) \times (\text{当該復元倍率の労働者数})\}}{\text{（範囲内の労働者数）}}$$

- ② ①の対象にならなかった復元倍率については、事業所数が3以上の場合、復元倍率はそのままとする（調整案①(第10回)）
- ③ 復元倍率が（極めて）大きく、事業所数が1、2のものは削除（抽出対象としない）（調整案③(第10回)）

- ※1 範囲の選定については、復元倍率の分布の状況により、できるだけ削除対象を少なくすることに留意
- ※2 ①、②はともに考慮し、調整を行う

今後について

●次回の作成方法WG（第12回）

➤日時

- 8月中旬以降を予定

➤議題

- WG案のとりまとめ