

第10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）

# 個人項目のトップコーディング及びボトムコーディングについて

- 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	<b>44年以上</b>	<b>43年以上</b>
• 実労働日数	ボトムコーディング	<b>13日以下</b>	<b>なし</b>
	トップコーディング	<b>28日以上</b>	<b>28日以上</b>
• 所定内実労働時間数	ボトムコーディング	<b>101時間以下</b>	<b>なし</b>
	トップコーディング	<b>215時間以上</b>	<b>176時間以上</b>
• 超過実労働時間数	トップコーディング	<b>83時間以上</b>	<b>40時間以上</b>
• きまって支給する現金給与額	トップコーディング	<b>96万円以上</b>	<b>37万円以上</b>
• 超過労働給与額	トップコーディング	<b>20万円以上</b>	<b>5万円以上</b>
• 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	<b>546万円以上</b>	<b>118万円以上</b>
• 年齢	ボトムコーディング	<b>なし</b>	
	トップコーディング	<b>75歳以上</b>	

※年齢のグルーピングは、5歳階級（ただし、24歳以下は、15～17、18～19、20～21、22～24歳）

# リサンプリング

- リサンプリング率は「40%」とし、労働者を等確率で抽出
- 提示した案の問題点
  - (案1) 『労働者』を無作為に抽出
    - ⇒ 復元倍率から企業規模が推測され、特定化のリスク
  - (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企業規模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、値の幅も大きいいため、最大値に合わせると抽出される労働者数が少なくなる
- 問題点のための調整案
  - 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
  - 3事業所以上をまとめ、調整（原則、1、2事業所では行わない）
    - ・ 調整案① → 事業所数が3以上の場合、復元倍率はそのまま
    - ・ 調整案② → 復元倍率の値に対して、1及び2事業所数の場合
      - ✓ その1：復元倍率を近傍の労働者数が多い値に集約
      - ✓ その2：復元倍率をある値に集約し、労働者数を調整
      - ✓ その3：復元倍率を近い事業所をまとめ復元倍率の平均値を算出し調整、労働者数をそのまま
    - ・ 調整案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）
- 抽出後の歪みを検証

# 今後について

---

## ●次回の作成方法WG（第11回）

### ➤日時

- 7月上旬以降を予定

### ➤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