

平成 23 年 6 月 7 日

早稲田大学・黒田祥子

『労働力調査』に係る匿名データの作成について

●今回の提供に関して

① 地域区分について

失業率等には地域によって大きな違いがあることは公表値からも周知の通りであり、たとえば、データ提供予定期間の直近年である 2007 年平均の失業率は、東海地方が 2.7%に対して、北海道は 5%である。このように、労働市場の需給度合が地域間で相当程度異なることを鑑みると、精緻な分析を行うためには地域間の違いを考慮した分析のニーズは相当程度高いのではないかと思われる。

② 月末一週間に仕事をした時間について

90 時間以上でトップコーディングをすることを予定されているが、労働時間は外観の識別が難しいことから、トップコーディングをする必要性はそれほどないのではないかと考えられる。

③ 年齢の刻み

このほか、地域情報を一切出さないのであれば、他に原データのまま提供できるものがないのかどうかをご検討いただければありがたい。たとえば、年齢は 5 歳刻みでグルーピングしているが、居住地域が特定化できないのであれば、年齢はもう少し細かい刻みで提供することも可能ではないか。

●中長期的な要望

① データ提供期間

今回の『労働力調査』の匿名データ化は、提供期間を平成元年から 19 年まで予定しており、1990 年代の分析も可能となることからユーザーとしては大変ありがたい措置である。データのクリーニングコストなども相当程度かかることが推察されることから、直ちにと申し上げるつもりはないが、可能な限りさらに遡及をし、1980 年代分まで提供いただけるようにご尽力いただければ幸いである。

② 特定調査票の提供

今回は基礎調査票のみ提供とのことだが、特定調査票には学歴や収入等、分析を行うう

えでは非常に重要な情報が多数含まれており、ニーズが非常に高いことが推測される。特定調査票の情報も併せた提供を中長期的にはぜひご検討いただければありがたい。

③ パネル化

『労働力調査』は同一世帯を2か月連続で調査し、さらに翌年にも2か月連続で調査を実施していることから、政府統計としては2か月×2年のパネル化が可能となる希少な統計である。1年前に失業していた人が翌月あるいは翌年にどのような就業状態に移行したかといった遷移の分析はニーズが高く、ぜひ将来的にはパネルデータの提供も視野に入れていただきたい。

④ ノイズの付与やスワッピング等の検討

『労働力調査』に限られたことではないが、米国で広く実施されているとされるノイズの付与やスワッピングも将来的にご検討いただければ、結果として提供の幅が広がるのであればユーザーとしては大変ありがたいのではないかとと思われる。