

R パッケージによるメタデータの付与 とデータ特有の性質に適した分析方法の検証

一橋大学経済研究所 白川 清美
株式会社タクミインフォメーションテクノロジー 中松 建

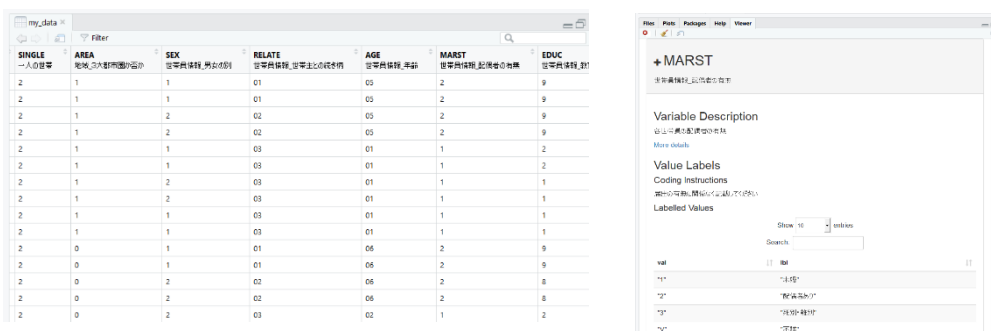
近年、政府統計の二次利用やオープンデータへの関心が特に高まっており、大学等の研究者による実証分析や公共サービスの利便性向上への活用が期待されている。しかしながら、これらのデータは、データの特長把握やデータ分析開始までのハンドリングが多く、準備に要する時間が多くなっている。これらの分析の開始までに時間を要する要因の一つとして、分析を行うには変数名と変数ラベルおよび符号表(コードリスト)だけでなく、分析の内容に応じてサンプリング情報や調査票情報などの附帯情報が必要になることが挙げられ、これらの情報をメタデータとして効率的に取り扱う方法が必要である。

この課題に対して、国際的に統合した人口マイクロデータを公開している IPUMS が開発・配布している R パッケージ「ipumsr」とそのメタデータ形式などの機能利用と他のデータへの拡張性について紹介する(図1参照)。

また他の要因として、データの構造が複雑な場合、データの分割や構造の変換が必要になることが挙げられる。特にデータ構造の変換が必要な場合には、時間がかかるだけでなく情報が十分に利用されずに埋もれてしまうことも考えられる。

そのためデータ特有の性質や含まれている情報を十分に活かすため、データのレベル別(世帯、個人など)の分割やデータ構造変換の検討を行っている。

当日は社会生活基本調査を例として、「ipumsr」の利用とメタデータの作成の際の課題、また分析に適したデータの構造や方法に関して検討したものを発表する。



SINGLE	AREA	SEX	RELATE	AGE	MARST	EDUC
一人の世帯	地域、国大県市町村の区分	世帯員性別、男女性別	世帯員関係、世帯主との関係	世帯員年齢、年齢	世帯員婚姻、結婚者の有無	世帯員学歴、学歴
2	1	1	01	05	2	9
2	1	1	01	05	2	9
2	1	2	02	05	2	9
2	1	2	02	05	2	9
2	1	1	03	01	1	2
2	1	1	03	01	1	2
2	1	2	03	01	1	1
2	1	2	03	01	1	1
2	1	1	03	01	1	1
2	1	1	03	01	1	1
2	0	1	01	06	2	9
2	0	1	01	06	2	9
2	0	2	02	06	2	8
2	0	2	02	06	2	8
2	0	2	03	02	1	2

図1 事例：ipumsr によるメタデータの付与

「ipumsr」 <https://github.com/mnpopcenter/ipumsr> (参照 2019 年 6 月 28 日)