

労働力調査における外れ値対応の試み

総務省統計局 小田 大輔
白松 千明

1. 労働力調査の乗率及び修正倍率

労働力調査は、国勢調査調査区を第1次抽出単位、住戸を第2次抽出単位とする標本調査で、国勢調査時点の人口を基に乗率（ウエイト）を算出している。また、国勢調査後に調査区内の世帯数が大幅に増加した場合は住戸の抽出間隔を変更することを認めており、その際は修正倍率（修正前後の抽出間隔の比）を乗じて乗率を調整しているが、調整された乗率が調査時点の母集団における当該調査区の代表性と一致しない可能性がある。さらに、代表性が正しくない世帯の特性が他の世帯とかけ離れていた場合は、推定値が真の値から大きく乖離する可能性がある。

本研究では、そのような事態を回避するため、国勢調査後に世帯数が増加したため住戸の抽出間隔を変更した調査区について①調査区の総人口線型推定値の上限を設けた場合、②修正倍率の上限値を2倍とした場合の総人口線型推定値を算出し、その効果の妥当性を視覚的に確認する。

2. 総人口線型推定値の地域別上限値 ($\hat{P}_{k,U}$) 及び調査区別修正倍率上限値 ($m_{j,U}$) の算出

調査区 (j) が労働力調査の調査区に抽出された場合の総人口線型推定値 (\hat{P}_j) を、国勢調査結果を基に算出し、その全国上位0.1%点 ($\hat{P}_{0.1\%}$)、第1四分位点 ($\hat{P}_{1/4}$) 及び第3四分位点 ($\hat{P}_{3/4}$) を基に定数 C_U を求める。同様に求めた各地域 (k) の第1四分位点 ($\hat{P}_{k,1/4}$)、第3四分位点 ($\hat{P}_{k,3/4}$) 及び定数 C_U を基に、地域別上限値 ($\hat{P}_{k,U}$) を求める。

$$\hat{P}_i = g_i \frac{P_i}{w_i}, \quad C_U = \frac{\hat{P}_{0.1\%} - \hat{P}_{3/4}}{\hat{P}_{3/4} - \hat{P}_{1/4}}, \quad \hat{P}_{k,U} = \hat{P}_{k,3/4} + C_U (\hat{P}_{k,3/4} - \hat{P}_{k,1/4})$$

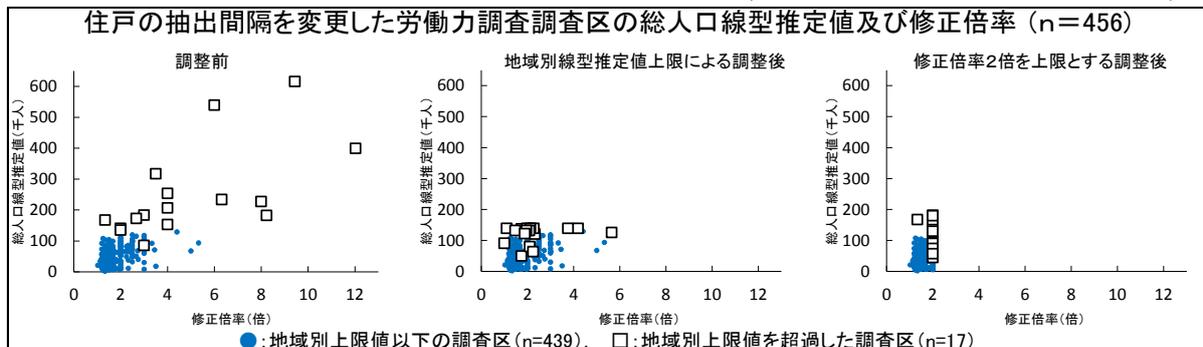
P_i : 総人口 (国勢調査結果)
 w_i : 住戸の抽出間隔
 g_i : 層の抽出間隔

次に、実際に住戸の抽出間隔を変更 ($w'_j > w_j$) した調査区 (j) について、総人口線型推定値 (\hat{P}'_j) が地域別上限値 ($\hat{P}_{k,U}$) と等しくなる調査区別修正倍率上限値 ($m_{j,U}$) を算出し、 $m_j \leq m_{j,U}$ とした場合及び $m_j \leq 2$ とした場合の総人口線型推定値 (\hat{P}'_j) 及び修正倍率 (m_j) の分布を確認する。

$$m_j = \frac{w'_j}{w_j}, \quad m_{j,U} = \frac{\hat{P}_{k,U} h_j w'_j}{g_j P_j h'_j}, \quad \hat{P}'_j = g_j m_j \frac{P_j h'_j / h_j}{w'_j}$$

h_j : 換算世帯数※ (国勢調査結果)
 h'_j : 換算世帯数※ (修正倍率変更後)
 w'_j : 住戸の抽出間隔 (修正倍率変更後)

※ 換算世帯数 = 世帯人員が二人以上の一般世帯数 + (世帯人員が一人の一般世帯数 + 施設等の世帯人員) / 3



平成22年国勢調査に基づく労働力調査の調査区。調査期間は平成25年5月から平成30年4月のいずれかの月に該当