

宿泊料の指数作成に関する新しい試み

総務省統計局 尾中 裕一

1. はじめに

近年、我が国においてもネットショッピングが急激に増加しており、消費者物価指数にネット販売価格を採用する必要性が高まっている。現在、消費者物価指数ではネット販売価格を取り入れている品目もあるが、原則として実店舗で売られている小売価格を採用している。また、諸外国ではネット販売価格等の多種多様かつ大量の情報をウェブサイトから自動的に収集する技術（ウェブスクレイピング）によって収集した価格データを用いて価格指数を作成する取組も進んでいる。

このような状況の中、総務省統計局では、消費者物価指数の2020年基準改定においてネット販売価格を消費者物価指数に更に取り入れるため、旅行サービス（航空運賃、宿泊料、外国パック旅行費）についてウェブスクレイピングを活用して価格データを収集し、価格指数を作成する検討を行っている。

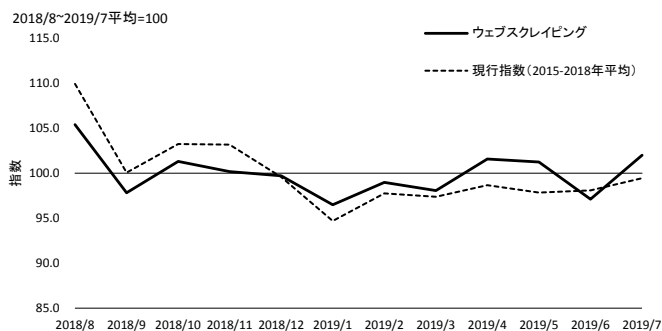
本報告では、宿泊料について、ウェブスクレイピングを用いた価格収集から外れ値処理やデータ補完の統計的処理及び指数作成方法までの試みを紹介する。

2. 価格データの収集

今回の試算に用いた宿泊料の価格データは、複数の総合予約サイトにおいて宿泊施設を指定して検索を行い、その結果をウェブスクレイピングにより収集した。その際、価格データのほか、宿泊日、宿泊プラン名、部屋の間取り、食事の有無などの附帯情報も合わせて収集した。

3. 価格指数の試算

ウェブスクレイピングで収集した価格データには、検索条件に合致する全てのプランが含まれており、スイートルームのような大きな品質差のある高価格帯のものや、タイムセールのような一時的な低価格帯のものなどが混在していたため、外れ値として除外した。また、満室や休業などによりサイト上に価格が表示されない場合、データに欠測が生じる。欠測値を無視して指数計算を行うと偏りが生じるおそれがあるため、データの補完を行った。報告では、外れ値の除外手法の検討や実測値データセットの回帰分析から欠測値を推定し補完する手法の検討、補完後のデータによる指数の試算結果等を紹介する。



図：試算結果と現行指数との比較