

表計算ソフトウェア演習への公開データの活用

Utilization of public data for Spreadsheet Software Exercise

吉根 勝美*1

Katsumi YOSHINE*1

*1 南山大学経済学部

*1 Faculty of Economics, Nanzan University

Email: kyoshine@nanzan-u.ac.jp

あらまし：大学新入生を対象にした表計算ソフトウェアの演習では、架空のデータではなく、その後の専門的な学習につながる実際のデータを用いる方が望ましい。本稿では、経済学部への新入生を対象にして、ウェブスクレイピングの技術を用いて、公開データから表計算演習のための教材を開発した2つの事例を紹介する。具体的には、公開されている地価情報や求人情報の個別データから、表計算演習の例題として使用可能な教材を生成する。

キーワード：公開データ、公的統計、データ処理演習、統計教育、ウェブスクレイピング

1. はじめに

大学生に対する統計教育において実際のデータを使用すること—Use real data—の重要性は、アメリカ統計学会による「統計教育における評価と指導のガイドライン」の College Report 2016 でも示されている⁽¹⁾。筆者が所属する学部の1年次必修科目「データ処理入門」では、行政機関や地方自治体が発表している公的統計を使用している。例えば、愛知県統計年鑑から百貨店の販売額を抽出して、表計算ソフトウェアの基本操作演習の教材としているが、単位をわざと伏せておき、県内の年間百貨店販売額の規模を推測させる教材でもある(図1)。

公開されている公的統計の多くは、調査統計や業務統計であり、集計される前の個々のデータが有する“reality”が失われてしまうのはやむを得ない。しかし、ウェブ公開されている公的な統計データの中には、情報提供を目的とする検索機能を用いて、個々のデータに1件1件アクセスすることが可能な場合がある。そこで本稿では、ウェブスクレイピングの技術で自動収集して教材化する2つの事例を示し、統計教育教材としての可能性を考えたい。

	A	B	C	D	E	F	G
1	百貨店販売額 (愛知県、平成29年)						
2	紳士服・洋装	婦人・子供服	その他の衣類	身の回り品	飲食料品	その他	
3	1月	3325	11560	880	6122	8906	9582
4	2月	1885	7701	526	4356	9388	9081
5	3月	2601	11321	826	5616	9208	12059
6	4月	2587	9220	638	4867	7369	9167
7	5月	2562	9000	1570	5064	7420	9327
8	6月	2322	8158	615	5083	9191	10101
9	7月	2641	9667	840	6011	11632	10939
10	8月	1705	6961	548	4535	7432	9520
11	9月	1991	8600	1250	4728	7643	10234
12	10月	2918	9388	668	5194	7701	9512
13	11月	3186	9587	961	5116	10138	10296
14	12月	3691	10729	664	7317	16904	13763

図1 公的統計による表計算ソフト演習用教材の例

2. 事例1 (地価情報)

国土交通省の地価公示のうち、「令和2年地価公示にみる地価の状況」⁽²⁾を参照すると、全国の主要都

市の住宅地の平均価格等がまとめられている。例えば、名古屋市千種区について、標準地数26、平均価格248,900(円/㎡)、上位の価格385,000(円/㎡)、下位の価格159,000(円/㎡)という記載があり、基本統計量の説明題材にふさわしい“real data”である。

この個別データは、「標準地・基準地検索システム～国土交通省地価公示・都道府県地価調査～」⁽³⁾から得ることができる。例えば、図2のように「名古屋市千種区の住宅地」を検索条件として指定すると、図3のように26か所の調査地点の地価公示情報が2ページにわたって表示される。

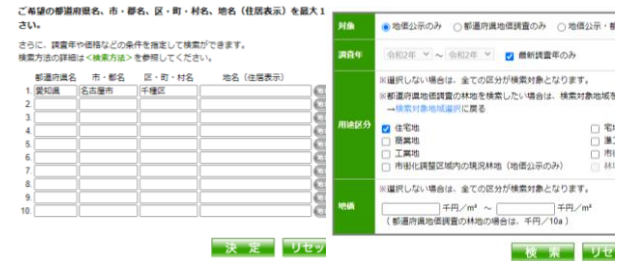


図2 公示地価の検索条件指定画面



図3 公示地価の検索結果表示

この検索過程をウェブスクレイピングで自動実行するプログラムを作成した。言語はPython 3.7.5で、使用ライブラリは、Requests 2.22.0、BeautifulSoup 4.8.1、Selenium 3.141.0である。使用ブラウザは

Google Chrome 83.0.4103.97 (64 ビット) で, Selenium WebDriver は ChromeDriver 83.0.4103.39 である.

作成したソースプログラムはおおよそ 50 行からなる. 住所と地価を抽出した結果を CSV ファイルに出力した. その一部を図 4 に示す. 平均, 最小値, 最大値を計算するための例題であるとともに, 文字列の一括置換機能 ((円/m²)の削除) やソート機能 (価格の昇順, 降順) を学ぶ演習にも利用可能である.

	A	B
1	address	price
2	愛知県名古屋千種区新西1丁目613番	159,000(円/m ²)
3	愛知県名古屋千種区清明山1丁目907番	196,000(円/m ²)
26	愛知県名古屋千種区神田2丁目904番	264,000(円/m ²)
27	愛知県名古屋千種区丸山町2丁目63番3外	288,000(円/m ²)

図 4 個別の地価情報から生成された CSV ファイル

形態, 賃金額) のデータを抽出して, CSV ファイルを出力する Python プログラムを作成した. ソースプログラムはおおよそ 80 行である. このプログラムを 2020 年 6 月 9 日に実行したところ, 検索結果は 79 件あった. 図 7 に CSV ファイルの一部を示す. このデータを用いて, 文字列操作による賃金額の下限, 上限の抽出, フィルター機能によるデータ抽出の演習が可能である.

A	B	C	D	E
求人区分	雇用形態	求人種別	賃金形態	賃金額
フルタイム	正社員	一般	月給	245,000円~245,000円
パート	パート労働者	一般	時給	980円~1,075円
フルタイム	正社員	一般	月給	200,000円~200,000円
フルタイム	正社員	一般	月給	230,000円~240,000円
パート	パート労働者	一般	時給	1,300円~1,500円
フルタイム	正社員	一般	月給	185,000円~215,000円

図 7 個別の求人情報から生成された CSV ファイル

3. 事例 2 (求人情報)

厚生労働省は, 公共職業安定所 (ハローワーク) における求人, 求職, 就職の状況をとりまとめ, 求人倍率などの指標を作成し, 一般職業紹介状況として毎月公表している. このようなデータは, 学習者にマクロ的な視点で観察させるのに必要ではあるが, “reality” を感じさせるのは難しい. そこで, 厚生労働省職業安定局が提供する「ハローワークインターネットサービス」⁽⁴⁾ から地域の実際の求人情報を抽出して, 表計算ソフト演習用の教材としたい.

図 5 求人情報の検索条件指定画面

図 6 求人情報の検索結果 1 件分の表示

就業場所と希望職種を「名古屋市瑞穂区」「介護、福祉」と指定すると (図 5), 求人情報が 1 ページに 30 件ずつ表示される (図 6). この検索過程を自動実行し, 求人区分, 雇用形態, 賃金 (求人種別, 賃金

4. おわりに

本稿では, 学習者が “reality” をより感じられるデータから教材を作成することを目的として, ウェブスクレイピングの技術を用いて, 地価公示データや求人情報データを収集して, 表計算ソフトウェア演習で利用可能な教材データの作成事例を紹介した. いずれも Python プログラムにより, データの自動収集が可能となった.

ウェブスクレイピングによりデータを自動的に収集するときには, Web サービスの利用規約でスクレイピングが禁止されていないかどうかを確認する必要がある. また, Web サービスに過度な負担をかけるようにする必要はない. 今回作成したプログラムでは, 待機時間を随時設けている.

政府が全大学生の受講を目指している初級水準の人工知能教育では, 表計算ソフトによるデータ分析実習用に, 実際の商品の購買データなどを実社会から集め, 各大学向けに公開する計画があるという新聞報道もある⁽⁵⁾. このように, 大学生の統計教育における “Use real data” の重要性は変わることないので, 同様の事例を積み上げてアーカイブ化し, 教材データの共同利用の促進を目指したい.

参考文献

- (1) GAISE College Report ASA Revision Committee: “Guidelines for Assessment and Instruction in Statistics Education College Report 2016”, <http://www.amstat.org/education/gaise> (2016)
- (2) 国土交通省: “令和 2 年地価公示”, https://www.mlit.go.jp/totikensangyo/totikensangyo_fr4_000380.html
- (3) 国土交通省: “標準地・基準地検索システム～国土交通省地価公示・都道府県地価調査～”, <https://www.land.mlit.go.jp/landPrice/AriaServlet?MOD=0&TYP=0>
- (4) 厚生労働省職業安定局: “ハローワークインターネットサービス”, <https://www.hellowork.mhlw.go.jp/>
- (5) 日本経済新聞: “AI「普段使い」へ教育一政府・大学などモデル案、学生に基礎力 担い手拡大”, 2020 年 2 月 4 日付夕刊 (2020)