

重回帰分析によるゴキブリ発生リスクの推定

日付	2021 年 09 月 15 日
名前	藤田憲伸

1 自由研究の背景

夏休み中に勤続 4 年のバイト先で初めてゴキブリに遭遇し、真剣にバイトを辞めるか悩んだことで自分はとても虫が嫌いだということを再認識した。

本自由研究では、重回帰分析を用いて都道府県毎のゴキブリ発生リスクの学習を行い、世界の主要な都市のゴキブリ発生リスクを推定することで将来の勤務地候補を考える。

ゴキブリ発生リスクとは、ゴキブリ駆除業界で最大級の Web サイトであるゴキブリ駆除マイスター [1] での相談件数を都道府県別世帯数で割り、平均 0 分散 1 に標準化したものとする。

2 自由研究内容

本自由研究では Python の機械学習ライブラリの scikit-learn を用いて重回帰分析を行う。説明変数には各都道府県の、人口 (人)、面積 (km²)、1km²あたりのごみ排出量 (t/km²)、年平均気温 (°C)、年平均湿度 (%) を、目的変数には各都道府県のゴキブリ発生リスクを用いて学習を行う。作成した線形モデルを用いて、アメリカ (ワシントン D.C.)、ロシア (モスクワ)、ドイツ (ベルリン)、韓国 (ソウル)、オーストラリア (キャンベラ)、の 5 つの都市についてゴキブリ発生リスクを推定する。

また、作成した線形モデルの標準偏回帰係数からそれぞれの説明変数が目的変数に与える影響について考察する。

3 数値例

3.1 実験条件

表 1 に実験に用いたデータセットの詳細を示す。

3.2 実行結果

scikit-learn の LinearRegression を用いて作成したモデルで、世界の主要な都市のゴキブリ発生リスクを推定した結果を表 2 に示す。また、日本の都道府県のゴキブリ発生リスク上位と下位それぞれ 2 都道府県を表 3 に示す。

世界の主要な都市の説明変数を表 4 に示す。また、作成した線形モデルの標準偏回帰係数を表 5 に示す。

3.3 考察

表 2 の全ての都市のゴキブリ発生リスクが表 3 の大阪府と比較して低いことから現在住んでいる大阪よりはゴキブ

表 1 実験に用いたデータセットの詳細

引用元 Web サイト	項目
総務省統計局国勢調査 [2]	都道府県人口 都道府県世帯数
環境省 [3]	都道府県ごみ排出量
国土交通省国土地理院 [4]	都道府県面積
国土交通省気象庁 [5]	都道府県年平均気温 都道府県年平均湿度 外国年平均気温
ゴキブリ駆除マイスター [1]	都道府県ゴキブリ相談件数
アメリカ合衆国国勢調査局 [6]	ワシントン D.C. 人口 モスクワ人口 外国面積
ユーロスタット [7]	ベルリン人口
国際連合 [8]	ソウル人口 キャンベラ人口
経済協力開発機構 [9]	外国ごみ排出量
Weatherbase[10]	外国年平均湿度

表 2 推定した世界の主要な都市のゴキブリ発生リスク

都市	ゴキブリ発生リスク
ワシントン D.C.	-0.33929
モスクワ	1.74012
ベルリン	0.12651
ソウル	1.60007
キャンベラ	-0.40518

リが少ないと推測できる。しかし本来ゴキブリは熱帯雨林を中心に高温多湿な環境を好むため、表 4 のモスクワの年間平均気温からするとゴキブリは生息していない、または生息数が少ないはずでありゴキブリ発生リスク 1.74012 という数字は高すぎると考えられる。表 5 の年平均気温の標準偏回帰係数が他に比べて極端に低いことから、今回作成したモデルにおいて、年平均気温がゴキブリ発生リスクに与える影響が低くなってしまっていることがわかる。これは、学習にゴキブリが全国的に分布している日本の都道府県しか使用していないためだと考えられる。今回、ゴキブリの生息数に関するデータが日本国内のものしか見つけることができなかったが、もし外国のデータがあれば、学習

表3 ゴキブリ発生リスク上位下位2都道府県

都道府県	ゴキブリ発生リスク
東京都	4.46522
大阪府	2.27283
長野県	-0.68619
北海道	-0.76841

表4 世界の主要な都市の説明変数

都市	人口(万人)	面積(km ²)
ワシントン D.C	64.75	177.0
モスクワ	1192.00	2511.0
ベルリン	347.00	891.8
ソウル	986.00	605.2
キャンベラ	39.58	814.2
都市	ごみ排出量(t/km ²)	年平均気温(°C)
ワシントン D.C	23.29	13.14
モスクワ	4.71	-9.35
ベルリン	139.37	9.97
ソウル	179.31	13.58
キャンベラ	1.75	13.55
都市	年平均湿度(%)	
ワシントン D.C	68.1	
モスクワ	80.8	
ベルリン	75.5	
ソウル	69.0	
キャンベラ	67.5	

に世界のさまざまな気候のデータを使用することで、更に正確な推定ができるだろう。

参考文献

- [1] ゴキブリ駆除マイスター, 都道府県別ゴキブリ発生リスク<図解> 2015 年版, <https://t-meister.jp/gokiburi/lab/highrisk-area2015.html>, (Accessed: 2021-9-15)
- [2] 総務省統計局国勢調査, 第2章 人口・世帯, <https://www.stat.go.jp/data/nihon/02.html>, (Accessed: 2021-9-15)
- [3] 環境省, 廃棄物処理技術情報, http://www.env.go.jp/recycle/waste_tech/ippan/h27/index.html, (Accessed: 2021-9-15)

表5 作成したモデルの標準偏回帰係数

説明変数	標準偏回帰係数
人口	0.53857
面積	-0.20174
ごみ排出量	0.32184
年平均気温	-0.02312
年平均湿度	-0.11016

html, (Accessed: 2021-9-15)

- [4] 国土交通省国土地理院, 全国都道府県市区町村別面積調, <https://www.gsi.go.jp/KOKUJYOHO/MENCHO-title.htm>, (Accessed: 2021-9-15)
- [5] 国土交通省気象庁, <https://www.jma.go.jp/jma/index.html>, (Accessed: 2021-9-15)
- [6] United States Census Bureau, <https://www.census.gov/en.html>, (Accessed: 2021-9-15)
- [7] Eurostat, <https://ec.europa.eu/eurostat>, (Accessed: 2021-9-15)
- [8] United Nations, Statistics Division, <https://unstats.un.org/unsd/demographic-social/products/dyb/>, (Accessed: 2021-9-15)
- [9] OECD, OECD Data, <https://data.oecd.org/waste/municipal-waste.htm>, (Accessed: 2021-9-15)
- [10] Weatherbase, <https://www.weatherbase.com/weather/countryall.php3>, (Accessed: 2021-9-15)