

データサイエンス演習 追加課題

中央大学理工学部
ビジネスデータサイエンス学科
23D7104001I 高木悠人

1. 取り扱う時系列データ

本課題では、東京都が公開している各区の外国人人口推移データを利用した。ただし、期間は昭和 54 年から令和 7 年で記録周期は 1 年である。また、前処理として西暦の変換や転置を実施した。そして今回、解析したファイルや進め方等は GitHub でまとめ、付録の章に URL で示すこととした。

2. 解析方法

本課題は、Python3.11.13 を利用し仮想環境で実施した。具体的なフローについては GitHub に示すこととした。ただし、利用する解析手法は講義で扱った内容である。

3. 全体像の可視化

はじめに、全体を折線グラフで可視化した。出力したものを図 1 に示す。

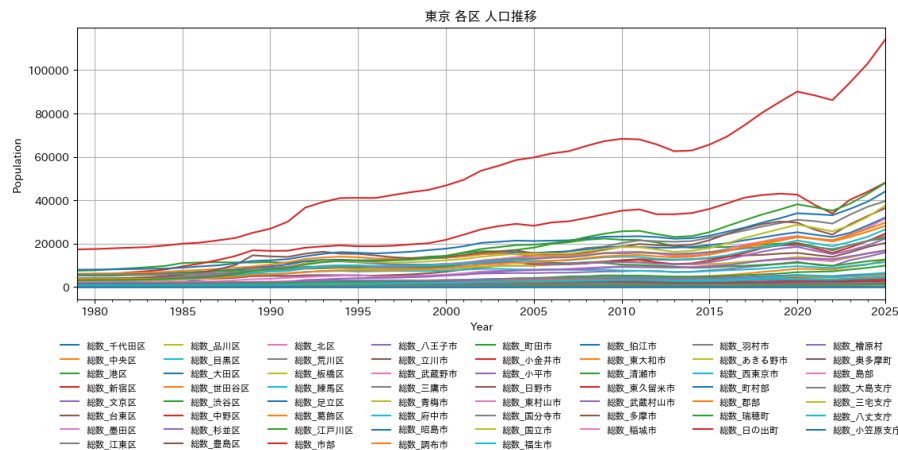


図 1. 東京各区 外国人人口推移

図 1 を見ると、各区で外国人人口が増加傾向にあることがわかる。特に、総数_市部を除くと新宿区の人口が多いことがわかる。人口増加推移を観察するため、新宿に絞りプロットした結果を図 2 に示す。

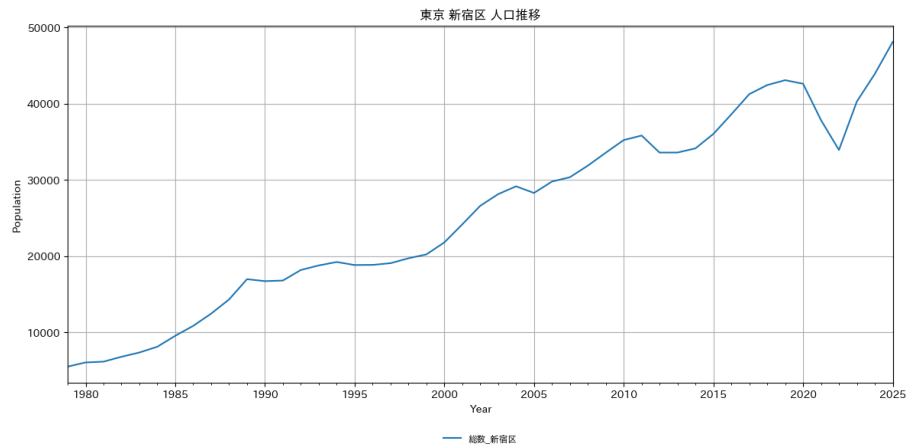


図 2. 新宿区 外国人人口推移

上記を見ると、2022 年に局所的に下に凸となっていることから新型コロナウイルス感染症拡大の影響を大きく受けていると推測できる。それ以外は、大きな周期性は見られず右肩上がりのトレンドがあると見える。

4. 一回差分の可視化

次に、一回差分の可視化を行う。これは、増減量における季節性やトレンドを確認するためである。まずは、全市区町村で全体像を把握するため、出力する。

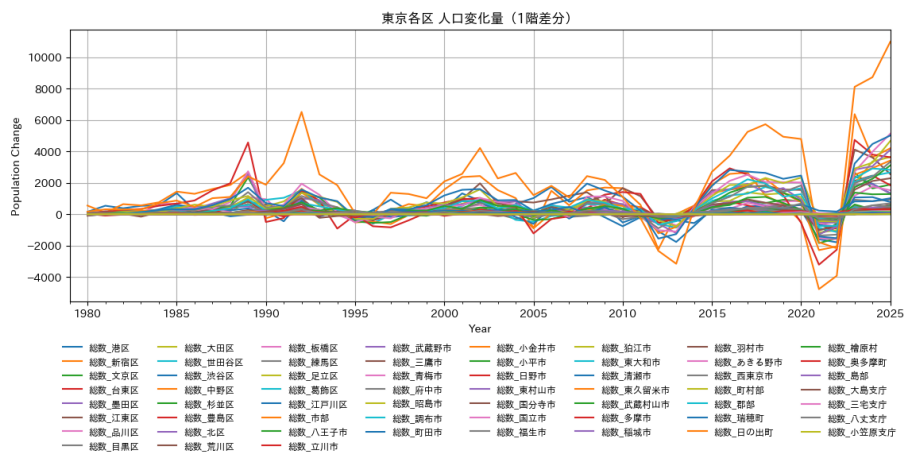


図 3. 東京都全市区町村における外国人人口の一回差分

上記を見ると、全体的に正で推移していることから上昇傾向であると言える。そして、ほとんどの市区町村で同様の推移をしていることが見受けられる。具体的には、2001 年前後や 2017 年前後ではほとんど正であるため外国人人口が流入または国内で出生していることが予想できる。そのため、何らかの外国人流入政策等が実施された

可能性がある」と推測できる。次に、最も外国人人口が多かった新宿に絞り、一回差分を出力した結果を図4に示す。

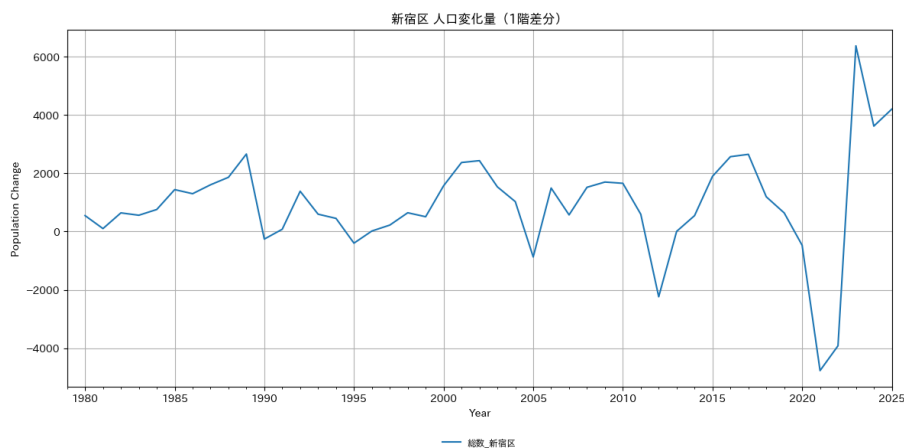


図4. 新宿区の外国人人口の一回差分

上記のグラフを見ると、目立った傾向がなく平均・分散が均一であると推測できるため定常であると予測できる。ただし定常性に関する検定として ADF と Ljung-Box で実施した結果については自明であると判断し、notebook に示すこととした。

5. 自己相関関数と偏自己相関関数

新宿区の外国人人口推移における自己相関関数と偏自己相関関数の出力を以下に示す。

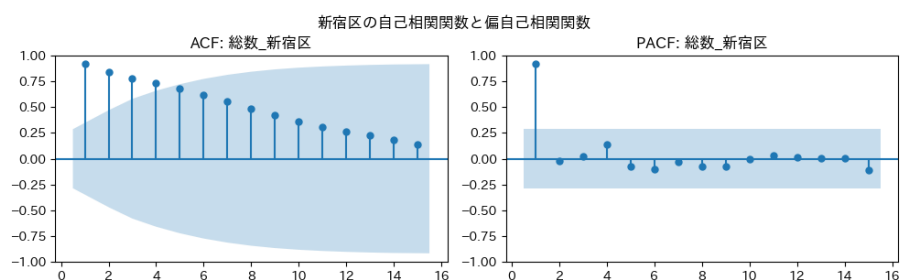


図5. 新宿区の自己相関関数と偏自己相関関数

まず、自己相関関数では、ラグが進むに従って穏やかに減衰する特徴が視認できる。つまり、過去の人口推移が将来値に継続的な影響を与えていることを示しており、穏やかな上昇トレンドを有する今回のような非定常性を反映している。

一方、偏自己相関関数では、ラグ1で正の相関が見られそれ以降急減している。つまり、周期性、季節性は見受けられず直前のデータに強い依存が見られると評価できる。

要するに、新宿区の外国人人口推移は、前年の水準を基準として徐々に変化していく傾向が強いと言える。

次に、地方的地域である三宅支庁における自己相関関数と偏自己相関関数を以下に示す。

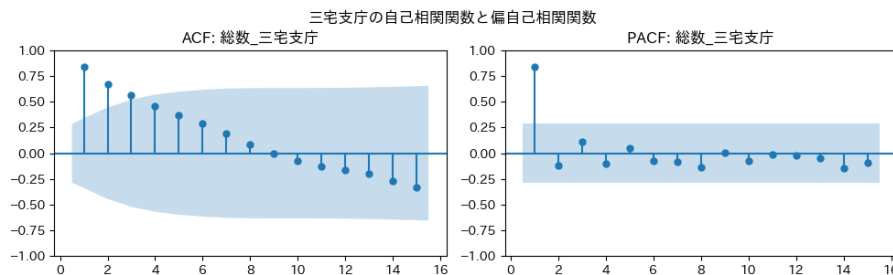


図 6. 三宅支庁の自己相関関数と偏自己相関関数

上記を見ると、形状は新宿区の推移と類似しているが ACF 及び自己相関関数において負に振れていることがわかる。これは、三宅支庁のような人口規模の小さい地域において、一時的な増減が翌年に反転し元の水準へと戻る動きが強く現れるためであると考え。具体的には、地方的地域において年間の外国人移住者数が一時的に増加した場合でもその変化が継続的な増加につながりにくい傾向がある。この傾向は、都市部のように就業機会や居住環境、教育機関等が集積している地域では、流入が安定しやすく、起こりにくい。しかし、地方的地域では雇用や生活インフラが限定的であるため、一時的な流入、政策的影響が人口変動に直接影響されやすい地域特性に起因すると考える。よって、三宅支庁では外生的ショックや短期的なイベントによる人口変動が強く現れ、これが ACF の負方向への振れとして可視化されたと考える。

以上より、都市部と地方的地域では外国人人口推移の時間構造が異なり、前者は持続的な増加傾向、後者は短期変動と平均回帰性という異なる特徴を示していると言える。

6. モデリング

次に、時系列モデルの推定を行った。モデル選択には、AIC 最小化に基づく逐次探索 (auto_arima) を用い、ARIMA モデルの候補を比較した。その結果、ARIMA(0,1,1) モデルが最適と選択された (AIC=818.304)。このモデルは 1 階差分を取ることでトレンドを除去し、1 期前の誤差項を用いて現在の変動を説明する構造であり、非定常なランダムウォーク的挙動に対して適切なモデル化を行うものである。推定結果を表 1 に示す。

表 1. ARIMA モデル推定結果

```

Performing stepwise search to minimize aic
ARIMA(2,1,2)(0,0,0)[0] intercept : AIC=inf, Time=0.04 sec
ARIMA(0,1,0)(0,0,0)[0] intercept : AIC=823.022, Time=0.00 sec
ARIMA(1,1,0)(0,0,0)[0] intercept : AIC=819.742, Time=0.00 sec
ARIMA(0,1,1)(0,0,0)[0] intercept : AIC=818.304, Time=0.01 sec
ARIMA(0,1,0)(0,0,0)[0] : AIC=832.069, Time=0.00 sec
ARIMA(1,1,1)(0,0,0)[0] intercept : AIC=820.503, Time=0.01 sec
ARIMA(0,1,2)(0,0,0)[0] intercept : AIC=820.559, Time=0.01 sec
ARIMA(1,1,2)(0,0,0)[0] intercept : AIC=inf, Time=0.02 sec
ARIMA(0,1,1)(0,0,0)[0] : AIC=822.952, Time=0.00 sec

Best model: ARIMA(0,1,1)(0,0,0)[0] intercept
Total fit time: 0.094 seconds

=====
SARIMAX Results
=====
Dep. Variable:          y      No. Observations:          47
Model:                SARIMAX(0, 1, 1)  Log Likelihood        -406.152
Date:                 Fri, 31 Oct 2025  AIC                  818.304
Time:                  00:59:36    BIC                  823.790
Sample:                0      HQIC                  820.359
                        - 47
Covariance Type:      opg
=====
              coef    std err          z      P>|z|      [0.025    0.975]
-----
...
=====

```

MA(1)項は有意 ($p < 0.05$) であり、直前の誤差が現在の変動に影響を与えていることが確認された。また、定数項は 1% 有意水準ではないものの、有意傾向 ($p = 0.059$) が見られ、長期的な平均的増加傾向が存在する可能性が示唆された。一方で、残差の正規性検定および分散不均一性検定の結果より、残差に非正規性および異分散性が確認され、外生要因や政策・社会情勢等の構造変化が存在する可能性が示唆される。さらに、Ljung-Box 検定では残差に自己相関が存在しないことが確認されたため、モデルが時系列依存を十分に捉えていると評価できる。以上の結果から、対象データに対する ARIMA(0,1,1)モデルの適合性が確認された一方、残差の非正規性および異分散性により、将来予測においては政策変更、災害、国際的な移動制限等を考慮する必要があると考える。

7. 参考文献

1. データサイエンス演習，講義資料。
2. 東京都，東京都の統計「外国人人口」，
(<https://www.toukei.metro.tokyo.lg.jp/gaikoku/ga-index.htm>),
2025/10/30 取得。

8. 付録

1. GitHub リポジトリ
(<https://github.com/yut0takagi/ass-ds-exercises>)