

2022統計関連学会連合大会資料

人流メッシュ統計データを使用した 月次宿泊旅行統計調査の推計について

2022年9月

総務省統計改革実行推進室 松井伸司

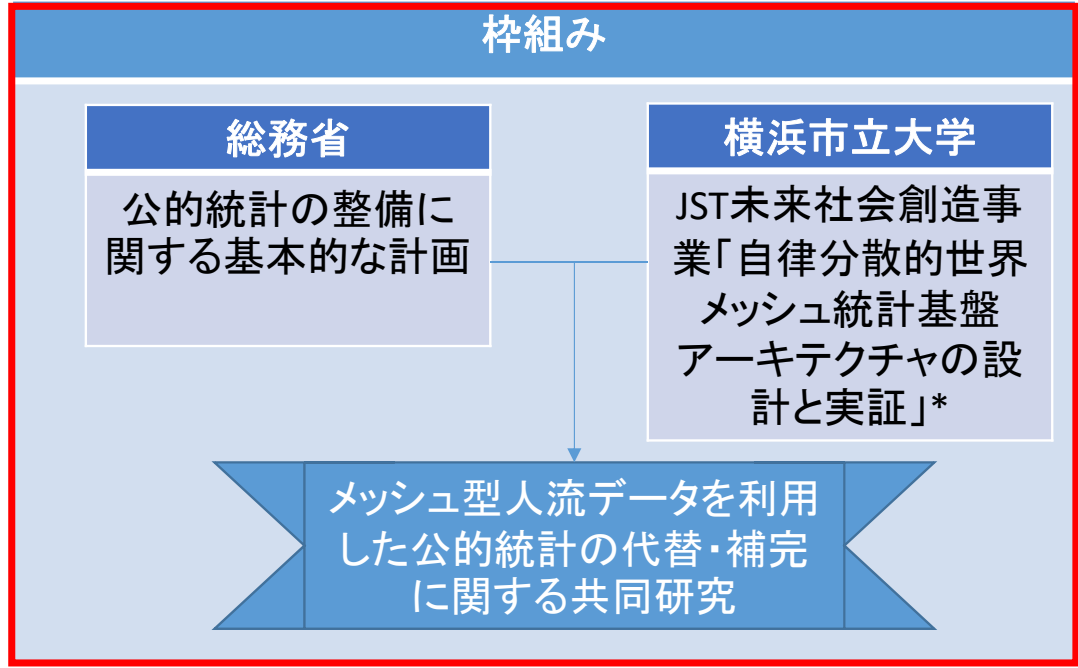
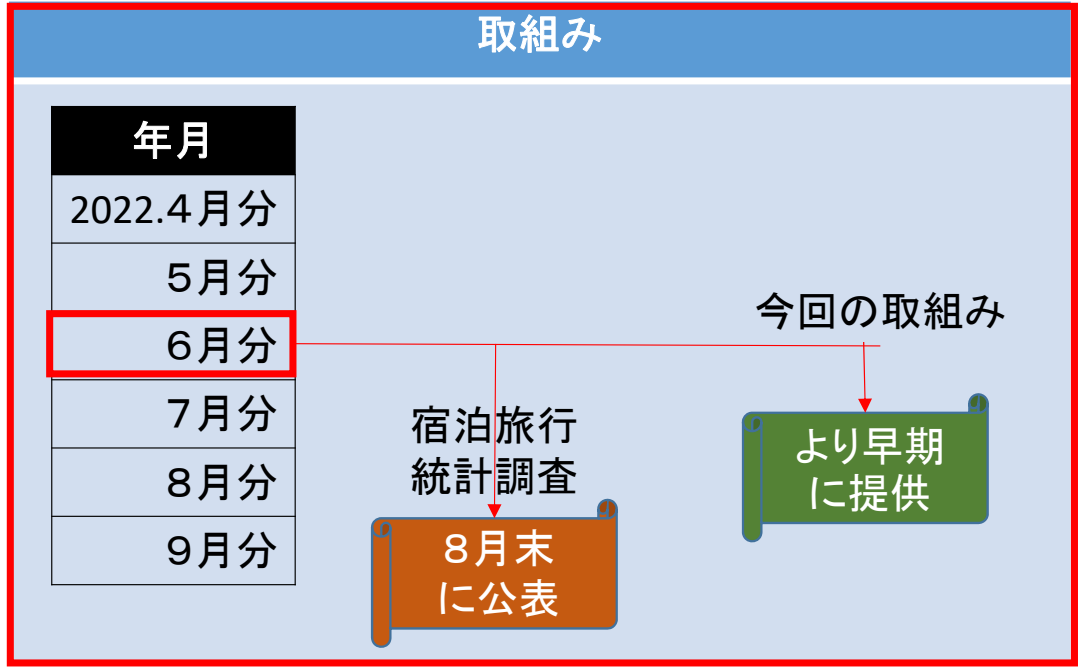
横浜市立大学データサイエンス学部 佐藤彰洋

目次

1. はじめに
2. 概要
3. 推計方法
4. 推計結果
5. まとめ

1. はじめに

- 【目的】ビッグデータ活用による公的統計の代替・補完の可能性の検討
- 【取組】国土交通省観光庁が実施する宿泊旅行統計調査は、都道府県単位の情報は翌々月の末日の公表まで約2ヶ月のタイムラグが発生している。人流データを活用して延べ宿泊者数の推計を行い、オルタナティブデータとして、より早期に統計値を提供する方法を研究
- 【枠組】総務省と横浜市立大学が締結した「メッシュ型人流データを利用した公的統計の代替・補完に関する共同研究」



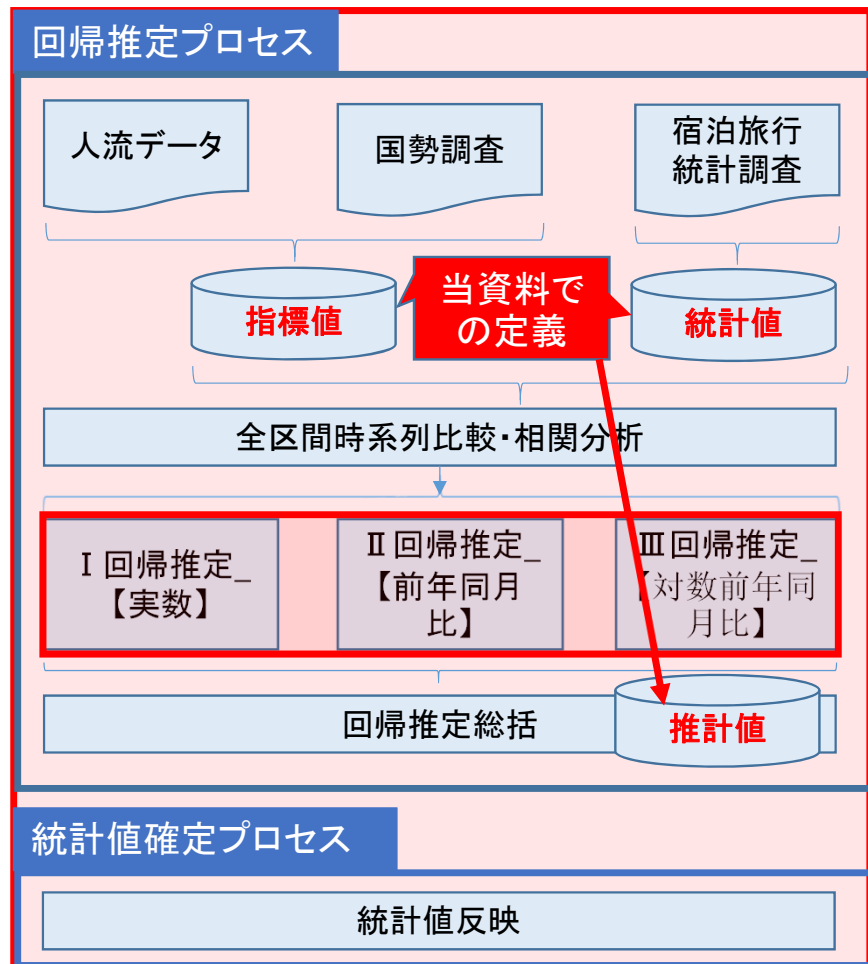
*研究課題番号: 2020 JPMJMI20B6、研究代表者 佐藤彰洋、研究期間: 2020年度～)

2. 概要_全体

- 都道府県単位に人流データを基に宿泊旅行統計調査の日本人延べ宿泊者数（延べ宿泊者数－外国人延べ宿泊者数）を推計
- コロナ禍で安定的な期間が確保できない時でも、有効な推計方法を検討

使用データ		
データ名	利用項目	説明
宿泊旅行統計調査（第2次速報）	（都道府県別） ・ 延べ宿泊者数 ・ 外国人延べ宿泊者数	全国の宿泊施設を対象に毎月調査した結果を復元・集計した国土交通省観光庁が公表する一般統計
人流データ*	（500m4次メッシュコード別・時間帯別） ・ 流動人口（4時台）（月単位、休日／平日）	スマートフォンアプリのGPSデータのユーザー数を、国勢調査市区町村単位の人口を基に拡大推計
国勢調査（2015年）	（500m4次メッシュコード別） ・ 常住人口	全国民を対象に5年に一回、年齢・世帯など人口の基礎的属性を調査した結果を総務省が公表する基幹統計
OSM（オープンストリートマップ）	・ 宿泊施設位置情報	英国OSM財団で運営されるインターネット上で自由に編集・書き込み・利用できる地図

*(株)Agoopの流動人口データ

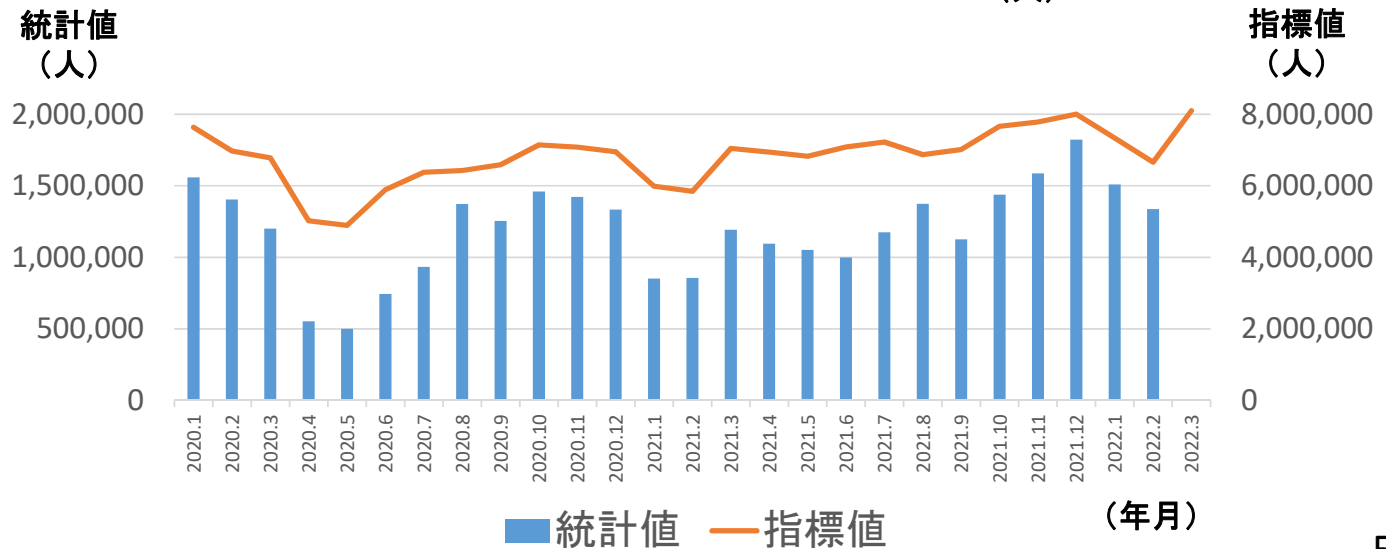
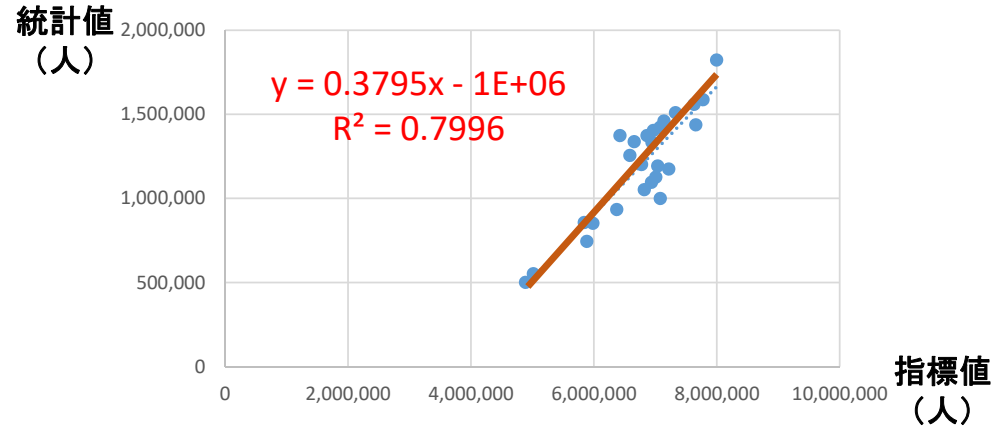


2. 概要__推計イメージ

- 都道府県単位に人流データを基に宿泊旅行統計調査の日本人延べ宿泊者数（延べ宿泊者数－外国人延べ宿泊者数）を推計
- ①時系列の統計値・指標値→②過去データで回帰分析→③未公表の統計値を推計→④統計値を反映

年月	統計値	指標値	推計値	乖離率
2020.1	1,558,660	7,634,559		
2020.2	1,404,080	6,973,832		
2020.3	1,200,760	6,781,331		
2020.4	552,400	5,019,855		
2020.5	500,300	4,892,221		
2020.6	744,440	5,890,392		
2020.7	933,160	6,376,496		
2020.8	1,372,930	6,428,540		
2020.9	1,254,860	6,589,556		
2020.10	1,460,010	7,144,585		
2020.11	1,422,230	7,078,922		
2020.12	1,333,240	6,949,600		
2021.1	851,430	5,988,528		
2021.2	856,340	5,847,008		
2021.3	1,192,520	7,043,353		
2021.4	1,095,390	6,942,426		
2021.5	1,052,060	6,823,242		
2021.6	999,440	7,084,630		
2021.7	1,174,370	7,224,239		
2021.8	1,373,760	6,867,605		
2021.9	1,125,990	7,012,320		
2021.10	1,437,380	7,662,925		
2021.11	1,586,670	7,780,740		
2021.12	1,822,220	8,001,985		
2022.1	1,509,410	7,333,431		
2022.2	1,337,520	6,659,958		
2022.3		8,098,059	1,609,093	

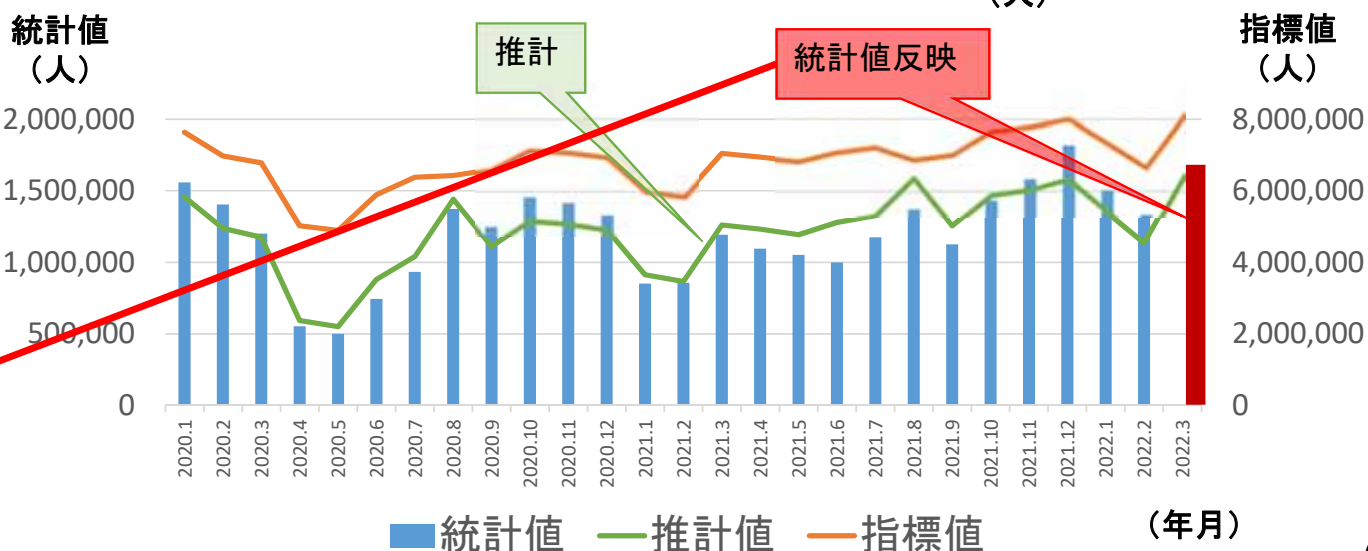
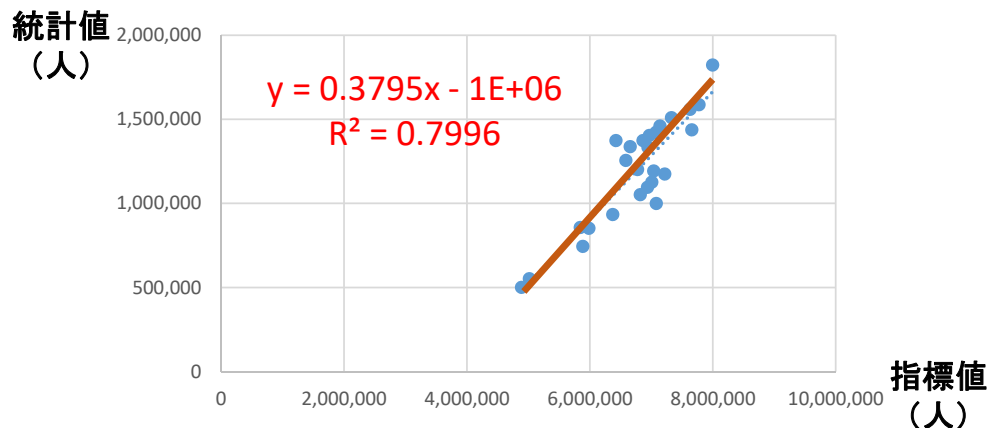
回帰分析対象
期間(ex:2020.1
～2022.2)での
回帰式から翌
月(ex:2022.3)の
統計値を推計



2. 概要__推計イメージ

- 都道府県単位に人流データを基に宿泊旅行統計調査の日本人延べ宿泊者数（延べ宿泊者数－外国人延べ宿泊者数）を推計
- ①時系列の統計値・指標値→②過去データで回帰分析→③未公表の統計値を推計→④統計値を反映

年月	統計値	指標値	推計値	乖離率
2020.1	1,558,660	7,634,559	1,456,057	-6.6%
2020.2	1,404,080	6,973,832	1,237,903	-11.8%
2020.3	1,200,760	6,781,331	1,174,344	-2.2%
2020.4	552,400	5,019,855	592,752	7.3%
2020.5	500,300	4,892,221	550,610	10.1%
2020.6	744,440	5,890,392	880,180	18.2%
2020.7	933,160	6,376,496	1,040,678	11.5%
2020.8	1,372,930	6,428,540	1,440,886	4.9%
2020.9	1,254,860	6,589,556	1,111,025	-11.5%
2020.10	1,460,010	7,144,585	1,294,281	-11.4%
2020.11	1,422,230	7,078,922	1,272,601	-10.5%
2020.12	1,333,240	6,949,600	1,229,902	-7.8%
2021.1	851,430	5,988,528	912,582	7.2%
2021.2	856,340	5,847,008	865,855	1.1%
2021.3	1,192,520	7,043,353	1,260,857	5.7%
2021.4	1,095,390	6,942,426	1,231,636	12.4%
2021.5	1,052,060	6,823,242	1,192,578	13.4%
2021.6	999,440	7,084,630	1,279,376	28.0%
2021.7	1,174,370	7,224,239	1,325,843	12.9%
2021.8	1,373,760	6,867,605	1,591,391	15.8%
2021.9	1,125,990	7,012,320	1,255,989	11.5%
2021.10	1,437,380	7,662,925	1,471,400	2.4%
2021.11	1,586,670	7,780,740	1,510,269	-1.8%
2021.12	1,822,220	8,001,985	1,582,906	-13.1%
2022.1	1,509,410	7,333,431	1,356,633	-10.1%
2022.2	1,337,520	6,659,958	1,134,270	-15.2%
2022.3	1,639,050	8,098,059	1,609,093	-1.8%



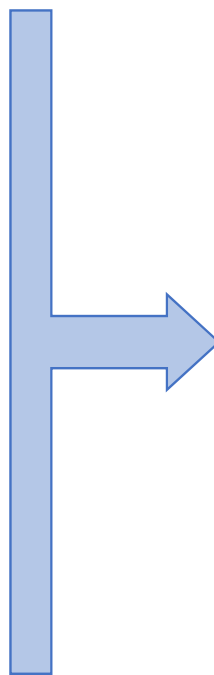
3 . 推計方法_推計ポイント

- 1. 指標値設定 : 人流データ、国勢調査データ
- 2. 回帰分析パターン(14)設定
 - a. 推計方式 : 実数、前年同月比、対数前年同月比
 - b. 分析区間設定 : 固定区間、自動区間（自動区分線形回帰）
- 3. 可視化のための指標の設定
 - a. 推計時の信頼性 : 6ヶ月二乗平均平方根誤差率（RMSPE）
 - b. 推計精度（結果） : 乖離率

当資料での位置づけ
 ・6ヶ月RMSPE→推計時の推計値の信頼性
 ・乖離率 →結果としての推計精度

回帰分析パターン				
	パターン		条件	
I 固定①	実数	固定	期始 2017.1	
I 固定②			期始 2019.10	
I 固定③			期始 2020.04	
I 自動①		自動 区分 線形 回帰	最低期間6ヶ月	
I 自動②			最低期間12ヶ月	
I 自動③			最低期間18ヶ月	
II 固定①	前年 同月 比	固定	期始 2018.1	
II 自動①			自動 区分 線形 回帰	最低期間6ヶ月
II 自動②				最低期間12ヶ月
II 自動③		最低期間18ヶ月		
III 固定①		対数 前年 同月 比	固定	期始 2018.1
III 自動①				自動 区分 線形 回帰
III 自動②	最低期間12ヶ月			
III 自動③	最低期間18ヶ月			

分析設定項目	
カテゴリ	設定数値*
分析区間	From、To、月数
回帰分析結果	決定係数(相関係数)、切片、傾き
回帰推計結果	推計値、6ヶ月RMSPE(RMSE)
統計値確定	統計値、乖離率



- パターン毎に推計値を算出し、信頼性の高いパターンを選定
- 統計値を反映・乖離率を把握し、推計の評価

3 . 推計方法_指標値作成

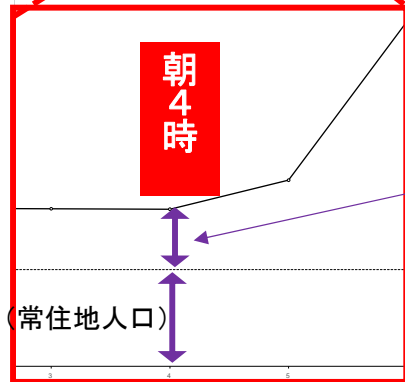
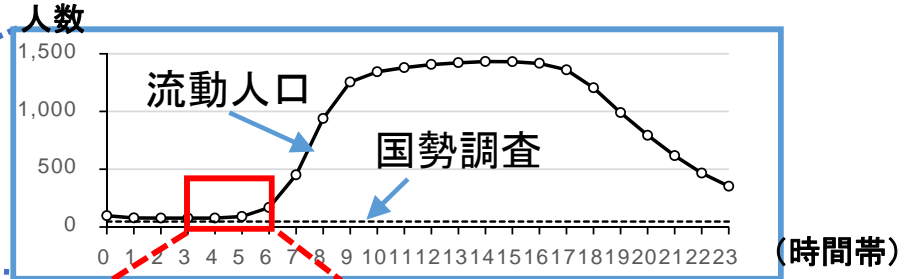
都道府県単位に宿泊施設が存在する500mメッシュを特定し、朝4:00の流動人口と国勢調査から指標値を算出する（休日・平日の日数を重付け）

$$\text{指標値} = \sum_{i \in I} (4\text{時流動人口}(i) - \text{常住地人口}(i))$$

(I : 宿泊施設が存在する500mメッシュコードの集合)



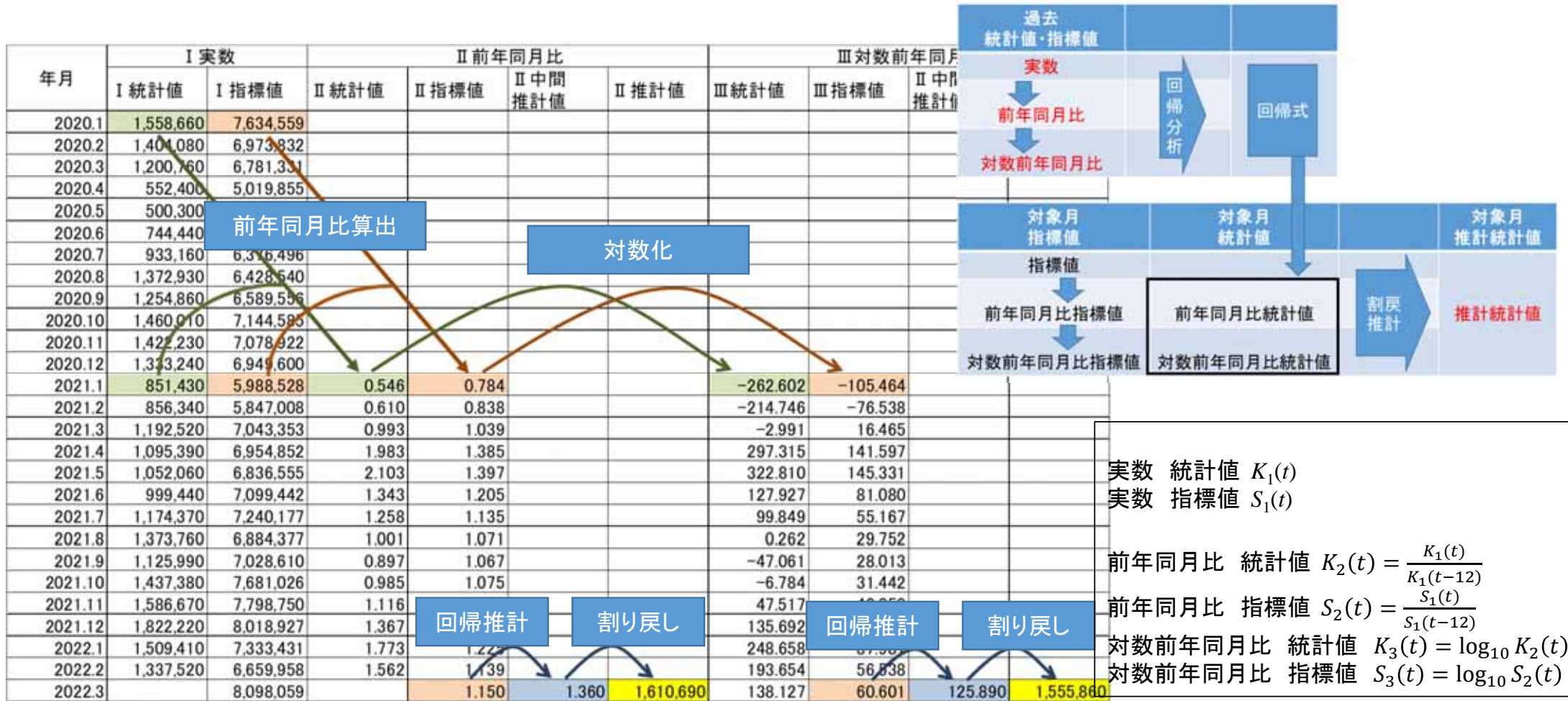
(国勢調査と人流データ)



朝4時の流動人口は主として、居住者と宿泊旅行者で構成されると考える

3 . 推計方法_推計方式

季節性を排除するため統計値・指標値それぞれの前年同月比・対数前年同月比を算出し、回帰推計と実数への割戻しを実施



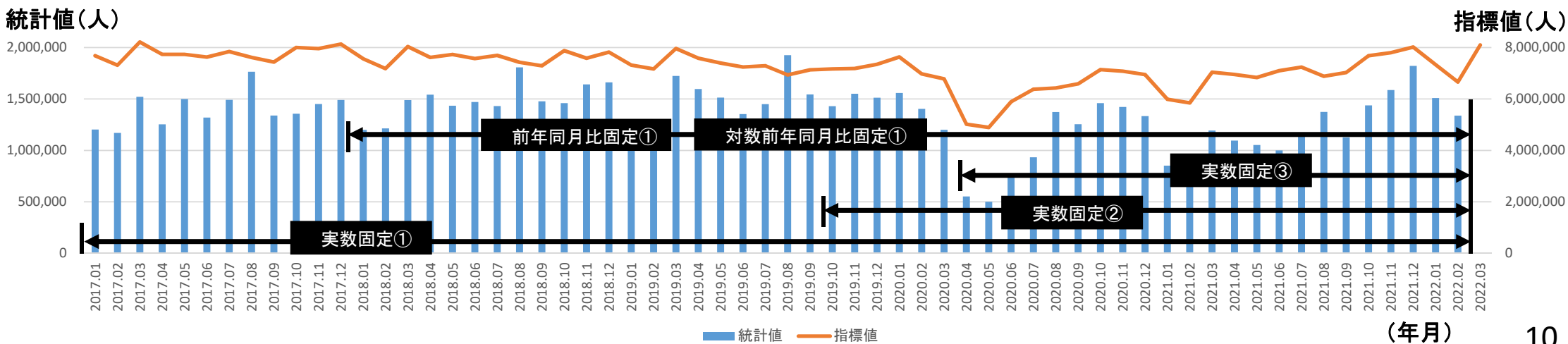
3 . 推計方法_分析区間設定 (固定区間)

社会環境・データのトレンド変更時にも、適切な分析区間の設定が必要

- 固定区間 : 社会環境・データ特性のトレンド変更を外的要因からみて区間を設定
- 自動区間 (自動区分線形回帰) : 有意なトレンド変化を尤度比検定 (有意水準0.05) で検出し、直近区間を設定

固定区間設定

パターン		期始	区間設定理由
実数	固定①	2017.1	人流データの基となるスマホアプリユーザー数が 一定数(100万人)超過 (データの信頼性確保)
	固定②	2019.10	スマホアプリの 大規模な入替え (データ特性変化によるデータ断層への対応)
	固定③	2020.4	コロナ禍の拡大による 宿泊者数の大幅減 (社会環境変化への対応)
前年同月比	固定①	2018.1	2017.1にアプリユーザー数が一定数(100万人)超過(データの信頼性確保)
対数前年同月比	固定①	2018.1	2017.1にアプリユーザー数が一定数(100万人)超過(データの信頼性確保)



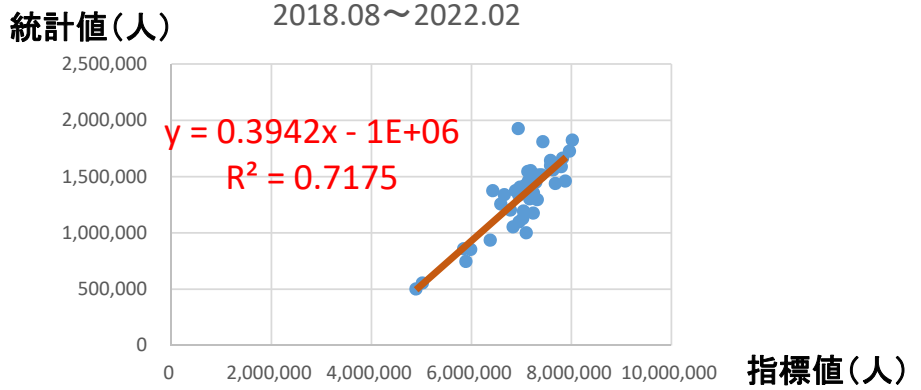
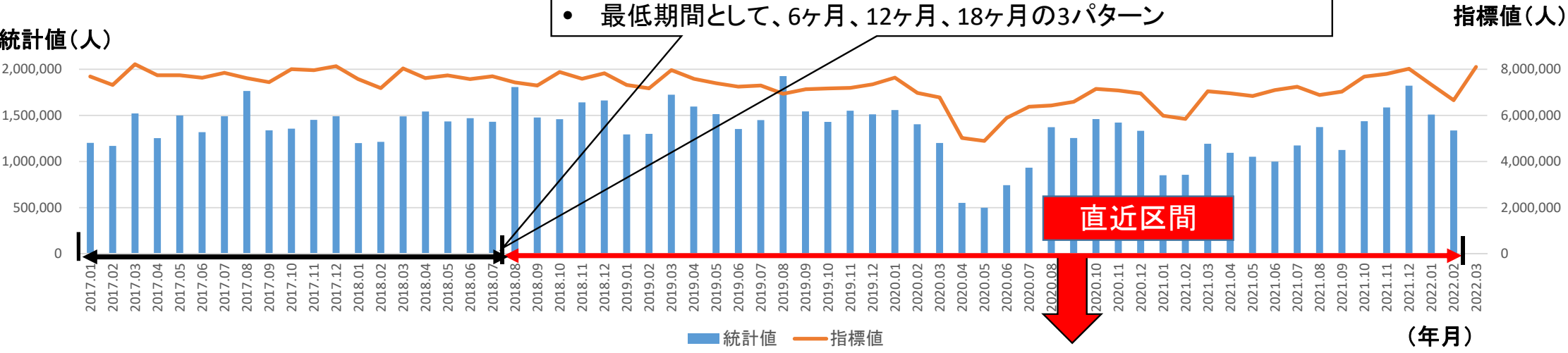
3 . 推計方法_分析区間設定 (自動区分線形回帰)

社会環境・データ特性のトレンド変更時にも、適切な分析区間の設定が必要

- 固定区間：社会環境・データ特性のトレンド変更を外的要因からみて設定
- 自動区間：自動区分線形回帰により有意なトレンド変化を尤度比検定 (有意水準0.05) で検出

自動区分線形回帰

- 有意水準0.05で尤度比検定
- 最低期間として、6ヶ月、12ヶ月、18ヶ月の3パターン



3 . 推計方法_分析区間設定 (自動区分線形回帰) 参考

尤度比検定

誤差は正規分布に従うとする

$$p(\varepsilon; \mu, \sigma^2) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(\varepsilon - \mu)^2}{2\sigma^2}\right)$$

分割あり

左側	右側
$y_t = a_L + b_L x_t + \varepsilon_t$	$y_t = a_R + b_R x_t + \varepsilon_t$

分割なし

$$y_t = a + b x_t + \varepsilon_t$$

対数尤度差の2倍

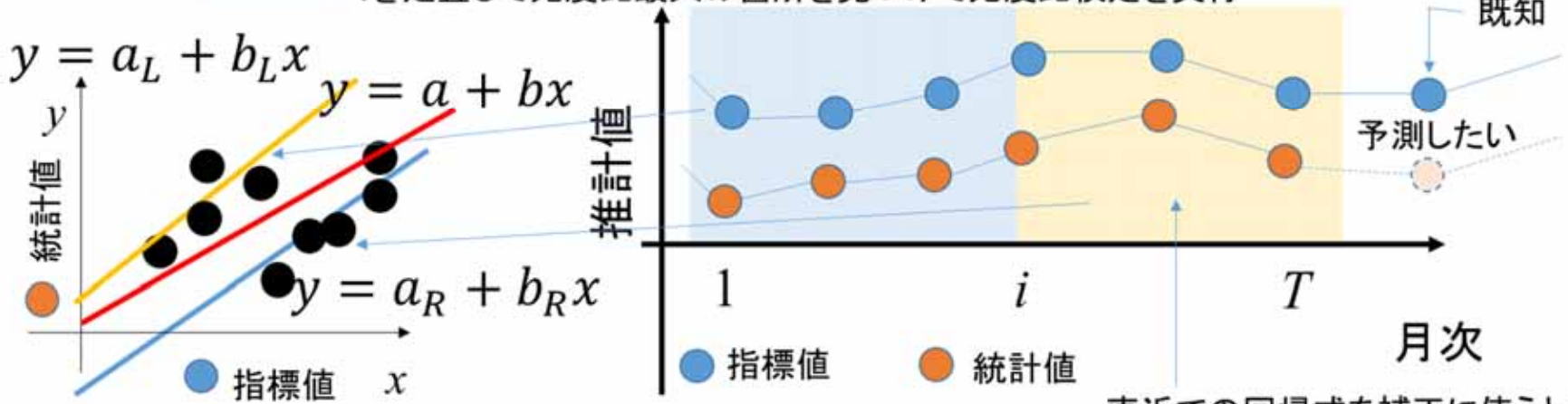
$$D(i) = 2 \left(\sum_{t=1}^i \ln p(\varepsilon_t; \mu_L, \sigma_L^2) + \sum_{t=i+1}^T \ln p(\varepsilon_t; \mu_R, \sigma_R^2) - \sum_{t=1}^T \ln p(\varepsilon_t; \mu, \sigma^2) \right) \approx -i \ln \sigma_L^2 - (T-i) \ln \sigma_R^2 + T \ln \sigma^2$$

p値

$$p = \exp\left(-\frac{\max_i D(i)}{2}\right)$$

(尤度比検定のp値はモデルパラメータ数の差(2=4-2)を自由度とするカイ二乗分布)

iを走査して尤度比最大の箇所を見つけて尤度比検定を実行

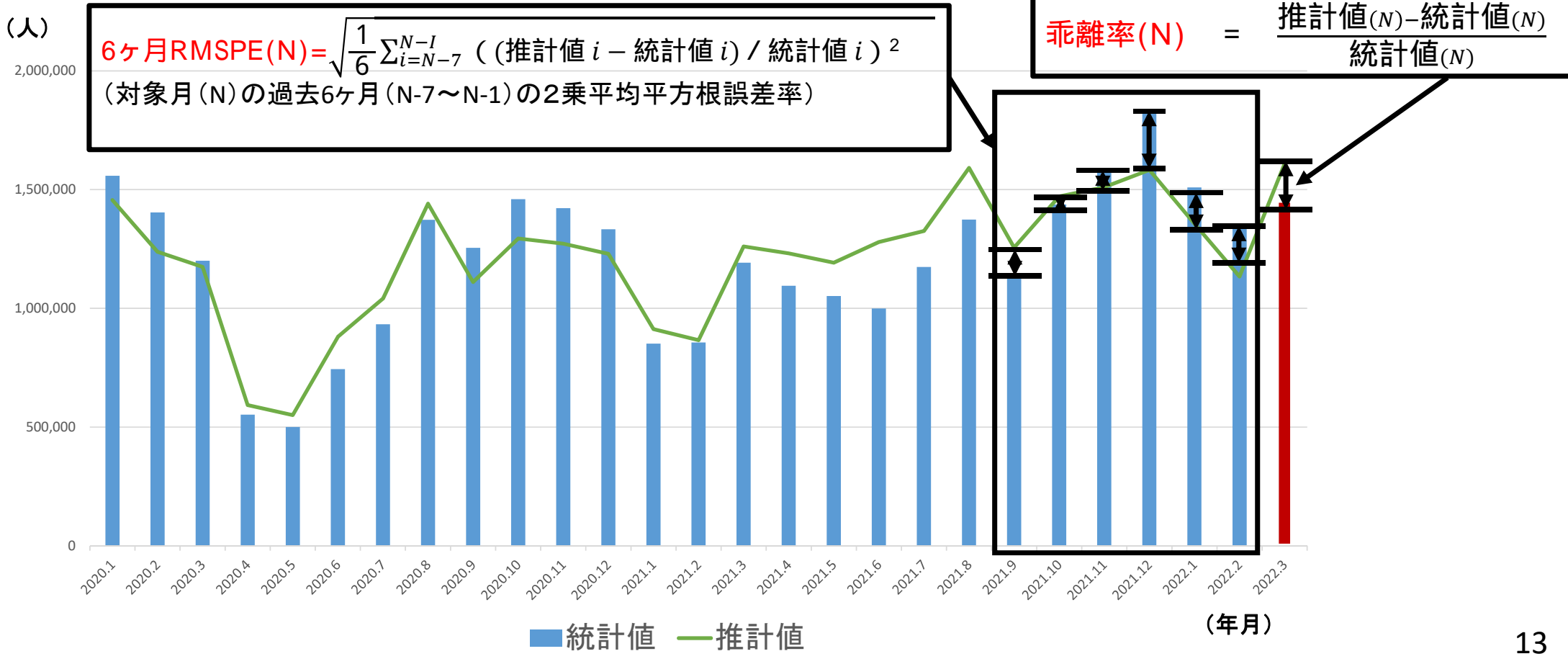


直近での回帰式を補正に使うとよい

3 . 推計方法_RMSPE・乖離率

推計状況を可視化するための指標の設定

- 推計時の信頼性 : 6ヶ月二乗平均平方根誤差率 (RMSPE) : 統計値と推計値の差の割合を二乗平均して平方根
- 推計精度 (結果) : 乖離率 : 推計値と統計値の差の割合



4 . 推計結果_2022.3京都府推計時

- 【推計時】
- 信頼性が高い候補として、6ヶ月RMSPE0.1未満が7パターン
 - 6ヶ月RMSPEが最小のものを方式毎に候補選定（Ⅰ固定②、Ⅱ自動①、Ⅲ自動①）

6ヶ月RMSPE

パターン	期始 From	6ヶ月 RMSPE	6ヶ月RMSPE						
			0	0.05	0.1	0.15	0.2	0.25	0.3
Ⅰ実数－固定①	2017.01-	0.084							
Ⅰ実数－固定②	2019.10-	0.076							
Ⅰ実数－固定③	2020.04-	0.080							
Ⅰ実数－自動①	2020.08-	0.078							
Ⅰ実数－自動②	2017.01-	0.084							
Ⅰ実数－自動③	2017.01-	0.084							
Ⅱ前年同月比－固定①	2018.01-	0.148							
Ⅱ前年同月比－自動①	2021.07-	0.101							
Ⅱ前年同月比－自動②	2021.03-	0.174							
Ⅱ前年同月比－自動③	2020.01-	0.271							
Ⅲ対数前年同月比－固定①	2018.01-	0.111							
Ⅲ対数前年同月比－自動①	2021.07-	0.083							
Ⅲ対数前年同月比－自動②	2021.02-	0.131							
Ⅲ対数前年同月比－自動③	2020.01-	0.206							

4 . 推計結果_2022.3京都府統計確定時

京都府2022年3月

【統計確定時】

- 候補として選定した I 固定②の乖離率10%程度、II 自動①、III 自動①については10%以下
- 候補として選定したものの乖離率が最小となるわけではないが、全体として一定レベルに収まっている

パターン	期始 From	6ヶ月 RMSPE	6ヶ月RMSPE							乖離率	乖離率(%)			
			0	0.05	0.1	0.15	0.2	0.25	0.3		0	5	10	15
I 実数-固定①	2017.01-	0.084								-11.7%				
I 実数-固定②	2019.10-	0.076								-10.7%				
I 実数-固定③	2020.04-	0.080								-4.0%				
I 実数-自動①	2020.08-	0.078								-2.1%				
I 実数-自動②	2017.01-	0.084								-11.7%				
I 実数-自動③	2017.01-	0.084								-11.7%				
II 前年同月比-固定①	2018.01-	0.148								1.1%				
II 前年同月比-自動①	2021.07-	0.101								-4.4%				
II 前年同月比-自動②	2021.03-	0.174								1.6%				
II 前年同月比-自動③	2020.01-	0.271								13.4%				
III 対数前年同月比-固定①	2018.01-	0.111								-2.9%				
III 対数前年同月比-自動①	2021.07-	0.083								-6.8%				
III 対数前年同月比-自動②	2021.02-	0.131								-0.4%				
III 対数前年同月比-自動③	2020.01-	0.206								7.4%				

(データ提供元:(株)Agoop)

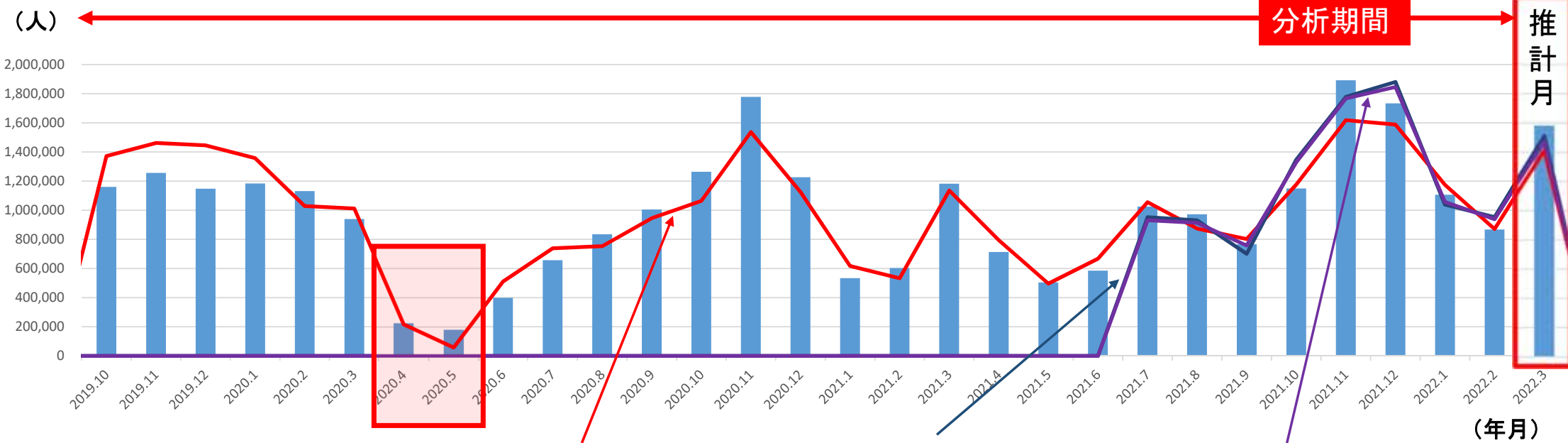
4 . 推計結果_2022.3京都府グラフ

京都府2022年3月

【連動性】

- 推計値は統計値に対して一定の連動性を持っており、コロナ禍の宿泊者数の激減にも良く連動している

(データ提供元: (株)Agoop)



統計値	固定②	自動①	自動①
RMSPE	0.076	0.101	0.083
決定係数	0.897	0.919	0.928
乖離率	-10.7%	-4.4%	-6.8%

4 . 推計結果_2022.3全国展開

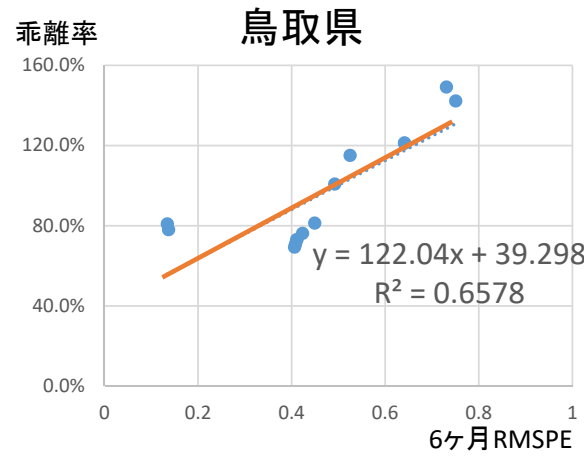
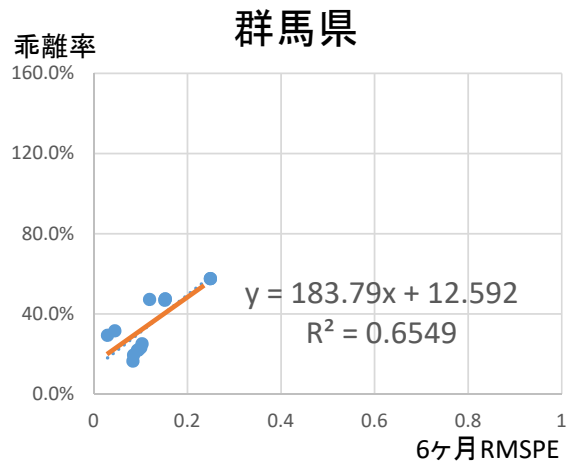
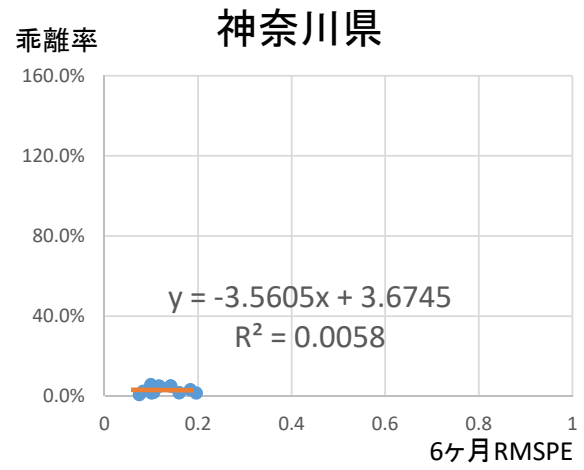
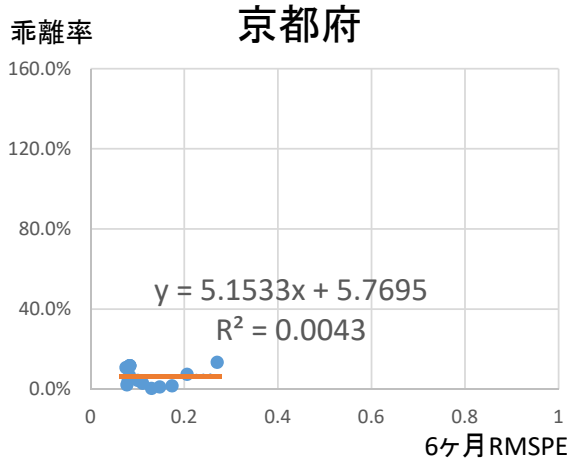
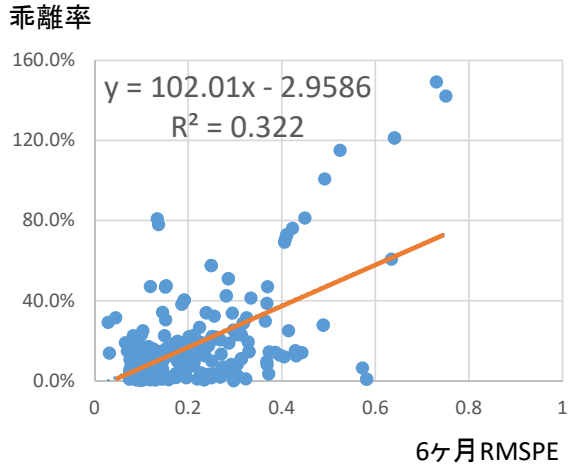
19都道府県2022年3月

19都道府県に対して2022年3月単月で回帰推計を実施し、6ヶ月RMSPEと乖離率の関連性を分析→6か月RMSPEによる選定の合理性

- 全体として6ヶ月RMSPEと乖離率は一定の強さの相関がある
- 京都府・神奈川県では全てのサンプルが乖離率・6ヶ月RMSPEとも小さい数値に収斂
- 群馬県・鳥取県など乖離率が大きくバラついている県では強めの相関を確認

(データ提供元:(株)Agoop)

19都道府県(サンプル数266)



4. 推計結果_2022.3全国展開

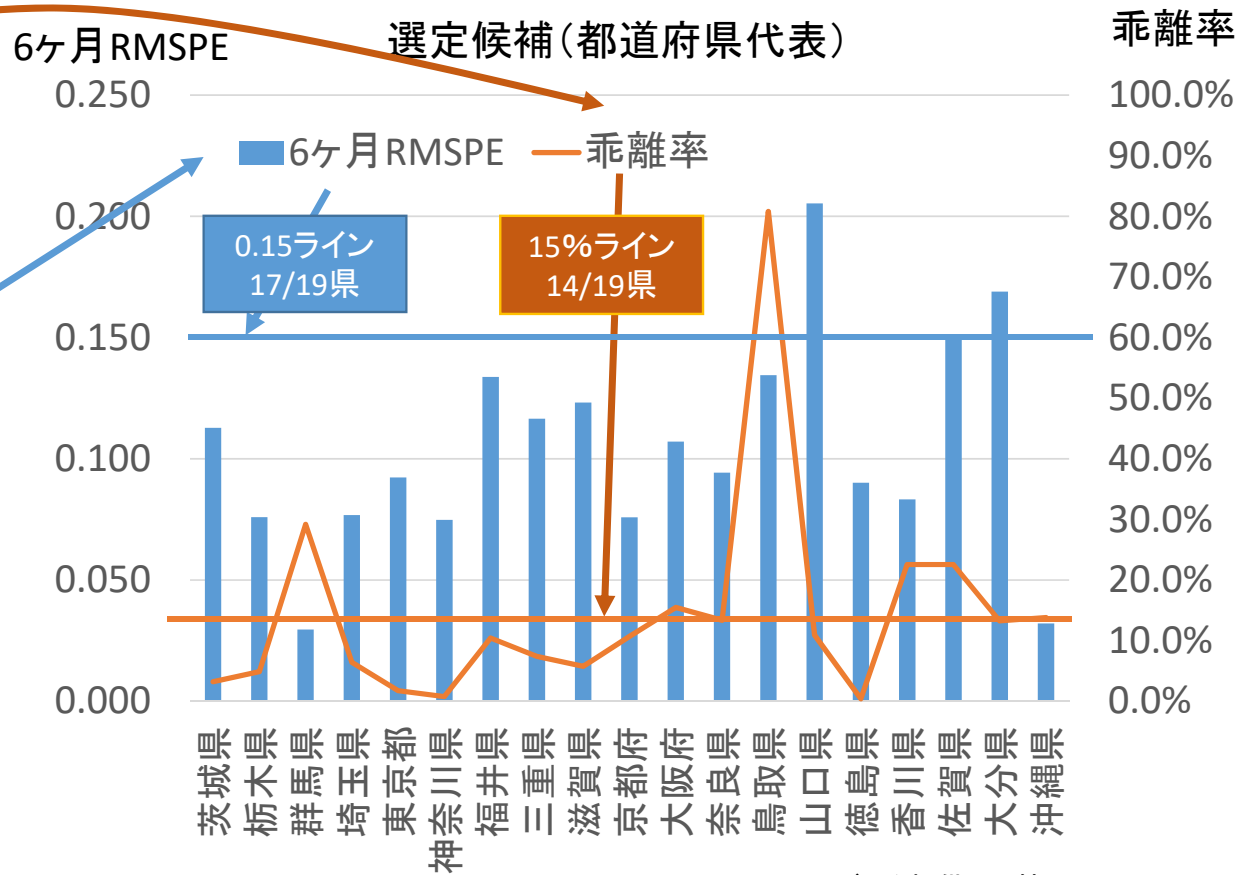
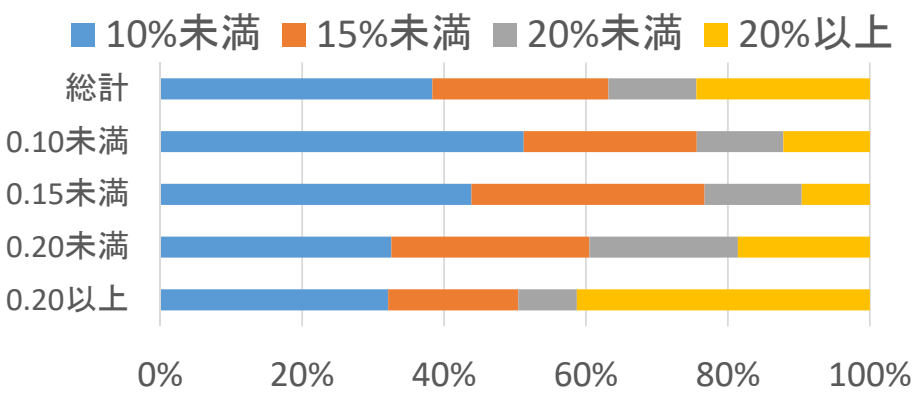
19都道府県2022年3月

都道府県を層別して俯瞰できるように6ヶ月RMSPEと乖離率の目標値（閾値）設定に向けた分析

- 6ヶ月RMSPEが0.15未満のパターンは全体の43%（114/266）→17都道府県をカバー（0.15未満のパターンが存在）
- 6ヶ月RMSPEが0.15未満のパターンにおいて、乖離率が15%未満のものが76%（87/114）

ランク別件数(RMSPE/乖離率)

	RMSPE	乖離率				総計	
		10%未満	15%未満	20%未満	20%以上		
R	0.10未満	21	10	5	5	41	15%
M	0.15未満	32	24	10	7	73	28%
S	0.20未満	14	12	9	8	43	16%
P	0.20以上	35	20	9	45	109	41%
E	総計	102	66	33	65	266	100%



5 . まとめ

まとめと今後の課題

- 推計方法全体としては有効に機能していると考える
 - ✓ 推計方式 4時人流ー常住人口、(対数)前年同月比、自動区分線形回帰
 - ✓ 評価方法 6ヶ月RMSPEによる選定
- 6ヶ月RMSPEと乖離率の目標等についてはユーザーサイドの視点も含めて継続検証が必要
 - ✓ 6ヶ月RMSPEの閾値 0.15未満
 - ✓ 乖離率目標 15%未満
- 精度向上に向けて以下の分析・検討が必要
 - ✓ 指標値と統計値の乖離要因の分析(ex:人流データ、調査統計、OSM宿泊施設)
 - ✓ 地域による推計信頼性や精度に差がある要因に関する分析(ex:人口、人口密度、宿泊者数)
 - ✓ 推計方式の高度化の検討(ex:過去の推計結果を学習し信頼性に反映)