

## 参議院常任委員会調査室・特別調査室

<b>論題</b>	統計分析の手法により正反対の結果等が生じる例 ～月別の延べ宿泊者数と新型コロナ感染者数の関係から～
<b>著者 / 所属</b>	前田 泰伸 / 調査情報担当室
<b>雑誌名 / ISSN</b>	経済のプリズム / 1882-062X
<b>編集・発行</b>	参議院事務局 企画調整室（調査情報担当室）
<b>通号</b>	199号
<b>刊行日</b>	2021-5-20
<b>頁</b>	47-58
<b>URL</b>	<a href="https://www.sangiin.go.jp/japanese/annai/chousa/keizai_prism/backnumber/r03pdf/202119903.pdf">https://www.sangiin.go.jp/japanese/annai/chousa/keizai_prism/backnumber/r03pdf/202119903.pdf</a>

※ 本文中の意見にわたる部分は、執筆者個人の見解です。

※ 本稿を転載する場合には、事前に参議院事務局企画調整室までご連絡ください（TEL 03-3581-3111（内線 75044） / 03-5521-7683（直通））。

# 統計分析の手法により正反対の結果等が生じる例 ～月別の延べ宿泊者数と新型コロナ感染者数の関係から～

調査情報担当室 前田 泰伸

## 《要旨》

本稿では、統計分析の手法やデータの期間の違いにより、場合によっては正反対の結論を導くこともあり得ることについて、観光庁「宿泊旅行統計調査」の延べ宿泊者数と新型コロナウイルス感染症の感染者数（いずれも、月別・都道府県別）の分析から紹介する。

2020年10月から2021年1月までの各月の両者の関係についてクロスセクションデータの回帰分析を行うと、両者は順相関の関係にあり、秋から冬にかけては月効果によって感染者数が増加する。しかし、同じデータをパネルデータとして固定効果モデルで分析を行うと、係数は有意なマイナスの値となり、両者は逆相関の関係となってしまう。また、固定効果モデルでも、データの期間を2020年4月から9月までとすると両者は順相関の関係となり、さらに、固定効果モデルに月効果（時間固定効果）を加えた分析では、両者の関係に有意性はなくなってしまうというように、それぞれ異なった結果となる。統計分析の際には、一つの分析により何らかの結果が得られたとしても、その結果を絶対視するのではなく、他の可能性についても検討を行い、可能であれば別の分析方法を試してみるといったことが重要かと思われる。

## 1. はじめに<sup>1</sup>

本稿では、「統計分析の手法により正反対の結果等が生じる例」と題し、観光庁「宿泊旅行統計調査」における延べ宿泊者数と新型コロナウイルス感染症（以下、「新型コロナ」という）の感染者数（いずれも、月別・都道府県別）<sup>2</sup>の分析から、統計分析の手法やデータの期間の違いにより、場合によっては正反対

<sup>1</sup> 本稿は、2021年5月10日までの公開情報に基づいて執筆している。

<sup>2</sup> 厳密には、「感染者」というより「陽性者」であるが、「感染者」という表現が報道等において使用され、より一般的かと思われるため、本稿では「感染者」の用語を使用する。

の結論を導くこともあり得るといった例について紹介することとしたい。拙稿「回帰分析の不適切使用に注意」<sup>3</sup>では、都道府県別クロスセクションデータから回帰分析を行った場合には関係がなさそうに見えても、パネルデータを構築して固定効果モデルで分析を行った場合には統計的に有意な関係となる例について、また、拙稿「犯罪と社会経済情勢に関するシンプルな計量分析」<sup>4</sup>では、クロスセクションデータの回帰分析では統計的に有意な関係が見られても、パネルデータとして固定効果モデルで分析を行った場合には、その有意性が失われてしまう例について取り上げてみた。本稿では、クロスセクションデータの分析とパネルデータの分析で係数の符号が逆になり、また、同じ手法によるパネルデータの分析でもデータの期間が異なれば係数の符号が逆転し、しかも、これらのいずれの場合も係数の  $t$  値が統計的に有意な値となっていることなどについて紹介することとしたい<sup>5</sup>。

なお、本稿で取り上げる題材は、2020年12月末から全国的に停止されているGOTOトラベル事業との関係では、少々重なる部分がある。本稿は、GOTOトラベル事業そのものの評価や同事業の再開の是非等について論じることを目的とするものではないが、筆者としては、GOTOトラベル事業と新型コロナの感染拡大の関係等については、憶測や偏見を排し、科学的・客観的な視点・方法論に基づいたエビデンスによって検証を行うことが不可欠であると考えている<sup>6</sup>。

## 2. 月別に見た日本全国の延べ宿泊者数と新型コロナ感染者数

まずは、月別の延べ宿泊者数と新型コロナ感染者数（「新規感染者」のことである。以下、同じ）について、日本全国の状態を確認しておくこととしたい。

図表1は、新型コロナの感染拡大が始まり最初の緊急事態宣言が出された

<sup>3</sup> 参議院企画調整室『経済のプリズム』第187号（2020.5）45頁参照。

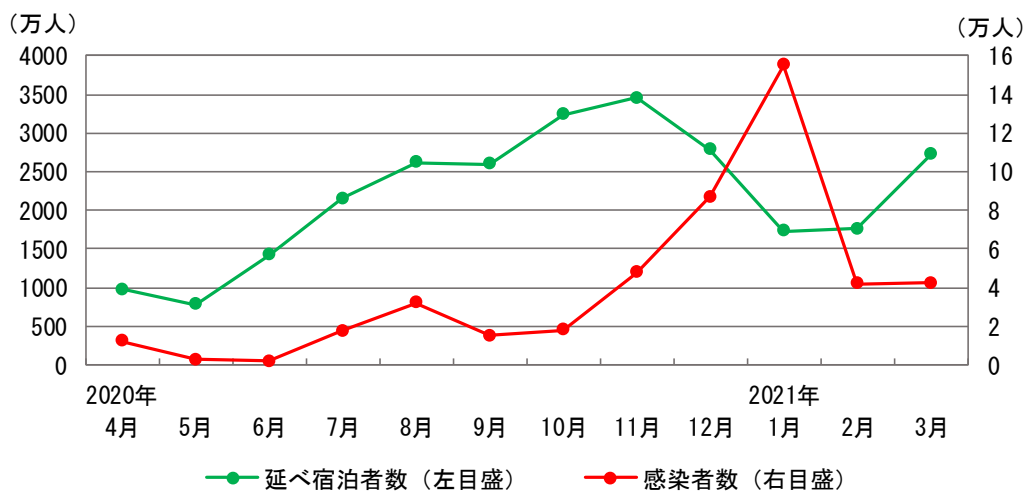
<sup>4</sup> 参議院企画調整室『経済のプリズム』第193号（2020.12）9頁参照。

<sup>5</sup> クロスセクションデータ（横断面データともいう）とは、時間をある一時点で固定して、その時点における場所、人、世帯等の状況を項目として記録したものである。これに対し、ある1つの項目について時間の経過に沿って記録したデータを、時系列データという。これらを合わせ、幾つかの項目（場所、人、世帯等）について時間の経過に沿って記録したデータを、パネルデータという。

<sup>6</sup> GOTOトラベル事業と新型コロナの感染拡大の関係については、京都大学のグループの研究（<https://www3.nhk.or.jp/news/html/20210125/k10012831381000.html>）や国立感染症研究所等の研究者の研究（<https://www.traicy.com/posts/20210321201010/>）などが公表されている。なお、政府は、航空旅客数と感染者数の増加に統計的な因果関係は確認できないとしている（第14回新型コロナウイルス感染症対策アドバイザリーボード（2020.11.19）参考資料（<https://www.mhlw.go.jp/content/10900000/000696903.pdf>））。

2020年4月（GOTOトラベル事業が始まる前である）から、2回目の緊急事態宣言が一部の都県で継続中であった2021年3月について、日本全国の延べ宿泊者数<sup>7</sup>と新型コロナ感染者数の推移を示したものである。これを見ると、月別の延べ宿泊者数については、全国で緊急事態宣言が解除された後の2021年6月から11月までは増加傾向であるが、12月から1月にかけては減少しており、その後は増加に転じている。他方で、新型コロナの感染者数については、このグラフ上では2020年4月から10月まではほぼ横ばいの動きであるが、10月を過ぎると2021年1月までは大きく増加し（いわゆる第3波）、その後の2月、3月になると、今度は大きく減少している。

図表1 月別の延べ宿泊者数と新型コロナ感染者数の推移



(注) 1. 感染者数は、下記のNHKサイト (<https://www3.nhk.or.jp/news/special/coronavirus/>) におけるNHKの取りまとめによる（以下の図表、推計式でも同様である）。  
 2. 延べ宿泊者数は、2021年3月分は第1次速報に、それ以前は第2次速報による。  
 (出所) 観光庁「宿泊旅行統計調査」、NHK「特設サイト 新型コロナウイルス」より作成

### 3. 都道府県における延べ宿泊者数と新型コロナ感染者数の単回帰分析

#### (1) クロスセクションデータの回帰分析

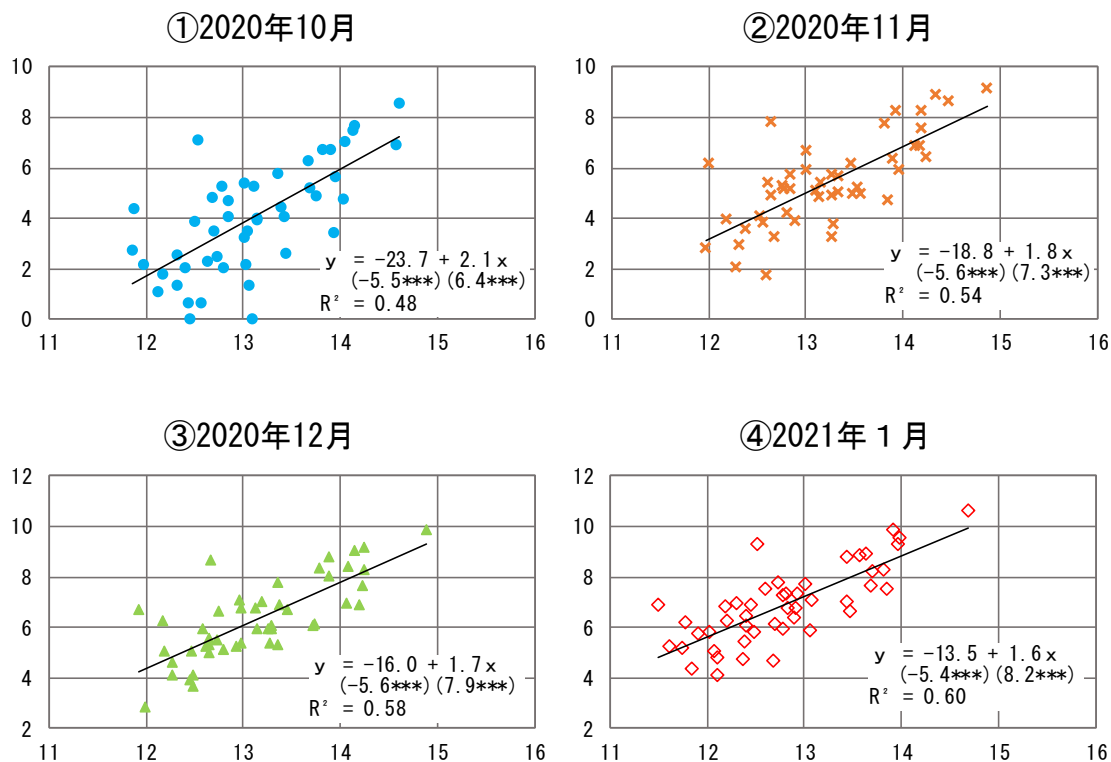
以上が日本全国における月別の延べ宿泊者数と新型コロナ感染者数の状況であるが、ここからは、都道府県における月別の延べ宿泊者数と新型コロナ感染者数のデータから、両者の関係について考えていくこととしたい。そこで、ま

<sup>7</sup> 延べ宿泊者数とは、各日の全宿泊者数（子供や乳幼児も1人として計算する）を足し合わせた数のことであり、例えば10人がそれぞれ3泊した場合には、 $10 \times 3 = 30$ の計算で、延べ宿泊者数は30人となる。

ずは、図表1で新型コロナウイルス感染者数の増加が始まる2020年10月から2回目の緊急事態宣言が出された2021年1月までの両者の関係について(恣意的であるが、本稿のテーマとの関係により、あえてこの期間を1つの区切りとしたい)、クロスセクションデータの回帰分析を行うこととする。図表1では、傾向として、2020年10月から2021年1月には延べ宿泊者数と新型コロナウイルス感染者数が反対の動き(逆相関<sup>8</sup>といえる)となっているが、都道府県別のクロスセクションデータでの回帰分析では、両者はどのような関係となるだろうか。

図表2①から④は、2020年10月から2021年1月の各月の延べ宿泊者数(横

図表2 都道府県における延べ宿泊者数と新型コロナウイルス感染者数の関係



- (注) 1. いずれも、縦軸は感染者数(対数値)、横軸は延べ宿泊者数(対数値)である。  
 2. 感染者数は当月の数値、延べ宿泊者数は前月の半数と当月の半数の合計値である(以下の図表、推計式でも同様である)。  
 3. 定数項及びxの係数の下の()内の数値はt値である。「\*\*\*」はt値が1%の有意水準を満たすことを示す。

(出所) 観光庁「宿泊旅行統計調査」、NHK「特設サイト 新型コロナウイルス」より作成

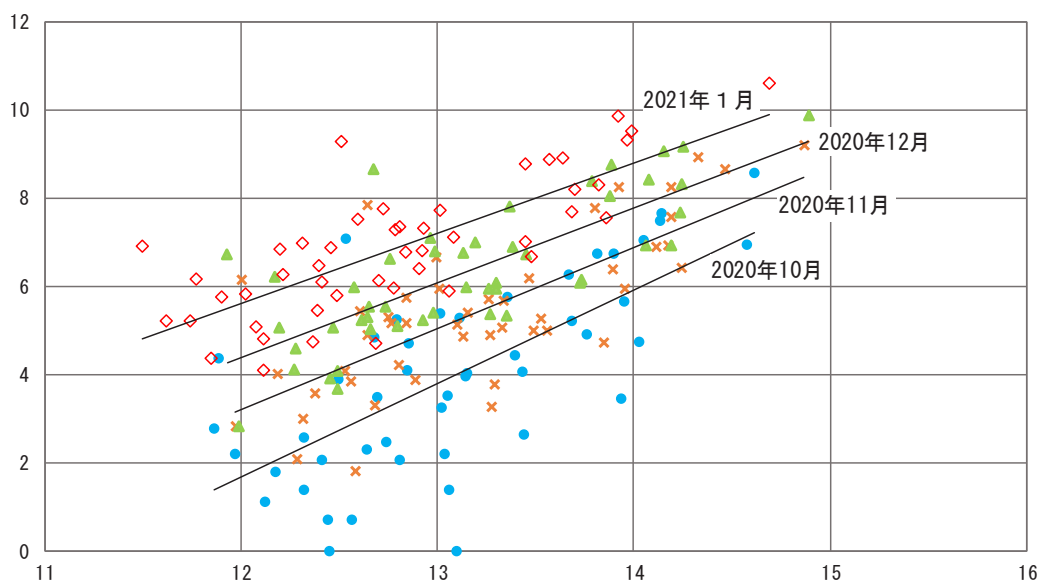
<sup>8</sup> 逆相関とは、一方が増加すれば他方が減少する関係のことであり、負の相関ともいう。これに対し、一方が増加すれば他方も増加する関係を順相関(正の相関)という。

軸)と新型コロナウイルス感染者数(縦軸)について(いずれも対数値に変換)、両者の関係を散布図に描き、回帰直線及び回帰式を示したものである<sup>9</sup>。

図表2を見ると、2020年10月から2021年1月のいずれの月においても、延べ宿泊者数が多い都道府県ほど新型コロナウイルス感染者数も多くなっており、このことに鑑みれば、延べ宿泊者数からうかがえるような人の移動が新型コロナウイルスの感染拡大の要因となっている可能性が考えられる。また、季節的に秋から冬へと気候が厳しくなっていくにつれ、切片のマイナス幅が徐々に縮小してきている(つまり、月の経過に伴って切片が大きくなってきている)。これは、図形的には回帰直線が上方に移動してきているということであり、すなわち、月が進むにつれて感染者数も増加していくことを意味している。

また、図表2①から④について、これらを同一の座標平面上の1つの図表として表したものが図表3である。このように回帰直線が上方に移動する理由に

図表3 図表2①から④の散布図と回帰直線を1つの図表にした場合



(注) 縦軸は感染者数(対数値)、横軸は延べ宿泊者数(対数値)である。

(出所) 観光庁「宿泊旅行統計調査」、NHK「特設サイト 新型コロナウイルス」より作成

<sup>9</sup> 新型コロナウイルスには、ウイルスが体内に侵入した後、増殖して発症に至るまでにタイムラグがあり、その期間は目安として2週間とされているようである。そこで、図表2以降の分析ではこの点を考慮し、便宜的であるが、月別の延べ宿泊者数における「月」を半月分ずらして計算することとする。具体的にいうと、例えば2020年10月の場合には、新型コロナウイルスの感染者数はそのまま2020年10月の数値であるが、それに対応する延べ宿泊者数については、2020年9月の数値の半数と10月の数値の半数の合計値をもって10月の数値とする。

については、一般論・常識論としては、冬になって気候が厳しくなれば新型コロナ感染者も増加するであろうという季節性の要因が考えられる。本稿は月単位であるので、これをとりあえず「月効果」と呼ぶこととして、さらに考察を進めることとしたい。

## (2) ダミー変数(月効果)を加えた回帰分析

次に、こうした月効果を組み入れた回帰式・推計式について検討することとする。これは、発想としては、図表2①から④の回帰式の切片を月効果に見立て、それぞれに各月(2020年10月、11月、12月、2021年1月)のダミー変数を充てることで、4つの回帰式を1つの推計式にまとめて表現しようというものである。つまり、図表2①から④のそれぞれのデータを1つにまとめ(図表3の元になっているデータとなる)、被説明変数を各都道府県の新型コロナ感染者数(対数値)、説明変数を延べ宿泊者数(対数値)及び各月のダミー変数として回帰分析を行うものである。なお、この場合の推計式では、多重共線性の問題を避けるため切片(定数項)を加えていない<sup>10</sup>。そうしてできた推計式は、次のような形となる(推計式1)。

### 推計式1 新型コロナ感染者数と延べ宿泊者数及び月効果の関係

$$y_{it} = 1.8x_{it} - 19.6D_1 - 18.4D_2 - 17.4D_3 - 16.2D_4 + u_{it} \quad (R^2 = 0.66)$$

(14.4<sup>\*\*\*</sup>)    (-11.7<sup>\*\*\*</sup>)    (-10.8<sup>\*\*\*</sup>)    (-10.3<sup>\*\*\*</sup>)    (-9.9<sup>\*\*\*</sup>)

- (注) 1.  $y$  は新型コロナ感染者数(対数値)、 $x$  は延べ宿泊者数(対数値)、 $u$  は誤差項であり、それぞれの変数の右下の添字  $i$  は都道府県、 $t$  は時間(月)を示す。また、 $D_1$  は10月、 $D_2$  は11月、 $D_3$  は12月、 $D_4$  は1月に、それぞれ1の値を取り、それ以外の月には0となるダミー変数である。
2.  $x_{it}$  と各ダミー変数の係数の下の()内の数値は  $t$  値(不均一分散に対して頑健な標準誤差による)。右肩の「\*\*\*」は、 $t$  値が1%の有意水準で有意であることを示す。 $R^2$  は決定係数である。
3. データの出所は、観光庁「宿泊旅行統計調査」、NHK「特設サイト 新型コロナウイルス」による。

推計式1は、延べ宿泊者数が増加すれば感染者数も増加するという順相関の関係があり、さらに、2020年10月から2021年1月に、つまり秋から冬へと気

<sup>10</sup> 多重共線性とは、互いに関連性の高い説明変数が存在すると計算が不安定となり、回帰式の精度が極端に悪くなるなどのことである。ダミー変数を使用する場合には、影響の度合いを知りたい項目の全てをダミー変数とすると完全な多重共線性が発生するため、ダミー変数の個数を1減らすか、又は本稿のように切片(定数項)を加えない方法が採られる。

候が厳しくなっていくにつれ、月効果によっても感染者数が増加するという結果となっている。ダミー変数の符号は、計算の結果としてはすべてがマイナスとなっているが、数値の大きさは $D_4$ （1月）、 $D_3$ （12月）、 $D_2$ （11月）、 $D_1$ （10月）の順（降順）であり、これは、 $D_4$ 、 $D_3$ 、 $D_2$ 、 $D_1$ の順に月効果が大きいことを意味している。つまり、前述の図表3と同様に、気候が厳しくなるにつれ感染者が増加する傾向にあるということである。

したがって、これまでのところからは、延べ宿泊者数からうかがえるような人の移動が増えるにつれ、また、秋から冬にかけて気候が厳しくなっていくにつれ、新型コロナ感染者数も増加していくという関係が見られるという結論となる。このことは、常識的に見てもそれなりに説得力がありそうだがということができよう。

#### 4. パネルデータの固定効果モデルによる分析

##### （1）2020年10月から2021年1月までのデータから

次に、図表2①から④（及び推計式1）で用いたデータをパネルデータとして、分析を行うこととする。パネルデータの分析には幾つかの手法があるが、ここでは、そのうちの代表的なものとして、固定効果モデルによる分析を行う。固定効果モデルとは、観察个体（ここでは都道府県）ごとに異なるが時間を通じて一定である変数（これを「固定効果」という）を回帰式に加えて分析を行うというものである。すなわち、固定効果モデルの基本的な考え方については、この場合には、都道府県ごとに固有の47個のダミー変数を加えて回帰分析を行ったのと同様ということが出来る<sup>11</sup>。

2020年10月から2021年1月までの各都道府県の新型コロナ感染者（対数値）を被説明変数、各都道府県の延べ宿泊者数（対数値）を説明変数として、都道府県効果（固定効果のことである）を加えた固定効果モデルで分析を行った結果は、次のようになる（推計式2）。

---

<sup>11</sup> 前述の推計式1についても、使用しているデータは推計式2と同じものであることから、パネルデータに対する分析ということもできる。このように考える場合、推計式1において月効果として加えたダミー変数は、観察个体（都道府県）では一定であるが、時間を通じて異なる変数であることから、時間固定効果ということが出来る。すなわち、推計式1は、パネルデータに対する時間固定効果モデルともいえるものである。なお、推計式1の月効果（時間固定効果）と、ここでいう都道府県効果（固定効果）の両方を加えて推計を行うことも可能であり、こうした方法については、本文後述の5（都道府県効果ともに月効果を加えた固定効果モデル）において説明している。



### 推計式2 新型コロナ感染者数と延べ宿泊者数の関係

$$y_{it} = \alpha_i - 2.9 x_{it} + u_{it} \quad (R^2 = 0.69)$$

(-10.7\*\*\*)

- (注) 1.  $y$  は新型コロナ感染者数 (対数値)、 $x$  は延べ宿泊者数 (対数値)、 $\alpha$  は都道府県効果、 $u$  は誤差項であり、それぞれの変数の右下の添字  $i$  は都道府県、 $t$  は時間 (月) を示す。  
2.  $x_{it}$  の係数の下の () 内の数値は  $t$  値 (クラスター構造に対して頑健な標準誤差による)。右肩の「\*\*\*」は、 $t$  値が 1% の有意水準で有意であることを示す。 $R^2$  は決定係数。  
3. データの出所は、観光庁「宿泊旅行統計調査」、NHK「特設サイト 新型コロナウイルス」による。

推計式2を見ると、延べ宿泊者数 (対数値) ( $x$ ) の係数の  $t$  値は統計的に有意な値であるが、係数の符号がマイナスとなっていることが分かる。つまり、図表2や推計式1では延べ宿泊者数が増加すると新型コロナ感染者数が増加する順相関であったのに対し、推計式2では、これとは反対に、延べ宿泊者数が増加すると新型コロナ感染者数が減少するという逆相関の関係となっているのである。そうであれば、推計式2の意味は、2020年10月から2021年1月までの間は、延べ宿泊者数からうかがえるような人の移動が新型コロナ感染者数を減少させる方向に働いていたということになる。しかし、この結論は、推計式1とはいわば正反対であり、奇異に感じられるのではなかろうか。

### (2) 2020年4月から9月までのデータから

そこで、今度は少し見方を変え、2020年の秋から冬に感染者数が大きく増加する前の4月から9月までのデータについても、これをパネルデータとして、手法としては推計式2と同様に都道府県効果を加えた固定効果モデルによって分析を行うこととする。そうした結果が、次の推計式3である。

### 推計式3 新型コロナ感染者数と延べ宿泊者数の関係 (その2)

$$y_{it} = \alpha_i + 2.6 x_{it} + u_{it} \quad (R^2 = 0.76)$$

(11.2\*\*\*)

- (注) 1.  $y$  は新型コロナ感染者数 (対数値)、 $x$  は延べ宿泊者数 (対数値)、 $\alpha$  は都道府県効果、 $u$  は誤差項であり、それぞれの変数の右下の添字  $i$  は都道府県、 $t$  は時間 (月) を示す。  
2. 感染者数が0の月は対数値に変換できないため、便宜的に、各都道府県の各月の感染者数に1を加えたうえで対数値に変換している。こうした場合も分析の結果は影響を受けない。  
3.  $x_{it}$  の係数の下の () 内の数値は  $t$  値 (クラスター構造に対して頑健な標準誤差による)。右肩の「\*\*\*」は、 $t$  値が 1% の有意水準で有意であることを示す。 $R^2$  は決定係数。  
4. データの出所は、観光庁「宿泊旅行統計調査」、NHK「特設サイト 新型コロナウイルス」による。

推計式3を見ると、延べ宿泊者数 (対数値) ( $x$ ) の係数がプラスとなってお

り、t 値も 1% の有意水準で有意となっている。したがって、2020 年 4 月から 9 月の間は、延べ宿泊者数が増加すれば、新型コロナ感染者も同様に増加するということであるが、このことは、図表 2 や推計式 1 とは整合的である一方で、今度は推計式 2 とは正反対となるのではないかという、かなり理解しがたい結果となっている。

### (3) どうして同じ固定効果モデルでも正反対の結果が出てくるのか

推計式 2 と推計式 3 では、同じ固定効果モデルでありながら、説明変数の効果が逆方向を向いており、しかも、ともに統計的に有意という結果となっている。ただ、こうしたことが起こる理由は、理論的に考えれば、それほど複雑なことではない。端的に言えば、その理由は、図表 1 にあるといえる。

推計式 2 や推計式 3 のように都道府県効果を加えた固定効果モデルでは、被説明変数に対する影響・効果のうち、都道府県（一般的に言えば観察个体）ごとに異なるが時間を通じて一定であるものは、都道府県ごとの固定効果（推計式上は「 $\alpha$ 」）に含められてしまう<sup>12</sup>。そのため、推計式 2、推計式 3 における延べ宿泊者数（対数値）の係数は、意味的には、感染者数（対数値）に対する、時間を通じて変化する影響・効果を数値で表したものとなる。ここで、もう一度、図表 1 を見ると、2020 年 4 月から 9 月（推計式 3 のデータの期間）は、延べ宿泊者数と感染者数の動きは同じ方向を向いており、延べ宿泊者数が増加すれば感染者数も増加する順相関の関係にあるといえるのに対し、2020 年 10 月から 2021 年 1 月（推計式 2 のデータの期間）は、感染者数が大きく増加する一方で、延べ宿泊者数は減少傾向となっており（2020 年 12 月と 2021 年 1 月の下げ幅が大きい）、逆相関の関係となっていることが分かる。都道府県効果を加えた固定効果モデルでは、こうした時系列的な関係が変数の係数に反映されることとなるため、結果的に、推計式 2 では係数がマイナスの逆相関、推計式 3 では係数がプラスの順相関となったと考えられる。なお、図表 1 は日本全国の合計値であるが、いずれの都道府県でも状況はさほど変わらないと考えられ、むしろ、そうした状況が都道府県別のパネルデータの固定効果モデルでの分析結果に現れていると考えるのが妥当であろう。

---

<sup>12</sup> 固定効果モデルでは、係数等の推計の際に第 1 段階として、変数の変換によって固定効果を一旦除去する作業（これを固定効果変換という）を行う。詳細については、西山義彦ほか『計量経済学』（有斐閣（2019））216 頁などを参照。

## 5. 都道府県効果とともに月効果を加えた固定効果モデル

### (1) 推計式2に月効果を加えるとどうなるか

推計式2と推計式3のように、月別で見た時系列的な動きに異なった傾向がある変数の場合には、いつからいつまでを分析の対象とするかによって、正反対の結果となることがあり得る。そうであれば、厳密な分析のためには、こうした時系列的な傾向の影響を取り除く方法が必要であるということになる。

そこで、本稿ではその一つの方法として、固定効果モデルにおいて、都道府県効果とともに、更に時間固定効果を加えた推計式によって分析を行うこととする。具体的には、2020年10月から2021年1月までの延べ宿泊者数(対数値)と新型コロナウイルス感染者数(対数値)のパネルデータについて、前述のものと同様の都道府県効果とともに、月効果(都道府県(一般的に言えば観察个体)では一定であるが、月(一般的に言えば時間)を通じて変化する変数(時間固定効果))を更に加えたモデルで推計を行うということである。そうした結果が、次の推計式4である。

#### 推計式4 新型コロナウイルス感染者数と延べ宿泊者数の関係(その3)

$$y_{it} = \alpha_i + \beta_t - 0.12 x_{it} + u_{it} \quad (R^2 = 0.91) \\ (-0.12)$$

- (注) 1.  $y$  は新型コロナウイルス感染者数(対数値)、 $x$  は延べ宿泊者数(対数値)、 $\alpha$  は都道府県効果、 $\beta$  は月効果、 $u$  は誤差項であり、それぞれの変数の右下の添字  $i$  は都道府県、 $t$  は時間(月)を示す。
2.  $x_{it}$  の係数の下の()内の数値は  $t$  値(クラスター構造に対して頑健な標準誤差による)。なお、推計式4では、 $t$  値は有意水準を仮に10%とした場合でも有意ではない。
3. データの出所は、観光庁「宿泊旅行統計調査」、NHK「特設サイト 新型コロナウイルス」による。

推計式4を見ると、延べ宿泊者数(対数値)( $x$ )の係数の $t$ 値は統計的に有意とはいえずなくなっており、このことは、延べ宿泊者数(対数値)と新型コロナウイルス感染者数(対数値)の関係は無関係という帰無仮説を棄却することができないことを意味している。つまり、都道府県効果や月効果を除いてみると、延べ宿泊者数(対数値)は新型コロナウイルス感染者数(対数値)に影響を与えていないのではないかということになる<sup>13</sup>。

<sup>13</sup> なお、推計式4の $p$ 値( $t$ 値をパーセント表示としたものと考えることができる)は0.90であり、これは、延べ宿泊者数(対数値)が新型コロナウイルス感染者数(対数値)とは無関係(本当のところは係数が0である)という可能性が90%程度ということを意味している。

## (2) 延べ宿泊者数（対数値）の係数の t 値が有意でなくなる理由

推計式 4 の結果は、ここに来て議論が振り出しに戻されるような話となっているが、今までの分析では統計的に有意となっていた延べ宿泊者数（対数値）と新型コロナウイルス感染者数（対数値）の関係が、都道府県効果と月効果を加えると、どうして有意でなくなってしまうのだろうか。

その理由をかいつまんでいうと、推計式 4 では月効果が加わり、推計式 2 での時系列的な関係が月効果として独立した変数のように扱われることになったため、都道府県効果と延べ宿泊者数（対数値）の間に一種の多重共線性<sup>14</sup>の問題が顕在化したことが考えられる。推計式 4 から都道府県効果を取り出して 2020 年 10 月から 2021 年 1 月の各都道府県の延べ宿泊者数（対数値）との間の相関係数<sup>15</sup>を計算すると、10 月 0.77、11 月 0.78、12 月 0.80、1 月 0.82 と、かなり高い数値となる。このように、時系列的な効果が除かれた各都道府県の延べ宿泊者数（対数値）と都道府県効果との間に線型的に強い（似通った数値となる）関係が生じるため、回帰式の計算が不安定となっている可能性が考えられるのである。これを比喩的にいうと、被説明変数である感染者数（対数値）の大部分が都道府県効果及び月効果から説明されてしまうため、延べ宿泊者数（対数値）が説明できる余地は残されたごくわずかな部分となり、そのために係数がゼロに近く、また、t 値が低くなり、有意性が失われることになったということになる。

なお、多少余談であるが、推計式 4 の都道府県効果は、数値としては都道府県別の人口（対数値）<sup>16</sup>とも非常に似通った値になっている。この両者の相関係数を計算すると 0.94 となり、都道府県効果と延べ宿泊者数の場合よりも高い値となる。そこで、少し見方を変えて考えてみると、推計式 4 の都道府県効果については、これを実質的に都道府県別の人口（対数値）と見ても、あながち間違いではないのではないかとも思われる<sup>17</sup>。そうした場合、推計式 4 の意味は、新型コロナウイルス感染者数（対数値）に影響を及ぼす要因は都道府県別の人口（推

<sup>14</sup> 多重共線性については、前掲注 10 参照。

<sup>15</sup> 相関係数とは、2 つの変数の間にある線形な関係の強弱を図る指標であり（プラスマイナス 1 の範囲を取る）、相関係数が正のときには正の相関が、負のときには負の相関があるとする。また、相関係数が 0 のとき両者は無相関である。

<sup>16</sup> 総務省「人口推計」による 2019 年 10 月 1 日現在の都道府県別人口である。

<sup>17</sup> なお、前述の図表 2 ①から④（2020 年 10 月から 2021 年 1 月）について、説明変数に都道府県人口（対数値）を加えた重回帰分析を行った場合には、延べ宿泊者数（対数値）の係数の t 値は、仮に有意水準を 10% とした場合でも有意ではなくなってしまう。これに対し、都道府県人口（対数値）の係数の t 値は、1% の水準で有意となる。

計値) と月効果であるということとなり、結局のところ、延べ宿泊者数や人の移動はともあれ、人口が多い都道府県では新型コロナ感染者数も多くなり、冬になって気候が厳しくなれば感染者数は更に増えるであろうという、少々拍子抜けするような結論となってしまふ。

## 6. おわりに

以上述べてきたように、延べ宿泊者数(対数値)と新型コロナ感染者数(対数値)の関係については、同じデータを使用しても一方では順相関の関係(図表2、推計式1)、他方では逆相関(推計式2)の関係となるとともに、データの期間を変えたり(推計式3)分析手法を別のものにすると(推計式4)、更に異なった結果となっている。したがって、一般論としていうと、統計分析の際には、一つの分析により何らかの結果が得られたとしても、その結果を絶対視するのではなく、他の可能性についても検討を行い、可能であれば別の分析方法を試してみるといったことが重要ということになる<sup>18</sup>。

また、延べ宿泊者数と新型コロナ感染者数の関係が本当のところはどうかという点については、本稿の分析からは確定的なところまではいえないと思われる。確かに、都道府県効果と月効果を取り入れた推計式4では、延べ宿泊者数によって説明できる部分はゼロかもしれないという結果となった。とはいえ、論理的に考えれば、観光であれビジネスであれ、宿泊を伴う旅行により人の移動が多くなれば、旅行先で感染させたりさせられたりという具合に感染者数が増えることもあり得るとするのが自然の成り行きであろう。この点については、今のところ筆者としては明確な回答を持っているわけではないが、GOTOトラベル事業と新型コロナの感染拡大の関係に関する議論も含め<sup>19</sup>、今後の検討課題とすることとしたい。

(内線75044)

---

<sup>18</sup> 本稿では、統計ソフトとしてEViews11を使用した。こうしたソフトがあれば、いわばクリック1つで手軽に様々な分析を行うことができるが、その一方で、適切な分析方法の選択については、ソフトの使用者の判断による。

<sup>19</sup> なお、2021年3月下旬以降は気候は暖かくなっているが、新型コロナ感染者数は増加傾向である(第4波)。また、2020年10月から2021年1月(第3波)とそれ以前(2020年4月の第1波や8月の第2波)についても、PCR検査の実施人数(厚生労働省ホームページ(<https://www.mhlw.go.jp/stf/covid-19/kokunainohasseijoukyou.html>)参照)等を見てもうかがえるように、それぞれ状況は異なっている。厳密に人の移動と新型コロナの感染拡大の関係を見ようという場合には、適切な変数を加えるなど、分析方法に工夫が必要かと思われる。