

回帰分析の不適切使用に注意

～完全失業率と自殺死亡率の関係から～

調査情報担当室 前田 泰伸

《要旨》

政府の白書や経済論文等では、都道府県別のクロスセクションデータから説明変数をX軸、被説明変数をY軸とした回帰分析を行い、図表として散布図や回帰直線を示すことがよく行われる。こうした方法は、視覚的効果により両変数の関係を分かりやすく説得的に読者に伝えることができる反面、場合によっては誤解やミスリードを招く原因にもなり得る。例えば、完全失業率と自殺死亡率については、時系列データから見ると両者の間に強い関係がありそうに見えるが、1年限りのクロスセクションデータで回帰分析を行うと、統計的に無関係といわざるを得ないという、互いに矛盾するような結果となる。こうした場合の解決策・考え方の1つとして、本稿ではパネルデータ分析の手法である固定効果モデルで分析を行い、その結果、完全失業率と自殺死亡率の間にはやはり統計的に関係がありそうだという結論を得た。

1. はじめに

政府の白書や経済論文等では、都道府県別のクロスセクションデータを用いて、説明変数をX軸、被説明変数をY軸とした回帰分析¹を行い、散布図や回帰直線を図表として描くことにより、両変数の関係を視覚的に分かりやすく示すということが、比較的よく行われている。こうした回帰分析を、例えば、都道府県別の労働生産性（説明変数）と労働者の収入（被説明変数）の関係について行ってみると²、図表による視覚的な効果が加わることにより、労働者の収入

¹ 回帰分析には、説明変数が1つの単回帰分析と2つ以上の重回帰分析がある。本稿で回帰分析という場合は、特に重回帰分析と断りを入れていない限り、単回帰分析である。

² 厚生労働省「労働経済の分析」（2015）193頁では、労働生産性と一般労働者の平均年収額について、こうした回帰分析を行っている。なお、拙稿「労働時間、賃金、労働生産性の関係について」（『経済のプリズム』第184号（参議院事務局企画調整室（2020.2））37頁）では、労働生産性と現金給与総額の関係等について、同様の分析を行った。

を上げるためには基本的に労働生産性を高める必要があるというメッセージを、より効果的・説得的に読者に伝えることができる。ところが、こうした回帰分析は、使い方によっては、思わぬ誤解やミスリードを招くこともあり得る。本稿では、不適切な回帰分析の例として、完全失業率と自殺死亡率の関係をとり上げることとし、安直な回帰分析に対する注意喚起とともに、自らに対しても今後の戒めとしたい。

2. 失業と自殺の関係

まずは、失業と自殺について、一般的によくいわれているところを示すこととする。自殺とは、多様かつ複合的な原因及び背景を有し、様々な要因が連鎖する中で起きるものであるが³、こうした原因や背景の中でもとりわけ失業については、従来から自殺との強い関係が指摘されてきたところである⁴。失業、すなわち働く意欲も能力もあるのに仕事がないという状況は、収入が途絶えることにより自分や家族の生活が成り立たなくなるという経済的なダメージだけではなく、「自分は立派な社会人、職業人である」という自信やプライドを打ち砕き、「自分は役に立たないダメな人間だ」などと、人を精神的に追い詰めるものともなり得る。

そこで、完全失業率と自殺死亡率について時系列的にグラフを追ってみると、不況時などで完全失業率が上昇すると、それに対応するように自殺死亡率も上昇する関係が見られる。図表1は、1978年から2019年の完全失業率と自殺死亡率の推移（時系列データ）を同じグラフ上に示したものであるが、これを見ると、大きな流れとしては、バブル崩壊後に雇用情勢の悪化を受けて完全失業率が上昇すると自殺死亡率も上昇し、2010年代の雇用情勢の改善により完全失業率が低下すると自殺死亡率も低下するという動きとなっている。厳密な時系列分析の手続はさておき⁵、完全失業率（説明変数：X）が自殺死亡率（被説明変

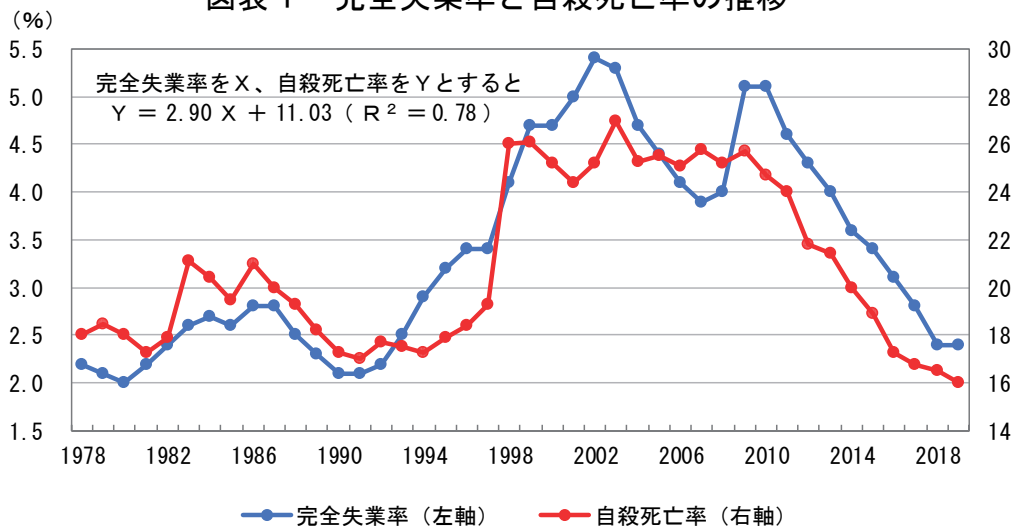
³ 毎年の厚生労働省「自殺対策白書」、警察庁「(各年中における)自殺の状況」において、こうした言及がある。また、具体的な自殺の原因・背景としては、家庭問題、健康問題、経済・生活問題、勤務問題、男女問題、学校問題の6つが挙げられている。

⁴ なお、自殺死亡率を男女別に見ると、一般的に男性は女性に比べて自殺死亡率が高く、2019年の自殺死亡率は男性22.9、女性9.4となっている（警察庁「令和元(2019)年中における自殺の状況」参照）。自殺死亡率には男女間の性差も関係していると考えられるが、自殺死亡率そのものの分析については別稿に委ねることとしたい。

⁵ 時系列データを扱う際には、単位根過程にある変数同士を回帰させると、本当は無関係なのに決定係数等が高い数値となり、あたかも関係がありそうな結果となる「見せかけの回帰」に注意する必要がある。厳密な分析のためには、単位根検定など複雑な手法が必要となる（沖本竜義『計量時系列分析』朝倉書店(2010)124頁参照）。

数: Y) に影響を及ぼすという前提で両者の関係を回帰式で示すと、「 $Y = 2.90 X + 11.03$ 」となり、これは、完全失業率が1%上昇すると自殺死亡率が2.9上昇することを示している。また、便宜的に回帰式の決定係数 (R^2) を算出すると0.78となり、決定係数の定義上、自殺死亡率のかなりの部分が完全失業率によって説明されることになる⁶。

図表1 完全失業率と自殺死亡率の推移



(注) 自殺死亡率とは、人口10万人当たりの自殺者数である (以下同じ)。
(出所) 警察庁「(各年中における) 自殺の状況」(1978~2019)、総務省「労働力調査」より作成

3. 都道府県別データから見た完全失業率と自殺死亡率

(1) 都道府県別のクロスセクションデータから回帰分析をする

失業と自殺の間には、社会学的な観点のみならず、図表1のように統計的にもかなり強い関係がうかがえる。そうであれば、今度は時系列データからではなく、都道府県別の完全失業率と自殺死亡率のクロスセクションデータ⁷から、被説明変数 (Y) を自殺死亡率、説明変数 (X) を完全失業率として回帰分析を行うとすると、この場合においても、両者の間に統計的に有意な関係が見出せることが予想される場所である。ところが、実際にこうした回帰分析を行ってみると、時系列データの場合とは異なり、両者の間には統計的に関係がなさそうな結果となってしまう。

⁶ 決定係数とは回帰式の当てはまりのよさを示す指標であり、0から1までの数値を取り、1に近いほど回帰式の当てはまりがよい。

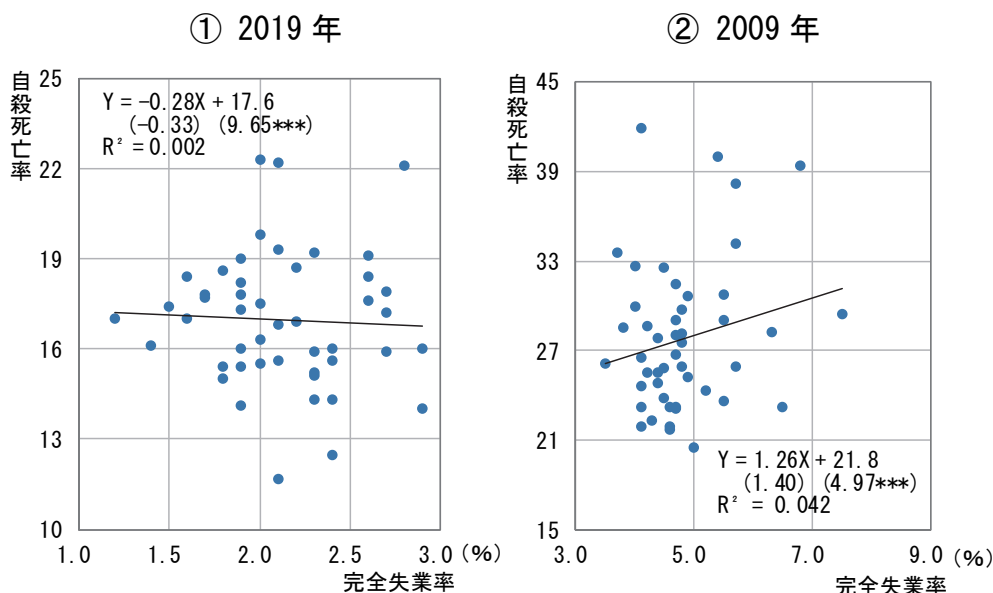
⁷ クロスセクションとは横断面という意味であり、時間を一時点に固定し、その時点における各地点、場所、人や世帯等 (観察個体) において発生したデータを収集・記録したものがクロスセクションデータである。

図表2①は、2019年の都道府県別の自殺死亡率を被説明変数（Y）、完全失業率を説明変数（X）として回帰分析を行ったものである。2019年については、回帰式は「 $Y = -0.28X + 17.6$ 」とXの係数がマイナスであり、これは完全失業率が高い都道府県ほど自殺死亡率が低くなることを示しているが、そもそもXの係数のt値は有意水準を10%としても有意ではなく、決定係数（ R^2 ）も0.002と、ほとんどゼロといってよい数値である。このように、景気がよく雇用情勢も好調であった2019年は、完全失業率と自殺死亡率の間で統計的に有意な関係は見いだしがたい結果となっている。

では、景気よかった2019年ではなく、リーマン・ショック直後で景気も雇用情勢もかなり悪化していた2009年では、どういった結果となるだろうか。図表2②は、2009年について、2019年と同様に都道府県データから回帰分析を行ったものである。この場合も、Xの係数のt値が有意ではなく（なお、①では符号がマイナスであるが、②では逆転してプラスとなっている）、決定係数（ R^2 ）も0.042とかなり低い値となっている。

このように、図表2では、完全失業率が高い（低い）都道府県では自殺死亡率が高い（低い）という関係は統計的に見られず、図表2だけを根拠とすれば完全失業率と自殺死亡率は無関係ではないかとも考えられるのである。

図表2 都道府県別に見た完全失業率と自殺死亡率の関係



(注) 回帰式の係数及び定数項の下の数値はt値。「***」はt値が1%の有意水準を満たすことを示す。R²は決定係数。なお、回帰式の誤差項については、記載を省略した。

(出所) 警察庁「令和元(2019)年中における自殺の状況」、「平成21(2009)年中における自殺の状況」、総務省「労働力調査(都道府県別結果・モデル推計値)」より作成

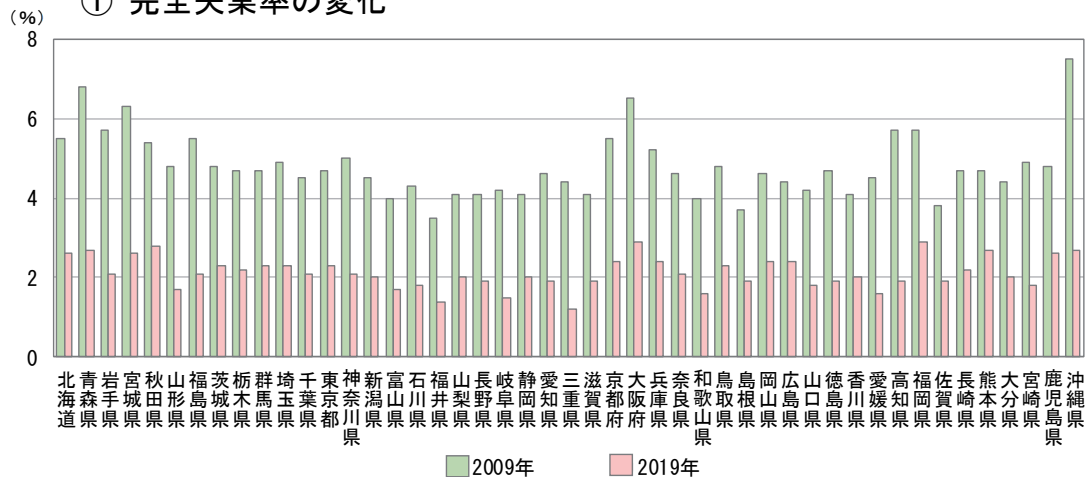
(2) 完全失業率と自殺死亡率の変化をそれぞれ詳しく見てみると

ここで少々視点を変え、散布図ではなく、具体的な都道府県名を示しつつ、各都道府県の完全失業率と自殺死亡率が2009年から2019年の間でどのように変化したのか見ることとする。

図表3は、各都道府県の完全失業率と自殺死亡率についてそれぞれ、2009年と2019年の数値を棒グラフで示し、変化の幅が分かるようにしたものである。これを見ると、いずれの都道府県も例外なく、完全失業率、自殺死亡率ともに、2019年は2009年に比べて改善（数値が低下）していることが分かる。改善の度合いは都道府県によってばらつきはあるが、この図表から単純に考えるとすれば、完全失業率が改善し、自殺死亡率も同様に改善しているのであれば、傾向

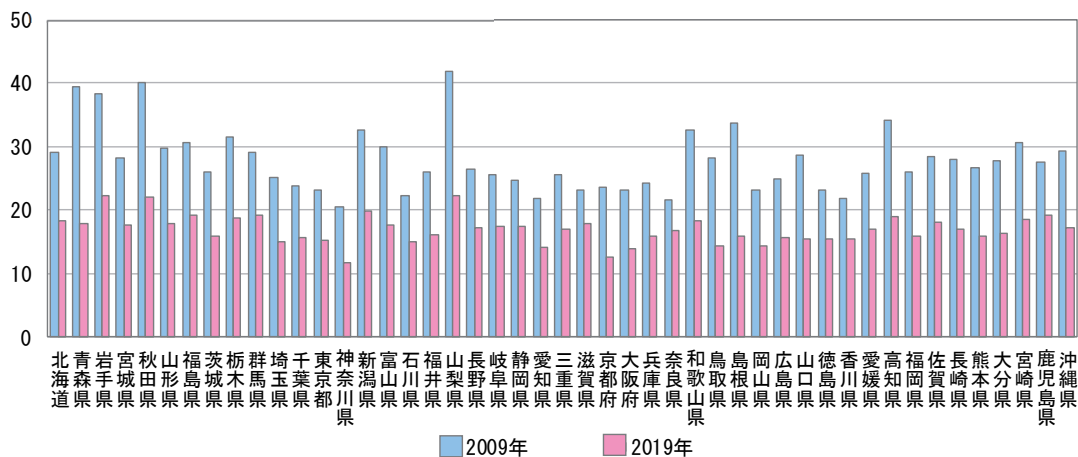
図表3 2009年から2019年の完全失業率と自殺死亡率の変化

① 完全失業率の変化



(出所) 総務省「労働力調査（都道府県別結果・モデル推計値）」より作成

② 自殺死亡率の変化



(出所) 警察庁「平成21（2009）年中における自殺の状況」、「令和元（2019）年中における自殺の状況」より作成

として、両者の間で一方が改善（低下）すればもう一方も改善（低下）するという関係を、ひとまずは肯定してよさそうにも見える。

ただし、その一方で、図表3をもう一度よく見てみると、2009年、2019年のいずれも、例えば大阪府では他の都道府県に比べると完全失業率は相対的に高くなっているが自殺率はそれほど高くなく、また、これとは反対に、山梨県では完全失業率は高いといえないが自殺死亡率はかなり高くなっているなど、図表2で見たように、完全失業率が低い（高い）都道府県と自殺死亡率が低い（高い）都道府県が完全に対応しているとはいえない結果となっている。

（3）自殺死亡率と都道府県それぞれの個別事情について

さらに、各都道府県の自殺死亡率の順位に着目することにより、自殺死亡率に影響を及ぼす各都道府県それぞれの個別事情との関係について考える。

図表4は、2009年、2014年、2019年の都道府県別の自殺死亡率について、その数値が低い順に、上位（低い）と下位（高い）の10の都道府県（2014年と2019年は10位に同率で2つの都道府県が並んでいる）を取り出して並べたものである。これを見ると、自殺死亡率の水準は全体的に低下傾向にあるが、都道府県

図表4 自殺死亡率の都道府県別の順位（2009年、2014年、2019年）

	2009年	2014年	2019年
低い ↑	1: 神奈川県 20.5	1: 神奈川県 15.6	1: 神奈川県 11.7
	2: 奈良 21.7	2: 大阪 15.7	2: 京都 12.5
	3: 愛知 21.9	3: 岡山 16.9	3: 大阪 14.0
	3: 香川 21.9	4: 石川 17.2	4: 愛知 14.1
	5: 石川 22.3	5: 京都 18.0	5: 岡山 14.3
	6: 徳島 23.1	6: 奈良 18.2	5: 鳥取 14.3
	7: 東京 23.2	7: 香川 18.3	7: 石川 15.0
	7: 滋賀 23.2	8: 福井 18.4	8: 埼玉 15.1
	7: 大阪 23.2	9: 愛知 18.7	9: 東京 15.2
	7: 岡山 23.2	10: 埼玉 19.0	10: 山口 15.4
...	10: 熊本 19.0	10: 徳島 15.4	
...	
↓ 高い	38: 福島 30.7	38: 鹿児島 22.9	38: 宮崎 18.6
	39: 栃木 31.4	39: 島根 23.8	39: 栃木 18.7
	40: 新潟 32.6	40: 高知 24.0	40: 高知 19.0
	41: 和歌山 32.7	41: 福島 24.7	41: 鹿児島 19.1
	42: 島根 33.6	42: 富山 24.9	42: 群馬 19.2
	43: 高知 34.2	43: 宮崎 25.0	43: 福島 19.3
	44: 岩手 38.2	44: 新潟 26.3	44: 新潟 19.8
	45: 青森 39.4	45: 秋田 26.7	45: 秋田 22.1
	46: 秋田 40.0	46: 岩手 29.1	46: 岩手 22.2
	47: 山梨 41.9	47: 山梨 30.6	47: 山梨 22.3

(注) 1. 数値は小数第1位まで示した。都道府県名の左側の数値は順位である。
 2. 2009年、2014年、2019年のいずれの年も自殺死亡率が低い又は高い10位以内の都道府県は濃い網掛け、そのうち2年について順位が10位以内の都道府県は薄い網掛けで示した。
 (出所) 警察庁「平成21（2009）年中における自殺の状況」、「平成26（2014）年中における自殺の状況」、「令和元（2019）年中における自殺の状況」より作成

別の順位については、2009年、2014年、2019年のいずれの年も、神奈川県、大阪府、愛知県などは自殺死亡率が低い10位以内の位置にあり、反対に山梨県、秋田県、岩手県などは自殺死亡率が高い10位以内となっている。つまり、2009年から2019年にかけて、いずれの都道府県でも完全失業率は低下し、自殺死亡率も水準としては低下しているが、都道府県で自殺死亡率の順位を付けてみると、その順位はいずれの年も大きな変化がなく、ある程度は固定されたものとなっているのである。これには、完全失業率以外の各都道府県それぞれの個別事情（その具体的内容はさておき）の影響が考えられるところである。

以上のことから考えると、都道府県の自殺死亡率は、図表1や図表3からうかがえるような完全失業率の低下に伴って低下する部分とともに、図表4のような、完全失業率とは無関係の各都道府県それぞれの個別事情によって影響を受ける部分から構成されるといえるのではないだろうか。そうであれば、完全失業率という1つの説明変数によって自殺死亡率を説明することは(図表2参照)、方法としては適切でなかったということになるだろう。

4. 結局どうすればよいのか～パネルデータ分析の手法

(1) 固定効果モデルとは

以上のように、完全失業率と自殺死亡率の関係については、都道府県別のクロスセクションデータから説明変数を1つ（完全失業率）として回帰分析を行うと、適切な分析とならない可能性がある。そこで、本稿では、これまでの論述を踏まえ、1つの方法として、パネルデータ分析の手法である固定効果モデルについて紹介することとしたい⁸。

パネルデータとは、時系列データとクロスセクションデータを組み合わせたものであり、各地点、場所、人や世帯等（観察個体）において発生したデータを複数の時点において収集・記録したものである。パネルデータでは、時系列データやクロスセクションデータと比べて情報量が多くなり推計の精度が向上するほか、とりわけ固定効果モデルでは、前述の都道府県それぞれの個別事情のような、観察個体の異質性を考慮できるという利点がある。固定効果モデルは⁹、具体的には、観察個体ごとに異なるが時間を通じて一定である変数を仮に

⁸ 別の方法としては、説明変数が2つ以上の重回帰分析が考えられる。完全失業率以外に自殺死亡率に影響を及ぼす変数（具体的にその変数が何かはさておき）を見つけ出し、重回帰分析を行えば、これらの変数と自殺死亡率の間で統計的に有意な関係を見いだす可能性がある。

⁹ パネルデータ分析には、他には変量効果モデル等の手法も用いられるが、議論が複雑になるため、本稿では割愛する。なお、奥井亮「固定効果と変量効果」(『日本労働研究雑誌』No. 657 (労働政策研究・研修機構 (2015.4)) 8頁) 参照。

α として（これを「固定効果」という）、この α をモデルに加えて分析を行うものであり、回帰式は「 $Y_{it} = \alpha_i + \beta X_{it} + u_{it}$ 」のような形で表される。都道府県の自殺死亡率と完全失業率の関係に即していうと、 Y （被説明変数）は自殺死亡率、 X （説明変数）は完全失業率、 α が固定効果であり、 u は誤差項に当たる。それぞれの変数の右下の添字は、 i が各都道府県、 t が時間（調査年）を示すものである。 α の添字が i だけで t がないのは、 α が時間を通じて一定であることを表現している。

なお、図表2のような回帰分析の回帰式は「 $Y_i = C + \beta X_i + u_i$ 」（ Y （被説明変数）は自殺死亡率、 X （説明変数）は完全失業率、 C は定数項、 u は誤差項であり、右下の添字 i は各都道府県を示す）と表すことができるが、固定効果モデルとの違いは、固定効果モデルでは固定効果 α は都道府県ごとに別々の値を取るのに対し、図表2のような回帰分析の定数項 C は、すべての都道府県において同じ値となることである¹⁰。

（2）分析結果～やはり完全失業率と自殺死亡率には関係があるのではないか

2009年から2019年の毎年の各都道府県における完全失業率と自殺死亡率のデータから固定効果モデルで分析を行うと、結果は次のようになる¹¹。

完全失業率と自殺死亡率の関係（固定効果モデル）

$$Y_{it} = \alpha_i + 3.53 X_{it} + u_{it}$$

[0.16]
(22.07***)

$$R^2 = 0.84$$

- (注) 1. Y （被説明変数）は自殺死亡率、 X （説明変数）は完全失業率、 α は各都道府県の固定効果であり、 u は誤差項に当たる。それぞれの変数の右下の添字は、 i が各都道府県、 t が時間（調査年）を示す。
2. X_{it} の下の[]内の数値は標準誤差。標準誤差は、自己相関と不均一分散に頑健な漸近分散推定量に基づく標準誤差（クラスター構造に頑健な標準誤差）による。
3. X_{it} の下の()内の数値は t 値。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(自由度調整済み)。
4. データの出所は、警察庁「(各年中における)自殺の状況」(2009～2019)、総務省「労働力調査(都道府県別結果・モデル推計値)」

このように、固定効果モデルで分析を行うと、説明変数（完全失業率）の係数は3.53と、各都道府県において完全失業率が1%上昇すれば自殺死亡率が

¹⁰ なお、固定効果モデルでも定数項を想定することはできるが、これを固定効果 α の中に含めて考えて、省略されることも多い。

¹¹ 使用した統計ソフトは、EViews11である。

3.53上昇する結果となり、説明変数に係る t 値は1%の有意水準でも有意となるなど、やはり完全失業率と自殺死亡率の間には統計的にも相当に強い関係があることがうかがえるものとなる(各都道府県の固定効果については、補論図表参照)。なお、決定係数 (R^2) もかなり高い数値となっている¹²。

5. むすび

以上のように、都道府県別のクロスセクションデータから説明変数をX軸、被説明変数をY軸として回帰分析を行い散布図を示すという方法は、図表による視覚的な効果を伴って説得力が増すこともあり、比較的によく用いられている。しかし、本稿で示した完全失業率と自殺死亡率の関係では、クロスセクションデータでの回帰分析では統計的に有意な関係が見られず、これだけを根拠に速断してしまうと、両者の間に関係はないであろうと、適切とはいえない結論を導く可能性がある。また、これとは反対に、クロスセクションデータでは有意な関係が見られるのに、パネルデータとして分析を行うと有意な関係がなくなってしまうということもあり得る¹³。結局、必要なことは、単に都道府県別のクロスセクションデータから回帰分析を行ったところ有意な関係が見られた、あるいは見られなかったというだけで終わるのではなく、分析の結果について、それが経済的・社会的にどのような意味を持つのかなど様々な角度から吟味しつつ、総合的に判断することによって、妥当な結論を導くということであろう。

補論～各都道府県の固定効果

補論として、各都道府県の固定効果(前述の α_i)を示す(補論図表)。固定効果は都道府県で別々の値を取り、都道府県によって高いところ、低いところなど様々であるが、基本的には、自殺死亡率は、固定効果の数値が低い都道府県では引き下げられ、その数値が高い都道府県では引き上げられることになる。

¹² 固定効果モデルでは、固定効果(α_i)として各都道府県にそれぞれ1つのダミー変数を充てたと考えることもでき、こうした固定効果(ダミー変数)の説明力が高いことが多いため、決定係数(R^2)も高くなる傾向があり、その解釈には注意が必要である(西山慶彦ほか『計量経済学』有斐閣(2019)246頁参照)。

¹³ 例えば、基本的には保育所定員率が高い都道府県では女性の有業率も高いことが考えられるが(例えば、厚生労働省「厚生労働白書」(2017)124頁参照)、総務省「国勢調査」の都道府県別データからパネルデータを構築し、都道府県の固定効果を加味したモデルで分析を行うと、保育所定員率と母親の就業の間にかような正の相関関係が見られなくなってしまうという研究もある(朝井友紀子ほか「保育所整備と母親の就業率」(内閣府経済総合研究所『経済分析』第191号(2016)121頁)参照)。

固定効果の数値が低い都道府県としては、大阪府（1.1）、沖縄県（1.8）、神奈川県（3.5）などがあり、大阪府や神奈川県は図表4において自殺死亡率が低い都道府県として登場していたところである。なお、沖縄県の固定効果は大阪府に次いで全国で2番目に小さいが、沖縄県の自殺死亡率はそれほど低いわけではない（図表3②参照）。沖縄県の場合は、県内の雇用情勢を背景として完全失業率が比較的に高くなっていることが影響していると考えられる（図表3①参照）。また、固定効果の数値が高い都道府県は、山梨県（19.4）、岩手県（15.8）、島根県（15.2）などとなっている¹⁴。

補論図表 都道府県の固定効果

都道府県	固定効果	都道府県	固定効果
北海道	7.8	滋賀	10.3
青森	9.5	京都	4.2
岩手	15.8	大阪	1.1
宮城	6.4	兵庫	6.5
秋田	14.5	奈良	5.2
山形	12.5	和歌山	13.7
福島	11.1	鳥取	10.2
茨城	7.9	島根	15.2
栃木	11.6	岡山	6.8
群馬	12.1	広島	8.2
埼玉	6.7	山口	10.3
千葉	7.3	徳島	7.4
東京	5.8	香川	8.3
神奈川	3.5	愛媛	10.8
新潟	14.9	高知	11.4
富山	13.5	福岡	5.2
石川	9.3	佐賀	11.0
福井	11.7	長崎	8.5
山梨	19.4	熊本	7.3
長野	10.6	大分	9.5
岐阜	11.8	宮崎	12.7
静岡	10.8	鹿児島	8.9
愛知	7.6	沖縄	1.8
三重	10.6		

- (注) 1. 前述の固定効果モデル ($Y_{it} = \alpha_i + 3.53 X_{it} + u_{it}$) のうち、各都道府県の固定効果 (α_i) を取り出して図表としたものである。
 2. EViews11では、各都道府県の固定効果とともに都道府県で共通の定数項も計算されるため（注10参照）、補論図表ではそれらの数値を合算した。
 (出所) 警察庁「(各年中における)自殺の状況」(2009～2019)、総務省「労働力調査(都道府県別結果・モデル推計値)」より作成

(内線75044)

¹⁴ 警察庁「(各年中における)自殺の状況」でいう都道府県とは、自殺者の住所地がある都道府県ではなく死体が発見された都道府県のことである。