

# 多重回帰分析と回帰不連続デザイン

安藤 道人

(国立社会保障・人口問題研究所研究員)

## I RDDの(不遇な)歴史

Regression Discontinuity Design (RDD: 回帰不連続デザイン)は、ミクロ計量経済学における統計的因果推論において、操作変数法や差の差分析と並んでスタンダードな分析手法の1つである。RDDの最も基本的な形は、観察可能な変数  $V$  がある閾値を超えたときに社会的状況  $X$  が不連続的に変化する(あるいはトリートメント  $X$  が導入される)場合において、その閾値前後でのアウトカム変数  $Y$  の不連続的な変化の大きさを推定することにより、 $X$  が  $Y$  に与える影響を推定することである。

RDDの歴史は古く、Thistlethwaite and Campbell (1960) という教育心理学の論文において初めて用いられ、紆余曲折の末にようやく標準的な計量分析手法として(とくにミクロ計量経済学において)定着した<sup>1)</sup>。RDDは「ミクロ計量経済学発の新しい計量分析手法」と誤解されがちだが、実は50年以上のやや不遇な歴史がある学際的な統計的因果推論の手法である。

## II 多重回帰分析で「シフト」を推定する

本稿では、RDDを多重回帰分析と対比しながら解説する。まず多重回帰分析の特徴を理解するために、以下のような仮想的社会の分析を試行してみよう。

- 設定1. 人口は10万人であり、それぞれ異なる能力を有し、大卒と非大卒がいる。
- 設定2. 能力は0から100まで均一に分布する。
- 設定3. 能力が0で非大卒の場合、平均年収は200万円である。
- 設定4. 能力が1上昇すると年収は10万円上昇する。
- 設定5. 大卒ならば年収は500万円高くなる。

ここで設定5の「500万円」を「大卒効果」と呼び、多重回帰分析の目的は、この500万円をできるだけ正しく推定することであるとする。

また大卒か否かは、以下のケース1あるいは2によって決定されるとする。

- ケース1. 能力が80以上の約2万人の中から約1万人がランダムに選ばれて大卒となる。
- ケース2. 能力を部分的に反映した学力テストの点数が180点以上であれば大卒となる。<sup>2)</sup>

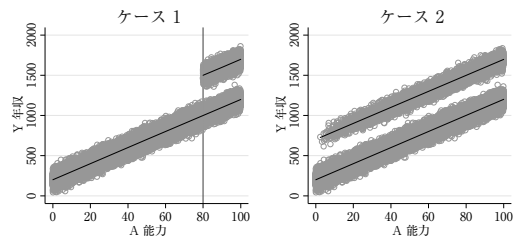
以上の設定2~5とケース1, 2を組み合わせた仮想的社会のデータ生成は、以下のようなデータ生成過程(Data generating process: DGP)で表現される。

$$Y = 200 \text{万} + 10 \text{万} \cdot A + 500 \text{万} \cdot X + \varepsilon \quad (1)$$

ここで  $Y$  は年収、 $A$  は能力、 $X$  は大卒ダミー、 $\varepsilon$  はランダム項であり、 $A$  は設定2、 $X$  はケース1あるいは2によって生成され、 $Y$  は(1)式によって生成される。

(1)式によってデータ生成を行い、横軸に能力  $A$ 、縦軸に所得  $Y$  をとった散布図が図1である<sup>3)</sup>。まずケース1をみると、能力  $A$  が80以上ではっきりと年収が高い層が出現している。これがケース1の選考プロセスで選ばれた大卒約1万人である。一方ケース2では、能力  $A$  の値に拘らず年収が高い層がいる。これは、ケース2における学力テストの点数  $V$  が能力  $A$  を必ずしも忠実に反映しないために、あらゆる能力層から大卒が出現し、高所得を実現するためである<sup>4)</sup>。

図1 能力と年収の関係



では多重回帰分析は大卒効果500万円を捉えられるだろうか? 結論からいえば、大卒効果500万円とは、図1の2つの散布図において示唆される非大卒層から大卒層への年収の「上方シフト」分であり、このシフト分は年収  $Y$  を能力  $A$  と大卒ダミー  $X$  に回帰した多重回帰分析によって、 $X$  の係数値として推定される<sup>5)</sup>。すなわち重回帰分析による大卒効果の推定とは、能力  $A$  が年収  $Y$  に与える影響を線形回帰という形でコントロールしながら、その回帰直線の上方シフト分を大卒ダミー変数  $X$  で捉えることである。紙面の都合で割愛するが、実際に  $X$  の係数値はほぼ500万円となる。

## III RDDで「ジャンプ」を推定する

上述した多重回帰分析による大卒効果の推定には重

大な前提条件がある。それは、能力 A が観察可能ということだ。しかし現実のデータでは能力 A が観察できることはほとんどなく、学力テストの点数 V のような代理変数が観察できるだけである。

それでは能力 A が観察不可能と仮定して、その代理変数として学力テストの点数 V を用いるとどうなるだろうか。それを図示したのが図 2 である。

図 2 学力テストの点数と年収の関係

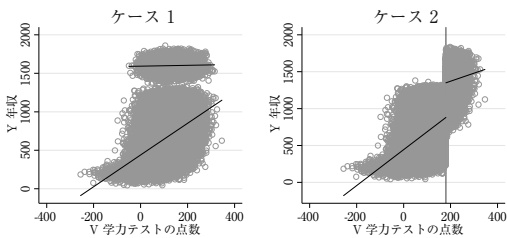


図 2 には図 1 のような年収の上方シフト（平行移動）が観察されない<sup>6)</sup>。実際、ケース 1、2 両方において、年収 Y を点数 V と大卒ダミー X に回帰しても X の係数値すなわち大卒効果の推定値は 500 万近傍とはならない。能力 A を正しく観察できなければ、多重回帰分析では大卒効果を推定できないのである。

ここで RDD の出番である。しかし残念ながら、ケース 1 の場合には RDD は使えない。RDD が使えるのは、ケース 2 のように観察可能な変数 V（学力テストの点数）の閾値によって X（大卒か否か）が振り分けられる場合だけである。RDD は、このような閾値を伴う DGP を利用した「調査設計（デザイン）」なのである。

ではケース 2 を実際に分析しよう。まず図 2 のケース 2 において学力テストの点数 V が 180 点のところ年に年収分布の明確な「ジャンプ」（不連続性）があることに着目する。180 点とは大卒となる制度的・人為的な閾値であり、この閾値前後で能力 A は連続的に分布しているはずである。また能力 A から年収 Y への影響度がこの閾値で唐突に変化するとは考えにくい。従って、この閾値において年収 Y の明確なジャンプがある理由は大卒効果以外には考えられない。これが RDD のロジックである。さらにいえば、閾値の直前直後においては、個々人が閾値の左右どちらに振り分けられるかはほぼランダムであり、局所的に「大卒」の割り当てがランダム化比較実験のように行われたとみなすこともできる。

そして実際、この年収ジャンプの大きさを閾値前後

のデータを用いた回帰分析で推定すると約 500 万円となり、DGP で設定した大卒効果とほぼ同じとなる。この年収ジャンプの推定値が RDD 推定値である<sup>7)</sup>。このように RDD を用いると、たとえ能力 A が観察できなくても、閾値でのデータの変化を利用して大卒効果を捉えることができる。

#### IV 終わりに

以上、単純化された設定下ではあるが、RDD について多重回帰分析と比較しながら解説した。今回は「学力テストの点数の閾値を利用した大卒効果の推定」を分析例としたが、実際の社会には様々な「制度的閾値」が存在し、それを利用した RDD による分析が行われている。最後に、本稿は RDD 推定のごく基本的な考え方を紹介したにすぎない。RDD 推定の入門書としては、Angrist and Pischke (2014) 第 4 章などを参照されたい。

- 1) RDD の心理学、教育学、統計学、経済学における紆余曲折の物語については、Cook (2008) がインサイダーとしての視点を交えながら、非常に興味深くまとめている。
- 2) ここで 180 点という閾値は、大卒がケース 1 と同様に 1 万人程度になるように便宜的に設定している。
- 3) データ生成に必要なその他の細かい設定の詳細は割愛する。また図 1 における 2 直線は大卒と非大卒の回帰直線である。
- 4) 実際には  $V = 30 \log(A) + \text{ランダム項}$  として V を生成した。
- 5) 回帰式は  $Y = a + \beta \cdot A + \gamma \cdot X + \epsilon$  であり、 $\epsilon$  は誤差項である。
- 6) 図 1 同様、2 直線は大卒と非大卒の回帰直線である。
- 7) 実際の RDD 推定法については Angrist and Pischke (2014) 等の入門書を参照されたい。

#### 参考文献

Angrist, JD. and Pischke, JS. (2014) *Mastering Metrics: The Path from Cause to Effect*, Princeton University Press.

Cook, TD. (2008) "Waiting for Life to Arrive": A History of the Regression-Discontinuity Design in Psychology. *Statistics and Economics, Journal of Econometrics*, Vol 142 (2), 636-654.

Thistlethwaite, DL. and Campbell, DT. (1960) Regression-Discontinuity Analysis: An Alternative to the Ex Post Facto Experiment. *Journal of Educational Psychology*, Vol 51 (6), 309-317.

あんどう・みちひと 国立社会保障・人口問題研究所研究員。最近の主な著作に "Dreams of Urbanization: Quantitative Case Studies on the Local Impacts of Nuclear Power Facilities Using the Synthetic Control Method," *Journal of Urban Economics*, Vol.85, 68-85, 2015. 公共経済学・応用ミクロ計量経済学専攻。