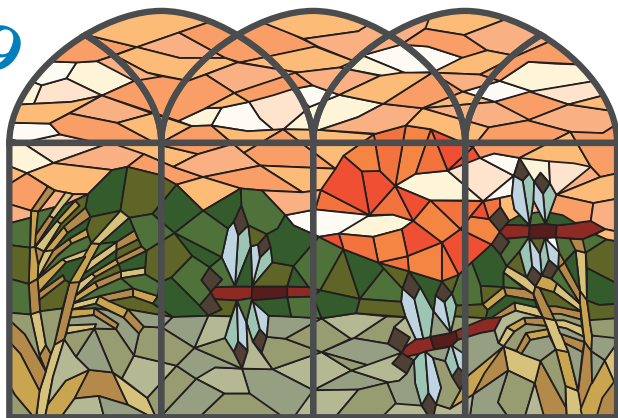


9



4

3

7

数理の窓

2

8

7

5

因果関係への挑戦

「丈の短いスカートが流行している年は景気が良くなる」という法則は、景気動向に関する俗説として知られている。スカートの丈と景気に偶然の相関関係はあるのだろうが、そこに因果関係があると本気で信じている人は少ないだろう。因果関係とは原因と結果の関係のことで、ここではもちろん、「原因＝スカートの丈」、「結果＝景気」である。

データ分析をする際、原因だと想定している変数（＝説明変数）と結果だと想定している変数（＝被説明変数）に本当に因果関係がある、ということを見極める作業は非常に重要であるため、様々な手法が確立されている。これらの手法を「因果推論」と呼ぶ。

因果推論の最も理想的な手法として、ランダム化比較試験が挙げられる。これは元々医療分野等で実施されている調査方法で、近年は経済学の分野でも注目されている。集団を無作為に分割し、一方の集団にはある操作（＝投薬・治療等）を施し、もう一方の集団には何もしないことで、その操作が有効だったか否かを判断するものである。ここで重要なことは、その操作を行うこと以外の条件がすべて同じになることである。そうすることで、純粋にその操作（＝原因）からのみの影響（＝結果）を知ることができる。しかし、試験環境の整備や人材の確保等の観点で莫大なコストがかかるため、これは容易には実施できない。

因果推論の手法の一つとして、パネル・データを利用することも検討されてきた。パネル・データとは、例えば「ある小学校で毎年夏に行っている学力テストでの個人成績の推移」のように、複数の調査対象に対して一定期間の情報を収集したものである。このデータから「そろばん塾に通うことでテストの成績が上がるか？」を調べたい場合、例えば3年生の冬からそろばん塾に通い始めた児童と、全く通っていない児童を分析対象として、3年生・4年生時点の成績を比較することで、因果関係を明らかにしようとする。しかしこれにも、対象間のトレンドの一致（＝そろばん塾という要因以外は同一の条件下にあることを示す）等、厳しい前提が必要である。

データ分析の結果は、様々な意思決定のインプットになるが、誤った因果関係の想定のもとで行われた分析結果は、正しい意思決定を導くことができない。また、分析対象のデータがどれだけ増えたとしても、前提としている因果関係の誤りを取り除くことはできないため、因果推論はビッグデータ分析の世界でも重要なものとなる。

AIによるデータ分析が注目されているが、AIによる因果推論は現時点では難しく、人間による判断が必要と考えられているようだ。任せられる仕事はAIに任せて、因果関係について少し考えてみるのはどうだろう。（松本 ゆかり）