

佐々木 弾 (著)『統計は暴走する』中公新書ラクレ (2017年)

伊藤 公一朗 (著)『データ分析の力 因果関係に迫る思考法』光文社新書 (2017年)

今国会で最重要法案とされる働き方改革関連法案に盛り込まれる裁量労働制の適用拡大に関連し、当初政府は裁量労働者が必ずしも一般労働者と比べて長時間労働でないことを示すために、厚生労働省が行った調査結果を引用して説明した。後に比較することが望ましくないデータであること、データに多くの不備が見つかったことなどから答弁を撤回し、法案から裁量労働制の適用拡大部分を削除するに至っているが、統計データの扱い方はあまりにもお粗末だったといえる。そこで、今回は統計データの扱い方を考えるための2冊を紹介しておきたい。

統計データを見るにあたって、まず『統計は暴走する』で統計リテラシーを向上させてみてはどうだろうか。“だます”、“盗む”、“迫害する”、“殺す”というセンセーショナルな見出しの各章において、著者はいずれも統計データを元にした例題を出しながら、Q&A方式でその真偽や見方を解説していく。例えば「国民投票前夜は、イギリスのEU残留派の勝利予想が80%超」は調査のサンプルに偏りがあったこと、「喫煙率が下がっても肺がんが減らないので、喫煙は肺がんの原因でない」は寿命が伸びたことによる肺がん罹患者の増加などの別の要因が無視されていることなどが指摘されている。データを恣意的に使い、あたかも統計的な裏付けがあるように見せることを防ぐためには、まずは統計を使う側の良心が求められよう。一方で、怪しげな統計データ、分析も少なくない現在、統計を見る側にも最低限の知識や批判的な視点が必要で、『統計は暴走する』にはそのための心構えが示されている。

心構えができた後は、データを活用していくために『データ分析の力 因果関係に迫る思考法』を手にとってもらいたい。本書は、統計学の分析を詳細に説明するものではなく、データ分析に当たって気をつけることを解説しながら統計的にデータ分析をするための基礎、センスを身につけるための入門書である。そのため、統計本にありがちな数式の羅列などはほとんどなく、読み物として手に取れる気軽さがある。しかしながらデータをどのように収集し、分析するかに関しては、非常に重要な示唆は随所にちりばめられている。

検討に当たってのスタートは、因果関係と相関関係の違いを指摘することから始まる。そして、因果関係を正確に把握するための方法としてランダム化比較実験(RCT)が紹介され、具体例として北九州市においてランダムに選んだ2つのグループに対して、片方には夕方から電力価格が上昇すると伝えた場合に、電力の使用に差が出るかどうかに関する実験結果が示されている。結果そのものも興味深いのが、結果を見出すための考え方が、データ収集や分析において大いに参考になる。このRCTの他にも、境界線をうまく使うRDデザインや階段状の変化を利用する集積分析、複数期間でデータが入手できる場合に活用できるパネル・データ分析などの解説もあり、これらの様々なデータ分析をビジネス戦略や政策形成に生かすための実例も紹介されている。アメリカのものが中心であるが、スーパーマーケットでの表示の違いによる売り上げの差、ウーバーにおける価格と需要のバランスの検討などがわかりやすく解説されている。そして、最後にどのようなデータ分析でも不完全性や限界があることについて言及がされており、これがデータ分析に関しての最も重要な示唆であろう。

ネット環境が急速に整い、統計データが容易に手に入るようになってきたからこそ、多くの人がデータの扱い方を知っていくことが必要になる。統計リテラシーを向上させ、分析の基礎・センスを身につけて、データを適切に活用していきたい。(加藤 健志)