

## ゴミの科学

朝のゴミ出しはご主人の仕事という家庭も多いのではないだろうか。ノーベル経済学賞を授与された Granger 教授も、知らせのあった翌朝、いつも通りゴミ出しをしていたのを近所の人に目撃されています。プラスチック、燃えるゴミ、燃えないゴミ、瓶カンペットボトル、資源回収ごみと分別するよりも、まず、果たしてゴミとして捨てていいものか、シュレッダーで粉砕する必要があるのかなどを判断する必要があり、結構面倒くさいですが手は抜けません。

統計学はゴミの科学であるというと怒り出す人もいるかもしれませんが、これはなにも嫌味ではなく、ある統計学関係の研究会で、物理学の研究者がコメントを求められて「まだゴミの科学ですか？」と強烈な一言を発せられたことを思い出したにすぎません。この一言は強烈ではあるものの、本質をついています。統計学の歴史はゴミとの闘いといってもよいでしょう。いかにゴミを見分けるか、どうしたらゴミまみれの品物から中身だけをうまく取り出せるかといった課題に取り組んできたのが統計学の中心です。ここでのゴミはいわゆるノイズであり、誤差、エラー、攪乱などと呼ばれることもあります。統計的検定はまさしく、ゴミかどうか判断する手段といってもよいのです。個人的には統計学を「ゴミの科学」と呼ぶのも悪くないと思っていますが、学生がそっぽを向いてしまうかもしれないし、「最強の学問」が「ゴミの科学」だったらベストセラーにはならなかったかも知れませんね。

最近、統計学の看板をデータサイエンスに架け替える動きがさまざまな大学で進んでいますが、「ゴミの科学」がそのままデータサイエンスとして通用するとは思えません。データから宝を見つけだすのに、ゴミをどう扱ったらよいか重要であることに変わりはありませんが、どこにどんな宝が眠っているのか知ることが関心事であることのほうが多いです。しかし学会などで、急遽データサイエンスの旗振りを始めた人に「あなたの提唱するデータサイエンスとはどんなものなんですか」と質問しても、いやな顔をされるだけでまともな答えが返ってきた試しがありません。それどころか、ある要職の人は「柴田君はそういうことを言うからいけないんだ」と怒る始末。それ以来、この質問は封印することにしました。

トーマス・クーンの科学革命（パラダイムシフト）を持ち出すまでもなく、ニューパラダイム「データサイエンス」は統計学の連続的な変化ではありえません。統計学を内包しながらも、新たな気持ちでのチャレンジが必要な若いデータの科学なのです。

2018-03-05 里程