



データ中心科学と統計思考力

横幹連合副会長 田村 義保*



横幹連合には設立前から、色々な立場で参加させていただいております。平成23年度からは副会長として会の運営を担当させていただいております。設立前の企業アンケートに関わった経験等から、横断型という修飾語のとらえ方については、分野、研究経歴ごとに、かなりの温度差があるように感じて来ました。ホームページには、『横断型基幹科学技術とは、論理を規範原理とし、自然科学、人文・社会科学、工学などを横断的に統合することを通して異分野の融合を促し、それにより新しい社会的価値の創出をもたらす基盤学術体系である〔補足説明〕たとえば、社会、人間、環境、生命、経営、組織マネジメントなどを扱うために生み出された、統計学、シミュレーション学、最適化手法、情報学、設計学などの学術体系である』とあります。最近では、この定義に違和感が無くなったと言うことは、私の横幹への理解が深まったせいかもしれません。しかし、「統計学帝国主義者」ではありませんが、カール・ピアソンが統計学を「科学の文法」と位置付けたように、統計学が横断型科学技術の中心であると考え続けています。

統計学はデータからその裡に潜む法則性、規則性を見出すための学問です。また、見つけた規則性、法則性を用いて予測や制御を行うための方法を提供するためのものです。予測結果を使って、何らかの対策、計画を実施するために用いることができます。また、効率的にデータを収集するためには、実験計画や調査計画を適切に立てる必要がありますが、このための方法の研究も統計学の範囲内であると考えています。

データ中心科学は、実験科学、理論科学、計算科学に続く第4の科学のパラダイムとして必要性が叫ばれています。新しいパラダイムであるかもしれ

ませんが、統計数理研究所の樋口所長が、この4つの科学を車の車輪に例えているように、どの一つが欠けても、目指す方向に科学は発展して行かないように思います。今までも、勿論、データは扱っており、ケプラーがティコ・ブラーエの観測データを用いて惑星が太陽を焦点とする楕円軌道で運動していることを見出したのは、まさしく、データ中心科学の成功例の一つだと思われます。ことさら、データ中心科学と言われるようになったのは、センサー技術の発展、通信網の充実等により、超大量のデータが取得できるようになり、それらを解析することにより、多くの利益が得られることを期待しているためであると考えます。

今回の津波の高さは「想定外」であったようで、多くの人的、物的被害がありました。しかし、本当に「想定外」だったのでしょうか？過去のデータを、謙虚に解析していたら、10メートルや20メートルの津波が来る可能性は推測できたはずで、一般的に言って、人間は直近の経験からしか判断できません。このため、最近ではそれほど大きな津波が来ていないので、今後も来ないだろうとの判断のもとで、堤防等の設計を行ったものと思われます。だったら、20メートルの堤防は必要でしょうか？500年や1000年に一度程度の災害のために、堤防等を用意することの適否を判断することは難しいことです。そんなに長く持つ堤防の材料はないことも判断を難しくしています。また、高い堤防を立てることは、海からの風の流れを変えますので、産業・生活に影響を与えすぎます。津波をすばやく予測して、すぐに逃げるができるようになるか、高台に住むという対策の方が良いと思います。予測すべきは波の高さの変化なので、潮位系を適切に設置し、分析・予測方法を開発しておけば、減災につながります。振り切れない地震計を設置するための予算が計上されたそうです。震源近くでは、

*統計数理研究所 データ科学研究系 教授

既存の地震計は振り切れてしまい、マグニチュードを正しく計算できなかったということが理由だそうですが、少しおかしいように思います。マグニチュードは遠方の地震計のデータを使っても計算できるのであり、今回も米国は、早い内に、マグニチュードは9クラスであると発表しています。米国のデータはすぐに日本にも提供されていたので、こちらを使えば、マグニチュードはすぐに計算でき、津波の高さもより適切に予測できたかもしれません。データがあっても、それを使わないのでは、全く意味がありません。また、目的のためには、どのようなデータを集めるべきかの視点が必要です。ふりきれない地震計があると、震源近くでも加速度が測れますが、減災には役立つまいに思います。

もう一つ、SPEEDIのデータを出さなかったのはおかしいという議論が今でもなされています。あれが無くて、ヨーロッパの各国の気象庁は放射能の拡散シミュレーションを動画で発表していました。放出量が分からないために、あくまでも、仮想の濃度分布でしたが、拡散の様子は分かったはずで、SPEEDIの最初の頃の発表で私が目にしたのは、二次元ベクトル図です。ページプリンター時代の出力であり、懐かしさで一杯でした。データ中心科学の柱の一つである可視化を念頭においた発表をして欲しかったように思います。

色々な機関が放射能の強度を発表しています。特に、気にかかるのが、食料品に含まれる放射能の量だと思います。ベクレルで書かれても、1年間に何ミリシーベルトになるかがよく分からないところがあります。1ミリシーベルト以下を基準にする

か、10ミリシーベルトを基準にするかは、政府が決めても良いのですが、基準を超えたら、直ちに危険ということは無いことを、丁寧に国民に説明する必要があります。日本人が特にそうなのかもしれないませんが、多くの方は○-×、Yes-Noの二元論的に考えているように思われます。統計学的、確率論的に判断しようとしていない感じがします。科学者の多くの発言も二元論であるのは悲しい限りです。これは、データが少なく、データから得られるであろう法則が見出されていないためです。今回の原発事故で、明らかに日本は世界中に迷惑をかけています。海はつながっているため、まだ、南極で福島由来のセシウムは観測されていないようですが、地球全体を汚染してしまう可能性は大です。この罪滅ぼしと今後の放射線暴露への対策のために、正しく計画された観測、疫学調査を行い、データを集積すべきです。日本人は医学でデータについては利己的で、今回も、「モルモットではない」という声が聞こえますが、全日本人がモルモットになって、放射線の影響データを少なくとも20年くらいは収集していくべきです。このようなことを国民に納得させるのは政府の役割ですが、政府や国民の統計思考力を高め、データを収集し横幹型の方法を用いて解析する必要があることを納得させるのは横幹型の研究者の使命であると考えています。

まとまりのない文章になった感が大です。しかし、統計的思考力を高めることで、日本は今回の地震、事故から復興できるだけでなく、さらなる発展もできるものと考えています。