

データ科学以前

小笠原 春 彦

大学の学部卒業を控えて進路については、なにも迷わず就職を決めていた。就活もそれなりに行き、分野もさまざまに数社を回った。当時は売り手市場で高を括っていたと思う。そうしてぶらぶらしているとき所属していた教育学部教育心理学科というところの20歳代の助手の方から、国鉄（日本国有鉄道、JRの前身）の研究所で採用があるから小笠原君受けてみないかと言われた。その頃は計量心理学の応用を卒論のテーマにしており、その助手の方の専門もそうだったので就職希望の私に薦めてくれたのだろう。これまた何も迷わず受けてみる気になった。

教育心理学科というのは進学先としては少数派であろう。大学の入学時から決めていたのではなかった。入学したのは経済学進学コースであった。文科的なことと理科的なことと両方できそうなので選んだという記憶がある。実際、経済学部では計量経済学やプログラミングのほかに工学の授業もあったようだ。しかし、あまり興味が持てず1、2年のころ所属していたクラブ活動の心理学研究会というところで勉強していた分野に惹かれていった。3、4年の専門課程へは何の抵抗もなく進路変更して教育心理学科を選んだ。いまになって思えば結果としてよかったものの安易だったかもしれない。

さて、就職予定先の国鉄には研究所が技術関係のものほかに鉄道労働科学研究所（国鉄労研といっていた）という100人規模のものがあり、後者が心理職の研究員を毎年1名ほど募集していた。他の各分野の大卒採用の応募者とともに入社試験を受けた。筆記試験は東京の駿台予備校で受けた。このほかにあるテーマについてディベートをさせるというのもあった。その得点がよかった

わけでもないと思うが運よく合格した。

入社してすぐに研究所に配属になるのではなく2年間ほど1974年の同年次に入社した事務系・技術系の新入職員とともに現業部門と非現業部門を実習した。3年目から研究所で国鉄職員に実施する心理適性検査の仕事を行った。国鉄は1987年に民営分割され、研究部門は財団法人として統合独立した。500人位の規模だった。鉄道（JR）総研の略称で呼んでいた。データに基づくJR社員の効率的な人的管理を目的とした適性検査の研究開発の仕事が継続してあり、これに従事した。

国鉄とJR総研で21年あまり勤めた。このときのデータ解析がその後の統計科学の研究に直接結びついた。データというのは不思議で思い入れの有無によって見え方が異なる。国鉄とJR総研の時は、人的適性管理のある部門の責任者として、全国各地に依頼して必要なデータを収集する一方、自分でも直接適性検査を実施して従業員の資質に関するデータを集めた。埼玉県大宮市の近くにJR東日本の研修センターというのがあり、本社の責任者に趣旨を話して同意を得てから、現場に直接依頼してそこによく出かけた。まず、JR総研の紹介をしてから適性検査を受けてもらった。負担のある検査を数百人の方々が真面目に受けてくれた。

それらのデータ解析には、どうしても新しい統計科学の手法が必要となった。このようなデータを用いた最初の統計科学の研究は国鉄時代で1986年に研究所の紀要に日本語で書いたものであった。その後、本学に赴任するまで様々な統計科学の手法の開発がほとんどこれらのデータ解析と結びついていた。最初の学位取得は1992年であったが統計手法の貴重な適用例としてそれらのデータを用いることができた。このような経緯は統計学ではある意味で「常識」である。天下り式に新しい手法を開発し、それを様々な分野に適用するということはありうるけれども、多くの場合は応用分野からの必要性によって優れた手法が開発されてきたのである。最近は「統計学」、「統計科学」に替えてデータサイエンス（データ科学）ということが多くなった。これは常識が表面化しただけと思う。日本にはほとんどなかった統計学部の代わりに滋賀大学ではデータサイ

エンス学部というのが2017年に生まれた。これに続き2018年春に横浜市立大学データサイエンス学部が誕生予定である。しかし、データの重要性は昔からあったのである。

統計科学で多数の指標（国語、算数の得点や身長、体重の測定値等で変量という。）の関連を調べる手法の分野として多変量解析という領域があり、その代表的なものに因子分析の手法がある。これは100年以上前に教育心理学において知能の解析から生まれたもので現在では医学、工学、経営学、心理学等様々な分野で用いられるようになってきている。また、統計学におけるモデル選択の規準として世界で広く使われている赤池情報量規準（AIC）というのがあるが、その導出の動機の一つは因子分析の応用だった。

本学に赴任後は授業の期末試験で成績を得るなどのほかは直接データを得ることはなくなった。新しい手法を発表するときは多くの場合その有効性を示すために解析例が必要となる。real data（実データ）という言い方があり、投稿した論文の審査ではそれを用いた解析を要求されることがよくある。計算機による数値実験データも存在しない虚構のデータではないので実データでなくもないが、はっきりした表現でreal-life data（実生活上のデータ）ということもある。これが必要な場合は現在では文献による他、データの使用を依頼することもある。しかし、以前に自らデータを集めた時の経験がそれらを用いる時の選択基準として活かしている。そのような約20年の経験があったことは統計科学の研究者として幸せなことだったとつくづく思う。