

モデルについて

篠崎 智大 (東京大学)

計量生物学の学問としての拠りどころは何と言っても統計学ですが、統計学に欠かせないのは「モデル」です。統計学と言わず、およそ「科学」と呼ばれる学問分野全般に、理論の基礎づけとして「モデル」が不可欠だと言えるかもしれません。いずれにしても、われわれは普段からデータに対峙して科学的営みを行おうとするとき、意識するかしないかを問わず「モデル」を使って事に当たることに慣れていきますし、それが上手く機能しているようにも思われます。一方で「モデル」とはなにか、なぜ「モデル」を使った推論が科学者どうしの合意に一役買うか、などと問うてみると、一筋縄ではいかないこの問いに往生することも想像に難くないと思われます。

この困難のもとを考えるには、そもそも「モデル」という言葉の多義性を看取する必要があるでしょう。われわれの座右にいつもある「回帰モデル」を例にとってみても、その物理的解釈には「測定誤差モデル」と「サンプリングモデル」が可能だと言われますが (たとえば Gelman & Hill, 2006), ここで既に「モデル」の意味が変化しています。「回帰モデル」と言う場合、字義通りには「条件付き期待値を (数式等で) 簡略化した表現」ですが、「測定誤差/サンプリングモデル」はむしろ、データの変動の原因を「個々の測定値の誤差」のせいにしてデータ自体が定まらないと考えるか、個々のデータ自体は変化しないがランダムサンプリングによって観測データが変動すると考えるか、単に「回帰モデル」と言うのとは階層のちがう、データ生成 (なにをわれわれが観測しているのか、という弱い意味) に関わる「ものの見方の枠組み」を指します。「一般化線形モデル」は回帰モデルの一種ですが、これらを構成する「構造モデル」は条件付き期待値の近似的表現という意味での「回帰モデル」ですし、もうひとつの構成要素である「サンプリングモデル」はデータの出方をどの分布関数で近似するかということ、今度はデータの背後にある「枠組み」を表すわけではありません。ジュディア・パール博士は「構造モデル」と「回帰モデル」を明確に区別されていますが (Pearl, 2009), 「構造 (方程式) モデル」に強い意味でのデータ生成の意味づけを与え、むしろ「枠組み」としての用法を採っています。「デザインベース」に対して「モデルベースの推測」と言う場合にも、特定の推測対象でデータ生成を物理的に制御するのではなく、仮定される不特定の推測対象からのデータ生成を確率分布で記述する態度を示す「枠組み」の用法です。

マイケル・ワイスバーグ博士は、科学におけるモデルを「具象モデル」(現実に物理的性質を模して作成される、いわゆる模型)、「数理モデル」(方程式; 回帰モデルもここに該当すると考えられる)、「数値計算モデル」(データ生成アルゴリズム) の 3 つに整理していま

す (Weisberg, 2015). このように分類した上で、それぞれの「モデリング」がちがった仕方でもどのように科学理論に貢献するのかの理論を打ち立てようとする、たいへんスリリングな展開を彼の本では楽しめるわけですが、印象的なのはそもそも「モデル」や「モデリング」の哲学が現在においても十分には発展していないという点です. 自らの興味の中心である因果推論分野を振り返ったとき、「因果モデル」についての理解をなおざりにして理論研究を指向することの限界を感じていたこともあり、「因果モデル」を「因果推論を支える存在論と認識論」(大塚, 2018)として大きく捉えた展開が、今後この分野でも本格的に求められて来るだろうと感じました. 「ルービンの因果モデル」(いわゆる潜在反応モデル)と「パールの因果モデル」(構造方程式モデル)は有名ですが、「 $Y(0), Y(1)$ 」と「 $f_j(\text{pa}(X_j))$ 」とか「割り付けメカニズム」と「グラフによる独立性条件」という象徴的だけれどやや表面的な記法や定理のちがいではなく、なにが存在し(実在性)、なにが原理的にデータから分からないのか(識別性)を整理する枠組みとしての「モデル」の理解が必要となる時代が、すぐそこに来ているように思えるのです. たとえば、臨床試験のエスティマンド. ここでモデル理論は机上の空論ではなく、それを通した科学的コンセンサスの基礎づけへの要求として位置づけられることで、現実に直結する実験研究であるからこそ、その議論に指針が与えられるように思います. 一例として、ジェームズ・ロビンス博士によれば「臨床試験」は単なるリサーチデザインではなく、現実医療の「(具象)モデル」として推論に役立ち、さらにエスティマンドの枠組みさえ与え得ると言えましょう (Robins, 1986). このように因果推論だけ眺めてみても、データを扱い科学的推論を行う諸分野、特に統計学やデータサイエンス分野の中で、計量生物学・生物統計学は「モデル」の哲学の発展が最も期待される分野のひとつなのかもしれません. (なお、「因果モデル」はワイスバークの3つのモデル分類には上手くはまらないように思われますが、別のクラスとして紹介されている「特定の対象を持たないモデル」を指向するうちのひとつ「仮説的モデリング」(現実に存在しないものを対象とする)という営為に近いように思われ、このギャップに関しては今後の理論展開が期待されます.)

さて、計量生物学の未来に向けて、最後に「生物統計家のモデル」について考えたいと思います. 言うまでもなくこの「モデル」は日常的な意味での「理想」や「模範」であり、ワイスバーク分類や科学的モデリングからなにか類推することに意味は無さそうですが、「枠組み」と「データ生成」という比喻は、研究者どうしお互いが「モデル」とする生物統計家像を披瀝し合う場合に有用かもしれません. 各人がどのような世界観や価値観を持っているか(枠組み)が、生物統計家としての自己実現(データ生成)の方向性に大いに影響しそうだと思われるからです. サンダー・グリーンランド博士は「統計モデルで現実を十分に記述し尽くすことはできないので、良くて推論に論理的に一貫した態度を与えるに過ぎない」ことや、モデルがあたかも現実に存在するような「物象化」が科学の世界でも認知バイアスを惹起することを指摘していますが (Greenland, 2017), これは日常の「モデル」に対しても同様かもしれません. 人は自分が歩んできた道に依存して行動や思考の

原則を決めてしまうもの。科学や統計的推論でも同様だから統計モデルによって必要以上に補強されてしまう認知の歪みに意識的であるべきだとグリーンランドは言います。日常でも、地に足をつけ（過度に「モデル」志向的にならず）、研究者仲間からよく学び、現実に根ざした理想主義に燃えた「生物統計家のモデリング」を実践していければというのが、研究者としての私のささやかな抱負です。

Gelman, A. & Hill, J. (2006). *Data Analysis Using Regression and Multilevel/Hierarchical Models*. Cambridge University Press.

Greenland, S. (2017). *American Journal of Epidemiology*, 186: 639–645.

大塚淳 (2018). 日本行動計量学会第 46 回大会抄録集. SB17-1.

Pearl, J. (2009). *Causality*, 2nd Ed. Cambridge University Press.

Robins, J. (1986). *Mathematical Modelling*, 7, 1393–1512.

Weisberg, M. (2015). *Simulation and Similarity: Using Models to Understand the World*. Oxford University Press. 松王政浩, 訳 (2017). *科学とモデル—シミュレーションの哲学入門—*. 名古屋大学出版会.