

計量生物学を理解したいと思って毎日挑戦しています。

長島 健悟 (統計数理研究所)

大学 3 年生の時に研究室を選ぶことになり、選択できる研究室の中では生物統計学が一番面白そうだなと思い、この世界に入門しました。物事が長続きしない性格で、多くのものを途中で投げ出してきましたが、生物統計学は飽きずに続けています。計算とコンピュータが好きなため、数理的な研究やソフトウェア開発に触れられる点が特に魅力的だと思っています。生物統計学では、未知のものを測る方法を考えるという事が根底にあると考えています。時にはどう手を付けてよいのか全く分からない事や挑戦的な課題がたくさんあり、これを解くツールを作っていく工程は楽しいです。

数理的な研究を行うためには数学を学ぶことが必須です。数学の面白さは、さっぱり分からない事であっても、ずっと考えているとふとした瞬間に分かるようになっていくところだと思っています。全然分からなかった本を 1 年後に読み返すと、全然違う景色が見えるといった事がよくあります。ゴールを定めずに本を読み切るまでコツコツと読んでいくことと、研究に使う道具を論文などの成果としてアウトプットにするときに、その時点でできる限界まで考えることを両方行うのが良いと思っています。面倒ですが、数式を操作したり、簡単な例を作ってみたり、数値を入れて具体例を計算したりする地道な作業も重要で、長続きしない人には酷なことかもしれません。新しい領域を知る事も重要で、基礎的な本を手にとってみたり、新しく出版された論文を眺めたりすると、意外なつながりが見えることもあります。完璧を目指しても無理なので、毎日挑戦することが良いと思います。

新たな手法を作ったとしたら、その手法をなるべく簡単に使えるソフトウェアを公開することが理想だと思っています。*Biometrics* や *Statistics in Medicine* などを含む多くの学術雑誌の投稿規定には、提案手法を実装したソフトウェアの公開を推奨するという記載があります(これとは別の考え方で、再現性のためにシミュレーションのコードの公開を指定している学術雑誌もあります)。論文と同時にソフトウェアを公開するとフィードバックをもらう事や引用につながる事があり、ソフトウェアの改善や研究者間のコミュニケーションにも役に立つこともあるのかなと思っています。個人的には 2010 年頃からは研究目的で R を良く使うようになりました。R はフリーの統計ソフトウェアとして最も普及したものの一つだと思います。2010 年代前半には統合開発環境の RStudio が普及し始め、R のパッケージ開発の効率が格段に上がりました。最近 R パッケージがついている論文を目にすることが増えています。最近正式リリースされた RStudio Cloud は、ネットワーク環境とブラウザがあれば R と RStudio をインストールする必要なしに R パッケージの開発環境が手に入ります(臨床研究のデータ解析をプロジェクト単位で管理したり、授業用に使う環境のコピーを学生に配布したり、色々な活用方法があります)。情報技術は日々進歩しており、一つのものに依存せずに新しい情報を毎日収集していくことが重要ではないかと思っています。

最近ではメタ科学(またはメタ研究)に少し興味を持っています。メタ科学とは、科学そのものを科学的な方法で研究するというものです。例えば、少し前に話題になった、統計的有意性の誤用や研究再現性の話には、あるメタ研究 (Ioannidis JPA. Why most published research findings are false. *PLOS Med.* 2005; 2(8): e124. ) が影響を与えていました。我々が日ごろ使っている方法を、いつもと違う広い視点で調べ

てみると、また別のものが見えてくる事があるのかもしれない。数年前に主要な生物統計学の雑誌に出版されていた論文をテキストマイニングしたことがあり、どのような方法がよく研究されていたのか、年代による傾向があるのかなどを検討してみたことがあります。なかなか面白い結果でした。