

## わがままな鳥類研究者のための統計入門書の紹介

- 生物統計学入門 (新版) 石居 進 1992年 培風館 定価2300円
- すぐわかる統計処理 石村貞夫 1994年 東京図書 224p. 定価2200円
- パソコンによるデータ解析 (ブルーボックス) 新村秀一 1995年 講談社 318p. 定価800円
- 生物学を学ぶ人のための統計のはなし 粕谷英一 1998年 文一総合出版 199p. 定価2400円
- らくらく生物統計学 足立堅一 1998年 中山書店 247p. 定価3200円
- 生物統計学 Sokal & Rohlf (藤井宏一訳) 1995年 共立出版 定価8400円

学術論文には統計処理がつきものです。これが割とややこしいので、鳥の観察データなどかなり蓄積できているのに、なかなか論文をまとめるまで行き着かないという会員も多いでしょう。第一に、どのような統計法があってどれを使えばいいのかがわからない。統計処理の方法を教えてくれる人が身近に居る人は幸せでしょう。多くの人はそれほど恵まれては居ませんから、教科書や入門書に頼ることになります。ところが、「統計学入門」などという本には、たいがいややこしい数式や理論がずらりと並んでいて、10ページくらいでギブアップすることになります。実際のデータ処理に使える統計的方法の原理と使用法を、高校数学程度の知識で理解が可能なように解説してある本という、虫のいいことを誰でも考えるでしょう。

「生物統計学」とか「生物学を学ぶ人のため」とかタイトルのついた本は数多く出ています。以前であれば、数学が苦手なことが生物学を選ばせた大きな理由だという研究者も多かったので、「生物(学)～」というタイトルの裏には、「数学の苦手な人のための～」と言う意味を含んでいるように思っていました。もちろんそうではなく、生物学分野(農学系、医学系等含んだ広義の)は統計処理を非常によく使い、また、他の分野とは少し異なった統計処理手法を必要としていることが、多くの生物統計学の本が出版されている理由です。でも、このように多くの本が出てきても、救われない人も多く、かえって苦労が増えるだけということもあります。

平均的な鳥の研究者は(「平均」は集団の代表として不適当という指摘もすでに多くの統計本に見られます)統計を苦手としています。鳥の論文でも統計的記述や検定を必要とする場合が必ずありますから、電卓かポケコンで平均値、標準偏差を計算し、 $t$ 検定か $\chi^2$ 検定を行い、回帰曲線を求めます。でも、レフリーは「データは正規分布しているか」とか、「回帰の検定はしたのか」とか、いろいろとクレームを付けてきます。こうなってくると、統計処理についてのレフリーの言い分を理解する知識が必要となり、また、そのクレームによってはもはや電卓では追いつかない事態にもなります。最近パソコンが普及し、統計ソフトを使ってデータ処理ができる環境にある幸せな人たちも現れました。ところが、データを入力すれば勝手に計算してくれるのは有り難いけれども、出力された結果のどこをどう読めばいいのかわからないという事態も現れます。さらに、「統計処理、統計処理」という強迫観念(?)から、検定結果だけを気にするような人も増えて来ます。適切に統計処理がなされた論文ができあがるには、適切な統計法の選択、結果の適切な解釈が必須です。そのためには、「統計的方法の理屈」(粕谷 前掲書)を知らねばなりません。統計学の訓練を受けていないものが、手軽に「統計的方法の理屈」を身につけるハウツー本はあるのでしょうか。適当な解説書をと望む鳥類研究者として、1) 使用する方法の指示と使用上の注意がほしいだけで、今更数学を勉強したくない、2) 数式大嫌いだが文章での論理展開には十分ついていける、3) ポケコンはもちろん持っているし、パソコンを使用できる場合もある、という条件の、極めてわがままな人を想定します。そんな

人には横に先生がついていて疑問点にはすぐ応えてくれるというのが一番理想的でしょう。でもそれは叶わぬ夢。そこで、ぴったりの本があればというのが、鳥の研究者の（だけではないけど）願いでしょう。もちろんありません。そのかわり、こちら当たりを丁寧に読んでいけば、統計法の選択とそれらに関する理屈を理解する助けになりそうな本をいくつか紹介します。

鳥の研究を計画するときや論文をまとめるときには、次の3つの難問にぶち当たるに違いありません。

- 1) どのような検定法を使うか
- 2) 統計ソフトの出力結果の読み方がわからない。
- 3) 使用する方法にはどのような制約があるか。どのような注意が必要か。

最初の検定法の選択については、すでに多くの人が石居進の「生物統計学入門」を利用していることでしょう。この本はデータの尺度水準（類別データ、順序データ、間隔データ、比率データ）に応じて、用いる統計法とその原理を解説しています。多くの方法を網羅していて、発行年は古いけれども定評があって、今でも版を重ねています。各方法毎に計算法と例題が示してあるので、これを参考にしながら、読者は自身のデータで検定を行うことができます。少数例のデータでの検定を行うことの多い、行動生態学でよく用いるノンパラメトリック検定法を多く紹介しています。これがあればポケコンでの計算もできて、大量データでなければ統計処理は済んでしまいます。自分で手計算でできないという環境の人にはこの本しか選択の余地はありません。でも、わがままな読者は「統計ソフトがあるから計算過程はパスでいい。使うときの注意だけわかればいい。」と、言い出すかもしれません。

石村貞夫の「すぐわかる統計処理」は、石居の本と同様に多くの統計法を紹介しています。この本の特徴的なところは、データを配列パターンによって分類していることです。複数の変数の配列かどうか、多群の比較かなどに基づいてデータの種類を15のグループに分け、それぞれに適用される統計法を紹介しています。石居の本に比べるとよりマニュアル的で、より視覚に訴える傾向が強く、記述も思いきり簡単にして、とにかくサンプルデータを実際に処理してみようという仕組みになっています。ここでは、コンピュータを用いてデータ処理することが前提であり、コンピュータ計算結果の出力に従って説明がされているので、結果のそれぞれがどのような意味を持つのかを理解するのに便利です（計算ソフトによっては表示されないものも多い）。こういうデータにはこの方法を使うのかということをお納得するのは、石居の本より早いかもしれません。ただ、方法の原理の説明は最小限にとどめてあるので、詳しくは別の教科書を参照する必要があります（同じ著者の「分散分析のはなし」「統計解析のはなし」「すぐわかる多変量解析」など（いずれも東京図書）を勧めます）。

徹底的に、パソコンを使った統計的方法の取り扱い説明書になっているのが、新村秀一の「パソコンによるデータ解析」です。この本では、SPSSやStatisticaなどの統計ソフトを使ってサンプルデータを実際に処理しながら、出力結果の読み方、統計法（主にパラメトリック検定と多変量解析）の解説をしています。必ずしも読者の持つソフトとは同じというわけではないでしょうが、出力結果はどれも似たり寄ったりなので、結果の読み方の解釈の手ほどきになります。ただ、本書ではいろんな場合の処理を解説するために、不適切なデータについても無理矢理統計処理を行っている場合もあることも考慮しておかねばなりません。また、ノンパラメトリック検定についての解説はありません。同じ著者の「パソコン楽々統計学」（講談社ブルーバックス：1500円）にはWindows版StatisticaのCD-ROMが付録についているので、実際にサンプルデータを処理しながら統計法をマスターできます。ただし、あくまでも演習用です。

著者は検定に入る前に、箱ヒゲ図や散布図を使ったデータの探索を行うことの重要性を強調しています。データ量が増えるほど、入力ミスなどによる外れ値が増える危険も高くなります。このような外れ値は結果に大きく影響します。その結果、計算過程での論理矛盾が起これば、コンピュータが誤

りを見つけてくれますが、そうでない限りは誤りのままで計算してしまい、結果を見ただけではその誤りに気づきません。これはかなりこわいことです。箱ヒゲ図や散布図を描いてみて外れ値を見つけ出すことは、統計ソフトを使っている限りはほとんど手間がかかりません。結果として外れ値の原因が入力ミスではなかった場合でも、外れ値の性質を検討することで新しい生物学的な発見がもたらされるかもしれません。コンピュータを使う利点はここにあると言えます。

3番目の統計処理に関わる制約の問題で大きいのは、1) データの分布の問題(正規性と関連); たとえば、「正規分布していると確認してt検定をしているのですか」という指摘、2) データの種類の問題; 「上中下というような類別にするなら、順序データとして処理した方がいいよ」とか、3) 多重性の問題; 「多くの群で総当たりのt検定を繰り返すときは要注意です」、4) 検出力の問題; 「2つの方法で検定したら一方だけに有意差が出た」、などです。コンピュータが計算してくれるとしても、どのような処理をするか、どのような結果を出力するかなど、使用者が決定しなければならないことが多くあります。例えば、両側検定か片側検定か、変数変換が必要かどうか、 $\chi^2$ 検定に十分なサンプル数か、などなど。これら、方法の制約に関わる問題はまさに「統計的方法の理屈」を知らなくては理解ができません。そのためにはある程度の統計理論の説明が必要です。統計的方法を紹介している本ではスペースの関係から、これらの制約、使用条件などを丁寧に解説しているものはありません。読者からすると、アンダーライン付きでの説明がほしいところです。

粕谷英一の「生物学を学ぶ人のための統計のはなし」は、統計処理を行う時によく見られる誤りを指摘し、なぜそれが誤りなのかを説明し、どのように解決するかを指し示す本であると言えます。著者は生態学研究者であり、取り上げられている間違いの例は、著者が実際に見聞したもので、鳥の研究者にも共通する問題があり、かなり役に立つ本だと言えます。この本には数式はほとんど出てきません。それでも適用方法の誤りやその理由がよく理解できます。この本のもう一つの特徴は、架空の大学の研究室における、教官と院生の対話形式で書かれていることです。統計的方法に関する問答集のようでもあります。それで、読者は横にいる先生に教えてもらうような感覚で説明を理解することができます。ただ、最近のメディアの多様化の流れで、会話文のほうが理解が早いという人たちが増えたのかもしれませんが、評者にはなんだかまどろしくて、通常の記事の方が要点をつかむのが楽なように思えます。また、著者独特のエスプリに慣れていない人は、混乱するかもしれません(副題の「きみにも出せる有意差」、特に、付録にある「君にもできるごまかし」など)。

足立堅一の「らくらく生物統計学」は統計学的な考え方、論理構築、思想について述べた本です。著者は薬理学関係の研究者で、医薬データの解析を専門としているので、大量データを処理するパラメトリック検定に関する記述が大部分を占めています。

章のタイトルは以下の通りです。

- 1) 母集団と標本
- 2) SD(標準偏差)とSE(標準誤差)
- 3) 仮説検定と2種類の過誤( $\alpha \cdot \beta$ )
- 4) 標本数と統計的結論
- 5) 分割表における $\chi^2$ 検定とU検定
- 6) 回帰と相関
- 7) 多重性に関する問題点
- 8) 平均値・分散の意義
- 9)  $E(aX \pm bY \pm c)$ と $V(aX \pm bY \pm c)$
- 10) 数学的モデルとしての分布
- 11) 検定と推定
- 12) t検定とU検定

この本の重要な点は、統計結果の解釈に見られるよくある思いこみに対して警鐘を鳴らしていることです。統計処理を気にするあまり、有意差をだすことだけに汲々として、データの示す本質的な問題を見忘れてしまうという検定至上主義に陥ることがよくあります。著者はこの問題を「significantosis (佐久間 (1977)「薬効評価 I-計画と解析」, 東京大学出版会:「有意差病」と訳する)」と「non-significantosis (著者による造語):「有意差なし病」(?)」の両語を使って説明しています。significantosisは統計的有意性とその差のもつ生物学的意義の混同に基づいています。もともとわずかでも差があるものであれば、標本数さえ増やせばそのうちに有意差が出るものです。この病気の症状は、その程度の有意差が生物学的な意味を持つのかどうかという検討なしに、有意差が出たことで安心してしまうことです。また、相関が有意であるという結果を因果関係の証明と短絡させてしまうこともsignificantosisの典型的な症例です。逆に、non-significantosisの方は、有意差がないことを両者が同等であるという結論に結びつけてしまう傾向で、こちらたびたび見かける症例です。有意差が出る出ないが標本数に依存することを忘れていることに起因します。言われてみると、このような病気は広く蔓延しているように感じます。この本には各章の冒頭に診断テストがついています。そのテストを試してみたところ、評者も軽い症状を呈していたようです。

各章で取り上げているテーマについても実際に現場(医学的データの統計処理)で生じる問題について詳しく解説して対処法を示しています。その分、内容がやや専門的になっていますが、数式などは極力少なくしているので、統計学の入門書を読んでいるときのように途中で投げ出したいくなることはないでしょう。一通り読んでみると、パラメトリック検定の原理について理解できたような気になります。理系の大学出身で統計数理の教育を受けた人には常識的なことばかりのように思えますが、統計処理に関する病気が広く蔓延していることを考えると、将来必要となるという意識を持って受けられない限り、教育は無になるということの実証かもしれません。評者は学部時代に統計を利用するための教育を選択せずに大学院に入り、そこで必要上やむを得ず、スネディカーとコクランの名著「統計的方法」をテキストに勉強しましたが、統計検定の仕方だけとりあえずマスターして、原理の理解は先送りのままでした。そのような人に、何のために統計処理をするのかということへの注意を喚起してくれる本だと思います。ただ、数値などのミスプリントがいくつか見られるので注意が必要です(大きなところは、117ページの外れ値の説明、157ページの幹葉図と箱ヒゲ図の指示など)。

以上が費用もお手頃で、比較的易しい解説書です。もし、費用は問題ではない、少々難しくても大丈夫という人には、迷わずソーカル&ロルフの「生物統計学」をお勧めします。

江口和洋(九州大学理学部生物学教室)