

無料統計アプリ「EZR」でらくらく本格統計解析

西川 洋史

埼玉県立進修館高等学校 nishikawa.hirofumi.01@st.spec.ed.jp

キーワード：分散分析，多重比較，生物統計学

(受付：2024 年 8 月 27 日)

I. 生物学における統計学の必要性

生物学を志す生徒（学生）の中に以下のような考え方をしている者は意外といっているのではなかろうか？

- ・学問的な志向は理系である。しかし、数学は苦手だ。暗記は得意だ。じゃあ、生物系にすすもう。生き物大好きだし。
- ・効果？たくさんデータを集めて差がある雰囲気のグラフが出ればいいんじゃない？10%程度の差が出たら効果ありと判断していいかなと。
- ・最後はKD（感と度胸）検定。

何を隠そう、かつての筆者であるが、生物学の探究（研究）における現実とは真逆である。例えば『ある物質 A をガン細胞に投与する実験を十数回繰り返し、平均値を求めたところ、ガン細胞数が 10%も減った。この結果から、この薬はガンを抑える効果があるに違いない！とゼミで嬉々として発表すると、教授に怒られる』のが落ちである。客観的に見れば「ばらつきが大きい中でたまたま減ったように見えるのでは？標本サイズと標本数は妥当なのか？ANOVA はやった？解析アプリケーションは何を使った？」という疑問がわいてくる。『内下線部「も」「違いない」というのは、ある事象について感覚的に物事を判断したり、思い込みが強すぎたりすることを表している。探究活動に情熱的に取り組み、印象的な結果が出るほど判断は狂いやすくなるものだ。とくに生物のように条件を制御することが難しい領域ではなおさらである。そのような不安定でばらつきの多い中で一般的な傾向を求めるには、統計解析がどうしても必要になる。それでは、中高生が ANOVA のような高度な統計学を必要とする場面はあるのだろうか？近年、中高生の理科系の部活やサークルの研究発表が活発になっているが、統計学的手法で解析した結果が数多く発表されている。会場でしばしば出るのが「適切な統計解析アプリケーションがない、あっても高額で買えない」という話題である。この問題を解決する方法として無料の統計アプリケーションを紹介させていただきたい。

II. 生物学で必要となる統計と解析アプリケーション

生物学でよく使われる統計解析には 2 群間比較であれば t 検定がある。いわゆる条

件 A と条件 B で結果に有意差があるかどうか(たまたまなのか, そうではないのか), という検定方法である。この統計解析は Microsoft Excel でカバーすることは可能だ。しかし, 3 群間以上の比較では t 検定を繰り返す (A と B, B と C, C と A について t 検定を繰り返す) ことは有意水準を下げるため実施することができない。3 群間以上の比較では一元配置分散分析を行った後に多重比較 (どこの群れに差があるのかを解析する) を行う。よく使われるものに Tukey 法, Dunnett 法, Bonferroni 法, Holm 法, Williams 法がある。また, 標本が正規分布していない場合は, Kruskal-Wallis 検定後に, それ用の多重比較を行う。例えば生育環境を同一とし, 5 種類の肥料 A~E をある植物に与えたとき, どの肥料が最も効果的かを判断したいとする。この場合, 成長量の平均値が最も高いものを選び「肥料 A が最も良い」と結論づけるのが従来の探究活動でのやり方だろう。しかし, 分散分析後に多重比較をすると「肥料 A と肥料 B では肥料 A のほうがやや成長量が大いだが, この 2 つには有意差がない。よって肥料 A と肥料 B のどちらでもよいが, 〇〇の理由から肥料 B のほうが使いやすいだろう」などと, より深い考察が可能となる。このような統計解析を実施できる統計アプリケーションには SAS, S-PLUS, SPSS, JMP など挙げることができる。いずれも市販されているが, 高額 (20 万円を超える) なので中高生が授業や課外活動で利用することは難しい。

Ⅲ. EZR は無料かつプロ仕様

統計アプリケーション「EZR」(Easy R) は上記のような導入コストの問題を解決できる。EZR は R コマンダーを基に開発された無料の統計解析アプリケーションである。R コマンダーとは, 統計解析に特化したプログラミング言語である「R 言語」を用いた解析を, マウス主体の操作によって行うことを可能にしたアプリケーションである。R コマンダーの利点は R 言語の習得なしに統計解析ができることにある。そして EZR はこの R コマンダーを医療統計用にカスタマイズしたものだ。つまり Microsoft Excel 操作に近い使用感で R 言語による統計解析が可能となる。開発者は自治医科大学附属さいたま医療センター血液科の神田善伸教授であり, 学術論文にもなっている信頼のおける統計解析アプリケーションである (Kanda, 2013)。

EZR では具体的にどのような統計解析ができるのか? t 検定はもちろん, 生物分野の探究活動で最も使われそうな分散分析や多重比較も網羅されている。多重比較については, 用量依存性を前提とする結果の分析に適用される Williams 法が使えないのが惜しいのだが, 十分だろう。さらに正規分布に従わない 3 群間以上の多重比較もでき, 正規分布したデータが集まらないことが多いであろう中高生の研究では, 大いに活用できそう。

Ⅳ. 入手方法と動作環境

EZR は自治医科大学附属さいたま医療センター血液科の Web サイトからダウンロードできる (<https://www.jichi.ac.jp/saitama-sct/SaitamaHP.files/statmed.html>) (2024 年 8 月 26 日確認) (図 1)。

Windows OS では, Windows XP / Vista/7/8/ 8.1/10/11 で動作する。Mac OS では



図1 EZRがダウンロードできるWebサイトのトップ画面

インストールの方法がやや煩雑になるが、Mac OS X 10.5～10.10で動作する。無料で始められるLinux OSではUbuntu 11.10～20.04で動作確認がされている。いずれのOSでもインストールの方法や動作環境の詳細についてはWebサイト上に掲載されているので参照するとよい。何分、無料であることから気兼ねなく試せるというの也有りがたい。

V. ソフト導入もICT環境整備に該当

学習指導要領には「各科目の指導に当たっては、観察、実験の過程での情報の収集・検索、計測・制御、結果の集計・処理などにおいて、コンピュータや情報通信ネットワークなどを積極的かつ適切に活用すること。」とある（文部科学省，2018）。統計解析は、結果の集計・処理に該当する。学校教育に即したICTのソフト面での環境整備という観点からもEZRは適切だろう。現在、生物教育実験における結果の集計・処理は単なる平均値比べから本格的な統計解析へと前進させる条件が整っていると言えよう。

文献

Kanda, Y. 2013. Investigation of the freely available easy-to-use software ‘EZR’ for medical statistics. Bone marrow transplantation, 48(3), 452-458.

文部科学省 2018. 高等学校学習指導要領（平成30年告示），p.186. Retrieved from https://www.mext.go.jp/content/20230120-mxt_kyoiku02-100002604_03.pdf (2024年8月26日閲覧)