

植物防疫基礎講座：

生物統計処理方法の注意点

中央農業総合研究センター ^{みつ}光 ^{なが}永 ^{たか}貴 ^{ゆき}之

はじめに

近年、農業研究における統計処理の必要性はますます増している。例えば、和文、英文を問わず、論文投稿の際にはデータに対して適切な統計処理を行っているかどうか論文受理の際の大きな判断基準になっているばかりではなく、農業判定試験などについても適切な統計処理が要求されていることから明らかである。

しかし、現場の研究者の多くは、大学時代、あるいは大学院時代においてさえ、体系的な統計手法についての講義を受けた経験がほとんどないのではないだろうか。これは筆者自身の経験であるが、1980年代中ごろの学部生の時代に、確かに“確率論・統計学”という講義を受講した記憶はある。しかしながら、その講義の内容は、“ポアソン分布の典型例は19世紀のプロシア軍で1年間に馬に蹴られて死んだ兵士の数である”といった豆知識としては面白いかもしれないが、どのように実用化すればよいのかわからない知識である場合や、あるいは“二項分布を例として中心極限定理を証明する”といった重要ではあるが数学的に偏った講義がほとんどであったことしか覚えていない。このような勉強を続けた挙げ句、1年の終わりになってようやく付け足しのよう¹に t 検定、カイ二乗検定、直線回帰などの古典的統計手法のいくつかを駆け足で紹介されて1年間の講義が終了するといった具合であった。その結果、“アノバ(ANOVA、分散分析)とは何か?”と聞かれた際に“原生動物の一種ですね”と真顔で答える大学院生ができあがるのである。もっとも、当時はPCがほとんど普及しておらず、現在用いられている統計手法は手計算するには膨大な計算量を要求するため却って当時は実用的ではなかったとする擁護意見も存在するが、米国で古典的な統計手法についての教科書である *Biometry* の第2版が1981年に出版されていることを考える限り、日本の教育システムがお粗末であった感は否めない。現在でも、応用動物昆虫学会などで大学院生から聞いている限

りでは、大学の講義は多少の例外があるものの本質的にはあまり変わっていないように思える。

大学、大学院で統計についての体系的な知識を得ることができなかった場合、さしあたって必要な統計学の知識を得る手段は第一に自学することである。最近では、自学しやすいようになるべく取り組みやすく平易に書かれた教科書(例えば、粕谷, 1998)も存在しており、米国でよく用いられている統計学の実用的な教科書が翻訳されており(例えば、SOKAL and ROHLF, 1983)、さらにはインターネットなどでも真贋はあるが容易に情報を得ることができるため、自学できる環境は以前よりも整っていると考えられる。しかし、大学受験後十年以上が経過し、数学などいまさら見るのも億劫になっている人が多いことも事実であろう。まして、日々の雑用に追われ、空いた時間で実験をせざるを得ない状況の中で新たに統計の勉強時間を作り出すことは決して言うほど簡単なことではないと思われる。

大学などで適切な授業を受けられず、現在自学のための時間も取れない場合、残された手段はつてを頼って統計に詳しい研究者に相談することであろう。そして、現状では極めて多くの人達がこのようにして急場をしのいでいると考えられる。このこと自体は決して悪いことではないのだが、一つだけ大きな問題が生じる場合が多い。それは、ほとんどの場合、相談者は“データを取った後に”相談をするので、データのとらえ方や取り方そのものの不備は相談をした時点で初めて発覚するということである。そこで、本稿では、統計があまり得意ではない研究者の方を対象に最低限知っておくべきデータのとらえ方や解析についての考え方と、その理由について簡単に述べていきたいと思います。実験設定、試験設定をする前にこれらのことを参考にいただければ幸いです。

I 5%水準とは何か?

統計があまり得意ではない人でも、5%水準という言葉は聞いた記憶があるであろう。これは、言ってしまうと、ある集団と別の集団が違うと判断できるかどうかの尺度基準である。より具体的には、得られたデータから推測する限り、集団同士が同じである確率は0.05であるということを示している。言い換えれば、5%水準を

Notes of Statistical Processing Methods. By Takayuki MITSUNAGA

(キーワード:生物統計, 5%水準, 事後検定, 分散分析, 計画的比較)

基準とした際には、本当は同じであるのにランダムなデータのずれによって違うと判断される確率（第一種の誤りの確率）は20回に1回であるということになる。ただし、これは、集団と集団が違うか否かという質的なことにのみ言及している基準であることは留意しておかなければならない。統計の教科書には、大抵何ページも割いてこれについての詳細な考え方が述べられているが、それぞれの教科書を参照していただきたい。ここで述べたいのは、少なくとも現在は、この5%水準を満たしているか否かということが、ある処理区と別の処理区で違うか違わないかのデファクトスタンダードになっているということである。だからこそ、ほとんどのデータ解析の際に、 p 値が0.05を超えるか否かでテンションが上がったり下がったりを繰り返すのである。重要なのは、5%水準で有意性を判断するということは論理的必然性というよりは、むしろコンセンサスであるので、異を唱えても始まらないということである。当たり前だと思われる方は、読み飛ばしていただきたいが、今でも、論文によっては p 値が0.051のときに“少々有意であった”とか、“有意傾向があった”といった、あきらめきれない表現が多々見られる。このような表現は賛否両論あるかもしれないが、いたずらに査読者の印象を悪くするおそれがあるので使わず、すっぱりとあきらめるほうが望ましいであろう。

5%水準についても一つ確認していただきたいのは、先にも述べたように、この基準は単にある集団（処理区）と別の集団（処理区）が同じか違うかを判断する基準に過ぎないということである。言い換えれば、 p 値が非常に小さいことと集団同士が大きく違うことは等価ではないということである。これも当然と思われる方は読み飛ばしていただいて結構なのだが、こちらもいまだに“より有意であった”とか“大きく有意であった”といった p 値が小さいことを喜ぶ表現がよく見られる。 p 値が小さいことが集団の違いがより大きいことを示しているわけではなく、単に集団同士は違う確率が高いだけであるということが実感できていないのであろうが、このような表現も論文などの発表の際には軋轢を呼ぶ可能性が高いので使わないほうがよいであろう。

II 事後検定とはどのような作業なのか？

実験、あるいは試験とは、簡単にまとめると、複数個の処理区を考えて処理区間の違いを確認し、そこから有益な情報を引き出すという作業である。例えば、ある病原菌について、複数個の殺菌剤を用意し、その病菌に最も良く効く殺菌剤を見つけるために実験を行うわけであ

る。このような場合、どのような薬剤が効くのかについて研究者はよくわからないことがほとんどであるので、何種類もの薬剤を用意して病菌に処理することになる。このようにして得られたデータについては、ほとんどの場合分散分析などで実験系全体の有意性を見た後、個別に各薬剤間で有意性について見ていくことになる。そして、最も良く効く薬剤が決定されて実験終了となる。この最後の個別に比較している部分が事後検定と呼ばれるもので、ほとんどの研究者はこれを見たいために実験を行っているのである。近年では、このような一連の流れそのものについて本当にこれでいいのか、という議論が起きていることは事実である（多重性の問題として、永田・吉田, 1997）が、少なくとも現状では上記のフローが一般的な流れであろう。

ここで注意すべきことは、先ほど紹介した5%水準というルールがこの一連の作業に大きく関係してくるということである。例えば、最初に行う分散分析（あるいは他の検定でも同じ）で実験系全体の有意性を評価する場合には、実験系全体での処理区間の違いを5%水準で評価する。これは簡単に言えば実験全体を見たときにある処理区と他の処理区が違うか否かということの評価しているのである。このような実験系全体での第一種の誤りの確率のことを **family-wise** な第一種の誤りの確率と呼ぶ。一方、事後検定においては、実験内の個々の処理区間での総当たりとして個別に検定を行う。この場合の第一種の誤りの確率は **comparison-wise** な第一種の誤りの確率と呼ばれる。

一般に5%水準のルールにおいては、重視されるのは **family-wise** な第一種の誤りの確率であり、**comparison-wise** な第一種の誤りの確率は **family-wise** にコントロールされなければならないとされている。例として、5区からなる実験系を考えてみよう。最初に全体を分散分析で検定を行った際に、5区が本当は同じなのにランダムなデータのばらつきによってどこかが違うと判断される確率は5%である。言い換えれば、このような実験を20回行った際に1回誤った判断をすることが予想される。今度は、事後検定として5区の総当たりで比較をすると、事後検定の回数は10回である（五つの中から二つを取り出す組み合わせの数）。もしも、事後検定の有意水準を5%のままに設定していたならば、5区からなる実験を2回行うとどこかで間違いが生じてしまうことになる。したがって、事後検定の際には、全体の5%水準を守るために個々の検定は5%よりも厳しい有意水準を設定しなければならない。

以前、事後検定の定番であった DUNCAN の多重検定が

近年禁じ手になっているのは、この検定が個々の有意水準をコントロールできないためである。DUNCANの多重検定では、本来事後検定で行うべき厳しい有意水準よりはるかに甘い水準（つまり5%）で個々の対比較を行うために、有意差が出やすいのだが、これがまやかしであることは、有意水準の本質を考えれば明らかである。

以上のことから事後検定の際には、処理区が多くなればなるほど、各処理区間の違いを検出する有意水準は厳しくなり、結果として違いを統計的に見だし難くなるということがわかる。これを簡単に表現しているものがBONFERRONIの方法と呼ばれるものであり、これは、事後検定の際には、個々の対比較での有意水準を0.05/(対比較の個数)で設定するというものである。例えば、先ほどの例として用いた5区からなる実験系の場合、対比較は10通り存在したので、個々の対比較の有意水準は $0.05/10 = 0.005$ となる。BONFERRONIの方法は設定が簡単なのだが、処理区の数が多くなるにつれて判断基準が急速に厳しすぎるものとなるので、処理区数が大きいときには、より条件を緩和させた他の手法を代用することが多い（HOLMの方法など）。また、現在、分散分析の事後検定として広く用いられているTUKEYのHSD法などはJMPで採用されていることもあるが、総当たりの順番を考慮するなどにより、BONFERRONIの基準よりはるかに緩和されているため優れた方法である。

ここまで述べてきたことから、事後検定というのは、その性質上いかなる方法を用いたとしても処理区が増えるにつれて有意水準は厳しくなっていくことがおわかりいただけたと思う。したがって、実験設定をする際には、むやみに処理区を増やさないということが重要であることがわかる。つまり、自分が本当に見たいと思った処理についてのみ実験を行うべきであり、サンプルが余ったから、圃場が少々余ったといった安易な理由で処理区を増やすということはデータ解析の際に自分の首を絞めることになりかねないということを理解していただきたい。あくまでも理想に過ぎないが最もエレガントな実験系は無処理区と一つの対象区であり、事後検定はできるならば避けるべきであると筆者は思っている。

データ解析に用いる検定が分散分析であることまでがわかっている場合には、事後検定を減らす有効な手段として計画的比較（planned comparison）という手法があることを紹介しておきたい。日本ではあまりこの手法は用いられていないが、総当たりの数を減らすためには大変有効な手段である。詳しくは藤井（1983）、またはSOKAL and ROHLF（1981）を参照していただきたいが具体的には以下の手法である。例として、ある害虫に良

く効く殺虫剤をスクリーニングする試験をしたいとする。殺虫剤にはいろいろな種類があるが、これらの殺虫剤はピレスロイド系であったりIGR系であったりとグループ化することができよう。そこで、無処理区とピレスロイド3種とIGR3種を選んで実験を行い、その害虫の生存日数を測ったとする。通常ならば、分散分析の後に事後検定を行い、対比較は21通りになる（7個の中から2個を取り出す組み合わせ）。しかし、計画的比較では分散分析時に、ピレスロイド系とIGR系、あるいは無処理区とピレスロイド系といったようにグループ間で比較を行う。これにより、無処理区、IGR系に対してピレスロイド系だけが有意に生存日数が短かったならば、事後検定はピレスロイド系の3種、あるいは場合によってはそれに加えて無処理区の間に対してのみ行えば十分となる。この場合、事後検定の組み合わせ数は大幅に軽減される。計画的比較のためには、実験設定の段階で、処理区が適切にグループ化されるように過不足なく選択をしなければならないが、それが正確にできれば、大変強力なツールであると思われる。

おわりに

現在、農業研究に携わっている研究者は、生物統計について大変詳しい人と逆にあまり得意ではない人に二極化しているように筆者には思える。この知識の二極化は、主として大学、大学院での統計の接し方に影響しているのではないだろうか。しかしながら、筆者には、この知識の二極化と携わっている研究のレベルには現時点を見る限り、あまり相関があるとは思えない。確かに、生物学とその応用分野である農学のような現象科学においては、優れた研究というのは不断の思想的努力に加え、対象に対する深い洞察から生まれるものであり、それに比して数学の一分野である統計的な知識というのは必ずしも必要ではないかもしれない。そうした点において、統計というのは単なるツールであり、業績発表の際の縛りの一つであるので、しっかりとした論理、観察の基に研究を行ってさえいけば、統計のようなテクニックについては得意な人に丸投げしてしまえばよい、という意見にもある程度同意する点がある。ただし、統計手法を知るということは、エレガントな実験設定につながり、さらには事実をよりコンフリクトに表現してくれるということもまた事実である。したがって、いつまでも統計の丸投げを続けるよりは、少しずつでも自学することで統計に通じていくこともまた重要ではないだろうか。付け加えて、SOKAL and ROHLFは“生物統計に対する理解度は、数学的な知識との相関よりはむしろ研究者が

対象としている生物に対する生物学的な知識との相関の方がずっと高い”といったことを述べていることも紹介しておきたい。

本稿は、最初は事後検定について以前はよく用いられてきた DUNCAN の多重検定の欠点から始まって現在主に用いられている TUKEY の HSD 法などの事後検定について概説する予定であった。しかしながら、筆者が日々感じている研究者の間での統計知識の二極化がもしも正しいならば、このような概説は、詳しい人には周知のことであり、逆に得意ではない人には単なる豆知識になってしまうのではないかと思われた。そこで、統計が得意ではない人にとってはどのような情報が実際の研究シーンで最も即効性が高いかを考え、本稿の内容に至った。様々な事後検定について興味がある方には永田・吉田(1997)をお勧めしておきたい。また、全面的な書き直しとそれに伴う原稿の遅れにより、本誌編集部には多大なご迷惑をおかけした。この場を借りて改めてお詫びを申し上げたい。

筆者自身、大学の学部生の際には、冒頭で述べたように実用的な統計についての講義を全くとっていいほど

受けてこなかった。これは、今考えれば、教育システムの大きな不備であるとしか言いようがないことである。したがって、筆者も統計を丸投げする研究者になっていた可能性はあったのである。幸いなことに、筑波大学大学院で藤井宏一教授にご指導をいただき、生物統計学については講義というよりはむしろ日常の中で、言うなれば生活実感として接することができたのは筆者にとって望外の喜びであった。藤井教授には、ただならぬご指導をいただき、大きなご学恩を受けたが、統計の知識もまたその一部である。ただし、本稿にもしも不備な点があれば、それはすべて筆者に帰すことは言うまでもない。また、本稿を鈴木芳人博士にご査読いただき、大変有益なコメントをいただいた。最後になるが、ここに厚く御礼申し上げる。

引用文献

- 1) 粕谷英一 (1998): 生物学を学ぶ人のための統計のはなし, 文一総合出版, 東京, 199 pp.
- 2) SOKAL, R. R. and F. J. ROHLF (1981): Biometry, Freeman, New York, 887 pp.
- 3) 藤井宏一 (訳) (1983): 生物統計学, 共立出版, 東京, 447 pp.
- 4) 永田 靖・吉田道弘 (1997): 統計的多重比較法の基礎, サイエンティスト社, 東京, 197 pp.

好評発売中

農薬取締法令・関連通達集

(社)日本植物防疫協会編 B5判 261 ページ
価格: 1,050 円(税込) 送料 340 円

<掲載内容>

農林水産省・環境省・厚生労働省関連の農薬に関する政令、省令、告示、関連通知、その他省令を網羅

- ・ 農薬取締法と関連の政・省令を見やすく 2 列に表示
- ・ 農薬関連の告示を取締法に関連付けてレイアウト
- ・ 関連する通知文およびその他関連法令(抄)も掲載

農業関係者必携の 1 冊です。

