講 座

使える統計学基礎講座 (第4回)

—分散分析と共分散分析—

光永貴之

独立行政法人農業・食品産業技術総合研究機構中央農業総合研究センター

Lectures of Statistical Logics for Beginners (Part IV)

Analysis of Variance and Analysis of Covariance

Takayuki MITSUNAGA

National Agricultural Research Center NARO

1. はじめに

第三回では、回帰分析について検定法や結果の解釈の妥当性を中心にお話ししました。また、分散分析についてはその基本原理であるデータ分散の要因分散や誤差分散ごとの分割について説明しました。今回は前回の続きとして分散分析の検定原理について述べていきます。続いて、分散分析と直線回帰を組み合わせた検定手法である共分散分析の原理等について述べていきます。

第一回でもお話ししました様に、この講座はできる限り数学をいじらない様に心がけています。数学的な計算過程、証明に興味のある方は藤井(1983)などを参考にすることをお勧めします。また、分散分析やその応用である実験計画法については既に本誌上で講座として優れた解説が行われています(執行, 2002)。従って本稿では、データを取り扱う際の注意点や解析上の問題点等を中心にお話ししていきます。試験設計、論文作成などの際の参考にして頂ければ幸いです。

2. F 検定の考え方

第三回では、分散分析の原理とはデータ全体の

分散を実験要因の分散と誤差由来の分散に分割し、その比を取ることで要因の効果を検出するというものであるというお話をしました。具体的には、実験要因によって生じる分散が誤差由来の分散よりもある基準に従って大きければ、データのばらつきは実験要因に由来するものである＝要因の効果が認められると結論する訳です。この論法が成り立つためには、要因分散、誤差分散の大きさの比を客観的に判断する指標が必要になります。そこで、今回は最初にこの指標を考えていきましょう。

一般に、データから計算された、あるいは推定された統計学的な数値のことを“統計量 (statistic)”と呼びます。最も身近な例としては平均値、標準偏差などが挙げられます。統計量にはたくさんの種類がありますが、その中に“検定統計量 (test statistic)”と呼ばれるものがあります。検定統計量とはある仮定の下にこの統計量を計算した場合、得られた数値がその事象が観察される確率と対応するようなものを指します。この性質を利用して、データから計算された検定統計量の値から仮定が成り立つかどうかを判断することが統計的検定の原理です。分散分析では、検定統計量としてF統計量というものを用います。

ある現象が正規分布に従っていると考えましょう。ここで、データの背景には世界全体で同じ分布(母分布と呼びます)が存在すると考えますが、この母分布の性質を知るために、全数調査を行うことはほとんどの場合不可能です。例えば、スズ

平成 22 年 8 月 6 日受付

平成 22 年 8 月 7 日受理

Corresponding author

光永貴之 Takayuki MITSUNAGA

〒305-8666 茨城県つくば市観音台 3-1-1

3-3-1 Kannondai, Tsukuba, Ibaraki, 305-8666, Japan

E-mail : aeiou@affrc.go.jp

メのクチバシの長さを知りたいと思って世界中的ズメを捕まえて測ることはできませんね。そこで、一定数をサンプリングして得られたデータによって作られる分布（標本分布と呼びます）から世界全体の分布を推定せざるを得ません。母分布と標本分布ではその平均値の期待値は変わりませんが、分散は標本分布の方がちょっと小さくなるという性質があります。具体的には以下の式で表されます。

$$E(V) = \frac{n}{n-1} v \dots\dots\dots(1)$$

ここで、 $E(V)$ は母分布の分散（母分散）の期待値、 v は標本分布の分散（標本分散）、 n は標本分布のサンプル数を示します。この式から、サンプル数が増えるほど母分散と標本分散のズレは小さくなると予想されますが基本的には標本分散の方が小さいことが分かります。統計学的推定はこの母分布と標本分布のずれをどのように解決するのかを常に考えていることであると言えます。

母分布が正規分布の時、簡単な演算によって別の統計量も導けます。一つ一つのサンプルと平均値の差を取って二乗したものを足し合わせて分散で割ってやります。言葉で書いてもピンときませんが次の式で表されます。

$$\chi^2 = \sum_{i=1}^n \frac{(x_i - m)^2}{V} \dots\dots\dots(2)$$

これがカイ二乗統計量と呼ばれるもので、統計学的には極めてポピュラーな検定統計量です。“カイ二乗検定”という言葉に皆さんもどこかで接したことがあると思います。ここで、 x_i は i 番目のサンプルデータ、 m はデータの平均値を表します。この式を見ると分子はサンプルデータによって作られ、分母は母分散ですので母分布由来の統計量です。もう少し変形すると次の式になります。

$$\chi^2 = (n-1) \frac{v}{V} \dots\dots\dots(3)$$

この変形の意味することは、カイ二乗統計量は母分散、標本分散、そして標本分布のサンプルサイズによって値が決定されるということです。このことは私たちが集めたデータの統計解析を考える上では大変重要なことを含んでいます。なぜなら、検定統計量はサンプルサイズに影響されることを

示しているからです。このような検定統計量を計算する際のサンプルサイズの影響分（この場合は $n-1$ です）を“自由度 (degree of freedom, df)”と呼びます。自由度には様々な解釈があり、多くの教科書で色々な説明がされていますがなかなか分かりにくいものです。この解説を理解の一助にしてみてください。

さて、この章の目的は二つの分散の大きさを客観的に比較することです。このことは二つのデータ分布の分散が同じかどうか、言い換えれば比較したいデータ分布が同じ母分布からサンプリングされているかどうかを判断することと同じことです。つまり、二つのデータセットがあった時、(1)～(3)式の中にある V が同じかどうかを判断するのが目標になります。そのためにはカイ二乗統計量だけでは不十分です。そこで、次の式を作ります。

$$\frac{\frac{\chi_1^2}{n_1-1}}{\frac{\chi_2^2}{n_2-1}} = \frac{\frac{v_1}{V}}{\frac{v_2}{V}} = \frac{v_1}{v_2} = F \dots\dots\dots(4)$$

これが F 統計量と呼ばれるものでこれも検定統計量の性質を持っています。従って、実験によって得られた二つの分散の比を計算し、それをある基準値（ほとんどの場合はその値が得られる確率が5%である値）と見比べることにより、母分散が同じというには無理があるならば“有意差あり”と判断できます。添字は一つ目のデータセット、二つ目のデータセットを表します。 F とは、R.A. フィッシャーの頭文字を取ったもので別名をフィッシャーの分散比と呼びます。この式で注意して頂きたいのは、 F 統計量はあくまでも本来はカイ二乗統計量/自由度の比であるということです。

F 統計量は二つのデータ分布から得られた分散を比の形で比較することにより、その比が同じ母分布から得られる確率を教えてくれる大変便利な指標です。ただし、 F 統計量の理論分布である F 分布は分子、分母の自由度の関数になっています。よりくどく説明すれば、同じ値の F 値であったとしてもデータ分布のサンプル数によってその値の得られる確率が違うということです（図1）。 F 値を用いることにより、分散分析の行いたい実験要因によって生じる分散と誤差由来の分散の大

きさの比較を行うことができます。より厳密に言うと“実験由来と思われる分散と誤差由来の分散は同じ母分布から取られているかどうか”を判断することができる訳です。

蛇足ですが、現在知られている検定統計量はそれほどたくさん種類がある訳ではありません。今回ご紹介したF統計量の他に現在よく使われるものとしてはカイ二乗統計量やG統計量（尤度比統計量）等が挙げられます。古くから用いられている検定統計は使っている検定統計量の名前をつけているものがほとんどです（t検定、カイ二乗検定など）。その一方で、PCの発達とともに演算量の多い新しい統計が用いられるようになってきました。これらの新しい統計は演算を繰り返すことによって用いたい検定統計量を利用可能な形までデータを変形していきます（一般化線形モデル等）。こうした演算量の多い統計手法では計算過

程の名前がつけられているものがほとんどですが、検定統計量としてG統計量やF統計量がよく用いられています。

3. 分散分析表の情報

分散分析の原理、そして判断基準となるF統計量についてここまでお話をしてきました。それでは実際に分散分析をデータ解析に用いた場合、どのような情報を論文や発表で示すべきなのでしょう。

図2は例としてカブトムシの寿命に及ぼす餌添加物の比較を行ったものです（仮想データ）。皆さんも夏休みにカブトムシを飼育した経験があると思います。最近は昆虫専用の餌が販売されていますが、私が子供の頃はもっぱらスイカを与えていました。スイカだけでは栄養が偏ると子供心に思ったので、スイカに様々な調味料を振りかけてみました。図2を見る限り、調味料Aや調味料Bを与えられた虫は長生きをしている様です。全ての実験区で10頭ずつ実験を行いました。この結果を分散分析で調べてみましょう。

分散分析を行った結果が図3になります。個々の計算過程は前回は参照してください。図3に示したような統計表を“分散分析表”と呼び、分散分析を行った際にはこれを示すのが一般的です。この表を見て第一に分かることは、全体のばらつきを添加物によるばらつきと誤差のばらつきに分けているということです。第二に、それぞれの要因の自由度が示されています。前回にお話ししました様に、要因間平方和は、それぞれの処理区の平均によって計算されます。従って、要因（この場合は添加物ですね）の自由度は5要因-1なので4となります。一方、誤差の平方和はバランスド

図1 異なる分子、分母自由度によるF分布の形の変化。横軸がデータより計算されるF値。縦軸がそのようなF値を取る確率。
a) 分子自由度=3, 分母自由度=5の場合。
b) 分子自由度=10, 分母自由度=100の場合。

図2 カブトムシの餌に様々な調味料を添加した場合の寿命の平均値（仮想データ）。

要因	自由度	平方和	平均平方	F値	p値
添加物	4	1077.32	269.33	49.37	<0.01
誤差	45	245.50	5.456		
全体	49	1322.82			

どちらか片方を載せれば良い

図3 分散分析表の例. データの解釈に必要な情報が整理されている. 平方和と平均平方はどちらかを記載すれば良い.

デザインなので処理区毎のばらつきを単純に足したのになります. それぞれの処理区で繰り返しは10ですから, 処理区内の自由度は9, それを5処理区分足しますと45となります. 全体ではカブトムシの総数は50ですから全体の自由度は49となります.

平方和を自由度で割ったものが平均平方 (Mean Square, MS) と呼ばれるもので推定分散になります. 一般には平方和と平均平方の2つを同時に載せる必要はありません (図3). ほとんどの場合, 平均平方だけを分散分析表では示します. 要因の平均平方と誤差の平均平方の比がF値になります. 要因の自由度を分子自由度, 誤差の自由度を分母自由度として, 得られたF値以上の値が何%の確率で観察されるかを示すのがp値になります. 今回の例では, 計算されたF値が観察される確率は1%以下ですと要因の分散と誤差分散が同じ母分布であるとは考えにくく, 要因の効果はあるであろうと判断されます.

分散分析が教えてくれるのは, 実験計画全体で要因間に違いがあるということだけです. しかし, 実際に知りたいのはほとんどの場合, どの処理とどの処理が違うのか, ということです. そのために, 通常は分散分析を行い, 要因の有意差を確かめた後に総当たりの検定をもう一度行います. これが“事後検定 (post-hoc test)”と呼ばれるものです. 以前はダンカンの多重検定 (Duncan's multiple range test) と呼ばれるものが広く用いられてきました. しかし, この検定手法は前回お話ししました多重性の問題が解消されていないために現在ではほとんど使われません. もしもこれ

を用いて論文を投稿したら, おそらくはレフェリーが勝ち誇って不備を指摘してくるでしょう. 本稿では多重性の問題については詳しく触れませんが, 興味がある方は光永 (2009) を参照してください.

現在最も広く使われている事後検定手法はチューキーのHSD法 (Tukey's HSD test) またはその拡張版であるチューキー・クレーマー法 (Tukey-Cramer test) です. オリジナルのチューキー法はバランスデザインにしか対応しませんが, チューキー・クレーマー法はサンプル数に対して制限が緩くなっています. その代わりに, ほんの少しオリジナルの方が精度が高いと言われています. ほとんどの統計ソフトで計算してくれますから, どちらかを用いるのがお勧めです.

注意して頂きたいのは, 分散分析と事後検定は基本的に別の検定であるということです. 従って, 例えば分散分析で有意差が見られなくても事後検定で有意差が見られることはあり得ます. 一般的に言って分散分析の方が精度は高いので, 事後検定は分散分析で実験系の有意差があることを確認した上で行うべきです. また, 同様の理由で特殊な場合を除き, 論文には分散分析の結果と事後検定の結果は併記されるべきです.

4. 共分散分析の原理と構造

最後に分散分析と直線回帰を組み合わせた“共分散分析 (Analysis of Covariance)”について紹介しましょう. 前回少し触れました様に, この手法により二つの回帰直線の傾きの違いを検定することができます.

第3回で直線回帰の精度を表す決定係数についてお話ししました. これはデータのばらつきのうち, 直線部分で何%説明できるかを示していると解説しました. つまり, 直線回帰もデータのばらつきを分割している訳です. 従って分散分析と計算過程は違いますが, 原理は実は一緒です. そこで, 分散分析と直線回帰を合わせて“一般線形モデル (general linear model)”と呼びます. このように考えると, データのばらつきを要因, 傾き, そして誤差に分けていくことが理論上は可能であると言えます. 分散分析は比較したい要因が質的なものであるのに対し, 直線回帰を組み込んだ共分散分析は比較したい要因に量的なものが含まれている (だから共分散ですね, 第二回を参照して

ください）ことが特徴です。

具体例として、図4を考えてみましょう。これは前回も前々回も使いましたウサギの体重と食べた餌の量の関係です。図ではオスの場合とメスの場合に分けていますが、これをまとめて、全体のデータのばらつきを雌雄によるばらつき、体重によるばらつき（回帰の傾きによって説明されるばらつき）、そして誤差のばらつきに分割すればよい訳です。ただし、雌雄毎に体重によるばらつきには違いがあるかもしれませんので、この部分のばらつきも加えて考えましょう。このような二変量の組み合わせ毎に違う発現をする効果を“交互作用（interaction）”と呼びます。

前回、分散分析の一般式を示した様に、共分散分析を数式化すると次の様になります。

$$Y = \mu + A_i + \text{Slope} + A_i \text{Slope} + \varepsilon \dots\dots\dots(5)$$

ここで、Aは要因の効果（この場合は雌雄ですね）、Slopeは体重による回帰の傾き、ASlopeは雌雄毎の体重による傾きの違いです。分散分析同様に μ は分散を持ちませんから、データ分散は雌雄の効果、体重による傾き、交互作用、そして誤差の4つに分割されます。

図4のデータについて共分散分析を行った結果が図5です。表の読み方は先ほどの分散分析表と全く同じです。最初に注目すべきは交互作用の有意性です。今回のデータではp値が0.04ですので有意であると言えます。つまり、雌雄で回帰直線の傾きには有意差があると言えます。全体としても傾きには有意差が見られますので、体重が増えるほど食べる量は増加すると判断できます。その一方で、雌雄には有意差がありません。これはどのように解釈すれば良いのでしょうか？この解釈は、雌雄という属性だけでは食べる量を判断するには不十分で、体重を考慮して始めて判断ができるということです。

一般に、分散分析、共分散分析のどちらでも、交互作用に有意差がある場合には、それ以上の詳しい解釈をするには無理があります。なぜなら、注目した要因単独の効果だけではなく、要因間の組み合わせによって観測している現象が影響されているからです。従って、交互作用が有意である場合には、事後検定は行わないのが普通です（今回の例では元々必要ありませんが）。

a)

b)

図4 ウサギの体重と一日に食べる草の量の散布図（仮想データ）。第二回の図6と同じ。雌雄で体重が見た目で違うので二つに分けた。どちらも有意な相関が見られる。
a) オスの場合。
b) メスの場合。

要因	自由度	平方和	平均平方	F値	p値
雌雄	1	16.32	16.32	1.97	0.17
体重	1	336.16	336.16	40.61	<0.01
交互作用	1	33.76	33.76	4.08	0.04
誤差	56	463.64	8.28		

共分散分析では最初にここに注目すること！

図5 図4のデータについて共分散分析を行った結果。要因の交互作用に有意差があることに注意。ウサギは雌雄で体重の増加と食べる量の増加の関係が違くと判断される。しかし、雌雄で食べる量に違いがあると統計的には言えない。

5. 第四回のまとめ

1) データから計算された、あるいは推定された統計学的な数値のことを統計量と呼ぶ。中でもある仮定の下に計算した場合、得られた数値が「その事象が観察される確率」と対応するような統計量を検定統計量と呼ぶ。分散分析の際には、検定統計量としてF統計量を用いる。計算される統計量は、データのサンプル数に影響を受ける。この影響を与える部分を自由度と呼ぶ。F統計量は分散比で表され、その分布の形は分子、分母の両自由度によって決定される。

2) 分散分析を行った場合には分散分析表を示すのが一般的である。分散分析表には、データ分散をどのように分割したか、それぞれの分割毎の自由度はいくらか、平均平方はいくらか、そしてそれから計算されるF統計量はいくらか、そしてF統計量は母分布が同じとしたら何%の確率で観察されるかと言った分散分析を行った結果を全て示している。

通常は、分散分析を行った後に個々の要因毎の比較のために事後検定を行う。以前はダンカンの多重検定が主流であったが、現在はチューキーのHSD検定やその拡張であるチューキー・クレーマー検定等が一般的である。分散分析と事後検定は別の検定であり、その精度は分散分析の方が一般には高いので、論文に記載する時には分散分析の結果を示した後に事後検定の結果を示すのが望ましい。

3) 第三回で触れた様に、直線回帰の決定係数とはデータ全体の分散がどのくらい直線回帰の傾きによって説明できるかの指標である。このことから、直線回帰も分散分析と同様にデータのばらつきを分割していることが分かる。そこで、分散分析と直線回帰をまとめて一般線形モデルと呼ぶ。共分散分析とは、一般線形モデルに質的な違いと量的な違いの両方が含まれているものを指す。

分散分析同様に、共分散分析も原理はデータ全体のばらつきを要因毎に分解していくものである。ただし、量的なばらつきは直線回帰の傾きと等価である。複数要因の組み合わせによるばらつきを交互作用と呼ぶ。共分散分析では通常、質的・量的要因毎の傾きの違いが交互作用で示される。

分散分析同様に、結果は分散分析表で示す。最

初に注目するのは交互作用の有意性である。ここが有意であるならば、二つの回帰直線の傾きは違うことが示される。交互作用が有意でない場合には、それぞれの要因の違いについて有意性を確かめるための事後検定も(必要ならば)可能である。

6. 終わりに

今回で“使える統計学基礎講座”は最終回です。本稿では、“農作業研究”に実際に投稿されている論文を参考にして、皆さんが接すると思われるデータ解析について原理やロジック、注意点等なるべく分かりやすく紹介することを目的としました。その一方で個々の統計手法の具体的な計算過程等はなるべく省きました。現在は、ほとんどの人がPCの統計ソフトで統計を行っているので計算過程を詳述してもそれほど助けにならないと考えたからです。シリーズを通じて私が述べたかったことは、サンプル数を十分に取ること、厳密な実験設定の重要性、そして得られたデータ分布の理解とそれに応じた統計手法の選択の必然性です。

今シリーズでは順位のデータ等、正規分布に従わないデータに用いるべきノンパラメトリック検定については解説ができませんでした。また近年主流となりつつあるロジスティック回帰等、一般化線形モデルと呼ばれるグループについても触れることができませんでした。これらの統計手法についてはもしも機会がありましたら改めて紹介したいと考えております。もちろん、現在は多くの統計入門書が発行されておりますので、興味を持たれたら、そうした本も参考にして頂きたいと思えます。

最後に“何のために統計を行うのか”ということを考えてみましょう。私が在籍しています中央農業総合研究センターには統計解析相談グループという部署があり、内外からの年間100件以上の統計相談に対応しています。ここに寄せられる相談でよく見られるのが、“Aという統計手法で解析した所、有意差が見られなかったので別の統計手法を紹介してほしい”というものです。最初に断ってしまいましたが、原則としてAという手法が適切に用いられている限り、別の統計手法はありません。むしろ、皆さんに考えて頂きたいのは、どうしてこのような質問が数多く寄せられるのか?ということです。私にはこのような質問が寄

せられる背景には、“業績原理主義”というものがある様に思えてなりません。試験を行ったからには、必ず論文を書かなければならず、そのためには統計的な有意性が必須であると言う考え方で、一昔前に比べて、研究者に課せられているハードルが高くなっているのです。このようなスタンスにならざるを得ないのも一理あります。しかし、そうであっても“業績原理主義”にはうかつに与したくないものがあります。

私の考えでは統計を用いる最大の理由は、客観的な尺度によって結果を保守的に解釈するためです。残念なことに私が見ている限りでは、統計をきちんと理解していない研究や業績原理主義の研究では結果をほとんどの場合甘く解釈（少しの違いでも違いがあると言いたがる）しています。論文生産を考えれば、この戦略の方が優れていますね。しかし、現在では科学技術は単独で完結せず、ある事実の上に次の事実を積み重ねるという点で情報の伝言ゲームになっています。しかも、その伝言ネットワークは世界に広がっている大変に複雑なものになってきています。このようなネットワークの中であって一点でも情報の信頼性に疑問がある場合、それが周囲に与える悪影響は莫大なものになる可能性があります。従って、厳密な統計とそれによる保守的な判断はネットワークを維

持するためには欠かせないと考えています。

日本では（日本に限らないのですが）個々の統計手法、特に新しい手法について勉強する機会は多いのですが、統計の厳密さについて考える機会は多くありません。学会等でも、新しい手法の紹介には質問が殺到しますが、その信頼性については余り質問されません。これは、日本人の（日本人に限らないのですが）結果を重視し、過程や裏付けをあまり重視しない性質と関係があるのかもしれない。統計の厳密さの指標として“検出力（statistical power）”というものがあります。今回は紹介できませんでしたが、機会がありましたら是非皆さんに考えて頂きたいと思います。以上、おつきあいを頂きましてありがとうございました！

7. 引用文献

- 執行盛之（2002）：遠くまで飛ぶ紙ひこうきを折る一直交表による実験計画法の適用一，農作業研究 37（2）；81-96.
- 藤井宏一（訳）（1983）：生物統計学（原本：Sokal, R.R. and F.J. Rohlf (1961) Introduction to biometry），共立出版，東京，pp.447.
- 光永貴之（2009）：生物統計処理方法の注意点，植物防疫 63（6）；377-380.