

改訂増補版： 統計検定を理解せずに使っている人のために III

池田郁男

東北大学未来科学技術共同研究センター

改訂増補にあたって

この総説は、「統計検定を理解せずに使っている人のために III」の改訂増補版であり、「改訂増補版：統計検定を理解せずに使っている人のために II」の続きである。改訂増補に当たっての詳細は、「改訂増補版：I」の冒頭をお読みいただきたい。この改訂増補版では、理解しにくい部分について、わかりやすい説明に努めた。また、研究者が比較的行う頻度が高いと考えられる多重比較や二元配置分散分析の対応のある関連多群の検定の考え方を主に書き足した。

「改訂増補版：I」では、母集団、標本、母分散、母標準偏差、標本分散、標本標準偏差、不偏分散、不偏標準偏差、正規性の検定について主に記述した。「改訂増補版：II」では、標準誤差、パラメトリック検定の基本、 t 検定の原理、有意水準、両側検定、片側検定、paired と unpaired t 検定の違い、等分散性の検定、ノンパラメトリック検定の原理や利点、欠点について述べた。この「改訂増補版：III」では、3群以上の場合の検定である、多重比較、一元配置分散分析、二元配置分散分析、および、多重性の問題、外れ値の取り扱いを中心に記述する。内容はさらに複雑になることをお許しいただきたい。なお、図番号は前回からの通し番号である。

冊子体は本文の文字が黒ですが、WEB版では一部カラー文字を使用しております。是非WEB版をご覧ください。

3群以上の統計検定

2群の場合は、最も単純な群構成であったが、3群以上になると複雑にならざるを得ない。3群以上の検定では、大きく分けて、要因が一つで横並びの試験で、その要因によりなんらかの変化が起こったかを知りたい場合と、要因が2つあり、それぞれの要因が独立して作用しているのか、互いに影響し合っているのかが知りたい試験の2種類がある。パラメトリック検定では、前者は、一元配置分散分析および多重比較、後者は二元配置分散分析および多重比較（場合によっては2群の検定）に係わる。ここでは主に、パラメトリック検定を中心に論じる。

3群以上の統計検定の前に：正規性の検定、等分散性の検定

正規性の検定：3群以上の場合でも、2群の場合と基本的な流れは同じである。検定にはパラメトリック検定とノンパラメトリック検定があるが（図44）、どちらを利用するかは2群の場合と同様である。

パラメトリック検定は、母集団が正規分布する前提があるので、正規性の検定で正規分布かどうかを調べることができる（改訂増補版：Iの図7）。正規分布でなければノンパラメトリック検定となる。しかし、「改訂増補版：I」で述べたように、正規性の検定は標本の大きさ n が小さいとあてにならない。 n が30前後以上であ

◇◇◇ コラム ◇◇◇

学生に生物統計学を講義すると、理解が難しいポイントが見えてくる。一応は基礎統計学や数理統計学を学習済みの場合が多いにもかかわらず、母集団と標本では、母集団から標本を取るとどうなるのか、あるいは、標本から母集団を推定するという発想が簡単ではないようで、その部分をしっかり理解させることがまずは必要となる。その次に、平方和、分散と標準偏差である。計算式は比較的簡単で、教える側はそれほど難しいと思っていないが、それらの意味がわからない学生が続出する。本セミナーでは、標準偏差が4種類登場するが、これらを区別して理解してもらうのはたいへんである。平方和を標本の大きさ n で割るまでは理解できるとしても、自由度 $n-1$ が登場すると俄然難しくなる。自由度の意味や算出方法は多くの統計書に書かれているが、なぜ自由度で割るかのわかりやすい解説がなく、学生に説明することは容易ではない。このあたりを何とか乗り越えないと、標準誤差を理解することは難しくなる。さらに、分散、標準偏差、標準誤差に関連して正規分布、F分

布やt分布が登場するが、これらの分布がどのようにして出現してきたのかも、きちんと説明する必要がある。この点も、わかりやすく書かれた統計書はなかなか見つからない。たとえば、統計書には、「計算した値はt分布に従うので」などと当然のように書かれているが、初学者はこの「t分布に従う」の意味がわからない。統計学ではいろいろな分布が登場するので、この意味がわからないと、その他の分布の意味もわかるわけがない。おそらく、統計学を専門としている方々は、初学者がこのような点をなぜ理解できないのかが、わからないのではないだろうか？

このような基本的内容をしっかりと理解しないと、そのあとのt検定や種々の統計検定を理解することは難しい。長く統計検定を利用してきても、この基本が理解できていない研究者が多いようである。生命科学系研究者の多くが統計に自信を持っていない理由は、このあたりにあると筆者は考える。そこで、このセミナーでは基本中の基本を中心に解説している。初学者はまず基本をしっかりと理解してほしい。難しい統計検定はその後の話である。

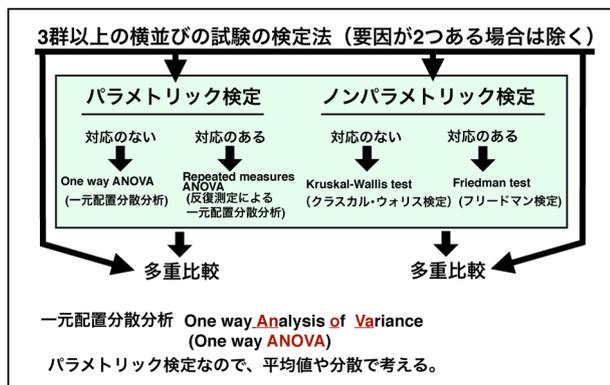


図44 ■ 3群以上の検定：一元配置分散分析，多重比較の一般的な手順

ば、正規性の検定で判断できるが、 n が小さいと判断はあいまいとなる。したがって、 n が小さい場合はパラメトリックでもノンパラメトリックでもどちらでも利用可能である。 n が小さいと正規分布と判定される場合が多いので、基本的にパラメトリック検定を利用することになるがノンパラメトリックも考慮してよい。

等分散性の検定：2群の場合の等分散性の検定に関しては前回説明した(改訂増補版：II, 図34)。3群以上でも基本的な考え方はそれに準じる。等分散性の検定も n が小さいとあてにならないので参考程度と考えたほうがよい。 n が小さいと等分散と判定される場合が多い。等

分散の場合とそうでない場合の検定法については、後述する。

3群以上の等分散性の検定はいくつか知られる。ハートレイの検定(各群の n が同じ場合)やパートレットの検定(n が異なってもよい場合)およびルビーンの検定などがある⁽¹⁾。筆者はこれらの検定法の原理をすべて把握しておらず、使い分け方にも詳しくない。ここでは、もっともわかりやすいハートレイの検定を簡単に紹介するにとどめる。

この検定の基本的考え方は、前回の2群の等分散性の検定とほぼ同じである。2群の場合、それぞれの群の不偏分散 u^2 の比をとってこれをF値として、F値がF分布することを利用して、F分布に当てはめて検定した(改訂増補版：II, 図34)。

3群以上の場合、正規分布する同じ母集団から必要な群数の標本をとり(ここでは3群とする)、不偏分散 u_1^2, u_2^2, u_3^2 を計算する。ここで、 $u_1^2 > u_2^2 > u_3^2$ と仮定すると、不偏分散の最大 u_1^2 を分子に、最小の u_3^2 を分母にして、その比をとる。これを F_{\max} とすると、 $F_{\max} = u_1^2 / u_3^2$ である。不偏分散は3つあるが、それぞれで比を計算すると、ほかの比は F_{\max} よりも小さくなる。

そこで、母集団から3群の標本を取って、 F_{\max} を計算することをなんども繰り返して分布を調べるのである。そうすると、F分布と同様の分布になることが知られて

いる。

あとの手順は、2群の場合の等分散性の検定と同様である。ある実験を行って、 F_{max} を計算し F_{max} がこの分布のどこに位置するかを考える。ここでは、図48のF分布の右下図を F_{max} の図と仮定して説明するが、図の右端方向の5%内に入れば、かなり外れた F_{max} 値となるので、等分散というには外れすぎており、等分散とはいえないと判定する。このように、ハートレイの検定では最もばらついている群とばらついていない群の比を基準とすることで、等分散かどうかを判断している。そのほかの F 値は F_{max} よりも小さい値となるので、 F_{max} で等分散と判定されれば、ほかの F 値も等分散との判定となる。2群の場合の等分散性の検定を理解していれば、容易に理解できるはずである。詳しくは文献1を参照してほしい。同じ母集団からの標本で計算した F_{max} であるから、等分散ではないと断定できるものではないことは、 F 検定やそのほかの検定の基本的考え方と同じである。

3群以上の統計検定：一元配置分散分析，多重比較 (図44, 45)

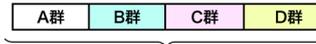
対応のないパラメトリック検定として一元配置分散分析が知られるが、それに相当する対応のある場合、および、ノンパラメトリック検定の対応のない場合と対応のある場合の検定法が存在する(図44)。このセミナーでは、主に、パラメトリック検定の対応のない場合(母集団は正規分布、各群は等分散)、すなわち一元配置分散分析で説明していく。対応のある場合やノンパラメトリック検定は後述する。検定法は多様であり、それらすべてを理解すれば理想的であるが、統計検定を利用する研究者にとって、そこまでは必要ないと筆者は考える。研究者は代表的な検定法の基本的な原理を知ること、それをいかに利用できるかを学ぶべきである。そのほかの多様な検定法は、どのような違いがあるのかを簡単に理解しておけばよい。

まずは、要因が一つで横並びの試験で説明する(図45)。これまでは2群間の比較について述べてきたが、実験研究では群数が3群以上になることはよくある。たとえば、対照群、試験物質1群、試験物質2群…と群数が増え、試験物質の影響を知りたい場合である(なお、二元配置分散分析に相当する試験の場合は後述する)。

まず、図45の①のように、4群全体で何らかの変化が起こったかを知りたい場合は、一元配置分散分析を用いる。一元配置分散分析は英語のone way analysis of varianceを略してone way ANOVAと呼ばれる。一方、②のように、どの群とどの群で有意差があるかを知りたい

3群以上の場合に研究者は何が知りたいのか？

①4群全体でなんらかの変化が起こっているのか？ → 一元配置分散分析



②どの群と、どの群間で差があるのか？ → 多重比較

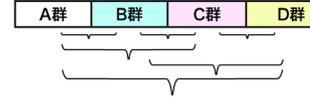


図45 ■ 3群以上の検定で何が知りたいのか？

一元配置分散分析の基本的考え方 (1)

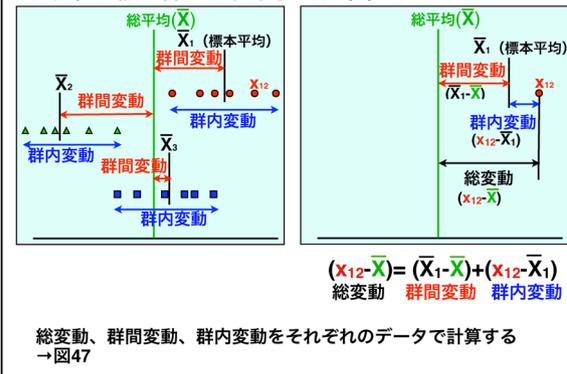


図46 ■ 一元配置分散分析の考え方 (1-1)

い場合は多重比較を用いる。②では4群あるが、すべての群間比較を行うと、比較回数は6回となる。これらの比較を、2群の検定法であるt検定などで繰り返して行なってはならない(理由は後述する)。これが、まずは基本である。2群間の検定の繰り返しはいまだに研究論文で見かけることがあるが、真似するべきではない。

図44にあるように、多重比較を行う場合、多くの統計書には、まずone way ANOVAを行って有意差があれば多重比較を行うと記述されている。統計ソフトで多重比較を行うと、one way ANOVAの検定結果も同時に出力される場合が多いが、その結果を無視し多重比較の結果だけを見ている方も多いのではないだろうか？しかし、重要な研究結果を見逃すことになるかもしれないので、どのような考え方で一元配置分散分析が行われているかは、理解しておく必要がある。

一元配置分散分析の考え方(1) (図46)

分散とはバラツキ方であることはすでに述べた。分散分析とは、3群以上ある場合、たとえば、対照群に対して2種類の試験物質を試験したとき(合計3群)、試験物質

が何らかの影響を与えたかどうかをバラツキ方を用いて調べようとする検定法である。あくまでも、何らかの影響があったかどうかの検定であって、群間比較ではない。

概念をつかんで欲しいが、それでも込み入った話になるので注意して読んで欲しい。一元配置分散分析はパラメトリック検定であるから、母集団は正規分布することが前提である。図46の左グラフを見て欲しい。

2群の検定の場合と同様に、統計検定では差がないことから考えるので、同じ母集団から3群を取り出すという考え方を。今、一つの母集団から6個の標本データを取り出す。これを3回繰り返して3つの群を作り、それぞれの群の標本平均を $\bar{X}_1, \bar{X}_2, \bar{X}_3$ とする。

これらの標本平均をどのように比較するかを考える。2群の比較の場合は、2つの標本平均の差を取ったが、3群以上ある場合は、全データ（ここでは3群の18個）の平均である総平均を基準にし、各標本平均が総平均からどれ位離れているかを考える。ここでは、総平均を \bar{X} と定義する。左図では3つの群のデータと標本平均をわかりやすいように上下にずらして示している（赤丸、緑三角、青四角の3群）。今第1群（赤丸）の6つのデータを $x_{11}, x_{12}, \dots, x_{16}$ と番号をつけ、右から2番目のデータ x_{12} を代表の標本データとして考える。

ここで、標本データ x_{12} だけを取り出した図46右グラフをみて欲しい。ここで、標本データ x_{12} が基準となる総平均 \bar{X} からどれだけ離れた（変動した）ところにあるかを考える。総平均 \bar{X} からの離れ方を「総変動」と名づけると、総変動は $(x_{12} - \bar{X})$ を計算すればよい。

次に、第1群の標本平均 \bar{X}_1 から x_{12} がどれくらい離れているかを考えると $(x_{12} - \bar{X}_1)$ を計算すればよい。これは第1群内での変動であるから「群内変動」と名づける。

さらに、 x_{12} が所属する第1群の標本平均 \bar{X}_1 が総平均 \bar{X} からどれくらい離れているかは $(\bar{X}_1 - \bar{X})$ で求めることができる。これを「群間変動」と名づける。そうすると、「総変動」は「群間変動」と「群内変動」を足したものであることがわかる（図46右下式）。そこで、この計算を全18個のデータで行う（図47の左側）。

「改訂増補版：I」の分散の計算方法の項で述べたが、これら変動はプラスになる場合とマイナスになる場合があるので、平方することに決まっている。

そこで、総変動、群内変動、群間変動をそれぞれ平方し、18個すべてを合計する（図47では、縦に合計する）。これはいわゆる平方和である。そうすると面白いことに、総変動²の総和 = 群間変動²の総和 + 群内変動²の総和の式が成り立つのである（図47下の式）。そんなバカな！と思われるかもしれないが、実際のデー

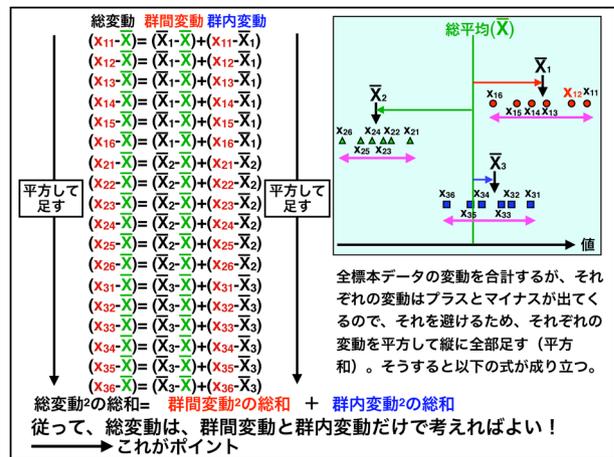


図47 ■ 一元配置分散分析の考え方 (1-2)

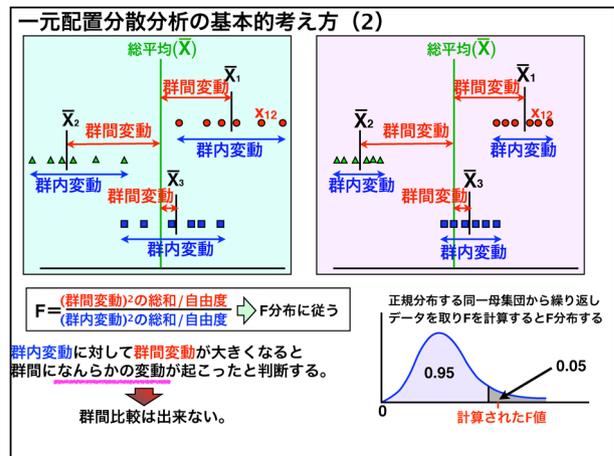


図48 ■ 一元配置分散分析の考え方(2)

タで計算すると必ず成り立つのである⁽¹⁾。何が言いたいのかというと、つまり、すべての標本データのバラツキ方の合計である「(総変動)²の総和」は「(群間変動)²の総和」と「(群内変動)²の総和」だけで考えればよく、そのほかの因子を考慮する必要はないということである。別の言い方をすると、「(総変動)²の総和」は「(群間変動)²の総和」と「(群内変動)²の総和」に分解できるのである⁽¹⁾。これはたいへん重要なポイントである。

一元配置分散分析の考え方(2) (図48)

それでは、群間変動と群内変動を用いて、どのようにして検定するのであろうか？ 図48の左のグラフは図46の左と同じである。では、図48の右上グラフを見て欲しい。3群の標本平均と総平均は左グラフと全く同じとする（すなわち群間変動は同じ）。しかし、群内のデータのバラツキが左グラフよりも小さいため、群内変動が小さ

い。したがって、左グラフに比べ右グラフでは、3つの群の間に明確な差があるように見える。この原理を利用したのが一元配置分散分析である。つまり、群内変動に対して群間変動が大きくなると群間に差があることにしようと考えるのである。具体的には、まず、(群間変動)²の総和と(群内変動)²の総和は群数や n が大きいほど大きな値となる平方和であるから、自由度で割って、平均の変動を計算する(これは不偏分散の計算と同様であるが、平均平方という名称が用いられる)。そして、以下の式のように F 値を計算する(図48左下式)。

$$F = \frac{\frac{(\text{群間変動})^2 \text{総和}}{\text{自由度}}}{\frac{(\text{群内変動})^2 \text{総和}}{\text{自由度}}}$$

ここで、同一母集団から3群のデータを取り出し F 値を計算することを、何度も繰り返して、 F 値がどのような分布をするか求めると、図48右下グラフにあるような分布になることがわかっている。これは F 分布と呼ばれる。そうすると、同一母集団から取り出した3群のデータであっても、群内変動に対して群間変動がかなり大きくなる場合が稀に起こりうる(すなわち、 F 値が大きくなる)。 F 分布グラフでは右端あたりになる。

有意差の考え方は、計算された F 値が F 分布の右端5%内にあると、同一母集団からのデータと考えるにはあまりにもかけ離れており、滅多に起こらないことが起こったと考えて、何らかの影響があったことにしようとするのである(これは F 検定と呼ばれる)。

ここで自由度であるが、群間変動では群数が3であり、それぞれの標本平均から総平均 \bar{X} を引いており、計算式に総平均 \bar{X} が入っているため、自由度は1減って、 $3-1=2$ となる。一般化すると、群数を a とすると $a-1$ である。

一方、群内変動では、それぞれの群で6個のデータがあり、計算式にそれぞれの群(3群)の標本平均 $\bar{X}_1, \bar{X}_2, \bar{X}_3$ が入っている(図47左側の式)。したがって、それぞれの群での計算で自由度が1ずつ減ることになる。そこで、各群の標本の大きさを n とすれば、総標本データ数 $a \times n$ (ここでは18)から a (ここでは3)減って、自由度15となる。一般化すると、 $an-a$ となる*1。

F 分布は自由度の違いで形が異なるので、ここでは自由度(3-1, 18-3)の F 分布を用いて検定することとなる。一般化すると、自由度($a-1, an-a$)の F 分布となる。

*1 自由度の考え方は、前回の脚注*7に記載したとおりである。

対応がある場合の分散分析

ラットにある物質を経口投与し、経時的に採血して物質量を測定した。

①時間と共に上昇したといえるのか? --- 反復測定による一元配置分散分析

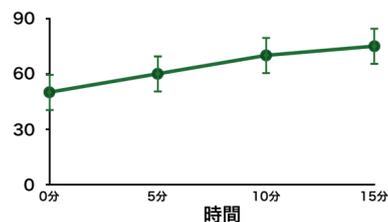


図49 ■ 反復測定による一元配置分散分析の例

その他の分散分析法 (図49)

ここまでは、図44左端のパラメトリック検定、対応のない場合の一元配置分散分析について説明した。パラメトリック検定では、対応のある場合の検定法は、反復測定による一元配置分散分析 (one way repeated measures ANOVA) がある(図44)。図49に例を示した。同じラットから経時的に採血しているため、「対応のある」実験となる。たとえば、0分と5分後の2点の比較であれば、paired t testを行うことができる。しかし、ここでは4点あるので、paired t testは使えない。分散分析は何らかの変化が起こったことがわかるので、この例では有意差が得られれば、統計的に有意に増加が起こったと判断される。

母集団が正規分布しない、あるいは等分散でない場合にはノンパラメトリック法を用い、対応がない場合はKruskal-Wallis法、対応がある場合はFriedman法がある(図44右側)。2群のノンパラメトリック法は前回説明したが、3群以上でも基本的な原理は同様であり、データを順位に置き換えて検定する。これらの検定法については、拙著にその原理を簡単に説明している²⁾。また、詳しく学びたい場合は、文献1を参照されたい。

多重比較の基本的考え方 (図50)

多くの生命科学系研究では、図45の②にあるように、どの群とどの群の間で有意差があるかを知りたい場合が多いのではないだろうか? この場合は多重比較を用いる。基本的に、 t 検定を繰り返し用いてはならない。

多重比較はたくさんの方が考案されており、それらすべての考え方を理解することは困難で、生命科学系研

3群以上の場合、すべての群で合算の u を計算する!

第1群のデータを x_1 から x_5 標本平均を \bar{X}_1

第2群のデータを x_6 から x_{10} 標本平均を \bar{X}_2 とする。

第3群のデータを x_{11} から x_{15} 標本平均を \bar{X}_3

$$u_1^2 = \frac{(x_1 - \bar{X}_1)^2 + (x_2 - \bar{X}_1)^2 + (x_3 - \bar{X}_1)^2 + (x_4 - \bar{X}_1)^2 + (x_5 - \bar{X}_1)^2}{n_1 - 1} \rightarrow \text{平方和1とする}$$

$$u_2^2 = \frac{(x_6 - \bar{X}_2)^2 + (x_7 - \bar{X}_2)^2 + (x_8 - \bar{X}_2)^2 + (x_9 - \bar{X}_2)^2 + (x_{10} - \bar{X}_2)^2}{n_2 - 1} \rightarrow \text{平方和2とする}$$

$$u_3^2 = \frac{(x_{11} - \bar{X}_3)^2 + (x_{12} - \bar{X}_3)^2 + (x_{13} - \bar{X}_3)^2 + (x_{14} - \bar{X}_3)^2 + (x_{15} - \bar{X}_3)^2}{n_3 - 1} \rightarrow \text{平方和3とする}$$

合算の不偏分散

$$v^2 = \frac{\text{平方和1} + \text{平方和2} + \text{平方和3}}{(n_1 - 1) + (n_2 - 1) + (n_3 - 1)} \rightarrow \text{平均平方と呼ばれる}$$

合算の不偏標準偏差

$$u = \sqrt{\frac{\text{平方和1} + \text{平方和2} + \text{平方和3}}{(n_1 - 1) + (n_2 - 1) + (n_3 - 1)}} \rightarrow \text{合算の不偏標準偏差}$$

図50 ■ 3群の場合の合算の不偏標準偏差の計算方法

研究者がもっとも知りたい、それらの使い分けを知ること
も容易ではない(筆者も同様である)。研究者は基本的
ないくつかの検定法についての考え方と使い方を理解す
れば、大半の実験で利用できるかと筆者は考える。

すでに、2群の場合の t 検定の原理は前回説明した。
多重比較だからといって全く異なる計算を行うわけでは
なく、原理は2群の場合とそれほど違いはないので安心
してほしい。 t 検定の場合の t 値を計算する式は以下で
あった。この式の誘導がわからない場合は、前回の図
24~26を参照して欲しい。

$$t = \frac{(\bar{X}_1 - \bar{X}_2)}{u \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}}$$

この式の分母は、同一母集団から2群の標本をとって
得られた2つの標本平均の差 $(\bar{X}_1 - \bar{X}_2)$ の標準誤差であ
り、これを基準(分母)として、標本平均の差 $(\bar{X}_1 - \bar{X}_2)$ (分子)
がどれくらい大きい(あるいは小さい)かを計算している。
ここで、 u は2群の合算の不偏標準偏差である(改訂増補版:IIの図26)。
簡単に言えば、標本平均の差 $(\bar{X}_1 - \bar{X}_2)$ のバラツキ $((\bar{X}_1 - \bar{X}_2)$
のありそうな範囲)を基準にして(分母)、分子の $(\bar{X}_1 - \bar{X}_2)$
が t 分布のゼロからどれくらい離れているかを計算
している。かなり離れていれば有意差ありと判定する。

3群以上ある場合でも、ほぼ同様の考え方であるが、
分母の不偏標準偏差が異なる。3群以上ある多群の場合
は、合算の不偏標準偏差は実験したすべての群で合算し
て計算する。つまり、3群であれば、3群分の平方和を
足し、3群分の自由度の合計で割って平均平方(不偏分
散の計算と同じ)を計算し、それをルートして不偏標準
偏差を求める(図50)。この値を v とする。すなわち、
3群分のバラツキを基準にして、ある2群の標本平均の

差、たとえば $(\bar{X}_1 - \bar{X}_2)$ を計算して、 t 値を求める。こ
の t 値は2群の場合と同様 t 分布するので、 t 分布に当て
はめれば、有意差があるかどうかわかる。このあたりは
 t 検定の手順と同様である。すなわち、計算式は以下と
なる。 $(\bar{X}_1 - \bar{X}_3)$ および $(\bar{X}_2 - \bar{X}_3)$ についても同様に
計算して t 値を計算し、 t 分布で検定する。この方法は、
最小有意差法(LSD法、あるいは、FisherのPLSD法)
と呼ばれる。

$$t = \frac{(\bar{X}_1 - \bar{X}_2)}{v \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}}$$

このように、多重比較は基本的には t 検定を拡張した
検定法であることがわかる。この検定法は、最も有意差
のでやすい検定法であるが、後述する多重性の問題があ
り、3群の場合でのみ利用することができる。後述する
が、この多重性の問題をクリアするために多くの多重比
較法が開発された。

■ 多重比較を行う前の一元配置分散分析は必要か?

(図51)

もともと、一元配置分散分析は「何らかの影響がある
かどうか」を調べることに役割をもった検定法であり、研
究分野によっては利用価値がある。一方で、どの群とど
の群同士で違いがあるのかを知りたい研究者は、多重比
較で検定する必要がある。統計ソフトで多重比較を用い
ると、一元配置分散分析の結果も同時に表示される。す
でに述べたが、多くの統計書には一元配置分散分析で有意
差が得られ「何らかの変動が起こった」と判断される
と、多重比較に進むと記されている。確かに、一元配置
分散分析で変動が検出されないのに、多重比較では、あ
る2群間に有意差があるというはおかしな話である。し
かし、現実ではこの現象はよく起こるのである。これは一
元配置分散分析と多重比較の考え方の違いが原因である。

一元配置分散分析では、(群内変動)²の総和/自由度
に対する((群間変動)²の総和/自由度)の比を求める
ことから、標本データのばらつき全体に対する標本平均
のばらつき全体で考えている。したがって、たとえば、
群数が多い場合、そのなかで1群の標本平均のみが変動
しても、全体としては埋没してしまう可能性がある(図
51)*²⁽³⁾。一元配置分散分析はまさしく「全体で何らかの
変動が起こった」ことを調べる方法であり、ある成分の

*²逆も起こりうる。つまり、一元配置分散分析で有意差があるの
に、多重比較では全く差がないという現象である。これも困った
ことであるが、群間比較が目的であれば、多重比較の結果を尊重
せざるを得ない。

確率が高くなる。すなわち、t検定を繰り返すと有意差が検出される確率が増加する。これが問題視され、多重性の問題と呼ばれる。このような多重性の問題を解決するために、有意水準の増加を極力小さくしたのが多重比較である。簡単に言えば、有意水準を狭めることで、全体の有意水準を5%に調整するのである。

多重比較法の検定原理：多重比較は2群間の比較よりも厳しくなる！（図53, 54）

多重比較では有意水準を狭めているから、当然のことながら、2群のみの比較よりも検定は厳しくなり、群数が増えれば増えるほど厳しくなる（図53）。したがって、研究を行う際に、あれもこれもと考えて多群で一度に実験してしまうのは考えものである。目的をはっきりもち、極力群数を減らす努力をしたほうがよい。

多重比較はいろいろな方法が考案されているが、上記の有意水準の狭め方がそれぞれで異なり、厳しかったり、緩かったりする。したがって、同じデータを用いて種々の多重比較を実施すると、有意差があったりなかったりすることになる。

最も単純でわかりやすいBonferroni法では、たとえば3群の場合、2群の差の検定は3回行う必要があるので、単純に有意水準0.05を3で割って、0.017を有意水準とする（図53）。このようにすれば、3回検定しても全体の有意水準は0.05に保つことができる。4群の場合は6回検定を行うので、 $0.05/6=0.0083$ を有意水準とする。5群の場合は10回の検定となるので、有意水準は0.005となる。このように、Bonferroni法の場合は群数が増えたと有意水準が飛躍的に小さくなり厳しい検定となるので、5群以上では用いるべきではないとされる。厳しすぎるのである。そこで、Bonferroni法の関連法であるHolmやShaffer法ではこの点が改良され、この厳しさが緩和されている。Holm法の簡単な原理は、拙著で説明している²⁾。

また、よく用いられる多重比較としてTukey-Kramer法がある。この検定法は、Bonferroni法とは異なる原理で有意水準を狭めている。

Tukey-Kramer法は最もよく用いられる検定法の一つである。この方法では、計算式は以下の最小有意差法の場合と基本的には同じである。

$$t = \frac{(\bar{X}_1 - \bar{X}_2)}{\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}}$$

たとえば、3群あるとして、標本平均の差は、 $(\bar{X}_1 - \bar{X}_2)$ 、 $(\bar{X}_1 - \bar{X}_3)$ および $(\bar{X}_2 - \bar{X}_3)$ が得られるが、そのうち、最大の差で考える検定法である。比較する2つの

多重性の問題を解決するためには、有意水準の増加を小さくする！
その方法が、それぞれの多重比較法で異なるので、検定結果が異なる場合が出てくる。

Bonferroni法では、有意水準を比較回数で割り算することで、有意水準を狭めている。

3群の場合、3回比較するので、有意水準は $0.05/3=0.017$

4群の場合、6回比較するので、有意水準は $0.05/6=0.008$

5群の場合、10回比較するので、有意水準は $0.05/10=0.005$

群数が増えるほど、飛躍的に比較回数が増えて、有意差は出にくくなる！

3群以上の実験を行うと有意差は出にくい。群数が増えれば増えるほど有意差は出にくいことを知っておくべき。

群数はできるだけ少なく！
2群が理想的！
いたずらに群数を増やすな！

Bonferroni法は5群以上では用いるべきではないとされる。

図53 ■ 多重比較はt検定よりも厳しい！

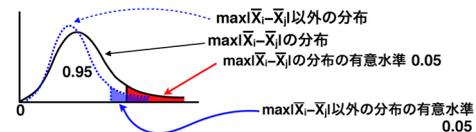
Tukey (Tukey-Kramer) 法では、

$\bar{X}_1 - \bar{X}_2$, $\bar{X}_1 - \bar{X}_3$, $\bar{X}_2 - \bar{X}_3$ のうち、最も大きな差で考える。
すなわち、

$$\max.t = \frac{\max|\bar{X}_i - \bar{X}_j|}{\sqrt{\frac{1}{n_i} + \frac{1}{n_j}}}$$

の分布を考える。

スチューデント化された範囲の分布



右端5%に入れば有意とする。

$\max|\bar{X}_i - \bar{X}_j|$ 以外の差もこの分布で検定。

$\max|\bar{X}_i - \bar{X}_j|$ 以外の差は、 $\max|\bar{X}_i - \bar{X}_j|$ よりも小さい値となるので、 $\max|\bar{X}_i - \bar{X}_j|$ の分布の有意水準5%を越えずに検定可能。

図54 ■ Tukey-Kramer法の原理

標本平均を \bar{X}_i および \bar{X}_j で表すとして、最大の差を、 $\max|\bar{X}_i - \bar{X}_j|$ で表すこととする。差の最大を示すためにマイナスがでないように絶対値を用いている。この点で、最小有意差法の式とは異なってくる。Tukey-Kramer法では、この $\max|\bar{X}_i - \bar{X}_j|$ の場合にt値がどのように分布するかを考える（最大のt値なので、ここではmax.tで表す）。すなわち、正規分布する母集団から3群とって、以下のmax.t値を計算することを、無限に繰り返してt値の分布を調べると、図54の実線で書いた分布になる。これはt検定ではt分布になったが、式の分子を絶対値としたのでマイナスにならず、このような分布になる。この分布はF分布に似ており、「スチューデント化された範囲の分布」と呼ばれる。

$$\max.t = \frac{\max|\bar{X}_i - \bar{X}_j|}{\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}}$$

あとは、F検定と同様で、ある実験を行って得られた

t 値がこの分布の外れたあたり（右方向の有意水準5%）にすれば、かなり外れているので、有意差ありと判定する。 $\max|\bar{X}_i - \bar{X}_j|$ 以外の標本平均の差 $|\bar{X}_i - \bar{X}_j|$ は、 $\max|\bar{X}_i - \bar{X}_j|$ よりも小さな値になるので、 t 値も小さくなり、その分布をとると図54の点線で書いた図となる。この点線の分布で有意水準5%で検定すると、図の青の部分となり、かなり小さな t 値でも有意差が得られるが、それをせず、 $\max|\bar{X}_i - \bar{X}_j|$ での分布で検定を行うと（図の赤の部分）、そのほかの差の検定も有意水準5%を超えずに検定できるという考え方である。

この検定法は、5群以上ではBonferroni法よりも厳しくないもので、よく用いられる（2群の検定よりも厳しいことに変わりはないが）。

このように、多重比較では多重性の問題をいろいろ工夫して解決しようとしていることがわかる。

多重比較の欠点（図55）

ここでは多重比較での、注意しなければならない点を挙げる。

多重比較は「対応のない場合」で開発され、ほとんどの場合「等分散」が仮定されている。「対応のある場合」は検定法がないので、対応のない場合の検定法で代用される（後述する）。したがって、paired t testのような厳密な検定は無理であることを知っておくべきである。等分散でない場合は、パラメトリック法があるにはあるが、ノンパラメトリック法も検討した方がよい。

多重比較は判定が厳しく、群数が増えれば増えるほど厳しくなる。しかも、いろいろな検定法で厳しさが異なるので、検定結果に違いが出る。したがって、ある検定法で有意差がないからといって有効性がないとはかぎらない。たとえば、多群で行った試験を、2群の試験でやり直すと有意差が得られることはありうる。有意差がないからといってあきらめないことである（図55上）。

多重比較では群数が増え、比較回数が増加するほど厳しい検定となるので、比較回数を減らす工夫が必要である。たとえば、機能性物質が5種類あって、それらのある細胞に添加したとき、特定の細胞内物質を増加させるかを知りたいとき、無添加群を対照として変化を見たい。この場合、5種の機能性物質は無添加群とのみ比較すればよく、機能性物質間の差を調べる必要がないという研究はよくある。対照群とのみ比較すれば、全群間で比較するよりもかなり比較回数が減る。このような検定法が存在し、パラメトリック検定では、Dunnnett法である。

最小有意差法の t 値の計算式は前項で示したが、分母

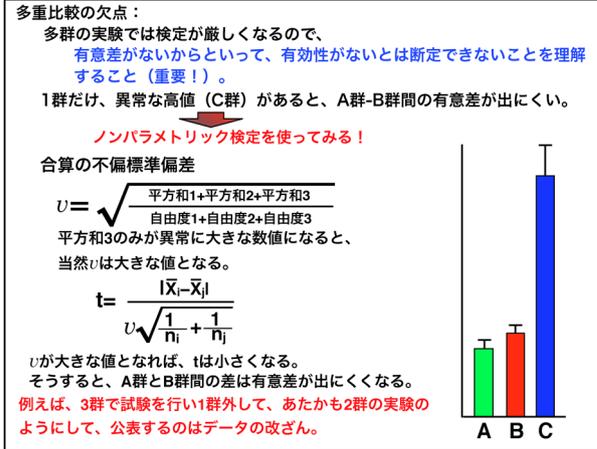


図55 ■ 多重比較の欠点

は標準誤差である。この分母の v は全群の合算の不偏標準偏差である。つまり、 t 値は全群のばらつき方を基準として、2つの標本平均の差がどれくらいあるかを示す値である。分母が大きな値になると相対的に t 値は小さくなるため、有意差は得られにくくなる。

そこで、たとえば図55右グラフのように、大きな標本平均と大きな標準偏差をもつ群（C群）が存在すると、合算の不偏標準偏差がかなり大きくなると考えられる。そうすると、A群とB群を比較する場合、標本平均の差が相対的に小さいので、 t 値は小さくなり、この群間の有意差はでにくくなる。そこで、A群とB群だけ取り出して、 t 検定を行うと有意差が得られるが、3群で検定するとA群とB群間には有意差がないということがありうる。実験目的がB群であったら大問題である。だからといって、C群を外し、あたかもA群とB群の2群の試験であったように見せかけて公表するのは、データ改ざんとなり許されない*4。このように、多群の試験では、各標本平均が大きく違うと（変動していると）、 t 値計算の基準となる分母が大きくなるため、差の小さい群同士の有意差は得られにくくなることを覚悟したほうが

*4ただし、たとえば、実験計画段階で、A群は対照群、B群が目的の試験物質群としてA群とB群だけで比較すると計画したとする。しかし、実験自体が妥当なものかどうかを判断する目的で、C群をポジティブコントロールとして、B群の試験物質と同様の効果があることがすでに知られている物質を与えた群を設けたと仮定する。この場合、C群を測定項目が妥当な応答をするかどうかを確認するためだけに用いるのであれば、多重比較を適用せず、A群とB群だけで2群間の比較をすることは可能と考えられる。あくまでも、実験計画段階でそれを宣言する必要がある。もちろん、決してA群とC群あるいはB群とC群を比較してはならない。さらに述べるならば、当初は多重比較の3群として計画したが、図55のグラフのように、C群があまりにも大きな値となってしまったために、実験終了後にC群を後づけで排除し、A群とB群だけで2群間比較を行うことは許されない。

よい。差の小さい群同士の違いを調べたいのであれば、それらの群のみを別に試験すべきである。

なお、このようなデータの場合、ノンパラメトリック検定が有効かも知れない。ノンパラメトリック検定では、データを順位に置き換えるので、数値上の大きな差がなくなる^{*5}。

多重性の問題はほかにもある！ (図56)

ここで多重性の問題についてもう少し述べる。図56の上側を読んでほしい。ある実験を行って思うようなデータがでず、再度同じ試験を行うことは、われわれにもある。再実験でよい結果が得られたことから、そのデータを論文にしたとする。よく考えてみると、1回目と2回目の試験でどちらがより真実に近いかは、実はわからない。これも多重性の問題に相当する。何度も同じ実験を繰り返すと、そのうちに本当は差がないのに有意となる可能性が高まる。2回同じ試験を行うと、有意差がでない確率が $(1-0.05) \times (1-0.05) = 0.90$ となり、有意水準が10%へ上昇する。簡単な *in vitro* 試験であれば、思うようなデータがでるまで、何度でもやり直せるが、実は多重性の問題をはらんでいることを認識すべきである^{*6}。

図56の下側を読んで欲しい。このような恐ろしい結果になると、笑っては済まされない。実は、われわれの身の回りでは結構起きている。筆者の印象では、たまたま有効性の得られた論文がかなり多く出回っている。研究論文とは必ずしも正しい内容のものばかりではない。publishされた論文を鵜呑みにしてはいけないし、1回の実験結果だけを論文にするのは危険である。研究者にとって再現性を調べることは必須である。

さらに多重性の問題！ (図57)

さらに、身近なところに多重性の問題は潜んでいる。動物や細胞試験では一度の試験で多くのパラメータを測定する場合がある。多重性の問題という観点から考えてみると、一つひとつの測定パラメータは、本当は差がな

^{*5} 筆者は図55のようなデータで、A群とB群間にTukey-Kramer法では有意差が得られなかったが、ノンパラメトリック検定の多重比較(Steel-Dwass法)を適用し、有意差をえた経験がある。

^{*6} 「改訂増補版：I」で言及したが、データには測定誤差が含まれる。測定誤差が大きいと判断されるデータの場合に全く同じ試験を繰り返し行うのは仕方がない。しかし、その場合、1回目と2回目に測定したデータのどちらが真実に近いかわからない。測定誤差の大きい測定者はまず訓練すべきである。なお、測定誤差もまた正規分布することが知られる。

多重性の問題は他にもある！

ある実験を行って、ぎりぎり有意差が得られなかった。そこで、もう一度同じ実験を行って、有意差を得たので論文にした。本当にそれでいいのか！

この場合、有意水準は5%ではない。有意水準は10%近くに上昇する。何度も同じ実験を繰り返すと有意となる確率が高まる。

20回実験を繰り返せば、本当は差がないのに、60%以上の確率で有意となる可能性あり。その実験だけを論文にしているのか？

有意水準5%は、100人の研究者が全く同じ試験を行った場合、5人は、本当は差がないにもかかわらず、有意差を得る確率でもある。この5人だけが研究結果を論文にするかもしれない。

→ 真実とは異なる研究結果が世の中に出回ることとなる！

→ 再現性のない論文が多い一因。

publishされた論文の内容を鵜呑みにしてはいけない！

1回の実験結果だけを論文にするのは危険！

図56 ■ そのほかの重要な多重性の問題

測定項目を増やすのも、問題あり！

有意水準5%での検定は、一つの実験で20項目のパラメータを測定すると、そのうち1項目は、本当は差がないはずなのに、有意となるかもしれない確率でもあることを認識すべき！
(マイクロアレイで1000項目測定すると50項目に有意差！)

何の影響も与えない成分でも、多くのパラメータを測定すると、有意差が出てくる可能性が高まる。



だからといって、測定項目を減らすわけにも行かない。このような有意差には再現性はないはず。

真実を見極めることが重要！

標本での試験とは、そのようなものである事を認識すべき！

研究者はいくつかの条件で試験して再現性をみて、真実を見極めるべき！

図57 ■ まだまだ多重性の問題！

くても有意水準5%の確率で有意となる可能性がある。そうすると、20項目測定すると、1項目は偶然に有意となる可能性がある。マイクロアレイでは多数のデータが得られるが、そのうちの5%は偶然の有意差かもしれないことを認識すべきである。パラメトリック検定の基本的な考え方からすれば、そうなることが宿命であることは、このセミナーの読者であればご理解いただけるはずである。

このようなことを避けるためには、測定項目は極力減らすのがよいが、実際問題として、それはできないのが普通である。われわれにできることは、別の角度からの実験を行って、同様の結果が得られるかを調べ、データの再現性をみることである。偶然の有意差には再現性はないはずであり、真実を見極める必要がある。上述のマイクロアレイであれば、重要な遺伝子は、必ず、PCRなどで確認すべきである。あるいは、別の角度から実験条件を変えて試験を行えばよい。間違いのない真実であれば、実験条件を変えても類似の結果が得られるはずで

ある。一回の実験だけを論文化する危険性がここでも見えてくる。良識ある研究者であれば、このような多重性の問題を頭の片隅に置いて、研究を行うべきである*7。

一般に用いられる多重比較

比較的よく用いられる多重比較を示す。

パラメトリック検定-対応のない場合

全群の比較では、等分散の場合は、Tukey-Kramer法、Bonferroni法とその関連法（Holm法、Shaffer法）、最小有意差法（3群のみ利用可能）がある。

等分散ではない場合は、Games-Howell法があるが、筆者は利用経験がない。この方法は標本の大きさ n が同じ場合に用いることができる。また、ノンパラメトリック法も用いられる（Steel-Dwass法）^(1,2)。

対照群とのみ比較する場合は、等分散の場合はDunnett法がある。等分散でない場合は、ノンパラメトリック法（Steel法）を用いる。また、たとえば、試験物質の投与量依存的にあるパラメータが次第に上昇するあるいは降下するような場合は、Williams法が有意差はでやすいとされる。

ノンパラメトリック検定-対応のない場合

全群の比較では、Steel-Dwass法があり、対照群とのみ比較する場合は、Steel法がある。Williams法のノンパラメトリック版として、Shirley-Williams法がある。上述のように、母集団が正規分布しているが等分散でない場合はノンパラメトリック法を用いることができる。

Steel-Dwass法やSteel法は馴染みがないかもしれないが、パラメトリック法と遜色ない検定法である。ノンパラメトリックの多重比較の原理はここでは述べないが、興味があれば拙著⁽²⁾および文献1を参照されたい。

パラメトリックおよびノンパラメトリック検定-対応のある場合

多重比較の欠点の項で述べたが、対応のある場合の多重比較は開発されていないので、対応のある場合は多重比較は使えないと書かれた統計書がある。どうしても検定したい場合は、対応のない場合の検定法で代用するしかない。したがって、パラメトリック法もノンパラメトリック法も全群での比較はBonferroni法とその関連法

が、対照群とのみの比較はDunnett法が用いられる。

2群の場合はpaired t testが使える、そのメリットはすでに述べた。しかし、3群以上では、対応のない場合で代用するので、有意差は得られにくくなる。したがって、検定に適しているとはいえず、このような実験系は避けたほうがよい。また、Tukey-Kramer法は、対応のある場合には適用できないとされる⁽⁶⁾。

なお、パラメトリック検定の対応のない場合で、いまだに一部の研究分野ではDuncanの検定法（Duncanの多重比較法、Duncan's multiple range test）が用いられている。われわれも多重性の問題が明らかになる前は利用したことがある。しかし、この方法は多重性の問題を考慮していないため、使用すべきではない⁽⁴⁾。

二元配置分散分析（図58）

二元配置分散分析はtwo way analysis of varianceを略して、two way ANOVAと呼ばれる。一元配置分散分析同様、群間比較を行う検定ではなく「何らかの影響がある」ことがわかる検定法である。一元配置分散分析では、対照群、試験物質1、試験物質2、…という群構成となり、群は横並びであった。しかし、二元配置分散分析では、試験したい因子が2つあり、それら2つの因子がそれぞれ独立して影響を与えるのか、相互に影響しあっているかどうかを知りたいという検定である⁽¹⁾。

具体的には、2つの因子の影響があるかないか、および、2つの因子の作用は相加作用か、相乗作用あるいは相殺作用かを知ることができる点の特徴である。図58の架空の例では、2つの因子としてA因子（発酵茶）とB因子（コレステロール、Chol）とした。横並びで示す群構成は、

Two way ANOVAで何が知りたいのか？

因子が二つある場合

例えば、因子A 発酵茶無添加、添加
因子B コレステロール (Chol) 無添加、添加

	Chol無添加(B ₁)	Chol添加(B ₂)
発酵茶無添加(A ₁)	A ₁ B ₁	A ₁ B ₂
発酵茶添加(A ₂)	A ₂ B ₁	A ₂ B ₂

二元配置分散分析で知りたいこと。

1. 因子Aの影響（発酵茶）
2. 因子Bの影響（Chol）
3. AとBは影響し合っているか？
（交互作用）
相殺作用、相乗作用それとも相加作用？

互いに影響を与えていなければ、それぞれの因子ごとに考えればよい。

互いに影響を与えていれば、2つの因子を切り離しては考えられない。---因子AとBの間には交互作用（interaction）があると表現する。(AxB)

*7 残念ながら、再現性のない論文が世の中には氾濫していることは、「改訂増補版：I」で指摘したとおりである。

図58 ■ 二元配置分散分析

1. 対照群—発酵茶無添加, Chol無添加群 (A1B1)
 2. 発酵茶群—発酵茶添加, Chol無添加群 (A2B1)
 3. Chol群—発酵茶無添加, Chol添加群 (A1B2)
 4. 発酵茶+Chol群—発酵茶添加, Chol添加群 (A2B2)
- の4群である。二元配置分散分析用に群を書き直したのが右側である。

これを図示すると、図58の左表となり、二次元の群構成となる。

このような群構成であるにもかかわらず、単に横らびに4群と捉えて多重比較で統計検定している例が多く見られるが、二元配置分散分析で処理し結果を考えたほうが、明解になる場合が多い。

ここでA因子は2つの水準に分かれ、発酵茶無添加(A1)と添加(A2)である。B因子も同様に2つの水準に分かれ、Chol無添加(B1)と添加(B2)である。対照群はA1B1群、発酵茶群はA2B1群、Chol群はA1B2群、発酵茶+Chol群はA2B2群と表記する。このような場合、縦がA因子の影響、横がB因子の影響となり、 2×2 の二元配置分散分析となる。A因子あるいはB因子は増やすことも可能である。たとえば、別の発酵茶(A3因子)を加えれば、縦が3、横が2の6群構成となり、 2×3 の二元配置分散分析となる。ここでは、最も簡単な 2×2 の二元配置分散分析で話を進める。

図58の例では、1. 因子A：発酵茶の影響があるかないか、2. 因子B：Cholの影響があるかないか、それに加えて、3. 因子AとB：発酵茶とCholは影響し合っているかどうかを知りたいことになる(図58右側)。この項目3は日本語では「交互作用」というわかりにくい単語が割り当てられているが、英語では「interaction」であり、「相互作用」のほうがイメージしやすい。この項目3を知ることができる点が、二元配置分散分析の最大の特徴である。もし、2つの因子が互いに影響していなければ、それぞれの因子ごとの影響を考えればよく、互いに影響を与えていれば、2つの因子を切り離しては考えられない。この場合、因子AとBの間には交互作用があると表現する。

二元配置分散分析で何がわかるのか？(図59)

二元配置分散分析にはいくつかの特徴があるが、まず、それを理解するために端的な例を示す。図59には架空のデータの標本平均の棒グラフがある。グラフ上のabcは多重比較を行った場合の検定結果を示している。abcは異なる文字間で有意差があることを意味し、対照群(a)に対して、発酵茶群とChol群は(b)なので有

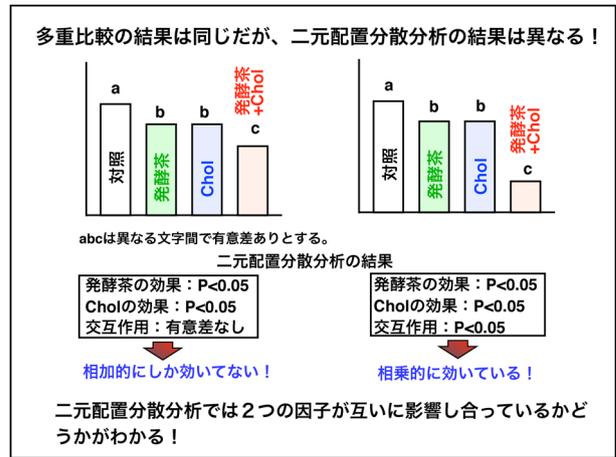


図59 ■ 二元配置分散分析でわかること(1)

意に低値を示し、発酵茶+Chol群は(c)であるから、対照群、発酵茶群、Chol群に対して有意差があり、発酵茶群、Chol群に対してさらに有意に低下したことを示す。右側の棒グラフでは、対照群、発酵茶群、Chol群のデータは左グラフと同じであるが、発酵茶+Chol群がかなりの低値を示している。しかし、多重比較での検定結果は左グラフと同じにしかならず、結論は同じになる。つまり、発酵茶とCholを同時に摂取すると有効性が高まるという結論である。

しかしよくみると、左グラフの結果は、発酵茶とCholで低下作用があり、発酵茶+Chol群では、相加的にさらに低下したと読み取れる。一方、右グラフの結果は、発酵茶+Chol群では劇的に(相乗的に)低値を示しているように見える。このような結果は、多重比較では表現することができないのである。ところが、二元配置分散分析では、図59下側に書いているように、右側では交互作用が有意となっている。これは、発酵茶とCholが相互作用して、相乗的に効果を発揮し低下させたことを意味する。

一方、左のデータでは、交互作用が有意ではない。これは、発酵茶とCholが独立して効果を発揮し、発酵茶+Chol群では相加的に低下したに過ぎないことを示している。

さらに、もし、右グラフの発酵茶+Chol群が、対照群と同じレベルであったとすると(図63のグラフ5を参照)、この場合も交互作用は有意となり、発酵茶とCholは単独では低下作用を示すのに、両者を一緒に摂取すると、相殺作用により低下作用が消失することを意味する。

このように、2つの物質が互いに影響し合っているかどうか、交互作用が有意か否かでわかり、相加作用か、相乗あるいは相殺作用かを判断できる。これが、二

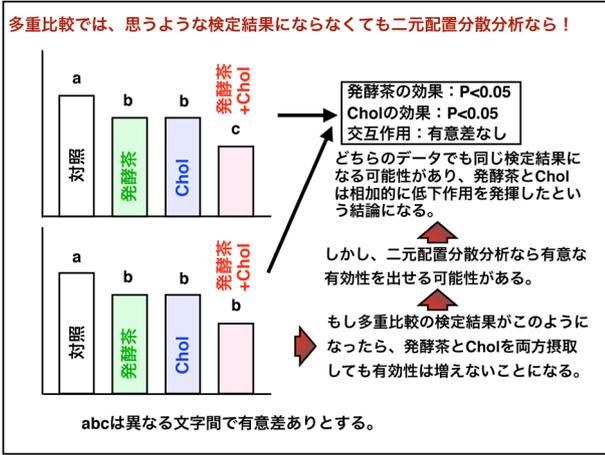


図60 ■ 二元配置分散分析でわかること(2)

元配置分散分析の最大の特徴である*8。

もう一つの例を示す。図60の2つの棒グラフデータは、標本平均は全く同じである。しかし、標本データのばらつきが異なるため、多重比較の結果が違う例である。下のグラフでは、発酵茶+Chol群で低下傾向があるにもかかわらず、発酵茶群とChol群に対して有意差がない。このような結果はよく起こる。多重比較では、群数が多いと検定が厳しくなり有意差が出づらくなることも関係する。

多重比較の検定結果から考察すると、上グラフでは、発酵茶とCholを一緒に摂取すると低下効果が増すといえるが、下グラフでは、発酵茶群、Chol群と発酵茶+Chol群間には有意差がないので、発酵茶とCholを同時に摂取しても有意には低下効果は強まらないという結果になる。そうすると、発酵茶+Chol群では数値的には低下傾向にあるので、「有意差はないが低下傾向がある」と苦しい言いわけをせざるを得なくなる。

このような場合、二元配置分散分析で解析すると、上および下グラフのデータは、どちらも同じ検定結果になる可能性があり、発酵茶+Chol群では相加的な低下作用が認められるという結果になりやすい(図60右上)*8。このように、二元配置分散分析と多重比較で、検定結果によっては、結果の解釈が異なる場合がありうる。どちらの検定結果が真実に近いかは、この1回の試験ではわからないが、基本的には二元配置分散分析の検定結果で解釈したほうが妥当と考えられる。

*8 もちろん、データのばらつき具合が検定結果に影響するので、このようにうまく検定できるかどうかはわからない。あくまでも、端的な例として示した。

二元配置分散分析の基本的考え方

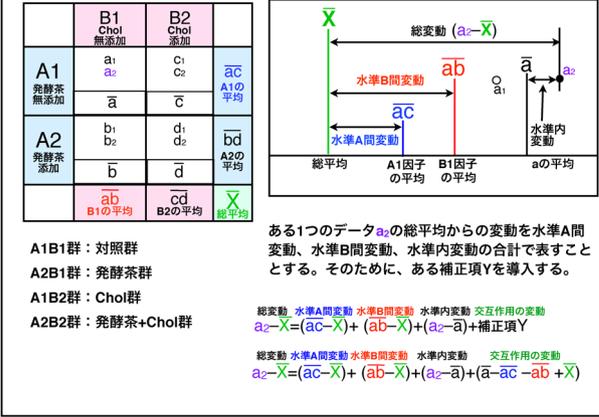


図61 ■ 二元配置分散分析の基本的考え方

二元配置分散分析の有意差の決め方もF検定

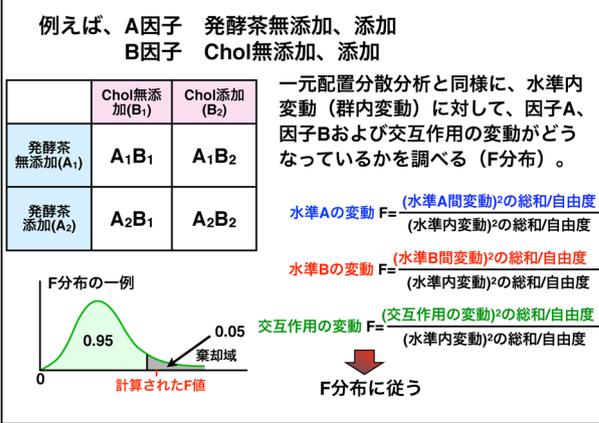


図62 ■ 二元配置分散分析の有意差の決め方

二元配置分散分析の基本的考え方⁽¹⁾ (図61, 62)

二元配置分散分析は難解であり、その原理を理解するには時間がかかる。難しい場合はこの項は読み飛ばしても結構であるが、原理を理解するとどのように利用することも理解しやすいはずである。

図58の左表に個々のデータや平均値を記入したのが図61左上表である。簡単にするために、各群の標本データ数をそれぞれ2としている。

理解のためには、一つひとつ図と見比べて進んで欲しい。

- データは以下の合計8個である。
A1B1群： a_1, a_2
A2B1群： b_1, b_2
A1B2群： c_1, c_2
A2B2群： d_1, d_2
- 平均値は、まず、各群の標本平均 $\bar{a}, \bar{b}, \bar{c}, \bar{d}$ の4つがある。

3. 各水準毎に考えると、水準A1の発酵茶無添加の2群は、A1B1とA1B2群であり、この2群の平均値 \overline{ac} が得られる。

同様に、水準A2の発酵茶添加の2群は、A2B1とA2B2群であり、平均値 \overline{bd} 。

同様に、水準B1のChol無添加の2群は、A1B1とA2B1群であり、平均値 \overline{ab} 。

同様に、水準B2のChol添加の2群は、A1B2とA2B2群であり、平均値 \overline{cd} 。

の合計4つの平均値が得られる。

4. 最後に全体の基準となる総平均を \bar{X} とする。このように、全部で9個の平均値が得られる。
5. そこで、今、A1B1群の標本データ a_2 を例にとって考える。考え方は一元配置分散分析の場合と基本的には同様である。しかし、一元配置とは異なり、因子が2つあるために複雑となる。

図61の右上図で a_2 の位置を確認してほしい。まず、標本データ a_2 が総平均 \bar{X} からどれくらい離れているかを考える。これは総変動と名付けるが、 $(a_2 - \bar{X})$ で表すことができる。基準となる値、ここでは \bar{X} を引き算する。 $(\bar{X} - a_2)$ とはしない) 以下同様。

6. 次に、 a_2 が所属する水準A1の平均値 \overline{ac} が総平均 \bar{X} からどれくらい離れているかを計算すると、 $(\overline{ac} - \bar{X})$ となる。これは水準A間変動と名付ける。
7. 次に、 a_2 が所属する水準B1の平均値 \overline{ab} が総平均 \bar{X} からどれくらい離れているかを計算すると、 $(\overline{ab} - \bar{X})$ となる。これを水準B間変動と名付ける。
8. 次に、 a_2 が、所属するA1B1群の標本平均 \bar{a} からどれくらい離れているかを計算すると、 $(a_2 - \bar{a})$ で計算できる。これは、一元配置分散分析では群内変動と名付けたが、ここでは、水準内変動と名付ける。
9. そこで、総変動を水準A間変動、水準B間変動、水準内変動に分けて考えることとする。一元配置分散分析では、総変動は群間変動と群内変動の合計となり、総変動は群間変動と群内変動に分けることができた。

そこで、総変動と水準A間変動+水準B間変動+水準内変動を式で書いてみる。

総変動は、 $(a_2 - \bar{X})$ であり、水準A間変動+水準B間変動+水準内変動は、 $(\overline{ac} - \bar{X}) + (\overline{ab} - \bar{X}) + (a_2 - \bar{a})$ である。

これら両者は=では結べない。そこで、補正項 Y を導入して、以下の式を作る。

$$(a_2 - \bar{X}) = (\overline{ac} - \bar{X}) + (\overline{ab} - \bar{X}) + (a_2 - \bar{a}) + \text{補正項 } Y \text{ (図61右下)}$$

ここで、式を変形して、補正項 Y を求めると、

$$\begin{aligned} \text{補正項 } Y &= (a_2 - \bar{X}) - (\overline{ac} - \bar{X}) - (\overline{ab} - \bar{X}) - (a_2 - \bar{a}) \\ &= \bar{a} - \overline{ac} - \overline{ab} + \bar{X} \end{aligned}$$

となる。

そこで元の式に補正項 Y の式を代入すると以下となる。(図61右下式)

$$(a_2 - \bar{X}) = (\overline{ac} - \bar{X}) + (\overline{ab} - \bar{X}) + (a_2 - \bar{a}) + (\bar{a} - \overline{ac} - \overline{ab} + \bar{X})$$

この補正項 Y が、交互作用の変動を表す*9。すなわち、

$$\begin{aligned} \text{総変動} &= \text{水準A間変動} + \text{水準B間変動} + \text{水準内変動} \\ &\quad + \text{交互作用の変動} \end{aligned}$$

このようにすることで、総変動は、水準A間変動、水準B間変動、水準内変動、交互作用の変動の4つに分解される。

10. これと同様の計算を、全標本データで行う。(ここでは8個の標本データ)
11. 一元配置分散分析での計算を思い出して欲しいが、次に、全標本データの総変動、水準A間変動、水準B間変動、水準内変動、交互作用の変動を平方して合計するのである(平方和)。そうすると、
- $$\begin{aligned} (\text{総変動})^2 \text{の総和} &= (\text{水準A間変動})^2 \text{の総和} \\ &\quad + (\text{水準B間変動})^2 \text{の総和} + (\text{水準内変動})^2 \text{の総和} \\ &\quad + (\text{交互作用の変動})^2 \text{の総和} \end{aligned}$$

が成り立つ。したがって、一元配置分散分析の場合と同様に、総変動は、水準A間変動、水準B間変動、水準内変動および交互作用の変動だけで考えればよいことになる。

12. この後の検定は一元配置分散分析と同様である(図62)。一元配置分散分析では、自由度を加味して、(群内変動)²の総和を基準(分母)として、(群間変動)²の総和を計算すると、F分布することから、F分布に当てはめて有意差を求めた。二元配置分散分析では、因子が2つあり、しかも、交互作用もあるので、それら3つについての検定が

*9 交互作用の変動を頭のなかでイメージすることは難しい。実例を用い、A, B 2つの因子が全く交互作用をしていない例を作成すると、交互作用の変動はゼロとなる。大きく交互作用している例を作成すると、交互作用の変動は大きな値を示す。実例は文献1を参照されたい。

できる。基準（分母）は、（水準内変動）²の総和であり、（水準A間変動）²の総和、（水準B間変動）²の総和および（交互作用の変動）²の総和を分子として、F値を計算する（図62右下の3つの式）。それぞれは自由度で割るが、これは一元配置分散分析の場合と同様である。このF値はF分布する。

13. F分布での有意差の求め方は一元配置分散分析と同様で、F値がかなり大きくなり、右端5%に入ると有意と判定する（図62左下図）*¹⁰。

二元配置分散分析の検定結果の解釈（図63, 64）

1. 交互作用に有意差がない場合

交互作用に有意差がなく、A因子かB因子あるいはその両方に有意差が得られた場合、AおよびB因子は互いに影響し合っていないことを示すので、A因子およびB因子のみの影響だけを考えればよい。

例として、図63のグラフ1では、水準A（発酵茶）のみが有意の場合、グラフ2では、水準B（Chol）のみが有意の場合、グラフ3では、水準AおよびBともに有意の場合の端的な例である。この場合、たとえばグラフ1では、「発酵茶摂取で統計的に有意な変動が認められた」と表現する。グラフ3では「発酵茶摂取およびChol摂取で統計的に有意な変動が認められたが、交互作用が有意ではないので、発酵茶とCholは相互に影響し合っていない」と表現することができる。

2×2の二元配置分散分析では、交互作用に有意差がない場合は、これ以上の検定を行う必要はない。発酵茶摂取で有意に影響があるということは、水準A1の2群

*¹⁰自由度の計算はかなり面倒である。水準A間変動では、水準数をaとすると、計算式に \bar{X} が入っているため、一つ減って、(a-1)となる。水準B間変動も同様で、水準数をbとすると、(b-1)となる。この例では、水準数はAもBも2なので、自由度はそれぞれ2-1=1である。水準内変動は、各群で標本データから標本平均を引いている。したがって、それぞれの群で自由度が1減るので、4群では自由度は4減ることになる。各群のデータ数はここでは2であるが、これをnとすると、総データ数は、 $a \times b \times n$ となり、そこから群数 $a \times b$ を引くと水準内変動の自由度となる($abn - ab$)。この例では、 $2 \times 2 \times 2 - 2 \times 2 = 4$ となる。問題は交互作用の自由度である。ここで、総変動の自由度を考える。総変動はすべての標本データから \bar{X} を引いているので、自由度は総データ数-1である。つまり、($abn - 1$)であり、この例では7である。総変動は、水準A間変動、水準B間変動、水準内変動、交互作用の変動に分解されたが、自由度も同様に分解できることが知られている。つまり、以下の式のように、総変動の自由度は水準A間変動、水準B間変動、水準内変動、交互作用の変動のそれぞれの自由度の合計となる。 $(abn - 1) = (a - 1) + (b - 1) + (abn - ab) + \text{交互作用の自由度}$ ここでは、 $7 = 1 + 1 + 4 + \text{交互作用の自由度}$ ここから、交互作用の自由度 $= ab - a - b + 1 = (a - 1) \times (b - 1)$ となり、この例では1が得られる。

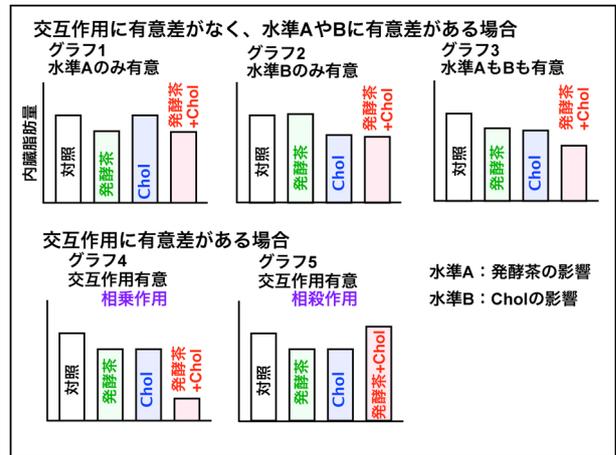


図63 ■ 交互作用に有意差がない場合、有意差がある場合の考え方(1)

交互作用がないとき

因子Aおよび因子Bの影響があるか、ないかが分かる。各群間比較はできない。

因子Aに有意差が出れば、水準A₁と水準A₂間の有意差検定は可能であるが、例えば、A₁B₁とA₂B₁間の検定はできない。

しかし、実際には多重比較を行っている論文あり→間違い。

	Chol無添加(B ₁)	Chol添加(B ₂)
発酵茶無添加(A ₁)	A ₁ B ₁	A ₁ B ₂
発酵茶添加(A ₂)	A ₂ B ₁	A ₂ B ₂

交互作用があるとき

因子Aおよび因子Bの影響は参考程度にしか分からないため、→全群の多重比較を行う（群間比較ができる）

	Chol無添加(B ₁)	Chol添加(B ₂)
発酵茶無添加(A ₁)	A ₁ B ₁	A ₁ B ₂
発酵茶1添加(A ₂)	A ₂ B ₁	A ₂ B ₂
発酵茶2添加(A ₃)	A ₃ B ₁	A ₃ B ₂

図64 ■ 交互作用に有意差がない場合、有意差がある場合の考え方(2)

(A₁B₁とA₁B₂)と、水準A₂の2群(A₂B₁とA₂B₂)の間に有意差があることを意味する。水準B₁とB₂の場合も同様である。なお、決して各群間（ここでは4群）の多重比較を行ってはならない（図64）。二元配置分散分析では発酵茶の有意な影響が認められたにもかかわらず、多重比較を行うと、たとえば、A₁B₂群とA₂B₂群に有意差がないといった検定結果になることが往々にして起こる。つまり、これら2群間では影響がないことになる。どちらの検定結果を採用して論ずればよいのか研究者は迷う。このような場合、多重比較の結果を中心に論じている論文が多いが、そうすると、二元配置分散分析では発酵茶の効果は認められるにもかかわらず、多重比較の結果からCholを一緒に摂取すると発酵茶の効果は消失するという、わけのわからない結論を述べる羽目に陥る。このようなケースでは、「二元配置分散分析の

結果から、発酵茶の有効性が認められた」と論ずればよい。

ただし、 2×3 や 3×3 の二元配置分散分析の場合、たとえば、A因子の発酵茶の水準が3つある場合（図64右下表）、二元配置分散分析で発酵茶の効果が有意となった場合、A1, A2, A3のどの水準同士が有意かを多重比較で検定することはできる。A1水準とA2水準は差がないが、A1水準とA3水準は差があるといった検定である。この検定では、A1水準（A1B1とA1B2）、A2水準（A2B1とA2B2）、A3水準（A3B1とA3B2）の3群として多重比較を行う。この場合も、 2×2 の場合と同様、各群間の多重比較を行ってはならない。

2. 交互作用に有意差がある場合

図63のグラフ4は相乗作用、グラフ5は相殺作用の端的な例である。交互作用に有意差が得られたとき、AとB因子は互いに影響し合っていることになるので、両者を切り離しては考えられない。したがって、この場合はA因子の影響やB因子の影響は参考程度となる。そこで、4群それぞれを独立した群と捉え、横並びの4群として4群の多重比較を行うのである（図64左下）。つまり、交互作用がある場合は多重比較による群間比較が可能となる。筆者の研究分野では、このような二元配置分散分析時の独特の検定方法を理解していない研究者はかなり多く、正しい検定を行った研究発表や論文のほうが少ないように思われる。論文において記述する場合は、「交互作用が有意であったため、全群の多重比較を行ったところ、……」という表現が適切である。

交互作用が有意であるにもかかわらず、全群の多重比較を行うとどの群間にも有意差がないということもありうる。これは群数が多いために多重比較での検定が厳しくなるためである。この場合、少なくとも交互作用は有意であるから、「2つの因子は影響しあっている」と結論することは可能である。

また、交互作用が有意の場合に、全群の多重比較を行うとすると、比較する必要のない群間比較がありうる。たとえば、図64の右下の 2×3 の群構成で、A2B1群とA3B2群間を実験上比較する必要がないという場合である。このような場合は、その検定結果は表示しなければよい。

このように、二元配置分散分析を行った場合、その分析結果を優先して論じ、次に多重比較で論ずるべきである。しかし、多重比較優先で論じた論文が多いのが現状で、二元配置分散分析を行った意味がない。

二元配置分散分析で水準が増えると、検定結果の判断は難しい！（図65）

2×2 の二元配置分散分析は、それぞれ水準は2つしかないで、値が高いか低いか比較的明確である。しかし、図64右下表のように、水準が3つ以上になると、結果がわかりにくくなる可能性がある。図65のグラフを見て欲しい。 2×3 では6群となる。いずれのグラフも、Chol無添加3群を左に、Chol添加3群を右に棒グラフで表示してある。Aでは、赤矢印で示すように、標本平均は並行して変動しているのので、交互作用はない例である。Cholを添加しても、発酵茶の影響は同じである。Bでは、Cholを添加すると発酵茶の影響が増加しているのので、交互作用有意となりうる。問題はCである。Cholを添加しても、発酵茶1では変化がなく、発酵茶2で影響が強まっている。この場合、研究者は交互作用を期待するが、交互作用が有意となるかどうかは、計算してみなければわからない。一元配置分散分析でも触れたが、分散分析では全群のばらつきを考慮するので、1群だけの変動は埋没する可能性がある。もし、交互作用なしと判定されると、発酵茶2の影響は見逃されてしまう可能性がある。たとえば、発酵茶1の2群を行わず、 2×2 で実験すると交互作用ありが明確に示されると考えられる。このように、二元配置分散分析で水準を増やすと、判定が難しくなることがあるので、むやみに水準数を増やさないほうがよい。あいまいな結果の場合は、水準を減らしてやり直すことも視野に入れるべきである。

対応のある二元配置分散分析（図66）

これまで述べてきた二元配置分散分析は、対応のない場合である。対応のある実験では、一方の因子は対応が

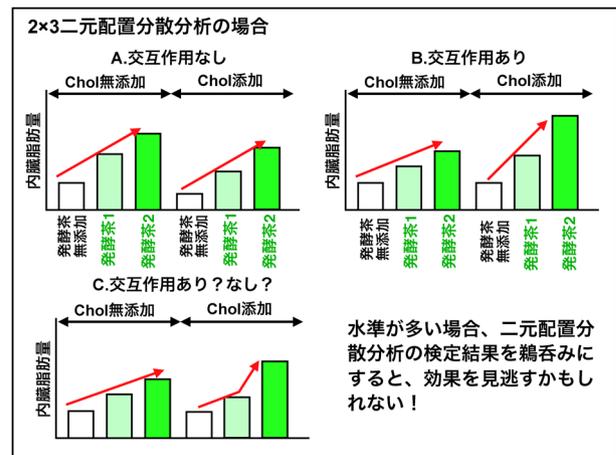


図65 ■ 水準が増えた場合の問題

あり、他方は対応がない場合、あるいは、両方の因子ともに対応がある場合がありうる。ここでは、研究者が頻繁に行うと考えられる実験である一方の因子は対応があり、他方は対応がない場合について簡単に説明する。この解析法が書かれている一般向け統計書は少ないが、文献6に実例とともに手順が示されている。しかし、理論的説明が少ないためかなり難解である。

図66のように、たとえば、マウスを3群とし経時的に採血して、あるパラメータを測定する場合である。3群間に対応がないが、経過時間には対応がある場合となる。

このような場合の二元配置分散分析法として、**重複測定→二元配置分散分析**がある。詳細は説明しないが、検定の考え方は通常の二元配置分散分析と同じである。しかし、二元配置分散分析の場合と同様に検定を行うと、研究者の意図しない検定結果になる可能性がある。

二元配置分散分析の検定では、**交互作用がない場合**は、多重比較を2つの因子（3群間の水準と4点の時間経過の水準）に対して行うことになる（水準毎の比較であり、各群間比較ではない）。対照群、A添加群、B添加群の3つの水準についての検定は、各水準の4点の時間点のデータを合計して、3水準で検定することになる。しかし、この検定は必ずしも研究者が意図する検定ではない場合が多い。研究者は各時間点の3群の比較ができればよいのであって、各時間点を合計する意味はないのである。しかし、交互作用がないと、正式には各時間点の3群の比較はできない。

時間経過の影響を知りたい場合では、「対応がある」ので多重比較は行えない。無理に行うとすると代用法を用いることはすでに述べた。この場合も、各時間点の3群の合計で検定するので、研究者の意図する検定ではな

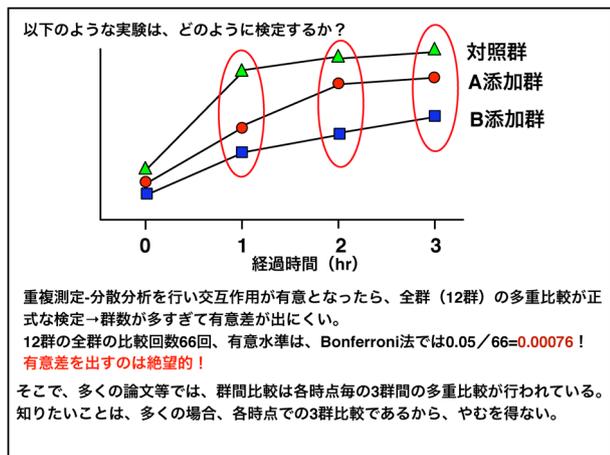


図66 ■ 対応のある二元配置分散分析での解析法

いかかもしれない。そもそも時間経過の影響を知りたいとは考えていない場合も多々ある。

一方、図66のグラフのようなデータでは、多くの場合、交互作用が有意となると考えられる。**交互作用がある場合は、二元配置分散分析では全群の多重比較を行うことになる。**しかし、図66の例では全群数は12群である。12群全群を比較すると比較回数は66回にもなる！Bonferroni法で検定すると有意水準 $0.05/66=0.00076$ となり、有意差をえるのは絶望的である。時間点がさらに増えると天文学的な比較回数となり、全群の多重比較はあまりにも検定が厳しくなりすぎる。

このように、通常の検定法の手順では、研究者の意図とかけ離れた検定となる場合があり、適切な検定とはいえない。しかし、残念ながらちょうどよい検定法は存在しない。

このような実験の場合、各時間毎に3群間の多重比較を繰り返し行う研究者が多いのではないだろうか？図66の赤の楕円内の比較の繰り返しである。その場合、**多重性の問題を考慮すると、検定が甘くなる**ことはこのセミナーの読者であればわかるはずである。筆者は、比較したいのが群間のみで、時間経過は参考程度で重要でなければ、群間（ここでは3群間）の多重比較（あるいは、2群であればt検定）の繰り返しでよいと割り切ることになっている。このように、複雑な実験系になると、適切な検定法が見当たらない場合はある。どうしても全群の比較が必要な実験では、群数や時間点数を減らすことを考えた方がよい*11。

残念ながら、この種の実験の検定法に関して詳しく解説した統計書はないようであり、これ以上解説することができない。この種の実験は比較的多く行われており、適当に時間点を振って実験を行うが、統計学的にはこのような問題をはらんでいることを認識しておくべきである。

なお、**二元配置分散分析に相当するノンパラメトリック検定は存在しない。**したがって、パラメトリック検定の二元配置分散分析で代用するしかない。正規分布しないことがすでにわかっている母集団からの標本を用いて、二元配置分散分析を必要とする試験は実施しないほうがよい。

*11 本来なら、予備試験で最も大きな差が認められる時間を決定し、0時間とその時間の2点（たとえば3時間後）で実験を行う方が明確な検定ができる。しかし、時間経過でのパラメータの変動を追跡する実験の場合、2点だけとるわけにいかない場合もある。

外れ値の取り扱い

主観で外れ値を外すことは許されない！

外すには理由が必要

- ・測定ミス、計算ミスなどのチェック
- ・再測定
- ・データの吟味（他のパラメータに異常はないか？など）
- ・棄却検定
- ・ノンパラメトリック検定

再実験！

外れ値は図示すると
わかりやすい。

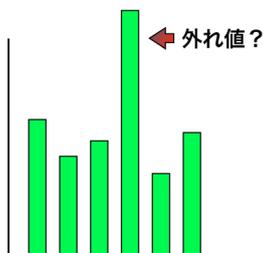


図67 ■ 外れ値の取り扱い

外れ値の取り扱い（図67）

外れ値は図示するとわかりやすい（図67）。研究データには外れ値はつきものであり、取り扱いには苦労する。動物試験では比較的よく起こる。外れ値のあるなしで、統計検定結果が異なる場合が問題となる。人間はミスの多い動物である。筆者は、外れ値があるとまず測定ミス、記入ミス、パソコンへの打ち込みミス、Excelでの計算ミスなどを疑う。測定ミスが疑われれば、再測定する。これで解決すれば問題はない。

次に、動物試験であれば、飼育中に摂食量や成長に異常がなかったか、絶水になったりしなかったかなどを調べる。また、同じ個体のそのほかの測定パラメータに異常がないかを考える。このように実験遂行上や個体独自の問題で、何らかの異常が認められれば、それを理由にデータを外す。場合によっては、その個体のすべての測定データを外す場合もある。

これに平行して実験結果を吟味する。外れ値があるために、有意差がでる場合と有意差がでない場合の2つが考えられるので、どちらが妥当な判断かを考える。そのための材料として、パラメトリック検定およびノンパラメトリック検定両方を行ってみて、検定結果から考えることもある。パラメトリックとノンパラメトリック法で検定結果が同じであれば、それを採用する。検定結果が異なる場合は、どちらが妥当かは簡単には判断できない。

また、棄却検定を行って、そのデータを外せるかどうかを判断し（外すことが妥当かどうかは別として、外す理由にはなる）、その結果に従うこともありうるが、生物材料の場合安易に外すことには異論がある³⁾。いずれにしても、できる限り主観が働かないように、客観的に

実験計画を立てるときに考えるべきポイント

- ・実験計画は統計的手法を考慮して立てるのが基本。
複雑な実験系は適切な検定法がない場合があり、有意差も得られにくくなるので、実験系はシンプルに！
群数は出来るだけ少なく（理想は2群）。
統計を理解すれば有意差の得られやすい計画は可能。
- ・nは多いほど有意差は出やすい。
- ・ノンパラメトリック法も考慮する。
- ・多重性の問題を心の隅に置こう！
むやみに測定項目を増やさない。
全く同じ実験は二度行うべきではない。
- ・再現性を調べる実験も必要なことを心に留めておく。

図68 ■ 実験計画を立てるときに考えるべきポイント

判断すべきである。どうにも判断がつかない場合は、角度を変えた別の試験を行って判断したほうが妥当である。

実験計画を立てるときに考えるべきポイント（図68）

図68に実験計画を立てるときのルールをまとめた。

ほとんどは述べてきたことばかりであり、読んでいただければわかると思う。基本的に、実験計画は統計的手法を考慮して立案すべきである。統計の原理をある程度理解すると、有意差の得やすい実験計画を立案することも可能であることは本セミナー全体を通じてご理解いただけたと思う。

おわりに

本セミナーでは筆者が通常用いる検定法を中心に執筆した。そのほかにも多種多様な検定法があるが、本セミナーの基本を身につけていれば、分布や計算式が違っていても、実は検定の流れは同じであり、理解はそれほど難しくはないはずである。

なお、本セミナーは統計を利用する生命科学系研究者がわかりやすいように、かなり明解に切り分けた書き方をしているが、実際はそれほど簡単ではない。興味があれば、より難易度の高い統計書に挑戦してほしい。

本セミナーが、読者の統計検定に対する理解に少しでも貢献できたならば、筆者にとって喜びである。

文献

- 1) 石村貞夫, 石村光資郎: “入門はじめての分散分析と多重比較”, 東京図書, 2008.

- 2) 池田郁男：“実験で使うとこだけ生物統計2 キホンのホン 改訂版”，羊土社，2017.
- 3) 足立堅一：“らくらく生物統計学”，中山書店，1998.
- 4) 永田 靖，吉田道弘：“統計的多重比較法の基礎”，サイエンティスト社，1997.
- 5) 林 智幸，新見直子：厳格化の観点からの多重比較法の整理，広島大学大学院教育研究科紀要，**54**, 189, (2005).
- 6) 石村貞夫，石村光資郎：“SPSSによる分散分析と多重比較の手順 第4版”，東京図書，2011.

プロフィール



池田 郁男 (Ikuo IKEDA)

＜略歴＞1975年九州大学農学部食糧化学工学科卒業／1980年同大学大学院農学研究科博士後期課程修了，農学博士／同年同大学農学部助手，助教授を経て2005年より東北大学大学院農学研究科教授／2018年より同大学未来科学技術共同研究センター教授，現在に至る＜研究テーマと抱負＞食品成分および栄養素の肥満，動脈硬化症予防作用に関する研究，ステロール吸収機構の解明＜趣味＞以前はバドミントンでしたが，今はドライブ&温泉巡り

Copyright © 2019 公益社団法人日本農芸化学会
DOI: 10.1271/kagakutoseibutsu.57.629