分散分析を始める前に

小杉 考司

分散分析は、三群以上の平均値の比較を行う検定法である。まずは分析の大まかな流れを理解し、その後で具体的な計算方法について学習しよう。

1　用語の統一と分析の流れ

1.1　何故 '分散' 分析なのか

既に述べたように、分散分析は**三群以上の平均値の比較**を行うものである。

賢明なる読者はお気づきのことだろうが、これらの狙いはあくまでも「各群の**平均値**に差があるかどうか。差が誤差ではないといって良いかどうか」にある。であるのに、分析方法の名前は**分散**分析なのである。平均なの？分散なの？どうしてそうなるの？となってしまうのではないだろうか。

平均値の差があるとして、それに統計的な意味があると考えられる根拠は、ではどこにあるのだろうか。表 1 を見てみよう。これは三つのクラスで試験をやったときの、点数の平均値である。一組が 65.74 点、二組が 55.47 点だから、ここには 10.27 点分の差があり、一組の方が二組よりも優れている、という結果に飛びつきたくなる。しかし問題は、これが統計的に意味がある差かどうかである。さて、各組とも生徒数が 1 ではないから、これらのデータは分散を持っている。

表 1　三群の平均値

	一組	二組	三組
テストの点	65.74	55.47	49.30

成績だけから見れば、一組は二組よりも出来がいい。しかし、一組の中でも点数が低かった子がいるだろう。逆に、二組にも点数が良かった子がいるに違いない。各組の点数を図にしてみたのが図 1 である。これを見ると、どの三つの組もたいして違いがないようである。二組で最も点数が良かった子は、一組で最も点数が良かった子よりも、上に位置している。一組で最も点数の悪かった子は、二組の最低点よりも更に悪い。こういうデータを見てしまうと、「ほんとうに一組は二組よりも優れている、なんていっていいのかな」と思えてくるだろう。では、平均点は表 1 と全く同じで、図 2 のような場合はどうだろうか？ この例は、一組の最高点は三つのクラスの最高点であり、二組の最高点でさえ届いていない。また、一組の最低点は二組や三組のそれよりは

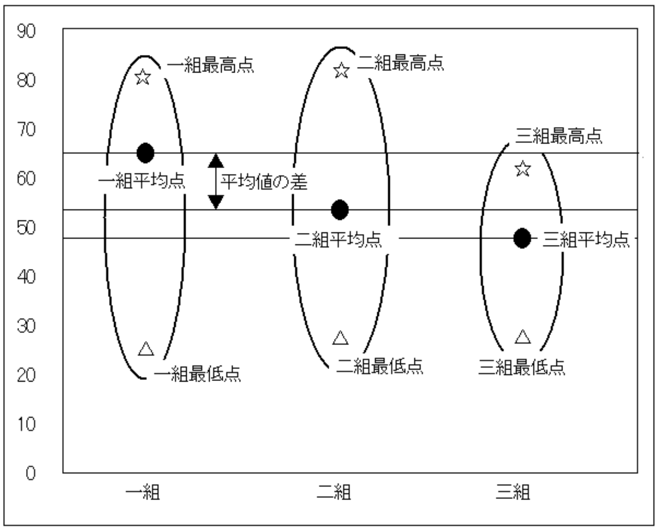

図 1　三つの組の平均値、最大値、最小値

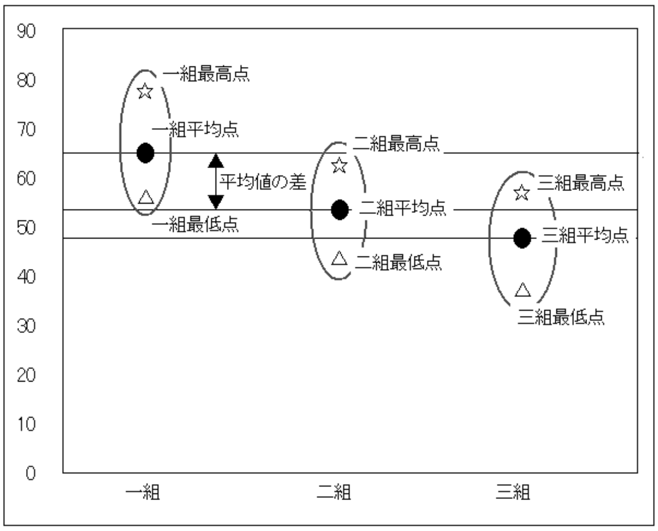

図 2　三つの組の平均値、最大値、最小値・その 2

グッとましである。こういうデータであれば、なるほど、一組は先生の教え方がいいとか何とかで、他よりも優れているんじゃないか、と言えそうだ。

図 1 や図 2 の楕円はデータの散らばりを表しているが、こういった、散らばりの大きさを比較しなければ、ほんとうに平均値の差に意味があるかどうかはわからない。**平均値の差があるかどうかは、データの分散にかかっている**、といえるのだ。

分散分析は、ここでいう一組、二組、三組というグループの違いによる散らばりが、クラスの中での散らばりよりも、十分に大きいかどうか、で「一組は優れている」といった結果を出そうとする。つまり、クラスの中での散らばりが大変大きなものであれば (最大値と最小値の幅が大

きい)、どこかが優れているかどうかといった説明はできない、と考える。心理学の実験に於ける「操作」や「誤差」という言葉を用いて言うならば、操作による効果 (の分散) が誤差 (の分散) に比して大きいかどうか、を見るのである。だから**分散**分析なのである。

1.2 要因と水準、間と内

1.2.1 要因と水準

よくある学生の問い方。「二要因三水準の分散分析をするのですが・・・。」実は、これだけ言われてもどのようにアドヴァイスするべきか、言葉が見あたらない。これだけでは十分な意味をなさないからだ。

要因、英語では factor だが、と呼ばれるのは、平均値を左右すると思われる要素のことである。先ほどの例では、一組、二組、三組といった組わけに何か意味がありそうだったから、これは「組」あるいは「組み分け」を要因としたデータであるといえる。一方、水準とは、英語で level であり、これが意味するのは「組」という要因をいくつに分けているか、という意味である。一組と二組しかないのであれば、これは二水準である。四組まであれば、四水準だ。これらを併せて、一要因三水準のデータ、という呼び方をする。

さて、分散分析は三群以上の平均値の差の検定をするものであるが[*1]、常に一要因とは限らない。ふたつ以上の要因が絡み合うデータが考えられるのである。例えば一組、二組、三組の成績がこんなにも違ってきたのは、担当教員の教え方に違いがあることも考えられるし、使っている教科書が違ったのかもしれない。とすると、要因は「教師」と「教科書」のふたつになる。それぞれ二水準だったとすると、全ての組み合わせ、すなわち教師 α で教科書 A の場合、教師 α で教科書 B の場合、教師 β で教科書 A の場合、教師 β で教科書 B の場合、に当たるデータを収集して、教師と教科書、どちらの効果があったのか (あるいはその両方の効果なのか) を確かめなければならない[*2]。

ともかく、群が三つ以上になるというのは水準が増えるだけではなくて、要因が増えて、その組み合わせの結果、比較すべき群が増えていくことも考えられる。水準は要因ひとつひとつにあるものだから、最初にあげた「二要因三水準の・・・」という表現では言葉が足りないのである。「二要因」ということであるなら、水準は「三水準と二水準」とか「二水準と四水準」といった表現になるだろう。英語では簡単に、2×3 の分散分析、という表現をすることが多い。

1.2.2 間と内

t 検定のところで「対応のある t 検定」と「対応のない t 検定」に区別したが、分散分析でも同様の区別が必要である。左記の例であった、一組、二組、三組という分け方は、それぞれの生徒が入れ替わるわけではないから、対応がない群になる。分散分析では特に被験者間計画 (Between

---

*1 一要因二水準の分析は t 検定に任せよう。

*2 データは必ずしも全ての組み合わせで集めなければならないというわけではない。大村著『実験計画と分散分析のはなし』を参照のこと。

Groups) という。

これに対して、勉強前と勉強後の成績の変化をみる、といった場合は、被験者が繰り返しデータを提供するわけだから、対応のある群になる。これを分散分析では、被験者内計画 (Within Groups) という。

これら「内」であるか、「間」であるか、は要因に関連する。クラスの違いというのは間の要因 (間要因) であり、試験前後の成績というのは内の要因 (内要因) である。ということは、二要因以上の計画であれば、間要因 × 間要因なのか、内要因 × 内要因なのか、間要因 × 内要因なのか、を区別しなければならない。特に最後のものは、**混合計画**とよばれる。

ということは、この節の冒頭にあった「二要因三水準の分散分析で・・・」という表現は、「2×3 の被験者間要因で・・・」とか「3 × 2 の混合計画で・・・」という表現をするのが正しかった、ということがわかってもらえるだろう[*3]。

実際の計算に当たっては、間と内の違いはどのように誤差を算出するか、という点で大きな違いが生じてくる。そのため、たかが用語といわずに、しっかりと使い分けられるようになっておく必要がある。

1.3 下位検定

分散分析が終わったら、下位検定を行う必要がある。

分散分析の帰無仮説は、「全ての群平均の間に、有意な差はない」というものである。言い換えるならば、$\mu_1 = \mu_2 = \mu_3$、というものになるわけだ。さて、分析の結果、この帰無仮説が棄却されたとしよう。なぜこれで終わりではないかというと、例えば三水準の場合でも、

$$\mu_1 \neq \mu_2 = \mu_3$$

$$\mu_1 = \mu_2 \neq \mu_3$$

$$\mu_1 = \mu_3 \neq \mu_2$$

$$\mu_1 \neq \mu_2 \neq \mu_3$$

のどの状態にあるのかがわからないからである。分散分析は、あくまでも $\mu_1 = \mu_2 = \mu_3$ じゃない、というだけで、どこに差があるのかまでは教えてくれないからだ。

「なあんだ、じゃあ t 検定を繰り返してやるよ。結局同じことだろう」と考えるのは早計である。一般に、繰り返し検定を行うと検定の多重性の問題、というのが生じる。

統計では、危険率 α% で帰無仮説を棄却するかどうかを考える。例えば 5% の危険率であれば、間違えてしまう可能性が 5% ある、というものだ。逆にあっている確率は $1 - 0.05 = 0.95$ だから、検定を 10 回繰り替えそうものなら、$1 - 0.95^{10} = 0.4013$ から、全部正しい結果が出る確率が約 40% しかないことになる。

[*3] 心理学の研究室では、卒論の時期が近づくと「2 × 3 のカンカンでな・・・」とか「うん、ナイナイカンの計画で」といった暗号 (?) が飛び交う。

そのため、ひとつずつのペアを繰り返して検定するのではなく、あるひとつの検定量を元に「有意な差があるかどうか」を検定する**多重比較**をするべきなのである。多重比較は様々な手法があり、詳しくは他書に譲る[*4]。

[*4] 永田靖・吉田道弘（著）統計的多重比較法の基礎, サイエンティスト社が良書。