

# 平均差を分散で吟味する

分散分析とは、数ある統計的手法の中でも用いられる頻度が最も高い手法である。t 検定のように、検定できるデータが 1 要因 2 水準（例えば、性別による反応の違いを検定する場合など）だけという制限もなく、1 要因以上 2 水準以上（例えば、性別と世代の組合せの効果を見る）の平均値を比較できる大変便利な手法である。詳しい解説は統計学の教科書に譲ることにして、分散分析の基本的な考え方・使い方を中心に説明してみよう。

## 分散とは

大量にあるデータがどのような性質をもっているかを把握するために、少数の代表値を用いることがある。たとえば、平均値 (mean) や、中央値 (median)、最頻値 (mode) などである。しかしながら、これらの代表値だけではデータの特徴を記述するのに不十分である。たとえば、ある条件とある条件との平均値が大きく異なっているように見えた場合にも、一方の条件にハズレ値があつてそのため平均が異なっていたならば、条件間で差があるとは言い難い。このような不具合を解消するために、データの分布の特徴を記述する方法として中心的データである平均値からの散らばり具合を把握する必要があり、その代表値として分散 (variance または  $s^2$ ) ならびに標準偏差 (Standard Deviation, STD または  $s$ ) が用いられる。平均値と分散は以下の式によって定義される。 $x_i$  とは全体で  $N$  人いる被験者の  $i$  人目の被験者のデータの意味である。

$$\text{平均値} \cdots \text{Mean} = \bar{x} = \frac{1}{N} \sum_{i=1}^N x_i, \quad \text{分散} \cdots \text{Variance} = s^2 = \frac{1}{N} \sum_{i=1}^N (x_i - \bar{x})^2$$

なお、散布度を表す値として、平均との差の絶対値を用いることも良さそうに見えるかも知れないが、絶対値を用いた計算は煩雑になりやすく式の展開がしにくいという欠点があるため、上記の平均との差の二乗和が採用されている。また、全体を被験者の人数である  $N$  で割るのは、標本の大きさに依存しない指標である必要があるからである。分散はズレが二乗されているので、その平方根をとったものである標準偏差が散らばり具合を代表する値として用いられることもある。入学試験などで用いられる偏差値とは、この標準偏差を数値変換して導出した値である。

## 分散分析の理屈

たとえば、以下のような実験データが得られたとしよう。

		試行数					平均値
		$T_1$	$T_2$	$T_3$	$T_4$	$T_5$	
被験者 ↓	$S_1$	10	5	$X_{13}$	3	3	$X_{1.}$
	$S_2$	20	4	$X_{23}$	3	3	$X_{2.}$
	$S_3$	15	10	$X_{33}$	3	3	$X_{3.}$
	⋮	⋮	⋮	⋮	⋮	⋮	⋮
	⋮	⋮	⋮	⋮	⋮	⋮	⋮
	$S_n$	20	8	$X_{n3}$	3	3	$X_{n.}$
平均値		$X_{.1}$	$X_{.2}$	$X_{.3}$	$X_{.4}$	$X_{.5}$	$X_{..}$

各個人の解答時間  $x_{ij}$  ( $j$  試行目の  $i$  さんのデータ) が以下のようなモデルによって表現可能であるとする。このとき、 $\mu$  を母集団 (被験者となりうる全ての人間) の平均値、 $\alpha_j$  を試行数の効果、 $\epsilon_{ij}$  を実験とは関係のない誤差 (個人差や、体調の変化、PC の不調など) とする。

$$x_{ij} = \mu + \alpha_j + \epsilon_{ij}$$

$\mu$  を全被験者の全試行分のデータの平均  $X_{..}$  (標本平均) に、 $\alpha_j = X_{.j} - X_{..}$  に置き換え、次に、 $j$  番目の試行の効果 (繰り返し学習の効果) が全て個人に同じ効果  $\alpha_j$  を及ぼすという仮定すると、上記の式を以下のような標本に関する式に変換することができる。

$$x_{ij} - x_{..} = (x_{.j} - x_{..}) + (x_{ij} - x_{.j})$$

この式の左辺はデータ全体の分散を、右辺第一項は水準間の変動 (繰り返し学習の効果 = 試行間の変動) を、右辺第二項は水準内部での変動 (試行内部での変動 = 個人差など = 誤差) を表しているといえる。上記の式の両辺を二乗して、 $i$  と  $j$  につ

いての総和をとると、以下の式になる。pは処理水準の数（試行数・今回は5）を、nは被験者を表している。

$$\begin{aligned} \sum_1^n \sum_1^p (x_{ij} - x_{..})^2 &= \sum_1^n \sum_1^p [(x_{.j} - x_{..}) + (x_{ij} - x_{.j})]^2 \\ &= \sum_1^n \sum_1^p [(x_{.j} - x_{..})^2 + 2(x_{.j} - x_{..})(x_{ij} - x_{.j}) + (x_{ij} - x_{.j})^2] \\ &= n \sum_1^p (x_{.j} - x_{..})^2 + \sum_1^n \sum_1^p (x_{ij} - x_{.j})^2 \end{aligned}$$

この式から実験データの全変動が第1項「実験統制による変動（水準間変動＝繰り返し学習の効果＝SSa）」と第2項「実験誤差による変動（SSw）」とに分解されることが分かる。このとき、分散分析の理屈とは、「水準間変動による分散と実験誤差による分散を比較して、前者がかなり大きなものであれば、実験でコントロールした処理の効果があつたといえるのではないだろうか？」という単純なものである。検定では、上記のそれぞれの変動をそれぞれの自由度（データの個数-1）で割った値の比を用いて推論する。その式とは以下のようなものである。

$$F[p-1, p(n-1)] = \frac{SSa/p-1}{SSw/p(n-1)}$$

すなわち、分子の項が分母の項と比較してとても大きければ条件操作が成功したと判断し、小さければ失敗したと判断するわけである。<sup>2</sup>たとえば、Fの値が1よりも小さい場合には、実験とは関係ない要因による変動の方が、条件操作による変動よりも大きかったことを表しており、実験が失敗であったことを示している。

## 分散分析で何が分かるのか

たとえば、分散分析の結果を論文などに記載するときは、以下のように記述する。

略) ... 試行の繰り返しの効果が有意であることが示された ( $F_{(4,90)}=5.63, p<.05$ ) ...

「 $F_{(4,90)}$ 」の括弧内の数字は、それぞれ試行数の自由度と誤差項の自由度である（各有意水準でのFの臨界値を求めるのに用いるF分布が、これらの2つの自由度によって定まるからである）。また、「 $p<.05$ 」は、生起確率<sup>3</sup>が5%よりも小さいという意味であり、5%水準で有意であることを示す（1%水準で有意である場合には、「 $p<.01$ 」と表記する）。あるいは、p値をそのまま表記する場合もある。その場合は、「 $p=.002$ 」と表記する。これとは異なり、有意ではなかった場合には、「 $p>.10$ 」などの代わりに、「n.s.」（no significance, 有意ではない）と表記したり、単に $F<1.0$ と記述することもある。さらに、場合によっては、誤差項の分散を「 $MSe=13.83$ 」のように表記し、F値に併記することもある。いずれにしても、統計的分析結果は、t検定、 $\chi^2$ 乗検定、分散分析など、どのような分析をしても「自由度、統計値、確率値」の3つを最低限掲載する必要がある。これは、これらの3つの数値から、誤差項となる分散、要因の効果を示す分散など、実験結果全体の大まかなプロフィールを、読者がつかむことができるからである。

ただし、ここで重要なことは、分散分析で分かることが「繰り返しの効果がある」ということだけである点である。この場合の帰無仮説は、「比較している全ての平均値が等しい」こと、すなわち、「 $A=B=C$ 」ということである。それに対して対立仮説は、「比較している平均値は等しくない」ことであるから、対立仮説を採択するといっても「 $A \neq B=C$ 」かも知れないし、「 $A=B \neq C$ 」かも知れないし、あるいは「 $A \neq B \neq C$ 」かも知れない。たとえば、今回の実験でも、試行を繰り返すことによって学習効果が上がるかどうかは事前には明らかでない。要するに、5つの平均値のパターンがどのようになるのかが未確定なので、それを明らかにする必要が生じてくるのである。そこで、3水準以上の要因で有意差が認められた場合は、AとB、BとC、そしてAとCというように2水準（ペア）ずつ取り出してその差を検定するという手続きを行う。これが多重比較（Multiple Comparison）である。多重比較には、様々な手法が用意されているが、分析の目的などから手法を選んでいくことになる。一般的には、全てのペアを総当たりで検定するTukey法やScheffe法を使うことが多い。<sup>4</sup>以下の式はTukey法を用い

<sup>1</sup>  $\sum_1^n \sum_1^p x_{ij} = n \times x_{.j} = \sum_1^n x_{i.j}$  であり  $\sum_1^n \sum_1^p (x_{.j} - x_{..}) = n \sum_1^p (x_{.j} - x_{..})$  であるため。

<sup>2</sup> 今回の実験を正しく分析するならば、乱塊法に基づく一要因の分散分析を行う必要がある。しかし、説明のわかりやすさを優先して完全無作為法を採用した。意欲的な方は教科書などを参考にして被験者を要因に含めた「2要因の分散分析（交互作用なし）」に挑戦してもらいたい。

<sup>3</sup> 生起確率とは、帰無仮説を間違えて棄却する確率、すなわち、本当は条件操作の効果がなかったのに関わらず間違えて効果があつたと判定してしまう確率を表している。

<sup>4</sup> 通常のt検定を繰り返し適用すれば問題ないと思うかも知れないが、これは間違いである。たとえば、信頼区間を95%（危険率=5%）としたt検定を3回繰り返すと、統計的推論の妥当性が最終的には $0.95^3=0.857$ となり、統計的推論の全体的な信頼性が大幅に低くなることが分かる。このような推論エラー（第1種の過誤）を防ぐために開発された統計手法が、Tukey法やScheffe法といった多重比較である。

た多重比較の計算式である.

$$HSD = q_{\alpha, m, df} \sqrt{\frac{MS_w}{n}}$$

計算式における  $\alpha$ ,  $m$ ,  $df$  は, それぞれ有意水準, 比較に用いる平均値の数 (水準数 = 5), 誤差項の自由度 (18 前後) を意味しており, 別表 E を見て係数  $q$  の値を調べるときに用いる. また,  $n$  には被験者数 (12 または 13) を,  $MS_w$  は誤差項 (Error 行の Mean) を代入する. 多重比較の結果の記述方法は以下の通りである. **HSD** や **MSe** が併記されることもある.

解答に要した時間について, 試行数を要因とした一要因の分散分析を行ったところ主効果が認められた ( $F_{(4,90)}=10.43, p<.01$ ). この解答時間の違いを明確にするため, 有意水準 5% で Tukey 法によって多重比較を行った. その結果, 第 1 試行と第 2 試行... との間に有意差が認められた. これらの結果から, 第 2 試行以降と比べて第 1 試行は有意に長い解答時間を要したことが分かる.

### 計算手順

		試行数 →				
		T <sub>1</sub>	T <sub>2</sub>	T <sub>3</sub>	T <sub>4</sub>	T <sub>5</sub>
被験者 ↓	S <sub>1</sub>	10	5	4	3	3
	S <sub>2</sub>	20	4	7	3	3
	⋮	⋮	⋮	⋮	⋮	⋮
	S <sub>10</sub>	20	8	4	3	3
計算手順	N	10	10	10	10	10
	$\sum_1^n x/n$	10	5	9	8	8
	$\sum_1^n (x - x_{.j})^2$	1000	500	405	320	320

基本的には,  $n \sum_1^p (x_{.j} - x_{..})^2$  と  $\sum_1^n \sum_1^p (x_{ij} - x_{.j})^2$  を求め, あとはそれぞれの数値をそれぞれの自由度である  $(p - 1)$  と

$p (n - 1)$  で割ってやればよい. 上記の表では,  $\sum_1^n x/n$  が  $X_{.j}$  に相当している.

参考のために, 誤差項の中から個体差を取り除く (各被験者の能力のばらつきを除外する) モデル式を下に記しておく.

$$x_{ij} - x_{..} = (x_{.j} - x_{..}) + (x_{i.} - x_{..}) + (x_{ij} - x_{.j} - x_{i.} + x_{..})$$

右辺第 1 項が要因差, 第 2 項が個体差, 第 3 項が誤差を表しており, 実験統制による変動の大きさを見るときには, 第 1 項を分子に第 3 項を分母にしたような分散比の検定を行えばよい. 自由度も, それぞれの分散を求めるときに使用したものと同じ値となる.