

先週の講義では、F検定の具体例として、下記の例を挙げて、計算手順を説明しました。念のため、再掲します。

表1に、実験区1に6個、実験区2群に7個、実験区3群に5個、実験区4に6個のデータセットの例を示しました。この平均値間に差があるかどうかを検討します。

表 1.1 要因分散分析の例

実験区			
1	2	3	4
2	10	8	9
5	2	7	15
3	4	3	8
8	9	4	12
9	13	5	13
4	14		4
	15		

計算手順を以下の通りです。

1. 全平方和(SS_{total})を計算する
2. 各水準ごとの平均値を計算する
3. 残差平方和計算する
4. 全平方和から残差平方和を差し引いて、水準間の平方和とする
5. 全自由度と水準間の自由度の差として、残差自由度を求める
6. 残差平方和を残差の自由度で割って、残差分散を求める
7. 水準間の平方和を水準間の自由度で割って、水準間の分散をもとめる
8. 水準間の分散を残差自由度で割って、これをF値とする
9. 判定のための危険率を定め、水準間の自由度を分子の自由度、残差の自由度を分母の自由度として、F臨界値の表などを使って、有意性を判定する。

表 2.計算のための Excel シート

	A	B	C	D	合計	
	2	10	8	9		
	5	2	7	15		
	3	4	3	8		
	8	9	4	12		
	9	13	5	13		
	4	14		4		
		15				
n_i	6	7	5	6	24	N
T_i	31	67	27	61	186	T
M_i	5.166667	9.571429	5.4	10.16667		
s_i	199	791	163	699	1852	S
T_i^2/n_i	160.1667	641.2857	145.8	620.1667	1567.419	$\sum t_i^2/n_i$

表中の記号の説明

一般化して、実験区を $i = 1, \dots, m$ とします。

n_i : 実験区 i のデータ数

T_i : 実験区 i のデータの合計

M_i : 実験区 i の平均値

S_i : 実験区 i のデータの平方和 $\sum x^2$

何回か出てきた SS の簡便な計算式を使います。

$$SS_i = \sum_{j=1}^{n_j} x_{ij}^2 - \frac{1}{n} \left(\sum_{j=1}^{n_j} x_{ij} \right)^2$$

ここで

$$S_i = \sum_{j=1}^{n_i} x_{ij}^2$$

と表すことにすると

$$SS_i = S_i - \frac{T_i^2}{n_i}$$

これを各グループの残差平方和といいます。

この値のすべてのグループについての和は

$$\begin{aligned} S_1 - \frac{T_1^2}{n_1} + S_2 - \frac{T_2^2}{n_2} + \dots + S_m - \frac{T_m^2}{n_m} \\ = S_1 + S_2 + \dots + S_m - \left(\frac{T_1^2}{n_1} + \frac{T_2^2}{n_2} + \dots + \frac{T_m^2}{n_m} \right) \end{aligned}$$

これは全体の残差平方和の合計です。

二乗和の総和は

$$S = S_1 + S_2 + \dots + S_m$$

ですから

全体の残差平方和は

$$SS_{residual} = S - \left(\frac{T_1^2}{n_1} + \frac{T_2^2}{n_2} + \dots + \frac{T_m^2}{n_m} \right) = S - \sum_{i=1}^m \frac{T_i^2}{n_i}$$

一方全体の SS は SS_{total}

$$SS_{total} = S - \frac{T^2}{N}$$

です。

$$SS_{total} = SS_{\text{実験区}} + SS_{residual}$$

ですから、

$$\begin{aligned} SS_{\text{実験区}} &= SS_{total} - SS_{residual} = S - \frac{T^2}{N} - \left\{ S - \left(\frac{T_1^2}{n_1} + \frac{T_2^2}{n_2} + \dots + \frac{T_m^2}{n_m} \right) \right\} \\ &= \left(\frac{T_1^2}{n_1} + \frac{T_2^2}{n_2} + \dots + \frac{T_m^2}{n_m} \right) - \frac{T^2}{N} \\ &= \sum_{i=1}^m \frac{T_i^2}{n_i} - \frac{T^2}{N} \end{aligned}$$

表 16 の計算例では $N = 24, T = 186, S = 1852, \sum_{i=1}^m \frac{T_i^2}{n_i} = 1567.419$

だから

$$SS_{tot} = 1852 - \frac{186^2}{24} = 410.5$$

$$SS_{residual} = S - \sum_{i=1}^m \frac{T_i^2}{n_i} = 1852 - 1567.419 = 284.581$$

$$SS_{\text{実験区}} = 1567.419 - \frac{186^2}{24} = 125.919$$

$$\sigma^2_{residual} = \frac{SS_{residual}}{df_{residual}} = \frac{SS_{residual}}{df_{total} - df_{\text{実験区}}} = \frac{284.581}{23 - 3} = 14.229$$

$$\sigma^2_{\text{実験区}} = \frac{SS_{\text{実験区}}}{df_{\text{実験区}}} = \frac{125.919}{3} = 41.973$$

$$F = \frac{41.973}{14.229} = 2.950$$

となります。α=0.05 の F 分布表で 分子の自由度 3。分母の自由度 20 で臨界値を調べると、3.098 です。わずかですが、臨界値に達していません。

分散分析の結果は、以下のような分散分析表で示します。

表 3. 分散分析表の例

	平方和 (SS)	自由度 (df)	平方平均 (MS)	分散比 (F)
変動源				
水準間	125.919	2	62.95952 ^a	4.6460 ^{*c}
残差	284.581	21	13.55147 ^b	
合計	410.5	23		

*はα = 0.05で有意の意味^b

$$a: \frac{125.919}{2} = 62.95952 \quad b: \frac{284.581}{21} = 13.55147$$

$$c: \frac{a}{c} = \frac{62.95952}{13.55147} = 4.6460$$

繰り返しのない 2 要因分散分析の例

1 要因分散分析の結果を拡張して、要因が組み合わさっている場合、たとえば、飼っている魚の給仕量を 5 段階に変え、飼育温度を 4 段階に変えて、それぞれの組み合わせについて、3 匹の魚を飼い、その成長率を比べた場合を、それぞれの要因が成長率に依拠しているかどうかを判定する場合に使われる分析を説明します。この場合、それぞれのレベルは、給仕量や温度のような連続変数である必要はなく、たとえば種の違いや、水槽の形状の違いでもかまいません。第一段階としてくり返しが無い場合を考えます。つまり、魚が一匹しかいない。あるいは 1 水槽の全ての魚の成長の平均値を 1 データとするという例を考えてください。たとえば表 4 の例は A が餌の種類で A1、A2、A3 の 3 種類の餌、B は水槽の形状で、B1、B2、B3、B4 の 4 タイプあり、成長に及ぼす餌、水槽の形状の影響の有無を論じたいというような場合を考えてください。このデータを各要因ごとの分散を切り分けるのですが、この作業はすでに和の分散のところでも説明しました。ここでは、具体的な計算例を示します。

表 4. 繰り返しのない 2 要因分散分析の例

	A ₁	A ₂	A ₃
B ₁	11	11	8
B ₂	10	13	19
B ₃	9	18	18
B ₄	14	18	19

表 4.計算手順

	A ₁	A ₂	A ₃	n	T _{ij}	S _j	T _{ij} ² /n	S _j -T _{ij} ² /n
B ₁	11	11	8	3	30 ^a	306 ^b	300 ^c	6 ^d
B ₂	10	13	19	3	42	630	588	42
B ₃	9	18	18	3	45	729	675	54
B ₄	14	18	19	3	51	881	867	14
n	4 ^e	4	4	12 ^h				
T _{ji}	44	60	64		168 ⁱ			
S _i	498 ^f	938	1110			2546 ^j	2430 ^k	116 ^l
T _{ji} ² /n	484 ^g	900	1024			2408 ^m	2352 ⁿ	
S _i -T _{ji} ² /n	14 ^h	38	86			138 ^o		194 ^p

$$T_{ij} = \left(\sum_{i=1}^{n_a} A_i, B_j \right), \quad S_j = \sum_{i=1}^{n_a} (A_i, B_j)^2, \quad T_{ij}^2/n = \frac{T_{ij}^2}{n_a}, \quad S_j - T_{ij}^2/n = \sum_{i=1}^{n_a} (A_i, B_j)^2 - \frac{T_{ij}^2}{n_b}$$

$$T_{ji} = \left(\sum_{j=1}^{n_b} A_i, B_j \right), \quad S_i = \sum_{j=1}^{n_b} (A_i, B_j)^2, \quad T_{ji}^2/n = \frac{T_{ji}^2}{n_b}, \quad S_i - T_{ji}^2/n = \sum_{j=1}^{n_b} (A_i, B_j)^2 - \frac{T_{ji}^2}{n_b}$$

a: $11 + 11 + 8 = 30$, b: $11^2 + 11^2 + 8^2 = 306$, c: $\frac{a^2}{n_a}$, d: $b - c$

e: $11 + 10 + 9 + 14 = 44$, f: $11^2 + 10^2 + 9^2 + 14^2 = 498$, g: $\frac{e^2}{n_b}$, h: $f - g$

h: $n_a n_b = 12$, i (総和) = $30 + 42 + 45 + 51 = 44 + 60 + 64 = 168$,

j (全二乗和) = $306 + 630 + 729 + 881 = 498 + 938 + 1110 = 2546$

k = $300 + 588 + 675 + 867 = 2430$, l = $j - k = 2546 - 2430 = 116$

m = $484 + 900 + 1024 = 2408$, o = $j - m = 2546 - 2408 = 138$

n = $\frac{i^2}{h} = \frac{168^2}{12} = 2352$, p = $2546 - 2352 = 194$

この表を取りまとめると

全平方和 $SS_{total} = 194$

$$n_b SS_A = 194 - 138 = 56$$

$$n_a SS_B = 194 - 116 = 78$$

$$SS_{residual} = 194 - (56 + 78) = 60$$

$$SS_A = \frac{56}{4} = 14$$

$$SS_B = \frac{78}{6} = 13$$

$$\sigma^2_A = \frac{14}{2} = 7, \quad \sigma^2_B = \frac{26}{3} = 8.6667, \quad \sigma^2_{residual} = \frac{60}{11 - (2 + 3)} = \frac{60}{6} = 10$$

結果を分散分析表によってとりまとめて示します。

表 5. 分散分析表

	平方和	自由度	平方平均	分散比
	SS	df	MS	F
変動源				
A	14	2	7	0.7
B	26	3	8.6667	0.86667
残差	60	6	10	
合計	194	11		

F 比は、A の分散と残渣分散の比 $F_{A/residual} = \frac{7}{10} = 0.7$

B の分散と残渣分散の比 $F_{B/residual} = \frac{8.6667}{10} = 0.86667$

このように計算しているのですが、いきなり、結果だけを見せられても、考え方がわかりませんね。わかることは、全体の平方和が、A、B、残渣の部分平方和の和だということを利用してのことだけです。それだけでは、残渣平方和や残渣分散の内容がわかりませんね。

表 4 で例えば、 A_i, B_j のマスのデータは、全平均からの隔たり（偏差 ε_{ij} ）と全体の平均で表すと

$$D(A_i, B_j) = M_{total} + \varepsilon_{ij}$$

と表せます。

例えば、

$$D(A_1, B_1) = \frac{168}{12} + \varepsilon_{ij} = 14 + \varepsilon_{ij}$$

で

$$\varepsilon_{ij} = -3$$

です。 ε_{ij}^2 の総和が全平方和ですから、

$$SS_{total} = \sum_{i=1}^{n_a} \sum_{j=1}^{n_b} \varepsilon_{ij}^2$$

でこの値が194です。この計算ですが、例の平方和の計算法

$$SS_x = \sum_{i=1}^{n_x} (x_i - M_x)^2 = \sum_{i=1}^{n_x} x_i^2 - 2M_x \sum_{i=1}^{n_x} x_i + M_x^2 \sum_{i=1}^{n_x} 1 = \sum_{i=1}^{n_x} x_i^2 - \frac{1}{n_x} \left(\sum_{i=1}^{n_x} x_i \right)^2$$

を使いますから、

$$SS_{total} = \sum_{i=1}^{n_a} \sum_{j=1}^{n_b} (A_i, B_j)^2$$

こんな風に計算します。

ε_{ij} を構成する成分について考えます。和の分散の説明の時は、

$$D(A_i) = M_a + e_i$$

$$D(B_j) = M_b + e_j$$

e_i は $D(A_i)$ のAの平均値 M_a からの隔たり、 e_j は $D(B_j)$ の平均値 M_b として、

$$D(A_i + B_j) = M_a + e_i + M_b + e_j$$

平均値は加法的で、和の平均は平均の和だと考えられるから、

$$D(A_i + B_j) = M_a + e_i + M_b + e_j = M_{total} + (e_i + e_j)$$

つまり、

$$\varepsilon_{ij} = e_i + e_j$$

と考えることが出来たのですが、それは私たちが、 A_i と B_j の値を知っていて、それ以外の変動要因がないことを仮定したからです。実際のデータは、その組み合わせの時に加わるランダムな変動があるから、それによる偏差が加わって、

$$\varepsilon_{ij} = e_i + e_j + e_{ij}$$

のようになっているはずですが。これが $D(A_i, B_j)$ の変動源の構成です。ですから、全平方和の構成は次のようになります。

$$\begin{aligned}
SS_{total} &= \sum_{i=1}^{n_a} \sum_{j=1}^{n_b} (e_i + e_j + e_{ij})^2 = \sum_{i=1}^{n_a} \sum_{j=1}^{n_b} ((e_i^2 + e_j^2 + e_{ij}^2) + 2(e_i e_j + e_i e_{ij} + e_j e_{ij})) \\
&= \sum_{i=1}^{n_a} \sum_{j=1}^{n_b} e_i^2 + \sum_{i=1}^{n_a} \sum_{j=1}^{n_b} e_j^2 + \sum_{i=1}^{n_a} \sum_{j=1}^{n_b} e_{ij}^2 + 2 \left(\sum_{i=1}^{n_a} \sum_{j=1}^{n_b} e_i e_j + \sum_{i=1}^{n_a} \sum_{j=1}^{n_b} e_i e_{ij} + \sum_{i=1}^{n_a} \sum_{j=1}^{n_b} e_j e_{ij} \right) \\
&= n_b \sum_{i=1}^{n_a} e_i^2 + n_b \sum_{j=1}^{n_b} e_j^2 + \sum_{i=1}^{n_a} \sum_{j=1}^{n_b} e_{ij}^2
\end{aligned}$$

偏差の総和は0

$\leftarrow SS_A$ $\leftarrow SS_B$

ですから、

$$SS_{total} = n_b SS_A + n_a SS_B + SS_{residual}$$

となります。これも、全平方和は部分平方和の総和という公式の一例にすぎません。

これを使って、どのように効率的に、それぞれの値を出すかということですが、どうやっても構いません。私が効率が良いと思うのは、個々に示した計算手順です。

表計算を縦方向に見ると、 A_i の列のSSは A_i が一定ですから、 A_i の変動を含んでいません。

そこから、列の平均値を差し引いた、列平均からの偏差は $e_j + e_{ij}$ です。

その偏差を二乗した列の平方和は

$$\begin{aligned}
\sum_{j=1}^{n_b} (e_j + e_{ij})^2 &= \sum_{j=1}^{n_b} e_j^2 + \sum_{j=1}^{n_b} e_j e_{ij} + \sum_{j=1}^{n_b} e_{ij}^2 \\
&= SS_B + \sum_{j=1}^{n_b} e_j e_{ij} + \sum_{j=1}^{n_b} e_{ij}^2
\end{aligned}$$

これを、横方向に足し合わせれば、

$$\begin{aligned}
\sum_{i=1}^{n_a} \sum_{j=1}^{n_b} (e_j + e_{ij})^2 &= \sum_{i=1}^{n_a} \left(SS_B + \sum_{j=1}^{n_b} e_j e_{ij} + \sum_{j=1}^{n_b} e_{ij}^2 \right) \\
&= n_a SS_B + \sum_{i=1}^{n_a} \sum_{j=1}^{n_b} e_j e_{ij} + \sum_{i=1}^{n_a} \sum_{j=1}^{n_b} e_{ij}^2 \\
&= n_a SS_B + SS_{residual}
\end{aligned}$$

です。

$$SS_{total} = n_b SS_A + n_a SS_B + SS_{residual}$$

という式を使えば、

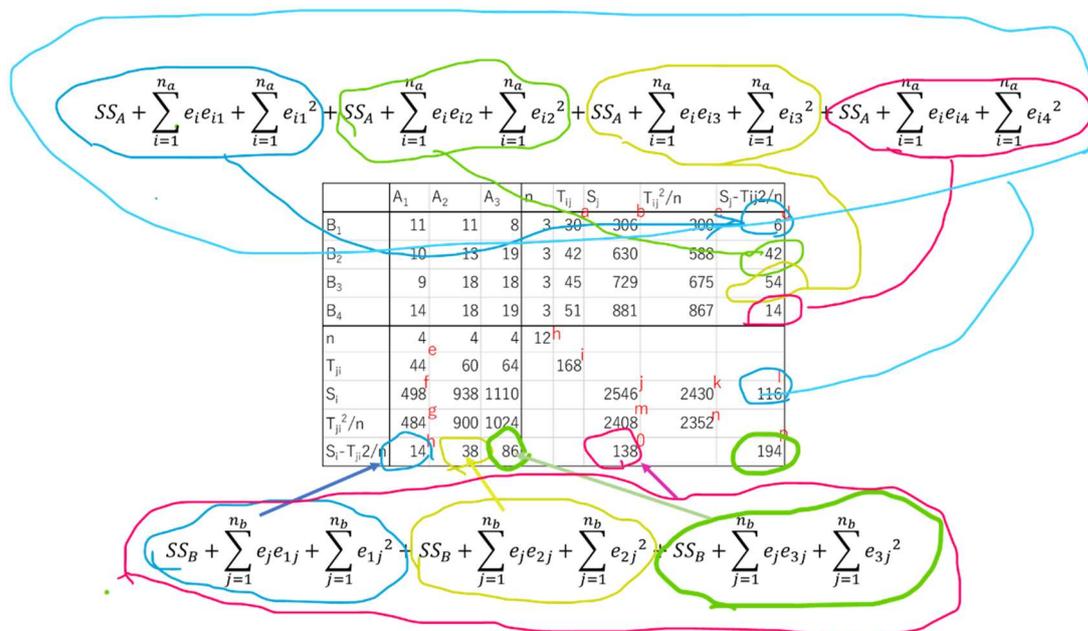
$$n_b SS_A = SS_{total} - (n_a SS_B + SS_{residual})$$

同じことを、 B_j を固定して、横う方向に平方和を求めて、縦方向に集計すれば、

$$n_a SS_B = SS_{total} - (n_b SS_A + SS_{residual})$$

となります。

表 4. 計算手順 (再掲)



次に自由度ですが、自由度についても、全自由度は部分自由度の和です。この場合は全自由度が 11 で、A の自由度が 2 で B の自由度が 3 ですから、残差自由度は

$$11 - (2 + 3) = 6$$

と計算するのが普通です。

この例の場合、A と B が直交してて、

$$n_a n_b - 1 - (n_a - 1) - (n_b - 1) = n_a n_b - n_a - n_b + 1 = (n_a - 1)(n_b - 1)$$

ですから、

$$2 \times 3 = 6$$

と計算しても間違いではありませんが、繰り返しのある複雑な形の多要因分散分析の場合、苦あり返し数が違っていたりして、 $(n_a - 1)(n_b - 1)$ のように単純化できない場合がありますから、全自由度は部分自由度の和だと記憶した方が良いでしょう。

それはそれとして、計算途中で、何か変だなと思いませんでしたか。以前にやった、和の分散の時の例で計算した時と、途中の計算の値が同じなのです。データセットはそうなるように、私が作ったのです。例にした表 4 は、2 要因分散分析に用いた総当たりの 2 つのサンプルグループを足し合わせて作ったデータ・シート (表 6) に表 7 のデータを足して作ったデータシートです。表 4 のままだと、A、B の標準からの偏差だけしかないので、

残

表 4. (再掲)

	A ₁	A ₂	A ₃
B ₁	11	11	8
B ₂	10	13	19
B ₃	9	18	18
B ₄	14	18	19

表 6 A₁ = 1, A₂ = 5, A₃ = 6, B₁ = 1, B₂ = 5, B₃ = 6, B₄ = 8 を総当たりのに足し合わせて作ったデータシート

	A1	A2	A3
B1	2	6	7
B2	6	10	11
B3	7	11	12
B4	9	13	14

表 7. 残差分散を作るために足し合わせたデータ

	A1	A2	A3
B1	9	5	1
B2	4	3	8
B3	2	7	6
B4	5	5	5

渣平方和が 0 になってしまいます。そこで残差平方和が出来るように、表 7 のデータを足したのです。このデータシートをよく見ると、3 行目までは、縦に足しても、横に足しても 15 になる、魔法陣になっていて、どの行でも列でも、平均が 5 になるようになっています。今まで使っていた表 6. の計算結果をそのまま使いたかったので、行の分散、列の分散が変わらないように、どの行でも、列でも平均的には同じ大きさのかさ上げになるように工夫したのです。表 7 で全 S S を計算すると、

$$9^2 + 5^2 + 1^2 + 4^2 + 3^2 + 8^2 + 2^2 + 7^2 + 6^2 + 5^2 + 5^2 + 5^2 - \frac{(9 + 5 + 1 + 4 + 3 + 8 + 2 + 7 + 6 + 5 + 5 + 5)^2}{12} = 360 - \frac{3600}{12} = 60$$

一方、すべての行と列の平均が等しいから、 $SS_A = 0$, $SS_B = 0$

$$SS_{total} = n_b SS_A + n_a SS_B + SS_{residual}$$

の式のこれらを入れれば、

$$SS_{total} = SS_{residual}$$

$SS_A = 0, SS_B = 0$ でも、それぞれ自由度があるから

$$df_{residual} = df_{total} - (df_A + df_B) = (12 - 1) - ((3 - 1) + (4 - 1)) = 6$$

$$\sigma^2_{residual} = \frac{60}{6} = 10$$

となります。全平方和が部分平方和の総和だということが、感覚的に納得できたでしょうか。

繰り返しのある2要因分散分析

2要因分散分析を発展させれば3要因分散分析等、さまざまな形式のデータを解析することが原理的に可能ですが、あまり複雑な形のもの进行分析しても結果の解釈に困るでしょう。最も単純な構造としては、よくある多要因分散分析の例として、1つの要因の中にいくつかの水準があり、もう一つの要因にもいくつかの水準があって、それぞれの要因の組み合わせのセル内にいくつかの繰り返しがあるという形式のデータについて考えます。表8に、要因Aに3水準、Bに4水準、それぞれの組み合わせに3つの繰り返しという例を示しました。

表8. 繰り返しのある2要因分散分析のデータ・シートの例

		A		
		A ₁	A ₂	A ₃
B	B ₁	10	8	6
		11	12	8
		12	13	10
	B ₂	9	12	18
		9	12	19
		12	15	20
	B ₃	8	15	17
		9	19	18
		10	20	19
	B ₄	13	14	18
		13	19	19
		16	21	20

この場合、各セル内の平均値を求めて、先に示した繰り返しのない2要因分散分析を行っても良いのですが、繰り返しの数が増えると、統計的な検出感度が上がるはずですが、できればこれを平均化せずに、繰り返しを生かした形で分散分析を行いたいところです。

まず、自由度のついて考えてみます。

Aの要因に m 個、Bの要因に n 個の水準（あるいは種類・条件・等々）、それぞれの組み合わせについて、 l 個の繰り返しという場合について考えます。

$$\text{全自由度 } df_{total} = mnl - 1 = 36 - 1 = 35$$

$$\text{Aの自由度 } df_A = m - 1 = 3 - 1 = 2$$

$$\text{Bの自由度 } df_B = n - 1 = 4 - 1 = 3$$

今までと同じに考えて

$$\text{(残渣)自由度 } df_{residual} = (m - 1)(n - 1) = 2 \times 3 = 6$$

と考えると、部分自由度の総和」という式に入れると

$$df_A + df_B + df_{residual} = 2 + 3 + 6 = 11$$

となって、全自由度と一致しません。つまり、繰り返しをすることによって、生まれた新たな変動源があるのです。この値と全自由度の差を考えます。

$$\begin{aligned} df_{total} - (df_A + df_B + df_{residual}) &= (mnl - 1) - ((m - 1) + (n - 1) + (m - 1)(n - 1)) \\ &= mnl - 1 - (m - 1 + n - 1 + mn - m + 1) = mnl - 1 - (mn - 1) \\ &= mnl - mn = mn(l - 1) \end{aligned}$$

つまり、この自由度に相当する、平方和と分散があるということです。 mn はAとBの水準（あるいは種類・条件・等々）の数の積で、実験ならば実験区の総数にあたるものですから、この式は、自由度が $l - 1$ の分散が、 mn 個あるという意味だろうという解釈が出来ます。 l が同一条件での繰り返しであることを思い出せば、こちらが、いわゆるランダムネスで、残渣に相当するものであり、 $(m - 1)(n - 1)$ に相当するものは、条件Aと条件Bの組み合わせに由来するもの（交互作用）だと考えられます。これは、繰り返しのない2要因分散分析では、これを残差分散としました。繰り返しのある、2要因分散分析では、交互作用を変動元とする変動と変動元を特定できないランダムな変動を取り分けて取り出すことが出来ます。これが繰り返しのある多要因分散分析の最大の利点なのです。繰り返しのない2要因分散分析では、交互作用を変動元とする変動とランダムな変動を取り分けることが出来ないから、自由度 $(m - 1)(n - 1)$ の変動を残差分散と呼ぶしかなかったのです。繰り返しのある2要因分散分析では、変動源は、要因A、要因B、要因Aと要因Bの交互作用と残渣があり、交互作用を $A \times B$ と表します。これを使って、自由度について整理すると、

$$\text{全自由度 } df_{total} = mnl - 1 = 36 - 1 = 35$$

$$\text{Aの自由度 } df_A = m - 1 = 3 - 1 = 2$$

$$\text{Bの自由度 } df_B = n - 1 = 4 - 1 = 3$$

$$\text{交互作用の自由度 } df_{A \times B} = (m - 1)(n - 1) = 2 \times 3 = 6$$

$$\text{残渣の自由度 } df_{residual} = mn(l - 1) = 2 \times 3 \times 2 = 24$$

となります。

複雑な多要因分散分析の場合、要因別に平方和をどのように分離すべきか、手順が頭に浮

かばないことがあります。そういう時に、まず、自由度について考えて、自由度の和が全自由度になるかどうかを確認します。

その上で、実際の計算をしますが、全平方和は部分平方和の総和だということを利用して、できるだけ効率よく計算します。この式をきちんと作るということも大切です。それの要因が何回繰り返されているのかを考えることがポイントです。この場合は、 $SS_{total} =$

$$nlSS_A + mlSS_B + lSS_{A \times B} + SS_{residual}$$

となっています。この式は、「要因 A については nl 回、要因 B については ml 回、交互作用については l 回の繰り返しがあり、その残りが、残渣平方和だ。」という意味です。実際の数値を入れると

$$SS_{total} = 4 \times 3 \times SS_A + 3 \times 3SS_B + 3 \times SS_{A \times B} + SS_{residual}$$

となります。

どのように計算するかは、その人がやりやすいように計算すれば良いのですが、込み入って、落ち着いて考えないと混乱します。エクセルにパターンを作って、それに入れて計算するのが間違いないでしょう。もっとも、今では、コンピュータのソフトを使えば、複雑な構造のデータシートでも、あっという間に計算してくれます。ちなみに、エクセルの分析ツールでも、繰り返しのある 2 要因分散分析まではあります。ですから、今では必要ないかもしれませんが、参考までに、私が作ったエクセルの計算シートを紹介します。具体的に何をしているのかわかるとと思います。

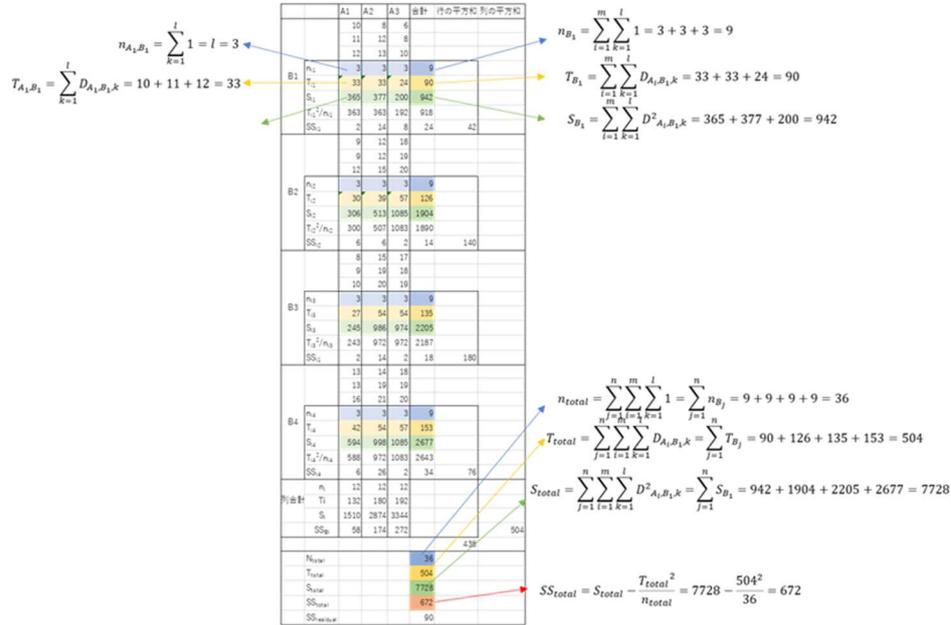
繰り返しのある 2 要因分散分析の計算手順

まず、簡便化した平方和の計算式を使って、全体の和 (T_{total} ; 504)、全体の 2 乗の和 (S_{total} 7728) 全データ数 (n_{total} 36) から、全平方和 (SS_{total}) を求めます。

$$SS_{total} = S_{total} - \frac{(T_{total})^2}{n_{total}} = 7728 - \frac{504^2}{36} = 672$$

具体的な手順がわかるように、私が作ったエクセルのシートで具体的な手順を示します。

表9. エクセルシートを使った平方和の計算 (全平方和)



次に各行の残差平方和 (24,14,18,34) から残差平方和を求めます。

$$SS_{residual} = \sum_{i=1}^m \sum_{j=1}^n \sum_{k=1}^l e_{residual:ijk}^2 = 24 + 14 + 18 + 34 = 90$$

表 10. エクセルシートを使った平方和の計算 (残渣平方和)

$SS_{residual_{A_1, B_1}} = S_{A_1, B_1} - \frac{T_{A_1, B_1}^2}{n_{A_1, B_1}} = 365 - \frac{33^2}{3} = 2$

$SS_{residual_{B_1}} = \sum_{i=1}^m S_{residual_{A_i, B_1}} = 2 + 14 + 8 = 24$

$SS_{residual} = \sum_{i=1}^m \sum_{j=1}^n SS_{residual_{A_i, B_j}} = \sum_{j=1}^n S_{residual_{B_j}} = 24 + 14 + 18 + 34 = 90$

次に、 B_1 の行について、 n_{i1}, T_{i1}, S_{i1} を列方向に加えた $\sum_{i=1}^m n_{i1}, \sum_{i=1}^m T_{i1}, \sum_{i=1}^m S_{i1}$ を計算します。

$$\sum_{i=1}^m n_{i1} = 3 + 3 + 3 = 9$$

$$\sum_{i=1}^m T_{i1} = 33 + 33 + 24 = 90$$

$$\sum_{i=1}^m S_{i1} = 365 + 377 + 200 = 942$$

それぞれ、 $n_{B_1}, T_{B_1}, S_{B_1}$ と名前を付けておきます平方和は、

$$S_{B_1} - \frac{T_{B_1}^2}{n_{B_1}} = 942 - \frac{90^2}{9} = 42$$

この SS はそれぞれの変動元からの偏差を考えると、次のような構成になっているはずす。

$$\sum_{i=1}^m \sum_{k=1}^l e_{A:i1k}^2 + \sum_{i=1}^m \sum_{k=1}^l e_{A \times B:i1k}^2 + \sum_{i=1}^m \sum_{k=1}^l e_{residual:i1k}^2$$

つまり、BをB₁に固定しているから、Bの変動を含んでいないのです。

B₂, B₃, B₄の行についても同じことをします。

$$S_{B_2} - \frac{T_{B_2}^2}{n_{B_2}} = 1904 - \frac{126^2}{9} = 140$$

$$\sum_{i=1}^m \sum_{k=1}^l e_{A:i2k}^2 + \sum_{i=1}^m \sum_{k=1}^l e_{A \times B:i2k}^2 + \sum_{i=1}^m \sum_{k=1}^l e_{residual:i2k}^2$$

$$S_{B_3} - \frac{T_{B_3}^2}{n_{B_3}} = 2205 - \frac{135^2}{9} = 180$$

$$\sum_{i=1}^m \sum_{k=1}^l e_{A:i3k}^2 + \sum_{i=1}^m \sum_{k=1}^l e_{A \times B:i3k}^2 + \sum_{i=1}^m \sum_{k=1}^l e_{residual:i3k}^2$$

$$\sum_{i=1}^m S_{B_4} - \frac{T_{B_4}^2}{n_{B_4}} = 2677 - \frac{153^2}{9} = 76$$

$$\sum_{i=1}^m \sum_{k=1}^l e_{A:i4k}^2 + \sum_{i=1}^m \sum_{k=1}^l e_{A \times B:i4k}^2 + \sum_{i=1}^m \sum_{k=1}^l e_{residual:i4k}^2$$

平方和を求めます。これも、簡便化した平方和の計算の公式を使います。

これらをすべて足し合わせると

$$\sum_{i=1}^m \sum_{k=1}^l e_{A:i1k}^2 + \sum_{i=1}^m \sum_{k=1}^l e_{A \times B:i1k}^2 + \sum_{i=1}^m \sum_{k=1}^l e_{residual:i1k}^2 = 42$$

$$\sum_{i=1}^m \sum_{k=1}^l e_{A:i2k}^2 + \sum_{i=1}^m \sum_{k=1}^l e_{A \times B:i2k}^2 + \sum_{i=1}^m \sum_{k=1}^l e_{residual:i2k}^2 = 140$$

$$\sum_{i=1}^m \sum_{k=1}^l e_{A:i3k}^2 + \sum_{i=1}^m \sum_{k=1}^l e_{A \times B:i3k}^2 + \sum_{i=1}^m \sum_{k=1}^l e_{residual:i3k}^2 = 180$$

$$\sum_{i=1}^m \sum_{k=1}^l e_{A:i4k}^2 + \sum_{i=1}^m \sum_{k=1}^l e_{A \times B:i4k}^2 + \sum_{i=1}^m \sum_{k=1}^l e_{residual:i4k}^2 = 76$$

+

$$\sum_{j=1}^n \sum_{i=1}^m \sum_{k=1}^l e_{A:ijk}^2 + \sum_{j=1}^n \sum_{i=1}^m \sum_{k=1}^l e_{A \times B:ijk}^2 + \sum_{j=1}^n \sum_{i=1}^m \sum_{k=1}^l e_{residual:ijk}^2 = 438$$

左辺、第1項の総和の順番を入れ替えると。

$$\sum_{j=1}^n \sum_{i=1}^m \sum_{k=1}^l e_{A:ijk}^2 = \sum_{j=1}^n \sum_{k=1}^l \sum_{i=1}^m e_{A:ijk}^2 = \sum_{j=1}^n \sum_{k=1}^l SS_A = n l SS_A$$

同様に第2項は、

$$\sum_{j=1}^n \sum_{i=1}^m \sum_{k=1}^l e_{A \times B:ijk}^2 = \sum_{k=1}^l \sum_{i=1}^m \sum_{j=1}^n e_{A \times B:ijk}^2 = \sum_{k=1}^l SS_{A \times B} = lSS_{A \times B}$$

第3項は、

$$\sum_{j=1}^n \sum_{i=1}^m \sum_{k=1}^l e_{residual:i4k}^2 = SS_{residual}$$

となるので、

$$nlSS_A + lSS_{A \times B} + SS_{residual} = 438$$

これは、Bの変動を含んでいない平方和の総和ですね。

これも、エクセルシートで計算手順を示します。

表 11. エクセルシートを使った平方和の計算 (Bの変動を除いた全平方和)

	A1	A2	A3	合計	行の平方和	列の平方和
	10	8	6			
	11	12	8			
	12	13	10			
B1	$n_{1.}$	3	3	3	9	
	$T_{1.}$	33	33	24	90	
	$S_{1.}$	365	377	200	942	
	$T_{1.}^2/n_{1.}$	363	363	192	918	
	$SS_{1.}$	2	14	8	24	42
	9	12	18			
	9	12	19			
	12	15	20			
B2	$n_{2.}$	3	3	3	9	
	$T_{2.}$	30	39	57	126	
	$S_{2.}$	306	513	1085	1904	
	$T_{2.}^2/n_{2.}$	300	507	1083	1890	
	$SS_{2.}$	6	6	2	14	140
	8	15	17			
	9	19	18			
	10	20	19			
B3	$n_{3.}$	3	3	3	9	
	$T_{3.}$	27	54	54	135	
	$S_{3.}$	245	986	974	2205	
	$T_{3.}^2/n_{3.}$	243	972	972	2187	
	$SS_{3.}$	2	14	2	18	180
	13	14	18			
	13	19	19			
	16	21	20			
B4	$n_{4.}$	3	3	3	9	
	$T_{4.}$	42	54	57	153	
	$S_{4.}$	594	998	1085	2677	
	$T_{4.}^2/n_{4.}$	588	972	1083	2643	
	$SS_{4.}$	6	26	2	34	76
列合計	$n_{.}$	12	12	12		
	$T_{.}$	132	180	192		
	$S_{.}$	1510	2874	3344		
	$SS_{B.}$	58	174	272		438
	N_{total}			36		
	T_{total}			504		
	S_{total}			7728		
	SS_{total}			672		
	$SS_{residual}$			90		

$n_{B_1} = \sum_{i=1}^m \sum_{k=1}^l 1 = 3 + 3 + 3 = 9$

$T_{B_1} = \sum_{i=1}^m \sum_{k=1}^l D_{A_i, B_1, k} = 33 + 33 + 24 = 90$

$S_{B_1} = \sum_{i=1}^m \sum_{k=1}^l D^2_{A_i, B_1, k} = 365 + 377 + 200 = 942$

$SS_{B_1} = S_{B_1} - \frac{T_{B_1}^2}{n_{B_1}} = 942 - \frac{90^2}{9} = 42$

$SS_{-B} = \sum_{j=1}^n SS_{B_j} = 42 + 140 + 180 + 76 = 438$

背景を小梅の赤で示した、行の平方和の列の438という値にはそんな意味があります。

「全平方和は部分平方和の総和」の式に、これを入れると、

$$\begin{aligned} SS_{total} &= nlSS_A + mlSS_B + lSS_{A \times B} + SS_{residual} \\ SS_{total} - (nlSS_A + lSS_{A \times B} + SS_{residual}) &= mlSS_B \\ mlSS_B &= 672 - 438 = 234 \\ SS_B &= \frac{234}{3 \times 3} = 26 \end{aligned}$$

同じ作業を列方向にします。A₁ の列について、n_{1j}, T_{1j}, S_{1j}を列方向に加えた $\sum_{j=1}^n n_{1j}$, $\sum_{j=1}^n T_{1j}$, $\sum_{j=1}^n S_{1j}$ を計算します。

$$\begin{aligned} \sum_{j=1}^n n_{1j} &= 3 + 3 + 3 + 3 = 12 \\ \sum_{j=1}^n T_{1j} &= 33 + 30 + 27 + 42 = 132 \\ \sum_{j=1}^n S_{1j} &= 365 + 306 + 245 + 594 = 1510 \end{aligned}$$

それぞれ、n_{A₁}, T_{A₁}, S_{A₁} と名前を付けておきます平方和は、

$$S_{A_1} - \frac{T_{A_1}^2}{n_{A_1}} = 1510 - \frac{132^2}{12} = 58$$

$$\sum_{j=1}^n \sum_{k=1}^l e_{B:1jk}^2 + \sum_{j=1}^n \sum_{k=1}^l e_{A \times B:1jk}^2 + \sum_{j=1}^n \sum_{k=1}^l e_{residual:i1k}^2$$

A を固定して A の変動を含まない変動和を作っています。

A₂, A₃の列についても同じことをします。

$$S_{A_2} - \frac{T_{A_2}^2}{n_{A_2}} = 2874 - \frac{180^2}{12} = 174$$

$$\sum_{j=1}^n \sum_{k=1}^l e_{B:2jk}^2 + \sum_{j=1}^n \sum_{k=1}^l e_{A \times B:2jk}^2 + \sum_{j=1}^n \sum_{k=1}^l e_{residual:2jk}^2$$

$$S_{A_3} - \frac{T_{A_3}^2}{n_{A_3}} = 3344 - \frac{192^2}{12} = 272$$

$$\sum_{j=1}^n \sum_{k=1}^l e_{B:3jk}^2 + \sum_{i=1}^m \sum_{k=1}^l e_{A \times B:3jk}^2 + \sum_{i=1}^m \sum_{k=1}^l e_{residual:i3k}^2$$

これらをすべて足し合わせると

$$\sum_{j=1}^n \sum_{k=1}^l e_{B:1jk}^2 + \sum_{j=1}^n \sum_{k=1}^l e_{A \times B:1jk}^2 + \sum_{j=1}^n \sum_{k=1}^l e_{residual:i1k}^2 = 58$$

$$\sum_{j=1}^n \sum_{k=1}^l e_{B:2jk}^2 + \sum_{j=1}^n \sum_{k=1}^l e_{A \times B:2jk}^2 + \sum_{j=1}^n \sum_{k=1}^l e_{residual:2jk}^2 = 174$$

$$\sum_{j=1}^n \sum_{k=1}^l e_{B:3jk}^2 + \sum_{j=1}^n \sum_{k=1}^l e_{A \times B:3jk}^2 + \sum_{j=1}^n \sum_{k=1}^l e_{residual:i3k}^2 = 272$$

+

$$\sum_{i=1}^m \sum_{j=1}^n \sum_{k=1}^l e_{B:ijk}^2 + \sum_{i=1}^m \sum_{j=1}^n \sum_{k=1}^l e_{A \times B:ijk}^2 + \sum_{i=1}^m \sum_{j=1}^n \sum_{k=1}^l e_{residual:i4k}^2 = 504$$

左辺、第1項の総和の順番を入れ替えると。

$$\sum_{i=1}^m \sum_{j=1}^n \sum_{k=1}^l e_{B:ijk}^2 = \sum_{i=1}^m \sum_{k=1}^l \sum_{j=1}^n e_{B:ijk}^2 = \sum_{i=1}^m \sum_{k=1}^l SS_B = mlSS_B$$

同様に第2項は、

$$\sum_{j=1}^n \sum_{i=1}^m \sum_{k=1}^l e_{A \times B:ijk}^2 = \sum_{k=1}^l \sum_{i=1}^m \sum_{j=1}^n e_{A \times B:ijk}^2 = \sum_{k=1}^l SS_{A \times B} = lSS_{A \times B}$$

第3項は、

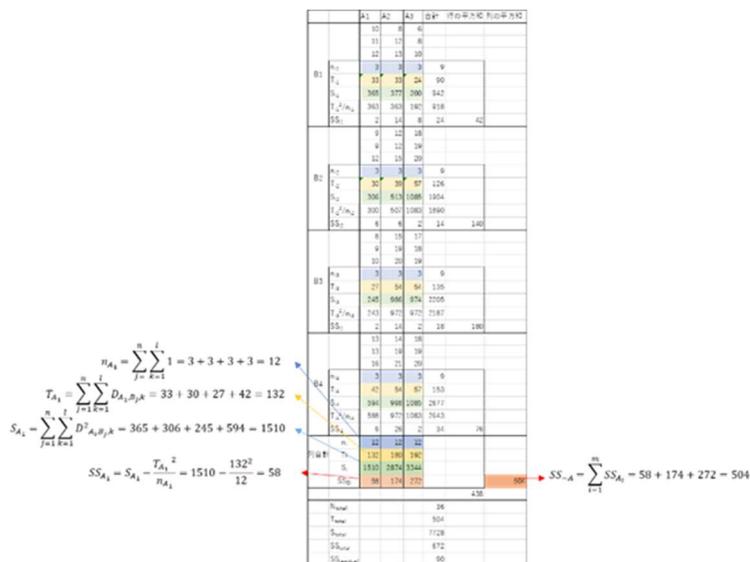
$$\sum_{j=1}^n \sum_{i=1}^m \sum_{k=1}^l e_{residual:i4k}^2 = SS_{residual}$$

となるので、

$$mlSS_B + lSS_{A \times B} + SS_{residual} = 504$$

Aの変動を含まない平方和の総和が出来ました。

表 12. エクセルシートを使った、平方和の計算 (A の変動を除いた全平方和)



「全平方和は部分平方和の総和」の式にこれを入れると、

$$SS_{total} = nLSS_A + mLSS_B + LSS_{A \times B} + SS_{residual}$$

$$SS_{total} - (mLSS_B + LSS_{A \times B} + SS_{residual}) = nLSS_A$$

$$nLSS_A = 672 - 504 = 168$$

$$SS_A = \frac{168}{4 \times 3} = 14$$

全平方和は部分平方和の総和の式にもどって。項誤作用の平方和を求めます

$$SS_{total} = nLSS_A + mLSS_B + LSS_{A \times B} + SS_{residual}$$

$$672 = 168 + 234 + LSS_{A \times B} + 90$$

$$LSS_{A \times B} = 180$$

$$SS_{A \times B} = \frac{180}{3} = 90$$

となります。

分散分析表で分析結果を示します (表 13)。

表 13 分散分析表

	平方和	自由度	分散	F 比 ¹	F 比 ²
	SS	d f	MS	F ¹	F ²
変動源					
A	14	2	7	1.866667	0.7
B	26	3	8.666667	2.311111	0.866667
交互作用AXB	60	6	10	2.666667 *	
残渣	90	24	3.75		
合計		35			

ここでは、分散比は、残渣分散に対する分散比 F^1 と交互作用に対する分散比 F^2 を示しました。普通、個々の要因による変動の有意性を論ずるだけであれば、 F^1 で有意性を論じて良いでしょう。交互作用が有意な場合、 F^2 で有意性を論ずる必要があるかもしれませんが、その前に交互作用があるとはどういうことなのかを考えます。分析したデータをグラフに描いてみます。

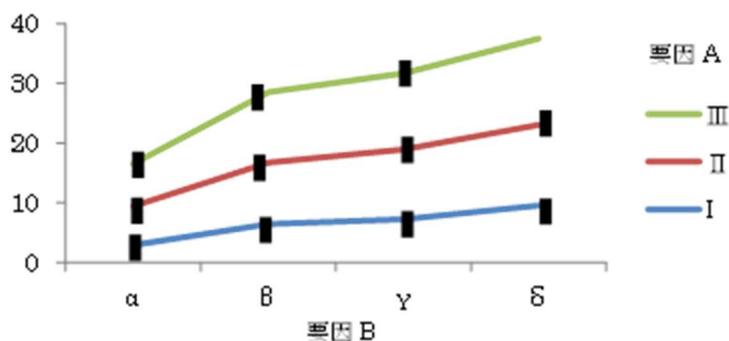


図2. 要因 A と B の関係 (原型)

黒いバーは、残渣によるデータの広がりを示しています。こういう図では、データのバラつきを標準偏差で示しますが、標準偏差の値は小さくて、この縮尺の図では見えなくなってしまうので、大きく拡大して書きました。具体的な大きさよりも、データが残渣による幅を持って存在することを示していると思ってください。要因 AI、要因 AII、要因 AIII で、要因 B による変動の大きさが違って見えるように見えませんか。要因 AIII では要因 AI に比べて、要因 B による変動が大きく出ています。もう少し極端にした方が分かりやすいので、交互作用として加えたデータを十倍にしてみます。

	A ₁	A ₂	A ₃			A ₁	A ₂	A ₃			A ₁	A ₂	A ₃
B ₁	2	6	7	+	B ₁	90	50	10	=	B ₁	92	56	17
B ₂	6	10	11		B ₂	40	30	80		B ₂	46	40	91
B ₃	7	11	12		B ₃	20	70	60		B ₃	27	81	72
B ₄	9	13	14		B ₄	50	50	50		B ₄	59	63	64

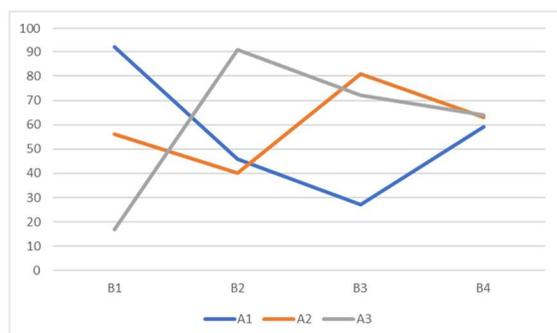


図3. 要因 A と要因 B の関係 (交互作用 10 倍)

交互作用のために、要因 B の影響の出方が、A1、A2、A3 で全く違ってきます。残渣分散が小さくて、残渣分散に対する要因 A、要因 B の分散比が十分に大きくても、どんな場合にも一定の法則で、要因 A、B の変動が表れるとは言えないでしょう。こういう場合、1 要因分散分析的に、A1、A2、A3 の個々の条件の中で、要因 B の影響を論ずることになるでしょう。

次に、交互作用が小さかった場合の極端な例とし交互作用を 10 分の 1 にしてみます。

	A ₁	A ₂	A ₃
B ₁	2	6	7
B ₂	6	10	11
B ₃	7	11	12
B ₄	9	13	14

+

	A ₁	A ₂	A ₃
B ₁	0.9	0.5	0.1
B ₂	0.4	0.3	0.8
B ₃	0.2	0.7	0.6
B ₄	0.5	0.5	0.5

=

	A ₁	A ₂	A ₃
B ₁	2.9	6.5	7.1
B ₂	6.4	10.3	11.8
B ₃	7.2	11.7	12.6
B ₄	9.5	13.5	14.5

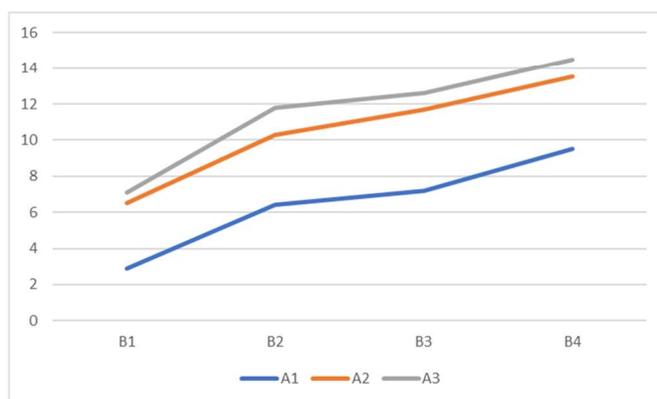


表 4. 要因 A と要因 B の関係 (交互作用 10 分の 1)

A1、A2、A3 の折れ線グラフの形がほとんど同じで、B の要因の影響が、A1、A2、A3 の条件にかかわらず、一定の強さで得依拠していることが分かります。

ここで行ったことは、一種の思考実験です。例えば、要因 A、要因 B の F^1 の分散比が有意で、 F^2 の分散比で、片方が有意でもう一方の有意性が否定された場合はどう考えるべきなのか等々、いくつかのケースが考えられと思います。幸い私はそういうケースに遭遇していません。そういう時はそういう時、その意味を具体的に考えれば良いのではないのでしょうか。あらかじめ一般的に考えておいても意味がないような気がします。ここで言えるのは、交互作用があった場合には、ここで例示したようなグラフを作って、交互作用の内容がどんなものかを確認して意味を考えるとということです。どのようにすればよいかは、統計学の問題ではありません。どうすべきかを知っているのは、専門知識・経験を持っている分析者自身です。グラフを作れば、その変動の意味がわかり、何を論ずべきかが決まると思っています。