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		

計算手順をは以下の通りです。

1．全平方和(SStotal)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	$\Sigma t_i^2/n_i$

表中の記号の説明

一般化して、実験区を$i = 1, \cdots m$とします。

n_i: 実験区i のデーター数

T_i: 実験区i のデーターの合計

M_i: 実験区i の平均値

S_i: 実験区 i のデーターの平方和$\sum x^2$

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

$$S_1 - \frac{T_1^2}{n_1} + S_2 - \frac{T_2^2}{n_2} + \cdots + S_m - \frac{T_m^2}{n_m}$$

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

これは全体の残差平方和の合計です。

二乗和の総和は

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

ですから

全体の残差平方和は

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

一方全体の SS はSS_{total}

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

です。

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

ですから、

$$SS_{実験区} = SS_{total} - SS_{residual} = S - \frac{T^2}{N} - \left\{S - \left(\frac{{T_1}^2}{n_1} + \frac{{T_2}^2}{n_2} + \cdots + \frac{{T_m}^2}{n_m}\right)\right\}$$

$$= \left(\frac{{T_1}^2}{n_1} + \frac{{T_2}^2}{n_2} + \cdots + \frac{{T_m}^2}{n_m}\right) - \frac{T^2}{N}$$

$$= \sum_{i=1}^{m} \frac{{T_i}^2}{n_i} - \frac{T^2}{N}$$

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

$$SS_{t0tal} = 1852 - \frac{186^2}{24} = 410.5$$

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

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

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

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

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

となります。$\alpha=0.05$ 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

変動源	平方和 (SS)	自由度 (df)	平方平均 (MS)	分散比 (F)
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$\alpha=0.05$で有意の意味 b

a : $\frac{125.919}{2}=62.95952$ 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ない場合を考えます。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-Tij2/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 f	60	64		168 i			
S_i	498 g	938	1110			2546 j	2430 k	116 l
$T_{ji}{}^2/n$	484 h	900	1024			2408 m	2352 n	
S_i-T_{ji}2/n	14	38	86			138 o		194 p

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

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

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

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

$\mathrm{h}=n_a n_b=12,\quad \mathrm{i}\left(\text{総和}\right)=30+42+45+51=44+60+64=168,$

$\mathrm{j}\left(\text{全二乗和}\right)=306+630+729+881=498+938+1110=2546$

$\mathrm{k}=300+588+675+867=2430,\ \mathrm{l}=\mathrm{j}-\mathrm{k}=2546-2430=116$

$\mathrm{m}=484+900+1024=2408,\ \mathrm{o}=\mathrm{j}-\mathrm{m}=2546-2408=138$

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

この表を取りまとめると

全平方和　SS_{total} =194

$$n_b SS_A = 194 - 138 = 56$$
$$n_a SS_B = 194 - 116 = 78$$
$$SS_{residual} = 194 - (56 + 78) = 60$$
$$SS_A = \frac{56}{4} = 14$$
$$SS_A = \frac{78}{6} = 26$$
$$\sigma^2{}_A = \frac{14}{2} = 7, \quad \sigma^2{}_B = \frac{26}{3} = 8.667, \quad \sigma^2{}_{residual} = \frac{60}{11 - (2+3)} = \frac{60}{6} = 10$$

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 SS	自由度 df	平方平均 MS	分散比 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いるということだけです。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、

$$D(A_1, B_1) = \frac{168}{12} + \varepsilon_{ij} = 11$$

で

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

です。$\varepsilon_{ij}{}^2$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

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

を使いますから、

n=2352

j ：2546

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

h=12

i=168

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、

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

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_j2 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。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。

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

$$= \sum_{i=1}^{n_a}\sum_{j=1}^{n_b}{e_i}^2 + \sum_{i=1}^{n_a}\sum_{j=1}^{n_b}{e_j}^2 + \sum_{i=1}^{n_a}\sum_{j=1}^{n_b}{e_{ij}}^2 + 2\left(\sum_{i=1}^{n_a}\sum_{j=1}^{n_b}e_i e_j + \sum_{i=1}^{n_a}\sum_{j=1}^{n_b}e_i e_{ij} + \sum_{i=1}^{n_a}\sum_{j=1}^{n_b}e_j e_{ij}\right)$$

偏差の総和は0

$$= n_b\sum_{i=1}^{n_a}{e_i}^2 + n_b\sum_{j=1}^{n_b}{e_i}^2 + \sum_{i=1}^{n_a}\sum_{j=1}^{n_b}{e_{ij}}^2$$

SS_A　SS_B　SS_r

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いうのは、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

$$\sum_{j=1}^{n_b}\left(e_j + e_{ij}\right)^2 = \sum_{j=1}^{n_b}{e_j}^2 + \sum_{j=1}^{n_b}e_j e_{ij} + \sum_{j=1}^{n_b}{e_{ij}}^2$$

$$= SS_B + \sum_{j=1}^{n_b}e_j e_{ij} + \sum_{j=1}^{n_b}{e_{ij}}^2$$

これを、横方向に足し合わせれば、

$$\sum_{i=1}^{n_a}\sum_{j=1}^{n_b}\left(e_j + e_{ij}\right)^2 = \sum_{i=1}^{n_a}\left(SS_B + \sum_{j=1}^{n_b}e_j e_{ij} + \sum_{j=1}^{n_b}{e_{ij}}^2\right)$$

$$= n_a SS_B + \sum_{i=1}^{n_a}\sum_{j=1}^{n_b}e_j e_{ij} + \sum_{i=1}^{n_a}\sum_{j=1}^{n_b}e_{ij}$$

$$= n_a SS_B + SS_{residual}$$

です。

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

という式を使えば、

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

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）

$$SS_A + \sum_{i=1}^{n_a} e_i e_{i1} + \sum_{i=1}^{n_a} {e_{i1}}^2 + SS_A + \sum_{i=1}^{n_a} e_i e_{i2} + \sum_{i=1}^{n_a} {e_{i2}}^2 + SS_A + \sum_{i=1}^{n_a} e_i e_{i3} + \sum_{i=1}^{n_a} {e_{i3}}^2 + SS_A + \sum_{i=1}^{n_a} e_i e_{i4} + \sum_{i=1}^{n_a} {e_{i4}}^2$$

	A_1	A_2	A_3	n	T_{ij}	S_j	T_{ij}^2/n	S_j-Tij2/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	4	4	12 [h]				
T_{ji}	44 [e]	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]

$$SS_B + \sum_{j=1}^{n_b} e_j e_{1j} + \sum_{j=1}^{n_b} {e_{1j}}^2 + SS_B + \sum_{j=1}^{n_b} e_j e_{2j} + \sum_{j=1}^{n_b} {e_{2j}}^2 + SS_B + \sum_{j=1}^{n_b} e_j e_{3j} + \sum_{j=1}^{n_b} {e_{3j}}^2$$

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的に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　$A_1 = 1$, $A_2 = 5$, $A_3 = 6$, $B_1 = 1, B_2 = 5$, $B_3 = 6$, $B_4 = 8$ を総当たり的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ＳＳ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を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。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。

全自由度 $df_{total} = mnl - 1$=36-1=35

A の自由度$df_A = m$-1=3-1=2

B の自由度$df_B = n - 1$=4-1=3

今までと同に考えて

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

と考えて、部分自由度の総和」という式に入れると

$$df_A + f_B + f_A df_B = 2 + 3 + 6 = 11$$

となって、全自由度と一致しません。つまり、繰り返しを作ることによって、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。

$$df_{total} - (df_A + df_B + df_A df_B) = (mnl - 1) - \big((m-1) + (n-1) + (m-1)(n-1)\big)$$
$$= mnl - 1 - (m - 1 + n - 1 + mn - n - m + 1) = mnl - 1 - (mn - 1)$$
$$= \ mnl - mn = mn(l-1)$$

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、

全自由度 $df_{total} = mnl - 1$=36-1=35

A の自由度$df_A = m$-1=3-1=2

B の自由度$df_B = n - 1$=4-1=3

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

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のも大切です。それぞ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、今では必要ないのかもしれませんが、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。
具体的になのをしているのかわかると思います。

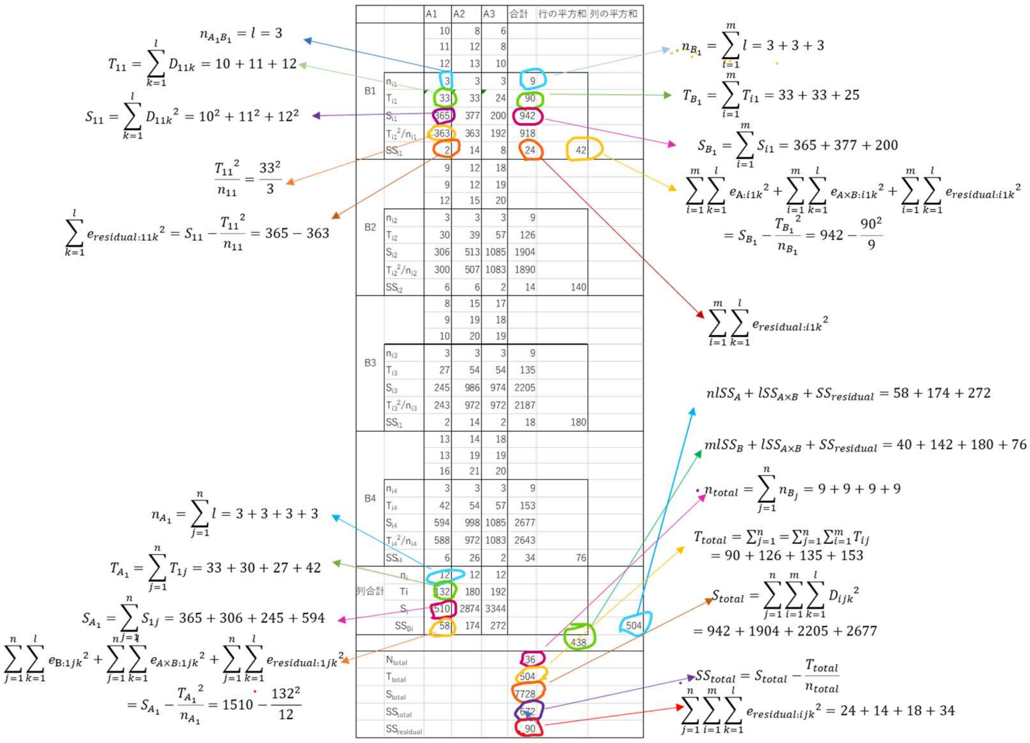

図１．繰り返しのある２要因分散分析の計算手順

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）全データー数(n_{total} 36)から、全平方和(SS_{total}) を求めます。

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

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。

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

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_{B1} , 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です。

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

つまり、B をB_1に固定しているから、B の変動を含んでいないのです。

B_2, B_3, B_4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。ここも、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

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

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

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

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

+

$$\sum_{j=1}^{n}\sum_{i=1}^{m}\sum_{k=1}^{l}{e_{A:ijk}}^2+\sum_{j=1}^{n}\sum_{i=1}^{m}\sum_{k=1}^{l}{e_{A\times B:ijk}}^2+\sum_{j=1}^{n}\sum_{i=1}^{m}\sum_{k=1}^{l}{e_{residual:i4k}}^2=438$$

左辺、第 1 項の総和の順番を入れ替えると。

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

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。

緑で囲った、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」の式のこれを入れると、

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

$$SS_{total} - (\,nlSS_A + lSS_{A\times B} + SS_{residual}) = mlSS_B$$

$$mlSS_B = 672 - 438 = 234$$

$$SS_B = \frac{234}{3\times 3} = 26$$

同じ作業を列方向にします。A_1 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。

$$\sum_{j=1}^{n} n_{1j} = 3+3+3+3 = 12$$

$$\sum_{j=1}^{n} T_{1j} = 33+30+27+42 = 132$$

$$\sum_{j=1}^{n} S_{1j} = 365+306+245+594 = 1510$$

それぞれ、n_{A_1}, T_{A_1}, S_{A_1} 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。

A_2, A_3,の列についても同じことをします。

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

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

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

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

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。

「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。

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

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

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

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

となります。

分散分析表で分析結果を示します（表９）。

表９分散分析表

	平方和	自由度	分散	Ｆ比^1		F比^2
	SS	ｄｆ	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。

ここで、分析したデータをグラフに描いてみます。

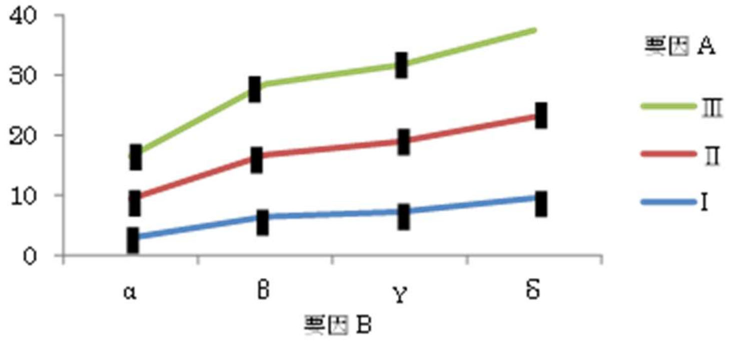

図２．要因 A と B の関係（原型）

黒いバーは、残差によるデータの広がりを示しています。こういう図では、データー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、データが残差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い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
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	90	50	10
B_2	40	30	80
B_3	20	70	60
B_4	50	50	50

＝

	A_1	A_2	A_3
B_1	92	56	17
B_2	46	40	91
B_3	27	81	72
B_4	59	63	64

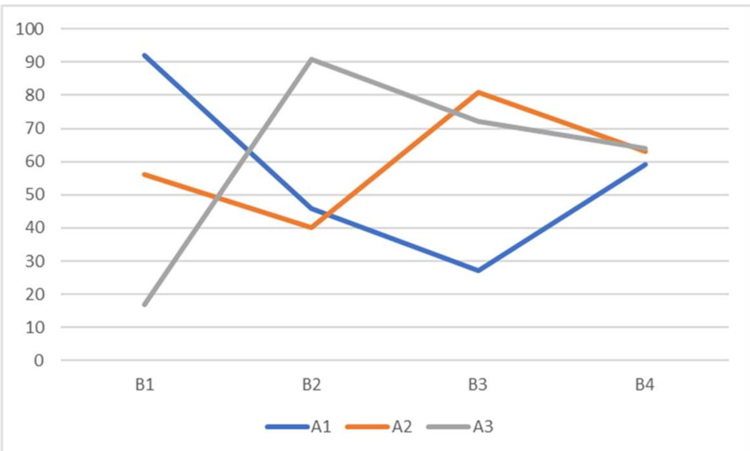

図 3. 要因 A と要因 B の関係（交互作用１０倍）

交互作用のために、要因 B の影響の出方が、A1、A2、A3 で全く違っています。仮に、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

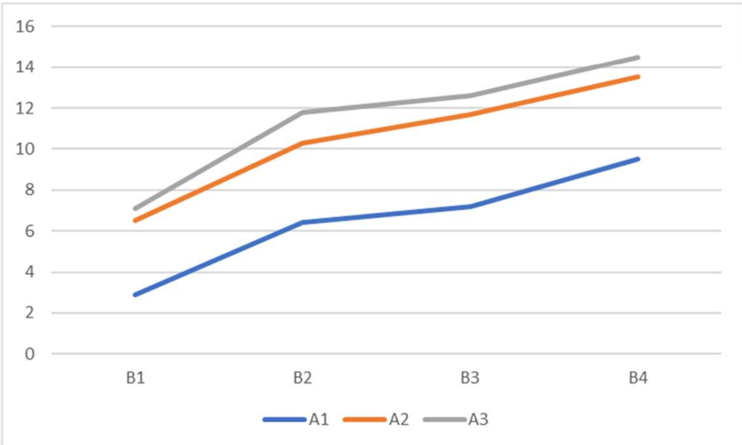

表 4. 要因 A と要因 B の関係（交互作用１０分の１）

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います。