IV-2-3. データの構造

実際のデータは複数の要因によって構成されています。まず、最も簡単な例として、2つの要因によって構成されているデータを考えます。しかし、すでに IV-2-2 で論じたように、完全に2つの要因のみによってデータが構成されているとすると、私たちは、その要因による変動の有意性を論ずることが出来ません。要因が有意であるとは、その要因による変動が、予測されないランダムな変動に比べて十分な大きさを持っているという意味だからです。そこで、この理論的な考察では、各要因の分散に影響を与えないランダムな変動を加えてみます。具体的には、それぞれの行列の平均値が等しくなるように、それぞれのセルの値に加えるということです。ここでは表 16 が加えるランダム変動のリストです。

表 16. 加えるランダムな変動

	A_1	A_2	A_3	Sum	Mean
B_1	0.9	0.5	0.1	1.5	0.5
B_2	0.4	0.3	0.8	1.5	0.5
B_3	0.2	0.7	0.6	1.5	0.5
B_4	0.5	0.5	0.5	1.5	0.5
Sum	2.0	2.0	2.0	6.0	0.5
Mean	0.5	0.5	0.5	0.5	

これならば、どの行についても列についても、同じ値が加えられますから、要因 A、要因 B には影響がないことになります。もちろん、新たに変動が加わったので、その分だけ、全体の変動は大きくなります。加わった変動を表 17、表 18 のように計算して、増加した変動の平方和は、SS_r =0.60です。

Table 17. 平均値からの距離の平方

	A_1	A_2	A_3
B_1	$(0.9-0.5)^2$	$(0.5-0.5)^2$	$(0.1-0.5)^2$
B_2	$(0.4-0.5)^2$	$(0.3-0.5)^2$	$(0.8-0.5)^2$
B_3	$(0.2-0.5)^2$	$(0.7-0.5)^2$	$(0.6-0.5)^2$
B_4	$(0.5-0.5)^2$	$(0.5-0.5)^2$	$(0.5-0.5)^2$

計算結果

	A1	A2	A3	Sum
B1	0.16	0	0.16	0.32
B2	0.01	0.04	0.09	0.14
B3	0.09	0.04	0.01	0.14
B4	0	0	0	0.
Sum	0.26	0.08	0.26	0.60

表 18. 合成されたデータ

	A1	A2	A3	Sum	Mean
B1	2.9	6.5	7.1	16.5	5.5
B2	6.4	10.3	11.8	28.5	9.5
B3	7.2	11.7	12.6	31.5	10.5
B4	9.5	13.5	14.5	37.5	12.5
Sum	26	42	46	114	9.5
Mean	6.5	10.5	11.5	9.5	

表 19. 平方和

	A1	A2	A3	Sum
B1	43.56	9	5.76	58.32
B2	9.61	0.64	5.29	15.54
B3	5.29	4.84	9.61	19.74
B4	0	16	25	41.00
Sum	58.46	30.48	45.66	134.60

全 SS は 134.60 です。この計算結果から、あることに気づきませんか。

$$134.60 = 56 + 78 + 0.60 = 4 \times 14 + 3 \times 78 + 0.60 = n_B SS_{\hat{A}} + n_A SS_{\hat{B}} + SS_r$$

という風に見えてきませんか。そうなるように作ったのだか、そんなこと当たり前だろうという感じかもしれません。あるいは、なんでそうなるんだろうという感じかもしれません。ここでは、実際の計算に沿ってなるほどという納得できる説明をします。

例えば A1B1 の全平均値からの距離の 2 乗の計算ですが。

$$(2.9 - 9.5)^2 = (6.6)^2 = 43.56$$

と計算しているのですが、もともとは、

$$\{(1 - 4) + (1 - 5) + (0.9 - 0.5)\}^2 = 43.56$$

と言う計算だったのですね。赤字のところが、要因 A の関するもので、青字のところが要因 B に関する部分、緑字のところが最後に加えた要因 r によるものです。

この式を展開すると、

$$(1 - 4)^2 + (1 - 5)^2 + (0.9 - 0.5)^2 + 2\{(1 - 4)(1 - 5) + (1 - 4)(0.9 - 0.5) + (1 - 5)(0.9 - 0.5)\}$$

となりますが、B 1 の行を横に見ていって、A2B1 セルでは、

$$(5 - 4)^2 + (1 - 5)^2 + (0.5 - 0.5)^2 + 2\{(5 - 4)(1 - 5) + (5 - 4)(0.5 - 0.5) + (1 - 5)(0.5 - 0.5)\}$$

A3B1 のセルでは

$$(6 - 4)^2 + (1 - 5)^2 + (0.1 - 0.5)^2 + 2\{(6 - 4)(1 - 5) + (6 - 4)(0.1 - 0.5) + (1 - 5)(0.1 - 0.5)\}$$

これらを、行 B1 についてたしあわせれば。

$$(1 - 4)^2 + (5 - 4)^2 + (6 - 4)^2 + 3(1 - 5)^2 + (0.9 - 0.5)^2 + (0.5 - 0.5)^2 + (0.1 - 0.5)^2 + 2\{(1 - 5)((1 - 4) + (5 - 4) + (6 - 4)) + (1 - 4)(0.9 - 0.5) + (5 - 4)(0.5 - 0.5) + (6 - 4)(0.1 - 0.5) + (1 - 5)((0.9 - 0.5) + (0.5 - 0.5) + (0.1 - 0.5))\}$$

$$(1-4)+(5-4)+(6-4)=0$$

$$(0.9-0.5)+(0.5-0.5)+(0.1-0.5)=0$$

だから、この計算は、

$(1-4)^2+(5-4)^2+(6-4)^2+3(1-5)^2+(0.9-0.5)^2+(0.5-0.5)^2+(0.1-0.5)^2+2\{(1-4)(0.9-0.5)+(5-4)(0.5-0.5)+(6-4)(0.1-0.5)\}$

$=SS_{\hat{A}}+3(1-5)^2+(0.9-0.5)^2+(0.5-0.5)^2+(0.1-0.5)^2+2\{(1-4)(0.9-0.5)+(5-4)(0.5-0.5)+(6-4)(0.1-0.5)\}$

この計算結果を B1 から B4 の行について縦に加えると。

$$SS_{total}=4SS_{\hat{A}}+3\{(1-5)^2+(5-5)^2+(6-5)^2+(8-5)^2\}$$
$$+\{(0.9-0.5)^2+(0.5-0.5)^2+(0.1-0.5)^2+(0.4-0.5)^2+(0.3-0.5)^2$$
$$+(0.8-0.5)^2+(0.2-0.5)^2+(0.7-0.5)^2+(0.6-0.5)^2+(0.5-0.5)^2+(0.5-0.5)^2$$
$$+(0.5-0.5)^2\}$$
$$+2\{(1-4)\big((0.9-0.5)+(0.4-0.5)+(0.2-0.5)+(0.5-0.5)\big)+(5-4)\big((0.5-0.5)+(0.3-0.5)+(0.7-0.5)+(0.5-0.5)\big)+(6-4)\big((0.1-0.5)+(0.8-0.5)+(0.6-0.5)+(0.5-0.5)\big)\}$$

$$(0.9-0.5)+(0.4-0.5)+(0.2-0.5)+(0.5-0.5)=0$$

$$(0.5-0.5)+(0.3-0.5)+(0.7-0.5)+(0.5-0.5)=0$$

$$(0.1-0.5)+(0.8-0.5)+(0.6-0.5)+(0.5-0.5)=0$$

だから、

$$SS_{total}=4SS_A+3SS_B+SS_r=n_BSS_A+n_ASS_B+SS_r$$

この式は、全平方和は、データを構成する個々の要因による部分平方和に分解できるということを示しています。そして、データーの構造を理解していれば、複雑な構造を持ったデーターでも、全平方和を個々の要因による平方和に分解できます。それに加えて、個々の要因の自由度がわかれば、個々の要因による分散をもとめることができて、それぞれの分散の大きさを比べることができます。ここでは、具体的な数値を使って計算していますから、一般的な証明にはなっていませんが、考え方はわかると思います。

さて、問題は自由度です。この場合、要因Ａの自由度２、要因Ｂの自由度３は。全自由度１１はわかっています。とすると、残された、ｒの自由度は$11-2-3=6$となります。これで正しいのですが、どうしてそうなるのかということは、若干説明が必要かもしれません。データに加えたｒの要因によるデータの表を見てください。横３、縦４の表になっています。横の数字についてみると、各行の平均が同じです、各列の平均も同じです。列が３つなのだから、列についての自由度は２ですね。同じように行についての自由度は３です。これが、お互いに直交しているのだから、全自由度は

$$(3-1)(4-1)-6$$

です。

もう少し、一般的な書き方をすると

$$全自由度 = n_A n_b - 1$$
$$要因 A の自由度 = n_A - 1$$
$$要因 B の自由度 = n_B - 1$$
$$要因 r の自由度 = \mathrm{df}_r$$
$$全自由度 = 要因 A の自由度＋要因 B の自由度＋要因 r の自由度$$
$$n_A n_b - 1 = (n_A - 1) + (n_B - 1) + \mathrm{df}_r$$
$$\mathrm{df}_r = n_A n_b - n_A - n_B + 1 = (n_A - 1)(n_B - 1)$$

従ってランダムな変動の分散は

$$\sigma_r = \frac{SS_r}{df_r} = \frac{0.6}{6} = 0.1$$

要因の分散はすでに知っている通り

$$\sigma_A = \frac{SS_A}{df_A} = \frac{14}{2} = 7$$

$$\sigma_B = \frac{SS_B}{df_B} = \frac{26}{3} = 8.6667$$

ですから

$$F_{A-r} = \frac{\sigma_A}{\sigma_r} = \frac{7}{0.1} = 70$$

$$F_{B-r} = \frac{\sigma_B}{\sigma_r} = \frac{8.6667}{0.1} = 86.667$$

F 表を見るまでもなく、この結果は、r(ここではランダムな変動を想定しています。r の要因による変動に比べて、要因 A および B による変動は十分な広がりを持っている（有意だ）と結論できるでしょう。ここで、想定した要因 A、B によって説明できない分散を残差（residual）分散といいます。

実際のデータでは、合成した元のデータの分散など知りませんから、どのようにそれぞれの平方和を計算するかが問題になります。直接、計算できそうなのは、すでに計算した全平方和と、行間の平方和、列間の平方和, 行ごとの平方和、列ごとの平方和だから、それらがどの様な要素から構成されるのか、検討してみましょう。

表 18（データセット、再掲）

	A1	A2	A3	SS	Mean
B1	2.9	6.5	7.1	16.5	5.5
B2	6.4	10.3	11.8	28.5	9.5
B3	7.2	11.7	12.6	31.5	10.5
B4	9.5	13.5	14.5	37.5	12.5
Sum	26	42	46	114	9.5
Mean	6.5	10.5	11.5	9.5	

行ごとの平方和の計算。行の平均値からの距離の２乗の表を作ります。

行ごとの平方和の計算

	A1	A2	A3	SS
B1	6.76	1.00	2.56	10.32
B2	9.61	0.64	5.29	15.54
B3	10.89	1.44	4.41	16.74
B4	9.00	1.00	4.00	14.00
Sum				56.60

行の平均値の分散

$$(5.5-9.5)^2+(9.5-9.5)^2+(10.5-9.5)^2+(12.5-9.5)^2\text{=}26$$

列ごとの平方和の計算。列の平均値からの距離の２乗の表を作ります。

列ごとの平方和の計算

	A1	A2	A3	SS
B1	12.96	16.00	19.36	
B2	0.01	0.04	0.09	
B3	0.49	1.44	1.21	
B4	9.00	9.00	9	
Sum	22.46	26.48	2966	78.60

列の平均値の分散

$$(6.5-9.5)^2+(10.5-9.5)^2-(11.5-9.5)^2=14$$

この計算結果を見ると、

行の平均値の平方和は列の要因の平方和SS_{B_i}とランダムな変動の和、列の平均値の分散は行の要因の平方和SS_{A_j}とランダムな変動の和。行ごとの平方和の和は、$n_B SS_A+SS_r$、列ごとの平方和の和は、$n_A SS_B+SS_r$だという予想がつきます。一応、計算の過程を追ってみましょう。

１行目の平均値は

5.5 ですが、この値は

$$\frac{1}{3}\{2.9+6.5+7.1\}=\frac{1}{3}\{(1+1+0.9)+(5+1+0.5)+(6+1+0.1)\}$$

$$=\frac{1}{3}\{(1+5+6)+3\times 1+(0.9+0.5+0.1)\}$$

$$=M_A+1+M_r$$

この値の平均値からの距離は

$$5.5-9.5=M_A+1+M_r-(M_A+M_B+M_r)=1-M_B$$

この値の２乗は、

$$(1-M_B)^2$$

この値の各行についての総和は

$$(1-M_B)^2+(5-M_B)^2+(6-M_B)^2+(8-M_B)^2=SS_{\hat{B}}$$

となります。

列についても同様で、列の平均値の平方和は、$SS_{\hat{A}}$です。

行ごとの平方和の和の計算では、

A1B1 のセルについてみると

$$(2.9-5.5)^2=\{(1-4)+(1-1)+(0.9-0.5)\}^2$$
$$=\{(1-4)^2+(0.9-0.5)^2+2(1-4)(0.9-0.5)\}$$

行を横に見ていき、A2B1 のセルでは

$$(6.5-5.5)^2=\{(5-4)^2+(0.5-0.5)^2+2(5-4)(0.5-0.5)\}$$

A3B1 では

$$(7.1-5.5)^2=\{(6-4)^2+(0.1-0.5)^2+2(6-4)(0.1-0.5)\}$$

これらをたしあわせた B1 行の総和は

$$(1-4)^2+(5-4)^2+(6-4)^2+(0.9-0.5)^2+(0.5-0.5)^2+(0.1-0.5)^2$$
$$+2\{(1-4)(0.9-0.5)+(5-4)(0.5-0.5)+(6-4)(0.1-0.5)\}$$

B2 行の総和は

$$(1-4)^2+(5-4)^2+(6-4)^2+(0.4-0.5)^2+(0.3-0.5)^2+(0.8-0.5)^2$$
$$+2\{(1-4)(0.4-0.5)+(5-4)(0.3-0.5)+(6-4)(0.8-0.5)\}$$

B3 行の総和は

$$(1-4)^2+(5-4)^2+(6-4)^2+(0.2-0.5)^2+(0.7-0.5)^2+(0.6-0.5)^2$$
$$+2\{(1-4)(0.2-0.5)+(5-4)(0.7-0.5)+(6-4)(0.6-0.5)\}$$

B4 行の総和は

$$(1-4)^2+(5-4)^2+(6-4)^2+(0.5-0.5)^2+(0.5-0.5)^2+(0.5-0.5)^2$$
$$+2\{(1-4)(0.5-0.5)+(5-4)(0.5-0.5)+(6-4)(0.5-0.5)\}$$

これらをたしあわせた総和は

$$\begin{aligned}
&4\{(1-4)^2+(5-4)^2+(6-4)^2\}\\
&+\{(0.1-0.5)^2+(0.2-0.5)^2+(0.3-0.5)^2+(0.4-0.5)^2\\
&+(0.5-0.5)^2+(0.6-0.5)^2+(0.7-0.5)^2+(0.8-0.5)^2\\
&+(0.9-0.5)^2\}\\
&+(1-4)\{(0.9-0.5)+(0.4-0.5)+(0.2-0.5)+(0.5-0.5)\}\\
&+(5-4)\{(0.5-0.5)+(0.3-0.5)+(0.7-0.5)+(0.5-0.5)\}\\
&+(6-4)\{(0.1-0.5)+(0.8-0.5)+(0.6-0.5)+(0.5-0.5)\}
\end{aligned}$$
$$=4SS_A+SS_r=n_BSS_A+SS_r$$

$$\begin{aligned}
\because\ &(0.9-0.5)+(0.4-0.5)+(0.2-0.5)+(0.5-0.5)\\
=&(0.5-0.5)+(0.3-0.5)+(0.7-0.5)+(0.5-0.5)\\
=&(0.1-0.5)+(0.8-0.5)+(0.6-0.5)+(0.5-0.5)\\
=&0
\end{aligned}$$

同様に列ごとの平方和の和は

$$n_A SS_B + SS_r$$

となります。

これらを使えば、SS_{total}、SS_A、SS_B、SS_rを求めることができます。実際の計算は、やりやすいようにやればよいのですが。

$$SS_{total} = n_B SS_A + n_A SS_B + SS_r$$

なのだから、

$$SS_{total} = n_B SS_A + (n_A SS_B + SS_r)$$

$$n_B SS_A = SS_{total} - (n_A SS_B + SS_r)$$

と変形して、括弧の中を列ごとの平方和の和としてもとめて、全体の変動可差し引き
同様に

$$n_A SS_B = SS_{total} - (n_B SS_A + +SS_r)$$

$$SS_r = SS_{total} - n_B SS_A - n_A SS_B$$

と計算するのが早そうな気がします。

これで、様々な要因を取り分けることができるようになったのですが、実は、もう一つ、検討しておくべき問題があります。今まで、検討してきたのは、ある要因が、一定の傾向を与えるかどうかと言うことでしたが、2つの要因を考えると、片方の要因が異なると、もう一つの要因の影響が異なるということがあります。たとえば、男女で薬の効果が違う。または反対の効果を持つとか、リンが一定以上ある場合には、肥料としての窒素が高い成長効果を持つが、リンが少ない場合には効果がないとか、特定の条件の時だけ、ある要因の影響が強く表れるという現象です。実際、こういう影響の現れ方は少なくないでしょう。こういうことを考えるにどうすればよいのかということです。実は、このことは、繰り返しの無い２要因分散分析では検討できません。何故かと言うと、繰り返しの無い２要因分散分析では、２つの要因で説明できない要因による分散を残差分散として一つにまとめています。この中を、普通のランダムな変動によるバラつきと、組み合わせの効果に分けることができません。この問題を解決するには、２つの要因を組み合わせた１つの条件で複数のデータがえられれば、その条件内（ここではセル内という表現を使います。）のデータのバラつきは条件に依存しないランダムなバラつきですから、これを使って、先ほどの繰り返しの無い２要因分散分析で、残差として取り扱ったものと区別すればよいことになり、セル内でのバラつきが残差になります。そして、先ほど残差として扱われたものは、２つの要因の組み合わせによる影響を含んでいるということで、交互作用という名前を与えられます。同じものの名前が変わるところが納得いかないかもしれません。私も納得いかないのですが、繰り返しがない場合には、交互作用の中にはランダムな変動という意味で「本来の残差」と「本来の交互」作用を区別することができないし、繰り返しの無い２要因分散分析では、「残差」と考えているのだと理解すればよいでしょう。

表 20. ランダムな変動として与える数値

	A_1	A_2	A_3
B_1	0.9	0.4	0.2
	0.5	0.3	0.7
	0.1	0.8	0.6
B_2	0.9	0.5	0.1
	0.4	0.3	0.8
	0.2	0.7	0.6
B_3	0.5	0.5	0.6
	0.4	0.5	0.3
	0.6	0.5	0.6
B_4	0.6	0.4	0.4
	0.4	0.4	0.2
	0.5	0.7	0.9

とにかく、繰り返しのあるモデルを考えます。ここでは、各セル内に３つの繰り返しを作ります。いままで、検討してきたモデル（というよりは事例）の各セルに、表 20 の値をたして、繰り返しを作ります。セル内の平均値に注目してください。全部等しく 0.5 ですから、個々の分散に影響を与えずに。全体の分散だけが増加します。増加する分散を計算しておきます。この計算には簡便な計算法があって、次のように平方和の式を変形します。

$$\sum_{i}^{n}(x_i-\bar{x})^2=\sum_{i}^{n}{x_i}^2-2\sum_{i}^{n}x_i\bar{x}+n\bar{x}^2$$

$\bar{x}$はx_iの平均値

このように展開できますが、ところで

$$\sum_{i}^{n}x_i\bar{x}=\bar{x}\sum_{i}^{n}x_i=\frac{\sum_i^n x_i}{n}\sum_{i}^{n}x_i=\frac{1}{n}\sum_{i}^{n}x_i{}^{2}$$

$$n\bar{x}^2=n\left(\frac{\sum_i^n x_i}{n}\right)^2$$

$$\because\ \bar{x}=\frac{\sum_i^n x_i}{n}$$

合計をデータ数で割れば平均値だといっているだけです。このままだと見にくいので、

$$\sum_{i}^{n}x_i=T$$

$$\sum_{i}^{n}{x_i}^2=S$$

と表すと、

$$SS = \sum_{i}^{n} (x_i - \bar{x})^2 = S - \frac{T^2}{n}$$

となります。これを使って、

Tを求める表

				合計
	0.9	0.4	0.2	1.5
	0.5	0.3	0.7	1.5
	0.1	0.8	0.6	1.5
	0.9	0.5	0.1	1.5
	0.4	0.3	0.8	1.5
	0.2	0.7	0.6	1.5
	0.5	0.5	0.6	1.6
	0.4	0.5	0.3	1.2
	0.6	0.5	0.6	1.7
	0.6	0.4	0.4	1.4
	0.4	0.4	0.2	1
	0.5	0.7	0.9	2.1
合計	6	6	6	18

Sを求める表				合計
	0.81	0.16	0.04	1.01
	0.25	0.09	0.49	0.83
	0.01	0.64	0.36	1.01
	0.81	0.25	0.01	1.07
	0.16	0.09	0.64	0.89
	0.04	0.49	0.36	0.89
	0.25	0.25	0.36	0.86
	0.16	0.25	0.09	0.5
	0.36	0.25	0.36	0.97
	0.36	0.16	0.16	0.68
	0.16	0.16	0.04	0.36
	0.25	0.49	0.81	1.55
合計	3.62	3.28	3.72	10.62

$$10.62 - \frac{18^2}{36} = 1.62$$

と計算します。コンピューターがなかった時代には、こういう計算過程の省略のためのテクニックも必要でしたが、今はもう必要ない技術かもしれません。でも、知識として知っておくと何かの役に立つかもしれません。

では、データの合成をします。

表 21. 合成されたデータ

	A_1	A_2	A_3
B_1	3.8	6.9	7.3
	3.4	6.8	7.8
	3.0	7.3	7.7
B_2	7.3	10.8	11.9
	6.8	10.6	12.6
	6.6	11.0	12.4
B_3	7.7	12.2	13.2
	7.6	12.2	12.9
	7.8	12.2	13.2
B_4	10.1	13.9	14.9
	9.9	13.9	14.7
	10.0	14.2	15.4

早速、今覚えたテクニックを使って計算します。

Tを求める表

				合計
	3.8	6.9	7.3	18
	3.4	6.8	7.8	18
	3	7.3	7.7	18
	7.3	10.8	11.9	30
	6.8	10.6	12.6	30
	6.6	11	12.4	30
	7.7	12.2	13.2	33.1
	7.6	12.2	12.9	32.7
	7.8	12.2	13.2	33.2
	10.1	13.9	14.9	38.9
	9.9	13.9	14.7	38.5
	10	14.2	15.4	39.6
合計	84	132	144	360

				合計
Sを求める表				
	14.44	47.61	53.29	115.34
	11.56	46.24	60.84	118.64
	9	53.29	59.29	121.58
	53.29	116.64	141.61	311.54
	46.24	112.36	158.76	317.36
	43.56	121	153.76	318.32
	59.29	148.84	174.24	382.37
	57.76	148.84	166.41	373.01
	60.84	148.84	174.24	383.92
	102.01	193.21	222.01	517.23
	98.01	193.21	216.09	507.31
	100	201.64	237.16	538.8
合計	656	1531.72	1817.7	4005.42

$$4005.42 - \frac{360^2}{36} = 405.42$$

この数字は、私たちの期待通りに、

$$3(4 \times 14 + 3 \times 26 + 0.6) + 1.62 = 405.42$$

となっています。つまり、繰り返しによる分散(残差分散)を$\sigma_{residual}$、交互作用の分散を$\sigma_{interaction}$とすると。

$$n_{residual}(n_B SS_A + n_A SS_B + SS_{\widehat{interaction}}) + SS_{residual} = SS_{total}$$

となっている。どうしてそうなるのかと言う説明は、ここでは、もう不要だと思います。

$$n_{residual} n_B SS_A = 168$$

$$n_{residual} n_A SS_B = 234$$

$$n_{residual} SS_{interaction} = 1.8$$

$$SS_{residual} = 1.62$$

これから

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

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

$$SS_{interaction} = \frac{1.8}{3} = 0.6$$

を求めて

$$\sigma_A = \frac{SS_A}{df_A} = \frac{14}{2} = 7$$

$$\sigma_B = \frac{SS_B}{\mathrm{df}_B} = \frac{26}{3} = 8.6667$$

$$\sigma_{interaction} = \frac{SS_{interaction}}{\mathrm{df}_{interaction}} = \frac{0.6}{6} = 0.1$$

$$\sigma_{residual} = \frac{SS_{residual}}{\mathrm{df}_{residual}} = \frac{SS_{residual}}{(n_{residual}n_A n_B - 1) - \mathrm{df}_A - \mathrm{df}_B - \mathrm{df}_{interaction}} = \frac{1.62}{35 - 2 - 3 - 6}$$

$$= \frac{1.62}{24} = 0.0675$$

分散比を計算すると

$$F_{A-residua} = \frac{7}{0.0675} = 103.7037$$

$$F_{B-residual} = \frac{8.6667}{0.0675} = 128.3956$$

$$F_{interaction-residua} = \frac{0.1}{0.0675} = 1.481$$

$$F_{A-interactio} = \frac{7}{0.1} = 70$$

$$F_{B-interacti} = \frac{8.6667}{0.1} = 86.667$$

この結果から、要因 A 要因 B によるデーターの広がりは、ランダムな分散（残差分散）およびデーターの組み合わせによる分散（交互作用）に比べて、十分な広がりを持って分布しており､観察された要因 A および要因 B による違いは､ランダムな変動の結果ではなく、統計的に意味のある違いである（統計的に有意）と結論できます。

ここまでの、説明は、どんな計算をしているのかを、理解してもらうために、論理的な説明と言うよりは、具体的な数字を使って具体的な計算をできるだけ想像しやすいように説明してきました。そのやり方は、普通の方法と違って、様々な、要因をたしあわせて、データを作るというやり方でした。そのために、どの行も列もセルも平均値として同じ値が加わるように工夫した数字を加えていきました。数字を加えることで、他の分散に変化が及ばないようにそうしたのですが、平均値を 0 にして、負の値も含めて考えたほうが考え方が簡単だったかもしれません。繰り返しになりますが、和の分散の説明で用いた図でさらに詳しい説明をします。

平均値を 0 にして考えるというところから解説を始めます。

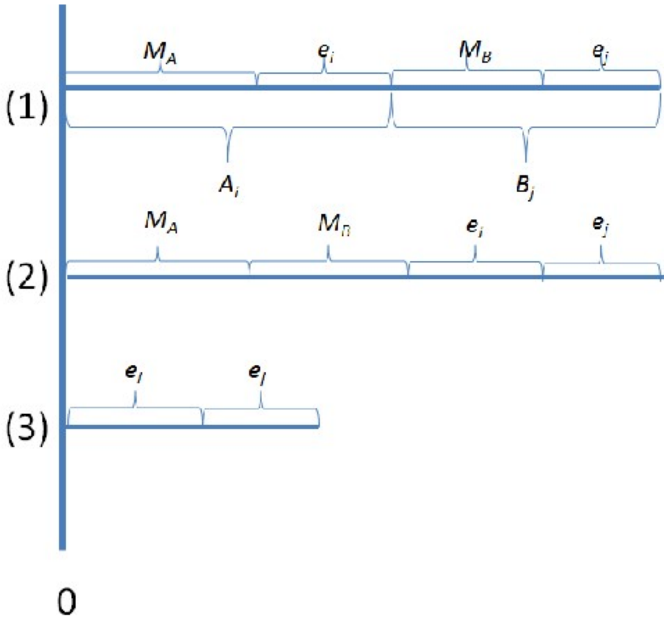

図 32. データをたしあわせるイメージ

（１）は、データーのたしあわせですが、たす順番を変えると（２）になり、平均値を差し引いてしまえば（３）になります。

ここで、この足し合わせたものの平均値からの隔たりをe_{A+B}と表しす (持ってまわった言い方ですが、簡単に言えば、（３）の直線の長さのをこのように表すということです。)。

$$e_{A+B_{ij}} = e_{A_i} + e_{B_j}$$

e_Aの平均を$\bar{e_A}$と表します。

$$\bar{e_A} = \frac{\sum_{i=1}^{n_A} e_{A_i}}{n_A}$$

ということです。この表現を使えば、標本データの２次の積率（平方和をデーター数で割った物）は、$\overline{e_A{}^2}$のように表せます。

$$\overline{e_A{}^2} = \frac{\sum_{i=1}^{n_A} e_{A_i}{}^2}{n_A}$$

ということです。

足し合わせてできた標本集団の 2 次の積率(分散)を考えます。すべてのデーター―を使っているので、データーの数は mn です。

どういう組み合わせ方になっているのかが問題ですが、すべてのデータの総和をどのように計算するかという計算の仕方で表すと

$$\sum_{i=1}^{n_A}\sum_{j}^{n_B} e_{A+B_{ij}} = \sum_{i=1}^{n_A}\sum_{j}^{n_B}\left(e_{A_i} + e_{B_j}\right) = \sum_{i=1}^{n_A}\left(n_B e_{A_i} + \sum_{j}^{n_B} e_{B_j}\right) = n_B\sum_{i=1}^{n_A} e_{A_i} + n_A\sum_{j}^{n_B} e_{B_j} = 0$$

という構造です。縦横のマトリックスを作って、総当り的にたしあわせるということです。

２乗の総和をデータ数で割って

$$\overline{e_{A+B}{}^2} = \frac{SS_{A+B}}{n_A n_B} = \frac{\sum_{i=1}^{n_A}\sum_{j}^{n_B} e_{A+B_{ij}}{}^2}{n_A n_B} = \frac{\sum_{i=1}^{n_A}\sum_{j=1}^{n_B}\left(e_{A_i} + e_{B_j}\right)^2}{n_A n_B}$$

この表記をつかうと

$$\overline{e_A{}^2} = \frac{\sum_{i=1}^{n_A}\sum_{j}^{n_B} e_{A_i}{}^2}{n_A n_B} = \frac{\sum_{i=1}^{n_A} e_{A_i}{}^2}{n_A} = \frac{SS_A}{n_A}$$

$$\overline{e_B{}^2} = \frac{\sum_{i=1}^{n_A}\sum_{j}^{n_B} e_{B_j}{}^2}{n_A n_B} = \frac{\sum_{j=1}^{n_B} e_{B_j}{}^2}{n_B} = \frac{SS_B}{n_B}$$

$$\because \sum_{j=1}^{n_B} e_{A_i}{}^2 = n_B e_{A_i}{}^2$$

$$\sum_{i=1}^{n_A}\sum_{j=1}^{n_B} e_{B_j}{}^2 = \sum_{j=1}^{n_B}\sum_{i=1}^{n_A} e_{B_j}{}^2$$

$\overline{e_{A+B}{}^2} = \frac{SS_{A+B}}{n_A n_B}$を展開します。$\frac{SS_{A+B}}{n_A n_B} = \frac{\sum_{i=1}^{n_A}\sum_{j=1}^{n_B}\left(e_{A_i}+e_{B_j}\right)^2}{n_A n_B} = \frac{1}{n_A n_B}\left\{\sum_{i=1}^{n_A}\sum_{j=1}^{n_B} e_{A_i}{}^2 + \right.$

$$\left. 2\sum_{i=1}^{n_A}\sum_{j=1}^{n_B} e_{A_i} e_{B_j} + \right\}$$

$$= \frac{1}{n_A n_B}\left\{ n_B \sum_{i=1}^{n_A} e_{A_i}{}^2 + 2\sum_{i=1}^{n_A} e_{A_i} \sum_{j=1}^{n_B} e_{B_j} + n_A \sum_{j=1}^{n_B} e_{B_j}{}^2 \right\}$$

$$= \frac{\sum_{i=1}^{n_A} e_{A_i}{}^2}{n_A} + \frac{\sum_{j=1}^{n_B} e_{B_j}{}^2}{n_B}$$

$$= \frac{SS_{\hat{A}}}{n_A} + \frac{SS_{\hat{B}}}{n_B}$$

整理して書くと

$$\frac{SS_{A+B}}{n_A n_B} = \frac{SS_A}{n_A} + \frac{SS_B}{n_B}$$

$$SS_{A+B} = n_B SS_A + n_A SS_B$$

(式 48)

となります。

あるいは

$$\sum_{i=1}^{n_A}\sum_{j=1}^{n_B} e_{A+B_{ij}} = n_B\sum_{i=1}^{n_A} e_{A_i} + n_A\sum_{j=1}^{n_B} e_{B_j} = 0$$

というデータ構造の時は

$$\sum_{i=1}^{n_A}\sum_{j=1}^{n_B} {e_{A+B_{ij}}}^2 = n_B\sum_{i=1}^{n_A} {e_{A_i}}^2 + n_A\sum_{j=1}^{n_B} {e_{B_j}}^2$$

という形で一般化すると応用範囲が広いかもしれません。

この場合は、総当りになので、要因 A については、n_B回、要因 B についてははn_B回の繰り返しがあるということです、しかし、データーの組み合わせ方によってはそうではない場合があります。要因 A+B と交互作用 I の組み合わせでは、

$$\sum_{i=1}^{n_A}\sum_{j=1}^{n_B} e_{A+B+I_{ij}} = n_B\sum_{i=1}^{n_A} e_{A_i} + n_A\sum_{j=1}^{n_B} e_{B_j} + \sum_{i=1}^{n_A}\sum_{j=1}^{n_B} e_{I_{ij}} = 0$$

$$1 \le i \le n_A, \qquad i\text{は整数}\quad i \in \mathbb{Z}$$

というデータの構成になっています。この SS を考えます。

$$\sum_{i=1}^{n_A}\sum_{j=1}^{n_B} {e_{A+B+I_{ij}}}^2 = \sum_{i=1}^{n_A}\sum_{j=1}^{n_B} \left(e_{A_i} + e_{B_j} + e_{I_{ij}}\right)^2$$

$$= \sum_{j=1}^{n_B}\sum_{i=1}^{n_A} {e_{A+B_{ij}}}^2 + 2\sum_{i=1}^{n_A}\sum_{j=1}^{n_B} e_{I_{ij}}\left(e_{A_i} + e_{B_j}\right) + \sum_{i=1}^{n_A}\sum_{j=1}^{n_B} {e_{I_{ij}}}^2$$

$$= n_B SS_A + n_A SS_B + 2\sum_{i=1}^{n_A} e_{A_i}\sum_{j=1}^{n_B} e_{I_{ij}} + 2\sum_{j=1}^{n_B} e_{B_j}\sum_{i=1}^{n_A} e_{I_{ij}} + + SS_I$$

$$= n_B SS_A + n_A SS_B + SS_r$$

$$\because\ \sum_{j=1}^{n_B} e_{I_{ij}} = \sum_{i=1}^{n_A} e_{I_{ij}} = 0$$

$\sum_{j=1}^{n_B} e_{I_{ij}} = \sum_{i=1}^{n_A} e_{I_{ij}} = 0$となるのは、私たちがデーター合成のときにそのようにデーターを作ったからです。実際の分析では、そのようになるモデルを考えるということです。

これらのデータとセル内の繰り返し、つまり、残差との関係を整理します。

$$\sum_{i=1}^{n_A}\sum_{j=1}^{n_B}\sum_{k=1}^{n_r} e_{A+B+I+r_{ijk}} = n_r n_B\sum_{i=1}^{n_A} e_{A_i} + n_r n_A\sum_{j=1}^{n_B} e_{B_j} + n_r\sum_{i=1}^{n_A}\sum_{j=1}^{n_B} e_{I_{ij}} + \sum_{i=1}^{n_A}\sum_{j=1}^{n_B}\sum_{k=1}^{n_r} r_{ijk} = 0$$

という構造です。全平方和を計算します。

$$\sum_{i=1}^{n_A}\sum_{j=1}^{n_B}\sum_{k=1}^{n_r} e_{A+B+I+r_{ijk}}{}^2 = \sum_{i=1}^{n_A}\sum_{j=1}^{n_B}\sum_{k=1}^{n_r}\left(e_{A+B+I_{ijk}} + e_{r_{ijk}}\right)^2$$

$$= \sum_{k=1}^{n_r}\sum_{i=1}^{n_A}\sum_{j=1}^{n_B} e_{A+B+I_{ijk}}{}^2 + 2\sum_{i=1}^{n_A}\sum_{j=1}^{n_B}\sum_{k=1}^{n_r} e_{A+B+I_{ij}}\, e_{r_{ijk}} + \sum_{i=1}^{n_A}\sum_{j=1}^{n_B}\sum_{k=1}^{n_r} e_{r_{ijk}}{}^2$$

$$= n_r(n_B SS_A + n_A SS_B + SS_I) + 2\sum_{i=1}^{n_A}\sum_{j=1}^{n_B} e_{A+B+I_{ij}} \sum_{k=1}^{n_r} e_{r_{ijk}} + \sum_{i=1}^{n_A}\sum_{j=1}^{n_B}\sum_{k=1}^{n_r} e_{r_{ijk}}{}^2$$

$$= n_r(n_B SS_A + n_A SS_B + SS_I) + SS_r$$

$$\because \text{ 各セル内で} \sum_{k=1}^{n_r} e_{r_{ijk}} = 0$$

各行ごとに平均値を求めると、平均値はB_jとなるので、行ごとの平均値の分散は、SS_B、 d
同様に、各列ごとの平均値はA_i、列ごとの平均値の分散はSS_A
各行ごとに平均値を求めて、その平均値からの距離の２乗を分散とした場合、行内には列間の分散がないので、これの値の全ての行の総和は

$$n_r(n_B SS_A + SS_I) + SS_r$$

これを全平方和から差し引けば、

$$SS_{A+B+I+r} - \{n_r(n_B SS_A + SS_I) + SS_r\} \ = n_r n_A SS_{\hat{B}}$$

同様に

$$SS_{A+B+I+r} - \{n_r(n_A SS_A + SS_I) + SS_r\} = n_r n_B SS_A$$

さらに別途SS_rを求めて、最後にSS_Iを求めればよいでしょう。

いずれにしてもデータの繰り返し数を考える必要があります。これについては

$$\sum_{i=1}^{n_A}\sum_{j=1}^{n_B}\sum_{k=1}^{n_r} e_{A+B+I+r_{ijk}} = \sum_{i=1}^{n_A}\sum_{j=1}^{n_B}\sum_{k=1}^{n_r}\left(e_{A_i} + e_{B_j} + e_{I_{ij}} + e_{r_{ijk}}\right)$$

のように、データの繰り返しがどんな構造になっているのかを考えれば良いでしょう。とくにそれぞれの要因についている添え字が重要です。ここでは添え字はi, j, kの３つです。e_{A_i}には３つの内j, kの添え字がありません。jの最大値はn_B、kの最大値はn_rです。それだけ繰り返しがあるということですから、A の要因の繰り返し数は$n_B n_r$です。同様に B の繰り返し数は$n_A n_r$、交互作用の繰り返し数はn_rで、すべてがそろっている、r の繰り返し数は１です。