分散分析 (その2:二元配置法)

福井 勉¹⁾

Statistical Method. Analysis of Variance. (Part 2)

Tsutomu FUKUI, RPT¹⁾.

1) Dept. of Physical Therapy, Tokyo Metropolitan College of Allied Medical Sciences, 7-2-10 Higashiogu Arakawa-ku Tokyo 116 Japan TEL 03-819-1211

J. Exerc. Physiol. 5(2):107-114, 1990. ISSN 0912-7100. Received April 20, 1990.

要旨

前回の分散分析一元配置に引続き、二元配置の考え方を述べます。PTの行う実験にはここまでの知識はかなり重要であると思われるので、概念を中心に述べ、その解析手順は簡略に解説しました。繰り返しのある場合と無い場合について述べます。計算の詳細は補足に述べました。

1) 東京都立医療技術短期大学 理学療法学科: 東京都荒川区東尾久7-2-10 (〒116)

TEL 03-819-1211

運動生理 5(2):107-114, 1990.

受付日1990年4月20日

1. 二元配置

職場でのPT同志の力関係はMMTでは判断できませんが、患者にはよく筋力測定を行います。その結果からこの患者は筋力があるかないかといった判断をします。仮に両手の握力が手関節の肢位によって異なるかどうかを調べたい場合があったとします。この時に例えば手関節角度を背屈20度, 10度, 0度の3水準で実験を行うとすれば一元配置法を用いれば良いことになり、前回解説した一元配置法を用いればめでたく結果が出ます。しかし、この時に

左右の手の握力の差も同時に調べたいという場合にはどうもできそうにはありません。このようなときに用いる方法、つまり因子を2つ割付けた実験のことを二元配置法といいます。

前回登場した構造模型で示すと、二元配置のデータ構造は因子Aの水準*i*、因子Bの水準*j*の時、

$$X_{ij} = \mu + \alpha_i + \beta_j + \varepsilon_{ij}$$

となります (繰り返し無し: 後述)。つまりデータは総平均とA*i*水準での効果 α_i とB*j*水準での効果 β_j および誤差 ε_{ij} に分解できるということです。

2. 二元配置のころ

両手の握力を背屈 20,10,0度で計測したところ表 1 のような結果でした。これについて考えてみます。この結果にもし何の差も無いとしたら全て同じ値が計測されるはずでした（表 2）。これは総平均を示したものです。表 1 から表 2 を差し引いた残りの部分が表 3 になります。そうすると、列の効果（背屈角）と行の効果（検側）さらに誤差の 3 つが含まれているはずですから、これらを分離していきます。

表 1 手関節肢位と検査側の違いによる握力
(単位;kg)

検側/背屈角	2 0	1 0	0
左手	39	38	37
右手	43	34	25

表 2 差が無いとすると

検側/背屈角	2 0	1 0	0
左手	36	36	36
右手	36	36	36

表 3 表 1 から表 2 を減じて

検側/背屈角	2 0	1 0	0
左手	3	2	1
右手	7	-2	-11

その過程を表 4 に示しました。行の平均及び列の平均は総平均からの違いの分だけの効果があるということになります。残りは誤差ですが、

$$\begin{aligned} (\text{誤差}) &= (\text{測定値}) - (\text{総平均}) - \\ &\quad (\text{行の効果; 検側}) - (\text{列の効果; 背屈角}) \end{aligned}$$

ですから、誤差を表 5 のように分離できました。

表 5 誤差の分離

検側/背屈角	2 0	1 0	0
左手	-4	0	4
右手	4	0	-4

この先の手順は、前回の分散分析一元配置とほぼ同様です。

$$\frac{\text{行の効果の不偏分散}}{\text{誤差の不偏分散}} = F 1$$

$$\frac{\text{列の効果の不偏分散}}{\text{誤差の不偏分散}} = F 2$$

を計算し、F 表から求めた値と比較して有意性を判断します。不偏分散 V は、

$$V = \frac{\sum (x_i - \bar{x})^2}{\phi}$$

ϕ は、

表 4 行の効果と列の効果

検側/背屈角	2 0	1 0	0	行の合計	行の平均	行の効果
左手	39	38	37	114	38	2
右手	43	34	25	102	34	-2
列の合計	82	72	62			
列の平均	41	36	31			
列の効果	5	0	-5			

総平均=36

誤差の自由度； $\phi_E = (\text{行の数} - 1) \cdot (\text{列の数} - 1)$

行の自由度； $\phi_A = (\text{行の数} - 1)$

列の自由度； $\phi_B = (\text{列の数} - 1)$

ですので、それぞれの因子の変動の不偏分散は次のようになります。

$$\text{行の効果の不偏分散} = \frac{2^2 \times 3 + (-2)^2 \times 3}{2-1} = 24$$

(列数を掛けています)

列の効果の不偏分散

$$= \frac{5^2 \times 2 + 0^2 \times 2 + (-5)^2 \times 2}{3-1} = 50$$

(行数を掛けています)

また、表5より、

$$\text{誤差の不偏分散} = \frac{(-4)^2 \times 2 + 4^2 \times 2}{(2-1) \times (3-1)} = 32$$

よって $F_1 = 0.75$, $F_2 = 1.56$ となります。あとはこの値 F_0 を F 表の値と比べればいいわけです。つまり、

$F_0 \geq F(\phi_A, \phi_E; \alpha)$ ならば、危険率 α で有意であり、

$F_0 < F(\phi_A, \phi_E; \alpha)$ ならば、危険率 α で有意でない

ということになります(列の効果の判定をするときは ϕ_A の所には ϕ_B をいれる)。

$F(1, 2; 0.05) = 18.5$, $F(2, 2; 0.05) = 19.0$ であるので、いずれも有意でないということになります。すなわち、握力は背屈角度や左手右手といった要因によって違うということは言えないということです。ですから、左右どちらの手で行っても背屈角度がこの中のどの角度でも握力が大きいとか小さいといったことは言えないことになります。これが二元配置のところです。これを下記のように示したものを分散分析表といいます(補足参照)。

表6 分散分析表

要因	S	ϕ	V	F_0
A	24	1	24	0.75
B	100	2	50	1.56
E	64	2	32	
計	188	5		

3. 繰り返しのある場合

私の友人でふられた相手が必ず獅子座のA型だというのがいます。前の例で「20度背屈位で特に左手握力が大きい」とか「右手0度で特に小さい」といったように組合せによる効果のことを交互作用といいます。組合せのすごさ、生まれ星座や血液型による相性占いといったところでしょうか。繰り返しが無い場合には、交互作用と誤差は分離ができないために当然交互作用は求めることができません。しかし、実験を繰り返すことによって交互作用を分離することができ、また精度も上がります。つまり全体ばらつきSの中身は次のようになります(図1)。

繰り返しのない場合

S_A	S_B	S_E
-------	-------	-------

繰り返しのある場合

S_{AB}			S_E
S_A	S_B	$S_{A \times B}$	

図1 総変動Sの中身

表 7 5回の繰り返し

検側/背屈角	2 0	1 0	0
左手	38	36	39
	46	54	52
	20	37	36
	31	30	32
	34	36	40
右手	33	33	37
	23	24	24
	25	29	23
	23	20	20
	30	32	33

では、実際に繰り返しの実験を行った場合について、進めて行きます。先ほどと同じ実験をそれぞれ5回ずつ繰り返す実験を行ったとします(表7)。実験回数は $2 \times 3 \times 5 = 30$ 回となるので、背屈角度と検側方向についてそれぞれランダム化することは一元配置となんら違いありません。実際の手順にはいります。検側を因子A、背屈角を因子Bとします。

$A : A_1, A_2, \dots, A_l$

$B : B_1, B_2, \dots, B_m$

で、 $A_i B_j$ での組合せときの結果で繰り返し数 r の k 番目のデータを x_{ijk} で表します。よって、下記のことを成り立ちます。

$\bar{x}_{...} = \sum \sum \sum x_{ijk} / lmr$

$\bar{x}_{i..} = \sum \sum x_{ijk} / mr$

$\bar{x}_{.j.} = \sum \sum x_{ijk} / lr$

$\bar{x}_{ij.} = \sum x_{ijk} / r$

(1) 数値変換を行う。計算を簡単にするために与えられたデータ (x_{ijk}) を簡単なデータ (X_{ijk}) に変換します。

$X_{ijk} = (x_{ijk} - x_0) \cdot h = (x_{ijk} - 30) \cdot 1$
変換したデータを表8に示しました。

そして各水準 (A_i) 及び (B_j) ごとのデータの和をとり、その和の2乗の表(表9)並びに表8の2乗表を作成する(表10)。さらに表9の各水準毎の値を2乗した補助表をつくる(表11)。

表 8 表7の変換データ

検側/背屈角	2 0	1 0	0
左手	8	6	9
	16	24	22
	-10	7	6
	1	0	2
	4	6	10
右手	3	3	7
	-7	-6	-6
	-5	-1	-7
	-7	-10	-10
	0	2	3

表 9 行和と列和とその2乗

検側/背屈角	2 0	1 0	0	行和	(行和) ²
左手	19	43	49	111	12321
右手	-16	-12	-13	-41	1681
列和	3	31	36	70	14002
(列和) ²	9	961	1296	2266	

表10 2乗表

検側/背屈角	2 0	1 0	0
左手	64	36	81
	256	576	484
	100	49	36
	1	0	4
	16	36	100
右手	9	9	49
	49	36	36
	25	1	49
	49	100	100
	0	4	9

合計=2364

表11 補助表

361	1849	2401
256	144	169

合計=5180

(2) 検定

下記の式にしたがって計算してみる。だいたい、計算ミスはここで起きるので、いかに辛抱強くかつまじめに計算するかということだけである。もうひとつは、有意差が出るようにと折りながらやるのがポイントである。

総平方和 $S_T = \sum \sum \sum (x_{ijk} - \bar{x} \dots)^2$

A B 級間平方和

$S_{AB} = r \sum \sum (\bar{x}_{ij.} - \bar{x} \dots)^2$

A 間平方和 $S_A = m r \sum (\bar{x}_{i..} - \bar{x} \dots)^2$

B 間平方和 $S_B = l r \sum (\bar{x}_{.j.} - \bar{x} \dots)^2$

交互作用平方和 $S_{A \times B} = S_{AB} - S_A - S_B$

誤差平方和 $S_E = S_T - S_{AB}$

これを計算して、

$S_T = 2200.7$

$S_A = 770.14$

$S_B = 63.27$

$S_{A \times B} = 39.26$

$S_E = 1328$

$\phi_T = lmr - 1 = 29$

$\phi_A = l - 1 = 1$

$\phi_B = m - 1 = 2$

$\phi_{A \times B} = (l - 1)(m - 1) = 2$

$\phi_E = lm(r - 1) = 24$

不偏分散は $V_A = S_A / \phi_A$

$V_B = S_B / \phi_B$

$V_{A \times B} = S_{A \times B} / \phi_{A \times B}$

であるので、分散分析表は、表12のようになる。

表12 分散分析表

要因	S	ϕ	V	F0
A	770.1	1	770.1	13.91**
B	63.27	2	31.64	0.57
A × B	39.26	2	19.63	0.35
E	1328	24	55.33	
	2200.7	29		

検定は、

$(F(1, 24, 0.01) = 7.82) < F_0 = 13.91$

$(F(2, 24, 0.05) = 3.40) > F_0 = 0.57$

$(F(2, 24, 0.05) = 3.40) > F_0 = 0.35$

よって左手と右手の筋力には違いがあるということになる。しかし、背屈角度による違いは認められない。さらに、組合せの効果であるところの交互作用は認められなかった。お疲れさまでした(補足参照)。

4. おわりに

今回の例では「実験に対する慣れ」あるいは「疲労」などといった要因は、統計学のみではわからないわけです。つまり、この実験を行う際の固有技術というべきものが実験者側に無いとできないということになります。固有技術が確立しているような分野では分散分析は非常に有効となってきますが、1989年の理学療法士学会でも分散分析はほとんど用いられていません。特に実験順序をランダム化していない演題などは数知れません。安易に統計ソフトを用いるといろいろな障害が生じます。しかしながらきつと、これからはこの手法が多く使われることになることと思います。また、ここでは触れませんでした実験計画法には、経済的に実験を進めるための数多くの手法が用意されています。これらを有効に用いるためにも理学療法そのものの、すくなくとも統計負けしないようにしたいものです。

※ 2 元配置のころ (補足)

背屈角や左右の手の違いは握力に影響を及ぼさないときの6個のデータはすべて同じ値つまり6個のデータの総平均 ($\bar{x}_{..} = \sum \sum x_{ij} / 1 \cdot m$) と考えられます (表2)。1は背屈角の水準数、mは左右の違いの水準数です。

しかし、いくら正確に計測したとしても人間のやることですからある程度の誤差は避けられません。この避けられない誤差は、測定値と総平均の差 ($x_{ij} - \bar{x}_{..}$) で求めることができます (表3)。これを2乗して合計したのが全体のばらつきで総変動 ($S_T = (\sum \sum x_{ij} - \bar{x}_{..})^2$) といえます。

しかし、背屈角や手の違いが握力に影響を及ぼす場合はどうなるでしょう。測定値は測定誤差だけでなく、背屈角の違いや左右の手の違いによる影響を被ることになり、総変動 $S_T =$ 因子Aの水準間の変動 (S_A) + 因子Bの水準間の変動 (S_B) + 誤差変動 (S_E) となります。そして、背屈角の違いにより握力が変化するとすれば、背屈角の水準間の変動が誤差変動より十分大きいとき、その因子の水準間には有意差があると判定します。では、表4をもとに具体的なばらつき分解方法を説明しましょう。

繰り返しの無い場合

(1) 総変動 (S_T)

総変動 (総平方和) は先に説明しましたように、測定値と総平均の差を2乗し、合計することにより求められます。

$$S_T = \sum \sum (x_{ij} - \bar{x}_{..})^2 \\ = \sum \sum x_{ij}^2 - (\sum \sum x_{ij})^2 / 1m$$

ここで、 $(\sum \sum x_{ij})^2 / 1m$ を CT (修正項: Correction Term) とおくと

$$S_T = \sum \sum x_{ij}^2 - CT \text{ とも書けます。}$$

(2) 因子の水準間の変動

因子の水準間の変動を求めるには、水準内におけるデータのばらつき (誤差) をなくすることが必要です。それには、水準間の平均を用

いれば良いことになります。したがって、因子の水準間の変動は、水準間の平均と総平均の差を2乗し合計することにより求められます。因子Aの水準間の変動 (平方和) をもとめる。

$$S_A = \sum \sum (\bar{x}_{i.} - \bar{x}_{..})^2 \\ = m \sum \bar{x}_{i.}^2 - 2m \sum \bar{x}_{i.} \bar{x}_{..} + 1 \cdot m \bar{x}_{..}^2 \\ \bar{x}_{i.} = T_{i.} / m \quad m \sum \bar{x}_{i.} = T_{..} \quad \text{より}$$

$$S_A = \sum T_{i.}^2 / m - (T_{..}^2 / 1m) \\ = \sum T_{i.}^2 / m - CT \\ \text{となります。}$$

因子Bの水準間の変動 (平方和) も同様となります。

$$S_B = \sum \sum (\bar{x}_{.j} - \bar{x}_{..})^2 \\ = \sum T_{.j}^2 / 1 - CT$$

となります。

(3) 誤差変動

誤差変動 (S_E) は、総変動と各因子の変動との差から求められます。

$$S_E = S_T - S_A - S_B$$

表13 繰り返し無し

	B1	B2	...	Bj	...	Bm	計	2乗	平均
A1	x ₁₁	x ₁₂	...	x _{1j}	...	x _{1m}	T _{1.}	T _{1.} ²	$\bar{x}_{1.}$
A2	x ₂₁	x ₂₂	...	x _{2j}	...	x _{2m}	T _{2.}	T _{2.} ²	$\bar{x}_{2.}$
.	.	.	...	.	...	.	.	.	.
.	.	.	...	.	...	.	.	.	.
Ai	x _{i1}	x _{i2}	...	x _{ij}	...	x _{im}	T _{i.}	T _{i.} ²	$\bar{x}_{i.}$
.	.	.	...	.	...	.	.	.	.
.	.	.	...	.	...	.	.	.	.
A1	x ₁₁	x ₁₂	...	x _{1j}	...	x _{1m}	T _{1.}	T _{1.} ²	$\bar{x}_{1.}$
計	T _{1.}	T _{2.}	...	T _{j.}	...	T _{m.}	T _{..}	$\sum T_{i.}^2$	$\bar{x}_{..}$
2乗	T _{1.} ²	T _{2.} ²	...	T _{j.} ²	...	T _{m.} ²	$\sum T_{i.}^2$		
平均	$\bar{x}_{1.}$	$\bar{x}_{2.}$	...	$\bar{x}_{j.}$	...	$\bar{x}_{m.}$			

(4) 計算の手順

表1のデータを使って実際の計算を手順を追って説明します。

(1) 数値変換、合計、2乗和、平均の計算 (表14)

計算を容易にするために数値変換を行い、総合計 ($\sum \sum x_{ij}$)、各因子ごとの2乗和 ($\sum T_{i.}^2$, $\sum T_{.j}^2$) を求めます。数値変換については「運動生理 (第4巻第4号) p226」を参照してください。

表14 補助表

	B 1	B 2	B 3	計	2 乗
A 1	3	2	1	6	36
A 2	7	-2	-11	-6	36
計	10	0	-10	0	72
2 乗	100	0	100	200	↑
				↑	$\sum T_{i.}^2$
				$\sum T_{.j}^2$	
	$\sum \sum x_{ij} = 0$				

(2) 2乗表

総変動を求めるため2乗表を作ります(表15)。

表15 2乗表

	B 1	B 2	B 3	計
A 1	9	4	1	14
A 2	49	4	121	174
計	58	8	122	188
				↑
	$\sum \sum x_{ij}$			

(3) 修正項を求める。

$$CT = (\sum \sum x_{ij})^2 / lm = 0^2 / 6 = 0$$

(4) 各変動を求める。

$$S_T = \sum \sum x_{ij}^2 - CT$$

$$= 188 - 0 = 188$$

$$S_A = \sum T_{i.}^2 / m - CT$$

$$= 72 / 3 - 0$$

$$= 24$$

$$S_B = \sum T_{.j}^2 / l - CT$$

$$= 200 / 2 - 0$$

$$= 100$$

$$S_E = S_T - S_A - S_B = 188 - 100 - 24 = 64$$

以上が計算手段です。

※. 繰り返しのある場合(補足)

繰り返しのある場合はすると各測定値は表16の様に表現できます。したがって、各変動は次のようにして求めることができます。

(1) 総変動 (S_T)

$$S_T = \sum \sum \sum (x_{ijk} - \bar{x} \dots)^2$$

$$= \sum \sum \sum x_{ijk}^2 - T_{\dots}^2 / lmn$$

$$= \sum \sum \sum x_{ijk}^2 - CT$$

(2) 因子 A、B の水準間の組合せの変動 (S_{AB})

$$S_{AB} = r \sum \sum (\bar{x}_{ij.} - \bar{x} \dots)^2$$

$$= \sum \sum T_{ij.}^2 / r - CT$$

(3) 各因子の水準間の変動

$$S_A = \sum \sum \sum (\bar{x}_{i..} - \bar{x} \dots)^2$$

$$= \sum T_{i..}^2 / mr - CT$$

$$S_B = \sum \sum \sum (\bar{x}_{.j.} - \bar{x} \dots)^2$$

$$= \sum T_{.j.}^2 / lr - CT$$

(4) 相互作用、誤差変動

$$S_A \times B = S_{AB} - S_A - S_B$$

$$S_E = S_T - S_{AB}$$

表16 繰り返しあり

	B 1	...	B j	...	B m	計	平均
A 1	x_{111}	...	x_{1j1}	...	x_{1m1}		
	.		.		.		
	x_{11r}	...	x_{1jr}	...	x_{1mr}		
小計	$t_{1.}$		$t_{j.}$		$t_{m.}$	$T_{1..}$	$\bar{x}_{1..}$
.....							
	x_{i11}	...	x_{ij1}	...	x_{im1}		
	.		.		.		
A i	x_{i1k}	...	x_{ijk}	...	x_{imk}		
	.		.		.		
	x_{i1r}	...	x_{ijr}	...	x_{imr}		
小計	$t_{i1.}$		$t_{ij.}$		$t_{im.}$	$T_{i..}$	$\bar{x}_{i..}$
.....							
	x_{111}	...	x_{1j1}	...	x_{1m1}		
	.		.		.		
A l	x_{l1r}	...	x_{ljr}	...	x_{lmr}		
	.		.		.		
小計	$t_{l1.}$		$t_{lj.}$		$t_{lm.}$	$T_{l..}$	$\bar{x}_{l..}$
計	$T_{.1.}$		$T_{.j.}$		$T_{.m.}$	$T_{...}$	

(5) 計算の手順

表7のデータを使って実際の計算を手順を追って説明します。

(1) 数値変換と2乗表

数値変換した結果を表8に示してあります。また全データの2乗表を作ります(表17)。

(2) 繰り返しを合計した表18およびその2乗表(表19)を作ります。

(3) 修正項を求める。

$$CT = T_{...}^2 / lmr$$

$$= 70^2 / 3 \cdot 2 \cdot 5$$

$$= 4900 / 30$$

$$= 163.33$$

(4) 各変動を求める。

$$S_T = \sum \sum \sum x_{ijk}^2 - CT$$

$$= 2364 - 163.33$$

$$= 2200.67$$

$$S_{AB} = \sum \sum T_{ij}^2 / r - CT$$

$$= 5180 / 5 - 163.33$$

$$= 872.67$$

$$S_A = \sum T_{i..}^2 / mr - CT$$

$$= 14002 / 3 \cdot 5 - 163.33$$

$$= 770.14$$

$$S_B = \sum T_{.j.}^2 / lr - CT$$

$$= 2266 / 2 \cdot 5 - 163.33$$

$$= 63.27$$

$$S_A \times B = S_{AB} - S_A - S_B$$

$$= 872.67 - 63.27 - 770.14$$

$$= 39.26$$

$$S_E = S_T - S_{AB}$$

$$= 2200.67 - 872.67$$

$$= 1328.0$$

以上です。

表18 補助表

	B 1	B 2	B 3	計	
A 1	19	43	49	111	12321
A 2	-16	-12	-13	-41	1681
計	3	31	36	70	14002
	9	961	1296	2266	↑
				↑	$\sum T_{i..}^2$
					$\sum T_{.j.}^2$
					$T_{...} = 70$

表17 2乗表

	B 1	B 2	B 3	計
	64	36	81	181
	256	576	484	1316
A 1	100	49	36	185
	1	0	4	5
	16	36	100	152
	9	9	49	67
	49	36	36	121
A 2	25	1	49	75
	49	100	100	249
	0	4	9	13
計	569	847	948	2364

↑

$$\sum \sum \sum x_{ijk}^2$$

表19 2乗表

	B 1	B 2	B 3	計
A 1	361	1849	2401	4611
A 2	256	144	169	569
計	617	1993	2570	5180
				↑
				$\sum \sum T_{ij}^2$

謝辞

最後に本論をまとめるにあたりご指導頂いた相沢直行氏はじめHuman Factors研究会の諸先生方に深謝致します。