

論文の書き方6

分散分析(その2)

星川佳広 CSCS, NSCA ジャパン編集委員, 東海学園大学准教授

本稿は、「論文の書き方1：投稿論文を書こう」(2013年10月号)の続編パート5です。NSCAジャパンウェブサイトに掲載中の「投稿論文(事例報告)の書き方」を参照しながら読み進めてください。

【投稿論文(事例報告)の書き方】

NSCAジャパンウェブサイトTOP → [指導者の育成] → [事例報告・研究論文]、[投稿要領]部分

前号、「分散分析(その1)」で紹介したのは、1つの要因(要因Ⅰ)の効果の有無を判断する一元配置の分散分析でした。今回紹介するのは二元配置分散分析で、2つの要因(要因Ⅰ、要因Ⅱ)の効果と、その2つの要因の交互作用の効果を判断する方法です。本記事では、二元配置分散分析を理解する上での基本事項について理解を深めた後、Excelの分析ツールを使った二元配置分散分析を紹介します。本稿の第4節、「繰り返しのない二元配置分散分析」は、対応する群による一元配置分散分析と実質同じで、私たちS&C専門職としては、その利用のほうが多いかと思しますので、それに沿った紹介をします。

1. 二元配置分散分析の概要

一元配置分散分析が1つの要因の効

果の有意性を調べるものであったので、二元配置分散分析が2つの要因の効果を調べるものということは、およそ想像がつくかと思います。例えば、異なるトレーニング方法を異なる対象に当てはめる場合を考えます。今、A、Bという2つのトレーニング方法(要因Ⅰ)があって、トレーニングを実施する対象が若者と高齢者の2つの年齢層(要因Ⅱ)であるとします。このときトレーニング効果(トレーニング前後の変化量)に対して二元配置分散分析を実施すると、①A、Bのトレーニング方法に違いがあるか(要因Ⅰの主効果)、②年齢(若者と高齢者)に違いがあるか(要因Ⅱの主効果)を調べることができます(図1)。

しかし、それだけではありません。要因が2つになると、若者はトレーニングBで効果が大きい、高齢者はト

レーニングAで効果が大きい(図1下段中央)、というような要因の組み合わせによる効果が考えられるからです。すなわち、二元配置分散分析では、③若者(あるいは高齢者)のみに変化が大きいトレーニング方法があるかも調べることができます。このことを**交互作用効果**(要因Ⅰ×要因Ⅱ)といいます。二元配置分散分析は、単にそれぞれの要因について一元配置分散分析を2回繰り返すことと同じではなくて、交互作用の効果も併せて検証することが特徴ともいえます。つまり、二元配置分散分析は、個々のデータを「個々のデータ=全体の平均+要因Ⅰの効果+要因Ⅱの効果+要因Ⅰ×Ⅱの交互作用の効果+誤差」とみなすことができるかどうかの判断基準を提供します(前号1節参照)。

2. Excelによる二元配置分散分析¹

Excelのデータ分析には、二元配置分散分析として、“繰り返しのある”二元配置と、“繰り返しのない”二元配置が用意されています。“繰り返しのある”とは、要因Ⅰと要因Ⅱの組み合わせに対して、データが複数ある場合です。“繰り返しのない”とは、要因Ⅰと要因Ⅱの組み合わせに対して1つのデータしかない場合です。繰り返しのある二元配置分散分析では、交互作用も同時に検定できますが、繰り返しのない二元配置分散分析では、計算上、交互作用効果を検定できません。したがって、繰り返しのない二元配置分散分析は、交互作用がないとわかっている場合にしか適用できません。

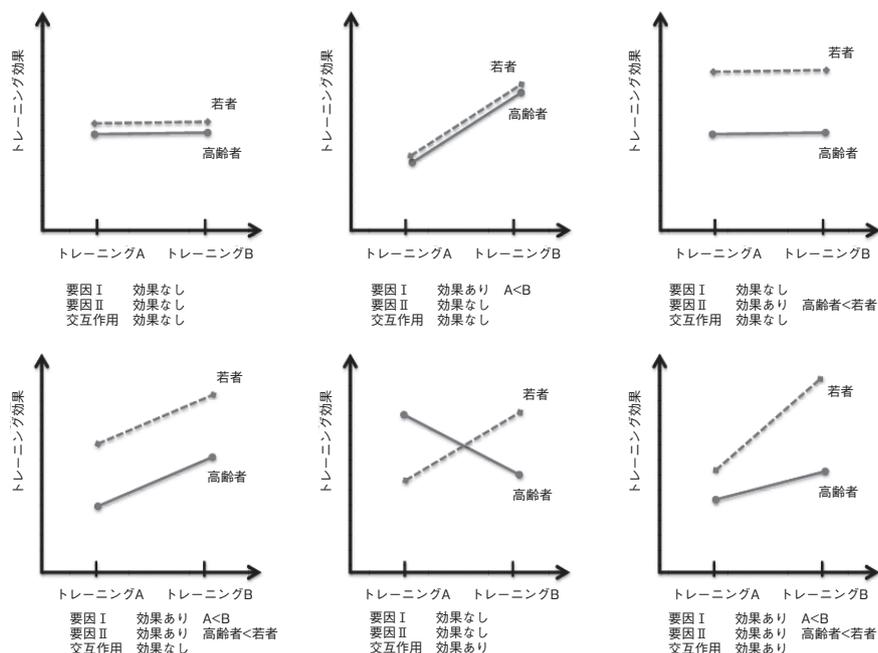


図1 トレーニング方法(要因Ⅰ)と年齢(要因Ⅱ)の二元配置分散分析のイメージ

3. 繰り返しのある二元配置分散分析

前号の続きで、あなたはエアロビクスのインストラクターで、レッスン内容としてA、B、Cの3パターンをもっているとします。前号の一元配置分散分析の結果は、レッスンCがレッスンAよりも体脂肪率の減少効果が有意に大きい、というものでした。ここで前号で分析したデータ例(前号表2)は、対象が大学生であったとします。そうすると、もしかしたらクライアントの年齢層が変わると、レッスン効果も異なる可能性があります。そこで新たに中高年のクライアント3群(各9名)を対象にして、同じようにレッスンA～Cを実施し、体脂肪率の減少効果を検証することにしました。そして、そのデータが表1のように得られたとします。

表1のデータ構造は、レッスン(要因Ⅰ)とクライアントの年齢層(要因Ⅱ)

	A	B	C	D
1		レッスンA	レッスンB	レッスンC
2	大学生	-1.5	0.7	-1.6
3		-0.2	-0.8	-1.2
4		0.9	-0.1	-1.8
5		0.3	0.7	-0.7
6		0.1	-0.9	0.2
7		-0.7	-0.2	0.5
8		-0.3	-0.9	-2.1
9		0.5	-0.4	-1.8
10		-0.4	-1.6	-2
11	中高年	-1.5	-0.3	-3.6
12		-1.2	-0.8	-2.2
13		-0.1	-0.1	-2.8
14		-0.7	-0.3	-1.7
15		0.1	-0.9	-1.6
16		-1.7	-0.2	-2
17		-0.3	-0.9	-3.1
18		0.5	-1.4	-2.8
19		-0.4	-1.0	-4

¹ 実際には、分散分析後の事後検定や多重比較、対応のある群での分散分析など、分散分析を十分かつ柔軟に実施しようとすると、Excelのデータ分析ツールのみでは力不足。もちろんExcelの関数を駆使して自分で計算することは可能だが、統計専用のソフト(SPSSなど)のほうが利便性は高い。

II)の組み合わせにデータが複数(9ずつ)あります。したがって、ここでは“繰り返しのある”二元配置分散分析を利用します。そうすることで、体脂肪率減少効果について、レッスンA、B、Cの違い(要因Iの主効果)とクライアントの年齢層の違い(要因IIの主効果)と、さらに、レッスンとクライアントの年齢層の交互作用効果の有無について、検証することができます。

Excelでデータ分析を呼び出し、分散分析：繰り返しのある二元配置を選択します。そうすると図2のボックスが出現します。表1の例では、入力範囲に「\$A\$1:\$D\$19」を指定します。1標本あたりの行数には、標本数(各グループの被検者数)9を入力します²。出力先に\$A\$25を指定してOKをします。その結果、表2(A25:G54の領域)が現れます。

概要(A27:E44の領域)には、レッスンA、B、Cそれぞれに関して、大学生、中高年、および合計(大学生と中高年の全データ)の標本数や平均、分散などの基本的な統計量が記されています。図3は、これらの基本統計量を使って、レッスン(要因I)と年齢層(要因II)の組み合わせの平均値をグラフ化したものです。

表2において大事なのは、分散分析表(A47:E54の領域)に出力されている内容です。変動要因(セルA49~A51)には、標本、列、交互作用と記されていますが、この場合、標本=年齢層(要因II)、列=レッスン(要因I)、交互作用=要因I×IIに相当します。セルD49~D51にそれぞれの分散が出力されていて、これと繰り返し誤差の分散(セルD52)の比、すなわちF値(前号参照)がセルE49~E51に出力

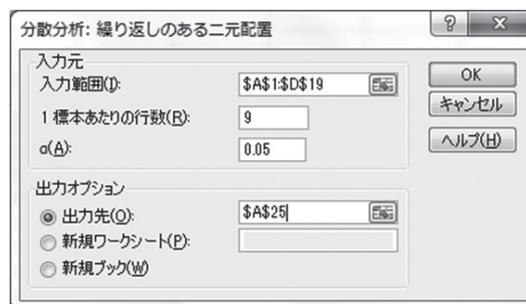


図2 分析ツール(繰り返しのある二元配置)

	A	B	C	D	E	F	G	H
24								
25	分散分析: 繰り返しのある二元配置							
26								
27	概要	レッスンA	レッスンB	レッスンC	合計			
28	大学生							
29	標本数	9	9	9	27			
30	合計	-1.3	-3.5	-10.5	-15.3			
31	平均	-0.14444	-0.38889	-1.16667	-0.56667			
32	分散	0.500278	0.581111	0.9275	0.815385			
33								
34	中高年							
35	標本数	9	9	9	27			
36	合計	-5.3	-5.9	-23.8	-35			
37	平均	-0.58889	-0.65556	-2.64444	-1.2963			
38	分散	0.558611	0.197778	0.700278	1.392678			
39								
40	合計							
41	標本数	18	18	18				
42	合計	-6.6	-9.4	-34.3				
43	平均	-0.36667	-0.52222	-1.90556				
44	分散	0.550588	0.385359	1.344085				
45								
46								
47	分散分析表							
48	変動要因	変動	自由度	分散	観測された分散比	P-値	F 境界値	
49	標本	7.186852	1	7.186852	12.44277	0.000935	4.042652	**
50	列	25.83593	2	12.91796	22.36518	1.36919E-07	3.190727	**
51	交互作用	3.849259	2	1.92463	3.332158	0.044146	3.190727	*
52	繰り返し誤差	27.72444	48	0.577593				
53								
54	合計	64.59648	53					
55								

されています。

そして分散分析表の中で最も重要なのがセルF49~F51にあるP-値です。表2の例だと、標本、列のP-値はいずれも非常に小さく³、1%水準($p < 0.01$)で有意な効果があるといえ

ます。すなわち、体脂肪率減少効果は、レッスンごとに有意差があり、また、年齢層によっても有意に異なると判断することができます。さらには交互作用についてもP-値=0.0441…なので、5%水準($p < 0.05$)で有意な効果

2 Excelのデータ分析での繰り返しのある二元配置分散分析は、繰り返しの数(標本数)が水準ごとに同一である必要があります。

3 1.37E-07(セルF50)は、 1.37×10^{-7} を表す。したがって、0.000000137...

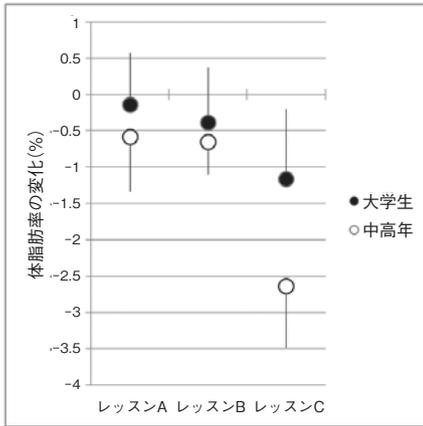


図3 表2概要のデータをグラフ化

があると判断できます。

交互作用の有意な効果とは、**図3**によって理解できると思います。レッスンCは、対象が大学生、中高年者いずれであっても、レッスンA、Bよりも体脂肪率の減少効果が大きいですが、レッスンCを中高年が受けた場合の体脂肪率減少効果は、大学生が受けた場合と比べるときわめて大きい。このような要因I、要因IIのある水準間の組み合わせではかとは異なる傾向のデータが得られているときに交互作用が有意になります。ちなみに、今回のデータとは異なり、レッスンCと中高年の組み合わせが、大学生の結果と異なりむしろ体脂肪率が増えてしまった、というデータであっても交互作用は有意になります。

4. 繰り返しのない二元配置分散分析

Excelに用意されている分散分析は、一元配置、二元配置いずれにおいても、水準間で独立した(対応のない)データであることを前提とします。前節の例でいうとレッスンA、B、Cを受けるクライアントは異なっていることが前提です。もし同じ人が反復してレッスン

を受けたとすると、その人がもつ特性がデータに影響し、異なるクライアントがレッスンを受けた場合に比べて誤差(バラつき)が小さくなるからです。

前節の要因IIが、例えば年代(大学生と中高年)ではなくて、レッスン時間(朝、昼、晩)であったとしましょう。体脂肪率減少効果が最も大きいのは、どのレッスンをどの時間帯に行なうと良いか、を調べるイメージです。この場合、同じ被検者がレッスンA、B、Cを受けて効果の違いを比較したり、あるいは同じ被検者がレッスンを朝、昼、晩と受けて効果の違いを比較したりしたくなるものです。しかし、Excelのデータ分析ツールを使って、2つの要因の効果を検証する場合には、“繰り返しのない二元配置分散分析”であろうと“繰り返しのある二元配置分散分析”であろうと、レッスンA～Cと、レッスンを行なう時間帯(朝、昼、晩)でクライアントは違ってなければなりません。この点には注意しましょう。

その一方で、私たちS&Cの分野にあっては、データ分析ツールの繰り返しのない二元配置分散分析が有用な場面が別にあります。それは、一元配置分散分析において1つの要因内のすべての水準に対して、同じ被検者から反復してデータが得られている場合です。例えばレッスンA、B、Cを同じクライアントが行なった場合で、すなわち要因間のデータに対応のある場合です。この場合、検討する事項はレッスン間(要因I)のみの比較ですが、実質的に繰り返しのない二元配置分散分析が利用できます⁴。私たちS&Cの分野では、むしろこちらのほうが利用頻度が高いと思います。ウェブサイト掲載中の「投稿論文(事例報告)の書き方」の論

文例では、シーズン間の体脂肪率を二元配置分散分析で分析したことが記されていますが、この方法によっています。そこで以下では、繰り返しのない二元配置分散分析というよりは、対応のある一元配置の分散分析の例として説明します。

あなたはレッスンCがAよりも体脂肪率減少の効果が大きいことを知り、次に、レッスンA、B、Cでどの程度の身体負荷がクライアントにかかっているかを定量的に確認したいと思いました。そこで、心拍計を使ってレッスン中の心拍数を調べることにしました。クライアントには、レッスンA、B、Cすべてを体験してもらうことにしました。

表3は、被検者No.1～8までの8名がレッスンA、B、Cすべてを受けて、そのレッスン中の平均的な心拍数データとします。このデータに対して、繰り返しのない二元配置分散分析によって、レッスン間に心拍数の差があるかどうかを検証します。計算上は、繰り返しのない二元配置分散分析と同じになりますが、ここでは、レッスン間(要因I)の差異のみを検証し、被検者間の差異は重要な意味をもちません。また、レッスンと被検者の交互作用もないものと考えます(あったとしても効果の有無を検証できません)。**図4**は**表3**のデータをグラフ化したものです。

Excelでデータ分析を呼び出し、分散分析：繰り返しのない二元配置を選択します。そうすると**図5**のボックスが出現します。**表3**の例では、入力範囲に\$A\$1:\$D\$9を指定します。ラベルをチェックすると、1行目がデータではなくてラベルとして利用され

4 一元配置分散分析において、同じヒトやものから反復してデータを取るようになると、異なるヒトやものからデータを取るよりも誤差が減って水準間の差異を表出させやすくなる。S&C分野では被検者がヒトであることが多いので被検者内要因というが、一般にはヒト以外のこともあるのでブロック要因という。

表3 ワークシート例(繰り返しのない二元配置)

	A	B	C	D
1	被検者	レッスンA	レッスンB	レッスンC
2	No.1	150	158	160
3	No.2	145	150	150
4	No.3	157	155	151
5	No.4	143	144	154
6	No.5	132	133	138
7	No.6	165	163	171
8	No.7	179	182	185
9	No.8	123	131	128

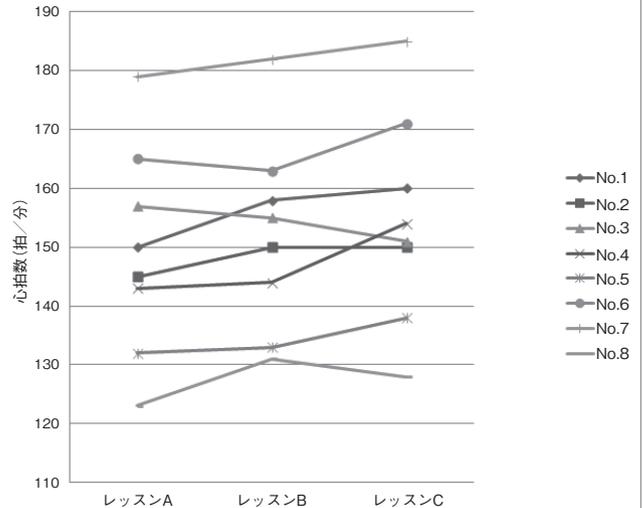


図4 表3データのグラフ

ます。出力先に\$A\$15を指定してOKをします。その結果、表4(A15:G38の領域)が現れます。

表4の分散分析表内の行(セルA34)は被検者間の差異になります。図4のグラフをみて当然ながら、被検者の個人差は非常に大きいので、セルF34のP-値は非常に小さい値になります。しかし「論文を書く」場面ではこのことは重要ではなく、論文内に記述する必要もありません。重要なのは「列(レッスン間の差異)」におけるP-値のセルF35の値です。今、列のP-値も0.020…ですから、5%水準でレッスンという要因に有意な効果を認める結果となりました。すなわち、レッスン間で心拍数は異なる、と判断することができます。

最後に、レッスン間の差異を最小有意差法で事後検定しましょう(前号参照)。

$$A43 : =TINV(0.05,C36)$$

$$D42 : =+D27-D28$$

$$E42 : =D42-A43*SQRT((1/B27+1/B28)*D36)$$

$$F42 : =D42+A43*SQRT((1/B27+1/B28)*D36)$$

E43 ~ F44は、E42, F42と同様

セルA43には、最小有意差法に用いるt値を求めています。このとき

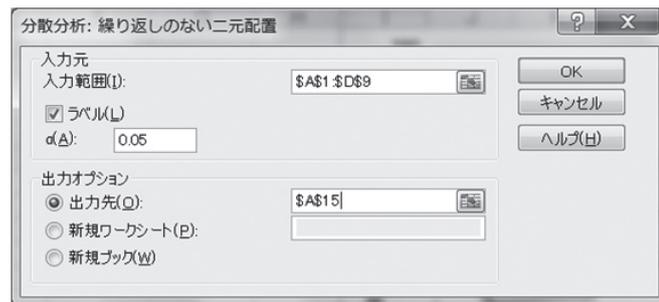


図5 分析ツール(繰り返しのない二元配置)

TINV関数で用いる自由度は、分散分析表内の“誤差要因”の自由度(セルC36)です。さらに平均の差の95%信頼区間を求めるとき(セルE42~F44)には、分散分析表内の“誤差要因”の分散(セルD36)を利用します。

結果的に、レッスンCはレッスンAよりも有意に心拍数が高いことが分かりました。

5. 二元配置分散分析以上

原理的には3つ以上の要因を比較検討する三元、四元配置の分散分析もありえます。しかし、例えば三元配置の分散分析は、要因I、II、IIIそれぞれの主効果に加えて、要因I×II、II×

III、I×IIIの交互作用効果、さらに要因I×II×IIIの交互作用効果の有意性を検証することになります。私たちS&C分野の研究対象としては、この構造は複雑すぎて何を明らかにしているかが分かりにくいです。「論文を書くこと」を念頭に何かしらの実験や測定を実施するならば、設定はなるべくシンプルにし、要因は2つくらいまでが妥当でしょう。したがって、二元配置分散分析までが現実的な使用範囲になります。

また、二元配置分散分析においても、要因の各水準のデータに、対応がある場合と対応がない場合が混在するときもあると思います。S&C分野で

いうと、トレーニング群と非トレーニング群(あるいはトレーニングA群とB群)において、トレーニング前、1ヵ月後、3ヵ月後のように、被検者の変化を追跡していくような実験設定の場合です。この実験設定では、トレーニング群と非トレーニング群で差異が出るのか、差異が出るとしたらいつからか(1ヵ月後? 3ヵ月後?)を明らかにできます。しかしこの実験設定では、トレーニング群と非トレーニング群では被検者が異なり対応がない一方で、トレーニング前、1ヵ月後、3ヵ月後の方は対応があるデータになります。このような構造をもつデータに対して適切に分散分析を実施するには、分散分析表の自由度や分散の計算方法を熟知していなければなりません。残念ながら、Excelのデータ分析ツールや本連載の目的の範疇を超えています。より詳しい本で勉強してもらえればと思います。◆

表4 分散分析(繰り返しのない二元配置)の出力結果

	A	B	C	D	E	F	G	H
14								
15	分散分析: 繰り返しのない二元配置							
16								
17	概要	標本数	合計	平均	分散			
18	No.1	3	468	156	28			
19	No.2	3	445	148.333333	8.33333333			
20	No.3	3	463	154.333333	9.33333333			
21	No.4	3	441	147	37			
22	No.5	3	403	134.333333	10.3333333			
23	No.6	3	499	166.333333	17.3333333			
24	No.7	3	546	182	9			
25	No.8	3	382	127.333333	16.3333333			
26								
27	レッスンA	8	1194	149.25	319.642857			
28	レッスンB	8	1216	152	276.571429			
29	レッスンC	8	1237	154.625	319.982143			
30								
31								
32	分散分析表							
33	変動要因	変動	自由度	分散	観測された分散比	P-値	F 境界値	
34	行	6257.625	7	893.946429	80.3547352	3.19452E-10	2.76419926	*
35	列	115.583333	2	57.7916667	5.19475655	0.02053404	3.73889183	*
36	誤差	155.75	14	11.125				
37								
38	合計	6528.95833	23					
39								
40					95%信頼区間			
41	最小有意差法			平均の差	下限	上限		
42	t(0.05)		A,B間	-2.75	-6.3268779	0.82687792		
43	2.1447867		B,C間	-2.625	-6.2018779	0.95187792		
44			A,C間	-5.375	-8.9518779	-1.7981221	*	
45								

CSCS/NSCA-CPT認定者の皆様へ

CEU インフォメーション
～ Vol.63 ～

- 2014年は今期CEU報告期間の最終年です
CEU報告の手続きは、「マイページ」から簡単に行なえるようになりました！
必要CEU数を取得された方は、お早めに手続きを行ないましょう。



CEU報告 最終期限 2014年12月31日