

8. 分散分析

分散分析(analysis of variance; ANOVA)とは、簡単にいえば平均からのズレ(分散)を分析する手法である。いくつかの群の平均値の差が、意味のあるものであるかどうかを見たい場合に用いる。具体的には、群間のズレ(標本集団間の違い)と群内のズレ(誤差や個人差)を比較し、群間のズレが大きければ、標本集団間の違いが大きいと判断する。群間比較の場合、2群間の比較であれば t 検定を用いることができるが、3群以上の比較の場合は、分散分析を用いることになる。SASで分散分析をおこなうためには、glm プロシージャを用いる。

注)たとえば A,B,C の3群間の比較をおこなう場合、AとB、BとC、AとCと総当たりで t 検定をおこなえばよいのではないかと思うかもしれない。しかしこうした手法で多重に検定をおこなうことは、第1種の過誤(検定が誤って有意となる誤り)の確率を増加させるので、決してやってはいけない事柄である。

8.1 1要因分散分析の概要

glmとは、一般線形モデル(general linear model)の略で、分散分析や回帰分析など、さまざまな多変量解析を包括する分析モデルを総称したものである。「glmのすべて」を語るには途方もない時間と労力が必要なので、ここではとりあえずそのうち「1要因分散分析」の方法について学習する

分散分析をおこなうプロシージャには anova もある。ただし、このプロシージャは要因の各水準に割り付けられたデータの数が等しくないと実行できない(つまりそれぞれの群に含まれているサンプルの数が等しくないとだめ)ので、glm を使う方が一般的である

8.2 被験者間(対応なし)要因の場合

タッチタイピングにおけるトレーニング効果の検証に関して、次のような(仮想)実験をおこなった

	トレーニング方法(method)		
	なし	Aメソッド	Bメソッド
テストの 得点 (point)	55	60	75
	60	73	68
	49	68	86
	45	58	65
	53	59	82
	62	63	79
	71	58	73
	38	64	69

左表は、タッチタイピングに関して

- 1) 特にトレーニングをおこなわなかったグループ(統制群)
- 2) 5日間Aメソッドでトレーニングしたグループ(条件a群)
- 3) 5日間Bメソッドでトレーニングしたグループ(条件b群)

の、5日目のタイピングテスト得点(point)を示したものである。各グループに8名が割り付けられたので、被験者は合計で24名となる

さて、2種類のトレーニングメソッドは効果があったのだろうか。また、AとBではどちらのメソッドがより効果的だったのだろうか。これを検証するためには、被験者間1要因の分散分析をおこなう必要がある。まず、講義Wikiページから、上記実験結果のデータセットをダウンロードしよう

glm プロシージャ;被験者間1要因分散分析の基本形

```
proc glm;
  class method;
  model point=method/ss3;
  means method/tukey;
run;
quit;
```

`class` ステートメントで、タッチタイピングの得点(従属変数;point)を説明する要因(独立変数;この場合トレーニングメソッド(method))を指定する

`model` ステートメントで、従属変数と要因、分析の際に用いる平方和(この場合タイプ3;大体いつもこれ)を指定する

`means` ステートメントで、3つのメソッドごとの得点の平均値を求め、その差を比較(多重比較)する `tukey` オプションは検定方法を指定している

glm プロシージャを実行する場合、終了のサインとして `run;`に加えて `quit;`と記述する必要がある

output の見方

glm プロシージャでは数多くの出力がなされるが、ここではもっとも重要な要因が従属変数に与える効果(主効果)の有意性に関する検定結果
 要因のもつ複数の水準ごとの平均値の差を比較(多重比較)した結果は、どこを見ればわかるのかを説明する

主効果の検定

Dependent Variable: point

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	2	1693.000000	846.500000	13.39	0.0002
Error	21	1327.625000	63.220238		
Corrected Total	23	3020.625000			

R-Square	Coeff Var	Root MSE	point Mean
0.560480	12.44793	7.951116	63.87500

Source	DF	Type III SS	Mean Square	F Value	Pr > F
method	2	1693.000000	846.500000	13.39	0.0002

Pr>F に示されているのが、トレーニングメソッドの効果に関する有意性の検定結果(すなわち有意確率)である。これは、左隣のF値(F Value)=13.39を自由度(2,21)のF分布にもとづいて検討したもので、高度に有意、すなわち「テスト得点に対してトレーニングメソッドの効果が見られた」ことを示している

2群の比較であればこれで終わりだが、今回は3群の比較なので、具体的にはどこにどのような「効果」が見られたのか、トレーニングメソッド間で得点に差はあるのか、というところまではこの結果からは分からない。そこで、次のステップとして「多重比較」をおこなう必要がある。outputの続きを見てみよう

多重比較

Tukey's Studentized Range (HSD) Test for point (point に関する Tukey の HSD 検定)

Means with the same letter are not significantly different. (同じアルファベット間は無差なし)

	Mean	N	method	
A	74.625	8	2	...メソッド B でトレーニングを受けた群の平均得点
B	62.875	8	1	...メソッド A で "
B				
B	54.125	8	0	...トレーニングを受けなかった群の平均得点

「トレーニングなし群」と「メソッド A でトレーニングを受けた群」には、同じアルファベット B がマーキングされていることがわかる。つまりこの2群間には「有意差がない」のである。それに対して「メソッド B 群」は A が付されているので、他の2群とは「有意差がある」ことが、そして得点が高いことがわかる。つまり、この実験の結果から言えることは、「(トレーニングなし群と得点の差がない)トレーニングメソッド A は効果があるとは言えないが、(トレーニングなし群と得点の差がある)メソッド B はトレーニング効果をもっている」ということである

do ループを用いたデータセットの読み込み方

この例では、これまでにおこなってきたように、実験参加者1名=1オブザベーション=1行として入力したデータセットを用いた。一方、8-1 ページにまとめたような表そのままの形のデータセットを用いることもできる。こちらの方が、実験計画を素直に反映した、直感的なデータのかたちとも言える

```
do method=1 to 3;
  input point @@; output;
end;

cards;
55 60 75
60 73 68
49 68 86
45 58 65
53 59 82
62 63 79
71 58 73
38 64 69
;
```

- do 文では「ある要因(この場合トレーニング方法; method)がいくつの水準から構成されているか(この場合3)」を指定する
- input 文では、要因のそれぞれの水準で測定されたデータ(この場合は得点; point)を指定する。
- @@; output; はとりあえず「呪文」だから必ずこう書くのだと覚えておこう
- do 文は必ず最後に end;を持ってくるルールになっている
- このように書けば、各オブザベーションの最初のデータは method=1 の時の point 変数の値、2 番目は method=2 の時、3 番目は method=3 の時と順番に読み込んでくれる。1 行のデータが終了したら次の行に移って同様に読み込んでくれる
- 外部データファイルを読み込む場合は、do ループの直前に infile ステートメントを置けばよい

なお、データの形が異なっても、被験者間要因の分散分析の場合は、glm プロシージャの書き方は同じである

8.3 被験者内(対応のある)要因の場合

先ほどの(仮想)実験で、メソッド B が効果的であることがわかったので、次に新たな実験をおこなった

	テスト実施時点(time)		
	トレーニング前(first)	3ヶ月後(second)	半年後(third)
1	30	52	73
2	32	48	75
3	25	50	73
4	28	49	76
5	31	55	80
6	35	58	72
7	30	50	80
8	27	47	78

1~8 までの 8 名の被験者に、メソッド B を用いたトレーニングを半年間おこなった。左表は、トレーニング前(first)、トレーニング開始 3 ヶ月後(second)、そして半年後(third)のタイピングテストの得点(point)を示している

トレーニングによって得点は伸びているだろうか。つまり、長期間のメソッド B によるトレーニングは、効果をもっていたらうか

先ほどとは異なり、要因の水準ごとのデータは「8 名について各々 3 回」取られたものである。平均値の差の検定で言えば「対応のある」ものと同様で、分散分析の場合はこのような要因のことを「被験者内要因」と呼ぶ

被験者内要因を用いた分散分析の場合、想定した要因に加えて「被験者要因」を組み込む必要がある。トレーニングの効果をなるべく正確に取り出すために、被験者の個人差の影響を考慮する必要があるからである。

また、被験者内要因の場合は、1 人あたり 3 回測定がおこなわれているため、実験参加者 1 人を 1 オブザベーションとみなし、個人 = 1 行でデータを記述するためには、各テスト時点での得点に別々の変数名をつける必要がある。これに対して、do ループを用いた記述方法であれば、同じ point という変数に 3 回のテスト得点を割り当てることができる。どちらでも分析は可能だが、プログラムが異なる。ここでは後者のみを説明する。

```

data glm_within;
do sub=1 to 8;
  do time=1 to 3;
    input point @@; output;
  end;
end;
cards;
30 52 73
32 48 75
25 50 73
28 49 76
31 55 80
35 58 72
30 50 80
27 47 78
;
run;

```

先ほどよりも do ループが少し入り組んだ構造になっていることがわかるだろう。これは、card で読み込むデータに、1 行ずつ同じ被験者(sub)のデータが入っていることを示すためのプログラムである。こう書かないと、先ほどのように 1 つずつ別の被験者のデータとして認識されてしまう。

こう書けば、第 1 行第 1 列のデータは sub=1, time=1 の得点(point), 同じく第 2 列は sub=1, time=2, 第 3 列は sub=1, time=3 の得点として読み込まれる

なお, do ループを用いる場合, end; は do ループの数だけ必要である。ここでは 2 つ必要である

glm プロシージャ;被験者内 1 要因分散分析の基本形

```

proc glm;
class sub time;
model point=sub time/ss3;
means time/tukey;
run;

```

class ステートメントで、タッチタイピングの得点(従属変数;point)を説明する要因(独立変数)を指定する。この場合はテスト実施時点(time)と被験者要因(sub)である
model ステートメントで、従属変数と要因を指定する
means ステートメントで、テスト実施時点ごとの得点の平均値を求め、その差を比較(多重比較)する。tukey オプションは検定方法を指定している

output の見方

出力パターンは被験者間要因の場合と同じだが、被験者要因が増えていることに注意しよう

主効果の検定

Source	DF	Type III SS	Mean Square	F Value	Pr > F
sub	7	94.500000	13.500000	1.34	0.3018 ...被験者要因
time	2	8525.250000	4262.625000	423.99	<.0001

この結果から、テスト実施時点の効果は有意だが、被験者要因は有意ではない(すなわち個人差はない)ことがわかる。では、テスト実施時点のどことどの間に有意差があるのか、詳しく見るために多重比較の結果を見てみよう

多重比較

	Mean	N	time
A	75.875	8	3
B	51.125	8	2
C	29.750	8	1

先ほどと同じく、異なるアルファベットがマーキングされている群に有意差があると判断できる。つまりこの場合は、「トレーニング開始前」「3 ヶ月後」「半年後」のいずれの間にも有意差があるといえる。実験の結果、トレーニングメソッド B の効果があがっていることが示されたといえる

8.4 被験者間2要因分散分析(完全無作為要因計画)

分散分析は、2つ以上の平均値の差を検定する場合に用いられる分析手法である。従属変数に対して影響をもつと考えられる要因が2つ以上ある場合にも同じように分散分析を適用することができる。2要因分散分析には、それぞれの要因が被験者間であるか被験者内であるかによって3つのパターン(要因計画)があり、それぞれ

- 1) 被験者間2要因(完全無作為要因計画)
- 2) 被験者内2要因(乱塊法計画)
- 3) 被験者内×被験者間2要因(分割法計画)

と呼ぶ。ここでは1)についてのみ解説する。

ある従属変数に対して影響を及ぼしているかどうかを調べたい要因がAとBの2つあり、これらの要因に実験参加者を完全にランダムに割り付ける計画を「完全無作為要因計画(Completely Randomized Factorial Design)」という。この場合、実験参加者は、いくつかある実験状況のうち、自分が割り付けられた1つの下でしか課題を遂行しない。よって要因はそれぞれ 対応がない/被験者間要因 であるといえる。たとえば、タッチタイピングのトレーニングをおこなう際に、男女5名ずつを、A~Dという4つのトレーニングメソッドに無作為に割り当て、事後のタイピングテストに対する性別とトレーニングメソッドの効果を検証する、といった場合がこれにあたる。調査などで事後的に回答者を分割する場合も、この要因計画に準じた分析をおこなうことが多い。

2要因分散分析では、以下のような検討をおこなう。1要因分散分析と異なるのは、要因AとBの組み合わせによって生じる(裏返して言えば、どちらか一方の要因を操作しただけでは出現しえないような)効果、すなわち 交互作用効果 の検討が加わる点である。

1) 主効果 の検討

有意な性別 and/or トレーニングメソッドの主効果あり: 男女間 and/or トレーニングメソッド間の多重比較

2) 交互作用 の検討

有意な性別×トレーニングメソッドの交互作用効果あり: 単純主効果 の検定と多重比較

男女別のトレーニングメソッド効果の検討と多重比較

(or トレーニングメソッド別の性別の効果の検討と多重比較...この場合はあまり一般的ではない?)

それでは、以下のようなデータを使って、SASによる分析をおこなってみよう。まず「主効果と交互作用の検討」と「単純主効果の検定」をおこなうプログラムを紹介する。基本形は、1要因分散分析と同じである。

		性別 (sex; 2水準)							
		男性				女性			
		トレーニングメソッド (method; 4水準)							
		A	B	C	D	A	B	C	D
テストの 得点 (point)		51	81	55	65	49	52	52	81
		45	78	65	53	59	73	62	73
		52	73	73	58	64	48	49	68
		64	83	45	63	58	65	62	80
		46	77	35	51	60	60	58	67

*各条件に男女5名ずつを割り当て、合計で $4 \times 5 \times 2 = 40$ 名のデータである

```

data crf24;
  do sex=1 to 2;
    do method=1 to 4;
      input point @@; output;
    end; end;

cards;
51 81 55 65 49 52 52 81
45 78 65 53 59 73 62 73
52 73 73 58 64 48 49 68
64 83 45 63 58 65 62 80
46 77 35 51 60 60 58 67
;
proc glm;
class sex method;
model point=sex method sex*method /ss3; /*検定する効果を並べ, 平方和を指定*/
means sex method /tukey; /*主効果の多重比較*/
means sex*method; /*各条件での平均値の出力*/
lsmeans sex*method /slice=sex; /*性別ごとのメソッド単純主効果の検定*/
lsmeans sex*method /slice=method; /*メソッドごとの性別単純主効果の検定*/

run;
quit;

```

```

data crf24;
  input sub sex method point;

cards;
1 1 1 51
2 1 2 81
...
5 2 1 49
6 2 2 52
...
49 2 3 58
50 2 4 67
;

```

上記の例では do ループを用いてデータを読み込んでいるが, 1 行に実験参加者 1 名のデータを記述する下記のスタイルであっても, プロシージャステップの記述は変わらない(つまり, 被験者間要因のみを扱う限り, データセットの並びがプログラムに影響を及ぼすことはない). この場合は左のようなデータになる.

output の見方

主効果と交互作用の検定

Dependent Variable: point

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	7	3414.975000	487.853571	6.07	0.0001
Error	32	2570.400000	80.325000		
Corrected Total	39	5985.375000			

R-Square	Coeff Var	Root MSE	point Mean
0.570553	14.84459	8.962422	60.37500

Source	DF	Type III SS	Mean Square	F Value	Pr > F	
sex	1	18.225000	18.225000	0.27	<u>0.6096</u>	非有意
method	3	1551.875000	517.291667	7.55	<u>0.0006</u>	有意!!
sex*method	3	1601.875000	533.958333	7.79	<u>0.0005</u>	有意!!

四角で囲んだ数値が、性別とトレーニングメソッドの主効果、そして交互作用に関する有意性の検定結果である。この結果から、性別の主効果は見られなかったが、トレーニングメソッドの主効果、および性別×トレーニングメソッドの交互作用が有意であることがわかるトレーニングメソッドによって、タッチタイピングスキルの向上には違いがあり、またその違いは実験参加者が男性であるか女性であるかによって異なる可能性が示されたわけだ

さて、ここまでの出力結果から、次に進むべき道が、以下の3つの選択肢のうちのいずれか1つに決まる

- 1) 交互作用が見られた場合: 単純主効果の検定と多重比較へ(主効果の多重比較はおこなわない)
- 2) 主効果が見られ、交互作用が見られなかった場合: 主効果の多重比較へ
- 3) 主効果も交互作用も見られなかった場合: 残念!!(ここで終了)

ここでは1)の道を選ぶことになるが、その前に2)の場合を説明しておこう

主効果のいずれかが有意で、交互作用が有意でなかった場合は「means sex method /tukey;」により実行される主効果の多重比較の出力結果を参照する。このデータの場合はこのように出力されている

主効果の多重比較

Tukey's Studentized Range (HSD) Test for point (point に関する Tukey の HSD 検定)

Means with the same letter are not significantly different. (同じアルファベット間は無差なし)

性別の主効果の多重比較

	Mean	N	sex	
A	62.000	20	2	...女性の平均得点
A	60.650	20	1	...男性の平均得点

主効果が有意でなかったことから分かるように、性別については平均得点に差はない

トレーニングメソッドの主効果の多重比較

	Mean	N	method
A	69.000	10	2 ...メソッド B の平均得点
A			
A	65.900	10	4 ...メソッド D の平均得点
B			
B	55.600	10	3 ...メソッド C の平均得点
B			
B	54.800	10	1 ...メソッド A の平均得点

同じでない符号が付されている条件間に有意差があるので、メソッド B, D とメソッド C, A の間にテスト得点の有意差が見られる (B と D および A と C の間には有意差がない) ことがわかる。どうやらメソッド B と D が効果的なトレーニングであるようだ。このように、メソッドの主効果のみが見られ、交互作用が見られなかった場合は、この多重比較結果を記述して分析終了となる。

では 1) に戻ることにして。性別 × トレーニングメソッドの交互作用が見られたということは、男女によってトレーニングメソッドの効果が異なるらしいということが予想される。具体的にはどのように「異なる」のだろうか。「男性の場合はトレーニングメソッドの違いによる効果が見られないが、女性では顕著に見られる」のかもしれないし、「男性の場合はメソッド A や C が効果的だが、女性の場合は B が効果的」なのかもしれない。これを読み解くためにおこなうのが、単純主効果の検定 (そしてその後におこなう多重比較) である。

交互作用の単純主効果の検定

「lsmeans sex*method /slice=sex;」により実行されるのが、性別ごとのメソッドの単純主効果の検定であり、出力は、次のようになる。

Least Squares Means

sex*method Effect Sliced by sex for point ...性別ごとのメソッドの単純主効果の検定

sex	DF	Sum of Squares	Mean Square	F Value	Pr > F
1	3	2202.950000	734.316667	10.72	<.0001
2	3	950.800000	316.933333	4.63	0.0085

この結果から、男性の場合も女性の場合も、メソッドの主効果が有意であることがわかる。もちろんここでどちらかの主効果が非有意となることもあり、その場合は「男性(女性)の場合はトレーニングメソッドの違いによる効果が見られないが、女性(女性)では顕著に見られる」ということになる。いずれにせよ、この交互作用がどのような意味を持っているのかを検討するために、次に単純主効果の多重比較に進もう。

単純主効果の多重比較

実は SAS には、この単純主効果の多重比較を正確におこなうためのオプションが用意されていない。検出力の高い(つまり信頼できる)分析をおこなうためには、後述する手順に従って手計算をする必要があるのだが、ここではまずそれに準じた(やや検出力が劣るがそれなりに見当をつけられる)方法を紹介する。

```
proc sort; by sex;
proc glm; by sex;
class method;
model point=method /ss3;
means method/ tukey;
```

ここでは、性別ごとにトレーニングメソッドを独立変数とした 1 要因分散分析をおこなう。そのためには、最初に proc sort で、性別でデータを並び替えることが必要である。

その後の GLM プロシジャは、by sex という並び替えコマンドが入っている以外は、1 要因分散分析で学習したのと同じである。

この分析結果は、次のよう出力される。

sex=1 ...男性データ

Tukey's Studentized Range (HSD) Test for point
Means with the same letter are not significantly different.

	Mean	N	method
A	78.400	5	2
B	58.000	5	4
B	54.600	5	3
B	51.600	5	1

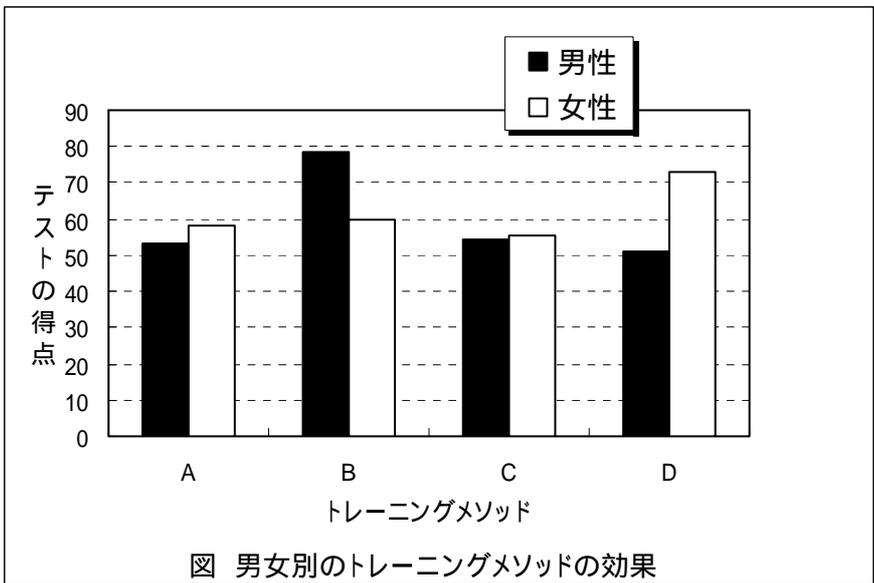
男性にもっとも効果的なのは
メソッド B,他のメソッドは Bと
比較するとあまり効果がなく,
メソッド間の有意差なし

sex=2 ...女性データ

	Mean	N	method
A	73.800	5	4
B	59.600	5	2
B	58.000	5	1
B	56.600	5	3

女性にもっとも効果的なのは
メソッド D,他のメソッドは Dと
比較するとあまり効果がなく,
メソッド間の有意差なし

これらの結果を記述する際は,グラフを併用するとより分かりやすくなる.その例が次の図である.どのような
グラフを作成すればよりの確で分かりやすいかについては,各自で工夫してみよう.



なお,もっと検出力の高い単純主効果の多重比較は,HSD 検定を用いておこなう.しかし SAS では直接の
出力が得られないので,出力で得られた数値をもとにして手計算することになる.この手法については,森・吉
田「心理学のためのデータ解析テクニカルブック」(北大路書房)などを参照のこと