

## 【資 料】

## GLM (General Linear Model) によるパラメトリック的統計解析の 統一的理解 (1)

高梨一彦

### Parametorical understanding through GLM (1)

Kazuhiko TAKANASHI

#### 要旨

パラメトリックな多変量の統計解析において用いられている(重)回帰分析や分散分析法等は、基本的な解析方法であると同時にコンピュータの利用が必須である。これらの方法は、従来、それぞれ別個の手法と考えられて教育上もそのように扱われてきている。しかしながら近年は理論的な枠組みも統一のかつ一般的になってきて、これらの方法をより一般性の高いものから考察しようという動きがある。それがGLM (General Linear Model; 一般線型モデル) である。本研究では、GLMの理論的な枠組みを外国文献ならびに邦文文献によって、統計学上の理論をまとめ、統計パッケージソフトによる分析手順の違いを示した。

**キーワード:** 分散分析 (ANOVA)、回帰分析 (Regression analysis)、一般線型モデル (General Linear Model)

心理学系の一般的な統計の分析方法について、二つの平均値の比較は t 検定、多数の平均値の比較は分散分析 (場合によっては実験計画法と対で)、要因同士に相関がある場合には共分散分析、また重回帰分析などが使われてきている。この分野において現状ではいわゆる一般線型モデル (General Linear Model; GLM) あるいは線形混交モデルなどが使われている<sup>1)2)3)</sup>。この流れは、例えばSPSSにおけるprocedureの種類が分散分析を行うためのANOVA手続きから、GLM手続きに変わっていくという歴史的な流れにも現れており、それがどのように変遷していったのか、そのことで具体的に何が変わっていったのかについて、明らかにしたいとの思いを持ってきた。

筆者は心理統計法関連の科目を教授する際に統計パッケージソフト (SPSSあるいはR言語) を利用することがほとんどであるが、その際に統計分析のひな形を作成して結果を確認しながら教育を行っている。学部時代に統計的分析方法に関して習ってきたことやその後に学んだことでは、少なくとも1990年代までは、前述したような分散分析・共分散分析や(重)回帰分析などといった手法は、統計書籍には別の章立てでまとまって記載されており、そのことはすなわち、それぞれ別個の手法として扱われてきていた<sup>3)4)5)6)7)8)9)</sup>。このことがいつの頃からか、理論的には同じものであること、そしてそのことによって手法としてはGLMのバリエーションに過ぎないということが認められてきたため、SPSS(-X)などでの分析

手続きの変化へとつながってきている。

### (1) 分散分析 (ANOVA ; Analysis of Variance) について

例えば、固定モデル2要因の分散分析は、以下のような数式で表される。<sup>10)</sup>

$$E(Y_{jkl}) = \mu + \alpha_j + \beta_k + (\alpha\beta)_{jk}$$

$E$  は期待値、 $\mu$  は母平均、 $\alpha$ 、 $\beta$  は定数項、 $Y$  は従属変数である。また  $\alpha\beta$  は交互作用である。二つの要因 ( $\alpha$  と  $\beta$  をもたらす) によって母平均の期待値にどのような変化をもたらされるかに関してそれらの要因の持つ分散を偶然による分散 (誤差項: 要因以外のもの) との比較によって意味があるかどうかを確かめるものである。ここでは、誤差項という名称は使われているが、あくまで扱われる分散は、母平均とA要因の効果 ( $\alpha$ )、B要因の効果 ( $\beta$ )、それにA要因とB要因の交互作用 ( $\alpha\beta$ ) のみであり、それ以外の部分については、明記していない。実際は残りの部分を誤差項として扱って、分散を求めて比較を行っている。

### (2) (重) 回帰分析 ((Multiple) Regression Analysis) について

一番単純な回帰分析のモデルは次のように表される。

$$Y = aX + b$$

そして、変数が二つの場合の重回帰分析のモデルは次のように表される。

$$Y = a_1X_1 + a_2X_2 + b_0$$

$Y$  は予測したい変数、 $a$  は係数、 $X$  は独立変数、 $b$  は切片である。 $Y$  と予測された  $Y'$  の値の違いを最小にするように  $a$  の係数を求めていくことになる。

### (3) 一般線型モデル (GLM ; General Linear Model) について

モデルをできるだけ簡単に表記すれば以下のようなになる (Rutherford, 2001)。<sup>11)</sup>

$$\text{data} = \text{model} + \text{error}$$

すなわち、一般的な数式では以下のように表される。

$$Y = XB + E$$

$Y$  は予測したい変数、 $X$  は独立変数、 $B$  は重み行列、 $E$  は誤差である。この誤差の部分がGLMの特徴となる。

ここで分散分析との関係についてはRutherford (2001) によれば、

モデルに誤差を加えることによって、分散分析についても考えることができる。ここでは従属変数の値はデータを、実験の条件はモデルを構成し、またモデルで調整できなかったデータ成分は誤差として表される。一般的に分散分析を用いている研究者は、実験条件で得られた従属変数の平均値の有意差がどのくらいあるのかということに興味を持っている。実験条件間での違いに関して、従属変数

がどのくらい変化したのかを決定することによって分散分析は成し遂げられ、そして、それぞれの実験条件のうちで起こる従属変数の変化と誤差の項を比較する。つまり、従属変数の変化の割合が実験変数を操作したことによる部分に興味を持っている。分散分析における従属変数は、量的な尺度で測定されることがもっとも好ましいが、量的な条件下での被験者群と元々がカテゴリカルな被験者群との間の比較の場合でさえも統計的な分析は行われる。それゆえに分散分析は、カテゴリカルな予測を量的な予測として用いる回帰分析の特別なタイプである。

と述べられており、ANOVAはGLMの一つの亜型であると考えてよい。

一方、重回帰分析についてRutherford (2001) は同様に

回帰分析は、独立変数あるいは予測変数 (モデル) そして残差 (誤差) によってデータ (従属変数の値) を説明することの試みである。一般的に回帰分析を用いる研究者は一つかそれ以上の量的な独立変数から一つの量的な従属変数を予測することや、予測にそれぞれの独立変数を割り当ててその相対的な寄与を決定することに関心がある。すなわち、従属変数の分散のどの程度の割合が (複数の) 独立変数の分散によるものであるのかを求めることに関心があるのである。回帰分析はカテゴリカルな (名義的あるいは質的なものとして知られている) 予測変数を用いることも可能である。すなわち、それらは性別、婚姻状況、教授方法の型などの独立変数を一般的に用いることである。

と述べており、(重) 回帰分析もGLMの一つの亜型であると考えてよい。

#### (4) 統計パッケージソフト (SPSSならびにR言語) による分散分析について

次にこれらの解析的な理解の下に二つのパッケージソフトでどのように分析手順が行われているのかについてGLMとの関連で述べる。用いるデータについては、佐々木ほか (1997)<sup>12)</sup>の分散分析の例を利用する。

##### 1) SPSSの場合

SPSSは、以前のバージョンであるSPSS-X<sup>8)13)14)15)</sup>の時代では、コマンドレベルでの分散分析の場合、ONEWAY、ANOVA (ただしSPSS-Xの時代まで)、GLM (SPSS-XからSPSSに変わってから)、UNIANOVA、MANOVAという分析手順を用いている。

##### a) 一要因分散分析について

4種類 (A ~ D) の教授方法の違いが独立変数であり、従属変数はテストの得点である (データの読み込み部分は省く)。この場合の数的な仮定は以下のようになる。

$$E(Y_{jk}) = \mu + \alpha_j$$

表1 4つの教授方法とその結果

名簿	教授方法			
	A	B	C	D
1	45	56	83	100
2	57	68	78	90
3	32	82	62	95
4	48	74	87	88
5	62	62	74	92
平均	48.8	68.4	76.8	93
SD	10.38	9.07	8.61	4.20

i) SPSS-Xまでの手順と結果 (SPSS/PC+<sup>6)</sup> による分析)

コマンド部分は次のようになる。一要因分散分析を行い、Scheffe法で多重比較を行う場合である。

```
-----
ONEWAY VARIABLES=SCORE BY GROUP(1,4)
/RANGES=SCHEFFE.
```

結果は次の通り (多重比較の結果は省く)。

```
-----
      - - - - - O N E W A Y - - - - -
Variable SCORE      テスト得点
By Variable GROUP   グループ

      Analysis of Variance
      Sum of          Mean
      Squares        Squares          F      F
      Source          D. F.          Ratio Prob.
Between Groups        3      5074.9500    1691.6500    19.2124 .0000
Within Groups        16      1408.8000     88.0500
Total                 19      6483.7500
-----
```

これは後述のGLM手続きではなく、ONEWAY手続きによって分析を行っている。分析の結果、級間要因 (Between Groups ; 教授方法) と級内要因 (Within Groups ; 個人差) に分散を分けて、級内要因の平均平方和で級間要因 (教授方法) を割った比率を求めるとその主効果は有意である ( $F=19.2124$ ,  $df=3, 16$ ,  $p<0.001$ )。

ii) SPSS-XからSPSSに変わってからの手順と結果 (SPSS 16.0以降の結果)

求める結果は同様であるが、ここからはGLM手続きの一要因を指定する (UNIANOVA)。この場合の数的な仮定は以下のようになる。

$$E(Y_{ij}) = \mu + \alpha_j + \epsilon_{ij}$$

$\epsilon_{ij}$  は誤差である。

```
-----
UNIANOVA SCORE BY GROUP
/METHOD=SSTYPE(3)
```

```

/INTERCEPT=INCLUDE
/POSTHOC=GROUP (SCHEFFE)
/CRITERIA=ALPHA (0.05)
/DESIGN=GROUP.

```

結果は以下の通り。

表2 被験者間効果の検定

従属変数:テスト得点

ソース	タイプ III 平方和	自由度	平均平方	F 値	有意確率
修正モデル	5074.950 <sup>a</sup>	3	1691.650	19.212	.000
切片	102961.250	1	102961.250	1169.350	.000
GROUP	5074.950	3	1691.650	19.212	.000
誤差	1408.800	16	88.050		
総和	109445.000	20			
修正総和	6483.750	19			

a. R2乗 = .783 (調整済みR2乗 = .742)

「切片」が示されており、GLMを用いた結果になっているのが分かる。前述の結果の級内要因（個人差）がここでは「誤差」という形で示されている。タイプIIIの平方和<sup>16)</sup>を用いた分析の結果、GROUP要因（教授方法）が有意であり（ $F=19.212$ ,  $df=3, 16$ ,  $p<0.001$ ）、前述の級内要因（個人差）は「誤差」として分けられている。当然ながら結果の数値は全く同一である。

#### b) 二要因分散分析の場合

男女(2) × クラス(4)の二要因で繰り返しなしの固定モデルである。従属変数はテスト得点になる（表3参照）。二要因分散分析を行う。

表3 男女20名ずつ20名に実施した学級別のテスト結果

		学級			
		A	B	C	D
男	1	66	52	75	88
	2	55	70	85	70
	3	80	48	62	58
	4	75	57	87	80
	5	79	63	70	76
女	1	74	83	56	55
	2	80	78	68	57
	3	96	62	54	40
	4	88	87	74	50
	5	77	74	62	60
平均		77.0	67.4	69.3	63.4

i) SPSS-Xまでの手順と結果 (SPSS/PC+<sup>6)</sup>による分析)

-----  
ANOVA VARIABLES=SCORE BY CLASS(1, 4) GENDER(1, 2).  
-----

結果は次の通り。

-----

* * * A N A L Y S I S   O F   V A R I A N C E * * *					
	SCORE	テスト得点			
BY	CLASS	クラス			
	GENDER	性別			
Source of Variation	Sum of	DF	Mean		Signif
	Squares		Square	F	of F
Main Effects	988.100	4	247.025	2.723	.047
CLASS	977.075	3	325.692	3.590	.024
GENDER	11.025	1	11.025	.122	.730
2-way Interactions	2865.075	3	955.025	10.528	.000
CLASS GENDER	2865.075	3	955.025	10.528	.000
Explained	3853.175	7	550.454	6.068	.000
Residual	2902.800	32	90.713		
Total	6755.975	39	173.230		

-----

この結果は、従来の分散分析表である。ANOVA手続きによって二要因分散分析を行った結果、学級 (CLASS) は有意 (F=325.692, df=3, 32, p<0.05)、性別 (GENDER) の主効果は非有意 (F=11.025, df=1, 32, n.s.)、学級×性別の交互作用は有意 (F=955.025, df=3, 32, p<0.001) であった。

ii) SPSS-XからSPSSに変わってからの手順と結果 (SPSS 16.0の結果)

ここからは、GLM手続きで変量が一つである場合のANOVAの指定をしている。また手順でデザインや切片の指定があり、GLMによる分析になっている。この場合の数的な仮定は以下ようになる。

$$E(Y_{jkl}) = \mu + \alpha_j + \beta_k + (\alpha\beta)_{jk} + \epsilon_{jkl}$$

$\epsilon_{jkl}$  は誤差である。

-----  
UNIANOVA SCORE BY GENDER CLASS  
  /METHOD=SSTYPE(3)  
  /INTERCEPT=INCLUDE  
  /CRITERIA=ALPHA(0.05)  
  /DESIGN=GENDER CLASS GENDER\*CLASS.  
-----

結果は次の通り。

表4 分散分析表a, b

	一意的な方法				
	平方和	自由度	平均平方	F 値	有意確率
テスト得点 主効果 (結合された)	988.100	4	247.025	2.723	.047
クラス	977.075	3	325.692	3.590	.024
性別	11.025	1	11.025	.122	.730
2次交互作用 クラス * 性別	2865.075	3	955.025	10.528	.000
モデル	3853.175	7	550.454	6.068	.000
残差	2902.800	32	90.713		
合計	6755.975	39	173.230		

a. クラス, 性別によるテスト得点

b. 全ての効果を同時に投入

「主効果」や「2次交互作用」といった項目は分散分析表の表記通りである。さらに「残差」があり、これが誤差項と考えてよい。前述のResidualがそれに該当する。分散分析結果としては、「モデル」は考慮しない。それ以外の結果については前述のものと同一である。

## 2) R言語の場合<sup>17)</sup>

これはprocedureで言えば、oneway、aovあるいはanova、そしてlmといったものであり、当初からこの手続きである。データは全て同一のものである。

### a) 一要因分散分析について

```
# aov で分散分析を行う
summary(aov(formula=score2~group2))
```

```
# anova と lm で分散分析を行う
anova(lm(score2~group2))
```

結果は次の通り。

```
-----
              Df Sum Sq Mean Sq F value    Pr(>F)
group2          3   5075   1691.7    19.21 1.49e-05 ***
Residuals      16   1409     88.1
-----
```

Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

Analysis of Variance Table

Response: score2

```
              Df Sum Sq Mean Sq F value    Pr(>F)
group2          3 5075.0  1691.65    19.212 1.49e-05 ***
Residuals      16 1408.8    88.05
-----
```

Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1



内部的な扱いがすべてGLMになっているためか、結果は同様になる。Residualsとあるため、分散分析の用語ではあるが、手続きはlm (Linear Model) である。当然ながら結果は同一である。

#### b) 二要因分散分析の場合

# aov で分散分析を行う

```
summary(aov(score2 ~ subjects2 * gender2))
```

# anova と lm で分散分析を行う

```
anova(lm(score2 ~ subjects2 + gender2 + gender2 * subjects2))
```

結果は次の通り。

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
subjects2	3	977.1	325.7	3.590	0.0242 *
gender2	1	11.0	11.0	0.122	0.7297
subjects2:gender2	3	2865.1	955.0	10.528	5.67e-05 ***
Residuals	32	2902.8	90.7		

Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

#### Analysis of Variance Table

Response: score2

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
subjects2	3	977.08	325.69	3.5904	0.02416 *
gender2	1	11.03	11.03	0.1215	0.72966
subjects2:gender2	3	2865.07	955.02	10.5280	5.673e-05 ***
Residuals	32	2902.80	90.71		

Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

結果はanovaで示しているが、内部的にlm手順を用いており、モデルの指定は"score2 ~ subjects2 + gender2 + gender2 \* subjects2"となり、前項で挙げた数式と同じになっており、性別とクラス（被験者）それにそれらの交互作用を組み込んだものになっている。もちろん、結果も同一である。

#### (5) まとめと今後について

上記のように固定モデル2要因までの分散分析を扱ってきたが、今後は3要因以上のもの、変量モデルを含むもの（混合モデルを含む）、repeated measurementの場合などを扱っていききたい。その後に（重）回帰分析についても触れていきたいと考えている。



## 謝辞

本研究は、平成20年度和洋女子大学研究奨励費 (GLM (General Linear Model) によるパラメトリック的統計解析の統一的理解—SPSSの実際的な処理プロシージャ作成とその教育的応用—) ならびに平成21年度和洋女子大学研究奨励費 (GLM (General Linear Model) によるパラメトリック的統計解析の統一的理解2—SPSSならびにRによる実際的な処理プロシージャ作成とその教育的応用—)、さらに平成23年度和洋女子大学研究奨励費 (GLM (General Linear Model) によるパラメトリック的統計解析の統一的理解3—SPSSならびにRを用いた数値的な理解とその互換性—) の助成を受けた。

## 参考文献

1. A. Grafen, R. Hails 野間口謙太郎・野間口眞太郎訳, 一般線形モデルによる生物科学のための現代統計学—あなたの実験をどのように解析するか, 共立出版, 2007.
2. Annette J. Dobson 田中豊・森川敏彦・山中竹春・富田誠 (訳), 一般線形モデル入門, 共立出版, 2008.
3. Gustav Levine, A guide to SPSS for analysis of variance. Hillsdale, N. J.: Lawrence Erlbaum Associates, 1991.
4. Howell, D. C., Statistical methods for psychology, 8th edition. Wadsworth Pub Co., 2012.
5. Jose M. Cortina and Hossein Nouri, Effect size for ANOVA designs SAGE Publications, 2000.
6. Keppel, G., Design & analysis, 2nd edition. A researcher's Handbook. London: Prentice-Hall, 1982.
7. SPSS/PC+ V3.0J Base Manual V2.0 SPSS Japan Inc, 1989.
8. SPSS-X user's guide 3rd edition. SPSS inc, 1988.
9. Winer, B. J., Statistical Principles in Experimental Design (Mcgraw-Hill Series in Psychology) 3rd edition. New York: McGraw-Hill, 1991.
10. 櫻井由香里, GLM (一般線形モデル) について 平成24年度和洋女子大学人文学群心理・社会学類人間発達学専修心理発達コース卒業論文, 2012.
11. Andrew Rutherford, Introducing Anova and Ancova (Introducing Statistical Methods series) SAGE Publications, 2001.
12. 佐々木保行・久米弘・高梨一彦・竹内史宗, 改訂版 心理・教育統計法—卒論・修論作成のために—, 高文堂出版, 1997. ISBN4-7707-0569-7
13. 佐々木保行監修, 久米弘・高梨一彦, 実務的SPSSによる多変量解析法, 高文堂出版, 1993, ISBN4-7707-0433-X
14. 三宅一郎・山本嘉一郎・白倉幸男・垂水共之・小野寺孝義, 新版SPSS X III解析編2, 東洋経済新報社, 1991. ISBN4-492-47043-3
15. 垂水共之・西脇二一・石田千代子・小野寺孝義, 新版SPSS X II解析編1, 東洋経済新報社, 1993. ISBN4-492-47042-5
16. “StatsBeginner: 初学者の統計学習ノート.”  
<http://statsbeginner.hatenablog.com/entry/2015/01/04/130108> (参照2016-08-30)
17. 間瀬茂, R基本関数マニュアル. <https://cran.r-project.org/doc/contrib/manuals-jp/Mase-Rstatman.pdf>. (参照2016-08-30)

高梨一彦 (和洋女子大学 人文社会科学系 教授)

(2016年10月11日受理)