

LS-Means 再考

– GLM と PLM によるモデル推定後のプロセス –

○魚住 龍史

京都大学大学院医学研究科 医学統計生物情報学

The current innovation for LS-Means: implementation by using both GLM and PLM procedures

Ryuji Uozumi

Department of Biomedical Statistics and Bioinformatics, Kyoto University Graduate School of Medicine

要旨

SAS では、線形モデルによる解析を行うためのプロシジャが多く実装されている。これらのプロシジャでは、モデルの推定後にパラメータの線形式に対する推定が必要となることがある。例えば、LSMEANS ステートメントによる各群の LS-Means 「最小 2 乗平均」を算出し、比較することが挙げられる。LSMEANS ステートメントや ESTIMATE ステートメント等の機能は、SAS/STAT V9.22 から大幅に拡張され、より多くのプロシジャで実行できるようになった。さらに、上述したモデル推定後のプロセスは、以前のバージョンではプロシジャの実行とともに記述する必要があったが、SAS/STAT V9.22 から、モデル情報をアイテムストアとして保存するための STORE ステートメント、及びアイテムストアを呼び出すための PLM プロシジャが追加された。本稿では、GLM プロシジャを用いた、LSMEANS ステートメントによる LS-Means の算出について、各パラメータに対する係数を ESTIMATE ステートメントで明示的に概説する。次に、オプション機能として、OBSMARGINS オプション、BYLEVEL オプション、AT オプションを用いた LSMEANS ステートメントによる LS-Means の算出について解説する。また、GLM プロシジャによるモデル情報から、PLM プロシジャを用いて、LSMESTIMATE ステートメントによる推定を行う実行手順を示す。

キーワード : LS-Means, GLM, ESTIMATE, OBSMARGINS, BYLEVEL, AT, モデル情報, PLM, LSMESTIMATE

1 はじめに

SAS では、GLM プロシジャや LOGISTIC プロシジャなどのように、線形モデルによる解析を行うためのプロシジャが多く実装されている。これらのプロシジャでは、MODEL ステートメントにおいて推定するモデル式を記述し、SOLUTION オプションによってパラメータ推定値を表示することができる。さらに、データ解析では、モデル推定後のプロセスとして、パラメータの線形式に対する推定が必要となることがある。

例えば、LSMEANS ステートメントによる各群の最小 2 乗平均 (least squares means, 以下、LS-Means) を算出し、比較することが挙げられる。LSMEANS ステートメントは、OBSMARGINS (OM) オプション、BYLEVEL オプション、AT オプションといったオプション機能がサポートされており、大変簡便かつ有用である。また、CONTRAST ステートメントや ESTIMATE ステートメントによる、任意の係数を用いた推定も考えられる。LSMEANS ステートメントの機能は、以前のバージョンにおいては GLM プロシジャなどの一部のプロシジャのみで利用できるものであった。しかし、SAS/STAT V9.22 から大幅に拡張され、表 1 のように、LOGISTIC プロシジャや PHREG プロシジャなどのより多くのプロシジャで実行できるようになった。また、新たに LSMESTIMATE ステートメントが追加されており、LS-Means の線形式に対する推定、検定を行うことができる。LSMESTIMATE ステートメントは、LSMEANS ステートメントと異なり、各 LS-Mean の値は算出されない。さらに、モデルパラメータの線形式を用いる ESTIMATE ステートメントと異なり、LSMESTIMATE ステートメントでは LS-Means の線形式を用いるため、より解釈しやすいステートメントといえる。本ユーザー総会においては、浜田 (2013) によって、PHREG プロシジャによる上記のステートメントの解説が行われ、決まりきった対比較等の不必要な項目まで冗長に出力されてしまう LSMEANS ステートメントに比べて、特定の群間比較のみ出力できる LSMESTIMATE ステートメントが推奨された^[7]。

表 1: 各プロシジャにおけるステートメントの使用可否^[3]

プロシジャ	LSMEANS	CONTRAST	ESTIMATE	LSMESTIMATE
GLM	○	○	○	-
MIXED	○	○	○	☆
GENMOD	☆	○	○	☆
GLIMMIX	○	○	○	☆
LOGISTIC	☆	○	☆	☆
PHREG	☆	○	☆	☆
ORTHOREG	☆	-	☆	☆
PLM	☆	-	☆	☆
SURVEYREG	☆	○	○	☆
SURVEYLOGISTIC	☆	○	☆	☆
SURVEYPHREG	☆	-	☆	☆

○: 以前のバージョンから利用可能, ☆: 新たに追加あるいは更新, -: サポートされていない

その他にも、SAS/STAT V9.22 から拡張された機能として、モデル情報をアイテムストアとして保存するための STORE ステートメント、及びアイテムストアを呼び出すための PLM プロシジャが追加された。上述した LS-Means を算出するモデル推定後のプロセスは、以前のバージョンではプロシジャの実行とともに記述する必要があり、オブザベーション数が多い場合やモデル式が複雑である場合、モデル推定に再度多くの時間を要することがあった。しかし、PLM プロシジャを用いると、推定されたモデル情報を呼び出し、統計量の算出、ODS GRAPHICS によるグラフ表示等を実施でき、再度の入力データセットの参照、線形モデルプロシジャの実行が不要であり、大変有用であるといえる。

本稿では、GLM プロシジャを用いた、LSMEANS ステートメントによる LS-Means の算出について、各パラメータに対する係数を ESTIMATE ステートメントで明示的に概説する。次に、オプション機能として、OM オプション、BYLEVEL オプション、AT オプションを用いた LSMEANS ステートメントによる LS-Means の算出について述べる。さらに、GLM プロシジャによるモデル情報を用いて、PLM プロシジャを実行する手順を述べ、GLM プロシジャではサポートされていない LSMESTIMATE ステートメントによる推定を行う手順を示す。

2 GLM プロシジャによる LS-Means の算出

2.1 GLM プロシジャによる LS-Means の数理

GLM プロシジャによる LS-Means の算出までのプロセスを示すために、表 2 の 2 元配置アンバランスデータを考える。

表 2: 2 元配置アンバランスデータ例 (データセット名: twoway)

		Block	
		ブロック 1	ブロック 2
Treatment	治療 1	17, 19, 19, 21, 22, 28 ($n_{11} = 6$)	30, 39, 43 ($n_{12} = 3$)
	治療 2	22, 26, 30, 31, 33, 46 ($n_{21} = 6$)	26, 29, 31, 33 ($n_{22} = 4$)

このとき、切片パラメータを μ 、治療効果を表すパラメータを α_i ($i = 1, 2$)、ブロック効果を表すパラメータを β_j ($j = 1, 2$)、治療 i とブロック j の交互作用効果を表すパラメータを γ_{ij} 、治療 i 、ブロック j における k 番目の観測パラメータを Y_{ijk} ($k = 1, 2, \dots, n_{ij}$)、誤差パラメータを ε_{ijk} (互いに独立に平均 0、分散 σ^2 の正規分布に従うと仮定) とすると、

$$Y_{ijk} = \mu + \alpha_i + \beta_j + \gamma_{ij} + \varepsilon_{ijk} \quad (1)$$

と表せ、治療 i 、ブロック j における母平均パラメータは

$$E[Y_{ij}] = \mu + \alpha_i + \beta_j + \gamma_{ij}$$

と表せる。また、治療 i 、ブロック j の周辺母平均パラメータはそれぞれ

$$E[Y_{i.}] = \frac{E[Y_{i1}] + E[Y_{i2}]}{2} = \mu + \alpha_i + \frac{\beta_1 + \beta_2}{2} + \frac{\gamma_{i1} + \gamma_{i2}}{2} = \mu + \alpha_i + \bar{\beta} + \bar{\gamma}_i \quad (2)$$

$$E[Y_{.j}] = \frac{E[Y_{1j}] + E[Y_{2j}]}{2} = \mu + \frac{\alpha_1 + \alpha_2}{2} + \beta_j + \frac{\gamma_{1j} + \gamma_{2j}}{2} = \mu + \bar{\alpha} + \beta_j + \bar{\gamma}_j$$

と表記できる。

さらに、治療効果の差の推定を考えると、式 (2) より、

$$E[Y_{1.}] - E[Y_{2.}] = \alpha_1 - \alpha_2 + \frac{\gamma_{11} + \gamma_{12} - \gamma_{21} - \gamma_{22}}{2} \quad (3)$$

となり、 γ_{ij} が含まれることに留意しなければならない。

ここで、パラメータベクトル β を

$$\beta = (\mu \ \alpha_1 \ \alpha_2 \ \beta_1 \ \beta_2 \ \gamma_{11} \ \gamma_{12} \ \gamma_{21} \ \gamma_{22})^T,$$

係数行列 L を

$$L = \begin{bmatrix} 1 & 1 & 0 & 0.5 & 0.5 & 0.5 & 0.5 & 0 & 0 \\ 1 & 0 & 1 & 0.5 & 0.5 & 0 & 0 & 0.5 & 0.5 \\ 1 & 0.5 & 0.5 & 1 & 0 & 0.5 & 0 & 0.5 & 0 \\ 1 & 0.5 & 0.5 & 0 & 1 & 0 & 0.5 & 0 & 0.5 \\ 1 & 1 & 0 & 1 & 0 & 1 & 0 & 0 & 0 \\ 1 & 1 & 0 & 0 & 1 & 0 & 1 & 0 & 0 \\ 1 & 0 & 1 & 1 & 0 & 0 & 0 & 1 & 0 \\ 1 & 0 & 1 & 0 & 1 & 0 & 0 & 0 & 1 \end{bmatrix} \quad (4)$$

として、(周辺) 母平均パラメータベクトルは $L\beta$ を計算することによって求まる。LS-Means は、推定パラメータベクトル $\hat{\beta}$ から求まる $L\hat{\beta}$ である。

2.2 LSMEANS ステートメントによる実行

GLM プロシジャでは、LSMEANS ステートメントを用いて、LS-Means を容易に算出することができる。表 2 のデータに対して、モデル式 (1) を用いて、LS-Means を求めるための SAS プログラム例及び実行結果を一部抜粋したものを図 1 に示す。

LSMEANS ステートメントを用いたプログラム						
<pre>proc glm data=twoway; class treatment block; model y=treatment block; lsmeans treatment / cl stderr e; lsmeans block / cl stderr e; lsmeans treatment*block / cl stderr e; run;quit;</pre>						
係数ベクトルの出力 (治療効果)			LS-Means 算出結果			
最小 2 乗平均 効果	Treatment の係数 水準		Treatment	Block	y の最小 2 乗平均	標準誤差
	1	2	1	.	29.167	2.091
Intercept	1	1	2	.	30.542	1.909
Treatment 1	1	0	.	1	26.167	1.707
Treatment 2	0	1	.	2	33.542	2.259
Block 1	0.5	0.5	1	1	21.000	2.415
Block 2	0.5	0.5	1	2	37.333	3.415
Treatment*Block 1 1	0.5	0	2	1	31.333	2.415
Treatment*Block 1 2	0.5	0	2	2	29.750	2.957
Treatment*Block 2 1	0	0.5				
Treatment*Block 2 2	0	0.5				

図 1: GLM プロシジャによる LS-Means の算出 (1)

LS-Means の計算は、CLASS ステートメント及び MODEL ステートメントによるモデル推定後のプロセスとなるため、GLM プロシジャ内の LSMEANS ステートメントは上記 2 つのステートメントの後に記述する。

図 1 の例では、交互作用効果を含んだモデル式 (1) から得られる治療効果、ブロック効果、交互作用効果の LS-Means をそれぞれ求めている。なお、LSMEANS ステートメントの E オプションは、推定に用いる係数ベクトルを出力させるための機能である。推定に用いる係数ベクトルの出力は、式 (2) に対応する場合のみ抜粋して示している ($i=1, 2$)。すなわち、式 (4) の \mathbf{L} における第 1 行、第 2 行が該当する。

なお、表 1 に示すように、SAS/STAT V9.22 以降、LSMEANS ステートメントは多くのプロシジャでサポートされている。

2.3 ESTIMATE ステートメントによる実行

図 1 のように、LSMEANS ステートメントの E オプションを指定すれば、式 (4) の \mathbf{L} のような係数 (行列) を出力させることは可能である。しかし、SAS プログラム上で明示的に確認はできない。そこで、ESTIMATE ステートメントを用いて、各パラメータに対する係数を指定した上で、LS-Means を求めるための SAS プログラム例及び実行結果を一部抜粋したものを図 2 に示す。

ESTIMATE ステートメントを用いたプログラム

```

proc glm data=twoway;
  class treatment block; model y=treatment | block;
  estimate "Treatment 1" intercept 1 treatment 1 0 block 0.5 0.5 treatment*block 0.5 0.5 0 0 / e;
  estimate "Treatment 2" intercept 1 treatment 0 1 block 0.5 0.5 treatment*block 0 0 0.5 0.5 / e;
  estimate "Block 1" intercept 1 treatment 0.5 0.5 block 1 0 treatment*block 0.5 0 0.5 0 / e;
  estimate "Block 2" intercept 1 treatment 0.5 0.5 block 0 1 treatment*block 0 0.5 0 0.5 / e;
  estimate "Treatment 1 Block 1" intercept 1 treatment 1 0 block 1 0 treatment*block 1 0 0 0 / e;
  estimate "Treatment 1 Block 2" intercept 1 treatment 1 0 block 0 1 treatment*block 0 1 0 0 / e;
  estimate "Treatment 2 Block 1" intercept 1 treatment 0 1 block 1 0 treatment*block 0 0 1 0 / e;
  estimate "Treatment 2 Block 2" intercept 1 treatment 0 1 block 0 1 treatment*block 0 0 0 1 / e;
run;quit;

```

係数ベクトルの出力 (治療効果)				LS-Means 算出結果				
推定 Treatment 1 の係数		推定 Treatment 2 の係数		パラメータ	推定値	標準誤差	t 値	Pr > t
行 1		行 1						
Intercept	1	Intercept	1	Treatment 1	29.167	2.091	13.95	<.0001
Treatment 1	1	Treatment 1	0	Treatment 2	30.542	1.909	16.00	<.0001
Treatment 2	0	Treatment 2	1	Block 1	26.167	1.707	15.33	<.0001
Block 1	0.5	Block 1	0.5	Block 2	33.542	2.259	14.85	<.0001
Block 2	0.5	Block 2	0.5	Treatment 1 Block 1	21.000	2.415	8.70	<.0001
Treatment*Block 1 1	0.5	Treatment*Block 1 1	0	Treatment 1 Block 2	37.333	3.415	10.93	<.0001
Treatment*Block 1 2	0.5	Treatment*Block 1 2	0	Treatment 2 Block 1	31.333	2.415	12.98	<.0001
Treatment*Block 2 1	0	Treatment*Block 2 1	0.5	Treatment 2 Block 2	29.750	2.957	10.06	<.0001
Treatment*Block 2 2	0	Treatment*Block 2 2	0.5					

図 2: GLM プロシジャによる LS-Means の算出 (2)

ESTIMATE ステートメントにおいても E オプションを指定でき、推定に用いる係数ベクトルを出力させることができる。図 1 と同様に、推定に用いる係数ベクトルの出力は、式 (2) に対応する場合のみ抜粋して示している ($i=1, 2$)。図 2 における LS-Means の結果は、LSMEANS ステートメントを用いて求めた図 1 の結果と一致していることを確認できる。

なお、LSMEANS ステートメント同様、ESTIMATE ステートメントも多くのプロシジャでサポートされている。しかし、LOGISTIC プロシジャ等では、CLASS ステートメントの PARAM オプションを指定でき、プロシジャによってデフォルトが異なる。そのため、ESTIMATE ステートメントにおける係数の指定に注意が必要である。図 2 のような LS-Means の推定を行う場合、GLM 法によるデザイン行列の指定がわかりやすい^[8]。

3 GLM プロシジャによるオプション機能を用いた LS-Means の算出

同様の表 2 のデータを用いて、LSMEANS ステートメントのオプション機能を考える。本セクションでは、モデル式 (1) よりシンプルな交互作用効果を含まないモデル式 (5) を用いて考える。

$$Y_{ijk} = \mu + \alpha_i + \beta_j + \varepsilon_{ijk} \quad (5)$$

このとき、治療 i 、ブロック j に対して、LSMEANS ステートメントによって求まる LS-Means はそれぞれ

$$E[Y_{i.}] = \mu + \alpha_i + \frac{\beta_1 + \beta_2}{2} = \mu + \alpha_i + \bar{\beta}. \quad (6)$$

$$E[Y_{.j}] = \mu + \frac{\alpha_1 + \alpha_2}{2} + \beta_j = \mu + \bar{\alpha} + \beta_j$$

と表記できる。

さらに、治療効果の差の推定を考えると、式 (6) より、

$$E[Y_{1.}] - E[Y_{2.}] = \alpha_1 - \alpha_2 \quad (7)$$

となり、式 (3) と異なり、 α_i のみで表せる。

パラメータベクトル $\boldsymbol{\beta}$ を

$$\boldsymbol{\beta} = (\mu \quad \alpha_1 \quad \alpha_2 \quad \beta_1 \quad \beta_2)^T,$$

係数行列 \mathbf{L} を

$$\mathbf{L} = \begin{bmatrix} 1 & 1 & 0 & 0.5 & 0.5 \\ 1 & 0 & 1 & 0.5 & 0.5 \\ 1 & 0.5 & 0.5 & 1 & 0 \\ 1 & 0.5 & 0.5 & 0 & 1 \end{bmatrix} \quad (8)$$

とすると、LS-Means は、推定パラメータベクトル $\hat{\boldsymbol{\beta}}$ から求まる $\mathbf{L}\hat{\boldsymbol{\beta}}$ である。

3.1 OM オプションの利用

治療効果に対する LS-Means を考えると、表 2 のデータのようにブロックが 2 水準の場合、 β_1, β_2 に対する係数 c_1, c_2 は

$$c_1 = 1/2 = 0.5, \quad c_2 = 1/2 = 0.5 \quad \left(\sum_k c_k = 1 \right)$$

となり、各水準に対して等しい割合が係数として用いられていることがわかる。

一方、LSMEANS ステートメントの OBSMARGINS (OM) オプションを用いると、入力データセットにおける各水準のオブザベーション数に依存して、係数が設定される。表 2 のデータでは、各ブロックにおけるデータ数は

$$n_{.1} = \sum_i n_{i1} = 12$$

$$n_{.2} = \sum_i n_{i2} = 7$$

であるため、係数 c_1, c_2 は

$$c_1 = n_{.1} / n_{..} = \sum_i n_{i1} / \sum_i \sum_j n_{ij} = 12/19 = 0.6316$$

$$c_2 = n_{.2} / n_{..} = \sum_i n_{i2} / \sum_i \sum_j n_{ij} = 7/19 = 0.3684$$

となる。

すなわち、OM オプションを用いると、治療 i の LS-Means は

$$E[Y_{i.}] = \mu + \alpha_i + \frac{12}{19}\beta_1 + \frac{7}{19}\beta_2$$

と表記でき、治療効果の差の LS-Means は

$$E[Y_{1.}] - E[Y_{2.}] = \alpha_1 - \alpha_2 \tag{9}$$

となり、式 (7) と同様に α_i のみで表せる。係数行列を考えると、式 (8) の代わりに

$$\mathbf{L} = \begin{bmatrix} 1 & 1 & 0 & 0.6316 & 0.3684 \\ 1 & 0 & 1 & 0.6316 & 0.3684 \\ 1 & 0.4737 & 0.5263 & 1 & 0 \\ 1 & 0.4737 & 0.5263 & 0 & 1 \end{bmatrix}$$

を用いて、 $\mathbf{L}\hat{\boldsymbol{\beta}}$ を計算している。

LSMEANS ステートメントの OM オプションを利用して、モデル式 (5) から LS-Means を求めるための SAS プログラム例、及び対応する ESTIMATE ステートメントによる SAS プログラム例を図 3 に示す。

図 3 の SAS プログラムを実行すると、LSMEANS ステートメント及び ESTIMATE ステートメントにより求めた LS-Means は一致していることが確認できる。

さらに、治療効果の差の推定を考えると、式 (9) のように表せ、交互作用効果を含まないモデルでは、OM オプションの使用の有無に関わらず、同様の結果となる。OM オプションの有無で LS-Means の結果を比較するための SAS プログラム例及び実行結果を一部抜粋したものを図 4 に示す。

図 4 の SAS プログラムを実行すると、LSMEANS ステートメントにより求めた LS-Means は、治療効果の差の推定に関しては一致していることが確認できる。しかし、各治療の LS-Means は異なった結果となる。

LSMEANS ステートメント (OM オプション) を用いたプログラム

```
proc glm data=twoway;
  class treatment block; model y=treatment block;
  lsmeans treatment / om cl stderr e;
run;quit;
```

ESTIMATE ステートメントを用いたプログラム

```
proc glm data=twoway;
  class treatment block; model y=treatment block;
  estimate "Treatment 1" intercept 1 treatment 1 0 block 0.6316 0.3684 / e;
  estimate "Treatment 2" intercept 1 treatment 0 1 block 0.6316 0.3684 / e;
run;quit;
```

図 3: GLM プロシジャによる LS-Means の算出 (3)

OM オプションなし (デフォルト)

```
proc glm data=twoway;
  class treatment block; model y=treatment block;
  lsmeans treatment / cl diff e;
run;quit;
```

Treatment		y の最小 2 乗平均		95% 信頼限界	
1		27.538		22.166	32.910
2		31.356		26.343	36.369
効果 Treatment に対する最小 2 乗平均					
i	j	平均の差	LSMean(i)-LSMean(j) の 95% 信頼限界		
1	2	-3.818		-11.038	3.402

OM オプションあり

```
proc glm data=twoway;
  class treatment block; model y=treatment block;
  lsmeans treatment / om cl diff e;
run;quit;
```

Treatment		y の最小 2 乗平均		95% 信頼限界	
1		26.675		21.443	31.907
2		30.493		25.530	35.456
効果 Treatment に対する最小 2 乗平均					
i	j	平均の差	LSMean(i)-LSMean(j) の 95% 信頼限界		
1	2	-3.818		-11.038	3.402

図 4: GLM プロシジャによる LS-Means の算出 (4)

3.2 OM BYLEVEL オプションの利用

表 2 のデータに対して, LSMEANS ステートメントの OM オプションに加えて, BYLEVEL オプションも用いると, 入力データセットにおける各水準のオブザベーション数に依存して, 以下のように係数が設定される.

$$\text{【治療 1】 } c_1 = n_{11} / n_{.1} = n_{11} / \sum_j n_{1j} = 6/9 = 0.6667$$

$$c_2 = n_{12} / n_{.1} = n_{12} / \sum_j n_{1j} = 3/9 = 0.3333$$

$$\text{【治療 2】 } c_1 = n_{21} / n_{.2} = n_{21} / \sum_j n_{2j} = 6/10 = 0.6$$

$$c_2 = n_{22} / n_{.2} = n_{22} / \sum_j n_{2j} = 4/10 = 0.4$$

OM オプションのみを使用した場合、各治療に対して共通の係数 c_1, c_2 を用いた。BYLEVEL オプションも加えると、各治療に対する係数をそれぞれ計算する。

すなわち、OM オプションを用いると、治療 i の LS-Means はそれぞれ

$$E[Y_{1.}] = \mu + \alpha_1 + \frac{6}{9}\beta_1 + \frac{3}{9}\beta_2$$

$$E[Y_{2.}] = \mu + \alpha_2 + \frac{6}{10}\beta_1 + \frac{4}{10}\beta_2$$

と表記でき、治療効果の差の LS-Means は

$$E[Y_{1.}] - E[Y_{2.}] = \alpha_1 - \alpha_2 + \left(\frac{6}{9} - \frac{6}{10}\right)\beta_1 + \left(\frac{3}{9} - \frac{4}{10}\right)\beta_2 \quad (10)$$

となり、式 (7) や式 (9) と異なり、 α_i のみで表せない。係数行列を考えると、式 (8) の代わりに

$$\mathbf{L} = \begin{bmatrix} 1 & 1 & 0 & 0.6667 & 0.3333 \\ 1 & 0 & 1 & 0.6 & 0.4 \\ 1 & 0.5 & 0.5 & 1 & 0 \\ 1 & 0.4286 & 0.5714 & 0 & 1 \end{bmatrix}$$

を用いて、 $\mathbf{L}\hat{\boldsymbol{\beta}}$ を計算している。

LSMEANS ステートメントの OM オプション及び BYLEVEL オプションを利用して、LS-Means を求めるための SAS プログラム例、及び対応する ESTIMATE ステートメントによる SAS プログラム例を図 5 に示す。

```

LSMEANS ステートメント (OM BYLEVEL オプション) を用いたプログラム
proc glm data=twoway;
  class treatment block; model y=treatment block;
  lsmeans treatment / om bylevel cl stderr e;
run;quit;

ESTIMATE ステートメントを用いたプログラム
proc glm data=twoway;
  class treatment block; model y=treatment block;
  estimate "Treatment 1" intercept 1 treatment 1 0 block 0.6667 0.3333 / e;
  estimate "Treatment 2" intercept 1 treatment 0 1 block 0.6 0.4 / e;
run;quit;

```

図 5: GLM プロシジャによる LS-Means の算出 (5)

図 5 の SAS プログラムを実行すると、LSMEANS ステートメント及び ESTIMATE ステートメントにより求めた LS-Means は一致していることが確認できる。

さらに、OM オプション及び BYLEVEL オプションとデフォルトの LS-Means の結果を比較するための SAS プログラム例及び実行結果を一部抜粋したものを図 6 に示す。

OM オプションなし (デフォルト)					
<code>proc glm data=twoway;</code>		Treatment y の最小 2 乗平均		95% 信頼限界	
<code>class treatment block; model y=treatment block;</code>		1	27.538	22.166	32.910
<code>lsmeans treatment / cl diff e;</code>		2	31.356	26.343	36.369
<code>run; quit;</code>		効果 Treatment に対する最小 2 乗平均			
		i	j	平均の差	LSMean(i)-LSMean(j) の 95% 信頼限界
		1	2	-3.818	-11.038 3.402

OM BYLEVEL オプションあり					
<code>proc glm data=twoway;</code>		Treatment y の最小 2 乗平均		95% 信頼限界	
<code>class treatment block; model y=treatment block;</code>		1	26.444	21.219	31.670
<code>lsmeans treatment / om bylevel cl diff e;</code>		2	30.700	25.743	35.657
<code>run; quit;</code>		効果 Treatment に対する最小 2 乗平均			
		i	j	平均の差	LSMean(i)-LSMean(j) の 95% 信頼限界
		1	2	-4.256	-11.458 2.947

図 6: GLM プロシジャによる LS-Means の算出 (6)

図 6 の SAS プログラムを実行すると、LSMEANS ステートメントにより求めた LS-Means は、OM オプション及び BYLEVEL オプションを利用すると、デフォルトと結果が一致しないことがわかる。治療効果の差の推定を考える場合、式 (10) のように表せ、式 (7) とは大きく異なる表記となるため、結果は一致しないことになる。

3.3 AT オプションの利用

モデル式 (4) に連続量 z_{ijk} を説明変数に含んだモデル (11) を考える。

$$Y_{ijk} = \mu + \alpha_i + \beta_j + \delta \cdot z_{ijk} + \varepsilon_{ijk} \quad (11)$$

このとき、治療 i 、ブロック j に対して、LSMEANS ステートメントによって求まる LS-Means は

$$E[Y_{.i}] = \mu + \alpha_i + \frac{\beta_1 + \beta_2}{2} + \bar{z} \cdot \delta \quad (12)$$

$$E[Y_{.j}] = \mu + \frac{\alpha_1 + \alpha_2}{2} + \beta_j + \bar{z} \cdot \delta$$

と表記でき、パラメータベクトル $\boldsymbol{\beta}$ を

$$\boldsymbol{\beta} = (\mu \ \alpha_1 \ \alpha_2 \ \beta_1 \ \beta_2 \ \delta)^T,$$

係数行列 \mathbf{L} を

$$\mathbf{L} = \begin{bmatrix} 1 & 1 & 0 & 0.5 & 0.5 & \bar{z} \\ 1 & 0 & 1 & 0.5 & 0.5 & \bar{z} \\ 1 & 0.5 & 0.5 & 1 & 0 & \bar{z} \\ 1 & 0.5 & 0.5 & 0 & 1 & \bar{z} \end{bmatrix}$$

とすると、LS-Means は、推定パラメータベクトル $\hat{\boldsymbol{\beta}}$ から求まる $\mathbf{L}\hat{\boldsymbol{\beta}}$ である。式 (12) におけるパラメータ δ の係数として、算術平均 \bar{z} が用いられていることがわかる。

ここで、表2のデータは $\bar{z}=14.5263$ であると仮定する。このとき、LSMEANS ステートメントの AT オプションを利用して、LS-Means を求めるための SAS プログラム例、及び対応する ESTIMATE ステートメントによる SAS プログラム例を図7に示す。

LSMEANS ステートメント	ESTIMATE ステートメント
<pre>proc glm data=twoway; class treatment block; model y=treatment block z; lsmeans treatment / cl stderr e; run;quit;</pre>	<pre>proc glm data=twoway; class treatment block; model y=treatment block z; estimate "Treatment 1" intercept 1 treatment 1 0 block 0.5 0.5 z 14.5263 / e; estimate "Treatment 2" intercept 1 treatment 0 1 block 0.5 0.5 z 14.5263 / e; run; quit;</pre>
<p>AT オプション</p> <pre>proc glm data=twoway; class treatment block; model y=treatment block z; lsmeans treatment / cl stderr e at z=14.5263; run;quit;</pre>	

図7: GLM プロシジャによる LS-Means の算出 (7)

LSMEANS ステートメントの AT オプションを用いると、図7における "at z=14.5263" の値を変えることによって、 \bar{z} の代わりに任意の値を係数として用いることができる。

4 PLM プロシジャによる LS-Means の算出

前節まで、モデル推定後の LS-Means の算出までのプロセスについて、GLM プロシジャ内の記述方法を示した。OM オプションのようなオプションを用いた結果が必要となった場合、オプションを追記した上で、再び GLM プロシジャを実行しなければならない。このような操作は、オブザベーション数が多い場合やモデル式が複雑である場合、モデル推定に再度多くの時間を要することになる。

4.1 PLM プロシジャの概要

SAS/STAT V9.22 より、GLM プロシジャ等のプロシジャを用いて推定されているモデル情報を呼び出し、統計量やグラフ表示などを行うことのできる、PLM プロシジャが追加された。

PLM プロシジャを利用するにあたっては、まず GLM プロシジャ等のプロシジャを実行する際に STORE ステートメントを用いて、プロシジャから得られるモデル情報、すなわちアイテムストアをバイナリファイルとして保存する。STORE ステートメントは、表1に示したプロシジャで実行できる。そして、PLM プロシジャの RESTORE ステートメントを用いて、アイテムストアとして保存したモデル情報を呼び出し、統計量やグラフ表示などを行うことができる。モデル情報に基づく実行であるため、再度の入力データセットの参照、線形モデルのプロシジャを実行することなく、モデル推定後のプロセスを実行できる。

4.2 LSMEANS ステートメントによる実行

GLM プロシジャによるモデル情報を "twowayfit" という名前のアイテムストアに保存し、PLM プロシジャで実行するまでの SAS プログラム例を図 8 に示す。表 2 のデータに対して、交互作用効果を含むモデル式 (1) による解析を考えている。

```
STORE ステートメントによるアイテムストアの保存  PLM プロシジャによる LS-Means の算出

proc glm data=twoway;
  class treatment block; model y=treatment | block;
  store twowayfit;
run;
quit

proc plm restore=twowayfit;
  lsmeans treatment / cl;
  lsmeans block / cl;
  lsmeans treatment*block / cl;
run;
```

図 8 : GLM 及び PLM プロシジャによる LS-Means の算出 (1)

図 8 の SAS プログラムの実行結果のうち、治療とブロックの交互作用効果に対する LS-Means として、ODS GRAPHICS による出力結果を図 9 に示す。

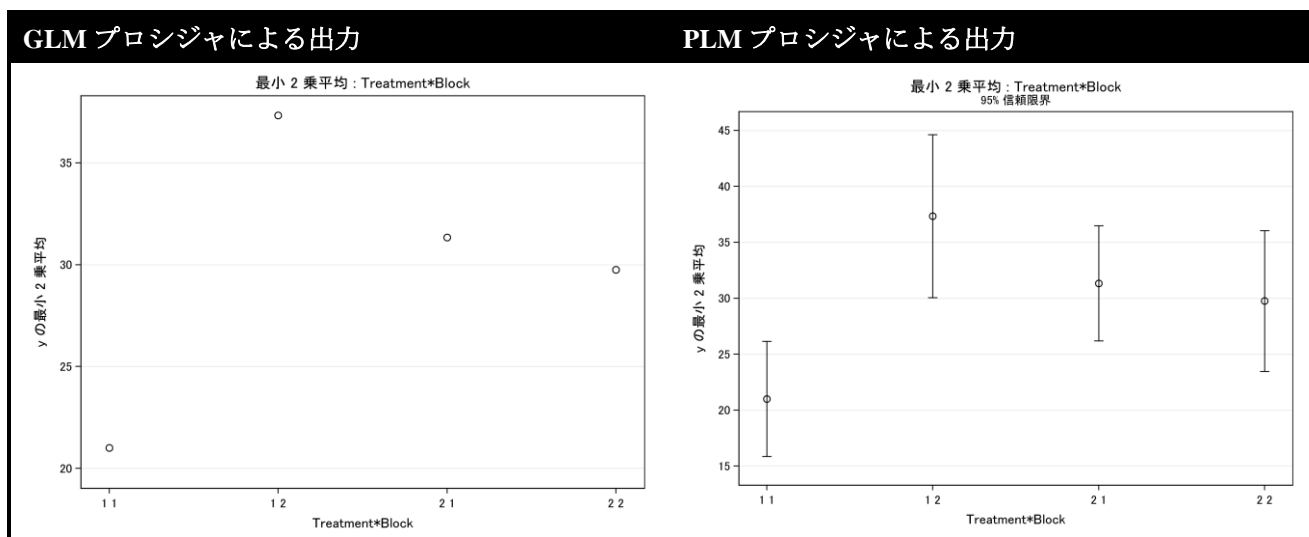


図 9 : GLM 及び PLM プロシジャによる LS-Means の算出 (2)

図 9 のように、GLM プロシジャでは各水準の LS-Means がプロットされる一方、PLM プロシジャでは信頼区間も出力される点で有用といえる。

なお、アイテムストアとして保存したモデル情報は、図 10 のような SAS プログラムを実行することによって確認できる。

```
proc plm restore=twowayfit;
  show all;
run;
```

図 10 : すべてのモデル情報の表示を行う SAS プログラム

4.3 LSMESTIMATE ステートメントによる実行

SAS/STAT V9.22 より、LSMESTIMATE ステートメントが追加された。モデルパラメータの線形式を用いる ESTIMATE ステートメントと異なり、LSMESTIMATE ステートメントでは LS-Means の線形式を用いる。浜田 (2013) は、決まりきった対比較等の不必要な項目まで冗長に出力されてしまう LSMEANS ステートメントに比べて、特定の群間比較の LS-Means のみを出力できる LSMESTIMATE ステートメントを推奨している。しかし、表 1 で示したように、GLM プロシジャでは LSMESTIMATE ステートメントはサポートされていない。そこで、GLM プロシジャによるモデル情報を用いて、PLM プロシジャで LSMESTIMATE ステートメントによる推定を行う方法を示す。

図 8 において保存したアイテムストア "twowayfit" を用いて、PLM プロシジャによって LSMESTIMATE ステートメントによる SAS プログラム例を図 11 に示す。

図 11 のように、LSMESTIMATE ステートメントでは、LS-Means を算出したい変数とその係数のみを指定すれば、ESTIMATE ステートメントを用いて求めた図 2 と同様の結果を得ることができる。

```
LSMESTIMATE ステートメントを用いたプログラム

proc plm restore=twowayfit;

  lsestimate treatment "Treatment 1" 1 0 / cl e;
  lsestimate treatment "Treatment 2" 0 1 / cl e;

  lsestimate block "Block 1" 1 0 / cl e; lsestimate block "Block 2" 0 1 / cl e;

  lsestimate treatment * block "Treatment 1 Block 1" 1 0 0 0 / cl e;
  lsestimate treatment * block "Treatment 1 Block 2" 0 1 0 0 / cl e;
  lsestimate treatment * block "Treatment 2 Block 1" 0 0 1 0 / cl e;
  lsestimate treatment * block "Treatment 2 Block 2" 0 0 0 1 / cl e;

run;
```

図 11 : GLM 及び PLM プロシジャによる LS-Means の算出 (3)

4.4 SLICE ステートメントによる実行

SLICE ステートメントは、LSMESTIMATE ステートメントと同様に、SAS/STAT V9.22 から追加された機能であり、カテゴリ変数が 2 次以上の交互作用項に対し、1 つの変数の水準をスライスした解析ができる。LSMEANS ステートメントと同じオプションを用いることができる。LSMESTIMATE ステートメントと同様に、GLM プロシジャではサポートされていない。

図 8 において保存したアイテムストア "twowayfit" を用いて、PLM プロシジャによって SLICE ステートメントによる SAS プログラム例及び出力結果を図 12 に示す。

図 12 は、GLM プロシジャによる出力結果を比較対照として出力している。GLM プロシジャによる対比較の結果においては、ODS GRAPHICS が冗長に出力されてしまう一方、PLM プロシジャの SLICE ステートメントを用いると、特定の水準に対する交互作用効果の結果のみ出力できる。

GLM プロシジャ LSMEANS ステートメント

PLM プロシジャ SLICE ステートメント

```
proc glm data=twoway;
  class treatment block; model y=treatment | block;
  lsmeans treatment * block
  / diff cl adjust=simulate(report seed=4989);
run;quit;
```

```
proc plm restore=twowayfit;
  slice treatment * block
  / diff cl sliceby(block="1");
run;
```

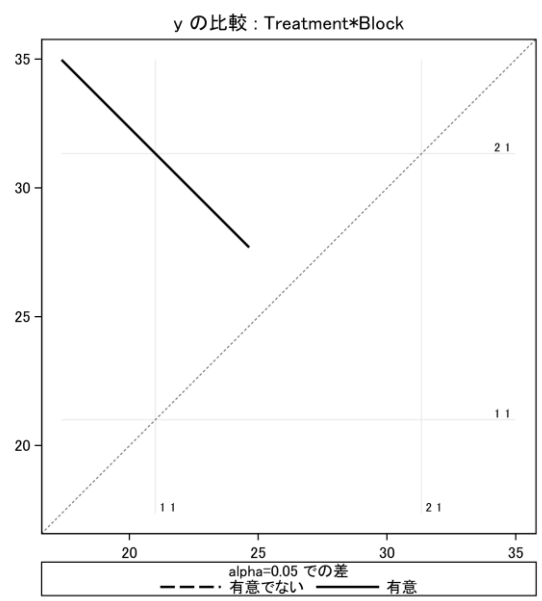
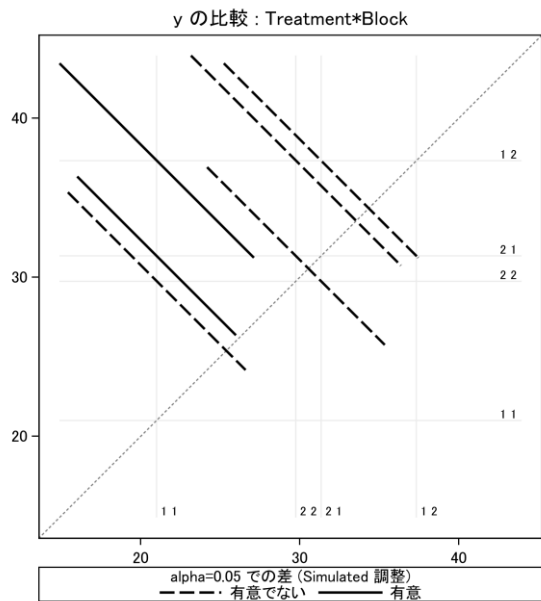


図 12 : GLM 及び PLM プロシジャによる LS-Means の算出 (4)

5 まとめ

本稿では、線形モデルによる解析を行うためのプロシジャによるモデルの推定後に算出する LS-Means について、LSMEANS ステートメントで算出される数値を示した上で、ESTIMATE ステートメントを用いて SAS プログラム上で明示的に概説した。次に、LSMEANS ステートメントのオプション機能として、OM オプション、BYLEVEL オプション、AT オプションを用いた LS-Means の算出方法を詳述した。さらに、線形モデルによる解析を行うためのプロシジャによって、推定されたモデル情報をアイテムストアとして保存した上で、PLM プロシジャによる推定を行う方法を解説した。加えて、LSMESTIMATE ステートメント及び SLICE ステートメントによる解析方法についても示した。

表 1 に示したように、現在の SAS のバージョンでは、LS-Means 算出のためのステートメントの選択肢は幅広く、多くのプロシジャで実行できる。特に、LSMEANS ステートメントは、ESTIMATE ステートメントよりもはるかに簡潔な SAS プログラムで LS-Means を算出できるといえる。さらに、OM オプションを指定すれば、入力データセットにおける各水準のオブザベーション数に依存して係数を設定できる。例えば、ランダム化臨床試験のデータ解析を考えた場合、割付因子として調整した共変量を説明変数に含めて、群間の LS-Means の差の推定を行うことが多い。このとき、LS-Means の差であれば、OM オプションを指定した結

果はデフォルトの結果と一致する。しかし、LS-Means の差ではなく、各群の LS-Means を求めることに関心がある場合、割付因子として調整している説明変数に対する係数を指定できる、LSMEANS ステートメントの OM オプションは有用であるといえる。また、非線形なデータに対して、ある特定の説明変数の値における LS-Means の推定を行う場合、LSMEANS ステートメントの AT オプションが有用になると考えられる。

参考文献

- [1] Cai W. Making Comparisons Fair: How LS-Means Unify the Analysis of Linear Models. *Proceedings of the SAS Global Forum*. Cary, NC: SAS Institute Inc., 2014. Available at <http://support.sas.com/resources/papers/proceedings14/SAS060-2014.pdf>.
- [2] High R. Plotting Differences among LSMEANS in Generalized Linear Models. *Proceedings of the SAS Global Forum*. Cary, NC: SAS Institute Inc., 2014. Available at <http://support.sas.com/resources/papers/proceedings14/1902-2014.pdf>.
- [3] Kiernan K, Tobias R, Gibbs P, Tao J. Making CONTRAST and ESTIMATE Statements Made Easy: The LSMESTIMATE Statement. *Proceedings of the SAS Global Forum*. Cary, NC: SAS Institute Inc., 2011. Available at <http://support.sas.com/resources/papers/proceedings11/351-2011.pdf>.
- [4] SAS Institute Inc. *SAS/STAT(R) 9.3 User's Guide*, Cary, NC, USA: SAS Institute Inc; 2011.
- [5] 関根暁史. PLM プロシジャによる回帰分析と予測の分離. SAS ユーザー総会 論文集 2013, 275-290.
- [6] 竹内啓, 高橋行雄, 大橋靖雄, 芳賀敏郎. SAS による実験データの解析. 東京大学出版会, 1989.
- [7] 浜田知久馬. SAS 生存時間解析プロシジャの最新の機能拡張. SAS ユーザー総会 論文集 2013, 3-72.
- [8] 吉田早織, 魚住龍史. 線形モデルにおける CLASS ステートメントの機能. SAS ユーザー総会 論文集 2014.

連絡先

E-mail : uozumi@kuhp.kyoto-u.ac.jp