glm における TypeI，TypeII および TypeIII の計算例

田辺健一郎

Sample Calculations of TypeI, TypeII, and TypeIII in glm

Kenichiro Tanabe

要旨

glm プロシジャにおける TypeI，TypeII および TypeIII 平方和の計算方法について小データを用いて例示する．また，これらの平方和の相違点について簡単に考察する．

キーワード：glm，平方和，TypeI，TypeII，TypeIII，バランス，アンバランス，直交

1 はじめに

glm プロシジャには，モデル平方和の分解方法として TypeI〜TypeIV の 4 種類が用意されている．4 つのうち，よく用いられるのは TypeI，TypeII および TypeIII の 3 つである．TypeI，TypeII および TypeIII という言葉はよく聞くが，どのように計算するのだろうか？また，どのような違いがあるのだろうか？という疑問から，小データを使ってそれらの計算方法について調べてみようと考えた．

※平方和，分散分析，回帰分析，最小二乗推定，ベクトルの直交，基底等の数理統計や線形代数の内容を少しだけ暗に使用しております．論文集の内容に若干項目を追加いたしました．ご質問，ご意見等ございましたら kenichiro_tanabe13@yahoo.co.jp までお願いいたします。

2 平方和

TypeI，TypeII および TypeIII は，回帰分析や分散分析に出てくる「平方和」に関連している．平方和については「総平方和」,「モデル平方和」および「残差平方和」があり，

$$\text{総平方和} \ = \ \text{モデル平方和} \ + \ \text{残差平方和} \tag{1}$$

という関係がある.「総平方和」とは応答変数の全体のバラツキ,「モデル平方和」とはモデルによって説明できた応答変数のバラツキ，そして「残差平方和」とはモデルによって説明することができなかった応答変数のバラツキ，とイメージできる．

TypeI，TypeII，および TypeIII は，これら平方和のうちの「モデル平方和」を分解する方法である.「モデル平方和」は，回帰分析であれば各説明変数，分散分析であれば各要因，さらにはそれらの交互作用項を考慮する場合は交互作用項も含めて，モデル全体で説明できた応答変数のバラツキのことである．今，例として，応答変数 y のバラツキを要因 A，要因 B，およびそれらの交互作用項 $A*B$ で説明する分散分析モデルを考えると,「モデル平方和」は，メンバー A，B，および $A*B$ の全てで行った (応答変数 y のバラツキに対する) 説明量とイメージできる.「分解」というのは，このメンバー全てで行った説明量を各メンバーごとの説明量に分けることである．簡単な図を以下に示す[1]．

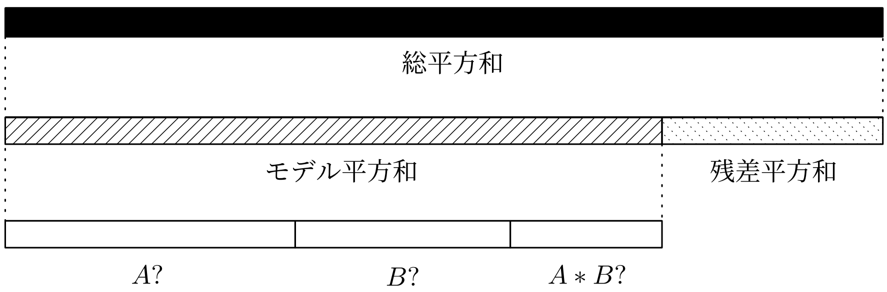


図 1:平方和分解のイメージ

要因 A が行った説明量はいくらと考えるのか？要因 B が行った説明量はいくらと考えるのか？そして，交互作用項 $A*B$ が行った説明量はいくらと考えるのか？図 1 では各メンバーに分けた説明量の合計がメンバー全てで行った説明量に一致しているように見えるが，そうならない場合もあるのではないか？

[1] n をオブザベーション数，$\bar{y}$ を応答変数の標本平均として，この図の総平方和は，$\sum_{i=1}^{n} y_i^2$ から切片のみのモデルのモデル平方和 $n\bar{y}^2$ を引いた，平均調整済み総平方和 $\sum_{i=1}^{n} y_i^2 - n\bar{y}^2$ となっている．

3 データとモデル

今回使用するデータおよびモデル式について述べる．

3.1 データ

以下の表 1 にある二元配置分散分析データを用いる．要因 A および要因 B の水準数はいずれも 2 である．欠損のセル (値が 1 つも入っていないセル) を無くし，少なくとも 1 つのセルに入っている値の数が他のセルと異なる (アンバランス) なデータを用意した[2]．また，簡単にするため，標本数は計算可能な限り少なくした．本データにおいては，A_2B_1 セルのみ値が 2 つ (2 と 4) 入っていて，他のセルでは値が 1 つだけ入っている．

表 1:サンプルデータ

	B_1	B_2
A_1	0	18
A_2	2, 4	6

3.2 モデル式

応答変数 (各セルの中にある値) を y で表すことにして，表 1 のデータに以下のような交互作用項を含む二元配置分散分析の線形モデル

$$y_{ijk} = \mu + \alpha_i + \beta_j + \gamma_{ij} + \varepsilon_{ijk} \tag{2}$$

を仮定する．ここで，i は要因 A の水準番号，j は要因 B の水準番号，k は該当するセル内のオブザベーション番号を表す添字であり，μ は総平均，α_i は A の水準 i の主効果，β_j は B の水準 j の主効果を示すパラメータ，γ_{ij} は A_iB_j の交互作用を示すパラメータ，そして ε_{ijk} はランダム誤差である．今は要因 A，要因 B，そして交互作用項 $A * B$ の順でモデルに組み込むと考える[3]．

[2]全てのセルに同じ数だけ値が入っている (繰り返し数が等しい) データはバランスデータである．

[3]高次の交互作用項は最後に組み込む

4 プログラムと出力結果

4.1 プログラム

以降に述べる内容に関連する結果を出力するための SAS プログラムである．ods select ステートメントは，必要な情報のみを出力するよう選択している．DesignPoints はデザイン行列 X，OverallANOVA は分散分析表，ModelANOVA は TypeI〜III 平方和，XPX は $X'X$，InvXPX は一般化逆行列 $(X'X)^-$ そして ParameterEstimates は正規方程式 $(X'X)b = X'y$ の解 b を選択する．

```
/*********************************************************************/
/************************* サンプルデータの作成 ************************/
/*********************************************************************/
data sampledata ;
  input A B Y @@ ;
datalines ;
1 1 0   1 2 18   2 1 2   2 1 4   2 2 6
;

/*********************************************************************/
/**************** glm 実行 (M4 モデルのみ glmmod も実行) ****************/
/*********************************************************************/
* (M0) 切片のみのモデル ;
ods select OverallANOVA ;
proc glm data = sampledata ;
  model Y = ;
run ;
quit ;

* (M1) 要因 A を組み込んだモデル ;
ods select OverallANOVA ;
proc glm data = sampledata ;
  class A ;
  model Y = A ;
run ;
quit ;

* (M2) 要因 B を組み込んだモデル ;
ods select OverallANOVA ;
proc glm data = sampledata ;
  class B ;
  model Y = B ;
run ;
quit ;

* (M3) 要因 A と要因 B を A, B の順で組み込んだモデル ;
ods select OverallANOVA ;
proc glm data = sampledata ;
  class A B ;
  model Y = A B ;
run ;
quit ;

* (M4) 要因 A, 要因 B および交互作用項 A*B を A, B, A*B の順で組み込んだモデル ;
ods select DesignPoints ;
proc glmmod data = sampledata ; /*** デザイン行列 X の表示 ***/
  class A B ;
  model Y = A B A*B ;
run ;
ods select OverallANOVA ModelANOVA ParameterEstimates XPX InvXPX ;
proc glm data = sampledata ;
  class A B ;
  model Y = A B A*B / solution xpx inverse ss1 ss2 ss3 ;
run ;
quit ;

* (M5) 要因 B と交互作用項 A*B を B, A*B の順で組み込んだモデル ;
ods select OverallANOVA ;
proc glm data = sampledata ;
  class A B ;
  model Y = B A*B ;
run ;
quit ;
```

4.2 出力結果

プログラムの出力結果である．最初に切片のみのモデル (これを前述プログラムの記述に倣って M0 モデルとする) の分散分析表が以下のように出力される．平均調整されないため,「Uncorrected Total」と出力されている．M0 モデルのモデル平方和は $n\bar{y}^2 = 5 * (0 + 18 + 2 + 4 + 6)^2/5^2 = 180$ であり，総平方和は $\sum_{i=1}^{5} y_i^2 = 0^2 + 18^2 + 2^2 + 4^2 + 6^2 = 380$ であることが分かる．

要因	自由度	平方和	平均平方	F 値	Pr > F
Model	1	180.0000000	180.0000000	3.60	0.1306
Error	4	200.0000000	50.0000000		
Uncorrected Total	5	380.0000000			

次に要因 A を組み込んだモデル (M1 モデルとする) の分散分析表が以下のように出力される．model ステートメントの等号の右に要因または説明変数等を記述すると，システムでは平均調整済みの値 (切片のみのモデルのモデル平方和を引いた値) をモデル平方和および総平方和として出力する．平均調整しているので「Corrected Total」と表示されている．このモデルの平均調整済みモデル平方和は 30(平均調整しない場合は 30+180=210) であり，平均調整済み総平方和は 200(平均調整しない場合は 200+180=380) である[4]．

要因	自由度	平方和	平均平方	F 値	Pr > F
Model	1	30.0000000	30.0000000	0.53	0.5195
Error	3	170.0000000	56.6666667		
Corrected Total	4	200.0000000			

次に要因 B を組み込んだモデル (M2 モデルとする) の分散分析表が以下のように出力される．このモデルの平均調整済みモデル平方和は 120(平均調整しない場合は 120+180=300) であることが分かる．

要因	自由度	平方和	平均平方	F 値	Pr > F
Model	1	120.0000000	120.0000000	4.50	0.1240
Error	3	80.0000000	26.6666667		
Corrected Total	4	200.0000000			

次に要因 A と要因 B を組み込んだモデル (M3 モデルとする) の分散分析表が以下のように出力される．このモデルの平均調整済みモデル平方和は $133\frac{5}{7}$(平均調整しない場合は $133\frac{5}{7}$+180=$313\frac{5}{7}$) であることが分かる[5]．

要因	自由度	平方和	平均平方	F 値	Pr > F
Model	2	133.7142857	66.8571429	2.02	0.3314
Error	2	66.2857143	33.1428571		
Corrected Total	4	200.0000000			

[4] M1～M5 のモデルに対して総平方和は同じ値である．

[5] $133\frac{5}{7}$ は $133 + \frac{5}{7}$ のことである．以降も帯分数で記述する．

次に要因 A，要因 B および交互作用項 $A*B$ を組み込んだモデル (M4 モデルとする) のデザイン行列 X(Y 列よりも右の列が X となる)，$X'X$，$X'X$ の一般化逆行列 $(X'X)^{-}$，分散分析表，TypeI〜III 平方和および正規方程式 $(X'X)b = X'y$ の解 b が以下のように出力される．このモデルの平均調整済みモデル平方和は 198(平均調整しない場合は 198+180=378) であることが分かる．M4 モデルが平方和計算対象モデルである．

デザイン点

オブザベーション番号	Y	1	2	3	4	5	6	7	8	9
		列の番号								
1	0	1	1	0	1	0	1	0	0	0
2	18	1	1	0	0	1	0	1	0	0
3	2	1	0	1	1	0	0	0	1	0
4	4	1	0	1	1	0	0	0	1	0
5	6	1	0	1	0	1	0	0	0	1

X'X 行列

	Intercept	A 1	A 2	B 1	B 2	A*B 1 1	A*B 1 2	A*B 2 1	A*B 2 2	Y
Intercept	5	2	3	3	2	1	1	2	1	30
A 1	2	2	0	1	1	1	1	0	0	18
A 2	3	0	3	2	1	0	0	2	1	12
B 1	3	1	2	3	0	1	0	2	0	6
B 2	2	1	1	0	2	0	1	0	1	24
A*B 1 1	1	1	0	1	0	1	0	0	0	0
A*B 1 2	1	1	0	0	1	0	1	0	0	18
A*B 2 1	2	0	2	2	0	0	0	2	0	6
A*B 2 2	1	0	1	0	1	0	0	0	1	6
Y	30	18	12	6	24	0	18	6	6	380

X'X 一般化逆行列 (g2)

	Intercept	A 1	A 2	B 1	B 2	A*B 1 1	A*B 1 2	A*B 2 1	A*B 2 2	Y
Intercept	1	-1	0	-1	0	1	0	0	0	6
A 1	-1	2	0	1	0	-2	0	0	0	12
A 2	0	0	0	0	0	0	0	0	0	0
B 1	-1	1	0	1.5	0	-1.5	0	0	0	-3
B 2	0	0	0	0	0	0	0	0	0	0
A*B 1 1	1	-2	0	-1.5	0	3.5	0	0	0	-15
A*B 1 2	0	0	0	0	0	0	0	0	0	0
A*B 2 1	0	0	0	0	0	0	0	0	0	0
A*B 2 2	0	0	0	0	0	0	0	0	0	0
Y	6	12	0	-3	0	-15	0	0	0	2

要因	自由度	平方和	平均平方	F 値	Pr > F
Model	3	198.0000000	66.0000000	33.00	0.1271
Error	1	2.0000000	2.0000000		
Corrected Total	4	200.0000000			

要因	自由度	Type I 平方和	平均平方	F 値	Pr > F
A	1	30.0000000	30.0000000	15.00	0.1609
B	1	103.7142857	103.7142857	51.86	0.0878
A*B	1	64.2857143	64.2857143	32.14	0.1111

要因	自由度	Type II 平方和	平均平方	F 値	Pr > F
A	1	13.7142857	13.7142857	6.86	0.2322
B	1	103.7142857	103.7142857	51.86	0.0878
A*B	1	64.2857143	64.2857143	32.14	0.1111

要因	自由度	Type III 平方和	平均平方	F 値	Pr > F
A	1	23.1428571	23.1428571	11.57	0.1820
B	1	126.0000000	126.0000000	63.00	0.0798
A*B	1	64.2857143	64.2857143	32.14	0.1111

パラメータ		推定値		標準誤差	t 値	Pr > |t|
Intercept		6.00000000	B	1.41421356	4.24	0.1474
A	1	12.00000000	B	2.00000000	6.00	0.1051
A	2	0.00000000	B	.	.	.
B	1	-3.00000000	B	1.73205081	-1.73	0.3333
B	2	0.00000000	B	.	.	.
A*B	1 1	-15.00000000	B	2.64575131	-5.67	0.1111
A*B	1 2	0.00000000	B	.	.	.
A*B	2 1	0.00000000	B	.	.	.
A*B	2 2	0.00000000	B	.	.	.

最後に要因 B と交互作用項 $A * B$ を組み込んだモデル (M5 モデルとする) の分散分析表が以下のように出力される．このモデルの平均調整済みモデル平方和は 198(平均調整しない場合は 198+180=378) であることが分かる[6]．

要因	自由度	平方和	平均平方	F 値	Pr > F
Model	3	198.0000000	66.0000000	33.00	0.1271
Error	1	2.0000000	2.0000000		
Corrected Total	4	200.0000000			

[6] M4 モデルから要因 A を除いても，モデル平方和が M4 モデルに等しい．

5 計算例

ここでは TypeI～TypeIII 平方和の計算を行う．平均調整しないモデル平方和に対し $R(\cdot)$ の記号を使用し，平均調整しないモデル平方和の差分に $R(\cdot|\cdot)$ の記号を使用する．例えば，$R(\mu)$ は切片のみのモデルのモデル平方和，$R(\mu,\alpha)$ は切片と要因 A を含む M1 モデルのモデル平方和，$R(\mu,\alpha,\beta)$ は切片，要因 A および要因 B を含む M3 モデルのモデル平方和，$R(\mu,\alpha,\beta,\gamma)$ は切片，要因 A，要因 B および交互作用項 $A*B$ を含む M4 モデルのモデル平方和，$R(\mu,\beta,\gamma)$ は切片，要因 B および交互作用項 $A*B$ を含む M5 モデルのモデル平方和を表す．また，$R(\alpha|\mu,\beta)$ は，$R(\mu,\alpha,\beta)-R(\mu,\beta)$，つまり，切片，要因 A および要因 B を含む M3 モデルのモデル平方和と切片と要因 B を含む M2 モデルのモデル平方和の差分を表す.

5.1 TypeI の計算

モデルに切片は既にあるものとして話を進める．TypeI は各要因に対する平方和を，モデル平方和の増加量によって，モデルに組み込んだ順 (glm プロシジャの model ステートメントに記述した要因および交互作用項の順) に評価する．簡単にいえば，先にモデルに組み込まれた要因が応答変数 y のバラツキを (複数の要因が説明できるような重複部分を説明できるという意味で) 多く説明できる[7].

最初に要因 A をモデルに組み込み，次に要因 B，最後に各要因よりも高次の交互作用項 $A*B$ を組み込む．最初に組み込まれた要因 A が，できる範囲の説明を全て行う．その量が全て要因 A の平方和となる．次にモデルに組み込まれた要因 B は，要因 A がやり残した説明の中から，できる範囲の説明 (それがモデル平方和の増加分となる) を全て行う．最後にモデルに組み込まれた交互作用項 $A*B$ は，要因 A および要因 B の両方がやり残した説明の中から，できる範囲の説明 (それがモデル平方和の増加分となる) を全て行う．つまり，以下のように計算する[8].

$$\begin{cases} SS_A & \triangleq R(\alpha|\mu)=R(\mu,\alpha)-R(\mu)=SSR_{\mathrm{M1}}-(180-180)=30 \\ SS_B & \triangleq R(\beta|\mu,\alpha)=R(\mu,\alpha,\beta)-R(\mu,\alpha)=SSR_{\mathrm{M3}}-SSR_{\mathrm{M1}}=133\frac{5}{7}-30=103\frac{5}{7} \\ SS_{A*B} & \triangleq R(\gamma|\mu,\alpha,\beta)=R(\mu,\alpha,\beta,\gamma)-R(\mu,\alpha,\beta)=SSR_{\mathrm{M4}}-SSR_{\mathrm{M3}}=198-133\frac{5}{7}=64\frac{2}{7} \end{cases}$$

となり，M4 モデルにおける glm の出力と一致する．イメージは図 2 のようになる．重複部分については，常にモデルに先に組み込まれた要因の説明量となる．

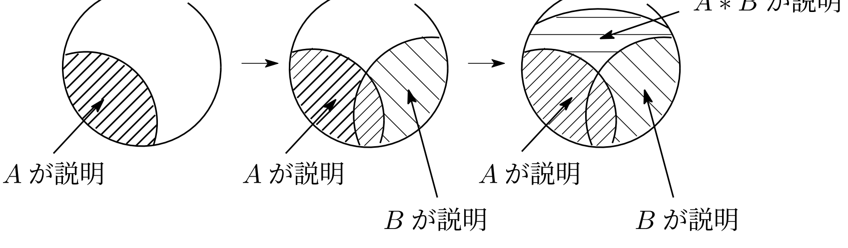


図 2:TypeI 平方和のイメージ

[7]多項式回帰等で用いる．一般的には低次の項からモデルに組み込み，有意性を検討してモデルの次数を決定する．
[8]平均調整済みモデル平方和を SSR_{M1} 等で表している．計算の際，切片のみのモデルの平方和の値が相殺されている．

5.1.1 重複部分が無いと (直交していると)

図 2 の中央図の重複している部分 (要因 A と要因 B のどちらもが説明できる部分) が空集合でない，つまり要因 A と要因 B が直交していない場合は，どちらが先にモデルに組み込まれるかによって要因 A および要因 B の平方和が異なる[9]．よってモデルへの要因投入順序に依存する．しかし，各要因が直交している場合には投入順序に依存しない．このことを図で表してみる．各要因が直交している場合は図 3 のようなイメージとなる．

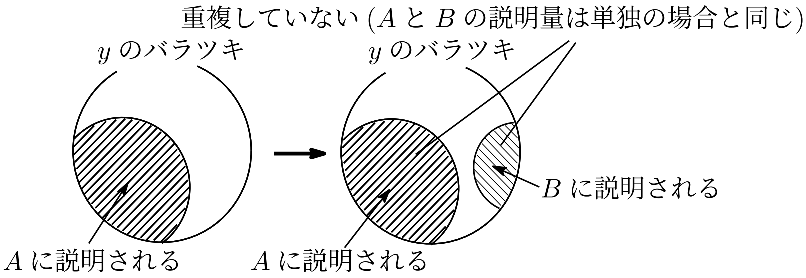
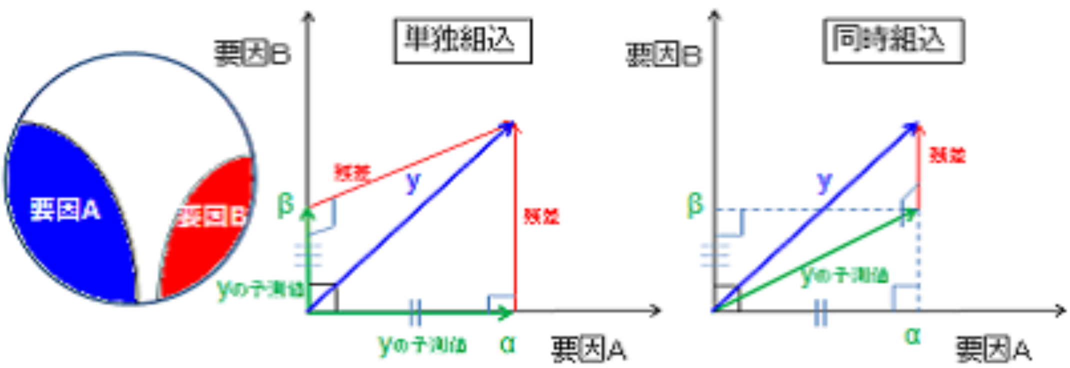


図 3:要因 A と要因 B が直交するイメージ

この場合は，要因 A と要因 B が説明できる部分が完全に分かれているので，要因 A を先にモデルに組み込んでも，要因 B を先にモデルに組み込んでも，説明できる部分の大きさ，つまり要因 A の平方和と要因 B の平方和は，モデルに要因を単独で組み込んだ場合の説明量と同じになる．これは，モデルに要因 A だけを組み込んだ場合とモデルに要因 A と要因 B の両方を組み込んだ場合を考えたとき，要因 A に対するモデルパラメータの推定値が同じ値になるということに由来する[10]．



[9]モデルに要因 B が先に組み込まれた場合，重複部分はの要因 B の説明量の一部となる．

[10]回帰分析の例では，x_1 と x_2 が直交する場合，単回帰 $y = \beta_0 + \beta_1 x_1 + \varepsilon$ モデルと重回帰 $y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \varepsilon$ モデルで回帰係数を比較すると，x_1 に対する単回帰の回帰係数 $\hat{\beta}_1$ と重回帰の回帰係数 $\hat{\hat{\beta}}_1$ が同じ値になる．

5.1.2 直交の条件

デザイン行列 X の要因 A，要因 B に関する部分について，それぞれの列の和が 0 になるように基準化 (各列ベクトルに対してそれぞれの列の平均を引く) し，これを $\tilde{X}_{A,B}$ とする．

$$
\begin{array}{cccc} A_1 & A_2 & B_1 & B_2 \end{array}
\qquad\qquad
\begin{array}{cccc} A_1 & A_2 & B_1 & B_2 \end{array}
$$

$$
\begin{pmatrix} 1 & 0 & 1 & 0 \\ 1 & 0 & 0 & 1 \\ 0 & 1 & 1 & 0 \\ 0 & 1 & 1 & 0 \\ 0 & 1 & 0 & 1 \end{pmatrix} \rightarrow \tilde{X}_{A,B} = \begin{pmatrix} \frac{3}{5} & -\frac{3}{5} & \frac{2}{5} & -\frac{2}{5} \\ \frac{3}{5} & -\frac{3}{5} & -\frac{3}{5} & \frac{3}{5} \\ -\frac{2}{5} & \frac{2}{5} & \frac{2}{5} & -\frac{2}{5} \\ -\frac{2}{5} & \frac{2}{5} & \frac{2}{5} & -\frac{2}{5} \\ -\frac{2}{5} & \frac{2}{5} & -\frac{3}{5} & \frac{3}{5} \end{pmatrix}
$$

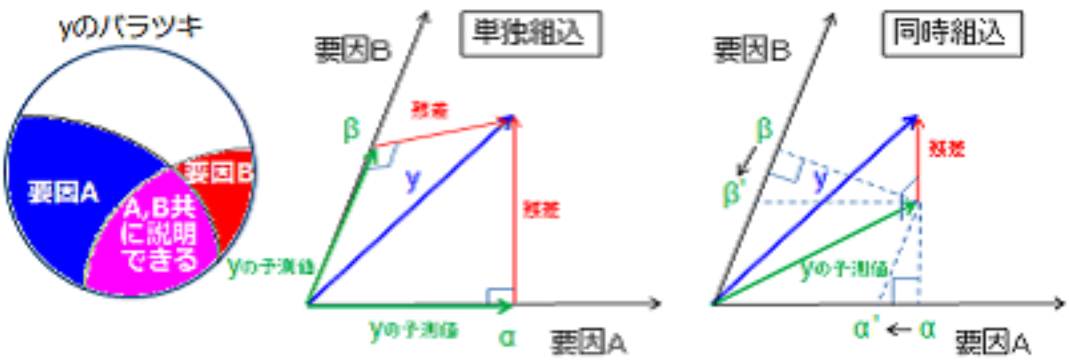
$\tilde{X}_{A,B}$ の”A に関する列ベクトル 対 B に関する列ベクトル”という全ての組み合わせについて内積が 0，つまり $A_1 \cdot B_1 = 0$，$A_1 \cdot B_2 = 0$，$A_2 \cdot B_1 = 0$ および $A_2 \cdot B_2 = 0$ になれば，要因 A と要因 B は直交していることになる[11]．このモデルでは $\tilde{X}'_{A,B}\tilde{X}_{A,B}$ が以下のような行列になれば直交となる．

$$
\tilde{X}'_{A,B}\tilde{X}_{A,B} = \begin{array}{c} \\ A_1 \\ A_2 \\ B_1 \\ B_2 \end{array}
\begin{array}{c} \begin{array}{cccc} A_1 & A_2 & B_1 & B_2 \end{array} \\
\begin{pmatrix} \tilde{x}_{11} & \tilde{x}_{12} & 0 & 0 \\ \tilde{x}_{21} & \tilde{x}_{22} & 0 & 0 \\ 0 & 0 & \tilde{x}_{33} & \tilde{x}_{34} \\ 0 & 0 & \tilde{x}_{43} & \tilde{x}_{44} \end{pmatrix} \end{array}
$$

しかしながら，今回のデータで計算すると以下のようになり，要因 A と要因 B は直交していないことが分かる[12]．

$$
\tilde{X}'_{A,B}\tilde{X}_{A,B} = \begin{array}{c} \\ A_1 \\ A_2 \\ B_1 \\ B_2 \end{array}
\begin{array}{c} \begin{array}{cccc} A_1 & A_2 & B_1 & B_2 \end{array} \\
\begin{pmatrix} \frac{6}{5} & -\frac{6}{5} & -\frac{1}{5} & \frac{1}{5} \\ -\frac{6}{5} & \frac{6}{5} & \frac{1}{5} & -\frac{1}{5} \\ -\frac{1}{5} & \frac{1}{5} & \frac{6}{5} & -\frac{6}{5} \\ \frac{1}{5} & -\frac{1}{5} & -\frac{6}{5} & \frac{6}{5} \end{pmatrix} \end{array}
$$

直交していないと，モデルに要因 A だけを組み込んだ場合とモデルに要因 A と要因 B の両方を組み込んだ場合を考えたとき，要因 A に対するモデルパラメータの推定値が同じ値にならない．



[11]基準化した列ベクトルの対の内積が 0 になるということは，$\tilde{X}_{A,B}$ に基準化する前の列ベクトルの対の共分散が 0 になることと同じである．

[12]バランスデータのときは直交する．アンバランスであっても要因が直交する場合がある．

5.2 TypeII の計算

ここでもモデルに切片は既にあるものとして話を進める．TypeII では，要因間において，要因 A しか説明できない部分を要因 A の説明量として，要因 B しか説明できない部分を要因 B の説明量として評価する．要因の平方和計算を行う際に交互作用は考慮しない．交互作用項の平方和計算は TypeI の場合と同じである．つまり，以下のように計算する．

$$
\begin{cases}
SS_A & \triangleq R(\alpha|\mu,\beta) = R(\mu,\alpha,\beta) - R(\mu,\beta) = SSR_{\mathrm{M3}} - SSR_{\mathrm{M2}} = 133\frac{5}{7} - 120 = 13\frac{5}{7} \\
SS_B & \triangleq R(\beta|\mu,\alpha) = R(\mu,\alpha,\beta) - R(\mu,\alpha) = SSR_{\mathrm{M3}} - SSR_{\mathrm{M1}} = 133\frac{5}{7} - 30 = 103\frac{5}{7} \\
SS_{A*B} & \triangleq R(\gamma|\mu,\alpha,\beta) = R(\mu,\alpha,\beta,\gamma) - R(\mu,\alpha,\beta) = SSR_{\mathrm{M4}} - SSR_{\mathrm{M3}} = 198 - 133\frac{5}{7} = 64\frac{2}{7}
\end{cases}
$$

となり，M4 モデルにおける glm の出力と一致する．イメージは図 4 のようになる．

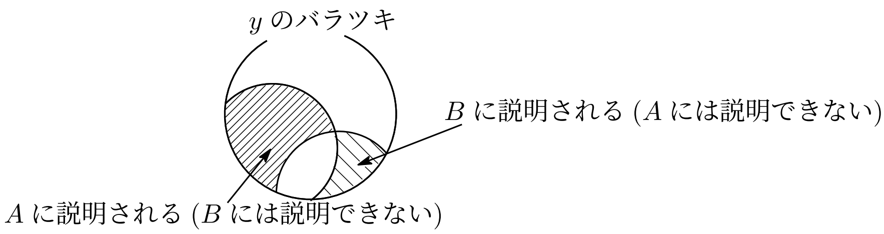


図 4:TypeII 平方和のイメージ

これより，要因の投入順序には依存しないことが分かる．また，図 3 の要因 A と要因 B が直交する場合を思い出すと，要因が直交する場合は TypeI と TypeII が同じ結果になり，図 2 または図 4 のように要因 A と要因 B が直交しない場合は同じ結果にならないことが分かる．

5.3 TypeIII の計算

TypeIII は TypeII と似ているが，要因の平方和計算を行う際に交互作用も考慮する点が TypeII と異なる．これだけを見ると，TypeII のときと同様に，例えば要因 A の平方和については $R(\alpha|\mu,\beta,\gamma)$=$R(\mu,\alpha,\beta,\gamma) - R(\mu,\beta,\gamma)$ を計算すれば良いのであるが，4.2 節の結果からでは計算できない．4.2 節の結果を使って $SSR_{\mathrm{M4}} - SSR_{\mathrm{M5}}$ を計算すると，$198 - 198 = 0$ となる．これは，デザイン行列の交互作用項に関する列ベクトルの線形結合で要因 A に関する列ベクトルが表せてしまうことが原因である[13]．つまり，デザイン行列 X の各列ベクトルは基底でない．これに対処するために，要因に対する制約条件 ($\sum_{i=1}^{2}\alpha_i = 0$ および $\sum_{j=1}^{2}\beta_j = 0$) だけでなく，交互作用に関して制約条件 ($\sum_{i=1}^{2}\gamma_{ij} = 0$ および $\sum_{j=1}^{2}\gamma_{ij} = 0$) を課して，各列ベクトルが基底となるデザイン行列を作成して計算する．4.2 節の出力で，元のデザイン行列 X は

$$
\begin{array}{ccccccccc}
V_\mu & V_{\alpha_1} & V_{\alpha_2} & V_{\beta_1} & V_{\beta_2} & V_{\gamma_{11}} & V_{\gamma_{12}} & V_{\gamma_{21}} & V_{\gamma_{22}}
\end{array}
$$
$$
\begin{pmatrix}
1 & 1 & 0 & 1 & 0 & 1 & 0 & 0 & 0 \\
1 & 1 & 0 & 0 & 1 & 0 & 1 & 0 & 0 \\
1 & 0 & 1 & 1 & 0 & 0 & 0 & 1 & 0 \\
1 & 0 & 1 & 1 & 0 & 0 & 0 & 1 & 0 \\
1 & 0 & 1 & 0 & 1 & 0 & 0 & 0 & 1
\end{pmatrix}
$$

であることが分かる．これに制約条件 $\alpha_1 + \alpha_2 = 0$(つまり $\alpha_2 = -\alpha_1$)，$\beta_1 + \beta_2 = 0$(つまり $\beta_2 = -\beta_1$)，$\gamma_{11} + \gamma_{12} = 0$，$\gamma_{11} + \gamma_{21} = 0$，$\gamma_{21} + \gamma_{22} = 0$ および $\gamma_{12} + \gamma_{22} = 0$(つまり $\gamma_{12} = \gamma_{21} = -\gamma_{11}$ および $\gamma_{11} = \gamma_{22}$) を用いてデザイン行列 ($X_r$ と書く) を作成すると

$$
X_r = \begin{array}{c}
\begin{array}{cccc} V_\mu & V_\alpha & V_\beta & V_\gamma \end{array} \\
\begin{pmatrix}
1 & 1 & 1 & 1 \\
1 & 1 & -1 & -1 \\
1 & -1 & 1 & -1 \\
1 & -1 & 1 & -1 \\
1 & -1 & -1 & 1
\end{pmatrix}
\end{array}
$$

となる．ベクトル V_μ，V_α，V_β および V_γ は基底となるので $R(\mu,\beta,\gamma)$ が計算できるようになる．ここからの計算は iml プロシジャ，data ステップあるいは手計算で確認する必要がある．以降，行列 X_r により計算された平方和およびその差分については $R_r(\cdot)$ および $R_r(\cdot|\cdot)$ と書くことにする．y の予測値 $\hat{y} = X_r(X_r'X_r)^{-}X_r'y$ を計算して，$R_r(\mu,\alpha,\beta,\gamma)$=$y'y - (y-\hat{y})'(y-\hat{y}) = 378$ を得る[14]．$R_r(\mu,\beta,\gamma)$ は行列 X_r の代わりに X_r から列ベクトル V_α を除いた行列 $X_{r|\alpha}$(列ベクトルが V_μ，V_β および V_γ のみ)

$$
X_{r|\alpha} = \begin{array}{c}
\begin{array}{ccc} V_\mu & V_\beta & V_\gamma \end{array} \\
\begin{pmatrix}
1 & 1 & 1 \\
1 & -1 & -1 \\
1 & 1 & -1 \\
1 & 1 & -1 \\
1 & -1 & 1
\end{pmatrix}
\end{array}
$$

で同様に計算して，$R_r(\mu,\beta,\gamma) = 354\frac{6}{7}$ を得る．同様に $R_r(\mu,\alpha,\gamma)$ は X_r から列ベクトル V_β を除いて計算し，$R_r(\mu,\alpha,\gamma) = 252$ を得る．また，$R_r(\mu,\alpha,\beta)$ は X_r から列ベクトル V_γ を除いて計算し，$R_r(\mu,\alpha,\beta) = 313\frac{5}{7}$ を得る[15]．よって，平方和は以下のように計算でき，M4 モデルにおける glm の出力と一致する[16]．

$$
\begin{cases}
SS_A & \triangleq R_r(\alpha|\mu,\beta,\gamma) = R_r(\mu,\alpha,\beta,\gamma) - R_r(\mu,\beta,\gamma) = 378 - 354\frac{6}{7} = 23\frac{1}{7} \\
SS_B & \triangleq R_r(\beta|\mu,\alpha,\gamma) = R_r(\mu,\alpha,\beta,\gamma) - R_r(\mu,\alpha,\gamma) = 378 - 252 = 126 \\
SS_{A*B} & \triangleq R_r(\gamma|\mu,\alpha,\beta) = R_r(\mu,\alpha,\beta,\gamma) - R_r(\mu,\alpha,\beta) = 378 - 313\frac{5}{7} = 64\frac{2}{7}
\end{cases}
$$

---

[13] V_{α_1} =$\mathrm{V}_{\gamma_{11}} + V_{\gamma_{12}}$ かつ V_{α_2} =$\mathrm{V}_{\gamma_{21}} + V_{\gamma_{22}}$ であり，M4 モデルの予測値と (M4 モデルから要因 A を除いた)M5 モデルの予測値が同じ値になってしまう．

[14] 4.2 節で見た M4 モデルの平均調整しないモデル平方和 $R(\mu,\alpha,\beta,\gamma)$ と一致する．

[15] 4.2 節で見た M3 モデルの平均調整しないモデル平方和 $R(\mu,\alpha,\beta)$ と一致する．

[16] ここでは書かないが X_r で TypeI，TypeII についても計算してみると結果は一致する．

5.3.1 バランスデータの場合

今回扱っているデータはバランスデータではない．しかしここで各セルに 2 つずつ値をもつバランスデータを例にして考えてみる．この場合のデザイン行列 X は

$$\begin{array}{c} \begin{array}{ccccccccc} V_\mu & V_{\alpha_1} & V_{\alpha_2} & V_{\beta_1} & V_{\beta_2} & V_{\gamma_{11}} & V_{\gamma_{12}} & V_{\gamma_{21}} & V_{\gamma_{22}} \end{array} \\ \begin{pmatrix} 1 & 1 & 0 & 1 & 0 & 1 & 0 & 0 & 0 \\ 1 & 1 & 0 & 1 & 0 & 1 & 0 & 0 & 0 \\ 1 & 1 & 0 & 0 & 1 & 0 & 1 & 0 & 0 \\ 1 & 1 & 0 & 0 & 1 & 0 & 1 & 0 & 0 \\ 1 & 0 & 1 & 1 & 0 & 0 & 0 & 1 & 0 \\ 1 & 0 & 1 & 1 & 0 & 0 & 0 & 1 & 0 \\ 1 & 0 & 1 & 0 & 1 & 0 & 0 & 0 & 1 \\ 1 & 0 & 1 & 0 & 1 & 0 & 0 & 0 & 1 \end{pmatrix} \end{array}$$

となる．これに制約条件を入れてデザイン行列 X_r を作成すると

$$\begin{array}{c} \begin{array}{cccc} V_\mu & V_\alpha & V_\beta & V_\gamma \end{array} \\ \begin{pmatrix} 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 \\ 1 & 1 & -1 & -1 \\ 1 & 1 & -1 & -1 \\ 1 & -1 & 1 & -1 \\ 1 & -1 & 1 & -1 \\ 1 & -1 & -1 & 1 \\ 1 & -1 & -1 & 1 \end{pmatrix} \end{array}$$

となり，$V_\mu \perp V_\alpha$，$V_\mu \perp V_\beta$，$V_\mu \perp V_\gamma$，$V_\alpha \perp V_\beta$，$V_\alpha \perp V_\gamma$ および $V_\beta \perp V_\gamma$，つまり直交基底となっている．よって，5.1 節の要因が直交する場合のところで述べたように，例えば,「切片および要因 A だけをモデルに含めた場合のモデル平方和」と「切片，要因および交互作用項を全て含めたモデルのモデル平方和 − 全て含めたモデルから要因 A だけを除いたモデルのモデル平方和」が一致する．要因 B についても同様である．つまり，バランスデータの場合には TypeI〜III の結果が全て一致する．

また，要因が直交していてもアンバランスなデータの場合には，要因が直交するので TypeI と TypeII の結果は一致する．しかし，モデルに交互作用項まで含めた場合は，交互作用に関する列ベクトル V_γ に対して直交しないような要因に関する列ベクトルが存在するため，TypeII と TypeIII の結果が一致しない．

5.3.2 制約条件を変更すると

セルの重み (オブザベーションの数) を考慮した以下のような制約条件に変更してみる．

- $2\alpha_1 + 3\alpha_2 = 0$（つまり $\alpha_2 = -\frac{2}{3}\alpha_1$）
- $3\beta_1 + 2\beta_2 = 0$（つまり $\beta_2 = -\frac{3}{2}\beta_1$）
- $\gamma_{11} + \gamma_{12} = 0$，$\gamma_{11} + 2\gamma_{21} = 0$，$2\gamma_{21} + \gamma_{22} = 0$ および $\gamma_{12} + \gamma_{22} = 0$
（つまり $\gamma_{12} = -\gamma_{11}$，$\gamma_{21} = -\frac{1}{2}\gamma_{11}$ および $\gamma_{22} = \gamma_{11}$）

このとき，

$$
X_r = \begin{array}{c} \begin{array}{cccc} V_\mu & V_\alpha & V_\beta & V_\gamma \end{array} \\ \begin{pmatrix} 1 & 1 & 1 & 1 \\ 1 & 1 & -\frac{3}{2} & -1 \\ 1 & -\frac{2}{3} & 1 & -\frac{1}{2} \\ 1 & -\frac{2}{3} & 1 & -\frac{1}{2} \\ 1 & -\frac{2}{3} & -\frac{3}{2} & 1 \end{pmatrix} \end{array}
$$

となるので，$V_\mu \perp V_\gamma$，$V_\alpha \perp V_\gamma$ および $V_\beta \perp V_\gamma$，つまり交互作用に関するベクトルは切片及び各要因のベクトルのいずれとも直交するので，要因の計算の際に交互作用を考慮しても，(要因計算の際に交互作用を考慮しない)TypeII と同じ結果となることが分かる．実は要因に関する制約条件の方は $\alpha_1 + \alpha_2 = 0$，$\beta_1 + \beta_2 = 0$ としても結果は変わらないので，交互作用に関する制約条件が TypeII と TypeIII の違いと考える．つまり，交互作用にセルの重みの違いを考慮した制約条件を課した結果が TypeII であり，セルの重みを同じとした制約条件を課した結果が TypeIII であると考える．ここで 2×2 以外の例についてもこれを確認し，さらに 2×2 の特別な場合を例示する．

5.3.2.1 (2×3) の例

A を 2 水準，B を 3 水準として，各セルの重み (オブザベーション数) が以下の場合を考える．

	B_1	B_2	B_3
A_1	1	2	3
A_2	3	2	1

交互作用にセルの重みを考慮した制約条件 ($\gamma_{11}+2\gamma_{12}+3\gamma_{13}=0$，$3\gamma_{21}+2\gamma_{22}+\gamma_{23}=0$，$\gamma_{11}+3\gamma_{21}=0$，$2\gamma_{12}+2\gamma_{22}=0$ および $3\gamma_{13}+\gamma_{23}=0$) を課すとデザイン行列は以下のようになり，$V_\mu \perp V_{\gamma_i}$，$V_\alpha \perp V_{\gamma_i}$，$V_{\beta_1} \perp V_{\gamma_i}$ および $V_{\beta_2} \perp V_{\gamma_i}(i=1,2)$ となるので TypeII の結果と一致する．

$$
X_r = \begin{array}{c}
\begin{array}{cccccc} V_\mu & V_\alpha & V_{\beta_1} & V_{\beta_2} & V_{\gamma_1} & V_{\gamma_2} \end{array} \\
\begin{pmatrix}
1 & 1 & 1 & 0 & 1 & 0 \\
1 & 1 & 0 & 1 & 0 & 1 \\
1 & 1 & 0 & 1 & 0 & 1 \\
1 & 1 & -1 & -1 & -\frac{1}{3} & -\frac{2}{3} \\
1 & 1 & -1 & -1 & -\frac{1}{3} & -\frac{2}{3} \\
1 & 1 & -1 & -1 & -\frac{1}{3} & -\frac{2}{3} \\
1 & -1 & 1 & 0 & -\frac{1}{3} & 0 \\
1 & -1 & 1 & 0 & -\frac{1}{3} & 0 \\
1 & -1 & 1 & 0 & -\frac{1}{3} & 0 \\
1 & -1 & 0 & 1 & 0 & -1 \\
1 & -1 & 0 & 1 & 0 & -1 \\
1 & -1 & -1 & -1 & 1 & 2
\end{pmatrix}
\end{array}
$$

次に，交互作用に制約条件 ($\gamma_{11}+\gamma_{12}+\gamma_{13}=0$，$\gamma_{21}+\gamma_{22}+\gamma_{23}=0$，$\gamma_{11}+\gamma_{21}=0$，$\gamma_{12}+\gamma_{22}=0$ および $\gamma_{13}+\gamma_{23}=0$) を課すとデザイン行列は以下のようになり，TypeIII の結果と一致する．

$$
X_r = \begin{array}{c}
\begin{array}{cccccc} V_\mu & V_\alpha & V_{\beta_1} & V_{\beta_2} & V_{\gamma_1} & V_{\gamma_2} \end{array} \\
\begin{pmatrix}
1 & 1 & 1 & 0 & 1 & 0 \\
1 & 1 & 0 & 1 & 0 & 1 \\
1 & 1 & 0 & 1 & 0 & 1 \\
1 & 1 & -1 & -1 & -1 & -1 \\
1 & 1 & -1 & -1 & -1 & -1 \\
1 & 1 & -1 & -1 & -1 & -1 \\
1 & -1 & 1 & 0 & -1 & 0 \\
1 & -1 & 1 & 0 & -1 & 0 \\
1 & -1 & 1 & 0 & -1 & 0 \\
1 & -1 & 0 & 1 & 0 & -1 \\
1 & -1 & 0 & 1 & 0 & -1 \\
1 & -1 & -1 & -1 & 1 & 1
\end{pmatrix}
\end{array}
$$

要因に関する制約条件を変更しても結果は変わらない．

5.3.2.2 (3×3) の例

A を 3 水準，B を 3 水準として，各セルの重み (オブザベーション数) が以下の場合を考える．

	B_1	B_2	B_3
A_1	1	1	1
A_2	1	2	1
A_3	1	1	1

交互作用にセルの重みを考慮した制約条件 ($\gamma_{11}+\gamma_{12}+\gamma_{13}=0$，$\gamma_{21}+2\gamma_{22}+\gamma_{23}=0$，$\gamma_{31}+\gamma_{32}+\gamma_{33}=0$，$\gamma_{11}+\gamma_{21}+\gamma_{31}=0$，$\gamma_{12}+2\gamma_{22}+\gamma_{32}=0$ および $\gamma_{13}+\gamma_{23}+\gamma_{33}=0$) を課すとデザイン行列は以下のようになり，$V_\mu \perp V_{\gamma_i}$，$V_{\alpha_1} \perp V_{\gamma_i}$，$V_{\alpha_2} \perp V_{\gamma_i}$，$V_{\beta_1} \perp V_{\gamma_i}$ および $V_{\beta_2} \perp V_{\gamma_i}$ $(i=1,2,3,4)$ となるので TypeII の結果と一致する．

$$
x_r = \begin{array}{c} \begin{array}{ccccccccc} V_\mu & V_{\alpha_1} & V_{\alpha_2} & V_{\beta_1} & V_{\beta_2} & V_{\gamma_1} & V_{\gamma_2} & V_{\gamma_3} & V_{\gamma_4} \end{array} \\ \begin{pmatrix}
1 & 1 & 0 & 1 & 0 & 1 & 0 & 0 & 0 \\
1 & 1 & 0 & 0 & 1 & 0 & 1 & 0 & 0 \\
1 & 1 & 0 & -1 & -\frac{4}{3} & -1 & -1 & 0 & 0 \\
1 & 0 & 1 & 1 & 0 & 0 & 0 & 1 & 0 \\
1 & 0 & 1 & 0 & 1 & 0 & 0 & 0 & 1 \\
1 & 0 & 1 & 0 & 1 & 0 & 0 & 0 & 1 \\
1 & 0 & 1 & -1 & -\frac{4}{3} & 0 & 0 & -1 & -2 \\
1 & -1 & -\frac{4}{3} & 1 & 0 & -1 & 0 & -1 & 0 \\
1 & -1 & -\frac{4}{3} & 0 & 1 & 0 & -1 & 0 & -2 \\
1 & -1 & -\frac{4}{3} & -1 & -\frac{4}{3} & 1 & 1 & 1 & 2
\end{pmatrix} \end{array}
$$

次に，交互作用に制約条件 ($\gamma_{11}+\gamma_{12}+\gamma_{13}=0$，$\gamma_{21}+\gamma_{22}+\gamma_{23}=0$，$\gamma_{31}+\gamma_{32}+\gamma_{33}=0$，$\gamma_{11}+\gamma_{21}+\gamma_{31}=0$，$\gamma_{12}+\gamma_{22}+\gamma_{32}=0$ および $\gamma_{13}+\gamma_{23}+\gamma_{33}=0$) を課すとデザイン行列は以下のようになり，TypeIII の結果と一致する．

$$
x_r = \begin{array}{c} \begin{array}{ccccccccc} V_\mu & V_{\alpha_1} & V_{\alpha_2} & V_{\beta_1} & V_{\beta_2} & V_{\gamma_1} & V_{\gamma_2} & V_{\gamma_3} & V_{\gamma_4} \end{array} \\ \begin{pmatrix}
1 & 1 & 0 & 1 & 0 & 1 & 0 & 0 & 0 \\
1 & 1 & 0 & 0 & 1 & 0 & 1 & 0 & 0 \\
1 & 1 & 0 & -1 & -1 & -1 & -1 & 0 & 0 \\
1 & 0 & 1 & 1 & 0 & 0 & 0 & 1 & 0 \\
1 & 0 & 1 & 0 & 1 & 0 & 0 & 0 & 1 \\
1 & 0 & 1 & 0 & 1 & 0 & 0 & 0 & 1 \\
1 & 0 & 1 & -1 & -1 & 0 & 0 & -1 & -1 \\
1 & -1 & -1 & 1 & 0 & -1 & 0 & -1 & 0 \\
1 & -1 & -1 & 0 & 1 & 0 & -1 & 0 & -1 \\
1 & -1 & -1 & -1 & -1 & 1 & 1 & 1 & 1
\end{pmatrix} \end{array}
$$

要因に関する制約条件を変更しても結果は変わらない．

5.3.2.3 (2×2) の特別な場合

セルの重みが以下のような場合 $(m \neq n)$,

	B_1	B_2
A_1	m	n
A_2	n	m

このとき，各要因は直交しないので TypeI と TypeII は異なる．しかし，交互作用の制約条件について以下の 2 つを考えてみる．

1. $\gamma_{11}+\gamma_{12}=0$，$\gamma_{11}+\gamma_{21}=0$，$\gamma_{21}+\gamma_{22}=0$ および $\gamma_{12}+\gamma_{22}=0$
2. $m\gamma_{11}+n\gamma_{12}=0$，$m\gamma_{11}+n\gamma_{21}=0$，$n\gamma_{21}+m\gamma_{22}=0$ および $n\gamma_{12}+m\gamma_{22}=0$
   (2. はセルの重みを考慮している)

簡単にするために $m=1$，$n=2$ とすると，今両者の X_r はそれぞれ

$$
\begin{array}{c}
\begin{array}{cccc} V_\mu & V_\alpha & V_\beta & V_\gamma \end{array} \\
\begin{pmatrix}
1 & 1 & 1 & 1 \\
1 & 1 & -1 & -1 \\
1 & 1 & -1 & -1 \\
1 & -1 & 1 & -1 \\
1 & -1 & 1 & -1 \\
1 & -1 & -1 & 1
\end{pmatrix}
\end{array},
\quad
\begin{array}{c}
\begin{array}{cccc} V_\mu & V_\alpha & V_\beta & V_\gamma \end{array} \\
\begin{pmatrix}
1 & 1 & 1 & 1 \\
1 & 1 & -1 & -\frac{1}{2} \\
1 & 1 & -1 & -\frac{1}{2} \\
1 & -1 & 1 & -\frac{1}{2} \\
1 & -1 & 1 & -\frac{1}{2} \\
1 & -1 & -1 & 1
\end{pmatrix}
\end{array}
$$

となり，一見違うようだが両者に対する計算結果が同じになる．つまり，アンバランスかつ要因が直交せず，モデルに交互作用まで含めた場合でも，TypeII と TypeIII が同じになる．

5.4 共通点および相違点のまとめ

バランスデータであれば I，II，および III は全て一致する．アンバランスだが要因が直交する場合は，I と II は一致するが交互作用をモデルに含めた場合は III と一致しない．また，交互作用をモデルに含めない場合は，どのようなデータに対してでも II と III は一致する．今回のように，アンバランスかつ要因が直交せず，モデルに交互作用まで含めた場合は I，II および III は一般的には全て異なる．しかし，交互作用項 $A*B$ の平方和については，交互作用まで含めたモデルの平方和と交互作用を含めないモデルの平方和の差分で定義され I，II，III 全て一致する．

6 SAS Technical Report の計算方法

参考までに SAS Technical Report R-101 の計算方法を紹介する．平方和は $(Lb)'(L(X'X)^{-}L')^{-1}(Lb)$ で計算される．ここで L は推定可能関数，b は正規方程式の解，そして X はデザイン行列である．全ての平方和の計算に対して

$$b = \begin{pmatrix} 6 \\ 12 \\ 0 \\ -3 \\ 0 \\ -15 \\ 0 \\ 0 \\ 0 \end{pmatrix}, \quad X = \begin{array}{c} \begin{array}{ccccccccc} \mu & A_1 & A_2 & B_1 & B_2 & A_1B_1 & A_1B_2 & A_2B_1 & A_2B_2 \end{array} \\ \begin{pmatrix} 1 & 1 & 0 & 1 & 0 & 1 & 0 & 0 & 0 \\ 1 & 1 & 0 & 0 & 1 & 0 & 1 & 0 & 0 \\ 1 & 0 & 1 & 1 & 0 & 0 & 0 & 1 & 0 \\ 1 & 0 & 1 & 1 & 0 & 0 & 0 & 1 & 0 \\ 1 & 0 & 1 & 0 & 1 & 0 & 0 & 0 & 1 \end{pmatrix} \end{array}$$

および

$$X'X = \begin{pmatrix} 5 & 2 & 3 & 3 & 2 & 1 & 1 & 2 & 1 \\ 2 & 2 & 0 & 1 & 1 & 1 & 1 & 0 & 0 \\ 3 & 0 & 3 & 2 & 1 & 0 & 0 & 2 & 1 \\ 3 & 1 & 2 & 3 & 0 & 1 & 0 & 2 & 0 \\ 2 & 1 & 1 & 0 & 2 & 0 & 1 & 0 & 1 \\ 1 & 1 & 0 & 1 & 0 & 1 & 0 & 0 & 0 \\ 1 & 1 & 0 & 0 & 1 & 0 & 1 & 0 & 0 \\ 2 & 0 & 2 & 2 & 0 & 0 & 0 & 2 & 0 \\ 1 & 0 & 1 & 0 & 1 & 0 & 0 & 0 & 1 \end{pmatrix}, \quad (X'X)^{-} = \begin{pmatrix} 1 & -1 & 0 & -1 & 0 & 1 & 0 & 0 & 0 \\ -1 & 2 & 0 & 1 & 0 & -2 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ -1 & 1 & 0 & 1.5 & 0 & -1.5 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 1 & -2 & 0 & -1.5 & 0 & 3.5 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{pmatrix}$$

は共通であり，L だけが異なる．L 以外は 4.2 節の出力結果で全て確認できる．

6.1 TypeI の計算

$X'X$ の上の行から順に (つまりモデルに組み込む μ，A，B，$A*B$ の順に) 対角成分を 1 にして，同じ列のそれよりも下の行の成分が 0(非 0 の対角成分が 1 となっている上三角行列) になるように，行による変換 (例えば最初の処理は 1 行目=1 行目 $\times\frac{1}{5}$，2 番目の処理は 2 行目=2 行目+(-2)×1 行目，etc...) を行うと，

$$
\begin{array}{cc}
 & \begin{array}{ccccccccc} \mu & A_1 & A_2 & B_1 & B_2 & A_1B_1 & A_1B_2 & A_2B_1 & A_2B_2 \end{array} \\
\begin{array}{c} \mu \\ A_1 \\ A_2 \\ B_1 \\ B_1 \\ A_1B_1 \\ A_1B_2 \\ A_2B_1 \\ A_2B_2 \end{array} &
\begin{pmatrix}
5 & 2 & 3 & 3 & 2 & 1 & 1 & 2 & 1 \\
2 & 2 & 0 & 1 & 1 & 1 & 1 & 0 & 0 \\
3 & 0 & 3 & 2 & 1 & 0 & 0 & 2 & 1 \\
3 & 1 & 2 & 3 & 0 & 1 & 0 & 2 & 0 \\
2 & 1 & 1 & 0 & 2 & 0 & 1 & 0 & 1 \\
1 & 1 & 0 & 1 & 0 & 1 & 0 & 0 & 0 \\
1 & 1 & 0 & 0 & 1 & 0 & 1 & 0 & 0 \\
2 & 0 & 2 & 2 & 0 & 0 & 0 & 2 & 0 \\
1 & 0 & 1 & 0 & 1 & 0 & 0 & 0 & 1
\end{pmatrix}
\end{array}
\rightarrow
\begin{array}{c}
\begin{array}{ccccccccc} \mu & A_1 & A_2 & B_1 & B_2 & A_1B_1 & A_1B_2 & A_2B_1 & A_2B_2 \end{array} \\
\begin{pmatrix}
1 & \frac{2}{5} & \frac{3}{5} & \frac{3}{5} & \frac{2}{5} & \frac{1}{5} & \frac{1}{5} & \frac{2}{5} & \frac{1}{5} \\
0 & 1 & -1 & -\frac{1}{6} & \frac{1}{6} & \frac{1}{2} & \frac{1}{2} & -\frac{2}{3} & -\frac{1}{3} \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 1 & -1 & \frac{3}{7} & -\frac{3}{7} & \frac{4}{7} & -\frac{4}{7} \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 1 & -1 & -1 & 1 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0
\end{pmatrix}
\end{array}
\tag{3}
$$

となる．変換後の行列の 2 行目が要因 A に関する L_α，4 行目が要因 B に関する L_β，そして 6 行目が交互作用項 $A*B$ に関する L_γ である[17]．これらの L により $(Lb)'(L(X'X)^-L')^{-1}(Lb)$ を計算すると，要因 A，要因 B および交互作用項 $A*B$ に対する TypeI 平方和 SS_A，SS_B および SS_{A*B} は

$$
\begin{cases}
SS_A & = (L_\alpha b)'(L_\alpha(X'X)^-L'_\alpha)^{-1}(L_\alpha b) = 30 \\
SS_B & = (L_\beta b)'(L_\beta(X'X)^-L'_\beta)^{-1}(L_\beta b) = 103\dfrac{5}{7} \\
SS_{A*B} & = (L_\gamma b)'(L_\gamma(X'X)^-L'_\gamma)^{-1}(L_\gamma b) = 64\dfrac{2}{7}
\end{cases}
$$

となり，glm の出力と一致する．また，この場合は

$$
SS_A + SS_B + SS_{A*B} = 30 + 103\frac{5}{7} + 64\frac{2}{7} = 198 \tag{4}
$$

で各メンバーの平方和の合計が平均調整済みモデル平方和と一致していることが分かる[18]．

[17]glm プロシジャの model ステートメントで e1 オプションを記述すると，これらの L が出力される．
[18]各メンバーの平方和の合計がモデル平方和と一致するのは，一般的には TypeI のみである．

6.2 TypeII の計算

TypeI 計算の際に用いた (3) 式の右側の行列を用いる．TypeI の L_α は B_1 列および B_2 列に対する値が $-\frac{1}{6}$ および $\frac{1}{6}$ であり 0 ではない．ここが 0 になるように，2 行目=2 行目+$(\frac{1}{6})\times$4 行目の変換を行うと，

$$\begin{array}{c|c} & \begin{array}{ccccccccc}\mu & A_1 & A_2 & B_1 & B_2 & A_1B_1 & A_1B_2 & A_2B_1 & A_2B_2\end{array} \\ \begin{array}{c}\mu \\ A_1 \\ A_2 \\ B_1 \\ B_2 \\ A_1B_1 \\ A_1B_2 \\ A_2B_1 \\ A_2B_2\end{array} & \begin{pmatrix} 1 & \frac{2}{5} & \frac{3}{5} & \frac{3}{5} & \frac{2}{5} & \frac{1}{5} & \frac{1}{5} & \frac{2}{5} & \frac{1}{5} \\ 0 & 1 & -1 & -\frac{1}{6} & \frac{1}{6} & \frac{1}{2} & \frac{1}{2} & -\frac{2}{3} & -\frac{1}{3} \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & -1 & \frac{3}{7} & -\frac{3}{7} & \frac{4}{7} & -\frac{4}{7} \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & -1 & -1 & 1 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{pmatrix} \end{array} \rightarrow \begin{array}{c} \begin{array}{ccccccccc}\mu & A_1 & A_2 & B_1 & B_2 & A_1B_1 & A_1B_2 & A_2B_1 & A_2B_2\end{array} \\ \begin{pmatrix} 1 & \frac{2}{5} & \frac{3}{5} & \frac{3}{5} & \frac{2}{5} & \frac{1}{5} & \frac{1}{5} & \frac{2}{5} & \frac{1}{5} \\ 0 & 1 & -1 & 0 & 0 & \frac{4}{7} & \frac{3}{7} & -\frac{4}{7} & -\frac{3}{7} \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & -1 & \frac{3}{7} & -\frac{3}{7} & \frac{4}{7} & -\frac{4}{7} \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & -1 & -1 & 1 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{pmatrix} \end{array} \tag{5}$$

となる．TypeI のときと同様に，変換後の行列の 2 行目が要因 A に関する L_α，4 行目が要因 B に関する L_β，そして 6 行目が交互作用項 $A * B$ に関する L_γ である[19]．これらの L により $(Lb)'(L(X'X)^{-}L')^{-1}(Lb)$ を計算すると，要因 A，要因 B および交互作用項 $A*B$ に対する TypeII 平方和 SS_A，SS_B および SS_{A*B} は

$$\begin{cases} SS_A &= (L_\alpha b)'(L_\alpha(X'X)^{-}L'_\alpha)^{-1}(L_\alpha b) = 13\dfrac{5}{7} \\ SS_B &= (L_\beta b)'(L_\beta(X'X)^{-}L'_\beta)^{-1}(L_\beta b) = 103\dfrac{5}{7} \\ SS_{A*B} &= (L_\gamma b)'(L_\gamma(X'X)^{-}L'_\gamma)^{-1}(L_\gamma b) = 64\dfrac{2}{7} \end{cases}$$

となり，glm の出力と一致する．また，この場合は

$$SS_A + SS_B + SS_{A*B} = 13\frac{5}{7} + 103\frac{5}{7} + 64\frac{2}{7} = 181\frac{5}{7} \neq 198 \tag{6}$$

で各メンバーの平方和の合計が平均調整済みモデル平方和と一致していないことが分かる．

[19]glm プロシジャの model ステートメントで e2 オプションを記述すると，これらの L が出力される．

6.3 TypeIII の計算

TypeIII では，TypeI や TypeII の場合のような $X'X$ 行列の変換は行わない．モデルを選択したときに決まっている．今回のモデルでの TypeIII に対する行列は，

$$\begin{array}{c} \\ \mu \\ A_1 \\ A_2 \\ B_1 \\ B_2 \\ A_1B_1 \\ A_1B_2 \\ A_2B_1 \\ A_2B_2 \end{array}
\begin{array}{c} \begin{array}{ccccccccc} \mu & A_1 & A_2 & B_1 & B_2 & A_1B_1 & A_1B_2 & A_2B_1 & A_2B_2 \end{array} \\
\begin{pmatrix}
1 & \frac{1}{2} & \frac{1}{2} & \frac{1}{2} & \frac{1}{2} & \frac{1}{4} & \frac{1}{4} & \frac{1}{4} & \frac{1}{4} \\
0 & 1 & -1 & 0 & 0 & \frac{1}{2} & \frac{1}{2} & -\frac{1}{2} & -\frac{1}{2} \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 1 & -1 & \frac{1}{2} & -\frac{1}{2} & \frac{1}{2} & -\frac{1}{2} \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 1 & -1 & -1 & 1 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0
\end{pmatrix} \end{array} \tag{7}$$

となる[20]．TypeI および TypeII のときと同様に，行列の 2 行目が要因 A に関する L_α，4 行目が要因 B に関する L_β，そして 6 行目が交互作用項 $A*B$ に関する L_γ である[21]．これらの L により $(Lb)'(L(X'X)^-L')^{-1}(Lb)$ を計算すると，要因 A，要因 B および交互作用項 $A*B$ に対する TypeIII 平方和 SS_A，SS_B および SS_{A*B} は

$$\begin{cases} SS_A &= (L_\alpha b)'(L_\alpha(X'X)^-L'_\alpha)^{-1}(L_\alpha b) = 23\dfrac{1}{7} \\ SS_B &= (L_\beta b)'(L_\beta(X'X)^-L'_\beta)^{-1}(L_\beta b) = 126 \\ SS_{A*B} &= (L_\gamma b)'(L_\gamma(X'X)^-L'_\gamma)^{-1}(L_\gamma b) = 64\dfrac{2}{7} \end{cases}$$

となり，glm の出力と一致する．また，この場合は

$$SS_A + SS_B + SS_{A*B} = 23\frac{1}{7} + 126 + 64\frac{2}{7} = 213\frac{3}{7} \neq 198 \tag{8}$$

で各メンバーの平方和の合計が平均調整済みモデル平方和と一致していないことが分かる．

[20]バランスデータのときの $X'X$ を変換するとこの行列になる．
[21]glm プロシジャの model ステートメントで e3 オプションを記述すると，これらの L が出力される．

7 おわりに

小データを用いて TypeI，TypeII および TypeIII の計算方法を例示した．今後もし機会があれば，TypeIV の計算例，Technical Report の計算方法との関連等について調べてみたい．

8 謝辞

本書作成にあたり，ご助言を頂きました 東レ株式会社 医薬開発推進室 土居正明様に深く感謝いたします．また，本書提出にあたり，レビューして頂きました皆様に深く感謝いたします．

参考文献

[1] http://support.sas.com/documentation/cdl/en/statug/63347/HTML/default/viewer.htm#introglmest_toc.htm

[2] 高橋行雄, 芳賀 敏郎, 大橋 靖雄.(1989). SAS による実験データの解析, 東京大学出版会.

[3] 土居正明.(2011). Fixed Effect model ８（２元配置：その６）, http://www012.upp.so-net.ne.jp/doi/math/anova/2way_fixed_6.pdf

[4] http://support.sas.com/documentation/onlinedoc/v82/techreport_r101.pdf

[5] Searle, S. R. (1987), *Linear Models for Unbalanced Data*, New York: John Wiley Sons, Inc.

[6] http://www.aichi-gakuin.ac.jp/~chino/anova/chapter1/sec1-3-7.html