JMP による各種分割実験入門
－変量効果モデルの基礎－

2011 年 12 月 8 日 v1.0

12 月 27 日 v1.1

高橋 行雄

12/27/2011 2:37:00 PM

表紙裏

2011 年 10 月 24 日，新規作成，11 月 29 日 v0.1　12 月 5 日 v0.2
2011 年 12 月 8 日 第 1 版,
ファイル名：F:¥Q3_2011¥Q3_JMP_分割実験_最終¥高橋_2011_分割実験の統計解析.doc

目　次

JMP による各種分割実験入門 －変量効果モデルの基礎－

1.　はじめに 1

第 2 版の改定の要旨 3

2.　様々なランダム化に対応する分散分析 5

2.1.　完全ランダム実験とみなした場合 5

完全にランダムの場合の分散分析 5

JMP による解析「モデルのあてはめ」 5

結果のグラフ化と解釈 7

多重性の配慮 9

2.2.　測定の繰り返し実験 10

測定の繰返しでの 12 回のランダム化 10

分散分析 11

JMP での解析手順と結果の読み方 11

EMS 法および REML 法のプロファイルの相違 13

変量効果と固定効果の使い分け 16

2.3.　反復実験 16

反復を伴うランダム化 16

変量効果としての反復 R 17

交互作用を含めた場合，含めない場合の平均平方の期待値 18

2.4.　1 段分割実験 19

枝分かれ構造でのランダム化 19

変動要因の可視化に役立つ変動性図／ゲージチャート 21

2.5.　反復を伴う 1 段分割実験 22

2.6.　2 方分割実験 24

ランダム化の更なる工夫 24

ランダム化が 4 回の場合の悩ましいデータ構造 26

2.7.　反復のある 2 方分割実験 27

3.　EMS（従来法）に変わる REML 法 31

3.1.　要因配置型の分割実験 31

EMS（従来法）の限界 31

分割実験に詳しい日本語の教科書 31

母平均の推定，95%信頼区間 33
JMP による母平均の 95%信頼区間の推定 35
3.2. 直交表を用いた分割実験 36
JMP での解析 39
交互作用の検討 42
最適水準の組み合わせ平均，95%信頼区間 43
4. 欠測値を含むクロスオーバー試験 45
4.1. 用量反応および食前食後を含む BE 試験 45
4.2. 直交表での分割実験と同じ構造 46
4.3. 90%信頼区間なども自在に 49
5. REML 法による分割実験の解析 50
5.1. 反復を伴う 1 段分割実験 50
データの構造 50
分散分析表 51
5.2. 固定効果の水準平均 52
プロファイル 52
因子 A の水準平均の推定 53
因子 A についての効果の差の推定 55
因子 A についての効果の *F* 検定 56
因子 B について 58
交互作用 A×B について 59
対比の応用 60
5.3. パラメータの推定値および予測値 60
パラメータの推定値 60
固定効果についての予測値 62
変量効果の予測値 63
変量効果の推定 64
5.4. 実験数の削減 65
6. スタッガード型枝分かれ実験 68
6.1. 完備型の枝分かれ実験の問題点 68
誤差の自由度のインフレーション 68
スタッカード型枝分れ実験 69

6.2. SAS および JMP による分散成分の推定 69
完備型データでの分散成分 69
スタッカード型データにでの分散成分 71
REML 法による分散成分の推定 73
SAS ユーザ会 2011 年のでの論文集に対する補足 75
6.3. 分散成分の推定方式（芳賀本の事例との対比） 76
6.4. 今後の課題 78
シミュレーション研究による分散成分の 95%信頼区間の妥当性 78
7. Excel ソルバーによる REML 法入門 79
7.1. Excel による追試の必要性 79
7.2. 最小 2 乗法による母平均の推定 80
最尤法による母平均と母分散の推定 81
REML（制限付き最尤法）による母平均と母分散の推定 83
7.3. REML（制限付き最尤法）による回帰分析のパラメータ推定 85
標準的な最小 2 乗法による回帰分析 85
最尤法による回帰分析 85
REML 法による回帰分析 87
7.4. 枝分かれ型のデータに対する REML 法 90
分散分析表から分散成分の推定 90
最尤法による分散成分の推定 90
REML 法による分散成分の推定 92
7.5. 繰返しを排除した REML 法 93
データ y についての尤度 93
固定効果のパラメータについての尤度 94
7.6. 繰返しが不揃いの場合の REML 法 95
7.7. 分散共分散行列の生成 96
8. 枝分れ実験における分散共分散構造 98
8.1. 複数の誤差構造 98
8.2. 枝分かれ実験データ 98
8.3. データ間の相関構造 99
8.4. 簡便な方法による相関構造の確認 101
8.5. 多次元正規分布の確率密度 103

9. Excel によるスタッカード型枝分かれ実験データの解析 105
9.1. 変量効果の推定 BLUP 108
9.2. 統計連合大会 2010 年での発表 抄録 110
9.3. 統計連合大会 2010 年での発表 スライド 111
10. Excel ソルバーによる反復のある 1 段分割実験の解析 115
11. ブロックに因子を割り付けたクロスオーバー法 117
11.1. JMP による解析 117
11.2. Excel による解析 118
12. 分割実験を活用した応答局面解析 120
12.1. トウモロコシの顆粒の製造実験 120
実験手順 120
実験結果 120
実験のランダム化の確認と反応のグラフ化 122
12.2. 山本達郎氏による解析事例 123
13. 欠測セルがある 2 方分割実験 127
13.1. 2 つの遺伝子型の小麦の収量 127
13.2. 要因配置実験とした解析 129
南北方向，東西方向 130
遺伝子の型の収量の差 130
13.3. 分割実験として解析 131
窒素肥料の量を名義尺度 131
13.4. 交互作用（東西×南北）を変量効果 133
窒素肥料の量を名義尺度 134
窒素肥料の量を連続量 135
13.5. 南北と東西を別々な変量効果 136
窒素肥料の量を名義尺度 136
窒素肥料の量を連続量 137
14. 経時データの REML 法による解析 139
14.1. 2 群 4 症例，2 時点の経時データ 139
14.2. Excel での計算 140
14.3. 症例を固定効果とし，投与前を群間で共通とした場合 147

14.4. 症例を変量効果とし，投与前を群間で共通とした場合 148
14.5. 症例を固定効果とした場合と変量効果とした場合の時期 1 での群間比較 149
14.6. 考察 150
14.7. 統計連合 2011 での発表 抄録 151
14.8. 統計連合 2011 での発表 スライド 152

15. 今後の課題 156

新しい流れ 158

図 表 目 次

図 2.1 完全ランダム実験の場合の 2 因子の交互作用プロット 7
図 2.2 水準間の差についての統計量（抜粋） 8
図 2.3 差の順位レポート（抜粋） 9
図 2.4 差についての詳細なレポート（抜粋） 9
図 2.5 Tukey の HSD 検定に基づく差の順位レポート（抜粋） 9
図 2.6 分散分析表と JMP の出力の対応付け 14
図 2.7 因子ごとの水準平均のプロファイル 15
図 2.8 水準の設定によるプロファイルの変化の比較 16
図 2.9 全ての因子の組み合わせによる変動性図 22
図 5.1 REML 法による解析 50
図 5.2 因子の水準の選択に伴うにプロファイルの変化 52
図 5.3 交互作用プロファイル 53
図 5.4 t 分布の 95%点 55
図 5.5 固定効果の予測式 63
図 6.1 繰返しが揃っている枝分かれ実験 69
図 6.2 スタッカード型枝分かれ実験 69
図 6.3 スタッカード型データに対する JMP の REML 法による分散成分の推定 76
図 7.1 Excel のソルバーを用いた最小 2 乗法による母平均の推定 80
図 7.2 ステップ 1 およびステップ 2 での偏差平方和 81
図 7.3 ステップ 4 で母平均を 2.5 からずらした場合の偏差平方和の変化 81
図 7.4 平均と分散の最尤法による同時推定 82
図 7.5 ステップ 2 およびステップ 3 でのデータごとの尤度 83
図 7.6 母平均と母分散を変化させた場合の尤度の同時分布 83
図 7.7 平均と分散の最尤法による同時推定 84
図 7.8 ステップ 1 およびステップ 2 でのデータと平均値の尤度 85
図 7.9 標準的な回帰分析 85
図 7.10 回帰分析での初期値に対する尤度の計算 86
図 7.11 初期値に対するデータ y_i の $\hat{y}_i$ に対する尤度 86
図 7.12 回帰分析での最尤解 87
図 7.13 最尤解に対するデータ y_i の $\hat{y}_i$ に対する尤度 87
図 7.14 回帰分析のデザイン行列 $\boldsymbol{X}$ を用いたパラメータの分散共分散 $\boldsymbol{\Sigma}$ の計算 88
図 7.15 REML 法による回帰分析（初期値） 89
図 7.16 REML 法による回帰分析（尤度の比を最大化） 89
図 7.17 繰返しが等しい場合の複数誤差の同時推定 92
図 7.18 繰返しが等しい場合の複数誤差の REML 法による同時推定 93
図 7.19 6 次元正規分布を用いたデータ y に関する尤度の計算 94

図 7.20　パラメータに関する分散共分散行列の計算 95
図 7.21　繰返しが不揃いの場合の REML 法による分散成分の計算 96
図 8.1　貨車ごとの硫黄含有率の変動 99
図 8.2　異なる試料間の散布図 100
図 8.3　試料間の散布図および確率楕円 103
図 9.1　行列計算による分散共分散行列の生成 107
図 9.2　REML 法によるスタッカード型の 2 段枝分かれ分散成分の推定 108
図 12.1　因子 A 内のバッチ別変動，4 因子の水準ごとの変動図 122
図 13.1　南北方向の収量に与える影響　　図 13.2　東西方向の収量に与える影響 128
図 13.3　窒素肥料の量と収量の関連 129
図 13.4　南北方向の収量の違い 130
図 13.5　遺伝子型の収量の差の変化 131
図 13.6　遺伝子型別の 2 次曲線のあてはめ 133
図 13.7　取り上げた因子のプロファイル 135
図 13.8　取り上げた因子のプロファイル 136
図 13.9　遺伝子型ごとの窒素肥料の量と収量の関連 137
図 13.10　対比による最大収量間の差の推定 137
図 13.11　遺伝子型ごとの窒素肥料の量と収量の関連 138

表 1.1　整理された実験データ 2
表 1.2　同一データセットに対する 7 種類の解析法 2
表 2.1　完全ランダム化の手順と実験データ 5
表 2.2　完全ランダム化の場合の分散分析表 5
表 2.3　測定の繰り返しがある実験データ 10
表 2.4　測定の繰り返しがある場合の分散分析表 11
表 2.5　反復実験でのランダム化 17
表 2.6　反復実験とした場合の分散分析表 18
表 2.7　交互作用項を含めた平均平方の期待値 19
表 2.8　1 段分割実験のデータ 20
表 2.9　1 段分割実験の場合の分散分析表 21
表 2.10　反復を伴う 1 段分割実験 23
表 2.11　反復のある 1 段分割実験の場合の分散分析表 23
表 2.12　2 法分割実験のデータ 24
表 2.13　2 方分割実験の分散分析表 26
表 2.14　悩ましいデータ 27
表 2.15　反復のある 2 方分割実験データ 27
表 2.16　反復を伴う 2 方分割（プーリング前） 30
表 2.17　反復を伴う 2 方分割（プーリング後） 30
表 3.1　乱塊法実験 39
表 5.1　反復を伴う 1 段分割実験 50
表 5.2　反復のある 1 段分割実験の REML 法での分散分析表 52
表 5.3　因子 A の水準平均の計算 61
表 5.4　因子 B の水準平均の計算 62
表 5.5　交互作用 A×B の効果 62
表 5.6　条件付き予測式 64
表 5.7　反復のある 1 段分割実験のデータの削減 66
表 5.8　平均平方の期待値の比較 66
表 5.9　反復のある 1 段分割実験の REML 法での分散分析表 67
表 6.1　枝分れが 2 段の繰り返しが揃っているデータ 68
表 6.2　スタッカード型の枝分かれデータ “ . ”は測定値なし 69
表 6.3　3 元配置分散分析による平方和の分解 70
表 7.1　分析データ 90

表 7.2　分散分析表 90
表 7.3　枝分かれ実験の分散共分散構造 91
表 7.4　繰返しが不揃いの場合の分散共分散分行列 96
表 8.1　硫化鉱の分析データ 98
表 8.2　異なる試料間のデータの組み合わせ表 100
表 8.3　試料間のデータの組み合わせ表 102
表 8.4　貨車を行方向に転置した表 102
表 9.1　スタッカード型の枝分かれデータ 105
表 9.2　スタッカード型の枝分かれデータの分散分析表 105
表 11.1　乱塊法のデータ（Brown 本） 117
表 11.2　JMP の REML 法での症例間の誤差の推定 118
表 12.1　トウモロコシ顆粒製造実験の結果 121
表 12.2　因子ごと水準ごとの実験数 121
表 12.3　因子 B と因子 C の水準ごとの組み合わせ実験数 122
表 13.1　窒素肥料の量と遺伝子型の組み合わせた小麦の収量 127
表 13.2　東西方向の窒素肥料の量 128
表 13.3　東西方向の窒素肥料の量の平均を出来るだけ均一にした割り付け 128
表 13.4　遺伝子ごとの窒素肥料の量を変化させた場合の収量の変化 129
表 13.5　完全ランダムとした分散分析表 129
表 13.6　南北方向の収量の違い 130
表 13.7　東西方向の収量 130
表 13.8　窒素肥料の量ごとに遺伝子型の差の変化 131
表 13.9　1 段分割実験とした分散分析表 132
表 13.10　窒素肥料の量を連続量とした 1 段分割実験 132
表 13.11　遺伝子型 G1 および G2 の場合の 2 次曲線のあてはめ 133
表 13.12　パラメータの推定値 134
表 13.13　分散分析表 134
表 13.14　パラメータの推定値 135
表 13.15　分散分析表 135
表 13.16　パラメータの推定値 136
表 13.17　分散分析表 136
表 13.18　パラメータの推定値 137
表 13.19　複数の変量効果を考慮した分散分析表 137

1. はじめに

2006 年 3 月 3 日の JMPer's Meeting で「JMP による各種の分割実験データの解析」というテーマで次のような内容の話をする機会があった.

> 実験上の制約により実験順序のランダム化が制約を受けた場合には，分割実験モデルでの解析を行う必要がある．実験データが綺麗に揃っている場合には，面倒ではあるが直交表あるいは完備型の分散分析表を手作業で組み替えることにより対応してきた．しかしながら，データが不揃いの場合には，お手上げとなる．JMP 6 では，データが不揃いな分割実験・枝分かれ実験の場合に対しても洗練された解析機能が「モデルのあてはめ」に備わっている．SAS の MIXED プロシジャで解析が可能であるが，JMP で解析できるようになったことは喜ばしい．今回の JMPer's Meeting では，分割実験・枝分かれ実験の簡単な事例について解析方法と結果の見方を紹介する．

スライド： http://www.jmp.com/japan/about/events/jmp_jmpers_2006_1.shtml#tokyo

その後，直交表を用いた分割実験，2 方分割実験の場合の解析について追加資料を作成し，これらのパワーポイント資料も，上記の Web アドレスからダウンロードできるようになっている．2011 年 11 月に幾つかの Web の検索エンジンで「分割実験」を検索語としてサーチすると，最初に「JMP による各種の分割実験データの解析」がヒットし，「直交表」が 2 番目，「2 法分割」はやや落ちる結果となる．「2 方」をキーワードに追加すると最初「2 方分割」の内容が最初にリストされる．

JMPer's Meeting 後に「JMP による分割表の解析についての勉強会」を何人かのメンバーと共に開始し，2009 年まで継続的な活動が行なってきたが，成果を公表することなく休眠状態となってしまった．理由はいろいろあるのだが，最大の問題は，JMP で分割表の解析を行なう際に，

REML（推奨）制限付き最尤法

EMS （従来）期待平均平方

と選択を迫られる．もちろん JMP が推奨している REML 法を選べばよいのであるが，従来の EMS 法と REML 法の違いについて，私（高橋）含め勉強会に集まったメンバーは，適切かつ説得性のある説明をすることができなかった．

自らが解析する場合に，JMP で推奨されている REML 法を用い報告書を作成しても，JMP で行なったといえば，それ以上の説明をせずに済むであろう．しかしながら，セミナーなどで，JMP を用いた分割実験の解析法として REML 法を取り上げた場合には，説明する責任が生ずる．セミナーを開催し，各メンバーが講師を勤められることを勉強会の暗黙の目的としていたこともあり，勉強会の継続が極めて困難となった．

その後，勉強会での議論を元に，Excel を用いて REML 法による計算にチャレンジし，その経験を SAS ユーザ会，統計の関連学会で発表し，分割実験に対する REML 法について説明することができるようになった．そこで，「JMP による分割表の解析についての勉強会」が挫折したまま自然消滅となることは忍びないので，その後の成果を含めて最終講義を企画した．

最終講義の前半は，1 組のデータが実験のランダム化の方法により，幾通りもの全く異なる解となることを解説し，各種のランダム化に対応する JMP の使い方，結果の見方，解析上の注意点などと合わせて紹介する．これは，2006 年の JMPer's Meeting のスライド 56 で次のようなデータを提示した．

表 1.1　整理された実験データ

	R	B1	B2	B3	B4
A1	R1	12.3	13.9	15.0	15.0
	R2	11.7	12.7	13.0	13.4
A2	R1	15.5	14.8	15.2	14.4
	R2	13.9	13.3	14.3	13.6
A3	R1	13.9	13.3	14.3	12.8
	R2	14.0	13.2	13.5	12.8

このデータに対して，表 1.2 に示すように多くの分割実験の方法があることをスライド 57 示したが，その解析方法および結果については触れなかった．なお，ここに示した解析の結果については，1994 年の SAS ユーザ会で，「各種分割実験モデルに対する MIXED プロシジャの活用」で発表したものに，今回は，JMP の卓越したグラフ機能を盛り込み，その後に得られた知見を含めて大幅な改定をおこなった．

表 1.2　同一データセットに対する 7 種類の解析法

実験モデル	固定効果	変量効果	R の役割
1）完全ランダム実験	A　B　A×B	なし	なし
2）測定の繰り返し	A　B	A×B	なし
3）反復実験	A　B　A×B	R	反復
4）1 段分割実験	A　B　A×B	R(A)	ブロック
5）反復のある 1 段分割実験	A　B　A×B	R　R×A	反復
6）2 方分割実験	A　B　A×B	R×A　R×B	ブロック
7）反復のある 2 方分割実験	A　B　A×B	R　R×A　R×B	反復

後半は，JMP による分割実験の計算原理の基礎で，これまでの実験データの解析の基礎である最小 2 乗法とは異なる REML の考え方について理解の助けにとなる詳細な資料を作成したので，そのエッセンスを解説する．一般の統計の教科書で取り上げられていない複数の分散の同時推定を実感してもらうことを目的としている．

分割実験を行なったときには複数の誤差がデータに張り付いている．それを一括して求める方法

はすでに確立していて SAS の MIXED プロシジャで取り入れられ，JMP でも使えるようになった．しかしながら，JMP の解説書では，どのような計算手順なのか，それを追試できるような解説は見当たらない．どのような誤差が個々のデータにどのようについているかは，REML 法の理解し，適切に使いこなす上でのキーポイントである．そのためには，Excel による計算過程を可視化し，だらもが追試できるようにすることが肝要である．

本書をまとめる切っ掛けは「最終講義」のためであったが，勉強会で取り上げられた例題，高橋が関連学会で発表したスライドなどについて解説を加えて含めることにした．なお，本書は，要因配置実験および分割実験の統計解析について基礎的な知識があることを前提としている．

v1.1 への改定の要旨

2011 年 12 月 8 日に「最終講義」を行なった．参加者から様々な質問，コメント，要望などに答えるために，改定作業を行なった．

1. 第 2 章 様々なランダム化に対応する分散分析では，節により解説が不十分であったので，加筆・訂正を行なった．2.2 節 測定の繰返し実験では，分散分析表を作成するために JMP の出力のどの部分を用いたかを図にして，本文中の説明が理解しやすいようにした．図 2.6 の因子ごとの水準平均のプロファイルの説明に対して，理解しやすいようにプロファイルの変化の図を加えた．2.3 節では，【モデル効果の構成】を加えた．2.7 節の反復のある 2 法分割実験では，プーリング前と後を複合したのを別々の表とした．

2. 第 3 章 EMS（従来法）に変わる REML 法では，直交表を用いた分割実験について，乱塊法実験での「ブロック番号」を変量効果とし，ブロックに 1 次因子を割り付ける考え方を付け加えた．

3. 第 4 章 欠測値を含むクロスオーバー試験では，EMS 法では因子 dose の平方和が 0 となり解析できないとしたが，枝分れ構造にすると解析ができることを追加した．

4. 第 7 章 Excel ソルバーによる REML 法入門で，7.2 節 最小 2 乗法による母平均の推定では，ソルバーで偏差平方和を最小化する前後のイメージ図などを追加した．7.2 節 最尤法による母平均と母分散の推定でもソルバーで尤度を最大化する前後のイメージ図を追加した．新たな節「REML（制限付き最尤法）による母平均と母分散の推定」を追加し，REML 法の理解の原点であることを強調し，理解の助けになるイメージ図も加えた．さらに，新たな節 「REML（制限付き最尤法）による回帰分析のパラメータ推定」を加え，誤差が一つしかない場合の従来の線形モデルに対する最小 2 乗法，ML 法，REML 法の適用方法について解説を加えた．初版 V1.0 では 7.4 節以後の「枝分かれ

型のデータ」で 2 つの誤差についての Compound Symmetry 型（複合対称型）の分散共分散行列の構造について解説をせずに用いた．枝分れ実験のデータが，Compound Symmetry 型（複合対称型）となることの実例を用いて解説した新たな章「枝分れ実験における分散共分散構造」として起こした．

5．第 6 章 スタッカード型枝分れ実験の 6.3 節で従来型の分散成分の 95%信頼区間に推定法方法と REML 法による推定方法について比較を行なった．新たな節「6.4.　シミュレーション研究による分散成分の 95%信頼区間の妥当性」を起こし，REML 法による推定方法が適切なのかをシミュレーション研究の必要性について追記した．

6．まだ書き足りない問題があるが，以後の改定で反映させたい．

2. 様々なランダム化に対応する分散分析

2.1. 完全ランダム実験とみなした場合

完全にランダムの場合の分散分析

繰り返しありの 2 元配置分散分析というのは，表 1.1 に示した 24 個のデータが'完全にランダム化されて得られたことを前提としている．しかしながら，この前提が満たされるような実験は，現実的にはほとんど行われていないと考えたほうがよい．それでは，完全なランダム化実験とは，どのよう一な実験なのだろうか．表 2.1 には，実際にどのような順番で実験データが得られたのかも合わせて示した．

表 2.1　完全ランダム化の手順と実験データ

	B1	B2	B3	B4
A1	(6) 12.3 (11) 11.7	(7) 13.9 (16) 12.7	(18) 15.0 (24) 13.0	(5) 15.0 (17) 13.4
A2	(19) 15.5 (21) 13.9	(1) 14.8 (4) 13.3	(3) 15.2 (10) 14.3	(8) 14.4 (23) 13.6
A3	(2) 13.9 (9) 14.0	(15) 13.3 (22) 13.2	(14) 14.3 (20) 13.5	(12) 12.8 (13) 12.8

括弧（ ）の中の数字が，実際に実験を行なった順番の一例である．第 1 番目の実験は（A2，B2）の組み合わせ，第 2 番目の実験は（A3，Bl）の組み合わせ, . . . , 第 24 番目の実験は（A1，B3）の組み合わせで逐次行なわれたことを示している．

このような手順で実験をしたならば，表 2.2 に示すように標準的な方法により，繰り返しありの 2 元配置の分散分析表を作成する．交互作用 A×B の p 値も 0.1818 と大きく，因子 A の p 値が 0.0523 で統計的に差が有りそうに見えるので，交互作用項を誤差にプールして主効果モデルとして分散分析を行なっても因子 A の p 値は 0.0523 と有意水準 $\alpha = 0.05$ で有意とはならない．

表 2.2　完全ランダム化の場合の分散分析表

要因	自由度	平方和	平均平方	F 値	p 値	平均平方の期待値
A	2	4.8533	2.4267	3.81	0.0523	$\sigma_e^2 + 8\,\sigma_A^2$
B	3	1.8683	0.6228	0.98	0.4353	$\sigma_e^2 + 6\,\sigma_B^2$
A×B	6	6.8767	1.1461	1.80	0.1818	$\sigma_e^2 + 2\,\sigma_{A\times B}^2$
誤差 e	12	7.6400	0.6367	1.00		σ_e^2
全体	23	21.2383				

JMPによる解析「モデルのあてはめ」

JMP には，1 元配置，2 元配置，分割実験，L_{16} 直交表，などの手法別のタイトルは全く，重回帰分析も含めた線形形モデル，または線形混合モデルなども「モデルのあてはめ」に 1 本化されており，各種の手法はユーザの使い方によって実現できるようになっている．ここでは，繰り返し 2 の

2 元配置分散分析を行なうための基本を示す.

表 2.2 のデータを次のような JMP データとし,【モデルの指定／役割の選択】で,【Y】として変数 y を,【モデルの構成】で，因子 A，因子 B を選択し【追加】すると【モデル効果の構成】に書き込まれる．交互作用 A×B は，因子 A と因子 B を同時に選択して【交差】を選択すると A*B として書き込まれるので,【実行】する.

	R	A	B	y
1	R1	A1	B1	12.3
2	R1	A1	B2	13.9
3	R1	A1	B3	15.0
4	R1	A1	B4	15.0
5	R1	A2	B1	15.5
6	R1	A2	B2	14.8
7	R1	A2	B3	15.2
8	R1	A2	B4	14.4
9	R1	A3	B1	13.9
10	R1	A3	B2	13.3
11	R1	A3	B3	14.3
12	R1	A3	B4	12.8
13	R2	A1	B1	11.7
14	R2	A1	B2	12.7
15	R2	A1	B3	13.0
16	R2	A1	B4	13.4
17	R2	A2	B1	13.9
18	R2	A2	B2	13.3
19	R2	A2	B3	14.3
20	R2	A2	B4	13.6
21	R2	A3	B1	14.0
22	R2	A3	B2	13.2
23	R2	A3	B3	13.5
24	R2	A3	B4	12.8

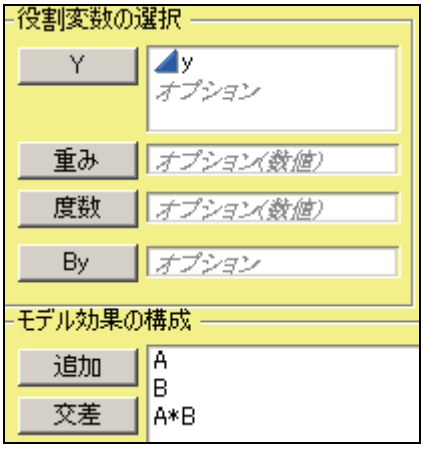


表 2.2 の分散分析表は,【モデルのあてはめ】の次の出力を整理しまとめ直した結果である.

分散分析

要因	自由度	平方和	平均平方	F値
モデル	11	13.5983	1.23621	1.9417
誤差	12	7.6400	0.63667	p値(Prob>F)
全体(修正済み)	23	21.2383		0.1348

効果の検定

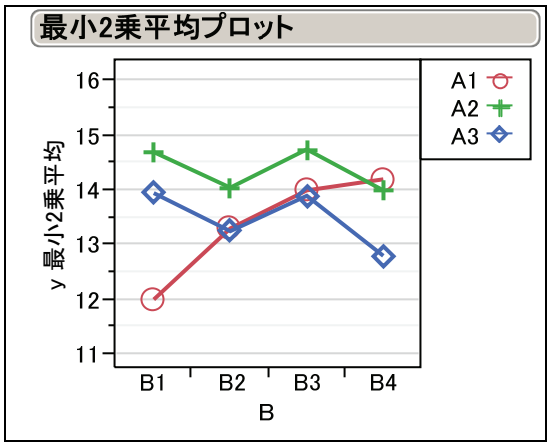
要因	パラメータ数	自由度	平方和	平均平方	F値	p値(Prob>F)
A	2	2	4.8533	2.4267	3.8115	0.0523
B	3	3	1.8683	0.6228	0.9782	0.4353
A*B	6	6	6.8767	1.1461	1.8002	0.1818

小数点以下は 4 桁，または 2 桁に設定し直してある.

結果のグラフ化と解釈

分散分析表の作成は，実験データの解析の第 1 歩に過ぎない．実験条件を変えて再実験する再にも，実験のランダム化に即した実験結果のグラフ表示が欠かせない．JMP のグラフ機能は，バージョンアップするごとに充実してきている．それらの使い方の詳細は冗長になるので省くが，最小限のヒントは残しつつ結果の解釈で最もスマートと思われるグラフと結果の解釈に有用と思われる出力を抜粋して示す．

因子 A を質的因子，因子 B を量的因子と想定したときには【効果の詳細／A*B／最小 2 乗平均プロット】を用いて因子 B を X 軸とした交互作用の図を作成することで結果の解釈がしやすい．反応が大きい方が望ましいとしたときに，因子 A の A2 水準が因子 B の水準に関係なく高くなっていることから，A2 水準が安定した結果となっている．B3 水準は，因子 A の条件にかかわらず安定して高いので，最適水準と思われる．再度，実験を計画して確認することが必要である．



最小2乗平均差のStudentのt検定

$\alpha = 0.050$ t= 2.17881

水準				最小2乗平均
A2,B3	A			14.7500
A2,B1	A			14.7000
A1,B4	A	B		14.2000
A2,B2	A	B		14.0500
A1,B3	A	B		14.0000
A2,B4	A	B		14.0000
A3,B1	A	B		13.9500
A3,B3	A	B		13.9000
A1,B2	A	B	C	13.3000
A3,B2	A	B	C	13.2500
A3,B4		B	C	12.8000
A1,B1			C	12.0000

同じ文字でつながっていない水準は有意に異なります。

図 2.1　完全ランダム実験の場合の 2 因子の交互作用プロット

実験結果のグラフを見ると，各ポイント間で最適水準と（A2，B3）と現行水準を（A1，B1）としたときの水準平均の比較を行ないたくなる．【効果の詳細／A*B／最小 2 乗平均差の Student の t 検定／文字の接続レポート】を用いると図 2.1 右のように全ての水準の組み合わせ同士の検定結果がコンパクトに出力される．

最も小さい（A1，B1）の推定平均（最小 2 乗平均）は 12.0 であり，最も大きい（A2，B3）は 14.75 であることが読み取れる．同じ文字の線（A，B，C）でつながっている水準の組み合わせ間は有意な差でなく，つながっていない水準の組み合わせ間は有意な差である．現行水準（A1，B1）12.0 と最適水準（A2，B3）14.75 は，文字 A の線，B の線，C の線，どれを見てもつながっていないので有意な差と判断される．

なお，このレポートで用いられている「最小 2 乗平均」は，データが揃っているので，セル平均に等しい．（A2，B3）の場合は，（15.2 + 14.3）/ 2 = 14.75 と同じ結果になっている．セルの繰返しが不揃いな場合は，最小 2 乗平均とセル平均が異なる．

【最小 2 平均】は，次のような考え方で計算されたものである．2 元配置分散分析の名義尺度の要因は，全て連続量のダミー変数として展開され「重回帰分析」のように解析されている．最小 2 平均】は，次に示す【推定値】から計算されたものである．

全水準の推定値

名義尺度の要因においては、全水準に対して推定値が求められている

項	推定値
切片	13.7417
A[A1]	-0.3667
A[A2]	0.6333
A[A3]	-0.2667
B[B1]	-0.1917
B[B2]	-0.2083
B[B3]	0.4750
B[B4]	-0.0750
A[A1]*B[B1]	-1.1833
A[A1]*B[B2]	0.1333
A[A1]*B[B3]	0.1500
A[A1]*B[B4]	0.9000
A[A2]*B[B1]	0.5167
A[A2]*B[B2]	-0.1167
A[A2]*B[B3]	-0.1000
A[A2]*B[B4]	-0.3000
A[A3]*B[B1]	0.6667
A[A3]*B[B2]	-0.0167
A[A3]*B[B3]	-0.0500
A[A3]*B[B4]	-0.6000

最小 2 乗平均が最も大きい A2B3 の支障 2 乗平均 14.75 は，【項】はダミー変数名で，該当する【推定値】を選んで，次のように計算されたものである．

$$
\begin{aligned}
(A_2B_3) &= \text{切片} + (A_2) + (B_3) + (AB)_{23} \\
&= 13.7417 + 0.6333 + 0.4750 - 0.1000 = 14.75
\end{aligned}
$$

【文字の接続レポート】の前には，図 2.2に示すように全ての水準間の差，差の標準誤差（SE），95%信頼区間について示されている．（A1，B1）と（A2，B3）の差が -2.75，SE が 0.7979，95%信頼区間が（-4.488，-1.0115）であり，95%信頼区間が 0.0 を含んでいないので有意な差であることがわかる．

	最小2乗平均[j]							
平均[i]-平均[j] 差の標準誤差 差の下側信頼限界 差の上側信頼限界	A1,B1	A1,B2	A1,B3	A1,B4	A2,B1	A2,B2	A2,B3	A2,B4
A1,B1	0	-1.3	-2	-2.2	-2.7	-2.05	-2.75	-2
	0	0.79791	0.79791	0.79791	0.79791	0.79791	0.79791	0.79791
	0	-3.0385	-3.7385	-3.9385	-4.4385	-3.7885	-4.4885	-3.7385
	0	0.43851	-0.2615	-0.4615	-0.9615	-0.3115	-1.0115	-0.2615

図 2.2　水準間の差についての統計量（抜粋）

図 2.3 に示す【差の順位レポート】も同様な統計量が示されている．（A2，B3）と（A1，B1）の差について，差が 2.75，差の 95%信頼区間が（1.0115，4.4885），および p 値は 0.0048 であることが読み取れる．

水準	－ 水準	差	差の標準誤差	下側信頼限界	上側信頼限界	p値
A2,B3	A1,B1	2.7500	0.7979	1.0115	4.4885	0.0048*
A2,B1	A1,B1	2.7000	0.7979	0.9615	4.4385	0.0054*
A1,B4	A1,B1	2.2000	0.7979	0.4615	3.9385	0.0174*
A2,B2	A1,B1	2.0500	0.7979	0.3115	3.7885	0.0246*
A1,B3	A1,B1	2.0000	0.7979	0.2615	3.7385	0.0276*
A2,B4	A1,B1	2.0000	0.7979	0.2615	3.7385	0.0276*
A3,B1	A1,B1	1.9500	0.7979	0.2115	3.6885	0.0309*
A2,B3	A3,B4	1.9500	0.7979	0.2115	3.6885	0.0309*
A3,B3	A1,B1	1.9000	0.7979	0.1615	3.6385	0.0347*
A2,B1	A3,B4	1.9000	0.7979	0.1615	3.6385	0.0347*
A2,B3	A3,B2	1.5000	0.7979	−0.2385	3.2385	0.0846
A2,B3	A1,B2	1.4500	0.7979	−0.2885	3.1885	0.0942

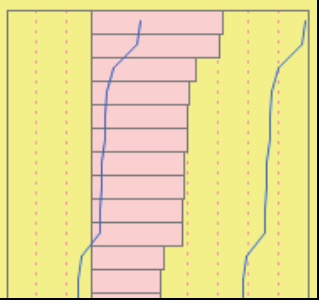


図 2.3　差の順位レポート（抜粋）

検定統計量など詳細な結果は，【詳細な比較】で出力される．（A1，B1）と（A2，B3）の差の p 値は，$t = 2.7500 / 0.7979 = 3.4565$ に対して自由度 12 の t 分布から求められたことがわかる．

A2,B3をA1,B1と比較

差	2.75000	t値	3.446487
差の標準誤差	0.79791	自由度	12
差の上側信頼限界	4.48851	p値(Prob>|t|)	0.0048*
差の下側信頼限界	1.01149	p値(Prob>t)	0.0024*
信頼率	0.95	p値(Prob<t)	0.9976

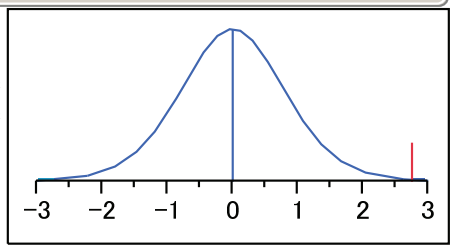


図 2.4　差についての詳細なレポート（抜粋）

多重性の配慮

ここで示した総当り形式の有意差検定は，検定の多重性を全く考慮していない 2 群間の差の検定結果である．探索的な検討のために用いることは大いに結構なのだが，実験の結果は様々な原因によって揺らぐので，この有意差検定の結果を検証目的とする場合には使用してはならない．図 2.5 に示す検定の多重性を考慮した【Tukey の HSD 検定／差の順位レポート】の結果では，全ての水準の組み合わせ間で有意な差がないことが確認できる．

水準	－ 水準	差	差の標準誤差	下側信頼限界	上側信頼限界	p値
A2,B3	A1,B1	2.750000	0.7979139	−0.41784	5.917839	0.1128
A2,B1	A1,B1	2.700000	0.7979139	−0.46784	5.867839	0.1241
A1,B4	A1,B1	2.200000	0.7979139	−0.96784	5.367839	0.3028
A2,B2	A1,B1	2.050000	0.7979139	−1.11784	5.217839	0.3841
A1,B3	A1,B1	2.000000	0.7979139	−1.16784	5.167839	0.4140
A2,B4	A1,B1	2.000000	0.7979139	−1.16784	5.167839	0.4140
A3,B1	A1,B1	1.950000	0.7979139	−1.21784	5.117839	0.4452
A2,B3	A3,B4	1.950000	0.7979139	−1.21784	5.117839	0.4452
A3,B3	A1,B1	1.900000	0.7979139	−1.26784	5.067839	0.4775
A2,B1	A3,B4	1.900000	0.7979139	−1.26784	5.067839	0.4775
A2,B3	A3,B2	1.500000	0.7979139	−1.66784	4.667839	0.7521

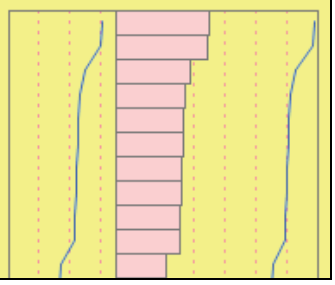


図 2.5　Tukey の HSD 検定に基づく差の順位レポート（抜粋）

2.2. 測定の繰り返し実験

測定の繰返しでの12 回のランダム化

しばしば発生するのは，表 2.3 に示すように，実験が 12 回のランダム化しか行なわれなかった場合である．第 1 回目の実験は（A2，B2）の組み合わせで，作成されたサンプルを，その場で 2 回測定して（14.8，13.3）が得られ，第 2 回目の実験は，（A1，B3）の組み合わせで（15.0，13.0）が得られ，順次第 12 回目まで実験が行なわれたものとする．

なお，実験の結果として液体のようなサンプルを分割できるようなものは，それを 2 等分して測定を繰り返しても同じ実験モデルと考えられる．さらに，2 等分した 24 個のサンプルを自動分析装置などで一括した順次測定しても同じ実験モデルが適用される．

表 2.3　測定の繰り返しがある実験データ

	B1	B2	B3	B4
A1	（10）12.3 11.7	（9）13.9 12.7	（2）15.0 13.0	（8）15.0 13.4
A2	（4）15.5 13.9	（1）14.8 13.3	（6）15.2 14.3	（12）14.4 13.6
A3	（5）13.9 14.0	（11）13.3 13.2	（3）14.3 13.5	（7）12.8 12.8

実験の順序が記載されていないならば，表 2.3 のデータは，表 2.1 とまったく同じで区別が付かない．違いは，各測定値に付いている目には見えない誤差の構造が異なるのである．完全ランダム実験の場合は，各データにただ 1 つの誤差しか付いていないのに対し，この実験には 2 種類の誤差，1 次誤差 $\varepsilon_{(1)}$ と 2 次誤差 $\varepsilon_{(2)}$ が，各測定値に付いている．因子 A の水準を i，因子 B の水準を j，測定の繰返しを k としたときに．因子 A の効果を α_i，因子 B の効果を β_j，因子 A と因子 B の交互作用を $(\alpha\beta)_{ij}$，因子 A と因子 B の変更に伴う 1 次誤差を $\varepsilon_{(1)ij}$，測定の繰返しに付く 2 次誤差を $\varepsilon_{(2)ijk}$ とする．測定値を y_{ijk} は，次のような構造で示すことができる．

$$y_{111} = \mu + \alpha_1 + \beta_1 + (\alpha\beta)_{11} + \varepsilon_{(1)11} + \varepsilon_{(2)111} = 12.3$$

$$y_{112} = \mu + \alpha_1 + \beta_1 + (\alpha\beta)_{11} + \varepsilon_{(1)11} + \varepsilon_{(2)112} = 11.7$$

$$\vdots$$

$$y_{342} = \mu + \alpha_3 + \beta_4 + (\alpha\beta)_{34} + \varepsilon_{(1)34} + \varepsilon_{(2)342} = 12.8$$

ただし，

$$\sum \alpha_i = 0, \qquad \sum \beta_j = 0, \qquad \varepsilon_{(1)ij} \sim \mathrm{N}(0, \sigma_{(1)}^2), \qquad \varepsilon_{(2)ijk} \sim \mathrm{N}(0, \sigma_{(2)}^2),$$

とする．1 次誤差 $\varepsilon_{(1)ij}$ は，因子 A と因子 B の水準の組み合わせを変えて実験するたびに付く誤差である．y_{111} と y_{112} は，同時に実験されたのであるから共通の 1 次誤差 $\varepsilon_{(1)11}$ が付き，それぞれのデータに別々の 2 次誤差 $\varepsilon_{(2)111}$ と $\varepsilon_{(2)112}$ が付いている．問題は交互作用 $(\alpha\beta)$ と 1 次誤差 $\varepsilon_{(1)}$ が，それぞれ

のデータの添字が完全一致，言い換えれば完全に交絡しているので分離できない．このために，因子 A および因子 B の検定は，交互作用 A×B を 1 次誤差 $\varepsilon_{(1)}$ とみなすことになる．

分散分析

標準的な解析法で得られた分散分析の結果を表 2.4 に示す．F 値の欄は検定のための分母を明示するために 2 つに分けてある．F 値が 1.00 の場合の平均平方が F 検定の分母である．因子 A および因子 B は，1 次誤差の平均平方を分母として F 値を求めている．1 次誤差の F 値は，2 次誤差の平均平方を分母として求めている．分散分析表の平均平方は，表 2.2 で示した完全ランダム実験の場合と同じであるが，平均平方の期待値に対応して F 値の分母が求め方が異なる．

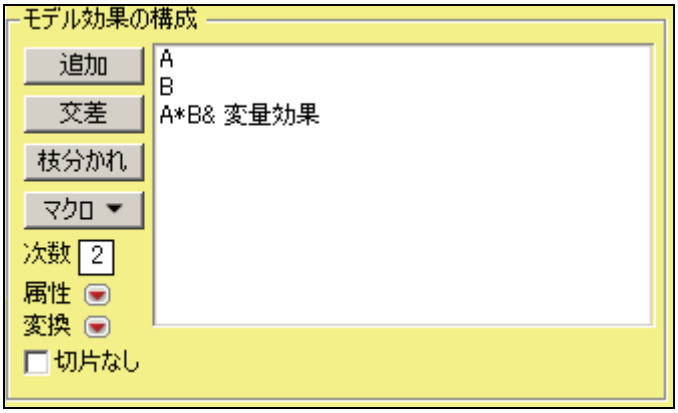
表 2.4　測定の繰り返しがある場合の分散分析表

要因	自由度	平方和	平均平方	F 値		p 値	平均平方の期待値
A	2	4.8533	2.4267	-	2.12	0.2015	$\sigma^2_{(2)}+2\sigma^2_{(1)}+8\sigma^2_A$
B	3	1.8683	0.6228	-	0.54	0.6703	$\sigma^2_{(2)}+2\sigma^2_{(1)}+6\sigma^2_B$
A×B:1 次誤差 $e_{(1)}$	6	6.8767	1.1461	1.80	1.00	0.1818	$\sigma^2_{(2)}+2\sigma^2_{(1)}$
2 次誤差 $e_{(2)}$	12	7.6400	0.6367	1.00			$\sigma^2_{(2)}$
全体	23	21.2383					

因子 A と因子 B の F 値は，2 次誤差の平均平方が 1.1461 と 2 次誤差の 0.6367 に比べて大きいので，表 2.2 の分散分析表の F 値に比べ因子 A も因子 B も小さくなっている．因子 A の p 値は 0.0523 から 0.2015 と増大し，有意な傾向（$p<0.1$）も見られなくなっている．

JMPでの解析手順と結果の読み方

JMP で測定の繰返しがある場合についての解析方法と結果の見方について示す．【分析／モデルのあてはめ】を選択し．【モデルの構成】で，因子 A，因子 B を【追加】し，交互作用 A*B を【交差】で追加し，A*B をアクティブにし【属性】の中の【変量効果】を選択し，交互作用 A*B を「A*B& 変量効果」と変える．【方法】を【EMS（従来）】に変え，【実行】する．



JMP の出力は，汎用性を考慮しているために，冗長な表示となっている．表 2.4 の分散分析表は，以下に示す結果を Excel にコピーして整形したものである．【分散分析】は，モデルに含めた主効果，変量効果をすべて 3 「モデル」としている．分散分析表の「2 次誤差 $\varepsilon_{(2)}$」と「全体」の行は，ここから抽出している．

分散分析

要因	自由度	平方和	平均平方	F値
モデル	11	13.5983	1.2362	1.9417
誤差	12	7.6400	0.6367	p値(Prob>F)
全体(修正済み)	23	21.2383		0.1348

【F 検定の分母】は，【モデル効果の構成】に含まれた全ての要因に対して，F 値を求めるときの分母とする誤差項が示されている．因子 A は 1 次誤差 $\varepsilon_{(1)}$ とした「A*B&変量効果」の平均平方 1.1461 を分母としていることが分かる．「A*B&変量効果」は，2 次誤差 $\varepsilon_{(2)}$ 「残差」の平均平方 0.6367，これは【分散分析】の「誤差」の平均平方に対応している．

F検定の分母

要因	分母平均平方	分母自由度	分母平均平方合成
A	1.1461	6	A*B&変量効果
B	1.1461	6	A*B&変量効果
A*B&変量効果	0.6367	12	残差

【変量効果を考慮した検定】は，【F 検定の分母】を用いて，それぞれの要因の平均平方を割っている．因子 A の場合では，2.4267/1.146 = 2.11 であり，A*B&変量効果は，1.146/0.6367 = 1.80 であることが確認できる．

変量効果を考慮した検定

要因	平方和	分子平均平方	分子自由度	F値	p値(Prob>F)
A	4.8533	2.4267	2	2.1173	0.2015
B	1.8683	0.6228	3	0.5434	0.6703
A*B&変量効果	6.8767	1.1461	6	1.8002	0.1818

【期待平均平方】は，表 2.4 の平均平方の期待値の欄に対応する．行方向の因子 A には，8 倍の因子 A の分散成分 σ_A^2 と 2 倍の A*B&変量効果の分散成分 $\sigma_{(1)}^2$ の合計となることが示されている．さらに脚注で【プラス 1.0 倍の残差誤差分散】$\sigma_{(2)}^2$ が付くことが示されているので，因子 A の平均平方の期待値は，$\sigma_{(2)}^2 + 2\sigma_{(1)}^2 + 8\sigma_A^2$ となる．行方向の A*B&変量効果には，2 倍の A*B&変量効果の分散成分 $\sigma_{(1)}^2$ と【プラス 1.0 倍の残差誤差分散】$\sigma_{(2)}^2$ が付くので，$\sigma_{(2)}^2 + 2\sigma_{(1)}^2$ となる．これらの平均平方の期待値から，【F 検定の分母】となる因子が特定できる．

期待平均平方

各行の平均平方の期待値を構成する各列の分散成分の係数
期待平均平方

	切片	A	B	A*B&変量効果
切片	0	0	0	0
A	0	8	0	2
B	0	0	6	2
A*B&変量効果	0	0	0	2

プラス1.0倍の残差誤差分散

【分散成分推定値】には，「A*B&変量効果」の 1 次誤差の分散成分 $\sigma^2_{(1)} = 0.2547$，1 次誤差の分散成分 $\sigma^2_{(2)} = 0.6367$ が示されている．分散成分 $\sigma^2_{(1)}$ は，平均平方の期待値が $\sigma^2_{(2)} + 2\sigma^2_{(1)}$ であるので分散成分 $\sigma^2_{(2)}$ を差し引いて 2 で割ればよい．表 2.4 の平均平方の期待値を用いて，実際に計算すると （1.1461 - 0.6373）/ 2 = 0.2547 が得られる．

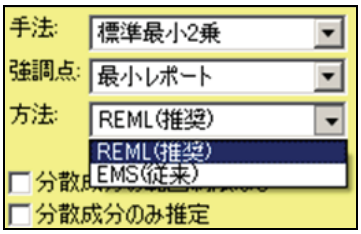
分散成分推定値

成分	分散成分推定値	全体に対する百分率
A*B&変量効果	0.2547	28.6
残差	0.6367	71.4
合計	0.8914	100.0

平均平方がその期待値に等しいものとして推定したものです。

EMS法およびREML法のプロファイルの相違

【モデル効果の構成】で，ある要因に対して変量効果に変更すると，解析【方法】が【REML（推奨)】となる．これまで【EMS（従来)】を選ぶように記述してきたのは，従来の日本語の実験計画法の教科書での分割実験の解析法が，【EMS（従来)】に対応しているからである．



平均平方の期待値を用いた分割法の分散分析表は，データが揃っている場合について【REML（推奨)】と同じ F 検定結果となり，【EMS（従来)】での解析法を示してきた．ところが，水準平均に関連する結果は，【EMS（従来)】と【REML（推奨)】と異なる．REML 法を用いた解析については，次章（第 3 章）で示すが，どのような違いがあるかを【予測プロファイル】の結果を用いて示す．

期待平均平方

各行の平均平方の期待値を構成する各列の分散成分の係数
期待平均平方

	切片	A	B	A*B&変量効果
切片	0	0	0	0
A	0	8	0	2
B	0	0	6	2
A*B&変量効果	0	0	0	2

プラス1.0倍の残差誤差分散

変量効果を考慮した検定

要因	平方和	分子平均平方	分子自由度	F値	p値(Prob>F)
A	4.8533	2.4267	2	2.1173	0.2015
B	1.8683	0.6228	3	0.5434	0.6703
A*B&変量効果	6.8767	1.1461	6	1.8002	0.1818

測定の繰り返しがある場合の分散分析表

要因	自由度	平方和	平均平方	F 値		p 値	平均平方の期待値
A	2	4.8533	2.4267	-	2.12	0.2015	$\sigma^2_{(2)}+2\,\sigma^2_{(1)}+8\,\sigma^2_A$
B	3	1.8683	0.6228	-	0.54	0.6703	$\sigma^2_{(2)}+2\,\sigma^2_{(1)}+6\,\sigma^2_B$
A×B : 1 次誤差 $e_{(1)}$	6	6.8767	1.1461	1.80	1.00	0.1818	$\sigma^2_{(2)}+2\,\sigma^2_{(1)}$
2 次誤差 $e_{(2)}$	12	7.6400	0.6367	1.00			$\sigma^2_{(2)}$
全体	23	21.2383					

分散成分推定値

0.2547
0.6367

分散分析

要因	自由度	平方和	平均平方	F値
モデル	11	13.5983	1.2362	1.9417
誤差	12	7.6400	0.6367	p値(Prob>F)
全体(修正済み)	23	21.2383		0.1348

変量効果を考慮した検定

要因	平方和	分子平均平方	分子自由度	F値	p値(Prob>F)
A	4.8533	2.4267	2	2.1173	0.2015
B	1.8683	0.6228	3	0.5434	0.6703
A*B&変量効果	6.8767	1.1461	6	1.8002	0.1818

分散成分推定値

成分	分散成分推定値	全体に対する百分率
A*B&変量効果	0.2547	28.6
残差	0.6367	71.4
合計	0.8914	100.0

平均平方がその期待値に等しいものとして推定したものです。

図 2.6 分散分析表と JMP の出力の対応付け

図 2.7 に因子 A と因子 B について【予測プロファイル】を示す．因子 A の A2 水準，因子 B の B3 水準を選択した場合のそれぞれの因子のプロファイルが表示され，選択された水準の組み合せ平均，95%信頼区間が表示されるので，結果の解釈をする場合に有益である．

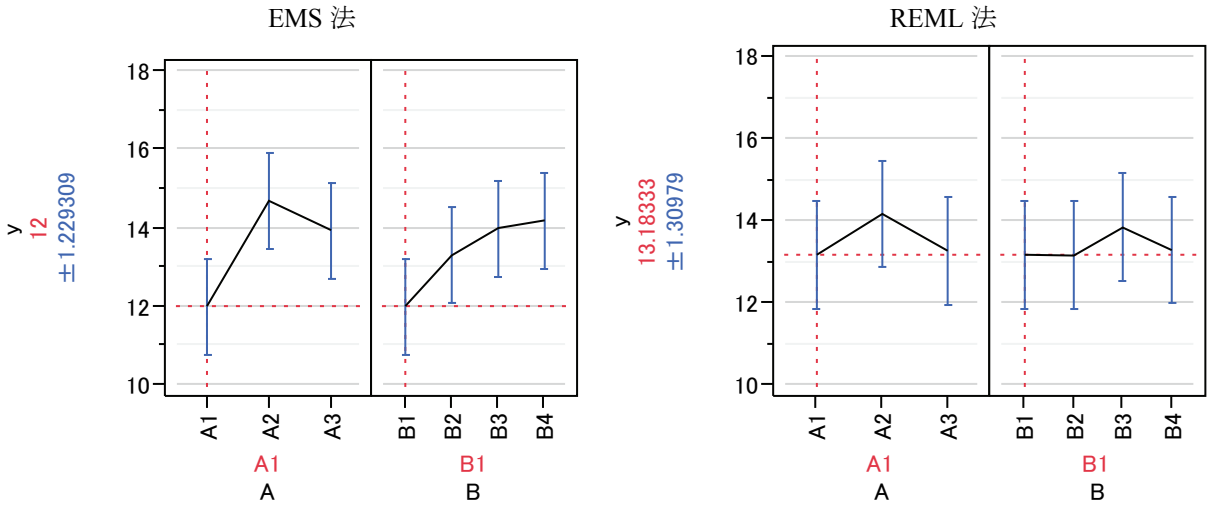


図 2.7　因子ごとの水準平均のプロファイル

EMS 法による水準平均は交互作用 A×B を考慮しているので，交互作用 A×B を変量効果（1 次誤差）とした REML 法による水準平均とは異なる．95%信頼区間も異なる．

EMS 法でのプロファイルは，変量効果とした交互作用 A×B があるとした結果となっている．したがって，因子 A を A1 水準から A2 水準に変更すると，それに対応して因子 B のプロファイルは（A2，B1），. . . ，（A2，B4）となり，プロファイルの形が変化する．REML 法の場合は，因子 A を

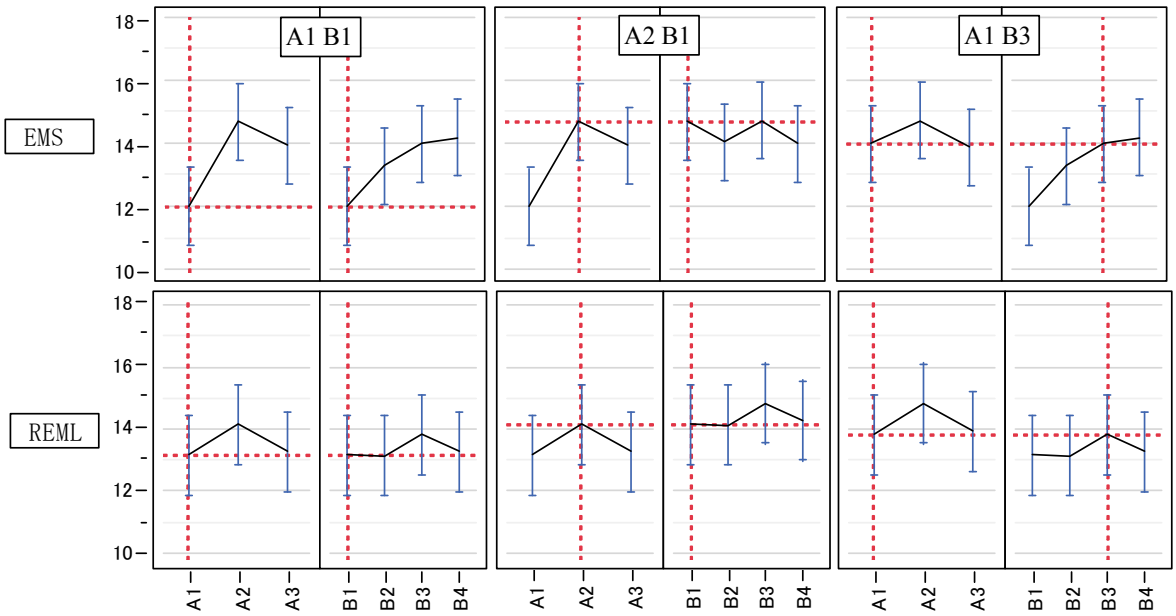


図 2.8　水準の設定によるプロファイルの変化の比較

A1 水準から A2 水準に変更しても，因子 B のプロファイルは上下に移動するだけであり，交互作用 A×B を交互作用としてではなく誤差として扱っていることがわかる．したがって，分割実験で水準平均に関連する解析は，EMS 法ではなく REML 法を使う必要がある．

変量効果と固定効果の使い分け

測定値 y_{ijk} についている誤差は，完全ランダム実験の場合には，ただ 1 つの誤差のみであり，【モデル効果の構成】には含める必要性はない．測定の繰り返し実験の場合の例では，1 次誤差は交互作用 A×B と完全に交絡しており，これを 1 次誤差とみなすことになる．JMP で自動的に付けられる低次の誤差以外は，ユーザが【モデル効果の構成】で変量効果として明示する必要がある．

交互作用 A×B を 1 次誤差とするためには，固定効果としての交互作用を変量効果に変更した．従来の分散分析では，1 次誤差，2 次誤差, . . . として表現してきたが，最低次の誤差以外を変量効果と総称することもある．JMP では，何次の変量効果という指定ではないので，複数の変量効果を設定したときに，それが何次の誤差となるかはユーザが判別する必要がある．

【モデル効果の構成】でとりあえず変量効果としたい要因を固定効果として設定し，【属性】の中から【変量効果】を選んで，属性を変量効果に変更する．どの因子が変量効果になるかは，実験のランダム化に伴い変化する．どのように変量効果を設定しても， JMP は忠実にユーザの設定に従って解析をし，ほとんどの場合は何らかの結果を出してくる．したがって，モデルの構成の適切性を示すためには，ランダム化の手順を正確に詳細に示し，データの構造を合わせて示す必要がある．少しでも油断しミスると，まったく異なる結果となりかねない．

2.3. 反復実験

反復を伴うランダム化

完全ランダム化実験よりも，繰り返しのない 2 元配置実験を 2 回反復する方が，いろいろな意味で好ましい性質を持つ．ある実験が，1 週間で 12 回の実験しか物理的にできないような場合に，1 週間単位で 12 回のランダム化を 2 回反復すれば，実験時期に起因するさまざまな，実験の場の変化を別枠で取り出すことが可能となり，解析の精度を高めることができる．

具体的には，表 2.5 に示すように第 1 反復を R1 として，最初の 12 回の実験を因子 A と因子 B の全ての組み合わせに対しランダムに行なう，次に，第 2 反復を R2 として，（13）から（24）まで

のランダム化に行なう．もしも，第 3 反復を行なうのであるならば，さらに（25）から（36）と実験を追加すればよい．

測定の「繰返し」も，実験の「反復」も一般的には同義語とみなされているが，実験計画法では，別の意味で使われている．英語でも繰返しを repetition，反復を replication と区別することもあるようだが，紛らわしい．反復を丁寧にいえば，「週を一塊，ブロックとした中で因子 A と因子 B の 12 通りの組み合わせをランダムに実験し，次週に同様な実験を反復して行なった」，といえば正確であるが，本書では，反復実験ということにする．

表 2.5　反復実験でのランダム化

R：反復	A	B1	B2	B3	B4
R1	A1	(10) 12.3	(9) 13.9	(2) 15.0	(8) 15.0
	A2	(4) 15.5	(1) 14.8	(6) 15.2	(12) 14.4
	A3	(5) 13.9	(11) 13.3	(3) 14.3	(7) 12.8
R2	A1	(15) 11.7	(21) 12.7	(13) 13.0	(18) 13.4
	A2	(20) 13.9	(16) 13.3	(22) 14.3	(17) 13.6
	A3	(23) 14.0	(14) 13.2	(19) 13.5	(24) 12.8

表 2.5 は，一見すると繰り返しのない 3 元配置分散分析の様に見える．このままで単純に分散分析をすると，反復 R と他の因子との 2 因子交互作用 R×A，R×B，および．3 因子交互作用 R×A×B が得られる．反復実験では，反復 R と反復 R が含まれている交互作用をプーリングして，それを誤差とする．交互作用 R×A は，因子 A の反復 R1 と反復 R2 の変更に伴う食い違いの大きさを示すもので，交互作用 R×B も同様であり，誤差とみなすべき性質のものである．

変量効果としての反復R

この実験には 2 種類の誤差，反復 R の変更に伴う反復誤差 $\varepsilon_{(0)}$ と 1 次誤差 $\varepsilon_{(1)}$ が，各測定値に付いている．反復 R の水準を i，因子 A の水準を j，因子 B の水準を k としたときに，反復 R の効果 γ_i の変更に伴う反復誤差を $\varepsilon_{(0)i}$，因子 A の効果を α_j，因子 B の効果を β_k，因子 A と因子 B の交互作用を $(\alpha\beta)_{jk}$，1 次誤差を $\varepsilon_{(1)ijk}$ とする．測定値を y_{ijk} は，次のような構造で示すことができる．

$$y_{111} = \mu + \gamma_1 + \varepsilon_{(0)1} + \alpha_1 + \beta_1 + (\alpha\beta)_{11} + \varepsilon_{(1)111} = 12.3$$
$$y_{112} = \mu + \gamma_1 + \varepsilon_{(0)1} + \alpha_1 + \beta_2 + (\alpha\beta)_{12} + \varepsilon_{(1)112} = 13.9$$
$$\vdots$$
$$y_{234} = \mu + \gamma_2 + \varepsilon_{(0)2} + \alpha_3 + \beta_4 + (\alpha\beta)_{34} + \varepsilon_{(1)234} = 12.8$$

ただし，

$$\sum \alpha_j = 0, \qquad \sum \beta_k = 0, \qquad \varepsilon_{(0)i} \sim \mathrm{N}(0, \sigma_{(0)}^2), \qquad \varepsilon_{(1)ijk} \sim \mathrm{N}(0, \sigma_{(1)}^2),$$

とする．反復誤差 $\varepsilon_{(0)i}$ は，反復ごとに付く誤差である．y_{111} と y_{112} は，同じ反復なので共通の反復誤差 $\varepsilon_{(0)1}$ が付き，それぞれのデータに因子 A と因子 B の変更に伴う 1 次誤差 $\varepsilon_{(1)111}$ と $\varepsilon_{(1)112}$ が付いて

いる．反復 γ_i と反復誤差 $\varepsilon_{(0)i}$ は，それぞれのデータの添字が完全一致，言い換えれば完全に交絡しているので分離できない．したがって，反復 R を反復誤差とみなす．

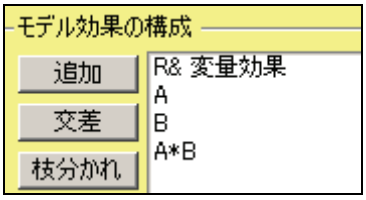


表 2.6　反復実験とした場合の分散分析表

要因	自由度	平方和	平均平方	F 値	p 値	平均平方の期待値
R:反復誤差 $e_{(0)}$	1	5.0417	5.0417	21.34	0.0007	$\sigma^2_{(1)}+12\,\sigma^2_{(0)}$
A	2	4.8533	2.4267	10.27	0.0030	$\sigma^2_{(1)}+8\,\sigma^2_{A}$
B	3	1.8683	0.6228	2.64	0.1019	$\sigma^2_{(1)}+6\,\sigma^2_{B}$
A×B	6	6.8767	1.1461	4.85	0.0117	$\sigma^2_{(1)}+2\,\sigma^2_{A\times B}$
1 次誤差 $e_{(1)}$	11	2.5983	0.2362	1.00		$\sigma^2_{(1)}$
全体	23	21.2383				

反復 R の平均平方は 5.0417 と誤差の平均平方 0.2362 に比べてに比べ極めて大きいことから第 1 反復 R1 と第 2 反復 R2 の間で実験全体に影響する何らかの変化があったことが推測される．【効果の詳細／R&変量効果】の【最小 2 乗平均表】を示す．第 2 反復の全体の平均が 13.28 と第 1 反復の 14.2 に対して小さくなっている．

最小2乗平均表

水準	最小2乗平均	標準誤差	平均
R1	14.2000	0.1403	14.2000
R2	13.2833	0.1403	13.2833

表 2.2 の完全ランダム実験の誤差の自由度が 12，平均和が 7.6400 であるのに対し，表 2.6 の誤差の自由度が 11 と自由度 1 分だけ小さくなっているに対し，平均和は，2.5983 と激減し平均平方が 0.2362 と小さくなり，交互作用 A×B が有意な差（p = 0.0117）と変化している．これは，完全ランダム実験の平均和が 7.6400 から反復 R の平方和 5.0417 を差し引いた結果となっている．

交互作用を含めた場合，含めない場合の平均平方の期待値

表 2.6 で示した平均平方の期待値と JMP で出力される【期待平均平方】とは異なる．【期待平均平方】では，交互作用が有るという前提での構造が示されている．

期待平均平方

各行の平均平方の期待値を構成する各列の分散成分の係数
期待平均平方

	切片	R&変量効果	A	B	A*B
切片	0	0	0	0	0
R&変量効果	0	12	0	0	0
A	0	0	8	0	2
B	0	0	0	6	2
A*B	0	0	0	0	2

プラス1.0倍の残差誤差分散

表 2.7 に交互作用を考慮した平均平方の期待値を示した．交互作用が有意な場合に，因子 A および因子 B の主効果についての平均平方の期待値は，交互作用を含んでいること明示されている．したがって，有意な交互作用項を残したままでの主効果に対する F 検定は，不適切と思われる．

表 2.6 に示したように交互作用がないことを前提とした因子 A と因子 B の F 検定の結果は，有意な交互作用がある場合は，単独で用いてはならない．

因子 A の p 値は，交互作用項を含めているけれども無視した場合は，0.0030 と高度に有意であるが，交互作用項を誤差にプーリングした場合に p 値は 0.0297 と多く変化する．場合によっては有意な差ではなくなる場合もある．

表 2.7　交互作用項を含めた平均平方の期待値

要因	自由度	平均平方	p 値	交互作用を含めた平均平方の期待値	自由度	平均平方	p 値	交互作用を含めない平均平方の期待値
R:反復誤差 $e_{(0)}$	1	5.0417	0.0007	$\sigma^2_{(1)}+12\sigma^2_{(0)}$	1	5.0417	0.0079	$\sigma^2_{(1)}+12\sigma^2_{(0)}$
A	2	2.4267	0.0030	$\sigma^2_{(1)}+2\sigma^2_{A\times B}+8\,\sigma^2_A$	2	2.4267	0.0297	$\sigma^2_{(1)}+8\,\sigma^2_A$
B	3	0.6228	0.1019	$\sigma^2_{(1)}+2\sigma^2_{A\times B}+6\,\sigma^2_B$	3	0.6228	0.3697	$\sigma^2_{(1)}+6\,\sigma^2_B$
A×B	6	1.1461	0.0117	$\sigma^2_{(1)}+2\sigma^2_{A\times B}$	—	—	—	—
1 次誤差 $e_{(1)}$	11	0.2362		$\sigma^2_{(1)}$	17	0.5574		$\sigma^2_{(1)}$

2.4. 1 段分割実験

枝分かれ構造でのランダム化

表 2.8 に示すように，因子 A のランダム化と因子 B についてのランダム化が，枝分かれ的に行なわれることもある．ランダムな順序で，最初に選ばれた A2 水準の中で因子 B の水準をランダムな順序で実験を 4 回行ない，それぞれ 15.5，14.8，15.2，14.4 が得られ，次に，選択された A1 水準で同様に因子 B 内をランダム化を 4 回行ない 12.3，13.9，15.0，15.0 が得られ，順次，因子 A の水準を設定して行く．因子 A の同じ水準で 2 回繰返しが起きるので区別するためにブロック R を使う．ただし，A1 の中の R1 と A2 の中の R1 は，同じ水準番号を使っているが，R1 水準としての

共通点はない．これを因子 A のなかで枝分かれしたブロック R という．

見方をかえれば，因子 A について繰り返しが 2 の 1 元配置分散分析を計画し，そこで得られた中間製品を 4 等分して，因子 B の条件を変えて実験をし最終製品を作成したことになる．複数の工程からなる実験の場でいえば，上流の工程での実験が因子 A に相当し，下流ので実験が因子 B になる．

表 2.8　1 段分割実験のデータ

A	ブロック R	B1	B2	B3	B4
A1	R1（2）	12.3	13.9	15.0	15.0
	R2（3）	11.7	12.7	13.0	13.4
A2	R1（1）	15.5	14.8	15.2	14.4
	R2（5）	13.9	13.3	14.3	13.6
A3	R1（4）	13.9	13.3	14.3	12.8
	R2（6）	14.0	13.2	13.5	12.8

この実験には 2 種類の誤差，因子 A のブロック R のランダム化に伴う 1 誤差 $\varepsilon_{(1)}$，因子 B のランダム化に伴なう 2 次誤差 $\varepsilon_{(2)}$ が，各測定値に付いている．因子 A の水準を i，ブロック R の水準を j，因子 B の水準を k としたときに，因子 A の効果を α_i，ブロック R の変更に伴う 1 次誤差を $\varepsilon_{(1)ij}$，因子 B の効果を β_k，因子 A と因子 B の交互作用を $(\alpha\beta)_{ik}$，2 次誤差を $\varepsilon_{(2)ijk}$ とする．測定値を y_{ijk} は，次のような構造で示すことができる．

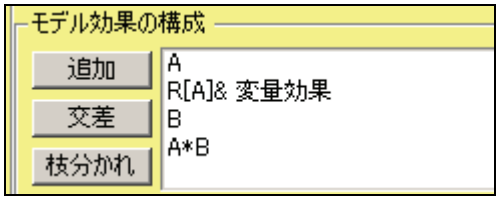
$$y_{111} = \mu + \alpha_1 + (\alpha\gamma)_{11} + \varepsilon_{(1)11} + \beta_1 + (\alpha\beta)_{11} + \varepsilon_{(2)111} = 12.3$$
$$y_{112} = \mu + \alpha_1 + (\alpha\gamma)_{11} + \varepsilon_{(1)11} + \beta_2 + (\alpha\beta)_{12} + \varepsilon_{(2)112} = 13.9$$
$$\vdots$$
$$y_{324} = \mu + \alpha_3 + (\alpha\gamma)_{32} + \varepsilon_{(1)32} + \beta_4 + (\alpha\beta)_{34} + \varepsilon_{(2)324} = 12.8$$

ただし，

$$\sum \alpha_i = 0, \qquad \sum \beta_k = 0, \qquad \varepsilon_{(1)ij} \sim \mathrm{N}(0, \sigma_{(1)}^2), \qquad \varepsilon_{(2)ijk} \sim \mathrm{N}(0, \sigma_{(2)}^2),$$

とする．1 次誤差 $\varepsilon_{(1)ij}$ は，因子 A の繰り返し付く誤差である．y_{111} と y_{112} は，同じ繰り返しで共通の 1 次誤差 $\varepsilon_{(1)11}$ が付き，それぞれのデータに因子 A と因子 B の変更に伴う 2 次誤差 $\varepsilon_{(2)111}$ と $\varepsilon_{(2)112}$ が付いている．因子 A とブロック R の交互作用と 1 次誤差 $\varepsilon_{(1)11}$ の添字が完全一致しているので分離できない．

1 次誤差には，ブロック R の主効果と交互作用 A×R を合わせるために【枝分かれ】を用いる．モデル効果の構成で，「R[A]&変量効果」が 1 次誤差となる．



「R[A]&変量効果」は，【モデル効果の構成】でブロック R を選択し，因子 A を【列の選択】で選択し，【枝分かれ】をブロック R に上書きする．さらに【属性】から【変量効果】を上書きする．因子 B および交互作用 A×B を追加して，1 段分割実験モデルを完成させる．

JMP の出力を整理した分散分析表を 表 2.9 に示す．交互作用 A×B は，平均平方の期待値から 2 次誤差を分母とすることから，$p = 0.0015$ と高度に有意な結果となる．これは，交互作用 A×B の平均平方は，これまでの分散分析表と同じ 1.1461 であるが，分母の誤差平方が 0.1150 と小さくなったためである．

表 2.9　1 段分割実験の場合の分散分析表

要因	自由度	平方和	平均平方	F 値		p 値	平均平方の期待値
A	2	4.8533	2.4267	-	1.10	0.4377	$\sigma^2_{(2)} + 4\,\sigma^2_{(1)} + 8\,\sigma^2_A$
R[A]:1 次誤差 e_1	3	6.6050	2.2017	19.14	1.00	0.0003	$\sigma^2_{(2)} + 4\,\sigma^2_{(1)}$
B	3	1.8683	0.6228	5.42	-	0.0210	$\sigma^2_{(2)} + 6\,\sigma^2_B$
A*B	6	6.8767	1.1461	9.97	-	0.0015	$\sigma^2_{(2)} + 2\sigma^2_{A\times B}$
2 次誤差 e_2	9	1.0350	0.1150	1.00	-		$\sigma^2_{(2)}$
全体	23	21.2383					

この結果は，これまでの分散分析の結果と大きく違い，因子 B の主効果，交互作用 A×B ともに統計的に有意な差となっている．どうしてなのであろうか．1 次誤差「R[A]&変量効果」の平均平方が 2.2017 と大きいことが，2 次誤差の平均平方 0.1150 が相対的に小さいことによる．

分散成分で検討してみよう．分散分析表から平均平方の期待値は $\sigma^2_{(2)} + 4\sigma^2_{(1)}$ であり，分散成分 $\sigma^2_{(1)}$ は，

$$\hat{\sigma}^2_{(1)} = \frac{(\hat{\sigma}^2_{(2)} + 4\hat{\sigma}^2_{(1)}) - \hat{\sigma}^2_{(2)}}{4} = \frac{2.2017 - 0.1150}{4} = 0.5217$$

と推定され，2 次誤差の分散成分 $\hat{\sigma}^2_{(2)} = 0.1150$ に比べて 因子 A の水準の変更に伴う分散が $\hat{\sigma}^2_{(1)} = 0.5217$ と大きい．【分散成分推定値】からも，この結果が正しいことが確かめられる．

分散成分推定値

成分	分散成分推定値	全体に対する百分率
R[A]&変量効果	0.5217	81.937
残差	0.1150	18.063
合計	0.6367	100.000

平均平方がその期待値に等しいものとして推定したものです。

変動要因の可視化に役立つ変動性図／ゲージチャート

実験結果のグラフ化は，【モデルのあてはめ】に先立って，JMP のグラフ機能を使うことを薦め

る．ここでは，【変動性図／ゲージチャート】をもちいて，因子 A の変更に伴う変動，その中で起きる因子 B の変更に伴う変動を確かめてみよう．

図 2.9 は，X 軸に，A・R・B を入れ子構造，Y 軸に測定値 y を設定し【セル平均をつなぐ】で描いた．因子 A とブロック R の変更に対し，測定値 y が大きく上下していることが読み取れる．それに対し，R1 内の因子 B の変化パターンと R2 内の変化パターンは，A の各水準内で R1 と R2 を視覚的に上下に動かしてみると，重なり合い，大きな差異がないことから，2 次誤差が小さいことが推測される．

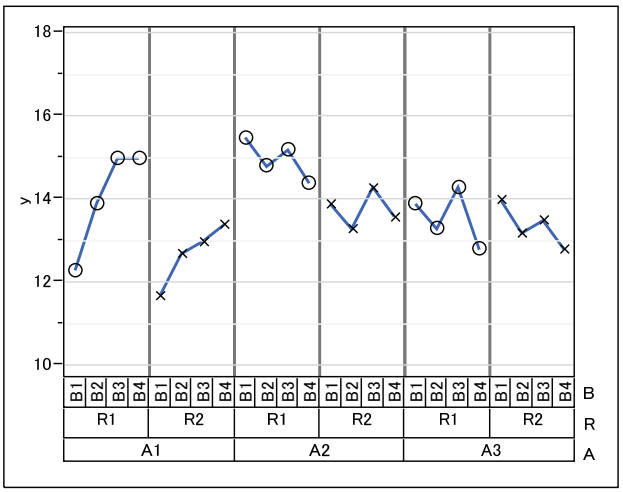


図 2.9　全ての因子の組み合わせによる変動性図

これまでのグラフは，【モデルのあてはめ】で作成されるものを示してきたのであるが，JMP には生データのグラフ表示に優れた機能があり，実験データのグラフ表示にチャレンジすることを薦める．多くの実験データについて適切なグラフ表示ができれば，【モデルのあてはめ】で統計的分析をすることなしに，実験結果の概要を容易につかむことができる．ただし，きちっとした報告書に実験結果をまとめる際には，適切な統計解析の結果を反映させることは，もちろんのことである．

2.5. 反復を伴う 1 段分割実験

2.3 節で取り上げた反復実験は，優れた実験計画であるが，ランダム化の範囲がまだ大きい．反復に 2.4 節で述べた 1 段分割と組み合わせてみよう．表 2.10 に示すようにランダム化は，反復 Rl の中で因子 A の水準変更をランダムに行ない，その中で，因子 B の 4 水準の中でランダム化を行なつている．反復 R2 でも同様なランダム化を行なう．実験が 2 日間かかるような場合に，第 1 日目に反復 R1 を，第 2 日目に R2 と設定することにより，実験日に起因する変動があっても反復 R がそれを吸収し，他の因子への影響を排除することができる．なお，第 5 章では，この実験モデルに対する EMS 法および REML 法による詳細な解析を示す．

表 2.10　反復を伴う 1 段分割実験

R：反復	A	B1	B2	B3	B4
R1	A1　(2)	12.3	13.9	15.0	15.0
	A2　(1)	15.5	14.8	15.2	14.4
	A3　(3)	13.9	13.3	14.3	12.8
R2	A1　(5)	11.7	12.7	13.0	13.4
	A2　(6)	13.9	13.3	14.3	13.6
	A3　(4)	14.0	13.2	13.5	12.8

この実験には 3 種類の誤差が，各測定値に付いている．反復 R の水準を i，因子 A の水準を j，因子 B の水準を k としたときに，反復 R の変更に伴う反復誤差を $\varepsilon_{(0)i}$，因子 A の効果を α_j，反復 R 内の因子 A の水準変更に伴う 1 次誤差を $\varepsilon_{(1)ij}$，因子 B の効果を β_k，因子 A と因子 B の交互作用を $(\alpha\beta)_{ik}$，2 次誤差を $\varepsilon_{(2)ijk}$ とする．測定値を y_{ijk} は，次のような構造で示すことができる．

$$y_{111} = \mu + \gamma_1 + \varepsilon_{(0)1} + \alpha_1 + (\alpha\gamma)_{11} + \varepsilon_{(0)11} + \beta_1 + (\alpha\beta)_{11} + \varepsilon_{(0)111} = 12.3$$

$$y_{111} = \mu + \gamma_1 + \varepsilon_{(0)1} + \alpha_1 + (\alpha\gamma)_{11} + \varepsilon_{(1)11} + \beta_2 + (\alpha\beta)_{12} + \varepsilon_{(2)112} = 13.9$$

$$\vdots$$

$$y_{234} = \mu + \gamma_2 + \varepsilon_{(0)2} + \alpha_3 + (\alpha\gamma)_{23} + \varepsilon_{(1)23} + \beta_4 + (\alpha\beta)_{34} + \varepsilon_{(2)234} = 12.8$$

ただし，

$$\sum \alpha_j = 0,\quad \sum \beta_k = 0,\quad \varepsilon_{(0)i} \sim \mathrm{N}(0, \sigma^2_{(0)}),\quad \varepsilon_{(1)ij} \sim \mathrm{N}(0, \sigma^2_{(1)}),\quad \varepsilon_{(2)ijk} \sim \mathrm{N}(0, \sigma^2_{(2)})$$

である．因子 A の後に付く 1 次誤差 $\varepsilon_{(2)ij}$ は，繰返しなしの 3 元配置分散分析とみなした場合の交互作用 R×A に対応する．

因子 A は，反復 R の変量効果を除去した R×A：1 次誤差で検定している．因子 B と交互作用 A×B は，2 次誤差で検定している．1 段分割実験に反復を含める目的は，実験日間の変動を取り除いて因子 A の評価の精度を上げたい目的もある．表 2.9 の 1 段分割の場合の因子 A の検定のための分母は R[A]で，自由度は 3，平均平方は 2.2017 であり，反復を含めた 表 2.11 の場合は，分母は

表 2.11　反復のある 1 段分割実験の場合の分散分析表

要因	自由度	平方和	平均平方	F 値		p 値	平均平方の期待値
R:反復誤差 e_0	1	5.0417	5.0417	-	6.45	0.1263	$\sigma^2_{(2)} + 4\,\sigma^2_{(1)} + 12\,\sigma^2_{(0)}$
A	2	4.8533	2.4267	-	3.10	0.2436	$\sigma^2_{(2)} + 4\,\sigma^2_{(1)} + 8\,\sigma^2_{A}$
R×A:1 次誤差 e_1	2	1.5633	0.7817	6.80	1.00	0.0159	$\sigma^2_{(2)} + 4\,\sigma^2_{(1)}$
B	3	1.8683	0.6228	5.42	-	0.0210	$\sigma^2_{(2)} + 6\,\sigma^2_{B}$
A×B	6	6.8767	1.1461	9.97	-	0.0015	$\sigma^2_{(2)} + 2\sigma^2_{A\times B}$
2 次誤差 e_2	9	1.0350	0.1150	1.00	-		$\sigma^2_{(2)}$
全体	23	21.2383					

R×A で，自由度は 2，平均平方は 0.7817 と減少している．因子 A の F 値は 3.10 と大きくなるが，分母の自由度が 2 と小さいために有意な差とはならない．

因子 A の主効果があるか検討したい場合には，この分割実験の計画は不適切である．因子 A の 3 水準に対して，それぞれ 2 回の繰返しかないごく小規模の 1 元配置の計画をしたことに相当している．それに対し因子 B の検定の分母の自由度は 9 とそこそこの大きさがあり，表 2.2 の完全ランダム化実験の分散分析表での分母の自由度 12 に比べても遜色がない．

分割実験では，因子 A の主効果があるかの検討が誤差項の自由度が小さいために期待できないが，因子 B との交互作用 A×B の検定のための誤差項の自由度は十分大きく，低次の分割に割り付けた因子 B と高次に割り付けた因子 A との交互作用があるかないかの検討には適している．

2.6. 2 方分割実験

ランダム化の更なる工夫

これまで，因子 A をランダム化の最初の分割の対象にしてきたのであるが，因子 A ではなく因子 B を最初の分割の対象としてもよい．どちらを対象にするかは，実際の実験に際して，水準の変更の難易度による．因子 A が上流の工程で，因子 B を下流の工程とした場合でも，どちらも水準の変更がしにくい場合は，同時に分割の対象とすることもできる．これは，2 方分割と呼ばれているものである．

表 2.12 は 2 方分割実験の例である．因子 A の 3 水準についての 2 回の繰返しについて，最初は A2，次は A1, . . . ，A3 とランダムな順序で実験を行ない，この時にまとめて 4 個の中間製品を（4 個の中間製品でなく 1 個の中間製品を作り，後で 4 等分しても同じである），合計 24 個の中間製品を作っておく．因子 B では，因子 A のそれぞれの水準で作成された 4 個の実験サンプルより 1 個ずつ抜き出し，合計 3 個の中間製品をまとめて因子 B の 4 水準で実験し，さらに繰り返して最終製品を作成する．

表 2.12　2 法分割実験のデータ

A	ブロック R	B1	B2	B3	B4
A1	R1（2）	12.3 [3]	13.9 [2]	15.0 [4]	15.0 [1]
	R2（3）	11.7 {5}	12.7 {8}	13.0 {6}	13.4 {7}
A2	R1（1）	15.5 [3]	14.8 [2]	15.0 [4]	14.4 [1]
	R2（5）	13.9 {5}	13.3 {8}	14.3 {6}	13.6 {7}
A3	R1（4）	13.9 [3]	13.3 [2]	14.3 [4]	12.8 [1]
	R2（6）	14.0 {5}	13.2 {8}	13.5 {6}	12.8 {7}

（ ）内の順に 6 回のランダム化，［ ］内で 4 回，{ }内で 4 回のランダム化．

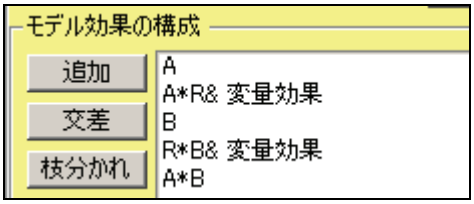
この実験には 3 種類の誤差が発生する．因子 A の変更に伴う A 方向の 1 次誤差，因子 B の変更に伴う B 方向の 1 次誤差，さらに 2 次誤差 $\varepsilon_{(2)}$ が各測定値に付いている．因子 A の水準を i，ブロック R の水準を j，因子 B の水準を k としたときに，因子 A の効果を α_i，因子 A 内のブロック R の変更に伴う 1 次誤差を $\varepsilon_{(1A)ij}$，因子 B の効果を β_k，因子 B とブロック R の変更に伴う 1 次誤差 $\varepsilon_{(1B)jk}$，最後に 2 次誤差 $\varepsilon_{(2)ijk}$ が付いている．測定値を y_{ijk} は，次のような構造で示すことができる．

$$y_{111} = \mu + \alpha_1 + (\alpha\gamma)_{11} + \varepsilon_{(1A)11} + \beta_1 + (\beta\gamma)_{11} + \varepsilon_{(1B)11} + (\alpha\beta)_{11} + \varepsilon_{(2)111} = 12.3$$
$$y_{112} = \mu + \alpha_1 + (\alpha\gamma)_{11} + \varepsilon_{(1A)11} + \beta_2 + (\beta\gamma)_{12} + \varepsilon_{(1B)12} + (\alpha\beta)_{12} + \varepsilon_{(2)112} = 13.9$$
$$\vdots$$
$$y_{324} = \mu + \alpha_3 + (\alpha\gamma)_{32} + \varepsilon_{(1A)32} + \beta_4 + (\beta\gamma)_{24} + \varepsilon_{(1B)24} + (\alpha\beta)_{34} + \varepsilon_{(2)324} = 12.8$$

ただし，

$$\sum \alpha_i = 0,\quad \sum \beta_k = 0,\quad \varepsilon_{(1A)ij} \sim \mathrm{N}(0, \sigma^2_{(1A)}),\quad \varepsilon_{(1B)jk} \sim \mathrm{N}(0, \sigma^2_{(1B)}),\quad \varepsilon_{(2)ijk} \sim \mathrm{N}(0, \sigma^2_{(2)})$$

である．【モデルの構成】は次のようする．



【期待平均平方】から因子 A は，A 方向の 1 次誤差「A*R&変量効果」を F 検定の分母，因子 B は，B 方向の 1 次誤差「B*R&変量効果」を F 検定の分母とし，交互作用 A×B は，2 次誤差での検定となることがわかる．異なる 1 次単位間の因子 A と因子 B の交互作用が 2 次単位となり，2 次誤差での F 検定となる．

期待平均平方

各行の平均平方の期待値を構成する各列の分散成分の係数
期待平均平方

	切片	A	A*R&変量効果	B	R*B&変量効果	A*B
切片	0	0	0	0	0	0
A	0	8	4	0	0	2
A*R&変量効果	0	0	4	0	0	0
B	0	0	0	6	3	2
R*B&変量効果	0	0	0	0	3	0
A*B	0	0	0	0	0	2

プラス1.0倍の残差誤差分散

【F 検定の分母】からもそれぞれの要因の分母が確認できる．2 方分割の場合の A 方向，B 方向のそれぞれの 1 次誤差は．2 次誤差を分母としていて，因子 A と因子 B の間に階層構造がないことが確認される．

F検定の分母

要因	分母平均平方	分母自由度	分母平均平方合成
A	0.7817	2	A*R&変量効果
A*R&変量効果	0.8336	7	残差
B	0.0806	3	R*B&変量効果
R*B&変量効果	0.8336	7	残差
A*B	0.8336	7	残差

【変量効果を考慮した検定】のそれぞれの要因の F 値から，分母となる平均平方を確認してみよう．因子 A の分母は 2.4267 / 3.1045 = 0.7817 と「A*R&変量効果「の平均平方，因子 B の分母は 0.6228 / 7.7310 = 0.0806 と「B*R&変量効果」の平均平方であることが確かめられる．

変量効果を考慮した検定

要因	平方和	分子平均平方	分子自由度	F値	p値(Prob>F)
A	4.8533	2.4267	2	3.1045	0.2436
A*R&変量効果	1.5633	0.7817	2	0.9377	0.4357
B	1.8683	0.6228	3	7.7310	0.0635
R*B&変量効果	0.2417	0.0806	3	0.0966	0.9594
A*B	6.8767	1.1461	6	1.3749	0.3406

これらの結果を整理して，表 2.13 の分散分析表にまとめる．2 方分割は，複雑な構造のようであるが，ブロック R を因子とみなした繰返しのない 3 元配置とと対比して

表 2.13　2 方分割実験の分散分析表

レベル	要因	自由度	平方和	平均平方		F 値		p 値
A 方向	A	2	4.8533	2.4267		-	3.10	0.2436
1 次単位	R×A：$e_{(1A)}$	2	1.5633	0.7817	0.94	-	1.00	0.4357
B 方向	B	3	1.8683	0.6228	-	7.73		0.0635
1 次単位	R×B：$e_{(1B)}$	3	0.2417	0.0806	0.10	1.00		0.9594
2 次単位	A×B	6	6.8767	1.1461	1.37	-		0.3406
	2 次誤差 $e_{(2)}$	7	5.8350	0.8336	1.00	-		
	全体	23	21.2383					

みるとよい．「R*A&変量効果」は，R×A に対応し，「R*B&変量効果」は R×B に対応している．2 次誤差は，R と R×A×B を合わせたものとなっている．

ランダム化が 4 回の場合の悩ましいデータ構造

表 2.12 の実験結果を得たときに，因子 B のランダム化は 6 個の中間製品をまとめて 4 回の水準変更でもいいのではないかと思われる．【モデル効果の構成】で，「R*B&変量効果」を除いても解析結果が出力される．因子 B の検定の分母は 2 次誤差となり，なんら問題はないように思われる．

表 2.14 に因子 A とブロック R を取り除いたデータを示す．繰返しが 6 の 1 元配置分散分析として解析していいのだろうか．ランダム化は因子 B について 4 回，24 個の中間製品を 4 等分し，6 個をまとめて処理したとしよう．因子 B が焼成工程であったとすれば温度設定に時間がかかるので，

まとめて焼成するのは当然と思われる．

実験で何を検討したいのか，結果が知りたいことのを正しく反映しているのかを吟味してみよう．6 個のデータがあるが，これは因子 B の各水準内での測定の繰り返しである．そうすると，実験結果としては 6 個のデータの平均値をデータとすべきではないのだろうか．そうすると，因子 B の各水準に対して 1 個のデータしか得られなかったことになり，繰り返し 6 の 1 元配置の問題として扱うことと矛盾する．

表 2.14　悩ましいデータ

R’	B1（2）	B2（4）	B3（1）	B4（3）
1	12.3	13.9	15.0	15.0
2	11.7	12.7	13.0	13.4
3	15.5	14.8	15.2	14.4
4	13.9	13.3	14.3	13.6
5	13.9	13.3	14.3	12.8
6	14.0	13.2	13.5	12.8

2.7. 反復のある 2 方分割実験

2.5 節で示した反復を伴う 1 段分割実験は，実験の場の揺らぎを避け，交互作用 A×B をうまく抽出できた．2.6 節で示した 2 方分割の場合に反復 R を組み込むことは容易である．反復．R1 の中で因子 A の水準変更を行い，第 1 回目は A2 水準であるが，この時にまとめて 4 個の中間製品を作っておく．次に A1 水準でまとめて 4 個の中間製品をさらに A3 水準で 4 個の実験サンプルを作り，合計 12 個の中間製品を作る（中間製品が分割可能ならば 1 個の中間製品を作り，後で 4 等分してもよい）．因子 B では，因子 A のそれぞれの水準で作成された 4 個の中間製品より 1 個つつ抜き出し，合計 3 個の中間製品に対し，まとめて実験して最終製品を作成し結果を得る．

表 2.15　反復のある 2 方分割実験データ

R	A	B1	B2	B3	B4
		[3]	[2]	[4]	[1]
R1	A1　(2)	12.3	13.9	15.0	15.0
	A2　(1)	15.5	14.8	15.2	14.4
	A3　(3)	13.9	13.3	14.3	12.8
		[7]	[5]	[8]	[6]
R2	A1　(5)	11.7	12.7	13.0	13.4
	A2　(6)	13.9	13.3	14.3	13.6
	A3　(4)	14.0	13.2	13.5	12.8

この実験には 4 種類の誤差が発生する．反復 R に伴う反復誤差，因子 A の変更に伴う A 方向の 1 次誤差，因子 B の変更に伴う B 方向の 1 誤差，さらに 2 次誤差が各測定値に付いている．反復 R

の水準をi，因子 A の水準をj，因子 B の水準をkとしたときに，反復 R の効果をγ_i　因子 A の効果をα_j，反復 R の変更に伴う 2 次誤差を$\varepsilon_{(A2)ij}$，因子 B の効果をβ_k，因子 B と反復 R の変更に伴う 2 次誤差$\varepsilon_{(B2)ik}$，最後に 3 次誤差$\varepsilon_{(3)ijk}$が付いている．測定値をy_{ijk}は，次のような構造で示すことができる．

$$y_{111} = \mu + \gamma_1 + +\varepsilon_{(0)1} + \alpha_1 + (\gamma\alpha)_{11} + \varepsilon_{(1A)11} + \beta_1 + (\gamma\beta)_{11} + \varepsilon_{(1B)11} + (\alpha\beta)_{11} + \varepsilon_{(2)111} = 12.3$$

$$y_{112} = \mu + \gamma_1 + +\varepsilon_{(0)1} + \alpha_1 + (\gamma\alpha)_{11} + \varepsilon_{(1A)11} + \beta_2 + (\gamma\beta)_{12} + \varepsilon_{(1B)12} + (\alpha\beta)_{12} + \varepsilon_{(2)112} = 13.9$$

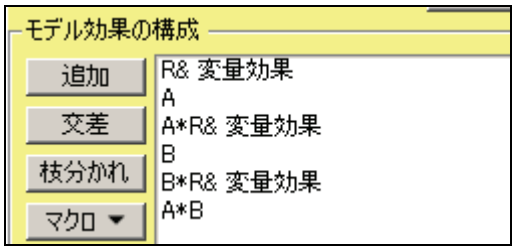
$$\vdots$$

$$y_{234} = \mu + \gamma_2 + +\varepsilon_{(0)2} + \alpha_3 + (\gamma\alpha)_{23} + \varepsilon_{(1A)23} + \beta_4 + (\gamma\beta)_{24} + \varepsilon_{(1B)24} + (\alpha\beta)_{34} + \varepsilon_{(2)234} = 12.8$$

ただし，

$$\sum \alpha_j = 0,\quad \sum \beta_k = 0,\quad \varepsilon_{(0)i} \sim \mathrm{N}(0, \sigma^2_{(0)}),\quad \varepsilon_{(1A)ij} \sim \mathrm{N}(0, \sigma^2_{(1A)}),\quad \varepsilon_{(1B)ik} \sim \mathrm{N}(0, \sigma^2_{(1B)}),\quad \varepsilon_{(2)ijk} \sim \mathrm{N}(0, \sigma^2_{(2)})$$

とする．【モデルの構成】は次のように設定する．



【期待平均平方】と【F 検定の分母】で，それぞれの因子の検定の分母を確認しよう．因子 A は自由度 2 の「A*R&変量効果」，因子 B は自由度 3 の「B*R&変量効果」，交互作用 A×B は自由度 6 の残差誤差分散が，それぞれの分母となることがわかる．反復 R の平均平方には，「A*R&変量効果＋B*R&変量効果」，「残差誤差分散」の 3 つが含まれているので，それらの平均平方の期待値を

A*R&変量効果 ＋B*R&変量効果－1*残差誤差分散

と合成さていることがわかる．「－1*残差誤差分散」となっているのは，「A*R&変量効果」と「B*R&変量効果」にそれぞれ，「プラス 1.0 倍の残差誤差分散」がついていて，2.0 倍となるので，1.0 倍となるようにマイナス 1.0 倍を差し引いているためである．

期待平均平方

各行の平均平方の期待値を構成する各列の分散成分の係数
期待平均平方

	切片	R&変量効果	A	A*R&変量効果	B	B*R&変量効果	A*B
切片	0	0	0	0	0	0	0
R&変量効果	0	12	0	4	0	3	0
A	0	0	8	4	0	0	2
A*R&変量効果	0	0	0	4	0	0	0
B	0	0	0	0	6	3	2
B*R&変量効果	0	0	0	0	0	3	0
A*B	0	0	0	0	0	0	2

プラス1.0倍の残差誤差分散

F検定の分母

要因	分母平均平方	分母自由度	分母平均平方合成
R&変量効果	0.73	1.7158	A*R&変量効果+B*R&変量効果-1*残差
A	0.78167	2	A*R&変量効果
A*R&変量効果	0.13222	6	残差
B	0.08056	3	B*R&変量効果
B*R&変量効果	0.13222	6	残差
A*B	0.13222	6	残差

分散成分の推定値を次に示す.「B*R&変量効果」の分散成分の推定値が負となっている. これは,「B*R&変量効果」の平均平方が，残差誤差分散に比べて小さい場合に，分散成分の計算をする際に負となるためである．このような場合には,」B*R&変量効果」の存在を見出せなかったことになり，【モデル効果の構成】から除いて再度解析を行なうことになる．この場合は,「B*R&変量効果」は残差誤差分散に自動的に含まれる.

分散成分推定値

成分	分散成分推定値	全体に対する百分率
R&変量効果	0.359306	56.435
A*R&変量効果	0.162361	25.502
B*R&変量効果	-0.01722	-2.705
残差	0.132222	20.768
合計	0.636667	100.000

平均平方がその期待値に等しいものとして推定したものです。

分散分析

要因	自由度	平方和	平均平方	F値
モデル	17	20.445000	1.20265	9.0957
誤差	6	0.793333	0.13222	p値(Prob>F)
全体(修正済み)	23	21.238333		0.0060*

変量効果を考慮した検定

要因	平方和	分子平均平方	分子自由度	F値	p値(Prob>F)
R&変量効果	5.04167	5.04167	1	6.9064	0.1394
A	4.85333	2.42667	2	3.1045	0.2436
A*R&変量効果	1.56333	0.78167	2	5.9118	0.0381*
B	1.86833	0.62278	3	7.7310	0.0635
B*R&変量効果	0.24167	0.08056	3	0.6092	0.6332
A*B	6.87667	1.14611	6	8.6681	0.0094*

これらの結果を表 2.16 に示すように分散分析表にまとめる．モデルを再構築した場合は，その前のモデルを分散分析表に残しておき，あらたに計算した結果を追加することが望ましい．ここでは，前の結果は片括弧（ で印を付けて残し，再計算された 3 次誤差の自由度は $3+6=9$ と変化しいたと解しやすい．この結果は，表 2.11 に示した反復のある 1 段分割実験の分散分析表に帰着する.

表 2.16　反復を伴う 2 方分割（プーリング前）

レベル	要因	自由度	平方和	平均平方	*F* 値			*p* 値
	R : 反復誤差 $e_{(0)}$	1	5.0417	5.0417	(6.91)			0.1394
A 方向	A	2	4.8533	2.4267	-		3.10	0.2436
1 次単位	誤差 $e_{(1A)}$	2	1.5633	0.7817	5.91		1.00	0.0381
B 方向	B	3	1.8683	0.6228	-	7.73		0.0635
1 次単位	誤差 $e_{(1B)}$	3	0.2417	0.0806	0.61	1.00		0.6332
2 次単位	A×B	6	6.8767	1.1461	8.67	-	-	0.0094
	2 次誤差 $e_{(2)}$	6	0.7933	0.1322	1.00	-	-	-
	全体	23	21.2383					

表 2.17　反復を伴う 2 方分割（プーリング後）

レベル	要因	自由度	平方和	平均平方	*F* 値		*p* 値
	R : 反復誤差 $e_{(0)}$	1	5.0417	5.0417	-	6.45	0.1263
1 次単位	A	2	4.8533	2.4267	-	3.10	0.2436
	誤差 $e_{(1A)}$	2	1.5633	0.7817	6.80	1.00	0.0159
2 次単位	B	3	1.8683	0.6228	5.42		0.0210
	A×B	6	6.8767	1.1461	9.97		0.0015
	2 次誤差 $e_{(2)}$	9	1.0350	0.1150	1.00		
	全体	23	21.2383				

プーリング後は，表 2.11 と同じ結果となる．

3. EMS（従来法）に変わる REML 法

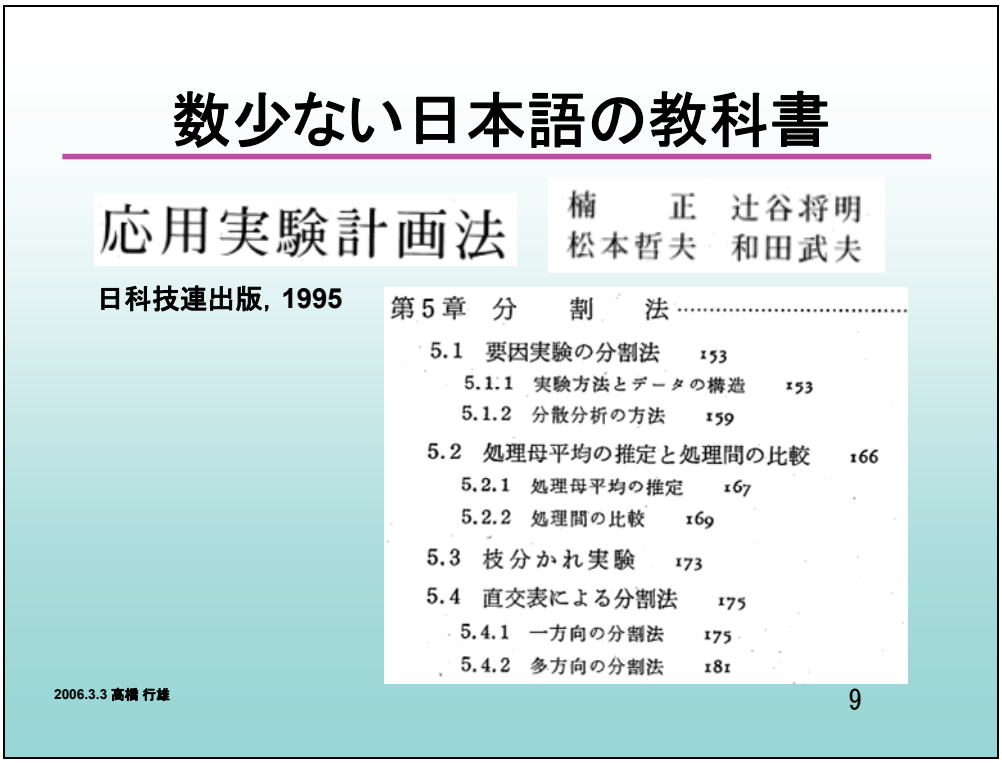
3.1. 要因配置型の分割実験

EMS（従来法）の限界

第 2 章で各種の分割実験が変量効果の使い分けによって実現でき，分散分析表を作成する方法を述べてきた. データが揃っている要因配置型をベースにした分割実験に対して, 因子 A, 因子 B, あるいは，交互作用 A×B など固定効果の F 検定の結果は，REML 法（制限つき最尤法）の結果と一致する．しかしながら，図 2.7 で指摘したように，EMS 法の場合は，水準平均，水準間の差，それらの標準誤差 SE の算出に際して，変量効果を適切に反映していないこともあり，95%信頼区間，および検定結果も不適切であることを指摘した.

2006 年の JMP er's Meeting での「JMP による各種の分割実験データの解析」において, 水準平均, その 95%信頼区間の求め方について，楠ら（1995）が示している方法と結果を引用し，REML 法による水準平均の 95%信頼区間が，従来の方法と一致するが，EMS 法の場合には一致しないことを示した．スライドを引用して説明をする.

分割実験に詳しい日本語の教科書



スライド 9 で，楠ら（1995）に出版された「応用実験計画法」が分割法について詳しく書かれている入手可能な教科書であることをし示した．以下，楠本という.

[例題 5.1] 単一分割法

生薬エキスの抽出実験で，有効成分の含有率を高めるため，抽出温度Aを4水準，添加剤Bを3水準にとり実験することになった．抽出温度の水準の変更には時間を要するので，Aを1次因子，Bを2次因子とする単一分割法を採用して2反復の実験を行った．測定結果を表5.1に示す．

表 5.1 データと2元表(エキス固形分1g当たりの有効成分の含量 mg/g)

	R_1				R_2				$R_1+R_2(T_{ij.})$				計
	A_1	A_2	A_3	A_4	A_1	A_2	A_3	A_4	A_1	A_2	A_3	A_4	$T_{.j.}$
B_1	19	23	27	26	21	26	30	24	40	49	57	50	196
B_2	23	21	19	27	21	25	23	26	44	46	42	53	185
B_3	21	18	22	24	20	20	25	27	41	38	47	51	177
1次プロット計 $T_{i\cdot k}$	63	62	68	77	62	71	78	77	125	133	146	154	
反復計 $T_{\cdot\cdot k}$	270				288								$T=558$

2006.3.3 高橋 行雄 15

楠本での分割実験の例題として，反復 R が 2 回ある単一分割が示されている．実験は，因子 A をランダムに設定し，それぞれごとに因子 B をランダムに設定し，それを 2 回反復した．

分散分析表

分散分析

要因	自由度	平方和	平均平方	F値
モデル	15	211.83333	14.1222	8.9193
誤差	8	12.66667	1.5833	p値(Prob>F)
全体(修正済み)	23	224.50000		0.0020＊

変量効果を考慮した検定

要因	平方和	分子平方平均	分子自由度	F値	p値(Prob>F)
R	13.5	13.5	1	2.4059	0.2187
A	84.1667	28.0556	3	5.0000	0.1096
R*A&変量効果	16.8333	5.61111	3	3.5439	0.0677
B	22.75	11.375	2	7.1842	0.0164＊
A*B	74.5833	12.4306	6	7.8509	0.0052＊

- **1 次単位の因子は，R*A&変量効果で検定されている．**

2006.3.3 高橋 行雄 19

楠本で示されている分散分析表が，EMS 法による【分散分析】，【変量効果を考慮した検定】と一致することを示した．

Rも変量とした分散成分の推定

EMS法（期待平均平方）

分散成分推定値

成分	分散成分推定値
R&変量効果	0.657407
R*A&変量効果	1.342593
残差	1.583333
合計	3.583333

REML法

REML分散成分の推定値

変量効果	分散比	分散成分
R	0.4152047	0.6574074
R*A	0.8479532	1.3425926
残差		1.5833333
合計		3.5833333

どちらの方法でも分散成分の推定値は同じである.

2006.3.3 高橋 行雄　41

楠本では，反復 R を固定効果とみなす場合と，変量効果とする場合の 2 通りの方法が示されている．反復 R を変量効果とした場合について，EMS 法および REML 法で推定された分散成分が，楠本と一致することを示した．

母平均の推定，95%信頼区間

楠本で母平均の95%は

5.2.1　処理母平均の推定

A と B の 2 因子による単一分割法の 2 反復を例にとって説明する．データの構造は(5.1.4)であり，水準数は a と b，反復数を n とすると，1 次単位の総数は $N_1=an$，実験の総数は $N=abn$ である．例題は［例題 5.1］を使う．

図 5.4:　$y_{ijk}=\mu+r_k+\alpha_i+e_{(1)ik}+\beta_j+(\alpha\beta)_{ij}+e_{(2)ijk}$　(5.1.4)

図 5.5:　$y_{ijk}=\mu+\alpha_i+e_{(1)ik}+\beta_j+(\alpha\beta)_{ij}+e_{(2)ijk}$　(5.1.5)

2006.3.3 高橋 行雄　25

因子 A と因子 B の 2 因子についてそれらの母平均の 95%信頼区間を算出するために，データの構造が示されている．

推定の考え方

点推定量は $\hat{\mu}(A_iB_j)=\bar{y}_{ij.}$ であるが，その構造

$$\bar{y}_{ij.}=\mu(+\bar{r})+\alpha_i+\bar{e}_{(1)i.}+\beta_j+(\alpha\beta)_{ij}+\bar{e}_{(2)ij.} \qquad (5.2.1)$$

より，誤差は $\bar{e}_{(1)i.}$ と $\bar{e}_{(2)ij.}$ の分散を含む．反復 R の分散を考慮しない場合には，(5.2.1)の $\bar{r}$ を省く．分散分析表(表 5.2)の E(ms)欄より，$\hat{\sigma}_1^2=(V_{e(1)}-V_{e(2)})/b$，$\hat{\sigma}_2^2=V_{e(2)}$ であるから，推定量の分散は，

$$\hat{V}\{\hat{\mu}(A_iB_j)\}=\underline{\frac{1}{n}\hat{\sigma}_1^2+\frac{1}{n}\hat{\sigma}_2^2}=\frac{1}{n}\cdot\frac{V_{e(1)}-V_{e(2)}}{b}+\frac{V_{e(2)}}{n}$$

$$=\frac{a}{N}V_{e(1)}+\frac{a(b-1)}{N}V_{e(2)} \qquad (5.2.2)$$

となるが，R を考慮して $\bar{r}$ を加えると，$\hat{\sigma}_R^2=(V_R-V_{e(1)})/ab$ を用いて，

$$\hat{V}\{\hat{\mu}(A_iB_j)\}=\underline{\frac{1}{n}\hat{\sigma}_R^2+\frac{1}{n}\hat{\sigma}_1^2+\frac{1}{n}\hat{\sigma}_2^2}$$

$$=\frac{1}{N}V_R+\frac{a-1}{N}V_{e(1)}+\frac{a(b-1)}{N}V_{e(2)} \qquad (5.2.3)$$

2006.3.3 高橋 行雄

26

推定の考え方の例として，交互作用 A×B があった場合に，それらを組み合わせた点推定量（母平均）の構造が示され，反復 R の分散を考慮しない場合，考慮する場合の式が示されている．分散成分の推定値を用いた場合には，複数の分散成分の合計を反復数 n で割ることが示されている．

信頼区間の計算

［例題 5.1］では，$a=4, b=3, n=2, N=24$ である．表 5.1 および表 5.2 より，$\hat{\mu}(A_3B_1)=\bar{y}_{31.}=57/2=28.5$ であり，R を考慮すると，

$$\hat{V}\{\hat{\mu}(A_3B_1)\}=\frac{1}{24}\times 13.500+\frac{3}{24}\times 5.611+\frac{8}{24}\times 1.583$$

$$=0.5625+0.7014+0.5278=1.7917 \qquad (5.2.6)$$

$$\frac{1.792^2}{\phi^*}=\frac{0.5625^2}{1}+\frac{0.7014^2}{3}+\frac{0.5277^2}{8} \qquad (5.2.7)$$

$$\phi^*=6.23,\quad t(6.23, 0.05)=2.428 \qquad (5.2.8)$$

$$\{\mu(A_3B_1)\}_{U,L}=28.5\pm 2.428\sqrt{1.7917}=25.3, 31.7 \qquad (5.2.9)$$

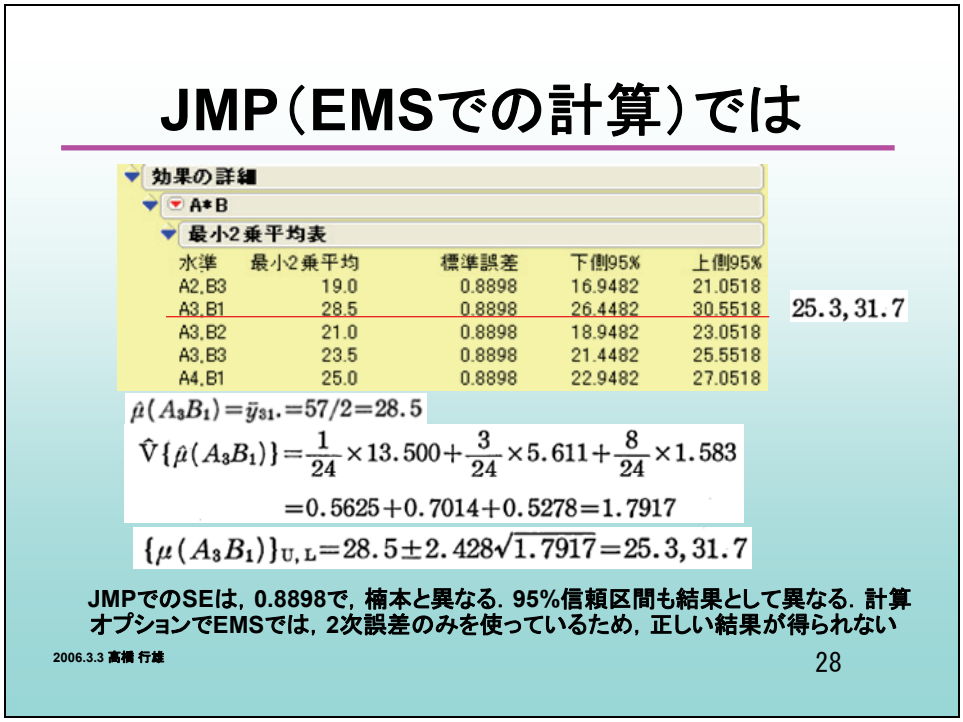
となる．(5.2.7)には(4.1.7)を用いている．

2006.3.3 高橋 行雄

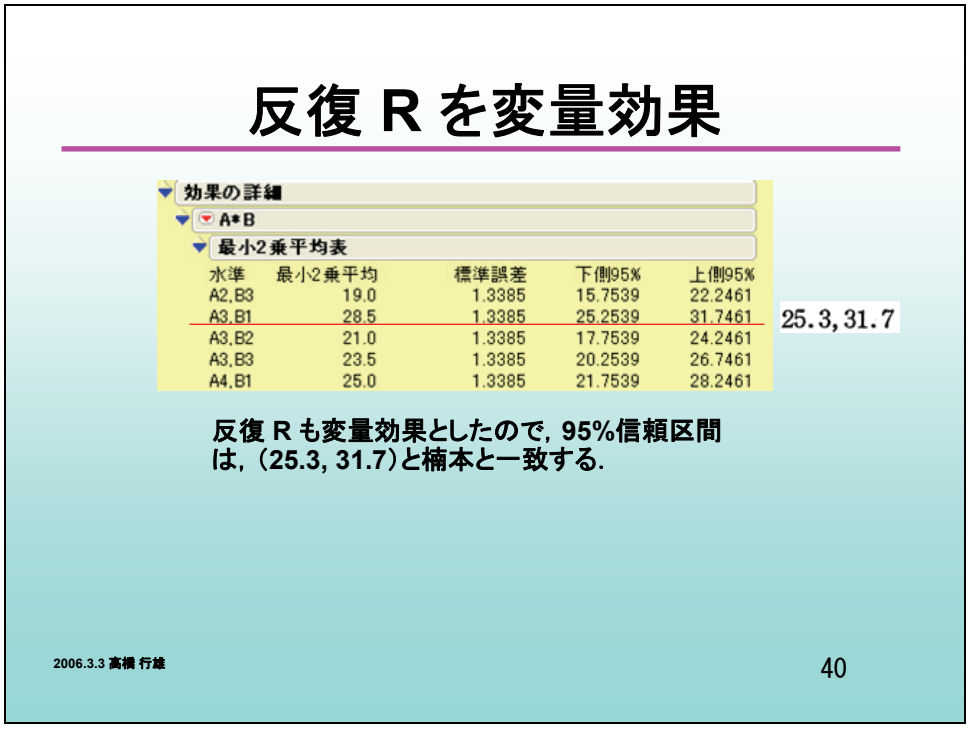
27

計算例としては，平均平方を用い，複数の分散を合成しているので，その自由度を Satterthwaite の方法で求めている． A_3B_1 の母平均は，28.5，95%信頼区間は（25.3，31.7）である．

JMPによる母平均の95%信頼区間の推定



EMS 法での 95%信頼区間は（26.4，30.6）で楠本の（25.3，31.7）より狭くなっている．これは，である．SE が 0.8898 となっているので，2 次誤差のみを使ってことがわかる．したがって，EMS 法での信頼区間は，過小評価となるので不適切である．

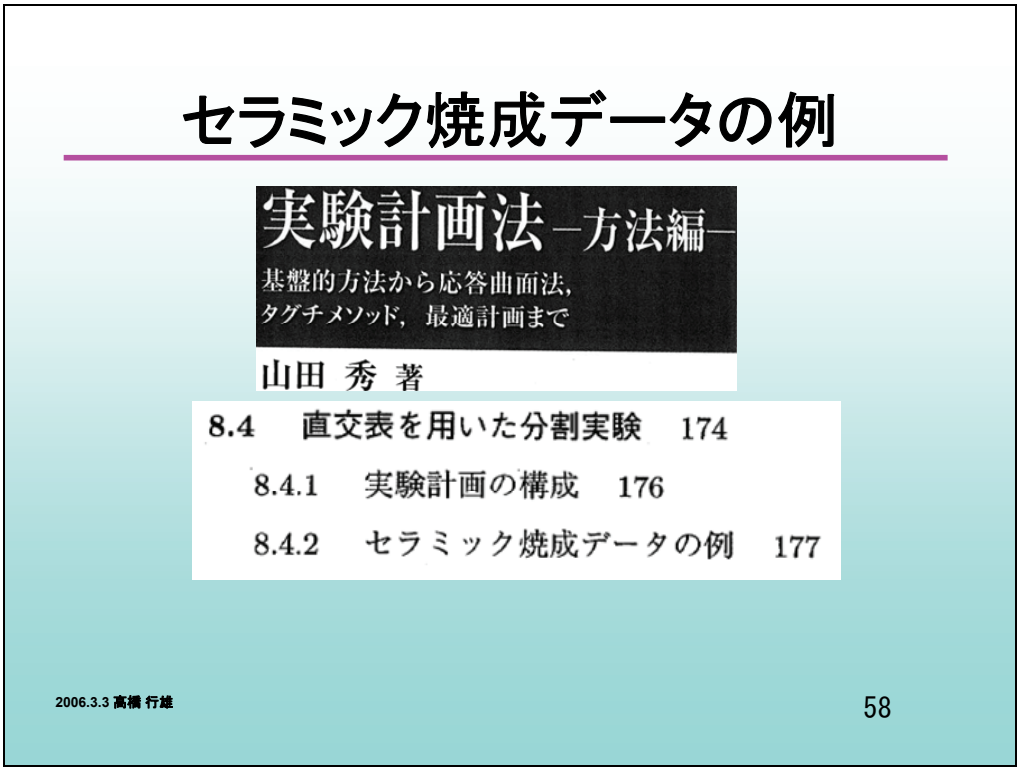


スライド 40 は，反復 R を変量効果とし REML 法を用いた場合の 95%信頼区間が楠本の結果と一致することを示した．

3.2. 直交表を用いた分割実験

直交表を用いた分割実験については，2006 年の JMP er's Meeting での「JMP による各種の分割実験データの解析」において，山田秀（2004）の「セラミック焼成データ」について 1）MP では，どのようにしたら解析できるのであろうか，2）現行水準と最適水準の母平均と 95%信頼区間はどのようにしたら求められるのだろうか，と今後の課題としただけであった．後日，この問題について JMP での解析方法について追加資料を作成した．楠本にも直交表を用いた分割実験の事例があるが，2006 年当時に出版された山田本を事例として示した．

L_{16} 直交表を用いたセラミックの焼成実験



山田本で，直交表を用いた分割実験の基本的な構成方法は次のように述べられている．.

1） 1 次因子，反復を上位の群に割付ける．その際 1 次誤差を求める列を確保するため，1 次因子，反復の列を割付けた群に何も割付けない列を確保しておく．また，反復ではなく繰返しを行う場合であっても，1 次誤差と 1 次因子の効果を分離するために，1 次因子のそれぞれの水準が少なくとも 2 回ずつは再設定されるようにする．

2） 2 次因子を，1 次因子や反復を割付けた群よりも高次の群に割付ける．その際，交互作用の出現のパターンが次のとおりであることを考慮する必要がある．（以下略）

表 8.14　直交表を用いた分割実験の例：セラミック焼成データ

	[1]	[2]	[3]	[4]	[5]	[6]	[7]	[8]	[9]	[10]	[11]	[12]	[13]	[14]	[15]		
	R	A	$E_{(1)}$	B	$E_{(1)}$	$A\times B$	$C\times D$	C	F	$A\times C$	$B\times D$	$B\times C$	$E_{(2)}$	$E_{(2)}$	D	順序	y
1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	5	80.2
2	1	1	1	1	1	1	1	2	2	2	2	2	2	2	2	6	79.9
3	1	1	1	2	2	2	2	1	1	1	1	2	2	2	2	1	85.4
4	1	1	1	2	2	2	2	2	2	2	2	1	1	1	1	2	80.9
5	1	2	2	1	1	2	2	1	1	2	2	1	1	2	2	3	83.1
6	1	2	2	1	1	2	2	2	2	1	1	2	2	1	1	4	84.3
7	1	2	2	2	2	1	1	1	1	2	2	2	2	1	1	8	80.4
8	1	2	2	2	2	1	1	2	2	1	1	1	1	2	2	7	86.9
9	2	1	2	1	2	1	2	1	2	1	2	1	2	1	2	12	80.8
10	2	1	2	1	2	1	2	2	1	2	1	2	1	2	1	11	73.8
11	2	1	2	2	1	2	1	1	2	1	2	2	1	2	1	9	80.8
12	2	1	2	2	1	2	1	2	1	2	1	1	2	1	2	10	81.9
13	2	2	1	1	2	2	1	1	2	2	1	1	2	2	1	15	80.7
14	2	2	1	1	2	2	1	2	1	1	2	2	1	1	2	16	86.6
15	2	2	1	2	1	1	2	1	2	2	1	2	1	1	2	13	80.4
16	2	2	1	2	1	1	2	2	1	1	2	1	2	2	1	14	81.0
成分	a	b	a b	c	a c	b c	a b c	d	a d	b d	a b d	c d	a c d	b c d	a b c d		
群	1	2		3				4									

2006.3.3 高橋 行雄

60

あるセラミック焼成工程では，焼成後のセラミック強度を改善するために 2 水準の因子として A：焼成温度，B：焼成機械種類，C：成分 C 配合量，D：成分 D 配合量，F：成分 F 配合量 を取り上げ実験を行い, 最も強度が高くなる焼成温度, 原料の配合を求めたい.水準設定の困難さから, A, B は 1 次因子として，残りを 2 次因子として分割実験を行いたい．また，1 日あたりの実験回数の上限が 8 回であることから，日をブロック因子としてこれを用いて反復 R を導入したい．さらに考慮すべき交互作用として A×B，A×C，B×C，B×D，C×D があげられる．応答 y は,強度の指数であり,値は大きいほど好ましい.

分散分析表

表 8.15　セラミック焼成データの分散分析表（プーリング前）

要因	S	ϕ	V	F	p
R	14.251	1	14.251	18.373	0.050
A	24.256	1	24.256	31.272	0.031
B	4.306	1	4.306	5.551	0.143
$A \times B$	25.756	1	25.756	33.206	0.029
$C \times D$	3.706	1	3.706	4.778	0.160
1 次誤差	1.551	2	0.776	1.371	0.422
C	0.766	1	0.766	1.354	0.365
F	0.331	1	0.331	0.585	0.524
D	32.776	1	32.776	57.946	0.017
$A \times C$	38.751	1	38.751	68.509	0.014
$B \times C$	0.951	1	0.951	1.681	0.324
$B \times D$	0.001	1	0.001	0.001	0.977
2 次誤差	1.131	2	0.566		
計	148.529	15			

- JMPでは，どのようにしたら解析できるのであろうか．
- 現行水準と最適水準の母平均と95%信頼区間はどのようにしたら求められるのだろうか．

2006.3.3 高橋 行雄

61

山田本では，プーリング前の分散分析表と，プーリング後の分散分析表が示されているだけで，あり，楠本の事例で示したような JMP での解析と対比することが出来ないために．「JMP では，どのようにしたら解析できるのであろうか．」とした．

直交表を用いた分割実験の手作業による解析が確立している．まず，［1］列から［15］列に対応する平方和を算出し，分散分析表に 1 次単位に割り付けた要因を拾い集め，残りの列の平方和を加えて 1 次誤差とし，F 検定をする．次に，2 次単位に割り付けた要因を拾い集め，残りを 2 次誤差とし，F 検定を行なう．直交表を用いた分割実験の解析を行なえる統計ソフトは，この手作業の手順に準拠することが一般的である．

JUM の「モデルのあてはめ」は，従来の手作業に対応する解析手順と異なる．1 次単位での 1 次誤差を変量効果とするのであるが，複数の列を合わせて「誤差」とする機能はない．L_{16} 直交表を 2×2×2×2 の要因配置実験とし，（2×2×2＝8）を 1 次単位としてランダム化し，その中で 2 水準の因子 C のランダム化を行なった考えよう．

表 3.1 に示すようにこの実験は大きさ 2 の乱塊法の実験となり，ブロック番号を変量効果する実験となる．直交表の実験では，ブロックに 1 次因子を割り付け，考慮する交互作用を決めたことに相当する．反復 R（固定効果とみなす）と因子 A および因子 B を割り付け【モデル効果の構成】順次追加すると変量効果としてのブロック番号は 1 から 8 まで 8 水準あるが，その自由度は 7 から 5 に自動的に減少する．さらに因子 B を追加し，考慮する交互作用 A×B，C×D を追加すると，変量効果としてのブロックの自由度は 5 から 2 に自動的に減少する．

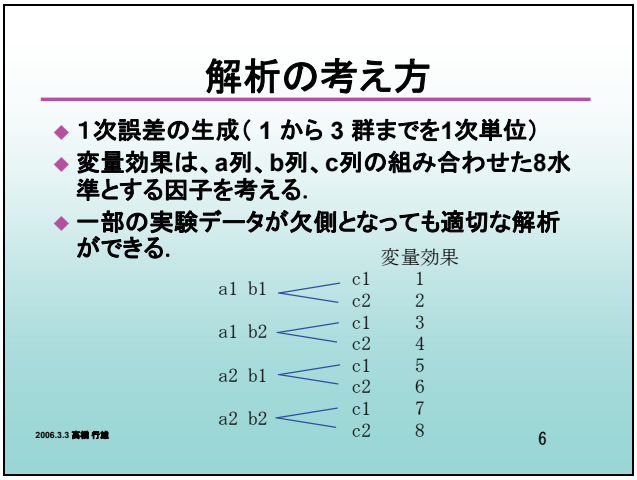
2 次単位には，最初は因子 C のみを割り付け，因子 F と因子 D を追加し，考慮する交互作用として，さらに A×C，B×C，B×D を追加する．2 列分ある 2 次誤差は，自動的に拾い集めてくれる．

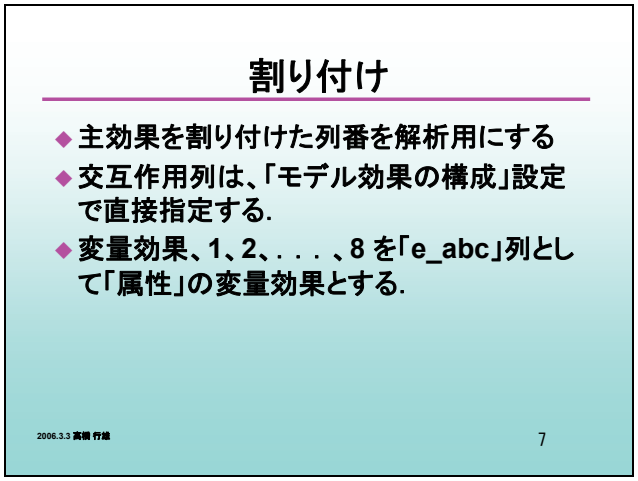
表 3.1　乱塊法実験

成分 a R	成分 b A	成分 c B	ブロック 番号	2 次単位 C1	成分 d C2
1	1	1	1	80.2	79.9
		2	2	85.4	80.9
	2	1	3	83.1	84.3
		2	4	80.4	86.9
2	1	1	5	80.8	73.8
		2	6	80.8	81.9
	2	1	7	80.7	86.6
		2	8	80.4	81.0

2006 年当時のスライドでは，「ブロック番号」ではなく，「変量効果は，a 列、b 列、c 列の組み合わせた 8 水準とする因子を考える」としたことと同じ考え方である．

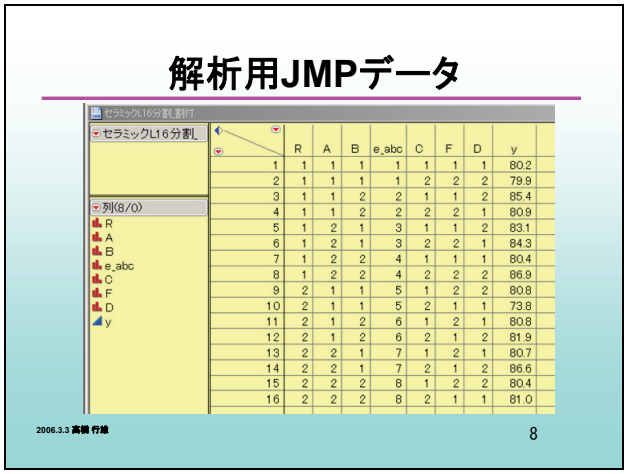
JMPでの解析

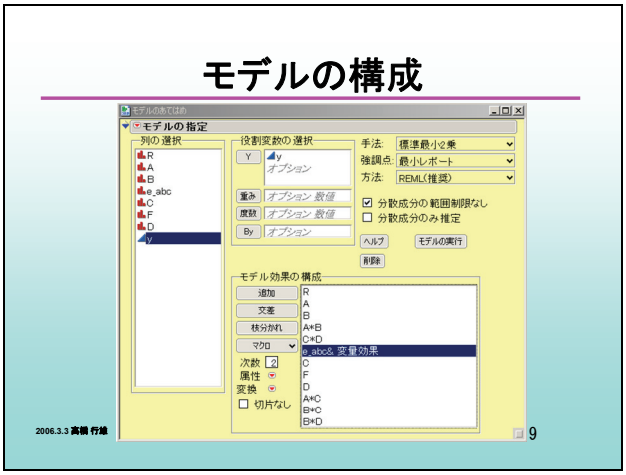




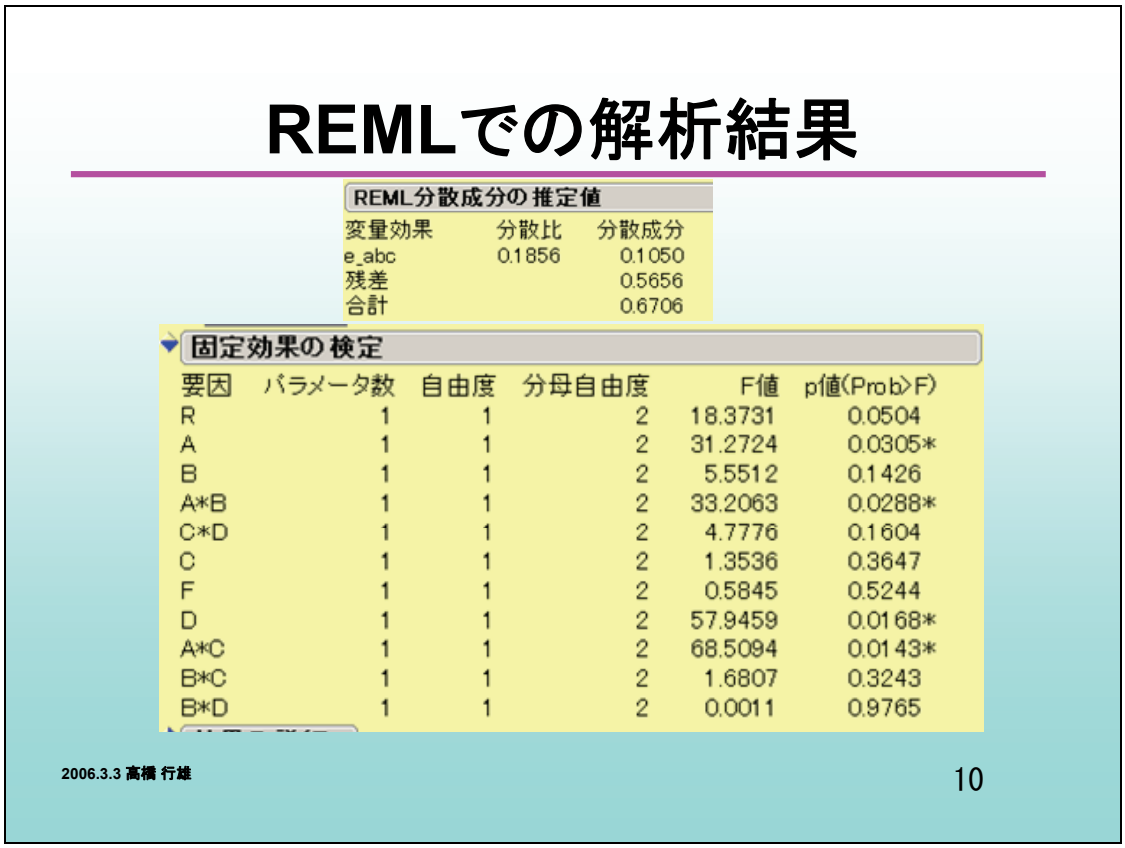
直交表を用いた分割実験データを JMP で解析するためには，1 次単位全体を変量効果とするような設定が必要となる．セラミック焼成実験では，3 つの成分 *a*，*b*，*c* の 8 通りの組み合わせについて通し番号をつけ，これに変数名をつけて変量効果とする．主効果を割り付けた列について解析用のデータとし，考慮する交互作用については，【モデル効果の構成】で交互作用を直接設定する．

2 次単位に割り付けた因子も主効果を割り付けた列のみ直交表から抜き出してデータとする．交互作用は，1 次単位と同様に直接設定する．





【モデル効果の構成】では，1 次単位の主効果，2 因子交互作用，*a*，*b*，*c* の組み合わせ変数を *e_abc* として変量効果として修飾し，引き続き 2 次単位の主効果，2 因子交互作用を設定する．直交表で誤差とする列は，これは無視するが．変量効果 *e_abc* が 1 次誤差として自動的に拾い集めてくれる．



REML 法での解析結果を示す．山田本の 2 次誤差の平均平方 V=0.566 は，REML 分散成分の推定値の分散成分 0.5656 と同じである．1 次誤差の平均平方 V=0.776 は，*e_abc* の分散成分（0.776 - 0.5656）/2 = 0.1050 に対応している．1 次単位の要因について F 検定のための分母の自由度 2 は，山田本の 1 次誤差の自由度 $\phi = 2$ に対応している．

1次単位内のプーリング

- 「モデルの効果の構成」で、有意でない効果を除く.
- 1次単位、C*D をモデルから除く.
- 2次単位、F、B*C、B*D もモデルから除く.

2006.3.3 高橋 行雄　11

直交表実験では，多くの要因をできるかぎり取り上げるために，誤差の自由度が極端に小さくなる．そのために，平均平方の小さい要因を誤差にプーリングし誤差の自由度を上げることが手順化されている．JMP では，【モデルの効果の構成】で、有意でない効果を除くことで，プーリングを実施する．1 次単位では，C×D をモデルから除き，2 次単位では， F、B×C、B×D をモデルから除く．

プーリング後の分散分析表

REML分散成分の推定値

変量効果	分散比	分散成分
e_abc	1.3154	0.6348
残差		0.4826
合計		1.1175

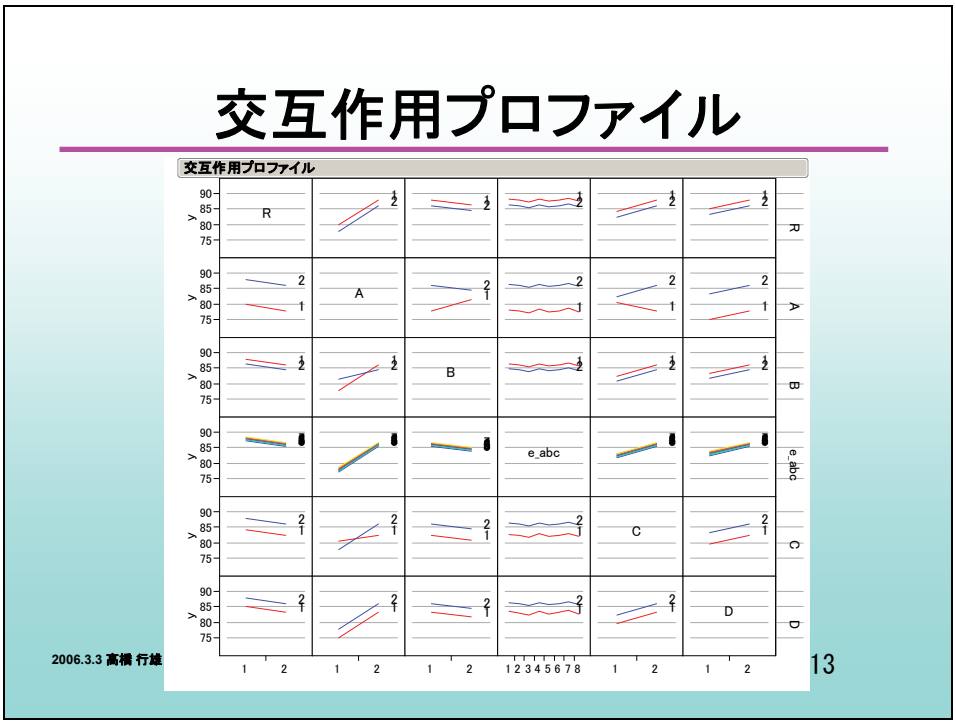
固定効果の検定

要因	パラメータ数	自由度	分母自由度	F値	p値(Prob>F)
R	1	1	3	8.1326	0.0650
A	1	1	3	13.8422	0.0338*
B	1	1	3	2.4571	0.2150
A*B	1	1	3	14.6983	0.0313*
C	1	1	5	1.5864	0.2634
D	1	1	5	67.9112	0.0004*
A*C	1	1	5	80.2914	0.0003*

2006.3.3 高橋 行雄　12

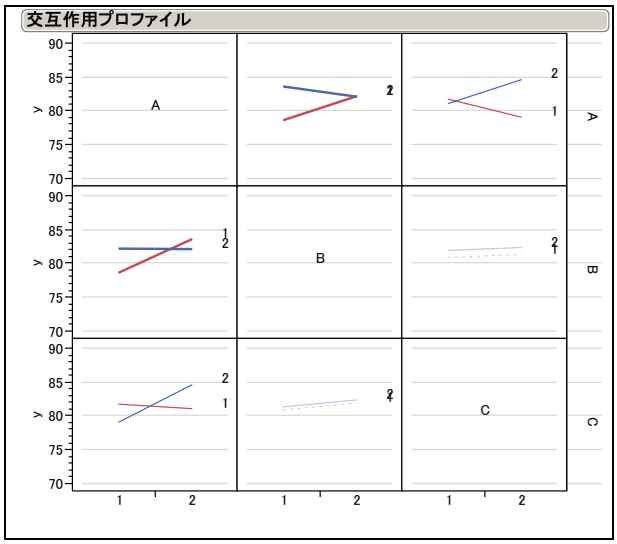
山田本では，1 次誤差が 2 次誤差に比べて有意ではないことから，1 次誤差を 2 次誤差にプーリングした結果が示されているが，ここでは，1 次単位を残した形で結果を表示した．

交互作用の検討

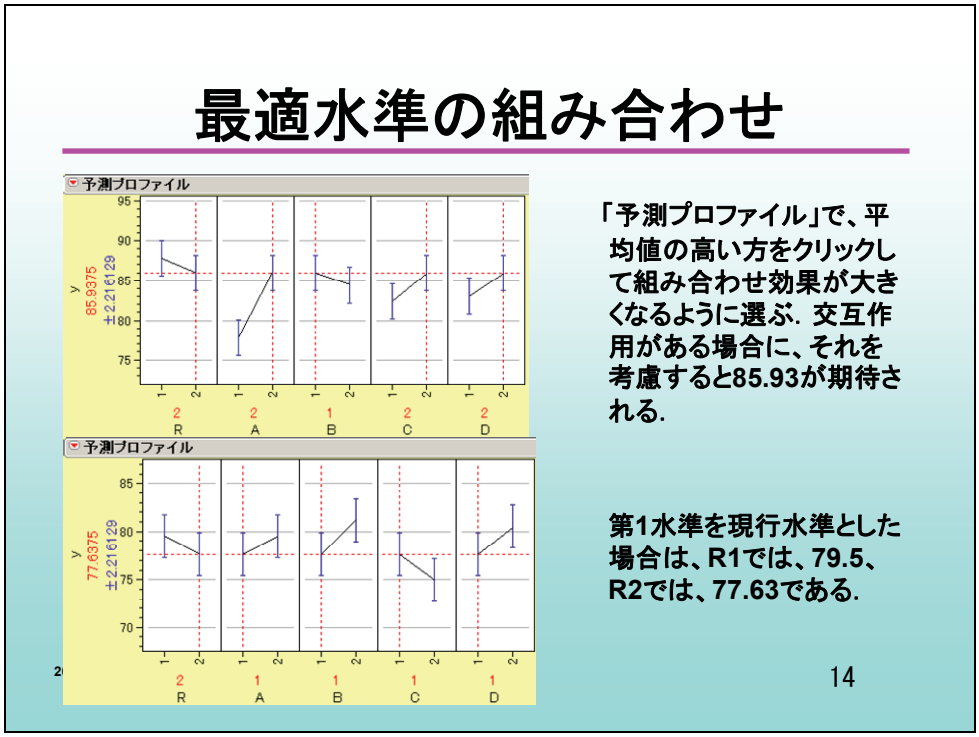


交互作用プロファイルで，主効果，交互作用を一覧する．交互作用項を設定しなかった因子の組み合わせについてもプロファイルが示されている．因子 A と因子 D の「交互作用」は平行な 2 本直線となっているので，交互作用をモデルにいれていないことが読み取れる．

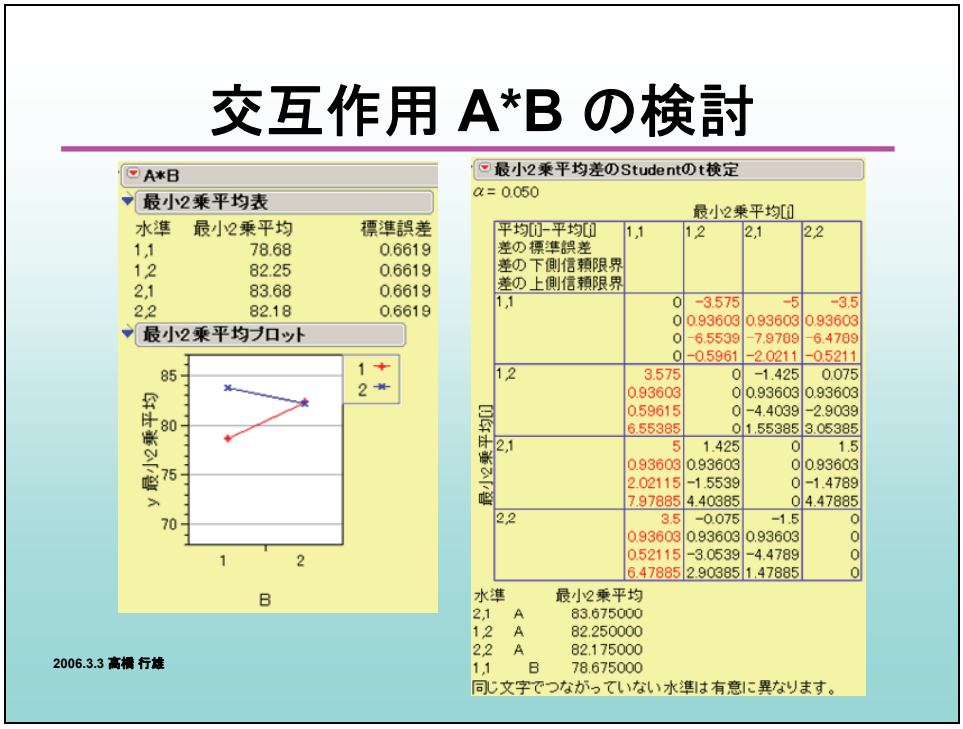
注）ここで示したのは，【因子プロファイル／予測プロファイル／交互作用プロファイル】の結果であり，予測プロファイルで選択した水準に連動して変化するので，結果を解釈する際に注意が必要であり，レポートに使うことは薦められない．正確には次に示す【因子プロファイル／交互作用プロット】を用いること．



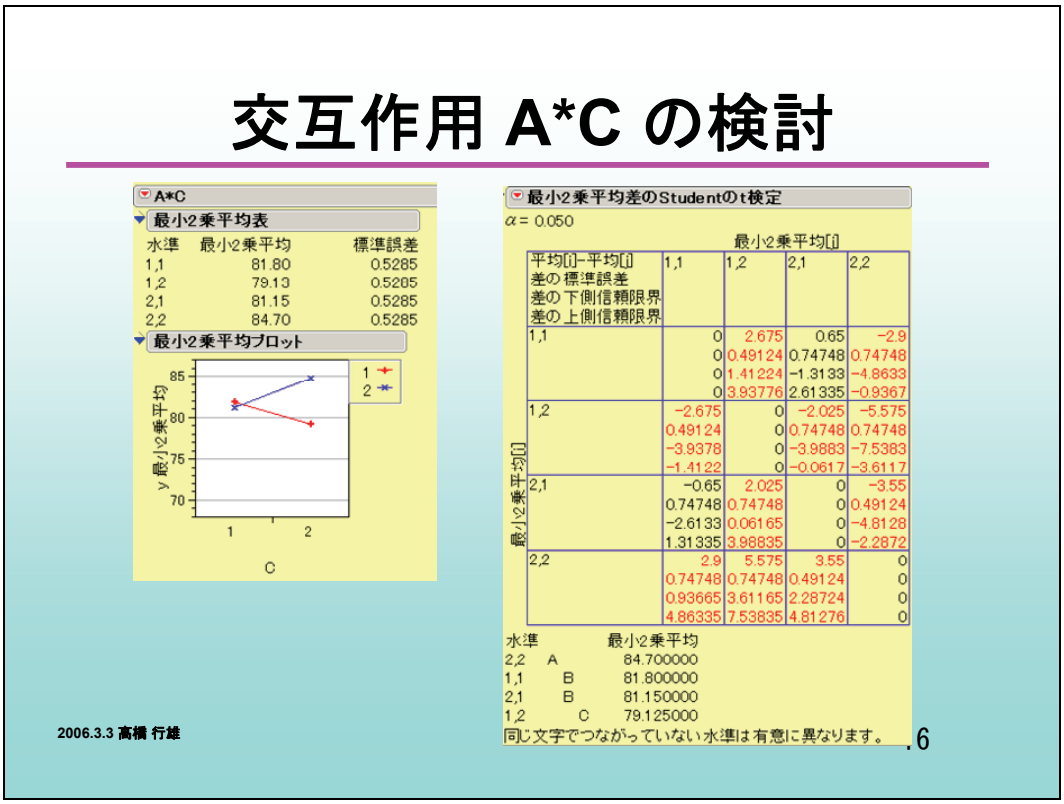
最適水準の組み合わせ平均，95%信頼区間



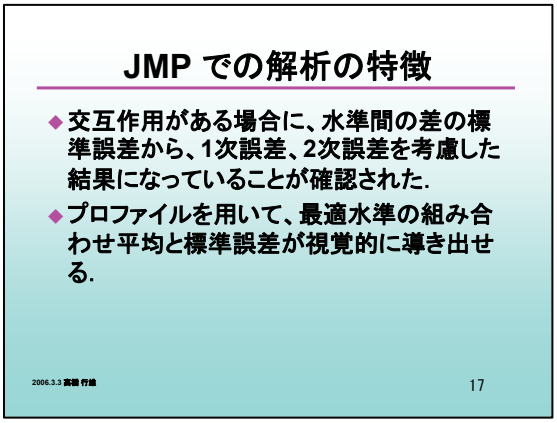
【予測プロファイル】で、平均値の高い方をクリックして組み合わせ効果が大きくなるように選ぶ．交互作用がある場合に，それを考慮すると 85.93 が推定値となる．第 1 水準を現行水準とした場合は，R1 では 79.5、R2 で 77.63 であることが読み取れる．



交互作用 A×B は 1 次単位内の因子の組み合わせであり，組み合わせ水準間の差の標準誤差は全て 0.9360 と一定である．



因子 A は 1 次単位，因子 C は 2 次単位なので，交互作用 A*C は 2 次単位となる．水準間の標準誤差は，1 次誤差成分と 2 次誤差成分を合成したものとなっている．組み合わせ水準間の差は，A1 および A2 内の C1 と C2 の差の場合には 2 次誤差のみを用いていて標準誤差は 0..4912 であるのに対し，(A1C1)－(A2C1) など異なる因子 A の水準にまたがる差は，1 次誤差成分とと 2 次誤差成分を合わせたもので，0.7475 と大きくなっている．



直交表を用いた分割実験の解析で，交互作用がある場合に，水準間の差の標準誤差から，1 次誤差，2 次誤差を考慮した結果になっていることが確認された．また，プロファイルを用いて、最適水準の組み合わせ平均と標準誤差が視覚的に手軽に導き出せた．これまでの，様々な統計ソフトで直交表での分割実験の解析を試みてきたのであるが，JMP の解析機能が最も優れていることを実感した．なお，現時点 2011 年 11 月現在でも，JMP より優れた統計ソフトには見当たらない．

4. 欠測値を含むクロスオーバー試験

4.1. 用量反応および食前食後を含む BE 試験

2006 年 3 月 3 日の JMPer's Meeting で「JMP による各種の分割実験データの解析」のスライド 43 囲以後に，人間を対象とした臨床研究でも，様々な実験計画法の活用事例を示した．こららのスライドに解説をくわえて再掲するとともに, JMP6 では不十分であった 90%信頼区間について JMP9.02 での追加を行なった結果を示す.

食前投与と食後投与の違い

2 要因 2×2 クロスオーバー

用量	投与順	被験者	実施 時期 1	実施 時期 2
10mg	1	1, 2, 3	食前	食後
	2	4, 5, 6	食後	食前
20mg	3	7, 8, 9	食前	食後
	4	10, 11, 12	食後	食前
30mg	5	13, 14, 15	食前	食後
	6	16, 17, 18	食後	食前

投与量ごとに6人，計18人，同一症例に食前投与と食後投与をクロスオーバー法で割りつける．主要変数は薬物濃度曲線下面積（AUC）．被験者が変量効果．計画はバランスしているが，欠測値がしばしば発生する．

2006.3.3 高橋 行雄

43

日本では，経口剤の薬では食後に服薬してもらうのが一般的であるのに対し，食前に服薬してもらうことが一般的な国々がある．食後に服薬した場合，食前に服薬した場合に，経口剤の有効成分が消化管で血液中に吸収され，体内で循環することによって効果が期待できる．食後か食前かによって, 消化管での吸収性能が異なると, 同じ用量の薬でも同じ効果が期待できなくなってしまう．そこで，同じ症例に時期を変えて食後，食前の服薬してもらい，血中の薬物濃度を測定して，食後でも，食前でも同様に有効成分が血液中に吸収されることを確かめる必要がある．

投与量ごとに 6 人，計 18 人，同一症例に食前投与と食後投与をクロスオーバー法で割りつけるた事例を示す．主要変数は薬物濃度曲線下面積（AUC）とする．解析モデルは，被験者を変量効果用量，食前食後，実施時期を固定効果となる．計画はバランスしているが，欠測値がしばしば発生する．

実験結果，AUC

dose	投与順	被験者	食前	食後
10mg	1	1	67.39	47.63
		2	19.48	9.58
		3	18.85	21.90
	2	4	35.38	22.79
		5	21.39	17.99
		6	23.44	20.18
20mg	1	7	31.30	44.41
		8	65.22	65.74
		9	43.81	37.02
	2	10	46.73	37.25
		11	27.71	20.62
		12	52.42	43.34
30mg	1	13	37.99	46.50
		14	−	53.58
		15	72.33	47.36
	2	16	77.89	58.46
		17	−	−
		18	61.27	68.70

30mg群のNo.14は，誤投与，No.17は，未投与で欠測値が発生．

2006.3.3 高橋 行雄　44

反応として AUC を示す．AUC は血中薬物濃度曲線下面積の略号で，それぞれの症例の薬物濃度曲線から計算した結果を示す．30mg 群の No.14 は，誤投与，No. 17 は，未投与で欠測値が発生してしまった．

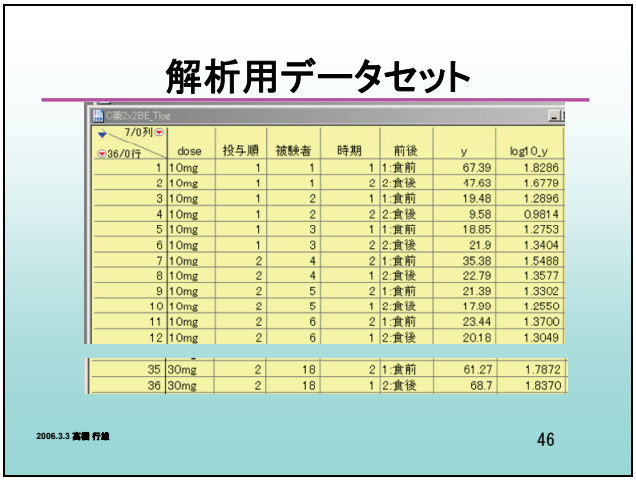
4.2. 直交表での分割実験と同じ構造

この実験は枝分かれ

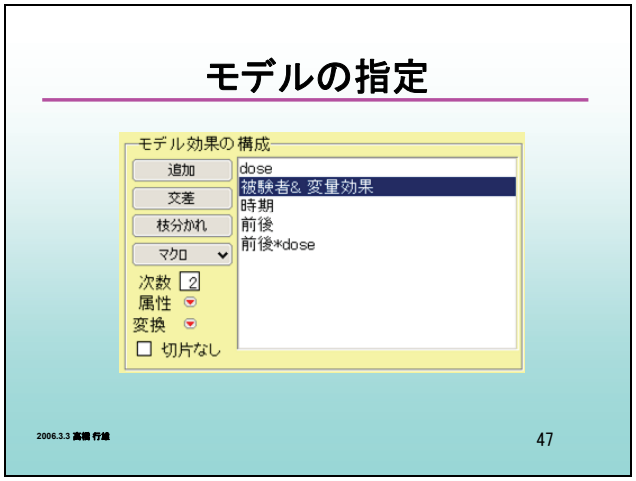
- EMS法（期待平均平方）では解析できない．
  - 18 症例と 3 用量が交絡している．
- 直交表を用いた分割実験でも同様の交絡が生ずる．
- REML法によって解くことができる．

2006.3.3 高橋 行雄　45

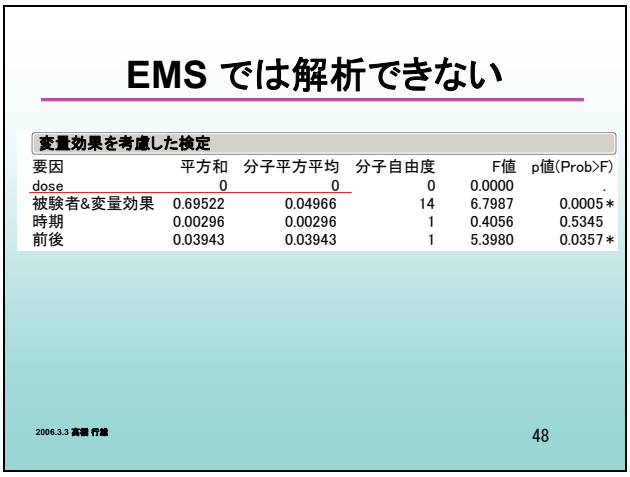
これまでは， SAS の MIXED プロシジャ（REML 法）で解析していたのであるが，JMP で適切に解析できるかを検証する．症例を変量効果とし，その中で 3 用量を割り付けている．このために，症例と用量が交絡して EMS 法（期待平均平方）では解析できない．直交表を用いた分割実験でも同様の交絡が生ずる．



	dose	投与順	被験者	時期	前後	y	log10_y
1	10mg	1	1	1	1:食前	67.39	1.8286
2	10mg	1	1	2	2:食後	47.63	1.6779
3	10mg	1	2	1	1:食前	19.48	1.2896
4	10mg	1	2	2	2:食後	9.58	0.9814
5	10mg	1	3	1	1:食前	18.85	1.2753
6	10mg	1	3	2	2:食後	21.9	1.3404
7	10mg	2	4	2	1:食前	35.38	1.5488
8	10mg	2	4	1	2:食後	22.79	1.3577
9	10mg	2	5	2	1:食前	21.39	1.3302
10	10mg	2	5	1	2:食後	17.99	1.2550
11	10mg	2	6	2	1:食前	23.44	1.3700
12	10mg	2	6	1	2:食後	20.18	1.3049
35	30mg	2	18	2	1:食前	61.27	1.7872
36	30mg	2	18	1	2:食後	68.7	1.8370

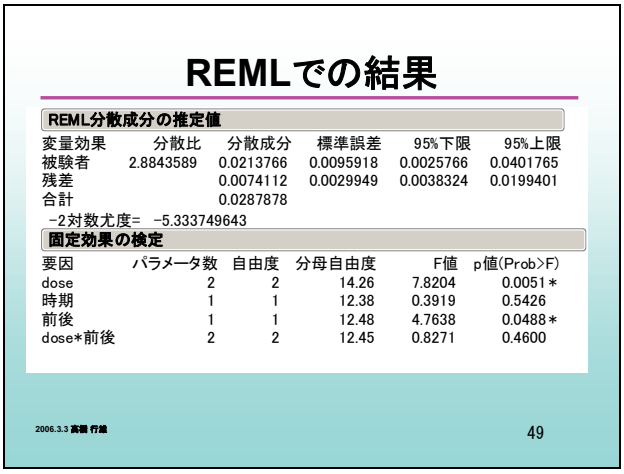


用量（dose）を 1 次因子，時期，投与前後を 2 次因子，交互作用として 2 次因子の投与前後と 1 次因子の dose を含むモデルを設定した．1 次誤差は，変量効果とした被験者となる．



変量効果を考慮した検定

要因	平方和	分子平方平均	分子自由度	F値	p値(Prob>F)
dose	0	0	0	0.0000	.
被験者&変量効果	0.69522	0.04966	14	6.7987	0.0005＊
時期	0.00296	0.00296	1	0.4056	0.5345
前後	0.03943	0.03943	1	5.3980	0.0357＊



REML分散成分の推定値

変量効果	分散比	分散成分	標準誤差	95%下限	95%上限
被験者	2.8843589	0.0213766	0.0095918	0.0025766	0.0401765
残差		0.0074112	0.0029949	0.0038324	0.0199401
合計		0.0287878			

−2対数尤度= −5.333749643

固定効果の検定

要因	パラメータ数	自由度	分母自由度	F値	p値(Prob>F)
dose	2	2	14.26	7.8204	0.0051＊
時期	1	1	12.38	0.3919	0.5426
前後	1	1	12.48	4.7638	0.0488＊
dose＊前後	2	2	12.45	0.8271	0.4600

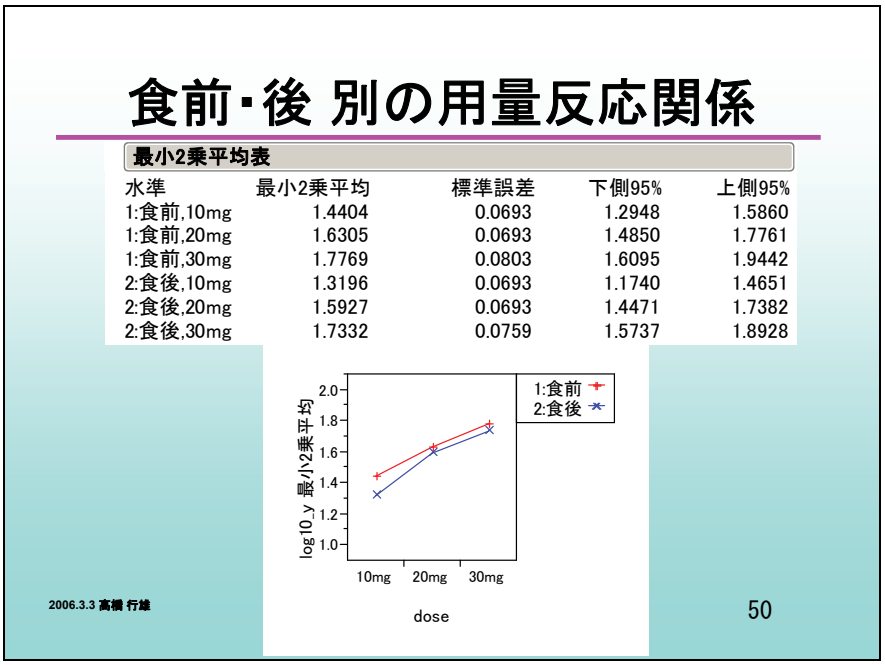
EMS 法では，用量 dose の平方和が 0 となり，解析できない．REML 法では，変量効果として被験者，固定効果の用量 dose が分離して示され，適切な解析となっている．なお，直交表での分割実験でも同様のデータの構造なので EMS 法で，dose の平方和が 0 となり，解くことができないとしたが，「dose に枝分れした被験者」と設定しなすことで，平方和の計算は行なえる．

EMS 枝分れ

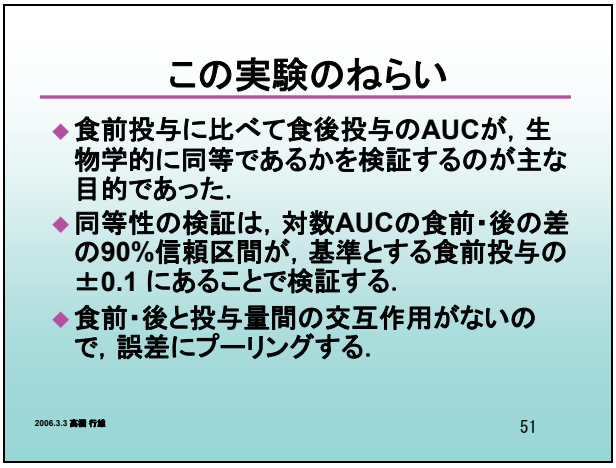
変量効果を考慮した検定

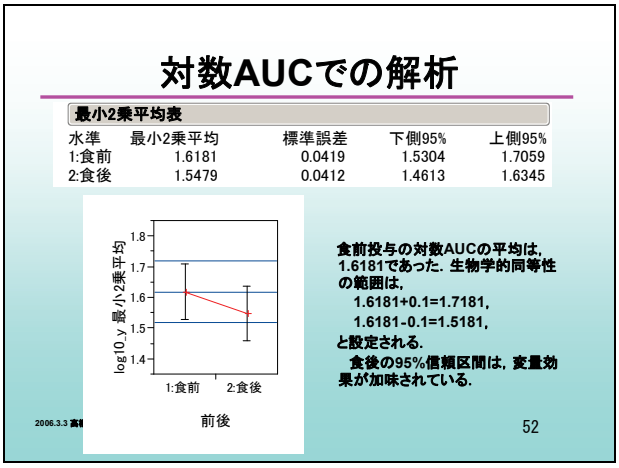
要因	平方和	分子平均平方	分子自由度	F値	p値(Prob>F)
dose	0.7208	0.3604	2	7.5308	0.0059*
被験者[dose]&変量効果	0.6954	0.0497	14	6.6290	0.0011*
時期	0.0030	0.0030	1	0.3953	0.5413
前後	0.0348	0.0348	1	4.6425	0.0522
前後*dose	0.0123	0.0062	2	0.8231	0.4624

【固定効果の検定】で，用量 dose，投与前後で有意な差があることが示されている．被験者の分散成分の 95%信頼区間が，t 分布の%点を用いた方式の結果を示してしまったが，この当時は，χ^2 分布を用いた適切な方法に切り替えられるこることの認識が乏しかったためである．これについては，6.3 節で詳しく述べる．



この実験は，用量の増加に伴ない反応 AUC がどのように変化するかの確認しつつ，食前と食後の AUC に差がないことを検証するのが目的であり，ここに示した交互作用プロットが，手軽に描けることは，解析コストの軽減となり，速やかな結果の解釈に役立つ．





食前投与に比べて食後投与の AUC が，生物学的に同等であるかを検証するのが主な目的であった．同等性の検証は，対数 AUC の食前・後の差の 90%信頼区間が，基準とする食前投与の±0.1 にあることで検証する．食前・後と投与量間の交互作用がないので，誤差にプーリングした解析を行なった．

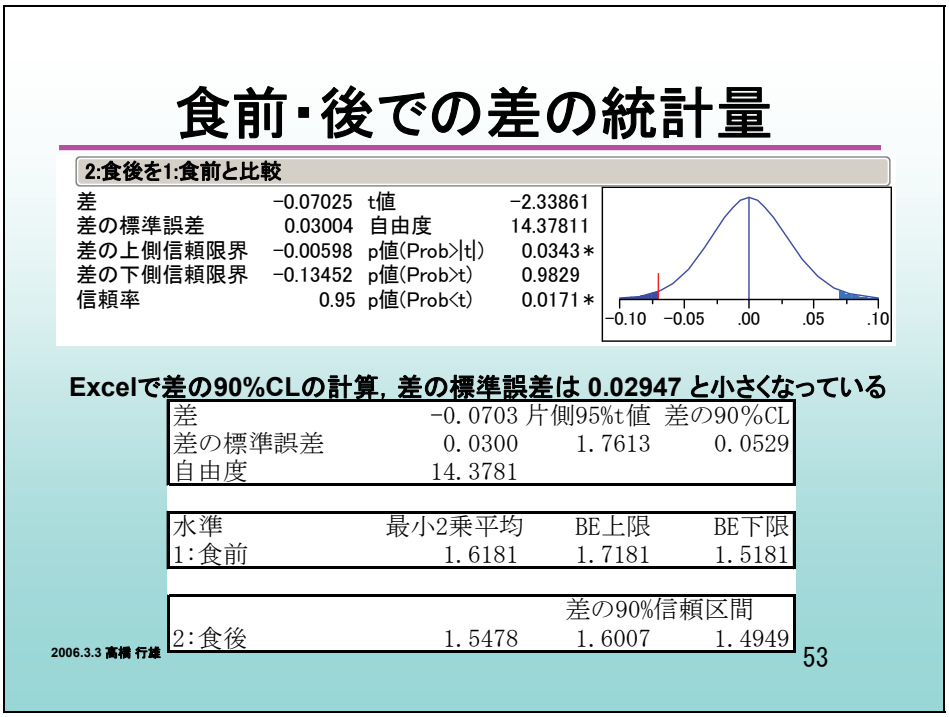
食前投与の対数 AUC の平均は，1.6181 であった．生物学的同等性の範囲は，

1.6181＋0.1=1.7181，

1.6181 - 0.1=1.5181，

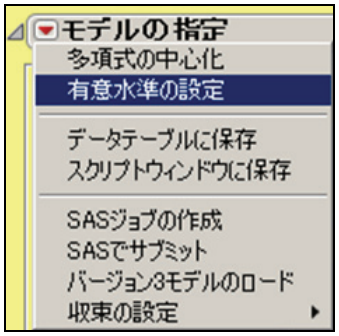
と設定される．食後の 95%信頼区間は，変量効果が加味されていることが確認された．

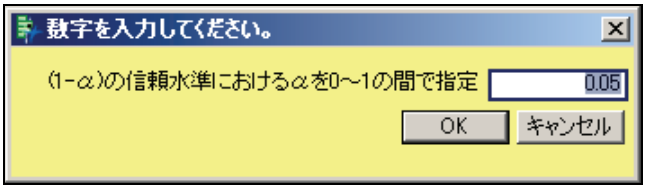
4.3.　90%信頼区間なども自在に

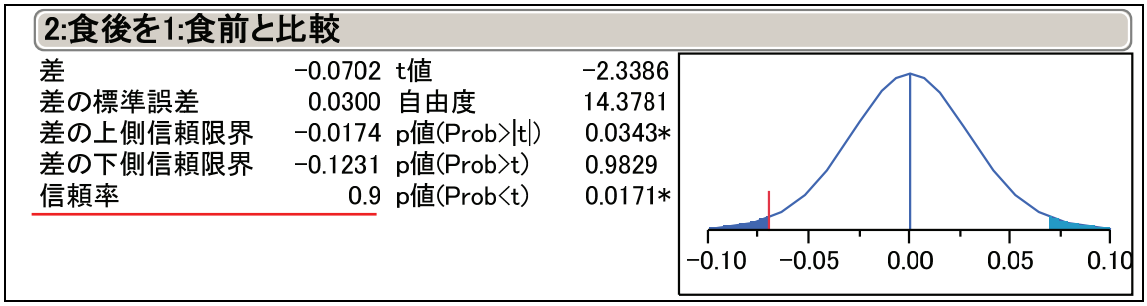


食前と食後の差について詳細な比較ができるのであるが，90%信頼区間を用いることが制度化されており，2006 年当時の JMP 6 では，対応できなかったので，JMP の関数計算を用いた計算結果を示した．

注）　現在の JMP 9 では，次のように有意水準の設定が任意に設定できるように機能拡張がされているので，別途計算する必要はなくなっている．







5. REML 法による分割実験の解析

5.1. 反復を伴う 1 段分割実験

データの構造

2.5 節で取り上げた「反復を伴う 1 段分割実験」についておさらいをしてみよう．表 5.1 に示す実験データは，反復 Rl の中で因子 A の水準をランダムに設定し，その中で，因子 B の 4 水準をランダムに設定し実験を行なう．反復 R2 でも同様なランダム化を行なう．

表 5.1　反復を伴う 1 段分割実験

	R1				R2			
A	B1	B2	B3	B4	B1	B2	B3	B4
A1	12.3	13.9	15.0	15.0	11.7	12.7	13.0	13.4
A2	15.5	14.8	15.2	14.4	13.9	13.3	14.3	13.6
A3	13.9	13.3	14.3	12.8	14.0	13.2	13.5	12.8

この実験には 3 種類の誤差が，各測定値に付いている．反復 R の水準を i，因子 A の水準を j，因子 B の水準を k としたときに，反復 R の変更に伴う反復誤差を $\varepsilon_{(0)i}$，因子 A の効果を α_j，反復 R 内の因子 A の水準変更に伴う 1 次誤差を $\varepsilon_{(2)ij}$，因子 B の効果を β_k，因子 A と因子 B の交互作用を $(\alpha\beta)_{ik}$，2 次誤差を $\varepsilon_{(3)ijk}$ とする．測定値を y_{ijk} は，次のような構造で示すことができる．

$$y_{ijk} = \mu + \gamma_i + \varepsilon_{(0)i} + \alpha_j + (\gamma\alpha)_{ij} + \varepsilon_{(1)ij} + \beta_k + (\alpha\beta)_{jk} + \varepsilon_{(2)ijk}$$

ただし，

$$\sum \alpha_j = 0,\quad \sum \beta_k = 0,\quad \varepsilon_{(0)i} \sim \mathrm{N}(0, \sigma^2_{(0)}),\quad \varepsilon_{(1)ij} \sim \mathrm{N}(0, \sigma^2_{(1)}),\quad \varepsilon_{(2)ijk} \sim \mathrm{N}(0, \sigma^2_{(2)})$$

である．EMS 法による解析と REML 法による解析結果を対比してみよう．REML 法を選択した場合に，モデル効果の設定を行なうと，解析方法が「REML（推奨）」に代わり，☑分散成分の範囲制限なし チェックボックスが現れる．そこににレ点が付いている．このままだと分散成分の 95%信頼区間の推定で，標準誤差のパーセント点を用いるために，下限が負となる場合も生じ不適切となるので，次のようにはずしておく．

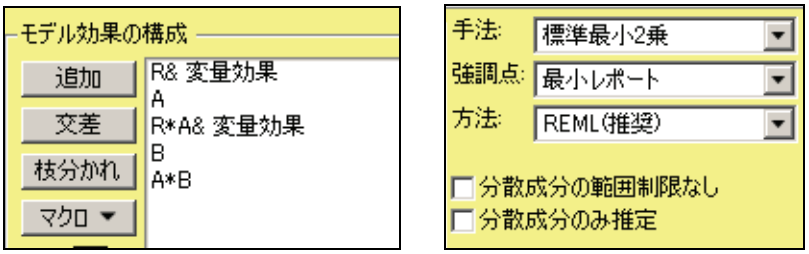


図 5.1　REML 法による解析

EMS 法は最小 2 乗法をベースとした解析で，全体の平方和を取り上げた因子に平方和を分解し，

それらの自由度で割った平均平方を求め，誤差の平均平方で割り F 値を求めて有意差検定を行なうという従来型の手順を踏襲している．分割実験での複数の分散成分の推定に EMS 法は，平均平方の期待値の構造から間接的に求めている．

REML 法は，固定効果および複数の分散成分を直接推定している．REML 法の基礎と応用については第 6 章で示すが，ここでは，分割実験に限定して REML 法での解析結果の見方を EMS 法と対比しつつ示す．

分散分析表

REML 法の解析結果には，従来型の分散分析表を作成するための出力はない．次に示す【固定効果の検定】が従来の分散分析表の一部ではあるが，REML 法では，因子 A，因子 B，交互作用 A×B の F 値を行列計算で直接計算しているので，平方和，自由度で割った平均平方の出力はない．

REML

固定効果の検定

要因	パラメータ数	自由度	分母自由度	F値	p値(Prob>F)
A	2	2	2	3.1045	0.2436
B	3	3	9	5.4155	0.0210*
A*B	6	6	9	9.9662	0.0015*

従来型の分散分析表は，データ総平均からの平方和を段階的に要因ごとの平方和に分解するといった計算手順であり，手計算でも追試できるといった安心感があった．しかしながら，REML 法では，要因の自由度，分母の自由度，F 値しか出力されないので，解析モデルが適切であったか不安になる．そのために REML 法に慣れるまでは，EMS 法による【変量効果を考慮した検定】と対比することで，REML 法の結果の妥当性をチェックすることを薦める．

REML 法での因子 A の F 値は 3.1045，EMS 法では，平方和 4.8533，自由度 2，平均平方 2.4267，F 値は 3.1045 と一致している．この F 値は，因子 A 下の「R*A&変量効果」の平均平方を分母とした 2.4267 / 0.7817 = 3.1045 である．

EMS

変量効果を考慮した検定

要因	平方和	分子平均平方	分子自由度	F値	p値(Prob>F)
R&変量効果	5.0417	5.0417	1	6.4499	0.1263
A	4.8533	2.4267	2	3.1045	0.2436
R*A&変量効果	1.5633	0.7817	2	6.7971	0.0159*
B	1.8683	0.6228	3	5.4155	0.0210*
A*B	6.8767	1.1461	6	9.9662	0.0015*

REML 法で直接推定した分散成分は【REML 法による分散成分推定値】の分散成分に出力されている．変量効果の R についての分散成分は 0.3550，変量効果の R*A の分散成分は 0.1667，残差の分散成分は 0.1150 となっている．分散成分の 95%信頼区間は，詳細は 6.3 節で示すが χ^2 分布のパーセント点を用いた方法が用いられていて，変量効果の R*A の場合分散成分が 0.1667，標準誤差が 0.1958 と分散成分よりも大きいが，95%下限は 0.0392 と 0 以下にならないように計算されている，

上限は 22.2617 と分散成分の 20 倍以上と裾広がりになっている．

REML

REML法による分散成分推定値

変量効果	分散比	分散成分	標準誤差	95%下側	95%上側	全体に対する百分率
R	3.09	0.3550	0.5977	0.0603	6049.5435	55.8
R*A	1.45	0.1667	0.1959	0.0392	22.2617	26.2
残差		0.1150	0.0542	0.0544	0.3833	18.1
合計		0.6367				100.0

-2対数尤度= 41.138630026

REML 法による出力結果は，冗長であるので，それらを 表 5.2 に示すように，従来の分散分析表の形式にまとめると，解析結果を一覧でき結果の解釈がしやすくなる．

表 5.2　反復のある 1 段分割実験の REML 法での分散分析表

要因	分散成分	自由度	分母自由度	F 値	p 値
R:反復誤差 e_0	0.3550	1			
A		2	2	3.10	0.2436
R×A:1 次誤差 e_1	0.1667	2			
B		3	9	5.42	0.0210
A×B		6	9	9.97	0.0015
2 次誤差 e_2	0.1150	9			
全体	0.6367	23			

5.2. 固定効果の水準平均

プロファイル

【因子プロファイル／プロファイル】では，全ての固定効果の水準平均および 95%信頼区間を一括して図 5.2 に示すようにグラフ化する．

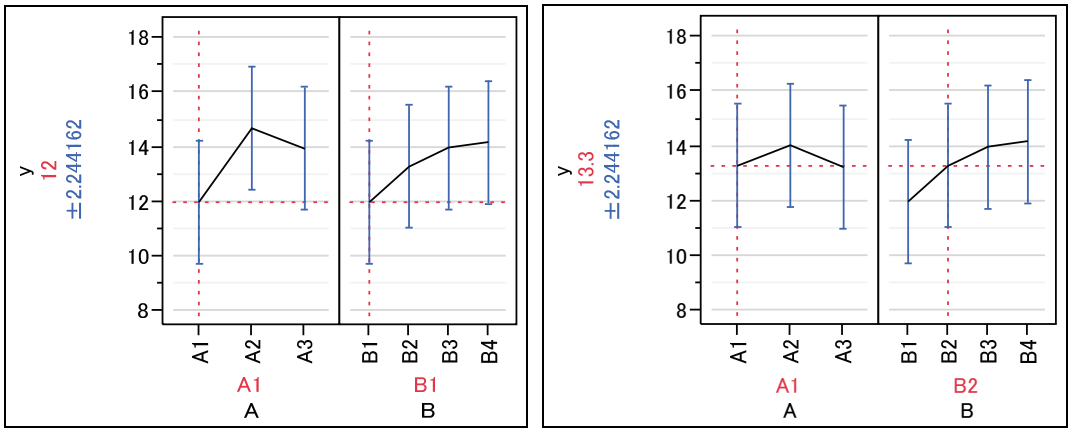


図 5.2　因子の水準の選択に伴うにプロファイルの変化

図の左側は，A1 と B1 水準が選択され，その平均値が 12，その 95%信頼区間が 2.24 であることが読み取れる．交互作用 A×B を解析モデルに含めているので，右側の図のように因子 B について B1 から B2 に水準の選択を変更（縦の点線）すると，因子 B のプロファイル自体は変化しないが，

因子 A のプロファイルは変化する．逆に，A1 から A2 と変化させても因子 A のプロファイルは変化しないが，因子 B のプロファイルは変化する．

交互作用を含めないモデルの場合には，片方のプロファイルは上下するだけで，プロファイルの形状は保たれる．最適水準を求めたい場合に，それぞれの因子の望ましい水準を選択すれば，それらの水準を組み合わせた場合の最適水準の推定平均と 95%信頼区間を求めることができる．交互作用がある場合に，どのようなプロファイルの変化があるかを概観するためには適しているが，報告書では，【因子プロファイル／交互作用プロット】で定性的な交互作用プロットを用いるべきである．

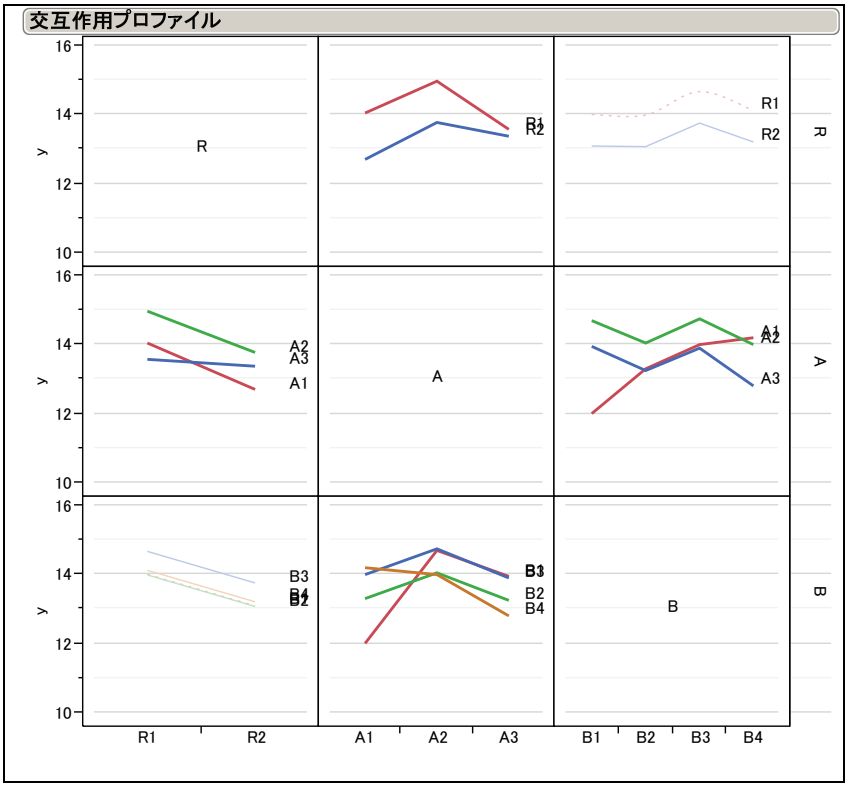


図 5.3　交互作用プロファイル

因子Aの水準平均の推定

【効果の詳細】には，モデルに含めた全ての要因について詳細な解析が行なえるようになっている．因子 A については，交互作用 A×B が有意な差なので，主効果単独での水準平均の検討は，不適切であるが，交互作用のない場合のことも想定して出力結果の見方を示す．

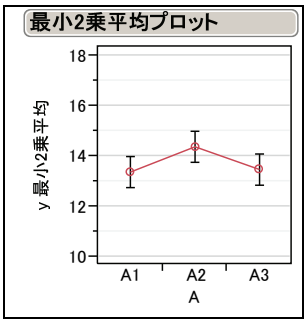
複数の誤差がある分割実験の水準平均の標準誤差は，2.1 節の完全ランダム実験とみなした場合とは全く異なる．完全ランダム実験の場合の因子 A の水準は 8 個のデータの平均で，誤差平方和が表 2.2 に示したように 0.6367 なので，その分散は 0.6367 / 8 = 0.07959，標準誤差はその平方根を取って 0.2821 となる．

反復を伴う 1 段分割実験の場合は，因子 A の各水準平均は，完全ランダム実験と同じであるが，標準誤差は 0.5246 と 0.2821 に比べ大きく異なっている．

完全ランダム

最小2乗平均表

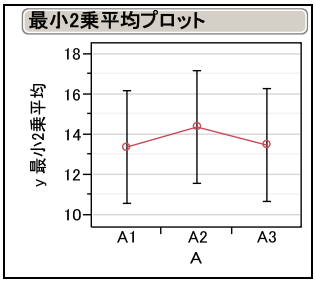
水準	最小2乗平均	標準誤差	下側95%	上側95%	平均
A1	13.3750	0.2821	12.7603	13.9897	13.3750
A2	14.3750	0.2821	13.7603	14.9897	14.3750
A3	13.4750	0.2821	12.8603	14.0897	13.4750



分割実験

最小2乗平均表

水準	最小2乗平均	標準誤差	下側95%	上側95%
A1	13.3750	0.5246	10.5645	16.1855
A2	14.3750	0.5246	11.5645	17.1855
A3	13.4750	0.5246	10.6645	16.2855



これは，因子 A 水準平均の誤差構造が完全ランダムの場合とは異なるためである．A_1 水準平均は，次の組み合わせで得られたデータ 8 個の合計の平均となっている．

$$R_1A_1B_1 \qquad R_1A_1B_2 \qquad R_1A_1B_3 \qquad R_1A_1B_4$$
$$R_2A_1B_1 \qquad R_2A_1B_2 \qquad R_2A_1B_3 \qquad R_2A_1B_4$$

A_1 の水準は，反復 R の中で 2 回ランダム化が行なわれているので,に過ぎない．因子 B は，それぞれの反復 R の中でランダム化が 8 回行なわれている．そのために 8 個のデータの平均の誤差分散は，

$$V(A_1)=V\left(\frac{\sum_{i=1}^{2}4(\hat{\sigma}_{(0)}^2+\hat{\sigma}_{(1)}^2)+\sum_{i=1}^{2}\sum_{k=1}^{4}\hat{\sigma}_{(2)}^2}{8}\right)=\frac{\hat{\sigma}_{(0)}^2+\hat{\sigma}_{(1)}^2}{2}+\frac{\hat{\sigma}_{(2)}^2}{8}$$

となる．表 5.2 で整理された分散成分を用いて

$$V(A_1)=\frac{\hat{\sigma}_{(0)}^2+\hat{\sigma}_{(1)}^2}{2}+\frac{\hat{\sigma}_{(2)}^2}{8}=\frac{0.3550+0.1667}{2}+\frac{0.1150}{8}=0.2752$$

$$SE(A_1)=\sqrt{0.2752}=0.5246$$

と，因子 A の標準誤差は 0.5234 となることが確認できる．この標準誤差の自由度は 3 種類の誤差分散を合成しているので，95%信頼区間を求めるためには誤差分散の自由度に対応する t 分布 95%点を必要とする．自由度は，表 2.11 から $df_{(0)}=1$，$df_{(1)}=2$，$df_{(2)}=9$ となり，自由度の近似計算には Satterthwaite 法を適用する．注）JMP では近似自由度が出力されていないので，95%信頼区間と標準誤差から

$$df_{(0,1,2)} = \frac{1}{2} \cdot \frac{U95\% - L95\%}{SE} = \frac{16.1855 - 10.5645}{2 \times 0.5346} = 5.36$$

となり，図 5.4 の自由度と 95%点の図から，自由度は 1.62 程度と推測される．

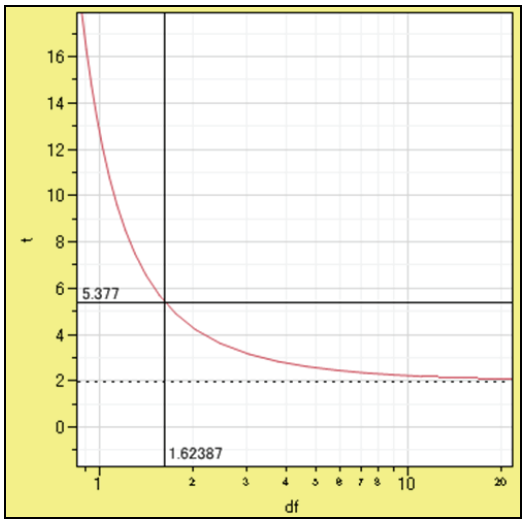


図 5.4　t 分布の 95%点

なお，異なる分散の合成は，2 群の t 検定で分散が異なる場合の Aspin-Welch 検定でも分散の合成が行なわれ，自由度の計算には Satterthwaite 法が用いられている．JMP では，近似の精度をあげた Kenward-Roger 法が採用されている．JMP のマニュアルで，次のような記載「単純で釣り合いの取れたモデルの場合、分散推定値の下限が 0 に設定されていなければ、Kenward-Roger の REML 法で得られる結果は従来の方法と一致します。」とあり，この事例では一致することがわかる

注）Satterthwaite 法を用いて自由度の計算をしたが，2.16 となり，95%点は 4.00 で再現できない．今後の検討課題とする．

因子Aについての効果の差の推定

完全ランダム実験の場合について，因子 A の水準間の平均値の差について，【最小 2 乗平均差の Student の t 検定／差の順位レポート】を適用した結果を次に示す．因子 A の水準間の差の大きさの順に並べかえられていて，信頼区間の下限と上限が全て 0 を含まない，あるいは p 値が 0.05 以下の A1 と A3，A2 と A3 の水準間に有意な差があることが分かる．

完全ランダム

$\alpha = 0.050$　t= 2.17881

水準	－水準	差	差の標準誤差	下側信頼限界	上側信頼限界	p値
A2	A1	1.00	0.3990	0.1307	1.8693	0.0276*
A2	A3	0.90	0.3990	0.0307	1.7693	0.0435*
A3	A1	0.10	0.3990	−0.7693	0.9693	0.8063

差の標準誤差は，各水準の $\sqrt{0.6367/8} = 0.2821$ に対し，2 つの分散の和なので $\sqrt{2 \times 0.6367/8} = 0.3990$，

と$\sqrt{2}$倍となっている．

それに対し反復のある 1 段分割実験の場合には，次に示すように

分割実験

$\alpha = 0.050$

水準	－ 水準	差	差の標準誤差	下側信頼限界	上側信頼限界	p値
A2	A1	1.00	0.4421	−0.9020	2.9020	0.1521
A2	A3	0.90	0.4421	−1.0020	2.8020	0.1787
A3	A1	0.10	0.4421	−1.8020	2.0020	0.8421

信頼区間の下限と上限が全て 0 を含むこと，あるいは p 値が 0.05 以上なので，水準間に有意な差がないことが分かる．水準間の標準誤差は 0.4421 と完全ランダム実験の場合の 0.3990 に比べ大きくなっていることが，有意な差とならない原因である．ただし，水準平均の標準誤差 0.5246 よりも差の標準誤差は$\sqrt{2}$倍ではなく逆に 0.4421 と小さくなっている．

これは，A1 と A2 の平均値の差は，反復 R1 内での A1 と A2 の差，反復 R2 内での A1 と A2 の差に分解でき，反復の誤差分散$\hat{\sigma}^2_{(0)} = 0.3550$の影響を受けない．したがって，差の標準誤差は

$$V(A_1 - A_2) = V\left(2 \times \frac{\sum_{i=1}^{2} 4\hat{\sigma}^2_{(1)} + \sum_{i=1}^{2}\sum_{k=1}^{4} \hat{\sigma}^2_{(2)}}{8}\right) = \hat{\sigma}^2_{(1)} + \frac{\hat{\sigma}^2_{(2)}}{4} = 0.1667 + \frac{0.1150}{4} = 0.1955$$

$$SE(A_1) = \sqrt{0.1955} = 0.4421$$

と計算されている．

因子Aについての効果の*F*検定

【差の順位レポート】では，個々の水準間の比較を総当で行なうのであるが，A_1水準と A_2および A_3水準の平均値を比較したい．【最小 2 乗平均の対比】は，対比を用いて実現できる．対比とは，水準平均（最小 2 乗平均）を $\hat{\mu}(A_j)$ としたときに，水準間の差を計算する際の係数で，それらを加えたときに 1 となるベクトルである．A_1水準と A_2および A_3水準の平均値の差は，次のように表すことができ，

$$\hat{\mu}(A_1) - \frac{\hat{\mu}(A_2) + \hat{\mu}(A_2)}{2} = 1 \times \hat{\mu}(A_1) - \frac{1}{2} \times \hat{\mu}(A_2) - \frac{1}{2} \times \hat{\mu}(A_3)$$

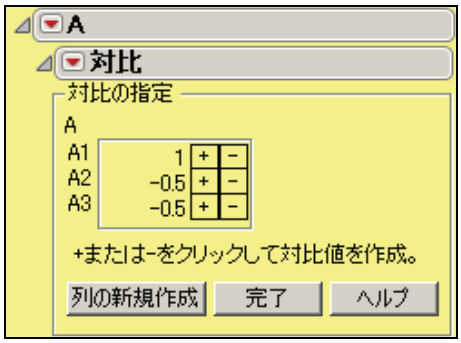
この場合の対比は（1，-1/2，-1/2）である．A_2と A_3の差の対比は（0，1，-1）である．それぞれの対比について【最小 2 乗平均の対比】を用いて計算してみよう．

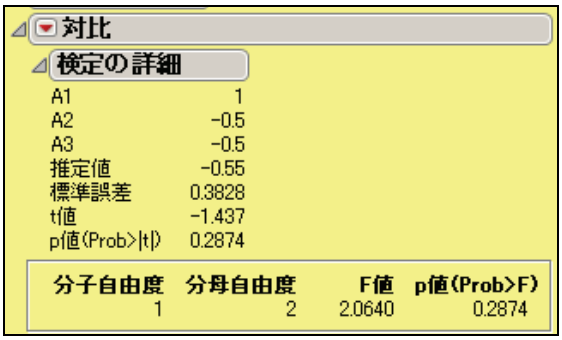
対比（1，-1/2，-1/2）の場合の推定値は，

$$1 \times 13.275 - \frac{1}{2} \times 14.374 - \frac{1}{2} \times 13.475 = -0.55$$

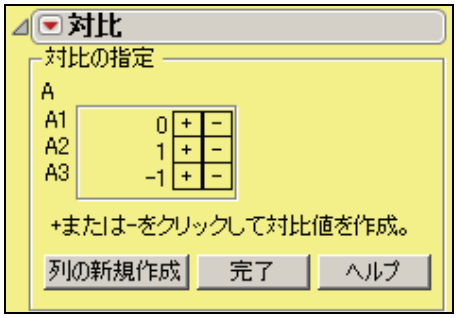
で，標準誤差は 0.3828，t 値は-1.437，p 値は自由度 2 の t 分布表から 0.2874 となる．F 値は t 値の 2 乗 2.0640，p 値は t *検定の*場合と同じで 0.2874 である．

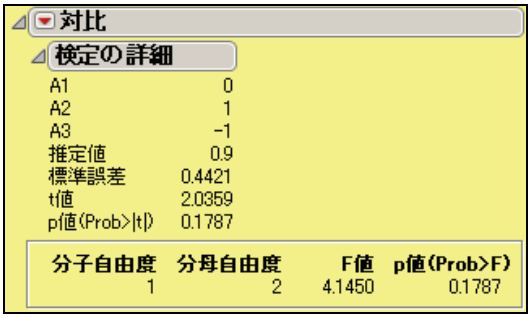
対比（1，-1/2，-1/2）





対比（0，1，-1）



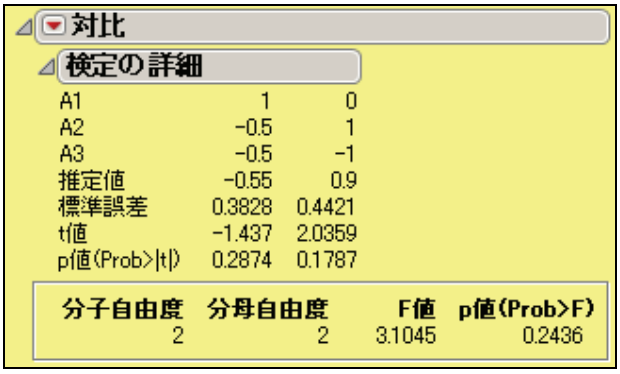


対比（0，1，-1）の場合の差は 0.9，標準誤差は 0.4421，p 値は 0.1787 である．これは，すでに示した【差の順位レポート】でし示した

水準	－ 水準	差	差の標準誤差	下側信頼限界	上側信頼限界	p値
A2	A3	0.90	0.4421	−1.0020	2.8020	0.1787

に一致する．

因子 A の 3 水準全体に対する F 検定は，これらの 2 つの対比をまとめて計算した結果と一致する．これは，1 つの対比を設定したあとに，【列の新規作成】を選択すると 2 つ目の対比を設定できるようになる．



このに示されている F 値が，すでに示した【固定効果の検定】に対応する．

REML

固定効果の検定

要因	パラメータ数	自由度	分母自由度	F値	p値(Prob>F)
A	2	2	2	3.1045	0.2436

因子 A の効果全体についての F 検定のためにここで示した 1 組の対比以外に，多数設定できる．（A_1 と A_2 の差）および（A_2 と A_3 の差）の 2 組でも同じ結果となる．

対比

検定の詳細

A1	1	0
A2	-1	1
A3	0	-1
推定値	-1	0.9
標準誤差	0.4421	0.4421
t値	-2.262	2.0359
p値(Prob>|t|)	0.1521	0.1787

分子自由度	分母自由度	F値	p値(Prob>F)
2	2	3.1045	0.2436

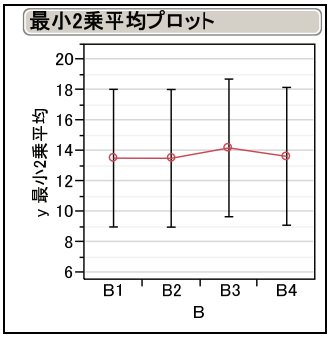
因子Bについて

因子 A は 1 次単位内の因子であるのに対し，因子 B は 2 次単位の因子である．このために因子 A で適用した標準誤差の計算が異なる．次に因子 B の【最小 2 乗平均】の結果を示す．

分割実験

最小2乗平均表

水準	最小2乗平均	標準誤差	下側95%	上側95%
B1	13.5500	0.4738	9.0322	18.0678
B2	13.5333	0.4738	9.0155	18.0512
B3	14.2167	0.4738	9.6988	18.7345
B4	13.6667	0.4738	9.1488	18.1845



B_1 の水準は，反復 R の中で 2 回ランダム化，因子 A について 3 回のランダム化が行なわれているので 6 個のデータの平均の誤差分散は，

$$V(B_1)=V\left(\frac{\sum_{i=1}^{2}3\hat{\sigma}_{(r)}^2+\sum_{i=1}^{2}\sum_{j=1}^{3}(\hat{\sigma}_{(1)}^2+\hat{\sigma}_{(2)}^2)}{6}\right)=\frac{\hat{\sigma}_{(r)}^2}{2}+\frac{\hat{\sigma}_{(1)}^2+\hat{\sigma}_{(2)}^2}{6}$$

$$SE(B_1)=\sqrt{\frac{0.3550}{2}+\frac{0.1667+0.115}{6}}=0.4738$$

と計算されている．最小 2 乗平均差の Student の t 検定】から【差の順位レポート】水準間の差をみると B3 水準と他の水準間で有意な差となっている．水準平均の標準誤差が大きく全く水準間に有意な差がないと判断しがちであるが，差の標準誤差は 0.1958 と因子 A に比べても小さくなっている．

分割実験

$\alpha = 0.050$

水準	－ 水準	差	差の標準誤差	下側信頼限界	上側信頼限界	p値
B3	B2	0.68	0.1958	0.2404	1.1262	0.0068*
B3	B1	0.67	0.1958	0.2238	1.1096	0.0078*
B3	B4	0.55	0.1958	0.1071	0.9929	0.0204*
B4	B2	0.13	0.1958	−0.3096	0.5762	0.5130
B4	B1	0.12	0.1958	−0.3262	0.5596	0.5659
B1	B2	0.02	0.1958	−0.4262	0.4596	0.9340

因子 A の水準間の差の場合は，反復 R1 内での A1 と A2 の差，反復 R2 内での A1 と A2 の差に分解でき，反復の誤差分散 $\hat{\sigma}^2_{(r)} = 0.3550$ の影響を受けない．因子 B の水準間の差は，反復 R 内の因子 A の水準内内の差となり，反復の誤差分散 $\hat{\sigma}^2_{(r)} = 0.3550$ も 1 次誤差 $\hat{\sigma}^2_{(1)} = 0.1667$ の影響を受けず，2 次誤差 $\hat{\sigma}^2_{(2)} = 0.1150$ のみとなり，差の標準誤差は，$SE(B_1 - B_2) = \sqrt{2 \times 0.1150/6} = 0.1958$ と反復誤差，1 次誤差の影響を全く受けない．

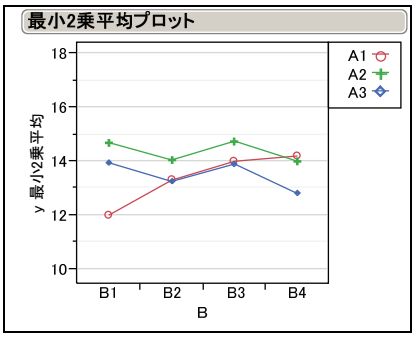
交互作用A×Bについて

交互作用 A×B については，表 5.2 から $p = 0.0051$ と有意であった．最小 2 乗平均プロットから A_1B_1 の組み合わせが最小値で， A_1 内の B_2，B_3，B_4，と平均値が大きくなっている．次に B_1 内の A_2，A_3，との間で平均値の差の信頼区間を出力してみよう．

分割

最小2乗平均表

水準	最小2乗平均	標準誤差	下側95%	上側95%
A1,B1	12.0000	0.5642	9.7558	14.2442
A1,B2	13.3000	0.5642	11.0558	15.5442
A1,B3	14.0000	0.5642	11.7558	16.2442
A1,B4	14.2000	0.5642	11.9558	16.4442
A2,B1	14.7000	0.5642	12.4558	16.9442
A2,B2	14.0500	0.5642	11.8058	16.2942
A2,B3	14.7500	0.5642	12.5058	16.9942
A2,B4	14.0000	0.5642	11.7558	16.2442
A3,B1	13.9500	0.5642	11.7058	16.1942
A3,B2	13.2500	0.5642	11.0058	15.4942
A3,B3	13.9000	0.5642	11.6558	16.1442
A3,B4	12.8000	0.5642	10.5558	15.0442



A_1B_1 の組み合わせの標準誤差は，

$$V(A_1B_1) = V\left(\frac{\sum_{i=1}^{2}(\hat{\sigma}^2_{(0)} + \hat{\sigma}^2_{(1)} + \hat{\sigma}^2_{(2)})}{2}\right) = \frac{\hat{\sigma}^2_{(0)} + \hat{\sigma}^2_{(1)} + \hat{\sigma}^2_{(2)}}{2}$$

$$SE(A_1B_1) = \sqrt{\frac{0.3550+0.1667+0.1150}{2}} = 0.5642$$

となる．差の順位レポートは，組みあわせ数が 66 通りもあり冗長なので，A_1B_1 と A_1 内の B_2，B_3，B_4 の組み合わせ，次に B_1 内の A_2，A_3 との間を抽出した結果を次に示す．

水準	- 水準	差	差の標準誤差	下側信頼限界	上側信頼限界	p 値	
A1,B2	A1,B1	1.30	0.3391	0.5329	2.0671	0.0040	*
A1,B3		2.00	0.3391	1.2329	2.7671	0.0002	*
A1,B4		2.20	0.3391	1.4329	2.9671	0.0001	*
A2,B1	A1,B1	2.70	0.5307	1.2239	4.1761	0.0071	*
A3,B1		1.95	0.5307	0.4739	3.4261	0.0215	*

これらの水準の組み合わせ間の差は，全に有意な結果であるが，差の標準誤差が異なる．A_1 水準内の因子 B 間の比較の場合は，0.3391 であるが，B_1 水準内の因子 A 間の比較の場合は，0.5307 大きくなっている．A_1 水準内の因子 B 間の比較では，反復誤差，1 次誤差の影響を受けずに 2 次誤差分散のみとなり，

$$SE(A_1B_1 - A_1B_2) = \sqrt{2 \times 0.1150/2} = 0.3391$$

であるのに対し，A_1 水準と A_2 水準にまたがる場合は，1 次誤差分散の影響を受けて

$$SE(A_1B_1 - A_2B_1) = \sqrt{2 \times (0.1150+0.1667)/2} = 0.5307$$

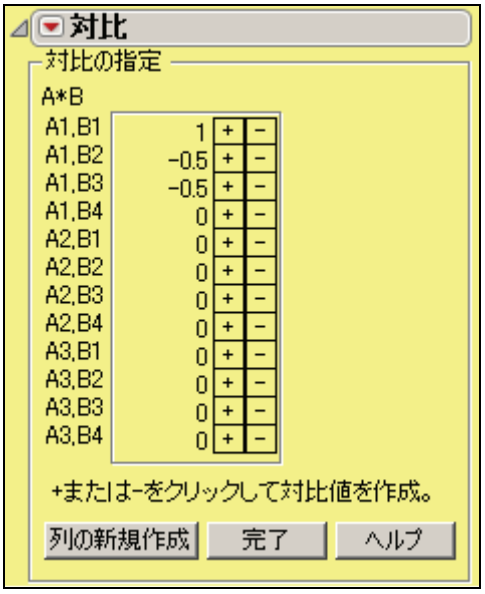
と大きくなる．

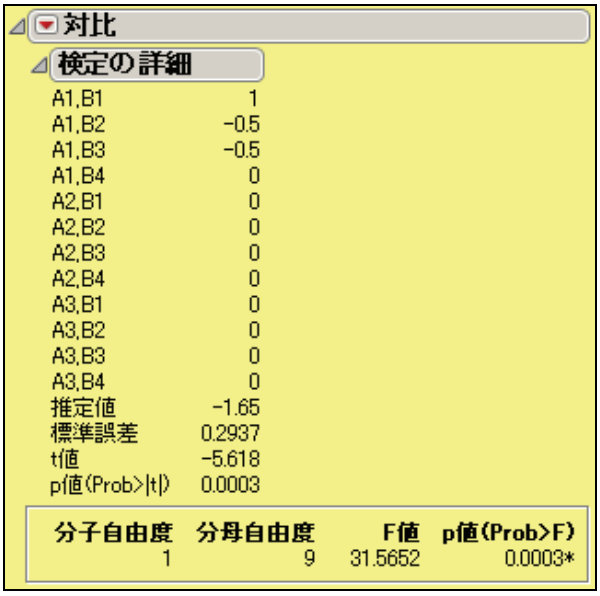
対比の応用

因子 A についての効果の F 検定の項で，対比によって任意の水準の組み合わせ平均の差についての検定統計量の出し方を説明した．A_1B_1 を現行水準とした場合に，因子 A の水準を変えた場合「A_2B_1 と A_3B_1 の平均」との検定統計量を計算してみよう．対比は（1，-0.5，-0.5）となり，推定値は，

$$1 \times 12.0 - 0.5 \times (13.3 + 14.0) = -1.65$$

で，標準誤差は 0.2937 との結果が得られる．





5.3. パラメータの推定値および予測値

パラメータの推定値

【パラメータ推定値】は，データの構造で設定したモデル式に対応した出力になっている．固定効果としては切片 μ，α_j，β_k，$(\alpha\beta)_{ik}$ であり，次に REML 法での推定結果を示す．この推定結果をもとに，これまでに示してきた因子 A，因子 B，交互作用 A×B の水準平均が計算されている．

どのような計算がなされているか理解することは，REML 法による解析手順を理解するために役に立つ．

パラメータ推定値

項	推定値	標準誤差	分母自由度	t値	p値(Prob>|t|)
切片	13.7417	0.4583	1	29.98	0.0212*
A[A1]	−0.3667	0.2552	2	−1.44	0.2874
A[A2]	0.6333	0.2552	2	2.48	0.1312
B[B1]	−0.1917	0.1199	9	−1.60	0.1444
B[B2]	−0.2083	0.1199	9	−1.74	0.1163
B[B3]	0.4750	0.1199	9	3.96	0.0033*
A[A1]*B[B1]	−1.1833	0.1696	9	−6.98	<.0001*
A[A1]*B[B2]	0.1333	0.1696	9	0.79	0.4519
A[A1]*B[B3]	0.1500	0.1696	9	0.88	0.3994
A[A2]*B[B1]	0.5167	0.1696	9	3.05	0.0139*
A[A2]*B[B2]	−0.1167	0.1696	9	−0.69	0.5088
A[A2]*B[B3]	−0.1000	0.1696	9	−0.59	0.5699

固定効果としては，24 個のデータの平均値としての切片 $\mu = 13.7417$，因子 A は 3 水準であるが A[A1]：α_1，A[A2]：α_2 と 2 水準分しか推定結果がない．A[A3]の効果は，$-\alpha_1 - \alpha_2$ として計算される．JMP の内部では，因子 A は名義尺度であるので，因子 A の 3 水準をダミー変数と一般的に言われる連続量 x_1，x_2 を用いて，A1 の場合：（$x_1 = 1$，$x_2 = 0$），A2 の場合：（$x_1 = 0$，$x_2 = 1$），A3 の場合：（$x_1 = -1$，$x_2 = -1$）と展開している．ダミー変数 x_1 のパラメータを $\alpha_1 = -0.3667$，ダミー変数 x_2 のパラメータを $\alpha_2 = 0.6333$ に対応する．因子 A の第 j 水準平均は，切片に対するダミー変数を $x_0 = 1$ としたときに，

$$\mu(A_j) = \mu \cdot x_0 + \alpha_1 \cdot x_1 + \alpha_2 \cdot x_2,$$

で計算されている．

表 5.3 に各水準平均の計算結果を示す．

表 5.3　因子 A の水準平均の計算

水準	μ	x_0	α_1	x_1	α_2	x_2	水準平均	効果
A1	13.7417	1	-0.3667	1	0.6333	0	13.3750	α_1
A2		1		0		1	14.3750	α_2
A3		1		-1		-1	13.4750	$-\alpha_1 - \alpha_2$

因子 B は 4 水準なので，ダミー変数は x_3，x_4，x_5 となり，パラメータは，β_1，β_2，β_3 で，各水準平均は，因子 B の第 k 水準平均は

$$\mu(B_k) = \mu \cdot x_0 + \beta_1 \cdot x_3 + \beta_2 \cdot x_3 + \beta_3 \cdot x_5,$$

となり，表 5.4 に計算結果を示す．

表 5.4　因子 B の水準平均の計算

水準	μ	x_0	β_1	x_3	β_2	x_4	β_3	x_5	水準平均	効果
B1	13.7417	1	-0.1917	1	-0.2083	0	0.4750	0	13.5500	β_1
B2		1		0		1		0	13.5333	β_2
B3		1		0		0		0	14.2167	β_3
B4		1		-1		-1		-1	13.6667	$-\beta_1-\beta_2-\beta_3$

交互作用 A×B は, 因子 A の 2 水準と因子 B の 3 水準を掛けた 6 水準のパラメータの推定結果が示されている．交互作用全体は 3×4 ＝ 12 通りあるので，残りの効果は 表 5.5 に示すように計算されている．

表 5.5　交互作用 A×B の効果

	B1	B2	B3	B4
A1	$(\alpha\beta)_{11}=-1.1833$	$(\alpha\beta)_{12}=0.1333$	$(\alpha\beta)_{13}=0.1500$	$-(\alpha\beta)_{11}-(\alpha\beta)_{12}-(\alpha\beta)_{13}=0.9000$
A2	$(\alpha\beta)_{21}=0.5167$	$(\alpha\beta)_{22}=-0.1167$	$(\alpha\beta)_{23}=-0.1100$	$-(\alpha\beta)_{21}-(\alpha\beta)_{22}-(\alpha\beta)_{23}=-0.3000$
A3	$-(\alpha\beta)_{11}-(\alpha\beta)_{21}$ $=0.6667$	$-(\alpha\beta)_{12}-(\alpha\beta)_{22}$ $=-0.0167$	$-(\alpha\beta)_{13}-(\alpha\beta)_{23}$ $=-0.0050$	$-(\alpha\beta)_{11}-\cdots-(\alpha\beta)_{23}$ $=-0.6000$

因子 A×B の水準平均は,

$$\hat{\mu}((AB)_{jk})=\text{総平均}\mu+A_j\text{の効果}+B_j\text{の効果}+(AB)_{jk}\text{の効果},$$

として計算されている．

固定効果についての予測値

結果の測定値 y_{ijk} に対する推定値 $\hat{y}_{ijk}$，残差は，【列の保存】で JMP データセットに追加出力される．【列の保存／予測式】を選択すると，変量効果を除いた固定効果のみについて，予測式を含んだ推定値がテーブルに保存される．図 5.5 に保存された【計算式】を示す．この結果は，前項で示した交互作用 A×B の水準平均の算出式に対応している．　なお，【推定値／予測式の推定】でも，予測式は表示される．

```
13.7417 + Match(A) { "A1" ⇒ -0.3667
                     "A2" ⇒ 0.6333
                     "A3" ⇒ -0.2667
                     else ⇒ . }
        + Match(B) { "B1" ⇒ -0.1917
                     "B2" ⇒ -0.2083
                     "B3" ⇒ 0.475
                     "B4" ⇒ -0.075
                     else ⇒ . }
        + Match(A) { "A1" ⇒ Match(B) { "B1" ⇒ -1.1833
                                       "B2" ⇒ 0.1333
                                       "B3" ⇒ 0.15
                                       "B4" ⇒ 0.9
                                       else ⇒ . }
                     "A2" ⇒ Match(B) { "B1" ⇒ 0.5167
                                       "B2" ⇒ -0.1167
                                       "B3" ⇒ -0.1
                                       "B4" ⇒ -0.3
                                       else ⇒ . }
                     "A3" ⇒ Match(B) { "B1" ⇒ 0.6667
                                       "B2" ⇒ -0.0167
                                       "B3" ⇒ -0.05
                                       "B4" ⇒ -0.6
                                       else ⇒ . }
                     else ⇒ . }
```

図 5.5　固定効果の予測式

変量効果の予測値

反復誤差$\varepsilon_{(0)i}$，および 1 次誤差$\varepsilon_{(1)ij}$についての予測結果である．固定効果については，パラメータの「推定値」としているが変量効果については，次に示すように，変量効果の「予測」と使い分けがなされている．予測値は「BLUP（最良線形不偏予測量)」と表示されている．

固定効果については, 各因子の最後の水準が示されていないが, 変量効果については全ての水準, および，全ての水準の組み合わせが示されている．

変量効果の予測

項	BLUP	標準誤差	分母自由度	t値	p値(Prob>|t|)
R[R1]	0.3873	0.4528	1	0.86	0.5496
R[R2]	−0.3873	0.4528	1	−0.86	0.5496
R[R1]*A[A1]	0.2454	0.3400	1.758	0.72	0.5542
R[R1]*A[A2]	0.1814	0.3400	1.758	0.53	0.6533
R[R1]*A[A3]	−0.2450	0.3400	1.758	−0.72	0.5548
R[R2]*A[A1]	−0.2454	0.3400	1.758	−0.72	0.5542
R[R2]*A[A2]	−0.1814	0.3400	1.758	−0.53	0.6533
R[R2]*A[A3]	0.2450	0.3400	1.758	0.72	0.5548

固定効果に変量効果の予測量（BLUP）を加えたy_{ijk}に対する推定値は，【列の保存／条件付き予測式】を選択すると JMP データセットに追加出力されるので，Excel にコピーし反復 R について並列した形式に整理した結果を 表 5.6 に示す．

表 5.6　条件付き予測式

		y		予測式 y		条件付き予測式 y	
A	B	R1	R2	R1	R2	R1	R2
A1	B1	12.3	11.7	12.00	12.00	12.63	11.37
	B2	13.9	12.7	13.30	13.30	13.93	12.67
	B3	15.0	13.0	14.00	14.00	14.63	13.37
	B4	15.0	13.4	14.20	14.20	14.83	13.57
A2	B1	15.5	13.9	14.70	14.70	15.27	14.13
	B2	14.8	13.3	14.05	14.05	14.62	13.48
	B3	15.2	14.3	14.75	14.75	15.32	14.18
	B4	14.4	13.6	14.00	14.00	14.57	13.43
A3	B1	13.9	14.0	13.95	13.95	14.09	13.81
	B2	13.3	13.2	13.25	13.25	13.39	13.11
	B3	14.3	13.5	13.90	13.90	14.04	13.76
	B4	12.8	12.8	12.80	12.80	12.94	12.66

固定効果は，反復 R1 と R2 で同じであるが，変量効果を加味した条件付き予測式は，図 5.5 に示した固定効果の計算式に，次の変量効果を加えた式である．

$$\hat{y}_{ijk} = \hat{\mu}((AB)_{jk}) + \text{If}\begin{cases} R==1 \Rightarrow 0.3873 \\ R==2 \Rightarrow -0.3873 \\ \text{else} \Rightarrow 0 \end{cases} + \text{If}\begin{cases} R==1\ \&\ A==\text{"A1"} \Rightarrow 0.2454 \\ R==1\ \&\ A==\text{"A2"} \Rightarrow 0.1814 \\ R==1\ \&\ A==\text{"A3"} \Rightarrow -0.245 \\ R==2\ \&\ A==\text{"A1"} \Rightarrow -0.2454 \\ R==2\ \&\ A==\text{"A2"} \Rightarrow -0.1814 \\ R==2\ \&\ A==\text{"A3"} \Rightarrow 0.245 \\ \text{else} \Rightarrow 0 \end{cases}$$

実際に計算してみると，

$$\hat{y}_{111} = 12.00 + 0.3873 + 0.2454 = 12.63$$
$$\hat{y}_{112} = 12.00 - 0.3873 + 0.2454 = 12.37$$
$$\vdots$$
$$\hat{y}_{234} = 12.80 - 0.3873 + 0.2454 = 12.66$$

と，データセットに追加された値と一致する．

変量効果の推定

変量効果とした R および R×A についても，【効果の詳細】で固定効果と同様に水準平均と標準誤差が出力される．変量効果 R について次の結果が得られる．R1 と R2 の水準平均は，14.13 と 13.35 となっているが，この推定値は R1 と R2 について 12 個のデータの平均値 14.20 と 13.28 は異なり，R1 はやや低めに，R2 は高めに推定されている．

REML

最小2乗平均表

水準	最小2乗平均	標準誤差	下側95%	上側95%
R1	14.13	0.2451	13.2096	15.0483
R2	13.35	0.2451	12.4351	14.2737

これは，総平均 13.7417 に反復 R の BLUP 推定（0.3873，-0.3873）を加えた結果になっている．変量効果 R×A については，

$$\hat{\mu}((RA)_{ij}) = 13.7417 + \mathrm{Match}(A)\begin{cases} \text{"A1"} \Rightarrow -0.3667 \\ \text{"A2"} \Rightarrow 0.6333 \\ \text{"A3"} \Rightarrow -0.2667 \\ \text{else} \Rightarrow . \end{cases} + \mathrm{If}\begin{cases} R==1 \Rightarrow 0.3873 \\ R==2 \Rightarrow -0.3873 \\ \text{else} \Rightarrow 0 \end{cases} + \mathrm{If}\begin{cases} R==1\ \&\ A==\text{"A1"} \Rightarrow 0.2454 \\ R==1\ \&\ A==\text{"A2"} \Rightarrow 0.1814 \\ R==1\ \&\ A==\text{"A3"} \Rightarrow -0.245 \\ R==2\ \&\ A==\text{"A1"} \Rightarrow -0.2454 \\ R==2\ \&\ A==\text{"A2"} \Rightarrow -0.1814 \\ R==2\ \&\ A==\text{"A3"} \Rightarrow 0.245 \\ \text{else} \Rightarrow 0 \end{cases}$$

によって計算され，次の結果として出力されている．ただしこの結果は，反復 R でも述べたように，算術平均の結果とは異なり，組み合わせ平均が総平均に対して高めの場合は，やや低めに，低めの場合はやや高めになる．

最小2乗平均表

水準	最小2乗平均	標準誤差	下側95%	上側95%
R1,A1	14.0077	0.1650	13.6398	14.3756
R1,A2	14.9437	0.1650	14.5758	15.3116
R1,A3	13.6173	0.1650	13.2494	13.9852
R2,A1	12.7423	0.1650	12.3744	13.1102
R2,A2	13.8063	0.1650	13.4384	14.1742
R2,A3	13.3327	0.1650	12.9648	13.7006

5.4. 実験数の削減

要因配置型の実験計画をベースに，各種の分割実験を実施することは，ランダム化の回数を減らすことになり，実験に要する時間，実験の費用の軽減に役に立つ．直交表を用いた実験計画は，高次の交互作を無視することにより，少ない実験数で多くの因子を同時に検討することを実現した．

分割実験でも，実験精度を劣化させない程度に実験回数を減らすことを考えてみよう．2.5 節で取り上げた「反復を伴う分割実験」についての分散分析表から，3 次誤差の自由度は 9 と 2 次誤差の自由度 2 に比べて大きいくアンバランスである．3 次誤差の自由度が 3 まで落とすことが出来ないものか．表 5.7 に示すように，反復 R2 の場合に，因子 B の 4 水準を意図的に間引いて 2 水準に限定したとしよう．この結果，実験データ数は，24 から 18 に減少させることができる．

表 5.7　反復のある 1 段分割実験のデータの削減

R	A	B1	B2	B3	B4	実験回数
R1	A1（2）	12.3	13.9	15.0	15.0	4
	A2（1）	15.5	14.8	15.2	14.4	4
	A3（3）	13.9	13.3	14.3	12.8	4
R2	A1（5）	11.7	12.7	(13.0)	(13.4)	2
	A2（6）	13.9	(13.3)	(14.3)	13.6	2
	A3（4）	(14.0)	13.2	13.5	(12.8)	2
実験回数		5	5	4	4	18

意図的に実験回数を減らした場合に，平均平方の期待値はどう変化するのであろうか．EMS 法により平均平方の期待値を次に示す．因子 B および交互作用 A×B の成分の係数は整数倍ではなくなり，端数がでている．

EMS

期待平均平方

各行の平均平方の期待値を構成する各列の分散成分の係数
期待平均平方

	切片	反復R&変量効果	A	A*反復R&変量効果	B	A*B
切片	0	0	0	0	0	0
反復R&変量効果	0	6	0	2	0	0
A	0	0	4	2	0	1
A*反復R&変量効果	0	0	0	2	0	0
B	0	0	0	0	3.64898	1.21633
A*B	0	0	0	0	0	1.26613

プラス1.0倍の残差誤差分散

表 5.8 に平均平方の期待値を，完備型の場合と並列して示す．反復 R の場合 $\sigma^2_{(0)}$ の係数は，完備型の 12 から 6 に減少している．これは反復 R2 でのデータ数が 12 から 6 に減少したことに対応している．因子 A の場合は 8 から 4 に減少しているが，これは，R1 と R2 のデータ数 8 から，R1 のみのデータ数 4 に対応していると思われる．1 次誤差 $\sigma^2_{(1)}$ の係数は 2 は，反復 R2 の繰返し数の 2 に対応していると思われる．

表 5.8　平均平方の期待値の比較

	完備型		欠測セルが 6 個ある場合	
要因	自由度	平均平方の期待値	自由度	平均平方の期待値
R:反復誤差 e_0	1	$\sigma^2_{(2)}+4\,\sigma^2_{(1)}+12\,\sigma^2_{(0)}$	1	$\sigma^2_{(2)}+2\,\sigma^2_{(1)}+6\,\sigma^2_{(0)}$
A	2	$\sigma^2_{(2)}+4\,\sigma^2_{(1)}+8\,\sigma^2_{A}$	2	$\sigma^2_{(2)}+2\sigma^2_{(1)}+4\,\sigma^2_{A}$
R×A:1 次誤差 e_1	2	$\sigma^2_{(2)}+4\,\sigma^2_{(1)}$	2	$\sigma^2_{(2)}+2\,\sigma^2_{(1)}$
B	3	$\sigma^2_{(2)}+6\,\sigma^2_{B}$	3	$\sigma^2_{(2)}+3.65\,\sigma^2_{B}$
A×B	6	$\sigma^2_{(2)}+2\sigma^2_{A\times B}$	***6***	$\sigma^2_{(2)}+1.27\,\sigma^2_{A\times B}$
2 次誤差 e_2	9	$\sigma^2_{(2)}$	***3***	$\sigma^2_{(2)}$
全体	23		***17***	

REML 法による変量効果の成分，固定効果についての検定結果を示し，それらを 表 5.9 の分散

分析表としてまとめる．

REML

変量効果	分散比	分散成分	標準誤差	95%下側	95%上側	全体に対する百分率
反復R	2.72	0.3375	0.5114	0.0631	926.5830	71.7
A*反復R	0.07	0.0092	0.0874	0.0017	1.07e+142	1.9
残差		0.1242	0.1014	0.0398	1.7262	26.4
合計		0.4708				100.0

REML

要因	パラメータ数	自由度	分母自由度	F値	p値(Prob>F)
A	2	2	0.87	7.26	0.2870
B	3	3	3.49	4.39	0.1086
A*B	6	6	3.32	5.97	0.0727

表 5.9　反復のある 1 段分割実験の REML 法での分散分析表

要因	分散成分	自由度	分母自由度	F 値	p 値
R:反復誤差 e_0	0.3375	1			
A		2	0.87	7.26	0.2870
R×A:1 次誤差 e_1	0.0092	2			
B		3	3.49	4.39	0.1086
A×B		6	3.32	5.97	0.0727
2 次誤差 e_2	0.1242	3			
全体	0.4708	17			

反復 R2 内の因子 B について，実験数を半分に削減しても，実験コストの削減にさほど寄与しないのであれは，反復 R2 での因子 A のある水準自体を除いてしまうことも考えられる．仮のデータをいれて，分割実験の解析をしてみて判断することを薦める．

6. スタッガード型枝分かれ実験

6.1. 完備型の枝分かれ実験の問題点

誤差の自由度のインフレーション

枝分かれ実験計画は, 固定効果が総平均一つで, 複数の変量効果を含んでいる. そのために, REML 法の本質を理解し，計算プロセスを習得するために適した題材である．また，多くの実験計画法の教科書で取り上げられているテーマの一つでもあり，分析法のバリデーションの領域では、JIS 規格 Z 8404 も制定されている．JIS 規格も含めて一般的な実験計画法の教科書には，手計算でも分散成分の推定ができるように複数の因子に対して完備型の事例が示されている.

ところが，完備型の枝分かれ実験は，一般的に分散成分が小さくなる下位の「分析」などの因子に多くの繰り返しが求められ，その結果として自由度のインフレーションが起きる．スタッガード型（じぐざぐ型）枝分かれ実験は，複数の誤差因子の自由度が均一となるように工夫された不完備型の実験計画で，手計算で分散成分を計算する手順も考案されいる.

分散成分の推定は， JMP の EMS 法と同様の解析は，GLM プロシジャによって行なうことができた．また VARCOMP プロシジャは分散成分の推定に特化しており，REML 法も含め各種の分散成分の推定が行なえるようになっている．JMP の「モデルのあてはめ」での REML 法による解析は，SAS の MIXED プロシジャによって実施できる.

JMP の「モデルのあてはめ」での REML 法による解析が適切に行なわれているのか，使い方は適切なのか，新しい統計ソフトを使うときには，使用経験のある統計ソフトと比較検討しておくことも必要である．スタッガード型枝分かれ実験の解説をしつつ，SAS のプロシジャ，および，JMP の「モデルのあてはめ」と対比しながら分散成分の推定, 95%信頼区間の計算方法についても示す.

同一ロット A から 2 個の試料 B をランダムに採取し, この試料に対して分析 C を 2 回繰り返す. この実験を 3 ロットについて行ない，3×2×2 ＝ 12 回 の分析を行なった．繰返し 2 の 2 元配置分散分析を行い，分散分析表の平均平方の期待値の構造から分散成分の推定を行なうことが容易にできる.

表 6.1　枝分れが 2 段の繰り返しが揃っているデータ

B	C	A_1	A_2	A_3
B_1	C_1	94	107	86
	C_2	96	108	87
B_2	C_1	90	117	85
	C_2	92	116	83

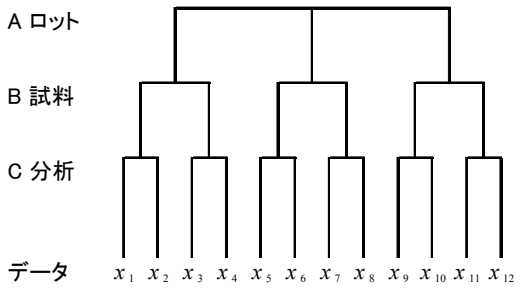


図 6.1 繰返しが揃っている枝分かれ実験

スタッカード型枝分れ実験

分析に時間と費用がかかる場合には，試料 B の第 2 水準の分析 C を 2 回ではなく 1 回で済ませたい．この実験計画は，ロット A，試料 B，分析 C に関する平均平方の自由度を均一化させる効果がある．このような実験計画は，スタッカード（じぐざぐ）型計画として知られている．

表 6.2　スタッカード型の枝分かれデータ　“ . ”は測定値なし

B	C	A_1	A_2	A_3
B_1	C_1	94	107	86
	C_2	96	108	87
B_2	C_1	90	117	85
	C_2	.	.	.

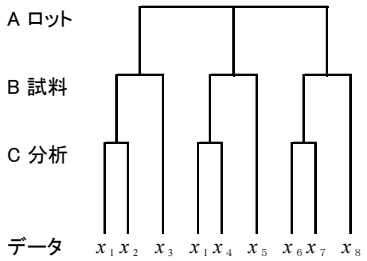


図 6.2 スタッカード型枝分かれ実験

6.2. SAS および JMP による分散成分の推定

完備型データでの分散成分

SAS には古くから分散成分を推定するための専用の VARCOMP プロシジャが提供されている．バージョン 9.1 では，分散成分の推定に 4 種類の推定方法（Type1，MIVQUE0，ML，REML）が選択できるようになっている．Default は，MIVQUE0 であるが，GLM プロシジャの random ステートメントで出力される平均平方の期待値の構造を出力するためには type1（逐次型平方和）を用いる．表 6.1 に示した完備型の枝分かれ実験データについて VARCOMP プロシジャで分散成分の推定を

行なってみよう．

Program 1　完備型のデータ

```
data d01 ;
    input B $ C$ @@ ;
    do A = 'A1', 'A2', 'A3' ;
      input y @@ ;
      output ;
    end ;
datalines ;
B1  C1  94  107  86
B1  C2  96  108  87
B2  C1  90  117  85
B2  C2  92  116  83
;
proc varcomp method=type1 ;
    class        A  B  C ;
    model  y  =  A  B(A) ;
    run ;
```

Output 1　完備型のデータに対する Type1 での分散成分の推定

Type 1 Analysis of Variance

Source	DF	Sum of Squares	Mean Square	Expected Mean Square
A	2	1515.500000	757.750000	Var(Error) + 2 Var(B(A)) + 4 Var(A)
B(A)	3	103.250000	34.416667	Var(Error) + 2 Var(B(A))
Error	6	7.500000	1.250000	Var(Error)
Corrected Total	11	1626.250000	.	.

Type 1 Estimates

Variance Component	Estimate
Var(A)	180.83333
Var(B(A))	16.58333
Var(Error)	1.25000

枝分かれ型分散分析を行なうために　model　y = A　B(A) ; のように因子 A を括弧で括っている．これは，因子 A の水準ごとに枝分かれしている因子 B の水準は単なる繰返しであり，因子 B の主効果および交互作用 A*B は意味を持たないので，それらを合わせた平方和を求めることを意味している．Error 項の平方和 7.25 は，表 6.3 に示すように繰返しがない 3 元配置分散分析を形式的に行なったときの因子 C の主効果，および，2 因子交互作用 A*C，B*C ，3 因子交互作用の平方和の小計となっている．

表 6.3　3元配置分散分析による平方和の分解

Source	DF	Type I SS	小計	枝分かれ因子
A	2	1515.500000	1515.50	A
B	1	2.083333		
A*B	2	101.166667	103.25	B(A)
C	1	0.750000		
A*C	2	3.500000		
B*C	1	2.083333		
A*B*C	2	1.166667	7.25	Error, C(A B)
全体	11		1626.25	C.Total

平均平方の期待値の構造が output 1 に示されているので，これから分散成分を次式のように推定できる．

$$\hat{\sigma}_e^2 = \mathrm{Var(Error)} = \mathrm{Var(C(A\,B))} = 1.2500$$

$$\hat{\sigma}^2_{B(A)} = \mathrm{Var(B(A))} = \frac{34.4167 - 1.25}{2} = 16.5833$$

$$\hat{\sigma}^2_{A} = \mathrm{Var(A)} = \frac{757.75 - 34.4167}{4} = 180.8333$$

試料 B(A)の自由度 3 に対して，Error 項，分析 C(A B)の自由度は 6 と倍になる．試料 B(A)の自由度が 3 と小さく推定値として不安定なので，自由度を 10 になるよにするためには，因子 A を 10 ロットにする必要がある．すると，分析 C の自由度は 20 となり，元々分散が相対的に小さい分析 C の分散成分の自由度が 20 もある必要はなく，自由度を 10 程度に抑えたい．どのようにしたらよいのだろうか.

JMP/EMS 法による分散成分の推定値を求めた．結果を次に示す．VARCOMP プロシジャの type1 法による結果と一致している.

期待平均平方

各行の平均平方の期待値を構成する各列の分散成分の係数
期待平均平方

	切片	A&変量効果	B[A]&変量効果
切片	0	0	0
A&変量効果	0	4	2
B[A]&変量効果	0	0	2

プラス1.0倍の残差誤差分散

分散成分推定値

成分	分散成分推定値	全体に対する百分率
A&変量効果	180.8333	91.0
B[A]&変量効果	16.5833	8.3
残差	1.2500	0.6
合計	198.6667	100.0

平均平方がその期待値に等しいものとして推定したものです。

スタッカード型データにでの分散成分

表 6.2 に示したスタッカード型の分かれ実験データについて VARCOMP プロシジャで分散成分の推定を行なってみよう．Data ステップで B_2C_2 の行のデータをピリオド“ . ”にして欠測値に変更する.

Program 2　スタッカード型のデータ

```
data d01 ;
    input B $ C$ @@ ;
    do A = 'A1', 'A2', 'A3' ;
      input y @@ ;
      output ;
    end ;
datalines ;
B1  C1  94  107  86
B1  C2  96  108  87
B2  C1  90  117  85
B2  C2   .    .   .
;
proc varcomp method=type1 ;
    class       A  B  C ;
    model  y  =  A  B(A) ;
    run ;
```

完備型データの場合には，平均平方の期待値の構造は，分散成分を整数倍した形式であったが，スタッカード型の場合には，output 2 に示すように実数となり分散成分の計算も煩雑になる.

Output 2　スタッカード型のデータに対する type1 での分散成分の推定

Type 1 Analysis of Variance

Source	DF	Sum of Squares	Mean Square	Expected Mean Square
A	2	962.666667	481.333333	Var(Error) + 1.6667 Var(B(A)) + 3 Var(A)
B(A)	3	78.333333	26.111111	Var(Error) + 1.3333 Var(B(A))
Error	3	3.000000	1.000000	Var(Error)
Corrected Total	8	1044.000000	.	.

Type 1 Estimates

Variance Component	Estimate
Var(A)	149.64815
Var(B(A))	18.83333
Var(Error)	1.00000

$$\hat{\sigma}_e^2 = \mathrm{Var(Error)} = \mathrm{Var(C(A\ B))} = 1.0000$$

$$\hat{\sigma}_{B(A)}^2 = \mathrm{Var(B(A))} = \frac{26.11111 - 1.0}{1.3333} = 18.8333$$

$$\hat{\sigma}_A^2 = \mathrm{Var(A)} = \frac{481.3333 - 1.6667 \times 18.8333 - 1.0}{3} = 149.6481$$

試料 B(A)の自由度 3 に対して，error 項，分析 C(A B)の自由度は 3 と同じ自由度となることが確認できる．分散成分の推定方法としては，type1 の推定方法は，最小 2 乗法を基盤とする古い時代の便宜的な方法であるが，多くのテキストで取り上げられていることもあり，分散成分の計算手順の理解を深めるために示した．

JMP/EMS 法での分散成分の推定を次に示す．因子 A&変量効果の平均平方の期待値の係数が 2.6667 となっているのに対し，VARCOM プロシジャでは，Var(A)の係数 3 と異なる．これは，VARCOM プロシジャでは，逐次型（type1）の平方和に対して解析しているのに対して，EMS 法は type3 の平方和を用いているためと思われる．

期待平均平方

各行の平均平方の期待値を構成する各列の分散成分の係数

期待平均平方

	切片	A&変量効果	B[A]&変量効果
切片	0	0	0
A&変量効果	0	2.66667	1.33333
B[A]&変量効果	0	0	1.33333

プラス1.0倍の残差誤差分散

分散成分推定値

成分	分散成分推定値	全体に対する百分率
A&変量効果	179.8542	90.1
B[A]&変量効果	18.8333	9.4
残差	1.0000	0.5
合計	199.6875	100.0

平均平方がその期待値に等しいものとして推定したものです。

EMS 法での平方和を次に示す．逐次型平方和は，指定した要因を順番にモデルに入れて，モデルの平方和の増加分を「平方和」としている．VARCOMP プロシジャで最初にモデルの入れたのは因子 A であり，これは B(A)の影響を受けていない．そのために平方和は 926.6667 となり，EMS 法の 1011.4444 と異なる．Type3 の平方和は，全ての要因をモデルに含めた状態から，因子 A を除い

たときの平方和の減少分としているためである．

変量効果を考慮した検定

要因	平方和	分子平均平方	分子自由度	F値	p値(Prob>F)
A&変量効果	1011.4444	505.7222	2	19.37	0.0193*
B[A]&変量効果	78.3333	26.1111	3	26.11	0.0119*

REML法による分散成分の推定

VARCOMP プロシジャでは，REML 法による分散成分の推定ができるようになっている．線形混合モデルに対する MIXED プロシジャで標準的に用いられている推定方法でもある．

表 6.1 および
表 6.2 で示してきた枝分かれ実験データは，総平均のみを固定効果，因子 A を変量効果，枝分れ因子 B(A) も変量効果とする線形混合モデルとして定義できる．VARCOMP プロシジャと MIXED プロシジャで REML 法による分散成分を推定してみよう．VARCOMP プロシジャのオプションでは method=reml と指定する．

Program 3　VARCOM プロシジャの REML オプション

```
proc varcomp data=d02 method=reml ;
    class        A  B  C ;
    model  y  =  A  B(A) ;
    run ;
```

Output 3 に示すように REML 法で推定された分散成分は，Var(A)=178.7098，Var(B(A))=18.7484，Var(Error)=1.0042 と Type1 の場合 Var(A)=149.6481, Var(B(A))=18.8333, Var(Error)=1.0000 と異なる．

VARCOMP プロシジャでは, model ステートメントで変量効果として因子 A, 枝分かれ因子 B(A) を指定したが，MIXED プロシジャでは，model ステートメントでは固定効果のみを，変量効果をすべて random ステートメントで指定する．

Output 3　スタッカード型のデータに対する REML 法（VARCOMP）での分散成分の推定

REML Estimates

Variance Component	Estimate
Var(A)	178.70983
Var(B(A))	18.74847
Var(Error)	1.00421

Asymptotic Covariance Matrix of Estimates

	Var(A)	Var(B(A))	Var(Error)
Var(A)	35566.6	-110.75127	-0.77320
Var(B(A))	-110.75127	254.01340	-0.56378
Var(Error)	-0.77320	-0.56378	0.67799

枝分かれ実験の場合の固定効果は，総平均（切片項）のみなので，model ステートメントでは何も指定しない．これは，線形モデルに対する統計ソフトの通例として切片項（総平均）を解析用の

デザイン行列 $\boldsymbol{X}$ に自動的に加えることを default としているためである．Random ステートメントでは，変量効果として因子 A，枝分かれ因子 B(A)を指定する．プロシジャのオプションで分散成分の 95%信頼区間を出力するために cl=wald，SE を出すために covtest を追加しておく．MIXED プロシジャは REML 法が標準なので推定方法の method オプションで reml を指定する必要はない．

Program 4　MIXED プロシジャによる分散成分の推定

```
proc mixed data=d02 cl=wald covtest ;
    class        A  B  C ;
    model  y  =    ;
    random       A  B(A) ;
    run ;
```

Output 4　スタッカード型のデータに対する REML 法（MIXED）での分散成分の推定

Covariance Parameter Estimates

Cov Parm	Estimate	Standard Error	Z Value	Pr Z	Alpha	Lower	Upper
A	178.71	188.59	0.95	0.1717	0.05	46.2398	10118
B(A)	18.7485	15.9378	1.18	0.1197	0.05	5.8236	310.35
Residual	1.0042	0.8234	1.22	0.1113	0.05	0.3212	14.2087

MIXED プロシジャでの分散成分の推定値は，VARCOMP プロシジャの REML 法による推定結果と一致している．分散成分の 95%信頼区間は，χ^2 分布のパーセント点を用いた次式が用いられている．

$$\frac{\nu\hat{\sigma}^2}{\chi^2_{\nu,1-\alpha/2}} \le \sigma^2 \le \frac{\nu\hat{\sigma}^2}{\chi^2_{\nu,\alpha/2}}\text{，ただし，}\ \nu = 2\left(\hat{\sigma}^2 / SE(\hat{\sigma}^2)\right)^2$$

因子 A について 95%信頼区間の下限 46.24 は，

$$\nu = 2\left(\hat{\sigma}^2_A / SE(\hat{\sigma}^2_A)\right)^2 = 2\times(178.71/188.59)^2 = 1.7959$$

$$\frac{\nu\hat{\sigma}^2}{\chi^2_{\nu,1-\alpha/2}} = \frac{1.7959\times 178.71}{6.9410} = 46.24$$

と計算されている．VARCOMP プロシジャでは，Output 3 に示すように分散成分の 95%信頼区間は計算されていないが，分散共分散行列が出力されているので，対角要素の平方根が $SE(\hat{\sigma}^2_A)$ となることから別途計算が可能である．因子 A の場合

$$SE(\hat{\sigma}^2_A) = \sqrt{Var(\hat{\sigma}^2_A)} = \sqrt{35566.6} = 188.59$$

が，MIXED プロシジャの因子 A の SE で出力された結果 188.59 と一致する．

JMP は，Excel シート上のデータとカット&ペーストでのデータ交換も自在で，SAS データセットを JMP で取り込むこと，JMP データセットを SAS データセットとして出力することも標準機能として備わっている．これらの機能を使い，SAS で加工あるいはシミュレーションなどを行なった結果を SAS データセットに出力し，JMP でそのまま読み込み様々な高品質のグラフを作成することにより統計解析のスピードおよび質の向上がはかられる．JMP もバージョンを重ねるごとに統計解析の機能が充実し，現在では，スタッカード型データに対する分散成分の推定も容易にできる．

JMP で枝分かれ型データに対する分散成分の推定は「モデルのあてはめ」のモデル効果の構成ボックスに，SAS と同様に因子 A，枝分かれ因子 B（A）をセットし「属性▼」のプルダウンメニューから変量効果を選択して因子の属性変換を行なう．変量効果を因子に含めると，計算方法は REML（推奨）法に自動的に切り替わる．

☑分散成分の範囲制限なし にレ点が付いている．このままだと分散成分の 95%信頼区間の推定で，標準誤差のパーセント点を用いるために，下限が負となる場合も生じ不適切となるので，外す必要がある．次に結果を示すが，SAS と一致している

REML

変量効果	分散比	分散成分	標準誤差	95%下側	95%上側	全体に対する百分率
A	177.96	178.7098	188.5911	46.2398	10118.282	90.0
B[A]	18.67	18.7485	15.9378	5.8236	310.3477	9.4
残差		1.0042	0.8234	0.3212	14.2087	0.5
合計		198.4625				100.0

−2対数尤度= 47.362362396

SAS の MIXED プロシジャの計算式を用いて，3 つの分散成分の 95%信頼区間の計算について追試を行なった結果を示す．結果は一致した．

	σ2^	se	ν	chi(0.975, ν)	chi(0.025, ν)	95%L	95%U
1	178.7100	188.5900	1.7959	6.9410	0.0317	46.2401	10117.7943
2	18.7485	15.9378	2.7676	8.9101	0.1672	5.8236	310.3458
3	1.0042	0.8234	2.9747	9.3012	0.2102	0.3212	14.2090

$$\nu = \left[\left[\frac{\sigma 2\hat{}}{se}\right]^2 * 2\right], \quad L95\% = \frac{[\nu * \sigma 2\hat{}]}{chi(0.975, \nu)} \quad U95\% = \frac{[\nu * \sigma 2\hat{}]}{chi(0.025, \nu)}$$

SASユーザ会2011年のでの論文集に対する補足

☑分散成分の範囲制限なし のままの解析を行なうと，次のような不都合が起きる．

図 6.3 に示すように分散成分の欄から因子 A の分散成分 178.71 となり，SAS の VARCOMP および MIXED プロシジャの結果と一致する．ところが 95%信頼区間は因子 A の場合に（-190.92～548.34）となり $SE(\hat{\sigma}_A^2)$ と t 分布のパーセント点から便宜的に計算された結果が示されている．*Var*

(B(A))も同様であるが，*Var*(Error)は，χ^2分布のパーセント点を用いた計算をしているので，95%信頼区間の下限が負とはなっていない．JMP を使っていて分散成分の 95%信頼区間が必要な場合は，分散共分散行列が出力されているので，その対角要素から別途χ^2分布のパーセント点を用いた計算を行なう必要がある．

REML法による分散成分推定値

変量効果	分散比	分散成分	標準誤差	95%下側	95%上側	全体に対する百分率
A	177.9607	178.7096	188.5907	-190.92	548.34	90.0
B[A]	18.6699	18.7485	15.9378	-12.49	49.99	9.4
残差		1.0042	0.8234	0.32	14.21	0.5
合計		198.4623				100.0

-2対数尤度= 47.362362396

分散成分推定値の共分散行列

変量効果	A	B[A]	残差
A	35566.459	-110.7515	-0.7730
B[A]	-110.7515	254.0137	-0.5638
残差	-0.7730	-0.5638	0.6780

図 6.3　スタッカード型データに対する JMP の REML 法による分散成分の推定

6.3. 分散成分の推定方式（芳賀本の事例との対比）

JMP の分散成分の 95%信頼区間の推定方式については，どのような推定式を用いているかが JMP のドキュメントに示されていない．分散成分の 95%信頼区間について，幾つかの方式が知られていて，芳賀（2010）の「医薬品開発のための統計解析 第 2 部実験計画法」の「第 6 章 変量模型，枝分れ実験」に詳しく述べられている．

芳賀本では，JMP の REML 法による分散成分の 95%信頼区間が従来の方法と異なり，下限も上限も大きめに推定されていると述べている．そこで，芳賀の示した事例を追試し，JMP での計算方式を適用した結果とを比較する．

EMS

分散分析

要因	自由度	平方和	平均平方	F値
モデル	5	3352.9444	670.589	7.4556
誤差	12	1079.3333	89.944	p値(Prob>F)
全体(修正済み)	17	4432.2778		0.0022*

期待平均平方

各行の平均平方の期待値を構成する各列の分散成分の係数
期待平均平方

	切片	B&変量効果
切片	0	0
B&変量効果	0	3

プラス1.0倍の残差誤差分散

分散成分推定値

成分	分散成分推定値	全体に対する百分率
B&変量効果	193.5481	68.3
残差	89.9444	31.7
合計	283.4926	100.0

平均平方がその期待値に等しいものとして推定したものです。

	B	W	y
1	B1	1	120
2	B1	2	125
3	B1	3	137
4	B2	1	139
5	B2	2	138
6	B2	3	130
7	B3	1	118
8	B3	2	125
9	B3	3	109
10	B4	1	132
11	B4	2	148
12	B4	3	135
13	B5	1	159
14	B5	2	146
15	B5	3	160
16	B6	1	102
17	B6	2	132
18	B6	3	110

芳賀本

分散分析表

要因	平方和	自由度	平均平方	F 比	p 値	分散成分	標準偏差	E[平均平方]
水準間 (B)	3352.94	5	670.59	7.456	0.002	193.55	13.91	$\sigma_W^2 + r\sigma_B^2$
水準内 (W)	1079.33	12	89.94	1.000		89.94	9.48	σ_W^2
全体	4432.28	17	260.72					
(検算)	4432.28	17						

EMS 法による分散分析表，分散成分の推定値は，芳賀本の結果に一致している．REML 法により，分散成分も EMS 法と同じ結果になっている．

REML

変量効果	分散比	分散成分	標準誤差	95%下側	95%上側	全体に対する百分率
B	2.15	193.5481	141.9014	67.5895	1797.8454	68.3
残差		89.9444	36.7197	46.2505	245.0920	31.7
合計		283.4926				100.0

JMP で「誤差」項の分散成分の 95%信頼区間は，(46.25，245.09) である．JMP の計算方式は異なるが，平方和の自由度が 12 となり，次に示す芳賀本の結果と一致する．

芳賀本

$\hat{\sigma}_W^2$ の区間推定は，第 1 部 §2.5(2) で説明した方法で得られる．

$$\frac{S_W}{\chi^2(\nu_W, \alpha/2)} < \sigma_W^2 < \frac{S_W}{\chi^2(\nu_W, 1-\alpha/2)} \tag{6.1.10}$$

$$\frac{1079.33}{23.337} < \sigma_W^2 < \frac{1079.33}{4.404}$$

$$46.25 < \sigma_W^2 < 245.09$$

REML 法の変量効果 B の分散成分の 95%信頼区間は，(67.59，1794.85) であるのに対し，芳賀本

では，（57.11，1424.62）と下限・上限ともに低めに推定されている．

芳賀本

$$\frac{S_B/\chi^2(\nu_B,\alpha/2)-V_W}{r} < \sigma_B^2 < \frac{S_B/\chi^2(\nu_B,1-\alpha/2)-V_W}{r} \tag{6.1.12}$$

$$\frac{3352.94/12.832-89.94}{3} < \sigma_B^2 < \frac{3352.94/0.831-89.94}{3}$$

$$\frac{261.29-89.94}{3} < \sigma_B^2 < \frac{4033.82-89.94}{3} \tag{6.1.13}$$

$$57.11 < \sigma_B^2 < 1314.62$$

芳賀本では計算式が示されていないが，森口の式での結果は 49.04 < σ_B^2 < 1314.67 としている．これとも REML 法の結果は，異なる

そこで，芳賀本の事例で，REML 法での計算方式を適用した結果を次に示す．分散成分の 95%信頼区間の計算方式については，SAS の MIXED プロシジャで用いられている計算式を適用した．

分散成分の 95%信頼区間の計算

	要因	σ2^	se	ν	chi(0.975, ν)	chi(0.025, ν)	95%L	95%U
1	B	193.5482	141.9014	3.7208	10.6548	0.4006	67.5895	1797.8453
2	W	89.9444	36.7197	12.0000	23.3367	4.4038	46.2505	245.0920

$$\nu = \left[\left(\frac{\sigma 2\hat{}}{se}\right)^2 * 2\right], \quad L95\% = \frac{(\nu * \sigma 2\hat{})}{chi(0.975, \nu)} \quad U95\% = \frac{(\nu * \sigma 2\hat{})}{chi(0.025, \nu)}$$

結果は，因子 B の分散成分の 95%信頼区間は（67.58，179784），因子 W の分散成分の 95%信頼区間も同じ計算式で（46.25，245.09）と一致した．

6.4. 今後の課題

注）分散成分推定値に対する分散共分散行列の計算方法が，SAS のマニュアルを調べたが 2011 年 12 月 13 日現在不明である．枝分れ実験では，分散成分の信頼区間を計算するのが通常なので，Excel でも計算方法を押さえる必要性がある．

シミュレーション研究による分散成分の95%信頼区間の妥当性

REML 法で用いられている分散成分の 95%信頼区間の計算式は，日本語で書かれた文献が見当たらない．そのために，REML 法の妥当性について，文献レベルでの検討，シミュレーションによる検討が必要である．

7. Excel ソルバーによる REML 法入門

7.1. Excel による追試の必要性

最小 2 乗法は，反応が計量値の場合の回帰分析，分散分析などで標準的に用いられているが，誤差が単一の場合を前提としている．したがって，複数の誤差を持つ実験データの解析には本質的に対応できない.

枝分かれ実験は，複数の誤差を持つ典型的な事例である．繰返し数が一定の場合には，1 元配置あるいは 2 元配置の分散分析を適用し，平均平方期待値の構造から複数の誤差の分散成分を推定する方法が定式化されている．意図的にバランスを崩した枝分かれ(スタッカード型)実験は，平均平方の自由度を均一化することにより，バランスの良い分散成分の推定を行なう方法として知られてはいるが，計算が煩雑であることから実際の実験の場で用いられることはまれである.

実験の局所管理の典型例として知られている乱塊法は，複数の誤差を内在している．実験因子の変更の困難さを考慮した分割実験も，複数の誤差が内在する．どちらの場合も完備した実験の場合には，分散の期待値が簡潔に示すことができ，誤差成分の計算方法が定式化されている．しかしながら，欠測値がある場合には，最小 2 乗法を用いた解析方法では，分散成分を求めることが困難になる.

複数の誤差が内在する場合の問題の解析に，制限付き最尤法（REML 法）が開発され，多くの統計ソフトに実装されるようになってきた．しかしながら，その計算過程は複雑かつ難解であり，解析の結果をブラックボックスからの出力として鵜呑みにした対応を迫られがちである．そのため，REML 法を用いた解析方法を積極的に使うこと，薦めることにためらいが生ずる．制限付きでない「最尤法」による計算方法もブラックボックス的であるのに，さらにどのような「制限」が付いているのかは，さらに説明が困難となる.

そこで，現代の算盤 Excel を用いて REML 法の計算原理を簡単な例によって示し，誰もが追試ができるような入門書の必要性を感じた．これにより，REML 法の基礎の理解を深め，各種の応用問題に対して統計ソフトを適切に使いこなして，より柔軟な実験計画力および統計解析力を付けてもらいたいと願っている.

Excel による統計解析は，結果の再現性などの保障が極めて困難なために実務で使うことは絶対に避けなければならない．しかしながら，数式を自ら入力することによる能動的な理解を期待でき，計算過程，および結果を同時に計算シート上で見ることができることが，統計解析の原理を学習するために適している。

7.2. 最小 2 乗法による母平均の推定

最も簡単な事例として，平均値と分散の推定を取り上げる．データ y_i，$i=1,2,...,n$ に対する母平均 μ と母分散 σ^2 の推定は，

$$\hat{\mu}=\bar{y}=\frac{\sum_{i=1}^{n} y_i}{n},\quad \hat{\sigma}^2=\frac{\sum_{i=1}^{n}(y_i-\bar{y})^2}{n-1}$$

として求められる．母平均の推定値 $\hat{\mu}$ は算術平均 $\bar{y}$ である．母分散の推定値 $\hat{\sigma}^2$ は y_i の $\hat{\mu}$ に対する偏差平方和を $n-1$ で割ったいわゆる不偏分散として推定している．

母平均 μ を未知とした場合に，Excel のソルバーで最小 2 乗法により $\hat{\mu}$ を推定してみよう．最小 2 乗法を用いた場合に，偏差 $(y_i-\hat{\mu})$ の平方和 S を最小にするような推定値 $\hat{\mu}$ を求めればよい．

$$S_e=\sum_{i=1}^{n}(y_i-\hat{\mu})^2$$

図 7.1 にデータ（1，2，3，4）についてソルバーを用いて母平均 $\hat{\mu}$ を最小 2 乗法によって推定した結果を示す．

ステップ 1：母平均の推定値 $\hat{\mu}$ は未知とする．

ステップ 2：母平均を仮に 0.5 として偏差平方和 S_e を計算すると

$$S_e=(1-0.5)^2+(2-0.5)^2+(3-0.5)^2+(4-0.5)^2=21.0$$

となる．

ステップ 3：S_e を最小にするように $\hat{\mu}$ をソルバーで逐次変えて行くと，$\hat{\mu}=2.5$ が得られる．

ステップ 4：$\hat{\mu}$ を手作業 $\hat{\mu}=2.4$ に変更すると S_e が 5.04，$\hat{\mu}=2.6$ に変更すると S_e が 5.04 とステップ 3 で求めた $S_e=5.00$ よりも大きくなり，最小値らしいことが確認できる．

ステップ1	ステップ2		ステップ3		ステップ4	
y	y	$(y_i-\mu\hat{})^2$	y	$(y_i-\mu\hat{})^2$	$(y_i-\mu\hat{})^2$	$(y_i-\mu\hat{})^2$
1	1	0.25	1	2.25	1.96	2.56
2	2	2.25	2	0.25	0.16	0.36
3	3	6.25	3	0.25	0.36	0.16
4	4	12.25	4	2.25	2.56	1.96
	$Se=$	21.00	$Se=$	5.00	5.04	5.04
$\mu\hat{}=?$	仮の$\mu\hat{}=$	0.5	Se を最小にする$\mu\hat{}$	2.50	2.40	2.60
			ソルバーで S が最小となるように $\mu\hat{}$を変化させた結果		$\mu\hat{}$を手作業で変化させた結果	

図 7.1　Excel のソルバーを用いた最小 2 乗法による母平均の推定

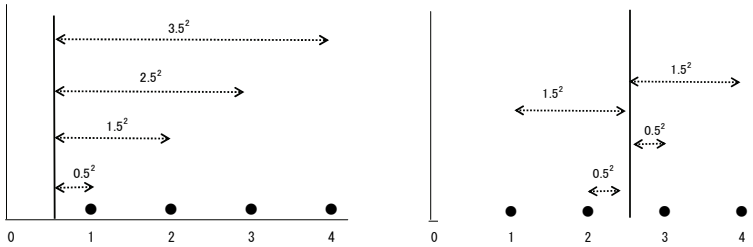


図 7.2　ステップ 1 およびステップ 2 での偏差平方和

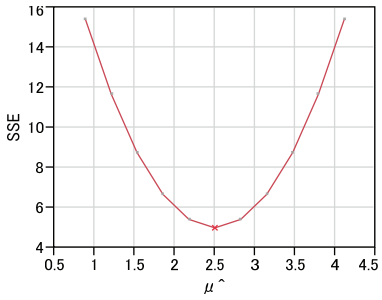


図 7.3　ステップ 4 で母平均を 2.5 からずらした場合の偏差平方和の変化

最尤法による母平均と母分散の推定

最小 2 乗法では，推定したパラメータは「平均値」のみであった．もう一つ推定したい分散は，平均値が推定された後に，平均値周りの偏差の平方を加えた偏差平方和をその自由度で割って，別途求めている．

最尤法は，データ y_i， $i=1,2,...,n$ に対して母平均の推定値を $\hat{\mu}$，母分散の推定値を $\tilde{\sigma}^2$（最尤推定により推定される分散なので “ ~ ” を付けている）を同時に推定する方法である．そのために，尤度という概念を用いて，同時推定をおこなう．母平均 μ の周りでデータが，正規分布に従うと仮定する．しかしながら，その正規分布の母分散 σ^2 も未知であると仮定する．

そのために，データ y_i が平均 $\hat{\mu}$，分散 $\tilde{\sigma}^2$ の正規分布に従うとしたとき， y_i に対応する正規分布の確率密度 $L_i = \mathrm{N}(y_i;\hat{\mu},\tilde{\sigma}^2)$ を考え，それらの積として尤度 L を定義する．

$$L = \prod_{i=1}^{n} \mathrm{N}(y_i;\hat{\mu},\tilde{\sigma}^2) = \prod_{i=1}^{n} \frac{1}{\sqrt{2\pi\tilde{\sigma}^2}} \exp\left(-\frac{1}{2\tilde{\sigma}^2}(y_i-\hat{\mu})^2\right)$$

Excel のソルバーは，適当な $\hat{\mu}$ と $\tilde{\sigma}^2$ のパラメータの推定値を出発点とし，尤度 L が大きくなるようにパラメータの値を変化させ，尤度 L がこれ以上大きくならないと判定したときに，ストップする．このとき，推定されたパラメータを最尤解という．Excel には正規分布の確率密度を計算する Normdist 関数があり，これを用いて尤度 L_i を簡単に計算できる．図 7.4 に示すように，データ（1，2，3，4）について計算する方法を示す．

ステップ 1：母平均と母分散は未知とする．

ステップ 2：仮の平均を 1.00，仮の分散を$1.2247^2 = 1.50$としたときに，それぞれのデータの正規分布（平均=1.0，標準偏差=1.2247）の確率密度は（0.3257，0.2333，0.0858，0.0162）と計算結果が示されている．これらを尤度L_iとしよう．それらの積Lは Excel のシート上で，

$$L = \prod_{i=1}^{n} L_i = 0.000106$$

と計算されている．

ステップ 3：尤度Lを最大となるように$\hat{\mu}$と$\tilde{\sigma}^2$をソルバーで自動的に変化させると，$\hat{\mu} = 2.5$と$\tilde{\sigma}^2 = 1.25$が得られる．最尤法による母分散の推定値$\tilde{\sigma}^2$は，偏差平方和S_eをデータ数nで割ったものであることが知られているので，不偏分散$\hat{\sigma}^2$は，

$$\hat{\sigma}^2 = \frac{n\tilde{\sigma}^2}{n-1} = \frac{4 \times 1.25}{4-1} = 1.67$$

として計算することができる．最小 2 乗法の場合と同じように手作業で$\hat{\mu}$を 2.5 の周りで変化させると尤度Lは小さくなる．同様に$\tilde{\sigma}^2$を 1.25 の周りで変化させても尤度Lは小さくなる．

最尤法による平均と分散の推定

ステップ1	ステップ2 (n= 4)		ステップ3 (n= 4)	
y	y	尤度 N(μ,σ~)	y	尤度 N(μ,σ~)
1	1	0.325735	1	0.145075
2	2	0.233399	2	0.322869
3	3	0.085863	3	0.322868
4	4	0.016217	4	0.145074
	L=	0.000106	L=	0.002194

μ^=？	仮のμ^=	1.00	L を最大にするμ^=	2.50
σ^{2}^=？	仮の σ^{2}~=	1.50	L を最大にする σ^{2}~=	1.25
			σ^{2}^=(n σ^{2}~)/(n−1)	1.67

ソルバーで尤度L が最大となるように
$\mu^と\sigma^{2}$~を変化させた結果

尤度関数L(y=1) = NORMDIST(H6,I12,SQRT(I13),FALSE)＝0.145075
尤度 L ＝ L1×L2×L3×L4 = 0.002194

図 7.4　平均と分散の最尤法による同時推定

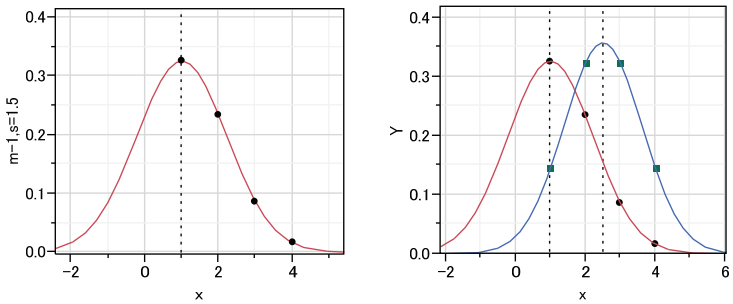


図 7.5　ステップ 2 およびステップ 3 でのデータごとの尤度

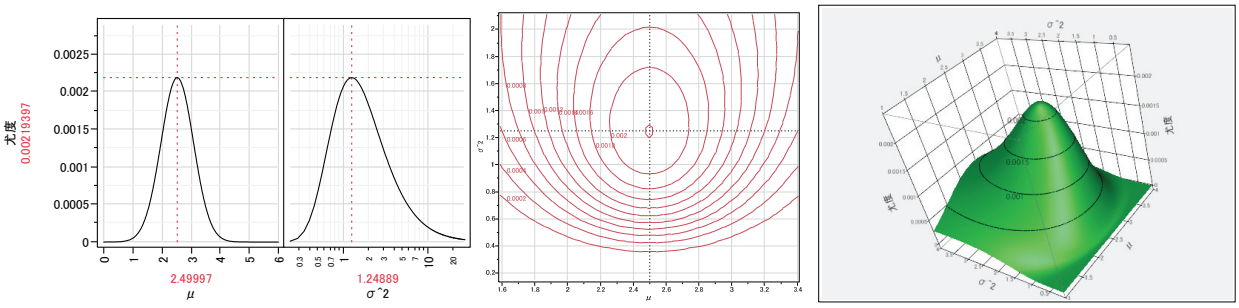


図 7.6　母平均と母分散を変化させた場合の尤度の同時分布

REML（制限付き最尤法）による母平均と母分散の推定

最尤法では，データ y_i，$i=1,2,...,n$ に対して母平均の推定値を $\hat{\mu}$，母分散の推定値を $\tilde{\sigma}^2$ を同時に推定したのであるが，母平均 $\hat{\mu}$ を「真の母平均」とみなした方法であり，$\hat{\mu}$ 自体も揺らぐことを考慮していないないために，母分散の推定値が過小評価となる．

REML 法は，母平均 $\hat{\mu}$ の分散も考慮した方法で，不偏分散を直接推定できる．制限付き最尤法を用いなくても，母平均 $\hat{\mu}$ 周りの偏差平方和を $(n-1)$ で割れば不偏分散の推定値を計算できるので，わざわざ REML 法を用いる必要は全くないが，REML 法の原理を理解するための最初の問題として適している．

最尤法は，0節で示したように，データ y_i が平均 $\hat{\mu}$，分散 $\tilde{\sigma}^2$ の正規分布に従うとしたとき，y_i に対応する正規分布の確率密度 $L_i = \mathrm{N}(y_i;\hat{\mu},\tilde{\sigma}^2)$ を考え，それらの積を尤度 L ではなく尤度 $f(\boldsymbol{y})$ とする．

$$f(y)=\prod_{i=1}^{n}\mathrm{N}(y_i;\hat{\mu},\tilde{\sigma}^2)=\prod_{i=1}^{n}\frac{1}{\sqrt{2\pi\tilde{\sigma}^2}}\exp\left(-\frac{1}{2\tilde{\sigma}^2}(y_i-\hat{\mu})^2\right)$$

不偏分散の推定値を $\hat{\sigma}^2$ としたときに，母平均 $\hat{\mu}$ の分散は，データ数 n で割った $\hat{\sigma}^2/n$ となる．最尤法によって推定された平均値を $\hat{\mu}$ ではなく $\tilde{\mu}$ と置き換え，$\tilde{\mu}$ の尤度 $g(\tilde{\mu})$ を考えよう．尤度 $g(\tilde{\mu})$ は，平均 $\hat{\mu}$，分散 $\hat{\sigma}^2/n$ とする正規分布の確率密度，

$$g(\tilde{\mu}) = \frac{1}{\sqrt{2\pi(\hat{\sigma}^2/n)}} \exp\left(-\frac{1}{2(\hat{\sigma}^2/n)} (\tilde{\mu} - \hat{\mu})^2 \right)$$

とする．RELM 法は，データに関する尤度 $f(\boldsymbol{y})$ と平均（パラメータ）に関する尤度 $g(\tilde{\mu})$ の比

$$L_{REML} = \frac{f(\boldsymbol{y})}{g(\tilde{\mu})} =$$

を最大化する方法である．図 7.7 に，データ（1，2，3，4）について計算する方法を示す．

ステップ 1：仮の平均を 1.00，仮の分散を $1.2247^2 = 1.50$ としたときに，それぞれのデータの正規分布（平均=1.0，標準偏差=1.2247）の確率密度は（0.3257，0.2333，0.0858，0.0162）と計算結果が示されている．これらを尤度 L_i としよう．それらの積 L は Excel のシート上で，

$$f(y) = \prod_{i=1}^{n} L_i = 0.000106$$

と計算されている．ここまでは，最尤法の場合と同じである．仮の平均と $\tilde{\mu}$ は同じと仮定し，平均値に関する正規分布（平均=1.0，標準偏差=1.2247/4）の確率密度は，$g(\tilde{\mu}) = 0.6515$ であり，それらの比は，$f(\boldsymbol{y})/g(\tilde{\mu}) = 0.0002$ と計算されている．

ステップ 2：尤度の比 $f(\boldsymbol{y})/g(\tilde{\mu})$ を最大にするように $\hat{\mu}$ と $\hat{\sigma}^2$ をソルバーで自動的に変化させると，$\hat{\mu} = 2.5$ と $\tilde{\sigma}^2 = 1.6667$ が得られる．REML 法による母分散の推定値 $\hat{\sigma}^2$ は，

$$\hat{\sigma}^2 = \frac{S_e}{n-1} = \frac{(1-2.5)^2 + (2-2.5)^2 + (3-2.5)^2 + (4-2.5)^2}{4-1} = \frac{5}{3} = 1.6667$$

と，偏差平方和 S_e をデータ数 $n-1$ で割った不偏分散に一致する．

REML法（制限付き最尤法）による平均と分散の推定

ステップ1 初期値

			y	尤度 L_i
データ数	$n=$	4		
固定効果	μ^=	1.0000	1	0.3257
分散	σ^2^=	1.5000	2	0.2334
			3	0.0859
f ：データの尤度	$f(y)=$	0.0001	4	0.0162
g ：平均値の尤度	$g(\mu\tilde{})=$	0.6515	$\mu\tilde{}=$ 1.0000	
			σ^2^/$n=$ 0.3750	平均値の分散
尤度の比 f/g	$f/g=$	0.0002		

ステップ2 尤度の比 f/g を最大化

			y	尤度 L_i
	$n=$	4		
	μ^=	2.5000	1	0.1573
	σ^2^=	1.6667	2	0.2867
			3	0.2867
	$f(y)=$	0.0020	4	0.1573
	$g(\mu\tilde{})=$	0.6180	$\mu\tilde{}=$ 2.5000	
			σ^2^/$n=$ 0.4167	平均値の分散
	$f/g=$	0.0033		

図 7.7　平均と分散の最尤法による同時推定

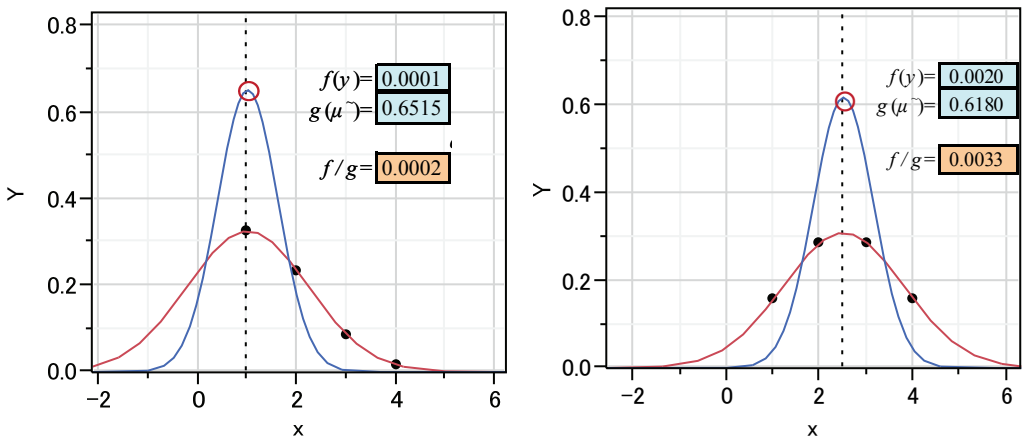
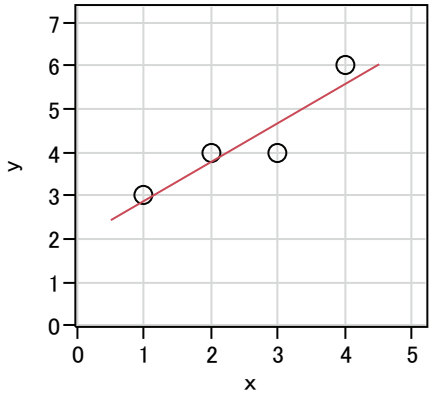


図 7.8　ステップ 1 およびステップ 2 でのデータと平均値の尤度

7.3. REML（制限付き最尤法）による回帰分析のパラメータ推定

標準的な最小2乗法による回帰分析

4 組のデータ（1，3），（2，4），（3，4），（4，6）に対して JMP の【二変量の関係】を用いた結果を図 7.9 に示す. 推定された回帰係数は $\hat{\beta}_0 = 2.0$，$\hat{\beta}_1 = 2.0$，誤差分散の推定値は $\hat{\sigma}_e^2 = 0.350$ である.



分散分析

要因	自由度	平方和	平均平方	F値
モデル	1	4.0500	4.0500	11.5714
誤差	2	0.7000	0.3500	p値(Prob>F)
全体(修正済み)	3	4.7500		0.0766

パラメータ推定値

項	推定値	標準誤差	t値	p値(Prob>|t|)
切片	2.0000	0.7246	2.7603	0.1100
x	0.9000	0.2646	3.4017	0.0766

図 7.9　標準的な回帰分析

最尤法による回帰分析

エラー! 参照元が見つかりません。 は Excel シート上で 4 組のデータ (y_i, x_i)，i =1, . . . , 4 に対して，おおよその切片 $\hat{\beta}_0 = 0.9$ と傾き $\hat{\beta}_1 = 1.0$ と初期値を設定し，誤差分散の初期値を $\tilde{\sigma}_e^2 = 0.5$ を与える．データ y_i の推定値は，$\hat{y}_i = \hat{\beta}_0 + \hat{\beta}_1 x_i$ として計算する．それぞれのデータ y_i の尤度 L_i は，平均が $\hat{y}_i$，分散が $\tilde{\sigma}_e^2 = 0.5$ の正規分布に従うとして，それらの積，尤度 L

$$L = \prod_{i=1}^{4} L_i = \prod_{i=1}^{4} \mathrm{N}(y_i; \hat{y}_i, \hat{\sigma}_e)$$

を計算する．1 番目のデータ（1，3）の推定値は，$\hat{y}_1 = 0.9 + 1 \times 1 = 1.9$ となり $y_1 = 3$ が，平均 $\hat{y}_1 = 1.9$，

標準偏差が $\sqrt{0.5}$ となる正規分布の確率密度として，$L_1 = 0.1682$ が計算されている．2 番目のデータ（2，4）の推定値は，$\hat{y}_2 = 0.9 + 1 \times 2 = 2.9$ となり，$y_2 = 4$ が，平均 $\hat{y}_2 = 2.9$，標準偏差が $\sqrt{0.5}$ となる正規分布の確率密度として，$L_2 = 0.1682$ が計算されている．順次 $L_3 = 0.5568$，$L_4 = 0.1682$ が計算され，尤度 L は，

$$f(\boldsymbol{y}) = L = \prod_{i=1}^{4} L_i = 0.1682 \times 0.1682 \times 0.5586 \times 0.1682 = 0.0027$$

となる．

最尤法	データの尤度 f を最大化						
			X				
データ数	$n=$	4	x_0	x_1	y	y^	尤度 L_i
固定効果	β_0^=	0.9000	1	1	3	1.90	0.1682
	β_1^=	1.0000	1	2	4	2.90	0.1682
分散	$\sigma_e^{2\sim}=$	0.5000	1	3	4	3.90	0.5586
			1	4	6	4.90	0.1682
f ：データの尤度	$f(\boldsymbol{y})=$	0.0027					

図 7.10　回帰分析での初期値に対する尤度の計算

図 7.11 の左に初期値に対する回帰直線を示す．散布図上の $x_1 = 1$ に対して回帰直線の推定値 $\hat{y}_1 = 1.9$ が×印で示されている．図の右は X 軸を y，Y 軸に正規分布の確率密度（尤度）とし，$\hat{y}_1 = 1.9$ を期待値，標準偏差を $\sqrt{5}$ とした正規分布が描かれ，その曲線上の ① が $y_1 = 3$ に対応した尤度 0.1682 である．

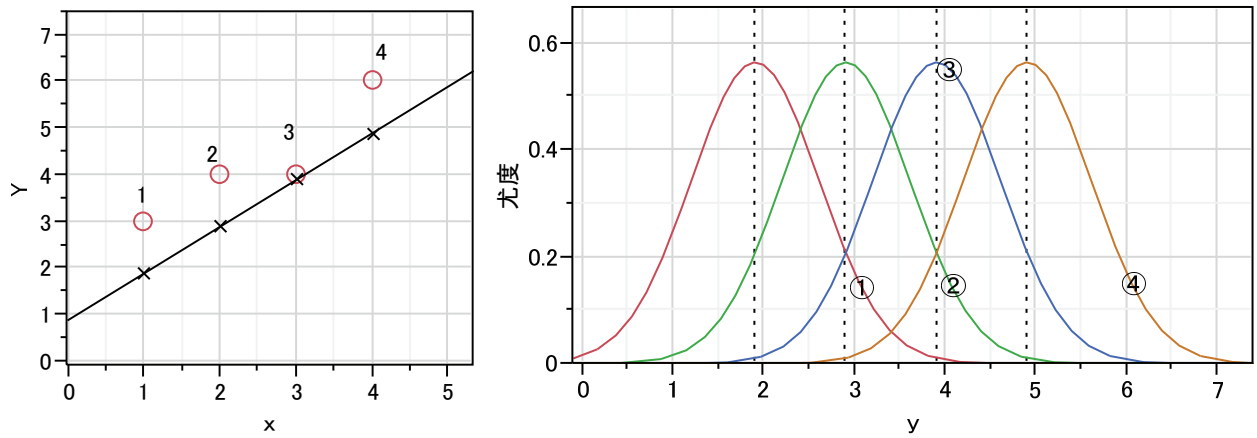


図 7.11　初期値に対するデータ y_i の $\hat{y}_i$ に対する尤度

Excel のソルバーで，$f(\boldsymbol{y})$ が最大となるように，（$\hat{\beta}_0$，$\hat{\beta}_1$，$\tilde{\sigma}_e^2$）を動かすと，図 7.12 に示すように，切片 $\hat{\beta}_0 = 2.0$ と傾き $\hat{\beta}_1 = 0.9$ が推定され，誤差分散の初は $\tilde{\sigma}_e^2 = 0.1119$ が推定される．1 番目の

データ（1，3）の推定値は，$\hat{y}_1 = 2 + 0.9 \times 2 = 2.9$ となり $y_1 = 3$ が，平均 $\hat{y}_1 = 2.9$，標準偏差が $\sqrt{0.1119}$ となる正規分布の確率密度として，$L_1 = 0.9268$ が計算されている．2 番目のデータ（2，4）の推定値は，$\hat{y}_2 = 2 + 0.9 \times 2 = 3.8$ となり，$y_2 = 4$ が，平均 $\hat{y}_2 = 3.8$，標準偏差が $\sqrt{0.1119}$ となる正規分布の確率密度として，$L_2 = 0.8507$ が計算されている．順次 $L_3 = 0.2352$，$L_4 = 0.6037$ が計算され，尤度 L は，

$$f(\boldsymbol{y}) = L = \prod_{i=1}^{4} L_i = 0.9268 \times 0.8507 \times 0.2352 \times 0.6037 = 0.1119$$

となる．パラメータの推定値 $\hat{\beta}_0 = 2.0$，$\hat{\beta}_1 = 0.9$ は，最小 2 乗法での解と一致するが，誤差分散は 0.1750 と 0.3500 に比べ小さくなっている．回帰分析の誤差の自由度は $n-2$ であるので，次式，

$$\hat{\sigma}_e^2 = \frac{n\tilde{\sigma}^2}{n-2} = \frac{4 \times 0.1750}{4-2} = 0.35$$

により，不偏分散の推定が行なえる．

最尤法	データの尤度 f を最大化							
				X				
データ数	$n=$	4		x_0	x_1	y	y^	尤度 L_i
固定効果	β_0^=	2.0000		1	1	3	2.90	0.9268
	β_1^=	0.9000		1	2	4	3.80	0.8507
分散	$\sigma_e^{2\sim}=$	0.1750		1	3	4	4.70	0.2352
				1	4	6	5.60	0.6037
f ：データの尤度	$f(\boldsymbol{y})=$	0.1119						

図 7.12　回帰分析での最尤解

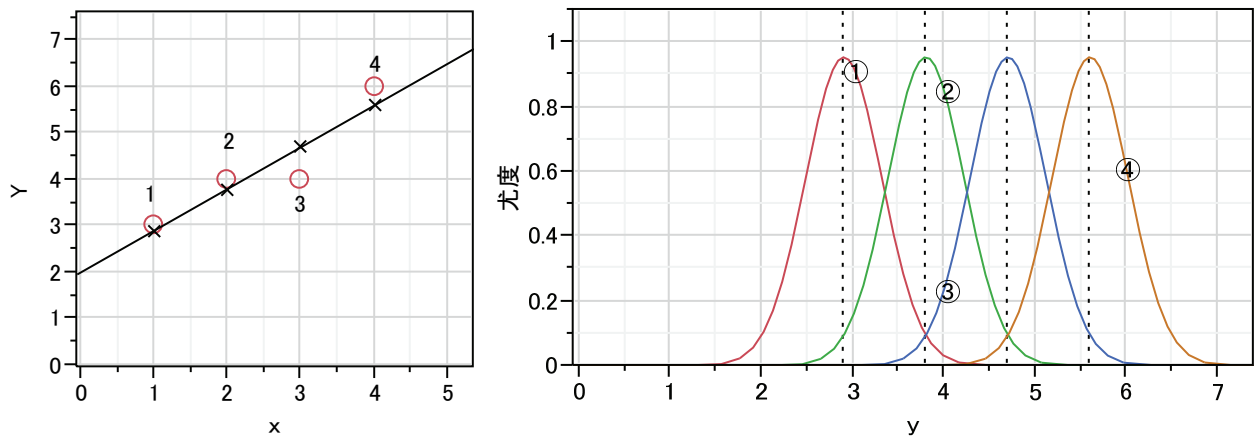


図 7.13　最尤解に対するデータ y_i の $\hat{y}_i$ に対する尤度

REML法による回帰分析

REML 法では，推定したいパラメータについての分散を求めて，パラメータの尤度を計算する必

要がある．回帰分析では，切片 β_0 と傾き β_1 の分散を計算したい．図 7.9 に示した標準的な回帰分析を Excel の行列関数を用いて計算した結果を 図 7.14 に示す．分散共分散行列 $\boldsymbol{\Sigma}$ の対角要素から切片 $\hat{\beta}_0 = 2.0$ の分散は 0.525，傾き $\hat{\beta}_1 = 0.9$ の分散は 0.007 となる．このデザイン行列 $\boldsymbol{X}$ を用いた計算方法を REML 法で用いることになる．

$\boldsymbol{X}$			$\hat{\beta}_j$			
x_0	x_1	y_i	$(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{y}$	y_i^	e_i	e_i^2
1	1	3	2	2.9	0.1	0.01
1	2	4	0.9	3.8	0.2	0.04
1	3	4		4.7	−0.7	0.49
1	4	6		5.6	0.4	0.16
					$Se=$	0.7
					$\sigma e^{2\wedge}=$	0.35

$\boldsymbol{X}'\boldsymbol{X}$		$(\boldsymbol{X}'\boldsymbol{X})^{-1}$		$\boldsymbol{\Sigma}=(\boldsymbol{X}'\boldsymbol{X})^{-1}\sigma_e^{2\wedge}$	
4	10	1.5	−0.5	0.525	−0.175
10	30	−0.5	0.2	−0.175	0.07

図 7.14　回帰分析のデザイン行列 $\boldsymbol{X}$ を用いたパラメータの分散共分散 $\boldsymbol{\Sigma}$ の計算

図 7.15 はデータに関する最尤法に加えて，回帰パラメータに関する尤度の計算を加えたものである．Excel シート上で 4 組のデータ (y_i, x_i)，i =1, . . . , 4 に対して，おおよその切片 $\hat{\beta}_0 = 0.9$ と傾き $\hat{\beta}_1 = 1.0$ と初期値を設定し，誤差分散の初期値を $\tilde{\sigma}_e^2 = 0.5$ を与える．データ y_i の推定値は，$\hat{y}_i = \hat{\beta}_0 + \hat{\beta}_1 x_i$ として計算する．尤度 L は，

$$f(\boldsymbol{y}) = L = \prod_{i=1}^{4} L_i = 0.1682 \times 0.1682 \times 0.5586 \times 0.1682 = 0.0027$$

と最尤法での結果と一致している．

切片 $\tilde{\beta}_0 = 0.9$ が，$\hat{\beta}_0 = 0.9$ を期待値，分散を $V(\hat{\beta}_0) = 0.750$ とする正規分布に従うとしたときの確率密度を尤度は 0.4607 と計算されている．傾き $\tilde{\beta}_1 = 1.0$ についても同様に $\hat{\beta}_1 = 1.0$ を期待値，分散を $V(\tilde{\beta}_1) = 1.0$ としたときの尤度 1.2616 を求め，それらの積が，

$$g(\tilde{\beta}) = 0.4607 \times 1.2616 = 0.5812$$

と計算されている．それらの尤度の比は，

$$\frac{f(\boldsymbol{y})}{g(\tilde{\boldsymbol{\beta}})} = \frac{0.0027}{0.5812} = 0.0046$$

となる．

REML　　データの尤度 f と回帰パラメータ尤度 g の比を最大化

			β_i~	β_i^分散	β_i^尤度
データ数	$N=$	4			
固定効果	β_0^=	0.9000	0.9000	0.7500	0.4607
	β_1^=	1.0000	1.0000	0.1000	1.2616
分散	σ_e^2^=	0.5000			
f :データの尤度	$f(y)=$	0.0027			
g :パラメータの尤度	$g(\mu$^$)=$	0.5812			
尤度の比 f/g	$f/g=$	0.0046			

X

x_0	x_1	y	y^	尤度 L_i
1	1	3	1.90	0.1682
1	2	4	2.90	0.1682
1	3	4	3.90	0.5586
1	4	6	4.90	0.1682

$\Sigma=(X'X)^{-1}\sigma_e^{2\wedge}$

0.7500	-0.2500
-0.2500	0.1000

図 7.15　REML 法による回帰分析（初期値）

ソルバーで「尤度の比 f/g が最大となるようにソルバーで $\hat{\beta}_0$，$\hat{\beta}_1$，および $\tilde{\sigma}^2$ を試行錯誤的に動かせ」と指示することにより 図 7.16 に示すように解が求められる．最尤法で推定された誤差分散は $\tilde{\sigma}_e^2=0.1750$ であり，REML 法で推定された誤差分散は $\hat{\sigma}_e^2=0.3500$ となる．これは，最小 2 乗法での不偏分散に一致する．

ただし，最小 2 乗法では，推定された回帰パラメータから推定値 $\hat{y}_i$ を求めデータ y_i との偏差平方を求め，その自由度で割った平均平方を誤差分散としているのに対し，REML 法では，回帰パラメータと誤差分散を同時推定している．

REML　　データの尤度 f と回帰パラメータ尤度 g の比を最大化

			β_i~	β_i^分散	β_i^尤度
データ数	$N=$	4			
固定効果	β_0^=	2.0000	2.0000	0.5250	0.5506
	β_1^=	0.9000	0.9000	0.0700	1.5079
分散	σ_e^2^=	0.3500			
f :データの尤度	$f(y)=$	0.0761			
g :パラメータの尤度	$g(\mu$^$)=$	0.8302			
尤度の比 f/g	$f/g=$	0.0916			

X

x_0	x_1	y	y^	尤度 L_i
1	1	3	2.90	0.6648
1	2	4	3.80	0.6369
1	3	4	4.70	0.3349
1	4	6	5.60	0.5365

$\Sigma=(X'X)^{-1}\sigma_e^{2\wedge}$

0.5250	-0.1750
-0.1750	0.0700

図 7.16　REML 法による回帰分析（尤度の比を最大化）

ここで示した回帰分析に対する REML 法での解析方法は，ダミー変数を用いた要因配置実験，重回帰分析などの一般線形モデルに対して適用可能である．誤差項が 1 つしかない場合一般線形モデルに対し実用上は，最小 2 乗法を適用すればよく，制限付き最尤法を適用する必要性はまったくない．しかしながら，REML 法の考え方を学習するためには，誤差項が 1 つしかない単純な問題から一歩一歩積み上げて理解を深めて行くことを薦める．

7.4. 枝分かれ型のデータに対する REML 法

枝分かれ型の実験データは，7.2 節で示した REML 法による母平均と 1 つの母分散の推定の問題から, 2 つの母分散への拡張となっていて, REML 法の本質を理解するための基本的な問題である.

分散分析表から分散成分の推定

最も基本的な 1 段階の枝分かれ実験データを用いよう．3 台の貨車 A から 2 個の硫化鉱の試料 B を分析した結果を表 7.1 示す.

表 7.1　分析データ

試料番号	貨車		
	A_1	A_2	A_3
1	42.0	41.4	41.1
2	41.8	41.5	40.8

繰返しが等しい場合の 1 元配置分散分析の分散分析の基本的な計算手順により分散分析表を作成する. 貨車間 A の平均平方の期待値は，それぞれの繰返し数は $m=2$ であるので， $\sigma_e^2+m\sigma_A^2=\sigma_e^2+2\sigma_A^2$ となる.

表 7.2　分散分析表

要因	自由度	平方和	平均平方	F 値	p 値	平均平方の期待値
貨車間 A	2	0.9033	0.4517	19.36	0.0193	$\sigma_e^2+2\sigma_A^2$
貨車内 e	3	0.0700	0.0233			σ_e^2
全体　T	5	0.9733				

貨車間のばらつき σ_A^2 および誤差分散 σ_e^2 の推定値は平均平方の期待値から，次のようにして求められる.

$$\hat{\sigma}_e^2=V_e=0.0233$$

$$\hat{\sigma}_A^2=\frac{0.4517-0.0233}{2}=0.2142$$

最尤法による分散成分の推定

枝分かれ実験データに対して貨車間の誤差を $\varepsilon_i^{(1)}$，貨車内の誤差を $\varepsilon_{ij}^{(2)}$ とした 2 つの誤差をもつデータ構造を考える. データ y_{ij} は, 総平均 μ に貨車間 A の貨車内 B の誤差を $\varepsilon_i^{(1)}$ と貨車内の誤差 $\varepsilon_{ij}^{(2)}$ を加えたものとして次式のように表すことができる.

（データ）＝ （総平均）＋ （貨車間 A の誤差） ＋ （貨車内 B の誤差）

$$y_{ij} = \mu + \varepsilon_i^{(1)} + \varepsilon_{ij}^{(2)}$$

この 2 つの誤差の分散 $\hat{\sigma}_A^2$ と $\hat{\sigma}_e^2$ を直接推定する方法として最尤法を用いる．最尤法は，得られたデータに対して正規分布の確率密度の積を最大化する方法であり，誤差が 1 つだけの場合には通常

の正規分布の確率密度関数 Normdist 関数を使うことにより簡単に求めることができた. しかしながら，複数の誤差がある場合には，表 7.3 に示すように貨車内 B のデータは互いに独立ではないので，それを考慮した多次元正規分布の確率密度の計算の必要がある.

表 7.3　枝分かれ実験の分散共分散構造

			分散共分散構造 $\boldsymbol{V}_i$	
A_i	r_{ij}	y_{ij}	$r_{ij}=1$	$r_{ij}=2$
A1	1	42.0	$\sigma_A^2+\sigma_e^2$	σ_A^2
	2	41.8	σ_A^2	$\sigma_A^2+\sigma_e^2$
A2	1	41.4	$\sigma_A^2+\sigma_e^2$	σ_A^2
	2	41.5	σ_A^2	$\sigma_A^2+\sigma_e^2$
A3	1	41.1	$\sigma_A^2+\sigma_e^2$	σ_A^2
	2	40.8	σ_A^2	$\sigma_A^2+\sigma_e^2$

枝分かれ実験データ y_{ij} についての分散共分散構造については, 同一貨車内の 2 つの繰返しデータに対して，2 次元正規分布の確率密度を計算する．データ（y_{i1}，y_{i2}）が，それぞれの共通の総平均 $\hat{\mu}$ に対し分散共分散行列 $\boldsymbol{V}_i$ が,

$$\boldsymbol{V}_i=\begin{bmatrix}\sigma_A^2+\sigma_e^2 & \sigma_A^2\\ \sigma_A^2 & \sigma_A^2+\sigma_e^2\end{bmatrix}$$

となる確率密度 L_i を計算し，$L=L_1\cdot L_2\cdot L_3$ を最大にするような（$\hat{\mu}$，$\tilde{\sigma}_A^2$，$\tilde{\sigma}_e^2$）を図 7.17 に示すように Excel のソルバーで求める.

手順 1：Excel シート上で，総平均 $\hat{\mu}$，貨車間分散 $\tilde{\sigma}_A^2$．貨車内の誤差分散 $\tilde{\sigma}_e^2$ のおおよその初期値を入れるセルを用意する.

手順 2：貨車 A_i ごとに 2×2 の分散共分散行列のセルを用意し，対角線上のセルには $\tilde{\sigma}_A^2$ と $\tilde{\sigma}_e^2$ の値を足す計算式を用意する．それ以外は $\tilde{\sigma}_A^2$ の値を代入する.

手順 3：尤度は貨車 A_i ごとに 2 次元正規分布の確率密度

$$f(y;\hat{\mu},\boldsymbol{V}_i)=\frac{1}{(2\pi)^{p/2}\left|\boldsymbol{V}_i\right|^{1/2}}\exp\left[-\frac{1}{2}(y-\hat{\mu})^T\boldsymbol{V}_i^{-1}(y-\hat{\mu})\right]$$

を $p=2$ として Excel の行列式 Mdeterm 関数，行列の転置 Transpose 関数，行列の積 Mmult 関数，逆行列 Minverse 関数を用いて計算式を定義する．貨車 A_1，貨車 A_2，および貨車 A_3 の尤度を計算し，それらの積を $L=L_1\cdot L_2\cdot L_3$ を Excel のシート上に定義する.

計算シートで 2 次元正規分布から求めた尤度が定義できれば, Excel のソルバーで,「尤度 L が最大になるように，総平均 $\hat{\mu}$，貨車間分散 $\tilde{\sigma}_A^2$．貨車内の誤差分散 $\tilde{\sigma}_e^2$ を変化させよ」とすると，それぞれの尤度が計算され，それらの尤度の積 L が

$$L=0.6001\times1.7043\times0.3333=0.3408$$

となり，最尤法による推定値 $\hat{\mu}=41.4333$，$\tilde{\sigma}_A^2=0.1389$，$\tilde{\sigma}_e^2=0.0233$ が得られる．この貨車間 A の分散の推定値 $\tilde{\sigma}_A^2=0.1389$ は，最尤解であり，表 7.2 の平均平方の期待値から

求めた貨車間の分散の推定値 $\hat{\sigma}_A^2 = 0.2142$ とは異なる．

	A	B	C	D	E	F	G	H
1			最尤法による平均値と分散成分の推定					
2								
3				μ^=	41.4333			
4				σ~$_A^2$=	0.1389			
5				σ~$_e^2$=	0.0233			
6							L =	0.3408
7								
8								
9		A_i	r_j	y_{ij}	V_i		y_{ij}-μ^	L_i
10		A1	1	42.0	0.1622	0.1389	0.5667	0.6001
11			2	41.8	0.1389	0.1622	0.3667	
12		A2	1	41.4	0.1622	0.1389	-0.0333	1.7043
13			2	41.5	0.1389	0.1622	0.0667	
14		A3	1	41.1	0.1622	0.1389	-0.3333	0.3333
15			2	40.8	0.1389	0.1622	-0.6333	
16								
17			多変量正規分布					
18								
19				$f(y,\hat{\mu},V) = \frac{1}{(2\pi)^{p/2}\lvert V\rvert^{1/2}}\exp\left[-\frac{1}{2}(y-\hat{\mu})^t V^{-1}(y-\hat{\mu})\right]$				
20								
21				p= 2				

図 7.17　繰返しが等しい場合の複数誤差の同時推定

尤度 L_1 の計算式　=(1/((2*pi())^(2/2)*Mdeterm(E10:F11)^(1/2))) * Exp(-(1/2) *
Mmult(Transpose(G10:G11), Mmult(Minverse(E10:F11),(G10:G11))))

REML法による分散成分の推定

枝分かれ実験データでの固定効果に関するパラメータは，総平均 μ のみである．最尤法では固定効果は，最小 2 乗法による推定値と一致することを前提としている．REML 法は，固定効果にもゆらぎがあると考え，固定効果の推定値に対する尤度も考える．尤度の計算には，総平均 $\hat{\mu}$ もゆらぐものとして $\tilde{\mu}$ とおき，その分散は，表 7.3 から，次のように求められる．

$$\tilde{\mu} = \frac{y_{11} + y_{12} + y_{21} + y_{22} + y_{31} + y_{32}}{6}$$

$$\begin{aligned} V(\tilde{\mu}) &= \frac{1}{36}\{V(y_{11}+y_{12}) + V(y_{21}+y_{22}) + V(y_{31}+y_{32})\} \\ &= \frac{1}{36}\sum_{i=1}^{3}\{(V(y_{i1}) + V(y_{i2}) + 2Cov(y_{i1}, y_{i2})\} \\ &= \frac{(2\hat{\sigma}_A^2 + \hat{\sigma}_e^2)}{6} \end{aligned}$$

総平均 $\hat{\mu}$ が分散 $V(\tilde{\mu})$ の正規分布に従うとしたときの正規確率密度をパラメータに関する尤度 $g(\tilde{\mu})$ とする．

$$g(\tilde{\mu}) = \frac{1}{\sqrt{2\pi\{(2\hat{\sigma}_A^2 + \hat{\sigma}_e^2)/6\}}}\exp\left[-\frac{(\tilde{\mu}-\hat{\mu})^2}{2\sqrt{(2\hat{\sigma}_A^2 + \hat{\sigma}_e^2)/6}}\right]$$

パラメータに関する尤度を $g(\tilde{\mu})$ とし，データに関する尤度を $f(y)$ としたときに，その比を

$$L(y,\tilde{\mu}) = \frac{f(y)}{g(\tilde{\mu})}$$

考える．$L(y,\tilde{\mu})$ が最大になるような総平均 $\hat{\mu}$，貨車間分散 $\hat{\sigma}_A^2$．貨車内の分散 $\hat{\sigma}_e^2$ を求める．これが REML 法といわれている方法である．図 7.18 に示した Excel シート上に $f(y)$，$g(\tilde{\mu})$ とそれらの比 $L(y,\tilde{\mu})$ の計算手順を定義し，ソルバーで尤度の比 $L(y,\tilde{\mu})$ を最大にするような $\hat{\mu}$，$\hat{\sigma}_A^2$，$\hat{\sigma}_e^2$ を求めさせる．

図 7.18 にはパラメータ $\tilde{\mu}$ についての尤度の計算シートを追加してある．最小 2 乗法で求めたパラメータ μ の推定値 $\hat{\mu}$ は，REML 法で求めた $\hat{\mu}$ に一致することから，計算シート上では，便宜的に $\hat{\mu} = \tilde{\mu}$ を代入している．正規分布の分散が $(2\hat{\sigma}_A^2 + \hat{\sigma}_e^2)/6$ であるので Normdist 関数の確率密度を求めている．

データ y に関する尤度 $f(y)$ とパラメータに関する尤度 $g(\tilde{\mu})$ の比 $L(y,\tilde{\mu})$ を計算シート上におき，ソルバーでこの尤度の比を最大にするような $\hat{\mu}$，$\hat{\sigma}_A^2$，$\hat{\sigma}_e^2$ を探索させる．この結果，$\hat{\sigma}_A^2 = 0.2141$ が得られ，表 7.2 の平均平方の分散成分から求めた $\hat{\sigma}_A^2 = 0.2141$ に一致する．

	B	C	D	E	F	G	H	I
1		REML法による平均値と分散成分の推定						
3				μ^=	41.4333			
4				σ^$_A^2$=	0.2142			
5				σ^$_e^2$=	0.0233			
7				データ y に関する尤度の計算シート				
8	A_i	r_j	X	y_{ij}	V_i		$y_{ij}-\mu$^	L_i
9	A1	1	1	42.0	0.2375	0.2142	0.5667	0.6236
10		2	1	41.8	0.2142	0.2375	0.3667	
11	A2	1	1	41.4	0.2375	0.2142	-0.0333	1.3920
12		2	1	41.5	0.2142	0.2375	0.0667	
13	A3	1	1	41.1	0.2375	0.2142	-0.3333	0.3524
14		2	1	40.8	0.2142	0.2375	-0.6333	
15					データの尤度		$f(y)$=	0.3059
16								
17				パラメータ μ に関する尤度の計算シート				
18			パラメータ	推定値	分散		μ~-μ^	L_μ
19			μ^	41.4333	0.0753		0.0000	1.4540
20					パラメータの尤度		$g(\mu$~$)$=	1.4540
22					制限付き尤度		$f(y)/g(\mu$~$)$=	0.2104

図 7.18　繰返しが等しい場合の複数誤差の REML 法による同時推定

μ の尤度：Normdist(F3,E19,Sqrt(F19),false)=1.454041

7.5. 繰返しを排除した REML 法

データ y についての尤度

繰返しが不揃いの場合には，その次元に合わせた多次元正規分布を定義する必要があり，さらに

パラメータに関する分散をその都度計算することも厄介である．これらの問題を解決する方法として，2×2 の分散共分散行列を大きさをデータ数とする $n\times n$ の行列の対角方向に配置し，n 次元正規分布を使うことによって，尤度をまとめて計算する方法が一般的に用いられている．これまでは，分散共分散行列を $\boldsymbol{V}$ で表記してきたが，慣習的に使われている $\boldsymbol{H}$ を使うことにする．

$$f(\boldsymbol{y};\hat{\mu},\boldsymbol{H})=\frac{1}{(2\pi)^{p/2}\left|\boldsymbol{H}\right|^{1/2}}\exp\left(-\frac{1}{2}(\boldsymbol{y}-\hat{\mu})^T\boldsymbol{H}^{-1}(\boldsymbol{y}-\hat{\mu})\right)$$

図 7.17 で繰返しが等しい場合の複数誤差の同時推定のための Excel シートを示したが，A_1，A_2，および A_3 についてそれぞれの 2×2 の分散共分散行列を図 7.19 に示すように対角上に配置する．この 6×6 の行列を 6 個のデータの分散共分散行列として，6 次元の正規分布の確率密度の計算を行なう．それを尤度として，ソルバーで最大化するように，総平均 $\hat{\mu}$，貨車間分散 $\tilde{\sigma}_A^2$．貨車内の誤差分散 $\tilde{\sigma}_e^2$ を変化させる．

図 7.17 で求めた尤度は $L=0.3408$ と図 7.19で求めた尤度は一致する．2 次元以上であっても Excel の計算方法は次元数 n の変更と引用セルの範囲を設定し直すのみであり，計算の手順が簡素化される．

	A	B	C	D	E	F	G	H	I	J	K
1			最尤法による平均値と分散成分の推定								
2											
3				$\mu=$	41.4333						
4				$\sigma^{\sim}{}_A^2=$	0.1389						
5				$\sigma^{\sim}{}_e^2=$	0.0233						
6											
7			データ y に関する尤度								
8		A_i	r_j	y_{ij}	Σ_i						$y_{ij}-\mu$
9		A1	1	42.0	0.1622	0.1389	0	0	0	0	0.5667
10			2	41.8	0.1389	0.1622	0	0	0	0	0.3667
11		A2	1	41.4	0	0	0.1622	0.1389	0	0	-0.0333
12			2	41.5	0	0	0.1389	0.1622	0	0	0.0667
13		A3	1	41.1	0	0	0	0	0.1622	0.1389	-0.3333
14			2	40.8	0	0	0	0	0.1389	0.1622	-0.6333
15										$f(y)=$	0.3408
16				多次元正規分布		$p=6$					
17				$f(\boldsymbol{y},\hat{\mu},\boldsymbol{H})=\frac{1}{(2\pi)^{p/2}\left\|\boldsymbol{H}\right\|^{1/2}}\exp\left(-\frac{1}{2}(\boldsymbol{y}-\hat{\mu})^T\boldsymbol{H}^{-1}(\boldsymbol{y}-\hat{\mu})\right)$						
18											

図 7.19　6 次元正規分布を用いたデータ y に関する尤度の計算

6 次元の確率密度：=(1/((2*pi())^(6/2)* Mdeterm(E9:J14)^(1/2))) * Exp(-(1/2) * Mmult(Transpose(K9:K14), Mmult(Minverse(E9:J14),(K9:K14))))

固定効果のパラメータについての尤度

枝分かれ実験の場合には，パラメータは総平均 μ のみであったので，その分散を別途求めて，1 次元の正規分布の確率密度を Normdist 関数からパラメータに関する尤度を計算した．図 7.20 に示すように，この手順も 6 次元の分散共分散行列を使うことによって，次の行列計算により求めることができる．

$$\hat{\sigma}_{\hat{\mu}}^2=(\boldsymbol{X}'\boldsymbol{H}^{-1}\boldsymbol{X})^{-1}$$

ここで $\boldsymbol{X}$ は総平均 $\hat{\mu}$ を求めるためのデザイン行列である．誤差分散が $\hat{\sigma}_e^2$ のみで，貨車間分散 $\hat{\sigma}_A^2$ なければ，通常の線形モデルとなり，パラメータの分散は

$$\hat{\sigma}_{\tilde{\mu}}^2 = (\boldsymbol{X}'\boldsymbol{X})^{-1}\hat{\sigma}_e^2 = \frac{\hat{\sigma}_e^2}{6}$$

となる．

パラメータが総平均 μ しかない場合は，1 次元の正規分布からパラメータに関する尤度の計算方法を図 7.18 で示した．しかしながら，複数のパラメータ $\boldsymbol{\beta}$ が存在するような場合には，データに関する尤度 $f(\boldsymbol{y};\hat{\mu},\boldsymbol{H})$ と同様に，パラメータについて p 多次元正規分布を用い尤度の計算が必要となる．データに関する分散共分散行列 $\boldsymbol{H}$ をパラメータに関する分散共分散行列 $(\boldsymbol{X}'\boldsymbol{H}^{-1}\boldsymbol{X})^{-1}$ で置き換え，p 次元正規分布を用いて尤度の計算を行う．

$$p\text{ 次元：}\ g(\tilde{\boldsymbol{\beta}};\hat{\boldsymbol{\beta}},(\boldsymbol{X}'\boldsymbol{H}^{-1}\boldsymbol{X})^{-1}) = \frac{1}{(2\pi)^{p/2}\left|(\boldsymbol{X}'\boldsymbol{H}^{-1}\boldsymbol{X})^{-1}\right|^{1/2}}\exp\left[-\frac{1}{2}(\tilde{\boldsymbol{\beta}}-\hat{\boldsymbol{\beta}})^T(\boldsymbol{X}'\boldsymbol{H}^{-1}\boldsymbol{X})(\tilde{\boldsymbol{\beta}}-\hat{\boldsymbol{\beta}})\right]$$

データに関する尤度 $f(\boldsymbol{y};\hat{\mu},\boldsymbol{H})$ とパラメータに関する尤度 $g(\tilde{\boldsymbol{\beta}};\hat{\boldsymbol{\beta}},(\boldsymbol{X}'\boldsymbol{H}^{-1}\boldsymbol{X})^{-1})$ の比の尤度を最大にする方法が REML 法である．

$$L(y,\tilde{\boldsymbol{\beta}}) = \frac{f(\boldsymbol{y};\hat{\boldsymbol{\beta}},\boldsymbol{H})}{g(\tilde{\boldsymbol{\beta}};\hat{\boldsymbol{\beta}},(\boldsymbol{X}'\boldsymbol{H}^{-1}\boldsymbol{X})^{-1})}$$

	A	B	C	D	E	F	G	H	I	J	K
1			パラメータに関する分散共分散行列								
2											
3					μ^=	41.4333					
4					σ^$_A^2$=	0.2142					
5					σ^$_e^2$=	0.0233					
6											
7			データ y に関する尤度								
8		A_i	r_j	デザイン行列 X	y_{ij}	H					
9		A1	1	1	42.0	0.2375	0.2142	0	0	0	0
10			2	1	41.8	0.2142	0.2375	0	0	0	0
11		A2	1	1	41.4	0	0	0.2375	0.2142	0	0
12			2	1	41.5	0	0	0.2142	0.2375	0	0
13		A3	1	1	41.1	0	0	0	0	0.2375	0.2142
14			2	1	40.8	0	0	0	0	0.2142	0.2375
15											
16					パラメータに関する分散共分散行列の計算						
17											
18					$V(\mu) = (X^T H^{-1} X)^{-1}$	0.0753					

図 7.20　パラメータに関する分散共分散行列の計算

=Minverse(Mmult(Transpose(D9:D14),Mmult(Minverse(F9:K14),D9:D14)))

7.6. 繰返しが不揃いの場合の REML 法

繰返しが不揃いの場合でも，REML 法の計算の考え方は同じである．貨車 A_3 の繰返しが 1 回し

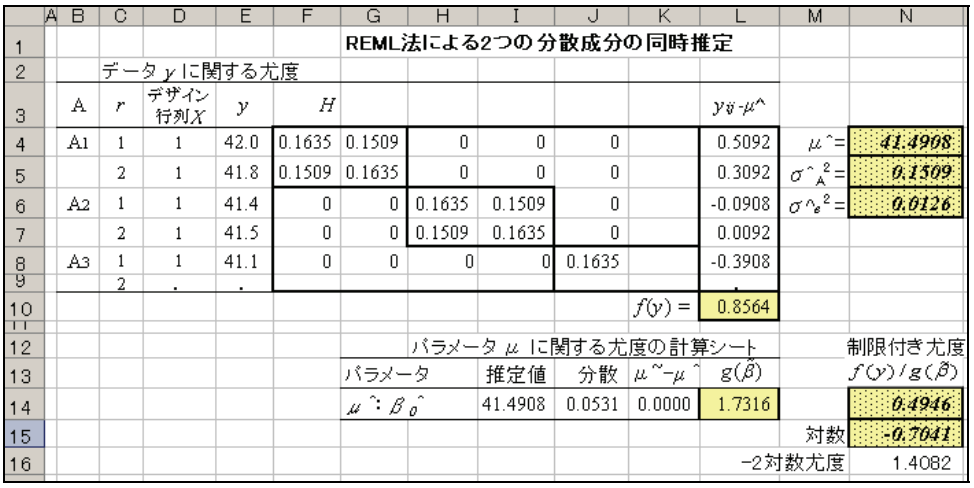
かないとすると，その分散は$\sigma_A^2+\sigma_e^2$だけとなる．

繰返し$r_i=2$の場合を除いた 5 次元正規分布の確率密度を計算するように計算式を変更する．パラメータμの分散もp次元正規分布を用いた尤度の計算に変更する．これらを図 7.21 に示す Excel の表にまとめ，Excel のソルバーで，データに関する尤度とパラメータに関する尤度の比を最大にするような $(\hat{\mu},\hat{\sigma}_A^2,\hat{\sigma}_e^2)$ を探索させる．

表 7.4　繰返しが不揃いの場合の分散共分散分行列

			A_1		A_2		A_3
A_i	r_j	y_{ij}	$r_j=1$	$r_i=2$	$r_j=1$	$r_i=2$	$r_j=1$
A_1	1	42.0	$\sigma_A^2+\sigma_e^2$	σ_A^2			
	2	41.8	σ_A^2	$\sigma_A^2+\sigma_e^2$			
A_2	1	41.4			$\sigma_A^2+\sigma_e^2$	σ_A^2	
	2	41.5			σ_A^2	$\sigma_A^2+\sigma_e^2$	
A_3	1	41.1					$\sigma_A^2+\sigma_e^2$

その結果，総平均は$\mu=41.4908$，貨車間の分散$\hat{\sigma}_A^2=0.1509$，誤差分散$\hat{\sigma}_e^2=0.0126$が推定される．負の 2 倍の対数尤度を計算すると 1.4082 となる．



	A	B	C	D	E	F	G	H	I	J	K	L	M	N
1							REML法による2つの分散成分の同時推定							
2			データ y に関する尤度											
3		A	r	デザイン行列X	y	H						yij-μ^		
4		A1	1	1	42.0	0.1635	0.1509	0	0	0		0.5092	μ^=	41.4908
5			2	1	41.8	0.1509	0.1635	0	0	0		0.3092	σ^A2=	0.1509
6		A2	1	1	41.4	0	0	0.1635	0.1509	0		-0.0908	σ^e2=	0.0126
7			2	1	41.5	0	0	0.1509	0.1635	0		0.0092		
8		A3	1	1	41.1	0	0	0	0	0.1635		-0.3908		
9			2	.	.							.		
10											f(y) =	0.8564		
11														
12							パラメータ μ に関する尤度の計算シート							制限付き尤度
13							パラメータ		推定値	分散	μ~-μ^	g(β^)		f(y)/g(β^)
14							μ^: β0^		41.4908	0.0531	0.0000	1.7316		0.4946
15													対数	-0.7041
16													−2対数尤度	1.4082

図 7.21　繰返しが不揃いの場合の REML 法による分散成分の計算

7.7. 分散共分散行列の生成

表 7.4 で示した分散共分散行列$\boldsymbol{H}$は，初期値として設定した$\hat{\sigma}_A^2$と$\hat{\sigma}_e^2$から Excel のセルごとの計算式を用いた．これを行列計算で求める方法を示す．まず，貨車の 3 水準について右のような 5×3 のインジケータ型デザイン行列$\boldsymbol{Z}$を準備する．

Z

A1	A2	A3
1	0	0
1	0	0
0	1	0
0	1	0
0	0	1

貨車の 3 水準について分散 $\hat{\sigma}_A^2$ の初期値を，右に示すような 3×3 の対角要素に展開した行列 $\boldsymbol{G}$ を準備する．

G

0.1509	0	0
0	0.1509	0
0	0	0.1509

デザイン行列 $\boldsymbol{Z}$ と貨車間分散行列 $\boldsymbol{G}$ について $\boldsymbol{ZGZ'}$ を計算すると，表 7.4 で示した貨車の水準に対応した 5×5 のブロック型の対角行列が求められる．

ZGZ'

σ_A^2	σ_A^2	0	0	0
σ_A^2	σ_A^2	0	0	0
0	0	σ_A^2	σ_A^2	0
0	0	σ_A^2	σ_A^2	0
0	0	0	0	σ_A^2

ZGZ'

0.1509	0.1509	0	0	0
0.1509	0.1509	0	0	0
0	0	0.1509	0.1509	0
0	0	0.1509	0.1509	0
0	0	0	0	0.1509

行列 $\boldsymbol{ZGZ'}$ の対角要素に誤差分散の初期値 $\hat{\sigma}_e^2$ を加えるために，5×5 の誤差分散行列 $\boldsymbol{R}$ を右のように作成する．

R

0.0126	0	0	0	0
0	0.0126	0	0	0
0	0	0.0126	0	0
0	0	0	0.0126	0
0	0	0	0	0.0126

表 7.4 で示した分散共分散行列 $\boldsymbol{H}$ は右のように $\boldsymbol{ZGZ'}$ に $\boldsymbol{R}$ を加えて計算される．

ZGZ'+R

0.1635	0.1509	0	0	0
0.1509	0.1635	0	0	0
0	0	0.1635	0.1509	0
0	0	0.1509	0.1635	0
0	0	0	0	0.1635

この程度の大きさの分散共分散行列 $\boldsymbol{H}$ であれば，σ_A^2 と σ_e^2 の初期値に対して，セルごとに計算式を逐次的にセットすることが簡便である．しかしながら，変量効果が複数あるような場合には，次式，

$$\boldsymbol{H} = \boldsymbol{ZGZ'} + \boldsymbol{R}$$

による行列計算を行うことが望ましい．この例で誤差分散行列 $\boldsymbol{R}$ は誤差分散が均一の対角行列としたが，REML 法では誤差構造を自由に与えることができる．複雑な誤差構造を仮定したモデルについても，行列 $\boldsymbol{ZGZ'}$ と行列 $\boldsymbol{R}$ を組み合わせて自由に設定できる．

8. 枝分れ実験における分散共分散構造

8.1. 複数の誤差構造

枝分かれ実験, 乱塊法実験, 分割実験など本質的に複数の誤差構造を持つ実験データ解析に対し, 制限付き最尤法は，最小 2 乗法に代えて標準的に用いるべき解析方法である．制限付き最尤法は, 欠測値がある場合，実験の一部を意図的に省略する場合，探索的なデータ解析などのデータ解析でも汎用的に使える.

誤差が 1 つの場合の最尤法は，最小 2 乗法の「データ y_i の偏差平方和を最小にする」考え方を, 「データ y_i の尤度の積を最大化する」考え方に置き換えてソルバーで解くことが出来た．複数の誤差がある場合には, データ y_i の尤度だけでなくデータ $y_{i'}$ との共分散も考慮しつつ尤度を計算する必要がある．このためにデータ間の共分散を考慮した尤度のを求めるために多変量正規分布の密度関数を用いる必要がある.

多変量正規分布の密度関数の計算では，行列計算が不可欠である．Excel の行列計算の関数を使うことにより，多変量正規分布の密度関数から尤度を計算する．これによりデータの尤度の積 L とパラメータの尤度の積の G を定義し，その比 L/G について Excel のソルバーで「回帰パラメータと複数の誤差成分を試行錯誤的に動かして L/G を最大せよ」指示することにより, パラメータと誤差成分を同時推定する.

8.2. 枝分かれ実験データ

枝分かれ実験は，複数の分散成分を推定する実験計画の典型例である．硫化鉱がたくさんの貨車に積まれて入荷した．全体について平均硫黄含有率を推定したい．貨車 6 台をランダムに選び，1 貨車あたり 5 個の試料をランダムに抜き取る．各試料ごとに縮分して分析した結果を 表 8.1 に示す.

表 8.1　硫化鉱の分析データ

試料番号	A1	A2	A3	A4	A5	A6
1	42.0	41.4	41.1	40.5	41.2	40.5
2	41.8	41.5	40.8	40.4	40.9	40.3
3	40.8	41.1	40.2	39.9	40.7	41.0
4	41.4	41.6	41.5	39.7	41.3	39.9
5	41.0	41.9	41.4	41.0	41.0	40.4

このデータ y_{ij} は次の構造を持っている.

（データ）＝ （総平均）＋ （貨車間のばらつき） ＋ （誤差）

$$y_{ij} = \mu + \alpha_i + \varepsilon_{ij}$$

このデータ構造は，貨車を母数因子型とした場合と同じであるが，α_i の意味が異なる．貨車はランダムにサンプリングしたので母数模型ではなく変量因子とする変量模型である．変量模型では，α_i の母平均 0，分散 σ_A^2 の母集団（ここでは貨車のあつまり）の中からランダムに抜き取られた偏差であって，α_i の値そのものを推定するのが目的ではなく，α_i のばらつきを通して母集団の分散 σ_A^2 を推定するのが目的である［1］.

分散 σ_A^2 を推定するために，1 元配置分散分析を行ない，その平均平方の期待値の構造から，分散 σ_A^2 を容易に推定することができる．ただし，試料の数が不揃いの場合には，その推定は容易ではない.

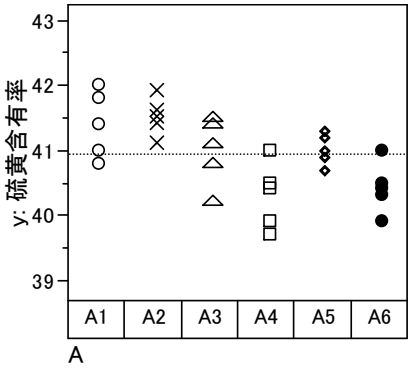


図 8.1　貨車ごとの硫黄含有率の変動

母数模型では α_i は，総平均 μ から貨車ごとの水準平均を差し引いて求めるのであるが，変量模型に対する最尤法，および制限つき最尤法では，α_i が母平均 0，分散 σ_A^2 の正規分布に従い，ε_{ij} が母平均 0，分散 σ_e^2 の正規分布に従うと仮定して，データ y_{ij} の尤度を最大化する．したがって，個々の α_i の推定は行わない.

8.3. データ間の相関構造

枝分かれ実験のデータ y_{ij} 相互には独立とみなせない状況が発生する．貨車が異なる y_{ij} と $y_{i'j}$ とのデータは互いに独立，したがって共分散 $Cov(y_{ij}, y_{i'j})$ は 0 とみなせる．しかしながら，貨車内の試料間 y_{ij} と $y_{ij'}$ は, 互いに独立とはいえない. 同じ貨車 i に限定すれば y_{ij} と $y_{ij'}$ は互いに独立とみなせる.

確率的には α_i が母平均 0，分散 σ_A^2 の正規分布に従って変動する．したがって，$\mu+\alpha_i$ が確率的に変動する上での $\mu+\alpha_i+\varepsilon_{ij}$ と $\mu+\alpha_i+\varepsilon_{ij'}$ の関係は，X 軸と Y 軸が共に $\mu+\alpha_i$ であるような直線上の点に対して，X 軸の $\mu+\alpha_i$ に ε_{ij} を加えたものと，Y 軸の $\mu+\alpha_i$ に $\varepsilon_{ij'}$ を加えたものを想定する．分散 σ_A^2 が分散 σ_e^2 に比べて大きいときには 45 度に傾いた細長い楕円になり，分散 σ_A^2 が相対的に小さい

場合には太目の楕円となる．

実際のデータで確かめてみよう．貨車$i=1$に対して，異なるjの組み合わせは，（1，2），（1，3），（1，4），（1，5），（2，3），（2，4），（2，5），（3，4），（3，5），（4，5）の 10 通りとなり，貨車$i=2$に対しても 10 通りあり，6 貨車全体で 60 通りのデータの組み合わせが得られる．

表 8.2　異なる試料間のデータの組み合わせ表

ペア番号	貨車 i	試料 j	$y{:}ij$	試料 j'	$y{:}ij'$
1	A1	1	42.0	2	41.8
2	A1		42.0	3	40.8
3	A1		42.0	4	41.4
4	A1		42.0	5	41.0
5	A1	2	41.8	3	40.8
6	A1		41.8	4	41.4
7	A1		41.8	5	41.0
8	A1	3	40.8	4	41.4
9	A1		40.8	5	41.0
10	A1	4	41.4	5	41.0
:					
60	A6	4	39.9	5	40.4

図 8.1 に，この 60 通りの組み合わせデータに対する確率楕円を上書きした散布図を示す．相関係数は 0.48 となるので，σ_A^2 と σ_e^2 がほぼ同じであることが推測される．

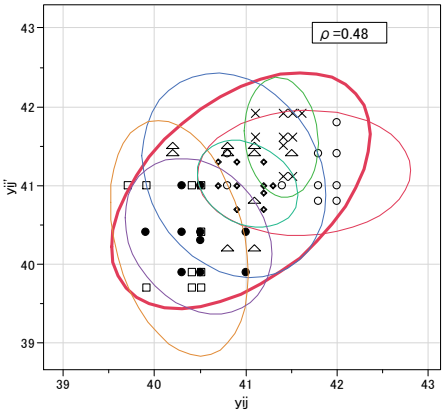


図 8.2　異なる試料間の散布図

全データを用いた場合に相関係数は 0.48 となるが，

個々の貨車iごとには相関関係は見られない．

この楕円の共分散は $Cov(y_{ij}, y_{ij'}) = \sigma_A^2$ となることが，α_i と ε_{ij} の共分散が 0 であることを用いて求められる．

$$
\begin{aligned}
Cov(y_{ij}, y_{ij'}) &= E((y_{ij} - \mu)(y_{ij'} - \mu)) \\
&= E((\alpha_i + \varepsilon_{ij})(\alpha_i + \varepsilon_{ij'})) \\
&= E(\alpha_i^2 + \alpha_i \varepsilon_{ij} + \alpha_i \varepsilon_{ij'} + \varepsilon_{ij} \varepsilon_{ij'}) \\
&= E(\alpha_i^2) \\
&= V(\alpha_i) \\
&= \sigma_A^2
\end{aligned}
$$

したがって，同じ貨車内の 5 個の試料間の分散共分散行列 Σ_i は，次のように 5×5 の対称構造の行列として表せる．貨車 i の試料が不揃いの場合には，4×4 あるいは 3×3 分散共分散行列 $\mathbf{\Sigma}_i$ を用いる．試料の繰返しが 1 の場合には，分散共分散行列 $\mathbf{\Sigma}_i$ が縮小して 1×1 の $[\sigma_A^2 + \sigma_e^2]$ のみとなる．

$$
\mathbf{\Sigma}_i = \begin{bmatrix}
\sigma_A^2 + \sigma_e^2 & \sigma_A^2 & \sigma_A^2 & \sigma_A^2 & \sigma_A^2 \\
\sigma_A^2 & \sigma_A^2 + \sigma_e^2 & \sigma_A^2 & \sigma_A^2 & \sigma_A^2 \\
\sigma_A^2 & \sigma_A^2 & \sigma_A^2 + \sigma_e^2 & \sigma_A^2 & \sigma_A^2 \\
\sigma_A^2 & \sigma_A^2 & \sigma_A^2 & \sigma_A^2 + \sigma_e^2 & \sigma_A^2 \\
\sigma_A^2 & \sigma_A^2 & \sigma_A^2 & \sigma_A^2 & \sigma_A^2 + \sigma_e^2
\end{bmatrix}
$$

8.4. 簡便な方法による相関構造の確認

前節で示した相関構造を確認するための簡便な方法を示す．表 8.2 のデータは貨車 i ごとに異なる試料 j の番号のペア 10 組を作り，6 貨車分で 60 組のデータを作成した．これを，ペア番号（1，11，21，31，41，51），（2，12，22，32，42，52）,…，（10，20，30，40，50，60）のように順番を変えると，異なる試料の組み合わせごとに貨車が（A1，A2, . . . ，A6）がセットになる．

表 8.3　試料間のデータの組み合わせ表

ペア番号	貨車 *i*	試料 *j*	*y:ij*	試料 *j'*	*y:ij '*
1	A1	1	42.0	2	41.8
11	A2	1	41.4	2	41.5
21	A3	1	41.1	2	40.8
31	A4	1	40.5	2	40.4
41	A5	1	41.2	2	40.9
51	A6	1	40.5	2	40.3
2	A1	1	42.0	3	40.8
12	A2	1	41.4	3	41.1
22	A3	1	41.1	3	40.2
32	A4	1	40.5	3	39.9
42	A5	1	41.2	3	40.7
52	A6	1	40.5	3	41.0
:			:		:
10	A1	4	41.4	5	41.0
20	A2	4	41.6	5	41.9
30	A3	4	41.5	5	41.4
40	A4	4	39.7	5	41.0
50	A5	4	41.3	5	41.0
60	A6	4	39.9	5	40.4

見方を変えると，貨車を行方向，試料を列方向にまとめ直して列方向の試料 5 変数として，全ての組みあわせについて散布図を作ることになる．

表 8.4　貨車を行方向に転置した表

	試料番号				
A	1	2	3	4	5
A1	42.0	41.8	40.8	41.4	41.0
A2	41.4	41.5	41.1	41.6	41.9
A3	41.1	40.8	40.2	41.5	41.4
A4	40.5	40.4	39.9	39.7	41.0
A5	41.2	40.9	40.7	41.3	41.0
A6	40.5	40.3	41.0	39.9	40.4

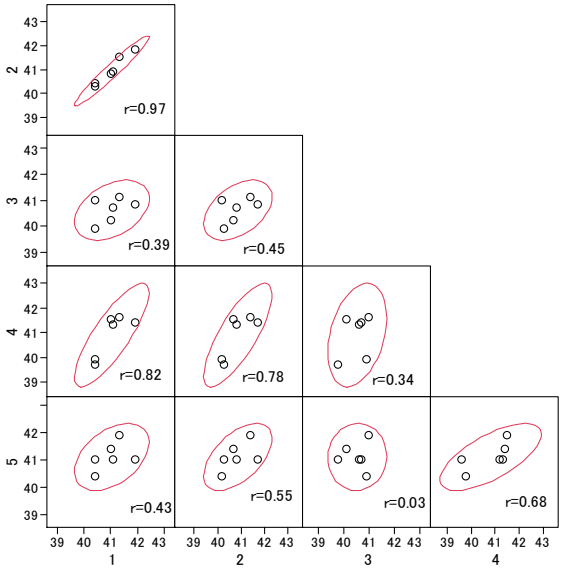


図 8.3　試料間の散布図および確率楕円

8.5. 多次元正規分布の確率密度

正規分布の確率密度関数は，μ を平均，σ^2 を分散としたときに次式で与えられる．

$$f(x;\mu,\sigma)=\frac{1}{\sqrt{2\pi\sigma^2}}\exp\left(-\frac{1}{2\sigma^2}(x-\mu)^2\right)$$

これを p 次元のデータ $\boldsymbol{y}_i=[y_{i1},y_{i2},\ldots,y_{ip}]^T$ に拡張した p 次元正規分布は，μ を平均，$\boldsymbol{\Sigma}_i$ を分散共分散行列としたときに，次式で与えられる．

$$f(\boldsymbol{y}_i;\mu,\boldsymbol{\Sigma}_i)=\frac{1}{(2\pi)^{p/2}\left|\boldsymbol{\Sigma}_i\right|^{1/2}}\exp\left(-\frac{1}{2}(\boldsymbol{y}_i-\mu)^T\boldsymbol{\Sigma}_i^{-1}(\boldsymbol{y}_i-\mu)\right)$$

硫化鉱の枝分かれデータで i 番目の貨車の試料は繰返しが 5 なので，$\boldsymbol{\Sigma}_i$ は 5×5 の行列となり，$\boldsymbol{y}_i=[y_{i1},y_{i2},\ldots,y_{i5}]^T$ について 5 次元の確率密度を尤度 L_i が求められる．全ての貨車について尤度を求め，それらの積を尤度 L とする．

$$L=L_1\times L_2\times L_3\times L_4\times L_5\times L_6$$

貨車ごとに尤度 L_i を求めてその積を計算する代わりに，全てのデータについて一括して尤度 L を計算する方法が，統計ソフトの計算では用いられている．これは，i 番目の分散共分散行列 $\boldsymbol{\Sigma}_i$ を一つの行列に 6 貨車分を次のようにまとめて，30×30 の行列 $\boldsymbol{\Sigma}$ とする．

$$\boldsymbol{\Sigma} = \begin{bmatrix} \boldsymbol{\Sigma}_1 & 0 & 0 & 0 & 0 & 0 \\ 0 & \boldsymbol{\Sigma}_2 & 0 & 0 & 0 & 0 \\ 0 & 0 & \boldsymbol{\Sigma}_3 & 0 & 0 & 0 \\ 0 & 0 & 0 & \boldsymbol{\Sigma}_4 & 0 & 0 \\ 0 & 0 & 0 & 0 & \boldsymbol{\Sigma}_5 & 0 \\ 0 & 0 & 0 & 0 & 0 & \boldsymbol{\Sigma}_6 \end{bmatrix}$$

データを $\boldsymbol{y} = [(y_{11},...,y_{15}),(y_{21},...,y_{15}),...,(y_{61},...,y_{65}),]^T$ のように大きさ 30 のベクトルとしてまとめて，次のように 30 次元多変量正規分布の確率密度から

$$f(\boldsymbol{y};\mu,\boldsymbol{\Sigma}) = \frac{1}{(2\pi)^{30/2}\left|\boldsymbol{\Sigma}\right|^{1/2}} \exp\left(-\frac{1}{2}(\boldsymbol{y}-\mu)^T \boldsymbol{\Sigma}_i^{-1}(\boldsymbol{y}-\mu)\right)$$

尤度 L が求められる.

9. Excel によるスタッカード型枝分かれ実験データの解析

表 6.2 でスタッカード型の枝分かれデータの例示をした．Exel で解析するためにデータの形式を 表 9.1 に示すように整理した

表 9.1　スタッカード型の枝分かれデータ

No.	A：ロット	B：試料	C：分析	y_{ijk}
1	A_1	B_1	C_1	94
2	A_1	B_1	C_2	96
3	A_1	B_2	C_1	90
4	A_2	B_1	C_1	107
5	A_2	B_1	C_2	108
6	A_2	B_2	C_1	117
7	A_3	B_1	C_1	86
8	A_3	B_1	C_2	87
9	A_3	B_2	C_1	85

このデータに対する JMP「モデルのあてはめ」で EMS 法による解析結果は，表 9.2 のようにまとめられる．

表 9.2　スタッカード型の枝分かれデータの分散分析表

要因	自由度	平方和	平均平方	F 値		p 値	平均平方の期待値
A： e_A	2	1011.44	505.72	-	19.37	0.0193	$\sigma_e^2+1.33\,\sigma_{B(A)}^2+2.67\,\sigma_A^2$
B(A)：$e_{B(A)}$	3	78.33	26.11	26.11	1.00	0.0119	$\sigma_e^2+1.33\,\sigma_{B(A)}^2$
誤差 e	3	3.00	1.00	1.00	-		σ_e^2
全体	8	1044.00					

EMS 法での分散成分は，$\sigma_A^2=179.85$，$\sigma_{B(A)}^2=18.83$，$\sigma_e^2=1.00$ となる．REML 法では，

変量効果	分散比	分散成分	標準誤差	95%下側	95%上側	全体に対する百分率
A	177.96	178.71	188.59	46.24	10118.28	90.0
B[A]	18.67	18.75	15.94	5.82	310.35	9.4
残差		1.00	0.82	0.32	14.21	0.5
合計		198.46				100.0
-2対数尤度= 47.362362396						

に示すように，$\sigma_A^2=178.71$，$\sigma_{B(A)}^2=18.75$，$\sigma_e^2=1.00$ と，EMS 法の場合と幾分小さくなっている．

スタッカード型の枝分かれ実験の場合に REML 法で分散成分の推定を行なうためには，データ相互の相関構造を定義する必要がある．2 つの変量因子に対して，分散共分散構造は次のような考えで求めることができる．

1) データ数をnとしたときに，対角要素が全て$\hat{\sigma}_e^2$の$n \times n$行列$\boldsymbol{R}$を用意する．
2) $n \times n$行の行列$\boldsymbol{H}$を用意し，同じロットA内の同じ水準内のデータ相互間のセルに$\hat{\sigma}_A^2$を行列$\boldsymbol{H}$に加える．
3) 試料 B(A)内の同じ水準内のデータ相互のセルに$\sigma_{B(A)}^2$を$\boldsymbol{H}$に加える．ただし，B(A)は，ロットAと試料Bのインジケータ型デザイン行列の積とする．

実際の計算は，7.7 節で示した行列計算と同じ方法で，それぞれの分散共分散行列に対する加算分を生成することができる．図 9.1 に$\hat{\sigma}_A^2 = 4$，$\hat{\sigma}_{B(A)}^2 = 2$，$\hat{\sigma}_e^2 = 1$とした場合の分散共分散行列の作成過程を示す．

ロット A について9×3のインジケータ型のデザイン行列$\boldsymbol{Z}_A$，ロットの 3 水準に対応する3×3の分散成分行列$\boldsymbol{G}_A$からデータに関する9×9の分散共分散行列$\boldsymbol{H}_A$を計算する．

$$\boldsymbol{H}_A = \boldsymbol{Z}_A \boldsymbol{G}_A \boldsymbol{Z}'_A$$

これと同様に，ロットAから枝分かれになった調整試料B(A)についての9×6のインジケータ型のデザイン行列$\boldsymbol{Z}_{B(A)}$と6×6の分散成分行列から9×9の$\boldsymbol{H}_{B(A)}$を計算する．

$$\boldsymbol{H}_{B(A)} = \boldsymbol{Z}_{B(A)} \boldsymbol{G}_{B(A)} \boldsymbol{Z}'_{B(A)}$$

分析 C は繰返しがないので，9×9の角要素が全てσ_e^2の$n \times n$の誤差行列$\boldsymbol{R}$とする．最後にこれらをすべて加算し，分散共分散行列$\boldsymbol{H}$を計算する．

$$\boldsymbol{H} = \boldsymbol{H}_A + \boldsymbol{H}_{B(A)} + \boldsymbol{R}$$

説明のために$\boldsymbol{Z}_A$と$\boldsymbol{Z}_{B(A)}$を分けた計算法を示したが，$\boldsymbol{Z} = \boldsymbol{Z}_A + \boldsymbol{Z}_{B(A)}$のデザイン行列，これに対応する対角要素に分散成分を並べた9×9の行列$\boldsymbol{G}$とし，分散共分散行列Vを

$$\boldsymbol{H} = \boldsymbol{Z}\boldsymbol{G}\boldsymbol{Z}' + \boldsymbol{R}$$

として計算するのが一般的である．

$\hat{\sigma}^2_A =$ 4
$\hat{\sigma}^2_{B(A)} =$ 2
$\hat{\sigma}^2_e =$ 1

$H = H_A + H_{B(A)} + R$

	1	2	3	4	5	6	7	8	9
1	7	6	4	0	0	0	0	0	0
2	6	7	4	0	0	0	0	0	0
3	4	4	7	0	0	0	0	0	0
4	0	0	0	7	6	4	0	0	0
5	0	0	0	6	7	4	0	0	0
6	0	0	0	4	4	7	0	0	0
7	0	0	0	0	0	0	7	6	4
8	0	0	0	0	0	0	6	7	4
9	0	0	0	0	0	0	4	4	7

Z_A

	A1	A2	A3
1	1	0	0
2	1	0	0
3	1	0	0
4	0	1	0
5	0	1	0
6	0	1	0
7	0	0	1
8	0	0	1
9	0	0	1
	1	2	2

G_A $\hat{\sigma}^2_A$

	1	2	3
1	4	0	0
2	0	4	0
3	0	0	4

$H_A = Z_A G_A Z_A'$

	1	2	3	4	5	6	7	8	9
1	4	4	4	0	0	0	0	0	0
2	4	4	4	0	0	0	0	0	0
3	4	4	4	0	0	0	0	0	0
4	0	0	0	4	4	4	0	0	0
5	0	0	0	4	4	4	0	0	0
6	0	0	0	4	4	4	0	0	0
7	0	0	0	0	0	0	4	4	4
8	0	0	0	0	0	0	4	4	4
9	0	0	0	0	0	0	4	4	4

$Z_{B(A)}$

	A1		A2		A3	
	B1	B2	B1	B2	B1	B2
1	1	0	0	0	0	0
2	1	0	0	0	0	0
3	0	1	0	0	0	0
4	0	0	1	0	0	0
5	0	0	1	0	0	0
6	0	0	0	1	0	0
7	0	0	0	0	1	0
8	0	0	0	0	1	0
9	0	0	0	0	0	1
	1	2	3	4	5	6

$G_{B(A)}$ $\hat{\sigma}^2_{B(A)}$

	1	2	3	4	5	6
1	2	0	0	0	0	0
2	0	2	0	0	0	0
3	0	0	2	0	0	0
4	0	0	0	2	0	0
5	0	0	0	0	2	0
6	0	0	0	0	0	2

$H_{A(B)}: Z_{B(A)} G_{B(A)} Z_{B(A)}'$

	1	2	3	4	5	6	7	8	9
1	2	2	0	0	0	0	0	0	0
2	2	2	0	0	0	0	0	0	0
3	0	0	2	0	0	0	0	0	0
4	0	0	0	2	2	0	0	0	0
5	0	0	0	2	2	0	0	0	0
6	0	0	0	0	0	2	0	0	0
7	0	0	0	0	0	0	2	2	0
8	0	0	0	0	0	0	2	2	0
9	0	0	0	0	0	0	0	0	2

$\hat{\sigma}^2_e$

1

R

	1	2	3	4	5	6	7	8	9
1	1	0	0	0	0	0	0	0	0
2	0	1	0	0	0	0	0	0	0
3	0	0	1	0	0	0	0	0	0
4	0	0	0	1	0	0	0	0	0
5	0	0	0	0	1	0	0	0	0
6	0	0	0	0	0	1	0	0	0
7	0	0	0	0	0	0	1	0	0
8	0	0	0	0	0	0	0	1	0
9	0	0	0	0	0	0	0	0	1

図 9.1　行列計算による分散共分散行列の生成

このように一般化した分散共分散行列の構成法を用いた Excel による REML 法による分散成分の推定結果を 図 9.2 に示す．推定された分散成分は $\hat{\sigma}_A^2 = 178.7482$， $\hat{\sigma}_{B(A)}^2 = 18.7483$， $\hat{\sigma}_e^2 = 1.0042$ である．

$\hat{\mu}$ =	96.6667
$\hat{\sigma}_A^2$ =	178.7482
$\hat{\sigma}_{B(A)}^2$ =	18.7483
$\hat{\sigma}_e^2$ =	1.0042

パラメータ μ に関する尤度の計算シート

パラメータ	推定値	分散	$\tilde{\mu}-\hat{\mu}$	$\log g(\hat{\beta})$
$\tilde{\mu}:\tilde{\beta}_0$	96.7	62.8	0.0	-2.99

制限付き尤	
$\log f(y)/g(\hat{\beta})$	-23.68
−2 log(L)	47.36

	X μ	Z A1	A2	A3	A1B1	A1B2	A2B1	A2B2	A3B1	A3B2	y_{ijk}	H=ZGZ'+R									$y_{ijk}-\hat{\mu}$
1	1	1	0	0	1	0	0	0	0	0	94	198.5	197.5	178.7	0.0	0.0	0.0	0.0	0.0	0.0	−2.7
2	1	1	0	0	1	0	0	0	0	0	96	197.5	198.5	178.7	0.0	0.0	0.0	0.0	0.0	0.0	−0.7
3	1	1	0	0	0	1	0	0	0	0	90	178.7	178.7	198.5	0.0	0.0	0.0	0.0	0.0	0.0	−6.7
4	1	0	1	0	0	0	1	0	0	0	107	0.0	0.0	0.0	198.5	197.5	178.7	0.0	0.0	0.0	10.3
5	1	0	1	0	0	0	1	0	0	0	108	0.0	0.0	0.0	197.5	198.5	178.7	0.0	0.0	0.0	11.3
6	1	0	1	0	0	0	0	1	0	0	117	0.0	0.0	0.0	178.7	178.7	198.5	0.0	0.0	0.0	20.3
7	1	0	0	1	0	0	0	0	1	0	86	0.0	0.0	0.0	0.0	0.0	0.0	198.5	197.5	178.7	−10.7
8	1	0	0	1	0	0	0	0	1	0	87	0.0	0.0	0.0	0.0	0.0	0.0	197.5	198.5	178.7	−9.7
9	1	0	0	1	0	0	0	0	0	1	85	0.0	0.0	0.0	0.0	0.0	0.0	178.7	178.7	198.5	−11.7
																				$\log f(y)$	-26.67

G

1	178.7	0	0	0	0	0	0	0	0
2	0	178.7	0	0	0	0	0	0	0
3	0	0	178.7	0	0	0	0	0	0
4	0	0	0	18.748	0	0	0	0	0
5	0	0	0	0	18.748	0	0	0	0
6	0	0	0	0	0	18.748	0	0	0
7	0	0	0	0	0	0	18.748	0	0
8	0	0	0	0	0	0	0	18.748	0
9	0	0	0	0	0	0	0	0	18.748

R

1.0	0	0	0	0	0	0	0	0
0	1.0	0	0	0	0	0	0	0
0	0	1.0	0	0	0	0	0	0
0	0	0	1.0	0	0	0	0	0
0	0	0	0	1.0	0	0	0	0
0	0	0	0	0	1.0	0	0	0
0	0	0	0	0	0	1.0	0	0
0	0	0	0	0	0	0	1.0	0
0	0	0	0	0	0	0	0	1.0

図 9.2　REML 法によるスタッカード型の 2 段枝分かれ分散成分の推定

9.1. 変量効果の推定 BLUP

変量効果に関する推定は，BLUP 法で推定されている．JMP のでの結果は，次のようである．

変量効果の予測

項	BLUP	標準誤差	分母自由度	t値	p値(Prob>|t|)
A[A1]	−4.0725	8.1076	2.163	−0.50	0.6620
A[A2]	14.5673	8.1076	2.163	1.80	0.2046
A[A3]	−10.4947	8.1076	2.163	−1.29	0.3164
A[A1]:B[B1]	2.1871	3.0688	3.198	0.71	0.5245
A[A1]:B[B2]	−2.6143	3.0714	3.198	−0.85	0.4537
A[A2]:B[B1]	−3.7926	3.0688	3.198	−1.24	0.2995
A[A2]:B[B2]	5.3208	3.0714	3.198	1.73	0.1759
A[A3]:B[B1]	0.1634	3.0688	3.198	0.05	0.9607
A[A3]:B[B2]	−1.2644	3.0714	3.198	−0.41	0.7066

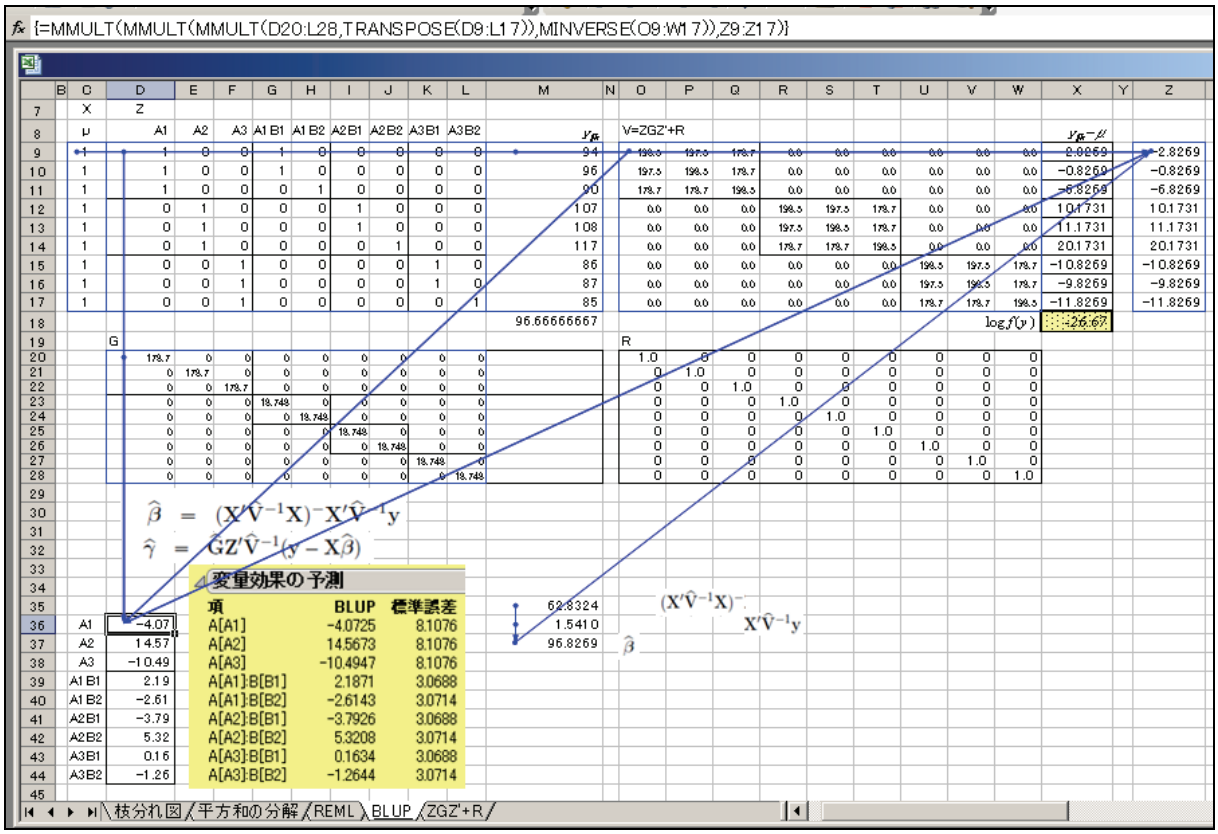
固定効果は，

$$\hat{\boldsymbol{\beta}} = (\boldsymbol{X}'\hat{\boldsymbol{V}}^{-1}\boldsymbol{X})^{-1}\boldsymbol{X}'\hat{\boldsymbol{V}}^{-1}\boldsymbol{y}$$

としたときに，変量効果については，次式で与えられる．

$$\hat{\gamma} = \hat{G}Z'\hat{V}^{-1}(y - X\hat{\beta})$$

実際に．Excel で計算した結果を次に示す．JMP での結果と一致する．



なお，計算式は，次に示す SAS/STAT 9.1 Users Guide，p2739 の式を用いた．

Estimating β and γ in the Mixed Model

ML, REML, MIVQUE0, or Type1–Type3 provide estimates of $\mathbf{G}$ and $\mathbf{R}$, which are denoted $\widehat{\mathbf{G}}$ and $\widehat{\mathbf{R}}$, respectively. To obtain estimates of $\boldsymbol{\beta}$ and $\boldsymbol{\gamma}$, the standard method is to solve the *mixed model equations* (Henderson 1984):

$$\begin{bmatrix} \mathbf{X}'\widehat{\mathbf{R}}^{-1}\mathbf{X} & \mathbf{X}'\widehat{\mathbf{R}}^{-1}\mathbf{Z} \\ \mathbf{Z}'\widehat{\mathbf{R}}^{-1}\mathbf{X} & \mathbf{Z}'\widehat{\mathbf{R}}^{-1}\mathbf{Z} + \widehat{\mathbf{G}}^{-1} \end{bmatrix} \begin{bmatrix} \widehat{\boldsymbol{\beta}} \\ \widehat{\boldsymbol{\gamma}} \end{bmatrix} = \begin{bmatrix} \mathbf{X}'\widehat{\mathbf{R}}^{-1}\mathbf{y} \\ \mathbf{Z}'\widehat{\mathbf{R}}^{-1}\mathbf{y} \end{bmatrix}$$

The solutions can also be written as

$$\begin{aligned} \widehat{\boldsymbol{\beta}} &= (\mathbf{X}'\widehat{\mathbf{V}}^{-1}\mathbf{X})^{-}\mathbf{X}'\widehat{\mathbf{V}}^{-1}\mathbf{y} \\ \widehat{\boldsymbol{\gamma}} &= \widehat{\mathbf{G}}\mathbf{Z}'\widehat{\mathbf{V}}^{-1}(\mathbf{y} - \mathbf{X}\widehat{\boldsymbol{\beta}}) \end{aligned}$$

and have connections with empirical Bayes estimators (Laird and Ware 1982, Carlin and Louis 1996).

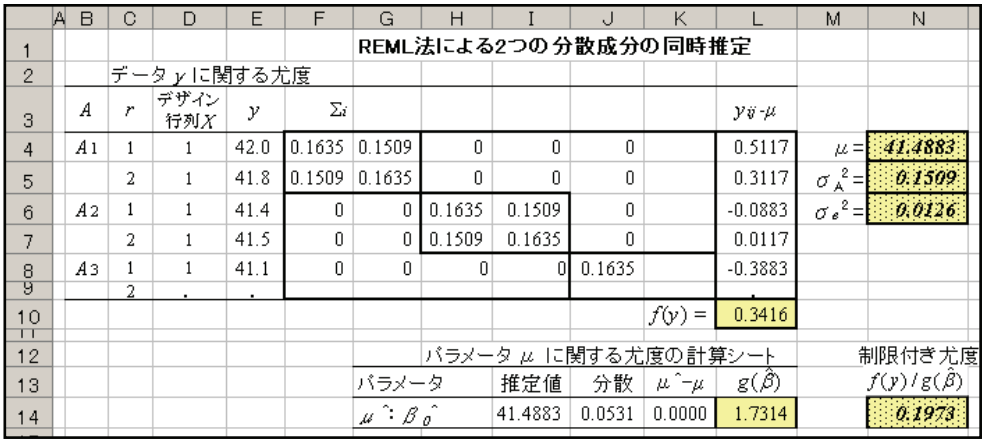
9.2. 統計連合大会 2010 年での発表 抄録

Excel ソルバーによる複数の誤差成分の制限付き最尤法(REML)法による同時推定

要約　複数の誤差構造を持つ枝分かれ実験，乱塊法実験，経時データなどの解析を行うために，制限付き最尤法(REML)法が開発され，多くの統計ソフトで利用できるようになった．しかしながら，その計算過程はブラックボックス的であり，簡単なデータに対してさえ自ら計算することが困難であった．Excel のソルバーは，複数のパラメータの値を変化させながら尤度を最大化するための機能を持っており，これを用いて REML 法の計算が可能と思われた．REML 法は，最尤法をベースにしており，最尤法によるパラメータの推定を Excel で行うことを前提とする．誤差成分が複数の場合には，データ間の分散共分散行列を考慮した多次元正規分布の確率密度の計算も必要となる．これらも Excel の行列関数で計算可能であり，Excel の1枚のシート上で REML 法による計算プロセスおよび解析結果について「見える化」を実現した．

解析の考え方　Diggle ら(2002)は，REML の解析法の原理として，データについての尤度を $f(y)$，固定効果(パラメータ)についての尤度を $g(\hat{\beta})$ としたときに，その比 $f(y)/g(\hat{\beta})$ を最大化することによって，固定効果と分散成分を同時推定する方法を示している．芳賀(2007)の Excel のソルバーを用いた最尤法の考え方を参考に，尤度 $f(y)/g(\hat{\beta})$ を最大化する方法により，複数誤差の同時推定を試みた．

見える化



REML法による2つの分散成分の同時推定

データ y に関する尤度

	A	r	デザイン行列X	y	Σi					yij-μ
A1	1	1	42.0	0.1635	0.1509	0	0	0		0.5117
	2	1	41.8	0.1509	0.1635	0	0	0		0.3117
A2	1	1	41.4	0	0	0.1635	0.1509	0		-0.0883
	2	1	41.5	0	0	0.1509	0.1635	0		0.0117
A3	1	1	41.1	0	0	0	0	0.1635		-0.3883
	2	.	.							.
									$f(y)$ =	0.3416

μ = 41.4883
σ_A^2 = 0.1509
σ_e^2 = 0.0126

パラメータ μ に関する尤度の計算シート

パラメータ	推定値	分散	$\hat{\mu}-\mu$	$g(\hat{\beta})$
$\hat{\mu}$: $\hat{\beta}_0$	41.4883	0.0531	0.0000	1.7314

制限付き尤度 $f(y)/g(\hat{\beta})$: 0.1973

枝分かれ実験で繰返しが不揃いのデータを用いて2つの分散成分の同時推定を試みた．1)総平均 μ，ロット A 間の分散成分 σ_A^2，誤差分散 σ_e^2 に適当な初期値をセット，2) データ y に関する分散共分散行列を定義，3)尤度 $f(y)$ を 5 次元正規分布の確率密度として Excel の行列関数で定義，4)尤度 $g(\hat{\beta})$ は，パラメータ $\mu:\beta_0$ のデザイン行列 X と y の分散共分散行列から定義，5)比 $f(y)/g(\hat{\beta})$ を定義．6)ソルバーで比 $f(y)/g(\hat{\beta})$ を最大にするように，μ，σ_A^2，σ_e^2 を逐次的に変化させる．7)初期値が $\hat{\mu}=41.4883$，$\hat{\sigma}_A^2=0.1509$，$\hat{\sigma}_e^2=0.0126$ となり，一枚の Excel シート上で解を求めることができた．

文献：　1) Diggle P J., et al. (2002), Analysis of Longitudinal Data 2ed., Oxford Press.
2) 芳賀敏郎(2007)，Excel-ソルバーによる非線形最小 2 乗法・最尤法・ロバスト回帰，
http://www.med.kurume-u.ac.jp/med/biostat/forum/haga.pdf

9.3. 統計連合大会 2010 年での発表 スライド

Excelソルバーによる
複数の誤差成分の制限付き最尤法
（REML）法による同時推定

高橋 行雄（中外製薬）

2010.9.7 高橋行雄 1

Excelソルバーによる
複数の誤差成分の制限付き最尤法
（REML）法による同時推定

高橋 行雄（中外製薬）

2010.9.7 高橋行雄 1

線形混合モデル

- 枝分かれ実験は複数の分散成分を持つ
- 実験の繰返しが一定ならば，平均平方の期待値から，分散成分の推定は容易
- JIS規格 Z8404-2
  - 測定の不確かさ－第2 部:
  - 測定の不確かさの評価における繰返し測定及び枝分かれ実験の利用の指針

2010.9.7 高橋行雄 3

ばらつきを評価する実験

- 繰返しが一定の枝分かれ実験

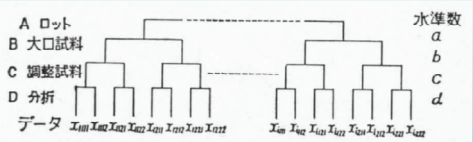


図 7.34 繰返し数の等しい 4 段階の枝分かれ配置

$a = 10$, $b = 2$, $c = 2$, $d = 2$

（新版 品質管理便覧，規格協会，p496）

2010.9.7 高橋行雄 4

自由度がアンバランス

- 繰返し一定の枝分かれ実験では，推定したい誤差成分が小さいほうに自由度がインフレーション

表 7.143 分散分析表（バランス型枝分かれ配置）

要　因	ν	E[V]
A	9	$\sigma_D^2+2\sigma_{C(AB)}^2+4\sigma_{B(A)}^2+8\sigma_A^2$
B(A)	10	$\sigma_D^2+2\sigma_{C(AB)}^2+4\sigma_{B(A)}^2$
C(AB)	20	$\sigma_D^2+2\sigma_{C(AB)}^2$
D(ABC)	40	σ_D^2
総	79	

2010.9.7 高橋行雄 5

じぐざぐ（staggered）型計画

- 分散成分の自由度の均一化

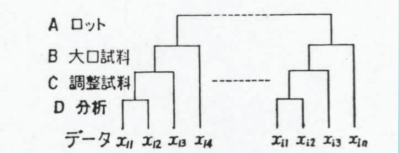


図 7.35 Staggered 型枝分かれ配置（最初と最後の単位を示してある）

$a = 20$, $b = 2$, $c = 2$, $d = 2$

2010.9.7 高橋行雄 6

均一な自由度

- ロットを20水準と倍増，最終の分析数は80回と同じ，B(A)の自由度も10から20と倍増

要　因	ν	E[V]
A	19	$\sigma_D^2+\frac{3}{2}\sigma_C^2+\frac{5}{2}\sigma_B^2+4\sigma_A^2$
B(A)	20	$\sigma_D^2+\frac{7}{6}\sigma_C^2+\frac{3}{2}\sigma_B^2$
C(B)	20	$\sigma_D^2+\frac{4}{3}\sigma_C^2$
D(C)	20	σ_D^2
計	79	

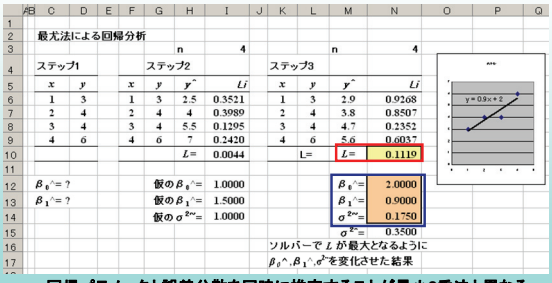
線形混合モデルの適用

- 繰返しが一定の枝分かれ実験は，多くの分野で実施されている．
- じぐざぐ(staggered)型の枝分かれ実験の適用事例を見出すことができない．
- 解説書は，新版 品質管理便覧以外，他には見出せない．
- 線形混合モデルの適用で，じぐざぐ(staggered)型計画でも手軽に解析が行なえる．
- 統計ソフトを使うことをいさぎよし潔しとしない民族的感情が，統計ソフトの活用を避けているようだ．

Excelのソルバーで最尤法

- Excelのソルバーは，複数のパラメータの値を変化させながら尤度を最大化するための機能を持っている．
- ソルバーを用いて REML法の計算が可能と思われた．REML法は，最尤法をベースにしており，最尤法によるパラメータの推定をExcelで行うことが前提となる．

最尤法で回帰分析



最尤法による回帰分析

ステップ1		ステップ2				ステップ3			
				n	4			n	4
x	y	x	y	$y^\wedge$	Li	x	y	$y^\wedge$	Li
1	3	1	3	2.5	0.3521	1	3	2.9	0.9268
2	4	2	4	4	0.3989	2	4	3.8	0.8507
3	4	3	4	5.5	0.1295	3	4	4.7	0.2352
4	6	4	6	7	0.2420	4	6	5.6	0.6037
				L=	0.0044		L=	L=	0.1119

β_0^=?　β_1^=?

仮のβ_0^= 1.0000
仮のβ_1^= 1.5000
仮の$\sigma^{2\sim}$= 1.0000

β_0^= 2.0000
β_1^= 0.9000
$\sigma^{2\sim}$= 0.1750
$\sigma^{2\sim}$= 0.3500

ソルバーで L が最大となるように
β_0,β_1,σ^2を変化させた結果

回帰パラメータと誤差分散を同時に推定することが最小2乗法と異なる

枝分かれ実験データの例示

表 4.1　分析データ

試料番号	貨車 A1	A2	A3
1	42.0	41.4	41.1
2	41.8	41.5	40.8

完備型ならば容易に分散成分の計算ができる

表 4.2　分散分析表

要因	自由度	平方和	平均平方	F 値	p 値	平均平方の期待値
貨車間 A	2	0.9033	0.4517	19.36	0.0193	$\sigma_e^2+2\sigma_A^2$
貨車内 e	3	0.0700	0.0233			σ_e^2
全体　T	5	0.9733				

$$\hat{\sigma}_e^2=V_e=0.0233$$

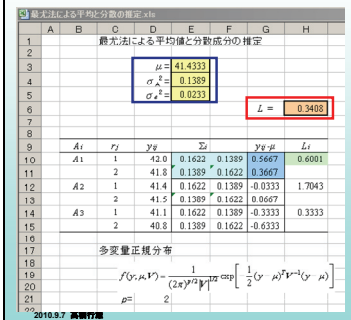
$$\hat{\sigma}_A^2=\frac{0.4517-0.0233}{2}=0.2142$$

分散共分散構造

表 4.3　分散共分散構造

A_i	r_{ij}	y_{ij}	分散共分散構造 V_i $r_{ij}=1$	$r_{ij}=2$
$A1$	1	42.0	$\sigma_A^2+\sigma_e^2$	σ_A^2
	2	41.8	σ_A^2	$\sigma_A^2+\sigma_e^2$
$A2$	1	41.4	$\sigma_A^2+\sigma_e^2$	σ_A^2
	2	41.5	σ_A^2	$\sigma_A^2+\sigma_e^2$
$A3$	1	41.1	$\sigma_A^2+\sigma_e^2$	σ_A^2
	2	40.8	σ_A^2	$\sigma_A^2+\sigma_e^2$

最尤法の適用



最尤法による平均値と分散成分の推定

μ =	41.4333
σ_A^2 =	0.1389
σ_e^2 =	0.0233
L =	0.3408

A_i	r_j	y_{ij}	Σ_i		$y_{ij}-\mu$	L_i
$A1$	1	42.0	0.1622	0.1389	0.5667	0.6001
	2	41.8	0.1389	0.1622	0.3667	
$A2$	1	41.4	0.1622	0.1389	-0.0333	1.7043
	2	41.5	0.1389	0.1622	0.0667	
$A3$	1	41.1	0.1622	0.1389	-0.3333	0.3333
	2	40.8	0.1389	0.1622	-0.6333	

多変量正規分布

$$f(y,\mu,V)=\frac{1}{(2\pi)^{p/2}|V|^{1/2}}\exp\left[\frac{1}{2}(y-\mu)^T V^{-1}(y-\mu)\right]$$

p= 2

最尤法による平均値と分散成分の推定.xls

貨車ごとに2次元正規分布の確率密度の積から尤度を計算

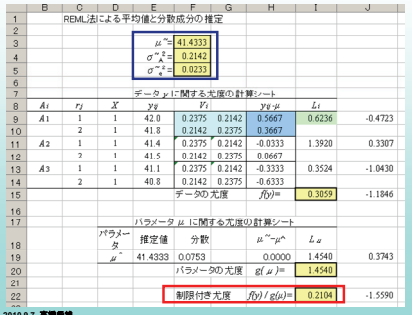
=(1/((2*pi())^(2/2)
*Mdeterm(E10:F11)^(1/2))) *
Exp(-(1/2) *
Mmult(Transpose(G10:G11),
Mmult(Minverse(E10:F11),(G10:
G11))))

制限付き最尤法（REML法）

- データについての尤度 $f(y)$ ➤最尤法
- 固定効果（パラメータ）についての尤度 $g(\hat{\beta})$
- 制限付き最尤法

比 $f(y)/g(\hat{\beta})$ を最大化

尤度の比をソルバーで最大化



REML法による平均値と分散成分の推定

$\mu^{\sim}$=	41.4333
$\sigma_A^{\sim 2}$=	0.2142
$\sigma_e^{\sim 2}$=	0.0233

データ y に関する尤度の計算シート

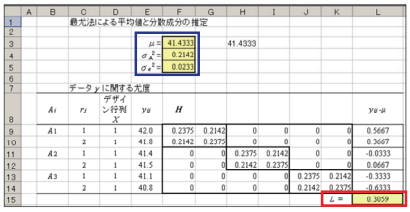
A_i	r_j	X	y_{ij}	V_i		$y_{ij}-\mu$	L_i	
$A1$	1	1	42.0	0.2375	0.2142	0.5667	0.6236	-0.4723
	2	1	41.8	0.2142	0.2375	0.3667		
$A2$	1	1	41.4	0.2375	0.2142	-0.0333	1.3920	0.3307
	2	1	41.5	0.2142	0.2375	0.0667		
$A3$	1	1	41.1	0.2375	0.2142	-0.3333	0.3524	-1.0430
	2	1	40.8	0.2142	0.2375	-0.6333		
				データの尤度		$f(y)$=	0.3059	-1.1846

パラメータ μ に関する尤度の計算シート

パラメータ	推定値	分散		$\mu^{\sim}-\mu^{\wedge}$	L_μ	
$\mu^{\wedge}$	41.4333	0.0753		0.0000	1.4540	0.3743
		パラメータの尤度		$g(\mu)$=	1.4540	
		制限付き尤度		$f(y)/g(\mu)$=	0.2104	-1.5590

REML法により分散成分を直接推定することができる.

繰返し演算を排除した行列計算



最尤法による平均値と分散成分の推定

μ =	41.4333	41.4333
σ_A^2 =	0.2142	
σ_e^2 =	0.0233	

データ y に関する尤度

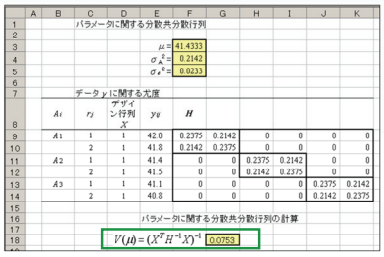
A_i	r_j	デザイン行列 X	y_{ij}	H						$y_{ij}-\mu$
$A1$	1	1	42.0	0.2375	0.2142	0	0	0	0	0.5667
	2	1	41.8	0.2142	0.2375	0	0	0	0	0.3667
$A2$	1	1	41.4	0	0	0.2375	0.2142	0	0	-0.0333
	2	1	41.5	0	0	0.2142	0.2375	0	0	0.0667
$A3$	1	1	41.1	0	0	0	0	0.2375	0.2142	-0.3333
	2	1	40.8	0	0	0	0	0.2142	0.2375	-0.6333
									L =	0.3059

図 5.1 6 次元正規分布を用いたデータ y に関する尤度の計算

6次元の確率密度：=(1/((2*Pi())^(6/2)*Mdeterm(E9:J14)^(1/2)))＊Exp(-(1/2)＊
Mmult(Transpose(K9:K14), Mmult(Minverse(E9:J14),(K9:K14))))

6次元の正規分布の確率密度を最尤法で直接計算する

パラメータに関する分散の計算



パラメータに関する分散共分散行列

μ =	41.4333
σ_A^2 =	0.2142
σ_e^2 =	0.0233

データ y に関する尤度

A_i	r_j	デザイン行列 X	y_{ij}	H					
$A1$	1	1	42.0	0.2375	0.2142	0	0	0	0
	2	1	41.8	0.2142	0.2375	0	0	0	0
$A2$	1	1	41.4	0	0	0.2375	0.2142	0	0
	2	1	41.5	0	0	0.2142	0.2375	0	0
$A3$	1	1	41.1	0	0	0	0	0.2375	0.2142
	2	1	40.8	0	0	0	0	0.2142	0.2375

パラメータに関する分散共分散行列の計算

$V(\mu)=(X^T H^{-1} X)^{-1}$ 0.0753

図 5.2 パラメータに関する分散共分散行列の計算

=Minverse(Mmult(Transpose(D9:D14),Mmult(Minverse(F9:K14),D9:D14)))

- パラメータに関する分散共分散行列の計算.
- パラメータ数が 1 なので分散のみとなる.

繰返しが不揃いの場合の分散共分散

表 5.1 繰返しが不揃いの場合の分散共分散分行列

			$A1$		$A2$		$A3$
A_i	r_j	y_{ij}	$r_j=1$	$r_j=2$	$r_j=1$	$r_j=2$	$r_j=1$
$A1$	1	42.0	$\sigma_A^2+\sigma_e^2$	σ_A^2			
	2	41.8	σ_A^2	$\sigma_A^2+\sigma_e^2$			
$A2$	1	41.4			$\sigma_A^2+\sigma_e^2$	σ_A^2	
	2	41.5			σ_A^2	$\sigma_A^2+\sigma_e^2$	
$A3$	1	41.1					$\sigma_A^2+\sigma_e^2$

- 貨車A3は繰返しが 1 であり，分散 $\sigma_A^2+\sigma_e^2$ のみとなる.

ソルバーによるREML法

	A	r	デザイン行列X	y	H						$y_{ij}-\mu$		
1	REML法による2つの分散成分の同時推定												
2	データyに関する尤度												
4	A1	1	1	42.0	0.1635	0.1509	0	0	0		0.5092	μ=	41.4908
5		2	1	41.8	0.1509	0.1635	0	0	0		0.3092	σ_A^2=	0.1509
6	A2	1	1	41.4	0	0	0.1635	0.1509	0		-0.0908	σ_e^2=	0.0126
7		2	1	41.5	0	0	0.1509	0.1635	0		0.0092		
8	A3	1	1	41.1	0	0	0	0	0.1635		-0.3908		
9		2	.	.							.		
10										f(y) =	0.8564		

パラメータμに関する尤度の計算シート					制限付き尤度
パラメータ	推定値	分散	μ^−μ	$g(\hat{\beta})$	$f(y)/g(\hat{\beta})$
μ^: β0^	41.4908	0.0531	0.0000	1.7316	0.4946
				対数	-0.7041
				−2対数尤度	1.4082

繰返しが不揃いでも計算方法は同じである

統計ソフトJMPによる検証

パラメータ推定値

項	推定値	標準誤差	分母
切片	41.488304	0.230568	

変量効果の予測

REML分散成分の推定値

変量効果	分散比	分散成分
Ai	12.018752	0.1509499
残差		0.0125595
合計		0.1635094

−2対数尤度= 1.4080474523

- 有効数で4桁まで一致している．負の対数尤度は1.408047でありExcelで計算した場合も1.4082と4桁目まで一致している．
- これは，Excelのソルバーの収束条件がディフォルトではJMPよりもゆるいことに起因する数値計算上の問題で，「合っていない」のではなく，ほぼ一致していると判断される

じぐざぐ型2段枝分かれ実験

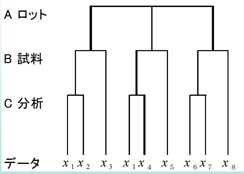


表 6.3　じぐざぐ型の枝分かれデータ

B	C	A1	A2	A3
B1	C1	94	107	86
	C2	96	108	87
B2	C1	90	117	85
	C2	.	.	.

- 試料Bの第2水準目について分析Cの繰返し数が 1 である．
- じぐざぐ型の枝分かれ実験といわれている．

じぐざぐ型をREML法で解析

μ=	96.6667
σ_A^2=	178.7482
σ_B^2=	18.7483
σ_e^2=	1.0042

パラメータμに関する尤度の計算シート					制限付き尤度
パラメータ	推定値	分散	μ^−μ	$\log g(\hat{\beta})$	$\log f(y)/g(\hat{\beta})$
μ^: β0^	96.7	62.8	0.0	-2.99	-23.68
				−2 log(L)	47.36

	X: μ	Z: A1	A2	A3	A1B1	A1B2	A2B1	A2B2	A3B1	A3B2	y_{ijk}	V=ZGZ'+R									$y_{ijk}-\mu$
1	1	1	0	0	1	0	0	0	0	0	94	198.5	197.5	178.7	0.0	0.0	0.0	0.0	0.0	0.0	−2.7
2	1	1	0	0	1	0	0	0	0	0	96	197.5	198.5	178.7	0.0	0.0	0.0	0.0	0.0	0.0	−0.7
3	1	1	0	0	0	1	0	0	0	0	90	178.7	178.7	198.5	0.0	0.0	0.0	0.0	0.0	0.0	−6.7
4	1	0	1	0	0	0	1	0	0	0	107	0.0	0.0	0.0	198.5	197.5	178.7	0.0	0.0	0.0	10.3
5	1	0	1	0	0	0	1	0	0	0	108	0.0	0.0	0.0	197.5	198.5	178.7	0.0	0.0	0.0	11.3
6	1	0	1	0	0	0	0	1	0	0	117	0.0	0.0	0.0	178.7	178.7	198.5	0.0	0.0	0.0	20.3
7	1	0	0	1	0	0	0	0	1	0	86	0.0	0.0	0.0	0.0	0.0	0.0	198.5	197.5	178.7	−10.7
8	1	0	0	1	0	0	0	0	1	0	87	0.0	0.0	0.0	0.0	0.0	0.0	197.5	198.5	178.7	−9.7
9	1	0	0	1	0	0	0	0	0	1	85	0.0	0.0	0.0	0.0	0.0	0.0	178.7	178.7	198.5	−11.7
																				log f(y)	-26.67

G

1	178.7	0	0	0	0	0	0	0	0
2	0	178.7	0	0	0	0	0	0	0
3	0	0	178.7	0	0	0	0	0	0
4	0	0	0	18.7	0	0	0	0	0
5	0	0	0	0	18.7	0	0	0	0
6	0	0	0	0	0	18.7	0	0	0
7	0	0	0	0	0	0	18.7	0	0
8	0	0	0	0	0	0	0	18.7	0
9	0	0	0	0	0	0	0	0	18.7

R

1.0	0	0	0	0	0	0	0	0
0	1.0	0	0	0	0	0	0	0
0	0	1.0	0	0	0	0	0	0
0	0	0	1.0	0	0	0	0	0
0	0	0	0	1.0	0	0	0	0
0	0	0	0	0	1.0	0	0	0
0	0	0	0	0	0	1.0	0	0
0	0	0	0	0	0	0	1.0	0
0	0	0	0	0	0	0	0	1.0

今後の課題

様々な複数の誤差を持つ

実験データの解析について

Excelのソルバーを用いた

線形混合モデルの解析事例を

充実させたい．

10. Excel ソルバーによる反復のある 1 段分割実験の解析

第 5 章で取り上げた事例について Excel のソルバーで解析した結果を示す．JMP による結果と一致している．

反復のある1段分割実験，REML法による解析結果　　2011年11月20日，高橋行雄

N		24
p		12
ln(f)		-9.7142
ln(g)		10.8551
ln(f/g)		-20.5693

$$f(y)=(2\pi)^{-\frac{1}{2}nm}\mid H\mid^{-\frac{1}{2}}\exp\{-\tfrac{1}{2}(y-X\beta)'H^{-1}(y-X\beta)\}\qquad g(\hat{\beta})=(2\pi)^{-\frac{1}{2}p}\mid X'H^{-1}X\mid^{\frac{1}{2}}\exp\{-\tfrac{1}{2}(\hat{\beta}-\beta)'(X'H^{-1}X)(\hat{\beta}-\beta)\}.$$

								X												Z										
									A		B			A×B						R		R×A								
								0	1	2	1	2	3	11	12	13	21	22	23	1	2	11	12	13	21	22	23			
		推定値β^	β^−β		R	A	B	μ	x1	x2	x3	x4	x5	x6	x7	x8	x9	x10	x11	z1	z2	z3	z4	z5	z6	z7	z8	y	Xβ	e
μ	0	13.7417	0	1	1	1	1	1	1	0	1	0	0	1	0	0	0	0	0	1	0	1	0	0	0	0	0	12.3	12.00	0.30
A	1	-0.3667	0	2	1	1	2	1	1	0	0	1	0	0	1	0	0	0	0	1	0	1	0	0	0	0	0	13.9	13.30	0.60
	2	0.6333	0	3	1	1	3	1	1	0	0	0	1	0	0	1	0	0	0	1	0	1	0	0	0	0	0	15.0	14.00	1.00
B	1	-0.1917	0	4	1	1	4	1	1	0	−1	−1	−1	−1	−1	−1	0	0	0	1	0	1	0	0	0	0	0	15.0	14.20	0.80
	2	-0.2083	0	5	1	2	1	1	0	1	1	0	0	0	0	0	1	0	0	1	0	0	1	0	0	0	0	15.5	14.70	0.80
	3	0.4750	0	6	1	2	2	1	0	1	0	1	0	0	0	0	0	1	0	1	0	0	1	0	0	0	0	14.8	14.05	0.75
A×B	11	-1.1833	0	7	1	2	3	1	0	1	0	0	1	0	0	0	0	0	1	1	0	0	1	0	0	0	0	15.2	14.75	0.45
	12	0.1333	0	8	1	2	4	1	0	1	−1	−1	−1	0	0	0	−1	−1	−1	1	0	0	1	0	0	0	0	14.4	14.00	0.40
	13	0.1500	0	9	1	2	1	1	−1	−1	1	0	0	−1	0	0	−1	0	0	1	0	0	0	1	0	0	0	13.9	13.95	−0.05
	21	0.5167	0	10	1	3	2	1	−1	−1	0	1	0	0	−1	0	0	−1	0	1	0	0	0	1	0	0	0	13.3	13.25	0.05
	22	-0.1167	0	11	1	3	3	1	−1	−1	0	0	1	0	0	−1	0	0	−1	1	0	0	0	1	0	0	0	14.3	13.90	0.40
	23	-0.1000	0	12	1	3	4	1	−1	−1	−1	−1	−1	1	1	1	1	1	1	1	0	0	0	1	0	0	0	12.8	12.80	0.00
σ_R^2		0.3550		13	2	1	1	1	1	0	1	0	0	1	0	0	0	0	0	0	1	0	0	0	1	0	0	11.7	12.00	−0.30
σ_{RA}^2		0.1667		14	2	1	2	1	1	0	0	1	0	0	1	0	0	0	0	0	1	0	0	0	1	0	0	12.7	13.30	−0.60
σ_e^2		0.1150		15	2	1	3	1	1	0	0	0	1	0	0	1	0	0	0	0	1	0	0	0	1	0	0	13.0	14.00	−1.00
				16	2	1	4	1	1	0	−1	−1	−1	−1	−1	−1	0	0	0	0	1	0	0	0	1	0	0	13.4	14.20	−0.80
				17	2	2	1	1	0	1	1	0	0	0	0	0	1	0	0	0	1	0	0	0	0	1	0	13.9	14.70	−0.80
				18	2	2	2	1	0	1	0	1	0	0	0	0	0	1	0	0	1	0	0	0	0	1	0	13.3	14.05	−0.75
				19	2	2	3	1	0	1	0	0	1	0	0	0	0	0	1	0	1	0	0	0	0	1	0	14.3	14.75	−0.45
				20	2	2	4	1	0	1	−1	−1	−1	0	0	0	−1	−1	−1	0	1	0	0	0	0	1	0	13.6	14.00	−0.40
				21	2	2	1	1	−1	−1	1	0	0	−1	0	0	−1	0	0	0	1	0	0	0	0	0	1	14.0	13.95	0.05
				22	2	3	2	1	−1	−1	0	1	0	0	−1	0	0	−1	0	0	1	0	0	0	0	0	1	13.2	13.25	−0.05
				23	2	3	3	1	−1	−1	0	0	1	0	0	−1	0	0	−1	0	1	0	0	0	0	0	1	13.5	13.90	−0.40
				24	2	3	4	1	−1	−1	−1	−1	−1	1	1	1	1	1	1	0	1	0	0	0	0	0	1	12.8	12.80	0.00

H=ZGZ'+R

1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24
0.637	1	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
0.522	1	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
0.522	1	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
0.522	1	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
0.355	0	0	0	1	1	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
0.355	0	0	0	1	1	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
0.355	0	0	0	1	1	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
0.355	0	0	0	1	1	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
0.355	0	0	0	0	0	0	0	1	1	1	1	0	0	0	0	0	0	0	0	0	0	0	0
0.355	0	0	0	0	0	0	0	1	1	1	1	0	0	0	0	0	0	0	0	0	0	0	0
0.355	0	0	0	0	0	0	0	1	1	1	1	0	0	0	0	0	0	0	0	0	0	0	0
0.355	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	0.637	1	1	1	0	0	0	0	0	0	0	0.355
0	0	0	0	0	0	0	0	0	0	0	0	0.522	1	1	1	0	0	0	0	0	0	0	0.355
0	0	0	0	0	0	0	0	0	0	0	0	0.522	1	1	1	0	0	0	0	0	0	0	0.355
0	0	0	0	0	0	0	0	0	0	0	0	0.522	1	1	1	0	0	0	0	0	0	0	0.355
0	0	0	0	0	0	0	0	0	0	0	0	0.355	0	0	0	1	1	1	1	0	0	0	0.355
0	0	0	0	0	0	0	0	0	0	0	0	0.355	0	0	0	1	1	1	1	0	0	0	0.355
0	0	0	0	0	0	0	0	0	0	0	0	0.355	0	0	0	1	1	1	1	0	0	0	0.355
0	0	0	0	0	0	0	0	0	0	0	0	0.355	0	0	0	1	1	1	1	0	0	0	0.355
0	0	0	0	0	0	0	0	0	0	0	0	0.355	0	0	0	0	0	0	0	1	1	1	0.522
0	0	0	0	0	0	0	0	0	0	0	0	0.355	0	0	0	0	0	0	0	1	1	1	0.522
0	0	0	0	0	0	0	0	0	0	0	0	0.355	0	0	0	0	0	0	0	1	1	1	0.522
0	0	0	0	0	0	0	0	0	0	0	0	0.355	0	0	0	0	0	0	0	1	1	1	0.637

X'H⁻¹X

	1	2	3	4	5	6	7	8	9	10	11	12
1	4.8	−0	−0	−0	−0	−0	−0	−0	−0	0	0	0
2	0	20	10	−0	−0	−0	−0	−0	−0	−0	0	0
3	0	10	20	−0	0	0	−0	0	0	−0	0	0
4	0	0	0	104	52	52	0	0	0	0	0	0
5	0	0	−0	52	104	52	0	0	0	0	0	0
6	0	0	−0	52	52	104	0	0	0	0	0	0
7	0	0	0	0	0	0	70	35	35	35	17.4	17
8	0	0	0	0	0	0	35	70	35	17	34.8	17
9	0	0	0	0	0	0	35	35	70	17	17.4	35
10	−0	−0	−0	0	0	0	35	17	17	70	34.8	35
11	−0	0	0	0	0	0	17	35	17	35	69.6	35
12	−0	0	−0	0	0	0	17	17	35	35	34.8	70

G

1	2	3	4	5	6	7	8
0.4	0.0	0.0	0.0	0.0	0.0	0.0	0.0
0.0	0.4	0.0	0.0	0.0	0.0	0.0	0.0
0.0	0.0	0.2	0.0	0.0	0.0	0.0	0.0
0.0	0.0	0.0	0.2	0.0	0.0	0.0	0.0
0.0	0.0	0.0	0.0	0.2	0.0	0.0	0.0
0.0	0.0	0.0	0.0	0.0	0.2	0.0	0.0
0.0	0.0	0.0	0.0	0.0	0.0	0.2	0.0
0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.2

R

1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24
0.1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
0	0.1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
0	0	0.1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
0	0	0	0.1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
0	0	0	0	0.1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0.1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0.1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0.1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0.1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0.1	0	0	0	0	0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	0.1	0	0	0	0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	0	0.1	0	0	0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	0	0	0.1	0	0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	0	0	0	0.1	0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0.1	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0.1	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0.1	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0.1	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0.1	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0.1	0	0	0	0
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0.1	0	0	0
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0.1	0	0
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0.1	0
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0.1

JMP による固定効果の推定（再掲）

パラメータ推定値

項	推定値	標準誤差	分母自由度	t値	p値(Prob>|t|)
切片	13.7417	0.4583	1	29.98	0.0212*
A[A1]	−0.3667	0.2552	2	−1.44	0.2874
A[A2]	0.6333	0.2552	2	2.48	0.1312
B[B1]	−0.1917	0.1199	9	−1.60	0.1444
B[B2]	−0.2083	0.1199	9	−1.74	0.1163
B[B3]	0.4750	0.1199	9	3.96	0.0033*
A[A1]*B[B1]	−1.1833	0.1696	9	−6.98	<.0001*
A[A1]*B[B2]	0.1333	0.1696	9	0.79	0.4519
A[A1]*B[B3]	0.1500	0.1696	9	0.88	0.3994
A[A2]*B[B1]	0.5167	0.1696	9	3.05	0.0139*
A[A2]*B[B2]	−0.1167	0.1696	9	−0.69	0.5088
A[A2]*B[B3]	−0.1000	0.1696	9	−0.59	0.5699

μ	0	13.7417
A	1	-0.3667
	2	0.6333
B	1	-0.1917
	2	-0.2083
	3	0.4750
A×B	11	-1.1833
	12	0.1333
	13	0.1500
	21	0.5167
	22	-0.1167
	23	-0.1000

Excel の固定効果についての REML 法による推定値は，JMP の結果と一致している

JMP の REML 法による分散成分の推定（再掲）

REML法による分散成分推定値

変量効果	分散比	分散成分	標準誤差	95%下側	95%上側	全体に対する百分率
R	3.09	0.3550	0.5977	0.0603	6049.5435	55.8
R*A	1.45	0.1667	0.1959	0.0392	22.2617	26.2
残差		0.1150	0.0542	0.0544	0.3833	18.1
合計		0.6367				100.0

−2対数尤度= 41.138630026

σ_R^2	0.3550
σ_{RA}^2	0.1667
σ_e^2	0.1150

分散成分の推定値も一致している．

11. ブロックに因子を割り付けたクロスオーバー法

11.1. JMP による解析

誤差が複数あるような場合についてまず EMS 法で複数の誤差を同時推定してみよう.

表 11.1　乱塊法のデータ（Brown 本）

Patient	Treatment A	Treatment B	Difference A- B	Patient mean
1	20	12	8	16.0
2	26	24	2	25.0
3	16	17	-1	16.5
4	29	21	8	25.0
5	22	21	1	21.5
6	24	17	7	20.5
Mean	22.83	18.67	4.17	20.75

JMP「モデルのあてあめ」の EMS 法による解析結果を次に示す.

分散分析

要因	自由度	平方和	平均平方	F値
モデル	6	206.83333	34.4722	4.3728
誤差	5	39.41667	7.8833	p値(Prob>F)
全体(修正済み)	11	246.25000		0.0634

分散成分推定値

成分	分散成分推定値	全体に対する百分率
Patient&変量効果	11.53333	59.399
残差	7.883333	40.601
合計	19.41667	100.000

平均平方がその期待値に等しいものとして推定したものです。

変量効果を考慮した検定

要因	平方和	分子平均平方	分子自由度	F値	p値(Prob>F)
Patient&変量効果	154.75	30.95	5	3.9260	0.0798
Treatment	52.0833	52.0833	1	6.6068	0.0500

Patient を変量, Tatment を固定効果とした結果, $\hat{\sigma}_{\mathrm{P}}^2 = 11.54$, 誤差 $\hat{\sigma}_e^2 = 7.88$ が推定されている.

REML 法によって症例間の誤差を接求めてみよう. Treatment (A, B) の 2 水準のダミー変数 (1, -1) を x, 症例間の誤差を $\varepsilon_{(p)i}$ としたときに,

$$y_{ij} = \beta_0 + \beta_1 x_j + \varepsilon_{(p)i} + \varepsilon_{ij}, \quad i = 1,2,...6, \quad j = 1,2,$$

を考える. ここで,

$\varepsilon_{(p)i}$ ：症例間の誤差, 正規分布（0, σ_p^2）に従う

ε_{ij} ：症例内の誤差, 正規分布（0, σ_e^2）に従う

とする.

表 11.2　JMP の REML 法での症例間の誤差の推定

パラメータ推定値

項	推定値	標準誤差	分母自由度	t値	p値(Prob>|t|)
切片	20.7500	1.6060	5	12.92	<.0001*
Treatment[A]	2.0833	0.8105	5	2.57	0.0500

変量効果	分散比	分散成分	標準誤差	95%下側	95%上側	全体に対する百分率
Patient	1.46	11.5333	10.0997	3.4963	219.1545	59.4
残差		7.8833	4.9859	3.0716	47.4207	40.6
合計		19.4167				100.0

-2対数尤度= 60.834203091

最小2乗平均表

水準	最小2乗平均	標準誤差	下側95%	上側95%
A	22.8333	1.7989	18.6249	27.0418
B	18.6667	1.7989	14.4582	22.8751

薬剤間の差の SE は 0.8105 であるが，薬剤ごとの SE は 1.7989 と

症例間の誤差を含むので大きくなっている

11.2. Excel による解析

1 段階目　最尤法　<< パラメータの推定は OK>>

$f(y)$ による最尤解，$\hat{\sigma}_p^2 = 9.611$，$\hat{\sigma}_e^2 = 6.569$ が得られる．

多変量正規分布_芳賀_高橋追加_完成.xls

	A	B	C	D	E	F	G	H	I	J	K
1											
2		クロスオーバー法									
3		最尤法によるσeとσi^(p)の推定									
4		p	2								
5		b0	20.7500							対数尤度	
6		b1	2.0833						（データ）－（パラメータ）		
7		σp^2	9.6111	cov	16.18	9.61	行列式	169.437		-32.7728	2段階最大化
8		σe^2	6.5694		9.61	16.18					
9		Se=12σe^2	78.8334								
10											
11			生データの尤度								
12			Treatment_j		推定値		残差		2変量正規		
13		Patient_i	A	B	y^_i1	y^_i2	r_i1	r_i2	L	ln(L)	
14		1	20	12	22.83	18.67	-2.83	-6.67	0.0029	-5.8381	
15		2	26	24	22.83	18.67	3.17	5.33	0.0051	-5.2831	
16		3	16	17	22.83	18.67	-6.83	-1.67	0.0022	-6.1203	
17		4	29	21	22.83	18.67	6.17	2.33	0.0035	-5.6636	
18		5	22	21	22.83	18.67	-0.83	2.33	0.0082	-4.8075	
19		6	24	17	22.83	18.67	1.17	-1.67	0.0090	-4.7120	
20									最尤解	-32.4247	1段階最大化

2 段階目　REML 法

固定効果についてのパラメータについて多変量正規分布を使って推定しなおした．パラメータの分散共分散行列は簡便的に求めた．

コピー複数誤差をREMLできた.xls

	A	B	C	D	E	F	G	H	I	J	K	L
1												
2		クロスオーバー法										
3		最尤法によるσeとσi^(p)の推定										
4		p	2									
5		b0	20.7500							対数尤度		
6		b1	2.0833						(データ) − (パラメータ)			
7		σp^2	11.5333	cov	19.42	11.53	行列式	243.988		-30.4171	1-2段階 ln(L1)-ln(L2)	
8		σe^2	7.8833		11.53	19.42						
9		Se=12σe^2	94.5999						パラメータの分散共分散行列			
10				X	1	1			6*X'inv(cov)*X		0.3877	0.0000
11					1	-1					0.0000	1.5222
12									行列式		0.5902	
13			生データの尤度									
14			Treatment_j		推定値		残差		2変量正規			
15		Patient_i	A	B	y^_i1	y^_i2	r_i1	r_i2	L	ln(L)		
16		1	20	12	22.83	18.67	-2.83	-6.67	0.0031	-5.7814		
17		2	26	24	22.83	18.67	3.17	5.33	0.0049	-5.3189		
18		3	16	17	22.83	18.67	-6.83	-1.67	0.0024	-6.0166		
19		4	29	21	22.83	18.67	6.17	2.33	0.0036	-5.6360		
20		5	22	21	22.83	18.67	-0.83	2.33	0.0073	-4.9226		
21		6	24	17	22.83	18.67	1.17	-1.67	0.0079	-4.8430		
22									最尤解	-32.5186	1段階 ln(L1)	
23												
27			b0	b1	b0^	b1^	b0^-b0	b1^-b1	L	ln(L)		
28			20.75	2.08	20.75	2.08	0.00	0.00	0.1223	-2.1015	2段回 ln(L2)	

(rows 24–26)

$$f(x;\mu,\Sigma)=\frac{1}{(2\pi)^{p/2}|\Sigma|^{1/2}}\exp\left[-\frac{1}{2}(x-\mu)^T\Sigma^{-1}(x-\mu)\right]$$

JMP による計算結果．尤度は −2×–30.4171=60.8342 と一致．$\hat{\sigma}_p^2$=11.5333，$\hat{\sigma}_e^2$=7.8833 と 2 つの誤差分散が推定できた．

REML_枝分かれ複数誤差をREMLで.xls

$$f(y)=(2\pi)^{-\frac{1}{2}nm}\mid H\mid^{-\frac{1}{2}}\exp\{-\tfrac{1}{2}(y-X\beta)'H^{-1}(y-X\beta)\}$$

$$g(\hat{\beta})=(2\pi)^{-\frac{1}{2}p}\mid X'H^{-1}X\mid^{\frac{1}{2}}\exp\{-\tfrac{1}{2}(\hat{\beta}-\beta)'(X'H^{-1}X)(\hat{\beta}-\beta)\}.$$

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U
1		クロスオーバー法																			
2		最尤法によるσeとσi^(p)の推定																			
3		nm	12																		
4		p	2																		
5		b0	20.7500	0																	
6		b1	2.0833	0																	
7		σp^2	11.5333		ln f(x)		=	-27.5	+	-5	=	-32.5									
8		σe^2	7.8833											-30.4							
9	Se=12σe^2		94.60		ln g(β)		=	-2.1	+	0	=	-2.1									
10																					
11		Patien	repeat	y	X			H											Xβ	y-Xβ	|H|
12		1	1	20	1	1	19.4	11.5	0	0	0	0	0	0	0	0	0	0	22.83	-2.83	2.1E+14
13		1	2	12	1	-1	11.5	19.4	0	0	0	0	0	0	0	0	0	0	18.67	-6.67	
14		2	1	26	1	1	0	0	19.4	11.5	0	0	0	0	0	0	0	0	22.8	3.17	
15		2	2	24	1	-1	0	0	11.5	19.4	0	0	0	0	0	0	0	0	18.7	5.33	
16		3	1	16	1	1	0	0	0	0	19.4	11.5	0	0	0	0	0	0	22.8	-6.83	
17		3	2	17	1	-1	0	0	0	0	11.5	19.4	0	0	0	0	0	0	18.7	-1.67	
18		4	1	29	1	1	0	0	0	0	0	0	19.4	11.5	0	0	0	0	22.8	6.17	
19		4	2	21	1	-1	0	0	0	0	0	0	11.5	19.4	0	0	0	0	18.7	2.33	
20		5	1	22	1	1	0	0	0	0	0	0	0	0	19.4	11.5	0	0	22.8	-0.83	
21		5	2	21	1	-1	0	0	0	0	0	0	0	0	11.5	19.4	0	0	18.7	2.33	
22		6	1	24	1	1	0	0	0	0	0	0	0	0	0	0	19.4	11.5	22.8	1.17	
23		6	2	17	1	-1	0	0	0	0	0	0	0	0	0	0	11.5	19.4	18.7	-1.67	
24							H^-1												X'H^-1 X		|XHX|
25		1	1				0.08	-0.05	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.4	0.0	0.59019
26		1	2				-0.05	0.08	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.0	1.5	
27		2	1				0.00	0.00	0.08	-0.05	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00			
28		2	2				0.00	0.00	-0.05	0.08	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00			
29		3	1				0.00	0.00	0.00	0.00	0.08	-0.05	0.00	0.00	0.00	0.00	0.00	0.00			
30		3	2				0.00	0.00	0.00	0.00	-0.05	0.08	0.00	0.00	0.00	0.00	0.00	0.00			
31		4	1				0.00	0.00	0.00	0.00	0.00	0.00	0.08	-0.05	0.00	0.00	0.00	0.00			
32		4	2				0.00	0.00	0.00	0.00	0.00	0.00	-0.05	0.08	0.00	0.00	0.00	0.00			
33		5	1				0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.08	-0.05	0.00	0.00			
34		5	2				0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	-0.05	0.08	0.00	0.00			
35		6	1				0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.08	-0.05			
36		6	2				0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	-0.05	0.08			

12. 分割実験を活用した応答局面解析

この事例は，Little（2006）の Section16.2「Response Surface experiment in a Split-Plot Design」のものであるが，例示されている実験の背景の説明が十分でないので，実験の背景，実験の意図，実験の割り付けなどをより具体的になるように全面的に書き直した．これは，読者が具体的な実験のイメージをもち．解析の方法について思いめぐらせるような配慮をした．これにより，解析の結果の解釈がしやすくなること期待している．

12.1. トウモロコシの顆粒の製造実験

実験手順

全粒トウモロコシを粉砕してトウモロコシの顆粒を製造しているが，顆粒サイズを均一にして品質の向上をしたい．全粒トウモロコシの含有水分量によって粉砕条件をこれまでの経験により微調整しているが，製品のムラが目立ってきた．そこで，製造条件による顆粒サイズがどのように変化するか実験をして確認することにした．

全粒トウモロコシ (30 kg の袋) の含有水分量を調整して 3 等分し, それぞれ 10 kg ごとに, 粉砕機のローラの間隔，スクリーンのサイズ，ローラのスピードを調整した．調整は，現行で行っている微調整の範囲を参考にして，その下限，平均，上限の 3 水準を適宜組み合わせることにした．

実験結果は，10 kg から製造される顆粒の数とした．顆粒数が多くなるとロスが増大する．逆に顆粒数が少ないと顆粒のサイズにムラがでる．そこで，10 kg から製造される顆粒の数が 500 unit となるような操業条件の組み合わせを探したい．

実験は，30 kg の全粒トウモロコシ 10 袋を用いた．取り上げた因子と水準は，次の通りである．

A：含有水分量，（12%，13%，14%）
B：ローラの間隔，（1.2 mm，1.3 mm，1.4 mm）
C：スクリーンのサイズ（0.9 mm，1 mm，1.1 mm）
D：ローラ回転数（60 rpm，65 rpm，70rpm）

実験結果

実験は，トウモロコシの 10 袋に対して A : 含有水分量 12%を 4 袋，13%に対して 2 袋，14%に対して 4 袋をランダムに割りつけた. それぞれの袋を 10 kg ごとに分け, B : ローラの間隔, C：スクリーンのサイズ，D：ローラ回転数の水準の組み合わせは，（下限，中央，上限）の

頻度が（2：1：2）となるようにあらかじめ確率的に割り付けた．最終的に表 12.1 に示すように全体で 30 個のデータを得た．

表 12.1　トウモロコシ顆粒製造実験の結果

実験順	batch	A:水分量 %	B:間隔 mm	C:網目 mm	D:回転数 rpm	y unit
1	1	14	1.4	1.1	70	505
2	1	14	1.2	0.9	60	493
3	1	14	1.2	1.1	60	491
4	2	14	1.4	0.9	65	498
5	2	14	1.4	0.9	60	504
6	2	14	1.2	1.1	65	500
7	3	12	1.3	0.9	60	494
8	3	12	1.3	1.1	65	498
9	3	12	1.2	1.0	70	498
10	4	13	1.2	0.9	65	496
11	4	13	1.3	1.1	70	503
12	4	13	1.2	1.0	60	496
13	5	12	1.2	1.1	70	503
14	5	12	1.4	1.1	60	495
15	5	12	1.2	1.1	60	494
16	6	13	1.3	1.0	65	486
17	6	13	1.4	1.1	60	501
18	6	13	1.4	0.9	70	490
19	7	12	1.4	1.0	65	494
20	7	12	1.4	1.1	70	497
21	7	12	1.2	1.1	60	492
22	8	14	1.2	1.1	70	503
23	8	14	1.3	1.0	60	499
24	8	14	1.3	0.9	70	493
25	9	14	1.4	1.1	60	505
26	9	14	1.4	1.0	70	500
27	9	14	1.2	0.9	70	490
28	10	12	1.2	0.9	70	494
29	10	12	1.4	0.9	60	497
30	10	12	1.2	0.9	60	495

表 12.2　因子ごと水準ごとの実験数

因子	水準	N
A:水分量	12	12
	13	6
	14	12
B:間隔	1.2	13
	1.3	6
	1.4	11
C:網目	0.9	11
	1	6
	1.1	13
D:回転数	60	13
	65	6
	70	11

実験のランダム化の確認と反応のグラフ化

4 つの因子について水準の選択が（2：1：2）と確率的なランダム化を行ったので，水準ごとの実験回数を示す．ランダム化に際して，同じ水準の組み合わせが出た場合には，それをスキップするといったなどの方法でおこなった．因子ごとの水準ごとの実験回数を表 12.2 に示す．おおむね（2：1：2）になっていることが確認された．

2 つの因子ごとの組み合わせ実験数を 表 12.3 示す．おおむね 4 隅の実験数が多くなっていることが確認された．

表 12.3　　因子 B と因子 C の水準ごとの組み合わせ実験数

		A:水分量			B:間隔			C:網目			D:回転数		
		12	13	14	1.2	1.3	1.4	0.9	1.0	1.1	60	65	70
B:間隔	1.2	6	2	5	13	0	0	5	2	6	6	2	5
	1.3	2	2	2	0	6	0	2	2	2	2	2	2
	1.4	4	2	5	0	0	11	4	2	5	5	2	4
C:網目	0.9	4	2	5	5	2	4	11	0	0	5	2	4
	1.0	2	2	2	2	2	2	0	6	0	2	2	2
	1.1	6	2	5	6	2	5	0	0	13	6	2	5

次に，因子ごと水準ごとのデータのプロットおよび平均値の推移を 図 12.1 示す．結果は，486 から 505 unit の間の広く分布していて，単因子だけでは，目標の 500 unit となるような制御は困難であることが示唆されている．

原料となる全粒トウモロコシの水分含有量ごとに操業条件にしたら目標の 500 unit のトウモロコシ顆粒が製造できるのであろうか．

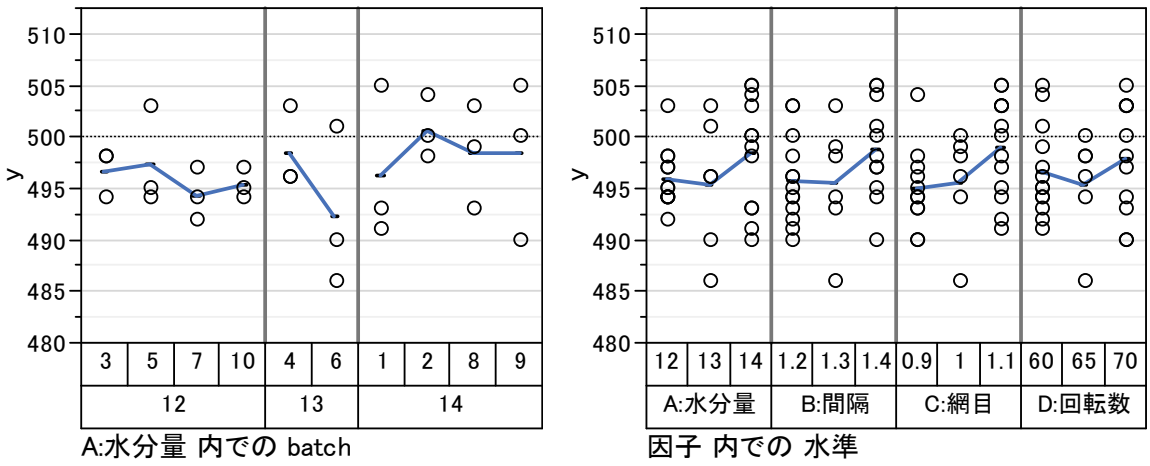


図 12.1　因子 A 内のバッチ別変動，4 因子の水準ごとの変動図

12.2. 山本達郎氏による解析事例

ジャパンゴアテックス株式会社　山本達郎

2010/01/27（水）　2010/03/01(月)改訂

C16_02_トウモロコシ顆粒

ジャパンゴアテックス株式会社
山本達郎
2010/01/27(水)
2010/03/01(月)改訂

1

実験の概要

実験の目的

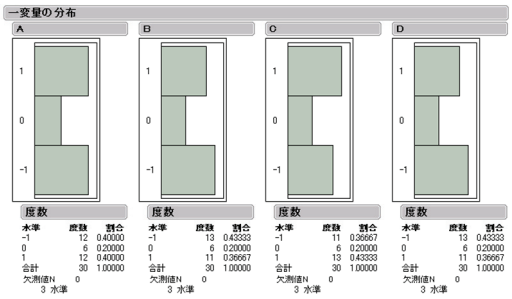
［顆粒の数］が500 unit となるような操業条件の探索

要因と水準　　すべて連続量！

A：含有水分量，（12%，13%，14%）
B：ローラの間隔，（1.2 mm，1.3 mm，1.4 mm）
C：スクリーンのサイズ（0.9 mm，1 mm，1.1 mm）
D：ローラ回転数（60 rpm，65 rpm，70rpm）

2

各要因の水準のバランス



3

データ

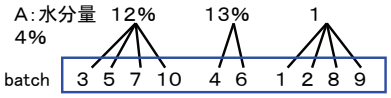
実験順	batch	A:水分量 %	B:間隔 mm	C:網目 mm	D:回転数 rpm	y unit
1	1	14	1.4	1.1	70	505
2	1	14	1.2	0.9	60	493
3	1	14	1.2	1.1	60	491
4	2	14	1.4	0.9	65	498
5	2	14	1.4	0.9	60	504
6	2	14	1.2	1.1	65	500
7	3	12	1.3	0.9	60	494
8	3	12	1.3	1.1	65	498
9	3	12				
10	4	13				
11	4	13				
12	4	13				
13	5	12				
14	5	12				
15	5	12				
16	6	13				
17	6	13				
18	6	13				
19	7	12				
20	7	12				
21	7	12	1.2	1.1	60	492
22	8	14	1.2	1.1	70	503
23	8	14	1.3	1	60	499
24	8	14	1.3	0.9	70	493
25	9	14	1.4	1.1	60	505
26	9	14	1.4	1	70	500
27	9	14	1.2	0.9	70	490
28	10	12	1.2	0.9	70	494
29	10	12	1.4	0.9	60	497
30	10	12	1.2	0.9	60	495

四因子すべて連続量で三水準
Aはブロック化（10ブロック）
B，C，Dはランダムに変更

4

データの構造

[A:水分量]とbatchは枝分かれ構造

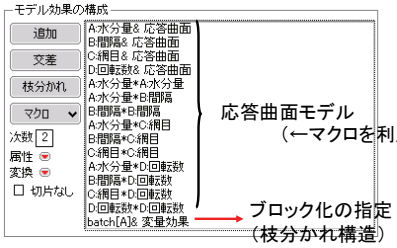


水準の対応がない！

5

[モデルのあてはめ]の指定

すべての変数が連続変数かつ3水準なので，応答曲面モデルで解析．



6

解析結果（フルモデル）

固定効果の検定

要因	自由度	分母自由度	F値	p値(Prob>F)
A:水分量	1	4.41	3.5095	0.1277
B:間隔	1	10.09	5.3664	0.0428*
C:網目	1	13.20	18.6566	0.0008*
D:回転数	1	8.69	0.0878	0.7739
A:水分量*A:水分量	1	4.68	0.0008	0.9787
A:水分量*B:間隔	1	10.31	12.8340	0.0048*
B:間隔*B:間隔	1	13.90	2.8096	0.1160
A:水分量*C:網目	1	14.53	0.3621	0.5566
B:間隔*C:網目	1	11.81	0.3008	0.5936
C:網目*C:網目	1	10.61	1.1730	0.3028
A:水分量*D:回転数	1	10.25	0.3839	0.5491
B:間隔*D:回転数	1	13.51	5.4606	0.0354*
C:網目*D:回転数	1	10.58	17.5019	0.0017*
D:回転数*D:回転数	1	12.53	0.1398	0.7148

ブロック化した[A:水分量]は分母自由度が小さい

7

解析結果（プーリング後）

固定効果の検定

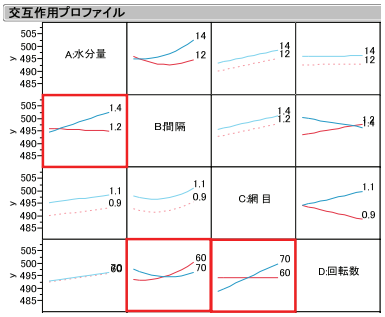
要因	自由度	分母自由度	F値	p値(Prob>F)
A:水分量	1	4.157	5.6581	0.0737
B:間隔	1	14.74	4.6109	0.0488*
C:網目	1	17.28	21.5224	0.0002*
D:回転数	1	11.77	0.1939	0.6677
A:水分量*B:間隔	1	13.51	15.9148	0.0014*
B:間隔*B:間隔	1	20.7	3.4573	0.0772
B:間隔*D:回転数	1	19.12	10.3891	0.0044*
C:網目*D:回転数	1	15.96	21.6061	0.0003*

p値が0．1以下の交互作用をプーリング

一般的な基準ではp値0．2以下あるいはF値2．0以下の要因をプーリングする．

8

交互作用



A*B，B*D，C*D の交互作用は質的交互作用

9

最適化

[顆粒の数]500を実現する条件組み合わせは複数ある．

どの条件を選択するか？→生産性を考慮すべき！

↓

四因子の中で[水分量(A)]は誤差因子である．

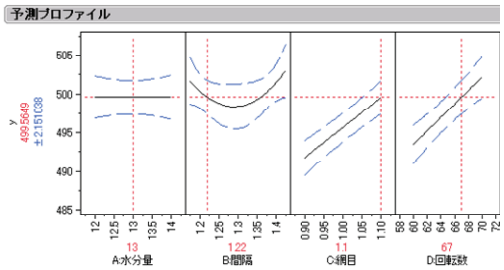
この因子を調整せずに生産できればいいのでは？

A*Bの交互作用を利用して[A:水分量]が[顆粒の数]に影響をしない[B:間隔]の条件を決定する

10

最適化結果

[B:間隔]を1．22に設定することで[A:水分量]が[顆粒の数]に与える影響を減衰できる．



最終結果

因子	設定
A:水分量	任意
B:間隔	1．22
C:網目	1．1
D:回転数	67

微調整は回転数で行う．

補足

最適化方法その2

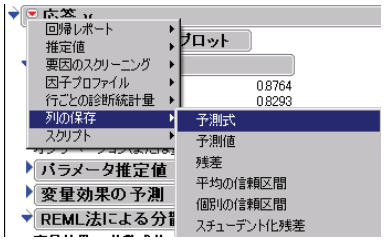
[満足度関数]

[満足度関数]を利用して最適化を行おうとした場合，[顆粒の数]500を実現する条件組み合わせが無数にあり，解が求まらない．

1つの方法として，[グラフ]の[プロファイル]を利用して[A;水分量]の影響を減衰できる条件を探索する方法を考えた．

予測式の保存

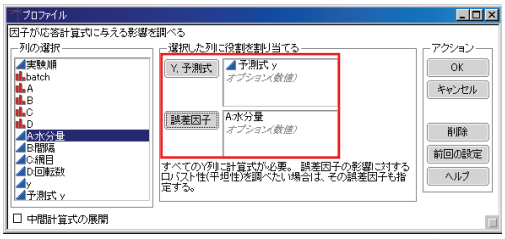
最適化されたモデルを一旦，[予測式]としてデータシートに保存する．



この操作でデータシートに予測式 yという列が追加される．

[グラフ]→[プロファイル]

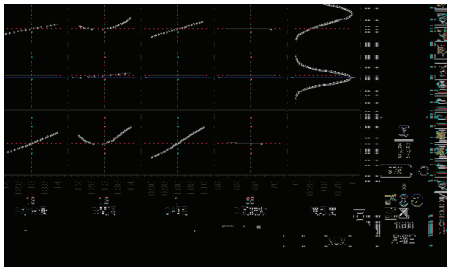
メニューバーの[グラフ]をプルダウンして，[プロファイル]を選択する．



[予測式 y]をY,予測式に，[A;水分量]を誤差因子に指定する

[満足度関数]の設定

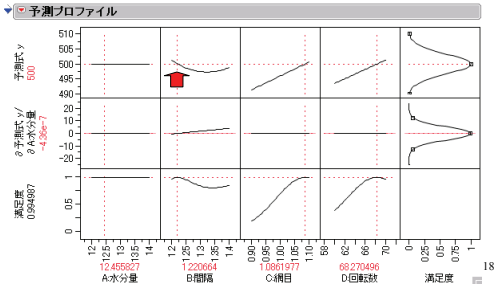
[予測式 y]の満足度が500の時最大になるように設定して[満足度の最大化]を選択する．このとき，二段目は[予測式 y]を[A;水分量]で偏微分した式の予測値．になっており，0が最大に設定されているので．．．



17

最適化後

[A：水分量]が[顆粒の数]に与える影響を最も減衰できる条件である[B;間隔]＝1．22になる．



18

おしまい！

19

13. 欠測セルがある 2 方分割実験

13.1. 2 つの遺伝子型の小麦の収量

窒素肥料の量を変化させて 2 つの遺伝子型の小麦の収量を評価したい．実験圃場を東西方向に 5 区分，南北方向に 3 区分とし，全体で 15 区画とした．5 水準の窒素肥料の量を 5×5 のラテン法格の一部分を用いて割り付けた．東西方向には 5 水準の窒素肥料の量を均等に割り付けられ，南北方向には 5 水準の内 3 水準の窒素肥料の量が割り付けられた．15 区分を 2 つのサブ区画に分け，遺伝子型が異なる 2 つの小麦がランダムに割り付けられた．

表 13.1　窒素肥料の量と遺伝子型の組み合わせた小麦の収量

	東西 B1			東西 B2			東西 B3			東西 B4			東西 B5		
	N	G	y	N	G	y	N	G	y	N	G	y	N	G	y
南北 A1	1	1	20.1	2	1	18.1	5	1	23.1	4	1	22.0	3	1	22.7
（北）		2	20.4		2	16.7		2	18.1		2	18.0		2	19.5
南北 A2	2	1	16.2	1	1	10.3	3	1	17.8	5	1	18.1	4	1	18.9
		2	14.0		2	8.1		2	16.1		2	12.9		2	14.3
南北 A3	3	1	16.2	4	1	15.1	1	1	9.8	2	1	14.0	5	1	17.0
（南）		2	12.6		2	10.5		2	9.0		2	12.1		2	10.9

N：窒素肥料の量，G:遺伝子型，y：収量

Littell（2006）の SAS for Mixed Models 2ed.，Section16.5.

窒素肥料の量を変化させると収量が多くなるが，多すぎても逆に収量が減少することが考えられので，収量を最大にする窒素肥料の量を知りたい．遺伝子型が異なると最適な窒素肥料の量が異なるかもしれない．同じ収量を上げるための窒素肥料の量はどのくらいなのか．少ない窒素肥料の量で収量が高いのはどちらなのか．窒素肥料の量にかかわらず収量が高い遺伝型があるのだろうか．

遺伝子型，窒素肥料の量の組み合わせにより実験データを概観してみよう．遺伝子型，窒素肥料の A:度組み合わせに対しても，南北方向は，A1：北側の収量がよく，A3:南側の収量が落ちることがわかる．東西方向は，一定した関連は見出せない．

実験圃場の場所によって収量が異なることから，窒素肥料の量が 5 水準なのでのそれらの影響を除去できるよう圃場を東西方向と南北方向に 5×5＝25 の区画にして，5 水準の窒素肥料の量を南北方向にも東西方向にもバランスをとったラテン法格で割り付けたい．しかしながら，実験に用いる圃場の東西方向は 5 区分取れたが，南北方向には 3 区分しか取れない．この制約の中で圃場の場所による収量の違いを除去して，窒素肥料の量と小麦の収量の関係を知りたい．

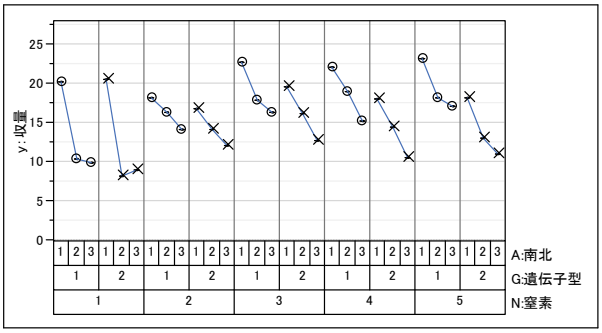


図 13.1　南北方向の収量に与える影響

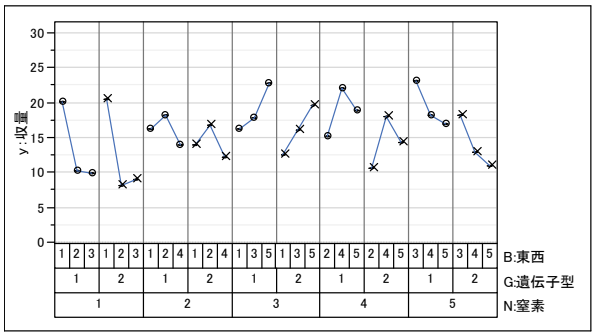


図 13.2　東西方向の収量に与える影響

窒素肥料の量の割付は，3 つの南北方向の区分に対して東西方向の 5 区分された区画には 5 水準の窒素肥料の量がすべて揃っている．東西方向の 5 区分に対して，3 区分の南北方向には 5 水準の窒素肥料の量から 3 水準を選ぶ必要があり，できるだけ 5 区分の東西方向に窒素肥料の量の平均を揃えることが考えられる．

表 13.2 に割付された窒素肥料の量は，西から東に向けて窒素肥料の量の平均が増加している．これは，南北の第 1 区分の窒素肥料の量を，第 2 区分で単純に 1 段階ずらしただけで，窒素肥料の量のバランスを考慮していないためである．

表 13.2　東西方向の窒素肥料の量

A:南北	B:東西 B1	B2	B3	B4	B5
A1	1	2	5	4	3
A2	2	1	3	5	4
A3	3	4	1	2	5
合計	6	7	9	11	12
平均	2.0	2.3	3.0	3.7	4.0

窒素肥料の量の合計がすべて 9 となるのが理想であるが，東西方向でも出来るだけ窒素肥料の量の平均が均一になるような割り付けを表 13.3 に示す．

表 13.3　東西方向の窒素肥料の量の平均を出来るだけ均一にした割り付け

A:南北	B:東西 B1	B2	B3	B4	B5
A1	3	4	5	1	2
A2	4	1	3	2	5
A3	1	5	2	4	3
合計	8	10	10	7	10
平均	2.7	3.3	3.3	2.3	3.3

13.2. 要因配置実験とした解析

圃場の場所が収量に与える影響を無視して実験データの解析を行なってみる．これは，圃場の場所による収量に与える系統的な影響がないならば，完全ランダム実験として解析することができる．15 区画のデータを 5 水準の窒素肥料の量を 2 つの遺伝子型の 2 因子として 表 13.4 のようにまとめなおすと，2 元配置で繰り返しありの分散分析を適用することが妥当かのように錯覚する．遺伝子型 G1 では，窒素肥料を増やすと収量が単調に増加するが，遺伝子型 G2 では，N3 がピークとなっている．

表 13.4　遺伝子ごとの窒素肥料の量を変化させた場合の収量の変化

G:遺伝子型	N:窒素	収量 y			平均
G 1	N1	20.1	10.3	9.8	13.4
	N2	16.2	18.1	14.0	16.1
	N3	16.2	17.8	17.8	17.3
	N4	15.1	22.0	18.9	18.7
	N5	23.1	18.1	17.0	19.4
G 2	N1	20.4	8.1	9.0	12.5
	N2	14.0	16.7	12.1	14.3
	N3	12.6	16.1	16.1	14.9
	N4	10.5	18.0	14.3	14.3
	N5	18.1	12.9	10.9	14.0

分散分析の結果は 表 13.5 に示すように，誤差分散が 16.41 と大きく主効果も交互作用もはっきりとした差が見出せない．

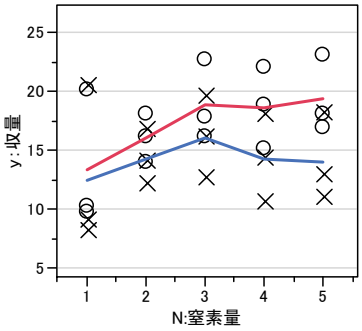


図 13.3　窒素肥料の量と収量の関連
○：G1，×：G2

表 13.5　完全ランダムとした分散分析表

要因	*df*	*s*	*ms*	*F* 値	*p* 値
N:窒素	4	75.30	18.83	1.15	0.3632
G:遺伝子型	1	71.15	71.15	4.34	0.0503
N*G	4	20.47	5.12	0.31	0.8666
誤差	20	328.11	16.41		
全体	29	495.03			

南北方向，東西方向

南北方向の 3 区分では東西方向の 5 区分に窒素肥料の量の 5 水準がすべて割り付けられているので, 南北方向を因子 A, 遺伝子型を G とし繰り返し 5 の 2 元配置分散分析を試みよう.

表 13.6　南北方向の収量の違い

要因	*df*	*s*	*ms*	*F* 値	*p* 値
A:南北	2	273.22	136.61	21.86	<.0001
G:遺伝子	1	71.15	71.15	11.39	0.0025
A*G	2	0.72	0.36	0.06	0.9440
誤差	24	149.95			
全体	29	495.03			

北側の収量が最も大きく南側が小さくなっており，A:南北の平均平方 ms が G:遺伝子型の ms より倍程度大きい.

東西方向の違いは，窒素肥料の量が 5 区分の中でアンバランスなために単純な解析では見出せないが，東西方向で整理した表 13.7 からは，同じ窒素肥料 N1 の中で B3 区画では 9.8 と B1 の半分であるが，窒素肥料 N3 の B3 区画 17.8 は B1 区画 16.2 よりも多く，一定の関連は見出すことが出来ない.

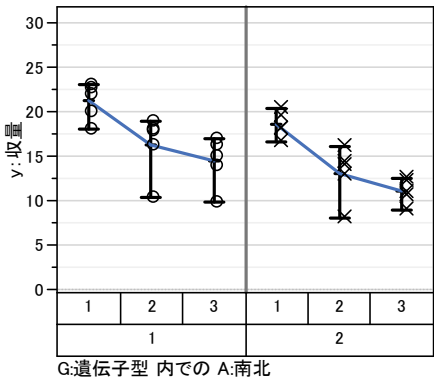


図 13.4　南北方向の収量の違い

表 13.7　東西方向の収量

		窒素肥料の量				
G:遺伝子	B:東西	N1	N2	N3	N4	N5
G1	B1	20.1	16.2	16.2	.	.
	B2	10.3	18.1	.	15.1	.
	B3	9.8	.	17.8	.	23.1
	B4	.	14.0	.	22.0	18.1
	B5	.	.	22.7	18.9	17.0
G1	B1	20.4	14.0	12.6	.	.
	B2	8.1	16.7	.	10.5	.
	B3	9.0	.	16.1	.	18.1
	B4	.	12.1	.	18.0	12.9
	B5	.	.	19.5	14.3	10.9

遺伝子の型の収量の差

東西・南北で区切られた 15 区画に対して窒素肥料の量の 5 水準をランダムに 3 区画に割り付，その中で 2 つの遺伝子型の作物を割り付けたとすると，窒素肥料の量について繰り返し

3 の 1 元配置分散分析とみなせる．

遺伝子型は，15 区画それぞれの中でランダムに G1 か G2 を割り付けているので，その収量の差を計算すると，表 13.8 に示すように窒素肥料の量が増えるにつれて差が広がって行く．遺伝子型の収量の差について 1 元配置分散分析を行うと N:窒素肥料の量の P 値 0.0002 と高度に有意な差となる．

15 のサブ区画の中に割り付けた遺伝子型の収量の差については，窒素肥料の量によって，収量の差が拡大することが統計的にも捉えられた．このことは，15 の区画間による地味の差が収量に与える影響が強いために，その影響を統計的に取り除かないかぎり収量そのものを用いた場合には窒素肥料の量による変化による収量の差を統計的に見出しがたいのである．

遺伝子型ごとの窒素肥料の量による収量の変化の違いを統計的に見出したいのであるが，一筋縄ではいかないことがわかった．

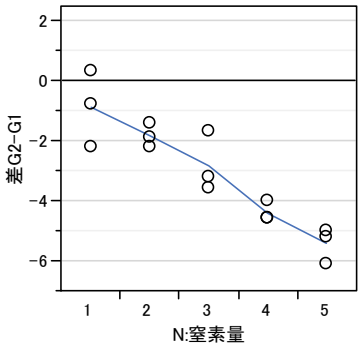


図 13.5　遺伝子型の収量の差の変化

表 13.8　窒素肥料の量ごとに遺伝子型の差の変化

N:窒素	A:南北	B:東西	G1	G2	差 G2-G1	差の平均
N1	1	1	20.1	20.4	0.3	
	2	2	10.3	8.1	-2.2	
	3	3	9.8	9.0	-0.8	-0.9
N2	1	2	18.1	16.7	-1.4	
	2	1	16.2	14.0	-2.2	
	3	4	14.0	12.1	-1.9	-1.8
N3	1	5	22.7	19.5	-3.2	
	2	3	17.8	16.1	-1.7	
	3	1	16.2	12.6	-3.6	-2.8
N4	1	4	22.0	18.0	-4.0	
	2	5	18.9	14.3	-4.6	
	3	2	15.1	10.5	-4.6	-4.4
N5	1	3	23.1	18.1	-5.0	
	2	4	18.1	12.9	-5.2	
	3	5	17.0	10.9	-6.1	-5.4

13.3. 分割実験として解析

窒素肥料の量を名義尺度

この実験は，A:南北，B:東西，N;窒素肥料の量について不完備なラテン方格での割り付け

が 15 区画でなされ，それぞれの区画を 2 つに分割して G:遺伝子型を割り付けているので，枝分かれ型の分割型実験モデルとなっている．このような場合に，因子を逐次的に解析モデルに入れて誤差平方和を求め，その差分から表 13.9 のような分散分析表を作成することができる．

A:南北による収量の変化が大きく，B:東西および N:窒素肥料の効果は認められなかった．G:遺伝子型は 2 次誤差の平方和が小さいので，F = 222.34 と大きな F 値となっている．交互作用 G×N も高度に有意となっている．

表 13.9　1 段分割実験とした分散分析表

要因	水準数	*df*	逐次 *ss*	*ms*	*F*（1 次）	*F*（2 次）	*p* 値
A:南北	3	2	273.22	136.61	31.59		0.0035
B:東西	5	4	59.34	14.84	3.43		0.1298
N:窒素	5	4	50.36	12.59	2.91		0.1626
区画（1 次誤差）	15	4	17.30	4.32	1.00	13.51	0.0005
G:遺伝子型	2	1	71.15	71.15		222.34	0.0000
G×N	10	4	20.47	5.12		15.99	0.0002
2 次誤差		10	3.20	0.32		1.00	
全体		29	495.03				

窒素肥料の量を連続量

窒素肥料の量について 5 水準の名義尺度として取り扱っているために誤差の自由度が 4 と小さいので，連続量（1，2, . . . , 5）として 2 次曲線をあてはめ，誤差の自由度も上げ，遺伝子型との交互作用も含めて，窒素肥料の量と収量の関係を詳細に検討してみよう．

表 13.10 に示すように窒素肥料の 2 次 N×N の P 値は 0.0422 と有意な差となり曲線関係となることがわかる．遺伝子型との交互作用は 1 次 G×N が有意で，2 次 G×N×N は有意でないので，遺伝子型により窒素肥料の量による収量の 1 次的な関連に差があるものの 2 次的な曲線の形状には差がないこと推測できる．

表 13.10　窒素肥料の量を連続量とした 1 段分割実験

要因	水準数	*df*	*ss*	*ms*	*F*（1 次）	*F*（2 次）	*p* 値
A:南北	3	2	273.22	136.61	47.32		0.0002
B:東西	5	4	59.34	14.84	5.14		0.0383
N:窒素	1	1	31.22	31.22	10.82		0.0166
N×N	1	1	19.11	19.11	6.62		0.0422
区画（1 次誤差）	15	6	17.32	2.89	1.00	13.51	0.0001
G:遺伝子	2	1	71.15	71.15		258.03	0.0000
G×N	1	1	20.30	20.30		73.62	0.0000
G×N×N	1	1	0.06	0.06		0.23	0.6413
2 次誤差	15	12	3.31	0.28		1.00	
全体	30	29	495.03				

図 13.6 に遺伝子型別に窒素肥料の量を連続量とした場合の 2 次曲線のあてはめを行った結果を図示する．しかし，この図は区画に関連する地味をまったく無視した結果であり，表 13.11 に示すが，2 次の項は有意ではなく，統計的には 2 次曲線のあてはめは統計的には無理がある．

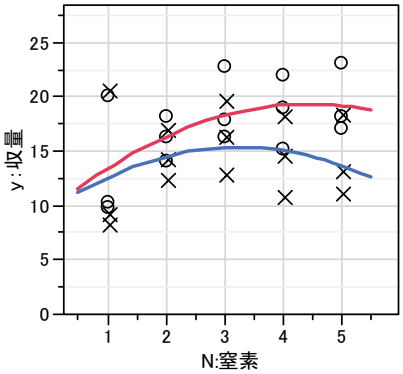


図 13.6　遺伝子型別の 2 次曲線のあてはめ

○ : G1，y : 収量= 13.92+ 1.46*N:窒素 - 0.50*(N:窒素-3)^2

× : G2，y : 収量= 14.44 + 0.29*N:窒素 - 0.55*(N:窒素-3)^2

表 13.11　遺伝子型 G1 および G2 の場合の 2 次曲線のあてはめ

パラメータ推定値

項	推定値	標準誤差	t値	p値(Prob>|t|)
切片	13.92	2.37	5.87	<.0001*
N:窒素	1.46	0.64	2.29	0.0413*
(N:窒素-3)^2	-0.50	0.54	-0.92	0.3738

パラメータ推定値

項	推定値	標準誤差	t値	p値(Prob>|t|)
切片	14.44	2.69	5.36	0.0002*
N:窒素	0.29	0.72	0.41	0.6921
(N:窒素-3)^2	-0.55	0.61	-0.90	0.3838

逐次型の平方和を用いて分散分析表にまとめることはできても，遺伝子型別の窒素肥料との 2 次式を推定することができない．どのようにすれば，南北方向，東西方向の地味の差を考慮しつつ 2 次曲線の推定ができるのだろうか．

これまでの解析は，すべての因子を固定効果とした解析法をいろいろ試みてきたのであるが，この実験には 15 区画に割り付けた因子，それぞれの区画内で割り付けた因子があることから 15 区画の割り付けから 1 次誤差が，区画内の割り付けから 2 次誤差と 2 つの誤差が含まれている．従来の分散分析では，誤差が一つの場合についての解析方法が定式化されて，誤差が 2 つの場合については，計算された逐次平方和から 2 次誤差とみなす因子を特定して，分散分析表をあらためてまとめなおす必要があった．

13.4. 交互作用（東西×南北）を変量効果

JMP には，複数の誤差を取り扱う機能が備わっている．誤差とみなす因子について，それを変量効果とすることにより解析可能となっている．

窒素肥料の量を名義尺度

解析モデルは，1 次因子として A:南北，B：東西，N：窒素（5 水準），1 次誤差として，枝分かれ型とした区画［A:南北，B:東西］を変量効果として指定し，2 次単位に G:遺伝子型，G×N を指定する．

A:南北
B:東西
N:窒素
区画[A:南北,B:東西]& 変量効果
G:遺伝子型
G:遺伝子型*N:窒素

G 遺伝子型と A:南北 および B:東西 との交互作用も考えられるのであるが，これらは，実験緒の局所管理のためのものであり,水準の違いについて技術的な意味付けをするものではないので，交互作用は誤差的な変動と考えられるので，交互作用は解析モデルに含めないことにする．

パラメータの推定値および変量効果の推定値を表 13.12 に示す．解析のためのデザイン行列は最後の水準を（-1）とする対比型で与えられているので, N:窒素肥料の第 5 水準は, N5 = - (-1.09 - 0.62 - 0.12 + 0.66）= 1.17 となり，第 1 水準から徐々に収量が上がっていることがわかる．

表 13.12　パラメータの推定値

パラメータ推定値

項	推定値	標準誤差
切片	15.75	0.38
A:南北[1]	4.12	0.54
A:南北[2]	-1.08	0.54
B:東西[1]	1.51	0.84
B:東西[2]	-2.03	0.84
B:東西[3]	0.21	0.84
B:東西[4]	-0.13	0.84
N:窒素[1]	-2.70	0.84
N:窒素[2]	-0.35	0.84
N:窒素[3]	1.01	0.84
N:窒素[4]	1.29	0.84
G:遺伝子型[1]	1.54	0.10
G:遺伝子型[1]*N:窒素[1]	-1.09	0.21
G:遺伝子型[1]*N:窒素[2]	-0.62	0.21
G:遺伝子型[1]*N:窒素[3]	-0.12	0.21
G:遺伝子型[1]*N:窒素[4]	0.66	0.21

表 13.13 に示した REML 分散成分の推定値から，1 次誤差は 2.00，2 次誤差が 0.32 と約 1/6 と小さいことがわかる．これは，表 13.9 に示した逐次平方和に基づく誤差 4.32 と 0.32 に対応する．REML による推定では 1 次誤差が小さいことがわかる．固定効果の P 値は，B:東西意外は同じ結果となっている．

表 13.13　分散分析表

REML分散成分の推定値

変量効果	分散比	分散成分
区画[A:南北,B:東西]	6.26	2.00
残差		0.32
合計		2.32

固定効果の検定

要因	パラメータ数	自由度	分母自由度	F値	p値(Prob>F)
A:南北	2	2	4	31.59	0.0035*
B:東西	4	4	4	1.99	0.2609
N:窒素	4	4	4	2.91	0.1626
G:遺伝子型	1	1	10	222.34	<.0001*
G:遺伝子型*N:窒素	4	4	10	15.99	0.0002*

主効果のプロファイルを図 13.7 示す．A:南北に比べ B：東西は地味の変化が小さいことがわかる．N:遺伝子型と N:窒素の交互作用が有意なので，そのプロファイルも合わせて示す．G1 の場合には窒素量の増加により N4 まで収量が伸び, N5 で頭打ちになっていることがわか

る．G2 は，N3 をピークとする 2 次曲線となっている．このプロファイルは，図 13.6 で示した 2 次曲線の形状とほぼ同じである．

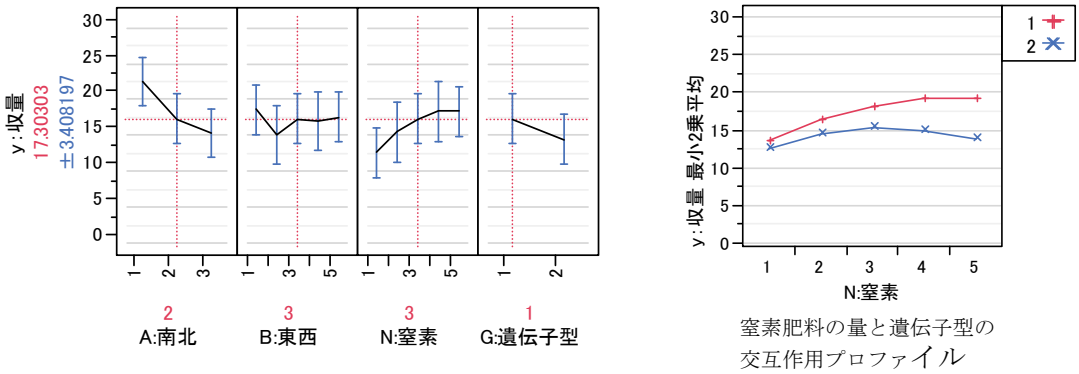


図 13.7　取り上げた因子のプロファイル

窒素肥料の量を連続量

解析モデルは，1 次因子として A:南北，B：東西，N：窒素（5 水準），1 次誤差として，枝分かれ型とした区画［A:南北，B:東西］を変量効果として指定し，2 次単位に G:遺伝子型，G×N を指定する．

A:南北
B:東西
N:窒素
N:窒素*N:窒素
区画[A:南北,B:東西]& 変量効果
G:遺伝子型
N:窒素*G:遺伝子型

表 13.14　パラメータの推定値

パラメータ推定値

項	推定値	標準誤差
切片	9.77	1.57
A:南北[1]	4.12	0.44
A:南北[2]	-1.08	0.44
B:東西[1]	1.52	0.68
B:東西[2]	-2.05	0.64
B:東西[3]	0.22	0.63
B:東西[4]	-0.14	0.64
N:窒素	3.79	1.17
N:窒素*N:窒素	-0.49	0.19
G:遺伝子型[1]	-0.20	0.22
N:窒素*G:遺伝子型[1]	0.58	0.07

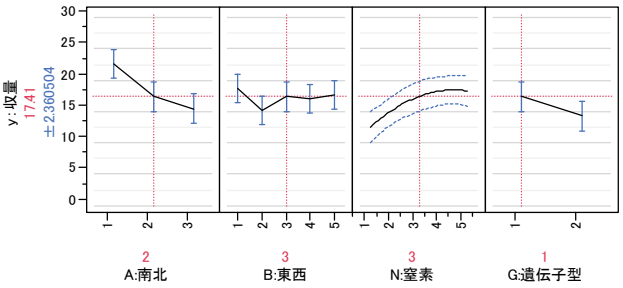
表 13.15　分散分析表

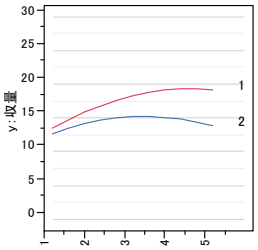
REML分散成分の推定値

変量効果	分散比	分散成分
区画[A:南北,B:東西]	5.07	1.31
残差		0.26
合計		1.57

固定効果の検定

要因	パラメータ数	自由度	分母自由度	F値	p値(Prob>F)
A:南北	2	2	6	47.32	0.0002*
B:東西	4	4	6	3.51	0.0830
N:窒素	1	1	6	10.50	0.0177*
N:窒素*N:窒素	1	1	6	6.62	0.0422*
G:遺伝子型	1	1	13	0.88	0.3643
N:窒素*G:遺伝子型	1	1	13	78.27	<.0001*





窒素肥料の量と遺伝子型の交互作用プロファイル

図 13.8　取り上げた因子のプロファイル

13.5. 南北と東西を別々な変量効果

窒素肥料の量を名義尺度

JMP では実験誤差が複数個あっても解析可能なように拡張されているので，圃場の A:南北と B:東西 を変量効果としてみよう．

A:南北& 変量効果
B:東西& 変量効果
N:窒素
区画[A:南北,B:東西]& 変量効果
G:遺伝子型
G:遺伝子型*N:窒素

表 13.16　パラメータの推定値

パラメータ推定値

項	推定値	標準誤差
切片	15.75	2.17
N:窒素[1]	-2.74	0.86
N:窒素[2]	-0.49	0.86
N:窒素[3]	1.38	0.86
N:窒素[4]	0.99	0.86
G:遺伝子型[1]	1.54	0.10
G:遺伝子型[1]*N:窒素[1]	-1.09	0.21
G:遺伝子型[1]*N:窒素[2]	-0.62	0.21
G:遺伝子型[1]*N:窒素[3]	-0.12	0.21
G:遺伝子型[1]*N:窒素[4]	0.66	0.21

表 13.17　分散分析表

REML分散成分の推定値

変量効果	分散比	分散成分
A:南北	41.37	13.24
B:東西	2.72	0.87
区画[A:南北,B:東西]	6.12	1.96
残差		0.32
合計		16.39

固定効果の検定

要因	パラメータ数	自由度	分母自由度	F値	p値(Prob>F)
N:窒素	4	4	4.95	3.01	0.1302
G:遺伝子型	1	1	10.00	222.34	<.0001*
G:遺伝子型*N:窒素	4	4	10.00	15.99	0.0002*

遺伝子型 1 の場合は，窒素肥料の量が 4 から 5 で収量が最大になるが，遺伝子型 2 の場合は窒素肥料の量が 3 で収量がピークになる．対比による収量の差は 3.41，SE は 1.22 となり，有意な差（p=0.032）である．

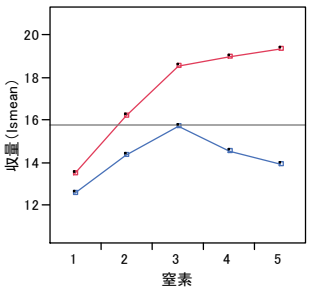


図 13.9　遺伝子型ごとの窒素肥料の量と収量の関連

最小2乗平均表

水準	最小2乗平均	標準誤差
1,1	13.47	2.35
1,2	16.18	2.35
1,3	18.55	2.35
1,4	18.94	2.35
1,5	19.32	2.35
2,1	12.57	2.35
2,2	14.35	2.35
2,3	15.72	2.35
2,4	14.54	2.35
2,5	13.89	2.35

対比

検定の詳細

1,1	0
1,2	0
1,3	0
1,4	0
1,5	1
2,1	0
2,2	0
2,3	−1
2,4	0
2,5	0
推定値	3.6032
標準誤差	1.3403
t値	2.6884
p値(Prob>|t|)	0.0399

分子自由度	1.0000
分母自由度	5.4234
F値	7.2276
p値(Prob>F)	0.0399

対比

検定の詳細

1,1	0
1,2	0
1,3	0
1,4	0.5
1,5	0.5
2,1	0
2,2	0
2,3	−1
2,4	0
2,5	0
推定値	3.4153
標準誤差	1.2227
t値	2.7933
p値(Prob>|t|)	0.0322

分子自由度	1
分母自由度	5.8629246762
F値	7.8022926017
p値(Prob>F)	0.0322247363

図 13.10　対比による最大収量間の差の推定

窒素肥料の量を連続量

窒素肥料の量の 1 から 5 までの等間隔であるとみなして, 連続量であるとした解析を行なう. それぞれの遺伝子型にたいして 2 次の項まで考慮する.

A:南北& 変量効果
B:東西& 変量効果
N:窒素
N:窒素*N:窒素
区画[A:南北,B:東西]& 変量効果
G:遺伝子型
G:遺伝子型*N:窒素

表 13.18　パラメータの推定値

パラメータ推定値

項	推定値	標準誤差
切片	9.67	2.68
N:窒素	3.86	1.16
N:窒素*N:窒素	−0.50	0.19
G:遺伝子型[1]	−0.20	0.22
G:遺伝子型[1]*N:窒素	0.58	0.07

表 13.19　複数の変量効果を考慮した分散分析表

REML分散成分の推定値

変量効果	分散比	分散成分
A:南北	51.58	13.38
B:東西	4.90	1.27
区画[A:南北,B:東西]	4.97	1.29
残差		0.26
合計		16.19

固定効果の検定

要因	パラメータ数	自由度	分母自由度	F値	p値(Prob>F)
N:窒素	1	1	6.42	10.98	0.0146*
N:窒素*N:窒素	1	1	6.37	6.97	0.0364*
G:遺伝子型	1	1	13.00	0.88	0.3643
G:遺伝子型*N:窒素	1	1	13.00	78.27	<.0001*

窒素肥料 N について，2 次の項の p 値が 0.364 となり，2 次式のあてはめがようやく支持された. 遺伝子型 G の主効果は有意でないが, 窒素肥料 N との交互作用は $p < 0.0001$ となった. このことから，遺伝子型 G の種別により窒素肥料 N の量による収量が反応が異なることが確認できた. 図 13.11 に遺伝子型ごとの窒素肥料の量と収量に 2 次式をあてはめた結果を示す.

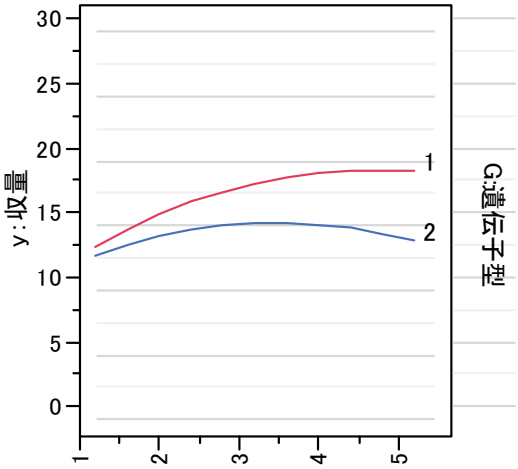


図 13.11　遺伝子型ごとの窒素肥料の量と収量の関連

遺伝子型 1 の場合：$y = (9.67 - 0.21) + (3.86 + 0.58)\ x + (-0.50)\ x^2$

遺伝子型 2 の場合：$y = (9.67 + 0.21) + (3.86 - 0.58)\ x + (-0.50)\ x^2$

14. 経時データの REML 法による解析

14.1. 2 群 4 症例，2 時点の経時データ

群は g，症例を id，時期を t，データを y とする．時点 0 を投与前とし，時点 1 を投与後とする．投与前後差を d とする．ダミー変数 z0 は投与前 $t = 0$ の場合で，z1 は群 g = 1 の場合の投与後，z2 は群 g = 2 の場合の投与後の場合に，1，それ以外は 0 とするダミー変数である．

2 群 4 症例，2 時点のデータ

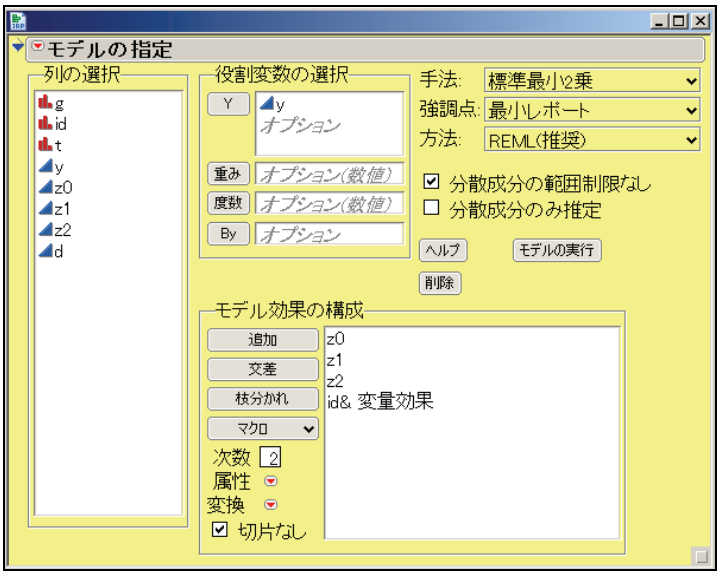
	g	id	t	y	d	z0	z1	z2
1	1	1	0	11	0	1	0	0
2	1	1	1	10	−1	0	1	0
3	1	2	0	12	0	1	0	0
4	1	2	1	10.5	−1.5	0	1	0
5	2	3	0	9	0	1	0	0
6	2	3	1	10	1	0	0	1
7	2	4	0	8	0	1	0	0
8	2	4	1	9.5	1.5	0	0	1

群ごと時点ごとの平均値

	g	t	行数	平均(y)	平均(d)
1	1	0	2	11.5	0
2	1	1	2	10.25	−1.25
3	2	0	2	8.5	0
4	2	1	2	9.75	1.25

y での $t = 1$ の群平均の差は 0.5，d では -2.5

投与前を共通にした混合効果モデル



REML

応答 y

あてはめの要約

R2乗	0.987745
自由度調整R2乗	0.982843
誤差の標準偏差(RMSE)	0.251152
Yの平均	10
オブザベーション(または重みの合計)	8

パラメータ推定値

項	推定値	標準誤差	分母自由度	t値	p値(Prob>|t|)
z0	10.0000	0.890232	3.003	11.23	0.0015*
z1	8.7798	0.908186	3.237	9.67	0.0017*
z2	11.2202	0.908186	3.237	12.35	0.0008*

変量効果の予測

項	BLUP	標準誤差	分母自由度	t値	p値(Prob>|t|)
id[1]	1.0989216	0.898782	3.116	1.22	0.3058
id[2]	1.8413849	0.898782	3.116	2.05	0.1295
id[3]	-1.098922	0.898782	3.116	-1.22	0.3058
id[4]	-1.841385	0.898782	3.116	-2.05	0.1295

REML分散成分の推定値

変量効果	分散比	分散成分	標準誤差	95%下限	95%上限
id	49.256756	3.1069723	2.5881335	-1.965769	8.1797141
残差		0.0630771	0.0636664	0.0169629	2.6436471
合計		3.1700494			

-2対数尤度= 16.262716836

z1 と z2 の差は 2.4403 となり，投与前との差 d の-2.5 に対し小さくなり，平均への回帰現象を若干でも和らげている．

14.2. Excel での計算

手順 1．g を群，id を症例番号，t を時点とするデータを定義する．

i	g	id	t
1	1	1	0
2	1	1	1
4	1	2	0
4	1	2	1
4	2	3	0
6	2	3	1
7	2	4	0
8	2	4	1

手順 2．投与前値 t=0 を群 g で共通とするデザイン行列 X を作成する．

X

x0	x1	x2
1	0	0
0	1	0
1	0	0
0	1	0
1	0	0
0	0	1
1	0	0
0	0	1

手順 3．適当な初期値を入れた固定効果の推定値を β として y^=$X\beta$ を計算する．

β 0=	10.00
β 1=	8.00
β 2=	11.00

y	y^	y−y^
11.00	10.00	1.00
10.00	8.00	2.00
12.00	10.00	2.00
10.50	8.00	2.50
9.00	10.00	−1.00
10.00	11.00	−1.00
8.00	10.00	−2.00
9.50	11.00	−1.50

手順 4. 変量効果（症例）についてインジケータ型の行列 Z を作成する．

id
1
1
2
2
3
3
4
4

r1	r2	r3	r4
1	0	0	0
1	0	0	0
0	1	0	0
0	1	0	0
0	0	1	0
0	0	1	0
0	0	0	1
0	0	0	1

手順 5. 症例についての分散共分散行列 G を作成する．行列 G の対角要素は症例間の分散 σ_A^2 について適当に定めた初期値を代入する．共分散は，症例が互いに独立なので 0 とする．

σ_A^2 =	3.00
σ_e^2 =	0.50

G

3.00	0.00	0.00	0.00
0.00	3.00	0.00	0.00
0.00	0.00	3.00	0.00
0.00	0.00	0.00	3.00

手順 6．症例ごと各時点ごとの症例内誤差マトリックス R を定義する．行列 R の対角要素 σ_e^2 については適当に定めた初期値を代入する．症例の分散を除いた時点間，他の症例の時点間の共分散は互いに独立なので 0 とする．

σ_A^2 =	3.00
σ_e^2 =	0.50

R

1	0.50	0.00	0.00	0.00	0.00	0.00	0.00	0.00
2	0.00	0.50	0.00	0.00	0.00	0.00	0.00	0.00
3	0.00	0.00	0.50	0.00	0.00	0.00	0.00	0.00
4	0.00	0.00	0.00	0.50	0.00	0.00	0.00	0.00
5	0.00	0.00	0.00	0.00	0.50	0.00	0.00	0.00
6	0.00	0.00	0.00	0.00	0.00	0.50	0.00	0.00
7	0.00	0.00	0.00	0.00	0.00	0.00	0.50	0.00
8	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.50

手順 7．症例・時点ついて分散共分散行列 H を計算する．行列 H は，次式で求められる．

$$H = ZGZ' + R$$

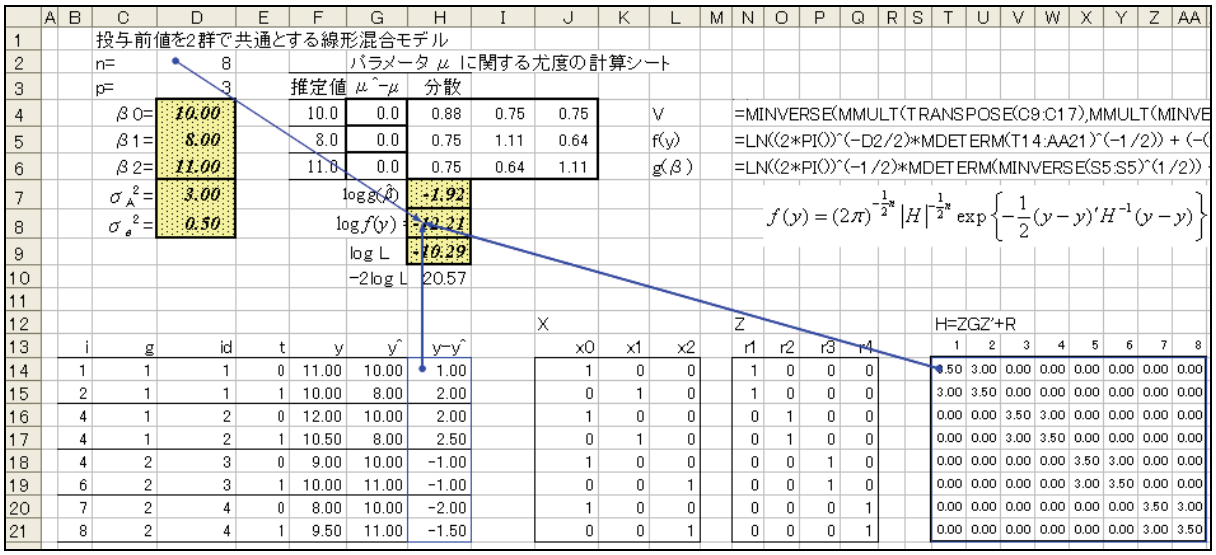
作成された行列 H は，各症例内では Compound Symmetry（対称型）といわれる構造となっている．これを元にすべての症例・時点間の分散共分散行列にしたのが行列 H である．

$$\begin{bmatrix} \begin{matrix} \sigma_A^2+\sigma_e^2 & \sigma_A^2 \\ \sigma_A^2 & \sigma_A^2+\sigma_e^2 \end{matrix} & & \boldsymbol{0} \\ & \ddots & \\ \boldsymbol{0} & & \begin{matrix} \sigma_A^2+\sigma_e^2 & \sigma_A^2 \\ \sigma_A^2 & \sigma_A^2+\sigma_e^2 \end{matrix} \end{bmatrix}$$

手順 8．推定値 ベクトル $(y-\hat{y})$ が 期待値 0，分散共分散 H の 8 次元の多次元正規分布に従うとしたときの対数尤度 $\log f(y)$ を計算する．計算は，次の $n=8$ の多次元正規分布の確率密度 $f(y)$ の対数をとる．

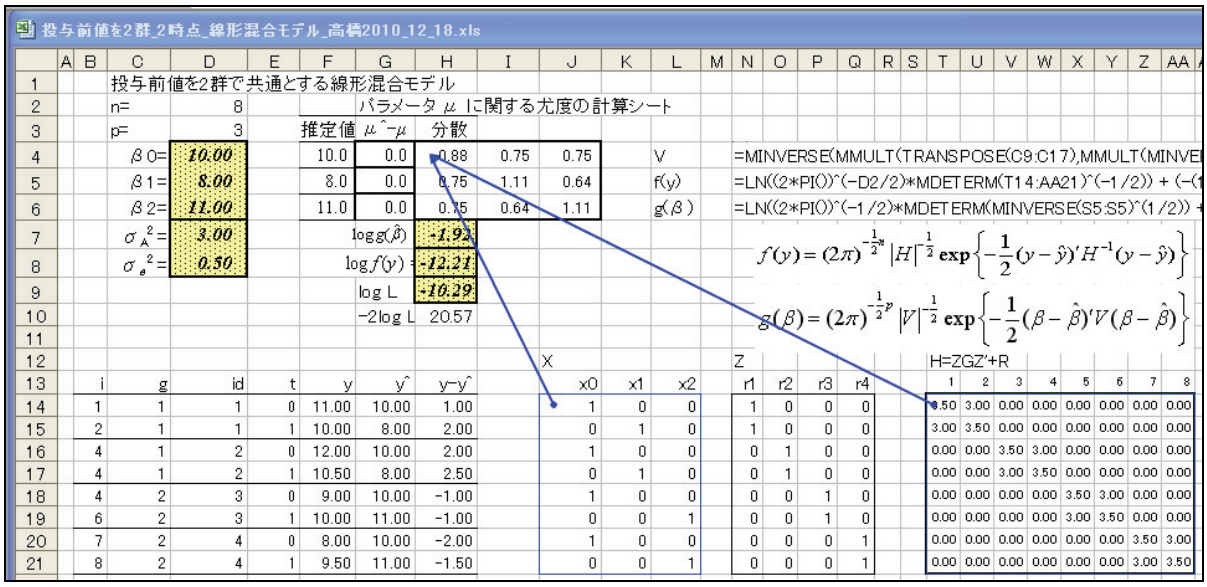
$$\log f(y) = \log\left[(2\pi)^{-\frac{1}{2}n}|H|^{-\frac{1}{2}}\exp\left\{-\frac{1}{2}(y-\hat{y})'H^{-1}(y-\hat{y})\right\}\right]$$

このデータ y に関する尤度 $f(y)$ を最大化する方法が，最尤法 ML といわれているものである．

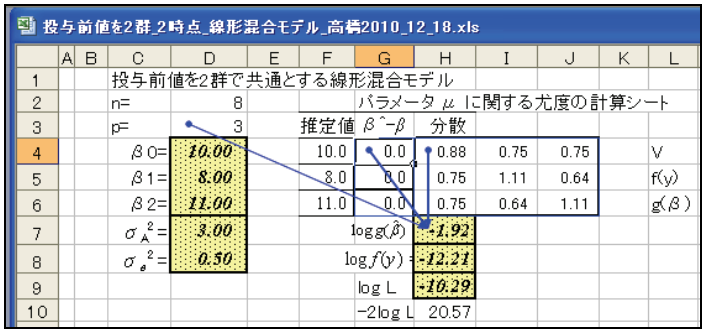


手順 9．固定効果の推定値ベクトル β についての対数尤度を計算を $(\beta-\hat{\beta})$ が 期待値 0，分散共分散 $V=(X'H^{-1}X)^{-1}$ 3 次元の多次元正規分布に従うとしたときの対数尤度 $\log g(y)$ を計算する．計算は，次の $p=3$ の多次元正規分布の確率密度 $g(y)$ の対数をとる．

$$V = (X'H^{-1}X)^{-1}$$



$$g(\beta) = (2\pi)^{-\frac{1}{2}p} |V|^{-\frac{1}{2}} \exp\left\{-\frac{1}{2}(\beta-\hat{\beta})'V(\beta-\hat{\beta})\right\}$$



手順 10. データ y についての尤度 $f(y)$，とパラメタ β についての尤度 $g(y)$ の比

$$L = \frac{f(y)}{g(\beta)}$$

の対数を最大にするように暫定的に決めたパラメタ β と 2 つの分散をソルバーを用いて探索的に変化させる．このプロセスでデータについて求めた尤度 $f(y)$ をパラメータについての尤度 $g(y)$ で制限（除算）することから，制限付き最尤法 REML 問われているゆえんである．与えた初期値と $f(y)$ の $g(y)$ の対数尤度は，次の通り．

β0=	10.00
β1=	8.00
β2=	11.00
σA2=	3.00
σe2=	0.50

log g(β̂)	-1.92
log f(y)	-12.21
log L	-10.29
−2log L	20.57

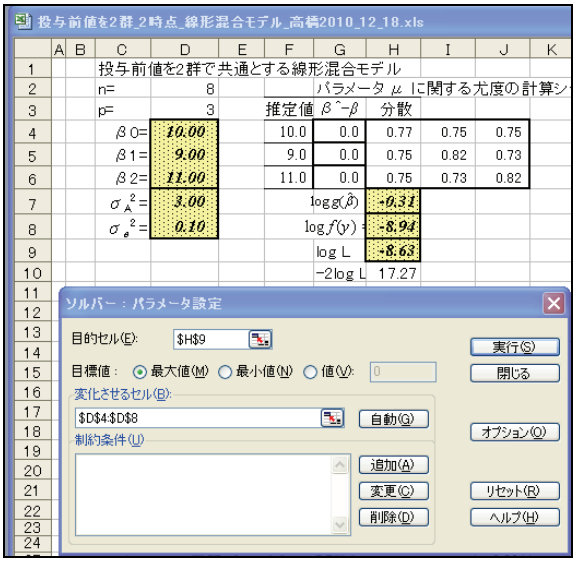
手順 11. データ数 8 に対してパラメータ数 3, 分散成分 2, 計 5 個もあり，ソルバーで解がすぐに求まらない.

手順 12. 手作業で尤度が大きくなるように手動でパラメータと分散成分を動かす. $\beta_1 = 9.0$, $\sigma_e^2 = 0.1$ と変化させると，対数尤度は，−10.29 から -8.63 と大きくなる.

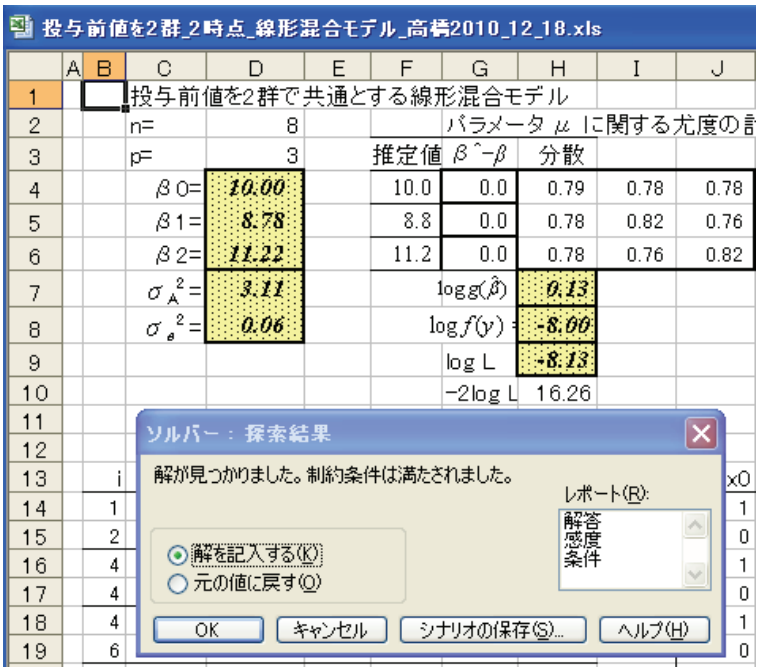
β0=	10.00
β1=	9.00
β2=	11.00
σA2=	3.00
σe2=	0.10

log g(β̂)	-0.31
log f(y)	-8.94
log L	-8.63
−2log L	17.27

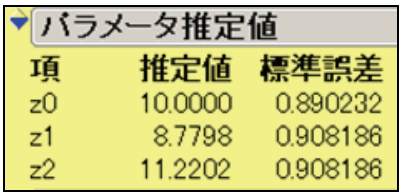
手順 13. ソルバーで Log L を「目的セル」に，β_0 から σ_e^2 を「変化させるセル」にセットし，「実行」する.



手順 14．解が求まったので，「解を記入」で「OK」とする．JMP の結果と一致することが確かめられる．

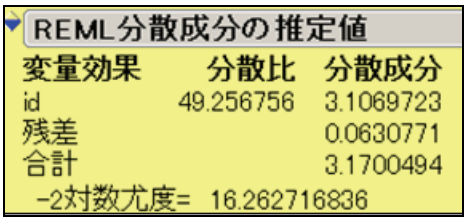


JMP の結果



パラメータ推定値

項	推定値	標準誤差
z0	10.0000	0.890232
z1	8.7798	0.908186
z2	11.2202	0.908186



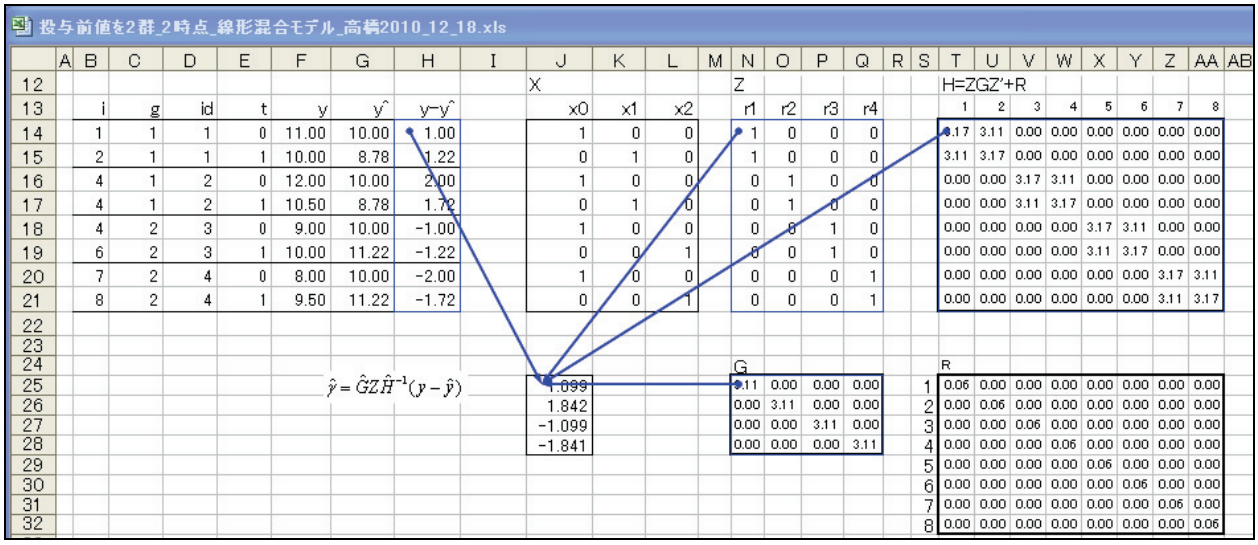
REML分散成分の推定値

変量効果	分散比	分散成分
id	49.256756	3.1069723
残差		0.0630771
合計		3.1700494

−2対数尤度= 16.262716836

手順 15．症例についての変量効果 γ の推定値は，次式で求められる．

$$\hat{\gamma} = \hat{G}Z\hat{H}^{-1}(y - \hat{y})$$



Excel で求めた変量効果の推定値が, JMP で求めた推定値と一致していることが確かめられる.

$$\hat{\gamma} = \hat{G}Z\hat{H}^{-1}(y - \hat{y})$$

id	
1	1.099
2	1.842
3	−1.099
4	−1.841

変量効果の予測

項	BLUP	標準誤差
id[1]	1.0989216	0.898782
id[2]	1.8413849	0.898782
id[3]	−1.098922	0.898782
id[4]	−1.841385	0.898782

手順 16. 群・時期ごとの推定値は, $\hat{y}_{ij} = X\hat{\beta}$ で, 次のように推定されている.

y	y^	y−y^
11.00	10.00	1.00
10.00	8.78	1.22
12.00	10.00	2.00
10.50	8.78	1.72
9.00	10.00	−1.00
10.00	11.22	−1.22
8.00	10.00	−2.00
9.50	11.22	−1.72

手順 17. 症例・時期ごとの推定値は, 変量効果を加えて求めることができる. JMP では条件付の予測値として表示される.

$$\hat{y}'_{ij} = X\hat{\beta} + \hat{\gamma}_i$$

投与前値を2群で共通とする線形混合モデル_追加_高橋2010_12_17.xls

投与前後データの線形混合モデル

変量効果γは、「変量効果の予測」
yの予測値 y^＋γ が「条件付予測値として、JMPの列の保存で行える。

g	id	t	y	d	z0	z1	z2	予測式 y	変量効果γ	y^＋γ	条件付き予測式 y	条件付き残差 y
1	1	0	11	0	1	0	0	10.00	1.10	11.10	11.10	−0.10
1	1	1	10	−1	0	1	0	8.78	1.10	9.88	9.88	0.12
1	2	0	12	0	1	0	0	10.00	1.84	11.84	11.84	0.16
1	2	1	10.5	−1.5	0	1	0	8.78	1.84	10.62	10.62	−0.12
2	3	0	9	0	1	0	0	10.00	−1.10	8.90	8.90	0.10
2	3	1	10	1	0	0	1	11.22	−1.10	10.12	10.12	−0.12
2	4	0	8	0	1	0	0	10.00	−1.84	8.16	8.16	−0.16
2	4	1	9.5	1.5	0	0	1	11.22	−1.84	9.38	9.38	0.12
								計	0.00		計	0.00

変量効果の予測

項	BLUP	標準誤差	分母自由度	t値	p値(Prob>|t|)
id[1]	1.0989	0.8988	3.116	1.22	0.3058
id[2]	1.8414	0.8988	3.116	2.05	0.1295
id[3]	−1.0989	0.8988	3.116	−1.22	0.3058
id[4]	−1.8414	0.8988	3.116	−2.05	0.1295

この変量効果を加えた条件付き予測値 $\hat{y}'_{ij}$ と測定値 y_{ij} との差が誤差 e_{ij} となる. 変量効果 $\hat{\gamma}_i$, 誤差 e_{ij} を加えると 0 になりことが確かめられる.

14.3. 症例を固定効果とし，投与前を群間で共通とした場合

症例を固定効果をみなした最小 2 乗法で，投与前を群間で共通とする解析も考えられる．ただし，この解析法は，投与前からの差のデータで群間比較を行った結果と一致し，平均への回帰を緩和する方法とは言いがたい．

差についての群間比較 t 検定と対比してみよう．症例について対比型のダミー変数として一般線形モデル（最小 2 乗法）で解いてみよう．

症例を対比型のダミー変数（c1，c2，c3）で与えた場合の最小２回乗解

パラメータ推定値

項	推定値	標準誤差
z0	10	0.13
z1	8.75	0.22
z2	11.25	0.22
c1	1.125	0.18
c2	1.875	0.18
c3	-1.125	0.18

群・時点の効果は，群で共通の投与前値 z0=10，1 群の投与後 z1=8.75，2 群の投与後 z2=11.15 が得られている．これらの推定値は，症例ごとの投与前から投与後の差の平均に，共通の投与前値 z0=10 を加えた

$$z1 = \frac{11+12+9+8}{4} + \frac{(10-11)+(10.5-12)}{2} = 10 - 1.25 = 8.75$$

$$z2 = 10 + \frac{(10-9)+(9.5-8)}{2} = 10 + 1.25 = 11.25$$

となっている．従って，これらの推定値は，投与前からの差についての平均値に z0 を加えたものとなっている．

デザイン行列 X・推定値 β ＝予測式 $\hat{y}$

z0	z1	z2	c1	c2	c3
1	0	0	1	0	0
0	1	0	1	0	0
1	0	0	0	1	0
0	1	0	0	1	0
1	0	0	0	0	1
0	0	1	0	0	1
1	0	0	-1	-1	-1
0	0	1	-1	-1	-1

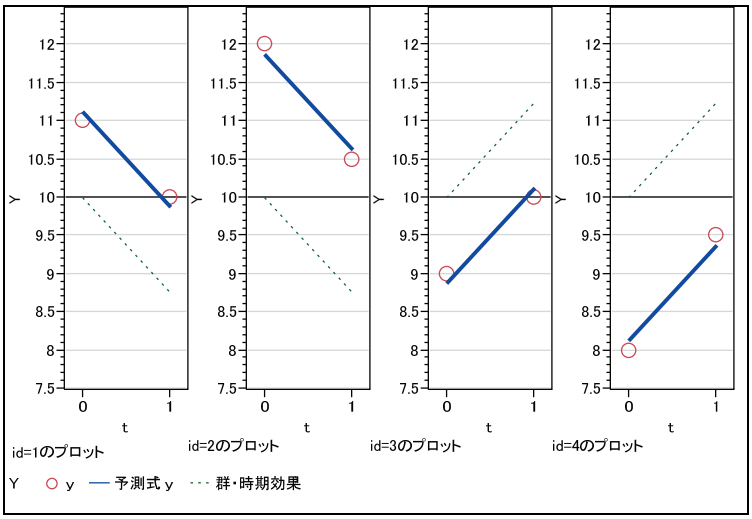
推定値
10
8.75
11.25
1.125
1.875
-1.125

=

予測式 y
11.125
9.875
11.875
10.625
8.875
10.125
8.125
9.375

症例に対する効果は，c1=1.125，c2=1.875，c3=-1.125，c4=-c1-c2-c3=-1.875 となる． これらの実データ y と推定値 $\hat{y}$ の重ね書き図から，群・時期効果の推定値に症例ごとの時点の推定値までの距離が症例の効果（c1，c2，c3，c4）が加算されている．

実データ y と予測値の重ね書き図



14.4. 症例を変量効果とし，投与前を群間で共通とした場合

制限付き最尤法（REML）による固定効果 X に対する推定値 β，変量効果 Z に対する推定値 γ から各症例の時期ごとの推定値が得られる．

z0	z1	z2
1	0	0
0	1	0
1	0	0
0	1	0
1	0	0
0	0	1
1	0	0
0	0	1

推定値
10.00
8.78
11.22

＋

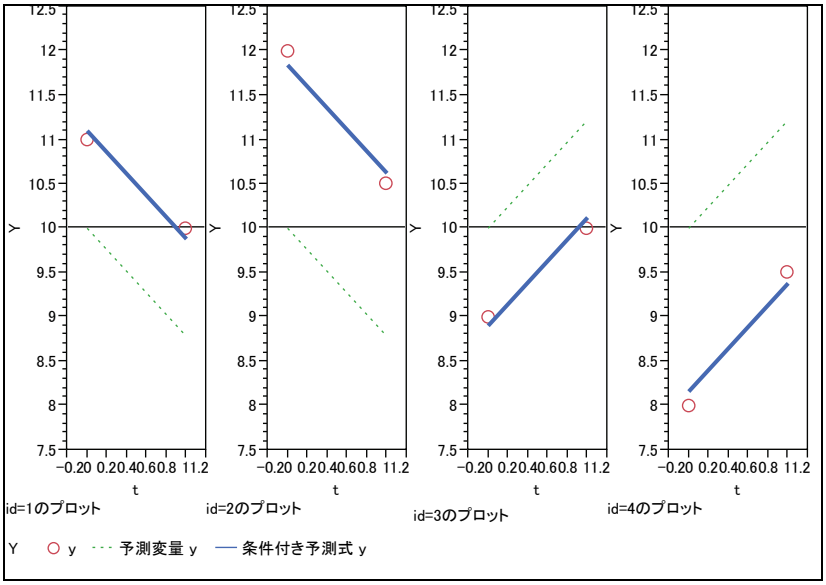
c1	c2	c3	c4
1	0	0	0
1	0	0	0
0	1	0	0
0	1	0	0
0	0	1	0
0	0	1	0
0	0	0	1
0	0	0	1

BLUP
1.10
1.84
-1.10
-1.84

＝

条件付き予測式 y
11.10
9.88
11.84
10.62
8.90
10.12
8.16
9.38

症例を固定効果とた最小２乗法に対して，固定効果および変量効果ともに予測値が期待値方向に縮小していることが観察される．このような予測に関する振る舞いが，投与前の群平均の違いに起因する「平均への回帰」現象を和らげる働きをしている．



14.5. 症例を固定効果とした場合と変量効果とした場合の時期 1 での群間比較

デザイン行列 X を，セル平均型から，群間差型に変更して，群間差を計算する．

セル平均型

z0	z1	z2
1	0	0
0	1	0
1	0	0
0	1	0
1	0	0
0	0	1
1	0	0
0	0	1

==>

群間差型

z0	z1	z2
1	0	0
0	1	0
1	0	0
0	1	0
1	0	0
0	1	1
1	0	0
0	1	1

症例を固定効果モデルとした場合には z2=2.5 と，症例ごとの差

$$d = \frac{(1+1.5)}{2} - \frac{(-1-1.5)}{2} = 1.25 - (-1.25) = 2.5$$

について群間比較を行った場合に検定統計量は一致する．

症例ごとの前後差での群間の t 検定

t検定

2-1
分散が等しいと仮定

差	2.50000	t値	7.071068
差の標準誤差	0.35355	自由度	2
差の上側信頼限界	4.02122	p値(Prob>|t|)	0.0194*
差の下側信頼限界	0.97878	p値(Prob>t)	0.0097*
信頼率	0.95	p値(Prob<t)	0.9903

症例を固定効果とした場合の群間差

パラメータ推定値

項	推定値	標準誤差	t値	p値(Prob>|t|)
z0	10.00	0.125	80.00	0.0002*
z1	8.75	0.216506	40.41	0.0006*
z2	2.50	0.353553	7.07	0.0194*
c1	1.13	0.176777	6.36	0.0238*
c2	1.88	0.176777	10.61	0.0088*
c3	-1.13	0.176777	-6.36	0.0238*

症例を変量効果とした場合は，自由度は 2 から 2.003 と微増するが群間差は 2.5 から 2.44 に縮小し, 誤差 SE も 0.353 から 0.359 と増大するために p 値は 0.194 から 0.209 と増大する.

パラメータ推定値

項	推定値	標準誤差	分母自由度	t値	p値(Prob>|t|)
z0	10	0.890232	3.003	11.23	0.0015*
z1	8.7798467	0.908186	3.237	9.67	0.0017*
z2	2.4403065	0.359383	2.003	6.79	0.0209*

14.6. 考察

この 4 症例の仮想データは，第 1 群の投与前値が高い症例が集まり，第 2 群で低い症例を意図的に割り付けたので，平均への回帰現象を想定した投与後のデータをセットした．症例を固定効果とみなした最小 2 乗法の適用は，症例ごとの差の群間比較と同様の結果となり，平均への回帰現象をまったく考慮しない解析となっている.

他方，症例を変量効果とする REML 法では，変量効果および固定効果にも期待値への縮小効果があり，2 時点の場合でも平均への回帰現象を和らげる働きをしたものと思われる.

14.7. 統計連合 2011 での発表 抄録

はじめに　「平均値と分散の尤度に分解する」との REML 法の解説も散見するが，経時データの解析で直面する複数の固定効果と複数の変量効果（分散）の推定にどのよう一般化するのであろうか．Diggle ら（2002）には，REML 法に関してデータに関する尤度 $f(y)$ と固定効果（パラメータ）に関する尤度 $g(\hat{\beta})$ の比を最大化する考え方が示されている．2 群の群間比較から得られる経時データに対し Excel の行列関数とソルバーのみで REML 法の解析が手軽に行なえることを示し，ブラック･ボックス的である REML 法の可視化を試みた．

経時データ　2 群 2 時点，各群 2 例，計 4 例について t=0 を群間で共通とするモデルとする．

投与前値を2群で共通とする線形混合モデル

n= 8　　パラメータ μ に関する尤度の計算シート

p= 3

		推定値 $\hat{\beta}-\beta$		分散 V	
$\hat{\beta}_0$=	10.00	10.0	0.0	0.76	0.75	0.75
$\hat{\beta}_1$=	9.00	9.0	0.0	0.75	0.79	0.74
$\hat{\beta}_2$=	11.00	11.0	0.0	0.75	0.74	0.79
$\hat{\sigma}_A^2$=	3.00					
$\hat{\sigma}_e^2$=	0.05					

$\ln g(\hat{\beta})$ =	0.38
$\ln f(y)$ =	-8.80
$\ln L$	-9.19
$-2\ln L$=	18.37

G

3.00	0.00	0.00	0.00
0.00	3.00	0.00	0.00
0.00	0.00	3.00	0.00
0.00	0.00	0.00	3.00

R

1	0.05	0.00	0.00	0.00	0.00	0.00	0.00	0.00
2	0.00	0.05	0.00	0.00	0.00	0.00	0.00	0.00
3	0.00	0.00	0.05	0.00	0.00	0.00	0.00	0.00
4	0.00	0.00	0.00	0.05	0.00	0.00	0.00	0.00
5	0.00	0.00	0.00	0.00	0.05	0.00	0.00	0.00
6	0.00	0.00	0.00	0.00	0.00	0.05	0.00	0.00
7	0.00	0.00	0.00	0.00	0.00	0.00	0.05	0.00
8	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.05

i	g	id	t	y	$y^\wedge$	y-$y^\wedge$	X x_0	x_1	x_2	Z a_1	a_2	a_3	a_4
1	1	1	0	11.00	10.00	1.00	1	0	0	1	0	0	0
2	1	1	1	10.00	9.00	1.00	0	1	0	1	0	0	0
4	1	2	0	12.00	10.00	2.00	1	0	0	0	1	0	0
4	1	2	1	10.50	9.00	1.50	0	1	0	0	1	0	0
4	2	3	0	9.00	10.00	-1.00	1	0	0	0	0	1	0
6	2	3	1	10.00	11.00	-1.00	0	0	1	0	0	1	0
7	2	4	0	8.00	10.00	-2.00	1	0	0	0	0	0	1
8	2	4	1	9.50	11.00	-1.50	0	0	1	0	0	0	1

H=ZGZ'+R

	1	2	3	4	5	6	7	8
1	3.05	3.00	0.00	0.00	0.00	0.00	0.00	0.00
2	3.00	3.05	0.00	0.00	0.00	0.00	0.00	0.00
3	0.00	0.00	3.05	3.00	0.00	0.00	0.00	0.00
4	0.00	0.00	3.00	3.05	0.00	0.00	0.00	0.00
5	0.00	0.00	0.00	0.00	3.05	3.00	0.00	0.00
6	0.00	0.00	0.00	0.00	3.00	3.05	0.00	0.00
7	0.00	0.00	0.00	0.00	0.00	0.00	3.05	3.00
8	0.00	0.00	0.00	0.00	0.00	0.00	3.00	3.05

解析手順　固定効果に関する計画行列を X，仮の推定値を $\hat{\boldsymbol{\beta}}$ として，残差 $\boldsymbol{y}-\boldsymbol{X}\hat{\boldsymbol{\beta}}$ を計算する．変量効果（症例）に関する計画行列を Z，変量効果の分散の仮の推定値（$\hat{\sigma}_A^2$=3.0）を G の対角要素へ代入，誤差分散の仮の推定値（$\hat{\sigma}_e^2$=00.5）を R の対角要素へ代入し，残差 $\boldsymbol{y}-\boldsymbol{X}\hat{\boldsymbol{\beta}}$ に関する分散共分散行列 $\boldsymbol{H}=\boldsymbol{ZGZ}'+\boldsymbol{R}$ を計算する．データに関する対数尤度 $\ln f(y)$ は，$\boldsymbol{y}-\boldsymbol{X}\hat{\boldsymbol{\beta}}$ が分散共分散 H となる n=8 次元正規分布の確率密度で -8.80 と計算されている．パラメータに関する対数尤度 $\ln g(\hat{\boldsymbol{\beta}})$ は，$\hat{\boldsymbol{\beta}}$ に関する分散共分散行列を $\boldsymbol{V}=(\boldsymbol{X}'\boldsymbol{H}^{-1}\boldsymbol{X})^{-1}$ とし，$\hat{\boldsymbol{\beta}}-\boldsymbol{\beta}$ の 3 次元正規分布の確率密度で 0.38 と計算されている．REML の対数尤度は $\ln(L)=\ln\{f(\boldsymbol{y})/g(\hat{\boldsymbol{\beta}})\}=-9.19$ と計算されている．Excel のソルバーで $\ln(L)$ が最大になるように $\hat{\beta}_0$ から $\hat{\sigma}_e^2$ まで 5 変数を同時に変化させると REML 法での尤度 $\ln(L)$ は，-9.19 から -8.13 と最大になり，$\hat{\beta}_0=10.0$，$\hat{\beta}_1=8.78$，$\hat{\beta}_2=11.22$，$\hat{\sigma}_A^2=3.11$，$\hat{\sigma}_e^2=0.06$ が求まる．計算時間は標準的な PC で 1 秒程度であった．なお，この結果は SAS の MIXED モデルの結果と一致した．

文献：　1) Diggle P J., et al. (2002), Analysis of Longitudinal Data 2ed., Oxford Press.
2) 芳賀敏郎（2007），Excel-ソルバーによる非線形最小 2 乗法・最尤法・ロバスト回帰，http://www.med.kurume-u.ac.jp/med/biostat/forum/haga.pdf

14.8. 統計連合 2011 での発表 スライド

Excel のソルバーを用いた 制限付き最尤法（REML法）による 経時測定データの解析

高橋 行雄
BioStat研究所（株）

対象とする経時測定データ

- 2群間あるいは多群間比較
- 測定時点はあらかじめ定められている
- 時点間に正の相関
- 線形混合効果モデル
  - 群と時期を固定効果
  - 症例を変量効果

線形混合効果モデル

- 多くの統計ソフトで制限付き最尤法（REML法）による解析が可能
- REML法の理論的な背景は多くの書物に
- ガウス・ニュートン法など反復計算による最適化が必須
- 簡単なデータに対しても追試することが困難
- 統計ソフトをブラックボックス的に使わざるをえなくなる

経時測定データの解析

- 制限付き最尤法
  - 最尤法によるデータ解析
  - どのような“制限”なのかをはっきりさせる
  - その上で制限付き最尤法によるデータ解析を実施する
- Excel による制限付き最尤法による解析
  - 逆行列，行列式などの関数
  - Excelのソルバーは，ガウス・ニュートン法による最適化ソフト
  - 制限付き最尤法の実施可能がある

2群4症例，2時点のデータ

2 群 4 症例，2 時点のデータ

群	症例	前値	後値
1	1	11.0	10.0
	2	12.0	10.5
2	3	9.0	10.0
	4	8.0	9.5

固定効果 および 変量効果

- 解析モデル $y = X\beta + Z\gamma + \varepsilon$

			データ y	固定効果 群x時点 X			変量効果 症例 Z			
群	症例	時点	y	x_0	x_1	x_2	z_1	z_2	z_3	z_4
1	1	0	11.0	1	0	0	1	0	0	0
1	1	1	10.0	0	1	0	1	0	0	0
1	2	0	12.0	1	0	0	0	1	0	0
1	2	1	10.5	0	1	0	0	1	0	0
2	3	0	9.0	1	0	0	0	0	1	0
2	3	1	10.0	0	0	1	0	0	1	0
2	4	0	8.0	1	0	0	0	0	0	1
2	4	1	9.5	0	0	1	0	0	0	1

データに関する尤度関数

$(y-\hat{y})$ が 期待値 0，分散共分散行列 H

データ y に関する対数尤度

n 次元正規分布の確率密度

$$f(y)=(2\pi)^{-\frac{1}{2}n}|H|^{-\frac{1}{2}}\exp\left\{-\frac{1}{2}(y-\hat{y})'H^{-1}(y-\hat{y})\right\}$$

尤度 $f(y)$ を最大化（この場合は最尤法となる）

仮の固定効果

固定効果に対する適当な初期値

固定効果の推定値を $\hat{\beta}$ として $\hat{y}=X\hat{\beta}$

$\hat{\beta}_0=$ 10.00
$\hat{\beta}_1=$ 9.00
$\hat{\beta}_2=$ 11.00

y	y^	y-y^
11.00	10.00	1.00
10.00	9.00	1.00
12.00	10.00	2.00
10.50	9.00	1.50
9.00	10.00	−1.00
10.00	11.00	−1.00
8.00	10.00	−2.00
9.50	11.00	−1.50

変量効果に関する分散共分散

症例についての分散共分散行列 G

行列 G の対角要素は症例間の仮の分散 $\hat{\sigma}_A^2$

共分散は，症例が互いに独立なので 0

$\hat{\sigma}_A^2=$ 3.00
$\hat{\sigma}_e^2=$ 0.05

G

3.00	0.00	0.00	0.00
0.00	3.00	0.00	0.00
0.00	0.00	3.00	0.00
0.00	0.00	0.00	3.00

誤差分散

症例ごと時点ごとの分散共分散行列 R

行列 R の対角要素に仮の分散 $\hat{\sigma}_e^2$，共分散は全て 0

$\hat{\sigma}_A^2=$ 3.00
$\hat{\sigma}_e^2=$ 0.05

R

1	0.05	0.00	0.00	0.00	0.00	0.00	0.00	0.00
2	0.00	0.05	0.00	0.00	0.00	0.00	0.00	0.00
3	0.00	0.00	0.05	0.00	0.00	0.00	0.00	0.00
4	0.00	0.00	0.00	0.05	0.00	0.00	0.00	0.00
5	0.00	0.00	0.00	0.00	0.05	0.00	0.00	0.00
6	0.00	0.00	0.00	0.00	0.00	0.05	0.00	0.00
7	0.00	0.00	0.00	0.00	0.00	0.00	0.05	0.00
8	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.05

データにつての分散共分散

症例・時点ついて分散共分散行列 H

$H=ZGZ'+R$

行列 H は，各症例内では Compound Symmetry（対称型）

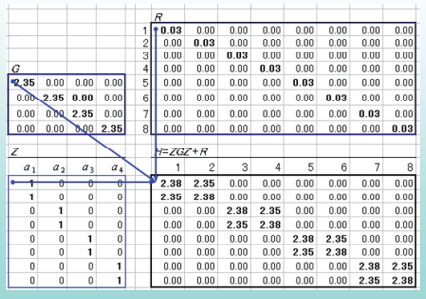
$$\begin{bmatrix} \begin{matrix}\sigma_A^2+\sigma_e^2 & \sigma_A^2 \\ \sigma_A^2 & \sigma_A^2+\sigma_e^2\end{matrix} & & 0 \\ & \ddots & \\ 0 & & \begin{matrix}\sigma_A^2+\sigma_e^2 & \sigma_A^2 \\ \sigma_A^2 & \sigma_A^2+\sigma_e^2\end{matrix} \end{bmatrix}$$

分散共分散行列の生成法

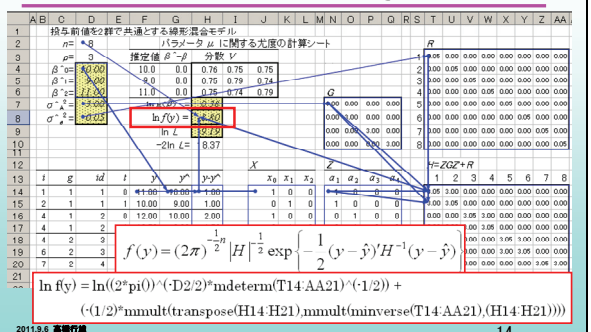
- モデルに含まれる変量効果の分散を行列 G の対角要素に
- 変量効果についてインジケータ型の計画行列 Z とする
- 誤差分散を R とする（任意の時点間の相関構造を持たせても良い）.
- 分散共分散行列 H は

$$H=ZGZ'+R$$

$H = Z\,G\,Z' + R$



データに関する尤度 ln $f(y)$ の最大化



最尤解

ln $f(y)$ を最大化するように

$\hat{\beta}_0$=	10.00
$\hat{\beta}_1$=	9.00
$\hat{\beta}_2$=	11.00
$\hat{\sigma}_A^2$ =	3.00
$\hat{\sigma}_e^2$ =	0.05

を変化させると最尤解が得られる

ln $f(y)$ = -7.54

n=	8
$\hat{\beta}_0$=	10.00
$\hat{\beta}_1$=	8.77
$\hat{\beta}_2$=	11.23
$\hat{\sigma}_A^2$ =	2.35
$\hat{\sigma}_e^2$ =	0.03

固定効果に関する尤度

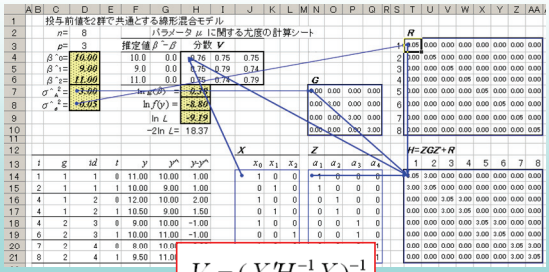
推定値 $\hat{\beta}$ についての尤度 $g(\hat{\beta})$

$(\hat{\beta}-\beta)$ が 期待値 0, 分散共分散 $V=(X'H^{-1}X)^{-1}$

p 次元正規分布に従う

$$g(\hat{\beta}) = (2\pi)^{-\frac{1}{2}p} |V|^{-\frac{1}{2}} \exp\left\{-\frac{1}{2}(\hat{\beta}-\beta)V(\hat{\beta}-\beta)\right\}$$

固定効果の分散共分散行列

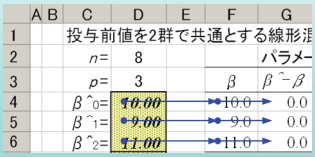


$$V=(X'H^{-1}X)^{-1}$$

固定効果の推定値

制限付き最尤法では，固定効果 β は，$f(y)$ の尤度を最大化したときの最尤解 $\hat{\beta}$ と同じと仮定

$\beta = \hat{\beta}$



固定効果に関する尤度 $g(\hat{\beta})$

$p=$	3	β	$\hat{\beta}-\beta$	分散 V		
$\hat{\beta}_0=$	10.00	10.0	0.0	0.76	0.75	0.75
$\hat{\beta}_1=$	9.00	9.0	0.0	0.75	0.79	0.74
$\hat{\beta}_2=$	11.00	11.0	0.0	0.75	0.74	0.79
$\hat{\sigma}^2_A=$	3.00	$\ln g(\hat{\beta})$ =		0.38		

$$g(\hat{\beta}) = (2\pi)^{-\frac{1}{2}p}|V|^{-\frac{1}{2}}\exp\left\{-\frac{1}{2}(\hat{\beta}-\beta)V(\hat{\beta}-\beta)\right\}$$

$$= (2\pi)^{-\frac{1}{2}p}|V|^{\frac{1}{2}}$$

3次元正規分布の確率密度

REML法での尤度

データ y についての尤度 $f(y)$

固定効果 β についての尤度 $g(\hat{\beta})$

それらの比を最大化

$$L = \frac{f(y)}{g(\hat{\beta})}$$

適当な推定値での尤度

◆ 対数尤度 L は -9.19 と計算されている

	A B	C	D	E	F	G	H	I	J
1		投与前値を2群で共通とする線形混合モデル							
2		$n=$	8			パラメータ μ に関する尤度の			
3		$p=$	3		β	$\hat{\beta}-\beta$	分散 V		
4		$\hat{\beta}_0=$	10.00		10.0	0.0	0.76	0.75	0.75
5		$\hat{\beta}_1=$	9.00		9.0	0.0	0.75	0.79	0.74
6		$\hat{\beta}_2=$	11.00		11.0	0.0	0.75	0.74	0.79
7		$\hat{\sigma}^2_A=$	3.00			$\ln g(\hat{\beta})$ =	0.38		
8		$\hat{\sigma}^2_e=$	0.05			$\ln f(y)$ =	-8.80		
9						$\ln L$	-9.19		
10						$-2\ln L$ =	18.37		

対数尤度 L の最大化

◆ 対数尤度 L は -8.13 と大きくなっている

	A B	C	D	E	F	G	H	I	J
1		投与前値を2群で共通とする線形混合モデル							
2		$n=$	8			パラメータ μ に関する尤度の			
3		$p=$	3		β	$\hat{\beta}-\beta$	分散 V		
4		$\hat{\beta}_0=$	10.00		10.0	0.0	0.79	0.78	0.78
5		$\hat{\beta}_1=$	8.78		8.8	0.0	0.78	0.82	0.76
6		$\hat{\beta}_2=$	11.22		11.2	0.0	0.78	0.76	0.82
7		$\hat{\sigma}^2_A=$	3.11			$\ln g(\hat{\beta})$ =	0.13		
8		$\hat{\sigma}^2_e=$	0.06			$\ln f(y)$ =	-8.00		
9						$\ln L$	-8.13		
10						$-2\ln L$ =	16.26		

統計ソフトによる検証

統計ソフト JMP の結果

項	推定値	標準誤差
x0	10.0000	0.8902
x1	8.7798	0.9082
x2	11.2202	0.9082

変量効果	分散比	分散成分
id	49.256756	3.1069723
残差		0.0630771
合計		3.1700494

−2対数尤度= 16.262716836

Excel の結果

$\hat{\beta}_0=$	10.00
$\hat{\beta}_1=$	8.78
$\hat{\beta}_2=$	11.22

$\hat{\sigma}^2_A=$	3.107
$\hat{\sigma}^2_e=$	0.063

$\ln L$	-8.1314
$-2\ln L$ =	16.2627

変量効果の推定

症例についての変量効果 γ の推定

$$\hat{\gamma} = \hat{G}Z\hat{H}^{-1}(y-\hat{y})$$

◆ JMPによる結果

項	BLUP	標準誤差
id[1]	1.099	0.899
id[2]	1.841	0.899
id[3]	−1.099	0.899
id[4]	−1.841	0.899

Excelによる計算

id	
1	1.099
2	1.842
3	−1.099
4	−1.841

まとめ

- 一般的な統計ソフトでは線形混合モデルの解析に制限付き最尤法が用いられているが，その計算過程はブラックボックス的であるので，現代の算盤ともいえるExcel を用いて追試を行なった．
- 制限付き最尤法の計算過程の可視化は，線形混合モデルの適切な活用に寄与する．

15. 今後の課題

古典的な様々な実験計画に対して，多くの統計ソフトは，たとえば 1 因子実験，2 因子実験，L_{16} 直交表実験など手法ごとにプログラム化している．SAS はこのような対応はせずに，一般線形モデルをベースにした GLM プロシジャのみで古典的なさまざまな実験計画の解析ができるだけでなく，欠測値を含むデータの解析が自在にできるようになったのは画期的であった．

しかしながら，実験データの解析に直面している実務家が，汎用的な SAS の GLM プロシジャを使いこなし，成果を挙げられるようになるためには，相当な努力を必要とする．そのようなこともあり，実務家を対象にした実験計画法関連のセミナーなどで，SAS が使われることは見い出せない．

JMP の「モデルのあてはめ」も SAS の GLM プロシジャと同様なコンセプトであり，さらに，優れた GUI（グラフィカル・ユーザ・インターフェイス）のみならず，生データのグラフ表示，解析結果のグラフ表示などに傑出している．

そのために，実務家を対象にした実験計画法・品質管理関連の研修などで，JMP を用いることにより研修の成果を短時間で上げられるようになる．2010 年 5 月 26 日の JMPer's で，「中外製薬における統計教育の新たな展開～ 品質管理教育における JMP の活用～」と題して，次のような実践教育の経験を発表した．

「中外製薬でエンドユーザが用いる標準的な統計ソフトとして JMP のライセンス数が，350 を超えた．多くの部門のメンバーにさまざまなデータ解析の研修を繰り返し行なってきた結果が実ったと感慨深い．ここ 3 年間，品質管理のための統計教育を毎年 4 ヶ所の工場から 3～4 人の希望者を募って継続的に行なっている．どのようなデータでも JMP・Excel でグラフを作成することが何らかの発見，新たな興味につながることを繰り返し体験させている．関係者に対する説得力の向上のために，数理統計の知識を基盤にした応用力を身に付けさせることもこの教育の目的である．今年の統計教育の目標設定は「目指せ：品質管理（QC）検定 1 級」である．これらの教育の経験に基づいて，品質管理のための統計教育のあり方について考察したい．

http://www.jmp.com/japan/about/events/jmpers20100526.shtml

日本における実験計画法の教育・様々なセミナーなどで取り上げられているのは，手計算の時代に開発された方法が主体である．このような事例にたいして，JMP の「モデルのあてはめ」の使い方を説明することから始めることも必要ではあるが，本書は，JMP の「モデルのあてはめ」が，他の統計ソフトにない画期的な機能を用いたアドバンスコースの教材を目指した．

繰返しを意図的に不揃いにしたスタッカード型の枝分かれ実験は，実験数を大幅に減少し分析コストを下げても，解析の精度を保つ方法として知られている．JMP は，SAS の GLM プロシジャ，あるいは，VARCOMP プロシジャと同様にスタッカード型の枝分かれ実験の解析ができることを示した．

枝分かれ実験は, 分析法バリデーションでの必須の統計手法であるが, SAS を用いたスタッカード型の枝分かれ実験の事例を日本語の Web 上で見出すことはできなかった．また，日本語で書かれた統計の教科書にもスタッカード型の枝分かれ実験の例示は見いだせなかった．これは，手計算でも追試できることを基本としているためと推測される．

日本では，SAS が一部の統計専門家たちの使う道具として定着してしまい，枝分かれ実験を計画し，解析する人達の現場の道具としてはまったく支持されていないと思われる．SAS を日常的に使うことができる環境であれば，スタッカード型だけではなく，必要最小限の計画を立てて実験コストの削減を図れれば，統計ソフトのコストなどすぐに回収できると思われる．

手軽な計算手段として Excel が普及し，このことが統計ソフトの普及を阻害しているかのようである．そして，Excel でできる範囲内の統計計算で問題解決を図ろうとし，高価な統計ソフトなど使わなくてよいと思いんでしまう．日本語で書かれた統計の教科書にも事例が示されていないような実験を計画し，統計ソフトでの解決を求めるのは統計の非専門家には酷な要求である．

SAS の MIXED プロシジャは，GLM プロシジャでは便宜的な対応をしていた複数の誤差構造をもつ種々の実験計画の問題に対して，根本的な解決ができるようになった．しかしながら，その解法に制限付き最尤法(REML 法)が用いられていて，これが難解なために MIXED プロシジャは統計の専門家が使うものとの認識が蔓延し，多くの身近にある統計的問題に対しする適用事例が出てこない．事例がないと使う動機が起こるはずがない．

SAS に遅れること 10 年，ようやく JMP で複数の誤差構造をもつ種々の実験計画の解析ができるようになった．ただし，実際の計算の根本は，SAS と同じであり，統計の専門家でも難解である．そのために，現実に良くある実験計画ではあるが，従来の解析法では便宜的な対応を迫られるような事例を主体にした．

JMP を使うことだけを前提にすると統計ソフトの「ブラックボックス」使用となり，表層的なテキストになりかねない．きちっと理論を理解し，小さな典型的な事例に対して，手計算レベルでも追試ができることは望ましいのでるが, JMP 自体は行列計算を前提にしている．そのために,「手計算」レベルでななく，Excel の「行列計算・ソルバー」レベルで理論の理解ができるようになることを目指した．

従来の実験計画法は，最小 2 乗法を基盤とする「平方和の分解」で構築されいた．JMP では，最小 2 乗法はもちろんのこと，制限付き最尤法が標準機能をして使われ，もはや「平方和の分解」での説明は不可能となっている．

幸い Excel には必要最小限の行列関数，多くの統計的諸問題を解くことができるソルバーが備わっている．これらを活用して，線形混合モデルの最も基本である完備型の枝分かれ実験データに対して，Excel による REML 法の基本的な考え方，計算方法を示すことにした．さらに，スタッカード型の枝分かれ実験データに対しても応用できることを示した．

これらの簡単な典型的なデータに対して，読者が自ら，追試できるような解析手順を丁寧に示した．追試をすることは，新しい方法の本質的な理解の助けになり，多くの便宜的な対応でしのいでいる統計的な諸問題に対して，自信をもって新しい統計手法を活用ができるようになる．

新しい流れ

新しい流れを 2010 年 10 月 28 日（金）の JMPer’s で感じた．これは,「東芝グループにおける JMP の活用概況」について赤間氏の講演である．抄録を次に示す．

「東芝グループでは MI（Management Innovation）活動を基軸にプロセス／バリューイノベーションを実行し、利益ある持続的成長をスローガンとして邁進しています。社員全員が受講する MI 教育の中にはシックスシグマも導入され、統計の基礎と応用を学習しています。この

ような背景の中、最近では、課題解決のための統計解析に JMP を有効活用するユーザーが増えており、コーポレートとしても高度統計ツールとして推奨しています。今回は、このような状況となる過程を通して、JMP の有効活用例を紹介させていただきます。」

中断した勉強会で，日本ゴアテックスの山本氏がまとめた，新しい実験計画法の研修（案）を示す．今後，新たなセミナーを企画する際のたたき台として活用したい．

目次

1．バランスのとれた連続量二因子3水準の実験

1－1．繰り返しのある総当たり実験

1－2．繰り返しのない総当たり実験

		B		
		B1	B2	B3
A	A1	○	○	○
		○	○	○
	A2	○	○	○
		○	○	○
	A3	○	○	○
		○	○	○

1－1．繰り返しのある総当たり実験

		B		
		B1	B2	B3
A	A1	○	○	○
	A2	○	○	○
	A3	○	○	○

1－2．繰り返しのない総当たり実験

繰り返しがある場合とない場合の違いについて理解する．

2．バランスのとれた連続量三因子3水準の実験

2－1．繰り返しのある総当たり実験

2－2．繰り返しのない総当たり実験

2－3．一部実施実験（応答曲面法）

B→		B1			B2			B3		
C→		C1	C2	C3	C1	C2	C3	C1	C2	C3
A	A1	○	○	○	○	○	○	○	○	○
		○	○	○	○	○	○	○	○	○
	A2	○	○	○	○	○	○	○	○	○
		○	○	○	○	○	○	○	○	○
	A3	○	○	○	○	○	○	○	○	○
		○	○	○	○	○	○	○	○	○

B→		B1			B2			B3		
C→		C1	C2	C3	C1	C2	C3	C1	C2	C3
A	A1	○	○	○	○	○	○	○	○	○
	A2	○	○	○	○	○	○	○	○	○
	A3	○	○	○	○	○	○	○	○	○

2－1. 繰り返しのある総当たり実験

B→		B1			B2			B3		
C→		C1	C2	C3	C1	C2	C3	C1	C2	C3
A	A1	○		○		○		○		○
	A2		○		○	◎	○		○	
	A3	○		○		○		○		○

2－2. 繰り返しのない総当たり実験

B→		B1			B2			B3		
C→		C1	C2	C3	C1	C2	C3	C1	C2	C3
A	A1		○		○		○		○	
	A2	○		○		◎		○		○
	A3		○		○		○		○	

2－3（1）. 一部実施実験（応答曲面法=CCD 法）

2－3（2）一部実施実験（応答曲面法=BH 法）

繰り返しがある場合とない場合，および一部実施実験の違いについて理解する.
応答曲面法の考え方を理解する.

3. アンバランスな連続量二因子3水準の実験

3－1. 繰り返し数が不揃いな場合

3－2. 欠測組み合わせがある場合

		B		
		B1	B2	B3
A	A1	○	○	○
			○	○
	A2	○	○	○
		○	○	
	A3	○	○	○
		○		○

3－1. 繰り返し数が不揃いな場合

		B		
		B1	B2	B3
A	A1	○	○	○
		○	○	○
	A2	○	○	○
		○	○	○
	A3	○	○	
		○	○	

3－2. 欠測組み合わせがある場合

データがアンバランスな場合の解析について理解する.
繰り返し数が不揃いな場合と欠測組み合わせがある場合の違いを理解する.

4. ブロック化した連続量二因子3水準の実験

4－1. バランスがとれたブロック化実験

4－2. アンバランスなブロック化実験

		B		
ブロック	A	B1	B2	B3
1	A1	○	○	○
2	A1	○	○	○
3	A2	○	○	○
4	A2	○	○	○
5	A3	○	○	○
6	A3	○	○	○

		B		
ブロック	A	B1	B2	B3
1	A1	○		○
2	A1	○	○	
3	A2		○	○
4	A2	○	○	○
5	A3	○		○
6	A3	○	○	○

4－1．バランスがとれた（完備型）ブロック化実験　　**4－2．アンバランスな（不完備型）ブロック化実験**

ブロック化構造の影響と解析方法について理解する．
枝分かれと変量効果の意味について理解する．

5．ブロック化した連続量三因子3水準の実験

5－1．バランスがとれたブロック化実験

5－2．アンバランスなブロック化実験

	B	B1			B2			B3		
ブロック	A　C	C1	C2	C3	C1	C2	C3	C1	C2	C3
1	A1	○	○	○	○	○	○	○	○	○
2	A1	○	○	○	○	○	○	○	○	○
3	A2	○	○	○	○	○	○	○	○	○
4	A2	○	○	○	○	○	○	○	○	○
5	A3	○	○	○	○	○	○	○	○	○
6	A3	○	○	○	○	○	○	○	○	○

5－1．バランスがとれた（完備型）ブロック化実験

	B	B1			B2			B3		
ブロック	A　C	C1	C2	C3	C1	C2	C3	C1	C2	C3
1	A1	○	○	○	○	○	○	○	○	○
2	A1	○		○	○	○	○	○	○	○
3	A2	○	○	○		○	○	○	○	○
4	A2	○	○	○	○	○	○	○	○	○
5	A3	○	○	○	○	○	○	○	○	
6	A3	○	○	○	○	○	○	○	○	○

5－2．アンバランスな（不完備型）ブロック化実験

5章は4章の変数を増やしただけなのでいらない？
1～4章が理解できているかどうかの練習問題でもいいかもしれない．

補足：各実験の実験計画法

1～5章の各実験計画をJMPで立案する方法について理解する．