

mixed - 混合効果線形回帰 【 評価版 】

mixed コマンドは線形混合効果モデル (linear mixed-effects models) をフィットさせる機能を提供します。

| | |
|-----------------|------------|
| 1. 線形混合効果モデル | |
| 2. 2 階層モデル | Example 1 |
| | Example 2 |
| 3. 共分散構造 | Example 3 |
| 4. REML 推定法 | |
| 5. 3 階層モデル | Example 4 |
| 6. ブロック対角な共分散構造 | Example 5 |
| 7. 分散不均一な変量効果 | Example 6 |
| 8. 分散不均一な残差誤差 | Example 7 |
| 9. その他の残差誤差構造 | Example 8 |
| | Example 9 |
| 10. 交差効果モデル | Example 10 |
| | Example 11 |
| 11. 収束上の問題 | |
| 12. サーベイデータ | Example 12 |
| 13. 小標本時の推定 | Example 13 |
| | Example 14 |
| 補足 1 | |

1. 線形混合効果モデル

混合効果モデル (mixed models) とは固定効果 (fixed effects) と変量効果 (random effects) の双方を含んだモデルのことを言います。線形混合効果モデルの定義式は

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{Z}\mathbf{u} + \boldsymbol{\epsilon} \quad (1)$$

のように表現されます。一般的な線形回帰モデルに比べると、全般的な誤差項 $\boldsymbol{\epsilon}$ の他にランダムな変動 (効果) を表す $\mathbf{Z}\mathbf{u}$ という項が加わっている点に特徴があります。ここに

- ・ \mathbf{y} は $n \times 1$ ベクトル
- ・ \mathbf{X} は固定効果 β に対する $n \times p$ のデザイン/共変量行列
- ・ \mathbf{Z} は変量効果 \mathbf{u} に対する $n \times q$ のデザイン/共変量行列

です。また、 $n \times 1$ の誤差ベクトル ϵ は平均値が 0 で分散行列が $\sigma_\epsilon^2 \mathbf{R}$ であるような多変量正規分布に従うものとしします。

(1) 式の固定効果部分 $\mathbf{X}\beta$ は標準的な OLS 回帰における線形予測項に該当するものです。これに対し (1) 式の変量効果部分 $\mathbf{Z}\mathbf{u} + \epsilon$ に対しては、 \mathbf{u} が分散共分散行列 \mathbf{G} を有し、それは ϵ と直交関係にあること、すなわち

$$\text{Var} \begin{bmatrix} \mathbf{u} \\ \epsilon \end{bmatrix} = \begin{bmatrix} \mathbf{G} & \mathbf{0} \\ \mathbf{0} & \sigma_\epsilon^2 \mathbf{R} \end{bmatrix}$$

であることを仮定します。変量効果 \mathbf{u} は予測はできても直接推定されるわけではありません。ただしそれは分散成分 (variance components) として知られる \mathbf{G} の要素によって特徴付けられるわけですが、それらは一般的な残差分散 σ_ϵ^2 と \mathbf{R} 内に含まれる残差分散パラメータと共に推定されます。

デザイン行列 \mathbf{X} と \mathbf{Z} の形式によってさまざまなデザインに対応した線形モデルが構成されます。それらはまたクラスター内相関をモデリングする上での種々の手法を提供するものでもあります。すなわちランダム切片やランダム傾きを共用することによってクラスター内相関を表現することができます。さらに \mathbf{G} や \mathbf{R} の規定のしかたによってもさまざまな柔軟性がもたらされます。

混合モデルをフィットさせる上でキーとなるのは分散成分の推定であり、種々の手法が提唱されてきました。そのうちで最もポピュラーなものが最尤法 (ML: maximum likelihood) と制限付き最尤法 (REML: restricted maximum-likelihood) によるもので、その双方が mixed ではサポートされています。

n 個の観測データが M 個の独立なグループ (クラスター) から構成される場合には、(1) 式は

$$\mathbf{y}_j = \mathbf{X}_j \beta + \mathbf{Z}_j \mathbf{u}_j + \epsilon_j, \quad j = 1, \dots, M \quad (2)$$

のように表現できます。この場合、 \mathbf{y}_j はクラスター j に対応する \mathbf{y} の行からなるものとしします。 \mathbf{X}_j, ϵ_j についても同様に定義されます。変量効果 \mathbf{u}_j については平均が 0、分散行列が Σ の正規分布に従う $q \times 1$ ベクトルの M 個の実現値とみなすことができます。行列 \mathbf{Z}_j はクラスター j の変量効果に対応した $n_j \times q$ のデザイン行列です (クラスター j に含まれる観測データの数を n_j とする)。これらを (1) 式と対応付けると次のようになります。

$$\mathbf{Z} = \begin{bmatrix} \mathbf{Z}_1 & \mathbf{0} & \cdots & \mathbf{0} \\ \mathbf{0} & \mathbf{Z}_2 & \cdots & \mathbf{0} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{0} & \mathbf{0} & \cdots & \mathbf{Z}_M \end{bmatrix}; \quad \mathbf{u} = \begin{bmatrix} \mathbf{u}_1 \\ \vdots \\ \mathbf{u}_M \end{bmatrix}; \quad \mathbf{G} = \mathbf{I}_M \otimes \Sigma; \quad \mathbf{R} = \mathbf{I}_M \otimes \Lambda \quad (3)$$

ただし Λ はレベル 1 残差の分散行列であり、 \otimes は Kronecker 積を意味します。

モデル式 (2) は Laird and Ware (1982) に由来するものですが、それは次のような 2 つの利点を持ちます。第 1 に変量効果の規定をより容易なものとしします。例えばクラスターが学校を意味するのであれば、学校のレベルにおける変量効果を考えれば良いことになるわけです。第 2 にそれは変量効果の多段化を可能にします。例えば学校内にネストされたクラスを考えた場合、学校レベルでの変量効果とクラスレベルでの変量効果を表現できるモデルへの拡張がより容易に行えるようになります。

このような多階層モデルの扱いは後述するとして、(2) 式で表現されるモデルを 2 階層モデル (two-level model) と呼ぶことにします。クラス j 内の個体 i に対応する観測データを y_{ij} とするとき、個体はモデルの第 1 レベルを構成し、クラスは第 2 レベルを構成することになります。学校内にクラスがネストされた 3 階層モデルの場合には、学生が第 1 レベルを、クラスが第 2 レベルを、学校が第 3 レベルを構成することになります。

後続のセクションのうち前半においては残差は独立で分散が均一である場合を想定します。すなわち (3) 式において $\Lambda = \mathbf{I}$ (単位行列) であるとし、単一の残差分散値 σ_e^2 を推定するものとします。

2. 2 階層モデル

最初に (2) 式の簡単な用例から見て行くことにします。1 階層モデルをスキップする理由はそれが単なる標準的な OLS 回帰と変わらないからです。

▷ Example 1

ここでは Example データセット pig.dta を用いた用例を紹介します。

```
. use https://www.stata-press.com/data/r18/pig.dta *1
(Longitudinal analysis of pig weights)
```

このデータセット中には 48 頭の豚の体重が 9 週間にわたって記録されています。参考までに先頭の豚についてのデータをリスト出力しておきます。

```
. list id week weight if id == 1 *2
```

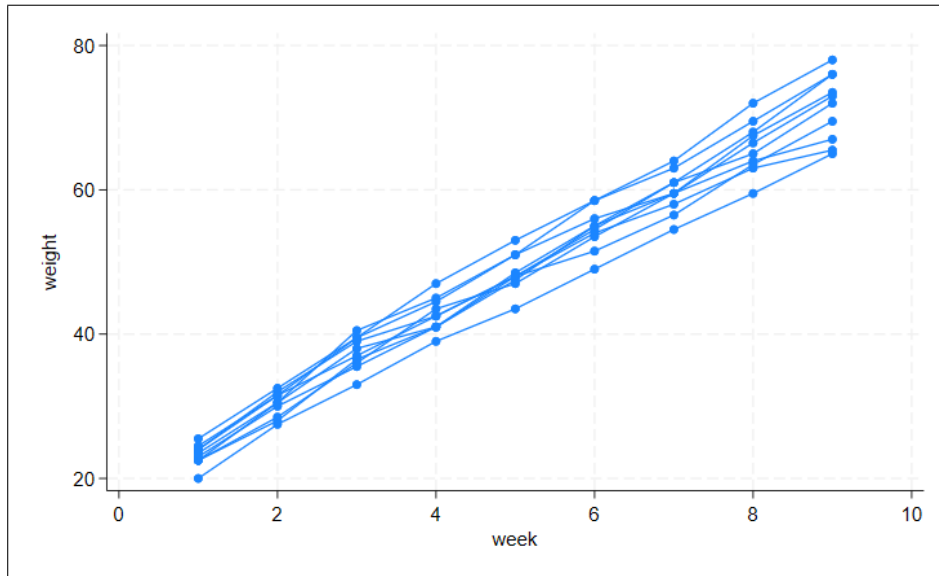
| | id | week | weight |
|----|----|------|--------|
| 1. | 1 | 1 | 24 |
| 2. | 1 | 2 | 32 |
| 3. | 1 | 3 | 39 |
| 4. | 1 | 4 | 42.5 |
| 5. | 1 | 5 | 48 |
| 6. | 1 | 6 | 54.5 |
| 7. | 1 | 7 | 61 |
| 8. | 1 | 8 | 65 |
| 9. | 1 | 9 | 72 |

*1 メニュー操作 : File ▷ Example Datasets ▷ Stata 18 manual datasets と操作、Multilevel Mixed-Effects Reference Manual [ME] の mixed の項よりダウンロードする。

*2 メニュー操作 : Data ▷ Describe data ▷ List data

id = 1 から 10 までの範囲の豚についてその体重の変化をグラフ化してみると次のようになります。

```
. twoway connected weight week if id<=10, connect(L) *3
```



それぞれの豚の体重の変化は線形であり、また個体差もそれなりにあることがわかります。ところで我々は pig.dta に含まれている特定の 48 頭の豚について関心があるわけではない点に注意してください。すなわちそれら 48 頭はより大きな母集団からランダムに抽出された標本に過ぎないととらえ、個体間の変動を変量効果としてモデル化することにします。具体的にはモデル式 (2) における変量効果項 $Z_j u_j$ として定数項、すなわちランダム切片 (random-intercept) 項のみを想定するわけです。このときのモデル式は次のようになります。

$$\text{weight}_{ij} = \beta_0 + \beta_1 \text{week}_{ij} + u_j + \epsilon_{ij}, \quad i = 1, \dots, 9; j = 1, \dots, 48 \quad (4)$$

2 階層モデルの場合、第 2 レベルを表す j がクラスタに対応するため、豚を識別する変数 id に対応し、第 1 レベルを表す i は変数 week に対応する点に注意してください。

このモデル式における固定効果部 $\beta_0 + \beta_1 \text{week}_{ij}$ は母集団平均に対応した 1 本の回帰直線を表します。これに対し変量効果項 u_j はその回帰直線を平行に上下させる個体ごとの効果を表すこととなります。このようなモデルをフィットさせるには、クラスタを識別する j が変数 id に対応する点を考慮し、次のように操作します。

*3 メニュー操作 : Graphics ▷ Tway graph (scatter, line, etc.)

- Statistics ▷ Multilevel mixed-effects models ▷ Linear regression と操作
- Model タブ: Estimation method: Maximum likelihood (デフォルト)
Fixed-effects model: Dependent variable: weight
Independent variables: week
Random-effects model: Create...

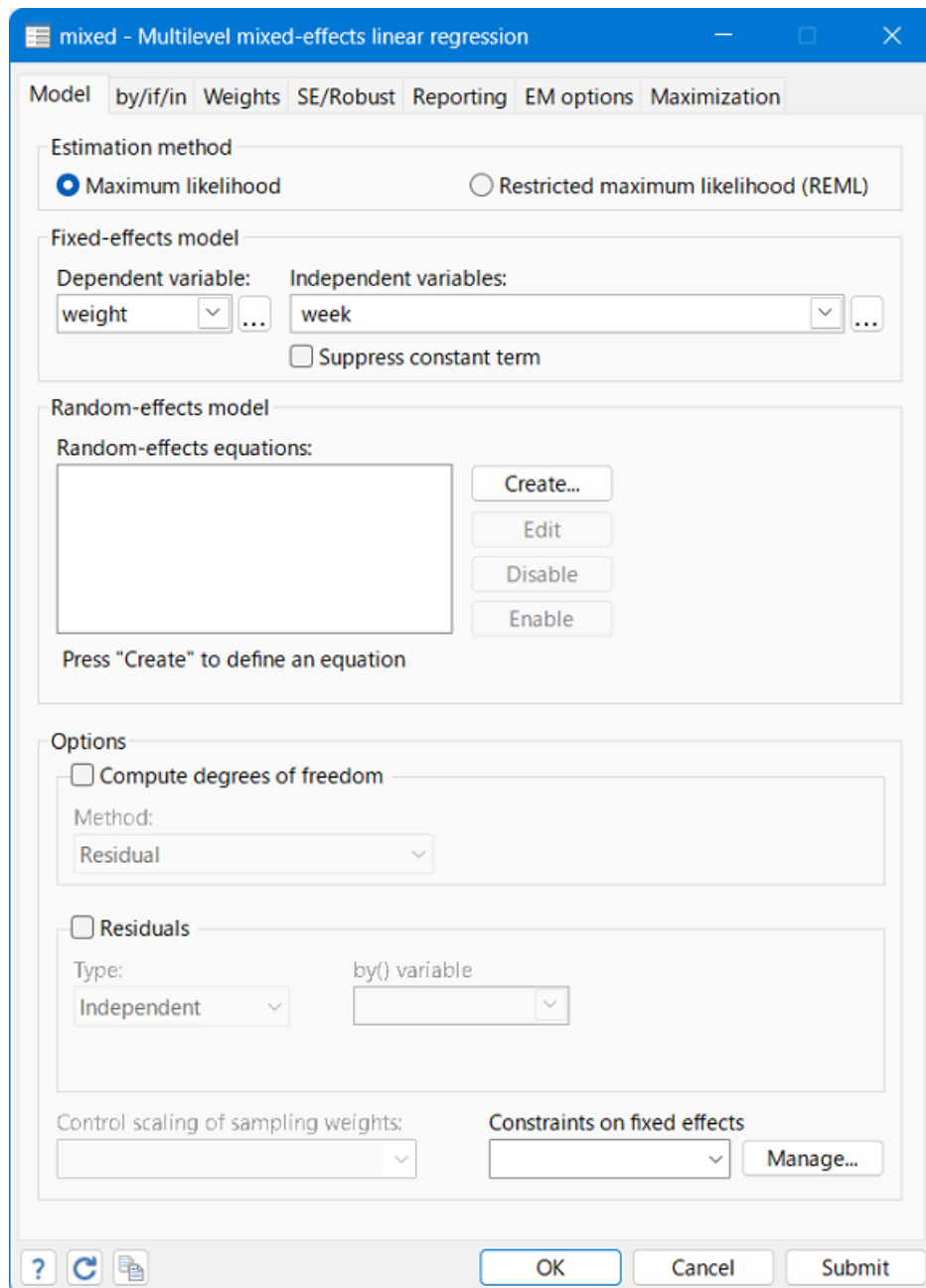


図 1 mixed ダイアログ- Model タブ

- Equation 1 ダイアログ: Level variable for equation: id

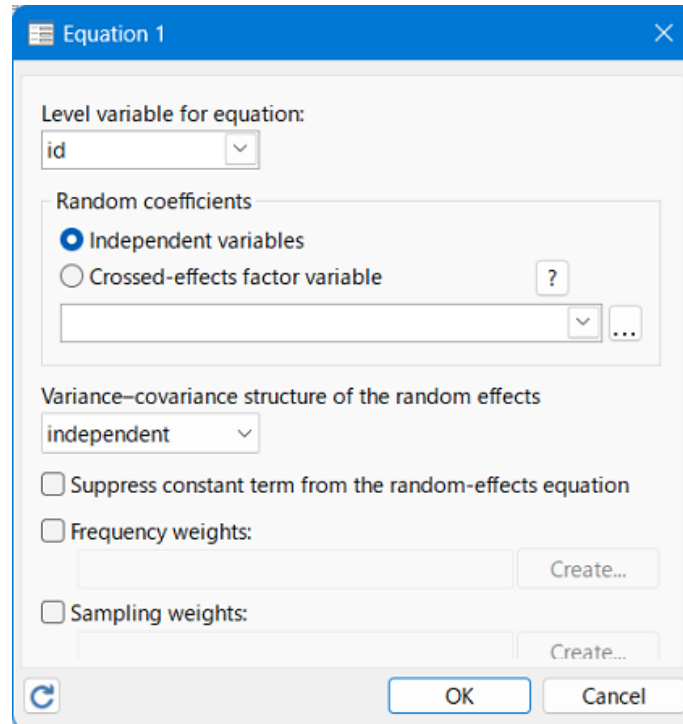


図 2 Equation 1 ダイアログ – ランダム切片

```
. mixed weight week || id:
Performing EM optimization ...

Performing gradient-based optimization:
Iteration 0:  Log likelihood = -1014.9268
Iteration 1:  Log likelihood = -1014.9268

Computing standard errors ...

Mixed-effects ML regression              Number of obs   =    432
Group variable: id                      Number of groups =    48
                                         Obs per group:
                                         min =          9
                                         avg =         9.0
                                         max =          9
                                         Wald chi2(1)   = 25337.49
Log likelihood = -1014.9268              Prob > chi2     =  0.0000
```

| weight | Coefficient | Std. err. | z | P> z | [95% conf. interval] | |
|--------|-------------|-----------|--------|-------|----------------------|----------|
| week | 6.209896 | .0390124 | 159.18 | 0.000 | 6.133433 | 6.286359 |
| _cons | 19.35561 | .5974059 | 32.40 | 0.000 | 18.18472 | 20.52651 |

| Random-effects parameters | | Estimate | Std. err. | [95% conf. interval] | |
|---------------------------|---------------|----------|-----------|----------------------|----------|
| id: Identity | | | | | |
| | var(_cons) | 14.81751 | 3.124226 | 9.801716 | 22.40002 |
| | var(Residual) | 4.383264 | .3163348 | 3.805112 | 5.04926 |

LR test vs. linear model: $\text{chibar2}(01) = 472.65$ Prob >= chibar2 = 0.0000

なお、Equation 1 ダイアログ上で変量効果の分散共分散構造として independent がデフォルトとして選択されている点に注意してください。

コマンドインタフェースの場合には

```
. mixed weight week || id:
```

のように入力します。この構文と mixed からの出力について情報を補足しておくようになります。

評価版では割愛しています。

▷ Example 2

評価版では割愛しています。

3. 共分散構造

▷ Example 3

評価版では割愛しています。

4. REML 推定法

評価版では割愛しています。

5. 3 階層モデル

▷ Example 4

評価版では割愛しています。

6. ブロック対角な共分散構造

▷ Example 5

評価版では割愛しています。

7. 分散不均一な変量効果

▷ Example 6

評価版では割愛しています。

8. 分散不均一な残差誤差

▷ Example 7

評価版では割愛しています。

9. その他の残差誤差構造

▷ Example 8

評価版では割愛しています。

▷ Example 9

評価版では割愛しています。

10. 交差効果モデル

▷ Example 10

評価版では割愛しています。

▷ Example 11

評価版では割愛しています。

11. 収束上の問題

評価版では割愛しています。

12. サーベイデータ

本セクションについては英文マニュアルをご参照ください。

13. 小標本時の推定

▷ Example 13

評価版では割愛しています。

▷ Example 14

評価版では割愛しています。

補足 1 – グラフ作成コマンド操作

評価版では割愛しています。

