

t 検定・分散分析から混合モデルへ：
文理解研究の導入事例から学ぶ

**Shift from *t*-tests and ANOVAs to Mixed Effects Models:
Practical Applications in Psycholinguistic Research**

企画・話題提供：神長 伸幸（理化学研究所）
井上 雅勝（武庫川女子大学）
話題提供：新井 学（東京大学・日本学術振興会）

mixed.model.jp@gmail.com
http://mixedmodeljp.blogspot.jp

Abstract

In this workshop, we aim to introduce basic and practical information about the applications of linear mixed effects models in psycholinguistic research. The first talk (Jincho) discusses limitations of conventional analyses (*t*-tests and ANOVAs) and advantages of mixed effects models in analyzing data from psycholinguistic experiments. It also provides a step-by-step description of a standard procedure for applying these models. The second talk (Inoue) offers hands-on illustrations of how to apply mixed effects models to the data from self-paced reading experiments. The third talk (Arai) introduces a method for analyzing the data from visual world eye-tracking studies using generalized mixed effects models.

Keywords — Mixed Effects Models, Statistical Methods, Psycholinguistics

1. 開催趣旨

実験的なアプローチに基づく心理言語学研究では、主な解析手法として *t* 検定や分散分析が用いられてきた。しかし近年、心理言語学のデータや実験デザインをより適切に捉えるには混合モデルを用いるべきであると海外の研究者の間で議論されている。混合モデルを用いた研究事例がほとんどないという国内の現状を考えると、混合モデル解析について知っている人、使っている人、解析をいかした実験デザインを議論できる人を一人でも増やすことが、我が国の心理言語学の発展にとって急務であると考え。そこで本ワークショップでは、継続的な議論に向けての第一歩として混合モデル解析の基礎および心理言語学研究における導入事例を学ぶ場を提供したい。

2. 混合モデル解析の基礎と有用性

神長伸幸

本発表は、混合モデル解析を用いるための基礎と有用性を紹介する。話題として(1)従来の解析を用いる際の限界や問題点(2)混合モデルの基本的な考え方(3)混合モデルを学ぶ上でのポイント(4)混合モデルを用いるメリット(5)実際の解析手順例を取り上げる。

(1) *t* 検定や分散分析などの従来の解析方法の限界と問題点

実験的なアプローチによる心理言語学研究では、*t* 検定や分散分析を盛んに利用してきた。これはデータの性質や実験デザインを考慮すると本来適当ではない場合にまで及んでいた。例えば正答・誤答の2値でコード化される理解テストのカテゴリデータに対して *t* 検定や分散分析を用いることは、分散の正規性、等分散性といった解析データについての前提条件から逸脱しているため問題である。また、被験者内要因を含む実験デザインデータを解析する際、従来の分散分析は計算の都合のために被験者の繰り返しの影響をランダムサンプルの扱いにしないことを許容してきた。これらの問題点を解消できるのが混合モデルである。

(2) 混合モデルの基本的な考え方

混合モデルは、固定因子とランダム因子の両者を一つのモデルに組み込む手法である。*t* 検定もしくは分散分析における要因は固定因子として扱われる。また、被験者や刺激項目のような各水準に研究上の意味がなく母集団からのサンプルで構成

される変数をランダム因子として扱う。固定因子は効果量とその分散を推定するが、ランダム因子は効果量を0と仮定して分散のみを推定する。この解析方法は、従属変数データの分散の正規性や条件ごとの等分散性を前提としておらず、複数のランダム因子を一つのモデルに含められる。

(3) 混合モデルを学ぶポイント

混合モデルは t 検定・分散分析の拡張と位置付けられるが、それらにない考え方や手続きの把握が重要である。第1に t 検定や分散分析で扱う条件間の「差」は混合モデルの回帰方程式における各因子の「係数」、グラフ上の回帰直線の「傾き」に対応する。条件は因子にダミーコーディング(2水準の固定因子の場合、統制条件を0、実験条件を1とする)もしくはエフェクトコーディング(統制条件を-0.5、実験条件を0.5)することで表現される。第2に混合モデルを使った解析では、自ら回帰方程式を構成し、構成した複数のモデルのいずれがより妥当であるかを検証する。モデルの比較には逸脱度、AIC、BICなどの適合度指標が用いられる。第3に混合モデルは固定因子とランダム因子を明示的に設定する。ランダム因子は、得られるデータや研究の種類に応じて適宜設定しなければならない。第4にランダム因子が回帰方程式上の切片のみに影響するのか、固定因子の係数にも影響するのかを適合度を基準に決定する。

(4) 混合モデル解析のメリット

実験にもとづく心理言語学にとって混合モデル解析の第1のメリットは、被験者と刺激項目という二つのランダム因子を同時に考慮できることである。また、従属変数や独立変数のデータが正規分布していることを解析の前提としないため、異なる形式の分布を持つデータを扱う場合にも、データの性質を考慮に入れた上で固定因子の効果を検証できる。さらに、従来の分析では操作した要因以外の要因の影響を刺激項目の作成の段階でなるべく小さくするなどの対処が必要であった。このような対処は刺激項目の選定をより困難にするだけでなく、実験で得られた知見の一般化可能性を低めている可能性もある。混合モデルでは、関心

のある固定因子以外の因子も共変量として回帰方程式に組み込むことで、固定因子の効果を高い精度で検証できる。つまり、刺激として選定する項目の制限が緩和されると同時に、得られた知見の一般化可能性も高められる。

(5) 混合モデルの解析手順

最後に混合モデル解析の手順の一例を示す。発表では仮想データを用いて一連の手続きを具体的に説明する。

- ① データの性質と実際の分布を検討し、従属変数の誤差が正規分布に従うなら線形混合モデル、それ以外のデータには一般化混合モデルを適用する。
- ② 研究の興味となる要因とそれとは異なるが従属変数に影響を及ぼし得る要因を固定因子、研究上、水準間の差に意味がなく、従属変数の分散にのみ影響する要因をランダム因子として回帰方程式を作成する。
- ③ 制限最尤法を用いた混合モデル解析を実施する。初めにランダム因子について切片と各固定因子の係数への影響を最大限考慮したモデルを構築して解析する。
- ④ 固定因子の係数に対するランダム因子の影響を取り除いたモデルと③の最大モデルとの適合度の違いを対数尤度に基づいて検証し、適合度に有意差がなければシンプルなモデルを採用する。
- ⑤ 上記④で得られたモデルで固定因子の t 値の絶対値が2を超えている場合、その因子の効果は十分に0より大きいと推測できる。また可能であればマルコフ連鎖モンテカルロシミュレーションを用いて各因子の効果の有意性を検証する。

3. R を用いた線形混合モデル分析ー 読み時間データの分析手順と要点

井上雅勝

本発表の目的は、心理言語学的研究における主要な測定法としての自己ペース読み課題から得られる従属変数（読み時間, RT）について、R を用いた線形混合モデル分析の具体的な手順を示すことである。

題材となるデータとして、文の構造曖昧性にもとづくガーデンパス（GP）効果の量が固有名詞の識別性によって異なることを見出した筆者の研究（井上, 2010）を取り上げる。この研究の主たる知見は、(1a, b) のような一時的構造曖昧性をもつ文の関係節主要部（e.g., 吉田を）で生じる GP 効果が、主節主語の固有名詞の識別性が低い (1a) よりも、識別性が高い (1b) の方が小さくなるという現象である。この結果は、識別性の高い固有名詞（e.g., タモリ）は記憶負荷がより小さいことから、(1b) では [タモリが今井を殴った...] という単文解釈と、[タモリが [今井を殴った] (関係節主要部) ...] という関係節解釈の両方が保持される可能性が高くなり (1a) と比較して相対的に GP 効果が小さくなるという、作動記憶容量と複数解釈保持可能性との相互作用（Just & Carpenter, 1992）から考察される。この現象に対する統計分析の主眼は、関係節主要部（e.g., 吉田を）の RT について、(1a) > (1b) > (2a) = (2b) のようなパターンの交互作用が現れることを見出すことである。

(1a) 藤田が/今井を/殴った/吉田を/誉めた。

(1b) タモリが/今井を/殴った/吉田を/誉めた。

(2a) 今井を/殴った/吉田を/藤田が/誉めた。

(2b) 今井を/殴った/吉田を/タモリが/誉めた。

線形混合モデル分析に先立って、以下のようなデータ処理を施す。具体例として、上の実験で行われた操作を付記する。

1. 独立変数（カテゴリ）のコード化：主たる要因（固定因子）である (i) 構造的曖昧性 (1 vs. 2) および (ii) 主語固有名詞の識別性 (a

vs. b) の各条件（カテゴリ）を、それぞれ、(0.5, -0.5) のようにコードする。

2. 独立変数（連続量）の中心化：RT に影響すると考えられる 2 つの共変量（固定因子）として、固有名詞の識別性（事前の評定調査に基づく数値であるが、詳細は省略）。ないし動詞の親密度、各語句のモーラ数の値を準備する。いずれの連続量変数も、平均が 0 になるよう中心化する。
3. ランダム因子の設定：実験参加者の識別番号 (ss)、刺激のセット番号（4 つの条件文毎に割り振られた数値: set）ないし刺激番号 (item) を、ランダム因子に割りあてる。
4. 従属変数の調整：読文後の理解テストの誤答試行を分析から削除する。また、(i) 200ms 未満および 4000ms 以上の操作不具合にもとづく外れ値を含む試行を分析から削除する（Roland et al, 2012）、(ii) 領域毎・条件毎に、平均値 + 2.5~3SD 以上の数値を削除する、(iii) 対数変換する、といった方法によってデータ分布の正規性を高める。
5. データ配置：測定された各試行の実験条件記号、実験参加者番号、共変量の数量、raw RT、中心化された変数の値といった情報が行毎に配置されたデータを分析に用いる。従来の被験者分析 (F1) / 項目分析 (F2) のように、被験者毎 / 項目毎の平均値に基づいて分析するわけではない。

線形混合モデルの処理の指針は以下の通りである（以下、本実験の分析で実際に使用された記号を例として用いる）。

1. 記号の説明：(i) 従属変数: 第 4 語句の RT (r4)、(ii) 実験仮説にかかわる固定因子（要因）とその交互作用 (f1c * f2c: なお lmer 関数の書式で "f1c * f2c" は、f1c および f2c の主効果と、f1c と f2c 間の交互作用を表す)、(iii) 従属変数に影響すると考えられる共変量としての固定因子（e.g., R4 領域の固有名詞の識別性 (fam4c) や動詞の親密度、語句の

モーラ数 (mora4c) など)、(iv) 実験参加者記号 (ss) と刺激セット番号 (set) のようなランダム因子、(v) ランダム因子項 (e.g., (1 | ss), (1 | set), (f1c*f2c | ss) など)。

- 固定因子とランダム因子からなる最大の回帰式 (3) を作成する。

(3) $\text{lmer}(r4 \sim f1c*f2c + (1+f1c*f2c | ss) + (1+f1c*f2c | set) + fam4c + mora4c)$

- 最大モデル (3) に対する線形混合モデル分析を実施し、推定係数 (Estimate) の t 値が 2 未満の共変量があれば、モデルに寄与しないとみなして (Gelman & Hill, 2007)、回帰式から削除する (ただし、この方法は暫定的なものである)。一方、実験仮説に関連する実験変数 (ここでは、f1c, f2c) とその交互作用は、最後まで削除しないのが普通である。なお、実験変数としての固定因子と共変量固定因子間の交互作用を検討するかどうかは、実験目的によって異なる。

- 共変量を適宜削除した最大モデル (4a) から、後進ステップワイズの様式に基づき、ランダムスロープ項を順次削除してより単純なモデルを作成し、最小モデル (4g) までの式リスト (e.g., 4a~g) を作成する (Jaeger, 2009 に基づく)。

(4a) $\text{model1} = \text{lmer}(r4 \sim f1c*f2c + (1+f1c*f2c|ss) + (1+f1c*f2c|set))$

(4b) $\text{model2} = \text{lmer}(r4 \sim f1c*f2c + (1+f1c*f2c|ss) + (1+f1c+f2c|set))$

(4c) $\text{model3} = \text{lmer}(r4 \sim f1c*f2c + (1+f1c+f2c|ss) + (1+f1c+f2c|set))$

(4d) $\text{model4} = \text{lmer}(r4 \sim f1c*f2c + (1+f1c+f2c|ss) + (1+f1c|set))$

(4e) $\text{model5} = \text{lmer}(r4 \sim f1c*f2c + (1+f1c|ss) + (1+f1c|set))$

(4f) $\text{model6} = \text{lmer}(r4 \sim f1c*f2c + (1+f1c|ss) + (1|set))$

(4g) $\text{model7} = \text{lmer}(r4 \sim f1c*f2c + (1|ss) + (1|set))$

- 尤度比検定 (the likelihood ratio test) を用いて、

複雑なモデルと次に単純なモデルとを比較する (Baayen, Davidson, & Bates, 2008; Jaeger, 2009)。具体的には、2つのモデルを、anova (モデル名 1, モデル名 2) 関数によって比較し、 χ^2 検定の結果が有意でない場合、ないし単純なモデルの対数尤度比 (LogLik) が有意に大きい場合は、より単純な方を選択し、さらに検定を続ける。複雑なモデルの対数尤度比の方が有意に大きい場合はこれを採用し、検定を打ち切る。

- 確定した最終モデルを、線形回帰モデル結果に類似した表として報告する。ここでは、推定係数 (Estimate)、標準誤差 (Standard error)、 t 値が記載される (Bates, 2006)。
- t 値が 2 を超える固定因子および交互作用を、従属変数に対して有意な効果があると解釈する (Gelman & Hill, 2007)。

発表では、以上の実験データの分析過程を実際に示すとともに、分析や結果を論文上でどのように記載するかサンプルを、本発表の事例に基づいた資料論文として配布する予定である。

参考文献

Baayen, R. H., Davidson, D. J., & Bates, D. M. (2008). Mixed-effects modeling with crossed random effects for subjects and items. *Journal of Memory and Language*, 59, 390–412.

Bates, D. (2006). lmer: p-values and all that. [Post on a mailing list] <https://stat.ethz.ch/pipermail/r-help/2006-May/094765.html>

Gelman, A., & Hill, J. (2007). *Data analysis using regression and multilevel/hierarchical models*. New York: Cambridge University Press.

井上雅勝 (2010). 文の意味的曖昧性が構造的曖昧性の解消と保留に及ぼす影響 日本認知心理学会第 8 回大会発表論文集, 81.

Jaeger, T. F. (2009). Random effect: Should I stay or should I go? [Web log post] <http://hlplab.wordpress.com/2009/05/14/randomoeff>

ect-structure/ Retrieved 24.07.11.

Just, M. A., & Carpenter, P. A. (1992). A capacity theory of comprehension: Individual differences in working memory. *Psychological Review*, 99, 122-149.

4. 線形混合モデルを用いた眼球運動データ解析

新井学

現在眼球運動測定を用いた言語理解実験は、文を構成する文字を画面上に提示し、それを読んで理解する際の眼球運動を計測する手法（リーディングパラダイム）と、画面上に絵刺激を提示し、音声として提示された言語刺激を理解する際の眼球運動を計測する手法（視覚世界パラダイム、以下VWP）に大きく分けられる。どちらの手法においても測定機から得られるデータは、各データポイントが記録された時間の情報、そして画面の解像度に従ったX-Y座標値の位置の情報の二つである。そして実験で得られるデータポイントの数は計測機器のサンプリングレートによって決定される（サンプリングレートが500Hzであれば1秒あたり500個のデータが記録される）。

リーディングパラダイムの場合、視点が同じ場所に連続して留まるデータポイントを停留として、ある特定の語句における停留の持続時間を指標に、その語句の、またはその語句までの言語情報の処理にかかる負荷を推定することができる。この場合、停留時間は連続変数として自己ペース読み実験で得られる反応時間データと同様の従属変数として扱える。

一方、VWPの場合にはまだ決定的な分析方法は確立しておらず、現在も議論が継続中である。その原因として(1)従属変数がカテゴリ変数であること(2)データポイント間に依存関係が在ることの二つが挙げられる。

VWPにおけるデータは、被験者の視線の位置とオブジェクトの位置とのマッピングで得られたカテゴリ変数として扱われる。実際には分析対象として特定のオブジェクトに対する注視の量に実験

条件間で差が見られるか検定を行うのが一般的であり、この場合ターゲットオブジェクトに対して注視がある(1)、なし(0)という二項変数データとして扱われる。二項変数の分布は正規分布とは異なる為、そのままでは線形モデルを適用することができない。

また各試行で得られるデータは同じ被験者・項目から得られ、時間的に連続している為、隣り合うデータポイントの間には依存関係が存在する。つまり、ある時点でオブジェクトAを見ていた場合、次のサンプリングの時点でそのままオブジェクトAを見続けている確率は、無作為に選んだある一時点におけるオブジェクトAへの注視の確率よりも遙かに高くなる。これはつまり対象物に対する注視の確率が、一つ前にサンプルされた注視に依存していることを意味しており、これをそのまま従属変数として扱うことは線型回帰モデルにおける独立性の仮定にそぐわない。

本発表で紹介するのは、VWPにおけるデータに線形混合モデルを適用する為に、ある特定の時間内のデータポイントを合算することでデータポイント間の非独立性の問題を解決し、更に特定のオブジェクトに対する注視を対数変換したオッズとして計算することでカテゴリ変数としての分布の問題を解決するという手法である。

この手法を用いる際、リンク関数として Barr (2008)で紹介されている以下の関数を使用する。

$$\eta' = \ln\left(\frac{y + 0.5}{n - y + 0.5}\right)$$

y は興味の対象であるオブジェクトへの一定時間内の注視回数、そして n は背景を含むすべてのオブジェクトへの注視回数を示す。これによって得られる値は「画面上に向けられたすべての注視回数の内どれだけがターゲットオブジェクトに向けられたか」を表すことになる。分子分母両方に0.5が足されているのは、0の対数は定義されていない為、分子分母どちらかに0を含む場合対数変換できないという問題を回避する為である。

参考として紹介する Nakamura, Arai, and Mazuka (2012)は「男の子が三輪車に乗っていた女の子を

見つめた」のような関係節文を理解する際に、関係節上に文脈と整合する韻律情報が与えられた場合、関係節構造の予測が起こることを実証した研究である。ここで報告されているデータ分析では、「乗っていた」という関係節内の動詞のオンセットからオフセットの長さに相当する時間内で観測された関係節主部に対応するオブジェクト（女の子）への注視回数の対数オッズを計算し、線形混合モデルを用いて分析を行っている。

本発表ではより具体的にランダム因子の選択方法として最小モデルから変数を増加させる方法と最大モデルから変数を減少させる方法の2通りを紹介・比較し、更にランダムスロープを含めたモデルでは現時点で p 値が出力されないという問題への対処法も議論する。

参考文献

- Barr, D. J. (2008). Analyzing 'visual world' eyetracking data using multilevel logistic regression. *Journal of Memory and Language*, 59, 457-474.
- Nakamura, C., Arai, M., & Mazuka R. (2012). Immediate use of prosody and context in predicting a syntactic structure. *Cognition*, 125, 317-323.