

МЕТОДЫ

Линейные модели со смешанными эффектами в когнитивных исследованиях

Андрей Анатольевич Четвериков

Лаборатория когнитивных исследований ИОН РАНХИГС, Москва;
факультет психологии СПбГУ, Санкт-Петербург, Россия

Аннотация. Анализ данных в когнитивных исследованиях всегда сопряжен с определенными трудностями: отсутствие полностью сбалансированного дизайна, невозможность подобрать полностью эквивалентные стимулы, пропущенные значения в ответах и т. д. Для контроля случайной ошибки чаще всего применяется усреднение данных по испытуемым либо стимулам, а после этого усредненные данные анализируются с помощью традиционных методов, таких как дисперсионный анализ. В статье рассматривается альтернативный подход, основанный на применении регрессионных моделей со смешанными эффектами. Этот вид моделей позволяет анализировать данные без усреднения, включая в модель в явном виде допущения о существующих между отдельными наблюдениями взаимосвязях. В качестве примера анализа используются результаты квазиэксперимента, в котором оценивается связь времени ответа в задаче узнавания с правильностью ответа, полом испытуемого и полом человека на предъявленной фотографии. Сравнение результатов анализа с помощью смешанных моделей и различных вариантов дисперсионного анализа (с усреднением по стимулам, по испытуемым и без усреднения) показывает схожесть результатов смешанных моделей с результатами дисперсионного анализа по испытуемым и результатами дисперсионного анализа по наблюдениям. Однако в случае смешанных моделей интерпретация результатов становится более обоснованной, так как и различия между стимулами, и различия между испытуемыми включены в модель. В заключении рассматриваются преимущества смешанных моделей и даются рекомендации по выбору случайных эффектов для включения в модель.

Контактная информация: Четвериков Андрей Анатольевич, a.chetverikov@psy.spbu.ru; факультет психологии СПбГУ, наб. Макарова, д. 6, 199034 Санкт-Петербург, Россия.

Ключевые слова: смешанные модели, регрессионный анализ, дисперсионный анализ, время реакции, квазиэксперимент, ошибки

© 2015 Четвериков Андрей Анатольевич. Данная статья доступна по лицензии [Creative Commons "Attribution" \(«Атрибуция»\) 4.0. всемирная](https://creativecommons.org/licenses/by/4.0/), согласно которой возможно неограниченное распространение и воспроизведение этой статьи на любых носителях при условии указания автора и ссылки на исходную публикацию статьи в данном журнале в соответствии с канонами научного цитирования.

Статья частично основана на материалах более ранней публикации в сборнике материалов I Международной научной конференции «Формирование основных направлений развития современной статистики и эконометрики»: Четвериков А. А. Опыт применения смешанных регрессионных моделей для анализа квазиэкспериментальных данных // Материалы I Международной научной конференции «Формирование основных направлений развития современной статистики и эконометрики» / Под ред. В. Н. Афанасьева. Оренбург: ООО ИПУ «Университет», 2013. Т. 1. С. 332–338.

Благодарности. Работа выполнена при поддержке РФФИ (#15-06-07417 А) и СПбГУ (НИР #8.38.287.2014).

Статья поступила в редакцию 18 февраля 2015 г. Принята в печать 21 марта 2015 г.

Введение

В психологическом эксперименте основным объектом интереса являются различия в значении зависимой переменной между уровнями одной или нескольких независимых переменных. Все остальное — температура в помещении, время суток, индивидуальные различия испытуемых, свойства стимулов и множество других факторов — считаются «шумом» или случайной ошибкой, влияния которой экспериментатор старается избежать всеми доступными способами. Для этого испытуемые случайным образом распределяются между группами, стимулы — между условиями, испытуемые изолируются от внешних факторов в лаборатории, применяется специальное оборудование, увеличивается количество наблюдений или опытов, а также используются другие методические приемы, хорошо известные любому исследователю. После сбора данных наступает пора их анализа. В идеальной ситуации перед исследователем предстает прекрасная картина сбалансированного дизайна, в котором отсутствуют пропущенные значения, количество измерений для каждой комбинации испытуемого и независимых переменных достаточно велико и все измерения проводились в одних и тех же условиях.

В реальности, разумеется, ситуация отличается. Появляются пропущенные значения, например, часто из анализа исключаются ошибочные ответы испытуемого, выясняется, что из-за усталости и научения испытуемого количество наблюдений достаточно ограничено, на предполагавшиеся одинаковыми стимулы испытуемые дают разные реакции и т.д. Хуже всего дело обстоит в тех случаях, когда какие-либо переменные оказываются неподконтрольными экспериментатору. Например, пытаясь изучить связь между частотой употребления слова и вероятностью его опознания при краткосрочном предъявлении, исследователь может сколь угодно долго выравнять группы частот и редкоупотребляемых слов по всем характеристикам, однако в результате перед ним все равно окажутся разные группы слов. Профессиональная область интересов автора связана с изучением ошибок и эмоциональных реакций, и, как можно предположить, и то, и другое в достаточно слабой степени может контролироваться экспериментатором.

Подобные ситуации возникают наиболее часто в квазиэкспериментальных исследованиях, однако и в других видах исследований может быть полезно учесть влияние переменных, которые невозможно контролировать. Например, данные в сбалансированном психофизическом эксперименте будут тем не менее подвержены влиянию научения и усталости, то есть будет влиять порядковый номер пробы. Психологи, проводящие исследования в школах, могут захотеть учесть различия между испытуемыми, набранными в разных классах или разных школах. Физиологические данные, собираемые в лонгитюдных исследованиях, могут различаться в зависимости от времени суток и т.д.

Классическим методом для контроля ошибки в ситуациях, когда невозможно отделить независимую переменную от испытуемого (иногда подобные переменные также называют субъектными) или стимула,

являются тесты для двух связанных выборок и дисперсионный анализ (ANOVA) с повторными измерениями. Однако у анализа с повторными измерениями есть некоторые недостатки. Во-первых, часто оказывается неочевидным, что нужно выбрать в качестве единицы анализа. Например, при анализе реакций испытуемого на разные стимулы в зависимости от экспериментальных условий можно использовать в качестве единицы анализа как стимул, так и испытуемого. Чаще всего в таких ситуациях предлагается проводить анализ после усреднения реакций «по стимулам» или «по испытуемым». Однако при усреднении будет потеряна часть данных. Более того, у нас нет никаких оснований для того, чтобы выбрать из этих двух вариантов более оптимальный. По этой причине в психолингвистических исследованиях рекомендуется приводить результаты обоих вариантов анализа и рассчитывать на их основе критерий $\min F'$: $\min F' = (F1+F2)/F1 \cdot F2$, где $F1$ и $F2$ относятся к результатам анализа по испытуемым и по стимулам соответственно (Clark, 1973; Raaijmakers et al., 1999; Raaijmakers, 2003).

Смешанные линейные модели

Возможной альтернативой дисперсионному анализу с повторными измерениями является регрессионный анализ с использованием смешанных линейных моделей. Сущность данного метода заключается в следующем. Предполагается, что эффекты (факторы), оказывающие влияние на зависимую переменную, можно условно разделить на два типа: фиксированные и случайные. Хотя по поводу философских и методологических аспектов разделения эффектов на фиксированные и случайные идут споры (Gelman, 2005; r-sig-mixed-models FAQ–GLMM, 2015), в данной работе я буду рассматривать как случайные те эффекты, которые случайным образом варьируются в исследовании и не отражают все возможные значения фактора в генеральной совокупности. Фиксированные эффекты, с другой стороны, — это то, что обычно является предметом интереса исследователя, то есть те независимые переменные, уровни которых он устанавливает или контролирует. С практической точки зрения необходимо также учитывать, что для оценки случайных эффектов необходимо, чтобы количество уровней соответствующего фактора было относительно велико — 5–6 уровней минимум (r-sig-mixed-models FAQ–GLMM, 2015).

По сути смешанные модели — это расширение регрессионных моделей. Обычная линейная регрессионная модель выглядит следующим образом:

$$y_i = \mu + \sum_{k=1}^p \beta_k x_{ki} + \varepsilon_i, \quad (1)$$

где y_i — значение зависимой переменной для i -го наблюдения, μ — константа, β_k — значение коэффициента для k -го фактора, x_{ki} — значение k -го фактора для i -го наблюдения, ε_i — значение остатка для i -го наблюдения.

Возьмем для примера простой эксперимент с задачей лексического решения, где время ответа измеряется в зависимости от типа стимула (слово – не слово) и количества букв в стимуле. Тогда y_i будут обозначать время ответа в каждой пробе, β_1 — различие по времени ответа между словом и не словом, x_{1i} — тип стимула в данной пробе, β_2 — изменение времени ответа при увеличении длины слова на одну букву, x_{2i} — длину слова в данной пробе, ε_i — отличие времени ответа, предсказанного на основе модели, от реального.

В случае смешанных моделей уравнение меняется следующим образом:

$$y_{ij} = \mu + \sum_{k=1}^p \beta_k x_{kji} + \sum_{m=1}^q b_{mj} z_{mji} + \varepsilon_{ji} \quad (2).$$

В отличие от β , коэффициентов для фиксированных эффектов, которые одинаковы для всех наблюдений, коэффициенты для случайных эффектов, b , определены для j -й группы наблюдений, а z_{mji} задает значение случайного эффекта для данной группы и данного наблюдения. Продолжая пример, допустим, каждая группа наблюдений — отдельный испытуемый и случайный эффект для каждого испытуемого только один — константа. Тогда b_j — среднее¹ отклонение времени ответа для данного испытуемого от общегруппового среднего, а z_{ji} — идентификатор испытуемого.

Группой наблюдений может являться не только испытуемый, но и стимул, или и то, и другое. Другими словами, можно предположить, например, что каждому испытуемому или стимулу соответствует свое время ответа, то есть наличие случайного свободного члена в уравнении регрессии. Кроме того, можно также предположить, что время правильных и неправильных ответов у разных испытуемых различается или различается время ответа на мужские лица в сравнении с женскими, то есть предположить случайный эффект наклона прямой регрессии для каких-либо внутригрупповых факторов. В случае более сложных дизайнов исследования испытуемые сами могут быть объединены в какие-то большие группы, например, школы или больницы, и тогда в уравнение регрессии добавляется еще один уровень.

Основное требования к данным в случае смешанных моделей — наличие групп взаимосвязанных наблюдений. В дополнение к этому смешанные линейные модели требуют тех же допущений, что и обычные линейные модели:

1. Эффекты в модели аддитивны, то есть влияние одного предиктора не зависит от уровня другого предиктора. Если это требование нарушается, необходимо преобразование переменных или включение в модель взаимодействий.

2. Линейная взаимосвязь между независимыми и зависимой переменной. Случаи нелинейной связи также можно анализировать, преобразовав переменные или включив в модель полином $(x+x^2+x^3+\dots)$.

3. Ошибки (остатки) имеют равную дисперсию и распределены нормально. Это наименее важные из ограничений, практически никогда не соблюдаемые на практике, и их нарушения обычно не приводят к существенным проблемам (Gelman, Hill, 2007).

Независимость ошибок, существенная для обычных линейных моделей, в случае смешанных моделей не обязательна (Pinheiro, Bates, 2000). Более того, отдельные статистические пакеты, например, *nlme* (Sarkar et al., 2008), позволяют легко задавать структуру взаимосвязей ошибок в явном виде, хотя эта тема выходит за пределы данной статьи.

Кроме того, число наблюдений должно быть не меньше числа уровней случайных эффектов в модели для того, чтобы анализ с помощью смешанных моделей стал возможен хотя бы теоретически. Например, если речь идет об испытуемых как случайном эффекте и в модели также присутствует случайный эффект для взаимодействия испытуемого и какого-либо еще фактора, то число наблюдений должно быть не меньше, чем число испытуемых, умноженное на число уровней фактора. Проводя аналогию с классическими тестами, провести тест Стьюдента для двух связанных выборок невозможно, если число наблюдений для каждого испытуемого меньше двух. Однако это лишь ограничение на применимость метода: осмысленность и надежность полученных результатов будет тем выше, чем больше наблюдений анализируется.

Как и для других регрессионных моделей, для смешанных моделей предикторами могут выступать как количественные (включая ранговые), так и категориальные переменные. Зависимая переменная также может измеряться в различных шкалах при условии, что для нее указано соответствующее распределение. Отдельно стоит отметить, что для биномиальных данных (в том числе точности ответов) использование смешанных биномиальных моделей приводит к более точным оценкам, чем дисперсионный анализ (Jaeger, 2008). Для ранговых переменных можно использовать порядковую регрессию со смешанными эффектами, например, с помощью пакета *DPPackage* (Jara et al., 2011), хотя во многих случаях вполне оправданным будет применение обычных смешанных моделей.

Благодаря подобным возможностям смешанные модели в последние годы начинают использоваться в различных областях: в психолингвистике (Mousikou et al., 2015), в исследованиях восприятия и внимания (Kristjánsson, Jóhannesson, 2014; Laubrock et al., 2008; Королькова, 2014), в работах по проблеме сознания (Sandberg et al., 2010), в социальной психологии (Biesanz, Human, 2010; Wright et al., 2010), в мета-анализе (Bakker, Wicherts, 2011), в исследованиях памяти (Oberauer, Kliegl, 2006; Wright, London, 2009), клинической психологии (Baldwin et al., 2009), психофизиологии (Gulbinaite et al., 2014; Wu, Clark, 2014) и т. д.

Сложность применяемых моделей также различается. Например, Сэндберг и коллеги (Sandberg et al., 2010) сравнивали различные шкалы осознанности, используя внутригрупповой дизайн и контролируя избыточную дисперсию между испытуемыми через включение в модель случайного эффекта для испытуемых. Мусику и коллеги (Mousikou et al., 2015) исследовали влияние

¹ Корректнее было бы говорить о предсказанном среднем отклонении времени ответа испытуемого, но для краткости здесь и далее — среднее.

прайминга в задаче лексического решения и включали в модель случайные эффекты для испытуемого, стимула и типа прайма с группировкой по стимулу. Четвериков (Chetverikov, 2014) использовал смешанную модель для контроля вариативности стимулов и индивидуальных различий между испытуемыми в задаче оценки привлекательности лиц. Применяются и более сложные модели. Например, в работе Болдуина и коллег (Baldwin et al., 2009) рассматривалась эффективность психотерапии с включением случайного эффекта пациента, линейного и квадратичного эффектов количества психотерапевтических сессий с группировкой по пациенту, эффекта психотерапевта и линейного эффекта количества сессий с группировкой по психотерапевту.

При этом количество статей с использованием смешанных моделей постоянно увеличивается. На момент написания этой статьи Google Scholar по запросу «mixed effects model psychology» выдавал 1150 результатов за 2012 год, 1440 — за 2013 год, и 1890 — за 2014. Далеко не все авторы упоминают применяемый метод именно с формулировкой “mixed effects model”, поэтому реальное число статей с использованием смешанных моделей еще выше.

В оставшейся части статьи я описываю различия в подходах и результатах обработки квазиэкспериментальных данных с помощью дисперсионного анализа и смешанных регрессионных моделей на примере реального исследования.

Анализ времени ответа в квазиэксперименте

Дисперсионный анализ с усреднением по испытуемым / стимулам

Рассмотрим следующий квазиэксперимент. Использовалась задача узнавания: испытуемому на краткий промежуток времени предъявляется фотография лица, затем через какое-то время эта фотография предъявляется вместе с другой, новой, и испытуемого просят выбрать фотографию, которую он уже видел ранее. Исследователя интересует, как соотносится между собой время правильных и неправильных ответов. Правильность ответа испытуемого невозможно задать экспериментально, не влияя на другие параметры задачи. Например, можно сделать так, что будут только неправильные ответы, сделав задачу практически нерешаемой, но это будет уже фактически другая задача. Поэтому в данном случае речь идет о классическом квазиэксперименте: независимая переменная «правильность ответа» не контролируется экспериментатором. Кроме того, ему необходимо учитывать фактор пола испытуемого и пола человека на фотографии (фактор, далее обозначаемый как тип стимула), так как есть предположение, что данные факторы могут мешать корректной оценке эффекта правильности ответа в генеральной совокупности.

Есть несколько возможных вариантов проведения дисперсионного анализа: анализ «по испытуемым», «по стимулам» и «по наблюдениям». В первом случае рассчитывается среднее для каждого испытуемого

по каждому типу стимула, во втором случае, наоборот, рассчитывается среднее для каждого стимула по каждому типу испытуемых. Эти виды анализа особенно проблематичны в случаях, подобных рассматриваемому, поскольку распределение стимулов по правильным и неправильным ответам не является случайным, а определяется сочетанием свойств испытуемого и стимула. Наконец, в третьем случае можно отказаться от анализа с повторными изменениями в пользу обычного дисперсионного анализа в надежде, что избыточная дисперсия будет скомпенсирована увеличением количества измерений за счет отсутствия усреднения.

Для данной статьи я использовал данные эксперимента, цель которого не заключалась в анализе эффекта правильности ответов, но который тем не менее позволяет провести подобный анализ. В рамках данного эксперимента по описанной выше процедуре испытуемые ($N = 60$) выполняли 43 пробы каждый. Всего использовалось 86 стимулов, для каждого испытуемого 43 из них выбирались случайным образом. Таким образом, набор данных состоял из 2580 наблюдений (данные доступны для анализа в онлайн-приложении к статье).

На рисунке 1 приведены медианы времен ответов в зависимости от пола испытуемого, типа стимула и правильности ответа. Как видно из данного графика, распределение ответов достаточно сильно отличается от равномерного. Мужских фотографий предъявлялось больше, чем женских (это связано с соотношением числа мужских и женских фотографий в использованной базе стимулов), правильных ответов заметно больше, чем неправильных. В результате у испытуемых номер 5, 6, 16, 19, 24, 28 практически нет неправильных ответов при предъявлении женских фотографий. Анализ распределения времен правильных и неправильных ответов в целом по группе (правая часть графика) позволяет предположить, что будет выражен эффект правильности ответа и, возможно, типа стимула.

Попробуем сначала провести дисперсионный анализ с усреднением данных «по испытуемым» и «по стимулам». Анализ проводился в системе статистических вычислений R (R Core Team, 2014) с применением суммы квадратов типа II и контрастов с нулевой суммой (*sum-to-zero*). Результаты анализа приведены в таблице 1. Здесь и далее переменные обозначены следующим образом: пол испытуемого — *Subject Gender*, тип стимула — *Stimulus Gender*, правильность ответа — *Accuracy*. Взаимодействия переменных обозначены через « \times ». Статистически значимыми оказались эффекты правильности ответа и взаимодействия правильности ответа и типа стимула. Кроме того, на уровне тенденции проявился основной эффект типа стимула. При попытке провести анализ «по стимулам» оказалось, что для 6 из 86 стимулов не хватает данных, т.е. для них не представлены те или иные комбинации пола испытуемого и правильности ответа. Результаты анализа по оставшимся стимулам представлены в таблице 1. Как можно заметить, они сильно отличаются от анализа «по испытуемым» — значимыми остались только эффекты типа стимула и правильности ответа.

Далее я использовал третий подход — дисперсионный анализ «по наблюдениям», без усреднения. Хотя при использовании этого подхода отсутствует потеря

	По испытуемым		По стимулам		По наблюдениям		Смешанная модель
	<i>F</i>	<i>p</i>	<i>F</i>	<i>p</i>	<i>F</i>	<i>p</i>	<i>p</i>
Accuracy	60.29	<.001	53.15	<.001	80.75	<.001	<.001
Subj. Gender	0.23	.636	0.12	.729	0.01	.904	.644
Stim. Gender	3.84	.055	7.64	.007	8.30	.004	.011
Accuracy × Subj. Gender	5.32	.025	2.28	.135	5.52	.019	.026
Accuracy × Stim. Gender	4.77	.033	0.77	.382	4.60	.032	.018
Subj. Gender × Stim. Gender	0.76	.387	1.66	.201	2.37	.124	.292
Accuracy × Stim. Type × Subj. Gender	0.23	.632	0.01	.928	1.25	.263	
Residual DF	58		78		2572		

Таблица 1. Результаты дисперсионного анализа и анализа с применением смешанной модели.

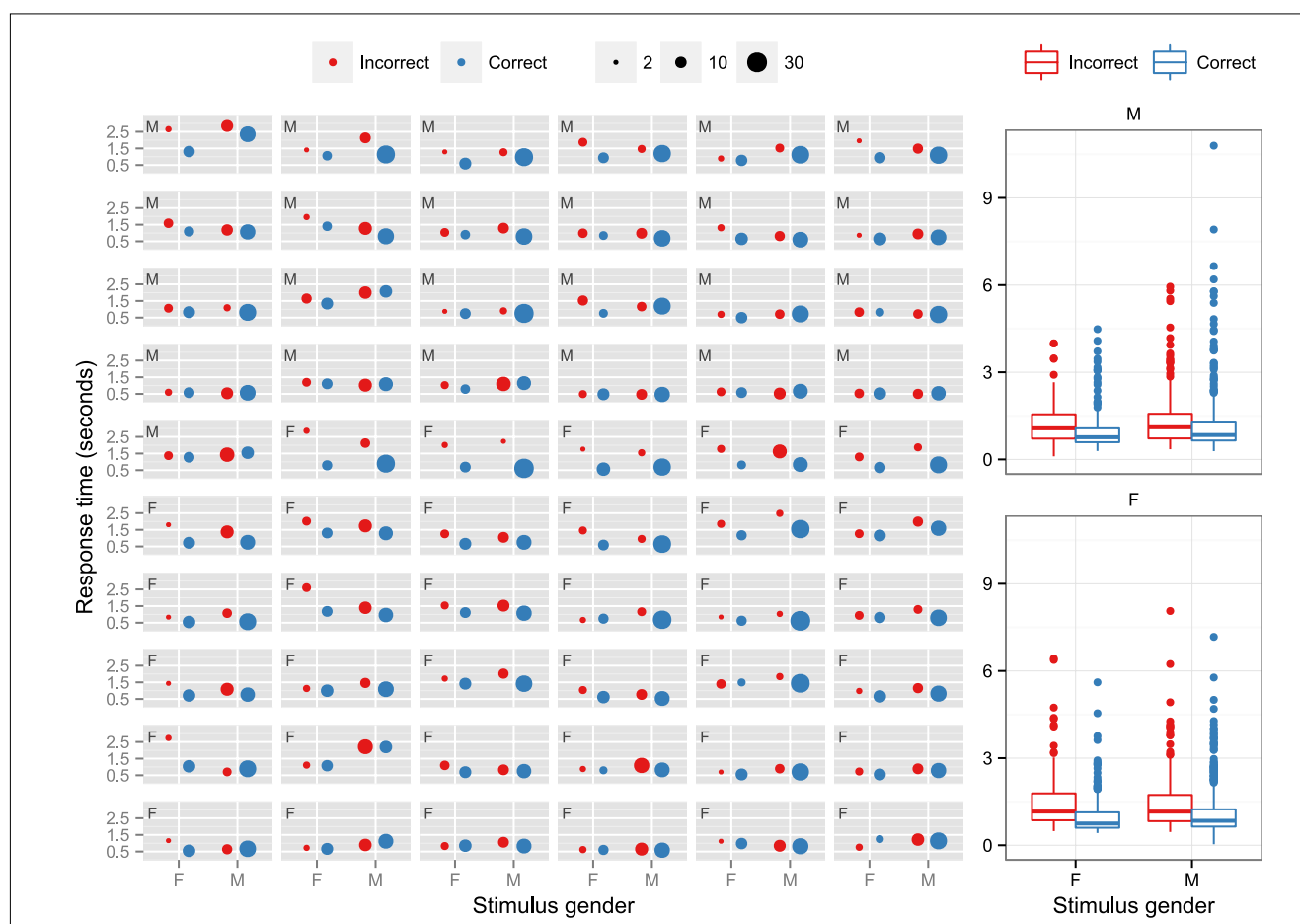


Рисунок 1. Медианы времен правильных и неправильных ответов (в секундах) по испытуемым (слева) и в целом по выборке (справа). Индексы F и M под графиком означают женские и мужские лица соответственно. Индексы F и M над графиками справа и в углах графиков слева означают пол испытуемых. Красным цветом отображены неправильные ответы, синим — правильные. Площадь кругов на графиках слева отражает количество наблюдений. На графиках справа верхняя и нижняя граница «ящика» показывает первый и третий квартили, верхняя и нижняя граница вертикальных линий — ±полтора межквартильного расстояния, точки — выбросы, выходящие за ±полтора межквартильного расстояния.

данных вследствие усреднения, возникает опасность того, что результаты будут искажены из-за несбалансированности данных. Например, низкая средняя точность и скорость ответов нескольких отдельно взятых испытуемых может привести к тому, что будет проявляться снижение скорости неправильных ответов по всей выборке. Кроме того, не учитывается взаимосвязь результатов каждого испытуемого, то есть нарушается допущение дисперсионного анализа о независимости измерений. Как видно из таблицы 1, результаты

дисперсионного анализа сырых данных согласуются с результатами анализа «по испытуемым», но не согласуются с результатами анализа «по стимулам».

Смешанные линейные модели

В смешанную регрессионную модель были включены шесть фиксированных эффектов: пол испытуемого, тип стимула, правильности ответа и их взаимодействия второго уровня. Кроме того, в модель были включены четыре случайных эффекта: (1) испытуе-

Случайный эффект в lme4	Формальная интерпретация	Пример содержательной интерпретации в исследовании
$(1 z)$	случайное значение среднего для группы z	испытуемые различаются по времени ответа
$(x z) = (1+x z)$	случайное значение константы для группы z, случайное значение угла наклона x внутри группы z, которые могут зависеть друг от друга	испытуемые различаются по времени ответа, по-разному реагируют на x, и эти два эффекта могут быть взаимосвязаны
$(0+x z) = (-1+x z)$	только случайное значение угла наклона x внутри группы z, средние не различаются	испытуемые в среднем не различаются по времени ответа, но могут по-разному реагировать на x
$(1 z) + (0+x z) = (x z)$	случайное значение константы для группы z, случайное значение угла наклона x внутри группы z, не зависящие друг от друга	испытуемые различаются по времени ответа, по-разному реагируют на x, но эти два эффекта являются независимыми
$(1 z/block) = (1 z)+(1 z:block)$	случайное значение среднего для группы z, которое может меняться между блоками block внутри группы z (вложенные случайные эффекты)	испытуемые различаются по времени ответа, причем для каждого испытуемого среднее может варьироваться между блоками (или стимулами, если каждый стимул повторяется по несколько раз)
$(x site/block) = (x site)+(x site:block) = (1+x site)+(1+x site:block)$	случайное значение среднего и угла наклона x для групп z, которое может меняться между блоками block	испытуемые различаются по среднему времени ответа и времени ответа на x, причем для каждого испытуемого среднее и эффект x могут по-разному варьироваться между блоками, возможна взаимосвязь между ними
$(1 z1)+(1 z2)$	случайное значение среднего для групп z1 и групп z2	испытуемые и стимулы различаются по времени ответа, эти эффекты не взаимосвязаны
$(x z1)+(x z2) = (1+x z1)+(1+x z2)$	случайное значение среднего и угла наклона x для групп z1 и z2, внутри каждой группы среднее и угол наклона могут зависеть друг от друга	испытуемые различаются по времени ответа, по-разному реагируют на x, и эти два эффекта могут быть взаимосвязаны; стимулы также различаются по времени ответа на них в среднем и в зависимости от x, эти эффекты также могут быть взаимосвязаны

Таблица 2. Примеры описания случайных эффектов в пакете lme4.

Group	Variable 1	Variable 2	Std. Dev.	Correlation
Stimulus	(Intercept)		0.04	
Subject	(Intercept)		0.43	
	Accuracy		0.16	
	Stim. Type		0.18	
Subject	(Intercept)	Accuracy		-.86
	(Intercept)	Stim. Type		.33
	Accuracy	Stim. Type		.11
Residual			0.74	

Таблица 3. Анализ смешанной модели, случайные эффекты.

	χ^2	df	p
Accuracy x Subject	11.64	3	.009
Stim. Type x Subject	17.41	3	<.001
Stimulus	0.18	1	.671

Таблица 4. Анализ необходимости включения случайных эффектов в модель. Примечание: χ^2 — значение статистики χ^2 для теста отношения правдоподобия при сравнении моделей с включением случайного эффекта и без него, df — разница в степенях свободы между моделями.

	LRT		Parametric bootstrap	Kenward-Roger			
	B	SE		F	p		
Accuracy	-0.30	0.07	39.51	<.001	<.001	53.94	<.001
Subj. Gender	0.17	0.13	0.18	.672	.644	0.17	.677
Stim. Type	0.04	0.08	6.61	.010	.011	6.93	.011
Accuracy x Subj. Gender	-0.17	0.08	4.84	.028	.026	4.81	.032
Accuracy x Stim. Type	0.16	0.07	5.66	.017	.018	5.59	.018
Subj. Gender x Stim. Type	-0.08	0.08	1.11	.292	.292	1.08	.304

Таблица 5. Анализ смешанной модели, фиксированные эффекты. Примечание: B — регрессионный коэффициент, SE — стандартная ошибка коэффициента, LRT — тест отношения правдоподобия, Parametric Bootstrap — оценка p с помощью параметрического бутстрапа, Kenward-Roger — расчет степеней свободы по методу Кенварда и Роджера.

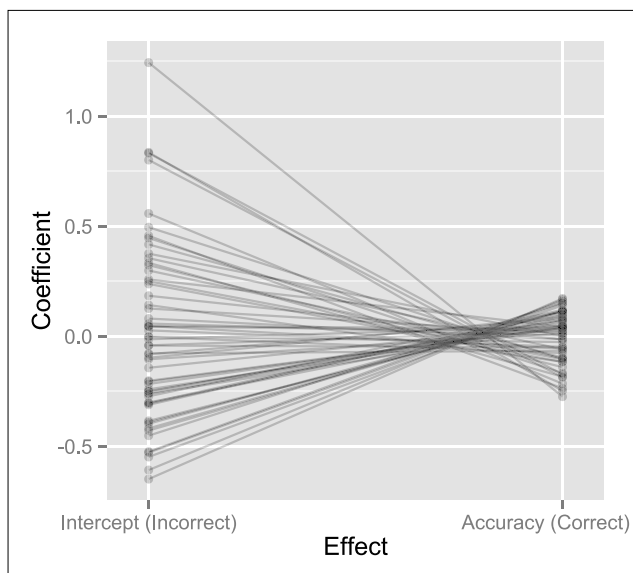


Рисунок 2. Взаимосвязь коэффициентов случайных эффектов по испытуемым. *Intercept* — константа в использованном уравнении регрессии — отражает отклонение времени неправильных ответов испытуемого от среднего времени неправильных ответов, *Accuracy* — случайный эффект для правильности ответа — отражает отклонение разницы между правильными и неправильными ответами испытуемого по времени от средней разницы.

мого; (2) случайного угла наклона для правильности ответа по испытуемым; (3) стимула; (4) случайного угла наклона для пола испытуемого по стимулам.

Данная модель предполагает, что испытуемые отличаются по среднему времени ответа и могут по-разному реагировать на стимулы разного пола и что время ответа на конкретные стимулы также отличается и может быть различным в зависимости от пола испытуемого. Я не включал в модель взаимодействие третьего уровня, так как все три предшествующих анализа показали, что оно не значимо, а его исключение упрощает расчеты модели.

Для проведения анализа была использована библиотека *lme4* (Bates et al., 2014). В R существует несколько альтернативных библиотек для анализа смешанных моделей, однако наиболее легкой в использовании представляется именно эта. Современные версии других программ для статистических расчетов, таких как STATA, SAS и SPSS, также включают в себя команды для регрессионного анализа со смешанными моделями (краткое сравнение статистических пакетов для смешанных моделей можно найти здесь: GLMM package comparison, 2015; r-sig-mixed-models FAQ – GLMM, 2015).

В *lme4*, как и во многих других пакетах в R, для задания модели используется синтаксис в виде «формулы», схожей с регрессионным уравнением. Например, для модели, включающей фиксированный эффект для *x*, фиксированный эффект для *z* и случайный эффект для *x* и среднего с группировкой по испытуемым, формула будет выглядеть как $y \sim x + z + (x | Subject)$. Здесь видно, что зависимая переменная отделяется от предикторов тильдой, предикторы разделяются знаком плюс, а группировка случайных эффектов задается в скобках с вертикальной чертой, отделяющей группирующую переменную. Случайные и фиксиро-

ванные эффекты для среднего (*Intercept*) предполагаются по умолчанию, поэтому в явном виде их указывать необязательно. В таблице 2 представлены примеры формул для записи некоторых часто встречающихся вариантов случайных эффектов в *lme4*.

Смешанная модель (*Akaike information criterion*, $AIC = 5999$) в целом лучше описывала данные, чем простой дисперсионный анализ без усреднения ($AIC = 6500$). Сначала рассмотрим данные о случайных эффектах. В таблице 3 в первом столбце представлены данные о «группах» случайных эффектов. В полученной модели случайные эффекты разбиты на две группы: к первой группе (*Stimulus*) относится только эффект стимула, ко второй (*Subject*) — эффект испытуемого и эффекты разного угла наклона прямой регрессии по испытуемым для правильных и неправильных ответов (*Accuracy*) и стимулов разного типа (*Stimulus Type*). В данной модели были применены уровневые (treatment) контрасты. В четвертом и пятом столбцах показаны стандартные отклонения и корреляции между случайными эффектами. Доля изменчивости по испытуемым (.43) значительно выше, чем по стимулам (.04). Это означает, что время ответа в значительно большей степени варьируется между испытуемыми, нежели между стимулами.

Кроме того, существует высокая негативная корреляция (–.86) между случайным эффектом испытуемого (*Intercept*) и случайным эффектом для правильности ответа по испытуемым (*Accuracy*). В то же время взаимосвязь случайного эффекта испытуемого и случайного эффекта типа стимула по испытуемым низкая (.33). Поскольку используются уровневые контрасты, то случайный эффект *Intercept* в данной модели отражает сдвиг времени ответа испытуемого относительно группового среднего при условии, что ответ неправильный ($Accuracy = 0$) и предъявлена фотография женщины ($Stimulus Type = F$). Это затрудняет непосредственный анализ обнаруженной негативной корреляции.

Для того чтобы прояснить ее природу, был проведен повторный анализ с исключением случайного эффекта взаимодействия испытуемого и типа стимула. Корреляция эффектов *Intercept* и *Accuracy* осталась высокой (–.76). Поскольку в модели осталось всего два случайных эффекта для испытуемого, можно говорить о том, что время неправильных ответов негативно взаимосвязано с различием времени правильных и неправильных ответов. Иначе говоря, если неправильные ответы испытуемого медленнее неправильных ответов по группе (значение *Intercept* будет соответственно выше), то разница по времени правильных и неправильных ответов будет меньше. Эта взаимосвязь отражена на рисунке 2.

Применение смешанных моделей позволяет узнать, насколько необходимо добавление случайного эффекта, путем сравнения моделей с включенным и исключенным случайным эффектом. В таблице 4 представлены результаты такого сравнения для случайных эффектов угла наклона прямой регрессии по испытуемым и эффекта стимула. Удаление эффекта стимула не приводит к значимым различиям в объяснительной силе модели, в отличие от эффектов угла наклона. Это неудивительно, учитывая низкие значения диспер-

сии эффекта стимула (таблица 3). Поэтому данный эффект можно было бы исключить из дальнейшего анализа, но поскольку эффекты для включения в модель выбирались исходя из теоретических соображений, дальнейший анализ также проводился с включением случайного эффекта стимула.

Значения фиксированных эффектов для итоговой модели представлены в таблице 5. В библиотеке *lme4* не реализован расчет p -уровней для моделей с гауссовым распределением зависимой переменной, что обусловлено теоретической неоднозначностью расчетов числа степеней свободы (Bates, 2006). Тем не менее существуют возможные пути решения данной проблемы, в частности:

1. Сравнение моделей на основе отношения правдоподобия (*likelihood ratio test*, LRT).
2. Оценка значимости с расчетом степеней свободы с использованием аппроксимации по методу Кенварда-Роджера, реализованная в библиотеке *pbkrtest* (Halekoh, Nøjsgaard, 2014).
3. Параметрические методы бутстрепа, также реализованные в библиотеке *pbkrtest*.
4. Оценка значимости с расчетом степеней свободы с использованием аппроксимации по методу Саттертуэйта, реализованная в библиотеке *lmerTest* для проверки гипотез с использованием суммы квадратов типа III (Kuznetsova et al., 2014).

LRT обычно считается менее надежным подходом, чем аппроксимация по методу Саттертуэйта, которая в свою очередь менее надежна, чем метод Кенварда-Роджера или параметрический бутстреп (*r-sig-mixed-models FAQ–GLMM*, 2015).

Я использовал три первых метода. Сравнение моделей производилось по аналогии с вычислением суммы квадратов типа II в дисперсионном анализе. Все методы расчета p -уровня показали практически идентичные результаты.

Таким образом, результат применения смешанных моделей оказался близок к результатам дисперсионного анализа «по испытуемым» и результатам анализа сырых данных, что говорит о надежности данного метода.

Заключение

Когда есть вероятность влияния различий между стимулами или между испытуемыми на полученные эффекты, для анализа данных можно применить несколько подходов. Дисперсионный анализ «по стимулам» позволяет учесть различия между стимулами, однако не учитываются различия между испытуемыми. Анализ «по испытуемым» позволяет учесть различия между испытуемыми, но теряются различия между стимулами. При этом в обоих случаях используется предварительное усреднение данных, что потенциально может приводить к потерям в мощности анализа. Дисперсионный анализ сырых данных не имеет проблем с потерей мощности в результате усреднения, но и не учитывает наличия связанных групп наблюдений. Смешанные модели позволяют учесть оба типа различий. Это не делает квазиэксперимент экспери-

ментом, но снижает вероятность получения результатов, искаженных за счет различий между стимулами или различий между испытуемыми. Кроме того, даже в обычном эксперименте это позволит проконтролировать возможное влияние побочных переменных.

Проведенное сравнение методов на материале квазиэксперимента с измерением времени ответа показывает, что результаты анализа с помощью смешанных линейных моделей в целом согласуются с результатами дисперсионного анализа с усреднением по испытуемым. Однако у смешанных моделей есть ряд преимуществ. В частности, они позволяют проверить предположение о необходимости учета тех или иных случайных факторов (в рассмотренном случае фактор стимула оказался излишним) и получить данные о взаимосвязях между случайными факторами, что позволило показать негативную корреляцию между отклонением времени неправильных ответов испытуемого от среднего времени неправильных ответов и отклонением от среднего разницы по времени между правильными и неправильными ответами. Чем медленнее ошибочные ответы испытуемых, тем меньше различие по времени правильных и неправильных ответов, то есть тем слабее эффект замедления на ошибочных ответах. Кроме того, поскольку в смешанных моделях в явном виде учитываются возможные источники искажения в данных, это позволяет заранее отвергнуть критику, связанную с этими источниками, возможную в случае дисперсионного анализа сырых данных. Таким образом, в описанном случае смешанная модель не показывает различий с дисперсионным анализом в результатах статистических тестов для отдельных факторов, однако позволяет с большей уверенностью делать выводы о результатах исследования. Кроме того, необходимо помнить, что априорных оснований предпочитать анализ «по стимулам» анализу «по испытуемым» нет, хотя мы с тем же успехом могли получить ситуацию, когда бы сильное влияние оказывал фактор стимула, а не фактор испытуемого. В этом случае анализ «по испытуемым» дал бы ошибочные результаты. Исходя из этого, использование смешанных моделей с точки зрения автора оказывается вполне целесообразным — в худшем случае они покажут те же результаты, что и дисперсионный анализ, а в лучшем позволят избежать грубых ошибок.

Применение смешанной регрессии оставляет исследователю значительно больше свободы в выборе эффектов для включения в модель. При этом в случае простых моделей, где случайные эффекты сгруппированы только по одной переменной (например, по испытуемым), рекомендуется включать в случайные эффекты все те переменные, которые используются среди фиксированных эффектов и для которых есть повторные измерения. Например, если речь идет об обычном внутригрупповом плане 2×2 , то обе независимые переменные и их взаимодействие должны быть включены в модель как в качестве фиксированных, так и в качестве случайных эффектов с группировкой по испытуемым. В противном случае увеличивается вероятность ложнопозитивных результатов (Barr et al., 2013). Однако когда речь идет о более сложных моделях, то возникает ряд затруднений. При увеличении числа случай-

ных эффектов растет время расчетов и вероятность проблем подгонки моделей. Рекомендация автора — исходить из теоретических представлений о том, какие факторы будут оказывать влияние, и использовать один и тот же подход для всей серии исследований. Может возникнуть впечатление, что подобные рекомендации оставляют слишком большой простор для манипуляций. Однако нужно помнить, что проблема выбора модели характерна не только для регрессии со смешанными эффектами. Например, дисперсионный анализ также позволяет включать в модель дополнительные предикторы (например, пол испытуемого, возраст, среднее время ответа, номер пробы и т. д.). Регрессионный подход всего лишь делает эту проблему явной, вынуждая исследователя задумываться о предпосылках, лежащих в основе его действий. Наконец, один статистический результат сам по себе не подтверждает и не опровергает теоретическую гипотезу. Только серия исследований может позволить сделать это, а в таком случае модель, произвольно подогнанная ради получения заветной значимости в одном исследовании, вряд ли покажет тот же результат в другом исследовании.

Литература

- Королькова О. А. Перцептивное пространство и предикторы различия эмоциональных экспрессий лица // Российский журнал когнитивной науки. 2014. Т. 1. № 4. С. 82–97. URL: <http://cogjournal.org/1/4/pdf/KorolkovaRJCS2014b.pdf>.
- Bakker M., Wicherts J. M. The (mis)reporting of statistical results in psychology journals // Behavior Research Methods. 2011. Vol. 43. No. 3. P. 666–678. doi:10.3758/s13428-011-0089-5
- Baldwin S. A., Berkefjon A., Atkins D. C., Olsen J. A., Nielsen S. L. Rates of change in naturalistic psychotherapy: Contrasting dose-effect and good-enough level models of change // Journal of Consulting and Clinical Psychology. 2009. Vol. 77. No. 2. P. 203–211. doi:10.1037/a0015235
- Barr D. J., Levy R., Scheepers C., Tily H. J. Random effects structure for confirmatory hypothesis testing: Keep it maximal // Journal of Memory and Language. 2013. Vol. 68. No. 3. P. 255–278. doi:10.1016/j.jml.2012.11.001
- Bates D. [R] lmer, p-values and all that // R-help mailing list. 2006. URL: <https://stat.ethz.ch/pipermail/r-help/2006-May/094765.html>.
- Bates D., Maechler M., Bolker B., Walker S. lme4: Linear mixed-effects models using Eigen and S4 2014. URL: <http://cran.r-project.org/package=lme4>
- Biesanz J. C., Human L. J. The cost of forming more accurate impressions: accuracy-motivated perceivers see the personality of others more distinctively but less normatively than perceivers without an explicit goal // Psychological Science. 2010. Vol. 21. No. 4. P. 589–594. doi:10.1177/0956797610364121
- Chetverikov A. Warmth of familiarity and chill of error: Affective consequences of recognition decisions // Cognition & Emotion. 2014. Vol. 28. No. 3. P. 385–415. doi:10.1080/02699931.2013.833085
- Clark H. H. The language-as-fixed-effect fallacy: A critique of language statistics in psychological research // Journal of Verbal Learning and Verbal Behavior. 1973. Vol. 12. P. 335–359. doi:10.1016/S0022-5371(73)80014-3
- Gelman A. Analysis of variance: Why it is more important than ever // The Annals of Statistics. 2005. Vol. 33. No. 1. P. 1–53. doi:10.1214/009053604000001048
- Gelman A., Hill J. Data analysis using regression and multi-level/hierarchical models. New York, USA: Cambridge Univ Press, 2007.
- GLMM package comparison 2015. URL: <http://glmm.wikidot.com/pkg-comparison>
- Gulbinaitė R., Johnson A., de Jong R., Morey C. C., van Rijn H. Dissociable mechanisms underlying individual differences in visual working memory capacity // NeuroImage. 2014. Vol. 99. P. 197–206. doi:10.1016/j.neuroimage.2014.05.060
- Halekoh U., Højsgaard S. A Kenward-Roger approximation and parametric bootstrap methods for tests in linear mixed models – The R package pbrtest // Journal of Statistical Software. 2014. Vol. 59. No. 9. P. 1–30. URL: <http://www.jstatsoft.org/v59/i09/>
- Jaeger T. F. Categorical data analysis: Away from ANOVAs (transformation or not) and towards Logit Mixed Models // Journal of Memory and Language. 2008. Vol. 59. No. 4. P. 434–446. doi:10.1016/j.jml.2007.11.007
- Jara A., Hanson T., Quintana F., Müller P., Rosner G. DDPackage: Bayesian semi- and nonparametric modeling in R // Journal of Statistical Software. 2011. Vol. 40. No. 5. P. 1–30. URL: <http://www.jstatsoft.org/v40/i05/>.
- Kristjánsson Árnir Jóhannesson Ó. I. How priming in visual search affects response time distributions: Analyses with ex-Gaussian fits // Attention, Perception & Psychophysics. 2014. Vol. 76. No. 8. P. 2199–2211. doi:10.3758/s13414-014-0735-y
- Kuznetsova A., Brockhoff P. B., Christensen R. H. B. lmerTest: Tests in Linear Mixed Effects Models 2014. URL: <http://cran.r-project.org/package=lmerTest>.
- Laubrock J., Engbert R., Kliegl R. Fixational eye movements predict the perceived direction of ambiguous apparent motion // Journal of Vision. 2008. Vol. 8. No. 14. P. 13.1–13.17. doi:10.1167/8.14.13
- Mousikou P., Kinoshita S., Wu S., Norris D. Transposed-letter priming effects in reading aloud words and nonwords // Psychonomic Bulletin & Review. 2015. Vol. Online first. P. 1–6. doi:10.3758/s13423-015-0806-7
- Oberauer K., Kliegl R. A formal model of capacity limits in working memory // Journal of Memory and Language. 2006. Vol. 55. P. 601–626. doi:10.1016/j.jml.2006.08.009
- Pinheiro J. C., Bates D. M. Mixed effects models in S and S-Plus. New York: Springer Verlag, 2000. doi:10.1198/tech.2001.s574
- R Core Team R: A language and environment for statistical computing. Vienna, Austria: R Foundation for Statistical Computing, 2014. URL: <http://www.r-project.org/>.
- Raaijmakers J. G. W. A further look at the “language-as-fixed-effect fallacy” // Canadian Journal of Experimental Psychology. 2003. Vol. 57. No. 3. P. 141–151. URL: <http://www.ncbi.nlm.nih.gov/pubmed/14596473>
- Raaijmakers J. G. W., Schrijnemakers J. M. C., Gremmen F. How to deal with “the language-as-fixed-effect fallacy”: Common misconceptions and alternative solutions // Journal of Memory and Language. 1999. Vol. 41. No. 3. P. 416–426.
- r-sig-mixed-models FAQ–GLMM 2015. URL: <http://glmm.wikidot.com/faq>
- Sandberg K., Timmermans B., Overgaard M., Cleeremans A. Measuring consciousness: is one measure better than the other? // Consciousness and Cognition. 2010. Vol. 19. No. 4. P. 1069–1078. doi:10.1016/j.concog.2009.12.013
- Sarkar D., DebRoy S., Bates D., Pinheiro J., the R Core team nlme: Linear and nonlinear mixed effects models // R package version 3.1-90. R package version 3.1-90, 2008.
- Wright D. B., London K. Multilevel modelling: Beyond the basic applications // The British Journal of Mathematical and Statistical Psychology. 2009. Vol. 62. No. Pt 2. P. 439–456. doi:10.1348/000711008X327632
- Wright D. B., London K., Waechter M. Social anxiety moderates memory conformity in adolescents // Applied Cognitive Psychology. 2010. Vol. 24. No. 7. P. 1034–1045. doi:10.1002/acp
- Wu Y., Clark L. Disappointment and regret enhance corrugator reactivity in a gambling task // Psychophysiology. 2014. Vol. 52. No. 4. P. 518–523. doi:10.1111/psyp.12371

methods

Linear Mixed Effects Regression in Cognitive Studies

Andrey Chetverikov

Cognitive Research Lab, Russian Presidential Academy of National Economy and Public Administration (RANEPA), Moscow; Saint Petersburg State University, Saint Petersburg, Russia

Abstract. Data from cognitive studies can be difficult to analyze. Often, there is a lack of balanced design, omnipresent differences in stimuli or missing data cells. To control for the erroneous variance, researchers frequently aggregate the data by participant or by stimulus, and then analyze the averaged data using conventionally selected methods such as ANOVA. The present paper describes an alternative approach that utilizes linear mixed-effects regression (LMER). This approach does not involve data aggregation and its accompanying loss of statistical power. Instead, it provides researchers with an opportunity to incorporate assumptions about data covariance into the model. To demonstrate the benefits of this approach, I use data from a quasi-experimental study with response accuracy, participants' gender, and the gender of a recognized face as predictors of response times in a face recognition task. LMER results are compared to the results of ANOVA with aggregation by participant and by stimulus, and without aggregation. LMER results were similar to ANOVA with aggregation by participant and to ANOVA based on raw data. However, LMER allows researchers to make more solid conclusions since possible sources of error — differences between stimuli and between participants — are controlled within the regression model. The implications of using LMER for data analysis and some practical issues related to it are discussed.

Correspondence: Andrey A. Chetverikov, a.chetverikov@psy.spbu.ru; Department of Psychology, St. Petersburg State University, 6 Makarova nab., 199034 St. Petersburg, Russia

Keywords: mixed models, regression, ANOVA, response times, quasi-experiment, errors

Copyright © 2015. Andrey A. Chetverikov. This is an open-access article distributed under the terms of the [Creative Commons Attribution License](https://creativecommons.org/licenses/by/4.0/) (CC BY), which permits unrestricted use, distribution, and reproduction in any medium, provided that the original author is credited and that the original publication in this journal is cited, in accordance with accepted academic practice.

Note. The article is partially based on an earlier paper in the conference proceedings volume: Chetverikov, A. A. (2013). [Application of the linear mixed models for the analysis of the quasi-experimental data]. In V. N. Afanasyev (Ed.) Proceedings of the 1st International Scientific Conference [“Formation of the main directions of development in contemporary statistics and econometrics”]. Vol. 1. (pp. 332–338). Orenburg: Universitet.

Acknowledgments. The research is supported by the Russian foundation for basic research (project # 15-06-07417 A) and Saint Petersburg State University (research project #8.38.287.2014).

Received 18 February 2015, accepted 21 March 2015.

References

- Bakker, M., & Wicherts, J.M. (2011). The (mis)reporting of statistical results in psychology journals. *Behavior Research Methods*, 43 (3), 666–678. doi:10.3758/s13428-011-0089-5
- Baldwin, S.A., Berkeljon, A., Atkins, D.C., Olsen, J.A., & Nielsen, S.L. (2009). Rates of change in naturalistic psychotherapy: Contrasting dose-effect and good-enough level models of change. *Journal of Consulting and Clinical Psychology*, 77 (2), 203–211. doi:10.1037/a0015235
- Barr, D.J., Levy, R., Scheepers, C., & Tily, H.J. (2013). Random effects structure for confirmatory hypothesis testing: Keep it maximal. *Journal of Memory and Language*, 68 (3), 255–278. doi:10.1016/j.jml.2012.11.001
- Bates, D. (2006). [R] lmer, p-values and all that. *R-help mailing list*. Retrieved January 18, 2015 from <https://stat.ethz.ch/pipermail/r-help/2006-May/094765.html>
- Bates, D., Maechler, M., Bolker, B., & Walker, S. (2014). lme4: Linear mixed-effects models using Eigen and S4. Retrieved from <http://cran.r-project.org/package=lme4>
- Biesanz, J.C., & Human, L.J. (2010). The cost of forming more accurate impressions: accuracy-motivated perceivers see the personality of others more distinctively but less normatively than perceivers without an explicit goal. *Psychological Science*, 21 (4), 589–594. doi:10.1177/0956797610364121
- Chetverikov, A. (2014). Warmth of familiarity and chill of error: Affective consequences of recognition decisions. *Cognition & Emotion*, 28 (3), 385–415. doi:10.1080/02699931.2013.833085

- Clark, H.H. (1973). The language-as-fixed-effect fallacy: A critique of language statistics in psychological research. *Journal of Verbal Learning and Verbal Behavior*, 12, 335–359. doi:10.1016/S0022-5371(73)80014-3
- Gelman, A. (2005). Analysis of variance: Why it is more important than ever. *The Annals of Statistics*, 33 (1), 1–53. doi:10.1214/009053604000001048
- Gelman, A., & Hill, J. (2007). *Data analysis using regression and multilevel/hierarchical models*. New York, USA: Cambridge Univ Press.
- GLMM package comparison (2015). Retrieved March 01, 2015 from <http://glmm.wikidot.com/pkg-comparison>
- Gulbinaite, R., Johnson, A., de Jong, R., Morey, C.C., & van Rijn, H. (2014). Dissociable mechanisms underlying individual differences in visual working memory capacity. *NeuroImage*, 99, 197–206. doi:10.1016/j.neuroimage.2014.05.060
- Halekoh, U., & Højsgaard, S. (2014). A Kenward-Roger approximation and parametric bootstrap methods for tests in linear mixed models — The R package pbkrtest. *Journal of Statistical Software*, 59 (9), 1–30. Retrieved from <http://www.jstatsoft.org/v59/i09/>
- Jaeger, T.F. (2008). Categorical data analysis: Away from ANOVAs (transformation or not) and towards Logit Mixed Models. *Journal of Memory and Language*, 59 (4), 434–446. doi:10.1016/j.jml.2007.11.007
- Jara, A., Hanson, T., Quintana, F., Müller, P., & Rosner, G. (2011). DPPackage: Bayesian semi- and nonparametric modeling in R. *Journal of Statistical Software*, 40 (5), 1–30. Retrieved from <http://www.jstatsoft.org/v40/i05/>
- Korolkova, O.A. (2014). Perceptual space and predictors of emotional facial expression discrimination. *The Russian Journal of Cognitive Science*, 1 (4), 82–97. Retrieved from <http://cogjournal.org/1/4/pdf/KorolkovaRJCS2014b.pdf>
- Kristjánsson Árne, & Jóhannesson, Ó.I. (2014). How priming in visual search affects response time distributions: Analyses with ex-Gaussian fits. *Attention, Perception & Psychophysics*, 76 (8), 2199–2211. doi:10.3758/s13414-014-0735-y
- Kuznetsova, A., Brockhoff, P.B., & Christensen, R.H.B. (2014). lmerTest: Tests in Linear Mixed Effects Models. Retrieved from <http://cran.r-project.org/package=lmerTest>
- Laubrock, J., Engbert, R., & Kliegl, R. (2008). Fixational eye movements predict the perceived direction of ambiguous apparent motion. *Journal of Vision*, 8 (14), 13.1–13.17. doi:10.1167/8.14.13
- Mousikou, P., Kinoshita, S., Wu, S., & Norris, D. (2015). Transposed-letter priming effects in reading aloud words and nonwords. *Psychonomic Bulletin & Review*, Online first, 1–6. doi:10.3758/s13423-015-0806-7
- Oberauer, K., & Kliegl, R. (2006). A formal model of capacity limits in working memory. *Journal of Memory and Language*, 55, 601–626. doi:10.1016/j.jml.2006.08.009
- Pinheiro, J.C., & Bates, D.M. (2000). *Mixed effects models in S and S-Plus*. New York: Springer Verlag. doi:10.1198/tech.2001.s574
- R Core Team (2014). *R: A language and environment for statistical computing*. Vienna, Austria: R Foundation for Statistical Computing. Retrieved from <http://www.r-project.org/>.
- Raaijmakers, J.G.W. (2003). A further look at the “language-as-fixed-effect fallacy”. *Canadian Journal of Experimental Psychology*, 57 (3), 141–151. Retrieved from <http://www.ncbi.nlm.nih.gov/pubmed/14596473>
- Raaijmakers, J.G.W., Schrijnemakers, J.M.C., & Gremmen, F. (1999). How to deal with “the language-as-fixed-effect fallacy”: Common misconceptions and alternative solutions. *Journal of Memory and Language*, 41 (3), 416–426.
- r-sig-mixed-models FAQ–GLMM (2015). Retrieved January 18, 2015 from <http://glmm.wikidot.com/faq>
- Sandberg, K., Timmermans, B., Overgaard, M., & Cleeremans, A. (2010). Measuring consciousness: is one measure better than the other? *Consciousness and Cognition*, 19 (4), 1069–1078. doi:10.1016/j.concog.2009.12.013
- Sarkar, D., DebRoy, S., Bates, D., Pinheiro, J., & the R Core team (2008). nlme: Linear and nonlinear mixed effects models. *R package version 3.1-90*
- Wright, D.B., & London, K. (2009). Multilevel modelling: Beyond the basic applications. *The British Journal of Mathematical and Statistical Psychology*, 62 (Pt 2), 439–456. doi:10.1348/000711008X327632
- Wright, D.B., London, K., & Waechter, M. (2010). Social anxiety moderates memory conformity in adolescents. *Applied Cognitive Psychology*, 24 (7), 1034–1045. doi:10.1002/acp.1604
- Wu, Y., & Clark, L. (2014). Disappointment and regret enhance corrugator reactivity in a gambling task. *Psychophysiology*, 52 (4), 518–523. doi:10.1111/psyp.12371