Animal Behaviour 77 (2009) 753–758

Contents lists available at ScienceDirect

Animal Behaviour

journal homepage: www.elsevier.com/locate/yanbe

# A simple method for distinguishing within- versus between-subject effects using mixed models

Martijn van de Pol*, Jonathan Wright

*Centre for Conservation Biology, Department of Biology, Norwegian University of Science and Technology (NTNU)*

ARTICLE INFO

*Article history:*
Received 15 September 2008
Initial acceptance 29 October 2008
Final acceptance 5 November 2008
Published online 12 December 2008
MS. number: 08-00599

*Keywords:*
adaptive hypothesis
ecological fallacy
hierarchical model
hypothesis testing
individual heterogeneity
multilevel data
random effect
reaction norm
repeated measure
within-group centering

Here we describe a statistical procedure called within-subject centering (not to be confused with grand-mean centering; e.g. Kreft et al. 1995). This simple technique can be used in mixed models to separate within-subject effects (i.e. phenotypically plastic or facultative behavioural responses) from between-subject effects (i.e. evolutionarily fixed behavioural responses based on the individual or its class). Such a separation is important as it allows us to distinguish between alternative biological hypotheses and prevents us from erroneously generalizing within-subject relationships to between-subject relationships, or vice versa. We claim no originality for this statistical technique, which is commonly used in the social sciences (e.g. Davis et al. 1961; Raudenbush 1989; Kreft et al. 1995; Snijders & Bosker 1999; see also van de Pol & Verhulst 2006). However, we offer it as a piece of overlooked statistical methodology that we think is crucial to many researchers in animal behaviour, and in various other areas of biology as well. We illustrate our explanation of the technique with several biological examples and simulated data, but this method is widely applicable and most readers will probably be able to identify appropriate examples from their own research.

## LEVELS OF AGGREGATION AND MIXED MODELS

In biology in general and in the study of animal behaviour in particular we usually collect data that have different levels of aggregation. For example, one might collect multiple measurements of the same individual (i.e. subject), measure multiple members (e.g. offspring, helpers) of the same collective (e.g. nest, social group), or measure multiple individuals from the same area or time period. An important feature of such multilevel data is that measurements within a level of aggregation are often not independent. For example, some individuals might consistently behave differently from other individuals, and therefore the values of measurements of the same individuals are intercorrelated. This issue becomes particularly important when testing adaptive evolutionary hypotheses, because typically the individual subject represents *the* unit of analyses. We therefore have to be careful to avoid the problem of pseudoreplication (artificially inflating one's degrees of freedom) in any statistical analysis involving repeated observations of the same subjects (e.g. Snijders & Bosker 1999; Goldstein 2003).

---

* Correspondence: M. van de Pol, Centre for Conservation Biology, Department of Biology, NTNU, Trondheim N-7491, Norway.

*E-mail address:* m.van.de.pol@myscience.eu (M. van de Pol).

0003-3472/$38.00 © 2008 The Association for the Study of Animal Behaviour. Published by Elsevier Ltd. All rights reserved.
doi:10.1016/j.anbehav.2008.11.006

The simplest solution to overcome this problem is to aggregate all measurements of the same subject into an average value and make these the unit of analysis. However, this approach is usually undesirable, first because it omits potentially useful information on variation within subjects, and second because it rules out any investigation of hypotheses that specifically address questions concerning within-subject variation. Luckily two types of statistical models are available to account for interdependency and structuring of data: repeated measures models and mixed models. Of these two types of models, mixed models are by far the most generally applicable, because they can (1) easily account for complex data structures with various levels of aggregation (i.e. hierarchical linear mixed models) and (2) effectively deal with unbalanced data sets (i.e. situations with varying numbers of measurements per subject; Snijders & Bosker 1999; Goldstein 2003; Gelman & Hill 2006).

The basic idea of mixed models is that when data have different levels of aggregation (e.g. multiple measurements of the same individual), the identifier of the level of aggregation (e.g. individual identity) is included as a random intercept (and sometimes also as a random slope; see later). This random intercept allows us to separate the total variance in the dependent variable into a within- and between-subject (e.g. individual) variance component. If we then add a fixed predictor variable to this model we can investigate how much of the total variation in the dependent variable is explained by the fixed predictor variable, while accounting for the fact that measurements of the same subject are intercorrelated (Gelman & Hill 2006).

## THE PROBLEM

Mixed models are now becoming widely used in biology, as they are ideally suited for analysing behavioural, ecological and evolutionary data (because such data sets typically have one or more levels of aggregation). Evolutionary and ecological hypotheses often specifically address either within- or between-subject differences, and thus it is important to be able to construct statistical models that produce unambiguous and unbiased estimates of such within- and between-subject effects. However, when talking to biologists that use mixed models we have noticed that there is some confusion concerning the extent to which fixed predictor variables quantify within- or between-subject effects on response variables. In fact, there seems to be a common misconception that because random effects account for consistent differences between subjects in the value of the dependent variable ($y$), this automatically means that any effect of a predictor variable ($x$) must reflect a within-subject effect. However, as we show below this is not necessarily the case, and it can lead to what sociologists and economists call an ‘ecological fallacy’, in which a relationship described at one level of aggregation is erroneously generalized to another level (Robinson 1950).

The problem lies in the fact that fixed predictor variables ($x$), as well as dependent variables ($y$), can vary at multiple levels of aggregation. This is because in nonexperimental studies we typically cannot choose the value of the predictor variable to be the same for all subjects, and we have to rely on natural variation in the predictor variable. For example, it is very possible that we will only be able to measure $y$ for some subjects at low values of $x$, while for other subjects we will only be able to measure $y$ at high values of $x$. Although random effects (intercepts) in mixed models can account for between-subject variation in $y$, they do not automatically account for possible between-subject variation in $x$. Thus, in cases where there is between-subject variation in $x$, an association between a predictor and response variable could be caused by a within-subject effect of $x$ on $y$, but also by a between-subject effect of $x$ on $y$ (subjects with a high $\bar{x}$ also have a high $\bar{y}$). As long as the within- and between-subject effects of $x$ on $y$ are similar, estimates from standard mixed models will indeed reflect the within-subject effect in the presence of between-subject variation in $x$ (Fig. 1a). However, in cases where we have between-subject variation in $x$, and the within- and between-subject effects of $x$ on $y$ are different (Fig. 1b–d), predictor variables in standard mixed models do not simply reflect within-subject effects (as commonly thought), but reflect a combination of within- and between-subject effects.

The biological relevance of this problem is illustrated by the following example from the classic question in cooperative breeding of whether genetic relatedness affects helping behaviour (see Brown 1987; Cockburn 1998; Koenig & Dickinson 2004). Typically, such studies have measured relatedness to the brood ($x$) and provisioning effort ($y$) of the same helpers (subjects) over several breeding attempts (i.e. unbalanced repeated measures data). Usually, some helpers will have a consistently higher provisioning effort than others and helper identity is included as a random intercept in the model to account for the fact that there is between-helper variation in provisioning effort (i.e. measurements of the same helper at different nests are interdependent). Moreover, usually some helpers will have a higher overall genetic relatedness ($x$) to the broods at which they help than others, for example, because some helpers are local offspring while others are immigrants. Consequently, if a positive relationship between relatedness and helping effort is found, this could be caused by: (1) a between-subject effect in which some more-related helpers always help at high levels, while other less-related helpers always help at low levels (as in Fig. 1b); or (2) a within-subject effect in which each helper provisions at higher levels at nests where it is more related and at lower levels at other nests where it is less related (as in Fig. 1c); or (3) a combination of both these within- and between-subject effects (as in Fig. 1a). From a biological perspective these patterns represent two different but not mutually exclusive biological mechanisms and hypotheses: (1) that helpers that are more related (e.g. those in their natal group) always help at higher levels irrespective of what nest they attend in that group; and (2) that helpers use specific cues concerning variation in kinship at different nests within the same group and facultatively adjust their helping effort accordingly. Importantly, neither of these hypotheses is actually explicitly tested in itself when using relatedness as a predictor variable in standard mixed models, because the result is potentially a mix of the two. In Table 1, we give some additional biological examples in which distinguishing between within- and between-subject hypotheses has been shown to be important (for a very illustrative example see Salomons et al. 2006). These examples aim to illustrate how general this problem is in ecological, evolutionary and behavioural studies and to help readers to identify similar examples from their own research.

Whether the effect of a predictor variable in standard mixed models is dominated by a within- or a between-subject effect depends on the structure of the unbalanced data set, and in particular how many different observations we have for how many different subjects, and over what range of circumstances per subject relative to the full range of values for the predictor variable. It is also important to note that there is no guarantee that the within- and between-subject effects contained within a single predictor variable go in the same direction. This raises the interesting possibility that a nonsignificant fixed effect may actually contain two significant effects (see Fig. 1d): a within-subject effect in one direction, and a between-subject effect in the opposite direction, with the two effectively cancelling each other out in the main association reported by the standard mixed model (see Table 2, Fig. 2d). Opposing within- and between-subject effects may in fact not be uncommon. For example (life history) trade-offs result in negative within-individual association between two variables, but between-individual quality/acquisition differences often cause some individuals to score high on both variables and