Rethinking analyses of crossed effects experiments in marketing communications research

Jose Tummers[1,2], Kim Janssens[1,3]

[1] Leuven University College
[2] KULeuven, Quantitative Lexicology and Variational Linguistics
[3] KULeuven, School of Mass Communication Research

Although Repeated Measures ANOVA is generally used to analyze experimental designs, this technique does not suffice to describe all variance in a crossed effects experiment. Responses are generated from the same subjects and simultaneously those responses will be collected for the same stimuli, exposing the independence of the observations and the generalizability of the results. We address this methodological concern by fitting a mixed-effects model to reanalyze the outcomes of an experiment in marketing communications. In that experiment, a RM ANOVA was used to analyze the impact of a condition and a treatment factor on the recall of products displayed on a computer screen and where the within-subject variance was a random effect (Janssens et al., 2011). Although no major impact on the fixed effects was identified, the interaction between the experimental condition and the treatment remained significant, the mixed-effects model with two random effect terms outweighs a RM ANOVA with only one random effect term for subject. It significantly reduces the overall variance and significantly improves the predictive power of the model. Additionally, the intra-class correlation reveals that the random effect term for the stimuli explains 49.14% of the variance compared to only 7.93% for the subjects.

1 Introduction

In marketing communication, there is a burgeoning interest in experimental approaches (Alhabash & Wise, 2012; Janssens et al., 2011; Krakowiak & Oliver; 2012; Mazodier et al., 2012) which allow to detect more conclusive causal relations than it is generally the case in observational and survey-based research. It is common practice to use a Repeated Measures ANOVA (RM ANOVA) to account for the correlation or non-independence of the experimental outcomes. In the current study, we argue that RM ANOVA is not the technique best suited to analyze the results of a crossed effects experiment, since it only allows to model one grouping factor, generally the subjects, discarding the variance caused by the stimuli and hence jeopardizing the generalizability of the results, since one source of variance is excluded from the model.

We will reanalyze an existing experiment in marketing communication research (Janssens et al., 2011) to demonstrate that mixed-effects models (Gelman & Hill, 2007; Pinheiro & Bates, 2010) are better suited to analyze crossed effects experiments. That is, both subjects and stimuli can simultaneously be included in the model as random effect terms, next to the fixed effect terms representing the experimental condition and treatment factor (Baayen et al., 2008; Quené & Van den Berg, 2008; Richter, 2006).

2 Problem statement: sources of variance in experimental designs with crossed effects

In a crossed effects design, all subjects participating in the experiment respond to the same set of stimuli. Consequently, the observations display a twofold dependency since they are grouped under the subjects, each subject responding to the same set of stimuli, as well as under the stimuli, each stimulus being responded to by all subjects participating in the experiment. Therefore a statistical model is required that includes both sources of variance simultaneously in order to generalize the results beyond the sample of subjects as well as beyond the sample of stimuli used in the experiment. This is impossible in a RM ANOVA where only one source of variance is modeled, generally the subjects. In psychological research, different solutions have been proposed to overcome this problem, such as the F1 and F2 statistics and the F'_{min} statistic. However, all these solutions face different problems that can be overcome by using mixed-effects models (see Baayen et al. (2008), Quené & Van den Bergh (2008) and Richter (2006) for a discussion). Finally, the sphericity and homoscedasticity assumptions of RM ANOVA, which appear not always to be explicitly tested in actual research, do not apply to mixed-effects models.

Mixed-effects models are characterized by the combination of fixed and random effect terms (Gelman & Hill, 2007; Pinheiro & Bates, 2000). Fixed effect terms exhaust all levels of a parameter. Hence, their values cover all values in the population, such as gender whose values can be male or female. Random effect terms are sampled from a larger population and therefore only represent a sample of the actual population, that will probably differ in a replication study. Fixed effect terms are modeled by means of contrasts and random effect terms are modeled as random variables with 0 as mean and an unknown variance ($N(0,\sigma^2)$). Mixed-effects models allow to account for the non-independence of observations by inclusion of random effects corresponding to the grouping variables, viz. the subjects and the stimuli in the experiment, so that correlations between observations are directly modeled.

In this contribution, we will fit a random intercept model, where a separate intercept is estimated for every value of the grouping variables, viz. the subjects and the stimuli in the experimental design. This boils down to a correction of the intercept for each subject and each stimulus according to the deviation of their variance to the overall variance (represented by the fixed intercept which is the mean of all random intercepts).

3 Case study

3.1 Original experiment

The main purpose of the original experiment was to investigate whether exposure to a mating cue induces perceptual readiness (Janssens et al., 2011). To this end, subjects were briefly exposed to ten visual stimuli. Each stimulus was a display consisting of six products. Subsequently the subjects were instructed to list as many products as they could recall (Roskos-Ewoldsen & Fazio, 1992). Of the six products within a given display, one product was associated with a high status. The hypothesis predicts that exposure to a mating cue will automatically divert attention towards that object in the visual display that evokes a high social status and that this will be especially the case for single men as opposed to men involved in a committed relationship. Male subjects were exposed to either a sexily dressed female experimenter or a plainly dressed female experimenter before engaging in a visual status display task.

Fournier & Richins (1991) found that materialists often prefer consumption of high status associated goods. In a similar vein, Richins (1994) showed that people high in materialism are more likely to place greater importance to expensive goods and goods that communicate prestige than people low in materialism. Thus, it seems very reasonable that attention to high status goods, often expensive luxury items (Richins & Dawson, 1992; Wang & Wallendorf, 2006), may be a suitable indicator of materialism. Therefore, a computerized visual display task was used to measure participants' attention to status (Roskos-Ewoldsen & Fazio, 1992). This task measures the attention people have for certain displayed objects. To be more precise, this visual attention task was used to measure people's attention to and interest in high status goods. Participants were instructed that they would be exposed to six products displayed on a computer screen for a short period of time and were asked to recall as many products as possible. They were exposed to ten different displays, each consisting of six different product pictures. Each display remained on the screen for one second and comprised one picture of a status product (e.g. Breitling watch, Porsche, exclusive mansion) and five pictures of functional products (e.g. stapler, towel), randomly arranged in a circle on the computer screen. After exposure to each display, participants had 25 seconds to write down as many products as possible after which they were exposed to the next display.

The experimental condition, viz. the presence or absence of a mating cue, was manipulated through the clothing of the female who led the experiment. Two conditions were created, to which subjects were randomly assigned: the experimenter was either plainly (control condition) or sexily dressed (mating cue condition).

To check to what extent the subjects were involved in a (serious) relationship, their relationship status was asked. Responses were made using a asymmetrical 7-point Likert scale, ranging from 1 (*I am single*) to 7 (*I am married*). Each participant received a dichotomous relationship status score: single (n = 72) (responses < 3) or in a committed relationship (n = 61) (responses > 3). Subjects who responded 3 (n = 5) were eliminated from the dataset.

One hundred and thirty-three male heterosexual students participated, varying in age from 17 to 32 years (M = 20, SD = 1.79). Of these participants 70 (52.6%) were assigned to the control condition (exposed to plainly dressed young woman), and 63 (47.4%) to the mating cue condition (exposed to sexily dressed young woman).

3.2 Data matrix adaptation

The original data matrix has to be adapted since both the subjects and the stimuli have to be simultaneously included as parameters in the model equation in order to appropriately structure the crossed effects design of the experiment. In the RM ANOVA with an error term for subjects, an aggregate stimulus score is computed per subject over the ten stimuli by adding the successful recalls of the status product. This procedure suggests that every subject was presented only one stimulus rather than that the same ten stimuli were presented to every subject, as it was actually the case in the original experiment.

To fit a mixed-effects model, every combination of subjects and stimuli has to be made explicit in the data matrix to reflect the variance due to both the subjects and the stimuli. This results in an expanded data matrix: instead of *i* rows (*n* = 133), viz. one per subject, the matrix consists of *i* x *j* rows (*n* = 1,330), viz. one for each subject x stimulus combination. When generating a contingency table for subjects (row variable) and stimuli (column variable), the crossed effects design of the experiment – every subject responding once to every stimulus – now clearly shows up, as illustrated in table 1:

	s01	s02	s03	s04	s05	s06	s07	s08	s09	s10
p1097	1	1	1	1	1	1	1	1	1	1
p4420	1	1	1	1	1	1	1	1	1	1
p5626	1	1	1	1	1	1	1	1	1	1
p6001	1	1	1	1	1	1	1	1	1	1
p6313	1	1	1	1	1	1	1	1	1	1
…	…	…	…	…	…	…	…	…	…	…

Table 1: Contingency table summarizing the crossed effects design in adapted data matrix

Moreover, the level of measurement of the response variable is changed from numerical (sum of successful recalls of the 10 high status products per subject) into binomial (success vs. failure for the recall of the high status product for each stimulus by each subject). Consequently, we need the mixed-effects variant of a logistic regression, a generalized linear mixed-effects model (GLMM) (Gelman & Hill, 2007), with the odds $^{\text{success}}/_{\text{failure}}$ for the recall of the high status product as response variable.

3.3 Case study reanalyzed

The original study using the RM ANOVA with subjects as random effect term will be reanalyzed by means of a GLMM with relationship status (`relation`), mating cue (`condition`) and their interaction as fixed effect terms and subject (`subject`) and stimulus (`stimulus`) as random effect terms. The GLMMs will be fitted by means of the `lme4` library (Bates, 2005; Bates et al., 2013) in R (R Core Team, 2012). To be more precise, two GLMM models will be estimated. First, a model with only subject as random effect (`glmm1`) will be fitted as non-numerical pendant of the RM ANOVA. Next, the full GLMM model will be fitted with both subject and stimulus as random effects (`glmm2`). Both models are random intercept models, where only the intercept can vary over the values of the random effect terms.

We start by inspecting the coefficients of the fixed effect terms. The data in table 2 show that the major split is caused by the removal of the aggregation over the stimuli, since the main effect of relation is significant in both GMLLs as opposed to the RM ANOVA. Furthermore, the differences between both GLMMs are limited to moderate modifications of the size of the regression coefficients.

Parameter	**`rm.anova`**	**`glmm1`**	**`glmm2`**

condition	NS	NS (0.13)	NS (0.17)
relation	NS	** (0.43)	** (0.56)
condition*relation	*	* (-0.51)	* (-0.66)

Table 2: Significance (coefficients) of fixed effect terms in RM ANOVA and GLMMs

Although the data in table 2 seem to suggest a minor impact of the inclusion of a random effect term for stimulus in the model, the model statistics present a different image. First of all, the inclusion of the random effect term for stimulus realizes a highly significant reduction of the variance. An ANOVA with χ^2 as test statistic was computed to compare the difference between the deviance[1] of both GLMMs. The outcome of this test yields a significant decrease of the residual variance ($\chi^2 = 251.53$, df = 1, $p < 2.2e\text{-}16$).

These findings are corroborated by the comparison of the index of concordance (*c* index) of both models[2]: whereas the model with only subject as random effect term has a rather unsatisfactory *c* index of 0.5933 (95% CI = [0.5915;0.5951]), the model with two random effect terms displays a highly satisfactory *c* index of 0.8078 (95% CI = [0.8065;0.8091]).

The model statistics clearly support the model with both random effects (`glmm2`). Let us now have a closer look at the random part of this model, as summarized in table 3.

```
Random effects:
 Groups      Name        Variance Std.Dev.
 subject    (Intercept)  0.18367  0.42857
 stimulus   (Intercept)  1.13734  1.06646
Number of obs: 1330, groups: subject, 133; stimulus, 10
```

Table 3: Random effects in `glmm2`

In order to gage the proportion of the total variance explained by the random effects, we will compute the intra-class correlation coefficient (ρ) for both random effects. Again, the figures show that the impact of the subjects (ρ = 0.0793) is outreached by the impact of the stimuli (ρ = 0.4914), which are responsible for almost half of the variance in the model. It is common in experimental research that random effects account for substantial proportions of the overall variance, but generally it are the subjects rather than the stimuli who account for the major part of the variance.

The effect of both random parameters is shown in figure 1, where the random intercepts of the subjects (plot above) and the stimuli (plot below) are visualized with their 95% CI in grey. Please notify that the random effects rather than the coefficients of the intercepts are plotted, which implies that they identify the deviations from the overall intercept (dotted black line in both plots) which is the mean of all the random intercepts and equals 0 (cf. section 2). In the plot above, we can observe that the subjects display moderate variance and that all CIs cross the overall intercept, indicating that the average recall rate of no subject significantly differs from the overall baseline recall rate. In line with the above-mentioned results, more clear-cut differences emerge from the plot below. First, the intercept of 6 out of 10 stimuli significantly differs from the overall intercept: stimuli s01, s03 and s08 entail average recall rates that are significantly higher than the overall baseline, whereas the average recall rates of stimuli s06, s07 and s09 are significantly lower than the overall baseline recall rate. Second, two distinct clusters of stimuli with significantly different random intercepts arise: s01, s03 and s08 display significantly higher recall rates than stimuli s02, s04, s05, s06, s07, s09 and s10.

---

[1] The deviance equals -2 * log likelihood of a model and follows a chi-squared distribution.

[2] The *c* index computes the area under the Receiver Operating Characteristic curve, that plots the true positive rate against the true negative rate. When interpreting the *c* index ($c \in [0.5;1.0]$), $c \geq 0.8$ is considered to be indicative of a very good model.

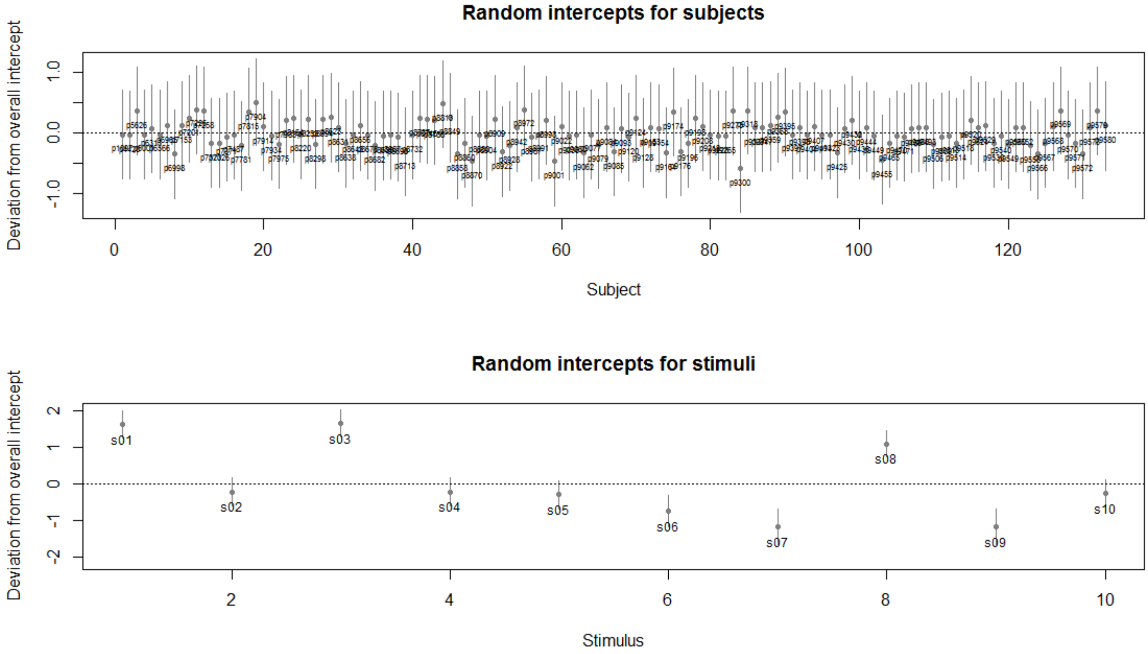

Figure 1: Deviations of random intercepts for subjects and stimuli from overall intercept

4 Discussion

The present research hypothesized that RM ANOVA is not the technique best suited to model variance in an experiment with crossed effects. Although coefficients and their p-values hardly differ, the inclusion of the second random effect term for the stimuli significantly improved the descriptive and predictive power of the model due to the further structuring of the error term. Results show that the baseline values for the stimuli display significant differences from the overall baseline value of the model and that baseline values of the stimuli mutually differ significantly. These findings are overlooked in a RM ANOVA with a random effect for subjects only. Furthermore, the high intra-class correlation for the stimuli clearly proves that the subjects' responses show a high degree of intra-group closeness: observations for the same stimulus are similar on different subjects and simultaneously different from the observations for other stimuli. In short, the GLMM has unveiled a significant effect of the stimulus on the successful recall of high status goods, suggesting an impact of the screen position of that good.

Future research will proceed along the following lines. First, it will be tested whether the present GLMM model can be improved by including random slopes allowing the fixed effect terms to vary across subjects and stimuli. Second, to fully understand the role of the stimuli, the impact of the display position on the successful recall rate will be further investigated.

The results argue for a well-considered selection of the statistical technique in experimental research in order to represent the characteristics, especially the non-independence of the observations, of the research design to obtain the best fitting model and to maximize the generalizability of the results.

References

Alhabash, S. & Wise, K. (2012). PeaceMaker: Changing Students' Attitudes Toward Palestinians and Israelis Through Video Game Play. *International Journal of Communication*, 6, 356-380.

Baayen, H., Davidson, D. & Bates, D. (2008). Mixed-effects modeling with crossed random effects for subjects and items. *Journal of Memory and Language*, 59 (4), 390-412.

Bates, D. (2005). Fitting linear mixed models in R. *R News*, 5, 27-30.

Bates, D., Maechler, M. & Bolker, B. (2013). lme4: Linear mixed-effects models using S4 classes. R package version 0.999999-2. http://cran.R-project.org/package=lme4.

Fournier, S. & Richins, M. (1991). Some theoretical and popular notion concerning materialism. *Journal of Social Behavior and Personality*, 6 (6), 403-414.

Gelman, A. & Hill, J. (2007). *Data Analysis Using Regression and Multilevel/Hierarchical Models*. Cambridge: Cambridge University Press.

Janssens, K., Pandelaere, M., Van den Bergh, B., Millet, K. Lens, I. & Roe, K. (2011). Can buy me love. Mate attraction goals lead to perceptual readiness for status products. *Journal of Experimental Social Psychology*, 47, 254-258.

Krakowiak, K. & Oliver, M. (2012) When Good Characters Do Bad Things: Examining the Effect of Moral Ambiguity on Enjoyment. *Journal of Communications*, 62, 117-135.

Mazodier, M., Quester, P. & Chandon, J.-L. (2012) Unmasking the ambushers: Conceptual framework and empirical evidence. *European Journal of Marketing*, 46 (1/2), 192-214.

Pinheiro, J. & Bates, D. (2000). *Mixed-Effects Models in S and S-PLUS*. New York: Springer.

Quené, H. & van den Berg, H. (2008). Examples of mixed-effects modeling with crossed random effects and with binomial data. *Journal of Memory and Language*, 59, 413-425.

R Core Team (2012). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing, Vienna, Austria, http://www.R-project.org.

Richins, M. (1994). Special possessions and the expression of material values. *Journal of Consumer Research*, 21 (3), 522-533.

Richins, M. & Dawson, S. (1992). A consumer values orientation for materialism and its measurement: Scale development and validation. *Journal of Consumer Research*, 19 (3), 303-316.

Richter, T. (2006). What is Wrong With ANOVA and Multiple Regression? Analyzing Sentence Reading Times With Hierarchical Linear Models. *Discourse Processes*, 41 (3), 221-250.

Roskos-Ewoldsen, D. & Fazio, R. (1992). On the orienting value of attitudes: Attitude accessibility of an object's attraction of visual attention. *Journal of Personality and Social Psychology*, 63 (2), 198-211.

Wang, J. & Wallendorf, M. (2006). Materialism, status signaling, and product satisfaction. *Journal of the Academy of Marketing Science*, 34, 494-505.