



Generalizing across stimuli as well as subjects: A non-mathematical tutorial on mixed-effects models

Yu-Hsuan A. Chang^{a,b,c} and David M. Lane^{a,✉}

^aDepartment of Psychology, Rice University

^bMultimodal Imaging Laboratory, University of California, San Diego

^cDepartment of Psychiatry, University of California

Abstract ■ Although it has long been known that analyses that treat stimuli as a fixed effect do not permit generalization from the sample of stimuli to the population of stimuli, surprisingly little attention has been paid to this issue outside of the field of psycholinguistics. The purposes of the article are (a) to present a non-technical explanation of why it is critical to provide a statistical basis for generalizing to both the population subjects and the population of stimuli and (b) to provide instructions for doing analyses that allows this generalization using four common statistical analysis programs (JMP, R, SAS, and SPSS).

Keywords ■ mixed-effects models, tutorials. **Tools** ■ JMP, SAS, SPSS, R.

✉ lane@rice.edu

 *YHAC*: 0000-0003-0953-0094; *DML*: 0000-0002-6364-9945

 [10.20982/tqmp.12.3.p201](https://doi.org/10.20982/tqmp.12.3.p201)

Acting Editor ■ Denis Cousineau (Université d'Ottawa)

Reviewers

■ One anonymous reviewer

Introduction

In a recent study on perception, Davis and Abrams (2016) presented subjects with general knowledge questions about celebrity names and a distractor that was either visually similar or dissimilar to the name. Questions that contained a visually-similar distractor led to more illusions than did the questions that contained a visual dissimilar distractor. One way to have analyzed these data would have been to average over all names and do the analyses on these averages. However, this method does not take the variability of names or the possibility that the treatment may have different effects for different names into account. As a result, this analysis would not have provided a statistical basis to generalize from the names used in the study to the population of names. In order to support this generalization, textciteda16 computed a statistic called F2 which, as we discuss later in this article, is better than the common practice of ignoring the variability of stimuli but is still not entirely satisfactory.

Although the importance of providing a statistical basis for generalizing to a population of stimuli has been apparent for many years (Clark, 1973; Coleman, 1964), researchers outside of the field of psycholinguistics, unlike

textciteda16, typically fail to do this. Clark (1973), in a now-classic paper, showed convincingly that in the analysis of data from psycholinguistic research, it is important to provide a statistical basis to generalize from the sample of words to the population of words from which they were sampled. A critical point made by Clark (1973) is that if one does not consider the variability of the stimuli in the statistical analysis, then the probability of a Type I error approaches 1.0 as the number of subjects increases. Clark's (1973) article was very influential and the vast majority of articles in psycholinguistics now provide a statistical basis to generalize not only to the population of subjects but also to the population of stimuli. However, surprisingly little attention has been paid to this issue in other areas in psychology.

One common erroneous belief is that if stimuli are counterbalanced across conditions, then it is not necessary to take additional steps to generalize to the population of stimuli. One source of this error is a misinterpretation of Raaijmakers, Schrijnemakers, and Gremen's (1999) article which appeared to make that point. In 2003, Raaijmakers emphasized the importance of generalization from the sample of stimuli to the population of stimuli by restating that “the claim that counterbalanced designs don't need



item analyses . . . is absolutely untrue. Such a design removes one problem that normally requires an item analysis (inadequate matching of materials across conditions), but it does not eliminate the issue of generality” (p. 146).

In spite of the publication of articles exhorting researchers to provide a statistical basis to generalize to the population of stimuli in fields including industrial and organizational psychology (Fontenelle, Phillips, & Lane, 1985), clinical psychology (Martindale, 1978), and social psychology (Judd, Westfall, & Kenny, 2012), these articles appear to have had a relatively limited impact. In the most recent two issues of *Journal of Experimental Psychology: Learning, Memory, Cognition* (Vol. 41, No. 5 & No. 6, 2015), there were 34 out of 54 articles (63%) aimed to generalize their findings from the samples of subjects and stimuli to the population of subjects and stimuli, but only 12 of them (35%) used the appropriate mixed effects model in their analyses. This article makes additional contributions to this field by illustrating the importance of using mixed-effects model and by providing an easy instruction for the readers to use mixed-effects model. The purposes of this article are (a) to present a non-technical explanation of why it is critical to provide a statistical basis for generalizing to both the population of subjects and the population of stimuli and (b) to provide instructions for doing an analysis that allows this generalization. The structure of the article is as follows: First, we use fictitious data to illustrate how the effects of different types of systematic variation across stimuli can lead to unjustified conclusions when the analysis that does not consider this variation is used. The fictitious data we use are purposely very small so that the consequences of ignoring stimulus variability can be seen simply by inspecting the raw data themselves. Second, we analyze these data using mixed-effects models, and, third, we provide detailed instructions for analyzing data with mixed-effects models using four common statistical analysis programs (SAS, SPSS, JMP, and R). We do not attempt to resolve theoretical disputes concerning the best ways to conduct a mixed-effects analysis. Instead, readers are referred to technical articles that address these issues.

Before proceeding, it is important to specify how a few terms will be used in this article. In accordance with common usage, if a separate group of subjects is used for each treatment condition, the design is called a “between-subject” design whereas if the same group of subjects is used for all treatment conditions, the design is called a “within-subjects” design. If different stimuli are used for each treatment condition, the design is called a “between-stimuli” design whereas if the same stimuli are used for all treatment conditions, the design is called a “within-stimuli” design.

It is important to distinguish between fixed and ran-

dom effects. An effect is a fixed effect if the levels of a variable are chosen by the experimenter. The analysis of fixed effects does not allow a statistical basis to generalize beyond the levels used in the experiment. An effect is a random effect if the levels of the variable are chosen randomly. The proper analysis of a random effect allows the statistical analysis is generalizable to the population levels.

Analyzing a Between-Subjects and Between-Stimuli Design

Suppose a researcher was interested in understanding whether people prefer ice cream or cake. One experimental design would be to randomly select a set of flavors for ice-cream and a different set of flavors for cake. In the example that follows, three flavors are selected for each food type (cake and ice cream) and the tastes of the foods are rated by different groups of three subjects.

The fictitious data in Table 1a show a possible outcome of the experiment. A typical way to analyze these data would be to compute the mean rating across flavors for each subject and conduct a one-way between-subjects ANOVA on these means. For these data, there is a significant effect of food type, $F(1, 4) = 12.45, p = 0.024, 95\%CI[-1.89, -0.22]$, supporting the conclusion that the effect of food type is real. However, since different flavors were selected for cake and for ice cream, it is possible that, by chance, the ice-cream flavors were better tasting than the cake flavors. An informal assessment of the variation among flavors within food type for these data suggests that differences among flavors are small relative to differences between food types and therefore the inference that there is a food-type effect may be justified.

The data in Table 1b tell a different story. The mean ratings of the subjects are the same as for the data in Table 1a and, therefore, the F values from the ANOVA are identical. However, there is much greater variability among the flavors within each food type in Table 1b. This variability should make a careful researcher extremely cautious about generalizing the results to the population of flavors for it is clear that a different choice of flavors might have made a big difference.

One way to take variation in stimuli into account is to compute the mean for each stimulus across subjects. For the data in Table 1a, the scores for ice-cream are 6.33, 6.33, and 6.00 and the scores for cake are 4.00, 4.00, and 4.33. An ANOVA comparing these conditions is significant, $F(1, 4) = 180.50, p < 0.001, 95\%CI[-1.27, -0.84]$. In contrast, the same analysis for the data in Table 1b results in a non-significant effect, $F(1, 4) = 0.97, p = 0.380, 95\%CI[-4.03, 1.92]$.

The F computed by averaging over stimuli is called F_1 , and the F computed by averaging over subjects is called F_2



Table 1 ■ Fictitious Data from Between-Subjects, Between-Stimuli Design.

Table 1a

Condition	Subject ID	F1	F2	F3	F4	F5	F6	Mean
Ice Cream	S1	6	5	5				5.33
Ice Cream	S2	6	7	7				6.67
Ice Cream	S3	7	7	6				6.67
Cake	S4				3	3	4	3.33
Cake	S5				5	5	4	4.67
Cake	S6				4	4	5	4.33
	Mean	6.33	6.33	6.00	4.00	4.00	4.33	

Table 1b

Condition	Subject ID	F1	F2	F3	F4	F5	F6	Mean
Ice Cream	S1	8	4	4				5.33
Ice Cream	S2	8	6	6				6.67
Ice Cream	S3	8	6	6				6.67
Cake	S4				1	1	8	3.33
Cake	S5				3	4	7	4.67
Cake	S6				2	2	9	4.33
	Mean	8	5.33	5.33	2.00	2.33	8.00	

(Clark, 1973). The former F tests whether the results for the specific stimuli (flavors in this example) used in the experiment generalize to the population of subjects. As such, “subjects” is a random effect and “stimuli” is a fixed effect. The latter F tests whether, for the specific subjects used in the experiment, the results generalize to the population of stimuli. In this analysis, “subjects” is a fixed effect and “stimuli” is a random effect. Naturally, it is important to use a method that allows one to generalize to the population of stimuli in the population of subjects. Note that it is possible for (a) the effect to be present in the population of subjects for the stimuli used in the experiment and (b) in the population of stimuli for the subjects used in the experiment but not (c) for the population of stimuli in the population of subjects. Clark (1973) showed an approximate method for generalizing to the population of stimuli in the population of subjects with quasi- F tests. In recent years, mixed model software capable of analyzing complex combinations of random and fixed effects have become widely available and these analyses are preferable to quasi- F tests. The advantages of mixed effects models include their ability to (a) handle missing and unbalanced data, (2) handle continuous and categorical predictors and responses, and (3) avoid inflated Type I errors (Baayen, 2008; Baayen, Davidson, & Bates, 2008; Judd et al., 2012).

Analyzing the data in Table 1a using a mixed model analysis results in $F(1, 4) = 12.45, p = 0.024, 95\%CI[-1.89, -0.22]$, whereas the same analysis for the data in Table 1b results in $F(1, 4.2) = 0.94, p = 0.384, 95\%CI[-4.01, 1.90]$ (See the Appendix for instruc-

tions for doing this and other mixed model analyses). Thus, for the data in Table 1a, one can conclude that there is an effect of food type in the population of subjects and the population of flavors. This conclusion is not justified for the data in Table 1b.

Analyzing a Between-Subjects and Within-Stimuli Design

In between-subjects/between-stimuli designs, the between-subjects effect is confounded with the effect of stimuli. It might appear that if the same stimuli were used in each condition, then there would be no need to consider the variability of stimuli. Although this would be true if the differences between conditions were the same for all stimuli, it is not true when there is a Condition (Food Type) \times Stimulus (Flavor) interaction.

The fictitious data in Table 2a are from a between-subjects/within-stimuli design in which the researcher randomly selected three flavors and these flavors were paired with both food types (cake and ice cream) so that each subject rated each of the six food type \times flavor combinations.

The data in Table 2a show a possible outcome of this experiment. A typical way of analyzing these data would be to compute the mean rating across flavors for each subject and conduct a one-way between-subjects ANOVA on these means. For these data, there is a significant effect of food type, $F(1, 4) = 64, p = 0.001, 95\%CI[-1.80, -0.87]$, which would lead to the conclusion that the effect of food type is real. However, even though the same flavors are used in both food types, it is possible that subjects prefer



Table 2 ■ Fictitious Data from Between-Subjects, Within-Stimuli Design.

Table 2a

Condition	Subject ID	F1	F2	F3	F1	F2	F3	Mean
Ice Cream	S1	4	8	9				7
Ice Cream	S2	5	7	9				7
Ice Cream	S3	3	7	8				6
Cake	S4				1	4	7	4
Cake	S5				1	4	7	4
Cake	S6				1	4	7	4
	Mean	4	7.33	8.67	1	4	7	

Table 2b

Condition	Subject ID	F1	F2	F3	F1	F2	F3	Mean
Ice Cream	S1	4	8	9				7
Ice Cream	S2	5	7	9				7
Ice Cream	S3	3	7	8				6
Cake	S4				6	5	1	4
Cake	S5				5	4	3	4
Cake	S6				5	4	3	4
	Mean	4	7.33	8.67	5.33	4.33	2.33	

cakes for some flavors and ice cream for others. Informally, this does not appear to be a serious problem for the data in Table 2a, but does appear to be a problem for the data in Table 2b. Since the mean ratings of the subjects are the same in Tables 3 and 4, the *F* values from the ANOVA are identical. However, there is a much greater Food Type × Flavor interaction for the data in Table 2b, and this interaction should make a careful researcher extremely cautious about generalizing the results to the population of flavors.

One way to take this interaction of Food Type × Flavor into account is to compute the mean for each stimulus across subjects. For the data in Table 2a, an ANOVA with repeated measures on flavors is significant, $F(1, 2) = 27.43, p = 0.035, 95\%CI[-2.43, -0.24]$. By comparison, the same analysis for the data in Table 2b results in not significant effect, $F(1, 2) = 1.44, p = 0.353, 95\%CI[-6.11, 3.44]$.

Analyzing these data using a mixed model analysis results in $F(1, 3) = 21.33, p = 0.019, 95\%CI[-2.25, -0.42]$, for the data in Table 2a, and $F(1, 4) = 2.63, p = 0.180, 95\%CI[-3.62, 0.95]$, for the data in Table 2b (see Section 4 in the Appendix for analysis instructions). Thus, for the data in Table 2a, one can conclude that there is an effect of food type in the population of subjects and the population of flavors. This conclusion is not justified for the data in Table 2b.

Analyzing a Within-Subjects and Between-Stimuli Design

The data in Table 2a are from a within-subjects/between-stimuli design in which the researchers randomly selected a set of flavors for the ice-cream and a different set of flavors for the cake, and each subject rated all these food flavors under both two food types.

The fictitious data in Table 3a show a possible outcome of this experiment. A typical way of analyzing these data would be to compute the mean rating across three flavors in each food type for each subject and conduct a one-way within-subjects ANOVA on these means. For these data, there is a significant effect of food type, $F(1, 2) = 49, p = 0.020, 95\%CI[-1.88, -0.45]$, which would lead to the conclusion that the effect of food type is real. This effect of food types is strong, because all the subjects have higher mean ratings for ice cream than for the cake.

For the data in Table 3b, the mean ratings of the subjects are the same as in Table 3a and therefore the *F* values from the ANOVA are identical. However, there is much greater variability among the flavors.

As in the previous analyses, one should analyze these data using a mixed model analysis, which results in $F(1, 16) = 12.25, p = 0.003, 95\%CI[-1.87, -0.46]$, for the data in Table 3a, and $F(1, 4) = 3.87, p = 0.121, 95\%CI[-2.81, -0.48]$, for the data in Table 3b (See Section 5 in the Appendix for analysis instructions). Thus, for the data in Table 3a, one can conclude that there is an effect of food type in the population of subjects and the



Table 3 ■ Fictitious Data from Within-Subjects, Between-Stimuli Design.

Table 3a

Condition	Ice Cream			Cake			Mean	
Subject ID	F1	F2	F3	F4	F5	F6	Ice Cream	Cake
S1	5	4	6	5	3	1	5	3
S2	7	6	5	3	4	5	6	4
S3	4	7	7	1	3	5	6	3
Mean	5.33	5.67	6	3	3.33	3.67		

Table 3b

Condition	Ice Cream			Cake			Mean	
Subject ID	F1	F2	F3	F4	F5	F6	Ice Cream	Cake
S1	9	3	3	5	3	1	5	3
S2	7	5	6	3	4	5	6	4
S3	8	6	4	1	3	5	6	3
Mean	8	4.67	4.33	3	3.33	3.67		

population of flavors. This conclusion is not justified for the data in Table 3b.

Analyzing a Within-Subjects and Within-Stimuli Design

It might appear that in a within-subjects/within-stimuli design, both the variability of subjects and variability of stimuli are controlled. However, this design still requires subjects and stimuli to be analyzed as random effects in order for the results to generalize to the populations of subjects and stimuli.

The data in Tables 7 and 8 are from a within-subjects/within-stimuli design in which the researcher randomly selected three flavors and these flavors were paired with both food types (cake and ice cream) so that each subject rated all six food stimuli.

The fictitious data in Table 4a show a possible outcome of the experiment. A typical way of analyzing these data would be to compute the mean rating across three flavors in each food type condition for each subject and conduct a one-way within-subjects ANOVA on these means. For these data, there is a significant effect of food type, $F(1, 2) = 49, p = 0.020, 95\%CI[-1.88, -0.45]$, which would lead to the conclusion that the effect of food type is real. However, even though the same flavors are used in both food types, it is possible that flavor preferences differ across subjects. In other words, there might be a Subjects \times Flavor interaction. For example, in Table 4b, S1 prefers F1 over the other flavors whereas S2 prefers F3 over the other flavors. Informally, this does not appear to be a serious problem for the data in Table 4a, but does appear to be a problem for the data in Table 4b. Since the mean ratings of the subjects are the same in Tables 7 and 8, the F values from the ANOVA are identical. However, there is a much greater Subject \times Flavor interaction for the data in Table 4b, which should

make a careful researcher extremely cautious about generalizing the results to the population of flavors.

Analyzing these data using a mixed model analysis results in $F(1, 16) = 24.50, p < 0.001, 95\%CI[-1.67, -0.67]$. for the data in Table 4a, and $F(1, 2.3) = 7.50, p = 0.096, 95\%CI[-2.79, 0.46]$, for the data in Table 4b (see Section 6 in the Appendix for analysis instructions). Thus, for the data in Table 4a, one can conclude that there is an effect of food type in the population of participants and the population of flavors. This conclusion is not justified for the data in Table 4b.

Conclusion

The example data in this article highlight the danger of not taking the variability of stimuli and possible interactions between variables into account. In order to be able to generalize the results from the samples of stimuli and subjects to the populations of stimuli and subjects, both stimuli and subjects should be treated as random effects simultaneously in a mixed-effects model analysis. The Appendix in this article provides detailed instructions for analyzing the sample data with mixed-effects models with four most popular statistical programs currently in psychological use (JMP, R, SAS, and SPSS).

This article does not present a technical mathematical discussion of mixed-effects models and therefore researchers who are seeking either more theoretical information or discussions of more complex designs should consult the following sources: For further discussion of how failure to generalize the results to the population of stimuli can threaten the validity of the analyses, we refer the readers to Forster and Dickinson (1976), Kenny (1985), Wells and Windschitl (1999). For more technical discussions of mixed-effects models, Baayen et al. (2008), Judd



Table 4 ■ Fictitious Data from Within-Subjects, Within-Stimuli Design.

Table 4a

Condition	Ice Cream			Cake			Mean	
Subject ID	F1	F2	F3	F1	F2	F3	Ice Cream	Cake
S1	6	6	6	4	4	4	6	4
S2	7	6	5	5	4	3	6	4
S3	4	7	7	3	1	5	6	3
Mean	5.67	6.33	6	4	3	4		

Table 4b

Condition	Ice Cream			Cake			Mean	
Subject ID	F1	F2	F3	F1	F2	F3	Ice Cream	Cake
S1	10	4	4	10	1	1	6	4
S2	4	5	9	1	2	9	6	4
S3	3	7	8	3	1	5	6	3
Mean	5.67	5.33	7	4.67	1.33	5		

et al. (2012), Raaijmakers et al. (1999) and Raaijmakers's (2003) articles are recommended. For readers who are interested in using R to analyze data with mixed-effect models, Baayen's (2008) book titled "Analyzing linguistic data: A practical introduction to statistics" provides a detailed step-by-step instructions in Chapter 7. For readers who are interested in using SAS to analyze data with mixed-effects models, Littell, Milliken, Stroup, Wolfinger, and Schabenberger's (2006) book titled "SAS for Mixed Models, Second Edition" provides very detailed instructions.

References

Arnau, J., Bono, R., & Vallejo, G. (2009). Analyzing small samples of repeated measures data with the mixed-model adjusted f test. *Communications in Statistics - Simulation and Computation*, 38, 1083–1103. doi:10.1080/03610910902785746

Baayen, R. H. (2008). *Analyzing linguistic data: a practical introduction to statistics using r*. Cambridge, England: Cambridge University Press.

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. doi:10.1016/j.jml.2007.12.005

Barton, K. (2016). Mumin: multi-model inference (Version 1.15.6). Retrieved from <https://cran.r-project.org/web/packages/MuMIn/MuMIn.pdf>

Bates, D. (2014). Lme4 for sas proc mixed users. Retrieved from <https://cran.r-project.org/web/packages/SASmixed/vignettes/Usinglmer.pdf>

Bates, D., Maechler, M., Bolker, B., & Walker, S. (2015). Lme4: linear mixed-effects models using eigen and s4 (r package) (Version 1,1-9). Retrieved from <https://CRAN.R-project.org/package=lme4>

Clark, H. (1973). The language as fixed-effect fallacy: a critique of language statistics in psychological research. *Journal of Verbal Learning and Verbal Behavior*, 12(4), 335–359. doi:10.1016/S0022-5371(73)80014-3

Coleman, E. B. (1964). Generalizing to a language population. *Psychological Reports*, 14(1), 219–226. doi:10.2466/pr0.1964.14.1.219

Davis, D. K. & Abrams, L. (2016). Here's looking at you: visual similarity exacerbates the moose illusion for semantically similar celebrities. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 42, 75–90. doi:doi:10.1037/xlm0000144

Fai, A. H. T. & Cornelius, P. L. (1996). Approximate f-tests of multiple degree of freedom hypotheses in generalized least squares analyses of unbalanced split-plot experiments. *Journal of Statistical Computation and Simulation*, 54(4), 363–378. doi:10.1080/00949659608811740

Fontenelle, G. A., Phillips, A. P., & Lane, D. M. (1985). Generalizing across stimuli as well as subjects: a neglected aspect of external validity. *Journal of Applied Psychology*, 70(1), 101–107. doi:10.1037/0021-9010.70.1.101

Forster, K. I. & Dickinson, R. G. (1976). More on the language-as-fixed-effect fallacy: monte carlo estimates of error rates for f1, f2, f', and min f'. *Journal of Verbal Learning and Verbal Behavior*, 15, 135–142. doi:10.1016/0022-5371(76)90014-1

Hocking, R. R., Green, J. W., & Bremer, R. H. (1989). Variance-component estimation with model-based diagnostics. *Technometrics*, 31(2), 227–239. doi:10.1080/00401706.1989.10488516

Johnson, P. C. D. (2014). Extension nakagawa & schielzeth's r2 glmm to random slopes models. *Methods in Ecology and Evolution*, 5(9), 44–946. doi:10.1111/2041-210x.12225



- Judd, C. M., Westfall, J., & Kenny, D. A. (2012). Treating stimuli as a random factor in social psychology: a new and comprehensive solution to a pervasive but largely ignored problem. *Journal of Personality and Social Psychology, 103*(1), 54–69. doi:[10.1037/a0028347](https://doi.org/10.1037/a0028347)
- Kenny, D. A. (1985). Quantitative methods for social psychology. In G. Lindzey & E. Aronson (Eds.), *Handbook of social psychology (vol. 1, 3rd ed.)* (pp. 487–508). New York, NY: Random House.
- Kenward, M. G. & Roger, J. H. (1997). Small sample inference for fixed effects from restricted maximum likelihood. *Biometrics, 53*(3), 983–997. doi:[10.2307/2533558](https://doi.org/10.2307/2533558)
- Kuznetsova, A., Brockhoff, P. B., & Christensen, R. H. B. (2015). Lmertest: tests in linear mixed effects models (r package) (Version 2.0-29). Retrieved from <https://cran.r-project.org/web/packages/lmerTest/index.html>
- Littell, R. C., Milliken, G. A., Stroup, W. W., Wolfinger, R. D., & Schabenberger, O. (2006). *Sas for mixed models (2nd ed.)* Cary, NC: SAS Institute.
- Liu, H. H., Zheng, Y., & Shen, J. (2008). Goodness-of-fit measures of r^2 for repeated measures mixed effect models. *Journal of Applied Statistics, 35*(10), 1081–1092. doi:[10.1080/02664760802124422](https://doi.org/10.1080/02664760802124422)
- Martindale, C. (1978). The therapist-as-fixed-effect fallacy in psychotherapy research. *Journal of Consulting and Clinical Psychology, 46*(6), 1526–1530. doi:[10.1037/0022-006X.46.6.1526](https://doi.org/10.1037/0022-006X.46.6.1526)
- Müller, S., Scealy, J. L., & Welsh, A. H. (2013). Model selection in linear mixed models. *Statistical Science, 28*(2), 135–167. doi:[10.1214/12-sts410](https://doi.org/10.1214/12-sts410)
- Nakagawa, S. & Schielzeth, H. (2013). A general and simple method for obtaining r^2 from generalized linear mixed effects models. *Methods in Ecology and Evolution, 4*(2), 133–142. doi:[10.1111/j.2041-210x.2012.00261.x](https://doi.org/10.1111/j.2041-210x.2012.00261.x)
- Orelien, J. G. & Edwards, L. J. (2008). Fixed-effect variable selection in linear mixed models using r^2 statistics. *Computational Statistics & Data Analysis, 52*(4), 1896–1907. doi:[10.1016/j.csda.2007.06.006](https://doi.org/10.1016/j.csda.2007.06.006)
- Raaijmakers, J. G. W. (2003). A further look at the language-as-a-fixed-effect fallacy. *Canadian Journal of Experimental Psychology, 57*(3), 141–151. doi:[10.1037/h0087421](https://doi.org/10.1037/h0087421)
- Raaijmakers, J. G. W., Schrijnemakers, J. M. C., & Gremen, 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. doi:[10.1006/jmla.1999.2650](https://doi.org/10.1006/jmla.1999.2650)
- Satterthwaite, F. E. (1946). An approximate distribution of estimates of variance components. *Biometrics Bulletin, 2*(6), 110–114. doi:[10.2307/3002019](https://doi.org/10.2307/3002019)
- Schaalje, G., McBride, J., & Fellingham, G. (2001). Approximations to distributions of test statistics in complex mixed linear models using sas proc mixed. In *Proceedings of the sas users group international 26th annual conference* (p. 262).
- Smith, D. W. & Murray, L. W. (1984). An alternative to eisenhart’s model ii and mixed model in the case of negative variance estimates. *Journal of the American Statistical Association, 79*(385), 145–151. doi:[10.2307/2288349](https://doi.org/10.2307/2288349)
- Snijders, T. A. & Bosker, R. J. (1994). Modeled variance in two-level models. *Sociological methods & research, 22*(3), 342–363. doi:[10.1177/0049124194022003004](https://doi.org/10.1177/0049124194022003004)
- Spilke, J., Piepho, H. P., & Hu, X. (2005). Analysis of unbalanced data by mixed linear models using the mixed procedure of the sas system. *Journal of Agronomy and Crop Science, 191*(1), 47–54. doi:[10.1111/j.1439-037X.2004.00120.x](https://doi.org/10.1111/j.1439-037X.2004.00120.x)
- Thompson, W. A. & Moore, J. R. (1963). Non-negative estimates of variance components. *Technometrics, 5*(4), 441–449. doi:[10.2307/1266020](https://doi.org/10.2307/1266020)
- Wells, G. L. & Windschitl, P. D. (1999). Stimulus sampling and social psychological experimentation. *Personality and Social Psychological Bulletin, 25*, 1115–1125. doi:[10.1177/01461672992512005](https://doi.org/10.1177/01461672992512005)
- Xu, R. H. (2003). Measuring explained variation in linear mixed effects models. *Statistics in Medicine, 22*(22), 3527–3541. doi:[10.1002/sim.1572](https://doi.org/10.1002/sim.1572)

Appendix

In this appendix, we illustrate how to do mixed model analysis for the experimental designs discussed in the article. We give instructions for using four popular statistical analysis programs: JMP, R, SAS, and SPSS.

Section 1: Technical Details.

There are two ways of estimating variance components: constrained and unbounded. With constrained estimates, negative estimates of variance are set to zero. In unbounded estimates, estimates can be negative in order for the estimates to be unbiased. A technical discussion of the arguments for using either constrained or unbounded estimates is beyond the scope of this paper. Technical discussions can be found in the following sources (Hocking, Green, & Bremer, 1989; Smith & Murray, 1984; Thompson & Moore, 1963). Although both SAS and JMP allow options for constrained or unbounded variance estimates, R’s `lme4` package (Bates, Maechler, Bolker, & Walker, 2015) and SPSS only do analyses using constrained



Table 5 ■ Notations used in JMP, R, SAS and SPSS

	Flavor Nested within Food Type	Flavor Crossed with Food Type
JMP	Flavor[Food Type]	Flavor * Food Type
R	FoodType/Flavor	Flavor : Food Type
SAS	Flavor(FoodType)	Flavor * Food Type
SPSS	Flavor(FoodType)	Flavor * Food Type

variance components. To ensure that all programs used the same methods in our examples, we used constrained variance estimates throughout. However, we do not take a position here on whether constrained or unbounded estimates are preferable.

The denominator degrees of freedom and p-values of F tests in this tutorial were computed using Satterthwaite’s approximation (Fai & Cornelius, 1996; Satterthwaite, 1946), the preferred method when a constrained variance components model is used. An alternative method, Kenward-Roger Approximation (Kenward & Roger, 1997), is recommended when sample sizes are moderate to small and design is reasonably balanced (Schaalje, McBride, & Fellingham, 2001), although which method is better has not been fully resolved (Arnau, Bono, & Vallejo, 2009; Spilke, Piepho, & Hu, 2005).

There are many ways to create contrast matrices to fit the mixed effects model, and different program may use different strategies to create the contrast matrices. A contrast matrix contains the contrasts used for coding the factor in the mixed effects model. The default contrast matrix for `lmer` contains a contrast column for each level of the factor with the baseline level. There is no separate contrast for the baseline level, and the contrast matrix is not orthogonal to the intercept. The default contrast matrix for SAS and SPSS sets the baseline level to the last level of the factor instead of the first, and the contrast matrix is not orthogonal to the intercept. The default for JMP applied is to use contrasts that are orthogonal to the intercept. Since the F -values and p-values of the fixed effect are the same for different contrast matrices, we used the default contrasts provided in each program. Researchers can choose reasonable contrasts based on their experimental design and hypotheses. In R, researchers can easily apply the following codes before running model to change the default of contrast to be the same with the contrasts that SAS (Bates, 2014, p. 6), SPSS, and JMP used:

```
>> contrasts(mydata$FoodType) <- contr.SAS # SAS and SPSS
>> contrasts(mydata$FoodType) <- contr.sum # JMP
>> contrasts(mydata$FoodType) <- contr.treatment # R
```

If the levels of one factor (e.g., A) occur only within a single level of another factor (e.g., B), then the factor A is nested within B. If all levels of one factor (e.g., effect A) occur in combination with each level of another factor (e.g., B), then A and B are said to be crossed. Table 5 shows the notations used in the four programs.

Finally, the analyses in both JMP and SPSS can be performed by either using a graphical interface or by writing code whereas the analyses in both SAS and R are performed by writing code. In this tutorial, we show the graphical interface in JMP and we provide the codes for all of the four statistical tools.

Before conducting any analyses, the data should be organized so that there are four variables: Subject, FoodType, Flavor, and Rating.

Section 2: Program Specific Instructions

JMP-Specific Instructions

Choose “Fit Model” by clicking on “Analyze” in the menu, and the following dialog is shown. Choose “Mixed Model” in “Personality,” uncheck “The Unbounded Variance Components,” indicate the “Rating” in “Y” (dependent variable), and then add the “FoodType” in “Fixed Effects” as seen in Figure 1

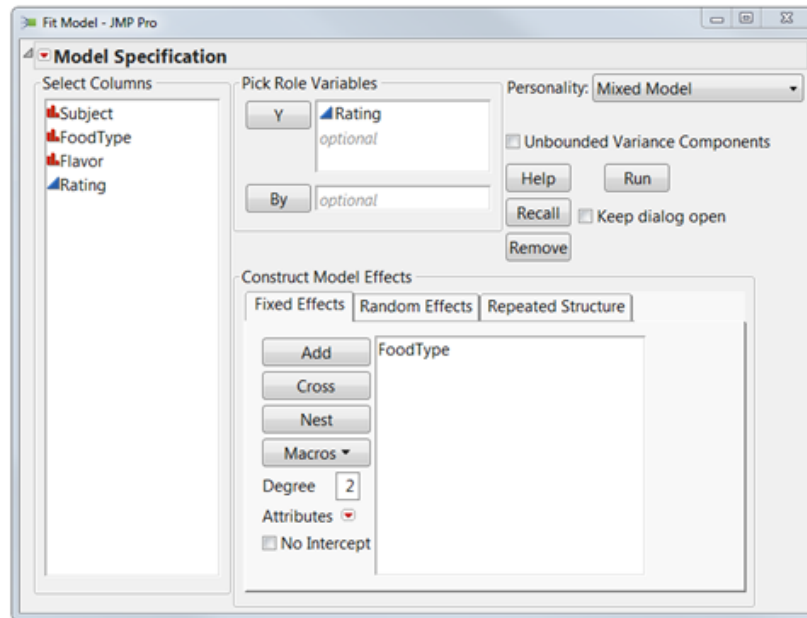
R-Specific Instructions

In R, `lme4` package (Bates et al., 2015) and `lmerTest` package (Kuznetsova, Brockhoff, & Christensen, 2015) are applied to analyze the data with mixed-effects models. In Section 7, `MuMIn` packages (Barton, 2016) is used to calculate the R^2 in the mixed effects model.

Read the data, name it “mydata,” and then create a new object which here called “data.model” from the `lmer()`



Figure 1 ■ Performing analyses with JMP



function. The mixed model has three essential elements: the dependent variable (Rating), a fixed effect (FoodType), and the random effects. In the `lmer()` function, the dependent variable is specified on the left side of the `~` operator. The independent variables are assigned on the right side of the `~` operator. Random effects indicated by placing them in parentheses, as `(1|variable)`, which indicates the intercept is random with respect to the variable. The format for `lmer()` function is `lmer(independent variable ~ fixed effect variable + (1|random effect variable))`.

SAS-Specific Instructions

In SAS, use the PROC MIXED command to compute the mixed-effects analysis. To include the negative variance components in the model, use the NOBOUND option. The data in the examples are in the dataset named as mydata.

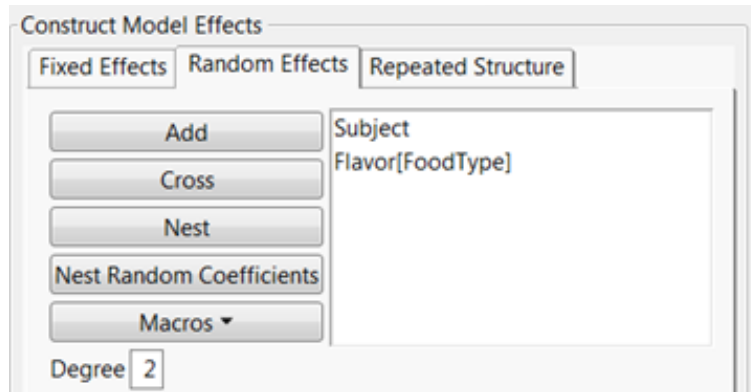
There are three essential elements in this code: CLASS, MODEL, and RANDOM. The CLASS statement is used to name the classification variables that would be used in the analysis. Therefore, these three variables are named: Subject, FoodType, and Flavor. The MODEL statement is used to identify the dependent variable (Rating) and the fixed effects (FoodType) by using the formula of dependent variable = fixed effect variable. In addition, the DDFM command to specify the approximation to estimate the denominator degrees of freedom and p-values in *F* tests. In this tutorial, SATTERTHWAIT approximation was used. The RANDOM statement is used to define the random effects that included in the models.

SPSS-Specific Instructions

In SPSS, use the MIXED command to do a mixed-effects model analysis. There are three essential elements in this syntax: BY, FIXED, and RANDOM. The BY statement is used to indicate the dependent variable and other variables that are used in the model. The dependent variable (Rating) is placed in the left side of BY, and all other independent variables (Subject, Flavor, FoodType) are placed in the right side of BY. The /FIXED statement is used to indicate the fixed effect variable: FoodType. The /RANDOM statement is used to define the random effects that included in the model.



Figure 2 ■ Model effects for Between-Subjects and Between-Stimuli Design within JMP



Section 3: Between-Subjects and Between-Stimuli Design (Table 1b data)

JMP

In this design, there are two random effects: Subject and Flavor nested within FoodType. To indicate this, click on the “Random Effects” tab under “Construct Model Effects,” and then use “Add” button for “Subject” and for “Flavor.” Then, use the “Nest” button to create Flavor[FoodType]. Finally, click the “Run” function. Figure 2 illustrates this step

The following code performs the same analysis. Once the code is entered, click on “Run Script” button to execute the script.

```
Fit Model(
  Y( :Rating ),
  Effects( :FoodType ),
  Random Effects( :Subject, :Flavor[:FoodType] ),
  NoBounds( 0 ),
  Personality( ``Mixed Model`` ),
  Run( Repeated Effects Covariance Parameter Estimates( 0 ) )
);
```

The output of fixed effects tests is contained in Figure 3, top panel.

R

In this design, there are two random effects: Subject and Flavor nested within FoodType. To indicate this, (1|Subject) and (1| Flavor) are used. The nested effect would be identified by lmer function. The code to generate the complete mixed-effects model in this design as follows:

```
> data.model = lmer(Rating ~ FoodType + (1|Subject) + (1|Flavor), data= mydata)
```

To view the results of the fixed-effects model, enter the name of fitted model object as a command or using summary() command. To obtain the estimates of degrees of freedom and p-values of the F-values for the fixed effects, use anova() function from lmerTest package with the command:

```
> anova(data.model)
```

The output of fixed effect is contained as followed:

```
Analysis of Variance Table of type III with Satterthwaite
approximation for degrees of freedom
Sum Sq Mean Sq NumDF DenDF F.value Pr(>F)
```



Figure 3 ■ Outputs from JMP (panel a), SAS (panel b) and SPSS (panel c) for the data of Table 1b

(a)

Fixed Effects Tests

Source	Nparm	DFNum	DFDen	F Ratio	Prob > F
FoodType	1	1	4.2	0.9426628	0.3837

(b)

Type 3 Tests of Fixed Effects				
Effect	Num DF	Den DF	F Value	Pr > F
FoodType	1	4.23	0.94	0.3838

(c)

Fixed Effects

Type III Tests of Fixed Effects^a

Source	Numerator df	Denominator df	F	Sig.
Intercept	1	4.233	22.582	.008
FoodType	1	4.233	.943	.384

a. Dependent Variable: Rating.

FoodType 0.89019 0.89019 1 4.2326 0.94256 0.3838

SAS

In this design, there are two random effects: Subject and Flavor nested within FoodType. The code to generate the complete mixed-effects model in this design as follows:

```
PROC MIXED DATA=mydata;
  CLASS Subject FoodType Flavor;
MODEL Rating = FoodType / DDFM=SATTERTHWAITE SOLUTION;
  RANDOM Subject;
  RANDOM Flavor(FoodType);
RUN;
```

The output of fixed effect is contained in Figure 3, middle panel.

SPSS

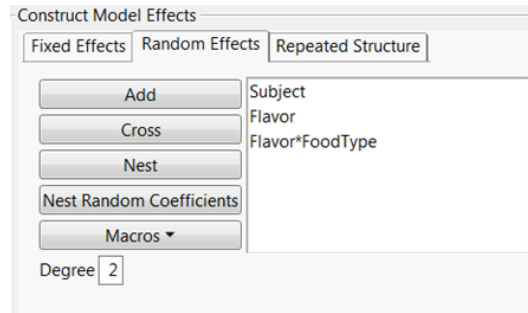
In this design, there are two random effects: Subject and Flavor nested within FoodType. The code to generate the complete mixed-effects model in this design as follows:

```
MIXED Rating BY Subject FoodType Flavor
  /FIXED = FoodType
  /RANDOM Subject Flavor(FoodType)
  /PRINT SOLUTION.
```

The output of fixed effect is contained in Figure 3, bottom panel.



Figure 4 ■ Model effects for Between-Subjects and Within-Stimuli Design within JMP



Section 4: Between-Subjects and Within-Stimuli Design (Table 2b data)

JMP

In this design, there are three random effects: Subject, Flavor, and the interaction between Flavor and FoodType. To indicate this, click on the “Random Effects” tab under “Construct Model Effects,” and then use “Add” button for “Subject” and for “Flavor.” Then, use the “Cross” button to create Flavor*FoodType. Finally, click the “Run” function. Figure 4 illustrates this step.

The following code performs the same analysis. Once the code is entered, click on “Run Script” button to execute the script.

```
Fit Model(
  Y( :Rating ),
  Effects( :FoodType ),
  Random Effects( :Subject, :Flavor, :Flavor * :FoodType ),
  NoBounds( 0 ),
  Personality( ``Mixed Model`` ),
  Run( Repeated Effects Covariance Parameter Estimates( 0 ) ),
);
```

The output of fixed effect is shown in Figure 5, top panel.

R

In this design, there are three random effects: Subject, Flavor and the interaction between Flavor and FoodType. To indicate this, (1|Subject), (1|Flavor), and (1|Flavor:FoodType) are used. The code to generate the complete mixed-effects model in this design as follows:

```
> data.model = lmer(Rating ~ FoodType + (1|Subject) + (1|Flavor) + (1|FoodType:
  Flavor), data= mydata)
```

To view the results of the fixed-effects model, enter the name of fitted model object as a command or using summary() command. To obtain the estimates of degrees of freedom and p-values of the F-values for the fixed effects, use anova() function from lmerTest package.

```
> anova(data.model)
```

The output of fixed effect is contained as followed.

```
Analysis of Variance Table of type III with Satterthwaite
approximation for degrees of freedom
Sum Sq Mean Sq NumDF DenDF F.value Pr(>F)
```



Figure 5 ■ Outputs from JMP (panel a), SAS (panel b) and SPSS (panel c) for the data of Table 2b

(a)

Fixed Effects Tests

Source	Nparm	DFNum	DFDen	F Ratio	Prob > F
FoodType	1	1	4.0	2.630137	0.1802

(b)

Type 3 Tests of Fixed Effects				
Effect	Num DF	Den DF	F Value	Pr > F
FoodType	1	4	2.63	0.1802

(c)

Fixed Effects

Type III Tests of Fixed Effects^a

Source	Numerator df	Denominator df	F	Sig.
Intercept	1	4	42.082	.003
FoodType	1	4	2.630	.180

a. Dependent Variable: Rating.

FoodType 1.6073 1.6073 1 4 2.6301 0.1802

SAS

In this design, three are random effects: Subjects, Flavor and the interaction between Flavor and FoodType. The code to generate the complete mixed-effects model in this design as follows:

```
PROC MIXED DATA=mydata;
  CLASS Subject FoodType Flavor;
MODEL Rating = FoodType / DDFM=SATTERTHWAITE SOLUTION;
  RANDOM Subject;
  RANDOM Flavor;
  RANDOM Flavor*FoodType;
RUN;
```

The output of fixed effect is shown in Figure 5, middle panel.

SPSS

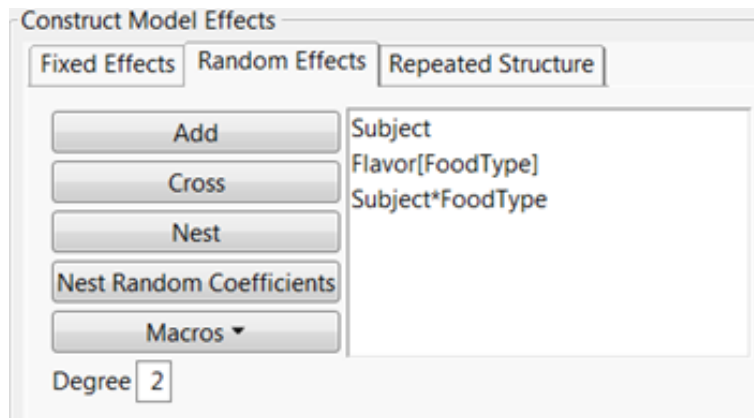
In this design, three are random effects: Subject, Flavor and the interaction between Flavor and FoodType. The code to generate the complete mixed-effects model in this design as follows:

```
MIXED Rating BY Subject FoodType Flavor
  /FIXED = FoodType
  /RANDOM Subject Flavor FoodType*Flavor
  /PRINT SOLUTION.
```

The output of fixed effect is shown in Figure 5, bottom panel.



Figure 6 ■ Model effects for Within-Subjects and Between-Stimuli Design within JMP



Section 5: Within-Subjects and Between-Stimuli Design (Table 3b data)

JMP

In this design, there are three random effects: Subject, Flavor nested within FoodType, and the interaction between Subject and FoodType. To indicate this, click on the “Random Effects” tab under “Construct Model Effects,” and then use “Add” button for “Subject” and for “Flavor.” Then, use the “Nest” button to create Flavor[FoodType], and use the “Cross” button to create Subject*FoodType. Finally, click the “Run” function. Figure 6 illustrates this step.

The following code performs the same analysis. Once the code is entered, click on “Run Script” button to execute the script.

```
Fit Model(
  Y( :Rating ),
  Effects( :FoodType ),
  Random Effects( :Subject, :Flavor[:FoodType], :Subject * :FoodType ),
  NoBounds( 0 ),
  Personality( ``Mixed Model`` ),
  Run( Repeated Effects Covariance Parameter Estimates( 0 ) ),
);
```

The output of fixed effect is shown in Figure 7, top panel. .

R

In this design, there are three random effects: Subject, Flavor nested within FoodType, and the interaction between Subject and FoodType. To indicate this, (1|Subject), (1|Flavor), and (1|Subject:FoodType) are used. The code to generate the complete mixed-effects model in this design as follows:

```
> data.model = lmer(Rating ~ FoodType + (1|Subject)+(1|Flavor) + (1|Subject:
  FoodType), data= mydata)
```

To view the results of the fixed-effects model, enter the name of fitted model object as a command or using summary() command. To obtain the estimates of degrees of freedom and p-values of the F-values for the fixed effects, use anova() function from lmerTest package.

```
> anova(data.model)
```

The output of fixed effect is contained as followed.



Figure 7 ■ Outputs from JMP (panel a), SAS (panel b) and SPSS (panel c) for the data of Table 3b

(a)

Fixed Effects Tests

Source	Nparm	DFNum	DFDen	F Ratio	Prob > F
FoodType	1	1	4.0	3.8684211	0.1206

(b)

Type 3 Tests of Fixed Effects				
Effect	Num DF	Den DF	F Value	Pr > F
FoodType	1	4	3.87	0.1206

(c)

Fixed Effects

Type III Tests of Fixed Effects^a

Source	Numerator df	Denominator df	F	Sig.
Intercept	1	4.000	57.553	.002
FoodType	1	4.000	3.868	.121

a. Dependent Variable: Rating.

Analysis of Variance Table of type III with Satterthwaite approximation for degrees of freedom

	Sum Sq	Mean Sq	NumDF	DenDF	F.value	Pr(>F)
FoodType	9.886	9.886	1	4	3.8684	0.1206

SAS

In this design, three are random effects: Subject, Flavor and the interaction between Flavor and FoodType. The code to generate the complete mixed-effects model in this design as follows:

```
PROC MIXED DATA=mydata;
  CLASS Subject FoodType Flavor;
MODEL Rating = FoodType / DDFM=SATTERTHWAITE SOLUTION;
  RANDOM Subject;
  RANDOM Flavor(FoodType);
  RANDOM Subject*FoodType;
RUN;
```

The output of fixed effect is shown in Figure 7, middle panel.

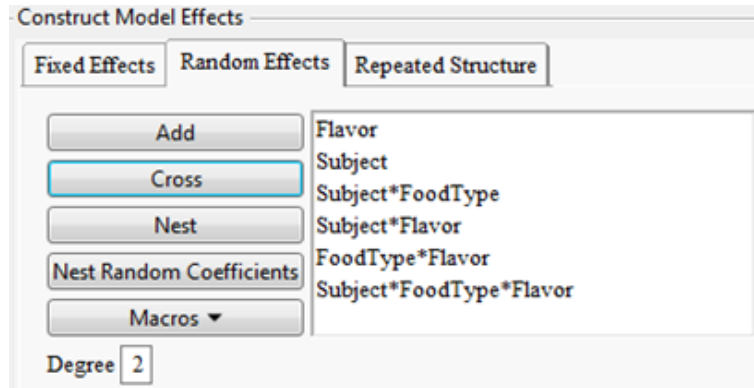
SPSS

In this design, there are three random effects: Subject, Flavor nested within FoodType, and the interaction between Subject and FoodType. The code to generate the complete mixed-effects model in this design as follows:

```
MIXED Rating BY Subject FoodType Flavor
  /FIXED = FoodType
  /RANDOM Subject Flavor(FoodType) FoodType*Subject
  /PRINT SOLUTION.
```



Figure 8 ■ Model effects for Within-Subjects and Within-Stimuli Design within JMP



The output of fixed effect is shown in Figure 7, bottom panel.

Section 6: Within-Subjects and Within-Stimuli Design (Table 4b data)

In this design, we would like to make a note that since Subject × FoodType × Flavor interaction consists of 3 levels of the grouping factor, which greater than the number of observations, R gives an error message. This issue is very unlikely to occur with real data using a reasonable sample size. Therefore, the error or warning messages regarding this are ignored and a way to override the error is given for the purpose of this tutorial.

JMP

In this design, there are six random effects: Subject, Flavor, the interactions between Subject and FoodType, Subject and Flavor, FoodType and Flavor, and among Subject, FoodType and Flavor. To indicate this, click on the “Random Effects” tab under “Construct Model Effects,” and then use “Add” button for “Subject” and for “Flavor.” Then, use the “Cross” button to create Subject*FoodType, Subject*Flavor, FoodType*Flavor, and Subject*FoodType*Flavor. Finally, click the “Run” function. Figure 8 illustrates this step.

The following code performs the same analysis. Once the code is entered, click on “Run Script” button to execute the script.

```
Fit Model(
  Y( :Rating ),
  Effects( :FoodType ),
  Random Effects(
    :Subject,
    :Flavor,
    :Subject * :FoodType,
    :Subject * :Flavor,
    :FoodType * :Flavor
    :FoodType * :Flavor * :Subject
  ),
  NoBounds( 0 ),
  Personality( ``Mixed Model`` ),
  Run( Repeated Effects Covariance Parameter Estimates( 0 ) )
);
```

The output of fixed effect is shown in Figure 9, top panel.



Figure 9 ■ Outputs from JMP (panel a), SAS (panel b) and SPSS (panel c) for the data of Table 4b

(a)

Fixed Effects Tests

Source	Nparm	DFNum	DFDen	F Ratio	Prob > F
FoodType	1	1	2.3	7.4954192	0.0961

(b)

Type 3 Tests of Fixed Effects				
Effect	Num DF	Den DF	F Value	Pr > F
FoodType	1	2.29	7.50	0.0961

(c)

Fixed Effects

Type III Tests of Fixed Effects^a

Source	Numerator df	Denominator df	F	Sig.
Intercept	1	8.416	22.241	.001
FoodType	1	2.294	7.495	.096

a. Dependent Variable: Rating.

R

In this design, there are six random effects: Subject, Flavor, the interactions between Subject and FoodType, Subject and Flavor, FoodType and Flavor, and among Subject, FoodType and Flavor. To indicate this, (1|Subject), (1|Flavor), (1|Subject:FoodType), (1|Subject:Flavor), (1|FoodType*Flavor), and (1|Subject:FoodType:Flavor) are used. The code to generate the complete mixed-effects model in this design as follows:

```
> data.model = lmer(Rating ~ FoodType + (1|Subject) + (1|Flavor) + (1|FoodType:
  Flavor) + (1|Subject:FoodType) + (1|Subject:Flavor) + (1|Subject:Flavor:FoodType)
  , data= mydata)
```

General speaking, the above codes would be successfully to produce the results. However, since the data in Table 4b, the number of subject (observation) is less than the number of factors, the R gave an error message as following.

```
Error: number of levels of each grouping factor must be < number of observations
```

In order to override the errors, the following codes were used to override the issues of grouping factor > number of observations. For more details about each code, please refer the package of lme4.

```
> data.model = lmer(Rating ~ FoodType + (1|Subject) + (1|Flavor) + (1|FoodType:
  Flavor) + (1|FoodType:Subject) + (1|Flavor:Subject) + (1|Flavor:Subject:FoodType
  ), data= mydata, control = lmerControl(check.nobs.vs.nlev="ignore", check.nobs.
  vs.nRE="ignore"))
```

To view the results of the fixed-effects model, enter the name of fitted model object as a command or using summary() command. To obtain the estimates of degrees of freedom and p-values of the F-values for the fixed effects, use anova() function from lmerTest package.

```
> anova(data.model)
```

The output of fixed effect is contained as followed.



Analysis of Variance Table of type III with Satterthwaite approximation for degrees of freedom

	Sum Sq	Mean Sq	NumDF	DenDF	F.value	Pr(>F)
FoodType	9.7672	9.7672	1	2.2935	7.4954	0.09613 .

SAS

In this design, there are six random effects: Subject, Flavor, the interactions between Subject and FoodType, Subject and Flavor, FoodType and Flavor, and among Subject, FoodType and Flavor. The code to generate the complete mixed-effects model in this design as follows:

```
PROC MIXED DATA=mydata;
  CLASS Subject FoodType Flavor;
  MODEL Rating = FoodType / DDFM=SATTERTHWAITE SOLUTION;
  RANDOM Subject;
  RANDOM Flavor;
  RANDOM Flavor*FoodType;
  RANDOM Subject*FoodType;
  RANDOM Subject*Flavor;
  RANDOM Subject*Flavor*FoodType;
RUN;
```

The output of fixed effect is shown in Figure 9, middle panel.

SPSS

In this design, there are six random effects: Subject, Flavor, the interactions between Subject and FoodType, Subject and Flavor, FoodType and Flavor, and among Subject, FoodType and Flavor. The code to generate the complete mixed-effects model in this design as follows:

```
MIXED Rating BY Subject FoodType Flavor
  /FIXED = FoodType
  /RANDOM Subject Flavor FoodType*Subject Flavor*Subject FoodType*Flavor Subject
  *FoodType*Flavor
  /PRINT SOLUTION.
```

The output of fixed effect is shown in Figure 9, bottom panel.

Section 7: Additional Information from Statistical Reports

In the previous sections, we focused on the statistical outputs of fixed effect, the primary interest of this paper. In this section, we discuss three additional statistical reports which might help some readers to interpret their results: fit statistics, random effects covariance parameter, and fixed effects parameter estimates. The fit statistics report gives statistics used for model comparison: information criteria such as Akaike Information Criterion (AIC) and Bayesian Information Criteria (BIC). AIC and BIC are information-based criteria that assess the model fit penalized for the number of estimated parameter. Due to the complication of computational issues in mixed effects model, the model selection is less straightforward than linear regression. Müller, Scealy, and Welsh (2013) discussed and compared four major model selection methods in mixed effects model. The second useful statistical report is the random effect covariance parameter estimates. This report provides details for the estimated covariance components of random effects in the model, which helps the readers understand the variances of assigned random effect factors in the model. For example, the estimated covariance of random effects in Table 1b were 0.22 for Subject, 6.55 for Flavor nested within FoodType, and 0.94 for the residuals. These suggest that Flavor nested within FoodType has larger variance than the Subject in the model. The third useful statistical report contains the fixed effects parameter estimates, which provides details for the fixed effect parameters specified in the model. It gives the estimate coefficient of the specified fixed effect factor in the model. However, because different contrasts produce different estimated coefficients, it is important for the readers to carefully consider the contrasts that are used to define the effects of factors.

Coefficient of determination, R^2 , is frequently used as a summary statistic to quantify the goodness-of-fit of fixed



effects models: what proportion of the dependent variable(s) can be explained by the independent variable(s). However, R^2 is not often reported in mixed effect models because there is no consensus of how to define R^2 in mixed effects model. A technical discussion of the definition of R^2 and the methods of calculating it are beyond the scope of this paper. Technical discussions and different methods to obtain R^2 can be found in the following sources (Liu, Zheng, & Shen, 2008; Nakagawa & Schielzeth, 2013; Orelieu & Edwards, 2008; Snijders & Bosker, 1994; Xu, 2003). SPSS, SAS, and JMP do not provide R^2 for mixed effects model, but R's `MuMIn` package (Barton, 2016) calculates R^2 based on Nakagawa and Schielzeth's (2013) and Johnson's (2014) methods. Because SPSS, SAS and JMP do not provide R^2 , we only demonstrated it in R. Nakagawa and Schielzeth (2013) categorized R^2 into two types: marginal and conditional R^2 . Marginal R^2 represents the variance explained by fixed effects, and conditional R^2 represents the variance explained by both fixed and random factors. Although we adopt Nakagawa and Schielzeth's (2013) methods to calculate the marginal R^2 and conditional R^2 , we do not take a position here on which methods are preferable.

We used the data from Table 1b as an example to demonstrate the calculation of conditional and marginal R^2 : `r.squaredGLMM()` function from `MuMIn` package.

```
> r.squaredGLMM(data.model)
```

The output of marginal R^2 is 0.13, and the conditional R^2 is 0.89. This means that 13% of the variances of Rating can be explained by FoodType, and 76% (89% - 13%) of the variance of Rating can be explained by the random factors (subjects and flavors). The estimated covariance of random effects in Table 1b shows that the variance for Subject is 0.2222, the variance for Flavor nested within FoodType is 6.5556, and the variance of residuals is 0.9444. Combining the outputs of covariance of random effects and `r.squaredGLMM`, the variance for the fixed effect is 1.1797 $((0.2222 + 6.5556 + 0.9444) \times 0.1325257 / (1 - 0.1325257) = 1.1797)$. Thus, the variance in the whole model is 8.90 $(0.22 + 6.56 + 0.94 + 1.18 = 8.90)$. The variance can be explained by subjects is 0.02 $(0.22 / 8.90 = 0.02)$, by Flavor is 0.74 $(6.56 / 8.90 = 0.74)$, and by FoodType is 0.13 $(1.18 / 8.90 = 0.13)$.

R^2_m	R^2_c
0.133	0.894

Open practices

- The *Open Data* badge was earned because the data of the experiment(s) are available on the [journal's web site](#).
- The *Open Material* badge was earned because supplementary material(s) are available on the [journal's web site](#).

Citation

Chang, Y.-H. A. & Lane, D. M. (2016). Generalizing across stimuli as well as subjects: A non-mathematical tutorial on mixed-effects models. *The Quantitative Methods for Psychology*, 12(3), 201–219. doi:[10.20982/tqmp.12.3.p201](https://doi.org/10.20982/tqmp.12.3.p201)

Copyright © 2016, Chang, Lane . This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) or licensor are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

Received: 09/06/2016 ~ Accepted: 23/09/2016