The Report Committee for Maliki Eyvonne Ghossainy
Certifies that this is the approved version of the following report:

The Utility of Hierarchical Logistic Regression for Predicting Repeated Measures Binary Responses

APPROVED BY
SUPERVISING COMMITTEE:

Supervisor: _____
Susan Natasha Beretvas

_____
Keenan Pituch

The Utility of Hierarchical Logistic Regression for Predicting Repeated Measures Binary Responses

by

Maliki Eyvonne Ghossainy, B.A., M.A., Ph.D.

Report

Presented to the Faculty of the Graduate School of

The University of Texas at Austin

in Partial Fulfillment

of the Requirements

for the Degree of

Master of Science in Statistics

The University of Texas at Austin

May 2016

Abstract

The Utility of Hierarchical Logistic Regression for Predicting Repeated Measures Binary Responses

Maliki Evyonne Ghossainy, M.S. Stat

The University of Texas at Austin, 2016

Supervisor: Susan Natasha Beretvas

This report will employ a hierarchical logistic regression model with mixed effects as an alternative to the traditional analysis of variance (ANOVA) approach that is often used when repeated observations are taken for each treatment. In the case of a binary response variable, ANOVA approaches typically require the user to first convert responses to an appropriate continuous variable, often a total score. The data used in this report include responses from 83 participants and are coded binary. Each participant was asked to make 12 separate decisions based on information received from videos of adult informants. The original purpose of the study was to determine the effect of subjects' age (between-subjects) and of video characteristics (within-subjects) on the likelihood that children will make the correct choice. The purpose of this report was instead methodological in nature. The results of the hierarchical model are compared to the results of a traditional mixed design analysis of variance to illustrate the strengths gained from applying hierarchical models to data that includes repeated observations per subject and to compare results when the dichotomous nature of the outcomes is appropriately modeled.

Table of Contents

List of Tables

List of Figures

Introduction

Repeated measures designs are common and useful tool for investigating changes in behavior over time or for testing differences across conditions where all subjects are exposed to all conditions. Repeated measures designs are used when all participants have more than one score on the same dependent variable at multiple measurement occasions. A repeated measures analysis of variance (RM ANOVA) allows for testing whether there are differences in related group means. The null hypothesis tests that all group means are equal. Rejecting the null hypothesis suggests that not all group means are equal. Importantly, the test of the null hypothesis is an omnibus test that simply signals some group differences without explicitly identifying which groups are statistically significantly different.

Repeated measures ANOVA can be an attractive alternative to independent samples ANOVA because of its ability to reduce unexplained variance. Whereas independent samples ANOVA can partition variance attributable to differences between conditions/treatments separately from error variance that is attributable to differences within each group (see Figure 1a), repeated measures ANOVA further accounts for variation that stems from individual differences. Because all of the same subjects comprise every condition, some of the overall variation in the data can be attributed to differences within individuals. Repeated measures ANOVA allows for modeling this within-subject variation, thereby reducing unexplained (error) variance (see Figure 1b). Estimating an additional parameter for within subject variation leads to a reduction in the error degrees of freedom. This, in turn, affects the calculation of the F-statistic. So long as the reduction in error variance is sufficiently large, the overall reduction in residual MSE can lead to a more powerful test of group differences.

Insofar as ANOVA techniques are meant to measure differences in means between treatments, this technique is most suitable for analyzing the relationship between a categorical predictor that has more than 2 levels and a quantitative outcome variable. Four main assumptions underlie the reliability of ANOVA results. The residuals are assumed to be independent and to follow a multivariate normal distribution. Repeated measures ANOVA further assume that subject-averaged pooled variances across repeated measures are equal across groups. Furthermore, repeated measures ANOVA designs assume that the variance-covariance matric is homogenous across groups. The condition of compound symmetry assumes that covariances between factor levels are all equal. This can be a restrictive assumption insofar as it implies that all subjects are affected by the treatment conditions in the same way. Such rigid constraints are not always met and severely affect the accuracy of results.

This type of analytical technique is quite popular within experimental psychology, particularly the field interested in measuring changes across development. Commonly, researchers within the field of developmental psychology create categories based on age, in order to accommodate the requirements of repeated measures ANOVA techniques. A study published in 2008 investigated the development of the number approximation ability (Halberda & Feigenson, 2008). The researchers recruited a group of children between 3- and 6-years of age and a group of adults to participate in a study in which they saw two displays that differed in the number of objects they contained. The subjects were asked to identify which display contained the most objects. Participants completed 66 trials and their responses were recorded and dummy coded as 1 and if they were correct and 0 if they were incorrect. To examine whether there were significant age differences in number estimation accuracy, researches categorized participants based on age (3-year-old vs. 4-year-old vs. 5-year old vs. 6-year old vs. adults) and computed a

total percent correct for each subject. Then, using within- and between-subjects ANOVA, the researchers compared the average percent correct across each age group. Throughout this remainder of this report, ANOVA models that include within- and between-subject variables will be referred to as a mixed ANOVA.

Another recent study in the *Journal of Experimental Child Psychology* was conducted by researchers interested in the role of parental emotional cues on infants smiling behaviors (Mireault, Crockenberg, Sparrow, Cousineau, Pettinato, & Woodard, 2015). Researchers compared smiling frequency between three groups of infant ages (namely, 5-, 6-, and 7-month old infants) across 6 within-subjects conditions. Through the use of a mixed ANOVA, the researchers concluded that the oldest (7-month old) infants were more likely to smile differentially as a function of parental cues than either group of younger infants (5- and 6-month olds).

While the strengths of mixed ANOVA make it an improvement over independent ANOVA designs, there are many limitations that are often overlooked or ignored. Many of these limitations can be addressed by using hierarchical models. Whereas ANOVA designs cannot accommodate missing data and typical ANOVA analyses instead exclude all cases with missing data, the estimation procedures used to estimate hierarchical models use all the available data.

The timing of repeated measures also presents a methodological constraint on repeated measures ANOVA designs but not on hierarchical models. Specifically, the ANOVA approach requires that repeated measurements be recorded at equal intervals for all subjects. Hierarchical models pose no such restriction – the interval between measurements is allowed to vary across subjects.

The purpose of the present report is to demonstrate differences in the estimates produced and to illustrate the strengths of using hierarchical models when repeated

observations are recorded for every subject. A typical design within experimental psychology is one that includes within- and between-subject variables as well as a binary response variable. Traditionally, a mixed ANOVA approach is the test of choice for such data, particularly within the field of developmental psychology. However, this test requires that a binary response data be transformed into a continuous measure, often through the computation of a total score. Such results are thus interpreted in terms of average score differences between groups. Hierarchical logistic regression approaches are an attractive alternative in that they respect the original format of the data; they provide better estimates of the variability due to within subject differences and provide estimates of the likelihood of success as a function of the included predictor variables.

Methods

PARTICIPANTS

The data were collected as part of a study investigating the development of children's ability to use verbal and nonverbal information in their decisions of trust. In the experimental study, 26 typically developing 4-year-olds (M = 4.3 years, range = 4.03-4.74), 29 5-year-olds (M = 5.5, range = 5.01 – 5.96), and 28 6-year-olds (M = 6.6, range = 6.08 – 7.05) were recruited to participate. These ages were chosen based on previous research which strongly suggests that the interval between 4 and 6 years is associated with dramatic improvements in children's selective trust across a variety of factors (e.g., Koenig & Jaswal, 2011; Pasquini, Corriveau, Koenig, & Harris, 2007). Participants included 40 females and 43 males.

MATERIALS

As part of the design of the study, children were asked to watch a series of testimony videos. There were 4 different types of videos (described below), with three trials for each type. Importantly, all videos showed an adult sitting behind a table, facing the camera with a clear, full frontal view of his/her face, arms, and upper body. Two different colored boxes of equal size and shape were shown resting on the table in front of the adult. The position of the boxes was randomized in order to prevent response patterns based on the color or position of the boxes. To minimize any unintended effects of speaker, no two videos had the same speaker and all actors were asked to wear a solid grey t-shirt provided by the researchers. Of the 12 testimony videos, five were filmed with male actors.

Types of Videos.

During the verbal testimony videos, the adult opened each box in sequence and looked in, keeping a neutral expression. After looking into both boxes, the adult looked up at the camera and said, "you should look in the (color) box", verbally suggesting one of the two boxes.

In the nonverbal testimony videos, the adult opened each box, in sequence, and looked inside. The adult looked into one of the boxes with a neutral expression but reacted excitedly upon looking into the other box. This nonverbal expression of excitement, happiness, and eagerness was achieved through a gasp, a smile, and raised eyebrows, and was meant to indicate that the contents of the box were interesting and highly desirable.

In the consistent testimony videos, the adult expressed excitement nonverbally towards the contents of one box and verbally suggested that the object was in that same box.

During the inconsistent videos, the adult expressed excitement nonverbally towards one of the boxes but stated that the object was in another.

PROCEDURE

After each video was played, children were asked to determine where the toy was hidden. Children's responses were coded dichotomously. In the verbal testimony videos, responses in favor of the verbally indicated box received a 1, whereas responses in favor of the alternative received a 0. In the nonverbal testimony videos, responses in favor of the nonverbally indicated box received a 1, whereas responses in favor of the alternative received a 0. In the consistent testimony videos, responses in favor of the box indicated both sources received a 1, whereas responses in favor of the alternative received a 0. In

the inconsistent testimony videos, responses in favor of the nonverbally indicated box received a 1, whereas responses in favor of the verbally indicated box received a 0. For ease of discussion throughout the remainder of the paper, all choices coded as 1 were considered the correct response for that video type.

The goal of this study was to examine the effect of nonverbal behavior in children's decisions of trust. If children choose in favor of the nonverbal testimony in the inconsistent trials, this would suggest that they were, indeed, able to draw on nonverbal behaviors when deciding whether to trust a speaker's words and would provide further evidence that children are equipped with mechanisms that protect against being deceived.

ANALYSES

In the present report, four distinct statistical models were estimated to detect differences between age groups on the decisions made across the 4 types of videos. Two models (Model 1a and 1b) were based on an ANOVA approach and used composite scores of the total number of correct responses as the outcome variable. Both ANOVA models provide a measure of the average differences in subjects' scores as a function of age group, type of video, and their interaction. In Model 1a, a mixed ANOVA was applied with age group as a between-group factor and type of video as a within-group factor. This model accounts for the repeated measures obtained on each subject. In Model 1b, an independent samples ANOVA was conducted. This model did not account for the nesting of responses within individual and instead treated all responses as independent outcomes. The estimates for the effect of each treatment were compared between models,

with a particular focus on the differences in explained error variance and the implications for detecting true differences between treatments.

The two remaining models (Model 2a and 2b) were based on logistic regression wherein binary outcome variables are adequately fitted. Results of the logistic regression analyses provided estimates of the likelihood of answering correctly as a function of age group and type of video. In Model 2a, a hierarchical logistic regression was built to account for the nesting of responses within subjects. Conversely Model 2b applied a simple logistic regression in which responses were treated as independent outcomes. Both logistic regression models were estimated using a binomial distribution. The likelihood estimates of Model 2a and 2b were compared, with an additional focus on differences in their standard errors. Finally, the differences between the ANOVA approach and the logistic regression approach were discussed.

Results

Preliminary analyses revealed that one type of video produced almost no variability in children's choices. Specifically, nearly all children chose correctly following the consistent videos. Because there were no differences in children's performance for this video type, it was excluded from further analyses.

MODEL 1A. ANOVA APPROACH WITH BETWEEN- AND WITHIN-SUBJECT VARIABLES

To predict the average number of correct responses as a function of age and the type of video, a mixed ANOVA was conducted in R Studio (version .99) using the aov command and specifying that responses for each video were nested within subject. The results in Table 1 show a significant interaction between age group and the type of video watched on children's average number of correct responses, $F(4,160) = 18.86$, $p<0.01$. In addition, the estimates for the main effect of age and video type were significant, $F(2,80) = 16.58$, $p<0.01$ and $F(2,160)=122.24$, $p<0.01$, respectively. Importantly, the estimate for the main effect of age is based on a residual variance estimate of 0.58, which is modeled to account for the repeated measures for each subject. This residual variance is the result of dividing the sum of squares, 46.44, by 80 degrees. This is different from the residual variance estimated for the effects of video and the interaction of video and age group. The estimate used there is computed separately by dividing the residual sums of squares by 160 degrees of freedom, leading to a residual variance of 0.54.

MODEL 1B. ANOVA APPROACH IGNORING NESTING OF OBSERVATIONS WITHIN INDIVIDUALS

The results shown in Table 2 are of a model that ignores the nesting of responses within individuals. This would be the model used if all observations were independent. The results of this model show significant main effects for age and type of video, $F(2,240)=17.47$, $p<0.01$ and $F(2,240)=18.35$, $p<0.01$. Additionally they show a significant interaction effect, $F(4,240)=18.35$, $p<0.01$. These results are based on a single estimate of residual variance of 0.5.

MODEL 2A. HIERARCHICAL LOGISTIC REGRESSION WITH RANDOM INTERCEPTS

The results in Table 3 are for testing the hierarchical logistic regression model with a random intercept to accommodate the multiple observations per individual subject. The logistic regression estimates the log odds of choosing the correct box as a function of age, video type, and their interaction. Log odds can be converted to odds and/or probabilities using simple algebraic calculations. The following equation represents the current model:

$$\text{Log}\,(p/1\text{-}p)_{ij} = \gamma_{00} + \gamma_{01}(4\text{yo})_{ij} + \gamma_{02}(6\text{yo})_{ij} + \gamma_{10}(\text{I})_{ij} + \gamma_{11}(4\text{yo})(\text{I})_{ij} + \gamma_{12}(6\text{yo})(\text{I})_{ij} + \gamma_{20}(\text{NV})_{ij} + \gamma_{21}(4\text{yo})(\text{NV})_{ij} + \gamma_{22}(6\text{yo})(\text{NV})_{ij} + \upsilon_{ij}$$

The results of the hierarchical model provide rich information about the differences in correctly choosing the box as a function of age group and type of video. Starting with the interaction between age group and video type, the results show a significant improvement in odds of success for 6 year olds responding to inconsistent videos relative to 5 year olds, $z=5.02$, $p<0.01$. No significant difference was observed in the odds of success for 4 year olds responding to inconsistent videos compared to 5 year

olds. Regarding the simple effects, results show that among 5 year olds (the reference age group), there is a significant decrease in the odds of success for inconsistent videos relative to verbal videos, $z=-5.68$, $p<0.01$. Finally, the test of the intercept is significant, suggesting the 5 year olds odds responding to verbal videos (reference age group and video type), have odds of success that are significantly different from chance, $z=5.59$, $p<0.01$.

MODEL 2B. LOGISTIC REGRESSION IGNORING THE NESTING OF OBSERVATIONS WITHIN INDIVIDUALS

Results of a logistic regression ignoring the nesting of measurement observations within individuals are reported in Table 4. Although many of the significance tests lead to similar inferences, the estimates are different from those observed in Model 2a. The test of the intercept suggests that the baseline condition, 5 year olds in response to verbal videos, have odds of success that are significantly different from chance, $z=5.24$, $p<0.01$. The results show a significant improvement in the odds of success for 6 year olds responding to inconsistent videos relative to 5 year olds, $z= 4.58$, $p<0.01$. There appears to be no significant change in odds of success for 4 year olds in response to inconsistent videos relative to 5 year olds. Among 5 year olds, there is no significant change in odds for nonverbal videos relative to verbal videos but there is significant decrease in odds of success for inconsistent videos relative to verbal videos among this reference age, $z=-6.39$, $p<0.01$. Finally, unlike the results from Model 2a, the results of Model 2b show a significant decrease in the odds of success for 6 year olds relative to 5 year olds in response to the verbal videos (reference video type), $z=-2.04$, $p=0.04$.

Discussion

The results from Model 1a and 1b illustrate some important differences that emerge when multiple responses are recorded for each individual rather than having observations that are completely independent. To start, one notable feature in Model 1a is in the way error variance is computed for each model. In Model 1a, the estimate of residual variance for the differences between age groups accounts for the fact that each subject has multiple observations. Thus, the test of significance for the between group factor can be evaluated using 80 denominator degrees of freedom rather than 240, with a resulting decrease in the MSE from .58 to .55. This leads to a more conservative, and arguably more accurate, test of the between group differences. Regarding the effect of video type and the interaction effect, Model 1a is more powerful than Model 1b for a number of reasons. First, specifying that responses be nested reduces the amount of unexplained variation from 132.23 to 85.79. Additionally, in Model 1a, the significance tests for video type and for the interaction term is evaluated using 160 denominator degrees of freedom rather than 240 and the residual mean squared error is slightly smaller compared to the independent model. This affects the calculation for each F test. The calculated F statistics for the differences between videos and for the interaction term are underestimated in the model without nesting. Had the effects of the treatments been smaller, the model without nesting would have increased the risk committing a Type 2 error.

Another limitation of these ANOVA models is that they provide limited information about the nature of the treatment effects. We can only conclude that there are significant differences in the number of correct responses as a function of the type of video, age, and their interaction, but additional analyses are needed to identify where these significant differences lie. As discussed before, the F-test is an omnibus test and signals overall differences between groups without specifying the specific direction of the effect and does not clarify which groups or combinations thereof actually might differ significantly on the outcome of interest. Pairwise comparisons are required to fully identify the nature of the treatment effects.

Finally, the ANOVA approach requires that the response variable be quantitative, thus, accommodations have to be made when the variable is in fact a binary response variable. In order to conduct the mixed ANOVA on this data, an aggregate of the number of correct responses (out of 3) for each type of video was computed. Thus, the analysis can only be interpreted in terms of average number of correct responses as a function of age and the type of video. Averaging across the number of correct responses diminishes the detectable differences between treatments; especially in scenarios in which the total possible number of correct responses is small. Here, the highest score a subject could obtain was 3. Furthermore, some significant differences might be difficult to interpret in a meaningful way. For example, if one treatment group had an average score of 1.4 (out of 3) and another had an average score of 1.7, what would one say about the importance of this difference? Arguably, one could claim that, in the context of the research question,

both groups' averages are indicative of poor performance, despite the statistical significance of the statistical test.

An alternative approach uses a multilevel logistic regression model to better respect the format in which the data was collected. This is because logistic regression models assume a binomial distribution. In the present report, the logistic regression provides a measure of the log odds of choosing correctly for each age group across each type of video.

The output from Model 2a provides a rich picture of how the groups differ relative to the reference categories. With the current reference group being 5 year olds and verbal videos, the results show us that the odds of success for 4 year olds responding to verbal videos is comparable to 5 year olds responding to this same kind of video. Likewise, 6-year-olds have comparable odds of success to 5-year-olds for verbal videos.

If we wanted to compare odds of success across the videos for a certain age group, we can look at the simple effects of each video type. Given that the reference group is 5 year olds and verbal videos, the simple effect of inconsistent videos suggests that 5-year-olds have worse odds of success for inconsistent videos compared to their odds on verbal videos. We can easily make similar comparisons for 4 year olds and 6 year olds by changing the reference category.

Examining the interaction effects, we observe a change in the odds for 6 year olds responding to inconsistent videos compared to 5 year olds responding to inconsistent videos. Older children have higher odds of success for inconsistent videos relative to 5-year-olds. There were no other significant changes in the odds for the interaction effects.

Changing the reference category easily allows us to examine differences in odds between different age groups as a function of different video types.

The hierarchical logistic regression clearly provides more information about the nature of the effects than does the mixed ANOVA analysis. In addition to the providing significance tests on the estimated odds of success for each parameter, it also provides easily interpretable information about the predicted probabilities for each age group across each video type. Table 5 presents these predicted probabilities for the current data. We see the predicted probability that children from each age group would respond correctly to each type of video - this is arguably more informative than a table of means that is provided by a mixed ANOVA technique.

Model 2b was conducted to show the differences in parameter estimates when nesting is excluded and responses are treated as independent. Although many of the statistical inferences were comparable between Model 2a and 2b, the estimates for the log odds of choosing correctly differ. Ignoring the nesting of responses within individuals led to an underestimation of the standard errors for all variables in Model 2b compared to Model 2a. This underestimation of the standard errors results from an inflated estimate of the sample size in Model 2b. Specifically, ignoring the nesting of responses within individuals implies that each observation is independent. These underestimated values of standard error increase the risk of committing a type 1 error; that is, detecting a false effect. Indeed, Model 2b suggests a significant difference in the log odds of choosing correctly in the verbal videos for 6 year olds relative to 5 year olds. This difference is not significant in the results from Model 2a.

The models shown in this report highlight only some of the differences that emerge from applying a hierarchical logistic regression instead of a mixed ANOVA when multiple binary responses are collected for each subject. The hierarchical logistic regression is more flexible, particularly in model specification, and can accommodate different variance-covariance relationships between responses. Whereas ANOVA techniques rely on the assumption that variances between treatments are equal, hierarchical models do not impose such a condition.

Moreover, hierarchical logistic regression models make use of all data, even when some missing data exists (assuming the data were missing at random). Had the data used here included subjects who only responded to some trials, those subjects would have been completely excluded from the mixed ANOVA analysis. Their data would not have been excluded from the hierarchical model, however. As long as missing observations could be assumed to be missing at random, any existing data would continue to be be utilized for estimating the parameters. In sum, researchers interested in measuring the effect of treatment variables for which repeated measures are collected are encouraged to consider the strengths of hierarchical regression models for estimating the effects of interest.

An additional strength of regression models over ANOVA approaches, in general, was not explicitly assessed in this report but warrants discussion, nonetheless. That is, unlike ANOVA approaches, logistic regression models (and regression models in general) do not require that predictor variables strictly be categorical. Using a logistic regression approach, a variable such as age, which is continuous in nature, can be kept in its original unit of measurement. In an ANOVA model, age needs to be artificially

categorized into different groups. Researchers who are interested in understanding developmental changes in behavior, for example, commonly measure age as a predictor variable. Applying a logistic regression analysis with age as a continuous predictor would provide a measure of the estimated change in the outcome for every unit change in age. This can be far more elucidating for understanding developmental trajectories than the results of an ANOVA model, which reports mean differences between age groups.

	df	Sum of Squares	Mean Squares	F	p-value
Age (yrs.)	2	19.25	9.62	16.58	<0.01
Residuals	80	46.44	0.58		

	df	Sum of Squares	Mean Squares	F	p-value
Video Type	2	131.09	65.55	122.24	<0.01
Video Type x Age	4	40.45	10.11	18.86	<0.01
Residuals	160	85.79	0.54		

Table 1: ANOVA with age as a between group factor and video as a within subject variable

	df	Sum of Squares	Mean Squares	F	p-value
Age (yrs.)	2	19.25	9.62	17.47	<0.01
Video Type	2	131.09	65.55	118.97	<0.01
Video Type x Age	4	40.45	10.11	18.35	<0.01
Residuals	240	132.23	0.55		

Table 2: ANOVA ignoring the nesting of observations within individuals

Random Effects				
Random Intercept for Subject Variance = 1.44				
Fixed Effects	Estimate	S.E.	Z	p-value
Intercept	4.27	0.76	5.59	>0.01
4-year-olds	-0.07	1.07	-0.07	0.94
6-year-olds	-1.66	0.87	-1.91	0.06
Inconsistent videos	-5.68	0.82	-6.90	<0.01
Nonverbal videos	-1.26	0.81	-1.56	0.12
4yo x Inconsistent	-0.54	1.14	-0.48	0.63
6yo x Inconsistent	4.73	0.94	5.02	<0.01
4yo x Nonverbal	-1.22	1.13	-1.08	0.28
6yo x Nonverbal	1.74	0.99	1.75	0.08
AIC = 519.6				

Table 3: Hierarchical Logistic Regression with Random Intercepts

	Estimate	S.E.	Z	p-value
Intercept	3.75	0.72	5.24	<0.01
4-year-olds	-0.11	1.01	-0.11	0.91
6-year-olds	-1.63	0.80	-2.04	0.04
Inconsistent videos	-4.83	0.76	-6.39	<0.01
Nonverbal videos	-1.31	0.82	-1.61	0.11
4yo x Inconsistent	-0.41	1.09	-0.38	0.70
6yo x Inconsistent	4.01	0.88	4.58	<0.01
4yo x Nonverbal	-0.97	1.12	-0.86	0.39
6yo x Nonverbal	1.76	0.99	1.78	0.07
AIC=539.91				

Table 4: Logistic Regression ignoring the nesting of observations within individuals

	Verbal	Nonverbal	Inconsistent
4-year-olds	0.99	0.86	0.13
5-year-olds	0.96	0.92	0.12
6-year-olds	0.86	0.93	0.75

Table 5: Predicted Probabilities from hierarchical logistic regression

Partitioning of Total Variability in Independent Measures Designs

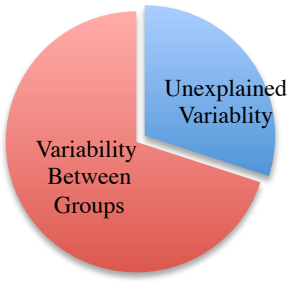

Figure 1a: Partitioning of Total Variability in Independent Measures Designs

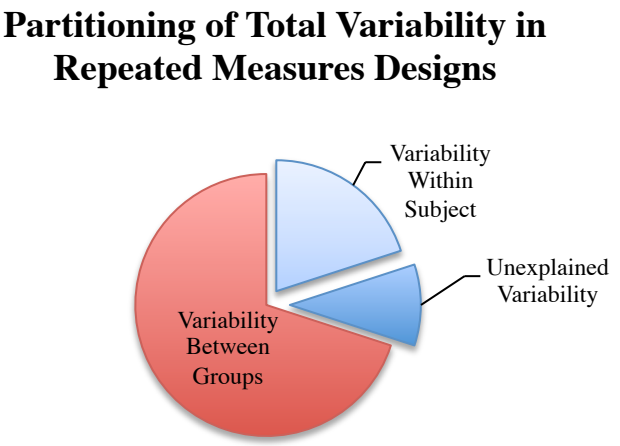

Figure 1b: Partitioning of Total Variability in Repeated Measures Designs

References

Halberda, J. & Feigenson, L. (2008). Developmental change in the acuity of the "number sense": The approximate number system in 3-, 4-, 5-, and 6-year-olds and adults. *Developmental Psychology*, *44(5),*1457-1465.

Mireault, G. C., Crockenberg, S. C., Sparrow, J. E., Cousineau, K., Pettinato, C., & Woodard, K. (2015). Laughing matters: Infant humor in the context of parental affect. *Journal of Experimental Child Psycology. 136,* 30-41.

Koenig, M. A., & Jaswal, V. K. (2011). Characterizing children's expectations about expertise and incompetence: Halo or pitchfork effects? *Child Development*, *82*(5), 1634-1647.

Pasquini, E. S., Corriveau, K. H., Koenig, M., & Harris, P. L. (2007). Preschoolers monitor the relative accuracy of informants. *Developmental Psychology, 43*(5), 1216-1226.