DOCUMENT RESUME

ED 474 271 TM 034 806

AUTHOR	Lamb, Gordon D.
TITLE	Understanding "Within" versus "Between" ANOVA Designs: Benefits and Requirements of Repeated Measures.
PUB DATE	2003-02-00
NOTE	37p.; Paper presented at the Annual Meeting of the Southwest Educational Research Association (San Antonio, TX, February 13-15, 2003).
PUB TYPE	Reports - Descriptive (141) -- Speeches/Meeting Papers (150)
EDRS PRICE	EDRS Price MF01/PC02 Plus Postage.
DESCRIPTORS	*Analysis of Variance; *Comparative Analysis; *Research Design
IDENTIFIERS	*Repeated Measures Design; Statistical Package for the Social Sciences

ABSTRACT

This paper discusses the basics of repeated measures designs. Within-subjects designs are compared to between-subjects designs, discussing the advantages and disadvantages of each. Further discussion compares a univariate one-way analysis of variance (ANOVA) with the between-subjects ANOVA and multivariate repeated measures ANOVA. Limitations of the univariate repeated measures ANOVA and their corrections are explained. This paper also demonstrates that the univariate repeated measures ANOVA is a form of linear regression. The advantages of linear regression over ANOVA are discussed briefly. The discussion concludes with examples of how to compute univariate, multivariate, and linear regression ANOVAs using the Statistical Package for the Social Sciences. Five appendixes contain tables of study results. (Contains 5 tables and 27 references.) (SLD)

Running head: UNDERSTANDING "WITHIN" VERSUS "BETWEEN" ANOVA

Understanding "Within" versus "Between" ANOVA Designs:

Benefits and Requirements of Repeated Measures

Gordon D. Lamb

Texas A&M University

Paper presented at the annual meeting of the Southwest Educational Research Association, San Antonio, TX, February 13-15, 2003.

Abstract

This paper discusses the basics of repeated measures designs. Within-subjects designs are compared to between-subjects designs, discussing the advantages and disadvantages of each. Further discussion compares a univariate one-way ANOVA with the between-subjects ANOVA and multivariate repeated measures ANOVA. Limitations of the univariate repeated measures ANOVA and their corrections are explained. This paper also demonstrates that the univariate repeated measures ANOVA is a form of linear regression. The advantages of linear regression over ANOVAs are discussed briefly. Discussion concludes with examples of how to compute univariate, multivariate, and linear regression ANOVAs using SPSS.

Understanding "Within" versus "Between" ANOVA Designs:

Benefits and Requirements of Repeated Measures

A repeated measure design is an experimental design that measures each participant on the dependent variable multiple times (Girden, 1992; Minke, 1997). Each time the participant is measured, he or she experiences different levels of the independent variable or factor (Heiman, 1999). This is known as a within-subjects factor (Stevens, 1996; Wells, 1998).

Repeated Measures Design

Types of Repeated Measures Designs

There are three ways of acquiring these multiple measures (Huck, 2000). Participants could perform one task during testing periods that are separated by a specified amount of time, such as when students take the same test at the beginning and end of a course. Participants could be measured several times during one testing period, performing a different treatment or activity each time (Huck, 2000). For example, in thought suppression studies participants are typically asked to not think about a specified thought and then to think about the thought and are measured on the number of times they think about the thought (Wegner, Schneider, Carter, & White, 1987). Participants could be measured on multiple characteristics during one testing period, such as the participant's views on various types of abuse (Huck, 2000).

Repeated measures designs are known by many names. A repeated measures design may also be called a within-subjects design (Girden, 1992). If the design contains both between and within-subject factors, it could be called a mixed-model design, a randomized blocks design, or a split-plot design (Barcikowski & Robey, 1984; Huynh & Feldt, 1970). This paper will only discuss fully within-subject designs.

Advantages of Repeated Measures Designs

Within-subject designs require fewer participants than between-subjects designs (Huck, 2000; Keselman & Algina, 1996; Minke, 1997; Tanguma, 1999). This is advantageous when random assignment is not possible, obtaining participants is expensive, or participants are hard to find (Keselman & Algina; Tanguma; Wells, 1998).

These designs require fewer participants since participants serve as their own control (Greenwald, 1976; Winer, 1962). The error variance attributed to individual variation is removed, resulting in more statistical power (Tanguma, 1999). Keppel and Saufley (1980) argue, "the primary source of error variance is the subjects" (p. 176). Taking out the variance due to individual differences increases the likelihood that differences between levels are due to the treatment itself and not the participants (Keppel & Zedeck, 1989; Keselman & Algina, 1996; Stevens, 1996). The statistical power of repeated measures designs will be discussed in greater detail later in this paper.

Disadvantages of Repeated Measures Designs

Although repeated measures designs have several advantages over between-subjects designs, they have several limitations. This design can be more time consuming than a study using separate groups for each level (Kogos, 2000). The additional time requirement could affect attrition rates (Girden, 1992). Because there are typically fewer participants in a repeated measures design, the results may not be as generalizable to other populations (Huck, 2000).

Another potential problem with repeated measures designs is the effect one treatment could have on subsequent treatments, a phenomenon known as a carry-over or practice effect (Huck, 2000; Keppel & Zedeck, 1989). Carry-over effects could cause biased estimates of the effect of the treatment (Keppel & Zedeck, 1989). Practice effects can be negative (deflating scores) or positive (inflating scores; Lewis, 1993, as cited in Wells, 1998).

There are several ways to minimize carry-over effects. In the case of boredom, monetary incentives may increase motivation, and rest periods may reduce fatigue (Keppel & Zedeck, 1989; Tanguma, 1999). To control for positive practice effects, Keppel and Saufley (1980) suggest having intervals between treatments long enough to allow the previous treatment's effect to dissipate or to bring the participant back to an agreed upon performance level before implementing the next treatment.

Counterbalancing

One of the most popular ways to control for carry-over and practice effects is counterbalancing. In counterbalancing, each of the treatments is given the same number of times at each level, and each treatment precedes the other treatments an equal number of times (Girden, 1992; Huck, 2000; Keppel & Zedeck, 1989).

Girden (1992) outlines two methods of counterbalancing, assuming a balanced design. For an even number of levels, the order for the levels of the first participant is 1, 2, n, 3, n-1, 4, n-2, etc., where the numbers refer to a level. The order for the second participant is found by adding 1 to each level in the first participant's order (because there is not a fifth level, 5 becomes 1). Table 1 gives an example of counterbalancing with four levels and four participants.

Table 1

Counterbalancing with an Even Number of Levels

	Level			
Participant	1	2	3	4
A	1	2	4	3
B	2	3	1	4
C	3	4	2	1
D	4	1	3	2

For an odd number of treatment levels, the order for the first participant follows the same pattern as for an even number of levels. The order for the second participant is found by reversing the order of the first participant. The order for the third and subsequent participants follows the same pattern as subsequent participants with an even number of levels.

Though counterbalancing is a useful tool, it does not completely prevent one treatment from affecting another (Kieffer, 1998, as cited in Wells, 1998). Latency effects can also be a problem. Girden (1992) defined latency effect as "an effect of treatment that is not evident until a second treatment is introduced" (p. 3). Allowing adequate time between treatments may prevent latency effects (Girden, 1992; Kogos, 2000; Tanguma, 1999).

Data Analysis

Univariate Repeated Measures and Between-subjects ANOVAs

Data from a repeated measures design can be analyzed through the use of a special univariate analysis of variance (ANOVA; Tanguma, 1999). This ANOVA is known as a repeated measures ANOVA or a within-subjects ANOVA (Huck, 2000). The purpose of a repeated measures ANOVA is the same as a between-subjects ANOVA. Both are used to see "whether the sample data cast doubt upon the null hypothesis" (Huck, 2000, p. 471). Understanding a repeated measures ANOVA can be accomplished by comparing it to a between-subjects one-way ANOVA.

Table 2

Example Data Set

	Treatment				
Participant	A	B	C	D	Indv $\overline{Y}$
1	1	2	3	4	2.5
2	2	2	4	8	4
3	3	5	6	7	5.25
4	5	5	6	8	6
$\overline{Y}_J$	2.75	3.5	4.75	6.75	$\overline{Y}_G = 4.44$

Sums of Squares

In a between-subjects ANOVA there are "three sources of variability… treatment effects, individual differences, and experimental error" (Tanguma, 1999, p. 243). Because a repeated measures ANOVA removes the variance due to individual differences, there are only two "sources of variability." The reduction of the error term decreases the chance of a Type II error (Stevens, 1996). Greenwald (1976) argued that it is possible to have the same statistical power using a within-subjects design with $1/J$ subjects fewer than a between-subjects design (J represents the number of treatments).

Table 2 provides data to compare the between-subjects ANOVA to the within-subjects ANOVA. Table 2 provides the data for the within-subjects ANOVA in which all four participants receive each of four treatment conditions, although they do so in a counterbalanced order.

Partitioning sums of squares begins the same way as in a between-subjects ANOVA. The total sums of squares, SOS_{tot},, is computed with Equation 1:

$$SOS_{tot} = \sum(Y - \overline{Y}_G)^2, \tag{1}$$

where Y equals an individual score and $\overline{Y}_G$ equals the grand mean. SOS_{tot} measures the variability of the individual scores around the grand mean (Haase & Thompson, 1992)

The between-groups sum of squares, SOS_B, or treatment sum of squares for a repeated measure, SOS_{treat}, indicates the proportion of the total variance that is due to the treatment (Bartz, 1999). It is found using Equation 2:

$$SOS_B = n\sum(\overline{Y}_k - \overline{Y}_G)^2 = SOS_{treat}, \tag{2}$$

where n equals the number of participants in each group or treatment and $\overline{Y}_k$ equals a group or treatment mean. If the null hypothesis is true, the group means should be equal to the grand mean. SOS_B and SOS_{treat} are measures of the deviation of the groups form the grand mean (Haase & Thompson, 1992). The larger the SOS_B or SOS_{treat}, the more likely the results will be statistically significant (Hinkle, Wiersma, & Jurs, 1998).

The error or residual sum of squares, SOS_{res}, represents the uncontrollable variability of the study (Keppel & Zedeck, 1989). This score is a measure of how much the individual scores deviate from their respective means, or the variability within the group or treatment. It is obtained by using Equation 3:

$$SOS_{res} = \sum (Y - \overline{Y}_k)^2 . \tag{3}$$

In the repeated measures ANOVA, the sum of squares due to individual differences, or subject sum of squares, SOS_s is also calculated (Equation 4):

$$SOS_s = k\sum (\overline{Y}_S - \overline{Y}_G)^2 , \tag{4}$$

where $\overline{Y}_S$ equals a participant's mean score across treatment conditions. SOS_s measures the variability of a participant's score around his or her mean across treatments. The SOS_s is subtracted from the SOS_{res}, resulting in a larger Fcalc.

Table 3

One-way Between-subjects ANOVA Summary

Source	SS	df	MS	F	Eta Sq
SOS_B	36.69	3	12.23	4.16*	51%
SOS_{res}	35.25	12	2.94		
SOS_{tot}	71.94	15			

*$p < .05$.

Computing Fcalc

Table 3 is a summary of the one-way between-subjects ANOVA. The between-groups degrees of freedom (*df*) is $k - 1$, residual *df* is $n - k$, and the total *df* is $N - 1$ (N = total participants). The *df* are additive in that $df_{tot} = df_B + df_{res}$.

The mean square, *MS*, for each row is obtained by dividing the sums of squares by their corresponding *df*, resulting in the *MS* between, MS_B, and *MS* residual, MS_{res}. The *MS* total is not needed. *F*calc equals MS_B / MS_{res}. The effect size, eta squared, η^2, for the treatment is found by taking $SOS_B / SOS_{tot} * 100$. Eta squared allows the reader to know the percentage of the variance explained by the sums of squares from which it was calculated (Cohen, 2001).

Table 4 is a summary for the one-way repeated measures ANOVA. The *df* for subjects is $n - 1$, *df* for treatment is $k - 1$, *df* residual is $(n - 1)(k - 1)$, and *df* total is $n_T - 1$ (n_T equals the total number of scores).

Table 4

One-way Repeated Measures ANOVA Summary

Source	*SS*	*df*	*MS*	*F*	Eta Sq
SOS_S	28.19	3	9.40		39%
SOS_{treat}	36.69	3	12.23	15.68**	51%
SOS_{res}	7.06	9	.78		
SOS_{tot}	71.94	15			

**$p < .01$

The between-subjects design had four times as many participants; however, the *F*calc for the repeated measures ANOVA was almost four times as large. Despite the reduced degrees of freedom (3, 9), the repeated measures ANOVA had a smaller statistical probability ($p = .001$) than the between-subjects ANOVA (3, 12, $p = .031$).

Assumptions of Univariate Repeated Measures ANOVA

Independence of Observations

In order for the results of a repeated measures ANOVA to be accurate, three assumptions must be met (Cohen, 2001; Huynh & Feldt, 1970; Stevens, 1996). The violation of these assumptions can lead to an increased Type I error rate (Hinkle, Wiersma, & Jurs, 1998). The first assumption, independence of observations, is typically assumed through random selection (Keppel & Zedeck, 1989). There are some instances, however, where dependent observations are made, such as in cooperative learning (Stevens, 1996). In this example, interaction of the group is intended to affect the scores of its members. Correlated observations typically cause an overestimation of the true probability and can be resolved through using a more conservative probability level (Stevens, 1996).

Multivariate Normality

The repeated measures ANOVA is robust to violations of the second assumption, multivariate normality. "The ANOVA *F* test are robust to nonnormality in the sense that the actual probability of a Type I error would be close to the nominal level" (Wilcox, 1997, p. 7). This assumption would have to be "severely violated" (Cohen, 2001, p. 451) with a small sample size to have a marked effect on the test statistic. In this rare situation, Cohen suggests using a nonparametric test or a data transformation.

Sphericity Assumption

Assessing sphericity. The third assumption, sphericity, is the requirement "that variances of differences for all treatment combinations be homogenous (i.e. $\sigma^2_{y1-y2} = \sigma^2_{y2-y3}$, etc.;" Girden, 1992, p. 6). In other words, the variances should meet "a set of acceptable patterns" (Huck, 2000, p. 477) or "people should respond similarly across treatments" (Kogos, 2000, p. 8).

If the variance of the differences of treatment levels is not equal, the *F*calc would tend to overestimate the statistical significance level (Box, 1954; Huck, 2000; Stevens, 1996). This could potentially lead to an increased Type I error rate (Stevens, 1996).

Girden (1992) argued that it is rare for homogeneity to exist among variance differences when studies have more than two levels. When there are only two levels of the repeated measure, sphericity is not an issue (Edwards, 1985). In this case, there is not another variance of difference to compare against, thus "homogeneity must exist" (Girden, p. 18).

The variance of differences between pairs of scores, $\sigma^2_{y_1-y_2}$, can be found two ways. First:

$$\sigma^2_{y_1-y_2} = \sigma^2_1 + \sigma^2_2 - COV, \tag{5}$$

where σ_1^2 equals the variance of one set of scores, σ_2^2 equals the variance of the paired set of scores, and COV equals the covariance (Girden, 1992). Covariance is computed by $[\sum(X-\overline{X})(Y-\overline{Y}_G)]/n$, where X equals the group or level number and $\overline{X}$ equals the average of the group numbers.

The second way of computing the difference between pairs of scores is to subtract the individual scores of one treatment from another to obtain a difference scores, then compute the variance of the difference scores (Girden, 1992). Variance can be computed by $[\sum(D-\overline{D})^2]/n-1$, where D equals an individual difference score and $\overline{D}$ equals the mean of the difference scores.

Conservative F. When the sphericity assumption is violated, there are several corrections that can be made. The most popular corrections involve decreasing the degrees of freedom, and thus the *F*calc (Huck, 2000). The Geisser and Greenhouse conservative *F*-test is the simplest correction. For this adjustment, the degrees of freedom would be 1 for the numerator and $n-1$ for the denominator (Girden, 1992; Huck, 2000). This is assuming that sphericity has been violated to highest extent; therefore, this is a very conservative test. Stevens (1996) explained "this makes the test very conservative, since adjustment is made for the worst possible case, and we don't recommend it" (p. 460). This procedure often overcorrects for violations of sphericity (Huck, 2000).

Another method of correcting the degrees of freedom is multiplying the degrees of freedom by the correction factor epsilon, ε (Girden, 1992; Huck, 2000; Huynh & Feldt, 1976; Stevens, 1996). O'Brien and Kaiser (1985) explain "ε is a measure of nonsphericity" (p. 319), a smaller epsilon means a further departure from sphericity. The range of epsilon is from 1.0 to $1/J$-1 (Box, 1954; Greenhouse & Geisser, 1959). If the variances of difference are not exactly the same, epsilon will be less than 1.0 (Huynh & Feldt).

Epsilon hat adjustment. The Geisser-Greenhouse adjustment, or epsilon hat, $\hat{\varepsilon}$, is an estimation of epsilon which ranges from 1.0 to $1/J-1$ (Girden, 1992). Once computed, epsilon hat is then multiplied by both degrees of freedom to more closely estimate *F*crit (Huynh & Feldt, 1976).

Table 5

Variance-covariance Matrix

	Treatment			
Treatment	1	2	3	4
1	2.92	2.5	2.25	2.25
2	2.5	3	2.5	1.5
3	2.25	2.5	2.25	1.92
4	2.25	1.5	1.92	3.58

A variance-covariance matrix must be constructed to compute epsilon hat. Using the data from Table 2, calculate the variances for each treatment condition. Starting in the upper right hand corner with treatment 1, place the variances along the diagonal axis. Next compute the covariance for each possible pair of treatments. Place each covariance in the cells where both treatments intersect. Table 5 provides the completed variance-covariance matrix.

Epsilon hat is computed using Equation 6:

$$\hat{\varepsilon} = \frac{J^2(\overline{D} - \overline{Cov}_T)^2}{(J-1)(\sum Cov_{ij}^2 - 2J\sum \overline{Cov}_i^2 + J^2\overline{Cov}_T^2)}, \qquad (6)$$

where $\overline{D}$ equals the mean of variances along the diagonal, $\overline{Cov}_T$ equals the mean of all entries in the matrix, Cov_{ij}^2 equals a squared entry in the matrix, and $\overline{Cov}_i$ equals the mean of the entries of a row in the matrix. In this instance, $\hat{\varepsilon} = 0.47$.

Our calculated epsilon hat is then multiplied by each degree of freedom resulting in the new $F(1.4, 4.2) = 15.58$, $p = .013$. It is also interesting to note that the eta squared effect size is unaffected by this correction, although a small epsilon value may suggest the necessity of using a different effect size estimate.

Epsilon tilde. Another correction for a violation of the sphericity assumption is to use the Huynh-Feldt epsilon tilde, $\tilde{\varepsilon}$. Equation 7 gives the formula for epsilon tilde:

$$\tilde{\varepsilon} = \frac{[N(J-1)\hat{\varepsilon}]-2}{(J-1)[N-k-(J-1)\hat{\varepsilon}]}, \tag{7}$$

where k equals the number of groups, or 1 for a single-factor study (Girden, 1992). Using the sample data from Table 2, epsilon tilde equals 0.76.

The epsilon tilde is multiplied by each degree of freedom producing a new *F*crit but not a new *F*calc. However, the lower degrees of freedom associated with epsilon tilde increased the *F*crit, and therefore makes obtaining statistical significance more difficult.

Epsilon hat or tilde. Each correction reduces the Type I error rate, compared to an unadjusted *F*crit, but they also have their drawbacks. The conservative *F* test is good for making a quick evaluation of the power of the test statistic, but it is often too conservative (Huck, 2000). Epsilon hat is the best estimator of epsilon when epsilon is less than .75, but it tends to underestimate epsilon if "epsilon is near or a little above .75" (Huynh & Feldt, 1976, p. 71). Epsilon tilde is the best predictor of epsilon when epsilon is near or above .75, but as epsilon falls below .75, epsilon tilde tends to overestimate epsilon. As Huynh and Feldt (1976) argued, "the difference between epsilon tilde and epsilon hat decreases with increasing *N*" (p. 75). "It would be desirable to have an unbiased estimator for ε. Such an estimator, unfortunately, is not known" (Huynh & Feldt, 1976, p. 73). Girden (1992) and Stevens (1996) recommend averaging epsilon tilde and epsilon hat to obtain a more accurate epsilon.

Univariate and Multivariate Repeated Measures ANOVAs

Another way to deal with potential violations of the sphericity assumption is to use a multivariate repeated measures ANOVA. Sphericity is not necessary with this ANOVA because the test statistic uses transformed variables instead of the raw scores (Girden, 1992; Stevens, 1996; Wells, 1998). This procedure treats the different treatments for the individuals as separate dependent variables and the treatment scores can be come correlated with each other (Kogos, 2000; Minke, 1997).

In cases where the sphericity assumption is violated, the multivariate ANOVA may have more statistical power against Type II errors (Girden, 1992; Stevens, 1996). In using transformed scores, "the researcher has lost the advantage of repeatedly measuring participants because now each measurement is a separate dependent variable" (Kogos, 2000, p. 10). The multivariate approach may be more statistically powerful with larger sample sizes (Stevens, 1996). A good rule of thumb is to have at least ten participants more than the number of levels when using the multivariate repeated measures approach (Stevens, 1996).

If the sphericity assumption is not violated, the univariate ANOVA is more powerful because it has a higher degree of freedom than Hotellings T^2 (Girden, 1992; Stevens, 1996). For example, the multivariate repeated measures

ANOVA has a $F(3, 1)$, = 6.33, p = .282. To have statistically significant results, with the sample data, F would need to equal 261. In deciding between using the univariate or multivariate approach, one must consider the sample size and the possibility or predicted extent of violation of the sphericity assumption.

Linear Regression Repeated Measures ANOVA

Basics of Linear Regression

A univariate repeated measures ANOVA can be run using linear regression. In linear regression what is known about one variable is used to make predictions about the other variable (Keppel & Zedeck,1989) and "a less frequent but equally plausible use is to test hypotheses" (p. 58). The linear regression equation is $\hat{Y} = B_0 + B_1X$, where B_0 and B_1 are the constants (Y intercept and slope, respectively), and $\hat{Y}$ is the predicted Y value for a given X value.

The slope is found using Equation 8:

$$B_1 = \frac{\sum[(X - \bar{X})(Y - \bar{Y})]}{\sum(X - \bar{X})^2}, \tag{8}$$

where $\overline{X}$ equals the mean of all the X scores. Using the data from Table 2, the X score would be the number corresponding to the treatment. The result is four 1s, four 2s, and so forth. Once the slope is calculated, the regression equation can be used to find the *Y* intercept. The *Y* intercept is calculated $B_o = \overline{Y_G} - B_1(\overline{X})$.

Next, the regression formula can be used to calculate the predicted *Y* values. Once the constants, B_o, B_1, and the predicted *Y* values are found, the sums of squares can be partitioned. The total sums of squares is found using Equation 1. The regression sum of squares is found using Equation 9:

$$SOS_{reg} = \sum(\hat{Y} - \overline{Y}_G)^2 . \qquad (9)$$

This sum of squares is used in the same way as the treatment sum of squares in the ANOVA. This is similar to the previous formula for the treatment sum of squares, except *Y* predicted is used instead of the group mean and the equation is not multiplied by the sample size. Because a *Y* predicted for each person will enter the equation, there is no need to multiply by *n*.

The residual sum of squares is obtained by subtracting each observed *Y* from its respective predicted value, then squaring and summing the difference scores, as seen in Equation 10:

$$SOS_{res} = \sum(Y - \hat{Y})^2 . \qquad (10)$$

This is the same formula used earlier, except *Y* predicted is used instead of the group mean. The further *Y* observed deviates from the *Y* predicted on the regression line, the larger the error term will be (Neter, Kutner, Nachtsheim, & Wasserman, 1996).

Least Squares Method

If the data in the regression equation were a perfectly linear relationship, the predicted *Y* for each score would equal its treatment mean. This revelation is intuitive because the best predictor for *Y* without knowing about *X* is the mean of *Y* (Keppel & Zedeck, 1989). In this situation the treatment sums of squares is at its maximum and the residual sum of squares is at its minimum (Edwards, 1985). This regression line is termed the method of least squares (Edwards). Using this method, the sums of squares in the repeated measures ANOVA can be translated into the ones used previously.

Advantages of Linear Regression

An ANOVA is a simplified form of linear regression (Edwards, 1985). Linear regression has a major advantage over an ANOVA. An ANOVA uses a nominal or ordinal scale for the independent variable whereas linear regression uses data at any scale for the independent variables (Cohen, 2001). For example, if a researcher were studying how well a test score predicts future performance using an ANOVA, he or she would have to turn interval data (test scores) into a nominal scale by chunking groups of scores together, 100-95 points, 94-90 points, and so forth. In doing so, valuable information would be lost (Pedhazur, 1982, as cited in Haase & Thompson, 1992). Haase and Thompson (1992) argue that changing interval variables to nominal dichotomies or trichotomies, distorts the shape, variability, and relationships between variables.

Disadvantages of Linear Regression

The disadvantage of linear regression is that its statistical power decreases the further the treatment means move from a straight line (Cohen, 2001). This causes the between sum of squares to be greater than the regression sums of squares and subsequently, a larger residual sum of squares (Cohen).

To maintain the ability to use continuous independent variables while studying nonlinear relationships, Cohen (2001) suggests multiple regression. Multiple regression can tell the researcher about the shape of the relationship in addition to doing everything an ANOVA can do (Haase & Thompson, 1992).

Data Analysis with SPSS

For large data sets, performing computations by hand is very tedious and time consuming. Thankfully, there are commercial statistical packages that can assist in these situations. When using statistical programs, it is important to remember they were written by humans, and thus, may have mistakes. A statistical package does not replace a statistically sophisticated mind. For the purpose of this paper, analysis of data will be discussed using SPSS version 11.0.

Linear Regression Repeated Measures ANOVA

To use SPSS for linear regression, the dependent variables must first be made into $J-1$ orthogonal contrasts (Minke, 1997). To understand the process by which the contrasts are developed, the reader is referred to Edwards (1985), Neter et al. (1996), or Keppel and Zedeck (1989). Appendix A is a modification of the coding table presented by Edwards (1985, p. 124).

In Appendix A, V1 is the coding for the linear model explained above. V2 and V3 are the quadratic and cubic model vectors respectively. These vectors allow the researcher to study nonlinear relationships. The vector in the summary table with the largest sum of squares is the best fit for the data, or best explains the shape of the relationship (Keppel & Zedeck, 1989). The last vector contains the sum for each participant. This represents the variability of the subjects (Keppel & Zedeck, 1989).

The dependent variable and vectors can be typed directly onto the data editor, which is a spreadsheet similar in appearance to Lotus or Excel. For those who prefer the use of a mouse, SPSS is very accommodating. To do this analysis, go to the pull down menu and click *Analyze,* then go down to *Regression* and click *Linear*. In the linear regression dialog box, highlight the dependent variable *y* and click on the arrow beside the *Dependent* box. Then highlight the vectors and click the arrow beside the *Independent(s)* box and then click *OK*. An ANOVA summary table is then created in the output viewer (See Appendix B). This same analysis can be done using the syntax editor. Once one understands how to use syntax, it can be used to tailor the analysis to his or her specific needs. (See Appendix C for the syntax of this analysis). A good way to begin using syntax is to use the paste command located in the dialog box.

Computing Univariate and Multivariate Repeated Measures ANOVAs

To run a univariate or multivariate repeated measures ANOVA, place the scores for each treatment in its own column. Click *Analyze*, go to *General Linear Model*, and under that menu click *Repeated measures*. In the Repeated Measures Define Factor(s) dialog box enter 4 in the box labeled *Number of Levels*, click *Add*, and then *Define*. In the Repeated measures dialog box, highlight the variables representing each treatment level, click the arrow beside the *Within-Subjects Variable* box and click *OK*. The output viewer will display summary tables for the multivariate and univariate repeated measures ANOVAs. (See Appendix D for the output and Appendix E for the syntax.)

Conclusion

Repeated measures designs have several advantages over between-subjects designs, including greater statistical power with fewer participants. Counterbalancing is suggested to minimize carryover effects. The repeated measures ANOVA has greater power against Type II errors, because it explains more of the variance than between-subjects ANOVAs.

If the sphericity assumption is violated the *F*calc can become inaccurate. For this reason, it is suggested that the degrees of freedom be adjusted by an estimator of epsilon. The best estimator of epsilon depends on how much the sphericity assumption is violated. In cases where sphericity is violated, one may consider using a multivariate repeated measures ANOVA. An ANOVA is a general form of linear regression.

Linear regression may be preferred over an ANOVA because it is not limited to testing nominal or ordinal independent variables. Calculating test statistics are easier with the use of statistical programs, such as SPSS; however, a firm understanding of the statistics being used is still required.

References

Barcikowski, R. S., & Robey, R. R. (1984). Decisions in single group repeated measures analysis: Statistical tests and three computer packages. *The American Statistician, 38*, 148-150.

Bartz, A. E. (1999). *Basic statistical concepts* (4th ed.). Upper Saddle River, NJ: Prentice Hall.

Box, G. E. P. (1954). Some theorems on quadratic forms applied in the study of analysis of variance problems, II. Effects of inequality of variance and of correlation between errors in the two-way classification. *Annals of Mathematical Statistics, 25*, 484-498.

Cohen, B. H. (2001). *Explaining psychological statistics* (2nd ed.). New York: John Wiley.

Edwards, A. L. (1985). *Multiple regression and the analysis of variance and covariance* (2nd ed.). New York: W. H. Freeman.

Girden, E. R. (1992). *ANOVA: Repeated measures*. Newbury Park, CA: Sage.

Greenhouse, S. W., & Geisser, S. (1959). On methods in the analysis of profile data. *Psychometrika, 24*, 95-112.

Greenwald, A. G. (1976). Within-subjects designs: To use or not to use?. *Psychological Bulletin, 83*, 314-320.

Haase, T., & Thompson, B. (1992, January). *The homogeneity of variance assumption in ANOVA: What it is and why it is required.* Paper presented at the annual meeting of the Southwest Educational Research Association, Houston, TX.

Heiman, G. W. (1999). *Research methods in psychology* (2^{nd} ed.) Boston, MA: Houghton Mifflin.

Hinkle, D. E., Wiersma, W., & Jurs, S. G. (1998). *Applied statistics for the behavioral sciences* (4^{th} ed.). Boston, MA: Houghton Mifflin.

Huck, S. W. (2000). *Reading statistics and research* (3^{rd} ed.). New York: Longman.

Huynh, H., & Feldt, L. S. (1970). Conditions under which mean square ratios in repeated measurements designs have exact F-distributions. *Journal of the American Statistical Association, 65*, 1582-1589.

Huynh, H., & Feldt, L. S. (1976). Estimation of the Box correction for degrees of freedom from sample data in randomized block and split-plot designs. *Journal of Educational Statistics, 1*, 69-82.

Keppel, G., & Saufley, W. H., Jr. (1980). *Introduction to design and analysis: A student's handbook*. San Francisco, CA: W. H. Freeman.

Keppel, G., & Zedeck, S. (1989). *Data analysis for research designs: Analysis of variance and multiple regression/correlation approaches*. New York: W. H. Freeman.

Keselman, H. J., & Algina, J. (1996). The analysis of higher-order repeated measures designs. In B. Thompson (Ed.), *Advances in social science methodology* (Vol. 4, pp. 45-70). Greenwich, CT: JAI Press.

Kogos, S. C., Jr. (2000). R*epeated measures designs and the sphericity assumption* (TM 033 280). (ERIC Document Reproduction Service No. ED457184)

Minke, A. (1997, January). *Conducting repeated measures analyses: Experimental design considerations* (TM 026 436). Paper presented at the annual meeting of the Southwest Educational Research Association, Austin, TX. (ERIC Document Reproduction Service No. ED407415)

Neter, J., Kutner, M. H., Nachtsheim, C. J., & Wasserman, W. (1996). *Applied linear statistical models* (4th ed.). Chicago, IL: Irwin.

O'Brien, R. G., & Kaiser, M. K. (1985). MANOVA method for analyzing repeated measures designs: An extensive primer. *Psychological Bulletin, 97*, 316-333.

Stevens, J. (1996). *Applied multivariate statistics for the social science* (3rd ed.). Mahwah, NJ: Erlbaum.

Tanguma, J. (1999). Analyzing repeated measures designs using univariate and multivariate methods: A primer. In B. Thompson (Ed.), *Advances in social science methodology* (Vol. 5, pp. 233-250). Stamford, CT: JAI Press.

Wegner, D. M., Schneider, D. J., Carter, S. R., III, & White, T. L. (1987). Paradoxical effects of thought suppression. *Journal of Personality and Social Psychology, 53,* 5-13.

Wells, R. D. (1998, November). *Conducting repeated measures analyses using regression: The General Linear Model lives* (TM 029 320). Paper presented at the annual meeting of the Mid-South Educational Research Association, New Orleans, LA. (ERIC Document Reproduction Service No. ED426091)

Wilcox, R. R. (1997). *Introduction to robust estimation and hypothesis testing.* San Diego, CA: Academic Press.

Winer, B. J. (1962). *Statistical principles in experimental design.* New York: McGraw-Hill.

Appendix A

Orthogonal Coding of the Sample Data from Table 2

	Vectors				
Participant	V1	V2	V3	Sum	Y
1	-3	1	-1	10	1
2	-3	1	-1	16	2
3	-3	1	-1	21	3
4	-3	1	-1	24	5
1	-1	-1	3	10	2
2	-1	-1	3	16	2
3	-1	-1	3	21	5
4	-1	-1	3	24	5
1	1	-1	-3	10	3
2	1	-1	-3	16	4
3	1	-1	-3	21	6
4	1	-1	-3	24	6
1	3	1	1	10	4
2	3	1	1	16	8
3	3	1	1	21	7
4	3	1	1	24	8

Appendix B

Linear Regression ANOVA Output

Regression

Variables Entered/Removed[b]

Model	Variables Entered	Variables Removed	Method
1	SUM, V3, V2, V1[a]	.	Enter

a. All requested variables entered.

b. Dependent Variable: Y

Model Summary

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	.950[a]	.902	.866	.80128

a. Predictors: (Constant), SUM, V3, V2, V1

ANOVA[b]

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	64.875	4	16.219	25.261	.000[a]
	Residual	7.062	11	.642		
	Total	71.938	15			

a. Predictors: (Constant), SUM, V3, V2, V1

b. Dependent Variable: Y

Coefficients[a]

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.
		B	Std. Error	Beta		
1	(Constant)	.000	.699		.000	1.000
	V1	.663	.090	.699	7.395	.000
	V2	.313	.200	.147	1.560	.147
	V3	1.250E-02	.090	.013	.140	.892
	SUM	.250	.038	.626	6.626	.000

a. Dependent Variable: Y

Appendix C

Linear Regression ANOVA Syntax

```
REGRESSION
  /MISSING LISTWISE
  /STATISTICS COEFF OUTS R ANOVA
  /CRITERIA=PIN(.05) POUT(.10)
  /NOORIGIN
  /DEPENDENT y
  /METHOD=ENTER v1 v2 v3 sum.
```

Appendix D

Multivariate and Univariate Repeated Measures ANOVA Output

General Linear Model

Within-Subjects Factors

Measure: MEASURE_1

FACTOR1	Dependent Variable
1	LEVEL1
2	LEVEL2
3	LEVEL3
4	LEVEL4

Multivariate Tests[b]

Effect		Value	F	Hypothesis df	Error df	Sig.
FACTOR1	Pillai's Trace	.950	6.333[a]	3.000	1.000	.282
	Wilks' Lambda	.050	6.333[a]	3.000	1.000	.282
	Hotelling's Trace	19.000	6.333[a]	3.000	1.000	.282
	Roy's Largest Root	19.000	6.333[a]	3.000	1.000	.282

a. Exact statistic

b.

Design: Intercept
Within Subjects Design: FACTOR1

Mauchly's Test of Sphericity[b]

Measure: MEASURE_1

					Epsilon[a]		
Within Subjects Effe	Mauchly's W	Approx. Chi-Square	df	Sig.	Greenhouse-Geisser	Huynh-Feldt	Lower-bound
FACTOR1	.005	9.199	5	.151	.467	.750	.333

Tests the null hypothesis that the error covariance matrix of the orthonormalized transformed dependent variables is proportional to an identity matrix.

a. May be used to adjust the degrees of freedom for the averaged tests of significance. Corrected tests are display Tests of Within-Subjects Effects table.

b.

Design: Intercept
Within Subjects Design: FACTOR1

Tests of Within-Subjects Effects

Measure: MEASURE_1

Source		Type III Sum of Squares	df	Mean Square	F	Sig.
FACTOR1	Sphericity Assumed	36.688	3	12.229	15.584	.001
	Greenhouse-Geisser	36.688	1.400	26.206	15.584	.013
	Huynh-Feldt	36.688	2.250	16.307	15.584	.003
	Lower-bound	36.688	1.000	36.688	15.584	.029
Error(FACTOR1)	Sphericity Assumed	7.063	9	.785		
	Greenhouse-Geisser	7.063	4.200	1.682		
	Huynh-Feldt	7.063	6.749	1.046		
	Lower-bound	7.063	3.000	2.354		

Tests of Within-Subjects Contrasts

Measure: MEASURE_1

Source	FACTOR1	Type III Sum of Squares	df	Mean Square	F	Sig.
FACTOR1	Linear	35.113	1	35.113	31.562	.011
	Quadratic	1.563	1	1.563	1.271	.342
	Cubic	1.250E-02	1	1.250E-02	1.000	.391
Error(FACTOR1)	Linear	3.337	3	1.112		
	Quadratic	3.688	3	1.229		
	Cubic	3.750E-02	3	1.250E-02		

Tests of Between-Subjects Effects

Measure: MEASURE_1
Transformed Variable: Average

Source	Type III Sum of Squares	df	Mean Square	F	Sig.
Intercept	315.063	1	315.063	33.532	.010
Error	28.188	3	9.396		

Appendix E

Repeated Multivariate and Univariate ANOVA Syntax

```
GLM
  level1 level2 level3 level4
  /WSFACTOR = factor1 4 Polynomial
  /METHOD = SSTYPE(3)
  /CRITERIA = ALPHA(.05)
  /WSDESIGN = factor1 .
```

LAMB

U.S. Department of Education
Office of Educational Research and Improvement (OERI)
National Library of Education (NLE)
Educational Resources Information Center (ERIC)

ERIC®

REPRODUCTION RELEASE

(Specific Document)

TM034806

I. DOCUMENT IDENTIFICATION:

Title: UNDERSTANDING "WITHIN" VERSUS "BETWEEN" ANOVA DESIGNS: BENEFITS AND REQUIREMENTS OF REPEATED MEASURES

Author(s): GORDON D. LAMB

Corporate Source:

Publication Date: 2/14/03

II. REPRODUCTION RELEASE:

In order to disseminate as widely as possible timely and significant materials of interest to the educational community, documents announced in the monthly abstract journal of the ERIC system, *Resources in Education* (RIE), are usually made available to users in microfiche, reproduced paper copy, and electronic media, and sold through the ERIC Document Reproduction Service (EDRS). Credit is given to the source of each document, and, if reproduction release is granted, one of the following notices is affixed to the document.

If permission is granted to reproduce and disseminate the identified document, please CHECK ONE of the following three options and sign at the bottom of the page.

The sample sticker shown below will be affixed to all Level 1 documents	The sample sticker shown below will be affixed to all Level 2A documents	The sample sticker shown below will be affixed to all Level 2B documents
PERMISSION TO REPRODUCE AND DISSEMINATE THIS MATERIAL HAS BEEN GRANTED BY Sample TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC) 1	PERMISSION TO REPRODUCE AND DISSEMINATE THIS MATERIAL IN MICROFICHE, AND IN ELECTRONIC MEDIA FOR ERIC COLLECTION SUBSCRIBERS ONLY, HAS BEEN GRANTED BY Sample TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC) 2A	PERMISSION TO REPRODUCE AND DISSEMINATE THIS MATERIAL IN MICROFICHE ONLY HAS BEEN GRANTED BY Sample TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC) 2B
Level 1	Level 2A	Level 2B
XX		
Check here for Level 1 release, permitting reproduction and dissemination in microfiche or other ERIC archival media (e.g., electronic) *and* paper copy.	Check here for Level 2A release, permitting reproduction and dissemination in microfiche and in electronic media for ERIC archival collection subscribers only	Check here for Level 2B release, permitting reproduction and dissemination in microfiche only

Documents will be processed as indicated provided reproduction quality permits.
If permission to reproduce is granted, but no box is checked, documents will be processed at Level 1.

I hereby grant to the Educational Resources Information Center (ERIC) nonexclusive permission to reproduce and disseminate this document as indicated above. Reproduction from the ERIC microfiche or electronic media by persons other than ERIC employees and its system contractors requires permission from the copyright holder. Exception is made for non-profit reproduction by libraries and other service agencies to satisfy information needs of educators in response to discrete inquiries.

Sign here, → please

Signature: Gordon Lamb	Printed Name/Position/Title: GORDON D. LAMB	
Organization/Address: TAMU DEPT EDUC PSYC COLLEGE STATION, TX 77843-4225	Telephone: 979/845-1831	FAX:
	E-Mail Address:	Date: 3/4/03

(over)

III. DOCUMENT AVAILABILITY INFORMATION (FROM NON-ERIC SOURCE):

If permission to reproduce is not granted to ERIC, *or*, if you wish ERIC to cite the availability of the document from another source, please provide the following information regarding the availability of the document. (ERIC will not announce a document unless it is publicly available, and a dependable source can be specified. Contributors should also be aware that ERIC selection criteria are significantly more stringent for documents that cannot be made available through EDRS.)

Publisher/Distributor:
Address:
Price:

IV. REFERRAL OF ERIC TO COPYRIGHT/REPRODUCTION RIGHTS HOLDER:

If the right to grant this reproduction release is held by someone other than the addressee, please provide the appropriate name and address:

Name:
Address:

V. WHERE TO SEND THIS FORM:

Send this form to the following ERIC Clearinghouse:

University of Maryland
ERIC Clearinghouse on Assessment and Evaluation
1129 Shriver Laboratory
College Park, MD 20742
Attn: Acquisitions

However, if solicited by the ERIC Facility, or if making an unsolicited contribution to ERIC, return this form (and the document being contributed) to:

ERIC Processing and Reference Facility
1100 West Street, 2nd Floor
Laurel, Maryland 20707-3598

Telephone: 301-497-4080
Toll Free: 800-799-3742
FAX: 301-953-0263
e-mail: ericfac@inet.ed.gov
WWW: http://ericfac.piccard.csc.com