UNDERSTANDING ANALYSIS OF COVARIANCE (ANCOVA)

In general, research is conducted for the purpose of explaining the effects of the independent variable on the dependent variable, and the purpose of research design is to provide a structure for the research. In the **research design**, the researcher identifies and controls independent variables that can help to explain the observed variation in the dependent variable, which in turn reduces error variance (unexplained variation). Since the research design is structured before the research begins, this method of control is called **experimental control**.

> **Research design** – the science (and art) of planning procedures for conducting studies so as to get the most valid findings. Called "design" for short. When designing a research study, one draws up a set of instructions for gathering evidence and for interpreting it. Experiments, quasi-experiments, double-blind procedures, and correlated groups design are examples of types of research design (Vogt, 1999).

> **Control for** – to subtract statistically the effects of a variable (a control variable) to see what a relationship would be without it (Vogt, 1999).

> **Hold constant** – to "subtract" the effects of a variable from a complex relationship so as to study what the relationship would be if the variable were in fact a constant. Holding a variable constant essentially means assigning it an average value (Vogt, 1999).

In addition to controlling and explaining variation through research design, it is also possible to use **statistical control** to explain variation in the dependent variable. Statistical control, used when experimental control is difficult, if not impossible, can be achieved by measuring one or more variables in addition to the independent variables of primary interest and by controlling the variation attributed to these variables through statistical analysis rather than through research design. The analysis procedure employed in this statistical control is **analysis of covariance (ANCOVA)**.

> **Statistical control** – using statistical techniques to isolate or "subtract" variance in the dependent variable attributable to variables that are not the subject of the study (Vogt, 1999).

> **Analysis of Covariance (ANCOVA)** – an extension of ANOVA that provides a way of statistically controlling the (linear) effect of variables one does not want to examine in a study. These extraneous variables are called covariates, or control variables. (Covariates should be measured on an interval or ratio scale.) ANCOVA allows you to remove covariates from the list of possible explanations of variance in the dependent variable. ANCOVA does this by using statistical techniques (such as regression to partial out the effects of covariates) rather than direct experimental methods to control extraneous variables. ANCOVA is used in experimental studies when researchers want to remove the effects of some antecedent variable. For example, pretest scores are used as covariates in pretest-posttest experimental designs. ANCOVA is also used in non-experimental research, such as surveys or nonrandom samples, or in quasi-experiments when subjects cannot be assigned randomly to control and experimental groups. Although fairly common, the use of ANCOVA for non-experimental research is controversial (Vogt, 1999).

A **one-way analysis of covariance (ANCOVA)** evaluates whether population means on the dependent variable are the same across levels of a factor (independent variable), adjusting for differences on the covariate, or more simply stated, whether the adjusted group means differ significantly from each other. With a one-way analysis of covariance, each individual or case must have scores on three variables: a factor or independent variable, a covariate, and a dependent variable. The factor divides individuals into two or more groups or levels, while the covariate and the dependent variable differentiate individuals on quantitative dimensions. The one-way ANCOVA is used to analyze data from several types of studies; including studies with a pretest and random assignment of subjects to factor levels, studies with a pretest and assignment to factor levels based on the pretest, studies with a pretest, matching based on the pretest, and random assignment to factor levels, and studies with potential confounding (Green & Salkind, 2003).

The **analysis of covariance (ANCOVA)** is typically used to adjust or control for differences between the groups based on another, typically interval level, variable called the covariate. The ANCOVA is an extension of ANOVA that typically provides a way of statistically controlling for the effects of continuous or scale variables that you are concerned about but that are not the focal point or independent variable(s) in the study. For example, imagine that we found that boys and girls differ on math achievement. However, this could be due to the fact that boys take more math courses in high school. ANCOVA allows us to adjust the math achievement scores based on the relationship between number of math courses taken and math achievement. We can then determine if boys and girls still have different math achievement scores after making the adjustment (Leech, Barrett, & Morgan, 2005).

STATISTICAL CONTROL USING ANCOVA

Controlling and explaining variation in the dependent variable can be accomplished with either experimental control, using research design, or statistical control, using analysis of covariance. Analysis of covariance is used primarily as a procedure for the statistical control of an extraneous variable. ANCOVA, which combines regression analysis and analysis of variance (ANOVA), controls for the effects of this extraneous variable, called a **covariate**, by partitioning out the variation attributed to this additional variable. In this way, the researcher is better able to investigate the effects of the primary independent variable. The ANCOVA F test evaluates whether the population means on the dependent variable, adjusted for differences on the covariate, differ across levels of a factor. If a factor has more than two levels and the F is significant, follow-up tests should be conducted to determine where there are differences on the adjusted means between groups. For example, if a factor has three levels, three pairwise comparisons among the adjusted means can be conducted: Group 1 versus Group 2, Group 1 versus Group 3, and Group 2 versus Group 3.

Covariate – (also called a "concomitant" or "confound" variable) a variable that a researcher seeks to control for (statistically subtract the effects of) by using such techniques as multiple regression analysis (MRA) or analysis of covariance (ANCOVA) (Leech, Barrett, & Morgan, 2005; Vogt, 1999).

Extraneous variable – (sometimes called "nuisance variable.") any condition not part of a study (that is, one in which researchers have no interest) but that could have an effect on the study's dependent variable. (Note that, in this context, extraneous does not mean unimportant.) Researchers usually try to control for extraneous variables by experimental isolation, by randomization, or by statistical techniques such as analysis of covariance (Vogt, 1999).

An ANCOVA will be superior to its ANOVA counterpart in two distinct respects (i.e., increased statistical power and control), so long as a good covariate is used. The covariate role is to reduce the probability of a Type II error when tests are made of main or interaction effects, or when comparisons are made within planned or post hoc investigations. Since the probability of a Type II error is inversely related to statistical power, the ANCOVA will be more powerful than its ANOVA counterpart, presuming that other things are held constant and that a good covariate has been used within the ANCOVA. As you have seen, the F-tests associated with a standard ANOVA are computed by dividing the MS for error into the MS for the main effect. If MS_{error} can somehow be made smaller, then the calculated F will be larger, the calculated p will also be smaller, and as a result, there's a better chance that null hypotheses will be rejected. When a good covariate is used within a covariance analysis, this is exactly what happens. Data on the covariate function to explain away a portion of within-group variability, thus resulting in a smaller value for MS_{error}. Recall that mean square is often referred to as "error variance." In addition to its power function, the covariate in an ANCOVA has another function. This second function can be summed up the word *control*. In fact, some researchers will refer to the covariate of their ANCOVA studies as the *control variable*.

The logic behind the control feature of the ANCOVA is simple. To bring about the desired control, ANCOVA adjusts each group mean on the dependent variable. Although the precise formula used to make these adjustments are somewhat complicated, the rationale behind the adjustment process is easy to understand. If one of the comparison groups had an *above-average* mean on the *control variable* (as compared with the other groups in the study), then that group's mean score on the *dependent variable* will be *lowered*. In contrast, any group that has a *below-average* mean on the *covariate* will have its mean score on the *dependent variable raised.* The degree to which any group's mean score on the dependent variable is adjusted depends on how far above or below average that group stands on the control variable. By adjusting the mean scores on the dependent variable in this fashion, ANCOVA provides the best estimates of how the comparison groups would have performed if they had all possessed identical (statistically equivalent) means on the control variable(s).

The researcher must carefully select the covariate. In order for ANCOVA to be effective, the covariate must be linearly related to the dependent variable. In addition, the covariate must be unaffected by other independent variables. For example, in an experiment, it must be unaffected by the manipulation of the experimental variable. By statistically controlling for the variation attributed to the covariate, the researcher increases the precision (accuracy) of the research by reducing the error variance, which is illustrated in the Schema of Partitioning Variation in ANCOVA figure. The area inside the circle represents the total variation of the scores on the dependent variable. The proportion of the variation attributed to the treatment effect is shown, along with the variation attributed to the covariate. Note that, if the effects of the covariate were not considered, the amount of error variance would be considerably larger. However, with the ANCOVA, this variation can be controlled statistically and partitioned out of the error variance.

Schema of Partitioning Variation in ANCOVA
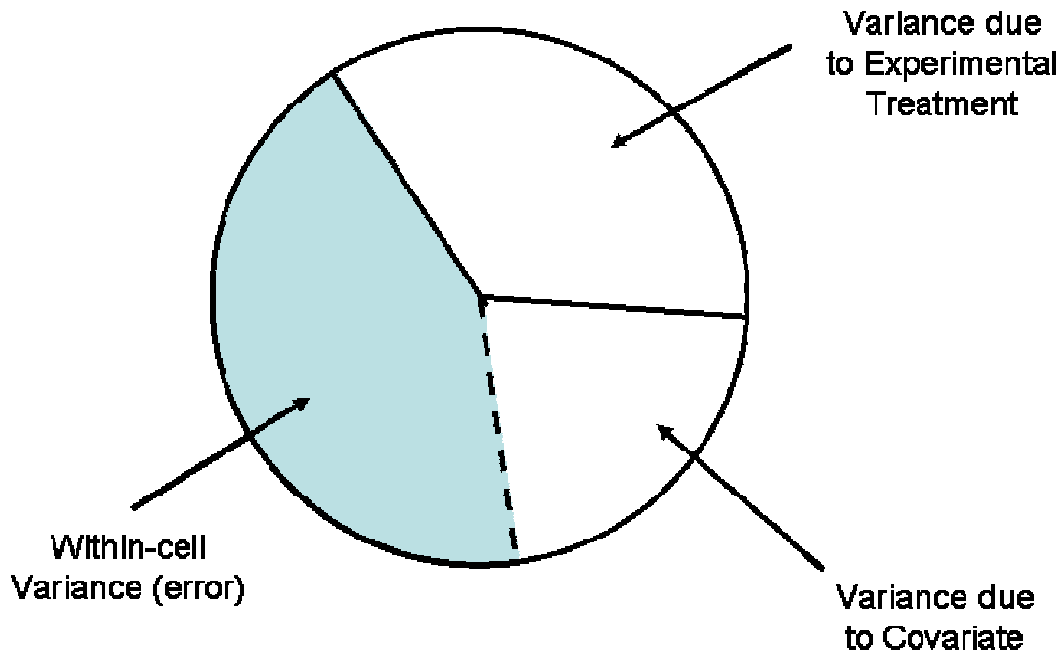
NULL AND ALTERNATIVE HYPOTHESES

The null hypothesis and the alternative hypothesis for ANCOVA are similar to those for ANOVA. Conceptually, however, these population means have been adjusted for the covariate. Thus, in reality, the null hypothesis of ANCOVA is of no difference among the *adjusted* population means.

$$H_0: \mu_1^{'} = \mu_2^{'} = ... = \mu_k^{'} \qquad\qquad H_a: \mu_i^{'} \neq \mu_k^{'} \quad \text{for some } i, k$$

ANCOVA SUMMARY TABLE

The format of the summary table for ANCOVA is similar to that for ANOVA; the difference is that the values for the sums of squares and degrees of freedom have been adjusted for the effects of the covariate. The between-groups degrees of freedom are still $K - 1$, but the within-groups degrees of freedom and the total degrees of freedom are $N - K - 1$ and $N - 1$, respectively. This reflects the loss of a degree of freedom when controlling for the covariate; this control places an additional restriction on the data. The test statistic for ANCOVA (F) is the ratio of the adjusted between-groups mean squares ($MS_B^{'}$) to the adjusted within-groups mean square ($MS_W^{'}$). The underlying distribution of this test statistic is the F distribution with $K - 1$ and $N - K - 1$ degrees of freedom.

$$F = \frac{MS_B^{'}}{MS_W^{'}}$$

Summary ANOVA

Source	Sum of Squares	Degrees of Freedom	Variance Estimate (Mean Square)	F Ratio
Covariate	SS_{Cov}	1	MS_{Cov}	$\dfrac{MS_{Cov}}{MS'_W}$
Between	SS'_B	$K-1$	$MS'_B = \dfrac{SS'_B}{K-1}$	$\dfrac{MS'_B}{MS'_W}$
Within	SS'_W	$N-K-1$	$MS'_W = \dfrac{SS'_W}{N-K-1}$	
Total	SS'_T	$N-1$		

The summary table produced in SPSS contains several additional lines. Below is a sample SPSS printout with the applicable lines marked through to reflect the above table.

Tests of Between-Subjects Effects

Dependent Variable: Total DVD assessment

Source	Type III Sum of Squares	df	Mean Square	F	Sig.
Corrected Model	1656.073[a]	4	414.018	9.340	.000
Intercept	17565.017	1	17565.017	864.018	.000
age	249.233	1	249.233	5.623	.020
promotion	1323.306	3	441.102	9.951	.000
Error	4210.927	95	44.326		
Total	126276.000	100			
Corrected Total	5867.000	99			

a. R Squared = .282 (Adjusted R Squared = .252)

ASSUMPTIONS FOR ANCOVA

In addition to the assumptions underlying the ANOVA, there are two major assumptions that underlie the use of ANCOVA; both concern the nature of the relationship between the dependent variable and the covariate.

> **The first is that the relationship is linear.** If the relationship is nonlinear, the adjustments made in the ANCOVA will be biased; the magnitude of this bias depends on the degree of departure from linearity, especially when there are substantial differences between the groups on the covariate. Thus it is important for the researcher, in preliminary analyses, to investigate the nature of the relationship between the dependent variable and the covariate (by looking at a scatter plot of the data points), in addition to conducting an ANOVA on the covariate.

The second assumption has to do with the regression lines within each of the groups. We assume the relationship to be linear. Additionally, however, the regression lines for these individual groups are assumed to be parallel; in other words, they have the same slope. This assumption is often called homogeneity of regression slopes or parallelism and is necessary in order to use the pooled within-groups regression coefficient for adjusting the sample means and is one of the most important assumptions for the ANCOVA. Failure to meet this assumption implies that there is an interaction between the covariate and the treatment. This assumption can be checked with an *F* test on the interaction of the independent variable(s) with the covariate(s). If the *F* test is significant (i.e., significant interaction) then this assumption has been violated and the covariate should not be used as is. A possible solution is converting the continuous scale of the covariate to a categorical (discrete) variable and making it a subsequent independent variable, and then use a factorial ANOVA to analyze the data.

The assumptions underlying the ANCOVA had a slight modification from those for the ANOVA, however, conceptually, they are the same.

Assumption 1: The cases represent a random sample from the population, and the scores on the dependent variable are independent of each other, known as the assumption of independence.

> The test will yield inaccurate results if the independence assumption is violated. This is a design issue that should be addressed prior to data collection. Using random sampling is the best way of ensuring that the observations are independent; however, this is not always possible. The most important thing to avoid is having known relationships among participants in the study.

Assumption 2: The dependent variable is normally distributed in the population for any specific value of the covariate and for any one level of a factor (independent variable), known as the assumption of normality.

> This assumption describes multiple conditional distributions of the dependent variable, one for every combination of values of the covariate and levels of the factor, and requires them all to be normally distributed. To the extent that population distributions are not normal and sample sizes are small, *p* values may be invalid. In addition, the power of ANCOVA tests may be reduced considerably if the population distributions are non-normal and, more specifically, thick-tailed or heavily skewed. The assumption of normality can be checked with skewness values (e.g., within $\pm$3.29 standard deviations).

Assumption 3: The variances of the dependent variable for the conditional distributions are equal, known as the assumption of homogeneity of variance.

> To the extent that this assumption is violated and the group sample sizes differ, the validity of the results of the one-way ANCOVA should be questioned. Even with equal sample sizes, the results of the standard post hoc tests should be mistrusted if the population variances differ. The assumption of homogeneity of variance can be checked with the Levene's *F* test.

INTERPRETING AN ANALYSIS OF COVARIANCE

Interpreting an analysis of covariance can present certain problems, depending on the nature of the data and, more important, the design of the experiment. The ideal application for an analysis of covariance is an experiment in which subjects are randomly assigned to treatments (or cells of a factorial design). In that situation, the *expected value* of the covariate mean for each group or cell is the same, and any differences can be attributed only to chance, assuming that the covariate was measured before the treatments were applied. In this situation, the analysis of covariance will primarily reduce the error term, but it will also, properly, remove any bias in the dependent variable means caused by change group differences on the covariate.

In a randomized experiment in which the covariate is measured *after* the treatment has been applied and has affected the covariate, interpreting the results of an analysis of covariance is difficult at best. In this situation the expected values of the group covariate means are not equal, even though the subjects were assigned randomly. It is difficult to interpret the results of the analysis because you are asking what the groups would have been like had they not differed on the covariate, when in fact the covariate differences may be an integral part of the treatment effect. This problem is particularly severe if the covariate was measured in error (i.e., if it is not perfectly reliable). In this case an alternative analysis, called the **true-score analysis of covariance**, may be appropriate if the other interpretive problems can be overcome.

When subjects are not assigned to the treatment groups at random, interpreting the analysis of covariance can be particularly troublesome. The most common example of this problem is what is called the **nonequivalent group design**. In this design, two (or more) intact groups are chosen (e.g., schools or classrooms of children), a pretest measure is obtained from subjects in both groups, the treatment is applied to one of the groups, and the two groups are then compared on some posttest measure. Since subjects are not assigned to the groups at random, we have no basis for assuming that any differences that exist on the pretest are to be attributed to chance. Similarly, we have no basis for expecting the two groups to have the same mean on the posttest in the absence of a real treatment effect.

The problem of interpreting results of designs in which subjects are not randomly assigned to the treatment groups is not easily overcome. This is one of the reasons why random assignment is even more important than random selection of subjects. It is difficult to overestimate the virtues of randomization, both for interpreting data and for making causal statements about the relationship between variables. Anyone using covariance analysis must think carefully about their data and the practical validity of the conclusions they draw.

EFFECT SIZES IN AN ANALYSIS OF COVARIANCE

Computing effect sizes for an ANCOVA is a bit more complicated compared to the One-way ANOVA. We have choices to make in the means we compare and the error term that we use. We can look at effect size with *r*-family and *d*-family measures. It is common to use the *r*-family measures when looking at an omnibus (overall) *F* test and a *d*-family measure when looking at specific contrasts (e.g., pairwise comparisons). Keep in mind that there are several options when it comes to effect size measures. What is shown here is only one example of each family.

r-FAMILY MEASURE

A common *r*-family measure of association (effect size index) is the omega square. Calculating the omega squared (ω^2) for the ANCOVA is very similar to that for the One-way ANOVA. We only need to make a few minor adjustments to the formula – to account for the adjusted values of interest…

$$\omega^2 = \frac{SS'_B - (K-1)MS'_W}{SS'_T + MS'_W}$$

Where SS'_B is the sums of square for the adjusted treatment (independent variable).

The omega square ranges in value from 0 to 1, and is interpreted as the proportion of variance of the dependent variable related to the factor (independent variable), holding constant (partialling out) the covariate. That is, the proportion of total variance in the dependent variable accounted for by the independent variable, controlling for the effect of the covariate.

d-FAMILY MEASURE

A common *d*-family measure (effect size index) is Cohen's *d*. Like with the One-way ANOVA, the effect size consists of the (numerator) mean difference between the contrasts (pairs) divided by the (denominator) averaged or pooled variability. We can estimate the average variability by taking the square root of *MS'*$_{error}$ (MS'_W) from the analysis of covariance, which would standardize the mean difference in the metric of the adjusted scores. As such, we can estimate Cohen's *d* ($\hat{d}$) with the following formula:

$$\hat{d} = \frac{\overline{X'}_i - \overline{X'}_k}{\sqrt{MS'_{error}}}$$

References

Green, S. B., & Salkind, N. J. (2003). *Using SPSS for Windows and Macintosh: Analyzing and Understanding Data* (3rd ed.). Upper Saddle River, NJ: Prentice Hall.

Hinkle, D. E., Wiersma, W., & Jurs, S. G. (2003). *Applied Statistics for the Behavioral Sciences* (5th ed.). Boston, MA: Houghton Mifflin Company.

Howell, D. C. (2002). *Statistical Methods for Psychology* (5th ed.). Pacific Grove, CA: Duxbury.

Huck, S. W. (2004). *Reading Statistics and Research* (4th ed.). Boston, MA: Allyn and Bacon.

Leech, N. L., Barrett, K. C., & Morgan, G. A. (2005). *SPSS for Intermediate Statistics: Use and Interpretation* (2nd ed.). Mahwah, NJ: Lawrence Erlbaum Associates.

Vogt, W. P. (1999). *Dictionary of Statistics and Methodology: A Nontechnical Guide for the Social Sciences* (2nd ed.). Thousand Oaks, CA: Sage Publications.