Analysis of Covariance (ANCOVA)

ANCOVA is a simple extension of ANOVA, where ANCOVA is just an ANOVA that has an added covariate. Statistical packages have a special analysis command for ANCOVA, but, just as ANOVA and simple regression are equivalent, so are ANCOVA and multiple regression. In regression model terms, instead of just using a dichotomous independent variable, X, as the predictor, we also include another predictor, X_2, in the model. X_2 can be (but does not have to be) a continuous predictor. The multiple regression model below would be equal to an ANCOVA if X_1 was binary (e.g., treatment vs. control), and X_2 was some covariate of interest.

$$Y = B_0 + B_1X_1 + B_2X_2 + e$$

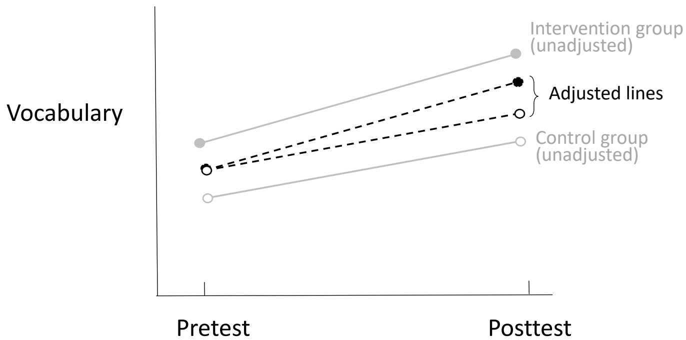
The means are not directly evident from the output of the multiple regression analysis, but the key is in the interpretation of the intercept value and the slope for the binary predictor. The intercept represents the expected value (or mean) of Y when X_1 and X_2 are both equal to zero. If X_1 is binary with values 0 and 1, then the intercept is the average of Y for the 0 group when X_2 also equals zero. The slope, B_1, has a special interpretation and represents the mean difference between the 0 and 1 group—the average increase in as X_1 increases by one unit, controlling for X_2.

In either the ANCOVA command or regression modeling approach, the difference between two groups (e.g., preschool boys and girls) “adjusts” for or controls for the other independent variable (e.g., age). Any number of covariates can be included, of course, so we are not limited to just X_2. An independent variable with any number of nominal values is acceptable, but just as with any regression model with a categorical independent variable, we must divide multiple nominal categories up into a set of dummy code variables (or other coding scheme) in order to test them in a regression analysis.

When Is ANCOVA Used?

ANCOVA is commonly used for analysis of quasi-experimental studies, when the treatment groups are not randomly assigned and the researcher wishes to statistically "equate" groups on one or more variables which may differ across groups. Pre-existing differences in treatment groups may include any number of different variables, such as greater or lesser motivation for improvement (e.g., in therapeutic trials), income or education, or differences in severity of illness or need for the treatment (e.g., low academic achievement). There are a number of potential pitfalls (see Miller & Chapman, 2001, for an extensive critique) and precautions about the interpretation of the analysis (Campbell & Kenny, 1999), perhaps most importantly, that the correct reasons for the initial differences are included and well-measured.

One particular application along these lines is the analysis of pretest-posttest designs in which groups are compared at posttest, using pretest scores as the covariate to control for pre-existing differences on the dependent variable. For example, in the hypothetical example below, a kindergarten readiness intervention is given to one group of kids and that group is compared to a control group that does not get the treatment. The graph illustrates a higher level of readiness on a vocabulary test at pretest. The ANCOVA analysis attempts to control for the pre-existing advantage of kids who wind up in the treatment group. The dotted line represents the values after the adjustment for pretest scores. Of primary interest in the analysis is the treatment effect that compares the mean posttest vocabulary scores in the two groups (intervention vs. control) after adjustment for the vocabulary pretest scores.



Should there be a significant effect for treatment, it suggests that the treatment had an effect after taking into account the pre-existing vocabulary differences.

ANCOVA is sometimes recommended with experimental research (e.g., Huitema, 2011; Maxwell & Delaney, 1990), where inclusion of covariates can have the effect of reducing the mean square error and thus increasing the power of the analysis. The extent to which this power advantage occurs certainly varies based on the relationship of the covariates to the outcome and the other predictors, the number of predictors, the reliability of the covariates (Tabachnick & Fidell, 2013), the sample size, and the overall error in prediction in the model.

Relation to Repeated Measures ANOVA

The pretest-posttest design is also commonly analyzed with repeated measures ANOVA. As we saw last term, repeated measures ANOVA (and the paired t-test) is equivalent to the test of whether the average difference score (gain score) is different from zero. The repeated measures test and the ANCOVA test are not equivalent, however. Results will often differ in the two analysis approaches, something referred to as *Lord's paradox* (Lord, 1967). The reason the two methods differ is that they represent two different ways of conceptualizing change (Newsom, 2024, Chapter 4). There is a perennial debate about whether repeated measures ANOVA (or difference scores) or ANCOVA is better for investigating pretest-posttest differences (e.g., Edwards, 1995; Rogosa, 1988), which we will not get into here but will revisit when we discuss longitudinal analysis.

Adjusted Means

ANCOVA procedures will produce adjusted means, which represents the means of each group once the covariate(s) has been controlled. In regression terms, these adjusted means come from the regression intercept, B_0. The intercept is really just an adjusted mean, representing the average value of the dependent variable taking into account the covariate and its relationship to the independent variable and the dependent variable. The equation for the intercept for a multiple regression with two predictors shows the adjustment to the mean, where the last two terms below are subtracted off.

$$B_0 = \bar{Y} - B_1\bar{X}_1 - B_2\bar{X}_2$$

Because the intercept represents the average value of Y when all predictors equal zero, however, we have to pay attention to the scaling of the predictors. If the grouping variable, X_1, has two values, 0 and 1, the adjustment leads to an intercept for the 0 group where the covariate, X_2, is also equal to zero. In many cases, the estimated mean when covariates are equal to zero is not meaningful because the zero point on the covariate is either not meaningful or outside of the observed points. Take, for example a 1-to-7 Likert scale, which has no valid value of 0, or a study of adults, for which an age of 0 is not relevant. Such circumstances are quite common if not nearly always relevant. So, if using regression analysis and there is interest in obtaining the adjusted means, it is common to rescale the covariates (not the dummy variable) by subtracting out the mean, where $x = X - \bar{X}$, what we have called deviation scores until this point. The rescaling is also commonly known in regression analysis as "centering." ANCOVA procedures in statistical packages like SPSS or R automatically center the covariates when the adjusted means are generated.

In regression analysis, the adjusted means can be computed by using the regression coefficients and inserting values for X_1, such as 0 and 1.

$$\bar{Y}_{adj} = B_0 + B_1 X_1 - B_2 \bar{X}_2$$

When dummy coding is used, the intercept from the regression model already represents the adjusted means for the 0 group. If the covariate has been centered, it is exactly equivalent to that obtained from an ANCOVA for that group. If the coding of the group variables is switched, so that 0 becomes 1 and 1 becomes 0, and the model is retested, the intercept will be the adjusted mean for the group originally coded 1.

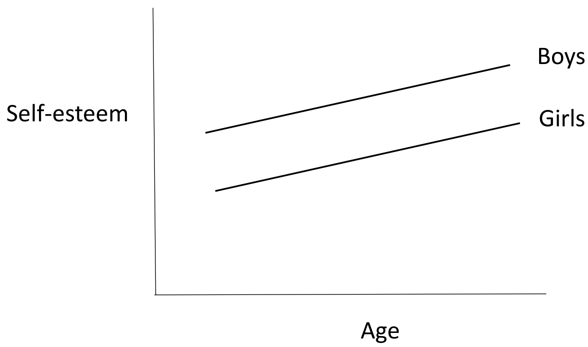
Should I use ANCOVA or Regression?
Whether you test hypothesis about group differences using an ANCOVA procedure or a regression analysis in a statistical package is arbitrary. The result will be the same with either approach.[1] When there are several categories for the independent variable or interactions are of interest, it may be more convenient to use an ANCOVA procedure, which does not require construction of a set of dummy variables. Adjusted means are also easier to obtain and report with ANCOVA procedures, because covariates are rescaled automatically.

Sum of Squares
The issue the type of sum of squares comes up in ANCOVA as it does in ANOVA. There are three types of sum of squares commonly discussed, Type I, Type II, and Type III.[2] In Type I sum of squares, each effect partials out or controls for only those effects entered before it. If effect A is entered first, it does not partial out B or A × B added after it. But B partials out A and the A × B interaction partials out A and B. In Type II sum of squares, each effect entered at the same step or before is controlled but not for later steps. Say A and B are entered first and then A × B is added. A controls for B and B controls for A. Neither controls for A × B, but A × B controls for both A and B. With Type III sum of squares, each effect adjusts for all other factors or variables in the model. For experimental studies in which the factors are randomly assigned and cell sizes are equal, we can assume the effects are already independent of one another as long as the sample size is equal across groups. It therefore, in theory, should not make any difference whether Type I, Type II, or Type III is used in the equal N experimental study.[3] Unequal samples sizes make the factors non-independent even in an experiment, however. So, some texts argue that for experimental studies with equal *n*s Type I sum of squares should or could be used. Type III sum of squares does not assume equal sample sizes and is the method commonly employed in regression analysis. Because ANOVA and regression are the same statistical model and because there does not seem to me any disadvantage to using Type III sum of squares, that is what I always suggest. Type III sum of squares is almost universally used in regression analysis, so it makes considerable sense to me to use the same method whether an ANOVA/ANCOVA procedure or a regression procedure is requested from a statistical package.

Homogeneity of Slopes Assumption
The homogeneity of slopes assumption in ANCOVA states that the relationship (slope) between the covariate (say X_2) and the dependent variable is the same for all groups being compared. If we were

[1] Interpretation of the means in the regression model approach requires some attention to the scoring of the independent variable (see above). Moreover, when interactions are involved, coding of the independent variables is also critical for demonstrating equivalence of the results obtained in the ANOVA/ANCOVA procedures and regression procedures (Aiken & West, 1991; Cohen, Cohen, West, & Aiken, 2003).
[2] You may also hear of Type IV sum of squares, which is designed for incomplete designs with no cases in some of the cells of a factorial design.
[3] Howell has a nice illustration of the three sum of squares types on p. 558: http://vulstats.ucsd.edu/pdf/howell.ch-16.pdf

interested in sex differences in self-esteem controlling for age in a sample of preschool kids, age is assumed to have the same magnitude of relationship to self-esteem for boys and girls. In the figure below (showing hypothetical results), the slope for self-esteem regressed on age is the same for boys and girls. Although boys appear to have higher self-esteem on average, both groups tend to increase in self-esteem at the same rate as they age.



Another way of stating this is that the relationship between age and self-esteem does not depend on sex. That should sound familiar from the definition of an interaction, where a result with parallel lines illustrates the absence of an interaction. So, the homogeneity of slopes assumption for the above example assumes that there is no interaction between age and gender. ANCOVA procedures will, either by default or by option, test for the interaction effect to determine whether the assumption has been met. If the interaction is significant and the assumption has not been met, the interaction could be incorporated into the analysis. We will cover this kind of regression model in coming weeks.

References

Aiken, L. S., & West, S. G. (1991). *Multiple regression: Testing and interpreting interactions*. Newbury Park, CA: Sage.
Cohen, J., Cohen, P., West, S.G., & Aiken, L.S. (2003). A*pplied multiple regression/correlation analysis in the behavioral sciences (Third Edition*). Mahwah, NJ: Erlbaum.
Campbell, D. T., & Kenny, D. A. (1999). *A primer on regression artifacts.* New York: Guilford Publications.
Edwards, J. R. (1995). Alternatives to difference scores as dependent variables in the study of congruence in organizational research. *Organizational Behavior and Human Decision Processes, 64*, 307-324.
Huitema, B. (2011). *The analysis of covariance and alternatives: Statistical methods for experiments, quasi-experiments, and single-case studies* (Vol. 608). New York: John Wiley & Sons.
Lord, E. M. (1967). A paradox in the interpretation of group comparisons. *Psychological Bulletin, 68*, 304–305
Maxwell, S. E., & Delaney, H. D. (1990). Designing experiments and analyzing data: A model comparison approach. Belmont, CA: Wadsworth.
Miller, G. A., & Chapman, J. P. (2001). *Misunderstanding analysis of covariance. Journal of abnormal psychology, 110*(1), 40-48.
Newsom, J.T. (2024). *Longitudinal Structural Equation Modeling: A Comprehensive Introduction, second edition*. Routledge.
Rogosa, D. (1988). Myths about longitudinal research. In K. W. Schaie,R. T. Campbell, W. Meredith, & S. C. Rawlings (Eds.), *Methodological issues in aging research* (pp.171-210). New York: Springer.
Tabachnick, B. G., and Fidell, L. S. (2013). *Using multivariate statistics, 6th ed.* Boston: Pearson.