ANCOVA

Overview

Analysis of covariance is used to test the main and interaction effects of categorical variables on a continuous dependent variable, controlling for the effects of selected other continuous variables which covary with the dependent.The control variable is called the "covariate." There may be more than one covariate. One may also perform planned comparison or post hoc comparisons to see which values of a factor contribute most to the explanation of the dependent.

ANCOVA is used for three purposes:

- In quasi-experimental (observational) designs, to remove the effects of variables which modify the relationship of the categorical independents to the interval dependent.
- In experimental designs, to control for factors which cannot be randomized but which can be measured on an interval scale. Since randomization in principle can control for all unmeasured variables, in principle the addition of covariates to a model is rarely or never needed in experimental research. If a covariate is added and it is uncorrelated with the treatment (independent) variable, it is difficult to interpret as in principle it is controlling for something already controlled for by randomization. If the covariate is correlated with the treatment/independent, then it will lead to underestimate of the effect size of the treatment/independent.
- In regression models, to fit regressions where there are both categorical and interval independents. (This third purpose has become displaced by logistic regression and other methods. On ANCOVA regression models, see Wildt and Ahtola, 1978: 52-54).

All three purposes have the goal of reducing the error term in the model. Like other control procedures, ANCOVA can be seen as a form of "what if" analysis, asking what would happen if all cases scored equally on the covariates, so that the effect of the factors over and beyond the covariates can be isolated. Note: ANCOVA can be used in all ANOVA designs and much of the discussion in the ANOVA section applies to ANCOVA as well, including the discussion of assumptions.

Key Concepts and Terms

- **Covariate:** An interval-level independent. If there are no covariates, ANOVA must be used instead of ANCOVA, and if there are covariates, ANCOVA is used instead of ANOVA. Covariates are commonly used as control variables. For instance, use of a baseline pre-test score can be used as a covariate to control for initial group differences on math ability or whatever is being assessed in the ANCOVA study. That is, in ANCOVA we look at the effects of the categorical independents on an interval dependent, after effects of interval covariates are controlled. (This is similar to regression, where the beta weights of categorical independents represented as dummy variables entered after interval independents reflect the control effect of these independents).

- **F-test:** The F-test of significance is used to test each main and interaction effect, for the case of a single interval dependent and multiple (>2) groups formed by a categorical independent. F is between-groups mean square divided by within-groups mean square. If the computed F score is greater than 1, then there is more variation between groups than within groups, from which we infer that the grouping variable does make a difference. If the F score is enough above 1, it will be found to be significant in a table of F values, using df=k-1 and df=N-k-1, where N is sample size and k is the number of groups formed by the factor(s)..
- **Adjusted means** are usually part of ANCOVA output and are examined if the F-test demonstrates significant relationships exist. Comparison of the original and adjusted group means can provide insight into the role of the covariates. For k groups formed by categories of the categorical independents and measured on the dependent variable, the adjustment shows how these k means were altered to control for the covariates. Typically, this adjustment is one of linear regression of the type: $Y_{adj.mean} = Y_{mean} - b(X_{ith.mean}-X_{mean})$, where Y is the interval dependent, X is the covariate, i is one of the k groups, and b is the regression coefficient. There is no constant when Y is standardized. For multiple covariates, of course, there are additional similar X terms in the equation.
- **Sphericity test** is a special case of a homogeneity of variance test for repeated measures ANCOVA when using a repeated measures factor with three or more levels. (Two levels always display sphericity). ANCOVA assumes sphericity, which is when the variance of the difference between the estimated means for any pair of different groups is the same as for any other pair. SPSS automatically does the sphericity test for ANCOVA repeated measures designs with repeated measures factors with three or more levels. In SAS, the PRINTE option in PROC GLM tests sphericity. Sphericity is tested because ANCOVA repeated measures designs are sensitive to this violation of assumptions -- that is, to a spherical within-subjects covariance matrix . If the significance of the sphericity test is less than 0.05 then the researcher accepts the null hypothesis that the data are not spherical, thereby violating the sphericity assumption, and must correct for sphericity or must use multivariate ANCOVA tests (Wilks Lambda, Pillai's Trace, Hotelling-Lawley Trace, Roy's Greatest Root).

  Epsilon. If the researcher wishes to correct the univariate F test, this is done by using Huynh-Feldt or Greenhouse-Geisser Epsilon. Recall that F is the ratio of between-groups to within-groups mean square variance. The degrees of freedom for between-groups is (k-1), where k = the number of groups. The degrees of freedom for within-groups is k(n-1), where n is the number of cases in each group. To correct F given a finding of lack of sphericity, the researcher multiplies the between-groups degrees of freedom by the value of epsilon. SPSS supplies Huynh-Feldt epsilon, and the more conservative Greenhouse-Geisser epsilon [which in turn is an extension of Box's epsilon, no longer widely used]). For more severe departures from sphericity (epsilon < .75), the more conservative Greenhouse-Geisser epsilon is used, while Huynh-Feldt epsilong is used for less severe violations of the sphericity assumption. The researcher rounds down to the nearest whole number and looks up the corrected F value in a table using the corrected degrees of freedom.
- **Hotelling's T-Square** is a multivariate significance test of mean differences, for the case multiple interval dependents and two groups formed by a categorical independent. As of version 8, SPSS does not compute this directly, but does compute the related statistic, **Hotelling's Trace** (a.k.a. Lawley-Hotelling or Hotelling-Lawley Trace). To convert from the Trace coefficient to the T-Square coefficient, multiply the Trace coefficient by (N-L), where N is the sample size across all groups and L is the number of groups. The T-Square result will still have the same F value, degrees of freedom, and significance level as the Trace coefficient.

- **Wilks's lambda** is a multivariate significance test of mean differences, for the case of multiple interval dependents and multiple (>2) groups formed by the independent(s). The t-test, Hotelling's T, and the F test are special cases of Wilks's lambda.
- **t-test:** A test of significance of the difference in the means of a single interval dependent, for the case of two groups formed by a categorical independent. n effect.
- **Eta-square** is a measure of effect size, equal to the ratio of the between-groups sum of squares to the sum of squares summed for all main, interaction, and error effects (but not covariate effects). Eta-square is interpreted as the percent of variance in the dependent variable explained by the factor(s). When there are curvilinear relations of the factor to the dependent, eta-square will be higher than the corresponding coefficient of multiple correlation (R^2).

Assumptions

- **At least one categorical and at least one interval independent**. The independent variable(s) may be categorical, except at least one must be a covariate (interval level). Likewise, at least one independent must be categorical.
- **Interval dependent**. The dependent variable is continuous and interval level.
- **Limited number of covariates**. The more the covariates, the greater the likelihood that an additional covariate will have little residual correlation with the dependent after other covariates are controlled. The marginal gain in explanatory power is offset by loss of statistical power (a degree of freedom is lost for each added covariate).
- **Low measurement error of the covariate**. The covariate variables are continuous and interval level, and are assumed to be measured without error. Imperfect measurement reduces the statistical power of significance tests for ANCOVA and for experimental data, there is a conservative bias (increased likelihood of Type II errors: thinking there is no relationship when in fact there is a relationship) . As a rule of thumb, covariates should have a reliability coefficient of .80 or higher.
- **Covariates are linearly related or in a known relationship to the dependent**. The form of the relationship between the covariate and the dependent must be known and most computer programs assume this relationship is linear, adjusting the dependent mean based on linear regression. Scatterplots of the covariate and the dependent for each of the k groups formed by the independents is one way to assess violations of this assumption. Covariates may be transformed (ex., log transform) to establish a linear relationship.
- **Homogeneity of covariate regression coefficients**. The covariate coefficients (the slopes of the regression lines) are the same for each group formed by the categorical variables and measured on the dependent. The more this assumption is violated, the more conservative ANCOVA becomes (the more likely it is to make Type I errors - accepting a false null hypothesis). There is a statistical test of the assumption of homogeneity of regression coefficients (see Wildt and Ahtola, 1978: 27). Violation of the homogeneity of regressions assumption indicates an interaction effect between the covariate(s) and the factor(s).

  Homogeneity of regression in SPSS can be tested in the SPSS MANOVA module, in the DESIGN statement in syntax (not in the menu as of SPSS 11.0). An effect can be modeled which is the pooled covariates by each of the factors and each interaction of factors. If this pooled covariate effect is significant, then the homogeneity of regressions assumption is violated. For the case of two factors: DESIGN {list of covariates, factors, and interactions},POOL(list of covariates) BY factor1+ POOL(list of covariates) BY factor2 + POOL(list of covariates) BY factor1 by factor2.

- **No covariate outliers**. ANCOVA is highly sensitive to outliers in the covariates.
- **No high multicollinearity of the covariates**. ANCOVA is sensitive to multicollinearity among the covariates and also loses statistical power as redundant covariates are added to the model. Some researchers recomment dropping from analysis any added covariates whose squared correlation with prior covariates is .50 or higher.
- **Additivity**. The values of the dependent are an additive combination of its overall mean, the effect of the categorical independents, the covariate effect, and an error term. ANCOVA is robust against violations of additivity but in severe violations the researcher may transform the data, as by using a logarithmic transformation to change a multiplicative model into an additive model. Note, however, that ANCOVA automatically handles interaction effects and thus is not an additive procedure in the sense of regression models without interaction terms.
- **Independence of the error term**. The error term is independent of the covariates and the categorical independents. Randomization in experimental designs assures this assumption will be met.
- **Independent variables orthogonal to covariates**. In traditional ANCOVA, the independents area assumed to be orthogonal to the factors. If the covariate is influenced by the categorical independents, then the control adjustment ANCOVA makes on the dependent variable prior to assessing the effects of the categorical independents will be biased since some indirect effects of the independents will be removed from the dependent. However, in GLM ANCOVA, the values of the factors are adjusted for interactions with the covariates.
- **Homogeneity of variances**. It is assumed there is homogeneity of variances of the dependent and of the covariates in the cells formed by the factors *Heteroscedasticity* is lack of homogeneity of variances, in violation of this assumption. When this assumption is violated, the offending covariate may be dropped or the researcher may adopt a more stringent alpha significance level (ex., .01 instead of .05).
- **Normal distribution within groups**.The dependent variable should be normally distributed within groups formed by the factors. Deviations from this assumption are unimportant by the central limit theorem when group size is large (as a rule of thumb, > 20; more if group sizes are unequal or there are outliers).
- **Compound sphericity**. The groups display sphericity (the variance of the difference between the estimated means for any two different groups is the same) A more restrictive assumption, called *compound symmetry*, is that the correlations between any two different groups are the same value. If compound symmetry exists, sphericity exists. Tests or adjustments for lack of sphericity are usually actually based on possible lack of compound symmetry.
- See also the "Assumptions" section for ANOVA.
- See also the "Testing Assumptions" section of the Prophet Statguide for ANCOVA.

Frequently Asked Questions

- **Do you use the same designs (between groups, repeated measures, etc.) with ANCOVA as you do with ANOVA?**
- **Where is ANCOVA in SPSS?**
- **How is GLM ANCOVA different from traditional ANCOVA?**

- **Where is ANCOVA in SPSS?**

  It is part of the ANOVA procedure. It is invoked when interval-level independents (covariates) are listed in the VARIABLES subcommand after the WITH keyword. The maximum number of covariates SPSS will process is 10.

- **Do you use the same designs (between groups, repeated measures, etc.) with ANCOVA as you do with ANOVA?**

  By and large, yes. Be warned, however, that for repeated measures designs with more than two levels of the repeated measure factor, if the covariates are also measured repeatedly, only the univariate tests output is appropriate (in SPSS, the ones labeled "AVERAGED tests of significance". SPSS will also print the multivariate tests, but they are not appropriate because the multivariate tests partial each of the covariates from the entire set of dependent variables. The appropriate univariate approach partials variance on the basis of dependent-variable/covariate matched pairs.

- **How is GLM ANCOVA different from traditional ANCOVA?**

  The traditional method assumes that the covariates are uncorrelated with the factors. The GLM (general linear model) approach adjusts for interactions of the covariates with the factors.

- **What are paired comparisons (planned or post hoc) in ANCOVA?**

  After the omnibus F test establishes an overall relationship, the researcher can test differences between pairs of group means (assuming the independent has more than two levels) to determine which groups are most involved with significant effects. Ideally these comparisons are based on a priori theory and there are just a few of them. But if the researcher wants to investigate all possible paired comparisons on a post hoc basis, some will be found significant just be chance, so there are various adjustments (Bonferroni, Tukey, Scheffe) which make it harder to find significance.

- **Can ANCOVA be modeled using regression?**

  Yes, if dummy variables are used for the categorical independents. When creating dummy variables, one must use one less category than there are values of each independent. For full ANCOVA one would also add the interaction crossproduct terms for each pair of independents included in the equation, including the dummies. Then one computes multiple regression. The resulting F tests will be the same as in classical ANCOVA.

- **How does blocking with ANOVA compare to ANCOVA?**

  In blocking under ANOVA, what would have been a continuous covariate in ANCOVA is classified (ex., high, medium, low) and used as an additional factor in an ANOVA. The main effect of this factor is similar to the effect of the covariate in ANCOVA. If there is an interaction effect involving this factor, this shows the homogeneity of regressions assumption would have been violated in ANCOVA. This has the advantage compared to ANCOVA that one need not assume the relationship between the covariate and the dependent variable is linear. However, classification involves loss of information and attenuation of correlation. If the covariate is related to the dependent in a linear manner, ANCOVA will be more powerful than ANOVA with blocking and is preferred. Also, blocking after data are collected may involve unequal group sample sizes, which also makes ANOVA less robust.

Bibliography

- Wildt, Albert R. and Olli T. Ahtola (1978). *Analysis of covariance*. Quantitative Applications in the Social Sciences series #12. Thousand Oaks, CA: Sage Publications.