APPLIED STATISTICS INSTRUCTION SHEET

ANCOVA MODELS IN SPSS AND PSPP

Instructions

The Analysis of Covariance (ANCOVA) uses a combination of numeric and categorical variables to predict a numeric dependent variable. Interactions can be included, though their appropriate use is not discussed here. There are at least three different ways to run the procedure in SPSS—Regression, Univariate, Generalized Linear Model. PSPP can use only the Regresion procedure.

Regression (Analyze> Regression>Linear)

- The dependent variable goes into the Dependent box.
- The independent and control variables go into the Independent(s) box.
- Categorical variables must be entered into the Independent(s) box as a block of dummy variables, which are created outside the procedure using one of the procedures in the Transform menu. Each category should be transformed into a 0/1 binary. When the block is entered, one variable should be dropped as the reference category.
- For binary variables that are in pairs, such as Male or Female, only one item should be used in a 0/1 format.

Univariate (Analyze>General Linear Model>Univariate)

- The dependent variable goes into the Dependent box.
- Categorical variables go into the Fixed Factor(s). They do not need entering as binary variables, as is the case with the Regression procedure. The procedure automatically withholds the last category as the reference category. If you want a different reference category, you need to create a block of dummy variables and enter them as binary with the reference category withheld.
- Numeric variables go into the Covariate(s) box.
- The Model should be set. Use the Build Custom Terms option. Set the Type in the middle of the dialog box to Interaction or Main Effect, depending on what you want to do. Slide the variables you want from the left box to the right box. If you want to construct an interaction term, slide both variables together.
- To get the output you want, you probably need to select some item under the Options button. I am most interested in Descriptive Statistics, Estimates of Effect Size, Observed Power, Parameter Estimates, and Heteroskedasticity Tests.

Generalized Linear Models (Analyze>Generalized Linear Models>Univariate)

- Type of Model tab: Under Scale Response, select Linear. (This is the default.)
- Response tab: Select the dependent variable.
- Predictors tab:
    - Categorical variables go into the Fixed Factor(s). They do not need entering as binary variables, as is the case with the Regression procedure. Right click on the variable to indicate whether you want to use the first category the last category as the reference category. If you want a different reference category, you need to

create a block of dummy variables and enter them as binary with the reference category withheld.
    - Numeric variables go into the Covariate(s) box.
- Model tab: Set the Type in the middle of the dialog box to Interaction or Main Effect, depending on what you want to do. Slide the variables you want from the left box to the right box. If you want to construct an interaction term, slide both variables together.
- Statistics tab: I usually select Descriptive Statistics, Model Information, Goodness of Fit Statistics, Model Summary Statistics, and Parameter Estimates.

Key Statistics

The Coefficients tables contains parameters of the regression line and is where the results related to most hypothesis testing are found. The key statistics are as follows:

- The column B contains the parameter estimates for the y-intercept (Constant) and individual variables that are used to build the regression equation. The sign of B determines whether the direction is positive or negative.
- Depending on procedure, the standardized Betas or partial eta-squared column provides an effect size for each individual coefficient.
    - Regression: The interpretation of the effect size using the standardized Beta is based on the following ranges: weak relationships have a standardized Beta between .1 and .29, medium or moderate relationship are between .30 and .49, and strong relationships equal or exceed .50.
    - Univariate and Generalized Linear Model: The interpretation of the effect size using the partial eta-squared is based on the following ranges, which are the squares of their equivalents for the standardized Beta: weak relationships have a partial eta-squared between .01 and .08, medium or moderate relationship are between .09 and .24, and strong relationships equal or exceed .25.
- The Sig. column gives the p-value used in determining statistical significance. The null hypothesis is that B=0.

The R^2 and Adjusted R^2 are available only in the Regression and Univariate procedures. In Regression, the Model Summary and ANOVA tables provide statistics to interpret the overall fit of the model. In the Univariate procedure, use the Tests of Between-Subjects Effects, which contains these statistics in a footnote.

- The R^2 is the proportion of the variance of the model explained by the regression. When multiple by 100, it can be described as the percent of variance explained. The interpretation of the effect size using the R^2 is based on the following ranges: weak relationships have an R^2 between .01 and .08, medium or moderate relationship are between .09 and .24, and strong relationships equal or exceed .25.
- The statistical significance of the goodness of fit is the significance of the F-statistics in the ANOVA table in Regression and the statistical significance of the Corrected Model in the Test of Between-Subjects Effects in the Univariate output.

Written Interpretation

Written comments should highlight the direction, effect size, and statistical significance of the individual coefficients and the overall fit and statistical significance of model as a whole.

Statistics can be included in the text parenthetically when accompanied by a statement that conveys the key result in plain English (B=-.53, p<.01). If there is no substantive or significant effect, this fact should be stated in the text.