APPLIED STATISTICS INSTRUCTION SHEET

OLS REGRESSION IN SPSS AND PSPP

Instructions

Ordinary least-squares (OLS) regression is a form of the general linear model that describes the relationship between an independent and dependent variable. Multiple regression means there are two or more independent or control variables in the model. Software packages do not distinguish whether a variable is independent or control—the designation is driven by whether the variable is related to the hypothesis to be tested. The path in both SPSS and PSPP is Analyze> Regression>Linear.

Instructions for both SPSS and PSPP

- Toggle the dependent variable into the Dependent box on the upper right-hand side of the dialog box.
- Toggle the independent variable into the Independent(s) box on the right-hand side. One can have more than one independent variable in the model, making the model a multiple regression

Regression Diagnostics in SPSS

- Under the Save button are a variety of options for calculating predicted values, residuals, and several distance and leverage statistics, including but not limited to Cook's D, leverage values, and DfFits. Calculated values for selected statistics appear at the end for the data set (last columns in Data View, bottom rows in Variable View), from which they can be plotted.
- The Plots button allows for the creating of residual plots without saving the calculated values.

Regression Diagnostics in PSPP

- The Save button allows for the creation of residuals and predicted values, which appear at the end of the data set (last columns in Data View, bottom rows in Variable View), from which they can be plotted.

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.
- The standardized Betas column provides an effect size for each individual coefficient. The interpretation of the effect size is based on the following ranges: weak relationships have a standardized Beta between .10 and .29, medium or moderate relationship are between .30 and .49, and strong relationships equal or exceed .50.
- The Sig. column gives the p-value used in determining statistical significance. The null hypothesis is that B=0.

The Model Summary and ANOVA tables provide statistics to interpret the overall fit of the model.

- 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 R, which is the square root of R^2, provides the measure of the effect size for the overall model. The interpretation of the effect size is based on the following ranges: weak relationships have a standardized Beta between .10 and .29, medium or moderate relationship are between .30 and .49, and strong relationships equal or exceed .50.
- In the ANOVA table, the Sig. value is related to the F-distribution, from which the statistical significance of the fit of the overall model is determined. The null hypothesis is that F=0.

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.