A.4 SPSS

Chapters 1–3: Introduction, Two-Way Contingency Tables

The DESCRIPTIVE STATISTICS option on the ANALYZE menu has a suboption called CROSSTABS, which provides several methods for contingency tables. After identifying the row and column variables in CROSSTABS, clicking on STATISTICS provides a wide variety of options, including the chi-squared test and measures of association. The output lists the Pearson statistic, its degrees of freedom, and its P-value (labeled Asymp. Sig.). If any expected frequencies in a 2×2 table are less than 5, Fisher's exact test results. It can also be requested by clicking on Exact in the CROSSTABS dialog box and selecting the exact test. SPSS also has an advanced module for small-sample inference (called *SPSS Exact Tests*) that provides exact P-values for various tests in CROSSTABS and NPAR TESTS procedures. For instance, the *SPSS Exact Tests* module provides exact tests of independence for $I \times J$ contingency tables with nominal or ordinal classifications. See

www-01.ibm.com/software/analytics/spss/products/statistics/exact-tests/features.html?S_CMP=rnav

In CROSSTABS, clicking on CELLS provides options for displaying observed and expected frequencies, as well as the standardized residuals, labeled as "Adjusted standardized". Clicking on STATISTICS in CROSSTABS provides options of a wide variety of statistics other than chi-squared, including gamma and Kendall's tau-b. The output shows the measures and their standard errors (labeled Asymp. Std. Error), which you can use to construct confidence intervals. It also provides a test statistic for testing that the true measure equals zero, which is the ratio of the estimate to its standard error. This test uses a simpler standard error that only applies under independence and is inappropriate for confidence intervals. One option in the list of statistics, labeled Risk, provides as output the odds ratio and its confidence interval.

Suppose you enter the data as cell counts for the various combinations of the two variables, rather than as responses on the two variables for individual subjects; for instance, perhaps you call COUNT the variable that contains these counts. Then, select the WEIGHT CASES option on the DATA menu in the Data Editor window, instruct SPSS to weight cases by COUNT.

Chapter 4: Generalized Linear Models

To fit generalized linear models, on the ANALYZE menu select the GENERALIZED LINEAR MODELS option and the GENERALIZED LINEAR MODELS suboption. Select the Dependent Variable and then the Distribution and Link Function. Click on the Predictors tab at the top of the dialog box and then enter quantitative variables as Covariates and categorical variables as Factors. Click on the Model tab at the top of the dialog box and enter these variables as main effects, and construct any interactions that you want in the model. Click on OK to run the model.

Chapters 5–7: Logistic Regression and Binary Response Methods

To fit logistic regression models, on the ANALYZE menu select the REGRESSION option and the BINARY LOGISTIC suboption. In the LOGISTIC REGRESSION dialog box, identify the binary response (dependent) variable and the explanatory predictors (covariates). Highlight variables in the source list and click on a*b to create an interaction term. Identify the explanatory variables that are categorical and for which you want indicator variables by clicking on Categorical and declaring such a covariate to be a Categorical Covariate in the LOGISTIC REGRESSION: DEFINE CATEGORICAL VARIABLES dialog box. Highlight the categorical covariate and under Change Contrast you will see several options for setting up indicator variables. The Simple contrast constructs them as in this text, in which the final category is the baseline.

In the LOGISTIC REGRESSION dialog box, click on Method for stepwise model selection procedures, such as backward elimination. Click on Save to save predicted probabilities, measures of influence such as leverage values and DFBETAS, and standardized residuals. Click on Options to open a dialog box that contains an option to construct confidence intervals for exponentiated parameters.

Another way to fit logistic regression models is with the GENERALIZED LINEAR MODELS option and suboption on the ANALYZE menu. You pick the binomial distribution and logit link function. It is also possible there to enter the data as the number of successes out of a certain number of trials, which is useful when the data are in contingency table form. One can also fit such models using the LOGLINEAR option with the LOGIT suboption in the ANALYZE menu. One identifies the dependent variable, selects categorical predictors as factors, and selects quantitative predictors as cell covariates. The default fit is the saturated model for the factors, without including any covariates. To change this, click on Model and select a Custom model, entering the predictors and relevant interactions as terms in a customized (unsaturated) model. Clicking on Options, one can also display standardized residuals (called adjusted residuals) for model fits. This approach is well suited for logit models with categorical predictors, since standard output includes observed and expected frequencies. When the data file contains the data as cell counts, such as binomial numbers of successes and failures, one weights each cell by the cell count using the WEIGHT CASES option in the DATA menu.

Chapter 8: Multinomial Response Models

SPSS can fit logistic models for multinomial response variables. On the ANALYZE menu, choose the REGRESSION option and then the ORDINAL suboption for a cumulative logit model. Select the MULTINOMIAL LOGISTIC suboption for a baseline-category logit model. In the latter, click on Statistics and check Likelihood-ratio tests under Parameters to obtain results of likelihood-ratio tests for the effects of the predictors.

SPSS Regression is an add-on module for performing logistic regression, ordinal regression, multinomial models, and mixed models. See

www-01.ibm.com/software/analytics/spss/products/statistics/regression/features.html?S_CMP=rnav

Chapters 9–10: Loglinear Models

For loglinear models, one uses the LOGLINEAR option with GENERAL suboption in the ANALYZE menu. One enters the factors for the model. The default is the saturated model, so click on Model and select a Custom model. Enter the factors as terms in a customized (unsaturated) model and then select additional interaction effects. Click on Options to show options for displaying observed and expected frequencies and adjusted residuals. When the data file contains the data as cell counts for the various combinations of factors rather than as responses listed for individual subjects, weight each cell by the cell count using the WEIGHT CASES option in the DATA menu.

SPSS Categories is an add-on module that provides optimal scaling procedures such as categorical principal components analysis and multidimensional scaling, and some reduction-dimension techniques such as correspondence analysis, biplots, and canonical correlation analysis. See

> www-01.ibm.com/software/analytics/spss/products/statistics/categories/features.html?S_CMP=rnav

Chapter 11: Models for Matched Pairs

The models discussed in this chapter are almost all generalized linear models and can be fitted as described above for Chapter 4. The LOGLINEAR option just mentioned for Chapters 9–10 can also be used.

Chapters 12–14: Clustered Categorical Responses

For GEE methods, on the ANALYZE menu choose the GENERALIZED LINEAR MODELS option and the GENERALIZED ESTIMATING EQUATIONS suboption. You can then select structure for the working correlation matrix and identify the between-subject and within-subject variables.

For random effects models, on the ANALYZE menu choose the MIXED MODELS option and the GENERALIZED LINEAR suboption.

Generalized linear mixed models can be fitted with the *SPSS Advanced Statistics* add-on module. See

> www-01.ibm.com/software/analytics/spss/products/statistics/advanced-statistics/features.html?S_CMP=

Chapter 15: Non-Model-Based Classification and Clustering

Discriminant analysis methods are available in the base version of SPSS. From the ANALYZE menu choose CLASSIFY and then DISCRIMINANT. For details, see

> www.cs.uu.nl/docs/vakken/arm/SPSS/spss6.pdf

SPSS Decision Trees is an add-on module for constructing classification trees. See

> www.uky.edu/ComputingCenter/SSTARS/SPSS/19%20Manuals/IBM%20SPSS%20Decision%20Trees%2019.pdf

and

> www-01.ibm.com/software/analytics/spss/products/statistics/decision-trees/features.html?S_CMP=rnav

Cluster analysis methods are available in the base version of SPSS. From the AN-ALYZE menu choose CLASSIFY and then HIERARCHICAL CLUSTER. For details, see

`www.cs.uu.nl/docs/vakken/arm/SPSS/spss8.pdf`

and `www.norusis.com/pdf/SPC_v13.pdf`

Chapter 16: Large- and Small-Sample Theory for Multinomial Models

SPSS Exact Tests is an add-on module for small-sample analyses with contingency tables. See

`www-01.ibm.com/software/analytics/spss/products/statistics/exact-tests/features.html?S_CMP=rnav`