Complex Samples: Model Testing

This document describes the methods used for conducting linear hypothesis tests based on the estimated parameters in CS models.

Required input is a set of the linear hypothesis, parameter estimates and their covariance matrix estimated for the complex sample design. Some methods require an estimate of the parameter covariance matrix under the simple random sampling assumption as well. Also needed is the number of degrees of freedom for the complex sample design; typically this will be the difference between the number of primary sampling units and the number of strata in the first stage of sampling.

Given consistent estimates of the above constructs, no additional restrictions are imposed on the complex sample design.

Notations

p Number of regression parameters in the model.

r The number of linear hypothesis considered.

$\mathbf{L}$ Generalized linear hypothesis matrix with r rows and p columns.

$\mathbf{K}$ Hypothesis value vector with r elements.

$\mathbf{B}$ Vector of p unknown population parameters.

$\hat{\mathbf{B}}$ Vector of p estimated population parameters (solution).

$\hat{\mathbf{V}}(\hat{\mathbf{B}})$ Estimated covariance matrix for $\hat{\mathbf{B}}$ given the complex sample design.

v The number of sampling design degrees of freedom.

Hypothesis Testing

Given matrix $\mathbf{L}$ with r rows and p columns, and vector $\mathbf{K}$ with r elements, the following test of generalized linear hypothesis is performed:

$$H_0 : \mathbf{LB} = \mathbf{K}.$$

It is assumed that $\mathbf{LB}$ is estimable.

Wald chi-square test

Wald $\mathbf{X}^2$ statistic proposed by Koch et al. (1975) is defined by

$$\mathbf{X}^2 = (\mathbf{L}\hat{\mathbf{B}} - \mathbf{K})'(\mathbf{L}\hat{\mathbf{V}}(\hat{\mathbf{B}})\mathbf{L}')^-(\mathbf{L}\hat{\mathbf{B}} - \mathbf{K}).$$

Asymptotic distribution of the $\mathbf{X}^2$ test statistic is chi-square with r_I degrees of freedom, where $r_I = rank(\mathbf{L}\hat{\mathbf{V}}(\hat{\mathbf{B}})\mathbf{L}')$. If $r_I < r$, $(\mathbf{L}\hat{\mathbf{V}}(\hat{\mathbf{B}})\mathbf{L}')^-$ is a generalized inverse such that Wald tests are effective for restricted set of hypothesis $\mathbf{L}_I\mathbf{B} = \mathbf{K}_I$ containing a particular subset I of independent rows from H_0.

Wald F test

Wald F statistic suggested by Fellegi (1980) is computed by the formula

$$F = \frac{v - r_I + 1}{r_I v}\mathbf{X}^2.$$

This statistic is asymptotically approximated by the F-distribution $F(r_I, v - r_I + 1)$, where v is the number of the sampling design degrees of freedom. The statistic is undefined if $v < r_I$. See Korn and Graubard (1990) for the properties of this statistic.

Adjusted Wald chi-square test

Wald $\mathbf{X}^2_{srs}$ statistic under the simple random sampling assumption is given by the following expression:

$$\mathbf{X}^2_{srs} = (\mathbf{L}\hat{\mathbf{B}} - \mathbf{K})'(\mathbf{L}\hat{\mathbf{V}}_{srs}(\hat{\mathbf{B}})\mathbf{L}')^-(\mathbf{L}\hat{\mathbf{B}} - \mathbf{K}).$$

$\hat{\mathbf{V}}_{srs}(\hat{\mathbf{B}})$ is an asymptotic covariance matrix estimated under the simple random sampling assumption. If $rank(\mathbf{L}\hat{\mathbf{V}}_{srs}(\hat{\mathbf{B}})\mathbf{L}') < r$, adjusted Wald tests are effective for restricted set of hypothesis $\mathbf{L}_I\mathbf{B} = \mathbf{K}_I$ containing a particular subset I of independent rows from H_0.

Since the asymptotic distribution of $\mathbf{X}^2_{srs}$ is generally not a chi-square distribution, we apply an adjustment using the $\hat{\mathbf{\Delta}}$ matrix given by:

$$\hat{\mathbf{\Delta}} = (\mathbf{L}\hat{\mathbf{V}}_{srs}(\hat{\mathbf{B}})\mathbf{L}')^{-}(\mathbf{L}\hat{\mathbf{V}}(\hat{\mathbf{B}})\mathbf{L}').$$

$\hat{\mathbf{V}}(\hat{\mathbf{B}})$ is an estimated asymptotic covariance matrix under the complex sample design. We use second-order adjustment as in Rao and Scott's (1984) given by

$$\mathbf{X}^2_{adj} = \frac{\mathbf{X}^2_{srs}}{tr\hat{\mathbf{\Delta}}/d}$$

with

$$d = \frac{(tr\hat{\mathbf{\Delta}})^2}{tr\hat{\mathbf{\Delta}}^2}.$$

Asymptotic distribution of $\mathbf{X}^2_{adj}$ is approximated by the chi-square distribution with d degrees of freedom. See Graubard and Korn (1993) for properties of this statistic in reference to regression problems.

Adjusted Wald F test

We also use the F-based variant of the Rao and Scott's (1984) second-order adjustment defined by the following:

$$F_{adj} = \frac{\mathbf{X}^2_{srs}}{tr\hat{\mathbf{\Delta}}}.$$

Asymptotic distribution of F_{adj} is approximated by the F-distribution $F(d, d\nu)$ where d is defined as above and ν is the number of the sample design degrees of freedom. See Thomas and Rao (1987) for the heuristic derivation of this test, and Rao and Thomas (2003) for a review of the related simulation studies.

Individual tests

Each row l' of matrix $\mathbf{L}$ may also be tested separately. For such tests, or when matrix $\mathbf{L}$ contains a single row, the statistics above simplify as follows:

$$X^2 = \frac{(l'\hat{\mathbf{B}} - k)^2}{l'\hat{\mathbf{V}}(\hat{\mathbf{B}})l}$$

and

$$X^2 = F = X^2_{adj} = F_{adj}.$$

Asymptotic distribution used for test statistics X^2 and X^2_{adj} is the chi-square distribution with 1 degree of freedom. Test statistics F and F_{adj} are approximated by the F-distribution $F(1,\nu)$ with ν as defined earlier. Tests are undefined if $l'\hat{\mathbf{V}}(\hat{\mathbf{B}})l$ is not positive.

P-values

Given a value of test statistic T and a corresponding cumulative distribution function G as specified above, the p-value p of the given test is computed as $p = 1 - G(T)$.

Multiple comparison tests

In addition to the testing methods mentioned in the previous section, the hypothesis $H_0 : \mathbf{LB} = \mathbf{K}$ can also be tested using the multiple row hypotheses testing technique. Let l'_i be the i-th row vector of matrix $\mathbf{L}$, and k_i be the i-th element of vector $\mathbf{K}$. The i-th row hypothesis is $H_{0i} : l'_i\mathbf{B} = k_i$. Testing H_0 is the same as testing multiple hypotheses $\{H_{0i}\}_{i=1}^{R}$ simultaneously, where R is the number of non-redundant row hypotheses. A hypothesis H_{0i} is redundant if there exists another hypothesis $H_{0j}, j \neq i$ such that $l_i = cl_j, k_i = ck_j, c \neq 0$.

For each individual hypothesis H_{0i}, tests described in the previous section can be performed. Let p_i denotes the p-value for testing H_{0i}, and p_i^* denotes the adjusted p-value. The conclusion from the multiple testing is, at level α (the family-wise type I error),

reject $H_{0i} : l'_i\mathbf{B} = k_i$ if $p_i^* < \alpha$;

reject $H_0 : \mathbf{LB} = \mathbf{K}$ if $\min_i(p_i^*) < \alpha$.

There are different methods to adjust p-values. Five methods are provided here. Please note that if the adjusted p-value is bigger than 1, it is set to 1 in all the methods.

LSD (Least Significant Difference)

The adjusted p-values are the same as the original p-values:

$$p_i^* = p_i.$$

BONFERRONI

The adjusted *p*-values are

$$p_i^* = R p_i.$$

SIDAK

The adjusted *p*-values are

$$p_i^* = 1 - (1 - p_i)^R.$$

Sequential Bonferroni test (Holm)

In sequential test, the *p*-values are first ordered from the smallest to the biggest, and then adjusted depending on the order. Let the ordered *p*-values be $p_{(1)} \leq p_{(2)} \leq \cdots \leq p_{(R)}$ with corresponding hypotheses being $H_{0(1)}, H_{0(2)}, \cdots, H_{0(R)}$.

The adjusted *p*-value of $p_{(i)}$ is

$$p_{(i)}^* = \begin{cases} R p_{(1)} & i = 1 \\ \max\left((R - i + 1) p_{(i)}, p_{(i-1)}^*\right) & i \geq 2 \end{cases}.$$

Sequential Sidak test

The adjusted *p*-value of $p_{(i)}$ is

$$p_{(i)}^* = \begin{cases} 1 - (1 - p_{(1)})^R & i = 1 \\ \max\left(1 - (1 - p_{(i)})^{R-i+1}, p_{(i-1)}^*\right) & i \geq 2 \end{cases}.$$

Comparison of adjustment methods

A multiple testing procedure tells not only if H_0 is rejected, but also if each individual H_{0i} is rejected. All the methods, except LSD, control the family-wise type I error for testing H_0, i.e. the probability of rejecting at least one individual hypothesis under H_0. In addition, sequential methods also control the family-wise type I error for testing any subset of $\{H_{0i}\}_{i=1}^{R}$.

LSD is the one without any adjustment, it rejects H_0 too often. It does not control the family-wise type I error and should never be used to test H_0. It is provided here mainly for reference.

Bonferroni is conservative (i.e. reject less often than it should be). In some situations, it becomes extremely conservative when test statistics are highly correlated.

Sidak is also conservative in most cases, but is less conservative than Bonferroni. It gives the exact type I error when test statistics are independent.

Sequential Bonferroni (Holm) is as conservative as the Bonferroni in terms of testing H_0 because the smallest adjusted p-value used in making decision is the same in both methods. But in term of testing individual H_{0i}, it is less conservative than the Bonferroni. Sequential Bonferroni rejects at least as many individual hypotheses as Bonferroni.

Sequential Sidak is as conservative as the Sidak in terms of testing H_0, but less conservative than the Sidak in terms of testing individual H_{0i}. Sequential Sidak is less conservative than sequential Bonferroni.

References

Fellegi, I. P. (1980), "Approximate tests of independence and goodness of fit based on stratified multistage samples", Journal of the American Statistical Association, volume 75, 261-268.

Graubard, B. I. and Korn, E.L. (1993), "Hypothesis testing with complex survey data: The use of classical quadratic test statistics with particular reference to regression problems", Journal of the American Statistical Association, 88, 629-41.

Holm, S. (1979). A simple sequentially rejective multiple test procedure. Scandinavian Journal of Statistics, 6, 65–70.

Koch, G. G., Freeman, D. H. and Freeman, J. L. (1975), "Strategies in the multivariate analysis of data from complex surveys", International Statistical Review, 43, 59-78.

Korn, E.L. and Graubard, B. I. (1990), "Simultaneous testing of regression coefficients with complex survey data: Use of Bonferroni t statistics", American Statistician, 44, 270-76.

Rao, J. N. K., and Scott, A. J. (1984), "On chi-squared tests for multiway contingency tables with cell proportions estimated from survey data", Annals of Statistics, volume 12, pages 46-60.

Rao, J. N. K., and Thomas, D. R. (2003), "Analysis of categorical response data from complex surveys: an upraisal and update", In Analysis of Survey Data, ed. R.Chambers and C. Skinner. New York: John Wiley & Sons.

Thomas, D. R., and Rao, J. N. K. (1987), "Small-sample comparisons of level and power for simple goodness-of-fit statistics under cluster sampling", Journal of the American Statistical Association, volume 82, pages 630-636.

Wright, S. P. (1992). Adjusted P-values for simultaneous inference. Biometrics, 48, 1005–1013.