NONASYMPTOTIC GOODNESS-OF-FIT TESTS FOR CATEGORICAL DATA

KENNETH J. BERRY AND PAUL W. MIELKE, JR.

Colorado State University, Fort Collins

NONASYMPTOTIC GOODNESS-OF-FIT TESTS FOR CATEGORICAL DATA

[Abstract]

A FORTRAN-77 subroutine is presented to calculate two test statistics for assessing goodness-of-fit in categorical data. The test statistic inferences are based on the exact first three cumulants of the Pearson chi-square statistic and a recently published modification of the Pearson chi-square statistic.

NONASYMPTOTIC GOODNESS-OF-FIT TESTS FOR CATEGORICAL DATA

It is often necessary to test a null hypothesis that n observations follow category probabilities that are specified *a priori*, i.e., a goodness-of-fit test. The usual asymptotic goodness-of-fit tests are known to perform poorly for cases with small expected category frequencies. Recently, two nonasymptotic methods for testing *a priori* specified category probabilities have been developed by Mielke and Berry (1988). The purpose of this paper is to present a FORTRAN-77 subroutine for goodness-of-fit (GOF), which uses the methods of Mielke and Berry (1988) to test for goodness-of-fit between the observed category frequencies and the corresponding *a priori* category probabilities.

Subroutine

Let O_i denote the observed frequency for the ith of k mutually-exclusive and exhaustive categories where

$$\sum_{i=1}^{k} O_i = n$$

and n is the frequency total. Also let $p_i > 0$ denote the occurrence probability of the ith of the k categories where

$$\sum_{i=1}^{k} p_i = 1 .$$

The null hypothesis, H_0, simply states that the p_i values are the correct occurrence probabilities. The classical goodness-of-fit chi-square test statistic corresponds to

$$T = \sum_{i=1}^{k} \frac{O_i^2}{E_i}$$

and the modified goodness-of-fit chi-square test statistic, due to Zelterman (1986, 1987), is given by

$$S = \sum_{i=1}^{k} \frac{O_i(O_i-1)}{E_i}$$

where $E_i = np_i$ is the expected frequency of the ith of the k categories. Subroutine GOF begins by calculating T and S and then computes the first three exact cumulants of T and S (Mielke and Berry, 1988) to obtain the exact means (μ_T, μ_S), variances (σ_T^2, σ_S^2), and skewness terms (γ_T, γ_S) of the sampling distributions of T and S, respectively. Finally, subroutine GOF computes the probability values associated with T and S under H_0 given by $P_{T_o} = P(T \geq T_o | H_0)$ and $P_{S_o} = P(S \geq S_o | H_0)$ where T_o and S_o are the observed values of T and S, respectively.

The expected value of S under an alternative hypothesis, H_1, is always greater than the expected value of S under H_0. However, the expected value of T under H_1 sometimes can be less than the expected value of T under H_0, a most undesirable outcome. If $2 \leq k \leq 6$ and

$$\binom{n+k-1}{k-1}$$

is not too large (i.e., less than 10^7), then Fisher's exact goodness-of-fit test (Mielke and Berry, 1993) is the method of choice. The nonasymptotic goodness-of-fit tests based on the statistics T and S are apropos for $k \geq 2$ and any value of

n. The adjustment involving the three exact cumulants under H_0 obviates the usual concerns regarding small expected category frequencies when using the commonly-employed asymptotic versions of T (i.e., the Pearson chi-square distributions and their degree-of-freedom conditions), even when parameters are fit from the observed frequency data.

It should be noted that while T and S appear quite similar, they behave very differently when small sample sizes are encountered. Whenever $k \le 6$ and

$$\binom{n+k-1}{k-1}$$

is not too large, Fisher's exact goodness-of-fit test (Mielke and Berry, 1993) should be used. However, when either $k \ge 7$ and/or

$$\binom{n+k-1}{k-1}$$

is large, the use of the test based on statistic S rather than statistic T is most definitely the preferred choice for the reason noted above regarding the expected value of T under H_1. A nonasymptotic test analogous to S (Berry and

Mielke, 1989) for analyzing independence in r-way contingency tables eliminates the degrees-of-freedom problem inherent in the classical asymptotic chi-square test.

Example

Consider an example where n = 208 northern European adults are classified into k = 7 categorical blood groupings on the Rh system. Table 1 lists seven Rh blood groupings using the Fisher-Race notation, along with the category

---

Insert Table 1 about here

---

probabilities (Race and Sanger, 1975, p. 184). Since the Rh haplotype *CdE* has such a low expected value among northern Europeans, it has been dropped from the list of the eight possible Fisher-Race categories for this example. Under the null hypothesis of no difference between the observed and expected frequencies, $T_o = 13.1010$, $\mu_T = 6.0000$, $\sigma_T^2 = 14.8495$, $\gamma_T = 2.0217$, and $P_{T_o} = 0.0583$, while $S_o = 212.7500$, $\mu_S = 207.000$, $\sigma_S^2 = 11.9423$, $\gamma_S = 1.4407$, and $P_{S_o} = 0.0690$.

Subroutine Language

Subroutine GOF is written in ANSI FORTRAN-77 for an IBM RISC/6000-930 in double precision. Comment lines provide input/output specification and documentation. An interactive driver program, supplied with subroutine GOF, prompts for the input. The input consists of the number of categories (k), the observed cell frequencies ($O_1, \ldots, O_k$) and the expected cell probabilities ($p_1, \ldots, p_k$). The output consists of a summary of the input, the observed test statistics T_o and S_o, the exact mean, variance, and skewness values for T and S, and the associated probability values P_{S_o} and P_{T_o}.

Availability

A listing of subroutine GOF and an appropriate driver program are available from Kenneth J. Berry, Department of Sociology, Colorado State University, Fort Collins, CO 80523, or by e-mail from Internet address: berry@lamar.colostate.edu.

REFERENCES

Berry, K. J. and Mielke, P. W. (1989). Analyzing independence in *r*-way contingency tables. EDUCATIONAL AND PSYCHOLOGICAL MEASUREMENT, 49, 605-607.

Mielke, P. W. and Berry, K. J. (1988). Cumulant methods for analyzing independence of *r*-way contingency tables and goodness-of-fit frequency data. *Biometrika*, 75, 790-793.

Mielke, P. W. and Berry, K. J. (1993). Exact goodness-of-fit probability tests for analyzing categorical data. EDUCATIONAL AND PSYCHOLOGICAL MEASUREMENT, 54, xxx-xxx.

Race, R. R. and Sanger, R. (1975). *Blood Groups in Man, 6th Edition*. Oxford: Blackwell Scientific Publishers.

Zelterman, D. (1986). The log likelihood ratio for sparse multinomial mixtures. *Statistics and Probability Letters*, 4, 95-99.

Zelterman, D. (1987). Goodness-of-fit tests for large sparse multinomial distributions. *Journal of the American Statistical Association*, 82, 624-629.

TABLE 1

Observed Frequencies and Expected Probabilities for Seven Blood Groupings on the Rh System for 208 Northern European Adults (Fisher-Race Notation)

Haplotype	Observed Frequency	Expected Probability
CDe	98	0.4205
cDE	24	0.1411
cDe	12	0.0257
CDE	1	0.0024
Cde	1	0.0098
cdE	2	0.0119
cde	70	0.3886