Contingency table testing for categorical data: SAS implementation

LYNN M. VEATCH and LARRY P. GONZALEZ
University of Oklahoma Health Sciences Center
Oklahoma City, Oklahoma

Nonparametric and distribution-free tests of categorical data provide an evaluation of statistical significance between groups of subjects differing in their assignment to a set of categories. This paper describes an implementation in the SAS programming language of three tests to evaluate categorical data. One of these tests, the Contingency Table Test for Ordered Categories evaluates data assessed on at least an ordinal scale where the categories are in ascending or descending rank order. The remaining two tests, Fisher's Fourfold-Table Test for Variables with Two Categories and Fisher's Contingency Table Test for Variables with More than Two Categories, evaluate data assessed on either a nominal or an ordinal scale. The program described completes analysis of a 2 × C categorical contingency table as would be obtained from the application of a multiple-level rating scale to the behavior of a treatment and a control group.

Many areas of research evaluate change in behavior as a dependent variable and use observational rating scales to measure these behavioral changes. In psychopharmacological research, rating scales are often used to assess the acute or chronic effects of drugs in laboratory animals (Jacobs & Falgoust, 1984; Kabes, 1972; Kurlan, Kim, & Gash, 1991; Rubin, 1978), where degree of motor dysfunction or impairment, seizure severity, or other behavioral changes can be rated by an experimental observer. Similarly, clinical research applications use rating scales in the evaluation of disease severity, drug side-effect profiles, and treatment success (Banger, Philipp, Herth, Hebenstreit, & Aldenhoff, 1992; Bech, 1988; Fankhauser & German, 1987; Moses, Emerson, & Hosseini, 1988).

Statistical evaluation of the measurement data obtained from use of these instruments requires consideration as to the actual type of information obtained and proper statistical analysis required. Since many of these rating instruments are based upon nominal or ordinal scales (rather than interval or ratio scales) and the underlying distribution of data obtained using such instruments is not always known, nonparametric or distribution-free statistical procedures often provide the most appropriate tests of group differences (Siegel, 1956). Krauth (1988) described a variety of distribution-free statistical procedures that are appropriate for the analysis of this type of data. These tests have had limited use, however, because

they are mathematically tedious and are not readily available in most common statistical analysis computer software systems.

In this paper, we describe a program written in the SAS language (SAS Institute, 1987) to perform three tests of categorical data as described by Krauth (1988). The program evaluates data from two or more nominal or ordinal categories and provides significance tests for Fisher's Fourfold-Table Test for Variables With Two Categories, or for the Contingency Table Test for Ordered Categories and Fisher's Contingency Table Test for Variables With More Than Two Categories. Examples are presented of the application of these tests to data illustrated in Krauth (1988) and also to data from the authors' laboratory examining changes in the seizure sensitivity of animals withdrawn from chronic ethanol exposure.

Method

The statistical procedures described here analyze categorical data obtained when a rating scale is used to evaluate the responses of two groups of subjects—usually, one treatment group and one control group. Data consist of the number of subjects within each group receiving a particular rating or assigned to a particular response category. It is assumed that data are measured at least on a nominal scale. Data may also be measured on an ordinal scale—in which case, category assignment is also assumed to indicate response strength or severity.

In general, these contingency table tests are based upon a determination of the probability of obtaining a contingency table with marginal sums identical to those of the test table merely by chance. The program evaluates data following procedures outlined by Krauth (1988) for the three specific tests. Figure 1 provides an overview of the program flow. First, the initial value from the input file is read to establish the number of columns in the table, the table is then read, and the multiple hypergeometric probability statistic and marginal sums are calculated for the input contingency table.

Alternate contingency tables are then determined following the procedures of Krauth (1988) for evaluating small samples by determining the minimum and maximum treatment cell values and the resultant control cell values. Once an alternate table has been determined, the calculated table is compared with the original input table using either cumulative marginal sums or specific cell values as specified by each test.

In the final module of the program, summary statistics are generated. The specific statistics presented by the program are determined by the number of columns, or rating categories, in the input table.

Examples

Two examples of problem contingency tables and the resultant computer program output are included as Fig-
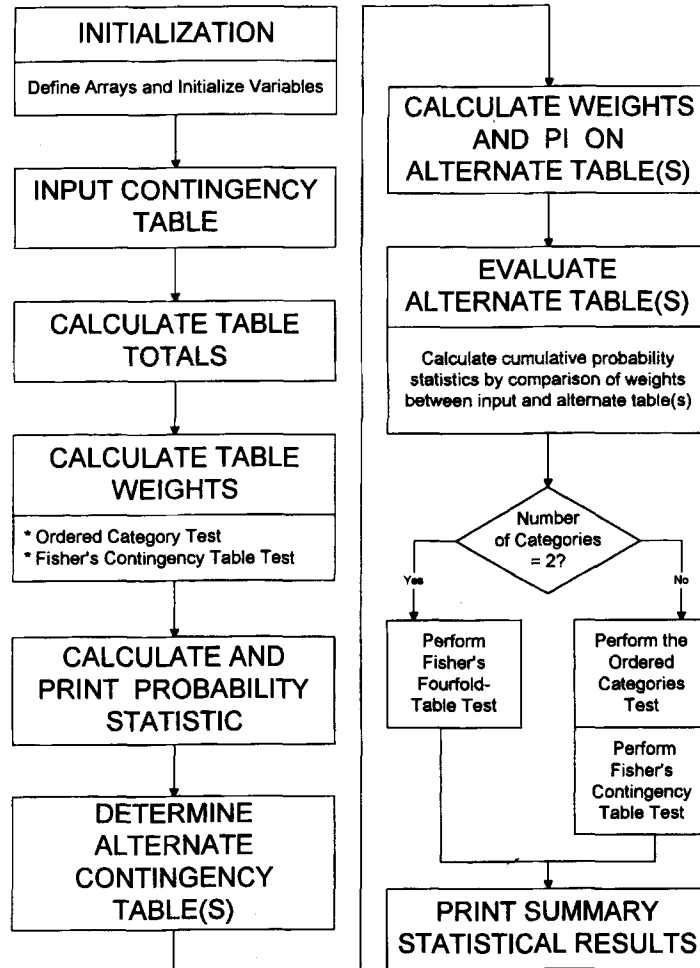
Figure 1. Overview of the SAS Program for Contingency Table Testing of Categorical Data.

ures 2 and 3. These are examples used by Krauth (1988) in discussion of the test procedures.

Figure 2 shows a 2 × 2 contingency table evaluated using Fisher's Fourfold-Table Test for Variables With Two Categories. As discussed by Krauth (1988, p. 80), this test is appropriate for the analysis of 2 × 2 contingency tables composed of nominal or ordinal scale data. The significance of group differences is determined by comparing the calculated test probabilities with an experimenter-selected alpha level for a one-sided test or with one half the alpha level for a two-sided test. With $\alpha = 0.05$, the groups shown here do not differ significantly from one another.

A 2 × 3 contingency table is shown in Figure 3 and is evaluated using both Fisher's Contingency Table Test for Variables With More Than Two Categories and the Contingency Table Test for Ordered Categories. Krauth (1988, p. 86) used this table in discussion of the Contingency Table Test for Ordered Categories. Since correct

interpretation of the results depends upon whether the input data are nominal or ordinal, the computer program output indicates the type of values for which each statistic is meaningful. This example illustrates the greater statistical power in testing ordinal data relative to nominal data. The groups would be considered significantly different ($\alpha = 0.05$) if the data are ordinal, but not if the data are measured on a nominal scale.

A final example of the use of this contingency table test is presented in Figure 4. The data were obtained from studies conducted in the authors' laboratory, examining the effects of chronic ethanol exposure and withdrawal on picrotoxin-induced seizure thresholds in laboratory rats. The procedures for ethanol exposure, withdrawal, and seizure testing are presented in detail elsewhere (Gonzalez, 1993; Gonzalez, Czachura, & Brewer, 1989). In general, male Sprague-Dawley rats received chronic ethanol exposure in ethanol-vapor inhalation chambers for a period of 21 days, with blood

CONTINGENCY TABLE

TEST RESULTS

Figure 2. Example of the problem contingency table and computer program output evaluating a 2 × 2 contingency table. Results of the one-sided test are evaluated in comparison with an experimenter-selected alpha level; results of the two-sided test are compared with alpha/2.

ethanol levels approaching 400 mg/dl at the time of removal from the ethanol chambers. Control animals received similar handling, but no ethanol exposure.

After 21 days, animals were removed from the vapor chambers and tested 10 h later for their response to the convulsant picrotoxin. For seizure testing, animals received a single, acute injection of picrotoxin (3.0 mg/kg, i.p.). Picrotoxin, at various doses, is observed to induce a series of responses in some animals, which progress from (1) myoclonic jerks, (2) partial seizures, (3) generalized tonic/clonic seizures, (4) tonic extension of the limbs, and, finally, to (5) severe tonic seizures. Progression to each stage of these responses is dose dependent, and, thus, the stage of behavioral response exhibited can serve as an ordinal measurement of seizure severity. Responses to picrotoxin were rated on a scale of 0 (*no response*) to 5 (*severe tonic seizures*), indicating the stage of behavioral response exhibited. Data obtained from two groups of animals are summarized in Figure 4. The groups include an ethanol-naive control and a group of ethanol-treated animals withdrawn from ethanol for 10 h. As shown in the figure, the contingency table test for ordered categories indicates that the responses of ethanol-withdrawn animals to picrotoxin were significantly more severe ($p \leq .016$) than were those of ethanol-naive controls.

Conclusions

Nonparametric tests of categorical data provide the researcher with powerful tools for statistical evaluation of frequency data. The contingency table tests discussed above are particularly appropriate for the evaluation of observational rating scales as are often used in psychopharmacological research. The SAS program described here analyzes a 2 × C contingency table using (1) Fisher's Fourfold-Table Test for Variables with Two Categories, (2) Fisher's Contingency Table Test for Variables with More than Two Categories, and (3) the Contingency Table Test for Ordered Categories. These tests provide for the statistical analysis of categorical data without the requirement for a priori knowledge about the distribution of the data. Where data do meet the more strict requirements of parametric tests, these nonparametric tests have the disadvantage of being lower in statistical power in relation to analogous parametric tests, thus requiring a larger number of observations in the data pool to obtain the same relative statistical power. In contrast to analogous parametric tests, however, nonparametric contingency tests are more sensitive to data medians than to means and are thus less affected by outlying data points (Zimmerman & Zumbo, 1990). See Krauth (1988) for a more thorough discussion of the statistical advantages and limitations of contingency table tests.

Program Availability

A listing of the SAS source code described in this paper is available on request from the authors. The listing also includes documentation of procedures and variables used. An ASCII disk file of the SAS program source code can be obtained by sending a formatted, MS-DOS diskette (5.25 or 3.5 in.) to the authors.

CONTINGENCY TABLE

TEST RESULTS

Figure 3. Example of the problem contingency table and computer program output evaluating a 2 × 3 contingency table.

CONTINGENCY TABLE

Rating

	0	1	2	3	4	5
Chronic Ethanol Treatment (T)	0	0	0	4	1	3
Ethanol-Naive Control (C)	0	0	2	6	0	0

TEST RESULTS

MULTIPLE HYPERGEOMETRIC PROBABILITY : 0.016

••••••••••••••••••••••••••••••••• NOMINAL DATA ••
FISHER'S CONTINGENCY TABLE TEST FOR VARIABLES WITH 3+ CATEGORIES

T IS DIFFERENT FROM C: $p <= 0.58$

••••••••••••••••••••••••••••••••• ORDINAL DATA •••
CONTINGENCY TABLE TEST FOR ORDERED CATEGORIES:

T IS SMALLER THAN C: $p <= 1.00$
T IS GREATER THAN C: $p <= 0.016$

Figure 4. Results of analyzing the effects of chronic ethanol exposure and withdrawal on responses to picrotoxin. The results indicate that, on a nominal scale, the distribution of scores for the chronic-ethanol treatment group does not differ from that of the ethanol-naive control group. On an ordinal scale, the results indicate that the ethanol group demonstrates statistically higher seizure severity than does the control group.

REFERENCES

BANGER, M., PHILIPP, M., HERTH, T., HEBENSTREIT, M., & ALDEN-HOFF, J. (1992). Development of a rating scale for quantitative measurement of the alcohol withdrawal syndrome. *European Archives of Psychiatry & Clinical Neuroscience*, **241**, 241-246.

BECH, P. (1988). Rating scales in psychopharmacology: Statistical aspects. *Acta Psychiatrica Belgica*, **88**, 291-302.

FANKHAUSER, M. P., & GERMAN, M. L. (1987). Understanding the use of behavioral rating scales in studies evaluating the efficacy of antianxiety and antidepressant drugs. *American Journal of Hospital Pharmacy*, **44**, 2087-2100.

GONZALEZ, L. P. (1993). Sensitivity to strychnine seizures is unaltered during ethanol withdrawal. *Alcoholism: Clinical & Experimental Research*, **17**, 1029-1034.

GONZALEZ, L. P., CZACHURA, J. F., & BREWER, K. W. (1989). Spontaneous versus elicited seizures following ethanol withdrawal: Differential time course. *Alcohol*, **6**, 481-487.

JACOBS, K. W., & FALGOUST, L. (1984). Psychometric properties of King's Rodent Emotionality Rating Scale. *Psychological Reports*, **55**, 951-958.

KABES, J. (1972). Rating scale for autonomic, motor and central effects in dogs: Effects of anticholinergic and cholinergic drugs. *Activitas Nervosa Superior*, **14**, 164-165.

KRAUTH, J. (1988). *Distribution-free statistics: An application-oriented approach*. Amsterdam: Elsevier.

KURLAN, R., KIM, M. H., & GASH, D. M. (1991). Oral levodopa dose-response study in MPTP-induced hemiparkinsonian monkeys: Assessment with a new rating scale for monkey parkinsonism. *Movement Disorders*, **6**, 111-118.

MOSES, L. E., EMERSON, J. D., & HOSSEINI, H. (1988). Statistics in practice analyzing data from ordered categories. *New England Journal of Medicine*, **311**, 442-448.

RUBIN, A. A. (1978). *New drugs: Discovery and development*. New York: Marcel Dekker.

SAS INSTITUTE, INC. (1987). *Statistical Analysis System (SAS)* [Proprietary Software Program Release 6.04]. Cary, NC: SAS Institute.

SIEGEL, S. (1956). *Nonparametric statistics for the behavioral sciences*. New York: McGraw-Hill.

ZIMMERMAN, D., & ZUMBO, B. D. (1990). Effect of outliers on the relative power of parametric and nonparametric statistical tests. *Perceptual & Motor Skills*, **7**, 339-349.