PROGRAM ABSTRACTS/ALGORITHMS

An Applesoft BASIC program for the G statistic: An alternative to the chi-square test

BERNARD C. BEINS
Ithaca College, Ithaca, New York

The analysis of nominal data has developed beyond the traditional Pearson chi-square test that has dominated statistics based on frequency counts. Different discussions have presented the logic and rationale for using these new statistical applications (Overall, Rhoades, & Starbuck, 1987; Reynolds, 1977; Sokal & Rohlf, 1981). The G statistic presented here is one such development. It is described in detail in Sokal and Rohlf (1981, pp. 691-746).

The G statistic is a log likelihood ratio test, a class of statistics largely initiated by Goodman (1970). It is conceptually quite similar to the traditional chi-square test in that both can be used for goodness of fit tests, as well as tests of independence for row $\times$ column arrays of data. The formula for finding G is given in Equation 1:

$$G = 2 \times f_i \sum \ln(f_i/\hat{f}_i), \quad (1)$$

where f_i and $\hat{f}_i$ indicate the observed and expected frequencies in cell i .

The G test holds one minor and one major advantage over chi-square. The minor advantage is that, although the sampling distributions of the G and chi-square tests are based on the theoretical chi-square distribution with $k-1$ degrees of freedom, G conforms more closely to the theoretical chi-square distribution than the empirically based chi-square test (Sokal & Rohlf, 1981). Because the magnitude of this advantage is not large, however, it is probably not of much practical significance. The major advantage of the G test is that it can identify departures from expectations for subgroups of data, as well as for an entire data set. If an investigator replicates a given design to test for goodness of fit, several different, useful G values can be obtained from a single data set.

Pooled G . The pooled G value is obtained when the data from individual replicates of a goodness of fit test are combined and a G test is computed on the aggregate data. The pooled G value assesses whether or not all data, combined, produce a significant departure from expectations, even if individual test results are not significant.

Total G . One can also sum the independent, calculated G values to see how likely the pattern of individual

G values really is. This result is useful for a research program involving multiple replicates of the same design. The pooled G value might not be significant because different replicates can show different patterns of departure from expectation, thus offsetting one another. The total G value indicates whether or not the results of individual tests, taken together, are significant, even if their observed patterns differ.

Heterogeneity G . The discrepancy between the sum of individual and pooled G values reflects heterogeneity in the data and can be tested in what is called an *interaction*, or *heterogeneity*, G test. A significant value for the heterogeneity G test reflects the fact that the data for at least one replicate depart from the pattern of the others. This value is analogous to an interaction effect in the analysis of variance.

This partitioning across groups is not completely particular to the G test. For example, as Siegel and Castellan (1988) note, it is reasonable to partition the results of a row $\times$ column design into separate chi-square tests, each having 1 degree of freedom. In this way, one can assess the locus of significant departures from expectation. Because the a priori sequence of separate tests is not fixed in any theoretical sense, however, different orders of partitioned chi-square tests may result in different values (Siegel & Castellan, 1988, pp. 194-198). The way to interpret these discrepancies may not always be clear. With the G test, on the other hand, the user is constrained in the structure of the analysis, and the outcome may be more tractable conceptually because each G test originally computed is directly represented in the output. The G test is appropriate when marginal values either are not fixed or are fixed for one variable (Sokal & Rohlf, 1981). (When marginal values are fixed for both variables in a 2×2 design, Sokal and Rohlf recommend Fisher's exact probability test, although as Overall, Rhoades, & Starbuck, 1987, point out, Fisher's test is extremely conservative in its actual alpha level.) The G test also appears to be more useful (and more accurate) than the chi-square test when the number of cells in a design is greater than or equal to 5 and the expected values equal 3, or when expected values in general exceed 10 (Conahan, 1970, cited in Sokal & Rohlf, 1981). Furthermore, there is a correction appropriate for goodness of fit and contingency tests that renders the G test more accurate with respect to actual and nominal alpha levels (Williams, 1976). The question of the match between a given statistic and its underlying theoretical distribution is important here because, as Overall, Rhoades, and Starbuck (1987) have noted, with smaller samples, the actual and nominal levels may not match very closely in some instances. Dunlap, Myers, and Silver (1984) provide exact values in multinomial

I thank Bill Porter for his useful comments on the program and on this manuscript. I am also grateful to two anonymous reviewers who also contributed helpful suggestions on the program and on the theoretical background. Correspondence may be sent to Bernard C. Beins, Psychology Department, Ithaca College, Ithaca, NY 14850.

cases, in those instances in which exact values are especially desirable.

The Program

This program is written in Applesoft BASIC (DOS 3.3). It is listable and not copy protected and runs on any of the Apple II family of computers, as well as Franklin 1200s and Laser systems.

The user specifies the numbers of rows and columns, and enters the data for each test, including expected values if they are known. (Expected values of zero are not permitted.) Default values are computed and used when expected values are not specified; these values are obtained just as for the traditional chi-square test. During all data entry, the program will not accept nonnumerical values. In addition, either the reverse arrow or the delete key can be used in making changes during data entry. Furthermore, a correction subroutine will allow the user to change either the design (i.e., row $\times$ column specifications) or individual data points that have already been entered. Throughout the program, the user presses the return key to continue and the space bar to make corrections. Finally, the output can be printed to monitor alone or to both monitor and printer.

There is also an option to enter data on multiple replicates of the same goodness of fit design (i.e., same number of columns in a one-way table) to generate a pooled G test and a heterogeneity G test. If the user performs only one G test, the results include two G values. One is based on Formula 1 above; the other is an adjusted value, G (Adj), that corrects for the fact that the nominal value of alpha in a formal hypothesis will underestimate the true value of alpha. The adjusted value is obtained by computing G in the standard fashion and then dividing by the appropriate correction factor, q , as shown in Formulas 2 and 3 below. For multiple replicates, the adjusted value of G is omitted because these values have been shown not to be additive. Specifically,

$$q = 1 + \frac{(C - 1)}{6 \times \text{Total } N} \quad (2)$$

for a one-way design, or

$$q = 1 + \frac{[n \sum^R (1/\sum f_c) - 1][n \sum^C (1/\sum f_r) - 1]}{6 \times N(R - 1)(C - 1)} \quad (3)$$

for a row $\times$ column design, where R is the number of rows, C is the number of columns, f_r is the observed frequency in row r and f_c is the observed frequency in column c .

There are several subroutines in the program. They include the introduction for the user (lines 100–190), instructions (220–270), input for column (1,000–1,099), and row (1,100–1,182) data, observed (1,190–1,299) and expected values when unknown (1,500–1,599) and when known (10,200–10,370), data correction (11,000–12,170), and printing (16,000–16,240).

Example. As an example of the use of the program, I investigated whether excessive absence in an introductory psychology class was associated with grades. On the basis

Table 1
Example of Observed Frequencies of Students Excessively Absent from Class According to Final Grade, the Resultant G Values, and the Summary Table Generated by the Program

Example Data:

Class	Grade				
	F	D	C	B	A
1	2 (0.8814)	7 (2.8644)	10 (8.1525)	6 (8.8136)	1 (5.2881)
2	2 (0.6897)	6 (3.7931)	15 (10.3488)	6 (12.0690)	1 (3.1034)
3	3 (0.8769)	2 (0.8769)	7 (5.8462)	6 (9.6462)	1 (1.7588)

Program Output:

Test Number	G	G (Adj)	df
1	11.93	11.63	4
2	10.25	10.02	4
3	6.37	6.16	4
Pooled	26.53		4
Heterogeneity	2.02		8
Total	28.55		12

Note—Expected values are in parentheses.

of the number of students receiving each grade (A, B, C, D, F), I calculated the expected number of student absences in each grading category for classes across three semesters. One would expect more absences among C students because there are more of them. A and F students are rarer and should show fewer absences. I tested to see whether or not the proportion of students in each category matched the predicted values. A source table similar to that for an ANOVA indicates the patterns of the G values, as shown in Table 1. The program does not inform the user of the significance level of the results; for this information, one can consult a table of chi-square values. As revealed by individual G tests, absences for the first two courses did not match the expectations; instead, students with high grades showed fewer absences than expected, and lower grades were associated with greater absence; for the third class, the numbers of absences of students in each grade category did not differ significantly from what was predicted. The significant pooled G value reflects the fact that when attendance data for all students from the three classes were combined, the pattern of absences according to grade departs from expectations. The failure to find a significant heterogeneity G value indicates that, even though the third class showed no significant departure from expectations, the pattern from this class did not deviate significantly from the other two.

Availability. A listing of the program is available without cost. To receive a copy of the program, please send a 5.25-in. diskette and a self-addressed stamped mailer to the author.

REFERENCES

- DUNLAP, W. P., MYERS, L., & SILVER, N. D. (1984). Exact multinomial probabilities for one-way contingency tables. *Behavior Research Methods, Instruments, & Computers*, **16**, 54-56.
- GOODMAN, L. A. (1970). The multivariate analysis of qualitative data: Interactions among multiple classifications. *Journal of the American Statistical Association*, **65**, 226-256.
- OVERALL, J. E., RHOADES, H. M., & STARBUCK, R. R. (1987). Small-sample tests for homogeneity of response probabilities in 2×2 contingency tables. *Psychological Bulletin*, **102**, 307-314.
- REYNOLDS, H. T. (1977). *Analysis of nominal data*. Beverly Hills, CA: Sage Publications.
- SIEGEL, S., & CASTELLAN, N. J., JR. (1988). *Nonparametric statistics for the behavioral sciences* (2nd ed.). New York: McGraw-Hill.
- SOKAL, R. R., & ROHLF, F. J. (1981). *Biometry* (2nd ed.). San Francisco, CA: W. H. Freeman.
- WILLIAMS, D. A. (1976). Improved likelihood ratio tests for complete contingency tables. *Biometrika*, **63**, 33-37.

(Manuscript received March 14, 1989;
revision accepted for publication October 2, 1989.)