Small Sample Comparisons of Exact Levels for Chi-Square Goodness-of-Fit Statistics

TR #242

by Kinley Larntz*

Technical Report No. 242

February, 1975

*Kinley Larntz is Assistant Professor, School of Statistics, University of Minnesota, St. Paul, Minnesota 55108. Research for this paper was supported in part by a Grand-in-Aid from the Graduate School, University of Minnesota. This paper is based in part on the author's doctoral dissertation written under Professor Stephen E. Fienberg at the University of Chicago. Support for this research at the University of Chicago was provided by a National Defense Education Act Fellowship and a Warner-Lambert Research Institute Fellowship, with additional support from the Shell Companies Foundation, Inc. The author is indebted to S.E. Fienberg for guidance and to R.R. Bahadar, S.J. Haberman, D.L. Wallace and the referees for valuable comments.

SUMMARY

The small sample properties of three goodness-of-fit statistics are examined with respect to the adequacy of the asymptotic chi-square approximation. In general, the approximate tests based on the likelihood-ratio and Freeman-Tukey statistics yield exact levels that are in excess of the nominal levels. In contrast the Pearson statistic attains exact levels that are quite close to the nominal values. The reason for the large number of rejections for the likelihood-ratio and Freeman-Tukey statistics is related to their handling of small observed counts.

1. INTRODUCTION

Several statistics are commonly used to judge the goodness-of-fit for counted data models. In this paper, three of these statistics will be compared with respect to their small sample properties under the null hypothesis. The usual chi-square statistic (Pearson statistic) is defined by

$$X^2 = \sum_{\text{all cells}} \frac{(\text{Observed} - \text{Expected})^2}{\text{Expected}} . \tag{1.1}$$

A suggested alternative statistic that has some asymptotically optimal properties (see [1] and [12]) is the likelihood-ratio statistic

$$G^2 = 2 \sum_{\text{all cells}} \text{Observed} \log_e (\text{Observed/Expected}). \tag{1.2}$$

Another alternative is the Freeman-Tukey chi-square statistic [2]

$$T^2 = \sum_{\text{all cells}} (\sqrt{\text{Observed}} + \sqrt{\text{Observed} +1} - \sqrt{4 \text{ Expected} +1})^2 \tag{1.3}$$

Many statisticians prefer the use of one or the other of these statistics, although among everyday users the Pearson statistic is by far the most popular. Also, some statistcians follow the practice of reporting two or more statistics (see, for example, [10]), but little guidance is available concerning the occurrence of large discrepancies between the statistics.

In Section 2 we introduce the two models used for comparison. One model is the standard multinomial goodness-of-fit case. The other is a particular parametric model that arises naturally from data in a problem solving experiment. Section 3 provides the small sample comparison of the statistics, and in Section 4 some particular properties of the statistics are examined in more detail. Conclusions are given in the final section.

2. THE MODELS

Comparisons between the statistics are made for two diverse models. Since the results are similar for both models, we believe that the conclusions will apply to the entire range of counted data problems.

2.1 Multinomial Goodness-of-Fit

We consider first the simplest chi-square test which arises in testing the null hypothesis that the observed frequency vector $(n_1,n_2,\ldots,n_k)$ follows a multinomial distribution with specified probability vector $(p_1,p_2,\ldots,p_k)$. If the total sample size is $N = \sum_{i=1}^{k} n_i$, the expected cell frequencies are simply Np_i and the chi-square statistics may be calculated using (1.1) - (1.3).

There are several reasons for selecting this model for making comparisons:

(a) This is the simplest counted data model, with the expecteds depending only on a prespecified probability vector. Comparisons are therefore made for a simple null hypothesis.

(b) The amount of computation is not excessive. For $k = 2$ or 3 and $N \leq 100$, exact computations are easily carried out. For larger numbers of cells, Monte Carlo procedures can be easily programmed.

(c) By varying the probability vector $(p_1,p_2,\ldots,p_k)$, the entire range from skewed to equal cell probabilities can be considered. Previous studies ([4],[18],[19]) have indicated that for small expected values the Pearson statistic does not follow the chi-square distribution well, while some suggestion has been indicated (cf. [3], p. 38) that the likelihood-ratio statistic would be better in such situations. Recently it has been suggested that for certain models the Freeman-Tukey statistic has true significance levels closer to the nominal levels than does either the Pearson or likelihood-ratio statistic [17].

2.2 Group Helping Model

In this example individuals or groups are given the opportunity to help another individual in distress. The degree of help is graded I, II, or III: I for not helping, III for actively helping, and II for an intermediate action. Further details on this particular application can be found in [8] or [16]. Similar models are also used in component testing problems (see [6]).

Consider the case where data were gathered for individuals and groups of size two. Let p_1, p_2, and p_3 be the probabilities of observing an individual with help graded I, II, and III, respectively. Then if the individuals in a group act independently and if only the higher grade of help is scored, p_1^2, $p_2^2 + 2p_1p_2$, and $p_3^2 + 2p_1p_3 + 2p_2p_3$ are the respective probabilities of observing I, II, and III for groups of size two.

Suppose N_1 individuals and N_2 groups are tested. The results can be summarized in a 3 X 2 contingency table with column totals fixed as in Table 1.

Table 1 goes about here

Under the above assumptions, (n_{11},n_{21},n_{31}) follows a multinomial distribution with probability vector (p_1,p_2,p_3) and (n_{12},n_{22},n_{32}) follows a multinomial distribution with probability vector (g_1,g_2,g_3) where

$$\begin{aligned} g_1 &= p_1^2 \\ g_2 &= p_2^2 + 2p_1p_2 \\ g_3 &= p_3^2 + 2p_1p_3 + 2p_2p_3 \;. \end{aligned} \tag{2.1}$$

For this case the unique maximum likelihood estimates for (p_1,p_2,p_3) can be written down directly [8] as

$$\begin{aligned} \hat{p}_1 &= (-n_{31} + \sqrt{n_{31}^2 + 4ac}\,)/2a \\ \hat{p}_2 &= r\hat{p}_1 \\ p_3 &= 1 - (1+r)\,\hat{p}_1 \end{aligned} \tag{2.2}$$

where

$$r = \frac{n_{21} - 2n_{11} - 4n_{12} + \sqrt{(2n_{11}+4n_{12}-n_{21})^2 + 8(n_{21}+n_{22})(n_{11}+2n_{12})}}{2(n_{11}+2n_{12})}\;, \tag{2.3}$$

$$a = (1+r)\;\left[(n_{11}+2n_{12})(1+r) + (n_{31}+2n_{32}) + 2n_{22}(1+r)/(2+r)\right], \tag{2.4}$$

and

$$c = n_{11} + 2n_{12} + 2n_{22}/(2+r)\;. \tag{2.5}$$

(If the i-th row total (R_i in Table 1) is zero, the maximum likelihood estimates are derived conditional on the zero total. The estimates in such a case are just the extension by continuity of the estimates given by (2.2) through (2.5).)

The selection of this model for comparing the chi-square statistics provides several advantages:

(a) The model depends upon two parameters, p_1 and p_2, and thus the goodness-of-fit test for the null hypothesis involves the estimation of these parameters. Comparisons can therefore be made for a composite null hypothesis.

(b) Since the maximum likelihood estimates can be written down in closed form, iteration is not necessary for finding the estimates. This is important when considering the feasibility of doing large amounts of computation.

(c) Examining (2.1), note that the probability of Help Grade I for groups is p_1^2. When p_1 is small, p_1^2 is quite small. Thus, the selection of this model allows for comparisons of very skew multinomials, which means comparisons can be made for small as well as moderate minimum cell expectations.

In the next sections, the properties of the three chi-square statistics for these two models will be presented. Because of the diversity of these models we believe similar results hold for other counted data models.

3. SMALL SAMPLE PROPERTIES UNDER THE NULL HYPOTHESIS

Under the null hypothesis, the goodness-of-fit statistics, X^2, G^2, and T^2 have asymptotic chi-square distributions. However, for small samples the chi-square approximation in many cases does not agree well with the actual distribution. Several studies ([4],[9],[15],[19]) have given conflicting points of view as to at what point the approximation is "reasonable" for the Pearson chi-square statistic. Standard rules specify that the minimum cell expectation should be 5, with possibly a few smaller. The emphasis here will not be on finding such a rule, but in comparing the likelihood-ratio, Freeman-Tukey, and Pearson statistics with regard to the approximation. In other words we ask, for small samples, which of the three statistics is best approximated by the asymptotic chi-square distribution?

3.1 Results for the Multinomial Goodness-of-Fit Model

Both exact and Monte Carlo computations were carried out in the study of the multinomial goodness-of-fit model. For number of cells $k = 2$ and 3, exact enumeration was made of the small sample distributions of the chi-square statistics. For $k = 5$ and 10 cells, Monte Carlo methods [11] were used to simulate the small sample distributions. 1000 trials were run for each null hypothesis and sample size. Table 2 lists the probability vectors used as null hypotheses in this part of the study. For each null hypothesis, computations were carried out for sample sizes $N = 10$ (1) 100.

Tables 2 & 3 go about here

The minimum cell expectation (M.C.E.) for each case can be computed as

$$\text{M.C.E.} = N \times \min\{p_1,\ldots,p_k\} \;. \qquad (3.1)$$

Some evidence has been given that the minimum cell expectation governs the closeness of the small sample distribution to asymptotic theory for several chi-square problems (see, for example, [4],[5],[14],[19]). For each value of k, the cases were divided into classes according to their minimum cell expectations. Table 3 gives the range of actual small sample rejection rates when the nominal .05 level test is used. For example, for k = 3 the statistics are asymptotically distributed as chi-square variables with two degrees of freedom and the asymptotic .05 critical value is 5.991. The values in Table 3 for k = 3 give the minimum and maximum of

$$\alpha_{.05}\,(p_1,p_2,p_3) = P(\text{Statistic} > 5.991) \qquad (3.2)$$

for all cases with minimum cell expecteds within the class boundaries. Although the spread is quite large for many classes, it is clear that for smaller minimum cell expected values ($1 \leq \text{M.C.E.} \leq 4$) the Pearson statistic has an exact size closer to the nominal .05 than either the likelihood-ratio or Freeman-Tukey statistic. For larger M.C.E.'s there does not appear to be a great difference between the three statistics. Also, recall that the values for k = 5 and 10 are based on Monte Carlo results and thus the true spread would be somewhat less than that given in Table 3.

Figures A, B, C and D go about here

Figures A, B, C, and D present graphs of the medians of the exact size for the groupings based on minimum cell expectations. All four graphs are similar in the sense that for small MCE's the likelihood-ratio and Freeman-Tukey statistics are conservative; while for moderate MCE's G^2 and T^2 have rejection rates considerably in excess of the nominal .05 and the excess is greater for small degrees of freedom. In all cases, the Pearson statistic has a median size that is very close to the nominal .05. Results for the .10 and .01 levels give the same general impression. The basic conclusion is that a P-value based on the asymptotic chi-square approximation is "on average" about right for the Pearson statistic, but is understated for the likelihood-ratio and Freeman-Tukey statistics when there are some cell expectations under 5.0. Further exploration of the problems associated with the likelihood-ratio and Freeman-Tukey statistics is given in Section 4.

3.2 Results for the Parametric Model

The model presented in Section 2.2 assumes that the data consist of two independent trinomials: the individuals have probability vector (p_1,p_2,p_3) and the pairs have probability vector (g_1,g_2,g_3), where the g's are given by (2.1). The maximum likelihood estimates for sample sizes N_1 and N_2 are given by equations (2.2) through (2.5). The number of possible outcomes for two trinomials with sample sizes N_1 and N_2 is given (see [7]) by

$$\text{Outcomes} = \binom{N_1 + 2}{2}\binom{N_2 + 2}{2}. \qquad (3.3)$$

For $N_1 = N_2 = 8$, the number of possible outcomes is 2025. For the 36 values of (p_1,p_2,p_3) listed in Table 4 and for $N_1 = N_2 = 4$, 6, 8 12 and 16, the distributions of X^2, G^2 and T^2 were determined by enumeration on the computer.

Table 4 goes about here

One question that arises in the use of this method is how to deal with zero cell counts and zero expected values. As indicated previously, the maximum likelihood estimates were extended by continuity to provide well-defined procedures. In the same manner, when a cell had a zero expected value, it contributed zero to the chi-square statistics.

Figures E, F, and G go about here

For $N_1 = N_2 = 8$, Figure E gives a contour plot of the exact size for the Pearson statistic when the nominal .05 level test is used. Barycentric coordinates were chosen to represent the 3 probabilities (see [13]). Each corner of the triangle represents one of the probability vectors (1,0,0), (0,1,0) and (0,0,1), while a general point in the triangle corresponds to the probability vector (p_1,p_2,p_3). Figures F and G give similar plots for the likelihood-ratio and Freeman-Tukey statistics. Figures F and G show that for $N_1 = N_2 = 8$, both G^2 and T^2 reject the null hypothesis more often than the nominal .05 level. Figure E for X^2 shows that the Pearson statistic does not reject too often and, in fact, taking the size for a composite null hypothesis as the maximum over the possible parameter values of the probability of rejection, the size is about .048. Table 5 gives the corresponding maxima for all statistics for $N_1 = N_2 = 4$, 6 8, 12, and 16 and for nominal levels .01, .05, and .10.

Table 5 goes about here

The results from Figures E, F, and G as well as Table 5 show that the Pearson statistic has exact sizes that are close to the nominal values for $N_1 = N_2 \geq 8$ and is conservative for smaller sample sizes. For the sample sizes considered the likelihood-ratio statistic has exact sizes that are 46 - 92% higher than the nominal .05; for the Freeman-Tukey statistic the exact sizes are 27 - 39% too high. A similar picture is found for other levels with the exception that the Freeman-Tukey statistic is conservative for small sample sizes at the nominal .01 level.

Comparing the above results with those given in the previous section, similar patterns are noted for each of the three statistics. The Pearson statistic attains exact levels that are close to the nominal levels for a wide range of sample sizes and parameter values. The Type I error rates of the approximate chi-square test based on the likelihood-ratio statistic are too high for both simple and composite null hypotheses with moderate cell expectations. For the Freeman-Tukey statistic it appears that the chi-square approximation works slightly better in the composite null case as opposed to the simple multinomial. This is true at least in comparing T^2 with the likelihood-ratio statistic G^2. However, T^2 does indeed have higher than nominal Type I error rates for both simple and composite null hypotheses.

In sum, the Pearson chi-square is the best statistic in terms of having Type I error rates that are closest to the nominal levels based on the asymptotic chi-square approximation. This result holds for simple multinomial null hypotheses as well as the more complex parametric Group Helping Model. The next section will offer an explanation of the results for the likelihood-ratio and Freeman-Tukey statistics.

4. EFFECTS OF VERY SMALL COUNTS ON THE CHI-SQUARE STATISTICS

In the last section it was concluded that for moderate minimum expected values, the Pearson statistic attained an exact level closer to the nominal than did either the likelihood-ratio or Freeman-Tukey statistics. The principal thesis of this section is that this discrepancy in behavior is due to the differing influence given to very small observed counts by the statistics.

Tables 6 and 7 go about here

Consider the case where a table has a cell with an observed value of 0. If the expected value is positive, then each of the three statistics X^2, G^2, and T^2 will give some weight to this discrepancy between the observed and expected cell frequencies. The minimum contributions to the three statistics caused by a zero count can be calculated; these minimum contributions are given in Table 6 for a range of expected cell frequencies. Thus, when a zero count appears with an expected cell frequency of 2.0, X^2 is at least 2.00, G^2 is at least 4.00, and T^2 is at least 4.00. The last column in Table 6 gives the probabilities of a zero observed value for a Poisson random variable with mean equal to the corresponding expected cell frequency. Under multinomial sampling with a large sample size, this is a good approximation to the probability of getting an observed zero in that cell. Table 7 gives similar minimum contributions and Poisson probabilities for the influence of an observed count of 1.

Looking at the case of a zero observed with expected cell frequency of 2.0, the corresponding Poisson probability of a zero count is 0.13534.

In the special case of a chi-square test with one degree of freedom, the nominal .05 level test would reject for statistic values in excess of 3.84146. Thus, both G^2 and T^2 would have exact levels somewhat in excess of 0.13534 compared to a nominal value of .05. For chi-square tests with more degrees of freedom, the effects are not so strong, but nonetheless, Table 6 and 7 illustrate the general pattern that very small observed counts increase G^2 and T^2 to a much greater extent than X^2. In addition, for expected cell frequencies in the range of 2.0 - 5.0, "zero" and "one" counts are very common occurrences as is shown by the Poisson approximation.

Figures H, I, and J go about here

To illustrate the above statements, Figures H, I, and J give the exact levels for the nominal .05 tests based on X^2, G^2, and T^2, respectively, for the case of a binomial null hypothesis with cell probabilities of (.05,.95). In each figure exact levels were computed for N = 10 (1) 100; thus, minimum cell expectations were .50 (.05) 5.00. Figure H for the Pearson statistic X^2 shows that the exact level fluctuates in a fairly regular pattern around the nominal value of .05. In contrast, looking at Figures I and J, G^2 has two fairly large increases in exact level at N = 38 and N = 87 while T^2 has large increases at N = 38 and N = 89. These are precisely the points at which zero and one counts become "significant at .05" for these statistics.

5. RECOMMENDATIONS AND CONCLUSIONS

For the case of a specified null hypothesis in the multinomial goodness-of-fit problem, the small sample distributions of three chi-square statistics were examined. Using as criterion the closeness of small sample distribution to the asymptotic chi-square approximation, the Pearson chi-square statistic is by far the most desirable. Both the likelihood-ratio and Freeman-Tukey statistics yield far too many rejections under the null distribution. In addition, a special model with a composite null hypothesis requiring estimation of two parameters was considered. The conclusions for this model were identical to those for the completely specified multinomial case. Given the diversity of the two models studied, it is clear that the statistic of choice, as far as null hypothesis behavior is concerned, is the Pearson chi-square statistic.

The high Type I Error rates for the likelihood-ratio and Freeman-Tukey statistics result from the large contributions to the chi-square value for very small counts in cells with moderate expected values. If a user desires to use one of these statistics in a case with moderate cell expectations, then he should be aware that the P-values based upon the nominal tests will, on the average, be somewhat understated.

REFERENCES

[1] Bahadur, R. R., "An Optimal Property of the Likelihood Ratio Statistic," Proceedings of Fifth Berkeley Symposium on Mathematical Statistics and Probability, Berkeley: University of California Press, 1967, 13 - 26.

[2] Bishop, Y. M. M., "Calculating Smoothed Contingency Tables," in J. Bunker, W. Forrest, Jr., F. Mosteller, and L. Vandam, ed., The National Halothane Study, National Institutes of Health, 1969.

[3] Bliss, C. I., Statistics in Biology, Volume 1, New York: McGraw-Hill Book Co., 1967.

[4] Cochran, W. G., "The χ^2 Test of Goodness of Fit," Annals of Mathematical Statistics, 23 (1952), 315 - 346.

[5] Cramer, H., Mathematical Methods of Statistics, Princeton, N. J.: Princeton University Press, 1945.

[6] Easterling, R. G. and Prairie, R. R., "Combining Component and System Information," Technometrics, 13 (1971), 271 - 280.

[7] Feller, W., An Introduction to Probability Theory and Its Applications, Volume 1, 2nd ed., New York: John Wiley & Sons, Inc., 1957.

[8] Fienberg, S. E. and Larntz, K., "Some Models for Individual-Group Comparisons and Group Behavior," Psychometrika, 36 (1971), 349 - 367.

[9] Fisher, R. A., Statistical Methods for Research Workers, 13th ed., New York: Hafner Publishing Company, 1958.

[10] Goodman, L. A., "Guided and Unguided Methods for Selecting Models for a Set of T Multidimensional Contingency Tables," Journal of the American Statistical Association, 68 (1973), 165 - 175.

[11] Hammersley, J. M. and Handscomb, D. C., Monte Carlo Methods, London: Methuen & Co., 1964.

[12] Hoeffding, W., "Asymptotically Optimal Tests for the Multinomial Distribution," Annals of Mathematical Statistics, 36 (1965), 369 - 401.

[13] Mosteller, F. and Tukey, J. W., "Data Analysis, Including Statistics," in G. Lindsey and E. Aronson, ed., Handbook of Social Psychology, Volume 2, Reading, Mass.: Addison-Wesley, 1968.

[14] Odoroff, E. L., "A Comparison of Minimum Logit Chi-Square Estimation and Maximum Likelihood Estimation in 2 x 2 x 2 and 3 x 2 x 2 Contingency Tables: Tests for Interactions," Journal of the American Statistical Association, 65 (1970), 1617 - 1631.

[15] Roscoe, J. T. and Byars, J. A., "Sample Size Restraints Commonly Imposed on the Use of the Chi-Square Statistic," Journal of the American Statistical Association, 66 (1971), 755 - 759.

[16] Staub, E., "A Child in Distress: The Influence of Age and Number of Witnesses on Children's Attempts to Help," Journal of Personality and Social Psychology, 14 (1970), 130 - 140.

[17] Sylwester, D., Abstract: "A Monte Carlo Study of Multidimensional Contingency Table Analysis," Biometrics, 30 (1974), 386.

[18] Tate, M. W. and Hyer, L. A., "Inaccuracy of the X^2 Test of Goodness of Fit When Expected Frequencies Are Small," Journal of The American Statistical Association, 68 (1973), 836 - 841.

[19] Yarnold, J. K., "The Minimum Expectation in X^2 Goodness of Fit Tests and the Accuracy of Approximations for the Null Distribution," Journal of the American Statistical Association, 65 (1970), 864 - 886.

1. OBSERVED TABLE FOR N_1 INDIVIDUALS AND N_2 GROUPS

Help Grade	Individuals	Groups	Total
I	n_{11}	n_{12}	R_1
II	n_{21}	n_{22}	R_2
III	n_{31}	n_{32}	R_3
Total	N_1	N_2	

2. CELL PROBABILITIES FOR MULTINOMIAL GOODNESS-OF-FIT COMPUTATIONS

---

k=2 cells

(.01,.99)	(.02,.98)	(.03,.97)
(.04,.96)	(.05,.95)	(.10,.90)
(.15,.85)	(.25,.75)	(.35,.65)
(.50,.50)		

---

k=3 cells

(.02,.02,.96)	(.02,.49,.49)	(.05,.05,.90)
(.05,.45,.45)	(.10,.10,.80)	(.10,.20,.70)
(.10,.30,.60)	(.10,.40,.50)	(.20,.20,.60)
(.20,.30,.50)	(.20,.40,.40)	(.30,.30,.40)
(1/3,1/3,1/3)		

---

k=5 cells

(.05,.05,.05,.05,.80)	(.05,.05,.05,.425,.425)
(.05,.05,.30,.30,.30)	(.05,.2375,.2375,.2375,.2375)
(.10,.10,.10,.10,.60)	(.10,.10,.10,.20,.50)
(.10,.10,.10,.30,.40)	(.10,.10,.20,.20,.40)
(.10,.10,.20,.30,.30)	(.10,.20,.20,.30,.30)
(.20,.20,.20,.20,.20)	

---

k=10 cells

(.05,.05,.05,.05,.05,.05,.05,.05,.05,.55)
(.05,.05,.05,.05,.05,.05,.05.,05.,30.,30)
(.05,.05,.05,.05,.05,.05,.05,.20,.20,.25)
(.05,.05,.05,.05,.05,.05,.15,.15,.20,.20)
(.05,.05,.05,.05,.05,.15,.15,.15,.15,.15)
(.05,.05,.05,.05,.10,.10,.15,.15,.15,.15)
(.05,.05,.05,.10,.10,.10,.10,.15,.15,.15)
(.05,.05,.10,.10,.10,.10,.10,.10,.15,.15)
(.05,.10,.10,.10,.10,.10,.10,.10,.10,.15)
(.10,.10,.10,.10,.10,.10,.10,.10,.10,.10)

---

3. MINIMUM AND MAXIMUM SIGNIFICANCE LEVELS FOR NOMINAL .05 TESTS FOR MULTINOMIAL GOODNESS-OF-FIT MODEL

Number of Cells	Min. Cell Expectation	X^2	G^2	T^2
2	.50-1.49	(.018,.105)	(.007,.043)	(.002,.026)
	1.50-1.99	(.020,.071)	(.011,.170)	(.007,.155)
	2.00-2.49	(.022,.065)	(.091,.148)	(.091,.148)
	2.50-2.99	(.020,.060)	(.062,.105)	(.059,.095)
	3.00-3.99	(.022,.086)	(.030,.079)	(.021,.068)
	4.00-4.99	(.025,.076)	(.026,.095)	(.023,.085)
	5.00-6.99	(.021,.071)	(.038,.109)	(.022,.079)
	7.00-9.99	(.031,.077)	(.031,.077)	(.031,.077)
	10.00-	(.031,.078)	(.035,.078)	(.034,.078)
3	.50- .99	(.027,.090)	(.012,.056)	(.001,.037)
	1.00-1.99	(.034,.074)	(.022,.090)	(.006,.068)
	2.00-2.49	(.030,.062)	(.040,.086)	(.033,.111)
	2.50-3.49	(.034,.060)	(.053,.114)	(.057,.139)
	3.50-4.99	(.038,.070)	(.044,.078)	(.031,.072)
	5.00-5.99	(.041,.061)	(.047,.074)	(.044,.072)
	6.00-9.99	(.041,.060)	(.039,.066)	(.039,.059)
	10.00-	(.044,.060)	(.046,.062)	(.041,.058)
5[a]	.50- .99	(.035,.095)	(.031,.060)	(.008,.043)
	1.00-1.49	(.033,.074)	(.021,.080)	).005,.075)
	1.50-2.49	(.031,.074)	(.038,.105)	(.038,.095)
	2.50-2.99	(.036,.071)	(.044,.082)	(.045,.085)
	3.00-3.99	(.031,.073)	(.042,.098)	(.045,.112)
	4.00-4.99	(.033,.068)	(.041,.098)	(.032,.092)
	5.00-	(.033,.068)	(.037,.074)	(.031,.071)
10[a]	.50- .99	(.036,.088)	(.012,.086)	(.000,.058)
	1.00-1.49	(.028,.072)	(.025,.101	(.000,.084)
	1.50-1.99	(.028,.071)	(.058,.107)	(.027,.093)
	2.00-2.99	(.029,.074)	(.051,.108)	(.045,.100)
	3.00-3.99	(.033,.071)	(.047,.094)	(043,.091)
	4.00-4.99	(.030,.071)	(.034,.096)	(.028,.099)
	5.00-	(.033,.063)	(.041,.075)	(.037,.074)

[a]Monte Carlo results based on 1000 trials per case.

4. INDIVIDUAL PROBABILITIES FOR GROUP HELPING MODEL COMPUTATIONS

(.1,.1,.8)	(.1,.2,.7)	(.1,.3,.6)	(.1,.4,.5)
(.1,.5,.4)	(.1,.6,.3)	(.1,.7,.2)	(.1,.8,.1)
(.2,.1,.7)	(.2,.2,.6)	(.2,.3,.5)	(.2,.4,.4)
(.2,.5,.3)	(.2,.6,.2)	(.2,.7,.1)	(.3,.1,.6)
(.3,.2,.5)	(.3,.3,.4)	(.3,.4,.3)	(.3,.5,.2)
(.3,.6,.1)	(.4,.1,.5)	(.4,.2,.4)	(.4,.3,.3)
(.4,.4,.2)	(.4,.5,.1)	(.5,.1,.4)	(.5,.2,.3)
(.5,.3,.2)	(.5,.4,.1)	(.6,.1,.3)	(.6,.2,.2)
(.6,.3,.1)	(.7,.1,.2)	(.7,.2,.1)	(.8,.1,.1)

5. MAXIMUM REJECTION RATES FOR THE GROUP HELPING MODEL

Nominal Level	$N_I=N_G$	X^2	G^2	T^2
.10	4	.0999	.1723	.1091
	6	.1159	.1711	.1269
	8	.1083	.1613	.1302
	12	.1063	.1528	.1278
	16	.1053	.1567	.1254
.05	4	.0351	.0802	.0675
	6	.0412	.0961	.0679
	8	.0483	.0907	.0695
	12	.0491	.0804	.0682
	16	.0494	.0729	.0636
.01	4	.0036	.0112	.0077
	6	.0054	.0170	.0145
	8	.0077	.0194	.0171
	12	.0088	.0186	.0172
	16	.0092	.0174	.0163

6. MINIMUM CONTRIBUTIONS TO CHI-SQUARE FOR AN OBSERVED COUNT OF ZERO

Cell Expectation	Minimum Contributions X^2	G^2	T^2	Prob(Zero Count) Under Poisson
1.0	1.00	2.00	1.53	.36788
1.5	1.50	3.00	2.71	.22313
2.0	2.00	4.00	4.00	.13534
2.5	2.50	5.00	5.37	.08208
3.0	3.00	6.00	6.79	.04979
3.5	3.50	7.00	8.25	.03020
4.0	4.00	8.00	9.75	.01832
4.5	4.50	9.00	11.28	.01111
5.0	5.00	10.00	12.83	.00674

7. MINIMUM CONTRIBUTIONS TO CHI-SQUARE FOR AN OBSERVED COUNT OF ONE

Cell Expectation	Minimum Contributions X^2	G^2	T^2	P(Zero or One Count) Under Poisson
2.0	0.50	0.61	0.34	.40601
2.5	0.90	1.17	0.81	.28730
3.0	1.33	1.80	1.42	.19915
3.5	1.79	2.49	2.13	.13589
4.0	2.25	3.23	2.92	.09158
4.5	2.72	3.99	3.78	.06110
5.0	3.20	4.78	4.70	.04043
5.5	3.68	5.59	5.67	.02657
6.0	4.17	6.42	6.69	.01735
6.5	4.65	7.26	7.74	.01127
7.0	5.14	8.11	8.83	.00730

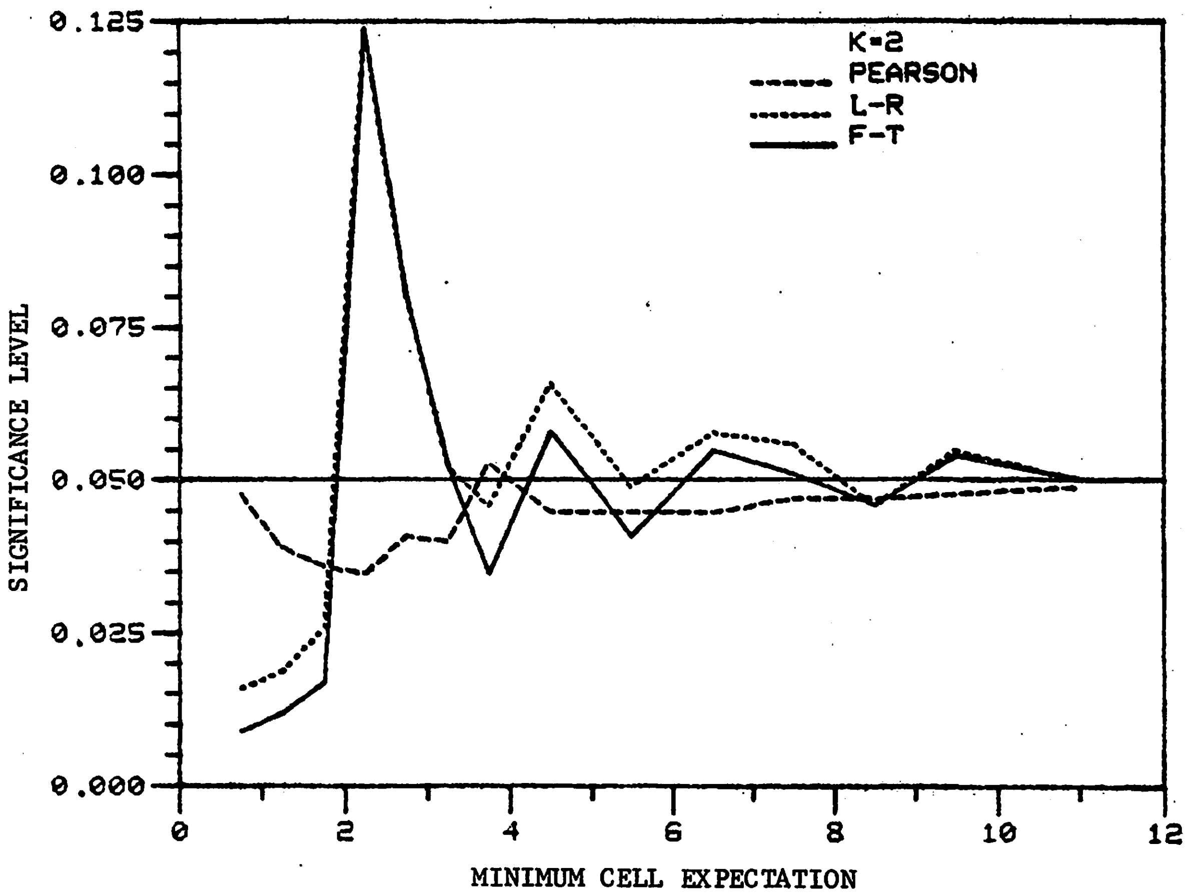

Figure A. Median Levels of Significance for Multinomial Null Hypotheses, 2 cells.

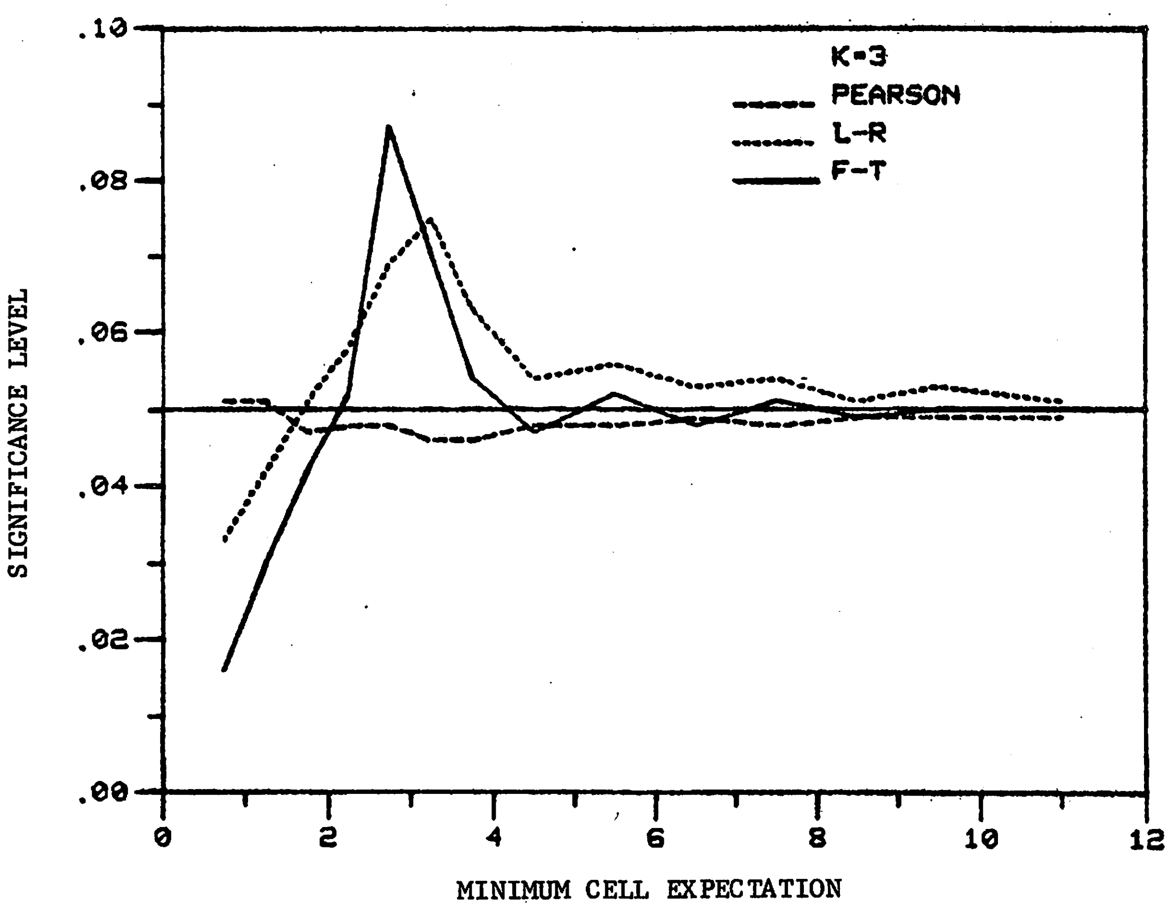

Figure B. Median Levels of Significance for Multinomial Null Hypotheses, 3 cells.

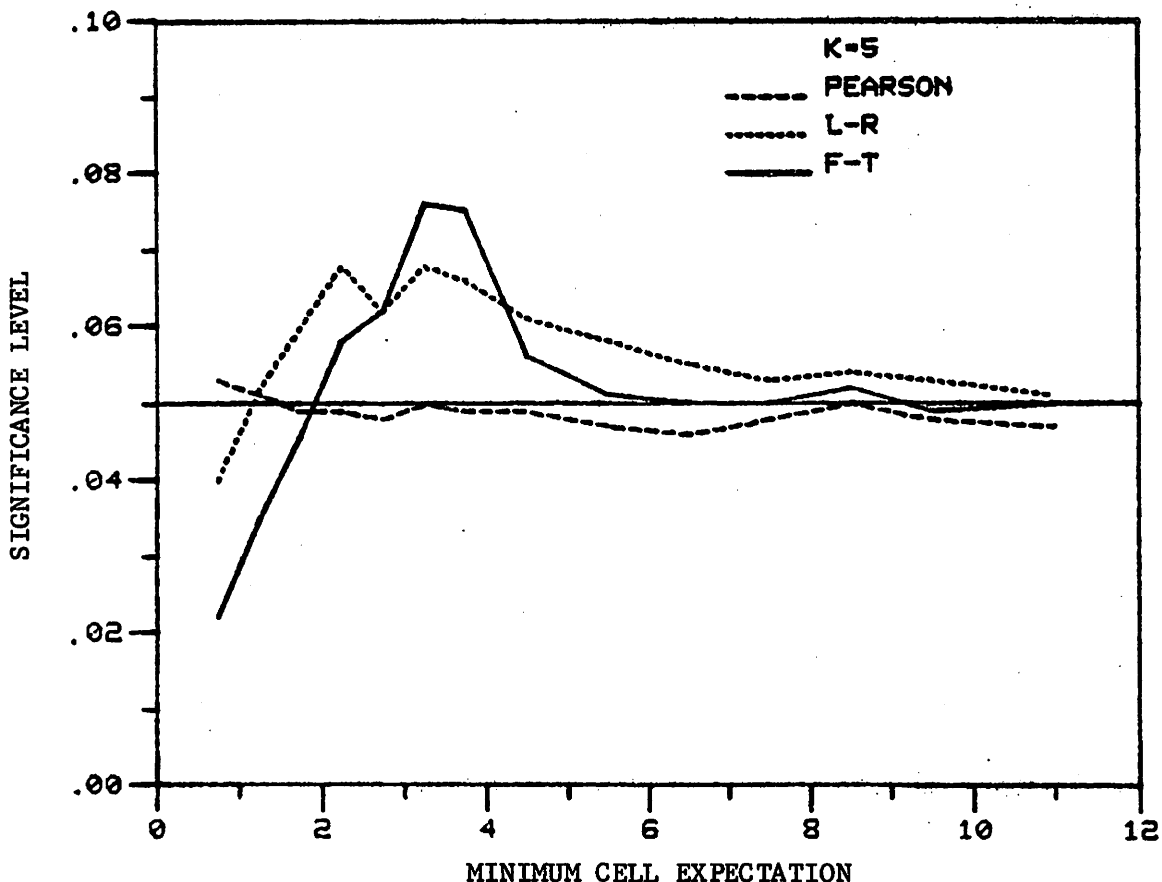

Figure C. Median Levels of Significance for Multinomial Null Hypotheses, 5 cells.

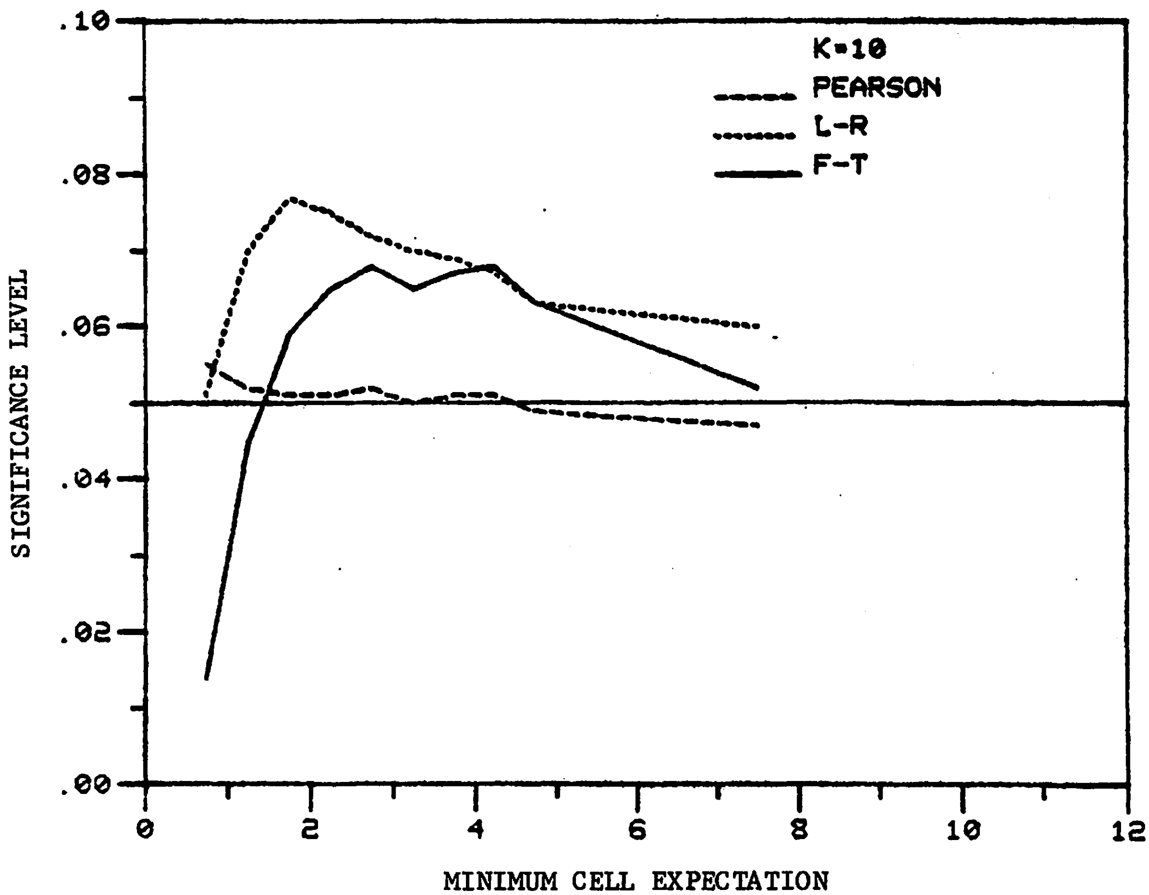

Figure D. Median Levels of Significance for Multinomial Null Hypotheses, 10 cells.

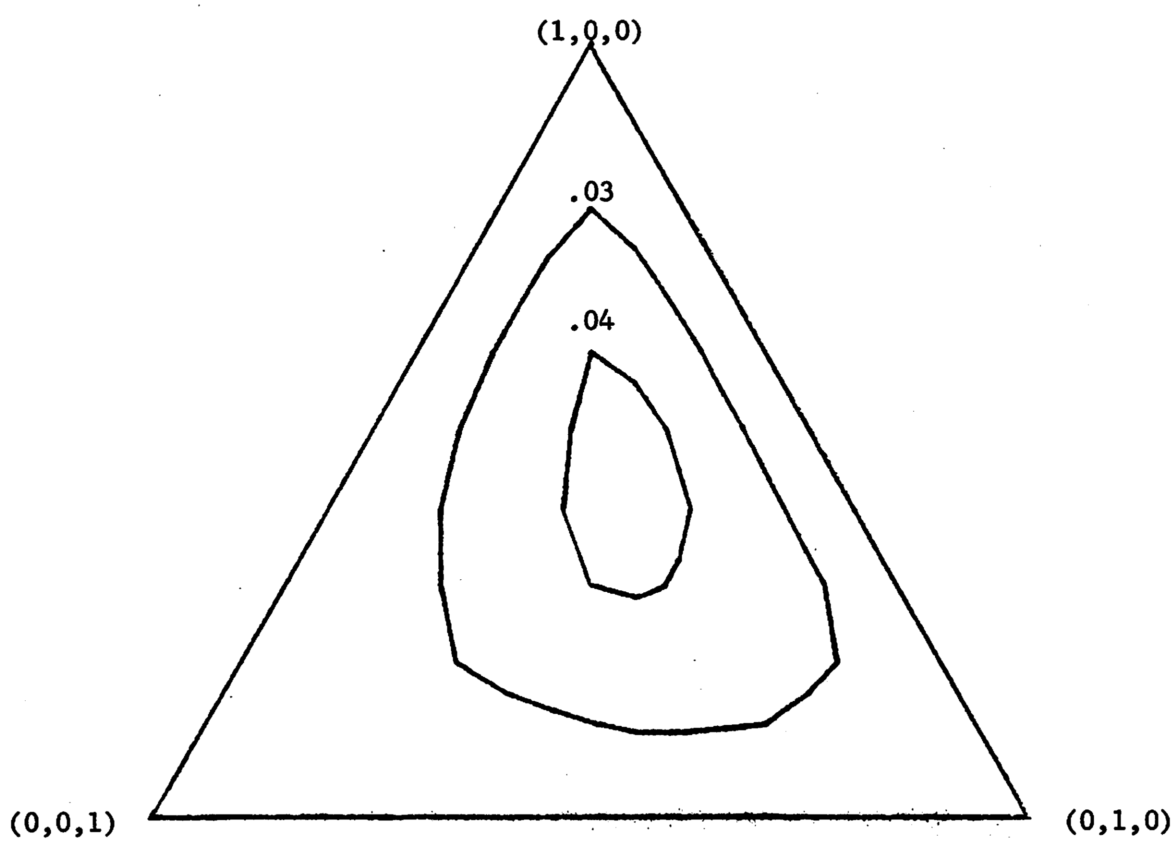

Figure E. Level of Significance of Pearson Chi-Square .05 Test for $N_I = N_G = 8$, Group Helping Model.[a]

[a]Contour plots drawn by computer program written by Daniel Laliberte under the supervision of Christopher Bingham, School of Statistics, University of Minnesota.

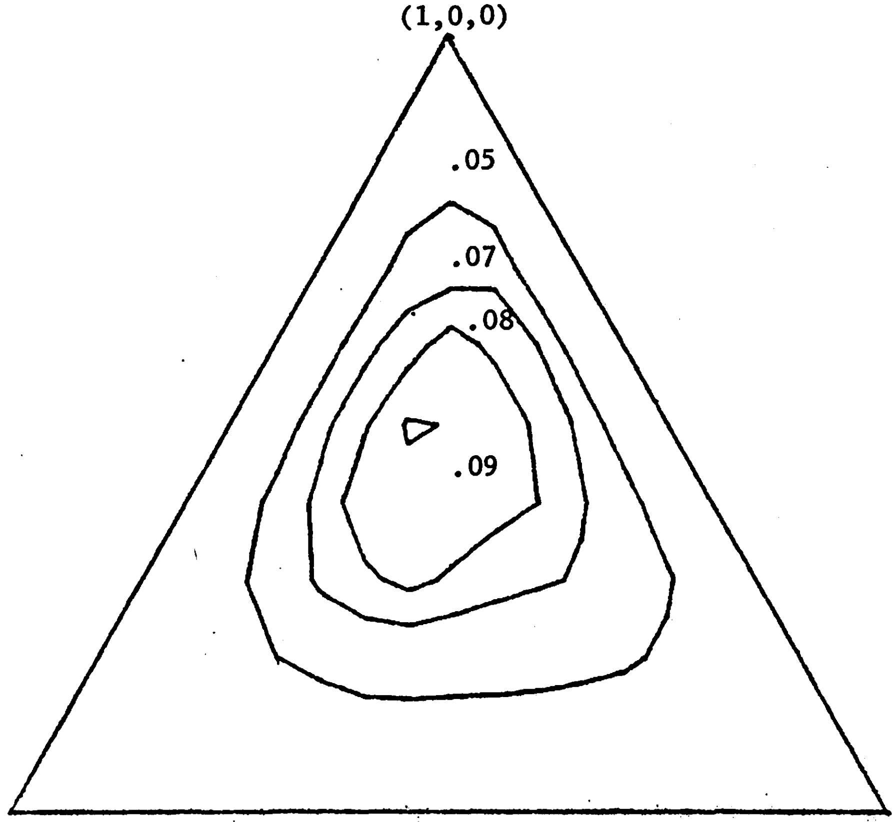

Figure F. Level of Significance of Likelihood Ratio Chi-Square .05 Test for $N_I = N_G = 8$, Group Helping Model.[a]

[a]See footnote a, Figure E.

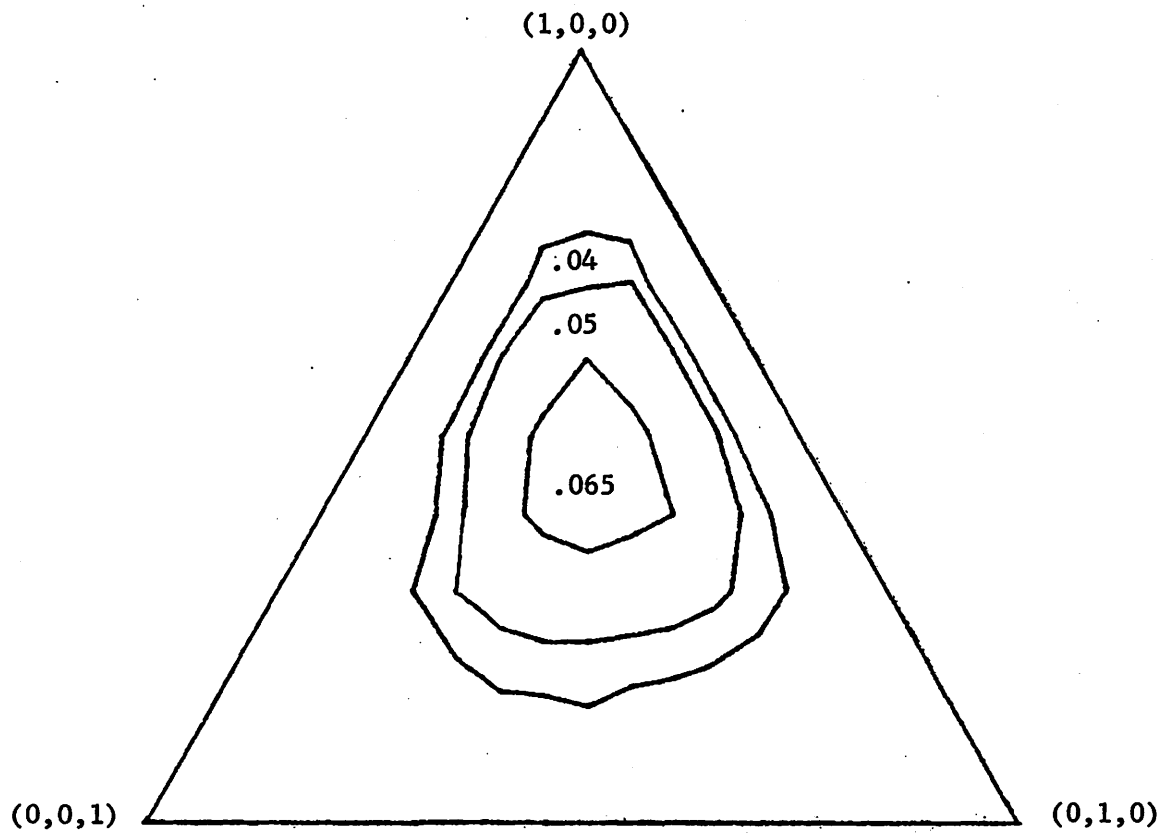

Figure G. Level of Significance of Freeman-Tukey Chi-Square .05 Test for $N_I = N_G = 8$, Group Helping Model.[a]

[a]See footnote a, Figure E.

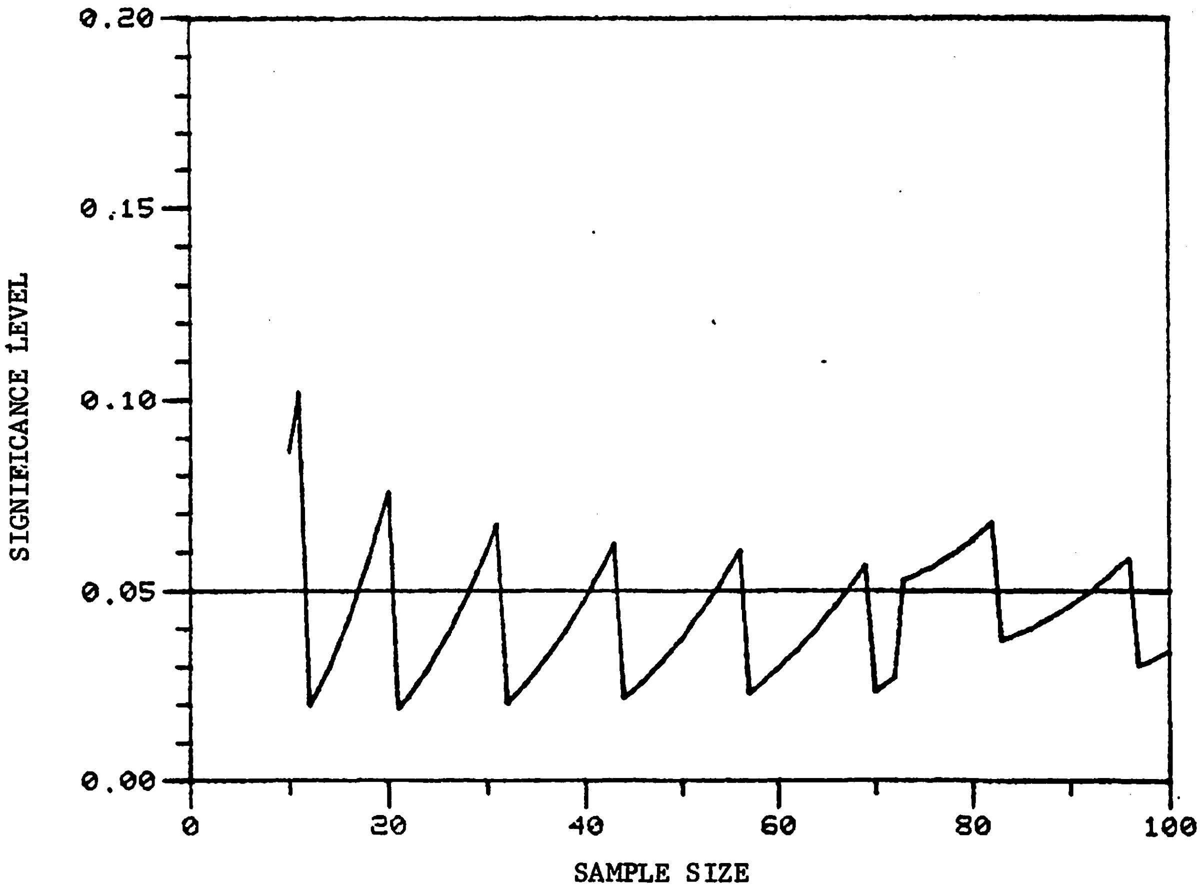

Figure H. Level of Significance for Pearson Chi-Square .05 Test, Binomial H_0: p = .05.

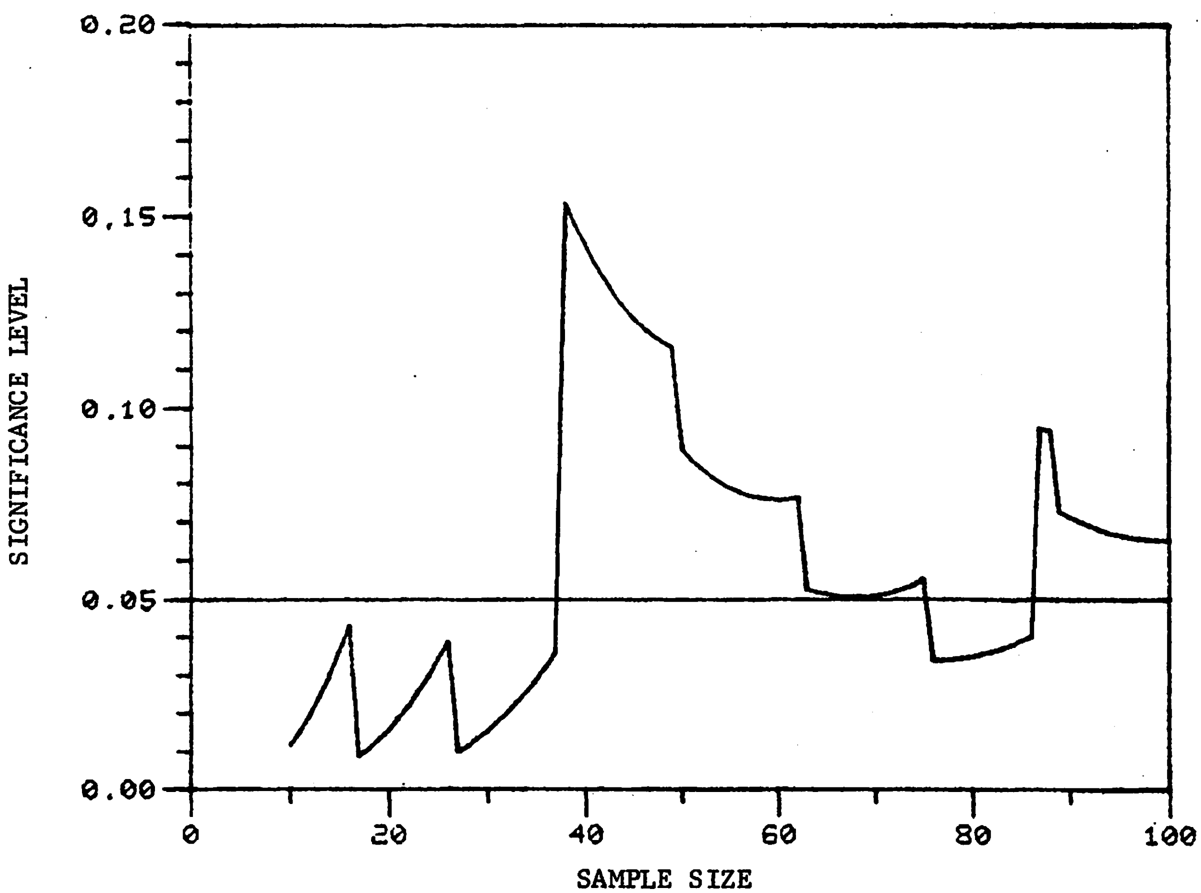

Figure I. Level of Significance for Likelihood Ratio Chi-Square .05 Test, Binomial H_0: p = .05.

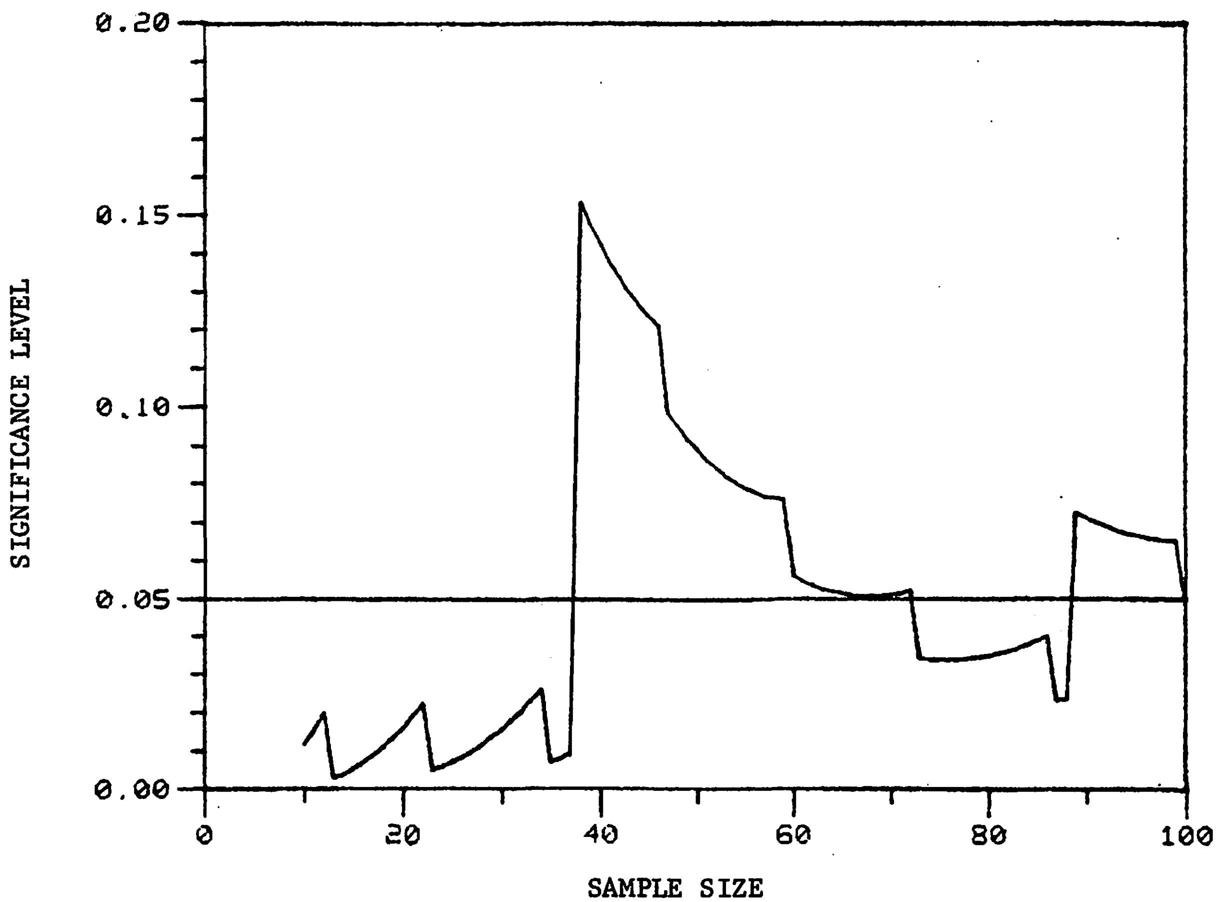

Figure J. Level of Significance for Freeman-Tukey Chi-Square .05 Test, Binomial H_0: p = .05.