THE ANALYSIS OF CONTINGENCY TABLES OF DATA FROM COMPLEX SAMPLES

Thomas J. Tomberlin
Harvard University[1]

1. Introduction

The analysis of data in contingency tables has traditionally been focused on tests of independence between variables using techniques which assume independent indentically distributed (iid) observations. In general, the iid assumption leads to multinomial and Poisson distributions of cell totals, and tests can be carried out using likelihood techniques. This may be reasonable for workers in biological, medical and other experimental fields (though even here it can sometimes be called into question), but for most users of data obtained from sample surveys, it is quite unrealistic. In the context of finite population sampling, the iid assumption is only reasonable for simple random samples from large populations. As most survey data are obtained from complex sample designs, alternative means of analyzing these data need to be developed.

When faced with this problem, many analysts have resorted to using the traditional tools, eg, Pearson's Chi-Squared Test for independence, and noting that the violation of the iid assumptions forces one to be extremely cautious about the validity of any conclusions drawn from their analyses. See for example Little (1978). The danger in using the traditional techniques of analysis when the iid assumptions are invalid because the data are drawn in clusters has been demonstrated empirically by Monte Carlo simulation techniques by Cowan and Binder (1978) and analytically by Fellegi (1978) and Scott (1978). A method for analyzing data from simple random samples of clusters of size two has been developed by Cohen (1976) and extended to clusters of arbitrary size k by Altham (1976). This work has not yet been extended to the more usual case (at least for sample survey data) of unequal probability, multi-stage selection of clusters of unequal sizes.

Recently, analytic techniques yielding more information than simple tests of hypotheses of independence have become available. For example, Grizzle, Starmer and Koch (1969) and Bishop Fienberg and Holland (1974) have developed methods utilizing techniques based on linear and log-linear models of cell probabilities. A wide range of hypotheses can be framed in terms of the parameters of these models. In particular, the log-linear model which is discussed in both of these works allows for traditional tests of hypotheses of independence, as well as more informative analyses under the same framework.

At this time, methods for testing hypotheses based on parameters of these models are dependent on the assumption of simple random sampling (or at most, stratified random sampling). Bishop, Fienberg and Holland (1974) for example, limit their discussions to the analysis of contingency tables where the cell totals follow Poisson, multinomial or product multinomial sampling distributions. However, in the more general framework of Grizzle, Starmer and Koch (1968), the tools for analysis of data obtained from complex sample designs are developed, but are not actually applied to this problem. These tools are further expanded by Koch, Freeman and Freeman (1975), Freeman and Koch (1976), Freeman, Freeman and Brock (1977), and Freeman, Freeman, Brock and Koch (1976). It seems that as yet, no one has applied these tools to the general problem of testing the fit of log-linear models (allowing for the usual tests of independence) to data from complex sample surveys. Fellegi (1978) and Scott (1978) consider similar techniques for specifically testing independence in an r × c classification and simple goodness of fit tests, but nothing more complex.

2. Framework for Analysis

Let us briefly consider the general model developed by Grizzle, Starmer and Koch (1969), and the analytic techniques appropriate for the model. As previously noted, this model contains, as a special case, the log-linear models discussed by Bishop, et al (1974) which allow for tests of independence.

Let $\underset{\sim}{\Pi}' = (\pi_1, \pi_2, \ldots, \pi_c)$ be a multidimensional table of cell classification probabilities strung out into a one dimensional vector, for the population. We assume that we have a sample drawn from this population from which we can estimate $\underset{\sim}{\Pi}$. About the sample design, we assume:

(1) The existence of a consistent estimator $\hat{\underset{\sim}{\Pi}}' = (\pi_1, \ldots, \pi_c)$ for $\underset{\sim}{\Pi}$.

(2) The existence of a consistent estimator for the sample covariance matrix of $\underset{\sim}{\Pi}$, $\hat{\underset{\sim}{V}}(\underset{\sim}{\Pi})$, say.

(3) The asymptotic multivariate normality of the estimator of $\underset{\sim}{\Pi}$.

Consider a family of functions of the form: $\underset{\sim}{F}(\underset{\sim}{\Pi}) = \underset{\sim}{K} \ln(\underset{\sim\sim}{A}\underset{\sim}{\Pi})$, where, $\underset{\sim}{\Pi}$ is defined as before, $\underset{\sim\sim}{A}$ and $\underset{\sim}{K}$ are u×c and t×u matrices of constants, respectively. By $\ln(\underset{\sim\sim}{A}\underset{\sim}{\Pi})$ we mean the vector of natural logarithms of the components of $\underset{\sim\sim}{A}\underset{\sim}{\Pi}$.

The αth element of the vector $\underset{\sim}{F}(\underset{\sim}{\Pi})$ is given by, $F_\alpha(\underset{\sim}{\Pi}) = \sum_{\gamma=1}^{u} K_{(\alpha,\gamma)} \ln\{\Sigma\, a_i \pi_i\}$, where $t \le u \le c$.

We can obtain a consistent estimator of $\underset{\sim}{F}(\underset{\sim}{\Pi})$ by substituting $\hat{\underset{\sim}{\Pi}}$ for $\underset{\sim}{\Pi}$ in the definition of $\underset{\sim}{F}(\underset{\sim}{\Pi})$ above. Using the Taylor Series expansion method, Grizzle et al, (1969) (or any of several other methods discussed in Kish and Frankel (1974) or in Kalton (1977)), we can estimate the sample covariance matrix, $\underset{\sim}{S}$, of $\underset{\sim}{F}(\hat{\underset{\sim}{\Pi}})$ using the estimator $\hat{\underset{\sim}{V}}(\hat{\underset{\sim}{\Pi}})$. That is, $\underset{\sim}{S} = \underset{\sim}{H}\hat{\underset{\sim}{V}}(\hat{\underset{\sim}{\Pi}})\underset{\sim}{H}'$ where the i,jth element of $\underset{\sim}{H}$ is

$$\left.\frac{\partial F_i \underset{\sim}{\Pi}}{\partial \pi_j}\right|_{\underset{\sim}{\Pi}=\hat{\underset{\sim}{\Pi}}}$$

Thus, if we can express the null hypothesis in the form $H_0: \underset{\sim}{F}(\hat{\underset{\sim}{\Pi}}) = 0$, then, under H_0, $\underset{\sim}{F}(\hat{\underset{\sim}{\Pi}})\, \underset{\sim}{S}^{-1}\, \underset{\sim}{F}'(\hat{\underset{\sim}{\Pi}})$ is asymptotically the same as that proposed by Wald (1943) and is approximately distributed as a $\chi^2_{(u)}$, where u is the rank of $\underset{\sim}{S}$.

For $\underset{\sim}{S}$ based on a small number of degrees of freedom, it seems reasonable that $\underset{\sim}{F}(\hat{\underset{\sim}{\pi}})\ \underset{\sim}{S}^{-1}\ \underset{\sim}{F}'(\hat{\underset{\sim}{\pi}})$ would be approximately distributed as an $F_{(u,df)}$, where df is the number of degrees of freedom used in estimating $\underset{\sim}{S}$.

For example, consider the case of testing for independence in a 2x2 table, with population cell probabilities given by:

π_{11}	π_{12}
π_{21}	π_{22}

The usual expression of the null hypothesis of independence is $H_o: \pi_{11}\pi_{22} = \pi_{12}\pi_{22}$. We can re-express this in the above format as $H_o \underset{\sim}{F}(\underset{\sim}{\pi}) = 0$, where $\underset{\sim}{A} = \underset{\sim}{I}_4$ and $\underset{\sim}{K} = (1, -1, -1, 1)$. It can be easily seen that in this case,

$$S = \sum_{ij,kl} \frac{K_{ij}\, K_{kl}\, \mathrm{Cov}(\hat{\pi}_{ij}, \hat{\pi}_{kl})}{\hat{\pi}_{ij}\, \hat{\pi}_{kl}},$$

so that, under the null hypothesis,

$$\underset{\sim}{F}(\hat{\underset{\sim}{\Pi}})\ \underset{\sim}{S}^{-1}\ \underset{\sim}{F}'(\hat{\underset{\sim}{\Pi}}) = \frac{(\ln \hat{\pi}_{11} + \ln \hat{\pi}_{22} - \ln \hat{\pi}_{12} - \ln \hat{\pi}_{21})^2}{S} \overset{\cdot}{\sim} \chi^2_{(1)},$$

or as $F_{(1,df)}$ if the number of degrees of freedom is small.

In the case of a 2X2 table, as shown here, the mathematics of the analysis are quite simple. This is not true for the general rxs table, and certainly not for a higher dimensional table. However, in terms of matrix algebra, the tests of independence, or in the language of loglinear modelers, tests for zero interactions, remain tests of linear combinations of logarithms of cell proportions. By using a computer for the calculations of test statistics, and specification of contrasts, this complexity need not be a limiting factor in the analysis.

3. An Example

The following example is based on data obtained from the Integration Test of the Canada Health Survey, which was conducted in June and July of 1977 in the provinces of Nova Scotia, New Brunswick and parts of Quebec. Briefly, the sample design consisted of a stratified, multi-stage selection of clusters of roughly equal size with equal probability within strata. A disproportionate allocation (square root) was used among provinces and regions in Quebec in order to achieve a certain degree of accuracy at these subnational levels. The total sample size was around 500 for the variables we will use. A more detailed discussion of the sample design can be found in Chinappa (1978).

For the purposes of this example, a simple formula for calculating variances and covariances for ratio estimates, taking into account sample differences between first stage units only, was utilized. Some strata with only one primary selection had to be collapsed for purposes of variance calculation, yielding six strata made up of 5, 3, 3, 4, 2, and 2 primary sampling units (PSU's) each. The sampling fractions (and thus, the sample weights) varied by a factor of ten, with the urban Quebec stratum having the smallest sampling fraction and the New Brunswick stratum the largest. Normally, the number of degrees of freedom for estimating variances is given by $\Sigma(n_h - 1)$, where n_h is the number of PSU's in the h^{th} stratum. However, because of unequal rates of sampling, this is in fact an over estimate of the "effective" number of degrees of freedom for this calculation. Cochran (1977) discusses a procedure due to Satterthwaite for estimating an effective number of degrees of freedom for this case. Assuming variability within strata is relatively constant among strata, and that the sampling fractions are all very small, using the Satterthwaite procedure, we arrive at an effective number of degrees of freedom of 3, a reduction from $\Sigma(n_h - 1) = 13$ which we would have had if proportionate allocation had been used.

Three variables are used in this example, each being an indicator of hypertension. They are:
INTERVIEW: A response to an interviewer administered questionnaire, possibly a proxy response. One person responds for the whole household.
PHYSICAL: A physical measure of blood pressure, yielding a classification into the presence or absence of hypertension.
PERSONAL: A personal response to a self-administered qestionnaire.
The unweighted sample totals and proportions for those who responded to all three questions are given in Table I. Weighted estimates of the same are given in Table II. These variables were chosen, not so much for their substantive interest as to demonstrate the use of the proposed analytical procedures. There should be no doubt that the three variables are dependent, and with a large enough sample, this dependence could be detected.

First, using the ratio estimates in Table II, and estimates of their sample variances and covariances, we can proceed with fitting a log-linear model as discussed in Section 2. If we string out the 2X2X2 table row by row as found in Tables I and II, we can test the three factor interaction term by setting $\underset{\sim}{A} = \underset{\sim}{I}_8$ and $\underset{\sim}{K} = (-1, 1, 1, -1, 1, -1, -1, 1)$, which yields, $F(\hat{\underset{\sim}{\Pi}}) = -2.428$, with a sample variance of $S = 1.381$, and a Wald statistic of 4.271. Here, F is the linear contrast of logs of the cell proportions appropriate for testing the three factor interaction, and can be obtained using the usual techniques of analyzing the analysis of variance type models. Following the framework set up in Section 2, we would expect this statistic to be distributed as an $F_{(1,3)}$ which has a 95% cut off point of 7.71, so the three factor interaction is not significant. If we conduct the same analysis assuming a simple random sample of size 476, we get a Chi-squared statistic of 3.44, which is just shy of the 95% point of 3.84. In either case, we would accept the null hypothesis of no three factor interaction at a .05 significance level. Table III contains a summary of a series of hierarchical tests using the proposed framework and the traditional "naive approach".

In fact, inferences based on the two approaches are not drastically different. That is probably because there is not a dramatic intra-class correlation for any of these variables. In fact, the average of the estimated design effects turns out to be slightly less than one, so we

would not expect there to be drastic differences. However, even here, testing at a .05 level, the traditional method gives two significant two factor interactions where the proposed method would allow collapsing over the INTERVIEW response.

4. An Empirical Study

One drawback of this method is that it requires the calculation of a variance-covariance matrix for cell proportion estimates. For small tables, this need not be a problem. For larger tables, as the number of variances and covariances increases with the square of the number of cells, which itself increases with the product of the numbers of categories in each classification variable, this matrix can quickly get out of hand.

Fellegi (1978) suggests an heurestically reasonable statistic for testing independence in an rxc classification which utilizes an "average design effect", and might be helpful in this situation.[2] The statistic is:

$$X = 1/\bar{b} \sum\sum_{ij} \frac{(\hat{\Pi}_{ij} - \hat{\Pi}_{i\cdot}\,\hat{\Pi}_{\cdot j})^2}{\hat{\Pi}_{i\cdot}\,\hat{\Pi}_{\cdot j}} \quad \text{, where}$$

$$\bar{b} = \frac{n}{rc} \sum_i \sum_j \frac{\mathrm{var}(\hat{\Pi}_{ij})}{\hat{\Pi}_{ij}\,(1-\hat{\Pi}_{ij})} \quad .$$

This is nothing more than the usual Pearson's Chi-Squared statistic divided by an average of the estimated design effects, averaging over all cells in the table. It is conjectured that X should follow a distribution which could reasonably be approximated by an ordinary χ^2 distribution with appropriate degrees of freedom. This statistic seemed to perform reasonably well in his empirical studies as well as in those presented in this paper. It does not perform as well as the Wald statistic previously described, and should only be used when this statistic is unavailable.

The example in Section 2 used to illustrate the technique is not especially useful for pointing out the difference between the various proposed techniques since the design effects for the various cells are not large. In fact, as previously stated, the average design effect (Fellegi's $\bar{b}$) is slightly less than one. (Recall, the variance estimates are based on only approximately 3 degrees of freedom.) In order to compare Fellegi's proposed X and the Wald statistic, and also to see if it is reasonable to use an F-distribution for these statistics in cases where the design effects are of some consequence, a small Monte Carlo simulation study was undertaken.

The simulation study was restricted to a two-stage equal probability sample of equal size clusters with equal size subsamples. The analysis was restricted to a test of independence of two variables in a 2x2 classification. The data were generated in such a way as to have expected values of cell proportions equal to various 2x2 tables having independent classification variables. The models thus satisfied the null hypotheses of independence so that the empirical distributions of the statistics could be compared with the theoretical null χ^2 and F distributions with the appropriate degrees of freedom.

In order to simulate a design effect, a distribution was imposed on selected cluster classifaction probabilities having expected values the Π_{ij}'s satisfying the null hypothesis. The obvious multivariate distribution to impose on these p_{ij}'s is a Dirichlet distribution (see Johnson and Kotz (1970).) The parameters of this distribution, together with the size of the cluster subsample determine the intraclass correlation, and thus the design effects for estimating cell proportions. This allowed for relatively easy manipulation of the characteristics studied but also resulted in a serious simplification when compared with many realistic sample designs. The intraclass correlations for all variances and covariances are equal for all cells in the table. This certainly need not be the case in an actual population. A summary of some of the results, organized by the number of sampled clusters, is presented in Table IV. These figures are based on 1000 replications.

The three tables of expected cell proportions are shown at the top of Tables A, B, and C. There were four independent simulation experiments for each of the three sample designs 1, 2, and 3. The sample design is described at the top of each of the sub-tables.

For each experiment, both the Wald statistic $\underset{\sim}{F}' \underset{\sim}{S}^{-1} \underset{\sim}{F}$, and Fellegi's X were computed, and the proportion of observations larger than some of the usual critical values of the appropriate χ^2 and F distributions are reported as the empirically observed significance levels corresponding to the nominal levels given at the top of each sub-table.

The results for design 1 are not terribly interesting since we have a large sample size (2000), and a large number of degrees of freedom for variance estimates (199), so that the χ^2 and F distribution critical values are very close. The Wald statistic with F distribution critical values with a few exceptions seems to match the nominal levels better than the same with χ^2 critcal vlaues, but there does not seem to be much to choose between X and the Wald statistic. There are no obvious trends with regards to small cell sizes or different design effects.

Results for design 2 indicate that we can dismiss the use of χ^2 cutoff points in favor of F distribution cutoff points for the Wald statistic which performs rather well except for table C. With regards to Fellegi's X, we can dismiss the use of F distribution cutoff points. For the most part, the Wald statistic out-performs Fellegi's X, though again only marginally. Both statistics have trouble with table C, because of its small expected cell sizes. It should be said here that in the case where the simulation returned a sample table with an empty cell, no analysis was carried out. This seemed to match usual practice, but obviously results in biased estimates of cell proportions.

In design 3, the Wald statistic with F distribution cutoff points again performs reasonably well except for table C, and outperforms Fellegi's X fairly consistently. Fellegi's X tends to have actual significance levels larger than the nominal ones, and in all cases, larger than those of the Wald statistic.

In summary, this small study indicaties that when available, the Wald statistic, using cutoff points from the F distribution should be the preferred procedure. When this statistic is unavailable, Fellegi's X with χ^2 cutoff points performs surprizingly well. Again, it should be emphasized

that this study tests Fellegi's X under most favorable conditions. If there were some variability among the individual cell design effects, one would expect its performance to suffer. In any case, one should avoid this type of analysis when expected cell sizes are small.

5. Conclusions and Discussion

It can now be stated that one no longer need use simple random sampling assumptions for analysing categorical data obtained from complex samples. Goodness of fit test, log-linear model significance tests, and in general, any test whose null hypothesis can be formulated in terms of a linear contrast of cell proportions or logs of cell proportions can be analysed using techniques suggested by Grizzle, et al. These techniques require that sample estimates of sample variances and covariances of the estimated cell proportions be available. In the event that such estimates are not available, either for lack of information or because of the size of the tables, reasonable procedures can be based on average design effects as suggested by Fellegi (1978) and modified by Rao and Scott (1979). If design effects are all close to one, ordinary simple random sample procedures are probably good enough.

With regards to research organizations, governments and other agencies responsible for producing data from surveys, two points seem relevant. First, in order to carry out proper analysis of contingency tables, it is necessary to have sample variances and if possible covariances. Since in order to compute these estimates, one must have PSU level data, this means that these agencies should either be prepared to release more data to data analysts (and deal with the confidentiality problems that this creates) or become more heavily involved in the analysis of such data.

It should also be noted that the precision of variance estimates has a large effect on inferential analysis. If this type of analysis is a primary goal of a survey, then providing adequately precise variance estimates should be considered when designing the sample. This is all too often ignored.

A last, more serious problem which still requires research is that of estimating cell proportions for reduced models. Bishop, et al, discuss the problem of predicting cell proportions once a reduced model has been fitted. Assuming a simple random sample, one can use an iterative proportionate fitting algorithm to get maximum likelihood estimates for cell proportions under the reduced model. These can be useful for "smoothing" cell estimates and reducing sampling variance by in effect pooling information across cells. At this point there does not seem to be a way of doing this for data from complex samples.

Footnotes

1. This work was, for the most part, carried out while the author was employed by Statistics Canada.
2. Rao and Scott (1979) in a paper to be presented to the ASA in August 1979, give a theoretical justification for a similar statistic which is the same in the case that all cell frequencies and/or all cell design effects are equal. The latter is the case for the empirical study which follows.

References

1. Altham, P.M.E. (1976), "Discrete variable analysis for individuals grouped into families", Biometrika, 63, 263-369.
2. Bishop, Fienberg, and Holland (1974), Discrete Multivariate Analysis, Theory and Practice, Cambridge, Mass:MIT Press.
3. Chinappa, Nanjamma, "The design of the Integration Test of the Canada Health Survey", unpublished Statistics Canada memorandum. Institutional and Agricultural Survey Methods Division, Statistics Canada, Ottawa, 1978.
4. Cochran, W. (1977), Sampling Techniques, 3rd ed., New York:John Wiley and Sons.
5. Cohen, J.E. (1976), "The distribution of the chi-squared statistic under clustered sampling from contingency tables", J. Amer. Statist. Assoc., 71, 665-670.
6. Cowan, J. and Binder, D. (1978), "The effect of a two-stage sample design on tests of independence of a 2x2 table", Survey Methodology, 4.
7. Fellegi, I. (1978), "Approximate tests of independence and goodness of fit based on stratified multi-stage samples", Survey Methodology, 4, 29-56.
8. Freeman, D. and Koch, G. (1976), "An asymptotic covariance structure for testing hypotheses on raked contingency tables from complex surveys", Amer. Statist. Assoc., Proceedings of the Social Statistics Section, Boston.
9. Freeman, Freeman and Brock (1977), "Modularisation for the analysis of complex sample survey data", Invited Paper, Int. Assoc. of Surv. Statist., ISI meetings, New Delhi.
10. Freeman, Freeman, Brock and Koch (1976), "Strategies in the multivariate analysis of data from complex surveys II", Int. Statist. Rev. 44, 317-330.
11. Grizzle, Starmer and Koch (1969), "Analysis of categorical data by linear models", Biometrics, 489-504.
12. Johnson, N.L. and Kotz, S. (1970), Continuous Multivariate Distributions, Boston:Houghton Mifflin Company.
13. Kalton, G. (1977), "Practical Methods for Estimating Sampling Errors", Proceedings of the ISI, New Delhi.
14. Kish, L. and Frankel, M. (1974), "Inference from complex samples", J.R. Statist. Soc. b, 36, 1-37.
15. Koch, G., Freeman, D. and Freeman, J. (1975), "Strategies in the Multivariate analysis of data from complex surveys", Int. Statist. Rev., 43, 59-78.
16. Little, R.J.A. (1978), "Generalized linear models for cross-classified data from the WFS", World Fertility Survey Technical Bulletin No. 5/Tech. 834., London.
17. Nathan, G. (1975), "Tests of independence in contingency tables from stratified proportional samples", Sankhya C, 37, 77-87.
18. Rao, J.N.K. and Scott, A.J. (1979), "Chi-Square Tests for Analysis of Categorical Data from Complex Surveys", Amer. Stat. Assoc., Proceedings of the Section on Survey Research Methods, Washington, D.C.
19. Scott, A.J. (1978), "Statistical Analysis of data from complex samples", Invited talk at the European Conference of Statisticians, Oslo.
20. Wald, A. (1943), "Tests of statistical hypotheses concerning several parameters when the number of observations is large", Transactions of the American Mathematical Society, 54, 426-482.

TABLE I

Unweighted sample totals and proportions of a classification of three separate measures of hypertension. (Absent and present refer to the absence or presence of hypertension.)

Physical	Personal	INTERVIEW Absent	INTERVIEW Present	
Absent	Absent	.838 (399)	.004 (2)	401
	Present	.053 (25)	.034 (16)	41
		424	18	442
		Absent	Present	
Present	Absent	.032 (15)	.002 (1)	16
	Present	.013 (6)	.025 (12)	18
		21	13	34

TABLE II

Weighted estimates of cell proportions of a cross classification of three separate measures of hypertension.

Physical	Personal	INTERVIEW Absent	INTERVIEW Present
Absent	Absent	.851	.003
	Present	.052	.029
		Absent	Present
Present	Absent	.024	.004
	Present	.012	.025

TABLE III

Hierarachical tests of parameters of the log-linear model for data given in Table II.

Parameter	Test Statistic: Wald Statistic (null distribution)	Test Statistic: SRS Statistic (null distribution)
Three way interaction	4.271 ($F_{1,3}$) NS	3.44 (χ^2_1) *
all three conditional two way interactions	58.384 ($F_{3,3}$) ***	90.130 (χ^2_3) ***
conditional Physical-Interview interactions	1.298 ($F_{1,3}$) NS	1.864 (χ^2_1) NS
conditional Physical-Personal interactions	24.818 ($F_{1,3}$) **	33.952 (χ^2_1) ***
conditional Interview-Personal interactions	7.294 ($F_{1,3}$) *	14.441 (χ^2_1) ***
after collapsing over the Interview variable Physical-Personal	35.894 ($F_{1,3}$) ***	59.772 (χ^2_1) ***

NS indicates not significant
* .10 level significance
** .05 level significance
*** .01 significance

TABLE IV

EMPIRICAL RESULTS

1000 Replicates

DESIGN 1. 100 clusters of size 100 selected with subsamples of size 20, total sample size of 2000.

DESIGN 2. 10 clusters of size 100 selected with subsamples of size 20, total sample size of 200.

DESIGN 3. 5 clusters of size 100 selected with subsamples of size 20, total sample size of 100.

Table A

.25	.25
.25	.25

Table B

.0625	.1875
.1875	.5625

Table C

.01	.09
.09	.81

DESIGN 1

Table	Intra-Class Correlation	Deff	Statistic	Distribution	Nominal Levels .10	.05	.025	.01	.005
A	.3266	7.2281	Wald	$F^{\chi^2(1)}_{(1,99)}$	.096	.053	.023	.011	.005
				$F_{(1,99)}$	.092	.050	.021	.010	.004
			X	$F^{\chi^2(1)}_{(1,99)}$	.098	.054	.025	.010	.005
				$F_{(1,99)}$	.096	.050	.026	.010	.004
A	.102132	2.97076	Wald	$F^{\chi^2(1)}_{(1,99)}$	.119	.060	.030	.017	.008
				$F_{(1,99)}$	.113	.055	.028	.017	.006
			X	$F^{\chi^2(1)}_{(1,99)}$	.119	.057	.028	.011	.008
				$F_{(1,99)}$	.113	.053	.025	.011	.006
B	.102132	2.97076	Wald	$F^{\chi^2(1)}_{(1,99)}$	.117	.059	.027	.009	.003
				$F_{(1,99)}$	.111	.055	.022	.009	.003
			X	$F^{\chi^2(1)}_{(1,99)}$	.110	.053	.027	.009	.003
				$F_{(1,99)}$	.102	.049	.027	.008	.003
C	.00970	1.21775	Wald	$F^{\chi^2(1)}_{(1,99)}$	.098	.046	.023	.011	.003
				$F_{(1,99)}$	.095	.044	.023	.011	.003
			X	$F^{\chi^2(1)}_{(1,99)}$	.094	.048	.025	.009	.006
				$F_{(1,99)}$	.090	.045	.024	.009	.006

DESIGN 2

Statistic	Distribution	Nominal Levels .10	.05	.025	.01	.005
Wald	$F^{\chi^2(1)}_{(1,9)}$	.116	.081	.053	.034	.024
	$F_{(1,9)}$	.100	.051	.031	.011	.005
X	$F^{\chi^2(1)}_{(1,9)}$	.116	.063	.038	.018	.010
	$F_{(1,9)}$	.078	.035	.013	.002	.001
Wald	$F^{\chi^2(1)}_{(1,9)}$	.111	.079	.052	.025	.015
	$F_{(1,9)}$	.095	.050	.017	.006	.003
X	$F^{\chi^2(1)}_{(1,9)}$	.098	.048	.022	.007	.005
	$F_{(1,9)}$	.071	.020	.007	.002	.001
Wald	$F^{\chi^2(1)}_{(1,9)}$	.138	.076	.045	.029	.019
	$F_{(1,9)}$	.096	.043	.024	.009	.009
X	$F^{\chi^2(1)}_{(1,9)}$	.092	.056	.029	.012	.007
	$F_{(1,9)}$	.069	.025	.009	.003	.001
Wald	$F^{\chi^2(1)}_{(1,9)}$	.081	.052	.026	.018	.013
	$F_{(1,9)}$	.059	.026	.016	.007	.005
X	$F^{\chi^2(1)}_{(1,9)}$	.077	.046	.027	.017	.016
	$F_{(1,9)}$	.060	.026	.016	.007	.004

DESIGN 3

Statistic	Distribution	Nominal Levels .10	.05	.025	.01	.005
Wald	$F^{\chi^2(1)}_{(1,4)}$	.199	.150	.112	.082	.069
	$F_{(1,4)}$	.123	.070	.035	.010	.005
X	$F^{\chi^2(1)}_{(1,4)}$	.125	.083	.055	.023	.014
	$F_{(1,4)}$	.065	.015	.003	.001	-
Wald	$F^{\chi^2(1)}_{(1,4)}$	.161	.115	.083	.061	.051
	$F_{(1,4)}$	.095	.053	.026	.005	.001
X	$F^{\chi^2(1)}_{(1,4)}$	.109	.062	.037	.021	.011
	$F_{(1,4)}$	.043	.013	.002	.001	-
Wald	$F^{\chi^2(1)}_{(1,4)}$	.197	.135	.100	.073	.057
	$F_{(1,4)}$	.113	.059	.028	.008	.004
X	$F^{\chi^2(1)}_{(1,4)}$	.117	.070	.040	.017	.010
	$F_{(1,4)}$	.049	.010	.004	-	-
Wald	$F^{\chi^2(1)}_{(1,4)}$	.134	.089	.062	.042	.033
	$F_{(1,4)}$	.071	.033	.018	.004	.003
X	$F^{\chi^2(1)}_{(1,4)}$	.136	.085	.060	.029	.018
	$F_{(1,4)}$	.066	.018	.004	-	-