TESTING FOR CONTEXTUAL EFFECT

C. H. Proctor, North Carolina State University

1. Introduction

In a pair of recent papers appearing in the journal Sociological Methods and Research (Farkas, 1974 and Hauser, 1974) the two authors disagreed on the attribution of belief in a, so called, contextual effect for a given set of data. Farkas decided it was there, and Hauser maintained it was not established. They said much more and, in fact, the two papers constitute an excellent review of contextual effect analysis. Hauser admitted that "...the effects are statistically significant at the usual α-levels, since the sample has about 5,300 cases," (p.368) and then went on to argue that statistical significance is not practical significance. However, that t-value was around 10, and I feel that such statistical significance as $t = 10$ exhibits (even with 5,300 cases) almost has to be of practical significance as well. The largeness of this t-value, however, may be due to their use of an inappropriate model equation. In short, although there were 5,300 students, they came from only 99 colleges and estimation of a context effect as a contrast on college means may better be viewed as based on a sample size of 99 than of 5,300.

The following discussion is an attempt to summarize some current methods available for testing for contextual effect. The first two are truly rough and ready approaches. The first one struggles to obtain balanced data by deleting and duplicating at random, while the other is a version of the celebrated Tukey Jackknife. The next two use variance component models and exhibit the rather involved considerations that seem to arise when estimating components of variance with unbalanced data along with testing a regression coefficient in a mixed model. One of these latter two estimates the variance components by equating mean squares to their expectations and then tests by use of generalized least squares, while the final method is simply likelihood all the way. The situation is an instructive example of a data handling problem that is quite common in social survey analysis and although no new methods are here proposed, it may be useful to see what already-existing statistical practice seems to suggest can be done.

2. Model Equations

In a concrete example treated in the two papers, scores were given for his answers on a questionnaire to each student on his attitude toward drinking (the independent or X-variable) and on the extent of his drinking behavior (the dependent or Y-variable). As one would expect, these two measures were rather closely related. The question posed was whether the between colleges regression slope of behavior on attitude was equal to the within colleges slope. There is suspicion that the between colleges slope may be steeper due to some cumulative, reasonance interpersonal phenomena as described by (Farkas,1974, p.339ff). This is called contextual effect. A difference in slopes may be due to enumerator or sample selection biases or to omitted variable effects or to many other causes but the data handling methods would still be called testing for contextual effect.

It is by no means uncommon to do regression analysis in the presence of a cluster or nested sample design. One simply removes the cluster or "block" effects (schools, counties, area segments or whatever) and operates upon the within cluster cross products with perhaps a pious expression of hope that there are no interactions between slopes and clusters (at any rate I have done so in S-63, 1974, p.85). In the present case of being interested in the contextual effect there is no way to avoid the possibly upsetting presence of confounded college effects. To see how they enter let's consider model equations.

Regression calculations would be governed by a model equation such as·

$$(1)\quad y_{ij} = \mu' + x_{ij}\beta + \bar{x}_{i.}\delta + e_{ij}, \text{or (redefining } \mu)$$
$$= \mu + (x_{ij} - \bar{x}_{..})\beta + (\bar{x}_{i.} - \bar{x}_{..})\delta + e_{ij},$$

where y_{ij} is the score on his drinking behavior reported by the j^{th} student among n_i at the i^{th} college where $i = 1,2,\ldots,t$. His attitude score is written x_{ij}; the college mean is written $\bar{x}_{i.}$ and the grand mean is denoted $\bar{x}_{..}$. The e_{ij} quantities ($n = \sum^t n_i$ of them) are taken to be independent, identically distributed with variance σ_e^2 . The objective of the analysis is to test H_0: $\delta = 0$.

I suspect that including a term to be written u, where the u_i's $i = 1,2,\ldots,t$ are independent, identically distributed with variance σ_u^2 , would allow the model to fit more closely to the actual data. This issue should be decided by examining the data although one's judgement based on past experiences with similar situations should enter also. The model equation then becomes

$$(2)\quad y_{ij} = \mu + (x_{ij} - \bar{x}_{..})\beta + (\bar{x}_{i.} - \bar{x}_{..})\delta + u_i + e_{ij} .$$

Suppose that the data have already been cleaned (of outliers, for example) and possibly some transformations made to assure linearity so that least squares estimation is not too far from optimal. There are two, somewhat idealistic, circumstances in which the test of $\delta = 0$ becomes relatively simple to compute. We consider these first and show how they can become the bases for two, more realistic although approximate, corresponding methods of analysis.

3. Two Rough and Ready Methods

One of the ideal cases is of balanced data. That is, if all colleges furnished the same number of students then the computational methods of the analysis of covariance (Fisher, 1958) become available. For example, Snedecor and Cochran's textbook (1967,pp.436-438) includes formulas for estimating σ_u^2 (called σ_B^2 there) and for testing equality of the between classes slope to the within classes slope. Although the present data set does not have equal sample sizes in every college it would likely be possible to consider a number of colleges having almost equal (say the ratio of largest to smallest is less than 3 to 1) sample sizes. Then one may randomly delete some cases in the schools with larger sample sizes and perhaps randomly duplicate some in the others and so end with a balanced data set without having damaged to any appreciable extent the information on the context effect. This possibility is discussed at some length by Searle (1971,p.364). Before recommending widespread use of this procedure one should investigate, at least through empirical sampling, the effects of varying amounts of deleting and duplicating.

The other ideal case involves the independent replication of the entire sample design in both the sample selection as well as the measurement and tabulation phases. Such designs have been advocated by Deming (1960, Part II, Replicated Sampling Designs) and are in common use, for example, for sampling business accounts whenever there is great, perhaps legal, need of an unbiased estimate of sampling variance. In such a case one would compute separate estimates of the parameter δ on each sub-sample and then compute a t-statistic to test that the population mean of these estimates is zero.

In practice, replicated designs may be deemed expensive and cumbersome to carry out or may not have been used because variance estimation was given low priority. In such cases the data can still be broken into pseudo-replicates (McCarthy, 1966) or portions and then one can use the above-described t-test procedure or some other method. One of these other methods is the Tukey Jackknife (Mosteller and Tukey, 1968) that proceeds as follows. One portion of the data is deleted, an estimate of the context effect is computed from the remainder and a pseudo-value is then formed. This portion is returned to the sample, another one deleted, and again a pseudo-value is computed. This is done until we have as many pseudo-values as portions. The pseudo-values are then used to compute a t-value as a test statistic for the hypothesis that the population average pseudo-value is zero. It is important to define "portions" so as to reflect the principal source of sampling variance in the application considered. In the present example a school is a portion. Each pseudo-value equals the-number-of-schools times the-estimate-of-context-effect-from-all-schools minus the-number-of-schools-less-one times the-estimate-based-on-all-schools-except-one.

4. Regression Approach

We return now to a more direct attack guided by model equation (2) and first using methods of estimating variance components described by Searle (1971) in conjunction with a generalized least squares procedure to estimate δ and to test that $\delta = 0$. First, however, we might ask how the test based on the least squares regression calculation that was employed in the cited papers performs when model equation (2) holds. If we adopt a more concise notation of $d_{ij} = x_{ij} - \bar{x}$ and $c_i = \bar{x}_i - \bar{x}_{..}$ then model equation (2) can be written in terms of deviates as:

$$(3)\quad y_{ij} = \mu + \beta\, d_{ij} + \delta\, c_i + u_i + e_{ij}\ ,$$

with $\sum\sum_{i,j} d_{ij} = 0$ and $\sum\sum_{i,j} c_i = 0 = \sum n_i c_i$. Notice that $\sum_{j=1}^{n_i} d_{ij}/n_i = c_i$.

The regression computations proceed from the 4 by 4 matrix of sums of squares and cross products. This is shown in partitioned form just prior to removing the β and μ effects as:

$$(4)\quad A = \begin{array}{c} \text{Effects:}\quad \begin{array}{cccc} y & \delta & \beta & \mu \end{array} \\ \left[\begin{array}{cc|cc} TYY & BXY & TXY & y_{++} \\ BXY & BXX & BXX & 0 \\ \hline TXY & BXX & TXX & 0 \\ y_{++} & 0 & 0 & n \end{array}\right] \end{array} = \begin{bmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{bmatrix}$$

Partitioning A into 2 by 2 sub-matrices as shown in (4) and sweeping out the μ and β effects corresponding to the last two rows and columns, leaves the following residual cross products matrix.

$$(5)\quad A_{11\cdot 2} = A_{11} - A_{12}\, A_{22}^{-1}\, A_{21}$$

$$= \begin{bmatrix} TYY - y_{++}^2/n - (TXY)^2/TXX & BXY - (TXY)(BXX)/TXX \\ BXY - (TXY)(BXX)/TXX & BXX - (BXX)^2/TXX \end{bmatrix} .$$

Next sweeping out the now-adjusted δ effect from $A_{11\cdot 2}$ yields an estimate of δ as:

$$\hat{\delta} = \frac{[BXY - (TXY)(BXX)/TXX]}{[BXX - (BXX)^2/TXX]}$$, and the following

regression sum of squares (call it DSS) for a test of $\delta = 0$:

$$(6)\quad DSS = [BXY - (TXY)(BXX)/TXX]^2/(BXX - BXX^2/TXX).$$

In the absence of the u_i terms this quantity, DSS, would be taken to have a single degree of freedom chi-square distribution and could be divided by the error mean square, say $RSS/(n-3) = RMS$, to furnish an F-ratio test (or, since $t = \sqrt{F}$ when numerator degrees of freedom are one, a t-test) statistic for the hypothesis $\delta = 0$. This is likely to have been the calculation that led to the t = 10 value in Farkas (1974), and will be denoted $t_1 = \sqrt{F_1} = \sqrt{DSS/RMS}$.

In case that model equation (2) holds with $\sigma_u^2 \neq 0$ we find E(DSS) as follows. First, express DSS as:

$$(7)\quad DSS = (TXX\ BSY - TXY\ BXX)^2/BXX(TXX-BXX)TXX\ .$$

Writing $WXX = TXX - BXX$, introducing the expectation, and substituting for y_{ij} from model equation (2) leads to:

$$(8)\quad BXX\ WXX\ TXX\ E(DSS) = (BXX\ WXX)^2\delta^2 + WXX^2 \sum n_i^2 c_i^2 \sigma_u^2 + BXX\ WXX\ TXX\ \sigma_e^2\ .$$

The final result is:

$$(9)\quad E(\mathrm{DSS}) = \frac{\mathrm{BXX\ WXX}}{\mathrm{TXX}}\,\delta^2 + \frac{\Sigma n_i^2 c_i^2 \mathrm{WXX}}{\mathrm{BXX\ TXX}}\,\sigma_u^2 + \sigma_e^2 \ .$$

On the other hand as we will see below in (15):

$$(9a)\quad E(\mathrm{RMS}) = \sigma_e^2 + [n - \mathrm{tr}(Z'X(X'X)^{-1}X'Z)]\sigma_u^2/(n-3),$$

where Z and X in this formula refer to data matrices. The coefficient of σ_u^2 in E(DSS) is generally larger than that of σ_u^2 in E(RMS), and thus the ratio F_1 is biased upward from 1 .

It is thus evident that only if $\sigma_u^2 = 0$ (i.e., if model equation (1) holds) will the regression approach be correct. If $\sigma_u^2 \neq 0$ then to form an F-ratio test statistic one needs to search elsewhere than the error mean square (RMS) for a denominator. If one examines the expected value of the residual sum of Y-squares after removal of DSS, namely of:

$$(10)\quad \mathrm{RSS} = \mathrm{TYY} - y_{++}^2/n - (\mathrm{TXY})^2/\mathrm{TXX} - \mathrm{DSS} \ ,$$

it is found that both σ_u^2 and σ_e^2 appear, but no fixed effects do. By next sweeping out the college effects as well, one can obtain an even more refined error sum of squares, (ESS, say) and thus calculate an unbiased estimate of σ_e^2 alone. Using this, along with the coefficients of σ_u^2 and σ_e^2 in E(RSS-ESS), allows one to get an estimate of σ_u^2. It bears emphasizing that one should examine the F-ratio in which the mean square (RSS-ESS)/(t-2) is divided by the error mean square, as a test of $\sigma_u^2 = 0$. If rejection of the hypothesis $\sigma_u^2 = 0$ is not possible even at the, say, 20% level, then one may proceed to act as if model equation (1) held and perform a t-test using $\hat{\delta}$ and the error mean square EMS=ESS/n-t-1 . I would be fairly reluctant to suppose that $\sigma_u^2 = 0$ and so would not recommend this test. At any rate it may be denoted $t_2 = \sqrt{\mathrm{DSS/EMS}}$.

5. Estimating Variance Components for the Test

A denominator for a F-ratio statistic with DSS as numerator can be constructed using method of moments estimates based on ESS and RSS_2 of σ_u^2 and σ_e^2 or more elaborate estimates of σ_u^2 and σ_e^2. The notation and most of the methods are to be found in Searle's textbook (1971), and we will first move toward that notation.

Model equation (2) can be written in matrix form as (see p.465 in Searle,1971)

$$(11)\quad \underset{\sim}{y} = X\underset{\sim}{b} + Z\underset{\sim}{u} + \underset{\sim}{e} \ ,$$

which in the present case becomes:

$$(12)\quad \begin{bmatrix} y_{11}\\ y_{12}\\ \vdots\\ y_{t,n_t}\end{bmatrix} = \begin{bmatrix} 1 & d_{11} & c_1\\ 1 & d_{12} & c_1\\ \vdots & \vdots & \vdots\\ 1 & d_{t,n_t} & c_t\end{bmatrix}\begin{bmatrix}\mu\\ \beta\\ \delta\end{bmatrix} + \begin{bmatrix} 1 & 0 & \cdots & 0\\ 1 & 0 & & 0\\ \vdots & \vdots & \cdots & \vdots\\ 0 & 0 & & 1\end{bmatrix}\begin{bmatrix}u_1\\ u_2\\ \vdots\\ u_t\end{bmatrix} + \begin{bmatrix}e_{11}\\ e_{12}\\ \vdots\\ e_{t,n_t}\end{bmatrix}$$

Searle assumes that the only linear relationship of columns of X to those of Z is by way of the first column of X equal to the sum of the columns of Z. In our case there is yet another relationship so that we must generalize his derivation as follows. Let r be the rank of X and t (the number of PSU's) is, of course, the rank of Z. Then define λ_1 to be

$$(13)\quad \lambda_1 = r + t - r[X\ Z] \ ,$$

where r[X Z] is the rank of the combined matrix [X Z]. In our example $\lambda_1 = 2$, but we might as well be prepared for dealing with numerous independent variables since the contextual effect may be multi-componential.

What had previously been denoted as TYY is now written as $\underset{\sim}{y}'\underset{\sim}{y}$. The quantity R(b) will refer to the reduction in sum of y-squares due to fitting the model:

$$(14)\quad \underset{\sim}{y} = Xb + e \ ,$$

while R(b,u) will refer to the reduction achieved by fitting the full model of rank r[X Z]. The additional reduction may be written $R(u|b) = R(b,u) - R(b)$ and upon using formula (79) in Searle (1971,p.445) its expectation is found to be:

$$(15)\quad E(R(u|b)) = \sigma_u^2(n - \mathrm{tr}[Z'X(X'X)^{-}X'Z]) + \sigma_e^2(t-\lambda_1) \ ,$$

while

$$(16)\quad E(\underset{\sim}{y}'\underset{\sim}{y}) - R(b,u) = [N - (r + t - \lambda_1)]\ \sigma_e^2 \ .$$

In our example, X is of full rank and upon extending the method to cover more than one independent variable when the observed values are obtained by survey methods (rather than from a balanced experiment), it will continue to be of full rank so that we could have written $(X'X)^{-1}$ rather than just $(X'X)^{-}$, the generalized inverse, as appears in the textbook.

To test $\sigma_u^2 = 0$ one computes the ratio

$$(17)\quad F = \frac{R(u|b)}{t-\lambda_1} \Bigg/ \frac{(\underset{\sim}{y}'\underset{\sim}{y} - R(b,u))}{N-(r+t-\lambda_1)}$$

and refers it to a table of the $F(t-\lambda_1, N-(r+t-\lambda_1))$ distribution.

The estimates can be written explicitly as:

$$(18)\quad \hat{\sigma}_e^2 = (\underset{\sim}{y}'\underset{\sim}{y} - R(b,u))/[N - (r + t - \lambda_1)]$$

$$(19)\quad \hat{\sigma}_u^2 = (R(u|b) - (t-\lambda_1)\hat{\sigma}_e^2)/(n - \mathrm{tr}[Z'X(X'X)^{-1}X'Z]).$$

At this point one could refer to expression (9) and using the values obtained for $\hat{\sigma}_e^2$ and $\hat{\sigma}_u^2$, compute $[\Sigma n_i^2 c_i^2\ \mathrm{WXX}\ \hat{\sigma}_u^2/\mathrm{BXX\ TXX} + \hat{\sigma}_e^2]$ which is then divided into DSS to get an F-ratio test statistic.

In practice, one could perhaps improve the estimates of σ_e^2 and σ_u^2 in accord with Thompson (1969), by first computing the ratio

$\hat{\lambda} = \hat{\sigma}_e^2/\hat{\sigma}_u^2$ and then new value of $R(b,u)$ and $R(u|b)$ as $R^*(b,u)$ and $R^*(u|b)$ based on the maximum likelihood equations for b (Searle,1971, section 11.7c). Upon iteration the estimators may be denoted with an upper squiggle as $\tilde{\sigma}_u^2$ and $\tilde{\sigma}_e^2$. They can be written explicitly as:

(18a) $\tilde{\sigma}_e^2 = (\underset{\sim}{y}'\underset{\sim}{y} - R^*(b,u))/(N - r)$

(19a) $\tilde{\sigma}_u^2 = R^*(u|b)/c$,

where c is the same denominator as for $\hat{\sigma}_u^2$ in equation (19).

We can now use generalized least squares to test H_0. The maximum likelihood equations for estimating the fixed effects are:

$$(20)\quad \begin{matrix} X'X & X'Z \\ Z'X & P \end{matrix} \quad \begin{matrix} \underset{\sim}{b}^* \\ \underset{\sim}{u}^* \end{matrix} = \begin{matrix} X'y \\ Z'y \end{matrix}$$

where $P = Z'Z + \lambda I$ and $\lambda = \sigma_e^2/\sigma_u^2$. The appearance of λI is the new ingredient. From

(21) $X'X\, \underset{\sim}{b}^* = X'(\underset{\sim}{y} - Z\, \underset{\sim}{u}^*)$ and $P\underset{\sim}{u}^* = Z'(\underset{\sim}{y} - X\, \underset{\sim}{b}^*)$

we get

(22) $X'X\, \underset{\sim}{b}^* = X'(\underset{\sim}{y} - Z\, P^{-1}Z'\underset{\sim}{y} + X'Z\, P^{-1}Z'X\, \underset{\sim}{b}^*$

or

(23) $X'(I - Z\, P^{-1}Z')X\, \underset{\sim}{b}^* = X'(I - Z\, P^{-1}Z')\underset{\sim}{y}$ or finally:

(24) $\underset{\sim}{b}^* = (X'T\,X)^- X'T\, \underset{\sim}{y}$ where

$T = (I - Z\, P^{-1}Z')'$.

The matrix $\underset{\sim}{b}^*$ contains estimates of the parameters μ, δ and β (call them μ^*, δ^* and β^*) and its covariance matrix is estimated by $\hat{\sigma}_u^{-2}(X'T\,X)^{-1}$. Therefore a test of $\delta = 0$ can be based on the ratio of δ^* and its standard error obtained from $\hat{\sigma}_u^{-2}(X'T\,X)^{-1}$.

6. Maximum Likelihood

It was most helpful to have someone do the differentiations for a maximum likelihood solution as Jennirch and Sampson (1976) have provided. It may be that the method of maximum likelihood is justly criticized for its small sample properties. In the present case of variance estimation it provides negatively biased estimators since division is by n rather than n-1 . It does however, furnish a lot of useful results in such a situation as the present one where nuisance parameters (σ_e^2 and σ_u^2 as well as μ and β) abound. Their (Jennirch and Sampson's) paper (and previous ones by Hartley and others that they cite) are complete and explicit so that I would only show my tenuous grasp of the material by paraphrasing it.

7. Discussion

There are a number of general observations on data handling that this problem of testing for contextual effect prompts. The first is the recognition of some difference in point of view or, better, of methods between the pseudo-replicate or jackknife methods and the variance component estimation methods. There are many sources of this difference. One is more data-based and the other is more model-based, although both have models and both worry about data collection and processing. One proceeds from the survey sampling side and the other from the analysis of experimental data.

My own bias is toward favoring the components of variance approach and this paper (in its relative neglect of the literature on pseudo-replication) reflects that bias. I do believe that in the face of the complexities of sample design and measurement techniques, not to mention the population distributions, of the variables included in most social surveys the pseudo-replicate approach is the only contender in practice. The method does demand a considerable amount of judgement and experience to form replicates which will faithfully represent the uncertainty in the data. One variation of the method that I confess I do use in practice is to run regression on individual subjects and then correct t-values and chi-squares with the ratio of actual sample size to effective sample size. I use the results of Kish and Frankel (1974) that showed such ratios of 1.4 or so for regression coefficients.

The reason that I favor the variance component approach is its explicit attention to the model and particularly to its distributional assumptions not just the systematic part. There are, I suspect, a good many more components of variance floating around that should be included both for making honest tests and also for improving on future surveys. The controversy over tests of significance that bubbled in the sociological literature a few years back may have been fueled in part from misjudgements about the modeling of variances. Given the level of complexity in variances that was mentioned, it is no wonder to me that there have been serious mistakes in calculating levels of significance and I'm sure I've contributed to them. The solution is certainly not to scrap the tests, but rather to study variances harder.

A brief point needs to be made about computing requirements of the methods. The balanced data case is the least demanding and should be chosen by the researcher who wants quick and cheap, but honest results. I think that the jackknife also can be done by an investigator operating with limited computing resources. In doing the calculations for maximum likelihood and the regression-type computations I used a procedure, PROC MATRIX, from the software package known as SAS (Barr, Goodnight, Sall, Helwig,1976), and I can report that it was only moderately painful to write, while recognizing that my skills at programming are quite limited.

Returning to the problem as an exercise in data handling, it bears remarking again that no new methods were developed. The model is the mixed one. It has the, perhaps novel, feature of an additional linear dependence between the X matrix and the Z matrix (or U matrix as written by Hartley, et.al.). That is, the sum of the columns of Z equaling the constant column of X is the usual dependency, while the context effect (being a linear combination of group means) is a second dependency.

REFERENCES

Barr, A. J., Goodnight, J. H., Sall, J. P. and Helwig, J. H. (1976) A User's Guide to SAS76. Sparks Press, Raleigh, N. C.

Deming, W. E. (1960) Sample Design in Business Research. New York: Wiley.

Farkas, G. (1974) "Specification, residuals and contextual effects", Sociological Methods and Research, 2 (Feb.): 333-363.

Fisher, R. A. (1958) Statistical Methods for Research Workers. 13th edition, New York: Hafner.

Hauser, R. M. (1974) "Contextual analysis revisited", Sociological Methods and Research, 2 (Feb.): 365-375.

Kish, L. and Frankel, M. R. (1974) "Inferences from Complex Samples" Journal of Royal Statistical Society, A, 137: 1 - 37.

Mosteller, F. and Tukey, J. W. (1968) "Data Analysis" in Lindsey, G. and Aronson, E.,Eds., The Handbook of Social Psychology, Vol. 2, second edition. New York: Addison-Wesley.

Searle, S. R. (1971) Linear Models, New York: Wiley.

Snedecor, G. W. and Cochran, W. G. (1967) Statistical Methods, 6th edition, Ames, Iowa: Iowa State University Press.

Southern Region Research Project S-63, (1974), Influences on Occupational Goals of Young People in Three Southern Sub-cultures, Mimeo, Agricultural Experiment Stations of Alabama, Kentucky, Mississippi, North Carolina, South Carolina, Tennessee, and Virginia.

Thompson, R. (1969) "Iterative estimation of variance components for non-orthogonal data: Biometrics, 25: 767 - 773.

McCarthy, P. J. (1966) "Replication: An approach to the analysis of data from complex surveys", Vital and Health Statistic. PHS Publication, No. 1000, Series 2, No. 14, Washington, D. C.

An expanded version of the paper with an appendix containing examples of the computations is available from the author upon request.