The p-Problem with Forward Selection Stepwise Regression: Algorithm for Controlling Type I Errors

T. Mark Beasley, St. John's University, New York
Dennis W. Leitner, Southern Illinois University at Carbondale

The use of forward selection stepwise regression has been criticized for both interpretive misunderstandings and statistical aberrations. A major statistical problem with stepwise regression and other procedures that involve multiple significance tests is the inflation of the Type I error rate. Common approaches to control the family-wise error rate (e.g., the Bonferroni and Sidak corrections) are based on the assumptions of independent tests which typically reduce power. Because the presence of correlated predictors is a more realistic situation, other algorithms based on the average correlation in the predictor matrix have been proposed. The present study proposes an algorithm based on the maximum eigenvalue and the determinant of the predictor matrix for controlling the family-wise Type I error rate for multiple, correlated tests in forward selection regression under the complete null hypothesis. A Monte Carlo simulation with 5,000 replications was performed to demonstrate the effectiveness of the proposed algorithm.

Most users of multiple regression techniques in educational research are attempting to reduce a set of k predictor variables in order to report a simplified model. Typically, if a set of k predictors are regressed on a dependent variable, Y, only those predictors that are found statistically significant will be considered substantively valuable. Furthermore, because of the nature of many educational and psychological measurement scales, researchers are less likely to estimate regression coefficients as a way of interpreting substantive findings. Rather, F-ratios or p-values are used in a dichotomous decision process such that the relationship between a predictor and a criterion variable is "significant or not" (e.g., Thompson, 1989b). Furthermore, it is possible to have a statistically significant model (i.e., significant full model R^2) when the component variables are individually nonsignificant in either a zero-order or partial manner. However, educational researchers are not likely to consider such a model in the development of theory. Therefore, the forward selection procedure of stepwise regression became popular among educational researchers because it begins with significance tests of zero-order correlations and proceeds to more complex models.

For several years now, applied statisticians (e.g., Thompson, 1989a; Wilkinson, 1979) have been calling attention to the abuses of stepwise regression in its common use by less statistically sophisticated researchers. But theses and dissertations continue to step (unwisely) across the desks of graduate educators, and articles with many of these same problems continue to appear in print. It is hoped that elaborating these limitations and proposing new methods for using stepwise regression will bring about its more appropriate use. Three statistical procedures are considered under the rubric of stepwise regression: Forward selection; backward elimination; and true stepwise (Draper & Smith, 1981). Specifically, the forward selection procedure forms a model from the set of k dependent variables by first selecting the single best predictor. The second best predictor is then chosen by the criteria of strongest contribution to the prediction of Y, while controlling for the effects of the first predictor entered. Thus, the first step involves k simultaneous tests of zero-order correlations, while the second step involves $(k - 1)$ simultaneous tests of first-order semi-partial correlations (Aitkin, 1974). The process continues so that at each step the variable selected for inclusion significantly increases the prediction of Y (i.e., full model R^2).

The use of the various stepwise regression procedures has been criticized for many interpretative misuses and statistical aberrations. First, researchers often interpret the final solution of a reduced set of g predictors as being the best subset of predictors overall and of that size. Also, there is a tendency to confuse the order of entry and variable importance (Huberty, 1989). Stepwise procedures suffer from the use of the largest partial F as a test of a potential entry variable which is not in the regression model at that stage. The correct null sampling distribution for this test in not the ordinary F distribution, but is a partial F distribution which is very difficult to obtain. (Draper & Smith, 1981). Moreover, researchers often proceed to test each stage in a stepwise regression as if the partial F distribution does not exist and as if the test at that step is the only test that has or will occur. Furthermore, the *degrees of freedom* (*df*) used for these tests are often incorrect. For example, in the forward selection procedure the *df*'s used for the first step, a test of zero-order correlations, is $(N - 2)$, while $(N - k - 1)$ would be more appropriate. These considerations, in general, tend to inflate the probability of at least one Type I error (i.e., the probability of forming an erroneous model).

Another interpretative problem arises when two or more predictor variables are highly correlated. In such situations, there is a strong probability that one of the variables will absorb the majority of the other variables' prediction power and therefore cause their exclusion from subsequent models. Not only does a set of correlated predictors lead to potential substantive misinterpretations, it also makes estimating the probability of a Type I error more complex. Thus, due to multiple tests, incorrect *df*s, misunderstood partial F tests, and correlated predictor variables, it is difficult to determine the correct Type I error rate in stepwise regression. To compound these problems, the p-value associated with each variable entered stepwise into a regression equation (except for the final step) is incorrect in many canned statistical packages.

MULTIPLE TESTING AND THE TYPE I ERROR RATE

As with any statistical procedure, two kinds of inferential errors can be made. A Type I error occurs if a variable is selected when the population regression weight is zero. A Type II error occurs when a variable is not selected when it has a non-zero population regression weight. Many educational researchers adopt one of the traditional fixed significance levels (i.e., α = .05 or .01) when evaluating an F-ratio. This significance level determines the *Type I error rate* for each test independently. However, it is rare that educational researchers test a single hypothesis. Several variables and multiple significance tests are common. Thus, a researcher must consider the probability of committing a Type I error when multiple hypotheses are tested (i.e., the *family-wise error rate*).

In the context of post-hoc tests in the analysis of variance (ANOVA), the true family-wise Type I error rate (α_T) for k independent (i.e., orthogonal) tests with the same alpha level (α_i) is defined by the following equation:

$$\alpha_T = 1 - (1 - \alpha_i)^k , \qquad (1)$$

assuming the *complete null hypothesis* (i.e., all groups have identical means). Thus, the *family size* of the tests performed is equal to k. In order to return the Type I error to the nominal alpha (α_i), one could adjust α_i by the Sidak method:

$$\alpha_{adj} = 1 - (1 - \alpha_i)^{1/k} . \qquad (2)$$

This correction would yield an alpha level smaller than the nominal alpha, but over the course of multiple tests, this adjusted alpha (2) is expected to yield a Type I error rate equivalent to the nominal alpha, α_i.

Similarly, the forward selection method in stepwise regression conducts no less than k simultaneous tests of significance as if multiple tests are not performed. That is, the first predictor is selected by the largest zero-order correlation of all k variables without consideration for the number of tests being conducted. Thus, if an educational researcher using forward selection regression were to commit a Type I error under the complete null hypothesis (i.e., all k zero-order correlations between Y and the predictors were null), it would occur on the first step. That is, when all predictors are not correlated with the dependent variable, testing the maximum of the k zero-order dependent variable-predictor correlations determines the Type I error rate of the forward selection procedure. Thus, assuming independent predictors, the probability of a Type I error on the first step is equal to (1). To adjust α_T so that the Type I error rate returned to the nominal alpha (α_i), one could assume the family size is equal to k and adjust α_i with (2). However, if the k predictors were all perfectly correlated, then the family size would be equal to one ($k = 1$) and the Type I error rate would equal the nominal alpha (i.e., $\alpha_T = \alpha_i$). In the more realistic situation of correlated predictors, the solution for the correct Type I error rate is considerably more complex and requires the integration of the correlated F distribution (Pope & Webster, 1972). Furthermore, only limited tables of critical values are available (e.g., Games, 1977), while a few Monte Carlo approximations based on averaged correlations have been proposed (i.e., Krishnaiah & Armitage, 1965; Pohlmann, 1979).

For example, Pohlmann (1979) proposed a method based on the average squared correlation in the predictor matrix to control the Type I error in forward selection regression. To elaborate, a value, c, can estimate *family size* and substitute for k in (2) in order to control the *family-wise* Type I error rate. Pohlmann suggested the following function:

$$c = k - (k - 1)\bar{r}_x^2, \qquad (3)$$

where k equals the number of predictors and $\bar{r}_x^2$ equals the averaged squared inter-predictor correlation. Pohlmann also suggested correcting $\bar{r}_x^2$ by using a less biased estimate of the squared correlation based on the McNemar (1969) shrinkage formula. Initially, each squared correlation in the predictor matrix is corrected by:

$$\hat{r}_{ij}^2 = 1 - (1 - r_{ij}^2)\frac{(N-1)}{(N-2)} , \qquad (4)$$

where N equals the number of cases and r_{ij}^2 equals the square of the ijth element of the predictor matrix. Then $\bar{r}_x^2$ is calculated by:

$$\bar{r}_x^2 = \frac{\sum_{i=1}^{k-1} \sum_{j=i+1}^{k} \hat{r}_{ij}^2}{k(k-1)/2} \quad (5)$$

and entered into (3). However, Pohlmann's study simulated cases in which all correlations within the predictor matrix were equal which is an unrealistic expectation. That is, a variety of correlation patterns may yield the same average squared correlation, but it is not likely that the family-wise Type I error rates would be equal for these matrices.

PROPOSED ALGORITHM FOR ESTIMATING FAMILY SIZE

To consider another perspective, however, the appropriate Type I error rate may approach (1) with b orthogonal factors rather than these algorithms based on the average correlation of k predictors. To elaborate, one possible approach to the p-value problem would be to perform a principal component analysis (PCA) on the predictor correlation matrix and extract b orthogonal components. In fact, it can easily be shown that such a linear transformation will not affect the full model R^2. That is, if all k variables and the k components extracted from the predictor matrix are used as separate models to predict a criterion variable, Y, then both models would have the same full model R^2. The expected Type I error rate when using the k orthogonal principal components, however, will equal (1) for the first step of a forward selection stepwise regression. Thus, decomposing the set of k predictors into b orthogonal components and modifying algorithms for correlated predictors may provide a better approximation of the family-wise Type I error rate. Importantly, this indicates a relationship between the transformation matrix and the family-wise Type I error rate. Thus, it is proposed that the maximum eigenvalue (λ_{max}) from an unrotated principal components analysis and the determinant, $|\mathbf{P}|$, of the predictor correlation matrix, $\mathbf{P}$, is related to the proportion of Type I errors on the first-step, which defines the probability of forming an erroneous model under the complete null hypothesis.

The eigenvalues of a correlation matrix, $\mathbf{P}$, are commonly used as indices of the number of factors that underlie a correlation matrix (e.g., Kaiser, 1970). Furthermore, the maximum eigenvalue provides an index for the proportion of variance accounted for by the largest principal component, the average correlation of $\mathbf{P}$, and the number of underlying factors (Tatsuoka, 1988). The determinant of a correlation matrix, $|\mathbf{P}|$ has been used in establishing the independence of variables in PCA (Nagarsenker, 1976). The determinant of the covariance matrix, $|\mathbf{C}|$, gives the generalized variance (Tatsuoka, 1988), and the determinant of the correlation matrix, $|\mathbf{P}|$, is equal to $|\mathbf{C}|$ divided by the determinant of the diagonal variance matrix $|\mathbf{V}|$,

$$|\mathbf{P}| = \frac{|\mathbf{C}|}{|\mathbf{V}|}\,. \quad (6)$$

Thus, it follows that the generalized proportion of variance in $\mathbf{P}$, that is the generalized $\mathbf{R}^2$, is equal to:

$$\mathbf{R}^2 = 1 - \frac{|\mathbf{C}|}{|\mathbf{V}|} = 1 - |\mathbf{P}|\,. \quad (7)$$

Therefore, in combination λ_{max} and $|\mathbf{P}|$ provide rather unique information about the inter-correlation of the predictor matrix. Specifically in PCA, λ_{max} divided by k gives the proportion of variance in $\mathbf{P}$ accounted for by the first and largest principal component. However, λ_{max} is known to always be greater than one even in random data matrices (Horn, 1965). In fact, when the variables are independent and all off-diagonal elements in $\mathbf{P}$ are zero then $\mathbf{P}$ is an Identity matrix, $\mathbf{I}$, and the expected value of λ_{max} equals one,

$$\lim_{\mathbf{P} \to \mathbf{I}} \lambda_{max} = 1\,. \quad (8)$$

Therefore, subtracting one from λ_{max} and dividing by k would provide a corrected proportion of variance for the largest principal component

$$\frac{\lambda_{max} - 1}{k} \quad (9)$$

Also, if the variables are independent, then the determinant, $|\mathbf{P}|$, equals one,

$$\lim_{\mathbf{P} \to \mathbf{I}} |\mathbf{P}| = 1\,. \quad (10)$$

Although it is left undefined because such a matrix is not invertable, one can imagine that if all predictor variables were perfectly correlated, then λ_{max} would equal k. That is, the limit of λ_{max} as all the elements of $\mathbf{P}$ approach unity is k:

$$\lim_{\mathbf{P} \to 1} \lambda_{max} = k\,. \quad (11)$$

Furthermore, since the product of the eigenvalues must equal the determinant, then under the same conditions specified for (11), the limit of $|\mathbf{P}|$ equals zero as λ_{max} approaches k :

$$\lim_{\mathbf{P} \to 1} |\mathbf{P}| = 0 \quad \text{and} \quad \lim_{\lambda_{max} \to k} |\mathbf{P}| = 0\,. \quad (12)$$

Given conditions (11) and (12), all predictors are perfectly correlated and there is only one "true" variable and the *family size* (denoted as c) should be equal to one, which can be described as:

$$c = k - (k - 1) \quad (13)$$

Thus, $(k - 1)$ multiplied by (9) results in the proportion of $(k - 1)$ that should be subtracted from k; however, c also depends on the correlations in **P** whose generalized estimate comes from $|\mathbf{P}|$. Thus, $(k - 1)$ should be multiplied by (9) and (7). Therefore, c can be estimated by:

$$c = k - \frac{(k - 1)(\lambda_{max} - 1)(1 - |\mathbf{P}|)}{k} \quad . \quad (14)$$

Thus under the conditions set in (8), (10), (11), and (12), as the relationship among the predictor variables approaches perfect multicollinearity, the estimat of family size in (14) approaches one. Also if the k predictors are independent then (14) equals k. Therefore, if a researcher can use k, λ_{max}, and $|\mathbf{P}|$ to estimate the independence of the predictors in **P** with c, then (14) could be substituted for k in equation (2) and used as an estimate of family size to adjust α_T so that it approximates the nominal alpha. Thus in the present study, a Monte Carlo simulation of a forward-selection stepwise procedure with no expected correlation between the dependent variable, Y, and the k predictors was used to estimate the correct Type I error rate (p-values) for $k = 2, 3, 4, 5, 7$, and 10 correlated variables under various inter-predictor correlation conditions. From these results, the proposed formulation of c (14) was substituted for k in (2) to determine whether it was useful in controlling the Type I error rate.

For comparison purposes, Pohlmann's (1979) algorithm (3) was also used. The Appendix provides numerical examples that demonstrate the differences between the two methods.

METHODS

Simulation Procedure

A Monte Carlo program was written in SAS/IML (SAS Institute, 1990) to simulate the forward selection process of stepwise regression. Initially, the RANNOR fucnction, which provides a pseudo-random clock generated values, was use to generate a normally distributed predictor matrix, **X**, with dimensions of n rows (cases) and k columns (variables). All predictor means were equal to zero and all variances were equal to one. Then by using the fundamental postulate of PCA (Tatsuoka, 1988) and a method described by Kaiser and Dickman (1962), a $k \times k$ matrix of principal component coefficients, **F**, was derived from a prespecified predictor correlation matrix, **P** and pre-multiplied by the transpose of **X** to create a transformed data matrix $\mathbf{Z_P}$ that simulates **P** (see Beasley, 1994):

$$\mathbf{Z_P} = \mathbf{F}\ \mathbf{X}^t \quad (15)$$

Then a normally distributed dependent variable vector, **Y**, was randomly generated and concatenated with the transpose of $\mathbf{Z_P}$ to form the entire data matrix, **M**. Thus, although there was correlation among the k variables in **P**, there was no expected correlation between the predictor variables and Y. This process was replicated 5,000 times. Since an infinite number of correlation matrices can be simulated, various combinations of λ_{max} and $|\mathbf{P}|$ were used for each level of k. Tables 1, 2, and 3 in the Results section reference the values of λ_{max} and $|\mathbf{P}|$ that were imposed on **X**. The number of predictors was manipulated from $k = 2$, 3, 4, 5, 7, and 10. The number of cases was held constant at a fairly large number of $N = 200$ in order to avoid extreme shrinkage of R^2 (Harris, 1975) and to reduce the residual error in the transpose of $\mathbf{Z_P}$ as it simulates the predictor correlation matrix, **P**.

Test Procedures

Under conditions of the complete null hypothesis, if an erroneous model is to be formed (i.e., Type I error committed) using a forward selection procedure then it will occur on the first step. Furthermore most packaged stepwise programs (i.e., SAS STEPWISE, SPSS REGRESSION) perform the first entry with $(N - 2)$ df's. Therefore, the maximum zero-order correlation in the predictor column of **M** was tested. If the calculated $F(1, 180)$ exceeded the critical values for F at the $\alpha = .05$ level of significance, then it was counted as a Type I error. The number of rejections divided by the 5,000 replications served as empirical p-values and estimates of the true family-wise Type I error rate, α_T. The results of this procedure were used to help estimate family size,c. That is, if α_T and α_i are known then family size, c, can be solved as follows:

$$c = \frac{ln(1 - \alpha_T)}{ln(1 - \alpha_i)} , \quad (16)$$

where ln refers to the Naperian logarithm.

The expected values of k, λ_{max}, and $|\mathbf{P}|$ using the formula described in (14) were regressed on c derived from the simulations and (16) to investigate the goodness of fit. These results were also compared to the results of Pohlmann's (1979) algorithm (3). Furthermore, the effectiveness of (14) in controlling the family-wise Type I error rate was assessed by substituting these estimates of c for k in (2) to set a more stringent α_i in each simulation. These corrected Type I error rates were compared to the nominal alpha of .05..

RESULTS

Using the expected values of k, λ_{max}, and $|\mathbf{P}|$, the family size estimates of c from (14) were regressed on the empirical values of c derived from the proportion of rejections at the $\alpha = .05$ level of significance during the 5,000 replications. Thus, the following model was tested

$$c_{emp} = b_1 k + b_2 \frac{(k - 1)(\lambda_{max} - 1)(1 - |\mathbf{P}|)}{k} \quad (17)$$

with the intercept restricted to zero and the coefficients $b1$ and b_2 restricted to one. The model R^2 with these restriction was 0.9858. Figure 1 (upper panel) shows a scatter plot of this analysis with different elements for each value of k. The model R^2 when using Pohlmann's (1979) algorithm (3) based on averaged squared correlations was 0.9147. A scatter plot of that regression is shown in the lower panel of Figure 1. The diagonals on each panel represent a perfect fit of the expected and empirical values of family size. As can be seen, many more estimates of family size, c, deviate from the perfect fit diagonal for the Pohlmann's average squared correlation estimate of c as compared to the current proposed algorithm. Using a dependent t-test for correlations, the proposed correction (14) was found to be significantly better than Pohlmann's estimate of c, $t(67) = 9.70$, $p < .001$.

Tables 1, 2, and 3 show the expected values for the average squared correlation within the predictor matrix, $\mathbf{P}$, the maximum eigenvalue (λ_{max}), and the determinant of $\mathbf{P}$, $|\mathbf{P}|$. These tables also show the empirical values of the family-wise Type I error rate (Empirical p-values), the estimated value of family size, c from (14), and the corrected p-values after controlling Type I errors with (14).

As can be seen by looking across Tables 1, 2, and 3, when the number of predictors increased from k = 2 to 10 the expected increase in the family-wise Type I error rate also occurred. Also, by examining the first entry for any number of predictors (k), when the average squared correlation of the predictor matrix is zero, the empirical p-values approximate their estimated value from (1). For example, for $k = 4$ independent predictors(i.e., $E(\overline{\rho}^2) = 0$), the expected family size is four. Using (1) the expected family-wise Type I error rate under the complete null hypothesis is 0.1855. In comparison, the simulation in this study estimated the family-wise Type I error rate with an empirical p-value of 0.1870. From (16), the estimated family size is c = 4.0361 (see Table 2, upper panel). One can also see by looking within any Table that as the expected average squared correlation increases the Type I error rate and family size. Yet, some matrices with the same average

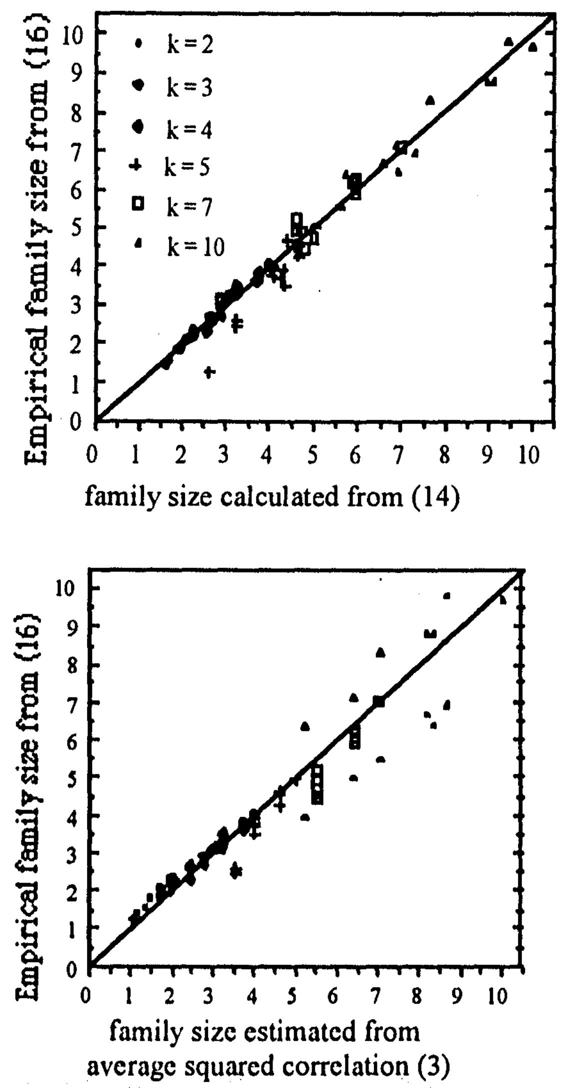

Figure 1. Empirical family size, c, derived from (16) as a function of the estimated family size from (14; upper panel) and from the average squared correlation (3; lower panel).

ρ^2 have different values for λ_{max}, and $|\mathbf{P}|$ and more importantly different empirical proportions of Type I errors. This is most notable in Table 3 with k = 10. It is important to note that when corrected with (14) the Type I error rates (corrected p-values) are reasonably close to the nominal alpha of .05 regardless of these discrepancies. Corrections based on average squared correlations (i.e., Pohlmann, 1979), however, would correct these discrepant configurations in the same manner. Thus, logically as well as statistically, the correction formula (14) provides a better adjustment for controlling Type I errors for multiple, correlated tests.

In any Monte Carlo study, one must consider the sampling error of the simulation process. Based on the nominal alpha of α = .05 and 5,000 replications, the standard error of each estimate is s_e = .003, which is used as a general heuristic to evaluate the proposed procedure. Although several corrected *p*-values exceed the + 2 standard error range, most are within the range of acceptability set by Bradley (1978). Explanation for these aberrations for the currently proposed correction may be twofold. One problem may be that some correction for sample size is necessary. Since sample size was held constant in this study, it should not have presented a serious problem. However, this possibility warrants further attention. A second problem is consistent with technical issues involving multicollinearity, in that the use of highly correlated predictor matrices yields extremely small determinants. In this case the accuracy of estimating such small values present a serious computational problem. That is, slight estimation errors can lead to rather large computational errors.

Table 1.
Expected values for the population average r^2, maximum eigenvalue (l_{max}) and determinant (|**P**|) with empirical Type I error rate, estimated *c* from (16) and corrected Type I error rate (14) for *k* = 2 and 3 predictors.

$E(\overline{\rho}^2)$	$E(\lambda_{max})$	E(|**P**|)	Empirical *p*-value	*c* (16)	Corrected *p*-value (14)
k = 2					
0.00	1.0000*	1.0000	0.0962	1.9719	0.0506
0.01	1.1000*	0.9900	0.0998	2.0497	0.0518
0.09	1.3000*	0.9100	0.0934	1.9116	0.0496
0.25	1.5000*	0.7500	0.0892	1.8215	0.0510
0.49	1.7000*	0.5100	0.0896	1.8301	0.0532
0.64	1.8000*	0.3600	0.0760	1.5410	0.0454
0.81	1.9000*	0.1900	0.0714	1.4442	0.0438
k = 3					
0.00	1.0000*	1.0000	0.1474	3.1089	0.0502
0.09	1.5695	0.7609	0.1296	2.7061	0.0496
	1.5984	0.7826	0.1394	2.9268	0.0494
	1.6000*	0.7840	0.1300	2.7150	0.0462
0.25	1.8922	0.2910	0.1120	2.3158	0.0452
	1.9860	0.4692	0.1286	2.6837	0.0516
	2.0000*	0.5000	0.1250	2.6033	0.0540
0.49	2.3658	0.0700	0.1052	2.1670	0.0532
	2.3986	0.2531	0.1090	2.2500	0.0504
	2.4000*	0.2160	0.1122	2.3202	0.0492
0.64	2.5885	0.0384	0.0910	1.8601	0.0462
	2.6000*	0.1040	0.1004	2.0627	0.0510

Note. * indicates that all correlations in **P** are equal.

Table 2.
Expected values for the population average r^2, maximum eigenvalue (l_{max}) and determinant ($|\mathbf{P}|$) with empirical Type I error rate, estimated *c* from (16) and corrected Type I error rate (14) for $k = 4$ and 5 predictors.

$k = 4$

$E(\bar{\rho}^2)$	$E(\lambda_{max})$	$E(\lvert\mathbf{P}\rvert)$	Empirical p-value	c (16)	Corrected p-value (14)
0.00	1.0000*	1.0000	0.1870	4.0361	0.0494
0.09	1.7926	0.5832	0.1744	3.7363	0.0470
	1.8016	0.5439	0.1706	3.6467	0.0474
	1.8964	0.6481	0.1746	3.7410	0.0478
	1.9000*	0.6517	0.1790	3.8452	0.0552
0.25	2.2670	0.1188	0.1548	3.2788	0.0512
	2.4150	0.1042	0.1508	3.1868	0.0528
	2.4995	0.3019	0.1542	3.2650	0.0582
	2.5000*	0.3125	0.1650	3.5155	0.0572
0.64	3.3696	0.0011	0.1042	2.1453	0.0538
	3.3984	0.0238	0.1106	2.2851	0.0520
	3.4000*	0.0272	0.1104	2.2807	0.0546
0.00	1.0000*	1.0000	0.2252	4.9743	0.0460
0.09	2.0462	0.3866	0.2068	4.5168	0.0516
	2.0558	0.4449	0.2078	4.5414	0.0524
	2.0954	0.3589	0.1970	4.2774	0.0514
	2.1946	0.5221	0.2112	4.6252	0.0486
	2.2000*	0.5282	0.2124	4.6549	0.0508
0.25	2.7652	0.0081	0.1648	3.5109	.0486
	2.8100	0.0768	0.1800	3.8689	0.0556
	2.9094	0.0112	0.1728	3.6985	0.0600
	2.9914	0.1745	0.1758	3.7693	0.0516
	3.0000*	0.1875	0.1818	3.9118	0.0628
0.64	4.1957	0.0039	0.1250	2.6033	0.0560
	4.2000*	0.0067	0.1184	2.4568	0.0598
0.98	4.9600	0.0001	0.0610	1.2271	0.0328

Note. * indicates that all correlations in **P** are equal.

Table 3.
Expected values for the population average r^2, maximum eigenvalue (l_{max}) and determinant ($|\mathbf{P}|$) with empirical Type I error rate, estimated c from (16) and corrected Type I error rate (14) for $k = 7$ and 10 predictors.

$k = 7$

$E(\overline{\rho}^2)$	$E(\lambda_{max})$	$E(\lvert\mathbf{P}\rvert)$	Empirical p-value	c (16)	Corrected p-value (14)
0.000	1.0000*	1.0000000	0.3024	7.0206	0.0476
0.090	2.3742	0.1097200	0.2664	6.0396	0.0548
	2.5193	0.1764100	0.2708	6.1569	0.0570
	2.5539	0.2311600	0.2738	6.2373	0.0546
	2.7745	0.3024000	0.2720	6.1890	0.0596
	2.8000*	0.3294200	0.2602	5.8755	0.0576
0.250	3.3159	0.0016268	0.2138	4.6896	0.0544
	3.5875	0.0004800	0.2040	4.4481	0.0486
	3.6922	0.0125600	0.2234	4.9291	0.0564
	3.7627	0.0001200	0.2072	4.5266	0.0540
	3.9779	0.0554100	0.2344	5.2072	0.0660
	4.0000*	0.0625000	0.2232	4.9241	0.0558
0.640	5.7938	0.0002600	0.1492	3.1501	0.0586
	5.8000*	0.0003700	0.1466	3.0906	0.0574

$k = 10$

$E(\overline{\rho}^2)$	$E(\lambda_{max})$	$E(\lvert\mathbf{P}\rvert)$	Empirical p-value	c (16)	Corrected p-value (14)
0.000	1.0000*	1.0000000	0.3920	9.7007	0.0456
0.153	2.3770*	0.5333101	0.3960	9.8289	0.0512
	4.0039	0.0003575	0.2986	6.9147	0.0516
0.187	2.6830*	0.4163279	0.3634	8.8045	0.0512
	4.4190	0.0000643	0.2794	6.3882	0.0514
0.205	2.8450*	0.3608996	0.3636	8.8107	0.0484
	4.7727	0.0000260	0.2900	6.6771	0.0532
0.327	3.9430*	0.1116766	0.3474	8.3206	0.0624
	5.8601	0.0000005	0.2462	5.51004	0.0548
0.400	4.6000*	0.0463574	0.3062	7.12708	0.0558
	6.5298	4.49^{e-8}	0.2268	5.01464	0.0536
0.532	5.7880*	0.0062336	0.2784	6.36115	0.0614
	7.4954	2.84^{e-9}	0.1840	3.96428	0.0550

Note. * indicates that all correlations in **P** are equal.

DISCUSSION

The behavioral science literature is replete with "significant" findings that fail the ultimate test of replication (Pedhazur, 1982; Rosnow & Rosenthal, 1989). One explanation for this conundrum lies in the family-wise Type I error rate that increases when stepwise regression or other multiple testing procedures are used. Faced with the problem of multiple tests that may be correlated, the researcher should take some action to correct the Type I error rate. Possible approaches to this problem include:

a). Prior to performing a stepwise regression, conduct an omnibus test with all potential predictors in the model.

b). When searching for a significant subset of predictors, use stepwise methods with backward elimination

c). When searching for a reduced subset of predictors through stepwise methods, perform a PCA and extract orthogonal components and use (1) to correct the family-wise Type I error rate.

d). In any multiple test situation, use one of several simultaneous inference tests (e.g., Games, 1977; Schafer, 1992; Schafer & Macready, 1975) to control Type I errors.

e). Use the Bonferroni inequality, however, one may over-correct the probability of a Type I error and lose power.

f). Use the algorithm (14) suggested here if multiple, correlated test are being performed.

It should be noted that there are practically an infinite number of configurations a correlation matrix can assume; therefore, there is no way to exhaust those possibilities. Therefore, these findings are limited to the specific correlation matrices simulated. Thus, although extensive replications of this study are needed to assume the generality of these findings, it is not unreasonable to assume that the proposed algorithm (14) will work in other situations.

Although the family-wise Type I error correction suggested here has been framed in terms of the forward selection procedure of stepwise regression, there is no reason for its exclusion from other situations that involve a single dependent variable and multiple tests that are correlated. For example, a set of nonorthogonal contrasts for an ANOVA, although based on coded vectors for means have correlations coefficients associated with them. Therefore, a matrix of correlations among contrasts could be analyzed with (14). In conclusion, the suggested algorithm shows adequate control of the family-wise Type I error rate and is based on more complete information than estimates based simply on the average squared correlation. Yet, in the results the suggested correction sometimes deviated from the nominal alpha. Thus, further investigation will focus on manipulating sample sizes and using a shrinkage correction for the determinant of the predictor correlation matrix.

REFERENCES

Aitkin, M. A. (1974). Simultaneous inference and the choice of variable subsets in multiple regression. *Technometrics, 16*, 221-27.

Beasley, T. M. (1994). CORRMTX: Generating correlated data matrices in SAS/IML, *Applied Psychological Measurement, 18*, 95.

Bradley, J. V. (1978). Robustness? *British Journal of Mathematical and Statistical Psychology, 31*, 144-152.

Draper, N., & Smith, H. (1981). *Applied regression analysis* (2nd ed.). New York: Wiley.

Games, P. A. (1977). An improved table for simultaneous control on *g* contrasts. *Journal of the American Statistical Association, 72*, 531-534.

Harris, R. (1975). *A primer of multivariate statistics.* New York: Academic Press.

Horn, J. L. (1965). A rationale and test for the number of factors in factor analysis. *Psychometrika, 30*, 179-185.

Huberty, C. J. (1989). Problems with stepwise methods - better alternatives. In B. Thompson (Ed.), *Advances in social science methodology* (vol. 1). Greenwich, CT: JAI Press.

Kaiser, H. F. (1970). A second generation Little Jiffy. *Psychometrika, 35*, 401-415.

Kaiser, H. F., & Dickman, K. (1962). Sample and population score matrices and sample correlation matrices from an arbitrary population correlation matrix. *Psychometrika, 27*, 179-182.

Krishnaiah R. K., & Armitage, R. B. (1965). *Probability integrals of the multivariate F distribution, with Tables and applications.* (Report No. ARL 65-236). Wright-Patterson AFB, OH: U. S. Air Force.

McNemar, Q, (1969). *Psychological statistics.* New York: Wiley.

Nagarsenker, B. N. (1976). The distribution of the determinant of correlation matrix useful in principal components analysis. *Communications in Statistics: Simulation and Computation, B5*, 1-13.

Pedhazur, E. J. (1982). *Multiple regression in behavioral research.* (2nd ed.). New York: Holt, Rinehart, and Winston.

Pohlmann, J. T. (1979). Controlling the Type I error rate in stepwise regression analysis. *Multiple Linear Regression Viewpoints, 10,* 46-60.

Pope, P. T., & Webster, J. T. (1972). The use of an *F*-statistic in stepwise regression procedures. *Technometrics, 14*, 327-340.

Rosnow, R. L., & Rosenthal, R. (1989). Statistical procedures and the justification of knowledge in psychological science. *American Psychologist, 44,* 1276-1284.

SAS Institute. (1990). *SAS/IML user's guide* (Release 6.04). Cary, NC: Author.

Schafer, W. D. (1992). Simultaneous inference options for statistical decision making. *Measurement and Evaluation in Counseling and Development, 25,* 98-101.

Schafer, W. D., & Macready, G. B. (1975). A modification of the Bonferroni procedure on contrasts which are grouped into internally independent sets. *Biometrics, 31*, 227-228.

Thompson, B. (1989a). Why won't stepwise methods die? Measurement *and Evaluation in Counseling and Development,, 21,* 146-148.

Thompson, B. (1989b). Statistical significance, result importance, and result generalizability: Three noteworthy but somewhat different issues. *Measurement and Evaluation in Counseling and Development,, 22*, 2-5.

Tatsuoka, M. M. (1988). *Multivariate analysis: Techniques for educational and psychological research* (2nd ed.). New York: Macmillan.

Wilkinson, L. (1979). Tests of significance in stepwise regression. *Psychological Bulletin, 86,* 168-174.