DOCUMENT RESUME

ED 257 860 TM 850 325

AUTHOR Barcikowski, Robert S.; Robey, Randall R.
TITLE Sample Size Selection In Single Group Repeated Measures Analysis.
PUB DATE Apr 85
NOTE 31p.; Paper presented at the Annual Meeting of the American Educational Research Association (69th, Chicago, IL, March 31-April 4, 1985). Small print in tables 2-6.
PUB TYPE Speeches/Conference Papers (150) -- Reports - Research/Technical (143)

EDRS PRICE MF01/PC02 Plus Postage.
DESCRIPTORS *Analysis of Variance; Effect Size; Hypothesis Testing; *Multivariate Analysis; Research Methodology; *Sample Size; Tables (Data)
IDENTIFIERS Hotellings t

ABSTRACT

This paper provides researchers with a method of determining sample size for a given power level in the preparation of a single group exploratory repeated measure analysis. The rationale for determining sample size which takes into consideration the powers and assumptions of both the adjusted univariate and multivariate repeated measures tests is presented. Six tables to determine sample size for a minimally acceptable power level (.80), at three levels of significance (.01, .05, and .10), and varying levels of repeated measures and effect size are given. The noncentrality parameters used in the FORTRAN program for the univariate and multivariate repeated measures tests to drive the sample sizes are presented in Appendix A. The noncentrality parameters are related to Cohen's effect size index (f), a commonly used measure of treatment differences. Three example analyses are given to illustrate the utility of this methodology. (BS)

Sample Size Selection In Single Group Repeated Measures Analysis

Robert S. Barcikowski

and

Randall R. Robey

Ohio University

Paper presented at the annual meeting of the American Educational Research Association, Chicago, April, 1985.

Abstract

In single group exploratory repeated measures analysis the use of both the univariate and multivariate repeated measures tests has been advocated. A method for determining the number of subjects necessary to achieve satisfactory power when both of these tests are considered is presented in this paper. Tables to determine sample size for a minimally acceptable power level (i.e., .80), given three levels of significance (.01, .05, and .10), and varying levels of repeated measures and of effect sizes are also presented.

Sample Size Selection
In Single Group
Repeated Measures Analysis

Introduction

Our intent in this paper is to provide researchers with a method of determining sample size for a given power level in the preparation of a single group exploratory repeated measures analysis. In so doing, we develope a rationale for determining sample size which takes into consideration the powers and assumptions of both the adjusted univariate, and the multivariate, repeated measures tests. In what follows we provide the backgound for this rationale, describe the rationale, and include a set of tables which will allow you to easily find sample sizes for single group repeated measures designs at a minimally acceptable power level. Examples in the use of these tables are also provided.

Background

The Use Of Two Tests: Fisher's F and Hotelling's T^2

In three recent papers (Barcikowski and Robey, 1984a, 1984b; Robey and Barcikowski, 1984) we have advocated the routine use of both the adjusted univariate F test and Hotelling's T^2, a multivariate test, in the analysis of single group exploratory repeated measures data. That is, we have recommended the use of both of these tests in situations where you are unable to determine *a priori* which test would be most powerful.

When both tests are used in an exploratory study, we generally recommend that you conduct each test at the level of significance that you would have used if you had conducted only one test. When you follow this advice, your experimentwise level of significance will be twice the level of significance used for each test, but you will not have sacrificed power.

If a significant result is found with either test, we recommend the use of a post hoc test based on individual error terms (Boik, 1981). Using individual error terms, Maxwell (1980) recommends the use of a Bonferroni dependent t-test approach to compare all pairs of means over several variations of the Tukey test. Also, for complex comparisons, Maxwell, Delaney, and Sternitzke (1984) recommend the use of individual error terms with either Roy-Bose simultaneous confidence intervals or with a Bonferroni dependent t-test over several variations of the Scheffé test.

The reason we have encouraged the routine use of both the univariate and multivariate tests is that it is possible for the univariate test (adjusted or unadjusted) to be nonsignificant, say at $p < .4444$, and for the multivariate test to be significant, say at $p < .01$. Although the multivariate test may not demonstrate such a dramatic power disadvantage, it can demonstrate power sufficiently low relative to the adjusted univariate test to miss treatment effects which the adjusted univariate test would detect (Barcikowski and Robey, 1984a; Davidson, 1972). For example, it is possible for the adjusted univariate test to be significant, say at $p < .05$, while the multivariate test could prove nonsignificant, say at $p < .1492$. These possibilities can occur when the univariate test's circularity (sphericity) assumption (i.e., that $\varepsilon = 1$, where ε is defined in Winer, 1971, p. 283) is violated, and can occur under a mild violation of this assumption, (e.g., when $\varepsilon = .95$).

You should note that under the condition where the circularity assumption holds, the univariate test will always be more powerful than the multivariate test. The reason for this is that both tests have the same numerator degrees of freedom (K-1), but the univariate test has larger denominator degrees of freedom ((K-1)(n-1)), as against (n-K+1) for the multivariate test.

In an exploratory study it would be unreasonable for a researcher to assume that circularity would hold. Therefore, prudent researchers routinely use an adjusted univariate F test to control the actual level of significance. This is because Collier, Mandeville, and Hays (1967) and Imhof (1962), among others, have shown that the actual level of significance will be inflated when circularity does not hold. The adjusted univariate F test is performed by estimating the circularity parameter, ε, from the data, and then multiplying it times the numerator and denominator degrees of freedom. The critical F value is then found using these "adjusted" degrees of freedom and compared with the calculated F statistic.

Two estimates of ε are available, $\hat{\varepsilon}$ recommended by Greenhouse and Geisser (1959), and $\tilde{\varepsilon}$ recommended by Huynh and Feldt (1976). Collier et al. (1967) have shown that the Greenhouse-Geisser $\hat{\varepsilon}$ yields a conservative adjusted F test when $\varepsilon \geq .90$. Huynh and Feldt have shown that their measure yields a liberal actual level of significance when $\varepsilon < .75$, but that the actual and nominal levels are close when ε is greater than or equal to .75. Therefore, they recommended $\tilde{\varepsilon}$ for use when $\varepsilon \geq .75$ and the Greenhouse-Geisser $\hat{\varepsilon}$ for use when $\varepsilon < .75$. However, in most exploratory studies the value of ε would be unknown, and so we recommend the routine use of the more conservative Greenhouse-Geisser $\hat{\varepsilon}$ in such studies. Therefore, in this paper when we refer to the adjusted test, we mean the Greenhouse-Geisser adjustment.

The Use Of Single Degree Of Freedom Contrasts

Because of the difficulties encountered when the circularity assumption is not met, Rouanet and Lepine (1970) and Rogan, Keselman, and Mendoza (1979) have recommended that researchers consider the use of the single degree of freedom contrast dependent t-test in place of the omnibus F and T^2 tests. That is, they recommend that researchers consider selecting a limited number of differences that they would like to investigate, and then test these differences without first using an overall test. This strategy is attractive because the dependent t-test for a single degree of freedom contrast does not require the circularity assumption.

There are three reasons why in general we think the use of both omnibus tests followed by a post hoc testing procedure is a better strategy than using single degree of freedom contrasts when dealing with an exploratory repeated measures analysis. First, in an exploratory study a small number of single degree of freedom contrasts may be difficult to formulate prior to the collection of the data. Second, when only a small number of contrasts are specified to be tested, other contrasts of interest may not be legitimately tested without playing havoc with Type I error. What does a researcher do when the selected contrasts are not significant, but other interesting contrasts appear in the data? Third, the powers of the two procedures can be close to each other (see Appendix A) when the number of comparisons is small, but the omnibus test will generally gain a power advantage as the number of contrasts increases.

We realize that there are situations where even in an exploratory study a researcher may *a priori* decide to ask only a limited number of questions of his/her data. For these situations, the tables presented in this paper can be of assistance in selecting sample size, see Appendix A. Note, however, that we strongly endorse the use of single degree of freedom contrasts in confirmatory studies where past research and/or theory enables you to make predictions that can best be tested using single degree of freedom contrasts.

A Rationale For Sample Size Selection

Sample sizes for the tables in this paper were derived using a modification of a FORTRAN program described in Robey and Barcikowski (1984). The noncentrality parameters used in this program for the univariate and multivariate repeated measures tests are described in Appendix A. In Appendix A the noncentrality parameters are related to a commonly used measure of treatment differences, Cohen's (1977) effect size index (f). In this paper Cohen's effect size index is labeled f_U for the univariate case and f_M for the multivariate case.

The sample size tables prepared for this paper were based on the multivariate effect size, f_M. In order to understand why f_M was used instead of f_U to determine sample size, consider the logic of the following statements.

1. When the univariate test's circularity assumption is met, the univariate test is more powerful than the multivariate test because of the univariate test's greater number of degrees of freedom, (n-1)(K-1) versus (n-K+1).

2. If the univariate effect size (f_U) is used to find sample size, it is reasonable to assume that the circularity assumption holds. This is because in an exploratory study one would have no basis for selecting different effect sizes for the univariate and multivariate tests, and the effect sizes for the two tests are equal under circularity, see Appendix A. In this case, because of the difference in denominator degrees of freedom, the sample size found for the univariate case will be less than that found using f_M for the multivariate case.

3. When the circularity assumption is violated the power of the adjusted univariate test varies from being more powerful than the multivariate test to being dramatically less powerful than the multivariate test (Barcikowski and Robey, 1984a; Jensen, 1982).

4. When the adjusted univariate test is less powerful than the multivariate test, then the multivariate test should be used. However, if sample size had been based on the unadjusted univariate test the power of the multivariate test could be too low (see statements #1 and #2).

5. When the univariate test's circularity assumption is violated, and the adjusted univariate test is more powerful than the multivariate test, having based the sample size on the multivariate test yields a power bonus.

Repeated Measures Effect Sizes

In exploratory analyses of variance (ANOVA's) Cohen (1977) provides three different effect sizes, "small", f = .10, "medium", f = .25, and "large", f = .40 that can be used as benchmarks to help determine the sample size needed for an experiment. Based on informed judgement, a researcher can select one of these benchmark measures of treatment differences and then use Cohen's tables to select a sample size for a given level of significance, and number of treatments.

Cohen (1977) does not provide benchmark effect sizes for repeated measures analyses. In Table 1, we provide such measures based on three possible intercorrelations among repeated measures. The repeated measures benchmark effect sizes shown in Table 1 are based on the assumption that the correlations among the repeated measures are constant.

Constant correlations among repeated measures describes a condition known as "compound symmetry" or as "uniformity". When this condition is met, we show in Appendix A that $f_U = f_M = f/\sqrt{1-\rho}$. Therefore, the measures shown in Table 1 are Cohen's benchmark effect sizes for analysis of variance divided by $\sqrt{1-\rho}$. Under noncircularity, a correlation in Table 1 might be considered to represent the population intraclass correlation for repeated measures data.

Table 1
Repeated Measures Benchmark Effect Sizes

Correlation	ANOVA Benchmark Effect Sizes small = .10	medium = .25	large = .40
.30	.12	.30	.49
.50	.14	.35	.57
.80	.22	.56	.89

The correlations, .30, .50, and .80 in Table 1 were subjectively selected. The correlation of .30 appears to be a reasonable lower limit that one would expect to find among the measures in a repeated measures design. The correlation of .50 seems to represent a reasonable conservative measure of the relationship among repeated measures. Correlations of .80 or higher are found in many repeated measures designs.

For most repeated measures studies we recommend the effect sizes associated with the correlation of .50. We make this recommendation because in most cases the effect sizes based on a correlation of .50 should slightly underestimate the actual effect size, and therefore, they will provide sample sizes which will yield high power.

Sample Size Tables

Five tables of sample sizes for single group exploratory repeated measures designs were generated. Table 2 contains sample sizes for the effect sizes shown in Table 1. The effect sizes in Table 2 represent our repeated measures equivalents to Cohen's "small," "medium," and "large" effect sizes. Tables 3 through 6, based respectively on the .01, .025, .05, and .10 levels of significance, contain sample sizes for a more general set of effect sizes.

In Tables 2 through 6 the number of repeated measures, K, is set at 2 through 10 inclusively with additional levels of 20 and 30. These tables contain the sample sizes that would be necessary to obtain a power value as close to .80 as is possible without becoming less than .80. Cohen (1977, p. 56) proposed that when a researcher "has no other basis for setting the desired power value, the value .80 be used;" this advice was taken in the construction of these tables.

To use Tables 2 through 6 to select a sample size for a single group exploratory repeated measures design you must consider:

a) a level of significance,
b) the expected correlation among the repeated measures,
c) the effect size that you would like to be able to detect, and
d) the number, K, of repeated measures.

Given this information, Tables 2 through 6 will yield sample sizes that will allow Hotelling's T^2 to have power of .80. And, although the adjusted univariate F test's power can be less than, equal to, or greater than .80, depending on the degree of noncircularity, you are sure of having adequate power to detect repeated measures differences if they exist.

Repeated Measures Effect Sizes: Table 2

Table 2 contains sample sizes for levels of

Table 2

Sample Sizes For Power At .80

Rho	Effect Size	2	3	4	5	6	7	8	9	10	20	30
		Number Of Repeated Measures (K)										
		Alpha = .005										
.30	0.12	462.	365.	306.	267.	238.	217.	201.	188.	177.	126.	110.
	0.30	78.	63.	55.	49.	46.	43.	41.	40.	39.	39.	44.
	0.49	32.	27.	25.	24.	23.	23.	22.	23.	23.	29.	37.
.50	0.14	340.	270.	227.	198.	178.	162.	150.	141.	133.	98.	89.
	0.35	58.	48.	42.	38.	36.	34.	33.	33.	32.	34.	41.
	0.57	25.	22.	20.	20.	19.	19.	19.	20.	20.	27.	36.
.80	0.22	141.	113.	96.	85.	77.	72.	67.	64.	62.	53.	54.
	0.56	26.	22.	21.	20.	20.	20.	20.	20.	20.	27.	36.
	0.89	13.	12.	12.	13.	13.	14.	14.	15.	16.	24.	34.
		Alpha = .01										
.30	0.12	404.	324.	273.	238.	214.	195.	181.	169.	160.	115.	102.
	0.30	68.	56.	49.	44.	41.	39.	37.	36.	36.	36.	42.
	0.49	28.	24.	22.	21.	21.	21.	21.	21.	21.	27.	36.
.50	0.14	298.	239.	202.	177.	159.	146.	136.	127.	121.	90.	82.
	0.35	51.	43.	38.	35.	33.	31.	30.	30.	29.	32.	39.
	0.57	22.	19.	18.	18.	18.	18.	18.	18.	19.	26.	35.
.80	0.22	123.	100.	86.	76.	69.	65.	61.	58.	56.	49.	51.
	0.56	22.	20.	19.	18.	18.	18.	18.	19.	19.	26.	35.
	0.89	11.	11.	11.	12.	12.	13.	13.	14.	15.	24.	33.
		Alpha = .025										
.30	0.12	327.	267.	227.	200.	180.	165.	154.	144.	136.	100.	90.
	0.30	55.	46.	41.	37.	35.	33.	32.	31.	31.	33.	40.
	0.49	23.	20.	19.	18.	18.	18.	18.	18.	19.	26.	34.
.50	0.14	241.	197.	168.	149.	134.	124.	115.	108.	103.	78.	73.
	0.35	41.	35.	32.	29.	28.	27.	26.	26.	26.	30.	37.
	0.57	18.	16.	15.	15.	15.	15.	16.	16.	17.	25.	34.
.80	0.22	100.	83.	71.	64.	59.	55.	52.	50.	48.	43.	47.
	0.56	18.	17.	16.	15.	15.	16.	16.	17.	17.	25.	34.
	0.89	9.	9.	10.	10.	11.	11.	12.	13.	14.	23.	32.

							Alpha = .05					
.30	0.12	268.	223.	192.	170.	154.	141.	132.	124.	117.	88.	80.
	0.30	45.	39.	35.	32.	30.	29.	28.	27.	27.	30.	37.
	0.49	19.	17.	16.	16.	16.	16.	16.	16.	17.	24.	33.
.50	0.14	199.	165.	142.	126.	114.	106.	99.	93.	89.	69.	66.
	0.35	34.	30.	27.	25.	24.	23.	23.	23.	23.	28.	36.
	0.57	14.	14.	13.	13.	13.	14.	14.	15.	15.	24.	33.
.80	0.22	82.	69.	60.	54.	50.	47.	45.	43.	42.	39.	44.
	0.56	15.	14.	13.	13.	14.	14.	14.	15.	16.	24.	33.
	0.89	8.	8.	8.	9.	10.	10.	11.	12.	13.	22.	32.
							Alpha = .10					
.30	0.12	209.	178.	154.	137.	125.	116.	108.	102.	97.	74.	69.
	0.30	35.	31.	28.	26.	25.	24.	23.	23.	23.	28.	35.
	0.49	14.	14.	13.	13.	13.	13.	14.	14.	15.	23.	32.
.50	0.14	154.	131.	114.	102.	93.	87.	81.	77.	73.	59.	58.
	0.35	26.	24.	22.	20.	20.	19.	19.	19.	20.	26.	34.
	0.57	11.	11.	11.	11.	11.	12.	13.	13.	14.	22.	32.
.80	0.22	64.	55.	49.	44.	41.	39.	37.	36.	35.	35.	40.
	0.56	12.	11.	11.	11.	12.	12.	13.	13.	14.	23.	32.
	0.89	6.	7.	7.	8.	9.	9.	10.	11.	12.	21.	31.
							Alpha = .20					
.30	0.12	149.	130.	114.	103.	94.	87.	82.	78.	74.	59.	58.
	0.30	25.	23.	21.	20.	19.	19.	19.	19.	19.	25.	33.
	0.49	10.	10.	10.	10.	11.	11.	12.	12.	13.	22.	31.
.50	0.14	110.	96.	85.	76.	70.	65.	62.	59.	56.	48.	49.
	0.35	19.	17.	16.	16.	15.	15.	16.	16.	16.	23.	32.
	0.57	8.	8.	8.	9.	9.	10.	11.	12.	12.	21.	31.
.80	0.22	45.	40.	36.	33.	31.	30.	29.	28.	28.	30.	36.
	0.56	8.	8.	9.	9.	10.	10.	11.	12.	12.	21.	31.
	0.89	4.	5.	6.	7.	8.	8.	9.	10.	11.	21.	31.

Note. This table contains the sample sizes necessary to detect small, medium and large effect sizes at the .90 power level. The effect sizes are given for the three magnitudes of intrclass correlation coefficients (i.e., .3, .5 and .8). Alpha levels are varied at .005, .025, .01, .05, .10 and .20.

Table 3

Sample Sizes Necessary To Detect Various Effect Sizes With Power At .80 And Alpha At .01

Effect Size	Number Of Repeated Measures (K) 2	3	4	5	6	7	8	9	10	20	30
0.05	2308.	1839.	1541.	1340.	1193.	1082.	924.	816.	737.	650.	503.
0.10	580.	464.	390.	340.	304.	277.	239.	213.	195.	175.	143.
0.15	260.	209.	177.	155.	140.	128.	112.	102.	94.	87.	77.
0.20	148.	120.	102.	91.	82.	76.	68.	63.	59.	56.	54.
0.25	96.	79.	68.	61.	56.	52.	47.	45.	43.	42.	44.
0.30	68.	56.	49.	44.	41.	39.	36.	35.	35.	35.	39.
0.35	51.	43.	38.	35.	33.	31.	30.	29.	29.	30.	36.
0.40	40.	34.	30.	28.	27.	26.	25.	26.	26.	28.	34.
0.45	32.	28.	25.	24.	23.	23.	23.	23.	24.	26.	32.
0.50	27.	24.	22.	21.	20.	20.	20.	21.	22.	24.	31.
0.55	23.	20.	19.	18.	18.	18.	19.	20.	21.	23.	31.
0.60	20.	18.	17.	17.	17.	17.	18.	19.	20.	22.	30.
0.65	18.	16.	15.	15.	15.	16.	17.	18.	19.	22.	30.
0.70	16.	15.	14.	14.	14.	15.	16.	17.	19.	21.	29.
0.75	14.	13.	13.	13.	14.	14.	15.	17.	18.	21.	29.
0.80	13	12.	12.	13.	13.	14.	15.	16.	18.	20.	29.
0.85	12.	12.	12.	12.	12.	13.	14.	16.	18.	20.	28.
0.90	11.	11.	11.	12.	12.	13.	14.	16.	17.	20.	28.
0.95	10.	10.	11.	11.	12.	12.	14.	15.	17.	20.	29.
1.00	10.	10.	10.	11.	11.	12.	14.	15.	17.	20.	28.
1.05	9.	9.	10.	10.	11.	12.	13.	15.	17.	19.	28.
1.10	9.	9.	9.	10.	11.	11.	13.	15.	17.	19.	28.
1.15	8.	9.	9.	10.	11.	11.	13.	15.	16.	19.	28.
1.20	8.	8.	9.	10.	10.	11.	13.	14.	16.	19.	28.
1.25	8.	8.	9.	9.	10.	11.	13.	14.	16.	19.	27.
1.30	7.	8.	9.	9.	10.	11.	12.	14.	16.	19.	27.
1.35	7.	8.	8.	9.	10.	11.	12.	14.	16.	19.	27.
1.40	7.	7.	8.	9.	10.	10.	12.	14.	16.	19.	27.
1.45	7.	7.	8.	9.	10.	10.	12.	14.	16.	19.	27.
1.50	6.	7.	8.	9.	9.	10.	12.	14.	16.	18.	27.
1.55	6.	7.	8.	9.	9.	10.	12.	14.	16.	19.	27.
1.60	6.	7.	8.	9.	9.	10.	12.	14.	16.	19.	27.
1.65	6.	7.	8.	8.	9.	10.	12.	14.	15.	18.	27.
1.70	6.	7.	7.	8.	9.	10.	12.	14.	15.	18.	27.
1.75	6.	7.	7.	8.	9.	10.	12.	13.	15.	18.	27.
1.80	6.	6.	7.	8.	9.	10.	12.	13.	15.	18.	27.
1.85	6.	6.	7.	8.	9.	10.	12.	13.	15.	18.	27.
1.90	5.	6.	7.	8.	9.	10.	11.	13.	15.	18.	27.
1.95	5.	6.	7.	8.	9.	10.	11.	13.	15.	18.	27.
2.00	5.	6.	7.	8.	9.	10.	11.	13.	15.	18.	27.

Table 4

Sample Sizes Necessary To Detect Various Effect Sizes With Power At .80 and Alpha At .025

Effect Size	Number Of Repeated Measures (K) 2	3	4	5	6	7	8	9	10	20	30
0.05	1869.	1519.	1285.	1123.	1005.	914.	843.	785.	736.	491.	395.
0.10	470.	383.	325.	285.	256.	235.	217.	203.	192.	136.	118.
0.15	210.	173.	147.	130.	118.	109.	101.	96.	91.	71.	67.
0.20	120.	99.	85.	76.	69.	65.	61.	58.	56.	48.	51.
0.25	78.	65.	57.	51.	47.	44.	42.	41.	40.	38.	43.
0.30	55.	46.	41.	37.	35.	33.	32.	31.	31.	33.	40.
0.35	41.	35.	32.	29.	28.	27.	26.	26.	26.	30.	37.
0.40	32.	28.	25.	24.	23.	22.	22.	22.	22.	28.	36.
0.45	26.	23.	21.	20.	20.	20.	20.	20.	20.	27.	35.
0.50	22.	20.	18.	19.	17.	17.	18.	18.	19.	26.	34.
0.55	19.	17.	16.	16.	16.	16.	16.	17.	17.	25.	34.
0.60	16.	15.	14.	14.	14.	15.	15.	16.	16.	24.	33.
0.65	14.	13.	13.	13.	13.	14.	14.	15.	16.	24.	33.
0.70	13.	12.	12.	12.	13.	13.	14.	14.	15.	24.	33.
0.75	12.	11.	11.	12.	12.	13.	13.	14.	15.	23.	33.
0.80	11.	10.	11.	11.	11.	12.	13.	14.	14.	23.	33.
0.85	10.	10.	10.	10.	11.	12.	12.	13.	14.	23.	32.
0.90	9.	9.	10.	10.	11.	11.	12.	13.	14.	23.	32.
0.95	8.	9.	9.	10.	10.	11.	12.	13.	13.	23.	32.
1.00	8.	8.	9.	9.	10.	11.	12.	12.	13.	22.	32.
1.05	8.	8.	8.	9.	10.	11.	11.	12.	13.	22.	32.
1.10	7.	8.	8.	9.	10.	10.	11.	12.	13.	22.	32.
1.15	7.	7.	8.	9.	9.	10.	11.	12.	13.	22.	32.
1.20	7.	7.	8.	9.	9.	10.	11.	12.	13.	22.	32.
1.25	6.	7.	8.	8.	9.	10.	11.	12.	13.	22.	32.
1.30	6.	7.	7.	8.	9.	10.	11.	12.	12.	22.	32.
1.35	6.	7.	7.	8.	9.	10.	11.	11.	12.	22.	32.
1.40	6.	6.	7.	8.	9.	10.	11.	11.	12.	22.	32.
1.45	6.	6.	7.	8.	9.	10.	10.	11.	12.	22.	31.
1.50	5.	6.	7.	8.	9.	9.	10.	11.	12.	22.	31.
1.55	5.	6.	7.	8.	9.	9.	10.	11.	12.	22.	31.
1.60	5.	6.	7.	8.	8.	9.	10.	11.	12.	22.	31.
1.65	5.	6.	7.	7.	8.	9.	10.	11.	12.	22.	31.
1.70	5.	6.	7.	7.	8.	9.	10.	11.	12.	22.	31.
1.75	5.	6.	7.	7.	8.	9.	10.	11.	12.	21.	31.
1.80	5.	6.	6.	7.	8.	9.	10.	11.	12.	21.	31.
1.85	5.	6.	6.	7.	8.	9.	10.	11.	12.	21.	31.
1.90	5.	5.	6.	7.	8.	9.	10.	11.	12.	21.	31.
1.95	4.	5.	6.	7.	8.	9.	10.	11.	12.	21.	31.
2.00	4.	5.	6.	7.	8.	9.	10.	11.	12.	21.	31.

Table 5

Sample Sizes Necessary To Detect Various Effect Sizes With Power At .80 And Alpha At .05

Effect Size	Number Of Repeated Measures (K) 2	3	4	5	6	7	8	9	10	20	30
0.05	1535.	1269.	1082.	951.	855.	780.	672.	598.	542.	482.	378.
0.10	385.	320.	274.	242.	218.	200.	174.	157.	144.	131.	110.
0.15	173.	144.	124.	111.	101.	93.	82.	75.	71.	66.	61.
0.20	99.	83.	72.	65.	59.	55.	50.	47.	45.	44.	44.
0.25	64.	54.	48.	43.	40.	38.	35.	34.	33.	34.	37.
0.30	45.	39.	35.	32.	30.	29.	27.	27.	27.	28.	34.
0.35	34.	30.	27.	25.	24.	23.	23.	23.	24.	25.	32.
0.40	26.	23.	22.	20.	20.	20.	20.	20.	21.	23.	30.
0.45	21.	19.	18.	17.	17.	17.	18.	19.	20.	22.	29.
0.50	19.	16.	16.	15.	15.	15.	16.	17.	19.	21.	29.
0.55	15.	14.	14.	14.	14.	14.	15.	16.	18.	20.	28.
0.60	13.	13.	12.	12.	13.	13.	14.	16.	17.	20.	28.
0.65	12.	11.	11.	12.	12.	12.	14.	15.	17.	19.	28.
0.70	10.	10.	10.	11.	11.	12.	13.	15.	16.	19.	27.
0.75	9.	10.	10.	10.	11.	11.	13.	14.	16.	19.	27.
0.80	9.	9.	9.	10.	10.	11.	13.	14.	16.	19.	27.
0.85	9.	8.	9.	9.	10.	11.	12.	14.	16.	18.	27.
0.90	7.	8.	8.	9.	10.	10.	12.	14.	15.	18.	27.
0.95	7.	7.	8.	9.	9.	10.	12.	14.	15.	18.	27.
1.00	7.	7.	8.	8.	9.	10.	12.	13.	15.	18.	27.
1.05	6.	7.	7.	8.	9.	10.	11.	13.	15.	18.	27.
1.10	6.	7.	7.	8.	9.	10.	11.	13.	15.	19.	26.
1.15	6.	6.	7.	8.	9.	9.	11.	13.	15.	18.	26.
1.20	5.	6.	7.	8.	9.	9.	11.	13.	15.	18.	26.
1.25	5.	6.	7.	8.	8.	9.	11.	13.	15.	18.	26.
1.30	5.	6.	7.	7.	8.	9.	11.	13.	15.	18.	25.
1.35	5.	6.	7.	7.	8.	9.	11.	13.	15.	17.	26.
1.40	5.	6.	6.	7.	8.	9.	11.	13.	15.	17.	25.
1.45	5.	5.	6.	7.	8.	9.	11.	13.	14.	17.	26.
1.50	5.	5.	6.	7.	8.	9.	11.	13.	14.	17.	26.
1.55	4.	5.	6.	7.	8.	9.	11.	12.	14.	17.	26.
1.60	4.	5.	6.	7.	8.	9.	11.	12.	14.	17.	26.
1.65	4.	5.	6.	7.	8.	9.	11.	12.	14.	17.	26.
1.70	4.	5.	6.	7.	8.	9.	10.	12.	14.	17.	25.
1.75	4.	5.	6.	7.	8.	9.	10.	12.	14.	17.	26.
1.80	4.	5.	6.	7.	8.	9.	10.	12.	14.	17.	25.
1.85	4.	5.	6.	7.	8.	9.	10.	12.	14.	17.	25.
1.90	4.	5.	6.	7.	8.	8.	10.	12.	14.	17.	25.
1.95	4.	5.	6.	7.	7.	9.	10.	12.	14.	17.	26.
2.00	4.	5.	6.	7.	7.	9.	10.	12.	14.	17.	26.

Table 6

Sample Sizes Necessary To Detect Various Effect Sizes With Power At .80 And Alpha At .10

Effect Size	Number Of Repeated Measures (K)									
	2	3	4	5	6	7	8	9	10	20	30
0.05	1193.	1010.	871.	770.	695.	637.	552.	492.	448.	399.	316.
0.10	300.	255.	220.	196.	178.	164.	143.	130.	120.	109.	93.
0.15	134.	115.	100.	90.	92.	76.	69.	53.	59.	56.	53.
0.20	76.	66.	58.	52.	49.	46.	42.	40.	38.	38.	40.
0.25	50.	43.	39.	35.	33.	32.	30.	29.	29.	29.	34.
0.30	35.	31.	28.	26.	25.	24.	23.	23.	24.	25.	31.
0.35	26.	24.	22.	20.	20.	19.	19.	20.	21.	23.	30.
0.40	21.	19.	18.	17.	17.	17.	17.	18.	19.	21.	29.
0.45	17.	16.	15.	14.	14.	15.	15.	17.	18.	20.	28.
0.50	14.	13.	13.	13.	13.	13.	14.	16.	17.	20.	29.
0.55	12.	11.	11.	11.	12.	12.	13.	15.	16.	19.	27.
0.60	10.	10.	10.	11.	11.	12.	13.	14.	16.	19.	27.
0.65	9.	9.	9.	10.	10.	11.	12.	14.	16.	19.	27.
0.70	8.	9.	9.	9.	10.	10.	12.	14.	15.	18.	27.
0.75	7.	8.	8.	9.	9.	10.	12.	13.	15.	18.	26.
0.80	7.	7.	8.	8.	9.	10.	11.	13.	15.	18.	26.
0.85	6.	7.	7.	8.	9.	10.	11.	13.	15.	18.	26.
0.90	6.	6.	7.	8.	9.	9.	11.	13.	15.	19.	26.
0.95	6.	6.	7.	8.	8.	9.	11.	13.	15.	17.	26.
1.00	5.	6.	7.	7.	8.	9.	11.	13.	14.	17.	26.
1.05	5.	6.	6.	7.	8.	9.	11.	13.	14.	17.	26.
1.10	5.	6.	6.	7.	8.	9.	11.	12.	14.	17.	26.
1.15	5.	5.	6.	7.	8.	9.	11.	12.	14.	17.	26.
1.20	4.	5.	6.	7.	8.	9.	10.	12.	14.	17.	26.
1.25	4.	5.	6	7.	8.	9.	10.	12.	14.	17.	26.
1.30	4.	5.	6.		8.	8.	10.	12.	14.	17.	26.
1.35	4.	5.	6.	7.	7.	9.	10.	12.	14.	17.	26.
1.40	4.	5.	6.	7.	7.	8.	10.	12.	14.	17.	26.
1.45	4.	5.	6.	6.	7.	8.	10.	12.	14.	17.	26.
1.50	4.	5.	6.	6.	7.	8.	10.	12.	14.	17.	26.
1.55	4.	5.	5.	6.	7.	8.	10.	12.	14.	17.	26.
1.60	4.	5.	5.	6.	7.	8.	10.	12.	14.	17.	26.
1.65	3.	4.	5.	6.	7.	9.	10.	12.	14.	17.	26.
1.70	3.	4.	5.	6.	7.	8.	10.	12.	14.	17.	26.
1.75	3.	4.	5.	6.	7.	8.	10.	12.	14.	17.	26.
1.80	3.	4.	5.	6.	7.	9.	10.	12.	14.	17.	26.
1.85	3.	4.	5.	6.	7.	9.	10.	12.	14.	17.	25.
1.90	3.	4.	5.	6.	7.	9.	10.	12.	14.	17.	26.
1.95	3.	4.	5.	6.	7.	9.	10.	12.	14.	17.	26.
2.00	3.	4.	5.	6.	7.	9.	10.	12.	14.	17.	26.

significance set at .005, .01, .025, .05, .1ˀ, and .20. Within each level of significance, sample sizes are tabled by the expected correlations among the repeated measures, and within these correlations, for the "small", "medium" and "large" repeated measures effect sizes shown in the rows of Table 1.

Example. Suppose that you are planning an exploratory repeated measures analysis with four repeated measures and that you are planning to set your level of significance at .05. In this case if you expect a correlation of .30 among your measures and you are interested in detecting a "medium" effect size (.30), Table 2 indicates that you should select 35 units. However, if you expect the correlation among your repeated measures to be .50, the medium effect size increases to .35, so that you now need only 27 units.

General Effect Sizes: Tables 3 through 6

Tables 3 through 6 were respectively based on the .01, .025, .05 and .10 levels of significance. Effect sizes in these tables are varied from .05 to 2.00 in increments of .05. An effect size can be choosen directly from these tables, but you should keep in mind that the effect size that you select should be larger than one found in an ANOVA with K independent levels. This suggests that an approach to selecting an effect size for these tables is to first select one of Cohen's effect sizes for an ANOVA, i.e., for a design where the measures are uncorrelated, and then divide this effect size by $\sqrt{1-\rho}$.

Example. Suppose that you are planning an exploratory repeated measures analysis with five repeated measures and that you are planning to set your level of significance at .01. Suppose also that you feel that a large effect size is possible and that the correlation among your measures will be about .85. In this case you would enter Table 3 with K = 5 and an effect size of $.40/\sqrt{1-.85} = 1.03 \approx 1.00$; here you find that you need 11 units.

Examples

Davidson's Three Cases

Consider a researcher who is planning to carry out a single group exploratory repeated measures design with three treatment levels at a .05 level of significance. Suppose also that a correlation of .80 was expected among the measures and that a "large" effect size was anticipated. Using Table 2 with the level of significance

= .05, ρ = .80, effect size = .89 and K = 3, the researcher finds that he needs 8 units. However, because 10 units can be easily sampled, he decides to take 10 units so that he can expect to have power slightly higher than .80.

The data in Table 7 represent three different possible results. These data were taken from Davidson (1972, p. 450, Cases B, C, and D) with the last measure, X3, in Case B modified here to dramatize the differences between the univariate and multivariate tests. The variance-covariance matrix of the measures is the same across cases (see Davidson's Table 5), however, each case has different differences between its repeated measures means. The Greenhouse-Geisser measure of circularity for each case is .5247, and the intraclass correlation for each case is .8572.

Table 7
Three Repeated Measures Data Sets Which Yield Different Significance Test Results[a]

	CASE B			CASE C			CASE D		
Subject	X1	X2	X3[b]	X1	X2	X3	X1	X2	X3
1	49	53	91	52	50	71	51	51	92
2	53	49	111	56	46	91	55	47	112
3	63	65	65	66	62	45	65	63	66
4	37	33	35	40	30	15	39	31	36
5	39	39	59	42	36	39	41	37	60
6	43	51	87	46	48	67	45	49	88
7	43	47	25	46	44	5	45	45	26
8	49	45	47	52	42	27	51	43	48
9	65	65	105	68	62	85	67	63	106
10	59	53	75	62	50	55	61	31	76
Mean	50	50	70	53	47	50	52	48	71

[a]Taken from Davidson (1972, p. 450, Table 4).
[b]Davidson's X3 - 9.

The analyses of Cases B, C and D (using BMDP4V, Dixon 1983) are shown in Tables 8, 9 and 10, respectively. In the analysis of Case B, Table 8, the adjusted univariate test is significant ($F = 6.32$; $p < .0309$) and the multivariate test is not significant ($F = 2.88$; $p < .1140$). In the analysis of Case C, Table 9, the adjusted univariate test is not significant ($F = .43$; $p < .5389$) and the multivariate test is significant ($F = 7.84$; $p < .0130$). In the analysis of Case D, Table 10, both the adjusted univariate ($F = 7.15$; $p < .0235$) and the multivariate ($F =$

Table 8
BMDP4V Output For Davidson's Modified Case B

```
--------------------------------------------------------------------------------
================================================================================

WITHIN EFFECT: D: DAVID

EFFECT   VARIATE          STATISTIC                      F        DF          P
--------------------------------------------------------------------------------

  D

       DEP_VAR
               TSQ=            6.48710              2.88      2,     8  0.1140
            WCP SS=            2666.67
            WCP MS=            1333.33              6.32      2,    18  0.0084
            GREENHOUSE-GEISSER ADJ. DF              6.32   1.05,  9.45  0.0309
            HUYNH-FELDT ADJUSTED DF                 6.32   1.07,  9.62  0.0301

 ERROR
       DEP_VAR
            WCP SS=           3800.0000
            WCP MS=           211.11111

            GGI EPSILON=        0.52474
            H-F EPSILON=        0.53423

--------------------------------------------------------------------------------
```

Note. The original data set was modified by subtracting 9 from each subject's X3 measure.

Table 9
BMDP4V Output For Davidson's Case C Data Set

```
---------------------------------------------------------------------------------
=================================================================================
WITHIN EFFECT: D: DAVID

EFFECT   VARIATE       STATISTIC                   F         DF            P
---------------------------------------------------------------------------------

 D

       DEP_VAR
                TSQ=         17.6484              7.84      2,       8  0.0130
            WCP SS=          180.000
            WCP MS=          90.0000              0.43      2,      18  0.6593
            GREENHOUSE-GEISSER ADJ. DF            0.43   1.05,    9.45  0.5389
            HUYNH-FELDT ADJUSTED DF               0.43   1.07,    9.62  0.5422

 ERROR
       DEP_VAR
            WCP SS=          3800.0000
            WCP MS=          211.11111

            GGI EPSILON=       0.52474
            H-F EPSILON=       0.53423

---------------------------------------------------------------------------------
```

Table 10
BMDP4V Output For Davidson's Case D Data Set

WITHIN EFFECT: D: DAVID

```
EFFECT   VARIATE        STATISTIC                    F         DF          P
----------------------------------------------------------------------------

  D

       DEP_VAR
                 TSQ=        15.7065             6.98      2,      8  0.0176
              WCP SS=       3020.00
              WCP MS=       1510.00              7.15      2,     18  0.0052
              GREENHOUSE-GEISSER ADJ. DF         7.15   1.05,   9.45  0.0235
              HUYNH-FELDT ADJUSTED DF            7.15   1.07,   9.62  0.0228

 ERROR
       DEP_VAR
              WCP SS=       3800.0000
              WCP MS=       211.11111

              GGI EPSILON=    0.52474
              H-F EPSILON=    0.53423
----------------------------------------------------------------------------
```

6.98; p < .0176) tests are significant.

Myers' Data

In this example we consider a researcher who is planning to conduct an exploratory repeated measures analysis using response time scores with a .05 level of significance and three responses per subject. This researcher expects a very high correlation among the responses, i.e., .999, and a large effect size. She calculates an effect size of $.40/\sqrt{1-.999} = 12.65$ which is not in Table 5. She therefore executes the program provided by Robey and Barcikowski (1984) and finds that her power will be .95 if she uses a sample size of three subjects. [This example is a bit bizarre, but it illustrates some interesting points.]

The data in Table 11 were taken from Myers' (1979, p. 175) to illustrate the results of this study. The analysis of these data is shown in Table 12. The results show that neither the adjusted univariate test (F = 2.87; p < .2312) nor the mutivariate test (F = .75; p < .6329) is significant. However, after plotting these data, the researcher found an ordinal interaction and she decided to remove this interaction by taking the reciprocal of each score. [See Myers (1979, Chapter 7) for a discussion of the analysis of repeated measures data when there is an interaction between the units and the repeated measures.]

Table 11
Myers' Data

1.7	1.9	2.0
4.4	4.5	5.7
6.6	7.4	10.5

The results of the analysis of the reciprocals of the data in Table 11 are shown in Table 13. These results indicate that the adjusted univariate test is not significant (F = 13.79; p < .0649), but that the multivariate test is significant (F = 711.77; p < .0265).

Examples: Summary

The preceding example analyses illustrate the attractiveness of the repeated measures sample size selection rationale described in this paper. Based on informed judgement of the expected correlation among the repeated measures and of the expected effect size, the researchers in the examples were able to select an

Table 12
BMDP4V Output For Myers' Example Data Set

```
-----------------------------------------------------------------------------------------
=========================================================================================

WITHIN EFFECT: M: MYERS

EFFECT   VARIATE        STATISTIC                        F        DF          P
-----------------------------------------------------------------------------------------

 M

        DEP_VAR
                TSQ=          2.99349                  0.75      2,       1  0.6329
             WCP SS=          5.64667
             WCP MS=          2.82333                  2.87      2,       4  0.1686
             GREENHOUSE-GEISSER ADJ. DF                2.87   1.01,    2.03  0.2312
             HUYNH-FELDT ADJUSTED DF                   2.87   1.05,    2.11  0.2280

 ERROR
        DEP_VAR
             WCP SS=          3.9333340
             WCP MS=         0.98333349

             GGI EPSILON=      0.50671
             H-F EPSILCN=      0.52720
-----------------------------------------------------------------------------------------
```

Table 13
BMDP4V Output For Myers' Example Data Set After Reciprocal Transformation

```
-----------------------------------------------------------------------------------------------

===============================================================================================

WITHIN EFFECT: M: MYERS

EFFECT    VARIATE           STATISTIC                      F         DF            P
-----------------------------------------------------------------------------------------------

  M

          DEP_VAR
                TSQ=          2847.07              711.77       2,         1  0.0265
             WCP SS=        0.647415D-02
             WCP MS=        0.323708D-02            13.79       2,         4  0.0161
             GREENHOUSE-GEISSER ADJ. DF             13.79    1.01,      2.01  0.0649
             HUYNH-FELDT ADJUSTED DF                13.79    1.03,      2.05  0.0630

 ERROR
          DEP_VAR
             WCP SS=        0.93924449D-03
             WCP MS=        0.23481112D-03

             GGI EPSILON=     0.50323
             H-F EPSILON=     0.51302

-----------------------------------------------------------------------------------------------
```

appropriate sample size using the tables and/or equations presented here.

Two points should be emphasized however. First, as we have stated elseware: "decriptive analysis of repeated measures data such as examination of the structure of the covariance matrix, scatterplots for pairs of responses, and trend curves is often invaluable." (Barcikowski and Robey, 1984b, p. 150). This point was emphasized in the example using Myers' (1979) data set. Significant results would not have been found had the researcher only conducted significance tests and had not considered scatter plots of the responses.

Second, the importance of conducting both the adjusted univariate and the multivariate statistical tests was demonstrated. In Case B of the Davidson data, if the univariate test had not been conducted, the multivariate test by itself would have found no significant result, and in Case C, if the multivariate test had not been conducted, the univariate test by itself would have found no significant result. Also, with the transformed Myers' data the adjusted univariate test was not significant, but the multivariate test was significant.

Educational and Scientific Importance of the Study

The advantages of having sufficient sample size to achieve a desired level of statistical power in an experiment are generally recognized (Cohen, 1977). Cohen (1977) provides many tables to determine sample size in factorial analyses of variance. However, similar tables for repeated measures analyses are generally not available. This paper provides researchers with the methodology to find appropriate sample sizes in single group exploratory repeated measures designs, and includes sample size tables for minimum power (.80).

References

Barcikowski, R.S., & Robey, R. R. (1984a). Exploratory repeated measures analysis. Paper presented at the annual meeting of the American Educational Research Association, New Orleans, April.

Barcikowski, R. S., & Robey, R. R. (1984b). Decisions in single group repeated measures analysis: Statistical tests and three computer packages. The American Statistician, 38, 148-150.

Cohen, J. (1977). Statistical power analysis for the behavioral sciences, (2nd ed.). New York: Academic Press.

Collier, R. O., JR., Baker, F. B., Mandeville, G. K., & Hayes, T. F. (1967). Estimates of test size for several test procedures on conventional variance ratios in the repeated measures design. Psychometrika 32, 339-353.

Davidson, M. L. (1972). Univariate versus multivariate tests in repeated-measurements experiments. Psychological Bulletin, 77, 446-452.

Dixon, J. W. (Ed.) (1983). BMDP statistical software, 1983 edition. Los Angeles: University of California Press.

Green, P. E., & Carroll, J. D. (1976). Mathematical tools for applied multivariate analysis. New York: Academic Press.

Greenhouse, S. W., & Geisser, S. (1959). On methods in analysis of profile data. Psychometrika, 24, 95-112.

Huynh, H., & Feldt, L. S. (1970). Conditions under which mean square ratios in repeated-measurement designs have exact F-distributions. Journal of the American Statistical Association, 65, 1582-1589.

Imhof, J. P. (1962). Testing the hypothesis of no fixed main-effects in Scheffé's mixed model. The Annals of Mathematical Statistics, 33, 1085-1095.

Jensen, D. R. (1982). Efficiency and robustness in the use of repeated measurements. Biometrics, 38, 813-825.

Maxwell, S. E. (1980). Pairwise multiple comparisons in repeated measures designs. Journal of Educational Statistics, 5, 269-287.

Maxwell, S.E., Delaney, D. D., & Sternitzke, M. E. (1984). Complex comparisons in repeated measures designs. Paper presented at the annual meeting of the American Educational Research Association, New Orleans, April.

Myers, J. L. (1979). Fundamentals of experimental design (3rd ed.). Boston: Allyn and Bacon.

Robey, R. R., & Barcikowski, R. S. (1984). Calculating the statistical power of the univariate and the multivariate repeated measures analysis of variance for the single group case under various conditions. Educational and Psychological Measurement, 44, 137-143.

Rogan, J. C., Keselman, H. J., & Mendoza, J. L. (1979) Analysis of repeated measurements. British Journal of Mathematical and Statistical Psychology, 32, 269-286.

Rouanet, H., & Lépine, D. (1970). Comparison between treatments in a repeated-measurement design: ANOVA and multivariate methods. British Journal of Mathematical and Statistical Psychology, 23, 147-163.

Winer, B. J. (1971) Statistical principles in experimental design (2nd ed.). New York: McGraw-Hill.

Appendix A
Single Group Univariate And Multivariate Effect Sizes

Noncentrality Parameters

The Univariate Noncentrality Parameter

We have shown (Barcikowski and Robey 1984a) that the univariate noncentrality parameter, δ^2, in a single group repeated measures design can be written as:

$$\delta^2_U = \frac{n(K-1)\sum_{i=1}^{K-1}\psi_i^2}{\sum_{i=1}^{K-1}\sigma^2_{\psi_i}} \tag{1}$$

where, K = the number of repeated measures, n = the number of units (subjects), ψ_i is the i^{th} orthonormal contrast among the population repeated measures means, and $\sigma^2_{\psi_i}$ is the variance of the i^{th} contrast.

The Multivariate Noncentrality Paramenter

The multivariate noncentrality paramenter noncentrality parameter is written (Morrison, 1967, p.150) as:

$$\delta^2_M = n\,\underline{u}'\,C'(C\,\Sigma\,C')^{-1}C\,\underline{u} \tag{2}$$

where, n is the number of units, $\underline{u}$ is a column vector of the repeated measures population means, C is a nonsingular matrix of (K-1) by K contrast coefficients, Σ is the nonsingular variance-covariance matrix of the multivariate normal distribution from which the repeated measures are selected.

In terms of contrasts, Equation 2 becomes:

$$\delta^2_M = n\,\underline{\Psi}'\,(C\,\Sigma\,C')^{-1}\,\underline{\Psi} \tag{3}$$

where, $\underline{\Psi}$ is a vector of (K-1) contrasts on the population means of the repeated measures.

Now, if we select the rows of C in Equation 3 to be orthonormal contrast coefficients such that C Σ C' is a diagonal matrix, then the diagonal elements of $C \Sigma C'$ will be the variances of the mean contrasts in $\underline{\Psi}$ (Green and Douglas, 1976, Chapter 5). Then, $\sigma_{\psi_i}^2$ is the i^{th} contrast variance, and Equation 3 can be written as:

$$\delta^2_M = n \sum_{i=1}^{K-1} \frac{\psi_i^2}{\sigma^2_{\psi_i}} \qquad (4)$$

Effect Sizes

Cohen (1977) determines power using a function of the noncentrality parameter which he calls "effect size". Effect size, f, can be written in terms of the preceding noncentrality parameters as:

$$f = \sqrt{\delta^2/(nK)} \qquad (5)$$

Univariate Effect Size

Substituting Equation 1 into Equation 5 we have that Cohen's effect size, f_U, for the univariate case is:

$$f_U = \sqrt{\frac{(K-1) \sum_{i=1}^{K-1} \psi_i^2}{K \sum_{i=1}^{K-1} \sigma_{\psi_i}^2}} \qquad (6)$$

This effect size is used with a noncentral F having (n-1) and (n-1)(K-1) degrees of freedom to determine power.

Multivariate Effect Size

Substituting Equation 4 into Equation 5 we have that Cohen's effect size, f_M, for the multivariate case is:

$$f_M = \sqrt{\frac{1}{K} \sum_{i=1}^{K-1} \frac{\psi_i^2}{\sigma^2_{\psi_i}}} \qquad (7)$$

28

This effect size is used with a noncentral F having (n-1) and (n-K+1) degrees of freedom to determine power.

In the following sections the univariate and multivariate effect sizes shown in the latter two equations will be considered under special conditions.

Single Degree Of Freedom Contrasts

For a single contrast, and using a little algebra, both Equation 6 and Equation 7 become

$$f_U = f_M = \sqrt{\frac{\psi_i^2}{2\,\sigma_{\psi_i}^2}} \qquad (8)$$

This is the effect size that is used with the noncentral F distribution having 1 and (n-1) degrees of freedom to determine power.

It is of interest to compare the single degree of freedom effect size in Equation 8 with the omnibus multivariate effect size in Equation 7, since the tables presented in this paper are based on the latter test. In so doing we made the following conclusions where each conclusion was reached independent of the others. However, in each conclusion we have kept in mind the fact that the contrasts in Equations 7 and 8 would probably not be the same. This is because multivariate contrasts are a special type of orthonormal contrast, while the single degree of freedom contrasts would probably be "obvious" contrasts of interest.

1) The single degree of freedom effect size is used with a noncentral F having fewer numerator degrees of freedom (1 versus K-1) but slightly larger denominator degrees of freedom (n-1 versus n-K+1) then the multivariate effect size. Given that n = K + 20 (Davidson, 1972), as K increases, the omnibus multivariate test will generally be more powerful than some of the single degree of freedom tests. However, for values of n close to K, the single degree of freedom tests will generally be more powerful.

2) The single degree of freedom effect size has a two in its denominator while the multivariate effect size has a K in its denominator. In general the single degree of freedom contrasts will have larger effect sizes. This will be especially true when large contrasts have been chosen.

3) As the number of single degree of freedom contrasts tested in a study increases, the per contrast level of significance must decrease if one is to maintain control of the experimentwise level of significance. However, the level of significance for the multivariate test remains at a single "wholesome" value. As the number of contrasts increases, the decrease in the per contrast level of significance will tend to give a power advantage to the multivariate test.

<u>General conclusion #1.</u> In general for a small number of contrasts, particularly for small n, the single degree of freedom contrasts will be more powerful than the omnibus Hotelling's T^2 test. Indeed, for $n < K$, the single degree of freedom contrasts represent a very attractive test strategy since the multivariate test cannot be done. Therefore, the power tables provided in this paper for the omnibus multivariate test should provide a conservative estimate of sample size for the single degree of freedom testing strategy.

<u>General conclusion #2.</u> In using the single degree of freedom testing strategy, a sample size could be estimated from the tables provided in this paper by choosing the contrast with the smallest effect size, dividing the experimentwise level of significance by the number of contrasts, and then using the table with the resulting level of significance (or close to it) with K set at two.

Under Circularity

When the circularity assumption is meet, all of the contrast variances on the diagonal of the matrix $C \Sigma C'$ are equal. Under this condition the univariate and multivariate effect sizes are equal. That is, using a little algebra Equations 6 and 7 become:

$$f_U = f_M = \sqrt{\frac{\sum_{i=1}^{K-1} \psi_i^2}{K \sigma_{\psi_i}^2}} \qquad (9)$$

Under Uniformity

Under uniformity the contrast variances on the diagonal of the matrix $C \Sigma C'$ are all equal; the variances of the original measures are all equal; and the correlations among the measures are all equal. Under this condition Davidson (1972, p. 448) shows that the noncentrality parameter for the univariate and for the multivariate tests is:

$$\delta^2_U = \delta^2_M = \frac{n \sum_{i=1}^{K} (\mu_i - \mu.)^2}{\sigma^2 (1-\rho)} \qquad (10)$$

Here, σ^2 is the population variance of each measure, μ_i is the mean of measure i, $\mu.$ is the overall population mean, and ρ is the common population correlation among the measures.

Substituting Equation 10 into Equation 5 we have the univariate and multivariate effect sizes under uniformity are:

$$f_U = f_M = \sqrt{\frac{\sum_{i=1}^{K} (\mu_i - \mu.)^2}{K\,\sigma^2 (1-\rho)}} \qquad (11)$$

Equation 11 is Cohen's (1977, p. 275) effect size for a one-way analysis of variance with K independent groups divided by $\sqrt{1-\rho}$.