

Statistical Tests of Group Differences in ALSCAL-Derived Subject Weights: Some Monte Carlo Results

Joseph Lee Rodgers
University of Oklahoma

Several techniques to test for group differences in weighted multidimensional scaling (MDS) subject weights have recently been proposed. The present study presents monte carlo results to evaluate the operating characteristics of two of these with ALSCAL-derived subject weights. The first uses the analysis of angular variation (ANAVA) on raw subject weights. The second applies the analysis of variance (ANOVA) to the flattened subject weights generated by ALSCAL. The ANOVA on flattened weights was less affected by the presence of error and by distortions caused by ALSCAL's normalization routine than was the ANAVA.

Multidimensional scaling (MDS) models have traditionally been concerned with subjects as replications (classical MDS), subjects as points in the stimulus space (unfolding), or subjects as vectors in a weight space (weighted MDS). A logical extension is to consider groups of subjects, and to ask questions about group differences. Recently, a statistical approach to test for group differences in MDS subject spaces was presented. Schiffman, Reynolds, and Young (1981) applied Mardia's (1972, 1975) statistics of directional data to test for differences between two or more groups in the same way that the analysis of variance (ANOVA) tests for group differences in linear data. The need for two different techniques arises because of dif-

ferences in the raw data; subject weights are most appropriately interpreted as angles (MacCallum, 1977), whereas ANOVA data are represented as points. Distances between points may be very different from angles between vectors (however, see Jones, 1983, who nevertheless recommended using linear methods on subject weight angles).

As developed by Mardia (1972) and explained by Schiffman et al. (1981) in the context of MDS data, the ANOVA theory applicable to angles (ANAVA) is based on the von Mises distribution, the closest counterpart on the unit circle to the normal distribution on the number line. The mean direction is defined by averaging the unit normalized weight vectors on each dimension. The length of the mean direction vector is denoted \bar{R} . The resultant length is defined to be $R = m\bar{R}$, where m is the number of subjects. The most useful measure of angular variation is a value \bar{S} , which equals $1 - \bar{R}$. Mardia (1972) showed how the total sum of squared angular deviations— $m\bar{S}$ —could be partitioned into within and between group variation, and he used the work of Watson and Williams (1956) to show further that the ratio of mean square between to mean square within is distributed as F with $(r - 1)(q - 1)$ and $(r - 1)(m - q)$ degrees of freedom when $\bar{R} > .67$ (r = number of dimensions, q = number of groups, and m = total sample size). Mardia's development assumed that the points were distributed as von Mises across the unit circle, and assumed the counterpart of homogeneity of variance.

APPLIED PSYCHOLOGICAL MEASUREMENT
Vol. 9, No. 3, September 1985, pp. 241-248
© Copyright 1985 Applied Psychological Measurement Inc.
0146-6216/85/030241-08\$1.65

Several problems exist in the application of ANAVA to subject weights, however. These were discussed in detail in Schiffman et al. (1981) and Jones (1983). The most relevant problem for the purposes of this study is that ANAVA results obtained from MDS subject spaces are not invariant across different normalization procedures. Since different procedures scale the points with different means and variances, and in particular, the lengths of the subject vectors have different meanings across MDS algorithms, the ANAVA may be influenced by what is an entirely arbitrary decision in the choice of normalization procedures.

Schiffman et al. (1981) and Young (1982) developed a technique that may preclude the necessity of using ANAVA. In their method, subject weights are "flattened" into a space of dimensionality one less than the original subject space. As presented in Young (1982) and produced in the SAS version of ALSCAL, flattened weights are derived by normalizing each subject's weights across dimensions to sum to one, dropping the last dimension, and then standardizing the remaining dimensions to have a mean of zero and variance of one. This has the effect of flattening arcs onto a line that is 45 degrees to all dimensions. Young (1982) suggested that these may be used with normal linear procedures, so that problems in the use and interpretation of ANAVA may be ignored altogether.

The purpose of the present study was to investigate the operating characteristics associated with the use of ANAVA on raw weights and ANOVA on flattened weights. Monte carlo procedures were used to assess the empirical levels of alpha and power in both null and noncentral cases. The procedure involved (1) generating well-defined subject weights, (2) using those to define subject spaces, (3) analyzing the data with ALSCAL, and (4) testing for group differences in the raw and flattened weights.

Because of the problems mentioned above—particularly the arbitrary normalization procedure—monte carlo methods are the only practical way to assess the operating characteristics of these techniques. It should be noted that only the normalization procedure used by ALSCAL was considered. The behavior of other normalization procedures in

other algorithms was not assessed, nor can the ALSCAL results be generalized to other algorithms.

Method and Results

Overview

The results reported in this article were based on a simulation written in SAS using the SAS macro language to iterate the simulation (SAS Institute, Inc., 1982, p. 441). In this simulation, samples of subject weights were generated as random von Mises deviates with known parameters. Each set of weights was applied to a common stimulus space, and the coordinate values were then perturbed with random normal error. These subject spaces were then submitted to metric ALSCAL, and two-dimensional solutions were obtained. The derived raw and flattened subject weights were analyzed with ANAVA and ANOVA, respectively.

The generation routine was based on the fact that the von Mises distribution can be closely approximated by a normal distribution (with μ = mean angle and $\sigma = 1/\sqrt{k}$, k = the concentration parameter) for large k (Mardia, 1972, p. 60). Normal deviates were generated randomly and transformed to von Mises vectors on the unit circle. A number of preliminary simulations were run to assure the quality of this approximation and the efficacy of the sample generation routine across a wide range of means and concentration parameters. Chi-square tests consistently supported that the generated angles were indeed von Mises distributed. These results were reported in detail in Rodgers (1983).

In each case, the stimulus space that was used was one obtained from Schiffman et al. (1981, p. 177, dimensions 1 and 2), and was chosen because points were nonsystematically arrayed and spread across both dimensions of the space. Both dimensions were scaled to have standard deviations of one and means of zero, so that in this case ALSCAL's normalization procedure (which scales its solution to these values on each dimension) was not taxed. In each condition, the simulated subjects were divided in half either randomly (in the null cases) or by design (in the noncentral cases). The simulation was run 100 times in each condition,

and the proportion of observed F values greater than the critical value was tabulated. These proportions corresponded to observed α in null cases and observed power in noncentral cases. They were tested statistically against theoretical α and power with a binomial test by defining a 95% confidence interval about the true value and observing whether the generated value fell inside the interval.

Level of error was defined using Young's (1970) and Takane, Young, and deLeeuw's (1977) monte carlo studies as models. Random normal error was added separately and independently to each personal space coordinate value on each dimension prior to submission to ALSICAL. This is equivalent to adding noncentral chi-square deviates directly to the distances. Error levels were defined in units of standard deviations of the original stimulus configuration.

The ratio of variances was manipulated to observe how the ALSICAL normalization routine handled true configurations in which the ratio of variances was different from unity. This tests how the two procedures handle different levels of distortion due to the normalization procedure. To achieve this, the second dimension of the original stimulus configuration was multiplied by an expansion factor that took on values of 1, 2, 4, and 8; each of these values reflected the ratio of the standard deviations, since the first dimension was not affected.

Of particular interest in the present context is how ALSICAL recovers the raw data. Table 1 shows the mean angles and concentration parameters of the ALSICAL-derived subject weights in errorless data. ALSICAL tends to move extreme angles slightly toward the center, and tends to decrease angles at the center. The concentration parameters are increased by ALSICAL, implying a slightly tighter distribution about the mean angle than in the raw data. Centralizing the mean angles would be expected to reduce power, since existing group differences are made smaller. Increased concentration would have the opposite effect.

The specific details and results of each simulation are reported together in the two sections that follow. First, results of manipulating error, expansion factors, noncentrality, and α are presented for the ANOVA. Next, the equivalent manipulations are shown for the ANOVA of flattened subject weights.

Tests of the ANOVA

The first simulation to test the ANOVA involved generating data for 20 subjects, who were randomly divided into two groups of 10. The manipulations included in this null case were error level, expansion factor, and α level. Five levels of error were included, 0, .05, .1, .2, and .5. Four levels

Table 1
ALSICAL-Derived Mean Angles and Concentrations, Given
Population Mean Angle and Concentration (N=25 in Each Cell)

Mean Angle	Concentration Parameter (k)		
	4	8	16
$\mu_0 = 22.5$	$\bar{x}_0 = 25.6$	$\bar{x}_0 = 23.0$	$\bar{x}_0 = 23.5$
	$k_{obs} = 7.5$	$k_{obs} = 11.7$	$k_{obs} = 20.6$
$\mu_0 = 45$	$\bar{x}_0 = 40.9$	$\bar{x}_0 = 41.9$	$\bar{x}_0 = 43.0$
	$k_{obs} = 5.4$	$k_{obs} = 9.7$	$k_{obs} = 18.9$

of expansion were included, 1, 2, 4, and 8. Finally, two α levels were tabulated, $\alpha = .05$ and $\alpha = .25$. The higher α level was used so that a more sensitive indicator of extreme conservatism would be available than for $\alpha = .05$. The concentration parameter was set to 8 and the mean angle to 45 degrees. The mean angle was chosen because it splits the positive quadrant, and the concentration parameter was equivalent to a standard deviation of .3536 radians or 20.3 degrees. Occasionally, angles were generated outside the interval from 0 to 90. When this occurred, the angle was set back to the boundary to prevent the possibility of imaginary distances. An extensive preliminary simulation showed that the ANOVA was highly robust to the deviations from a true von Mises distribution caused by this procedure, even with lower concentration parameters (up to standard deviations of 40 degrees).

Results of the first simulation are shown in Table 2. The observed α levels were statistically equal to the set α s for 10 of the 20 cells at the .05 level and 8 of the 20 at the .25 level; all others were conservative. The pattern of results suggests that with increasing error and/or increasing expansion factors the level of conservatism increases. By inspecting results for $\alpha = .25$, it is clear that extreme conservatism is reached for error levels of .2 and .5.

In the second ANOVA simulation, the same manipulations were used in the noncentral case. Here, the simulated subjects were separated into two groups, whose mean angles were on opposite sides and equal distances from 45 degrees. There were

two levels of noncentrality; the means were separated by either one or two standard deviations (20.3 degrees and 40.5 degrees, respectively), and the concentration parameter was set to 8. More subject weights were generated that fell outside of the first quadrant than before; again, these were placed back onto the boundary. Set α had two levels (.05 and .01), amount of error had two levels (.05 and .5), and amount of expansion had two levels (2 and 8). Finally, group size was 5 per group or 10 per group. In each cell, the power that would have been obtained using the ANOVA in an equivalent errorless setting was computed using the Pearson and Hartley charts (cf. Kirk, 1968, p. 540). A binomial test was run to determine if the observed power of the ANOVA significantly departed from that in the tables.

Results of this simulation are presented in Table 3. In the errorless case with no expansion, six of the eight powers were statistically equal to the traditional ANOVA power. Otherwise, power was consistently and significantly lower, and was particularly sensitive to increasing error level. These results are compared with those from the ANOVA on flattened weights simulations below.

Tests of the ANOVA on Flattened Weights

The two simulations to test the ANOVA on flattened weights were identical in all respects to those described above. In fact, the ANOVA and ANOVA

Table 2
 Proportion of ANOVA Rejections Out of 100 Samples with Null ALSCAL Data for Five Error Levels, Four Expansion Factors, and $\alpha=.05$ ($\alpha=.25$ in parentheses)

Error Level	Expansion Factor			
	1	2	4	8
0	.06 (.25)	.05 (.25)	.01 (.21)	.02 (.21)
.05	.05 (.24)	.03 (.23)	.01 (.16)	.00 (.17)
.1	.03 (.18)	.00 (.15)	.01 (.10)	.00 (.07)
.2	.00 (.06)	.01 (.14)	.00 (.05)	.00 (.00)
.5	.00 (.00)	.00 (.00)	.00 (.01)	.00 (.00)

Note. The 95% confidence interval about true $\alpha=.05$ is [.008, .093].
 The 95% confidence interval about true $\alpha=.25$ is [.165, .335].

Table 3
Power of ANAVA (100 Samples) for Non-Central ALSICAL Data,
Manipulating Two Levels of Expansion, Two Levels of Error,
Two Levels of Group Size, for $\alpha=.05$ ($\alpha=.01$ in parentheses)

Expansion	Error Level	N/Group	Non-Centrality	
			1 σ	2 σ
1	0	5	.35 (.18)	.75 (.45)
		10	.55 (.29)	1.0 (.94)
2	.05	5	.08 (.02)	.58 (.15)
		10	.36 (.13)	.93 (.68)
	.5	5	.00 (.00)	.08 (.01)
		10	.00 (.00)	.12 (.00)
8	.05	5	.04 (.00)	.20 (.07)
		10	.08 (.01)	.13 (.03)
	.5	5	.00 (.00)	.01 (.00)
		10	.00 (.00)	.00 (.00)
Traditional ANOVA Power ^a		5	.32 (.10)	.79 (.47)
		10	.68 (.34)	.99 (.93)

Note. The 95% confidence intervals about the traditional values were as follows:

$\alpha=.05$	$\alpha=.01$
.32 - [.229, .411]	.10 - [.041, .159]
.68 - [.589, .771]	.34 - [.247, .433]
.79 - [.710, .870]	.47 - [.372, .568]
.99 - [.970, 1.0]	.93 - [.880, .980]

^aThe traditional ANOVA powers came from the Pearson & Hartley charts.

tests were applied to the same data; the flattened subject weights were those produced from the same runs in which the raw subject weights were produced above. This was done so that the ANAVA and ANOVA solutions would be most directly comparable, and not subject to possible sampling differences.

The results of the first ANOVA simulation—with null data in which error level, expansion factor, and α were manipulated—are presented in Table 4. Results are considerably less conservative than in the ANAVA. All values fell in the 95% confidence interval defined about the set α s. Inspection of these results suggests no systematic trends across level of error. There does appear to be a slight increasing conservative bias as the expansion factor increases, however; the mean α s for set $\alpha = .05$ were .048 for an expansion factor of 1, .038 for 2, .020 for 4, and .020 for 8. For set α of .25, observed α s were .266 for 1, .266 for 2, .246 for 4, and .236 for 8.

The power of ANOVA on flattened subject weights

is shown in Table 5. Power decreased systematically for increasing expansion factor, and was relatively insensitive to error level. In the errorless case with an expansion factor of 1, seven of the eight powers were nonsignificantly different from traditional ANOVA power. In all other cases, power was significantly lower. However, attenuation of power by increasing expansion factors was less than for the equivalent ANOVA results presented in Table 3.

Discussion

Jones (1983) outlined several approaches that can be taken to analyze for group differences in the subject weights produced from individual differences MDS output. These include the following: (1) the ANAVA on the raw subject weights; (2) the analysis of the raw subject weights—angular data—as if they were linear; (3) the analysis of subject weights after they have been transformed (e.g., logarithmic transformations, see Coxon & Jones,

1978, or inverse tangent transformations, see Batschelet, 1981). Included in the third class is Young's (1982) flattening technique, in which raw subject weights are projected into a subspace of one fewer dimensions than the original.

The purpose of the present study was to investigate the operating characteristics of two of these methods that are based on ALSICAL-derived subject weights. Preliminary results presented in Rodgers (1983) showed that in the errorless case with equal standard deviations on the two dimensions, both the ANOVA of raw subject weights and the ANOVA of flattened weights preserve the set operating characteristics in both central and noncentral data. The more realistic setting is, of course, the one in which data are errorful and the normalization procedure distorts the true structure of the stimulus space.

In this more realistic setting, the ANOVA was rather conservative in the presence of large amounts of error or considerable distortion due to normalization. This was true particularly when the ratio of the standard deviations of the dimensions was greater than two, or when the level of normal error had a standard deviation greater than one-tenth of the less variable dimension. Whether the ANOVA would be useful to an applied researcher would depend on the amount of error expected and how much the normalization procedure distorts the true ratio of spreads between the dimensions. In the presence of only slight distortion in both factors, the ANOVA would be a useful technique.

The most useful result of these analyses for an applied researcher is the relative robustness in the flattened subject weights analysis. The use of the traditional linear analysis, ANOVA, on the flattened weights defined by ALSICAL was more robust than the ANOVA to both the presence of error and normalization distortion. In the null case, the technique was virtually insensitive to level of error. The technique was slightly sensitive to normalization distortion. The power of ANOVA on flattened weights was attenuated by either error or normalization distortion, as would be expected. The reduction in power was less than for the equivalent ANOVA setting, however. These results suggest that with either small or substantial error or normalization distortion, an MDS researcher will be better served analyzing the flattened subject weights than the angular data to assess group differences.

The limitations of the present analysis should be explicitly stated. First, all analyses were restricted to the two group and two-dimensional setting. The conduct of such monte carlo research is extremely expensive, so that cost prohibited more complete investigation of dimensionality and number of groups. There seems no obvious reason to believe results would not generalize to more groups and higher dimensional settings, however. Second, all MDS analyses were metric analyses; the consideration of appropriate analyses to test for group differences in the presence of monotonic (but nonlinear) transformations of the raw subject spaces will

Table 4
 Proportion of ANOVA Rejections for Flattened ALSICAL Subject Weights
 Out of 100 Samples, with Null Data for Five Error Levels,
 Four Expansion Factors, and $\alpha=.05$ ($\alpha=.25$ in parentheses)

Error Level	Expansion Factor			
	1	2	4	8
0	.05 (.24)	.05 (.25)	.01 (.23)	.03 (.25)
.05	.06 (.28)	.04 (.24)	.02 (.27)	.01 (.23)
.1	.04 (.25)	.01 (.23)	.03 (.25)	.01 (.22)
.2	.02 (.27)	.05 (.30)	.02 (.23)	.02 (.21)
.5	.07 (.29)	.04 (.31)	.02 (.25)	.03 (.27)

Note. The 95% confidence interval about true $\alpha=.05$ is [.008, .093].
 The 95% confidence interval about true $\alpha=.25$ is [.165, .335].

Table 5
Power of ANOVA on Flattened ALSICAL Subject Weights (100 Samples) for Non Central Data, Manipulating Two Levels of Expansion, Two Levels of Error, Two Levels of Group Size, for $\alpha=.05$ ($\alpha=.01$ in parentheses)

Expansion	Error Level	N/Group	Non-Centrality		
			1 σ	2 σ	
1	0	5	.35 (.14)	.76 (.46)	
		10	.54 (.28)	1.0 (.94)	
2	.05	5	.12 (.02)	.69 (.22)	
		10	.44 (.18)	.97 (.79)	
	.5	5	.14 (.02)	.63 (.35)	
		10	.42 (.19)	.91 (.82)	
8	.05	5	.13 (.03)	.23 (.10)	
		10	.15 (.04)	.24 (.07)	
	.5	5	.08 (.02)	.20 (.08)	
		10	.13 (.03)	.23 (.06)	
	Traditional ANOVA Power ^a		5	.32 (.10)	.79 (.47)
			10	.68 (.34)	.99 (.93)

Note. The 95% confidence intervals about the traditional values were as follows:

	$\alpha=.05$	$\alpha=.01$
.32	-.229, .411	.10 - [.041, .159]
.68	-.589, .771	.34 - [.247, .433]
.79	-.710, .870	.47 - [.372, .568]
.99	-.970, 1.0	.93 - [.880, .980]

^aTraditional ANOVA powers came from the Pearson & Hartley charts.

be left for future investigation. Third, only ALSICAL results were considered. Results obtained with other MDS algorithms (e.g., MULTISCALE or INDSICAL) might be different. Fourth, not all of the possible methods were investigated in the present study. The analysis of raw angular data as if they were linear, and a number of possible transformations besides the flattening procedure used in ALSICAL, might also be used. The two methods chosen for this study were those that had been previously proposed in connection with ALSICAL (Schiffman et al., 1981; Young, 1982).

Finally, the use of either ANAVA or ANOVA with MDS subject weights involves two analytic models. All of the assumptions of both the MDS model and the ANOVA or ANAVA model make it seem rather unlikely that the data really fit such a set of assumptions. Thus, when true operating characteristics are not preserved, it is not clear which of the models is inconsistent with the data. One possible solution would be to define an MDS distance model

for the express purpose of testing for group differences, in the same way that Bloxom (1968), Carroll and Chang (1970), and Horan (1969) expressly modeled individual differences.

References

- Batschelet, E. (1981). *Circular statistics in biology*. New York: Academic Press.
- Bloxom, B. (1968). *Individual differences in multidimensional scaling* (Research Bulletin 68-45). Princeton NJ: Educational Testing Service.
- Carroll, J. D., & Chang, J. J. (1970). Analysis of individual differences in multidimensional scaling via an *N*-way generalization of "Eckart-Young" decomposition. *Psychometrika*, 35, 238-319.
- Coxon, A. P. M., & Jones, C. L. (1978). *The images of occupational prestige*. London: Macmillan.
- Horan, C. B. (1969). Multidimensional scaling: Combining observations when individuals have different perceptual structures. *Psychometrika*, 34, 139-165.
- Jones, C. L. (1983). A note on the use of directional statistics in weighted Euclidean distances multidimensional scaling.

- mensional scaling models. *Psychometrika*, 48, 473–476.
- Kirk, R. E. (1968). *Experimental design: Procedures for the behavioral sciences*. Belmont CA: Brooks/Cole.
- MacCallum, R. C. (1977). Effects of conditionality on INDSCAL and ALSCAL weights. *Psychometrika*, 42, 297–305.
- Mardia, K. V. (1972). *Statistics of directional data*. New York: Academic Press.
- Mardia, K. V. (1975). Distribution theory for the von Mises-Fisher distribution. In G. D. Patil, S. Kotz, & J. K. Ortz (Eds.), *Statistical distributions in scientific work* (Vol. 1, pp. 113–130). Dordrecht, Holland: Reidel.
- Rodgers, J. L. (1983, July). *The validity of the analysis of angular variation with ALSCAL-derived subject weights: Some Monte Carlo results*. Paper presented at the European meeting of the Psychometric Society, Paris, France.
- SAS Institute, Inc. (1982). *SAS users guide: Basics*. Cary NC: Author.
- Schiffman, S. S., Reynolds, M. A., & Young, F. W. (1981). *Introduction to multidimensional scaling*. New York: Academic Press.
- Takane, Y., Young, F. W., & deLeeuw, J. (1977). Non-metric individual differences multidimensional scaling: An alternating least squares method with optimal scaling features. *Psychometrika*, 42, 7–67.
- Watson, G. S., & Williams, E. S. (1956). On the construction of significance tests on the circle and the sphere. *Biometrika*, 43, 344–352.
- Young, F. W. (1970). Nonmetric multidimensional scaling: Recovery of metric information. *Psychometrika*, 35, 455–473.
- Young, F. W. (1982). Enhancements in ALSCAL-82. *Proceedings of the Seventh Annual SAS Users Group* (pp. 633–642). Cary NC: SAS Institute, Inc.

Author's Address

Send requests for reprints or further information to Joseph Lee Rodgers, Department of Psychology, 455 West Lindsey, Room 705, University of Oklahoma, Norman OK 73019, U.S.A.