

# Chapter 40

## The MDS Procedure

### Chapter Table of Contents

---

|  |      |
|--|------|
| <b>OVERVIEW</b> . . . . .  | 2045 |
| <b>GETTING STARTED</b> . . . . .   | 2047 |
| <b>SYNTAX</b> . . . . .  | 2049 |
| PROC MDS Statement . . . . .   | 2050 |
| BY Statement . . . . .   | 2059 |
| ID Statement . . . . .   | 2059 |
| INVAR Statement . . . . .  | 2060 |
| MATRIX Statement . . . . .   | 2060 |
| VAR Statement . . . . .  | 2060 |
| WEIGHT Statement . . . . .   | 2061 |
| <b>DETAILS</b> . . . . .   | 2061 |
| Formulas . . . . .   | 2061 |
| OUT= Data Set . . . . .  | 2063 |
| OUTFIT= Data Set . . . . .   | 2064 |
| OUTRES= Data Set . . . . .   | 2064 |
| INITIAL= Data Set . . . . .  | 2065 |
| Missing Values . . . . .   | 2066 |
| Normalization of the Estimates . . . . .                                   | 2066 |
| Comparison with the ALSCAL Procedure . . . . .                             | 2067 |
| Comparison with the MLSCALE Procedure . . . . .                            | 2069 |
| Displayed Output . . . . .   | 2070 |
| ODS Table Names . . . . .  | 2071 |
| <b>EXAMPLE</b> . . . . .   | 2072 |
| Example 40.1 Jacobowitz Body Parts Data from Children and Adults . . . . . | 2072 |
| <b>REFERENCES</b> . . . . .  | 2081 |



# Chapter 40

## The MDS Procedure

---

### Overview

Multidimensional scaling (MDS) is a class of methods that estimates the coordinates of a set of objects in a space of specified dimensionality that come from data measuring the distances between pairs of objects. A variety of models can be used that include different ways of computing distances and various functions relating the distances to the actual data. The MDS procedure fits two- and three-way, metric and nonmetric multidimensional scaling models. PROC MDS shares many of the features of the ALSCAL procedure (Young, Lewyckyj, and Takane 1986; Young 1982), as well as some features of the MLSCALE procedure (Ramsay 1986). Both PROC ALSCAL and PROC MLSCALE are described in the *SUGI Supplemental Library User's Guide, Version 5 Edition*.

The data for the MDS procedure consist of one or more square symmetric or asymmetric matrices of similarities or dissimilarities between *objects* or *stimuli* (Kruskal and Wish 1978, pp. 7–11). Such data are also called *proximity* data. In psychometric applications, each matrix typically corresponds to a *subject*, and models that fit different parameters for each subject are called *individual difference* models.

Missing values are allowed. In particular, if the data are all missing except within some off-diagonal rectangle, the analysis is called *unfolding*. There are, however, many difficulties intrinsic to unfolding models (Heiser, 1981). PROC MDS does not perform external unfolding; for analyses requiring external unfolding, use the TRANSREG procedure instead.

The MDS procedure estimates the following parameters by nonlinear least squares:

|                        |  |
|------------------------|--|
| configuration          | the coordinates of each object in a Euclidean (Kruskal and Wish 1978, pp. 17–19) or weighted Euclidean space (Kruskal and Wish 1978, pp. 61–63) of one or more dimensions  |
| dimension coefficients | for each data matrix, the coefficients that multiply each coordinate of the <i>common</i> or <i>group</i> weighted Euclidean space to yield the <i>individual</i> unweighted Euclidean space. These coefficients are the square roots of the <i>subject weights</i> (Kruskal and Wish 1978, pp. 61–63). A plot of the dimension coefficients is directly interpretable in that it shows how a unit square in the group space is transformed to a rectangle in each individual space. A plot of subject weights has no such simple interpretation. The weighted Euclidean model is related to the INDSCAL model (Carroll and Chang 1970). |

transformation parameters intercept, slope, or exponent in a linear, affine, or power transformation relating the distances to the data (Kruskal and Wish 1978, pp. 19–22). For a nonmetric analysis, monotone transformations involving no explicit parameters are used (Kruskal and Wish 1978, pp. 22–25). For a discussion of metric versus nonmetric transformations, refer to Kruskal and Wish (1978, pp. 76–78).

Depending on the LEVEL= option, PROC MDS fits either a regression model of the form

$$fit(datum) = fit(trans(distance)) + error$$

or a measurement model of the form

$$fit(trans(datum)) = fit(distance) + error$$

where

|                 |  |
|-----------------|--|
| <i>fit</i>      | is a predetermined power or logarithmic transformation specified by the FIT= option.   |
| <i>trans</i>    | is an estimated (“optimal”) linear, affine, power, or monotone transformation specified by the LEVEL= option.  |
| <i>datum</i>    | is a measure of the similarity or dissimilarity of two objects or stimuli.   |
| <i>distance</i> | is a distance computed from the estimated coordinates of the two objects and estimated dimension coefficients in a space of one or more dimensions. If there are no dimension coefficients (COEF=IDENTITY), this is an unweighted Euclidean distance. If dimension coefficients are used (COEF=DIAGONAL), this is a weighted Euclidean distance where the weights are the squares of the dimension coefficients; alternatively, you can multiply each dimension by its coefficient and compute an unweighted Euclidean distance. |
| <i>error</i>    | is an error term assumed to have an approximately normal distribution and to be independently and identically distributed for all data. Under these assumptions, least-squares estimation is statistically appropriate.  |

For an introduction to multidimensional scaling, refer to Kruskal and Wish (1978) and Arabie, Carroll, and DeSarbo (1987). A more advanced treatment is given by Young (1987). Many practical issues of data collection and analysis are discussed in Schiffman, Reynolds, and Young (1981). The fundamentals of psychological measurement, including both unidimensional and multidimensional scaling, are expounded by Torgerson (1958). Nonlinear least-squares estimation of PROC MDS models is discussed in Null and Sarle (1982).

---

## Getting Started

The simplest application of PROC MDS is to reconstruct a map from a table of distances between points on the map (Kruskal and Wish 1978, pp. 7–9). For example, the following DATA step reads a table of flying mileages between ten U.S. cities:

```

data city;
  title 'Analysis of Flying Mileages Between Ten U.S. Cities';
  input (atlanta chicago denver houston losangeles
        miami newyork sanfran seattle washdc) (5.)
        @56 city $15.;
  datalines;
    0 Atlanta
  587 0 Chicago
 1212 920 0 Denver
  701 940 879 0 Houston
 1936 1745 831 1374 0 Los Angeles
  604 1188 1726 968 2339 0 Miami
  748 713 1631 1420 2451 1092 0 New York
 2139 1858 949 1645 347 2594 2571 0 San Francisco
 2182 1737 1021 1891 959 2734 2408 678 0 Seattle
  543 597 1494 1220 2300 923 205 2442 2329 0 Washington D.C.
;

```

Since the flying mileages are very good approximations to Euclidean distance, no transformation is needed to convert distances from the model to data. The analysis can therefore be done at the absolute level of measurement, as displayed in the following PROC MDS step (LEVEL=ABSOLUTE). An output data set containing the estimated configuration (coordinates on the map) is created with the OUT= option and then used to display the map using the %PLOTIT macro. The ID statement copies the names of the cities to the OUT= data set so that they can be used on the plot. The following statements produce Figure 40.1:

```

proc mds data=city level=absolute out=out;
  id city;
run;

```

By default, PROC MDS displays only the iteration history. In this example, only one iteration is required. The badness-of-fit criterion 0.001689 indicates that the data fit the model extremely well.

```

Analysis of Flying Mileages Between Ten U.S. Cities

Multidimensional Scaling: Data=WORK.CITY
Shape=TRIANGLE Condition=MATRIX Level=ABSOLUTE
Coef=IDENTITY Dimension=2 Formula=1 Fit=1

Gconverge=0.01 Maxiter=100 Over=1 Ridge=0.0001

Iteration      Type      Badness-
              of-Fit
              Criterion
-----
              Change in
              Criterion
              Convergence
              Measure
-----
0      Initial      0.003273      .      0.8562
1      Lev-Mar      0.001689      0.001584      0.005128

Convergence criterion is satisfied.

```

**Figure 40.1.** Iteration History from PROC MDS

In order for the plot to represent the geometry of the configuration correctly, the axes must be scaled so that a given physical distance on one axis represents the same number of units as the same physical distance on the other axis. You can use the %PLOTIT macro to obtain such a scaling by specifying the VTOH= option. The VTOH= option indicates the ratio of the vertical distance between lines to the horizontal distance between characters; obtaining a suitable value may require some experimentation. The %PLOTIT macro can create graphical scatter plots with iteratively derived optimal label placement. See Appendix B, “Using the %PLOTIT Macro,” for more information on the %PLOTIT macro.

The following statements produce Figure 40.2:

```

%plotit(data=out, datatype=mds, labelvar=city,
         vtoh=1.75, labfont=swissb);
run;

```

While PROC MDS can recover the relative positions of the cities, it cannot determine absolute location or orientation. In this case, north is toward the bottom of the plot.

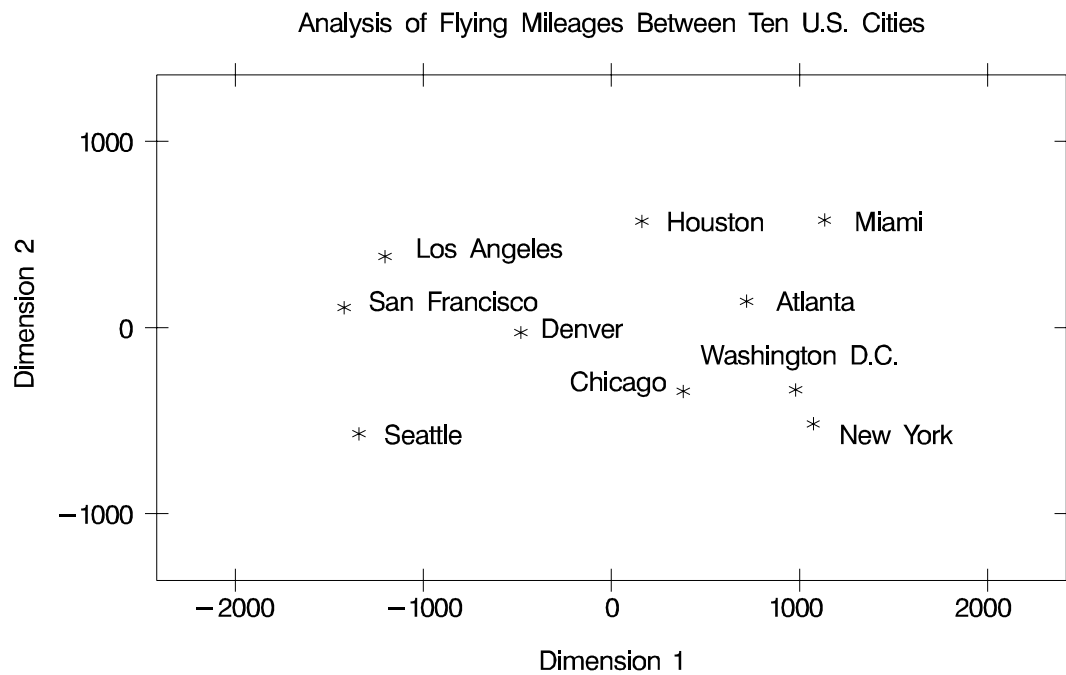


Figure 40.2. Plot of Estimated Configuration Using the %PLOTIT Macro

---

## Syntax

You can specify the following statements with the MDS procedure:

```

PROC MDS < options > ;
  VAR variables ;
  INVAR variables ;
  ID | OBJECT variable ;
  MATRIX | SUBJECT variable ;
  WEIGHT variables ;
  BY variables ;

```

The PROC MDS statement is required. All other statements are optional.

---

## PROC MDS Statement

**PROC MDS** < options > ;

By default, the only result produced by the MDS procedure is the iteration history. Hence, you should always specify one or more options for output data sets (OUT=, OUTFIT=, and OUTRES=) or displayed output (such as PFINAL). PROC MDS does not produce any plots; to produce plots, use the output data sets with PROC PLOT or PROC GPLOT.

The types of estimates written to the OUT= data set are determined by the OCONFIG, OCOEF, OTRANS, and OCRIT options. If you do not specify any of these four options, the estimates of all the parameters of the PROC MDS model and the value of the badness-of-fit criterion appear in the OUT= data set. If you specify one or more of these options, only the information requested by the specified options appear in the OUT= data set. Also, the OITER option causes these statistics to be written to the OUT= data set after initialization and on each iteration, as well as after the iterations have terminated.

Displayed output is controlled by the interaction of the PCONFIG, PCOEF, PTRANS, PFIT, and PFITROW options with the PININ, PINIT, PITER, and PFINAL options. The PCONFIG, PCOEF, PTRANS, PFIT, and PFITROW options specify *which* estimates and fit statistics are to be displayed. The PININ, PINIT, PITER, and PFINAL options specify *when* the estimates and fit statistics are to be displayed. If you specify at least one of the PCONFIG, PCOEF, PTRANS, PFIT and PFITROW options but none of the PININ, PINIT, PITER, and PFINAL options, the final results (PFINAL) are displayed. If you specify at least one of the PININ, PINIT, PITER, and PFINAL options but none of the PCONFIG, PCOEF, PTRANS, PFIT and PFITROW options, all estimates (PCONFIG, PCOEF, PTRANS) and the fit statistics for each matrix and for the entire sample (PFIT) are displayed. If you do not specify any of these nine options, no estimates or fit statistics are displayed (except the badness-of-fit criterion in the iteration history).

**ALTERNATE | ALT=NONE | NO | N**

**ALTERNATE | ALT=MATRIX | MAT | M | SUBJECT | SUB | S**

**ALTERNATE | ALT=ROW | R <=n>**

determines what form of alternating-least-squares algorithm is used. The default depends on the amount of memory available. The following ALTERNATE= options are listed in order of decreasing memory requirements:

**ALT=NONE** causes all parameters to be adjusted simultaneously on each iteration. This option is usually best for a small number of subjects and objects.

**ALT=MATRIX** adjusts all the parameters for the first subject, then all the parameters for the second subject, and so on, and finally adjusts all parameters that do not correspond to a subject, such as coordinates and unconditional transformations. This op-



tion usually works best for a large number of subjects with a small number of objects.

**ALT=ROW**

treats subject parameters the same way as the ALTERNATE=MATRIX option but also includes separate stages for unconditional parameters and for subsets of the objects. The ALT=ROW option usually works best for a large number of objects.

Specifying ALT=ROW= $n$  divides the objects into subsets of  $n$  objects each, except possibly for one subset when  $n$  does not divide the number of objects evenly. If you omit  $=n$ , the number of objects in the subsets is determined from the amount of memory available. The smaller the value of  $n$ , the less memory is required.

When you specify the LEVEL=ORDINAL option, the monotone transformation is always computed in a separate stage and is listed as a separate iteration in the iteration history. In this case, estimation is done by iteratively reweighted least squares. The weights are recomputed according to the FORMULA= option on each monotone iteration; hence, it is possible for the badness-of-fit criterion to increase after a monotone iteration.

**COEF=IDENTITY | IDEN | I****COEF=DIAGONAL | DIAG | D**

specifies the type of matrix for the dimension coefficients.

COEF=IDENTITY is the default, which yields Euclidean distances.

COEF=DIAGONAL produces weighted Euclidean distances, in which each subject is allowed differential weights for the dimensions. The dimension coefficients that PROC MDS outputs are related to the square roots of what are called subject weights in PROC ALSCAL; the normalization in PROC MDS also differs from that in PROC ALSCAL. The weighted Euclidean model is related to the INDSCAL model (Carroll and Chang 1970).

**CONDITION | COND=UN | U****CONDITION | COND=MATRIX | MAT | M | SUBJECT | SUB | S****CONDITION | COND=ROW | R**

specifies the conditionality of the data (Young 1987, pp. 60–63). The default is CONDITION=MATRIX.

The data are divided into disjoint subsets called *partitions*. Within each partition, a separate transformation is applied, as specified by the LEVEL= option.

COND=UN puts all the data into a single partition.

COND=MATRIX makes each data matrix a partition.

COND=ROW makes each row of each data matrix a partition.

The `CONDITION=` option also determines the default value for the `SHAPE=` option. If you specify the `CONDITION=ROW` option and omit the `SHAPE=` option, each data matrix is stored as a square and possibly asymmetric matrix. If you specify the `CONDITION=UN` or `CONDITION=MATRIX` option and omit the `SHAPE=` option, only one triangle is stored. See the `SHAPE=` option on page 2058 for details.

#### **CONVERGE | CONV=*p***

sets both the gradient convergence criterion and the monotone convergence criterion to  $p$ , where  $0 \leq p \leq 1$ . The default is `CONVERGE=.01`; smaller values may greatly increase the number of iterations required. Values less than .0001 may be impossible to satisfy because of the limits of machine precision. See the `GCONVERGE=` (page 2054) and `MCONVERGE=` (page 2055) options.

#### **CUTOFF=*n***

causes data less than  $n$  to be replaced by missing values. The default is `CUTOFF=0`.

#### **DATA=*SAS-data-set***

specifies the SAS data set containing one or more square matrices to be analyzed. In typical psychometric data, each matrix contains judgments from one subject, so there is a one-to-one correspondence between data matrices and subjects.

The data matrices contain similarity or dissimilarity measurements to be modeled and, optionally, weights for these data. The data are generally assumed to be dissimilarities unless you use the `SIMILAR` option. However, if there are nonmissing diagonal values and these values are predominantly larger than the off-diagonal values, the data are assumed to be similarities and are treated as if the `SIMILAR` option is specified. The diagonal elements are not otherwise used in fitting the model.

Each matrix must have exactly the same number of observations as the number of variables specified by the `VAR` statement or determined by defaults. This number is the number of objects or stimuli.

The first observation and variable are assumed to contain data for the first object, the second observation and variable are assumed to contain data for the second object, and so on.

When there are two or more matrices, the observations in each matrix must correspond to the same objects in the same order as in the first matrix.

The matrices can be symmetric or asymmetric, as specified by the `SHAPE=` option.

#### **DECIMALS | DEC=*n***

specifies how many decimal places to use when displaying the parameter estimates and fit statistics. The default is `DECIMALS=2`, which is generally reasonable except in conjunction with the `LEVEL=ABSOLUTE` option and very large or very small data.

#### **DIMENSION | DIMENS | DIM=*n* < TO *m* < BY=*i* >>**

specifies the number of dimensions to use in the MDS model, where  $1 \leq n, m < \text{number of objects}$ . The parameter  $i$  can be either positive or negative but not zero. If you specify different values for  $n$  and  $m$ , a separate model is fitted for each requested dimension. If you specify only `DIMENSION= $n$` , then only  $n$  dimensions are fitted.

The default is DIMENSION=2 if there are three or more objects; otherwise, DIMENSION=1 is the only valid specification. The analyses for each number of dimensions are done independently. For information on choosing the dimensionality, refer to Kruskal and Wish (1978, pp. 48–60).

**EPSILON | EPS=*n***

specifies a number  $n$ ,  $0 < n < 1$ , that determines the amount added to squared distances computed from the model to avoid numerical problems such as division by 0. This amount is computed as  $\epsilon$  equal to  $n$  times the mean squared distance in the initial configuration. The distance in the MDS model is thus computed as

$$distance = \sqrt{sqdist + \epsilon}$$

where  $sqdist$  is the squared Euclidean distance or the weighted squared Euclidean distance.

The default is EPSILON=1E–12, which is small enough to have no practical effect on the estimates unless the FIT= value is nonpositive and there are dissimilarities that are very close to 0. Hence, when the FIT= value is nonpositive, dissimilarities less than  $n$  times 100 times the maximum dissimilarity are disallowed.

**FIT=DISTANCE | DIS | D**

**FIT=SQUARED | SQU | S**

**FIT=LOG | L**

**FIT=*n***

specifies a predetermined (not estimated) transformation to apply to both sides of the MDS model before the error term is added.

The default is FIT=DISTANCE or, equivalently, FIT=1, which fits data to distances.

The option FIT=SQUARED or FIT=2 fits squared data to squared distances. This gives greater importance to large data and distances and lesser importance to small data and distances in fitting the model.

The FIT=LOG or FIT=0 option fits log data to log distances. This gives lesser importance to large data and distances and greater importance to small data and distances in fitting the model.

In general, the FIT= $n$  option fits  $n$ th-power data to  $n$ th-power distances. Values of  $n$  that are large in absolute value can cause floating-point overflows.

If the FIT= value is 0 or negative, the data must be strictly positive (see the EPSILON= option). Negative data may produce strange results with any value other than FIT=1.

**FORMULA | FOR=0 | OLS | O**

**FORMULA | FOR=1 | USS | U**

**FORMULA | FOR=2 | CSS | C**

determines how the badness-of-fit criterion is standardized in correspondence with stress formulas 1 and 2 (Kruskal and Wish 1978, pp. 24–26). The default is FORMULA=1 unless you specify FIT=LOG, in which case the default is FORMULA=2. Data partitions are defined by the CONDITION= option.

- FORMULA=0** fits a regression model by ordinary least squares (Null and Sarle 1982) without standardizing the partitions; this option cannot be used with the **LEVEL=ORDINAL** option. The badness-of-fit criterion is the square root of the error sum of squares.
- FORMULA=1** standardizes each partition by the uncorrected sum of squares of the (possibly transformed) data; this option should not be used with the **FIT=LOG** option. With the **FIT=DISTANCE** and **LEVEL=ORDINAL** options, this is equivalent to Kruskal's stress formula 1 or an obvious generalization thereof. With the **FIT=SQUARED** and **LEVEL=ORDINAL** options, this is equivalent to Young's s-stress formula 1 or an obvious generalization thereof. The badness-of-fit criterion is analogous to  $\sqrt{1 - R^2}$ , where  $R$  is a multiple correlation about the origin.
- FORMULA=2** standardizes each partition by the corrected sum of squares of the (possibly transformed) data; this option is the recommended method for unfolding. With the **FIT=DISTANCE** and **LEVEL=ORDINAL** options, this is equivalent to Kruskal's stress formula 2 or an obvious generalization thereof. With the **FIT=SQUARED** and **LEVEL=ORDINAL** options, this is equivalent to Young's s-stress formula 2 or an obvious generalization thereof. The badness-of-fit criterion is analogous to  $\sqrt{1 - R^2}$ , where  $R$  is a multiple correlation computed with a denominator corrected for the mean.

**GCONVERGE | GCONV= $p$** 

sets the gradient convergence criterion to  $p$ , where  $0 \leq p \leq 1$ . The default is **GCONVERGE=0.01**; smaller values may greatly increase the number of iterations required. Values less than 0.0001 may be impossible to satisfy because of the limits of machine precision.

The gradient convergence measure is the multiple correlation of the Jacobian matrix with the residual vector, uncorrected for the mean. See the **CONVERGE=** (page 2052) and **MCONVERGE=** (page 2055) options.

**INAV=DATA | D****INAV=SSCP | S**

affects the computation of initial coordinates. The default is **INAV=DATA**.

**INAV=DATA** computes a weighted average of the data matrices. Its value is estimated only if an element is missing from every data matrix. The weighted average of the data matrices with missing values filled in is then converted to a scalar products matrix (or what would be a scalar products matrix if the fit were perfect), from which the initial coordinates are computed.

**INAV=SSCP** estimates missing values in each data matrix and converts each data matrix to a scalar products matrix. The initial coordinates are computed from the unweighted average of the scalar products matrices.

**INITIAL | IN=SAS-data-set**

specifies a SAS data set containing initial values for some or all of the parameters of the MDS model. If the INITIAL= option is omitted, the initial values are computed from the data.

**LEVEL=ABSOLUTE | ABS | A****LEVEL=RATIO | RAT | R****LEVEL=INTERVAL | INT | I****LEVEL=LOGINTERVAL | LOG | L****LEVEL=ORDINAL | ORD | O**

specifies the measurement level of the data and hence the type of estimated (optimal) transformations applied to the data or distances (Young 1987, pp. 57–60; Krantz et. al. 1971, pp. 9–12) within each partition as specified by the CONDITION= option. The default is LEVEL=ORDINAL.

|                   |  |
|-------------------|--|
| LEVEL=ABSOLUTE    | allows no optimal transformations. Hence, the distinction between regression and measurement models is irrelevant.   |
| LEVEL=RATIO       | fits a regression model in which the distances are multiplied by a slope parameter in each partition (a linear transformation). In this case, the regression model is equivalent to the measurement model with the slope parameter reciprocated.                     |
| LEVEL=INTERVAL    | fits a regression model in which the distances are multiplied by a slope parameter and added to an intercept parameter in each partition (an affine transformation). In this case, the regression and measurement models differ if there is more than one partition. |
| LEVEL=LOGINTERVAL | fits a regression model in which the distances are raised to a power and multiplied by a slope parameter in each partition (a power transformation).   |
| LEVEL=ORDINAL     | fits a measurement model in which a least-squares monotone increasing transformation is applied to the data in each partition. At the ordinal measurement level, the regression and measurement models differ.   |

**MAXITER | ITER=*n***

specifies the maximum number of iterations, where  $n \geq 0$ . The default is MAXITER=100.

**MCONVERGE | MCONV=*p***

sets the monotone convergence criterion to  $p$ , where  $0 \leq p \leq 1$ , for use with the LEVEL=ORDINAL option. The default is MCONVERGE=0.01; if you want greater precision, MCONVERGE=0.001 is usually reasonable, but smaller values may greatly increase the number of iterations required.

The monotone convergence criterion is the Euclidean norm of the change in the optimally scaled data divided by the Euclidean norm of the optimally scaled data, aver-

aged across partitions defined by the `CONDITION=` option. See the `CONVERGE=` (page 2052) and `GCONVERGE=` (page 2054) options.

**MINCRIT | CRITMIN=*n***

causes iteration to terminate when the badness-of-fit criterion is less than or equal to *n*, where  $n \geq 0$ . The default is `MINCRIT=1E-6`.

**NEGATIVE**

allows slopes or powers to be negative with the `LEVEL=RATIO`, `INTERVAL`, or `LOGINTERVAL` option.

**NONORM**

suppresses normalization of the initial and final estimates.

**NOPHIST | NOPRINT | NOP**

suppresses the output of the iteration history.

**NOULB**

causes missing data to be estimated during initialization by the average nonmissing value, where the average is computed according to the `FIT=` option. Otherwise, missing data are estimated by interpolating between the Rabinowitz (1976) upper and lower bounds.

**OCOEF**

writes the dimension coefficients to the `OUT=` data set. See the `OUT=` option for interactions with other options.

**OCONFIG**

writes the coordinates of the objects to the `OUT=` data set. See the `OUT=` option for interactions with other options.

**OCRIT**

writes the badness-of-fit criterion to the `OUT=` data set. See the `OUT=` option for interactions with other options.

**OITER | OUTITER**

writes current values to the output data sets after initialization and on every iteration. Otherwise, only the final values are written to any output data sets. See the `OUT=`, `OUTFIT=`, and `OUTRES=` options.

**OTRANS**

writes the transformation parameter estimates to the `OUT=` data set if any such estimates are computed. There are no transformation parameters with the `LEVEL=ORDINAL` option. See the `OUT=` option for interactions with other options.

**OUT=SAS-data-set**

creates a SAS data set containing, by default, the estimates of all the parameters of the PROC MDS model and the value of the badness-of-fit criterion. However, if you specify one or more of the `OCONFIG`, `OCOEF`, `OTRANS`, and `OCRIT` options, only the information requested by the specified options appears in the `OUT=` data set. See also the `OITER` option.

**OUTFIT=SAS-data-set**

creates a SAS data set containing goodness-of-fit and badness-of-fit measures for each partition as well as for the entire data set. See also the OITER option.

**OUTRES=SAS-data-set**

creates a SAS data set containing one observation for each nonmissing datum from the DATA= data set. Each observation contains the original datum, the estimated distance computed from the MDS model, transformed data and distances, and the residual. See also the OITER option.

**OVER=*n***

specifies the maximum overrelaxation factor, where  $n \geq 1$ . Values between 1 and 2 are generally reasonable. The default is OVER=2 with the LEVEL=ORDINAL, ALTERNATE=MATRIX, or ALTERNATE=ROW option; otherwise, the default is OVER=1. Use this option only if you have convergence problems.

**PCOEF**

produces the estimated dimension coefficients.

**PCONFIG**

produces the estimated coordinates of the objects in the configuration.

**PDATA**

displays each data matrix.

**PFINAL**

displays final estimates.

**PFIT**

displays the badness-of-fit criterion and various types of correlations between the data and fitted values for each data matrix, as well as for the entire sample.

**PFITROW**

displays the badness-of-fit criterion and various types of correlations between the data and fitted values for each row as well as for each data matrix and for the entire sample. This option works only with the CONDITION=ROW option.

**PINAVDATA**

displays the sum of the weights and the weighted average of the data matrices computed during initialization with the INAV=DATA option.

**PINEIGVAL**

displays the eigenvalues computed during initialization.

**PINEIGVEC**

displays the eigenvectors computed during initialization.

**PININ**

displays values read from the INITIAL= data set. Since these values may be incomplete, the PFIT and PFITROW options do not apply.

**PINIT**

displays initial values.



**PITER**

displays estimates on each iteration.

**PTRANS**

displays the estimated transformation parameters if any are computed. There are no transformation parameters with the LEVEL=ORDINAL option.

**RANDOM**<=*seed*>

causes initial coordinate values to be pseudorandom numbers. In one dimension, the pseudorandom numbers are uniformly distributed on an interval. In two or more dimensions, the pseudorandom numbers are uniformly distributed on the circumference of a circle or the surface of a (hyper)sphere.

**RIDGE**=*n*

specifies the initial ridge value, where  $n \geq 0$ . The default is RIDGE=1E-4.

If you get a floating-point overflow in the first few iterations, specify a larger value such as RIDGE=0.01 or RIDGE=1 or RIDGE=100.

If you know that the initial estimates are very good, using RIDGE=0 may speed convergence.

**SHAPE=TRIANGULAR | TRIANGLE | TRI | T****SHAPE=SQUARE | SQU | S**

determines whether the entire data matrix for each subject or only one triangle of the matrix is stored and analyzed. If you specify the CONDITION=ROW option, the default is SHAPE=SQUARE. Otherwise, the default is SHAPE=TRIANGLE.

**SHAPE=SQUARE** causes the entire matrix to be stored and analyzed. The matrix can be asymmetric.

**SHAPE=TRIANGLE** causes only one triangle to be stored. However, PROC MDS reads both upper and lower triangles to look for nonmissing values and to symmetrize the data if needed. If corresponding elements in the upper and lower triangles both contain nonmissing values, only the average of the two values is stored and analyzed (Kruskal and Wish 1978, p. 74). Also, if an OUTRES= data set is requested, only the average of the two corresponding elements is output.

**SIMILAR | SIM**<=*max*>

causes the data to be treated as similarity measurements rather than dissimilarities. If =*max* is not specified, each datum is converted to a dissimilarity by subtracting it from the maximum value in the data set or BY group. If =*max* is specified, each datum is subtracted from the maximum of *max* and the data. The diagonal data are included in computing these maxima.

By default, the data are assumed to be dissimilarities unless there are nonmissing diagonal values and these values are predominantly larger than the off-diagonal values. In this case, the data are assumed to be similarities and are treated as if the SIMILAR option is specified.



**SINGULAR= $p$** 

specifies the singularity criterion  $p$ ,  $0 \leq p \leq 1$ . The default is SINGULAR=1E-8.

**UNTIE**

allows tied data to be assigned different optimally scaled values with the LEVEL=ORDINAL option. Otherwise, tied data are assigned equal optimally scaled values. The UNTIE option has no effect with values of the LEVEL= option other than LEVEL=ORDINAL.

---

## BY Statement

**BY variables ;**

You can specify a BY statement with PROC MDS to obtain separate analyses on observations in groups defined by the BY variables. When a BY statement appears, the procedure expects the input data set to be sorted in order of the BY variables.

If your input data set is not sorted in ascending order, use one of the following alternatives:

- Sort the data using the SORT procedure with a similar BY statement.
- Specify the BY statement option NOTSORTED or DESCENDING in the BY statement for the MDS procedure. The NOTSORTED option does not mean that the data are unsorted but rather that the data are arranged in groups (according to values of the BY variables) and that these groups are not necessarily in alphabetical or increasing numeric order.
- Create an index on the BY variables using the DATASETS procedure.

If the INITIAL= data set contains the BY variables, the BY groups must appear in the same order as in the DATA= data set. If the BY variables are not in the INITIAL= data set, the entire data set is used to provide initial values for each BY group in the DATA= data set.

For more information on the BY statement, refer to the discussion in *SAS Language Reference: Concepts*. For more information on the DATASETS procedure, refer to the discussion in the *SAS Procedures Guide*.

---

## ID Statement

**ID | OBJECT | OBJ variable ;**

The ID statement specifies a variable in the DATA= data set that contains descriptive labels for the objects. The labels are used in the output and are copied to the OUT= data set. If there is more than one data matrix, only the ID values from the observations containing the first data matrix are used.

The ID variable is not used to establish any correspondence between observations and variables.

If the ID statement is omitted, the variable labels or names are used as object labels.

---

## INVAR Statement

**INVAR** *variables* ;

The INVAR statement specifies the numeric variables in the INITIAL= data set that contain initial parameter estimates. The first variable corresponds to the first dimension, the second variable to the second dimension, and so on.

If the INVAR statement is omitted, the variables DIM1, . . . , DIM $m$  are used, where  $m$  is the maximum number of dimensions.

---

## MATRIX Statement

**MATRIX** | **MAT** | **SUBJECT** | **SUB** *variable* ;

The MATRIX statement specifies a variable in the DATA= data set that contains descriptive labels for the data matrices or subjects. The labels are used in the output and are copied to the OUT= and OUTRES= data sets. Only the first observation from each data matrix is used to obtain the label for that matrix.

If the MATRIX statement is omitted, the matrices are labeled 1, 2, 3, and so on.

---

## VAR Statement

**VAR** *variables* ;

The VAR statement specifies the numeric variables in the DATA= data set that contain similarity or dissimilarity measurements on a set of objects or stimuli. Each variable corresponds to one object.

If the VAR statement is omitted, all numeric variables that are not specified in another statement are used.

To analyze a subset of the objects in a data set, you can specify the variable names corresponding to the columns in the subset, but you must also use a DATA step or a WHERE clause to specify the rows in the subset. PROC MDS expects to read one or more square matrices, and you must ensure that the rows in the data set correctly correspond to the columns in number and order.

---

## WEIGHT Statement

### WEIGHT variables ;

The WEIGHT statement specifies numeric variables in the DATA= data set that contain weights for each similarity or dissimilarity measurement. These weights are used to compute weighted least-squares estimates. The number of WEIGHT variables must be the same as the number of VAR variables, and the variables in the WEIGHT statement must be in the same order as the corresponding variables in the VAR statement.

If the WEIGHT statement is omitted, all data within a partition are assigned equal weights.

Data with 0 or negative weights are ignored in fitting the model but are included in the OUTRES= data set and in monotone transformations.

---

## Details

---

### Formulas

The following notation is used:

|              |  |
|--------------|--|
| $A_p$        | intercept for partition $p$  |
| $B_p$        | slope for partition $p$  |
| $C_p$        | power for partition $p$  |
| $D_{rcs}$    | distance computed from the model between objects $r$ and $c$ for subject $s$   |
| $F_{rcs}$    | data weight for objects $r$ and $c$ for subject $s$ obtained from the $c$ th WEIGHT variable, or 1 if there is no WEIGHT statement |
| $f$          | value of the FIT= option   |
| $N$          | number of objects  |
| $O_{rcs}$    | observed dissimilarity between objects $r$ and $c$ for subject $s$   |
| $P_{rcs}$    | partition index for objects $r$ and $c$ for subject $s$  |
| $Q_{rcs}$    | dissimilarity after applying any applicable estimated transformation for objects $r$ and $c$ for subject $s$                       |
| $R_{rcs}$    | residual for objects $r$ and $c$ for subject $s$   |
| $S_p$        | standardization factor for partition $p$   |
| $T_p(\cdot)$ | estimated transformation for partition $p$   |
| $V_{sd}$     | coefficient for subject $s$ on dimension $d$   |
| $X_{nd}$     | coordinate for object $n$ on dimension $d$   |

Summations are taken over nonmissing values.

Distances are computed from the model as

$$\begin{aligned}
 D_{rcs} &= \sqrt{\sum_d (X_{rd} - X_{cd})^2} && \text{for COEF=IDENTITY:} \\
 &&& \text{Euclidean distance} \\
 &= \sqrt{\sum_d V_{sd}^2 (X_{rd} - X_{cd})^2} && \text{for COEF=DIAGONAL:} \\
 &&& \text{weighted Euclidean distance}
 \end{aligned}$$

Partition indexes are

$$\begin{aligned}
 P_{rcs} &= 1 && \text{for CONDITION=UN} \\
 &= s && \text{for CONDITION=MATRIX} \\
 &= (s - 1)N + r && \text{for CONDITION=ROW}
 \end{aligned}$$

The estimated transformation for each partition is

$$\begin{aligned}
 T_p(d) &= d && \text{for LEVEL=ABSOLUTE} \\
 &= B_p d && \text{for LEVEL=RATIO} \\
 &= A_p + B_p d && \text{for LEVEL=INTERVAL} \\
 &= B_p d^{C_p} && \text{for LEVEL=LOGINTERVAL}
 \end{aligned}$$

For LEVEL=ORDINAL,  $T_p(\cdot)$  is computed as a least-squares monotone transformation.

For LEVEL=ABSOLUTE, RATIO, or INTERVAL, the residuals are computed as

$$\begin{aligned}
 Q_{rcs} &= O_{rcs} \\
 R_{rcs} &= Q_{rcs}^f - [T_{P_{rcs}}(D_{rcs})]^f
 \end{aligned}$$

For LEVEL=ORDINAL, the residuals are computed as

$$\begin{aligned}
 Q_{rcs} &= T_{P_{rcs}}(O_{rcs}) \\
 R_{rcs} &= Q_{rcs}^f - D_{rcs}^f
 \end{aligned}$$

If  $f$  is 0, then natural logarithms are used in place of the  $f$ th powers.

For each partition, let

$$U_p = \frac{\sum_{r,c,s} F_{rcs}}{\sum_{r,c,s|P_{rcs}=p} F_{rcs}}$$

and

$$\bar{Q}_p = \frac{\sum_{r,c,s|P_{rcs}=p} Q_{rcs} F_{rcs}}{\sum_{r,c,s|P_{rcs}=p} F_{rcs}}$$

Then the standardization factor for each partition is

$$\begin{aligned}
 S_p &= 1 && \text{for FORMULA=0} \\
 &= U_p \sum_{r,c,s | P_{rcs}=p} Q_{rcs}^2 F_{rcs} && \text{for FORMULA=1} \\
 &= U_p \sum_{r,c,s | P_{rcs}=p} (Q_{rcs} - \bar{Q}_p)^2 F_{rcs} && \text{for FORMULA=2}
 \end{aligned}$$

The badness-of-fit criterion that the MDS procedure tries to minimize is

$$\sqrt{\sum_{r,c,s} \frac{R_{rcs}^2 F_{rcs}}{S_{P_{rcs}}}}$$

---

## OUT= Data Set

The OUT= data set contains the following variables:

- BY variables, if any
- `_ITER_` (if the OUTITER option is specified), a numeric variable containing the iteration number
- `_DIMENS_`, a numeric variable containing the number of dimensions
- `_MATRIX_` or the variable in the MATRIX statement, identifying the data matrix or subject to which the observation pertains. This variable contains a missing value for observations that pertain to the data set as a whole and not to a particular matrix, such as the coordinates (`_TYPE_='CONFIG'`).
- `_TYPE_`, a character variable of length 10 identifying the type of information in the observation

The values of `_TYPE_` are as follows:

|           |   |
|-----------|---|
| CONFIG    | the estimated coordinates of the configuration of objects |
| DIAGCOEF  | the estimated dimension coefficients for COEF=DIAGONAL    |
| INTERCEPT | the estimated intercept parameters                        |
| SLOPE     | the estimated slope parameters                            |
| POWER     | the estimated power parameters                            |
| CRITERION | the badness-of-fit criterion                              |

- `_LABEL_` or the variable in the ID statement, containing the variable label or value of the ID variable of the object to which the observation pertains. This variable contains a missing value for observations that do not pertain to a particular object or dimension.
- `_NAME_`, a character variable of length 8 containing the variable name of the object or dimension to which the observation pertains. This variable contains a missing value for observations that do not pertain to a particular object or dimension.
- DIM1, . . . , DIM $m$ , where  $m$  is the maximum number of dimensions

---

## OUTFIT= Data Set

The OUTFIT= data set contains various measures of goodness and badness of fit. There is one observation for the entire sample plus one observation for each matrix. For the CONDITION=ROW option, there is also one observation for each row.

The OUTFIT= data set contains the following variables:

- BY variables, if any
- `_ITER_` (if the OUTITER option is specified), a numeric variable containing the iteration number
- `_DIMENS_`, a numeric variable containing the number of dimensions
- `_MATRIX_` or the variable in the MATRIX statement, identifying the data matrix or subject to which the observation pertains
- `_LABEL_` or the variable in the ID statement, containing the variable label or value of the ID variable of the object to which the observation pertains when CONDITION=ROW
- `_NAME_`, a character variable of length 8 containing the variable name of the object or dimension to which the observation pertains when CONDITION=ROW
- N, the number of nonmissing data
- WEIGHT, the weight of the partition
- CRITER, the badness-of-fit criterion
- DISCORR, the correlation between the transformed data and the distances for LEVEL=ORDINAL or the correlation between the data and the transformed distances otherwise
- UDISCORR, the correlation uncorrected for the mean between the transformed data and the distances for LEVEL=ORDINAL or the correlation between the data and the transformed distances otherwise
- FITCORR, the correlation between the fit-transformed data and the fit-transformed distances
- UFITCORR, the correlation uncorrected for the mean between the fit-transformed data and the fit-transformed distances

---

## OUTRES= Data Set

The OUTRES= data set has one observation for each nonmissing datum. It contains the following variables:

- BY variables, if any
- `_ITER_` (if the OUTITER option is specified), a numeric variable containing the iteration number
- `_DIMENS_`, a numeric variable containing the number of dimensions

- `_MATRIX_` or the variable in the `MATRIX` statement, identifying the data matrix or subject to which the observation pertains
- `_ROW_`, containing the variable label or value of the ID variable of the row to which the observation pertains
- `_COL_`, containing the variable label or value of the ID variable of the column to which the observation pertains
- `DATA`, the original datum
- `TRANDATA`, the optimally transformed datum when `LEVEL=ORDINAL`
- `DISTANCE`, the distance computed from the `PROC MDS` model
- `TRANSDIST`, the optimally transformed distance when the `LEVEL=` option is not `ORDINAL` or `ABSOLUTE`
- `FITDATA`, the datum further transformed according to the `FIT=` option
- `FITDIST`, the distance further transformed according to the `FIT=` option
- `WEIGHT`, the combined weight of the datum based on the `WEIGHT` variable(s), if any, and the standardization specified by the `FORMULA=` option
- `RESIDUAL`, `FITDATA` minus `FITDIST`

To cause a datum to appear in the `OUTRES=` data set, yet be ignored in fitting the model, give the datum a nonmissing value but a 0 weight (see “`WEIGHT` Statement”).

---

## INITIAL= Data Set

The `INITIAL=` data set has the same structure as the `OUT=` data set but is not required to have all of the variables or observations that appear in the `OUT=` data set. You can use an `OUT=` data set previously created by `PROC MDS` (without the `OUTITER` option) as an `INITIAL=` data set in a subsequent invocation of the procedure.

The only variables that are required are `DIM1`, . . . , `DIM $m$`  (where  $m$  is the maximum number of dimensions) or equivalent variables specified in the `INVAR` statement. If these are the only variables, then all the observations are assumed to contain coordinates of the configuration; you cannot read dimension coefficients or transformation parameters.

To read initial values for the dimension coefficients or transformation parameters, the `INITIAL=` data set must contain the `_TYPE_` variable and either the variable specified in the `ID` statement or, if no `ID` statement is used, the variable `_NAME_`. In addition, if there is more than one data matrix, either the variable specified in the `MATRIX` statement or, if no `MATRIX` statement is used, the variable `_MATRIX_` or `_MATNUM_` is required.

If the `INITIAL=` data set contains the variable `_DIMENS_`, initial values are obtained from observations with the corresponding number of dimensions. If there is no `_DIMENS_` variable, the same observations are used for each number of dimensions analyzed.

If you want PROC MDS to read initial values from some but not all of the observations in the INITIAL= data set, use the WHERE= data set option to select the desired observations.

---

## Missing Values

Missing data in the similarity or dissimilarity matrices are ignored in fitting the model and are omitted from the OUTRES= data set. Any matrix that is completely missing is omitted from the analysis.

Missing weights are treated as 0.

Missing values are also allowed in the INITIAL= data set, but a large number of missing values may yield a degenerate initial configuration.

---

## Normalization of the Estimates

In multidimensional scaling models, the parameter estimates are not uniquely determined; the estimates can be transformed in various ways without changing their badness of fit. The initial and final estimates from PROC MDS are, therefore, normalized (unless you specify the NONORM option) to make it easier to compare results from different analyses.

The configuration always has a mean of 0 for each dimension.

With the COEF=IDENTITY option, the configuration is rotated to a principal-axis orientation. Unless you specify the LEVEL=ABSOLUTE option, the entire configuration is scaled so that the root-mean-square element is 1, and the transformations are adjusted to compensate.

With the COEF=DIAGONAL option, each dimension is scaled to a root-mean-square value of 1, and the dimension coefficients are adjusted to compensate. Unless you specify the LEVEL=ABSOLUTE option, the dimension coefficients are normalized as follows. If you specify the CONDITION=UN option, all of the dimension coefficients are scaled to a root-mean-square value of 1. For other values of the CONDITION= option, the dimension coefficients are scaled separately for each subject to a root-mean-square value of 1. In either case, the transformations are adjusted to compensate.

Each dimension is reflected to give a positive rank correlation with the order of the objects in the data set.

For the LEVEL=ORDINAL option, if the intercept, slope, or power parameters are fitted, the transformed data are normalized to eliminate these parameters if possible.



## Comparison with the ALSCAL Procedure

The MDS procedure generally produces results similar to those from the ALSCAL procedure (Young, Lewyckyj, and Takane 1986; Young 1982) if you use the following options in PROC MDS:

- FIT=SQUARED
- FORMULA=1 except for unfolding data, which require FORMULA=2
- PFINAL to get output similar to that from PROC ALSCAL

Unlike PROC ALSCAL, PROC MDS produces no plots, so you must use output data sets and PROC PLOT or PROC GPLOT.

The MDS and ALSCAL procedures may sometimes produce different results for the following reasons:

- With the LEVEL=INTERVAL option, PROC MDS fits a regression model while PROC ALSCAL fits a measurement model. These models are not equivalent if there is more than one partition, although the differences in the parameter estimates are usually minor.
- PROC MDS and PROC ALSCAL use different algorithms for initialization and optimization. Hence, different local optima may be found by PROC MDS and PROC ALSCAL for some data sets with poor fit. Using the INAV=SSCP option causes the initial estimates from PROC MDS to be more like those from PROC ALSCAL.
- The default convergence criteria in PROC MDS are more strict than those in PROC ALSCAL. The convergence measure in PROC ALSCAL may cause PROC ALSCAL to stop iterating because progress is slow rather than because a local optimum has been reached. Even if you run PROC ALSCAL with a very small convergence criterion and a very large iteration limit, PROC ALSCAL may never achieve the same degree of precision as PROC MDS. For most applications, this problem is of no practical consequence since two- or three-digit precision is sufficient. If the model does not fit well, obtaining higher precision may require hundreds of iterations.

PROC MDS accepts some PROC ALSCAL options as synonyms for the preceding options, as displayed in Table 40.1.

**Table 40.1.** PROC MDS Options Compared to PROC ALSCAL Options

| PROC ALSCAL Option | Accepted by PROC MDS? | Related PROC MDS Option or Comments     |
|--------------------|-----------------------|---|
| CONDITION=         | Yes                   |   |
| CONVERGE=          | Yes                   | Convergence measures are not comparable |
| CUTOFF=            | Yes                   |   |

Table 40.1. (continued)

| PROC ALSICAL Option | Accepted by PROC MDS? | Related PROC MDS Option or Comments  |
|---------------------|-----------------------|--|
| DATA=               | Yes                   |  |
| DEGREE=             | No                    |  |
| DIMENS=             | Yes                   |  |
| DIRECTIONS=         | No                    |  |
| HEADER              | Yes                   | Default in PROC MDS  |
| IN=                 | Yes                   |  |
| ITER=               | Yes                   | MAXITER=   |
| LEVEL=              | Yes                   | LEVEL=NOMINAL is not supported   |
| MAXDIM= <i>m</i>    | Yes                   | DIMENSION= <i>n</i> TO <i>m</i>  |
| MINDIM= <i>n</i>    | Yes                   | DIMENSION= <i>n</i> TO <i>m</i>  |
| MINSTRESS=          | Yes                   | MINCRIT=   |
| MODEL=EUCLID        | Yes                   | COEF=IDENTITY  |
| MODEL=INDSCAL       | Yes                   | COEF=DIAGONAL  |
| MODEL=GEMSCAL       | No                    |  |
| MODEL=ASYMSCAL      | No                    |  |
| MODEL=ASYMINDS      | No                    |  |
| NEGATIVE            | (Yes)                 | In PROC MDS, the NEGATIVE option affects slopes and powers, not subject weights. |
| NOULB               | Yes                   |  |
| OUT=                | Yes                   | Some differences in contents   |
| PLOT                | No                    |  |
| PLOTALL             | No                    |  |
| PRINT               | No                    |  |
| READV, etc.         | No                    | Use WHERE data set option  |
| READFIXV, etc.      | No                    |  |
| ROWS=               | No                    |  |
| SHAPE=SYMMETRI      | Yes                   | SHAPE=TRIANGLE   |
| SHAPE=ASYMMETR      | Yes                   | SHAPE=SQUARE   |
| SHAPE=RECTANGU      | No                    | Use SHAPE=TRIANGLE with extra missing values to fill out the matrix.             |
| SIMILAR             | Yes                   |  |
| TIESTORE=           | Yes                   | Ignored by PROC MDS  |
| UNTIE               | Yes                   |  |

## Comparison with the MLSCALE Procedure

Running the MDS procedure with the options

```
proc mds fit=log level=loginterval ... ;
```

generally produces results similar to using the MLSCALE procedure (Ramsay 1986) with the options

```
proc mlscale stvarnce=constant suvarnce=constant ... ;
```

Alternatively, using the FIT=DISTANCE option in the PROC MDS statement produces results similar to specifying the NORMAL option in the PROC MLSCALE statement.

The MDS procedure uses the least-squares method of estimation. The least-squares method is equivalent to the maximum-likelihood method if the error terms are assumed to be independent and identically distributed normal random variables. Unlike PROC MLSCALE, PROC MDS does not provide any options for unequal error variances.

PROC MDS accepts some PROC MLSCALE options as synonyms for the options described previously, as displayed in Table 40.2.

**Table 40.2.** PROC MDS Options Compared to PROC MLSCALE Options

| PROC MLSCALE Option | Accepted by PROC MDS? | Related PROC MDS Option or Comments |
|---------------------|-----------------------|-------------------------------------|
| SQUARE              | Yes                   | SHAPE=SQUARE                        |
| INPUT=MATRIX        | No                    | Default                             |
| INPUT=VECTOR        | No                    |                                     |
| STLABEL=            | No                    | ID statement                        |
| STLBDS              | No                    |                                     |
| SULABEL=            | No                    | MATRIX statement                    |
| SULBDS              | No                    |                                     |
| CONFIG              | No                    |                                     |
| CONFDS=             | No                    | IN= data set                        |
| NEQU=               | No                    |                                     |
| CONSDS=             | No                    |                                     |
| METVAL              | No                    |                                     |
| METVDS              | No                    | IN=                                 |
| SEWGTS              | No                    |                                     |
| SEWGDS=             | No                    |                                     |
| SPLVAL              | No                    |                                     |
| SLPVDS=             | No                    |                                     |
| DIMENS=             | Yes                   |                                     |
| METRIC=IDENTITY     | Yes                   | COEF=IDENTITY                       |
| METRIC=DIAGONAL     | Yes                   | COEF=DIAGONAL                       |
| METRIC=FULL         | No                    |                                     |

Table 40.2. (continued)

| PROC MLSCALE Option | Accepted by PROC MDS? | Related PROC MDS Option or Comments |
|---------------------|-----------------------|-------------------------------------|
| TRANSFRM=SCALE      | Yes                   | LEVEL=RATIO                         |
| TRANSFRM=POWER      | Yes                   | LEVEL=LOGINTERVAL                   |
| TRANSFRM=SPLINE     | No                    |                                     |
| STVARNCE=           | No                    |                                     |
| SUVARNCE=           | No                    |                                     |
| NORMAL              | No                    | Default (FIT=DISTANCE)              |
| ITMAX=              | Yes                   | MAXITER=                            |
| ITXMAX=             | No                    |                                     |
| ITWMAX=             | No                    |                                     |
| ITAMAX=             | No                    |                                     |
| ITPMAX=             | No                    |                                     |
| CONV=               | (Yes)                 | Meaning is different                |
| FACTOR=             | No                    |                                     |
| HISTORY             | No                    | PITER                               |
| ASYMP               | No                    |                                     |
| OUTCON              | No                    | OUT=                                |
| OUTDIS              | No                    |                                     |
| OUTMET              | No                    | OUT=                                |
| OUTSPL              | No                    |                                     |
| OUTRES              | (Yes)                 | OUTRES= data set                    |

---

## Displayed Output

Unless you specify the NOPHIST option, PROC MDS displays the iteration history containing

- Iteration number
- Type of iteration:
 

|          |                          |
|----------|--------------------------|
| Initial  | initial configuration    |
| Monotone | monotone transformation  |
| Gau-New  | Gauss-Newton step        |
| Lev-Mar  | Levenberg-Marquardt step |
- Badness-of-Fit Criterion
- Change in Criterion
- Convergence Measures:
 

|          |  |
|----------|--|
| Monotone | the Euclidean norm of the change in the optimally scaled data divided by the Euclidean norm of the optimally scaled data, averaged across partitions |
| Gradient | the multiple correlation of the Jacobian matrix with the residual vector, uncorrected for the mean   |

Depending on what options are specified, PROC MDS may also display the following tables:

- Data Matrix and possibly Weight Matrix for each subject
- Eigenvalues from the computation of the initial coordinates
- Sum of Data Weights and Pooled Data Matrix computed during initialization with INAV=DATA
- Configuration, the estimated coordinates of the objects
- Dimension Coefficients
- A table of transformation parameters, including one or more of the following:
  - Intercept
  - Slope
  - Power
- A table of fit statistics for each matrix and possibly each row, including
  - Number of Nonmissing Data
  - Weight of the matrix or row, allowing for both observation weights and standardization factors
  - Badness-of-Fit Criterion
  - Distance Correlation computed between the distances and data with optimal transformation
  - Uncorrected Distance Correlation not corrected for the mean
  - Fit Correlation computed after applying the FIT= transformation to both distances and data
  - Uncorrected Fit Correlation not corrected for the mean

---

## ODS Table Names

PROC MDS assigns a name to each table it creates. You can use these names to reference the table when using the Output Delivery System (ODS) to select tables and create output data sets. These names are listed in the following table. For more information on ODS, see Chapter 15, “Using the Output Delivery System.”

**Table 40.3.** ODS Tables Produced in PROC MDS

| ODS Table Name    | Description   | Option                     |
|-------------------|---|----------------------------|
| ConvergenceStatus | Convergence status  | default                    |
| DimensionCoef     | Dimension coefficients  | PCOEF w/COEF= not IDENTITY |
| FitMeasures       | Measures of fit   | PFIT                       |
| IterHistory       | Iteration history   | default                    |
| PConfig           | Estimated coordinates of the objects in the configuration                   | PCONFIG                    |
| PData             | Data matrices   | PDATA                      |
| PInAvData         | Initial sum of weights and weighted average of data matrices with INAV=DATA | PINAVDATA                  |

Table 40.3. (continued)

| ODS Table Name  | Description               | Option                                      |
|-----------------|---------------------------|---|
| PInEigval       | Initial eigenvalues       | PINEIGVAL                                   |
| PInEigvec       | Initial eigenvectors      | PINEIGVEC                                   |
| PInWeight       | Initialization weights    | PINWEIGHT                                   |
| Transformations | Transformation parameters | PTRANS w/LEVEL=RATIO, INTERVAL, LOGINTERVAL |

---

## Example

---

### Example 40.1. Jacobowitz Body Parts Data from Children and Adults

Jacobowitz (1975) collected conditional rank-order data regarding perceived similarity of parts of the body from children of ages 6, 8, and 10 years and from college sophomores. The following analysis includes data from 15 children (6-year-olds) and 15 sophomores. The method of data collection and some results of an analysis are also described by Young (1987, pp. 4–10). A portion of the data is included below. See “Mds Documentation Examples” in the SAS/STAT Sample Program Library for the complete data set.

```

data body;
  title 'Jacobowitz Body Parts Data from 6 Yr Olds and
        Adults';
  title2 'First 15 subjects (obs 1-225) are children';
  title3 'Second 15 subjects (obs 226-450) are adults';
  input (cheek face mouth head ear body arm elbow hand
        palm finger leg knee foot toe) (2.);
  if _n_ <= 225 then subject='C'; else subject='A';
  datalines;
0 2 1 3 410 5 9 6 7 811121314
2 012 113 3 81011 9 7 4 5 614
3 2 0 1 4 9 511 6 7 810131214
2 1 3 0 4 9 5 611 7 810121314
10 111 2 0 6 3 4 51213 7 814 9
1412 9 613 0 8 7 51011 1 4 2 3
1214111013 5 0 4 1 3 2 6 9 7 8
5 714 8 6 9 1 0 2 3 410111213
1311121014 9 3 4 0 1 2 6 5 7 8
8 6 7 9 4 5 310 1 0 212111314
14 513 6 912 3 4 1 2 0 7 81011
14121311 9 7 4 6 5 310 0 8 1 2
1211141013 4 5 8 6 7 9 1 0 2 3
1214101311 9 4 5 8 6 7 2 3 0 1
13 8 91114 3 6 5 71012 2 4 1 0
0 4 2 311 91412 1 713 8 6 510
7 011 9 1 2 8 3131410 612 4 5

```

```

1011 0 3 7 1 813 212 6 914 5 4
 4 811 0 1 2 3 9 71310 51412 6
14 61110 0 1 2 4 9 8 5 713 312
11 61412 3 013 2 1 9 5 410 7 8
1412 8 3 1 7 013 2 6 5 911 410
 610 914 3 412 0 813 5 711 1 2
131412 1 4 6 2 7 010 9 311 8 5
 6 9 711 3 5 112 2 013 810 414
10 6 9 81113 712 2 1 014 5 3 4
 6 9 710 4 21214 1 813 0 3 511
13 2 8 3 1 5 9141211 7 6 010 4
1411 8 413 212 9 5 310 6 7 0 1
14 3 4 713 6 2 8 91110 112 5 0

```

... 405 lines omitted ...

```

0 1 2 4 3141210 6 5 81311 7 9
2 0 3 1 414 810 5 612 911 713
2 1 0 4 31412 9 511 71310 6 8
2 1 4 0 314 810 61112 7 9 513
1 3 2 4 01412 9 8 7 5131011 6
131011 114 0 3 7 512 8 2 6 4 9
13 911101214 0 1 3 6 5 2 4 7 8
1012 813 914 1 0 611 4 3 2 5 7
12 911101314 5 4 0 1 3 8 6 2 7
 91211131014 5 4 1 0 2 6 7 3 8
10121113 914 6 5 1 3 0 8 7 4 2
121011 91314 2 5 6 8 7 0 1 3 4
1112 9131014 3 2 7 8 4 1 0 5 6
121011 91314 5 8 1 6 7 3 4 0 2
10121113 914 8 7 4 6 2 3 5 1 0
;

```

The data are analyzed as row conditional (CONDITION=ROW) at the ordinal level of measurement (LEVEL=ORDINAL) using the weighted Euclidean model (COEF=DIAGONAL) in three dimensions (DIMENSION=3). The final estimates are displayed (PFINAL). The estimates (OUT=OUT) and fitted values (OUTRES=RES) are saved in output data sets. The following statements produce Output 40.1.1:

```

options ps=60;
proc mds data=body
  condition=row
  level=ordinal
  coef=diagonal
  dimension=3
  pfinal
  out=out
  outres=res
;
subject subject;
title5 'Nonmetric Weighted MDS';
run;

```

## Output 40.1.1. Iteration History and Final Estimates for Body Parts Data

```

Jacobowitz Body Parts Data from 6 Yr Olds and Adults
First 15 subjects (obs 1-225) are children
Second 15 subjects (obs 226-450) are adults

Nonmetric Weighted MDS

Multidimensional Scaling: Data=WORK.BODY
Shape=SQUARE Condition=ROW Level=ORDINAL
Coef=DIAGONAL Dimension=3 Formula=1 Fit=1

Mconverge=0.01 Gconverge=0.01 Maxiter=100 Over=2 Ridge=0.0001 Alternate=MATRIX

```

| Iteration | Type     | Badness-<br>of-Fit<br>Criterion | Change in<br>Criterion | Convergence Measures |          |
|-----------|----------|---------------------------------|------------------------|----------------------|----------|
|           |          |                                 |                        | Monotone             | Gradient |
| 0         | Initial  | 0.4091                          | .                      | .                    | .        |
| 1         | Monotone | 0.2053                          | 0.2038                 | 0.3012               | 0.3190   |
| 2         | Gau-New  | 0.1937                          | 0.0116                 | .                    | .        |
| 3         | Monotone | 0.1862                          | 0.007533               | 0.0410               | 0.2314   |
| 4         | Gau-New  | 0.1847                          | 0.001496               | .                    | .        |
| 5         | Monotone | 0.1779                          | 0.006754               | 0.0372               | 0.1516   |
| 6         | Gau-New  | 0.1773                          | 0.000654               | .                    | .        |
| 7         | Monotone | 0.1758                          | 0.001488               | 0.0187               | 0.0975   |
| 8         | Gau-New  | 0.1755                          | 0.000253               | .                    | .        |
| 9         | Monotone | 0.1751                          | 0.000478               | 0.0104               | 0.0767   |
| 10        | Gau-New  | 0.1750                          | 0.000113               | .                    | .        |
| 11        | Monotone | 0.1748                          | 0.000199               | 0.006467             | 0.0582   |
| 12        | Gau-New  | 0.1747                          | 0.0000592              | .                    | 0.0384   |
| 13        | Gau-New  | 0.1747                          | 0.0000184              | .                    | 0.009952 |

Convergence criteria are satisfied.

```

Configuration

```

|        | Dim1  | Dim2  | Dim3  |
|--------|-------|-------|-------|
| cheek  | 1.46  | 0.77  | -0.30 |
| face   | 1.40  | -0.56 | 0.43  |
| mouth  | 1.25  | -0.97 | -0.51 |
| head   | 1.31  | 0.39  | 0.97  |
| ear    | 1.23  | 0.03  | -1.39 |
| body   | 0.24  | -0.18 | 2.24  |
| arm    | -0.70 | 1.23  | 1.08  |
| elbow  | -0.73 | 0.40  | 0.29  |
| hand   | -0.71 | 1.50  | -0.26 |
| palm   | -0.24 | 1.59  | -0.96 |
| finger | -0.74 | 0.57  | -1.35 |
| leg    | -0.69 | -1.44 | 0.88  |
| knee   | -1.44 | -0.64 | 0.42  |
| foot   | -0.78 | -1.51 | -0.36 |
| toe    | -0.87 | -1.20 | -1.19 |



Jacobowitz Body Parts Data from 6 Yr Olds and Adults  
 First 15 subjects (obs 1-225) are children  
 Second 15 subjects (obs 226-450) are adults

Nonmetric Weighted MDS

Multidimensional Scaling: Data=WORK.BODY  
 Shape=SQUARE Condition=ROW Level=ORDINAL  
 Coef=DIAGONAL Dimension=3 Formula=1 Fit=1

Dimension Coefficients

| subject | 1    | 2    | 3    |
|---------|------|------|------|
| C       | 1.18 | 1.02 | 0.76 |
| C       | 0.94 | 1.05 | 1.01 |
| C       | 0.94 | 1.01 | 1.05 |
| C       | 1.08 | 1.10 | 0.79 |
| C       | 0.98 | 1.05 | 0.97 |
| C       | 1.28 | 0.85 | 0.79 |
| C       | 0.90 | 1.03 | 1.06 |
| C       | 0.95 | 1.04 | 1.01 |
| C       | 1.18 | 1.06 | 0.69 |
| C       | 0.95 | 0.99 | 1.05 |
| C       | 1.07 | 1.13 | 0.76 |
| C       | 0.98 | 0.94 | 1.08 |
| C       | 1.34 | 0.82 | 0.73 |
| C       | 0.95 | 0.99 | 1.05 |
| C       | 1.03 | 1.09 | 0.86 |
| A       | 1.30 | 0.82 | 0.80 |
| A       | 1.05 | 0.64 | 1.22 |
| A       | 1.28 | 0.85 | 0.81 |
| A       | 0.95 | 0.88 | 1.15 |
| A       | 1.34 | 0.75 | 0.80 |
| A       | 1.27 | 0.81 | 0.85 |
| A       | 1.50 | 0.37 | 0.78 |
| A       | 1.38 | 0.82 | 0.64 |
| A       | 1.44 | 0.56 | 0.78 |
| A       | 1.53 | 0.73 | 0.36 |
| A       | 1.18 | 0.95 | 0.84 |
| A       | 0.85 | 0.86 | 1.24 |
| A       | 1.29 | 0.62 | 0.97 |
| A       | 1.20 | 0.58 | 1.10 |
| A       | 1.25 | 0.61 | 1.04 |

```

Jacobowitz Body Parts Data from 6 Yr Olds and Adults
First 15 subjects (obs 1-225) are children
Second 15 subjects (obs 226-450) are adults

Nonmetric Weighted MDS

Multidimensional Scaling: Data=WORK.BODY
Shape=SQUARE Condition=ROW Level=ORDINAL
Coef=DIAGONAL Dimension=3 Formula=1 Fit=1

```

| subject | Number of<br>Nonmissing<br>Data | Weight | Badness-of-<br>Fit<br>Criterion | Distance<br>Correlation | Uncorrected<br>Distance<br>Correlation |
|---------|---------------------------------|--------|---------------------------------|-------------------------|--|
| C       | 210                             | 0.03   | 0.16                            | 0.85                    | 0.99                                   |
| C       | 210                             | 0.03   | 0.25                            | 0.51                    | 0.97                                   |
| C       | 210                             | 0.03   | 0.23                            | 0.58                    | 0.97                                   |
| C       | 210                             | 0.03   | 0.16                            | 0.85                    | 0.99                                   |
| C       | 210                             | 0.03   | 0.21                            | 0.69                    | 0.98                                   |
| C       | 210                             | 0.03   | 0.15                            | 0.89                    | 0.99                                   |
| C       | 210                             | 0.03   | 0.26                            | 0.40                    | 0.96                                   |
| C       | 210                             | 0.03   | 0.25                            | 0.47                    | 0.97                                   |
| C       | 210                             | 0.03   | 0.13                            | 0.91                    | 0.99                                   |
| C       | 210                             | 0.03   | 0.24                            | 0.55                    | 0.97                                   |
| C       | 210                             | 0.03   | 0.15                            | 0.87                    | 0.99                                   |
| C       | 210                             | 0.03   | 0.23                            | 0.59                    | 0.97                                   |
| C       | 210                             | 0.03   | 0.15                            | 0.90                    | 0.99                                   |
| C       | 210                             | 0.03   | 0.21                            | 0.69                    | 0.98                                   |
| C       | 210                             | 0.03   | 0.19                            | 0.75                    | 0.98                                   |
| A       | 210                             | 0.03   | 0.12                            | 0.94                    | 0.99                                   |
| A       | 210                             | 0.03   | 0.17                            | 0.84                    | 0.98                                   |
| A       | 210                             | 0.03   | 0.12                            | 0.93                    | 0.99                                   |
| A       | 210                             | 0.03   | 0.21                            | 0.68                    | 0.98                                   |
| A       | 210                             | 0.03   | 0.14                            | 0.91                    | 0.99                                   |
| A       | 210                             | 0.03   | 0.16                            | 0.87                    | 0.99                                   |
| A       | 210                             | 0.03   | 0.09                            | 0.98                    | 1.00                                   |
| A       | 210                             | 0.03   | 0.11                            | 0.95                    | 0.99                                   |
| A       | 210                             | 0.03   | 0.11                            | 0.96                    | 0.99                                   |
| A       | 210                             | 0.03   | 0.13                            | 0.95                    | 0.99                                   |
| A       | 210                             | 0.03   | 0.13                            | 0.90                    | 0.99                                   |
| A       | 210                             | 0.03   | 0.21                            | 0.72                    | 0.98                                   |
| A       | 210                             | 0.03   | 0.12                            | 0.93                    | 0.99                                   |
| A       | 210                             | 0.03   | 0.10                            | 0.95                    | 0.99                                   |
| A       | 210                             | 0.03   | 0.11                            | 0.94                    | 0.99                                   |
| - All - | 6300                            | 1.00   | 0.17                            | 0.84                    | 0.98                                   |

The OUTRES= data set is used to produce a plot showing the overall fit of the model, with the transformed data on the vertical axis and the distances from the model on the horizontal axis. If the model fits perfectly, all points lie on a diagonal line from lower left to upper right. The vertical departure of each point from this diagonal line represents the residual of the corresponding observation. The HAXIS and VAXIS options in the PLOT statement specify that the horizontal and vertical axes use the definitions AXIS2 and AXIS1, respectively. The identical ORDER= options in each of these AXIS statements make physical distances on each axis comparable.

The following statements produce Output 40.1.2:

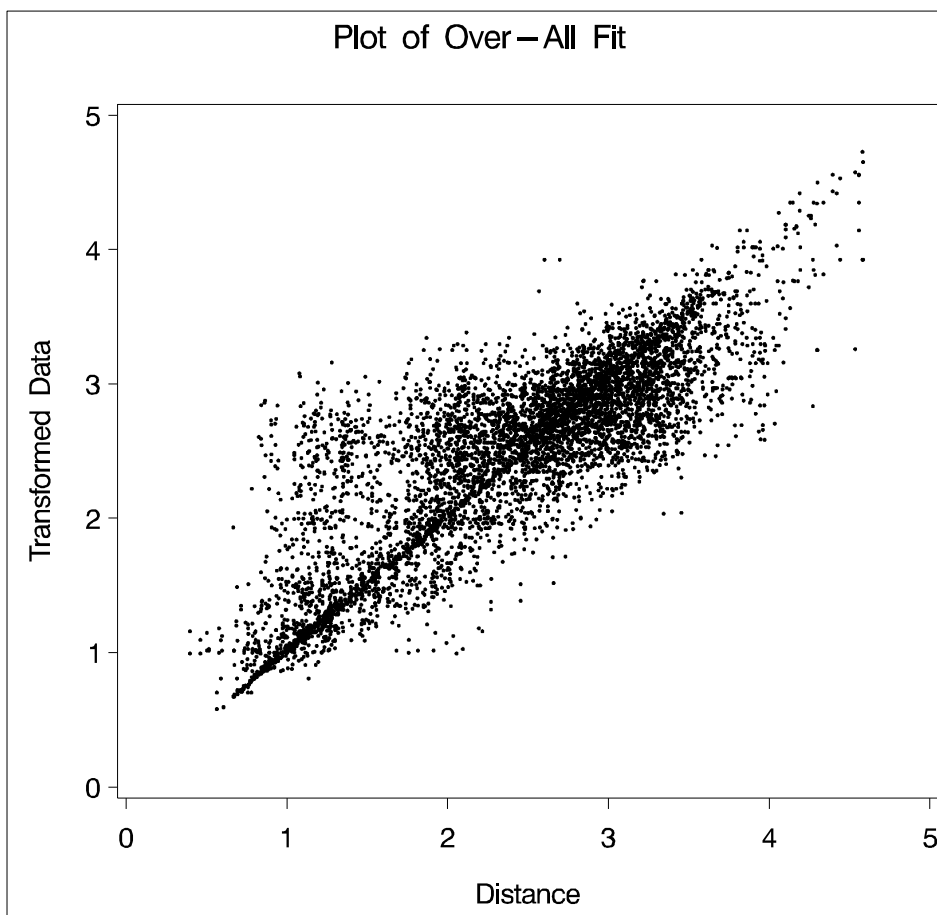
```

title1 'Plot of Over-All Fit';
axis1 label=(angle=90 rotate=0) minor=none
      order=(0 to 5 by 1);
axis2 minor=none order=(0 to 5 by 1);

proc gplot data=res;
  plot fitdata*fitdist/vaxis=axis1 haxis=axis2
      frame cframe=ligr;
run;

```

**Output 40.1.2.** Plot of Over-All Fit for Body Parts Data



The OUT= data set is used to plot the configuration and dimension coefficients using the %PLOTIT macro. Again it is necessary to use the VTOH= option to make the axes commensurable. The configuration is plotted by selecting observations having \_TYPE\_='CONFIG' with a WHERE statement and by using the \_NAME\_ variable to identify each body part on the plot.

The dimension coefficients are plotted by selecting observations having `_TYPE_='DIAGCOEF'` and by using the `Subject` variable to distinguish children from adults on the plot. The following statements produce Output 40.1.3:

```

title1 'Plot of configuration';
%plotit(data=out(where=( _type_='CONFIG')), datatype=mds,
        labelvar=_name_, vtoh=1.75);
%plotit(data=out(where=( _type_='CONFIG')), datatype=mds,
        plotvars=dim3 dim1, labelvar=_name_, vtoh=1.75);
run;

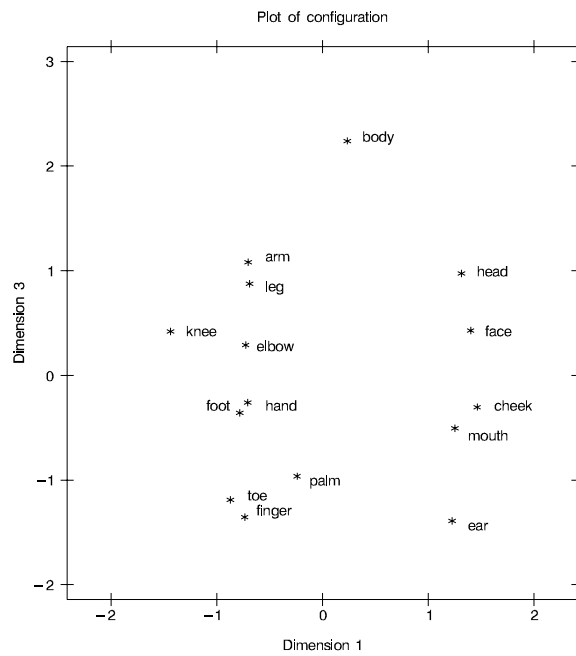
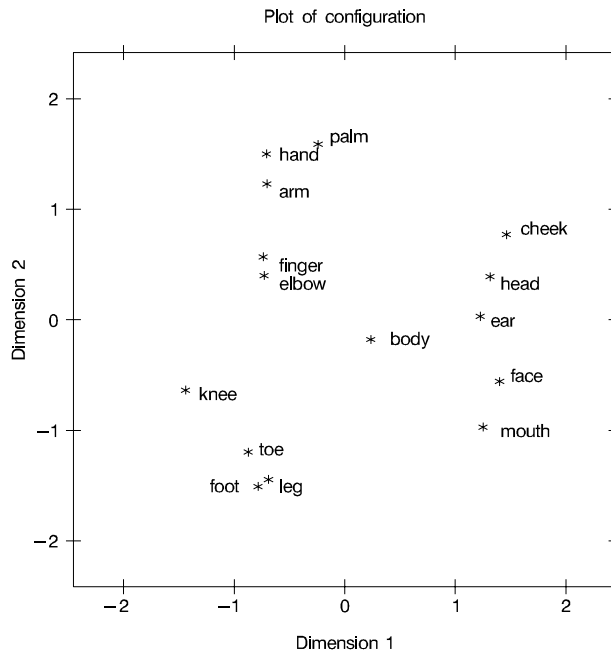
title1 'Plot of Dimension Coefficients for Each Subject';
%plotit(data=out(where=( _type_='DIAGCOEF')), symtype=vector,
        symbols='', datatype=mds, place=0, labelvar=subject,
        vechad=, ls=100, plotopts=hzero vzero);

%plotit(data=out(where=( _type_='DIAGCOEF')),
        plotvars=dim3 dim1, symtype=vector, symbols='',
        datatype=mds, place=0, labelvar=subject, ls=100,
        vechad=, plotopts=hzero vzero);
run;

```

The configuration displayed in Output 40.1.3 has a tripodal shape with `Body` at the apex. The three legs of the tripod can be distinguished in the plot of dimension 2 by dimension 1, which shows three distinct clusters with `Body` in the center. Dimension 1 separates head parts from arm and leg parts. Dimension 2 separates arm parts from leg parts. The plot of dimension 3 by dimension 1 shows the tripod from the side. Dimension 3 distinguishes the more inclusive body parts (at the top) from the less inclusive body parts (at the bottom).

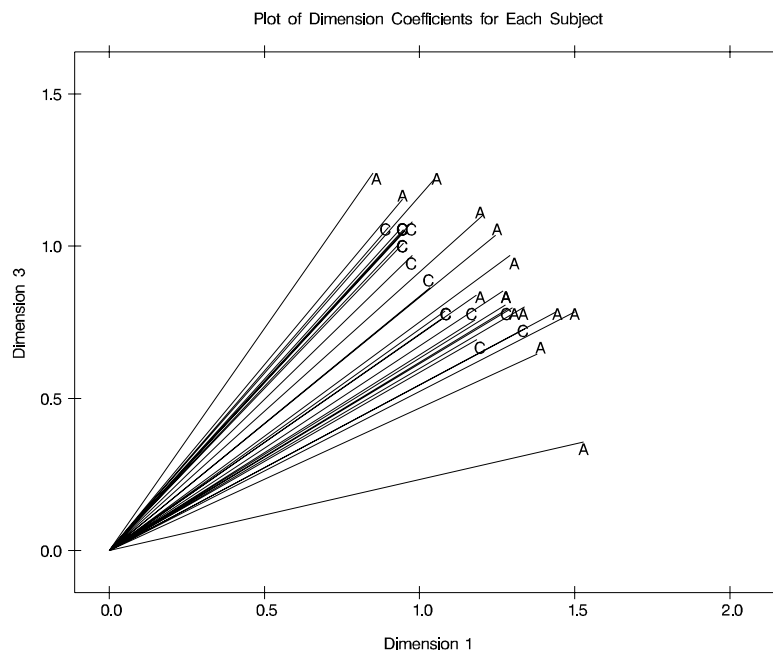
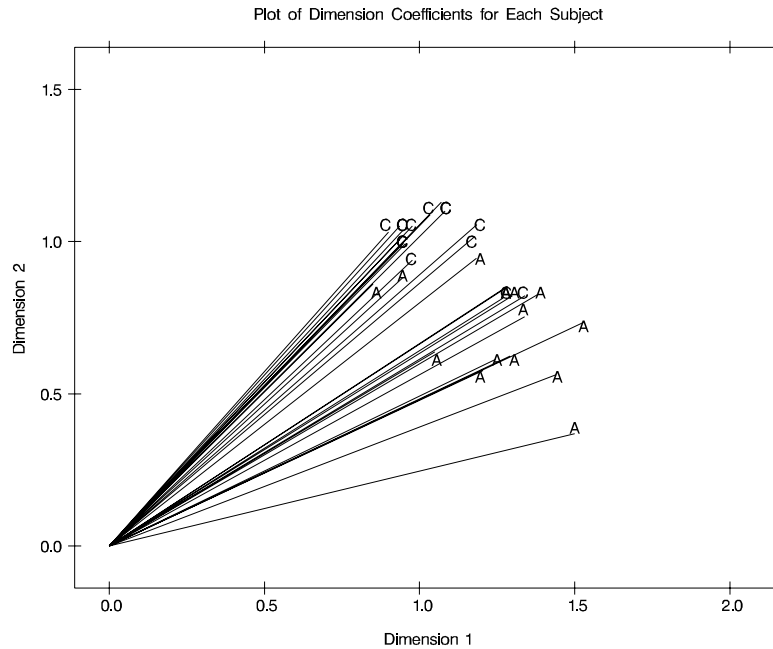
**Output 40.1.3.** Plot of Configuration for Body Parts Data



The plots of dimension coefficients in Output 40.1.4 show that children differ from adults primarily in the emphasis given to dimension 2. Children give about the same weight (approximately 1) to each dimension. Adults are much more variable than children, but all have coefficients less than 1.0 for dimension 2, with an average of

about 0.7. Referring back to the configuration plot, you can see that adults consider arm parts to be more similar to leg parts than do children. Many adults also give a high weight to dimension 1, indicating that they consider head parts to be more dissimilar from arm and leg parts than do children. Dimension 3 shows considerable variability for both children and adults.

**Output 40.1.4.** Plot of Dimension Coefficients for Body Parts Data



---

## References

- Arabie, P., Carroll, J.D., and DeSarbo, W.S. (1987), *Three-Way Scaling and Clustering*, Sage University Paper series on Quantitative Applications in the Social Sciences, 07-065, Beverly Hills and London: Sage Publications.
- Carroll, J.D. and Chang, J.J. (1970), "Analysis of Individual Differences in Multidimensional Scaling via an N-way Generalization of the 'Eckart-Young' Decomposition," *Psychometrika*, 35, 283–319.
- Davison, M.L. (1983), *Multidimensional Scaling*, New York: John Wiley & Sons, Inc.
- Heiser, W.J. (1981), *Unfolding Analysis of Proximity Data*, Leiden: Department of Datatheory, University of Leiden.
- Jacobowitz, D. (1975), *The Acquisition of Semantic Structures*, Doctoral dissertation, University of North Carolina at Chapel Hill.
- Krantz, D.H., Luce, R.D., Suppes, P., and Tversky, A. (1971), *Foundations of Measurement*, New York: Academic Press.
- Kruskal, J.B. and Wish, M. (1978), *Multidimensional Scaling*, Sage University Paper series on Quantitative Applications in the Social Sciences, 07–011, Beverly Hills and London: Sage Publications.
- Kuhfeld, W.F. (1994), "Graphical Scatter Plots of Labeled Points," *Observations*, Fourth Quarter, 1994, Cary, NC: SAS Institute Inc.
- Null, C.H. and Sarle, W.S. (1982), "Multidimensional Scaling by Least Squares," in *Proceedings of the Seventh Annual SAS Users Group International Conference*, Cary, NC: SAS Institute Inc.
- Rabinowitz, G. (1976), "A Procedure for Ordering Object Pairs Consistent with the Multidimensional Unfolding Model," *Psychometrika*, 41, 349–373.
- Ramsay, J.O. (1986), "The MLSCALE Procedure," in *SUGI Supplemental Library User's Guide, Version 5 Edition*, Cary, NC: SAS Institute Inc.
- Schiffman, S.S., Reynolds, M.L., and Young, F.W. (1981), *Introduction to Multidimensional Scaling*, New York: Academic Press.
- Torgerson, W.S. (1958), *Theory and Methods of Scaling*, New York: John Wiley & Sons, Inc.
- Young, F.W. (1982), "Enhancements in ALSCAL-82," in *Proceedings of the Seventh Annual SAS Users Group International Conference*, Cary, NC: SAS Institute Inc.
- Young, F.W. (1987), *Multidimensional Scaling: History, Theory, and Applications*, ed. R.M. Hamer, Hillsdale, NJ: Lawrence Erlbaum Associates.
- Young, F.W., Lewyckij, R., and Takane, Y. (1986), "The ALSCAL Procedure," in *SUGI Supplemental Library User's Guide, Version 5 Edition*, Cary, NC: SAS Institute Inc.

The correct bibliographic citation for this manual is as follows: SAS Institute Inc., *SAS/STAT® User's Guide, Version 8*, Cary, NC: SAS Institute Inc., 1999.

**SAS/STAT® User's Guide, Version 8**

Copyright © 1999 by SAS Institute Inc., Cary, NC, USA.

ISBN 1-58025-494-2

All rights reserved. Produced in the United States of America. No part of this publication may be reproduced, stored in a retrieval system, or transmitted, in any form or by any means, electronic, mechanical, photocopying, or otherwise, without the prior written permission of the publisher, SAS Institute Inc.

**U.S. Government Restricted Rights Notice.** Use, duplication, or disclosure of the software and related documentation by the U.S. government is subject to the Agreement with SAS Institute and the restrictions set forth in FAR 52.227-19 Commercial Computer Software-Restricted Rights (June 1987).

SAS Institute Inc., SAS Campus Drive, Cary, North Carolina 27513.

1st printing, October 1999

SAS® and all other SAS Institute Inc. product or service names are registered trademarks or trademarks of SAS Institute Inc. in the USA and other countries.® indicates USA registration.

Other brand and product names are registered trademarks or trademarks of their respective companies.

The Institute is a private company devoted to the support and further development of its software and related services.