

PROSCAL PROFESSIONAL: A Program for Probabilistic Scaling

January, 2014

David B. MacKay
Kelley School of Business
Indiana University
Bloomington, IN 47405
mackay@proscal.com

Joseph L. Zinnes
JDM Research
9731 Silver Dew St.
Las Vegas, NV 89123
zinnes@proscal.com

CONTENTS

I	Introduction	4
II	What's New	5
III	PROSCAL Parameter Estimation	5
IV	Tests of Hypotheses	7
	A. Variance Structure	8
	B. Location	10
	C. Dimensionality	10
	D. Metric Properties	11
	E. Common Space Assumption	11
	F. Sampling Properties	12
	G. Measurement Models	12
V	PROSCAL Options	12
	A. Location, Variance and Measurement Targets	12
	B. Proximities	13
	C. Standardization	13
	D. Hedonics	14
	E. Segmentation Analysis	14
	F. Internal and External Analysis	15
	G. Density Function Approximations	15
	H. Optimization Algorithms	15
VI	User's Guide	15
	A. Setup	15
	B. Data Sets	17
	C. PROSCAL Menu	18
	D. Generate Design	19
	D.1 The Project Tab	20
	D.2 The Parameters Tab	26
	D.3 The Targets Tab	30
	D.4 The Execution Tab	32
	E. Run PROSCAL	33
	F. Examine Output	33
	G. 2D Plot - Centroids	34
	H. 2D Plot - Standard Deviations	34
	I. 3D Plot – Centroids	38
	J. 3D Plot – Standard Deviations	39
	K. Change Labels	41
	L. Perceptual Share	42
	M. What-If Analysis	45

N.	About PROSCAL	45
O.	Limits	46
APPENDIX 1 Input Parameters for Data Examples		47
APPENDIX 2 Annotated Output of the Prox/Lik Data Set		54
APPENDIX 3 Segmentation Analysis Output		67
APPENDIX 4 Key for Instructional Variables' Nominal Output		68
REFERENCES		70

I. INTRODUCTION

PROSCAL[®] is a probabilistic multidimensional scaling program that represents objects by multivariate normal distributions and provides maximum likelihood estimates of each distribution's mean and variance parameters. The location of an object i is represented by the estimated mean of its coordinates $\hat{\mu}_{ik}$ on the $k = 1, \dots, r$ dimensions. The coordinates of all objects are referred to as the configuration. Objects may be real or ideal.

The probabilistic model upon which PROSCAL is based was first proposed by Hefner (1958) for use with same-different judgments. For objects i, j in a r dimensional space, the distances d_{ij} were assumed to have Euclidean properties, specifically,

$$d_{ij}^2 = \sum_{k=1}^r (x_{ik} - x_{jk})^2,$$

and the coordinates x_{ik} were assumed to be normally and independently distributed with mean μ_{ik} and variance σ_{ik}^2 . These assumptions have been relaxed. You are now allowed to specify unique variances σ_{ik}^2 for each object i on each dimension k in a Euclidean or city-block space.

PROSCAL input data now include similarities, dissimilarities, preference ratios and binary preferential choices for pairs of real objects and liking ratings for individual real objects. For dissimilarities, preference ratios and liking ratings, high values are positively related to expected distances or expected distance ratios. Thus, a large expected distance between two real objects means that the objects are very dissimilar. A large expected distance between a real and ideal object means that the real object has a high disutility (low liking). A large expected distance ratio indicates that the real object in the numerator is less preferred (farther from the ideal) than the real object in the denominator. For similarities and binary choices, the relationship is inverse – high values indicate low expected distances or expected distance ratios. Similarity judgments are assumed to range from zero (low similarity) to one (high similarity). Binary preferential choices may have the values zero or one with a one indicating the real object in the numerator is preferred to the real object in the denominator. (See Section VI.D.1 for more detail.) All distances and distance ratios are assumed to be non-negative.

When data are collected from one subject, $\hat{\sigma}_{ik}$ - the estimate of σ_{ik} , may be interpreted as the subject's uncertainty in locating object i on dimension k . A high standard deviation indicates that the subject's knowledge or familiarity of the object is relatively vague or uncertain. When data are for more than one subject, the standard deviations can be interpreted as measures of the heterogeneity of the subjects' evaluations of the objects. Incomplete, complete or replicated data may generally be used for each subject. The one exception to this is when liking ratings are evaluated using a dependent sampling model (see Section IV.F), in which case, incomplete data are disallowed.

II WHAT'S NEW

PROSCAL's capabilities have been broadened and its input format has been changed.

New features include:

- 1) A simpler, user friendly graphical interface
- 2) Simultaneous segmentation and scaling of hedonic (dominance) data – liking ratings, binary choices, preference ratios – using mixture maximum likelihood estimates from an expectation maximization (EM) algorithm
- 3) “Perceptual share” estimation
- 4) What-if analysis
- 5) Similarity or dissimilarity judgments of proximity
- 6) Ability to scale liking ratings, either by themselves or simultaneously with proximity judgments or binary choices
- 7) Ability to scale binary choices, either by themselves or simultaneously with proximity judgments or liking ratings
- 8) Inclusion of one, two or three parameter measurement models for transforming proximity judgments or liking ratings
- 9) Inclusion of a power transformation measurement model for preference ratio judgments
- 10) Unfolding analysis using independent or dependent sampling, for all types of hedonic judgments in a Euclidean space
- 11) Unfolding analysis, external *or* internal, using a single ideal object
- 12) External unfolding based upon real objects having fixed coordinates or fixed coordinates and variances
- 13) Estimation of configurations whose coordinates (for real or ideal objects) may be constrained to equal one another
- 14) Estimation of expected distances, in a Euclidean or city-block metric, for all pairs of objects
- 15) Incorporation of a very quick, highly accurate, Pearsonian transformation to speed up anisotropic unfolding analyses
- 16) Three dimensional plots of configurations and variance structures which may be rotated by the user for better visualization of the results
- 17) Provision of CAIC and BIC statistics for hypothesis testing
- 18) Improved initial estimates
- 19) Dynamic array allocation

III PROSCAL PARAMETER ESTIMATION

PROSCAL uses maximum likelihood (ML) procedures to estimate the mean location $\hat{\mu}_{ik}$ and variance $\hat{\sigma}_{ik}^2$ of an object i on dimension k . A significant advantage of using ML procedures to estimate the parameters is that it makes it possible to test specific hypotheses concerning the data.

To obtain the estimates, it is necessary to specify the density function of the distance random variable d_{ij} , the distance ratio random variable $r_{ijk} = d_{ij}/d_{ik}$, or the binary choice probability p_{ijk} of subject i choosing product j over product k . The selection of the appropriate density function depends upon the variance structure, the choice of a metric and the sampling process.

Variances may be defined in an isotropic or anisotropic space. In an isotropic space, variances must, for any object, have the same values on all dimensions. In an anisotropic space, variances may differ from dimension to dimension. In general, Euclidean metrics use density functions that are based upon quadratic forms in normal variables distributions. City-block metrics use density functions that are based upon folded normal distributions. When Euclidean distances are present in an isotropic space, special forms are used for the likelihood functions. For proximity judgments and liking ratings, the likelihoods are based upon the non-central chi-square distribution. For preference ratios, the likelihoods are based upon the doubly non-central F distribution. Binary preferential choices use the preceding distributions in conjunction with the binomial distribution function. (Details on how the ML functions are derived may be found in MacKay 1989, 2001, 2007; MacKay and Lilly, 2004; MacKay and Zinnes 1995, Zinnes and MacKay 1983.)

When preference ratios are evaluated, PROSCAL will print a *rescaled configuration* and *rescaled variances*. These values, which have the same likelihood as the original solution, are rescaled so that the minimum and maximum coordinate values are -1.0 and 1.0 respectively. Rescaled initial solution values are used to start the maximum likelihood analysis.

When independently sampled preference ratios, binary preference judgments or liking ratings are evaluated, PROSCAL will also print *adjusted variances and covariances*. Adjusted variances and covariances, which have the same likelihood as the original solution, are rescaled so that the real objects and ideal objects have the same minimum variance value on each dimension. (Independent and dependent sampling are covered in Section IV.F.)

Measurement models may be specified to express dissimilarity judgments δ_{ij} as a function of underlying latent distances d_{ij} . This process is needed when, for example, dissimilarity judgments are simultaneously evaluated with liking ratings. It may, though, also be used when dissimilarity judgments are evaluated by themselves or when they are used in conjunction with preference ratio judgments or binary choices. One, two or three-coefficient measurement models may be specified. The relationship of the underlying distances and the observed dissimilarity judgments takes the form $\delta_{ij} = a + bd_{ij}^c$. The user may specify which of the measurement model coefficients, a , b and c , are to be estimated. Default values of a , b and c are 0, 1, and 1 respectively. When proximities are expressed as similarities s_{ij} , additional transformations, exponential $s_{ij} = \exp(-\delta_{ij})$ or Gaussian $s_{ij} = \exp(-\delta_{ij}^2)$, are used by PROSCAL to relate dissimilarities to similarities. In the psychology literature, exponential transformations are usually associated with the city-block metric and Gaussian transformations are often associated with the Euclidean metric. However, some authors (Shepard, 1987) prefer to use the exponential transformation with all metrics.

Measurement model transformations may also be defined for liking ratings. When proximities are simultaneously evaluated with liking ratings, the measurement model transformation is applied only to the proximity judgments. A power transformation may be used when preference ratio judgments are evaluated. Measurement model transformations are not used with binary choice data.

Maximum likelihood estimates maximize the value of the likelihood function. For the models described here, iterative numerical optimization methods must be used to find a solution. PROSCAL employs an iterative process which, after calculating initial estimates, enters an alternating ML estimation sequence. The usual sequence is to first estimate the variances, while fixing the coordinates and measurement model coefficients, then to estimate the coordinates and measurement model coefficients, while fixing the variances. This two-stage alternating sequence is repeated until improvement in the likelihood function reaches a cutoff value, at which point a final estimate with no fixed values is computed. When, however, an external analysis is used that fixes the coordinates of the real objects or when a mixture model is used to reassign subjects to segments, then the process is changed and a three stage sequence in which the measurement model coefficients are estimated first, the variances are estimated second and the coordinates are estimated third is followed.

IV TESTS OF HYPOTHESES

PROSCAL provides for the testing of hypotheses concerning the variance structure of the objects, locations of the objects, dimensionality of the space, metric properties of the space, the existence of a common space for proximity and hedonic judgments, dependent or independent sampling and the structure of the measurement model. All of the tests involve comparing a complex model to a simple model. When a complex model does not fit the data significantly better than a simpler model, the simpler model should be used.

The question of whether a complex model is significantly better than the simpler model can be answered by using a likelihood ratio test or information criterion statistics. Likelihood ratio tests may only be used with nested models. Information criterion statistics may be used with nested and non-nested models.

To use a likelihood ratio test, start with the log-likelihoods that are given by PROSCAL for each model. If L_c is the log likelihood of the complex model and L_s is the log likelihood of the simple model, then the quantity $2(L_c - L_s)$ has an asymptotic chi-square distribution $\chi_{(a)}^2(df)$ where the degrees of freedom df is equal to the difference in the number of free parameters estimated by the models. When evaluating $\chi_{(a)}^2(df)$ it is usually best to select a conservative (low) value for α , since maximum likelihood theorems are asymptotically defined.

A formula for computing the number of free parameters k in a model is

$$k = m + q + p + (s - 1) - r - r(r-1)/2 - 1$$

where,

m = the number of “active” coordinates (see Section IV.B)
 q = the number of unique variances being estimated
 p = the number of measurement model parameters being estimated (p may be zero)
 s = the number of segments estimated when a mixture model is used, one otherwise
 r = the dimensionality of the space

The last three terms of the equation are subtracted for the centering, rotation and scale invariance of the solution. When anisotropic or city-block models are used, the rotational invariance term should be omitted. (See MacKay, Zinnes and Easley, 1995). When an external analysis is used, all three terms are omitted.

Information criterion statistics, in particular those by Akaike (1974), Schwarz (1978) and Bozdogan (1987), are also widely used. Bozdogan’s statistic, called CAIC, is

$$CAIC = -2L + ck$$

where,

L = natural log likelihood of the model

$c = \ln(j) + 1$

j = the number of judgments entering the estimation process

k = the number of free parameters.

Schwarz’s Bayesian Information Criterion (BIC) is very similar, the only difference being that $c = \ln(j)$.

For most models, the number of judgments is self evident. Thus, in the complete data **Prox/Lik** example of Appendix 1, the number of judgments is 2,160 $((8*7/2)*20 + 8*200)$. However, when binary choice data are used, the data are summarized over the subjects in each segment. Thus, the number of binary data points entering the estimation process is $(n*(n-1)/2)*s$ where n is the number of real objects and s is the number of segments. If a missing piece of data is found in liking ratings characterized by dependent sampling (see Section F below) then all of the judgments for the n real objects in that observation set are considered missing. Otherwise, a missing datum has no effect on other data.

$CAIC$ and BIC are calculated for both the simple and complex model. The model with the lower value is selected.

A. Variance Structure

PROSCAL permits the estimation of a hierarchy of probabilistic scaling models. Models within the hierarchy differ in how they portray the variance structure of the objects.

Models that directly assign a variance to each object are called random coordinate or stimulus models. Three models may be distinguished for either isotropic or anisotropic variance structures. In an isotropic space, for n objects, we have the following:

S1: $\sigma_i^2 = \sigma^2 \quad i = 1, \dots, n$

All objects have the same variance; one variance is estimated. In an isotropic space, this is the original Hefner (1958) model.

S2: $\sigma_i^2 = \sigma_k^2 \quad i = 1, \dots, n; \quad k = 1, \dots, m; \quad m < n$

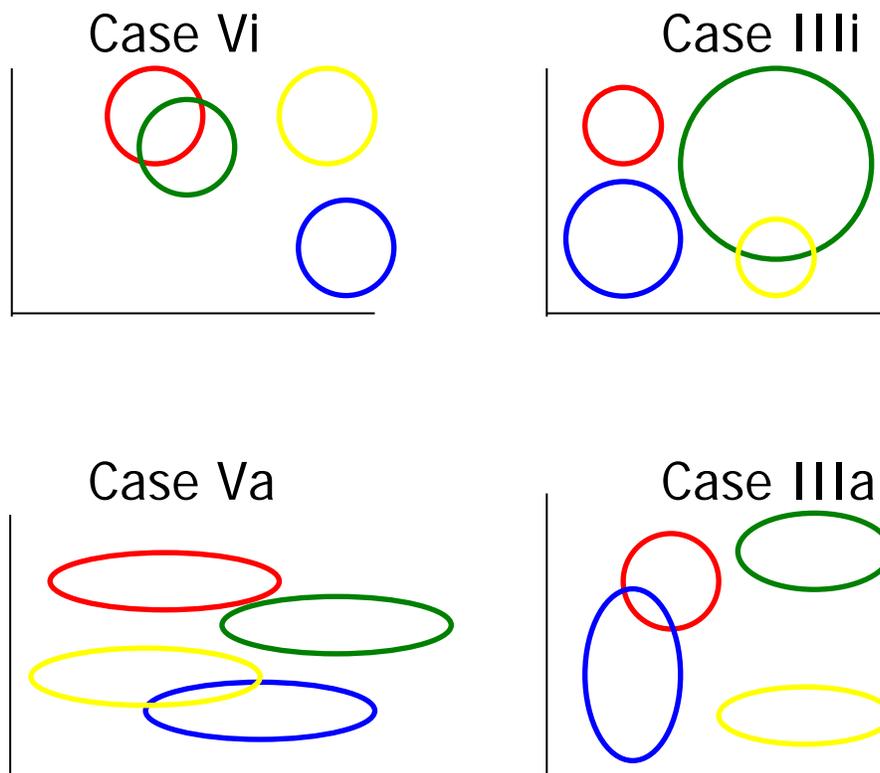
Objects are divided into m sets; m variances are estimated, one for each set. The number of sets and the composition of each set must be specified before parameter estimation begins.

S3: $\sigma_i^2 \quad i = 1, \dots, n$

Individual variances are estimated for each object; n variances are estimated. Using Thurstone's (1927) terminology, S1 refers to Case V models while S2 and S3 refer to Case III models.

Parallel distinctions may be made for anisotropic variance structures. The following figure illustrates the types of standard deviations that can occur when you use anisotropic (a) or isotropic (i) variance structures with Case V and Case III models.

Random coordinate model alternatives



B. Location

Coordinates of all real (and ideal) objects are usually active, which means that all of the coordinate values are to be estimated and that they are independent of each other. Occasionally, you may want to test whether two or more objects are perceived similarly. This situation occurs, for example, in product testing when you wish to see if consumers perceive a reformulated product in the same way that they perceive the original product. If one object is constrained to have the same coordinates as another, then the number of active coordinates is reduced by the number of dimensions.

Location constraints may also be used with dimensionality tests, as discussed below.

Inactive coordinates are often used when evaluating data from experimental designs. In a $p \times q$ design, for example, you may wish to constrain the row (column) coordinates of each object in the same row (column) to have the same value and estimate only $p + q$ coordinates.

C. Dimensionality

Information criterion statistics provide a convenient dimensionality test that can be used with isotropic and anisotropic analyses.

Example: Appendix 1 describes six example sets of data that may be obtained from either the PROSCAL CD or from www.proscal.com. Summary output files for the six sets are also included. Two of the examples, **Prox/Lik** and **3DProx/Lik**, are based upon the same simulated set of judgments, the parameters of which are two-dimensional. By running the examples or looking at their output files, it is seen that the log likelihood of the first (two-dimensional) analysis, is -2,933.3 and the log likelihood of the second (three-dimensional) analysis is -2,894.5. While the three-dimensional example has the higher likelihood, the two-dimensional analysis results in a lower value of CAIC — 6,127 vs. 6,171. The two-dimensional solution is thus correctly chosen as the better representation.

Likelihood ratio tests can be used with isotropic analyses. The test is a bit artificial. It requires estimating an unconstrained r dimensional solution and comparing the final log likelihood to the final log likelihood of a solution where the coordinates of the last k dimensions are constrained to have the same value. If you were testing a two versus a one dimensional solution, the constrained solution would be one where the coordinates are forced to be upon a line in a two dimensional space. Coordinate constraints are discussed in Section VI.D.2.

Example: The dependent liking rating example, **DepLkRat**, in Appendix 1, describes a two dimensional analysis of liking ratings for twelve real objects about a single ideal object. The output file, `deplikrato.dat`, reports a final log likelihood value of -1,884. Recovery of the known configuration is very good, with an R^2 of actual and estimated expected distances being 0.97. To estimate a one dimensional solution, go to the Coordinate Constraint window (See Section VI.D.2) and set the 13 values in the second column to 14. Run the program again. The new final log likelihood is -1,928 with a R^2 of only 0.82. Compare two times the difference in the log

likelihoods, 88, to the test statistic $\chi^2_{\alpha=0.001, df=12} = 32.91$. The two dimensional model is seen to be superior to the one dimensional model.

D. Metric Properties

Comparisons of Euclidean and city-block metrics are more difficult since one metric is not a nested version of the other. This is an active area of research in the mathematical literature. At this point, we recommend using information criterion statistics for hypothesis testing. Further information on the comparison of different metrics with PROSCAL is available in MacKay (2001) and MacKay and Lilly (2004).

Probabilistic city-block models are not as well developed as probabilistic Euclidean models. Dependent sampling, described below in Section IV.F, is available for Euclidean metrics but not city-block metrics.

E. Common Space Assumption

There has been a lot of discussion as to whether subjects use the same cognitive space when making proximity judgments as they do when making hedonic (liking rating, preference ratio or binary choice) judgments. With PROSCAL, it is possible to test whether or not a common space underlies the different types of judgments. To do this, three separate analyses are required. The first is an analysis of the proximity judgments; the second is an analysis of the hedonic judgments; and the third is joint analysis of the proximity and hedonic judgments. Since the separate proximity and hedonic judgments are independent of one another, you can compare the log likelihood of the jointly scaled judgments with the sum of the log likelihoods of the separately scaled judgments. The sum of the log likelihoods will be higher but the corresponding number of free parameters will be fewer. The comparison may be made with a likelihood ratio test or with an information criterion statistic.

Example: The **Prox/Lik** example of Appendix 2 illustrates a common space analysis of proximity and liking rating data. Using a three coefficient measurement model (see Section IV.G below) the resulting log likelihood is -2,933.27 and the CAIC statistic is 6,126.88. Splitting the data set in two parts, it is possible to do an analysis of the proximities and an analysis of the liking ratings. With a three coefficient measurement model, the proximity data have a log likelihood of -234.35 and a CAIC statistic of 615.26. Similarly, when evaluating the liking ratings by themselves, the log likelihood is -2,666.19 and the CAIC statistic is 5,558.58. To make the comparison analogous to the common space analysis, a measurement model with zero coefficients is used for the liking ratings. (In the common space model, the measurement model changes the dissimilarities but not the liking ratings.) Summing the independent proximity and liking rating analyses, the sum of the log likelihoods is -2,900.54, the sum of the free parameters is 47 (20 for the proximities and 27 for the liking ratings) and the number of judgments is 2,160. The corresponding CAIC statistic is 6,208.94. (This CAIC value must be calculated by the user with the equation of Section IV.) Thus, while the summed analyses have a higher likelihood, the common space analysis has a better (lower) CAIC statistic and the common space analysis is supported. (Another example of this test is provided in MacKay, Easley and Zinnes (1995).)

F. Sampling Properties

When preference ratios, binary choices or liking ratings are used with a Euclidean metric, dependent or independent sampling may be used. With dependent sampling, it is assumed that a subject samples the ideal object only once for each judgment (preference ratios and binary choices) or once for each set of n real object judgments (liking ratings). With independent sampling, it is assumed that the subject samples the ideal object twice, once for the numerator and once for the denominator, when using preference ratios or binary choices. When using liking ratings, n samples of the ideal object are assumed. Dependent and independent sampling may be used in either isotropic or anisotropic spaces. Dependent sampling is not available for the city-block metric.

If a solution is sought using only a single ideal object and using real objects whose locations are not fixed (internal unfolding), then dependent sampling must be used. Choice probabilities are also conditional upon whether dependent or independent sampling is assumed.

Dependent and independent sampling models have the same number of free parameters. The model with the higher likelihood should be selected.

G. Measurement Models

Likelihood ratio and information criterion statistics can also be used to determine whether the improvement in fit occasioned by going to a one, two or three-coefficient model is worth the loss in degrees of freedom. Our experience indicates that you should be cautious about using three-coefficient measurement models with small data sets and with liking rating scales.

V PROSCAL OPTIONS

In the previous section, the ability to specify alternative probabilistic scaling models and perform hypothesis tests was discussed. A number of other options is also available. The primary options are discussed in this section. Other options will be discussed in Section VI, User's Guide.

A. Location, Variance and Measurement Targets

Theory or previous studies will often suggest a particular set of target values. These targets can be used to define the initial estimates for the numerical optimization algorithms. Targets can also be compared to the final estimates, even when the targets are not used as initial estimates. To compare target and estimate locations in a Euclidean space, a proximity transformation (Cliff, 1966) is carried out on the estimates to optimally align the target and estimate coordinates. City-block spaces do not permit the arbitrary rotation of a configuration; dilation transformations are thus used to compare target and city-block coordinate estimates.

The ability to specify location, variance and measurement targets is also useful in hypothesis testing. If, for example, you have obtained an isotropic solution for Euclidean liking

rating data and you want to test this solution against a Euclidean anisotropic solution, you should first find the isotropic solution and then use that solution's final estimates as the initial solution in an anisotropic analysis. The initial and final likelihoods of the anisotropic analysis should then be used for hypothesis testing. Direct comparisons of isotropic and anisotropic likelihoods of Euclidean preference ratio data, binary choices and proximities should also be avoided.

One reason for avoiding direct comparisons of Euclidean isotropic and anisotropic likelihoods is that Euclidean isotropic solutions are based upon the non-central chi-square and doubly non-central F distributions while the Euclidean anisotropic solutions are based upon the more general quadratic forms in normal variables distribution. Theoretically, isotropic solutions have identical PDF (probability density function) values with the general and more specific distributions but slight computational differences can arise. A much more significant reason for avoiding direct comparisons occurs when calculating the doubly non-central F distribution, which is used with isotropic preference ratios and binary choices. The left tail of the doubly non-central F distribution is difficult to estimate. To overcome this problem, left-tail estimates are changed to right tail estimates by inverting the judgments and reversing the corresponding parameters. This is not done with anisotropic solutions. Isotropic and anisotropic likelihoods are thus non-comparable in a Euclidean space. (Inversion affects the likelihoods but not the estimates.)

B. Proximities

The proximity of two real objects may be measured by dissimilarity judgments (high values mean the two real objects are very different) or by similarity judgments (low values mean the two real objects are very different). Dissimilarity judgments and similarity judgments both have a lower bound of zero. Dissimilarities are unbounded on the high end and similarities are traditionally bounded by one.

To adapt similarities to a spatial model, functions are needed to transform the similarities to dissimilarities. PROSCAL provides two transformations, a Gaussian transformation $s_{ij} = \exp(-\delta_{ij}^2)$ and an exponential transformation $s_{ij} = \exp(-\delta_{ij})$. Either transformation may be used with a city-block or a Euclidean metric. However, it is traditional to associate the Gaussian transformation with the Euclidean metric and the exponential transformation with the city-block metric.

Measurement model transformations (Section III) may be used with either dissimilarities or similarities. In either case, it is the dissimilarity δ_{ij} that is being transformed by the measurement model.

C. Standardization

When data from more than one subject are simultaneously evaluated, standardization becomes an issue. Ideally, the experimental conditions from which the data are obtained will be such that standardization is not necessary. If standardization is required, this can, in some cases, be done by PROSCAL or by the user before the data are entered.

A standardization option for dissimilarity judgments provided in PROSCAL proportionally transforms the input data so that the root mean square of the dissimilarity judgments for each data set is equal to 1.0. If the data are incomplete (the number of distance judgments per data set is less than $n(n - 1)/2$ where n is the number of objects), this standardization procedure assumes that the expected values of the missing judgments do not vary significantly among data sets. If one data set consists of judgments about similar objects and another data set consists of judgments about dissimilar objects, other standardization procedures should be used. (Since standardization changes the data, likelihood, CAIC and BIC statistics cannot be compared with solutions derived from data that are not standardized.)

D. Hedonics

Hedonic judgments are evaluated with an unfolding model (Coombs, 1964). Unfolding models view utility as an inverse function of the distance from a real object to an ideal object. While individual ideal distributions may be estimated for each subject, this is not recommended since in situations where the number of parameters increases with the number of observations, the maximum likelihood estimates need not be consistent. Instead of estimating individual ideal distributions, ideal distributions are usually estimated for segments of subjects.

Many different bases may be used for segmentation. Geography, socio-economic variables and purchase history are common bases in consumer studies. When the hedonic judgments themselves are to be the basis of the segmentation, a simple procedure is to assign subjects with the same first-choice to the same segment. PROSCAL can also compute mixture maximum likelihood estimates using an expectation-maximization algorithm which automatically assigns subjects to segments on the basis of their hedonic evaluations. Segments can be of unequal size. The User's Guide, Section VI.D.2, describes how subjects are assigned to segments and Appendix 3 describes the output of the mixture model.

Liking ratings, preference ratios and binary preferential choices are the hedonic data types currently admitted by PROSCAL. Examples of each data type are provided in the User's Guide, Section VI.D.1. Hedonic judgments may be evaluated by themselves or in combination with proximity data. Internal or external analyses (Section V.F below) may be requested.

E. Segmentation Analysis

When beginning the analysis of hedonic data, the user must specify the segment or group to which each set of hedonic judgments belongs. Common bases for this specification include *a priori* assignment, cluster analysis of the hedonic data and the most frequently chosen or highly evaluated product. When segments are a function of the hedonic judgments, then cluster analysis or the most frequently chosen product is usually used. *A priori* assignment is more frequent when segments are not a function of the hedonic judgments.

Improved solutions can often be obtained if the program is allowed to make its own assignment of judgments to clusters. PROSCAL provides two ways of doing this. The first way is to use an EM algorithm to simultaneously rescale the data and estimate the probabilities of each

subject belonging to each of the segments. The second way is to take the probabilistic segment assignments of the EM analysis, deterministically assign subjects to their highest probability segments, and start the scaling over again from the beginning. With large data sets there will usually be no difference between the first and second methods but with small data sets it is usually worth doing the second analysis.

F. Internal and External Analysis

When you are evaluating hedonic data, the ML estimates of the coordinates (and variances if desired) of the real objects may be estimated from the data at hand (internal analysis) or they may be fixed to equal the values of the initial estimates (external analysis). The initial values may be read in as target values or estimated.

G. Density Function Approximations

PDFs for Euclidean anisotropic spaces are based upon the quadratic forms in normal variables distribution. There is no simple closed form expression for this density function. A time consuming procedure based upon numerical integration (Imhof, 1961) has been used by previous versions of PROSCAL. In the current version of PROSCAL, Imhof's method has been replaced by an adaptation of Pearson's (1959) approximation for binary choices and preference ratios and Jensen & Solomon's (1972) approximation for proximity and liking rating data. The use of approximations for city-block spaces is described in MacKay (2001).

H. Optimization Algorithms

Previous versions of PROSCAL allowed a variety of numerical optimization algorithms. Our experience has consistently favored the results obtained by the direct search method known as STEPIT (Chandler, 1969). As a result, PROSCAL now uses only STEPIT.

VI. USER'S GUIDE

A. Setup

PROSCAL is designed to work with 32 bit versions of S-PLUS. Supported versions include S-PLUS 2000, S-PLUS 6._, S-PLUS 7._ and S-PLUS 8._. In the following section, S-PLUS 8.2 will be used to illustrate how you should set up PROSCAL once S-PLUS is installed.

There are two methods for obtaining the PROSCAL files:

- From the PROSCAL web site located at www.proscal.com
 - Select **Downloads** from the PROSCAL home page.
 - Select **Proscal for S-PLUS**. This will download the file splus.zip, which should be placed in your project folder. If you are uncertain as to the name of the current project folder, click File►Open on the menu bar of S-PLUS. The name of your current

project folder appears in the “Open” Dialog Box opposite “Look in.” To see the entire path, click on the down wedge adjacent to the “Look in” option.

- Unzip the file `splus.zip`. This will generate four files: `Proscal`, `Pros51f.exe`, `Lf90.eer`, and `Readme.txt`. Be sure to save these files in the project folder.
- From the PROSCAL CD
 - Copy the files `Proscal`, `Pros51.exe`, `Lf90.eer` and `Readme.txt`, located in the Programs folder. Paste them into the project folder. Create an `S-PLUS.exe` shortcut, paste it in the project folder, modify its properties by adding the `S_PROJ` variable to the Target along with the project folder path. Then, specify the project folder path as the place in which to start the analysis.
- Open S-PLUS. The `SPLUS.exe` file may be found under `C:\Program Files (x86)\company\splusxx\cmd` where `company` is either `Insightful` or `TIBCO` and `xx` is the version of S-PLUS.
- Open the S-PLUS Commands Window. You can do this by clicking on the menu bar `Window` ► `Commands Window`.
- Enter

```
> source("PROSCAL")
> PSInstall()
```

Note that each line in the Commands Window begins with “>”. Be sure to type a lower case “s” in `source` and upper case letters “PSI” in `PSInstall`. Also note that `PSInstall` ends with both a left and right hand parenthesis.

PROSCAL is now ready. You should see `Proscal` in your menu bar. (If you are updating your version of PROSCAL, you will need to execute the `source` and `PSInstall` commands with the new `Proscal` file.)

If you are familiar with S-PLUS, you may want to create a Functions folder and use the Advanced Tab to set Classes equal to “function.” This will make all of the S-PLUS code available to you in case you should wish to customize the interface.

Caution. *When you are finished with the current S-PLUS session, be sure to save all files selected by S-PLUS.*

B. Data Sets

The following dialog boxes are illustrated using sample data sets. These sample data sets can be obtained from the web site by selecting **Data for S-PLUS Version**, located on the **Downloads** page. This will download the file `splusdata.zip`, which should be placed in your working S-PLUS folder. When it is unzipped, data files for various examples will be extracted. The sample data sets can also be obtained from the PROSCAL CD. They are located in the **Data** folder. You may want to place the data for each example in its own project folder.

Prox/Lik Example. This example illustrates the simultaneous analysis of proximity and liking rating data. The data files used for this example are: `proxlikd.dat`, `proxlikc.dat`, and `proxlikv.dat`. The file `proxlikd.dat` contains simulated dissimilarity judgments followed by liking ratings. The data are based on 8 stimuli (real objects) and 4 ideal objects. The file `proxlikc.dat` contains the coordinates that were used to generate these data. The file `proxlikv.dat` contains the variances of both the real and the ideal objects. All these variances are located in a single column. File `proxliko.dat`, containing the summary output, is annotated in Appendix 2. This example is discussed in more detail in the following sections and in Appendix 1.

Binary Choice Example. This example illustrates the unfolding of binary choice data. The data files used for this example are: `binaryd.dat`, `binaryc.dat`, and `binaryv.dat`. The file `binaryd.dat` contains simulated binary preferential choices. The data are based on 15 real objects and 6 ideal objects. The file `binaryc.dat` contains the coordinates that were used to generate these data. The file `binaryv.dat` contains the variances of both the real and the ideal objects. All these variances are located in a single column. File `binaryo.dat` contains the summary output. This example is discussed in more detail in Appendix 1.

Preference Ratio Example. This example illustrates the unfolding of preference ratio judgments. The data files used for this example are: `prefd.dat`, `prefc.dat`, and `prefv.dat`. The file `prefd.dat` contains simulated preference ratio judgments. The data are based on 12 real objects and 4 ideal objects. The file `prefc.dat` contains the coordinates that were used to generate these data. The file `prefv.dat` contains the variances of both the real and the ideal objects. All these variances are located in a single column. File `prefo.dat` contains the summary output. This example is discussed in more detail in Appendix 1.

Binary Choice and Proximity Data / Segmentation Example. This example illustrates the use of a mixture model to segment consumers. Following the mixture model analysis, the data are then reanalyzed with the new segment assignments. The data files used for this example are: `binsimsegd.dat`, `binsimsegc.dat`, and `binsimsegv.dat`. The file `binsimsegd.dat` contains simulated dissimilarity judgments and binary choices. The data are based on 8 real objects and 3 ideal objects. The file `binsimsegc.dat` contains the coordinates that were used to generate these data. The file `binsimsegv.dat` contains the variances of both the real and the ideal objects. All these variances are located in a single column. The second half of the data set is intentionally misclassified to illustrate the abilities of PROSCAL's segmentation algorithms. File `binsimsego.dat` contains the summary output. This example is discussed in more detail in Appendix 1.

3DProx/Lik Example. This example illustrates a three dimensional analysis of the **Prox/Lik** data previously analyzed in two dimensions. The data files used for this example are: proxlikd.dat, 3DProxLick.dat, 3DProxLikv.dat, and 3DProxLikm.dat. The last three files are based on files PSOut9.ASC, PSOut10.ASC and PSOut11.ASC saved at the end of the **Prox/Lik** analysis. File proxlikd.dat contains simulated dissimilarity judgments followed by liking ratings. File 3DProxLick.dat contains the final coordinate estimates of **Prox/Lik** with an extra column of zeros added for the third dimension. File 3DProxLikv.dat contains the variances estimated by **Prox/Lik**. Variances are located in a single column. Every third line contains a value of zero which has been added for the new dimension. The file 3DProxLikm.dat contains the measurement parameters estimated in **Prox/Lik**. File 3DProxLiko.dat contains the summary output. This example is discussed in more detail in Appendix 1.

Dependent Liking Ratings Example. This example illustrates internal unfolding about a single ideal object. The data files used for this example are: deplikatrd.dat, deplikratc.dat, and deplikratv.dat. The file deplikatrd.dat contains liking ratings that have been simulated using dependent sampling and then transformed to a 1 to 9 scale, where 1 indicates the most liked object. The data are based on 12 real objects and 1 ideal object. The file deplikratc.dat contains the coordinates that were used to generate these data. The file deplikratv.dat contains the variances of both the real and the ideal objects. All these variances are located in a single column. File deplikrato.dat contains the summary output. This example is discussed in more detail in Appendix 1.

C. PROSCAL Menu

To start, left-click on the Proscal menu item. You are presented with eleven choices. When opening a S-PLUS session, all choices are available. Data from the previous analysis will be used in responding to the choices. Once a design has been specified, some choices may not be enabled. For example, Perceptual Share will not be enabled unless Run Proscal has been previously selected and executed.

- Generate Design: Here you specify the type of type of analysis, data files, and analytical options you wish to pursue. Based on your responses, this dialog box generates the files that are read in by PROSCAL.
- Run PROSCAL Executes the estimation phase of PROSCAL. Progress reports of each step of the estimation procedure are displayed in an output window.
- Examine Output Opens the Window's Notebook and reads in the detailed output file of PROSCAL.
- 2D Plot - Centroids Plots the locations of either the target, estimate or transformed objects in a two dimensional space.

- **2D Plot – Std. Dev.** Plots ellipses representing the standard deviation of either the target, estimate or transformed objects in a two dimensional space.
- **3D Plot – Centroids** Plots the locations of either the target, estimate or transformed objects in a three dimensional space. The space may be rotated by the user.
- **3D Plot – Std. Dev.** Plots ellipsoids representing the standard deviations of either the target, estimate or transformed objects in a three dimensional space. The space may be rotated by the user.
- **Change Labels** Here you can change the labels and colors of the real and ideal objects.
- **Perceptual Share** Uses the analysis of the liking rating, preference ratio or binary choices to estimate the percentage of times the real objects would be selected as first choices.
- **What-if Analysis** Estimates the perceptual share if changes were made to an object's location or variance parameters of if new objects were added.
- **About PROSCAL** Displays the version number and date of PROSCAL

D. Generate Design

Choose **Proscal ►Generate Design**. The following dialog box appears. Default values, which may be overridden, are given for a basic analysis of proximities data. Note that not all options are enabled all the time.

*The values shown in the immediately following dialog boxes are for the **Prox/Lik** example. The data for this example can be obtained from the web site by going to the Downloads page or from the PROSCAL CD by going to the **Data** folder.*

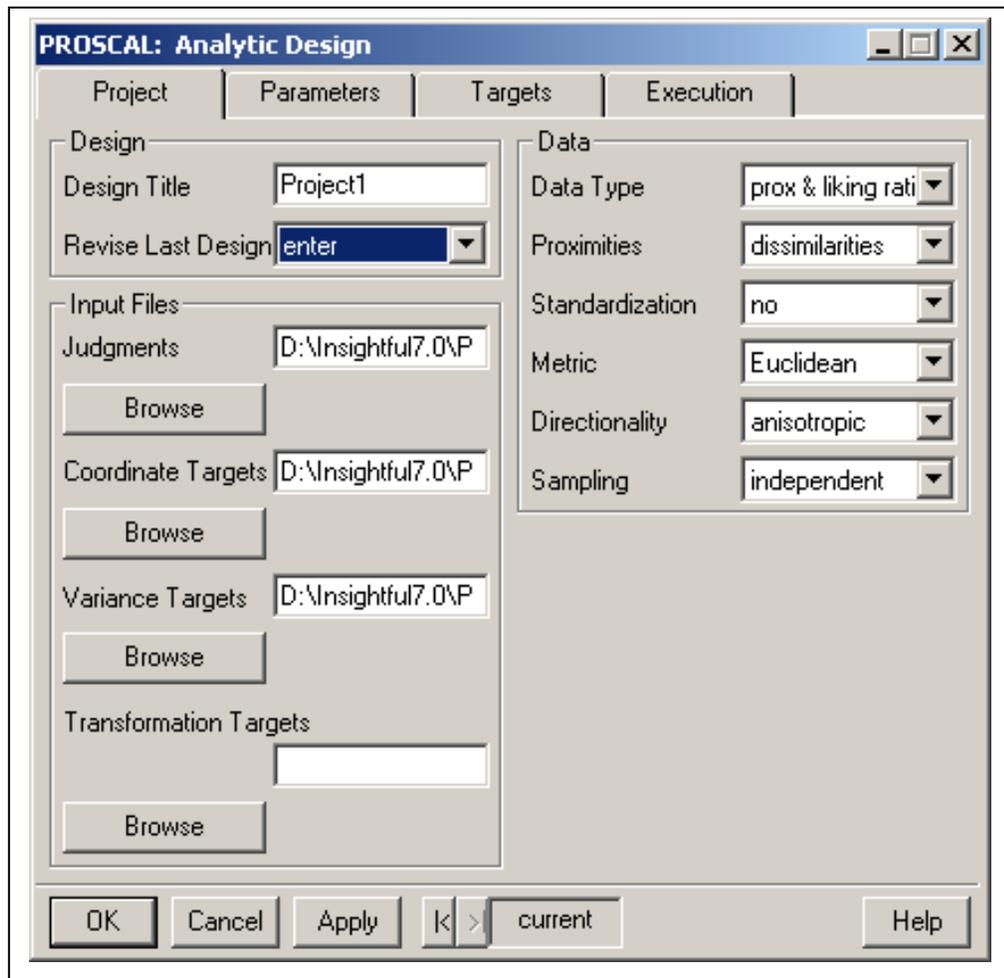
The Generate Design dialog box contains four tabs. They are labeled Project, Parameters, Targets and Execution. Each tab has its own dialog box. When you are through filling in these four dialog boxes, select **OK** or **Apply**. (**OK** closes the dialog, **Apply** does not.) A Report window will be generated with the message

[1] "Input file generated."

The Report window can be closed at this time.

Whenever you change the data or design parameters, you should fill in the Generate Design dialog box . Note that changing some parameters, such as the number of stimuli, will require that other parameters, such as the coordinate constraints, labels and colors be reset as well. Label and color information, which are stored in S-PLUS files PROSCAL.Labels and PROSCAL.Colors, may be edited directly from S-PLUS. To reset coordinate constraints, click on the coordinate constraint check box in the Parameters tab and modify the data frame that opens up.

D.1 The Project Tab



Design Title

You may enter any title that you wish. Usually it is desirable to use a title that describes the current project or the data being analyzed. The title will be reproduced in the PROSCAL output.

Revise Last Design

Select one from {yes, default, enter}

yes – restores the values used in the last analysis. This option assumes that the program is running in the same folder as the last analysis. Values from the last analysis are saved at *the end* of Run PROSCAL.

default – restores the values which were present when the Generate Design Window was first opened.

enter – indicates that you are entering new values.

Input Files

You may specify one, two, three or four files. If you give just the file name and not the full path, then it is assumed that the file is in your current working S-PLUS folder. Hit the enter key ↵ at the end of each entry.

In all the examples that follow, the letter **d** at the end of the name indicates that the file contains judgment data; the letter **c** indicates that the file contains coordinates; the letter **v** indicates that the file contains variances, and the letter **m** indicates that the file contains measurement transformation coefficients.

S-PLUS reads input files as “data frame” objects. It is assumed that each row of an ASCII input file is terminated with carriage return or a line feed.

Judgments

Enter the name of the input file where the judgment data reside. As described below under **Data Type**, judgments may include proximity data and hedonic data. The input file can also contain selected combinations of these data types.

Proximity Data

Proximity data are usually entered in the form of a lower-half matrix without a diagonal. All proximity judgments are positive. Negative entries indicate missing values. Proximities may be obtained directly from subjects or they may be derived from object profile data. As indicated in the next box, proximity judgments may take the form of dissimilarity or similarity judgments.

The following proximity data are dissimilarity judgments for the first subject in the proxlikd.dat data set. The proximity data are located at the beginning of this data set. The first value, 1.2542 is the dissimilarity for real objects 2 and 1. The value 1.6589 is for real objects 3 and 1. The greatest degree of similarity (least dissimilarity) is expressed for real objects 8 and 3. Note that the diagonal terms are missing, so that the first row corresponds to the second real object.

```

1.2542
1.6589 .7747
.8389 1.3700 1.0005
1.3991 1.6764 1.3997 1.0354
.8414 1.0816 .9523 .6273 .6632
1.4546 1.0631 .4944 .7789 1.0858 .4191
1.0527 1.2632 .1825 .8596 1.0584 1.5261 .6855

```

Preference Ratio Data

Preference ratio data are usually entered in the form of a lower-half matrix without a diagonal. All preference ratio judgments are positive. Negative entries indicate missing values. Preference ratio judgments greater than 1.0 indicate that the column object is preferred to the row object. Values between 0.0 and 1.0 indicate that the row object is preferred to the column object. Alternatively, you can think of an entry in the j^{th} row and k^{th} column as a ratio of disutilities for subject i ; i.e. $r_{ijk} = d_{ij}/d_{ik}$. Preference ratio data are often obtained from subjects using graphic rating scales.

The following preference ratio data are for the first subject in the prefd.dat data set. The first value of the fourth line indicates that the subject prefers the first real object almost twice as much as the fifth real object. (Alternatively, the disutility for object five is greater than the disutility for object one.) Note that the diagonal terms are missing, so that the first row corresponds to the second real object.

```

0.9775
0.7423 0.3426
1.1995 2.0228 5.3450
1.9923 1.1315 1.9936 1.0777
1.3864 1.1637 3.1746 0.7413 0.8745
0.8917 1.5122 3.1136 0.8046 0.8168 0.6253
1.6467 2.2642 3.4444 1.6325 1.1979 0.9487 1.7357
1.5765 2.3010 3.5762 1.3801 1.2176 1.5423 1.8687 1.0768
2.1924 1.3027 2.3300 1.0057 1.0362 1.4094 1.0424 0.7684 0.7299
0.5743 0.5010 0.7831 0.3687 0.4910 0.4340 0.7315 0.2915 0.3638 0.4596
0.4109 0.5256 1.0111 0.4166 0.4757 0.4732 0.3786 0.5554 0.3204 0.5829 0.7874

```

Liking Rating Data

Liking ratings are usually entered as a rectangular matrix with n columns, each row indicating a subject's liking for the n objects. High liking ratings indicate objects that are disliked, low values indicate objects that are liked. Alternatively, you can think of a liking rating as a judgment about the distance of a real object from an ideal object. Liking rating data are often obtained using integer scales.

The following liking rating data are for the first four subjects in the proxlikd.dat data set. The data indicate that the first subject has stated that the seventh real object is liked least and the second and fifth objects are liked most.

```
2 1 3 4 1 2 5 2
2 1 4 3 3 1 3 2
8 3 3 4 4 3 5 3
2 4 4 2 1 1 2 2
```

Binary Preferential Choice Data

Binary choice data are usually entered in the form of a lower-half matrix without a diagonal. All binary choice values must be zero or one. Negative entries indicate missing values. A binary choice value of 1 indicates that the row object is selected over the column object. (Binary choice data are thus an exception to the general PROSCAL rule that large values indicate greater distance or disutility.) Binary choice data are often obtained in choice experiments.

The following binary choice data are for the first subject in the binaryd.dat data set. The first value of the third line indicates that the subject chooses the fourth real object over the first real object. Note that the diagonal terms are missing, so that the first row corresponds to the second real object.

```
0
0 0
1 0 0
1 1 1 1
0 0 1 0 1
0 0 1 0 0 1
1 1 1 0 1 1 0
```

Multiple Data Sets

There are four possible combinations of data types:

- Proximity and Preference Ratio judgments
- Proximity and Liking Rating judgments
- Proximity judgments and Binary Choices
- Binary Choices and Liking Rating judgments

For the first three data combinations, where proximity data are present, the proximity data should occur first in the data file. For the last case, binary choices should precede the liking ratings.

Proximities

Proximity judgments may be either dissimilarities or similarities. For dissimilarity judgments, large values indicate pairs of objects that are dissimilar, small values pairs that are similar. For similarity judgments, large values indicate pairs of objects that are similar, small values indicate pairs that are dissimilar.

Similarities may be of a Gaussian or exponential form. A Gaussian form indicates that the expected relationship of a dissimilarity δ to a similarity is of the form $s = \exp(-\delta^2)$ while an exponential form indicates that the expected relationship of a dissimilarity to a similarity is of the form $s = \exp(-\delta)$. A Gaussian form is often specified when a Euclidean metric is being used and an exponential form is usually specified when a city-block metric is being used. When similarities are used, the values must range from 0 (low similarity) to 1 (high similarity). If the data are not within this range they are automatically rescaled and a warning message is issued.

Coordinate Targets

Coordinate targets are optional. If present, the coordinates for each object should be entered on one line. There should be as many items on a line as there are dimensions in the space. Each item should be separated by one or more spaces.

For an example of a coordinate file, see proxlikc.dat.

Variance Targets

Variance targets are optional. If present, variances should be entered one per line. An isotropic Case V model would consist of a single line. For an anisotropic Case III model with n objects in r dimensions there should be nr lines. The first r lines are for the first object.

For an example of a variance file, see proxlikv.dat.

Measurement Targets	Measurement targets are optional. If present, all three measurement coefficients should be entered, one per line. Measurement coefficients are described in the following section. Default values for the additive, scale and exponential coefficients are 0, 1 and 1. When, for example, an exponential transformation is used and the target value is not 1, the default values for the additive and scale transformation coefficients should be entered in the target file on the two lines before the exponential constant..
Browse	If you cannot find a file, select the Browse button to open the Windows Explorer.
Data Type	Select one from <ul style="list-style-type: none"> • proximities • preference ratios • proximities and preference ratios • liking ratings • proximities and liking ratings • binary choices • proximities and binary choices • binary choices and liking ratings
Standardization	Select one from {yes, no} to indicate whether the dissimilarity judgments should be standardized. (Only dissimilarities may be standardized.)
Metric	Select one from {Euclidean, city-block} to indicate the metric used for scaling the judgments.
Directionality	Select one from {isotropic, anisotropic} to indicate whether variances are the same or different from one dimension to another.
Sampling	Select one from {independent, dependent} to indicate which method of sampling is to be assumed when preference ratio judgments, binary choices or liking ratings are used. Independent sampling is often referred to as two-stage sampling and dependent sampling as one-stage sampling. Dependent sampling is only available for Euclidean spaces. Independent sampling with a single ideal point will produce meaningless results (MacKay, 2007).

D.2 The Parameters Tab

- Stimuli** Enter the number of real objects being evaluated.
- Ideals** Enter the number of ideal objects being evaluated.
- Prox. Sets** Enter the number of proximity data sets.
- Binary Choice Sets** Enter the number of binary choice data sets. (In the current example, there are no binary choice sets. The option is not enabled.)
- Pref./Liking Sets** Enter the number of sets of preference ratio or liking rating judgments.

*For the **Prox/Lik** example, there are four sets of 50 subjects, with each set having its own ideal object to be estimated. Thus, the number of Pref./Liking Sets is 200.*

Dimensions Enter the number of spatial dimensions.

Measurement Parameters The relationship of the dissimilarities δ_{ij} to the underlying or latent distances d_{ij} is modeled by the function $\delta_{ij} = a + bd_{ij}^c$ where a , b and c are additive, scale and exponent coefficients estimated by the program. Default values of a , b and c are 0, 1, and 1 respectively. Any combination of these

parameters may be selected. As mentioned earlier, this linear exponential transformation may also be used for similarities and liking ratings. An exponential transformation may be used with preference ratios. No transformations are used with binary choices.

*For the **proxlikd.dat** data set, which is a combination of dissimilarity and liking rating judgments, the function $\delta_{ij} = a + bd_{ij}^c$ was selected.*

Constraints

There are three constraint options: variances, ideals, and coordinates. The constraints become enabled after the number of stimuli and the number of ideals have been specified.

Constraint boxes must be checked. When a constraint box is checked, a data frame opens up with default values already filled in. These default values may be overridden by entering other values. Unchecking a box and rechecking the box again will reset the constraints to their default values. Whenever the number of stimuli, ideals or dimensions is changed, the constraints should be reset.

Constraint values are defined as a set of integers starting with the number 1. These values must be sequential in the sense that each value cannot be more than one unit greater than a previously entered value. Thus, values 1,2,3,1,2,3,... are allowed but 1,2,4,2,3,... are not. No missing values are allowed.

Constraints on Variances

Variance constraints are defined by object, not by object and dimension, as is done with **Constraints on Coordinates** (see below).

If a model is defined for n real objects and m ideal objects, a data frame with $n + m$ rows labeled O1 ... On, I1 ... Im will open with the first column filled with 1s, indicating that a Case V model in which all variances are the same is being estimated.

Suppose that you wish to estimate a Case III model in which the odd numbered real objects had one variance, the even numbered real objects had a second variance and the ideal objects had a third variance. Then, for 8 real objects and 4 ideal objects, you should fill in the first column with {1, 2, 1, 2, 1, 2, 1, 2, 3, 3, 3, 3}.

*For the **Prox/Lik** example, a Case III model was used. Variances were estimated separately for the first four real objects, the second four real objects and the four ideal objects. The presence of the proximity judgments in this example makes it possible to estimate unique variances for both the real and ideal objects. The variance constraint data frame is completed as follows.*

1		
VarSets		
1	O1	1
2	O2	1
3	O3	1
4	O4	1
5	O5	2
6	O6	2
7	O7	2
8	O8	2
9	I1	3
10	I2	3
11	I3	3
12	I4	3

Constraints on Ideals

Constraints on ideals are used to indicate which data sets are associated with which ideal objects. The data frame contains as many rows as there are hedonic data sets.

*In the **Prox/Lik** example, there are 4 ideal objects and 200 sets of liking ratings. The first 4 data sets are associated with the first 4 ideal objects. This pattern is repeated for the remaining 196 data sets. Therefore, you should fill in {1, 2, 3, 4, 1, 2, 3, 4, 1, ..., 4}. For the first 14 data sets, the ideal constraint data frame is completed as follows.*

	1	2
	IdealSets	
1		1
2		2
3		3
4		4
5		1
6		2
7		3
8		4
9		1
10		2
11		3
12		4
13		1
14		2

To make the task easier, you may fill in only the first 4 values and then specify that the **No. of Ideals per Pattern** is 4. The remaining values will then be generated by the program.

Constraints on Coordinates

Constraints on coordinates are used to indicate which coordinates are to have the same values. The opening data frame assigns a unique index number to each coordinate.

If constraints on coordinates are specified, then an **internal** model should be selected (Section D.3 below).

*For the **Prox/Lik** example, there are 12 objects in two dimensions. Accordingly, the index numbers 1 to 24 are assigned to the coordinates, with the numbers 1 to 12 for the first dimension and the numbers 13 to 24 for the second dimension. No coordinate constraints should be imposed for this data set.*

If, though, the first four real objects and the second four real objects were constrained to have the same coordinates, then the data frame should have the following values.

		1	2
		Dim.1	Dim.2
1	O1	1	9
2	O2	2	10
3	O3	3	11
4	O4	4	12
5	O5	1	9
6	O6	2	10
7	O7	3	11
8	O8	4	12
9	I1	5	13
10	I2	6	14
11	I3	7	15
12	I4	8	16

D.3 The Targets Tab

PROSCAL: Analytic Design

Project | Parameters | **Targets** | Execution

Targets: Initial Estimates:

Models: Model Selection:

OK Cancel Apply Help

Initial Estimates Select one from {estimated, target} to indicate whether the starting point for the ML analysis should be estimated by PROSCAL or provided by the target coordinates and variances. For target values to be initial values, the target files must first be specified in the Analytic Design Project Tab. Target files for coordinates and variances must both be specified but a target file for the measurement coefficients need not be specified. When a measurement coefficient file is not specified, the default values for the additive, scalar and exponential coefficients, 0, 1 and 1, are used.

Initial estimates have many uses. A common use is in the testing of dimensionality. Suppose, for example, that a two dimensional analysis has been conducted. To compare the results to those of a three dimensional analysis, start the three dimensional analysis with the results of the two dimensional analysis, entering values of 0.0 for the coordinates and variances of the third dimension. The initial likelihood value will be almost the same as the ending likelihood value for the two dimensional analysis. The reason for the discrepancy is that PROSCAL will replace the zero valued variance targets with very small positive values.

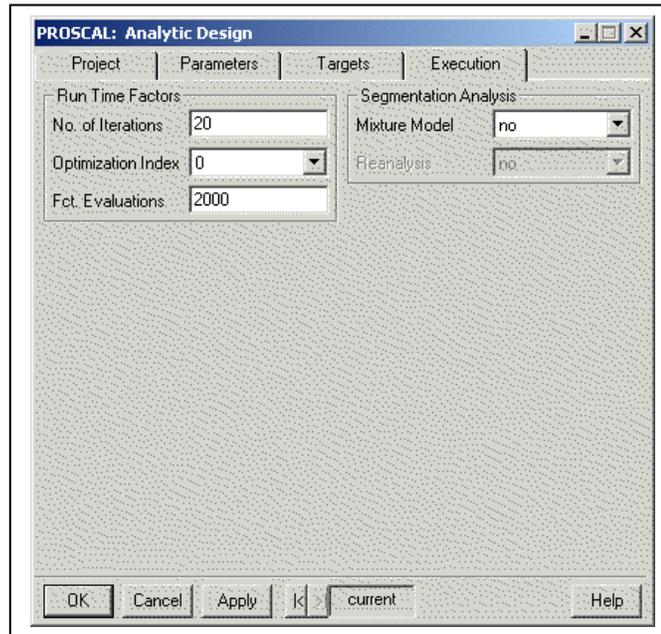
At the end of an analysis, text files PSOut9.ASC, PSOut10.ASC and PSOut11.ASC will contain the final estimates of the coordinates, variances and measurement parameters. Files PSOut9.ASC and PSOut10.ASC can be edited to include the values for the new dimension and then be read in along with PSOut11.ASC as new targets.

Model Selection Select one from {internal, external-C, external-CV} to indicate whether an internal or external analysis should be used. An internal analysis is one in which all maximum likelihood estimates are derived from the hedonic (and proximity) judgments. An external analysis is one in which the coordinates (external-C) or the coordinates and the variances (external-CV) of the real objects are fixed by the initial estimates.

Model Selection and Initial Estimates are independent of each other. Thus, if the Initial Estimates = estimated and Model Selection = external-C then it will be the initial estimated coordinates that remain fixed in the subsequent maximum likelihood analysis.

If an external analysis is used, then no constraints should be specified for the coordinates (Section D.2 above).

D.4 The Execution Tab



No. of Iterations Enter a positive integer. This value specifies the number of iterations that is to be used during the calculation phase. The default value of 20 is sufficient for most situations involving proximity data. PROSCAL will stop earlier if no progress is being made. If the output indicates that the log likelihood has been steadily increasing throughout all 20 iterations, then a higher value should be tried.

For other types of data, PROSCAL may choose other default values. If the selected number of iterations is insufficient, a “Maximum Number of Iterations Reached” message is printed.

If you specify a non-default value, PROSCAL will not stop earlier even if no progress is being made.

Optimization Index This index number affects the number of function evaluations allowed at each stage of each iteration. Select a number from $\{-2, -1, 0, 1, 2\}$. Negative numbers will speed up the optimization procedure, but will do so at the expense of less precise values. Positive numbers will usually slow down the optimization procedure and produce more precise values.

Fct. Evaluations The *maximum* number of function evaluations allowed on each stage of each iteration. The default values estimated by PROSCAL are usually sufficient but may be overridden by the user.

Mixture Model Select one from {yes, no} to indicate whether an EM analysis will follow the initial analysis. See Appendix 3 for more information.

Reanalysis Select one from {yes, no} to indicate whether a deterministic reassignment of subjects to segments will follow the EM analysis. The reassignment is based upon the EM output. Initial coordinate and variance estimates are based upon the original data and will differ from the estimates at the end of the EM analysis. See Appendix 3 for more information.

If, in the EM phase, the number of segments is diminished, then the Reanalysis should be omitted. The EM output can be used as the basis of a new analysis with a reduced number of segments.

Care must be taken when looking at the transformed estimates and recovery measures. If the order of segments after the EM analysis differs from the order of segments in the target configuration, then the transformed values and recovery measures will not be meaningful.

E. Run PROSCAL

Choose **Proscal** ► **Run PROSCAL**. A window will open and the estimation phase of PROSCAL will begin. Log likelihood values are displayed in the window to show the progress of the optimization. Values will generally increase in magnitude from one iteration to the next. The occasional appearance of lower values simply indicates that PROSCAL is trying a new search in a different direction.

At the end of the optimization process, a message will appear indicating that the window should be closed. When this occurs, all of the commands to PROSCAL are saved in a data frame called **PROSCAL.Commands**. This data frame is read in when you indicate that you wish to revise the last analysis. This option occurs in the Generate Design dialog box.

PROSCAL's execution time can vary from seconds to hours, depending upon the complexity of the design, options selected, the amount of data, and the configuration of your computer. Designs that request a segmentation analysis can be very time consuming, especially when there are a large number of segments. In most cases, though, the EM analysis takes no more than three cycles to finish.

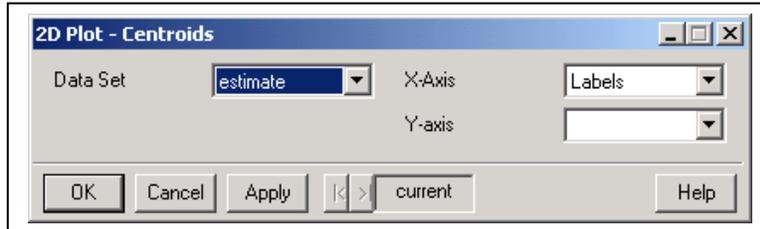
F. Examine Output

Choose **Proscal** ► **Examine Output**. Summary output from the PROSCAL analysis appears in the Notepad Window. For an annotated example of the summary output, see Appendix 2.

G. 2D Plot - Centroids

Choose **Proscal** ► **2D Plot - Centroids**. The following dialog box appears.

When through with the following steps, select **OK** or **Apply**.



Data Set

Select one of {estimate, target, transform}.

The **estimate** option plots the ML configuration estimated by PROSCAL.

The **target** option plots the target coordinates.

The **transform** option plots a transformation of the estimated configuration. The transformation aligns the estimated centroids with the target centroids. In a Euclidean space, the possible transformations of the estimated configuration include rotation, reflection, translation and dilation. In a city-block space, rotation and reflection are not allowed. City-block output should be carefully checked to determine the correct matching of transformed and target dimensions.

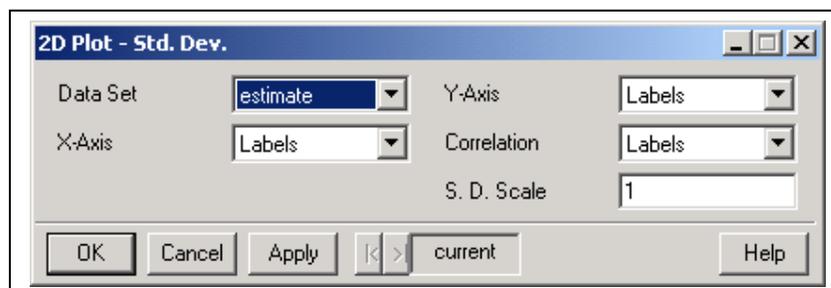
X-Axis Y-Axis

These two options contain a list of the column names as specified in the data frame you provided. Select whichever two columns you want to plot. Default values for column names are V1, V2, ..., V6.

H. 2D Plot - Standard Deviations

Choose **Proscal** ► **2D Plot - Std. Dev.**. The following dialog box appears.

When through with the following steps, select **OK** or **Apply**.



Data Set

Select one of {estimate, target, transform}.

The **estimate** option plots the centroids and standard deviations of the ML configuration estimated by PROSCAL. (Note that PROSCAL estimates variances, but plots standard deviations.)

The **target** option plots the centroids and standard deviations of the target objects.

The **transform** option plots the centroids and standard deviations of the transformed objects. The transformation aligns the estimated centroids with the target centroids. In a Euclidean space, the possible transformations of the standard deviations include rotation, reflection, translation and dilation. In a city-block space, rotation and reflection are not allowed.

Standard deviations are plotted as ellipses in a Euclidean space and as diamonds in a city-block space.

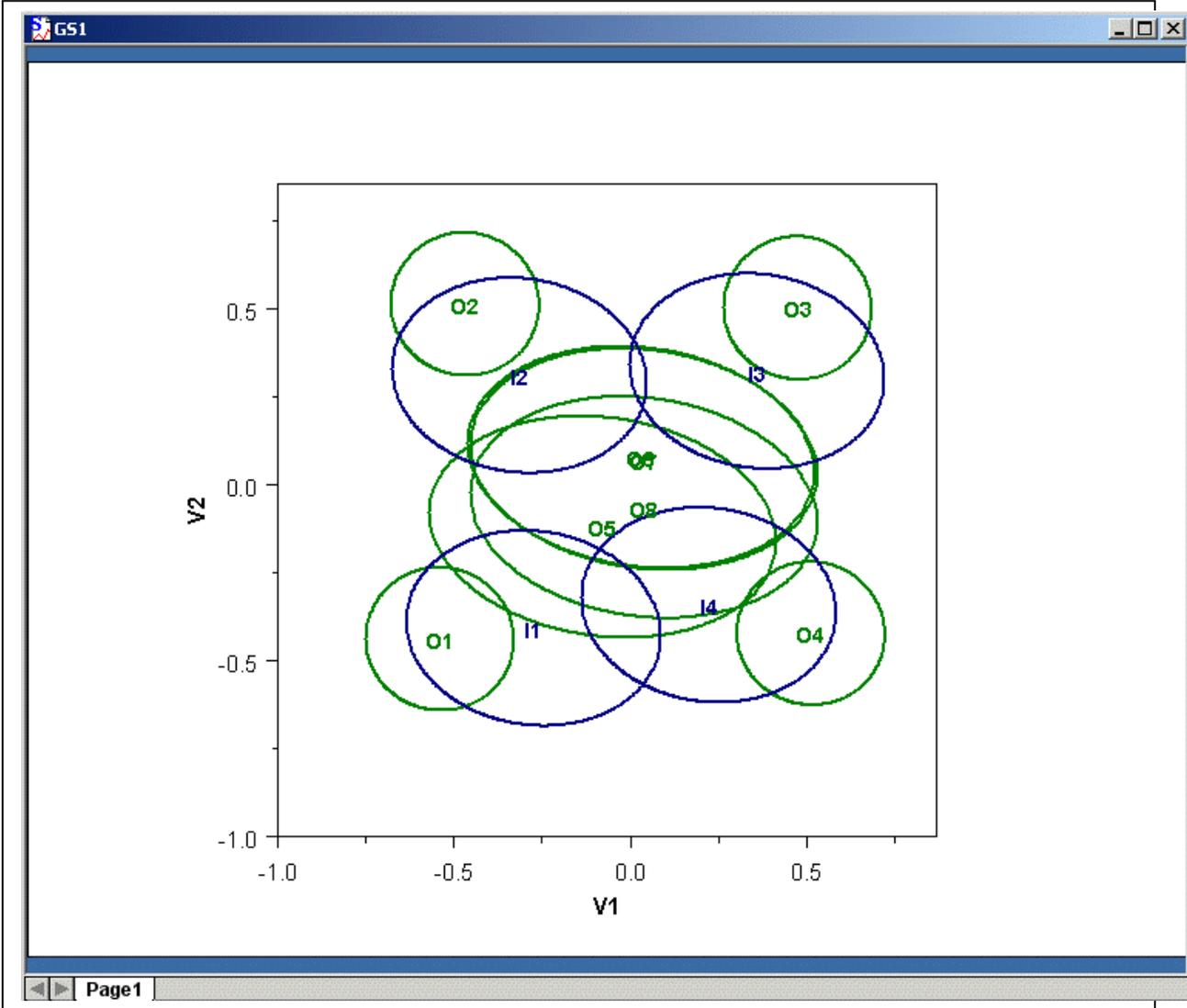
**X-Axis
Y-Axis**

These two options contain a list of the column names as specified in the data frame you provided. Select whichever two columns you want to plot. Default values for column names are V1, V2, ..., V6.

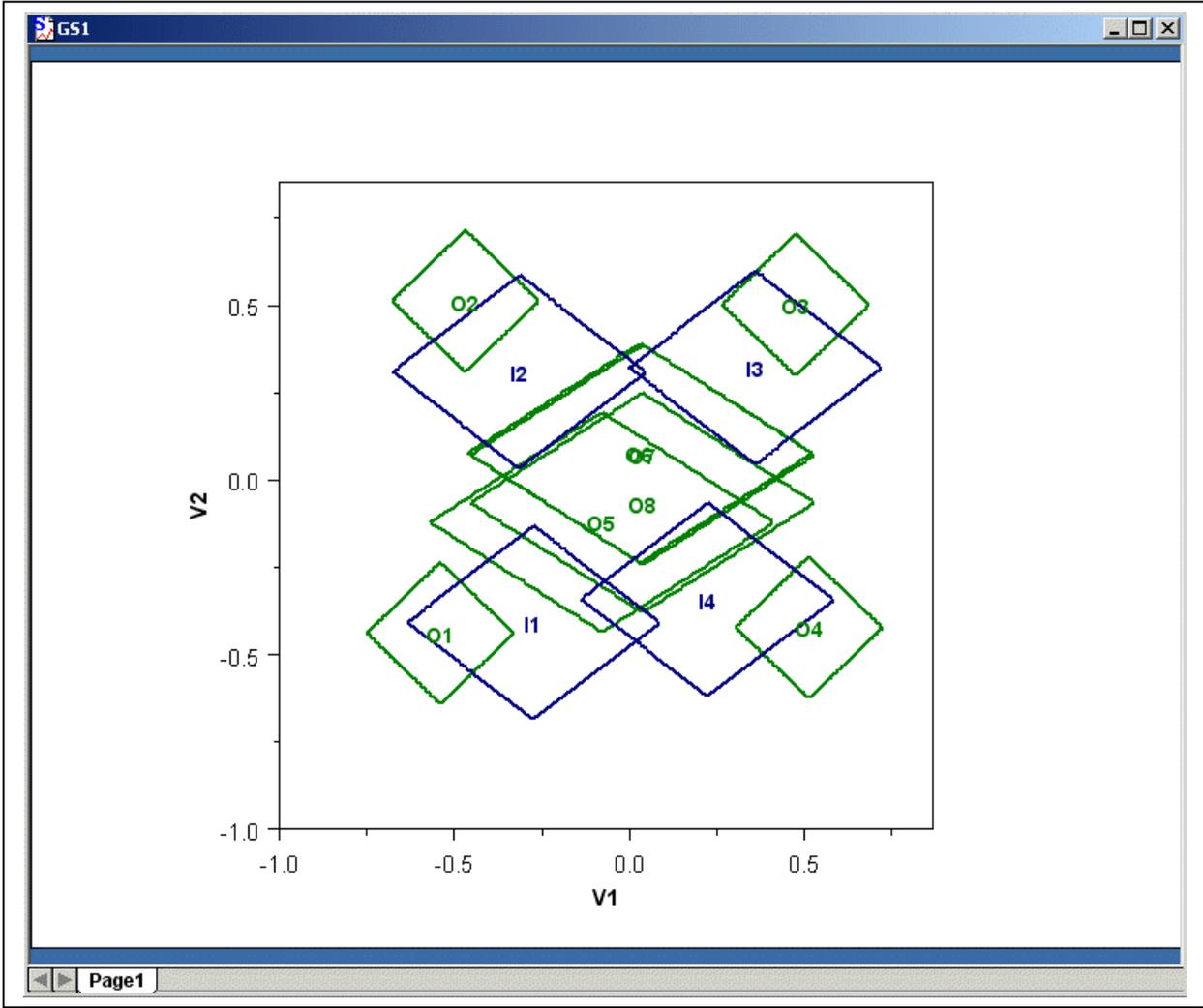
S. D. Scale

Select a positive real number to indicate the magnitude of the standard deviations which are to be plotted as ellipses. The default value of 1 means that an ellipse for a single standard deviation will be plotted.

Prox/Lik Example For this data set, the following plot portrays the transformed coordinates and standard deviations in a Euclidean space. Ellipses are in green for the eight real objects and blue for the four ideal objects.



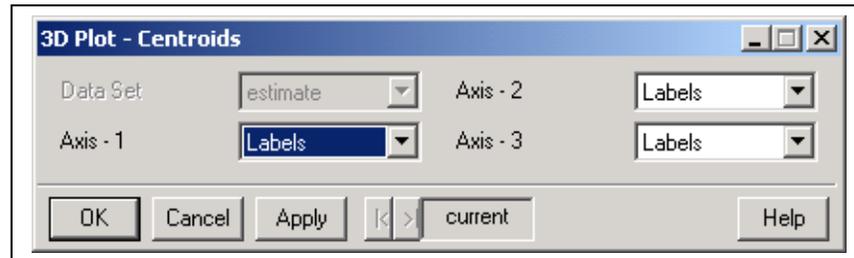
Prox/Lik Example *The following plot portrays the estimates of the previous plot in a city-block space. Diamonds are in green for the eight real objects and blue for the four ideal objects.*



I. 3D Plot – Centroids

Choose **Proscal ► 3D Plot - Centroids**. The following dialog box appears.

When through with the following steps, select **OK** or **Apply**.



Data Set

Select one of {estimate, target, transform}. (In the **3DProx/Lik** example provided with PROSCAL, no target values are read so the Data Set option is not enabled and **estimate** is automatically chosen.)

The **estimate** option plots the ML configuration estimated by PROSCAL.

The **target** option plots the target coordinates.

The **transform** option plots a transformation of the estimated configuration. The transformation aligns the estimated centroids with the target centroids. In a Euclidean space, the possible transformations of the estimated configuration include rotation, reflection, translation and dilation. In a city-block space, rotation and reflection are not allowed. City-block output should be carefully checked to determine the correct matching of transformed and target dimensions.

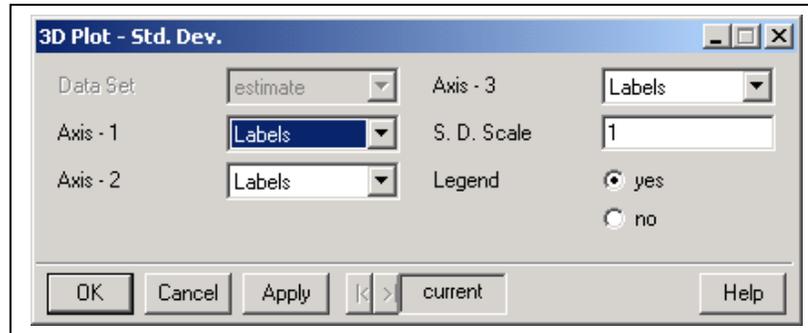
Axis-1 Axis-2 Axis-3

These three options contain a list of the column names as specified in the data frame you provided. Select whichever three columns you want to plot. Default values for column names are V1, V2, ..., V6. Unlike the 2D Plot, PROSCAL determines which column will appear on each axis. The column with the largest range of values is placed on the x-axis and the column with the smallest range of values is placed on the vertical z-axis.

J. 3D Plot – Standard Deviations

Choose **Proscal ► 3D Plot - Std. Dev.** The following dialog box appears.

When through, select **OK** or **Apply**.



Data Set

Select one of {estimate, target, transform}. (In the **3DProx/Lik** example provided with PROSCAL, no target values are read so the Data Set option is not enabled and **estimate** is automatically chosen.)

The **estimate** option plots the centroids and standard deviations of the ML configuration estimated by PROSCAL.

The **target** option plots the centroids and standard deviations of the target objects.

The **transform** option plots the centroids and standard deviations of the transformed objects. The transformation aligns the estimated centroids with the target centroids. Plots can only be made of solutions with zero valued dimensional correlations. (Three dimensional plots with non-zero valued correlations are difficult to comprehend. They are best viewed in two dimensional slices.)

Standard deviations are plotted as hyper-ellipses in a Euclidean space and as hyper-diamonds in a city-block space.

Axis-1 Axis-2 Axis-3

These three options contain a list of the column names as specified in the data frame you provided. Select whichever three columns you want to plot. Default values for column names are V1, V2, ..., V6. Unlike the 2D Plot, PROSCAL determines which column will appear on each axis. The column with the largest range of values is placed on the x-axis and the column with the smallest range of values is placed on the vertical z-axis.

S. D. Scale

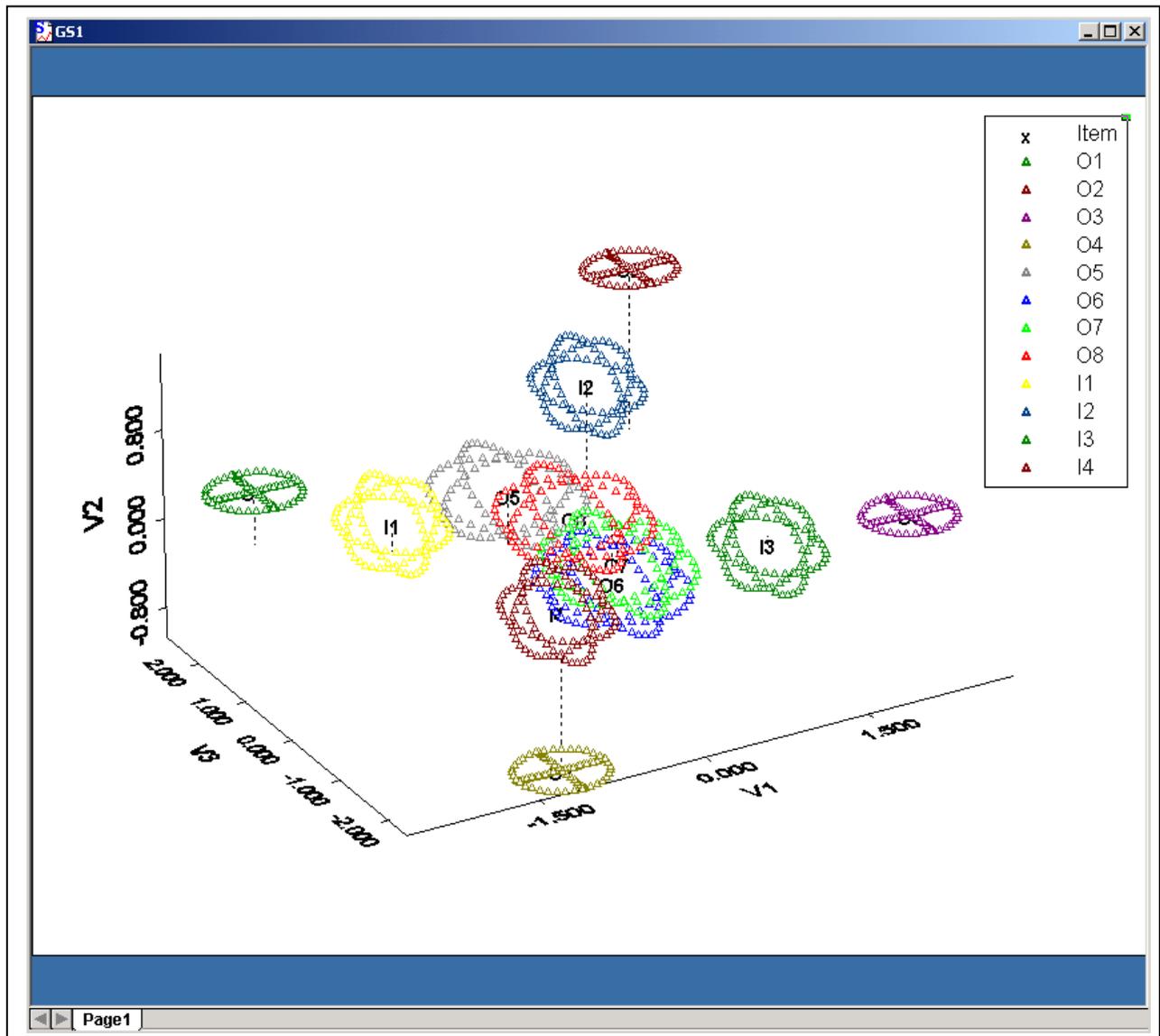
Select a positive real number to indicate the magnitude of the standard deviations which are to be plotted as ellipses. The default value of 1 means that an ellipsoid for a single standard deviation will be plotted. The visual

complexity of three dimensional plots is often reduced by using scale values less than one.

Legend

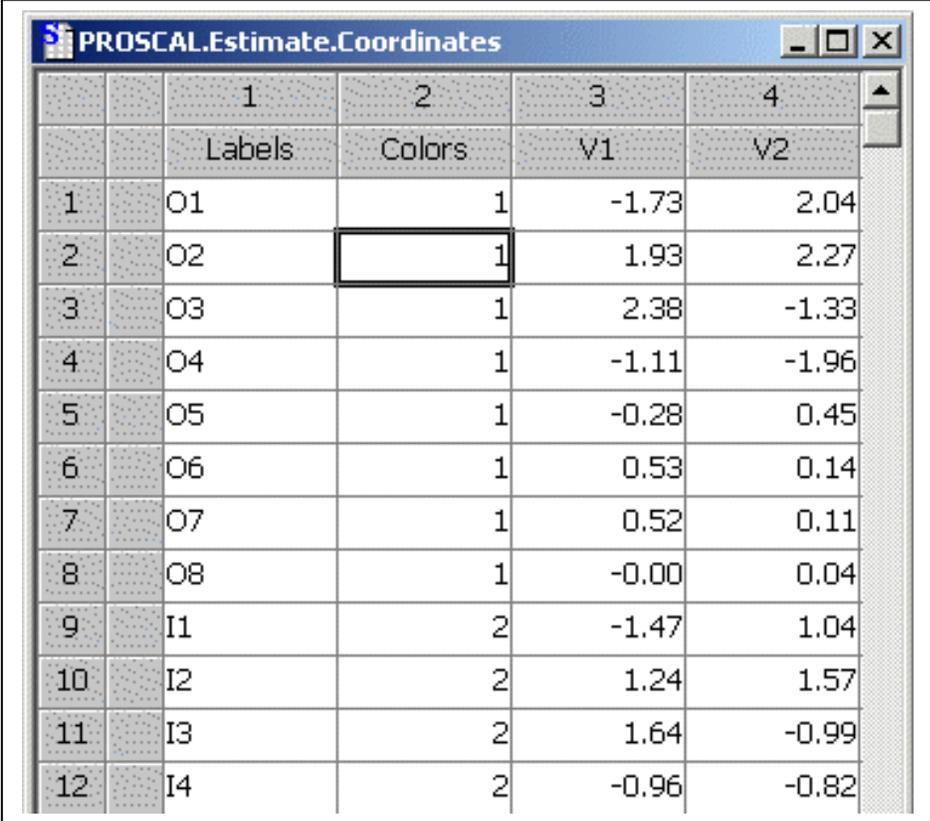
Legends may be deleted by selecting the No button. Rotation of the solution space is at times hindered by the presence of a legend.

3DProx/Lik Output *For this data set, the following plot portrays the estimated coordinates and standard deviations for a three dimensional solution in a Euclidean space. The standard deviation scale was set to 0.5 to enhance the clarity. When clicking on an empty space in the graph, four green circles and a green triangle will appear. If you drag horizontally on one of the green circles, a bounding box will appear and rotate in the direction you drag the mouse. When you release the mouse, the graph is redrawn at the new perspective. The green triangle can be dragged up and down to rotate the graph vertically.*



K. Change Labels

Choose **Proscal ► Change Labels**. The PROSCAL.Estimate.Coordinates data frame will open. The first column gives the default labels for the real (and ideal) objects. Enter whatever labels are desired. The new labels will appear in subsequent plots and in the Perceptual Share output. For the **Prox/Lik** example, the following is obtained.



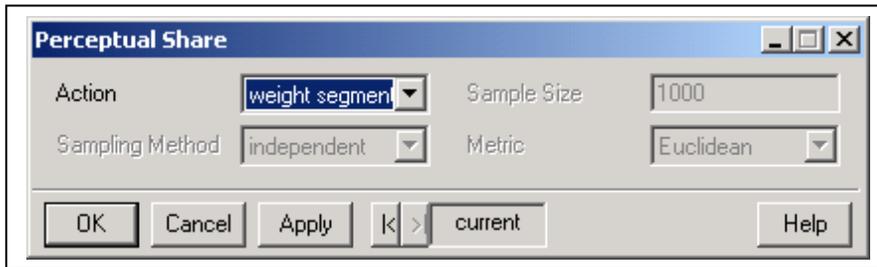
	1	2	3	4
	Labels	Colors	V1	V2
1	O1	1	-1.73	2.04
2	O2	1	1.93	2.27
3	O3	1	2.38	-1.33
4	O4	1	-1.11	-1.96
5	O5	1	-0.28	0.45
6	O6	1	0.53	0.14
7	O7	1	0.52	0.11
8	O8	1	-0.00	0.04
9	I1	2	-1.47	1.04
10	I2	2	1.24	1.57
11	I3	2	1.64	-0.99
12	I4	2	-0.96	-0.82

By default, two colors are displayed with hedonic data, green for real objects and blue for ideal objects. If more discrimination is desired, sequential integers (as in the specification of constraints) may be used to assign objects to colors.

If data are read in from a previous analysis and changes are made in the number of estimated segments, then a notice will appear with the information that the Labels and Colors saved from the last analysis are inappropriate. When this happens, correct the label and color information stored in S-PLUS files PROSCAL.Labels and PROSCAL.Colors. Then, enter `PSgenerateDataFrames()` in the S-PLUS command window. This will update the labels and colors for the new assignments.

L. Perceptual Share

Choose **Proscal** ► **Perceptual Share**. The following dialog box appears.



When through with the following steps, select **OK** or **Apply**.

Perceptual share may be calculated whenever you have hedonic data. PROSCAL calculates perceptual share by first estimating the proportion of first choices each real object would obtain for each segment and then sums the equally or unequally weighted results across the different segments. Each ideal object represents a different segment.

Action

Select one from {weight segments equally, weight segments unequally, calculate perceptual share, plot total share}.

When the option **weight segments equally** is selected, each segment contributes equally to the estimation of total perceptual share.

To weight segments differentially, select **weight segments unequally**, and then enter **OK** or **Apply**. A data frame with m rows, one for each segment, will then be opened. Enter the new weights, subject to the restriction that the weights should sum to 1.0.

After selecting the option **calculate perceptual share**, the remaining fields in the dialog box become enabled. Selecting **OK** or **Apply** produces a table with an estimated perceptual share for each segment as well as for the total market.

Selecting the option **Plot total share** produces a pie chart of the total perceptual share values.

Sampling Method Select one from {independent, dependent}. Generally, you should calculate perceptual share using the same sampling method that you used to estimate the parameters of the model. Since perceptual share may be very sensitive to the choice of a sampling method, you are given the opportunity of choosing a sampling method other than the one you used in the ML analysis.

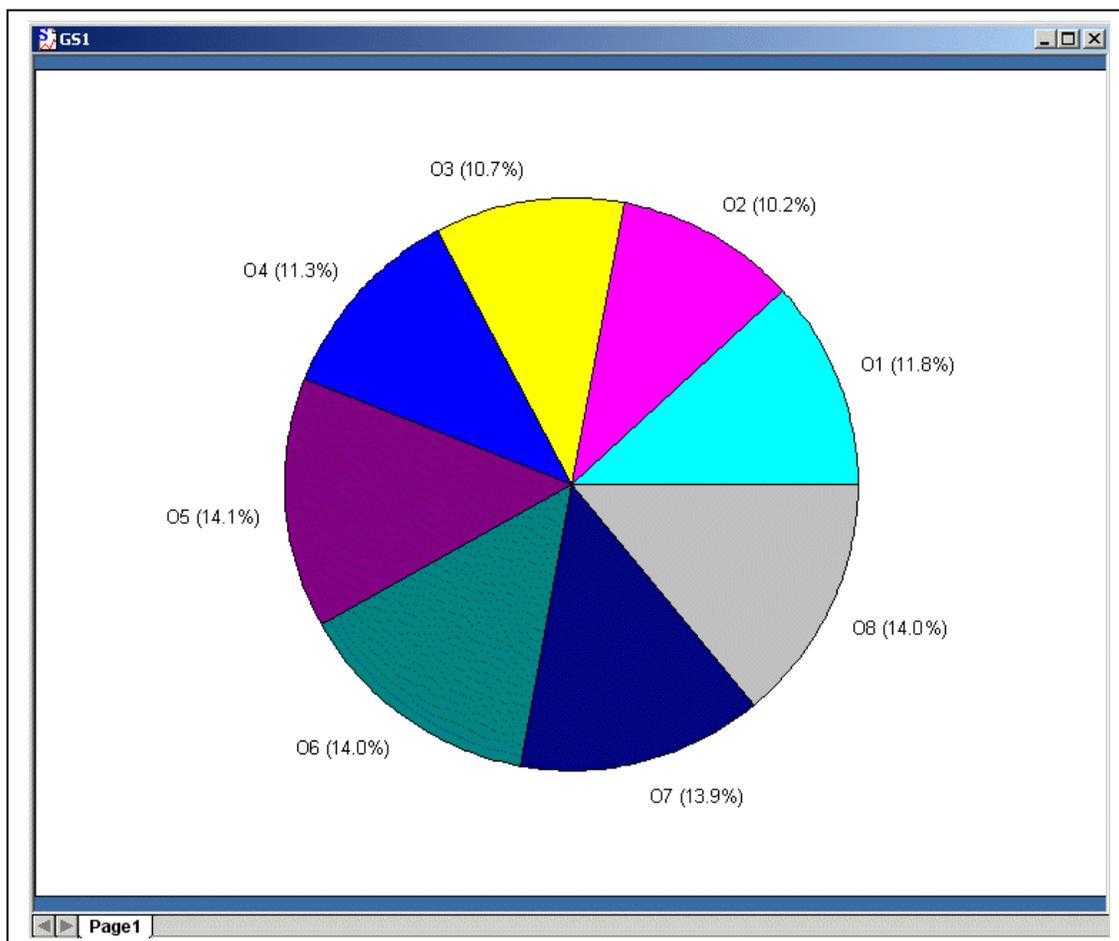
When data are obtained under artificial laboratory conditions, it may be prudent to estimate models under independent sampling assumptions and to project them under dependent sampling assumptions.

Sample Size For most situations, direct estimation of perceptual share is not practical and Monte Carlo methods must be used. Enter the sample size that you desire for the Monte Carlo analysis. For quick estimation, a sample size of 1000 is usually sufficient. Going beyond 20,000 is generally not necessary.

Metric Select one from {Euclidean, city-block}. As with the option **Sampling Method**, it is common to pick the metric that you used in the analysis of the data. You can, though, choose a metric other than the one you used in the ML analysis.

Prox/Lik Example: Perceptual Share Output. *Using equal segment weights, independent sampling, a sample size of 10,000, and a Euclidean metric, the following output is obtained. The results will differ slightly from machine to machine due to differences in how random numbers are generated.*

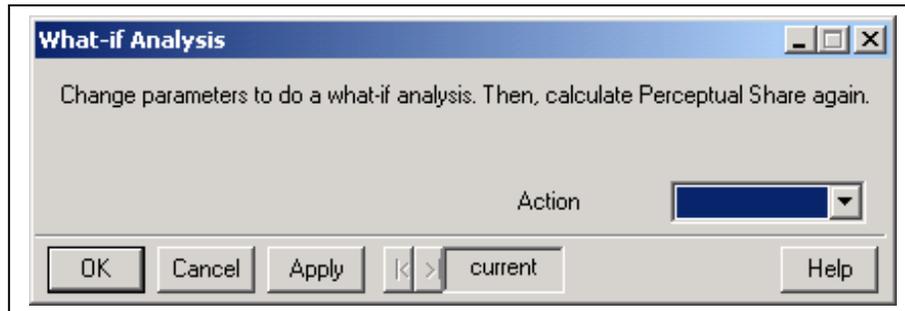
[1] "Share by Segment"								
	01	02	03	04	05	06	07	08
Segment I1	0.3340	0.0295	0.0075	0.0888	0.1689	0.1085	0.1147	0.1481
Segment I2	0.0494	0.3026	0.0666	0.0139	0.1287	0.1615	0.1532	0.1241
Segment I3	0.0081	0.0640	0.3250	0.0639	0.1002	0.1578	0.1566	0.1244
Segment I4	0.0806	0.0126	0.0305	0.2848	0.1661	0.1313	0.1315	0.1626
[1] "Total Share"								
	01	02	03	04	05	06	07	08
	0.118025	0.102175	0.1074	0.11285	0.140975	0.139775	0.139	0.1398



M. What-if Analysis

Choose **Proscal** ► **What-if Analysis**. The following dialog box appears.

Choose an Action and select **OK**.



- Action** Select one from {Change Centroids, Change Variances, Change Number of Objects}
- Change Coordinates** When you select the action **Change Centroids** and select **OK**, a data frame opens with the values of the current coordinate estimates. Change any of the values you wish. If you want to add one or more real objects, cut the coordinates for the ideal objects, enter the values for the new real objects, then paste the ideal object coordinates at the end of the real objects.
- Change Variances** When you select the action **Change Variances** and select **OK**, a data frame opens with the values of the current variance estimates. Change any of the values you wish. If you want to add one or more real objects, cut the variances for the ideal objects, enter the values for the new real objects, then paste the ideal object variances at the end of the real objects.
- Change Number of Objects** If you change the number of real or ideal objects, then you will need to update the values of NSTIM or NSUB in the PROSCAL.Commands file.

When you are through changing parameter values, recalculate the value of the Perceptual Share.

N. About PROSCAL

This gives the PROSCAL version number and date. This manual accompanies version 5.0.

O. Limits

The following table provides the current limitations of PROSCAL. There are no program limits on the number of judgments or data sets. If the number of judgments is too large for the computer's memory, the hard disk will be used to supplement the memory and the time degradation will be severe. Being able to estimate complex models with many parameters does not mean that it is desirable to do so. Our experience suggests that it is worthwhile trying to find as parsimonious a model as possible.

Parameter Limits

<i>Parameter</i>	<i>Limit</i>
<i>Number of spatial dimensions</i>	6
<i>Number of real and ideal objects</i>	60
<i>Number of parameters</i>	500

APPENDIX 1 Input Parameters for the Sample Data

Six test sets of simulated data are provided for use with the S-PLUS implementation of PROSCAL. Data used to generate the design for each set are provided in the following tables. Output files for the data sets, proxliko.dat, binaryo.dat, prefo.dat, binsimsego.dat, 3DProxLiko.dat, and deplikrato.dat may be found on the PROSCAL web site or the PROSCAL CD. The proxliko.dat output is annotated in Appendix 2. Shaded cells indicate disabled fields.

Prox/Lik Example

Project Tab	Judgments	proxlikd.dat
	Coordinate Targets	proxlikc.dat
	Variance Targets	proxlikv.dat
	Transformation Targets	
	Data Type	prox & liking ratings
	Proximities	dissimilarities
	Standardization	no
	Metric	Euclidean
	Directionality	anisotropic
	Sampling	independent
Parameters Tab	Stimuli	8
	Ideals	4
	Prox. Sets	20
	Binary Choice Sets	0
	Pref./Liking Sets	200
	Dimensions	2
	Additive	✓
	Scalar	✓
	Exponent	✓
	On Variances	✓ 1,1,1,1,2,2,2,2,3,3,3,3
	On Ideals	✓ 1,2,3,4, ...,
	No. of Ideals Per Pattern	4
	On Coordinates	✓ (accept default data frame)
Targets Tab	Initial Estimates	estimated
	Model Selection	internal

Execution Tab	No. of Iterations	20
	Optimization Index	0
	Fct. Evaluations	2000
	Mixture Model	no
	Reanalysis	no

Binary Choice Example

Project Tab	Judgments	binaryd.dat
	Coordinate Targets	binaryc.dat
	Variance Targets	binaryv.dat
	Transformation Targets	
	Data Type	binary
	Proximities	dissimilarities
	Standardization	no
	Metric	Euclidean
	Directionality	anisotropic
	Sampling	independent
Parameters Tab	Stimuli	15
	Ideals	6
	Binary Choice Sets	240
	Prox. Sets	0
	Pref./Liking Sets	0
	Dimensions	2
	Additive	
	Scalar	
	Exponent	
	On Variances	✓ 1111....111222
	On Ideals	✓ 1,2,3,4,5,6,...
	No. of Ideals Per Pattern	6
	On Coordinates	✓ (accept default data frame)
Targets Tab	Initial Estimates	estimated
	Model Selection	internal
Execution Tab	No. of Iterations	40
	Optimization Index	0
	Fct. Evaluations	4000
	Mixture Model	no
	Reanalysis	no

Preference Ratio Example

Project Tab	Judgments	prefd.dat
	Coordinate Targets	prefc.dat
	Variance Targets	prefv.dat
	Transformation Targets	
	Data Type	preference ratios
	Proximities	dissimilarities
	Standardization	no
	Metric	Euclidean
	Directionality	anisotropic
	Sampling	independent
Parameters Tab	Stimuli	12
	Ideals	4
	Prox. Sets	0
	Binary Choice Sets	0
	Pref./Liking Sets	248
	Dimensions	2
	Additive	
	Scalar	
	Exponent	
	On Variances	✓ 1,1,1,1,2,2,2,2,3,3,3,3,4,4,4,4
	On Ideals	✓ 1,2,3,4, ...,
	No. of Ideals Per Pattern	4
	On Coordinates	✓ (accept default data frame)
Targets Tab	Initial Estimates	estimated
	Model Selection	internal
Execution Tab	No. of Iterations	20
	Optimization Index	0
	Fct. Evaluations	2000
	Mixture Model	no
	Reanalysis	no

Binary Choice and Proximity Data / Segmentation Example

Project Tab	Judgments	binproxsegd.dat
	Coordinate Targets	binproxsegc.dat
	Variance Targets	binproxsegv.dat
	Transformation Targets	
	Data Type	prox & binary
	Proximities	dissimilarities
	Standardization	no
	Metric	Euclidean
	Directionality	isotropic
	Sampling	independent
Parameters Tab	Stimuli	8
	Ideals	3
	Prox. Sets	10
	Binary Choice Sets	48
	Pref./Liking Sets	0
	Dimensions	2
	Additive	
	Scalar	
	Exponent	
	On Variances	✓ 12122121222
	On Ideals	✓ 1,2,3,...,1,2,3 (first 24 ideals) 1,1, ..., 1 1 (last 24 ideals)
	No. of Ideals Per Pattern	0
	On Coordinates	✓ (accept default data frame)
Targets Tab	Initial Estimates	estimated
	Model Selection	internal
Execution Tab	No. of Iterations	40
	Optimization Index	0
	Fct. Evaluations	4000
	Mixture Model	yes
	Reanalysis	yes

3DProx/Lik Example

Project Tab	Judgments	proxlikd.dat
	Coordinate Targets	3DProxLikc.dat
	Variance Targets	3DProxLikv.dat
	Transformation Targets	3DProxLikm.dat
	Data Type	prox & liking ratings
	Proximities	dissimilarities
	Standardization	no
	Metric	Euclidean
	Directionality	anisotropic
	Sampling	independent
Parameters Tab	Stimuli	8
	Ideals	4
	Prox. Sets	20
	Binary Choice Sets	0
	Pref./Liking Sets	200
	Dimensions	3
	Additive	✓
	Scalar	✓
	Exponent	✓
	On Variances	✓ 1,1,1,1,2,2,2,2,3,3,3,3
	On Ideals	✓ 1,2,3,4, ...
	No. of Ideals Per Pattern	4
	On Coordinates	✓ (accept default data frame)
Targets Tab	Initial Estimates	target
	Model Selection	internal
Execution Tab	No. of Iterations	20
	Optimization Index	0
	Fct. Evaluations	2000
	Mixture Model	no
	Reanalysis	no

DepLikRat Example

Project Tab	Judgments	deplikratd.dat
	Coordinate Targets	deplikratc.dat
	Variance Targets	deplikratv.dat
	Transformation Targets	
	Data Type	liking ratings
	Proximities	dissimilarities
	Standardization	no
	Metric	Euclidean
	Directionality	isotropic
	Sampling	dependent
Parameters Tab	Stimuli	12
	Ideals	1
	Prox. Sets	0
	Binary Choice Sets	0
	Pref./Liking Sets	100
	Dimensions	2
	Additive	
	Scalar	
	Exponent	
	On Variances	✓ (accept default data frame)
	On Ideals	✓ (accept default data frame)
	No. of Ideals Per Pattern	0
	On Coordinates	✓ (accept default data frame)
Targets Tab	Initial Estimates	estimated
	Model Selection	internal
Execution Tab	No. of Iterations	30
	Optimization Index	0
	Fct. Evaluations	5000
	Mixture Model	no
	Reanalysis	no

APPENDIX 2 Annotated Output of the Prox/Lik Data Set

PROSCAL

Project1

Project Title

```

DATE      2010:10:05
TIME      16:38:10.118
NUMBER OF STIMULI          8
NUMBER OF ACTIVE COORD.   24
NUMBER OF DIMENSIONS       2
NUMBER OF VARIANCES        6
NUMBER OF IDEAL OBJECTS    4
NUMBER OF DATA SETS      220
NUMBER OF ML ITERATIONS    20
FUNCTION EVALUATIONS      2000
NUMBER OF DISTANCE SETS   20
TRANSFORMATION INDEX       7
OPTIMIZATION LEVEL         0
TARGET COORD. OPTION      TARG
TARGET VARIANCE OPT.      TARG
TARGET MEAS PARAM OPT.    TRNO
STANDARDIZATION OPT.      STNO
INITIALIZATION OPTION     COMP
DATA TYPE OPTION          PRRA
DISTRIBUTION OPTION       CVNO
FIXED POINT OPTION        FXNO
SAMPLING                   INDP
METRIC OPTION              ECLD
MIXTURE OPTION             NOEM
REANALYSIS OPTION         NORD
PROXIMITIES OPTION        NOSI
    
```

Date and time the run commenced.

Restatement of design parameters. Values of nominal variables refer to codes used in the executable PROSCAL program. A key to the codes is in APPENDIX 4.

TARGET CONFIGURATION

```

0.000      0.000
0.000      1.000
1.000      1.000
1.000      0.000
0.400      0.500
0.500      0.600
0.600      0.500
0.500      0.400
0.200      0.200
0.200      0.800
0.800      0.800
0.800      0.200
    
```

Target coordinates, read in from file proxlikc.dat. The first eight rows are coordinates for the real objects, the last four rows are the coordinates for the ideal objects.

In this design, target values are used only as a basis of comparison for the estimates. They do not enter the estimation process.

TARGET VARIANCES

0.040
 0.040
 0.200
 0.200
 0.200
 0.040

Target variances from file proxlikv.dat. The first pair of values is for the first four real objects, the second pair is for real objects five through eight, and the last pair is for the ideal objects. In each pair, the first value is for dimension V1, the second value is for dimension V2.

VARIANCE SET MEMBERSHIP

OBJECTS	1	2	3	4	5	6	7	8	9	10	11	12
SET	1	1	1	1	2	2	2	2	3	3	3	3

In this anisotropic Case III design, the first four real objects have the same variance, the second four real objects have the same variance, and the ideal objects have the same variance.

SUBJECT SET MEMBERSHIP

SUBJECT	SET
1	1
2	2
3	3
4	4
5	1
6	2
7	3
8	4
9	1
10	2
11	3
12	4
13	1
14	2
15	3
16	4
17	1
18	2
19	3
20	4
21	1
22	2
23	3
24	4
25	1
26	2
27	3
28	4
29	1
30	2
31	3
32	4
33	1
34	2
35	3
36	4
37	1

There are four ideal objects and 200 sets of liking rating judgments. Every fourth data set is assigned to the same ideal object.

38	2
39	3
40	4
41	1
42	2
43	3
44	4
45	1
46	2
47	3
48	4
49	1
50	2
51	3
52	4
53	1
54	2
55	3
56	4
57	1
58	2
59	3
60	4
61	1
62	2
63	3
64	4
65	1
66	2
67	3
68	4
69	1
70	2
71	3
72	4
73	1
74	2
75	3
76	4
77	1
78	2
79	3
80	4
81	1
82	2
83	3
84	4
85	1
86	2
87	3
88	4
89	1
90	2
91	3
92	4
93	1
94	2

95	3
96	4
97	1
98	2
99	3
100	4
101	1
102	2
103	3
104	4
105	1
106	2
107	3
108	4
109	1
110	2
111	3
112	4
113	1
114	2
115	3
116	4
117	1
118	2
119	3
120	4
121	1
122	2
123	3
124	4
125	1
126	2
127	3
128	4
129	1
130	2
131	3
132	4
133	1
134	2
135	3
136	4
137	1
138	2
139	3
140	4
141	1
142	2
143	3
144	4
145	1
146	2
147	3
148	4
149	1
150	2

151	3
152	4
153	1
154	2
155	3
156	4
157	1
158	2
159	3
160	4
161	1
162	2
163	3
164	4
165	1
166	2
167	3
168	4
169	1
170	2
171	3
172	4
173	1
174	2
175	3
176	4
177	1
178	2
179	3
180	4
181	1
182	2
183	3
184	4
185	1
186	2
187	3
188	4
189	1
190	2
191	3
192	4
193	1
194	2
195	3
196	4
197	1
198	2
199	3
200	4

A three parameter measurement model was requested. These two values are the initial estimates of the additive and scale coefficients respectively. *i.e.*,
 $\delta_{ij} = 0.862 + 3.44 d_{ij}$

INITIAL ADDITIVE AND SCALE COEFFICIENTS FOR DISSIMILARITY JUDGMENT
TRANSFORMATION 0.862263D+00 0.344344D+01

When preference ratios or liking ratings are present, PROSCAL begins the initial estimation process by estimating I scales (Coombs, 1964) for each of the ideal objects.

```

ISCALE FOR IDEAL OBJECT  1
    2.470   4.150   4.910   3.810   3.210   3.490   3.550   3.370

ISCALE FOR IDEAL OBJECT  2
    3.790   2.510   3.850   4.570   3.090   3.490   3.470   3.470

ISCALE FOR IDEAL OBJECT  3
    4.890   3.970   2.370   3.550   3.630   3.010   3.430   3.450

ISCALE FOR IDEAL OBJECT  4
    3.850   4.850   3.890   2.730   2.810   3.550   3.390   2.990

```

INITIAL CONFIGURATION

```

-3.082      0.654
 0.756      3.061
 3.183     -0.596
-0.675     -3.145
-1.220      0.521
 0.578      0.011
 0.786     -0.055
-0.326     -0.451
-3.082      0.654
 0.756      3.061
 3.183     -0.596
-0.675     -3.145

```

These are the initial coordinate values which are used to initiate the ML analysis. The first eight rows are for the real objects and the last four rows are for the ideal objects.

INITIAL STANDARD DEVIATIONS - ALL OBJECTS

```

1.348
1.367
1.306
1.485
1.315
0.987
1.319
1.010
1.028
1.024
0.987
1.017

```

As part of the initial estimation process, an isotropic standard deviation is estimated for each object, real and ideal.

```

MEAN OF STANDARD DEVIATIONS      1.183

```

INITIAL VARIANCES - ACTIVE VARIABLES

1.895
 1.895
 1.340
 1.340
 1.028
 1.028

Initial variances. These are of necessity isotropic, even when an anisotropic analysis is requested. Three sets of variances are estimated. The first two values are for the first set, etc.

INITIAL MEASUREMENT CONSTANTS -0.2504 0.2904 1.0000

TRANSFORMED CONFIGURATION

-0.407 -0.403
 -0.389 0.416
 0.405 0.419
 0.406 -0.417
 -0.202 -0.134
 0.054 0.086
 0.085 0.111
 0.034 -0.096
 -0.407 -0.403
 -0.389 0.416
 0.405 0.419
 0.406 -0.417

Transformed initial estimates are the result of a proximity transformation on the initial estimates. At this point, the initial values are Euclidean and Case V in nature.

Recovery of the targets is measured in two ways. The first is the product moment correlation of all pairs of inter-object (inter-point) distances for the target and transformed configurations. The second is the sum of the Euclidean distances between corresponding estimated and target objects.

CORRELATION OF TARGET AND ESTIMATED DISTANCES = 0.8377 SUM OF DISTANCE
 DIFFERENCES = 0.150595E+01

TRANSFORMED VARIANCES AND COVARIANCES

0.062
 0.000 0.062
 0.044
 0.000 0.044
 0.034
 0.000 0.034

The iterative ML procedure now begins.

INITIAL LOG LIKELIHOOD VALUE -0.34608970E+04

Estimate values are reported first by dimension across objects for the coordinates and then sequentially for the variances. With anisotropic Case III problems, the order of the variances is first across dimension and then across objects.

ESTIMATE VALUES ENTERING FIXED LOCATION PHASE

-3.082	0.756	3.183	-0.675	-1.220	0.578	0.786	-0.326	-3.082	
0.756									
3.183	-0.675	0.654	3.061	-0.596	-3.145	0.521	0.011	-0.055	-
0.451									
0.654	3.061	-0.596	-3.145	1.895	1.895	1.340	1.340	1.028	
1.028									

LOG LIKELIHOOD VALUE AT END OF PHASE -0.342901E+04

ESTIMATE VALUES ENTERING FIXED VARIANCE PHASE

-3.082	0.756	3.183	-0.675	-1.220	0.578	0.786	-0.326	-3.082	
0.756									
3.183	-0.675	0.654	3.061	-0.596	-3.145	0.521	0.011	-0.055	-
0.451									
0.654	3.061	-0.596	-3.145	0.999	1.204	1.763	2.066	1.762	
1.822									

LOG LIKELIHOOD VALUE AT END OF PHASE -0.295843E+04

Measurement model estimates are reported for the requested parameters. The order of the estimates is for the additive, linear and exponential terms.

MEASUREMENT PARAMETERS 0.400000E-01 0.290408E+00 0.976404E+00

ESTIMATE VALUES ENTERING FIXED LOCATION PHASE

-2.069	1.018	2.546	-0.284	-0.338	0.686	0.600	-0.070	-1.781	
0.558									
2.038	-0.503	0.917	2.345	-0.444	-2.039	0.240	0.153	0.187	-
0.215									
0.507	1.997	-0.679	-1.476	0.999	1.204	1.763	2.066	1.762	
1.822									

LOG LIKELIHOOD VALUE AT END OF PHASE -0.294308E+04

ESTIMATE VALUES ENTERING FIXED VARIANCE PHASE

-2.069	1.018	2.546	-0.284	-0.338	0.686	0.600	-0.070	-1.781	
0.558									
2.038	-0.503	0.917	2.345	-0.444	-2.039	0.240	0.153	0.187	-
0.215									
0.507	1.997	-0.679	-1.476	0.704	0.758	1.495	2.885	1.152	
2.185									

LOG LIKELIHOOD VALUE AT END OF PHASE -0.293759E+04

MEASUREMENT PARAMETERS 0.400000E-01 0.283205E+00 0.976404E+00

ESTIMATE VALUES ENTERING FIXED LOCATION PHASE

-1.999	1.510	2.511	-0.746	-0.357	0.594	0.571	-0.062	-1.631	
0.873									
1.822	-0.813	1.556	2.416	-0.851	-2.059	0.368	0.180	0.197	-
0.058									
0.846	1.878	-0.862	-1.115	0.704	0.758	1.495	2.885	1.152	
2.185									

LOG LIKELIHOOD VALUE AT END OF PHASE -0.293574E+04

ESTIMATE VALUES ENTERING FIXED VARIANCE PHASE

-1.999	1.510	2.511	-0.746	-0.357	0.594	0.571	-0.062	-1.631	
0.873									
1.822	-0.813	1.556	2.416	-0.851	-2.059	0.368	0.180	0.197	-
0.058									
0.846	1.878	-0.862	-1.115	0.629	0.660	1.428	3.288	1.133	
2.072									

LOG LIKELIHOOD VALUE AT END OF PHASE -0.293431E+04

MEASUREMENT PARAMETERS 0.400000E-01 0.278314E+00 0.976404E+00

ESTIMATE VALUES ENTERING FIXED LOCATION PHASE

-1.853	1.784	2.473	-0.971	-0.317	0.573	0.555	-0.024	-1.521	
1.114									
1.719	-0.885	1.864	2.361	-1.128	-2.007	0.418	0.161	0.144	
0.006									
0.979	1.708	-0.924	-0.915	0.629	0.660	1.428	3.288	1.133	
2.072									

LOG LIKELIHOOD VALUE AT END OF PHASE -0.293379E+04

ESTIMATE VALUES ENTERING FIXED VARIANCE PHASE

-1.853	1.784	2.473	-0.971	-0.317	0.573	0.555	-0.024	-1.521	
1.114									
1.719	-0.885	1.864	2.361	-1.128	-2.007	0.418	0.161	0.144	
0.006									
0.979	1.708	-0.924	-0.915	0.618	0.651	1.420	3.536	1.125	
1.979									

LOG LIKELIHOOD VALUE AT END OF PHASE -0.293340E+04

MEASUREMENT PARAMETERS 0.400000E-01 0.276896E+00 0.976404E+00

After finishing the alternating ML estimation process, a final phase in which all coordinate and variance estimates may be updated is conducted.

ESTIMATE VALUES ENTERING LAST PHASE

-1.763	1.895	2.412	-1.080	-0.295	0.550	0.531	-0.006	-1.483
1.206								
1.659	-0.938	2.005	2.300	-1.279	-1.966	0.438	0.150	0.121
0.037								
1.033	1.614	-0.962	-0.836	0.618	0.651	1.420	3.536	1.125
1.979								

Function evaluations, number of observations and use of exact or approximate ML estimates are the primary determinants of execution time. Constraint evaluations are usually 0. A high number of constraint evaluations may indicate the advisability of using a simpler model.

FINAL LOG LIKELIHOOD VALUE	-0.293327E+04
NUMBER OF FREE PARAMETERS	30
NUMBER OF JUDGMENTS	2160
CAIC	0.612688E+04
BIC	0.609688E+04

The final log likelihood along with the information criterion statistics CAIC and BIC of Bozdogan and Schwarz.

FUNCTION EVALUATIONS = 4006 CONSTRAINT EVALUATIONS = 0

FINAL CONFIGURATION

-1.731	2.039
1.927	2.268
2.384	-1.332
-1.114	-1.965
-0.277	0.447
0.534	0.138
0.518	0.107
-0.004	0.037
-1.475	1.042
1.237	1.575
1.639	-0.987
-0.961	-0.817

Final estimates, before transformation to the targets (if provided).

FINAL VARIANCES

0.608
0.653
1.420
3.615
1.118
1.943

FINAL MEASUREMENT TRANSFORMATION COEFFICIENTS

0.399999999000E-01
0.276896415243E+00
0.976403566657E+00

TRANSFORMED CONFIGURATION

-0.540	-0.438
-0.469	0.512
0.475	0.501
0.513	-0.422
-0.078	-0.120
0.030	0.077
0.038	0.072
0.037	-0.065
-0.274	-0.408
-0.315	0.310
0.360	0.322
0.223	-0.342

Final transformed coordinates and their recovery measures.

CORRELATION OF TARGET AND ESTIMATED DISTANCES = 0.9747 SUM OF DISTANCE DIFFERENCES = 0.797740E+00

TRANSFORMED VARIANCES AND COVARIANCES

0.044	
0.000	0.041
0.242	
-0.020	0.099
0.130	
-0.008	0.077

Final transformed variances and their correlation with the target values.

PRODUCT MOMENT CORRELATION OF TARGET AND FINAL TRANSFORMED STIMULUS VARIANCES = 0.8636

ESTIMATED TRUE AND EXPECTED DISTANCES

OBJECTS	TRUE	EXPECTED
2 1	0.366482E+01	0.384689E+01
3 1	0.531912E+01	0.543992E+01
3 2	0.362871E+01	0.380266E+01
4 1	0.405120E+01	0.420585E+01
4 2	0.521187E+01	0.533318E+01
4 3	0.355456E+01	0.374229E+01
5 1	0.215596E+01	0.296833E+01
5 2	0.285848E+01	0.348636E+01
5 3	0.320058E+01	0.377354E+01
5 4	0.255329E+01	0.319754E+01
6 1	0.295658E+01	0.356220E+01
6 2	0.254503E+01	0.321670E+01
6 3	0.236301E+01	0.312130E+01
6 4	0.267158E+01	0.331767E+01
6 5	0.867372E+00	0.288541E+01
7 1	0.296435E+01	0.356680E+01
7 2	0.257983E+01	0.324117E+01
7 3	0.235661E+01	0.311817E+01
7 4	0.263722E+01	0.329283E+01
7 5	0.864033E+00	0.288423E+01
7 6	0.349669E-01	0.277639E+01
8 1	0.264431E+01	0.330333E+01
8 2	0.295067E+01	0.353557E+01
8 3	0.275197E+01	0.342388E+01
8 4	0.228875E+01	0.303636E+01
8 5	0.493075E+00	0.281024E+01
8 6	0.546987E+00	0.281994E+01
8 7	0.526079E+00	0.281672E+01
9 1	0.102964E+01	0.205594E+01
9 2	0.361608E+01	0.396328E+01
9 3	0.453032E+01	0.479496E+01
9 4	0.302821E+01	0.337175E+01
9 5	0.133758E+01	0.278103E+01
9 6	0.220251E+01	0.324310E+01
9 7	0.220079E+01	0.324084E+01
9 8	0.178172E+01	0.298734E+01
10 1	0.300340E+01	0.342945E+01
10 2	0.978222E+00	0.203804E+01
10 3	0.312498E+01	0.346602E+01
10 4	0.424909E+01	0.449857E+01
10 5	0.188721E+01	0.304243E+01
10 6	0.159947E+01	0.288824E+01
10 7	0.163442E+01	0.290444E+01
10 8	0.197605E+01	0.308111E+01
10 9	0.276340E+01	0.342694E+01
11 1	0.452885E+01	0.478062E+01
11 2	0.326734E+01	0.357794E+01
11 3	0.820919E+00	0.198353E+01
11 4	0.292156E+01	0.335289E+01
11 5	0.239293E+01	0.334824E+01
11 6	0.157698E+01	0.287940E+01
11 7	0.156653E+01	0.287487E+01
11 8	0.193541E+01	0.307373E+01
11 9	0.371631E+01	0.418695E+01
11 10	0.259299E+01	0.320629E+01
12 1	0.295856E+01	0.331768E+01
12 2	0.422649E+01	0.449070E+01

“True” values refer to the distances between the means of the objects in the estimated (not transformed) configuration. “Expected” values are the corresponding expected distances. True and expected distances are calculated using the metric specified in the design.

12	3	0.338446E+01	0.376281E+01
12	4	0.115730E+01	0.211070E+01
12	5	0.143833E+01	0.281641E+01
12	6	0.177448E+01	0.298489E+01
12	7	0.174426E+01	0.296946E+01
12	8	0.128326E+01	0.275318E+01
12	9	0.192904E+01	0.278934E+01
12	10	0.324888E+01	0.376113E+01
12	11	0.260584E+01	0.330924E+01

ELAPSED MINUTES = 1

This time was obtained using a
3.2GHz Pentium IV processor.

APPENDIX 3 Segmentation Analysis Output

When an EM algorithm is specified for estimating the segment membership, the initial analysis, based upon the user's *a priori* estimate of segment membership, proceeds as it usually does. The output is unchanged.

After the initial analysis is completed, the EM phase begins. PROSCAL goes through cycles in which the probability of segment membership is first estimated and then the measurement model coefficient, centroid and variance estimates are updated. The cycles stop when the likelihoods no longer continue to increase.

The probabilities of each subject being in each segment are then printed. Deterministic segment assignments, based upon the segment with the highest probability of membership, are also printed.

When a reanalysis is requested, the deterministic segment assignments of the prior stage are used to begin a new analysis. Initial coordinate and variance estimates are based upon the original judgments or target values, not the EM estimates. The output is identical to what appears in a standard analysis.

APPENDIX 4 Key for Instructional Variables' Nominal Output

Variable	Code	Definition
Target Configuration	TARG	Target configuration to be read
	TRNO	No target configuration to be read
Target Variances	TARG	Target variances to be read
	TRNO	No target variances to be read
Target Measurement Parameters	TARG	Target transformation coefficients to be read
	TRNO	No target transformation coefficients to be read
Standardization of Dissimilarities	STDZ	Standardization performed
	STNO	No standardization performed
Initial Estimates	TARG	Defined by target values
	COMP	Computed independent of target values
Data Type	PROX	Proximities
	PREF	Preference ratios
	PRPR	Proximities and preference ratios
	RATE	Liking ratings
	PRRA	Proximities and liking ratings
	BINR	Binary preferential choices
	PRBR	Proximities and binary choices
	BRRA	Binary choices and liking ratings
Distribution (directionality)	CHIF	Isotropic analysis
	CVNO	Anisotropic analysis
External / Internal analysis	FXNO	Internal analysis
	STFC	External analysis – coordinates fixed
	STFA	External analysis – coordinates and variances fixed
Sampling	INDP	Independent sampling
	DEPN	Dependent sampling
Metric	ECLD	Euclidean metric
	CITY	City-block metric
Mixture model	EMMM	Expectation maximization mixture model chosen
	NOEM	No mixture model analysis chosen
Reanalysis	REDO	Reanalysis chosen
	NORD	No reanalysis chosen
Proximities	NOSI	Dissimilarities data
	GAUS	Gaussian similarities transformation chosen
	EXPN	Exponential similarities transformation chosen

Transformation model index	0	$\delta_{ij} = d_{ij}$
	1	$\delta_{ij} = a + d_{ij}$
	2	$\delta_{ij} = bd_{ij}$
	3	$\delta_{ij} = d_{ij}^c$
	4	$\delta_{ij} = a + bd_{ij}$
	5	$\delta_{ij} = a + d_{ij}^c$
	6	$\delta_{ij} = bd_{ij}^c$
	7	$\delta_{ij} = a + bd_{ij}^c$

REFERENCES

- Akaike, H. (1974). A new look at the statistical model identification. *IEEE Transactions on Automatic Control*, AC-19, 716-723.
- Bozdogan, H. (1987). Model selection and Akaike's information criterion (AIC): the general theory and its analytical extensions. *Psychometrika*, 52, 345-370.
- Chandler, J. P. (1969). STEPIT – Finds local minima of a smooth function of several parameters. *Behavioral Science*, 14, 81-82.
- Cliff, N. (1966). Orthogonal rotation to congruence." *Psychometrika*, 31, 33-42.
- Coombs, C. H. (1964). *A Theory of Data*. New York: Wiley.
- Hefner, R. (1958). *Extensions to the Law of Comparative Judgment to Discriminable and Multidimensional Stimuli*. Unpublished Ph.D. dissertation, Psychology Department, University of Michigan.
- Imhof, J. P. (1961). "Computing the distribution of quadratic forms in normal variables," *Biometrika*, 48, 419-426.
- Jensen, D. R. and Solomon, H. (1972). "A Gaussian approximation to the distribution of a definite quadratic form," *Journal of the American Statistical Association*, 67, 898-902.
- MacKay, D. B. (1983). "Alternative probabilistic scaling models for spatial data," *Geographical Analysis*, 15, 173-186.
- MacKay, D. B. (1989). "Probabilistic multidimensional scaling: an anisotropic model for distance judgments," *Journal of Mathematical Psychology*, 33, 187-205.
- MacKay, D. B. (2001). "Probabilistic multidimensional scaling using a city-block metric," *Journal of Mathematical Psychology*, 45, 249-264.
- MacKay, D. B. (2007). "Internal multidimensional unfolding about a single ideal – a probabilistic solution," *Journal of Mathematical Psychology*, 51, 305-318.
- MacKay, D. B., Easley, R. F. And Zinnes, J. L. (1995). A single ideal point model for market structure analysis, *Journal of Marketing Research*, 32, 433-443.
- MacKay, D. B. and Lilly, B. (2004). "Percept variance, subadditivity and the metric classification of similarity, and dissimilarity data, *Journal of Classification*, 21, 185-206.
- MacKay, D. B. and Zinnes, J. L. (1986). "A probabilistic model for the multidimensional scaling of proximity and preference data," *Marketing Science*, 5, 348-349.

- MacKay, D. B. And Zinnes, J. L. (1995). "Probabilistic multidimensional unfolding: an anisotropic model for preference ratio judgments," *Journal of Mathematical Psychology*, 99 - 111.
- Pearson, E. S. (1959). "Note on an approximation to the distribution of non-central χ^2 ," *Biometrika*, 46, 364.
- Schwarz, G. (1978), "Estimating the dimension of a model," *The Annals of Statistics*, 6, 461-464.
- Shepard, R. N. (1987), "Toward a Universal Law of Generalization for Psychological Science," *Science*, 237, 1317-1323.
- Thurstone, L. L. (1927). "A law of comparative judgment," *Psychological Review*, 34, 273-286.
- Zinnes, J. L. and MacKay, D. B. (1989). "Probabilistic multidimensional analysis of preference ratio judgments," in G. De Soete, H. Feger and K. C. Klauer (eds.), *New Developments in Psychological Choice Modeling*. New York: Elsevier, 177-205.
- Zinnes, J. L. and MacKay, D. B. (1983). "Probabilistic multidimensional scaling: complete and incomplete data," *Psychometrika*, 48, 27-48.