Glossary

agglomerative hierarchical clustering methods. Agglomerative hierarchical clustering methods are bottom-up methods for hierarchical clustering. Each observation begins in a separate group. The closest pair of groups is agglomerated or merged in each iteration until all the data are in one cluster. This process creates a hierarchy of clusters. Contrast to *divisive hierarchical clustering methods*.

anti-image correlation matrix or **anti-image covariance matrix**. The image of a variable is defined as that part which is predictable by regressing each variable on all the other variables; hence, the anti-image is the part of the variable that cannot be predicted. The anti-image correlation matrix $\mathbf{A}$ is a matrix of the negatives of the partial correlations among variables. Partial correlations represent the degree to which the factors explain each other in the results. The diagonal of the anti-image correlation matrix is the Kaiser–Meyer–Olkin measure of sampling adequacy for the individual variables. Variables with small values should be eliminated from the analysis. The anti-image covariance matrix $\mathbf{C}$ contains the negatives of the partial covariances and has one minus the squared multiple correlations in the principal diagonal. Most of the off-diagonal elements should be small in both anti-image matrices in a good factor model. Both anti-image matrices can be calculated from the inverse of the correlation matrix $\mathbf{R}$ via

$$\mathbf{A} = \{\text{diag}(\mathbf{R})\}^{-1}\mathbf{R}\{\text{diag}(\mathbf{R})\}^{-1}$$
$$\mathbf{C} = \{\text{diag}(\mathbf{R})\}^{-1/2}\mathbf{R}\{\text{diag}(\mathbf{R})\}^{-1/2}$$

Also see *Kaiser–Meyer–Olkin measure of sampling adequacy*.

average-linkage clustering. Average-linkage clustering is a hierarchical clustering method that uses the average proximity of observations between groups as the proximity measure between the two groups.

Bayes's theorem. Bayes's theorem states that the probability of an event, A, conditional on another event, B, is generally different from the probability of B conditional on A, although the two are related. Bayes's theorem is that

$$P(A|B) = \frac{P(B|A)P(A)}{P(B)}$$

where $P(A)$ is the marginal probability of A, and $P(A|B)$ is the conditional probability of A given B, and likewise for $P(B)$ and $P(B|A)$.

Bentler's invariant pattern simplicity rotation. Bentler's (1977) rotation maximizes the invariant pattern simplicity. It is an oblique rotation that minimizes the criterion function

$$c(\mathbf{\Lambda}) = -\log[|(\mathbf{\Lambda}^2)'\mathbf{\Lambda}^2|] + \log[|\text{diag}\{(\mathbf{\Lambda}^2)'\mathbf{\Lambda}^2\}|]$$

See *Crawford–Ferguson rotation* for a definition of $\mathbf{\Lambda}$. Also see *oblique rotation*.

between matrix and **within matrix**. The between and within matrices are SSCP matrices that measure the spread between groups and within groups, respectively. These matrices are used in multivariate analysis of variance and related hypothesis tests: Wilks's lambda, Roy's largest root, Lawley–Hotelling trace, and Pillai's trace.

Here we have k independent random samples of size n. The between matrix $\mathbf{H}$ is given by

$$\mathbf{H} = n\sum_{i=1}^{k}(\overline{\mathbf{y}}_{i\bullet} - \overline{\mathbf{y}}_{\bullet\bullet})(\overline{\mathbf{y}}_{i\bullet} - \overline{\mathbf{y}}_{\bullet\bullet})' = \sum_{i=1}^{k}\frac{1}{n}\mathbf{y}_{i\bullet}\mathbf{y}_{i\bullet}' - \frac{1}{kn}\mathbf{y}_{\bullet\bullet}\mathbf{y}_{\bullet\bullet}'$$

The within matrix $\mathbf{E}$ is defined as

$$\mathbf{E} = \sum_{i=1}^{k}\sum_{j=1}^{n}(\mathbf{y}_{ij} - \overline{\mathbf{y}}_{i\bullet})(\mathbf{y}_{ij} - \mathbf{y}_{i\bullet})' = \sum_{i=1}^{k}\sum_{j=1}^{n}\mathbf{y}_{ij}\mathbf{y}_{ij}' - \sum_{i=1}^{k}\frac{1}{n}\mathbf{y}_{i\bullet}\mathbf{y}_{i\bullet}'$$

Also see *SSCP matrix*.

biplot. A biplot is a scatterplot which represents both observations and variables simultaneously. There are many different biplots; variables in biplots are usually represented by arrows and observations are usually represented by points.

biquartimax rotation or **biquartimin rotation**. Biquartimax rotation and biquartimin rotation are synonyms. They put equal weight on the varimax and quartimax criteria, simplifying the columns and rows of the matrix. This is an oblique rotation equivalent to an oblimin rotation with $\gamma = 0.5$. Also see *varimax rotation*, *quartimax rotation*, and *oblimin rotation*.

boundary solution or **Heywood solution**. See *Heywood case*.

CA. See *correspondence analysis*.

canonical correlation analysis. Canonical correlation analysis attempts to describe the relationships between two sets of variables by finding linear combinations of each so that the correlation between the linear combinations is maximized.

canonical discriminant analysis. Canonical linear discriminant analysis is LDA where describing how groups are separated is of primary interest. Also see *linear discriminant analysis*.

canonical loadings. The canonical loadings are coefficients of canonical linear discriminant functions. Also see *canonical discriminant analysis* and *loading*.

canonical variate set. The canonical variate set is a linear combination or weighted sum of variables obtained from canonical correlation analysis. Two sets of variables are analyzed in canonical correlation analysis. The first canonical variate of the first variable set is the linear combination in standardized form that has maximal correlation with the first canonical variate from the second variable set. The subsequent canonical variates are uncorrelated to the previous and have maximal correlation under that constraint.

centered data. Centered data has zero mean. You can center data $\mathbf{x}$ by taking $\mathbf{x} - \overline{\mathbf{x}}$.

centroid-linkage clustering. Centroid-linkage clustering is a hierarchical clustering method that computes the proximity between two groups as the proximity between the group means.

classical scaling. Classical scaling is a method of performing MDS via an eigen decomposition. This is contrasted to modern MDS, which is achieved via the minimization of a loss function. Also see *multidimensional scaling* and *modern scaling*.

classification. Classification is the act of allocating or classifying observations to groups as part of discriminant analysis. In some sources, classification is synonymous with cluster analysis.

classification function. Classification functions can be obtained after LDA or QDA. They are functions based on Mahalanobis distance for classifying observations to the groups. See *discriminant function* for an alternative. Also see *linear discriminant analysis* and *quadratic discriminant analysis*.

classification table. A classification table, also known as a confusion matrix, gives the count of observations from each group that are classified into each of the groups as part of a discriminant analysis. The element at (i, j) gives the number of observations that belong to the ith group but were classified into the jth group. High counts are expected on the diagonal of the table where observations are correctly classified, and small values are expected off the diagonal. The columns of the matrix are categories of the predicted classification; the rows represent the actual group membership.

cluster analysis. Cluster analysis is a method for determining natural groupings or clusters of observations.

cluster tree. See *dendrogram*.

clustering. See *cluster analysis*.

common factors. Common factors are found by factor analysis. They linearly reconstruct the original variables. In factor analysis, reconstruction is defined in terms of prediction of the correlation matrix of the original variables.

communality. Communality is the proportion of a variable's variance explained by the common factors in factor analysis. It is also "1 − uniqueness". Also see *uniqueness*.

complete-linkage clustering. Complete-linkage clustering is a hierarchical clustering method that uses the farthest pair of observations between two groups to determine the proximity of the two groups.

component scores. Component scores are calculated after PCA. Component scores are the coordinates of the original variables in the space of principal components.

Comrey's tandem 1 and 2 rotations. Comrey (1967) describes two rotations, the first (tandem 1) to judge which "small" factors should be dropped, the second (tandem 2) for "polishing".

Tandem principle 1 minimizes the criterion

$$c(\mathbf{\Lambda}) = \left\langle \mathbf{\Lambda}^2, (\mathbf{\Lambda}\mathbf{\Lambda}')^2\mathbf{\Lambda}^2 \right\rangle$$

Tandem principle 2 minimizes the criterion

$$c(\mathbf{\Lambda}) = \left\langle \mathbf{\Lambda}^2, \{\mathbf{1}\mathbf{1}' - (\mathbf{\Lambda}\mathbf{\Lambda}')^2\}\mathbf{\Lambda}^2 \right\rangle$$

See *Crawford–Ferguson rotation* for a definition of $\mathbf{\Lambda}$.

configuration. The configuration in MDS is a representation in a low-dimensional (usually 2-dimensional) space with distances in the low-dimensional space approximating the dissimilarities or disparities in high-dimensional space. Also see *multidimensional scaling*, *dissimilarity*, and *disparity*.

configuration plot. A configuration plot after MDS is a (usually 2-dimensional) plot of labeled points showing the low-dimensional approximation to the dissimilarities or disparities in high-dimensional space. Also see *multidimensional scaling*, *dissimilarity*, and *disparity*.

confusion matrix. A confusion matrix is a synonym for a classification table after discriminant analysis. See *classification table*.

contrast or **contrasts**. In ANOVA, a contrast in k population means is defined as a linear combination

$$\delta = c_1\mu_1 + c_2\mu_2 + \cdots + c_k\mu_k$$

where the coefficients satisfy

$$\sum_{i=1}^{k} c_i = 0$$

In the multivariate setting (MANOVA), a contrast in k population mean vectors is defined as

$$\boldsymbol{\delta} = c_1\boldsymbol{\mu}_1 + c_2\boldsymbol{\mu}_2 + \cdots c_k\boldsymbol{\mu}_k$$

where the coefficients again satisfy

$$\sum_{i=1}^{k} c_i = 0$$

The univariate hypothesis $\delta = 0$ may be tested with `contrast` (or `test`) after ANOVA. The multivariate hypothesis $\boldsymbol{\delta} = 0$ may be tested with `manovatest` after MANOVA.

correspondence analysis. Correspondence analysis (CA) gives a geometric representation of the rows and columns of a two-way frequency table. The geometric representation is helpful in understanding the similarities between the categories of variables and associations between variables. CA is calculated by singular value decomposition. Also see *singular value decomposition*.

correspondence analysis projection. A correspondence analysis projection is a line plot of the row and column coordinates after CA. The goal of this graph is to show the ordering of row and column categories on each principal dimension of the analysis. Each principal dimension is represented by a vertical line; markers are plotted on the lines where the row and column categories project onto the dimensions. Also see *correspondence analysis*.

costs. Costs in discriminant analysis are the cost of misclassifying observations.

covarimin rotation. Covarimin rotation is an orthogonal rotation equivalent to varimax. Also see *varimax rotation*.

Crawford–Ferguson rotation. Crawford–Ferguson (1970) rotation is a general oblique rotation with several interesting special cases.

Special cases of the Crawford–Ferguson rotation include

κ	Special case
0	quartimax / quartimin
$1/p$	varimax / covarimin
$f/(2p)$	equamax
$(f-1)/(p+f-2)$	parsimax
1	factor parsimony

p = number of rows of $\mathbf{A}$.
f = number of columns of $\mathbf{A}$.

Where $\mathbf{A}$ is the matrix to be rotated, $\mathbf{T}$ is the rotation and $\boldsymbol{\Lambda} = \mathbf{AT}$. The Crawford–Ferguson rotation is achieved by minimizing the criterion

$$c(\boldsymbol{\Lambda}) = \frac{1-\kappa}{4}\left\langle \boldsymbol{\Lambda}^2, \boldsymbol{\Lambda}^2(\mathbf{11}' - \mathbf{I})\right\rangle + \frac{\kappa}{4}\left\langle \boldsymbol{\Lambda}^2, (\mathbf{11}' - \mathbf{I})\boldsymbol{\Lambda}^2\right\rangle$$

Also see *oblique rotation*.

crossed variables or **stacked variables**. In CA and MCA crossed categorical variables may be formed from the interactions of two or more existing categorical variables. Variables that contain these interactions are called crossed or stacked variables.

crossing variables or **stacking variables**. In CA and MCA, crossing or stacking variables are the existing categorical variables whose interactions make up a crossed or stacked variable.

curse of dimensionality. The curse of dimensionality is a term coined by Richard Bellman (1961) to describe the problem caused by the exponential increase in size associated with adding extra dimensions to a mathematical space. On the unit interval, 10 evenly spaced points suffice to sample with no more distance than 0.1 between them; however a unit square requires 100 points, and a unit cube requires 1000 points. Many multivariate statistical procedures suffer from the curse of dimensionality. Adding variables to an analysis without adding sufficient observations can lead to imprecision.

dendrogram or **cluster tree**. A dendrogram or cluster tree graphically presents information about how observations are grouped together at various levels of (dis)similarity in hierarchical cluster analysis. At the bottom of the dendrogram, each observation is considered its own cluster. Vertical lines extend up for each observation, and at various (dis)similarity values, these lines are connected to the lines from other observations with a horizontal line. The observations continue to combine until, at the top of the dendrogram, all observations are grouped together. Also see *hierarchical clustering*.

dilation. A dilation stretches or shrinks distances in Procrustes rotation.

dimension. A dimension is a parameter or measurement required to define a characteristic of an object or observation. Dimensions are the variables in the dataset. Weight, height, age, blood pressure, and drug dose are examples of dimensions in health data. Number of employees, gross income, net income, tax, and year are examples of dimensions in data about companies.

discriminant analysis. Discriminant analysis is used to describe the differences between groups and to exploit those differences when allocating (classifying) observations of unknown group membership. Discriminant analysis is also called classification in many references.

discriminant function. Discriminant functions are formed from the eigenvectors from Fisher's approach to LDA. See *linear discriminant analysis*. See *classification function* for an alternative.

discriminating variables. Discriminating variables in a discriminant analysis are analyzed to determine differences between groups where group membership is known. These differences between groups are then exploited when classifying observations to the groups.

disparity. Disparities are transformed dissimilarities, that is, dissimilarity values transformed by some function. The class of functions to transform dissimilarities to disparities may either be 1) a class of metric, or known functions such as linear functions or power functions that can be parameterized by real scalars or 2) a class of more general (nonmetric) functions, such as any monotonic function. Disparities are used in MDS. Also see *dissimilarity*, *multidimensional scaling*, *metric scaling*, and *nonmetric scaling*.

dissimilarity, **dissimilarity matrix**, and **dissimilarity measure**. Dissimilarity or a dissimilarity measure is a quantification of the difference between two things, such as observations or variables or groups of observations or a method for quantifying that difference. A dissimilarity matrix is a matrix containing dissimilarity measurements. Euclidean distance is one example of a dissimilarity measure. Contrast to *similarity*. Also see *proximity* and *Euclidean distance*.

divisive hierarchical clustering methods. Divisive hierarchical clustering methods are top-down methods for hierarchical clustering. All the data begin as a part of one large cluster; with each iteration, a cluster is broken into two to create two new clusters. At the first iteration there are two

clusters, then three, and so on. Divisive methods are very computationally expensive. Contrast to *agglomerative hierarchical clustering methods*.

eigenvalue. An eigenvalue is the scale factor by which an eigenvector is multiplied. For many multivariate techniques, the size of an eigenvalue indicates the importance of the corresponding eigenvector. Also see *eigenvector*.

eigenvector. An eigenvector of a linear transformation is a nonzero vector that is either left unaffected or simply multiplied by a scale factor after the transformation.

Here $\mathbf{x}$ is an eigenvector of linear transformation $\mathbf{A}$ with eigenvalue λ:

$$\mathbf{Ax} = \lambda\mathbf{x}$$

For many multivariate techniques, eigenvectors form the basis for analysis and interpretation. Also see *loading*.

equamax rotation. Equamax rotation is an orthogonal rotation whose criterion is a weighted sum of the varimax and quartimax criteria. Equamax reflects a concern for simple structure within the rows and columns of the matrix. It is equivalent to oblimin with $\gamma = p/2$, or to the Crawford–Ferguson family with $\kappa = f/2p$, where p is the number of rows of the matrix to be rotated, and f is the number of columns. Also see *orthogonal rotation*, *varimax rotation*, *quartimax rotation*, *oblimin rotation*, and *Crawford–Ferguson rotation*.

Euclidean distance. The Euclidean distance between two observations is the distance one would measure with a ruler. The distance between vector $\mathbf{P} = (P_1, P_2, \ldots, P_n)$ and $\mathbf{Q} = (Q_1, Q_2, \ldots, Q_n)$ is given by

$$D(\mathbf{P}, \mathbf{Q}) = \sqrt{(P_1 - Q_1)^2 + (P_2 - Q_2)^2 + \cdots + (P_n - Q_n)^2} = \sqrt{\sum_{i=1}^{n}(P_i - Q_i)^2}$$

factor. A factor is an unobserved random variable that is thought to explain variability among observed random variables.

factor analysis. Factor analysis is a statistical technique used to explain variability among observed random variables in terms of fewer unobserved random variables called factors. The observed variables are then linear combinations of the factors plus error terms.

If the correlation matrix of the observed variables is $\mathbf{R}$, then $\mathbf{R}$ is decomposed by factor analysis as

$$\mathbf{R} = \mathbf{\Lambda\Phi\Lambda}' + \mathbf{\Psi}$$

$\mathbf{\Lambda}$ is the loading matrix, and $\mathbf{\Psi}$ contains the specific variances, for example, the variance specific to the variable not explained by the factors. The default unrotated form assumes uncorrelated common factors, $\mathbf{\Phi} = \mathbf{I}$.

factor loading plot. A factor loading plot produces a scatter plot of the factor loadings after factor analysis.

factor loadings. Factor loadings are the regression coefficients which multiply the factors to produce the observed variables in the factor analysis.

factor parsimony. Factor parsimony is an oblique rotation, which maximizes the column simplicity of the matrix. It is equivalent to a Crawford–Ferguson rotation with $\kappa = 1$. Also see *oblique rotation* and *Crawford–Ferguson rotation*.

factor scores. Factor scores are computed after factor analysis. Factor scores are the coordinates of the original variables, $\mathbf{x}$, in the space of the factors. The two types of scoring are regression scoring (Thomson 1951) and Bartlett (1937, 1938) scoring.

Using the symbols defined in *factor analysis*, the formula for regression scoring is

$$\widehat{\mathbf{f}} = \mathbf{\Lambda}'\mathbf{R}^{-1}\mathbf{x}$$

In the case of oblique rotation the formula becomes

$$\widehat{\mathbf{f}} = \mathbf{\Phi}\mathbf{\Lambda}'\mathbf{R}^{-1}\mathbf{x}$$

The formula for Bartlett scoring is

$$\widehat{\mathbf{f}} = \mathbf{\Gamma}^{-1}\mathbf{\Lambda}'\mathbf{\Psi}^{-1}\mathbf{x}$$

where

$$\mathbf{\Gamma} = \mathbf{\Lambda}'\mathbf{\Psi}^{-1}\mathbf{\Lambda}$$

Also see *factor analysis*.

Heywood case or **Heywood solution**. A Heywood case can appear in factor analysis output; this indicates that a boundary solution, called a Heywood solution, was produced. The geometric assumptions underlying the likelihood-ratio test are violated, though the test may be useful if interpreted cautiously.

hierarchical clustering and **hierarchical clustering methods**. In hierarchical clustering, the data is placed into clusters via iterative steps. Contrast to *partition clustering*. Also see *agglomerative hierarchical clustering methods* and *divisive hierarchical clustering methods*.

Hotelling's T-squared generalized means test. Hotelling's T-squared generalized means test is a multivariate test that reduces to a standard t test if only one variable is specified. It tests whether one set of means is zero or if two sets of means are equal.

inertia. In CA, the inertia is related to the definition in applied mathematics of "moment of inertia", which is the integral of the mass times the squared distance to the centroid. Inertia is defined as the total Pearson chi-squared for the two-way table divided by the total number of observations, or the sum of the squared singular values found in the singular value decomposition.

$$\text{total inertia} = \frac{1}{n}\chi^2 = \sum_k \lambda_k^2$$

In MCA, the inertia is defined analogously. In the case of the indicator or Burt matrix approach, it is given by the formula

$$\text{total inertia} = \left(\frac{q}{q-1}\right)\sum \phi_t^2 - \frac{(J-q)}{q^2}$$

where q is the number of active variables, J is the number of categories and ϕ_t is the tth (unadjusted) eigenvalue of the eigen decomposition. In JCA the total inertia of the modified Burt matrix is defined as the sum of the inertias of the off-diagonal blocks. Also see *correspondence analysis* and *multiple correspondence analysis*.

iterated principal-factor method. The iterated principal-factor method is a method for performing factor analysis in which the communalities $\widehat{h}_i^2$ are estimated iteratively from the loadings in $\widehat{\mathbf{\Lambda}}$ using

$$\widehat{h}_i^2 = \sum_{j=1}^{m} \widehat{\lambda}_{ij}^2$$

Also see *factor analysis* and *communality*.

JCA. An acronym for joint correspondence analysis; see *multiple correspondence analysis*.

joint correspondence analysis. See *multiple correspondence analysis*.

Kaiser–Meyer–Olkin measure of sampling adequacy. The Kaiser–Meyer–Olkin (KMO) measure of sampling adequacy takes values between 0 and 1, with small values meaning that the variables have too little in common to warrant a factor analysis or PCA. Historically, the following labels have been given to values of KMO (Kaiser 1974):

0.00 to 0.49	unacceptable
0.50 to 0.59	miserable
0.60 to 0.69	mediocre
0.70 to 0.79	middling
0.80 to 0.89	meritorious
0.90 to 1.00	marvelous

kmeans. Kmeans is a method for performing partition cluster analysis. The user specifies the number of clusters, k, to create using an iterative process. Each observation is assigned to the group whose mean is closest, and then based on that categorization, new group means are determined. These steps continue until no observations change groups. The algorithm begins with k seed values, which act as the k group means. There are many ways to specify the beginning seed values. Also see *partition clustering*.

kmedians. Kmedians is a variation of kmeans. The same process is performed, except that medians instead of means are computed to represent the group centers at each step. Also see *kmeans* and *partition clustering*.

KMO. See *Kaiser–Meyer–Olkin measure of sampling adequacy*.

KNN. See *kth nearest neighbor*.

Kruskal stress. The Kruskal stress measure (Kruskal 1964; Cox and Cox 2001, 63) used in MDS is given by

$$\text{Kruskal}(\widehat{\mathbf{D}}, \mathbf{E}) = \left\{ \frac{\sum (E_{ij} - \widehat{D}_{ij})^2}{\sum E_{ij}^2} \right\}^{1/2}$$

where D_{ij} is the dissimilarity between objects i and j, $1 \leq i, j \leq n$, and $\widehat{D}_{ij}$ is the disparity, that is, the transformed dissimilarity, and E_{ij} is the Euclidean distance between rows i and j of the matching configuration. Kruskal stress is an example of a loss function in modern MDS. After classical MDS, `estat stress` gives the Kruskal stress. Also see *classical scaling*, *multidimensional scaling*, and *stress*.

kth nearest neighbor. kth-nearest-neighbor (KNN) discriminant analysis is a nonparametric discrimination method based on the k nearest neighbors of each observation. Both continuous and binary data can be handled through the different similarity and dissimilarity measures. KNN analysis can distinguish irregular-shaped groups, including groups with multiple modes. Also see *discriminant analysis* and *nonparametric methods*.

Lawley–Hotelling trace. The Lawley–Hotelling trace is a test statistic for the hypothesis test $H_0 : \boldsymbol{\mu}_1 = \boldsymbol{\mu}_2 = \cdots = \boldsymbol{\mu}_k$ based on the eigenvalues $\lambda_1, \lambda_2, \ldots, \lambda_s$ of $\mathbf{E}^{-1}\mathbf{H}$. It is defined as

$$U^{(s)} = \text{trace}(\mathbf{E}^{-1}\mathbf{H}) = \sum_{i=1}^{s} \lambda_i$$

where $\mathbf{H}$ is the between matrix and $\mathbf{E}$ is the within matrix, see *between matrix*.

LDA. See *linear discriminant analysis*.

leave one out. In discriminant analysis, classification of an observation while leaving it out of the estimation sample is done to check the robustness of the analysis; thus the phrase "leave one out" (LOO). Also see *discriminant analysis*.

linear discriminant analysis. Linear discriminant analysis (LDA) is a parametric form of discriminant analysis. In Fisher's (1936) approach to LDA, linear combinations of the discriminating variables provide maximal separation between the groups. The Mahalanobis (1936) formulation of LDA assumes that the observations come from multivariate normal distributions with equal covariance matrices. Also see *discriminant analysis* and *parametric methods*.

linkage. In cluster analysis, the linkage refers to the measure of proximity between groups or clusters.

loading. A loading is a coefficient or weight in a linear transformation. Loadings play an important role in many multivariate techniques, including factor analysis, PCA, MANOVA, LDA, and canonical correlations. In some settings, the loadings are of primary interest and are examined for interpretability. For many multivariate techniques, loadings are based on an eigenanalysis of a correlation or covariance matrix. Also see *eigenvector*.

loading plot. A loading plot is a scatter plot of the loadings after LDA, factor analysis or PCA.

logistic discriminant analysis. Logistic discriminant analysis is a form of discriminant analysis based on the assumption that the likelihood ratios of the groups have an exponential form. Multinomial logistic regression provides the basis for logistic discriminant analysis. Because multinomial logistic regression can handle binary and continuous regressors, logistic discriminant analysis is also appropriate for binary and continuous discriminating variables. Also see *discriminant analysis*.

LOO. See *leave one out*.

loss. Modern MDS is performed by minimizing a loss function, also called a loss criterion. The loss quantifies the difference between the disparities and the Euclidean distances.

Loss functions include Kruskal's stress and its square, both normalized with either disparities or distances, the strain criterion which is equivalent to classical metric scaling when the disparities equal the dissimilarities, and the Sammon (1969) mapping criterion which is the sum of the scaled, squared differences between the distances and the disparities, normalized by the sum of the disparities.

Also see *multidimensional scaling*, *Kruskal stress*, *classical scaling*, and *disparity*.

Mahalanobis distance. The Mahalanobis distance measure is a scale-invariant way of measuring distance. It takes into account the correlations of the dataset.

Mahalanobis transformation. The Mahalanobis transformation takes a Cholesky factorization of the inverse of the covariance matrix $\mathbf{S}^{-1}$ in the formula for Mahalanobis distance and uses it to transform the data. If we have the Cholesky factorization $\mathbf{S}^{-1} = \mathbf{L}'\mathbf{L}$, then the Mahalanobis transformation of $\mathbf{x}$ is $\mathbf{z} = \mathbf{L}\mathbf{x}$, and $\mathbf{z}'\mathbf{z} = D_M^2(\mathbf{x})$.

MANCOVA. MANCOVA is multivariate analysis of covariance. See *multivariate analysis of variance*.

MANOVA. *multivariate analysis of variance*.

mass. In CA and MCA, the mass is the marginal probability. The sum of the mass over the active row or column categories equals 1.

matching coefficient. The matching similarity coefficient is used to compare two binary variables. If a is the number of observations that both have value 1, and d is the number of observations that both have value 0, and b, c are the number of $(1,0)$ and $(0,1)$ observations, respectively, then the matching coefficient is given by

$$\frac{a+d}{a+b+c+d}$$

Also see *similarity measure*.

matching configuration. In MDS, the matching configuration is the low dimensional configuration whose distances approximate the high-dimensional dissimilarities or disparities. Also see *multidimensional scaling*, *dissimilarity*, and *disparity*.

matching configuration plot. After MDS, this is a scatter plot of the matching configuration.

maximum likelihood factor method. The maximum likelihood factor method is a method for performing factor analysis that assumes multivariate normal observations. It maximizes the determinant of the partial correlation matrix; thus, this solution is also meaningful as a descriptive method for nonnormal data. Also see *factor analysis*.

MCA. See *multiple correspondence analysis*.

MDS. See *multidimensional scaling*.

MDS configuration plot. See *configuration plot*.

measure. A measure is a quantity representing the proximity between objects or method for determining the proximity between objects. Also see *proximity*.

median-linkage clustering. Median-linkage clustering is a hierarchical clustering method that uses the distance between the medians of two groups to determine the similarity or dissimilarity of the two groups. Also see *cluster analysis* and *agglomerative hierarchical clustering methods*.

metric scaling. Metric scaling is a type of MDS, in which the dissimilarities are transformed to disparities via a class of known functions. This is contrasted to *nonmetric scaling*. Also see *multidimensional scaling*.

minimum entropy rotation. The minimum entropy rotation is an orthogonal rotation achieved by minimizing the deviation from uniformity (entropy). The minimum entropy criterion (Jennrich 2004) is

$$c(\mathbf{\Lambda}) = -\frac{1}{2}\left\langle \mathbf{\Lambda}^2, \log \mathbf{\Lambda}^2 \right\rangle$$

See *Crawford–Ferguson rotation* for a definition of $\mathbf{\Lambda}$. Also see *orthogonal rotation*.

misclassification rate. The misclassification rate calculated after discriminant analysis is, in its simplest form, the fraction of observations incorrectly classified. See *discriminant analysis*.

modern scaling. Modern scaling is a form of MDS that is achieved via the minimization of a loss function that compares the disparities (transformed dissimilarities) in the higher-dimensional space and the distances in the lower-dimensional space. Contrast to *classical scaling*. Also see *dissimilarity*, *disparity*, *multidimensional scaling*, and *loss*.

multidimensional scaling. Multidimensional scaling (MDS) is a dimension-reduction and visualization technique. Dissimilarities (for instance, Euclidean distances) between observations in a high-dimensional space are represented in a lower-dimensional space which is typically two dimensions

so that the Euclidean distance in the lower-dimensional space approximates in some sense the dissimilarities in the higher-dimensional space. Often the higher-dimensional dissimilarities are first transformed to disparities, and the disparities are then approximated by the distances in the lower-dimensional space. Also see *dissimilarity*, *disparity*, *classical scaling*, *loss*, *modern scaling*, *metric scaling*, and *nonmetric scaling*.

multiple correspondence analysis. Multiple correspondence analysis (MCA) and joint correspondence analysis (JCA) are methods for analyzing observations on categorical variables. MCA and JCA analyze a multiway table and are usually viewed as an extension of CA. Also see *correspondence analysis*.

multivariate analysis of covariance. See *multivariate analysis of variance*.

multivariate analysis of variance. Multivariate analysis of variance (MANOVA) is used to test hypotheses about means. Four multivariate statistics are commonly computed in MANOVA: Wilks's lambda, Pillai's trace, Lawley–Hotelling trace, and Roy's largest root. Also see *Wilks's lambda*, *Pillai's trace*, *Lawley–Hotelling trace*, and *Roy's largest root*.

multivariate regression. Multivariate regression is a method of estimating a linear (matrix) model

$$\mathbf{Y} = \mathbf{XB} + \mathbf{\Xi}$$

Multivariate regression is estimated by least-squares regression, and it can be used to test hypotheses, much like MANOVA.

nearest neighbor. See *kth nearest neighbor*.

nonmetric scaling. Nonmetric scaling is a type of modern MDS in which the dissimilarities may be transformed to disparities via any monotonic function as opposed to a class of known functions. Contrast to *metric scaling*. Also see *multidimensional scaling*, *dissimilarity*, *disparity*, and *modern scaling*.

nonparametric methods. Nonparametric statistical methods, such as KNN discriminant analysis, do not assume the population fits any parameterized distribution.

normalization. Normalization presents information in a standard form for interpretation. In CA the row and column coordinates can be normalized in different ways depending on how one wishes to interpret the data. Normalization is also used in rotation, MDS, and MCA.

oblimax rotation. Oblimax rotation is a method for oblique rotation which maximizes the number of high and low loadings. When restricted to orthogonal rotation, oblimax is equivalent to quartimax rotation. Oblimax minimizes the oblimax criterion

$$c(\mathbf{\Lambda}) = -\log(\langle \mathbf{\Lambda}^2, \mathbf{\Lambda}^2 \rangle) + 2\log(\langle \mathbf{\Lambda}, \mathbf{\Lambda} \rangle)$$

See *Crawford–Ferguson rotation* for a definition of $\mathbf{\Lambda}$. Also see *oblique rotation*, *orthogonal rotation*, and *quartimax rotation*.

oblimin rotation. Oblimin rotation is a general method for oblique rotation, achieved by minimizing the oblimin criterion

$$c(\mathbf{\Lambda}) = \frac{1}{4}\langle \mathbf{\Lambda}^2, \{\mathbf{I} - (\gamma/p)\mathbf{1}\mathbf{1}'\}\mathbf{\Lambda}^2(\mathbf{1}\mathbf{1}' - \mathbf{I})\rangle$$

Oblimin has several interesting special cases:

γ	Special case
0	quartimax / quartimin
$1/2$	biquartimax / biquartimin
1	varimax / covarimin
$p/2$	equamax

p = number of rows of $\mathbf{A}$.

See *Crawford–Ferguson rotation* for a definition of $\mathbf{\Lambda}$ and $\mathbf{A}$. Also see *oblique rotation*.

oblique rotation or **oblique transformation**. An oblique rotation maintains the norms of the rows of the matrix but not their inner products. In geometric terms, this maintains the lengths of vectors, but not the angles between them. In contrast, in orthogonal rotation, both are preserved.

ordination. Ordination is the ordering of a set of data points with respect to one or more axes. MDS is a form of ordination.

orthogonal rotation or **orthogonal transformation**. Orthogonal rotation maintains both the norms of the rows of the matrix and also inner products of the rows of the matrix. In geometric terms, this maintains both the lengths of vectors and the angles between them. In contrast, oblique rotation maintains only the norms, that is, the lengths of vectors.

parametric methods. Parametric statistical methods, such as LDA and QDA, assume the population fits a parameterized distribution. For example, for LDA we assume the groups are multivariate normal with equal covariance matrices.

parsimax rotation. Parsimax rotation is an orthogonal rotation that balances complexity between the rows and the columns. It is equivalent to the Crawford–Ferguson family with $\kappa = (f-1)/(p+f-2)$, where p is the number of rows of the original matrix, and f is the number of columns. See *orthogonal rotation* and *Crawford–Ferguson rotation*.

partially specified target rotation. Partially specified target rotation minimizes the criterion

$$c(\mathbf{\Lambda}) = \|\mathbf{W} \otimes (\mathbf{\Lambda} - \mathbf{H})\|^2$$

for a given target matrix $\mathbf{H}$ and a nonnegative weighting matrix $\mathbf{W}$ (usually zero–one valued). See *Crawford–Ferguson rotation* for a definition of $\mathbf{\Lambda}$.

partition clustering and **partition cluster-analysis methods**. Partition clustering methods break the observations into a distinct number of nonoverlapping groups. This is accomplished in one step, unlike hierarchical cluster-analysis methods, in which an iterative procedure is used. Consequently, this method is quicker and will allow larger datasets than the hierarchical clustering methods. Contrast to *hierarchical clustering*. Also see *kmeans* and *kmedians*.

PCA. See *principal component analysis*.

Pillai's trace. Pillai's trace is a test statistic for the hypothesis test $H_0 : \boldsymbol{\mu}_1 = \boldsymbol{\mu}_2 = \cdots = \boldsymbol{\mu}_k$ based on the eigenvalues $\lambda_1, \ldots, \lambda_s$ of $\mathbf{E}^{-1}\mathbf{H}$. It is defined as

$$V^{(s)} = \text{trace}[(\mathbf{E} + \mathbf{H})^{-1}\mathbf{H}] = \sum_{i=1}^{s} \frac{\lambda_i}{1 + \lambda_i}$$

where $\mathbf{H}$ is the between matrix and $\mathbf{E}$ is the within matrix. See *between matrix*.

posterior probabilities. After discriminant analysis, the posterior probabilities are the probabilities of a given observation being assigned to each of the groups based on the prior probabilities, the training data, and the particular discriminant model. Contrast to *prior probabilities*.

principal component analysis. Principal component analysis (PCA) is a statistical technique used for data reduction. The leading eigenvectors from the eigen decomposition of the correlation or the covariance matrix of the variables describe a series of uncorrelated linear combinations of the variables that contain most of the variance. In addition to data reduction, the eigenvectors from a PCA are often inspected to learn more about the underlying structure of the data.

principal factor method. The principal factor method is a method for factor analysis in which the factor loadings, sometimes called factor patterns, are computed using the squared multiple correlations as estimates of the communality. Also see *factor analysis* and *communality*.

prior probabilities Prior probabilities in discriminant analysis are the probabilities of an observation belonging to a group before the discriminant analysis is performed. Prior probabilities are often based on the prevalence of the groups in the population as a whole. Contrast to *posterior probabilities*.

Procrustes rotation. A Procrustes rotation is an orthogonal or oblique transformation, that is, a restricted Procrustes transformation without translation or dilation (uniform scaling).

Procrustes transformation. The goal of Procrustes transformation is to transform the source matrix $\mathbf{X}$ to be as close as possible to the target $\mathbf{Y}$. The permitted transformations are any combination of dilation (uniform scaling), rotation and reflection (that is, orthogonal or oblique transformations), and translation. Closeness is measured by residual sum of squares. In some cases, unrestricted Procrustes transformation is desired; this allows the data to be transformed not just by orthogonal or oblique rotations, but by all conformable regular matrices $\mathbf{A}$. Unrestricted Procrustes transformation is equivalent to a multivariate regression.

The name comes from Procrustes of Greek mythology; Procrustes invited guests to try his iron bed. If the guest was too tall for the bed, Procrustes would amputate the guest's feet, and if the guest was too short, he would stretch the guest out on a rack.

Also see *orthogonal rotation*, *oblique rotation*, *dilation*, and *multivariate regression*.

promax power rotation. Promax power rotation is an oblique rotation. It does not fit in the minimizing-a-criterion framework that is at the core of most other rotations. The promax method (Hendrickson and White 1964) was proposed before computing power became widely available. The promax rotation consists of three steps:

1. Perform an orthogonal rotation.
2. Raise the elements of the rotated matrix to some power, preserving the sign of the elements. Typically the power is in the range $2 \leq \text{power} \leq 4$. This operation is meant to distinguish clearly between small and large values.
3. The matrix from step two is used as the target for an oblique Procrustean rotation from the original matrix.

proximity, **proximity matrix**, and **proximity measure**. Proximity or a proximity measure means the nearness or farness of two things, such as observations or variables or groups of observations or a method for quantifying the nearness or farness between two things. A proximity is measured by a similarity or dissimilarity. A proximity matrix is a matrix of proximities. Also see *similarity* and *dissimilarity*.

QDA. See *quadratic discriminant analysis*.

quadratic discriminant analysis. Quadratic discriminant analysis (QDA) is a parametric form of discriminant analysis and is a generalization of LDA. Like LDA, QDA assumes that the observations

come from a multivariate normal distribution, but unlike LDA, the groups are not assumed to have equal covariance matrices. Also see *discriminant analysis*, *linear discriminant analysis*, and *parametric methods*.

quartimax rotation. Quartimax rotation maximizes the variance of the squared loadings within the rows of the matrix. It is an orthogonal rotation that is equivalent to minimizing the criterion

$$c(\mathbf{\Lambda}) = \sum_i \sum_r \lambda_{ir}^4 = -\frac{1}{4}\left\langle \mathbf{\Lambda}^2, \mathbf{\Lambda}^2 \right\rangle$$

See *Crawford–Ferguson rotation* for a definition of $\mathbf{\Lambda}$.

quartimin rotation. Quartimin rotation is an oblique rotation that is equivalent to quartimax rotation when quartimin is restricted to orthogonal rotations. Quartimin is equivalent to oblimin rotation with $\gamma = 0$. Also see *quartimax rotation*, *oblique rotation*, *orthogonal rotation*, and *oblimin rotation*.

reflection. A reflection is an orientation reversing orthogonal transformation, that is, a transformation that involves negating coordinates in one or more dimensions. A reflection is a Procrustes transformation.

repeated measures. Repeated measures data have repeated measurements for the subjects over some dimension, such as time—for example test scores at the start, midway, and end of the class. The repeated observations are typically not independent. Repeated-measures ANOVA is one approach for analyzing repeated measures data, and MANOVA is another. Also see *sphericity*.

rotation. A rotation is an orientation preserving orthogonal transformation. A rotation is a Procrustes transformation.

Roy's largest root. Roy's largest root test is a test statistic for the hypothesis test $H_0 : \boldsymbol{\mu}_1 = \cdots = \boldsymbol{\mu}_k$ based on the largest eigenvalue of $\mathbf{E}^{-1}\mathbf{H}$. It is defined as

$$\theta = \frac{\lambda_1}{1 + \lambda_1}$$

Here $\mathbf{H}$ is the between matrix, and $\mathbf{E}$ is the within matrix. See *between matrix*.

Sammon mapping criterion. The Sammon (1969) mapping criterion is a loss criterion used with MDS; it is the sum of the scaled, squared differences between the distances and the disparities, normalized by the sum of the disparities. Also see *multidimensional scaling*, *modern scaling*, and *loss*.

score. A score for an observation after factor analysis, PCA, or LDA is derived from a column of the loading matrix and is obtained as the linear combination of that observation's data by using the coefficients found in the loading.

score plot. A score plot produces scatterplots of the score variables after factor analysis, PCA, or LDA.

scree plot. A scree plot is a plot of eigenvalues or singular values ordered from greatest to least after an eigen decomposition or singular value decomposition. Scree plots help determine the number of factors or components in an eigen analysis. Scree is the accumulation of loose stones or rocky debris lying on a slope or at the base of a hill or cliff; this plot is called a scree plot because it looks like a scree slope. The goal is to determine the point where the mountain gives way to the fallen rock.

Shepard diagram. A Shepard diagram after MDS is a 2-dimensional plot of high-dimensional dissimilarities or disparities versus the resulting low-dimensional distances. Also see *multidimensional scaling*.

similarity, **similarity matrix**, and **similarity measure**. A similarity or a similarity measure is a quantification of how alike two things are, such as observations or variables or groups of observations, or a method for quantifying that alikeness. A similarity matrix is a matrix containing similarity measurements. The matching coefficient is one example of a similarity measure. Contrast to *dissimilarity*. Also see *proximity* and *matching coefficient*.

single-linkage clustering. Single-linkage clustering is a hierarchical clustering method that computes the proximity between two groups as the proximity between the closest pair of observations between the two groups.

singular value decomposition. A singular value decomposition (SVD) is a factorization of a rectangular matrix. It says that if $\mathbf{M}$ is an $m \times n$ matrix, there exists a factorization of the form

$$\mathbf{M} = \mathbf{U}\mathbf{\Sigma}\mathbf{V}^*$$

where $\mathbf{U}$ is an $m \times m$ unitary matrix, $\mathbf{\Sigma}$ is an $m \times n$ matrix with nonnegative numbers on the diagonal and zeros off the diagonal, and $\mathbf{V}^*$ is the conjugate transpose of $\mathbf{V}$, an $n \times n$ unitary matrix. If $\mathbf{M}$ is a real matrix, then so is $\mathbf{V}$, and $\mathbf{V}^* = \mathbf{V}'$.

sphericity. Sphericity is the state or condition of being a sphere. In repeated measures ANOVA, sphericity concerns the equality of variance in the difference between successive levels of the repeated measure. The multivariate alternative to ANOVA, called MANOVA, does not require the assumption of sphericity. Also see *repeated measures*.

SSCP matrix. SSCP is an acronym for the sums of squares and cross products. Also see *between matrix*.

stacked variables. See *crossed variables*.

stacking variables. See *crossing variables*.

standardized data. Standardized data has a mean of zero and a standard deviation of one. You can standardize data $\mathbf{x}$ by taking $(\mathbf{x} - \overline{\mathbf{x}})/\sigma$, where σ is the standard deviation of the data.

stopping rules. Stopping rules for hierarchical cluster analysis are used to determine the number of clusters. A stopping-rule value (also called an index) is computed for each cluster solution, that is, at each level of the hierarchy in hierarchical cluster analysis. Also see *hierarchical clustering*.

stress. See *Kruskal stress* and *loss*.

structure. Structure, as in factor structure, is the correlations between the variables and the common factors after factor analysis. Structure matrices are available after factor analysis and LDA. Also see *factor analysis* and *linear discriminant analysis*.

supplementary rows or columns or **supplementary variables**. Supplementary rows or columns can be included in CA, and supplementary variables can be included in MCA. They do not affect the CA or MCA solution, but they are included in plots and tables with statistics of the corresponding row or column points. Also see *correspondence analysis* and *multiple correspondence analysis*.

SVD. See *singular value decomposition*.

target rotation. Target rotation minimizes the criterion

$$c(\mathbf{\Lambda}) = \frac{1}{2}\|\mathbf{\Lambda} - \mathbf{H}\|^2$$

for a given target matrix $\mathbf{H}$.

See *Crawford–Ferguson rotation* for a definition of $\mathbf{\Lambda}$.

taxonomy. Taxonomy is the study of the general principles of scientific classification. It also denotes classification, especially the classification of plants and animals according to their natural relationships. Cluster analysis is a tool used in creating a taxonomy and is synonymous with numerical taxonomy. Also see *cluster analysis*.

tetrachoric correlation. A tetrachoric correlation estimates the correlation coefficients of binary variables by assuming a latent bivariate normal distribution for each pair of variables, with a threshold model for manifest variables.

ties. After discriminant analysis, ties in classification occur when two or more posterior probabilities are equal for an observation. They are most common with KNN discriminant analysis.

total inertia or **total principal inertia**. The total (principal) inertia in CA and MCA is the sum of the principal inertias. In CA, total inertia is the Pearson χ^2/n. In CA, the principal inertias are the singular values; in MCA the principal inertias are the eigenvalues. Also see *correspondence analysis* and *multiple correspondence analysis*.

uniqueness. In factor analysis, the uniqueness is the percentage of a variable's variance that is not explained by the common factors. It is also "1 − communality". Also see *communality*.

unrestricted transformation. An unrestricted transformation is a Procrustes transformation that allows the data to be transformed, not just by orthogonal and oblique rotations, but by all conformable regular matrices. This is equivalent to a multivariate regression. Also see *Procrustes transformation* and *multivariate regression*.

varimax rotation. Varimax rotation maximizes the variance of the squared loadings within the columns of the matrix. It is an orthogonal rotation equivalent to oblimin with $\gamma = 1$ or to the Crawford–Ferguson family with $\kappa = 1/p$, where p is the number of rows of the matrix to be rotated. Also see *orthogonal rotation*, *oblimin rotation*, and *Crawford–Ferguson rotation*.

Ward's linkage clustering. Ward's-linkage clustering is a hierarchical clustering method that joins the two groups resulting in the minimum increase in the error sum of squares.

weighted-average linkage clustering. Weighted-average linkage clustering is a hierarchical clustering method that uses the weighted average similarity or dissimilarity of the two groups as the measure between the two groups.

Wilks's lambda. Wilks's lambda is a test statistic for the hypothesis test $H_0 : \boldsymbol{\mu}_1 = \boldsymbol{\mu}_2 = \cdots = \boldsymbol{\mu}_k$ based on the eigenvalues $\lambda_1, \ldots, \lambda_s$ of $\mathbf{E}^{-1}\mathbf{H}$. It is defined as

$$\Lambda = \frac{|\mathbf{E}|}{|\mathbf{E} + \mathbf{H}|} = \prod_{i=1}^{s} \frac{1}{1 + \lambda_i}$$

where $\mathbf{H}$ is the between matrix and $\mathbf{E}$ is the within matrix. See *between matrix*.

Wishart distribution. The Wishart distribution is a family of probability distributions for nonnegative-definite matrix-valued random variables ("random matrices"). These distributions are of great importance in the estimation of covariance matrices in multivariate statistics.

within matrix. See *between matrix*.

References

Bartlett, M. S. 1937. The statistical conception of mental factors. *British Journal of Psychology* 28: 97–104.

——. 1938. Methods of estimating mental factors. *Nature, London* 141: 609–610.

Bellman, R. E. 1961. *Adaptive Control Processes*. Princeton, NJ: Princeton University Press.

Bentler, P. M. 1977. Factor simplicity index and transformations. *Psychometrika* 42: 277–295.

Comrey, A. L. 1967. Tandem criteria for analytic rotation in factor analysis. *Psychometrika* 32: 277–295.

Cox, T. F., and M. A. A. Cox. 2001. *Multidimensional Scaling*. 2nd ed. Boca Raton, FL: Chapman & Hall/CRC.

Crawford, C. B., and G. A. Ferguson. 1970. A general rotation criterion and its use in orthogonal rotation. *Psychometrika* 35: 321–332.

Fisher, R. A. 1936. The use of multiple measurements in taxonomic problems. *Annals of Eugenics* 7: 179–188.

Hendrickson, A. E., and P. O. White. 1964. Promax: A quick method for rotation to oblique simple structure. *British Journal of Statistical Psychology* 17: 65–70.

Jennrich, R. I. 1979. Admissible values of γ in direct oblimin rotation. *Psychometrika* 44: 173–177.

——. 2004. Rotation to simple loadings using component loss functions: The orthogonal case. *Psychometrika* 69: 257–273.

Kaiser, H. F. 1974. An index of factor simplicity. *Psychometrika* 39: 31–36.

Kruskal, J. B. 1964. Multidimensional scaling by optimizing goodness of fit to a nonmetric hypothesis. *Psychometrika* 29: 1–27.

Mahalanobis, P. C. 1936. On the generalized distance in statistics. *National Institute of Science of India* 12: 49–55.

Sammon, J. W., Jr. 1969. A nonlinear mapping for data structure analysis. *IEEE Transactions on Computers* 18: 401–409.

Thomson, G. H. 1951. *The Factorial Analysis of Human Ability*. London: University of London Press.